



Kent Academic Repository

Moser, Steven (2017) *Coherence Gated Laser Ray Tracing Based on a High Speed FPGA Platform*. Doctor of Philosophy (PhD) thesis, University of Kent.

Downloaded from

<https://kar.kent.ac.uk/62046/> The University of Kent's Academic Repository KAR

The version of record is available from

This document version

Pre-print

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

COHERENCE GATED LASER RAY TRACING BASED ON A HIGH SPEED FPGA PLATFORM

Steven Moser
Embedded Systems Group
Department of Engineering
University of Kent

Under the supervision of
Dr. Peter Lee
&
Dr. Adrian Podoleanu

Panel Chair: Professor S K Spurgeon
Panel Member: W A J Waller
S.J.Moser@kent.ac.uk

June 7, 2017

Abstract

Wavefront sensing is a critical component in the adaptive optics systems used for modern high resolution imaging modalities. One limitation of current wavefront sensor designs is that they have a large depth of field, rendering them incapable of differentiating light returned from targets with many layers (such as the human retina). Instead, the measured wavefront is the superposition of the wavefronts returned from each layer. By combining principles from low-coherence interferometry and wavefront sensing, a depth-resolved wavefront sensor may be realised. This allows only light from within the *coherence-gate* of the interferometer to be measured by the wavefront sensing device. By adjusting the axial position of the coherence-gate, wavefronts from distinct layers of a multi-layer object may be measured. This method has been demonstrated for the Shack-Hartmann wavefront sensor but requires an external PC for image processing and wavefront reconstruction.

This dissertation presents, for the first time, a depth-resolved laser ray tracing wavefront sensor. Results are shown, in the form of Zernike modes, which demonstrate the ability of the instrument to resolve wavefronts from a multi-layer target (two stacked microscope slides and a mirror). Also, an FPGA based embedded system was developed for all command, control, image processing and wavefront reconstruction functions. This highly specialised system is able to perform these operations in real-time, limited only by the frame rate of the available camera.

Specific attention is given to the portion of the system focused on wavefront reconstruction. Zernike modes are commonly used in adaptive optics systems to represent optical wavefronts. However, real-time calculation of Zernike modes is time consuming due to two factors: the large factorial components in the radial polynomials used to define them, and the large inverse matrix calculation needed for the linear fit. This dissertation presents an efficient parallel method for calculating Zernike coefficients from phase gradients and its real-time implementation using an FPGA by pre-calculation and storage of subsections of the large inverse matrix. The architecture exploits symmetries within the Zernike modes to achieve a significant reduction in memory requirements and a speed-up of 2.9 when compared to published results utilising a 2D-FFT method for a grid size of 8 x 8. Analysis of the processor element's internal word length requirements show that 24-bit precision in pre-calculated values of the Zernike mode partial derivatives ensures less than 0.5% error per Zernike coefficient and an overall error of less than 1%. The design has been synthesized on a Xilinx Virtex-6 XCVLX240T FPGA.

Acknowledgements

Many friends, colleagues and family members have contributed to the amazing journey that has culminated in the completion of this research project. First and foremost, I need to thank my wife Shelby, and children Julian and Finneas. When Shelby and I decided to return to university to pursue our doctorates we coupled it with the decision to move our family to the UK. This made for not only a wonderful educational experience, but a rich three years full of new friends, new experiences, and a period of amazing growth for our children.

Thank you, Shelby, for all your support while I was engaged in the lab. You somehow managed to pull off the Herculean task of both working on your own research and home-schooling the kids. I'll always cherish our long discussions of computers, art, and philosophy over an ale.

Julian and Finneas, thanks to you both for coming along on this great adventure. It was a joy watching you both grow as you explored the English countryside. I hope that the time we spent here has shown you that the world is full of experiences to be had and great people to know.

To my parents, thanks for all your support throughout my years of educational pursuits. Your help with kids, housing, and moving has been invaluable and is appreciated more than you know.

I also want to thank my two advisors, Dr. Peter Lee and Dr. Adrian Podoleanu. In a cross-disciplinary project such as this, your varied viewpoints have been invaluable. Above and beyond the technical expertise, your support and friendship over the last three years has been a source of strength.

Thanks to all my fellow researchers in the Embedded Systems group including Mark Esdale who always made himself available for questions. Without your guidance I would have been lost in a sea of VHDL! To my good friend Robert Horne, your camaraderie made all the difference in the final stretch of this research. To Jingyu Wang, our chats in the lab and the use of your optical equipment are greatly appreciated. To my good friend and mentor Dr. Don Isaak, many thanks for getting me started on the research path from my early days as a physics undergraduate. The time I spent at your UCLA Geophysics lab was invaluable and taught me a great deal about being an independent researcher.

Lastly, I want to thank my friends and co-workers at the Jet Propulsion Lab; Bijan Nemati, Frank Dekens, Dan Ryan, and Ali Azizi. It's been many years since we've worked together but I will always have fond memories of the Kite days and the lessons learned there have been an invaluable tool in the proceeding years. It was the work we did together that set my path toward optics and interferometry and ultimately led me to pursue this research topic.

Contents

1	Introduction	14
1.1	Motivation	14
1.2	Outline	17
2	Wavefront Sensing	21
2.1	An Introduction to Wavefront Sensing	21
2.2	The Shack-Hartmann Wavefront Sensor	26
2.3	The Zonal and Modal Methods	30
2.4	Zernike modes as wavefront descriptors	32
2.5	Optical Aberrations	35
2.5.1	Tilt	36
2.5.2	Defocus	36
2.5.3	Coma	37
2.5.4	Astigmatism	37
2.5.5	Spherical	38
2.6	Conclusion	39
3	Coherence Gated Wavefront Sensing	40
3.1	Introduction to Coherence Gated Wavefront Sensing	40
3.2	CGWFS in Biological Samples	43
3.3	Virtual Shack-Hartmann WFS	44
3.4	Coherence-Gated Shack-Hartmann WFS	45
3.5	Conclusion	50
4	Laser Ray Tracing Wavefront Sensor	51
4.1	An Introduction to Laser Ray Tracing	51
4.1.1	Single-pass LRT	52
4.1.2	Double-Pass LRT	55
4.2	Standard LRT Experimental Setup	57
4.3	Coherence-Gated LRT Experimental Setup	75
4.4	Conclusion	79

5	Field Programmable Gate Arrays and Real Time CGLRT	92
5.1	An Introduction to Field Programmable Gate Arrays	92
5.2	An Embedded Systems Approach to Wavefront Sensing	94
5.3	FPGA design workflow	95
5.4	CGLRT System Architecture	97
5.4.1	Master State Controller	100
5.4.2	Scanner Controller	103
5.4.3	Clock Domains and Clock Management	104
5.4.4	Cameralink Interface	106
5.4.5	Frame Grabber	107
5.4.6	PSI Algorithm	108
5.4.7	Centroiding	112
5.5	Testing	114
5.5.1	Wavefront Reconstruction	120
5.5.2	Modal Method	120
5.5.3	Parallel Modal Method	123
5.5.4	Hardware Implementation	124
5.5.5	Memory Reduction	126
5.5.6	Wavefront Simulation	126
5.5.7	Resource Utilization and Performance	130
5.6	Conclusion	132
6	The CGLRT Instrument	135
6.1	LRT vs. CGLRT Results	143
6.2	CGLRT Results	148
6.3	Timing and Resource Utilisation	153
6.4	Conclusion	156
7	Conclusion	158
7.1	Future Work	159
7.2	Closing Comments	160

List of Figures

1	Research motivation roadmap	15
2	Cell layers of the human retina	18
3	Example of a planar wavefront	22
4	An example of a spherical wavefront	23
5	Wavefront distortion in ground based astronomy	24
6	Illustration of an adaptive optics system	25
7	A portrait of Johannes Hartmann	26
8	A basic Shack-Hartmann wavefront sensor	27
9	Functional illustration of a Shack-Hartmann wavefront sensor for an non-aberrated wavefront.	28
10	Functional illustration of a Shack-Hartmann wavefront sensor for an aberrated wavefront.	29
11	Large aberrations can lead to gridding errors	30
12	The zonal method	31
13	The first 10 Zernike modes	34
14	Tilt aberration	36
15	Coma aberration	37
16	Astigmatism aberration	38
17	Spherical aberration	39
18	A Michelson interferometer	41
19	A Mach-Zehnder interferometer	42
20	The experimental set up used in [59]	47
21	The experimental set up used in [61]	49
22	A basic single-pass LRT	52
23	A basic single-pass LRT with aberrations	53
24	A single-pass LRT from [35]	55
25	The Shack-Hartmann setup from [35].	55

26	A basic double-pass LRT	56
27	A block diagram of the initial double-pass LRT setup.	58
28	A quad-cell type PSD	59
29	A lateral effect type PSD	59
30	Initial LRT system.	61
31	Close-up view of the test object	63
32	LRT raw measured spots	65
33	LRT averaged measured spots.	66
34	Spot pattern for small tilt in X.	67
35	Reconstructed wavefront from 14 Zernike terms for a small X tilt.	68
36	Zernike coefficients for a small X tilt.	69
37	Spot pattern for large tilt in X.	70
38	Reconstructed wavefront from 14 Zernike terms for a large X tilt.	71
39	Zernike coefficients for a large X tilt.	72
40	Spot pattern for tilt in Y.	73
41	Reconstructed wavefront from 14 Zernike terms for a Y tilt.	74
42	Zernike coefficients for a Y tilt.	75
43	CGLRT experimental setup.	77
44	Illustration of the multi-layer test object.	80
45	Raw scanning beam image with 0° phase shift.	81
46	Raw scanning beam image with 90° phase shift.	82
47	Raw scanning beam image with 180° phase shift.	83
48	Raw scanning beam image with 270° phase shift.	84
49	Coherence-gated image from surface P1.	85
50	The first 14 Zernike coefficients for surfaces P1-P5.	86
51	The wavefront measured from surface P1.	87
52	The wavefront measured from surface P2.	88

53	The wavefront measured from surface P3.	89
54	The wavefront measured from surface P4.	90
55	The wavefront measured from surface P5.	91
56	Flow diagram of System Generator based FPGA development	97
57	CGLRT top level system block diagram	99
58	Master State Controller block diagram	102
59	Top-level block diagram of the Scanner Controller	104
60	Scan Pattern Generator	104
61	Clock domains utilised in the CGLRT system	106
62	Cameralink Interface	107
63	Frame Grabber state diagram	108
64	Streaming Image Subtraction	110
65	Pixel Squaring and Binarization	110
66	Moment calculation block diagram	113
67	Divider block diagram	113
68	Slope calculation block diagram	114
69	Centroid test image 1	116
70	Centroid test image 2	116
71	Centroid test image 3	116
72	Raw Coherence Gated Image 1.	118
73	Raw Coherence Gated Image 2.	118
74	Raw Coherence Gated Image 3.	118
75	Binary Coherence Gated Image 1 with Centroid.	119
76	Binary Coherence Gated Image 2 with Centroid.	119
77	Binary Coherence Gated Image 3 with Centroid.	119
78	Circular pupil mapped into a square grid	121
79	Block diagram for the top level view of Zernike processing blocks.	125

80	Block diagram for the Zernike Coefficient Processors.	126
81	Even symmetry in the A_{2x}^\dagger vector	127
82	Odd symmetry in the A_{3x}^\dagger vector	128
83	Simulated input wavefront	129
84	Percent error in calculated Zernike coefficients for given word lengths	131
85	Photograph of the CGLRT optical layout	136
86	ML605 based FPGA system	137
87	ML605 board	138
88	A Xilinx FMC-CE expansion board	139
89	A photo of the Multi-layer Object	140
90	Close-up of the galvo-scanning mirror	141
91	A photo of the PZT mounted reference mirror	142
92	PI E-463 high-voltage PZT amplifier	143
93	PZT control signal block diagram	143
94	Noise floor results comparing LRT and CGLRT	145
95	Comparison of standard LRT vs. CGLRT for a single layer target	146
96	Difference in Zernike modes reported by standard LRT vs. CGLRT for a single layer target	147
97	Screen shot of reflections from the MLO	149
98	Screen shot with differentiated reflections from the MLO	150
99	Interference fringes from surface P5	151
100	Interference fringes from surface P4.	151
101	Interference fringes from surface P3.	151
102	Interference fringes from surface P2.	152
103	Interference fringes from surface P1.	152
104	Zernike coefficients for wavefront returned from each surface of the MLO	153

List of Tables

1	The first 14 Zernike polynomials and their common names	35
2	CLB resources available in a variety of Xilinx devices, including the Virtex-6 XC6VLX240T	93
3	MATLAB vs. FPGA image centroid comparison	117
4	MATLAB vs. FPGA binary image centroid comparison	120
5	Zernike coefficients for the simulated input wavefront	129
6	Numerical error for calculated Zernike coefficients using 24-bit and 32-bit word lengths.	130
7	Resource utilization for 24 bit ROM values compiled for a Xilinx Spartan-6 XC6SLX45.	132
8	Resource utilization for 32 bit ROM values compiled for a Xilinx Spartan-6 XC6SLX45.	132
9	Resource utilisation for 32 bit ROM values compiled for a Xilinx Virtex-6 XC6VLX240T-1FFG1156.	133
10	Resource utilisation for 2D-FFT phase recovery circuit: 8 bit inputs	133
11	Resource utilisation for 2D-FFT phase recovery circuit: 16 bit inputs	133
12	CGLRT FPGA architecture resource utilisation for various scan sizes	154
13	Timing of the CGLRT instrument	155
14	Expected timing of the CGLRT instrument utilising a high-speed camera	155
15	Timing comparison of the CGLRT instrument to a theoretical CGSHS	158

Acronyms

AO adaptive optics. 12, 16, 22, 23, 42, 43, 48, 52, 81

AOM acousto-optic modulator. 51, 52

ASIC application specific integrated circuit. 79

BRAM block RAM. 80

CCD charge-coupled device. 50, 51, 57

CG coherence-gated. 85, 95, 97, 100, 103, 127

CGLRT coherence-gated laser ray tracing. 9, 10, 18, 37, 46, 48, 57, 59, 72, 75, 76, 80–82, 84, 90, 100, 113, 118, 120–123, 126–128, 131, 135, 136, 138, 139

CGSHS coherence-gated Shack-Hartmann wavefront sensor. 17, 43, 44

CGWFS coherence-gated wavefront sensing. 38, 42, 43, 48

CLB configurable logic block. 79, 80

DAC digital to analogue converter. 122

DM deformable mirror. 12, 13, 17

DPRAM dual-port RAM. 80, 93

DSP digital signal processing. 79–82

FIFO first-in/first-out. 80

FMC FPGA mezzanine card. 81, 87

FPGA field-programmable gate array. 10, 15, 18, 59, 76, 79–82, 89, 90, 92, 97, 100, 113, 114, 118, 119, 122, 125–127, 131, 135–137

FSM finite state machine. 89, 91

- GPU** graphics processing units. 81
- HDL** hardware description language. 79, 82
- IOB** input/output buffer. 79, 80
- LCI** low-coherence interferometry. 37, 38, 46, 72
- LRT** laser ray tracing. 6, 7, 9, 12–15, 17, 23, 25, 28, 49–59, 62, 63, 118, 126–128, 131
- LTS** linear translation stage. 131
- LUT** look-up table. 80
- MLO** multi-layer object. 123, 127, 131, 135
- MMCM** multi-mode clock manager. 89
- MSC** master state-controller. 84, 85, 87, 91, 94
- OCT** optical-coherence tomography. 12, 14, 16, 17, 38, 43, 81
- OM** object mirror. 131
- OPD** optical path difference. 29, 41, 57
- OPL** optical path length. 131, 137
- PLD** programmable logic device. 79
- PSD** position sensitive detector. 7, 50–52, 55–58
- PSF** point spread function. 54
- PSI** phase-shifting interferometry. 38, 41, 127
- PZT** piezo-electric transducer. 73, 137
- RAM** random-access memory. 80
- ROM** read-only memory. 80
- SHS** Shack-Hartmann wavefront sensor. 12–15, 17, 23–25, 28, 29, 37, 42, 43, 49, 51–53

SLD super luminescent diode. 57, 58, 118

SLO scanning laser ophthalmoscopy. 12

SPRAM single-port RAM. 80

WFS wavefront sensor. 12, 15, 23, 41, 118

1 Introduction

Optical wavefront sensing is a critical component in adaptive optics (AO) technology, with applications in diverse imaging fields such as astronomy and ophthalmology. While the applications are varied, the core method remains the same: measure and correct the optical aberrations of a wavefront. In astronomy this involves correcting wavefront distortions caused by the atmosphere of the Earth. With retinal imaging techniques, such as optical-coherence tomography (OCT) [20], scanning laser ophthalmoscopy (SLO) [62] and confocal microscopy [30] it involves correcting distortions due to imperfections in the optics of the eye. Both cases involve first measuring the distorted wavefront. This is done utilising a device known as a wavefront sensor (WFS). Once the distortions are known they can be corrected, typically by utilising either a deformable mirror (DM) [16, 17], spatial light modulator [24], or similar device. By minimising these distortions, the quality of the final image can be greatly improved [23]. In general, the combination of a wavefront sensor and correction device is known as AO [70, 3, 15].

1.1 Motivation

This research presents, for the first time, a depth-resolved laser ray-tracing wavefront sensor on an embedded systems platform. The motivation to perform this research stems from Adrian Podoleanu and Simon Tuohy’s 2013 patent: ”Method for depth resolved wavefront sensing, depth resolved wavefront sensors and method and apparatus for optical imaging” US 8451452 B2 [39]. In it, they begin with several types of well-known wavefront sensors, including the Shack-Hartmann wavefront sensor (SHS) and laser ray tracing (LRT) wavefront sensor; and describe improving them with the addition of depth-resolution. In essence, the standard wavefront sensor form-factors are modified to reduce the depth of field. The axial location (in depth) of the wavefront is able to be selected, either by time domain or frequency domain methods.

A natural starting point for realising a depth-resolved wavefront sensor, as described in [39], is to focus on the SHS. The SHS is the most widely used wavefront sensor with applications in many fields, as discussed in Chapter 2. As a result, the initial depth-resolved wavefront sensor implementations focused on the SHS. Figure 1 shows the current state of depth-resolved wavefront sensing research. Standard LRT and SHS belong to the general category of wavefront sensors. Patent [39] describes the depth resolved versions of both. The initial implementation of a depth-resolved SHS was published in 2010 in a paper entitled

“Depth-resolved wavefront aberrations using a coherence-gated Shack-Hartmann wavefront sensor” [59]. In this paper, depth-resolved wavefront sensing was demonstrated for a *single-layer* object (in this case a DM). This paper focused more on the system’s ability to remove stray reflections and increase the SNR than demonstrate measuring wavefronts returned from a multi-layer target. It was left to the 2012 paper “Demonstration of real-time depth-resolved Shack–Hartmann measurement” [61] to show the ability of this technique to measure multiple wavefronts, selected in depth, from a multi-layer target (in this case stacked microscope slides). Both of these papers will be discussed in detail in Chapter 3. Combined, these papers demonstrate the ability of the SHS modality to perform depth-resolved wavefront sensing.

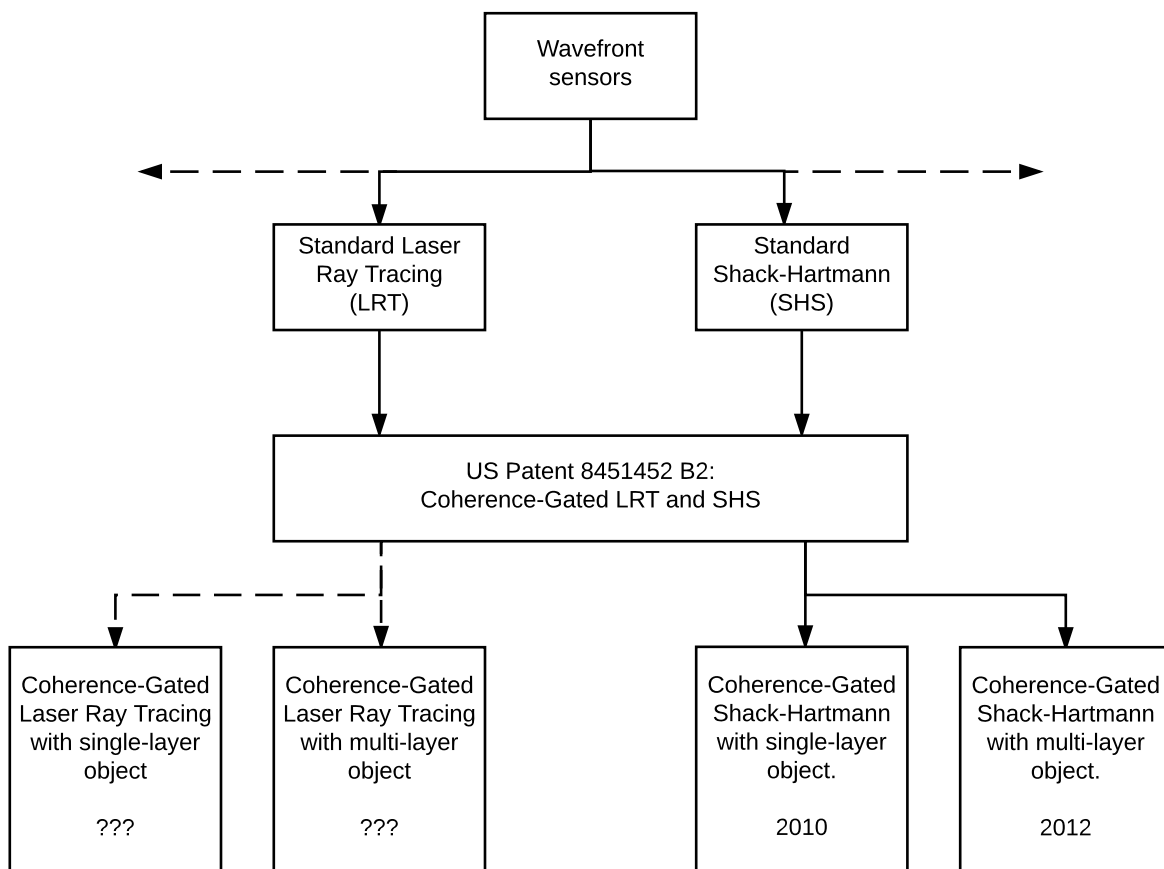


Figure 1: The general topic of wavefront sensors (shown at the top) can be broken down into many specific technologies, two of which are the SHS and LRT (shown in row 2). US Patent 8451452 B2 describes expanding the capabilities of these devices to include depth-resolved wavefront sensing. The ability of SHS to perform such measurements was demonstrated in 2010 and 2012 (shown on the right), but has yet to be shown for LRT (shown on the left).

The current situation presents the opportunity to explore realising depth-resolved wavefront sensing in other modalities. LRT is a well-established [31, 32, 35] wavefront sensing technique that closely resembles the process of geometric ray-tracing typically used to analyse optical systems. LRT sequentially scans a target object with a thin pencil laser beam producing a spot pattern similar to that of a SHS, the centroids of which may be used to reconstruct the wavefront. While the specifics of LRT will be discussed in Chapter 4, it is important here to note the motivation for adding depth-resolution to LRT:

- LRT devices exist as commercialised instruments already in use in ophthalmic applications [57]. Current realisations of this technology do not exhibit depth-resolved capabilities making them impractical for 3D imaging modalities such as OCT (discussed below).
- Many biological tissues such as the human retina, skin, and teeth are composed of distinct cellular layers which may be resolved by OCT type imaging systems. Wavefront sensors that work in tandem with such systems need to operate in a depth-resolved manner.
- Unlike the SHS, LRT is similar in form-factor to common OCT setups and will be simpler to integrate, as a wavefront sensor, into an existing OCT system.
- The serialised nature of data acquisition in LRT makes it well suited for implementation on an embedded systems platform.
- Including depth-resolution comes with the added benefit of removing stray reflections from the sensor. Only light selected from the limited depth of field is imaged, leading to other extraneous sources of light being removed.

In this new implementation, depth-resolution is added to a standard LRT by combining it with a low-coherence interferometer. By employing coherence-gating, the depth of field of the LRT is reduced to within the coherence-length of the optical source. This allows the system to reject light from outside the coherence-gate and perform depth-resolved wavefront sensing. It has the added benefit of rejecting stray reflections from the optical system which can often confuse or render unusable standard wavefront sensors.

This research attempts to fill both of the LRT boxes in Figure 1 by demonstrating depth-resolved LRT on both a single-layer object (a mirror) and a multi-layer object (stacked microscope slides). This will show that depth-resolved LRT agrees with standard LRT by

performing both types of measurements on the same single-layer object. In addition, it will show the ability for this method to resolve multiple wavefronts, returned in depth, from a multi-layer object. Lastly, this research takes an instrumentation-based approach to the LRT hardware development by employing an embedded systems field-programmable gate array (FPGA) platform for all system command and control, data acquisition and image processing tasks. In this manner, the use of a stand-alone PC for such tasks is rendered unnecessary, leading to a completely self-contained system which future research could refine with the goal of integration into an existing imaging system. The FPGA based system is also much faster than a standard CPU based computer allowing the system to run in near real-time. It should be noted that the use of the term *real-time* is somewhat ambiguous. In this research the term will be used to refer a computing system that is, in practice, limited not by its ability to quickly process data but by the external hardware resources, such as camera frame-rates, optical scanner slew-rates, and the like. It also excludes systems that gather data and then proceed with post-processing on a standard PC. This was the case with [59] and [61] where images were gathered and sent to a PC for processing after all data acquisition was complete. This will be in contrast to the LRT system demonstrated here, where all data processing is done in parallel with the data acquisition. For all practical purposes the result of the wavefront sensing is available as soon as the entirety of the images have been collected. With these motivations in place the discussion may now turn to a general outline of the research.

1.2 Outline

This research project is cross-disciplinary by nature, as the format of this thesis represents. The two main components focus on the optical portion of the depth-resolved LRT instrument and the FPGA based embedded system on which it rests. As a result, the relevant literature is reviewed on a chapter by chapter basis, as it applies to material being presented. For example, the literature relevant to LRT is presented in Chapter 4 where the initial LRT implementation is presented; while the literature relevant to the FPGA architecture is presented in Chapter 5 which details the design of the embedded systems platform.

The astronomical origins of WFS are explored in Chapter 2 but wavefront sensing also has a diverse history in ophthalmic applications [19]. While many methods have been studied including the Tscherning aberroscope [58] and Howland’s cross-cylinder aberroscope [18] both astronomical and ophthalmic applications have settled upon the SHS [52] as the method of choice. Details of the SHS will be discussed in Chapter 2. While the SHS is ubiquitous

1. INTRODUCTION

in its usage it does have limitations. As noted above one limitation of the SHS and other current methods is that they exhibit a large depth of field inhibiting their ability to resolve wavefronts returned from objects composed of multiple layers. Multi-layer objects are common in biological imaging where specimens are often composed of many distinct layers of cells, for example, the human retina and human skin. This is important in imaging modalities such as OCT that are commonly used to produce 3D volumetric images of such targets. For example, Figure 2, taken from [8] shows the ten layers of the human retina from a high resolution OCT image on the top and a standard histologic micrograph on the bottom (from [10]).

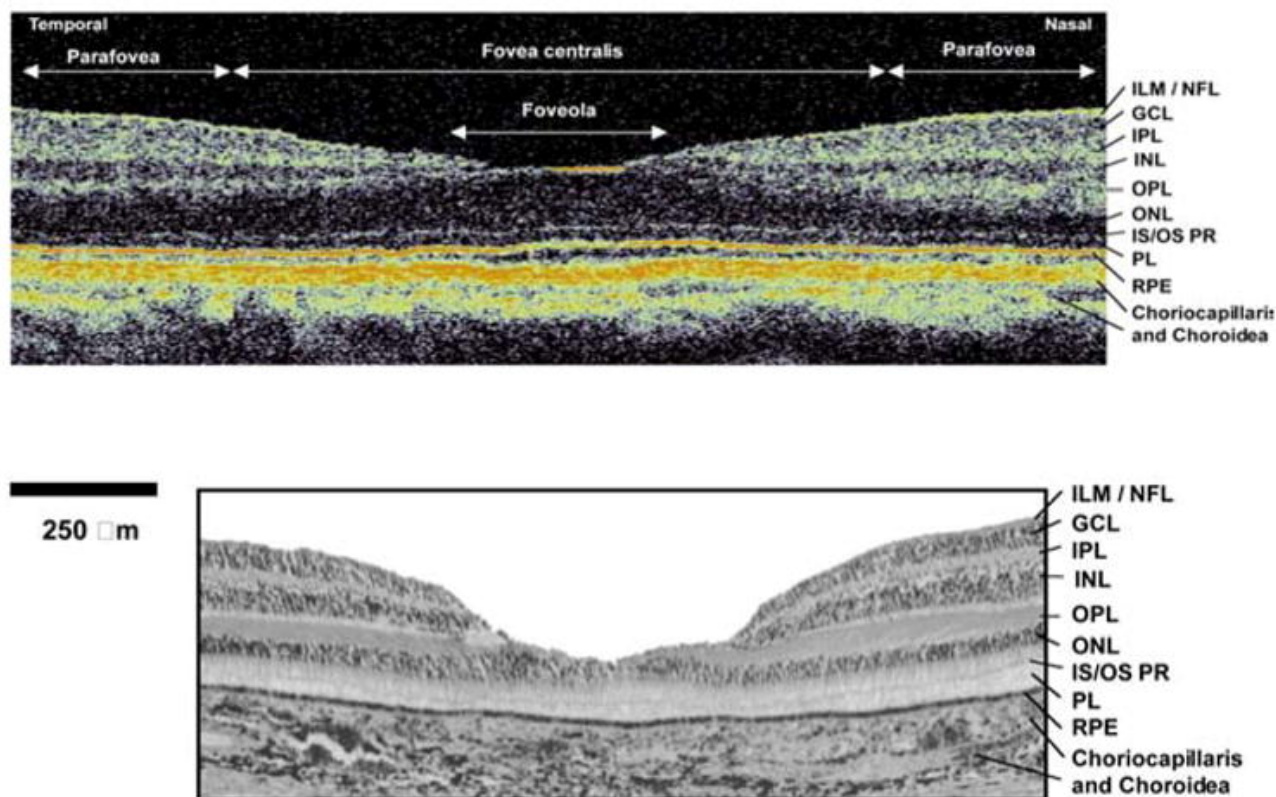


Figure 2: Cellular layers in the human retina are clearly visible, demonstrating the ability of modalities such as OCT to perform imaging on multi-layer targets. Top: High resolution OCT image taken from [8]. Bottom: A standard histologic micrograph taken from [10]

OCT is a non-invasive interferometric technique that can be employed *in vivo* to image bulk biological tissues. OCT was first developed in 1991 by Huang, et al. [20] and has rapidly developed over the past several decades into a high-speed imaging modality. Recently, OCT has benefited from the addition of the wavefront correcting abilities of AO. AO is well known in the astronomical community for its ability to compensate for phase distortions due

to atmospheric turbulence, pushing the functionality of ground based telescopes ever closer to their space based counterparts [11]. Based mainly on the classic combination of a DM and a SHS, AO-OCT systems have succeeded in increasing the quality of OCT images. In spite of this success, one short coming of the current breed of AO systems is their reliance on sensors that are unable to resolve wavefront aberrations as a function of depth. The ability to resolve depth based aberrations would be extremely useful in the field of microscopy and especially biological tissue imaging where aberrations often originate from scattering layers beyond the focal plane. A novel method utilized for resolving wavefront aberrations as a function of depth is the technique known as coherence-gated wavefront sensing. Utilizing the methods of low-coherence phase-shifting interferometry coherence-gated wavefront sensing is able to reject light acquired from outside the focal plane, focusing on a small area of interest in the sample. Currently, this methodology has been shown to function by Tuohy & Podoleanu in the form of a Coherence-Gated Shack Hartmann WFS (CGSHS) [59] and in a virtual form by Feierabend, Rüeckel and Denk [9, 49, 48, 47].

Refractive laser surgery revolutionized the explorations of wavefront aberrations in the human eye [19]. In contrast to psycho-physical subjective methods this catalysed a progression of objective measurement methodologies that characterized ocular wavefront aberrations directly [29]. This technology has branched out to general biological imaging, such as microscopy and tomography [3], and has had prolific uses in the dynamic field of OCT. Specialized wavefront sensing is currently critical in modalities such as OCT for high contrast imaging from multi-layer scattering samples using adaptive optics [70, 15].

As imaging modalities such as OCT approach real-time operation [63], wavefront sensing techniques are required to run at high-speed in order to be applicable. For this reason, an embedded systems platform was chosen for the control, image processing, and wavefront reconstruction of this instrument. A Virtex-6 FPGA architecture was developed which interfaces directly to a CCD camera for image acquisition and post-processing. The wavefront is reconstructed in parallel on the FPGA and presented in the form of Zernike modes. This all takes place at high-speed, limited only by the frame-rate of the CCD camera.

First, Chapter 2 will give a brief introduction to wavefront sensing including the common optical aberrations and the utilisation of Zernike modes to describe them. In Chapter 3 the topic of wavefront sensing is expanded to include coherence-gating which allows for depth-resolved wavefront sensing . This includes a discussion of implementations of a coherence-gated *virtual* SHS and a physical coherence-gated Shack-Hartmann wavefront sensor (CGSHS). Next, Chapter 4 gives an introduction to the basics of LRT including the single-pass and double-pass implementations. Descriptions and results of both non-

coherence-gated and coherence-gated LRT are given for non-real-time (Labview based) implementations. Next, the focus turns to the FPGA architecture used in the final coherence-gated laser ray tracing (CGLRT) instrument in Chapter 5. This includes descriptions of the general command and control functions, camera interface, and wavefront reconstruction modules. Lastly in Chapter 6, the embedded systems based CGLRT instrument is discussed and results are shown demonstrating the ability of the system to perform high-speed depth resolved wavefront sensing.

2 Wavefront Sensing

As discussed in Chapter 1, this cross-disciplinary project comprises the development of both optical and electronic systems. This chapter will focus on introducing the foundational topics needed to facilitate a more in-depth discussion of wavefront sensing in Chapters 3, 4 and 6. First, the concept of an optical wavefront is discussed. This leads naturally into the topic of adaptive optics of which the wavefront sensor is a critical component. The topic will then move to the Shack-Hartmann wavefront sensor, which is both ubiquitous, and useful for elaborating on the basic operating principles of wavefront sensors in general. After measuring, the topic of reconstructing the wavefront is covered, including both the zonal and modal methods. From there, Zernike modes are discussed as they are the most common method of mathematically representing wavefronts. Lastly, some common optical aberrations and their physical causes are discussed. This helps tie in the abstract notion of Zernike modes and the aberrations they describe to physical phenomenon. With these concepts in hand, the discussion may progress to the specific wavefront sensing techniques used in the research presented in this dissertation.

2.1 An Introduction to Wavefront Sensing

The twinkling stars that fill the night sky are a beautiful sight. However, to astronomers, the twinkle of a star embodies one of the great difficulties of ground based astronomy. The atmosphere of the Earth distorts the light from the distant stars decreasing the resolution of images obtained on the ground. It is this distortion, due to continuous variations in the refractive index of the atmosphere, that causes the familiar twinkle in starlight.

When discussing the distortion of light the concept of the *wavefront* is often employed. A wavefront is a surface of constant phase of a wave as it propagates through space. Typical wavefronts are either considered planar or spherical. Collimated light is a common example of a planar wavefront. Figure 3 shows several planar wavefronts superimposed on a light wave.

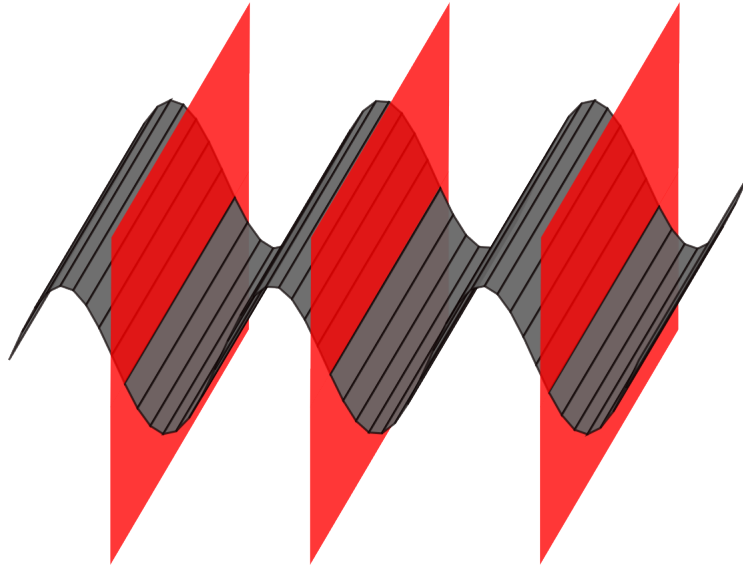


Figure 3: Plane waves or planar wavefronts. Planes of constant phase (wavefronts) are shown in red.

A spherical wavefront, on the other hand, is typically formed by a source such as stars and light-bulbs which emit waves in all directions. Figure 4 shows an illustration of a light-bulb emitting spherical wavefronts. The wavefront is only partially shown and in reality it is emitted as a spherical surface in all directions. Parts of planar wavefronts may be converted to spherical wavefronts and vice-versa via lenses. When a collimated beam of light (planar wave) is incident upon a lens the wavefront is converted from planar to spherical as the beam it focuses. The opposite is true for converging or diverging beams that are collimated by a lens, converting the spherical wavefront to planar.

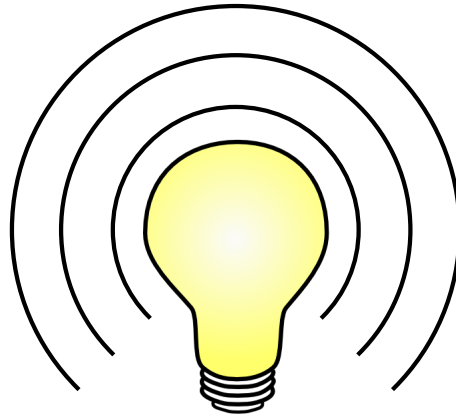


Figure 4: A spherical wavefront. Sources such as light-bulbs and stars, that emit waves in all directions, generate spherical wavefronts.

In practice, a localised small area on a spherical wavefront of very large radius can be approximated as planar. This is especially true for stars, in which the spherical wavefront has expanded to a great degree as it travels through space. Figure 5 shows a (now planar) illustration of a wavefront from a star (shown in red) travelling through the vacuum of space undistorted. As the wavefront travels through the turbulent atmosphere of the Earth the constantly varying refractive index distorts the wavefront from its original shape. This distortion gives the impression that the star is twinkling but also has the drawback of decreasing the resolution of images taken by ground based telescopic instruments.

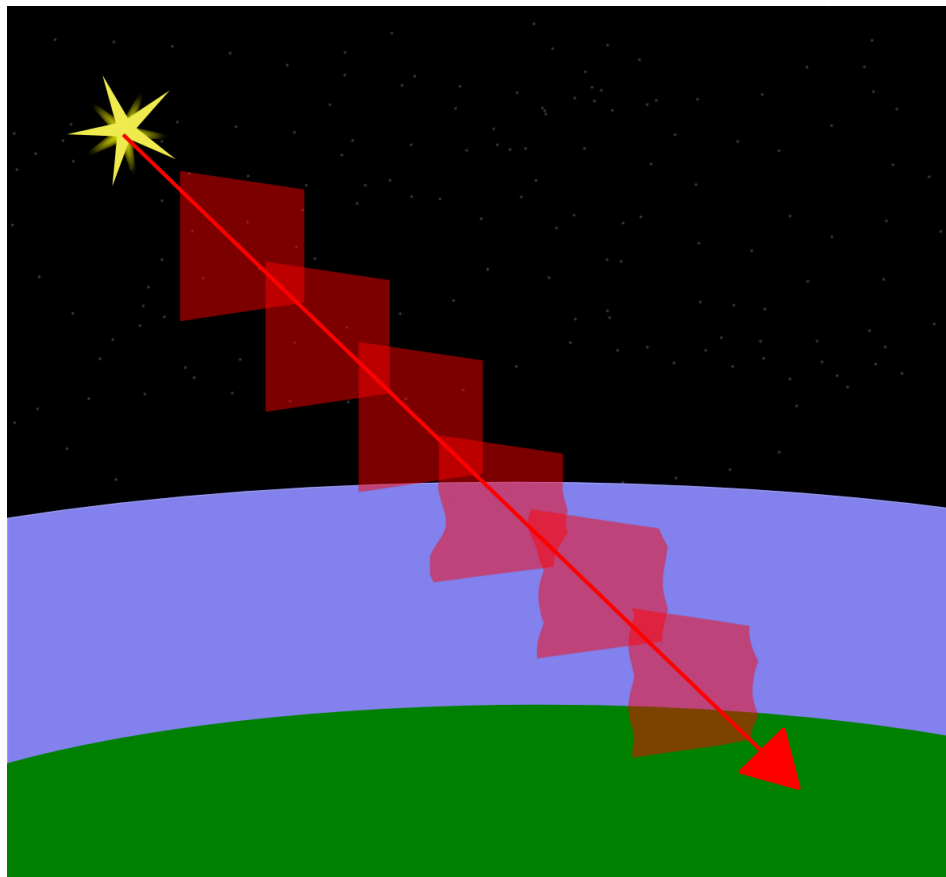


Figure 5: The wavefront from a star, shown in red, is distorted by the turbulence in the Earth's atmosphere.

Dealing with such distortion is critical, and astronomers are seemingly left with two options: place the telescope above the atmosphere (for example, the Hubble Space Telescope) or develop a method for correcting the distortions on the ground. The second option, first proposed by Babcock in 1953 [2], proposed manipulating an optical wavefront to reduce aberrations. Over time, this has developed into the field known as AO which attempts to measure and correct wavefront distortions in light. A simplified illustration of an ocular adaptive optics system is shown in Figure 6. In essence, a sensor is used to continuously measure the deviation of a wavefront and the deviation is corrected, in a closed-loop fashion, by another device, typically a deformable mirror. Over the years the art of AO has rapidly progressed such that current ground based, AO equipped telescopic systems are rivalling the performance of their space based counterparts [43]. As the technology has matured, it has also found its way beyond astronomy into other imaging modalities such as microscopy [42] and *in vivo* biological imaging [37].

A critical component in the art of AO is the ability to sense distortions of the incoming wavefront so that corrections may be applied. Many methods now exist for wavefront sensing, including SHS [52], LRT [35, 31], Modal Sensing such as Curvature [44] and Phase Diversity [12], Scanning Knife Edge tests [54], and Pyramid sensors [41]. This dissertation focuses on LRT, but it is prudent to first review the origins and function of SHS as it is by far the most common and well known WFS.

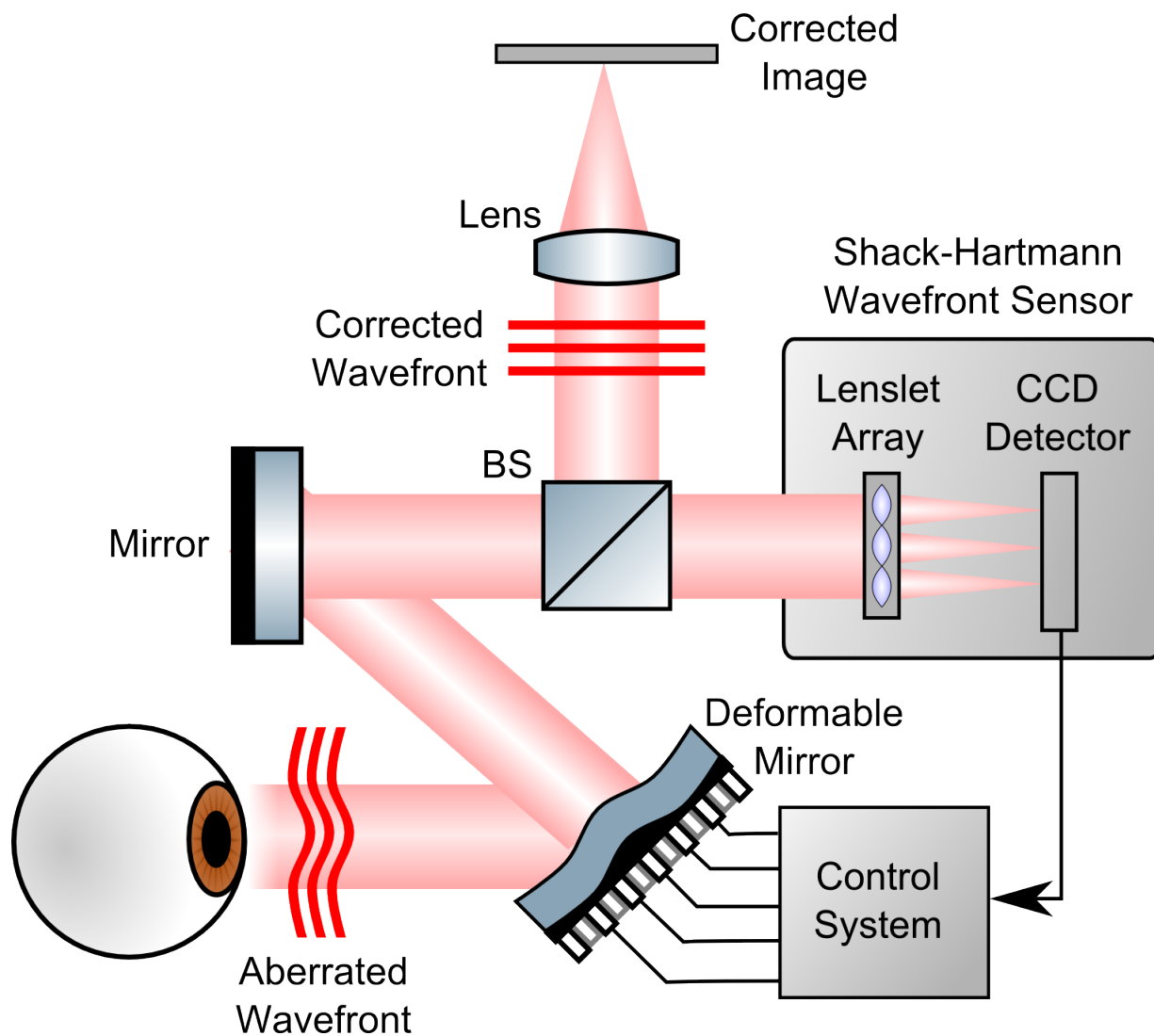


Figure 6: Simplified illustration of an adaptive optics system. In this example the system is measuring the aberrations of a human eye. In it, aberrated light is returned from the optics of the eye. The light is incident upon a deformable mirror (DM) which corrects for the distortions as commanded by the closed-loop control system. A mirror and beam splitter (BS) send the light to a wavefront sensor (in this example a Shack-Hartmann). The wavefront sensor measures the distortion in the wave and sends this information to the control system. The corrected wavefront is now available to the system.

2.2 The Shack-Hartmann Wavefront Sensor

Wavefront sensing had its origin in the fields of astronomy and the military. Perhaps the best known wavefront sensor, the SHS, had its genesis in the needs of the “Great Refractor” telescope in Potsdam, Germany. In the early 1900’s Johannes Hartmann (1865-1936) was a professor of astrophysics at Potsdam where he worked with the large telescope. Unfortunately, the telescope was incapable of performing the required imaging due to aberrations in its optics. In order to identify the faulty optic Hartmann devised a method of evaluating wavefront quality that would become known as the *Hartmann screen test* [14]

The test involved placing a plate with many evenly spaced holes in it over the aperture of the telescope. This plate created many small apertures that could be imaged via photographic plates at either focal plane of the given optic. By measuring where the ray bundles cross the optical axis, the quality of the optic can be inferred. High quality optics will have corresponding rays crossing at the same points on the optical axis, whereas in poor quality optics the ray location will vary. Using this technique Hartmann was able to identify the Great Refractor’s primary lens as the culprit [51]. The Hartmann screen test continues to be a useful diagnostic tool in the testing of optical elements, but the need to better characterize optical wavefronts soon pushed the Hartmann screen test into new territory.

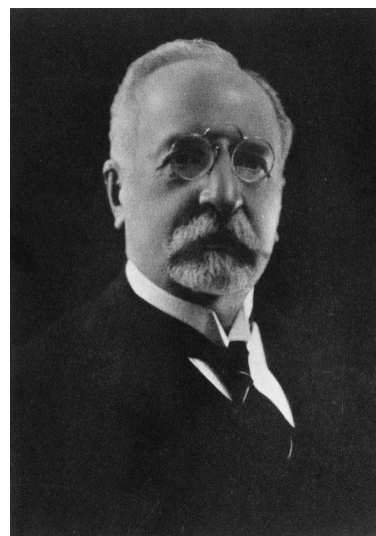


Figure 7: Johannes Hartmann

The catalyst for this next step came from the needs of the US Air Force during the Cold War era of the 1960’s. An attempt by the US Air Force to develop a method of imaging satellites from the ground led to the need to compensate for atmospheric turbulence in the images. The Air Force approached the Optical Sciences Center at the University of Arizona where one Aden Meinel was put on the project. Meinel suggested using the Hartmann screen test to characterize the atmospheric turbulence and his co-worker Roland Shack was given the task of testing the method’s feasibility. He soon found that the available light from the satellite images (most of the light was needed to image the dim satellites) was far too small to successfully perform the Hartmann screen test. Shack cleverly overcame this issue by placing small lenses in the holes of the Hartmann screen, focusing the light into spots, thereby increasing the intensity.

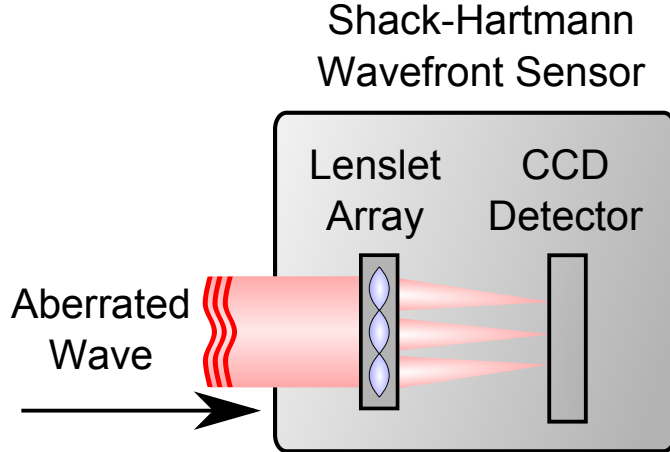


Figure 8: A basic Shack-Hartmann wavefront sensor.

The SHS continues to be a critical component in the field of wavefront sensing today. Not only has this device been utilized in the field of astronomy, but it has also been adapted for sensing applications in microscopy [13], ophthalmology [7] and *in vivo* imaging of biological tissue [23]. It is worth exploring the function of the SHS in detail as it provides a baseline for understanding wavefront sensing in general and LRT in particular.

The basic components of a SHS consist of a fixed lenslet array and an imaging camera, typically a CCD, as shown in Figure 8. The fixed lenslet array is an evenly spaced grid of micro-lenses each of which subsamples the incident wavefront and focuses it onto the detector surface. In the case of a non-aberrated wavefront the centroid of each focused spot will be located at the centre of the subaperture as shown in Figure 9. A spot pattern is created that forms a regular grid on the detector surface. The wavefront is then reconstructed from the image of the spots via a least-squares fit or similar method. Figure 10 illustrates an aberrated wavefront in which the spot locations will be offset from the lenslet centres by an amount proportional to the local wavefront gradient across the subaperture. The wavefront gradient over the i th subaperture $\partial W_i(x, y)$ is related to the local Δx_i and Δy_i displacements by

$$\begin{aligned} \frac{\partial W_i(x, y)}{\partial x} &= \frac{\Delta x_i}{f} \\ \frac{\partial W_i(x, y)}{\partial y} &= \frac{\Delta y_i}{f} \end{aligned} \tag{2.1}$$

where f is the focal length of the lenslet. In a similar manner to the non-aberrated case, the aberrated wavefront is reconstructed from the image of spot locations via a least-squares fit.

Wavefront reconstruction is discussed in detail in Section 2.3.

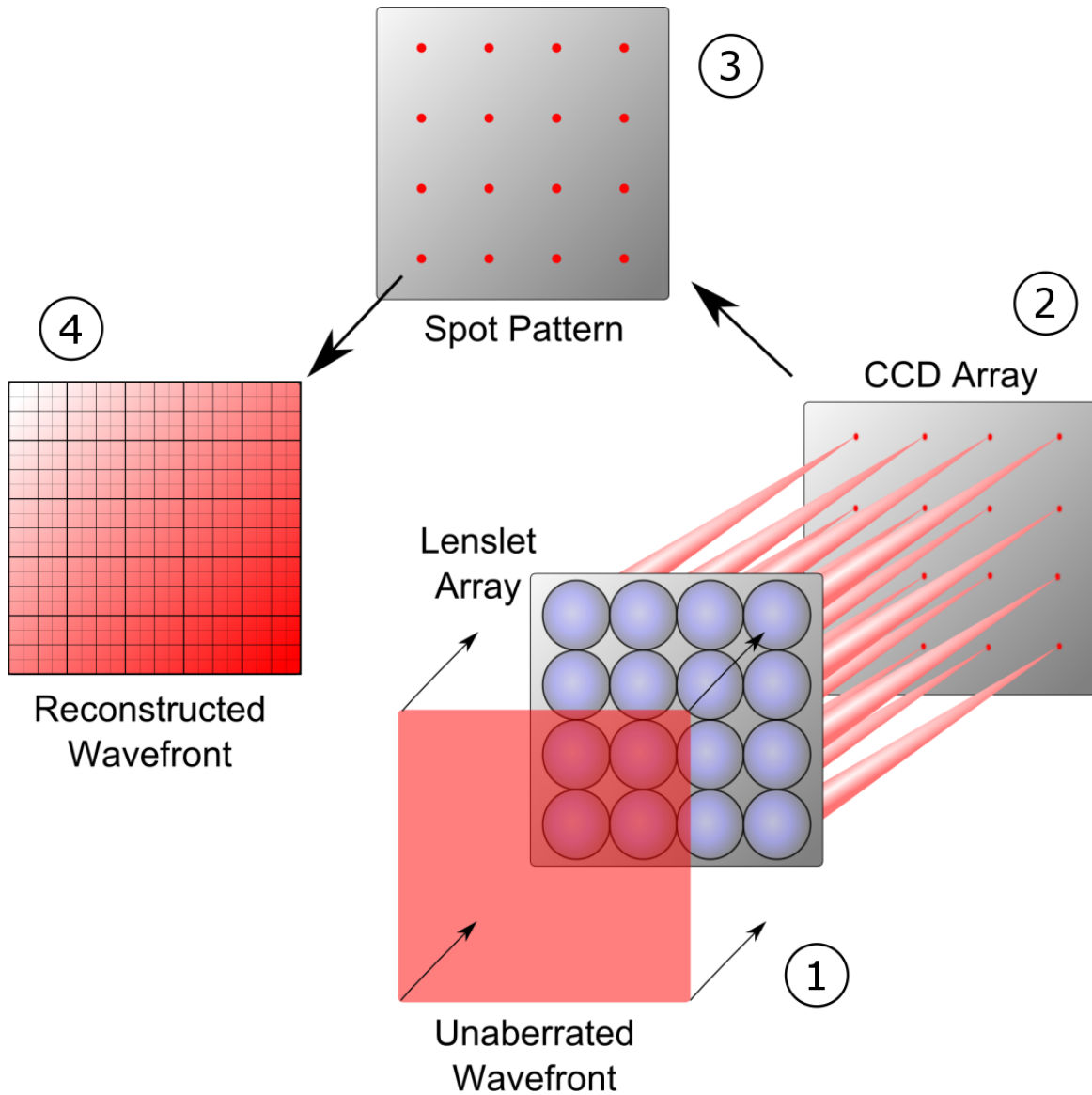


Figure 9: Functional illustration of a Shack-Hartmann wavefront sensor for an non-aberrated wavefront. 1. The non-aberrated wavefront is incident upon the lenslet array. 2. Each subaperture of the lenslet array focuses a portion of the wavefront onto the surface of a CCD (or similar) detector. 3. Due to the absence of aberrations the spot formed by each lenslet coincides with the centre location of the lenslet. 4. This forms a spot pattern from which the measured wavefront may be reconstructed.

