Low-Order Wavefront Sensing for Future Space-Based Coronagraphic Missions with a Sparse Aperture Mask

Hari B. Subedi

A Dissertation
Presented to the Faculty of Princeton University in Candidacy for the Degree of Doctor of Philosophy

Recommended for Acceptance by the Department of Mechanical and Aerospace Engineering Adviser: N. Jeremy Kasdin

June 2018
© Copyright by Hari B. Subedi, 2018.

All Rights Reserved.
Abstract

In the quest for extra-terrestrial life, we have landed on the moon and sent rovers to other planets in the Solar System. The discovery of the first Jupiter-sized exoplanet, two decades ago, has led to an exponential increase in exoplanet research and public interest in existence of life outside our Solar System. Since then, indirect techniques have led to the discovery of hundreds of exoplanets but are limited in their scope of exoplanet characterization. Direct imaging, however, will provide the capability to thoroughly characterize an exoplanet’s atmosphere, composition, temperature, size, and orbit. This would help scientists look for biomarkers of life. Diffraction due to the finite size of telescopes, $10^{10}$ difference in brightness between the host star and earth-like planets, and a small angular separation between them pose a huge challenge to direct imaging. The state of the art coronagraphs can create dark holes at smaller inner working angles, making it possible to image dim planets with smaller orbits. The contrast degradation caused by optical aberrations, however, limits the coronagraphs from achieving their theoretical limits. Robust estimation and control algorithms that can estimate and control these aberrations are being tested in various high-contrast imaging labs. However, to achieve the required contrast, it is also essential to estimate and control the dynamic aberrations that degrade the coronagraphic performance. This thesis focuses on a novel technique that estimates these dynamic aberrations. This technique utilizes the starlight rejected by the coronagraph and a sparse aperture mask (SAM) to infer the dynamic aberrations. In this work, we will present the working principle of the SAM wavefront sensor, detail the SAM optimization, and compare its performance with other existing wavefront sensors. We will also present an adaptive estimation and control technique that can be used with a Kalman filter for line-of-sight (LoS) jitter estimation and control in the absence of the jitter frequency data. The work presented in this thesis would help designers pick low-order wavefront sensing and control hardware and software for future coronagraphic missions.
Acknowledgements

I would like to start by thanking my advisor, Professor Jeremy Kasdin. Thank you for accepting me as your graduate student despite my minimal optical background. Your guidance and support over the last five years have been monumental. I am truly awed by your enthusiasm, commitment, and contribution to the high-contrast imaging field. Thank you for all the advice, words of encouragement, and the support you’ve provided throughout the years. I am grateful for your patience in editing my work and helping me grow as a researcher. I have learned a lot from you and thank you for your role in my academic life.

I also thank my committee members and readers, Professors Robert Stengel, Mike Littman, Robert Vanderbei, and Craig Arnold for their valuable time and commitment. I greatly appreciate your continued advice and suggestions. Most of the optimal estimation and control work I have been able to do can be contributed to Professor Stengel’s class and frequent discussions with him. I am grateful for Professor Vanderbei’s expertise in optimization and his willingness to guide me. I enjoyed taking classes from professor Littman and had a fun time teaching an undergraduate class for him. Professor Arnold’s encouragement in learning optics has been invaluable. I would also like to thank Jill Ray for the administrative assistance during the last five years.

Post-docs in our lab have been extremely supportive during my first few years. I have been fortunate to have overlapped with Dr. Tyler Groff’s and Dr. Neil Zimmerman’s time at Princeton. I am excited to continue working with both of you at NASA Goddard Space Flight Center. I have learned a great deal about almost everything from Dr. A J Eldorado Riggs. Thank you for being there for me. I would like to thank Dr. Jessica Gersch-Range, Mike Galvin, Katie Cavanagh, Dr. Anthony Harness, Dr. Dan Sirbu, Dr. Elizabeth Young, Aaron Lemmer, and Dr. Yungjong Kim for their suggestions. To the newer members in our group, it’s been fun having you around.
He Sun, I am happy that you joined our group. I am truly impressed by the progress you have made. I will miss our wide-ranging conversations. One day we will figure out the true price of freedom. Mia Hu, it was nice knowing you too. You will do great. I am glad for the time I have had with visiting students, Andreas Rousing and Peter Varnai. We sure did have fun.

Outside of our lab, I would like to thank all my MAE friends. I will always remember the lunch table conversations. I would especially like to thank Yibin Zhang and Peter Landgren for their continued friendship. Yibin, thanks for all the editing and mischiefs. I would like to thank the MAE intramural soccer team for the fun time we had. It was a great feeling winning the tournament.

For the Fall of 2018, I interned at NASA Goddard Space Flight Center, where I will be working after graduation. I would like to thank the Optics branch and the Wavefront Sensing Group for the warm welcome and support. Dr. Matthew Bolcar was extremely accommodating in helping me adjust at Goddard.

I would like to acknowledge the funding provided by the NASA Technology Development for Exoplanet Missions (TDEM) under Grant No. NNX09AB96G and the Jet Propulsion Laboratory of the California Institute of Technology. For the last three years, my research has been funded by National Science Foundation Graduate Research Fellowship.

I am thankful to Dr. Fang Shi from JPL for his advice. I am also grateful to Professor Chilok Chan from the University of Arizona for being an awesome mentor during my undergraduate years. I am also thankful to all my professors at University of Arizona and Pima Community College. I would also like to thank my teachers from Little Angels School and Blooming Lotus English School. It has been quite a journey.

I am thankful to my family for a lifetime of love and support. I would like to thank my parents, brothers, and sisters-in-law. I am grateful to my childhood friends
who are still in touch. I am always looking forward to our next adventure. Thank you *Hawa Daari* team for the friendship that transcends all barriers.

I dedicate this thesis to my wife-to-be, Deepa. I have no words to describe your presence and support through thick and thin. Without you, this would not have been possible. Thank you.

This dissertation carries T#3353 in the records of the Department of Mechanical and Aerospace Engineering.
“Look again at that dot. That’s here. That’s home. That’s us. On it everyone you
love, everyone you know, everyone you ever heard of, every human being who ever
was, lived out their lives. The aggregate of our joy and suffering, thousands of
confident religions, ideologies, and economic doctrines, every hunter and forager,
every hero and coward, every creator and destroyer of civilization, every king and
peasant, every young couple in love, every mother and father, hopeful child,
inventor and explorer, every teacher of morals, every corrupt politician, every
“superstar,” every “supreme leader,” every saint and sinner in the history of our
species lived there - on a mote of dust suspended in a sunbeam. The Earth is a very
small stage in a vast cosmic arena. Think of the rivers of blood spilled by all those
generals and emperors, so that, in glory and triumph, they could become the
momentary masters of a fraction of a dot. Think of the endless cruelties visited by
the inhabitants of one corner of this pixel on the scarcely distinguishable
inhabitants of some other corner, how frequent their misunderstandings, how eager
they are to kill one another, how fervent their hatreds.”

- Carl Sagan
To my family.
Contents

Abstract ............................................................... iii
Acknowledgements ....................................................... iv
List of Tables ............................................................... xiii
List of Figures ............................................................. xiv
List of Symbols ............................................................ xviii

1 Introduction ......................................................... 1
  1.1 Science Motivation .................................................. 1
  1.2 Direct Imaging ...................................................... 5
    1.2.1 Coronagraphs .................................................. 6
      1.2.1.1 Shaped Pupil Coronagraph ............................. 7
      1.2.1.2 Apodized Pupil Lyot Coronagraph ..................... 9
      1.2.1.3 Vector Vortex Coronagraph ........................... 10
      1.2.1.4 Band-limited Coronagraph ........................... 11
  1.3 Space Missions with Coronagraphs ............................. 12
    1.3.1 The Wide-Field Infrared Survey Telescope (WFIRST) .... 13
    1.3.2 Habitable Exoplanet Imaging Mission (HabEx) ............ 13
    1.3.3 The Large UV/Optical/IR Surveyor (LUVOIR) ............ 14
  1.4 Deformable mirror ................................................. 15
  1.5 Fourier Optics .................................................... 17
  1.6 Wavefront Sensing and Control .................................. 22
1.6.1  Zernike Polynomials ........................................... 23
1.6.2  Brief Introduction to AO ...................................... 26
1.6.3  Brief Introduction to FPWC .................................... 27
1.6.4  Brief Introduction to Low-Order Wavefront Aberration and LOWFS 28
1.7  Thesis outline .......................................................... 31

2  Detailed description of low-order wavefront sensors 33
2.1  Sparse Aperture Mask Wavefront Sensor .......................... 33
  2.1.1  Focal Plane mask (FPM) ...................................... 34
  2.1.2  Sparse Aperture Mask .......................................... 35
  2.1.3  SAM Sensing Equation ......................................... 36
  2.1.4  SAM initial design .............................................. 41
2.2  Zernike Wavefront Sensor ........................................ 42
  2.2.1  ZWFS algorithm ................................................. 43
2.3  Coronagraphic Low-Order Wavefront Sensor ........................ 48
2.4  Lyot-based Low-Order Wavefront Sensor .......................... 50

3  SAM WFS initial simulation verification and experiments 53
3.1  Monochromatic simulations ....................................... 53
  3.1.1  Fit Analysis .................................................... 55
  3.1.2  Phase reconstruction tests .................................... 57
  3.1.3  Response to a rapid pointing error ............................ 60
3.2  Intermediate Summary ............................................. 63

4  Experimental Results 64
4.1  Testbed design ..................................................... 64
  4.1.1  Masks ......................................................... 64
  4.1.2  DM .......................................................... 65
  4.1.3  Camera ...................................................... 66
List of Tables

2.1 Peak Differential Intensity of the SAM WFS ........................................ 42

3.1 Physical Parameters for Propagation .................................................... 55
3.2 Mean Residual Errors ........................................................................... 57
3.3 Tilt Error Statistics ............................................................................. 63

4.1 Camera Specifications ........................................................................... 67

6.1 Nominal Parameters for the SPC-integrated Sensors ......................... 84
6.2 Nominal Parameters for the VVC-integrated Sensors ......................... 84
6.3 Nominal Parameters for the BLC-integrated Sensors ......................... 84
6.4 Performance of the SPC-integrated Sensors .................................... 86
6.5 Performance of the VVC-integrated Sensors .................................... 87
6.6 Performance of the BLC-integrated Sensors .................................... 88
List of Figures

1.1 Telescope Diffraction ........................................... 3
1.2 Exoplanet Detection Methods ................................... 4
1.3 The Lyot Coronagraph Illustration .............................. 6
1.4 An SPC Image ..................................................... 7
1.5 Direct Imaging System with an SPC ............................. 8
1.6 APLC/SPLC System .............................................. 10
1.7 VVC Coronagraph ............................................... 10
1.8 Band-limited Coronagraph ..................................... 12
1.9 Deformable Mirror .............................................. 16
1.10 Free Space Light Propagation .................................. 18
1.11 Light Propagation in an Imaging System ..................... 21
1.12 Contrast Degradation .......................................... 23
1.13 Zernike Polynomials ........................................... 24
1.14 Complete Wavefront Estimation and Control Scheme ....... 25
1.15 Schematic of an AO System ................................... 26

2.1 The SAM WFS with an SPC .................................... 34
2.2 Bowtie Focal Plane Mask ....................................... 35
2.3 The SAM WFS for Phase Mask Coronagraphs ................. 40
2.4 SAM Images for Different Modes .............................. 41
2.5 The Zernike Wavefront Sensor ................................. 43
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>Performance of the BLC-integrated Sensors</td>
<td>88</td>
</tr>
<tr>
<td>6.5</td>
<td>WFS Noise Performance with an SPC</td>
<td>90</td>
</tr>
<tr>
<td>6.6</td>
<td>WFS Noise Performance with a VVC</td>
<td>91</td>
</tr>
<tr>
<td>6.7</td>
<td>WFS Noise Performance with a BLC</td>
<td>91</td>
</tr>
<tr>
<td>6.8</td>
<td>Readout Noise Effects</td>
<td>92</td>
</tr>
<tr>
<td>7.1</td>
<td>Comparison of Different Basis Sets</td>
<td>95</td>
</tr>
<tr>
<td>7.2</td>
<td>Estimation Error - New Basis Set</td>
<td>96</td>
</tr>
<tr>
<td>7.3</td>
<td>Singular Values of the SAM WFS Response</td>
<td>97</td>
</tr>
<tr>
<td>7.4</td>
<td>The SAM WFS Estimates with the DM Influence Function</td>
<td>98</td>
</tr>
<tr>
<td>8.1</td>
<td>Telescope Under Disturbance</td>
<td>103</td>
</tr>
<tr>
<td>8.2</td>
<td>Block Diagram for LoS Pointing Jitter</td>
<td>105</td>
</tr>
<tr>
<td>8.3</td>
<td>Error in LoS Disturbance Estimate</td>
<td>107</td>
</tr>
<tr>
<td>8.4</td>
<td>Block Diagram for Adaptive Estimation of LoS Pointing Jitter</td>
<td>108</td>
</tr>
<tr>
<td>8.5</td>
<td>LoS Pointing Jitter Estimate</td>
<td>111</td>
</tr>
<tr>
<td>8.6</td>
<td>LoS Pointing Jitter Parameter Estimate</td>
<td>112</td>
</tr>
<tr>
<td>8.7</td>
<td>LoS Disturbance with Changing Frequency</td>
<td>113</td>
</tr>
<tr>
<td>8.8</td>
<td>LOWFS Estimation and Control</td>
<td>115</td>
</tr>
<tr>
<td>9.1</td>
<td>The PHCIL lab setup</td>
<td>122</td>
</tr>
</tbody>
</table>
## List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$</td>
<td>Complex-valued electric field</td>
<td>17</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Wavelength</td>
<td>17</td>
</tr>
<tr>
<td>$i$</td>
<td>Imaginary number</td>
<td>17</td>
</tr>
<tr>
<td>$\mathcal{F}_r{\cdot}$</td>
<td>Fresnel propagator</td>
<td>17</td>
</tr>
<tr>
<td>$\mathcal{F}{\cdot}$</td>
<td>Fourier transform operator</td>
<td>17</td>
</tr>
<tr>
<td>$f$</td>
<td>Focal length</td>
<td>20</td>
</tr>
<tr>
<td>$\mathcal{L}$</td>
<td>Lens operator</td>
<td>20</td>
</tr>
<tr>
<td>$Z_m^n$</td>
<td>Zernike polynomial</td>
<td>23</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Phase aberration</td>
<td>36</td>
</tr>
<tr>
<td>$\mathcal{M}$</td>
<td>Sparse aperture mask</td>
<td>37</td>
</tr>
<tr>
<td>$I$</td>
<td>Intensity</td>
<td>38</td>
</tr>
<tr>
<td>$n$</td>
<td>Photon and readout noise</td>
<td>38</td>
</tr>
<tr>
<td>$Re$</td>
<td>Real part of a complex number</td>
<td>38</td>
</tr>
<tr>
<td>$H$</td>
<td>Response matrix</td>
<td>38</td>
</tr>
<tr>
<td>⊛</td>
<td>Convolution Operator</td>
<td>45</td>
</tr>
<tr>
<td>$f_z$</td>
<td>f-number</td>
<td>54</td>
</tr>
<tr>
<td>⊙</td>
<td>Hadamard product</td>
<td>73</td>
</tr>
<tr>
<td>$\mathbb{R}$</td>
<td>Real number</td>
<td>73</td>
</tr>
<tr>
<td>$r_n$</td>
<td>Readout noise</td>
<td>75</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Condition number</td>
<td>76</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>$h$</td>
<td>Planck’s constant</td>
<td>89</td>
</tr>
<tr>
<td>$\Phi_{\text{star}}$</td>
<td>Photon flux</td>
<td>89</td>
</tr>
<tr>
<td>$Q_e$</td>
<td>Quantum efficiency</td>
<td>89</td>
</tr>
<tr>
<td>$G$</td>
<td>Influence function matrix</td>
<td>93</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>DM voltage gain error</td>
<td>93</td>
</tr>
<tr>
<td>$K$</td>
<td>Kalman gain matrix</td>
<td>101</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The overall theme of this thesis is numerical and experimental validation of a new low-order wavefront sensor (LOWFS) that uses a sparse aperture mask (SAM). In this chapter, we will describe the science motivation behind the thesis, derive the fundamental Fourier propagation equations, and explain the basics of the low-order wavefront sensing problem. We will use the formulations developed in this chapter to prove the efficacy of the SAM wavefront sensor (SAM WFS) and compare its performance with other LOWFS techniques in use or proposed in the literature.

1.1 Science Motivation

The possibility of life beyond Earth has created excitement among the scientific community as well as the public. Major extrasolar planet (exoplanet) exploration programs have been initiated to discover and characterize the planetary systems in order to answer one of the everlasting questions: are we alone? The discoveries from these programs impact the understanding of our own solar system and provide a step forward towards the discovery of life beyond Earth. The field of exoplanet detection and characterization emerged quickly with the discoveries of the first exoplanets in 1992 and 1995. Wolszczan and Frail discovered three exoplanets orbiting a pulsar [1]
in 1992 and Mayor and Queloz detected the first exoplanet, 51 Pegasi b, orbiting the main sequence star, 51 Pegasi, in 1995 [2]. The discovery of 51 Pegasi b accelerated the search for more exoplanets. There have been 3,584 confirmed exoplanet discoveries, with 4,496 possible candidates based on the Kepler mission. The field of exoplanet discovery continues to evolve with new technological advancements. With the discoveries of new planets, we are closer to understanding the process of planet formation and the possibility of existence of life in those planets.

When looking for the possibility of life on other planets, we look for biomarkers that make life possible on our own planet. Most important is the existence of liquid water and oxygen. The planet must orbit at a distance from the star where liquid water could exist. If the planet is too close to the star, the water would boil off, and if it is too far away it would freeze. This region around a main-sequence star where water could exist in its liquid form is called a “habitable zone” (HZ). Since the host stars have a range of surface temperatures, the HZ ranges in distance from the host star: for a star hotter than our sun, the HZ would be farther away from the star as compared to the HZ around the Sun and closer from a cooler star. The small angular separation between the planets and the stars as seen from the Earth, the large difference in brightness between the bright star and dimmer planet, scattering and diffraction of light due to the atmospheric effects on the ground, and optical aberrations make it difficult to directly image exoplanets to obtain their spectra. It can be observed from Fig. 1.1 that the issue is not resolution but diffraction. In an ideal scenario, a 2 meter class telescope is capable of resolving two objects located 10 parsecs away from the Earth and 1 astronomical units (AU) away from each other. Figure 1.2 shows the exoplanets detected by using different methods. The majority of the exoplanets discovered so far have been detected using indirect methods — identified through their effects on the host star. The radial velocity method is one of the most used techniques to find exoplanets. Planets go around their host stars due to the gravitational pull
Figure 1.1: Cross section of a normalized intensity profile obtained by using a circular aperture with a 2 meter diameter (left) and a 10 meter diameter (right). Two sources, 1 AU in separation, located at 10 parsecs from Earth can be resolved by either telescopes if they have comparable brightness. The blue line represents the on-axis source and the red-dotted line is the off-axis source of equal brightness. On the other hand, if the off-axis source (green line) is billions of times dimmer than the on-axis source, the diffracted light from the on-axis source swamps the image making the off-axis source undetectable. Image obtained from Tyler Groff’s PhD thesis [3].

of the stars. Due to the laws of physics, planets also pull the stars toward them making the star wobble. When the star wobbles, it accelerates towards or away from the observer at a small velocity. Due to the Doppler effect, scientists can measure this velocity and deduce the presence of exoplanets. When the star moves away from the Earth, it looks redder and it seems bluer when it moves towards the Earth. The exoplanet 51 Pegasi b was detected using this method.

The Transit Method is another widely used indirect method to find exoplanets. In this method, scientists look at the decrease or dimming in light of the star. A planet revolves around its host star with a certain period of revolution. During its revolution, the planet comes in between the Earth and its star and blocks the light coming to the Earth from the star. When it happens, there is a dip in light coming from the star. If there is a planet orbiting a star, the light from the star decreases in brightness for
a period of time and then returns to the original brightness. To rule out the false positives in this dip caused by any object crossing between the star and the telescope, several occurrences of this dip in brightness are required to be recorded to show the periodic transit of an orbiting planet around the star. The Kepler space telescope uses the transit method to detect the exoplanets [4, 5]. Future spacecraft such as NASA’s Transiting Exoplanet Survey Satellite (TESS) and the James Webb Space Telescope (JWST) will use this technique to find more exoplanets. Other indirect methods include gravitational microlensing [6, 7], pulsar timing [8], and astrometry [9].

The indirect methods are limited on their scope in determining all the param-
eter space of exoplanet characterization. The transit method is limited to systems whose orbital plane are edge-on to our line-of-sight (LoS). The radial velocity is sensitive to the velocity component along the line-of-sight of the telescope, so any orbital inclination is unobservable and as a result, the planet’s actual mass and orbit cannot be determined. The indirect methods are incapable of spectrally characterizing exoplanets, hence the need to directly image exoplanets.

1.2 Direct Imaging

Direct imaging will provide the capability to thoroughly characterize an exoplanet’s atmosphere, composition, temperature, size, and orbit. The direct imaging of exoplanets has many technical challenges and difficulties: scattering and diffraction of light due to the atmospheric effects on the ground and optical aberrations in space, the small angular separation between the planets and the stars as seen from the Earth, and the large difference in brightness between the bright star and dimmer planet. To avoid the effects of dynamics of Earth’s atmosphere and small separation angles, most designs for high contrast have been space-based. Nevertheless, these problems do exist for the space-based telescopes too. As the brightness of Earth-like planets in the visible spectrum is $10^{-10}$ times that of their host stars, techniques have been developed to obtain high contrast levels in regions in the image where the planets are visible. The two major approaches used to create the necessary high contrast for exoplanet detection in space use either an occulter external to the telescope or a coronagraph internal to the telescope. The idea of an external starshade was proposed by Lyman Spitzer at Princeton University during the early sixties and has been highly researched recently [10, 11, 12, 13]. The main concept is to keep the starlight from ever entering the telescope and smearing everything in the image. While approximately the same number of exoplanets can be discovered and characterized using
both methods, each has its own challenges. An external occulter requires the precision engineering of the large starshade and controlling the separation of the occulter spacecraft and telescope over tens of thousands of kilometers [14]. The use of internal coronagraph poses optical challenges from the light diffracted and scattered within the telescope. This work focuses on solving the challenges of coronagraphic imaging.

1.2.1 Coronagraphs

A coronagraph is an optical device which blocks the starlight from the host star while permitting light from surrounding planets to pass through with minimum disturbance. The coronagraph was invented by the French astronomer Bernard Lyot in 1939 to enable astronomers to observe the outer atmosphere (the corona) of the sun, which otherwise was only visible during solar eclipses. Figure 1.3 is a schematic of a classical Lyot coronagraph. The light from the telescope is focused using a focusing optic. At the focal plane, the on-axis starlight is blocked by a focal plane mask (FPM), whereas, the off-axis planet light is not blocked. The light is then collimated using another powered optic (a lens or a concave mirror). Another mask, which is smaller in diameter than the telescope exit pupil, blocks most of the high-spatial frequency diffracted starlight. This mask is called a Lyot stop. A second powered optic then

Figure 1.3: Lyot coronagraph for exoplanet detection. A mask placed at the focal plane blocks the on-axis starlight, while the off-axis planet light passes through. Another mask, Lyot stop, placed at a relay pupil plane blocks the diffracted starlight making it possible to image the exoplanet. Image credit: Matthew Kenworthy.
reimages the pupil to form the final sky image on the detector, where the exoplanet can be seen. Since the original coronagraph, various coronagraphs have been proposed. In designing a coronagraph, the following metrics are employed: 1) Contrast, defined as the ratio of the image plane intensity at the planet location to the peak intensity of the point spread function (PSF). 2) Inner working angle (IWA), the smallest apparent separation between host and companion sources at which the companion is detectable. 3) Outer working angle (OWA), the largest such separation. 4) Airy Throughput, the amount of energy that falls into the main lobe of the PSF relative to the total energy conveyed through a fully open aperture. In the following subsections, I will briefly describe different coronagraphs used in this thesis.

