The Development of Metasurfaces for Manipulating Electromagnetic Waves

by

Mitchell Guy Kenney

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY (*Ph.D.*)

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

The work outlined in this thesis focuses on the development and fabrication of metasurfaces for manipulating electromagnetic waves, with the potential for applications in imaging and holography.

Metasurfaces are the Two-Dimensional counterpart of metamaterials, which are artificial materials used to invoke electromagnetic phenomena, not readily found in nature, through the use of periodic arrays of subwavelength 'meta-atoms'. Although they are a new and developing field, they have already secured a foothold as a meaningful and worthwhile focus of research, due to their straight-forward means of investigating fundamental physics, both theoretically and experimentally - owing to the simplicity of fabrication - whilst also being of great benefit to the realisation of novel optical technologies for real-world purposes. The main objective for the complete manipulation of light is being able to control, preferably simultaneously, the polarisation state, the amplitude, and the phase of electromagnetic waves. The work carried out in this thesis aims to satisfy these criteria, with a primary focus on the use of Geometric phase, or Pancharatnam-Berry phase. The first-principles designs are then used to realise proof-of-concept devices, capable of Circular Conversion Dichroism; broadband simultaneous control of phase and amplitude; and a high-efficiency, broadband, high-resolution hologram in the visible-to-infrared.

Acknowledgements

A great deal of help and support has allowed me to carry out my PhD and present the work shown in this thesis, with the following people helping especially so:

- My supervisor Shuang Zhang, for giving me the opportunity to undertake a PhD under his guidance, and for his knowledge, expertise, and intuitive approach to the research undertaken by me - I am extremely grateful and thankful.

- My colleagues and collaborators, for whom I owe a great deal of gratitude. Without these, the completion of projects (which ultimately led to the publication of research) would never have come to fruition. In the words of Charles Darwin – "*In the long history of humankind (and animal kind, too) those who learned to collaborate and improvise most effectively have prevailed.*"

- My university friends, who made my research experience such a memorable and worthwhile experience. Particular focus is put onto the daily 'religious' routine of coffee – "走吧!" - where we discussed many topics, both scientific and gossip-laden. I deserve I should name all of these friends, so thank you Mark, Biao, Teun-Teun, Yao Ting, Guixin, Wenlong, Qinghua, and to those who didn't involve themselves in coffee time, but 'food-time' – Shiyi, Fu, James, Lingling, Shumei, Lixiang, Ren, Guoxing, Xin, and Xianzhong – thank you also!

- Lastly, I would like to thank my friends and family, specifically my mother Debra, my father Peter, and my brothers, Myles and Rory, as well as my friends Benn and Scott, for who I am extremely grateful for just being there and giving a helping hand when needed – even if we didn't always see eye-to-eye. I would also like to thank equally as much Alex, Gaz, Ben, Dave and Chloé for having the "sense-of-humours" that you do…

Publication List

1. Zheng, G.; Mühlenbernd, H.; Kenney, M.; Li, G.; Zentgraf, T.; & Zhang, S.; "Metasurface holograms reaching 80% efficiency", *Nat. Nanotechnol.* **10,** 308–312 (2015)

2. Liu, L.; Zhang, X.; Kenney, M.; Su, X.; Xu, N.; Ouyang, C.; Shi, Y.; Han, J.; Zhang, W.; & Zhang, S.; "Broadband metasurfaces with simultaneous control of phase and amplitude", *Adv. Mater.* **26,** 5031–5036 (2014)

3. Kenney, M.; Zhang, X.; Kim, Teun-Teun.; Yu, D.; Zhang, W.; & Zhang, S.; "Silicon Herringbone Metasurface for Giant Chiroptical Response of Terahertz Light". *Awaiting Submission* (2016)

4. Xiao, S.; Mühlenbernd, H.; Li, G.;Kenney, M; Liu, F.; Zentgraf, T.; Zhang, S.; & Li, J.; "Helicity-Preserving Omnidirectional Plasmonic Mirror", *Adv. Opt. Mater.* January 2016 (Early View), DOI: 10.1002/adom.201500705

* The following publications are not affiliated with The University of Birmingham, but are included for completeness of the publication list to date

*5. Zheng, G.; Liu, G.; Kenney, M. G.; Li, Z.; He, P.; Li, S.; Ren, Z.; & Deng, Q.; "Ultracompact high-efficiency polarising beam splitter based on silicon nanobrick arrays", *Opt. Express.* **24**(6), 6749-6757 (2016)

*6. Hao, D.; Kenney, M. G.; & Cumming, D. R. S.; "Plasmonic Gold Discs using Piezoelectric Substrate Birefringence for Aqueous Biosensing", *Submitted to Applied Physics Letters* on 1ˢᵗ April 2016

Contents

Table of Figures

ix

Chapter 1

Introduction and Review

1.1 Motivation

The field of optics is, and always will be, at the forefront of technology and developments in science and scientific endeavour – that is, at least until the point where "evolution" (natural or otherwise) surpasses our physical method, or indeed requirement, of '*seeing*' – because, as humans we require light to see and what we can see, we can comprehend. Dating back to the dawn of civilisation, optical trickery and fascinations have been a key spoke in the wheel of scientific and economic development: starting with the study of the motions of the moon and stars in the night sky [1,2]; to the shiny appeal of gold, silver and precious stones as decoration and currency (which, not surprisingly, is still the case in this day and age) [3]; the use of small metal particles for use in ornamental glasses, of which the renowned Lycurgus Cup [4,5] (**Figure 1.1a**) and stained glass windows [6,7] (**Figure 1.1b**) are key examples; and the development of glass lenses to examine worlds beyond the
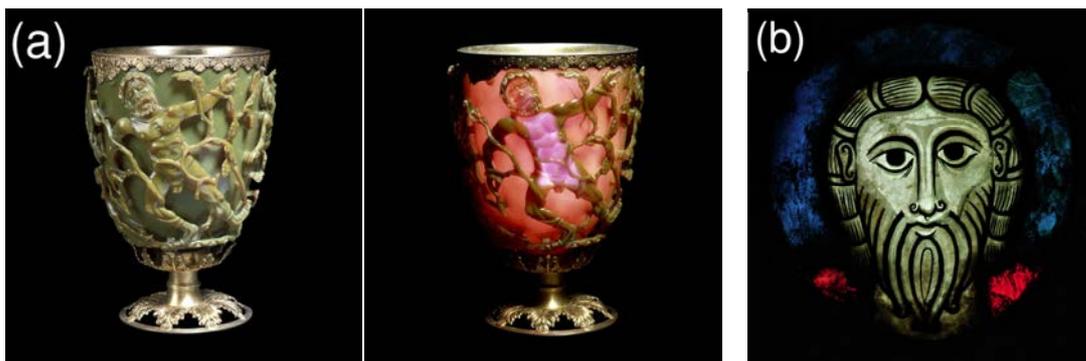
Figure 1.1: The "Lycurgus Cup" and "Christ of Wissembourg" stained glass

(a) Arguably the first commercial use of nanotechnology and plasmonics, the Lycurgus Cup (4th Century AD) makes use of small particles of gold to produce a red colour upon illumination from within [4,5,8]. **(b)** Stained glass window, named "Christ of Wissembourg", from 11th Century [9].

capability of our eyes alone – telescopes for seeing distant objects and microscopes for seeing the otherwise invisible objects too small to see, both developed around the same time in which Snell's Law was proposed [10,11]. In the last century or two, sight helped yield two of the (arguably) most important theories of our time, the first of which is the Special Theory of Relativity [12], which had its beginnings in the 16-year old mind of Albert Einstein as he pondered what it would be like to chase a beam of light [13], and the second of which was The Theory of Evolution by Natural Selection, where Charles Darwin found inspiration in the slight differences he saw in the markings of birds amidst travels in the Galapagos Islands [14].

Advancements in understanding light as an electromagnetic wave allowed the development of numerous modern day conveniences we take for granted, such as electricity, radio transmission, and lasers, where more recently things such as wireless

internet, full colour flatscreen displays, holography, and light emitting diodes have proven invaluable to society. One of the more notable turning points in the investigation of electromagnetism was the theoretical accumulation of the links between the charge density and current through a wire to the associated magnetic and electric fields (and also the link between the magnetic and electric fields for a wave propagating in free space) by James Clerk Maxwell in 1865 [15]; this is nowadays termed as classical electromagnetism or classical electrodynamics. Following this, work carried out by Jagadis Chunder Bose aided in the study of polarisation effects, whereby metal foil films were placed between the pages of a book and a linear polarisation effect observed [16] (where even a book without the foil showed this same effect, although less pronounced). He also experimented on twisted jute bundles, which he found to rotate the polarisation of light – this is now known as optical activity – the mechanism is that the twisted bundles represented an artificial 3D-chiral structure. Many natural substances exhibit a similar effect, such as sugar solution in water, due to the fact that such molecules exist as two mirror images of each other called stereoisomers. The term for D-glucose is actually *dextrose*, which is in relation to the fact that this particular handedness of glucose rotates linearly polarised light to the right (dexter) side. Interestingly, these solutions also impart different absorptions of Left and Right circular polarisations of light, which is referred to as Circular Dichroism and is a widely used technique to establish the handedness of biological and naturally occurring compounds. By examining the

transmittances of both left and right circular polarisations, it is possible to distinguish

the handedness of the compound. This is vitally important for some pharmaceutical

products, as the biological function can depend upon the handedness. This was the

case in the "Thalidomide Tragedy" due to the fact that one of the enantiomers of

this chiral molecule caused birth-defects whilst the other was effectively used to

alleviate morning sickness [17].

The advent of wireless technology is perhaps the single-most important turning

point in the development of modern day communications. Many of these advances

are primarily due to the expansive research carried out on the use of antennas for

transmission and detection of electromagnetic waves. Seminal work carried out in the

late 19th century, following the theoretical framework laid down by Maxwell, led to

Hertz constructing a rudimentary version of wireless transmission with the use of a

capacitor and a primitive version of a dipole antenna [18,19]. Following this, Marconi

(beating Tesla, who is now equally accredited and given just as much, if not more,

recognition) demonstrated the transmission of radio-signals over the Atlantic Ocean,

which is essentially the birth of global communications; this led to the emergence of

the study of electromagnetism that it is today. Many of these applications, although

military in origin, have paved the way towards what we now refer to as

metamaterials, through the periodic structuring of resonant structures to control the

selection of frequencies of electromagnetic waves [20–23].

One of the fall-backs for many of these works, though, is the fact that they are primarily bulky, due to their reliance upon naturally occurring materials (glasses, crystals), and have little tunability to work at optical wavelengths, where the antenna sizes are required to be smaller than the width of a human hair, and also at Terahertz wavelengths where the lack of sensing and emitting THz devices are scarce. Due to the recent advancement in fabrication techniques and nanoscience this can be overcome, with an astounding acceleration in the fields of nanoscience and metamaterials occurring in the past decade; as with many fields of science and technology, scaling down the size of a device is preferable for both cost and applicability, and the fact that a lot can be done with a 'surface'.

1.2   Metamaterials & Metasurfaces

As with all fields of science, there was not one defining moment where a new and emerging field sprang into existence – this is also the case for metamaterials, which has been reliant upon the ever increasing bank of research and knowledge accumulated in the past 100 years or so. As mentioned previously, the emergence of wireless communication led to the study of antennas for emitting and detecting electromagnetic signals. Scaling these antennae down to sizes smaller than the wavelength of the light we are interested in as well as arranging them periodically has great advantages over isolated antennas, such as scalability, efficiency, and simplification of the underlying physical modelling.

Natural optical devices are used to control the wavefront of light through means of polarisation, phase, and amplitude modification. In classical optics, using naturally occurring materials, the behaviour of the light is determined by the subwavelength structuring (namely the atoms or molecules) composing the medium. Many effects can arise including refraction, reflection, and diffraction, where the phase, polarisation, and directionality of propagation can be controlled by the refractive index of the media. However, the properties of naturally occurring materials are typically limited to small deviations in such optical manipulations.

Metamaterials (where *meta* means '*beyond')* are essentially man-made devices used to behave similarly to naturally occurring media and to control the wavefront of light (or generally any type of wave-like phenomenon) through means of user-defined constituent 'atoms' – where an 'atom', or meta-atom, implies that the element is smaller than the wavelength of interest, such that an averaged effect is encountered and the light does not see the small deviations of the meta-atom structure; the scattering and interference of such waves are then homogeneously applied and an *effective* resultant wave is recovered. Such works have led to the development of negative refractive index materials [24–27], zero-index [28–30], invisibility cloaking [31–33], and sub-diffraction imaging [34–36] – all of which would be impossible to achieve using the optical properties of naturally occurring materials. However, regardless of the fascinating and novel physical phenomena provided by

such metamaterials, the bulky volumetric arrangement has many fabrication difficulties preventing the direct availability to making useful devices.

In more recent years, a keen interest has been bestowed upon 2D planar metamaterials, namely metasurfaces, due to the fact that they can provide similar phenomena as metamaterials, except that they are only a fraction of a wavelength thin, are much easier to fabricate, theoretically more trivial to realise, and the associated losses are negligible. Many optical devices have been realised using metasurfaces, most of which use abrupt phase changes to engineer the wavefront of light.

