Polarisation microscopy of single emitters

James Hilton Clegg

Imperial College London
Department of Physics

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Ph.D) Imperial College of Science, Technology and Medicine
Abstract

This thesis contains a report on the development of a new type of confocal microscope. The microscope aims to allow the user to be able to determine the three dimensional orientation of single fluorescent emitters. The microscope has at its heart a binary spatial light modulator that allows us to control the excitation electric field in the pupil of the microscope objective. This allows us to exploit the fact that the excitation of, and emission from, a single fluorescent emitter is polarisation and orientation dependent. By changing the field in the excitation pupil we can generate a set of images that when taken together can be analysed to find the emitter orientation. We show that the microscope allows us to resolve the orientation of single fluorescent molecules and nitrogen vacancy centres in nanodiamond.

We designed the microscope from scratch using extensive mathematical modelling techniques. We anticipate that these models will be useful to other researchers. One example is that our model of the polarisation distortions introduced during scanning is relevant to any galvanometer-based scanning system. We also developed a full model of a confocal microscope that includes the dipole-like nature of many samples. We use this to calculate, amongst other things, the optical sectioning properties of confocal microscopes. This allows us to validate previous models that ignored polarisation distortions of high numerical aperture lenses and also to make calculations where previous models would have been inadequate, for example in calculating the sectioning strength of sheets of aligned dipoles.

As well as developing numerical models, we invented a new method for controlling the polarisation of light using a binary spatial light modulator. This work has applications in materials science, and industrial applications.
Author declaration

All the work presented in this thesis is my own, with the exception of three figures taken from the sources mentioned in the text.

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
To Mum and Dad
I would like to thank my supervisor, Mark A. A. Neil for giving me the opportunity to work on this project. Mark has always been able to help me when I needed it.

I originally submitted my PhD application to Paul French, and he suggested Mark as a supervisor as I expressed a keen interest in physics and optics (and wasn’t so enthusiastic about wet lab work). I am very grateful to Paul for this thoughtful act.

I would also like to thank Chris Dunsby, who has lent me a microscope frame, photomultiplier tube and control electronics. Without this kind loan my project would have been delayed.

The mechanical skills of Martin Kehoe and Simon Johnson have been invaluable, and they have always been able to make what I asked. Also thank-you to Judith, Sara and Marcia for all their help negotiating college administration issues. Academically, I have also had particular help from Sunil Kumar (how does he know so much?), Kwasi Kwakwa (single molecule sample preparation help), Hugo Sinclair (nanodiamond sample preparation help) and Sean Warren (tips and advice on writing software). All of my other colleagues have been helpful in ways that are harder to define. Thank you Romain, Dom, Lionel C, Lionel F, Alex, Wenjun, George, Hugh, Tom, Elizabeth, Fred, Mirella, Vincent, Doug, Stina, Ben, Claire and Youngchan.

I am grateful to Julian Lewis for all of our conversations about Physics, Optics and everything else. Julian suggested before I thought about doing a PhD that I might enjoy research in optical microscopy. He was right.

Finally thank you to all my friends and family, especially Sarah, Mum and Dad, for being so easy to talk to about work and everything else.

Funding was provided by an EPSRC doctoral training account.
# Table of Contents

## Abstract
- Author declaration 3
- Copyright declaration 3

## 0 Table of Contents
- List of abbreviations 9

## 1 Introduction
- 1.1. Thesis outline 11
- 1.2. Conventions in this thesis 14

## 2 Optical microscopy
- 2.1. Fundamental ideas in optical microscopy 15
- 2.2. Normalised optical coordinates 24
- 2.3. Methods of contrast in microscopy 26
- 2.4. Confocal microscopes 30
- 2.5. Multiphoton and superresolution techniques 34
- 2.6. Chapter summary 36

## 3 Spatial light modulators
- 3.1. SLM types 39
  - 3.1.1. Ferroelectric spatial light modulators 41
- 3.2. Holography 43
- 3.3. Phase and polarisation control with SLMs 48
- 3.4. Chapter summary 51

## 4 Polarisation control using an FLCSLM
- 4.1. Non-ideal operation of FLCSLMs 54
- 4.2. Binary holograms 56
- 4.3. Polarisation control scheme 59
- 4.4. Measured aberrations in the system 62
- 4.5. Experimental results 64
- 4.6. Control of holograms 67
4.7. Chapter summary 69

5 Existing methods for the determination of fluorophore orientation 71
  5.1. Single molecule techniques 71
  5.2. Dense labelling techniques 76
  5.3. Chapter summary 79

6 Theory and modelling of a confocal microscope 81
  6.1. From Maxwell’s equations to a bright focal spot 82
    6.1.1. Scalar diffraction theory 82
    6.1.2. The Kirchoff diffraction theory 84
    6.1.3. The Debye integral 87
    6.1.5. Generalisation of the Richards Wolf integrals to an arbitrary pupil 93
  6.2. Fast calculation of focal fields using Fourier transforms 94
    6.2.1. From polar integrals to Fourier transforms 94
    6.2.2. Zooming in on the focal plane: the chirp-z transform. 97
    6.2.3. Example focal field calculations 100
  6.3. Collection of light emitted from a dipole 101
    6.3.1. Pupil fields for a dipole at the Gaussian focal point 102
    6.3.2. Confocal image of a dipole 106
  6.4. Determination of dipole orientation 110
    6.4.1. 3D orientation determination in densely labelled samples 111
    6.4.2. Feasibility of orientation determination in 3D 112
    6.4.3. Aberrations and the need for a test dipole 114
  6.5. Vectorial calculation of optical sectioning 114
  6.6. Chapter summary 122

7 Design of a stationary pupil scanner 125
  7.1. Optical design of scanner 126
    7.1.1. Scan lens aberrations 132
  7.2. Modelling of polarisation changes during scanning 133
  7.3. Experimental results of pupil polarisations 142
  7.4. Chapter summary 144

8 Microscope assembly 146
  8.1. SLM-to-scan relay 148
  8.2. Detector path 150
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCD</td>
<td>Charge-coupled device</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complimentary metal oxide on silicon</td>
</tr>
<tr>
<td>cps</td>
<td>Counts per second</td>
</tr>
<tr>
<td>DAPI</td>
<td>4’,6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DAQ (box)</td>
<td>Data acquisition (box)</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DOF</td>
<td>Depth of focus</td>
</tr>
<tr>
<td>DOPI</td>
<td>Defocused orientation and position imaging</td>
</tr>
<tr>
<td>DPSS</td>
<td>Diode-pumped solid state (laser)</td>
</tr>
<tr>
<td>FLDL</td>
<td>Fluorescence-detected linear dichroism</td>
</tr>
<tr>
<td>FLCSLM</td>
<td>Ferroelectric liquid crystal SLM</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full width at half maximum</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical user interface</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid crystal</td>
</tr>
<tr>
<td>MTF</td>
<td>Modulation transfer function</td>
</tr>
<tr>
<td>NA</td>
<td>Numerical aperture</td>
</tr>
<tr>
<td>NADH</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>ND</td>
<td>Nanodiamond</td>
</tr>
<tr>
<td>NI</td>
<td>National instruments</td>
</tr>
<tr>
<td>NV</td>
<td>Nitrogen vacancy</td>
</tr>
<tr>
<td>OTF</td>
<td>Optical transfer function</td>
</tr>
<tr>
<td>PANSLM</td>
<td>Parallel-aligned nematic SLM</td>
</tr>
<tr>
<td>PBS</td>
<td>Polarising beamsplitter</td>
</tr>
<tr>
<td>PMMA</td>
<td>Poly(methyl methacrylate)</td>
</tr>
<tr>
<td>PMT</td>
<td>Photomultiplier tube</td>
</tr>
<tr>
<td>PSF</td>
<td>Point spread function</td>
</tr>
<tr>
<td>QD</td>
<td>Quantum dot</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>SCU</td>
<td>Scan control unit</td>
</tr>
<tr>
<td>SLM</td>
<td>Spatial light modulator</td>
</tr>
<tr>
<td>STED</td>
<td>Stimulated emission depletion</td>
</tr>
<tr>
<td>TIRF</td>
<td>Total internal reflection (microscopy)</td>
</tr>
<tr>
<td>λ</td>
<td>Wavelength</td>
</tr>
</tbody>
</table>
1 Introduction

Optical microscopes have been used since the mid-seventeenth century when Hooke and Leeuwenhoek saw cells and bacteria for the first time. Many have suggested the development of ideas in optics was held back in the 17th and 18th centuries because Newton, who commanded wide respect, believed in the corpuscular theory of light. Perhaps due to Newton then, it was not until 200 years later in 1876 that Abbé developed his now well-known theory that described the process of image formation in the microscope and led to his description of the resolution of the microscope. Around the same time, Carl Zeiss was able to use Abbé’s theory to develop the first lenses corrected for chromatic aberration and thus give clear images of objects illuminated with white light. Though the microscope might have been considered a finished design in 1900, the invention of the laser in 1960 [1,2] has dramatically changed the field, leading to the invention of the confocal microscope [3,4], and more recently an array of super-resolution techniques [5,6].

The example above of the laser illustrates an important point: research into optical microscopy is a highly interdisciplinary research area in which progress is made possible by advances in technology from other fields: semiconductor silicon-based electronics have led to the development of CCD and CMOS light detecting arrays for recording photomicrographs; and recent advances in Chemistry have led to a huge array of photostable fluorescent dyes which can be combined with antibodies to attach to very specific biological structures in cells [7]. This makes this research field very exciting: who knows from which area the next revolutionary change will come?

The many existing types of microscopes are all designed to make samples look characteristic in particular ways. Most simply, a sample can be illuminated with white light, and a magnified image is produced showing structure due to
absorption or scattering in the sample. This is the oldest form of microscopy, known as bright field imaging. It has a simple partner, dark-field imaging whereby light is scattered into rather than out of the collected region, which gives images with sharp bright features against a dark background. Many samples don’t exhibit strong scattering or absorption, so alternative contrast mechanisms have been invented. Today, research into microscopy often centres on finding new methods of contrast. Essentially, this means finding new ways of sensing the properties of a sample that aren’t visible under normal circumstances (for example bright field illumination). In this thesis, we present a microscope that can determine the orientation of single fluorescent emitters. Sensing the orientation of a fluorescent molecule is an example of a contrast mechanism.

We have carried out the development of our microscope with an open mind to its application. As things stand, most people who have wanted to determine the orientation of emitters have been biologists studying orientation effects in proteins, cell membranes and the cellular cytoskeleton. Over the course of this research, we have discussed potential applications with people from a wide range of fields. To give just two examples, a plasma physicist was interested in a microscope whose electric field at the focal spot could be switched very quickly for studying plasmonics in small resonant metal structures; and a colleague was interested in quantifying the degree of alignment of fluorescent collector molecules in his solar cells. These discussions give us confidence that our microscope will find a wide range of applications.

1.1. Thesis outline

We begin in chapter 2 with some background information about optical microscopy. We explain various fundamental ideas, particularly image formation by positive lenses and how the wave nature of light leads to a resolution limit. We then go on to discuss fluorescent confocal microscopes, and finally conclude with a brief comment on multiphoton and superresolution techniques.
Chapter 3 contains background information relating to a key component of our microscope: the spatial light modulator (SLM). We describe different types of SLMs, with particular focus on the ferroelectric liquid crystal SLM (FLCSLM). It is this type of SLM that we use in our microscope. We explain the principle of off-axis holography and use this explanation as the foundation of our discussion of methods of controlling the phase, amplitude and polarisation of light with SLMs.

Chapter 4 contains a description of our novel method of controlling the phase, amplitude and polarisation of light using an FLCSLM. This is the first chapter of this thesis that contains original work. Our method is not the first to allow this level of control of light using an FLCSLM but it is more robust and simple than previous techniques. We describe the details of calculating binary holograms given required output field states. We show experimental results that test the efficacy of our method in producing the output field states that we desire.

Chapter 5 contains background information describing existing optical methods of measuring dipole orientation. We discuss methods that operate in two regimes: single-molecule techniques and dense-labelling techniques. We describe the major techniques of both types and discuss their applications and limitations.

In chapter 6 we present our full numerical model of a confocal microscope that includes the effects of high NA focusing, collection of polarised light emitted from dipoles and sample scanning. The chapter begins by explaining, starting with Maxwell’s equations, how a high NA lens forms the focus of a light beam. The theory that explains this is known as the Richards-Wolf theory, and expresses the focal fields as integrals over polar coordinates. We explain how to compute these integrals very quickly using discrete Fourier transform algorithms. We then discuss how a high NA lens collects the light emitted from a radiating dipole. With the model of the excitation focal fields and the collected fields, we can model the excitation and the detection PSFs of a confocal microscope, and
therefore predict the image of a dipole at any angle in a microscope with any illumination pupil function. Using this model we discuss the feasibility of using a confocal microscope to determine the orientation of fluorescent dipoles in 3D. We close the chapter with a further application of our full vector model of a confocal microscope, which calculates the expected optical sectioning behaviour of thin sheets of randomly oriented dipoles and of highly aligned dipoles.

Next we begin to discuss the construction of our microscope. In chapter 7 we present the design of a 2D beam scanning system that keeps the beam stationary in the objective pupil. We construct a model of polarisation changes introduced by scanning units and show how these changes limit the performance of a scanning system.

Chapter 8 describes how all of the elements of our confocal microscope fit together. The previously described individual parts are the polarisation control scheme (chapter 4) and the beam scanning system (chapter 7). We also introduce our detection path, which uses a photon counting PMT and so allows us to detect very low light levels. We describe how these components are optically and electronically connected, and how we use a computer to control the entire system.

Chapter 9 contains our main experimental results. We begin by validating our confocal microscope model and measuring system aberrations using fluorescent beads. We also develop an extension to our model that allows us to predict the behaviour of focused fields at an interface. We then move on to show images of single molecules with our microscope. By comparing images taken with azimuthally, radially and circularly polarised light we show that we can determine the orientation of single molecules qualitatively. We then show images of single NV centres in nanodiamonds. A single NV centre in diamond contains two orthogonal dipoles. Our microscope can resolve the orientation of these dipoles. We conclude our imaging of single emitters with measurements of quantum dots. The quantum dot is a single emitter that behaves isotropically.
The progression from single molecules to quantum dots allows us to explore a range of fundamental emitter types.

We conclude with the 10th chapter. Here we sum up our results and provide an evaluation of the success of our microscope relative to its aims. We discuss the likely impact of our research and the next steps for the development of the instrument.

1.2. Conventions in this thesis

In the work that follows, the $z$ axis is always assumed to be along the optic axis of the microscope and the $xy$ plane is perpendicular to the optic axis. ‘In plane’ always means in the $xy$ plane. Longitudinal means along $z$, as does axial. Lateral means in the $xy$ plane. The polar angle is the angle to the $z$ axis. Unless otherwise stated modelling has been carried out at NA = 1.4 in oil ($n = 1.52$).
2 Optical microscopy

We provide this chapter to explain some standard results in optical microscopy and set out the language that we will use to evaluate other techniques in subsequent chapters. We begin with a brief history of the microscope, starting with Hooke in the 17th Century. We then lay out the foundations of imaging and microscopy, briefly discussing aberrations and optical design. We then develop the principle of the resolution limit in an optical system from basic consideration of interference, and then more formally using Fourier theory. This leads us to introduce the point-spread function (PSF) and optical transfer function (OTF) of an optical system. We briefly discuss the normalised optical coordinate system \((u, v)\). The axial normalised optical coordinate \(u\) is given special attention because it has multiple definitions that apply in different circumstances, and is often misused or used confusingly in the literature. Armed with a clear language, we can then discuss common methods of contrast in microscopy, particularly fluorescence, with a brief comment on photobleaching. While discussing fluorescence we note that the absorption and emission of light by a fluorophore are directional and polarisation dependent. We exploit this effect in the design of the microscope presented in this thesis. We describe the use of fluorescent dyes and proteins in biological imaging. We then proceed with a discussion of resolution and optical sectioning in the confocal microscope. We introduce the low numerical aperture (NA) theories, which contrast with the full high NA theories discussed in chapter 6. We close the chapter with descriptions of multiphoton microscopy and superresolution techniques.

2.1 Fundamental ideas in optical microscopy

A key component of any microscope is the objective lens. A lens is made of optically transparent material and is designed to bend light rays in a particular way. The simplest form of positive lens is made of just one glass element. When
light is incident on the lens surface, it refracts according to Snell’s law [8]. A single glass element is made of two curved surfaces that cause rays at higher distances from the optic axis to refract more. An example is shown in Figure 2.1. The shapes of the surfaces can be chosen so that the bending of rays is just right, causing light emitted from a point on one side of the lens, say A (object point) to be focussed down to a corresponding image point, B, on the other side. A and B are known as conjugates. The imaging property of a positive lens can also be understood in terms of an integral approach: by Fermat’s principle, light travels from A to B via the shortest optical path [9]. Light travels slower in glass than in air, and a positive lens is made thicker on the optic axis than at its edges by just the right amount that all rays from A to B experience the same optical path. It is pleasing that the Snell’s law approach and the Fermat’s principle approach both describe the operation of the lens. Given the axial position of A and the lens, the position of B can be found from Newton’s conjugate distance formula:

\[ zz’ = -f^2 \]  \hspace{1cm} (1)

Here the quantities are as indicated on the diagram in Figure 2.1. The negative sign ensures that \( z \) and \( z’ \) have opposite signs, in other words when A is to the left of the focal plane in object space, B is to the right of the focal plane in image space. The image is magnified by the transverse magnification \( M \) given by

\[ M = \frac{l’}{l} \]  \hspace{1cm} (2)
Figure 2.1. Simple single-element lens forming an image of point A at B. The image of objects in the same plane as A are magnified to the plane containing B as shown by the red ray.

Though a lens can be designed to give optimal focusing for a particular pair of conjugates, for example A and B as shown, it cannot be used for imaging extended objects perfectly. Rays collected by the lens from A’ will not cross at a single point B’ in image space, and the optical system is said to suffer from aberrations [10]. Aberrations can be classified based on how they degrade an image. One basic aberration is Spherical aberration, which causes rays passing at different heights through the lens to be focussed to different axial positions. Chromatic aberration arises because different wavelengths travel at different speeds in a material. This is known as dispersion and causes lenses to have slightly different focal lengths and magnifications for red and blue light. Spherical and chromatic aberration can be corrected by the use of lens pairs containing different glass types. There are many other aberrations, and the general way to quantify them is to compare the shape of a wavefront in the exit pupil of the system to an ideal spherical wavefront. The comparison is often made in one of two ways: using Seidel sums, which characterise the aberrations in a whole optical system; or using Zernike polynomials, which express the aberrations at a particular image point as an expansion in orthonormal functions in the exit pupil.
Imagine an optical designer trying to design a lens for a given application. Their goal is to introduce a number of refractive or diffractive elements to ensure that the rays all cross within an acceptably small volume in image space. The designer might need to know the type of object to be imaged, how bright it will be, where it might be situated, the location and pixel size of the detector, the acceptable mass of the lens, the range of operating temperatures and the available budget. The size of acceptably small will clearly depend on the given application. In the case of the microscope, the lenses are often diffraction limited, meaning that the intersection points of all rays for a given image point are contained within the finite volume of space that light would be focused to in the absence of aberrations. How small is this volume?

![Diagram](image)

**Figure 2.2.** A 4F system. We use this to introduce the idea of the diffraction limit. The dashed line is at the common focal plane of the two lenses and indicates the stop in the system.

To aid this discussion further, consider the system shown in Figure 2.2. In this setup, two lenses are separated by the sum of their focal lengths. The stop is placed at the common focal plane. The stop is the element that restricts the angular extent of rays emitted from A. Modern infinity corrected microscopes use this configuration where the first lens, the objective, has a short focal length and the second, the tube lens, a longer focal length. The magnification of the system is given by \( M = -f_2/f_1 \) (to see this consider the blue chief ray shown in the diagram) and is independent of the object position (ignoring aberrations).
Now, the optical paths to point B from all points in the stop are equal. What about the optical paths from the stop to point B? For now consider just two rays, one passing through the top of the stop, and one through the bottom. Let the path difference be \( \Delta \), and we can write

\[
\Delta = D\delta\theta .
\]

There will be constructive interference in the image plane as long as the path difference is smaller than \( \lambda/2 \). When \( \Delta = \lambda/2 \) there will be destructive interference. Now clearly \( |\overline{BB_d}| = f_2 \delta\theta \), so we can say that when

\[
\delta R_2 = |\overline{BB_d}| = \frac{\lambda f_2}{2D} ,
\]

there will be destructive interference in the image plane. Here we have introduced \( \delta R_2 \) as the size of the diffraction limited spot in image space. This gives us an expression for the resolution of the optical system. We can see that a single point emitter at A gives rise to a disc of light of radius \( \delta R_2 \) in image space, so if another point source a lateral distance of \( \delta R_1 \) from A is to be resolved then its image point must be further than \( \delta R_2 \) from B. Therefore in sample space the lateral resolution is approximately

\[
\delta R_1 = \frac{\delta R_2}{M} = \frac{\lambda f_1}{2D} .
\]

Similar arguments can be used to show that the axial resolution in sample space is \( \delta Z_1 = \frac{\lambda f_2^2}{D^2} \) [8]. The important point is that to maximise the resolution of an optical system, we must make \( \frac{D}{f_1} \) as large as possible.

Here we have calculated only the approximate form of the resolution of an optical system because we have considered adding phases from just two points in the stop. Of course in reality the contribution from all points must be integrated. It can be shown using diffraction theory (see chapter 6) that the field at the focus of a low numerical aperture lens is given by the Fourier transform of the pupil function, \( W \) [11]. The pupil function is the complex field in the plane of the stop in Figure 2.2. For a circular pupil with plane wave illumination, the integral can be written as a Fourier-Bessel transform:
Here \( J_n \) is the \( n \)th order Bessel function of the first kind and \( h_a \) refers to the field amplitude in the focal plane as a function of \( r \), the radius in the focal plane. \( A \) represents the strength of the incident field, \( f \) is the focal length of the lens, \( a \) is the radius of the pupil, \( \rho \) is the radial coordinate in the pupil plane and \( k = 2\pi/\lambda \) is the magnitude of the wavevector. The intensity in the focal plane is therefore given by

\[
h(r) = |h_a|^2 = \left( \frac{A}{\lambda f} \right)^2 \left[ \frac{J_1 \left( k \frac{a}{f} r \right)}{k \frac{a}{f} r} \right]^2,
\]

which is known as the Airy pattern and is plotted in Figure 2.3. The distribution is sharply peaked at zero and is surrounded by dimmer rings. 84\% of the energy lies inside the first zero, and 91\% inside the second zero. The zeros of \( J_1 (x) \) are not evenly spaced, but the first zero is at \( x = 1.22 \pi \). One Airy unit is defined as the radius in the focal plane for which \( k \frac{a}{f} r = 1.22 \pi \). Rayleigh’s definition of the resolution of a microscope says that two points are resolvable if the peak in the PSF of one point coincides with the first zero of the PSF of a second point. The distance between the two points is then the radius of the Airy pattern (shown in Figure 2.2). This is:

\[
r = \frac{0.61 f \lambda}{a}.
\]

This is surprisingly close to the approximate result in equation (5). The Airy pattern is the image of a point in an optical system such as the one shown in Figure 2.2, and is termed the point-spread function (PSF) of the system. We have implicitly assumed that the sample is illuminated with a uniform patch of light, and so our equation (7) is also the PSF of a wide-field microscope.
Figure 2.3. The Airy distribution. On the left is a 1D section through the Airy distribution. The red line shows the position of the first zero of the function at $\frac{k_0 x}{f} = 1.22\pi$. On the right is the Airy distribution in 2D. The right-hand portion shows the intensity on a logarithmic scale.

Figure 2.4. An infinity corrected microscope. The sample space is immersed in medium having refractive index $n$. The two black solid curved surfaces represent principal surfaces of the objective and tube lens. The black dashed surfaces represent the other principal surfaces, where the ray $A'P'Q'B'$ can be considered to refract. The focal lengths are indicated along the top of the diagram, but really are measured as radii of the principal surfaces. The blue ray shows that the magnification is given by $M = -\frac{f_T}{f_O} = -\frac{NA_0}{NA_T}$.

The results we have discussed up to now need to be clarified for a microscope. Here we refer to Figure 2.4, a diagram of the imaging arm of an infinity corrected microscope. The numerical aperture (NA) is defined as
\[ \text{NA} = n \sin \alpha , \]  

(9)

where \( n \) is the refractive index in sample space and \( \alpha \) is the angle between the marginal ray and the optic axis. For a system obeying the Abbé sine condition [12] the ray radius in the pupil plane \( \rho \) and angle that the ray makes with the optic axis, \( \theta \) are related by

\[ \rho = nf_0 \sin \theta , \]  

(10)

For the marginal ray we can write

\[ a = nf_0 \sin \alpha_0 = f_0 \text{NA}_0 \]  

(11)

Clearly equivalent relationships must hold in image space, giving

\[ f_0 \text{NA}_0 = f_T \text{NA}_T . \]  

(12)

Now with reference to Figure 2.4 the transverse magnification is given by

\[ M = -\frac{f_T}{f_0} = -\frac{\text{NA}_0}{\text{NA}_T} . \]  

(13)

Also, we note that \( \text{NA}_0 = \frac{a}{f_0} \). Using our ideas of resolution, we can rewrite equation (8) as

\[ r = \frac{0.61\lambda}{\text{NA}_0} , \]  

(14)

which is the conventional formula for the resolution of a microscope in sample space where the free-space wavelength is \( \lambda \). This formula is an oversimplification at high NA when polarisation effects are important. We discuss this situation in chapter 6. Using an immersion medium other than air has two advantages. First, it reduces reflection losses at the boundaries between the sample and first element of the objective lens. Second, it decreases the wavelength in sample space by a factor \( n \), thereby increasing the resolution by that same factor (for the same cone angle).

For a shift-invariant system an image is formed by convolving the object with the PSF of the optical system [11]. In reality, aberrations affect different portions of the image in different ways, meaning that many systems are only approximately shift-invariant over a small patch. Over any such patch, and by the convolution
theorem, the frequency content of the image is the frequency content of the object multiplied by some function $H$. $H$ is known as the optical transfer function of the system. For the infinity corrected microscope just described, the PSF is given by

$$h(r) = \vert h_a \vert^2,$$

where $h_a$ is the Fourier transform of the pupil function $W$. $W$ is the complex amplitude of the electric field in the pupil. We can now calculate

$$H(k) = F\{\vert h_a \vert^2\} = F\{h_a\} \ast F\{h_a\} = W \ast W,$$

where $\ast$ denotes correlation and we have invoked the autocorrelation theorem [11]. In words, equation (16) states that the OTF is the autocorrelation of the pupil function. Evaluating optical systems using their OTF is key because it is the frequency content of an image that is the most useful measure of resolution for many purposes. In general the OTF is complex because spatial frequencies can be changed in phase and amplitude by propagation through an optical system. The modulation transfer function (MTF) is the absolute value of the OTF. The MTF of a microscope as shown in Figure 2.4 is shown in Figure 2.5. This is the autocorrelation of a circle of radius $F_{\text{max}} = \frac{a}{\lambda f}$. Here $F_{\text{max}}$ is the maximum spatial frequency transmitted by the aperture, which has radius $a$. The transverse spatial frequency $F_\perp$ for a ray at $\theta$ to the optic axis is $F_\perp = \frac{\sin \theta}{\lambda}$. The normalised spatial frequency is $\frac{F_\perp}{F_{\text{max}}}$, so that a normalised spatial frequency of 1 corresponds to the maximum spatial frequency transmitted by the system aperture.
Figure 2.5. The MTF of a system having a circular pupil. We plot the amplitude of the transmitted frequency components against normalised spatial frequency, which is spatial frequency divided by $F_{\text{max}}$ as defined in the text.

2.2. Normalised optical coordinates

Equation (7) is a formula for the image of a point in an optical system. It is a function of $k \frac{a}{f} r$, which suggests introducing a scaled variable $v = k \frac{a}{f} r$. In terms of this coordinate, the width of the central lobe of the Airy distribution is always the same value, $v = 1.22\pi$. For an objective obeying the sine condition then $\frac{a}{f} = \text{NA}$, so

$$v = kr \text{ NA} . \quad (17)$$

We can also introduce an axial normalised optical coordinate, $u$ that also allows the PSF to be represented in a system-invariant way. The form of the axial normalised optical coordinate arises as follows. With reference to Figure 2.6, the extra optical path arising from an axial shift $z$ in observation point for a
particular ray at pupil height $\rho$, is given by $\delta P = n z \cdot \cos \theta$. The extra phase in the pupil is therefore given by

$$\phi = k n \delta P = n z k \cos \theta = n z k \left(1 - 2 \sin^2 \frac{\theta}{2}\right).$$  \hfill (18)

Note that there is a constant phase that we can neglect, and a term that varies across the pupil. So we write $\phi = \phi_0 + \phi_p(\theta)$, with

$$\phi_p(\theta) = 2 n z k \sin^2 \frac{\theta}{2}.$$  \hfill (19)

It is now clear that this does not represent a single term in the power series expansion of the pupil radius $\rho(\theta)$. However, introducing

$$s = \frac{\sin \frac{\theta}{2}}{\sin \frac{\alpha}{2}},$$  \hfill (20)

and

$$u = 4 n z k \sin^2 \frac{\alpha}{2}$$  \hfill (21)

allows us to write

$$\phi_p(s) = \frac{1}{2} u s^2,$$  \hfill (22)

Which is an analogue of the low NA formula for defocus, $\phi = \frac{1}{2} n k z \rho^2$. Sheppard shows that using this definition of $u$, the first zeros of a diffraction limited PSF are at $\pm 4\pi$ along the optical axis at low and high NA [13–15]. This allows us to introduce the axial airy unit which is given by $\frac{u}{4\pi}$. Sheppard gives a nice intuitive explanation of the form of $u$, which is that it is proportional to the axial extent of the pupil (see [14], figure 3 and Figure 2.7). This is closely related to McCutchen’s theorem, which states that the field along a line passing through the geometric focal point is proportional to the Fourier transform of the projection of the aperture onto that line [16,17]. At low NA we can approximate

$$u \approx n z k \alpha^2 \approx k z \frac{\text{NA}^2}{n}.$$  \hfill (23)

This is the form of the axial normalised optical coordinate given by Born & Wolf [12].
2.3. Methods of contrast in microscopy

We have mentioned bright field and dark field imaging modalities that are used to image scattering and absorption in samples. For samples that do not scatter or absorb, other approaches must be used. An important example is Zernike phase contrast microscopy, in which index variations in transparent samples, most importantly in cells, are converted into intensity variations in an image [18–20]. After transmission through a weak phase object, the scattered and unscattered light are different in phase by $\pi/2$. In Zernike phase contrast microscopy, another $\pi/2$ phase difference is added between the scattered and unscattered light which converts the phase changes into intensity variations. The extra quarter wave phase is added as follows. The sample is illuminated with a narrow range of angles by using a mask in the illumination pupil. When correctly aligned the illumination and imaging pupils are conjugate planes. Therefore the unscattered light in the imaging pupil will be in the corresponding place to the illumination mask in the illumination pupil, and all scattered light will fill the remainder of the pupil. Then we insert into the pupil a material giving quarter wave phase delay to only the unscattered light. Zernike phase contrast microscopy images contain a characteristic bright halo surrounding each cell. Nomarski differential interference microscopy (DIC) overcomes this artefact by recording the interference between two slightly shifted orthogonally polarised images.
We now turn to a method of contrast that is ubiquitous in biological imaging, fluorescence. Fluorescence occurs when an electron in a molecule absorbs a photon and a photon is subsequently emitted at a longer wavelength. The probability of absorption and wavelength of emission depend on the energy level structure of the fluorescent molecule in question. An example is shown Figure 2.8. Say the molecule is initially in a state $S_0$ and absorbs a photon of energy $E = h\omega = \frac{hc}{\lambda_{ex}}$. Over a time of roughly 1 ps the electron then relaxes through a narrow band of vibrational energy levels to the top of the visible energy gap. These relaxations are fast because the density of states is high in this dense band of energy levels and there is a large overlap of wavefunctions of the vibrational states (see for example the section on Fermi’s golden rule in [21]). The molecule then spends roughly 10 ns in the excited state, $S_1$, before relaxing down to somewhere in the lower band of vibrational states and then quickly falling back to $S_0$. The difference in wavelength between the excited and emitted photon is known as the Stokes’ shift [22]. The Stokes’ shift allows the detection of extremely weak fluorescence signals such as those from single molecules by the separation of the incident and emitted light using spectral filters [23]. A simplified energy level diagram and fluorescent spectrum of a real dye, Dil, are shown in Figure 2.8.

![Jablonski diagram and fluorescent spectrum](image)

Figure 2.8. Left: Jablonski diagram showing the process of fluorescence absorption and emission. Right: diagram showing the absorption and emission spectra for a Dil, a commonly encountered fluorescent molecule. The absorption peak is at 550 nm and the emission is at 570 nm.

A fluorescent molecule cannot undergo excitation and emission ad infinitum. When a molecule is excited, it has a small probability of undergoing a forbidden transition to a triplet state. The triplet state is long-lived because the transition
back to the ground state is also forbidden. Here forbidden means that the transition cannot proceed in accordance with the standard selection rules of quantum mechanics. These assume that total angular momentum must be conserved in the electron-photon system. Forbidden transitions can proceed if the atomic nucleus exchanges energy with the electron; now we say that total angular momentum must be conserved in the whole molecule-photon system. These transitions therefore require the coincidence of multiple processes and so are relatively improbable. When in a triplet state the molecule does not emit a photon at the expected emission wavelength, and remains unexcitable by the excitation light. Therefore the molecule appears dark. This is a form of reversible photobleaching also known as blinking [24]. In addition, fluorophores in their excited triplet states are more likely to chemically react with molecular Oxygen (which has a triplet ground state). This reaction permanently destroys the fluorescent properties of the molecule, a fact borne out by experiment: significant reduction in photobleaching is seen when Oxygen-scavenging species are introduced to a sample [25].

Now we try to understand a little more about the nature of transitions in fluorophores. Using time-dependent perturbation theory, transitions between electronic energy levels can be shown to occur at a rate governed by Fermi’s golden rule, which states that

\[ \Gamma \propto |E_{ex} \cdot \langle \psi_i | r | \psi_f \rangle |^2, \]  

(24)

where \( \Gamma \) is the transition rate, \( E_{ex} \) is the excitation electric field, \( \psi_i \) and \( \psi_f \) are the initial and final state wavefunctions and \( r \) is the position operator [21]. The key point is that the transition rate depends on the interaction of the electric field and the transition dipole matrix element, \( \mu = \langle \psi_i | r | \psi_f \rangle \). In spherically symmetric systems, for example the hydrogen atom, \( \mu \) is isotropic (meaning that if \( \mu = (\mu_x, \mu_y, \mu_z) \) then \( |\mu_x|^2 = |\mu_y|^2 = |\mu_z|^2 \)). The dependence of \( \mu \) on the wavefunctions of the system means that there is a close link between the symmetry of a molecule and its interaction with light. In many fluorophores, the molecules have an axis of symmetry, which makes one component of \( \mu \) much larger than the other two. In this case we say the fluorophore has an electric...
dipole moment along $\mathbf{u} = \mathbf{d}$. We can now understand that the absorption probability obeys

$$\text{absorption probability } \propto |\mathbf{d} \cdot \mathbf{E}_{ex}|^2.$$

(25)

In addition, the emission will be polarised, with the electric field $\mathbf{E}_{em}$ observed at position $\mathbf{r}$ from the dipole given by [9]

$$\mathbf{E}_{em} = \frac{\omega^2}{4\pi\varepsilon_0 c^2 r^3} \mathbf{r} \wedge \mathbf{p} \wedge \mathbf{r}.$$

(26)

Here $\mathbf{p}$ is the dipole moment, in units of Coulomb-metres, $\omega$ is the angular frequency of oscillation of the dipole, equal to the angular frequency of the emitted light, $\varepsilon_0$ is the permittivity of free space and $c$ is the speed of light in vacuum. We have ignored the non-propagating fields that exist in a small region surrounding the dipole. The key point to take from equations (25) and (26) is that the excitation and emission processes are directional and polarisation dependent. This is the basis of operation of the microscope described in this thesis and of many other techniques for determining the orientation of fluorescent molecules using microscopes [26,27].