1.2.1.1 Shaped Pupil Coronagraph

In contrast to a classical Lyot coronagraph described above, a shaped pupil coronagraph (SPC) modifies the pupil to create a PSF with the needed contrast at the planet location [15, 16]. An SPC, shown in Fig. 1.4a, is a binary apodizer that changes the PSF of the telescope to introduce regions of high contrast called dark holes, where the faint exoplanets can be imaged. Figure 1.5 is a sketch of an imaging system that has an SPC. The binary SPC, placed at the exit pupil, apodizes the light coming from the telescope to create the PSF shown. Due to the limited dynamic range of

![Figure 1.4: a) A shaped pupil coronagraph. The white region is open, i.e., 1 and the black region is opaque i.e., 0. b) PSF of the SPC in log scale. c) Re-imaged focal plane after the FPM blocks the rejected starlight.](image-url)
the imaging camera, the starlight rejected by the coronagraph is blocked by a focal plane mask (FPM). The light from the unblocked dark holes is re-imaged to an image plane where the detector is placed (Fig. 1.4c). The SPC uses the heritage from the two-dimensional apodization used to achieve contrast in the image plane suggested by David Slepian [18]. The prolate spheroidal function used by this apodizer concentrates energy in a single, central lobe at the image plane [17, 18]. The non-binary apodized pupils suggested by Slepian are difficult to manufacture. The SPC with binary openings, i.e., either zero or one, are easy to manufacture and have robust design. It is for these reasons that our group at Princeton High Contrast Imaging Lab (PHCIL) has been working on SPC designs since the early 2000s. David Spergel developed an SPC by using a Gaussian profile to achieve a region of high-contrast along one axis [19]. Jeremy Kasdin improved on Spergel’s and Slepian’s idea by converting the prolate spheroidal apodization into a binary mask [15, 16]. Robert Vanderbei has since used a linear optimization to design better SPCs [20, 21]. Alexis Carlotti,
A J Riggs, and Neil Zimmerman have further explored the optimization procedure [22, 23, 24, 25]. The goal of the optimization is to maximize the open area of the coronagraph while creating the desired contrast in the dark holes. This optimization problem is formulated as

\[
\text{maximize} \int_0^{1/2} A(r) 2\pi r dr \\
\text{s.t.} -10^{-5} E(0) \leq E(\rho) \leq 10^5 E(0), \quad \rho_{iwa} \leq \rho \leq \rho_{owa} \\
0 \leq A(r) \leq 1, \quad 0 \leq r \leq \frac{1}{2},
\]

where \(A(r)\) is the apodization function, \(E(\rho)\) is the image-plane electric field, \(\rho_{iwa}\) is the inner working angle (IWA), \(\rho_{owa}\) is the outer working angle (OWA), \(r\) is the radial coordinate in the pupil plane and \(\rho\) in the image plane. In this formulation, the Airy throughput is the ratio of open area of the shaped-pupil to that of an open circular aperture.

1.2.1.2 Apodized Pupil Lyot Coronagraph

The apodized pupil Lyot coronagraph [26, 27, 28, 29] uses a prolate apodization derived from the eigenvalue problem associated with the Lyot-style coronagraphic propagation for a given mask size. Since the eigenvalue of the problem gives only a single degree of freedom in the definition of the apodizer, the raw performance of the APLC is limited to \(10^{-7} - 10^{-8}\) contrast [30]. The recent designs of the APLC have SPC like optimization to design Lyot-style coronagraph with improved performance in terms of contrast and IWA [24]. This hybrid APLC is also referred to as a shaped pupil Lyot coronagraph (SPLC). Figure 1.6 shows an imaging system with an SPLC.
1.2.1.3 Vector Vortex Coronagraph

The vector vortex coronagraph (VVC) uses a phase mask, at the focal plane, that creates an optical delay (a “vortex” phase profile) of the form $e^{il\theta}$, where $l$ is the topological charge (the number of times the phase accumulates $2\pi$ radians along a closed path surrounding the center) that determines the polarization spin rate and the subsequent height of the mask phase after a full $2\pi$ rotation and $\theta$ is the azimuthal coordinate [31, 32, 33]. A vortex of topological charge $l$ means that upon a complete rotation about the center, it has undergone a total $2\pi l$ phase shift. Figure 1.7 shows the working principle of a charge 2 VVC ($l = 2$). In this mask, the polarization is
rotated locally (Fig. 1.7a and 1.7b) so that a rotation around the mask creates a phase shift of $4\pi$ (Fig. 1.7c). The electric field at the Lyot plane is the product of the Fourier transform of the electric field at the image plane and the mask phase $e^{i2\theta}$. This phase mask diffracts the residual starlight intensity downstream outside the geometrical pupil at the Lyot plane as seen in Fig. 1.7d. At the Lyot plane, a Lyot stop is placed to block this rejected starlight and the unobstructed light is imaged to a detector. For a non-ideal aperture, one with obscuration and spider arms, the residual starlight diffracted by the phase mask is not completely outside of the pupil diameter, therefore, the Lyot stop is undersized (as compared to the exit pupil diameter) to block the residual starlight.

1.2.1.4 Band-limited Coronagraph

The band-limited coronagraph (BLC) has a similar working principle as the classical Lyot type coronagraph. The only difference is that the occulting mask is modified so as to place all of the diffracted light from the on-axis source in the second pupil plane with narrow zones near the sharp edges of the entrance pupil as shown in Fig. 1.8 [34]. Then the Lyot stop would block all the straylight in this plane. Since the Fourier transform of the occulting mask is zero everywhere except the narrow range, it is called a band-limited mask. Figure 1.8 shows the working mechanism of the BLC in one-dimension. In this figure, the operation of the coronagraph is visualized entirely in the pupil plane. The entrance pupil is a tophat function represented as $(A(u))$, where $u$ is the pupil plane coordinate. The band-limited mask is placed at the focal plane and a corresponding Lyot mask at the Lyot plane. The final field has all the on-axis light blocked out.

Various ground-based telescopes are equipped with different families of coronagraphs that are capable of directly imaging self-luminous gas giants, but to achieve the contrast required to image cool rocky planets, it is essential to avoid the effects
Figure 1.8: An ideal band-limited coronagraph (image obtained from Kuchner et al.[34]). a) Aperture represented by a top hat, (b) the conjugate of the mask, (c) the second pupil field, (d) a Lyot stop that can block all the on-axis light, and (e) zero final field.

of Earth’s atmosphere. In the following section, I will briefly describe future space missions that will have coronagraphs.

1.3 Space Missions with Coronagraphs

The Hubble Space Telescope (HST) has a coronagraphic observing mode and the James Webb Space Telescope (JWST) will also have a coronagraph, but in this section I will briefly talk about space missions that will have a high contrast imaging coronagraph with active wavefront control.
1.3.1 The Wide-Field Infrared Survey Telescope (WFIRST)

The WFIRST is a 2.4-meter telescope with a launch target of mid-2020s. This obscured two-mirror system with a 2.4-m aperture and on-axis secondary mirrors has two different instrument volumes containing the wide field and coronagraph instruments (CGI) [35]. The goal of the wide field instrument is to perform dark energy, exoplanet microlensing, and near infrared (NIR) surveys. The CGI with an imaging and spectroscopic mode will support the exoplanet high contrast imaging and spectroscopic characterization of exoplanets and debris disks around nearby stars. WFIRST will carry the first active coronagraph in space and provide technological heritage for upcoming missions. WFIRST has two coronagraphic modes: 1) the hybrid Lyot coronagraph (HLC) for exoplanet detection and 2) SPC to demonstrate coronagraphic performance on-orbit and image zodiacal dust and debris disks. The spectroscopy mode will have an integral field spectrograph (IFS). This technology demonstration in space is critical for future missions for direct imaging of exoplanets.

1.3.2 Habitable Exoplanet Imaging Mission (HabEx)

The Habitable Exoplanet Imaging Mission (HabEx) is a mission concept exclusively planned to directly image Earth twins around Sun-like stars [36, 37]. By measuring the spectra of these planets, it will search for biomarkers of life. Since HabEx will have high UV, optical, and NIR resolution, it will also support a broad range of galactic and extragalactic astrophysics. The data collected by HabEx will help us understand the life cycle and deaths of the most massive stars, which ultimately supply the elements that are needed to support life as we know it. To directly image exoplanets, HabEx will be equipped with a VVC and also a starshade [37].
1.3.3 The Large UV/Optical/IR Surveyor (LUVOIR)

The Large UV/Optical/IR Surveyor (LUVOIR) is a multiwavelength serviceable telescope concept designed to enable a broad range of astrophysics [38]. One of the science goals of the mission is to search for habitable conditions and signs of life on dozens of planets in the HZ. The LUVOIR telescope will be able to gather valuable information about the atmospheres of these exoplanets. By analyzing the data obtained from directly imaging exoplanets, the LUVOIR would make it possible to develop and test theories of planetary formation and evolution. In addition to imaging exoplanets, it will provide better resolution imaging of planets in our own solar system and help monitor the atmospheric dynamics in Jupiter, Saturn, Uranus, and Neptune. LUVOIR will also provide sensitive, high resolution imaging and spectroscopy of comets, asteroids, and moons within our solar system. The scope of LUVOIR is not limited to planets within and outside of our solar system—it will help us understand galaxy formation and evolution, cosmology and structure, and star formation and stellar evolution. LUVOIR is in its initial concept design phase. There are two concepts that are being analyzed for the mission. The first architecture (LUVOIR-A) is a 15 m telescope and operates in the band between 100 nm and 2.5 μm. For the LUVOIR-A design, the following instruments are being developed: an optical / near-infrared apodized pupil Lyot coronagraph capable of $10^{-10}$ contrast at inner working angles $\leq 4\lambda/D$; the LUVOIR UV Multi-object Spectrograph (LUMOS), which will provide low and medium-resolution UV (100 - 400 nm) multi-object imaging spectroscopy in addition to far-UV imaging; the High Definition Imager (HDI), a high-resolution wide-field-of-view NUV-Optical-IR imager; and a high-resolution UV spectro-polarimeter being contributed by Centre National d’Etudes Spatiales (CNES) [39]. The second architecture (LUVOIR-B) being considered is a 9 m telescope. Even though it will have the same range as the 15-m, the coronagraph instrument will likely have a narrower bandpass. The LUVOIR-B also has coronagraphic instruments in all channels. The
scientific goals of the coronagraph instrument are to measure the occurrence rate of biomarkers in the atmospheres of rocky planets orbiting in the HZ of their host stars and study the diversity of exo-planetary systems. LUVOIR will be have an APLC/SPLC and a VVC [40].

All of the coronagraphs, including the ones to be used by the above space missions, are sensitive to wavefront aberrations. A wavefront is defined as a surface associated with a propagating wave passing through all points that have the same phase. A small amplitude and phase aberration causes leakage of starlight into the dark holes that can mask the planet. These aberrations are caused by the atmospheric turbulence, manufacturing error of optics, misalignments, and vibrations for ground-based telescopes. Space-based telescopes avoid the aberrations due to atmospheric turbulence, but suffer from aberrations caused by manufacturing error of optics, misalignments, and vibrations. These aberrations must be estimated and compensated through the use of adaptive optics (AO) and a deformable mirror (DM).

### 1.4 Deformable mirror

A deformable mirror (DM) is an adaptive element with a controllable reflective surface shape. The DM can be deformed into the desired shape by applying appropriate voltage to each actuator beneath the surface. Different techniques can be used for actuation, but the common techniques either use a piezo-electric actuation or a micro-electro-mechanical (MEMS) actuation. Figure 1.9a is a schematic of a MEMS deformable mirror made by Boston Micromachines Corporation (BMC). We exclusively use the BMC MEMS DM in our lab. To create a surface shape, we need to know the DM influence function (shown in Fig. 1.9b), which describes the shape to which the mirror will deform when voltage is applied to an actuator, and the height response of an actuator to the applied voltage. Using the influence function, height
Figure 1.9: Illustration of a deformable mirror. a) MEMS DM schematics (image obtained from BMC website) and b) an actuator influence function of a 12x12 BMC DM.

vs. voltage information, and the linear superposition principle, the voltage required to get the desired shape is calculated. The mirror deformation, $\delta_i$, due to a DM command, $u_i$, applied to an actuator is

$$\delta_i = G_i \alpha_i u_i,$$

(1.4.1)

where $G_i$ is a vector that represents the normalized influence function of an actuator and $\alpha_i$ is the height-voltage relationship of the actuator. The total mirror deformation, $\delta$, when a voltage command, $u$, is applied to the DM is

$$\delta = \sum_{i=1}^{i=N} G_i \alpha_i u_i,$$

$$= G\alpha u,$$

(1.4.2)

where N is the number of actuators, $G$ is the influence function matrix, and $\alpha$ is a diagonal matrix that has height-voltage relationship. The phase created by this
deformation is
\[ \phi = \frac{4\pi}{\lambda} \delta, \]
\[ = G\alpha \frac{4\pi}{\lambda} u, \] (1.4.3)
\[ = G\beta u, \]
where \(\lambda\) is the wavelength and \(\beta = \frac{4\pi}{\lambda} \alpha\). Using the above equation the voltage command required to create a phase map, \(\phi\), is obtained as

\[ u = (G\beta)^+ \phi, \] (1.4.4)

where the superscript, +, represents the pseudo-inverse of a matrix. Key parameters of a DM are surface type, actuation technology, actuator pitch, and number of actuators. In order to estimate and control both the amplitude and the phase aberrations, two DMs are required.

Before delving into the wavefront estimation and control problem, we need to understand the propagation of light through an imaging system. The following section, using Fourier optics, describes the propagation of light between different planes.

\section{1.5 Fourier Optics}

Maxwell’s equations fully describe the propagation of electromagnetic radiation. Solving Maxwell’s equations is cumbersome and computationally expensive. Hence, we will take advantage of the simplifying approximations that we can make and arrive at a simple relationship between the electric fields at two planes.

Figure 1.10 shows two planes separated by a distance \(\Delta z\). By the Huygens-Fresnel principle, the propagated electric field at the second plane, \(E_1(u_1, v_1)\), can be found by treating each point of the initial wavefront, \(E_0(u_0, v_0)\), as an amplitude-scaled point source and summing the resultant spherical wavefronts. Using this principle, the
Figure 1.10: Coordinate systems for free space propagation of electric field between two planes.

The electric field at the second plane is

\[ E_1(u_1, v_1) = \int \int E_0(u_0, v_0) \frac{e^{i \frac{2\pi}{\lambda} r}}{i \lambda r} \cos \theta du_0 dv_0, \] (1.5.1)

where \( r = |\mathbf{r}| = \sqrt{(\Delta z)^2 + (u_1 - u_0)^2 + (v_1 - v_0)^2} \), \( \lambda \) is the wavelength of light, and \( \theta \) is the angle between the optical propagation axis and the vector (\( \mathbf{r} \)) from point \((u_0, v_0)\) to \((u_1, v_1)\). The limits of integration are from \(-\infty\) to \(+\infty\). Substituting \( \cos \theta = \frac{\Delta z}{r} \) in the above equation, we get

\[ E_1(u_1, v_1) = \frac{\Delta z}{i \lambda} \int \int E_0(u_0, v_0) \frac{e^{i \frac{2\pi}{\lambda} r}}{r^2} du_0 dv_0. \] (1.5.2)

Equation 1.5.2 is still difficult to evaluate. We can further simplify the integral by simplifying the variable \( r \) by using a Taylor expansion:

\[
r = \sqrt{(\Delta z)^2 + (u_1 - u_0)^2 + (v_1 - v_0)^2} \\
= \Delta z \sqrt{1 + \left( \frac{u_1 - u_0}{\Delta z} \right)^2 + \left( \frac{v_1 - v_0}{\Delta z} \right)^2} \\
\approx \Delta z \left( 1 + \frac{1}{2} \left( \frac{u_1 - u_0}{\Delta z} \right)^2 + \left( \frac{v_1 - v_0}{\Delta z} \right)^2 - \mathcal{O}(2) + \ldots \right). \] (1.5.3)
The Fresnel approximation assumes that the propagation is large enough such that the higher order terms of the expansion can be ignored. As a consequence of this approximation, we get \( r^2 \simeq \Delta z^2 \) in the denominator of the integrand in Eq. 1.5.2. On the other hand, we cannot replace \( r \) in the numerator by \( \Delta z \) because of the factor \( 1/\lambda \).

As the wavelength of electromagnetic waves is of the order of microns, the \( 1/\lambda \) term is of order \( 10^6 - 10^7 \), hence this factor amplifies small errors due to the approximation and causes rapid \( 2\pi \) periodic errors in the phase of the integrand. Substituting these approximations in Eq. 1.5.2, we get

\[
E_1(u_1, v_1) = \frac{e^{i \frac{2\pi}{\lambda} \Delta z}}{i \lambda \Delta z} \int \int E_0(u_0, v_0) e^{i \frac{\pi}{\lambda \Delta z} \left( (u_1 - u_0)^2 + (v_1 - v_0)^2 \right)} du_0 dv_0
\]

(1.5.4)

Equation 1.5.4 is called the Fresnel equation and \( \mathcal{F}\{\cdot\} \) the Fresnel propagator. The Fresnel equation further simplifies as

\[
E_1(u_1, v_1) = \frac{e^{i \frac{2\pi}{\lambda} \Delta z}}{i \lambda \Delta z} \int \int E_0(u_0, v_0) e^{i \frac{\pi}{\lambda \Delta z} \left( u_0^2 + v_0^2 \right)} e^{-i \frac{2\pi}{\lambda \Delta z} \left( u_0 u_1 + v_0 v_1 \right)} du_0 dv_0
g(u_1, v_1) \mathcal{F}\{E_0 \cdot g\}(u_1, v_1),
\]

(1.5.5)

where

\[
g(u_1, v_1) = e^{i \frac{\pi}{\lambda \Delta z} \left( u_1^2 + v_1^2 \right)}
\]

(1.5.6)

and \( \mathcal{F} \) denotes the 2-dimensional Fourier transform:

\[
\mathcal{F}\{E\}(u_1, v_1) = \int \int E_0(u_0, v_0) e^{-i \frac{2\pi}{\lambda \Delta z} \left( u_0 u_1 + v_0 v_1 \right)} du_0 dv_0.
\]

(1.5.7)

For computational purposes, the Fresnel integral can be re-written as a convolution
of the electric field $E_0$ with a quadratic phase factor,

$$E_1(u_1, v_1) = \int \int E_0(u_0, v_0) h(u_1 - u_0, v_1 - v_0) du_0 dv_0$$

$$= \mathcal{F}^{-1} \left\{ \mathcal{F}\{E_0(u_0, v_0)\} e^{-i\pi \lambda \Delta z (u_1^2 + v_1^2)} \right\},$$

(1.5.8)

where $\mathcal{F}^{-\{\}}$ is the inverse Fourier transform operator and $h$ is the convolution kernel defined as

$$h(u_1, v_1) = \frac{e^{i \frac{2\pi}{\lambda \Delta z} \Delta z}}{i \lambda \Delta z} e^{i \frac{\pi \lambda \Delta z}{\Delta z} (u_1^2 + v_1^2)}.$$

(1.5.9)

For free space propagation between two planes, we will use Eq. 1.5.8. For most of the work presented in this thesis, we will be propagating electric field from pupil planes to image planes and vice-versa. A simple imaging system uses a lens or a collimating mirror to bring light to a focus. In the remainder of this section, we will develop the relationship between the pupil and image planes. Figure 1.11 shows the relevant planes and imaging optic used for this propagation. An arbitrary electric field, $E_0$, is propagated to a lens located at a distance $\Delta z = f$, where $f$ is the focal length of the lens. The lens is represented by an operator $\mathcal{L}$ and it focuses the incident field to an image plane. For a thin lens, the lens operator is defined as

$$\mathcal{L}\{E_{in}(\xi, \eta)\} = E_{in}(\xi, \eta) e^{-i \frac{\pi}{\lambda f} (\xi^2 + \eta^2)}.$$

(1.5.10)

Applying the Fresenal integral, Eq. 1.5.4, to obtain the electric field at distance $f$
Figure 1.11: Illustration of propagation of electric field from a pupil plane to an image plane. \( E_0 \) is an arbitrary electric field at a plane located one focal length, \( f \), upstream of a thin lens, represented by a linear operator \( \mathcal{L} \), and \( E_{im} \) is the electric field at the image plane which is also located at distance \( f \) from the lens.

downstream from the lens, we get

\[
E_{im}(x, y) = \mathcal{F}r\{L\{E_{in}(\xi, \eta)\}\}
\]
\[
= e^{i\frac{2\pi}{\lambda}f} \int \int \mathcal{L}\{E_1(\xi, \eta)\}e^{i\frac{\pi}{\lambda f}(x-\xi)^2+(y-\eta)^2} d\xi d\eta
\]
\[
= e^{i\frac{2\pi}{\lambda}f} \int \int E_1(\xi, \eta)e^{-i\frac{\pi}{\lambda f}(\xi^2+\eta^2)}e^{i\frac{\pi}{\lambda f}(x-\xi)^2+(y-\eta)^2} d\xi d\eta
\]
\[
= e^{i\frac{2\pi}{\lambda}f} \int \int E_1(\xi, \eta)e^{-i\frac{2\pi}{\lambda f}(\xi x+\eta y)}d\xi d\eta
\]
\[
= e^{i\frac{2\pi}{\lambda}f} \int \int E_1(\xi, \eta)e^{i\frac{\pi}{\lambda f}(x^2+y^2)}\mathcal{F}\{E_1(\xi, \eta)\}.
\]

It can be observed that the field at the image plane is the product of a piston phase, quadratic phase factor, and the FT of the incident field. This incident field is obtained by Fresnel propagation of the pupil plane field, \( E_0 \), by a distance \( f \). We know from Eq.1.5.4 that the Fresnel propagation is just a convolution of \( E_0 \) with the kernel \( h \).
Hence, the FT of $E_1(\xi, \eta)$ is given by

\[
\mathcal{F}\{E_1(\xi, \eta)\} = \mathcal{F}\{E_0(u, v)\}\mathcal{F}\{h\}
\]

\[
= \frac{e^{i\frac{2\pi f}{\lambda}f}}{i\lambda f} \mathcal{F}\{E_0(u, v)\} \int \int e^{i\frac{\pi}{\lambda f}(\xi^2 + \eta^2)} e^{-i\frac{2\pi f}{\lambda f}(\xi + \eta y)} d\xi d\eta
\]

\[
= \frac{e^{i\frac{2\pi f}{\lambda}f}}{i\lambda f} \mathcal{F}\{E_0(u, v)\} e^{-i\frac{\pi}{\lambda f}(x^2 + y^2)} \int \int e^{i\frac{\pi}{\lambda f}(x - \xi)^2 + (y - \eta)^2} d\xi d\eta
\]

\[
= e^{i\frac{2\pi f}{\lambda}f} e^{-i\frac{\pi}{\lambda f}(x^2 + y^2)} \mathcal{F}\{E_0(u, v)\}.
\]

Combining Eq. 1.5.12 and Eq. 1.5.11, we get

\[
E_{im}(x, y) = \frac{e^{i\frac{2\pi f}{\lambda}f}}{i\lambda f} \mathcal{F}\{E_0(u, v)\}.
\]

Other than the piston term, the electric field at the image plane is a FT of the electric field at the pupil plane —making them Fourier conjugates. As the original field, $E_0$, is arbitrarily selected from a reference field with zero phase, we can ignore the piston term without introducing any errors. The PSF of the SPC, shown in Fig. 1.4, is obtained by using Fourier transform. Now that we have established the relationship between different propagation planes of the electric field, we will briefly describe the wavefront sensing and control problem.

