

Adaptive optics for corrections of phase and polarisation state aberrations in microscopes



Qi Hu
St Hilda's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2021

This thesis is dedicated to
Hu Xing Long, Wang Xi Qin and Guo Shi Gong
for their love and support

Acknowledgements

The work in this thesis has only been possible through the help of many people.

Firstly, I am very grateful to Prof Martin J Booth for his knowledge, guidance and supervision throughout my undergraduate and doctoral studies.

I am also grateful to Dr Patrick Salter for his ideas and advice.

I am thankful to Dr Jacopo Antonello for his technical assistance with the Spatial Light Modulator, Dr Jingyu Wang for his technical expertise in microscopy, Drs Martin Hailstone and Raphaël Turcotte for providing the biological samples and advice, and Dr Matthew Wincott for his technical assistance and support in microscopy.

My gratitude goes to Dr Chao He and Debayan Saha (Max-Planck-Institut für molekulare Zellbiologie und Genetik, Dresden, Germany) for their ideas and advice, as well as Dr Asad Hussain for this support.

My thanks are also due to Dr Karen Hampson, Dr Chris Chong and Dr Ravi Shivaraman for their enthusiasm and encouragement.

I am also indebted to Prof Booth and the European Research Council for providing financial support.

Abstract

Phase and polarization aberrations can affect the focus of optical systems and thus cause image distortion and quality reduction. Adaptive optics (AO) such as objective lens correction collars, spatial light modulators (SLM) and deformable mirrors (DM) are used in microscopy imaging systems to modulate and correct phase and polarization state errors. Different adaptive elements have special properties which make them suitable for different systems. In certain applications such as laser machining or microscopic imaging through an interface between mismatched refractive indices, spherical aberration is the major factor that affects the focal spot quality; the correction collar attached to the objective lens is a suitable and simple form of adaptive elements for correcting spherical aberration. In some microscopes, system misalignment and sample induced phase aberrations, especially in deep imaging, greatly affect image quality; a DM can be a good option when correcting phase aberrations of both the input and output beam while a SLM can be used if only the input laser beam requires correction. In certain systems, such as gradient index (GRIN) lens and stimulated emission depletion (STED) microscopy imaging systems, the image quality is also affected by polarization aberrations which can potentially be corrected by SLM devices.

Despite years of development, the operations of these adaptive elements to fulfil different correction tasks effectively and efficiently remain challenging. The spherical compensations of objective lens correction collars

will introduce extra defocus at the same time, which makes it difficult to operate. In the cases when the fluorescence signal level is low, wavefront sensor-based operations of adaptive elements become infeasible so wavefront sensorless methods are used; yet there was no systematic comparison for various sensorless based methods to help people understand them better and advice people on method selections. Conventional sensorless based correction methods, which normally require multiple exposures per aberration mode and multiple rounds of corrections, are not the most efficient ways of operation. Polarisation states have a total of three degrees of freedom; corrections of polarisation states are more complex compared to phase aberrations, for which there is only one degree of freedom. Polarisation state conversion will induce extra geometric phase error which makes corrections more difficult.

In this thesis, I discussed and presented results trying to resolve the problems mentioned above. I first presented an automated control system for both spherical and defocus compensation using a motorised objective lens correction collar and an axially movable sample mounting stage. I then presented a universal framework for sensorless AO method to provide a better understanding of various sensorless approaches; this framework also allowed a systematic comparisons of various aberration representations used in different methods. I then presented a machine learning based sensorless AO method which shows promise to be more versatile than conventional sensor-less AO methods. Finally I presented theoretical analysis of vectorial state conversions (which includes both polarisation state and phase properties of light) and complex retarder compensations using a series of adaptive elements.

Statement of Originality

I hereby declare that except where specific references to the work of others are given, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except in the following:

1. The widefield microscope in chapter 3 was designed and built by Dr Bei Li and Dr Jingwen Xiao.
2. The two-photon microscope in chapter 5 was originally designed and built by Dr Jingyu Wang. The system was later modified and aligned by me.
3. The *Drosophila* melanogaster larvae sample in chapter 5 was prepared by Dalia Gala.
4. The fixed Thy1-GFP mouse brain tissue sample in chapter 5 was prepared by Dr Raphaël Turcotte.
5. The freshly-extracted, non-fixed *Drosophila* larval brain tissue sample in chapter 5 was prepared by Dr Martin Hailstone.

List of publications

1. **Q. Hu**, C. He, and M. J. Booth, “Arbitrary complex retarders using a sequence of spatial light modulators as the basis for adaptive polarisation compensation,” *J. Opt.*, 2021.
2. **Q. Hu**, Y. Dai, C. He, and M. J. Booth, “Arbitrary vectorial state conversion using liquid crystal spatial light modulators,” *Opt. Commun.*, vol. 459, p. 125028, 2020.
3. **Q. Hu**, J. Wang, J. Antonello, M. Hailstone, M. Wincott, R. Turcotte, D. Gala, and M. J. Booth, “A universal framework for microscope sensorless adaptive optics: Generalized aberration representations,” *APL Photonics*, vol. 5, no. 10, p. 100801, 2020.
4. D. Saha, U. Schmidt, Q. Zhang, A. Barbotin, **Q. Hu**, N. Ji, M. J. Booth, M. Weigert, and E. W. Myers, “Practical sensorless aberration estimation for 3d microscopy with deep learning,” *Opt. Express*, vol. 28, pp. 29044–29053, 2020.
5. Y. Dai, C. He, J. Wang, R. Turcotte, L. Fish, M. Wincott, **Q. Hu**, and M. J. Booth, “Active compensation of extrinsic polarization errors using adaptive optics,” *Opt. Express*, vol. 27, pp. 35797–35810, 2019.
6. C. He, J. Chang, **Q. Hu**, J. Wang, J. Antonello, H. He, S. Liu, J. Lin, B. Dai, D. S. Elson, P. Xi, H. Ma, and M. J. Booth, “Complex vectorial optics through gradient index lens cascades,” *Nat. Commun.*, vol. 10, no. 1, p. 4264, 2019.

Abbreviations

1-P single-photon. 56, 102, 110, 114–116, 118, 147, 149

2-P two-photon. 22–25, 72, 78, 79, 102, 104, 110, 112, 114, 116, 118, 149, 150, 155, 156

3-P three-photon. 22, 94

AO adaptive optics. 2–4, 8, 17, 24, 27, 29, 36, 37, 54–56, 62, 63, 67–70, 72, 94, 95, 136, 137, 145–151

CARS coherent anti-Stokes Raman scattering. 23, 24, 55

CNN convolutional neural network. 34, 97, 98, 100, 101, 106, 119, 151

DM deformable mirror. 2, 27, 28, 30, 36, 56, 63, 64, 94, 136, 146, 148

DNN deep neural network. 34, 99

FOV field of view. 48, 49, 52, 65, 100, 102, 113

GFP green fluorescence protein. 19

GPU graphics processing unit. 67, 151

GRIN graded-index. 2, 121, 150

IMPACT iterative multi-photon adaptive compensation technique. 55

iPSF inverse point spread function. 105, 107

LC liquid crystal. 28, 29

MEMS micro-electro-mechanical systems. 28

ML machine learning. 32–35, 98, 99, 149, 151

MSE mean squared error. 109

MSLE mean squared logarithmic error. 109, 111–117

NA numerical aperture. 1, 15–17, 27, 73, 74, 151

NN neural network. 4, 97–100, 149–151

OCM optical coherence microscopy. 55

OPL optical path length. 7, 28, 29

PB Pancharatnam-Berry. 11, 130, 133–136, 143, 171, 174, 175, 179–185

PMT photon multiplier tube. 84, 89, 91

PSF point spread function. 21, 40, 48, 56, 59, 62, 65, 97, 99, 100, 103–107, 110, 113, 117, 118, 124, 149

RMS root mean square. 67, 76, 80, 85, 111, 155, 156

SHG second harmonic generation. 22, 23

SIM structured illumination microscopy. 2, 25, 26, 55, 72, 121, 150

SLM spatial light modulator. 2, 3, 27–29, 36, 56, 72, 83, 87, 122–124, 126–146, 149, 150, 166–177, 179–185, 190, 191

SMLM single molecule localisation microscopy. 25

SMS single molecule switching. 55, 65, 72

SNR signal to noise ratio. 42, 56, 71, 72, 75, 81, 83–85, 88, 89, 91, 93, 95, 120

SOP state of polarisation. 125, 127, 128

STED stimulated emission depletion. 2, 25, 26, 55, 59, 65, 72, 121, 150

SVD Singular value decomposition. 64

THG third harmonic generation. 23

UV ultraviolet. 19

Contents

1	Introduction	1
1.1	Research contributions	2
1.2	Thesis outline	3
2	Background	5
2.1	Models of light	5
2.1.1	Wave model of light	6
2.1.2	Phase property	6
2.1.3	Polarisation state	8
2.1.3.1	Jones calculus	8
2.1.3.2	Stokes parameters and Müller matrices	9
2.1.3.3	Polarisation state transmission and phase difference .	11
2.1.3.4	Poincaré sphere	11
2.1.4	Diffraction and Fourier optics	12
2.1.4.1	Diffraction and approximations	12
2.1.4.2	Thin lens model	14
2.1.4.3	High numerical aperture lens model	15
2.2	Microscopy	17
2.2.1	The invention of optical compound microscopy	17
2.2.2	Fluorescence microscopy	19

2.2.3	Confocal microscopy	20
2.2.4	Nonlinear optical microscopy	22
2.2.5	High-resolution and super-resolution microscopy	25
2.3	Adaptive elements	27
2.3.1	Adaptive optics devices	27
2.3.1.1	Objective lens correction collar	27
2.3.1.2	Deformable mirror	28
2.3.1.3	Spatial light modulator	28
2.3.2	Adaptive optics operations	29
2.3.2.1	Sensored operations	29
2.3.2.2	Sensorless operations	30
2.3.2.3	Zernike polynomials	31
2.4	Machine learning	32
2.4.1	Supervised learning	33
2.4.2	Unsupervised learning	34
2.4.3	Reinforcement learning	34
3	An automated control system for spherical aberration and defocus compensation	36
3.1	The experimental setup	38
3.2	An automated control system	38
3.2.1	An image based assessment of spherical aberration and defocus correction	40
3.2.1.1	Some modifications to the frequency threshold metric	42
3.2.1.2	Variations of the image quality metric against the spherical aberration and defocus	44
3.2.2	Specification of the control system using the image quality metric	47

3.2.2.1	To locate the initial axial position of the movable sample mounting stage	47
3.2.2.2	To determine the gradient of the spherical-defocus linear model	48
3.2.2.3	To determine the maximum metric reading	50
3.3	Incorporating the control system into the new Zeiss objective lens . .	51
3.4	Conclusion and future outlook	53
4	A universal framework for sensorless adaptive optics	54
4.1	The mathematical framework for sensorless adaptive optics	55
4.2	Aberration representation modes	63
4.3	Optimisation metrics	64
4.4	Optimisation algorithms	66
4.5	Summary and discussion	67
5	Generalized aberration representations	69
5.1	Choice of microscope	72
5.2	Choice of optimisation metric and algorithm	72
5.3	Choice of simulated input aberrations for correction	76
5.4	Simulated and experimental comparisons of aberration representations over a square pupil	78
5.4.1	Square pupil aberration representations	78
5.4.2	Simulation and experimental results comparing the four aberration representations	80
5.4.3	Simulation results of multiple correction iterations comparing the four aberration representations	84
5.5	Experimental comparison of aberration representations over a circular pupil	86

5.5.1	Circular pupil aberration representations	87
5.5.2	Experimental results using tissue specimen and SLM-induced aberrations	87
5.5.3	Comparison of circular representation sets correcting specimen-induced aberration	89
5.6	Discussion	92
5.7	Conclusion	95
6	Machine learning based sensor-less adaptive optics	96
6.1	An untrained CNN for deducing pupil phase from a pair of images . .	100
6.1.1	A schematic illustration of the untrained network with physical meanings	104
6.1.2	A flowchart of the network with details	105
6.1.3	More details on the choices of the network parameters and designs	107
6.2	Demonstrations of the neural network on simulated data	110
6.2.1	The neural network modelling a 2-P microscope	110
6.2.1.1	Astigmatism phase retrieval	111
6.2.1.2	Coma phase retrieval	111
6.2.1.3	Trefoil phase retrieval	111
6.2.2	The network modelling a 1-P widefield microscope	114
6.2.2.1	Astigmatism phase retrieval	114
6.2.2.2	Coma phase retrieval	115
6.2.2.3	Trefoil phase retrieval	115
6.2.3	Discussion	117
6.3	Conclusion	118
7	Compensation of vectorial state errors using adaptive optics	121
7.1	Model system	122

7.2	Conversions between arbitrary polarisation	
	states	125
7.2.1	Conversions from a fixed to any arbitrary polarisation	
	state	126
7.2.2	Conversion from any arbitrary to a fixed polarisation	
	state	128
7.2.3	Conversions between arbitrary polarisation states	128
7.3	General conversions between arbitrary vectorial states	130
7.3.1	The PB phase of a single SOP transition	130
	7.3.1.1 The geometrical expression of the PB phase	130
	7.3.1.2 The PB phase derived through interference	132
7.3.2	The PB phase of the 3-SLM system	132
	7.3.2.1 The retardance settings of SLMs and the intermediate	
	transitional states	134
	7.3.2.2 A discussion of the P-B phase variations for the 3-	
	SLM system	135
7.3.3	The PB phase of a 4-SLM system to allow a full vectorial state	
	modulation	136
7.4	Compensation of the phase and polarisation state errors using a multi-	
	SLM system	137
7.4.1	The Jones matrix of any arbitrary homogeneous retarder for	
	compensation	138
7.4.2	Compensations of the polarisation state errors using a 3-SLM	
	system	139
	7.4.2.1 The configuration of a 3-SLM system with eigenvec-	
	tors same as the retarder	140

7.4.2.2	The configuration of a 3-SLM system with eigenvalues conjugate to that of the retarder	141
7.4.3	Compensations of both the polarisation state and phase errors using a 4-SLM system	142
7.4.4	Corrections of the vectorial state aberrations using a 4-SLM system on a full Poincaré beam	143
7.5	Conclusion	143
8	Conclusions and outlooks	146
8.1	Conclusions	146
8.2	Outlook	150
A	A discussion on the piecewise modal, piecewise constant and contin- uous modal schemes	153
A.1	Similarities between piecewise modal and piecewise constant schemes	153
A.2	Differences between piecewise modal and piecewise constant schemes	155
A.3	Differences between piecewise and continuous schemes	156
A.4	Orthogonality defined for a 2-P signal metric	156
B	Auto-correlation of various aberration representations	158
C	Derivations of different expressions for the PB phase	161
C.1	The geometrical and interference expression of PB phase	161
C.2	Derivations of the Stokes expression from the Jones expression of PB phase	164
D	Derivations of the SLM settings and intermediate SOPs when pass- ing through a 3-SLM system	166
E	Analysis of PB phase from a 3-SLM system	170

E.1	The PB phase variations with ψ_1 when $A_1 = D_1 \neq 0$	172
E.1.1	When $A_1 = D_1 = 1$	173
E.1.2	When $A_1 = D_1 = -1$	173
E.1.3	When $ A_1 = D_1 \neq 1 \neq 0$	173
E.2	The PB phase variations with ψ_1 when $A_1 = -D_1$	176
E.2.1	When $A_1 = -D_1 \neq 0$	176
E.2.2	When $A_1 = -D_1 = 0$	177
E.3	The PB phase variations with ψ_1 when $ A_1 > D_1 $	179
E.4	The PB phase variations with ψ_1 when $ A_1 < D_1 $	181
E.5	Conclusion	182
F	The PB phase analysis after passing through a 4-SLM system	184
G	The mathematical expressions for homogeneous retarders	186
G.1	The general expression for a homogeneous retarder	186
G.2	The general expression for a homogeneous matrix	188
G.3	The general expression for a sequence of homogeneous retarders . . .	189
H	The derivations of the retardance settings of the three SLMs for the polarisation compensations of an arbitrary retarder	190
	Bibliography	193

Chapter 1

Introduction

Microscopy is an important tool, particularly for biomedical imaging, as it reveals structure and function on a sub cellular scale, whilst being able to image multiple cells and tissues over volumes. Microscopes suffer from phase aberrations that arise from the optics and specimens [1–3]. Misaligned lenses result astigmatism and coma aberrations [4]. The coverslip on the specimen introduces a spherical aberration because the refractive index of glass (1.52) is different from that of common immersion fluids such as water (1.33) and glycerol (1.47) [5]. Specimens containing spatially varying structures and refractive index also induces scattering and aberrations which can significantly affect image quality especially for deep imaging. Refractive index variations change optical path lengths (OPL) across the beam and distort its wavefront. After passing through an objective, light no longer constructively interfere at a confined diffraction limited spot. The focal spot is distorted in both the lateral and axial direction and its centre intensity is reduced. This results a reduction in image resolution and contrast [1, 5].

When high numerical aperture (NA) lenses are used in microscopes, vectorial states (both phase and polarisation state) of light become important in the light propagation analysis [6–9]. Polarisation state errors can affect the optical system

imaging quality especially in graded-index (GRIN) optics fibre imaging [10, 11] and in some super-resolution microscopy such as stimulated emission depletion (STED) and structured illumination microscopy (SIM) systems [12–14]. Vectorial aberrations may be introduced focussing through birefringent crystals, through stressed optical elements or due to dielectric coatings [11, 15, 16].

Adaptive optical elements can be used to correct the vectorial aberrations introduced in microscopy systems. Common adaptive optics (AO) elements such as an objective lens correction collar can be deployed to correct spherical aberrations; a deformable mirror (DM) is a common choice for phase aberration compensation; a spatial light modulator (SLM) is wavelength and polarisation state dependent and can be used for both phase and polarisation state modulation. Despite the devices are available, their operations remained a difficult task.

Comparing to sensed AO operation methods where a dedicated wavefront sensor is used for aberration measurements, sensorless methods do not require extra hardware for wavefront sensing and can be cost-effective. In cases where fluorescent signal level is low or point like beacons are not available, sensorless methods become the only viable option. However, sensorless methods tend to be time consuming and more likely to cause photobleaching or phototoxicity due to their long convergence process which are not desirable. In this thesis, we are exploring various indirect sensorless methods, in controlling AO devices for efficient and effective vectorial state aberration corrections.

1.1 Research contributions

The aim of this DPhil research work consisted of implementing AO in microscopy systems of correcting both phase and polarisation state errors so as to improve the imaging quality in an efficient way. Two AO devices - a correction collar of an

objective lens and SLM devices are specifically discussed.

Correction collars correct spherical aberration but at the same time introduce extra defocus. The operations of correction collars thus require experienced microscopy users with careful adjustments and often lead to prolonged exposures of samples and inaccurate correction settings. An automated control scheme proposed in this thesis allowed an easy and accurate use of correction collar.

Sensorless AO operations are widely used for its simple optical design. Many different schemes were proposed in literature but no thorough comparison among them has been conducted. We suggested a universal sensorless AO framework which accommodates all sensorless methods and allowed a comprehensive discussion and comparison. The results can help to guide AO users on method selections and new method developments.

A new hand-crafted machine learning based sensorless AO method was proposed and shown to be effective, highly efficient and be able to easily adapt to different types of imaging systems with little adjustments. and shown to be effective.

Polarisation state modulations in literature often involved a fixed to arbitrary polarisation state conversion or vice versa but not an arbitrary to arbitrary state conversion. We discussed a pair of arbitrary state conversion and also compensations of errors introduced by any arbitrary waveplate regardless of the input state. This work is an essential step leading to full vectorial state error corrections in imaging system.

1.2 Thesis outline

Chapter 1. Introduction and research contributions.

Chapter 2. Background.

Chapter 3. An automated control system in operation a motorised objective lens

correction collar in spherical and defocus compensation.

Chapter 4. A thorough discussion on a mathematical framework accommodating various sensorless AO methods.

Chapter 5. A comprehensive discussion on various aberration representations in phase modulation and correction.

Chapter 6. A human-guided untrained neural network (NN) in efficient aberration corrections.

Chapter 7. A vectorial state modulation and error compensation scheme.

Chapter 8. Conclusion and outlook.

Chapter 2

Background

This background chapter outlines some of the essential previous knowledge and research that underpins the rest of the thesis. This includes mathematical models of light and optics, developments of microscopy and adaptive elements, and machine learning.

2.1 Models of light

Light is electromagnetic radiation that can be modelled as ray, wave or particle [5]. Different models have different limitations and are thus used in different situations. Ray models light travelling in a straight line [17]. This model can be derived from Maxwell's equations when taking a limit of infinitesimal wavelength and is usually used in analysis involving geometrical optics. Wave models light as propagating dynamic disturbance and can also be derived from Maxwell's equations [1, 5, 18–20]. This model is usually used in analysis for physical optics. Particle models light as energy packages (photon) which can only be absorbed (or emitted) as a whole to excite (or de-excite) a molecule from one state to a higher (or lower) energy [5, 21].

2.1.1 Wave model of light

From the four Maxwell's equations, it can be easily shown that

$$\nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} = -\mu\sigma \frac{\partial \vec{E}}{\partial t} - \mu\epsilon \frac{\partial^2 \vec{E}}{\partial t^2} \quad (2.1)$$

where $\vec{E}$ is the temporally-spatially varying electric field strength; μ is the permeability, ϵ is the permittivity and σ is the conductivity of the medium; t is the time variable.

In a medium with no free charge and conductivity is small enough to be ignored

$$\nabla(\nabla \cdot \vec{E}) = 0 \quad \mu\sigma \frac{\partial \vec{E}}{\partial t} \approx 0$$

The equation 2.1 can now be simplified as

$$\nabla^2 \vec{E} = \mu\epsilon \frac{\partial^2 \vec{E}}{\partial t^2} \quad (2.2)$$

To calculate harmonic solutions to the equation 2.3, we take the time dependent term as $\vec{E} = E' e^{j\omega t}$, where ω is the angular frequency. Assume the light propagates along the axis z ,

$$\vec{E} = E e^{j(\omega t - kz)} \quad (2.3)$$

where $k^2 = \omega^2 \mu\epsilon$.

2.1.2 Phase property

If the wave is frozen in time, the spatial term e^{jkz} of the expression 2.3 can also be written as

$$e^{jkz} = e^{j(\Phi + 2m\pi)} = e^{j\Phi} e^{j2m\pi} = e^{j\Phi} \quad \forall m \in \mathbb{Z}$$

where Φ is defined as the spatial phase of the light wave.

If we consider a collimated beam travelling through the space, the points with the same phase Φ forms an wavefront. A flat wavefront is descibed as aberration free; if the wavefront is deformed, it is called aberrated.

For the case when $\Phi = 0$,

$$kz = \frac{2\pi z}{\lambda} = 2m\pi$$

where λ is the wavelength of the wave and

$$k = \frac{2\pi}{\lambda}$$

is the angular wavenumber. In a similar way, the angular frequency ω relates to the period, T , as

$$\omega = \frac{2\pi}{T}$$

The phase velocity, v_p , of the light wave is defined as

$$v_p = \frac{\lambda}{T} = \frac{\omega}{k} = \frac{1}{\sqrt{\mu\epsilon}} \quad (2.4)$$

In vacuum, the phase velocity of the light equals to c . The ratio of c and the phase velocity when light travelling through a medium is the refractive index (n) of the medium.

$$n = \frac{c}{v_p}$$

When a collimated beam with a flat wavefront passes through a medium with spatially varying refractive index (n), the optical path length (OPL)s, which is defined as $OPL = nz$, travelled by different parts of the beam are different; the phase of the beam after passing throught the medium is no longer flat and aberration is introduced into the wavefront.

Polarisation state	Linear x	Linear y	Linear 45°	Linear -45°	Left-hand circular	Right-hand circular
Jones vector (J)	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -j \end{pmatrix}$	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ j \end{pmatrix}$

Table 2.1: Examples of Jones vectors (J) of selected polarisation states.

Aberrations existed in microscopes lead to a reduction in imaging quality which can be corrected using AO. Details of various aberration correction techniques are discussed in most parts of this thesis.

2.1.3 Polarisation state

2.1.3.1 Jones calculus

E in the expression 2.3 can be decomposed into x and y two components such that

$$\vec{E} = |E| \begin{pmatrix} |q_x| e^{j\theta_x} \\ |q_y| e^{j\theta_y} \end{pmatrix} e^{j(\omega t - kz)} = |E| e^{j(\omega t - kz)} \begin{pmatrix} q_x \\ q_y \end{pmatrix} = |E| e^{j(\omega t - kz)} J_Q \quad (2.5)$$

where $|E|$ is the amplitude of the wave and $|E|^2$ is the intensity of the light. q_x and q_y are the normalised x and y components such that $|q_x|^2 + |q_y|^2 = 1$; θ_x and θ_y are the phase of q_x and q_y respectively. $J_Q = (q_x, q_y)^T$ is the Jones vector which indicates the polarisation state of the light (see Table 2.1).

The Jones matrix can be used to represent polarisation property of optical element (see Table 2.2). Polarised light will have its polarisation state (represented by Jones vector J) modulated after passing through these elements.

$$J' = \mathcal{J} J$$

J' is the new polarisation state after passing through an element with polarisation

Optical elements	Rotators	Linear polarisers	Phase shifters
Jones matrix ($\mathcal{J}$)	$R = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$	$R \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} R^{-1}$	$R \begin{pmatrix} e^{j\phi/2} & 0 \\ 0 & e^{-j\phi/2} \end{pmatrix} R^{-1}$

Table 2.2: Jones matrix ($\mathcal{J}$) of polarising elements - rotators, polarisers and phase shifters. R is defined as rotators. α is the angle between the target axis and x-axis. A change of axis for polarisers and phase shifters can be realised through using rotators. ϕ is the retardance applied by the phase shifter between the fast and slow axis.

property defined in $\mathcal{J}$.

Note that Jones calculus is able to represent phase information so it is rather useful in coherent wave analysis. Jones calculus is unable to represent depolarisation.

2.1.3.2 Stokes parameters and Müller matrices

Polarisation states can also be expressed in the Stokes vector (S) such that

$$S = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} |q_x|^2 + |q_y|^2 \\ |q_x|^2 - |q_y|^2 \\ 2\text{Re}(q_x^* q_y) \\ -2\text{Im}(q_x^* q_y) \end{pmatrix} \quad (2.6)$$

For fully polarised light, $S_1^2 + S_2^2 + S_3^2 = S_0^2$. When the light is not fully polarised, $S_1^2 + S_2^2 + S_3^2 \leq S_0^2$. A Stokes vector can thus reflect how polarised the radiation is while a Jones vector cannot show this property. Some examples of Stokes vectors were shown in Table 2.3.

A Müller matrix can be used to represent polarisation property of optical element (see Table 2.4). Polarised light will have its polarisation state (represented by Stokes

Polarisation state	Linear x	Linear y	Linear 45°	Linear -45°	Left-hand circular	Right-hand circular
Stokes vector (S)	$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$

Table 2.3: Examples of Stokes vectors (S) of selected polarisation states.

Optical elements	Müller matrix ($\mathcal{M}$)
Rotators	$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\alpha & -\sin 2\alpha & 0 \\ 0 & \sin 2\alpha & \cos 2\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
Linear polarisers	$R \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} R^{-1}$
Phase shifters	$R \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \phi & -\sin \phi \\ 0 & 0 & \sin \phi & \cos \phi \end{pmatrix} R^{-1}$

Table 2.4: Müller matrix ($\mathcal{M}$) of polarising elements - rotators, polarisers and phase shifters. R is defined as rotators. α is the angle between the target axis and x-axis. The axis of polarisers and phase shifters can be realised through using rotators. ϕ is the retardance applied by the phase shifter between the fast and slow axis.

vector S) modulated after passing through these elements.

$$S' = \mathcal{M}S$$

S' is the new polarisation state after passing through an element with polarisation property defined in $\mathcal{M}$.

2.1.3.3 Polarisation state transmission and phase difference

When a system is subjected to cyclic adiabatic processes, a phase difference is acquired over the course of a cycle. This phenomenon was first discovered separately by Kato (1950) [22], Pancharatnam (1956) [23] and Longuet-Higgins (1958) [24] and later generalised by Berry (1984) [25,26] who obtained results not just limited to quantum mechanics but also extended to wave physics.

This phenomenon can thus be interpreted in the context of light properties. Light passing through a polarising elements will have its polarisation state together with its phase state changed. The change in the phase state is a sum of two components - dynamic phase and geometric phase (which is also called as Pancharatnam-Berry (PB) phase) where dynamic phase is subject to the wavelength, speed and distance of propagation which can be very sensitive to the experimental setup while geometric phase relates to the orientation of the eigenstate of the system. A more detailed discussion of this phase difference was included in chapter 7.

2.1.3.4 Poincaré sphere

Polarisation states can be virtually shown on a Poincaré sphere with x, y and z co-ordinates equals to S_1 , S_2 and S_3 which are defined as linear x, linear 45° and left circular polarisation respectively (see Figure 2.1). Therefore, all the polarisation states can be indicated on the Poincaré sphere. A full Poincaré beam means a spatially varying optical field that maps to the entire Poincaré sphere. Poincaré sphere is rather useful to visualise polarisation states and their modulations when a polarised beam passes through a polarising optics. For example, the polarisation state conversion after passing through a waveplate with extraordinary axis (Q) and retardance (Ψ) can be visualised on the Poincaré sphere as a rotation of the sphere about the axis, Q, with an angle of Ψ . The geometric phase introduced by the waveplate during the polarisation state modulation can also be visualised as the area of a geodesic

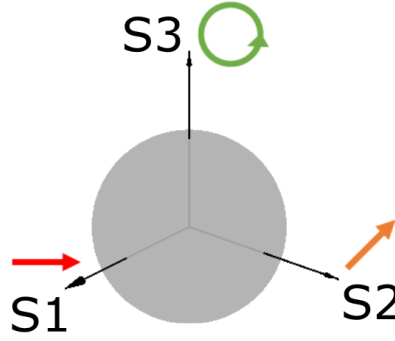


Figure 2.1: The three spatial co-ordinates on the Poincaré sphere, S1, S2 and S3, represent linear x, linear 45° and left circular respectively.

triangle connecting the initial and final polarisation state and the ordinary axis of the waveplate, -Q under the waveplate model described in the section 7.1.

Jones calculus, Stokes vectors, and the Poincaré sphere are important mathematical tools studying polarisation states and modulations which assisted the analysis and discussions greatly in the chapter 7.

2.1.4 Diffraction and Fourier optics

2.1.4.1 Diffraction and approximations

Huygens' principle stated that each point on a primary wavefront can be considered as the source of a secondary wave and a secondary wavefront can be constructed as the envelope of these secondary spherical wavefronts [27]. Fresnel provided a mathematical description of Huygens' principle which was later corrected by Kirchhoff to satisfy Maxwell's equations.

$$E(x_2, y_2) = -\frac{j}{\lambda} \iint E(x_1, y_1) \left[\frac{1 + \cos(r, z)}{2} \right] \frac{e^{jkr}}{r} dx_1 dy_1 \quad (2.7)$$

where $\cos(r, z)$ represents the cosine of the angle between r and z ; the term in the square bracket is the obliquity factor Kirchhoff added to account for the directionality of the secondary propagation. x_1, y_1 and x_2, y_2 are the spatial coordinates of the

primary and secondary plane. r is the distance between a pair of points (x_1, y_1) and (x_2, y_2) and defined as $r = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + z^2}$.

Fresnel further simplified the diffraction equation by assuming that (1) z is sufficiently large to approximate r in the denominator of the expression and (2) $|x_2 - x_1|$ and $|y_2 - y_1|$ are small enough so that $r \approx z + [(x_2 - x_1)^2 + (y_2 - y_1)^2]/2z$. The expression 2.7 is now simplified as

$$E(x_2, y_2) \approx -j \frac{e^{jkz}}{\lambda z} \iint E(x_1, y_1) e^{j \frac{k}{2z} [(x_2 - x_1)^2 + (y_2 - y_1)^2]} dx_1 dy_1 \quad (2.8)$$

Fraunhofer took some further simplifications by assuming the secondary plane is sufficiently far away such that $z \gg k(x_1^2 + y_1^2)_{max}/2$ and thus the expression can be approximated as

$$E(x_2, y_2) \approx -j \frac{e^{jkz}}{\lambda z} e^{j \frac{k}{2z} (x_2^2 + y_2^2)} \iint E(x_1, y_1) e^{-j \frac{k}{z} (x_2 x_1 + y_2 y_1)} dx_1 dy_1 \quad (2.9)$$

It is interesting to note that Fraunhofer approximation is in fact a Fourier transform with a premultiplying phase factor. If we create a new set of spatial coordinates such that $u = \frac{kx_2}{z}$ and $v = \frac{ky_2}{z}$, we can write

$$\begin{aligned} E(u, v) &\propto \iint E(x_1, y_1) e^{-j(u x_1 + v y_1)} dx_1 dy_1 \\ &\propto \mathcal{FT}[E(x_1, y_1)] \end{aligned}$$

where $\mathcal{FT}$ is the Fourier transform operation.

The Fresnel number, F is defined as

$$F = \frac{(x_1^2 + y_1^2)_{max}}{z \lambda} \quad (2.10)$$

where $(x_1^2 + y_1^2)_{max}$ is the largest value of $(x_1^2 + y_1^2)$ (radius of the incoming beam in the

case of a circular aperture). F can help to evaluate the validity of approximations; the Fresnel approximation (equation 2.8) is suitable to be used when $F \approx 1$ while the Fraunhofer approximation (equation 2.9) is valid when $F \ll 1$.

To allow a physical sense of F , suppose there is a beam with a wavelength of $500nm$ and a cross section radius of $5mm$ (which is similar to the back pupil size of an objective lens), for $F=1$, $z = 50m$. Therefore, for this beam, the Fresnel approximation is valid if the diffraction field is calculated at a distance of about several tens of metres; the Fraunhofer approximation is suitable if the distance is much greater than tens of metres.

2.1.4.2 Thin lens model

For a thin circular convex (converging) lens, using paraxial approximation, it can be shown that the transmission function of the lens, $t(x, y)$, is expressed as

$$t(x, y) = P_{amp}(x, y)e^{-j\frac{k}{2f}(x^2+y^2)}$$

where f is the focal length and $P_{amp}(x, y)$ is the pupil function of the lens.

$$P_{amp}(x, y) = \begin{cases} 1 & \text{for } \sqrt{x^2 + y^2} \leq 1 \\ 0 & \text{for } \sqrt{x^2 + y^2} > 1 \end{cases}$$

Using the Fresnel approximation of the diffraction theory (equation 2.8) and paraxial approximation of the transmission function of the lens [1], at a distance

of one focal length from the lens, i.e. $z = f$

$$\begin{aligned}
 E(x_2, y_2) &\approx j \frac{e^{-jkf}}{\lambda z} e^{j\frac{k}{2f}(x_2^2+y_2^2)} \iint E(x_1, y_1) t(x_1, y_1) e^{j\frac{k}{2f}[x_1^2+y_1^2-2x_1x_2-2y_1y_2]} dx_1 dy_1 \\
 &\propto \iint E(x_1, y_1) P_{amp}(x_1, y_1) e^{-j\frac{k}{f}(x_1x_2+y_1y_2)} dx_1 dy_1
 \end{aligned} \tag{2.11}$$

Once again, by introducing a new set of spatial coordinates $u = \frac{kx_2}{f}$ and $v = \frac{ky_2}{f}$, we can write

$$E(u, v) \propto \mathcal{FT}[E(x_1, y_1) P_{amp}(x_1, y_1)]$$

which is mathematically expressed as a Fourier transform operation.

The thin lens model (equation 2.11) is only valid when NA is smaller than ~ 0.5 and the paraxial approximation is valid [1, 7]. NA is defined as

$$\text{NA} = n \frac{\sqrt{(x^2 + y^2)_{max}}}{f} \tag{2.12}$$

2.1.4.3 High numerical aperture lens model

For high NA lenses, the contributions of the axially polarized E field in the focal volume cannot be ignored, while low NA models cannot describe the rotation of the E field from the entrance pupil to the focal volume. Vectorial theory is thus required for high NA lenses [6–9]. The general idea of the vectorial diffraction is to decompose the E field at the entrance pupil into x and y components (EX and EY) and the E field after passing through the high NA lens is decomposed into lateral (x-y) and axial (z) components and analyse each component separately. At the focal plane, the

total $E(u, v)$ field is

$$E(u, v) = \begin{pmatrix} E(u, v) \cdot \hat{i} \\ E(u, v) \cdot \hat{j} \\ E(u, v) \cdot \hat{k} \end{pmatrix} = \begin{pmatrix} EX(u, v) \cdot \hat{i} \\ EX(u, v) \cdot \hat{j} \\ EX(u, v) \cdot \hat{k} \end{pmatrix} + \begin{pmatrix} EY(u, v) \cdot \hat{i} \\ EY(u, v) \cdot \hat{j} \\ EY(u, v) \cdot \hat{k} \end{pmatrix} \quad (2.13)$$

where $\hat{i}$, $\hat{j}$ and $\hat{k}$ are three unit vectors indicating x, y and z directions; $E(u, v)$, $EX(u, v)$ and $EY(u, v)$ are three dimensional fields that can be expressed as functions of u and v ; u and v are the two-dimensional lateral spatial coordinates.

$EX(u, v)$ is the field at the focal plane calculated from the x component of E field at the entrance pupil which can be expressed as

$$EX(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[e^{jk_z f} E_{1x}(k_x, k_y) \frac{1}{\cos \theta} \right] e^{j(k_x u + k_y v)} dk_x dk_y \quad (2.14)$$

$$= \mathcal{FT} \left[e^{jk_z f} E_{1x}(k_x, k_y) \frac{1}{\cos \theta} \right] \quad (2.15)$$

where k_x , k_y and k_z are spatial coordinates in the k space on a spherical cap immediately after passing through the high NA lens such that $\sqrt{k_x^2 + k_y^2 + k_z^2} = k = \frac{2\pi}{\lambda}$. f is the focal length of the lens. $e^{jk_z f}$ is a phase constant for any lateral plane. θ is an angle and its cosine is defined as $\cos \theta = \frac{k_z}{k}$.

$E_{1x}(k_x, k_y)$ is the E field on the spherical cap at (k_x, k_y) which consists of three components

$$\begin{pmatrix} E_{1x}(k_x, k_y) \cdot \hat{i} \\ E_{1x}(k_x, k_y) \cdot \hat{j} \\ E_{1x}(k_x, k_y) \cdot \hat{k} \end{pmatrix} = E_{0x}(x, y) \begin{pmatrix} \sqrt{\cos \theta} [\cos \theta + \sin^2 \phi (1 - \cos \theta)] \\ \sqrt{\cos \theta} [(\cos \theta - 1) \sin \phi \cos \phi] \\ -\sqrt{\cos \theta} \sin \theta \cos \phi \end{pmatrix}$$

where ϕ is another angle and its cosine is defined as $\cos \phi = \frac{k_x}{k \sin \theta}$. $E_{0x}(x, y)$ is the E field at the entrance pupil; we normally assume a uniform illumination and thus

$|E_{0x}|$ is a constant across the pupil.

Similarly, we can derive $EY(u, v)$ from $E_{0y}(x, y)$ by simply rotating the axis by $\angle 90^\circ$ such that the new spatial coordinates $(x', y') = (-y, x)$.

We have thus shown the high NA lens model which can be used to calculate the three dimensional E field at the focal plane from the pupil plane. Though this vectorial diffraction theory should be used whenever a high NA lens is included for analysis, in practice when polarisation state is not included in discussion and especially in some applications (such as in some senseless AO methods discussed in the chapter 4 and 5 in this thesis) when total intensity is the only property of interest, the simple thin lens Fourier optics analysis would obtain very similar results and thus be favoured for its simplicity.

2.2 Microscopy

2.2.1 The invention of optical compound microscopy

Optical microscopy is an essential tool in biomedical imaging which reveals structure and function on a sub cellular scale, whilst being able to image multiple cells and tissues over volumes. The first compound microscopy was recorded to be invented by Dutch spectacle-maker Hans Jansen and his son Zacharias [28]. Hooke in *Micrographia* (1665) [29], the first book of microscopy by scientists, showed many drawings of biological samples, which were believed to be what he saw using his refined microscope. In the book, he also used the term ‘cells’ for the first time to address some porous structures. Van Leeuwenhoek (1632–1723) made more than 500 optical lenses and at least 25 refined microscopes of high quality at the time. He was also the first to describe protozoa (1674) [30] and bacteria (1683) [31]. Figure 2.2 shows the light beam passing through a compound microscope and forming a magnified virtual image of the object viewed by an eye.

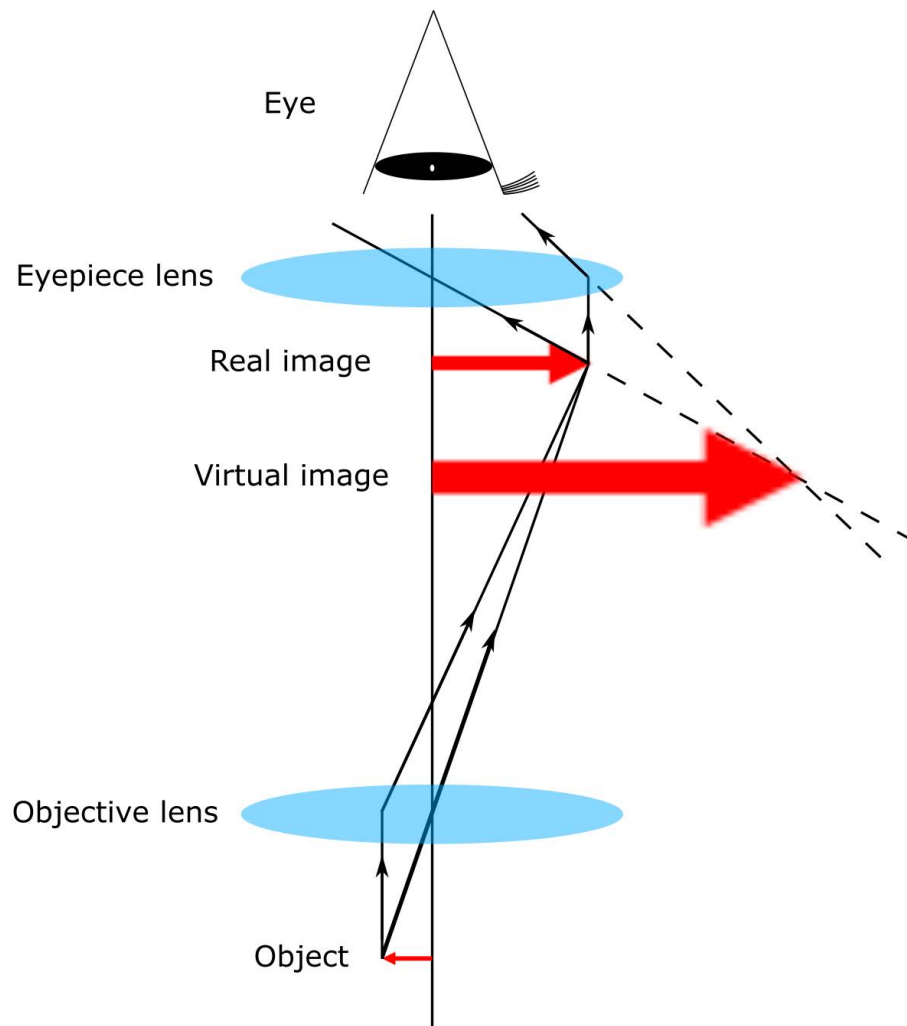


Figure 2.2: A schematic plot of a compound microscope when imaging a small object. The objective lens forms a magnified inverted real image from the object. The eyepiece lens forms a virtual magnified image from the real image. The virtual image is viewed by an eye. The magnification of the object by this compound microscope is the magnification by the objective lens times the the magnification by the eyepiece.

2.2.2 Fluorescence microscopy

The first recorded application of the fluorescence phenomenon was to detect counterfeit wood wrote by Monardes in 1565 [32]. Herschel described fluorescent properties of quinine sulphate in 1845 [33] which marked the foundation of modern fluorescence technology. The phenomenon was only later named by Stokes as “fluorescence” in 1852 [34] when he observed that illuminating the mineral fluorspar by ultraviolet (UV) excitation emitted red light. He also noted that fluorescence always emitted light with a longer wavelength than the excitation light (Stokes shift). Baeyer created Fluorescein as a fluorescent dye in 1871 [35] and the method of fluorescence labelling has been developed since then. The first green fluorescence protein (GFP) was isolated by Shimomura from the jellyfish in 1960s [36] and was later cloned and developed as a tracer molecule by Prasher *et al.* in 1980s and 90s [37, 38]. Fluorescence proteins led to a breakthrough in live cell imaging for their ability to label genetic protein regions selectively and thus to image protein mechanisms and functions.

The first fluorescence microscope was constructed in 1911 by Heimstädt [39] and this imaging technology were later fastly developed and widely used in biomedical imaging for its high image contrast. The main difficulties of building a fluorescence microscope were 1. concentrating sufficient illumination excitation light with the right wavelengths on the sample and 2. collecting emitted fluorescence free of noise. The first epi-fluorescence¹ microscope was designed by Ellinger and Hirt in 1929 [40] which tackled the two challenges. The excitation light passed through several filters for the right wavelength reaching the objective lens and illuminating the sample tissue. A yellow filter was positioned between the objective and eyepiece to block most reflected illumination light from interfering the observations. Later in about 40 years time, the invention of dichroic mirror led another major advance in fluorescence microscope [41].

¹The term “epi” is from Greek, meaning “same” and stands for both the excitation and emission light passing through the same objective lens.

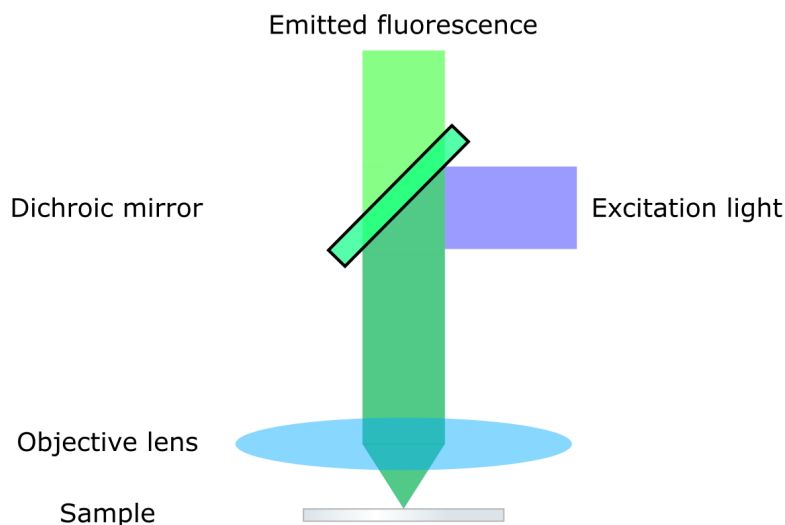


Figure 2.3: An schematic plot of a epi-fluorescence microscope with a dichroic mirror included. Excitation light is reflected by the dichroic mirror into the objective and the emitted fluorescence with a different wavelength is transmitted by the dichroic mirror.

Comparing to conventional filters which absorb certain wavelengths, dichroic mirrors work more efficiently by reflecting certain wavelengths while letting the others pass; the excitation beam is reflected by the dichroic into the objective and the emitted fluorescence of a different wavelength passes through the dichroic to reach the eyepiece (see Figure 2.3).

Fluorescence microscopes maximise signal to noise (background) ratio by attaching fluorescent molecules to the structures of interest and collecting light emitted only by those structures which led great advancements in microscopy.

2.2.3 Confocal microscopy

Widefield microscopy was the major form of microscopy in the early ages until the invention of confocal microscopy during the 1960s and further developments during 70s and 80s by the collective contributions of several researchers all over the world [42–44]. Widefield means the entire volume of interest in the sample is exposed to the light source and results the whole volume being imaged simultaneously with out of focus

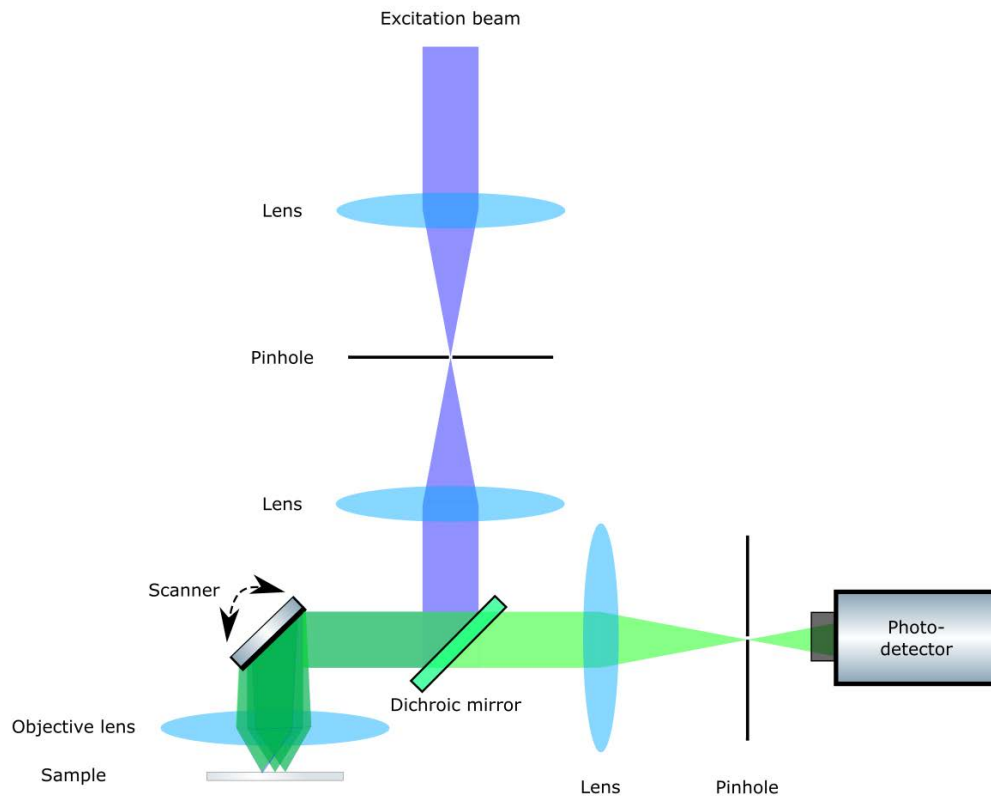


Figure 2.4: A schematic plot of a confocal microscope. Excitation beam passes through a pinhole before being reflected by a dichroic mirror onto a scanner which scans the beam across the sample. The emitted fluorescence is descanned and passes through the dichroic and another pinhole before reaching the signal detector. The scanner is normally positioned at a conjugate pupil plane which can be accessed through a 4-f system (which was not included in the figure) such that the back pupil plane is re-generated away from the objective.

planes being blurred. In contrary, confocal microscopy eliminates out of focus signal by introducing a pinhole which is comparative to the size of a diffraction limited point spread function (PSF); only a small area on a thin slice of the sample is illuminated each time and only fluorescence emitted from the illuminated region is collected. A scanning mechanism (either through scanning the beam across the sample or through moving the sample around the beam) is used so that the fluorescence signal is collected pixel by pixel and assembled into an image later. By doing so, it allows optical sectioning in axial direction. See Figure 2.4 for a design of a confocal microscope.

2.2.4 Nonlinear optical microscopy

Different from conventional microscopy which typically uses linear interactions, nonlinear microscopy uses nonlinear light-matter interactions to produce images. Two-photon (2-P) absorption was first discussed by Göppert-Mayer [45] in 1931 but was only observed thirty years later by Kaiser and Garrett [46]. 2-P absorption is a phenomenon that two photons of a longer wavelength are absorbed simultaneously by a molecule which then deexcite to release one photon of a short wavelength which is almost half of the wavelength of the excitation light. The first 2-P microscope was not feasible until 1990 by Denk *et al.* [47] when mode-locked femtosecond laser was available; femtosecond laser delivers much higher photon density than conventional laser to allow sufficient 2-P excitation events. The advantage of 2-P microscope is that it allows optical sectioning without a pinhole and it greatly reduced out-of-focus photobleaching. Long wavelength excitation beams also suffers less tissue scattering and travels deeper inside the tissues which allows deep tissue imaging. See Figure 2.5 for a design of a 2-P microscope.

Later, in 1996, Xu *et al.* first applied three-photon (3-P) excitation in laser scanning microscopy [48]. As the name suggested, 3-P excitation is similar to a 2-P excitation apart from three photons, instead of two, are absorbed simultaneously at each time. 3-P imaging can use even longer excitation wavelengths than 2-P which thus suffers less tissue scattering and is extremely useful especially in deep imaging. In 2013, Horton *et al.* demonstrated of using a 3-P microscope to image an intact mouse brain *in vivo* [49]. Another advantage of 3-P excitation than 2-P is that it has a even thinner optical sectioning as fluorescence signal drops faster for out-of-focus planes. However, since 3-P imaging requires a higher photon density, in-focus photobleaching is more likely to happen comparing to a 2-P system [50].