The non-propagating fields from a dipole are important in Förster Resonance Energy Transfer (FRET), a widely used tool in fluorescence microscopy [28]. FRET is based on the principle that the non-propagating fields are short-range [9]. When two molecules are very close together, typically below 10 nm [29], they can mutually interact through these non-propagating fields. This can be read out as a fluorescence signal, and provides a tool to study the interaction of features much smaller than the optical resolution limit of the microscope, for example in investigating protein-protein interactions [29].

Many naturally occurring molecules, for example NADH and collagen, are inherently fluorescent and therefore fluorescence imaging with one of these molecules is possible without staining or labelling the sample. This is referred to as endogenous fluorescence or autofluorescence and can be used to study tissue structure [30]. More commonly, fluorophores are introduced into a sample to highlight specific parts of a cell. For example the Carbocyanide dyes alter their
structure when embedded in cell membranes and in doing so become highly fluorescent [31] and DAPI is a fluorescent dye that becomes highly fluorescent when bound to DNA [32]. In addition, it is common practice to couple fluorescent molecules, for example rhodamine, to antibodies that bind to specific molecules inside cells [7]. Labelling a cell in this way requires the cell membrane to be made permeable, which usually kills the cell. Fluorescent proteins can be used to study processes inside living cells. GFP, the first fluorescent protein to be discovered, was noted in 1962 [33] and cloned and sequenced in 1992 [34]. After cloning the gene can be introduced into the cells of other species, meaning that cells make their own fluorescent tags. This wealth of methods makes fluorescence microscopy an extremely powerful tool, particularly in the field of biological imaging.

2.4. Confocal microscopes

![Confocal microscope diagram]

Figure 2.9. Diagram showing the confocal principle. Light emitted from the focal point passes through the pinhole, but light from away from the focal spot, both laterally and axially, is largely rejected. The image is built up by scanning the sample (or the illumination light) and detector and recording the intensity as a function of position.
The confocal microscope was invented by Minsky in 1955 [3] and the first confocal fluorescence image was published in 1972 [35]. Since then, confocal microscopy has found application in a wide range of materials science, semiconductor, medical and biological areas [36]. Commercialisation of the confocal microscope now means that many biology research laboratories have their own confocal microscope systems [37].

The basic layout of a confocal microscope is shown in Figure 2.9. Fluorescence microscopes are usually used in epi-illumination mode, in which the same objective is to both illuminate and collect light. This configuration is shown in Figure 2.10. In a confocal microscope the sample is illuminated with a diffraction-limited focal spot and imaged onto a pinhole. Light passing through the pinhole is integrated by a detector and the image is built up by scanning the excitation focal spot relative to the sample and recording the intensity as a function of position. The detector must also be scanned to match the excitation focal position. Early confocal microscopes used mechanical stages to scan the sample [38,39] or scan heads to move the beam [40]; beam scanning is now more popular for two reasons. First, it allows faster image acquisition because rotating small mirrors is easier than translating the sample. Second, the detector does not need to be scanned if the fluorescent light is returned through the same apparatus that scanned the excitation light. Normally scanning in the z direction is achieved by translating the sample or the objective. The first thing to notice

![Diagram showing the modified setup for an epi-illumination confocal microscope. This is the configuration most used in confocal fluorescence microscopy. A dichroic beamsplitter is used to direct the excitation light to the sample. In a reflectance confocal microscope this could just be a beamsplitter.](image-url)
about the image recorded in a confocal microscope is that it is optically sectioned. This means that light emitted from away from the focal spot it largely rejected. This is illustrated by the grey dashed lines in Figure 2.9. The property of optical sectioning allows images to be taken inside samples with enhanced contrast because the out-of-focus background is removed.

![Diagram of a confocal microscope used to explain the form of the PSF.](image)

Figure 2.11. Diagram of a confocal microscope used to explain the form of the PSF. In this setup, we consider scanning the sample (which is a point source). We have unfolded the epi-illumination setup for clarity. The contribution to the confocal image at the point shown, say \( r \), is given by \( h_\text{em}(r) [h_\text{em} * p](r) \).

Now we consider the PSF of the confocal microscope. We can describe the image forming process mathematically as follows. We make use of Figure 2.11 which shows the interaction of the focused excitation field, a point object and the emitted light and confocal pinhole. Consider a point object at position \( r \) in sample space. It is emits fluorescence with strength proportional to \( h_\text{ex}(r) \) where \( h_\text{ex} \) is the intensity point spread function of excitation. The image of the point in the plane of the tube lens is of the form \( h_\text{det}(r' - r) \) where \( r' \) is the position in the tube lens focal plane and \( h_\text{det} \) is the detection PSF. Only the light passing through the pinhole \( p \) is collected and integrated: \( \int h_\text{ex}(r)h_\text{det}(r' - r)p(r') \, dr' \). Therefore the PSF is given by

\[
\text{Confocal PSF}, h_c(r) = h_\text{ex}(r) \int h_\text{det}(r' - r)p(r') \, dr' .
\]

In words, the confocal PSF is the excitation PSF multiplied by: the detection PSF convolved with the pinhole. For an infinitely small pinhole we have that \( h_c(r) = h_\text{ex}(r)h_\text{det}(r) \), and if we ignore Stokes’ shifts then the excitation and detection PSFs are identical, so
Ignoring the Stokes’ shift is justified on the grounds that for many fluorescent molecules it is a small fraction of the incident frequency, for example in Dil (spectrum in Figure 2.8) the shift is 4%. We have simulated the effect of the Stokes’ shift on the PSF FWHM (using the model developed in chapter 6) and show the results in Figure 2.12. For Dil, the FWHM is only 1.8% larger than in the absence of a Stokes’ shift, and on this basis we ignore the Stokes’ shift from this point on. Figure 2.13 shows the PSF and MTF of the confocal microscope for an infinitely small pinhole. The FWHMs of the wide field and confocal PSFs are 0.84 and 0.61 Airy units, corresponding to \( \frac{0.51\lambda}{NA} \) and \( \frac{0.37\lambda}{NA} \) respectively. As a consequence of equation (28) and the autocorrelation theorem, the region of support of the confocal MTF is twice as large as the wide-field MTF. It is sometimes said that confocal microscopes have double the resolution of a wide-field system, but this isn’t accurate for a number of reasons. First, the MTF is almost zero above a normalised spatial frequency of 3. Second, in reality the pinhole must be opened to some extent to give sufficient fluorescent signal; Wilson showed that this causes the resolution improvement to be lost [41]. We mentioned that confocal microscopes also perform optical sectioning which allows 3D images to be reconstructed. One helpful way to quantify the optical sectioning is to imagine a thin fluorescent sheet being scanned along \( z \). The curve of the total intensity as a function of \( z \) will have a maximum in focus and tail off away from the focal plane. The FWHM is a measure of the sectioning strength. Wilson showed that opening the pinhole to 1 Airy unit does not significantly worsen the sectioning strength [42]. We discuss optical sectioning further in chapter 6.
2.5. Multiphoton and superresolution techniques

In 1931 it was predicted that two photons could be absorbed simultaneously by a fluorophore [43] if they arrived in a short time frame (Diaspro suggests that the time between arrivals should be less than $10^{-16}$ s [37]). The probability of absorption, and therefore the PSF, depends on the square of the optical intensity [44]. This has a number of benefits. Firstly, the PSF is of the same form as the PSF of the confocal microscope with infinitely small pinhole. There is
therefore inherent optical sectioning in a two-photon system. For a given fluorophore the resolution is not increased because the excitation wavelength is twice as long as in the single photon case. The multiphoton microscope also causes less photodamage to the sample than the single photon microscope because fluorophores are unlikely to be excited outside the focal volume. In addition, the longer wavelength is less susceptible to scattering and can therefore be used to image deep into the sample.

Recently a number of methods have been invented that circumvent the diffraction limit. These methods fall broadly into two categories: first, there are linear methods that can at most double the wide-field resolution and second, some non-linear methods that offer theoretically unlimited resolution. In order to demonstrate that the diffraction limit has been exceeded, one must consider the spatial frequencies present in the image. Simply taking a line profile and measuring the FWHM is not acceptable, because the FWHM of a PSF can be made arbitrarily small by linear filtering in software. This does not constitute increased resolution because energy is distributed into outer lobes in the PSF, causing image artefacts and the appearance of spurious objects. From Fourier theory we have argued that the MTF is related to the pupil function. With a detection and excitation path, the confocal microscope extends the region of support of the MTF to 4 normalised spatial frequency units, twice the diffraction limit. Any linear method will also have a fixed relationship between pupil function and MTF, so the region of support cannot be extended beyond these same limits [45,46]. Some methods, for example super-resolution structured illumination microscopy (SIM) [47], or some annular pupil filters, fill this region of support better than the confocal case, and so a marginally higher resolution is observed. In these examples the linear filtering is being done optically rather than digitally, which results in better signal to noise. This also makes deconvolution more successful. Another technique, considered by Sheppard in 1988 [48] and constructed in 2010 [49] is termed image scanning microscopy. This technique achieves the ideal confocal resolution without sacrificing light by recording a pixelated image of the detection PSF and computationally reassigning the light based on the idea of confocal microscopy with an offset.
detector [50–53]. The use of a CCD makes the image recording slow, so the process has been parallelised using an array of pinholes [54]. A microscope has also been constructed that optically reassigns the light [55]. The second class of superresolution techniques offer the potential for unlimited resolution by exploiting nonlinearities in the sample. The nonlinearities break the Fourier relationship between pupil size and MTF extent. Stimulated emission depletion microscopy (STED) [5,56–58] is a point scanning technique that reduces the size of the effective PSF by using a strong doughnut shaped region surrounding the excitation PSF to suppress fluorescence by stimulated emission. The FWHM of the PSF is proportional to $1/\sqrt{I_{\text{STED}}}$ where $I_{\text{STED}}$ is the intensity in the depletion beam. The need for very high excitation powers makes this technique most useful in inorganic samples, where a resolution of 4-7 nm has been achieved in diamond with a solid immersion lens [59]. The best claimed resolution for STED in living cells is 70 nm [60], though this figure was obtained by taking a line profile and thus should be interpreted with caution. As a final note in this section, we point out that biologists often use electron microscopy in tandem with optical microscopy when very high resolution is needed [61].

2.6. Chapter summary

We hope that this chapter serves to underpin the language used and techniques discussed later in this thesis. We began by providing a potted history of microscopy from the 17th Century. We showed from first principles that an optical system must have a limited resolution, and we showed the origin of the formula for the resolution of a wide field microscope. We then briefly discussed normalised optical coordinates. Care should be exercised when using this coordinate system because the axial normalised optical coordinate $u$ is different at high and low NA, and for systems obeying the sine or Hershel condition, though this latter point was only presented implicitly. We then discussed fluorescence and photobleaching, and pointed out a key concept for this thesis: the absorption and emission of light by fluorophores are directional and polarisation dependent. We then discussed the confocal microscope and its
resolution. We argued that in practice the resolution of a confocal microscope is not better than a wide-field microscope. Of much more importance is that a confocal microscope provides optically sectioned images. We closed with a mention of multiphoton and superresolution techniques. Some superresolution techniques can claim theoretically unlimited resolution, though in practice this is not achievable, particularly in living samples.
3 Spatial light modulators

Our microscope has at its heart a binary ferroelectric liquid crystal spatial light modulator (FLCSLM) that is used to control and quickly change the polarisation of the excitation beam and correct system aberrations. In this chapter we provide an introduction to the types, uses and operating principles of spatial light modulators (SLMs), and we explain why the FLCSLM was chosen for our application.

An SLM is a programmable pixelated device that can be used to control the amplitude and phase of light beams [62]. SLMs are reconfigurable, so a system containing an SLM can exhibit different optical behaviour at different times. One important application of this technology is in astronomical adaptive optics where dramatic improvements in earth-based telescope performance have been achieved by correcting phase errors introduced by the Earth’s atmosphere [63]. In recent times there has been significant effort to apply adaptive optics to microscopy, though with only limited success [64–67]. It is likely that progress has been slow in microscopy due to significant sample-varying aberrations, the difficulty of finding a suitable guide star and the fact that many samples are more continuous than punctate. In addition the high NA of microscope objectives adds complications (see chapter 6). SLMs are often used in the excitation path of microscopes to display off-axis holograms, for example to steer beams [68,69] or engineer PSFs [70,71] or to make a series of reconfigurable optical traps [72–74]. SLMs have also been used in the detection path of microscopes, for example to shape the PSF and make it sensitive to phase gradients [75]. SLMs can be used to generate optical vortices, which are used in STED microscopy [76] and also in optical tweezers where the orbital angular momentum of these beams can cause rotation of small particles [77,78].
Recently, SLMs have been used to generate light beams with exotic polarisations, despite being phase-only devices (explained below). Exotically polarised beams are used in a wide range of areas [79], for example it has been shown that a radially polarised beam provides the most efficient cutting for laser machining [80]. Polarisation control could be incorporated into a holographic optical tweezer to allow orientational as well as positional control of small suspended objects [81]. In microscopy, control of beam polarisation is especially important because focal spots formed by high numerical aperture lenses have complicated focal polarisation patterns. Again the radially polarised beam is of particular interest here as its focal spot is polarised axially at the centre [82]. This has been used to determine the three-dimensional orientation of single molecules [83] and objects in biological specimens [84].

Before explaining how one can control the phase and polarisation of a light beam using an SLM, we review the types of SLM available and explain the principles of holography.

3.1. SLM types

The main commercially available SLMs fall into three main categories: digital micromirrors, deformable mirrors and liquid crystal displays. Digital micromirrors are binary amplitude devices that are made of an array of bistable mirrors. They can switch states at 30 kHz and have been used in a number of microscopy applications [85–87]. Deformable mirrors are made of a thin membrane behind which actuators move to control the membrane shape [88,89] and are commonly used in astronomical adaptive optics or in the detection path of a fluorescence microscope [90] because they can modulate incoherent light. In our work we use liquid crystal SLMs, and henceforth when we use the term SLM we mean a device of the liquid crystal type unless otherwise stated. A liquid crystal spatial light modulator is essentially a liquid crystal display designed specifically for precise wavefront control [91]. Liquid crystals (LCs) are a phase of matter that possess properties similar to both solids and liquids. LCs are
classified by the degree and type of order in the material. SLMs are usually made of nematic or smectic LCs. The nematic state is characterised by significant orientational order within the material but no positional order. In the smectic state the molecules are ordered directionally and arranged in layers but there is no positional order within layers [92]. In both types of LC the mean direction of the molecules is called the molecular director. The nematic state is interesting for the formation of phase delay SLMs because the orientational order makes the material anisotropic and therefore birefringent.

The most common type of SLM is the twisted nematic liquid crystal display (TN LCD). These devices are cheap and widely available as computer monitors, but are not suitable for applications where high precision is needed such as adaptive optics [75] because they modulate the polarisation and phase of the incident light simultaneously [93]. For example, the polarisation eigenstates for a TN LCD have been shown to be elliptically polarised, with the ellipticity depending on the voltage applied to the pixel [94].

The parallel aligned nematic SLM (PANSLM) is a type of spatial light modulator commonly used in adaptive optics and microscopy that gives continuous control of the phase delay applied to each pixel [91]. A PANSLM contains rod-shaped molecules whose director is constrained to lie in a single plane (see Figure 3.1). Varying the voltage applied to each cell causes the molecules to rotate by a calibrated amount, until at maximum applied voltage the molecules all line up along the light propagation direction. In this way the phase delay between x- and y- polarised light can be varied in an almost continuous manner [95]. The pseudocontinuous phase control allows high efficiency as gratings can be made blazed (see discussion on off-axis holography in section 3.2). The efficiency of an off-axis hologram is the ratio of power in the first diffracted order to total power at the output of the SLM.
3.1.1. **Ferroelectric spatial light modulators**

In this thesis we use a ferroelectric spatial light modulator (FLCSLM). The main advantage of an FLCSLM is that it can switch at very high rates, theoretically as fast as 30 µs [62] and in our device 40 µs (see chapter 8). The commercially available device that we use, the ForthDD SXGA-R3, can display up to 2040 frames per second. These fast switching rates make FLCSLMs useful in adaptive optics for real-time aberration correction [96] and optical tweezers for generation of many traps in parallel [72,73]. We use an FLCSLM so that we can change the polarisation state of the excitation light in our microscope on the same timescale as beam scanning.

The FLCSLM is based on liquid crystals in the smectic C* phase. In the smectic C phase, the molecular director is inclined at an angle $\sigma$ to the layer normal [92]. Normally the molecular director is the same in all layers of the material, but if the molecules are chiral, the direction of the director changes slightly between each layer, forming a helix over many layers. This phase is called smectic C*. The axis of the helix is parallel to the layer normal and the angle between the molecular director and the helical axis is always $\sigma$. The helix can be suppressed if the material thickness (in the direction normal to the layer normal) is restricted to roughly the helical pitch, about 15 µm [97], and in this case the material exhibits ferroelectricity [98].
The operation of an FLCSLM as a binary spatial light modulator is explained in Figure 3.2. Each molecule has a permanent dipole moment, \( p \), and is constrained to lie in one of two directions at \( \pm \sigma \) to the layer normal. By reversing the direction of an applied field \( E \), the molecules are switched between their two orientations [99]. Since the material is not isotropic it is birefringent so the action of applying a voltage to the pixel switches the fast axis between one of two positions. In ideal operating mode the phase depth of the pixel is \( \pi \) so that each pixel acts as a half-wave plate. This means that incident light has its polarisation state flipped about the fast axis direction. In this case, for linearly polarised incident light, there will be an angle of \( 4\sigma \) between the incident and output polarisation azimuths. The ForthDD SLMs have \( 4\sigma = 90^\circ \) (nominally) so when combined with a polarising element can be used as binary amplitude modulators in display applications. We describe the non-ideal operation of FLCSLMs in chapter 4.

We mentioned that the principle advantage of FLCSLMs is their speed of operation, (2 kHz frame rate in the ForthDD version that we use). The speed of operation is a result of the permanent dipole (nematic SLMs rely on induced dipoles). Because the FLCSLM is binary, achieving continuous phase control requires using off-axis holography to separate out the significant higher-order diffracted orders. This limits the efficiency of an FLCSLM to \( \epsilon = \frac{4}{\pi^2} = 41\% \), which is the maximum energy diffracted into the first order of a binary grating. In reality values much lower than this are measured because the switching angle
of the SLM is not ideal. We discuss this further and show some experimental results in chapter 4.

3.2. Holography

When used for wavefront shaping, it is often beneficial to display an off-axis hologram on an SLM instead of just directly modulating the phase. Holography was conceived by Dennis Gabor in 1948 [100]. A hologram can be concisely described as an interference pattern used as a diffraction grating. The purpose of a hologram is to simulate the presence of an object by generating the same propagating light field as the object itself would create [101]. Traditionally this would be done in two stages: recording and reconstructing. In the recording phase light from a reference wave is interfered with light scattered from the object. For this, a laser is needed with coherence length exceeding the path difference between source and reference wave. The plane of interference contains a photographic emulsion that records the interference pattern between these two waves. To reconstruct the object the original reference wave is used to illuminate the recorded interference pattern. The first diffracted order then contains the same field as would have been scattered by the object. A simple way to understand holograms is to consider a plane mirror. Here we are just trying to reconstruct an object that reflects one plane wave. The steps are shown in Figure 3.3. The interference between these two waves gives rise to an intensity distribution at the hologram:

\[ I(x) \sim 1 + \cos\left(\frac{2\pi}{\lambda} x (\sin \beta - \sin \alpha) + C\right), \]

Where \( \alpha \) and \( \beta \) are the angles that the reference and object waves make with the hologram normal and \( C \) is a constant accounting for the difference in optical path between the object and reference waves. The fringe spacing of this pattern, \( \Delta \), satisfies \( \Delta = \frac{\lambda}{\sin \beta - \sin \alpha} \). Comparing this with the well-known grating equation, \( \Delta(\sin \beta - \sin \alpha) = m\lambda \) [102], it is clear that when the reference beam illuminates the holographic plate in the absence of the object, the object wave is reconstructed after the hologram for \( m = 1 \). During reconstruction there are
three plane waves emanating from the hologram: the desired object wave in the first order, the zero order wave and the conjugate beam with \( m = -1 \). Now that we understand the two steps for plane mirror, we can imagine that the scattered field from an arbitrary object can be considered as a sum of plane waves.

![Diagram of hologram recording and reconstruction](image)

**Figure 3.3.** The hologram of a plane mirror. The left-hand figure shows the recording process. A coherent source of light is used to illuminate the screen directly (the reference wave) and illuminate a mirror. The waves illuminating the object then reflect from it and are incident on the screen. The sinusoidal intensity pattern is recorded on the screen. The right-hand figure shows the reconstruction process. The same reference wave is used to illuminate the hologram. It diffracts through the fringes on the screen, forming three beams after the hologram. The desired plane wave representing the object is in the +1 diffracted order.

For a more complicated object the first diffracted order covers a range of angles corresponding to the range of ray angles scattered by the object. In order to get good reproduction of the object, the first diffracted order must not overlap with any of the other diffracted orders. For this reason, holograms are usually constructed with the object off-axis [103]. The off-axis situation has important consequences when considering pixelated holograms that we discuss below. We now explain how to chose the angle of the reference beam on the screen. Let the reference wave be represented by \( R = r \exp(2\pi i x \xi_r) \) and object wave by \( U = u \exp(i\phi) \). Here \( r \) is the amplitude of the reference wave, \( x \) is the screen coordinate and \( \xi_r = \frac{\sin \alpha}{\lambda} \) is the spatial frequency of the reference wave. Let \( \xi_{max} \) be the maximum spatial frequency of the object. During the recording step, the
field at the screen is \( E_0 = U + R \) and the intensity is \( I_0 = E_0E_0^\ast \). During the reconstruction, the field transmitted by the hologram is \( E_0 = RI_0 \). This gives

\[
E_0 = r^2R + u^2R + R^2U^\ast + r^2U ,
\]

where the terms are: (1) directly transmitted field; (2) zero order of diffraction pattern; (3) -1 order and (4) the object field. The zero order has angular width \( \pm 2\xi_{\text{max}} \) by the autocorrelation theorem [11] and is centred at \( \xi = \xi_r \). The object field has angular width \( \pm \xi_{\text{max}} \) and is centred at \( \xi = 0 \). Therefore we require

\[
\xi_r > 3\xi_{\text{max}} .
\]  

If \( u = 1 \) then the zero order has no spatial extent, but this does not remove the limit on \( \xi_r \) because then we must consider separating the second and the first order. This consideration gives the same limit on \( \xi_r \) [101,104]. (31) is a lower limit on \( \xi_r \) that is equivalent to a lower limit on the number of fringes on the hologram.

![Figure 3.4. Pixellation and aliasing. On the left we show an example diffraction pattern from a pixelated square wave grating. The zero order is at the centre of each square and the first order is shown in red. The order intensity is encoded in the gray level of each order, with white indicating very weak. On the left we show the effect of sampling and aliasing: the pixels have spacing \( a \) in frequency space, which causes a periodicity in the diffraction pattern with periodicity \( a^{-1} \). On the right we show the effect of finite sized pixels, which reduce the intensity of the diffracted light outside the first zone with width \( 2\kappa_a = a^{-1} \). Here we have chosen the pixel size to be roughly the same as the pixel pitch.](image)

For a pixelated hologram the pixel pitch sets another limit on \( \xi_r \) because of aliasing. The effect of pixilation is described by multiplication of the ideal fringe pattern with a comb of delta functions representing the pixel spacing, then
convolving with a rectangular function representing the pixel shape. If the pixel spacing is \( a \), then all diffracted orders of the hologram will lie inside the zone in spatial frequency space with bounded by spatial frequency \( \xi_a = \pm 1/2a \). The other diffracted orders appear inside this zone at aliased positions. This aliasing effect is shown on the left in Figure 3.4. Aliasing sets a lower limit on the fringe size, which is equivalent to an upper limit on the number of fringes on the hologram (for straight fringes). Aliasing also complicates the separation of diffracted orders, which can lead to problems in optical testing where the phase in the isolated first order is important [105]. The effects of aliasing can be mitigated by choosing \( \xi_r \) to be well below \( \xi_a \). In our case we often use roughly 60 fringes on a hologram with 600 pixel diameter. Our holograms are two-dimensional, which gives us a choice of the \( x \) and \( y \) tilts, say \( \xi_x \) and \( \xi_y \). We choose the two tilts such that \( \xi_x / \xi_y \) is not almost equal to a simple rational number (a number \( p/q \) with \( p \) and \( q \) integer and both small but not zero), which ensures that if there is any overlap between the first and higher orders, the higher orders will be very weak at the position of the first order (the order intensity decreases as \( n^{-2} \) for order \( n \) [106]).

It is not only the pixel pitch that affects the performance of pixelated holograms: the pixel size, say \( b \), is important too. In frequency space, the the effect of pixellation is to convolve the hologram frequency spectrum with a comb of delta functions then multiply it by a sinc function. The width of the sinc function is proportional to \( b^{-1} \). Therefore the pixels low-pass filter the hologram output so it is not advisable to let either \( \xi_x \) or \( \xi_y \) approach \( \xi_b = b^{-1} \). The effect of the finite pixel size on the diffraction pattern is shown on the right in Figure 3.4. The diffraction efficiency \( \epsilon \), defined as the percentage of light that ends up in the first order divided by the total diffracted light, increases as \( \epsilon \propto 1 - m^{-2} \) [107], with \( m \) as the number of pixels per fringe, so for efficient reproduction of the object field, we require at least a handful of pixels per fringe in the hologram. PANSLMs only provide a discrete set of allowed phase levels \( (n_L) \), for example the Holoeye HEO 1080p PAL SLM has \( n_L = 256 \) [108]. The effect of this can be modelled as a quantisation noise and reduces the diffraction efficiency as \( \epsilon \propto 1 - n_L^{-2} \) [109].
To see how these two effects combine to limit the application of an SLM, consider trying to replace a lens with an SLM. What is the most powerful lens we could make? The most challenging part of the hologram to form is where the phase gradient is highest, which is near the hologram edge. For a circular hologram of diameter $D$, the maximum phase gradient is $\omega_{\text{max}} = \frac{\pi D}{\lambda f}$ where $f$ is the focal length of the converging object beam. Given the requirement for the fringe tilt, the minimum fringe size is $\Delta = \frac{\lambda f}{3D}$. Therefore the minimum number of pixels per fringe is $m = \frac{\Delta}{a} = \frac{\lambda f}{6Da}$. We said that we require a handful of pixels per fringe for reasonable diffraction efficiency, so if $m = 5$ then the shortest possible focal length is

$$f = \frac{6 Da}{\lambda} = \frac{30 Da}{\lambda}. \quad (32)$$

Sensible parameters are $D = 10$ mm, $\lambda = 0.5 \mu m$, $a = 15 \mu m$, giving a minimum focal length of 9 m; not a very powerful lens. Therefore SLMs, with their current pixel pitches, cannot be used to replace large lenses in optical systems. Since the shortest possible focal length is proportional to $D$, SLMs have been used to replace microlens arrays, for example in endoscopy [110].

All the holograms used in this thesis are computer-generated, that is to say that the recording step is simulated in a computer [109] and the interference pattern is computed and displayed on an SLM. To use the SLM we effectively control the voltage of each pixel which in turn modulates the phase of the incident light by changing the refractive index and therefore the optical path length through the pixel. This means that the SLM is displaying phase holograms instead of the traditional intensity holograms discussed above. With a PANSLM, which has 256 phase levels, we can blaze the hologram, which should in theory direct almost all light to the first order. In practice there is always some zero order and some higher orders observed.
3.3. Phase and polarisation control with SLMs

We have seen that off-axis computer-generated holograms allow the simulation of light fields. SLMs were first used to directly introduce variable phase delay into a wavefront [91] but this method is not ideal because the corrected wave contains sharp phase edges. Off-axis holograms are often preferable to direct phase modulation because the first order can be separated from the other diffracted orders. In the case of a blazed phase grating displayed on an SLM, there will not be a zero order of the same form as in equation (30) but there is a zero order term from the unmodulated light that may result from the incomplete coverage of the SLM pixels over the entire reflective area. Using an off-axis hologram has another important benefit: because the diffracted orders are spatially separated and sharp phase jumps are high frequency features, the smoothly-varying first order can be isolated. For the same reason, the off-axis hologram has the added benefit that small nonlinearities in the response of the spatial light modulator do not affect the overall wavefront shape: they only degrade the efficiency with which light is diffracted into the first order.

We have seen how continuous phase holograms can provide wavefront control, and now we turn to looking at how to control the amplitude of the transmitted light in addition to the phase. For some applications it is desirable to be able to modulate the amplitude as well as the phase of the transmitted field. Schemes giving full complex control to the diffracted field using SLMs are complicated [111], but a useful subset of this full complex control can be achieved with a phase-only SLM. This is reasonably simple when using a continuous phase SLM, and it is surprising that it took until 1999 for a solution to appear in the literature [112]. Instead of simply controlling the position of each fringe in the grating, the fringe modulation depth is also controlled. Mathematically, we multiply the original wrapped phase function \( \phi_w(x, y) \) by a modulation function \( M(x, y) \). Crudely this could be thought of as changing the diffraction efficiency of each fringe, so when we isolate the first order we will get a wavefront with shape \( \phi(x,y) \) and amplitude controlled by \( M(x,y) \). This method relies on \( M \) having a much lower maximum spatial frequency than \( \phi_w \), so
that the fringe positions are not significantly moved by the multiplication step. We will see examples of amplitude modulation in chapter 4. A note of caution: the modulation offered is not exactly $M$, so if very precise amplitude control is needed, further work should be undertaken to determine the required modulation depth for a required diffracted amplitude. We do not discuss this further here as all of the holograms used in this thesis are binary and do not suffer from this complication.

![Figure 3.5. Double-pass, common path scheme for controlling the polarisation of a light beam using two holograms on a single SLM. The operation is described in the text. In the setup shown, hologram A is imaged onto hologram B via a reflective 4F system.](image)

There have been many papers about using phase only SLMs to control the polarisation of light beams [113–119]. At heart most techniques start with one beam, divide it and independently control the amplitude and phase of two orthogonal polarisation components, and then recombine the beams. The most obvious way to achieve this is to use linearly polarised light incident on a beamsplitter to form two incident beams, reflect both beams from a PANSLM, use a half-wave plate to flip the polarisation state of one beam, then recombine the beams. This method was used in previous work in our research group [70,71,120,121], but suffers from significant alignment difficulties. A previous PhD student reported spending more than a day every week aligning and realigning this part of his microscope [121]. A recent paper by Moreno et al. showed a particularly elegant full polarisation control method using a reflective PANSLM [122]. The idea, probably first due to Love [123], is as follows. See the layout in Figure 3.5. A light beam, propagating along $z$, is incident on a PANSLM. The SLM only modulates the polarisation component of an incident beam that is parallel with the molecular director, say along $x$. The light beam is prepared so that half of its energy is polarised along $x$, and half perpendicular to $x$. For
example, it could be circularly polarised, linearly polarised at 45° to \(x\) (as shown in the figure) or have \(x\) and \(y\) components that are mutually incoherent (as we used in our STED microscope [76,120]). Now, the light polarised along \(x\) is incident on an off-axis computer generated hologram displayed on the SLM, labelled hologram A. Hologram A directs \(x\) polarised light to its first order, and \(y\) polarised light reflects to the zero order. Let the zeroth order of hologram A be denoted \(A_0\), and the first order \(A_1\). Hologram A is then imaged onto a second hologram, hologram B, through a reflective afocal system, and in the process passes through a quarter wave plate twice. This flips the polarisation of the unmodulated light, \(A_0\) to be along \(x\) and the modulated light, \(A_1\), to be along \(y\). Therefore after reflecting from B, \(A_1\) is not modulated further, so ends up in the zeroth order of B, \(B_0\). We label this beam \(A_1B_0\). \(A_0\) however is now directed to the first order of B, giving \(A_0B_1\). \(A_1B_0\) and \(A_0B_1\) are orthogonally polarised, and each has only been modulated by one hologram. If the fringe tilts on A and B are equal, full polarisation control in the common first order is realised. This is an elegant and practical method of controlling the polarisation. In 2011 we changed the polarisation control scheme in our STED microscope (see [76]) from the beamsplitter-based method to Moreno’s method [76]. Since the change we have noticed a dramatic reduction in maintenance times: we almost never need to realign the system and when we do, the process is fast. The advantage of this double path method is that the only splitting and recombining of the beam is done diffractively by the SLM. In addition, the recombined beam is made up of components that have travelled almost exactly the same path through space. We refer to it henceforth as a double-pass, common path technique. In chapter 4 we describe a double-pass, common path method for controlling the polarisation of a light beam using an FLCSLM [124]. This is incorporated in our microscope and described later in this thesis. In once recent paper, Kenny et al. argued that the Moreno method does not give full polarisation control (i.e. full coverage of the Poincaré sphere) [119] on the basis that only the relative phases of the two orthogonal polarisation components are modified. If using off-axis holograms then Kenny et al. are incorrect because the diffracted amplitudes can be can be controlled by the method described above.
Controlling the phase and polarisation of light beams in the excitation path of a confocal microscope has become known as PSF engineering, or vector PSF engineering in the literature. Work has been done on PSF engineering in many areas, for example to correct system aberrations [65,88], control the polarisation of the excitation focal spot [125] and to generate exotic double helix PSFs to give depth information in localisation super-resolution microscopy [126–128].

3.4. Chapter summary

In this chapter we have provided an overview of the types and uses of SLMs. In microscopy, SLMs have seen particularly wide usage in aberration correction, PSF engineering and in optical tweezers. We described the main types of SLM, with particular focus on liquid crystal SLMs. We described the operation of the PANSLM, which allows phase control of light polarised parallel to the molecular director while leaving the other polarisation component unmodulated. This is achieved by varying the effective refractive index of the material. We also described the ideal operation of the FLCSLM, which provides the phase control used in this thesis. The non-ideal operation will be discussed in detail in chapter 4.

We then explained how holography works, paying particular attention to off-axis holography. We showed that for the object beam to be well-separated from the other orders, the fringe tilt must be greater than three times the maximum spatial frequency in the object. This applies for amplitude holograms, but SLMs display phase holograms. When using a PANSLM, the pseudocontinuous phase control allows for blazed holograms to be generated which in theory means that all light should be directed to the object beam. In practice there is always some residual light in the zero order due to the incomplete fill factor of the pixels, and in the higher orders due to nonlinearities in the phase-voltage response, the finite number of phase levels and the finite number of pixels per fringe in the hologram. This means that the fringe tilt limit still applies.
We then combined the ideas of holography and SLM operation to describe a recent technique that gives full control of the phase and polarisation of light in an elegant manner. All polarisation control methods with phase-only SLMs use two holograms to independently modulate orthogonal polarisation components. Previous methods are difficult to align and maintain because they rely on extra optical components to split and recombine the beam. The method described above uses the SLM itself to perform the splitting and recombination while independently controlling the phases of the orthogonal polarisation components. We have implemented this technique on our STED microscope (not described in this thesis) and have seen vast improvements in alignment and realignment times [76]. The success of this double-pass, common path technique led us to seek a similar technique with an FLCSLM (see chapter 4).
4 Polarisation control using an FLCSLM

In this chapter we present a new method for controlling the polarisation state of a light beam using two holograms encoded on a single FLCSLM. This work led to a publication in Optics Letters in 2013 [124]. We use this technique to encode linear, circular, radial and azimuthal polarisations into the pupil of our microscope and to correct system aberrations. Using one SLM presents an advantage because the two holograms are automatically synchronised. This is important when we incorporate our polarisation control scheme into a beam-scanning microscope.