One of the most intriguing pieces of research was that carried out by Hasman [37–39] in using the concept of *Geometric phase* (or *Pancharatnam-Berry phase*) – which is the wavelength-insensitive phase modification of a transmitted wave dependent only upon the polarisation states traversed on the Poincaré sphere (*this is covered in greater detail in Chapter 2, section 2.4*). These works operated to produce a spatially varying phase modification of incident light through a simple means of rotating the planar optical elements, which in this case were subwavelength gratings. Their use of the *spatially-variant* phase control using Geometric phase was seminal work, where previous uses of geometric phase had been time-variant (and spatially-invariant, i.e. laterally isotropic media), such that the phase of the light was altered after completing a cyclic path around the Poincaré sphere by means of propagating

through optical devices (polarisers). This work was a key step on the ladder towards

the development of what we now refer to as metasurfaces.

Some years later, the concept of laterally abrupt phase-changes utilising

metasurfaces allowed a generalisation of Snell's law by incorporating a phase gradient

term, which was carried out by Capasso in [40]. This work used geometrically varying

V-shaped antennas arranged such that the individual phase of each one was

periodically increasing by a small amount. This resulted in the negative refraction of

light at an interface, work which had only been previously realised using

metamaterials. Following this work, and combining the functionality of geometric

phase, rather than antenna geometry, and phase-gradient metasurfaces, a plasmonic

flat lens was realised by Chen [41] which operated in the visible spectrum. This made

use of circular polarisation conversion, which traces out pole-to-pole arcs on the

Poincaré sphere, and results in a phase modification of the wavefront to obtain both

concave and convex lens functionality. It is known that dielectric media have

negligible absorption losses; to this end, and following on from the work carried out

on geometric-phase controlled lenses using metal nanostructures, a silicon nanofin

phase-gradient was proposed by Crozier [42] as a means of distinguishing the

handedness of incident circularly polarised light by diffracting the opposite

handednesses into different angles, which made use of the fact that left and right

circularly polarised light undergoes equal and opposite geometric phase

accumulations after interacting with a phase-gradient metasurface. In doing so, it was realised that high-efficiency metasurfaces can be realised and can behave in similar ways to those demonstrated using resonant metal structures.

From these works, it is understood that combining the functionality of geometric-phase along with metallic antennae and dielectric structures are a profound means of achieving many desirable optical phenomena, and have resulted in the potential real-world applications of lensing, beam-steering, beam-shaping, and holography; this forms the basis of this thesis, where silicon geometric-phase structures, phase gradient metallic antennae, and high-efficiency broadband holography devices are realised, all of which are dependent upon the phase-modification of the transmitted wavefront by the transverse spatially-variant abrupt phase changes (and dynamic time-variant smooth phase changes) as carried out in the seminal works explained above.

1.3   Thesis Plan and Overview

The field of research involved with electromagnetism is staggering. A particularly fresh and attractive topic during the past 15 years or so is that of metamaterials. This thesis aims to investigate planar metamaterials, namely metasurfaces, with potential applications for practical devices due to their compactness and novel properties.

Chapter 2 describes the theoretical and mathematical framework required in the development and understanding of metasurfaces. The classical optical physics required for the understanding of electromagnetic waves is discussed, including refraction, polarisation and Jones matrix formalism, which leads on to the function of media in the modification of light, with a focus on Fresnel's equations for a three-layer medium and anisotropy of crystals resulting in polarisation modifications. Structured subwavelength gratings are shown to exhibit similar, yet more pronounced effects to those of natural birefringent crystals. The concept of collective electron oscillations, termed plasmons, are shown to be applicable to resonant structures. Finally, the concept of geometric-phase is presented, where the phase of light is altered upon polarisation cycling, and this is applied to space-variant surface conditions to modify the phase of the wavefront.

Chapter 3 covers the experimental techniques used to fabricate and characterise metasurfaces. Some of these techniques include substrate cleaning and preparation, photoresist spincoating, photolithography, and development of these resists. We then move on to explain the process of reactive ion etching, which is a dry etch technique used to fabricate deep structures on a substrate (typically silicon), where the well-known Bosch process for very deep etching is covered. Next, metal deposition and lift-off process is explained, which is the method of manufacturing very small metallic structures atop a substrate. The final section of this chapter covers the

characterisation techniques used to retrieve both topographic and optical data for the sample. Optical microscopy and electron beam lithography are used to examine the sample, whilst surface profilometry is used to provide dimensional analysis of the sample structures. Terahertz optical examination is performed by a THz TDS system, where a basic explanation is provided.

Chapter 4 provides the theoretical and experimental verification of a monolithic silicon herringbone device, which was used to generate a strong circular conversion dichroism effect at 1THz. Subwavelength gratings were used to provide polarisation flipping between circularly polarised light, and arranged into a herringbone type pattern to incorporate geometric-phase. One of the angled sets of gratings were elevated to provide a fixed offset dynamic phase, whereby interference effects between the geometric and dynamic phases induce either destructive or unaffected interferences for opposite polarisation of CP light. A performance of 80% was theoretically realised, whilst the performance of the fabricated sample was reduced to 60% due to fabrication errors.

In Chapter 5 I discuss the development of a broadband metasurface array, composed of split ring resonators (SRRs), which are used to provide a simultaneous control of phase and amplitude of 1THz light to achieve complete control of the intensities of diffractive orders. A linear polarisation basis was used, where the SRRs had combined functionalities from rod-type antennas, to control amplitude dependent

upon orientation angle, and SRR-antenna geometry to control phase. Arrays of SRRs with smoothly varying amplitudes and phases were configured such that far-field diffraction orders were arbitrarily controlled. This also performed in a broadband manner.

In Chapter 6 I cover the development of high-efficiency polarisation conversion reflectarrays, which use the principle of geometric-phase and dynamic phase to result in dispersionless broadband operation. These elements were then utilised to form a holographic far-field image, capable of a window efficiency of 80%, and a zero-order of less than 5%.

Chapter 7 gives a summary of the previous chapters in this thesis, which touches on the key points acquired for each chapter. I then move on to describe how these metasurface designs covered in Chapters 4-6 could be utilised to provide polarisation sensing, or colour holography devices.

Chapter 2

Fundamental Concepts and Background

This chapter aims to provide the basis for the development of metasurfaces, or planar metamaterials, and the underlying mathematics used to describe the effects they exhibit. The majority of work undertaken in this thesis has a focus on wavefront modification of plane monochromatic waves owing to phase and polarisation control, with particular focus on the benefits imparted from geometric control by utilising Pancharatnam-Berry phase, which arises from the combination of the in-plane rotation angles of the metasurface antennas (or structures) along with circularly polarised light conversion. Additionally, interference effects between geometric and dynamic phases, where dynamic phase control is dependent on propagation depth, plays a role in the modification of the total phase of the wavefront.

2.1   Electromagnetism Basics

2.1.1   Maxwell's Equations

Light propagating in any arbitrary isotropic medium can be fully described by the collection of equations known as ***Maxwell's Equations,*** linking the electric and magnetic field disturbances, as given below:

$$\textbf{M1:} \qquad \nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \qquad (2.1)$$

$$\textbf{M2:} \qquad \nabla \times \boldsymbol{H} = \mathbf{J} + \frac{\partial \boldsymbol{D}}{\partial t} \qquad (2.2)$$

$$\textbf{M3:} \qquad \nabla \cdot \boldsymbol{D} = \rho \qquad (2.3)$$

$$\textit{M4:} \qquad \nabla \cdot \boldsymbol{B} = 0 \qquad (2.4)$$

with **J** being the current density, ρ being the charge density, and where the electric field **E**, electric displacement **D**, magnetic field **H**, and the magnetic flux density **B** can be described by the following relationships, known as the *constitutive equations:*

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \qquad (2.5)$$

$$\boldsymbol{B} = \mu \boldsymbol{H} \qquad (2.6)$$

with the terms ε, μ representing the electric permittivity and magnetic permeability, respectively. These are described by $\varepsilon = \varepsilon_r \varepsilon_0$ and $\mu = \mu_r \mu_0$; where ε_0 is the *permittivity* of vacuum (or free-space) equal to $8.85 \times 10^{-12} F/m$, μ_0 is the *permeability* of vacuum (or free-space) equal to $4\pi \times 10^{-7} H/m$, and ε_r, μ_r are the relative permittivity and permeability, respectively, of the medium. In (naturally-occurring) non-magnetic media, $\mu_r = 1$. Now, by utilising the vector triple-product

rule for the curl of the curl, $\mathbf{\nabla} \times (\mathbf{\nabla} \times \mathbf{A}) = \mathbf{\nabla}(\mathbf{\nabla} \cdot \mathbf{A}) - \mathbf{\nabla}^2 \mathbf{A}$, we apply the del

operator, $\mathbf{\nabla}$, to Equation 2.1 (M1), which gives us:

$$\mathbf{\nabla} \times (\mathbf{\nabla} \times \mathbf{E}) = \mathbf{\nabla}(\mathbf{\nabla} \cdot \mathbf{E}) - \mathbf{\nabla}^2 \mathbf{E} = \mathbf{\nabla} \times \left(-\frac{\partial \mathbf{B}}{\partial t}\right) = -\frac{\partial (\mathbf{\nabla} \times \mathbf{B})}{\partial t} \tag{2.7}$$

Now, in the absence of any free charges, ρ=0, and in the absence of any currents,

$\mathbf{J} = 0$. We know from equation 2.6 that $\mathbf{B} = \mu \mathbf{H}$, and when $\mu_r = 1$ this gives us

$\mathbf{B} = \mu_0 \mathbf{H}$, which we then substitute above into equation 2.7, to give us:

$$\mathbf{\nabla} \times (\mathbf{\nabla} \times \mathbf{E}) = -\frac{\partial (\mathbf{\nabla} \times \mathbf{B})}{\partial t} = -\mu_0 \frac{\partial (\mathbf{\nabla} \times \mathbf{H})}{\partial t} \tag{2.8}$$

An alternative way of writing M2 (equation 2.2) is by incorporating the constitutive

equation for **D** and realising that **J**=0:

$$\mathbf{\nabla} \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} = 0 + \varepsilon_r \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{2.9}$$

By substituting equation 2.9 into 2.8 we now have that:

$$\mathbf{\nabla} \times (\mathbf{\nabla} \times \mathbf{E}) = -\mu_0 \frac{\partial (\mathbf{\nabla} \times \mathbf{H})}{\partial t} = -\mu_0 \frac{\partial \left(\varepsilon_r \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right)}{\partial t} = -\mu_0 \varepsilon_r \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{2.10}$$

If we then look back and use the expanded form of equation 2.7, realising that the

term in $\mathbf{\nabla} \cdot \mathbf{E} = 0$ due to the absence of free charges (from M3) we can now have

that:

$$-\mathbf{\nabla}^2 \mathbf{E} = -\mu_0 \varepsilon_r \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{2.11}$$

This can be rewritten as:

$$\mathbf{\nabla}^2 \mathbf{E} = \frac{n^2}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = \frac{1}{v^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{2.12}$$

which is known as the characteristic *Wave Equation* of light in a medium, where we label $c = 1/\sqrt{\mu_0 \varepsilon_0} = 3 \times 10^8 \text{ms}^{-1}$, which is the speed of light in a vacuum, and $n = \sqrt{\varepsilon_r}$, which is the refractive index of the medium – a reminder is that the relationship between the velocity of a wave in a medium and in a vacuum is given by $c = n_i v_i$, with i indicating the medium. In the case of the light being in a vacuum, the value of $\varepsilon_r = 1$ and the wave equation in 2.12 reverts back to that for a vacuum with the speed of light simply being c.

2.1.2   Propagation and Polarisation

By examining the expression for the wave equation of the electric field in a medium, we can see that the solution that the **E**-field must take will be both dependent upon time and position. Additionally, by instead taking the curl of equation 2.2 (M2) we obtain the wave equation for the **H**-field in a medium [43]. This therefore implies that there exists a relationship between the **E**-field and the **H**-field of an electromagnetic wave, which is implied through the Maxwell's Equations themselves. Therefore, we can instead modify the wave equation in 2.12 to be general, and given as:

$$\nabla^2 \boldsymbol{U} = \frac{n^2}{c^2}\frac{\partial^2 \boldsymbol{U}}{\partial t^2} = \frac{1}{v^2}\frac{\partial^2 \boldsymbol{U}}{\partial t^2} \tag{2.13}$$

There must exist a solution for $\boldsymbol{U}$ which satisfies the above equation, and the simplest solutions are trigonometric functions:

$$U(r,t) = U_0 \sin(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t) \tag{2.14}$$

Or a more general way to write this is as an exponential, using Euler's Theorem:

$$U(r,t) = U_0 e^{i(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)} \tag{2.15}$$

which is the equation for a plane harmonic wave of light, where U_0 is the maximum amplitude of the wave, $\mathbf{k}$ is the wavevector (defined as $k = 2\pi/\lambda$, λ is the wavelength of light), $\mathbf{r}$ is the position vector ($\boldsymbol{r} = x\hat{\boldsymbol{e}}_x + y\hat{\boldsymbol{e}}_y + z\hat{\boldsymbol{e}}_z$ with the terms in $\hat{\boldsymbol{e}}_i$ representing the unit vector in the x,y,z directions respectively), ω is the angular frequency of the wave ($\omega = 2\pi/f$, where f is the frequency of the wave), and t is time.