The phenomenon of second harmonic generation (SHG) was first demonstrated by Franken *et al.* in 1961 [51] and later observed in biological samples by Fine and

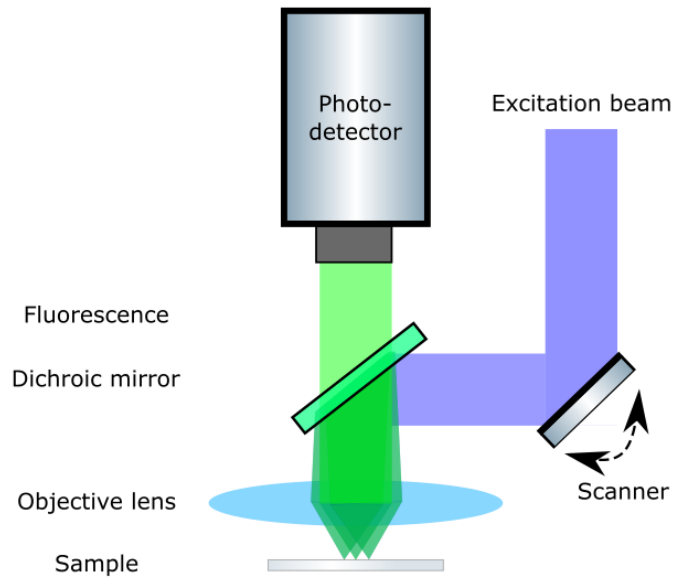


Figure 2.5: A schematic plot of a 2-P microscope. Excitation beam is scanned by a scanner before being reflected by a dichroic mirror into the objective. The emitted fluorescence passes through the dichroic before reaching the signal detector. The scanner is normally positioned at a conjugate pupil plane which can be accessed through a 4-f system (which was not included in the figure).

Hansen in 1971 [52]. The first SHG microscopy was demonstrated by Hellwarth and Christensen in 1971 [53] and was later used for biological imaging in 1986 by Freund and Deutsch [54]. Although SHG also requires two photon conversion, it is significantly different from 2-P; it does not require fluorophore labelled samples; after the conversion, the generated new photon contains exactly twice the energy (half the wavelength) and conserves the coherence of the excitation beam. Similar to 2-P imaging, SHG is also suitable for deep imaging and allows optical sectioning. Since no fluorophore was used, SHG does not suffer from photobleaching. An disadvantage of SHG is that only limited materials are capable of generating second harmonics. Note that there are also third harmonic generation (THG) optical microscopes which are not elaborated in this section since they are not commonly used in biomedical imaging.

The first coherent anti-Stokes Raman scattering (CARS) observation was reported

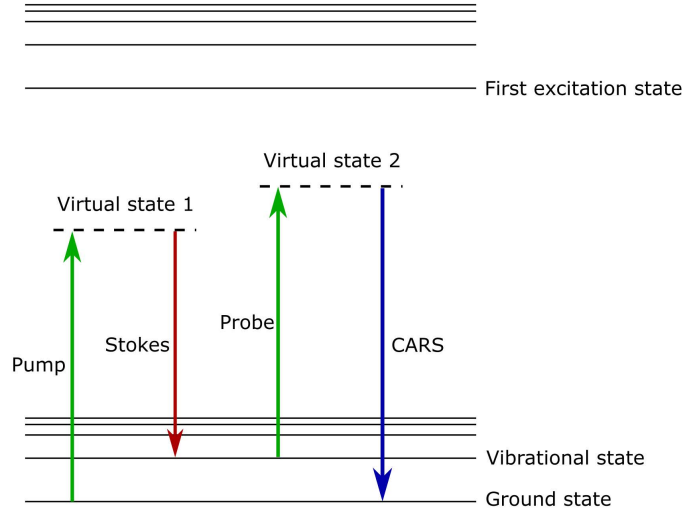


Figure 2.6: Energy band diagram for CARS imaging process. A pump beam first raise molecules from ground state to the first virtual state. A Stokes beam, which has a smaller energy than the pump beam, stimulate the depopulation of the virtual state to a vibrational state. A probe beam, which has the same wavelength as the pump beam, bring molecules from vibrational state to the second virtual state. The molecules relax from the second virtual state back to the ground state and release CARS signals for imaging.

by Maker and Terhune in 1965 of which they referred to as “three wave mixing experiments” [55]. This phenomenon is better explained by an energy band diagram in Figure 2.6 which can be simply described as molecules emit light with higher energy and a shorter wavelength than that of a pump beam, thus anti-Stokes scattering. CARS was first applied on biological imaging by Begley *et al.* in 1974 [56]. The advantage of CARS microscope is that no labelling of specimens is required which avoids phototoxicity and photobleaching.

Because non-linear microscopes normally use long excitation wavelengths, they are good candidates for deep tissue imaging which also make them more likely to suffer from sample induced scattering and aberrations. In this thesis, a 2-P microscope was used for sensorless AO methods demonstration in chapter 5. Since many of the non-linear microscopes (apart from CARS) have very similar optical designs, we are expecting the results and conclusions obtained from the 2-P microscope can be readily

extended to other non-linear microscopes.

2.2.5 High-resolution and super-resolution microscopy

In the late twentieth and twenty-first centuries, high-resolution microscopes like 4-pi microscope and super-resolution microscopes like STED, SIM and single molecule localisation microscopy (SMLM) create images with resolution better than the diffraction limit and improve the resolution to a few or tens of nanometers.

The 4-pi microscope was first described by Hell and Stelzer in 1992 [57] which has an axial resolution up to seven times higher and also holds a higher lateral resolution than that of a confocal fluorescence microscope. The main idea is to have two objective lens facing each other so that the illumination beams coming from both objectives interfere with each other and thus generate an effective focal spot which is smaller than the focal spot generated from a single objective. The 4-pi optical design can be implemented in conjunction with a confocal [58, 59], 2-P [60], STED [61] and SMLM [62].

The STED microscope was first described by Hell and Wichmann in 1994 [63] and demonstrated by Klar *et al.* in 1999 [64]. A ring shaped depletion beam was generated at the focal plane to overlap with the excitation focal spot. The excitation beam with a diffraction limited size excites molecules which are then selectively depleted by the depletion beam. Stimulated emission happens before fluorescence emission which relax excited molecules to emit photons of a different wavelength as normal fluorescence emissions. Since the depletion beam has a ring shape, molecules at the central void of the depletion beam are more likely to emit fluorescence than molecules on the depletion ring. By increasing the intensity of the depletion beam, fewer molecules can emit fluorescence; eventually the depletion beam saturates such that all the molecules on the ring are depleted and only molecules in the central void have a chance of emit fluorescence. By separating the fluorescence emission from the

stimulated emission, we can collect signals only from the central void of the depletion beam which is smaller than the diffraction limited size of the excitation focal spot.

Gustafsson proposed and realised a SIM microscope using spatially structured illumination beam to extract high frequency imaging components [65]. In a conventional widefield microscope, the maximum spatial frequency of the sample that can be observed depends on the diffraction limited size of the system focal spot. Structured illumination encodes a new set of high spatial frequency information in the form of moiré fringes. Several images can be collected when different illumination structures with different orientations and frequencies are used for imaging. High spatial frequency information of all directions are thus obtained and reconstructed into a super-resolution image. The high temporal resolution and the low illumination power allow SIM to be suitable in live-cell imaging.

Exploiting the known ability to localise single molecules with nanometer precision, a few similar single molecule approaches were proposed independently by different researchers in 2006 to achieve super-resolution imaging [66–68]. By separating different fluorophore emissions in time, one molecule can be distinguished from the neighbouring molecules. Fluorophores are activated stochastically by light and a large number of images each containing different active molecules are collected. For each image, the density of excited molecules is kept low so that no molecule overlaps with its neighbours. The locations of the molecules in the sequence of images are identified and used to build up a super-resolution image.

High-resolution and super-resolution microscopes tend to be more sensitive to aberrations than conventional microscopes. For example, piston, tip and tilt normally do not affect imaging quality in conventional microscopes but they can seriously influence the 4-pi and STED microscopes [69–71]. Furthermore, STED and SIM microscopes are also affected by polarisation state errors [12–14]. These observations are discussed in chapter 4 and 7.

2.3 Adaptive elements

2.3.1 Adaptive optics devices

Microscopy imaging quality suffers from phase and polarisation state errors [11, 19, 72–75]. With aberrations (especially phase aberrations), the focal spot is deformed and is no longer diffraction limited which reduce the imaging resolution of the system². These aberrations (both phase and polarisation state) may be corrected by suitably chosen adaptive elements. In this section, we discussed some major AO devices which are widely used in microscopes. Note that there are also other kinds of adaptive elements such as fluid-filled devices [76–78] which are not elaborated since they are less widely used in practice comparing to the major devices mentioned in this section.

2.3.1.1 Objective lens correction collar

Objective lens correction collar is one of the simplest AO which corrects spherical aberration introduced by sample coverslips or immersion fluid refractive index mismatch through adjusting central lens group position. This simple adaptive element can be rather useful when more complex AO devices (such as DM and SLM) are unavailable; it is also often desirable to be used in conjunction with continuous membrane DMs which have limited actuator stroke while unable to display wrapped phases and thus less desirable when correcting a large spherical aberration. Conventional objective lenses have manually tunable correction collar. In recent years, there are objective lenses equipped with motorised digitally controllable correction collars which assist an easier and more accurate operation.

²The effects of aberrations on the focal spots can be easily shown by the (thin or high NA) lens models in the section 2.1.4.2 and 2.1.4.3

2.3.1.2 Deformable mirror

A DM corrects phase aberrations through deforming the mirror shape and thus modulate the wavefront [79]. Its performance does not depend on polarisation states or wavelengths which makes it useful in fluorescence imaging. micro-electro-mechanical systems (MEMS) DMs with continuous reflective membrane normally do not have confined actuator influence which makes it difficult to be operated directly. Calibrations are often needed to before the device used for aberration corrections.

The aim of DM calibration is to calculate the relationship between the control signal applied to the mirror and the changes of the mirror shape. A control signal is applied to each actuator one at a time in sequence and the deformations of the DM are measured using a wavefront sensor. A control matrix, connecting the control signal to DM deformation, is calculated and inverted for later aberration correction usage.

There are also segmented DM consisting of many small rigid mirror sections that can be controlled individually. Therefore, the influence function of one actuator is localised. There are unavoidable gaps between mirror segments cause light and diffraction losses especially when neighbouring mirror segments have step changes on the surface. This can be reduced if there are three actuators per mirror such that each mirror can be tilted to minimise gaps.

There are also other DM concepts (such as bimorph or ferrofluid concepts) which are not discussed here in details since they are not commonly used in optical microscopy.

2.3.1.3 Spatial light modulator

A liquid crystal (LC) SLM contains a two-dimensional panel of independently controllable uniaxial birefringent pixels [80]. When a varying voltage applied, each pixel can change the OPL along its extraordinary axis . Light with components along the

extraordinary axis is thus modulated. Therefore, a LC SLM can be used as a polarisation state modulator. When a SLM is used as a phase only device, the input beam needs to have a linear polarisation state along the extraordinary axis. Since SLMs are polarisation state and wavelength dependent, they are mostly suitable of correcting aberrations of the laser beams but not incoherent light source or fluorescence.

2.3.2 Adaptive optics operations

The general operation principle of AO as a phase aberration compensator is to position the AO at the conjugate pupil plane and introduce a spatially varying OPL to the beam such that sections of the beam with phases ahead of the rest of the beam pass a longer OPL when reflected by and transmitted through the AO; the aberrated pupil wavefront is thus restored to a flat wavefront by the AO and the focal spot is shaped to its diffraction limited size. Many approaches have been examined to implement AO in different types of microscopes, which may be classified generally as sensed and sensorless aberration correction methods.

2.3.2.1 Sensed operations

As the name suggested, sensed methods include employ wavefront sensors for phase aberration measurements. There are commonly two types of wavefront sensors used in optical systems: Shack–Hartmann and interferometric wavefront sensor.

A Hartmann-Shack sensor [81,82] consists of a lenslet array, where each lens has the same focal length. A camera is used to image all the focal spots generated by the lenslet array when light with a wavefront for measurements passes through the lenslets. If the wavefront is totally flat, all the focal spots should position on the optical axes of all the lenslets (see Figure 2.7 (a)). The distances of the shifts of the focal spots represent the average local gradient of that section of the wavefront (see Figure 2.7 (b)).

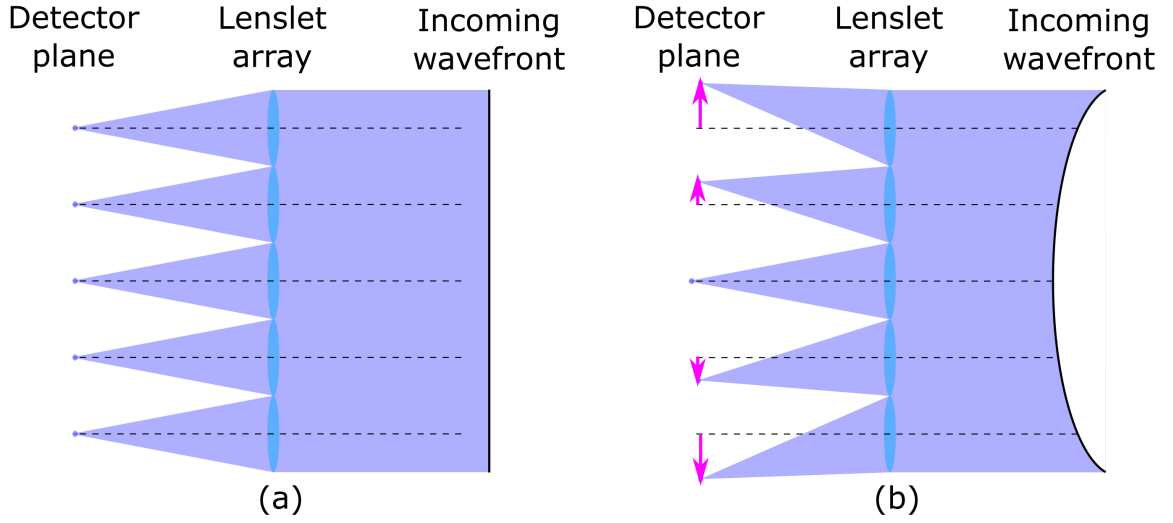


Figure 2.7: A demonstration of a Hartmann-Shack sensor used for wavefront sensing. (a) is when a beam with a flat wavefront comes into the sensor; all the focal spots are positioned on the optical axes of the lenslets. (b) is when a beam with a non-flat wavefront passes through the lenslets; the focal spots are shifted proportional to the average gradients of the section of the wavefront.

The working principle of an interferometric wavefront sensor is to make the beam for measurement interfere with a reference beam which has a flat wavefront. The interference patterns contain wavefront information which can be decoded through various ways. One way is to Fourier transform the interference patterns and the wavefront of the beam for measurement is encoded in the areas around the first order side lobes of the Fourier transformed interference pattern (which correspond to the frequency of the interference patterns). Shift one of the side lobes back to the origin and inverse Fourier transform reveals the wavefront information of the beam of interest [83]. Note that the interferometric wavefront sensor normally implemented on measuring the wavefront of the illumination beam or DM calibrations as the operation requires spatially coherent light.

2.3.2.2 Sensorless operations

Indirect wavefront sensing or sensorless operations describe techniques of correcting aberrations using adaptive elements without a dedicated wavefront sensor mentioned

above. Output images are used to calculate some quality metric which serves as a feedback signal in order to control the optics. Sensorless operations normally correct aberrations iteratively and thus require more sample exposures. However, because these kinds of methods require no extra hardware on direct wavefront sensing and hence allow simpler optical design, they are often chosen in many microscopy designs and applications. There are also cases when fluorescence signal level is too low, sensed methods become infeasible, leaving sensorless methods the only option.

2.3.2.3 Zernike polynomials

Zernike polynomials were first proposed by Zernike in 1934 [84] which are derived from classical optical imaging through systems containing lenses and mirrors of finite diameter. The polynomials are orthogonal over a unit circular area which are often used as the basis modes for aberration corrections.

For a given radial and angular order m and n , Zernike modes can be defined as

$$\begin{cases} Z_n^m(\rho, \varphi) &= \sqrt{2n+1} R_n^m(\rho) \cos(m\varphi) \\ Z_n^{-m}(\rho, \varphi) &= \sqrt{2n+1} R_n^m(\rho) \sin(m\varphi) \\ Z_n^0(\rho, \varphi) &= \sqrt{n+1} R_n^0(\rho) \end{cases} \quad (2.16)$$

where $n \geq m \geq 0$. ρ is the radial coordinate ($0 \leq \rho \leq 1$). φ is the azimuthal angle ($-\pi < \varphi \leq \pi$). Z_n^m are called even modes because of the cosine term and similarly Z_n^{-m} are called odd modes due to the sine term.

The radial polynomial $R_n^m(\rho)$ is defined as

$$R_n^m(\rho) = \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k! \left(\frac{n+m}{2}\right)! \left(\frac{n-m}{2}\right)!} \rho^{n-2k} \quad (2.17)$$

Noll introduced a single valued index, j , to sequence the Zernike polynomials of

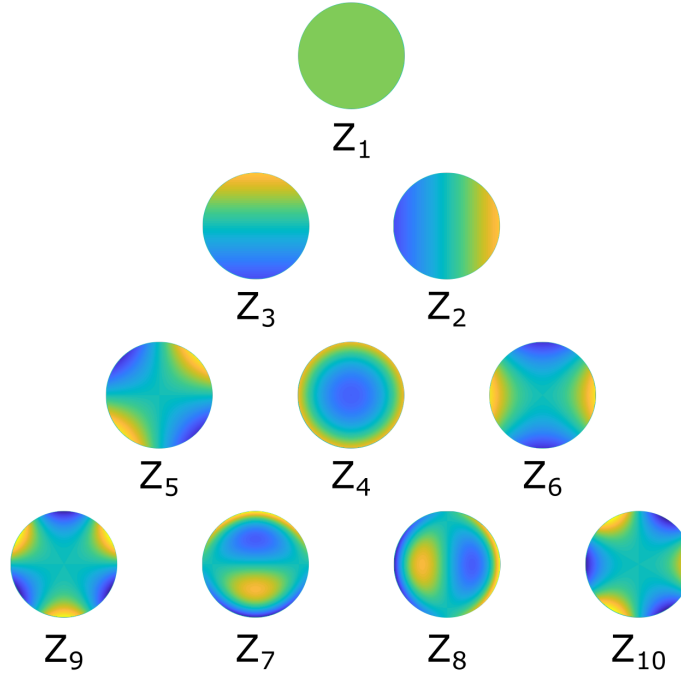


Figure 2.8: The first 10 Zernike polynomials in Noll's sequential index. The modes in the same row have the same radial order n . The modes sit in the central columns have small angular order m . Left-hand side are odd functions while right-hand side are even functions. For the equations to generate these modes and their names and orders, please see the table 2.5.

different radial, n , and angular, l , orders. Note that l is the angular order with signs and m is angular order without signs where signs indicate whether the modes are odd (negative signs) or even (positive signs). $|l| = m$. The exact definition of the indices is included in [85].

2.4 Machine learning

The machine learning (ML) technology arises from the urge of processing and making a better usage of the vastly growing number of available data. Conventionally, people write explicit algorithms instructing computers to process data in certain ways; by doing so, only a limited number of useful information can be obtained since the extracted information depends highly on the defined algorithm people put into the

Noll's indice (j)	Z_n^l	Name	Functional expression
1	Z_0^0	Piston	1
2	Z_1^1	Tip	$2\rho \cos \theta$
3	Z_1^{-1}	Tilt	$2\rho \sin \theta$
4	Z_2^0	Defocus	$\sqrt{3}(2\rho^2 - 1)$
5	Z_2^{-2}	Oblique Astigmatism	$\sqrt{6}\rho^2 \sin(2\theta)$
6	Z_2^{-2}	Vertical Astigmatism	$\sqrt{6}\rho^2 \cos(2\theta)$
7	Z_3^{-1}	Vertical coma	$\sqrt{8}(3\rho^3 - 2\rho) \sin(\theta)$
8	Z_3^1	Horizontal coma	$\sqrt{8}(3\rho^3 - 2\rho) \cos(\theta)$
9	Z_3^{-3}	Vertical trefoil	$\sqrt{8}\rho^3 \sin(3\theta)$
10	Z_3^3	Oblique trefoil	$\sqrt{8}\rho^3 \cos(3\theta)$

Table 2.5: The first 10 Zernike polynomials.

computer; computer is also only capable of doing simple tasks identified by people through those explicit algorithms. These kinds of data processing are not making full use of the available data and thus deemed as not ideal.

ML based studies and technologies provided an alternative way for data processing. When using ML based methods, people only identify tasks for the computer without providing any explicit algorithms [86, 87]. By feeding in a vast number of training datasets, the computer studies, extracts useful common features and tries to map a complex model linking the inputs to their outputs. Therefore, many ML base technologies are closely related to computational statistics.

ML approaches can be generally classified into three categories: supervised learning, unsupervised learning and reinforcement learning.

2.4.1 Supervised learning

Supervised learning is when the learning process of the computer is supervised by humans. This is achieved through feeding in labelled datasets during the training process. The labels of the training data are the ground truth, which are normally the desired outputs. The computer uses a complex network structure, consists of linear and nonlinear operations, to try to map the inputs (data) to the outputs (labels).

Through iterative convergence, the accuracy for the computer to predict the correct outputs (data labels) increases; when the network achieves a high accuracy, it is then ready to help us predict outputs from some other non-labelled inputs.

Some examples of supervised learning including deep neural network (DNN) and convolutional neural network (CNN) which were shown to be rather effective especially in processing a large number of data and accomplishing complex tasks.

2.4.2 Unsupervised learning

As the name suggests, unsupervised learning is opposite to the supervised learning where the training data was not labelled. In many unsupervised learning, the tasks were also not identified for the computer. The computer studies the data (such as the sequence and the appearance frequencies) and try to reveal hidden patterns.

One important application of unsupervised learning is natural language processing where a large number of articles, paragraphs and texts are fed into the computer to allow it to work out the relationships among different words and try to decipher the meanings of human languages.

2.4.3 Reinforcement learning

Reinforcement learning is when the computer learns through continuously interacting with the dynamic environment and thus navigating the problem space. The task is set through providing positive feedback when the computer provides desirable moves and negative feedback when the moves are not wanted. Reinforcement learning requires both positive and negative feedback and thus is normally used in simulated environments (since in a real environment negative feedback may bring devastating consequences which are not desirable). Some examples of using reinforcement learning are learning to drive a vehicle or play a game.

In recent years, ML, especially deep learning has been studied and demonstrated

in image processing (such as denoise and deconvolution) and phase retrieval. This will be further discussed in chapter 6 where we also presented a ML based phase retrieval method which is largely different from other ML based methods in literature and has distinct advantages.

Chapter 3

An automated control system for spherical aberration and defocus compensation

Correction collars have been widely used on high numerical aperture microscope objective lenses to enable corrections of spherical aberration induced by coverglasses and immersion fluid mismatches. Although they have simple designs and functionalities, correction collars have their distinct advantages over other more complex aberration compensation AO devices such as DMs and SLMs. Correction collar implementations do not require access to back pupil plane which lead to simple optical designs; they are in general cheaper than other more complex adaptive elements; they are also often used in conjunction with continuous membrane DMs in the same optical systems because DMs have limited actuator strokes while unable to display wrapped phases and thus are less effective when used to correct large amplitude aberrations such as a large spherical aberration.

The operations of correction collars when compensating spherical aberrations are through moving the central lens group. However, the adjustments to correction collar

settings not just apply spherical aberration compensations but also introduce extra defocus [88]. We therefore need an axially movable stage to bring the sample back to focus. The correction collar can be considered as a simple AO element; the aim of it is to apply a phase correction that is opposite to the spherical aberration in the system.

A conventional correction collar is controlled manually. Inaccurate markings and mismatched immersion fluid (which would in effect lead to a different correction collar scale) are often problematic. Besides, as we mentioned above, tuning the correction collar leads to an axial shift of the focal spot (which is equivalent to adding a defocus into the back pupil plane of system). Since both spherical aberration and defocus result blurred images, it is difficult to distinguish which factor leads to a degraded output, and thus makes the operations of conventional correction collars even harder.

Objective lenses with motorised correction collars started to be manufactured by Zeiss in recent years; this type of optical lens allows digital controls of correction collars. To harness the advantages of this motorised objective lens, an automated aberration and focus correction system for microscopes was presented in this chapter. In conjunction with an image-based control system, this new objective lens can now be more easily operated than a conventional objective lens.

There are previous works done for a similar purpose in the literature about correcting spherical aberration and defocus using an objective lens correction collar and an axial direction movable stage [89,90]. However, the method relied on the variations of the total intensity of the images and thus the method is not applicable to imaging systems like a widefield microscope (of which the total intensity does not vary with spherical or defocus aberrations). The method in the literature also requires x-z (or y-z) imaging scanning which may not be available in some imaging systems which were designed mainly for x-y lateral plane imaging. It is thus the motivation of this chapter to present a different, more general and efficient method which can be

adapted to any imaging system without such limitation.

In this chapter, we first presented the optical system setup. An image based quality metric was introduced and discussed. An automated control of spherical and defocus correction which required a minimum of six measurements was then presented. Outputs from the correction process were included to show that the control system performed promisingly.

3.1 The experimental setup

A widefield fluorescence and reflection microscope was used in this project. This was chosen for simplicity of the set up and demonstration of the control algorithms. The setup of this microscope is shown in Figure 3.1. Two degrees of freedom were allowed in this setup. A motorised objective lens (Zeiss Objective LD LCI Plan-Apochromat $40\times/1.2$ Imm) was used to apply spherical aberration correction. A 3D stage (Piezo TRITOR 102 SG) was used to adjust sample position and hence refocus the optical system.

Note that the aberration correction is specified in the unit of mm in a range of $0.15\text{ }mm$ to $0.19\text{ }mm$; this means that the negative aberration applied by the correction collar corrected an equivalent aberration induced by a coverglass with thickness in a range of $0.15\text{ }mm$ to $0.19\text{ }mm$.

The Piezo 3D stage actuator used has an axial working range of $80\text{ }\mu m$.

3.2 An automated control system

The aim of this project was to design an automated control system that allowed a convenient use of the motorised objective lens. As mentioned before, adjusting an aberration correction collar also moves the focus [89]. This control system was hence required to be capable of both applying the best spherical correction and locating the

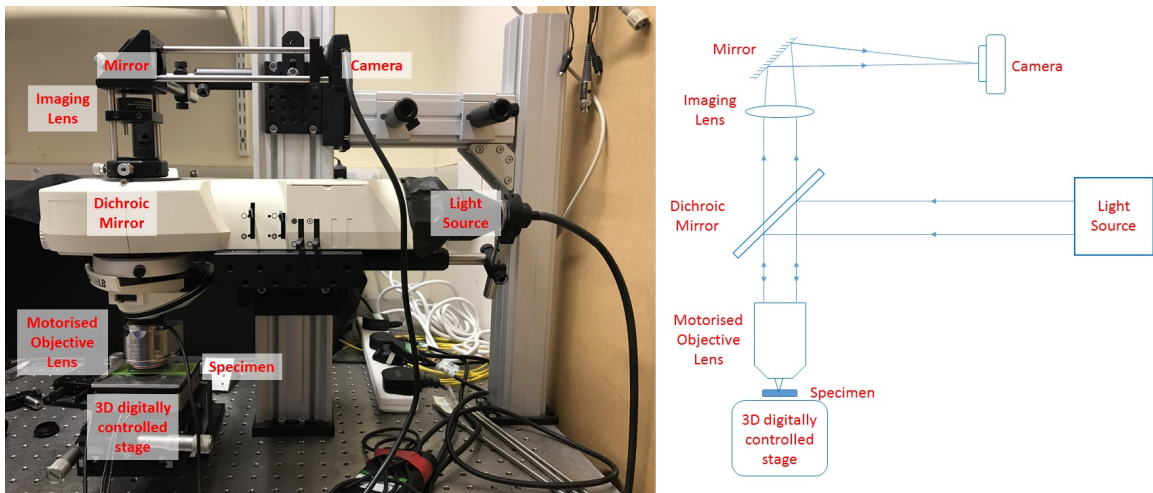


Figure 3.1: Left-hand side is the widefield microscopy setup for experiments. Right-hand side is a schematic plot of the setup. The light source generates visible light with a mixture of all wavelengths. Light with short wavelengths reflected down into the motorised objective lens (LD LCI Plan-Apochromat $40\times/1.2$ Imm) by a 510nm longpass dichroic mirror. Fluorescence emitted by the sample goes back into the objective lens, passes through the dichroic mirror and focused onto the camera (DCC1545M CMOS) by the imaging (tube) lens ($f=250\text{ mm}$). The microscope may also operate as a reflection microscope since the dichroic lets through some of the back reflected light. See the arrows on the schematic plot for the direction of the light propagation.

optimal focusing position for any sample of interest.

Before designing a control system to operate the devices, we needed to define a metric, calculated from the images, to quantify how well the the spherical aberration and defocus were corrected. The metric could then be optimised using a suitable algorithm.

3.2.1 An image based assessment of spherical aberration and defocus correction

Both spherical aberration and defocus affect the image quality. It is expected that the best images would be obtained when aberrations are minimised and when the system is in focus. Therefore, it is possible to determine an optimal aberration correction via optimising an appropriate image-based quality metric. [91]

Advantages of using such a metric are that the metric is easy to calculate and very effective. Since the outputs of all types of microscopes are images (or sequences of images), an image-based metric can be easily adapted to many other types of microscopes with little changes. This metric also requires no assumption about the optical system nor the sample structure for imaging.

The visibility of an image is affected by the contrast of the sample features to the background, the sizes of the features and the quality of the PSF; an object which has a higher contrast from its background, is larger in size and is imaged by a sharper PSF is more visible [92]. (See Figure 3.2.) The contrast of the image is affected by the illumination conditions; the size of features present depends on the choice of sample; only the quality of the imaging PSF is affected by the spherical aberration and defocus. The image based quality metric thus needs to be independent of the illumination condition and the sample structure. The image sharpness was chosen as the quality metric since smaller features and clearer edges can be resolved when the PSF quality is better.

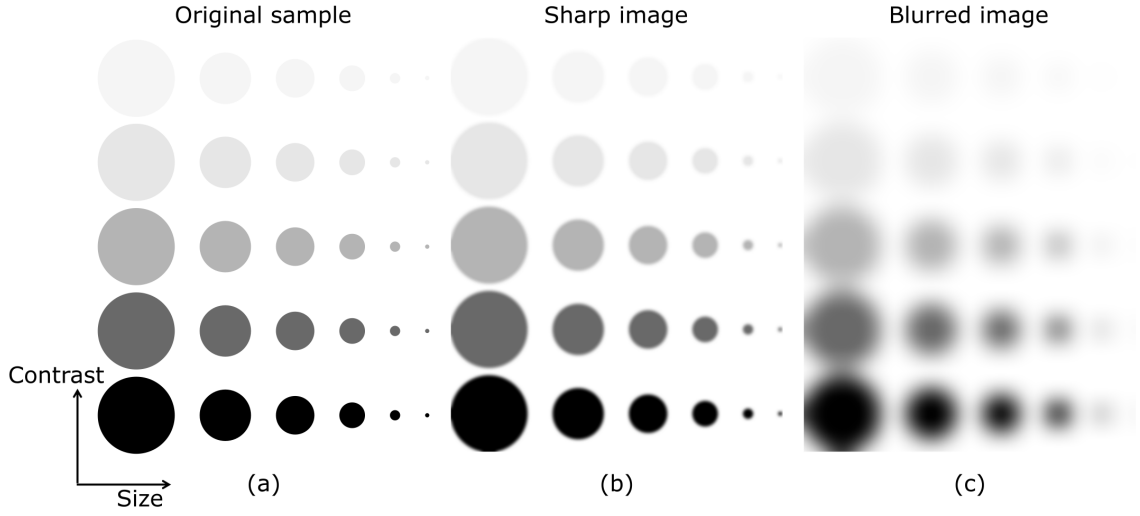


Figure 3.2: The horizontal axis of the images are the sizes of the circular objects. The vertical axis are the relative contrast of the objects to the background. (a) is the simulated original sample of which all the circular structures are visible. (b) is a simulated sharp image of the sample (a). Most structures are still visible but the edges of the circular structures are blurred. (c) is a simulated blurred image of the sample (a). Small sized structures with low contrasts are less visible than large structures with high contrasts.

Many mathematical expressions have been developed to define image sharpness [93]. Among all expressions, we chose a frequency threshold metric, f_{fq} , as our control system metric. The frequency threshold metric is defined as the sum of all frequency components above a certain frequency threshold α (see expression 3.1); it works similar to a high pass filter, of which only high frequency components contribute to the calculation of the image metric. This metric performs consistently [93]; besides, it also provides us with one degree of freedom to choose: α , the threshold value. This property can be very beneficial as we can adjust α based on the features of the specimens that we are interested in. For samples with coarse features, we will choose a smaller α while for samples with fine features, we can adjust α to a larger value. This adjustment can help us to customise the control system metric and create a

metric that is more accurate for each specific sample.

$$f_{fq} = \int_{\alpha}^{\infty} \int_0^{2\pi} |\mathcal{FT}[I(\theta, r)]| d\theta dr \quad (3.1)$$

where α is the chosen frequency threshold; $\mathcal{FT}[I(\theta, r)]$ is the Fourier transformed image intensity distribution in the frequency domain which is equivalent to the image spectrum; θ and r are polar coordinates in the spatial frequency domain.

3.2.1.1 Some modifications to the frequency threshold metric

If the noise is assumed to be uniformly distributed across all frequency components, as the threshold α increases, the signal to noise ratio (SNR) decreases; the metric becomes more susceptible to noise [93]. The practical values of α were also limited by the spatial frequency cut off of the imaging system. It was also noticed that this metric could also be influenced by other image properties such as contrast. To apply this metric and make it more suitable for our project, some modifications were required.

Noise threshold. To reduce the effect of noise, we defined a modified metric, f_n , with one more threshold β as upper frequency limit; any frequency components above β would be considered as noise and would not be summed in the metric calculation. An upper limit on β could be decided by the spatial frequency cut-off of the imaging system. The frequencies of the features were included in the range smaller than β and larger than α (see Figure 3.3).

$$f_n = \int_{\alpha}^{\beta} \int_0^{2\pi} |\mathcal{FT}[I(\theta, r)]| d\theta dr \quad (3.2)$$

Ratio metric. Contrast could also affect this metric calculation; an image with a larger contrast but not necessarily a better sharpness could also lead to a higher metric value. (see Figure 3.4) To reduce the influence of contrast, instead of summing

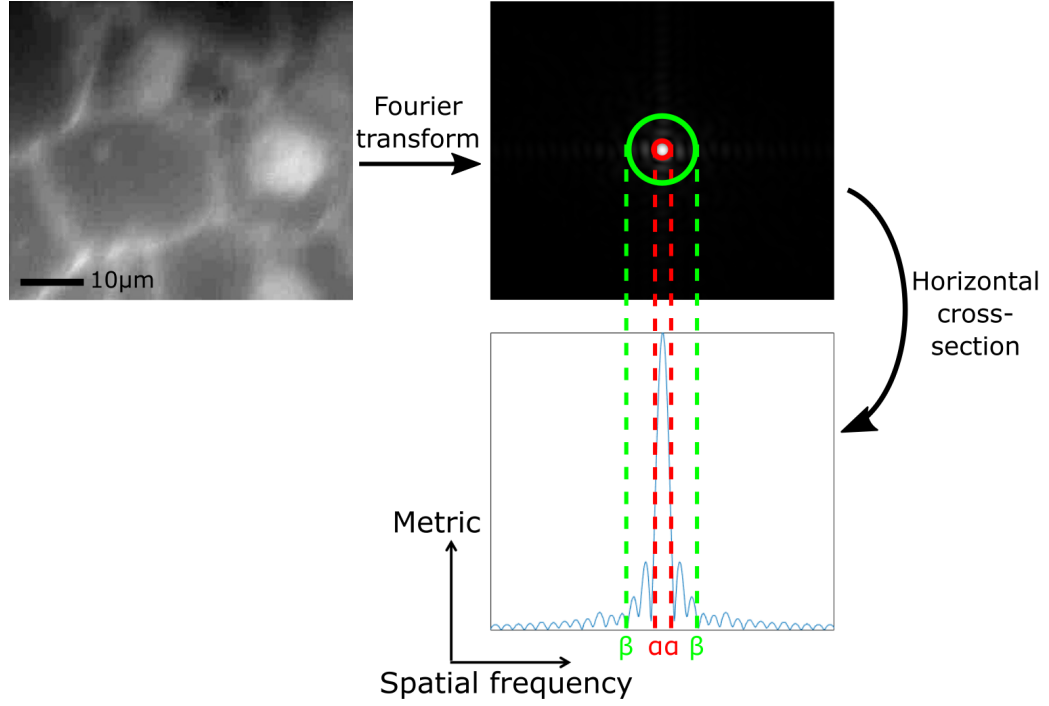


Figure 3.3: The output image (an image of a lilium anther sample) was Fourier transformed into the frequency domain. The area contained within the red ring (frequency components below α , $r < \alpha$) are the selected low frequency components. The area contained within the green ring but outside the red ring (frequency components between β to α , $\alpha < r < \beta$) were the selected high frequency components. The area outside of the green ring (frequency components larger than β , $\beta < r$) were the spatial frequency cut off of the imaging system; the frequency components contained mainly noise. Note that the both Fourier figures were calculated when zero paddings were added to the image which effectively increased the number frequency bins that were more closely spaced in frequency; this was done purely for display purposes.

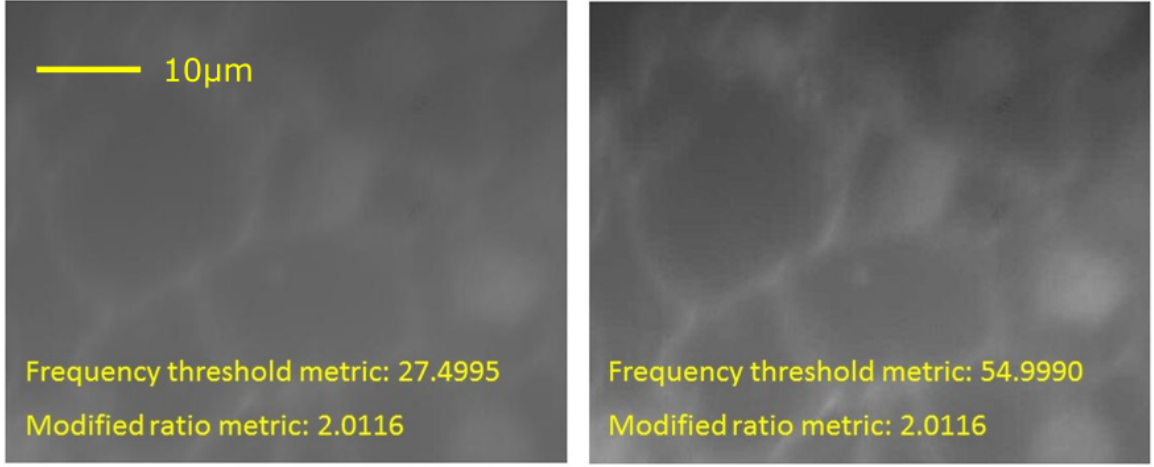


Figure 3.4: Two images of a liliun anther sample had the same sharpness but different contrasts. Contrast affects the frequency threshold metric (metric_{fre}) defined in the expression 3.1 but not the ratio metric (metric_{ratio}) defined in the expression 3.3.

high frequency components between α and β , we calculated a ratio metric, f_r defined as the sum of high frequency components between α and β over the sum of low frequency components smaller than α .

$$f_r = \frac{\int_{\alpha}^{\beta} \int_0^{2\pi} |\mathcal{FT}[\mathbf{I}(\theta, r)]| d\theta dr}{\int_0^{\alpha} \int_0^{2\pi} |\mathcal{FT}[\mathbf{I}(\theta, r)]| d\theta dr} \quad (3.3)$$

3.2.1.2 Variations of the image quality metric against the spherical aberration and defocus

We have defined an image-based quality metric to assess the spherical aberration and defocus correction according to the expression 3.3. A series of images was captured when the optical system stepped through a range of spherical aberration corrections and a range of axial movable stage depth positions. The images were analysed and the corresponding quality metric values were calculated. The system quality metric values varying against both the spherical aberration correction and the focus position are displayed in figure 3.5.

These results show that this system quality metric varies with both the spherical and defocus correction; a high metric value indicates that a good spherical aberration

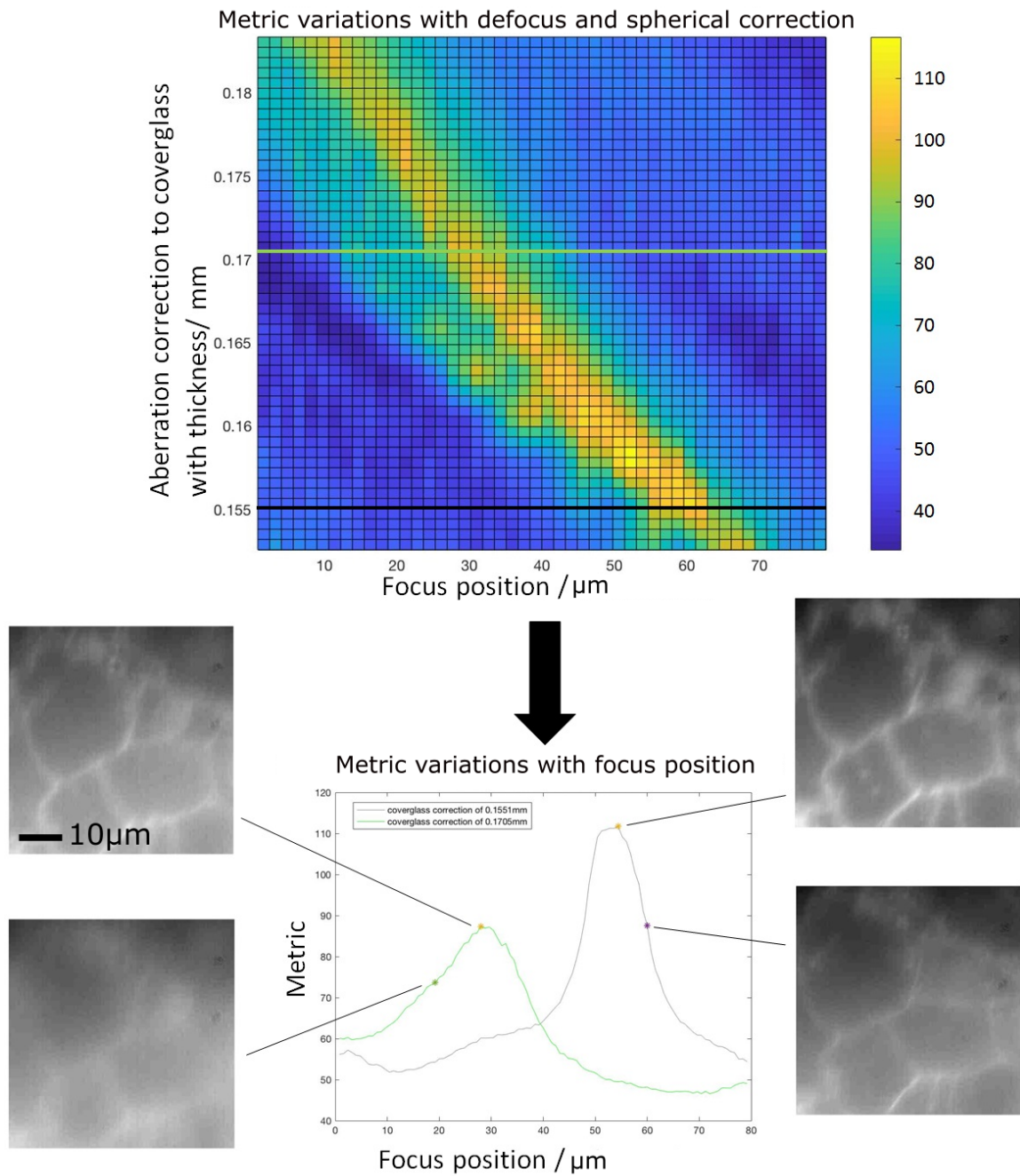


Figure 3.5: The metric as a function of both the spherical aberration correction and the focus position. Two aberration corrections to a coverglass with thickness of 0.1551 mm (the black coloured plot) and 0.1731 mm (the green coloured plot) were selected and the metric plots across all focal planes for both cases were compared. Four images of a lilium anther sample captured with different metric values were also included for demonstration.

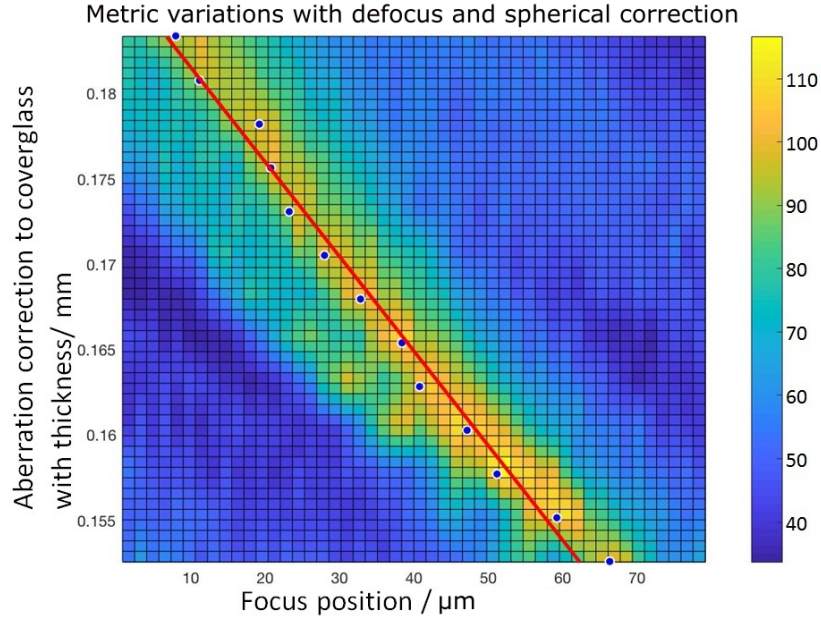


Figure 3.6: The linear approximation to the focal shift effect. For each objective lens correction collar setting, the maximum of the metric plot (when the system is in focus) was marked with a blue dot. A linear model was used to fit these marked points to determine the relationship between the correction collar settings and the corresponding best focusing positions.

correction and a focused optical system. When the spherical aberration is corrected perfectly, the plot of the metric against focus position is also observed to be have the highest peak, be the sharpest and most symmetrical roughly Gaussian curve shape. While when the spherical aberration is not fully corrected, the plot of the metric against the focus position is seen to be lower, wider and in the shape of a skewed normal distribution curve. It can be explained as the spherical aberration increases the depth of focus which causes skewness and a widening of the metric plot.

It can also be observed that the objective lens correction collar shifts the focal plane linearly as on the metric map (Figure 3.6), the high metric values (in the yellow colour and marked by blue dots) are distributed roughly on a diagonal line (fitted by a red line). This approximately linear relationship can be used to simplify the control system to make it more efficient.

3.2.2 Specification of the control system using the image quality metric

With the image quality metric, a full control strategy can now be specified. From the Figure 3.6, it is learnt that the objective correction collar settings have a linear relationship with the axial movable stage settings; it is therefore logical to search along the diagonal line (the red line) for the maximal metric value.

The automated control system thus consists of three steps: to determine the initial axial position of the focal spot, to determine the gradient of the linear model and to search for the maximal metric reading along the linear model.

3.2.2.1 To locate the initial axial position of the movable sample mounting stage

To start the control scheme, we first needed to identify the axial location of the focal spot for an initial correction collar setting. Mathematically, it is equivalent to determine the constant term of the linear model. This could be achieved through moving the sample mounting stage while seeking for the plane where the sharpest image and highest metric value was obtained. As shown in the Figure 3.5, for a collar setting, the metric varied against the focal position where the peak of one metric plot corresponded to the best focus for that particular collar setting. It was also observed that the curve could vary in shape due to spherical aberrations and it might be skewed (see the green metric plot in the Figure 3.5).

A polynomial up to order three could be used to fit the metric variations and help to determine the metric peak. Polynomials up to order three allowed the model to capture the skewness of the metric plot; at the same time avoided overfitting and ensured the efficiency: a minimum of four $(n + 1)$ data points were required for fitting an order three (n) polynomial model. This could promise a high calculation speed when processing as part of the control system. Figure 3.7 showed that the third order

polynomials fitting provided promising approximations of the metric peak. Note that a polynomial of up to order two may also be used to fit the metric but may result a less accurate prediction; for a quadratic fitting, a minimum of three data points were required for the optimisation.

3.2.2.2 To determine the gradient of the spherical-defocus linear model

The correction collar settings and their best focal planes have a linear relationship (as shown in the Figure 3.6). The gradient varied with different samples over different FOV but in a range of -2.3 to -1.6 which was around -2 . We could hence use -2 as an initial gradient and update it when more measurements became available.

One possible explanation for the varying gradient was that when the system had a large spherical aberration, the PSF was elongated in the axial direction (a large depth of focus) which could result a less defined focal plane. There was an interesting observation when the experiments were repeated ten times by imaging ten different regions of the same specimen. Ten series of images were analysed and ten sets of results were displayed in Figure 3.8. It was noticed that the approximated focal plane was the most distinct when the spherical aberration was best corrected and thus the focal spot was the sharpest with the smallest depth of focus. When the spherical aberration was not fully corrected, the focal plane was difficult to be identified since an aberrated focal spot was elongated with a large depth of focus. This observation explained why there was a variation in the gradient measurement of the spherical-defocus linear model; though all the linear model passed through the same point when spherical aberration was fully corrected, the gradient varied due to the inaccurate identifications of the focal plane when spherical aberration was far from corrected.

After the first two steps described in the above section 3.2.2.1 and in this section, we obtained the spherical-defocus linear model. Making use of this model, we could thus hold the system in focus while vary its spherical compensation setting

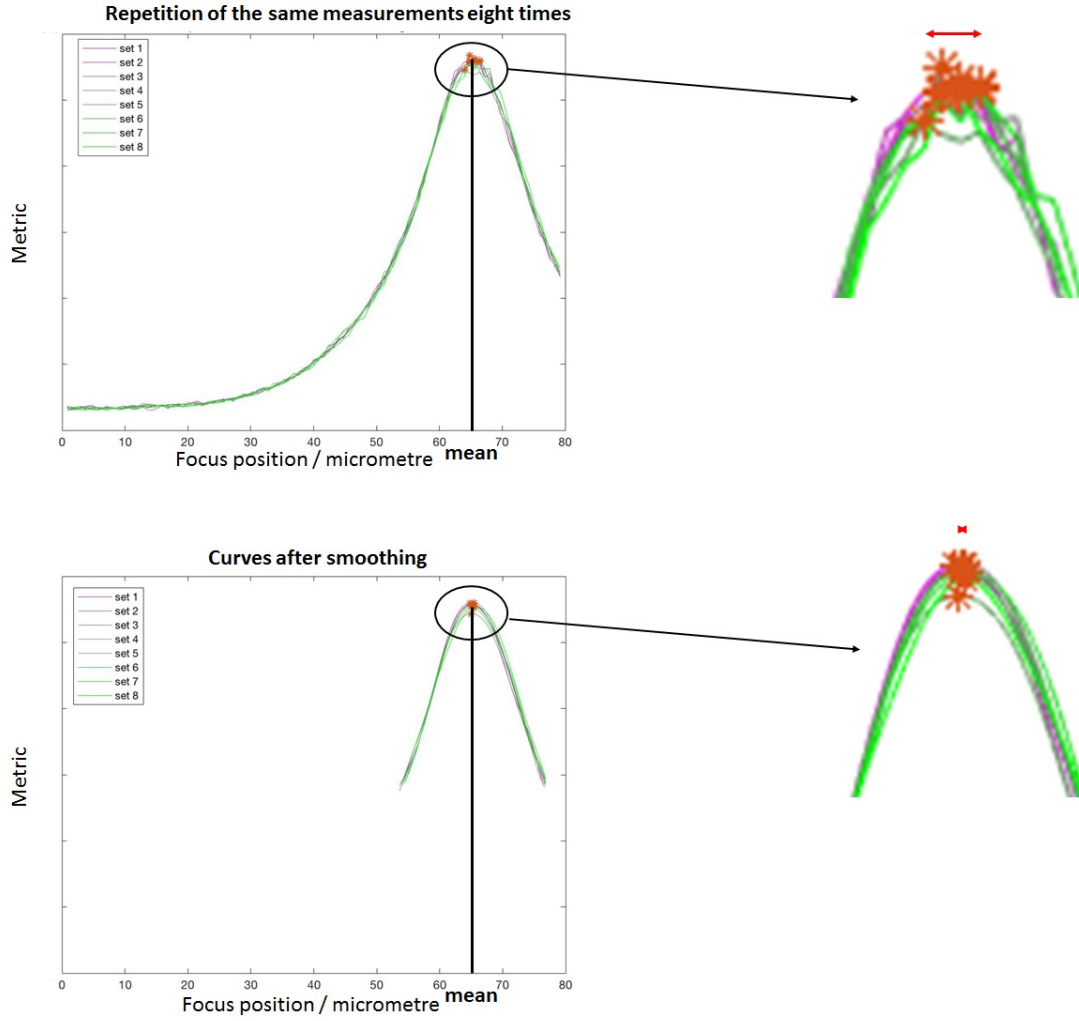


Figure 3.7: Eight experiments on the same sample over the same FOV with exactly the same correction collar settings and the same range of the focal positions were repeated. Eight identical outputs were expected from the eight repeated experiments. The maximum metric readings of the eight experiments varied slightly due to noise. However the third order polynomial curve fitting provided very similar focal plane approximations.