In chapter 3 we explained the steps in achieving full polarisation control of a light beam using a grayscale PANSLM. First, we explained holography and how a blazed grating leads to efficient diffraction of programmable light with phase $\phi(x, y)$ (if $(x, y)$ are spatial coordinates in the hologram plane) into the first diffracted order. Next, we described how multiplication of $\phi$ by $M(x, y)$, where $M$ has a significantly lower spatial frequency than $\phi$, gives amplitude control in the first order. Then, on combination of two orthogonally polarised beams, each having its own $\phi$ and $M$, we can realise full polarisation control in the first diffracted order. We also explained in chapter 3 that to give a polarisation controlled beam, we must start with a single beam, somehow split it, then independently control the $x$ and $y$ polarisations, then recombine the beam. We described an elegant method [122] that uses the SLM itself to perform the splitting and recombination, which significantly simplifies alignment and adjustment of the scheme in practice. The scheme works because a PANSLM only modulates one polarisation component of an incident light beam [123].

This chapter describes our development of a sibling double-pass, common path scheme but for use with an FLCSLM. This is not a trivial extension of the PANSLM scheme because an FLCSLM modulates all polarisation components. Achieving phase and amplitude control with an FLCSLM requires conversion
from \( M(x,y) \exp(i \phi(x,y)) \) to the binary phase values \{0, \pi\}. These conversion schemes are termed binarisation schemes, and are discussed in section 4.2.

After describing the operating principles, we present experimental results demonstrating the formation of radially, azimuthally and circularly polarised beams at low numerical aperture. We also evaluate the performance of our system in terms of the degree of polarisation of the output and measure the influence of aberrations introduced by the SLM itself, which doesn't have an optically flat surface, and by the 4F imaging system. We focus on the production of azimuthal and radial beams, otherwise known as cylindrical vector beams because of their properties when focused at high NA in a microscope. Cylindrical vector beams have many other uses [79]: in optical trapping, where a radially polarised beam has been used to trap metallic spheres [129]; laser machining, where a radially polarised beam was shown to double the cutting efficiency [80,130]; and in plasmonics where the strong p-polarised component can couple efficiently to surface plasmon modes [131,132]. Cylindrical vector beams can be generated in many ways other than using an SLM, for example by inserting an interferometer into a laser cavity [133], or by using a segmented half wave plate [134], or by using a Mach-Zender interferometer [135,136]. However, using an SLM is necessary when the beam must be reconfigurable.

### 4.1. Non-ideal operation of FLCSLMs

Previously we described the ideal operation of an FLCSLM with linearly polarised input light as a half wave plate with switchable fast axis. A given SLM is only designed for half-wave retardation at one wavelength, so here we generalise this description to any wavelength and also to any coherent input state. The modulation offered by an FLCSLM is well approximated by considering each pixel to be a wave plate with retardation \( \delta \) and fast axis switchable between two distinct orientations. The wave plate fast axis orientations are at angles \( +\theta \) or \( -\theta \) to their bisector depending on the pixel state and are both in the plane of the SLM. A waveplate with fast axis at \( \theta \) to \( x \) and with retardation \( \delta \) can be described by the Jones matrix.
\[ P(\delta, \theta) = \begin{pmatrix} a & b \\ b & a^* \end{pmatrix}, \]  
\[ \text{with} \]
\[ a = e^{-i\frac{\theta}{2}} \cos^2 \theta + e^{i\frac{\theta}{2}} \sin^2 \theta, \text{ and} \]
\[ b = -i \sin 2\theta \sin \frac{\delta}{2}. \]  
So the SLM pixels are represented by
\[ P_\pm(\delta, \theta) = \begin{pmatrix} a \pm b \\ \pm a^* \end{pmatrix}, \]  
with the sign flipping depending on the state of the pixel. This follows because \( a \) is even in \( \theta \) and \( b \) is odd in \( \theta \). The state of the pixel is denoted \(+/-\) when the fast axis is at \( \pm \theta \) to the \( x \) axis. Now, consider illuminating the SLM with the input state \( E_0 = (E_x, E_y) \). The output state \( E_1 \) will be one of two values corresponding to the pixel value, giving
\[ E_{1,\pm} = \begin{pmatrix} aE_x \\ a^* E_y \end{pmatrix} \pm b \begin{pmatrix} E_y \\ E_x \end{pmatrix}. \]  
Now, any grating displayed on this SLM will have a fraction \(|a|^2 = (1 - \eta)\) of the light unmodulated and in the zero order, with a fraction \( \eta = |b|^2 \) modulated. Of the diffracted light, a fraction \( \frac{4}{\pi^2} \) is in the first order, giving a diffraction efficiency
\[ \epsilon = \frac{4|b|^2}{\pi^2} = \frac{4\eta}{\pi^2}. \]  
The SLM used in this work is a ForthDD SXGA-R3 and has \( \theta \approx 17.5^\circ \) and \( \delta = \pi \) for \( \lambda = 550 \) nm. In our work we use \( \lambda = 532 \) nm, giving \( \eta = 0.31 \) and \( \epsilon = 0.13 \). Therefore there is significant unmodulated light in the zero order that can be reused in a second diffraction step. It is well known that FLCSLMS suffer from poor diffraction efficiency. In previous work we showed how to increase the diffraction efficiency of an FLCSLM by nearly a factor of 3 [107,137] by using a double pass to simulate a larger switching angle, but this relies on pixel-level imaging of first hologram onto second and so we only managed to demonstrate it with small and very simple holograms (for example straight fringe patterns). With careful custom lens design this could be improved.
The normal method of operation of an FLCSLM for phase modulation is to use linearly polarised light polarised along the bisector of the two fast axis directions [138]. This gives \( E_0 = (E_x, 0) \) and so \( E_1 = \left( \frac{a}{\pm b} \right) \). Therefore the diffracted light and undiffracted light are of orthogonal linear polarisations and can be separated with a polariser. Before we describe the operation of our polarisation control scheme, we explain how to use binary holograms to control the diffracted amplitude and phase of a light beam.

4.2. Binary holograms

Suppose we want to create a field \( u(x, y) = M(x, y) \exp(\imath \phi(x, y)) \) using a binary spatial light modulator. With a grayscale hologram, we add a linear phase tilt \( \tau \) to \( \phi \), giving the required phase function to be displayed on the SLM \( H = M \cdot (\phi + \tau) \). The tilt separates the zero, first and higher diffracted orders, with \( u \) present in the first order. An FLCSLM displays binary phase holograms, so we must convert \( H \) to a binary function \( H_b(x, y) \) in the SLM plane that gives \( u \) in the first order. First consider the simple case of \( M = 1 \), i.e. where no amplitude control is needed. For the grayscale SLM we would have \( H = (\phi + \tau) \).

Converting to a binary hologram can simply be achieved by mapping \( H \) to \( H_b \) such that \( H_b = +1 \) if \( \cos(\phi + \tau) > 0 \) or \(-1\) otherwise [106]. This hologram has \( \phi \) in its first order, and higher odd orders which result from the sharp edges of the binary function.
Figure 4.1. The double-D binarisation scheme. Varying the fringe width $\alpha$ gives control over the diffracted amplitude. On the right is the binarisation disc, which represents the mapping from the desired complex value $M \exp(i\phi)$ to pixel value in the binary hologram. The boundary shape is found by traversing the complex plane at constant $M$ and ensuring that boundary of the $+1$ and $-1$ regions matches the fringe width of the chart on the left.

Now, what about controlling the diffracted amplitude? The most obvious approach is to modify the fringe width in the binary hologram according to the required diffracted amplitude. To work out the binary pixel value at a given point in the hologram given a desired phase and amplitude, we investigate the first term in the Fourier series of this function. The square wave pattern (simplified to one dimension) is represented by

$$DD(x) = \begin{cases} 1, & 0 < |x| \leq \alpha \\ -1, & \text{otherwise} \end{cases} \quad \text{with } DD(x + 2\pi) = DD(x). \quad (38)$$

Expressed as a Fourier series, this becomes

$$DD(x) = \frac{2}{\pi} \left( \alpha - \frac{\pi}{2} + \sum_{k \neq 0} \frac{1}{k} \sin(k\alpha) \exp(i\ k\ x) \right). \quad (39)$$

So the first diffracted order has amplitude

$$c_1 = \frac{2}{\pi} \sin \alpha \quad (40)$$

which illustrates that changing the fringe width changes the diffracted amplitude. So given a desired $M$ (remembering that $0 \leq M \leq 1$), we pick the fringe width $\alpha = \sin^{-1} M$ to give us our required diffracted amplitude. $DD(x)$ is shown on the left in Figure 4.1. The right-hand half of Figure 4.1 shows the mapping from the unit disc in the complex plane (i.e. the required amplitude and phase) to the binary pixel value $+\ or -\$, and is known as a binarisation disc. We call the particular choice of function shown in Figure 4.1 the Double-D
binarisation scheme. The idea is described by Lee [109] and is implemented in several more recent papers [71,138].

The Double-D binarisation scheme is not the only binary function that gives control of the diffracted amplitude. Many binary functions with an adjustable parameter have variable amplitude in the first order. There are some restrictions on the functions we could use. We must not change the phase of the first order coefficient as we change the adjustable parameter. Even functions satisfy this requirement because they have only real Fourier coefficients. An even function has a binarisation disc that is the same when reflected about the real line. In our polarisation control scheme, described below, we reuse the undiffracted light in the zeroth order of the hologram, and so we don’t want the zero order amplitude to change as the fringe parameter is adjusted. Therefore we seek a binarisation scheme having no zero order, i.e. a function whose average value is always zero. We also require the common first order of both holograms to contain only $A0B1$ and $A1B0$; here $A_i$ labels the $i$th diffracted order of $A$ and similarly for $B$. A binary hologram with only odd orders satisfies this requirement. So we seek a binarisation scheme whose binarisation disc is odd under inversion about the imaginary axis. The simplest function satisfying these requirements is

$$AA(x) = \begin{cases} 
1, & 0 < x \leq \alpha \\
-1, & \alpha < x \leq \frac{\pi}{2} \\
1, & \frac{\pi}{2} < x \leq \pi - \alpha \\
-1, & \pi - \alpha < x \leq \pi
\end{cases} \quad (41)$$

with

$$AA(-x) = AA(x), AA(x + 2\pi) = AA(x).$$

We call this the ‘alien’ binarisation scheme. The first order diffracted amplitude is given by

$$c_1 = \frac{2}{\pi} (2 \sin \alpha - 1) \quad (42)$$

There is no first order when $\alpha = \frac{\pi}{6}$ and the first order has maximum intensity when $\alpha = 0$ or $\frac{\pi}{2}$. This gives two variants of the binarisation disc with a phase shift of $\pi$ between them. The two discs are shown in Figure 4.2. The variant on
the right is preferable because we can adjust $\alpha$ over twice as large a range to vary $c_1$, so the amplitude control is finer when dealing with pixelated holograms.

![Two variants of the alien binarisation scheme](image)

Figure 4.2. The two variants of the alien binarisation scheme. Both of these functions give no phase change as $\alpha$ is adjusted (symmetric about real axis); have no even orders (anti symmetric about imaginary axis); and have no zero order (equal area of black and white). In previous work in this group the left-hand version was used [139] but we use the right-hand version.

### 4.3. Polarisation control scheme

![Experimental setup to obtain full polarisation control of a diffracted beam in a double-pass, common path setup using an FLCSLM](image)

Figure 4.3. Experimental setup to obtain full polarisation control of a diffracted beam in a double-pass, common path setup using an FLCSLM. The input state is prepared with a waveplate that needs at least $\pi/2$ or $\lambda/4$ retardation. The diffracted orders from hologram A are polarised before the light is imaged back onto hologram B. Full polarisation control is realised when the fringe tilts on A and B are equal.
Figure 4.4. Sketch of the diffraction from two binary holograms, A then B, with orthogonal straight fringes. After hologram A we place a polariser to make all diffracted orders have equal polarisation state. After hologram B, which can be considered to be in normal FLCSLM operation mode, the zeroth orders of B (whole red line) and the diffracted orders of B (in blue) are of orthogonal linear polarisations.

Here we describe how to form two independently modulated outputs with orthogonal polarisations using an FLCSLM by passing light through two holograms A and B in succession. We image hologram A to hologram B using a reflective 4F system shown in Figure 4.3. Achieving the independent complex control of two orders is not as simple as in the case of the PANSLM because there is no unmodulated polarisation state for an FLCSLM. We consider the zeroth order of A followed by the first order of B, A0B1, and the first-order of A followed by the zeroth order of B, A1B0. A0B1 and A1B0 will have orthogonal polarisations if A0 and A1 have the same polarisation state and B0 and B1 have orthogonal polarisation states (or vice versa). Achieving the condition of orthogonal polarisation in B0 and B1 is trivial, as this is the mode in which FLCSLM devices are commonly used for phase modulation (described in section 4.1 above): if the input light is linearly polarised along the FLCSLM state bisector, then the output B0 will be linearly polarised in that same direction and B1 will be polarised in the orthogonal direction.

To obtain states of equal output polarisation after A is slightly more complicated. First, we write down the general expression for the output polarisation state after the SLM, equation (36):

\[
E_{1,±} = \left( \frac{aE_x}{a^*E_y} \right) \pm b \left( \frac{E_y}{E_x} \right).
\]
Now let us pass this state through a polariser oriented along $x$:

$$E_{4,\pm}^{(p)} = \begin{pmatrix} aE_x \pm bE_y \\ 0 \end{pmatrix}. \quad (44)$$

Remembering that $\eta = |b|^2$, we can see that if we chose an input state with $|E_x| = |E_y|$ then the unmodulated light will have strength $(1 - \eta)$ and the modulated light will have strength $\eta$. So we have obtained a linearly polarised output from A with the same ratio of intensities in A0 and A1 as will be generated by B in its orders B0 and B1. The diffraction pattern from two simple gratings with orthogonal fringes is shown in Figure 4.4. To achieve full polarisation control in a single diffracted order, we then simply overlap the orthogonally polarised A0B1 and A1B0 by choosing equal fringe spacing and orientation on A and B.

The required input condition $|E_x| = |E_y|$ is satisfied for any polarisation with an azimuth at an angle of $\pi/4$ to the FLC state bisector. This makes the scheme easy to setup, as a particular input state (point on the Poincaré sphere) is not required. We lose 50% of the light due to the polariser, but a more complex scheme can be realised as for one specific elliptical polarisation oriented at $\pi/4$ to the state bisector, both possible pixel outputs after A have the same polarisation but with opposite handedness to the input. In this case, all diffracted and undiffracted orders will have this polarisation too. This removes the need for a polariser between A and B but adds complexity because two wave plates are required to generate the specific input elliptical polarisation and two wave plates are required to convert the output from A to the linear input required for B. We discussed this further in previous work [107].

The system is aligned as follows. First, arrange a linearly polarised input along $x$ in the lab frame and display a simple grating on the SLM. From equation (36) we can see that the output states will only be linearly polarised if the SLM state bisector is along or perpendicular to $x$. Our SLM is mounted on a rotation mount. We place a second polariser to transmit only $y$-polarised components. Observing the diffraction pattern from the first hologram, we rotate the SLM to null the zero order. It is important to set the rotation of the SLM first because
this shifts the hologram positions and so the phase differences between the outputs (see section 4.5). Next we prepare the input state to have half its energy along and half perpendicular to $x$. We insert a waveplate before A with at least $\frac{\pi}{2}$ retardation. We set holograms A and B to have slightly different fringe tilts so the diffraction pattern has two adjacent dots in the first order. We then rotate the waveplate until the two first orders have equal intensity. As with the scheme for the PANSLM, this method will also work with unpolarised input light, producing two independently modulated outputs of orthogonal polarisation that are mutually incoherent.

Since $A0B1$ and $A1B0$ are both formed from the zeroth and then the first order of two binary holograms, they each have strength $\epsilon = 4\eta(1 - \eta)/\pi^2$. Overlapping them gives a maximum efficiency of $8\eta(1 - \eta)/\pi^2$ for a circularly polarised or a linearly polarised beam at $\pm45^\circ$ to $x$. In reusing the light, we make the performance of this scheme optimal for $\eta = 0.5$. For the switching angle and wavelength used here, $4\eta(1 - \eta)/\pi^2 = 8.7\%$.

### 4.4. Measured aberrations in the system

Hologram A is imaged onto B using a reflective 4F system. The lens is an achromatic doublet with a 30 mm clear aperture and 200 mm focal length. Between the lens and the mirror is a 1 cm polarising beamsplitter (PBS) cube used as a polariser. We measured the aberrations using a Zygo phase-shifting interferometer. The reconstructed wavefront is shown in Figure 4.5. The RMS wavefront amplitude is $0.03\lambda$ ($\lambda = 633$ nm), with the main contributions from $-0.39\lambda$ defocus and $0.33\lambda$ spherical (individual contributions are reported as peak-valley amplitudes). The only relevant portion of the wavefront is where the beam from the hologram passes through, which is indicated in white circles of diameter 8 mm on the plot. The wavefront just appears tilted in these regions, and has an RMS aberration of less than $0.01\lambda$ (not resolved by the instrument). Therefore we can conclude that the 4F system does not introduce significant aberration.
Figure 4.5. Aberrations in the 4F relay from hologram A to B. Reconstructed wavefront measured by phase-shifting Zygo interferometer. The diameter of the wavefront is 30 mm, equal to the clear aperture of the focusing lens in the 4F system. The main aberration visible in the wavefront is spherical (balanced by defocus). The white rectangle shows the relative size of the SLM, and the white circles show the areas covered by the holograms.

The SLM surface itself is another source of aberrations. We measured the shape of a wavefront reflecting directly from the SLM and show the results in Figure 4.6. The wavefront amplitude is twice the physical surface height. Over the whole surface the RMS wavefront error is $0.22\lambda$. Over the area of one hologram the wavefront error is $0.04\lambda$, with contributions of $0.13\lambda$ defocus and $0.06\lambda$ astigmatism (again contributions are reported as peak-valley amplitudes). Therefore we can conclude that the whole system should reach diffraction-limited performance.

Figure 4.6. Surface flatness of ForthDD R3 SLM (display J09904). Reconstructed wavefront measured by phase-shifting Zygo interferometer.
4.5. Experimental results

To demonstrate the validity of our method we focused and magnified the output beam from the SLM onto a fluorescent slide and then imaged the slide onto a CCD camera. The focusing was performed at low NA so that the focal field had the same polarisation as the output beam from the holograms [140]. The input beam used was a filtered DPSS laser at 532 nm, expanded to 10 mm beam diameter. The ForthDD SLM has an active area 17.5 by 14 mm, allowing for a maximum circular hologram radius of 4.7 mm. We use a radius of 3-4 mm to allow space for a pick-off mirror between input and output beams.

Figure 4.7 shows the images of the focal spot of a circularly polarised beam, produced by setting A and B to be simple gratings with a phase difference of $\frac{\pi}{2}$.

Figure 4.8 shows the holograms used to generate a radially polarised beam, and Figure 4.9 shows the focal spots of a radially polarised beam and an azimuthally polarised beam. A radially polarised beam has linear polarisation everywhere and polarisation vector pointing along the beam radius. An azimuthally polarised beam is linearly polarised everywhere but with polarisation vectors at right angles to the beam radius.

To set a given polarisation state at the output, we need to specify the phase difference between the two holograms. We find that giving both holograms equal phases does not result in zero phase difference between the output beams in the common first order. This is most likely because the image of A cannot be perfectly aligned to B. To see why a misalignment gives a phase difference, consider that the phase difference between the beams can be adjusted from 0 to $2\pi$ by adjusting the fringe positions. Therefore if the beam and either hologram are not perfectly aligned, there will be a phase difference between the two output states. To get an estimate of the magnitude of this effect, consider a hologram with 600 pixel diameter. We can position holograms to approximately
5 pixel accuracy (70 µm). A typical hologram has 60 fringes across its diameter, so the misalignment moves about 0.5 fringes. This corresponds to a phase error of \( \pi \). Though this seems like a great sensitivity of the output state to the hologram alignment, we observe the phase offset between A and B to be stable over several months.

In order to find the offset phase \( \Delta \phi \) between A and B we set both holograms to have straight fringes with equal fringe spacing and adjust the hologram phase difference until we get a null with a 45° polariser: this indicates that we are producing light polarised at \(-45°\) so the phase difference between A and B is 0 or \( \pi \). We find \( \Delta \phi = -0.65\pi \). We then set the holograms to produce radially polarised light, and rotate an analyser through the focal spot. If the lobes rotate in the same sense as the analyser, then we have produced a radially polarised beam. Otherwise, the lobes will rotate in the opposite sense to the polariser and then we need to add \( \pi \) to \( \Delta \phi \). We measure the diffraction efficiency for an \( x \)-polarised beam be \( \epsilon = 8.4\% \pm 0.2\% \) (should be 8.7\% in theory). The diffraction efficiency here is defined as the fraction of energy in the first diffracted order of the total energy in all diffracted orders, rather than of the incident energy. With the SLM off, we measured the fraction of incident light in the reflected zero order to be 25\%. This is low because of the combined effects of the native reflectivity of the SLM (doubly important as there are two SLM reflections), the pixel fill-factor and the 50\% loss due to the beamsplitter. So, combining diffractive and reflective losses, we only get 2.1\% of the incident energy in the first diffracted order.

![Figure 4.7](image.png)  
**Figure 4.7.** Low NA focal spot of a circularly polarised beam. The brightness has been adjusted separately for the image without polariser.
Figure 4.8. Holograms used to generate radially polarised light. The regions with higher fringe spacing correspond to regions of low amplitude diffracted to the first order: instead, light is being diffracted to the third order.

Figure 4.9. Low NA focal spots of radially (upper) and azimuthally (lower) polarised beams.

We measured the degree of polarisation using an optical power meter and a rotating analyser. The simplest beam is the $x$-polarised beam because it is generated by just one hologram. We measured the degree of polarisation in this case to be $1000 \pm 700$ (given as the ratio of maximum to minimum intensity measured as the analyser rotates). There is a large error on this result because the intensity at null is dominated by noise. We also measured the degree of polarisation of a linearly polarised beam at 45° to $x$. This beam is generated by setting holograms A and B to have the same tilt but with a phase difference $\Delta \phi$ (which must be set as described above). We measured the degree of polarisation of a 45°-polarised beam is $110 \pm 30$. This is limited by the accuracy by which we can set $\Delta \phi$. To get an idea of how accurately we have set $\Delta \phi$, we compare the ellipticity of the polarisation state with Jones vector. The measured ellipticity of roughly 10% (remembering that a degree of polarisation 1:100 is measured in terms of intensity) corresponds to a Jones vector.
\[ E = \left( \exp \left( i \cdot 0.06\pi \right) \right), \quad (45) \]
suggesting that we have determined \( \Delta \phi \) to within 0.06\( \pi \) or 3% of a cycle. We then added \( \frac{\pi}{2} \) to the phase difference between A and B to generate a circularly polarised beam. We rotated an analyser through the beam and measured the ratio of maximum to minimum intensities to be 1.2:1. This is consistent with our previously estimated value for the error on the phase difference being roughly 3% of a cycle.

### 4.6. Control of holograms

The ForthDD R3 is connected to PC via DVI connection and appears as an external monitor. We have written a MATLAB graphical user interface (GUI) to control the holograms. This GUI allows us to set the polarisation state of the beam, adjust aberrations and position the holograms on the SLM. The software is written in the model-view-controller architecture so allows us to control the holograms programmatically as well as graphically. A screenshot is shown in Figure 4.10 with the software configured to display a radially polarised beam. We showed above that the polarisation forming mechanism does not introduce significant aberration, so we expect to be able to correct aberrations from elsewhere in our system by adding the same amplitude of a given Zernike mode to both holograms. To add equal amplitudes of aberration correction to both holograms we must consider the symmetry of the imaging system. Let \( W(r, \theta) \) be a wavefront at the position of hologram A. Let \( W \rightarrow W' \) through the imaging system. Then

\[ W'(r, \theta) = -W(r, \theta + \pi), \quad (46) \]
i.e. \( W \) is inverted about the origin (equivalent to rotation by \( \pi \) about the origin in 2D) and has its magnitude changed (from Newton’s conjugate distance equation). This is shown in Figure 4.11. Now consider the effect of this operation on each Zernike mode. The Zernike modes in the Noll representation [141] are defined as [142]
Even set: \( Z_n^m(r, \theta) = R_n^m(r) \cos m\theta \), and
\[
(47)
\]
Odd set: \( Z_n^{-m}(r, \theta) = R_n^m(r) \sin m\theta \), where \( n \) can be any positive integer or zero, \( |m| \leq n \) and \( n - |m| \) is even. From these definitions, we can easily derive
\[
-Z_n^{\pm m}(r, \theta + \pi) = (-1)^{m+1}Z_n^{\pm m}(r, \theta) ,
\]
so therefore the signs of the polynomials with odd \( m \) do not change sign, but those with even \( m \) do. Since \( n \) and \( m \) are both odd or both even, then the polynomials with odd \( n \) do not change sign, and those with even \( n \) do. For example me must change the sign of tilt and coma, but not defocus, astigmatism or spherical aberration.

Figure 4.10. Screenshot of the MATLAB GUI used to control holograms on the SLM.
4.7. Chapter summary

We have described a robust method for control of the polarisation of a light beam using an FLCSLM. We built up to this description in stages, first describing the operation of an FLCSLM pixel on a general input state. We showed that roughly 69% of the input light is unmodulated for the particular model of SLM used here. We described the normal operation mode of an FLCSLM for achieving binary phase control: use linearly polarised input light polarised along the FLCSLM state bisector. In this mode the modulated and unmodulated light are of orthogonal linear polarisations. We then described how to control diffracted amplitudes with binary holograms by adjusting the fringe shapes. Our polarisation control scheme requires holograms with a constant zero order, no even orders, and whose first order phase doesn’t change as we vary the diffracted amplitude. We present two such binarisation schemes. We showed that we can generate a polarisation controlled beam by using two holograms, A and B, each controlling one linear polarisation component of the output beam. Hologram A is imaged onto B using a reflective 4F system, which means that the SLM itself is effectively splitting and recombining the diffracted orders and so alignment and adjustment are simple. Practically, we showed the generation of beams having circular, radial and azimuthal polarisation profiles. We then described measurements that demonstrate that our 4F imaging arrangement does not introduce significant aberrations, so we can apply the same aberration correction to A and B when correcting the output. We briefly described our GUI that allows adjustment of parameters including the correction of low-order
Zernike aberrations. The fast reconfigurable polarisation control afforded by our method forms a key part of our microscope, described in later chapters of this thesis.
5 Existing methods for the determination of fluorophore orientation

A central goal of this work is to develop a microscope that is capable of 3D orientation determination of fluorescent dipoles. We are not the first to propose using an optical microscope to determine the orientation of fluorescent dipoles in 3D, indeed many techniques have been proposed and demonstrated to work on a wide range of applications. In this chapter we review these techniques and hope to provide a sense of why fluorescent dipole orientation is a useful parameter to measure. We also explain the limitations of the existing techniques and how our new method should fill these limitations.

When determining the orientation of a single fluorescent dipole or ensemble of dipoles one exploits the fact that the excitation of and emission from a fluorescent dipole have characteristic spatial and polarisation dependencies [143]. We show this in our own simulations in chapter 6. In this chapter we refer to techniques operating in two regimes: in the single molecule regime the image of each fluorophore is well separated and in the dense-labelling regime, there are many fluorophores in a resolution element.

5.1. Single molecule techniques

To date, most methods that have been demonstrated to determine the 3D orientation of molecules have been single molecule techniques that attempt to recognise the patterns of the collected fluorescence from molecules. In 1993 Betzig and Chichester proposed and demonstrated a near field scanning technique that allowed qualitative determination of dipole orientation [144]. Five years later Dickson et al. described a technique that looked at the emission patterns of single molecules in the presence of introduced spherical aberration, in both total internal reflection and far field modalities [145–147]. The
aberration was introduced by imaging through 5 μm of water with an oil immersion objective. This technique was applied to measurement of polymer dynamics using Dil molecules in PMMA films [148]. A similar technique, also using pattern recognition of single molecule images was demonstrated by Osborne [149]. Jasny and Sepiol proposed introducing defocus instead of spherical aberration [150,151] and Böhmer and Enderlein demonstrated a wide field technique with a conventional microscope, termed defocused orientation and position imaging (DOPI). This uses spatially coherent Köhler illumination and a slightly defocused sample to image the in-plane orientation of single molecules [152]. We show examples of images taken with DOPI in Figure 5.1 and Figure 5.2 below. Figure 5.1 shows the comparison between an experimentally measured dipole image (on the left in the square) and three computed images (in the circles). The computed images are of dipoles at different inclinations to the z axis: $\frac{\pi}{2}$, $\frac{\pi}{4}$ and 0 from left to right. Figure 5.2 shows some example images of dipoles taken with DOPI. Notice that only in-plane dipoles are visible because the highly inclined dipoles are not excited by the Köhler illumination. Both figures are taken from Böhmer and Enderlein’s original paper [152]. The defocus is approximately 1 μm at an NA of 1.4, which is about twice the axial extent of the PSF.

Figure 5.1. DOPI: measured and theoretical images of single dipoles at a defocus distance of 1.2 μm. The scale bars are 100 μm and the images have been taken at 60× magnification. Taken from [152] with permission. On the left, in the square, is an experimentally measured image. There is significant noise and the centre of the pattern appears messy. On the right, in the circles, are three different calculated dipole images. The dipole angles are at $\theta = \frac{\pi}{2}, \frac{\pi}{4}$ and 0 to $z$ and are all in the $yz$ plane. The in-focus image would have a FWHM of approximately 14 μm ($\lambda = 647$ nm).
This amount of defocus has a number of important consequences. Firstly, the technique relies on being able to separate the image of each dipole, and at a defocus distance of 1 µm, the diameter of the PSF is approximately 10 times larger diameter than for the in-focus case. This means that molecules must be very sparsely spaced. Secondly, the peak intensity of the PSF at 1 µm is roughly 100 times lower than in focus. This means that long exposure times must be used to gather sufficient signal. Typical exposure times are on the order of a few seconds using a high-end cooled CCD camera [153]. Thirdly, the sample must be very thin in z because there must be no new in-focus objects when the imaging system is in its defocused position. In addition, the polar angle is not determined for molecules closer than about 45° to the optic axis because these molecules are not efficiently excited by Köhler illumination. A pattern-matching algorithm for DOPI was later developed by the same researchers to allow quantitative determination of the dipole orientation [154]. DOPI has been applied in a wide range of studies. One main area has been to study polymer dynamics in thin films [155–159] and model membranes [160–163]. For example in [157] the authors measure the glass transition temperature of a thin-film of Polystyrene by observing changes in the orientation of molecules attached to polymer chains.
Other studies have used DOPI to study rotational diffusion in glassy polymers on timescales of a few seconds [164] and on timescales of hundreds of seconds [165]. In addition the technique has been used in vitro in combination with conventional localisation microscopy to study the rotational motion of myosin as it moves along actin filaments [166–168] and to study orientation effects in Förster resonance energy transfer [169]. In [167] the authors were able to show that Myosin molecules walk along actin with a 180° rotation of the molecule for each footprint.

There are also a number of methods that have been described that do not require a defocused sample. One of these allows simultaneous measurement of 3D position and orientation using double-helix localisation microscopy [170]. The double helix PSF gives information about the position of the emitter in three dimensions [126–128], and images are collected with two excitation polarisations and with detection in two orthogonally polarised channels to give in-plane orientation estimates. In addition, the relative intensity of the two lobes encodes the out-of-plane dipole orientation. Another technique uses a spatial light modulator in the emission pathway to direct light from four quadrants of the pupil to slightly different points in the image plane [171]. The relative intensities of these four points then give information about the dipole orientation. A conceptually similar technique analyses the collected fluorescence polarisation patterns in the pupil of the microscope objective using a Bertrand lens [172]. This has the disadvantage that only one molecule can be observed in a field-of-view at any time. Therefore this technique could only be used in a point-scanning regime because signals from widely spaced sample points are all coincident in the pupil plane. Such a point scanning regime was proposed in 2005 [173,174] and demonstrated in 2012 [175]. Another technique attempts to fit in-focus PSFs by using TIRF illumination exploiting the asymmetrical structure of high NA focusing (explained in a later chapter in this thesis) and detection of polarised light [176]. Novotny et al. used pattern recognition of confocal images of fluorescent dipoles to determine the 3D orientation of single molecules. Their initial technique [177] used annular illumination to enhance the longitudinal field components relative to the xy components. This was then
improved to use the full spatial frequency content of the imaging system by use of a radially polarised pupil [83]. Ishitobi et al. implemented a natural addition to this technique by also using an azimuthally polarised pupil to determine the in-plane dipole angle [178]. Finally we mention a recent pattern matching technique which demonstrated the determination of the 3D orientation of nitrogen vacancies in nanodiamonds by scanning a radially polarised excitation beam and performing pattern recognition on the resulting PSFs [179].

There are also a class of methods that can detect single molecule orientation without resorting to pattern matching techniques [180]. In theory these should be more robust to aberrations and noise. First we consider techniques that use polarised detection, which are based on fluorescence anisotropy. In 2001 Fourkas proposed a technique for determining fluorophore orientation in 3D. This is a single point technique for measuring the orientation of single fluorescent molecules without ambiguity [181]. The angle of the dipole in the focal plane is determined from the axis of maximum polarisation, and the angle to the z axis is determined from degree of polarisation. Fourkas shows that measurements from three polarised detectors at 0, 90 and 45 degrees will enable the orientation of the dipole to be fully determined, though as with many techniques effective excitation of z-oriented dipoles was not considered so full 3D orientations may not be determined. This technique does not involve any fitting and so is predicted to be able to determine dipole orientations in ~1 ms. Further computational work was performed by Lu and co-workers [182]. They proposed a 50-50 beamsplitter and linear polariser to extract the 45° signal, and then a polarising beamsplitter to separate the 0° and 90° signals. Lu et al. also formulated a scheme for calculating the dipole orientation from the three detector signals in the presence of shot noise. One final technique in this class was proposed by Foreman et al. in 2008 [183]. This is a confocal scanning technique for measuring orientations of single molecules. A phase mask is used to engineer constructive interference at a confocal detector for the z-oriented dipoles. This is predicted to offer good signal to noise performance for z-oriented dipoles. Again, the authors do not consider the method of excitation of z-oriented dipoles. A practical implementation would require two, or probably
three, excitation polarisations in addition to the polarised detection scheme. In addition, the authors propose determining the in-plane orientation with just two detector signals at 0° and 90°. In fact, this does not completely determine the in-plane orientation because there is ambiguity between ±45° dipoles. A more recent technique used a rotating half-wave plate in the emission path to vary the signal between two detectors measuring orthogonal linear polarisation states [184]. As with the Fourkas technique, they determine the in-plane orientation from the relative strengths of the two polarised detection arms, and the out-of-plane orientation from the degree of polarisation.