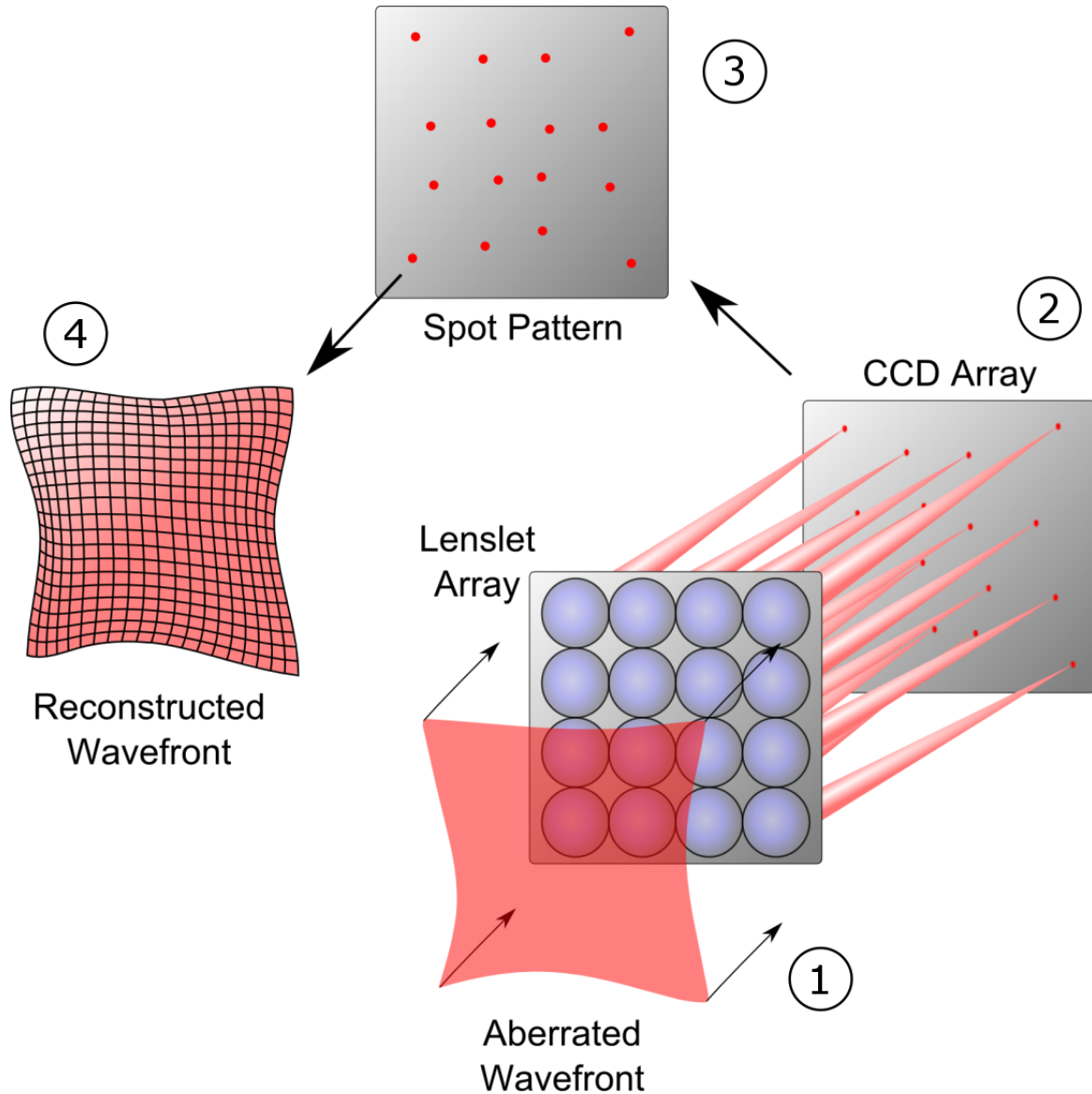
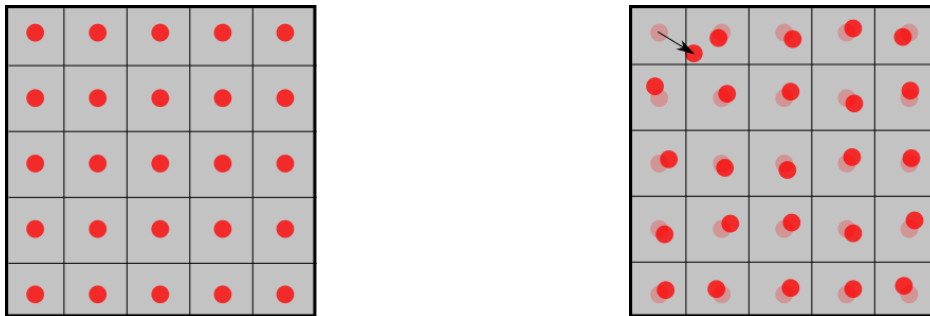


Figure 10: Functional illustration of a Shack-Hartmann wavefront sensor for an aberrated wavefront. 1. The aberrated wavefront is incident upon the lenslet array. 2. Each subaperture of the lenslet array focuses a portion of the wavefront onto the surface of a CCD (or similar) detector. 3. The spot locations are offset from the centre of each lenslet by an amount proportional to the slope of the wavefront at that location. 4. This forms a spot pattern from which the measured wavefront may be reconstructed.

SHSs have become ubiquitous for wavefront sensing with many versions available commercially [56]. They are often bundled with software for wavefront phase reconstruction or even as a complete adaptive optics kit. For example, the high resolution (1.3 Megapixels) SHS from [56] has a maximum frame rate of 15 hertz while the high-frame rate (450 hertz) version [55] offers a small frame (180×180 pixels) and only for onboard camera processing

of up to 5 Zernike modes. Though SHSs have general usage in the wavefront sensing field there are several drawbacks and/or limitations that justify exploring other modalities.

First, SHSs obtain all of the imaged spots in parallel. While this is desirable in terms of speed, it requires that the image be segmented (to produce separate images of all of the individual spots) before the spot centroids may be calculated. This is not only time-consuming but can lead to errors if, for instance, a large aberration causes a spot to be displaced from one lenslet area to the adjacent one as shown in Figure 11. This is especially troublesome for automated wavefront sensing systems which are unable to determine which spot belongs to the correct grid location. Errors such as this do not occur with LRT since the spots are imaged one at a time, serially. As such, there can be no ambiguity as to which spot belongs to a given scan location no matter how large the aberration.



(a) Each spot is within the lenslet grid. (b) Large aberrations can cause gridding errors.

Figure 11: In (a) a non-aberrated (or slightly aberrated) wavefront produces a spot pattern where each spot is located at the centre of its respective lenslet grid. In (b) large aberrations are present and a spot can move outside the grid associated with its lenslet area. Here, the top left spot (indicated by the arrow) has moved into the adjacent lenslet space, causing an error.

Another drawback of SHS is that the available spot patterns are fixed by the lenslet array such that neither the pattern nor the number of measurement locations may be changed. Modalities such as LRT are more flexible in this respect, with the number and pattern of measurement locations being user selectable.

2.3 The Zonal and Modal Methods

Once a given wavefront has been measured, the task turns to reconstructing it in a form that can be represented mathematically. As noted in Section 2.2 a SHS outputs an intensity image of spots that correspond to the local wavefront gradient over the lenslet subaperture.

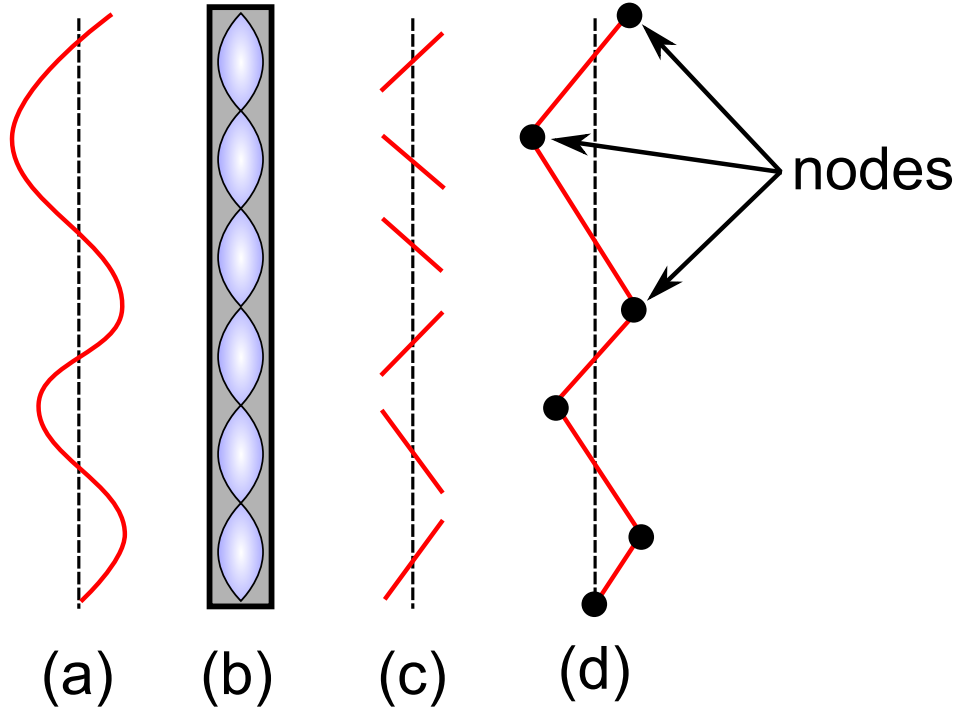


Figure 12: The zonal method of wavefront reconstruction. The original wavefront (a) is sampled by lenslet array (b). For each subaperture a local tilt is measured (c). The OPD for each subaperture is calculated via a least-squares, or similar method, resulting in the reconstructed wavefront (d).

Two common methods for converting the measured gradients into the original wavefront are the zonal method and the modal method.

The zonal method involves dividing the wavefront into N small subsections and a local tilt or optical path difference (OPD) is determined for each subsection. As $N \rightarrow \infty$ the wavefront may be reconstructed fully, though in practice N is limited to some finite value that depends on the physical setup of the sensor. Figure 12 illustrated the concept of the zonal method. The original wavefront (a) is sampled by a lenslet array (b). In practice the intensity image is then captured by a CCD camera or similar device and the tilt for each subaperture is obtained using the methods discussed in Section 2.2. After each local tilt is obtained (c) the OPD to each node is calculated via a least-squares or similar method, and the reconstructed wavefront is obtained (d). In this way, the final wavefront reconstruction is represented by an array of discrete OPD values, one for each node.

Another option for reconstructing a measured wavefront is the modal method. In contrast to the zonal approach, this method represents the measured wavefront as a sum of 2D-surfaces. This analytic approach has several advantages over the zonal approach in that it

- provides infinite resolution of the reconstructed wavefront due to it being represented by continuous functions, opposed to discretely, as in the zonal approach, and
- may represent the wavefront as a sum of intuitive components such as tilt, astigmatism and defocus.

Though there are many possible basis sets that one may choose to represent a given wavefront (e.g. Fourier, Seidel) the one generally used in optics are known as Zernike modes and will be discussed at length in Section 2.4.

2.4 Zernike modes as wavefront descriptors

If the modal method of wavefront reconstruction is utilized, it still remains to choose an appropriate polynomial basis set. There are several characteristics that make a given basis desirable:

1. The set should form an orthonormal basis.
2. The set should capture aberrations common to optical systems.
3. The set should be defined over a circular pupil.

First, it is desirable to have the basis set be orthonormal so that the addition or removal of terms do not effect the remaining terms. One would not wish to have the magnitude of the 3rd coefficient change because the expansion was made to include 20 terms instead of 10. Secondly, there are aberrations that are common to most optical systems such as defocus, astigmatism, coma, etc. It is highly desirable that each of these aberrations be represented uniquely in the basis set. These types of aberrations are discussed in more detail in Section 2.5. Lastly, most optical systems utilize circular beam profiles and it is desirable that the basis set captures this inherently.

Zernike polynomials are often chosen as a basis set for representing optical wavefronts as they meet all of the above criteria. Each Zernike term, often referred to as a mode, is orthogonal to the other modes, and defined over a normalized circular pupil. Generally, Zernike modes are defined over the polar coordinates (r, θ) and the i th Zernike mode is defined as

$$Z_{mn}(\rho, \theta) = \begin{cases} R_{mn}(\rho) & m = 0 \\ R_{mn}(\rho) \cos(m\theta) & m > 0 \\ R_{mn}(\rho) \sin(m\theta) & m < 0 \end{cases} \quad (2.2)$$

where m and n for the i th Zernike mode are defined in [36] as shown in Table 1). $R_{mn}(\rho)$ is referred to as the radial polynomial and is defined as

$$R_{mn}(\rho) = \sum_{k=0}^{(n-m)/2} \frac{(-1)^k (n-k)!}{k! \binom{n+m}{2-k}! \binom{n-m}{2-k}!} \rho^{n-2k} \quad (2.3)$$

The $\sin(m\theta)$ and $\cos(m\theta)$ terms are referred to as the azimuthal components where m is the azimuthal frequency. Table 1 lists the first 15 Zernike modes including the normalized polynomial and the common name. The first ten Zernike modes (normalised over a pupil radius of 1) are shown in Figure 13. Each mode, when normalised has a maximum of +1 (red) and a minimum of -1 (blue).

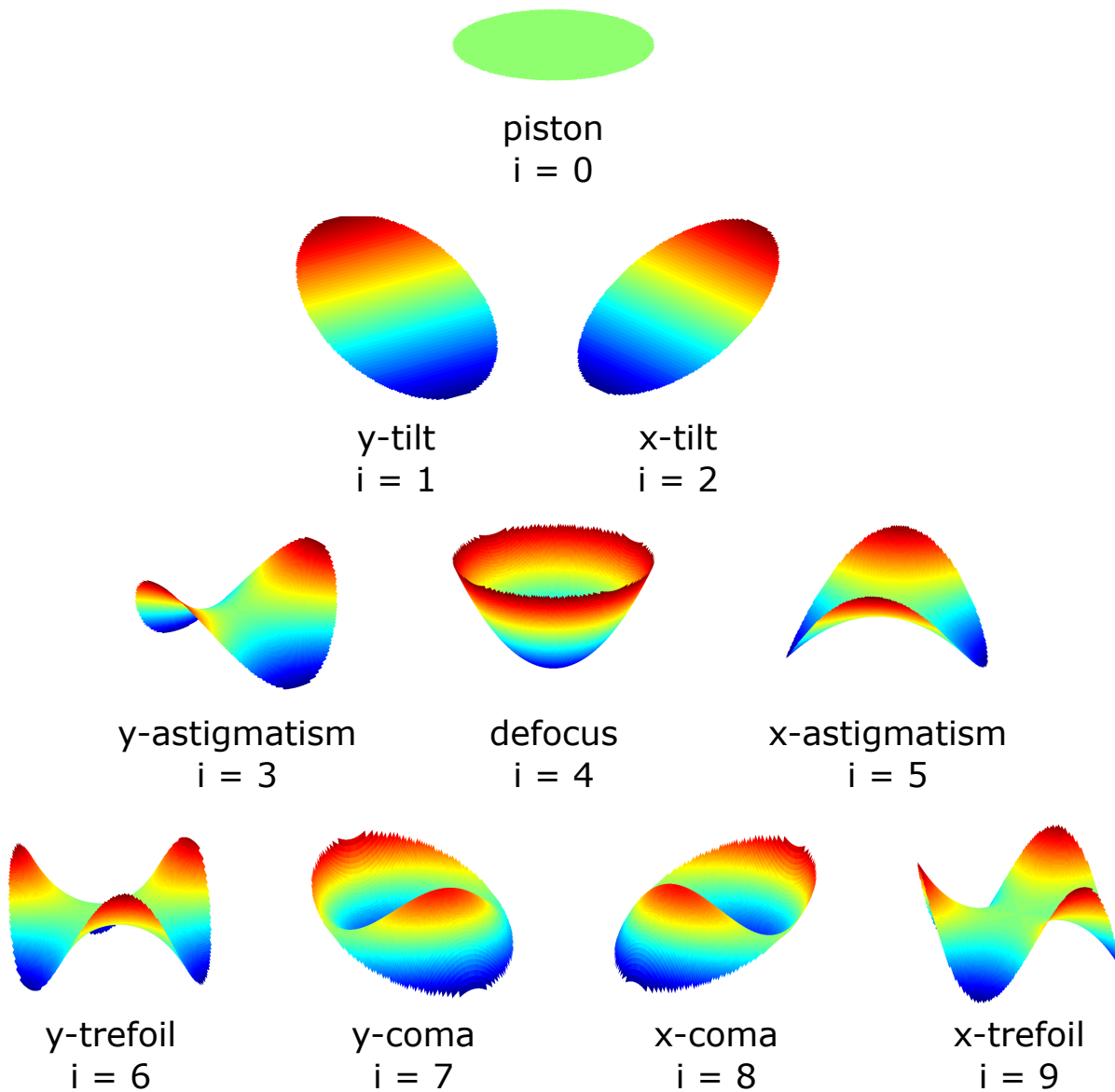


Figure 13: The first 10 Zernike modes.

i	n	m	Zernike Polynomial	Name
0	0	0	1	piston
1	1	-1	$2\rho \sin \theta$	y -tilt
2	1	1	$2\rho \cos \theta$	x -tilt
3	2	-2	$\sqrt{6}\rho^2 \sin 2\theta$	y -astigmatism
4	2	0	$\sqrt{3}(2\rho^2 - 1)$	defocus
5	2	2	$\sqrt{6}\rho^2 \cos 2\theta$	x -astigmatism
6	3	-3	$\sqrt{8}\rho^3 \sin 3\theta$	y -trefoil
7	3	-1	$\sqrt{8}(3\rho^3 - 2\rho) \sin \theta$	y -coma
8	3	1	$\sqrt{8}(3\rho^3 - 2\rho) \cos \theta$	x -coma
9	3	3	$\sqrt{8}\rho^3 \cos 3\theta$	x -trefoil
10	4	-4	$\sqrt{10}\rho^4 \sin 4\theta$	y -quadrafoil
11	4	-2	$\sqrt{10}(4\rho^4 - 3\rho^2) \sin 2\theta$	y -secondary astigmatism
12	4	0	$\sqrt{5}(6\rho^4 - 6\rho^2 + 1)$	spherical aberration
13	4	2	$\sqrt{10}(4\rho^4 - 3\rho^2) \cos 2\theta$	x -secondary astigmatism
14	4	4	$\sqrt{10}\rho^4 \cos 4\theta$	x -quadrafoil

Table 1: The first 14 Zernike polynomials and their common names.

2.5 Optical Aberrations

A major benefit of choosing the Zernike polynomials as a basis set is that the lower order modes correspond to common optical aberrations such as tilt, defocus, astigmatism and spherical. In this section, several of these aberrations will be discussed in detail and related to their physical causes in optical systems.

2.5.1 Tilt

Tilt is caused when there is a misalignment between the object and the optical axis. Figure 14 shows an example of tilt. Optical axis A is denoted by the solid black line to which object O is perpendicular. Lens L is tilted with respect to A by approximately 3° forming a tilted optical axis A_t indicated by the dashed lines. The blue and red rays originate from the bottom and top of object O , respectively. Image I is tilted with respect to the original axis A .

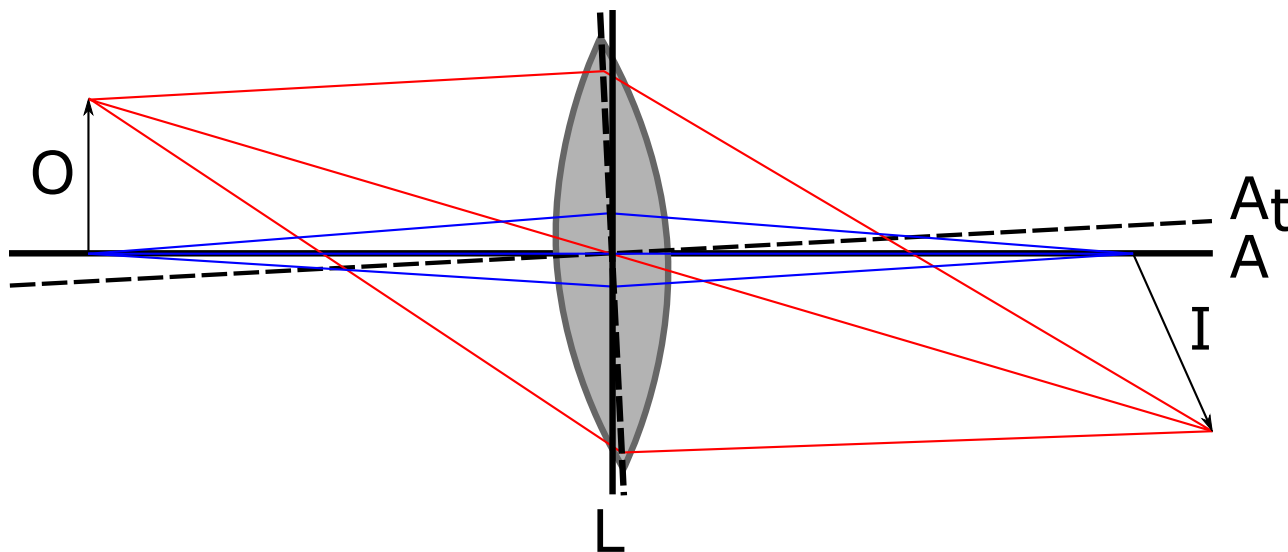


Figure 14: Tilt aberration. Optical axis A_t (dashed line) is tilted with respect to axis A (solid line) causing image I to be tilted with respect to object O .

In terms of Zernike modes, tilt is broken out into x and y modes with indexes $i = 1, 2$. The polynomials for x and y are $2\rho \sin \theta$ and $2\rho \cos \theta$ respectively as shown in Table 1.

2.5.2 Defocus

Defocus is a familiar aberration in which an image is out of focus. This occurs when the image plane is not coincident with the plane of best focus, typically caused by a translation along the optical axis. The Zernike polynomial $\sqrt{3}(2\rho^2 - 1)$ models the defocus aberration as shown in Table 1.

2.5.3 Coma

Coma is a variation of focus with field angle off-axis. Thus, point objects near the edge of the entrance pupil will obtain a 'comet-like' tail when an image is formed. Figure 15 shows a ray tracing of the coma aberration. Rays near the edge of the pupil are spread out whereas rays near the centre of the pupil are not. The bottom of the figure shows a point source and its image, with the comet tail aberration depicted. The coma aberration is modelled by Zernike modes $i = 7$ and 8 with y -coma polynomial being $\sqrt{8}(3\rho^3 - 2\rho)\sin\theta$ and the x -coma polynomial being $\sqrt{8}(3\rho^3 - 2\rho)\cos\theta$.

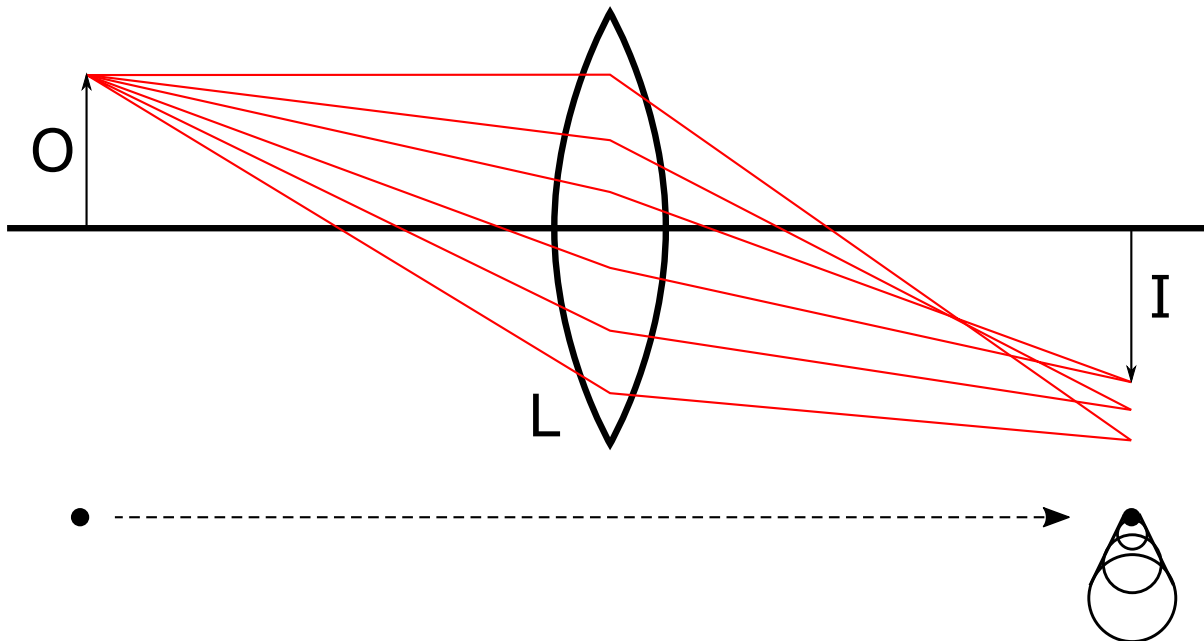


Figure 15: Coma aberration. Point source object O obtains a comet-like tail in image I due to coma aberration from lens L . The top part of the image illustrates the ray tracing while a depiction of the point source and image is shown below.

2.5.4 Astigmatism

Figure 16 shows a lens that exhibits astigmatism. Rays originating from point source O on the optical axis come to two different focal points f_x and f_y in the orthogonal directions. This can be due to defects in the lens or other effects such as mismatches in system alignment, etc. In the case of LRT astigmatism is introduced into the system by the spatial separation of the x and y galvo-scanner mirrors. Astigmatism can be purposely introduced into a system (or compensated for) by utilising a cylindrical lens.

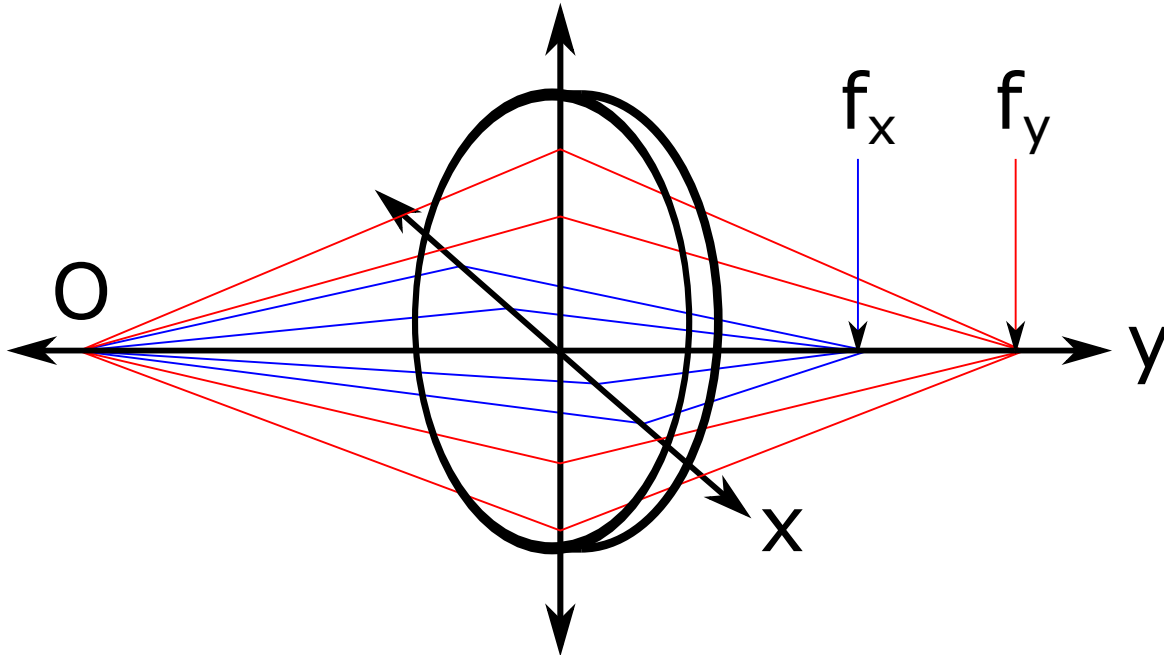


Figure 16: Astigmatism aberration. Rays originating from point source O are brought to a different focus in the x and y directions. Focal point f_x is depicted with blue rays while focal point f_y is depicted with the red rays.

2.5.5 Spherical

The spherical aberration due to a spherically shaped lens or mirror which causes rays incident near the edge of the pupil to be focused to a different location on the optical axis than rays incident near the centre of the pupil. This can be corrected with the use of parabolic lenses and mirrors but due to the ease of manufacturing spherical optics, this aberration is still commonly observed. Figure 17 shows rays emanating from a point source and being focused to differing locations due to a lens with spherical aberration. Zernike mode $i = 12$ with polynomial $\sqrt{5}(6\rho^4 - 6\rho^2 + 1)$ models the spherical aberration.

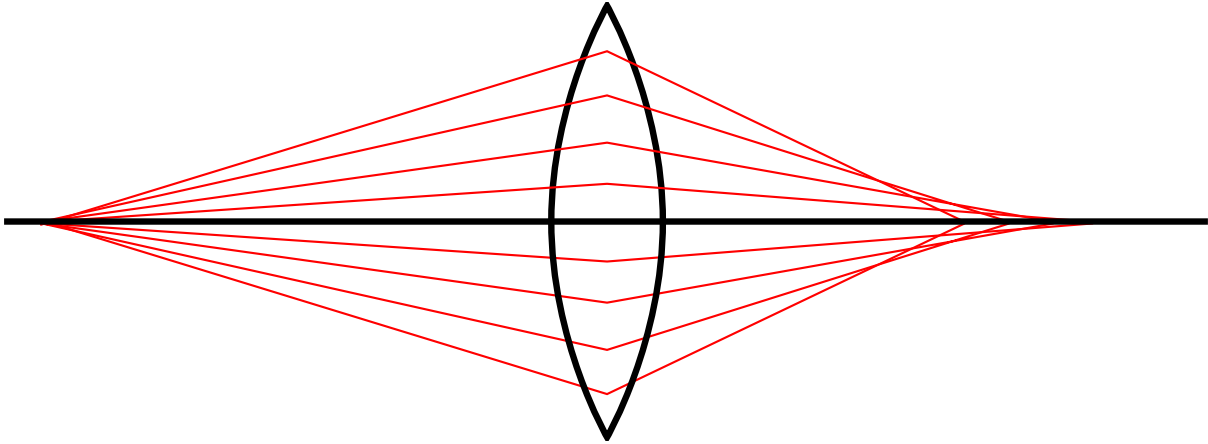


Figure 17: Spherical aberration. Spherical shape of the optic causes the focal point to differ for rays entering near the edge of the pupil opposed to those entering near the centre.

2.6 Conclusion

This chapter discussed the basic principles of wavefront sensing and provided the foundational material needed for an in-depth look at coherence-gated wavefront sensing. The SHS was used as a starting point, from which the general operating principles of wavefront sensing were presented. Next, the zonal and modal method of wavefront recovery were touched on, with emphasis given to the modal method, as that is the technique used in the CGLRT instrument. This led into a discussion of Zernike modes as wavefront descriptors and finally, how those modes relate to common physical aberrations in optical systems. Next, Chapter 3 will show how the combination of low-coherence interferometry (LCI) and basic wavefront sensing allows for the realisation of depth-resolved wavefront sensing. First, a brief discussion of interferometry and the common interferometer configurations are presented. Then, the major literature on coherence-gated wavefront sensing is discussed including its origins in biological imaging. Again, the focus on SHS is appropriate, as the major literature deals exclusively with this modality of wavefront sensing. After this survey of the literature and principles of coherence-gated wavefront sensing, the discussion will turn to topic of the laser ray tracing wavefront sensor in the next chapter.

3 Coherence Gated Wavefront Sensing

3.1 Introduction to Coherence Gated Wavefront Sensing

The concept of coherence gating in microscopy and wavefront sensing derives from the principles of LCI, phase-shifting interferometry (PSI) [25], and OCT [20]. In each case, an interferometer (typically a Michelson or Mach-Zehnder), is combined with a standard wavefront sensor. The previous chapter focused on the basics of wavefront sensing, whereas this chapter will begin with a review of the basic techniques of interferometry. From there the discussion will continue with a review of the pertinent literature on coherence-gated wavefront sensing (CGWFS).

Perhaps the most well known interferometer is the Michelson interferometer, made famous by its use in the Michelson-Morley experiment which provided evidence for the non-existence of the *aether* [27]. The basic setup of the interferometer is shown in Figure 18. In it an optical source, either monochromatic or broadband, delivers a collimated beam to a beam splitter BS which splits the beam into a reference and an object arm (blue and red, respectively in Figure 18). These beams reflect off of the reference mirror RM and the object mirror OM and are recombined by BS. The recombined beam is incident upon the detector where interference fringes may be measured. The output of the detector may be digitized and processed by a computer.

Another common interferometer configuration, known as a Mach-Zehnder is shown in Figure 19. In it, optical source OS launches a beam (purple) to beam splitter BS1. This splits the beam into the two arms of the interferometer, shown as red and blue. Mirrors M1 and M2 direct the beams to beam splitter BS2 where they are recombined. Half of the light is incident upon detector DET and the other half is absorbed by beam block BLK. A modified form of the Mach-Zehnder interferometer is utilised in the realization of the CGLRT instrument.

An interferometer with a monochromatic coherent source will exhibit an interference pattern that repeats over 2π intervals as the path length of the interferometer (usually in the reference arm) is modulated. The coherence length of an optical source is defined as the distance over which the temporal coherence degrades by a given amount, depending on the application. For a Michelson interferometer an optical path difference of one coherence length will cause the fringe visibility to drop by 37% [1]. In the case of a monochromatic source the coherence length is typically quite long and may range from 10s of cm to 100m or more depending on the type.

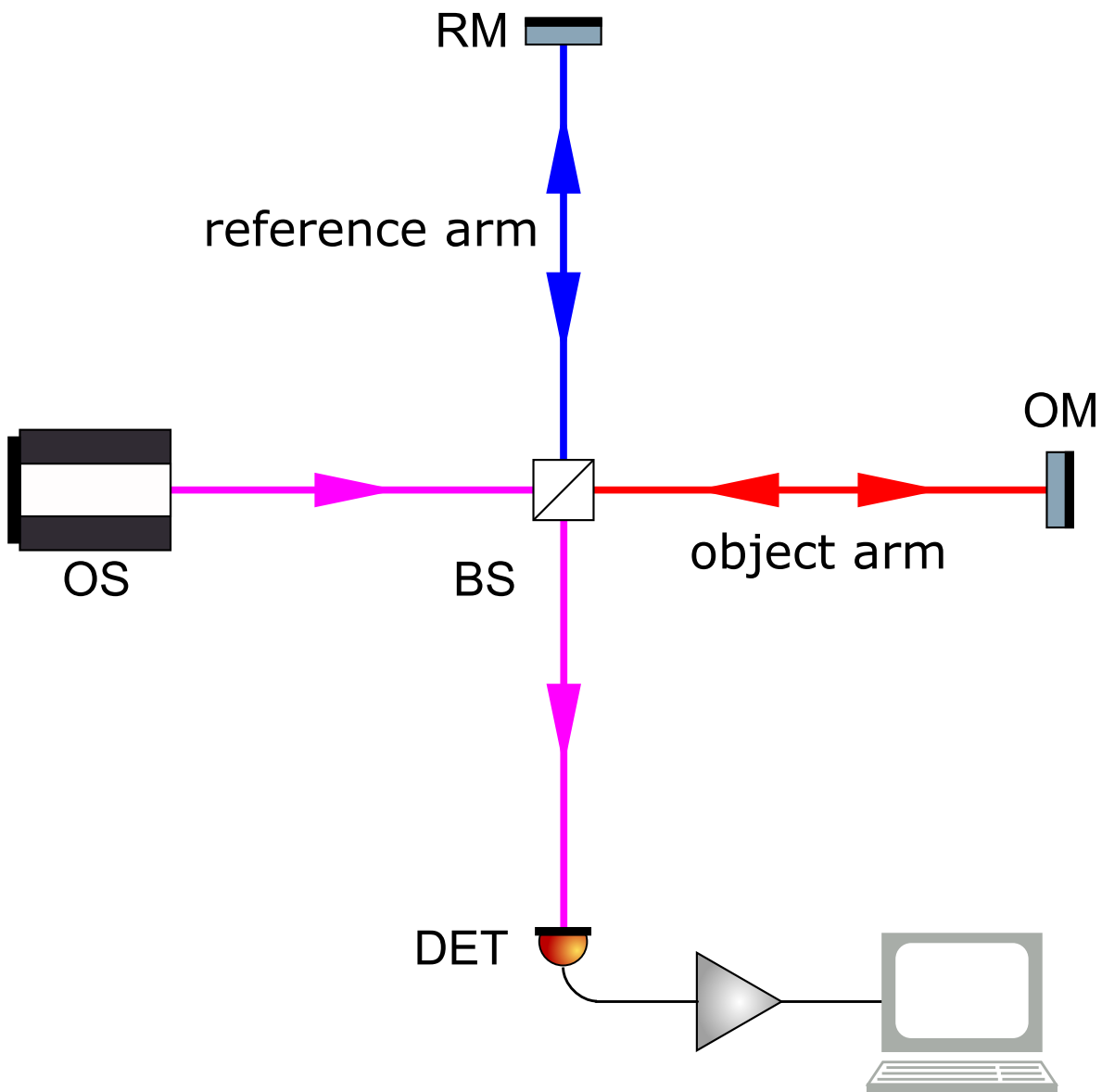


Figure 18: A Michelson interferometer. Optical source OS launches a beam (purple) to beam splitter BS. BS splits the beam into the reference (blue) and object (red) arms of the interferometer. The reference mirror RM and object mirror OM reflect the beams back to BS where the beam are recombined and sent to detector DET. The signal is amplified, digitized, etc., before being read into a computer for post processing.

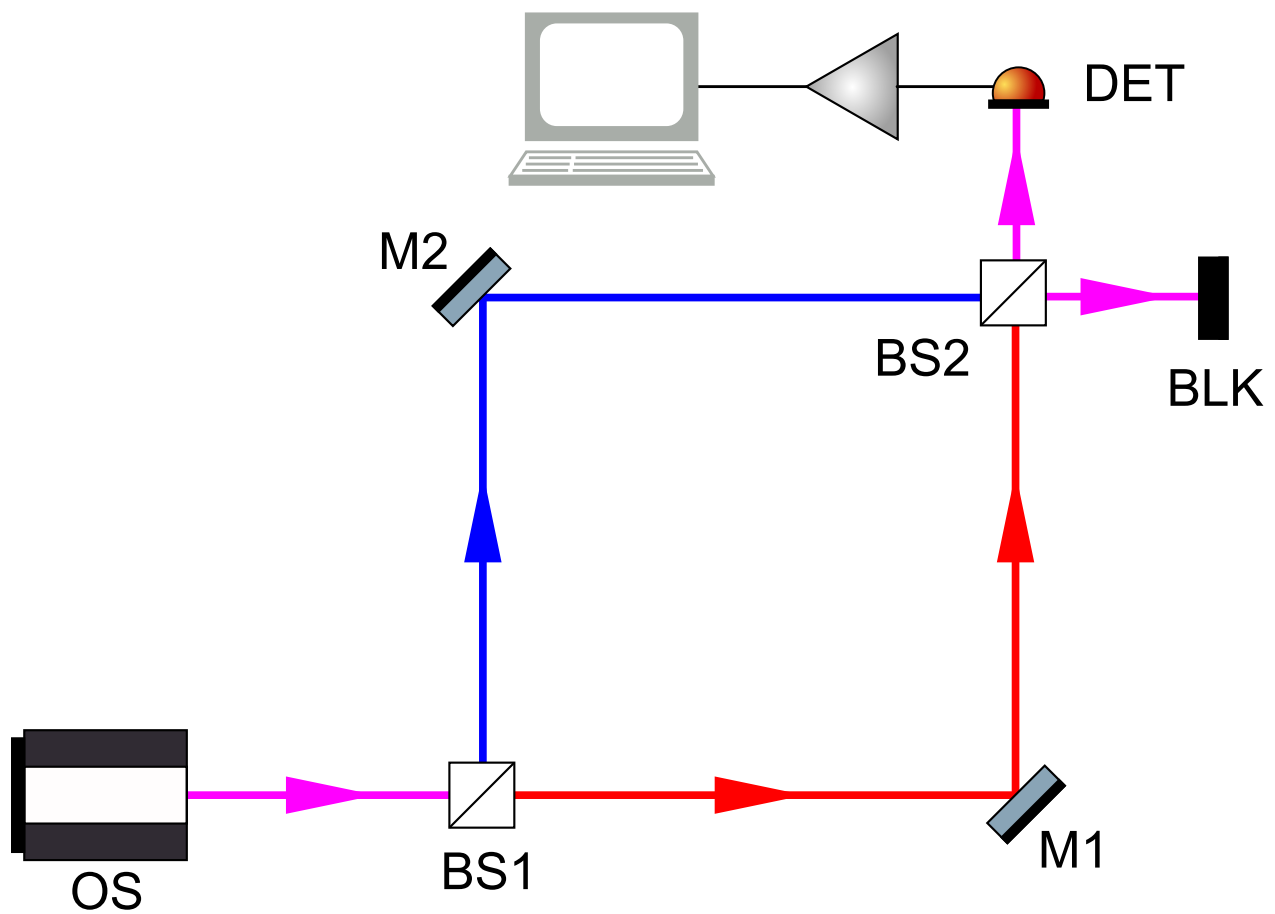


Figure 19: A Mach-Zehnder interferometer. Optical source OS launches a beam (purple) to beam splitter BS1. This splits the beam into the two arms of the interferometer, shown as red and blue. Mirrors M1 and M2 direct the beams to beam splitter BS2 where they are recombined. Half of the light is incident upon detector DET and the other half is absorbed by beam block BLK. BLK may be replaced by an alternate detector if desired.

If, instead, the interferometer is fed by a broad-band source, significant interference will only occur where the optical path difference is within the coherence length of the source. The coherence length l_c of a broad band source is defined as:

$$l_c = \frac{4 \ln 2}{\pi n} \frac{\lambda_0^2}{\Delta\lambda} \quad (3.1)$$

where λ_0 is the centre frequency of the broad band source, $\Delta\lambda$ is the full-width half-maximum (FWHM) bandwidth, and n is the refractive index.

For a low-coherence source with a central wavelength $\lambda = 675 \text{ nm}$ l_c is on the order of tens of microns and, clearly, as the bandwidth of the source increases the coherence length decreases. It is this very property of low-coherence interferometry, which naturally selects a very limited OPD, that allows selection of light returned solely from within the *coherence gate*. For example, coherence gating can be used to enhance axial resolution in *en face* microscopic imaging [28] by eliminating spurious light from outside the focal plane. For WFS, coherence gating may be employed to select only the light being returned from the layer of interest in a multilayer sample, allowing wavefront correction to ignore aberrations from layers beyond the one of interest [9]. Coherence gated WFS has been employed in both strongly scattering biological samples [9] and in two-photon microscopy for imaging of biological tissue [49].

3.2 CGWFS in Biological Samples

A general approach to coherence-gated WFS in scattering biological tissues is described in the paper, “Coherence-gated wave-front sensing in strongly scattering samples” by Feierabend, Rückel and Denk [9]. This paper describes an experimental setup with the intention of verifying the feasibility of a CGWFS utilising a low-coherence interferometer. The authors note the desire to perform wavefront sensing, similar to that performed by astronomers, but in strongly scattering biological samples. In astronomy, a guide star is often used to sense wavefront distortions in the atmosphere, but in biological samples no such option exists outside of fluorescence which can be damaging to tissues. They demonstrate that light returned only from the focal region of a scattering sample may be selected by coherence-gating which, instead of utilising a guide-star, matches time-of-flight between photons returned from the sample (object arm) and photons in the reference arm of the interferometer.

Coherence-gating is achieved via the techniques of phase shifting interferometry [4, 40], which is able to extract phase information from a series of interferograms. The four-step

PSI algorithm, as described by Malacara [25] was used to compute the complex amplitude from a quadruplet of images (I_1, I_2, I_3, I_4) . For each image, the reference arm path length is shifted by a quarter wave: $0, \lambda/4, 2\lambda/4$ and $3\lambda/4$. The complex amplitude A may then be calculated from the four interferograms:

$$A = (I_2 - I_0) + i(I_3 - I_1). \quad (3.2)$$

This may also be expressed in terms of the intensity only:

$$I_D = \sqrt{(I_2 - I_0)^2 + (I_3 - I_1)^2}, \quad (3.3)$$

where I_D is the intensity profile of the image.

Interestingly, the authors did not utilise a physical SHS but instead processed the four interferograms with a *virtual Shack-Hartmann sensor* (vSHS). The vSHS outputs a spot pattern, the same as one would expect from a physical Shack-Hartmann sensor, which may then be analyzed in the typical fashion via a least squares fit to Zernike polynomials. The vSHS is discussed in more detail in 3.3.

Next, the authors introduced known aberrations into the system and verified that CG-WFS behaved as expected. First, they introduced defocus by changing the reference-arm length. As the coherence-gate position changes relative to the focus, defocus is increased. As expected, the defocus term dominates while all other Zernike coefficients remain small and unchanged. In a second test, a glass coverslip (BK7, 180 μm thick) was placed between the sample and objective and tilted by $\sim 15^\circ$. This introduced astigmatism, coma, and spherical aberrations. The authors once again shifted the CG position (in the presence of the coverslip) and verified that only the defocus term varied. The authors conclude that CGWFS is able to successfully utilise backscattered light, isolated via coherence-gating, to reconstruct a measured wavefront while rejecting light from outside the coherence-gate. Another important paper, by the same authors [47], discusses the functioning of the vSHS, which will be treated in the following section.

3.3 Virtual Shack-Hartmann WFS

In [47] the authors discuss the implementation of a vSHS with an emphasis on noise performance in comparison to a physical SHS. They note that most applications of AO for confocal

microscopy require monitoring fluorescing light from the biological sample (similar in function to the ‘guide star’ in astronomical AO). This approach can be problematic though, as the specimen must be stained, and there is often accompanying damage to the tissue from photo-bleaching. As an alternative to fluorescence, light that is back-scattered from the focal plane of the confocal microscope may be utilised for wavefront sensing. The critical issue though, is that light from outside the focal plane must be rejected by the wavefront sensing system. If this were not the case, the AO system would be attempting to correct for distortions in wavefronts which were not being imaged. By employing coherence-gating, only the light back-scattered from the focal-plane of the confocal imaging system may be selected. This ensures that the AO system is only responding to wavefronts that are currently being imaged.

As in a physical SHS the vSHS divides the aperture up into $m \times m$ subapertures, but numerically instead of with a lenslet array. A discrete FFT is performed on A (from Equation 3.2) for each subaperture. This produces a diffraction pattern for each subaperture analogous to that formed by the lenslet of a physical SHS. The centroid of each diffraction pattern may then be calculated, slopes determined, and the wavefront reconstructed. The authors used a standard least-squares fitting method to reconstruct the wavefront with 28 Zernike modes.

Though the vSHS is fully parallel and does not require hardware it does have drawbacks, specifically, sensitivity to photon and speckle noise. In [47] they note that below photon numbers (returned from the sample) of approximately 10^4 the wavefront error was very large, on the order of one to three microns for an input wavefront with deviation of $0.1 \mu\text{m}$. This can be dealt with, partially, by increasing the diameter of the individual lenslets, but at the cost of sacrificing the ability to calculate higher-order Zernike modes. Also, according to [47] and [9] this method will fail in the presence of speckle noise. In [9] the authors were required to take 600 image quadruplets (for the four-step algorithm) at 24 different positions (shifted by $0.1 \mu\text{m}$) to reduce the speckle noise. The authors do not provide a quantitative value for the reduction. In terms of processing time neither [47] nor [9] provide any discussion regarding the computational time required to implement the vSHS wavefront recovery.

3.4 Coherence-Gated Shack-Hartmann WFS

Thus far, the motivation for utilising CGWFS has been to perform wavefront sensing on strongly scattering biological samples. While this is true other benefits exist such as rejecting stray reflections from the optical imaging systems themselves. In microscopy systems, for

example, there may be many sources of stray reflections including multifaceted microscope objectives, multiple reflections off of microscope slides, and for retinal imaging the eye lens and cornea. This also means that CGWFS systems can more easily cope with complicated optical imaging systems.

Ideally, it is desirable for CGWFS systems to operate at a high-speed so that they may correct for distortions in real-time imaging systems such as high-speed OCT. To achieve this it is preferable for the wavefront sensing system to minimize the calculations necessary to reconstruct the wavefront. For this reason, it is prudent to explore hardware based alternatives to the mathematically complex vSHS. By transferring the virtual lenslet array calculation to a hardware based SHS with a physical lenslet array one may realise a CGSHS. This does away with the need to:

1. Partition the image into subapertures, forming the virtual lenslet array.
2. Calculate the Fourier transform over each lenslet (including zero-padding).

The implementation of CGSHS is first explored by Tuohy and Podoleanu [59]. The authors first discuss some of the limitations of current (non-coherence-gated) SHSs, including the inability to differentiate aberrations from differing layers of a multi-layer specimen, as discussed above, and sensitivity to stray reflections.

The physical implementation of a CGSHS does away with a majority of the calculations above, trading off computational complexity for optical hardware complexity. The system is realised in the form of a modified Michelson interferometer. The experimental setup is shown in Figure 20. The interferometer is fed by a broad band optical source OS. Light from the source is collimated by lens L1. Beam splitter BS1 splits the beam into reference and object arms. The object beam traverses beam splitter BS2, lenses L2 and L3 and is incident upon deformable mirror DM which is used to introduce aberrations into the system. Lenses L2 and L3 were utilised to introduce strong stray reflections into the system. The reflected beam is picked off by splitter BS2 where it encounters the micor-lenslet array MLA, forming a grid of pencil beams. In the reference arm the beam is incident upon mirror RM that can be modulated by a piezo-electric transducer PZT. The reference beam is recombined with the object arm pencil beams via steering mirror M1 and splitter BS3. L4 and L5 form a telescope that image the combined beams onto the CCD camera.

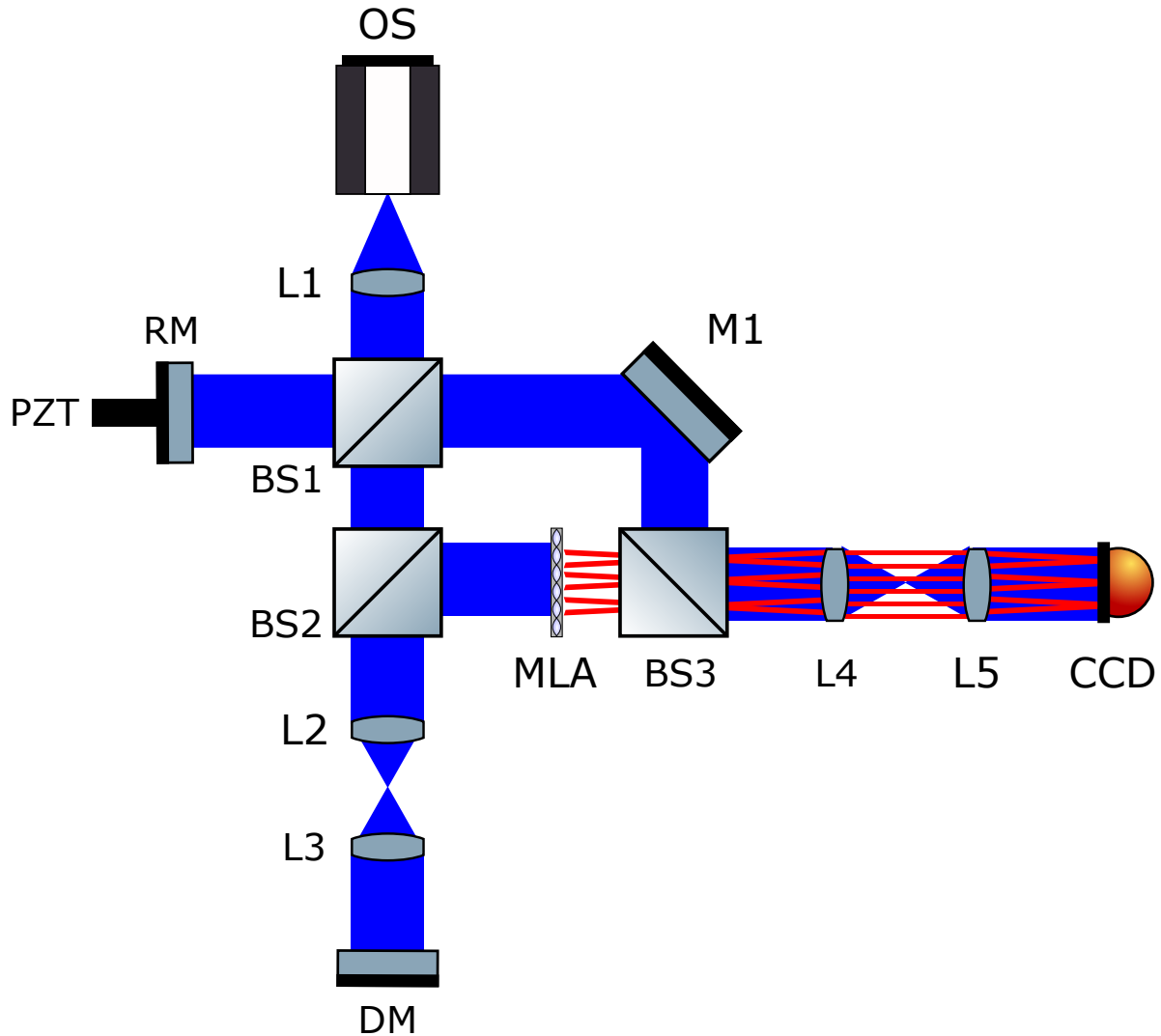


Figure 20: The experimental set up used in [59]. OS: Optical source. BS1-BS3: Beam splitters. L1: 10x microscope objective L2-L5 achromatic doublet lenses. DM: Deformable mirror. RM: Reference mirror. PZT: Piezo-electric transducer. CCD: Camera. MLA: Micro-lens array. M1: mirror.

For the time-domain implementation the reference mirror RM is mounted on a piezo-electric actuator which is used to translate the mirror a distance equal to the central wavelength λ of the source. As discussed in Section 3.2 this translation is implemented such that four images, each with a phase shift of $\lambda/4$ is obtained. The piezo control voltage is synchronized via a PC to the CCD acquisition signal so that images are obtained at the appropriate phase offset. Using the principles of PSI [25] the interferogram is obtained from Equation 3.3.

In discussing results, the authors first look at the rejection of stray light by comparing spot patterns of CG and non-CG SHS. In the non-CG case, the reference beam of the interferometer was blocked, reducing the operation to that of a normal SHS. In this case, there are many stray reflections in the image. When the CG is applied and the image generated (via the method discussed in Section 3.2) the stray reflections have been removed and the spot pattern is clearly improved. In the case of both unaberrated (DM is flat) and aberrated wavefronts the spot pattern is cleaner, yet retains the spot deviations one would expect. The authors also note that the overall signal level is enhanced in the CG case. This is due to the combination of the relatively weak signal from the object arm being combined with the relatively strong signal from the reference arm.