Let us now perform some operations on the equation $U(r,t)$ to help provide some insight as to the transversality of the electric and magnetic fields of a propagating electromagnetic wave. Now, by taking the time derivative, $\partial/\partial t$, of $U(r,t)$, we extract the term $-i\omega$ to give us:

$$\frac{\partial U(r,t)}{\partial t} = \frac{\partial}{\partial t}[U_0 e^{i(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)}] = -i\omega U_0 e^{i(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)} = -i\omega U(r,t) \tag{2.16}$$

In a similar fashion, we can perform the del operation, $\boldsymbol{\nabla}$, to extract the term $i\boldsymbol{k}$, which gives us:

$$\boldsymbol{\nabla}U(r,t) = \boldsymbol{\nabla}U_0 e^{i(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)} = ikU_0 e^{i(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)} = ikU(r,t) \tag{2.17}$$

From these operations it can be implied that $\partial/\partial t \rightarrow -i\omega$ and $\boldsymbol{\nabla} \rightarrow \mathbf{ik}$. Now,

upon revisiting the Maxwell equations (M1-M4), for charge free and current free

media, we then obtain the following relationships:

$$\boldsymbol{k} \times \mathbf{E} = \omega\boldsymbol{B} = \omega\mu\boldsymbol{H} \tag{2.18a}$$

$$\boldsymbol{k} \times \mathbf{H} = -\omega\boldsymbol{D} = -\omega\varepsilon\boldsymbol{E} \tag{2.18b}$$

$$\boldsymbol{k} \cdot \mathbf{D} = \varepsilon\boldsymbol{k} \cdot \boldsymbol{E} = 0 \tag{2.18c}$$

$$\boldsymbol{k} \cdot \boldsymbol{B} = \mu\boldsymbol{k} \cdot \boldsymbol{H} = 0 \tag{2.18d}$$

It is clear that the above results show that ($\boldsymbol{E,D}$), ($\boldsymbol{H,B}$), and $\boldsymbol{k}$ are all mutually

orthogonal to each other and exist in the three orthogonal directions simultaneously.

This implies that the electric and magnetic fields are perpendicular to each other,

and both of these are perpendicular to the direction of travel represented by the

wavevector $\boldsymbol{k}$. This also implies that electromagnetic waves are transverse to the

direction of travel, and cannot be longitudinal.

When dealing with cross products relating to electric and magnetic fields and

the wavevector $\boldsymbol{k}$, a critical concept is that of the *Poynting vector*, which is given by

$$S = E \times H \tag{2.19}$$

and has the explicit definition of being the directional energy flux, or rate of

electromagnetic energy transfer per unit area, which acts in the direction of

propagation of the oscillating wave and is in the same direction as the wavevector $\boldsymbol{k}$

(in isotropic media).

For the case of plane harmonic waves, represented by $\boldsymbol{E}$ and $\boldsymbol{H}$ fields, with

the same form as that in equation 2.15, we can obtain the Poynting vector value for

$\boldsymbol{S}$ as being:

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H} = \boldsymbol{E_0} \times \boldsymbol{H_0} \cos^2(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t) \tag{2.20}$$

where we have taken the real part of equation 2.15 to give us cosines; one issue arises

– the cosine-squared function is oscillatory in amplitude, where a maximum of 1 and

a minimum of 0 occurs for each cycle of 2π, which means that the value of $\boldsymbol{S}$ is

constantly changing from a maximum to a minimum value. Now, the average value

of a cosine is equal to 0, but cosine squared has an average value of ½. Therefore, we

can conclude that the average value of the Poynting vector, $\langle\boldsymbol{S}\rangle$, is given as:

$$\langle S \rangle = \frac{1}{2} \boldsymbol{E_0} \times \boldsymbol{H_0} \tag{2.21}$$

and the average magnitude of the Poynting vector is simply ½E_0H_0, which is

equivalent to intensity being the square of the electric field amplitude.

If we imagine a wave travelling in the z-axis direction, we intuitively ascertain

that the electric and magnetic field vectors *must* lie perpendicular to the direction of

travel, and so lie in the *x-y* plane. These two field vectors must also be orthogonal to

each other. However, there is no constraint that either of these fields must be aligned

<u>with</u> the x or y axes, and can exist along any vector lying in the x-y plane, providing

that they remain perpendicular, and in turn define the ***polarisation*** of the

electromagnetic wave. It is typical in optics to label the direction of the polarisation

as being parallel to that of the electric field vector $\boldsymbol{E}$. Taking the general form of an

electromagnetic wave in Equation 2.15, and providing that the maximum amplitude

of the electromagnetic wave is constant ($U_0 = E_0 = const$), we label this wave as

being *linearly polarised* due to the fact that the electric field vector simply traces out

a straight line as it propagates through space, when looking along the direction of

propagation. Even though the electric field vector itself oscillates periodically, the

amplitude of this wave at any point in time will always exist on a straight line

between the two maximum points of amplitude $\pm E_0$. A diagram of this is shown in

Figure 2.1a.

Now, let us consider the case where two coherent linearly polarised waves

travel in the same direction (along the z direction), albeit orthogonal to each other.

If they are in phase, the resultant electric field vector would simply be the vector

sum of the individual polarisation vector, due to the principle of superposition, and

thus would be 45° to either of the polarised waves. This is apparent by writing

$$\text{Wave 1 - } \boldsymbol{E_1}(\boldsymbol{r}, \boldsymbol{t}) = \hat{e}_x E_0 \sin(kz - \omega t) \tag{2.22a}$$

$$\text{Wave 2 - } \boldsymbol{E_2}(\boldsymbol{r}, \boldsymbol{t}) = \hat{e}_y E_0 \sin(kz - \omega t) \tag{2.22b}$$

$$\text{TOTAL: } \boldsymbol{E} = \boldsymbol{E_1} + \boldsymbol{E_2} = (\hat{e}_x + \hat{e}_y) E_0 \sin(kz - \omega t) \tag{2.22c}$$

Now, if we propose that one of the waves is out of phase with the other, by a factor

of π/2, which corresponds to turning the sine function into a cosine, we have

$$\boldsymbol{E} = \boldsymbol{E_1} + \boldsymbol{E_2} = \hat{e}_x E_0 \sin(kz - \omega t) + \hat{e}_y E_0 \cos(kz - \omega t) \qquad (2.23)$$

This is still a valid solution of the wave equation 2.12, and corresponds to the

resultant vector of the polarisation tracing a circle as it propagates along z – this

wave will trace out the circle with a frequency of **ω**. Depending upon the sign of the

phase, and whether one of these linear polarisations leads or lags the other, it can

lead to the revolution of the resultant vector having a certain *handedness* – we refer

to these as being either *Left* or *Right* circularly polarised light (LCP or RCP), as

shown in **Figure 2.1b**. There can be two 'stances' when describing the handedness
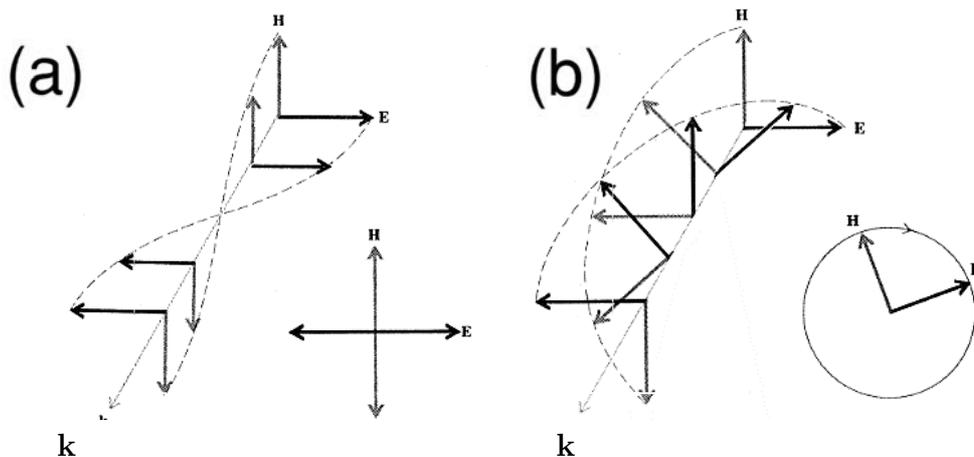
***Figure 2.1: Visualisation of both Linearly (a) and Circularly (b) polarised light (images from** [43])*

(a) For a linearly polarised wave, the electric field vector traces out a straight line between the maximum amplitude values as it propagates through space, where the magnetic field **H** is always perpendicular to the electric field **E**. **(b)** a circularly polarised wave is the vector sum of two orthogonal linearly polarised waves of equal amplitude, except where one of these waves is π/2 out of phase with the other. Thus, the tip of the resultant vector field traces out a circle as it propagates in space.

of a CP wave – either by looking at the rotation direction of the *incoming* wave or looking at the wave when it is *outgoing.* For this thesis, I shall use the latter, and describe a *right* handed wave as that which circulates in a clockwise direction when propagating *away* from me (where 'I' am the source).

The general solution to the superposition of two linearly polarised waves occurs for those which have arbitrary phases and differing amplitudes. This is a similar case to the circularly polarised light, except that now the resultant vector may not trace out a circular path, but an *elliptical* one instead. A diagram illustrating this is shown in **Figure 2.2.** The equation governing elliptically polarised light, without proof [44], is given by:

$$\frac{E_x^2}{E_{0x}^2} + \frac{E_y^2}{E_{0y}^2} - 2\frac{E_x}{E_{0x}}\frac{E_y}{E_{0y}}\cos\phi = \sin^2\phi \qquad (2.24)$$

 which is also known as the *polarisation ellipse*, where x and y correspond to the orthogonal electric field vector directions, the '0' subscript represents the maximum amplitude, and $\phi = \phi_y - \phi_x$ is the difference between the phases of the x and y polarised fields. It is apparent that if the phase difference $\phi = 0$, then equation 2.24 simply reduces to $E_x = E_y \frac{E_{0x}}{E_{0y}}$ which is simply linearly polarised light with the resultant vector being rotated by some angle depending on the fraction of the maximum amplitudes. And additionally, if $E_{0x} = E_{0y}$ then the above reduces to linearly polarised light at an angle of 45°. If we keep $E_{0x} = E_{0y}$ and instead now choose $\phi = \pi/2$, equation 2.24 reduces to $E_x^2 + E_y^2 = E_{0x}^2$ which is simply the equation
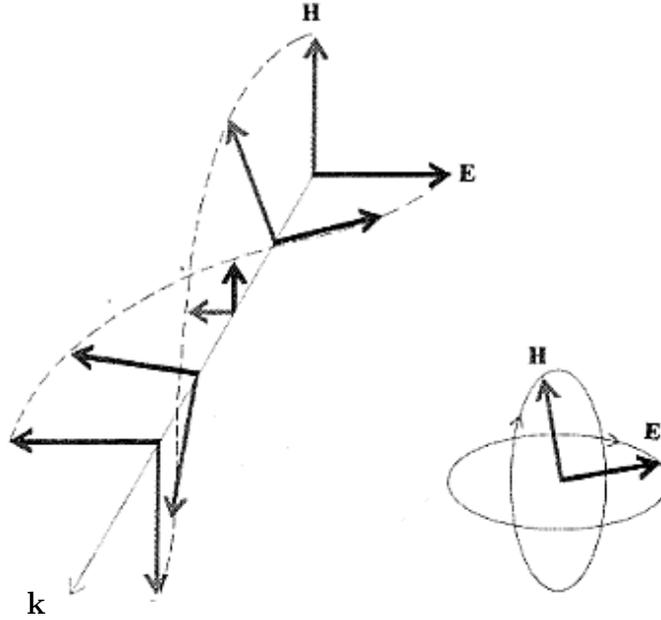
Figure 2.2: Visualisation of elliptically polarised light
(*image from* [43])

The ellipses are traced out due to the differing amplitudes and phases of the two electric (and magnetic) field vectors in their corresponding orthogonal directions.

of a circle with radius of E_{0x}, i.e. circularly polarised light. In the context of this thesis, elliptically polarised light will not be examined due to the metasurfaces covered later on operating only in either a linear (for Chapter 5) or circular (for Chapters 4 and 6) basis.