### 1.6 Wavefront Sensing and Control

As described in earlier section, amplitude and phase aberrations cause leakage of starlight, known as speckles, into the dark holes that can mask the planet. Figure 1.12 shows the effects of optical aberrations —the planet is not visible due to the speckles. The phase aberrations can be represented by Zernike polynomials. In the following subsection, we will briefly discuss the Zernike polynomials.

---

Although we have explained the relationship between different propagation planes of the electric field, the following sections assume the reader to be familiar with the fundamentals of optics.
Figure 1.12: Effect of optical aberrations. a) Ideal scenario with no aberrations. An exoplanet is imaged in the dark zone. b) Starlight leakage (speckle) masking out the exoplanet.

1.6.1 Zernike Polynomials

Zernike polynomials are a set of orthogonal polynomials defined on the unit circle as

\[
Z_n^m = \begin{cases} 
R_n^m(r)\cos(m\theta), & \text{if } m \geq 0 \\
R_n^m(r)\sin(m\theta), & \text{if } m < 0 
\end{cases}
\]  
(1.6.1)

where

\[
R_n^m = \sum_{k=0}^{(n-m)/2} \frac{(-1)^k(n-k)!}{k!((n+m)/2-k)!((n-m)/2-k)!} r^{n-2k}
\]  
(1.6.2)

and \(n\) and \(m\) are the radial degree and azimuthal order, respectively. At each radial degree \(n\) the number of Zernike polynomials at that degree is \(n+1\) and the azimuthal order ranges between \(-m\) to \(m\) in increments of 2. Zernike polynomials up to radial degree 4 are shown in the Fig. 1.13. Zernike polynomials have been used to represent the Kolmogorov spectrum of atmospheric disturbance [41], analyze wavefront aberrations [42], and also to simulate aberrations of the eye. For their unique properties over a circular aperture [43] and convenience of describing classical low-order aberrations
such as tip-tilt, defocus, coma, and astigmatism using them, the phase aberrations in this thesis are expressed in terms of Zernike polynomials. For convenience, the indices $n$ and $m$ are mapped to a single index $k$ using the convention introduced by Noll [44]. The rule is that the even Zernikes (with even azimuthal part $m$), $\cos(m\phi)$, obtain even indices $k$, the odd Zernikes obtain odd indices $k$. Within a given $n$, lower values of $m$ obtain lower $k$. We will use the Noll notation in the thesis.

The aberrations caused by telescope jitter, thermal fluctuations, and mechanical stresses can be represented by Zernike polynomials $Z_2 - Z_{15}$ (the low-order Zernike polynomials) and the aberrations due to the imperfections in the optics are of higher-order ($> Z_{15}$). Based on the time-scales of aberrations, for a ground-based system, there are three main categories of wavefront sensing and control. First, adaptive optics (AO) that compensates for atmospheric turbulence at timescales on the order of 1 millisecond (ms). Second, focal plane wavefront correction (FPWC) that mitigates quasi-static phase and amplitude aberrations from imperfections in manufacturing of optics and misalignments at timescales of hours and longer. Third, low-order wavefront sensing and control (LOWFS/C) that estimates and compensates for the dynamic aberrations caused by LoS jitter, mechanical flexure, and thermal fluctua-
Figure 1.14: Complete estimation and control scheme. The outer-most control loop is the higher-order control loop operating at a slower rate. It uses the image plane measurements to estimate quasi-static aberrations which are composed mainly of Zernike polynomials greater than $Z_{15}$. The LOWFS loop, depicted inside the red dotted box, uses the light rejected from the coronagraphic focal plane and estimates the lower-order aberrations. This loop has two control and estimation rates. The inner-most loop for tip-tilt control at 100 Hz - 1 KHz based on the system, and the middle loop that estimates and controls other aberrations such as astigmatism, focus, coma, trefoil, and sphere at mHz time scale.

At 100 hertz (Hz) to a few millihertz (mHz). For a spaced-based system, we have the latter two categories. The LOWFS/C is further divided into two different control loops based on the bandwidth of the control loops. Figure 1.14 is a schematic of the complete wavefront estimation and control scheme for a space-based system that employs two control loops, a faster inner loop to estimate and control the dynamic low-order aberrations and a slower outer loop to control the quasi-static higher order aberrations.
1.6.2 Brief Introduction to AO

Adaptive optics is employed in ground-based system to cancel out the effects of atmospheric turbulence. Typically, in an AO system, a beam splitter is placed in the optical path and used to reflect part of the light onto a wavefront sensor (WFS) and the remaining light to the science camera as illustrated in Fig. 1.15. An AO system employs a WFS, most commonly a Shark-Hartmann wavefront sensor (SHWFS), capable of characterizing the shape of an incident wavefront. Based on the information derived from the SHWFS, the shape of the DM needed to correct any wavefront distortion is calculated. This data is then fed back into the control system of the mirror which in turn changes the shape of the mirror. This system is a closed loop system that can continuously sample and measure the wavefront quality and feed back this information to control the mirror shape. One of the drawbacks of the AO system is

Figure 1.15: Schematics of an AO system. The light from the telescope is divided into sensing path and the science path. A wavefront sensor in the sensing path measures, at a rate of hundreds of times a second, how the star’s light is distorted by atmospheric turbulence. This information is sent to a fast computer, which calculates the voltage map that can be applied to the DM to conjugate the distortion. This mirror cancels out the distortions due to turbulence. Image obtained from Lawrence Livermore National Laboratory and NSF Center for Adaptive Optics.
the non-common path (NCP) errors. In an AO system, shown in Fig. 1.15, the NCP errors are the aberration differences between the WFS and the science camera; the WFS is not subject to the same light path as the science camera (and any aberration inducing optical elements along this path).

1.6.3 Brief Introduction to FPWC

The focal plane wavefront control (FPWC) uses the science camera image for wavefront sensing and control to avoid the NCP errors. The first FPWC technique was the model-free scheme called speckle nulling [45]. In this technique, Trauger et al. [45], applied sinusoids with different phases to the DM to suppress the stellar speckles at the targeted spatial frequency. Even though multiple localized speckles are corrected in each iteration, it requires hundreds or thousands of correction iterations to suppress the entire dark hole and is thus too slow for a space mission. The first model-based FPWC used the science image plane measurements and surface modulations applied to the DMs, in effect solving for a “dark hole” actuator solution [46]. This led the way for faster model-based estimation and control methods. This technique, however, involved a nonlinear optimization that was numerically unstable and too time consuming for real-time correction. Borde and Traub [47] proposed a linearized controller to minimize the energy in the dark hole as a quadratic cost function, but this direct minimization was also numerically unstable. Give’on developed the optimal controller Electric Field Conjugation (EFC) and the least squares wavefront estimation algorithm (DM diversity) that used a Tikhonov regularization to stabilize the problem [48]. To improve the accuracy of the linearized DM model, Pueyo invented the stroke minimization controller [49]. Groff utilized a Kalman filter to improve the DM Diversity algorithm [50] and Riggs used an extended Kalman filter for faster and optimal estimation of both incoherent and coherent sources [51].

These image plane methods, which estimate and control static or quasi-static
aberrations, are blind to short time-scale variations [52]. As discussed earlier, due to the dynamic range of the imaging camera, the FPM blocks most of the rejected starlight. The lack of the core of the stellar PSF makes it difficult to estimate the low-spatial frequency aberrations by using the science image. Therefore, a number of sensing schemes have been proposed that use the rejected star light or rely on instruments outside of the coronagraph and the science path to reduce the demand placed on the much slower image plane calibration loop. In the following section, we will briefly discuss different methods that use the rejected starlight to sense the low-order aberrations.

1.6.4 Brief Introduction to Low-Order Wavefront Aberration and LOWFS

The low-order aberrations, which change at shorter time-scale, are usually represented by low-order Zernike polynomials ($Z_2 - Z_{15}$), hence the name low-order wavefront sensor (LOWFS) for the sensor used to measure them. In this work, the LOWFS refers to both the sensor and all accompanying algorithms.

The first project that used rejected starlight for LOWFS was the Lyot project [53, 54]. It was implemented as a through-hole in a flat mirror to sense tip-tilt errors only. Guyon et al. [55] built on this idea to develop the Coronagraphic Low Order Wavefront Sensor (CLOWFS) approach to estimate additional low-order Zernike modes such as focus, astigmatism, and coma. In the CLOWFS scheme, the coronagraph focal plane mask is an oblique transmissive plane with an opaque occulting spot. The outer ring of the occulting spot is reflective, directing light from an annulus of the star’s PSF (with inner and outer radii $\sim 0.7$ and $1.8 \, \lambda/D$) to a lens outside of the science path. Guyon et al. place a CCD detector in the converging beam after this lens, but offset slightly from its focus. Tip, tilt, defocus, and two orthogonal astigmatism modes are fitted to the critically sampled intensity distribution in the defocused image [55, 56].
The displacement of the sensing plane from true focus is necessary to retrieve the
defocus component of the wavefront when the magnitude of that error is in the small
phase regime (\(\ll 1\) radian). In fact, it can be shown with additional analysis that
estimating any phase mode with even symmetry requires the CLOWFS sensor to be
defocused [57]. A similar technique is used for coronagraphs that employ a phase
mask, but the information is obtained from the Lyot plane and the sensor is called a
Lyot Low-order Wavefront sensor (LLOWFS) [58].

Other solutions for coronagraph wavefront sensing rely on external interferometric
calibration [59, 60, 61, 62]. One approach in this category that avoids some of the
complexities typically associated with constructing an interferometer is the phase
contrast technique. F. Zernike first developed the phase contrast sensor in 1942
to convert optical path differences in translucent microscope specimens to intensity
signals [63]. It was proposed for astronomical applications by R.H. Dicke [64], and
since then the concept has been adapted for coronagraph wavefront sensing [65] and
evaluated on several instrument testbeds [66, 67, 68]. A phase contrast wavefront
sensor (also known as a Zernike wavefront sensor (ZWFS)) also uses the light extracted
from the occulted region of the focal plane as in CLOWFS. The core of the beam is
then focused on a \(\pi/2\) phase-shifting spot of diameter \(\sim 1\)–\(2\) \(\lambda/D\). This phase-shifting
spot can also be built into the focal plane mask of the coronagraph to eliminate
unnecessary optics. For small phase errors, the intensity of the interference pattern
in the subsequent pupil is a linear function of the phase distribution in the original
pupil. The baseline architecture of the WFIRST mission includes a ZWFS for sensing
the low-order wavefront aberrations [35].

At PHCIL, Katie Cavanagh worked on a LOWFS technique that utilized inter-
ferometry to extract the low-order aberrations [57]. Cavanagh developed a one-
dimensional model that used two slits in the mask (to make the 1-D mask/plate
non-redundant). In a non-redundant mask (NRM), the sub-apertures (holes/slits)
are arranged so that the baselines between each pair of holes is unique, hence a 1-D NRM can only have two slits. This technique estimated the aberration based on the visibility pattern of different aberration modes (Legendre polynomials, as they are the 1-D version of Zernike polynomials). Since the number of sub-apertures was restricted to two, the only parameter that could be varied was their size. This limitation adversely affected the mask’s performance. Improving on the interferometric technique, in Subedi et al. [69], we showed that a sparse aperture mask wavefront sensor (SAM WFS) is an effective approach for estimating the dynamic low-order aberrations. Like the CLOWFS and the Zernike phase mask sensor, the SAM WFS method uses the rejected starlight from the coronagraphic focal plane. The FPM reflects light from the core of the PSF towards the SAM WFS and transmits the remaining light to the science camera optics. The collimating optic following the FPM forms a re-imaged pupil at the SAM. The light diffracted by the pattern of holes in the SAM is brought to a focus on a dedicated detector. The low-order aberration modes can be inferred from the intensity pattern at the detector. This thesis demonstrates a proof of concept for the SAM WFS.

Aperture masking with both sparse aperture and non-redundant masks have been used widely for wavefront sensing for diffraction-limited, high-dynamic-range astronomical observations [70, 71, 72, 73, 74, 75]. In this technique, the non-redundant mask is placed at a relay pupil and the light is brought to a focus on a detector forming an interference pattern. Fourier analysis of this interference pattern can retrieve closure phase observables that are invariant to time-varying low-order aberrations, and otherwise inaccessible with conventional filled aperture imaging [76]. Retrieval of the closure phase provides a robust way of co-phasing the segments of a segmented telescope and also provides a powerful diagnostics of AO residuals [77, 78, 79, 80]. An asymmetric sparse aperture mask can also be used instead of an NRM to retrieve the closure phase and has been recently implemented on the Subaru Coronagraphic
Extreme AO (SCExAO) instrument and on PALM-3000 extreme AO system on the Palomar 200-inch telescope [81, 82]. These closure phase techniques require access to the full PSF, hence the implementation of the aperture mask wavefront sensing till now has been without a coronagraph in the optical path.

1.7 Thesis outline

We begin in Ch. 2 with a detailed overview of the Zernike wavefront sensor, the coronagraphic low-order wavefront sensor, the Lyot-based low-order wavefront sensor, and the sparse aperture mask wavefront sensor. In this overview, we will present the working mechanism of each WFS, derive their Fourier propagation model, and the corresponding algorithm used to sense the low-order aberrations. The initial SAM design that is used to serve as proof of the SAM WFS concept is presented in this chapter. In Ch. 3 we use the initial SAM design and the Fourier model to numerically verify the efficacy of the SAM WFS in monochromatic simulations. We use the CLOWFS as a reference point for the performance evaluation. We do a fit analysis between the estimates made by the wavefront sensors and the true input values. We also present the linearity regime of the SAM WFS and compare it with linearity regimes of other wavefronts sensors. The chapter also includes the ability to translate the coefficient estimates to a reconstructed phase distribution and the ability to sense rapid tip-tilt error. Chapter 4 shows the results of the SAM WFS experiments. It includes a description of the SAM WFS lab, its layout and design iterations, and open-loop estimation results. In Ch. 5 we present the optimization formulation and the procedure used to optimize the SAM. Using the optimized mask obtained in Ch. 5, we present a detailed comparison of the optimized SAM WFS with other sensors in terms of estimation accuracy and robustness to noise for the different coronagraphs in Ch. 6. In Ch. 7 we explore the possibility and advantages of using a
different basis set than the Zernike polynomials as the LOWFS modes. The control section of the LOWFS/C is in Ch. 8. This chapter describes a new adaptive technique that can be used with a Kalman filter to estimate and control the LoS pointing error. We present the future direction of the SAM WFS work in Ch. 9 and then conclude the thesis.

There are three main new ideas in this thesis: 1) the SAM WFS technique and optimization of the SAM, 2) change of basis sets to represent the low-order modes, and 3) adaptive estimation and control of LoS pointing jitter. Therefore, all the work presented in this thesis is intended to serve as proof of concept of these ideas and compare them with existing approaches.
Chapter 2

Detailed description of low-order wavefront sensors

In this chapter, I will discuss different low-order wavefront sensors in detail and derive the mathematical formulation to estimate the low-order aberrations. The LOWFS configurations considered in this chapter use an SPC when the rejected starlight is obtained from the focal plane and a VVC when the LOWFS picks up the light from the Lyot-plane. In Section 2.1, the SAM WFS is discussed, the ZWFS is presented in Section 2.2, the CLOWFS in Section 2.3, and the LLOWFS in Section 2.4.

2.1 Sparse Aperture Mask Wavefront Sensor

A diagram of the sparse aperture mask wavefront sensor with an SPC is given in Fig. 2.1. Incoming starlight, apodized by the shaped pupil, is brought to a focus at a focal plane mask. This oblique mask reflects light from the core of the PSF towards the SAM WFS and transmits the remaining light to the science camera optics. The collimating optic following the reflective focal plane mask forms a re-imaged pupil at the SAM. The light diffracted by the pattern of holes in the SAM is brought to a focus on a dedicated detector. The right hand side of Fig. 2.1 depicts the intensity
pattern at the three critical planes: the first focal plane of the coronagraph, the SAM plane, and the sensor focus.

Figure 2.1: Diagram of the sparse aperture mask wavefront sensor, integrated with a shaped pupil coronagraph. On the right hand side, we plot the intensity at three critical planes on a logarithmic flux scale: the coronagraph focal plane, the sparse aperture mask, and the sensor.

2.1.1 Focal Plane mask (FPM)

The reflective focal plane mask directs discarded starlight toward the wavefront sensor. The radius of this region can be as small as the first zero of the PSF core, or as wide as the inner working angle of the coronagraph permits. For the model presented in this thesis, we fix the radius at $4\lambda/D$, to match the inner working angle of one of our ripple SPC designs.

Unlike conventional coronagraph apodizations, a shaped pupil can produce a PSF with significant energy outside the main lobe in certain directions. This is because
the shaped pupil optimization procedure can be tailored to create dark search zones restricted to a finite region of the image plane, in exchange for deeper contrast and small inner working angle [16, 23]. This is the case for the ripple SPC, as evident by the vertically extended sidelobe wedges of the nominal star PSF, \( I_{\text{foc}} \), shown in Fig. 2.1. Therefore, in addition to the simple reflecting spot, we tested a variation on the focal plane mask that utilizes the starlight outside the central lobe of the PSF. The only modification, as illustrated in Fig. 2.2, is an extension of the reflective region to a 95-degree wide wedge above and below the center, creating a “bowtie”-like shape out to a separation of 10 \( \lambda/D \) (Fig. 2.2).

Figure 2.2: Diagram of the alternative focal mask design for the SAM WFS, a reflective “bowtie” taking advantage of the shape of the PSF created by the ripple shaped pupil.

2.1.2 Sparse Aperture Mask

The light reflected from the focal plane mask is collimated to a relay pupil where the wavefront sensing aperture mask is located. The starlight re-imaged after propagating through the mask produces an interference pattern characteristic of the baselines between each pair of holes. The difference between the aberrated and nominal intensity
patterns provides the estimation signal.

For this illustration, we normalize the nominal intensity pattern (labeled $I_0$ in Fig. 2.1) to unity to show the relative scale of the differential signal. The aperture mask $\mathcal{M}$, over a two-dimensional Cartesian vector $\mathbf{u}$ in the telescope pupil plane (with the origin at the center of the pupil), is defined as it is in Zimmerman et al. [83], according to

$$\mathcal{M}(\mathbf{u}) = \Pi(\mathbf{u}/a) \otimes \sum_{i=1}^{N_h} \delta(\mathbf{u} - \mathbf{h}_i),$$

(2.1.1)

where $\otimes$ is the convolution operator, $N_h$ is the number of holes, $\mathbf{h}_i$ is the vector position of the center of each subaperture, $\Pi$ is the subaperture hole function, and $\delta$ is the Dirac delta function. For circular subapertures of radius $a$, $\Pi$ is defined via

$$\Pi(\mathbf{u}/a) = \begin{cases} 1, & \text{if } |\mathbf{u}|/a \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

(2.1.2)

### 2.1.3 SAM Sensing Equation

In this section, we develop an equation to relate the fringe pattern in the sensor to the aberrations in the pupil plane. The field at the first focal plane, $E_{foc}$, is the Fourier transform of the electric field at the pupil plane given by

$$E_{foc}(\nu) = \mathcal{F}\{A(r, \theta)e^{i\phi(r, \theta)}\}.$$ 

(2.1.3)

Here $A$ is the SPC and $\phi(r, \theta)$ is the phase aberration at the SPC plane. In the small aberration regime, the function $e^{i\phi(r, \theta)}$ can be well approximated by $1 + i\phi(r, \theta)$ and $\phi(r, \theta)$ can be expressed as a sum of Zernike polynomials, $Z_k(r, \theta)$, of Noll order $k$. 

36
Using these approximations, we get

\[
E_{foc}(\nu) \simeq \mathcal{F}\{A(r)\} + i \sum_k a_k \mathcal{F}\{A(r)Z_k(r, \theta)\}, \tag{2.1.4}
\]

where \(a_k\) is the coefficient of the Zernike polynomial \(Z_k\).