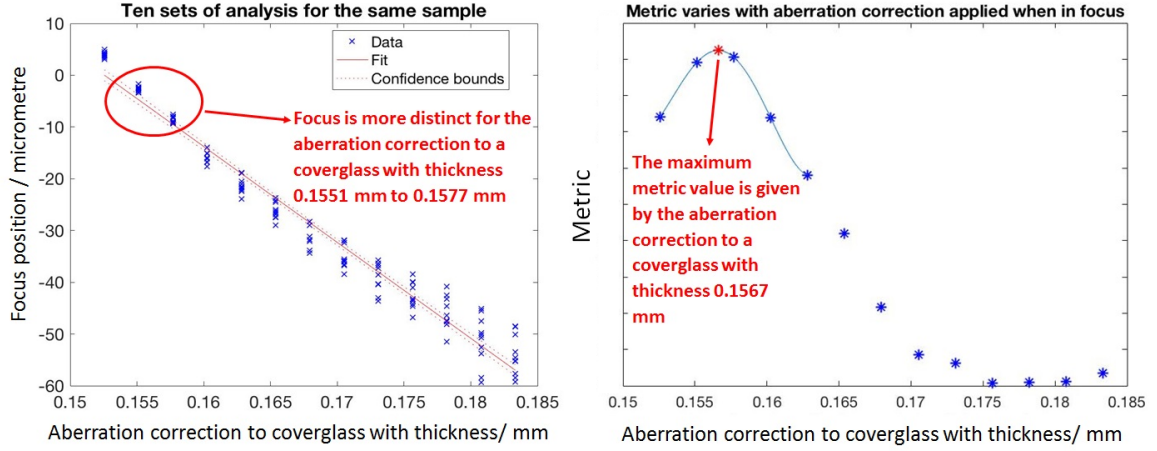


Figure 3.8: Ten sets of results from analysing ten different regions of the same lily anther sample were displayed to show that the relationship between the spherical aberration correction applied and the focal plane deduced was approximately linear. The left plot showed that the approximate focus was more distinct when the corrections to coverglass thickness were 0.1551mm to 0.1577mm ; the right plot showed that the system metric was the highest when the correction to coverglass thickness was 0.1567mm . The approximate focal plane was more distinct when the spherical aberration was corrected better.

independently to search for the metric optimum.

3.2.2.3 To determine the maximum metric reading

The best spherical aberration and defocus correction was determined by comparing the metric values on the spherical defocus linear model (see Figure 3.9). A polynomial model up to order three was once again being used for its simplicity and effectiveness (and a second order polynomial could be used but may result a less accurate result). A local search for a maximum was then applied after the metric plot was fit into a smooth curve based on our assumption that there was only one best correction collar setting; i.e. a local maximum also corresponded to the global maximum. The assumption is valid when the sample for imaging is very thin (in the order of $10\text{s } \mu\text{m}$).

An interesting observation was that for samples with fine features, the metric was more sensitive; a clear maximum of the plot could be easily determined. For samples only containing coarse features, the metric was less sensitive and the metric plot was

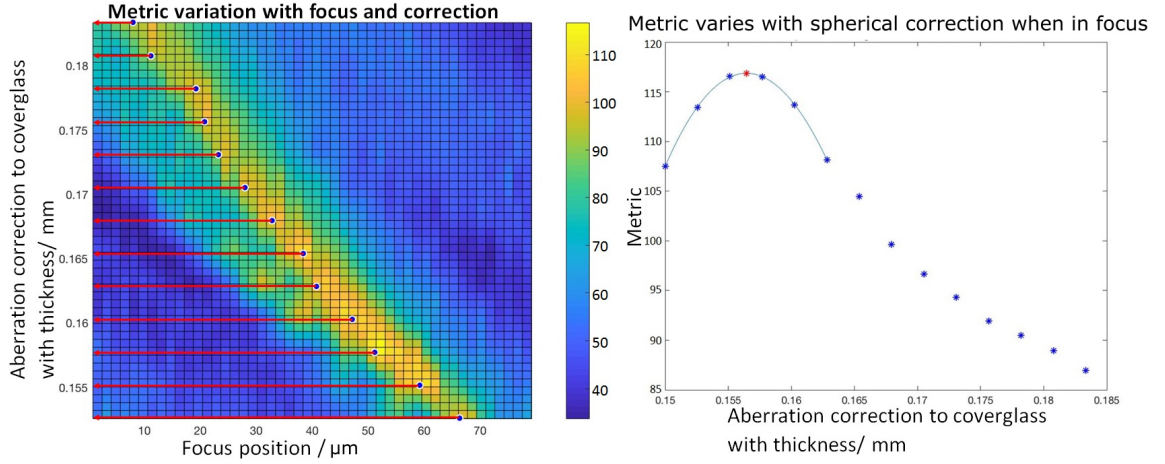


Figure 3.9: The metric values when the optical system was in focus were projected onto spherical aberration correction axis. The metric plot against correction collar settings was fitted into an order three polynomial to obtain a smooth curve.

flatter (see Figure 3.10). The sensitivity of the metric can be increased by changing the metric power used in the metric calculation; instead of using the absolute value of the frequency components to calculate the metric ratio, squared or cubed values of the frequency components can be used.

3.3 Incorporating the control system into the new Zeiss objective lens

By equipping the new Zeiss objective lens and using the new automated control system we designed, several samples were examined. Satisfactory results were obtained; after the correction procedure, the output image became sharper (see Figure 3.11). Consistent results were produced when we repeated experiments on the same samples; the best correction collar settings and the corresponding best focus positions worked out by the control system for the same sample stayed the same for the repetitive experiments.

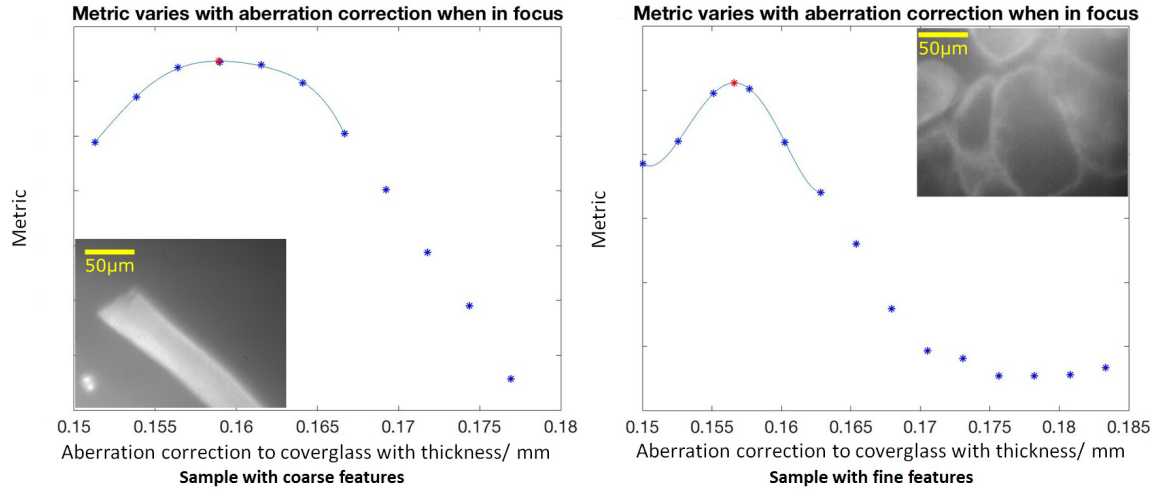


Figure 3.10: The metric variation with spherical aberration correction for FOV containing coarse features (left-hand side) and fine features (right-hand side). Both images are captured from the same lily anther sample but at different area. The insets are the images of different areas of the same sample used for the analysis. It could be clearly seen that the metric variation was less sensitive when the FOV contained more coarse features than fine.

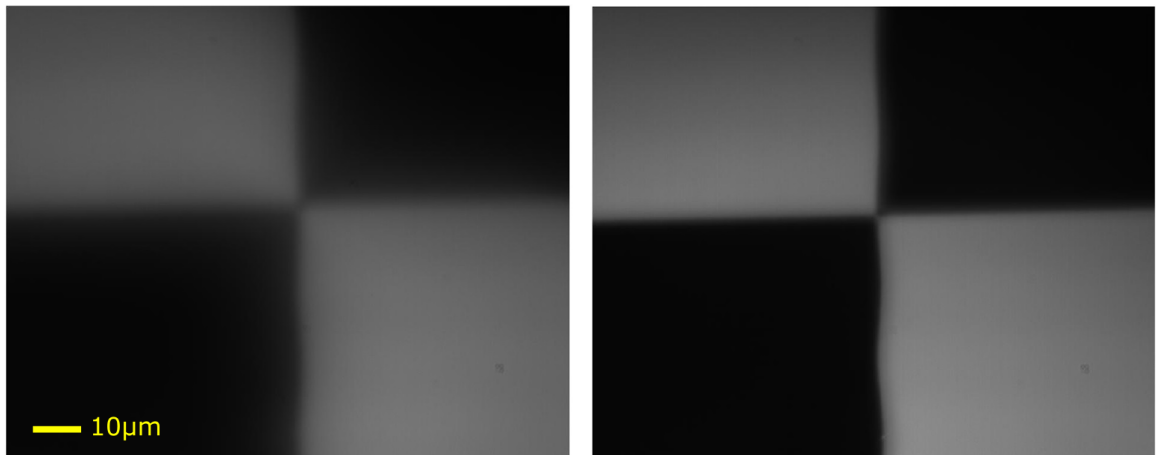


Figure 3.11: A glass plate with metal grids printed on was imaged to demonstrate the control scheme. Left and right are images of the sample before and after using the control scheme.

3.4 Conclusion and future outlook

In this chapter, we presented a control system for both the spherical aberration correction and the optical system focus to allow an easy usage of the Zeiss motorised objective lens. The system requires a minimum of $2n + 1$ measurements where n is the order of polynomial used for metric approximations; the results presented in this section were when a third order polynomial was used. The results produced by the control system were shown to be promising.

The automated aberration correction and focusing system together with the Zeiss objective lens have only been used on the widefield microscope with a small range of samples with glycerine as the immersion fluid for the purpose of this project. In the future, the system can be tested on more samples with different immersion fluid in different microscopy systems. The widefield microscope was chosen for simplicity of demonstration, but it is expected that the principles of operation will be transferable to other microscope modalities.

In this project, all of the samples used were relatively thin. Thick samples may possibly cause the metric to be less sensitive and create extra challenges for the control system. However, if the thick sample contain thin structures, the control system presented should still be able to provide promising results.

Chapter 4

A universal framework for sensorless adaptive optics

AO is often used in microscopes to compensate phase aberrations introduced by optical inhomogeneities in specimens or by optical systems and to restore near-diffraction-limited operation [3, 94–97]. This is achieved by estimating aberrations through use of a wavefront sensor or through indirect optimisation approaches. Instead of measuring phase aberrations directly, wavefront-sensorless aberration correction methods (known concisely as “sensorless AO”) infer aberration correction from a set of image measurements that are obtained when intensity and/or phase perturbations are introduced into the system. Sensorless AO methods are commonly used because they require no extra hardware for direct wavefront sensing and hence allow simpler optical designs and avoid non-common path sensing errors. There are also cases where the signal level is too low for sensor-based methods to be feasible, leaving sensorless methods the only viable option.

Many sensorless AO approaches have proved to be effective on various types of microscopes including widefield [98–107], confocal [107–120], multiphoton [103, 109, 121–146], multiharmonic [123, 147], light sheet [148], fibre imaging [115, 127, 133, 149],

CARS [150], optical coherence microscopy (OCM) [151], SIM [152, 153], STED [70, 71, 154, 155], single molecule switching (SMS) [156–159], and 4Pi [159, 160] microscopes. While some of the implementations were designed specifically for certain types of microscopy (such as the iterative multi-photon adaptive compensation technique (IMPACT) method [128, 137, 140, 151] were designed for multiphoton microscopes and later applied to OCM), some other methods were demonstrated on multiple different systems (such as the modal methods [70, 71, 99, 108, 109, 114, 124, 146–148, 152–154, 157, 159, 160] were applied on widefield, confocal, multiphoton, multiharmonic, light sheet, SIM, STED and SMS microscopes). Since different approaches were all developed separately for different applications, meaning a fair comparison among the methods was difficult to conduct, and no study has systematically compared, discussed and summarized the trade-offs of these methods. Furthermore, the apparent conceptual separations between these correction schemes meant that there were no practical guidelines to advise AO users about choosing the more effective method for a certain application.

In this chapter, we introduced a universal framework in which all of these sensorless AO methods could be interpreted. The framework requires the definitions of three components: an aberration representation, an optimisation metric, and an estimation algorithm. Different sensorless AO methods have different choices of these three components which make them suitable for different applications.

4.1 The mathematical framework for sensorless adaptive optics

We present in this section a generalized framework that is broad enough to encompass all published image based sensorless AO systems. The only measurement available to a sensorless AO scheme is the image, which consists of an ordered collection of in-

tensity measurements from the imaging plane. Using a widefield single-photon (1-P) fluorescence microscopy as an example, the image output of the microscope is a fluorescence intensity distribution which has a mathematical expression of a convolution between the PSF of the system and the fluorophore density of the sample. Since the effects of the phase aberration are confined to the PSF [19, 20] and normally sample structures are at least to some degree unknown, more than one measurement is commonly required to deduce the PSF unambiguously from image data. The typical SNR in microscopes makes the retrieval of phase even more challenging. Therefore, all existing variations of the sensorless AO methods require in practice multiple exposures. Each of these exposures is obtained with a different perturbation to the system, which is typically in the form of a pre-determined bias phase aberration introduced by AO devices, such as the DMs and SLMs. This set of measurements is then used to calculate a metric, which quantifies the quality of each recorded image (see Figure 4.1). Since only an aberration free pupil phase¹ can result a diffraction limited performance, by optimising an image quality metric we can infer the phase aberration correction to the pupil plane. Mathematically, this process can be expressed as an optimisation problem in the form:

$$\operatorname{argmax}_{\boldsymbol{\varphi}} \{f[O, \boldsymbol{\Psi}(\boldsymbol{\varphi})]\} \quad (4.1)$$

This expression can be interpreted as finding a set of optimised values of the set of vectors $\boldsymbol{\varphi}$, representing the aberrations, which maximises f . The metric f defines the quality of the output images; in many cases, f can be expressed explicitly as a function of the object structure, O , and the microscope PSF, which in turn is a function of

¹An aberration free pupil phase may be defined differently for different microscopy systems. It will be discussed in details later in this chapter.

the total aberration $\Psi(\boldsymbol{\varphi})$ [20]. Ψ is a function of $\boldsymbol{\varphi}$ which can be expressed as

$$\Psi(\boldsymbol{\varphi}) = \Phi(x, y) + \sum_i^N \varphi_i(x, y) = \Phi(x, y) + \mathbf{1} \cdot \boldsymbol{\varphi} \quad (4.2)$$

where $\Phi(x, y)$ is the combination of system and sample induced aberrations that need to be corrected; it is a function of the pupil plane spatial vectors x and y . The summation term represents the aberration correction imparted by a wavefront shaping device: $\varphi_i(x, y)$ is the i th mode of a set of aberration representation modes $\boldsymbol{\varphi}$; there are a total of N modes included in the set. The summation term can also be expressed as $\mathbf{1} \cdot \boldsymbol{\varphi}$ for simplicity where $\mathbf{1}$ is a $N \times 1$ vector with all elements equal to 1.

We now define a unit vector $\hat{\varphi}_i$ such that

$$\hat{\varphi}_i = \frac{\varphi_i}{\sqrt{\langle \varphi_i, \varphi_i \rangle}} \quad (4.3)$$

where $\langle \varphi_i, \varphi_i \rangle$ is the inner product² between two φ_i . The inner product can be used to define orthogonality.

Let us first define a matrix expression $\langle \hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\varphi}} \rangle$ which indicates a $N \times N$ symmetrical matrix; the i^{th} row and j^{th} column of the matrix indicates the element $\langle \hat{\varphi}_i, \hat{\varphi}_j \rangle$; $\langle \hat{\varphi}_i, \hat{\boldsymbol{\varphi}} \rangle$ expresses the i^{th} row and $\langle \hat{\boldsymbol{\varphi}}, \hat{\varphi}_j \rangle$ expresses the j^{th} column of the matrix $\langle \hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\varphi}} \rangle$. If $\hat{\boldsymbol{\varphi}}$ is a set of orthogonal (orthonormal) vectors, the matrix $\langle \hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\varphi}} \rangle$ is a diagonal (identity) matrix. If $\hat{\boldsymbol{\varphi}}$ is a set of non-orthogonal vectors, the matrix $\langle \hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\varphi}} \rangle$ has non-zero off-diagonal elements.

Using the unit vector expression 4.3, we can now write $\sum_i^N \varphi_i$ as

$$\sum_i^N \varphi_i = \sum_i^N \langle \varphi_i, \hat{\varphi}_i \rangle \hat{\varphi}_i = \sum_i^N a_i \hat{\varphi}_i = \mathbf{a} \cdot \hat{\boldsymbol{\varphi}}$$

²The inner product can be defined differently depending on the applications [161]. It is also discussed in more details later in the section.

where a_i is the scalar value for φ_i ; the vectors $\mathbf{a}$ and $\hat{\varphi}$ contain all a_i and $\hat{\varphi}_i$ components up to order N respectively. We may also write Φ as

$$\Phi = \sum_i^N a'_i \hat{\varphi}_i + \varepsilon = \mathbf{a}' \cdot \hat{\varphi} + \varepsilon$$

where $\mathbf{a}'$ is a set of scalar coefficients of modes $\hat{\varphi}$ contained within the system and sample induced aberrations Φ ; ε is the remaining non-correctible aberration that is outside and orthogonal to the set $\hat{\varphi}$.

$\Psi(\varphi)$ can now be written as

$$\Psi(\varphi) = \mathbf{a}' \cdot \hat{\varphi} + \varepsilon + \mathbf{a} \cdot \hat{\varphi} \quad (4.4)$$

To optimise equation (4.1), we can take the partial differentiation of the metric f with respect to each mode φ_i and make it equals to zero. For a general case when modes are not necessarily orthogonal to each other

$$\frac{\partial f}{\partial \varphi_i} = \frac{\partial f}{\partial \Psi} \sum_j^N \left(\frac{\partial \Psi}{\partial \hat{\varphi}_j} \frac{d \hat{\varphi}_j}{d \varphi_i} \right) = \frac{\partial f}{\partial \Psi} \sum_j^N \{ [(\mathbf{a}' + \mathbf{a}) \cdot \langle \hat{\varphi}_j, \hat{\varphi} \rangle] \frac{\langle \hat{\varphi}_j, \hat{\varphi}_i \rangle}{a_i} \} = 0 \quad (4.5)$$

To optimise the metric f with respect to one mode φ_i , the best aberration correction possible is when all the scalar coefficients $\mathbf{a} = -\mathbf{a}'$.

For a more specific case when the set of modes $\hat{\varphi}$ are orthogonal and thus the matrix $\langle \hat{\varphi}, \hat{\varphi} \rangle$ is a diagonal matrix,

$$\frac{\partial f}{\partial \varphi_i} = \frac{\partial f}{\partial \Psi} \frac{\partial \Psi}{\partial \hat{\varphi}} \frac{d \hat{\varphi}}{d \varphi_i} = \frac{\partial f}{\partial \Psi} (a'_i + a_i) \frac{1}{a_i} = 0 \quad (4.6)$$

In this case, to optimise the metric f with respect to one mode φ_i , the best aberration correction possible is when $a_i = -a'_i$.

It is thus interesting to see that for an orthogonal set of modes, the optimisation process for one mode is decoupled from other modes; the optimisation process for an

orthogonal set can be conducted mode by mode. If a set of non-orthogonal modes are used, global optimisation mechanisms, of which all modes are optimised together, may be necessary.

In both cases, to optimise the metric f , the best possible aberration correction is when $\mathbf{a} = -\mathbf{a}'$ and thus $\Psi(\boldsymbol{\varphi}) = \varepsilon$.

The highest system quality metric can then be written as

$$f(O, \varepsilon) \tag{4.7}$$

f is ideally chosen to reflect the image quality and in general monotonically decreases as $\langle \Psi, \Psi \rangle$ increases; f is maximised only when the aberration is minimised. Different choices of $\hat{\boldsymbol{\varphi}}$ affect the size of the remaining non-correctable aberrations ε and thus the best possible imaging quality. A minimal phase aberration may be defined differently depending on applications. For example, the piston aberration (a flat constant) affects the interference and thus jeopardise the PSF quality in a 4Pi system [69]; however the piston normally does not affect the imaging quality in other systems. The tip / tilt (a constant slope) aberration affects the quality of a STED microscope [70, 71], but does not affect the widefield and non-linear systems. The defocus aberration also decreases the quality when imaging a thin sample but does not influence much when imaging a thick sample.

In fact, the relationship between the quality metric f and different aberrations are embedded in the equation 4.6 (with orthogonal basis modes) involving the partial differentiation term $\partial f / \partial \Psi$. As elaborated above, f may take different mathematical forms which lead to different analysis. For the purpose of illustration, we chose one of the simplest form of f : the Strehl ratio to demonstrate how the choice of the quality metric is related to the aberrations and how the optimal modes can be chosen accordingly.

For a metric defining the Strehl ratio [162,163], it can be mathematically expressed as

$$f = \left| \iint_P e^{j\Psi} dx dy \right|^2 = \left(\iint_P e^{j\Psi} dx dy \right) \left(\iint_P e^{-j\Psi} dx dy \right)$$

where P is the total spatial area of the illuminated back pupil which is set to equal to 1, i.e. $\iint_P dx dy = 1$, so that the maximum metric value f equals to 1 which is consistent with the definition of the Strehl ratio. f in this case can be interpreted as the intensity of the on axis focal spot which is a valid definition of the Strehl ratio assuming a small aberration and thus the peak focal intensity is on axis.

Let us now define a unit vector $\hat{\Psi}$ such that

$$\hat{\Psi} = \frac{\Psi}{\sqrt{\langle \Psi, \Psi \rangle}}$$

where $\hat{\Psi}$ can be interpreted as the general shape of the wavefront. Ψ can be expressed as

$$\Psi = \sqrt{\langle \Psi, \Psi \rangle} \hat{\Psi} = a_\Psi \hat{\Psi}$$

where $a_\Psi = \langle \Psi, \hat{\Psi} \rangle$ is a scalar variable that can be interpreted as the amplitude of the aberration.

Differentiate the Strehl ratio metric f

$$\frac{df}{d\Psi} = \frac{\partial f}{\partial \hat{\Psi}} \frac{d\hat{\Psi}}{d\Psi} + \frac{\partial f}{\partial a_\Psi} \frac{da_\Psi}{d\Psi} \quad (4.8)$$

Differentiating f with respect to $\hat{\Psi}$

$$\begin{aligned} \frac{\partial f}{\partial \hat{\Psi}} &= \left(\iint_P e^{j\Psi} dx dy \right) \left(\iint_P -j a_\Psi e^{-j\Psi} dx dy \right) + \left(\iint_P j a_\Psi e^{j\Psi} dx dy \right) \left(\iint_P e^{-j\Psi} dx dy \right) \\ &= 0 \end{aligned}$$

It is interesting to see that $\frac{\partial f}{\partial \hat{\Psi}}$ always equals to 0 which means that for this choice of

f , it is independent on $\hat{\Psi}$. In fact, provided f is a one-dimensional scalar variable, it should only depend on the amplitude of the aberration if the $\hat{\Psi}$ is normalised using a suitable inner product definition.

The differentiation of f with respect to a_Ψ is given by

$$\frac{df}{da_\Psi} = \left(\iint_P e^{j\Psi} dx dy \right) \left(\iint_P -j\hat{\Psi} e^{-j\Psi} dx dy \right) + \left(\iint_P j\hat{\Psi} e^{j\Psi} dx dy \right) \left(\iint_P e^{-j\Psi} dx dy \right)$$

Take the Taylor expansion for $e^{j\Psi}$ and assume a small Ψ so that any term higher than second order is ignored.

$$\frac{df}{da_\Psi} \approx -2a_\Psi \iint_P \left(\hat{\Psi} - \iint_P \hat{\Psi} dx dy \right)^2 dx dy$$

$\iint_P \hat{\Psi} dx dy$ is the spatial mean of $\hat{\Psi}$. From the expression, if $\hat{\Psi}$ contains only piston (totally flat), $\hat{\Psi} - \iint_P \hat{\Psi} dx dy = 0$ and f is at its maximum; alternatively, $a_\Psi = 0$ also optimises f .

Finally, if the inner product is defined as $\langle \hat{\Psi}, \hat{\Psi} \rangle = \iint_P (\hat{\Psi} - \iint_P \hat{\Psi} dx dy)^2 dx dy = 1$ which can be interpreted as the variance of the unit vector $\hat{\Psi}$, we can write the differentiation of f with respect to a_Ψ as

$$\frac{df}{da_\Psi} = -2a_\Psi \tag{4.9}$$

$$f = f_0 - a_\Psi^2 = 1 - a_\Psi^2 \tag{4.10}$$

where f_0 is a constant indicating the maximum value of f when no aberration exists; $f_0 = 1$ in this case since the maximum Strehl ratio equals to 1. We have thus shown that for a quality metric defined as the Strehl ratio, the metric is maximised when the aberration is flat (contains only piston); the inner product is ideally defined as the variance so that the metric can vary isotropically regardless of the aberration mode. The metric can be approximated as a parabola to the aberration amplitude, a_Ψ .

In a lot of previous works, it has also been shown that the single valued metric f is desirable to be approximated in a quadratic form [99, 108, 109, 111, 152, 161] when the size of the aberration a_Ψ is small. For different definitions of f (such as the intensity of the PSF or the sharpness of the image), the inner product is defined accordingly to fit into the quadratic model [161]. By doing so, the quality metric can be equally sensitive to different aberration modes. The definition of the inner product affects the aberration representation modes decomposition as the orthogonality of the modes are examined differently [99, 111, 152, 161, 164]. The quadratic model with its convex shape helps to develop more efficient model based optimisation algorithms compared to stochastic search based algorithms.

There are also cases when f is defined as a multi-dimensional metric. For example, the PSF shape related multi-dimensional metrics are used in Gerchberg-Saxton [101] and Machine Learning [107, 120, 165, 166] based phase retrieval. In these cases, the multi-dimensional f varies with $\hat{\Psi}$. One of the advantages is that comparing to those single-valued metrics (such as the average intensity or sharpness), more information is extracted from one frame and thus is more efficient. More details are included in the section 4.4.

The above formulation provides the basis for the universal framework. All the aberration correction approaches contain three major components – a set of aberration representation modes, $\hat{\varphi}$, an optimisation metric, f , and an algorithm for the optimisation process, $\text{argmax}_\varphi[f(\varphi)]$. The fundamental differences between all sensorless AO methods consist of variations of one or more of these three components. In the following sections, we discussed how these three components of the framework varied in different sensorless AO methods for different applications.

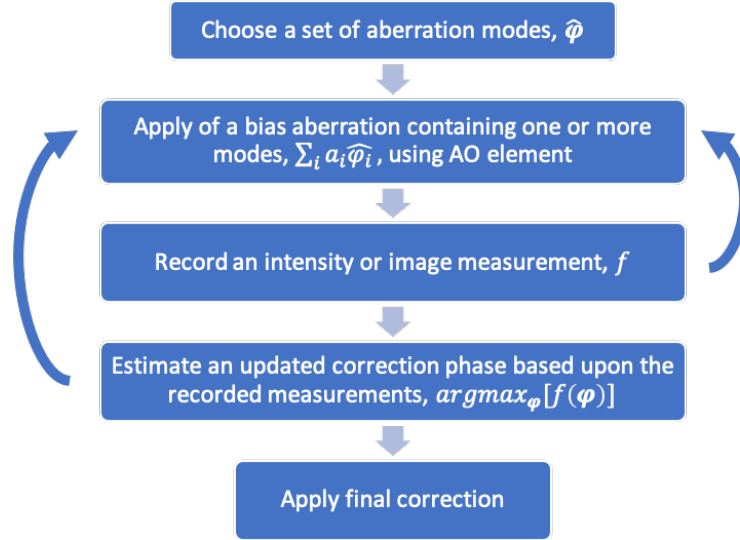


Figure 4.1: Flowchart of the generalized sensorless AO aberration correction process.

4.2 Aberration representation modes

As defined in the section 4.1, a set of aberration representation modes are said to be orthogonal if the inner product matrix $\langle \hat{\varphi}, \hat{\varphi} \rangle$ is a diagonal matrix, i.e. $\langle \hat{\varphi}_i, \hat{\varphi}_j \rangle \neq 0$ for $i \neq j$. Equation 4.5 and 4.6 showed that $\hat{\varphi}$ are ideally selected to be orthogonal; there are still cases when the modes used for phase modulations and corrections are not orthogonal. For example, a continuous membrane DM with a small number of actuators may have non localised influence functions which overlap with the influence functions of other actuators [167–169]. If non-orthogonal modes are used for aberration corrections, the modes cannot be optimised individually; global stochastic optimisation algorithms are needed [110, 112, 122, 150] which normally require more measurements for convergence [121] and the randomness results in a non-reliable convergence speed [110, 122]. Alternatively, the DM can be calibrated to work out the conversion control matrix from the actuator voltages to a set of orthogonal modes which are then used for the DM operations [108, 111, 161, 170, 171].

For most of the cases, orthogonal aberration representation modes are used for aberration modulations and corrections. These include continuous modes such as

Zernike polynomials [70, 71, 84, 100, 103, 105–109, 114–117, 120, 124, 129, 139, 141, 146–148, 152–157, 172], deformable mirror modes [111, 123, 127, 159–161] and Lukosz modes [99, 173, 174], of which the mirror modes and Lukosz modes are variants derived from the Zernike modes to better suited for specific applications; piecewise modal modes such as the Hadamard-Walsh modes [102, 119, 132, 138, 139, 142] and the polar Walsh modes [149]; piecewise linear zonal representations [100, 104, 125, 126, 130, 131, 133, 135, 136]; and piecewise constant zonal representations [128, 137, 140, 151].

There are already works dedicated to comparing various modes about their effectiveness. Wang *et al.* [161] compared two sets of deformable mirror modes when displayed on the DMs with different actuator arrangements; the differences of the two modes are not significant when the DM has a large number of actuators. Soloviev [164] compared the performances of a few variations of Zernike polynomials, including the Lukosz modes, Singular value decomposition (SVD) modes and Neumann modes; the results showed that the performances are similar with slight differences. There are also works combining different sets of modes for aberration corrections. Warber *et al.* [100] demonstrated a method of using Zernike polynomials to correct system aberrations and using piecewise linear zonal modes to correct sample induced aberrations. Galwaduge *et al.* [139] demonstrated of using both Zernike and Hadamard modes to correct aberrations and scattering. However, a systematic comparison among the continuous modes and piecewise modes was not discussed until the paper [175]. In the next chapter 5, the simulation and experimental results of comparing various continuous and piecewise modes in the paper [175] were reproduced and discussed.

4.3 Optimisation metrics

The choice of the optimisation metric for convergence, f , depends largely on the optical system. While the average image / fluorescence intensity is a common choice

for confocal and non-linear fluorescence microscopes [102, 103, 108–112, 114, 116–119, 121, 122, 124, 126, 128, 129, 132, 134, 136–142, 146, 147, 149–151, 153, 156], it is not a suitable metric for imaging systems such as a single photon widefield microscope since the image intensity does not vary with the size of aberrations [20]; it is also not suitable for super-resolution microscopes like SMS, of which the number of blinking events varies with time [157], and STED, of which a better quality depletion beam leads to a smaller excitation PSF and thus a lower average fluorescence intensity [70, 155]. In these cases, image sharpness, in the forms of spatial frequency components [70, 99, 115, 123, 148, 152, 154, 157, 158] and the size of the PSF [112, 127, 174], are used to measure image quality; the peak intensity of the FOV / the Strehl ratio is also used for the image quality assessment [159, 160].

Ideally, the optimisation metric is desired to have only one local optimum which corresponds to the global maximum; this is when the size of the pupil aberration, $\langle \Psi, \hat{\Psi} \rangle$, is minimised. Ideally, f monotonically decreases as the size of the aberration $\langle \Psi, \hat{\Psi} \rangle$ increases. In practice, this is true only when $\langle \Psi, \hat{\Psi} \rangle$ is within a limited range close to zero [108]. For large aberrations, f typically becomes very small and flat, which makes it susceptible to noise [19]; f may also have small local maxima (sidelobes) which do not correspond to the global maximum. These can result a failed optimisation process makes the system unable to converge to its global optimum. Furthermore, these one-dimensional metric only shows the size of the aberration, but unable to indicate the shapes of the aberration; this makes the single valued metric less efficient.

One way to overcome the problems mentioned above is to use multi-dimensional optimisation metrics. Obvious advantages of a multi-dimensional metric over a single valued metric are that it is unlikely to have local optimums - a local optimum can only happen if all dimensions of the metric have local optimums at the same point. Multi-dimensional metric also in general extract more information from one frame.

Antonello *et al.* [155] demonstrated a multi-scale wavelet decomposition based metric to extract localised features of different sizes; it shows that the optima of the metrics with the selected wavelet sizes all agree at the same point when aberrations are corrected the best. Recently, several groups dedicated to using machine learning into phase retrieval and aberration corrections in the microscopy systems. The multi-dimensional metric is thus encoded, in effect, into the multiple layers of the neural network. In this case, the multi-dimensional metric is not specified but learnt by the machine through numerous sets of training data [106, 107, 120, 165, 166]. There are also studies using the intensity readings of all pixels on a stack of sample images to optimise an untrained neural network and work out the pupil aberration [172].

In the later chapter 6, an untrained neural network encapsulating a physical model of microscope image formation is shown to work out the pupil aberration from a minimum of two images of the same sample. All the pixel readings on the images are used for convergence and thus it utilises a multi-dimensional metric.

4.4 Optimisation algorithms

Model free global stochastic optimisation algorithms such as genetic algorithms [110, 112, 118, 121, 134, 156], hill climbing [103, 110, 122] and random search [110, 150] are normally used when aberration representation modes are not orthogonal and thus a global optimisation is required. Model free local optimisation algorithms such as systematic search algorithm [104, 119, 125, 126, 128, 130, 131, 133, 135–138, 144, 151] are normally used when aberration representation modes are orthogonal. These model free algorithms are easy to implement and shown to be effective to converge, but they are often less efficient and require more image acquisitions. The stochastic algorithms also have less reliable convergence speeds [110, 122].

Booth *et al.* [108, 109, 111] showed that the intensity based metric varying with

the variance value of the aberration can be approximated as a quadratic polynomial within a limited range around zero; it was then shown that the spatial frequency based metric varying with the root mean square (RMS) gradient of the aberration can be approximated as a quadratic polynomial within a limited range [99, 152]. A minimum of $N + 1$ images each with a different bias modes are needed to calculate the optimum amplitude of N orthogonal modes in a global optimisation approach [116, 149]; a minimum of $2N + 1$ images are needed when the coefficients are calculated and corrected at the end of one iteration [99, 123, 124, 129, 147, 152]; a minimum of $3N$ images are needed when the coefficient of each mode is calculated and corrected before applying the next mode [71, 102, 117, 129, 132, 146, 148, 153–155, 157–160].

In machine learning based sensorless AO methods, the back propagation (gradient descent) algorithm is used for the network convergence [106, 107, 166, 172, 176]. It is a global algorithm optimising all kernels and weights of the connected layers with a gradient linearly related to the loss function and a pre-set learning rate. Normally hundreds of epochs are required before the network converges and a graphics processing unit (GPU) can greatly reduce the processing time.

In the later chapter 6, an untrained neural network was presented which is also optimised using the back propagation algorithm.

4.5 Summary and discussion

In this chapter, a mathematical framework for all sensorless AO methods was presented. The framework consists of aberration representation modes, optimisation metrics and algorithms. It was shown that sensorless AO methods may have different choices of these three components but can all be illustrated by this framework. We also provided mathematical interpretations on various sensorless AO method designs such as a comparison between orthogonal and non-orthogonal aberration represen-

tations, how orthogonality can affect the choice of the optimisation algorithm and a discussion on how the choice of the system quality metric can affect the definition of the orthogonality as well as how the choice of the quality metric affects the choice of the optimisation algorithm.

With the understanding of this universal framework, different sensorless AO approaches are no longer viewed as separate methods but variations of the unified aberration correction mechanism. Different methods now can be placed together in the framework and each of the components used in various methods can be systematically compared. When designing a sensorless AO method for a particular application, instead of choosing one from existing various methods, each of the three components can be selected separately based on different criteria and thus create a better suited method.

Ideally, the aberration representation modes are orthogonal; the optimisation metric has only one local optimum; the algorithm can converge to the global optimum in a fast and reliable way. In reality, due to limitations of the devices and the computational power, the requirements are often not satisfied. Various studies have worked on overcoming these obstacles to design better methods for sensorless aberration corrections.

In the next two chapters, a systematic comparison among various aberration representation modes through both simulation and experiments was presented; a novel untrained neural network encapsulating physical models of imaging system was demonstrated.

Chapter 5

Generalized aberration representations

In this chapter we discussed the use of a generalized aberration representation to allow a fair comparison between various aberration representations used in different sensorless AO methods within the universal sensorless AO framework defined in the chapter 4. A systematic comparison can be conducted to search for a more effective choice of representation for certain imaging applications. Before explaining the general representations, a brief review of existing methods is provided for the necessary context.

In the pixellated piston-based methods, the pupil is divided into a large number of separate pixels within which the phase is modelled as a constant piston value [128, 137, 140, 151]. This constitutes a piecewise constant model for the phase aberration. The applied bias aberration consists of a single pixel, whose phase is changed between measurements [Figure 5.1(a)]. The phase of a portion of the pixels each varies at a different frequency and a sampled sequence of measurements over time is recorded. The varying intensity is Fourier transformed with respect to time to determine its phase which corresponds to the phase of each pixel.

In the “pupil segmentation” zonal methods, the pupil is separated into a number of zones within which the phase is defined by three parameters corresponding to the piston, tip and tilt components. This constitutes a piecewise linear model of the aberration. There are two implementations of the method - single zone illumination [100, 104, 125, 130, 131, 133, 135] and full pupil illumination [126, 136]. For both implementations, a series of varying piston, tip and tilt components were applied in different pupil zones to modulate the pupil wavefront [Figure 5.1(b)]. Optimisation was performed either through estimation of the lateral location of the object [100, 104, 125, 130, 131, 133, 135] or through maximisation of fluorescence intensity depending on the implementations [126, 136].

In the modal sensorless AO methods, the aberrations are represented by a series of modes defined over the whole pupil. Common choices of bias aberrations used in modal methods include the Zernike polynomials [70, 71, 84, 100, 103, 105–109, 114–117, 120, 124, 127, 129, 139, 141, 146–148, 152, 153, 155–157, 172, 177], deformable mirror modes [111, 123, 159–161] and Lukosz modes [99, 173, 174], of which the mirror modes and Lukosz modes are variants derived from the Zernike modes to better suited for specific applications.¹ In these methods, a mode is chosen and a short sequence of different bias amplitudes is applied in order to provide an estimate correction for that mode through the chosen optimisation algorithm to maximise an image-based quality metric [71, 102, 117, 129, 132, 146, 148, 153–155, 157–160]; alternatively, a sequence of bias aberration modes are applied before optimisation in order to estimate several modes simultaneously [99, 116, 123, 124, 129, 147, 152].

These methods, as originally introduced, were conceived as different sensorless AO concepts. However, by placing them into the proposed framework defined in the

¹The methods using piecewise modal modes like the Hadamard-Walsh and polar Walsh modes are not discussed in this chapter because Hadamard-Walsh modes represent, in effect, the same forms of aberrations as the piecewise constant scheme. For a more detailed discussion on the similarities and differences between Hadamard-Walsh modes and the piecewise constant scheme please see appendix A

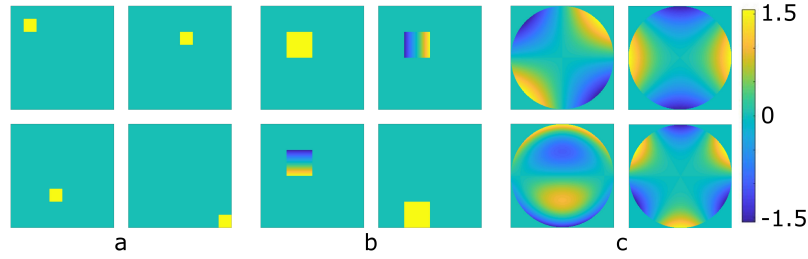


Figure 5.1: Four examples of modes used in: (a) the pixellated piston based method, (b) the “pupil segmented” zonal method, and (c) the Zernike based method.

chapter 4, we can see that the one of the primary differences between them is contained within the aberration representation. In the first case, the aberration is represented by a large number of separate zones with only piston applied within each zone [Figure 5.1(a)]; in the second case, there is a number of zones with 3 modes applied within each zone [Figure 5.1(b)]; in the final case, only one zone and multiple modes applied within the zone [Figure 5.1(c)]. These three seemingly different methods can now be considered as special cases of a more general scheme in which aberrations are defined both in terms of zones within the pupil and modes within each of the zones. A wider range of hybrid zonal/modal representations can be defined that form a spectrum ranging from purely zonal through hybrids to purely modal methods, while maintaining the same degrees of freedom in each case.

In order to ascertain whether, for a particular task, a more suitable basis set exists in the whole spectrum of aberration representations, we chose to explore this broader range of hybrid representations, rather than just the specific cases that have been presented previously in the literature.

Dependent upon the imaging application, including the nature of the specimen and microscope design, the nature of aberrations can vary: they can have smooth low-order shapes (due to refractive index mismatch and system misalignments) or complex high-order variations (such as when imaging deep inside a scattering tissue); the amplitude of aberration can be small or large. Moreover, there can be significant variations in SNR: the signal can be high or low, dependent on numerous factors such

as fluorophore density and imaging efficiency. Thus, we narrow this wide parameter space by choosing representative scenarios for modelling that encompass different combinations of simple vs complex aberration shapes, small vs large aberration amplitudes, and high vs low SNR.

5.1 Choice of microscope

For the purpose of comparing different aberration representations, we chose to simulate aberration correction processes and perform experiments on a 2-P excitation fluorescence microscope (Figure 5.2). The 2-P microscope has been the subject of most adaptive optics developments including the various methods mentioned above. The 2-P microscope has often been used for deep tissue imaging [47, 178] where sample induced aberrations can be significant. The SLM is used to introduce both the input aberrations and corrections (which will be explained later in section 5.3 and 5.4.1) for the AO scheme.

While the image formation processes vary between different microscope modalities, we expect the general trends revealed through studying the 2-P microscope will also be observed in other microscope systems. This approach is justified, as many of the approaches already used in sensorless adaptive optics have been shown to function in a similar manner across different types of microscope (for example the modal method described above has been tested on confocal [108, 109, 111, 115–117], non-linear [109, 123, 124, 129, 146, 147], widefield [99], fibre imaging [115], SIM [152, 153], STED [70, 71, 154, 155], SMS [157–160] and 4Pi [159, 160] microscopes).

5.2 Choice of optimisation metric and algorithm

In this chapter, the two-photon fluorescence intensity is chosen to quantify the output image quality. It is a common choice for many sensorless AO applications in

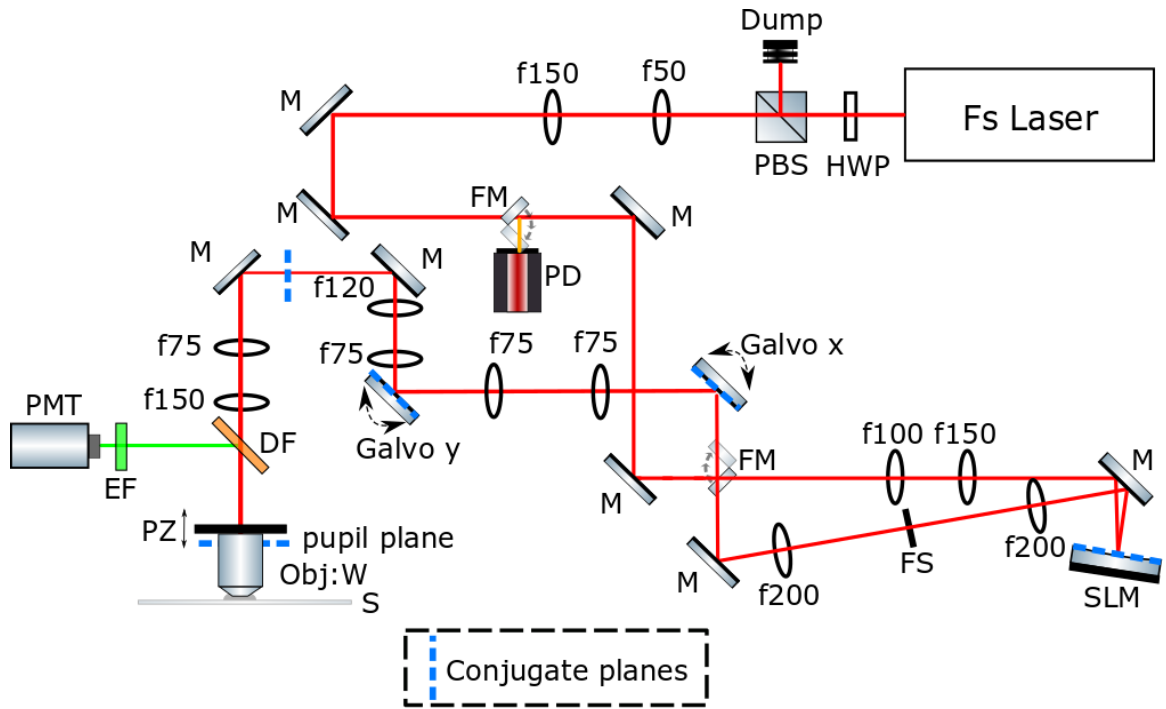


Figure 5.2: Configuration of the 2-P microscope: Ti:Sapphire femtosecond Laser; half wave plate (HWP); polarisation beam splitter (PBS); laser beam dump (Dump); lens with focal length = x mm (f_x); broadband dielectric mirror (M); motorized flip mirror (FM); photo diode for alignment (PD); spatial light modulator (SLM); field stopper (FS); X galvanometer (Galvo x); Y galvanometer (Galvo y); high NA (1.2) water immersion objective lens (Obj:W) mounted on a Z-piezo translation stage (PZ); Dichroic filter (DF) allow two photon signal from fluorophores to be reflected through emission filter (EF) into a photo-multiplier tube (PMT).

non-linear fluorescence microscopy [103, 109, 116, 121, 122, 124, 126, 128, 129, 134, 136, 139, 142, 146, 147]. It could be calculated using the common approximation of the focal field through a scaled Fourier transform of the pupil field. Although a high NA lens is used in the microscope, since the illumination beam is linear x polarised light and the contributions of axial field is small comparing to lateral field, using vectorial diffraction theory for fluorescence intensity analysis would obtain very similar results to the simple thin lens Fourier optics analysis while the later is much easier to calculate. The intensity distribution was found as the square of the modulus of the focal field [179, 180]. The pupil phase was described on a circular pupil P by the phase function $E(\mathbf{x})$, where $\mathbf{x} = (x, y, z)$ is the Cartesian coordinate vector in the pupil space; $|E(\mathbf{x})| = 1$ only over P since we assumed a uniform illumination; $\arg[E(\mathbf{x})]$ is the pupil phase of the illumination beam. The intensity distribution in the focal plane was then calculated as

$$I(\mathbf{v}) = |\mathcal{F}\{E(\mathbf{x})\}|^2 \quad (5.1)$$

where $\mathbf{v} = (u, v, l)$ is the Cartesian coordinate vector in the focal region and $\mathcal{F}$ represents the three-dimensional Fourier transform between the $\mathbf{x}$ and $\mathbf{v}$ domains. If the specimen for imaging is described by the fluorophore density $s(\mathbf{v})$, then the integrated two-photon fluorescence signal detected by a large area detector (for a parking PSF) would be given by [181]

$$f = \iiint_P s(\mathbf{v}) I(\mathbf{v})^2 d\mathbf{v}$$

The simulations presented in this chapter assumed imaging of a uniform volume specimen, for which $s(\mathbf{v}) = 1$ and

$$f = \iiint_P I(\mathbf{v})^2 d\mathbf{v} \quad (5.2)$$

Noise was simulated to follow a Poisson distribution with a mean and variance (σ) proportional to the square root of the photon count [182]. The simulation results presented in this chapter are acquired when no noise is added to the output metric f (high SNR case) and when Poisson noise with $\sigma = \sqrt{500}$ was added to f (low SNR case).

To optimize the quality metric, we chose a three-point quadratic optimisation algorithm (3N algorithm) for the convergence of each aberration representation mode [129] For the optimisation for the single i th mode $\varphi_i(\mathbf{x})$, three measurements were taken of the fluorescence emission. The first measurement f_0 was taken with no additional aberration applied by the adaptive element; for the second measurement f_1 , a bias aberration of $b\varphi_i(\mathbf{x})$ was applied; for the third measurement f_2 , a bias aberration of $-b\varphi_i(\mathbf{x})$ was applied. The correction amplitude was then calculated as [99, 147, 152]

$$a_{i0} = -\frac{b(f_1 - f_2)}{2(f_1 - 2f_0 + f_2)} \quad (5.3)$$

Following this, a correction of $a_{i0}\varphi_i(\mathbf{x})$ was applied to the adaptive element and the correction procedure moved on to the $(i + 1)$ th mode. This algorithm permits precise correction of aberrations if the signal follows closely a quadratic approximation as a function of the input mode amplitude. Such an approximation is appropriate if the sampling spacing around the peak of the function is even over a suitable range. Optimisation schemes employing more measurements can be useful to cover a larger range of aberration amplitudes, although this comes at the expense of requiring more specimen exposure and time [129].

In the simulations, the results of after one and two iterations of correction were presented. In the experiments, the discussions were limited to only one iteration of optimisation for each mode. Though some schemes perform better with more iterations (shown in the simulations), the additional specimen exposure and time required generally has a detrimental effect on the imaging task [183].

5.3 Choice of simulated input aberrations for correction

It is important to prevent bias when comparing different aberration representations. For example, it should be expected that input aberrations comprising a sum of aberration modes would be corrected most precisely by an aberration representation set defined using the same modes as the input aberration. However, in most practical situations, the aberrations experienced in the microscope contain components that cannot be compensated perfectly by the adaptive element. For this reason, we chose to model the input aberration using Fourier components, which was a different basis to any set of modes demonstrated for corrections later in the chapter. This ensured that the residual uncorrectable aberration $\varepsilon \neq 0$ for all demonstrations.

An input aberration $\Phi(\mathbf{x})$ can be generated as a linear sum of several chosen unit Fourier components ($\hat{\Phi}_w$) at frequency w and each component with a chosen RMS amplitude A_w

$$\Phi(\mathbf{x}) = \sum_w A_w \hat{\Phi}_w(\mathbf{x}) \quad (5.4)$$

A single unit component $\hat{\Phi}_w(\mathbf{x})$ with frequency equals to w was given by

$$\hat{\Phi}_w(\mathbf{x}) = \frac{\Phi_w(\mathbf{x})}{\sqrt{\langle \Phi_w(\mathbf{x}), \Phi_w(\mathbf{x}) \rangle}}$$

where $\langle \Phi_w(\mathbf{x}), \Phi_w(\mathbf{x}) \rangle$ is the inner product which is defined in this case as

$$\langle \Phi_w(\mathbf{x}), \Phi_w(\mathbf{x}) \rangle = \frac{\int_P [\Phi_w(\mathbf{x})]^2 d\mathbf{x}}{\int_P d\mathbf{x}}$$

$\Phi_w(\mathbf{x})$ is calculated from $\Phi'_w(\mathbf{x})$ by removing its mean

$$\Phi_w(\mathbf{x}) = \Phi'_w(\mathbf{x}) - \frac{\int_P \Phi'_w(\mathbf{x}) d\mathbf{x}}{\int_P d\mathbf{x}}$$

Note that this step is necessary because the frequency, w , of the Fourier series selected as the input aberrations may not be a whole number and thus the aberration $\Phi_w(\mathbf{x})$ may not be multiples of sinusoidal cycles across the aperture.

$\Phi'_w(\mathbf{x})$ is defined as

$$\Phi'_w(\mathbf{x}) = \mathbf{A}^T \sin_{2D}(w \mathbf{R} \cdot \mathbf{x} + \boldsymbol{\tau}) \quad (5.5)$$

where $\mathbf{A}^T$ is a 1×2 horizontal vector representing random amplitudes in two dimensions (A_x and A_y); $\mathbf{A}$ has a modulus of 1. $\sin_{2D}([a, b]^T)$ is a function taking a vertical vector as an input which provides an output as a vector equal to $[\sin(a), \sin(b)]^T$. w is the chosen frequencies of the sinusoids. $\mathbf{R}$ was the 2×2 rotation matrix showing the random orientation of the sinusoids. $\mathbf{x}$ was a 2×1 vertical vector indicating the spatial coordinates on the pupil (x and y). $\boldsymbol{\tau}$ was a 2×1 vertical vector giving the random spatial offset of the sinusoidal pattern (τ_x and τ_y).

In this model, low frequency Fourier components are analogous to the low-order aberrations, such as imaging through a medium with smooth refractive index variations, encountered in microscopes [Figure 5.3(a)]. Aberrations induced from highly disordered scattering media, such as when imaging deep through thick tissue, contain high frequency components and can be simulated by high order Fourier components [Figure 5.3(b)].

For the discussion of this chapter, we used Fourier series with order below 8π in one-dimensional ($w \leq 8\pi$), i.e. a maximum of 4×4 periods across the x-y pupil plane.