It is important to note, with regards to CGLRT that the setup above did not utilise a multi-layer object. The use of deformable mirror DM to introduce aberrations was useful in demonstrating the functionality of the system but only for a single-layer target. It was left to the next paper, “Demonstration of real-time depth resolved Shack-Hartmann measurements” by Wang and Podoleanu [61], to demonstrate depth-resolved wavefront sensing for a multi-layer object.

In [61] the authors demonstrate real-time operation of a time-domain CGSHS. In this paper Shack-Hartmann spots are obtained from a multi-layer sample. This setup was used to show that individual wavefronts returned from a multi-layer sample could be resolved while also rejecting stray reflections from system optics.

The setup for the system is shown in Figure 21. In it, a Super Luminescent Diode (SLD) source feeds the LCI. The beam is collimated and split via BS1 into reference and object arms. In the object arm the multi-layer sample OM is interrogated by the beam and sent via splitter BS3 to micro-lenslet array MLA. Here, the pencil beams are overlaid on the reference beam via splitter BS4. Lenses L3 and L4 form a telescope which images the spots on the sensor (CMOS camera). In the reference arm mirror RM is mounted to piezo PZT to provide path length modulation.

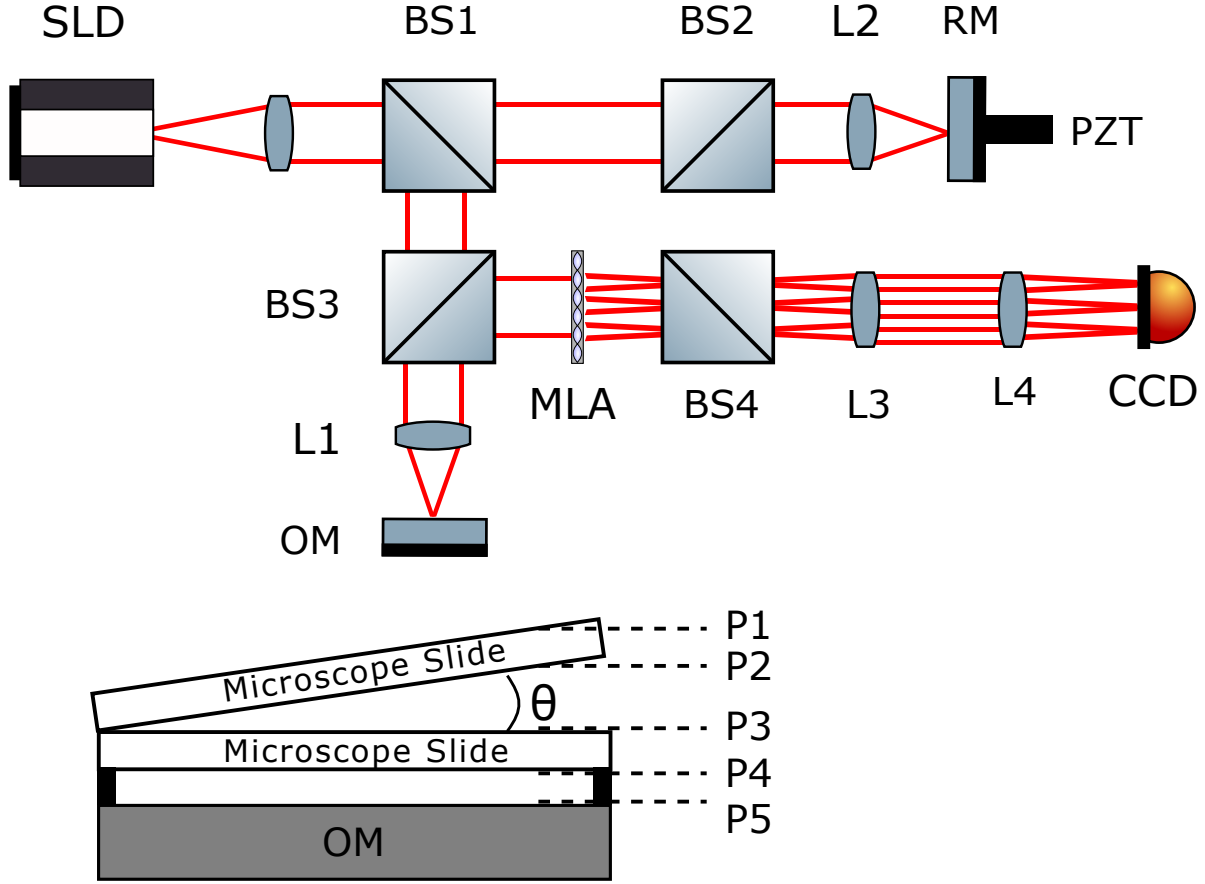


Figure 21: The experimental set up used in [61]. SLD: Super luminescent diode. BS1-BS4: Beam splitters. L1-L4 achromatic doublet lenses. OM: Object mirror. RM: Reference mirror. PZT: Piezo-electric transducer. CCD: Camera. MLA: Micro-lens array. Inset: The multi-layer object consists of two microscope slides mounted in front of the object mirror OM. This forms surfaces P1 through P5 as shown.

To verify the functionality of the CGSHS a multi-layer sample was constructed from two microscope slides and a mirror as shown in the subset of Figure 21. The two microscope slides, combined with the mirror, may be interrogated in depth. Each glass/air boundary (P1 - P4) creates a reflective surface from which a wavefront will be returned. With the object mirror OM providing the final surface P5. Spot images were obtained for P1 - P5, both with and without CG. The first microscope slide is intentionally tilted by a small amount to introduce a low-order aberration to the object. In each case the reference mirror was translated to the $OPD = 0$ position for the associated surface. For the non-CG cases, each image was clearly dominated by the strong reflection from the object mirror with small contributions from P1 - P4. In contrast, in the CG version each spot image contains only contributions from the associated $OPD = 0$ surface, with complete suppression of the dominant OM signal.

Finally, Zernike polynomials were used to evaluate each set of CG spots. As expected the tilt in y term dominates due to tilted slide. Small amounts of defocus were also present due to a mismatch of the focal point and the coherence gate location. In this way the authors demonstrate a functioning CGSHS capable of measuring wavefronts from a multi-layer object while rejecting strong reflections from other layers.

3.5 Conclusion

This chapter has covered the basic principles and literature regarding coherence-gated wavefront sensing. The principle of coherence-gating allows for the selection, in depth, of light returned from a multi-layer object by exploiting the short coherence-length of a broad-band source. In each case, the use of an interferometer led to the realisation of a depth-resolved wavefront sensor. In [9], the motivation for CGWFS is found in the need to add AO to systems which image strongly-scattering biological tissue. By using coherence-gating, the use of damaging fluorescence techniques may be avoided. The use of a virtual Shack-Hartmann WFS is discussed in [47]. While this technique was innovative, it is numerically complex. This led to the development of hardware based approaches to CGWFS, focusing on the marriage of a low-coherence interferometer with a standard Shack-Hartmann WFS. In [59], coherence-gating is employed to remove stray reflections from the system and a single-layer target is utilised. In [61], the system is expanded by use of a Mach-Zehnder configuration and coherence-gating is employed to resolve wavefronts from a multi-layer target.

In many ways, the systems discussed in this chapter illuminate the motivation for exploring coherence-gated laser-ray tracing. CGLRT should be able to employ coherence-gating to remove stray reflections from the system. It should also be able to resolve wavefronts from single-layer and multi-layer targets. With this in mind the upcoming chapter discusses the initial implementation of the CGLRT system. After a brief discussion of the basic principles of laser ray tracing, the implementation of both single-layer target and multi-layer target versions of the system are discussed. For the first time, results are shown which demonstrate the ability of LRT to perform depth-resolved wavefront sensing for a multi-layer object. It is important to note the distinction, however, between this initial system and the final instrument discussed in Chapter 6. All of the data in Chapter 4 was processed using Labview and *not* on an embedded system. It will be left to Chapter 5 to discuss the embedded system architecture used to perform the measurements in real-time. The combined CGLRT instrument will be discussed in the final chapter.

4 Laser Ray Tracing Wavefront Sensor

4.1 An Introduction to Laser Ray Tracing

While the Shack-Hartmann design is perhaps the most widely known WFS other methodologies do exist. One such methodology, known as LRT, [35] is a wavefront sensing modality that more closely resembles the process of mathematical ray-tracing typically used to design optical systems. LRT functions by interrogating a target sequentially with a thin pencil beam. The beam is articulated by a scanner (usually an x-y galvo, gimbal mounted mirror, acousto-optic modulator, or similar device) and the location of the returned beam is sensed via a position sensitive detector, CCD camera, or similar device. In the absence of aberrations the scanning beam will return to its nominal position on the detector, but in the presence of aberrations the beam will be shifted by an amount proportional to the gradient of the aberration at the point of interrogation. By scanning many points on the surface of interest a spot pattern similar to that of a SHS is produced. LRT devices are commercially available and are used in ophthalmic applications such laser ablative surgery and corneal tomography [57]. LRT has several benefits over the more common SHS including:

- A user definable scanning pattern, in contrast to a SHS, in which it is fixed due to the lenslet array.
- The physical implementation is a closer approximation to geometric optics methods typically used to design optical systems making the system easier to conceptualise.
- Serialized data output which is well suited for embedded systems processing using a Field Programmable Gate Array (FPGA).
- No need for segmentation and indexing tasks associated with SHS, which are time intensive and prone to error if, for instance, a spot moves outside its allotted segment due to large aberrations.

LRT also exhibits several disadvantages when compared to SHS:

- The serialised nature of the scanning is slower than the parallel acquisition of SHS due to the time needed to move the scanning beam mechanically.
- The need for scanning makes the system more mechanically complex by adding the need for a beam actuator such as a galvo-scanning mirror.

In its basic form LRT exhibits a large depth of field which makes it incapable of differentiating wavefronts returned from an object with multiple layers. This is important in imaging modalities such as Optical Coherence Tomography (OCT) that produce 3D volumetric images of multi-layer targets, for example, the human retina. Additionally, the large depth of field means that LRT is susceptible to stray reflections in the optical system. These limitations can be overcome by incorporating interferometric coherence-gating into a standard LRT which will be discussed in detail in Section 4.3 and Chapter 6.

LRT comes in two basic forms: single-pass and double-pass, depending on how many times the interrogating beam passes through the aberrating optics. In order to understand the basic principles of LRT let the discussion first examine the single-pass setup.

4.1.1 Single-pass LRT

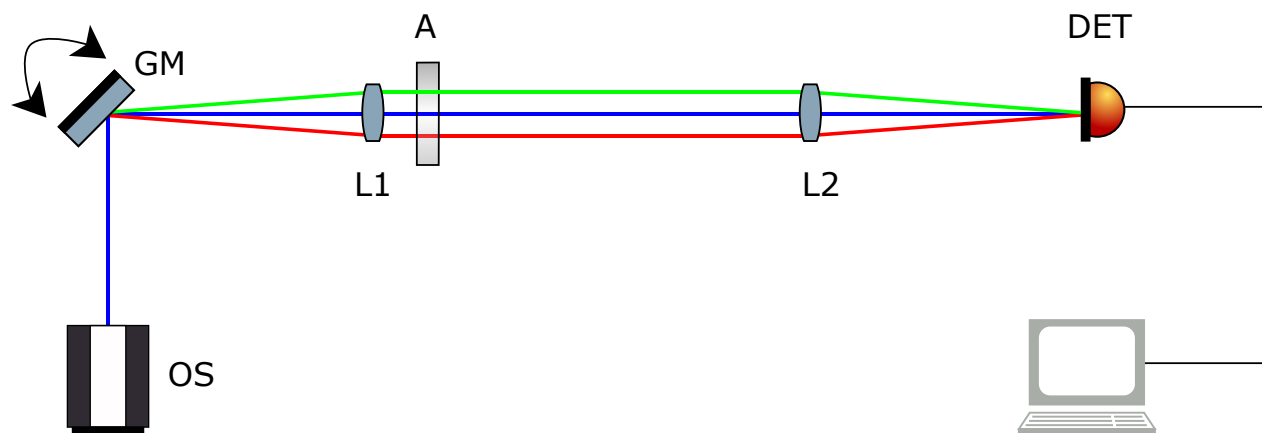


Figure 22: A basic single-pass LRT. OS: Optical Source, typically a laser or super-luminescent diode. GM: Galvanometer-Scanning Mirror. L1 & L2: Lenses of equal but arbitrary focal length. A: Aberrating plate, in this example, with zero aberrations. DET: The detector, which may consist of a position sensitive detector (PSD), quad-cell, charge-coupled device (CCD) camera, or similar device. Three different field angles are shown for reference.

Figure 22 illustrates a simplified single-pass LRT with no aberrations present. A thin ($\approx 500 \mu\text{m}$ diameter) collimated laser beam is launched from the optical source OS. This may be monochromatic, in the case of non-coherence gated LRT, or broadband, in the coherence-gated case. The beam is incident upon a scanning mirror GM which may be a galvanometer $x - y$ mirror, gimble mounted mirror, acousto-optic modulator (AOM), or the like (galvo-mirror shown in Figure 22). For reference, three beams are shown (green, blue and red), which represent the beam direction for three different scan locations of the mirror. The

scanner is at the focal point of lens L1, which collimates the articulated beams. In this example, the optic being interrogated is free of aberrations so the beam passes through with no change to its propagation direction. Lens L2 focuses the beam onto the detector DET which senses the beam position. In the aberration-free case the beam position on the surface of DET will remain constant as the various locations are scanned. Contrast this with the case shown in Figure 23 where the optic A is aberrated.

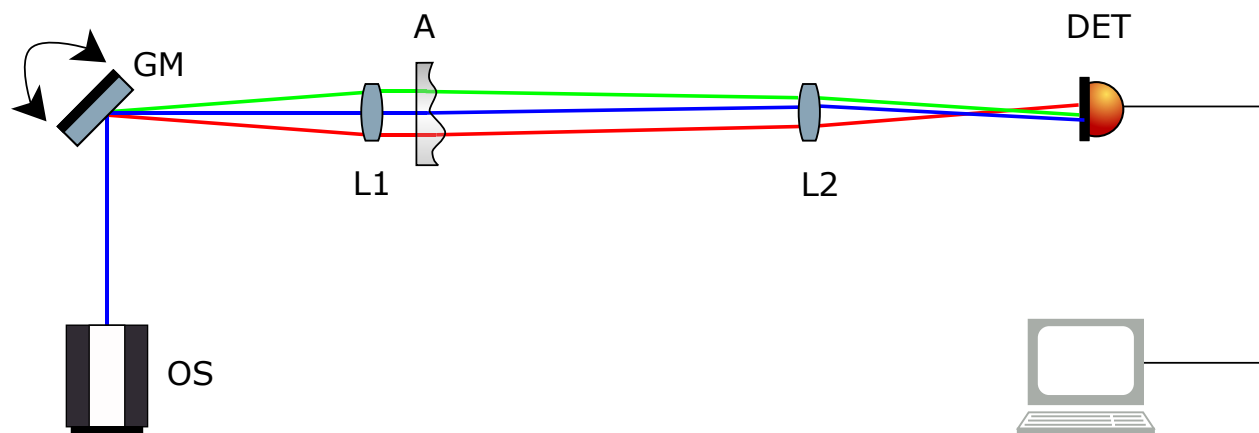


Figure 23: A basic single-pass LRT with aberrations. OS: Optical Source, typically a laser or super-luminescent diode. GM: Galvanometer-Scanning Mirror. L1 & L2: Lenses of equal but arbitrary focal length. A: Aberrating plate, in this example, imparting aberrations to the system. DET: The detector, which may consist of a PSD, quad-cell, CCD camera, or similar device. Three beams are shown for reference.

In the aberrated case the parallel beams exiting lens L1 are deflected from their nominal path by the local surface figure of A. Hence, when each beam is focused onto DET by lens L2, the location on the detector surface is offset by an amount proportional to the local wavefront aberration of A. In this way, lens L2 and DET act, in essence, as a single subaperture of a SHS. Once the complete set of beam offsets has been measured, the process of reconstructing the wavefront is analogous to the SHS case.

Molebny, et al., in 1997 [31], laid out an early iteration of the LRT concept which they dubbed *retina ray-tracing*. In this paper, the authors describe the retina ray-tracing device, which is similar to the double-pass LRT in Figure 26. For scanning, an AOM was utilised (opposed to a galvo-scanning mirror) which provided the system with a frame rate under 10 ms for a 65 point scan. For each scan position an image of the laser spot on the retina is projected onto a segmented PSD (quad-cell) which is discussed in detail in Section 4.2. The position of the laser spot on the quad-cell will be shifted away from the nominal location due to aberrations in the cornea. The magnitude of the shifts in the laser spot position

may be used to reconstruct the aberration of the eye in terms of Zernike polynomials. In Section 3 of the paper the authors show that an imperfect eye may be modelled as a perfect eye combined with a distorted wavefront W . In this way, the standard wavefront analysis and reconstruction techniques (Section 2.4) may be used. Finally, the authors show that the x and y displacement of the laser spot on the quad-cell relates to the gradients of W by Equations 2.1.

In a follow up paper “Principles of Ray Tracing Aberrometry” [32], the same authors describe an LRT instrument developed at the Institute of Biomedical Engineering at Kiev with testing performed at the Vardinoyannion Eye Institute of Crete. This technology was eventually commercialized and sold as the *iTrace* by the company Tracey [57]. A detailed diagram of the instrument was not given, only a brief description. They describe an instrument similar to that in [31] where an acousto-optic modulator (AOM) is used to deflect a 0.3 mm diameter scanning beam sourced from a 10 mW diode laser ($\lambda = 650$ nm). The deflected beam interrogates the eye through a beam splitter and is kept parallel to the optical axis of the eye using a lens. A PSD is used to detect the traverse position of the back-scattered light returned from the eye. The final instrument is stated as having a CCD camera and a fixation target to aid in aligning the eye with the wavefront sensing system. It is able to perform scans of 60 to 400 points in 10 to 20 ms. Reconstruction of the measured aberration is done in post processing with Zernike modes. It should be noted that the purpose of this instrument is to measure aberrations of the human eye for use in laser ablative surgery. It is not functioning as a wavefront sensor for the purposes of AO.

Another set of papers, considered to be the ‘classic’ LRT publications, are by Navarro and Moreno-Barriuso. This first of these, published in 1999 [35], describes a simple, single-pass, implementation of an LRT as shown in Figure 22. This paper does not focus on measuring the aberrations of the eye, but instead compares the results of LRT with a SHS integrated into the same setup. The LRT setup used in this paper is shown in Figure 24. The SHS setup is shown in Figure 25. For the LRT, a He-Ne laser (OS) with a beam diameter of ≈ 0.6 mm is sent to XY scanner GM. This directs the beam (3 possible deflection angles shown) to high-quality collimating lens L1 and aberrating plate A. Lens L2 focuses the beam onto the surface of a CCD camera. For the SHS, optical source OS is sent through a spatial filter and pinhole (not shown). The beam is collimated by lens L1 and sent through aberrating plate A. Lenses L2 and L3 form an image of L1’s pupil plane on the micro-lens array MLA. The micro-lens array is composed of 91 microlenses, each with a diameter of 0.5 mm and focal length $f=50$ mm. The spots formed by MLA are imaged onto the CCD camera by lenses L4 and L5.

The LRT was programmed to deliver a scan pattern identical to that produced by the MLA in the SHS setup, so the comparison was ideal. To obtain a reference wavefront, plate A was removed from the setup. Then A can be placed in the setup and aberrations measured. Different aberrations could be induced by inserting spherical and cylindrical lenses or warped pieces of glass. In this way, wavefronts measured by the LRT and SHS are compared directly. Results are reported in terms of Zernike modes. The differences between the two systems was small, with an RMS error of approximately 12%. The LRT exhibited consistently smaller aberrations, which the author theorizes is due to the simplicity of the setup. The additional optics of the SHS may contribute to the slightly higher aberrations. This paper validated the concept of single-pass LRT and showed that the resulting aberrations measured by both LRT and SHS are consistent.

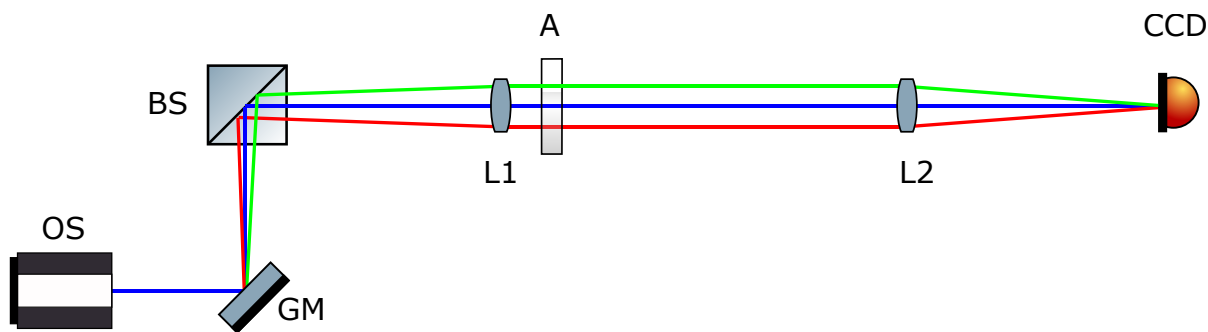


Figure 24: Single-pass LRT from [35]. Optical source OS launches a thin pencil beam at galvo-scanning mirror GM. The beam is directed toward beam splitter BS, through collimating lens L1 and optional aberrating plate A then to lens L2 to focus it onto the CCD camera. Three beams are shown for reference.

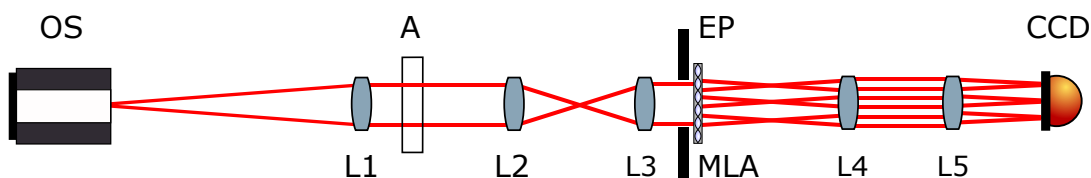


Figure 25: Shack-Hartmann setup from [35] Optical source OS launches a beam to collimating lens L1. The collimated beam passes through optional aberrating plate A. Lenses L2 and L3 form an image on the micro-lens array MLA through pupil stop EP. Lenses L4 and L5 form an image of the spots on the CCD camera.

4.1.2 Double-Pass LRT

While the single-pass case can be useful for measuring transmissive optical components it is unable to function in situations, such as the human eye, where transmission is not possible.

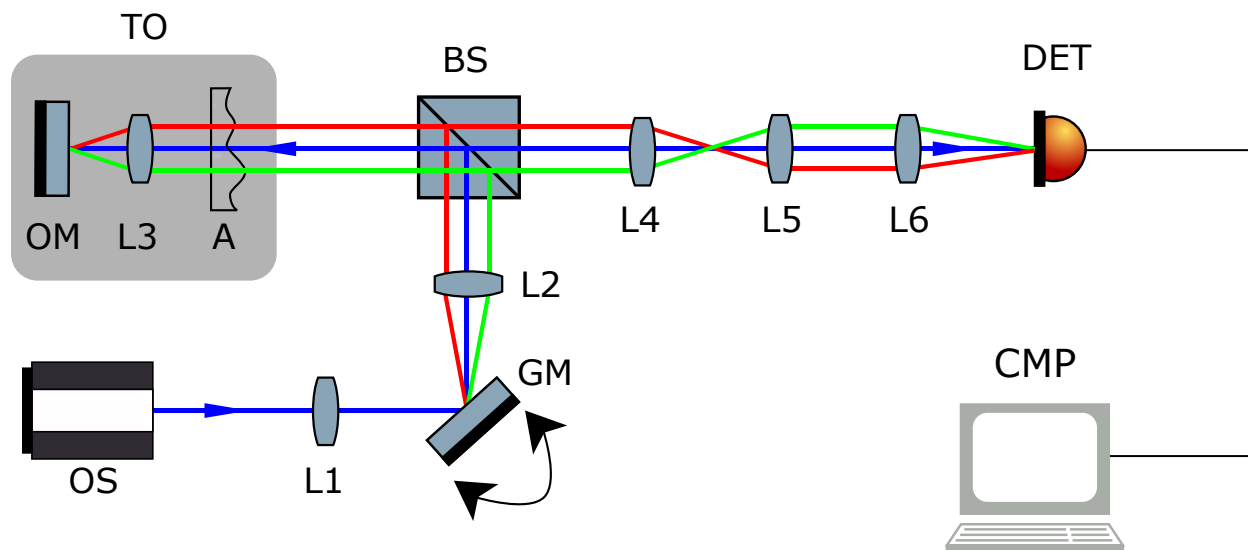


Figure 26: A basic double-pass LRT. Optical source OS launches a thin collimated beam to lens L1 which focuses the beam on galvo-scanning mirror GM. The beam is re-collimated by lens L2 and sent to beam splitter BS. The beam then interrogates the test object TO. TO is composed of aberrating plate A, lens L3, and object mirror OM. The beam reflects off OM and returns through BS to BS to lenses L4 and L5 which form an image of the pupil plane of TO. Lens L6 focuses the beam onto detector DET. Three field angles are shown for reference.

In this situation it is necessary to utilize a *double-pass* setup as shown in Figure 26. In this version, the collimated pencil beam is focused by lens L1 on the galvo-scanning mirror GM and, as in the single-pass example, the articulated beams are made paraxial by lens L2. Each beam is sent to the beam splitter BS which directs the light to the ‘eye model’ TO. The model consists of an (optional) aberrating optic A, short focal length lens L3 ($f = 35$ mm) and an object mirror OM. The pupil plane of the eye model is imaged by lenses L4 and L5. Lens L6 focuses the beam onto the detector DET. As in the single-pass case, the beam will be displaced off of the nominal position due to aberrations at the scan location.

To verify the functionality of a double-pass LRT, Moreno-Barriuso and Navarro compared wavefront measurements from LRT and SHS in both artificial and human eyes [33]. In contrast to the experiment in the previous paper, this setup incorporates both single-pass and double-pass measurements. The authors verified that the wavefronts obtained for single-pass and double-pass setups were equivalent, even though the interrogating light passes through the aberrated optics twice in the double-pass version. This is due to the fact that on the first pass the optics of the eye form a retinal image of the Gaussian laser spot. Upon reflection from the retina, the eye forms an aerial image of the spot in the pupil plane. The image is the cross-correlation between the Gaussian spot and the point spread function (PSF) of the

eye, so while the spot is blurred due to the PSF, the spot location (centroid) is not affected.

Together, these papers describe the foundational aspects of LRT as a wavefront sensing modality. With these tools in hand, as an initial setup, a standard double-pass LRT was built. This was used as a test-bed of sorts to build up the software and hardware resources necessary to begin exploring a coherence-gated LRT.

4.2 Standard LRT Experimental Setup

As an initial step, an LRT was constructed, without coherence gating, to allow for development of the hardware and software resources. A simplified block diagram of the setup is shown in Figure 27. In this implementation, National Instruments Labview software was utilised on a standard PC. Outboard hardware connections were made to the PC via a PCI-card and breakout board (NI PIC-6024E). The PCI card contains two, 12-bit digital-to-analogue converters (DACs) and 16 analogue-to-digital converters (ADCs). The breakout board contained all the necessary connections, via terminal blocks, to connect the ADCs and DACs to the LRT system. The two DAC signals were used to generate control signals for the galvo-scanning mirror, which is incorporated into the main body of the LRT optics setup (Figure 30 shows the internal optical layout). The output of the LRT is detected by a dual-axis position-sensitive detector. The current signal output is amplified and converted into a voltage by a transimpedance amplifier circuit and sampled by the Labview ADC (discussed above). The resulting samples are sent back to Labview for storage and post processing.

Initially, the LRT utilised a PSD as the detector, which is an analogue device capable of measuring the location of a light spot, on its surface, to very high accuracy. PSD functions are based on the well known photovoltaic properties of photodiodes but unlike a photodiode the active surface of PSD is diffuse over a large planar area.

There are two main types of PSD: segmented and lateral effect. A segmented PSD, often referred to as a quad-cell, is composed of four distinct detecting segments separated by a ‘dead zone.’ When the beam is incident across any of the segments a current proportional to the illumination on that segment is produced and by comparing these currents the position of the spot may be determined. The drawback with this approach is that if the beam happens to be placed totally in one segment then the position can no longer be determined. For example, the position of the red beam in Figure 28 can be determined, whereas, the position of the blue beam cannot as it is located completely in one quadrant. For this reason quad-cells tend to work best for monitoring a fairly static beam position confined to a small central

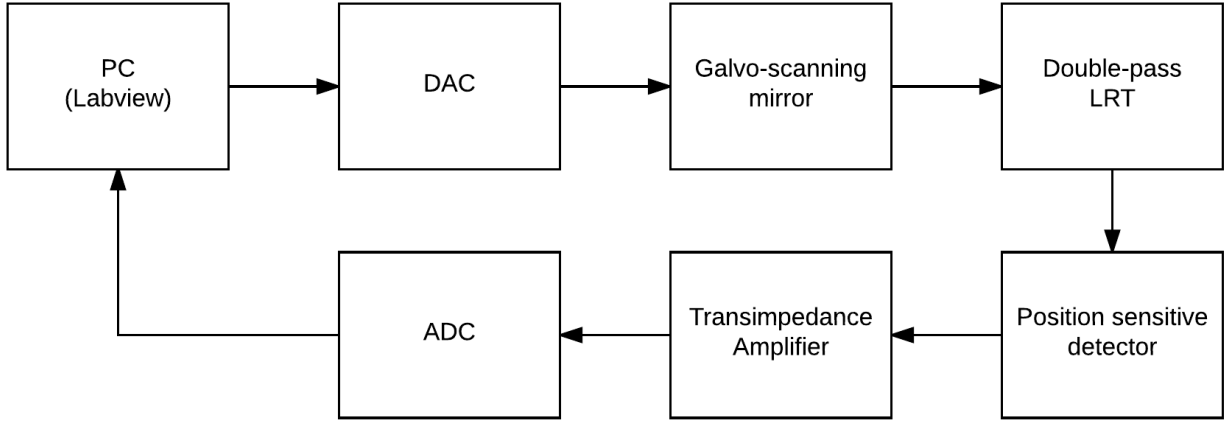


Figure 27: A block diagram of the initial double-pass LRT setup. A standard PC runs Labview. The outboard DAC and ADC are connected to the PC via a breakout board and PCI card (NI PIC-6024E). The DAC generates control signals for the galvo-scanning mirror. This steers the scanning beam through the internal optics of the LRT (not shown). The output of the LRT is sensed by a position-sensitive detector (PSD). The outputs of the PSD are amplified by a transimpedance amplifier and sampled by the ADC. The results are sent back to Labview for storage and post processing.

location. This does not fit the requirements for LRT so the other type of PSD, lateral effect, was chosen.

A lateral effect PSD, in contrast to the quad-cell PSD, has a continuous active area over which the light spot location may be determined. A duo-lateral PSD allows the determination of both x and y position by stacking two photodiode layers at a 90° angle as shown in Figure 29. Unlike a quad-cell PSD the lateral effect PSD divides the photo-generated current I_0 between the two contacts on each layer. One layer acts as a current source while the other acts as a sink so current enters one layer and exits from the other. Each layer basically acts as a current divider with the proportion of current entering/exiting each pair of contacts proportional to the light spot location on the surface and the overall intensity. The current I appearing at each contact is

$$I = I_0 \frac{\sinh(a[L - S])}{\sinh(aL)} \quad (4.1)$$

where I_0 is the induced photocurrent, a is known as the Lucovsky falloff parameter, S is the distance from the centroid of the light spot to the edge of the active area, and L is the total length of the active area. In practice, a is a function of the N-type region of the layer and approaches 0, so 4.1 becomes

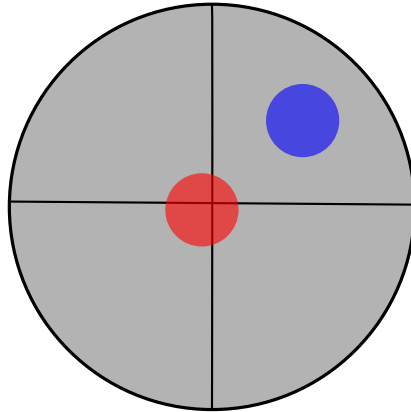


Figure 28: A quad-cell type PSD. The red beam's position is registered while the blue beam is not.

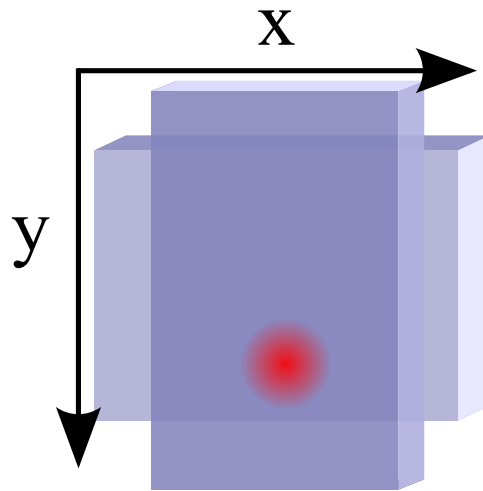


Figure 29: A lateral effect type PSD. Two diffuse photodiodes are stacked, one for the x-axis and one for the y-axis.

$$I = I_0 \left(1 - \frac{S}{L} \right) \quad (4.2)$$

which guarantees that the PSD will exhibit good linearity over the entire active area. The PSD chosen for the initial setup, sourced from Edmund Optics (part number 58-282), was duo-lateral with an active area of 4 mm^2 . The small active area gives the PSD a fast rise time of $0.025 \mu\text{sec}$. The sensitivity of the PSD is 400 mA W^{-1} which is more than sufficient for the typical optical power available at the detector (easily $50 \mu\text{W}$). The PSD x and y current outputs are converted to voltages and amplified by a transimpedance amplifier with a gain of 10 M . The voltages are then sampled by the ADC and stored in the PC. While a PSD

is well suited for non-coherence-gated LRT it was determined that it would be unsuitable for CGLRT due to its inability to differentiate interference fringes outside a very small area around $OPD = 0$. The average intensity of the interference fringes, when spread across a large portion of the PSD surface, is unchanging as the fringes are modulated. As such, the PSD is unable to differentiate between the modulating "AC" component of the fringes and the "DC" component due to stray reflections or light returned from outside the coherence gate. For this reason the PSD was eventually replaced by a high-speed CCD camera. For this initial measurements though, a PSD was utilised for all spot position measurements.

The initial LRT setup is shown in Figure 30. The measurement beam is launched by optical source OS (675 nm super luminescent diode (SLD)) and collimated by lens LC. The SLD exhibits a high degree of spatial coherence, as one would expect from a standard laser, but has a much lower temporal coherence. Lenses L1 ($f = 200$ mm) and L2 ($f = 30$ mm) form a beam reducer which decreases the beam diameter from ≈ 210 mm to ≈ 0.5 mm. Lens L3 ($f = 75$ mm) focuses the beam onto the galvo-scanning mirror GM. The beam is re-collimated by lens L4 ($f = 75$ mm) and sent to beam splitter BS. BS directs the beam to the test object TO which consists of a short focal length lens L5 ($f = 35$ mm) and object mirror OM. The reflected beam passes through BS and the pupil plane of L5 is imaged by lenses L6 and L7 (both $f = 75$ mm). The beam position is then sensed by the detector DET, in this case a PSD.

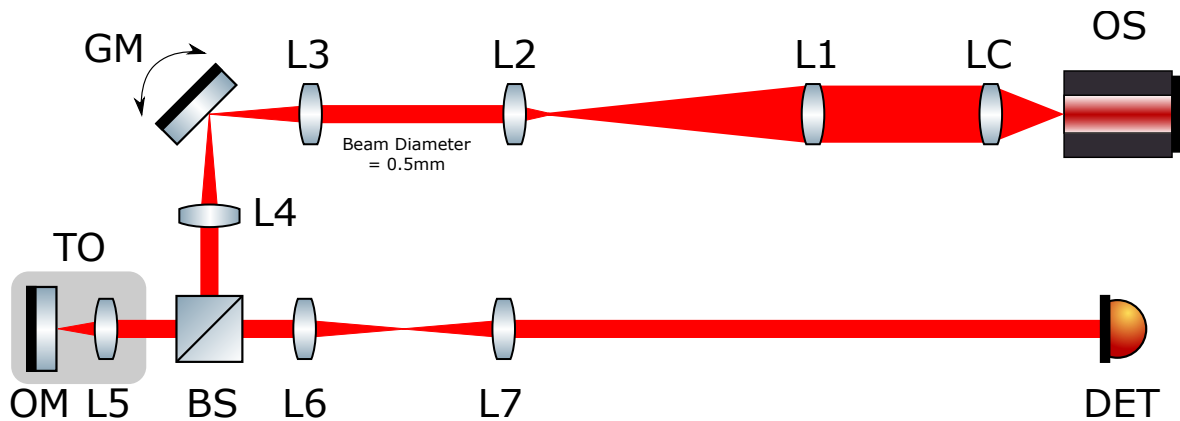


Figure 30: Initial LRT system. The measurement beam is launched by optical source OS (675 nm SLD) and collimated by lens LC. Lenses L1 ($f = 200$ mm) and L2 ($f = 30$ mm) form a beam reducer which decreases the beam diameter from ≈ 210 mm to ≈ 0.5 mm. Lens L3 ($f = 75$ mm) focuses the beam onto the galvo-scanning mirror GM. The beam is re-collimated by lens L4 ($f = 75$ mm) and sent to beam splitter BS. BS directs the beam to the test object TO which consists of a short focal length lens L5 ($f = 35$ mm) and object mirror OM. The reflected beam passes through BS and the pupil plane of L5 is imaged by lenses L6 and L7 (both $f = 75$ mm). The beam position is then sensed by the detector DET, in this case a PSD. The option exists to place a transparent aberrating optic in the test object between BS and L5.

A typical LRT scan is as follows:

- Upon system startup GM is centered and the system enters an idle state.
- A scan is initiated.
- The GM is commanded to move to location 1 (command voltages specified in Labview) and given time to move and settle.
- Labview acquires the current x and y voltages from the PSD via the ADC.
- After data acquisition is complete the GM is commanded to move to location 2 and the process repeats.
- When the last location data has been acquired the GM is re-centered and the system is idle.
- All voltage data is saved to the hard disk in a delimited text file for post-processing in MATLAB.

It is important to note that for this incarnation of LRT all data analysis and wavefront reconstruction was done in post-processing via MATLAB. This is in contrast with the FPGA based CGLRT implementation in which all data processing is performed in real-time on the embedded system. The real-time FPGA architecture is discussed in detail in Chapter 5 and the CGLRT system implementation and results are discussed in Chapter 6.

In order to test the functionality of the LRT aberrations were introduced into the test object TO via insertion of a microscope slide. The slide was placed between the beam splitter BS and lens L5 in several tilted variations, as shown in Figure 31. The following aberrations were introduced:

- small tilt in X (approximately 1°).
- large tilt in X (approximately $2^\circ - 3^\circ$).
- tilt in Y (approximately 1°).

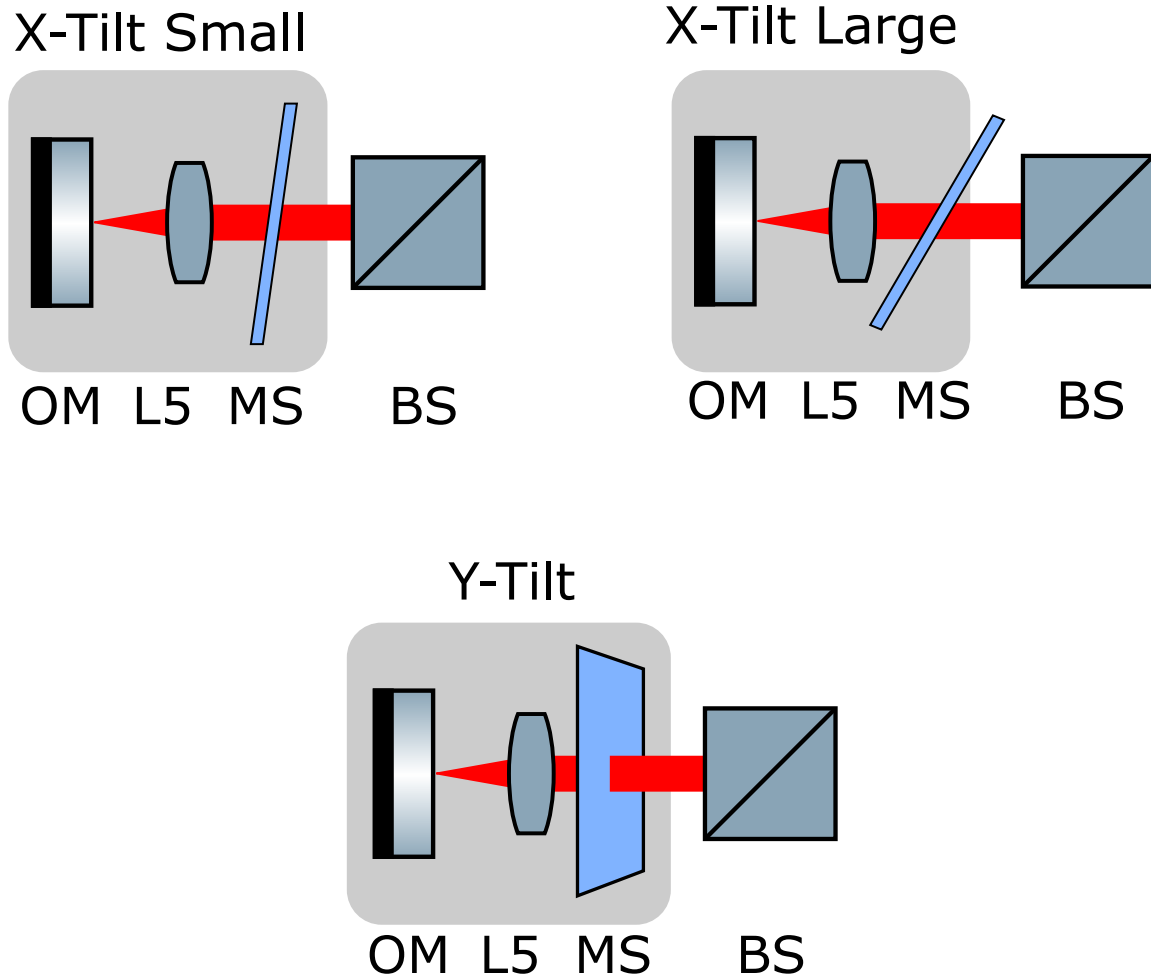


Figure 31: Close-up view of the test object TO in Figure 30 including the beam splitter BS. In each view, a microscope slide MS is inserted between BS and lens L5 to introduce a tilt aberration. Top left: Small tilt in x . Top right: Large tilt in x . Bottom centre: Tilt in y .

For each measurement a 9-by-9 spot pattern was generated via the scanning mirror. Galvo control voltages were stepped from -200mV to 200mV in 50mV steps. Each spot was allowed to settle for approximately 200ms before data was acquired. The spot position on the PSD was sampled at 1kHz and ten consecutive data points were acquired at each spot position and averaged.

First, a calibration spot pattern was generated by scanning without the slide inserted. The raw spot locations are shown in Figure 32, with ten measurements at each scan position. It is clear from the plot that the spot locations were very stable over the course of the scan. Figure 33 shows the spot locations after each of the ten measurements was averaged. This is used for the reference locations from which the Δx and Δy 's are calculated.

Figures 34, 37, and 40 show the measured spot locations (blue triangles) compared to the calibration spot locations (red circles) for each induced aberration. Figures 35, 38, and 41 show the reconstructed wavefront for each scan utilising the first 14 Zernike modes (piston removed). Refer to section 5.5.1 for a details on the calculation of the Zernike modes. Figures 36, 39, and 42 show the value of the calculated Zernike coefficients for each scan.

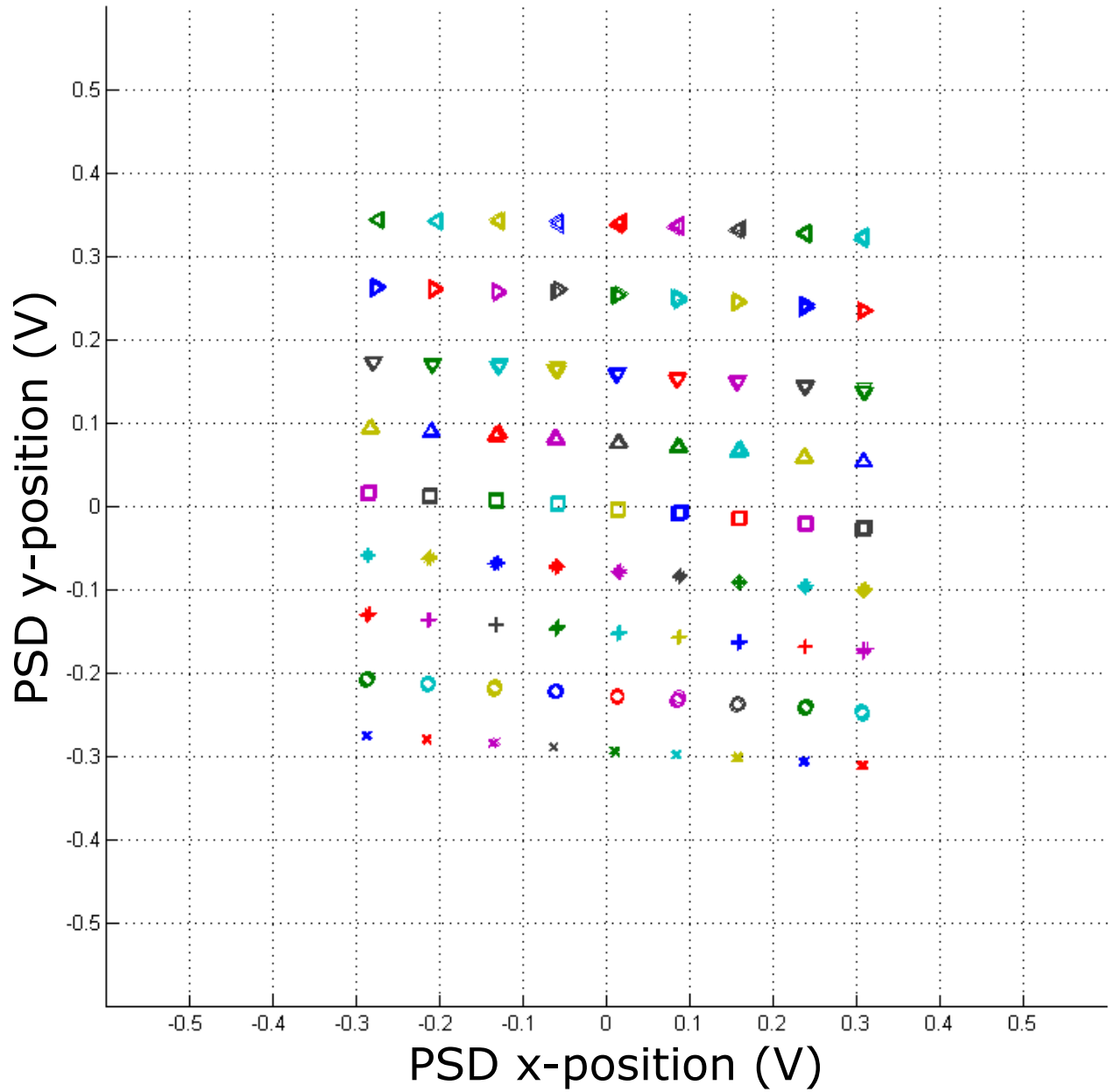


Figure 32: LRT raw measured spots. Ten measurements are taken at each location. Markers at each location vary in shape and colour to clearly indicate if any measurement is erroneous. The consistency of the markers shows that the measurements did not vary from location to location.

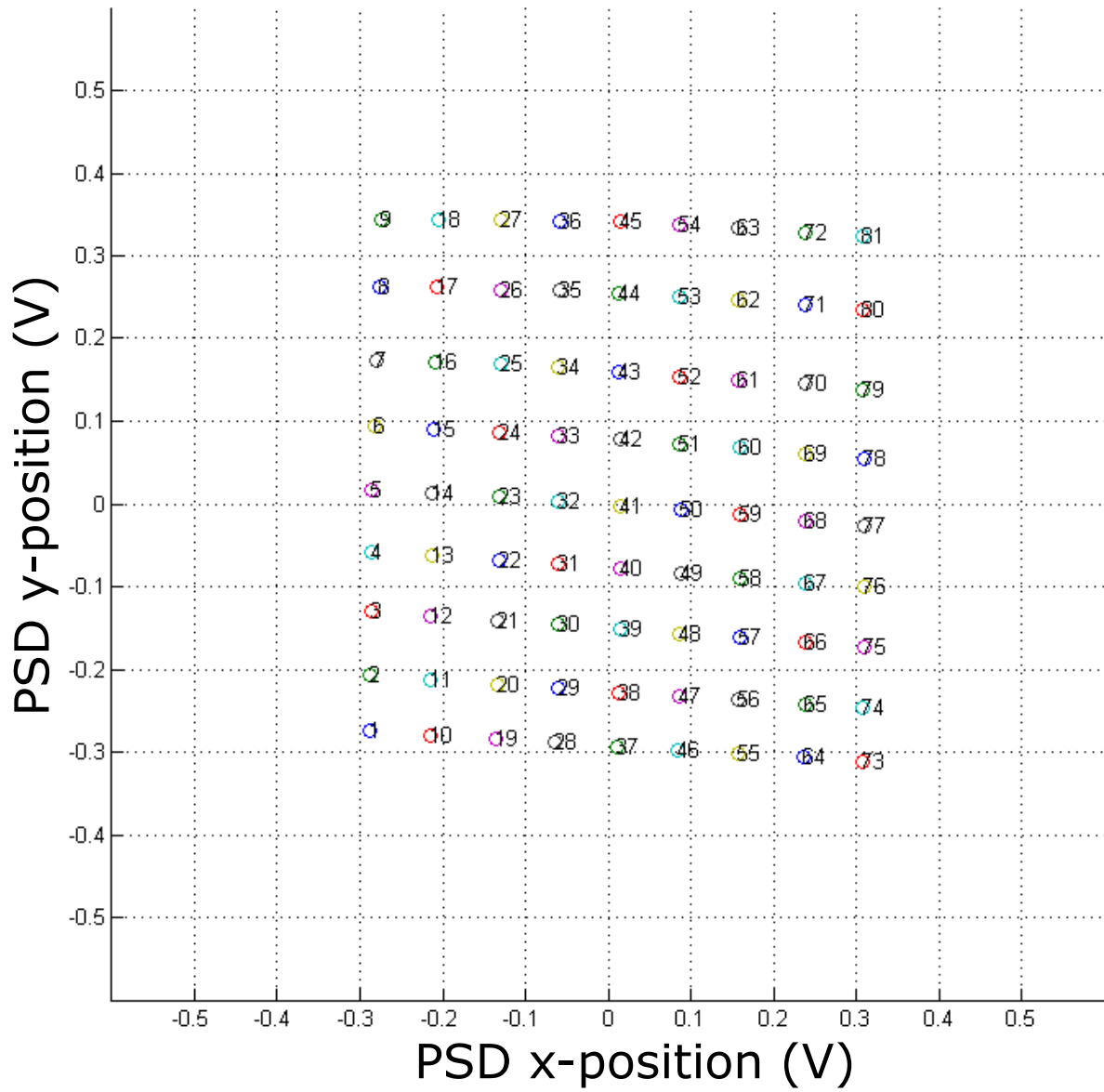


Figure 33: LRT averaged measured spots.

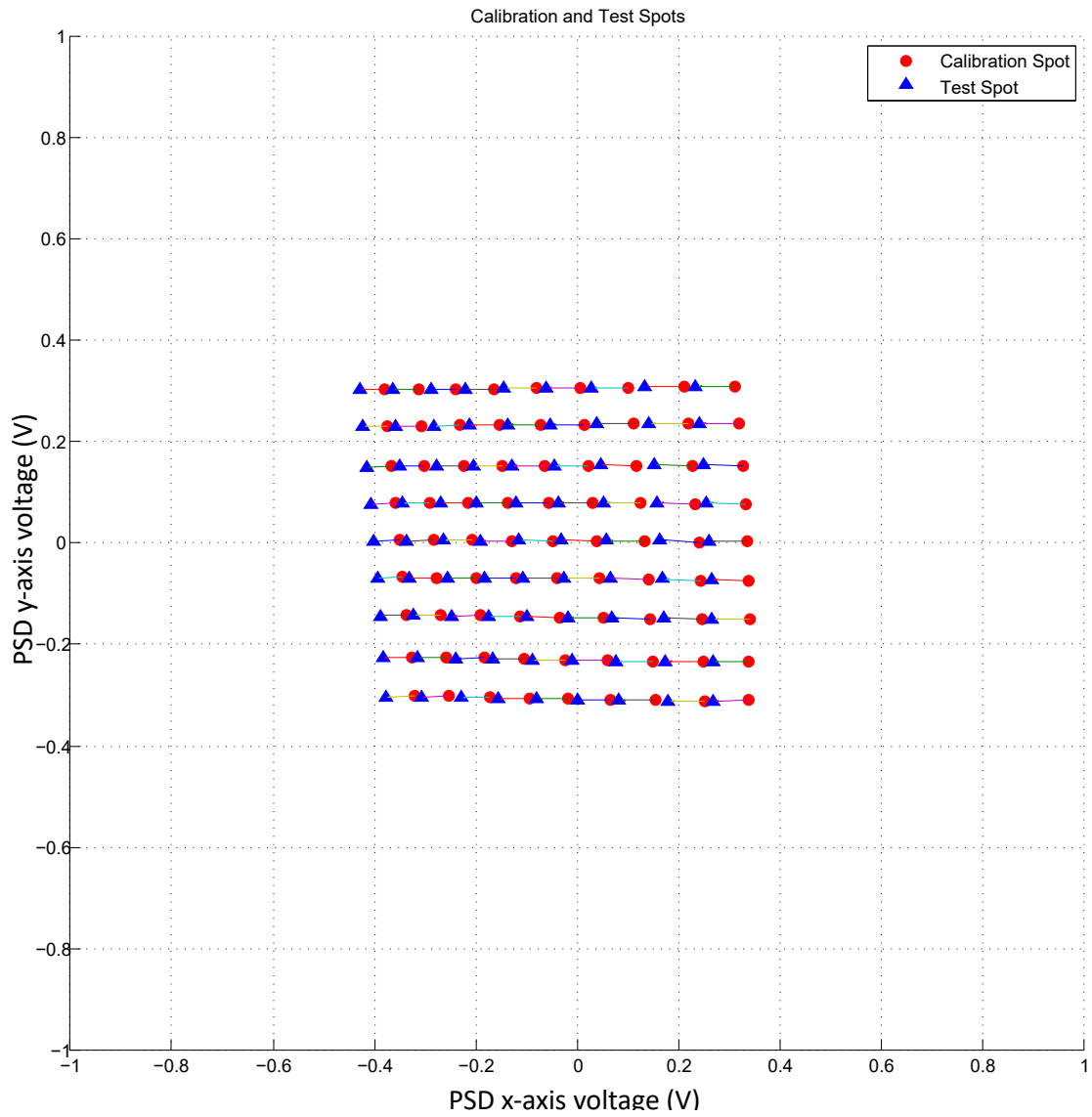
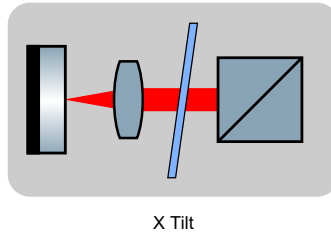


Figure 34: Spot pattern for small tilt in X.