The binary-valued focal plane mask \(\mu_{\text{SAM}}(x)\) reflects some of the light to the wavefront sensor, where \(x\) is a two-dimensional Cartesian vector in the focal plane. The arguments \(r, \theta, u\) and \(x\) are dropped hereafter to simplify the reading. The field at the relay pupil plane following the reflection, \(E_{\text{pup}}\), is given by the Fourier transform of the product of the mask and the field at the focal plane such that

\[
E_{\text{pup}} = \mathcal{F}\{\mu_{\text{SAM}}E_{\text{foc}}\}
= \mathcal{F}\{\mu_{\text{SAM}}\mathcal{F}\{A\}\} + i \sum_k a_k \mathcal{F}\{\mu_{\text{SAM}}\mathcal{F}\{AZ_k\}\}. \tag{2.1.5}
\]

The field at the sensor plane, obtained after the light passing through the SAM is focused by an optic, is then given by the product of the SAM transmission function and the field at the relay pupil plane. The field at the sensor plane is

\[
E_s = \mathcal{F}\{\mathcal{M}E_{\text{pup}}\}
= \mathcal{F}\{\mathcal{M}\mathcal{F}\{\mu_{\text{SAM}}\mathcal{F}\{A\}\}\} + i \sum_k a_k \mathcal{F}\{\mathcal{M}\mathcal{F}\{\mu_{\text{SAM}}\mathcal{F}\{AZ_k\}\}\}. \tag{2.1.6}
\]

This field can be represented as \(E_s = E_0 + E_{ab}\) where the nominal field \(E_0\) is given by

\[
E_0 = \mathcal{F}\{\mathcal{M}\mathcal{F}\{\mu_{\text{SAM}}\mathcal{F}\{A\}\}\} \tag{2.1.7}
\]

and the effect of the aberrations on the field, \(E_{ab}\), is given by

\[
E_{ab} = i \sum_k a_k \mathcal{F}\{\mathcal{M}\mathcal{F}\{\mu_{\text{SAM}}\mathcal{F}\{AZ_k\}\}\}. \tag{2.1.8}
\]
The intensity at the sensor plane is given by

\[ I_s = E_s \bar{E}_s + n \]

\[ = (E_0 + E_{ab})(\bar{E}_0 + \bar{E}_{ab}) + n \]

\[ = E_0 \bar{E}_0 + 2Re(E_0 \bar{E}_{ab}) + E_{ab} \bar{E}_{ab} + n \]

\[ \simeq I_0 + 2Re(E_0 \bar{E}_{ab}) + n, \]  

where \( n \) is the photon and readout noise, \( I_0 \) is the nominal intensity at the sensor plane and \( E_{ab} \bar{E}_{ab} \) is neglected because \( a_k^2 \ll 1 \). Therefore, the resulting equation is linear and given by

\[ \frac{(I_s - I_0)}{2} = Re \{ E_0 \bar{E}_{ab} \} + n/2. \]  

(2.1.10)

We relabel the variable \( n/2 \) as \( n \) hereafter to simplify the reading. Substituting Eq. 2.1.8 in Eq. 2.1.10 and rearranging the 2-dimensional electric field distribution of each mode into a column vector we get,

\[ \frac{(I_s - I_0)}{2} = Re \{ E_0 \bar{E}_{ab} \} + n \]

\[ = \sum_k Re \left\{ E_0 i a_k \mathcal{M} \mathcal{F} \left\{ \mu_{SAM} \mathcal{F} \left\{ AZ_k \right\} \right\} \right\} + n \]

\[ = \left[ Re \left\{ E_0 i \mathcal{F} \left\{ \mathcal{M} \mathcal{F} \left\{ \mu_{SAM} \mathcal{F} \left\{ AZ_2 \right\} \right\} \right\} \right\} ... ... \right] \begin{bmatrix} a_2 \\ a_3 \\ \vdots \\ a_k \end{bmatrix} + n. \]  

(2.1.11)

Equation 2.1.11 can be written as

\[ \frac{(I_s - I_0)}{2} = H_{SAM} x + n. \]  

(2.1.12)
where the response or the modal matrix

$$H_{SAM} = \left[ \text{Re} \left\{ E_0 i \mathcal{F} \{ \mathcal{M} \mathcal{F} \{ \mu_{SAM} \mathcal{F} \{ AZ_2 \} \} \} \right\} \right]_{j=1,...,k}$$

(2.1.13)

and $x = [a_j], j = 1,...,k$. The modal matrix $H_{SAM}$ and the nominal image $I_0$ are calculated analytically based on the optical configuration. The intensities are vectorized and have dimensions $m^2 \times 1$, and $H_{SAM}$ is a $m^2 \times k$ matrix, where $m$ is the number of pixels in one dimension and $k$ is the number of Zernike modes to be estimated. The modal matrix can also be obtained by propagating different Zernike modes through the system and observing the resulting images at the detector. As it is impossible to perfectly model a complicated optical system, thus for experiments the modal matrix is obtained by using the latter technique.

The images obtained at the detector are affected by readout noise and shot noise. The readout noise in the camera chip is random, uncorrelated noise. All cameras add additional noise when reading the image from the camera chip and converting it into a digital image. The shot noise is due to the discrete nature of photons arriving at the detector. Photon noise or the shot noise is modeled as a Poisson distribution to determine the numbers of photons that arrive on a pixel in a given time frame. If $X_{ij}$ is a random variable, the probability that the photon count equals a particular value $x_{ij}$ is given by

$$Pr[X_{ij} = x_{ij}] = e^{-\lambda_{ij}t} \frac{(\lambda_{ij}t)^{x_{ij}}}{x_{ij}!},$$

(2.1.14)

where $\lambda_{ij}$ is the mean arrival rate. The mean and variance for this distribution are $\lambda_{ij}t$. The signal to noise ratio (SNR) is the ratio of the signal to the background noise. Therefore a large signal to noise ratio represents that more signal is received by the sensor while a small SNR shows that the sensor is heavily influenced by noise. For photon noise, the signal to noise ratio is given by $SNR = \sqrt{\lambda_{ij}t}$. 
The solution to 2.1.12, \( \hat{x} \), is obtained using a least square fit,

\[
\hat{x} = (H_{SAM}^T H_{SAM})^{-1} H_{SAM} (I_s - I_0)/2,
\]

(2.1.15)

where superscript \( T \) denotes the transpose of a matrix.

When a VVC or a BLC is used, the configuration is modified such that the LOWFS optical path obtains the light from the Lyot plane as shown in Fig. 2.3. The Lyot mask in this case is the SAM itself. The SAM in this configuration has three regions: 1) the white transparent region that lets light through to the science camera, 2) the red opaque region that blocks the light, and 3) the black reflective region that reflects light toward the LOWFS camera. The mathematical analysis for this configuration

![Diagram](image.png)

Figure 2.3: Diagram of the sparse aperture mask wavefront sensor, integrated with a phase mask coronagraph. In this configuration, the starlight rejected by the coronagraph is directed from the Lyot plane to the LOWFS path.

is identical to the configuration with the SPC. However in this configuration, \( A \) is the exit pupil, \( \mu_{SAM} \) is the phase mask at the focal plane instead of the binary FPM, and \( M \) is the modified Lyot plane SAM. We will explain how the modified Lyot plane
2.1.4 SAM initial design

The design parameters of the initial aperture mask, used with an SPC, are the number, sizes, and locations of the subaperture holes. The need to sense even-symmetric Zernike modes such as defocus requires the pattern to be asymmetric with respect to the center of the pupil. Beyond this constraint, there is a large parameter space to explore. For the initial demonstration we tailored the SAM design for the ripple shaped pupil shown in Fig. 2.1, aiming to recover Zernike phase modes of radial degree up to $n = 4$ (14 modes in total, excluding piston). Table 2.1 provides the peak differential intensity for the first 14 Zernike aberrations after piston. The difference between the aberrated and nominal intensity patterns for different Zernike modes for this initial design is shown in Fig. 2.4.

![Figure 2.4: The differential intensity patterns for three aberration cases: (a) defocus ($\lambda/30$ waves RMS), (b) astigmatism ($\lambda/30$ waves RMS), and (c) coma ($\lambda/30$ waves RMS).](image)

All the quantitative results presented for monochromatic variations are based on the SAM design shown in Fig. 2.1, which uses 27 subaperture holes, each with a diameter 7% of the pupil. This specific mask is intended only to serve as a proof of concept. The numerical optimization of the SAM geometry will be a subject of later
Table 2.1: Peak differential intensity $(I_{ab} - I_0)/I_0$ of SAMWFS for a fixed $\lambda/30$ RMS wavefront error.

<table>
<thead>
<tr>
<th>Noll index (after piston Z1)</th>
<th>peak differential intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5.42e-02</td>
</tr>
<tr>
<td>3</td>
<td>5.34e-02</td>
</tr>
<tr>
<td>4</td>
<td>7.58e-03</td>
</tr>
<tr>
<td>5</td>
<td>8.37e-03</td>
</tr>
<tr>
<td>6</td>
<td>7.31e-03</td>
</tr>
<tr>
<td>7</td>
<td>1.03e-01</td>
</tr>
<tr>
<td>8</td>
<td>1.02e-01</td>
</tr>
<tr>
<td>9</td>
<td>4.74e-03</td>
</tr>
<tr>
<td>10</td>
<td>3.96e-03</td>
</tr>
<tr>
<td>11</td>
<td>1.92e-02</td>
</tr>
<tr>
<td>12</td>
<td>1.88e-02</td>
</tr>
<tr>
<td>13</td>
<td>2.57e-02</td>
</tr>
<tr>
<td>14</td>
<td>1.72e-03</td>
</tr>
<tr>
<td>15</td>
<td>1.78e-03</td>
</tr>
</tbody>
</table>

2.2 Zernike Wavefront Sensor

The ZWFS is also a differential wavefront sensor and uses light rejected by the coronagraph, but the diversity is introduced by using a phase mask. The ZWFS uses the Zernike phase contrast concept and converts the phase aberrations into intensity variations. Figure 2.5 depicts the concept of the Zernike wavefront sensor intended to be used for telescopic applications. The wavefront error is represented by the wavefront phase variation at the entrance pupil plane. The light from the telescope pupil plane is focused by using a powered optic to a focal plane. A ZPM with a diameter $\sim 1\lambda/D$ is placed at the focal plane. The light that passes through the phase disk of the ZPM acts as a reference wavefront (Reference in Fig. 2.5 a) and interferes with the light passing outside the phase disk which contains the information on wavefront aberrations (Aberrated in Fig. 2.5 a). The interfered light is then relayed to another pupil plane, where the detector records the intensity pattern. This intensity pattern
is related to the phase variation of the telescope pupil plane and depends on the size of the phase mask and the phase change introduced by this mask. If the phase aberrations at the telescope pupil plane are $\ll 1$ radian, a phase difference of about $\pi/2$ introduced by the phase mask results in a linear relationship between the pupil plane phase variation and the intensity recorded at the detector. The ZWFS configuration used with SPC is shown in Fig. 2.5 b. The apodized light is focused by using a powered optic to a focal plane. The reflective phase disk of the ZPM (Fig. 2.6) has a diameter of $1.22\lambda/D$, a reflective region outside the phase disk that has a radius of $4\lambda/D$ and the outer transparent region that lets light to the science path.

2.2.1 ZWFS algorithm

For a detailed analysis of the ZWFS, we refer readers to Wallace et al. and Shi et al [66, 84]. For this analysis, we strictly follow the procedure used for WFIRST and
Figure 2.6: FPM/ZPM for the ZWFS. The central reflective phase disk has diameter of \(1.22\lambda/D\) and the reflective region outside the phase disk has a diameter of \(4\lambda/D\).

detailed in Shi et al [84]. The electric field at the entrance pupil is by

\[
E(u) = P(u).A[1 + \epsilon(u)].e^{i\phi(u)},
\]

where \(u\) is the two-dimensional Cartesian vector in the telescope pupil plane (with the origin at the center of the pupil), \(P\) is the pupil geometry, \(A\) is the mean electric field amplitude, \(\epsilon\) is the amplitude variation across the entrance pupil, and \(\phi\) is the phase error at the pupil. In the small aberration regime, the field at the first focal plane, \(E_{\text{foc}}\), is the Fourier transform of Eq. 2.2.1 given by

\[
E_{\text{foc}}(x) = \mathcal{F}\left\{P(u).A[1 + \epsilon(u)].e^{i\phi(u)}\right\}
\]

\[
\simeq \mathcal{F}\{P(u)\} \otimes \mathcal{F}\{A[1 + \epsilon(u) + i\phi(u)]\},
\]

44
where $\mathcal{F}$ is the Fourier operator and $\otimes$ is the convolution operation. Representing the Fourier transform of $P$ by a complex PSF, $C$, then Eqn. 2.2.2 becomes

$$E_{foc}(x) = A.C(x) + A.C(x) \otimes \mathcal{F}\{\epsilon(u) + i\phi(u)\}. \quad (2.2.3)$$

It can be noticed from above equation that the electric field at the focal plane is the sum of the ideal complex PSF ($A.C(x)$) and the ideal PSF convolved with phase and amplitude errors ($A.C(x) \otimes \mathcal{F}\{\epsilon(u) + i\phi(u)\}$). At the focal plane a ZPM is placed and is represented by the transmission function

$$t(x) = 1 - (1 - e^{i\theta})M, \quad (2.2.4)$$

where $\theta$ is the $\pi/2$ phase shift and $M$ is a top hat with $M = 1$ for the area inside the phase dimple and $M = 0$ for the area outside of the disk. The arguments $u$ and $x$ are dropped hereafter to simplify the reading. The size of the phase disk is comparable with the complex PSF core size, where most of the energy of an ideal PSF is located. The energy distribution from either the amplitude or phase error component tends to distribute outside the PSF core area, where the phase shift is zero. Therefore, the phase shift $e^{i\theta}$ is applied only to the ideal PSF component ($A.C(x)$) and the phase shifted electric field at the focal plane is given by

$$E_{foc} = A.Ce^{i\theta} + A.C \otimes \mathcal{F}\{\epsilon + i\phi\}. \quad (2.2.5)$$

The field at the sensor plane, located at another pupil plane, after the focal plane electric field is Fourier propagated to a relay pupil is

$$E_s = P.A.(e^{i\theta} + \epsilon + i\phi). \quad (2.2.6)$$
As $\theta = \pi/2$, the field at the pupil plane becomes

$$E_s = P.A.(e^{i\pi/2} + \epsilon + i\phi)$$

$$= P.A.(i + \epsilon + i\phi).$$

(2.2.7)

The intensity at the sensor plane is given by

$$I_s = E_s \bar{E}_s$$

$$= (P.A)^2.(1 + \epsilon^2 + 2\phi + \phi^2).$$

(2.2.8)

For a small aberration regime, the intensity is approximated as

$$I_s = (P.A)^2.(1 + \epsilon^2 + 2\phi).$$

(2.2.9)

For a reference wavefront with amplitude $\epsilon_0$ and phase $\phi_0$, the nominal/reference image is obtained by using Eqn.2.2.6 as

$$I_0 = (P.A)^2.(1 + \epsilon_0^2 + 2\phi_0).$$

(2.2.10)

After a small phase change, $\delta \phi$, the aberrated intensity due to the phase aberration $\phi_{ab} = \phi_0 + \delta \phi$ is given by

$$I_{ab} = (P.A)^2.(1 + \epsilon_0^2 + 2\phi_{ab})$$

$$= (P.A)^2.(1 + \epsilon^2 + 2(\phi_0 + \delta \phi)).$$

(2.2.11)

Using Eqn.2.2.10 and Eqn.2.2.11, the phase change we want to estimate is given by

$$\delta \phi = (I_{ab} - I_0)/(2P^2.A^2).$$

(2.2.12)

In this treatment, the dynamic low-order wavefront aberration is decomposed as
a linear combination of Zernike polynomials i.e., \( \delta \phi = \sum_k a_k Z_k \), where \( a_k \) is the coefficient of each Zernike mode \( Z_k \). The linearized relationship between the differential images at the detector and the coefficient of the Zernike polynomials is given by

\[
(I_{ab} - I_0) = \sum_k H_k a_k + n, \tag{2.2.13}
\]

where \( H_k \) is the modal response for each Zernike mode, which is generated with differential images with the known input amount ("training point"), \( \alpha \), of each Zernike mode applied to the system and \( n \) is the noise. Eq. 2.2.13 can be vectorized as

\[
\Delta I = H_{ZWFS} x, \tag{2.2.14}
\]

where \( H_{ZWFS} \) is the response matrix of the ZWFS. The intensities have \( m^2 \times 1 \) dimensions and \( H_{ZWFS} \) is a \( m^2 \times k \) matrix, where \( m \) is the number of pixels in one dimension and \( k \) is the number of Zernike modes to be estimated. The coefficients of the Zernike modes can be solved by using the method of least squares, such that

\[
\hat{x} = \alpha^{-1}(H_{ZWFS}^T H_{ZWFS})^{-1} H_{ZWFS}^T \Delta I. \tag{2.2.15}
\]

The "training point", \( \alpha \), is chosen such that it is close to the anticipated aberrations. In addition, both the reference and aberrated images are normalized before the images are differentiated to keep the scale factor consistent. If the wavefront aberrations are significantly different from the "training points", then linearity assumptions do not hold and the estimations are erroneous. The schematic presented in Fig. 2.7 shows the modification of the ZWFS for a phase mask coronagraph. In this configuration, the information for the LOWFS sensor is obtained from the Lyot plane and the ZPM is placed at another focal plane. The modified ZPM has a central phase disk of \( 1.22 \lambda/D \) and an outside transparent region.
Even though this configuration has two extra planes from the SPC configuration, the estimation technique is similar to the one presented above, i.e., the response matrix is generated with differential images with the known input amount of each mode applied to the system.

2.3 Coronagraphic Low-Order Wavefront Sensor

The CLOWFS devised by Guyon et al. is similar to the SAM WFS (Section 2.1), in that it also relies on re-imaged starlight discarded by the focal plane stop [55]. The CLOWFS as presented by Guyon et al. used a different coronagraph and had an annular FPM. For the SPC configuration, we found that using the FPM similar to the FPM used by SAM WFS gives a better performance. Hence, the CLOWFS used for the SPC (Fig. 2.8) also has an oblique FPM that reflects light from the
core of the PSF towards the LOWFS path and transmits the remaining light to the science camera optics. However, there are two essential differences between the CLOWFS and the SAM WFS. First, rather than collimating the extracted light to form a re-imaged pupil, the light is simply directed to a single converging lens or mirror. Second, the detector in the CLOWFS sensing plane must be offset from true focus of the last converging optic, to enable estimation of aberration modes with even symmetry [55, 57].

Figure 2.8: Diagram of the coronagraphic low-order wavefront sensor, integrated with a shaped pupil coronagraph.

With the intent of using CLOWFS as a benchmark for our initial concept, here we develop an analogous sensing equation for that system, which we later employ in our numerical model. Optically, the field at the CLOWFS sensor is equivalent to the field in the plane at a small distance after the reflective focal plane mask $\mu_{CS}$. This defocused field is computed by the Fresnel integral, which we abbreviate as the operator $\mathcal{F}_r\{ \}$. We again symbolize the reflective annulus by the binary variable $\mu_{CS}$. Then, analogous to Equation 2.1.6, for the CLOWFS sensor electric field we
have

\[ E_s = \mathcal{F}_r\{\mu_{CSF}\{A(r, \theta)e^{i\phi(r, \theta)}\}\} \]
\[ = \mathcal{F}_r\{\mu_{CSF}\{A\}\} + i \sum_k a_k \mathcal{F}_r\{\mu_{CSF}\{AZ_k\}\} \]
\[ = E_0 + E_{ab}. \]  

Using the same approximation as in Equation 2.1.9, dropping the \( E_{ab} \) term in the intensity expression, we again find a system of linear equations relating the aberration coefficients to the differential intensity pattern:

\[ \frac{(I_s - I_0)}{2} = Re\left\{E_0 \overline{E_{ab}}\right\} + n \]
\[ = \sum_k Re\left\{E_0 \overline{a_k} \mathcal{F}_r\{\mu_{CSF}\{AZ_k\}\}\right\} + n \]
\[ = \left[ Re\left\{E_0 \mathcal{F}_r\{\mu_{CSF}\{AZ_2\}\}\right\} \ldots \right] \begin{bmatrix} a_2 \\ a_3 \\ \vdots \\ a_k \end{bmatrix} + n \]  
\[ = H_{CLOWFS}x + n \]

The response matrix \( H_{CLOWFS} \) can also be obtained by propagating different Zernike modes through the system and observing the resulting images at the detector. To obtain the aberration coefficients for CLOWFS, the same approach from 2.1.3 was used.

### 2.4 Lyot-based Low-Order Wavefront Sensor

For coronagraphs that employ a phase mask, the CLOWFS can be adapted such that the information is obtained from the Lyot plane and the sensor is called the Lyot Low-
order Wavefront Sensor (LLOWFS) [58]. Figure 2.9 is an illustration of the LLOWFS technique. At the focal plane of the telescope, a phase mask diffracts starlight in a re-imaged pupil plane. The starlight is then directed by the reflective Lyot stop towards a detector, at a small defocus length from another focal plane, which is used for low-order wavefront sensing. Assuming that the aberrations are small enough, the linearized relationship between the differential images at the detector and the coefficient of the Zernike polynomials is given by

$$(I_{ab} - I_0) = \sum_i H_i a_i + n$$

$$= H_{LL}x + n,$$  \hspace{1cm} (2.4.1)

where $I_{ab}$ is the intensity after the phase variation, $I_0$ is the nominal intensity (intensity before the phase change), $H_i$ is the modal response for each Zernike mode, and $a_i$ is the coefficient of each Zernike mode. The solution of this equation is also
obtained by using a least squares fit.

The LOWFS techniques presented in this chapter use the rejected star light to estimate the dynamic low-order wavefront aberrations. In addition, all the techniques are differential, i.e., they all use the estimation algorithms that calculate the phase variation from a nominal point. This differential approach mitigates the problem of non-common path errors.
Chapter 3

SAM WFS initial simulation verification and experiments

3.1 Monochromatic simulations

To verify the SAM WFS technique, we developed a numerical Fourier propagation model. Our numerical Fourier propagation model of the SAM WFS uses a 512-point diameter representation of the ripple shaped pupil shown in Fig. 2.1. We sample the focal planes (both coronagraph and sensor) with 8 points across each $\lambda/D$ resolution element. As mentioned in Ch.2, we test two focal plane mask shapes for the SAM WFS: a disk of radius $4\lambda/D$ and a disk with “bowtie” wedges (Fig. 2.2). The array representing the CCD in the sensor plane bins the intensity points $2 \times 2$, resulting in a final image resolution of 4 pixels per $\lambda/D$. Bit quantization, photon counting noise, and 3 analog-to-digital unit (ADU) read noise are then applied to give a simulated intensity image.

Throughout our monochromatic validation analysis we used the CLOWFS as a

The work presented in this chapter was done in collaboration with Dr. Neil Zimmerman and the results were published in Journal of Astronomical Telescopes, Instruments, and Systems (Reference [69]). Dr. Zimmerman made substantial contributions toward obtaining the simulation results and in editing the paper.
reference point for performance, for two reasons: (1) it is a relatively mature low-
order wavefront sensing concept with strong experimental verification and (2) due
to the similarities in the optical configuration, a direct comparison required only
straightforward modifications to the SAM WFS Fourier propagation model. A de-
tailed comparison of the SAM WFS, the CLOWFS/LLOWFS, and the ZWFS for a
realistic mission scenario using broadband light is presented in later chapters. The
CLOWFS propagation model is similar to the SAM model, but requires a near-field
Fresnel integral to compute the defocused sensor field after the coronagraph focal
plane. The Fresnel propagation requires us to define several physical dimensions:
pupil diameter, focal length, propagation distance (defocus), and wavelength. These
parameters, listed in Table 3.1, were selected to match the testbed planned for the
experimental verification of the SAM WFS. The outer radius of the reflective focal
annulus was fixed at $4\lambda/D$, identical to the SAM WFS. We then tuned both the defo-
cus and the inner radius of the annulus by trial and error to give the best performance
in terms of $R^2$, a metric described in Section 3.1.1. For our coronagraph and range of
mode estimation, we found that the CLOWFS worked best without any inner radius,
instead leaving the reflective region as a simple disk. We found the best performance
at a defocus of 3% of the 200 mm focal length, or 6 mm. For our 550 nm wavelength
and 10 mm diameter pupil, this corresponds to a defocus aberration of 1.88 microns
peak-to-valley, equivalent to 3.4 waves. This calculation was done using the relation

\[ p = \frac{-\Delta z}{8f_z^2}, \tag{3.1.1} \]

where $p$ is the defocus aberration coefficient at pupil plane, $\Delta z$ is the detector defocus
length, and $f_z$ is the f-number of the system. We note this is very close to the 3.3
waves used by Guyon et al. in their published design [55].
Table 3.1: Nominal physical parameters of the wavefront sensor optical propagation model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pupil diameter</td>
<td>10 mm</td>
</tr>
<tr>
<td>Pupil sampling</td>
<td>512 points / diameter</td>
</tr>
<tr>
<td>Focal length of imaging optics</td>
<td>200 mm</td>
</tr>
<tr>
<td>Wavelength</td>
<td>550 nm</td>
</tr>
<tr>
<td>Spatial sampling at sensor image plane</td>
<td>4 pixels per $\lambda/D$</td>
</tr>
<tr>
<td>Sensor image width</td>
<td>80 pixels (20 $\lambda/D$)</td>
</tr>
</tbody>
</table>

3.1.1 Fit Analysis

To evaluate the SAM WFS concept, we constructed 100 aberrated wavefront realizations from random combinations of Zernike polynomials. This was repeated for two ranges of mode input and estimation: Zernike radial degree $n = 1–4$ and $n = 1–5$. The polynomial coefficients were drawn from uniform distributions, with larger bounds for the lower order terms than the higher order terms: 0.04 waves RMS for $n = 1$ and 2 and 0.016 waves RMS for larger radial degree. These coefficients are all small enough to satisfy the linearity assumption. Since lower order modes such as tip-tilt and defocus are strongest and quickly varying, we chose a higher value for these modes ($n=1–2$). Nevertheless, the propagation model used the full expression for the field at the each plane, rather than the first-order expansion used for the estimation equation. In these trials, the pixel with the maximum intensity was assumed to have 40,000 counts (near full-well on a 16-bit CCD) while other pixels received photons based on their relative intensity. We plot the estimated coefficients versus the input coefficients of these trials in Fig. 3.1, for the case of the SAM WFS with a circular focal plane mask, estimating modes up to radial degree 4. The tight clustering of all these points to the $y = x$ line is a good indication of the accuracy of the sensor and indicates that there is no significant cross-talk or degeneracy between modes in the system.