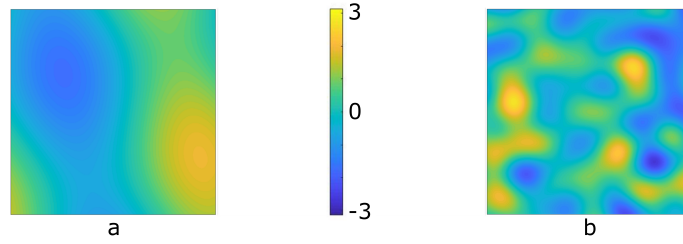


Figure 5.3: An example of (a) randomly generated low order ($w = 1.6\pi$) Fourier based aberration; (b) a random combination of Fourier components containing high orders. The maximum order (w) appearing in this pupil is 7.6π .

5.4 Simulated and experimental comparisons of aberration representations over a square pupil

In order to perform a fair comparison between the different approaches, we need to define a spectrum of aberration representations that permits variations of the aberration expansion while maintaining constant total degrees of freedom. To enable a rigorous comparison at the modelling stage, we initially chose a 2-P system with a square pupil, which can be subdivided into equal sized square zones; different layouts allow us to assess aberration representations with sub-zones of different sizes; different sized sub-zones support a different number of modes within each zone. This arrangement facilitates comparisons across different zonal/modal ratios in a way that would be difficult to formulate using circular pupils. While this square pupil is not normally used in practical microscopes, the difference should not affect the overall conclusions drawn from the results concerning the relative merits of the different zonal/modal aberration representations. Circular pupils are addressed in later experiments.

5.4.1 Square pupil aberration representations

We defined a particular set of modes to have Z zones, each of the zones supported M modes, such that the total degrees of freedom were given by the product $N = ZM$. To facilitate comparisons between different sets of representations we selected the number

of degrees of freedom N to be constant. To maintain square symmetry, we set $Z = 4^\zeta$ and $M = 4^\mu$, where ζ and μ are non-negative integers such that $\zeta + \mu = \text{constant}$.

Given the square symmetry of the zones, we chose to use a 2D Legendre expansion within each zone. The 2D polynomials are defined as $P_j^{2D}(x, y) = P_m(x)P_l(y)$, where $P_m(x)$ is a Legendre polynomial of order m . Legendre polynomials $P_n(x)$ in one dimension can be generated recursively by $(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$ with $P_0(x) = 1$ and $P_1(x) = x$. The restrictions on the values of the indices j , m and l are given by $j = (\max(m, l) + 1)^2 - \max(m, l) + l - m - 1$, where $\max(m, l)$ returns the larger value between m and l [184]. (See Table 5.1)

Table 5.1: Examples of the first 9 two-dimensional Legendre polynomials.

j	(m, l)	$P_j^{2D}(x, y)$
0	(0,0)	1
1	(1,0)	x
2	(1,1)	xy
3	(0,1)	y
4	(2,0)	$(3x^2 - 1)/2$
5	(2,1)	$(3x^2 - 1)y/2$
6	(2,2)	$(3x^2 - 1)(3y^2 - 1)/4$
7	(1,2)	$(3y^2 - 1)x/2$
8	(0,2)	$(3y^2 - 1)/2$
...	...	...

These Legendre modes demonstrate over a square region of support the same orthogonality² as do Zernike polynomials over a circular region. For the case where the number of zones $Z = 1$, the set of modes across the full pupil is defined by the first

²The orthogonality of aberrations for the 2-P signal metric used in this chapter is defined and discussed in the Appendix A.4

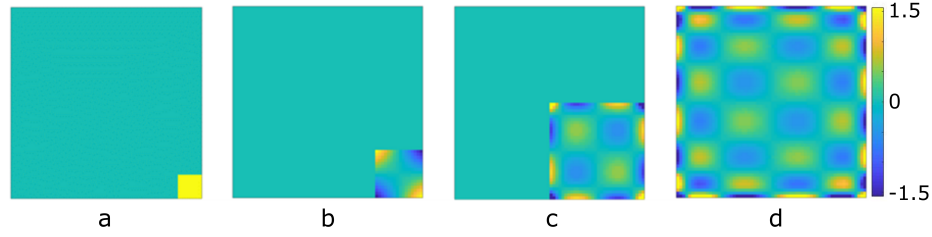


Figure 5.4: An example of one mode from each aberration representation set (a) Z64M1, (b) Z16M4, (c) Z4M16 and (d) Z1M64 respectively.

$M = N$ 2D Legendre polynomials. For cases where $Z \neq 1$, each of the full pupil modes is defined such that $Z-1$ of the zones have zero value; the remaining zone contains one of the first M 2D Legendre modes, where $M = N / Z$. In all of these cases, there are N modes in total defined across the full pupil.

In the numerical modelling, the total degrees of freedom N was chosen to be 64. We, therefore, included four sets of representations into this analysis: Z64M1, Z16M4, Z4M16 and Z1M64, where $ZpMq$ is the notation used for the set with p number of zones in the pupil and q number of modes in each zone (Figure 5.4).

5.4.2 Simulation and experimental results comparing the four aberration representations

We investigated the effectiveness of four aberration representations (Z64M1, Z16M4, Z4M16 and Z1M64) for correcting input aberrations with varying frequencies ($0.2\pi \leq w \leq 8\pi$) and RMS amplitudes ($0.1\pi \leq A_w \leq \pi$). For each selection of w and A_w , 20 aberrations were generated by randomly choosing the amplitude ratio A , the orientation R , and the spatial offset τ . (For the definitions of w , A_w , A , R and τ , please see the section 5.3). The statistics were then derived from this ensemble of results by averaging them for each frequency w and amplitude A_w setting. The four correction models were simulated to correct the aberrations and the results for one iteration of corrections were presented below where two scenarios for high and low

SNR were shown in the Figure 5.5(a) and (b) respectively to observe which correction model was more robust towards noise.

From the noise-free results [Figure 5.5 (a)], we saw that the single zone set of modes (Z1M64 in red) gave the best correction for most of the simulated aberrations. Figure 5.5 (a1) showed profiles of the mean corrected intensities of correcting 20 randomly generated aberrations for each set of representations along the line labelled ‘a1’ on Figure 5.5 (a) where frequency $w = 1.4\pi$. The black hexagons indicated the average initial intensity before correction. We note that the input aberration could contain significant tip, tilt and defocus components, particularly at lower values of w , which did not affect the intensity since we assumed a uniform volume specimen. The colored markers indicated the mean corrected intensity for the different aberration representation sets. In all cases, correction succeeded, although the benefit was the greatest for Z1M64 (red stars). Figure 5.5 (a2) showed the equivalent profiles along line ‘a2’ on Figure 5.5 (a) which corresponded to more complex input aberrations ($w = 7.2\pi$). In this case, the average corrected intensity was much lower than that shown in (a1). At higher amplitudes, above $A_w = 1$ rad, there was little benefit in correction from any of the four methods. However, at lower amplitudes, it could be seen that all methods provided some level of correction with Z64M1 (blue crosses) performing marginally better than the others.

As many practical imaging scenarios worked at low SNR, we also explored the case when Poisson noise was introduced to the system in our simulation. Comparing Figure 5.5(b) to (a), it was observed that the small sized zone set (Z64M1 in blue) was more susceptible to the effects of noise and the medium sized zone sets Z4M16 and Z16M4 performed better for more complex, low amplitude aberrations. Furthermore, all sets failed when correcting wavefronts with a large A_w and w at low SNR.

The trends observed in the simulations were confirmed through experiments on fluorescent bead specimens. $2\mu m$ fluorescent beads were diluted and dried on a cov-

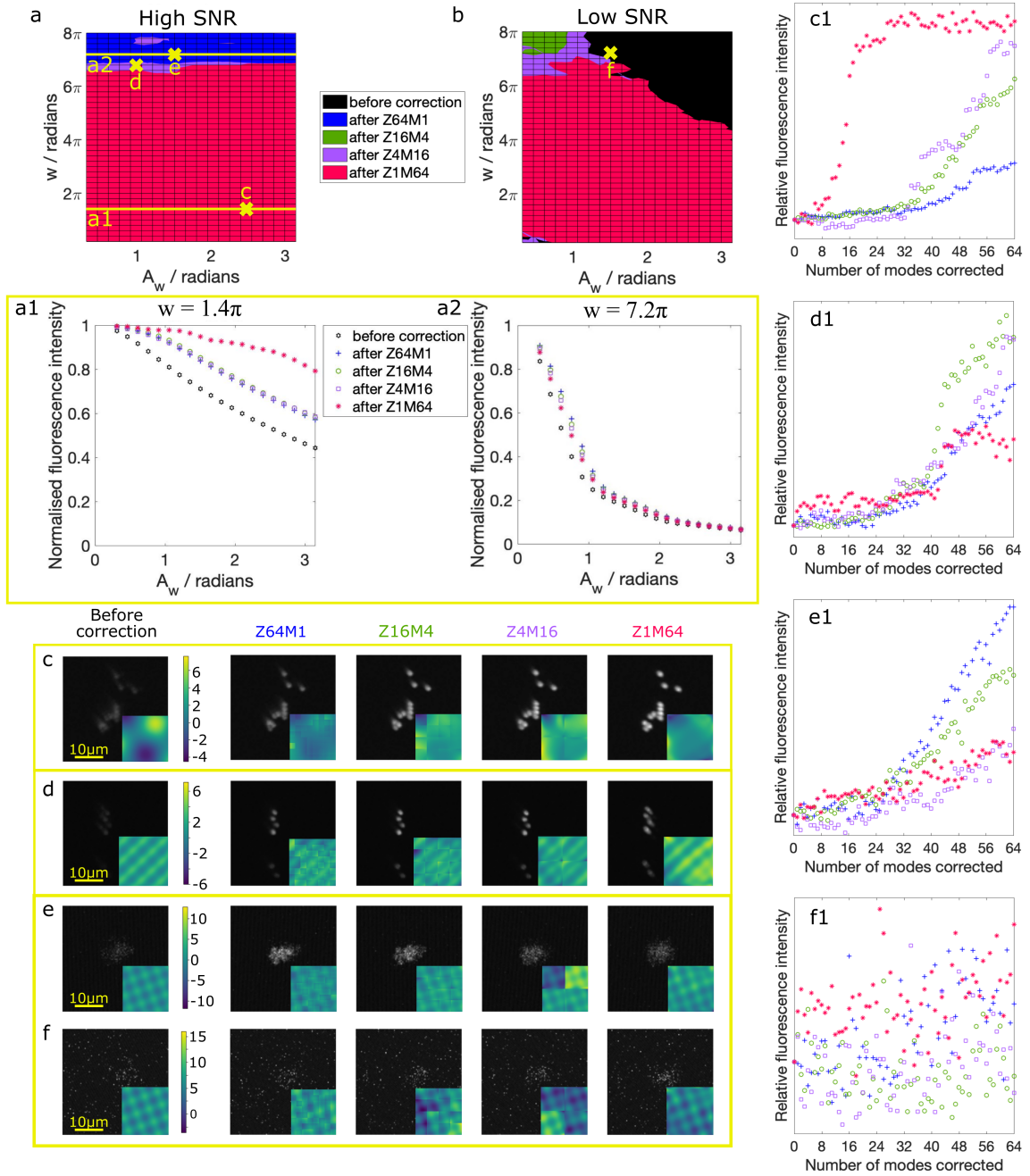


Figure 5.5: (a) and (b). Simulation results comparing correction using the four aberration representations. The colour indicated the method that provided the greatest mean signal (of 20 simulated experiments) after one cycle of correction. The region colored black indicated no correction using any representation set was effective. The labels 'c', 'd', 'e', 'f', 'a1' and 'a2' showed the properties of the input aberration used in the corresponding sections of this figure. (a1) and (a2) represented lines through the plot as shown in (a). They showed the mean signal (of 20 simulated experiments) after correction for each representation. (c), (d), (e), and (f) showed experimental results imaging $2\mu\text{m}$ fluorescence beads before and after corrections with the different representations; the insets (bottom right of each image) showed the wavefronts for each case. The colorbars showed the phase in radians for each row. The four plots (c1), (d1), (e1), and (f1) displayed the correction process when correcting the aberrations from (c), (d), (e) and (f) respectively.

erslip before sealing to a sample slide. The spherical aberration introduced by the coverslip was compensated mostly by the correction collar of the objective lens; any residual system aberration was compensated by the SLM using Zernike polynomials over the whole circular pupil before experiments being carried out. During the experiment, only light from the square pupil enclosed by the full circular objective pupil was allowed to enter the objective lens; light outside of the square pupil was deflected away by a grating pattern displayed on the SLM. When an aberration with $A_w = 2.5$ and $w = 1.4\pi$ was introduced into the system [Figure 5.5(c)] for correction, the images were best after Z1M64 correction; the correction process plot [Figure 5.5(c1)] also suggested that Z1M64 was the fastest as the fluorescence intensity was close to the optimised value after correcting the first 32 modes while the other three methods only reached the maximum after correcting all the 64 modes; thus the single zone method (Z1M64 in red) was shown to be the most suitable model for correcting simple shaped wavefronts with small frequency components w . Figure 5.5 (d) showed results with input aberration $A_w = 1.0$ and $w = 6.8\pi$; the medium sized zone sets (Z4M16 in purple and Z16M4 in green) gave better corrections than the other two; the correction using Z16M4 was also observed to be the fastest as seen in Figure 5.5 (d1), after correcting the first 45 modes, the fluorescence intensity of Z16M4 was much higher than the other three. Figure 5.5 (e) presented results with input aberration $A_w = 1.5$ and $w = 7.2\pi$; the shapes of the beads were revealed after small zone correction (Z64M1 in blue), but remained blurred after using the others; Z64M1 also gave the fastest correction process among the four as the slope of the relative fluorescence intensity growth is the steepest [Figure 5.5(e1)]. Figure 5.5(f) shows the same scenario as Figure 5.5(e), but with lower SNR, which was effected by reducing the excitation laser power; no representation set provided effective correction and only the single zone set (Z1M64 in red) gave marginal improvement of the image brightness after correction; it was similarly shown on Figure 5.5(f1) where no improvement

in fluorescence intensity was observed throughout the correction process. Note that for the four experiments (c), (d) (e) and (f), the photon multiplier tube (PMT) gain was adjusted so that the signals for each experiment can be displayed and compared under the same intensity scale.

The noise conditions presented when obtaining the results of Figure 5.5(f) were not directly comparable to those used in the simulation of Figure 5.5(b). In a real system, many factors, such as the laser power, fluorophore density, aberration amplitude, scanning speed, quantum efficiency of the photodetector and background noise [182], can all affect the SNR. However, the results were representative of the trends to be expected in a noisy system.

5.4.3 Simulation results of multiple correction iterations comparing the four aberration representations

After one iteration of correction as described in section 5.2 and 5.4, the obtained correction $\sum_i^N a_{i0}\varphi_i$ for all N modes (as defined in section 5.4.1) was applied to the pupil. The correction process was repeated again starting from the first aberration representation mode for the second correction iteration. Figure 5.6 (a), (b), (a1) and (a2) showed results after two iterations of correction in each case. Figure 5.6 (c) and (d) showed correction processes during five iterations of correction when correcting a simple ($w= 1.4\pi$, $A_w = 2.5$) and complex ($w= 7.2\pi$, $A_w = 1.5$) shaped aberration respectively.

Comparing Figure 5.6(a) and Figure 5.5 (a), it was observed that aberration was still corrected better in general by the large sized zone representation Z1M64. Comparing Figure 5.6 (a1) and Figure 5.5 (a1), it was observed that the fluorescence intensity was improved for all representations after 2 iterations compared to after 1 iteration and Z1M64 still corrected much better than the others. When correcting aberrations with more complex shapes (larger w), the medium sized zone represen-

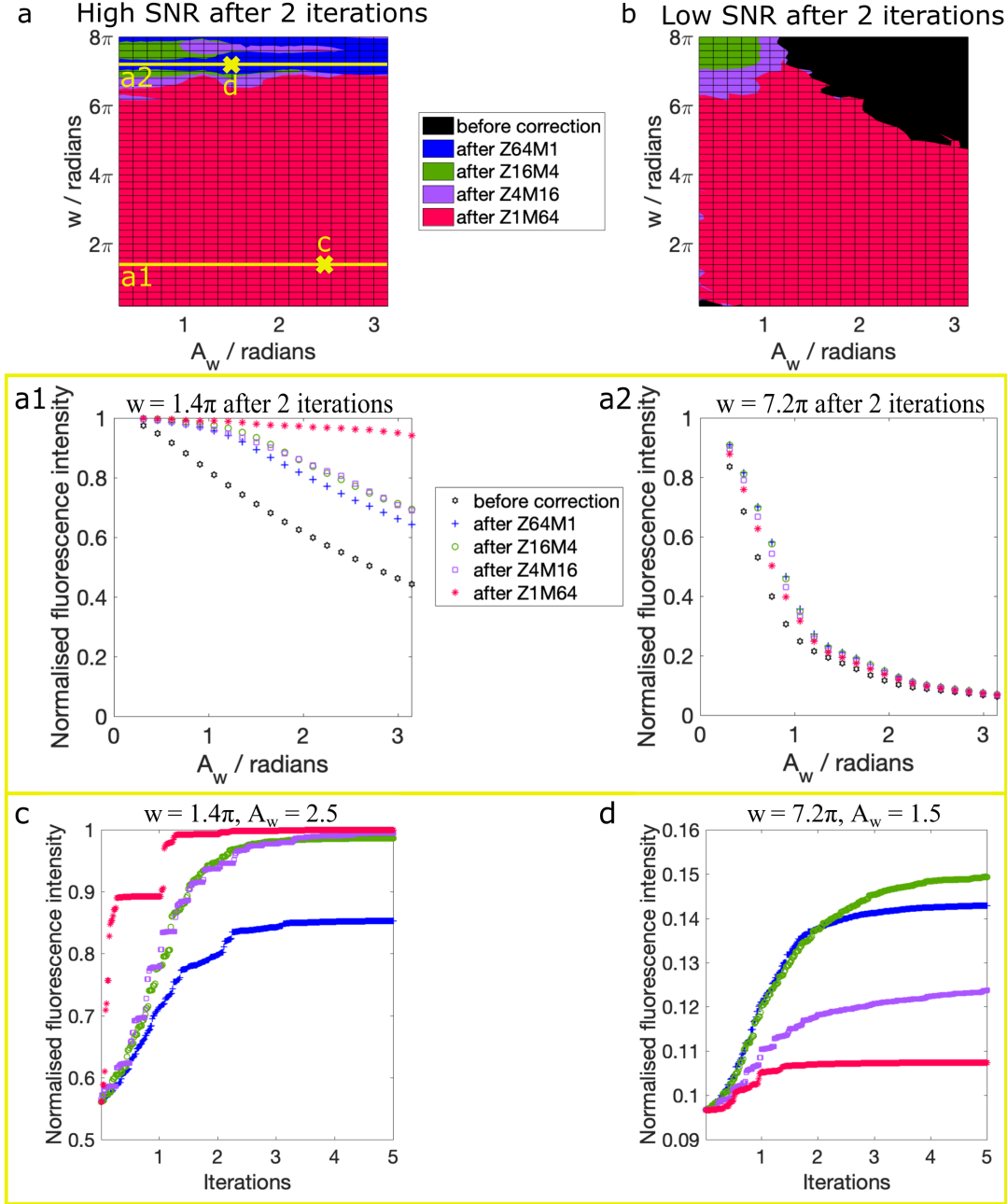


Figure 5.6: (a) and (b) Simulation results comparing corrections after 2 iterations using the four different aberration representations at high and low SNR respectively. The colour indicated the method that provided the greatest mean signal (of 20 experiments) after two cycles of correction, as a function of the RMS aberration amplitude A_w and the complexity, defined as the frequency component w . The region colored black indicated no correction using the four aberration representations was effective. (a1) and (a2) represented sections through the plot as shown in (a). They showed the mean (of 20 experiments) signal level after 2 rounds of correction for each of the representations. (c) and (d) represented simulated corrections of input aberrations with properties labelled on (a). A total of 5 iterations were corrected for each method.

tations Z4M16 and Z16M4 started to perform slightly better than the small sized zone representation; however, Figure 5.6 (a2) suggested that representations with the best performances might only be marginally better than the others after 2 iterations. Comparing Figure 5.6 (b) and Figure 5.5 (b), no significant difference was observed for one and two iterations of correction when SNR was low.

Figure 5.6 (c) and (d) showed that when correcting both a simple and a complex shaped aberration, the increase in fluorescence intensity was the fastest during the first iteration; after 2 rounds of correction, only little improvement could be observed. When the aberration has a simple shape, apart from zonal method Z64M1, all the other three methods achieved really good aberration compensation after around three iterations (as shown in Figure 5.6 (c)); when the aberration has a complex shape, small sized zone representations (Z16M4 and Z64M1) give in general better corrections than large sized zone representations (Z4M16 and Z1M64) with Z16M4 having the best performance after 3 rounds of corrections.

5.5 Experimental comparison of aberration representations over a circular pupil

The results in the previous section 5.4 provided mathematically complete and systematic like-for-like comparisons of the effectiveness of different aberration representations for correction while maintaining fixed degrees of freedom. However, the square pupils used in the analysis are not commonly used in practice for microscopy. We therefore performed further comparisons among the most common three aberration representation sets over circular (or near-circular) pupils, using a similar conceptual approach. In these cases, the approach necessitated slight variations in pupil shapes and degrees of freedom. For these experimental demonstrations we used thin fluorescently labelled biological specimens and simulated aberrations through introductions

of controlled phase patterns using the SLM.

5.5.1 Circular pupil aberration representations

Similar to the approach in section 5.4.1, we defined a spectrum containing different aberration representation sets over a circular pupil that permitted variations of the aberration expansion while maintaining the total degrees of freedom N . In this case, we aimed for N to be as close to 36 as possible. 36 is chosen for convenience in the zonal/modal expansion. We also aimed to create aberration representation sets to cover a plane as close to a circular pupil as possible. For the small sized zone set, we had a total of 37 square zones ($Z=37$) mapped to a circular pupil with piston ($M=1$) applied to each zone [Figure 5.7(a)]. For the medium sized zone set, we had 12 square zones ($Z=12$) mapped to a circular pupil with piston, tip and tilt ($M=3$) applied to each zone [Figure 5.7(b)]. For the large sized single zone set, we chose to use Zernike polynomials up to mode $M=36$ (Noll index [185]) over the whole circular pupil [Figure 5.7(c)]. Note that the small sized zone Z37M1 set corresponds to the representation used in the pixellated method [128,137,140,151] and the medium sized zone Z12M3 set closely modelled the representation used in the ‘pupil segmentation’ zonal method [100,104,125,126,130,131,133,135,136] mentioned earlier in this chapter. We note that for the sets Z37M1 and Z12M3 some areas of the circular pupil are missing. However, as they cover about 94% and 91% respectively of the full pupil area, this would not have a significant effect on our comparison.

5.5.2 Experimental results using tissue specimen and SLM-induced aberrations

For these experiments, the input aberration was defined as in the previous section 5.3 using a Fourier basis and applied using the SLM. The slide for imaging was a thin

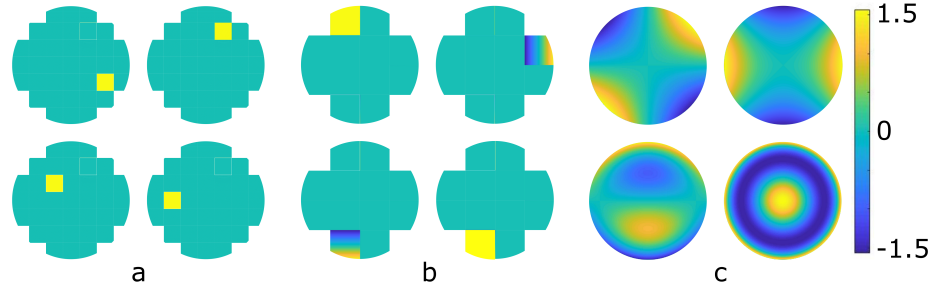


Figure 5.7: (a), (b), and (c). Four examples of each of the aberration representation sets Z37M1, Z12M3 and Z1M36 respectively.

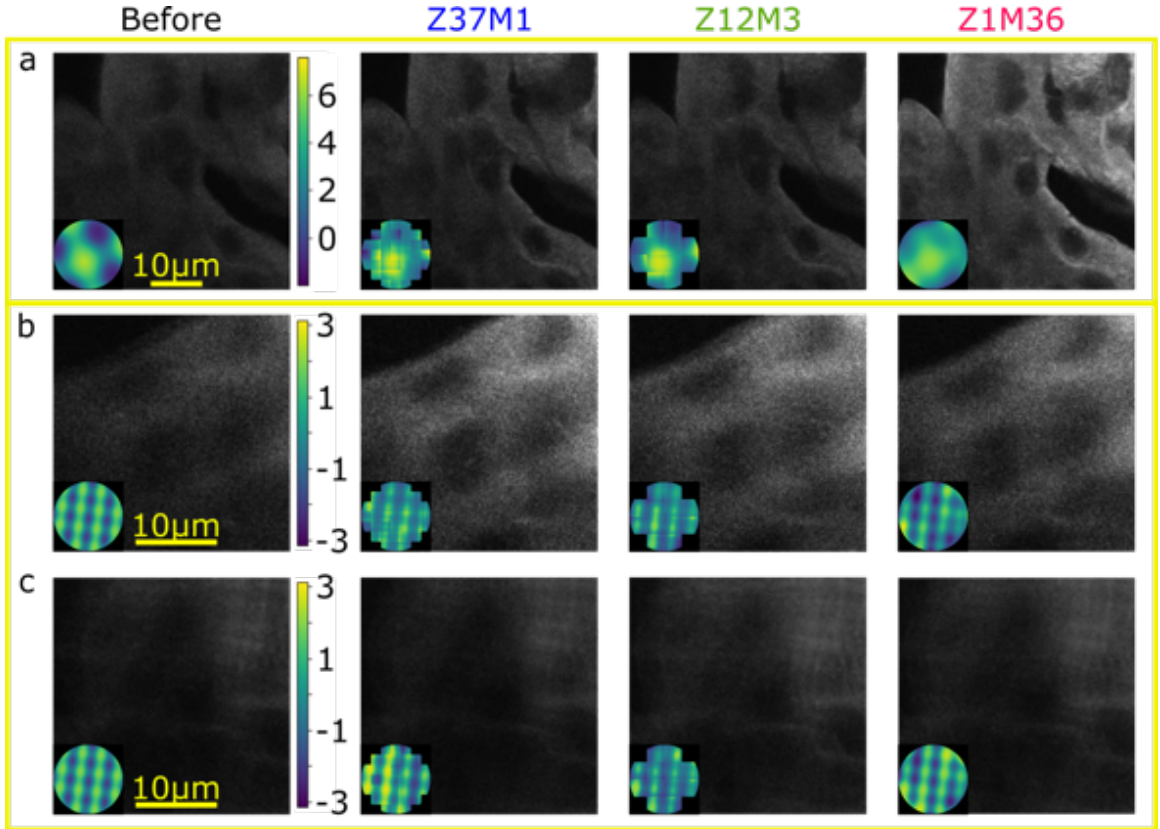


Figure 5.8: Correction of aberrations over a circular pupil using a FluoCells prepared slide containing $16\mu m$ cryostat section of mouse kidney stained with Alexa Fluor 488 wheat germ agglutinin (W-11261), Alexa Fluor 568 phalloidin (A-12380) and DAPI (D-1306). These demonstrations show the relative performance of the different aberration representation sets for variations in input aberration (a) a low frequency input aberration; (b) and (c) a high frequency input aberration. In (b) imaging was performed at high SNR; in (c) imaging was performed with low SNR. The inset in the bottom left of each image showed the wavefront before (the first column) and after (the second to fourth column) correction. The colorbar showed the wavefront phase in radians for each row.

($16\mu m$) mouse kidney specimen. Two particular demonstrations were performed using two different types of input aberration: one with low spatial frequency components [Figure 5.8(a)] and one with high frequency components [5.8(b) and (c)]. In the latter case, the laser power was adjusted to obtain images with large [Figure 5.8(b)] and small SNR [Figure 5.8(c)]. Note that again the PMT gain was adjusted for each demonstration; each row in the figure 5.8 is shown under the same intensity scale for comparison.

The experimental results agreed with our previous conclusions that the large sized zone representation (Z1M36) corrected the best when aberration had simple shapes (small w) [Figure 5.8(a)]. The small sized zone set (Z37M1) performed slightly better when the input aberration had complex shapes (large w) but was more susceptible towards noise. [Figure 5.8(b)] For low SNR, no set gave effective correction. [Figure 5.8(c)]

5.5.3 Comparison of circular representation sets correcting specimen-induced aberration

To further validate the above observations, we extended the experimental investigation to cover the correction of specimen-induced aberrations in thick samples. We tested the three aberration representation sets described in the section 5.5.1 in three different biological samples, where aberrations had been caused by focusing through the tissue at different depths.

Figure 5.9(a) shows images of denticle cells at $100\mu m$ deep inside a *Drosophila melanogaster* larva. 3rd instar Oregon-R *Drosophila melanogaster* larvae were dissected in 0.3 mM Ca^{2+} HL3 to prepare a larval fillet; larvae were stained with HRP (conjugated to Alexa568 fluorophore), DLG (detected with secondary antibody conjugated to Alexa488 fluorophore) and DAPI to detect nuclei. The larvae were mounted in 65% vectashield. Specimen-induced aberrations were corrected using the three

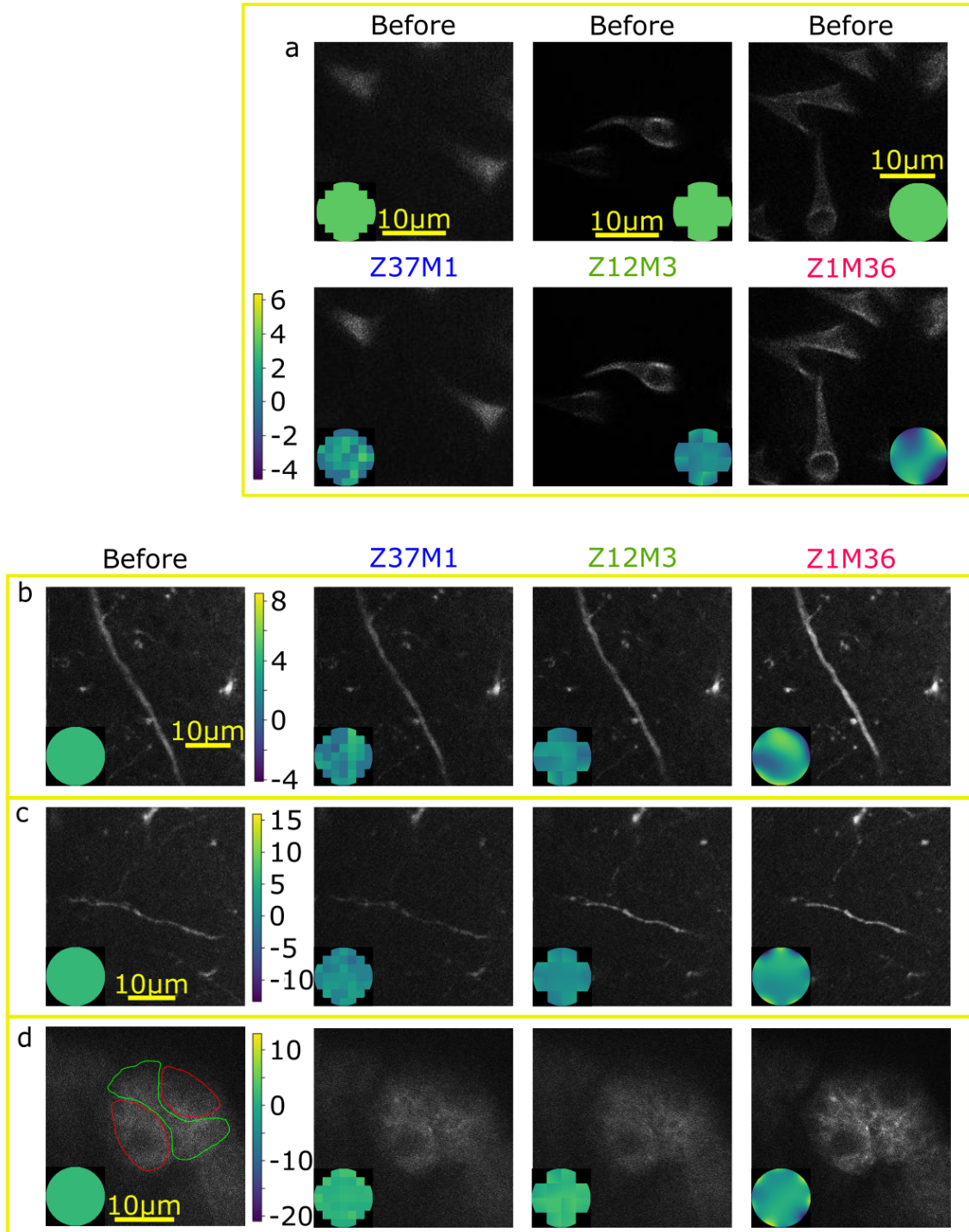


Figure 5.9: (a) Denticle cells imaged at $100\mu\text{m}$ deep inside a *Drosophila* larval fillets sample. The first row are images before aberration correction and the second row include images after correcting using representation sets Z1M36, Z12M3 and Z37M1. Wavefronts before and after correction are shown in the insets. The colorbar shows wavefront phase in radians. (b) and (c) neuronal processes at $183\mu\text{m}$ and $225\mu\text{m}$ respectively in the fixed, extracted brain of a Thy1-YFP mouse in a region where a dendrite was contained within the field of view. The first column are images before aberration correction and the second to fourth columns include images after correcting using representation sets Z1M36, Z12M3 and Z37M1. (d) *Drosophila* larval brain neuroblast (enclosed by red line) and progeny (enclosed by green lines).

aberration representations at three different sites, to obviate the effects of photo bleaching. As these sites were adjacent, the sample induced aberrations were expected to be similar in terms of shapes and amplitudes. The same laser power and PMT gain were used for all three experiments to ensure a fair comparison. All three sets improved the image quality; as seen in the images after correction, which were brighter and showed sharper features. The large sized zone Z1M36 set gave a more effective correction, as seen in the image improvement.

Figure 5.9(b) showed neuronal processes at $180\mu m$ depth in a fixed Thy1-GFP mouse brain. The brain was fixed by transcardiac perfusion with a solution of 4% paraformaldehyde. The whole brain was then extracted, immersed overnight in the paraformaldehyde solution, and glued onto a petri dish with the cortex remaining accessible to the objective lens. For imaging, the sample was submerged in water and a coverglass was balanced on the curved surface of the brain. In this case, the large sized zone Z1M36 set provided the best aberration correction. After Z1M36 correction, more specimen details were revealed and the brightness of the neuronal process was clearly increased. For the other correction schemes, the improvement was less obvious.

Figure 5.9(c) showed correction for imaging of neuronal processes at a depth of $225\mu m$ in the same fixed Thy1-GFP mouse brain. At this location, the wavefront correction appeared to contain higher order components with large amplitudes and the initial SNR reduced significantly. In this case, the single zone Z1M36 set gave the best correction compared to the other two methods, possibly because it is more robust in noisy scenarios, although Z12M3 also generated an improvement.

Figure 5.9(d), we imaged neuroblasts at about $100\mu m$ deep inside a freshly extracted, non-fixed *Drosophila* larval brain. *Drosophila* larval brains expressing Insc-Gal4 and mCD8-GFP were dissected from wandering L3 larvae in supplemented Schneiders medium, as described in [186], except for mounting, which was in agar

wells on glass slides and covered with a coverslip. After Z1M36 correction, the definition of membranes (marked by mCD8-GFP) was greatly improved, with the outlines of the small progeny cells and subcellular structures visible. The Z12M3 correction made some progeny outlines visible, whilst the Z37M1 correction achieved sharper membrane and subcellular structures relative to pre-correction. Though all methods provided some level of correction, Z1M36 performed the best for this specimen.

It is interesting to observe that in Figure 5.9(c) and (d), the wavefronts after correction using the single zone scheme Z1M36 can contain large amounts of some high order modes, despite successful correction overall. This is manifest in the ‘wrinkles’ that can be observed around the edge of the pupil, which are due to the presence of high azimuthal order Zernike polynomials in the correction aberration. These high order modes are likely to be artefacts due to incorrect optimisation, as it is improbable that these modes are actually introduced by the specimen. The real specimen induced aberration is likely to be smaller in amplitude than the correction aberration shown. It is probable that the method has incorrectly introduced these modes in an attempt to correct high order spherical-like aberrations, which are likely present. In advanced schemes, prior information on the nature of the specimen-induced aberrations could be used to regularize such unwanted phenomena.

5.6 Discussion

Overall, the modelling and experimental results point to the conclusion that large sized zone sets are better at correcting aberrations with simpler shapes and they are also less sensitive to noise; small sized zone sets work slightly better when aberrations have more complex shapes, while they are more susceptible to noise.

The results can be explained by the auto-correlations (defined in Appendix B) where both simple shaped aberrations with low spatial frequency components and

modal representations with full pupil expansions have high spatial auto-correlations; in comparison, both complex shaped aberrations with high spatial frequency components and zonal representations have low spatial auto-correlations. A high spatial auto-correlation means a shape changes gradually with gentle gradients or looks continuous; when zonal representations are used in attempts to correct these kinds of simple gradual aberrations, mismatches occur between any two discrete zones and hence result a poor mapping comparing to using continuous modal representations. On the other hand, a low spatial auto-correlation means a shape has a high spatial variations with steep gradients or looks discontinuous; when modal representations are used in attempt to correct these kinds of complex spatially varying aberrations, the first few low order modes contribute little to the corrections; furthermore, aberrations cause a low signal level which results a high possibility of erroneous predictions of low order modes. In contrast, each zone in a set of zonal representations contributed equally to the correction process; moreover, due to its discrete property and 2π phase wrapping, each component in the zonal set has a limited optimisation searching range [187] since its modulation to the metric reading is periodic (see Appendix A.3); limited searching ranges are desirable especially when correcting complex shaped aberrations. Therefore, zonal representations have a potential to correct more complex shaped aberrations better while the modal methods possibly correct simpler shaped aberrations better. Furthermore, large sized zone sets are expected to be more robust against the effect of noise as they create a greater varying change in the fluorescence signal when phase modulation is applied to the pupil plane (see Appendix A.2).

Results following the correction of specimen-induced aberrations suggest that [Figure 5.9(a), (b), (c) and (d)], the specimen-induced aberration can be better corrected by the large sized zone Z1M36 set. It is possibly because of limitations of the other representations in low SNR scenarios. This is also possibly the reason why Rodríguez

et al. [144] managed to image at around $900\ \mu\text{m}$ inside a brain tissue with the help of a pupil segmented AO method while with a modal based AO method Streich *et al.* [146] can image at around $1200\ \mu\text{m}$ inside a brain tissue; in both cases, the imagings were reported to happen in a 3-P system with illumination wavelengths also at $1300\ \text{nm}$.

It is known that specimen-induced aberrations can take many forms, including smooth forms, high spatial frequency components and discontinuities [75]. A Fourier representation for modelling input aberrations was chosen in order to cover a range of possibilities, while not being biased towards any choice of correction sets. All sets have limitations in the form of aberrations that they can adequately correct, as they are limited by the representation used. For example, in all of the zone-based sets, there exist discontinuities at the zone boundaries. This might be alleviated by inter-zone phasing [125], if assumptions can be made about phase continuity. However, in other situations, sudden steps in phase may be required for optimal aberration correction [75].

The systematic analysis in this chapter provides fair comparisons across the spectrum of different aberration representation bases to illustrate their relative capabilities for realistic scenarios. The detailed analysis has concentrated on a single cycle of correction, to provide a like-for-like comparison that seeks to minimize measurements and hence specimen exposure. Several demonstrations in the literature have used multiple iterative correction cycles. Through modelling of these scenarios, we found that multiple cycles increased the effectiveness of all sized zone sets, although at the cost of higher specimen exposure.

The conclusions drawn here could also be readily extended to other aberration representation bases such as DM deformation modes, Hadamard-Walsh modes or other arrangements of pupil zones, since in this chapter we demonstrated phenomena using both Legendre and Zernike based modes in both square and circular pupils, re-

spectively, to correct both Fourier based introduced aberration and sample-generated aberration.

5.7 Conclusion

We provided generalised aberration representations which permitted for the first time a systematic fair comparison of different representation configurations for sensorless AO schemes. This was illustrated through simulation and experimental demonstration of their operation in a two-photon microscope. Furthermore, we have shown that these representations are all special cases of a broader spectrum of zonal/modal hybrid methods. This opens up the possibility of other aberration representations beyond those already presented in previous research and can be useful in specification of future AO schemes.

Overall, the results suggested that the large zone schemes generally provided better correction in a single cycle and were more robust against noise; the small zone schemes exhibited better performance when the input aberration had complex high order shapes, but they were more susceptible to error in low SNR imaging.

The results have been presented in a manner that decoupled the essential phenomena from the chosen aberration representations. We therefore expect that the major conclusions will be broadly applicable and are not limited to the particular aberration representations employed here. Moreover, we expect that many of the phenomena observed here through demonstration on a two-photon microscope will also be common to other microscope modalities. Hence, the results will be useful in design of future sensorless AO schemes for a wide range of applications

Chapter 6

Machine learning based sensor-less adaptive optics

As discussed in chapter 4, sensorless AO consists of three key components: the aberration representations, the system quality metric and the optimisation algorithm for convergence. Conventional sensorless AO methods normally defined a single valued system quality metric, such as the total fluorescence intensity [102, 103, 108–112, 114, 116–119, 121, 122, 124, 126, 128, 129, 132, 134, 136–142, 146, 147, 149–151, 153, 156, 177] and the frequency components which reflects the image sharpness [70, 99, 115, 123, 148, 152, 154, 157, 158], to quantify the image quality; by introducing various perturbations to the system, the variations of the quality metric are monitored and optimised using an algorithm. These single valued system quality metrics are effective, but not the most efficient. Since they are one-dimensional metrics, they only show the amplitude but no other property of the aberration. (For more detailed analysis, please see chapter 4.) This is not desirable as images contain phase information and by condensing the whole image into a single valued metric, the useful information from the image is not fully extracted.

For most of the imaging systems, the output images can be mathematically

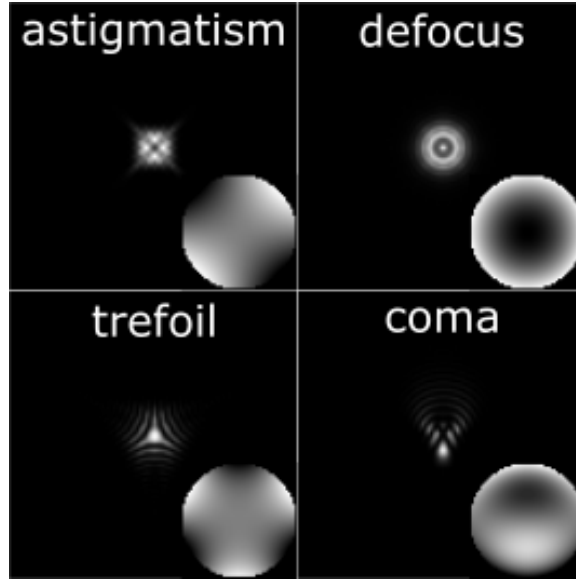


Figure 6.1: The synthetic PSFs of different Zernike modes by modelling a widefield system. A simple thin lens Fourier optics model is used by assuming a low NA lens.

expressed as a convolution between the sample structure and the PSF of the system [19, 20, 182]. Since the deformation of the PSF is distinctive for different pupil wavefront shapes (see Figure 6.1), the output images should contain information which helps us to identify not just the amplitude, but also the shape of the pupil aberration.

In each image, sample structures vary spatially while the PSF can be assumed to stay relatively similar across the field of view. When imaging a uniform area, PSF deformations are not obviously captured in the image. PSF information is mostly captured when there is a sharp edge or step change in the sample structure. Useful information can be extracted if we can differentiate the spatially varying sample structures from the relatively stable PSF especially close to sample edges and thus deduce the pupil phase information. CNN is a suitable option in fulfilling this task since it is convolutional based and can extract common features from the whole image.

A NN is a way for machine to learn the relationships between inputs and outputs in a statistical manner [188]. Different from conventional method of which we are trying to derive explicit mathematical expressions to calculate the outputs from the

inputs, a NN uses weights, biases and non-linear activations to connect the inputs to the outputs. A lot of training data sets (input-output pairs) are fed into a NN; the NN generates outputs from the inputs which then compared to ground truth outputs; the differences between the outputs from the NN and the ground truth are used in a back-propagation gradient descent algorithm to update the weights and biases in the NN. After several iterations and trainings on a number of data sets, weights and biases in the NN finally converge. This trained NN can then be used to derive outputs for new inputs.

A CNN is a form of a NN used specifically for image processing [188, 189]. As each image contains much more data points than normal inputs of a NN, using a NN to process image is computationally expensive. Furthermore, in an image, adjacent data points are normally more related than distant data points; fully connected layers, which analyse all the relationships between any two data points, are computationally inefficient. In each image, there are also common useful features which may exist at multiple locations; convolution based NN can thus effectively extract those common features using only a few kernels.

There are already many attempts of applying ML based methods to improve image quality through post image processing in terms of denoising and deconvolution [190–196], since ML based methods are relatively robust against random errors. The methods have been shown effective, but also have their intrinsic disadvantages. These ML based methods are observed to be limited to certain types of noise; they cannot be generalized to cover a wide range of noise forms. ML based methods are normally in the form of a “black box”; the intermediate layers are not directly interpretable and thus it is hard to verify the credibility of these computer processed or generated images. These kinds of methods are thus very useful when processing normal daily photos, such as pictures of a cat, but less useful when processing scientific images, where the accuracy and credibility are their priorities.

Recently, more research focused on using ML to deduce pupil phase aberrations [106,107,120,165,166,172]. This is an important step to eventually including ML into image formation and aberration correction loop as it now has the potential to be used in beam shaping; the pupil phase aberrations predicted by ML may be corrected using adaptive elements and the beam thus shaped for image generation. The feedback signal collected by the imaging system will also allow the results from ML to be verified.

Typically, a DNN is designed in an end-to-end style where the inputs and outputs of the network are decided; A NN framework is designed with multiple layers and neurons of which the size of the network is believed to be large enough to represent a close approximation to the relationship between the inputs and the outputs [188]. The network is thus treated as a black box; the network converges purely based on the training dataset that is fed in. The advantage of this style is that the network is thus unbiased; there is no pre-knowledge about the relationship is imposed on the network; the network itself decides on which important features to extract and how the features are connected based on the training dataset. The disadvantages are that the network only works in the domain covered by the training dataset; there is no guarantee that it can be adapted to other similar scenarios which are outside the scope of the data used for training. This is possibly the reason why all the published end-to-end style beam shaping networks were demonstrated only on images of point-like objects (i.e. PSFs) but not on extended objects since it is not feasible to always obtain a large number of labelled training datasets from the specimen of our interest to fully train the network. Another example is the denoising system we mentioned above: one denoising method only works for certain types of noise included by the training dataset. Furthermore, though NNs are relatively robust against random errors, they can be erroneously trained if there is any systematic error existing in the training data. Moreover, this end-to-end style can sometimes be computationally inefficient

especially when we hold some knowledge for parts or the whole of the relationship between the inputs and the outputs; by designing an end-to-end network, we are asking the computer to reproducing some existing knowledge.

On the other end of this NN style spectrum, there are networks designed in a human-guided way [172, 176]. Opposite to the end-to-end approach, it is usually designed incorporating available mathematical models of parts or the whole of the relationship. This type of network has pros and cons opposite to the end-to-end style listed above. It is more computationally efficient but more human-guided and thus biased. This style also may not be feasible in the cases when the relationship is so complicated that no explicit mathematical model for the relationship is available.

Since there is a well defined mathematical model for image generation process in microscopy systems, we worked on a human-guided style NN which were designed to predict the pupil phase aberrations from a minimum of two images of the same FOV; Two or more images, each with a different bias aberration introduced, were fed into the NN and allow the network to converge and thus predict the aberration correction. The simulation results showed that it is a possible approach which will lead to a substantial piece of work.

6.1 An untrained CNN for deducing pupil phase from a pair of images

As we discussed above, there is a mathematical model to work out the PSF if the pupil phase is known; however, the backward conversion is less straightforward since the phase information of the PSF is unavailable. The section above mentioned one approach to work out the pupil phase information from the images through a trained CNN to formulate this backward conversion. The CNN used an end-to-end style which means that no pre-knowledge about this backward conversion is included in

this network. In this section, we present another network in a human-guided style which makes full use of the mathematical model of the imaging system. The only part that CNN needs to predict is the pupil phase aberration which varies from experiment to experiment. The network is thus untrained and there is no training dataset required for the convergence of the network.

There are already human-guided untrained networks that have been used to improve image quality and recover un-aberrated images. Ulyanov *et al.* designed an untrained network (in a U-Net architecture [197]) that was shown to be effective in denoising, inpainting and super-resolution [196]; though this deep image prior network opens the field of using untrained networks for image processing, this network is limited to be used in computer image processing as there is no physical imaging system to verify its results. Inspired by this deep image prior, Wang *et al.* incorporated the image formation physics into a U-net to work out a phase only object from its diffraction pattern (PhysenNet) [176]. However, because the deep image prior network is over-parameterised and can possibly suffer from overfitting, early stopping is thus an important regulariser. Subsequently, Heckel and Hand designed a deep decoder where the network is under-parameterised and thus does not need early stopping [198]. Bostan *et al.* encapsulated the deep decoder and the system physics to reconstruct the object phase and phase aberrations from the captured images (deep phase decoder) [172]. With the physical constraints added onto the network, both PhysenNet and deep phase decoder were successfully demonstrated on simple imaging systems and showed their potential of being used in microscopes.

In both PhysenNet [176] and deep phase decoder [172], the network generated the object structure¹; the forward propagation from the network predicted object structures to the output images is formulated by the mathematical model of the

¹PhysenNet is different from deep phase decoder where PhysenNet generates the object structure from a captured image of the object while deep phase decoder generates object structure from a randomly chosen and fixed tensor.

physical system and the output images calculated were compared to the real captured images; the differences between the calculated images and the captured images were used to update the network to make a better object structure prediction. To allow a network to generate an object structure, the number of parameters in the network needs to be comparative to the number of pixels in the image, which requires a lot of computational power and time for convergence.

In microscopes, the adaptive elements are equipped to correct phase aberrations. The image quality can be improved through focal spot shaping rather than computer processing. We, therefore, designed a network to work out the aberrations from the captured images. Since the aberrations vary from experiment to experiment, the network is untrained. The number of parameters is comparative to the number of Zernike modes included in the analysis, which is much smaller than the number of parameters requested in PhysenNet and deep phase decoder. In this section, we presented a hand-crafted network incorporating the mathematical model of a 2-P and a 1-P microscope to predict the phase aberrations from a minimum of two captured images of the same FOV and with different bias modes applied. Since the image formation is represented by the mathematical model of the microscope, in this case a 2-P microscope and a 1-P widefield microscope, this network can be easily adapted and work on other microscopes by simply changing this mathematical image formation model while avoiding the trouble of re-collecting training data sets and re-training the whole network. The results showed that the network can effectively converge to the right phase aberrations which may be used to control an adaptive element to remove aberrations and improve the image quality.

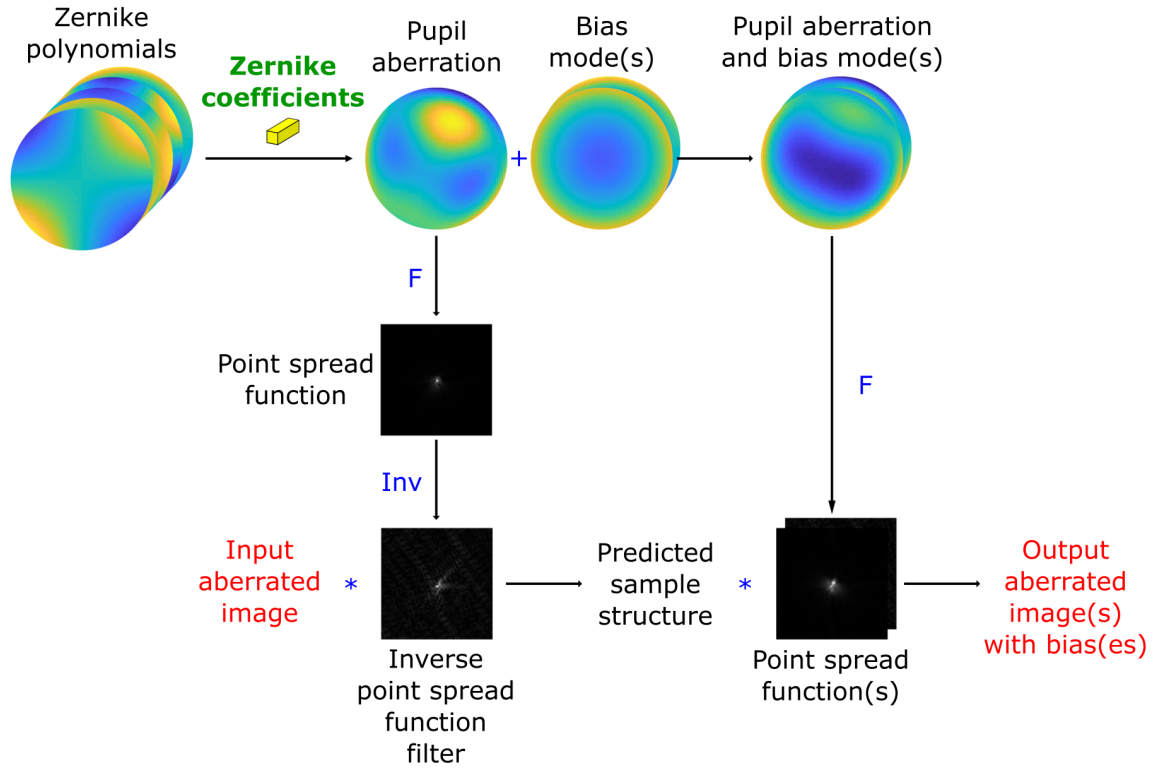


Figure 6.2: A schematic illustration of the untrained network with physical meanings for phase aberration predictions from a set images (two or more images where marked red as shown in the figure). One image without bias phase mode applied is used as the input aberrated image and the other n ($n \leq 1$) images each with a different bias phase mode applied are used as the output aberrated image(s) with bias(es). F is a Fourier transform based function calculating PSF from complex pupil field defined in equation 6.1. Inv is a function defined in equation 6.2, converting PSF to inverse PSF filter such that a convolution between the input aberrated image and the inverse PSF filter is equivalent to a deconvolution between the image and the PSF. The only trainable kernel in this architecture is the Zernike coefficients (marked yellow and stated green).