2.1.3   Jones Matrix Formalism

Suppose we have two linearly polarised plane waves propagating along the z-axis given by

$$E_x(z,t) = E_{0x}e^{i(kz-\omega t+\phi_x)} \tag{2.25a}$$

$$E_y(z,t) = E_{0y}e^{i(kz-\omega t+\phi_y)} \tag{2.25b}$$

We can ignore the *propagator* $(kz - \omega t)$, and then write the equations in a 2×1 column matrix for **E**, which is given as:

$$\boldsymbol{E} = \begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} E_{0x} e^{i\phi_x} \\ E_{0y} e^{i\phi_y} \end{pmatrix} \tag{2.26}$$

This is known as the *Jones Vector* representation of an electromagnetic wave, and is the general form for (elliptically) polarised light. It is custom to normalise the Jones Vector against its intensity, such that the intensity, *I,* is given by $I = \boldsymbol{E}^\dagger \boldsymbol{E}$, where the complex transpose of **E** is given as $\boldsymbol{E}^\dagger = (E_x^* \quad E_y^*)$, therefore, it is given that

$$\boldsymbol{E}^\dagger \boldsymbol{E} = I = \ E_x^* E_x + E_y^* E_y \tag{2.27}$$

And consequently, this then gives us

$$E_x^* E_x + E_y^* E_y = \ E_{0x}^2 + E_{0y}^2 = 1 = E_0^2 \tag{2.28}$$

where it is chosen that the total electric field amplitude squared, E_0^2, is normalised to be equal to unity (as is the intensity). For light polarised in only the *x*-axis, we have from Equation 2.28 that $E_{0x}^2 = 1$ (as $E_{0y} = 0$) and therefore the Jones vector for *x*-polarised light (linearly polarised in the *x*-direction) is given by:

$$\boldsymbol{E} = \begin{pmatrix} E_x \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{2.29a}$$

In a similar sense, for *y*-polarised light, we have:

$$\boldsymbol{E} = \begin{pmatrix} 0 \\ E_y \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{2.29b}$$

and for light polarised at $\pm 45°$ we have:

$$\boldsymbol{E} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} \tag{2.29c}$$

where the root factor of 2 comes from the fact that $E_x = \pm E_y$ and using Equation

2.27 gives $\Rightarrow 2E_{0x}^2 = 1$. Similarly, we can show the Jones vectors for right or left

circularly polarised light, given as

$$\textbf{RIGHT:} \qquad \boldsymbol{E} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ +i \end{pmatrix} \tag{2.30a}$$

$$\textbf{LEFT:} \qquad \boldsymbol{E} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix} \tag{2.30b}$$

where the +/- signs arise from the phase difference between the two linear

polarisations as $\phi = \phi_y - \phi_x = \pm \pi/2$, and therefore $exp(i\phi) = exp(\pm i\,\pi/2) = \pm i$.

Now, if we assume that these Jones vectors are passed through an optical

element, such as a polariser or a wave-plate, and that the emergent transmitted Jones

vector is simply a linear relation to the incident Jones vector, we have:

$$E_x' = T_{xx}E_x + T_{xy}E_y \tag{2.31a}$$

$$E_y' = T_{yx}E_x + T_{yy}E_y \tag{2.31b}$$

These can be rewritten in the Jones vector form, except now we have a matrix

composed of the T_{ij} terms:

$$\begin{pmatrix} E_x^{out} \\ E_y^{out} \end{pmatrix} = \begin{pmatrix} T_{xx} & T_{xy} \\ T_{yx} & T_{yy} \end{pmatrix}\begin{pmatrix} E_x^{in} \\ E_y^{in} \end{pmatrix} \tag{2.32}$$

The 2×2 matrix is termed the *Jones matrix* and is very useful for the later works

on metasurfaces, as this shows us that not only can a polarisation be transmitted through an element with the same polarisation but can also *induce* a polarisation that is orthogonal to it. It is worth noting that the subscripts of the Jones matrix components have an intrinsic relationship between the incident and transmitted polarisation; we have that the *ij* in T_{ij} correspond to (*i – Transmitted, j – Incident)* polarisation. Thus, as an example, T_{yx} is the conversion efficiency (of an optical device) between *x*-incidence and *y*-transmitted, and these Jones matrix components can be assumed to be the *Transmission coefficients* of a device.

As an analogy, it can be assumed that when working in the circular polarisation basis, equation 2.32 is changed to give:

$$\begin{pmatrix} E_R^{out} \\ E_L^{out} \end{pmatrix} = \begin{pmatrix} T_{RR} & T_{RL} \\ T_{LR} & T_{LL} \end{pmatrix} \begin{pmatrix} E_R^{in} \\ E_L^{in} \end{pmatrix} \tag{2.33}$$

However, it is sometimes experimentally complex and difficult to have both a circularly polarised input and measurable output, as more optical devices will be necessary (and it is not possible to measure circular polarisation directly using a detector). From [45] we can deduce the circular transmission coefficients in 2.33 simply from first obtaining the linear transmission components, where we have the following relationships:

$$T_{RR} = \left[T_{xx} + T_{yy} + i(T_{xy} - T_{yx}) \right]/2 \tag{2.34a}$$

$$T_{RL} = \left[T_{xx} - T_{yy} - i(T_{xy} + T_{yx}) \right]/2 \tag{2.34b}$$

$$T_{LR} = \left[T_{xx} - T_{yy} + i(T_{xy} + T_{yx})\right]/2 \tag{2.34c}$$

$$T_{LL} = \left[T_{xx} + T_{yy} - i(T_{xy} - T_{yx})\right]/2 \tag{2.34d}$$

Therefore, if we know the linear transmission coefficients of a device, we can predict what the operation of the device will be in a circular basis. This type of transformation is carried out for both experimental and simulation results.

2.1.4   Fresnel's Equations

As we saw at the start of this Chapter, the propagation of light at an arbitrary incident angle when encountering a boundary between two media of differing indices of refraction is governed by Snell's Law. This law works relatively well for unpolarised light, where it is understood that some of the light is refracted at an angle proportional to the $\sin^{-1}$ ratio of the refractive indices of the two media whilst the remaining light is reflected at an angle equal to that of the incident light. However, one problem that arises with this description is when the incident light is polarised; two special cases arise when the electric field vector (polarisation) is either parallel to the boundary (perpendicular to the plane of incidence), or parallel to the plane of incidence (with a magnetic field parallel to the boundary) – these two cases are referred to as *Transverse Electric* (electric field parallel to boundary) and *Transverse Magnetic* (magnetic field parallel to boundary) polarisations, termed TE and TM modes/polarised light. By employing boundary conditions at the interface between

two media, we state that **both the transverse electric and transverse magnetic**

fields must be continuous (derived from Ampère's Law). Combining this criteria

with Equations 2.18a-d, we then obtain equations for both TE and TM polarisations

(known alternatively as *'s'* and *'p'* polarisations, respectively, where *s* stands for

'senkrecht', meaning *perpendicular* in German, and *p* stands for parallel – these

definitions now refer instead to the plane of incidence, rather than the boundary, for

the polarisation direction) for a wave incident onto a boundary between medium 1

and 2 at an incident angle of θ_1 and transmitted angle of θ_2. For TE, we have:

$$E_i + E_r = E_t \tag{2.35a}$$

$$-H_i \cos \theta_1 + H_r \cos \theta_1 = -H_t \cos \theta_2 \tag{2.35b}$$

$$-k_i E_i \cos \theta_1 + k_r E_r \cos \theta_1 = -k_t E_t \cos \theta_2 \tag{2.35c}$$

And for TM polarisation we have:

$$H_i - H_r = H_t \tag{2.36a}$$

$$k_i E_i - k_r E_r = k_t E_t \tag{2.36b}$$

$$E_i \cos \theta_1 + E_r \cos \theta_1 = E_t \cos \theta_2 \tag{2.36c}$$

Now, if we declare that the reflection amplitude is simply the ratio between the

reflected and incident electric field amplitudes E_r/E_i, and the transmission amplitude

as being the ratio between the transmitted and incident amplitudes E_t/E_i, we can

rearrange the above equations to obtain the reflection and transmission amplitudes

for TE and TM polarised waves, respectively. For TE (subscript *s)*, using equations

2.35a,c, and eliminating E_t, we obtain:

$$r_s = \frac{E_r}{E_i} = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \qquad (2.37a)$$

$$t_s = \frac{E_t}{E_i} = \frac{2n_1 \cos\theta_1}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \qquad (2.37b)$$

And for TM (subscript *p*):

$$r_p = \frac{E_r}{E_i} = \frac{n_2 \cos\theta_1 - n_1 \cos\theta_2}{n_1 \cos\theta_2 + n_2 \cos\theta_1} \qquad (2.38a)$$

$$t_p = \frac{E_t}{E_i} = \frac{2n_1 \cos\theta_1}{n_2 \cos\theta_1 + n_1 \cos\theta_2} \qquad (2.38b)$$

We define the *reflectance* as being the fraction of incident energy reflected and is the

modulus-squared value of the reflection coefficients as follows:

$$R_s = |r_s|^2 = \left|\frac{E_r}{E_i}\right|^2_{TE} \qquad (2.39a)$$

$$R_p = |r_p|^2 = \left|\frac{E_r}{E_i}\right|^2_{TM} \qquad (2.39b)$$

An interesting case is when the incident angle is 0 (normal incidence) and then the

transmitted angle will also be 0, then the reflectance reduces to:

$$R = R_s = R_p = \left|\frac{n_1 - n_2}{n_1 + n_2}\right|^2 \qquad (2.40)$$

If we take the example of light incident from air to glass (using $n_{glass} = 1.5$), we get

that $R = 0.04$.

Now, for the transmittance, it should be slightly modified due to the fact that there must be a conservation of energy (if the media are non-absorbing, then R+T=1) and the light transmitted will travel at a different velocity depending on the refractive indices:

$$T_s = \frac{n_2 \cos\theta_2}{n_1 \cos\theta_1} |t_s|^2 = \frac{n_2 \cos\theta_2}{n_1 \cos\theta_1} \left|\frac{E_t}{E_i}\right|^2_{TE} \tag{2.41a}$$

$$T_p = \frac{n_2 \cos\theta_2}{n_1 \cos\theta_1} |t_p|^2 = \frac{n_2 \cos\theta_2}{n_1 \cos\theta_1} \left|\frac{E_t}{E_i}\right|^2_{TM} \tag{2.41b}$$

Again, a special case is for normal incidence where the angles are 0, and therefore the result for the transmittance is:

$$T = T_s = T_p = \frac{n_2}{n_1} \left|\frac{2n_1}{n_1+n_2}\right|^2 = \frac{4n_1 n_2}{(n_1+n_2)^2} \tag{2.42}$$

If we choose the case for light being transmitted from air to glass (n=1.5), then we have that: $T = 6/6.25 = 0.96$, which satisfies the energy conservation of T+R=1 for R=0.04, as calculated from equation 2.40.

In the modelling of metasurfaces, it is common for the system to be composed of three distinct layers (such as air/metasurface/substrate, with corresponding refractive indices n_1, n_2, n_3). It can be given [46], without derivation here, that the Fresnel equations are modified for a three-layer system to:

$$r = \frac{r_{12}+r_{23}e^{-2i\delta}}{1+r_{12}r_{23}e^{-2i\delta}} \tag{2.43a}$$

$$t = \frac{t_{12}t_{23}e^{-i\delta}}{1+r_{12}r_{23}e^{-2i\delta}} \tag{2.43b}$$

Where $\delta = \frac{2\pi d}{\lambda} n_2 \cos \theta_2$ (d is the thickness of the middle layer/metasurface layer

with corresponding refractive index n_2, θ_2 is the refracted angle between medium 1

and 2), r_{ij} is simply the normal two-layer Fresnel equation for reflection and can

exist as two forms, due to the incident polarisation being either TE or TM, where

the same occurs for t_{ij}. A simplification of this is for normal incidence, and so

$\delta = \frac{2\pi d}{\lambda} n_2$, $r_{ij} = (n_i - n_j)/(n_i + n_j)$ and $t_{ij} = (2n_i/(n_i + n_j))$, whilst equations 2.43

are unchanged.

2.2   Birefringence and Anisotropy

Up until now, we have dealt with refractive indices as being a constant value suitable for describing media, regardless of the direction or frequency of light propagating through them. In reality, however, most media are dispersive, anisotropic, and possibly inhomogeneous. Generally, the most interesting of these media will be ordered crystals, where the anisotropy is typically referred to as *birefringence* whereby light experiences two different refractive indices, corresponding to the *extraordinary axis* and the *ordinary axis* (which would be for two of the possible three orthogonal axes).