In Table 3.2 we collect the mean estimation errors across the ensemble of wavefront realizations, for each WFS configuration and for both ranges of mode estimations. The phase is compared only within the open area of the shaped pupil. For a typical
wavefront in the $n=1–4$ trial ensemble, the subtracted SAM WFS estimate reduces the RMS wavefront error by a factor of 30 and the peak-to-valley by a factor of 10. The estimate with the bowtie FPM is only marginally better than the circular case. For the CLOWFS, the RMS residuals are typically a factor of 2–3 worse than the SAM. For the $n=1–5$ trial, the SAM estimates reduce the typical RMS error by a factor of 20 and the improvement offered by the bowtie FPM over the circular FPM remains slight.

We also analyzed the random wavefront estimates with the coefficient of determina-
Table 3.2: Mean residual errors of the estimates of 100 random wavefronts, as measured within the open area of the shaped pupil. For each WFS configuration and radial degree estimation range, we give the mean RMS residual and the mean peak-to-valley residual. The top row gives the mean wavefront errors for the uncorrected input realizations. All values are in units of wavelength.

<table>
<thead>
<tr>
<th></th>
<th>n = 1–4</th>
<th></th>
<th>n = 1–5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMS in</td>
<td>P–V in</td>
<td>RMS in</td>
<td>P–V in</td>
</tr>
<tr>
<td>SAM, circular FPM</td>
<td>0.015</td>
<td>0.097</td>
<td>0.018</td>
<td>0.11</td>
</tr>
<tr>
<td>SAM, bowtie FPM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLOWFS</td>
<td>5.1 × 10^{-4}</td>
<td>8.9 × 10^{-3}</td>
<td>8.0 × 10^{-4}</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>4.6 × 10^{-4}</td>
<td>8.0 × 10^{-3}</td>
<td>6.8 × 10^{-4}</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>1.3 × 10^{-4}</td>
<td>0.022</td>
<td>2.3 × 10^{-4}</td>
<td>0.066</td>
</tr>
</tbody>
</table>

nation method ($R^2$ fit analysis). The fit assessment value was obtained by

$$R^2 = 1 - \frac{RSS}{TSS}.$$ (3.1.2)

RSS (residual sum of squares) is the sum of the squared errors of the inferred mode coefficients versus “truth” over all wavefront realizations, measuring the discrepancy between the data and the estimation model. TSS (total sum of squares) is the sum of the squared differences of the inferred coefficients from their respective means over all realizations. An aberration mode is considered to be estimated well if $R^2 > 0.5$.

Both wavefront sensors maintain $R^2$ values well above this cutoff for all phase modes (Fig. 3.2). However, the accuracy advantage of the SAM in this regime of full-well signal is apparent by the fact that the lowest $R^2$ value for the SAM is 0.95 with an average $R^2$ value of 0.97 versus 0.79 for CLOWFS with an average of 0.9.

### 3.1.2 Phase reconstruction tests

We illustrate the ability to translate the coefficient estimates to a reconstructed phase distribution in Fig. 3.3. The input phase map in Fig. 3.3a is comprised of the same combination of $\lambda/30$ defocus and coma modes used to simulate the aberrated PSF and contrast curve in Fig. 3.4. The residual error of the SAM WFS estimate is $6 \times$
Figure 3.2: The $R^2$ fit metric of the SAM WFS and CLOWFS, averaged for each mode over 100 random wavefront realizations. The SAM results are plotted with both tested FPM shapes, circular and bowtie.

Figure 3.3: (a) Input phase map at pupil plane with a defocus and coma mode, each of amplitude $\lambda/30$ waves RMS; (b) the corresponding reconstructed phase from the SAM WFS model; and (c) residual error of the estimated phase, with peak-to-valley $6 \times 10^{-3}$ waves. In all plots the phase is displayed in units of wavelength.

We also considered the response of the system to three more cases: (1) wavefronts consisting of pure Zernike modes of varying amplitude (2) a random phase screen obtained by a linear combinations of Zernike modes ($n = 1–4$), and (3) a Kolmogorov phase map.

For case (1), an increasing amount of each Zernike mode is applied to a flat
Figure 3.4: (a) Nominal (flat wavefront) point spread function, plotted on a log scale; (b) point spread function with defocus and coma modes added to the wavefront, each with strength $\lambda/30$ RMS waves; (c) Azimuthally averaged contrast curve for cases of flat wavefront, defocus and coma modes, and after removing the aberration estimated by the SAM WFS model.

wavefront to analyze the response of the sensor to low-order aberrations (Fig. 3.5a). The response of both wavefront sensors is linear for Zernike coefficients up to 0.07 waves RMS (other than Coma response for CLOWFS) and deviates about 10% from linear near $\lambda/10$ waves RMS. When compared with a similar calculation published for a Zernike WFS [85], our SAM and CLOWFS models exhibited a somewhat larger range of perfect linearity response up to 0.05$\lambda$ waves RMS as compared to 0.03$\lambda$ waves RMS for Zernike WFS. For case (2), we created a phase map composed of a linear

Figure 3.5: (a) Response of the SAM wavefront sensor to pure Zernike modes, showing the range of linearity, and (b) an example of the Zernike coefficient estimation for one random wavefront realization.
combination of Zernike modes with random coefficients, with the same constraints as the evaluations in Section 3.1.1. In Fig. 3.5b, we plot the value of the 14 Zernike coefficients, comparing the true input values and the estimates side by side. At each mode, the coefficient estimate is accurate to within $\lambda/100$, consistent with Table 3.2.

The case of a Kolmogorov phase map is an interesting test for the concept because in general we expect real optical surfaces to create a full spectrum of phase errors, with spatial frequency far above what the SAM WFS can estimate. Using a program to generate a random phase distribution with a Kolmogorov power spectrum (power law exponent $-11/3$), we tested several realizations with the resulting phase scaled up to $0.05\lambda$ RMS over the open area of the ripple SPC. One example of this reconstruction is shown in Fig. 3.6. The left hand panel (Fig. 3.6a) shows the original phase map at the entrance of the SPC. The reconstructed phase (Fig. 3.6b) has $0.038\lambda$ RMS and there is $0.030\lambda$ RMS of residual high order aberrations (Fig. 3.6c) after the estimated phase was subtracted from the input.

![Figure 3.6](image)

Figure 3.6: (a) Input phase map with a Kolmogorov power spectrum, (b) phase estimate for modes up to radial degree $n=4$, and (c) the residual error after subtracting the estimate from the input. The phase is plotted here in units of radians.

### 3.1.3 Response to a rapid pointing error

We expect tip-tilt errors arising from line-of-sight pointing oscillations to be the most quickly varying aberration mode for a space telescope and also one of the most im-
portant types of wavefront error to correct. Therefore, we assess the ability of the SAM WFS and the CLOWFS to sense a rapid pointing error in a regime of a fixed, realistic exposure time.

We assume the space telescope has an open, 2-meter diameter circular aperture, and that the coronagraph operates at a central wavelength of 0.55 µm. The target star has a V-band apparent magnitude of 4.83, appropriate for a Sun-like star at a distance of 10 pc. We collect light over a 20% bandwidth, and model this bandwidth in our Fourier propagation by averaging the sensor intensity pattern computed at 5 wavelengths spanning the passband. We assume losses due to reflections upstream of the coronagraph accumulate to 50% of the energy incident on the telescope primary, and a detector quantum efficiency of 0.8 e−/photon.

The ripple shaped pupil transmits 14.7% of the incident energy relative to an open circular aperture, and the FPM mask (in this test fixed as a disk of radius 4λ/D) transmits 56.7% of the energy incident on the coronagraph focal plane. For the SAM, due to the aperture mask there is an additional ratio of 9.1% between the energy arriving at the sensor and the energy reflected by the FPM. The distribution of the nominal intensity patterns of SAM and CLOWFS also differ significantly. The CLOWFS intensity is more centrally concentrated: for a CCD sampling of 4 pixels per λ/D the ratio of energy in the peak pixel to the full image is 1.10%, versus 0.22% for the SAM WFS.

We simulate the differential wavefront tilt signal collected over a 0.05 sec integration time for 1 and 10 milliarcsec (mas) pointing errors. The simulation model includes the diffraction propagation of 5 different wavelengths for 20% bandwidth centered at wavelength of 0.55 µm. This integration time is appropriate for controlling a jitter oscillation with a temporal frequency near or below 1 Hz. The phase gradient for a 1 mas pointing error is 0.0176 waves peak to valley across the telescope diameter. We simulate each CCD array with a read noise of standard deviation 1 ADU, although
in practice this could be lower if the instrument used an Electron Multiplying Charge Coupled Device (EMCCD) customized for low flux operation [86]. In Fig. 3.7, we show the noisy differential CCD image (aberrated minus nominal intensity) for the 1 mas pointing error, at a resolution of 4 pixels per $\lambda/D$. We tested the response of both wavefront sensors to 100 signal realizations for each pointing error.

The lower throughput of the SAM WFS is partially compensated for by the fact that the differential tilt signal is proportionally larger than the peak intensity. For a 1 mas tilt, the peak-to-valley difference signal is 2.5% of the peak SAM intensity, versus 1.1% for the CLOWFS. However, after taking into account the throughput to the sensor and the distribution of the intensity, the difference signal for the wavefront tilt is still 13 times higher for CLOWFS than for SAM. For short exposure times with low SNR, this naturally results in poorer estimation. For both sensors, the estimate distribution is centered on the true value. Therefore, in Table 3.3 we simply list the sample standard deviations to indicate the typical errors. The scatter in the SAM estimate is a factor of $\sim 5$ wider for both tilt levels.

Figure 3.7: Simulated CCD image differences between nominal and aberrated intensities for the SAM and CLOWFS wavefront sensors due to a 1 milliarcsec pointing error, observed with a 0.05 s integration time over a 20% bandwidth.
Table 3.3: Standard deviations of the tilt estimates for 100 realizations of a 0.05 sec exposure, for wavefronts with 1 mas and 10 mas pointing errors.

<table>
<thead>
<tr>
<th>Estimate std. dev. (milliarcsec)</th>
<th>1 milliarcsec tilt</th>
<th>10 milliarcsec tilt</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM WFS</td>
<td>0.36</td>
<td>0.37</td>
</tr>
<tr>
<td>CLOWFS</td>
<td>0.080</td>
<td>0.082</td>
</tr>
</tbody>
</table>

3.2 Intermediate Summary

In our comparisons with CLOWFS, we found that the accuracy of the SAM estimation was promising, but the light loss through the aperture hindered the performance in estimation scenarios with short exposure times. However, we knew from trials with different mask shapes that there was a tradeoff that could be explored between mask throughput and estimation accuracy. Hence, the need to optimize the SAM. After this initial monochromatic validation, the SAM WFS research differentiated into two primary goals: 1) experimental validation with the initial mask design and 2) SAM optimization and comparison of SAM WFS with other sensors using a realistic scenario and different families of coronagraphs. In the next three chapters, we will discuss the steps taken to achieve these goals.
Chapter 4

Experimental Results

After the numerical validation, we designed a LOWFS testbed to experimentally verify the efficacy of the SAM WFS. In this chapter, we will describe the lab setup and the experimental results.

4.1 Testbed design

We will start this section by describing the optical elements present in the setup.

4.1.1 Masks

The shaped pupil used on the testbed is shown in Fig. 4.1a and was manufactured using Deep Reactive Ion Etching (DRIE) out of a silicon wafer at the Jet Propulsion Laboratory (JPL) in Pasadena, CA. This mask was designed for a circular aperture (4.4 mm in diameter) to achieve two-sided DHs and a mean contrast of $10^{-7}$. The FPM and the SAM (shown in Fig. 4.1 b and c) were also manufactured using the same technique. The FPM had multiple holes of diameter 1.15 mm and a rectangular cut out for alignment. The SAM, manufactured according to the initial design specifications presented in Section 2.1.4, had 27 subaperture holes, each with a diameter
of 0.308 mm.

![Manufactured masks and their holding mechanisms: a) shaped pupil coronograph, b) focal plane mask, and c) the sparse aperture mask.](image)

4.1.2 DM

We have a $12 \times 12$ BMC MEMS DM in the lab. With 400 µm pitch between actuators, the DM has a 4.4 mm diameter. Since the DM was used to introduce the low-order aberrations, it was essential to characterize it. We used a Phasics wavefront sensor to take the necessary measurements. To obtain the influence function, all the DM actuators were actuated to 50 V and the response was saved. Then, one of the central actuators was actuated to 60 V while all the actuators were still commanded to 50 V input. The influence function was obtained by subtracting these two responses. Similarly, to obtain the voltage to height characteristics, one of the central actuators was actuated from 0 V to 100 V with an increment of 5 V and the response subtracted from the unpowered surface. The Phasics measurement shows that the unpowered DM has more than 1 wave peak-to-valley wavefront aberration, Fig. 4.2a. To flatten the DM, voltage commands necessary to conjugate the surface aberration are required. Fitting the obtained influence function, Fig. 4.2b, to a 2-D gaussian function, the full width at half maximum was calculated to be 0.78 actuators. Using the surface map
and influence function, the surface height for the actuators was calculated. Then, using the voltage to height relation (Fig. 4.2d), the voltage map required to flatten the DM was calculated and applied to the DM. Using similar procedure, the DM was used to create the low-order aberrations.

Figure 4.2: a) Unpowered DM shape in units of waves, b) influence function of a DM actuator poke, c) influence function in one dimension, and d) voltage to height relation of a DM actuator.

4.1.3 Camera

We used a Trius - SX9 Starlight Xpress monochrome camera to take the LOWFS images. The properties of this CCD camera are listed in Table 4.1. The LOWFS images taken by a CCD should be calibrated to reduce errors. Dark frames correct for the gradual accumulation of electrons (the dark current) in the pixels of the detector. The electric pre-charge applied to the CCD chip by the camera electronics to activate its photon collecting ability also creates bias in the images. Using a dark frame of the same duration and temperature as the data frames also accounts for the
bias element. Hence, these frames should be subtracted from the LOWFS images. Figure 4.3a shows a dark frame taken for the experiment exposure time and Fig. 4.3b is a frame showing readout noise of the CCD. All the images taken for the experiments were dark subtracted.

Table 4.1: Specifications of the Starlight Xpress camera used for the SAM WFS experiments.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel size</td>
<td>6.45 $\mu$m x 6.45 $\mu$m</td>
</tr>
<tr>
<td>Image format</td>
<td>1392 x 1040 pixels</td>
</tr>
<tr>
<td>CCD Image area</td>
<td>8.98mm (Horizontal) x 6.7mm (Vertical)</td>
</tr>
<tr>
<td>Quantum Efficiency</td>
<td>$\geq$ 50% (400 nm to 750 nm)</td>
</tr>
<tr>
<td>Readout Noise</td>
<td>Less than 7 electrons RMS</td>
</tr>
<tr>
<td>Data format</td>
<td>16 bits</td>
</tr>
<tr>
<td>System gain</td>
<td>0.35 electrons per ADU</td>
</tr>
<tr>
<td>Computer Interface</td>
<td>Built-in USB 2.0</td>
</tr>
<tr>
<td>Image download time</td>
<td>0.6 seconds at full resolution using USB 2.0</td>
</tr>
</tbody>
</table>

Figure 4.3: LOWFS camera images. a) A dark frame taken in the lab and b) readout noise.

4.1.4 SAM WFS testbed

The testbed, shown in Fig. 4.4, is the testbed designed to verify the SAM WFS performance. Using achromatic lenses, $f = 200$ mm, two relay pupil planes were
Figure 4.4: SAM WFS testbed design. a) Sketch of the experimental layout and b) aligned testbed. This layout shows the features of the LOWFS testbed at Princeton. Using a series of achromatic lenses with focal length of 200 mm, the DM, the SPC, and the SAM are placed in conjugate pupil planes.

created. The BMC 12x12 - actuator DM placed at the first pupil plane determined the entrance pupil diameter of 4.4 mm and we used it to introduce low-order aberrations. The SPC, also 4.4 mm in diameter, and the SAM were placed at subsequent relay pupil planes, respectively. The FPM was placed at the image plane after the SPC. The SPC and the SAM were mounted on circular mounts with linear translation capability in X-Y direction, the FPM was mounted on a motorized X-Y linear stage and a manual linear stage in the Z direction for proper alignment, and the camera was mounted on a motorized linear stage. To align the powered optics, we used an optical flat instead of the DM. We used a shearing interferometer and a Phasics wavefront sensor for alignment. Figure 4.5 shows the PSF of the system with a 3.5 mm diameter aperture used at the SPC plane. We removed the SAM and the FPM to obtain this PSF. After the testbed was aligned, we carried out the following open loop experiments.
4.2 Experimental procedure and Results

The following procedure was carried out for the open loop experiments.

1. A 50V bias was applied to the DM.

2. An image was taken without added aberrations and saved as the nominal or reference image ($I_0$).

3. A voltage map to create each Zernike mode was applied to the DM. For each mode, 0.2 radians was selected as the training coefficient.

4. To obtain the response matrix, an image was taken for each voltage map and the nominal image subtracted from it.

5. To obtain the differential aberrated image, aberrations of 0-0.5 radians were introduced and $I_0$ was subtracted from the resulting image.

6. The preceding step was carried out 25 times to obtain 25 sets of images for each input.
Figure 4.6: Simulated and measured response of SAM WFS to different Zernike modes. Simulations: a) tip, b) tilt, c) astigmatism, d) focus, and e) oblique astigmatism. Experiments: f) tip, g) tilt, h) astigmatism, i) focus, and j) oblique astigmatism.

Figure 4.7: Simulated and measured response of SAM WFS to different Zernike modes. Simulations: a) coma x, b) coma y, c) trefoil, and d) sphere. Experiments: e) coma x, f) coma y, g) trefoil, and h) sphere.

The responses for different modes are presented in Fig. 4.6 and Fig. 4.7. Qualitatively, the experimental responses look similar to the simulated responses with some
error introduced due to the alignment errors and surface aberrations of the lenses in the system. The experimental responses are also noisier than the simulations. After obtaining all the images, we carried out two sets of experiments. For the first experiment, we used the response matrix to fit the 25 different sets of differential aberrated images obtained by varying each mode from 0 radians to 0.5 radians. These results are presented in the linearity plots in Fig. 4.8. The plots include $3\sigma$ error bars.

![Figure 4.8: Performance of the SAM WFS. The plots show the linearity of SAM WFS response with $3\sigma$ error bar. The experiments were carried out 25 times to calculate the error.](image)

For the tip/tilt estimation, the SAM WFS estimations were linear up to $\sim0.8$ radians with a slope of $\sim0.982$ and the estimations were consistent with the simulation results—an order improvement from the previous setup. For other modes, the linearity held up to $\sim0.05$ waves rms. A deviation of about 5-10% was seen from 0.05 waves to 0.1 waves. The second-order effects became significant after 0.8 radians for tip/tilt and $\sim0.1$ waves for other modes. This onset of non-linear effects degraded the accuracy and the consistency of the estimations. The experimental results were
also limited by wavefront residual after flattening the DM and the linearity properties of DM actuators themselves. In addition, the system was aligned using a two-inch flat. Replacing the flat with the DM might have caused some collimation issues downstream.

4.3 Brief Chapter Summary

In this chapter, we demonstrated in a lab that the SAM WFS can make precise estimates of the low-order wavefront aberrations. In the desired aberration regime, the linearity response of the SAM WFS was consistent with the simulation results.
Chapter 5

SAM WFS Optimization

5.1 SAM Optimization

To increase the throughput of the SAM, we tried various optimization techniques. In this chapter, we will discuss the different optimization procedures used to optimize the SAM. Wavefront estimation using the SAM is done by solving the linear Eq. 2.1.12, by taking the pseudo-inverse of the response/modal matrix $H_{SAM}$. The $k$th column of the matrix $H_{SAM}$ as shown in Eq. 2.1.13 can be expressed as:

$$H_{SAMk}(\mathcal{M}) = a_1(\mathcal{M}) \odot a_{2,k}(\mathcal{M}) + a_3(\mathcal{M}) \odot a_{4,k}(\mathcal{M}), \quad (5.1.1)$$

where

$$a_1(\mathcal{M}) = \text{vec} \left[ K_I (E_{0SAM} \odot \mathcal{M}) K_I^T - K_R (E_{0SAM} \odot \mathcal{M}) K_R^T \right],$$

$$a_{2,k}(\mathcal{M}) = \text{vec} \left[ K_I (E_{SAMk} \odot \mathcal{M}) K_R^T + K_R (E_{SAMk} \odot \mathcal{M}) K_I^T \right],$$

$$a_3(\mathcal{M}) = \text{vec} \left[ K_I (E_{0SAM} \odot \mathcal{M}) K_R^T + K_R (E_{0SAM} \odot \mathcal{M}) K_I^T \right],$$

$$a_{4,k}(\mathcal{M}) = \text{vec} \left[ K_R (E_{SAMk} \odot \mathcal{M}) K_R^T - K_I (E_{SAMk} \odot \mathcal{M}) K_I^T \right].$$

The work presented in this chapter was done in collaboration with Peter Varnai from TU Delft.
⊙ is the Hadamard product, and $K_R \in \mathbb{R}^{m \times n}$ and $K_I \in \mathbb{R}^{m \times n}$ are the real and imaginary parts of the Fourier transform matrix $K \in \mathbb{C}^{m \times n}$, which is used for efficiently calculating 2D Fourier transforms, i.e. $\hat{F} = KFK^T$ [87]. Finally, the matrices $E_{0SAM}$ and each $E_{SAM_k}$ for $k = 1, \ldots, p$ give the electric field produced by the nominal phase and the Zernike perturbations propagated to the SAM plane, respectively. Since the ⊙ operator in Eq. 5.1.1 represents element-wise product, $H_{SAM}$ is a quadratic function of the mask $M$.