6.1.1 A schematic illustration of the untrained network with physical meanings

Figure 6.2 is a schematic illustration of this hand-crafted network and the physical meaning of each component. It is designed to work out the phase aberrations from a set of $1 + n$ images of the same sample, where 1 image captured when zero bias is applied and the other n images with n different known bias phase modulations applied by the adaptive element into the system (where $n \geq 1$). The only trainable kernel is the yellow kernel represented the coefficients of the first M Zernike polynomials which were used in the analysis. Multiplying and summing the Zernike polynomials and the coefficients obtains a pupil phase aberration (Ψ_0), which is then added to bias modes to obtain the other pupil phase aberration(s) with bias(es) (Ψ_b). The two (or more) pupil wavefronts (Ψ_0 and Ψ_b) then form PSFs through a Fourier transform based function (F); this conversion function F can be varied to model different imaging systems. Take the 2-P and widefield system as an example, F can be expressed as

$$\text{PSF} = F(\Psi) = |\mathcal{FT}(P_{amp} e^{i\Psi})|^l \quad (6.1)$$

where l is the power of the complex focal field to generate the PSF for imaging; $l = 4$ for a 2-P system and $l = 2$ for a single photon widefield microscope. $\mathcal{FT}$ is the Fourier transform operation. We used a simple thin lens model here P_{amp} is the amplitude of the pupil function representing the intensity information of the pupil field; in this case we assume a uniform illumination across a pupil field such that

$$P_{amp} = \begin{cases} 1 & \text{for } r \leq 1 \\ 0 & \text{for } r > 1 \end{cases}$$

where r is the normalised distance on the pupil plane from the optical axis.

The PSF without any bias is then converted using the Inv function to calculate

its inverse point spread function (iPSF) such that a deconvolution between an image and a PSF can be achieved through a convolution between the image and the iPSF. The Inv function is defined as

$$\text{iPSF} = \text{Inv}(\text{PSF}) = \mathcal{FT}^{-1}\left[\frac{1}{\mathcal{FT}(\text{PSF}) + \delta}\right] \quad (6.2)$$

where δ is a small term to avoid a “0” denominator.

If correct Zernike coefficients are predicted, the deconvolution between the input aberrated image and the PSF (or convolution with the iPSF) will reveal the sample structure which then convolves with the other PSF to form the output aberrated image with bias. If the predicted Zernike coefficients are not correct, the error between the output image and the ground truth image will be used to update the trainable kernel, i.e. the Zernike coefficients.

The advantage of this network is that it is a very flexible framework. It can be applied to various imaging systems through simply changing the F function; Zernike polynomials can also be switched to other orthogonal aberration representation sets which may be better suited for certain applications (as discussed in the chapter 5).

6.1.2 A flowchart of the network with details

In practice, to ensure that the network can converge at a fast rate, multiple layers of kernels and non-linear activations are needed. Non-linear activations are required between any two convolution layers since convolution is a linear operation and without non-linear operations the convolution layers would collapse and condense into one single layer [188, 189]. An easy demonstration is that suppose we have an input x followed by two convolutional layers with trainable kernels a and b . Without any non-linear activations between the two convolutional layers, the output y can be expressed as $y = x * a * b = x * (a * b) = x * c$ where $c = (a * b)$ and $*$ is the convolution operation;

this results the two convolutional layers condensed into one and we can not build the complexity or take the advantages of the neural network. By introducing a non-linear activation such as a tanh function between the two convolutional layers, the output y is now written as $y = \tanh(x * a) * b$ and thus the trainings of the kernel a are separated to the trainings of the kernel b ; this two-layer CNN hence has a potential to map a more complex function and do a more complicated task.

For images with sparse and coarse structures, to increase the sensitivity of the gradient descent algorithm for convergence, the starlet wavelet [199,200] decomposed images instead of the raw images are used for the loss function calculations. Starlets are used to extract isotropic spatially localised fine features which correspond the most to the PSF information. This is realised by a $1 \times 17 \times 17 \times 4$ kernel where 4 starlet wavelets with different scales are used to conduct wavelet decomposition of the output image(s); 17 is the largest size of one starlet wavelet and this size is chosen empirically. This kernel can be adjusted based on the scanning speed and resolution of the imaging system so that the wavelet scale is selected to correspond to the size of the PSF features contained in the images. In general, for small sampling rate, small sized starlets are sufficient to extract PSF information while for large sampling rate, large sized starlets are needed.

A range of network designs of differing complexities were tested. A suitable network design was selected on a heuristic basis that has the right level of complexity to both allow fast convergence and be simple enough to operate. Figure 6.3 shows the details of the network when used in applications. A total of 7 trainable kernels and 6 trainable biases were connected together to generate the predicted pupil phase. After the first convolution kernel the path was split into $1 + n$: one without and n with applying the different bias Zernike modes. $1 + n$ trainable biases were added to all paths with non-linear activations applied; the activation 1 was selected to be rectified linear unit (ReLU) function [201]. After the second convolution kernel and

bias for each path, the non-linear activation 2 was selected to be a sigmoid function [202]. It then convolved with the third kernel and added the third bias, with the non-linear activation 3 to be a tanh function. All three non-linear activations were commonly used in neural network design. The tanh function was chosen to be the last activation because it ranges from -1 to 1 while the other two activations have no negative part. After the fourth convolution for each path, $1 + n$ pupil phases were generated. The pupil phases from the paths with the bias Zernike modes subtracted the corresponding bias mode and were averaged with the pupil phase from the path without the bias mode to generate a single predicted pupil phase; the single predicted pupil phase then added different bias Zernike modes back in each path to generate the pupil phases with bias. The predicted phase without bias was used to calculate iPSF using the F and Inv functions defined in equation 6.1 and 6.2. The phases with bias calculated PSFs using the equation 6.1. The input image captured using a microscope with unknown aberrations present is convolved with the iPSF followed by convolution with the PSF to generate an output image. The calculated output image was wavelet transformed and compared to the ground truth, which was also wavelet transformed from the image captured using the same microscopy system with bias mode applied. The difference between the output and the ground truth was used to generate a loss function for network convergence.

6.1.3 More details on the choices of the network parameters and designs

The trainable kernels were initialized using the Glorot uniform (or Xavier uniform) initializer [203] to randomize the starting values of various kernels. The kernels were regularized using the L1L2 regularizer so that both the first and second moment of the kernels are used for regularization to avoid any kernel possessing extreme values and thus to avoid overfitting.

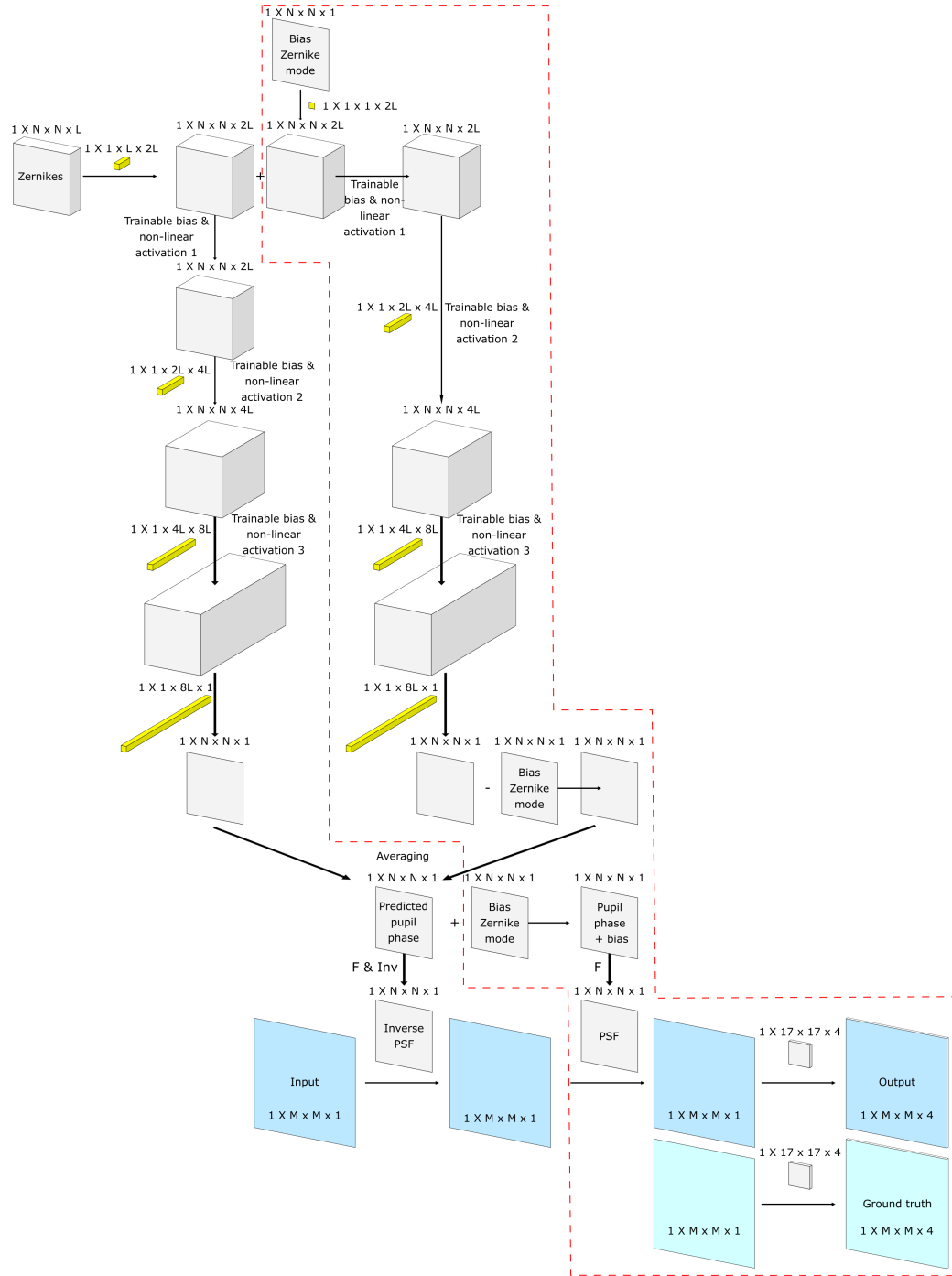


Figure 6.3: A detailed flowchart with dimensions of the different layers and kernels of the network when two (or more) images are used for phase aberration retrieval. Arrows indicate the directions of flow. Blocks by the side of arrows are the kernels for convolutions. Trainable biases are added and non-linear activations are applied after convolutions. Non-linear activation 1, 2 and 3 corresponds to ReLU, sigmoid and tanh respectively. Yellow blocks are trainable kernels for convergence during the training process. Gray blocks are pre-determined or calculated layers or kernels. Blue blocks are the images fed in and calculated by the network while cyan blocks are the ground truth images as references. N is the x-y size of the pupil plane; L is the number of aberration representation modes for analysis; M is the x-y size of the images. The area enclosed by the red box may be replicated by n times if n images with n different known bias phase modulations are fed into the network for phase retrieval.

The loss function is generated through starlet wavelet transforming both the calculated and captured images. After processing the calculated and captured images, the output and ground truth are used to calculate a mean squared logarithmic error (MSLE) as its loss function. The MSLE, $\mathcal{L}_{MSLE}$, is defined as

$$\begin{aligned}\mathcal{L}_{MSLE}(I, I') &= \frac{\iint (\log(I + 1) - \log(I' + 1))^2 dx dy}{\iint dx dy} \\ &= \frac{1}{\iint dx dy} \iint \left(\log \left(\frac{I + 1}{I' + 1} \right) \right)^2 dx dy\end{aligned}\quad (6.3)$$

where I is the ground truth and I' is the output from the network. x and y are spatial coordinates. In comparison, the mean squared error (MSE), $\mathcal{L}_{MSE}$, is defined as

$$\mathcal{L}_{MSE}(I, I') = \frac{\iint (I - I')^2 dx dy}{\iint dx dy}\quad (6.4)$$

Comparing to $\mathcal{L}_{MSE}$, $\mathcal{L}_{MSLE}$ relates mostly to proportional differences rather than absolute values. Since imaging samples with densely distributed structures will result in a larger MSE than sparsely distributed structures, using MSLE allows a consistent performance of the network when imaging different samples.

The Adam optimizer [204] was chosen to be the optimisation algorithm to update the trainable kernels and biases as iterations. The Adam optimizer adapts the learning rate on a parameter basis and it also averages the most recent few gradients to allow a smoother learning process. The effectiveness of the Adam optimizer makes it a suitable choice for this application. In addition, the learning rate was also set to decay exponentially as the number of training iterations to allow a faster and better convergence.

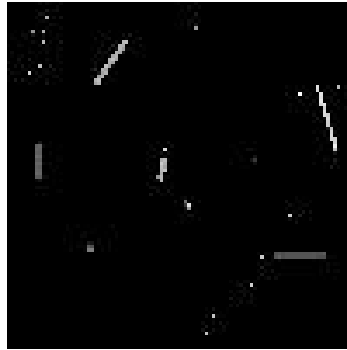


Figure 6.4: A simulated sample with random blocks, lines and points structures.

6.2 Demonstrations of the neural network on simulated data

To demonstrate the network, we generated a sample with random blocks, lines and points as shown in Figure 6.4. The sample was then convolved with PSFs (generated from the F function defined in equation 6.1) to simulate images captured by the imaging systems. These sets of images were fed into the network to test whether the network could correctly predict the unknown input aberrations. A total of 8 modes (defocus, two astigmatism, two coma, two trefoil and spherical) were included for analysis. The network was demonstrated by modelling both a 2-P and a 1-P widefield microscope.

6.2.1 The neural network modelling a 2-P microscope

Two (or more) images were introduced into the network modelling a 2-P microscope (where $l = 4$ for the F function defined in the equation 6.1) for the phase retrieval. One image was generated when the pupil phase was aberrated but without any known bias introduced; the other image(s) was(were) generated when known bias(es) was(were) introduced into the same aberrated pupil phase; the choices of the bias modes were different if more than one image with bias modes were fed into the network. We tested

the network to predict one radians of astigmatism, coma and trefoil aberrations RMS. The known bias added to assist the prediction was defocus when predicting astigmatism and coma; when predicting trefoil, both defocus and spherical were introduced and thus a total of three images were fed into the network.

6.2.1.1 Astigmatism phase retrieval

A unknown aberration consisting only astigmatism in this case was used for prediction. Defocus was introduced as the bias mode which was known to the network. An image generated without bias [Figure 6.5(a)] and an image with bias [Figure 6.5(b)] were fed into the network; insets of Figure 6.5(a) and (b) were the corresponding pupil phases when the images were generated. Results were presented in Figure 6.5(c). The MSLE loss function decreased in general as the number of iterations increased. After about 200 convergence loops, the loss function decreased at a slower speed and the network started to converge to a shape containing mainly astigmatism which was the unknown aberration for the network to predict.

6.2.1.2 Coma phase retrieval

Coma was used in this case as the unknown aberration for prediction. Defocus was again used as the bias mode which was known to the network. An image without bias [Figure 6.6(a)] and an image with bias [Figure 6.6(b)] were fed into the network. Results were presented in Figure 6.6(c). The MSLE loss function decreased as the number of iterations increased and the network gradually converged and identified the unknown aberration with mainly coma.

6.2.1.3 Trefoil phase retrieval

Trefoil was used as the unknown aberration for prediction in this case. Both defocus and spherical were used as the bias modes which were known to the network. Two

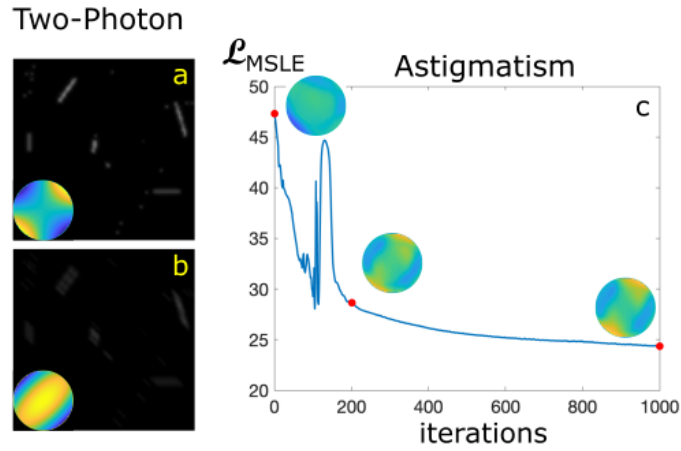


Figure 6.5: Results from the network when modelling a 2-P microscope. (a) and (b) were the two images generated when the system had a unknown pupil aberration of astigmatism, where (b) had an additional known bias mode of defocus introduced. The insets (bottom left) were the pupil wavefronts when the images were generated. The two images (a) and (b) were fed into the network to predict the unknown aberration. (c) was a plot of the MSLE loss function against the number of iterations for convergence. The insets were the pupil wavefronts predicted by the network at the number of iterations indicated by the red dots on the plot.

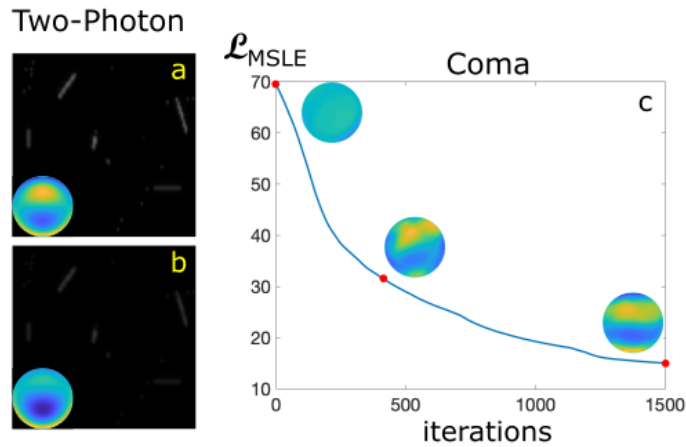


Figure 6.6: As explained in the caption of the Figure 6.5 (a) and (b) were the two images fed into the network for phase retrieval where (a) was generated when the 2-P system had an unknown aberration of coma while (b) had an extra known bias mode of defocus in addition to the unknown aberration. The insets (bottom left) were the pupil wavefronts when the images were generated. (c) was the plot of the MSLE loss function against the number of iterations for convergence. The insets were the pupil wavefronts predicted by the network at the number of iterations indicated by the red dots on the plot.

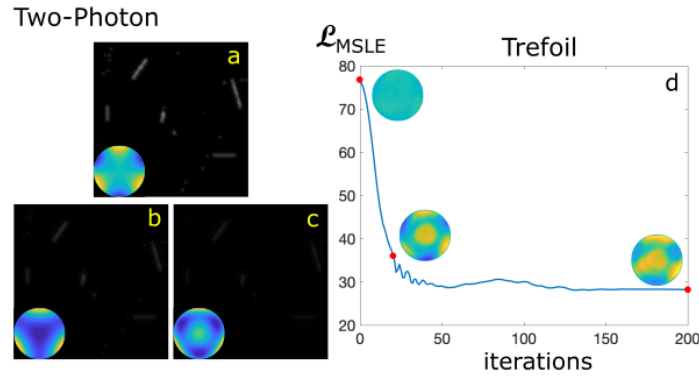


Figure 6.7: Slightly different from Figure 6.5 and 6.6, in this case three images, (a), (b) and (c) were generated when the system had a unknown pupil aberration of trefoil, where (b) and (c) had known bias modes of defocus and spherical introduced respectively. The insets (bottom left) were the pupil wavefronts when the images were generated. (d) is a plot of the MSLE loss function against the number of iterations for convergence. The insets are the pupil wavefronts predicted by the network at the number of iterations indicated by the red dots on the plot.

bias modes instead of one were used in this case because results from several attempts suggested that the performance of the network was more consistent if spherical in addition to defocus were used for phase retrieval. This could be because trefoil has a higher angular order than both astigmatism and coma which results a more complex deformation to the PSF in the lateral plane and thus harder to be identified. By introducing more bias modes and collecting more images of the same FOV, more information could possibly be more easily extracted.

An image without any bias [Figure 6.7 (a)] and two images each with a different bias - defocus [Figure 6.7 (b)] and spherical [Figure 6.7 (c)] were fed into the network. Results are presented in Figure 6.7 (d) where the MSLE loss function decreased for the first 40 iterations and no longer decreased afterwards. The insets showed that the network predicted a wavefront containing a combination of trefoil and spherical after about 20 iterations but more convergence loops did not lead to a better prediction as the loss function was not observed to decrease and the shape of the predicted wavefront did not converge to a shape containing only trefoil.

From results displayed in all three Figures 6.5, 6.6 and 6.7, we see the network

converged towards the correct prediction and the loss function was reduced with iterations in general. The three sets of results suggested that the network has the potential of identifying aberrations (out of eight modes) using a two or three images when modelling a 2-P system.

6.2.2 The network modelling a 1-P widefield microscope

Similar to modelling a 2-P microscope, two (or more) images were introduced into the network for phase retrieval; the network modelled a 1-P widefield microscope in this case where $l = 2$ for the F function defined in the equation 6.1. One difference could be easily observed from the generated images when comparing the 1-P to the 2-P model was that the average brightness of the images was not changed with the aberration for the 1-P model but decreased significantly with an increasing amount of aberration for the 2-P model. By applying the network to demonstrate on a 1-P system, we tested whether the network could correctly predict aberrations when the input image average brightness was not affected by the aberration.

6.2.2.1 Astigmatism phase retrieval

Astigmatism was used as the unknown aberration for prediction. Defocus was used again as the bias mode which was known to the network. An image without bias [Figure 6.8(a)] and an image with bias [Figure 6.8(b)] were fed into the network. Results were presented in Figure 6.8(c). The MSLE loss function decreased in general as the number of iterations increased. After about 500 iterations, the loss function decreased at a slower speed and the network gradually converged to the shape of astigmatism which was the unknown aberration for prediction.

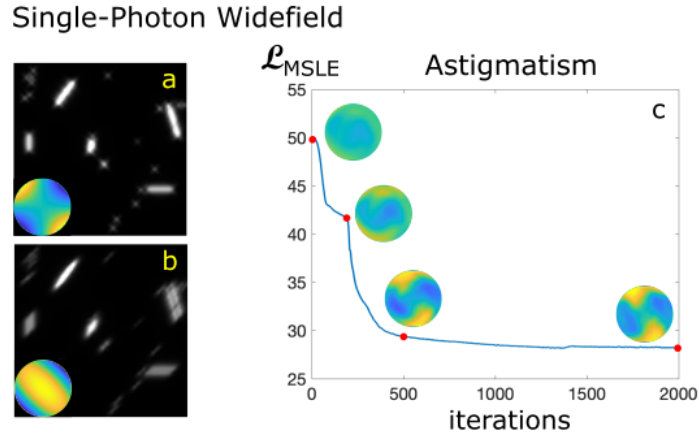


Figure 6.8: Results from the network when modelling a 1-P microscope. (a) and (b) were the two images generated when the system had a unknown pupil aberration of astigmatism, where (b) had a known bias mode of defocus introduced. The insets (bottom left) were the pupil wavefronts when the images were generated. The two images (a) and (b) were fed into the network to predict the unknown aberration. (c) was a plot of the MSLE loss function against the number of iterations for convergence. The insets were the pupil wavefronts predicted by the network at the number of iterations indicated by the red dots on the plot.

6.2.2.2 Coma phase retrieval

Coma was used in this case as the unknown aberration for prediction. Defocus was again used as the bias mode which was known to the network. An image without bias [Figure 6.9(a)] and an image with bias [Figure 6.9(b)] were fed into the network. Results were presented in Figure 6.9(c). The MSLE loss function decreased much more slowly than the previous demonstrations (by a factor of 10); after about 5000 iterations the network started to converge at a slower speed to identify the unknown aberration which was detected mainly as coma but also contained some spurious trefoil components.

6.2.2.3 Trefoil phase retrieval

Trefoil was used as the unknown aberration for prediction. As explained in the section 6.2.1.3, both defocus and spherical were again chosen as the bias modes which were known to the network. An image without any bias [Figure 6.10 (a)] and two

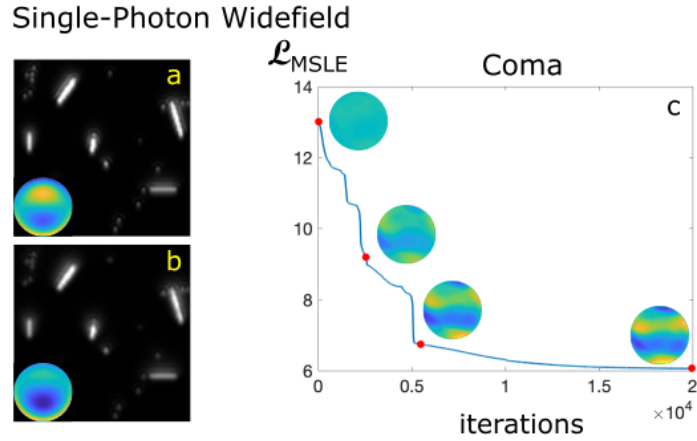


Figure 6.9: As explained in the caption of the Figure 6.8 (a) and (b) were the two images fed into the network for phase retrieval where (a) was generated when the 1-P system had an unknown aberration of coma while (b) had an extra known bias mode of defocus in addition to the unknown aberration. The insets (bottom left) were the pupil wavefronts when the images were generated. (c) was the plot of the MSLE loss function against the number of iterations for convergence. The insets were the pupil wavefronts predicted by the network at the number of iterations indicated by the red dots on the plot.

images each with a different bias - defocus [Figure 6.10 (b)] and spherical [Figure 6.10 (c)] were fed into the network. Results are presented in Figure 6.10 (d) where the MSLE loss function gradually decreased as iterations, but at a slower speed as the results presented when modelling the 2-P microscope. The insets showed that the network slowly converged to a wavefront containing mainly trefoil after about 20,000 iterations.

From results shown in all three Figures 6.8, 6.9 and 6.10, we see the network converged towards the correct prediction and the loss function was reduced with iterations in general. The three sets of results suggested that the network has the potential to identify aberrations (out of eight modes) using a two or three images when modelling a 1-P system.

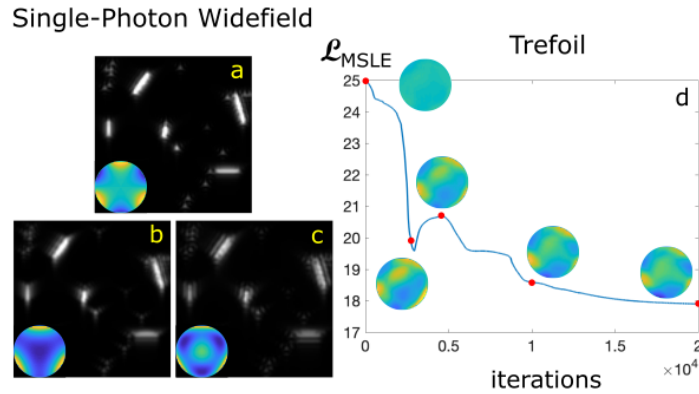


Figure 6.10: Slightly different from Figure 6.8 and 6.9, in this case three images, (a), (b) and (c) were generated when the system had a unknown pupil aberration of trefoil, where (b) and (c) had known bias modes of defocus and spherical introduced respectively. The insets (bottom left) were the pupil wavefronts when the images were generated. (d) is a plot of the MSLE loss function against the number of iterations for convergence. The insets are the pupil wavefronts predicted by the network at the number of iterations indicated by the red dots on the plot.

6.2.3 Discussion

Defocus in all six cases was used as (one of) the bias mode(s) because it reveals the three dimensional shape of a PSF and is a common choice in a lot of machine learning based phase retrieval methods [107, 172, 176]. Spherical was selected in addition to defocus when predicting trefoil because trefoil has a higher angular order than astigmatism and coma while spherical has zero angular order but a higher radial order than defocus; aberrations with high radial orders but zero angular order affect the PSF more strongly in the axial direction, and create only rotationally symmetric effects in the lateral plane. Intuitively spherical was chosen to expose the lateral deformation of the PSF resulting from the high angular order of the trefoil.

Comparing the results presented in the Figure 6.5, 6.6 and 6.7, the loss function decreased at varying rate despite similar learning rate was used. When predicting astigmatism [Figure 6.5] and trefoil [Figure 6.10], the learning rate chosen for convergence seemed to be too large as the loss function did not decrease monotonically as iterations and spikes were observed; on the other hand, the learning rate seemed to

be a bit too small when predicting coma [Figure 6.6] as the loss function decreased at a rather slow speed and did not fully converge after 1,500 iterations. This suggested that the loss function (quality metric) used in this network did not vary isotropically for different aberration modes. This was not desirable because the performance of the network was thus less predictable. This observation may lead to a conclusion that Zernike polynomials used for demonstrations were not optimised for this network (as discussed in chapter 4) and a more suitable set of aberration representation may allow a more consistent performance.

Comparing the results presented in the section 6.2.2 to the results in the section 6.2.1, the network converged in general faster when modelling a 2-P system than modelling a 1-P system. This may be explained as the average brightness of the images generated from a 2-P microscope simulation decreased greatly as the amplitude of aberrations increased; in contrast, the average brightness of images generated from a 1-P microscope simulation stayed constant as aberrations changed; the network may extract not just PSF deformation but also the brightness information which was not relevant to a 1-P system. The results in the section 6.2.2 still suggested that the network can retrieve phase purely based on PSF deformation but at a slower speed.

6.3 Conclusion

We demonstrated a untrained human-guided network for identifying pupil aberrations (out of eight modes) using a two or three images when modelling a 2-P and a 1-P microscopes. Results suggested that this network had a potential of extracting more useful information from images than conventional aberration correction algorithms which normally required a minimum of $N + 1$ images [116, 149] (and more often $2N + 1$ [99, 123, 124, 129, 147, 152] or $3N$ [71, 102, 117, 129, 132, 146, 148, 153–155, 157–160] images) for correcting N modes. This network also had advantages over the

other end-to-end style phase retrieval CNN presented in literature since this untrained network required no training dataset while the other CNNs required a large number of training data which was hard to obtain from microscopy systems; this network was also more versatile as it can be easily adapted to various types of microscopes by simply changing the physical model encapsulated in the network while the other end-to-end style CNN can normally work on only one microscope where its training data was collected from and the network was optimised for; adapting the same network to work on a separate microscope could in practical situations require re-collecting the training dataset from the different system and re-train the network [107]. The network proposed in this chapter also had advantage over other untrained network proposed in literature [172,176] as the number of parameters in this network was compatible to the number of aberration modes and thus much smaller than the networks in literature where the number of parameters was related to the number of pixels in the images and thus our network was expected to have the potential to converge faster.

Although the network only demonstrated on simulated data for now, it has potentials to work well on experimental data. The network is highly under-parameterized (comparing to the number of image pixels) and thus unlikely to fit any noise data; it is thus expected to be robust against random noise and system parameter variation in real systems.

In the future, a more detailed analysis on a suitable set of aberration representations for this network may need to be discussed; although Zernike polynomials used in this work were a common choice for a lot of aberration correction methods, the loss function in this network did not vary isotropically to different modes which made its performance less predictable; an ideal set of aberration representations should allow the network to converge at a relatively similar speed which can greatly ease the operation of the network. Furthermore, a detailed discussion on the choice of bias modes may also help the network to extract more useful information from the im-

ages and thus improve its performance; currently defocus and spherical modes were chosen intuitively since they have a zero angular order; different bias modes may be tested and compared to obtain a more accurate phase prediction from fewer number of sample exposures. Background noise may also need to be added into the testing data to assess the robustness of the method under low SNR conditions.

Chapter 7

Compensation of vectorial state errors using adaptive optics

A range of applications in classical and quantum optics require the control of the spatially variant vectorial state of light - encompassing both polarisation and phase [205–207]. These applications include focusing through high numerical aperture lenses; using optical elements like Q plates [208, 209], metasurfaces and GRIN optics [10, 11]; plasmonic or waveguiding structures [210, 211]; and laser machining [212].

Some emerging applications require compensation of spatially variant perturbations of the vectorial state (vectorial aberrations) when light propagates through an optical system [11, 213, 214]. This can occur, for example, in imaging systems like microscopes or lithography systems whose performance is sensitive to the polarisation state of the light. In addition to suffering from the effects of phase aberrations, which vary from specimen to specimen, the performance of these microscopes also degrades when the light deviates from the ideal polarisation state. This includes super-resolution methods, such as STED and SIM [12–14]. By their nature, these systems require adaptive compensation of phase and polarisation errors. Vectorial aberrations can also be introduced when focussing through birefringent crystals, through stressed

optical elements or due to dielectric coatings [11, 15, 16].

Adaptive optical elements, whose high pixel density provides fine spatial control of the beam profile, could be used to correct the vectorial aberrations introduced in such systems. While adaptive optics has been widely used for phase correction, its use for polarisation control is more challenging, due to more degrees of freedom being required to control the vectorial field. This would require a complex optical system with multiple adaptive elements. Despite this complexity, many systems have already been implemented that use multiple adaptive elements [155, 215, 216]. Note that multiple passes could also be implemented by different areas of the single SLM, so fewer devices could be used in practice [217].

Existing systems for vectorial beam control have mostly been concerned with the generation of a chosen beam profile from a fixed [218–222], known state, such as the linearly polarised output of a laser source. For broader implementation of vectorial beam control, a more general problem must be solved: the conversion between two arbitrary vectorial states. We present a new configuration, which uses four SLM passes, that permits full conversion between arbitrary vectorial states. We further provide a method and analytical expression of using four SLM passes to compensate the effects of any homogeneous retarder on any fully polarized light with arbitrary vectorial states.

7.1 Model system

In order to investigate the capabilities of SLMs in controlling the vectorial state (including polarisation and phase), a system shown in Figure 7.1 is used for analysis. This consists of a sequence of transmissive SLMs positioned along the beam path at normal incidence, such that the distance between the SLMs is small enough so that diffraction effects between elements can be considered negligible. We note also that

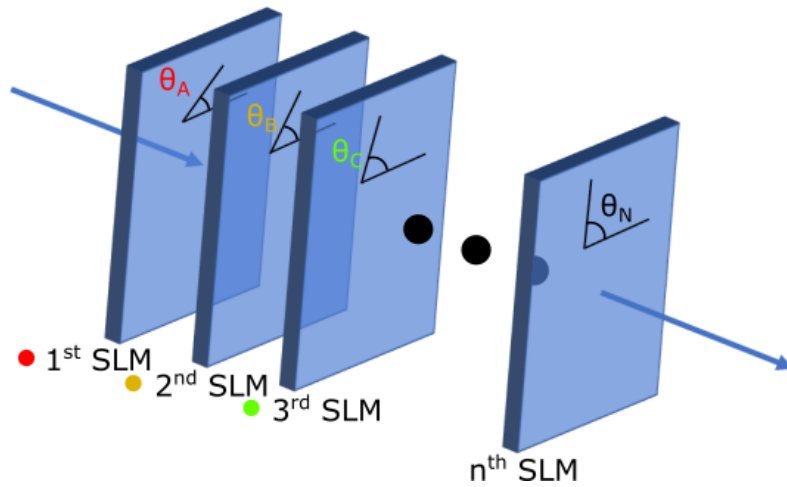


Figure 7.1: Schematic of the multiple transmissive SLM system. Note that in practice most commercial SLMs are operated in the reflective mode. Transmissive devices are drawn here for convenience. The modulation axis of each SLM is set at a chosen angle θ_A , etc.

similar operations could be achieved if suitable 4f imaging systems had been used between each pair of SLMs, although they are omitted here for clarity. While many modern SLMs operate in a reflection rather than transmission configuration, which sometimes requires operation at a slight deviation from normal incidence. We make the approximation here that all beams are at normal incidence.

The illumination is assumed to be a collimated beam of monochromatic, fully polarised light, although the polarisation state may vary across the beam profile. Phase variations may also be present in the incoming beam.

In order to model the operation of the most commonly used, parallel aligned nematic liquid crystal devices, the SLMs are assumed to operate as waveplates (linear retarders) with fixed axis, but variable retardance within a range of $\pm\pi$ radians. The SLMs can be rotated around the optical axis, such that the equivalent waveplate axis of each SLM can be at any chosen orientation. In many practical systems of this type, the SLMs are not mechanically rotated, but additional waveplates are used instead to transform the polarisation state between SLM passes.

The action of each of the SLM pixels is to introduce phase shifts along axis corre-

sponding to the ordinary (fixed phase shift) and extraordinary (variable phase shift) axes of the variable waveplate. The polarisation components projected on the ordinary and extraordinary axes are referred as the eigenmodes of the equivalent waveplate. To assist analysis we define an offset phase shift Φ_δ , which will be the same across the whole beam. The effects of each pixel are represented by Φ_δ absolute phase shift along the ordinary axis and the addition of $\Phi_\delta + \psi$ radians for the extraordinary axis. As the fixed offset Φ_δ will have no effect on the PSF and hence no consequence for our analysis, we will neglect this term from further expressions. The action of the SLM is thus equivalent to the addition of a relative phase of ψ between the modes. The relative phase corresponds to the retardance ψ of the SLM pixel that is related to the change in polarisation state.

A change in the polarisation state induced by the SLM pixel can be visualized as a transition on the Poincaré sphere. If the pixel can be represented by a Müller matrix with primary eigenvector given by the normalised reduced Stokes 3-vector $\mathbf{Q}$ ¹, corresponding to the extraordinary axis of the equivalent retarder, and retardance ψ , then the transition between the input state and the output state corresponds to a rotation through angle ψ , around the axis defined by $\mathbf{Q}$ (Figure 7.2 shows a case when $\mathbf{Q}$ aligns with $\mathbf{S}_1$). Hence, cascaded SLMs can be modelled geometrically by sequences of rotations about different axes on the Poincaré sphere, which correspond to a collection of circular arcs on the surface. We note the special case where the input polarisation is aligned with the primary eigenmode (extraordinary axis) of the SLM and hence the input state lies on the axis of rotation; in this case there is no change in polarisation state, although there is a phase modulation. This corresponds to the operation of the SLM as a pure phase modulator. We can also note when the input is aligned with the secondary eigenmode, there is no polarisation or phase change.

¹A reduced Stokes vector contains the last three elements (the first to third element) of a normal Stokes vector and ignored the zeroth element (as defined in the section 2.1.3.2).

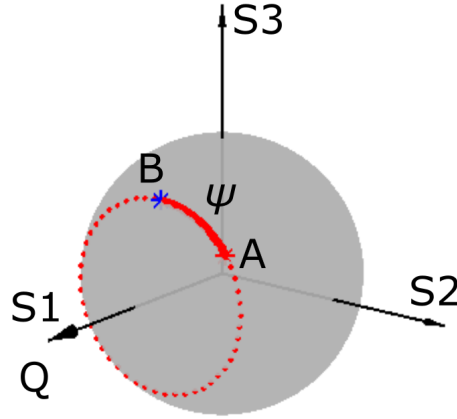


Figure 7.2: A visualisation of a SOP transition due to a retarder on a Poincaré sphere. The extraordinary axis Q of a retarder aligns with $S1$ and the ordinary axis $-Q$ aligns with $-S1$. The initial polarisation state, marked by A , is converted to another polarisation state, marked by B , after passing through a retarder with extraordinary axis Q and retardance ψ . The transition is equivalent to a rotation through angle ψ from A to B about the axis Q .

7.2 Conversions between arbitrary polarisation states

The problem of achieving general conversions between arbitrary polarisation states can be considered into two steps: 1. a conversion from any arbitrary polarisation state to a fixed state and 2. a conversion from a fixed state to any arbitrary state. Since most systems proposed in the literature are designed by assuming a known input state, we first present a brief review and a solution of how to convert from a fixed state to an arbitrary state, followed by a solution of converting from an arbitrary state to a fixed state and finally combining the two steps to achieve a conversion between arbitrary polarisation states.

7.2.1 Conversions from a fixed to any arbitrary polarisation state

As we mentioned above, there are already systems proposed in the literature to generate a defined polarisation and/or phase state from a known input [213–216, 223–225, 225–227]. We provide here a brief review of the operation of these systems, as the basis for later discussion. For all of these systems, the input beam is chosen to have a uniform, linear polarisation state. For generation of arbitrary polarisation states, a minimum of two SLM modulations is required, as the transition in polarisation state requires two free variables. In the optimal configuration, for full polarisation modulation capability, the extraordinary optical axis of the first SLM must be oriented at 45° to the polarisation of the input light. The second SLM is then oriented along the polarisation of the input light. On the Poincaré sphere, the action of the first SLM is to move the polarisation state along a great circle (see the blue path in Figure 7.3 (a)); the second SLM then moves the state along a circle in a perpendicular plane to the plane of the first SLM path (see the red paths in Figure 7.3 (a)). These two transitions can reach any point on the sphere and hence permit the generation of any polarisation state. Although we have assumed that all SLMs are capable of full $\pm\pi$ radians retardance, this same modulation range could also be achieved if the second SLM were limited to $\pm\pi/2$ radians. Note that if the orientation of either SLM is not perfectly aligned, the paths on the Poincaré sphere will not be in perpendicular planes, so there will be regions of the sphere that cannot be reached (see Figure 7.3 (b)); hence, there will be polarisation states that cannot be generated by the system.

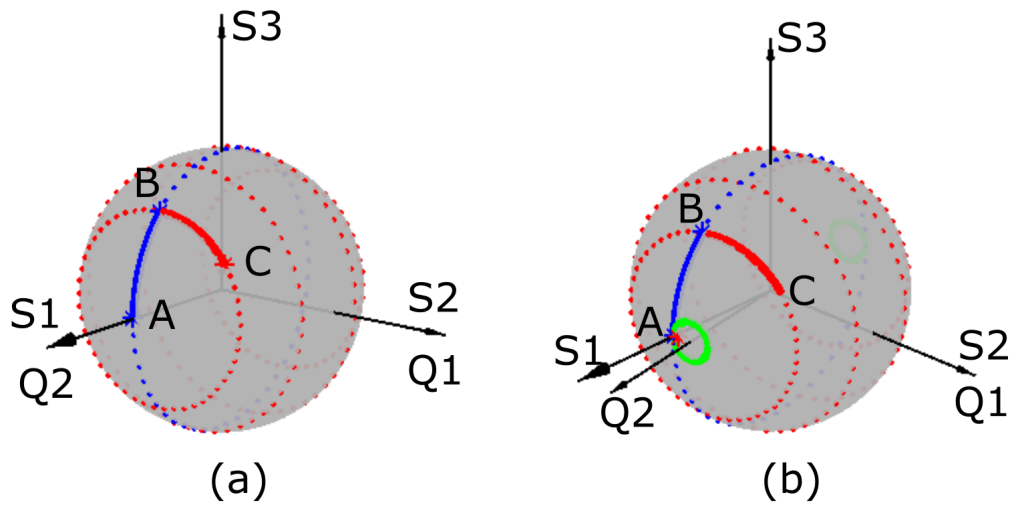


Figure 7.3: SOP transitions on the Poincaré sphere for the 2-SLM system. $S1$, $S2$ and $S3$ are the axes representing the reduced Stokes vector components. Point A represents the linear polarised input state; point B is the state after the first SLM; point C shows the output state. (a) The first SLM has extraordinary axis ($Q1$) aligned with $S2$ and the second SLM has extraordinary axis ($Q2$) aligned with $S1$. Variation of the retardance of the two SLMs permits the translation of point C to any point on the surface of the sphere. (b) The first SLM has extraordinary axis ($Q1$) aligned with $S2$ while the second SLM has extraordinary axis ($Q2$) aligned at an angle to $S1$. Therefore, $Q2$ is not orthogonal to $Q1$. Variation of the retardance of the two SLMs still permits the translation of most points on the surface of the sphere; however, the two areas enclosed by the solid green circle around $Q2$ and $-Q2$ can never be reached by this not perfectly aligned 2-SLM system.

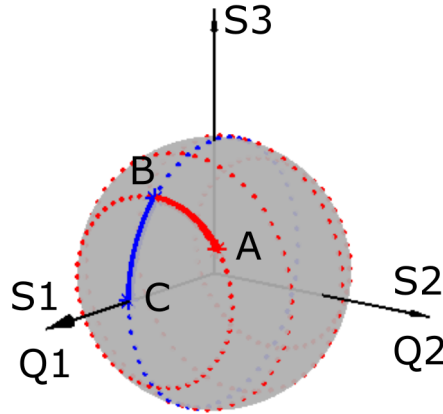


Figure 7.4: SOP transitions on the Poincaré sphere for the 2-SLM system. $S1$, $S2$ and $S3$ are the axes representing the reduced Stokes vector components. Point A represents any arbitrary input state; point B is the state after the first SLM; point C shows the output state. The first SLM has extraordinary axis ($Q1$) aligns with $S1$ and the second SLM has extraordinary axis ($Q2$) aligns with $S2$. Variation of the retardance of the two SLMs permits any point on the surface of the sphere to transmit to the fixed point C .

7.2.2 Conversion from any arbitrary to a fixed polarisation state

Similar to section 7.2.1, this reverse conversion also requires a minimum of two SLM modulations. In the optimal configuration, the extraordinary optical axis of the first SLM must be oriented along the polarisation of the output light. The second SLM is then oriented at 45° to the first SLM. On the Poincaré sphere, the action of the first SLM is to move any polarisation state to the great circle passing through the output polarisation perpendicular to the extraordinary axis of the second SLM (see the red paths in Figure 7.4); the second SLM then moves the state along the great circle to the fixed output state (see the blue path in Figure 7.4).

7.2.3 Conversions between arbitrary polarisation states

The discussions in section 7.2.1 and 7.2.2 can be used as the basis and lead to the solution of achieving a general conversion between arbitrary polarisation states. Since

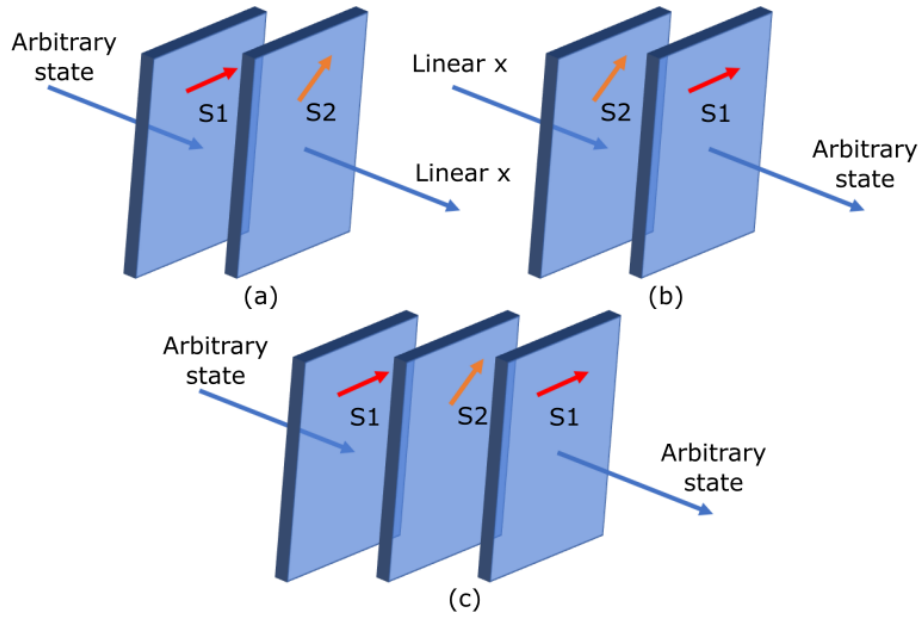


Figure 7.5: (a) A 2-SLM system to convert any arbitrary state to a linear x state. (b) A 2-SLM system to convert a linear x state to any arbitrary state. (c) A 3-SLM system to convert any arbitrary state to any other arbitrary state.

the transmission of an arbitrary state to a fixed state (linear x) can be achieved by passing through a SLM with extraordinary axis oriented at $S1$ followed by another SLM oriented at $S2$ (see Figure 7.5(a)), and the conversion of a fixed state (linear x) to any arbitrary state can be realised by a SLM oriented at $S2$ followed by another SLM oriented at $S1$ (see Figure 7.5(b)), a conversion from any arbitrary state to any other arbitrary state can be achieved through passing through these two 2-SLM systems in sequence to convert an arbitrary state to a linear x state and back to any other arbitrary state; i.e. a 4-SLM system with SLMs oriented at $S1$, $S2$, $S2$ and $S1$ in sequence. The two $S2$ oriented SLMs in the middle of this 4-SLM cascade can be combined as one SLM since we modelled the SLM to have a full working range of $\pm\pi$. Therefore, a 3-SLM system, oriented at $S1$, $S2$ and $S1$ respectively (see Figure 7.5(c)), can be used to convert any arbitrary polarisation state to any arbitrary state.

Note that for the purpose of demonstration, only one example of possible arrangements is presented here and analysed in the later parts of this thesis. In theory,

any 3-SLM system with the second SLM oriented at 45° to the first SLM and the third SLM aligns with the first SLM allows conversions between arbitrary polarisation states.

7.3 General conversions between arbitrary vectorial states

A vectorial state contains both the polarisation and phase state. By modulating the polarisation state, the phase state of light was also changed (see section 2.1.3.3) [23, 25, 26, 228]. To allow a full vectorial state modulation, we need to analyse the PB phase added to the system due to polarisation modulations. Note that based on the SLM models we adopted in the section 7.1, the dynamic phase introduced by the SLM system is a constant (Φ_δ) and thus excluded from our discussion. In this chapter, we will also show that in the cases when the PB phase cannot be modulated to cover a full range of $\pm\pi$, extra modulation devices need to be included for phase compensations.

7.3.1 The PB phase of a single SOP transition

The PB phase defines the phase difference between two beams with the same or different polarisation states subjected to the orientation of the eigenvector of the system. This phase can be calculated using a geometrical expression or through interference.

7.3.1.1 The geometrical expression of the PB phase

If the polarisation state A is in phase with -Q and -Q is also in phase with B, the phase difference Φ between A and B can be calculated as half of the area Ω of the

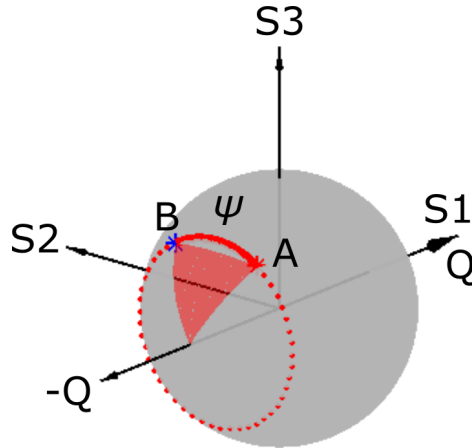


Figure 7.6: A transition from A to B after passing through a SLM with its extraordinary axis Q oriented at $S1$ and a retardance of Ψ . The phase difference between A and B after the transition can be calculated as half of the area of the geodesic triangle A-QB.

geodesic triangle A-QB [228] [Figure 7.6]

$$\Phi_{A-QB} = \frac{1}{2}\Omega_{A-QB} = \tan^{-1} \left[\frac{A \cdot (-Q \times B)}{1 + A \cdot -Q + -Q \cdot B + B \cdot A} \right] \quad (7.1)$$

where A , $-Q$ and B are normalised 3×1 reduced Stokes vectors; the vectors also indicate the three dimensional coordinates of the states on the Poincaré sphere (see section 2.1.3.4).

As discussed in section 7.1, the SLM is modeled such that the retardance is applied to the extraordinary axis while it has no effect on the ordinary axis. Therefore after the transition of the polarisation state from A to B, the input beam A is in phase with the ordinary axis $-Q$ and $-Q$ is also in phase with the output beam B. Φ_{AQB} in the equation 7.1 can therefore calculate the phase difference between the input and output beam A and B after passing through an SLM with extraordinary axis oriented at Q .

7.3.1.2 The PB phase derived through interference

If two beams, each with a different polarisation state J_A and J_B , constructively interfere with each other, the two beams will be in phase. The phase difference between two beams thus equals to the phase added to one of the beams for constructive interference. This phase difference can thus be calculated as

$$\Phi_{A-QB} = \arg(J_A \cdot J_B) \quad (7.2)$$

where J_A and J_B are Jones vectors of polarisation states A and B.