Another class of techniques is those that modulate the excitation polarisation and do not rely solely on polarisation sensitive detection. In 1996 Ha et al. demonstrated a technique to accurately determine the in-plane angle of a fluorescent dipole at a single point to 0.2° by continuously rotating their linearly polarised excitation light at about 1 Hz and detecting the resulting fluorescence intensity until photobleaching occurred (typically 4 s) [185]. Polarised detection was later added to the technique to study reorientation of dipoles between excitation and emission [186]. Moerner’s group used four in-plane excitation polarisations to remove ambiguities in molecule orientation [187,188] and measure motions of kinesin along microtubules. Forkey et al. extended this technique to 3D orientation measurement of fluorophores relative to actin filaments [189–191]. Their method used TIRF excitation to generate strong axially polarised fields which affords efficient excitation of the z-oriented dipoles. They achieve an angular precision of 5-10°, limited by shot noise, with a temporal precision of 40 ms [192]. Vacha and Kotani proposed a similar variant [193].

5.2. Dense labelling techniques

Traditionally, the orientations of fluorophores in densely labelled samples are studied using fluorescence anisotropy or FDLD [27]. Imaging in this regime is harder than in the single molecule regime because every resolution volume will contain a collection of molecules. Therefore one has to determine the
distribution of molecule orientations, or at least something about this
distribution. Nevertheless, these two traditional polarisation sensitive
techniques have been applied to a wide range of biological systems [194,195].

In standard fluorescence anisotropy measurements, a single linear polarisation
state is used to excite fluorophores. Light is then detected in two orthogonal
polarisation channels [36]. Those dipoles aligned parallel to the incident
polarisation vector will be preferentially excited and will emit light with a
polarisation depending on their orientation. This technique is typically used to
detect rotation between excitation and emission because if the sample contains
fluorophores with a random orientational distribution and molecules can
significantly rotate (for example by diffusion) before they radiate then there will
be no difference in signal between the two detection channels. Alternatively, if
the molecules cannot rotate between excitation and emission then there will be a
greater signal in the channel parallel to the excitation polarisation. Anisotropy
measurements could also be used to infer molecular direction in a sample with
highly aligned fluorophores, though the degree of alignment would need to be
known in advance. Detection of polarised fluorescence has been used to study
alignments in the cytoskeleton, for example examining the relative orientations
of actin and myosin [196,197], and to study how the cytoskeleton becomes
aligned during muscle contraction [198]. In 1979 Axelrod used linearly
polarised excitation in a wide field epifluorescence microscope and a rotating
analyser in front of the detector to measure the orientation of DiIC molecules in
Erythrocyte ghosts [31] (these are red blood cells that have been killed and have
had their contents extracted). He was able to report the molecular orientation in
2D in the cell membrane and confirm that the DiIC molecules preferentially
orient themselves normal to the cell membrane surface. In a careful theoretical
analysis he considered how many phenomena, including high-NA collection, non-
zero angle between excitation and emission dipoles, rotational diffusion, and a
range of molecular orientations within a focal volume, affected the
measurements. The depolarising effects of high-NA focusing were not
considered. Kinosita showed another wide field technique that was used to
measure the orientation of fluorescent dyes rigidly attached to actin
filaments [199,200]. This technique used an arrangement of beamsplitters in the emission path to simultaneously generate a pair of images.

FDLD is a similar technique using polarised excitation instead of polarised detection [201]. FDLD has been combined with fluorescence anisotropy (i.e. using polarised excitation and polarised detection) to measure the orientation of actin fibres and dyes in lipid membranes [26,202]. DeMay et al. [203] describe a wide field trans illumination modification to FDLD using four linear excitation polarisations. From images taken with incident polarisations at 0°, 45°, 90° and 135° they plot the intensity \( I \) at each pixel and infer the phase, amplitude and offset of a fitted sinusoid (note the similarity to single molecule techniques [187,188]). The polarisation states are generated using non-pixelated liquid crystal elements. No attempt is made to extract information from the amplitude and offset of the sinusoid, but they report the ratio \( I_{\text{min}}/I_{\text{max}} \), which is a function of the level of alignment of fluorophores in a focal volume but also the out-of-plane angle. This technique has been used to determine the organisation of septins in yeast cells [204].

These techniques are only sensitive to orientations in the \( xy \) plane, and suffer from ambiguities between orientational order and out-of-plane dipole angle. To circumvent this, Kress et al. [27,205] image only in the equatorial plane of cells and so can be reasonably confident that low variation in signal is due to a wide distribution of fluorophore angles (rather than dipoles oriented along \( z \)). They also acquire 90 images (instead of the four as described above) each with a different excitation linear polarisation state (angle \( \alpha \) to the \( x \) axis). The polarisation azimuth angle is rotated by a motorised half wave plate. They extract the average molecular direction and the average variation in molecular direction using a Fourier series method. This provides measurement of the amplitude, phase and offset of the intensity as a function of \( \alpha \). This is equivalent to extracting the parameters from a fitted sinusoid, as described above [203]. This technique has been applied to an investigation of the role of septins in cytokinesis [206].
Orientational effects are more pronounced in nonlinear and multi-photon microscopy than in single-photon microscopy because of increased photoselectivity of excitation. For example, the excitation probability in two-photon imaging will depend on \((d \cdot E)^4\). Brasselet conducted an extensive review into the use of polarisation microscopy in nonlinear imaging [207] and her group has made many advances in the field. For example, they have used two-photon microscopy with polarisation sensitive excitation to study orientation in artificial membranes [208]. A number of techniques for detecting two-dimensional [209,210] and three-dimensional [84,211] molecular orientation in second harmonic generation microscopy have also been proposed. The three dimensional technique used a spatial light modulator to generate three orthogonal excitation polarisations \((x, y, \text{ and } z)\) and was demonstrated in collagen fibres.

5.3. Chapter summary

We have given a brief overview of a large range of techniques, in both the single molecule and dense labelling regimes, for determining the orientations of fluorophores. In the single molecule regime, a large class of methods depend on pattern recognition of the image of fluorophores in the presence of aberrations. The most widely adopted technique is defocused orientation and position imaging, DOPI. DOPI has been widely used to study biological and physical problems despite many drawbacks. These are that very sparse samples and very sensitive cameras are needed to image the defocused fluorophore images, and that the sample must be very thin, so that there is no new light being emitted from the focal plane. DOPI also has the drawback that the emission patterns must be fitted with computer modelled PSFs, and so accurate results are limited to aberration-free samples. DOPI is not really a 3D technique because the sample is illuminated by Köhler illumination so the z-dipoles are not excited. Other techniques analyse the polarisation of the emitted light [181] and the behaviour under different excitation polarisations [190] by only looking at relative intensities not emission patterns and so would be expected to be more robust. In the densely labelled sample regime, many techniques are based on
fluorescence anisotropy or FDLD. Both of these techniques have been used extensively to study orientational effects in lipid and cell membranes and to study the alignment and rearrangement of the cytoskeleton. In general, there is a problem with the measurement of 3D fluorophore orientation in densely labelled samples: both a collection of fluorophores well aligned along the optic axis and a set of fluorophores with no alignment will display isotropic polarisation response.

A problem with many of these methods for 3D orientation determination is that dipoles aligned along, or close to, the optic axis are not efficiently excited. One solution is to use TIRF illumination, which restricts studies to be at the cover slip. The only other way to effectively excite these dipoles is to use radially polarised illumination [83], though the radially polarised beam does not facilitate the determination of the in-plane angle. There has been work combining radial and other excitation polarisations [178] but the switching of excitation polarisation was manual and slow. Therefore, we see a gap in the current range of techniques. Our microscope has switchable and reconfigurable excitation polarisation so that we can quickly change between radial and in-plane excitation modes. It presents a major advantage over DOPI in terms of signal strength because molecules are observed in focus. We hope that this chapter sets the design aims for our microscope in context.
6 Theory and modelling of a confocal microscope

In this chapter we introduce a full theoretical description of a confocal microscope. The chapter consists of some explanation of standard results from the literature and some new work. The model of the confocal microscope enables us to predict the behaviour of single fluorescent emitters under polarised excitation.

Starting from Maxwell’s equations, we show how a concentrated spot is formed at the focal plane of a lens. We include effects arising from the vector nature of the electromagnetic field, which Richards and Wolf showed are important at high NA [212]. In the Richards-Wolf theory, the expressions for the focal fields are written in polar integral form. We show how these focal fields can be computed efficiently by rewriting the integrals as Fourier transforms, and how these Fourier transforms can be computed with a chirp-z transform algorithm to allow arbitrary levels of zooming on the Fourier plane. We focus on explaining physical concepts and assumptions rather than providing a fully rigorous derivation. We then consider how light emitted from a Hertzian dipole is collected by a high NA microscope objective (in an epifluorescence microscope this is the same objective as was used for illumination) and focused onto a confocal pinhole. Taken with the focusing theory, this leads to a full model of a confocal microscope. This model allows us to investigate the interaction of single emitters at different orientations with different excitation polarisations and we discuss the possibility of using different excitation polarisations to resolve emitter orientation. The model also allows us to build a numerical model of optical sectioning in confocal microscopes that, unlike previous work, accounts for the dipole-like nature of emitters and the polarisation of excitation beam.
The work on the fast calculation of the focal field patterns of arbitrarily polarised pupils using FFTs is not new and was first undertaken by Boruah [213]. We are not the first to suggest speeding up the calculation of focal fields using the chirp-z transform [214]. The calculation of the collection of polarised light from dipoles is also not new [215], but to our knowledge we are the first to pair this with a fully vectorial model of excitation to produce simulations of a vectorial confocal microscope.

6.1. From Maxwell’s equations to a bright focal spot

The goal here is to describe the theoretical framework that allows the calculation of the focal fields of high numerical aperture microscope objectives. The theory was largely developed by Richards and Wolf before 1960 [212,216]. The Richards and Wolf theory, also known as vectorial Debye theory [179,213,217,218], is a generalisation of the Debye diffraction theory to high numerical aperture and accounts for the polarisation of light. The Debye theory is essentially just the Fresnel-Kirchoff diffraction integral for a focusing system [12,219].

With recent advances in computing power it is possible to quickly calculate focal fields by numerical integration or using fast Fourier transform (FFT) algorithms. Perhaps one day it will be efficient to simply describe a pupil field and aperture and solve the whole diffraction problem using Maxwell’s equations in their full form using a finite element package. For now, the integral method seems most appropriate.

6.1.1. Scalar diffraction theory

Suppose that we are seeking to find the form of the electromagnetic field in some region given its known distribution on another surface, which is often a plane. This is how Kirchhoff formulated the diffraction problem [11]. Our specific problem is to find the field near the focus of a lens given the field in the pupil
plane. The field at the pupil plane is assumed to have a finite, non-zero extent due to the microscope objective pupil.

Any description of electromagnetic phenomena must start with Maxwell’s equations. In the absence of free charge and current, they are [220]

\[
\begin{align*}
\nabla \cdot \epsilon \mathbf{E} &= 0 \\
\nabla \cdot \mu \mathbf{H} &= 0 \\
\n\nabla \times \mathbf{E} &= -\mu \frac{\partial \mathbf{H}}{\partial t} \\
\n\nabla \times \mathbf{H} &= \epsilon \frac{\partial \mathbf{E}}{\partial t}.
\end{align*}
\]

(49)

Here \( \mathbf{E} \) and \( \mathbf{H} \) are the electric and magnetic fields and \( \epsilon \) and \( \mu \) are the permittivity and permeability. Usually the equations are referred to as “Maxwell 1” to “Maxwell 4” in the order written. Taking the curl of Maxwell 3, and substituting in Maxwell 4 gives a wave equation for \( \mathbf{E} \):

\[
\nabla^2 \mathbf{E} = \frac{n^2}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}.
\]

(50)

In reducing \( \nabla \times \nabla \times \mathbf{E} \) to \( -\nabla^2 \mathbf{E} \) we have neglected a term involving \( \nabla \cdot \mathbf{E} \) that is only zero in a linear, isotropic and homogenous medium. Here we have introduced \( n \), the refractive index and \( c \), the speed of light in vacuum. A similar procedure can be performed to give a wave equation for \( \mathbf{H} \). Indeed, in a nonmagnetic material, \( \mathbf{H} \) obeys the same wave equation as \( \mathbf{E} \). Since each component of \( \mathbf{E} \) and \( \mathbf{H} \) must obey this wave equation, we can expect every component to obey a scalar wave equation,

\[
\nabla^2 u = \frac{n^2}{c^2} \frac{\partial^2 u}{\partial t^2},
\]

(51)

where \( u \) is a scalar field representing any component of \( \mathbf{E} \) or \( \mathbf{H} \). This is the essence of the scalar theory of light. From this point forward, even when we consider the polarisation of light in the Richards and Wolf theory, we are really only dealing with a scalar theory of light because we are neglecting the coupling of \( \mathbf{E} \) and \( \mathbf{H} \) at boundaries (for example at the edge of the pupil). It has been well demonstrated in experiment that the coupling of fields is not important if the diffraction aperture is large compared with the irradiating wavelength and the observation point is not close to the diffracting aperture [221]. Rudolf et al. state
that the scalar theory is appropriate when the smallest features in the aperture are greater than three to four times the incident wavelength \[222\]. Otherwise, a rigorous vector theory is needed. There is an entire chapter in Born & Wolf\[12\] devoted to rigorous diffraction theory, where the coupling of fields is not neglected. For now we proceed with the scalar version.

### 6.1.2. The Kirchoff diffraction theory

In 1678 Huygens proposed the idea that in a wave, each point on a wavefront could be considered to give rise to propagating secondary disturbances. Thomas Young then proposed the idea of interference in 1804, and in 1818 Fresnel combined Young’s and Huygens’ ideas to give a first theory of diffraction. In this section, we describe the integral theorem of Helmholtz and Kirchoff. This was developed in 1882 to put the ideas of Huygens, Young and Fresnel on sound mathematical footing.

For a monochromatic wave, we can write the field component \( u \) as separable in time and space, so that

\[
u(P, t) = U(P) \exp i \omega t .\] (52)

Where \( U(P) \) contains all of the spatial information about the scalar field at point \( P \) and \( \omega \) is the angular frequency of the wave oscillation. In this case, the scalar wave equation, equation (51) reduces to the Helmholtz equation

\[
(\nabla^2 + k^2)U(P) = 0 .\] (53)

Here we have introduced the wavenumber \( k \), given by \( k = \frac{\omega n}{c} = \frac{2\pi}{\lambda} \), with \( \lambda \) being the wavelength.

Kirchoff proceeded from here by using a Green’s function approach to the problem. The development of the Kirchoff diffraction theory is well explained by Goodman [11], and will not be fully reproduced here. In essence Kirchoff used the free space Green’s function, an expanding spherical wave \( G = \frac{\exp(ikr)}{r} \), and used Green’s second theorem to write an integral solution for \( U(P) \) given known
values of $U$ on some arbitrary surface $S$ that encloses $P$. The result, the integral theorem of Helmholtz and Kirchoff, is

$$U(P) = \frac{1}{4\pi} \int_S \left( G \frac{\partial U}{\partial n} - U \frac{\partial G}{\partial n} \right) dS,$$  \hspace{1cm} (54)

where $G$ is the Green’s function mentioned above, and $\partial/\partial n$ is a partial derivative along the outward normal to $S$. This form of the Helmholtz-Kirchoff diffraction integral gives a solution in a very general form, but if we further restrict ourselves to the case of an aperture in a plane screen, which takes us nicely towards our end goal of describing focusing through a lens, we can make further advances. This will lead us to the more familiar form of Kirchoff’s diffraction integral.

The geometry used by Kirchoff is shown in Figure 2.1. The surface $S$ is broken into a planar region $S_1$ and a spherical cap $S_2$. We consider an incident field on a diffracting aperture, $\Sigma$ in a screen. As $R$, the radius of the spherical cap is increased, the surface integral over $S_2$ vanishes if and only if the fields are decaying as fast, or faster than, a diverging spherical wave. This is known as the Sommerfeld radiation condition. Since we are dealing only with incoming
radiation through the planar screen, this condition will be satisfied. There is now one further step, a little associated controversy, and a partial resolution: Kirchoff assumed that outside $\Sigma$, $U$ and $\partial U/\partial n$ are both zero. This assumption causes some debate because it implies that $U$ is zero inside the diffracting aperture [223, 224]. An interesting philosophical question arises: how can an inconsistent theory give rise to excellent predictions [225]?

These inconsistencies are resolved by the Rayleigh-Sommerfeld theory. Happily, under further assumptions (detailed below) the Rayleigh Sommerfeld and Kirchoff diffraction integrals are the same. Following Kirchoff again, assuming that $U$ and $\partial U/\partial n$ are both zero outside $\Sigma$, the Helmholtz-Kirchoff integral reduces to

$$U(P) = \frac{1}{4\pi} \int_\Sigma \left( G \frac{\partial U}{\partial n} - U \frac{\partial G}{\partial n} \right) dS,$$

where the integral is now only taken over the finite planar region $\Sigma$. If we assume that we are investigating the field at some point that is much further from the screen than $\lambda$, we can make further progress. Now by definition, $\frac{\partial G}{\partial n} = \nabla G \cdot n$ and also clearly $\nabla G = r \frac{\partial G}{\partial r} = r \left( ik \frac{1}{r} \right) G$. Under the assumption that $r \gg \lambda$, we can neglect the second term here, and we therefore have $\frac{\partial G}{\partial n} = ikG \cos \rho$, with $\rho$ the angle between $r$ and $n$. Now consider spatially coherent illumination $U = A \exp(ikt)$ from a point source at $T$ on the left of the screen, define $t$ as the vector from $T$ to $P_1$ ($P_1$ is a point in the diffracting aperture that will be integrated over) and let then $t = |t| \gg \lambda$, giving $\frac{\partial U}{\partial n} = ikU \cos \theta$, with $\theta$ the angle between $t$ and $n$. Equation (55) becomes

$$U_0(P) = \frac{ikA}{4\pi} \int_\Sigma \frac{\exp ik(r+t)}{rt} (\cos \theta - \cos \rho) dS,$$

or

$$U_0(P) = \frac{A}{i\lambda} \int_\Sigma \frac{\exp ik(r+t)}{rt} (\cos \theta - \cos \rho) dS,$$

which is the Fresnel-Kirchoff diffraction formula. It is worth mentioning at this point that Fraunhofer and Fresnel diffraction formulae derive from the Fresnel-Kirchoff diffraction formula with further simplifications. Essentially one neglects the angle-dependent terms, known as the obliquity factors, and Taylor expands the exponents. A more rigorous theory was derived by Rayleigh and
Sommerfeld. They give two solutions RS1 and RS2, and the Kirchoff solution is the arithmetic mean of RS1 and RS2.

6.1.3. The Debye integral

Now we proceed to evaluating the focal fields of a lens. For this, we consider a converging spherical wave in an aperture, and use the Fresnel-Kirchoff theory described above. The geometry and variables are set out in Figure 6.2. A spherical wavefront converges on O at a distance \( f \) from C. If the numerical aperture is low, and we are observing the field reasonably close to \( O \) then we can ignore the obliquity factors. The negative sign on \( f \) indicates an incoming spherical wave. Equation (56) becomes

\[
U_0(P) = \frac{A}{i\lambda f} \int_{\Omega} \frac{\exp{ik(r - f)}}{r} \, dS .
\]

At low numerical aperture, the wavefront size is small compared to the focal length. We also only investigate the field at positions that are close to the geometric focal point relative to focal length. Under these assumptions we can rewrite \( r - f \) as follows. Note that \( PQ = -R + f \mathbf{q} \) and then

\[
|PQ| = r = \sqrt{R^2 + f^2 - 2f q \cdot R} = f \left( 1 - \frac{q \cdot R}{f} + O((R/f)^2) \right) ,
\]

So, keeping only the first order term and rearranging slightly, we have

\[
r - f = -q \cdot R .
\]

Now replacing \( dS = f^2 d\Omega \) and using \( rf \approx f^2 \), we have

\[
U_0(P) = \frac{A}{i\lambda} \int_{\Omega} \exp(-i k q \cdot R) \, d\Omega ,
\]

which is the Debye diffraction integral. We have converted the Fresnel-Kirchoff diffraction integral into a sum over plane waves. At this point it is worth making a brief comment on the validity of the Debye integral. Wolf showed that the Debye integral is only valid in systems where the Fresnel number is much greater than unity [226,227]. The Fresnel number is defined as

\[
Fr = \frac{a^2}{\lambda f} .
\]
In finding a scalar solution to Maxwell’s equations, we have already assumed that the aperture is large compared to the wavelength so requiring \( Fr \gg 1 \) is essentially saying that the numerical aperture of the system cannot be too small. As a concrete example, imagine green visible light having a wavelength of 0.5 \( \mu \text{m} \) incident on a lens a pupil radius of 5 mm. Then if we require \( Fr > 100 \), say, then the focal length of the lens must be less than 500 mm. It is interesting to note that we are placing an upper bound on the focal length. This amounts to saying that there must be some appreciable convergence for the Debye theory to hold. For example, the Debye theory cannot be used to predict the diffraction of plane wavefronts from an aperture. As we will see below, and as pointed out by Wolf during his discussion on the validity of the Debye theory [227], there is also a lower limit on the focal length (an upper limit on the NA) because at high NA polarisation effects distort the symmetry of the focal spot [212].

![Figure 6.2](image.png)

Figure 6.2. Coordinates and notation for the Debye integral. \( W \), shown in blue, is a spherical wavefront converging on the image point \( O \). It is restricted by the aperture \( \Sigma \) shown in grey. \( q \) is a unit vector along \( OQ \), \( f \) is the focal length of the lens.

### 6.1.4. High NA vector theories: the Richards and Wolf theory of focusing

Now we turn to modification of the Debye diffraction integral to account for polarisation effects in high NA lenses. This is known as the Richards and Wolf or vectorial Debye theory [212,216]. Ignatowsky also developed this theory in 1919, though his papers are only available in Russian [228,229]. The Richards
and Wolf theory was originally developed for a linearly polarised beam and has been extended for arbitrary phases and polarisations in the objective pupil [213,227]. The reader should be reminded that even though we are partially accounting for the vector nature of light, it is not a full vector theory because we are still dealing with Kirchoff’s solution to the scalar Helmholtz equation and in doing so are neglecting the coupling of fields at the boundary edge. We begin by stating the Richards and Wolf integrals. They are:

\[
E(x, y, z) = \frac{1}{i\lambda} \int_{\Omega} \frac{a(s_x, s_y)}{s_z} \exp(ik s \cdot r) \, ds_x ds_y ,
\]

and

\[
H(x, y, z) = \frac{1}{i\lambda} \int_{\Omega} \frac{b(s_x, s_y)}{s_z} \exp(ik s \cdot r) \, ds_x ds_y .
\]

These equations are simply a vector form of the Debye diffraction integral (60). Note that we now use the Richards-Wolf notation and replace \( R \rightarrow r \). The vector strength factors \( a \) and \( b \) replace the scalar field amplitude \( A \), \( s \) is a unit vector along the ray in the direction of propagation (opposite to \( q \)), and \( \frac{ds_x ds_y}{s_z} \) is a unit of solid angle. The form of this is explained in Figure 6.3.

![Figure 6.3](image.png)

Figure 6.3. Illustration of the form of the angular part of the Richards-Wolf diffraction integral. The diagram shows the ray arrangement when \( s_x = 0 \). In this case, \( ds_x = \sin \theta \, d\phi \). We can see that \( d\theta = ds_y /\cos \theta \), and by definition \( \cos \theta = s_x \). Therefore \( d\Omega = ds_x ds_y /s_z \).

To evaluate the vector strength factors, we need to consider how a ray \( s \) bends as it propagates through the system. We define the focal sphere as the sphere of radius \( f \) centered on the Gaussian image point \( O \), and then for a system obeying
the Abbé sine condition the incident ray height and ray angle in image space are related by $h = f \sin \theta$ [12]. In the following section, variables subscripted 0 refer to quantities in the pupil and those subscripted 1 refer to quantities on the focal sphere. The coordinate system is depicted in Figure 6.4. Here we consider a linearly polarised pupil field with uniform amplitude and constant phase, $e_0 = l_0 \hat{e}_0$. When this ray is traced through to the focal sphere, its field is $e_1 = l_1 \hat{e}_1$. Imagine a small region $dA_0$ surrounding the ray: the energy before focusing is given by $l_0^2 dA_0$, and the corresponding energy after focusing $l_1^2 dA_1$. By conservation of energy, $dA_0 = dA_1 \cos \theta$ and therefore $l_1 = l_0 \sqrt{\cos \theta}$. Therefore $a = f l_0 \sqrt{\cos \theta} \hat{e}_1$. Now we know the strength of the focused field we must work out the polarisation direction.

Figure 6.4 shows the propagation of a wavefront through an optical system. The electric field in the pupil plane (green disc) is polarised in the direction $e_0$ which defines the coordinate axes in object space.

Richards and Wolf assume that the angle of incidence at each interface in the physical lens is small, and then state that in this case it can be assumed that the angle which the polarisation vector makes with the meridional plane (that containing the optic axis and the ray) is constant through each surface, and therefore equal before and after focusing. This assumption neglects the unequal
transmission amplitudes of s and p polarised components, which is reasonable if the angle of incidence is small at each surface. Therefore

\[ \hat{e}_1 = (g_0 \cdot i)g_1 + (g_0 \cdot j)(g_1 \wedge s). \] (64)

Here \( g_0 \) is the radial vector in the pupil plane, \( i \) and \( j \) are unit vectors along and perpendicular to \( e_0 \) (as a reminder, this is the polarisation direction), both in the pupil plane, \( s \) is the unit vector along the ray in the focal space and \( g_1 \) is in the meridional plane and perpendicular to \( s \). We can now write the vector amplitude as

\[ a = fl_0 \sqrt{\cos \theta \left( (g_0 \cdot i)g_1 + (g_0 \cdot j)(g_1 \wedge s) \right)} . \] (65)

Now all that remains before evaluating the diffraction integrals is to change to a sensible spherically symmetric coordinate system. Richards and Wolf pick a system with polar axis in the +\( z \) direction and azimuth \( \phi = 0 \) containing the polarisation vector in object space. This makes sense for a linearly polarised beam. The pupil coordinates \((x, y, z)\) are also written in this same coordinate system, but the angles are labelled with the subscript \( P \) to distinguish these quantities from ray angles. This gives

\[
\begin{align*}
s_x &= \sin \theta \cos \phi, \quad s_y = \sin \theta \sin \phi, \quad s_z = \cos \phi, \\
x &= r \sin \theta \rho \cos \phi_p, \quad y = r \sin \theta \rho \sin \phi_p, \quad z = r \cos \phi_p, \\
and \quad s \cdot r &= r (\cos \theta \cos \theta_p + \sin \theta \theta_p \cos (\phi - \phi_p)) .
\end{align*}
\] (66)

Now we need to work out \( g_0 \) and \( g_1 \) in this coordinate system too. We have defined \( s_\perp = (s_x, s_y) \) which is the component of \( s \) in the \( xy \) plane. \( s_\perp \) is antiparallel to \( g_0 \) by definition. Therefore, since \( g_0 \) is a unit vector, \( g_0 = - (\cos \phi, \sin \phi) \). Now, since the ray doesn’t move out of the meridional plane, the projection of \( g_1 \) into the \( x - y \) plane is parallel to \( g_0 \), and since \( g_1 \) is a unit vector, its \( z \) component must be equal to \( \sin \theta \). Therefore \( g_1 = (-\cos \theta \cos \phi, -\cos \theta \sin \phi, \sin \theta) \) and so

\[
\begin{align*}
a_x &= fl_0 \sqrt{\cos \theta \left( \cos \theta + \sin^2 \phi (1 - \cos \theta) \right)} , \\
a_y &= -fl_0 \sqrt{\cos \theta \left( \sin \phi \cos \phi (1 - \cos \theta) \right)} , \\
a_z &= -fl_0 \sqrt{\cos \theta \sin \theta \cos \phi} .
\end{align*}
\] (67)

The final diffraction integrals can now be written down. They are
\[ E_x = -\frac{iff_0}{\lambda} \int_0^\alpha \int_0^{2\pi} \sqrt{\cos \theta} (\cos \theta \\
+ \sin^2 \phi (1 - \cos \theta)) e^{ikx - r \sin \theta \, d\theta \, d\phi} , \]
\[ E_y = \frac{iff_0}{\lambda} \int_0^\alpha \int_0^{2\pi} \sqrt{\cos \theta} (\sin \phi \cos \phi (1 - \cos \theta)) e^{iky - r \sin \theta \, d\theta \, d\phi} , \]
\[ E_z = \frac{iff_0}{\lambda} \int_0^\alpha \int_0^{2\pi} \sqrt{\cos \theta} (\sin \theta \cos \phi) e^{ikz - r \sin \theta \, d\theta \, d\phi} , \]

The formulae for the H-fields can be derived from \( b = s \times a \).

The solution of these equations at high numerical aperture leads to some effects now known as depolarisation [230,231]. As an example, the electric field energy densities in the focal plane of an objective of NA = 0.92 with an x-polarised pupil are shown in Figure 6.5. Roughly 77% of the energy is in the x-polarised component, 1% in the y-polarised component and 22% in the z component. In general, the field near focus is elliptically polarised with azimuth in any direction [212]. These results were confirmed experimentally in 1998 [217].
6.1.5. Generalisation of the Richards Wolf integrals to an arbitrary pupil

In their work, Richards and Wolf take the strength factor of their incident field to be uniform across the objective pupil. In order to generalise to a linearly polarised field with some spatial dependence (for example a Gaussian intensity profile or a wavefront with some aberration) Wolf points out that these strength factors can easily be modified to be complex functions of $\theta$ and $\phi$ in the pupil [227].

When considering arbitrary polarisations, we note that the Helmholtz equation is linear, so we can just decompose the 2D pupil field into its $x$- and $y$- components, work out the results of the focal fields for each (including variations in amplitude and phase), and then add the focal fields together.

To write down expressions for the full focal fields, we introduce a new notation. We put $EA_b$ where $A$ and $b$ are Cartesian components, and $A$ refers to the pupil polarisation and $B$ to the focal field component. We now rewrite the Richards and Wolf focal field expressions (68) as
\[ EX_x = -\frac{i}{\lambda} \int_{0}^{\alpha} \int_{0}^{2\pi} l_x(\theta, \phi)e^{iksr} \sin \theta \, d\theta \, d\phi, \]
\[ EX_y = -\frac{i}{\lambda} \int_{0}^{\alpha} \int_{0}^{2\pi} l_y(\theta, \phi)e^{iksr} \sin \theta \, d\theta \, d\phi, \quad (69) \]
\[ EX_z = -\frac{i}{\lambda} \int_{0}^{\alpha} \int_{0}^{2\pi} l_z(\theta, \phi)e^{iksr} \sin \theta \, d\theta \, d\phi. \]

Here we have also introduced the complex amplitude dependence of the pupil function, so that
\[ l_x = W_x(\theta, \phi)a_x, \quad (70) \]

and similarly for the other components of \( l \). We have introduced \( W_x \) which is the pupil function (i.e. the amplitude and phase of the field in the pupil). The subscript is a reminder that at this stage we are dealing with a linearly polarised pupil (polarised along \( x \)). Now we turn to the focal field generated by the \( y \) polarisation components in the pupil. There is clearly no new physics involved in this calculation: if we rotate the \( y \) component of the pupil function by \( -\frac{\pi}{2} \) about \( z \), including the polarisation vector, then calculate the fields, then undo the rotation we must end up with the correct result. This calculation can be found in detail in other works [71,213].

6.2. Fast calculation of focal fields using Fourier transforms

6.2.1. From polar integrals to Fourier transforms

The Richards-Wolf integrals, equations (68), are 2D integrals in spherical polar coordinates. Modern computers mean that it is computationally cheap to perform these integrals numerically, so many analytical works involving series solutions in terms of Lommel functions [12,232] or a multipole expansion [230,233,234] or evaluation of the Richards-Wolf integrals in the presence of specific aberrations [235–237] are no longer so relevant. In 1963, McCutchen [16,17] introduced the idea of converting the Debye integral (remember this is a low-NA theory) from a 2D integration over polar coordinates to a Fourier transform of a discrete matrix (known as a discrete Fourier
transform, DFT). There exist very fast algorithms for computing Fourier transforms of discrete matrices (fast Fourier transform, FFT). It has been reported that computation with FFTs speeds up the integration of the Richards-Wolf equations by a factor of 100 [214]. For these reasons it is advantageous to recast these equations in terms of Fourier transforms. This also allows us to compute the focal fields of arbitrary pupil functions, as the pupil function is just specified as a complex array. McCutchen’s theory was later updated to express the full Richards-Wolf integrals in terms of Fourier transforms [213,226]. Again, here we follow a similar style as before and do not provide a full derivation of the results, but an outline of the arguments. First, rewrite equations (69) as

$$EX_b = -\frac{i}{\lambda k^2} \int_{\Omega} \left. \left. \int \right|_{k'=0}^{\infty} l_b \exp(ik' \cdot r) \delta(k' - k) k'^2 dk' d\Omega \right.,$$

(71)

Where $\delta(x)$ is the Dirac delta function and the subscript $b$ indicates any Cartesian field component. Clearly (71) is equivalent to (69) by the sifting property of $\delta(x)$. Now we notice that $k'^2 dk' d\Omega$ is a $k$-space 3D volume element, so

$$EX_b = -\frac{i}{\lambda k^2} \int_{k_x,k_y,k_z} l_b \exp(ik' \cdot r) \delta(k' - k) dk_x dk_y dk_z .$$

(72)

Here the integral is taken over the half space given by $k_z \geq 0$ which excludes backwards-propagating waves. We can use the sifting property of the delta function to perform the integration over $k_z$. In physical terms this is equivalent to restricting ourselves to considering purely monochromatic sources. When $k' - k = 0, k_z = \sqrt{k'^2 - k^2}$, where $k^2 = k^2_x + k^2_y$. Picking only the positive square root ensures only forward-propagating waves. In addition, we must account for the projection of the spherical cap onto the $(k_x, k_y)$ plane giving a factor of $k/k_z$, so

$$EX_b =$$

$$-\frac{i}{\lambda k^2} \int_{k_x,k_y} l_b \left(k_x,k_y,\sqrt{k'^2 - k^2} \right) \frac{k}{\sqrt{k'^2 - k^2}} e^{ik' \cdot r} dk_x dk_y .$$

(73)
Now, writing the strength factors $l_b$ in terms of the Cartesian ray directions, using $\cos \theta = \frac{k_x}{k_0}$, $\sin \theta = \frac{k_x}{k_0}$, $\cos \phi = \frac{k_y}{k_r}$ and $\sin \phi = \frac{k_y}{k_r}$ and replacing $ks \rightarrow k$ we get

\[
\begin{align*}
    l'_x &= f e^{ik_z z} \sqrt{\frac{k_0}{k_z} \frac{k_0 k_y^2 + k_x k_z^2}{k_0 k_r^2}} W_x(k_x, k_y), \\
    l'_y &= -f e^{ik_z z} \sqrt{\frac{k_0}{k_z} \frac{(k_0 - k_z)k_x k_y^2}{k_0 k_r^2}} W_x(k_x, k_y), \quad (74) \\
    l'_z &= -f e^{ik_z z} \sqrt{\frac{k_0}{k_z} \frac{k_x}{k_0}} W_x(k_x, k_y).
\end{align*}
\]

Here $k \cdot r$ has been expanded, so that we can now write equation (73) as

\[
EX_b = -\frac{i \lambda}{(2\pi)^2} \iint_{k_x, k_y} l'_b e^{i(k_x x + k_y y)} \, dk_x dk_y. \quad (75)
\]

Now we have almost arrived at our desired result: we have an expression for the focal fields of an arbitrary complex pupil valid at high NA in terms of a 2D inverse Fourier transform (up to multiplicative constants). There is just one more step to get the focal fields of a full complex vector pupil, and that is to find the fields due to the y-polarised components of the pupil, $EY_b$. We described at the end of section 6.1.5 how to get these field expressions. Boruah shows [71,213] that the final expressions are

\[
E_b = -\frac{i \lambda}{(2\pi)^2} \iint_{k_x, k_y} L'_b e^{i(k_x x + k_y y)} \, dk_x dk_y, \quad (76)
\]

where

\[
\begin{align*}
    L'_x &= f e^{ik_z z} \sqrt{\frac{k_0}{k_z} \frac{1}{(k_z^2)}} \left( k_y^2 + k_x^2 \frac{k_z^2}{k_0^2} \right) W_x + k_x k_y \left( \frac{k_z}{k_0} - 1 \right) W_y, \\
    L'_y &= -f e^{ik_z z} \sqrt{\frac{k_0}{k_z} \frac{1}{(k_z^2)}} \left( k_y^2 + k_x^2 \frac{k_z^2}{k_0^2} \right) W_y + k_x k_y \left( \frac{k_z}{k_0} - 1 \right) W_x, \quad (77)
\end{align*}
\]
Here the vector pupil function $W$ has been decomposed into its components $W_x$ and $W_y$. Note the symmetry here: the $y$ field produced by the $y$ component of the pupil is the same as the $x$ field produced by the $x$ component of the pupil.