To be able to explain the effect of this, we will refer back to Maxwell's equations 2.5-2.8, except now we replace $\mathbf{D}$ with $\boldsymbol{D} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P}$, where $\mathbf{P}$ is the polarisability related to the volume density of electric dipoles, and replacing $\mathbf{B}$ with $\boldsymbol{B} = \mu_0\boldsymbol{H} + \boldsymbol{M}$, where $\mathbf{M}$ is the magnetisation related to the volume density of magnetic dipoles. We can relate the polarisability $\mathbf{P}$ to the electric field $\mathbf{E}$ by:

$$\boldsymbol{P} = \varepsilon_0\chi_e\boldsymbol{E} \tag{2.44}$$

where χ_e is called the e*lectric susceptibility* and is related to the relative permittivity $\chi_e = \varepsilon_r - 1$. Therefore, we can use this information to give the new definition of $\mathbf{D}$ as:

$$\boldsymbol{D} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P} = \varepsilon_0\boldsymbol{E} + \varepsilon_0\chi_e\boldsymbol{E} = \varepsilon_0(1 + \chi_e)\boldsymbol{E} = \varepsilon_0\varepsilon_r\boldsymbol{E} \tag{2.45}$$

this returns us back to the generic definition of the displacement current as given

earlier in this chapter. It is this susceptibility which governs the anisotropy of a

crystal, and in general is a tensor and not a scalar, which leads to:

$$\boldsymbol{P} = \varepsilon_0 \boldsymbol{\chi} \boldsymbol{E} \tag{2.46}$$

and in vector/matrix form, is given as:

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \varepsilon_0 \begin{bmatrix} \chi_{11} & 0 & 0 \\ 0 & \chi_{22} & 0 \\ 0 & 0 & \chi_{33} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{2.47}$$

Where the off-diagonal elements of the tensor are zero due to the crystal being treated

here as linear, uniform, lossless, and non-optically active. We saw earlier in this

chapter that the refractive index of a medium is related to its relative permittivity

by $n = \sqrt{\varepsilon_r}$. As we know that $\varepsilon_r = 1 + \chi$, we can then assign three refractive indices

to the crystal, corresponding to the three orthogonal direction, as:

$$n_1 = \sqrt{\varepsilon_{r1}} = \sqrt{1 + \chi_{11}} \tag{2.48a}$$

$$n_2 = \sqrt{\varepsilon_{r2}} = \sqrt{1 + \chi_{22}} \tag{2.48b}$$

$$n_3 = \sqrt{\varepsilon_{r3}} = \sqrt{1 + \chi_{33}} \tag{2.48c}$$

where these refractive indices are along the three principle axes. Now, if we apply

different constraints to the values of the tensor elements, we can classify media that

we may come across. In the case of *isotropic* media, such as air or vacuum, all of the

tensor elements χ_{ij} are equal, and as an example the tensor $\boldsymbol{\chi}$ is given below as:

$$\chi = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix} \tag{2.49}$$

with $\chi_{11} = \chi_{22} = \chi_{33} = a$, and $n_1 = n_2 = n_3 = \sqrt{1+a}$, where a is simply a constant.

A more special case exists where two of the tensor elements are equal and one differs, given as:

$$\chi = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix} \tag{2.50}$$

with $\chi_{11} = \chi_{22} = a$, $\chi_{33} = b$, where the refractive indices for χ_{11} and χ_{22} are referred to as being along the *ordinary* optical axes, for which $n_O = n_1 = n_2 = \sqrt{1+a}$, and the refractive index for χ_{33} as being along the *extraordinary* optical axis, for which $n_E = n_3 = \sqrt{1+b}$, (*noticing the subscripts 'O' and 'E' for the refractive indices corresponding to 'ordinary' and 'extraordinary', respectively*). This specific configuration is present in what is called a *uniaxial* crystal. However, there is no constraint as to which of the optical axis, O or E, has the greater value of refractive index. The typical labelling is that when $n_O < n_E$ it is termed a *positive* uniaxial crystal, whilst the opposite, for $n_O > n_E$, is termed a *negative* uniaxial crystal. **Table 2.1** below gives some refractive index values for both positive and negative uniaxial crystals. An alternative name for these crystals is *birefringent.* It is noticeable from the last column in the table that there is a finite value of the difference between the refractive indices in the ordinary and extraordinary axis directions. If we imagine a linearly polarised beam of light travelling perpendicular to both of these axis (so

Table 2.1: Refractive index values of Uniaxial media (λ=590nm)

Material	n_E	n_O	$\Delta n = n_E - n_O$
Calcite $(-)$	1.4864	1.6584	-0.172
Tourmaline $(-)$	1.638	1.669	-0.031
Beryl $(-)$	1.557	1.602	-0.045
Quartz $(+)$	1.5534	1.5443	$+0.009$
Ice $(+)$	1.313	1.309	$+0.004$
Rutile (TiO₂) $(+)$	2.903	2.616	$+0.287$

along one of the two ordinary axes) and the polarisation is along 45°, then we can

intuitively deduce that a phase delay will occur between the vertical and horizontal

resolved polarisation components due to the difference in refractive indices along the

optical axes. This phenomenon of birefringent crystals can be exploited to make many

devices, such as wave-plates, used to alter the polarisation state of incident light after

being transmitted. A further distance travelled by a polarised light beam through

such a crystal will result in an increased phase difference, which is governed by the

equation:

$$\Delta\varphi_{Dynamic} = \Delta n_{E-O} k d \tag{2.51}$$

where we refer to the phase difference $\Delta\varphi$ as being *dynamic*, $\Delta n_{E-O} = n_E - n_O$ is the

refractive index difference between the extraordinary and ordinary optical axis, $k = 2\pi/\lambda$ is the wavevector of the incident light, and d is simply the distance or thickness

of the crystal through which the light traverses. For example, if we wish to construct a half-wave plate (phase difference of **π)**, which converts the handedness of an incident circular polarisation into the opposite one, using calcite, we have:

$$d = \frac{\Delta\varphi_{Dynamic}}{\Delta n_{E-O}k} = \frac{\pi}{0.172(2\pi/\lambda)} = \frac{\lambda}{2\times0.172} \approx 3\lambda \qquad (2.52)$$

which implies that the thickness of the calcite must be 3 times larger than the free-space wavelength of the circularly polarised light we wish to flip; As the refractive index difference occurs at 590nm, this gives a distance of $d \approx 1.7\mu m$. This result means that the calcite crystal must be much larger than the wavelength we are interested in, and the fact we require very high-quality calcite crystals for use in optical experiments means they are very expensive and hard to come by. Additionally, as the crystals are undoubtedly dispersive, and so have wavelength-dependent refractive indices, a specific wave plate will only work for the wavelength it is fabricated for, making their usefulness deteriorate.

2.2.1   Form Birefringence and Subwavelength Gratings

As mentioned previously, we can construct useful devices using birefringent crystals. However, due to the fact they are larger than the wavelength in question; must be of a high-quality and therefore expensive; and only have limited use due to the constraint of the thickness-dependent phase and small refractive index differences; it is difficult to combine them with useful devices. One way around this

issue is to construct periodic gratings of alternating media to generate *form* birefringence. This is very desirable as the materials required need not even be birefringent themselves (or only very weakly birefringent) and so normal isotropic dielectric media can be used, which means they are much cheaper and more available than the high quality birefringent crystal devices. Instead of using alternating media, it is more easily obtainable in using a single dielectric medium and simply etching a periodic grating structure into it, such that the alternation is between air and dielectric (providing the refractive index of the dielectric is larger than air). One constraint is that the gratings are subwavelength, such that the periodicity is 'averaged' by the waves of incident light and is termed *effective medium theory (EMT).* The benefit of having these gratings subwavelength is that the roughness of the grating is seen as 'smooth' by the light, and so is treated as a continuous and homogeneous medium. Additionally, the differences between the refractive indices of the extraordinary and ordinary axes of such a subwavelength grating (SWG) can be much larger than naturally occurring birefringent media, and therefore exhibit much more pronounced optical effects.

A schematic diagram of a subwavelength grating is shown below in **Figure 2.3**. We assume that the periodicity of the grating must satisfy the following grating equation:

$$\Lambda \leq \frac{\lambda}{n_{Diel}} \tag{2.53}$$
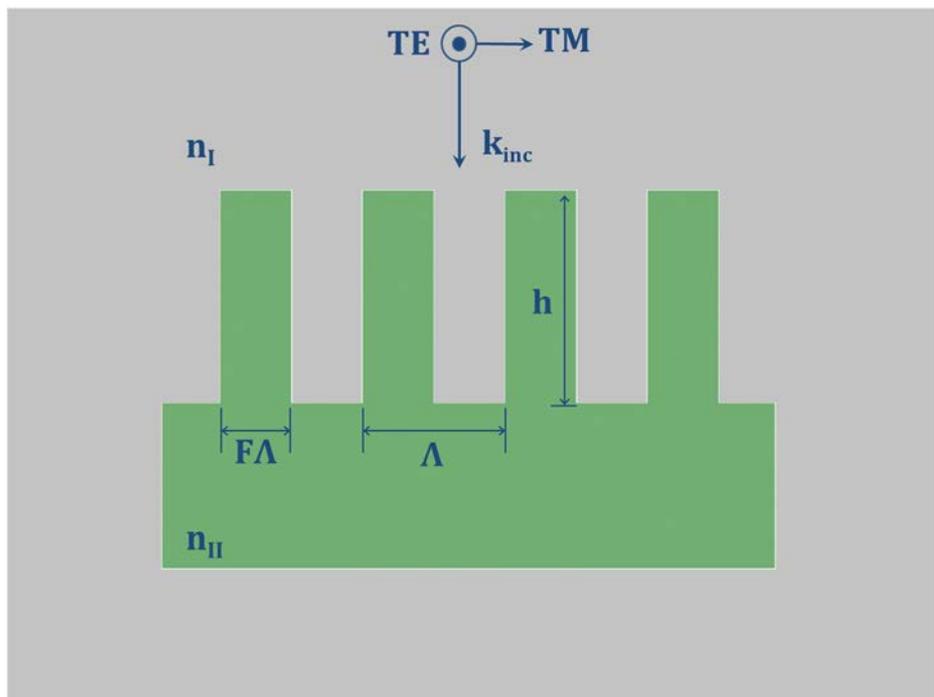
Figure 2.3: Basic diagram of a subwavelength grating

A subwavelength grating made from a dielectric substrate of refractive index n_{II} whilst the grooves consist of air (n_I). The filling factor is F, corresponding to the ratio of width of substrate to air, with a periodicity of Λ. The depth of the gratings is given by h. The directions of the polarisations are chosen as TE being parallel to the gratings whilst TM is perpendicular to the gratings.

in order for the grating to be homogeneous to the incident light (and therefore subwavelength) and assuming that the light is normally incident. In a similar expression as for the phase accumulation of light through a birefringent crystal in equation 2.51, the equation governing the phase for an SWG is given as:

$$\Delta\Phi_{TE-TM}(\lambda) = \left(\frac{2\pi h}{\lambda}\right)\Delta n_{form}(\lambda) \tag{2.54}$$

where the phase $\Delta\Phi$ is now dependent on the polarisation directions being TE and TM (analogous to the extraordinary and ordinary directions in crystals), h is simply

the depth of the gratings, and $\Delta n_{form}(\lambda) = n_{TE}(\lambda) - n_{TM}(\lambda)$ is the difference

between the refractive indices of the grating in the TE and TM directions.

If we wish to obtain these refractive indices, we must first derive the

expressions for the dielectric constants in the TE and TM directions of the grating.

We know from the start of this chapter that $\boldsymbol{D} = \varepsilon\boldsymbol{E}$, but because the dielectric

constant depends on the polarisation direction for the grating we instead use the

average **D** and **E** fields, to give us:

$$\varepsilon_{eff} \cong \frac{\boldsymbol{D}_{ave}}{\boldsymbol{E}_{ave}} \qquad\qquad (2.55)$$

where we treat the dielectric constant as being an *effective* constant as it depends on

the approximation of the **D** and **E** fields being averaged over a grating period. We

also know from the start of this chapter that the constraints of the electric field

polarisations are such that the components of the **E** field *parallel* to a boundary are

continuous, and the **D** field components *perpendicular* to a boundary are continuous.

Firstly, we will look at the case for TE polarised light, where the electric field

is parallel to the gratings as shown in **Figure 2.3**. We know that the parallel

component of an electric field is continuous across a boundary (which, in this case is

the grating tooth and air gap between). As such, we can assume that the TE polarised

electric field is simply $\boldsymbol{E}_{ave}$ as it is the same both in the grating and in the air gap.

Now, if we use equation 2.5, we have two equations for the **D** field in the two regions:

$$\boldsymbol{D}_I = \varepsilon_I \boldsymbol{E}_I = \varepsilon_I \boldsymbol{E}_{ave} \tag{2.56a}$$

$$\boldsymbol{D}_{II} = \varepsilon_{II} \boldsymbol{E}_{II} = \varepsilon_{II} \boldsymbol{E}_{ave} \tag{2.56b}$$

Because the averaged **D** field is not continuous across the boundary, we must

calculate the weighted average of the parallel **D** field components within each region

of the grating. This is carried out by using the fill factor F and the grating periodicity

Λ to give us:

$$\boldsymbol{D}_{ave} = \frac{F\Lambda \boldsymbol{D}_I + (\Lambda - F\Lambda)\boldsymbol{D}_{II}}{\Lambda} \tag{2.57}$$

Which then simplifies to:

$$\boldsymbol{D}_{ave} = F\boldsymbol{D}_I + (1 - F)\boldsymbol{D}_{II} \tag{2.58}$$

However, we have already calculated the expressions for the D fields in regions 1 (*I*)

and 2 (*II)* in equations 2.56a,b, which we can then substitute into equation 2.58 to

give us:

$$\boldsymbol{D}_{ave} = F\varepsilon_I \boldsymbol{E}_{ave} + (1 - F)\varepsilon_{II} \boldsymbol{E}_{ave} \tag{2.59}$$

If we look back at equation 2.55, the effective dielectric constant is defined as the

ratio between the averaged **D** and **E** fields in the grating. We achieve this by dividing

both sides of equation 2.59 through by $\boldsymbol{E}_{ave}$, to give us:

$$\frac{\boldsymbol{D}_{ave}}{\boldsymbol{E}_{ave}} \cong \varepsilon_{eff} = \varepsilon_{TE} = F\varepsilon_I + (1 - F)\varepsilon_{II} \tag{2.60}$$