Equation 7.2 can also be written in the form of the reduced Stokes vectors [228] as

$$\Phi_{A-QB} = \arg[\mu_1 + \mu_2 + (\mu_1 - \mu_2) Q \cdot A] = \arg[e^{i\psi} + 1 + (e^{i\psi} - 1) Q \cdot A] \quad (7.3)$$

where μ_1 and μ_2 are the eigenvalues of the extraordinary and ordinary axis of the SLM respectively. In our model system described in section 7.1, $\mu_1 = \exp(i\psi)$ and $\mu_2 = 1$.

Appendix C includes justifications and derivations of how the equation 7.1, 7.2 and 7.3 lead to the same result.

7.3.2 The PB phase of the 3-SLM system

After passing through a 3-SLM system, the input beam with polarisation state A is converted to D through three transitions (A to B, B to C and C to D) (Figure 7.7 (a)).

The SLM model in section 7.1 suggested that after the conversion from A to B, A is in phase with -Q1 and -Q1 is in phase with B. The phase difference between A and B can thus be calculated as Φ_{A-Q1B} ; same for Φ_{B-Q2C} and Φ_{C-Q3D} .

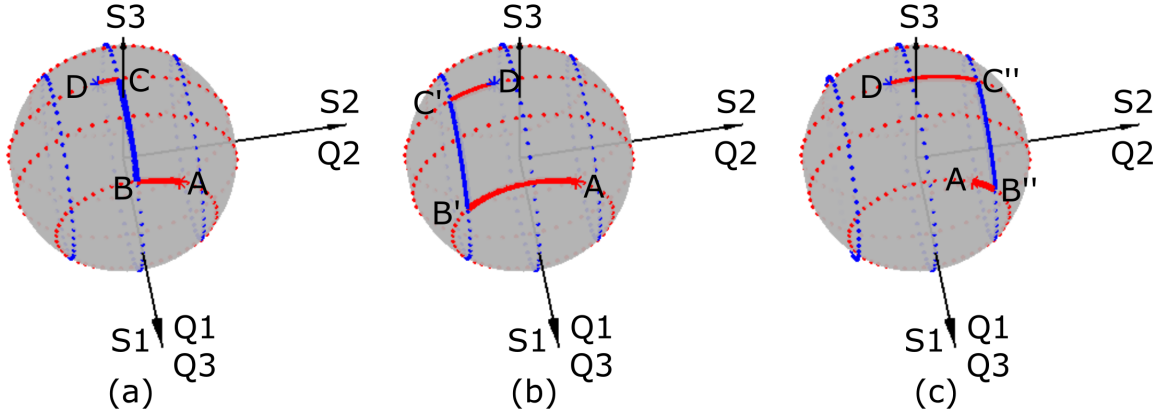


Figure 7.7: Three examples of the conversion from the polarisation state A to D after passing through three SLMs with extraordinary axis oriented at S1, S2 and S1 respectively. In each case, the three SLMs have different retardance settings and the transitions trace different paths on the Poincaré sphere. The resultant polarisation state conversions are the same but the PB phases introduced are different for the three examples.

Upon adding the phase difference Φ_{A-Q1B} to the Jones vector of A, J_A , adding Φ_{B-Q2C} to both J_A and J_B , and adding Φ_{C-Q3D} to J_A , J_B and J_C , we obtain three new vectorial states J'_A , J'_B and J'_C

$$J'_A = \exp [i(\Phi_{A-Q1B} + \Phi_{B-Q2C} + \Phi_{C-Q3D})] J_A$$

$$J'_B = \exp [i(\Phi_{B-Q2C} + \Phi_{C-Q3D})] J_B$$

$$J'_C = \exp (i\Phi_{C-Q3D}) J_C$$

such that J'_A is in phase with J'_B , J'_B is in phase with J'_C and J'_C is in phase with J_D . Using the same geometric phase definition we mentioned above in section 7.3.1, the phase difference between the new vectorial state J'_A and J_D can now be calculated as $\Phi_{ABC} + \Phi_{ACD}$.

Therefore, the total phase difference between A and D after passing through the 3-SLM system, Φ , can be calculated as

$$\Phi = \Phi_{A-Q1B} + \Phi_{B-Q2C} + \Phi_{C-Q3D} + \Phi_{ABC} + \Phi_{ACD} \quad (7.4)$$

Each individual component of this equation 7.4 can be calculated using either the equation 7.1, 7.2 or 7.3; but these calculations require the knowledge of the retardance settings of the second and third SLM and/or the Stokes vectors of the intermediate points B and C. The discussions on SLMs retardance settings and Stokes vectors of the intermediate points are presented in the next section 7.3.2.1.

7.3.2.1 The retardance settings of SLMs and the intermediate transitional states

To calculate the phase difference between A and D using the equation 7.4, the retardance settings of the second and third SLM, if equation 7.3 is used, and/or the Stokes vectors of the intermediate point B and C, if equation 7.1 is used, will need to be known. As shown in figure 7.7, the conversion from A to D is not unique and the resultant PB phase is different when different paths on the Poincaré sphere are taken. Though A and D are fixed chosen states, B and C may vary on the circles perpendicular to S1 and passing through A and D respectively. After some derivations, it is shown that the retardance settings of the second and third SLM, ψ_2 and ψ_3 , and the reduced Stokes vectors of B and C depend on A, D and the retardance

of the first SLM ψ_1 .

$$\psi_2 = \sin^{-1} \left(\frac{D_1}{\sqrt{1 - B_2^2}} \right) - \sin^{-1} \left(\frac{A_1}{\sqrt{1 - B_2^2}} \right) \quad (7.5)$$

$$\psi_3 = \sin^{-1} \left(\frac{D_2}{\sqrt{1 - D_1^2}} \right) - \sin^{-1} \left(\frac{B_2}{\sqrt{1 - D_1^2}} \right) \quad (7.6)$$

$$B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \cos \psi_1 + A_3 \sin \psi_1 \\ -A_2 \sin \psi_1 + A_3 \cos \psi_1 \end{bmatrix} \quad (7.7)$$

$$C = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} D_1 \\ D_2 \cos \psi_3 - D_3 \sin \psi_3 \\ D_2 \sin \psi_3 + D_3 \cos \psi_3 \end{bmatrix} \quad (7.8)$$

where S_n is the n^{th} entry of the Stokes vector of S . For the derivations of the above equations, see Appendix D for more details.

7.3.2.2 A discussion of the P-B phase variations for the 3-SLM system

As the equations shown in the above section 7.3.2.1, the retardance settings of the second and third SLM and the intermediate transitional states, B and C, are related only to the Stokes vector of the initial, final polarisation states, A and D, and the first SLM retardance settings, ψ_1 . The PB phase can be calculated by substituting the equations in section 7.3.2.1 into the equation 7.4; the total PB phase, Φ , is thus also a function of A, D and ψ_1 .

After some detailed analysis (see Appendix E), it is shown that the total PB phase Φ has a full range of $-\pi$ to π only if the Stokes vector of the initial and final polarisation states have an equal but opposite first element, i.e. $A_1 = -D_1$, or if both A_1 and D_1 equals to 1, i.e. $A_1 = D_1 = 1$.

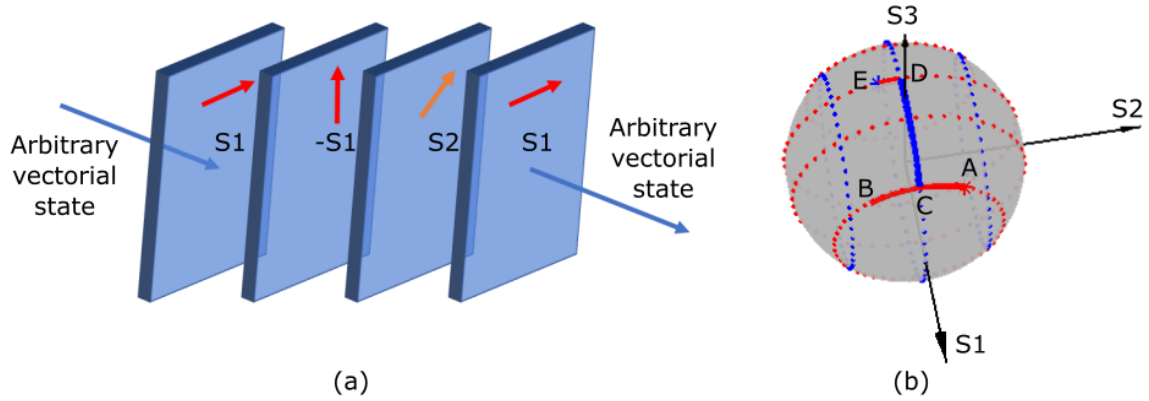


Figure 7.8: (a) A 4-SLM system proposed for a full range of vectorial state modulation. (b) An example of the conversion from the polarisation state A to E after passing through the 4-SLM system shown in (a).

7.3.3 The PB phase of a 4-SLM system to allow a full vectorial state modulation

As discussed in the section 7.3.2, the PB phase variations of a 3-SLM system only has a full range of $\pm\pi$ in limited cases. To ensure a full range of vectorial state modulation, an extra AO device needs to be included into the system.

One configuration that allows sufficient phase control would include an additional polarisation insensitive device, such as a DM or mirror-based SLM, which would introduce a polarisation independent phase shift. This device could be placed at any point in the optical system, before, between or after the 3-SLM passes.

An alternative way is using an additional SLM pass with its extraordinary axis aligned with -Q1 to achieve a similar goal. An advantage of using an extra SLM pass rather than including an additional polarisation insensitive AO device is to avoid a cumbersome optical design; the extra SLM pass can be implemented together with the other three SLM passes on different areas of the same device [155,214,216].

The sequence of the 4-SLM orientations would be S1, -S1, S2, S1, where the additional pass at -S1 is inserted in the second position (see Figure 7.8 (a)).

Considering the action of the first two SLM passes, we see that the locus of the

possible polarisation state outputs from each pass follows the same circle on the Poincaré sphere (see Figure 7.8 (b)). Hence, the accessible polarisation states are the same as would be possible for a single pass with SLM orientation at S1. However, the phases introduced in these two passes can vary with a full range of 2π by selecting different retardance settings of the first SLM. For a detailed analysis, please see Appendix F.

Note that this particular arrangement of SLMs is not unique and the same performance can be achieved with certain other angular arrangements of SLMs. One obvious configuration would be S1, S2, S1 and -S1 orientations.

7.4 Compensation of the phase and polarisation state errors using a multi-SLM system

In the previous sections of the chapter, we discussed of a conversion between any two arbitrary vectorial states of light. We showed that a minimum of four SLMs are required to achieve this arbitrary conversion. A more practical application, which is to correct the polarisation state and phase errors, however, is a slightly different question that needs to be tackled. Errors are introduced when light propagates through a medium with polarisation properties; the corrections of the errors should compensate the effects of the medium regardless of the vectorial states of the light. While in the previous discussions, the retardance settings of the four AO devices are dependent on the initial and final vectorial states of light; we need a more thorough configuration to correct the errors depends only on the properties of the medium.

One way to achieve such an error compensation is using Jones calculus. For non-diattenuating and non-depolarising homogeneous systems, the vectorial aberrations can be described by spatially variant retarders in the form of Jones matrices. A sequence of SLMs can be designed and serve as an inverse to the Jones matrix of the

aberrated system, and thus compensate the effect of the propagating medium.

To show that this new configuration is independent to the polarisation states of the light, we will demonstrate the system on a full Poincaré beam, which contains all possible polarisation states on the Poincaré sphere. The aberrations introduced by the medium on any polarisation state can therefore be corrected by the multi-SLM system proposed in this section.

7.4.1 The Jones matrix of any arbitrary homogeneous retarder for compensation

A homogeneous retarder has its fast axis orthogonal to its slow axis. Since light with the polarisation state aligned with the fast or the slow axis of the retarder will only have its phase but not polarisation state modulated, the Jones vectors of the fast and slow axis are the two eigenvectors of the Jones matrix of the homogeneous retarder. The retardance settings are thus the eigenvalues of the Jones matrix. A general Jones matrix expression for an arbitrary homogeneous retarder, $\mathcal{J}$, can be expressed as

$$\mathcal{J} = \begin{bmatrix} \lambda_1 |q_x|^2 + \lambda_2 |q_y|^2 & (\lambda_1 - \lambda_2) q_x q_y^* \\ (\lambda_1 - \lambda_2) q_x^* q_y & \lambda_2 |q_x|^2 + \lambda_1 |q_y|^2 \end{bmatrix}$$

where λ_1 and λ_2 are the eigenvalues of the matrix in the form of $\exp(i\psi_x)$ and $\exp(i\psi_y)$ where ψ_x and ψ_y are the retardance applied along the extraordinary and ordinary axes respectively; q_x and q_y are the components of the extraordinary and ordinary axis (q_1 and q_2) of the retarder such that

$$q_1 = \begin{pmatrix} q_x \\ q_y \end{pmatrix} \quad q_2 = \begin{pmatrix} -q_y^* \\ q_x^* \end{pmatrix}$$

$\mathcal{J}$ can also be written as

$$\mathcal{J} = \sqrt{\lambda_1 \lambda_2} \begin{bmatrix} \lambda |q_x|^2 + \lambda^* |q_y|^2 & (\lambda - \lambda^*) q_x q_y^* \\ (\lambda - \lambda^*) q_x^* q_y & \lambda^* |q_x|^2 + \lambda |q_y|^2 \end{bmatrix} = \sqrt{\det(\mathcal{J})} \mathcal{J}' \quad (7.9)$$

where $\sqrt{\lambda_1 \lambda_2} = \sqrt{\det(\mathcal{J})}$, $\lambda = \sqrt{\lambda_1/\lambda_2}$ and $\lambda^* = \sqrt{\lambda_2/\lambda_1}$. $\arg[\sqrt{\det(\mathcal{J})}]$ equals to the dynamic phase of the retarder. $\mathcal{J}'$ is the Jones matrix without the dynamic phase component, $\mathcal{J}' = \mathcal{J}/\sqrt{\det(\mathcal{J})}$.

For the expression 7.9 see the Appendix G.1 for more detailed derivations.

7.4.2 Compensations of the polarisation state errors using a 3-SLM system

A SLM is an adaptable homogeneous retarder with fixed axes orientations but a variable extraordinary axis retardance (see the section 7.1 for the model of SLM used in this thesis). A sequence of a few homogeneous retarders result a collective Jones matrix which is also homogeneous (see Appendix G.2 and G.3 for more details). This allows a multi-SLM system to compensate the effects of a homogeneous retarder.

From the analysis in the previous section 7.2, a 3-SLM system with extraordinary axes oriented at S1, S2 and S1 can fully modulate the polarisation state (but not the phase). If the 3-SLM system can have an equivalent Jones matrix such that $\mathcal{J}'_{3slm} = \mathcal{J}'^{-1}$, the 3-SLM system can compensate the effects of the homogeneous retarder. For a retarder with Jones matrix stated in equation 7.9, the Jones matrix of the 3-SLM system needs to be

$$\mathcal{J}'_{3slm} = \mathcal{J}'^{-1} = \begin{bmatrix} \lambda^* |q_x|^2 + \lambda |q_y|^2 & (\lambda^* - \lambda) q_x q_y^* \\ (\lambda^* - \lambda) q_x^* q_y & \lambda |q_x|^2 + \lambda^* |q_y|^2 \end{bmatrix} \quad (7.10)$$

to fully compensate the polarisation modulation introduced by the retarder.

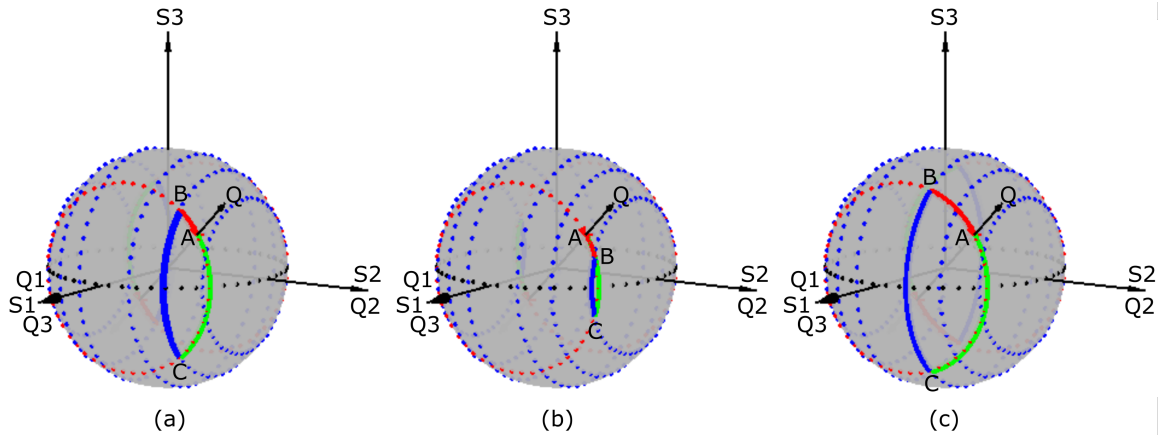


Figure 7.9: Light with polarisation state A is converted to B after passing through the first SLM with the extraordinary axis Q1 aligned with S1 and a retardance of ψ_1 (the red path). Its polarisation state is converted from B to C after passing through the second SLM with the axis Q2 oriented at S2 a retardance of ψ_2 (the blue path). The polarisation state is then converted back to A after passing through the third SLM with axis Q3 aligned with S1 and a retardance of ψ_3 (the green path). The resultant extraordinary axis of the 3-SLM system (or the eigenvector of the 3-SLM Jones matrix), Q, aligned with A. (a), (b) and (c) are three arrangements having the same eigenvectors (Q) but different eigenvalues.

From the above equation H.1, the sequence of the three SLMs needed to have two effective eigenvectors with the same orientations as the eigenvectors of the retarder; the two eigenvalues of the three SLMs need to be conjugate to the eigenvalues of the retarder.

7.4.2.1 The configuration of a 3-SLM system with eigenvectors same as the retarder

One configuration to create a 3-SLM system with eigenvectors the same as the retarder for compensation is shown in the figure 7.9 (a). A beam with polarisation state A returns back to A after passing through the 3-SLM system which suggests the three SLMs has a resultant eigenvector, Q, aligns with A.

Suppose the reduced Stokes vector of A (or Q) is defined as

$$A = Q = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \end{pmatrix} = \begin{bmatrix} |q_x|^2 - |q_y|^2 \\ 2 \operatorname{Re}(q_x^* q_y) \\ 2 \operatorname{Im}(q_x^* q_y) \end{bmatrix}$$

By geometry, it can be shown that if the retardance of the first and third SLM, ψ_1 and ψ_3 , has a relationship of

$$\frac{\psi_3 - \psi_1}{2} = \tan^{-1} \left(\frac{A_3}{A_2} \right) \quad (7.11)$$

and the second slm has a retardance of

$$\frac{\psi_2}{2} = \tan^{-1} \left(\frac{\sqrt{1 - A_1^2} \sin(\frac{\psi_1 + \psi_3}{2})}{A_1} \right) \quad (7.12)$$

this 3-SLM system can have resultant eigenvectors the same as the retarder for compensation.

7.4.2.2 The configuration of a 3-SLM system with eigenvalues conjugate to that of the retarder

As shown in the figure 7.9, multiple SLM configurations can achieve the same resultant eigenvectors. The differences among these configurations are their resultant eigenvalues.

After some analysis, it can be shown that to have eigenvalues conjugate to that

of the retarder, the three SLMs needed to have retardance settings expressed as

$$\psi_1 = \tan^{-1}\{A_1 \tan[\arg(\lambda^*)]\} - \tan^{-1}\left(\frac{A_3}{A_2}\right) \quad (7.13)$$

$$\psi_3 = \tan^{-1}\{A_1 \tan[\arg(\lambda^*)]\} + \tan^{-1}\left(\frac{A_3}{A_2}\right) \quad (7.14)$$

$$\psi_2 = 2 \sin^{-1}\{\sqrt{1 - A_1^2} \sin[\arg(\lambda^*)]\} \quad (7.15)$$

where λ and λ^* are the eigenvalues of the Jones matrix of the retarder, J' , defined in the equation 7.9. For the derivations of the equation 7.13, 7.14 and 7.15, please see Appendix H for more details.

7.4.3 Compensations of both the polarisation state and phase errors using a 4-SLM system

As discussed in the section 7.3.3, an arrangement of four SLMs with extraordinary axis oriented at S1, -S1, S2, S1 allows a full vectorial state modulation, including both the polarisation state and phase.

The first two SLM passes are along the same circular path on the Poincaré sphere but towards opposite directions since the extraordinary axes of the two SLM passes are opposite. It is easy to write down the retardance settings of the four SLMs as

$$\psi_1 - \psi_2 = \tan^{-1}\{A_1 \tan[\arg(\lambda^*)]\} - \tan^{-1}\left(\frac{A_3}{A_2}\right) \quad (7.16)$$

$$\psi_4 = \tan^{-1}\{A_1 \tan[\arg(\lambda^*)]\} + \tan^{-1}\left(\frac{A_3}{A_2}\right) \quad (7.17)$$

$$\psi_3 = 2 \sin^{-1}\{\sqrt{1 - A_1^2} \sin[\arg(\lambda^*)]\} \quad (7.18)$$

We can therefore see that the extra SLM pass allows one degree of freedom in the system for phase modulations.

Both the arbitrary retarder for compensation and the 4-SLM system for correction

have the same extraordinary and ordinary axes orientations; their retardance settings are equal and opposite to each other. We can therefore write down

$$\frac{\mathcal{J}_{4slm}}{\sqrt{\det(\mathcal{J}_{4slm})}} = \mathcal{J}'_{4slm} = \mathcal{J}'^{-1} = \frac{\mathcal{J}^{-1}}{\sqrt{\det(\mathcal{J}^{-1})}} = \mathcal{J}^{-1} \sqrt{\det(\mathcal{J})}$$

Using the definitions of the PB phase in the section 7.3.1.1, for any input polarisation state, the PB phase after passing through a retarder with Jones matrix of $\mathcal{J}'$ is equal and opposite to the PB phase after passing through a 4-SLM system with Jones matrix of $\mathcal{J}'_{4slm}$ since $\mathcal{J}'_{4slm} = \mathcal{J}'^{-1}$ and $\mathcal{J}' \mathcal{J}'_{4slm} = I$ where I is a 2×2 identity matrix.

To make $\mathcal{J}_{4slm} = \mathcal{J}^{-1}$, we thus need $\sqrt{\det(\mathcal{J}_{4slm})} = 1/\sqrt{\det(\mathcal{J})}$. For the derivations of $\det(\mathcal{J}_{4slm})$, please see Appendix G.2. Hence,

$$-(\psi_1 + \psi_2 + \psi_3 + \psi_4) = \arg(\lambda_1 \lambda_2) \quad (7.19)$$

We have thus derived the retardance settings of the four SLMs to allow compensations of full vectorial state errors introduced by any arbitrary homogeneous retarder.

7.4.4 Corrections of the vectorial state aberrations using a 4-SLM system on a full Poincaré beam

The 4-SLM configuration discussed above is demonstrated on a full Poincaré beam in simulation (see Figure 7.10). It is shown that the 4-SLM system can successfully correct the aberration introduced by an arbitrary retarder.

7.5 Conclusion

In this chapter, we proposed a 3-SLM system for an arbitrary to arbitrary polarisation state transmission and a 4-SLM system for a full vectorial state (including both

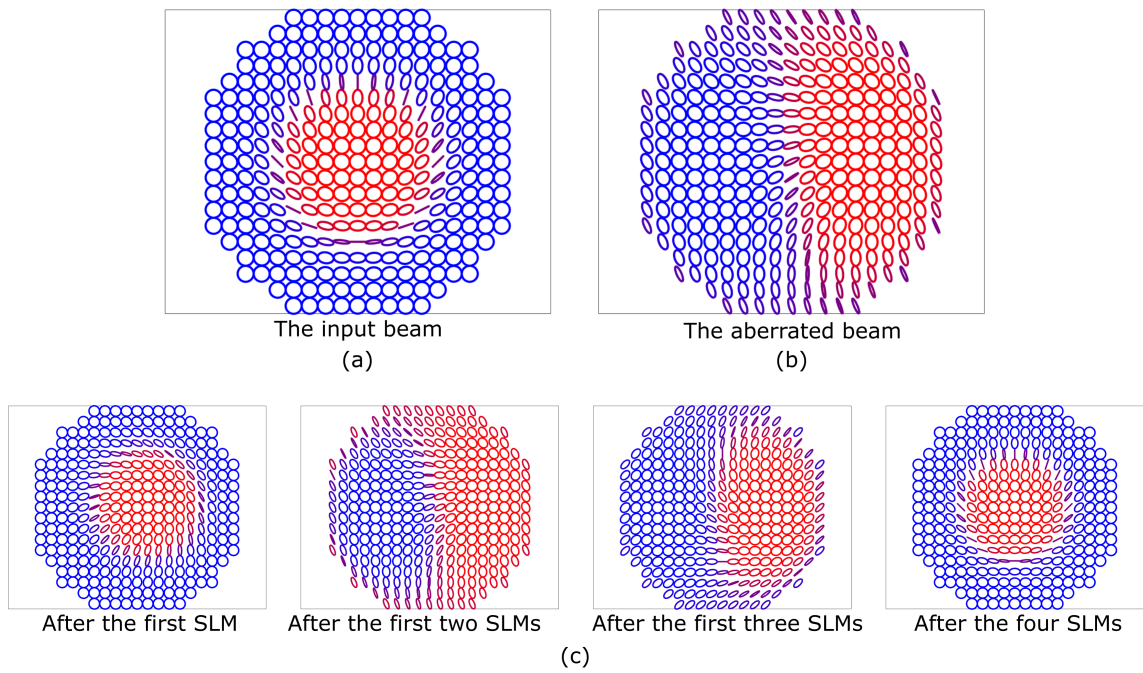


Figure 7.10: Red marks indicate right hand circular polarisation while blue marks indicate left hand circular polarisation. (a) is the input full Poincaré beam. (b) is the beam after passing through a randomly generated arbitrary waveplate with a Jones matrix of $\mathcal{J} = \begin{pmatrix} 0.3540e^{-j1.3916} & 0.9353e^{-j0.0157} \\ 0.9353e^{j1.5551} & 0.3540e^{-i0.2106} \end{pmatrix}$. The beam is aberrated by the waveplate. (c) is the beam after passing through each SLM of the 4-SLM system showing that the procedure does indeed compensate for the polarisation aberration.

the polarisation and phase state) conversion. We also showed that a 4-SLM configuration can correct both polarisation state and phase aberration introduced by an arbitrary retarder regardless the vectorial state of the light. We wanted to emphasise that although both discussions proposed very similar multi-SLM configurations, their applications are different as the first discussion focused on light vectorial state conversions while the second discussion focused on compensating the effect of polarising elements (retarders in this case). We also demonstrated an example on a simulated full Poincaré beam where the proposed 4-SLM system correctly compensated the aberrations introduced by an randomly generated retarder. This combination of adaptive elements could form the basis of an optically efficient correction device for vectorial aberrations in an AO scheme.

In practice, the performance of SLM based compensation systems would be affected by limitations of practical liquid crystal devices, such as pixellation, fill-factor and reflection efficiency of the SLMs. However, within these limitations, the configuration modelled here has the potential to operate with high throughput efficiency. The practicality of this is backed up by demonstrations in the existing literature that have used multiple SLMs for beam modulation [155,215,216]. Implementation of such a system with reflective SLMs is feasible, but cumbersome. Future implementations might use a stack of simpler transmission mode modulators tailored, for example, to low-order vectorial aberrations encountered in the system. The models derived in this paper could also be applied to these systems.

Chapter 8

Conclusions and outlooks

8.1 Conclusions

In this thesis we discussed various implementations of different AO devices to correct both phase and polarisation state aberrations in microscopy systems to improve imaging qualities.

We proposed an automated control scheme for a motorised objective lens correction collar incorporating a axially movable sample mounting stage to compensate both spherical aberration and defocus simultaneously. Objective lens correction collar is especially useful in optical systems when conjugate pupil planes are not accessible and no other adaptive elements (such as DM and SLM) is available. Objective correction collars are also useful to work in conjunction with a continuous membrane DM since a DM has limited actuator strokes, limited number of actuators and is unable to display wrapped phase; by correcting spherical aberrations using a correction collar, the correction power of a DM can be dedicated to correct other modes and thus improves the AO system performance as a whole. A correction collar to compensate spherical aberrations are through adjusting the central lens group position which not just applies spherical aberration compensations but also introduces extra defocus.

Conventionally, the operations of a correction collar with an axially movable sample stage are not easy which require careful controls and adjustments from experienced microscopy users; this could often result prolonged exposures to the sample which increase the chances of photobleaching and phototoxicity. Unclear markings on the correction collar and immersion fluid mismatches also lead to inaccurate operations. The automated scheme proposed in this thesis was thus demonstrated to resolve these problems and greatly ease the use of the optics. Making use of the linear relationship between the correction collar settings and axial focal plane positions, the correction scheme required a minimum of $2n + 1$ measurements (if n^{th} order polynomial fitting is used and $n \geq 2$) which is more efficient than normal operations by hand. Comparing to other correction collar operations in literature which were based on an intensity metric calculated from a three-dimensional volume, this proposed operation was based on a Fourier frequency metric calculated from a two-dimensional image which was thus less demanding (as some systems were unable to generate three-dimensional volume images) and more generally adaptable (as in some systems such as a 1-P widefield microscope the average intensity does not reflect the imaging quality).

We proposed a framework, which consists of aberration representations, system quality metric and optimisation algorithm, to incorporate different sensorless AO methods to be considered under the same phase aberration compensation mechanism. We suggested mathematical analysis on the sensorless aberration correction process which was formulated as a system quality metric optimisation process with physical models of imaging systems encapsulated. By doing so, we provided mathematical interpretations on various method designs such as a comparison between orthogonal and non-orthogonal aberration representations, how orthogonality can affect the choice of the optimisation algorithm and a discussion on how the choice of the system quality metric can affect the definition of the orthogonality as well as how the choice of the quality metric affects the choice of the optimisation algorithm.

The mathematical analysis presented offered a comprehensive overview of sensorless AO schemes. Since all the sensorless AO methods in literature were developed based upon fundamentally different concepts for different applications in different optical systems, there was no systematic fair comparison among different methods before our work. This framework allowed us to view different approaches as variations of a unified aberration correction mechanism and allowed fair and systematic comparisons among them to be possible. Inspirations drawn from this framework can also help in new sensorless AO method designs.

The framework permitted for the first time a systematic fair comparison of various aberration representations used in different sensorless AO approaches. Simulated and experimental results on fluorescence beads and biological samples showed that zonal representations can correct aberrations with high frequency components (scattering) better but are more susceptible to noise while modal representations can correct aberrations with gradual shapes better and are more robust against noise. This can be explained as aberrations with gradual shapes have high spatial auto-correlations and thus modal representations, which also have high spatial correlations, are better at representing and correcting these modes. On the other hand, aberrations with complex shapes have low spatial auto-correlations and thus are better corrected by zonal representations which are discrete in nature and also have low spatial auto-correlations. Furthermore, since modal representations are expansions over the whole pupil, they have high modulation depths of which signal variations with phase modulations are large. Modulation signals are thus less likely to be lost in noise comparing to the modulation signals of zonal representations where phase is defined over a small portion of the pupil and signal variations are small and easily lost in noise (see Appendix A for more details). We want to point out that the conclusions drawn from the simulations and experiments should be generalised and readily extended to other aberration representation bases such as DM deformation modes, Hadamard-Walsh

modes or other arrangements of pupil zones, since we demonstrated phenomena using both Legendre and Zernike expansions in both square and circular pupils, respectively, to correct both Fourier based introduced aberration and sample-generated aberrations; therefore the essential observations should be decoupled from the chosen aberration representations. Our analysis also demonstrated a broader spectrum of zonal/modal hybrid methods beyond those presented in literature and can be useful in specifications of future AO schemes.

We proposed and demonstrated a hand-crafted untrained NN on simulated data to show that it can correctly identify the pupil wavefront with two or three images when a total of eight modes were included in the analysis. This demonstration showed the potential of ML based aberration correction methods to extract more PSF related information from images than conventional sensorless AO methods which required a minimum of $N + 1$ exposures when analysing N number of modes. Requiring no training dataset or pre-training process, this network is much more easily implemented than normal NNs which require a large training datasets and a long training process before functioning. The proposed NN was also successfully demonstrated on both a simulated 2-P and 1-P widefield system which showed its high adaptability; the NN can be easily adjusted to work on a different system by simply change the physical model used in the network, while other NNs presented in literature required a whole new training process before applying to a separate system. This NN also had a potential to converge at a much faster speed than other untrained phase retrieval NNs presented in literature as our network contained a number of parameters comparative to the number of modes rather than the number of image pixels; a smaller number of parameters lead to a smaller network size and thus a smaller processing time.

We provided a first time complete discussion on full vectorial state modulation (including both phase and polarisation state) by using a multi-SLM configuration which allowed light with any arbitrary vectorial state to be converted to any other arbitrary

state. We also proposed a multi-SLM arrangement which could compensate errors introduced by any arbitrary waveplate through mapping the eigenstate of the multi-SLM system the same as that of the polarising element while generating eigenvalues conjugate to that of the polarising element for compensation. There is an increasing demand for full vectorial state modulation or vectorial state aberration compensation in various imaging applications like GRIN lens fibre imaging and super-resolution systems such as STED and SIM; yet no demonstration on full vectorial state error compensation has been presented in literature. Analysis shown in this thesis could form a key step to achieve an AO scheme to realise vectorial aberration correction in imaging systems in the future.

8.2 Outlook

The novel NN based sensorless AO method has a lot of potential for future development especially when a lot of researchers are currently dedicated to developing and creating various new NN frameworks and tools in this fast evolving and powerful technology. Currently positive results collected from working on simulated data suggested a lot of promise of the network but its capability can only be tested and explored if the method is demonstrated in real imaging systems on experimental data. A following up project on examining and optimising this NN based method on a 2-P microscope will be carried out in the near future.

Most NN based phase retrieval methods in literature used Zernike polynomials as the aberration representation basis. Zernike polynomials are a convenient choice since they are derived from classical optical imaging and are well-defined over a circular pupil. Although Zernike polynomials were shown to be a very common and sensible choice in conventional aberration correction methods, it remained questionable whether the same representations were optimised for NN based methods; in fact,

results shown in this thesis suggested that the NN proposed here had a unstable convergence speed upon using Zernike polynomials. An analysis on deriving more suitable aberration representation sets can be very beneficial for a lot of NN based phase retrieval approaches and may lead to more consistent and promising performances of the NNs.

Currently, most ML based methods were limited to post data processing; humans select, (possibly label) and feed data into the network and let machine to learn in a supervised way. To fully harness the advantage of ML, it may also be used in data gathering process. We have seen works from Horstmeyer *et. al* [229] where illumination models were advised by a trained CNN such that task-specific LED patterns were displayed to allow a microscope to capture images containing useful information in ways that can be more easily extracted by computers. A similar concept may be applied to NN based aberration correction methods where the applied bias modes may be advised by the network to assist phase retrieval.

Full vectorial state aberration compensation remained a difficult task especially in a sensorless manner. Comparing to phase aberration correction, vectorial state errors contain more dimensions and thus are more difficult to be corrected. NNs can be a possible candidate for solving this complex problem. An end-to-end style NN may be trained to approach this problem in a statistical way; alternatively, a human-guided style NN may also be used incorporating the vectorial analysis of the high NA lens model (as described in Section 2.1.4.3). This model can include both phase property and polarisation state of light into analysis and may have the potential to extract the vectorial state of light with a few sample exposures.

We would also like to develop a sharing toolkit for this NN based AO method in the future so that other researchers can also try out and take advantage of this powerful computational tool. Google Colab is a Python development environment that runs in the browser using Google Cloud and offers free access to their GPU usage; this can

be a great platform for us to disseminate our network since the hardware requirement on the user side can be minimised. This toolkit can be very beneficial for our research community as it not just welcomes more researchers into this field but also allows us to gather useful feedback and collecting more microscope image data from other users to upgrade the toolkit.

Appendix A

A discussion on the piecewise modal, piecewise constant and continuous modal schemes

A.1 Similarities between piecewise modal and piecewise constant schemes

As discussed in the chapter 4, piecewise modals such as Hadamard-Walsh modes (φ_H) and polar Walsh modes are binary piecewise expansions defined across the whole pupil [see Figure A.1 (a)] while piecewise constant schemes such as pixellated piston based representations (φ_P) consists of a series pixel arrays [see Figure A.1 (b)].

It thus can be easily shown that the Hadamard-Walsh modes form exactly the same aberration representation space as the pixellated piston based representations where any aberration being represented by Hadamard-Walsh modes can also be represented by pixellated piston based sets. An easy demonstration is that all the Hadamard-Walsh modes can be represented by a combination of pixellated piston

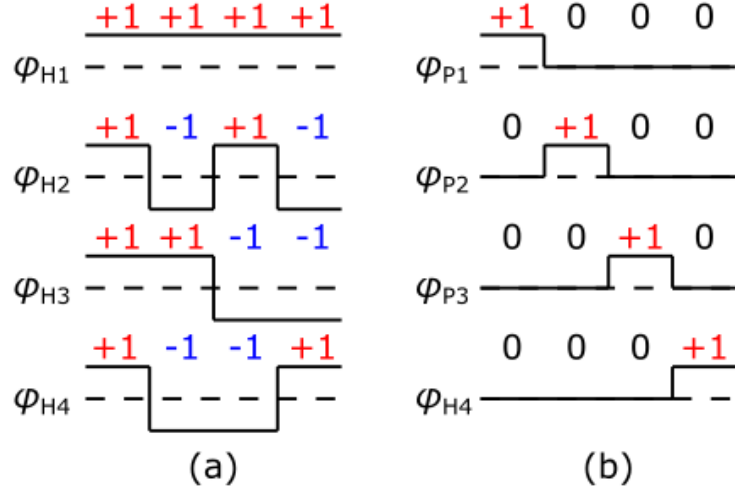


Figure A.1: (a) A set of four one dimensional Hadamard-Walsh modes φ_{Hi} , $i \in \{1, 2, 3, 4\}$; (b) A set of four one dimensional pixellated piston based representations φ_{Pi} , $i \in \{1, 2, 3, 4\}$

based representations with the same degrees of freedom

$$\varphi_{H1} = \varphi_{P1} + \varphi_{P2} + \varphi_{P3} + \varphi_{P4}$$

$$\varphi_{H2} = \varphi_{P1} - \varphi_{P2} + \varphi_{P3} - \varphi_{P4}$$

$$\varphi_{H3} = \varphi_{P1} + \varphi_{P2} - \varphi_{P3} - \varphi_{P4}$$

$$\varphi_{H4} = \varphi_{P1} - \varphi_{P2} - \varphi_{P3} + \varphi_{P4}$$

and vice versa.

$$\varphi_{P1} = \frac{1}{4}(\varphi_{H1} + \varphi_{H2} + \varphi_{H3} + \varphi_{H4})$$

$$\varphi_{P2} = \frac{1}{4}(\varphi_{H1} - \varphi_{H2} + \varphi_{H3} - \varphi_{H4})$$

$$\varphi_{P3} = \frac{1}{4}(\varphi_{H1} + \varphi_{H2} - \varphi_{H3} - \varphi_{H4})$$

$$\varphi_{P4} = \frac{1}{4}(\varphi_{H1} - \varphi_{H2} - \varphi_{H3} + \varphi_{H4})$$

Therefore, in an ideal noise free system, after one round of aberration corrections using Hadamard-Walsh modes would provide the same correction as the pixellated

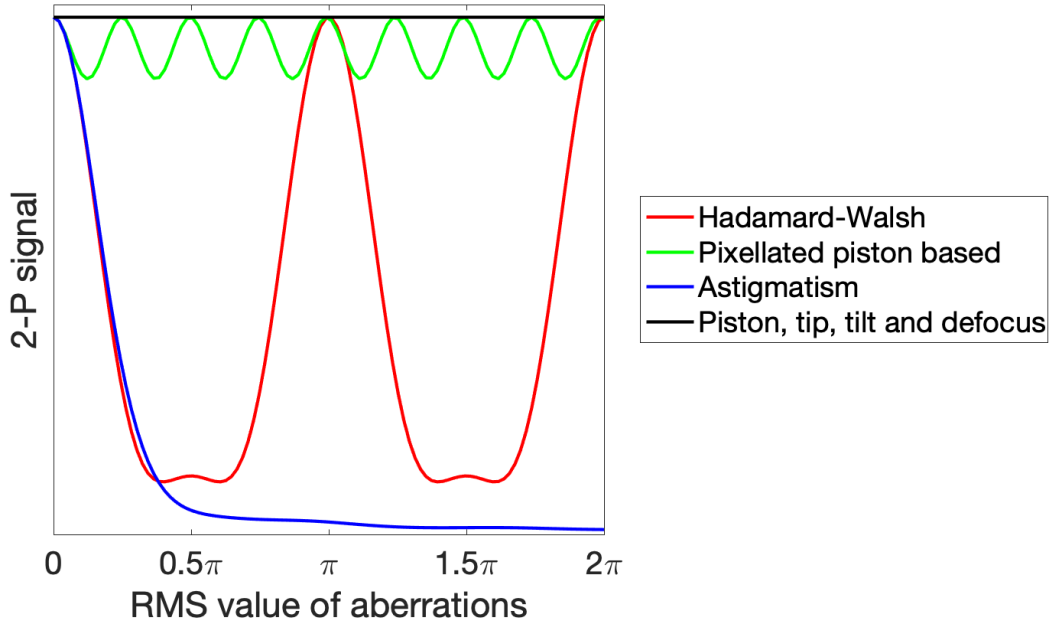


Figure A.2: 2-P signal was calculated when varying the RMS amplitude of a Hadamard-Walsh mode (over an 8×8 square pupil), a pixellated piston based mode (over an 8×8 square pupil), astigmatism (over a circular pupil enclosed within the 8×8 square pupil) and an aberration consisting piston, tip, tilt and defocus (over a circular pupil enclosed within the 8×8 square pupil).

piston based representations with the same degrees of freedom.

A.2 Differences between piecewise modal and piecewise constant schemes

However, similar to continuous modal modes, Hadamard-Walsh modes are expansions over the whole pupil and thus the modulation depth is larger comparing to the pixellated piston based representations where each component is defined only over one element zone of the whole pupil. This can be shown if we calculate the 2-P signal variations using the equation 5.1 and 5.2 when one component of each set is modulated. (see Figure A.2). With a larger modulation depth, Hadamard-Walsh modes are more robust against noise than pixellated piston based schemes of which

the signal modulation can be easily lost in noise.

A.3 Differences between piecewise and continuous schemes

Another interesting discussion is that comparing continuous modes (such as the astigmatism mode shown in the Figure A.2), both Hadamard-Walsh modes and pixellated piston based schemes modulate 2-P signals periodically. This is due to their discrete property and 2π phase wrapping. Therefore, for piecewise schemes, they have multiple global maxima where they can converge to and thus reduce the optimisation searching space when these kinds of piecewise schemes are used for aberration corrections. On the other hand, when continuous modes are used for aberration corrections, there is only one global maximum where an optimisation algorithm should converge to and thus the searching space is infinite.

A.4 Orthogonality defined for a 2-P signal metric

One final interesting discussion is on the orthogonality of aberration representations that should be defined for a 2-P signal metric. In the chapter 4, we discussed and concluded that the definitions of the inner product (which defined the orthogonality of aberration representations) were affected by the choice of the system quality metric. A suitable inner product should be defined such that the quality metric varied isotropically with any mode.

From the Figure A.2, it can be easily shown that for a small RMS amplitude, the 2-P signal decreases isotropically regardless of whether Hadamard-Walsh, pixellated piston or astigmatism is used. The inner product for the 2-P signal quality metric defined in the equation 5.1 and 5.2 can thus be chosen as the RMS value of an

aberration after removing piston, tip, tilt and defocus components.

Appendix B

Auto-correlation of various aberration representations

Here we discussed the auto-correlations of gradual and complex Fourier based input aberrations as well as two aberration representations, Z64M1 and Z1M64 used in chapter 5. By doing so, we attempted to explain why large sized zone methods are in general better at correcting simple shaped aberrations while small sized zone methods are better at correcting complex shaped aberrations. Auto-correlation ($R(\mathbf{s})$) measures the similarity of a pattern ($\Psi(\mathbf{x})$) with itself as a function of a spatial shift ($\mathbf{s}$) between them. Mathematically, it can be expressed as

$$R(\mathbf{s}) = \iint_{-\infty}^{\infty} \Psi(\mathbf{x})\Psi(\mathbf{x} + \mathbf{s}) d\mathbf{x} \quad (\text{B.1})$$

Figure B.1 shows the variations of R when $s_y = 0$ and s_x varies in a small range close to 0. R is the highest when $\mathbf{s} = (s_x, s_y) = (0, 0)$ that is when there is no spatial shift. $\Psi(\mathbf{x})$ with a gradual shape (containing small frequency orders w) has R decreases slowly as $\mathbf{s}$ increases while a pattern with quickly varying changes (containing large frequency orders w) has R drops sharply as $\mathbf{s}$ increases. This is also verified by the blue plot and red plot in the Figure B.1 where the blue plot calculated from a gradual

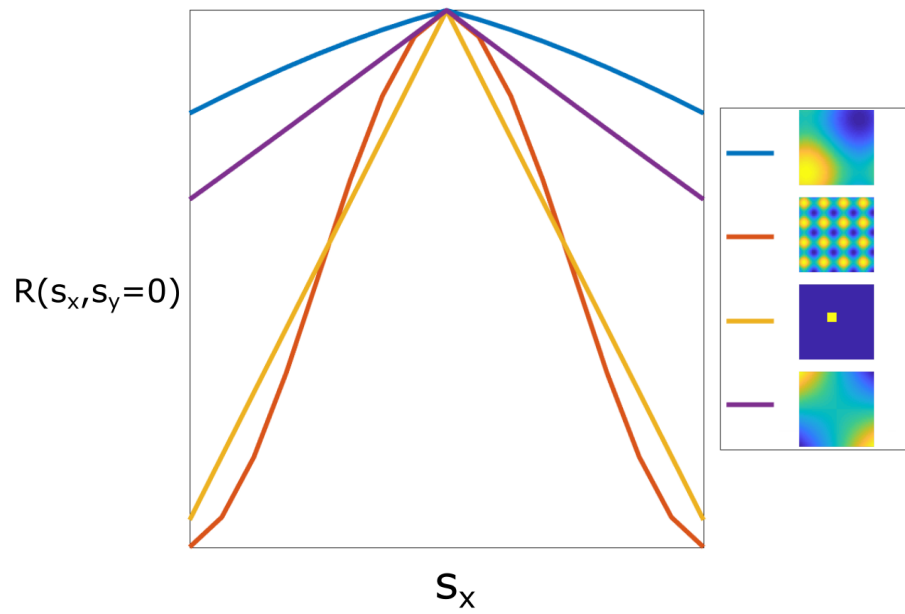


Figure B.1: The plots of $R(s_x, s_y = 0)$ against s_x when s_x varies in a small range close to 0. The blue plot is generated for an example of an aberration with a gradual shape (small frequency order w). The red plot is generated for an example of an aberration with a complex shape (large frequency order w). The yellow plot is generated for an example of the pixellated piston set Z64M1. The purple plot is generated for an example of the modal set Z1M64.

shape has R decreases slower than the red plot which is calculated from a shape with high frequency orders. Comparing the yellow plot (one example from the Z64M1 pixellated piston set) and the purple plot (one example from the Z1M64 modal set), it is shown that the R decreases faster for aberration representation sets with a high number of zones comparing to the modal sets.

Appendix C

Derivations of different expressions for the PB phase

C.1 The geometrical and interference expression of PB phase

This analysis follows the SLM model presented in section 7.1. A beam with polarisation state A passes through the first SLM with extraordinary axis oriented at Q and retardance of ψ , becoming polarization state B. A is thus in phase with -Q and -Q is in phase with B. Using equation 7.1 in the chapter 7, the phase difference between A and B can be calculated as

$$\frac{1}{2} \Omega_{A-QB} = \tan^{-1} \left[\frac{A \cdot (-Q \times B)}{1 + A \cdot -Q + -Q \cdot B + B \cdot A} \right] \quad (\text{C.1})$$

Let us use an SLM with Q aligned with S1 as an example. The ordinary axis -Q,

polarization A and B is then given as

$$-Q = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \quad A = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} \quad B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}$$

$$\frac{1}{2} \Omega_{A-QB} = \tan^{-1} \frac{A_3 B_2 - A_2 B_3}{1 - A_1 - B_1 + A_1 B_1 + A_2 B_2 + A_3 B_3} \quad (\text{C.2})$$

We now define the beam polarization state A in Jones vector (see Equation 2.5), J_A , as

$$J_A = \begin{bmatrix} a_x \\ a_y \end{bmatrix} = \begin{bmatrix} |a_x| e^{i\theta} \\ |a_y| \end{bmatrix}$$

where $\theta = \arg(a_x/a_y)$

Since A converts to B after passing through a SLM oriented at S1, a retardance of ψ is applied to the x component. The polarization state B in Jones vector (J_B) is

$$J_B = \begin{bmatrix} b_x \\ b_y \end{bmatrix} = \begin{bmatrix} |a_x| e^{i(\theta+\psi)} \\ |a_y| \end{bmatrix}$$

Substituting the elements from J_A and J_B , we can express the normalised Stokes vectors of A and B (Equation 2.6) as

$$A = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} |a_x|^2 - |a_y|^2 \\ 2\text{Re}(a_x^* a_y) \\ 2\text{Im}(a_x^* a_y) \end{bmatrix} = \begin{bmatrix} |a_x|^2 - |a_y|^2 \\ 2|a_x||a_y| \cos(\theta) \\ 2|a_x||a_y| \sin(-\theta) \end{bmatrix} \quad (\text{C.3})$$

$$B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} |b_x|^2 - |b_y|^2 \\ 2\text{Re}(b_x^* b_y) \\ 2\text{Im}(b_x^* b_y) \end{bmatrix} = \begin{bmatrix} |a_x|^2 - |a_y|^2 \\ 2|a_x||a_y|\cos(\theta + \psi) \\ 2|a_x||a_y|\sin(-\theta - \psi) \end{bmatrix} \quad (\text{C.4})$$

Equation C.2 can now be written as

$$\begin{aligned} \frac{1}{2} \Omega_{A-QB} &= \tan^{-1} \left[\frac{4|a_x|^2|a_y|^2 \sin(\psi)}{(1 - |a_x|^2 + |a_y|^2)^2 + 4|a_x|^2|a_y|^2 \cos(\psi)} \right] \\ &= \arg \left[(1 - |a_x|^2 + |a_y|^2)^2 + 4|a_x|^2|a_y|^2 e^{i\psi} \right] \end{aligned}$$

Since A and B are normalised polarization states, $|a_x|^2 + |a_y|^2 = 1$.

$$\frac{1}{2} \Omega_{A-QB} = \arg [4|a_y|^2] + \arg [|a_y|^2 + |a_x|^2 e^{i\psi}]$$

Since $\arg [4|a_y|^2] = 0$, we finally reached an expression for the phase difference between A and B as

$$\frac{1}{2} \Omega_{A-QB} = \arg [|a_x|^2 e^{i\psi} + |a_y|^2] \quad (\text{C.5})$$

The phase difference between A and B can also be derived through interference using expression 7.2. Constructive interference occurs when the two beams, each with a different polarization state, are in phase. Therefore, the phase difference equals to the amount of phase added to A for a constructive interference with B.

$$\arg (J_A \cdot J_B) = \arg (a_x^* b_x + a_y^* b_y) = \arg [|a_x|^2 e^{i\psi} + |a_y|^2] \quad (\text{C.6})$$

which is the same as the phase derived from the geometric expression $\Omega_{A-QB}/2$ in equation C.5 and thus 7.1 in the chapter 7 and C.1.

C.2 Derivations of the Stokes expression from the Jones expression of PB phase

Let us now derive the expression 7.3 from the expression 7.2 in the chapter 7. This part of derivation is based on Gutierrez paper [228] with more intermediate steps and details added and adjusted for this thesis. Since the polarization state B is transmitted from A by passing through an SLM, the Jones vector of B can be written as

$$J_B = \mathcal{J} J_A$$

where $\mathcal{J}$ is the Jones matrix of the SLM. Let us define the extraordinary axis J_Q and ordinary axis J_{-Q} of the SLM and the eigenvalues of the two axis μ_1 and μ_2 as

$$J_Q = \begin{bmatrix} q_x \\ q_y \end{bmatrix} \quad J_{-Q} = \begin{bmatrix} -q_y^* \\ q_x^* \end{bmatrix} \quad \mu_1 = e^{i\psi} \quad \mu_2 = 1$$

where J_Q and J_{-Q} are orthogonal to each other; μ_1 and μ_2 are defined based on our model system in section 7.1. Using the definitions, $\mathcal{J}$ can be defined as

$$\begin{aligned} \mathcal{J} &= \begin{bmatrix} J_Q & J_{-Q} \end{bmatrix} \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix} \begin{bmatrix} J_Q & J_{-Q} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} q_x & -q_y^* \\ q_y & q_x^* \end{bmatrix} \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix} \frac{1}{|q_x|^2 + |q_y|^2} \begin{bmatrix} q_x^* & q_y^* \\ -q_y & q_x \end{bmatrix} \end{aligned}$$

Since Q is normalised and $|q_x|^2 + |q_y|^2 = 1$

$$\mathcal{J} = \begin{bmatrix} \mu_1|q_x|^2 + \mu_2|q_y|^2 & (\mu_1 - \mu_2)q_xq_y^* \\ (\mu_1 - \mu_2)q_x^*q_y & \mu_1|q_y|^2 + \mu_2|q_x|^2 \end{bmatrix} = \begin{bmatrix} j_1 & j_2 \\ j_3 & j_4 \end{bmatrix} \quad (\text{C.7})$$

$\mathcal{J}$ can be expanded in a basis of Pauli matrices

$$\begin{aligned}
\mathcal{J} &= \frac{j_1 + j_4}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{j_1 - j_4}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \frac{j_2 + j_3}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + i \frac{j_2 - j_3}{2} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \\
&= \frac{\mu_1 + \mu_2}{2} \sigma_0 + \frac{(\mu_1 - \mu_2)(|q_x|^2 - |q_y|^2)}{2} \sigma_1 \\
&\quad + \frac{(\mu_1 - \mu_2)(q_x q_y^* + q_x^* q_y)}{2} \sigma_2 + i \frac{(\mu_1 - \mu_2)(q_x q_y^* - q_x^* q_y)}{2} \sigma_3 \\
&= \frac{\mu_1 + \mu_2}{2} Q_0 \sigma_0 + \frac{(\mu_1 - \mu_2)}{2} \sum_{n=1}^3 Q_n \sigma_n
\end{aligned}$$

where σ_n is the n th Pauli matrix; Q_n is the n th element of the normalised Stokes vector of Q .