There are many phenomenological papers studying the solutions of these integrals in special cases, for example the analytical solution of focusing radially and azimuthally polarised beams [82] and focusing of spirally polarised beams (which can be represented by the sum of a radial and an azimuthally polarised beam) [238]. Other authors have looked at the effects of aberrations on vortex beams [213,239] and radially polarised beams [240], and many more special cases, not of immediate interest to microscopy [241–243]. As an example of the flexibility offered by the DFT approach described above, Sheppard et al. considered the effect of Gaussian intensity distributions in the pupil [244] and showed that the axial polarisation component is weaker than in the case of a flat-top beam profile. This is to be expected as the Gaussian beam has proportionally less energy at towards the edges of the pupil where ray angles, and therefore polarisation rotation effects, are most prominent.

6.2.2. Zooming in on the focal plane: the chirp-z transform.

We now briefly turn to our implementation of the calculation of equation (76). The standard way to compute a DFT is using the fast Fourier transform (FFT) algorithm [245]. This is not ideally suited to the problem of calculating focal fields because a huge area of the focal plane needs to be calculated and then disregarded. We illustrate this as follows. Say that the pupil function $W$ is specified as two complex square arrays, $W_x$ and $W_y$. Let the number of elements along each dimension be $N$. The radius of the pupil is $r_p = f \cdot NA$ where $f$ is the focal length of the focusing lens, so the pupil is sampled in steps of size $P_s = \frac{2fNA}{N}$. The discrete Fourier transform algorithm returns positive and negative frequency components up to half of the sampling frequency, so the total width of
the frequency plane in either dimension is \( 2F_{\text{max}} = \frac{1}{P_s} = \frac{N}{2f\text{NA}} \). Now, in the focusing problem we will typically only be interested in computing to the focal fields where there is a reasonable energy density. In the absence of aberrations (and at low NA), 91% of the focal plane energy is concentrated inside the second diffraction zero [246], so we are probably only interested in calculating up to a few times the Airy disc diameter. The diameter of the airy disc is \( D_{\text{Airy}} = \frac{1.22 \lambda}{\text{NA}} \).

To scale from position in the focal plane to lateral spatial frequency, we multiply by \((f\lambda)^{-1}\), obtaining the frequency space diameter of the Airy disc, \( F_{\text{Airy}} = \frac{1.22}{f\text{NA}} \).

So, given that the interesting part of frequency space is only a few multiples of \( \frac{F_{\text{Airy}}}{2F_{\text{max}}} \) about the zero frequency component, only about 1 pixel in the output plane is useful. We find that \( N = 100 \) gives reasonable sampling of the pupil for our purposes (this agrees with suggested values in the literature, for example [71,214]). The FFT algorithm allows us to specify the sampling frequency in frequency space by zero-padding the input. If we zero pad the input arrays from \( N \times N \) to \( M \times M \) then we would expect to have roughly \( \frac{M}{N} \) pixels across the airy disc. If we want to see, say 50-100 pixels across the output PSF, then we need to use \( \frac{M}{N} = 10^4 \) which makes the FFT time consuming to calculate.

The chirp-z transform, also known as Bluestein’s FFT algorithm, is an alternative method for calculating DFTs and can be used for calculating DFTs over a subregion of frequency space to arbitrary sampling precision [247–249]. Let \( x \) and \( \hat{x} \) be transform pairs having length \( N \) and \( K \) respectively. The chirp-z transform allows \( x \) and \( \hat{x} \) to have different lengths. The algorithm work as follows. First, start with the definition of the DFT of \( x \) [245],

\[
\hat{x}(k) = \sum_{n=0}^{N-1} x(n)S^{-nk},
\]

where \( S = \exp(2\pi i/N) \). Now, notice that \( nk = \frac{1}{2}n^2 + \frac{1}{2}k^2 - \frac{1}{2}(n-k)^2 \) and therefore

\[
\hat{x}(k) = S^2k^2 \sum_{n=0}^{N-1} x(n)S^{-\frac{1}{2}n^2} \cdot S^{\frac{1}{2}(n-k)^2},
\]
which has the form of a convolution. So we can compute $\mathcal{F}$ by performing the discrete convolution of $A = x(n)5^{-\frac{1}{2}n^2}$ and $B = 5^{\frac{1}{2}n^2}$. By the convolution theorem this can be efficiently performed by computing the DFT of both $A$ and $B$, then multiplying the results and performing an inverse DFT. We have replaced one DFT of size $M$ with three of size $K + N - 1$ [214,250]. As a typical example, we computed the PSF of a pupil with $N = 128$ sampling points in each direction in the pupil. Requiring 128 sample points across the central region means that the FFT method uses padded 2D arrays of size $16384 \times 16384$ pixels, and takes 10 sec to compute on a laptop PC with a quad-core Intel i7 processor and 16 Gb of RAM. By comparison, the chirp-z transform performs its 3 DFTs on arrays of size $257 \times 257$ and takes 150 ms, a speed improvement of near 100 times. At this speed of computation, we can easily calculate 3D PSFs, for example as shown in Figure 6.6. The focal fields are calculated over a volume containing $50 \times 50 \times 50$ pixels, and the computation takes about 0.6 s per 3D PSF. The form of the pupil function is not important here, we just aim to demonstrate the capability of our approach.
Figure 6.6. Visualisation of the focal field energy density at the focus of a microscope objective having NA = 0.92. The pupil is the coherent sum of a LCP step beam and an RCP helix, which at focus has a dark spot surrounded by bright regions and is used in STED microscopy. The surfaces shown are contours of constant electric field energy density and represent 50%, 65% and 80% of the peak values. The results were computed in 1.8 seconds on a quad-core Intel i7 laptop with 16 Gb RAM.

6.2.3. Example focal field calculations

As a quick illustrative example we show the time averaged focal field energy densities of a radially polarised beam illuminating the pupil of a microscope with NA = 0.95 in Figure 6.7. As reported widely elsewhere (for example [79]) there is a strong axial electric field component at focus. There is no violation of Maxwell’s equations here because the magnetic field is zero at focus, so there is no Poynting flux. As will be seen below, the radially polarised beam is useful for determination of the orientation of a single fluorescent emitter in 3D. We have calculated these fields using 100×100 points in the pupil plane and 256×256 points in the focal plane. The calculations took 32 ms.
6.3. **Collection of light emitted from a dipole**

In this chapter we aim to build a full model of a confocal fluorescence microscope. Up to this point we have described an efficient method for computation of the focal fields of the illuminating objective. In this section we refer to these fields as $E_{\text{exc}}$ to distinguish them from radiated dipole fields. We now turn our attention to the collection and detection of light from fluorescent samples. We consider a model of the behaviour of a scanned dipole in a microscope with finite sized pinhole. The microscope consists of a high NA illumination objective and high NA collection objective (which in practice are usually the same objective). The objective is in a 4F arrangement with a tube.
lens and at the focus of the tube lens is a confocal pinhole. Light passing through this pinhole is integrated.

Figure 6.8 shows the geometry for transforming the electric field vector from the focal sphere (shown in blue) to the pupil plane (shown in green). A ray, travelling along $s$, has polarisation along $E_1$. The ray is refracted to be perpendicular to the optic axis. The polarisation vector in the pupil is $E_0$.

### 6.3.1. Pupil fields for a dipole at the Gaussian focal point

Suppose there is a dipole at the Gaussian focus of an objective. Let the dipole orientation be along the unit vector $d$. We wish to calculate the field from this dipole on the focal sphere, and then project this field to the pupil plane. The radiating field from a dipole observed at position $s$ on the focal sphere is [9,181,230,251,252] (up to multiplicative factors)

$$E_1 = s \wedge d \wedge s$$  \hspace{1cm} (80)

Again, here we use the subscript 1 to refer to quantities on the focal sphere, and the subscript 0 for the pupil plane. Now, since the cross product is distributive over addition, the electric field on the focal sphere can be written as a sum of fields from three orthogonal dipoles with appropriate strengths. The
transformation of $E_1$ onto the pupil plane field $E_0$ and the focusing down of $E_0$ to the focus of the tube lens are all linear operations, so we can write the fields at the focus of the tube lens as the sum of the fields from three basis dipoles at the focus of the objective. The strength of the contribution from the $x$ dipole would be given by $d \cdot x$ and similarly for $y$ and $z$.

To convert the fields on the focal sphere to the fields in the pupil plane, we decompose $E_1$ into its radial and angular parts and state that the projections onto the radial and azimuthal vectors in the pupil plane are the same as on the focal sphere. This is equivalent to the Richards-Wolf assumption that the angle that the polarisation vector makes with the meridional plane remains the same during the transformation. We can therefore write

$$E_0 \propto (E_1 \cdot r_1) r_0 + (E_1 \cdot \phi_1)\phi_0,$$  

where $r_1$ is the unit vector perpendicular to the ray propagation direction $s$ and points in direction of increasing $\theta_1$, $\phi_1 = s \wedge r_1$ is the unit vector in the direction of increasing $\phi_1$ and $r_0$ and $\phi_0$ are the radial and azimuthal vectors in the pupil plane. These are all shown in Figure 6.8. Now, since $(\hat{d} \cdot \hat{b})\hat{c} = (\hat{c} \otimes \hat{b})\hat{d}$ where $\otimes$ denotes the outer product, we can rewrite (81) as

$$E_0 = \frac{(r_0 \otimes r_1) E_1 + (\phi_0 \otimes \phi_1) E_1}{\sqrt{\cos \theta}} = \frac{r_{01} E_1}{\sqrt{\cos \theta}},$$  

where $r_{01}$ is a matrix that transforms $E_1 \rightarrow E_0$ and includes the proportionality constant that accounts for the conservation of flux in an objective obeying the sine condition [253]. This was erroneously omitted in some previous work [181], but has been correctly included elsewhere [252,254,255]. Now, $r_1 = (\cos \theta_1 \cos \phi, \cos \theta_1 \sin \phi, -\sin \theta_1)$ and $r_0 = (\cos \phi, \sin \phi, 0)$, where the subscripts on $\phi$ have been dropped as $\phi_1 = \phi_0$. Also, $\phi_0 = \phi_1 = (-\sin \phi, \cos \phi, 0)$. Using these relations, we can write

$$r_{01} = \begin{pmatrix} \cos^2 \phi \cos \theta + \sin^2 \phi & (\cos \theta - 1) \sin \phi \cos \phi & -\sin \theta \cos \phi \\ (\cos \theta - 1) \sin \phi \cos \phi & \sin^2 \phi \cos \theta + \cos^2 \phi & -\sin \theta \sin \phi \\ 0 & 0 & 0 \end{pmatrix}.$$  

We note here that a ray originating from the Gaussian focal point will always end up travelling parallel to the optic axis. If we want $r_{01}$ to represent a rotation
matrix then we can find the third row of \( r_{01} \) by imposing \( \det(r_{01}) = 1 \). In that case, we find

\[
r_{01} = \begin{pmatrix}
\cos^2 \phi \cos \theta + \sin^2 \phi \\
(c \cos \theta - 1) \sin \phi \cos \phi \\
\sin \theta \cos \phi
\end{pmatrix} \begin{pmatrix}
\cos \theta - 1 \sin \phi \cos \phi \\
\sin^2 \phi \cos \theta + \cos^2 \phi \\
\sin \theta \sin \phi
\end{pmatrix}
\]

Which agrees with previous results [13,181]. Of course (83) and (84) give the same results for a ray originating from the Gaussian focal point. We account for dipoles away from the focal point later in this section. Combining the expression for the dipole fields on the focal sphere, equation (80), and the transformation of fields from the focal sphere to the pupil plane, equation (82), we can calculate fields on the pupil plane from a radiating dipole at the objective focus.

Figure 6.9 shows the computed pupil fields for an objective with \( \sin \alpha = 0.92 \) for dipoles along \( x \) and \( z \). Notice that the amplitude of the pupil function is not uniform in either case. Roughly 35% of the total radiated energy is collected in the case of the \( x \) dipole, but only 22% in the case of the \( z \) dipole. By integrating \( |s \wedge d \wedge s|^2 \) over the pupil it is easy to show that the fraction of dipole radiation collected is

\[
\eta_c = \frac{3}{4} (1 - \cos \alpha) - \frac{1}{4} \cos^2 \theta_d (1 - \cos^3 \alpha)
- \frac{1}{4} \sin^2 \theta_d \left( 1 - \frac{3}{2} \cos \alpha + \frac{1}{2} \cos^3 \alpha \right),
\]

where \( \alpha \) is the semi-angle of the aperture of the objective and \( \theta_d \) is the angle that \( d \) makes to \( z \). \( \eta_c \) is known as the collection efficiency and is plotted in Figure 6.10. On the left, we show the collection efficiency for an objective with \( \sin \alpha = 0.92 \) and on the right we show the ratio of the collection efficiencies for dipoles lying along \( z \) and \( x \) as a function of \( \sin \alpha \). The insets show example dipole radiation patterns (cross section in \( xz \) plane) for \( x \) and \( z \) oriented dipoles in blue overlaid with the flux collected by an oil immersion objective with \( \sin \alpha = 0.92 \) in grey. As a further example, see Figure 6.11. Here we have plotted the fields on the focal sphere and in the pupil plane from a dipole at 60 degrees to the \( z \) axis. These can be computed directly from (80) or we can add fields from the \( x \) and \( z \) dipoles in the appropriate weights.
Figure 6.9 shows the collected pupil fields of an objective with NA = 0.92 with a radiating dipole at the Geometric focus. On the left, the dipole is along the x axis, and on the right the dipole is along z. These fields are computed by computing $E_1$ on the focal sphere using equation (80) and then transforming to the pupil using equation (82).

Figure 6.10. (L) Collection efficiency vs dipole angle for an objective with, $\sin \alpha = 0.92$. Insets show radiation patterns in blue and flux collected in grey. (R) Ratio of z collection efficiency to x collection efficiency as a function of NA/n. NA of 1.4 in oil is shown in blue.
6.3.2. Confocal image of a dipole

Now we turn our attention to computation of the confocal image of a dipole. As mentioned before, a dipole in a general direction can be considered as the sum on three orthogonal dipoles. We can therefore pre-calculate the tube lens focal fields, say $E_T$, for dipoles along $x$, $y$ and $z$, and add them in the correct proportions when we want to see the image of a specific dipole. An infinity corrected microscope has the objective and tube lens separated by the sum of their focal lengths. We have shown above how to calculate the field at the pupil plane of the objective, which is also the pupil plane of the tube lens. So we can calculate the field at the focal plane of the tube lens in the same way as for an objective, though the depolarising high-NA effects are not important. We can see that the depolarising effects are not important with the help of equation (84) as follows. We wish to get an idea of the strength of the depolarisation in going from pupil to focal sphere. Now, $r_{01}$ gives the rotation of rays or polarisation vectors from focal sphere to pupil plane, so we are interested in $r_{01}^{-1} = r_{01}^T$ (because $r_{01}$ is a rotation matrix). So, if we investigate the strength of $y$ and $z$ polarised fields introduced for an $x$ polarised beam, we need to refer to the top row of $r_{01}$. Now we can see that the strength of the $y$ component is proportional to $\sin^2(\theta/2)$ and the strength of the $z$ polarised component is proportional to...
sin $\theta$. For the tube lens, the maximum value of $\sin \theta$ is $\text{NA}_T = \text{NA}_0/M \approx 10^{-2}$. After adding the electric fields in the appropriate strengths we work out the total electric energy density. This gives the wide-field image of a dipole at the Gaussian focus of the objective. The confocal detector integrates all signal in the tube lens focal plane passing through a pinhole having radius $p$. In the confocal setup an image is formed by scanning the sample (or the excitation beam and detector) and recording the signal as a function of sample position (or beam/detector position). We now wish to calculate the confocal image as the sample or beam is scanned in the $xy$ plane (we will consider movement in the $z$ direction in 6.5). Imagine the sample is scanned through the excitation beam (for a system obeying the sine condition we could equivalently think about scanning the beam). As the sample moves to a scan position $r_s = (x_s, y_s)$, the image of the dipole moves to $r_i = (-Mx_s, -My_s)$ where $M$ is the magnification [15]. At each scan position, we collect and integrate all light passing through the pinhole and reject the remainder. This is equivalent to convolving the low NA image of the dipole with a disc of radius $p$. Finally, we multiply the resulting image (which is an intensity as a function of $(x_s, y_s)$) by the strength of the dipole radiation at that particular scan position, given by $|d \cdot E_{\text{exc}}|^2$. The overall image of the dipole is then

$$\text{Image of dipole} = h_D(x_s, y_s) = |d \cdot E_{\text{exc}}|^2 \cdot (|E_T|^2 \ast \text{Pinhole}).$$

(86)

where $\ast$ denotes convolution. Here we are effectively assuming that the dipole orientation remains fixed in the time between excitation and emission. This is likely to be valid if the fluorescent dipole is fixed in space or coated onto a substrate, as is investigated in this thesis (see section 9.1) but for molecules in solution we would need to consider rotation of the dye molecule between excitation and emission [252].

Figure 6.12 shows calculated confocal images of six dipoles when the illumination pupil is $x$-polarised for a microscope with illumination and collection $\text{NA} = 0.92$. We have set the pinhole radius to be $1.5 \times 1.22 \pi$ in normalised optical coordinates. In confocal microscopes normally the pinhole radius is set to be $1.22 \pi$ but our microscope requires a large pinhole to capture
signal from axial dipoles, which have a doughnut-shaped image. We have also assumed that the emission and excitation wavelengths are the same. We justified this assumption in section 2.4. For the dye used in our experimental work, DiI, the Stokes’ shift is approximately 50 nm with an excitation wavelength of 530 nm. Therefore the detection PSF is only 10% larger than the excitation PSF, which will not affect the calculated images significantly. The importance of the pinhole radius is explained in section 6.4. The dipole orientations are shown in the titles of each subfigure, given as $[\theta, \phi]$ where $\theta$ is the angle that $\mathbf{d}$ makes with $z$ and $\phi$ is the angle that $\mathbf{d}$ makes with $x$. As expected, the peak signal is strongest when the dipole is oriented along the $x$ axis, having a value of approximately 100% (arbitrary scale), and gets weaker as the dipole turns in the $xy$ plane towards $y$ where the signal drops to 1%. As the dipole turns from $x$ to $z$ the signal also gets weaker, but only falls to 10% because some $z$-oriented dipoles are excited by the $z$-component of the focal field (see Figure 6.5).

Figure 6.13 shows the same six dipoles illuminated with a radially polarised pupil. The focus of the radially polarised beam is longitudinal at centre but is surrounded by a ring of in-plane polarised light in the focal plane, as shown in Figure 6.7. The dipole appears with greatest peak intensity when it is along the $z$ direction and is excited by the strong longitudinal focus of the radially polarised beam. However, now consider the integrated intensity, important for dipole orientation determination in densely labelled samples. It is roughly 25% for the dipole along $z$, and rises to 35% for the dipole along $x$ (on the same intensity scale as the values for the $x$-polarised beam above). This presents a difficulty in determining $z$-dipole orientation and is discussed further in section 6.4.
Figure 6.12. Confocal images of six dipoles imaged with an $x$-polarised pupil at $\text{NA}/n = 0.92$. The orientation is specified by a vector $[\theta, \phi]$ where $\theta$ is the angle to the $z$ axis and $\phi$ is the angle to the $x$ axis. $v_x$ and $v_y$ are transverse optical coordinates. Colour scale is normalised to the maximum intensity in the set. The dipole appears brightest when it is aligned with the excitation polarisation direction along $x$.

Figure 6.13. Confocal images of six dipoles imaged with a radially polarised pupil. Parameters are the same as those in Figure 6.12. The intensity scale is different from in Figure 6.12. The dipole now appears with greatest peak intensity when it is aligned to the $z$ axis, however the greatest total signal in the image is for an in-plane dipole. This presents a difficulty in dipole orientation determination in 3D.
6.4. **Determination of dipole orientation**

We propose determining the orientation of dipoles by obtaining a number of images with different illumination polarisations. The scheme for determination of dipole orientation would be different in sparsely and densely labelled samples. Here, sparse means that the image of each individual dipole is well separated from other dipoles so that the characteristic shape can be observed. On the other hand, in a densely labelled sample there would be many fluorophores inside any given resolution element.

In the sparsely labelled regime dipole images from single fluorophores will be well separated so that it is easy to judge which pixels correspond to each molecule. We could then integrate all pixels corresponding to each dipole to give a total response for each molecule under each illumination pattern. These intensities would be compared for each molecule in the image, similar to the description by Fourkas [181]. This provides a benefit over pattern matching techniques because no fitting is required. For a qualitative idea of dipole orientation we could simply estimate dipole orientation by comparing the observed dipole images with those predicted by our model.

If the sample were densely labelled, so that there are many fluorophores contributing to each pixel, the intensity at each pixel would be directly used to form the pertinent ratios. In both qualitative methods it is the integrated intensity from each dipole image in Figure 6.12 and Figure 6.13 that is the important parameter. With these integrated intensities we might proceed to calculate the orientation of the fluorophores ratiometrically in a manner similar to FDLD [31,181,191]. Orientation determination in densely labelled samples adds a further difficulty namely, unless known a-priori, the angular distribution must be determined as well as the average orientation [27,207].
6.4.1.  **3D orientation determination in densely labelled samples**

As mentioned above, in order to determine the dipole orientation in a densely labelled sample we would need to compare pixel values between a set of images taken with different illumination polarisations. For the in-plane case, simulation results in Figure 6.12 show that a dipole will show up as brightest in the image where the illumination polarisation is aligned with its dipole moment. For orientation determination out of the plane, Figure 6.13 shows that a dipole along \( z \) has highest intensity in the image taken with the radially polarised pupil. The signal from the \( z \) dipole is not as strong as might be expected because a lower proportion of its radiation is collected by the objective as highlighted in Figure 6.10. In addition, there is a strong background from the in-plane dipoles that are excited by the ring of in-plane polarised light surrounding the longitudinal focal spot. There is a further difficulty, which is that by symmetry the low NA image of a \( z \)-oriented dipole has a zero on axis \([183,256]\). Therefore, if we use a small pinhole, the signal from a \( z \)-oriented dipole at position \( r_z \) does not contribute to the signal at \( r_i \), the corresponding pixel in the image \([217,257]\). In addition, the signal at \( r_i \) could have contributions from \( z \)-fluorophores at \( r_z \) or in-plane fluorophores a small distance away from \( r_z \). Using a larger confocal pinhole partly negates these effects, because a large pinhole will result in the peak signal from a \( z \)-dipole at \( r_z \) being observed at the corresponding image point. These effects are illustrated in Figure 6.14. On the left we show the ratio of the peak \( z \) signal to peak \( x \) signal for radially polarised illumination and on the right we plot the total confocal signal as a function of angle to \( z \) axis for a pinhole radius of 3 Airy units. A problem presents itself: the total confocal signal is actually larger for an \( x \) dipole than for a \( z \) dipole for all pinhole radii less than about 4 Airy units as the \( z \) dipoles are not being efficiently detected. The signal from the in-plane dipoles comes from the ring of excitation surrounding the focus (see Figure 6.13). These effects show that radial illumination cannot be used to determine the polar angle of the dipoles in a confocal microscope by simply comparing intensities.
Figure 6.14. (L) The effect of pinhole size on the ratio of peak confocal signal strengths with a radially polarised pupil. (R) The total confocal signal as a function of dipole angle to the z axis (0 = along z) for a pinhole with size 3 Airy units.

6.4.2. Feasibility of orientation determination in 3D

We have shown that in determining the three dimensional orientation of a dipole, it is the angle to the optic axis that is difficult to determine. In order to measure the angle to the optic axis well one must be able to excite and collect fluorescence efficiently from molecules whose fluorescent dipoles are oriented along or nearly along the optic axis. Here we suggest a number of possible modifications that may allow the dipole polar angle to be determined.

Apodization of radial beam. One simple option is to apodize the radial beam. A greater contrast between \( xy \) and \( z \) excitation can be engineered by reducing the intensity near the centre of the pupil. This option is of practical interest because it can easily be achieved with a spatial light modulator. A natural choice of apodization function is to use the pupil field that would be collected from a \( z \) oriented dipole as the pupil field for excitation (henceforth referred to as the \( z \) dipole pupil). This has radial polarisation but an amplitude profile that increases with radius from the beam. This is plotted in Figure 6.15.
Figure 6.15. The ‘z dipole beam’: field amplitude as a function of radius.

Figure 6.16. (L) The effect of pinhole size on the ratio of peak confocal signal strengths with a z dipole pupil. (R) The total confocal signal as a function of dipole angle to the z axis for a pinhole with size 3 Airy units.

Figure 6.16 shows the same information as Figure 6.14 but for the z dipole pupil instead of the basic radially polarised beam. With this illumination the signal is stronger for a z-oriented dipole than for an in-plane dipole (for all but the smallest pinhole sizes). The right hand plot shows that the signal is roughly constant with polar angle for highly inclined dipoles, which means we will only be able to determine the polar angle when it is greater than about 30 degrees using this technique. There has been an overall reduction in signal compared to the radially polarised beam.

**Increase objective numerical aperture.** Figure 6.10 shows the dependence of the collection efficiency as a function of objective numerical aperture. Any increase in numerical aperture will help to increase the relative strength of the highly inclined dipoles. For the highest NA commercial objective, $NA = 1.47$ in Oil, the signal improvement is not as good as using a z dipole beam for excitation, though there is no loss of signal.
**Engineering of pinhole.** By adding a central obstructive disk to the pinhole we could preferentially select the $z$ illumination. This would also reduce the signal from the in plane dipoles with linearly polarised illumination, though this may be a penalty worth paying.

**Engineering of emission pupil.** An alternative to adding a central mask to the pinhole would be to introduce a $\pi$ phase step (as in [183,256]) into the emission pupil, or introduce a polarisation mode converter (as in [258]). Both of these cause the signal from the $z$ oriented dipoles to be located on axis in this pinhole plane, and the signal from the in plane dipoles to be located off axis.

**STED microscopy.** We have seen that a radially polarised beam excited axial dipoles at the Gaussian focal point and in-plane dipoles just outside the focal point (in the focal plane). We could use a circularly polarised doughnut mode (as used in STED microscopy [56,58]) to deplete these same dipoles. Using STED and a radially polarised beam would significantly reduce contributions from unwanted dipoles, though at some inconvenience and expense.

6.4.3. **Aberrations and the need for a test dipole**

Before continuing we highlight that an incident beam is unlikely to reach its intended position without the introduction of aberrations. The pupil functions for the different excitation polarisations have different symmetries and so they will be affected in different ways by aberrations [240]. We can therefore expect the relative signal strengths to alter in the presence of aberrations. This will affect our angle determination, and for a careful quantitative determination of dipole orientation we would need test dipoles with known orientations to calibrate the system [183].

6.5. **Vectorial calculation of optical sectioning**

In section 6.3 we considered the image of a fluorescent dipole at arbitrary angle as it is scanned in the focal plane. We now consider moving the sample in the axial direction to study the phenomenon of optical sectioning.
Figure 6.17 shows the relationship between the phase accumulated for a given plane wave and the axial displacement. The thick black line is the focal sphere, radius \( f \). The radius in the pupil plane is given by \( r = \rho f \sin \alpha \) where \( 0 \leq \rho \leq 1 \) is the scaled pupil radius.

In an aplanatic system, where the sine condition is obeyed [259], lateral movement of the sample causes a lateral shift in image position. Let us investigate what happens when the sample is moved along the optic axis. Consider a point source displaced by a distance \( \delta z \) from the focal plane in image space and a plane wave at angle \( \theta \) to the optic axis. The extra phase in the pupil for this plane wave is given by

\[
\Phi = n k \delta P = n k \delta z \cos \theta = n k \delta z \sqrt{1 - \sin^2 \theta},
\]

where \( n \) is the refractive index in the sample space and \( k \) is the magnitude of the free space wavevector. See Figure 6.17 for a geometric illustration of the coordinates. If we now introduce a scaled pupil radius \( \rho \) so that \( \rho(\theta) = (f/R) \sin \theta \) and \( \rho(\alpha) = 1 \) then \( \rho \sin \alpha = \sin \theta \) and we can rewrite

\[
\Phi = n \delta z k \sqrt{1 - \rho^2 \sin^2 \alpha} = \delta z k \sqrt{n^2 - \rho^2 NA^2}.
\]

Now in the case of a high NA microscope objective and low NA tube lens in a 4F arrangement, the wavefront shape in the objective pupil, which is also the pupil of the tube lens, is

\[
\Phi_O = \delta z k \sqrt{n^2 - \rho^2 NA^2_O}.
\]

Here \( NA_O \) is the objective numerical aperture and \( n \) is the refractive index in sample space, which is assumed to be uniform. Now we consider the pupil field if the point source were in image space at \( \delta z' \) and calculate the pupil field of the tube lens. The NAs of the tube lens and the objective are related by \( NA_O = M \cdot NA_T \) where the subscripts denote objective lens (\( O \)) and tube lens (\( T \)). We have
\[ \phi_r = \delta z' k \sqrt{1 - \rho^2 \frac{NA_0^2}{M^2}} . \]  

(90)

Consider a Taylor expansion of \( \phi \) and compare terms. The quadratic terms give us the relation

\[ \delta z' = \frac{M^2}{n} \delta z , \]

(91)

which is the expression for the axial magnification of the system. Now consider the next terms, which are fourth order in \( \rho \). We get

\[ \phi_0^{(4)} = -\frac{1}{8} \delta z k \rho^4 NA_0^4 \cdot \frac{1}{n^3} \]

(92)

for the objective, where the superscript in parenthesis indicates the fourth order term, and

\[ \phi_r^{(4)} = -\frac{1}{8} \delta z k \rho^4 NA_0^4 \cdot \frac{1}{n M^2} = \frac{n^2}{M^2} \phi_0^{(4)} \]

(93)

for the tube lens. So displacement of the object point in object space is not equivalent to displacement of the image point (by a magnified amount) in image space because fourth order, and higher orders, of spherical aberration are present in the image [15,253,260]. In this derivation we have assumed that both the objective and tube lens obey the sine condition. If instead we had assumed the Herschel condition then we would find an equivalence between axial displacements, but lateral object displacements would no longer just cause shifts in image space [253].

Now we turn to the problem of calculating the optical sectioning properties of our confocal microscope. We imagine imaging a uniform thin fluorescent sheet extending in the \( xy \) plane and translating it along the \( z \) axis while recording the total signal in our confocal image. Our full focusing and collection model allows us to study the effects of different illumination pupil functions and also of fluorescent dipole orientation in the sample. To our knowledge, such a study has not been performed up to now. Wilson has made an extensive study of the optical sectioning in high NA systems [41], but by using Sheppard’s pseudoparaxial approximation [14] he did not account for vectorial effects. To calculate the optical section, we start with equation (86), the expression for the in-focus image of a single dipole \( h_D(x_o, y_o, z_t) \). We have now included \( z_t \) which is
the $z$ position of the dipole in object space. To find the image of a fluorophore distribution $f(x, y, z)$, we calculate the convolution of $f$ and $h_D$

$$I(x_s, y_s, z_1) = \iiint_{x,y,z} h_D(x, y, z) \cdot f(x_s - x, y_s - y, z_1 - z) \, dx \, dy \, dz . \quad (94)$$

For a uniform fluorescent sheet, we write $f = \delta(z)$, so (94) becomes

$$I(x_s, y_s, z_1) = S(z_1) = \iiint_{x,y} h_D(x, y, z_1) \, dx \, dy . \quad (95)$$

This shows that to calculate the signal strength $S$ for a uniform sheet at $z_1$ we can just integrate the pixels in the image of the single dipole at position $z_1$. Given that there is a pinhole we expect $S$ to be a maximum for the in focus case when $z_1 = 0$ and then to decrease as we move away from focus. Figure 6.18 shows some example plots of $S(z_1)$ for three different pinhole sizes, in the pseudoparaxial approximation as in Wilson’s work [41,621]. In this figure we have plotted the axial normalised optical coordinate on the horizontal axis, as defined in 2.2. We have also separately normalised the vertical axes to be 1 in focus, though of course the signal is stronger when the pinhole is larger. It is seen that the best sectioning is obtained for an infinitely small pinhole (solid line), though not much sectioning is lost as the pinhole size is increased to 1 Airy unit. For these two cases, we have true optical sectioning in that $S$ falls to zero at approximately the same $z$ values as the axial FWHM of the point spread function [259], so all out-of-focus light is rejected from the image. The sectioning is appreciably worse when the pinhole is 3 Airy units and there will be some out-of-focus light in the final image.

\footnote{A quick reminder: 1 Airy unit is 1 unit of transverse normalised optical coordinate divided by $1.22\pi$, so that 1 Airy unit is the radius of the Airy disc.}
Figure 6.18 shows some example plots of the signal strength as a thin fluorescent sheet is scanned through the focus of a confocal microscope. The three lines show the signal strength with different pinhole radii $p$, given in Airy units in the legend. The legend also shows the FWHM of the function, a measure of the sectioning strength. The horizontal axis is in units of the axial normalized optical coordinate $u_{\text{Im}}/4\pi$. These calculations ignore vectorial effects and use the pseudoparaxial approximation as defined by Sheppard [14].

Now we turn to our full vectorial optical sectioning calculations. First, imagine our sheet contains dipoles with random orientations. In that case we calculate the 3D PSF, obtained by equation (86), then average over the dipole orientation and integrate over $xy$. We then find the FWHM of the resulting functions as a function of the pinhole radius. The results are shown in Figure 6.19 for radial and right-circularly polarised pupils for a microscope with an oil immersion objective having NA = 1.4. As in our calculations of the images of dipoles, we have assumed that the emission and excitation wavelength are equal. Based on previous work, we would expect the optical sectioning performance to be slightly worse for realistic values of Stokes’ shifts [41]. Wilson’s result [261] for equal excitation and emission wavelengths is also plotted for comparison. First, notice that there is good agreement between Wilson’s result and the full vector theory at small pinhole sizes. At larger pinhole sizes Wilson underestimates the optical sectioning strength. We believe that this is because his calculation uses the low NA form of defocus (i.e. a parabola) whereas the actual phase introduced is a spherical shape (see for example equation (88)). At small pinhole sizes, radially polarised illumination results in worse optical sectioning performance
than circularly polarised illumination but there is little difference as the pinhole size increases. At first sight it feels odd that for radially polarised illumination the optical sectioning improves as the pinhole size is increased to 1 Airy unit. This can be explained by noting that for the radially polarised illumination pupil, the detection PSF will have a large contribution for z-oriented dipoles. When the pinhole is small, the detection PSF is just the low NA image of a z dipole, which is annular in planes of constant z. When we take the product of the annular detection PSF and the excitation PSF a larger axial extent of the PSF will be expected, which results in worse sectioning strength. As we increase the pinhole size the low NA annular shape will start to fill in the centre, so we will start sampling the central part of the excitation PSF, which has a smaller axial extent.