The error in the obtained Zernike coefficient vector determines the accuracy of the sensor. Including photon and readout noise in the sensing equation and assuming that the sensor operates within the linear regime, the $\hat{x}$ coefficients are obtained as the solution to

$$y = (I_s - I_0)/2 = H_{SAM}(M)x + n$$

(5.1.3)
as shown in Eq. 2.1.15. Let $y^*$, for which the true $x^*$ coefficients satisfy $y^* = H_{SAM}(M)x^*$ explicitly. The sensitivity of the estimate error to the measurement noise, $n$, can be bounded by the following inequality [88]:

$$\frac{||\hat{x} - x^*||}{||x^*||} \leq \kappa(H_{SAM}(M)) \frac{||n||}{||y^*||},$$

(5.1.4)

where $\kappa(\cdot)$ denotes the matrix condition number, defined in terms of its singular values as:

$$\kappa(\cdot) = \frac{\sigma_{\text{max}}(\cdot)}{\sigma_{\text{min}}(\cdot)}.$$  

(5.1.5)

The condition number is thus an indicator of how sensitive the solution of a linear equation is to small errors in the measurement; it should be as close to one as possible.

In this situation, the measurement noise $n$ is due to both photon and read-out noise, modeled as a Poisson and a Gaussian distribution, respectively. For a total of $N$ incoming photons, the effect of these two components is given by the following
equation:

\[ E(||n(N)||^2) = N + r_n, \quad (5.1.6) \]

where \( r_n \) is the variance of the readout noise. Furthermore, we can assume that the norm of the true measurement is proportional to the number of incoming photons, i.e., no saturation occurs and \( ||y^*(N)||^2 = \alpha N^2 \), where \( \alpha \) is a constant. Thus, the relative measurement error diminishes with the number of photons as \( \frac{1}{N} \times \sqrt{\frac{N + r_n}{\alpha}} \). This analysis and the sensitivity bound given by (5.1.4) suggests the following observations:

1. The modal matrix \( H_{SAM}(\mathcal{M}) \) should be well-conditioned in order to minimize the sensitivity of the estimated Zernike modes to measurement errors due to readout noise and photon noise.

2. A higher throughput implies better signal-to-noise ratio, leading to better estimation of Zernike coefficients as well.

From the above observations, we formulate an optimization procedure described in the following subsection.

### 5.1.1 Optimization procedure

In this section, we present different approaches used to formulate the design objectives highlighted above through an optimization problem. The first approach was based on the observation that a matrix is well-conditioned if its columns are orthogonal. This led to the following formulation of the SAM optimization problem:

\[
\begin{align*}
\min_{t, \mathcal{M} \in \mathcal{M}} & \quad t \\
\text{s.t.} & \quad ||tH_{SAM}^T H_{SAM} - I||_F \leq c.||I||_F \\
& \quad (5.1.7)
\end{align*}
\]
for some constant $c$. Here $t$ is a parameter introduced as a scaling factor to prevent the columns of $H_{SAM}$ having a norm of 1, $I$ is an identity matrix, and $||.||_F$ is the Frobenius norm. The value of the parameter $t$ gives the amount of light passing through the SAM. A mask with a higher throughput will yield a modal matrix with larger singular values; hence it will require a small scaling factor to minimize the norm shown in Eq. 5.1.7. Therefore, lower values of $t$ correspond to a higher average of singular values and thus a greater throughput. We found that the SAM obtained from this procedure was not as optimal as the second approach that used condition number of $H_{SAM}(\mathcal{M})$ as a measure. Nevertheless, for a faster convergence, the Frobenius norm solver could be used to find an appropriate initial mask for the condition number based solver.

In condition number based approach, to achieve the first goal of having a well-conditioned modal matrix (enumerated in the above subsection), we used the condition number of $H_{SAM}(\mathcal{M})$ as a measure. As for the second goal, maximizing the mask throughput, we tried different approaches. Since the SAM is located at a relay pupil plane, we tried maximizing the open area of SAM with more weights at the locations where the exit pupil is open. We also tried maximizing the open area of the mask with all locations weighted equally. However, it quickly became apparent that using open area of the SAM was a poor proxy for throughput - due to diffraction, the intensity of the light reaching the SAM is not evenly distributed along its surface. Better indicators for this measure are the singular values of $H_{SAM}(\mathcal{M})$. A higher average, $\bar{\sigma}(H_{SAM})$, of the singular values implies a higher throughput. Hence the optimization problem is formulated as

$$\max_{\mathcal{M} \in \mathcal{M}} \bar{\sigma}(H_{SAM}(\mathcal{M}))$$

$$\text{s.t. } \kappa(H_{SAM}(\mathcal{M})) \leq \kappa_0$$

(5.1.8)

for some constant $\kappa_0$. In practice, we found that condition numbers of $\kappa_0 = 20-25$
yield quite good results.

Even though this formulation is straightforward and better suited to achieving the design objectives, it has disadvantages too. It can been seen from Eqs. 5.1.1 and 5.1.2 that the modal matrix is a quadratic function of the mask to be optimized. The singular values and the condition number of the modal matrix are non-linear functions of the modal matrix. This makes the resulting optimization problem highly nonlinear with possibly many local minima. Most convex optimization techniques found in the literature involving the condition numbers are only applicable to matrices which are positive definite [89, 90]. Since the modal matrix we have is a quadratic function of the mask and is not positive definite, these techniques are not applicable. Chen et al. present a procedure based on a smooth approximation of the condition number which could serve as a good starting point for a more efficient implementation of the solution to the optimization problem in Eq. 5.1.8 [91]. In this work, we solved the problem using an interior-point method implemented in the \texttt{fmincon} function of \textsc{Matlab}®, without supplying a user-defined gradient for either the objective or the constraint. This nonlinear programming solver finds the minimum of a problem specified by

$$
\min_x \ f(x) \\
\text{s.t.} \quad c(x) \leq 0 \\
\quad \quad ceq(x) = 0 \\
\quad \quad A.x \leq b \\
\quad \quad Aeq.x = beq \\
\quad \quad lb \leq x \leq ub,
$$

(5.1.9)

where $b$ and $beq$ are vectors, $A$ and $Aeq$ are matrices, $c(x)$ and $ceq(x)$ are functions that return vectors, $f(x)$ is a function that returns a scalar, $lb$ is the lower bound, and $ub$ is the upper bound. The functions $f(x)$, $c(x)$, and $ceq(x)$ can be nonlinear.
functions.

To use this non-linear solver, we changed our optimization problem to a minimization problem by using \(\bar{\sigma}(H_{SAM}(\mathcal{M}))\) as the function to be minimized. We put the lower bound on all the mask points to be zero and the upper bound to one. The inequality constraints for our problem was \(\kappa(H_{SAM}(\mathcal{M}))-\kappa_0 \leq 0\). We found out that the solution converged to a binary mask and had a vertical symmetry. So to speed up the optimization and possibly avoid local minima, horizontal symmetry was explicitly enforced on the mask. A comparison of one of our fully optimized masks with the initial SAM design is depicted in Fig. 5.1. The optimized mask was significantly better in achieving both the design goals of having (i) a well-conditioned matrix and (ii) having a higher throughput. The \(\kappa(H)\) decreased by 60% while \(\sigma(H)\) increased by 350%.

Figure 5.1: Comparison of initial SAM design and the optimized SAM. a) Original mask, \(\kappa(H) = 49.6, \sigma(H) = 0.0844\). b) Optimized mask, \(\kappa(H) = 18.4, \sigma(H) = 0.297\).

From Eq. 5.1.4, we observed that the sensitivity of the SAM estimate depends on the product of the condition number and the relative measurement error. To find the optimal condition number for the SAM, we optimized the SAM for the SPC configuration using different condition numbers. Once we obtained the optimized mask, we calculated the photons arriving at the detector (details of photon calculations are
Figure 5.2: Optimized Lyot stop for the SAM WFS. a) Lyot stop for an ideal BLC and b) Lyot stop for a charge 6 VVC. The transparent section of the masks let the light to the science path, the dark region is the reflective region that sends the light to the LOWFS camera, and the red regions are opaque that block the light.

in Ch. 6) and calculated the product $\kappa(H_{SAM}(M)) \frac{||n||}{||n'||}$. For better estimates, i.e., for robustness to noise, we want the product to be small. Figure 5.3 shows condition number vs the product of condition number and the relative error. It can be observed that the product is small for $\kappa_0 = 20 - 30$.

Figure 5.3: Condition number vs the product of condition number and relative measurement error.

We chose $\kappa_0 = 25$ for the comparison analysis presented in the following chapter. Using similar methodology, we also optimized the Lyot stop for the BLC and the VVC;
the resulting masks are shown in Fig. 5.2. However, we found that the SAM used at the Lyot plane significantly underperforms other wavefront sensors when optimized using the constraint $\kappa_0 = 25$. We found out that $\kappa_0 = 10 - 15$ yields smaller relative error. Even though the constraints on condition number for these Lyot-based SAM masks are more stringent than the focal plane based SAM WFS, these masks let more light through to the LOWFS path. After we optimized the SAM for different coronagraphs, we compared the SAM WFS performance with other wavefronts sensors for a realistic mission scenario.
Chapter 6

Comparison of different LOWFS techniques using a broadband source

6.1 Comparison procedure and results

To compare different LOWFS techniques in a realistic scenario, we performed an analysis similar to the one presented in Ch. 3, but more extensive. In Ch. 3, we compared the SAM WFS and the CLOWFS techniques used with an SPC and analyzed their performance assuming the same photon level at the brightest pixel. The analysis presented in this chapter, however, assumes the same integration time and the optimized SAM to make the analysis fair and realistic. To evaluate the effectiveness of the sensors, we further divided the performance of the sensors into two sections: 1) sensor accuracy without noise and 2) robustness to noise. We also extended the analysis to include the ZWFS technique and used multiple families of coronagraphs. The coronagraphs used for this study are an SPC, a VVC, and a modified BLC. Since constructing a band-limited mask requires a mask of infinite extent, we used
a sine squared function with a cutoff at $16\lambda/D$ (Fig. 6.1) to make the mask manufacturable. Hence the mask was not perfectly band-limited —the light rejected at

![Figure 6.1: Modified BLC Mask that uses a sine squared function. The dark region of the mask blocks the light at the focal plane.](image)

the Lyot plane was more spread out as compared to the schematic shown in Fig. 1.8. The WFS configurations with an SPC was presented in an earlier section (Section 2) in Figs. 2.1, 2.5, and 2.8 and the WFS configurations for Lyot type coronagraphs are shown in Figs. 2.3 and 2.7. This performance evaluation will let the designers locate the cause of estimation error and help them select or optimize a sensor based on the problem at hand. In Section 6.1.1 and Section 6.1.2, we describe in detail the methodology used for the analysis and present the simulation results.

### 6.1.1 Accuracy of the sensors

To evaluate the inherent accuracy of the sensors, we simulated how the sensors perform in the absence of noise. In this section, we will describe the simulation procedure and present the results.
6.1.1.1 Simulation Procedure

For this evaluation, we constructed 100 aberrated wavefront realizations from random combinations of Zernike polynomials. The polynomial coefficients were drawn from uniform distributions, with larger bounds for the lower-order terms than the higher-order terms: 0.02 waves RMS for \( n = 1 \) and 2, 0.008 waves RMS for larger radial degree. These coefficients are all small enough to satisfy the linearity assumption: the input wavefronts were < 4 nm RMS. These aberrations were propagated through the system to obtain aberrated images at the LOWFS camera. Using the procedure described in Sections 2.1, 2.2, 2.3, and 2.4, the coefficients of the Zernike modes were estimated for each sensor. Once the estimates were made, for each mode, we obtained the slope of a linear fit between the input coefficients and the estimated coefficients. For a prefect system the slope would be unity. Just comparing the slope to one, however, does not give a complete picture of the efficacy of the sensors. To understand the goodness of the fit, we find the coefficient of determination \((R^2 \text{ fit analysis})\). As described earlier in Ch. 3, the fit assessment value was obtained via

\[
R^2 = 1 - \frac{RSS}{TSS},
\]  

(6.1.1)

where \( RSS \) (residual sum of squares) is the sum of the squared errors of the inferred mode coefficients versus “truth” over all wavefront realizations, measuring the discrepancy between the data and the estimation model, and \( TSS \) (total sum of squares) is the sum of the squared differences of the inferred coefficients from their respective means over all realizations. An \( R^2 \) value of 1 implies no spread and an intercept of 0. Therefore, the sensor with the slope and \( R^2 \) values for each Zernike modes closer to unity is more accurate and has less cross-talk between modes. The simulations use 100% broadband light centered at 550 nm for the SAM WFS and the CLOWFS. For the ZWFS, we found that using bandwidth broader than 80% resulted in esti-
mation error and convergence issues in closed-loop control. Therefore, we used 80% bandwidth for the ZWFS simulations.

Table 6.1: Nominal physical parameters of the wavefront sensor optical propagation model for the configuration with an SPC.

<table>
<thead>
<tr>
<th>Coronagraph: SPC</th>
<th>SAM</th>
<th>ZWFS</th>
<th>CLOWFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pupil diameter (mm)</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pupil sampling (points / diameter)</td>
<td>256</td>
<td>256</td>
<td>256</td>
</tr>
<tr>
<td>Wavelength (nm)</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Sampling at sensor image plane</td>
<td>2 pixels per $\lambda_0/D$</td>
<td>16 pixels/D</td>
<td>2 pixels per $\lambda_0/D$</td>
</tr>
<tr>
<td>Sensor image width (pixels)</td>
<td>32</td>
<td>16</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 6.2: Nominal physical parameters of the wavefront sensor optical propagation model for the configuration with a VVC.

<table>
<thead>
<tr>
<th>Coronagraph: VVC</th>
<th>SAM</th>
<th>ZWFS</th>
<th>LLOWFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pupil diameter (mm)</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pupil sampling (points / diameter)</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>Lyot plane reflective (points)</td>
<td>62-192</td>
<td>62-192</td>
<td>62-192</td>
</tr>
<tr>
<td>Wavelength (nm)</td>
<td>550</td>
<td>550</td>
<td>550</td>
</tr>
<tr>
<td>Sampling at sensor image plane</td>
<td>2 pixels per $\lambda_0/D$</td>
<td>16 pixels/D</td>
<td>2 pixels per $\lambda_0/D$</td>
</tr>
<tr>
<td>Sensor image width (pixels)</td>
<td>32</td>
<td>16</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 6.3: Nominal physical parameters of the wavefront sensor optical propagation model for the configuration with a BLC.

<table>
<thead>
<tr>
<th>Coronagraph: BLC</th>
<th>SAM</th>
<th>ZWFS</th>
<th>LLOWFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pupil diameter (mm)</td>
<td>10</td>
<td>NA</td>
<td>10</td>
</tr>
<tr>
<td>Pupil sampling (points / diameter)</td>
<td>64</td>
<td>NA</td>
<td>64</td>
</tr>
<tr>
<td>Lyot plane reflective (points)</td>
<td>34-96</td>
<td>NA</td>
<td>34-96</td>
</tr>
<tr>
<td>Wavelength (nm)</td>
<td>550</td>
<td>NA</td>
<td>550</td>
</tr>
<tr>
<td>Sampling at sensor image plane</td>
<td>2 pixels per $\lambda_0/D$</td>
<td>NA</td>
<td>2 pixels per $\lambda_0/D$</td>
</tr>
<tr>
<td>Sensor image width (pixels)</td>
<td>32</td>
<td>NA</td>
<td>32</td>
</tr>
</tbody>
</table>

For these simulations, we sampled the SAM WFS detector image with 2 pixels per $\lambda/D$ resolution element. The CLOWFS propagation model requires a near-field Fresnel integral to compute the defocused sensor field after the coronagraph focal plane. The Fresnel propagation requires us to define several physical dimensions: pupil diameter, focal length, propagation distance (defocus), and wavelength. We
defined the dimensions such that the image at the CLOWFS detector was sampled at 2 pixels per $\lambda_0/D$, where $\lambda_0$ is the central frequency. For the configuration with SPC, both the SAM WFS and the CLOWFS had the FPM with the reflective region fixed at $4\lambda_r/D$, where $\lambda_r$ is the largest wavelength used by the higher-order wavefront control loop. For the CLOWFS, we found the best performance was at a defocus of 3.4 waves. This is very close to the 3.3 waves used by Guyon et al. in their published design [55].

For the configuration with VVC and BLC, the best defocus length for LLOWFS was 0.8 waves. This defocus value is also close to the 5 radians wave defocus used by Singh et al. [58]. For the ZWFS with the SPC architecture, the reflective region was also fixed at $4\lambda_r/D$, but it had a phase dimple of $1.22\lambda_0/D$ that produced a phase shift of $\pi/2$ at the central wavelength. The nominal physical parameters of our propagation model are summarized in Tables 6.1, 6.2, and 6.3. The VVC Lyot stops are annular with inner diameter of 0.96 of the entrance pupil diameter (D) and outer diameter of 3D. For simplicity, the BLC Lyot stops are also annular but with a reflective region from 0.52D to 1.5D. A modified band-limited mask (modified from Fig. 6.1) used here blocks the central part of the PSF. Since the information from the central region of the image plane, which is supposed to go through the phase disk, is already blocked upstream, a ZWFS cannot be used. During these simulations, we made some interesting observations. For the BLC and VVC configurations (light obtained from the Lyot plane for LOWFS), the SAM WFS was limited by the larger range of detector pixel illumination, i.e., a few pixels were significantly brighter than the others. Ignoring the dimmer pixels did not remedy the problem since these pixels provided the distinguishing features of different modal responses. We also found out that when the SAM WFS detector was placed at a small optimal defocus distance like the LLOWFS, it solved the limit imposed by the dynamic range. To reduce the cross-talk between the modes and to make the estimates precise, we did a binary search to find the optimal SAM WFS detector defocus location. For this search, we minimized
the mean and the standard deviation of the $R^2$ values. A more robust method would be to modify the optimization problem described in Ch. 5 to include the defocus distance as a parameter to be optimized. We will discuss about possible modification of the optimization problem in a later chapter that discusses future direction of the SAM WFS work.

6.1.1.2 Simulation Results

Figure 6.2: Accuracy of the sensors in a configuration with an SPC. The simulations do not include noise. a) Slope of input vs estimated coefficients of 100 aberrated wavefront realizations from random combinations of Zernike polynomials for 100% bandwidth for the SAM WFS and CLOWFS and 80% bandwidth for the ZWFS. b) $R^2$ fit of the estimates in a).

Table 6.4: Inherent accuracy of different wavefront sensors with an SPC.

<table>
<thead>
<tr>
<th>Coronagraph: SPC</th>
<th>SAM</th>
<th>ZWFS</th>
<th>CLOWFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean slope</td>
<td>0.989</td>
<td>0.982</td>
<td>0.987</td>
</tr>
<tr>
<td>Std of slope</td>
<td>0.034</td>
<td>0.037</td>
<td>0.0365</td>
</tr>
<tr>
<td>Mean $R^2$</td>
<td>0.983</td>
<td>0.894</td>
<td>0.903</td>
</tr>
<tr>
<td>Std of $R^2$</td>
<td>0.021</td>
<td>0.0782</td>
<td>0.077</td>
</tr>
</tbody>
</table>

The results for the SPC are shown in Fig. 6.2 and Table 6.4, for the VVC in Fig. 6.3 and Table 6.5, and for the BLC in Fig. 6.4 and Table 6.6. All the sensors
Figure 6.3: Accuracy of the sensors in a configuration with a VVC. The simulations do not include noise. a) Slope of input vs estimated coefficients of 100 aberrated wavefront realizations from random combinations of Zernike polynomials for 100% bandwidth for the SAM WFS and CLOWFS and 80% bandwidth for the ZWFS. b) $R^2$ fit of the estimates in a).

Table 6.5: Inherent accuracy of different wavefront sensors with a VVC.

<table>
<thead>
<tr>
<th>Coronagraph: VVC</th>
<th>SAM</th>
<th>ZWFS</th>
<th>LLOWFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean slope</td>
<td>0.987</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Std of slope</td>
<td>0.011</td>
<td>0.020</td>
<td>0.0103</td>
</tr>
<tr>
<td>Mean $R^2$</td>
<td>0.986</td>
<td>0.9542</td>
<td>0.986</td>
</tr>
<tr>
<td>Std of $R^2$</td>
<td>0.007</td>
<td>0.048</td>
<td>0.01</td>
</tr>
</tbody>
</table>

perform well and make accurate estimates; however, with mean values of the metrics considered closest to unity and smaller deviations from the mean values, the SAM WFS is more precise for the configurations with the SPC. The LLOWFS and the SAM WFS have similar precision (similar mean values and standard deviation of $R^2$ and slopes) for the configuration with the VVC and the BLC. To get a complete knowledge of estimation errors for each sensor, we analyzed the estimation error due to photon and readout noise.
Figure 6.4: Accuracy of the sensors in a configuration with a BLC. a) Slope of input vs estimated coefficients of 100 aberrated wavefront realizations from random combinations of Zernike polynomials for 100% bandwidth. b) $R^2$ fit of the estimates in a).

Table 6.6: Inherent accuracy of different wavefront sensors with a BLC.

<table>
<thead>
<tr>
<th>Coronagraph: BLC</th>
<th>SAM</th>
<th>LLOWFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean slope</td>
<td>0.975</td>
<td>0.976</td>
</tr>
<tr>
<td>Std of slope</td>
<td>0.022</td>
<td>0.025</td>
</tr>
<tr>
<td>Mean $R^2$</td>
<td>0.931</td>
<td>0.93</td>
</tr>
<tr>
<td>Std of $R^2$</td>
<td>0.058</td>
<td>0.0605</td>
</tr>
</tbody>
</table>

6.1.2 Effects of Noise

Before we delve into the noise error analysis, we will explain the general procedure of calculating the photons arriving at a telescope detector.

6.1.2.1 Photon Calculations

We start from a magnitude zero ($f_\lambda$) point of a star. This is the magnitude of a source having a flux of 1 erg/s/cm$^2$/Angstrom. Then the energy per photon at the central frequency is calculated as

$$E_\lambda = \frac{hc}{\lambda}, \quad (6.1.2)$$
where $h$ is the Planck’s constant and $c$ is the velocity of light. The photon flux from the star is given by

$$\Phi = f_\lambda \Delta \lambda / E_\lambda \cdot 10^{-m_V/2.5},$$

(6.1.3)

where $\Delta \lambda$ is the bandwidth and $m_V$ is the apparent magnitude of the star. The photoelectron count rate arriving at the detector is

$$n_\star = \Phi A_{eff} \eta T Q_e,$$

(6.1.4)

where $A_{eff}$ is the effective area of the telescope, $\eta$ is the reflective losses due to the mirrors, $T$ is the fraction of non-scattered starlight incident on the primary that arrives at the detector, and $Q_e$ is the quantum efficiency. The count rate, $n_\star$, has the units of electrons/s. Based on the exposure time, the total number of count at the detector is calculated and based on the relative intensity of individual pixel, the count at each pixel is determined.