We can now write the phase difference between A and B as

$$\begin{aligned}
\arg(J_A \cdot J_B) &= \arg(J_A \cdot (\mathcal{J} J_A)) \\
&= \arg \left[\frac{\mu_1 + \mu_2}{2} Q_0 (J_A \cdot (\sigma_0 J_A)) + \frac{(\mu_1 - \mu_2)}{2} \sum_{n=1}^3 Q_n (J_A \cdot (\sigma_n J_A)) \right]
\end{aligned}$$

By the definition of the Stokes vector, $(J_A \cdot (\sigma_n J_A)) = A_n$. We finally reached an expression for the phase difference between A and B in the form of Stokes vector

$$\begin{aligned}
\arg(J_A \cdot J_B) &= \arg \left[\frac{\mu_1 + \mu_2}{2} Q_0 A_0 + \frac{(\mu_1 - \mu_2)}{2} (Q_1 A_1 + Q_2 A_2 + Q_3 A_3) \right] \\
&= \arg [\mu_1 + \mu_2 + (\mu_1 - \mu_2)(Q \cdot A)] \tag{C.8}
\end{aligned}$$

We thus derived the expression 7.3 from the expression 7.2.

Appendix D

Derivations of the SLM settings and intermediate SOPs when passing through a 3-SLM system

As shown in figure 7.7, a beam with a polarization state A is transmitted to D after passing through a 3-SLM system; the extraordinary axis of the first SLM oriented at S1, the second at S2 and the third at S1. A and D are desired selected initial and final polarization states; the Jones vectors, J_A and J_D , and the normalised Stokes vectors, A and D, are defined as

$$J_A = \begin{bmatrix} |a_x|e^{i\theta_1} \\ |a_y| \end{bmatrix} = \begin{bmatrix} |a_x|e^{i\theta_1} \\ \sqrt{1 - |a_x|^2} \end{bmatrix}$$
$$A = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} |a_x|^2 - |a_y|^2 \\ 2|a_x||a_y|\cos\theta_1 \\ 2|a_x||a_y|\sin\theta_1 \end{bmatrix} = \begin{bmatrix} A_1 \\ \sqrt{1 - A_1^2}\cos\theta_1 \\ -\sqrt{1 - A_1^2}\sin\theta_1 \end{bmatrix} \quad (\text{D.1})$$

$$J_D = \begin{bmatrix} |d_x|e^{i\theta_2} \\ |d_y| \end{bmatrix} = \begin{bmatrix} |d_x|e^{i\theta_2} \\ \sqrt{1-|d_x|^2} \end{bmatrix}$$

$$D = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} |d_x|^2 - |d_y|^2 \\ 2|d_x||d_y|\cos\theta_2 \\ 2|d_x||d_y|\sin\theta_2 \end{bmatrix} = \begin{bmatrix} D_1 \\ \sqrt{1-D_1^2}\cos\theta_2 \\ -\sqrt{1-D_1^2}\sin\theta_2 \end{bmatrix} \quad (\text{D.2})$$

Therefore, the normalised Stokes vector A and D has two degrees of freedom and can be written in terms of A_1 , θ_1 , D_1 and θ_2 .

Since B is when a light with an initial polarization state A passes through the first SLM, C is when B passes through the second SLM, and D is when C passes through the third SLM, the normalised Stokes vector of B, C and D can be expressed in terms of A and the retardance of the first, second and third SLM, ψ_1 , ψ_2 and ψ_3 .

$$B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} A_1 \\ \sqrt{1-A_1^2}\cos(\theta_1+\psi_1) \\ -\sqrt{1-A_1^2}\sin(\theta_1+\psi_1) \end{bmatrix} \quad (\text{D.3})$$

$$\begin{aligned} C = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} &= \begin{bmatrix} B_1\cos\psi_2 - B_3\sin\psi_2 \\ B_2 \\ B_1\sin\psi_2 + B_3\cos\psi_2 \end{bmatrix} \\ &= \begin{bmatrix} A_1\cos\psi_2 + \sqrt{1-A_1^2}\sin(\theta_1+\psi_1)\sin\psi_2 \\ \sqrt{1-A_1^2}\cos(\theta_1+\psi_1) \\ A_1\sin\psi_2 - \sqrt{1-A_1^2}\sin(\theta_1+\psi_1)\cos\psi_2 \end{bmatrix} \end{aligned} \quad (\text{D.4})$$

$$\begin{aligned}
D = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} &= \begin{bmatrix} C_1 \\ C_2 \cos \psi_3 + C_3 \sin \psi_3 \\ -C_2 \sin \psi_3 + C_3 \cos \psi_3 \end{bmatrix} \\
&= \begin{bmatrix} B_1 \cos \psi_2 - B_3 \sin \psi_2 \\ B_2 \cos \psi_3 + (B_1 \sin \psi_2 + B_3 \cos \psi_2) \sin \psi_3 \\ -B_2 \sin \psi_3 + (B_1 \sin \psi_2 + B_3 \cos \psi_2) \cos \psi_3 \end{bmatrix} \\
&= \begin{bmatrix} A_1 \cos \psi_2 + \sqrt{1 - A_1^2} \sin(\theta_1 + \psi_1) \sin \psi_2 \\ \sqrt{1 - A_1^2} \cos(\theta_1 + \psi_1) \cos \psi_3 + (A_1 \sin \psi_2 - \sqrt{1 - A_1^2} \sin(\theta_1 + \psi_1) \cos \psi_2) \sin \psi_3 \\ -\sqrt{1 - A_1^2} \cos(\theta_1 + \psi_1) \sin \psi_3 + (A_1 \sin \psi_2 - \sqrt{1 - A_1^2} \sin(\theta_1 + \psi_1) \cos \psi_2) \cos \psi_3 \end{bmatrix} \tag{D.5}
\end{aligned}$$

Since the equations D.2 and D.5 are equal to each other

$$\begin{aligned}
D_1 &= A_1 \cos \psi_2 + \sqrt{1 - A_1^2} \sin(\theta_1 + \psi_1) \sin \psi_2 \\
&= \sqrt{A_1^2 + (1 - A_1^2) \sin^2(\theta_1 + \psi_1)} \sin \left[\psi_2 + \tan^{-1} \left(\frac{A_1}{\sqrt{1 - A_1^2} \sin(\theta_1 + \psi_1)} \right) \right] \\
&= \sqrt{A_1^2 + (1 - A_1^2) \sin^2(\theta_1 + \psi_1)} \sin \left[\psi_2 + \sin^{-1} \left(\frac{A_1}{\sqrt{A_1^2 + (1 - A_1^2) \sin^2(\theta_1 + \psi_1)}} \right) \right]
\end{aligned}$$

We thus derived the retardance setting of the second SLM, ψ_2 in terms of A and D and ψ_1

$$\begin{aligned}
\psi_2 &= \sin^{-1} \left(\frac{D_1}{\sqrt{A_1^2 + (1 - A_1^2) \sin^2(\theta_1 + \psi_1)}} \right) - \sin^{-1} \left(\frac{A_1}{\sqrt{A_1^2 + (1 - A_1^2) \sin^2(\theta_1 + \psi_1)}} \right) \\
&= \sin^{-1} \left(\frac{D_1}{\sqrt{1 - B_2^2}} \right) - \sin^{-1} \left(\frac{A_1}{\sqrt{1 - B_2^2}} \right) \tag{D.6}
\end{aligned}$$

Note that B_2 can be expressed in terms of A and ψ_1 (see expression D.3).

$$\begin{aligned}
D_2 &= B_2 \cos \psi_3 + (B_1 \sin \psi_2 + B_3 \cos \psi_2) \sin \psi_3 \\
&= \sqrt{B_2^2 + (B_1 \sin \psi_2 + B_3 \cos \psi_2)^2} \sin \left[\psi_3 + \tan^{-1} \left(\frac{B_2}{(B_1 \sin \psi_2 + B_3 \cos \psi_2)} \right) \right] \\
&= \sqrt{B_2^2 + (B_1 \sin \psi_2 + B_3 \cos \psi_2)^2} \sin \left[\psi_3 + \sin^{-1} \left(\frac{B_2}{\sqrt{B_2^2 + (B_1 \sin \psi_2 + B_3 \cos \psi_2)^2}} \right) \right]
\end{aligned}$$

We thus derived the retardance setting of the third SLM, ψ_3 in terms of A and D and

ψ_1

$$\begin{aligned}
\psi_3 &= \sin^{-1} \left(\frac{D_2}{\sqrt{B_2^2 + (B_1 \sin \psi_2 + B_3 \cos \psi_2)^2}} \right) - \sin^{-1} \left(\frac{B_2}{\sqrt{B_2^2 + (B_1 \sin \psi_2 + B_3 \cos \psi_2)^2}} \right) \\
&= \sin^{-1} \left(\frac{D_2}{\sqrt{1 - C_1^2}} \right) - \sin^{-1} \left(\frac{B_2}{\sqrt{1 - C_1^2}} \right) \\
&= \sin^{-1} \left(\frac{D_2}{\sqrt{1 - D_1^2}} \right) - \sin^{-1} \left(\frac{B_2}{\sqrt{1 - D_1^2}} \right) \tag{D.7}
\end{aligned}$$

Note that again B_2 can be expressed in terms of A and ψ_1 (see expression D.3).

Appendix E

Analysis of PB phase from a 3-SLM system

Section 7.3.2.1 stated the expressions for the retardance of the second and third SLM (equation 7.5 and 7.6) and the Stokes vector of the intermediate transitional points B and C (equation 7.7 and 7.8). For the convenience, I restate the equations below.

$$\begin{aligned}\psi_2 &= \sin^{-1} \left(\frac{D_1}{\sqrt{1 - B_2^2}} \right) - \sin^{-1} \left(\frac{A_1}{\sqrt{1 - B_2^2}} \right) \\ \psi_3 &= \sin^{-1} \left(\frac{D_2}{\sqrt{1 - D_1^2}} \right) - \sin^{-1} \left(\frac{B_2}{\sqrt{1 - D_1^2}} \right) \\ B &= \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \cos \psi_1 + A_3 \sin \psi_1 \\ -A_2 \sin \psi_1 + A_3 \cos \psi_1 \end{bmatrix} \\ C &= \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} D_1 \\ D_2 \cos \psi_3 - D_3 \sin \psi_3 \\ D_2 \sin \psi_3 + D_3 \cos \psi_3 \end{bmatrix}\end{aligned}$$

where S_n is the n^{th} entry of the Stokes vector of S ; ψ_1 is the retardance of the first SLM.

In section 7.3.2 and 7.3.2.2, it is shown that the total PB phase of the polarisation transitions after passing through the 3-SLM system consists of five components (equation 7.4). Substituting the expressions for ψ_2 , ψ_3 , B and C into the equation 7.4, the five components can be expressed as below

$$\Phi_{A-Q1B} = \tan^{-1} \left(\frac{\sin(\psi_1)}{\cos(\psi_1) + \frac{1-A_1}{1+A_1}} \right) \quad (\text{E.1})$$

$$\Phi_{B-Q2C} = \tan^{-1} \left(\frac{\sin(\psi_2)}{\cos(\psi_2) + \frac{1-B_2}{1+B_2}} \right) \quad (\text{E.2})$$

$$\Phi_{C-Q3D} = \tan^{-1} \left(\frac{\sin(\psi_3)}{\cos(\psi_3) + \frac{1-C_1}{1+C_1}} \right) \quad (\text{E.3})$$

$$\Phi_{ABC} = \tan^{-1} \left[\frac{\sigma A_1 (\sin(\alpha + \psi_1) - \sin \alpha) - D_1 (1 - A_1^2) \sin \psi_1}{1 + A_1^2 + 2A_1 D_1 + \sigma (\cos(\alpha + \psi_1) + \cos \alpha) + (1 - A_1^2) \cos \psi_1} \right] \quad (\text{E.4})$$

$$\Phi_{ACD} = \tan^{-1} \left[\frac{\sigma D_1 (\sin(\alpha) - \sin(\alpha - \psi_3)) - A_1 (1 - D_1^2) \sin \psi_3}{1 + D_1^2 + 2A_1 D_1 + \sigma (\cos \alpha + \cos(\alpha - \psi_3)) + (1 - D_1^2) \cos \psi_3} \right] \quad (\text{E.5})$$

where $\sigma = \sqrt{(1 - A_1^2)(1 - D_1^2)}$ and $\alpha = \psi_3 - \tan^{-1}(A_3/A_2) + \tan^{-1}(D_3/D_2)$.

Note that since the second SLM has its extraordinary axis that aligns with S2, $B_2 = C_2$. Therefore

$$\sin \left[-\psi_3 + \tan^{-1} \left(\frac{D_2}{D_3} \right) \right] = \frac{\sqrt{1 - A_1^2}}{\sqrt{1 - D_1^2}} \sin \left[\psi_1 + \tan^{-1} \left(\frac{A_2}{A_3} \right) \right] \quad (\text{E.6})$$

From the equation E.6, it is shown that the range of possible values for ψ_1 depends on the ratio $\sqrt{(1 - A_1^2)}/\sqrt{(1 - D_1^2)}$ because the sine function cannot have an amplitude exceed 1; when $|A_1| \geq |D_1|$, $\sqrt{(1 - A_1^2)}/\sqrt{(1 - D_1^2)} < 1$; ψ_1 can access a whole range of $-\pi$ to π while ψ_3 has only a limited range of values; when $|A_1| < |D_1|$, $\sqrt{(1 - A_1^2)}/\sqrt{(1 - D_1^2)} > 1$; ψ_1 can only access a limited range of values and ψ_3 has a whole range of $-\pi$ to π .

I will discuss how the PB phase varies with ψ_1 in four cases, when $A_1 = D_1 \neq 0$,

$A_1 = -D_1$, $|A_1| > |D_1|$ and $|A_1| < |D_1|$, to provide a complete discussion.

E.1 The PB phase variations with ψ_1 when $A_1 = D_1 \neq 0$

Using the equations 7.5, 7.6 and E.6

$$\begin{aligned}\psi_2 &= 0 \\ \psi_3 &= \tan^{-1} \left(\frac{A_3}{A_2} \right) - \psi_1 - \tan^{-1} \left(\frac{D_3}{D_2} \right) = \theta - \psi_1\end{aligned}$$

where $\theta = \tan^{-1} (A_3/A_2) - \tan^{-1} (D_3/D_2)$.

Since $A_1 = -D_1$, $\sigma = 1 - A_1^2 = 1 - D_1^2$. Substitute ψ_3 into α , $\alpha = -\psi_1$. The five components of Φ thus equal to

$$\begin{aligned}\Phi_{A-Q1B} &= \tan^{-1} \left(\frac{\sin(\psi_1)}{\cos(\psi_1) + \zeta} \right) \\ \Phi_{B-Q2C} &= 0 \\ \Phi_{C-Q3D} &= -\tan^{-1} \left(\frac{\sin(\psi_1 - \theta)}{\cos(\psi_1 - \theta) + \zeta} \right) \\ \Phi_{ABC} &= 0 \\ \Phi_{ACD} &= -\tan^{-1} \left\{ \frac{A_1(1 - A_1^2)[\sin \theta - \sin \psi_1 + \sin(\psi_1 - \theta)]}{1 + 3A_1^2 + (1 - A_1^2)[\cos \theta + \cos \psi_1 + \cos(\psi_1 - \theta)]} \right\}\end{aligned}$$

where ψ_1 is the retardance of the first SLM; $\theta = \tan^{-1} (A_3/A_2) - \tan^{-1} (D_3/D_2)$; $\zeta = (1 - A_1)/(1 + A_1)$.

E.1.1 When $A_1 = D_1 = 1$

$$\zeta = 0$$

$$\Phi_{ACD} = 0$$

$$\Phi = \psi_1 - (\psi_1 - \theta) = \theta$$

This is when both the initial and final polarisation state lies along the S1 which is the orientation of the extraordinary axis of both the first and third SLM. In this case, there is no polarisation modulation between the initial and final states; both the first and third SLM can serve as phase modulators for a full $\pm\pi$ range of phase correction.

E.1.2 When $A_1 = D_1 = -1$

$$\zeta = \infty$$

$$\Phi_{ACD} = 0$$

$$\Phi = 0$$

This is when both the initial and final polarisation state lies along the -S1 which is the orientation of the ordinary axis of both the first and third SLM. Neither the first nor the third SLM can have any effect on the beam; the phase property of the light cannot be modulated.

E.1.3 When $|A_1| = |D_1| \neq 1 \neq 0$

Φ_{ACD} always has a small value such that

$$-\frac{1}{2}\pi < \Phi_{ACD} < \frac{1}{2}\pi$$

This is because for any choice of θ and ψ_1

$$-1 < \frac{A_1(1 - A_1^2)[\sin \theta - \sin \psi_1 + \sin(\psi_1 - \theta)]}{1 + 3A_1^2 + (1 - A_1^2)[\cos \theta + \cos \psi_1 + \cos(\psi_1 - \theta)]} < 1$$

and hence

$$-\frac{1}{2}\pi < \Phi_{ACD} = -\tan^{-1} \left\{ \frac{A_1(1 - A_1^2)[\sin \theta - \sin \psi_1 + \sin(\psi_1 - \theta)]}{1 + 3A_1^2 + (1 - A_1^2)[\cos \theta + \cos \psi_1 + \cos(\psi_1 - \theta)]} \right\} < \frac{1}{2}\pi$$

Φ_{A-Q1B} equals to the negative of Φ_{C-Q3Q} with a shift θ . When $-1 < A_1 = D_1 < 0$ [see Figure E.1(a)]

$$\zeta > 1$$

$$\begin{aligned} -\pi/2 < \Phi_{A-Q1B} &= \tan^{-1} \left[\frac{\sin(\psi_1)}{\cos(\psi_1) + \zeta} \right] < \pi/2 \\ -\pi/2 < \Phi_{C-Q3D} &= -\tan^{-1} \left[\frac{\sin(\psi_1 - \theta)}{\cos(\psi_1 - \theta) + \zeta} \right] < \pi/2 \end{aligned}$$

and when $0 < A_1 = D_1 < 1$ [see Figure E.1(b)]

$$\zeta < 1$$

$$\begin{aligned} -\pi < \Phi_{A-Q1B} &= \tan^{-1} \left[\frac{\sin(\psi_1)}{\cos(\psi_1) + \zeta} \right] < \pi \\ -\pi < \Phi_{C-Q3D} &= -\tan^{-1} \left[\frac{\sin(\psi_1 - \theta)}{\cos(\psi_1 - \theta) + \zeta} \right] < \pi \end{aligned}$$

For both cases, the variations of the sum of Φ_{A-Q1B} and Φ_{C-Q3Q} with respect to a varying ψ_1 always have a range smaller than $\pm\pi/2$. In fact, the total PB phase, $\Phi = \Phi_{ACD} + \Phi_{A-Q1B} + \Phi_{C-Q3D}$, is a constant value for any ψ_1 . Since the second SLM in this case has 0 retardance, and the first and third SLM, with the same extraordinary axis, have a total retardance of θ and a resultant eigenvector of S1, the total PB phase introduced by the 3-SLM system is thus suppose to be a constant regardless of the

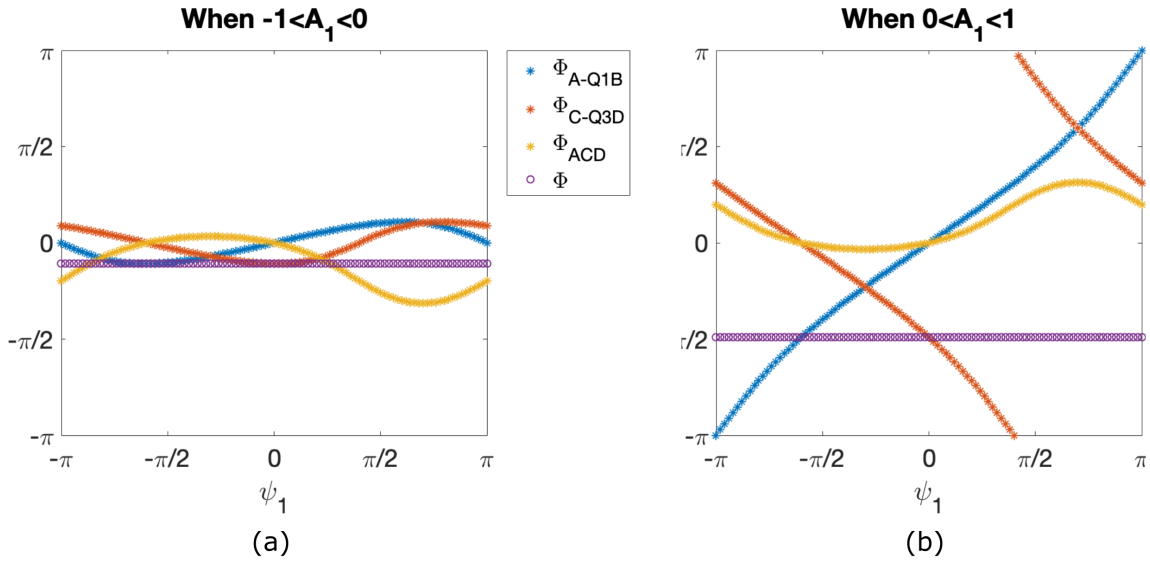


Figure E.1: The variations of Φ_{A-Q1B} , Φ_{C-Q3Q} , Φ_{ACD} and Φ against ψ_1 , where Φ is the sum of all the components of the total PB phase. The plots show two cases when $-1 < A_1 < 0$ (a) and when $0 < A_1 < 1$ (b). In either case, Φ is a constant value for any ψ_1 .

choice of the first SLM retardance.

Therefore, Φ has a full range of 2π only when $A_1 = D_1 = 1$; this is when the first and third SLM have their extraordinary axes align with the input polarisation state A and thus the SLMs only serve as phase modulators. When $A_1 = D_1 \neq 0$ or 1 , the total PB phase introduced by the 3-SLM system is a constant regardless of the choice of the first SLM retardance.

E.2 The PB phase variations with ψ_1 when $A_1 = -D_1$

Using the equations 7.5, 7.6 and E.6

$$\begin{aligned}\psi_2 &= -2 \sin^{-1} \left(\frac{A_1}{\sqrt{1 - B_2^2}} \right) \\ \psi_3 &= \tan^{-1} \left(\frac{A_3}{A_2} \right) - \psi_1 - \tan^{-1} \left(\frac{D_3}{D_2} \right) = \theta - \psi_1\end{aligned}$$

Since $A_1 = -D_1$, $\sigma = 1 - A_1^2 = 1 - D_1^2$. Substitute ψ_3 into α , $\alpha = -\psi_1$. The five components of Φ thus equal to

$$\begin{aligned}\Phi_{A-Q1B} &= \tan^{-1} \left(\frac{\sin \psi_1}{\cos \psi_1 + \zeta} \right) \\ \Phi_{B-Q2C} &= \tan^{-1} \left[\frac{A_1 \sin(\psi_1 + \theta_1)}{\sqrt{1 - A_1^2} - \cos(\psi_1 + \theta_1)} \right] \\ \Phi_{C-Q3D} &= -\tan^{-1} \left[\frac{\sin(\psi_1 - \theta)}{\cos(\psi_1 - \theta) + \frac{1}{\zeta}} \right] \\ \Phi_{ABC} &= \tan^{-1} \left(\frac{A_1 \sin \psi_1}{\cos \psi_1 + \frac{1}{2}} \right) \\ \Phi_{ACD} &= \tan^{-1} \left\{ \frac{A_1 [\sin \psi_1 - \sin \theta - \sin(\psi_1 - \theta)]}{1 + \cos \psi_1 + \cos \theta + \cos(\psi_1 - \theta)} \right\}\end{aligned}$$

where ψ_1 is the retardance of the first SLM; $\theta_1 = -\tan^{-1}(A_3/A_2)$; $\theta = \tan^{-1}(A_3/A_2) - \tan^{-1}(D_3/D_2)$; $\zeta = (1 - A_1)/(1 + A_1)$.

E.2.1 When $A_1 = -D_1 \neq 0$

It can be easily seen that

$$-\pi \leq \Phi_{ABC} = \tan^{-1} \left(\frac{A_1 \sin \psi_1}{\cos \psi_1 + \frac{1}{2}} \right) \leq \pi$$

and since $0 \leq \sqrt{1 - A_1^2} \leq 1$,

$$-\pi \leq \Phi_{B-Q2C} = \tan^{-1} \left[\frac{A_1 \sin(\psi_1 + \theta_1)}{\sqrt{1 - A_1^2} - \cos(\psi_1 + \theta_1)} \right] \leq \pi$$

It should also be easy to see that

$$-\pi/2 < \Phi_{ACD} = \tan^{-1} \left\{ \frac{A_1 [\sin \psi_1 - \sin \theta - \sin(\psi_1 - \theta)]}{1 + \cos \psi_1 + \cos \theta + \cos(\psi_1 - \theta)} \right\} < \pi/2$$

since $-1 \leq A_1 \leq 1$ [see Figure E.2 (c) and (d)].

When $A_1 = -D_1 < 0$ [see Figure E.2 (a)]

$$\begin{aligned} \zeta &> 1 \\ -\pi/2 < \Phi_{A-Q1B} &= \tan^{-1} \left[\frac{\sin(\psi_1)}{\cos(\psi_1) + \zeta} \right] < \pi/2 \\ -\pi \leq -\tan^{-1} \left[\frac{\sin(\psi_1 - \theta)}{\cos(\psi_1 - \theta) + \frac{1}{\zeta}} \right] &= \Phi_{C-Q3D} \leq \pi \end{aligned}$$

and when $A_1 = -D_1 > 0$ [see Figure E.2 (b)]

$$\begin{aligned} \zeta &< 1 \\ -\pi \leq \Phi_{A-Q1B} &= \tan^{-1} \left[\frac{\sin(\psi_1)}{\cos(\psi_1) + \zeta} \right] \leq \pi \\ -\pi/2 < \Phi_{C-Q3D} &= -\tan^{-1} \left[\frac{\sin(\psi_1 - \theta)}{\cos(\psi_1 - \theta) + \frac{1}{\zeta}} \right] < \pi/2 \end{aligned}$$

Hence, when $A_1 = -D_1 \neq 0$, the sum of the five components of Φ has a total range of $-\pi$ to π .

E.2.2 When $A_1 = -D_1 = 0$

Both points A and D fall on the great circle passing through S2 and S3. The second SLM can serve as a phase modulator if the polarisation state of point B after passing

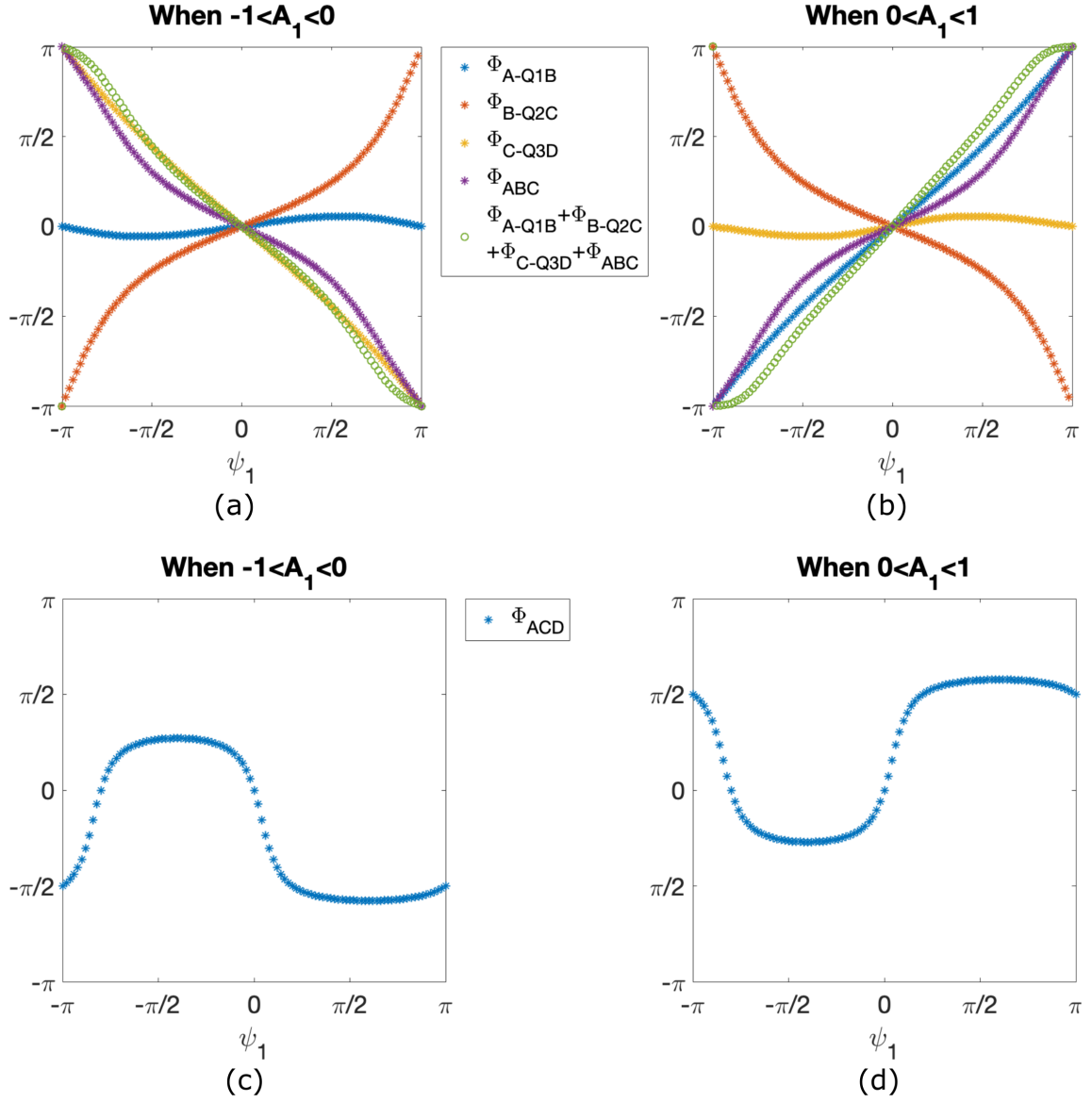


Figure E.2: The plot of Φ_{A-Q1B} , Φ_{B-Q2C} , Φ_{C-Q3D} , Φ_{ABC} and the sum of these four components of Φ when (a) $-1 < A_1 < 0$ and (b) $0 < A_1 < 1$. The plot of Φ_{ACD} when (c) $-1 < A_1 < 0$ and (d) $0 < A_1 < 1$.

through the first SLM lies along S2.

Therefore, when $A_1 = -D_1$, the total PB phase, Φ , introduced by the 3-SLM system always has a full range of 2π .

E.3 The PB phase variations with ψ_1 when $|A_1| > |D_1|$

To analyse the PB phase when $|A_1| > |D_1|$, let us define an input polarisation state A and three output polarisation states D, D' and D'' such that

$$A = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \quad D = \begin{pmatrix} A_1 \\ D_2 \\ D_3 \end{pmatrix} \quad D' = \begin{pmatrix} D'_1 \\ D_2 \\ D'_3 \end{pmatrix} \quad D'' = \begin{pmatrix} -A_1 \\ D_2 \\ D_3 \end{pmatrix}$$

where $|D'_1| < |A_1|$ (see Figure E.3). The three paths (A to B to C to D, A to B to C' to D' and A to B to C'' to D'') are thus the conversions of A to D, A to D' and A to D'' after passing through a 3-SLM system with different retardance settings respectively.

The total PB phase of the conversion of A to D, A to D' and A to D'' can be calculated as

$$\Phi_{AD} = \Phi_{A-Q1B} + \Phi_{B-Q1D} + \Phi_{ABD}$$

$$\Phi_{AD'} = \Phi_{A-Q1B} + \Phi_{B-Q2C'} + \Phi_{C'-Q1D'} + \Phi_{ABC'} + \Phi_{AC'D'}$$

$$\Phi_{AD''} = \Phi_{A-Q1B} + \Phi_{B-Q2C''} + \Phi_{C''-Q1D''} + \Phi_{ABC''} + \Phi_{AC''D''}$$

where Φ_{AD} , $\Phi_{AD'}$ and $\Phi_{AD''}$ are the PB phase difference of D to A, D' to A and D'' to A respectively.

From section E.1, Φ_{AD} is a constant for a selected A and D invariant to ψ_1 . From

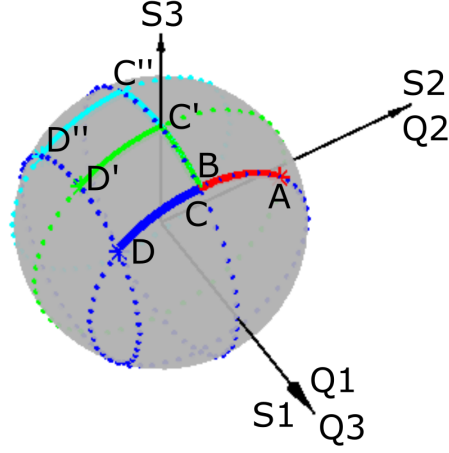


Figure E.3: Transitions of polarisation state A to D, D' and D'' after passing through a 3-SLM system with different retardance settings. The red path shows a conversion of A to an intermediate point B after passing through the first SLM with a retardance of ψ_1 . The blue path shows a conversion of B to D. The green path shows a conversion of B to D'. The cyan path shows a conversion of B to D''.

section E.2, $\Phi_{AD''}$ has a full range of $-\pi$ to π when ψ_1 varies.

Using the definition in section 7.3.1.1, the PB phase can be visualised as the spherical triangle enclosed on the Poincaré sphere (see Figure E.4 as an example). Therefore, it is easy to show that

$$|\Phi_{B-Q2C''}| > |\Phi_{B-Q2C'}| > 0 \quad (\text{E.7})$$

$$|\Phi_{C''-Q1D''}| < |\Phi_{C'-Q1D'}| < |\Phi_{C-Q1D}| \quad (\text{E.8})$$

$$|\Phi_{ABC''}| > |\Phi_{ABC'}| > 0 \quad (\text{E.9})$$

$$|\Phi_{AC''D''}| < |\Phi_{AC'D'}| < |\Phi_{ACD}| \quad (\text{E.10})$$

It can therefore be shown that the range of variations of the total PB phase $\Phi_{AD'}$ is always larger than Φ_{AD} and smaller than $\Phi_{AD''}$.

Therefore, when $|A_1| > |D_1|$, the total PB phase, Φ , introduced by the 3-SLM system has a range smaller than 2π .

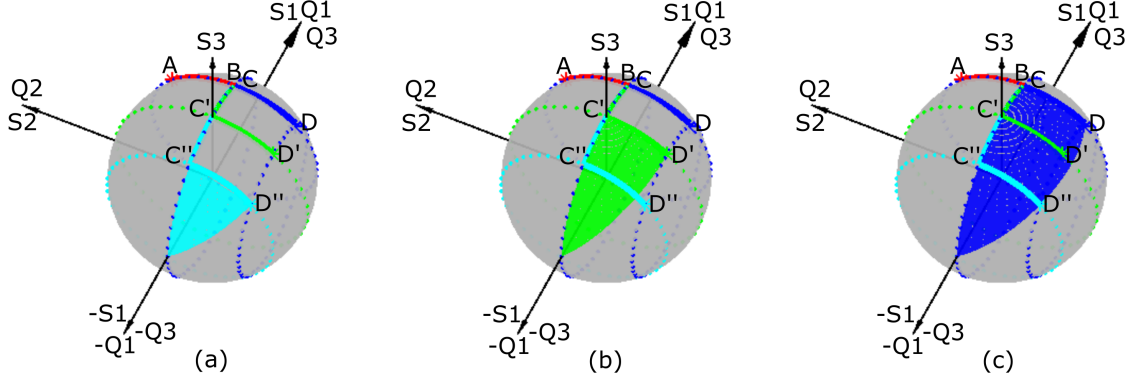


Figure E.4: The cyan area marked in (a) is a visualisation of the PB phase of D'' to C'' , $\Phi_{C''-Q3D''}$ on the Poincaré sphere. The green area marked in (b) is the PB phase of D' to C' , $\Phi_{C'-Q3D'}$. The blue area marked in (c) is the PB phase of D to C , Φ_{C-Q3D} . It can be clearly seen that $|\Phi_{C''-Q3D''}|$ is always smaller than $|\Phi_{C'-Q3D'}|$ and $|\Phi_{C'-Q3D'}|$ is always smaller than $|\Phi_{C-Q3D}|$ regardless the choice of the retardance setting of the first SLM, ψ_1 (hence the equation E.8).

E.4 The PB phase variations with ψ_1 when $|A_1| < |D_1|$

The analysis of the PB phase when $|A_1| < |D_1|$ is very similar to the analysis in the section E.3 when $|A_1| > |D_1|$. Let us defined three input polarisation states A , A' and A'' and an output polarisation state D such that

$$A = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \quad A' = \begin{pmatrix} A'_1 \\ A_2 \\ A'_3 \end{pmatrix} \quad A'' = \begin{pmatrix} -A_1 \\ A_2 \\ A_3 \end{pmatrix} \quad D = \begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix}$$

where $|A'_1| < |D_1|$ (see Figure E.5). The three paths (A to B to C to D , A' to B' to C to D and A'' to B'' to C to D) are thus the conversions of A to D , A' to D and A'' to D after passing through a 3-SLM system with different retardance settings respectively.

After some analysis similar to the section E.3, it can be shown that the range of variations of the total PB phase $\Phi_{A'D}$ is always larger than Φ_{AD} and smaller than

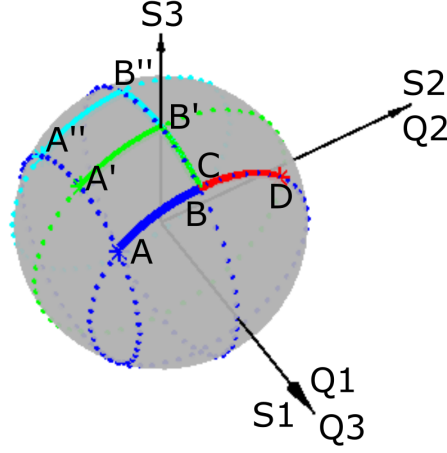


Figure E.5: Transitions of polarisation state A to D, A' to D and A'' to D after passing through a 3-SLM system with different retardance settings. The blue path shows a conversion of A to an intermediate point C after passing through the first and second SLM. The green path shows a conversion of A' to C. The cyan path shows a conversion of A'' to C. The red path shows a conversion of C to D after passing through the third SLM.

$\Phi_{A''D}$.

Therefore, when $|A_1| < |D_1|$, the total PB phase, Φ , introduced by the 3-SLM system again has a range smaller than 2π .

E.5 Conclusion

The PB phase introduced by the 3-SLM system is a constant value when the first element of the initial polarisation state equals to that of the final polarisation state and not equal to zero or one, i.e. $A_1 = D_1 \neq 0$ or 1. The PB phase has a full range of 2π when the first element of the initial polarisation state is equal and opposite to that of the final polarisation state or when the first element of both the initial and final state equals to 1, i.e. $A_1 = -D_1$ or $A_1 = D_1 = 1$. The PB phase has a range less than 2π if $|A_1| \neq |D_1|$. (Table E.1)

	$A_1 = D_1 \neq 0$ or $\neq 1$	$A_1 = -D_1$ or $A_1 = D_1 = 1$	$ A_1 \neq D_1 $
The range of the PB phase variations	0	2π	$< 2\pi$

Table E.1: The range of the PB phase variations with the retardance of the first SLM.

Appendix F

The PB phase analysis after passing through a 4-SLM system

The proposed configuration for a full vectorial state conversion consists of a sequence of the 4-SLM with orientations align to S1, -S1, S2, S1. Considering the first two SLM passes (polarisation state A to B to C), the PB phases are given by

$$\begin{aligned}
& \Phi_{A-Q1B} + \Phi_{B-Q2C} + \Phi_{ABC} \\
&= \tan^{-1} \left(\frac{\sin \psi_1}{\cos \psi_1 + \frac{1-A_1}{1+A_1}} \right) + \tan^{-1} \left(\frac{\sin \psi_2}{\cos \psi_2 + \frac{1+A_1}{1-A_1}} \right) \\
&\quad - \tan^{-1} \left\{ \frac{A_1 [\sin(\psi_1 + \psi_2) - \sin \psi_1 - \sin \psi_2]}{\frac{1+3A_1^2}{1-A_1^2} + [\cos(\psi_1 + \psi_2) + \cos \psi_1 + \cos \psi_2]} \right\} \\
&= \tan^{-1} \left[\frac{\sin(\psi_1)}{\cos(\psi_1) + \zeta} \right] + \tan^{-1} \left[\frac{\sin(\vartheta - \psi_1)}{\cos(\vartheta - \psi_1) + \frac{1}{\zeta}} \right] \\
&\quad - \tan^{-1} \left\{ \frac{A_1 [\sin \vartheta - \sin \psi_1 - \sin(\vartheta - \psi_1)]}{\frac{1+3A_1^2}{1-A_1^2} + [\cos \vartheta + \cos \psi_1 + \cos(\vartheta - \psi_1)]} \right\} \tag{F.1}
\end{aligned}$$

where ψ_1 and ψ_2 are the retardance setting of the first and second SLM respectively; $\vartheta = \psi_1 + \psi_2$; $\zeta = (1 - A_1)/(1 + A_1)$.

When $A_1 < 0$, $1 < \zeta$ and $0 < 1/\zeta < 1$. Φ_{A-Q1B} has a range smaller than π and

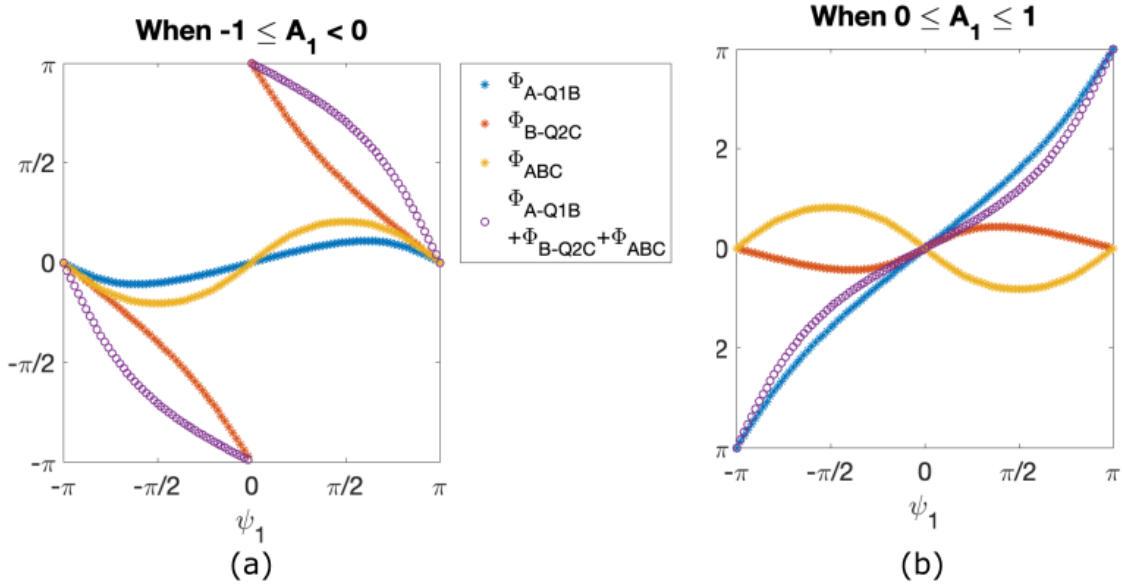


Figure F.1: Two plots of the PB phase, Φ_{A-Q1B} , Φ_{B-Q2C} and Φ_{ABC} , varying with the choice of the first SLM retardance setting, ψ_1 , after passing through the first two SLMs when (a) $A_1 < 0$ and (b) $0 \leq A_1$.

Φ_{B-Q2C} has a full range of $-\pi$ to π . The sum of the Φ_{A-Q1B} , Φ_{B-Q2C} and Φ_{ABC} therefore has a range of 2π (see Figure F.1 (a)).

When $A_1 \geq 0$, $0 < \zeta \leq 1$ and $1 \leq 1/\zeta$. Φ_{A-Q1B} has a full range of $-\pi$ to π and Φ_{B-Q2C} has a range smaller than π . The sum of the Φ_{A-Q1B} , Φ_{B-Q2C} and Φ_{ABC} therefore has a range of 2π (see Figure F.1 (b)).

For a fixed ϑ , the phase introduced by the third and fourth pass is a constant value; i.e. $\Phi_{C-Q3D} + \Phi_{D-Q4E} + \Phi_{ACD} + \Phi_{ADE} = \text{constant}$. Therefore, the total PB phase of the 4-SLM system always has a full range of 2π .

We conclude therefore that the proposed four pass SLM system can act as a full vectorial state transformer.

Appendix G

The mathematical expressions for homogeneous retarders

G.1 The general expression for a homogeneous retarder

This section was reproduced from Gutierrez paper [228] with some adjustments to assist the discussion in this appendix. A homogeneous retarder has two orthonormal eigenvectors. Suppose one of the eigenvector is defined as

$$q_1 = \begin{pmatrix} q_x \\ q_y \end{pmatrix}$$

The other eigenvector, q_2 is orthogonal to q_1 and thus defined as

$$q_2 = \begin{pmatrix} -q_y^* \\ q_x^* \end{pmatrix}$$

where $|q_x|^2 + |q_y|^2 = 1$ since the two eigenvectors are normalised; $q_1 \cdot q_2^* = 0$ since the two eigenvectors are orthonormal to one another.

Suppose the two eigenvalues correspond to the two eigenvectors are λ_1 and λ_2 . For a homogeneous retarder, the eigenvalues indicate the retardance applied to each axis. λ_1 and λ_2 are thus in the form of $\exp(i\psi_x)$ and $\exp(i\psi_y)$ where ψ_x and ψ_y are the retardance each applied along the axis respectively. The Jones matrix of the homogeneous retarder, J , can be expressed as

$$\begin{aligned} \mathcal{J} &= Q\Lambda Q^{-1} \\ &= \begin{pmatrix} q_x & -q_y^* \\ q_y & q_x^* \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} q_x^* & q_y^* \\ -q_y & q_x \end{pmatrix} \\ &= \sqrt{\lambda_1 \lambda_2} \begin{pmatrix} q_x & -q_y^* \\ q_y & q_x^* \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^* \end{pmatrix} \begin{pmatrix} q_x^* & q_y^* \\ -q_y & q_x \end{pmatrix} \end{aligned} \quad (\text{G.1})$$

where $\lambda = \sqrt{\lambda_1/\lambda_2}$; $\sqrt{\lambda_2/\lambda_1}$ therefore equals to λ^* . $\sqrt{\lambda_1 \lambda_2} = \exp[i(\psi_x + \psi_y)/2]$ which is the square root of the determinant of Λ ; $\arg[\sqrt{\det(\Lambda)}] = (\psi_x + \psi_y)/2$ is the dynamic phase of the retarder $\mathcal{J}$.

After multiply Q , Λ and Q^{-1} , a general expression for a homogeneous retarder can be written as

$$\mathcal{J} = \sqrt{\lambda_1 \lambda_2} \begin{bmatrix} \lambda|q_x|^2 + \lambda^*|q_y|^2 & (\lambda - \lambda^*)q_x q_y^* \\ (\lambda - \lambda^*)q_x^* q_y & \lambda^*|q_x|^2 + \lambda|q_y|^2 \end{bmatrix} \quad (\text{G.2})$$

G.2 The general expression for a homogeneous matrix

It is interesting to see that all the four matrix in the previous section, Q , $\Lambda/\sqrt{\det(\Lambda)}$, Q^{-1} and $\mathcal{J}/\sqrt{\det(\Lambda)}$ are in the same form of

$$\begin{pmatrix} A & -B^* \\ B & A^* \end{pmatrix} \quad (\text{G.3})$$

In fact, a matrix in this form is a homogeneous matrix with a set of orthogonal eigenvectors. It is easy to show that the eigenvalues of the matrix G.3 are

$$\lambda_{1,2} = \frac{A + A^* \pm \sqrt{(A - A^*)^2 - 4BB^*}}{2} \quad (\text{G.4})$$

The eigenvectors of the matrix G.3 are

$$q_1 = \begin{bmatrix} 1 \\ \frac{A - A^* + \sqrt{(A - A^*)^2 - 4BB^*}}{2B^*} \end{bmatrix} \quad q_2 = \begin{bmatrix} 1 \\ \frac{2B}{A - A^* + \sqrt{(A - A^*)^2 - 4BB^*}} \end{bmatrix}$$

and $q_1 \cdot q_2^* = 0$.

It is therefore shown that any matrix in the form of G.3 is a homogeneous matrix with orthogonal eigenvectors.

It is also interesting to note that the determinant of the matrix G.3 is $|A|^2 + |B|^2$ which equals to $\lambda_1 \lambda_2$. Therefore, the determinant of a homogeneous matrix equals to the multiplication of its two eigenvalues.

G.3 The general expression for a sequence of homogeneous retarders

As shown in the equation G.1 and G.3, when three homogeneous matrices multiply together, the resultant matrix is also homogeneous. In fact, the resultant matrix is always homogeneous when multiply homogeneous matrices multiply together.

Suppose light passing through two sequential homogeneous retarder with Jones matrices expressed as

$$\mathcal{J}_1 = \begin{pmatrix} A & -B^* \\ B & A^* \end{pmatrix} \quad \mathcal{J}_2 = \begin{pmatrix} C & -D^* \\ D & C^* \end{pmatrix}$$

The resultant system of these two homogeneous retarder can be written as

$$\mathcal{J} = \begin{pmatrix} AC - B^*D & -AD^* - B^*C^* \\ BC + A^*D & -BD^* + A^*C^* \end{pmatrix}$$

which is also in the form of a homogeneous matrix G.3.

It is thus shown that the system contains only homogeneous retarders is also a homogeneous system.

Appendix H

The derivations of the retardance settings of the three SLMs for the polarisation compensations of an arbitrary retarder

As discussed in the section 7.4.2 and I restated below, the combined Jones matrix of the three SLMs needs to equal to the inverse matrix of the retarder for correction.

$$\mathcal{J}'_{3slm} = \mathcal{J}'^{-1} = \begin{bmatrix} \lambda^*|q_x|^2 + \lambda|q_y|^2 & (\lambda^* - \lambda)q_xq_y^* \\ (\lambda^* - \lambda)q_x^*q_y & \lambda|q_x|^2 + \lambda^*|q_y|^2 \end{bmatrix} \quad (\text{H.1})$$

$\mathcal{J}_{3slm}$ can also be written by multiplying the Jones matrices of the three SLMs. The SLM model used in this thesis is defined in the section 7.1. Since the first, second and third SLM have extraordinary axes oriented at S1, S2 and S1 and their

retardance settings are ψ_1 , ψ_2 and ψ_3 respectively,

$$\begin{aligned}
\mathcal{J}_{3slm} &= \frac{1}{2} \begin{pmatrix} e^{i\psi_1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\psi_2} + 1 & e^{i\psi_2} - 1 \\ e^{i\psi_2} - 1 & e^{i\psi_2} + 1 \end{pmatrix} \begin{pmatrix} e^{i\psi_3} & 0 \\ 0 & 1 \end{pmatrix} \\
&= \frac{e^{i(\psi_1+\psi_2+\psi_3)/2}}{2} \begin{bmatrix} (e^{i\psi_2/2} + e^{-i\psi_2/2})e^{i(\psi_1+\psi_3)/2} & (e^{i\psi_2/2} - e^{-i\psi_2/2})e^{i(\psi_1-\psi_3)/2} \\ (e^{i\psi_2/2} - e^{-i\psi_2/2})e^{i(-\psi_1+\psi_3)/2} & (e^{i\psi_2/2} + e^{-i\psi_2/2})e^{i(-\psi_1-\psi_3)/2} \end{bmatrix} \\
&= \sqrt{\det(\mathcal{J}_{3slm})} \mathcal{J}'_{3slm}
\end{aligned} \tag{H.2}$$

where the argument of $\sqrt{\det(\mathcal{J}_{3slm})}$ is the dynamic phase of the 3-SLM system. $\mathcal{J}'_{3slm}$ is the Jones matrix of the system without the dynamic phase term.