This result is the *effective medium approximation* for the dielectric constant of a subwavelength grating for TE polarised light. In a similar fashion as above, we can derive the effective dielectric constant for a grating using TM polarised light instead. The difference in this case is that now the electric field is perpendicular to the gratings, and we then have to use the fact that the perpendicular **D** field is continuous across a boundary as opposed to the **E** field. This leads to the perpendicular **D** field occurring in the two regions as being $\boldsymbol{D}_{ave}$, and so we can use equation 2.5 to give us the equations of the electric field within the two regions as:

$$\boldsymbol{E}_I = \frac{\boldsymbol{D}_I}{\varepsilon_I} = \frac{\boldsymbol{D}_{ave}}{\varepsilon_I} \tag{2.61a}$$

$$\boldsymbol{E}_{II} = \frac{\boldsymbol{D}_{II}}{\varepsilon_{II}} = \frac{\boldsymbol{D}_{ave}}{\varepsilon_{II}} \tag{2.61b}$$

In a similar fashion as before, we must also calculate the weighted average of the perpendicular electric field in the grating, which again involves the use of the fill factor F and the grating period Λ. The equation is then identical to that for $\boldsymbol{D}_{ave}$ in equation 2.57 except we switch around the D's and E's, to give us:

$$\boldsymbol{E}_{ave} = F\boldsymbol{E}_I + (1 - \text{F})\boldsymbol{E}_{II} \tag{2.62}$$

Now, by substitution of the individual electric field expressions in regions 1 and 2, from equations 2.61a,b, into equation 2.62, we obtain:

$$\boldsymbol{E}_{ave} = F\frac{\boldsymbol{D}_{ave}}{\varepsilon_I} + (1 - \text{F})\frac{\boldsymbol{D}_{ave}}{\varepsilon_{II}} \tag{2.63}$$

Dividing both sides through by $\boldsymbol{D}_{ave}$ then gives us:

$$\frac{E_{ave}}{D_{ave}} \cong \frac{1}{\varepsilon_{eff}} = \frac{1}{\varepsilon_{TM}} = F\frac{1}{\varepsilon_I} + (1 - \text{F})\frac{1}{\varepsilon_{II}} \tag{2.64}$$

With a bit of tidying, we then have the effective dielectric constant of a subwavelength grating for TM polarised light as being:

$$\varepsilon_{TM}^{-1} = F\varepsilon_I^{-1} + (1 - \text{F})\varepsilon_{II}^{-1} \tag{2.65}$$

We know from earlier in the chapter than $n = \sqrt{\varepsilon_r}$, so we can then obtain the effective refractive indices for TE and TM polarisations by taking the square root of equations 2.60 and 2.65 to give us:

$$n_{TE} = \sqrt{Fn_I^2 + (1 - F)n_{II}^2} \tag{2.66a}$$

$$n_{TM} = \frac{1}{\sqrt{\frac{F}{n_I^2} + \frac{(1-F)}{n_{II}^2}}} \tag{2.66b}$$

where F is simply the filling factor (the ratio between width of the grating teeth and gap) and n_I, n_{II} are the refractive indices of the air and substrate, respectively.

As an example, with reference to the application of SWGs in Chapter 4, we have a substrate of silicon ($n_{II} = n_{\text{Si}} = 3.418$) and use a wavelength of 300μm (corresponding to a frequency of 1THz). Putting these values into equation 2.53 gives us a periodicity of 87.7μm, and so with the constraint that the periodicity must be less than this value a grating period of 86μm was chosen. Keeping the filling fraction as 0.5 for ease of use, we obtain values for the refractive indices parallel and perpendicular to the grating, with $n_{TE} = 2.52$ and $n_{TM} = 1.36$, and an index

difference of $\Delta n_{form} = 1.16$, which is approximately ~7 times the refractive index

difference for calcite (albeit at a different wavelength). Again, if we choose a value

of phase such that the SWG functions as a half-wave plate $(\Delta\Phi = \pi)$, then putting

this value and the value for Δn_{form} into equation 2.54 we obtain a grating depth, h,

of:

$$h = \frac{\lambda}{2\pi}\frac{\Delta\Phi_{TE-TM}}{\Delta n_{form}} = \frac{\lambda}{2\times1.16} \approx 0.4\lambda \qquad (2.67)$$

and using our previous choice of wavelength of 300μm we have $h \approx 129\mu m$. It is clear

that such a grating does indeed function in the subwavelength regime, and therefore

can be treated as an effective homogeneous medium.

2.3   Excitations at Metal-Dielectric boundaries

2.3.1   Dispersion

If we refer back to section 2.1, we have the modified macroscopic Maxwell Equations. Now, considering the case for nonmagnetic, electrically-neutral media, we can set the magnetisation $\mathbf{M}$ and free charge density ρ to zero. The general wave equation which governs the Maxwell equations containing terms of polarisation $\mathbf{P}$ and current density $\mathbf{J}$ is then given by:

$$\nabla^2 \boldsymbol{E} + \frac{1}{c^2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = -\mu_0 \frac{\partial^2 \boldsymbol{P}}{\partial t^2} - \mu_0 \frac{\partial \boldsymbol{J}}{\partial t} \tag{2.68}$$

where the terms on the right hand side are of great importance. The term in $\mathbf{P}$ relates to the polarisation charges in a medium, whilst the term in $\mathbf{J}$ relates to the conduction charges in a medium. In non-conducting media, it is the polarisation term $\mathbf{P}$ which is dominant, and plays a key role in the explanation of physical properties such as dispersion and absorption, whilst for conducting media it is now the current density $\mathbf{J}$ which is the dominant factor, and explains the high absorption and large opacity of metals.

We now look more closely at dielectric media, which are nonconducting. In these, electrons are not free but bound to the constituent atoms of the medium without preferential direction (isotropic), such as glass. If we assume that each of these bound electrons, with a charge of *-e*, can be displaced from its equilibrium

position by a distance **r**, then the resulting macroscopic polarisation, given by **P**, for

such a medium is simply:

$$\boldsymbol{P} = -Ne\boldsymbol{r} \tag{2.69}$$

with N is the electron density (number of electrons per unit volume). If we now

assume that the displacement of these electrons is driven by an external (static) field

E, then the force equation is given by:

$$-e\boldsymbol{E} = K\boldsymbol{r} \tag{2.59}$$

(where K is the force constant) and substituting equation 2.59 into 2.58, we obtain

the static polarisation:

$$\boldsymbol{P} = \frac{Ne^2}{K} \tag{2.60}$$

This expression is only valid for static electric fields, however, so for an electric field

which varies in time we must modify equation 2.59 to give the differential equation

of motion as:

$$-e\boldsymbol{E} = m\frac{d^2\boldsymbol{r}}{dt^2} + m\gamma\frac{d\boldsymbol{r}}{dt} + K\boldsymbol{r} \tag{2.61}$$

where the term in $m\gamma$ corresponds to a frictional-damping force, which is related to

the velocity of the electrons and arises from electron-ion collisions. If we now assume

that the applied electric field is harmonic and oscillates in time with a form $e^{-i\omega t}$

then by employing the simplification of derivatives in equations 2.22, we simply

extract the term in ω and equation 2.61 reduces to:

$$-e\boldsymbol{E} = (-m\omega^2 - i\omega m\gamma + K)\boldsymbol{r} \tag{2.62}$$

Substituting this into the original expression for polarisation in equation 2.58, we have:

$$\boldsymbol{P} = \frac{Ne^2/m}{\omega_0{}^2 - \omega^2 - i\omega\gamma}\boldsymbol{E} \tag{2.63}$$

where we have set $\omega_0 = \sqrt{K/m}$, which is the effective frequency of the electrons that are bound. Equation 2.63 is simply analogous to the equation for a driven harmonic oscillator, for which the solution is described by a *Lorentzian* resonance, and therefore the result will provide some type of resonance condition corresponding to the intrinsic resonance frequency ω_0 – this is intuitive from the bound electrons being elastically driven about an equilibrium point.

If we now substitute equation 2.63 back into the wave equation in 2.57, and removing the term in **J** due to the medium being dielectric, we can then see that a solution of this is:

$$\boldsymbol{E} = \boldsymbol{E_0}e^{i(\tilde{k}z - \omega t)} \tag{2.64}$$

and substitution of this into the wave equation yields us with an expression for the wavevector $\tilde{k}$:

$$\tilde{k}^2 = \frac{\omega^2}{c^2}\left(1 + \frac{Ne^2}{m\varepsilon_0} \cdot \frac{1}{\omega_0{}^2 - \omega^2 - i\omega\gamma}\right) \tag{2.65}$$

The presence of i in the denominator implies that $\tilde{k}$ is a complex function, and can be represented by

$$\tilde{k} = k' + ik'' \tag{2.66}$$

and is directly translatable to writing a complex refractive index:

$$\tilde{n} = n + i\kappa \tag{2.67}$$

where we use that $\tilde{k} = \tilde{n}\omega/c$. This gives us a general expression for the oscillating

harmonic electric field in 2.64 as:

$$\boldsymbol{E} = \boldsymbol{E_0} e^{i(k'z - \omega t)} e^{-k''z} \tag{2.68}$$

Where the $\exp(-k''z)$ term implies that the wave is *physically* decaying in amplitude

with increasing distance into the medium, and is explained through the process of

absorption of the applied electromagnetic wave.

2.3.2   Plasmons

2.3.2.1   *Bulk Plasmons*

We now look at a similar case for metals, where now it is the **J** term which

dominates, and we can ignore the polarisation **P.** In a similar manner, we derive the

differential equation of motion for the conduction electrons as:

$$-e\boldsymbol{E} = \frac{m}{\tau}\boldsymbol{v} + m\frac{d\boldsymbol{v}}{dt} \tag{2.69}$$

where the velocity of the electron is $\boldsymbol{v}$. The new frictional constant is given by m/τ,

and relates to the static conductivity (where the static conductivity is σ in **J** = σ**E**).

We set the current density as $\boldsymbol{J} = -Ne\boldsymbol{v}$, and consequently can present equation 2.69

in terms of **J** rather than $\boldsymbol{v}$. Solving this **J** formalism of equation 2.69 for a static

electric field (and therefore, a static **J** from **J** = σ**E**; hence, no term in d**J**/dt), we

obtain the expression:

$$\mathbf{J} = \frac{Ne^2}{m}\tau\boldsymbol{E} \qquad\qquad (2.70)$$

thereby giving us $\sigma = (Ne^2/m)\tau$. If we now presume that both the electric field (and

current density) have harmonic time dependence of $e^{-i\omega t}$, then we arrive at the result

(without explicit derivation) that:

$$\mathbf{J} = \frac{\sigma}{1-i\omega\tau}\boldsymbol{E} \qquad\qquad (2.71)$$

Again, substituting this into the general wave equation in 2.57, we obtain a complex

solution for $\tilde{k}$ given by:

$$\tilde{k}^2 \approx i\omega\mu_0\sigma \qquad\qquad (2.72)$$

(where we have assumed an approximation for dealing with very low frequencies). In

its complex form, we have $\tilde{k} = (1 + i)\sqrt{\omega\mu_0\sigma/2}$, and the real and imaginary parts of

$\tilde{k}$ are equal, such that $k' \approx k'' \approx \sqrt{\omega\mu_0\sigma/2}$. (Similarly, for the complex refractive

index $\tilde{n}$ we have $n \approx \kappa \approx \sqrt{\sigma/2\omega\varepsilon_0}$.) We now introduce the concept of the *skin depth*,

which is defined as $\delta = 1/k'' = \sqrt{2/\omega\mu_0\sigma} = \sqrt{\lambda_0/c\pi\mu_0\sigma}$, where λ_0 is simply the free

space wavelength of the incident electric field. For a high conductivity, which infers

a good conductor, we obtain a high absorbance, k'', such that the inverse leads to a

small skin depth δ; this result is concurrent with the everyday intuitive experience

of metals being highly opaque.

If we don't assume the approximation for low frequencies (in equation 2.72),

and using the relationship $\tilde{k} = \tilde{n}\omega/c$, we express the complex squared refractive index

as:

$$\tilde{n}^2 = \tilde{\varepsilon} = 1 - \frac{\omega_p{}^2}{\omega^2 + i\omega\gamma} \tag{2.73}$$

This is a very important result, namely the *dielectric function* of a metal, and is well

known as the *Drude Model*, which relates the complex permittivity of a medium to

its plasma frequency, ω_p, and its damping, γ, dependent upon the frequency of the

incident light (where $\gamma = \tau^{-1}$, and $\omega_p = \sqrt{Ne^2/m\varepsilon_0} = \sqrt{\mu_0\sigma\gamma c^2}$ is the intrinsic

plasma frequency of the medium). The plasma frequency is the term used to describe

the free conduction electrons as being akin to a plasma. For realistic modelling of

metals, such as gold and silver (widely used in plasmonic systems) the Drude model

must be modified to:

$$\tilde{\varepsilon} = \varepsilon_\infty - \frac{\omega_p{}^2}{\omega^2 + i\omega\gamma} \tag{2.74}$$

where the term in ε_∞ is a constant, akin to a 'bias', depending on the metal

investigated, in order to provide the correct offset; this is due to the highly polarised

environment of the bound electrons $\mathbf{P}$ (which was previously neglected).