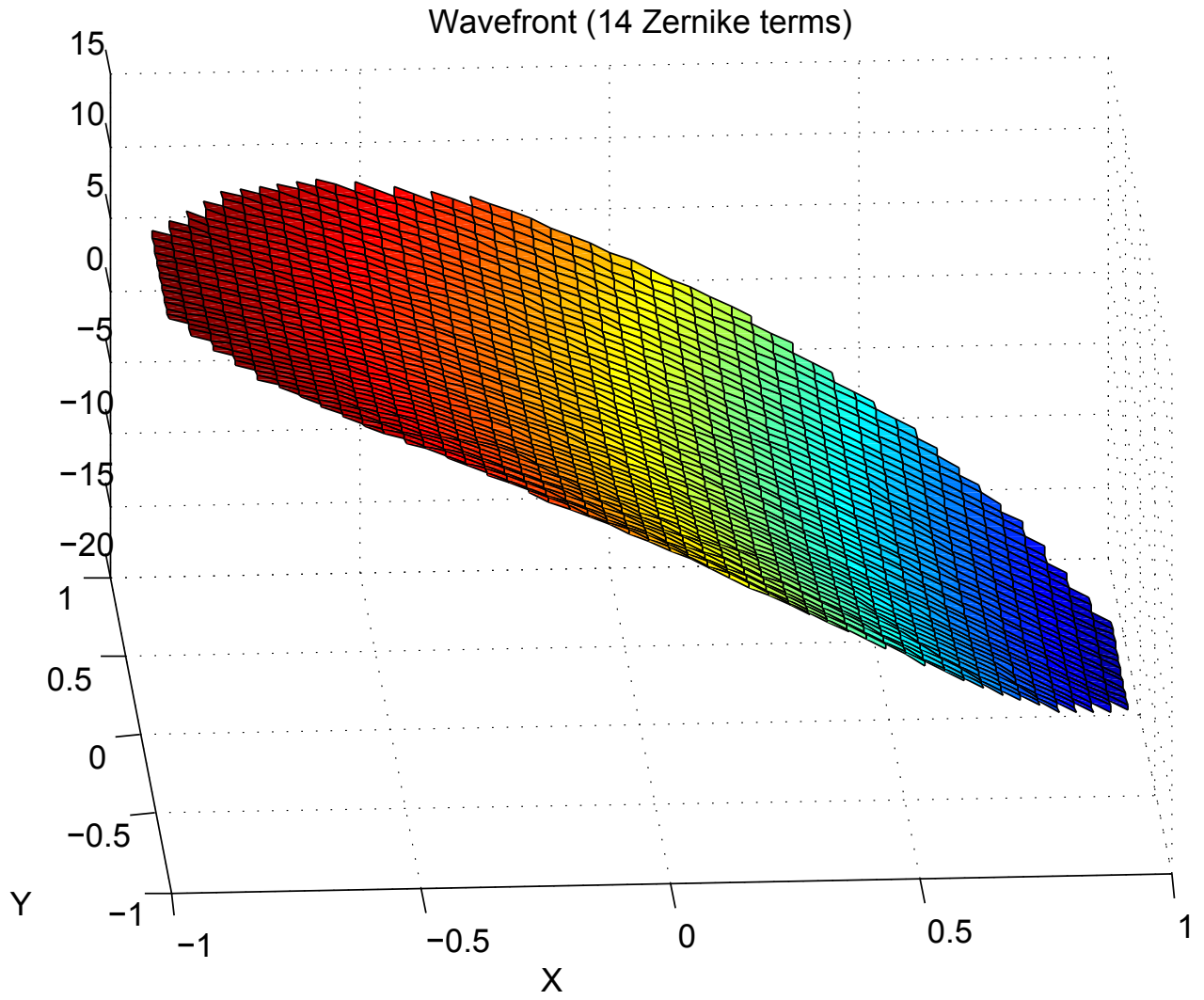


Figure 35: Reconstructed wavefront from 14 Zernike terms for a small X tilt.

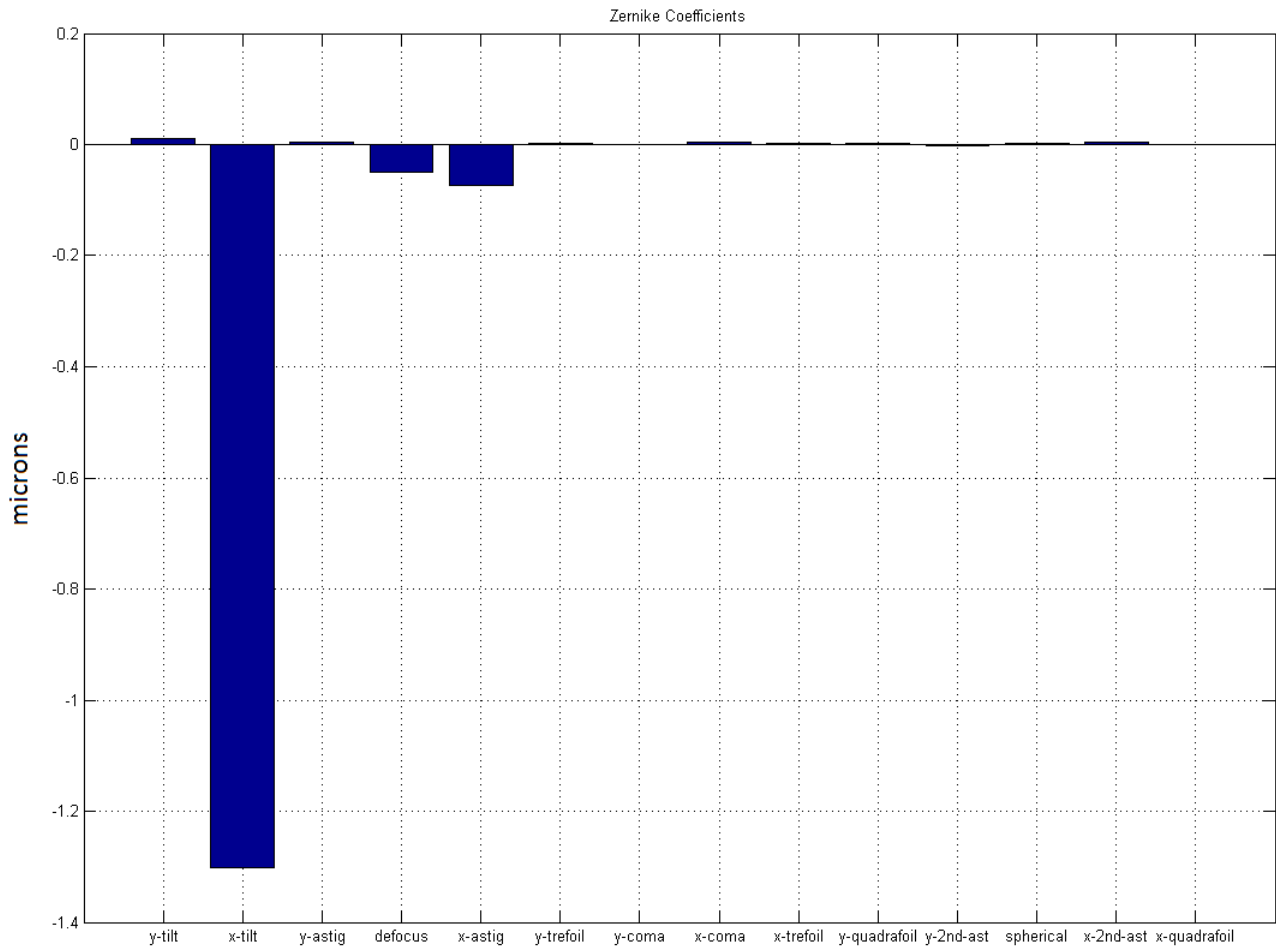


Figure 36: Zernike coefficients for a small X tilt.

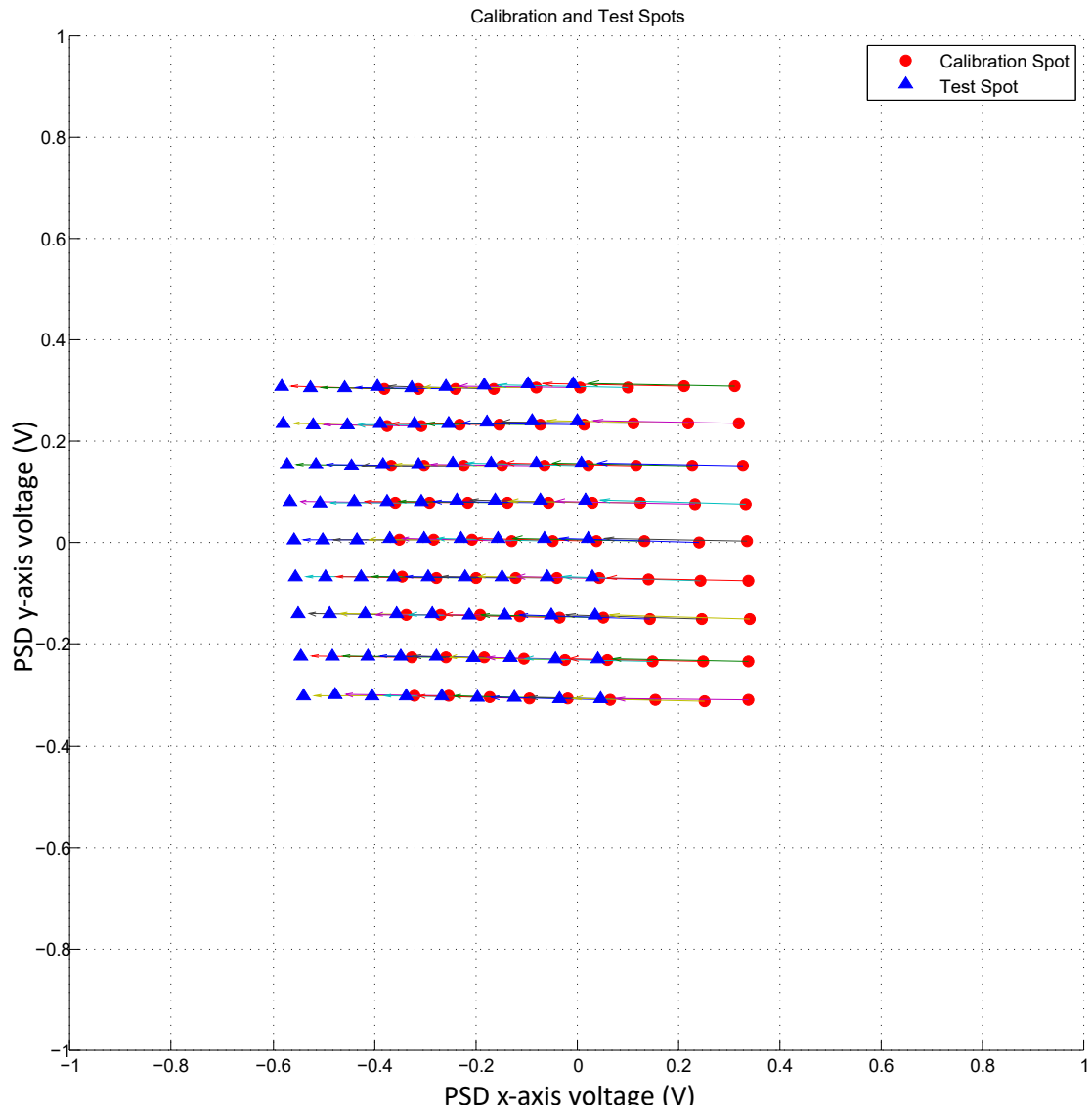
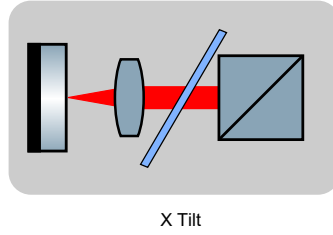


Figure 37: Spot pattern for large tilt in X.

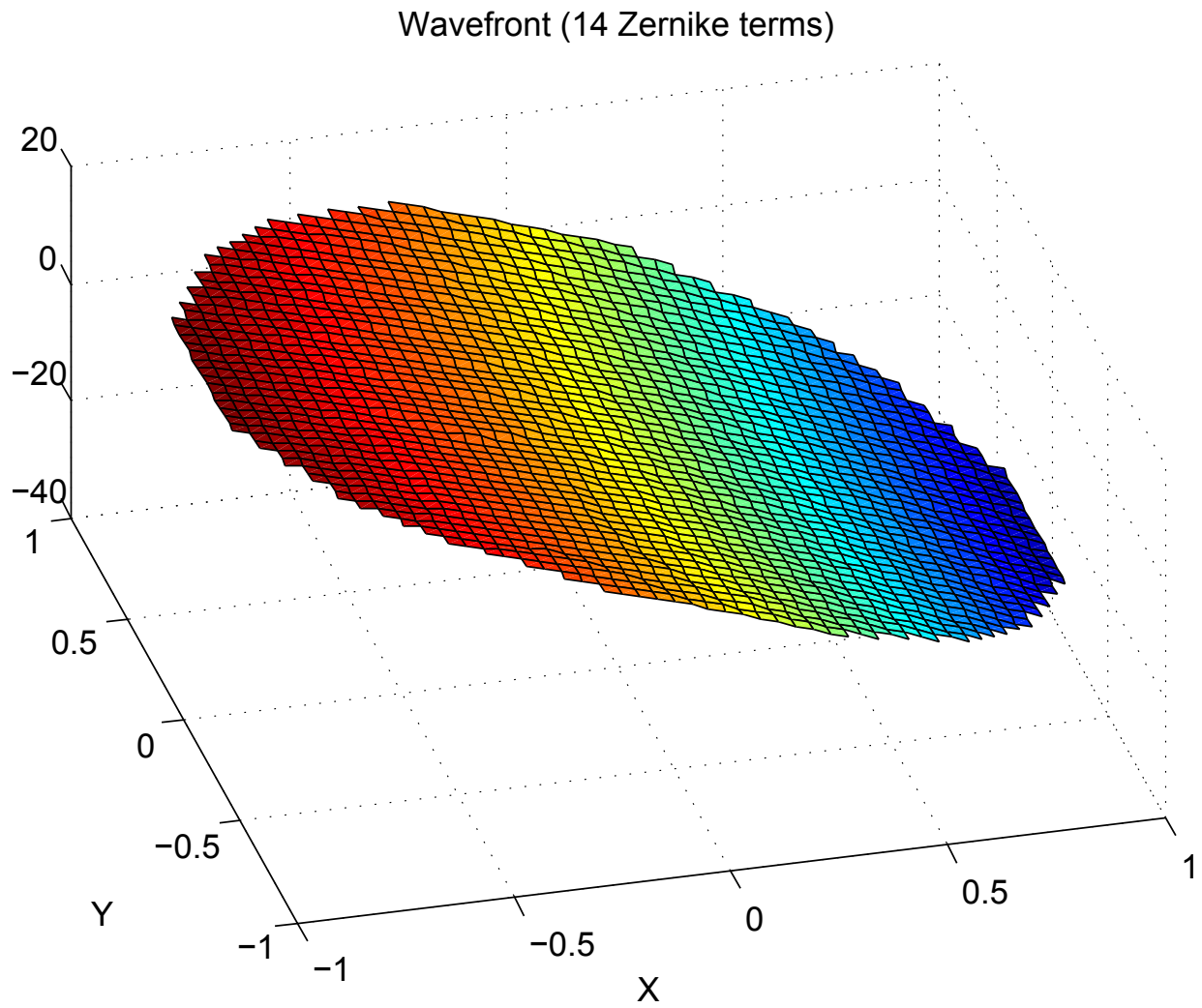


Figure 38: Reconstructed wavefront from 14 Zernike terms for a large X tilt.

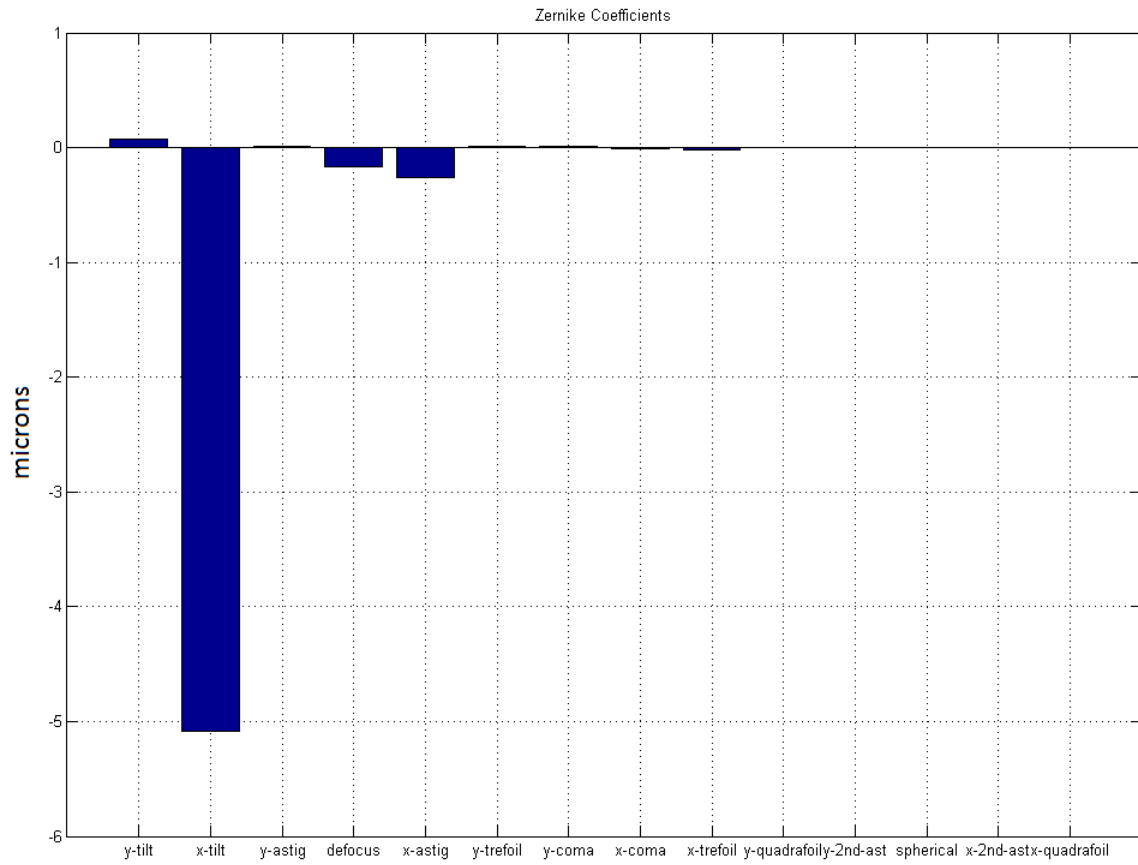
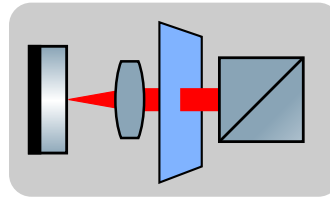


Figure 39: Zernike coefficients for a large X tilt.



Y Tilt

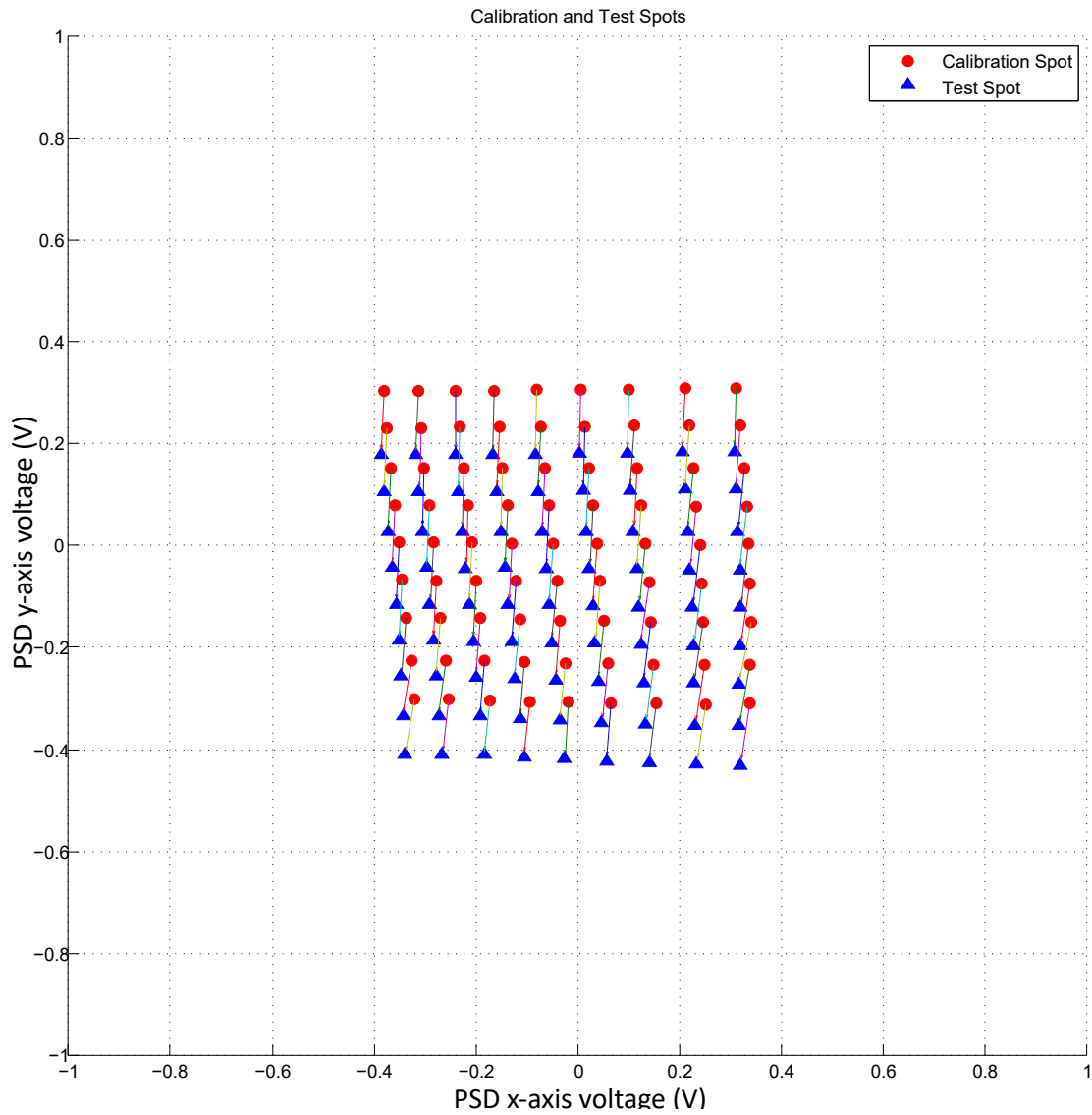


Figure 40: Spot pattern for tilt in Y.

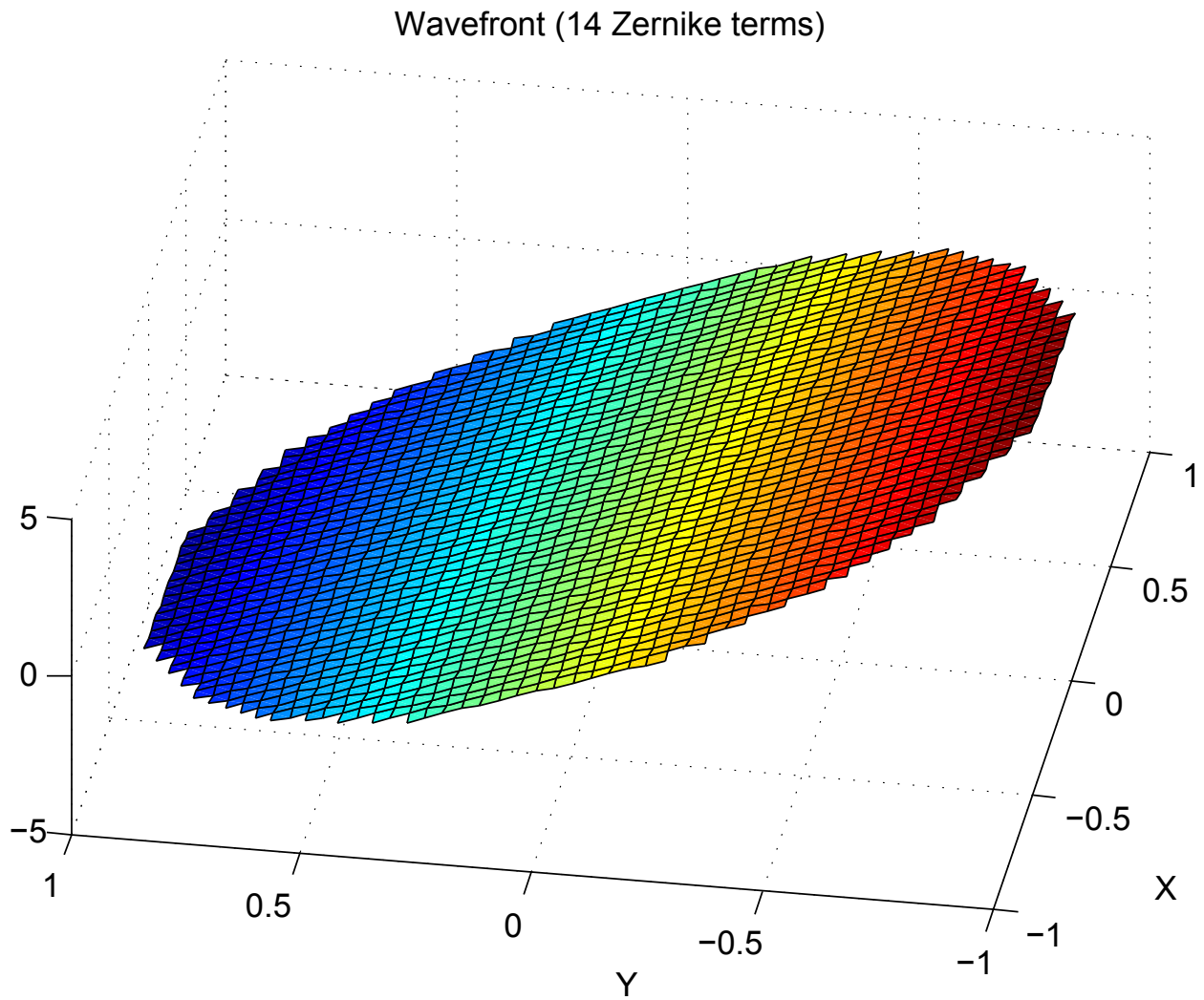


Figure 41: Reconstructed wavefront from 14 Zernike terms for a Y tilt.

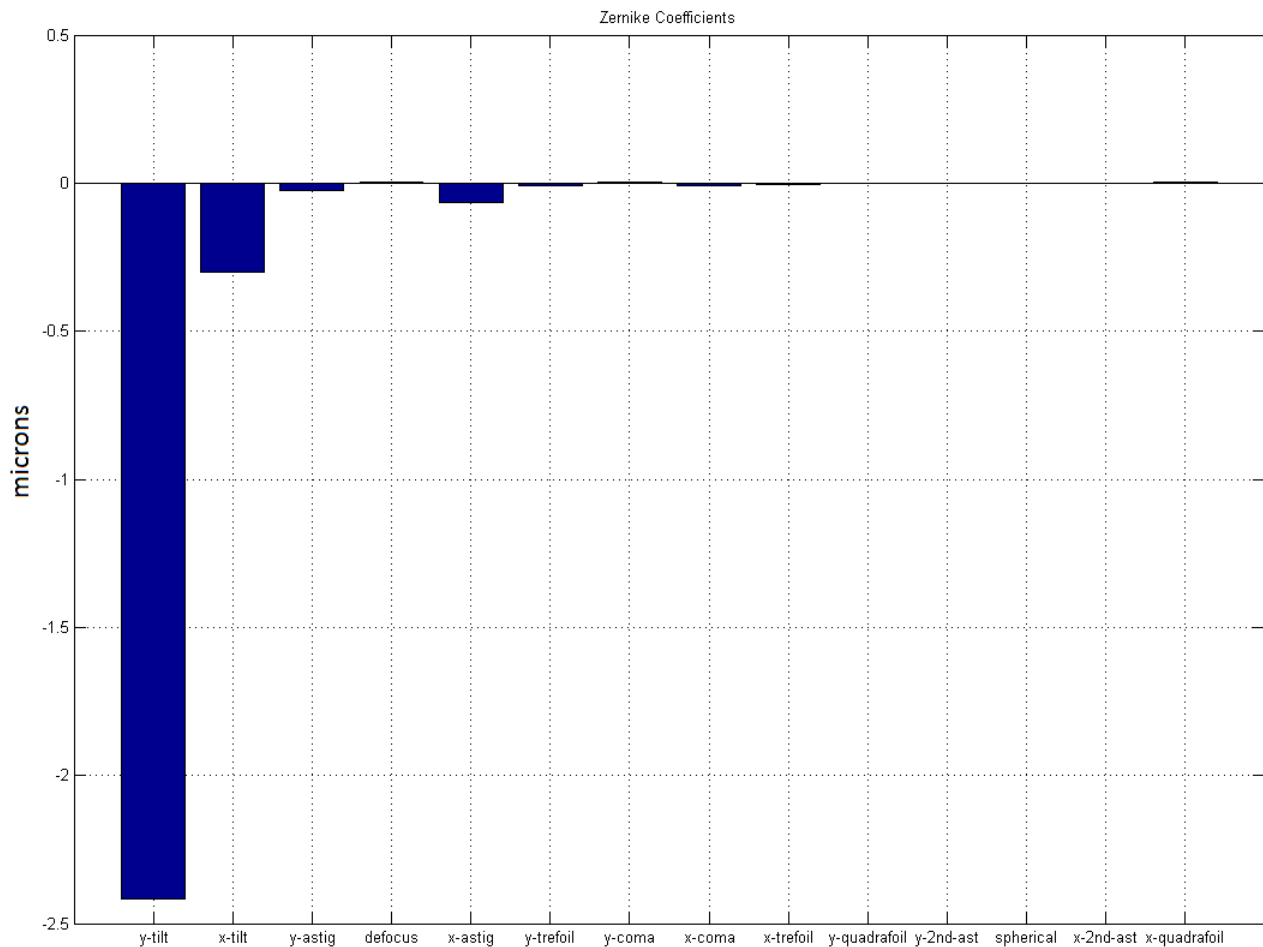


Figure 42: Zernike coefficients for a Y tilt.

4.3 Coherence-Gated LRT Experimental Setup

In its basic form LRT exhibits a large depth range making it ineffective for depth-resolved wavefront sensing. This also makes LRT sensitive to stray reflections such as those common in microscopy, which can make the determination of the spot centroid difficult if not impossible. By reducing the depth range of the sensed wavefront the use of LRT can be expanded to other imaging modalities that utilize adaptive optics such as optical coherence tomography. By combining LRT with LCI its effective depth range may be reduced to within the coherence-length of the interferometer. This section describes the development of the CGLRT wavefront sensor, based on a Mach-Zehnder interferometer, and presents the results of interrogating a multilayer object.

The experimental setup of the CGLRT developed during this research is illustrated in Fig. 43. A superluminescent diode (SLD) with a central wavelength of 675.5 nm and a FWHM bandwidth of 8.8 nm is utilized as an optical source. These characteristics correspond to an axial coherence gate of 23 μm in air (via Equation 3.1). A beam 5 mm in diameter is collimated by lens CL and sent to beam splitter BS1 forming the reference (blue) and object (red) arms of the interferometer. The reference beam is left uncompressed to illuminate the entire area of the camera sensor. This is needed, to ensure interference as the location of the scan beam moves on the detector surface due to aberrations in the test object. The reference beam is sent via mirror M1 to the piezo-electric transducer (PZT) mounted reference mirror RM. The PZT is used to introduce phase modulation when driven by a DAC (not shown). The RM and PZT are both mounted on a linear translation stage (LTS) that provides axial adjustment of the reference arm path length.

The object beam is first reduced to a diameter of approximately 700 μm by lenses L1 ($f = 200$ mm) and L2 ($f = 30$ mm). This decreases the beam diameter making it suitable for scanning. The beam is focused upon the galvo-scanning mirror GM by L3 and re-collimated by L4 (both $f = 75$ mm). The beam is then sent via beam splitter BS2 to multi-layer object MLO. The MLO consists of lens L5 ($f = 35$ mm), two microscope slides and object mirror OM. The reflected beam passes once again through L5 and BS2. Lenses L6 and L7 (both $f = 75$ mm) form a real image of the spot which is then focused onto the CMOS camera (Pulnix AccuPiXEL 1402CL, 1024×1024 pixels) by lens L8 ($f = 75$ mm). The uncompressed reference beam is recombined with the object beam at BS3 and incident upon the CMOS camera.

Interference will take place between the scanning pencil beam (object arm) and reference arm when the optical path lengths match. A four-step phase-shifting interferometry (PSI) algorithm [60] is used to recover an image of the scanning spot on the CMOS camera. For each scan location p , four images I_i where $i = 0$ to 3 are obtained. Each image corresponds to a PZT induced phase shift of $(\pi/2) i$ and the image intensity is given by

$$I(x, y) = \sqrt{(I_2 - I_0)^2 + (I_3 - I_1)^2} \quad (4.3)$$

The GM deflects the scanning beam $1^\circ/\text{V}$ in both x and y . When paired with L5 this produces a scan area of 2.6 mm by 2.6 mm. For the p th spot acquired by the LRT, the DAC commands the GM to position (x_p, y_p) in the range of -1 V to 1 V (in 0.25 V increments). This equates to $p = 81$ scan points arranged in a 9×9 grid. Though this grid size was

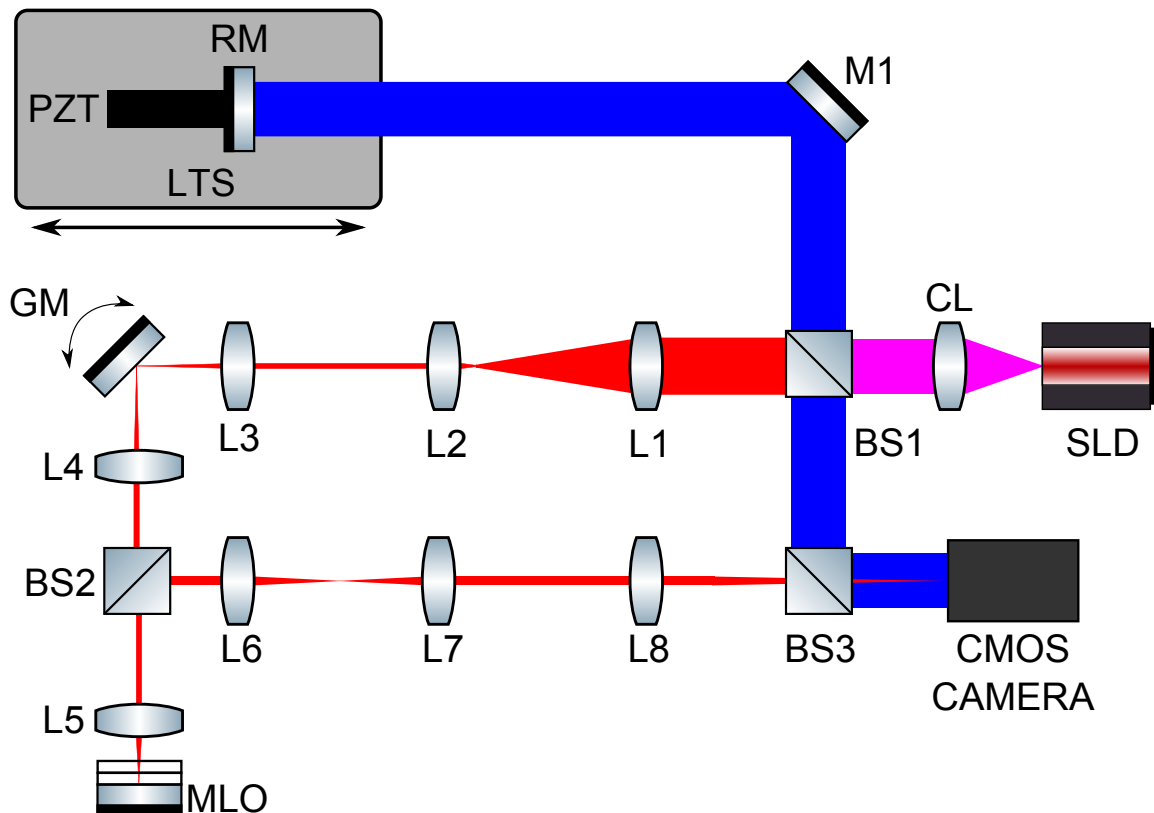


Figure 43: CGLRT experimental setup. SLD, superluminescent diode; CL, collimating lens; BS1, BS2, BS3, plate beam splitters; L1-L8, achromatic doublets; GM, galvo-scanning mirror; RM, reference mirror; PZT, piezoelectric actuator; MLO, multi-layer object composed of object mirror OM and 2 glass microscope slides; M1, mirror.

chosen for this demonstration, one of the advantages of LRT is that any scan pattern and/or number of points may be selected by the user simply by altering the DAC outputs.

Fig. 44 shows a zoomed in view of the MLO. Two microscope slides, approximately 1 mm thick, are inserted before the object mirror OM. To introduce aberrations the slides are tilted in the x-axis (parallel to the optical bench) in opposite directions ($\theta \approx 0.013$ rad). A best effort was made to minimize the y-axis tilt (normal to the optical bench) by manual adjustment. The four faces of the slides and OM form five surfaces, labelled P1 - P5, defined by the refractive index change from air to glass, which form the wavefronts of interest. The reflections from P1-P4 are much weaker than that strong reflection returned from the OM.

To begin with, the scanning beam is focused between P4 and P5 as this roughly equalizes the size of the returned beams. Next, the RM is adjusted axially using the LTS to acquire fringes for the OM by achieving $OPD = 0$. A spot pattern is then acquired by actuating the GM to each scan position and obtaining four images at each location in the four-step

PSI manner described above. For $p = 81$ scan locations this equates to 324 total images per surface. The LTS is then adjusted again to match path length with the P1 surface, a scan performed, and spot pattern obtained. This sequence is repeated for all surfaces to P5. In all measurements the MLO is kept in its fixed position and only the RM is adjusted to obtain $OPD = 0$. Lastly, a calibration scan is taken with the reference beam blocked and the microscope slides removed. The focal point of the scanning beam was not adjusted for the calibration scan. The calibration scan allows aberrations due to the system to be removed as this set of images was considered the reference for all the wavefront slope calculations. The calibration scan is only required once during a single run, as nothing in the setup is changed other than the OPD adjustment in the reference arm. Therefore, any aberrations present in the system (and not due to the sample) should be static over a single run.

The purpose of this initial implementation was solely to demonstrate the ability of CGLRT to perform depth-resolved wavefront sensing. As the embedded system portion of this research was yet to be built, timing of the current setup was not of great concern. That being the case, minor effort was to optimize the acquisition speed of the system. To capture the entire set of 324 images took the system approximately 54 seconds. Approximately 16 seconds of this time is due to the settle time allotted for the galvo-scanning mirror. The remaining 38 seconds are consumed by Labview acquiring and saving the images, which yields an effective frame rate of 8.6 fps. This does not include the time it takes for image post-processing and wavefront reconstruction in MATLAB, which is done after the scan is complete. This provides motivation for the development of the FPGA based embedded systems platform which is discussed in the next chapter.

Figures 45, 46, 47 and 48 shows a typical set of four images obtained at one scan location and the coherence-gated image 49. In images 45 through 48 the reference beam is visible as the large diameter illumination. The bright central reflection is due to the OM while the dimmer right-hand and left-hand side reflections are due to P1-P2 and P3-P4 respectively. The reflections from P1 and P2 are very closely overlaid and are indistinguishable in the image. The same is true of the P3 and P4 reflections. Image 49 is the coherence-gated spot after applying the four-step algorithm and shows the removal of all unwanted reflections including the bright central spot from the OM.

The spot patterns for each layer were then used to calculate a corresponding wavefront. Aberrations were calculated in terms of Zernike polynomials [45], the first 12 of which are shown in Fig. 50. There is a small amount of y-tilt, likely due to a small misalignment of the MLO. Of more interest are the x-tilt values, clearly showing the positive tilt from surfaces P1 and P2, and the negative tilt from P3 and P4. As expected P5 (the OM) has

negligible x-tilt. As the focus was fixed during the entire experiment the defocus term is greatest at P1 and decreases as P5 is approached. As noted in [35], a large amount of astigmatism is introduced into the system by the GM due to the 10 mm spacing between the x-axis and the y-axis scanning mirrors. This is mostly removed by the calibration but small residuals remain. These could be reduced to negligible levels in the future by the addition of a cylindrical lens after the GM. These results confirm that CGLRT is able to perform depth resolved wavefront sensing by measuring aberrations returned from multi-layer objects.

4.4 Conclusion

The Labview based LRT and CGLRT systems demonstrate the ability of the sensor to perform depth-resolved wavefront measurements, but the system does not operate in real-time. In order to implement real-time functionality the control, image acquisition, image processing and wavefront reconstruction are all performed on a high-speed FPGA. This allows the system to be realised in the form of a self-contained instrument with all computational processes performed in an embedded systems environment. Chapter 5 discusses the details of the FPGA architecture and its implementation while Chapter 6 discusses the results of the final FPGA based CGLRT instrument.

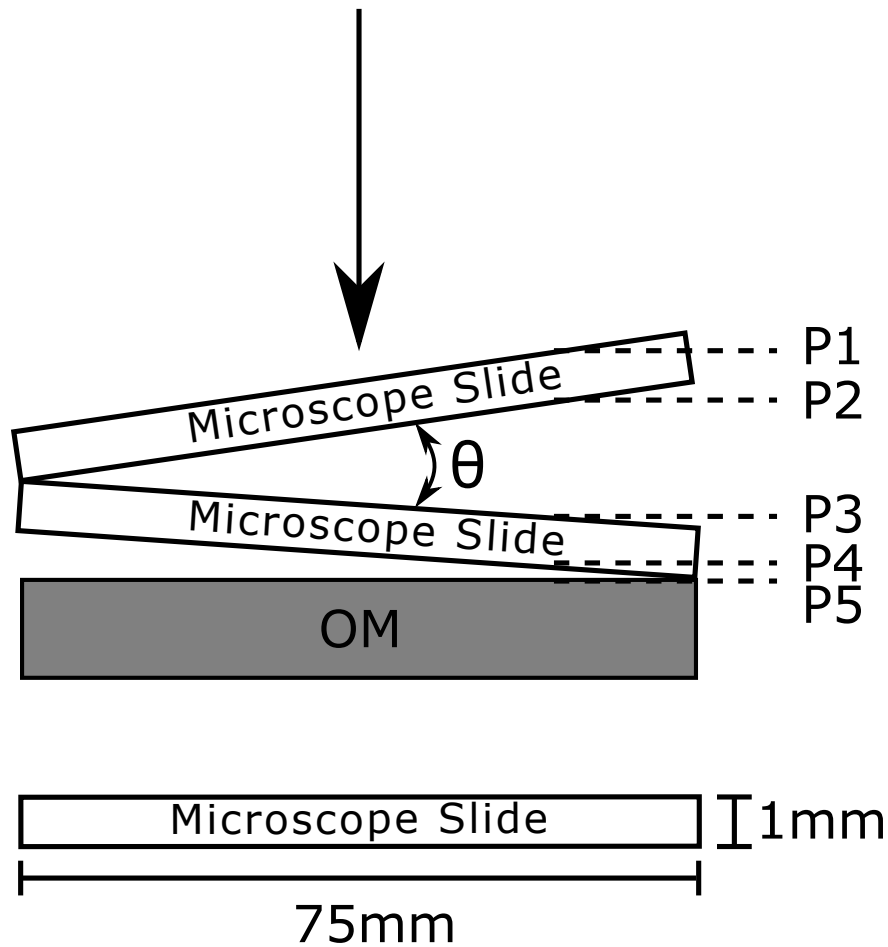


Figure 44: Microscope slides and mirror used to form the multi-layer test object. Surfaces P1-P4 are defined by the air-glass interface of each microscope slide. Surface P5 is defined by the object mirror OM. $\theta \approx 0.013$ rad. Maximum distance between the inner surfaces of the slides (P2 and P3) is approximately 1 mm. The chief ray is shown interrogating the test object from the top. Area of interrogation is approximately 2.6 mm by 2.6 mm centered on the slides. The axial coherence-width of the interferometer is 23 μm . Insert illustrates the size of the microscope slides which is approximately 75 mm by 25 mm by 1 mm.

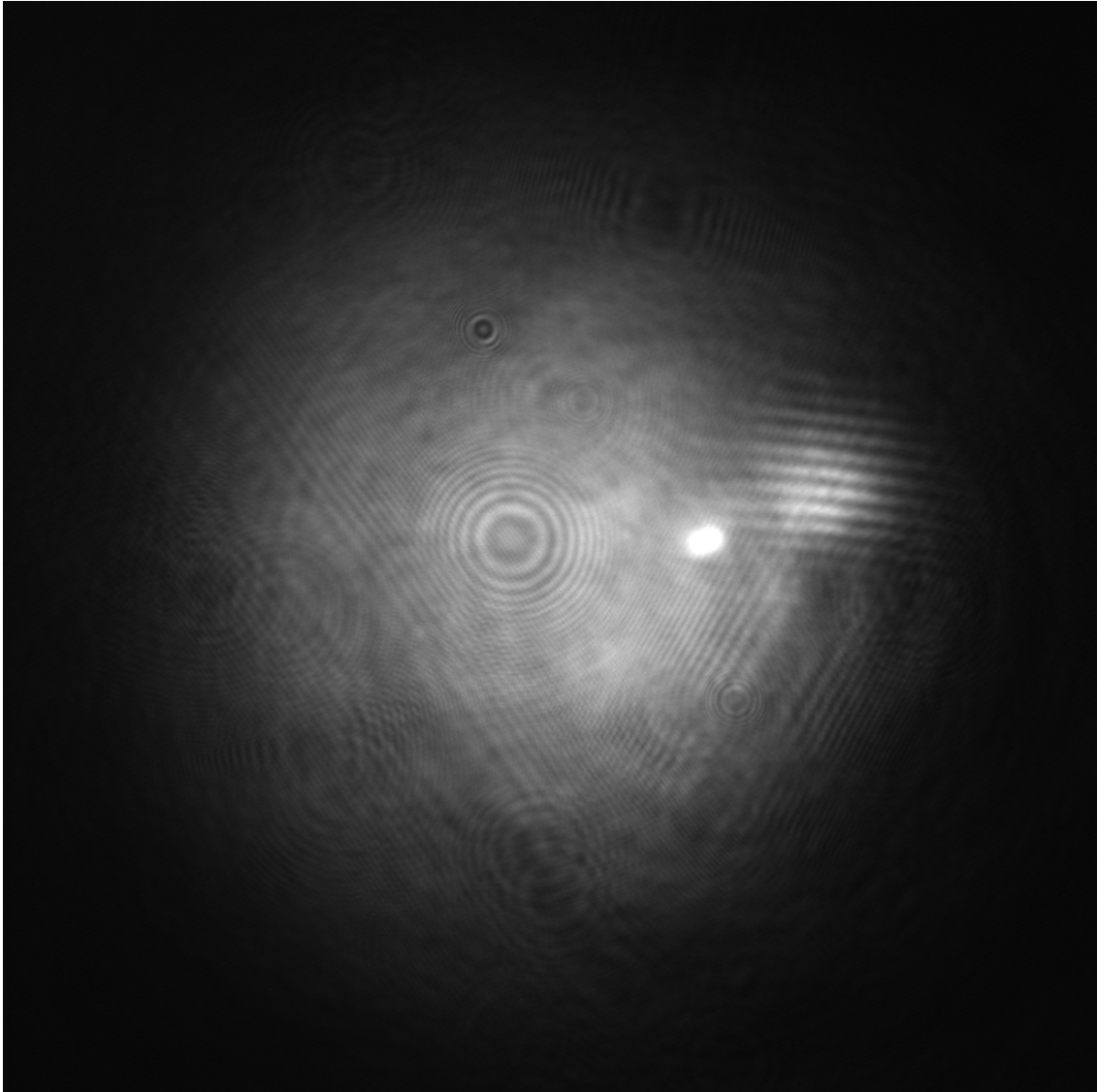


Figure 45: Raw scanning beam image with 0° phase shift. Reference beam, strong central reflection, and weaker side reflection from slides. Interference fringes clearly visible on side reflection from P1.

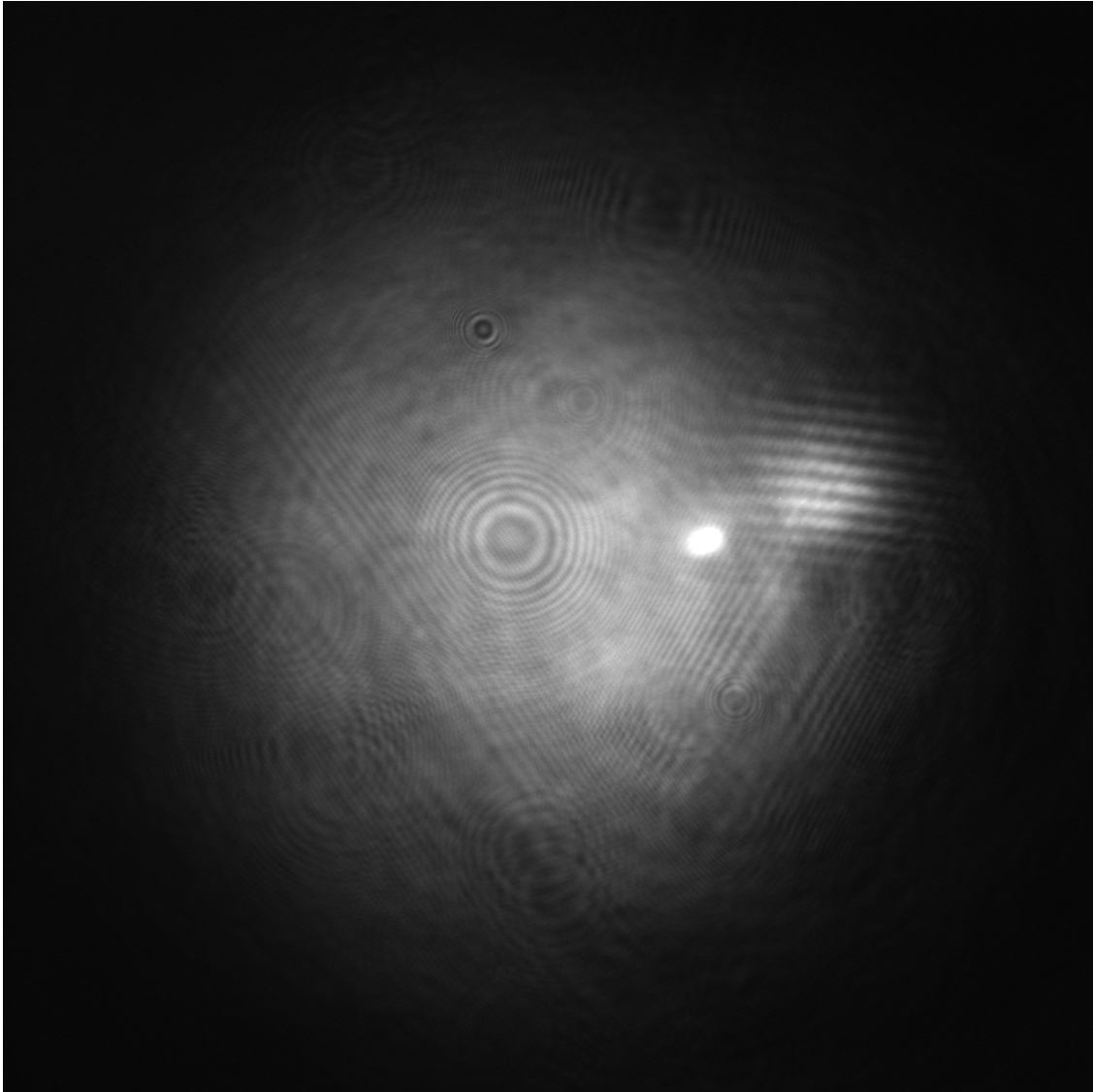


Figure 46: Raw scanning beam image with 90° phase shift. Reference beam, strong central reflection, and weaker side reflection from slides. Interference fringes clearly visible on side reflection from P1.

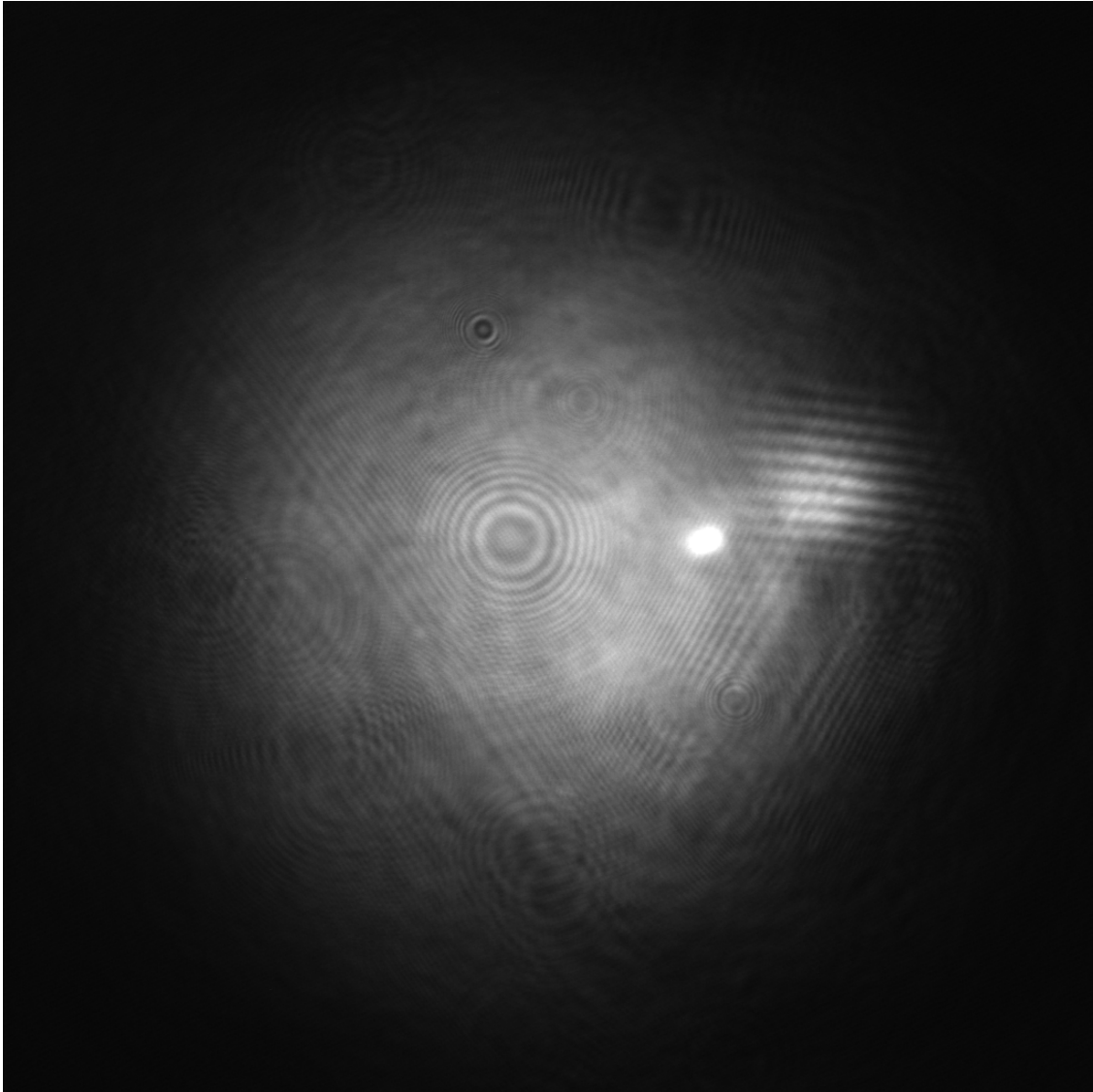


Figure 47: Raw scanning beam image with 180° phase shift. Reference beam, strong central reflection, and weaker side reflection from slides. Interference fringes clearly visible on side reflection from P1.