It can be easily shown that $\mathcal{J}'_{3slm}$ is also in the homogeneous matrix form defined in the equation G.3 in the appendix G.2. Using the equation G.4, the two eigenvalues, $\mu_{1,2}$ of $\mathcal{J}'_{3slm}$ are

$$\mu_{1,2} = \cos\left(\frac{\psi_2}{2}\right) \cos\left(\frac{\psi_1 + \psi_3}{2}\right) \pm i \sqrt{1 - \cos^2\left(\frac{\psi_2}{2}\right) \cos^2\left(\frac{\psi_1 + \psi_3}{2}\right)} \tag{H.3}$$

To make the retardance of the 3-SLM system equal and opposite to that of the retarder and thus the eigenvalues conjugate, $\mu_1 = \lambda^*$ and $\mu_2 = \lambda$. Therefore

$$\begin{aligned}
\arg(\lambda^*) &= \arg(\mu_1) = \tan^{-1} \left[\frac{\sqrt{1 - \cos^2\left(\frac{\psi_2}{2}\right) \cos^2\left(\frac{\psi_1 + \psi_3}{2}\right)}}{\cos\left(\frac{\psi_2}{2}\right) \cos\left(\frac{\psi_1 + \psi_3}{2}\right)} \right] \\
&= \cos^{-1} \left[\cos\left(\frac{\psi_2}{2}\right) \cos\left(\frac{\psi_1 + \psi_3}{2}\right) \right] \\
\cos[\arg(\lambda^*)] &= \cos\left(\frac{\psi_2}{2}\right) \cos\left(\frac{\psi_1 + \psi_3}{2}\right)
\end{aligned} \tag{H.4}$$

From the geometrical equation 7.12,

$$\cos\left(\frac{\psi_2}{2}\right) = \frac{A_1}{\sqrt{1 - (1 - A_1^2) \cos^2\left(\frac{\psi_1 + \psi_3}{2}\right)}} \tag{H.5}$$

Substitute the equation H.5 into H.4,

$$\begin{aligned}
(1 - (1 - A_1^2) \cos^2(\frac{\psi_1 + \psi_3}{2})) \cos^2 [\arg (\lambda^*)] &= A_1^2 \cos^2 \left(\frac{\psi_1 + \psi_3}{2} \right) \\
\cos^2 \left(\frac{\psi_1 + \psi_3}{2} \right) \{ (1 - A_1^2) \cos^2 [\arg (\lambda^*)] + A_1^2 \} &= \cos^2 [\arg (\lambda^*)] \\
\cos^2 \left(\frac{\psi_1 + \psi_3}{2} \right) &= \frac{\cos^2 [\arg (\lambda^*)]}{1 - (1 - A_1^2) \sin^2 [\arg (\lambda^*)]} \\
\cos \left(\frac{\psi_1 + \psi_3}{2} \right) &= \frac{\cos [\arg (\lambda^*)]}{\sqrt{\cos^2 [\arg (\lambda^*)] + A_1^2 \sin^2 [\arg (\lambda^*)]}} \\
\tan \left(\frac{\psi_1 + \psi_3}{2} \right) &= \frac{A_1 \sin [\arg (\lambda^*)]}{\cos [\arg (\lambda^*)]} = A_1 \tan [\arg (\lambda^*)] \\
\frac{\psi_1 + \psi_3}{2} &= \tan^{-1} \{ A_1 \tan [\arg (\lambda^*)] \}
\end{aligned} \tag{H.6}$$

From the geometrical equation 7.11

$$\psi_1 = \tan^{-1} \{ A_1 \tan [\arg (\lambda^*)] \} - \tan^{-1} \left(\frac{A_3}{A_2} \right) \tag{H.7}$$

$$\psi_3 = \tan^{-1} \{ A_1 \tan [\arg (\lambda^*)] \} + \tan^{-1} \left(\frac{A_3}{A_2} \right) \tag{H.8}$$

which are the equation 7.13 and 7.14 in the section 7.4.2.2.

Substitute the equation H.6 into H.5

$$\begin{aligned}
\cos \left(\frac{\psi_2}{2} \right) &= \frac{A_1}{\sqrt{A_1^2 \cos^2(\frac{\psi_1 + \psi_3}{2}) + \sin^2(\frac{\psi_1 + \psi_3}{2})}} \\
&= \frac{1}{\cos(\frac{\psi_1 + \psi_3}{2}) \sqrt{1 + \tan^2 [\arg (\lambda^*)]}} \\
&= \sqrt{1 - (1 - A_1^2) \sin^2 [\arg (\lambda^*)]} \\
\psi_2 &= 2 \sin^{-1} \{ \sqrt{1 - A_1^2 \sin^2 [\arg (\lambda^*)]} \}
\end{aligned} \tag{H.9}$$

which is the equation 7.15 in the section 7.4.2.2.

Bibliography

- [1] J. Goodman, *Introduction to Fourier Optics*. W. H. Freeman, 2017.
- [2] V. Mahajan, *Optical imaging and aberrations: Part II: Wave diffraction optics: Second edition*. January 2001.
- [3] M. J. Booth, “Adaptive optics in microscopy,” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1861, pp. 2829–2843, 2007.
- [4] F. Wright, *The Methods of Petrographic-microscopic Research, Their Relative Accuracy and Range of Application*. Carnegie Institution of Washington publication, Carnegie Institution of Washington, 1911.
- [5] E. Hecht and A. Zajac, *Optics*. Addison-Wesley Series in Physics, Addison-Wesley Publishing Company, 1974.
- [6] V. S. Ignatowski, “Diffraction by a lens of arbitrary aperture,” *Trans. Opt. Inst.*, vol. 1(4), p. 1, 1919.
- [7] B. Richards, E. Wolf, and D. Gabor, “Electromagnetic diffraction in optical systems, ii. structure of the image field in an aplanatic system,” *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 253, no. 1274, pp. 358–379, 1959.

- [8] J. J. Stamnes, *Waves in focal regions : propagation, diffraction, and focusing of light, sound, and water waves*. Adam Hilger series on optics and optoelectronics, Bristol: Adam Hilger, 1986.
- [9] B. Boruah and M. Neil, “Focal field computation of an arbitrarily polarized beam using fast fourier transforms,” *Optics Communications*, vol. 282, no. 24, pp. 4660–4667, 2009.
- [10] R. C. Devlin, A. Ambrosio, N. A. Rubin, J. P. B. Mueller, and F. Capasso, “Arbitrary spin-to-orbital angular momentum conversion of light,” *Science*, vol. 358, no. 6365, pp. 896–901, 2017.
- [11] C. He, J. Chang, Q. Hu, J. Wang, J. Antonello, H. He, S. Liu, J. Lin, B. Dai, D. S. Elson, P. Xi, H. Ma, and M. J. Booth, “Complex vectorial optics through gradient index lens cascades,” *Nature Communications*, vol. 10, no. 1, p. 4264, 2019.
- [12] K. I. Willig, S. O. Rizzoli, V. Westphal, R. Jahn, and S. W. Hell, “STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis,” *Nature*, vol. 440, no. 7086, pp. 935–939, 2006.
- [13] K. Wicker and R. Heintzmann, “Single-shot optical sectioning using polarization-coded structured illumination,” *Journal of Optics*, vol. 12, p. 084010, jul 2010.
- [14] M. Žurauskas, I. M. Dobbie, R. M. Parton, M. A. Phillips, A. Göhler, I. Davis, and M. J. Booth, “Isosense: frequency enhanced sensorless adaptive optics through structured illumination,” *Optica*, vol. 6, pp. 370–379, Mar 2019.
- [15] R. Chipman, “Polarization aberrations,” 09 1987.

- [16] A. Ariyawansa, K. Liang, and T. G. Brown, “Polarization singularities in a stress-engineered optic,” *J. Opt. Soc. Am. A*, vol. 36, pp. 312–319, Mar 2019.
- [17] A. Sommerfeld, “Chapter I - reflection and refraction of light,” in *Optics* (A. Sommerfeld, ed.), pp. 6–59, Academic Press, 1964.
- [18] A. Sommerfeld, “Chapter V - the theory of diffraction,” in *Optics* (A. Sommerfeld, ed.), pp. 179–272, Academic Press, 1964.
- [19] M. Born, E. Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, and W. L. Wilcock, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 7 ed., 1999.
- [20] T. Wilson and C. Sheppard, *Theory And Practice Of Scanning Optical Microscopy*, vol. 1. London: Academic Press, 10 1984.
- [21] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*. Basic Books, 2015.
- [22] T. Kato, “On the adiabatic theorem of quantum mechanics,” *Journal of the Physical Society of Japan*, vol. 5, pp. 435–439, 2021/04/06 1950.
- [23] S. Pancharatnam, “Generalized theory of interference, and its applications,” *Proceedings of the Indian Academy of Sciences - Section A*, vol. 44, no. 5, pp. 247–262, 1956.
- [24] H. C. Longuet-Higgins, U. Opik, M. H. L. Pryce, and R. A. Sack, “Studies of the Jahn-Teller Effect. II. The Dynamical Problem,” *Proceedings of the Royal Society of London Series A*, vol. 244, pp. 1–16, Feb. 1958.

- [25] M. V. Berry, “Quantal Phase Factors Accompanying Adiabatic Changes,” *Proceedings of the Royal Society of London Series A*, vol. 392, pp. 45–57, Mar 1984.
- [26] M. Berry, “The Adiabatic Phase and Pancharatnam’s Phase for Polarized Light,” *Journal of Modern Optics*, vol. 34, no. 11, pp. 1401–1407, 1987.
- [27] C. Huygens, *Traité de la Lumière: Où sont expliquées les causes de ce qui lui arrive dans la reflexion, & amp; dans la refraction et particulièrement dans l’/’etrange réfraction du cristal d’Islande*. A Leide: Chez Pierre vander Aa, marchand libraire, 1690.
- [28] N. Lockyer, *Nature*. No. v. 14 in Marilee E. Thomas and Robert C. Thomas Science and Related Subjects Collection, Macmillan Journals Limited, 1876.
- [29] R. Hooke, *Micrographia : or, Some physiological descriptions of minute bodies made by magnifying glasses. With observations and inquiries thereupon*. London, England: J. Allestry, 1665.
- [30] A. Van Leewenhoeck, “Observations, communicated to the publisher by mr. antony van leewenhoeck, in a dutch letter of the 9th of octob. 1676. here english’d: concerning little animals by him observed in rain-well-sea. and snow water; as also in water wherein pepper had lain infused,” *Philosophical Transactions (1665-1678)*, vol. 12, pp. 821–831.
- [31] A. Van Leewenhoeck, “Wrote letter 39 of 1683-09-17 (ab 76) to Francis Aston,”
- [32] N. Monardes, *Dos libros, el uno que trata de todas las cosas que traen de nuestras Indias Occidentales, que sirven al uso de la medicina, y el otro que trata de la piedra Bezaar y de la yerva escuerconera*. Diaz, 1565.

- [33] J. F. W. Herschel, "On a case of superficial colour presented by a homogeneous liquid internally colourless," *Philosophical Transactions of the Royal Society of London*, no. 135, pp. 143–145, 1845.
- [34] F. Rost, *Fluorescence Microscopy*. No. v. 1 in Fluorescence Microscopy 2 Volume Hardback Set, Cambridge University Press, 1992.
- [35] A. Baeyer, *Über ein neue Klasse von Farbstoffen*, vol. 11. Verlag Chemie, 1871.
- [36] O. Shimomura, F. H. Johnson, and Y. Saiga, "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusa, aequorea," *Journal of Cellular and Comparative Physiology*, vol. 59, no. 3, pp. 223–239, 1962.
- [37] D. Prasher, R. O. McCann, and M. J. Cormier, "Cloning and expression of the cDNA coding for aequorin, a bioluminescent calcium-binding protein," *Biochemical and Biophysical Research Communications*, vol. 126, no. 3, pp. 1259–1268, 1985.
- [38] M. Chalfie, Y. Tu, G. Euskirchen, W. Ward, and D. Prasher, "Green fluorescent protein as a marker for gene expression," *Science*, vol. 263, p. 802, 02 1994.
- [39] O. Heimstädt, "Das Fluoreszenzmikroskop," *Z Wiss Mikrosk*, vol. 28, pp. 330–337, 1911.
- [40] P. Ellinger and A. Hirt, "Mikroskopische Beobachtungen an lebenden Organen mit Demonstrationen (Intravitalmikroskopie)," *Arch. Exp. Pathol. Phar.*, vol. 147, 1929.
- [41] J. Ploem, "The use of a vertical illuminator with interchangeable dichroic mirrors for fluorescence microscopy with incidental light," *Zeitschrift für wis-*

- senschaftliche Mikroskopie und mikroskopische Technik*, vol. 68, pp. 129–142, November 1967.
- [42] M. Minsky, “Microscopy apparatus,” December 1961.
- [43] M. D. Egger and M. Petran, “New reflected-light microscope for viewing unstained brain and ganglion cells,” *Science*, vol. 157, p. 305, 07 1967.
- [44] C. J. R. Sheppard and T. Wilson, “The theory of the direct-view confocal microscope,” *Journal of Microscopy*, vol. 124, no. 2, pp. 107–117, 1981.
- [45] M. Göppert-Mayer, “Über Elementarakte mit zwei Quantensprüngen,” *Annalen der Physik*, vol. 401, no. 3, pp. 273–294, 1931.
- [46] W. Kaiser and C. G. B. Garrett, “Two-Photon Excitation in $\text{CaF}_2: \text{Eu}^{2+}$,” *Phys. Rev. Lett.*, vol. 7, pp. 229–231, Sep 1961.
- [47] W. Denk, J. Strickler, and W. Webb, “Two-photon laser scanning fluorescence microscopy,” *Science*, vol. 248, no. 4951, pp. 73–76, 1990.
- [48] C. Xu, W. Zipfel, J. B. Shear, R. M. Williams, and W. W. Webb, “Multiphoton fluorescence excitation: new spectral windows for biological nonlinear microscopy,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, pp. 10763–10768, 10 1996.
- [49] N. G. Horton, K. Wang, D. Kobat, C. G. Clark, F. W. Wise, C. B. Schaffer, and C. Xu, “In vivo three-photon microscopy of subcortical structures within an intact mouse brain,” *Nature Photonics*, vol. 7, no. 3, pp. 205–209, 2013.
- [50] C. J. Rowlands, D. Park, O. T. Bruns, K. D. Piatkevich, D. Fukumura, R. K. Jain, M. G. Bawendi, E. S. Boyden, and P. T. So, “Wide-field three-photon excitation in biological samples,” *Light: Science & Applications*, vol. 6, no. 5, pp. e16255–e16255, 2017.

- [51] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich, "Generation of optical harmonics," *Phys. Rev. Lett.*, vol. 7, pp. 118–119, Aug 1961.
- [52] S. Fine and W. P. Hansen, "Optical second harmonic generation in biological systems," *Appl. Opt.*, vol. 10, pp. 2350–2353, Oct. 1971.
- [53] R. Hellwarth and P. Christensen, "Nonlinear optical microscopic examination of structure in polycrystalline ZnSe," *Optics Communications*, vol. 12, pp. 318–322, Nov. 1974.
- [54] I. Freund and M. Deutsch, "Second-harmonic microscopy of biological tissue," *Optics Letters*, vol. 11, pp. 94–96, Feb. 1986.
- [55] P. D. Maker and R. W. Terhune, "Study of optical effects due to an induced polarization third order in the electric field strength," *Phys. Rev.*, vol. 137, pp. A801–A818, Feb 1965.
- [56] R. F. Begley, A. B. Harvey, and R. L. Byer, "Coherent anti-stokes raman spectroscopy," *Applied Physics Letters*, vol. 25, no. 7, pp. 387–390, 1974.
- [57] S. Hell and E. H. Stelzer, "Fundamental improvement of resolution with a 4pi-confocal fluorescence microscope using two-photon excitation," *Optics Communications*, vol. 93, no. 5, pp. 277–282, 1992.
- [58] M. Schrader, K. Bahlmann, G. Giese, and S. W. Hell, "4pi-confocal imaging in fixed biological specimens," *Biophysical Journal*, vol. 75, no. 4, pp. 1659–1668, 1998.
- [59] K. Bahlmann, S. Jakobs, and S. W. Hell, "4pi-confocal microscopy of live cells," *Ultramicroscopy*, vol. 87, no. 3, pp. 155–164, 2001.
- [60] P. E. Hänninen, S. W. Hell, J. Salo, E. Soini, and C. Cremer, "Two-photon excitation 4pi confocal microscope: Enhanced axial resolution microscope for

- biological research,” *Applied Physics Letters*, vol. 66, no. 13, pp. 1698–1700, 1995.
- [61] R. Schmidt, C. A. Wurm, S. Jakobs, J. Engelhardt, A. Egner, and S. W. Hell, “Spherical nanosized focal spot unravels the interior of cells,” *Nature Methods*, vol. 5, no. 6, pp. 539–544, 2008.
- [62] S. Liu and F. Huang, “Enhanced 4pi single-molecule localization microscopy with coherent pupil based localization,” *Communications Biology*, vol. 3, no. 1, p. 220, 2020.
- [63] S. W. Hell and J. Wichmann, “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy,” *Opt. Lett.*, vol. 19, pp. 780–782, Jun 1994.
- [64] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, “Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission,” *Proceedings of the National Academy of Sciences*, vol. 97, no. 15, pp. 8206–8210, 2000.
- [65] M. G. L. Gustafsson, “Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 37, pp. 13081–13086, 2005.
- [66] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, “Imaging intracellular fluorescent proteins at nanometer resolution,” *Science*, vol. 313, p. 1642, 09 2006.
- [67] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, “Ultra-high resolution imaging by fluorescence photoactivation localization microscopy,” *Biophysical Journal*, vol. 91, pp. 4258–4272, 2021/03/22 2006.

- [68] M. J. Rust, M. Bates, and X. Zhuang, “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),” *Nature Methods*, vol. 3, no. 10, pp. 793–796, 2006.
- [69] X. Hao, J. Antonello, E. S. Allgeyer, J. Bewersdorf, and M. J. Booth, “Aberrations in 4pi microscopy,” *Opt. Express*, vol. 25, pp. 14049–14058, Jun 2017.
- [70] T. J. Gould, D. Burke, J. Bewersdorf, and M. J. Booth, “Adaptive optics enables 3D STED microscopy in aberrating specimens,” *Opt. Express*, vol. 20, pp. 20998–21009, sep 2012.
- [71] A. Barbotin, S. Galiani, I. Urbančič, C. Eggeling, and M. J. Booth, “Adaptive optics allows STED-FCS measurements in the cytoplasm of living cells,” *Opt. Express*, vol. 27, pp. 23378–23395, Aug 2019.
- [72] C. Sheppard and C. Cogswell, “Effects of aberrating layers and tube length on confocal imaging properties,” *Optik - International Journal for Light and Electron Optics*, vol. 87, pp. 34–38, 01 1991.
- [73] A. Egner and S. W. Hell, “Equivalence of the huygens–fresnel and debye approach for the calculation of high aperture point-spread functions in the presence of refractive index mismatch,” *Journal of Microscopy*, vol. 193, no. 3, pp. 244–249, 1999.
- [74] M. Schwertner, M. J. Booth, M. A. A. Neil, and T. Wilson, “Measurement of specimen-induced aberrations of biological samples using phase stepping interferometry,” *Journal of Microscopy*, vol. 213, no. 1, pp. 11–19, 2004.
- [75] M. Schwertner, M. Booth, and T. Wilson, “Characterizing specimen induced aberrations for high NA adaptive optical microscopy,” *Optics Express*, vol. 12, p. 6540, dec 2004.

- [76] “Focus tunable lenses - Optotune.” <https://www.optotune.com/focus-tunable-lenses>. Accessed: 2021-03-27.
- [77] “Projektübersicht - Institut für Mikrosystemtechnik - IMTEK.” <https://www.imtek.de/forschung/projektuebersicht/projektuebersicht?projectId=10963>. Accessed: 2021-03-27.
- [78] K. Banerjee, P. Rajaeipour, Çağlar Ataman, and H. Zappe, “Optofluidic adaptive optics,” *Appl. Opt.*, vol. 57, pp. 6338–6344, Aug 2018.
- [79] M. A. Ealey and J. Wellman, “Fundamentals Of Deformable Mirror Design And Analysis,” in *Precision Engineering and Optomechanics* (D. Vukobratovich, ed.), vol. 1167, pp. 66 – 84, International Society for Optics and Photonics, SPIE, 1989.
- [80] Z. Zhang, Z. You, and D. Chu, “Fundamentals of phase-only liquid crystal on silicon (lcos) devices,” *Light: Science & Applications*, vol. 3, no. 10, pp. e213–e213, 2014.
- [81] J. Hartmann, *Zeitschrift für Instrumentenkunde*, pp. 1, 33, 97. Springer, 1904.
- [82] R. V. Shack and B. C. Platt, “Production and use of a lenticular hartmann screen,” *J. Opt. Soc. Am.*, vol. 61, p. 656, May 1971.
- [83] J. Antonello, J. Wang, C. He, M. Phillips, and M. Booth, “Interferometric calibration of a deformable mirror.” <https://doi.org/10.5281/zenodo.3714951>, Mar 2020.
- [84] F. von Zernike, “Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode,” *Physica*, vol. 1, no. 7, pp. 689–704, 1934.

- [85] R. J. Noll, “Zernike polynomials and atmospheric turbulence,” *Journal of the Optical Society of America (1917-1983)*, vol. 66, pp. 207–211, Mar. 1976.
- [86] T. Mitchell, *Machine Learning*. McGraw-Hill international editions - computer science series, McGraw-Hill Education, 1997.
- [87] E. Alpaydin, *Introduction to Machine Learning*. Adaptive Computation and Machine Learning series, MIT Press, 2020.
- [88] J. B. Pawley, “Limitations on optical sectioning in live-cell confocal microscopy,” *Scanning: The Journal of Scanning Microscopies*, vol. 24, no. 5, pp. 241–246, 2002.
- [89] M. Schwertner, M. J. Booth, and T. Wilson, “Simple optimization procedure for objective lens correction collar setting,” *Journal of Microscopy*, vol. 217, no. 3, pp. 184–187, 2005.
- [90] H. W. Yoo, M. E. van Royen, W. A. van Cappellen, A. B. Houtsmuller, M. Verhaegen, and G. Schitter, “Automated spherical aberration correction in scanning confocal microscopy,” *Review of Scientific Instruments*, vol. 85, no. 12, p. 123706, 2014.
- [91] R. A. Muller and A. Buffington, “Real-time correction of atmospherically degraded telescope images through image sharpening,” *J. Opt. Soc. Am.*, vol. 64, pp. 1200–1210, Sep 1974.
- [92] P. Sprawls, *The Physical Principles of Medical Imaging*. Medical Physics Publishing Corporation, 2nd ed., May 1995.
- [93] R. Ferzli and L. J. Karam, “A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB),” *IEEE Transactions on Image Processing*, vol. 18, pp. 717–728, April 2009.

- [94] M. J. Booth, “Adaptive optical microscopy: the ongoing quest for a perfect image,” *Light: Science & Applications*, vol. 3, no. 4, pp. e165–e165, 2014.
- [95] M. J. Booth and B. R. Patton, “Adaptive Optics for Fluorescence Microscopy,” in *Fluorescence Microscopy: Super-Resolution and other Novel Techniques* (A. Cornea and P. M. Conn, eds.), pp. 15–33, Boston: Academic Press, 2014.
- [96] N. Ji, “Adaptive optical fluorescence microscopy,” *Nature Methods*, vol. 14, no. 4, pp. 374–380, 2017.
- [97] M. Booth, D. Andrade, D. Burke, B. Patton, and M. Zurauskas, “Aberrations and adaptive optics in super-resolution microscopy,” *Microscopy*, vol. 64, pp. 251–261, 06 2015.
- [98] B. M. Hanser, M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, “Phase-retrieved pupil functions in wide-field fluorescence microscopy,” *Journal of Microscopy*, vol. 216, no. 1, pp. 32–48, 2004.
- [99] D. Débarre, M. J. Booth, and T. Wilson, “Image based adaptive optics through optimisation of low spatial frequencies,” *Opt. Express*, vol. 15, pp. 8176–8190, Jun 2007.
- [100] M. Warber, S. Maier, T. Haist, and W. Osten, “Combination of scene-based and stochastic measurement for wide-field aberration correction in microscopic imaging,” *Appl. Opt.*, vol. 49, pp. 5474–5479, Oct 2010.
- [101] P. Kner, L. Winoto, D. A. Agard, and J. W. Sedat, “Closed loop adaptive optics for microscopy without a wavefront sensor,” in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XVII* (J.-A. Conchello, C. J. Cogswell, T. Wilson, and T. G. Brown, eds.), vol. 7570, pp. 40 – 48, International Society for Optics and Photonics, SPIE, 2010.

- [102] C. Stockbridge, Y. Lu, J. Moore, S. Hoffman, R. Paxman, K. Toussaint, and T. Bifano, “Focusing through dynamic scattering media,” *Opt. Express*, vol. 20, pp. 15086–15092, Jul 2012.
- [103] C. Chang, L. Cheng, H. Su, W. Yen, and S.-J. Chen, “Widefield multiphoton microscopy with image-based adaptive optics,” in *Unconventional Imaging and Wavefront Sensing 2012*, Proceedings of SPIE - The International Society for Optical Engineering, Dec. 2012. null ; Conference date: 13-08-2012 Through 14-08-2012.
- [104] J. Scrimgeour and J. E. Curtis, “Aberration correction in wide-field fluorescence microscopy by segmented-pupil image interferometry,” *Opt. Express*, vol. 20, pp. 14534–14541, June 2012.
- [105] P. Kner, “Phase diversity for three-dimensional imaging,” *J. Opt. Soc. Am. A*, vol. 30, pp. 1980–1987, Oct 2013.
- [106] Y. Jin, Y. Zhang, L. Hu, H. Huang, Q. Xu, X. Zhu, L. Huang, Y. Zheng, H.-L. Shen, W. Gong, and K. Si, “Machine learning guided rapid focusing with sensor-less aberration corrections,” *Opt. Express*, vol. 26, pp. 30162–30171, Nov 2018.
- [107] D. Saha, U. Schmidt, Q. Zhang, A. Barbotin, Q. Hu, N. Ji, M. J. Booth, M. Weigert, and E. W. Myers, “Practical sensorless aberration estimation for 3D microscopy with deep learning,” *Opt. Express*, vol. 28, pp. 29044–29053, Sep 2020.
- [108] M. J. Booth, M. A. A. Neil, R. Juškaitis, and T. Wilson, “Adaptive aberration correction in a confocal microscope,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5788–5792, 2002.

- [109] M. J. Booth, M. A. A. Neil, and T. Wilson, “New modal wave-front sensor: application to adaptive confocal fluorescence microscopy and two-photon excitation fluorescence microscopy,” *J. Opt. Soc. Am. A*, vol. 19, pp. 2112–2120, Oct 2002.
- [110] A. J. Wright, D. Burns, B. A. Patterson, S. P. Poland, G. J. Valentine, and J. M. Girkin, “Exploration of the optimisation algorithms used in the implementation of adaptive optics in confocal and multiphoton microscopy,” *Microscopy Research and Technique*, vol. 67, pp. 36–44, 5 2005.
- [111] M. Booth, T. Wilson, H.-B. Sun, T. Ota, and S. Kawata, “Methods for the characterization of deformable membrane mirrors,” *Appl. Opt.*, vol. 44, pp. 5131–5139, Aug 2005.
- [112] S. P. Poland, A. J. Wright, and J. M. Girkin, “Evaluation of fitness parameters used in an iterative approach to aberration correction in optical sectioning microscopy,” *Appl. Opt.*, vol. 47, pp. 731–736, Feb 2008.
- [113] W. Lubeigt, S. P. Poland, G. J. Valentine, A. J. Wright, J. M. Girkin, and D. Burns, “Search-based active optic systems for aberration correction in time-independent applications,” *Appl. Opt.*, vol. 49, pp. 307–314, Jan 2010.
- [114] C.-E. Leroux, I. Wang, J. Derouard, and A. Delon, “Adaptive optics for fluorescence correlation spectroscopy,” *Opt. Express*, vol. 19, pp. 26839–26849, Dec 2011.
- [115] W. M. Lee and S. H. Yun, “Adaptive aberration correction of GRIN lenses for confocal endomicroscopy,” *Opt. Lett.*, vol. 36, pp. 4608–4610, Dec 2011.
- [116] J. Antonello, M. Verhaegen, R. Fraanje, T. van Werkhoven, H. C. Gerritsen, and C. U. Keller, “Semidefinite programming for model-based sensorless adaptive optics,” *J. Opt. Soc. Am. A*, vol. 29, pp. 2428–2438, Nov 2012.

- [117] C.-E. Leroux, A. Grichine, I. Wang, and A. Delon, “Correction of cell-induced optical aberrations in a fluorescence fluctuation microscope,” *Opt. Lett.*, vol. 38, pp. 2401–2403, July 2013.
- [118] X. Zhang and P. Kner, “Binary wavefront optimization using a genetic algorithm,” *Journal of Optics*, vol. 16, p. 125704, Oct 2014.
- [119] X. Tao, D. Bodington, M. Reinig, and J. Kubby, “High-speed scanning interferometric focusing by fast measurement of binary transmission matrix for channel demixing,” *Opt. Express*, vol. 23, pp. 14168–14187, Jun 2015.
- [120] I. Vishniakou and J. D. Seelig, “Wavefront correction for adaptive optics with reflected light and deep neural networks,” *Opt. Express*, vol. 28, pp. 15459–15471, May 2020.
- [121] L. Sherman, J. Y. Ye, O. Albert, and T. B. Norris, “Adaptive correction of depth-induced aberrations in multiphoton scanning microscopy using a deformable mirror,” *Journal of Microscopy*, vol. 206, no. 1, pp. 65–71, 2002.
- [122] P. N. Marsh, D. Burns, and J. M. Girkin, “Practical implementation of adaptive optics in multiphoton microscopy,” *Opt. Express*, vol. 11, pp. 1123–1130, May 2003.
- [123] N. Olivier, D. Débarre, and E. Beaurepaire, “Dynamic aberration correction for multiharmonic microscopy,” *Opt. Lett.*, vol. 34, pp. 3145–3147, Oct 2009.
- [124] D. Débarre, E. J. Botcherby, T. Watanabe, S. Srinivas, M. J. Booth, and T. Wilson, “Image-based adaptive optics for two-photon microscopy,” *Opt. Lett.*, vol. 34, pp. 2495–2497, Aug 2009.

- [125] N. Ji, D. E. Milkie, and E. Betzig, “Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues,” *Nature Methods*, vol. 7, pp. 141–147, Apr 2009.
- [126] D. E. Milkie, E. Betzig, and N. Ji, “Pupil-segmentation-based adaptive optical microscopy with full-pupil illumination,” *Opt. Lett.*, vol. 36, pp. 4206–4208, Nov 2011.
- [127] F. Bortoletto, C. Bonoli, P. Panizzolo, C. D. Ciubotaru, and F. Mammano, “Multiphoton fluorescence microscopy with GRIN objective aberration correction by low order adaptive optics,” *PLOS ONE*, vol. 6, pp. 1–8, 07 2011.
- [128] J. Tang, R. N. Germain, and M. Cui, “Superpenetration optical microscopy by iterative multiphoton adaptive compensation technique,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8434–8439, 2012.
- [129] A. Facomprez, E. Beaupaire, and D. Débarre, “Accuracy of correction in modal sensorless adaptive optics,” *Optics Express*, vol. 20, p. 2598, Jan 2012.
- [130] N. Ji, T. R. Sato, and E. Betzig, “Characterization and adaptive optical correction of aberrations during in vivo imaging in the mouse cortex,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 22–27, Jan 2012.
- [131] C. Wang and N. Ji, “Pupil-segmentation-based adaptive optical correction of a high-numerical-aperture gradient refractive index lens for two-photon fluorescence endoscopy,” *Optics Letters*, vol. 37, p. 2001, June 2012.
- [132] D. B. Conkey, A. M. Caravaca-Aguirre, and R. Piestun, “High-speed scattering medium characterization with application to focusing light through turbid media,” *Opt. Express*, vol. 20, pp. 1733–1740, Jan 2012.

- [133] C. Wang and N. Ji, "Characterization and improvement of three-dimensional imaging performance of GRIN-lens-based two-photon fluorescence endomicroscopes with adaptive optics," *Opt. Express*, vol. 21, pp. 27142–27154, Nov 2013.
- [134] O. Katz, E. Small, Y. Guan, and Y. Silberberg, "Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers," *Optica*, vol. 1, pp. 170–174, Sep 2014.
- [135] R. Liu, D. E. Milkie, A. Kerlin, B. MacLennan, and N. Ji, "Direct phase measurement in zonal wavefront reconstruction using multidither coherent optical adaptive technique," *Opt. Express*, vol. 22, pp. 1619–1628, Jan 2014.
- [136] C. Wang, R. Liu, D. E. Milkie, W. Sun, Z. Tan, A. Kerlin, T.-W. Chen, D. S. Kim, and N. Ji, "Multiplexed aberration measurement for deep tissue imaging in vivo," *Nature Methods*, vol. 11, no. 10, pp. 1037–1040, 2014.
- [137] L. Kong and M. Cui, "In vivo neuroimaging through the highly scattering tissue via iterative multi-photon adaptive compensation technique," *Opt. Express*, vol. 23, pp. 6145–6150, Mar 2015.
- [138] D. Sinefeld, H. P. Paudel, D. G. Ouzounov, T. G. Bifano, and C. Xu, "Adaptive optics in multiphoton microscopy: comparison of two, three and four photon fluorescence," *Opt. Express*, vol. 23, pp. 31472–31483, Nov 2015.
- [139] P. T. Galwaduge, S. H. Kim, L. E. Grosberg, and E. M. C. Hillman, "Simple wavefront correction framework for two-photon microscopy of in-vivo brain," *Biomed. Opt. Express*, vol. 6, pp. 2997–3013, Aug 2015.
- [140] J. H. Park, W. Sun, and M. Cui, "High-resolution in vivo imaging of mouse brain through the intact skull," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, pp. 9236–9241, July 2015.

- [141] M. Skorsetz, P. Artal, and J. M. Bueno, “Performance evaluation of a sensorless adaptive optics multiphoton microscope,” *Journal of Microscopy*, vol. 261, no. 3, pp. 249–258, 2016.
- [142] X. Tao, T. Lam, B. Zhu, Q. Li, M. R. Reinig, and J. Kubby, “Three-dimensional focusing through scattering media using conjugate adaptive optics with remote focusing (CAORF),” *Opt. Express*, vol. 25, pp. 10368–10383, May 2017.
- [143] I. N. Papadopoulos, J.-S. Jouhanneau, J. F. A. Poulet, and B. Judkewitz, “Scattering compensation by focus scanning holographic aberration probing (F-SHARP),” *Nature Photonics*, vol. 11, no. 2, pp. 116–123, 2017.
- [144] C. Rodríguez, A. Chen, J. Rivera, M. A. Mohr, Y. Liang, W. Sun, D. E. Milkie, T. G. Bifano, X. Chen, and N. Ji, “An adaptive optics module for deep tissue multiphoton imaging *in vivo*,” *bioRxiv*, p. 2020.11.25.397968, Nov 2020.
- [145] I. N. Papadopoulos, J.-S. Jouhanneau, N. Takahashi, D. Kaplan, M. Larkum, J. Poulet, and B. Judkewitz, “Dynamic conjugate F-SHARP microscopy,” *Light: Science & Applications*, vol. 9, no. 1, p. 110, 2020.
- [146] L. Streich, J. Boffi, L. Wang, K. Alhalaseh, M. Barbieri, R. Rehm, S. Deivasigamani, C. Gross, A. Agarwal, and R. Prevedel, “High-resolution structural and functional deep brain imaging using adaptive optics three-photon microscopy,” *bioRxiv*, p. 2021.01.12.426323, Jan 2021.
- [147] A. Jesacher, A. Thayil, K. Grieve, D. Débarre, T. Watanabe, T. Wilson, S. Srinivas, and M. Booth, “Adaptive harmonic generation microscopy of mammalian embryos,” *Opt. Lett.*, vol. 34, pp. 3154–3156, Oct 2009.
- [148] C. Bourgenot, C. D. Saunter, J. M. Taylor, J. M. Girkin, and G. D. Love, “3D adaptive optics in a light sheet microscope,” *Opt. Express*, vol. 20, pp. 13252–13261, Jun 2012.

- [149] F. Wang, “Wavefront sensing through measurements of binary aberration modes,” *Applied Optics*, vol. 48, pp. 2865–2870, May 2009.
- [150] A. J. Wright, S. P. Poland, J. M. Girkin, C. W. Freudiger, C. L. Evans, and X. S. Xie, “Adaptive optics for enhanced signal in CARS microscopy,” *Opt. Express*, vol. 15, pp. 18209–18219, Dec 2007.
- [151] R. Fiolka, K. Si, and M. Cui, “Complex wavefront corrections for deep tissue focusing using low coherence backscattered light,” *Opt. Express*, vol. 20, pp. 16532–16543, Jul 2012.
- [152] D. Débarre, E. J. Botcherby, M. J. Booth, and T. Wilson, “Adaptive optics for structured illumination microscopy,” *Opt. Express*, vol. 16, pp. 9290–9305, Jun 2008.
- [153] B. Thomas, A. Wolstenholme, S. N. Chaudhari, E. T. Kipreos, and P. Kner, “Enhanced resolution through thick tissue with structured illumination and adaptive optics,” *Journal of Biomedical Optics*, vol. 20, no. 2, pp. 1 – 8, 2015.
- [154] B. R. Patton, D. Burke, D. Oswald, T. J. Gould, J. Bewersdorf, and M. J. Booth, “Three-dimensional STED microscopy of aberrating tissue using dual adaptive optics,” *Opt. Express*, vol. 24, pp. 8862–8876, Apr 2016.
- [155] J. Antonello, A. Barbotin, E. Z. Chong, J. Rittscher, and M. J. Booth, “Multi-scale sensorless adaptive optics: application to stimulated emission depletion microscopy,” *Opt. Express*, vol. 28, pp. 16749–16763, May 2020.
- [156] I. Izeddin, M. E. Beheiry, J. Andilla, D. Ciepielewski, X. Darzacq, and M. Dahan, “PSF shaping using adaptive optics for three-dimensional single-molecule super-resolution imaging and tracking,” *Opt. Express*, vol. 20, pp. 4957–4967, Feb 2012.

- [157] C. Roider, A. Jesacher, S. Bernet, and M. Ritsch-Marte, “Axial super-localisation using rotating point spread functions shaped by polarisation-dependent phase modulation,” *Opt. Express*, vol. 22, pp. 4029–4037, Feb 2014.
- [158] D. Burke, B. Patton, F. Huang, J. Bewersdorf, and M. J. Booth, “Adaptive optics correction of specimen-induced aberrations in single-molecule switching microscopy,” *Optica*, vol. 2, pp. 177–185, Feb 2015.
- [159] F. Huang, G. Sirinakis, E. S. Allgeyer, L. K. Schroeder, W. C. Duim, E. B. Kromann, T. Phan, F. E. Rivera-Molina, J. R. Myers, I. Irnov, M. Lessard, Y. Zhang, M. A. Handel, C. Jacobs-Wagner, C. P. Lusk, J. E. Rothman, D. Toomre, M. J. Booth, and J. Bewersdorf, “Ultra-high resolution 3D imaging of whole cells,” *Cell*, vol. 166, pp. 1028–1040, 2021/02/19 2016.
- [160] “Implementation of a 4pi-SMS super-resolution microscope,” *Nature Protocols*, vol. 16, no. 2, pp. 677–727, 2021.
- [161] B. Wang and M. J. Booth, “Optimum deformable mirror modes for sensorless adaptive optics,” *Optics Communications*, vol. 282, pp. 4467–4474, 2009.
- [162] K. Strehl, “Aplanatische und fehlerhafte Abbildung im Fernrohr,” *Zeitschrift für Instrumentenkunde*, vol. 15, no. 7, pp. 362–370, 1895.
- [163] K. Strehl, “Über Luftschlieren und Zonenfehler,” *Zeitschrift für Instrumentenkunde*, vol. 22, pp. 213–217, 1902.
- [164] O. Soloviev, “Alias-free basis for modal sensorless adaptive optics using the second moment of intensity,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 0, no. 0, p. 2040008, 2020.

- [165] L. Möckl, P. N. Petrov, and W. E. Moerner, “Accurate phase retrieval of complex 3D point spread functions with deep residual neural networks,” *Applied Physics Letters*, vol. 115, no. 25, p. 251106, 2019.
- [166] B. P. Cumming and M. Gu, “Direct determination of aberration functions in microscopy by an artificial neural network,” *Opt. Express*, vol. 28, pp. 14511–14521, May 2020.
- [167] G. Vdovin, P. M. Sarro, and S. Middelhoek, “Technology and applications of micromachined adaptive mirrors,” *Journal of Micromechanics and Microengineering*, vol. 9, pp. R8–R20, Jan 1999.
- [168] L. Huang, C. Rao, and W. Jiang, “Modified Gaussian influence function of deformable mirror actuators,” *Opt. Express*, vol. 16, pp. 108–114, Jan 2008.
- [169] B. Roopashree, A. Vyas, and P. Budihal, “A novel model of influence function: calibration of a continuous membrane deformable mirror,” *International Journal on Control System and Instrumentation*, vol. 3, pp. 10–14, 01 2011.
- [170] L. Zhu, P.-C. Sun, D.-U. Bartsch, W. R. Freeman, and Y. Fainman, “Wavefront generation of Zernike polynomial modes with a micromachined membrane deformable mirror,” *Appl. Opt.*, vol. 38, pp. 6019–6026, Oct 1999.
- [171] T. Ota, S. Kawata, T. Sugiura, M. J. Booth, M. A. A. Neil, R. Juškaitis, and T. Wilson, “Dynamic axial-position control of a laser-trapped particle by wavefront modification,” *Opt. Lett.*, vol. 28, pp. 465–467, Mar 2003.
- [172] E. Bostan, R. Heckel, M. Chen, M. Kellman, and L. Waller, “Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network,” *Optica*, vol. 7, pp. 559–562, Jun 2020.

- [173] W. Lukosz, “Der Einfluß der Aberrationen auf die Optische Übertragungsfunktion bei Kleinen Orts-Frequenzen,” *Optica Acta: International Journal of Optics*, vol. 10, no. 1, pp. 1–19, 1963.
- [174] M. J. Booth, “Wavefront sensorless adaptive optics for large aberrations,” *Opt. Lett.*, vol. 32, pp. 5–7, Jan 2007.
- [175] Q. Hu, J. Wang, J. Antonello, M. Hailstone, M. Wincott, R. Turcotte, D. Gala, and M. J. Booth, “A universal framework for microscope sensorless adaptive optics: Generalized aberration representations,” *APL Photonics*, vol. 5, no. 10, p. 100801, 2020.
- [176] F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, and G. Situ, “Phase imaging with an untrained neural network,” *Light: Science & Applications*, vol. 9, no. 1, p. 77, 2020.
- [177] O. Albert, L. Sherman, G. Mourou, T. B. Norris, and G. Vdovin, “Smart microscope: an adaptive optics learning system for aberration correction in multiphoton confocal microscopy,” *Opt. Lett.*, vol. 25, pp. 52–54, Jan 2000.
- [178] R. K. Benninger and D. W. Piston, “Two-photon excitation microscopy for the study of living cells and tissues,” *Current Protocols in Cell Biology*, vol. 59, no. 1, pp. 4.11.1–4.11.24, 2013.
- [179] M. Gu and C. J. R. Sheppard, “Comparison of three-dimensional imaging properties between two-photon and single-photon fluorescence microscopy,” *Journal of Microscopy*, vol. 177, no. 2, pp. 128–137, 1995.
- [180] M. J. Booth and T. Wilson, “Refractive-index-mismatch induced aberrations in single-photon and two-photon microscopy and the use of aberration correction,” *Journal of Biomedical Optics*, vol. 6, pp. 6 – 6 – 7, 2001.

- [181] R. D. Simmonds, T. Wilson, and M. J. Booth, “Effects of aberrations and specimen structure in conventional, confocal and two-photon fluorescence microscopy,” *Journal of Microscopy*, vol. 245, no. 1, pp. 63–71, 2012.
- [182] J. Pawley, *Handbook of Biological Confocal Microscopy*. Springer US, 2006.
- [183] M. M. Frigault, J. Lacoste, J. L. Swift, and C. M. Brown, “Live-cell microscopy - Tips and tools,” *Journal of Cell Science*, vol. 122, pp. 753–767, mar 2009.
- [184] E. J. Borowski and J. M. Borwein, *Dictionary of Mathematics*. Collins, 1989.
- [185] R. J. Noll, “Zernike polynomials and atmospheric turbulence,” *J. Opt. Soc. Am.*, vol. 66, pp. 207–211, Mar 1976.
- [186] M. Hailstone, D. Waithe, T. J. Samuels, L. Yang, I. Costello, Y. Arava, E. Robertson, R. M. Parton, and I. Davis, “CytoCensus, mapping cell identity and division in tissues and organs using machine learning,” *eLife*, vol. 9, May 2020.
- [187] M. J. Booth, “Wave front sensor-less adaptive optics: a model-based approach using sphere packings,” *Opt. Express*, vol. 14, pp. 1339–1352, Feb 2006.
- [188] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
<http://www.deeplearningbook.org>.
- [189] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [190] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising: Can plain neural networks compete with BM3D?,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2392–2399, 2012.

- [191] C.-Y. Yang, C. Ma, and M.-H. Yang, “Single-image super-resolution: A benchmark,” in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), (Cham), pp. 372–386, Springer International Publishing, 2014.
- [192] S. Lefkimmiatis, “Non-local color image denoising with convolutional neural networks,” *CoRR*, vol. abs/1611.06757, 2016.
- [193] W. Lai, J. Huang, N. Ahuja, and M. Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5835–5843, 2017.
- [194] Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2790–2798, 2017.
- [195] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114, 2017.
- [196] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [197] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), (Cham), pp. 234–241, Springer International Publishing, 2015.

- [198] R. Heckel and P. Hand, “Deep decoder: Concise image representations from untrained non-convolutional networks,” *CoRR*, vol. abs/1810.03982, 2018.
- [199] J. Starck, J. Fadili, and F. Murtagh, “The undecimated wavelet decomposition and its reconstruction,” *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 297–309, 2007.
- [200] J. Starck, F. Murtagh, and M. Bertero, “Starlet transform in astronomical data processing,” in *Handbook of Mathematical Methods in Imaging* (O. Scherzer, ed.), pp. 2053–2098, Springer, 2015.
- [201] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML’10, (Madison, WI, USA), p. 807–814, Omnipress, 2010.
- [202] I. W. on Artificial Neural Networks, J. Mira, and F. Sandoval, *From Natural to Artificial Neural Computation: International Workshop on Artificial Neural Networks, Malaga-Torremolinos, Spain, June 7-9, 1995 : Proceedings*. Lecture Notes in Computer Science, U.S. Government Printing Office, 1995.
- [203] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feed-forward neural networks,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Y. W. Teh and M. Titterton, eds.), vol. 9 of *Proceedings of Machine Learning Research*, (Chia Laguna Resort, Sardinia, Italy), pp. 249–256, JMLR Workshop and Conference Proceedings, 13–15 May 2010.
- [204] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

- [205] R. Chipman, W. Lam, and G. Young, *Polarized Light and Optical Systems*. Optical Sciences and Applications of Light, CRC Press, 2018.
- [206] C. Rosales-Guzmán, B. Ndagano, and A. Forbes, “A review of complex vector light fields and their applications,” *Journal of Optics*, vol. 20, p. 123001, Nov 2018.
- [207] J. Wang, F. Castellucci, and S. Franke-Arnold, “Vectorial light–matter interaction: Exploring spatially structured complex light fields,” *AVS Quantum Science*, vol. 2, no. 3, p. 031702, 2020.
- [208] L. Marrucci, C. Manzo, and D. Paparo, “Optical spin-to-orbital angular momentum conversion in inhomogeneous anisotropic media,” *Phys. Rev. Lett.*, vol. 96, p. 163905, Apr 2006.
- [209] P. Chen, W. Ji, B.-Y. Wei, W. Hu, V. Chigrinov, and Y.-Q. Lu, “Generation of arbitrary vector beams with liquid crystal polarization converters and vector-photoaligned q-plates,” *Applied Physics Letters*, vol. 107, no. 24, p. 241102, 2015.
- [210] K. J. Moh, X.-C. Yuan, J. Bu, S. W. Zhu, and B. Z. Gao, “Radial polarization induced surface plasmon virtual probe for two-photon fluorescence microscopy,” *Opt. Lett.*, vol. 34, pp. 971–973, Apr 2009.
- [211] D. C. Hutchings, C. Zhang, B. M. Holmes, P. Dulal, A. D. Block, and B. J. H. Stadler, “Faraday polarisation mode conversion in semiconductor waveguides incorporating periodic garnet claddings,” in *Integrated Optics: Devices, Materials, and Technologies XX* (J.-E. Broquin and G. N. Conti, eds.), vol. 9750, pp. 153 – 159, International Society for Optics and Photonics, SPIE, 2016.

- [212] K. Venkatakrishnan and B. Tan, “Interconnect microvia drilling with a radially polarized laser beam,” *Journal of Micromechanics and Microengineering*, vol. 16, pp. 2603–2607, Nov 2006.
- [213] J. Chen, L. Kong, and Q. Zhan, “Demonstration of a vectorial optical field generator with adaptive close loop control,” *Review of Scientific Instruments*, vol. 88, no. 12, p. 125111, 2017.
- [214] Y. Dai, C. He, J. Wang, R. Turcotte, L. Fish, M. Wincott, Q. Hu, and M. J. Booth, “Active compensation of extrinsic polarization errors using adaptive optics,” *Opt. Express*, vol. 27, pp. 35797–35810, Nov 2019.
- [215] F. Kenny, D. Lara, O. G. Rodríguez-Herrera, and C. Dainty, “Complete polarization and phase control for focus-shaping in high-NA microscopy,” *Opt. Express*, vol. 20, pp. 14015–14029, Jun 2012.
- [216] A. Peña and M. F. Andersen, “Complete polarization and phase control with a single spatial light modulator for the generation of complex light fields,” *Laser Physics*, vol. 28, p. 076201, May 2018.
- [217] Y. Esumi, M. Kabir, and F. Kannari, “Spatiotemporal vector pulse shaping of femtosecond laser pulses with a multi-pass two-dimensional spatial light modulator,” *Opt. Express*, vol. 17, pp. 19153–19159, Oct 2009.
- [218] J. Chen, C. Wan, and Q. Zhan, “Vectorial optical fields: recent advances and future prospects,” *Science Bulletin*, vol. 63, no. 1, pp. 54 – 74, 2018.
- [219] M. T. Runyon, C. H. Nacke, A. Sit, M. Granados-Baez, L. Giner, and J. S. Lundeen, “Implementation of nearly arbitrary spatially varying polarization transformations: an in-principle lossless approach using spatial light modulators,” *Appl. Opt.*, vol. 57, pp. 5769–5778, Jul 2018.

- [220] Y. Xue, Y. Wang, S. Zhou, H. Chen, G. Rui, B. Gu, and Q. Zhan, “Focus shaping and optical manipulation using highly focused second-order full poincaré beam,” *J. Opt. Soc. Am. A*, vol. 35, pp. 953–958, Jun 2018.
- [221] J. Chen, D. Zhang, and Q. Zhan, “Effective iterative method for accurate amplitude modulation in complex optical field generation,” *Optical Engineering*, vol. 58, no. 8, pp. 1 – 9, 2019.
- [222] Y. Zhang, C. Chang, C. Yuan, S. Feng, S. Nie, and J. Ding, “Composite generation of independently controllable multiple three-dimensional vector focal curve beams,” *Optics Communications*, vol. 450, pp. 296 – 303, 2019.
- [223] W. Han, Y. Yang, W. Cheng, and Q. Zhan, “Vectorial optical field generator for the creation of arbitrarily complex fields,” *Opt. Express*, vol. 21, pp. 20692–20706, Sep 2013.
- [224] Z.-Y. Rong, Y.-J. Han, S.-Z. Wang, and C.-S. Guo, “Generation of arbitrary vector beams with cascaded liquid crystal spatial light modulators,” *Opt. Express*, vol. 22, pp. 1636–1644, Jan 2014.
- [225] I. Moreno, J. A. Davis, T. M. Hernandez, D. M. Cottrell, and D. Sand, “Complete polarization control of light from a liquid crystal spatial light modulator,” *Opt. Express*, vol. 20, pp. 364–376, Jan 2012.
- [226] X.-L. Wang, J. Ding, W.-J. Ni, C.-S. Guo, and H.-T. Wang, “Generation of arbitrary vector beams with a spatial light modulator and a common path interferometric arrangement,” *Opt. Lett.*, vol. 32, pp. 3549–3551, Dec 2007.
- [227] X. Zheng, A. Lizana, A. Peinado, C. Ramírez, J. L. Martínez, A. Márquez, I. Moreno, and J. Campos, “Compact LCOS–SLM Based Polarization Pattern Beam Generator,” *Journal of Lightwave Technology*, vol. 33, no. 10, pp. 2047–2055, 2015.

- [228] J. C. Gutiérrez-Vega, “Pancharatnam–Berry phase of optical systems,” *Opt. Lett.*, vol. 36, pp. 1143–1145, Apr 2011.
- [229] C. L. V. Cooke, F. Kong, A. Chaware, K. C. Zhou, K. Kim, R. Xu, D. M. Ando, S. J. Yang, P. C. Konda, and R. Horstmeyer, “Physics-enhanced machine learning for virtual fluorescence microscopy,” *CoRR*, vol. abs/2004.04306, 2020.