If we continue looking at *ideal* metals (i.e. no term in ε_∞) and for large frequencies, for which the damping term is negligible such that $\omega \gg \gamma$, then we can simplify equation 2.73 to being:

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega^2} \tag{2.75}$$

due to the fact that the term in $i\omega\gamma$ will be negligibly small compared to ω^2 (where we omit the tilde to signify that the quantities are no longer complex). Using the knowledge that we can relate k, n, and ε as being:

$$k^2 = \frac{n^2\omega^2}{c^2} = \frac{\varepsilon\omega^2}{c^2} \tag{2.76}$$

and substitution of 2.75 into 2.76, we obtain the *dispersion relation* of such a bulk metal, at large frequencies, giving us:

$$\omega^2 = \omega_p^2 + c^2 k^2 \tag{2.77}$$

We can intuitively see that the relationship between ω and k is offset by the term in ω_p, and by plotting a graph of ω/ω_p against ck/ω_p, as shown below in **Figure 2.4**, it is clear that there are no propagating electromagnetic waves inside a metal for $\omega < \omega_p$ which confirms the phenomenon of metals being opaque for low frequencies. An important conclusion can be shown for the case where $\omega = \omega_p$ such that $\varepsilon(\omega_p) = 0$ for equation 2.75. As derived in [47] there will then exist a solution to the wave-equation for which the vector product $\boldsymbol{K} \cdot \boldsymbol{E} \neq 0$ (where this dot product is typically zero for transverse electromagnetic waves), and is satisfied by $\varepsilon(\boldsymbol{K}, \omega) = 0$. This is
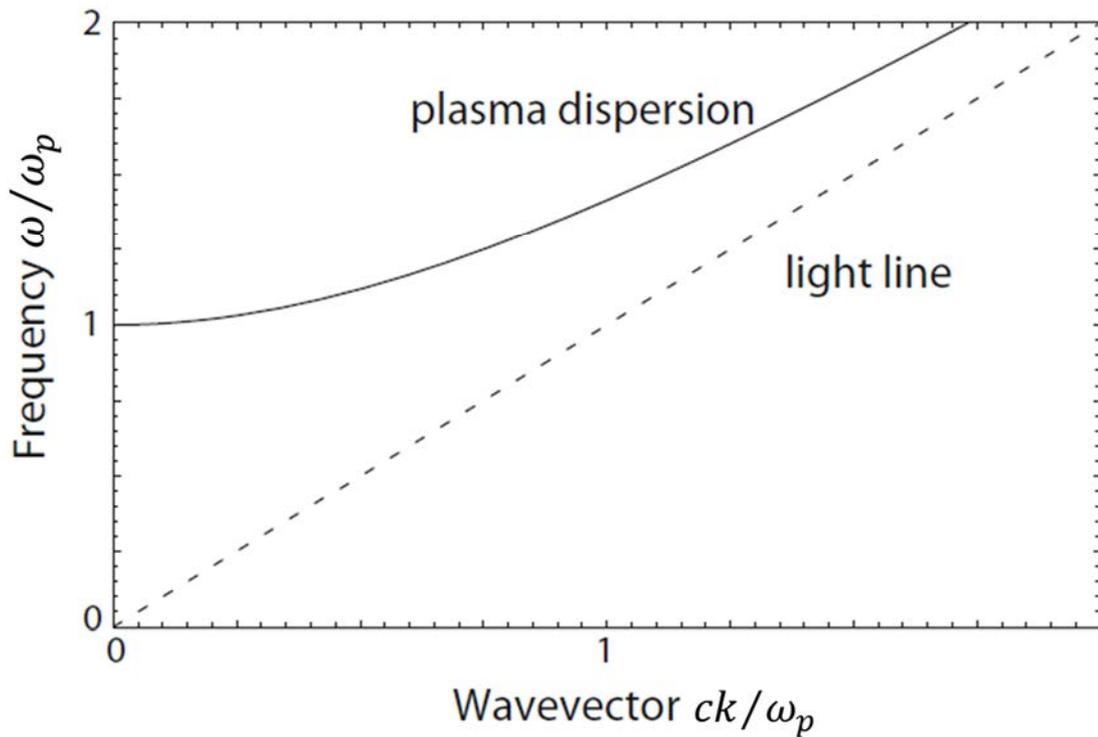
Figure 2.4: Dispersion relation of metals with a free-electron gas/plasma described by the Drude Model [47]

The relationship between ω and k as given in Eq. 2.77 - It is not possible for electromagnetic waves to propagate through a metal in the frequency regime of $\omega < \omega_p$, therefore implying strong opacity and reflection at the metal surface for low frequencies. Additionally, because the dispersion curve and light line do not cross, it is not possible to directly couple propagating electromagnetic radiation to excite a *Bulk Plasmon.*

equivalent to the existence of a longitudinal mode for the collective free electron plasma resonance, which have quantised modes referred to as *Bulk Plasmons.* These modes are longitudinal, and hence cannot couple to incident transverse electromagnetic waves, as illustrated by the fact that the metal dispersion curve in Figure 2.4 never crosses the dispersion for electromagnetic waves in free space (light line). For the sake of this thesis, bulk plasmons will not be explained in any greater depth; the reader can find further information on these in [47].

2.3.2.2   Surface Plasmon Polaritons

We now investigate the phenomenon of a *Surface Plasmon*, or S*urface Plasmon Polariton* (SPP), which is a mode that arises when the k-vector is matched on the boundary between a metal and dielectric, and a propagating surface wave is produced. Using Maxwell's equations from earlier in the chapter, and equating boundary conditions between a dielectric and metal (such that the tangential components of **E** and **H** are continuous and equal), we arrive at the dispersion relation (without rigorous derivation, further details can be found in [47]) between the permittivities of the metal/dielectric and the wavevector parallel to the boundary as:

$$\frac{k_1}{\varepsilon_1} + \frac{k_2}{\varepsilon_2} = 0 \tag{2.78}$$

where the terms in k_i (*i*=1,2 corresponding to either media*)* correspond to the wavevector perpendicular to the 2D metal-dielectric boundary, and is the real physical decay of the electromagnetic wave away from the boundary. Now, an additional equation linking the wavevectors and permittivities is obtained from the wave equation in both the metal and dielectric medium, which gives us:

$$k_i{}^2 = k_x{}^2 - k_0{}^2\varepsilon_i \tag{2.79}$$

where we label k_x as being the wavevector parallel to the boundary (so in-plane) and $k_0 = \omega/c$ is the wavevector of the impinging free-space wave. Upon combining

equations 2.78 and 2.79 (and with a lengthy algebraic derivation), we obtain the

dispersion relation of a surface plasmon polariton as given below:

$$k_x = k_0 \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}} \tag{2.80}$$

where we now use the subscripts m and d to signify the metal and dielectric

permittivities respectively. We can obtain the associated surface plasmon wavelength

by using $\lambda_{sp} = 2\pi / k_x$. This result is only valid for light which is TM polarised; SPP's

can only be excited by TM polarisation and not TE polarisation, due to the fact that

TM polarisation has a component of its electric field parallel to a component of the

incident k-vector (both of these components are tangent to the boundary) allowing

longitudinal wave excitation where $E \cdot K \neq 0$, and also has a component

perpendicular to the boundary between the metal and dielectric –which is necessary

for the excitation of charges having fields extending into either media – and thus

supply the necessary dispersion conditions to excite a plasmon. The dispersion curve,

ω vs k_x, for an SPP is given below in **Figure 2.5.** For small frequencies,

corresponding to small k_x, the dispersion curve is relatively linear. However, for much

larger values of k_x the response is asymptotic; this result is achieved by substitution

of equation 2.75 into 2.80, and by setting the dielectric medium to have a refractive

index of n=1 (hence, $\varepsilon_d = 1$) we see that the dispersion

tends to $\omega = \omega_p / \sqrt{2}$ (where more generally it is given as $\omega = \omega_p / \sqrt{1 + \varepsilon_d}$). This is

shown by the blue dotted line in Figure 2.5.

It can be seen from Figure 2.5 that the SPP dispersion does not cross the light

line for air/vacuum (given by the black dashed line). This infers that it is not possible

to directly couple free space light into a metal to excite an SPP. To this end, we

must use a dielectric with a refractive index greater than 1 in order to provide the

necessary matching conditions between the wavevectors along the boundary, as

demonstrated by the red dashed line corresponding to the modification of incident

light when a dielectric of glass is used next to the metal. This is the typical method
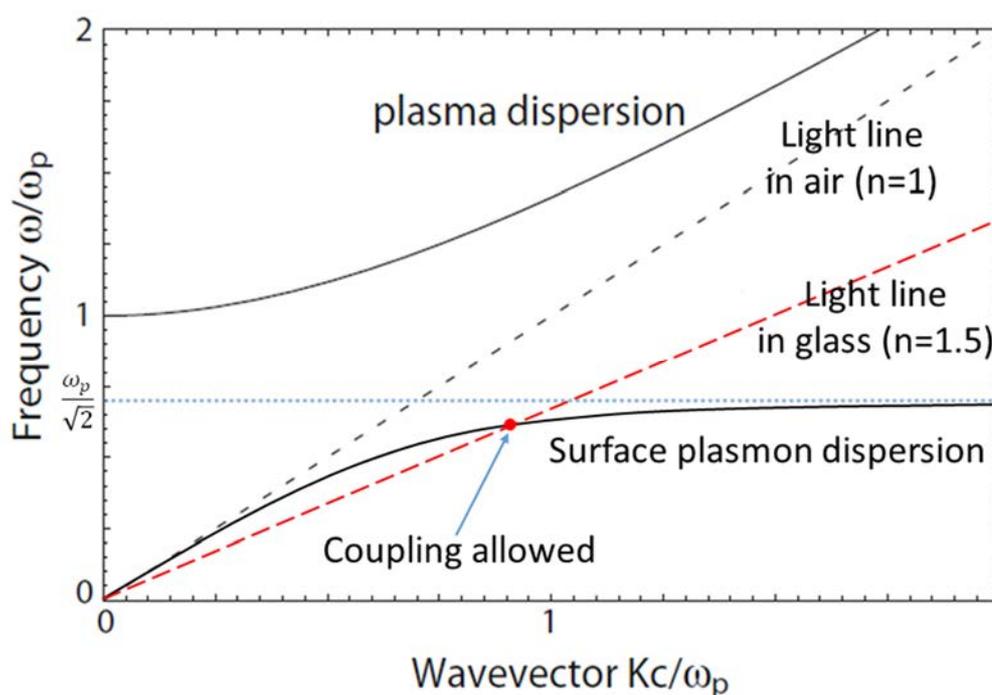
***Figure 2.5: Dispersion curves of both bulk and surface
plasmon polariton modes (figure modified from* [47])**

The dispersion for bulk, given in equation 2.77, is above the light line, whilst
the surface plasmon polariton dispersion, given in equation 2.80, lies below
the light line. This SPP dispersion is relatively linear at small frequencies
(small k_x), whilst at much larger k_x tends to the asymptote of $\omega_p/\sqrt{2}$ (given
by the dotted blue line). It is forbidden to couple free space light into a bulk
plasmon; however, it is possible to couple light into an SPP due to the fact
that we can obtain a 'crossing' point for a modified light line. This is achieved
by using a dielectric with an $\varepsilon_d > 1$ (such as glass), where the red dashed line
signifies the modification of propagating light and the allowed coupling.

used to investigate SPPs directly, for which the most popular setup is that of the

Kretschmann configuration [48], where a thin metal film is deposited onto a glass

prism, and the dispersion relation plotted dependent upon the angle of incidence

(which in turn affects the amount of light with wavevector parallel to the boundary).

More information on surface plasmon polariton coupling and experimental procedures

can be found in references [47–50].

2.3.2.3   *Localised Surface Plasmon Resonance*

We have investigated the phenomenon of surface plasmon polaritons, which are

propagating collective oscillations of the electrons at the boundary between a

dielectric and metal. In this regime, it is assumed that the boundary is that between

semi-infinite bulk media; however, this does not hold true when we deal with metallic

particles which are sub-wavelength in size. This is due to the fact that for such

particles, much smaller in comparison to the wavelength, upon illumination of an

oscillating harmonic electromagnetic wave the phase is considered to be constant over

the particle. This is referred to as the *quasi-static approximation* and we can assume

that the particle is in a static electric field.