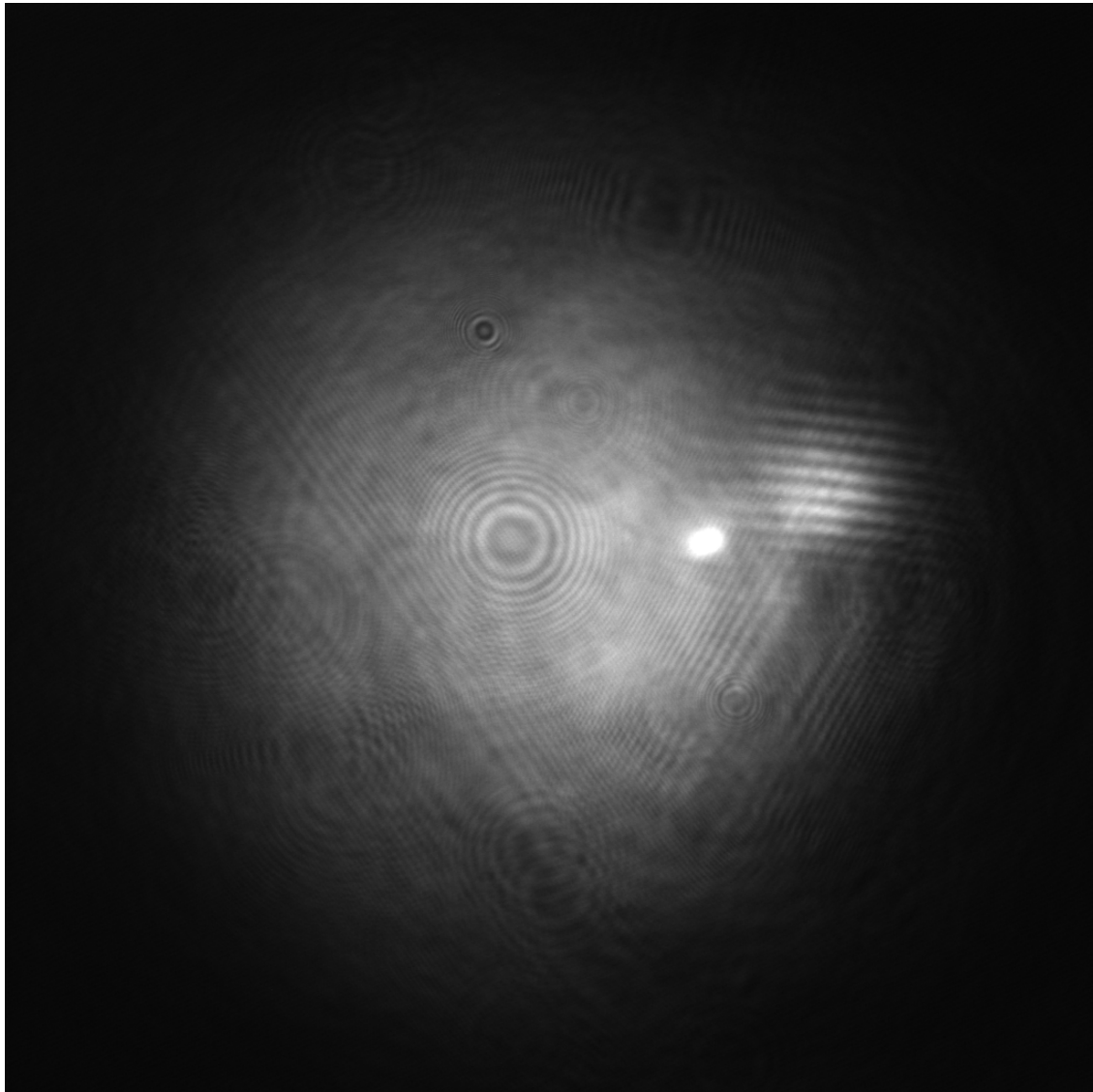


Figure 48: Raw scanning beam image with 270° phase shift. Reference beam, strong central reflection, and weaker side reflection from slides. Interference fringes clearly visible on side reflection from P1.

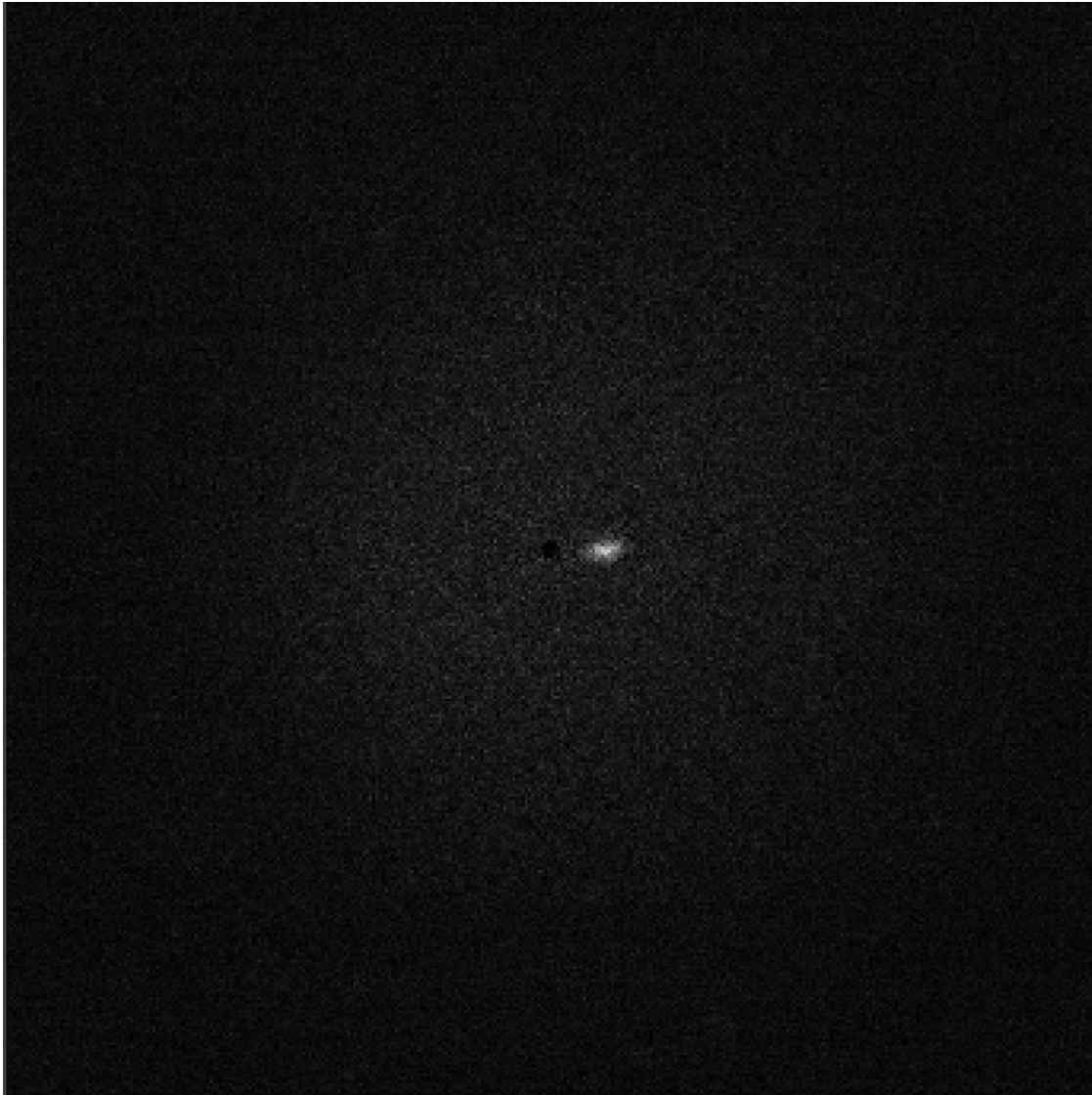


Figure 49: Coherence-gated image from surface P1. Light from sources other than P1 has been suppressed by the process of coherence-gating.

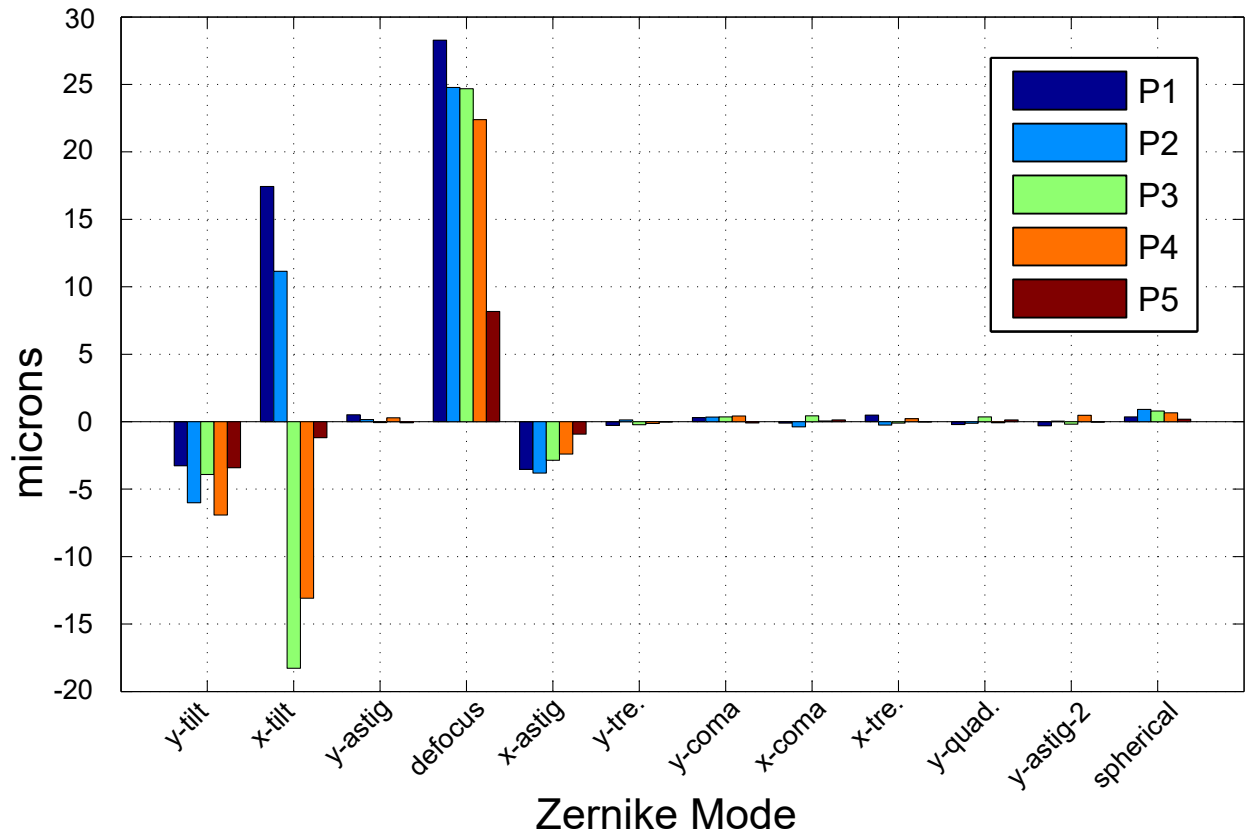


Figure 50: The first 14 Zernike coefficients for surfaces P1-P5.

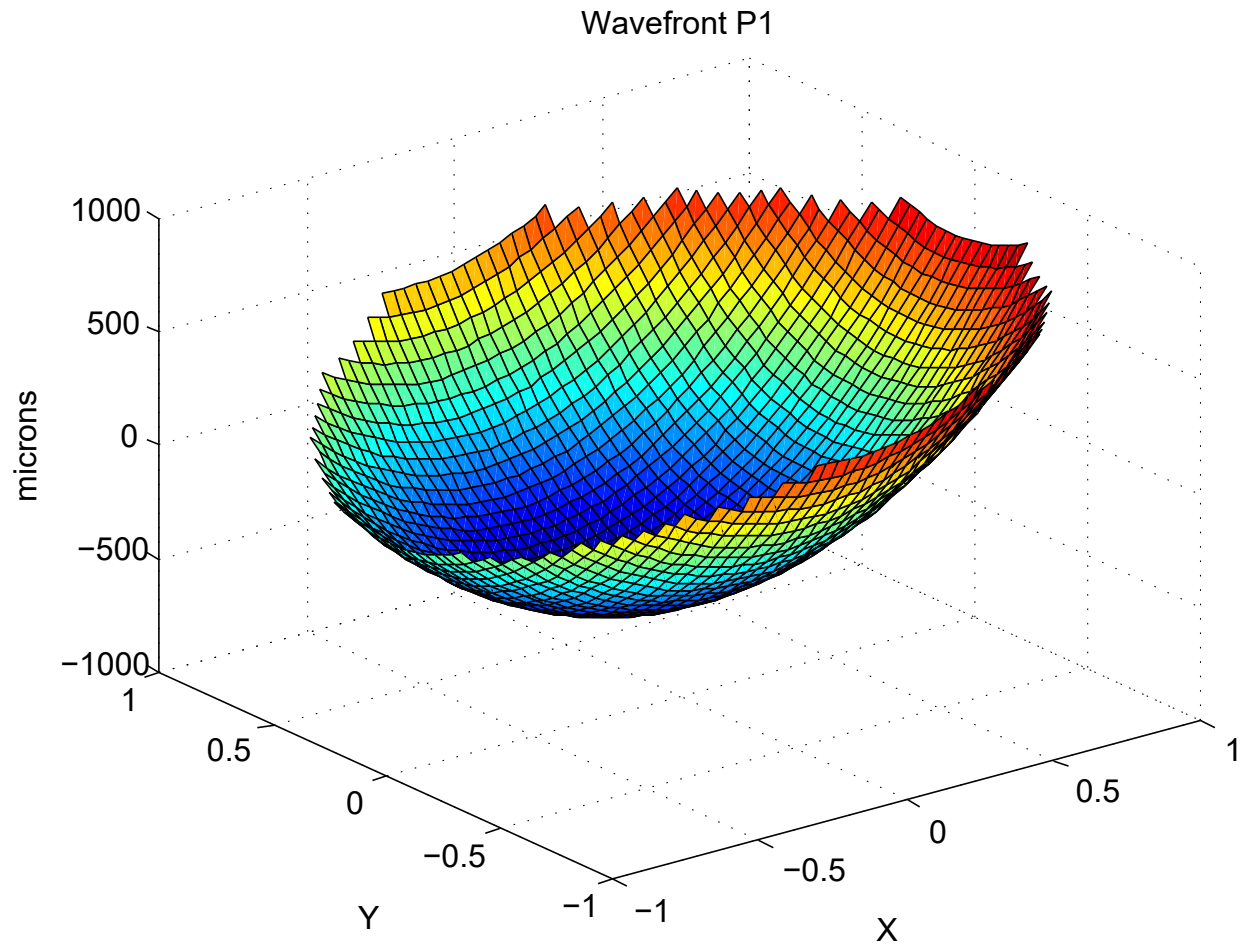


Figure 51: The wavefront measured from surface P1.

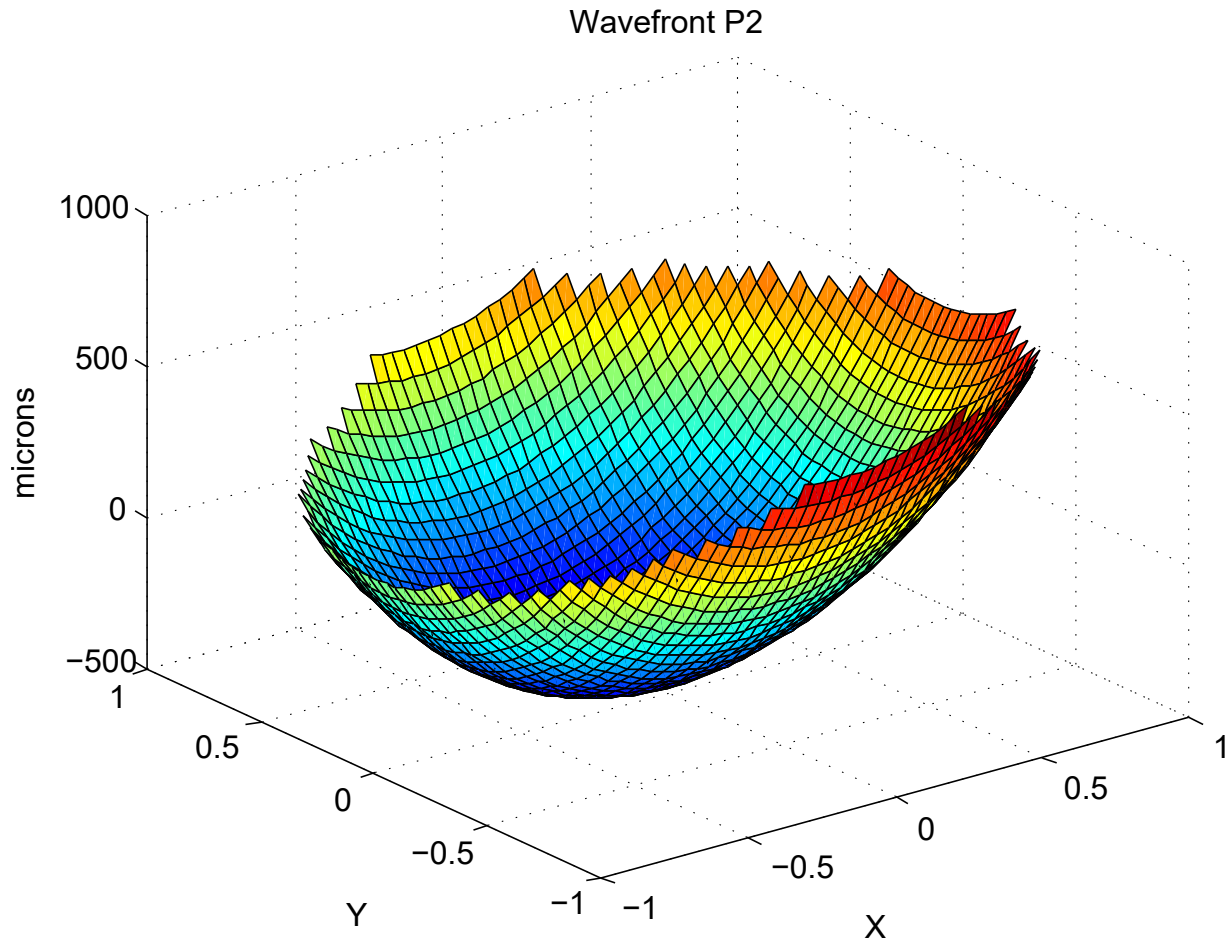


Figure 52: The wavefront measured from surface P2.

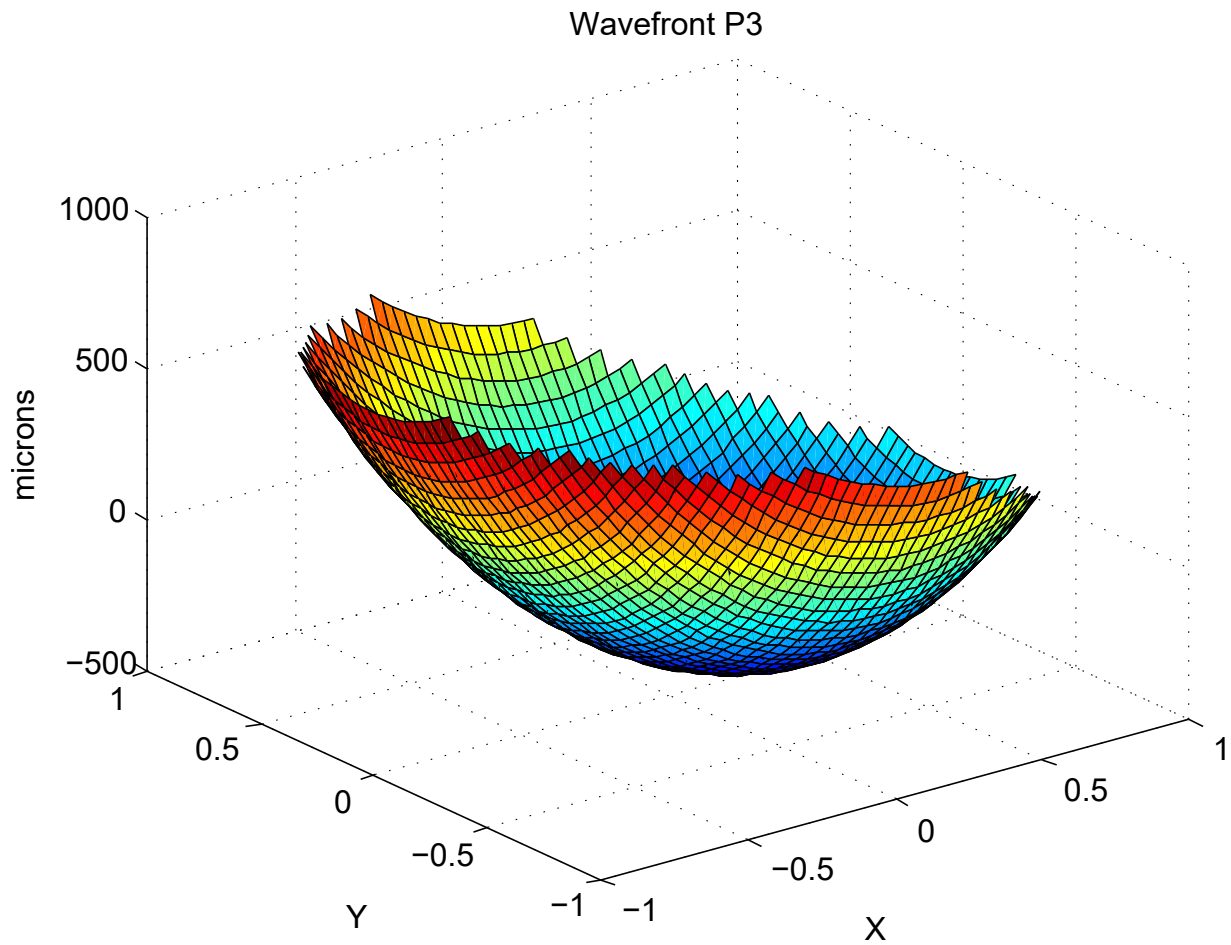


Figure 53: The wavefront measured from surface P3.

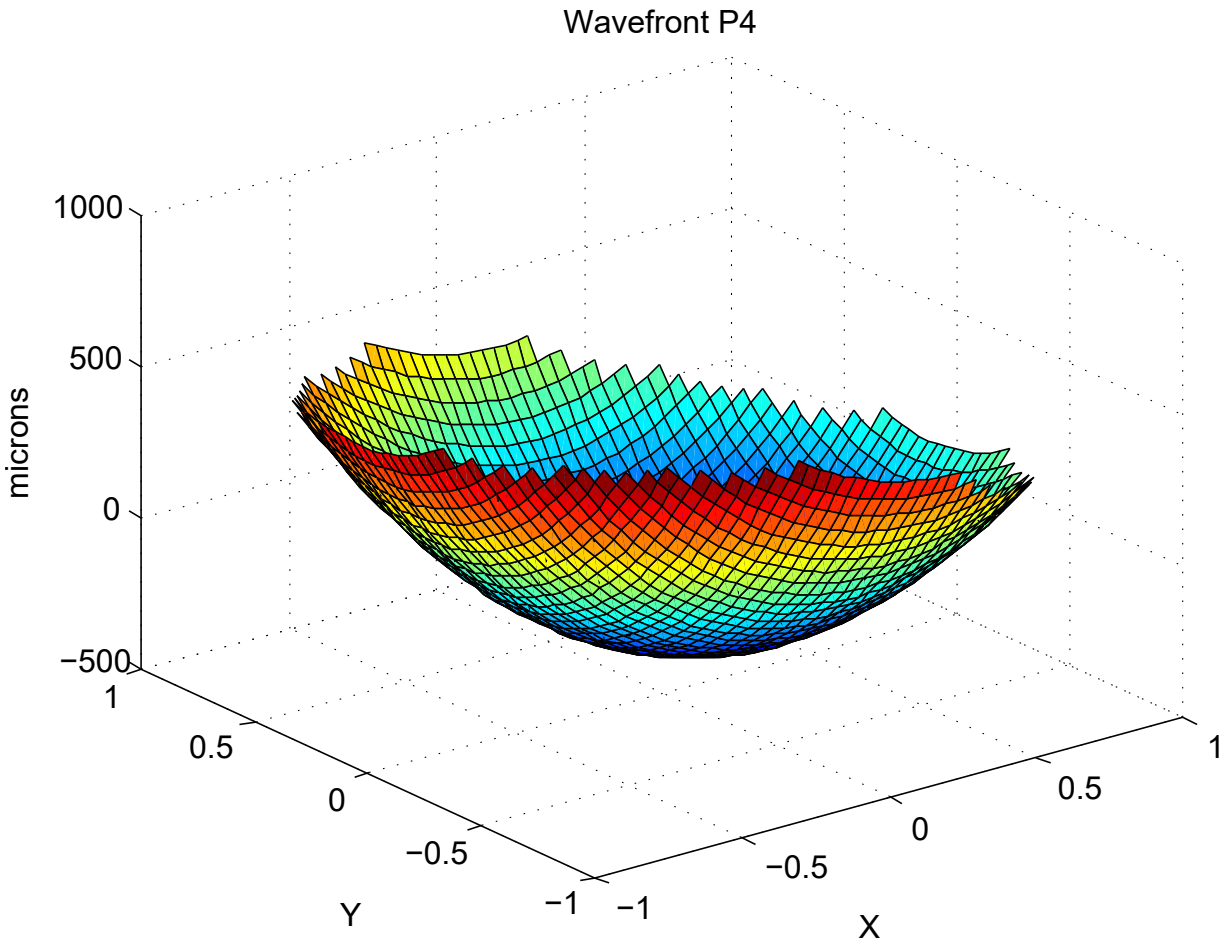


Figure 54: The wavefront measured from surface P4.

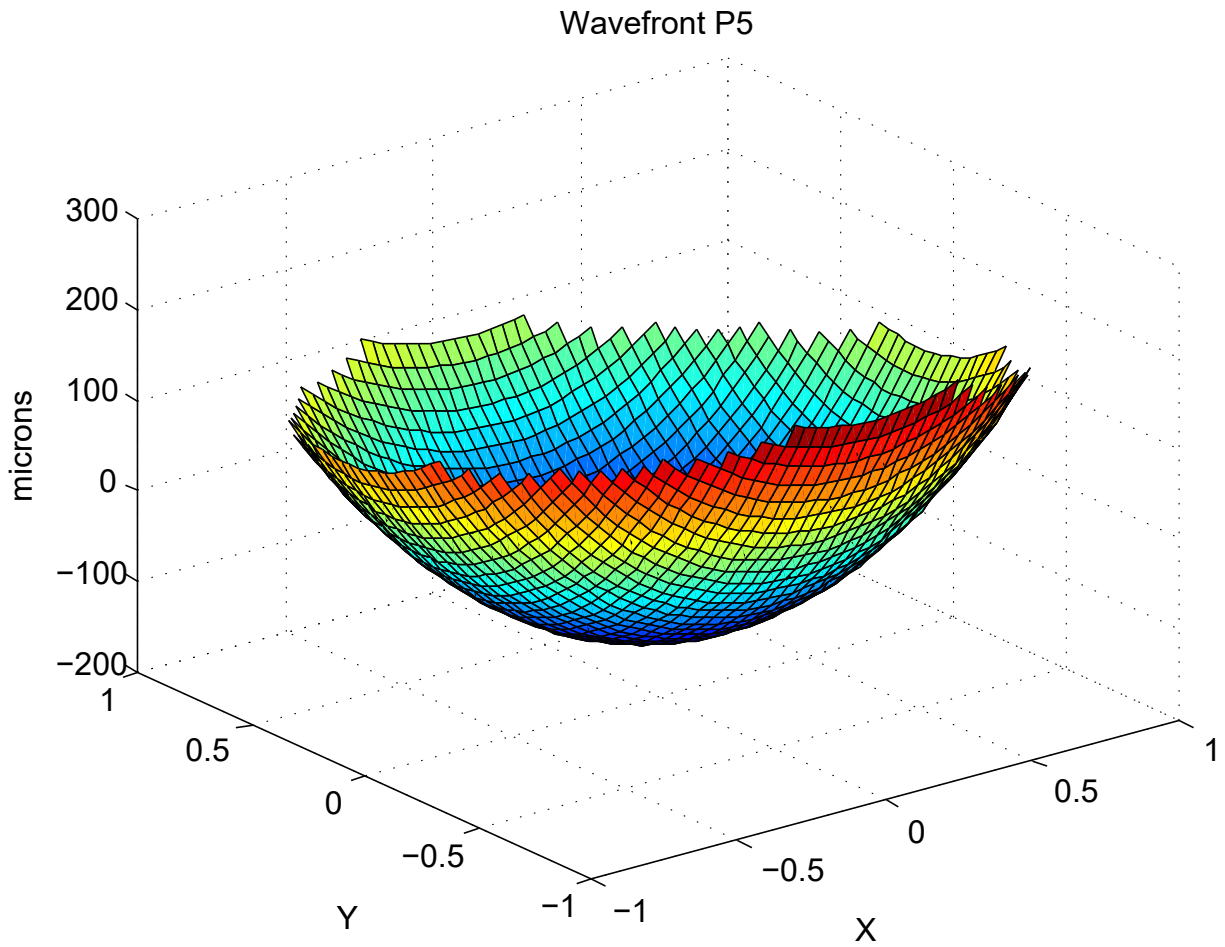


Figure 55: The wavefront measured from surface P5.

5 Field Programmable Gate Arrays and Real Time CGLRT

5.1 An Introduction to Field Programmable Gate Arrays

FPGAs are programmable logic devices (PLDs) that fill a niche between application specific integrated circuits (ASICs) full custom and general purpose integrated circuits (GPICs), such as CPUs and microcontrollers. ASICs are custom designed semiconductor devices that are manufactured for a specific purpose, (i.e. an MPEG-2 video encoder [21]) often for use in a specific device. ASICs are not reprogrammable and once manufactured, cannot be altered. This makes them extremely powerful tools for specific applications, but brings with it the drawback of being inflexible and cost-ineffective unless produced in mass quantities due to the fact that ASIC's require custom manufacturing of the circuit. On the other end of the scale are GPIC's which are well suited for many different applications but at a cost to performance due to lack of specialization. FPGAs, on the other hand, are well suited for applications that require a specialized device that can be reprogrammed as necessary, without the performance hit of a GPIC. This allows the engineer to design circuits at the logic-gate level for specific applications on a one-off basis. The drawbacks of FPGA design include coding in highly specialized hardware description language (HDL) with correspondingly long development times. Also, as designs get more complex, synthesis times can greatly increase making iteration on a design time consuming, at least compared to the short compile times associated with more familiar languages such as C++. The trade-offs are often well worth the effort though, as FPGAs are able to implement highly complex circuits, designed at the logic-gate level, allowing for fine-grain handling of digital signal processing (DSP) functions as well as a high degree of parallelization. This makes FPGAs a highly competitive platform for implementing complex, high-speed embedded systems. At the highest level, a typical FPGA consists of five main subsystems [67]:

1. Configurable logic blocks (CLBs)
2. Interconnects
3. Input/output buffers (IOBs)
4. Memory & DSP slices

Device	Slices	LUTs	Maximum Distributed RAM (Kb)	Flip-Flops
Virtex-6 XCVLX240T	37,680	150,720	3,770	301,440
Virtex-6 XC6VLX75T	11,640	46,560	1,045	93,120
Virtex-6 XC6VLX760	118,560	474,240	8,280	948,480
Spartan-6 XC6SLX4	600	2,400	75	4,800
Spartan-6 XC6SLX45	6,822	27,288	401	54,576
Spartan-6 XC6SLX150	23,038	92,152	1,355	184,304

Table 2: CLB resources available in a variety of Xilinx devices, including the Virtex-6 XCVLX240T which is utilised by the CGLRT instrument.

CLBs are basic building blocks of an FPGA. The exact structure of a CLB varies from device to device so here the discussion will focus on the Virtex-6 device [66] as it is utilised for CGLRT. Each CLB on the Virtex-6 contains two slices, where a slice consists four look-up tables (LUTs), eight flip-flop storage elements and multiplexers. Also, a subset of the slices support storing 256-bits of data as distributed random-access memory (RAM) and contain four 32-bit shift registers. Each slice LUT can be considered an independent function generator as it can implement any arbitrary six-input Boolean function. Multiple LUTs can be utilised in speciality slices to implement distributed RAM or read-only memory (ROM) elements as well. Table 2 summarizes the available CLB resources on the Virtex-6 XCVLX240T and other Xilinx devices. The Virtex-6 XC6VLX75T is the smallest device in the family while the Virtex-6 XC6VLX760 is the largest. The spread in values shows that the Virtex-6 XCVLX240T is a mid-range device. Resources from the Spartan-6 family of (smaller) FPGAs is also shown for the sake of comparison.

CLB signals are routed to each other and to external input/output connections via a series of interconnections that are controlled by the FPGA design software. This includes specialized interconnects for routing global clock signals, which are required to have low-skew, allowing the various distributed FPGA resources to remain in sync. The interconnects also provide access to the high-level IOBs which allow external signals to be sent/received to/from the FPGA. IOBs can be considered the way the FPGA ‘talks’ to the outside world. The number of IOBs varies per device with the Virtex-6 XCVLX240T having 600. For comparison the smallest Virtex-6 chip, the XC6VLX75T, has 360 IOBs whereas the largest, the XC6VLX760 has 1200.

FPGAs also contain dedicated memory resources known as block RAM (BRAM). Each BRAM instance on the Virtex-6 FPGA is 36 Kb wide and composed of two independently controlled 18 Kb RAMs. BRAM instances can be utilized for many purposes including single-port RAM (SPRAM), dual-port RAM (DPRAM), synchronous and asynchronous first-in/first-outs (FIFOs), and ROMs. The DPRAM instances are especially useful for crossing clock-domains as the read and write clocks can operate asynchronously. This is also true of the FIFO instances, which may be used to cross clock-domains as well. The Virtex-6 XCVLX240T has a total of 14,976 Kb of memory available as BRAM.

Another resource available on the Virtex-6 FPGAs is the DSP48E1 DSP slice which is a specialized slice composed of a 25×18 multiplier, an adder and an accumulator [65]. Due to their specific design these slices are able to run at a higher frequency than those implemented directly in fabric allowing the designer to take advantage of time-division multiplexing techniques and pipelining. The multiply-accumulate action is very common in DSP applications so the use to the DSP slices can free up general purpose FPGA resources. The Virtex-6 XCVLX240T has 768 DSP48E1 slices available.

5.2 An Embedded Systems Approach to Wavefront Sensing

In practice, wavefront sensors are not stand-alone instruments, but part of a larger imaging system. This is true for applications such as OCT where the wavefront sensor will be incorporated into the AO sub-system that performs aberration correction for the larger imaging system. In clinical applications these systems are most often found in the form of self-contained instruments with integrated optics, electronics, mechanical and computing sub-systems.

The demand for real-time performance of such systems has put a higher demand not only on the optical and mechanical portions of the system but also on the speed of the computational system. As the optical systems have become more sophisticated the limitations on real-time performance have shifted to computing system's ability to perform image reconstruction and associated tasks. This has led to a great interest in specialised parallel computing resources such as graphics processing units (GPUs) and FPGAs. In the case of CGLRT the decision was made to utilise an FPGA platform, opposed to a GPU, for several reasons. Firstly, the FPGA is better suited to interface with the external hardware required by CGLRT. The availability of abundant I/O and convenient form factors such as the FPGA mezzanine card (FMC) and PMOD facilitate (relatively) simple connections to external hardware. For example, in the CGLRT instrument the Xilinx FMC-CE expansion

card (shown in Figure 88) is used to connect two external digital to analogue converter (DAC)s and provide manual user inputs via slide-switches and push-buttons. The FMC and PMOD standards will be discussed in more detail in Chapter 6.

Secondly, the FPGA offers fine-grained parallel computing resources that allow for real-time wavefront reconstruction. The FPGA can easily interface with ADCs, DACs, and similar external systems which make it well suited for system command and control operations. GPUs also excel at parallel computing, especially in the case of floating-point operations, but still require a CPU to interface with and are not able to control external hardware on their own. Interfacing to the CPU requires the transfer of large amounts of data from the GPU memory to the CPU (typically over a PCI bus) which can often eliminate any gains in speed made by incorporating the GPU in the first place [22]. Lastly, it has been shown that FPGAs are better suited for OCT type imaging systems [22, 53] due to the scalability of combining several FPGAs into a single system.

The drawback of utilising an FPGA lies in the development time. Li, et al., [22] state, for example, that development of an OCT image reconstruction system on a GPU took approximately 3 months whereas development of a comparable system on an FPGA took approximately a year. Another drawback associated with FPGAs is the difficulty of generating test signals and debugging. Traditionally, these tasks were accomplished via a *testbench* which is a separate piece of HDL used to define input signals and drive the FPGA circuit with them at given times. Building a testbench is often very time consuming and an art in itself. To overcome these limitations the MATLAB suite of software was utilised for a major portion of the system development.

5.3 FPGA design workflow

MATLAB may be used to generate highly-efficient VHDL by utilising the well-known block diagram based Simulink environment. Xilinx provides specialized Simulink blocks that correspond directly to FPGA resources such as registers, adders, and DSP slices. The integrated MATLAB workspace may then be used to generate test signals which are sent to the Simulink FPGA design and processed. The outputs may be captured and sent back to the MATLAB workspace for analysis. This makes testing and debugging the design much simpler and time efficient. The final design may then be synthesised within the Simulink environment generating the files needed to program the FPGA hardware. These files may then be utilised as components in a higher-level design, etc.

A diagram of the System Generator based FPGA design workflow is shown in Figure 56. Within a given System Generator Simulink project there is a subsection which must be comprised completely of the Xilinx supplied blocks (shown as the grey section labelled Synthesizable). Each Xilinx block consists of some type of fundamental FPGA component whether that be a constant, adder, multiplier, comparator, or the like. These blocks are combined to form the main systems within the design. Only the Xilinx provided blocks may be used in this portion of the design because these blocks are able to be synthesized directly into hardware resources. Signals may be sent to the design as inputs and read from the design as outputs via the *input gateway* and *output gateway* blocks. These blocks map directly to the top level IOBs of the FPGA. Signals may be sent into, and read from the design from the MATLAB workspace using these blocks as well. The input and output gateways are the only allowable method for sending and receiving external signals to/from the design. From a design point of view this greatly enhances the engineer's ability to quickly test and debug the circuit. No testbench is needed as all testing may be done directly via the MATLAB workspace. External devices such as DACs and ADCs may be simulated within the MATLAB workspace as well, providing a way to integrate the FPGA architecture with external hardware. After the system has been tested within the MATLAB environment it may be synthesized into HDL which is used to program the hardware directly.

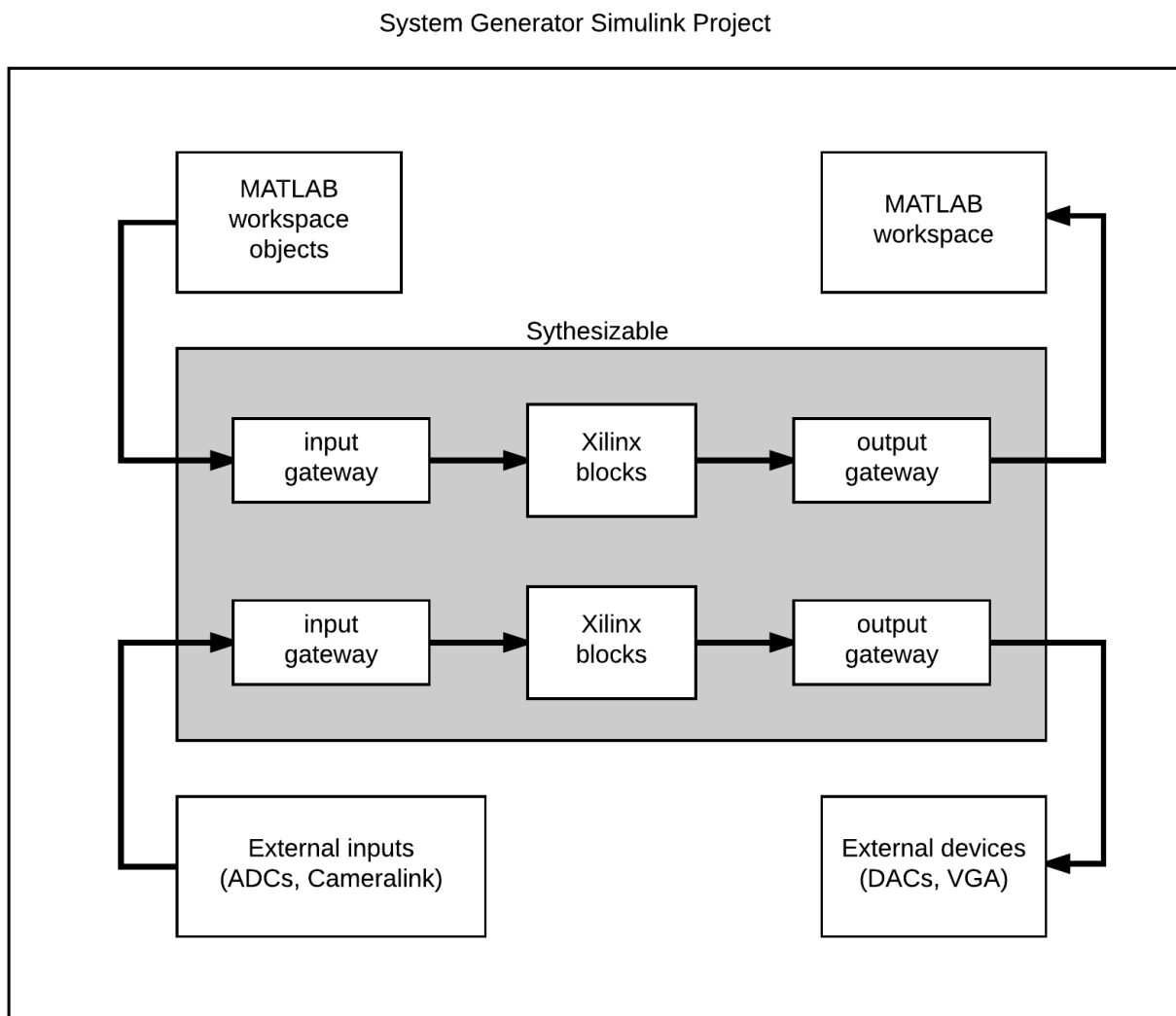


Figure 56: Flow diagram of System Generator based FPGA development. Input and output gateways send and receive signals from the MATLAB workspace. The grey portion of the illustration represents the part of the system that is synthesizable into hardware.

5.4 CGLRT System Architecture

Figure 57 shows a top-level block diagram of the CGLRT FPGA system. Each of the subsystems shown will be discussed in detail in the following sections, but a short overview is given here. In the centre of the diagram is the master state-controller (MSC) which is responsible for coordinating all of the subsystems in the design and generating the command signal to control them. In essence, the MSC is an finite state machine (FSM) which acts as

the “brain” of the CGLRT system. At the top of the diagram are the Scanner Controller and Scanner DAC subsystems. The Scanner Controller is responsible for generating the control signals that the Scanner DAC will convert into voltages supplied to the galvo-scanning mirror. The PZT DAC is responsible for generating the voltages necessary to step the reference mirror PZT which modulates the optical path-length of the interferometer. To the bottom right the external camera sends pixel data to the Frame Grabber. The MSC determines when a frame should be acquired, depending on the scan state, and directs the Frame Grabber accordingly. Pixel data from the Frame Grabber is saved in either Raw Image Memory A or Raw Image Memory B. The save location is determined by the MSC (control signals not shown for sake of clarity) and corresponds to the phase modulation as required by the four-step algorithm from Equation 4.3. Once the needed images are acquired the PSI algorithm, centroiding, and slope calculations are performed in the Image Post Process block. The slopes are then streamed to the Wavefront Recovery subsystem where the Zernike mode magnitudes are calculated. They may then be delivered to an external PC via the UART connection. The following sections provide details on the operation of each of these subsystems.

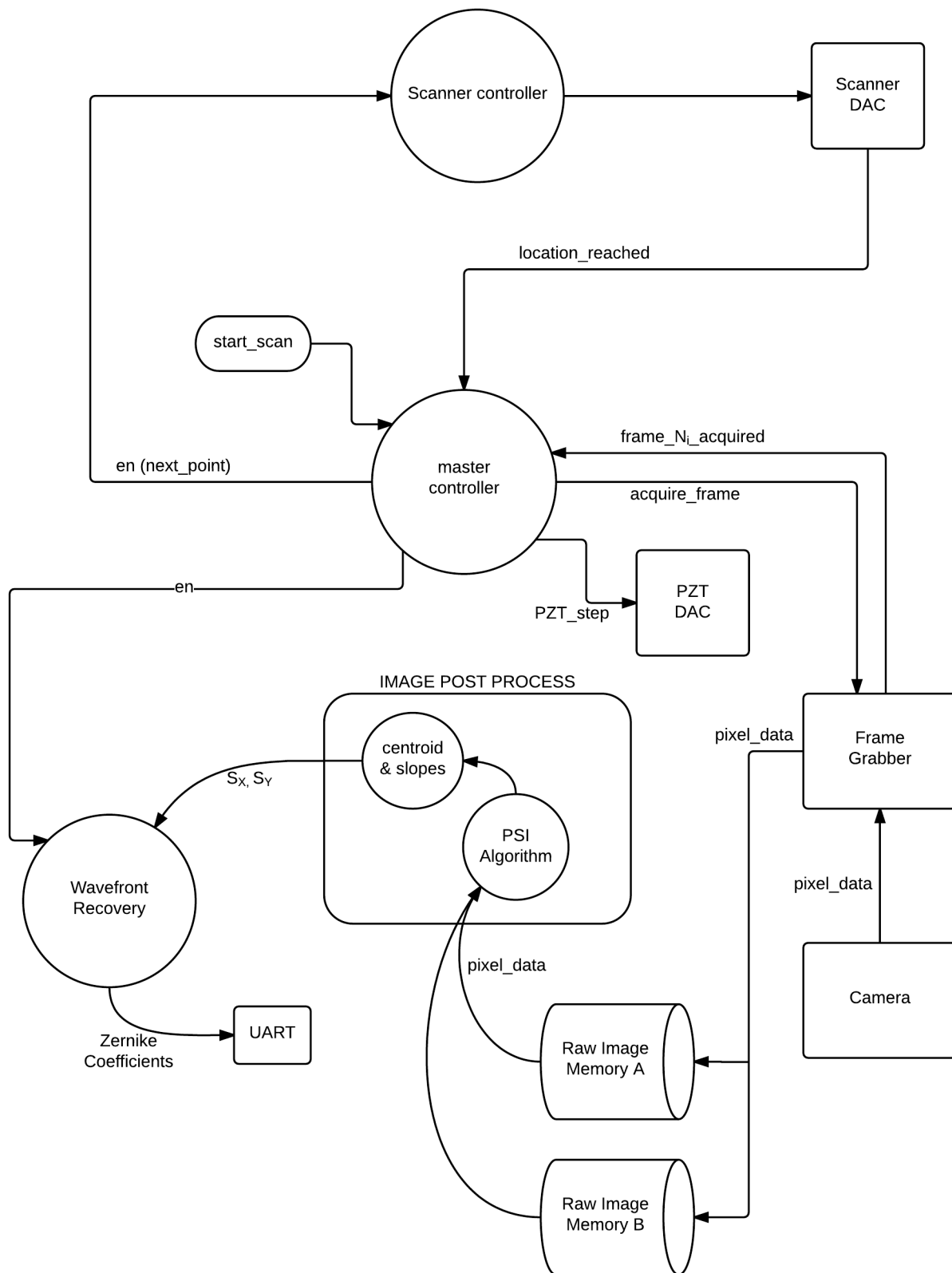


Figure 57: CGLRT top level system block diagram.

5.4.1 Master State Controller

The MSC is responsible for generating all the command signals used to coordinate the various CGLRT subsystems including the scanner controller (responsible for handling the scan patterns and galvo-scanner DAC), the frame grabber, the PZT DAC (responsible for modulating the reference arm path length) and the wavefront reconstructor.

Figure 58 shows a state diagram of the MSC. The following signals are utilised by the MSC:

- **start_scan** (bool): User input signal indicating a scan is to be performed. Held high for one clock cycle.
- **num_pts** (6-12 bits depending on scan size): Total number of points to be acquired in a scan.
- **N** (6-12 bits depending on scan size): Counter used to keep track of the number of points that have been acquired by the scan in progress.
- **i** (2-bits): Counter used to keep track of the four PSI images ($I_0 - I_3$).
- **next_point** (bool): Commands the Scanner Controller to move the galvo-scanner to the next point in the scan. Held high for one clock cycle.
- **point_acquired** (bool): Indicates to the MSC that the galvo-scanner has moved to the next point. Held high for one clock cycle.
- **acquire_frame** (bool): Commands the frame grabber to begin reading the next available frame from the camera.
- **frame_acquired** (bool): Indicates to the MSC that the frame grabber has completed acquiring the frame.
- **pzt_step** (bool): Commands the PZT DAC to modulate the reference arm path-length for PSI image acquisition.

When the system is powered up the MSC initialises to the **RESET** state where counters **N**, **i** and the **start_scan** signal are all set to 0. **N** is used to store the number of scan points that have been acquired and **i** is a 2-bit counter used to keep track of the PSI frames 0-3. After the **RESET** state is complete the system transitions to the **IDLE** state. Here the system

simply waits for the `start_scan` signal to be exerted. Once a scan is started the system repeats the following sequence of states until `N` equals the total number of points.

First, the system transitions to the `go_to_point` state where `next_point` is held high for one clock cycle. This indicates to the scanner controller to move the galvo-scanning mirrors (see Section 5.4.2 for details). The MSC transitions immediately to the `wait_for_point` state where the `point_acquired` signal is awaited. When `point_acquired` is high, this indicates that the galvo-scanner has finished moving and a frame may now be acquired. The MSC then transitions to the `grab_frame` state where the `acquire_frame` signal is set high, indicating to the frame grabber that it should be gathering pixel data from the camera. The MSC immediately moves to the `wait_for_frame` state. Once the frame is acquired the `frame_acquired` signal is held high by the frame grabber and the MSC transitions to the `move_pzt` state. Here the `pzt_step` signal is held high commanding the PZT DAC to step the PZT. Also, the `N` and `i` counters are incremented. If $i < 4$ then all coherence-gated (CG) images have not been acquired so the MSC returns to the `grab_frame` state to acquire the next CG image. If $i = 4$ and $N \neq \text{num_points}$ then the MSC returns to the `go_to_point` state to acquire the next point in the scan. If, on the other hand, $i = 4$ and $N = \text{num_points}$ then the entire scan is complete and the MSC transitions to the `reset` state where the counters are reset and the system returns to the `idle` state. The system will remain in the `idle` state until another `start_scan` signal is exerted.

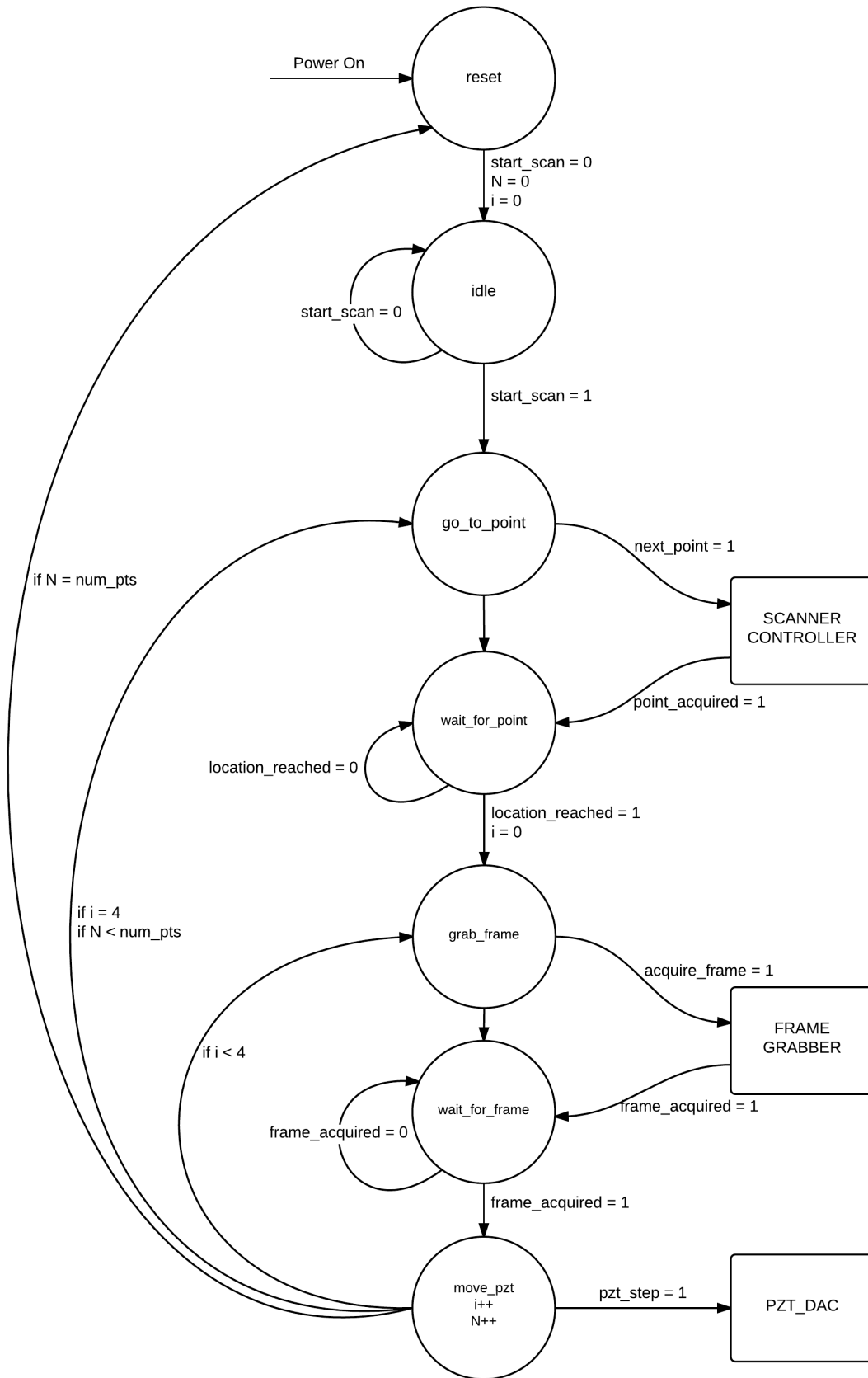


Figure 58: Master State Controller block diagram.