Figure 6.19 shows the FWHM optical section thickness for a sheet of randomly oriented dipoles as a function of pinhole radius for the full vectorial theory with dipoles, shown in blue and red, and Wilson’s calculation in yellow.

Now we consider how the models match as we vary the NA of the system. We refer to the plots shown in Figure 6.20. In each of the three plots shown the FWHM optical section thickness is calculated as a function of pinhole radius. The three plots show the calculations at NA = 0.05, NA = 0.5 and NA = 1.4. On each plot we show three functions. The first, in blue, is the full vectorial description with a sheet of randomly oriented dipoles scanned through the focus with circularly polarised excitation in the illumination pupil. The second, shown in red, is a simplification where we consider high NA effects in the fields but ignore dipole effects in the sample. This speeds up calculation as it removes the need to average over an ensemble of dipoles in the sheet. In this model the full PSF is modelled as the detection PSF multiplied by the excitation electric field density.
In yellow we have plotted the Wilson result using the pseudoparaxial approximation. At low and moderate NA, all three theories give very similar results, though Wilson predicts a marginally thicker optical section. At high NA there is a larger discrepancy, though the full vectorial theory and the simplified version ignoring dipole effects are indistinguishable. Given that the theories are all in reasonably good agreement with each other, one could view these results as a verification of the pseudoparaxial approximation for calculating optical sectioning thicknesses. The agreement between Wilson’s result and the full theory could be improved if Wilson had used the full spherical form for the phase error introduced with defocus.
Figure 6.20 shows plots of the optical section FWHM thickness as a function of pinhole radius for three different models, as explained in the text. The results are calculated at low NA (top), moderate NA (middle) and high NA (bottom).

Finally we imagine imaging thin sheets of uniformly aligned dipoles. Calculation of the optical sectioning properties in this case would not be possible without our full vectorial theory. We show the results in Figure 6.21. The upper plot shows the FWHM thickness as a function of pinhole radius for four cases: all combinations of two pupil functions and two dipole orientations. We predict that the optical section thickness is fairly similar in all cases except for a radially polarised beam exciting a sheet of x dipoles.
Figure 6.21 shows plots of the optical section FWHM thickness with sheets composed of aligned dipoles. We consider sheets made of uniformly aligned dipoles, along x and along z, and two illumination pupil functions, linearly polarised along x and radially polarised.

6.6. Chapter summary

In this chapter we have outlined some previous work and developed our own models. We started with the goal of understanding the fields at the focus of a high NA microscope objective from first principles. Our emphasis was on providing an understanding of the concepts rather than giving a full mathematical derivation. Starting with Maxwell’s equations, we described how Kirchoff arrived at his solution to the scalar Helmholz equation. This put the ideas of Huygens and Fresnel, formulated 200 years earlier, onto a sound mathematical footing. We then described the Debye theory for focusing light, which studies the focal fields of a focusing system, and its extension to high NA systems, the Richards-Wolf theory. This is a trusted theory and is well known to accurately describe the focal fields of a high NA lens. For quick reference, we provide Table 6.1 that shows a summary of the steps taken to reach the Richards-Wolf theory from Maxwell's equations, and the assumptions made along the way. We then described the extension of this theory to arbitrarily polarised pupils and reformulated the equations in terms of Fourier transforms. When computed with the chirp-z transform algorithm, we were able to calculate focal fields of a microscope having arbitrary pupil polarisation on a 256x256 grid in a speed of around 30 ms on a laptop computer.
We then described our development of a full vectorial model of a confocal fluorescence microscope that accounts for the high NA collection of fluorescent light from a radiating dipole, the focusing of this light by a tube lens, and the convolution with a confocal detector. This is new work: the models of focal fields formation and collection of polarised light from a fluorescent emitter have not been combined before. Based on this model we made some predictions about the images of dipoles under various illumination conditions, and were able to evaluate the feasibility of determining dipole orientation using a microscope with a controlled pupil function. We proposed illuminating a sample containing a collection of dipoles with a sequence of polarisations and recording the intensity from each dipole in each image. The proposal was then to determine the dipole orientation by comparing these intensities. We argued that it would be very difficult to determine the polar (out-of-plane) angle of the dipoles using a radially polarised beam and proposed some solutions. Based on these thoughts, we focused our experimental efforts on imaging sparsely labelled samples. We would hope to develop the method further for imaging in dense samples.

We then proceeded to calculate the optical sectioning properties of a confocal microscope using our full vectorial model. To our knowledge, this has not been performed before. Previous work had used the pseudoparaxial approximation that accounted for high ray angles but not polarisation effects during focusing or dipole effects during excitation and collection. We were able to show that the results of our full model match the pseudoparaxial approximation reasonably closely when the illumination pupil polarisation is circularly polarised and samples are made of a large collection of randomly oriented dipoles. Since our model accounts for more physical effects than the pseudoparaxial approximation, we can view our results as a validation of the pseudoparaxial approximation. Our model also allows us to calculate the optical sectioning in samples with many aligned dipoles, and we present some results of simulations comparing the sectioning strength for sheets made of x and z oriented dipoles.
<table>
<thead>
<tr>
<th>Theoretical framework</th>
<th>Key assumptions added/conceptual steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maxwell’s equations</td>
<td></td>
</tr>
<tr>
<td>Helmholtz equation for scalar fields</td>
<td>Linear, isotropic, homogenous media. Diffraction aperture $&gt;&gt;$ wavelength. Observation point ‘far from’ aperture. This removes effects of coupling of fields at the aperture.</td>
</tr>
<tr>
<td>Integral theorem of Helmholtz &amp; Kirchoff</td>
<td>Use free space Green’s function to solve the Helmholtz equation.</td>
</tr>
<tr>
<td>Fresnel-Kirchoff diffraction formula</td>
<td>Kirchoff boundary conditions, distance to aperture $&gt;&gt;$ lambda.</td>
</tr>
<tr>
<td>Debye diffraction integral</td>
<td>Low NA focusing and large Fresnel number. This amounts to saying there must be appreciable convergence, i.e. if diffraction from the aperture dominates over refraction by the lens the theory will not give accurate results. Valid for points close to the geometric focus compared to focal length.</td>
</tr>
<tr>
<td>Richards-Wolf diffraction integrals</td>
<td>Accounting for rotation of polarisation at high NA and conserve energy gives the Richards and Wolf theory. Assume that angle of incidence on any glass element is small.</td>
</tr>
</tbody>
</table>

Table 6.1 shows the steps taken and assumptions made to reach the Richards-Wolf diffraction theory from Maxwell’s equations.
7 Design of a stationary pupil scanner

In this chapter we present the design and implementation of a 2D beam scanning system. The design uses an off-the-shelf spherical mirror so is cheap, achromatic and introduces almost no aberrations. We first saw the design in the Yanus scan unit sold by T.I.L.L. Photonics GmbH, Gräfelfing, Germany (no longer commercially available). The chapter begins with an explanation of the optical design of the scanning system and the scan lens. We then present a model of polarisation changes introduced by galvanometric scanning systems that is validated with experimental data. These changes occur in the imaging from M1 to M2 because reflection of polarised light from a surface introduces phase changes between s and p polarised light. Finally we present some results showing the polarisation in the pupil of our microscope.

Our microscope has a focal spot whose polarisation and phase can be controlled by controlling the field in the pupil of the microscope (see chapter 4). We must relay the field pattern set by the SLM to the objective pupil through a scanning system so that we can build an image. Simple designs for scanning systems place the two mirrors, M1 and M2, near to each other, and image a plane mid-way between the two mirrors to the objective pupil [262]. With this approach the beam is not stationary on the pupil because neither mirror is imaged there. Nevertheless, it is satisfactory when used with an unstructured beam that overfills the pupil. For our system we must be sure that each hologram is imaged to the objective pupil and that they remain there as the beam is scanned. To do this there are three stages: (1) image the SLM to M1, (2) then image M1 to M2, (3) then image M2 to the objective pupil. Stage (1) is achieved by a slightly modified 4F system and stage (3) is achieved by a 4F system consisting of a scan lens and a tube lens. (1) and (3) are described in detail in chapter 8. In this chapter we begin by discussing our design of stage (2). One obvious way to perform this imaging is with a conventional 4F relay system. This approach
requires the use of costly custom optical components to reduce the field-dependent aberrations, and suffers from losses in throughput due to two extra lenses [262]. Field-dependent aberrations are problematic because the SLM is not fast enough to compensate them (the rates of scanning and SLM updating are discussed in chapter 8). Others have used custom made off-axis parabolic mirrors to image M1 to M2 [263,264].

Where possible in this chapter we have attempted to describe the scanning unit in isolation, however we have found it necessary at times to refer to other components of the microscope. For this reason the reader may find it helpful to consult parts of chapter 8 when reading this chapter. Unfortunately many concepts in that chapter depend on things explained in this chapter; one had to come first.

7.1. Optical design of scanner

Figure 7.1. Layout of the Galvo scanning system. M1 is imaged to M2 by the spherical imaging mirror, Mim. As M1 rotates, it scans the focal spot of L2 along a circle that is almost coincident with the focal surface (FS) of the spherical imaging mirror, so there are almost no field-dependent aberrations. M1 and M2 are mounted on perpendicular rotation axes so their movement allows the scanning of a 2D field.
The layout of the scanner is shown in Figure 7.1 and a photograph is shown in Figure 7.2. We now describe the beam path (shown in full in Figure 8.1). After reflection from the second hologram on the SLM, the beam passes through a pair of relay lenses L1 and L2. L1 and L2 are positive lenses that image the SLM onto M1. L2 also forms a focus of the beam at a distance $f_{im}$ away from M1 where $f_{im}$ is the focal length of the spherical imaging mirror, Mim. The L1/L2 relay system is described in chapter 8. For now we take it that M1 is conjugate to the SLM. M1 is positioned at a distance $2f_{im}$ from Mim, and the focus of L2 is very close to the focal sphere of the imaging mirror, denoted FS, so that the beam is collimated by the imaging mirror. M1 is imaged to M2 because M1 and M2 are both $2f_{im}$ away from the imaging mirror. Therefore the SLM is also conjugated onto M2. The scan lens and tube lens (not shown in Figure 7.1) form a conventional 4F relay system to image M2 onto the objective pupil plane. Therefore, this system affords a beam which is stationary in the objective pupil, with the SLM conjugated there. An important property of this design is that if the mirror separation, $\delta S$, and the beam diameter are very small with respect to $f_{im}$, the aberrations introduced are the same irrespective of the scan mirror angle because the mirrors are approximately at the centre of curvature of the spherical mirror, so the system is spherically symmetric.
The first design choice is to decide whether M1 should scan the beam in the same plane as the separation of M1 and M2 (in the plane of the page in Figure 7.1, we call this case 1) or perpendicular to this plane (out of the page, case 2). As M1 is scanned, the locus of points formed by the focus of L2 is a circle of radius $f_{im}$ centred on the axis of M1. Let this circle be donated C1. In the ideal case of $\delta S = 0$, C1 overlaps with FS so the aberrations do not change with scan position. In our system the separation between the beam on M1 and M2 is set to be 15 mm, which is as small as it could be, and $f_{im}$ is 150 mm, so C1 and FS do not quite overlap. In case 1, there is a lateral shift between C1 and FS whereas in case 2 C1 and FS have different radii and are tangent. Therefore the aberrations in case 1 depend on $\delta S$ but on $\delta S^2$ in case 2. For the design parameters we chose (rationalised below), we find that the RMS wavefront aberration is $0.013\lambda$ in case 1 and $0.007\lambda$ in case 2. Therefore we choose case 2 as our configuration (drawn in Figure 7.3). Note that we haven’t drawn L2 or the scan lens in this diagram. Figure 7.4 shows a sketch from 3D CAD of our custom-made mount and the mirror geometry. The mounting plate has two holes machined into the sides to allow secure fixing of the galvanometers so that $\delta S$ is as small as possible. The beam is positioned on the rotation axes of both mirrors, and near to the edge of M1 to minimise $\delta S$. We measured the surface figure of M1 on a phase-stepping interferometer and found it to have better than $\lambda/10$ flatness (at 633 nm) over its surface, ignoring parts within 0.5 mm of the mirror edge. The galvanometer scanners we use have relatively large mirrors (the $x$ mirror is 20.4 by 29.0 mm and the $y$ mirror is 22.9 by 34.5 mm). Ideally we would use smaller mirrors because this would enable faster scanning and also a shorter spherical imaging mirror focal length.
Figure 7.3. 3D view of scanning system showing the preferred layout where M1 scans the beam perpendicular to the direction of separation of M1 and M2. We have not shown L2 or the scan lens in these drawings, and have not shown the convergence of the beam in the 3D view. In both drawings we have shown the beam for various different positions of M1.
Now that we have discussed the geometrical layout of the scanner we discuss the choice of focal lengths $f_s$ and $f_{im}$. First we find the appropriate ranges of these two parameters. Consider $f_{im}$: the largest spherical imaging mirrors sold by Thorlabs have a 75 mm diameter. These mirrors are available in focal lengths between 60 mm and 500 mm. Next consider the scan lens: the scan lens and tube lens form a telecentric 4F system, so the chief ray is parallel to the optic axis in between the scan lens and the tube lens. Therefore the scan lens must have a radius bigger than the standard image radius, which is 10 mm. We chose our scan lens from the $\phi = 31.5$ mm Qioptic achromatic doublet range. The upper limit on the focal length of the scan lens is the scan pixel size. We wish to sample the PSF in fine detail and would like images with more than 20 scan pixels across the PSF in each dimension. Our scanner is controlled with a digital to analogue converter and has a minimum step size of $\delta \theta = 6.2 \mu \text{rad}$ (we measured this by measuring a beam deflection angle of 0.25 rad over 20100 scan steps). In the image plane the PSF has size approximately $\lambda M/\text{NA}_o$, where $M$ is the objective to tube lens magnification and $\text{NA}_o$ is the objective numerical aperture. The scan pixel size in the image plane is $2f_s \delta \theta$ so with $\text{NA}_o = 1.4$, $\lambda = 532$ nm and $M = 100$ a scan lens focal length of less than 170 mm will give images with more
than 20 pixels across the PSF. From the off-the-shelf list, this restricts our scan lens focal lengths to be between 60 mm and 160 mm.

Now we present our Zemax model of imaging from M1 to M2. The Zemax model is controlled by MATLAB, so that we can vary \( f_{im} \) and \( f_s \) independently and over many values and keep all other parameters consistent with the rest of the microscope. For example changing \( f_{im} \) causes the angle subtended by the scan mirrors at the scan mirror and the maximum scan angle to be changed and changing \( f_s \) changes the beam diameter at M1 and M2. Our MATLAB model accounts for these effects and constructs an appropriate Zemax layout for each pair of \( f_s \) and \( f_{im} \). MATLAB then calls Zemax’s optimise function and reads the value of the merit function. The merit function is the RMS wavefront error, averaged across the mirror scan angles. We show the results of the model in Figure 7.5. The main thing to notice is that many choices of \( f_{im} \) and \( f_s \) will yield diffraction limited performance. As \( f_{im} \) increases the aberrations improve because the separation of M1 and M2 subtends a progressively smaller angle at the imaging mirror, so FS and C1 overlap more closely. As \( f_s \) increases the aberrations worsen because the beam diameter increases in proportion to \( f_s \). The black line on the plot shows the diffraction limit (RMS wavefront aberration of \( \lambda/14 \)). Our chosen parameters are indicated by the position of the black cross, and are \( f_{im} = 150 \) mm and \( f_s = 80 \) mm. For these parameters Zemax predicts an RMS wavefront error of 0.007\( \lambda \) and beam diameter at M1 of 2.2 mm and maximum scanned beam angle of \( \pm 6.9^\circ \) (limited by imaging mirror size) giving an image radius of 9.9 mm. The maximum scan range permitted by the galvanometers is \( \pm 23^\circ \) (beam angle) so for all spherical imaging mirrors considered here it is the mirror, not the galvanometer, that restricts the maximum image size. For this reason we do not choose an even larger \( f_{im} \). We could choose a smaller \( f_{im} \) but this would worsen polarisation effects (see section 7.2). The choice of scan lens focal length is influenced by polarisation effects but also by the Zemax model of the scan lens, presented below.
Figure 7.5. Results of Zemax model of scanning system for imaging M1 to M2. The figure of merit is the RMS wavefront error. The black diagonal line shows the diffraction limit, so almost all available values of $f_{im}$ and $f_{s}$ give diffraction-limited performance. The cross shows the values we use in our design with an RMS wavefront error of 0.007\(\lambda\).

7.1.1. Scan lens aberrations

![Scan lens aberration graph](image)

Figure 7.6. Results of Zemax model of scan lens. The design is optimised for best performance in the region up to 4 mm image height. The diffraction limit is shown as a black dashed line.

We built a model of the scan lens, an 80 mm focal length achromatic doublet from Qioptic. Zemax was configured to show the aberrations for image heights at 0, 2, 3, 4 mm and at 9.5 mm and for three wavelengths, but only to optimise the focus up to 4 mm. The RMS wavefront aberration against image height is
shown in Figure 7.6. At these first four positions we can expect diffraction-limited performance. We had hoped for a 10 mm image height, but at an image height of 4 mm (so image diameter 8 mm) we can expect 200 by 200 resolution elements, and 8000 by 8000 pixels (choosing $f_e = 80$ mm gives roughly 40 by 40 pixels across a PSF). Astigmatism dominates the monochromatic point aberrations, causing a focal shift of 1.5 mm at image height 9.5 mm (DOF = 1.5 mm). There is roughly 2.5% distortion (relative to F-θ) at 9.5 mm, which is roughly twice the Airy disc diameter. If we restrict our image field to 4 mm this reduces to 0.4%, and further restricting to 1 mm distortion reduces to 0.03%. Over the wavelength range modelled we see a chromatic focal shift of 100 μm. The lateral chromatic aberration is more significant, with a 10 μm lateral shift for $\lambda = 633$ nm at 9.5 mm image height (the Airy disc diameter is 40 μm). We also modelled the 100 mm Qioptic achromatic doublet and found it to be an equally good choice. The model also predicts the aberrations to be almost independent of the orientation of the doublet. We place the crown element towards the scan mirror.

7.2. Modelling of polarisation changes during scanning

Our method of orientation determination relies on imprinting polarisation patterns onto the microscope pupil. After the second reflection from the SLM, the beam reflects from two pick off mirrors (at near normal incidence) and a dichroic at 45° incidence before reaching the scanning system. After the scanning system the beam reflects from two 45° mirrors: one to direct it into the side port of the microscope and another to turn the beam vertically into the objective. All of these components introduce phase differences between s and p polarised light that can be corrected by the SLM because these components are static. A greater problem is presented by reflections from M1 and M2 in the scanning system as they move faster than we can update our SLM (we will see in the next chapter that we can update the SLM once per scan line). In this section we model the polarisation distortions introduced by the scanning mirrors. The model applies to any 2D mirror-based scanning system.
Figure 7.7. Geometry of a plane wave reflecting from a scanning mirror. The ray travels along \( \mathbf{r}_0 \) and is incident on the mirror in the yz plane at an angle \( \alpha \) to the z axis. The mirror rotates about the y axis and has normal \( \mathbf{n} \). The angle between \( \mathbf{n} \) and \( \mathbf{z} \) is \( \beta \). After the ray reflects from the mirror it travels along \( \mathbf{r}_1 \). The polarisation before the mirror is in the plane perpendicular to \( \mathbf{r}_0 \) and is denoted \( \mathbf{E}_0 \) (shown linearly polarised in this figure). After reflection the polarisation state is \( \mathbf{E}_1 \).

First, consider the general case of polarised light reflecting from a single scan mirror that can rotate about a fixed axis. Figure 7.7 illustrates the coordinates of the problem. A ray travels along the unit vector \( \mathbf{r}_0 \) and is incident on the mirror in the yz plane at an angle \( \alpha \) to the z axis. The mirror rotates about the y axis and has normal \( \mathbf{n} \). The angle between \( \mathbf{n} \) and \( \mathbf{z} \) is \( \beta \). After the ray reflects from the mirror it travels along \( \mathbf{r}_1 \). The polarisation before the mirror is in the plane perpendicular to \( \mathbf{r}_0 \) and is denoted \( \mathbf{E}_0 \) (shown linearly polarised in this figure). After reflection the polarisation state is \( \mathbf{E}_1 \). To find \( \mathbf{E}_1 \) we project \( \mathbf{E}_0 \) onto the s and p vectors \( \mathbf{s}_0 \) and \( \mathbf{p}_0 \):

\[
\mathbf{E}_0 = (\mathbf{E}_0 \cdot \mathbf{s}_0)\mathbf{s}_0 + (\mathbf{E}_0 \cdot \mathbf{p}_0)\mathbf{p}_0,
\]

where \( \mathbf{s}_0 \) is a unit vector parallel to \( \mathbf{n} \times \mathbf{r}_0 \) and is perpendicular to the plane of incidence, and \( \mathbf{p}_0 \) is a unit vector parallel to \( \mathbf{s}_0 \times \mathbf{r}_0 \) and is in the plane of incidence. Then we calculate the Fresnel reflection coefficients for the s and p polarised light, given by [12]

\[
A_s = \frac{(\cos \theta_0 - n_1 \cos \theta_1)}{(\cos \theta_0 + n_1 \cos \theta_1)}
\]

(97)
and

\[ A_p = \frac{(n_1 \cos \theta_0 - \cos \theta_1)}{(n_1 \cos \theta_0 + \cos \theta_1)}. \]

Here \( \cos \theta_0 = \mathbf{r}_0 \cdot \mathbf{n} \), \( \theta_1 \) is the angle of the transmitted ray (inside the mirror), given by Snell’s law and \( n_1 \) is the refractive index of the mirror material. We can then write the field after reflection as

\[ E_1 = A_s (E_0 \cdot s_0) s_1 + A_p (E_0 \cdot p_0) p_1. \quad (98) \]

Before we show results of the simulation of the 2D scanning system, we show the effects of the Fresnel reflection coefficients for light of wavelength 532 nm reflecting from a silver mirror (our galvanometers are coated with an optically thick layer of silver). Silver has \( n_1 = 0.05 - 3.42i \) at 532 nm [265]. We show the reflectivity, \( R \), of s and p polarised components and the phase change between s and p polarised light in Figure 7.8. \( R \) is defined as \( R = |A|^2 \) where \( A \) is one of the two amplitude coefficients defined in (97). Figure 7.9 shows the output polarisation ellipses.

![Figure 7.8](image)

**Figure 7.8.** Fresnel coefficient data for light of wavelength \( \lambda = 532 \) nm reflecting a silver mirror. On the left we show the relative intensities of reflection for s- (in blue) and p- (in red) polarised light. The black dashed line shows the minimum in reflected intensity for p polarised light (which would be Brewster’s angle if the complex part of the refractive index was small). On the right we show the phase difference between s and p polarised light.
Figure 7.9. Visualisation of the output polarisation states when linearly polarised light at 45° to s and p is incident on a silver mirror. All these states have the same handedness.

Now we can proceed to present our model of a general 2D galvanometer scanning system. The model calculates the polarisation state across the image plane (which is the focal plane of the scan lens) given a fixed input polarisation. The model has three components. First, a beam is incident on a first scanning mirror, M1. This means that $a$ is fixed at M1 and $b$ varies as M1 is scanned. Next, the beam is incident on a second scanning mirror M2. We rotate the field state by 90° in between M1 and M2 because our model has a fixed axis of mirror rotation along $y$. Since the purpose of the scanner is to scan in 2D the axes of M1 and M2 must be perpendicular, so the plane in which the ray incident to M2 varies is perpendicular to the plane of rotation of the mirror normal. The geometry of this situation is shown in Figure 7.10. We ignore any polarisation changes that may be introduced between M1 and M2. Importantly this does not remove the validity of the model when there is an imaging step between M1 and M2 as in our scanning system design because here the beam is always almost exactly normal to the imaging mirror. The third and final step of the model accounts for the rotation of the rays and polarisation states through the scan lens to the microscope image plane. We rotate the rays and polarisation states using the calculations presented in section 6.3 for propagation of rays from the focal space of a high NA lens to the pupil plane. We assume the scan lens is positioned so that the central output ray passes through the axis of symmetry of the scan lens. Using a high NA model ensures that we can predict the electric field states in the focal plane for large scan angles.
We present the model results first for an idealised and unrealistic special case. Let $a_i$ and $b_i$ be the $a$ and $b$ angles shown in Figure 7.7 for mirror $i$. Imagine that $a_1 = 0$, and that $b_i$ varies by $\pm \theta$ ($\theta$ is the maximum scan angle of the mirror). At M2, $a_2$ varies by $\pm 2\theta$ as M1 moves. Let $b_2$ also vary by $\pm \theta$. This situation is unrealistic because the output beam is not separated from the input beam. The results are shown in Figure 7.11 for a vertically polarised input beam and for $\theta = 5^\circ$. The azimuth and ellipticity are both changed as the mirrors move. The polarisation azimuth is rotated whenever the input ray is not normal to the scan axis. Figure 7.12 illustrates this effect. The red input ray is vertically polarised, and the blue rays show the output rays for a range of mirror positions. The polarisation azimuth direction after the mirror is indicated by the tick marks on each ray. The view on the right clearly shows the rotation of polarisation. Not all of this rotation is undone as light is refracted through the scan lens. In fact, we can show that the azimuth rotation (ignoring Fresnel coefficient effects) after reflection from a single mirror is given by

$$\gamma = \arctan\left(\frac{\sin a \sin 2b}{\cos a + \cos 2b}\right) \approx ab,$$

where the approximation holds for small $a$ and $b$. The next terms in the Taylor expansion of $\gamma(a, b)$ are of fourth order in $a$ and $b$. For the example shown, we
expect about \( \pm 1^\circ \) of azimuth rotation for our scan range of \( \pm 5^\circ \). There is also ellipticity introduced due to the phase differences between s and p polarised components. We define the ellipticity as the ratio of minor to major axes of the polarisation ellipse. As the beam is scanned the angle of incidence changes at the first and second scan mirror, and the direction of the s and p vectors change at the second scan mirror.

![Figure 7.11. Polarisation states at the image plane (focal plane of scan lens) of an idealised scanner. The input beam is linearly polarised vertically (in lab coordinates). The azimuth rotation is mostly caused by geometrical factors (explained in the text), whereas the introduced ellipticity is mostly caused by the phase differences introduced between s and p polarised components.](image)
Figure 7.12. Visualisation of polarisation azimuth rotation during scanning. These two pictures show the same process each from a different viewpoint. The geometry is the same as in Figure 7.7. An incident ray, in red, is polarised in the yz plane. The red tick mark indicates the polarisation direction. The scan mirror (not shown) rotates about y. The scanner normal positions are shown as black dashed lines. There is a different output ray direction (shown blue) for each position. The output polarisation states are determined according to equation (98), though we ignore phase differences between s and p components for this illustration. It is evident that the polarisation azimuth is rotated in after reflection from the mirror.

Now we consider a more realistic situation where the input and output beams are offset, as in our design shown in Figure 7.3. A plan view of the layout is shown in Figure 7.13. The value of a at the first scan mirror is 5.5° and is limited by the lateral extent of L2. The value of the offset on b₂ is 12° and is limited by the lateral extent of the scan lens. This gives a total deflection between input and output beams of 35°. The results of the model are shown in Figure 7.14 (for a vertically polarised input state) and Figure 7.15 (for a 45° polarised beam). The form of the introduced ellipticity for the vertically polarised state is the same as that shown in Figure 7.11 but we have shifted away from the origin due to the offset in b₂. The azimuth plot has not shifted because the scan lens remains centred on the central ray of the output bundle. For both input states the introduced ellipticity is now up to 2% because the angles of incidence are greater than in the ideal case. This gives an important general design principle for a scanning system: to minimise polarisation distortions, minimise the angle between input and output beams. In our design we could significantly reduce a₁ by using a smaller diameter L2 as the beam diameter is 3 mm here (our current choice of L2 has a 25 mm diameter) but it would be hard to reduce the output beam offset (b₂ offset) as the scan lens needs to be large.
Figure 7.13. Top view of scanning layout showing offset input beam ($\alpha_1 = 5.5^\circ$) to allow space for L2 and offset output beam ($b_2$ offset = $11.7^\circ$) to allow space for the scan lens. These two offsets cause the polarisation effects during scanning to be more prominent.

Figure 7.14. Polarisation states at the image plane of our scanner. The input beam is linearly polarised vertically. The azimuth is rotated by roughly $\pm 0.5^\circ$, and there is a maximum of about 2% ellipticity. The second scan mirror is offset by $12^\circ$ to separate the input and output beams.
The calculations above are based on reflection from bulk silver. To check the validity of this assumption we measured the actual polarisation changes on reflection from one of the scan mirrors using linearly polarised light at 45° to the plane of incidence and measuring the output polarisation state by nulling (observing by eye) the transmitted intensity through a quarter waveplate (QWP) and analyser. When the fast axis of the QWP lines up with one of the axes of the ellipse in the output state then the output from the QWP is linearly polarised. The azimuth of this linear polarisation state indicates the ellipticity of the light after reflection from the mirror. The calculated azimuths and ellipticities are plotted in Figure 7.16 together with the predicted values based on the Fresnel reflection coefficients from a silver mirror. There is good agreement between theory and practice for low incidence angles, indicating that our predictions of ellipticities introduced during scanning will be accurate. For incidence angles greater than 40° the agreement between theory and practice is poor. Perhaps this is due to an unknown coating on the mirror. If this coating was largely optically transparent (i.e. no complex part of refractive index) then it would not have a significant effect at low incidence angles, but would become more significant at large incidence angles because there would be a significant fraction of incident energy reflected from the air-coating interface. We could have used an ellipsometer to determine the thickness and refractive index of a coating.
Figure 7.16. Experimentally measured and predicted ellipsometry data for a beam polarised at 45° to the plane of incidence. The experimental data is shown as black bars with ±1σ error bars. The blue lines are predicted from the Fresnel coefficient calculations assuming a silver mirror. There is good agreement between theory and practice for low incidence angles.

7.3. Experimental results of pupil polarisations

In chapter 4 we showed a method that gives control of the polarisation at the output of an SLM. The goal is to use this to control the polarisation in the objective pupil. As we will see in the next chapter, in between the SLM and the objective there are four mirrors, a dichroic and a scanning system. Of these components, the scanning system introduces dynamic polarisation changes, modelled above, and the other components introduce static polarisation changes. We measured these changes as follows. In section 4.5 we explained how to measure the phase offset between the two holograms on the SLM so that we can produce the required output states. At the output of the holograms the phase offset between A and B was reported to be −0.65π. To recap briefly, we set both holograms to simple gratings and observe the common first order. When there is a zero phase difference between the beams we would expect to find a linearly polarised output at 45°. We find this linearly polarised output using an analyser. We now repeat the process of finding the phase offset but
with the analyser placed in the objective pupil. The objective is removed and we observe the transmitted light by eye on a screen. For the measurement at the pupil we measure the phase offset to be $-0.02\pi$. Therefore the accumulated phase between s and p polarised light from SLM to pupil is $0.63\pi + 2m\pi$, where $m$ is an unknown integer. It is valid to repeat this phase adjustment process in the pupil because the outputs of holograms A and B remain linearly polarised at the pupil (to within measurement error). This is because, in lab coordinates, hologram A controls the polarisation in the plane of the optical table and hologram B controls the polarisation perpendicular to the table. There are four static mirrors and a dichroic between the output of the SLM and the objective pupil. The first three static mirrors and the dichroic deflect the beam in the plane of the optical table so their s and p vectors are vertical and horizontal respectively. The fourth mirror directs the beam vertically into the microscope objective, so the s and p vectors are horizontal and vertical. Therefore at every static reflection the light from hologram A remains linearly polarised and so does the light from hologram B; only the phase differences between these linearly polarised components changes.

![Figure 7.17](image)

Figure 7.17. Pupil intensity as a function of analyser angle for two different linearly polarised pupils: polarised at $0^\circ$ (top), and at $135^\circ$ (bottom). The numbers indicate the analyser position in degrees.
Figure 7.18. Pupil intensity as a function of analyser angle for four different pupil polarisations. From top to bottom: linear, 90°; circular; circular with helical phase; radial. The numbers indicate the analyser position in degrees.

We imaged a fluorescent slide placed in the objective pupil to show that we can control the pupil polarisation. To do this we placed a CCD camera (GuppyPro, AlliedVision) with an imaging system looking directly down on the objective pupil with no objective in place. We placed a fluorescent slide at the pupil plane. The results are shown in Figure 7.17 and Figure 7.18. Both figures show the recorded pupil intensity as an analyser is rotated through the beam. Figure 7.17 shows the generation of a 0° and a 135° linearly polarised pupil. Figure 7.18 shows, from top to bottom, a 90° polarised pupil, a circularly polarised pupil, a circularly polarised pupil with helical phase (note the dark spot in the centre of the pupil) and a radially polarised pupil. From these images we measure the degree of polarisation of the linearly polarised beams to be better than 100:1. The measurement is limited by noise on the CCD. The illuminated patches shown are limited by the extent of the holograms on the SLM. The intensity falls to less than 10% of the peak intensity outside a radius of 3.5 mm. The pupil radius is 2.8 mm. The intensity at the pupil edge is 63% of the peak intensity. Fitting a Gaussian intensity profile to this, we find that the 1/e beam diameter is 1.3 times the pupil diameter.

7.4. Chapter summary

We have introduced the beam scanning system for our confocal microscope. The scanning unit scans the laser beam in 2D while keeping the beam stationary in
the objective pupil. The first scan mirror is imaged onto the second by a spherical mirror. Both scan mirrors are situated near the centre of curvature of this mirror which means that there are effectively no aberrations introduced by the scanning system. Using a Zemax model we predict an RMS wavefront error of just 0.007\(\lambda\) (averaged over all scanning positions). After the scanner the beam is brought to focus by a scan lens. We demonstrate that using an 80 mm focal length scan lens can give diffraction-limited performance out to an image height of at least 4 mm. While we had hoped to be able to image to an image height of 10 mm, we expect that we will be able to gather images with 200×200 resolution elements and 8000×8000 pixels (limited by minimum scan step size) with this image size. We presented a detailed model of the polarisation changes introduced during scanning. We find that a 2D beam scanning system introduces changes in the polarisation state of the input beam. These changes have been modelled before [266], though the authors do not consider imaging optics between M1 and M2 and do not present their results clearly in terms of the predicted azimuth and ellipticity across the image plane. We find that the azimuth angle of polarisation is rotated by approximately \(ab\) at each mirror (\(a\) and \(b\) are defined in Figure 7.7), which amounts to 0.5˚ in our case. As well as the geometric rotation of polarisation, the scan mirrors introduce phase differences between s and p polarised light. These changes are complicated because as the mirrors move, not only do the angles of incidence change, but the directions of the s and p vectors rotate. We predict that a linearly polarised input state will accrue phase differences resulting in 2% ellipticity. For this model we assume the scan mirrors are made of silver. We measure the phase changes on reflection from one of the scan mirrors and find that for angles of incidence up to about 20˚ the polarisation changes introduced by the mirror are as if the mirror were made of silver. At larger angles of incidence there is a departure in the accuracy of the theory. Finally we showed that we can produce a variety of polarisation states in the pupil of the objective. To do this we correct static phase differences between s and p polarised components that are introduced by mirrors and a dichroic between the SLM and the objective pupil.
8 Microscope assembly

In previous chapters we have described some key components of our microscope in isolation. This chapter completes the description of the individual components of the optical system and explains how the components fit together. We also explain the electronics and software used to control the microscope.