To analyze the robustness of the sensors to noise, we considered a 4 m telescope observing three stars of different apparent magnitudes (0, 4.83, and 8 using Johnson-Cousins V-band zero point) operating at 100% bandwidth for the SAM WFS, the CLOWFS, and the LLOWFS configurations and 80% for the ZWFS. The central wavelength was 550 nm. We assumed losses due to reflections upstream of the coronagraph accumulated to 50% of the energy incident on the telescope primary and a charge-coupled device (CCD) with a quantum efficiency of 0.8 e$^-$/photon. We collected light over appropriate bandwidths and modeled the bandwidth in our Fourier propagation by averaging the sensor intensity pattern computed at 11 wavelengths spanning the passbands. We assumed sensor read noise of 3 analog-to-digital units (ADU) and assumed the same design parameters as in Section 6.1.1. For this analysis, photon and detector noise were added to an image obtained without any wavefront aberrations. Then the nominal image was subtracted from the noise added image.
to obtain the differential image and the corresponding Zernike coefficients were calculated. Since the images were generated without wavefront aberrations, the sensed wavefront is the noise equivalent sensing error. The line-of-sight (LoS) error refers to the errors in $Z_2$ and $Z_3$. For the LoS comparison, we assumed a camera operating at 1 kHz and the tip-tilt coefficients were calculated after each frame. On the other hand, to estimate the noise equivalent error for modes $Z_4 - Z_{15}$, we stacked images over 1 min. These integration times were used so that the control loops had the desired bandwidth described in Fig. 1.14 in Section 1 and the $Z_4 - Z_{15}$ aberrations were estimated to the desired sub-nanometer level. When we averaged camera images so that the equivalent readout rate was one frame per minute, the equivalent readout noise became $3\sqrt{N}$, where $N$ is the number of images stacked over 1 min at 1 kHz rate, i.e., 60,000. The results we present in this chapter are the averages of 100 random realizations used to obtain the mean RMS tip-tilt ($Z_2, Z_3$) sensing error and the RMS wavefront error due to $Z_4$ to $Z_{15}$ sensing error. It can be noticed in

![Figure 6.5](image)

**Figure 6.5:** Effects of noise with a LOWFS configuration with an SPC. a) The noise equivalent LoS angle and b) noise equivalent sensing error for $Z_4 - Z_{15}$.

Fig. 6.5 that for the SPC, the ZWFS is most robust to noise errors. Even though the SAM WFS throughput is smaller than the LLOWFS throughput, for the BLC configuration, their performance is similar (Fig. 6.7). For the configuration with the
VVC, the LLOWFS and the SAM WFS are more robust to noise (Fig. 6.6) than the ZWFS. Although, the SAM WFS and the LLOWFS have similar performance, the LLOWFS performs a little better than the SAM WFS. To analyze the effect of readout noise only, we calculated the noise equivalent sensing error for $Z_4 - Z_{15}$ with different readout noise without including the photon noise. For this simulation, we considered a 4 m telescope with a VVC observing a star with 0 apparent magnitude and readout noise with standard deviation of 1, 3, 5, and 7 ADUs. It can be observed from Fig. 6.8 that for lower readout noise the SAM WFS is as robust as the LLOWFS,
while ZWFS performs worst.

Figure 6.8: Effects of readout noise with a LOWFS configuration with a VVC for \( Z_4 - Z_{15} \).
Chapter 7

Change of basis sets

7.1 DM voltage map as basis set

Once the low-order aberrations are estimated, the voltage required to cancel the aberrations is calculated and applied to the DM. As described in Eq. 1.4.4 in Section 1.4, the voltage map, $u$, is calculated as

$$u = (G\beta)^+ \phi_{est}, \quad (7.1.1)$$

where $G$ is the matrix with the normalized influence function of each DM actuators, $\beta$ is a diagonal matrix that has height-voltage relationship of each actuator, $\phi_{est}$ is the estimated phase, and the superscript, $+$, represents the pseudo-inverse of a matrix. Even though the actuators are calibrated, there will still be some calibration/gain error $\epsilon$. When a voltage map, $u$, is applied to the DM, the phase realized by the DM is $G(\beta + \epsilon)u$ due to the calibration error. Hence, the residual phase error is

$$\Delta \phi = \phi_{est} - G(\beta + \epsilon)((G\beta)^+ \phi_{est}),$$

$$= \phi_{est} - (I + G\epsilon\beta^+G^+)\phi_{est}, \quad (7.1.2)$$

$$= -(G\epsilon\beta^+G^+)\phi_{est}.$$
As mentioned earlier in Ch. 1, this residual aberration has mid to high-order spatial frequency, is blind to LOWFS, and introduces speckles in the dark hole.

To mitigate this problem, we propose using DM voltage maps as the basis set rather than pure Zernikes. As a proof of concept and for simplicity, we use the DM voltage maps that attempt to reproduce the Zernike polynomials as the basis set. Let $u_i$ be one of the voltage maps used as the basis set and $a_i$ be the “training point” coefficient of this voltage map which produces the differential response $\Delta I$. For this voltage command, the realized training phase is

$$\phi_t = G(\beta + \epsilon)u_ia_i. \tag{7.1.3}$$

When the sensor measures an aberrated differential response of $k\Delta I$, where $k < $ some constant $\in \mathbb{R}$, it means a phase aberration of $k\phi_t$. For this phase aberration, the LOWFS would calculate the coefficient of the mode $u_i$ to be $ka_i$. Once the coefficient is estimated to be $ka_i$, a voltage $-ka_iu_i$ is applied to the DM. When a voltage $-ka_iu_i$ is applied, the actual phase realized by the DM is

$$-kG(\beta + \epsilon)u_ia_i = -k\phi_t. \tag{7.1.4}$$

Hence, applying the voltage command $-ka_iu_i$ eliminates the residual phase error caused by gain errors. This analysis can be extended to a complete set of $u_i$’s and their linear combinations. To numerically prove this concept, we used two sets of basis functions and compared the residual wavefront aberrations for different values of gain errors. The two basis sets used were: 1) pure Zernike modes and 2) the voltage map applied on a 12x12 DM to create Zernike-like shapes. The same 12x12 DM was used to control the wavefront aberration and the maximum gain errors were chosen from $\max_{\epsilon_i} = \{0\%, 5\%, 10\%, 15\%, 20\%\}$. For each element of $\max_{\epsilon_i} > 0$, we ran 100 different realizations of the DM voltage map. The gain error of each actuator
was chosen randomly from $0 \leq \epsilon_i \leq \max \epsilon_i$. The same input wavefront created by a random combination of Zernike polynomials was used for all scenarios. The results are presented in Fig. 7.1. In this plot, we show the residual phase error vs. maximum static gain error. When there is no gain error, the residual error is the same in both cases. This residual error is a combination of the LOWFS estimation error and the inevitable fitting error caused by fitting a DM shape using influence functions. The residual phase error increases linearly with the DM gain calibration errors when pure Zernike modes are used as basis sets, whereas the phase error remains the same for the changed basis formulation when $\max \epsilon_i < 10\%$. Even though the errors increase for this new formulation when the gain errors are greater than 10%, it still achieves a factor of 2 improvement from the pure Zernike formulation.

Figure 7.1: a) Input wavefront used for simulations. b) Residual wavefront aberration when a 12x12 DM is used to cancel the estimated aberration. Two sets of basis sets were used to estimate the aberrations: 1) pure Zernike modes and 2) Voltage map applied on a 12x12 DM to create Zernike-like shapes.

We performed an experiment to verify the effectiveness of using the Voltage maps/DM-created Zernike modes as a basis set. Since we do not have resources to create pure Zernikes, we simulated an aberrated response of the system by introducing an input wavefront obtained as a linear combination of pure Zernike modes. Then we
Figure 7.2: a) Simulated input wavefront created by linear combination of pure Zernike modes, b) wavefront estimated by using DM-created Zernike shapes using a 12x12 DM, and c) residual wavefront aberration after correction. All wavefronts are in the units of waves.

used the experimental response matrix to estimate the aberrations. Figure 7.2 shows the input wavefront, the estimated wavefront, and the residual error. The residual error is $\sim 18\%$ of the input wavefront. The errors are mainly caused by differences in the beam path between the simulated portion (aberrated image) and the measured portion (response matrix). As the DM and coronagraph are equal in size, a small misalignment might have contributed to the larger error. In addition, the aberrations due to the lenses were not included in the simulations. We could not quantify the alignment errors to include in the simulations. However, in a real mission, both the aberrated wavefront and the training wavefront propagate through the system avoiding this issue of different beam path.

As described earlier in Ch. 1 and Ch. 4, DM shapes are created by linear combination of the influence function of each actuator. We thus tried to expand the above analysis to use the DM influence functions as modes for the LOWFS. The DM influence function basis set is also used by the spatial linear dark field control (LDFC) and the spectral linear LDFC to maintain the high contrast dark holes [92, 93]. To estimate the low-order aberrations, first the response matrix, $H_{Inf}$, was obtained by
actuating each actuator and recording the response at the LOWFS detector. Then, the coefficients of the modes were obtained by solving the following equation:

\[ I_{ab} - I_0 = H_{inf}x, \]  

(7.1.5)

where \( I_{ab} \) is the aberrated image, \( I_0 \) is the nominal image, and \( x \) is the coefficients of the modes. The coefficients were obtained using the singular value decomposition (SVD). Using the SVD, we have

\[ H_{inf} = U\Sigma V^*, \]  

(7.1.6)

where \( U \) and \( V \) are singular matrices and \( \Sigma \) is the diagonal matrix with singular values as diagonal entries. It can be seen from Fig. 7.3a that the singular values are close to zero after certain modes. Hence to make the matrix invertible, only significant modes
(25 modes) were used. The estimates were obtained as

\[ \hat{x} = V_{\text{red}} \Sigma_{\text{red}}^{-1} U_{\text{red}}^* (I_{\text{ab}} - I_0), \]  

(7.1.7)

where the subscript \text{red} means reduced, i.e., includes only the significant modes. To verify the effectiveness of this method, an input wavefront, shown in Fig. 7.4a, was created using the linear combination of \( Z_2 - Z_{15} \). This wavefront was propagated through the SAM WFS system to obtain the aberrated image. Using Eq. 7.1.7, the coefficients were calculated and the wavefront reconstructed. It can be observed from Fig. 7.4b and c that the phase estimate is different from the phase input and the residual errors are bigger than the phase input itself, i.e., the estimate is inaccurate. We tried similar simulation using the ZWFS and got similarly inaccurate results. The inaccuracy in the results is due to the negligible difference in the response of different modes — as each individual poke has high-spatial frequency content, the LOWFS detector does not capture the higher spatial frequency response. Looking at the singular values, it can be observed that only 25 or fewer modes are significant. Trying to fit the low-order aberrations with just 25 actuator commands might be contributing to the error. To solve this issue, we tried to include more modes for estimation. However, including more modes meant that the ratio of singular value between the

Figure 7.4: The SAM WFS estimates with DM influence function as LOWFS modes. a) Input phase (in radians), b) the SAM WFS estimates, and c) estimate error.
most and the least significant modes exponentially increased, consequently, resulting in erroneous estimates. It can be concluded from these results that using the DM to create LOWFS modes helps mitigate the problem of calibration error, but the modes need to be selected carefully. Now we move on to the next chapter, where we will describe the closed loop estimation and control of the low-order wavefront aberrations using a Kalman filter.
Chapter 8

Closed loop control

The analysis presented up to this chapter used a least squares estimator to estimate the low-order aberrations. Since the Kalman filter uses a time averaging to reduce the noise floor, using a Kalman filter to estimate these aberrations would significantly improve the estimates and reduce the residual errors. In this chapter, we will use a Kalman filter to estimate the LOWF aberrations, use a tip-tilt mirror to control the line-of-sight errors ($Z_2$ and $Z_3$), and a DM to control $Z_4 - Z_{15}$. We begin in Section 8.1 with a brief introduction to the Kalman filter and how it could be used by the LOWFS/C system. In this section, we will describe the line-of-sight (LoS) pointing error problem as anticipated by the WFIRST and develop the LoS pointing error estimation and control scheme. In this scheme, we used a Kalman filter as an estimator and an off the shelf tip-tilt mirror to control the LoS pointing error. In Section 8.2, we will show how this scheme can be improved by using a gradient-based algorithm to estimate the system parameters. In subsection 8.2.1, we present the mathematical description of the adaptive estimation and control law. Following the mathematical description, we present the numerical verification of the estimation and control law (Section 8.3).

---

The work presented in this chapter was done in collaboration with Peter Varnai and presented at a conference (SPIE conference proceedings).
8.1 Applications of Kalman Filter in LOWFS/C

The low order wavefront aberrations can be directly compensated by adjusting the tip/tilt mirror or poking the Zernike shapes on the DMs. Given the low order wavefront control command, $u_k$, and process noise, $w_{k-1}$, the low order state transition model is

$$x_k = f(x_{k-1}, u_k) + w_{k-1}, \quad (8.1.1)$$

where the update function $f$ represents the state transition from one time step to the next. Including measurement noises, $n_k$, the low order observation model is,

$$z_k = H_k x_k + n_k, \quad (8.1.2)$$

where $z_k$ is the differential image at each iteration and $H_k$ is the LOWFS modal matrix. The Kalman filter equations for low order wavefront sensing are

State Estimate Extrapolation: $\hat{x}_k(-) = \Phi_k \hat{x}_{k-1}(+) + u_k, \quad (8.1.3)$

Covariance Estimate Extrapolation: $P_k(-) = \Phi_k \hat{P}_{k-1}(+) \Phi_k^T + Q_{k-1}, \quad (8.1.4)$

Filter Gain Computation: $K_k = P_k(-)H_k^T [H_k P_k(-)H_k^T + R_k]^{-1}, \quad (8.1.5)$

State Estimate Update: $\hat{x}_k(+) = \hat{x}_k(-) + K_k [z_k - H_k \hat{x}_k(-)], \quad (8.1.6)$

Covariance Estimate Update: $P_k(+) = [I - K_k H_k] P_k(-), \quad (8.1.7)$
where $\hat{x}_k(-)$ and $\hat{x}_k(+)\) are the priori and posteriori state estimates, $\Phi_k$ is the state transition matrix, $P_k(-)$ and $P_k(+)\) are the priori and posteriori state covariance, and $K_k$ is the Kalman gain. $R_k$ is a diagonal matrix with each diagonal element being the sum of variance of the readout noise and the photon noise and $Q = Q_0 I_{n \times n}$, where the parameter $Q_0$ can be tuned for better performance. Since state transition matrix, $\Phi_k$, depends on state dynamics, we derive/define $\Phi_k$ for specific problems in their respective sections.

8.1.1 Line-of-sight point error

For the WFIRST observatory, the wavefront dynamics can be decomposed into line-of-sight (LoS) pointing drift (with frequencies $< 2 \text{ Hz}$), LoS pointing jitter (with frequencies $> 2 \text{ Hz}$), low-order wavefront (LOWF) drift (with frequencies $< 2 \text{ Hz}$), and LOWF jitter (with frequencies $> 2 \text{ Hz}$) [84]. The slow LoS pointing drift is caused by the attitude control system (ACS) pointing error. The ACS design allows the telescope to drift up to 14 mas rms per axis [84]. The fast LoS pointing jitter is caused by unwanted forces and unbalanced torques from reaction wheel assemblies (RWAs), which induce sinusoidal vibrations at frequencies related to the wheel spin speed. At each RW speed, the LoS pointing jitter contains frequency equal to the wheel speed and multiple harmonic frequencies. For the WFIRST, the RW speed changes from 10 Hz to 40 Hz over 18 hours causing the LoS pointing to jitter with amplitudes up to 8 mas. The design of the WFIRST coronagraphs assumes that the residual LoS pointing error is between 0.4 mas and 1.6 mas rms per axis (tip and tilt modes are the axes considered). If left uncorrected, the LoS pointing error would increase the contrast instability and prevent the coronagraphs from achieving their design contrast. Therefore, it is essential to control the LoS pointing error to less than 0.4 mas per axis.

The slow pointing drift can be corrected by using a simple low-bandwidth feed-
back control that corrects for frequencies < 2 Hz. For the fast LoS pointing error, canceling the fundamental frequency of the LoS pointing jitter and its harmonics is necessary [94]. For the WFIRST observatory, the current estimation and control technique uses the wheel speed frequency as an input and a recursive least squares estimator (RLSE) to estimate the fast LoS pointing error [94]. LoS pointing error caused by mechanical vibrations limit the adaptive optics (AO) performance in ground telescopes as well. Ground-based systems, such as Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE) and Gemini Planet Imager (GPI), use a linear quadratic gaussian (LQG) control law [95, 96] to filter out these vibrations.

8.1.2 System Description

Figure 8.1: A simple representation of a telescopic system under disturbance. a) Blue line represents the incoming starlight (incidence angle \( i \)) to the tip/tilt mirror M1-O-M2 (normal \( n \)). The image formed at the camera is centered at c1. b) M1’-O-M2’ (normal \( n’ \)) is the new orientation after the telescope jitters due to a disturbance, \( d \). Now, the image formed at the camera is centered at a different location (c2). The angle, \( \xi \), between c2 and the desired image location (c1’) is the resulting state variable to be controlled to zero. We achieve this objective by applying control command, \( u \), to the mirror.

In this section, we present a high-level overview of the fast LoS pointing jitter
estimation and control. Here, we define the variables, derive the governing equations, and find the control command necessary to cancel the LoS pointing jitter. To achieve this objective, we consider a simple telescopic setting with a tip/tilt mirror for the LoS jitter control and an imaging camera. The incoming light from the observation star (blue line in Fig. 8.1) is incident to the tip/tilt mirror M1-O-M2 at an angle of $i$ and reflected towards the camera. The image formed at the camera is centered at c1. When the telescope jitters with a disturbance, $d$, the mirror rotates by an angle $d$ to a new location M1'-O-M2' (Fig. 8.1b). Since the whole optical system rotates together, the camera also rotates by an angle $d$, as shown in Fig. 8.1b. Now, the light from the star approaches the mirror at a different angle and the image formed at the camera is centered at a different location (c2). We force the angle, $\xi$, between the new image location and the desired image location to be zero by applying mirror command, $u$. Analyzing the system, we find that $u = \frac{\xi - d}{2}$.

With the relationship between the control variable and the command input established, we designed a block diagram for the estimation and control scheme as depicted in Fig. 8.2. The LoS pointing jitter is represented as the output of the spacecraft control system. At each iteration, $k$, the differential image, $z_k$, is the difference between the aberrated image formed due to the LoS pointing error ($\xi_k$) and the nominal image. The sensor noise is denoted as $n_k$. From Fig. 8.1b, we have $\xi_k = d_k + 2u_k$, where $d_k$ is the LoS pointing jitter due to the RWs and $u_k$ is the mirror command. The LOWFS estimator is a Kalman filter which estimates $\xi_k$ from the measurements. After the estimator makes the estimate, $\hat{\xi}_k$, the controller is used to conjugate the estimated LoS pointing jitter. The control information is then fed to an off the shelf tip-tilt mirror (represented in red dotted lines in Fig. 8.2).
Figure 8.2: Block diagram for LoS jitter estimation and control. The differential image measurement is denoted by $z_k$, the LOWFS camera noise by $n_k$. The LOWFS estimator is a Kalman filter which estimates $\xi$. The estimate, $\hat{\xi}$, is used to calculate the mirror command $u_k$.

8.1.3 State-Space Model

Since the tip-tilt mirror controllers work in discrete time and the camera exposure time was assumed to be 1 millisecond (ms), we discretized the equations using a sampling time of 1 ms. The sinusoidal LoS pointing jitter, $d$, with a frequency, $\omega$, represented by $\ddot{d} + \omega^2 d = 0$, can be discretized to obtain an autoregressive model as

$$d_{k+1} = a_1 d_k + a_2 d_{k-1},$$  \hspace{1cm} (8.1.8)

where $k$ is the time step, $d_{k+1}$ is the LoS pointing jitter at step $k+1$, $d_k$ is the LoS pointing jitter at step $k$, $d_{k-1}$ is the LoS pointing jitter at step $k-1$, $a_1 = 2\cos(\omega \Delta t)$, $a_2 = -1$, and $\Delta t$ is the sampling time.

Using the block diagram presented in Fig. 8.2, the state space model of the LoS
jitter control system can be represented by

\[
\begin{align*}
\xi_{k+1} &= a_1 d_k + a_2 d_{k-1} + 2u_k \\
d_{k+1} &= a_1 d_k + a_2 d_{k-1} \\
z_k &= H\xi_k + n_k
\end{align*}
\] (8.1.9)

where \(\xi_k\) is the controlled variable, \(u_k\) is the control command, \(n_k\) is the camera noise, \(z_k\) is the differential image. Considering \(\xi_k, d_k, \) and \(d_{k-1}\) as our states, in state-space form, Eq. 8.1.9 becomes

\[
\begin{bmatrix}
\xi_{k+1} \\
d_{k+1} \\
d_k
\end{bmatrix} =
\begin{bmatrix}
0 & a_1 & a_2 \\
0 & a_1 & a_2 \\
0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\xi_k \\
d_k \\
d_{k-1}
\end{bmatrix} + 
\begin{bmatrix}
2 \\
0 \\
0
\end{bmatrix} u_k
\]

\[
z_k = H\xi_k + n_k.
\] (8.1.10)

When the LoS pointing jitter is a combination of two sinusoids i.e. \(d = d_1 + d_2\), where \(d_1\) oscillates with a frequency \(w_1\) and \(d_2\) has a frequency of \(w_2\), the new system is represented by the following state-space model:

\[
\begin{bmatrix}
\xi_{k+1} \\
d_{1k+1} \\
d_{1k} \\
d_{2k+1} \\
d_{2k}
\end{bmatrix} =
\begin{bmatrix}
0 & a_{11} & a_{12} & a_{21} & a_{22} \\
0 & a_{11} & a_{12} & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & a_{21} & a_{22} \\
0 & 0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\xi_k \\
d_{1k} \\
d_{1k-1} \\
d_{2k} \\
d_{2k-1}
\end{bmatrix} + 
\begin{bmatrix}
2 \\
0 \\
0 \\
0 \\
0
\end{bmatrix} u_k
\]

\[
z_k = H\xi_k + n_k,
\] (8.1.11)

where \(a_{11} = 2\cos(\omega_1 \Delta t)\), \(a_{12} = -1\), \(a_{21} = 2\cos(\omega_2 \Delta t)\), and \(a_{22} = -1\). Following the pattern from Eq. 8.1.10 and Eq. 8.1.11, the state-space model can be created for a system to estimate and control both tip-tilt modes when the jitter has multiple
frequencies and harmonics.

![Figure 8.3: Tip estimate and control using a Kalman filter and a tip-tilt mirror a) when the RW speed input is correct and b) the RW speed input is incorrect.](image)

This control law, similar to the ones proposed in references [94, 95, 96], also assumes a priori knowledge of the wheel speed. The reliance of these techniques on a priori knowledge of the wheel speed can be problematic if the RW speed information is missing or incorrect as shown in Fig. 8.3b.

In the following section, we propose a new adaptive technique that estimates the parameters of the vibration frequency, and the LoS pointing error. This technique can be used for both ground and space-based systems to estimate and control LoS pointing error due to sinusoidal disturbances when the vibration frequency is unknown or inaccurate.

### 8.2 Adaptive estimation with Kalman filter for LoS jitter estimation

Incorrect tachometer readings (imperfect knowledge of the jitter frequency $\omega$) result in incorrect values of $a_{11}$ and $a_{12}$, which subsequently leads to erroneous estimation. As shown in the above section in Fig. 8.3b, at higher RW speeds, a 5% error in
Figure 8.4: Block diagram for adaptive LoS jitter estimation and control. The differential image is denoted by $z_k$ and the LOWFS camera noise by $n_k$. The LOWFS estimator is a Kalman filter which estimates $\xi$. The parameter estimator estimates the parameters of the system which are functions of the LoS RW speed. The estimate, $\hat{\xi}$, is used to calculate the mirror command $u_k$.