5.4.2 Scanner Controller

The scanner controller is responsible for handling the control of the galvanometer mirrors used to articulate the scanning beam position. The subsystem consists of three components, as shown in Figure 59, which are the scan pattern generator, the DAC controller, and the external DAC hardware.

The scan pattern generator, shown in Figure 60, is responsible for generating the scan pattern control signals that are sent to the DAC Controller. It can generate either a circle scan or a raster scan, depending on the `pattern_select` control input. It consists of four LUTs, two store the circle pattern x and y coordinates, and two the raster pattern x and y coordinates. The stored values are pre-computed at compile time. The LUTs are addressed by a counter which is reset by the `scan_en` signal at the beginning of a scan and only increments when the `next_point` signal is exerted. Two multiplexers, controlled by the `pattern_sel` signal, are utilised to select between the raster and circle pattern coordinates. Finally, the `next_point` signal is sent, after a delay of one clock-cycle (to account for the LUT delay) to the DAC Controller as a `DAC_enable` signal. This signal is exerted to inform the DAC Controller that the x and y coordinates should be written to the external DAC. Input control signals `scan_en`, `pattern_sel` and `next_point` originate from the MSC.

The DAC Controller consists of a simple two state FSM with an IDLE and WRITE state. It is responsible for taking the 12-bit parallel `scan_x` and `scan_y` writing them serially to the DAC hardware. It also generates the DAC hardware clock `SCLK` signal. On startup, it initialises to the IDLE state where the `SYNC` signal is held high, indicating to the DAC hardware that a write is not taking place. When the `DAC_enable` signal is exerted the current `scan_x` and `scan_y` values are loaded into two 16-bit registers, with the four MSBs being set to 0 (The DAC hardware requires a 16-bit input, with the four MSBs being control inputs, the 12 LSBs the signal). The controller then switches to the WRITE state, dropping the `SYNC` signal low and serially writing the 16 bits to the DAC hardware over the `scanner_x` and `scanner_y` signals. Once the write is complete the controller holds the `point_acquired` signal high for one clock cycle, indicating to the MSC that the value has been written to the DAC hardware. The state then returns to IDLE and awaits the next point.

The DAC hardware consists of two National Semiconductor DAC121S101 chips on a PMOD-DA2 board from Digilent. The PMOD standard from Digilent [6] is a peripheral interface composed of either 12-pin or 6-in connectors. Every PMOD connection provides VCC and GND connections. The board contains two 12-bit D/A converters with power and ground signals provided by the ML605 FMC interface. The chips accept `SCLK` clock signals

up to 30 MHz. For this application the SCLK is divided down by two in the 50 MHz clock domain to 25 MHz for simplicity. The output voltages of the DAC hardware are connected directly to the galvanometer scanning-mirror driver.

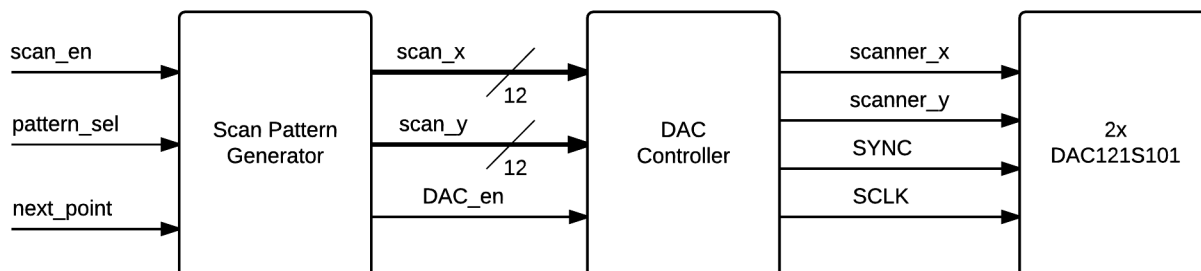


Figure 59: Top-level block diagram of the Scanner Controller.

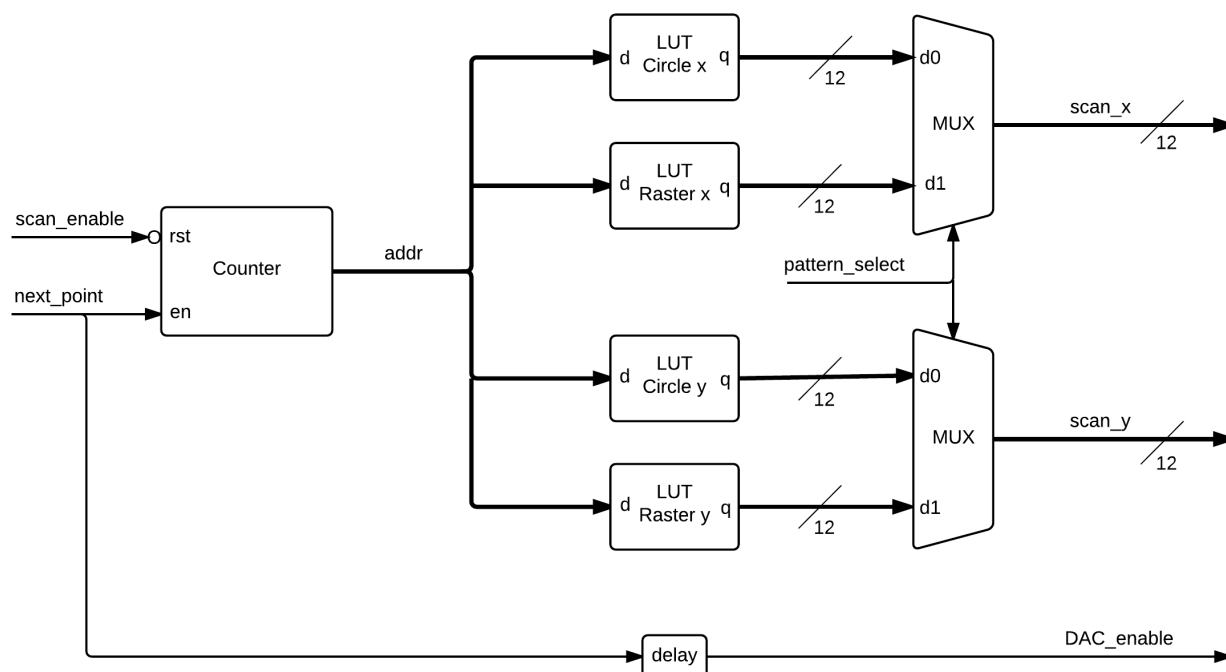


Figure 60: Scan Pattern Generator.

5.4.3 Clock Domains and Clock Management

Digital architectures do not require all subsystems to run at the same clock rate. It is quite common to have systems composed of several different ‘clock domains’ which run at different

rates. It is often preferable to implement clock domains via ‘clock enables’ which divide down the main system clock. This ensures that all clock signals are in sync by only routing the main system clock throughout the FPGAs specialized global clock interconnects. Another option is to generate separate clock signals for each clock domain using a clock manager, which on the Virtex-6 is accomplished using the multi-mode clock manager (MMCM) resource. While utilization of clock enables is often the preferred method, as it avoids the need to transmit data across parts of the circuit running at different frequencies, this approach was not chosen due to the requirement of using the external pixel clock from the camera. The pixel data coming in from the camera must be synchronized with the FPGA frame-grabber, which requires the frame-grabber to run off of the pixel clock provided by the camera. As this external clocking requirement already created two clock-domains (the pixel clock and system clock) the decision was made to keep to a single design methodology and utilize an MMCM to generate all the needed clock signals.

Figure 61 shows how the various subsystems are divided into four clock domains. The six subsystems within the camera clock domain are dependent upon pixel data to function and are therefore limited to 50 MHz by pixel clock. While it is possible to run some of these subsystems at a higher rate, there would be no advantage as the system would simply be idle while waiting for new pixels. The DAC clock domain contains the Scanner Controller FSM, galvo-scanner DAC interface and PZT DAC interface. All of these subsystems are run at 20 MHz which is compatible with the DAC121S101 chip. The VGA driver and video memory are contained in the VGA clock domain and run at 25.175 MHz which is dictated by the screen resolution (640×480). Lastly, is the clock domain utilised by the UART communication which runs at 100 MHz. This was chosen for convenience as the existing UART code was designed to run at this frequency.

Data is transferred across clock domains using either registers or shared-memory modules. When a single bit signal, such as a flag or enable/disable is transferred, then the signal is shared using a register. If a more complex signal is being transferred then a shared memory module is utilised which consists of independently clocked dual-port RAM modules.

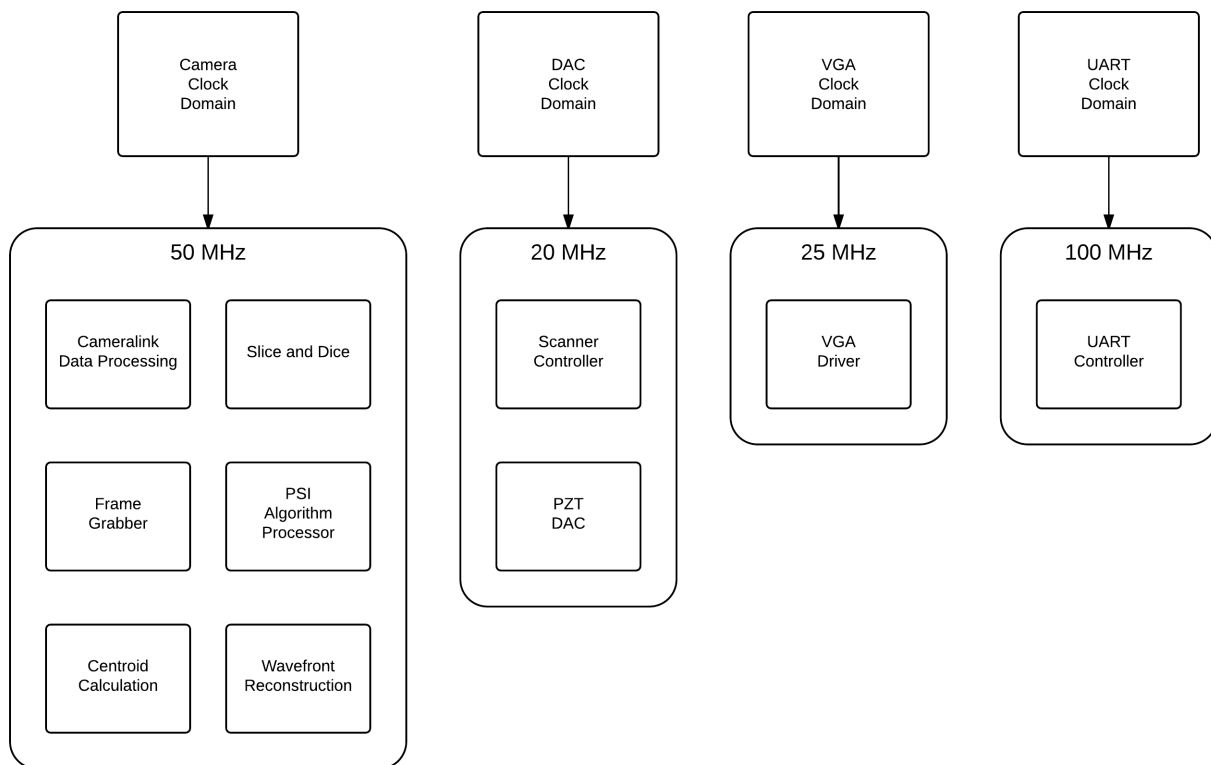


Figure 61: Clock domains utilised in the CGLRT system.

5.4.4 Cameralink Interface

Pixel data from the camera arrives at the FPGA as a 28-bit word `image_data` which is a standard for the Cameralink protocol. In base mode (Cameralink has several modes available) `image_data` has the capability of sending three 8-bit pixels, three control flags (`frame_valid`, `line_valid` and `data_valid`) and the pixel clock. As the CGLRT camera is greyscale, only one pixel is sent per 28-bit word and the other pixel values are zero.

The subsystem for handling the camera output is shown in Figure 62. The 28-bit word has the 8 LSBs sliced out forming `single_pixel`. The `frame_valid` and `line_valid` flags are sent to a 3-input AND gate along with the `cameralink_locked` flag all of which generate the `pixel_valid` flag. Finally, the `frame_valid` flag (which is held high by the camera for the entire frame) is sent to a rising-edge detector to generate a one-cycle pulse used to indicate a frame has started. The `single_pixel` and `pixel_valid` flags are delayed by one clock cycle to keep them in sync with the `frame_start` flag.

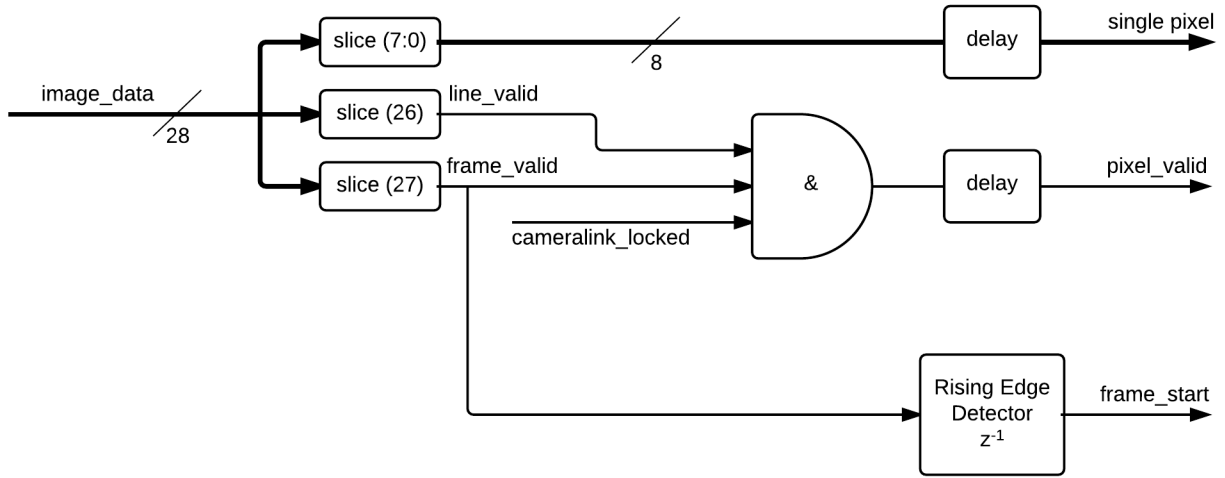


Figure 62: Cameralink Interface

5.4.5 Frame Grabber

In order to form the coherence-gated image I_c , four phase-shifted images, $I_0 - I_3$ are required. The acquired images are 250×250 pixels, which is the largest frame that the CCD camera can acquire at 96 fps. This image size was found to be sufficient for the observed deviations in the 2.5mm scan area when focused on the detector. Lower resolutions may be chosen (down to 25×25 pixels) to increase the frame-rate of the camera, but these were too small to capture the observed deviations from the MLO. In practice, the trade-off between frame-rate and image will need to be tailored to the individual imaging system. These images must be grabbed from the live camera feed, one frame at a time, when the `grab_frame` signal is exerted by the MSC. Figure 63 shows the state-flow diagram for the frame grabber. The FSM initializes to the `idle` state and waits for the `grab_frame` signal to be exerted. The `fg_ready` signal is held high indicating to the MSC that a frame can be obtained. Once `grab_frame` is exerted, the FSM transitions to the `wait for frame` state where a pixel address of 0 indicates the start of a new frame. The `fg_ready` signal is held low, signalling to the MSC that it must wait for this frame to be finished before acquiring another. At `addr == 0` the FSM transitions to the `acquire_frame` state and begins sending the frame data to the image binarization circuit. While the frame is being sent the `binarize_en` signal is held high to enable the binarization circuit. Once the entire frame has been acquired (`addr == w*h-1`) then `binarize_en` is de-exerted and the FSM returns to the `idle` state. Lastly, the `fg_ready` signal is once again held high, indicating that the next frame may be obtained.

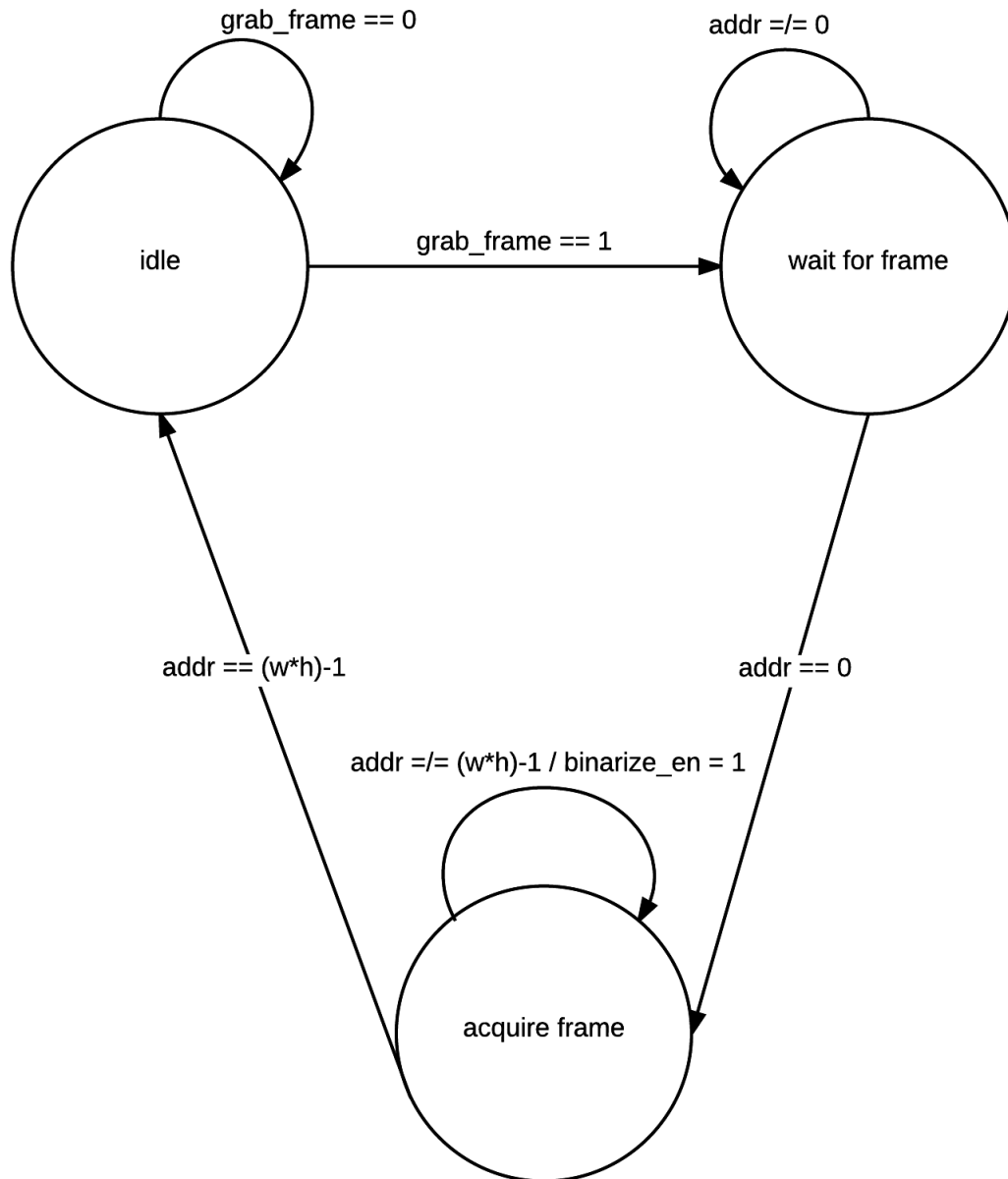


Figure 63: Frame Grabber state diagram.

5.4.6 PSI Algorithm

Due to memory limitations in the FPGA it is critical to minimize the amount of resources utilized for storing intermediate frames used for the PSI image processing and centroid calculation. Recall that obtaining the D th coherence-gated image I_D , requires four phase-shifted images as shown in Equation 3.3. Therefore, a total of 256 images need to be obtained for even the smallest scan size of 64 points. For 8-bit pixels and a 250×250 image this would

require 128Mb of memory to store all of the frames, much more than the roughly 15Mb of block RAM available on the XCVLX240T device. Even simply storing the four phase-shifted images $I_0 - I_3$ would require 2Mb ($\approx 15\%$ of block RAM). Block RAM availability is the major factor in determining how many Zernike modes may be reconstructed due to the use of RAM to store the precomputed A^\dagger matrices as discussed in Section 5.5.2. Therefore, it is critical for the design to minimize the memory used for image processing.

To this end, a subsystem was built that utilizes two DPRAM elements to process images I_{0-3} . One DPRAM processes the even images (I_0 and I_2) while the other DPRAM processes the odd. The functioning of the odd and even subsystems is effectively identical so only a single one is considered in detail. Figure 64 shows a block diagram of the PSI subsystem. The system is completely pipelined and operates on a pixel-by-pixel basis. On system startup each memory location in the DPRAM is initialized to zero. Pixel data (8-bit) enters the circuit on the `pixel_in` line. The `pixel_in` signal is delayed by one clock cycle so that it matches the `doutA` signal (due to the one cycle delay of the DPRAM). The stored value on `doutA` is then subtracted from the `pixel_in` value which increases the word length to 9 as the previous `unit8` pixel value is now signed. Next, the absolute value of the difference is taken and the 8 LSBs are sliced off and written into port B (`dinB`) of the DPRAM. The subtraction, absolute value, and slicing are pipelined and completed in one clock cycle, so the `addr` value sent to `addrB` is delayed by one clock cycle to keep the data in sync. Lastly, the `addrB` and `doutB` signals are concatenated and sent to the VGA driver circuit for external monitoring of the image.

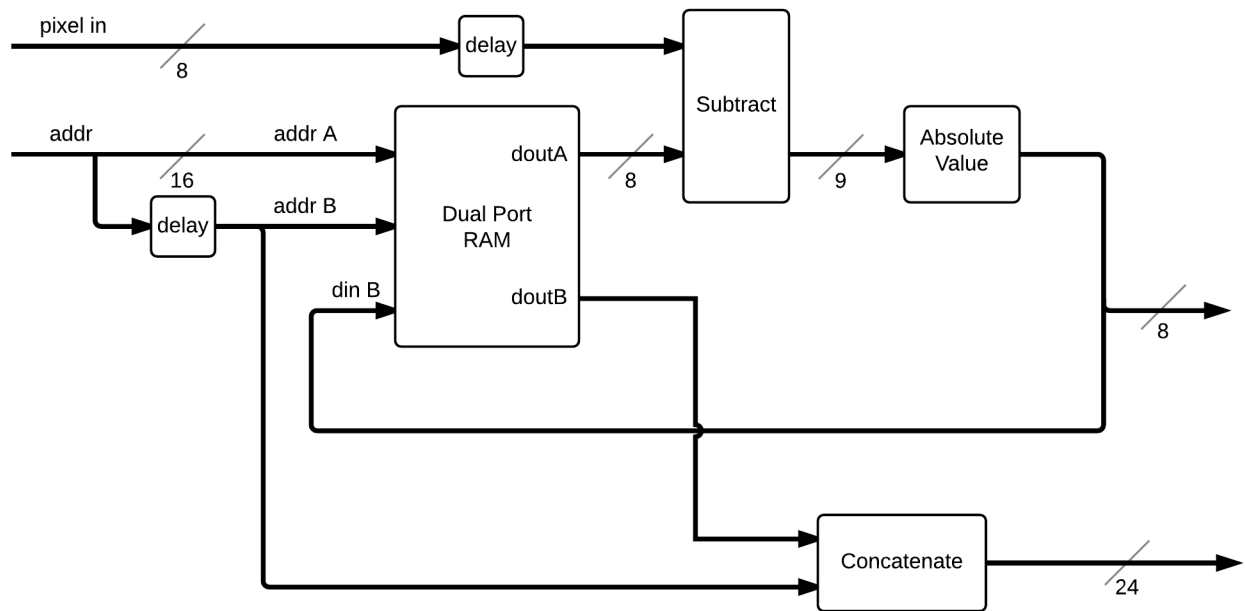


Figure 64: Streaming Image Subtraction.

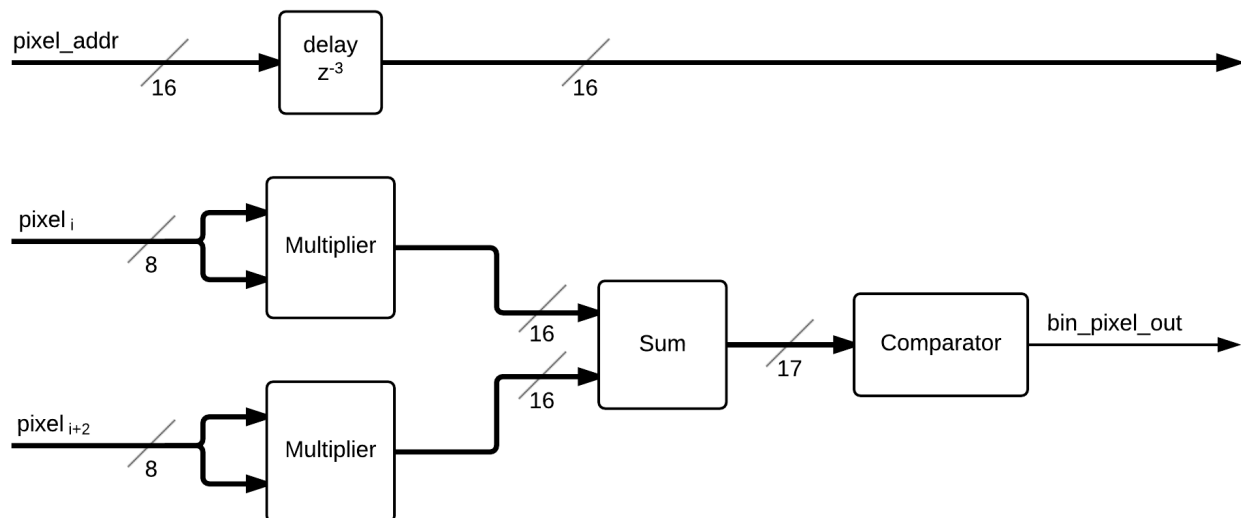


Figure 65: Pixel Squaring and Binarization.

The MSC enables the appropriate image subtraction subsystem depending on whether the even or odd images are being captured. When the two pairs of images have been subtracted the resulting pairs of pixels are squared and summed on a pixel by pixel basis according to Equation 4.3. Though the equation calls for square-rooting the sum, this is

not necessary, as the pixels are immediately binarized via a comparator. The comparator threshold is set manually. The value was determined by binarizing several images obtained from the test CGLRT setup. A best effort was made to minimize noise in the image while still maintaining as many pixels as possible in the CG spot. This is beneficial for the design as the square-root function is costly in hardware. Therefore, after the pixels are added together the sum is sent to a comparator resulting in the Boolean signal `bin_pixel_out`.

5.4.7 Centroiding

The centroiding subsystem takes a single frame of a CG image and calculates the 2D centroids \bar{x} and \bar{y} where

$$\bar{x} = \frac{M_{10}}{M_{00}}, \quad (5.1)$$

and

$$\bar{y} = \frac{M_{01}}{M_{00}} \quad (5.2)$$

where

$$M_{ij} = \sum_x \sum_y x^i y^j I_{CG}(x, y). \quad (5.3)$$

Figure 66 shows the section of the circuit responsible for calculating M_{00} , M_{01} and M_{10} . The important components in this circuit are:

- **bin_pixel_out** (bool): The n th binary pixel value of a given I_{CG} .
- **en** (bool): This signal is high every time valid pixel data is available on **bin_pixel_out**.
- **pixel_addr** (ufix16): Address of the currently valid pixel. For a 250×250 image this value will range from 0 to 62499.
- M_{00} , M_{01} and M_{10} (int19): The three image moments used to calculate the centroids in Equations 5.1 and 5.2.

For each I_{CG} the **pixel_addr** will be 0 at the start of a frame. Two 8-bit counters **counter_x** and **counter_y** keep track of the current pixel coordinate within the frame and count from 0-250. **counter_x** is incremented by the **en** signal each time a valid pixel is available. Each time **counter_x** completes a line **counter_y** is incremented by comparator C1. The x and y counter outputs are sent to accumulators A1 and A2 respectively that keep a running sum of the values. The accumulators are enabled only when **bin_pixel_out** = 1. Accumulator A3 keeps a running sum of all pixels when **bin_pixel_out** = 1 and therefore does not need the enable signal. On the last pixel of the frame (**pixel_addr** = 62499) comparator C2 is anded with the **en** signal generating a one clock cycle pulse which, after a delay, indicates that the frame is complete. This signal is used to reset the counters and accumulators, preparing them for the next available frame. It also forms the **centroid_valid** signal which is used to enable the registers shown in Figure 67.

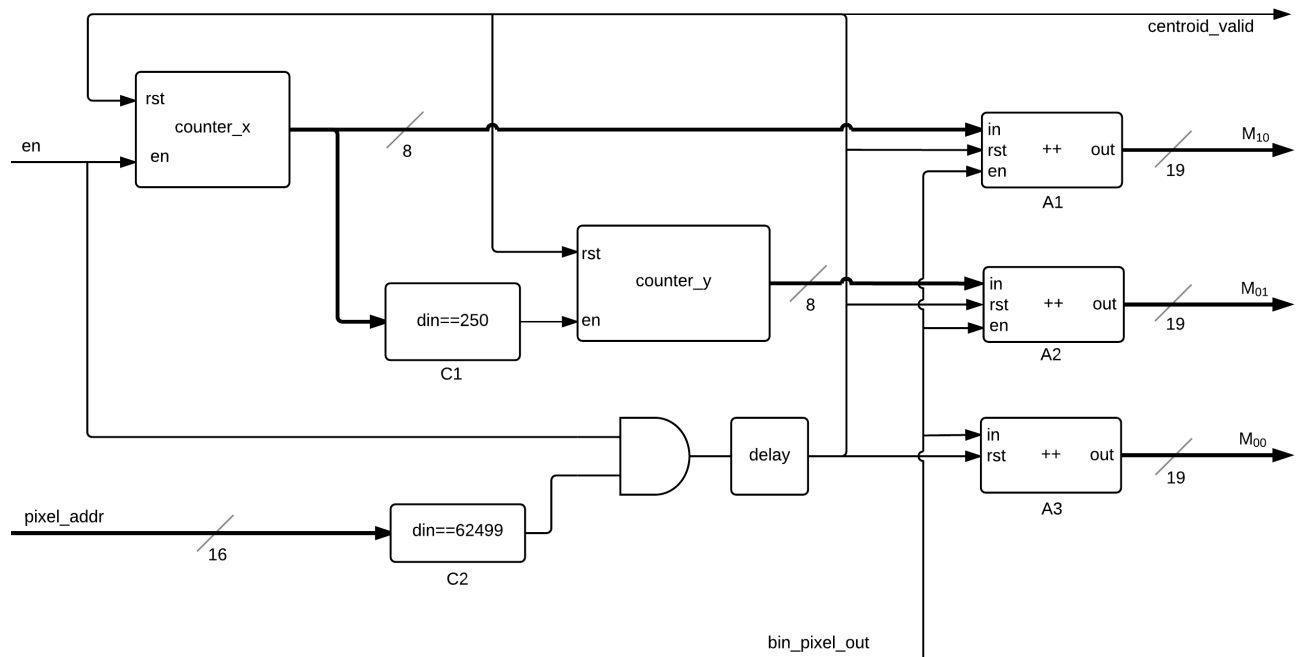


Figure 66: Circuit responsible for calculating the three moments required for the centroid.

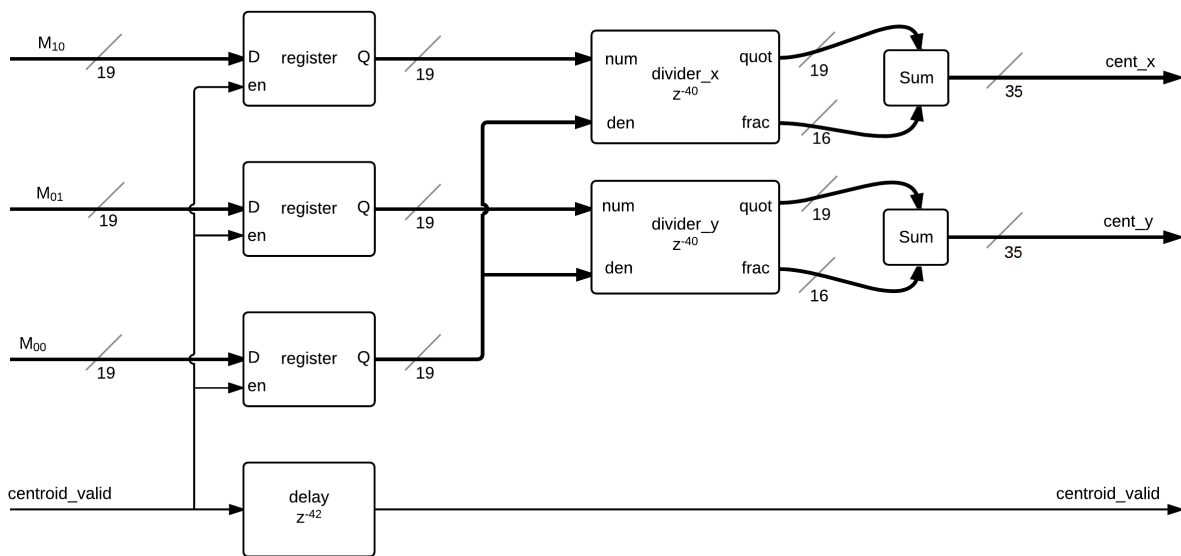


Figure 67: Division portion of the centroid calculation.

Each register is enabled by the `centroid_valid` pulse which loads the data for the three moments. On the next clock cycle the data is available to the two divider blocks on the numerator (`num`) and denominator (`den`) inputs. The dividers are implemented with the

Xilinx Divider 3.0 core. It has two outputs which are the quotient (`quot`) and fractional (`frac`) parts. These are then concatenated forming two 36-bit words (fraction part of 16-bits) which are the x and y centroids (`cent_x` and `cent_y` respectively). The bit-depth of the fractional and quotient parts are required by the architecture of the Divider 3.0 core and are later truncated to 12-bits, which is the input requirement for the wavefront recovery circuit. The `centroid_valid` flag is delayed by 42 clock cycles so that it remains in sync with the two centroid outputs (the divider core takes 40 cycles to complete the division. 40 cycles at 50 MHz = 0.8 μ sec which, in terms of system performance, is trivial).

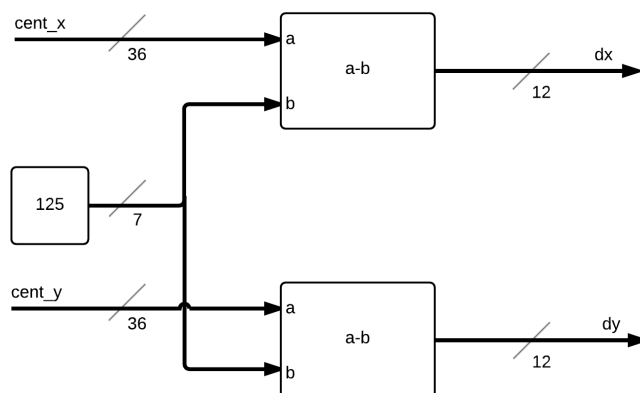


Figure 68: Calculation of the wavefront slope from the centroid.

Lastly, the wavefront slope is calculated by subtracting the centroids from the centre of the frame (125,125) as shown in Figure 68. A future iteration of the CGLRT instrument may include the ability for a calibration scan to be stored, where the centroid of each unaberrated spot location is stored for this subtraction. The final wavefront slopes dx and dy are formed by 12-bit words 4 of which are the fractional component. After the slopes are calculated they are sent to the wavefront recovery portion of the circuit, which uses the modal method to represent the wavefront as a sum of Zernike modes. Section 5.5.1 provides details of this subsystem.

5.5 Testing

To test the centroiding subsystem three black and white test images were generated in MATLAB. The centroids were calculated using MATLAB and compared to the centroids calculated by the centroiding subsystem (specifically signals `cent_x` and `cent_y`). Figures 69, 70 and 71 show the binary images that were generated in MATLAB. The MATLAB

calculated centroid is displayed on each image as a red circle. Centroid Test Images 1 and 2 are made up of two spots each. While this situation is not expected to happen in normal LRT operation it was used to demonstrate that the centroiding subsystem would produce the correct centroid for complex images. Centroid Test Image 3 is representative of a typical CG image. Table 3 compares the x and y centroids calculated by MATLAB with those computed by the FPGA centroiding subsystem along with the percent error for each component. Images 1 and 2 show error of less than 0.2% per component which is due to the limited numerical precision of the 4-bit fractional component utilised.

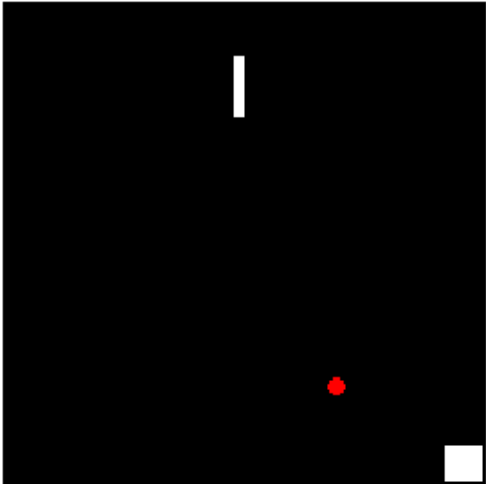


Figure 69: Centroid test image 1.

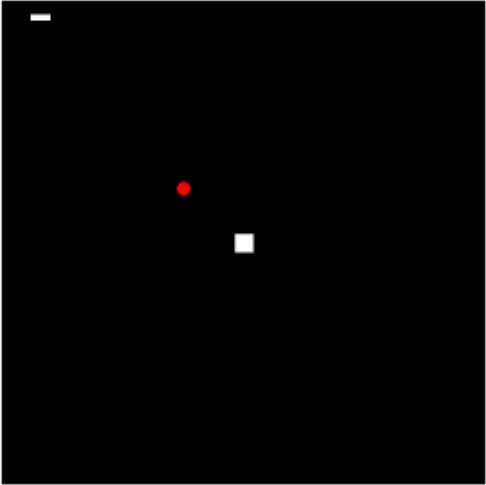


Figure 70: Centroid Test Image 2.

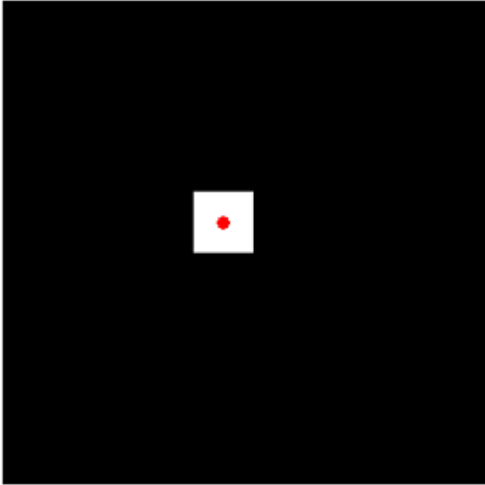


Figure 71: Centroid Test Image 3.

Image #	MATLAB x	MATLAB y	FPGA x	FPGA y	% error x	% error y
1	199.38	173.03	199.25	173.21	0.07%	0.11%
2	97.48	94.37	97.31	94.24	0.17%	0.14%
3	115.00	115.00	115.00	115.00	0.00%	0.00%

Table 3: Percent error for MATLAB and FPGA calculated image centroid for test images 1-3 in Figures 69, 70 and 71 respectively.

Next, three raw CG images were captured by the Labview based system (as described in Chapter 4) and the centroids were once again processed by MATLAB and the FPGA centroiding subsystem for comparison. Figures 72, 73 and 74 show the raw CG images and Figures 75, 76 and 77 show the binary images formed after thresholding. The MATLAB calculated centroid is indicated by the red ring. Table 4 shows the results of the centroid calculations from MATLAB versus the FPGA centroiding subsystem. The percent error between the two methods is below 0.35% for both x and y components which represents less than a single-pixel worth of disagreement. This testing shows that the centroiding subsystem is able to accurately calculate spot centroids for the CG images formed by the CGLRT system.

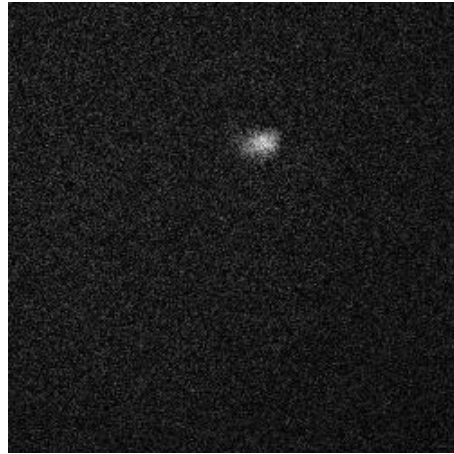


Figure 72: Raw Coherence Gated Image 1.

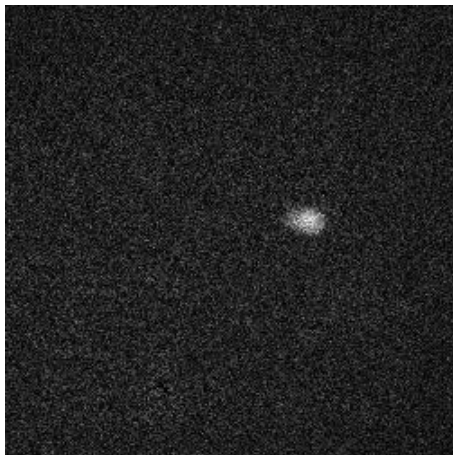


Figure 73: Raw Coherence Gated Image 2.

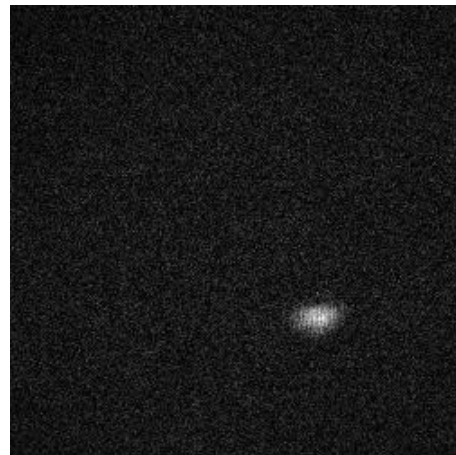


Figure 74: Raw Coherence Gated Image 3.

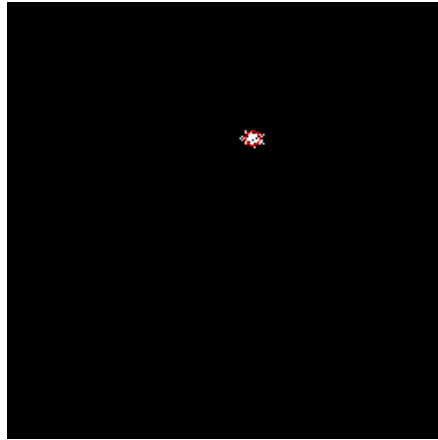


Figure 75: Binary Coherence Gated Image 1 with Centroid.

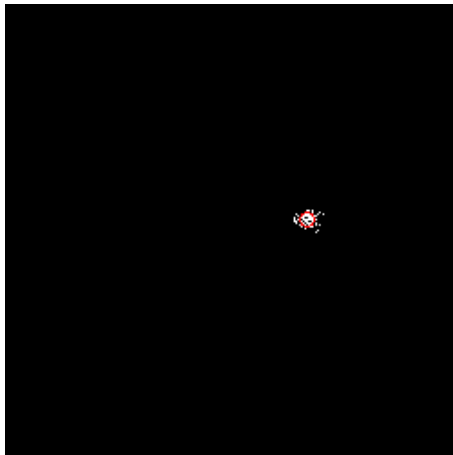


Figure 76: Binary Coherence Gated Image 2 with Centroid.

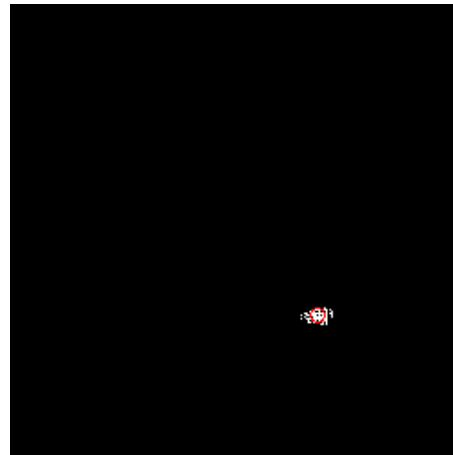


Figure 77: Binary Coherence Gated Image 3 with Centroid.

Image #	MATLAB x	MATLAB y	FPGA x	FPGA y	% error x	% error y
1	141.90	78.73	141.59	78.48	0.22%	0.32%
2	168.39	119.86	168.25	119.56	0.08%	0.25%
3	171.34	173.44	171.22	173.29	0.07%	0.09%

Table 4: Percent error for FPGA and MATLAB calculated image centroid for binary CG images 1-3 in Figures 75, 76 and 77.

5.5.1 Wavefront Reconstruction

Real-time estimation of optical wavefronts is critical for the use of adaptive optics (AO) in diverse fields such as astronomy, microscopy, and biomedical imaging. In the biomedical field, modalities such as Optical Coherence Tomography (OCT) [20] employ AO in high-speed depth resolved imaging [38] which requires processing of wavefront aberrations including the generation of Zernike coefficients. Zernike coefficients are useful not only for wavefront reconstruction but also for analysis of aberrations in an optical system.

Due to the factorial components in Equation 2.3, direct calculation of the Zernike modes requires many operations, and is impractical. Even excluding direction calculation, performing the inverse transform of the large matrix of measured data points is expensive and time-consuming. Processing in real-time requires a method that avoids direct calculation of these components and, ideally, avoids the calculation of the inverse transform needed to fit the measured data to the Zernike modes. In this section, a fast parallel method for calculating Zernike coefficients from measured phase gradients is discussed. Section 5.5.2 reviews the general modal method for obtaining coefficients via the singular value decomposition (SVD). Section 5.5.3 presents the parallel method as a restructuring of the general modal method by separating the large matrix multiplication of the singular value decomposition into inner products; two for each Zernike mode. In Section 5.5.4 hardware implementation is presented along with error analysis, timing and resource utilization.

5.5.2 Modal Method

In general, the modal method attempts to reconstruct the wavefront phase $W(x, y)$ at the sensor plane via a sum of orthogonal functions of the form

$$W(x, y) = \sum_{i=1}^j c_i Z_i(x, y) \quad (5.4)$$

where $Z_i(x, y)$ is the i th Zernike polynomial and c_i is the weighting coefficient. Note $i = 0$, the piston term (DC component), is omitted as it cannot be determined from phase derivative measurements [5]. The most common wavefront sensor, the Shack-Hartmann, typically has its subapertures arranged in a regularly spaced Cartesian grid. Zernike polynomials are typically defined in a polar coordinate system. However, they are represented in the Cartesian system to better match the Shack-Hartmann layout. This method assumes that the measured phase gradients are available over a circular pupil enclosed within a square grid. All values outside of that circle are ignored, as shown in Figure 78. For purposes of precalculation, points are generated over an evenly spaced Cartesian grid of size 250×250 pixels and converted to polar coordinates for calculation of the Zernike modes.

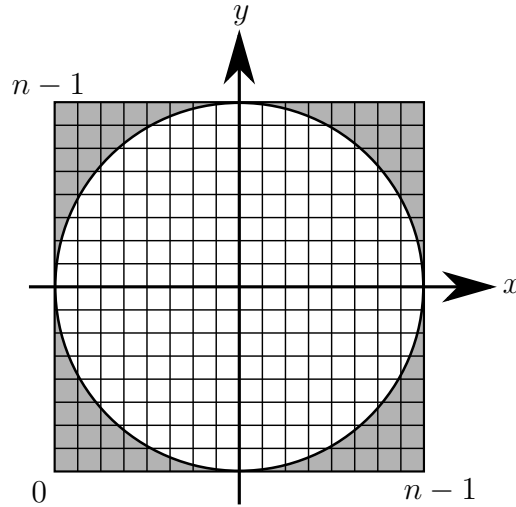


Figure 78: Circular pupil mapped into a square grid. Grey cells outside the pupil are ignored.

Common wavefront sensing methods, such as the Shack-Hartmann [52], or Laser Ray Tracing [35], provide measurements of local phase gradients at the sensor pupil, usually by division of the pupil into $n \times n$ sub-apertures. The gradient is related to Equation 5.4 via

$$S_x = \frac{\partial W(x, y)}{\partial x} = \sum_{i=1}^j c_i \frac{\partial Z_i(x, y)}{\partial x} \quad (5.5)$$

$$S_y = \frac{\partial W(x, y)}{\partial y} = \sum_{i=1}^j c_i \frac{\partial Z_i(x, y)}{\partial y}.$$

In matrix form Equation 5.5 may be expressed as

$$S_{xy} = A c \quad (5.6)$$

where S_{xy} is a vector of the measured phase gradients with alternating x and y values

$$S_{xy} = \left[\frac{\partial W(x_1, y_1)}{\partial x_1} \frac{\partial W(x_1, y_1)}{\partial y_1} \cdots \frac{\partial W(x_n, y_n)}{\partial x_n} \frac{\partial W(x_n, y_n)}{\partial y_n} \right]^T \quad (5.7)$$

A is a matrix of the partial derivatives of the Zernike modes

$$A = \begin{bmatrix} \frac{\partial Z_1(x_1, y_1)}{\partial x_1} & \frac{\partial Z_2(x_1, y_1)}{\partial x_1} & \cdots & \frac{\partial Z_j(x_1, y_1)}{\partial x_1} \\ \frac{\partial Z_1(x_1, y_1)}{\partial y_1} & \frac{\partial Z_2(x_1, y_1)}{\partial y_1} & \cdots & \frac{\partial Z_j(x_1, y_1)}{\partial y_1} \\ \frac{\partial Z_1(x_2, y_2)}{\partial x_2} & \frac{\partial Z_2(x_2, y_2)}{\partial x_2} & \cdots & \frac{\partial Z_j(x_2, y_2)}{\partial x_2} \\ \frac{\partial Z_1(x_2, y_2)}{\partial y_2} & \frac{\partial Z_2(x_2, y_2)}{\partial y_2} & \cdots & \frac{\partial Z_j(x_2, y_2)}{\partial y_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial Z_1(x_n, y_n)}{\partial x_n} & \frac{\partial Z_2(x_n, y_n)}{\partial x_n} & \cdots & \frac{\partial Z_j(x_n, y_n)}{\partial x_n} \\ \frac{\partial Z_1(x_n, y_n)}{\partial y_n} & \frac{\partial Z_2(x_n, y_n)}{\partial y_n} & \cdots & \frac{\partial Z_j(x_n, y_n)}{\partial y_n} \end{bmatrix} \quad (5.8)$$

and c is a column vector of Zernike coefficients

$$c = \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_j \end{bmatrix}^T \quad (5.9)$$

This may be solved for c , in the least-squares sense, by utilizing the pseudoinverse A^\dagger where A^\dagger is obtained via the SVD. From the SVD, $A = U_{mn} S_{nn} V_{mn}^T$ where the columns of

U_{nn} are the eigenvectors of $A A^T$, the columns of V_{nn} are orthonormal eigenvectors of $A^T A$, and S_{nn} is a diagonal matrix, which contains the singular values of A . The pseudoinverse is then, $A^\dagger = U_{nn} S_{nn}^\dagger V_{nn}^T$. Therefore, the Zernike coefficients may be obtained via

$$c = A^\dagger S_{xy} = U_{nn} S_{nn}^\dagger V_{nn}^T S_{xy}. \quad (5.10)$$

A parallel method for solving for the c_i 's in real-time is presented in the following section.

5.5.3 Parallel Modal Method

The proposed parallel method involves separating the x and y components of Equations 5.7 and 5.8. Two coefficients, c_{ix} and c_{iy} , may now be calculated solely from the x and y components, respectively, as shown in Equation 5.11

$$c_{ix} = A_{ix}^\dagger S_x, \quad (5.11)$$

$$c_{iy} = A_{iy}^\dagger S_y$$

where

$$S_x = \left[\frac{\partial W(x_1, y_1)}{\partial x_1} \quad \frac{\partial W(x_2, y_2)}{\partial x_2} \quad \dots \quad \frac{\partial W(x_n, y_n)}{\partial x_n} \right]^T, \quad (5.12)$$

$$S_y = \left[\frac{\partial W(x_1, y_1)}{\partial y_1} \quad \frac{\partial W(x_2, y_2)}{\partial y_2} \quad \dots \quad \frac{\partial W(x_n, y_n)}{\partial y_n} \right]^T, \quad (5.13)$$

$$A_{ix} = \left[\frac{\partial Z_i(x_1, y_1)}{\partial x_1} \quad \frac{\partial Z_i(x_2, y_2)}{\partial x_2} \quad \dots \quad \frac{\partial Z_i(x_n, y_n)}{\partial x_n} \right]^T, \quad (5.14)$$

and

$$A_{iy} = \left[\frac{\partial Z_i(x_1, y_1)}{\partial y_1} \quad \frac{\partial Z_i(x_2, y_2)}{\partial y_2} \quad \dots \quad \frac{\partial Z_i(x_n, y_n)}{\partial y_n} \right]^T \quad (5.15)$$

and A_{ix}^\dagger and A_{iy}^\dagger are now row vectors taken column-wise from the SVD of $\frac{\partial Z_i(x, y)}{\partial x}$ or $\frac{\partial Z_i(x, y)}{\partial y}$, respectively. In this manner, the large matrix multiplication in Equation 5.10 is decomposed into simple inner products, two for each desired mode. Finally, c_i can be obtained by

$$c_i = c_{ix} + c_{iy}. \quad (5.16)$$

Assuming the fixed grid size is known *a priori* each A_{ix}^\dagger and A_{iy}^\dagger vector may be precalculated and stored in memory. Two memory elements are required for each Zernike mode, one for A_{ix}^\dagger and another for A_{iy}^\dagger , allowing ∂x and ∂y to be processed independently, in parallel. The size of each memory element will vary directly with the number of sample points in S_{xy} and the word length of each precalculated A_{ix}^\dagger and A_{iy}^\dagger value. This method trades off long calculation times inherent in the direct method with larger memory requirements due to the precalculated vectors.

5.5.4 Hardware Implementation

Hardware implementation was undertaken using a Xilinx Spartan-6 FPGA device utilizing the Simulink System Generator suite of functions in MATLAB[®] [26]. The top level layout is shown in Figure 79 illustrating the Zernike Coefficient Processing (ZCP) blocks (labelled X-Tilt, Y-Tilt, Defocus, etc), Address Generator, dx and dy inputs and the c_i outputs. Each block accepts the address input from the Address Generator block. In each ZCP block 2 ROMs store the precalculated A_{ix}^\dagger and A_{iy}^\dagger values for the given Zernike mode. Figure 80 illustrates internal elements of the ZCP blocks. In each, the **addr** signal cycles from 0 to $(n^2/2) - 1$ and then from $(n^2/2) - 1$ back to 0 as each ROM holds half of the symmetric A_{ix}^\dagger and A_{iy}^\dagger vectors, respectively (see subsection 5.5.5). For the odd ZCP blocks the **select** signal toggles the negated signal through the mux when counting down. Each pair of gradients dx and dy is sent, in parallel, to each ZCP block. The inner product is formed via a multiply and accumulate, which is implemented using DSP slices. After the multiply and accumulate stage the c_{ix} and c_{iy} values are summed to produce the final c_i . The final c_i value is only valid once the entire n^2 input pairs are processed.