We begin by describing the whole optical setup, shown in Figure 8.1. The source is a 532 nm diode-pumped solid-state laser made by Shanghai Dream Lasers and outputs 40 mW of CW radiation. The beam is filtered and expanded to roughly 1 cm diameter by an objective, pinhole and achromatic doublet. The beam then passes through the polarisation control part of our microscope, described in detail in chapter 4. We then relay the second hologram, B, to the first scanning mirror, M1, reflecting from a dichroic on its way. This relay system is described in section 8.1 of this chapter. The scanning system images M1 to M2 and is described in chapter 7. After passing through the scanning system the beam is focused to the primary microscope image plane by the scan lens. The image plane is located just outside the side port of an inverted microscope frame (Nikon Eclipse TE300). We use the tube lens and objectives provided with the microscope frame. In between the tube lens and objective the beam is directed vertically by a 45° mirror (not shown in the figure). The tube lens and objective lens relay the focused beam from the scan lens to the sample plane. M2 is imaged to the objective pupil by the scan lens and tube lens, placed in a 4F arrangement as in a normal confocal microscope (as described in section 2.4). We use an 100X 1.4NA Oil immersion objective (Nikon PlanApo). The sample is mounted on a piezoelectric z-stage (Prior NanoScanZ). Light from the fluorescent sample returns through the microscope, is descanned by the scanning system, and passes through the dichroic to be detected by a photomultiplier tube.
In section 8.2 we describe the light detection path in our microscope, consisting of a fibre-coupled photon-counting photomultiplier tube (PMT). With this, we have largely completed the description of the optical setup. We then describe the electronics and software that allows images to be taken in section 8.3. All software is written in MATLAB and C/C++. There is a somewhat complex system designed to synchronise the scanning of image lines with the updating of the SLM frames. Our software also allows us to control the SLM, scanning unit, z stage and PMT with the PC.

Figure 8.1. Layout of the microscope. An expanded, filtered CW laser passes through the polarisation control scheme where it encounters two holograms, A and B. B is imaged to the first scan mirror M1 by a relay system consisting of two positive lenses. There is a 90° reflection from a dichroic between these two lenses. The beam is scanned in 2D (x y in the sample) by the scanner, which is designed to keep the beam stationary in the objective pupil. Return fluorescence is descanned and collected by a PMT. The SLM, scanners, microscope z stage and PMT are all controlled and synchronized using a PC and custom software written in MATLAB and C/C++.
8.1. SLM-to-scan relay

Figure 8.2. Imaging from second hologram on SLM (in green, left) to first scan mirror (in green, right) while ensuring a focusing beam at a distance \( d \) from the first scan mirror. The dichroic is situated about 1 cm to the left of where the marginal ray crosses the optic axis after L1 but is not shown in this diagram. The pinhole is situated at where the marginal ray crosses the optic axis after L1 and is also not shown.

The relay system for imaging hologram B to the M1 is shown in Figure 8.2. This is a modified 4F system where the separation of the two lenses, L1 and L2, is greater than the sum of their focal lengths, \( f_1 \) and \( f_2 \), by a distance \( D \) so that the beam converges after the second lens. The image of B is at the focal plane of L2, as shown by the dashed chief ray. The magnification of the system from B to M1 is given by \( \frac{h_3}{h_0} = -\frac{f_2}{f_1} \). This is shown by the constructed dotted ray that is parallel to the marginal ray between L1 and L2 so must meet the marginal ray at the back focal plane of L2. We construct this ray to pass through the front focal plane of L2 so that it is parallel to the optic axis after L2, and therefore it defines the height of the image of B and the magnification of the system.

The magnification from B to M1 must be so that the image of B fills the objective pupil. The radius of the image of the pupil on M1 is \( h_3 \) and is given by

\[
h_3 = \frac{f_s}{f_{T}} h_p = f_s \frac{\text{NA}_0}{M} = 1.12 \text{ mm},
\]

where \( f_s = 80 \text{ mm} \) and \( f_{T} = 100 \text{ mm} \) are the scan lens and tube lens focal lengths, \( M = 100 \) is the objective-tube lens magnification and \( \text{NA}_0 = 1.4 \) is the objective NA. Suitable holograms have a diameter of 400-600 pixels.
corresponding to \( h_1 = 3-4 \text{ mm} \). Since we would prefer to overfill the pupil, we seek a magnification of \(-\frac{1}{3}\) for the L1-L2 system. Selecting the lenses from the Qioptic range of achromatic doublets, we pick \( f_1 = 600 \text{ mm} \) to give enough space to pick off the beam after the SLM (there are two pick-off mirrors between the SLM and L1). This sets \( f_2 = 200 \text{ mm} \). We then find \( D \) by applying Newton’s conjugate distance equation at L2, giving
\[
D = \frac{f_2^2}{f_{im}} = 267 \text{ mm}, \tag{101}
\]
because \( d \) must equal \( f_{im} \). Using off-the shelf achromatic doublets from Qioptic, we simulated the system in Zemax and found that the RMS image plane wavefront aberration is 0.01\( \lambda \) and the RMS pupil plane wavefront aberration is 0.002\( \lambda \) for a hologram radius of 4 mm. These figures are reported for the lens configuration where the negative element of the first doublet faces the SLM and the negative element of the second doublet faces the scan system.

To conclude our description of the relay system, we briefly discuss the dichroic. We use a Semrock brightline FF562 filter. Photons with wavelengths shorter than 556 nm are reflected with greater than 90\% efficiency, and those with wavelengths longer than 568 nm are transmitted with more than 90\% efficiency. We measured the flatness of this dichroic in transmission using a phase stepping interferometer. The reconstructed wavefront results are shown in Figure 8.3. The dichroic introduces roughly 0.04\( \lambda \) peak-valley wavefront aberration but only about 0.004\( \lambda \) across the central 8 mm diameter disc. Because the dichroic is very flat we have a degree of freedom in deciding where to place it. We decide to position it near to the focus of L1.

At the focus of L1 is the Fourier plane of the SLM, which contains a series of spots along a line perpendicular to the fringe tilt on the holograms. A typical hologram has an 8 mm diameter and 50 fringes across the hologram. This means that the separation of zero and first order at the focal plane of L1 is 2.0 mm. The width of the first order (for a simple grating displayed on the SLM) to the first zero of the diffraction spot is 96 \( \mu \text{m} \) and the Strehl-limited depth of focus is 13 mm. We
select just the first order of A and B using one of three pinholes with diameters of 300, 400 and 900 µm.

![Figure 8.3. Reconstructed wavefront from phase stepping interferometer measurement of Semrock brightline FF562LP dichroic.](image)

8.2. Detector path

A schematic of the detector path is shown in Figure 8.4. Descanned light is focused to the image plane (this is conjugate to the primary image plane between tube lens and scan lens), and passes through the pinhole. This is the same pinhole that selects the first diffracted order from the SLM. The spot diameter at the image plane is 110 µm. The beam is imaged to the front face of a multimode fibre by a 10× objective with NA= 0.3. The fibre is a 1 m patch cable supplied by Thorlabs (M67L01). The beam diameter at the fibre is 5.4 µm. The fibre has a core diameter of 25 µm (i.e. 4.63 Airy disc diameters). The fibre core therefore forms the confocal pinhole (the pinhole selecting the first order of the holograms has a diameter of 900 µm, approximately 9 Airy disc diameters). The beam diameter just before the objective is 4.5 mm, which is smaller than the pupil diameter (9.6 mm). The NA of the fibre is 0.1. The NA of the beam at the fibre is 0.135, which is slightly too large so some photons are wasted. We should improve this in future work. We clad the fibre in aluminium foil to remove collection of photons through the fibre jacket. We estimate the modal dispersion in the fibre to cause pulse broadening of up to 30 ps, which is much shorter than the scan pixel time (1 to 30 µs) so this multimode fibre will not redistribute light.
between pixels. The fibre directs photons to the photocathode of a photon-counting PMT (Hamamatsu H7422P-40). The active area of the PMT has a 5 mm diameter and the beam has a diameter of 2.8 mm at the photocathode. The current pulses from the PMT are fed into an amplifier-discriminator (product AD-8 manufactured by Electron Tubes, UK) that outputs 17 ns 3.3 V square pulses. This unit has 50 Ω input impedance and a pre-set discriminator level of −2 mV. These pulses are counted by a data acquisition (DAQ) box (National Instruments 6351-USB). The amplification voltage of the PMT is adjusted with a control voltage of $V_C = 0$ to 0.8 V. We measure the signal/noise by measuring the number of counts in 100 ms time windows over a time of 10 s. We compute the ratio of the average signal to the standard deviation signal as a function of $V_C$. We start to see a dip in signal/noise above $V_C = 0.75$ V, so operate the PMT at $V_C = 0.71$ V. The gain at this voltage is $5 \times 10^5$ (from the PMT datasheet). We measure the dark counts to be 20-70 counts/sec with cooling turned on, and 9000 counts/sec with cooling off. We correct our signals for pulse pile up using $C' = C / (1 - \beta C)$ with $\beta = 25$ ns (taken from AD-8 data sheet).

Figure 8.4. Schematic of the detection path. We image the descanned spot at the image plane to the front face of a multimode fibre. The photons are then turned into current pulses by the PMT. The amplifier/discriminator outputs a 3.3 V pulse of 17 ns duration for each pulse above a pre-set level. These pulses are counted by the DAQ box.
8.3. Synchronisation and timing

There are five components in our microscope that are controlled by the PC: the SLM, the scan programming unit (SPU), the DAQ box, the z-stage and the PMT. To record an image we must synchronise these units. Before explaining the interaction of these components we explain the timing issues and methods of each one individually.

8.3.1. SLM timing

The SLM control board is connected to the PC via DVI connection and appears as an external monitor. We display holograms on the SLM by drawing windows on this external screen using the Psychtoolbox kit with MATLAB. We can display 24-bit RGB images on the monitor. The SLM itself is binary, so when displaying a 24-bit RGB image the control board separates the bit planes and displays them in quick succession on the SLM. Therefore by displaying a 24-bit RGB video at 85 Hz the SLM can display 2040 binary holograms (bit planes) per second. For our application we use just three bit planes per frame with a frame rate of 60 Hz, giving 180 bit planes per second. Figure 8.5 shows the operation of the SLM in the time domain with three bit planes per frame. If we request the pixel value \([R, G, B]\) where \(R, G,\) and \(B\) are either 0 or 1, the SLM displays each bit plane for a time \(t_{on}\). Here \(R, G,\) and \(B\) do not correspond to colours, but just to different binary bit planes. Between each of the three bit planes the SLM displays the complementary pixel value to achieve DC balancing. These DC-balanced planes are shown as \([\bar{R}, \bar{G}, \bar{B}]\) in Figure 8.5. Even though there are six bit planes displayed in the time of one frame we do not have a choice over what to display in the DC-balanced frames. In between each bit plane the pixel switches value over a time of \(t_s\). The SLM outputs two trigger signals indicating the times when a new frame starts (frame sync) and when a new bit plane is displayed (bit plane sync). We measure \(t_{on} = 2650\ \mu s\) and \(t_{off} = 2730\ \mu s\) using an oscilloscope. From this we can deduce that \(t_s = \frac{1}{2} (t_{off} - t_{on}) = 40\ \mu s\). In section 8.3.3 we explain
how this signal starts the scanners and pixel clock. Figure 8.6 shows an example RGB hologram that SLM displays as three binary holograms in quick succession.

Figure 8.5. Operation of FLCSLM in the time domain. The top chart shows a sequence of frames, which are sent from the PC at 60 Hz. The SLM is connected to the graphics card by DVI connection and is seen as an external monitor. Within each frame, the SLM displays a number of bit planes. We configure the SLM to show 3 bit planes per frame with duration \( t_{on} = 2650 \mu s \). The SLM displays each bit plane in sequence, interleaved with a series of inverted frames to DC-balance the pixel. Between each bit plane the SLM takes \( t_s = 40 \mu s \) to switch. There are pins on the SLM drive board that indicate the start of each frame and when each plane is in the stable ‘on’ position. We use these signals to synchronise the SLM and scanning unit.
Figure 8.6. Method of displaying three holograms in succession on the SLM: compute desired holograms for a given frame, assign each one to the R, G and B channel of an image, then show this at the appropriate location on the SLM. R, G and B do not correspond to colours, they just refer to the three binary bit planes that are displayed during each frame.

8.3.2. Scanner timing and control

The scanning system consists of an SC2000 scan programming unit (SPU), two MiniSAX scan drivers and two M-series M2 galvanometer scanners (one for each axis X and Y). All components are provided by Cambridge Technology, GSI group GmbH. The scan drivers take an analogue voltage between $\pm 3$ V and instruct the mirror to rotate to a position based on this voltage. The drivers stabilise the mirrors by reading the position feedback signals from the galvanometers. The scanners can be controlled directly by feeding analogue voltages to the MiniSAX controllers. The SPU has on-board memory and a serial interface to a PC. Using this interface we can load programs written in machine code onto the SPU to carry out predefined tasks. The SPU converts the requested programs into analogue voltages that are sent to the MiniSAX drivers. The machine code instruction set for the SPU is given in its manual. To simplify communication with the SPU we wrote a MATLAB program to read the SC2000 manual (converted to plain text format) and automatically write a corresponding MATLAB function for each SPU machine code instruction. This code only needs to be run once. The generated functions are methods of a class called the SC2000communicator. We show an example method in Figure 8.7. This function moves the beam (by rotating both mirrors) to a new position given by (xABPOS,
yABSPoS) over a time of COUNT. If we want to move to a new position (1000, 2000) in a time of 500 tic (one tic is 23 µs), we would send 6 1000 2000 500 to the SPU via serial connection. Our automatically-generated MATLAB function allows us to write slewXY(1000, 2000, 500) instead.

```matlab
function [ txData, rxData ] = slewXY( self, xABSPos, yABSPos, COUNT )
    % SLEWXY
    % Number of inputs: 4
    % Input 1: self.serialObj is an open serial port
    % Input 2: xABSPos is an LEWORD
    % Input 3: yABSPos is an LEWORD
    % Input 4: COUNT is an LEWORD
    % For use in vector mode.
    % Generated automatically by functionWriter class.
    % Source dictionary is at the end of SC2000 command reference document.
    % 14 February 2014. James Clegg.

    commandBit = 6;
    rxBytes = 0;

    b1 = self.convert2leWord( xABSPos );
    b2 = self.convert2leWord( yABSPos );
    b3 = self.convert2leWord( COUNT );
    txData = [ commandBit, b1, b2, b3 ];

    if self.transmit.statusB
        serialObj = self.serialObj;
        fwrite( serialObj, txData, 'uint8' );
        rxData = [];
    else
        rxData = [];
    end
end
```

Figure 8.7. Example function to allow easy communication between MATLAB and the SPU. This function was automatically generated by a separate MATLAB class that reads the SC2000 manual.

The SPU allows the scanning to be synchronised with other components via its waitSync and setSync functions. There is a repeatable delay of $t_c = 600 \, \mu s$ between the execution of a command on the SC2000 and the appearance of the corresponding voltage on the galvanometer position control pin. This delay arises from the digital to analogue converter on the SC2000. Figure 8.8 illustrates the timing issues with the fast axis (say x) of the scanner when scanning an image of size $L_{im} \times L_{im}$. We wish to scan a line of length $X_c$ and
collect photons only when the SLM is on and the mirror is moving at constant speed. To do this we must account for the delay between digital drive signal and analogue voltage on the command pin, $t_c$ and delay the start of the light collection by $t_d$ that allows for the time taken for the scanner to reach constant speed. We instruct the scanner to move a distance $X_S$ in a time $t_F$. Over the time of this movement, the scanner moves at a constant speed for a time $t_F - t_d$, and we run a pixel clock and collect light for a time $t_p$, which is slightly shorter than $t_F - t_d$ (in theory $t_p$ could equal $t_F - t_d$). During $t_p$, the scanner moves a distance $L_{im}$.

To calibrate the scanner we imaged a reflective metal disc with concentric circles etched into the surface. Uncalibrated, we see the circles distorted along the fast scan direction. We first adjust $t_d$ until the circles appear undistorted at the leading edge of the fast scan. We then adjust $t_F$ and $X_S$ in proportion until the circles are undistorted at the trailing edge of the fast scan. Then we have set the scanner so that we are only collecting light during the linear portion of the scan, and during this time the scanner moves a distance $L_{im}$. 
Figure 8.8. Schematic of the signals on the fast scanner when scanning an image of size $L_{im} \times L_{im}$. The red and green signals are the digital and analogue drive signals. There is a delay $t_c$ between these two. The digital drive signal starts almost instantaneously after the SLM bit plane on signal is received. We start the pixel clock at a time $t_d$ after the scanner receives its analogue instruction to move and adjust the scan line time $t_F$ so that we only collect light during the linear portion of the scanner motion. We adjust $X_S$ in proportion with $t_F$ so that the distance scanned during light collection, $L_{im}$, remains fixed.

8.3.3. Synchronisation of components

We have described the operation of the SLM and scanner in the time domains. Now we describe our method of synchronising these components. We show a schematic of the interaction of the electronic components of our microscope in Figure 8.9. The red rectangles indicate a connection to PC. The yellow triangles represent trigger signals where each triangle points along the direction of signal propagation. The green rectangles indicate programs saved on a component’s internal memory.

<table>
<thead>
<tr>
<th>Times in µs:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{on} = 2650$</td>
</tr>
<tr>
<td>$t_c = 600$</td>
</tr>
<tr>
<td>$t_d = 205$</td>
</tr>
<tr>
<td>$t_F = 2095$</td>
</tr>
<tr>
<td>$t_p = 1745$</td>
</tr>
</tbody>
</table>

$X_S = \frac{t_F}{t_p} \times L_{im}$
Using our system we can acquire images in two modes: black and white (BW) or RGB. To clarify, in RGB mode there is no colour but three binary bit planes are displayed per frame. In BW mode we display simple binary holograms on the SLM so the SLM displays the same holograms for every bit plane. This allows us to collect a single image or set of images with just one excitation polarisation. In RGB mode the SLM displays RGB holograms which means that excitation polarisation is changing. We first explain how our system is synchronised in BW mode.

We must ensure that we scan lines and collect photons only during the periods when the SLM is stable. Before scanning an image, we load the SPU and DAQ with appropriate image parameters, among which are the number of pixels in the image (assumed to be square) and the timing parameters in Figure 8.8. The following signals are sent to build up an image of $n \times n$ pixels and $N$ frames in BW mode. The bulleted numbers correspond to the yellow trigger signal triangles in Figure 8.9.

1. A command is sent via USB to the DAQ box to request a new image.
2. The SLM new frame signal triggers the start image program on the DAQ box, which immediately sends a pulse to the SPU.
3. This pulse starts the scan image program on the SPU. The SPU then waits for another signal before scanning a line.
4. The SLM new bit plane signal triggers the start line program on the DAQ box, which immediately sends a pulse to the SPU.
5. This pulse starts the scan line program on the SPU. The SPU instructs the X and Y galvanometer controllers to start scanning a new line.
6. When the mirrors have reached constant speed (after a calibrated delay \( t_c + t_d \)), the SPU sends a signal back to the DAQ.
7. This signal triggers a pixel clock on the DAQ. The pixel clock outputs \( n + 1 \) pulses with period \( t_p / n \). The extra pulse allows us to discard light collected during scanner flyback.
8. The counter counts pulses from the amplifier/discriminator, using the pixel clock to shift counts into a buffer.
9. At the end of a line the SPU instructs the X and Y galvanometer controllers to move to the start of the next line and we wait for the next SLM bit plane pulse (step 4). Steps 4-8 are repeated until we have recorded a whole image with \( n \) lines.
10. We return to step 2 to record the next frame.
11. Repeat steps 2-10 \( N \) times.

At any stage during this process we can run the read buffer program to display an image on the PC. The counts are read by a MATLAB imageGrabber class, which reads and counts and displays a live image on the PC as the frame is being scanned. A schematic shown the path of the focused laser spot in the image plane is shown in Figure 8.10.

![Schematic of the image plane during scanning in BW mode](image)

**Figure 8.10.** Schematic of the image plane during scanning in BW mode. Spatial quantities are shown in blue, and temporal quantities in orange. Each line scan is controlled by the 'scan line' program, which is triggered by a start line pulse from the DAQ.

To collect an \( n \times n \) pixel image with \( N \) frames in RGB mode a number of changes are made to the sequence described above. Since the SLM is displaying three sets of holograms in quick succession in every SLM frame, we have to redistribute the light collected during these times to different images. We still scan one line per SLM bit plane, but we only step down to the next scan line once
every three bit planes. If we denote line \( m \) as \( L_m^C \) where \( C \) indicates the SLM bit plane colour during this line, then the sequence of scanned lines is \( L_1^R, L_1^G, L_1^B, L_2^R, L_2^G, L_2^B, L_3^R, L_3^G, L_3^B, \ldots, L_n^R, L_n^G, L_n^B \). Software on the PC redistributes the count values to three images, one for each hologram colour (excitation polarisation state). We can then scan subsequent frames if necessary. This way of gathering light using three excitation polarisations in an interleaved way is novel and allows us to compare intensity levels between frames because there is roughly the same level of photobleaching in each image.

We now briefly summarise the software. Firstly, there is a MATLAB class that reads the SPU manual and outputs a set of methods for another MATLAB class that allows us to send programs to the SPU. We write a number of standard programs for the SPU that are uploaded the SPU (once every time the SPU has been turned on). They are then stored on SPU memory and run unsupervised by the PC. They are started and stopped by trigger signals as described above. The code that runs on the DAQ unit is written in C++ and is compiled as a MEX function so that it can be called from MATLAB. There is a final MATLAB class (imageGrabber) that calls the MEX function, and counts the photons, and then displays the confocal images.

8.4. Chapter summary

In this chapter we have completed the description of our microscope. We began by outlining the layout of the whole optical system. We described how the polarisation-controlled beam is relayed to the first mirror of the scanning system, reflecting from a dichroic mirror on its way. We showed that this imaging system introduces no significant aberration, and that the dichroic is optically flat over the relevant beam area. The track length of the imaging system is long: the distance between the SLM and M1 is given by \( 2f_1 + 2f_2 + \frac{f_2^2}{f_{im}} = 1.9 \) m. This could, and should, be reduced in a future design. The limiting factor here is that we required a demagnification of 3X from SLM to M1, while keeping a reasonable distance between L2 and M2. After the scanner the beam is
brought to focus by the scan lens at the image plane of our Nikon TE300 eclipse microscope. This image plane is located just outside the side port of the microscope. The internal proprietary tube lens and objective relay and demagnify the focused spot into the sample. The scan lens and tube lens also form a relay system that ensures that the beam is stationary on the objective pupil. We used a laser source manufactured by Shanghai Dream Lasers in our microscope. This intensity output over time is highly variable. This should be improved in a future iteration to allow comparison of image intensities for quantitative dipole orientation studies.

Next we described the detector path in our microscope. The fluorescence from the sample is directed to a fibre-coupled photon counting PMT. The current pulses from the PMT are amplified and compared to a threshold by the amplifier\discriminator, which outputs 17 ns pulses. We expect to be able to count up to $10^6$–$10^7$ s$^{-1}$ with this system. We measure dark count rates of 20-70 s$^{-1}$. We will see in the next chapter that these parameters give us the ability to detect fluorescence from single emitters such as fluorescent molecules and nanodiamonds.

We then described the system of synchronisation in our microscope. Synchronisation is needed because the SLM displays stable, valid holograms at only certain times. We use the signals from the SLM to trigger the scanner to trace a line in an image, and a pixel clock and counter to count photons during this process. All trigger signals are routed through a programmable DAQ box, which allows us to start and stop the image acquisition process with software. With this system we achieve a line rate of 180 Hz. The line rate is limited by the SPU, which has a 600 μs delay between receiving an instruction from the PC and instructing the galvanometers to move. We would remove this digital step in a future version of the microscope and drive the scanners directly with analogue voltages from the DAQ box. In this case we would expect to be able to reach the inertial limits of the scanners and might be able to double the line rate. This would require using 6 bit planes per frame of the SLM. We described our
method of controlling the SPU by programmatically translating the SPU command reference manual into MATLAB code.

We described how the synchronisation works in BW and RGB modes. The RGB mode is especially important because it allows us to obtain interleaved multi-frame images with three excitation polarisation states. We will see the benefits of this in the following chapter.
9 Interaction of polarised light with single emitters

In this chapter we present images obtained with our microscope. We begin by showing images of subresolution fluorescent beads. These beads respond isotropically to polarised light because they contain many randomly oriented dipoles and allow us to measure the PSF of our microscope as we change the excitation polarisation. We also use the beads to characterise the static (independent of scan position) and dynamic (depending on scan position) aberrations of the system. We present our own model of focusing through an interface of two dielectric media that allows us to predict the image of beads on a glass substrate. We validate our model experimentally. We also show images of beads under polarised light as we change the numerical aperture of our system.

We then present images of single molecules, nanodiamonds and quantum dots. We chose these samples because they represent three types of single emitters, with one, two and three orthogonal dipoles in each single emitter respectively. These images test out our microscope against its original design aim: to be able to determine the orientation of fluorescent emitters. The fluorescent centre in a fluorophore behaves as a single dipole in terms of absorption and emission. We show images that are able to resolve the orientation of this dipole. The fluorescent nitrogen vacancy centre in diamond consists of two orthogonal electric dipoles. A quantum dot is spherically symmetric and so contains three orthogonal dipoles. This means one would expect it to respond isotropically in polarised light.

In most cases we present the images as sets of three, taken in RGB mode with azimuthally, radially and circularly polarised pupils. The sets are obtained using a three-channel hologram that displays the required diffraction gratings to create azimuthally, radially and circularly polarised pupils in quick succession as described in section 8.3.1. This allows us to obtain these image sets interleaved,
so that the first line scanned has azimuthal polarisation and photons from this line are assigned to image 1. The next two lines scanned have radial and circular polarisation and the light gathered during these times is assigned to images 2 and 3 respectively. Only after each set of three holograms does the scanner move down one line. Obtaining the images in this interleaved way means that we can be sure that an equal degree of photobleaching has occurred for each image, so we can compare intensities between the three images reliably. This is especially important when imaging single molecules.

We frequently use a phrase such as “azimuthally polarised image”, by which we mean the image obtained with an azimuthally polarised excitation pupil. We also give the number of frames, \( N \), accumulated in each image. The time per line is always \( t_p = 1.745 \text{ ms} \) so the total time per pixel is \( T_p = t_p N / n \) where \( n \) is the number of pixels. Many images displayed below have \( n = 251 \) and \( N = 50 \), giving \( T_p = 347 \mu \text{s} \). Since the line rate is effectively 60 Hz in RGB mode, an RGB set of images with these parameters would take 210 s to acquire.

### 9.1. Fluorescent microspheres

To characterise the microscope we began by imaging samples of fluorescent beads. We prepared samples of 100 nm diameter TetraSpeck\(^2\) beads deposited onto cover slips. The stock beads were diluted by 1:10 with Ethanol Absolute\(^3\). This dilution was dried onto a cover slip. After drying, the cover slip was glued to a microscope slide using nail varnish. In some cases we placed a thin layer of glycerol (whose refractive index is 1.47 [267]) between the cover slip and slide to match the refractive index of the cover slip and immersion oil (whose refractive index is 1.52).

Each fluorescent bead contains many randomly oriented fluorescent dipoles. We show simulations of the effect of the interaction of a fluorescent bead with an

\(^2\)TetraSpeck is a trademark of ThermoFisher scientific. Catalog number T-7279.

\(^3\)Ethanol Absolute is provided by VWR Ltd. and has 99.8% purity.
azimuthal, radial and circularly polarised beam in Figure 9.1. For this calculation we have assumed that the bead is point-like. In a separate simulation we found that the 100 nm bead size had a small effect on the PSF sizes, for example by increasing the FWHM of the image of the bead under circularly polarised illumination by 10%. The three images have been normalised to the same value. On this scale the peak intensity of the image of the bead under azimuthally polarised light is 0.36, while under radially polarised light it is 0.32 and under circularly polarised illumination it is 1. The FWHM of the bead under radially polarised illumination is 380 nm and under circularly polarised illumination it is 210 nm. The diameter of the bright ring in the image of the bead under azimuthal illumination is 283 nm.

Figure 9.1. Simulated images of a point-like bead under azimuthal excitation (left), radial excitation (centre) and circularly polarised excitation (right). The relative maximum intensities of these three spots are 0.36, 0.32 and 1 (from left to right). The diameter of the ring in the azimuthal image is 283 nm. The FWHMs of the images taken under radial and azimuthally polarised light are 380 nm and 210 nm respectively.

Figure 9.2 shows images of 100 nm TetraSpeck beads under three different excitation polarisations. To gather this image, we set the SLM to display an RGB hologram to generate an azimuthal, radial and circularly polarised focus. We used 32 µW of excitation power (measured in the pupil), with a total pixel time of 350 ns over 50 frames. This corresponds to 3.0×10^7 photons arriving at each pixel, and the peak number of photons gathered is 379, giving a quantum efficiency of 0.0013%. We measure the peak intensity of the beads in the circularly polarised image to be 1.1×10^6 cps (counts per second), and the
relative intensities of the beads in the set to be 0.27 ± 0.03, 0.34 ± 0.1 and 1, reading from left to right. This matches simulation well but the azimuthally polarised beam is slightly dimmer than predicted above. We measure the FWHM of the image of each bead to be 360 ± 40 nm under radial illumination and 240 ± 40 nm under circularly polarised illumination. We measure the diameter of the rings to be 320 ± 40 nm under azimuthal illumination. These results are all larger than the predicted values by 15-30%. This is likely to be due partly to the finite size of the beads and partly to the fact that our pupil has a Gaussian intensity profile as described in the previous chapter.

Figure 9.2. Micrographs of 100 nm TetraSpeck beads obtained under azimuthal, radial and circularly polarised excitation. These images have aberrations corrected manually. Each image contains 50 frames and is 251×251 pixels and was obtained with 13 µW in the pupil.

The images shown in Figure 9.2 have been aberration compensated by manually adjusting Zernike mode amplitudes. The applied corrections are shown in Table 9.1. Figure 9.3 shows an image of the same field of view with no aberration correction applied. We correct the aberrations by zooming in on one or two beads and adjusting amplitudes one-by-one. To find the source of these aberrations we imaged the low NA focal spot formed at the focus of L1 (the first lens after the hologram) using a fluorescent slide and CCD camera. We also show the corrections applied to the L1 focus in Table 9.1. The only amplitude that differs from the correction applied at the sample is the coma in the Y direction. Therefore most of the aberration introduced in our microscope is introduced before L1.
Table 9.1. Table of aberration compensation amplitudes applied to the images shown in Figure 9.2. The Zernike mode magnitudes are shown as peak-valley amplitudes.

<table>
<thead>
<tr>
<th>Zernike mode</th>
<th>Aberration amplitude / waves</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>At sample</td>
</tr>
<tr>
<td>Astigmatism 1</td>
<td>0.63</td>
</tr>
<tr>
<td>Astigmatism 2</td>
<td>0.63</td>
</tr>
<tr>
<td>Coma X</td>
<td>0.13</td>
</tr>
<tr>
<td>Coma Y</td>
<td>−0.79</td>
</tr>
</tbody>
</table>

Figure 9.3. Micrographs of 100 nm TetraSpeck beads obtained under azimuthal, radial and circularly polarised excitation but with no aberration compensation applied. The Strehl ratio is approximately 0.8. The image parameters are the same as in Figure 9.2.

9.1.1. Aberrations introduced by scanning

In the figures above we showed images measured at the centre of the image plane. Figure 9.4 shows two images obtained with central position (+42.8 µm, +42.8 µm) in sample space. We first recorded a focused image taken centred on the origin in image space. We then moved the scanned image area to the new centre position. The image is shown on the left in the figure and is clearly degraded in quality. We then refocused the sample by −0.75 µm (nearer to the objective), and recorded another image, shown on the right in the figure. The image on the right appears nearly aberration free, so the dominant aberration in the system is field curvature. The diffraction-limited depth of focus is \( \pm 0.5n\lambda/NA^2 = \pm 0.21 \mu m \). This field curvature restricts the diffraction-limited
image size to be ±20 μm (e.g. Figure 9.5). Based on simulations, we had expected ±40 μm. Perhaps the source of this discrepancy is that the cover slip is not flat. We should test this in future work with beads coated onto a calibrated flat surface.

![Image](image.png)

**Figure 9.4.** Micrographs of 100 nm TetraSpeck beads obtained under circularly polarised excitation. We focused on beads at the centre of the scan field and then moved out to centre at (+42.8 μm, +42.8 μm). The image is shown on the left. We then refocused by −0.75 μm (negative z is towards the microscope objective). The beads appear aberrated, so the dominant aberration is field curvature. Each image contains 20 frames and is 251×251 pixels and was obtained with 13 µW in the pupil.
Figure 9.5. Micrographs of 100 nm TetraSpeck beads obtained under circularly polarised excitation. This shows roughly the largest diffraction-limited field-of-view that we can obtain. The image contains 20 frames and is 1001×1001 pixels and was obtained with 13 µW in the pupil.

9.1.2. Formation of radially polarised central maximum

We measured images of 100 nm beads while varying the beam diameter in the objective pupil using an iris to show the variation in the structure of the radially polarised focal spot as the NA increases. The results are shown in Figure 9.6. We placed the iris just before the SLM, and set the beam diameter to 1, 2, 3 and 4 mm. When accounting for the pupil magnification between SLM and pupil, a factor of 5/6, this gives beam radii in the pupil of 0.83, 1.67, 2.5 and 2.8 mm (the objective pupil has radius 2.8 mm). The corresponding values of NA are 0.42, 0.82, 1.25 and 1.4. As in the RGB image sets shown above, the left-hand image is
obtained with an azimuthally polarised beam. The azimuthally polarised beam has no axial field component and simply shrinks as the NA increases. When NA = 0.42, the radially polarised image and azimuthally polarised image look identical. As the NA increases, the spot in the centre of the radially polarised beam starts to fill in. The intensity scales on these images are not directly comparable because we have used different pupil powers for each. The pupil powers are 200 μW, 32 μW, 13 μW and 13 μW from top to bottom.
Figure 9.6. Micrographs of 100 nm TetraSpeck beads obtained under azimuthally, radially and circularly polarised excitation (reading from left to right). The four sets of images are obtained with NA=0.42, 0.83, 1.25, 1.4 (from top to bottom). Notice how the radially polarised beam changes shape as the NA increases. The images are all 251×251 pixels. The images contain 100, 50, 50 and 30 frames (from top to bottom). The pupil powers are 200, 32, 32, and 13 µW (from top to bottom).
9.1.3. Enhancement of axial fields at an interface

In previous work [83] other researchers have predicted an enhancement of the axial field component when focusing into a region of lower refractive index. In this subsection we describe our own model of focusing through a refractive index mismatch, present simulation results and also show images of fluorescent microspheres at an interface. Previous authors have developed an analytical theory of the focusing of light at high NA through an interface [268–272], but here we calculate the results numerically, which allows us to use our fast CZT-based approach presented in chapter 6.