Frequency readings causes the residual LoS pointing error to exceed the required 0.4 mas. In this section, we describe the gradient-based algorithm used to determine the parameters of the sinusoidal vibration.

### 8.2.1 Parameter estimation

In order to avoid the residual error due to erroneous frequency input, we use existing techniques from adaptive control in order to recover the true value of the parameters $a_{11}$ and $a_{12}$. Figure 8.4 shows the modified block diagram for the adaptive LoS jitter estimation and control scheme. It is similar to the previous block diagram, however, instead of using the frequency information as an input, there is a parameter estimator that uses $z_k$, $u_k$, and a gradient-based algorithm to estimate the parameters of the system which are functions of the vibration frequency. To begin, we rewrite
the measurement of (8.1.11) for the \((k+1)\)th iteration in the so-called linear-in-the-
parameters form:

\[
z_{k+1} - 2H_k u_k = H_k \left( \begin{bmatrix} a_{11} & a_{12} & a_{21} & a_{22} \\ \end{bmatrix} \begin{bmatrix} d_{1k} \\ d_{1k-1} \\ d_{2k} \\ d_{2k-1} \end{bmatrix} \right) + n_{k+1} \tag{8.2.1}
\]

or

\[
\zeta_k = H_k (\theta^T \cdot \varphi_k) \tag{8.2.2}
\]

for short. Here, \(\theta\) denotes the true parameters of the system, \(\zeta\) is the observation, and 
\(\varphi\) is the regressor which describes how the parameters affect the dynamical system.

Similarly, the predicted output, \(\hat{z}_{k+1}\), can be written in terms of the estimated states as:

\[
\hat{z}_{k+1} - 2H_k u_k = H_k \left( \begin{bmatrix} \hat{a}_{11k} & \hat{a}_{12k} & \hat{a}_{21k} & \hat{a}_{22k} \\ \end{bmatrix} \begin{bmatrix} \hat{d}_{1k} \\ \hat{d}_{1k-1} \\ \hat{d}_{2k} \\ \hat{d}_{2k-1} \end{bmatrix} \right) \tag{8.2.3}
\]

or

\[
\hat{\zeta}_k = H_k (\hat{\theta}_k^T \cdot \hat{\varphi}_k), \tag{8.2.4}
\]

where the superscript \(^\hat{}\) denotes an estimate of the variables. The observation error, 
\(\epsilon_k\), is given by \(\epsilon_k = \zeta_k - \hat{\zeta}_k\). Since the Kalman filter uses a time averaging to reduce
the noise floor, the error, \(\epsilon_k\), is mainly due to the erroneous parameters:

\[
\epsilon_k \approx H_k \left( (\theta^T - \hat{\theta}_k^T) \cdot \hat{\varphi}_k \right). \tag{8.2.5}
\]
As $H_k$ is time-invariant, we define a new error metric,

$$E_k = H_k^+ \epsilon_k,$$

$$= (\theta^T - \hat{\theta}_k^T) \cdot \hat{\phi}_k,$$  \hspace{1cm} (8.2.6)

where $H_k^+$ is the pseudo-inverse of the response matrix. We find the estimate $\hat{\theta}$ by minimizing the quadratic cost function:

$$J(\hat{\theta}_k) = \frac{1}{2} \| (\theta^T - \hat{\theta}_k^T) \cdot \hat{\phi}_k \|^2.$$  \hspace{1cm} (8.2.7)

Employing a gradient-based adaptive law [97], we update the parameter vector $\hat{\theta}$ in a direction that minimizes this cost function. The gradient with respect to $\hat{\theta}_k$ is:

$$\frac{\partial J(\hat{\theta}_k)}{\partial \hat{\theta}_k} = - \left[ (\theta^T - \hat{\theta}_k^T) \cdot \hat{\phi}_k \right] \hat{\phi}_k \approx -E_k \hat{\phi}_k.$$  \hspace{1cm} (8.2.8)

Our estimate is then updated in the opposite direction with a step size represented by the adaptive gain $\gamma$:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma \cdot E_k \hat{\phi}_k.$$  \hspace{1cm} (8.2.9)

Using Eq. 8.2.9 and a tuned $\gamma$, we can estimate the parameters $a_{11}$ and $a_{21}$.

### 8.3 Numerical Validation

To validate our adaptive estimation and control technique, we used WFIRST as the baseline. As the expected WFIRST RW speed changes from 10 Hz to 40 Hz, we estimated and controlled the LoS pointing error due to a 10 Hz, 20 Hz, 35 Hz, and 40 Hz vibrations and their respective sub-harmonic and a higher harmonic. We estimated and controlled both the tip and tilt error due to the LoS pointing error, with the goal of controlling the errors below 0.4 mas. As the baseline architecture for WFIRST has
an SPC, we used an SPC for our simulation.

The optical setup used for the simulation is similar to the one described in Ch. 1. The SPC apodizes the incoming light. Then, the FPM reflects the light from the core of the PSF \((4 \frac{\lambda}{D})\) towards the optimized SAM at a relay pupil. The light diffracted by the SAM is brought to a focus on a camera. The camera had a readout noise of 3ADU/pixel/frame. We sampled the image at \(2\frac{\text{pixels}}{\lambda} D\), so that the image was Nyquist sampled and it would be faster to read fewer pixels. To calculate the number of photons arriving at the LOWFS detector, we used a 4.83 apparent magnitude star as the observation target using a 2.4 m telescope over 100% bandwidth centered at 550 nm. We used an exposure time of 1 millisecond. Although the photon number calculations were made using broadband light, for simplicity and as initial proof of concept, the control simulations presented in this chapter are monochromatic.

Figure 8.5: Tip estimate and control using Kalman filter and parameter estimation. The RW has a fundamental tone of 40 Hz and its higher harmonic. a) The estimate converging to the true value in open loop. b) Open and closed loop estimation and control. c) Zoomed-in section of the residual error. The residual error is \(\sim 0.04\) mas which an order smaller than the required value.

For all wheel speeds considered, we were able to correct the LoS pointing error to an order smaller than the required 0.4 mas. In this chapter, we will show the results for a 40 Hz vibration and its higher harmonic. To calculate the initial values of the parameters, we assumed incorrect initial tachometer readings \((\omega_1 = 35 \text{ Hz} \text{ and } \omega_2 = 70 \text{ Hz}, \text{ an error } > 10\%\)\). We ran the open loop estimation for the first 0.25 s and
Figure 8.6: a) Tilt estimate and control using Kalman filter and parameter estimation. b) Estimates of the parameters $a_{11} = 2\cos(\omega_1 \Delta t)$ and $a_{11} = 2\cos(\omega_1 \Delta t)$. The initial values of the parameters are obtained using the frequency readings which are assumed to have an error of 12.5%.

turned on the controller. To reduce the effect of noise on the parameter estimation, we averaged the parameter estimates over the last 25 iterations, i.e., we averaged the parameters for 25 ms. Since the expected WFIRST disturbance frequency changes slowly, this averaging will be able to detect the changes in the parameters. We ran this simulation for 1500 iterations and assumed that the vibration frequency stayed the same in this time period. The Kalman filter and the adaptive gain were tuned to achieve a good performance. The LoS error estimates are shown in Figs. 8.5 and 8.6. Figure 8.6 also shows the estimates of the parameters $a_{11}$ and $a_{21}$. The parameters converged in around 0.5s and we were able to control the LoS pointing error in both directions to 0.04 mas.

As the WFIRST RW speed changes over time, we simulated a scenario where the disturbance frequency changed over time. Instead of simulating the entire 18 hour period in which the RW speed changes from 10 Hz to 40 Hz, we sped up the scenario and changed the wheel speed from 20 Hz to 30 Hz in 10 seconds (10,000 iterations). Similar to the previous case, we estimated the fundamental frequency and its higher harmonic. Since the true parameters were changing in each iteration, we did not
average the parameter estimates. Even though this led to some noisy estimates, the residual LoS pointing error (\(\sim 0.15\) mas) was significantly smaller than the required 0.4 mas. The estimates are shown in Fig. 8.7. For a real observatory case, we would not have this issue and we would get a better residual.

![Figure 8.7: LoS pointing error and RW speed parameter estimation for sinusoidal disturbance with changing frequency. a) Estimates of the parameters \(a_{11}\) and \(a_{21}\) when the wheel speed changes from 20 Hz to 30 Hz in 10 seconds. b) LoS pointing error in tip and the preferred 0.4 mas LoS pointing error. The residual after correction is less than the required LoS pointing error of 0.4 mas.](image)

8.4 Control of all low-order modes

After we closed the faster tip-tilt estimation and control loop, we integrated it with the slower outer LOWFS loop. The system equations for this scheme are

\[
x_k = \Phi_k x_{k-1} + \Gamma_k u_k, \\
z_k = H_k x_k + n_k,
\]

(8.4.1)

where \(k\) is the time step, \(x_k\) is the state vector that has the coefficients of \(Z_4 - Z_{15}\), \(\Phi_k\) is the state transition matrix and is equal to identity, \(\Gamma_k\) is the input matrix, \(u_k\) is the DM command, \(z_k\) is the differential image, \(H_k\) is the LOWFS response matrix,
and $n_k$ is the noise. Using the Kalman filter described in Section 8.1 and the system in Eq. 8.4.1, the Zernike coefficients are estimated. As the control bandwidth for the outer loop is different than the faster tip-tilt control scheme, the differential image $z_k$ for the $k^{th}$ outer loop time step is obtained by subtracting the nominal image from the time averaged LOWFS images which are obtained every 1 ms. Hence the differential image $z_k$ is obtained as

$$z_k = \frac{1}{N} \sum_{i=1}^{N} I_{abi} - I_0,$$  (8.4.2)

where $N$ is the number of aberrated LOWFS images ($I_{ab}$) stacked and $I_0$ is the nominal image.

Once the Zernike coefficients are estimated, a phase map is reconstructed using these coefficient estimates and a DM is used to conjugate the estimated phase. Using the following relations, the DM commands are calculated at each time step as

$$\phi_{estk} = Z\hat{x}_k,$$
$$u_k = (G\beta)^+\phi_{estk},$$
$$u_k = (G\beta)^+Z\hat{x}_k,$$  (8.4.3)

where $\phi_{estk}$ is the estimated phase at time step $k$, $Z$ is the Zernike polynomial matrix, $\hat{x}_k$ is the estimate of the Zernike coefficients at time step $k$, $G$ is the matrix with the normalized influence function of each DM actuators, $\beta$ is a diagonal matrix that has height-voltage relationship, and the superscript, $+$, represents the pseudo-inverse of a matrix. From the above equations, we get

$$\Gamma_k = ((G\beta)^+Z)^+,$$
$$= Z^+G\beta.$$  (8.4.4)

To simulate this scheme, we made the same assumptions as in Section 8.3 to
obtain each individual LOWFS images. However, to obtain the aberrated image for $Z_4 - Z_{15}$ coefficient estimations, we only averaged 100 LOWFS images, i.e., the outer loop bandwidth was faster than the 5 mHz level described in Ch. 1. As each LOWFS image had an exposure of 1 ms, stacking images for 5 mHz bandwidth would mean stacking 200000 images, which would be time and computer memory consuming to simulate. Even though we only stacked 100 LOWFS images for this outer loop simulations, we used equivalent noise statistics for the combined image such that it would be the same as stacking 200000 images. We simulated the outer loop for 25 iterations. For this simulation, we assumed similar LoS disturbance as in Section 8.3, but started the closed loop control from the first iteration. For other modes, we assumed an initial random phase $\sim 1$ nm RMS obtained by a linear combination of $Z_4 - Z_{15}$ and assumed $\Phi_k = I$. We used a tip-tilt mirror to control the LoS pointing jitter and a $48 \times 48$ Xinetics DM to control other modes. Figure 8.8 shows the results of the simulation. The LoS pointing error is controlled to less than 0.4 mas and other modes to less than 10 pm RMS.

Figure 8.8: Estimation of low-order aberrations for a spaced-based telescope. a) Estimation and control of LoS sight pointing error using a Kalman filter and a tip-tilt mirror. b) Estimation and control of phase aberrations (combination of $Z_4 - Z_{15}$) using a Kalman filter and a $48 \times 48$ Xinetics DM.
Chapter 9

Conclusion and future work

To directly image an Earth-like exoplanet, a space-based telescope must be equipped with a coronagraph capable of starlight suppression by a factor of $10^{10}$. Exquisite control of wavefront aberrations, caused by manufacturing errors, misalignments, line-of-sight jitter of the telescope, mechanical flexure, and thermal variation is necessary for a coronagraph to reach its designed performance. Wavefront estimation and control algorithms have been developed and tested in various testbeds to determine the DM actuators commands to create the high contrast at the image plane. These techniques only estimate and control the static or quasi-static aberrations and are blind to the dynamic low-order aberrations. We presented a novel coronagraph-integrated wavefront sensor, relying on a sparse aperture mask (SAM, to infer the dynamic low-order aberrations from a differential intensity distribution of the re-imaged, discarded starlight. We proposed a simple algorithm to sense the coefficients of the low-order aberrations and verified it with a Fourier propagation model. Wavefront estimation accuracy has been quantified for pure Zernike modes, random linear combinations of Zernike modes, and a phase screen with a Kolmogorov power spectrum. The SAM WFS is best suited for estimating dynamic aberrations. The estimation algorithm relies on the data acquired from a nominal point meeting the target wavefront goal.
(which is not necessarily flat phase, although in this thesis we have written our sensing equation in a way that assumes this). The initial reference point corresponding to the flat phase could be reached using any wavefront calibration method based on the science focal plane. Then the SAM WFS response matrix could be measured using the DM. We showed that using DM voltage maps as the basis set rather than pure Zernikes reduces the residual error, due to gain calibration errors, by a factor of 2.

The differential approach to estimation reduces the inherent vulnerability to non-common path errors. For a system with these limitations, a space coronagraph is the most compelling application and was the primary motivation for this thesis. In this situation, the high-order spatial frequency corrections require long integration times due to the faint signal in the relevant region of the science focal plane. For the high-order wavefront control system of a space-based exoplanet imaging instrument to work effectively, the low-order aberrations need to be stabilized between deformable mirror correction steps. Therefore, the SAM WFS, like the CLOWFS and the phase contrast/Zernike WFS, could be used to estimate changes in the wavefront on a time scale much shorter than the high-order wavefront control loop. The SAM WFS also has potential uses for ground-based, extreme AO exoplanet imaging instruments. In the case where flexure-induced, non-common path aberrations are a concern for the coronagraph performance, the SAM WFS could help to lock in on some initial low-order correction by offsetting the DM commands of the AO system.

Our initial SAM design for an SPC, used to serve as a proof of concept, used 27 subaperture holes, each with a diameter that is 7% of the pupil. In our initial comparisons with CLOWFS, we found that the accuracy of the SAM estimation was promising, but the light loss through the aperture hindered the performance in estimation scenarios with short exposure times. We found that the condition number of the response matrix was a useful diagnostic of the estimation accuracy of a given mask shape. For example, the maximally redundant aperture, an open circle,
has an extremely high condition number because the corresponding system matrix is ill-conditioned. To increase the throughput of the mask while maintaining its accuracy, we optimized the SAM. The optimization technique maximized the average of the singular values of the response matrix, which served as a proxy for the mask throughput, while constraining the response matrix condition number. Compared to the initial SAM design for the SPC, the optimized SAM reduced the condition number of the response matrix by 60% while increasing the throughput by 350%. Once we optimized the SAM, we did a realistic comparison between the SAM WFS, the CLOWFS, and the ZWFS for different families of coronagraphs. To obtain better estimates, we used 100% bandwidth light for the SAM WFS and the CLOWFS and 80% bandwidth for the ZWFS. At these bandwidths, the SAM WFS is more precise for the configurations with the SPC. For configurations where the LOWFS picked up the rejected light from the Lyot stop, the SAM WFS detector placed at a small optimal defocus distance, like the LLOWFS, performed better than the scenario in which the detector was placed at a re-imaged focal plane. The defocus length for the SAM WFS was obtained by a simple binary search. In the following future work subsection, we will discuss how the optimization problem described in Ch. 5 can be modified to solve for an optimal SAM and an optimal defocus distance. For the SPC configuration, the ZWFS estimation error due to noise was about 1.5 times smaller than other sensors. Hence for a coronagraphic mission that uses the rejected starlight from the focal plane for the LOWFS, we recommend the ZWFS. On the other hand, when used with a charge 6 VVC coronagraph, the ZWFS estimation error was twice as large as compared to the LLOWFS and the defocused SAM WFS. For a mission that would use the rejected starlight from the Lyot plane for the LOWFS, we suggest that a hybrid of the LLOWFS and the SAM WFS would be more accurate and robust.

We also experimentally verified the efficacy of the SAM WFS. For the tip/tilt estimation, the SAM WFS estimations were linear up to \( \sim 0.8 \) radians with a slope
of $\sim 0.982$ and the estimations were consistent with the simulation results. For other modes, the linearity held up to $\sim 0.05$ waves rms as predicted in simulations.

We also presented a new idea to estimate and control the line-of-sight pointing jitter. It is essential to control the line-of-sight pointing error to sub-milliarcsecond level for a coronagraph to achieve its desired contrast. For a space telescope, the largest pointing errors are caused by the reaction wheel disturbances, which are harmonic in nature. In the absence of accurate frequency measurements of the vibrations caused by the disturbances, it is difficult to estimate and control the LoS pointing error to the required level. In this thesis, we presented an adaptive estimation technique which uses a gradient-based method to estimate the parameters of the vibration and a Kalman filter to estimate the LoS pointing error. In simulation, we verified the effectiveness of this technique and were able to control the LoS error to an order of magnitude smaller than the desired value of 0.4 mas. We also verified in simulation that a Kalman filter can be used to estimate other low-order modes and a DM can be used to control these aberrations with residuals less than 10 pm. Even though we used the SAM WFS with an SPC to demonstrate this concept, it can be used with any of the LOWFS techniques with all families of coronagraph.

The SAM WFS is a nascent concept, hence it still needs to be explored for improved performance and tested in the lab for different experiments. We will conclude the thesis by exploring the possible SAM WFS future direction.

## 9.1 Future work

Combining the SAM WFS and the LLOWFS for a VVC or any other families of coronagraph that rejects stellar light at Lyot plane would yield an accurate and robust sensor. One of the possible formulations, as mentioned in Ch. 6, would be the modification of optimization procedure presented in Ch. 5 to optimize the mask
and find the optimal defocus length. In the following section, we will discuss how the modification could be made.

### 9.1.1 Improved Optimization of the defocused SAM WFS

When the SAM WFS detector is placed at a certain defocused length, $z$, the response matrix, $H_{SAM}(z)$, is

$$H_{SAM}(z) = \begin{bmatrix} Re \left\{ E_0 \mathcal{F} \{ \mathcal{MF} \{ \mu_{SAM} \mathcal{F} \{ AZ_2 \} } \right\} \ldots \ldots \end{bmatrix} ,$$

where $E_0$ is the nominal electric field at the SAM WFS detector, $\mathcal{F}R_z$ is the Fresnel propagator, $\mathcal{F}$ is the Fourier propagator, $\mathcal{M}$ is the SAM, $\mu_{SAM}$ is the focal plane mask, $A$ is the aperture, and $Z_i$ the Zernike polynomials. The $k$th column of the matrix $H_{SAM}(z)$ can be written as

$$H_{SAMk}(\mathcal{M}, z) = \text{vec} \left[ A_1(\mathcal{M}, z) \odot A_{2,k}(\mathcal{M}, z) - A_3(\mathcal{M}, z) \odot A_{4,k}(\mathcal{M}, z) \right] ,$$

where

$$A_1(\mathcal{M}, z) = a_1(z) (E_{0SAM} \odot \mathcal{M}) a_4 - a_2(z) (E_{0SAM} \odot \mathcal{M}) a_3 ,$$

$$A_{2,k}(\mathcal{M}, z) = a_1(z) (E_{SAMk} \odot \mathcal{M}) a_3 - a_2(z) (E_{SAMk} \odot \mathcal{M}) a_4 ,$$

$$A_3(\mathcal{M}, z) = a_1(z) (E_{0SAM} \odot \mathcal{M}) a_3 - a_2(z) (E_{0SAM} \odot \mathcal{M}) a_4 ,$$

$$A_{4,k}(\mathcal{M}, z) = a_1(z) (E_{SAMk} \odot \mathcal{M}) a_4 - a_2(z) (E_{SAMk} \odot \mathcal{M}) a_3 ,$$

Here $\odot$ is the Hadamard product, the matrices $E_{0SAM}$ and each $E_{SAMk}$ for $k = 1, \ldots, p$ give the electric field produced by the nominal phase and the Zernike perturbations.
propagated to the SAM plane, respectively, and

\[a_1(z) = [J_R H_R(z) - J_I H_I(z)] K_R,\]
\[a_2(z) = [J_I H_R(z) + J_R H_I(z)] K_I,\]
\[a_3 = (K_R J_R - K_I J_I)^T,\]
\[a_4 = (K_I J_R + K_R J_I)^T,\]

where \(K_R \in \mathbb{R}^{m \times n}\) and \(K_I \in \mathbb{R}^{m \times n}\) are the real and imaginary parts of the Fourier transform matrix \(K \in \mathbb{C}^{m \times n}\), \(J_R \in \mathbb{R}^{m \times n}\) and \(J_I \in \mathbb{R}^{m \times n}\) are the real and imaginary parts of the inverse Fourier transform matrix \(J \in \mathbb{C}^{m \times n}\), \(H_R(z)\) and \(H_I(z)\) are the real and imaginary parts of the kernel defined in Eq. 1.5.9 in Section 1.5, and the superscript, \(T\), is the transpose of a matrix. The optimization problem formulation for the defocused SAM configuration, similar to the one described in Ch. 5 becomes

\[
\max_{z, \mathcal{M} \in \mathcal{M}} \bar{\sigma}(H_{SAM}(\mathcal{M}, z)) \quad \text{s.t.} \quad \kappa(H_{SAM}(\mathcal{M}, z)) \leq \kappa_0,
\]

where \(\bar{\sigma}(.)\) is the average of the singular values of a matrix and \(\kappa(.)\) is the condition number of the matrix. The optimization formulations presented till now consider monochromatic light. Even though the SAM WFS is a binary mask that does not introduce any phase affects, the condition number of the SAM WFS response matrix increases with an increase in bandwidth. Hence, this optimization formulation should be extended to include a broadband source. It is also worth exploring different formulations to obtain an optimized mask and an optimal distance.

We would also like to integrate the SAM WFS in our existing focal plane wavefront control testbed at Princeton.
9.2 Integration of the SAM WFS with the FPWC

The PHCIL testbed is shown in Fig. 9.1. Collimated light is incident on two BMC MEMS DMs in series, which propagates to the ripple-shaped SPC, the core of the PSF is removed by a focal plane mask, and the dark holes are reimaged on a detector. To integrate the SAM WFS, the light blocked by the focal plane mask should be directed/reflected to a SAM placed at a relay pupil and then focused on a detector. In addition, one of the DMs should be placed at a pupil plane, conjugate to the SPC plane, to cancel the low-order aberrations. To compensate for the tip-tilt error, another pupil plane conjugate to the SPC plane should be created where a tip-tilt mirror would be placed.
Bibliography


[35] https://wfirst.gsfc.nasa.gov/, “.”


[38] https://asd.gsfc.nasa.gov/luvoir/, “.”