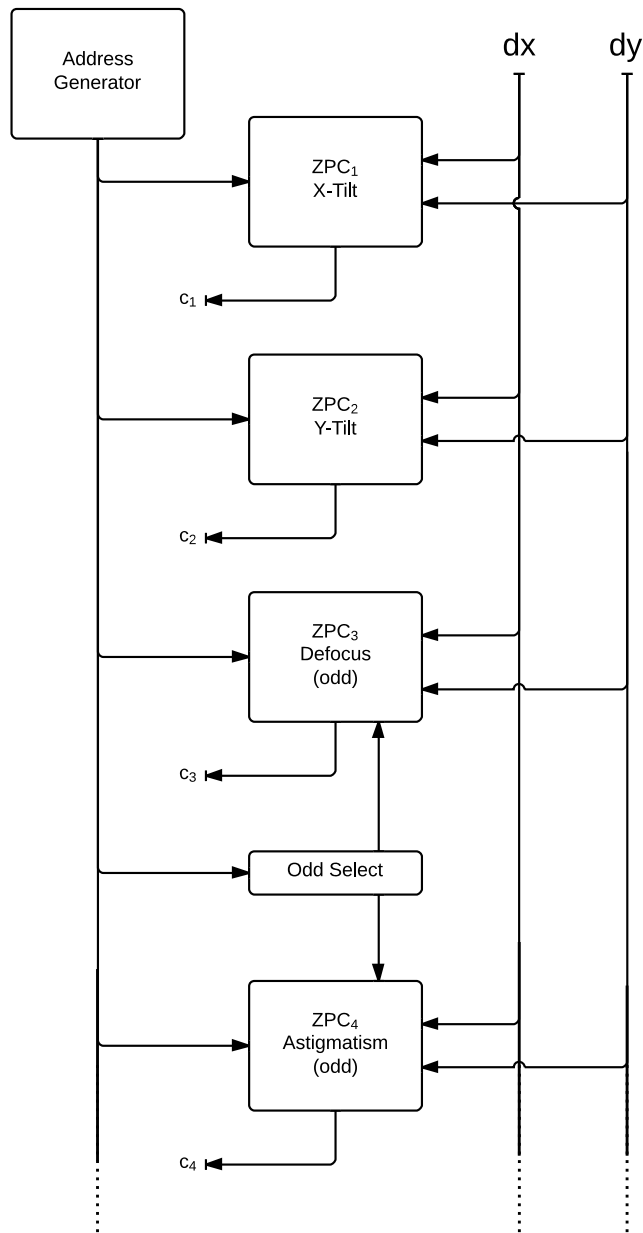


Figure 79: Block diagram for the top level view of Zernike processing blocks.

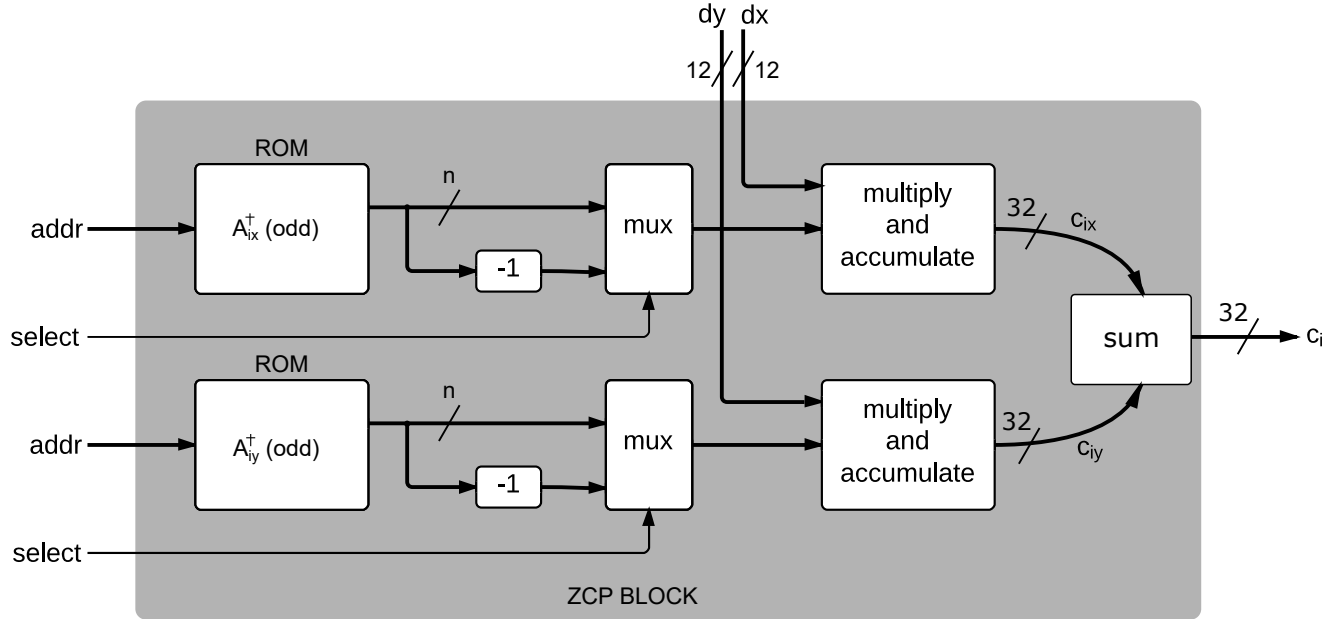


Figure 80: Block diagram for the Zernike Coefficient Processors.

5.5.5 Memory Reduction

Due to the symmetry inherent in the Zernike modes A_{ix}^{\dagger} and A_{iy}^{\dagger} also exhibit even or odd symmetry about the midpoint of the vector as shown in Figures 81 and 82. This redundancy enables storage of only half the A_{ix}^{\dagger} and A_{iy}^{\dagger} vectors in memory. One counter is used for the forward addressing of the inner product multiplications and a second counter is used for the reverse addressing. A master counter is used as a controller, enabling each counter as needed. For the A_{ix}^{\dagger} and A_{iy}^{\dagger} 's that exhibit odd symmetry a negation of each element is performed when the addressing is reversed. By exploiting this symmetry, the size of each memory element for the A_{ix}^{\dagger} and A_{iy}^{\dagger} 's is halved. This is similar to the techniques used in direct digital synthesis of symmetric waveforms.

5.5.6 Wavefront Simulation

To test the design a several wavefronts, one of which is shown in Figure 83, were generated in MATLAB. One arbitrary set of nine Zernike coefficients z_{in} is shown in Table 6. Note that the piston term is omitted as discussed in Section 5.5.2.

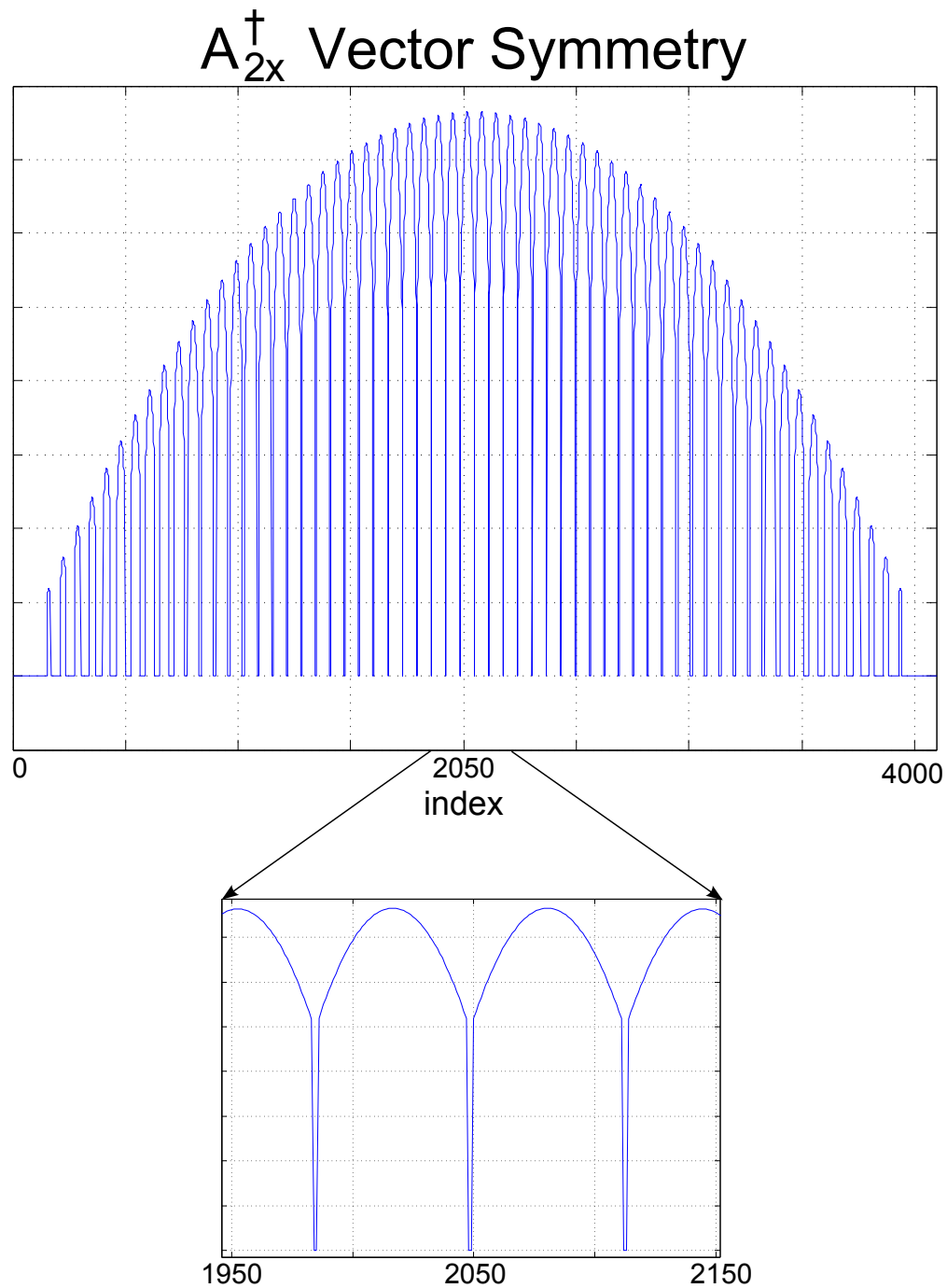


Figure 81: Symmetry in the A_{2x}^\dagger vector about the midpoint (index = 2048) is apparent. Zoomed in subimage shows the symmetry is even.

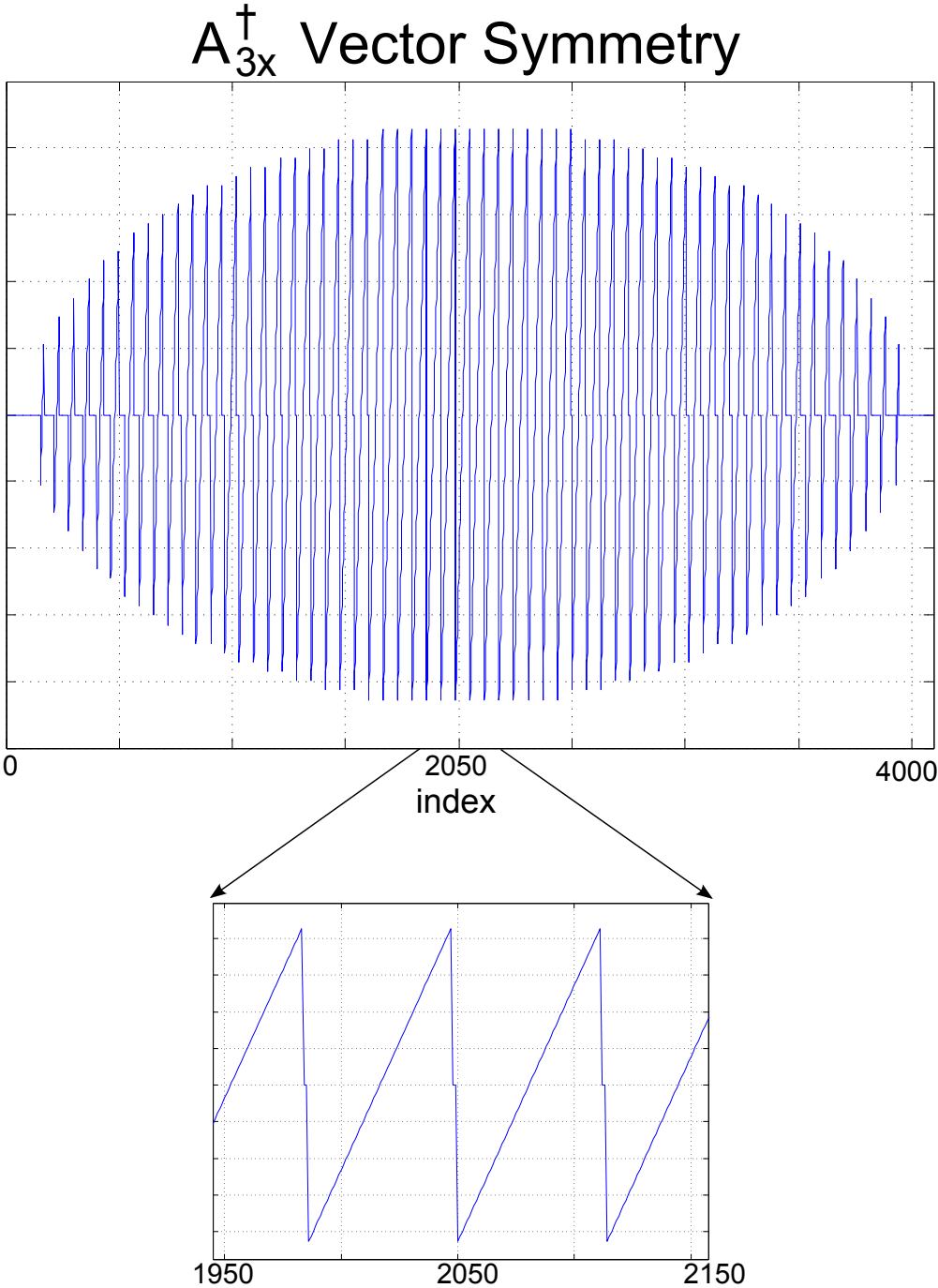


Figure 82: Symmetry in the A_{3x}^\dagger vector about the midpoint (index = 2048) is apparent. Zoomed in subimage shows the symmetry is odd.

Input wavefront Zernike coefficients								
X-Tilt	Y-Tilt	Y-Astig.	Defocus	X-Astig.	Y-Trefoil	Y-Coma	X-Coma	X-Trefoil
-0.6000	0.2000	0.7618	-0.2000	0.3000	0.1200	-0.4500	0.6780	0.9230

Table 5: Arbitrarily chosen Zernike coefficients for the simulated input wavefront.

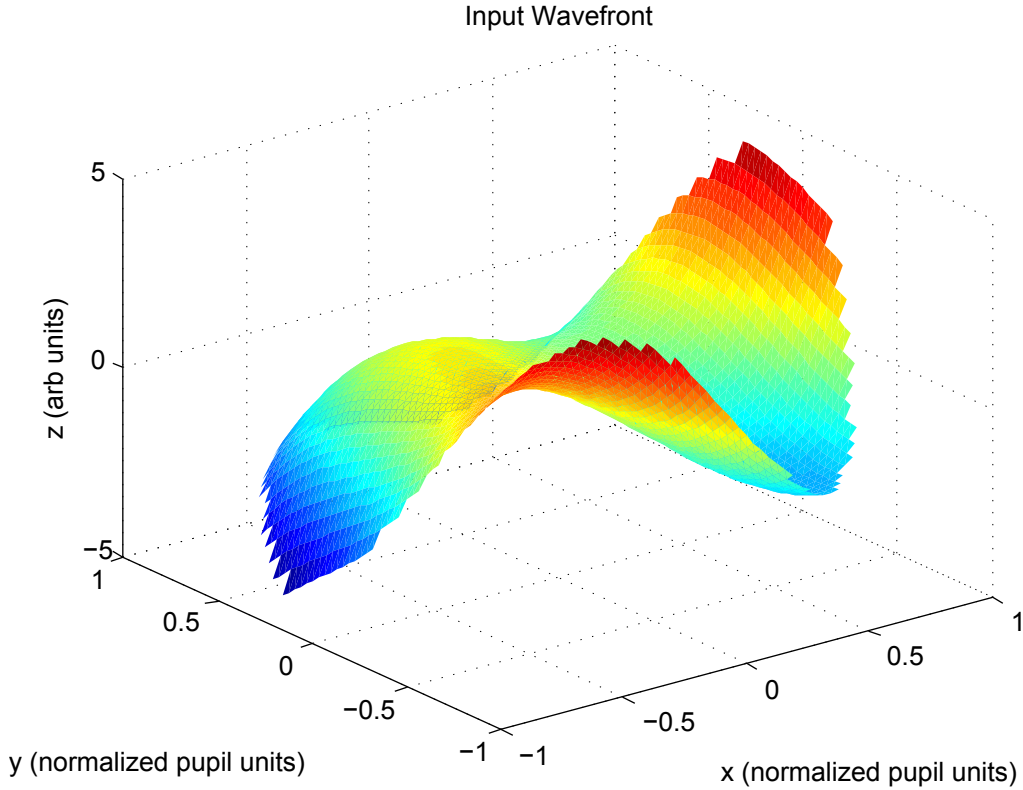


Figure 83: Simulated input wavefront

The gradient of the wavefront was calculated, serialized, and sent to Simulink to simulate outputs from the wavefront sensor. Inputs are converted from double precision format to fixed point 12 bit signed values with the binary point at 10 bits. The value of 12 bits was chosen due to our wavefront sensor having this precision, but the model can easily be scaled to accommodate other values. It is assumed that the inputs have been normalized to a value between 1 and -1 as the Zernike modes are also normalized between these values. Other input word lengths were tested and it was found that anything above 10 bits had a constant error contribution of less than 0.1%.

Zernike coefficients were calculated from the simulated measurements using word lengths of 18, 20, 22, 24, 26, 28, 30 and 32-bits for the precalculated A_{ix}^\dagger and A_{iy}^\dagger values respectively.

cin_i	c_i (24-bit)	% error	c_i (32-bit)	% error
-0.6000	-0.600558519	-0.09	-0.600044608	-0.01
0.2000	0.199260473	0.37	0.199770331	0.11
0.7618	0.761454582	0.05	0.761984587	-0.02
-0.2000	-0.200686216	-0.34	-0.200161457	-0.08
0.3000	0.299332142	0.22	0.299838722	0.05
0.1200	0.119457960	0.45	0.119983137	0.01
-0.4500	-0.450639725	-0.14	-0.450115860	-0.03
0.6780	0.677210808	0.12	0.677734435	0.04
0.9230	0.922606468	0.04	0.923126817	-0.01

Table 6: Numerical error for calculated Zernike coefficients using 24-bit and 32-bit word lengths.

These were tested to find the smallest word length (less resource utilization) that still gave acceptable accuracy (<1% error). For each word length the percent error was calculated via:

$$\% \text{ error} = \frac{|cin_i - c_i|}{cin_i} \times 100 \quad (5.17)$$

and results are shown in Figure 84. It is clear from the figure that word lengths above 22-bits ensure error of less than 1% for each of the 9 modes calculated. The difference in error between a 24-bit word length and 32-bit word length is small but not negligible and specific values are shown in Table 6. The RMS error between the reconstructed and input Zernike mode, for the entire wavefront is 6×10^{-4} .

5.5.7 Resource Utilization and Performance

Resource utilization was determined for implementation on a Xilinx Spartan-6 XC6SLX45 FPGA and a Xilinx Virtex-6 XCVLX240. For each implementation grid sizes of 8×8 , 16×16 , 32×32 , and 64×64 were considered. This is consistent with current commercial ray-tracing aberrometers such as the iTrace [57] which deliver a maximum of 256 (16×16) points. As the grid size increases, higher order Zernike modes may be more accurately calculated, but these are often unnecessary for the eye, which tends towards lower order aberrations.

The wavefront reconstruction system was implemented on the Spartan-6 FPGA during its initial development. This platform was chosen for its relatively low cost and to demonstrate

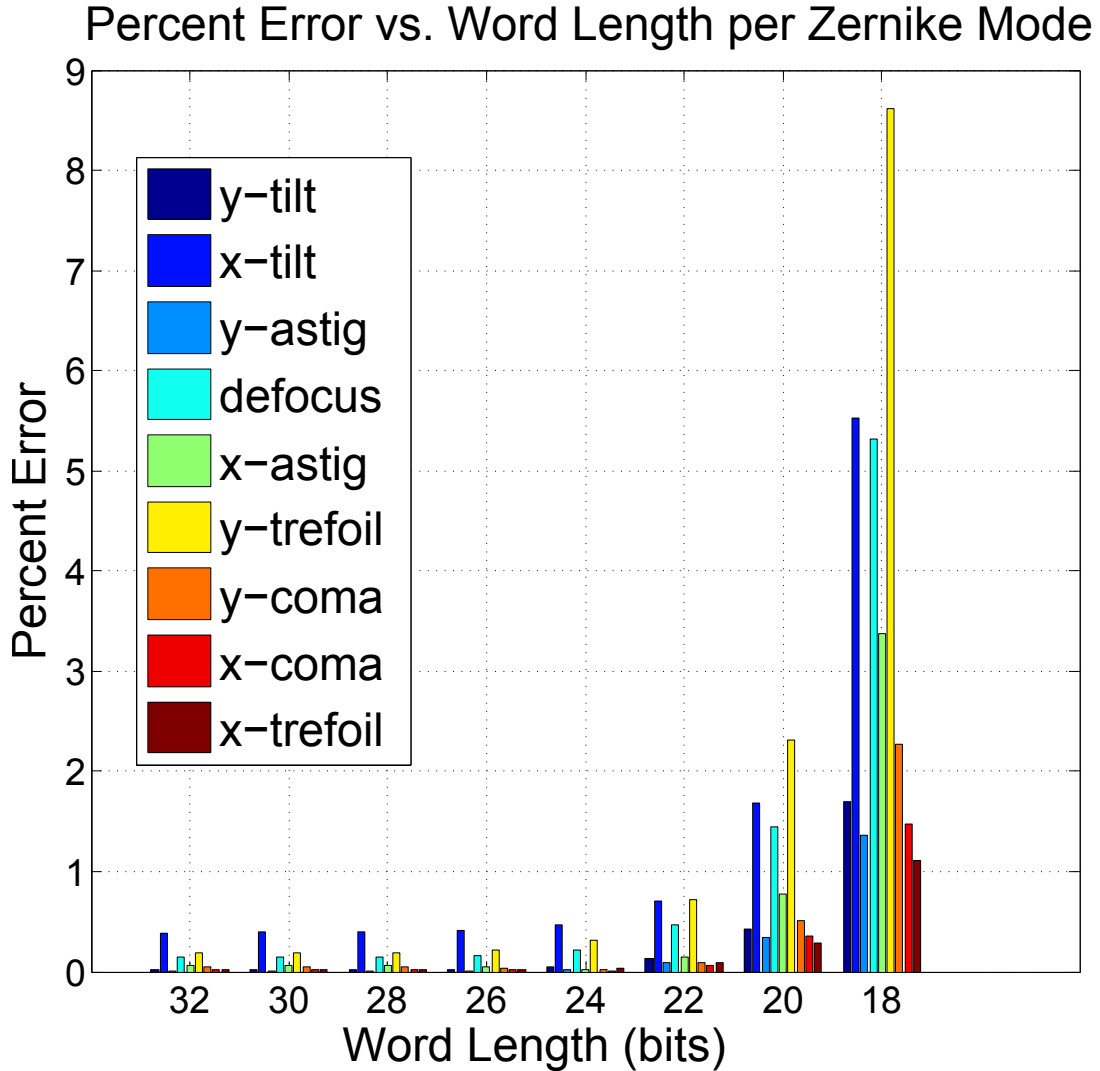


Figure 84: Percent error in calculated Zernike coefficients for given word lengths.

that this design can be implemented on a mid-grade FPGA. The final implementation was synthesised on the Virtex-6 FPGA as a part of the the complete CGLRT system. Table 7 shows the resource utilization for four grid sizes and 24-bit ROM values, while Table 8 shows the 32-bit version. The tables show that block RAM (RAMB8 and RAMB16) usage is unaffected by word length for grid sizes below 32×32 . Above that, RAMB16 usage increases dramatically. A grid size of 128×128 was attempted but was too large to fit on this device due to the system memory required for synthesis. DSP48 (specialized processing blocks with embedded multipliers) usage is independent of the grid size chosen, but will become a limiting factor as more Zernike modes are included in the calculation. Tables 7 and 8 also show that, other than a slight increase LUT usage, there is no penalty for utilizing 32-bit ROM values for grid sizes lower than 32×32 . Table 9 shows the resource utilisation

Grid size	Slice Registers	Slice LUTs	RAMB8	RAMB16	DSP48s	Clock Cycles	Duration (μs @ 100MHz)
8×8	1384	2049	18	0	28	57	0.57
16×16	1386	2060	18	0	28	239	2.39
32×32	1388	2068	0	18	28	984	9.84
64×64	1390	2083	0	54	28	4011	40.11

Table 7: Resource utilization for 24 bit ROM values compiled for a Xilinx Spartan-6 XC6SLX45.

Grid size	Slice Registers	Slice LUTs	RAMB8	RAMB16	DSP48s	Clock Cycles	Duration (μs @ 100MHz)
8×8	1832	4138	18	0	28	57	0.57
16×16	1834	4144	18	0	28	239	2.39
32×32	1836	4160	0	18	28	984	9.84
64×64	1838	4164	0	72	28	4011	40.11

Table 8: Resource utilization for 32 bit ROM values compiled for a Xilinx Spartan-6 XC6SLX45.

for the Virtex-6 FPGA (only 32-bit values were synthesised).

This method is shown to compare favourably with other methods of wavefront reconstruction. Rodríguez-Ramos, et. al, utilized a forward and inverse 2D-FFT method for grid sizes from 8×8 to 256×256 [46]. With this method, the Fourier coefficients from the forward 2D-FFT portion of the recoverer takes approximately $1.67\mu\text{s}$ compared to $0.57\mu\text{s}$ by the CGLRT architecture. These results are comparable due to both methods utilising the modal method of wavefront reconstruction. The generation of Fourier coefficients in the Rodríguez-Ramos method is analagous to the generation of Zernike coefficients in the method described in this thesis. Once the coefficients have been generated both utilise a linear fit to reconstruct the measured wavefront.

Saunter, et. al, [50] implemented an FPGA based wavefront reconstruction scheme (architecture details not provided), producing Zernike coefficients, and reporting a frame rate of $1.19\mu\text{s}$ for 64 Shack-Hartmann spots. The proposed method achieves these results in $0.57\mu\text{s}$.

5.6 Conclusion

This chapter began with a brief overview of FPGAs in general, and how they fit into the larger space of programmable logic devices. The unique workflow for FPGA development was also discussed

Grid size	Slice Registers	Slice LUTs	RAMB18	RAMB36	DSP48s	Clock Cycles	Duration (μs @ 200MHz)
8 \times 8	1656 (1%)	2557 (9%)	18 (2%)	0	28 (4%)	57	0.285
16 \times 16	1660 (1%)	2561 (9%)	18 (2%)	0	28 (4%)	239	1.195
32 \times 32	1664 (1%)	2566 (9%)	18 (2%)	0	28 (4%)	984	4.92
64 \times 64	1668 (1%)	2572 (9%)	0	36 (9%)	28 (4%)	4011	20.06

Table 9: Resource utilisation for 32 bit ROM values compiled for a Xilinx Virtex-6 XC6VLX240T-1FFG1156.

Grid size	Slice Registers	Slice LUTs	RAMB8	RAMB16	DSP48s	Clock Cycles	Duration (μs @ 100MHz)
8 \times 8	1141	XX	XX	4	6	162	1.62
16 \times 16	1667	XX	XX	4	6	412	4.12
32 \times 32	2855	XX	XX	4	15	1304	13.04
64 \times 64	4613	XX	XX	16	15	4516	45.16
128 \times 128	8135	XX	XX	64	25	17084	170.84

Table 10: Resource utilisation for 2D-FFT phase recovery circuit, 8 bit inputs, compiled for a Xilinx Virtex-4 XC4VSX35. Taken from [46]. “XX” indicates that slice LUTs and RAMB8 utilisation was not reported.

Grid size	Slice Registers	Slice LUTs	RAMB8	RAMB16	DSP48s	Clock Cycles	Duration (μs @ 100MHz)
8 \times 8	1650	XX	XX	4	10	162	1.62
16 \times 16	2302	XX	XX	4	10	412	4.12
32 \times 32	3992	XX	XX	8	22	1304	13.04
64 \times 64	6472	XX	XX	24	22	4516	45.16
128 \times 128	11326	XX	XX	24	36	17084	170.84

Table 11: Resource utilisation for 2D-FFT phase recovery circuit, 16 bit inputs, compiled for a Xilinx Virtex-4 XC4VSX35. Taken from [46]. “XX” indicates that slice LUTs and RAMB8 utilisation was not reported.

with an emphasis on the unique ability of the MATLAB based System Generator environment to simplify the design process. Next, an overview of the CGLRT FPGA system architecture was provided, followed by an in-depth discussion of the various subsystems. Special attention is paid to the novel parallel wavefront recovery subsystem used to generate Zernike coefficients from the measured wavefront slopes. Coefficients were fit to the phase gradients, in the least-squares sense, via a precomputed singular value decomposition. Symmetries in the precomputed vectors were exploited to halve the required memory. With this FPGA architecture in hand, the discussion may now continue to the final development of the CGLRT instrument.

6 The CGLRT Instrument

Chapter 4 presents the results of the non-real-time, Labview based LRT and CGLRT implementations. While these implementations were successful in demonstrating the ability of CGLRT to perform depth-resolved wavefront sensing it is important to combine the optical WFS system with the real-time processing FPGA architecture discussed in Chapter 5. With the addition of the FPGA based processing the CGLRT instrument may be realised in form which enables all of the data acquisition, system control, image processing, and wavefront reconstruction on the embedded system. The end goal of the CGLRT instrument is to be able to integrate into a high-speed OCT type imaging system, and the FPGA computing hardware provides the speed and resources that such an integration would require. This also opens up the possibility of combining the CGLRT and OCT processing systems on the same FPGA, given sufficient resources on the chip.

A photograph of the CGLRT optical setup is shown in Figure 85. This is functionally identical to the layout in Figure 43 (the placement of the camera has shifted, but this has no effect). Each component of the setup is discussed in detail in the following sections, so here a brief summary will suffice. Starting at the right, the optical source OS (675 nm SLD, including a collimating lens) sends a beam with an approximate diameter of 5 mm to the beam splitter BS1. Here the beam is split into the reference arm (blue) and the object arm (red) of the interferometer. The reference arm is directed to mirror M1 which sends the beam to the PZT mounted reference mirror RM. After reflecting off of RM the beam retraces its path to M1 and passes through BS1. At BS3 the reference arm recombines with the object arm and is incident upon the camera. In the object arm, after passing through BS1, the beam diameter is reduced to approximately 500 μm by lenses L1 and L2. The beam is then focused onto the galvo-scanning mirrors GM by lens L3. After traversing the GM lens L4 collimates the beam and sends it to beam splitter BS2. Here the beam interrogated the multi-layer object MLO which consists of lens L5, two microscope slides, and a mirror. After returning from the MLO the beam is deflected by BS2 to lenses L6 and L7 which form an image of the reflected spot in the pupil plane of L5. Lens L8 focuses the scanning beam, through beam splitter BS3, onto the detector of the camera. If the optical path lengths of the reference and object arms are matched, interference will occur between the two beams.

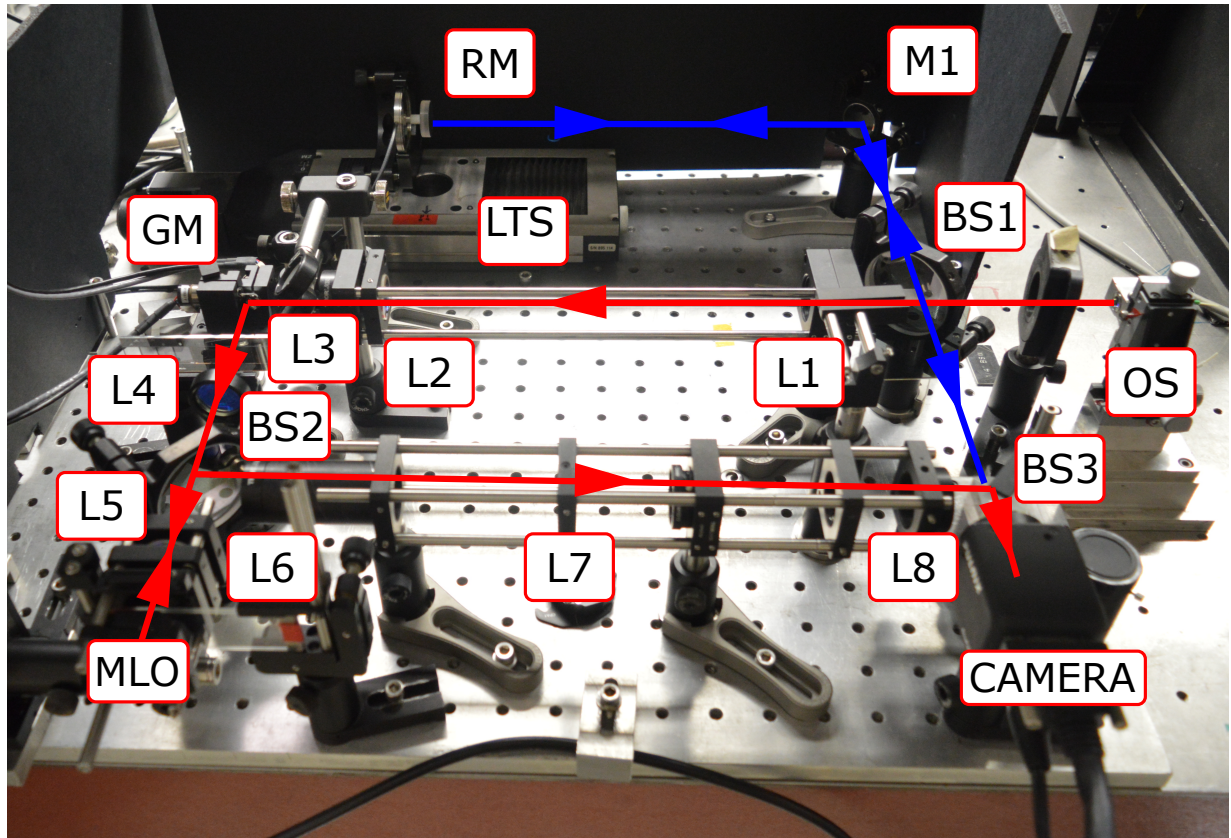


Figure 85: Photograph of the CGLRT optical layout. OS, optical source including a collimating lens; BS1, BS2, BS3, plate beam splitters; L1-L8, achromatic doublets; GM, galvo-scanning mirror; RM, reference mirror; MLO, multi-layer object composed of object mirror OM and 2 glass microscope slides; M1, mirror; CAMERA, CCD camera.

The FPGA system is based around the Xilinx ML605 development board, shown in Figure 86. The centre-piece of the board is the Virtex-6 FPGA but the board contains many other integrated peripherals including programmable DIP switches, and LCD display, 512 MB of DDR3 onboard RAM, ethernet, USB-JTAG, VGA, and 2 FMC expansion slots for connection Mezzanine cards. The FPGA is programmed directly from Xilinx’s ISE software on the PC via the USB-JTAG connection. A photograph of the ML605 board, take from [64] is shown in Figure 87.

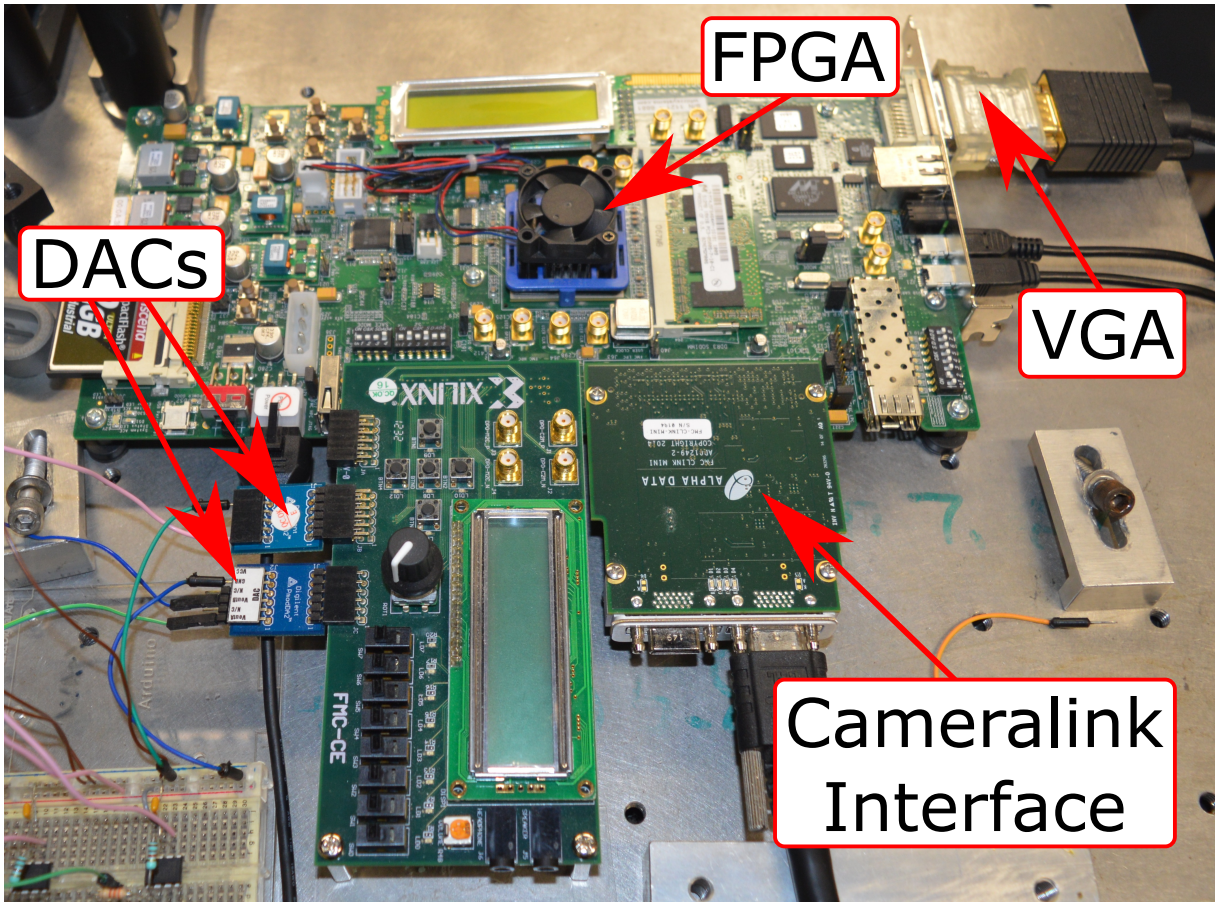


Figure 86: ML605 based FPGA system.

As the image shows, there are many components on the ML605 board, so only those that are applicable to the CGLRT instrument will be discussed here. Item 1) is the Xilinx Vertex-6 FPGA, model: XC6VLX240T-1FFG1156. The FPGA itself is not visible in the image as it is obscured by the cooling fan. Item 6) is the JTAG connector (mini-USB), used to download FPGA program file from the PC. Item 7a), on the back-side of the board, is the 200 MHz oscillator used as the main system clock. Item 14) is the video DVI connector. This was utilised in the experimental setup to drive a standard VGA monitor to view the output of the camera. The video interface is supported by a Chrontel CH7301C-TF video codec which interfaces to the FPGA. While many advanced features are available on the Chrontel chip, for this implementation, a simple VGA driver was implemented on the FPGA, which was sufficient. Items 19) and 20) are the two FMC (FPGA mezzanine card) connectors. FMC is an ANSI standard, developed by FPGA manufacturers [69] which allows for connecting a variety of peripherals to the main board. In this setup, FMC connector 20) was used to interface with the Cameralink peripheral that connects to the high-speed camera. FMC connector 19) was used to interface to a Xilinx's FMC-CE I/O board.

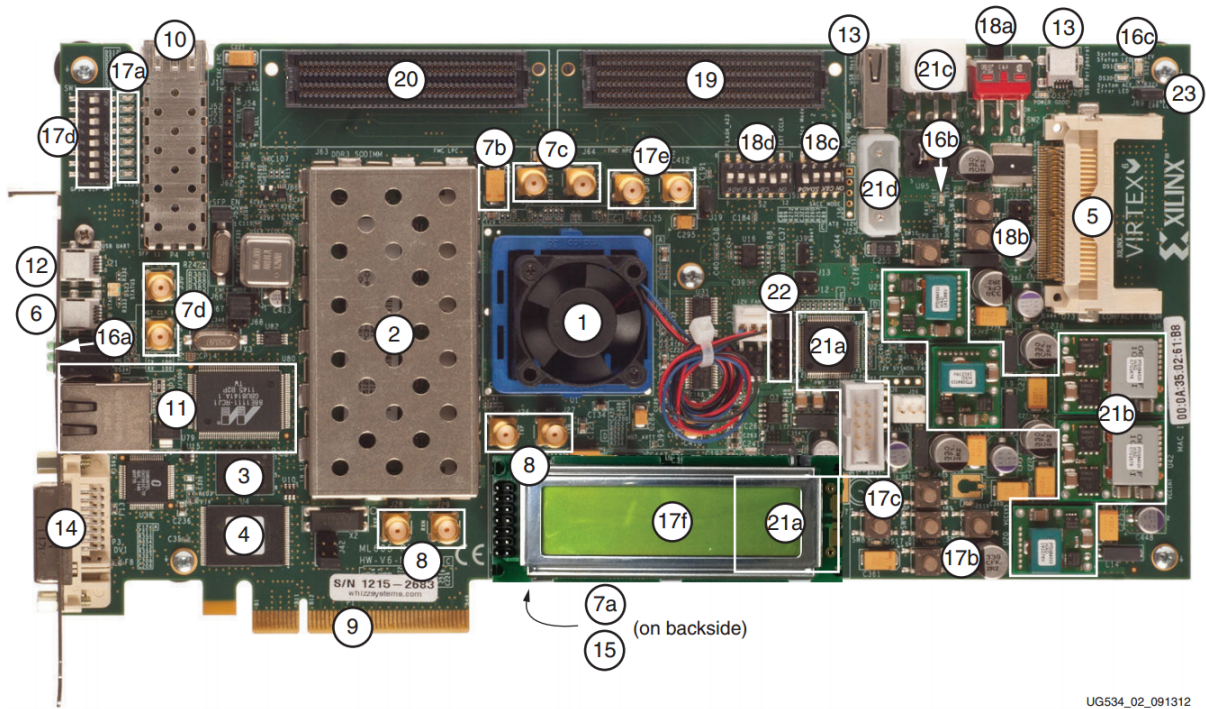


Figure 87: ML605 board. Image taken from [64]

An image of the FMC-CE board, taken from [68], is shown in Figure 88. As shown, the board is equipped with 4 SMA connectors, 5 push-buttons, an LCD display, a rotary knob, 8 slide-switches and associated LEDs, 2x12-pin PMOD connectors and 1x6 pin PMOD connector. For the CGLRT instrument, the central push-button was used to initiate a scan.



Figure 88: A Xilinx FMC-CE expansion board taken from [68].

The PMOD is an interface specification from Digilent used to connect low-frequency, low-power peripherals. For the CGLRT instrument two PMOD interfaces were used to connect three DACs to the ML605 host board. The three DACs are contained on two PmodDA2 boards from Digilent. Each PmodDA2 board contains two 12-bit digital-to-analog converters (Texas Instruments DAC121S101). These communicate with the FPGA using serial peripheral interface bus (SPI).

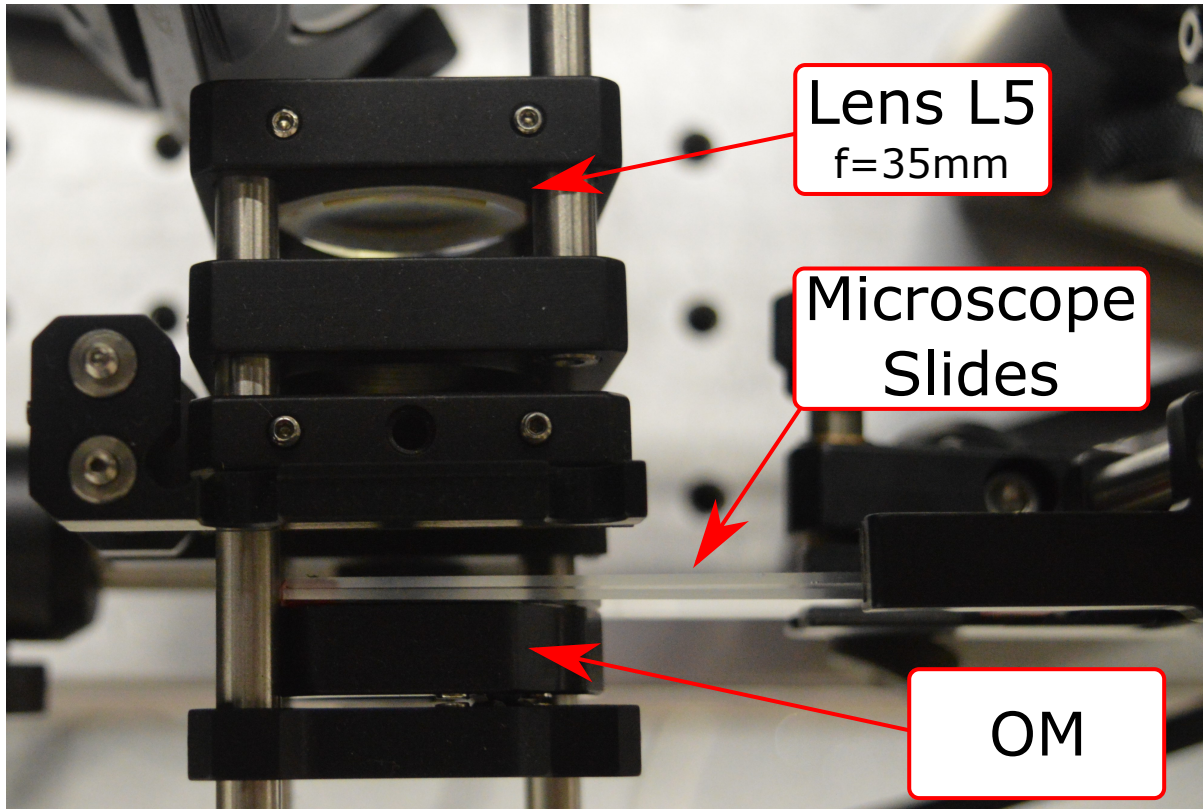


Figure 89: The Multi-layer Object (MLO) composed of lens L5 ($f = 35 \text{ mm}$), two microscope slides and object mirror OM.

Figure 89 shows a photograph of the multi-layer object (MLO) as installed in the CGLRT instrument. Lens L5 ($f=35 \text{ mm}$) imitates the lens of the eye. Two microscope slides form four minimally reflective surfaces while mirror OM form a fifth highly reflective surface.

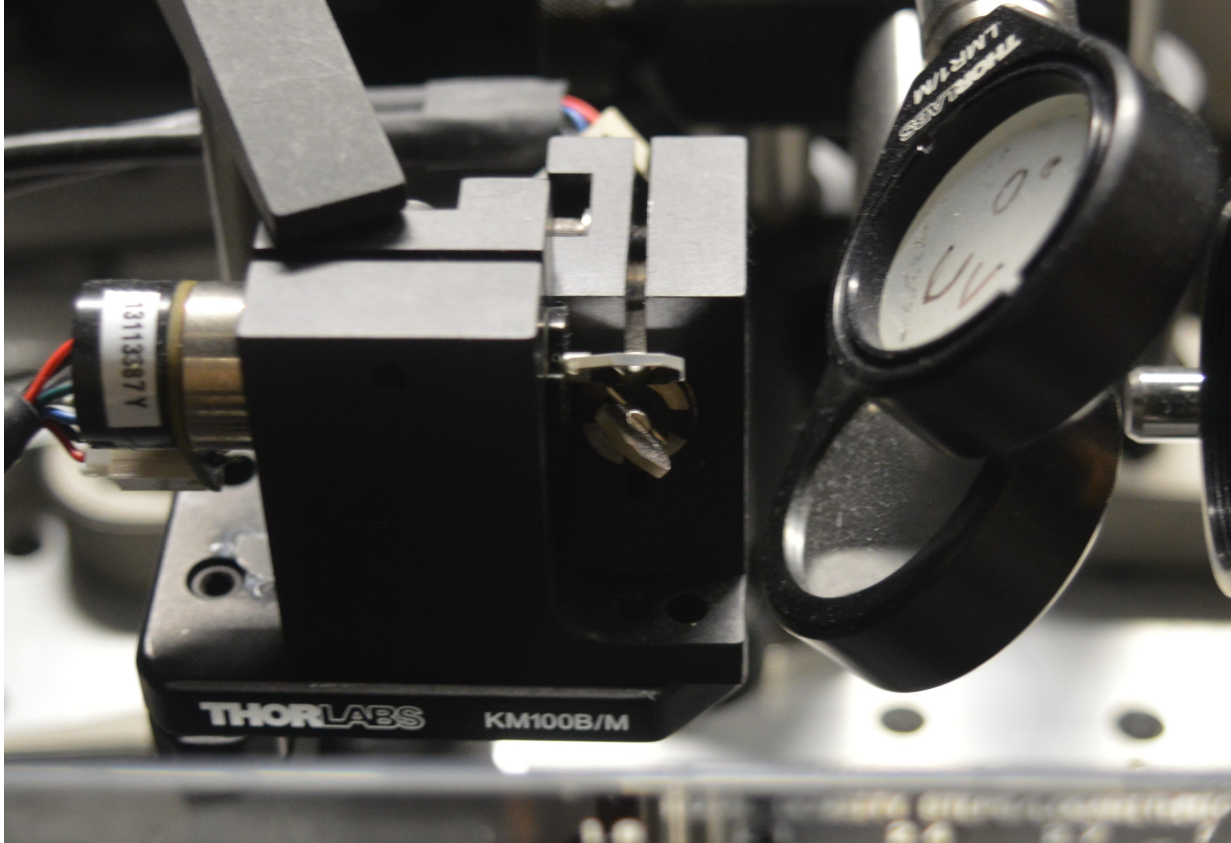


Figure 90: Close-up of the galvo-scanning mirror GM. Approximate separation between the x and y scanning mirrors is 10 mm.

The galvanometer scanning-mirror GM is shown in Figure 90. The GM consists of a Cambridge Technology MicroMaxTM 673 Series x - y scanning mirror and associated driver electronics. Each scanning mirror has a position output scale factor of 1V/deg and a maximum frequency of 500 Hz. The mirror motor housing was secured to the optical bench via Thorlabs a beam-splitter mount. This provided a suitable adjustable mount for the mirrors while also being stable.

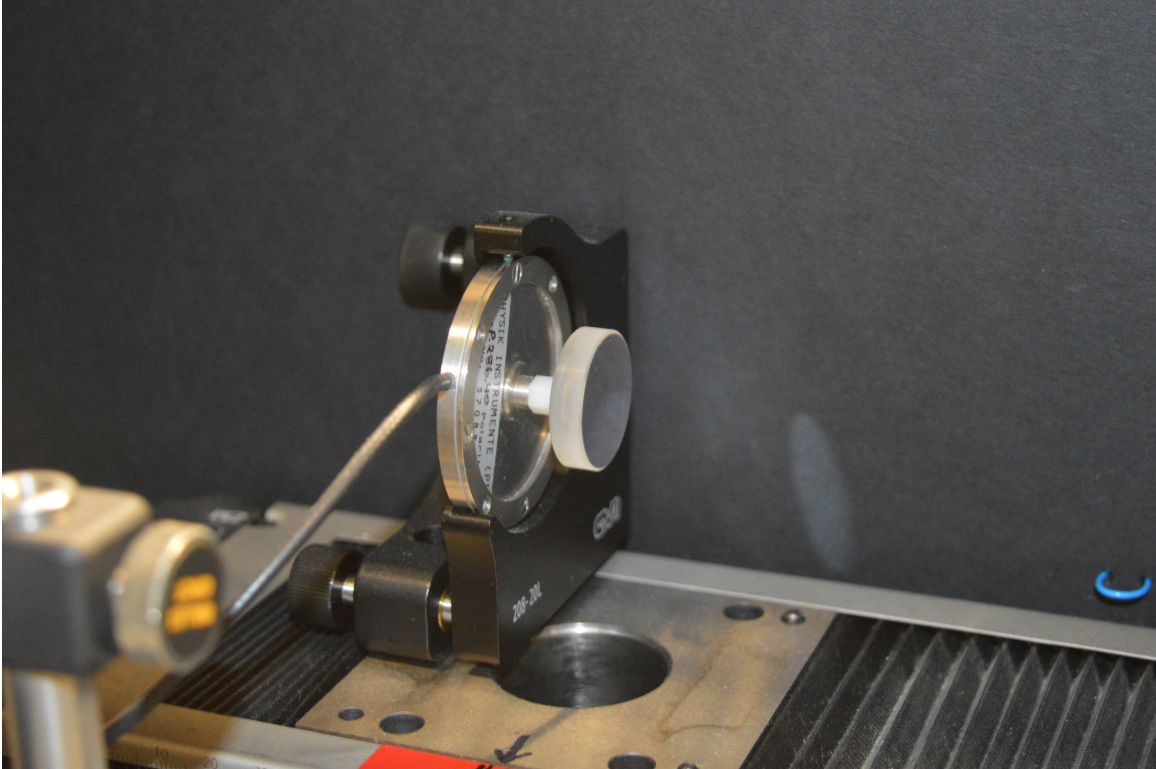


Figure 91: PZT mounted reference mirror RM.

The piezo mounted reference mirror RM is shown in Figure 91. The mirror is bonded directly to the PZT stack (silver disk). The PZT element is a Physik Instrumente (PI) P-286.40 which has a translation distance of $100\ \mu\text{m}$ with a control voltage range of 0 to $-1000\ \text{V}$. Obviously, this is much too large a voltage to be provided by low-voltage electronics so an amplifier is utilised. PI amplifier E-463 HVPZT-Amplifier is shown in Figure 92. This is a three-channel amplifier, only one channel of which was needed. It provides a manually adjustable DC-offset knob along with a control input voltage BNC connector. Control voltages were provided to the PZT amplifier by the DAC outputs of the FPGA board as shown in Figure 93.



Figure 92: PI E-463 high-voltage PZT amplifier.

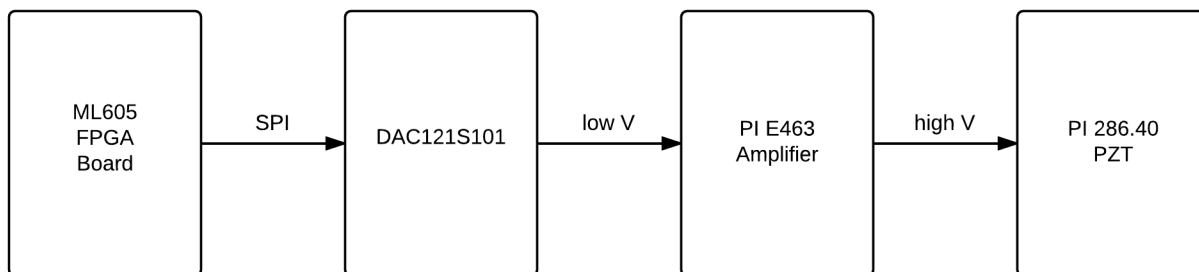


Figure 93: PZT control signal block diagram. The ML605 FPGA board sends SPI control signals to the DAC board, which is connected to the ML605 via mezzanine and PMOD interfaces. The low voltage DAC output is amplified by the PI E463 high-voltage amplifier and sent to the PZT.