![Diagram showing focusing onto a sample of fluorescent beads.](image)

Figure 9.7. Showing the geometry of focusing onto a sample of fluorescent beads. All results described up to this point have been obtained with beads immersed in glycerol (n=1.47). In this circumstance there is no appreciable interface between the front face of the objective and the beads. On the right we show a sample where the beads have been dried onto a cover slip and are not immersed in glycerol. Rays outside a cone defined by NA = n_2 (shown in red) are converted to z-polarised evanescent fields in the region containing the beads.

We use equations (76) and (77) in chapter 6 for calculating the fields at the focus of a high NA lens. These equations are presented in terms of integrals over ray directions (k_x, k_y, k_z). We calculate the fields at an interface between a medium of refractive index n_1 and n_2 as follows. n_1 is the index of the immersion medium and cover slip and n_2 is the refractive index where the sample is situated. The geometry is shown in Figure 9.7. To work out the focal fields after an interface, we apply three transformations. As before, we work out the fields on the focal sphere by applying a rotation to the fields in the pupil. We then work out the
strength of the field after the interface by calculating the s and p polarised components of the field on the focal sphere, calculating Fresnel transmission coefficients as a function of pupil position and recombining the s and p fields to find the fields in the second medium. If $\text{NA} > n_2$ then all the rays with angles of incidence at the interface greater than the critical angle, $\theta_c = \sin^{-1}(n_2/n_1)$ are totally internally reflected and give rise to evanescent fields in the sample medium. We also calculate the phase delays introduced by focusing through the refractive index mismatch as shown in Figure 9.8. After the interface, the wavefront propagates a distance $L_b + \delta_{bc}$ in refractive index $n_2$ so the optical path is $n_2L_b + n_2\delta_{bc}$. In the absence of an interface, the wavefront would have experienced an optical path difference $n_1L_a$. The total phase accumulated is therefore

$$\phi(\theta_1) = k(n_2L_b - n_1L_a + n_2\delta_{bc}) = kd(n_2 \cos \theta_2 - n_1 \cos \theta_1) + k \delta z n_2 \cos \theta_2,$$

(102)

where $n_1 \sin \theta_1 = n_2 \sin \theta_2$. This is the same as Booth, Neil and Wilson’s result [273] with an extra additive term for the displacement away from focus. We display the result of the focal field of a radially polarised beam at a mismatched refractive index boundary in Figure 9.9. The field is predominantly axially polarised. The amplification of the z-polarised component relative to the in-plane components arises for the radially polarised beam because all of the energy is p polarised at the interface, and all p polarised components become completely axially polarised for rays outside of a normalised pupil radius $n_2/\text{NA}$. The fraction of the pupil that gives rise to axial evanescent waves is given by $(1 - n_2/\text{NA})^2 = 0.49$. Furthermore the amplitude of the evanescent waves is amplified by a factor $t_s = 2$ for s-polarised waves and $t_p = 2n_1/n_2 = 3$ for p-polarised waves at the critical angle. The amplification is lower for angles either side of the critical angle.
Figure 9.8. Geometry of phase error introduced when focusing through a refractive index mismatch. We consider the phase change introduced to a particular wavefront propagating at $\theta_1$ to the optic axis in the space with refractive index $n_1$ when attempting to focus at $F$, a distance $d$ into the medium of refractive index $n_2$. We also consider the phase error introduced with an axial displacement to $P$ a distance $\delta z$ from $F$.

We also modelled the collection of light by the same objective. The interface allows fields propagating in the sample space to couple to rays propagating towards the objective. Despite this, the numerical aperture of collection is not effectively increased because the transmission coefficients are very small for highly inclined rays.
Figure 9.9. Focal fields just inside a medium of low refractive index at the focus of a high NA lens with a radially polarised pupil. We have used $\text{NA} = 1.4$, $n_1=1.52$ and $n_2 = 1$. There is very little energy in the $x$ and $y$ polarised components of the focal field and the field is predominantly axially polarised throughout the focal plane. Figure 6.7 shows the focal fields of a radially polarised beam in the absence of an interface, where the in-plane polarisations carry more energy relative to the axially-polarised component.

To round off our simulations, we show simulated images of point-like beads in Figure 9.10. These simulations combine the calculations of focal field intensity and collection of fluorescence described above. This figure should be compared to Figure 9.1. The relative maximum intensities in these simulated images are 0.61, 1 and 0.85 for azimuthally, radially and circularly polarised light. The FWHMs of the spots under radial and circular excitation are 190 nm and 270 nm respectively. The change in relative strengths of these is due to the amplification of the axial field component, which supresses the in-plane ring around the axial focus of the image with the radially polarised excitation pupil and amplifies the axial ring around the in-plane focus of the image with a circularly polarised excitation pupil.
Figure 9.10. Simulations of a point-like bead situated at an air-glass boundary, as shown on the right in Figure 9.7. The relative intensities of the three images are 0.62, 1 and 0.85 respectively. The diameter of the bright ring in the simulation with azimuthally polarised light is 300 nm. The FWHMs of the simulations with radially and circularly polarised light are 190 nm and 270 nm respectively.

We prepared samples of fluorescent beads dried onto a cover slip to test out these simulations. We show the results in Figure 9.11. These results should be compared to those shown in Figure 9.1. Notice that the beads in the image taken with radially polarised light are brightest. Relative to this, the beads in the circularly polarised image have a brightness of 0.6 ± 0.1, and those in the azimuthally polarised image 0.4 ± 0.1. The FWHMs of the beads in the images are 240 ± 40 nm for the radially polarised image and 400 ± 40 nm for the azimuthally polarised image. The FWHM of the radially polarised beads is larger than predicted, though not by a significant amount compared to measurement error. The FWHM of the circularly polarised beads is significantly larger than predicted. The larger size of the circularly polarised focal spot and the lower maximum intensity (relative to the radially polarised image) both suggest that there is a greater degree of axial field amplification than predicted in our simulation. We have not been able to explain this result.
Figure 9.11. Micrographs of 100 nm TetraSpeck beads at an air-glass interface obtained under azimuthal, radial and circular illumination polarisations. Compare with Figure 9.3 which has images taken of beads immersed in glycerol. The relative intensities are 0.4 ± 0.1, 1 and 0.6 ± 0.1 respectively. The diameter of the bright rings in the image taken with azimuthally polarised illumination is 360 ± 40 nm. The FWHMs of the spots in the images taken with radially and circularly polarised illumination are 240 ± 40 nm and 400 ± 40 nm respectively. Each image contains 30 frames and is 251×251 pixels and was obtained with 13 µW in the pupil.

9.2. Single molecules

In the previous section we presented results of imaging sub resolution fluorescent beads, each one of which contains many fluorophores. In this section we present images of single fluorescent dipoles. Single fluorophores are commonly modelled as electric dipoles. We presented a review of fluorophore orientation determination techniques in chapter 5, and tried to explain the limitations of the existing techniques. Here we present results of imaging single molecules with our polarisation confocal microscope. As in the previous section, each image is presented as a set of three that was taken with an RGB hologram set to form an azimuthal, radial and circularly polarised focal spot. The fact that the three images are obtained interleaved is particularly important when imaging single molecules because it mostly negates the effects of photobleaching.

We prepared our samples by diluting Dil molecules in a 1% PMMA solution in 10% Acetone and 90% Toluene. The Dil molecules were diluted down to a
concentration of $10^{-9}$ or $10^{-10}$ Mol L$^{-1}$. We spin coated the samples onto clean cover slips at 3000 rpm, then glued the cover slips onto glass slides. We deposited 6 μL of this solution onto a cover slip. Assuming that all of this solution becomes evenly deposited onto the cover slip, which is 20 mm × 20 mm, we should have roughly 10 or 1 molecules per square micron in the sample.

Figure 9.12 shows an image of a single molecule sample imaged with radially polarised light. We predicted the images of single molecules imaged in radially polarised light in chapter 6, figure Figure 6.13. To summarise, an in-plane molecule should appear as a pair of lobes, with the direction of the dipole pointing along a line joining the brightest points in the two lobes. A molecule pointing along the z axis should appear as a circularly symmetric spot surrounded by a ring. As the molecule turns away from the z axis, the relative intensity of the ring increases and the intensity of the central spot decreases. There are some of all of these features shown in Figure 9.12. On the right, we have annotated the image with white arrows to show the inferred orientation of in-plane molecules. We have also labelled three molecules in pink. These three molecules have bright central spots, indicating that they are largely oriented along the optical axis.
Now we turn our attention to Figure 9.13. This shows a single molecule imaged with an RGB hologram. From the three images obtained with this RGB hologram mode we can be more precise about our inferences of single molecule orientation. First, look at the central image. This shows a bright spot, with no discernible mixture of side lobes of the correct degree of angular symmetry to indicate a deviation from a vertical molecule. We can gain extra information about the molecule orientation from the image taken with azimuthally polarised light. This image shows, albeit very faintly, a double-lobed structure. Therefore the molecule must be at least partly rotated away from the optical axis to point along $\mathbf{y}$. The image obtained under circularly polarised light has a dark central spot because the focus of a circularly polarised beam is surrounded by an axially polarised doughnut mode. Figure 9.14 shows three simulated images of a dipole in a confocal microscope with azimuthally, radially or circularly polarised excitation pupils. The molecule in this simulation is at $\theta = 20^\circ$, $\phi = 270$ where $\theta$ is the polar angle (the $z$ axis is the pole), and $\phi$ is the azimuthal angle measured from $x$. There is a very good correlation between the simulated and experimental images, suggesting that the dipole lies close to $\theta = 20^\circ$, $\phi = 270$. The asymmetry in the circularly polarised image seems at first counterintuitive, especially when we consider the field strength of the excitation field, shown in
Figure 9.15. The spatial distributions of all components of the field strengths are circularly symmetric. The asymmetry in the image arises because there is a $\pi$ phase change in the $z$ field component across the PSF, but no phase change in the $y$ field component. Therefore there is a 90° difference in the polarisation azimuth vector, so the dipole is almost parallel to the polarisation azimuth on the left hand side of the image, but almost perpendicular on the right hand side. Notice that with the azimuthally polarised image alone we could not have determined if the dipole orientation was at $\theta = 20^\circ, \phi = 270$ or $\theta = 20^\circ, \phi = 90$. The circularly polarised image removes this $\pi$ ambiguity.

Figure 9.13. Shows a single molecule imaged using an RGB hologram, configured to give azimuthally polarised light (left), radially polarised light (centre) and circularly polarised light (right). This is an example of a molecule that is largely oriented along the optic axis, but tilted slightly towards the $y$ axis.
Figure 9.14. Simulations of a single molecule illuminated by azimuthally, radially and circularly polarised light. The molecule is at 20° to the $z$ axis, and has no component along $y$. Each image contains 200 frames and is $26 \times 26$ pixels and was obtained with 200 $\mu$W in the pupil.
Figure 9.15. Simulations showing the field intensities at the focus of a circularly polarised beam. Notice how the distribution of every field component is circularly symmetric and yet the image of the dipole excited by this field is not circularly symmetric because the polarisation azimuth of the field influences the dipole excitation strength.

Figure 9.16 shows another set of images taken with RGB holograms set to produce azimuthally, radially and circularly polarised beams. In Figure 9.16 we have provided two copies of the images. On one copy of the image taken with radially polarised light we have drawn white arrows and labelled some molecules. As above, the white arrows indicate molecules that lie near to perpendicular to the optic axis ($\theta = \pi/2$). We can infer this because there is no discernible intensity (over background) in the centre of the PSFs in the radially polarised image. Note that the lobes in the azimuthally polarised images always lie perpendicular to the corresponding lobes in the radially polarised image. By comparing the radially polarised and azimuthally polarised images, we can infer that some molecules are oriented almost completely along the optic axis. For example, molecules 1 and 3 do not show up in the azimuthally polarised image.
Molecule 2 appears similar to the molecule in Figure 9.13: the PSF in the radially polarised image looks circularly symmetric but there is a noticeable double-lobed feature in the azimuthally polarised image. The largely axial orientation of molecules 1-3 is further confirmed by the presence of the dark spot at the correct locations in the circularly polarised image (as in Figure 9.13). Molecules 4 and 5 are both more highly inclined, as they appear as double-lobe structures with a dim but noticeable central peak. By comparison with simulation, we estimate that these molecules are both between 60° and 70° to the optic axis.

![Images of single molecules taken with RGB holograms set to azimuthal, radial and circular polarisations.](image)

**Figure 9.16.** Images of single molecules taken with RGB holograms set to azimuthal, radial and circular polarisations. The lower set of images is a copy of the upper set with orientations marked and some molecules labeled. We discuss the inferences of the molecule orientations in the text. Each image contains 100 frames and is 151×151 pixels and was obtained with 200 µW in the pupil.

Figure 9.17 shows one final set of images objects that appear like single molecules. There is a linear feature and the azimuthally polarised image (also shown enlarged) shows that there is a degree of alignment along the direction of the line. This can be seen because the line has a dark centre in the azimuthal
image, but appears as bright points in the circularly polarised image. We cannot confidently say that this is indeed a set of single molecules because it would seem highly unlikely that such a deposition would occur with all the molecules aligned.

Figure 9.17. Single molecule image taken with RGB holograms. We have enlarged the azimuthally polarised image to show a strange feature: all of the objects along this linear feature seem to be aligned with the local direction of the line. Each image contains 50 frames and is 251×251 pixels and was obtained with a pupil power of 200 µW.
9.3. Nitrogen vacancies in nanodiamond

In this section we discuss the interaction of polarised light with fluorescent nanodiamonds containing NV centres. We have already shown images of small objects containing many fluorescent dipoles in section 9.1 and images of single molecules in section 9.2. The NV centre is an intermediate between these extremes because a single NV centre contains two fluorescent dipoles [274]. Our microscope can resolve these two dipoles.

The NV centre is just one of around 100 optically active luminescent defects present in diamond [275]. It consists of a nitrogen atom substituted for a carbon atom, with one of the adjacent carbon atoms vacant. The structure of the NV centre is shown in Figure 9.18. The centre exists in two charge states: NV0 and NV-. In this work we focus on NV- as it is excited by green light. The zero phonon line of the NV- centre is at 637 nm, with wide side bands in absorption extending up to 520 nm and in emission down to 680 nm [274]. Despite this, experimental work in our group has determined that the quantum efficiency of emission is low when the NV centre is excited near 637 nm, and highest fluorescence efficiency is observed for excitation between 520 nm and 540 nm [276], which is ideal for observation on our microscope. This could be explained by the observation that NV- centres are converted to NV0 centres when excited near 637 nm [277]. In this section we present images of nanodiamonds, which are small crystals of diamond and can be obtained in 10 nm-100 nm sizes from Adámas nanotechnologies (North Carolina, USA). Nanodiamonds can be manufactured by detonation of carbon-based explosives or by grinding bulk diamond. Nanodiamonds have found uses in biological imaging because they display a very high degree of photostability [278]. In-vivo imaging has been performed with nanodiamonds in C. elegans [279] and nanodiamonds can be used for biological labelling [280].

The NV centre has an axis of symmetry along the vector joining the nitrogen atom to the vacant lattice site. There is a plane of three-fold rotational symmetry about this axis. This symmetry gives rise to two orthogonal dipoles in the plane
perpendicular to the NV axis [274,281]. At room temperature the excited state is a thermal mixture of these two dipoles [282]. The drawing on the right in Figure 9.18 shows one possible orientation for the two dipole axes.

There has been previous work using a radially polarised beam to determine the orientation of single NV centres [179]. Since there are two orthogonal dipoles in an NV centre, its orientation is best characterised by the orientation of the normal to the plane containing both of these dipoles. Let this normal be $\mathbf{v}$. The authors image nanodiamonds with radially polarised light and then perform pattern matching on the resulting image to determine dipole orientations. The authors try to match the obtained images with one from a set of computed images of the nanodiamond emitter under radially polarised light. Unfortunately the nanodiamond emitter is modelled as a coherent combination of two orthogonal dipoles; this doesn’t make sense as the field from a coherent combination of two orthogonal dipoles along $\mathbf{d}_1$ and $\mathbf{d}_2$ is the same as that from a single dipole along $\mathbf{d}_1 + \mathbf{d}_2$. The fluorescence from an NV is more correctly modelled as the incoherent sum of fluorescence from two orthogonal dipoles.

We have predicted the image of a single NV centre taken with radially polarised light and we present these results in Figure 9.19. This figure is deliberately produced in the same layout as figure 3 in reference [179]. We present the images as a function of the direction of the normal to the plane containing the two dipoles, which is also the NV axis $\mathbf{v}$. $\theta$ is the angle between $\mathbf{v}$ and $\mathbf{z}$. $\phi$ is the
azimuthal angle with $\phi = 0$ along $x$. When $v$ is along $z$ (top left image) the two dipoles lie in the $xy$ plane and so are excited by the ring of in-plane polarised light surrounding the axial focus (for a reminder of the field structure at the focus of a high NA lens with a radially polarised pupil see Figure 6.7). As $v$ rotates away from $z$ towards $x$, one dipole starts to interact with the axial component of the focal field. When $v$ is along $(\theta, \phi) = (90,0)$, the dipoles are along $y$ and $z$. Rotation about the $z$ axis produces a rotation in the emission pattern, as expected.

![Figure 9.19. Simulated image of an NV centre illuminated with radially polarised light at the focus of a high NA lens. The angles measure the direction of the normal to the plane containing the two orthogonal dipoles.](image)

We performed our own imaging of nanodiamond samples with the aim of determining the NV axis orientation. We obtained samples of fluorescent nanodiamond crystals from Adámas Nanotechnologies. We purchased 30 nm nanodiamond crystals in solution with one to three NV centres per particle (advertised). Technical data from the supplier indicates that roughly 10% of the nanodiamond crystals should contain only one NV. We prepared our samples by diluting the NDs at 1:10 in a 1% solution of PMMA in Ethyl Lactate. 20 µL of this solution was then spin coated at 3000 rpm onto a clean cover slip. With these parameters we can predict roughly a 30-100 nm film thickness [283,284]. We observe that samples made in this way contain only one in-focus plane, so the film is optically thin. The PMMA layer has refractive index 1.49 [285] so we can
safely neglect interface effects between it and the cover slip. Beyond the PMMA layer is an air gap, which will cause reflection of light back towards the sample. We show the layout of this system in Figure 9.20. We developed a model of imaging in a layer such as this, explained in Figure 9.21. We imagine that there is a second virtual objective behind the interface, and calculate its pupil function by applying the appropriate Fresnel reflection coefficients. We then add the fields from the real and virtual objectives, remembering to reverse the direction of the axial component of the field. The real pupil function shown in Figure 9.21 is the absolute value of the $x$ component of the pupil function of a radially polarised beam. The virtual pupil function is calculated, and shows the region of weak reflection for small angles of incidence, surrounded by a strongly reflected annulus for incidence angles greater than the critical angle.

Figure 9.20. Geometry of imaging of NDs in a thin PMMA layer.
Using our model we calculated the image of single NV centres under azimuthal, radial and circularly polarised light. We show the results in Figure 9.22. These results, like those in Figure 9.19 are presented as a function of the angle that the NV axis. First look at the azimuthal image set. When the NV axis is along $z$ (top left), the two dipoles are both in-plane and so the NV image appears symmetric. As $\theta$ increases, from left to right, the NV axis turns towards $x$ and so the component of one dipole interacts less strongly with the azimuthally polarised focus. When $\theta = 90^\circ$, one dipole lies along $z$ but this is not seen because the azimuthally polarised focus has no axial component. As $\phi$ changes the PSF rotates, as would be expected. Next turn to the radially polarised image. Here all images of the NV appear almost circularly symmetric; the effect of the interface has been to amplify the axial components of the field, similarly to the effect we saw in the region after the interface in section 9.1.3. Because of this we mostly see the dipole with an axial component. In the images taken with circularly
polarised light, the focus appears strange and distorted. We explain this with reference to Figure 9.23, which shows the field components at the focus of an objective with circularly polarised pupil. Though the total PSF, shown middle bottom, is circularly symmetric, the $x$ and $y$ field components are not. In some orientations the dipole interacts with just one of these components, so the resulting image appears as calculated in Figure 9.22.

![Figure 9.22. Simulations of single NV centres under azimuthal, radial and circularly polarised light.](image)

<table>
<thead>
<tr>
<th>Coordinates are $(\theta, \phi)$</th>
<th>(90,45)</th>
<th>(90,90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0)</td>
<td>(30,0)</td>
<td>(60,0)</td>
</tr>
<tr>
<td>(90,0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 9.23. The electric field components at focus of a circularly polarised beam just inside an interface boundary as described in Figure 9.20. The $x$ and $y$ field components pick up a characteristic spiral shape which results in distorted images as shown in Figure 9.22.

Figure 9.24 shows images of a candidate single NV taken with our microscope. We obtained these images as an RGB set, with azimuthally, radially and circularly polarised light. We believe these to be NVs because they were measured to be exceptionally photostable. We measured the photobleaching in single molecule samples and compared it to that in ND samples. First, we recorded an image, then delivered a light dose to the sample by repeatedly scanning the same area, then recorded a second image. With a single molecule sample, we delivered a light dose of $2 \times 10^5$ J cm$^{-2}$ to the sample and observed the photoluminescence reduce to about 10% of the original intensity after the light dose was delivered (without background subtraction), whereas for the images shown in Figure 9.24 we delivered a dose of $4 \times 10^6$ J cm$^{-2}$ to the sample and observed no change in image intensity after the delivered dose. These doses were both delivered over a time of roughly 1000 s. Gruber et. al measured no change in the luminescence.
intensity of NV centres when imaged with a power density of 5 MW cm$^{-2}$ [278]. Now we comment on the images shown in the figure. If these were objects containing many dipoles, we would see a circularly symmetric shape in the azimuthally polarised image as seen in section 9.1. Therefore neither object shown in these images responds isotropically, or bead-like, to polarised light. If they were single molecules, there would be a double-lobed structure in the azimuthally polarised image as seen in section 9.2. If they appeared strongly in the azimuthally polarised image they would also appear double-lobed in the radially polarised image because the transition dipole would be substantially in-plane. Therefore, since both objects appear strongly in the azimuthal and the radial image they cannot be single dipoles either. We propose that both objects are NVs, with the NV axis oriented at roughly 135° to $x$ ($x$ is horizontal in the images). The upper left species has a smaller angle between its normal and the $z$ axis because it appears relatively stronger in the azimuthally polarised image, and because it appears with a less distinct minimum in the azimuthal ring. We conclude that the lower right species has its axis near normal to the $z$ axis because the minimum in the azimuthal ring is almost at background intensity levels. We show an example simulated image of an NV at $\theta = 60^\circ$, $\phi = 135^\circ$ in Figure 9.25 (simulations as in Figure 9.22). This is the most plausible match to the object shown in the lower right of Figure 9.24. The comma shape in the circularly polarised image is reproduced, though not in the correct orientation.
Figure 9.24. Images of a candidate single NV centre taken with azimuthally, radially and circularly polarised light. Each image contains 100 frames and is 101×101 pixels and was obtained with 200 µW in the pupil.

Figure 9.25. Simulations of a single NV centre imaged with azimuthally, radially and circularly polarised light. The NV axis is at $\theta = 60^\circ, \phi = 135^\circ$.

9.4. Quantum Dots

In this section, we briefly present images of fluorescent quantum dots (QDs), QD 625 supplied by ThermoFisher. These particular quantum dots have have diameters of 15-20 nm. A quantum dot is a spherical particle and therefore by symmetry has three degenerate excited states and hence three degenerate transition dipoles. For these reasons we expect quantum dots to behave similarly to fluorescent beads. We prepared samples of these QDs by diluting
them by 1:10 in H₂O and then 1:10 in Ethanol. We then dried 3 µL of this solution onto cleaned cover slips. The images are shown in Figure 9.26. As expected, each QD appears circularly symmetric in the azimuthally polarised image. The images show the same features as those of beads at an interface as shown in Figure 9.11. We obtained these images using a pupil power of just 12.6 µW. At higher pupil powers we observed saturation of the quantum dots, visible by increased apparent relative intensity of PSF sidebands.

![Image of Quantum dots taken with polarised light.](image)

**Figure 9.26.** Images of Quantum dots taken with polarised light. As expected, the QDs behave isotropically. Each image contains 100 frames and is 121×121 pixels and was obtained with 5 µW in the pupil.

### 9.5. Chapter summary

In this chapter we have presented a collection of images taken with our confocal polarisation microscope. We have presented the images of four simple types of objects: subresolution fluorescent beads, single molecules, nanodiamonds and quantum dots. Importantly, the final three of these are single emitters, and we have shown that they behave in characteristic ways that are consistent with their own fluorescent dipole characteristics.

We first used the fluorescent beads to measure the aberrations in our system. We found that the system was nearly diffraction limited without aberration correction, and that most of the system aberrations were introduced at or before the SLM. Even with no aberration correction, our system achieved almost
diffraction-limited performance. We found that the aberration introduced by
beam scanning is mostly field curvature. This limits the maximum scanned
sample size to be $40 \times 40 \, \mu m$. We were able to use the beads to show the
formation of the central bright spot in an radially polarised PSF as we increased
the NA of our microscope.

The fluorescent beads and QDs respond isotropically to the exciting field. In the
case of the fluorescent beads this is because there are many fluorescent dipoles
in each bead, whereas in the case of the QDs their excitation and emission is
spherically symmetric because the QDs themselves are spherically symmetric.
The single molecule images show orientation effects that let us determine
(qualitatively, for now) the direction of the transition dipole moment for each
molecule. The NDs show more complicated orientation effects because each
fluorescent nitrogen vacancy contains two orthogonal dipoles. We were able to
show that we can observe these single NV centres, and determine their
orientation.

We estimate the orientation of these objects by comparing their images with
azimuthally, radially and circularly polarised light with simulated results. In
future work we should develop pattern-matching techniques. As in DOPI, these
will allow quantitative determination of dipole angles. To have an accurate
library of images we needed to develop accurate simulations to correctly
characterise our results. We added interface effects into our model of a
polarisation microscope presented in chapter 6. This model proved to be very
accurate in predicting the image of fluorescent beads at a glass-air boundary.
Here we see large amplification of the axial fields due to total internal reflection.
Our model also allowed us to predict the fields inside a spin-coated layer of
PMMA next to an air gap.
10 Conclusions

Our microscope is not the first to sense the 3D orientation of molecules. We review many other techniques in chapter 5 but an important thing is missing from all of these techniques: there is no existing method that allows determination of the 3D orientation of a molecule in the far field (i.e. away from the cover slip). Our microscope addresses this inadequacy by using a radially polarised beam to excite axially oriented molecules. The images taken with radially polarised light are compared to images taken with other polarisation states and this comparison allows the 3D orientation to be found. Our microscope has a particular advantage, afforded by the incorporation of a fast binary SLM: we can switch the excitation field on the same timescale as the beam scanning and so build up sets of images taken with different excitation polarisations, where the sets of images are obtained almost in parallel. This means intensities can be compared between images in a set. This may be a key step in the further development of this technique to the application of imaging in dense samples. Our method offers a key advantage over existing pattern-recognition techniques, particularly DOPI, in that the molecules are observed in focus, so we can image approximately 100 times higher molecule densities (Figure 5.1 shows that the DOPI PSFs are 10 times larger in each dimension than in-focus PSFs) and have many more counts per pixel (the peak intensity for a PSF with 1 µm of defocus is about 40 times lower than in focus), so the influence of shot noise is reduced. We were also able to show that the quasi-parallel acquisition of images taken with different excitation polarisations yields more information than just acquiring one image with say, radially polarised light. For example, when a dipole is nearly aligned along the optical axis, we can sense the azimuthal angle using the image taken with azimuthally polarised light. Also, the image with circularly polarised light resolves the ambiguity in the polar angle (which could be at θ or π − θ if we didn’t have the circularly polarised beam).
We also sense the 3D orientation of single NV centres in nanodiamond. One other research group [179] has proposed and attempted this, but they used only a radial beam. We showed that the approach using just the radially polarised beam would not yield information about NV orientation when the NVs are situated near an interface because the signal from the axial dipole component swamps the in-plane signal. In future work we should use an extra PMT to construct a Hanbury-Brown-Twiss apparatus [286]. This would allow us to conclusively identify single emitters. As it stands, we identify single emitters based on their image under different excitation polarisations. Most nanodiamonds respond isotropically (i.e. they look like fluorescent beads), presumably because there are often several NV centres in each nanodiamond.

Our microscope only allows qualitative determination of the dipole angle. In future work there could be two promising avenues for obtaining quantitative results. Firstly, in single molecule samples, we could implement an algorithm, like is used in DOPI, to automatically match dipole images to computed PSFs. This could be done using a maximum likelihood-method, or using phase diversity techniques such as the Gerchberg-Saxton algorithm. We showed in chapter 9 that our models produce very accurate predictions of the images of dipoles, so we could expect good accuracy from this approach. Secondly, we would like to develop this technique to work without pattern recognition. This could be done in the single molecule and in the dense labelling regime. We discussed this possibility briefly in chapter 6, and can now comment further as we have presented our results. In the single molecule regime, the idea is to segment the images into regions containing only one fluorophore, and then integrate the total intensity for each fluorophore and report it as a function of excitation polarisation. Perhaps the full 3D fluorophore orientation could be determined with four excitation polarisations, with polarisation azimuths pointing along the four corners of a tetrahedron. This technique could also be applied to image densely labelled samples, though reassigning light from the pixel position in image space to the molecule position would be impossible, so the spatial resolution of the orientation determination would be degraded.
On the way to developing our microscope, we showed a number of interesting side results. The first new work was presented in chapter 4. Here we described a robust method for control of the polarisation of a light beam using an FLCSLM. We showed that we can generate a polarisation controlled beam by using two holograms that each control one linear polarisation component of the output beam. Hologram A is imaged onto B using a reflective 4F system, which means that the alignment and adjustment are simple. This polarisation control scheme has the potential to be useful wherever someone needs to generate an arbitrary polarisation state. The FLCSLM allow us to switch the polarisation states quickly and correct aberrations. As mentioned in the text, this work on using SLMs with reflective 4F systems also led us to develop a method for controlling light using a PANSLM. The STED system uses the two holograms to independently modulate two mutually incoherent components of a pulsed laser beam to produce lateral and axial resolution enhancement. This work has significantly improved the user-friendliness of our group's 3D-STED microscope [76].

We also showed the development of a full numerical model of a confocal microscope that includes polarisation effects during focusing of excitation light, excitation of fluorescent dipoles, and collection of polarised light from those dipoles. Importantly, the model allowed us to compute the images of fluorescent microspheres, dipoles, NVs and quantum dots under polarised excitation. It also allowed us to highlight some differences between previous non-vectorial methods and our theory, most notably in optical sectioning calculations. The work also allows us to predict the optical sectioning strength of sheets of aligned dipoles. We have also developed our model of polarisation-dependent excitation and emission to include polarisation-dependent depletion processes, which are important in STED microscopes, though we did not present that work in this thesis.

We then discussed the construction of our microscope. First, we focused on the beam-scanning system. We designed a galvanometer-based scanning system capable of producing diffraction-limited images over a 4 by 4 mm image plane. We found that we could not acquire images of this size, probably because the
coverslips are not flat. The most important property of the scanning system is that it scans the laser beam in 2D while keeping the beam stationary in the objective pupil. We also presented a detailed model of polarisation changes introduced during the scanning process. This is a novel model, and it gave an important design principle: one should always minimise the angle difference between the input and output beams of the scanner.

Next we described the construction of the microscope in terms of optics, electronics and software. The scanning mirrors are controlled using a scan control unit (SCU) that allows us to send digital commands from a PC to move the scanners. Unfortunately this introduced a latency of 600 µs per scan line to our system, which limited our scan line speed to 180 lines per second. In future we would use the DAQ box to send analogue commands to the scan mirrors, and would therefore be able to use more than three bit planes per frame on the SLM. This would enable us to obtain images more quickly, or to obtain images taken with more than three excitation polarisation states simultaneously.

The track length of the imaging system is long: the distance between the SLM and M1 is given by \(2f_1 + 2f_2 + \frac{f_2^2}{f_{im}} = 1.9\) m. This could, and should, be reduced in a future design. To do this we would have to use a significantly shorter focal length lens as the first lens of the relay system from the SLM to the first scan mirror. This is challenging because the beams from the SLM are close, and would probably require the design of a custom lens. We could also reduce the track length by reducing the size of the scan mirrors, which would permit a shorter focal length spherical imaging mirror to be used, and therefore L2 could also have a shorter focal length.

In conclusion, then, the work here represents important advances in the field of single emitter orientation determination. We have carried out our system design and experimental work backed up by detailed modelling that has led to a number of important new results. With a little more development, and collaboration with the right specialists, the instrument presented here could become a
powerful tool and could be used by biologists and others to determine single molecule orientation.

10.1. Publications

10.1.1. Journals


10.1.2. Conferences


Bibliography


19. F. Zernike, "Phase contrast, a new method for the microscopic observation


47. M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya,


90. T. J. Gould, D. Burke, J. Bewersdorf, and M. J. Booth, "Adaptive optics


225. J. Saatsi and P. Vickers, "Miraculous success? inconsistency and Untruth in


244. E. Y. S. Yew and C. J. R. Sheppard, "Tight focusing of radially polarized


230
(1994).


Reprint permissions

For Figure 9.18

| Title: | Anisotropic Interactions of a single spin and dark-spin spectroscopy in diamond |
| Author: | R. J. Epstein, M. Mendoza, Y. K. Kato and D. Awschalom |
| Publication: | Nature Physics |
| Publisher: | Nature Publishing Group |
| Date: | Nov 1, 2005 |

Copyright © 2005, Nature Publishing Group

Order Completed

Thank you very much for your order.

This is a License Agreement between James Clegg ("You") and Nature Publishing Group ("Nature Publishing Group"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

Get the printable license.

| License Number | 3724851339678 |
| License date | Oct 05, 2015 |
| Licensed content publisher | Nature Publishing Group |
| Licensed content title | Anisotropic Interactions of a single spin and dark-spin spectroscopy in diamond |
| Licensed content author | R. J. Epstein, M. Mendoza, Y. K. Kato and D. Awschalom |
| Licensed content date | Nov 1, 2005 |
| Type of Use | reuse in a dissertation/thesis |
| Volume number | 1 |
| Issue number | 2 |
| Requestor type | academic/educational |
| Format | print and electronic |
| Portion | figures/tables/illustrations |
| Number of figures/tables/illustrations | 1 |
| High-res required | no |
| Figures | Figure 1 and Figure 2 |
| Author of this NPG article | no |
| Your reference number | None |
| Title of your thesis/dissertation | Interaction of single emitters with polarised light |
| Expected completion date | Nov 2015 |
| Estimated size (number of pages) | 180 |
| Total | 0.00 GBP |
Dear Dr. Clegg,

Thank you for contacting The Optical Society.

OSA considers your requested use of its copyrighted material to be Fair Use under United States Copyright Law. It is requested that a complete citation of the original material be included in any publication.

Let me know if you have any questions.

Kind Regards,

Susannah Lehman

Susannah Lehman  
September 4, 2015  
Authorized Agent, The Optical Society

From: Clegg, James [mailto:j.clegg11@imperial.ac.uk]  
Sent: Monday, September 07, 2015 12:54 PM  
To: pubscopyright  
Subject: Figure for thesis

Dear Sir/Madam


I wish to reprint figure 4 and its caption. I will be talking about the work in the context of a literature review about orientation determination in microscopy.

Please could you provide formal confirmation that this is OK as I need to include this in my thesis.

Regards,

James Clegg  
Imperial College London