6.1 LRT vs. CGLRT Results

On a base level it is critical to show that CGLRT and standard LRT agree when measuring the same wavefront. Chapter 4 shows LRT and CGLRT performing wavefront measurements (not in

real-time) for differing setups. So a first step in validating the FPGA based CGLRT instrument was to compare the results of LRT versus CGLRT for the same, single-layer target. To this end the setup in Figure 85 was modified slightly by removing the two microscope slides from the MLO, thus creating a single-layer target, consisting only of the strongly reflecting object mirror OM.

The operation of the instrument was limited to that of a standard LRT by blocking the reference beam of the interferometer between BS1 and M1. In this way only the scanning beam was present to perform the measurement. The FPGA processor was set to only include the first of the four PSI images, setting the remaining three images to zero. This allows the rest of the FPGA architecture to be untouched. By unblocking the reference beam the instrument is returned to CG mode. Blocking the reference beam does not affect the intensity of the scanning beam so the binarization threshold can remain unchanged.

To begin with, the noise floor of the system was tested by disabling galvo-scanner GM and taking 10 stationary scans. In this way the beam was not articulated and the results will only be due to the inherent noise in the system including vibrations, air currents, detector noise, etc. Stationary scans were taken in standard LRT mode and CGLRT mode. Zernike modes were calculated in real-time by the FPGA architecture. Figure 94 shows the noise floor results. The blue triangles indicate the Zernike coefficients calculated for standard LRT (no CG). Error bars indicate the standard deviation of each Zernike mode. The red circles indicate the CGLRT results. As expected the CGLRT data is somewhat noisier due to using the fringes to calculate the spot position but the error is still well within acceptable limits ($< 0.02\mu m$ for most modes). The total RMS error for all nine Zernike modes in LRT mode was 0.0017 microns while the total RMS error in CGLRT mode was 0.0074 microns. Fringe noise is most likely induced by air currents which produce random fluctuations over the time that the four fringe images are acquired. These were minimized by building a rudimentary enclosure around the optical setup.

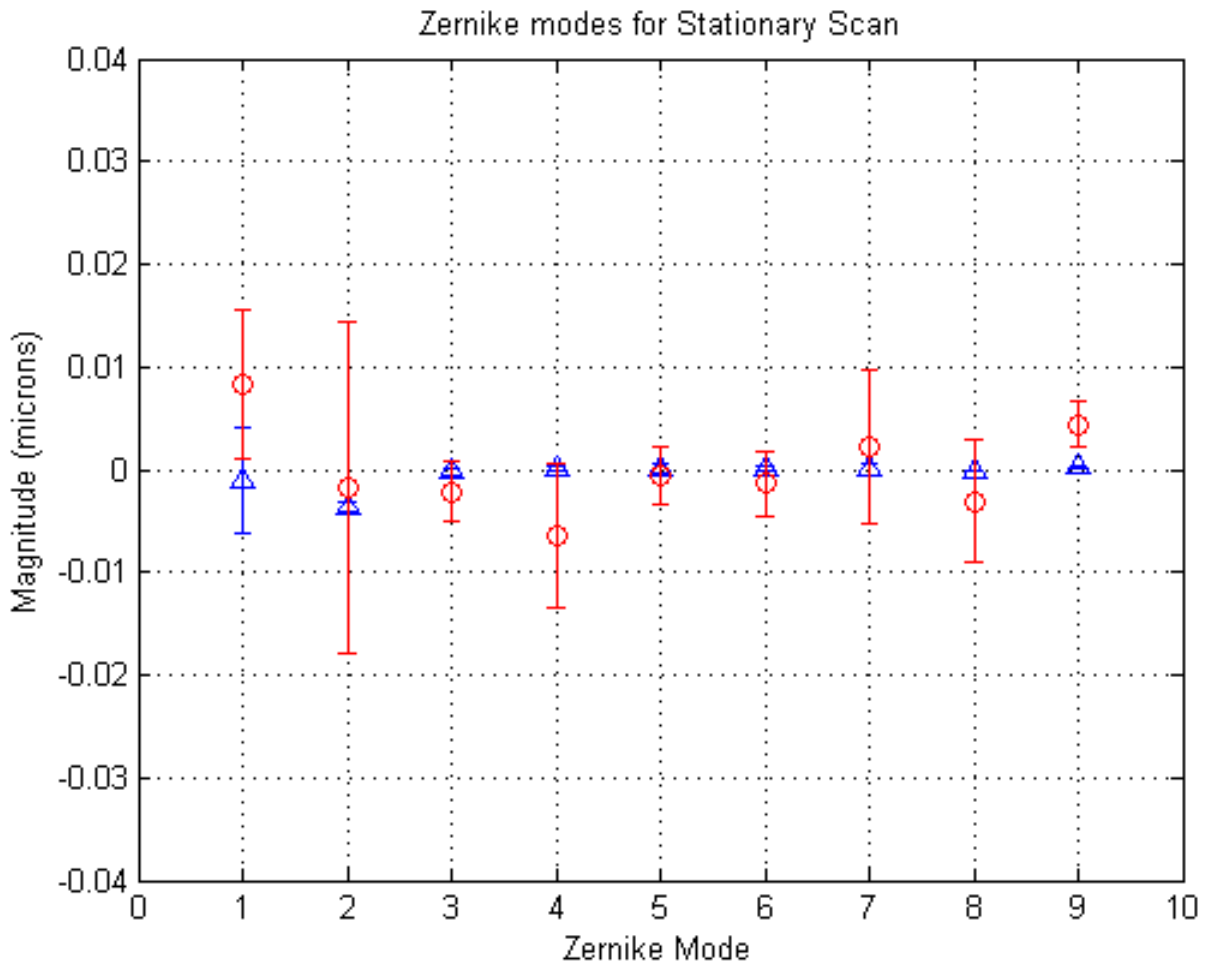


Figure 94: Noise floor results comparing LRT and CGLRT. Blue triangles indicate standard LRT while red circles indicate CGLRT. CGLRT results are noisier due to calculating the spot location via modulated fringes.

Next, single scans are performed in both LRT and CGLRT modes with the galvo-scanner GM re-enabled and a single-layer target. As noted above, the reference beam is blocked for LRT mode and the full interferometer is utilised in CGLRT mode. An 8×8 scan is performed in each case. Figure 95 shows the results. As before, the blue triangles denote the standard LRT results while the red circles denote the CGLRT results. The reported Zernike modes between the two methods agree to less than one micron, confirming that CGLRT agrees with standard LRT. Lastly, Figure 96 shows the difference in each Zernike mode between CGLRT and LRT. The largest difference in a single mode is approximately $-0.022 \mu\text{m}$.

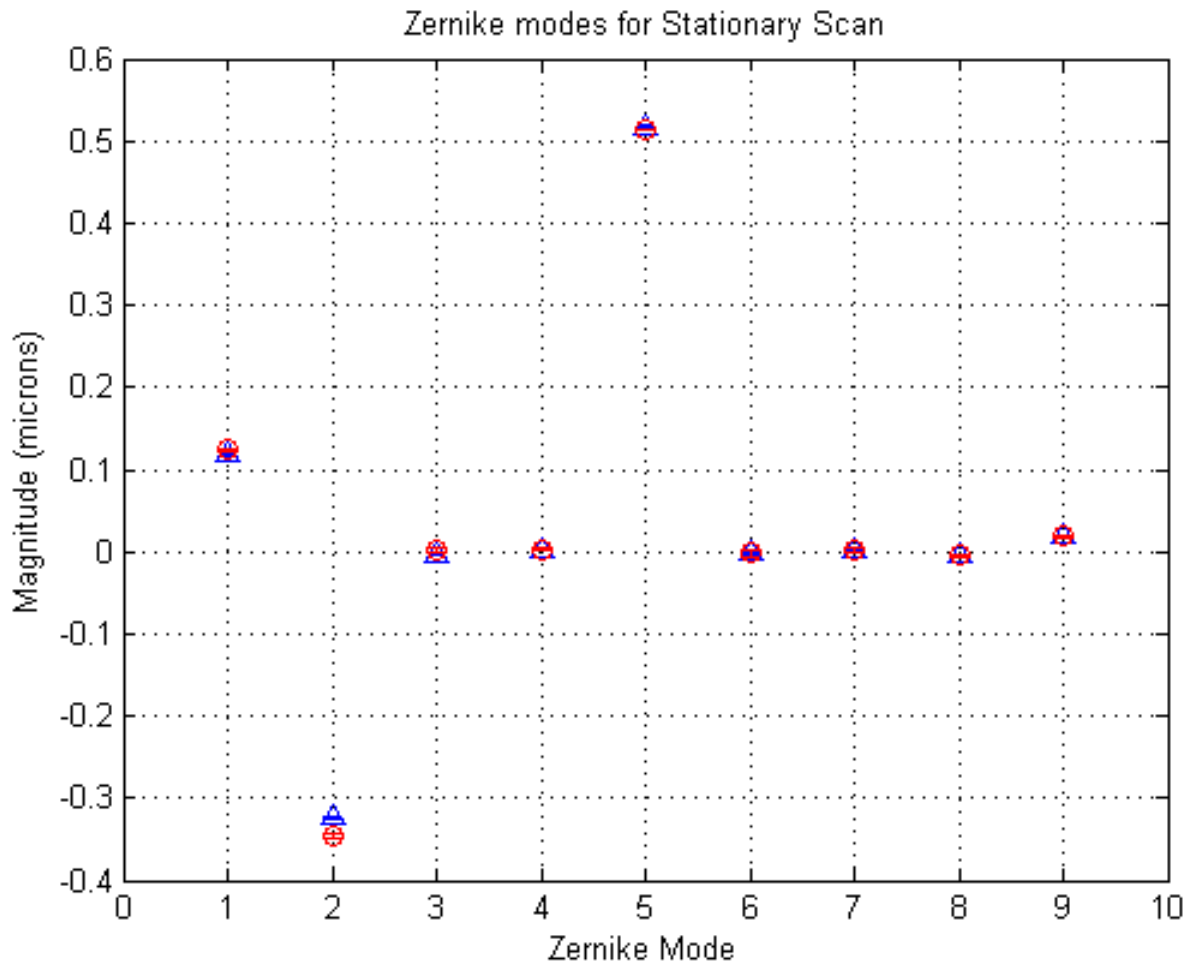


Figure 95: Comparison of standard LRT vs. CGLRT for a single layer target. Zernike modes from standard LRT are indicated by the blue triangles whereas Zernike modes from CGLRT are indicated by red circles.

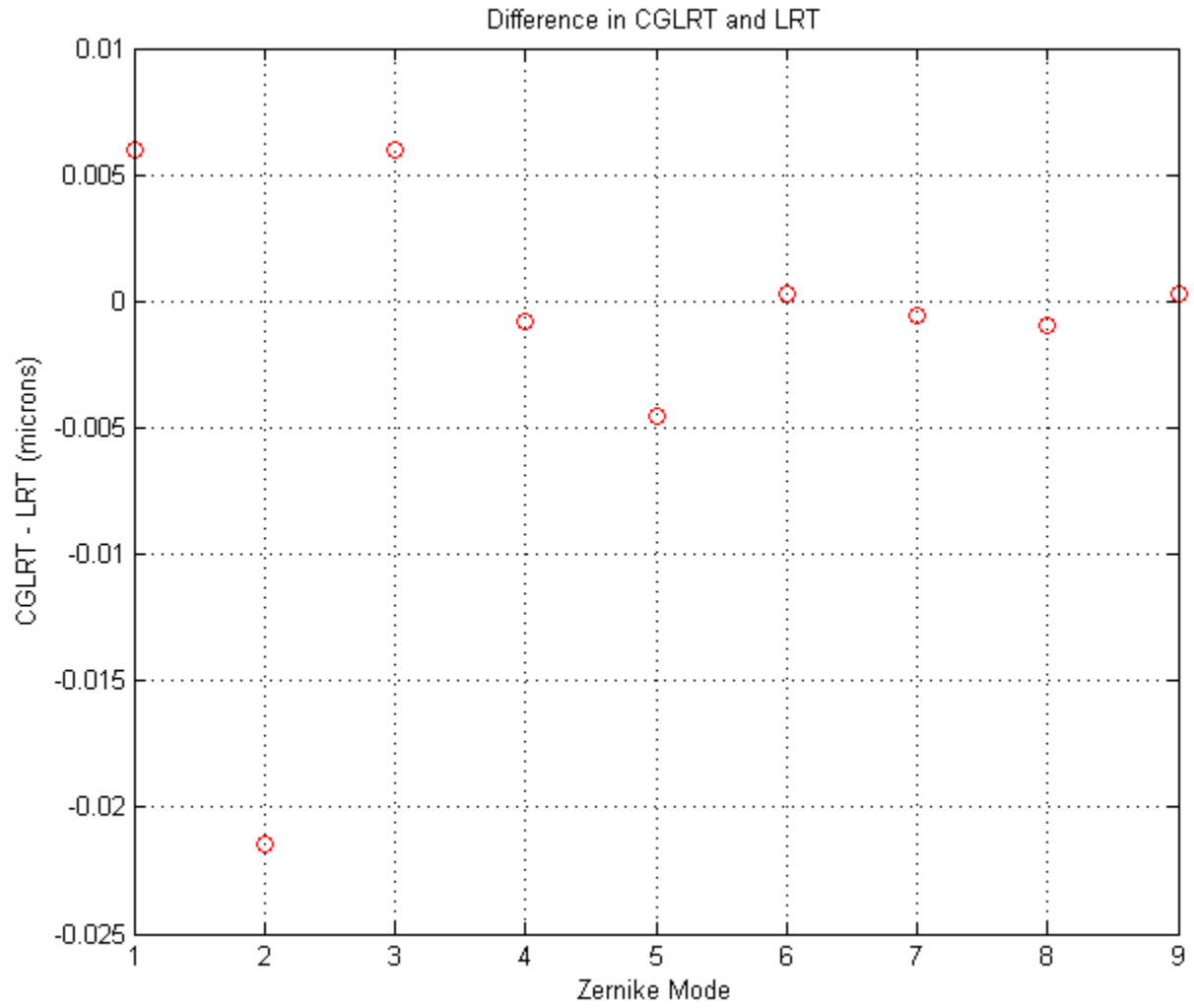


Figure 96: Difference in Zernike modes reported by standard LRT vs. CGLRT for a single layer target.

6.2 CGLRT Results

After it was shown that the CGLRT instrument agrees with standard LRT for a single-layer target, it remains to show the ability of the instrument to perform wavefront sensing for a multi-layer object utilising the FPGA platform. To this end, the microscope slides as shown in Figure 89 were reinserted into the MLO. Five separate measurements were performed, each with a grid size of 8×8 , one for each surface in the MLO. Each step of a measurement was performed as follows:

1. The focal point of the interrogating beam is adjusted to be between surfaces P5 and P4 (see Figure 44). This focus is held for all measurements.
2. The pathlength of the linear translation stage (LTS) was adjusted to obtain interference fringes for a single surface in the MLO.
3. An x, y scan is initiated by manually depressing a button on the FPGA board.
4. After the scan is performed the resulting Zernike coefficients are noted. The pathlength is then adjusted again to match the optical path length (OPL) for the next surface.

In this manner all five surfaces of the MLO are interrogated. All of the measurements are performed in the presence of the bright primary reflection from the object mirror (OM).

Figure 97 shows a screen shot of the spot images on the surface of the camera. The screen shots were obtained by photographing the display monitor as the FPGA does not provide a direct method for saving the output images. The bright primary reflection from P5 is clearly evident in the centre of the image. The spots from P3 and P4 can be seen, superimposed upon one another to the left of the P5 spot, whereas the P1 and P2 spots can be seen to the right. The spots from each microscope slide are overlapping because the slides are so thin that the deviation is small when the beam is interrogating the MLO paraxially.

Figure 98 shows a screen shot with the interrogating beam entering the MLO off-centre. In this manner the deviation between the returned beams is greater and the individual reflections from each surface are visible.

Figures 97 through 103 were obtained by photographing the VGA monitor used to view the output of the camera sensor. This was necessary as the FGPA architecture does not have a way to directly access the images as they are processed. Also, the small red square shown in Figures 100 through 103 is an artifact of the VGA driver circuit used to show the centroid of the current image. This is only valid for non-CG images as the CG-images have all of the stray light included in the image. This was used for validation purposes only and may be ignored.

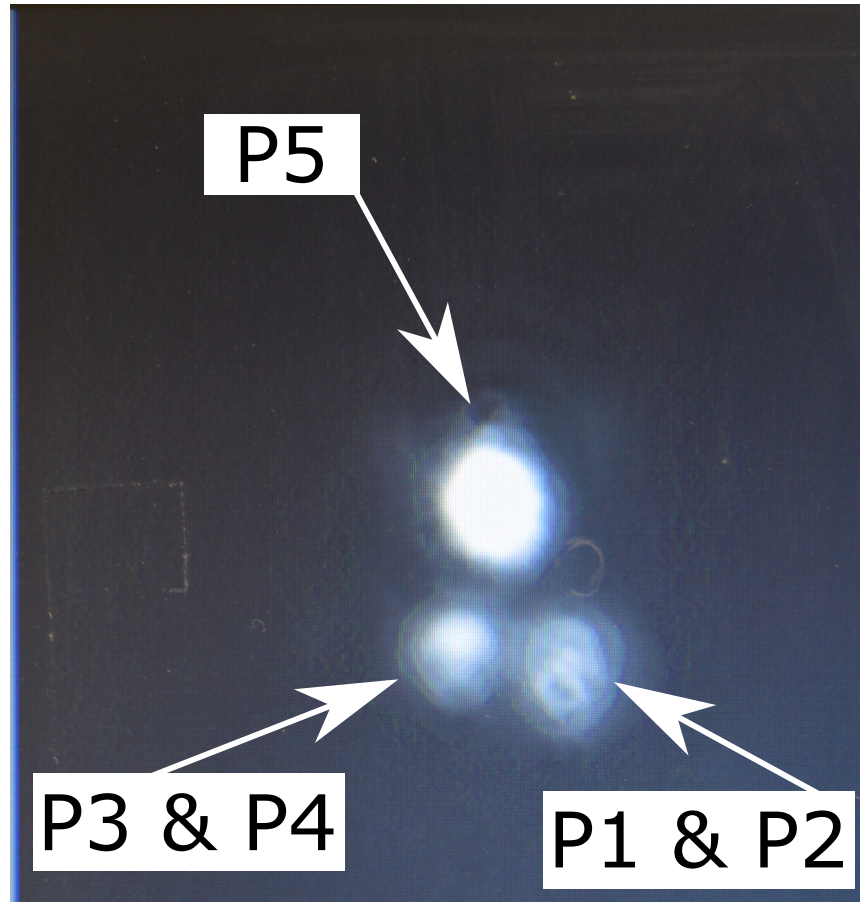


Figure 97: Screen shot of reflections from the MLO. The strong reflection from object surface P5 (object mirror) is evident in the centre of the image. Reflections from surfaces P1 and P2 (front and back of the front microscope slide) and surfaces P3 and P4 (front and back of the rear microscope slide) are superimposed upon one another to the left and right of the main reflection.

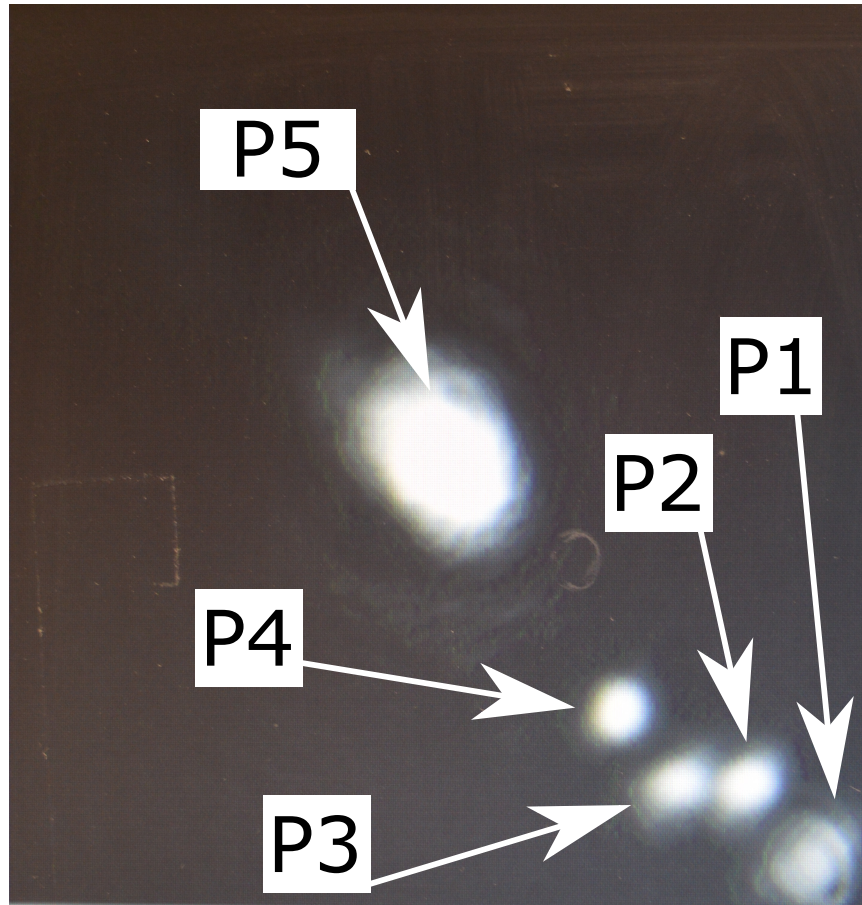


Figure 98: Screen shot similar to Figure 97, but here the interrogating beam is coming in at an angle to show the various reflections. Reflections from the front and back of each microscope slide are now evident.

Figures 99, 100, 101, 102, 103 are screen shots of the spot images for surfaces P5 through P1 with interference fringes visible. Note that in Figure 99 the intensity of the interrogating beam was reduced in order to make the fringes easily visible. In practice this is not required. Due to the less intense beam the side reflections from P4 through P1 are too dim to be seen.



Figure 99: Interference fringes from surface P5. The intensity of the beam is attenuated so as not to saturate the detector. Spots P1 - P4 are too dim to be visible.

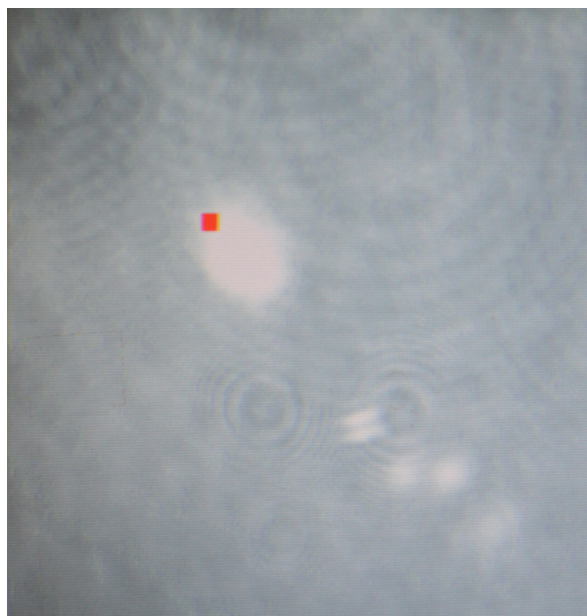


Figure 100: Interference fringes from surface P4.

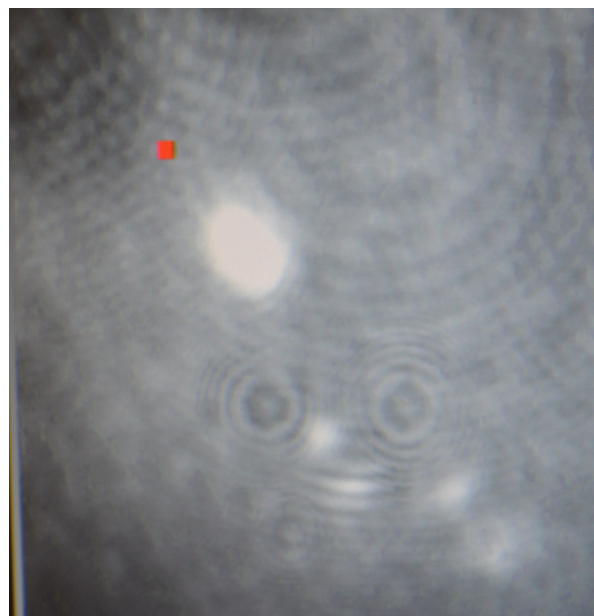


Figure 101: Interference fringes from surface P3.

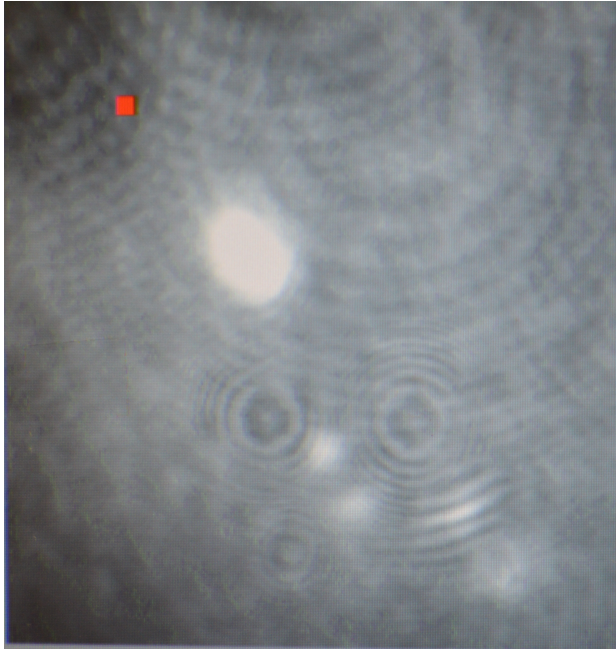


Figure 102: Interference fringes from surface P2.

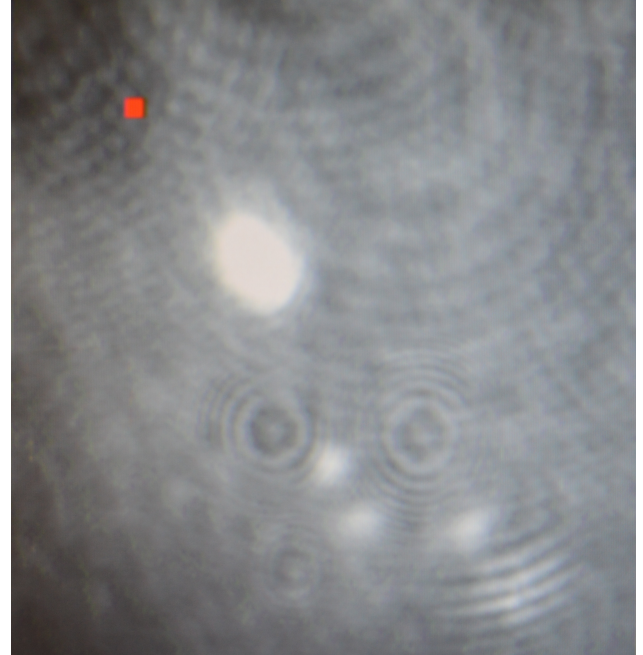


Figure 103: Interference fringes from surface P1.

Figure 104 shows the combined results from scanning each of the five surfaces of the MLO as Zernike coefficients. As expected the x-tilt is very small (due to imperfect alignment in the x-direction). The y-tilt values are positive for surfaces P2 and P2 and negative for P3 and P4, also as expected. The defocus is largest for surface P1 and approaches zero as the interface between P4 and P5 is reached, and flip sign from negative to positive for P5. This is also expected as the focal point of the interrogating beam was set between surfaces P4 and P5. Magnitudes of the remaining Zernike modes were negligible. These results demonstrate the ability of the CGLRT instrument to perform depth resolved wavefront sensing for a MLO all on a high-speed FPGA based embedded-systems platform.

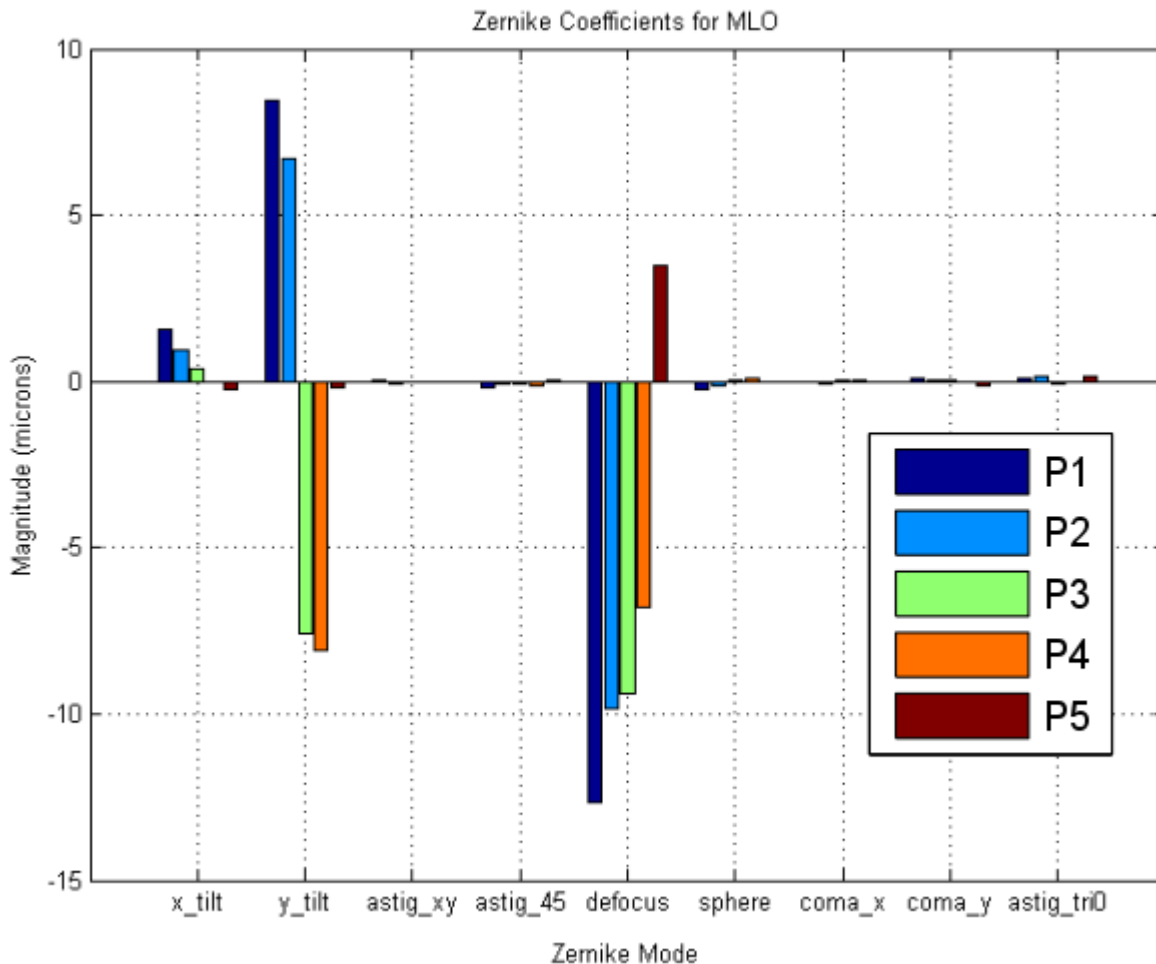


Figure 104: Magnitude of Zernike coefficients for wavefront returned from each surface of the MLO. As expected surfaces P1 and P2 have opposing y -tilt values from surfaces P3 and P4. Also, the defocus approaches zero as the interface between surface P4 and P5 is reached.

6.3 Timing and Resource Utilisation

Table 12 presents a summary of the device utilisation for the complete CGLRT system FPGA architecture. Results are given for four grid sizes (8×8 , 16×16 , 32×32 , 64×64) all with wavefront reconstruction consisting of 9 Zernike modes. The last column shows the resource utilisation for the maximum Zernike modes that can be included in the architecture for the largest grid size (121 Zernike modes). In practice, there would be little use in calculating this many modes but it demonstrates the maximum capability of the system architecture. It is notable that the architecture is very efficient in terms of resource utilisation including for the largest implementation (64×64 grid and 121 Zernike modes). Even this implementation only nears maxing out the memory (block RAM at 84%) while the remaining resources have plenty of headroom.

	8×8	16×16	32×32	64×64	64×64
Number of Zernike modes	9	9	9	9	121
Number of Slice Registers	2%	2%	2%	2%	4%
Number of Slice LUTs	4%	4%	4%	5%	18%
Number used as logic	3%	3%	4%	4%	17%
Number used as memory	1%	1%	1%	1%	1%
Number of occupied Slices	7%	7%	8%	7%	22%
Number with an unused LUT	33%	34%	36%	40%	55%
Number of fully used LUT-FF pairs	38%	35%	33%	48%	38%
Number of RAMB36E1/FIFO36E1s	11%	11%	11%	20%	84%
Number of DSP48E1s	3%	3%	3%	3%	29%

Table 12: CGLRT FPGA architecture resource utilisation for various scan sizes. The last column represents the most resource intensive version with a 64×64 grid and 121 Zernike modes.

Table 13 shows scan timing for the various grid sizes. The first row is for a 64 point grid (8×8) up to the last row which is 4096 points (64×64). The total time spent grabbing frames from the camera, in seconds, is shown in the first column. This varies from 2.59 seconds for the smallest grid size to almost 166 seconds for the largest grid size. The frame-grabbing was, by far, the largest contribution to scan time in the system. This will be discussed more below. The next column shows the contribution, in seconds, from the galvanometer scanning mirror. This delay is the total time required by the scanning mirrors to pivot and settle for each scan point. A single pivot and settle operation has a delay of approximately 3 ms which was determined experimentally and hard coded into the FPGA architecture. The PZT delay is the total time that the PZT spends modulating the OPL of the reference arm. This is also hard coded into the FPGA architecture with a single PZT move taking approximately 200 ns. The total scan time, in seconds, was measured experimentally by utilising a counter in the FPGA. The counter was started when the scan was initiated (by depressing a button on the FPGA board) and finished when the Zernike mode calculations were completed. The total measured time for each scan was approximately 13% larger than is expected by summing all of these contributions. It is expected that various sources, such as values being written to the DACs and delays in the analogue electronics are contributing factors. With the overwhelming delay being due to the frame grabbing it was not deemed critical to investigate these

other sources. Finally, the percentage of the total scan time due to frame grabbing is reported in the second-to-last column.

Grid points	Total Time Frame Grabbing (sec)	Galvo delay (sec)	PZT Delay (μ s)	Total scan time (sec)	Frame grab delay	Other Delay
64	2.59	0.192	51.2	3.188	81.3%	12%
256	10.37	0.768	204.8	12.92	80.3%	13%
1024	41.49	3.072	819.2	51.8	80.1%	14%
4096	165.97	12.29	3276.8	207.4	80.0%	14%

Table 13: Timing of the CGLRT instrument.

It is evident that approximately 80% of the time delay in the entire CGLRT instrument is due to the slow camera (96fps). Unfortunately, this was the only camera available at the time of this project but a faster camera would greatly reduce the total scan time. For example, another experimental setup in the department utilised a Mikroton EoSens CL high-speed camera. This camera is able to capture the same 250×250 pixel image with a frame rate of approximately 6000fps. Table 14 shows the expected timing of the CGLRT instrument if this high-speed camera were utilised.

Grid points	Total Time Frame Grabbing (ms)	Galvo delay (sec)	PZT Delay (μ s)	Total expected scan time (sec)	Frame grab delay
64	43	0.192	51.2	0.235	1.3%
256	171	0.768	204.8	0.939	1.3%
1024	683	3.072	819.2	3.76	1.3%
4096	2731	12.29	3276.8	15.02	1.3%

Table 14: Expected timing of the CGLRT instrument utilising a high-speed (6000fps) camera.

This would represent a massive decrease in scan time with the total time spent frame grabbing, for an 8×8 scan reduced from 3.188 seconds to 40 ms. The percentage of the total scan time spent frame-grabbing drops from 81.3% to 1.3%. The table shows that the 64 point and 256 point scans could both be performed in under 1 s. The 4096 point scan is still somewhat impractical at 15.02 s.

6.4 Conclusion

In summary, this chapter has presented the details of the CGLRT instrument's experimental setup, including the main optical and electrical systems. The optical system takes the form of a Mach-Zehnder interferometer, where the reference arm path-length is modulated by the PZT mounted reference mirror. The object arm of the interferometer is composed of the scanning beam, articulated by the galvo-scanning mirror. This interrogates the multi-layer object (MLO), composed of two weakly reflecting microscope slides and a strongly reflecting mirror. By employing coherence-gating the instrument is able to perform depth-resolved wavefront sensing, differentiating between the five wavefronts returned from the MLO. The instrument is able to do this in the presence of the strong reflection from the MLO mirror.

The FPGA based embedded system provides the command and control functions for the instrument, including image post-processing and wavefront reconstruction. Based on the Xilinx ML605 board, and utilising the Virtex-6 FPGA, the system is an elegant self-contained unit which is able to:

1. provide control signals to the galvo-scanning mirror via the PMOD connected DAC,
2. provide a control voltage to the PZT via the PMOD connected DAC,
3. read data from the camera via the FMC connected Cameralink card,
4. display the output of the camera via a VGA monitor,
5. perform image processing including image subtraction, binarization and centroiding and
6. perform wavefront reconstruction and deliver the results as Zernike coefficients.

The limiting factor in the speed of the system was the frame-rate of the available camera. Section 6.3 shows that the use of a high frame-rate camera would decrease the acquisition time by 92% for an 8×8 scan.

The instrument described in this chapter is the culmination of over three years of intense research and engineering. It is hoped that the cross-disciplinary effort of developing a combined optical and FPGA based system yielded an instrument that is more capable than either would

have produced alone. This novel CGLRT instrument is able to perform, for the first time with this modality, depth-resolved wavefront sensing, and to do so in real-time. Up until this point depth-resolved wavefront sensing (based on a SHS) has relied on image storage and PC based post-processing that is done off-line. Using an embedded systems based approach all of these functions are able to be completed in real-time, all within the FPGA.

7 Conclusion

The FPGA architecture for wavefront reconstruction detailed in Chapter 5 was published in an article entitled "An FPGA Architecture for Extracting Real-Time Zernike Coefficients from Measured Phase Gradients" [34].

The CGLRT acted as a suitable, though perhaps not ideal, testbed for the development of the FPGA based system, due to the inherent speed limitations of the serial, scanning based system. If the foundation laid for the embedded system were given minor alterations, it would be applicable not only to CGLRT but CGSHS as well. As designed, the embedded system described in this research utilises that same Cameralink interface, DAC controllers, VGA controller, image-processing and wavefront reconstruction subsystems that a CGSHS system would require. If the system were to be modified for operation with a CGSHS system it would need to:

1. remove the portions of the circuit responsible for control of the galvo-scanning mirror, as this is no longer required,
2. implement a subsystem to handle segmentation of the Shack-Hartmann image and
3. rework the centroiding subsystem to calculate the centroid of each image segment in parallel.

An implementation of this sort would do away with most of the time associated with frame-grabbing, while also removing the need for galvo-scanning. For example, Table 15 compares the time required for a 64 point scan (using the 6000 fps camera) in the CGLRT with a similar setup for a theoretical CGSHS instrument.

Modality	Total Time Frame Grabbing (ms)	Galvo delay (sec)	PZT Delay (μs)	Wavefront Recon. Delay (μs)	Total expected scan time (ms)
CGLRT	43	0.192	51.2	0.285	235
CGSHS	0.67	0	0.80	0.285	1.47

Table 15: Timing comparison of the CGLRT instrument to a theoretical CGSHS.

This comparison assumes that a single CGSHS scan is completed using four-frames and that the delay due to PZT scanning (to acquire four CG images) is comparable to that of CGLRT. It also assumes the use of the same FPGA based architecture for reconstructing the wavefront. This is

not a completely fair comparison, as the CGSHS architecture would need to account for segmenting tasks, but this may be able to be offset by increased parallelisation in the Zernike mode calculation circuit. As a rough comparison though, a CGSHS system may realise an increase in sensing time of up to 158 times over the CGLRT. This strengthens the case for extending the FPGA architecture in this research to the CGSHS modality.

It is also useful to note at what point the FPGA architecture will become a limiting factor for the performance of the theoretical CGSHS system. As in the case of the CGLRT, the time required to grab frames from the camera is the driving factor. Assuming a frame-rate of 6000 fps, the time to grab a single frame ($167 \mu s$) is 575 times greater than the time required to perform a full wavefront recovery. Even for the largest grid size (4096 points), the FPGA architecture, at $20.06 \mu s$ would require an increase of almost 10x in the frame rate of the camera to begin being a limiting factor. The Mikrotron camera referenced earlier, used in the current CGSHS setup, is able to run up to 33,390 fps but only as a reduced image size of 60 by 60 pixels. This shows the the FPGA architecture, as designed, leaves ample headroom for the remaining system components to be improved before it becomes a limiting factor.

7.1 Future Work

While the research presented here has demonstrated the ability for LRT to perform high-speed depth resolved wavefront sensing further work may be done to bring the system closer to a fully-functional adaptive optics (AO) system. As it stands, the current system fulfills the wavefront sensing portion of an AO system as shown in Figure 6. It still remains to incorporate this wavefront sensor into a fully-functional adaptive optics system.

A natural next step would be to expand the system to include a deformable-mirror (DM) and 'close the loop.' The FPGA based system, as presented, has ample resources (see 12) to implement a closed loop system for control of the DM. In this manner, depth-resolved wavefronts could not only be measured but also corrected. It may be interesting to investigate a zonal wavefront recovery method in this case. In this way, as each point in the LRT scan pattern is measured, the corresponding segments of the DM could be updated. This would allow the wavefront correction to be continuously updated without having to wait for a total scan to be completed.

Another next-step in this research would be to perform CGLRT for back-scattering targets such as biological tissue. This would be a much greater challenge due to the small amount of light typically returned from within the focal-plane of the object being imaged. This step would pave the way for CGLRT to be integrated into an existing ocular imaging system, such as OCT, to perform wavefront correction.

Outside of realising a full-fledged AO system, it would also be of interest to compare CGLRT to coherence-gated Shack-Hartmann (CGSHS) and show that the systems provide consistent results. In a similar fashion to the Navarro paper [35], a dual system which compared the output a CGLRT and CGSHS for a single, multi-layer target would be useful in demonstrating the self-consistency of the techniques.

7.2 Closing Comments

This dissertation has described the development of a coherence-gated laser ray tracing wavefront sensor with the ability to perform depth-resolved wavefront sensing. This cross-disciplinary project, taking an instrumentation approach, has also developed a highly-specialised, FPGA based embedded system on which the instrument rests. The motivation to perform this research stemmed from patent US 8451452 B2 entitled "Method for depth resolved wavefront sensing, depth resolved wavefront sensors and method and apparatus for optical imaging" [39]. This patent described the theoretical operational principles of depth-resolved versions of both the Shack-Hartmann and laser ray tracing type wavefront sensors. The Shack-Hartmann versions were built and described in the papers "Depth-resolved wavefront aberrations using a coherence-gated Shack-Hartmann wavefront sensor" [59] and "Demonstration of real-time depth-resolved Shack-Hartmann measurement" [61]. The research described in this dissertation has filled in the laser ray tracing gap by demonstrating not only a PC based depth-resolved LRT, but going one step further, a high-speed FPGA based depth-resolved LRT.

References

- [1] Ackermann, G. K. and J. Eichler (2007). *Holography: a practical approach*. John Wiley & Sons.
- [2] Babcock, H. W. (1953). The possibility of compensating astronomical seeing. *Publications of the Astronomical Society of the Pacific* 65(386), 229–236.
- [3] Booth, M. J., M. Neil, and T. Wilson (1998). Aberration correction for confocal imaging in refractive-index-mismatched media. *Journal of Microscopy* 192(2), 90–98.
- [4] Carré, P. (1966). Installation et utilisation du comparateur photoélectrique et interférentiel du Bureau International des Poids et Mesures. *Metrologia* 2(1), 13.
- [5] Cubalchini, R. (1979). Modal wave-front estimation from phase derivative measurements. *JOSA* 69(7), 972–977.
- [6] Digilent Inc. (2011, November). Digilent pmod interface specification. Accessed 20 September 2016.
- [7] Dreher, A., J. Bille, and R. Weinreb (1989). Active optical depth resolution improvement of the laser tomographic scanner. *Applied Optics* 28(4), 804–808.
- [8] Drexler, W., U. Morgner, R. K. Ghanta, F. X. Kärtner, J. S. Schuman, and J. G. Fujimoto (2001). Ultrahigh-resolution ophthalmic optical coherence tomography. *Nature medicine* 7(4), 502–507.
- [9] Feierabend, M., M. Rückel, and W. Denk (2004). Coherence-gated wave-front sensing in strongly scattering samples. *Optics letters* 29(19), 2255–2257.
- [10] Gass, J. D. M. (1997). *Stereoscopic atlas of macular diseases: diagnosis and treatment*. Mosby Inc.
- [11] Gavel, D. T. (2011). New adaptive-optics technology for ground-based astronomical telescopes.
- [12] Gonsalves, R. A. (1982). Phase retrieval and diversity in adaptive optics. *Optical Engineering* 21(5), 215829–215829.
- [13] Grimm, B., S. Goelz, J. F. Bille, et al. (1994). Objective measurement of wave aberrations of the human eye with the use of a Hartmann-Shack wave-front sensor. *JOSA A* 11(7), 1949–1957.
- [14] Hartmann, J. (1900). Bemerkungen über den bau und die justirung von spektrographen. *Z. Instrumentenkde* 20, 47.

- [15] Hermann, B., E. Fernández, A. Unterhuber, H. Sattmann, A. Fercher, W. Drexler, P. Prieto, and P. Artal (2004). Adaptive-optics ultrahigh-resolution optical coherence tomography. *Optics letters* 29(18), 2142–2144.
- [16] Hornbeck, L. J. (1984, April 10). Deformable mirror light modulator. US Patent 4,441,791.
- [17] Hornbeck, L. J. (1990). Deformable-mirror spatial light modulators. In *33rd Annual Technical Symposium*, pp. 86–103. International Society for Optics and Photonics.
- [18] Howland, B. (1968). Use of crossed cylinder lens in photographic lens evaluation. *Applied optics* 7(8), 1587–1600.
- [19] Howland, H. (2000). The history and methods of ophthalmic wavefront sensing. *Journal of Refractive Surgery* 16(5), 552–553.
- [20] Huang, D., E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto (1991). Optical coherence tomography. *Science* 254, 1178–1181.
- [21] Irvin, M., T. Kitazawa, and T. Suzuki (1996). A new generation of mpeg-2 video encoder asic and its application to new technology markets.
- [22] Li, J., M. Sarunic, and L. Shannon (2011). Scalable, high performance Fourier domain optical coherence tomography: Why FPGAs and not GPGPUs. In *Field-Programmable Custom Computing Machines (FCCM), 2011 IEEE 19th Annual International Symposium on*, pp. 49–56. IEEE.
- [23] Liang, J., D. R. Williams, and D. T. Miller (1997). Supernormal vision and high-resolution retinal imaging through adaptive optics. *JOSA A* 14(11), 2884–2892.
- [24] Love, G. D. (1997). Wave-front correction and production of zernike modes with a liquid-crystal spatial light modulator. *Applied optics* 36(7), 1517–1524.
- [25] Malacara, D. (2007). *Optical shop testing*, Volume 59. Wiley-Interscience.
- [26] Mathworks (2012). *version 7.10.0 (R2012a)*. Natick, Massachusetts: The MathWorks Inc.
- [27] Michelson, A. A. (1881). The relative motion of the earth and of the luminiferous ether. *American Journal of Science* (128), 120–129.
- [28] Miller, D., J. Qu, R. Jonnal, and K. Thorn (2003). Coherence gating and adaptive optics in the eye. In *Biomedical Optics 2003*, pp. 65–72. International Society for Optics and Photonics.
- [29] Miller, D. T., D. R. Williams, G. M. Morris, and J. Liang (1996). Images of cone photoreceptors in the living human eye. *Vision research* 36(8), 1067–1079.

- [30] Minsky, M. (1961). Microscopy apparatus us patent 3013467. *USP Office, Ed. US*.
- [31] Molebny, V., I. Pallikaris, L. Naoumidis, I. Chyzh, S. Molebny, and V. Sokurenko (1997). Retina ray-tracing technique for eye-refraction mapping. In *Proc. SPIE*, Volume 2971, pp. 175–183.
- [32] Molebny, V., S. Panagopoulou, S. Molebny, Y. Wakil, and I. Pallikaris (2000). Principles of ray tracing aberrometry. *Journal of Refractive Surgery* 16(5), 572–575.
- [33] Moreno-Barriuso, E. and R. Navarro (2000). Laser ray tracing versus Hartmann–Shack sensor for measuring optical aberrations in the human eye. *JOSA A* 17(6), 974–985.
- [34] Moser, S., P. Lee, and A. Podoleanu (2015). An fpga architecture for extracting real-time zernike coefficients from measured phase gradients. *Measurement Science Review* 15(2), 92–100.
- [35] Navarro, R. and E. Moreno-Barriuso (1999). Laser ray-tracing method for optical testing. *Optics letters* 24(14), 951–953.
- [36] Noll, R. J. (1976). Zernike polynomials and atmospheric turbulence. *JOsA* 66(3), 207–211.
- [37] Pedrazzani, M., V. Lorientte, P. Tchenio, S. Benrezzak, D. Nutarelli, and A. Fragola (2016). Sensorless adaptive optics implementation in widefield optical sectioning microscopy inside in vivo drosophila brain. *Journal of biomedical optics* 21(3), 036006–036006.
- [38] Podoleanu, A. (2005). Optical coherence tomography. *British journal of radiology* 78(935), 976–988.
- [39] Podoleanu, A. and S. Tuohy (2013, May 28). Method for depth resolved wavefront sensing, depth resolved wavefront sensors and method and apparatus for optical imaging. US Patent 8,451,452.
- [40] Polster, H. D., J. Pastor, R. M. Scott, R. Crane, P. Langenbeck, R. Pilston, and G. Steinberg (1969). New developments in interferometry. *Applied Optics* 8(3), 521–521.
- [41] Ragazzoni, R. (1996). Pupil plane wavefront sensing with an oscillating prism. *Journal of modern optics* 43(2), 289–293.
- [42] Reinig, M. R., S. W. Novack, X. Tao, F. Ermini, L. A. Bentolila, D. G. Roberts, A. MacKenzie-Graham, S. E. Godshalk, M. Raven, and J. Kubby (2016). Adaptive optics microscopy enhances image quality in deep layers of clarity processed brains of yfp-h mice. In *SPIE BiOS*, pp. 969008–969008. International Society for Optics and Photonics.
- [43] Roddier, F. (1999). *Adaptive optics in astronomy*. Cambridge university press.

- [44] Roddier, F. and C. Roddier (1988). Curvature sensing and compensation: a new concept in adaptive optics. In *Very Large Telescopes and their Instrumentation, Vol. 2*, Volume 2, pp. 667–673.
- [45] Roddier, N. A. (1990). Atmospheric wavefront simulation using zernike polynomials. *Optical Engineering* 29(10), 1174–1180.
- [46] Rodríguez-Ramos, J., E. M. Castelló, C. D. Conde, M. R. Valido, and J. Marichal-Hernández (2008). 2d-fft implementation on fpga for wavefront phase recovery from the cafadis camera. In *SPIE Astronomical Telescopes+ Instrumentation*, pp. 701539–701539. International Society for Optics and Photonics.
- [47] Rueckel, M. and W. Denk (2006). Coherence-gated wavefront sensing using a virtual Shack-Hartmann sensor. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Volume 6306, pp. 16.
- [48] Rueckel, M. and W. Denk (2007). Properties of coherence-gated wavefront sensing. *JOSA A* 24(11), 3517–3529.
- [49] Rueckel, M., J. Mack-Bucher, and W. Denk (2006). Adaptive wavefront correction in two-photon microscopy using coherence-gated wavefront sensing. *Proceedings of the National Academy of Sciences* 103(46), 17137–17142.
- [50] Saunter, C., G. Love, M. Johns, and J. Holmes (2005). Fpga technology for high-speed low-cost adaptive optics. In *5th International Workshop on Adaptive Optics for Industry and Medicine*, pp. 60181G–60181G. International Society for Optics and Photonics.
- [51] Schwiegerling, J. and D. Neal (2005). Historical development of the Shack-Hartmann wavefront sensor. *Legends in Applied Optics, R. Shannon, R. Shack, JE Harvey, and RB Hooker eds. (SPIE, Bellingham WA) PM148*, 132–139.
- [52] Shack, R. V. and B. Platt (1971). Production and use of a lenticular Hartmann screen. *J. Opt. Soc. Am* 61(5), 656.
- [53] Shannon, L., J. Li, M. Mohammadnia, and M. Sarunic (2011). Evaluating the scalability of high-performance, Fourier-domain optical coherence tomography on GPGPUs and FPGAs. In *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pp. 483–487. IEEE.
- [54] Smith, W. J. (2000). *Modern optical engineering*. Tata McGraw-Hill Education.
- [55] Thorlabs Inc. (2014a, July). Shack-hartmann wavefront sensor, 450 hz frame rate.
- [56] Thorlabs Inc. (2014b, July). Shack-hartmann wavefront sensors, 1.3 megapixel resolution.

- [57] Tracey Technologies Corporation (2013, April). Ray tracing technology: A background.
- [58] Tscherning, M. (1893). *Die monochromatischen Aberrationen des menschlichen Auges*.
- [59] Tuohy, S. and A. Podoleanu (2010). Depth-resolved wavefront aberrations using a coherence-gated Shack-Hartmann wavefront sensor. *Optics Express* 18(4), 3458–3476.
- [60] Vabre, L., A. Dubois, and A. C. Boccara (2002). Thermal-light full-field optical coherence tomography. *Optics letters* 27(7), 530–532.
- [61] Wang, J., A. G. Podoleanu, et al. (2012). Demonstration of real-time depth-resolved Shack–Hartmann measurements. *Optics letters* 37(23), 4862–4864.
- [62] Webb, R. H., G. W. Hughes, and F. C. Delori (1987). Confocal scanning laser ophthalmoscope. *Applied optics* 26(8), 1492–1499.
- [63] Wojtkowski, M., T. Bajraszewski, P. Targowski, and A. Kowalczyk (2003). Real-time in vivo imaging by high-speed spectral optical coherence tomography. *Optics letters* 28(19), 1745–1747.
- [64] Xilinx Inc. *ML605 Hardware User Guide*. Xilinx Inc.
- [65] Xilinx Inc. *Virtex-6 FPGA DSP48E1 Slice User Guide*. Xilinx Inc.
- [66] Xilinx Inc. (2012, February). *Virtex-6 FPGA Configurable Logic Block User Guide*. Xilinx Inc.
- [67] Xilinx Inc. (2015, August). What is a fpga?
- [68] Xilinx Inc. (2016a, September). Fmc-ce: Basic input/output expansion board. Accessed 18 September 2016.
- [69] Xilinx Inc. (2016b, September). Fpga mezzanine card (fmc) standard. Accessed 18 September 2016.
- [70] Zawadzki, R., S. Jones, S. Olivier, M. Zhao, B. Bower, J. Izatt, S. Choi, S. Laut, and J. Werner (2005). Adaptive-optics optical coherence tomography for high-resolution and high-speed 3d retinal in vivo imaging. *Optics express* 13(21), 8532.