In the simplest case, we deal with isotropic and homogeneous sub-wavelength

spherical nanoparticles, having a radius of a (where $a \ll \lambda$) and relative permittivity

of ε_a, which are placed into a static electric field and surrounding medium of

permittivity ε_d. One can then solve for the electric dipole moment of such a particle, as outlined in [47], yielding:

$$\boldsymbol{p} = 4\pi\varepsilon_d\varepsilon_0 a^3 \frac{\varepsilon_a - \varepsilon_d}{\varepsilon_a + 2\varepsilon_d} \boldsymbol{E} \qquad (2.81)$$

If we use the fact that the dipole moment can be expressed as $\boldsymbol{p} = \varepsilon_d\varepsilon_0\alpha\boldsymbol{E}$, we describe the term $\boldsymbol{\alpha}$ as being the *polarisability* of the sphere, given by

$$\alpha = 4\pi a^3 \frac{\varepsilon_a - \varepsilon_d}{\varepsilon_a + 2\varepsilon_d} \qquad (2.82)$$

It can be realised that when the denominator of equation 2.82 tends to zero the value of the polarisability $\boldsymbol{\alpha}$, and hence the dipole moment $\boldsymbol{p}$, tends to infinity; this corresponds to a strong absorption of the incident electromagnetic radiation and results in a resonance condition. This is explained in a similar means to the case for surface plasmon resonance, except that now the localised collection of the free electrons are affected by the (mostly curved) geometry of such small structures; this allows direct excitation of the electrons about the fixed ionic cores without the special requirement for wavevector matching as for SPPs. We term these local oscillations as *Localised Surface Plasmons* (LSPs). The condition for resonance is satisfied when $Re[\varepsilon_a] = -2\varepsilon_d$ and the denominator tends to zero (or in the case for real metals, when this value is a minimum). This is called the Fröhlich condition, and upon equating the permittivities with that for a metal described by equation 2.75 we obtain that the resonance frequency of an LSP is $\omega = \omega_p/\sqrt{3}$ where the surrounding dielectric media is chosen to be vacuum/air ($\varepsilon_d = 1$). Another noticeable observation

from the Fröhlich condition of $Re[\varepsilon_a] = -2\varepsilon_d$ and also equation 2.75, is that the dielectric medium in which the small nanoparticle is present has a strong influence upon its resonance frequency; an increase in the dielectric constant (or alternatively the refractive index) will lead to a red-shift of the LSP resonant frequency (red-shift implies an increase in wavelength, decrease in frequency). This is a profound result and is the reason that LSP systems have great applicability for sensing refractive index changes of the surrounding media. Additionally, in the derivation leading to equation 2.81 (from [47]) it can be seen that the electric field is simply the negative gradient of the potential. This infers that the electric field is also resonantly dependent upon the Fröhlich condition and therefore results in strong electric field enhancement within and outside the subwavelength particle, which also has interesting applications.

In the case for most metamaterials and plasmonics applications, spherical particles are rarely used due to their symmetric, and therefore isotropic, response to incident light and are undesirable as most metamaterials often rely on polarisation induced phase and amplitude effects. Instead, we look at the case for subwavelength ellipsoidal particles that have three distinct axis dimensions, which we can label a_x, a_y, a_z along the *x-, y-,* and *z*-axis directions, respectively. The modification of the nanoparticle geometry leads to a change of the polarisability [51] to:

$$\alpha_i = 4\pi a_1 a_2 a_3 \frac{\varepsilon(\omega) - \varepsilon_d}{3\varepsilon_d + 3L_i(\varepsilon(\omega) - \varepsilon_d)} \qquad (2.83)$$

The substrate i corresponds to the specific ellipsoid axis ($i = 1,2,3$ or x,y,z), ε_d is the dielectric constant of the surrounding media, $\varepsilon(\omega)$ is the frequency dependent dielectric function of the particle (presumed metallic; previously labelled ε_a for when the particle only had size a), and L_i is a geometrical factor used to describe the fraction of a principal axis compared to the other two, such that:

$$L_i = \frac{a_1 a_2 a_3}{2} \int_0^\infty \frac{dq}{(a_i^2 + q) f(q)} \tag{2.84}$$

where q is simply a dummy variable, where we have:

$$f(q) = \sqrt{(a_1^2 + q)(a_2^2 + q)(a_3^2 + q)} \tag{2.85}$$

$$\sum L_i = 1 \tag{2.86}$$

For the case of a sphere, we see that $L_1 = L_2 = L_3 = {}^1\!/_3$, and $a_1 = a_2 = a_3 = a$, which results in equation 2.83 simplifying to that of 2.82, due to the polarisabilities $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$.

Another notable difference between these subwavelength particles, exhibiting localised electron responses, and bulk surface modes is the fact that these particles also exhibit resonantly enhanced scattering (and absorption) properties. From the polarisability, we can obtain the scattering and absorption cross-sections, C_{sca} and C_{abs}, respectively, given by the following equations (for spherical particles):

$$C_{sca} = \frac{k^4}{6\pi} |\alpha|^2 = \frac{8\pi}{3} k^4 a^6 \left[\frac{\varepsilon_a - \varepsilon_d}{\varepsilon_a + 2\varepsilon_d} \right]^2 \tag{2.87}$$

$$C_{abs} = k\,Im|\alpha| = 4\pi k a^3 \left[\frac{\varepsilon_a - \varepsilon_d}{\varepsilon_a + 2\varepsilon_d}\right] \tag{2.88}$$

Due to the fact that the cross sections are dependent upon the size of the particles, a, arising from the polarisability, and the fact that they are very small (such that $a \ll \lambda$) we can deduce that the absorption (scaling with a^3) is dominant over the scattering (scaling with a^6). It is apparent that the absorption cross section presents the relationship $\frac{C_{abs}}{k} = C_{abs}\lambda \propto a^3$, and therefore the resonance wavelength of a nanoparticle is determined by its size and thus will red-shift for increasing particle sizes. In the case for ellipsoidal particles, the equations 2.87 and 2.88 are modified only through their polarisability. Interestingly, there then exist three distinct equations each for the absorption and scattering cross sections (due to $\alpha \rightarrow \alpha_i$), which leads to a splitting of the resonance and in turn implies different resonant frequency conditions for each axis. However, it is still a constraint that the particle dimensions are subwavelength ($a_1, a_2, a_3 \ll \lambda$). This shows promise for light which is normally incident onto such an ellipsoid, where the direction of the polarisation (and also the frequency) determines which of the resonances will be excited – many applications of such multi-resonance structures have been realised, with a large focus upon nanorods for geometric phase metasurfaces, as explained in the next section.

2.4   Geometric Phase

2.4.1   Poincaré Sphere

We saw at the start of this chapter that light can have well defined directionality of its orthogonal fields, namely polarisation. In this, we derived the general formula for polarisation as being akin to that of the equation for an ellipse, and so it is termed the polarisation ellipse. It is all well and good to use mathematical models to describe the polarisation of light, but of course with many things it is much better to be able to *visualise* such polarisation states.

In the late 19ᵗʰ century, Henri Poincaré developed a concept for which any polarisation of light can be represented geometrically on the surface of a sphere, termed the *Poincaré Sphere.* A schematic representation of this is shown below in **Figure 2.6**. The north and south poles correspond to Circularly polarised light, the equator corresponds to linearly polarised light (with x and y polarisations situated opposite) and any point between the poles and equator correspond to elliptically polarised light.

This visualisation is particularly useful when dealing with optical devices, for which a wave propagates through any number of these – each altering the polarisation state – and therefore traces out a path on the Poincaré sphere. For small changes of polarisation, the points situated on the sphere will remain close. If a beam of light experiences a change in polarisation state, then, because each point corresponds to a

state, an arc or path between two points implies a continuous polarisation state evolution and is considered to be slowly-varying and hence adiabatic.

The Poincaré sphere is closely linked to the Jones vector/matrix formalism: because the Poincaré sphere is a 3D object, yet Jones vectors are 2D, the correspondence is because the *surface* of the Poincaré sphere is the only place that a polarisation state is defined – a state cannot exist within the sphere – and therefore the 2D surface of the Poincaré sphere is described well by Jones vectors.
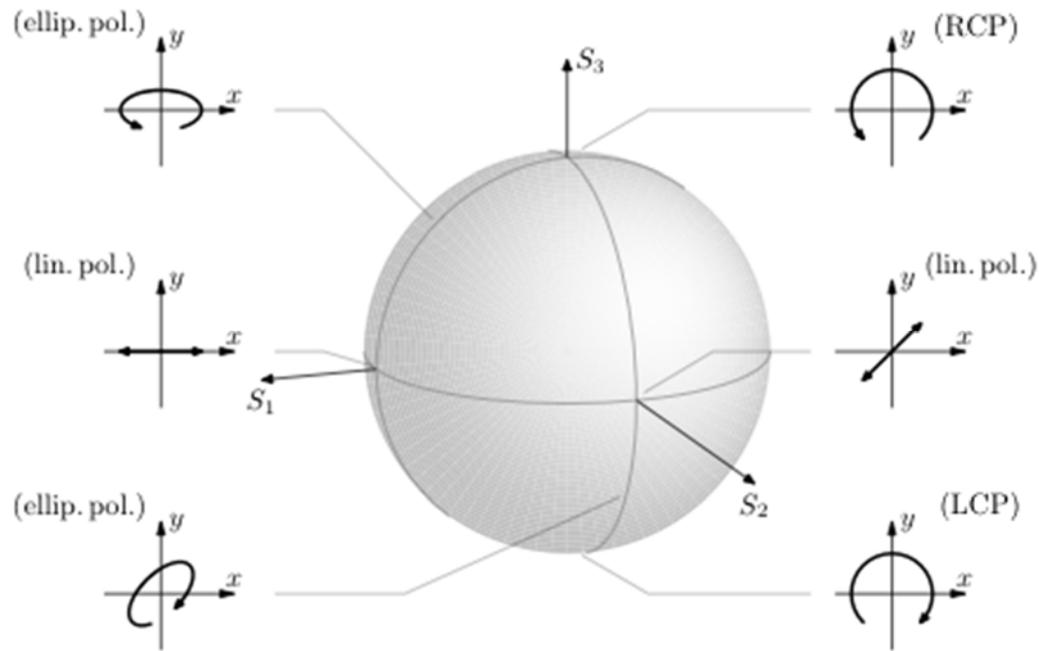
Figure 2.6: Visual illustration of the Poincaré Sphere for representing polarisation states of light [52]

2.4.2   Pancharatnam-Berry Phase

Along with typical and well-understood methods used to control the phase, amplitude, and polarisation of light, one such fundamental property related to the polarisation state and phase of light was discovered in the 1950's by Pancharatnam

[53]. This remarkable work summarised that the cyclic path taken around the Poincaré sphere by a polarised light beam, upon undergoing changes in its polarisation and returning to the original polarisation state, has an associated phase-change, which, surprisingly, is not zero but is in fact equal to half of the area enclosed by this cyclic-path, namely the solid-angle. This seminal work was further expanded to a quantum mechanical description by Michael Berry [54] which showed that this associated optical phase accumulation is analogous to the Aharonov-Bohm effect [55] experienced by electron beams which '*sense'* the magnetic vector potential when passing through two slits separated by a solenoid – even when there exists no net magnetic field outside of this solenoid – and is proportional to the magnetic flux enclosed. These two concepts are grouped together, known as either *Geometric phase* or *Pancharatnam-Berry phase* effects, due to the frequency-independent and geometric nature of the phases involved and the similarity between the works by Pancharatnam and Berry, respectively. In the context of this thesis, where the phenomena and operation of metasurfaces are described using a classical formalism, I will not cover the quantum description of Berry Phase in relation to the geometrical phase (which is necessary for the description of quantum mechanical phenomena, such as the Aharonov-Bohm effect), and will instead explain the result of Pancharatnam in terms of classical geometry and Jones matrix representations of polarisation; a very insightful review letter by Michael Berry on the result of Pancharatnam and the relation to quantum mechanical systems can be found in [56].

The seminal work published by Pancharatnam in 1956 [53] uncovered an amazing link between the polarisation state of light and the resultant phase accumulation. This explained that the phase acquired by a cyclic change of polarisation state of a light beam (where cyclic implies that the start and finish 'states' are the same polarisation) is not zero, and cannot be cancelled out through means of gauge-transformations. If this cyclic path was chosen to be that of a geodesic triangle upon the Poincaré sphere, with the initial and final polarisation states being on the 'North-pole' corresponding to RCP (or LCP, depending on convention), and the other two states being elsewhere (but not overlapping) then the difference of the phase induced in the final polarisation state compared to the initial phase is given in terms of half of the solid angle of the triangle, namely the area, which is simply:

$$\varphi_{A-A'} = e^{-i\frac{\Omega_{ABC}}{2}} \tag{2.89}$$

with the subscripts $i = (A, B, C, A')$ in φ_i and Ω_i corresponding to the states of the polarisation (with states A and A' being the same state, albeit with a phase difference) such that $\varphi_{A-A'}$ implies the phase difference between states A and A' (which are the same polarisation state) and Ω_{ABC} is the solid angle/area of the spherical triangle with vertices of states A → B → C → A'. From this, we can see that the only term which 'matters' is the term in Ω_{ABC} and so we can define the phase factor as being $\Omega_{ABC}/2$, which is the actual solid angle divided by two.

This result arises from the fact that the surface of the Poincaré sphere is indeed curved and not flat, and leads to a modification of a polarisation vector after traversing along a closed path such that the polarisation state is the same but has undergone a phase modification: One intuitive explanation of the factor of ½ for the phase is due to the fact that the real space angle θ corresponds to an angle of 2θ on the Poincaré sphere. If we refer back to Figure 2.6 we see that the north and south poles correspond to Right and Left circular polarisation states, respectively. These form an orthonormal set, as we consider Right and Left polarisations as being orthogonal to each other, and it is therefore intuitive to use these states as the basis for which to describe all other polarisation states. We can describe an arbitrary superposed wave, with equal amplitudes in the Right and Left bases, as [57]:

$$\boldsymbol{E} = \cos\left(\frac{\theta}{2}\right)\hat{e}_R + \sin\left(\frac{\theta}{2}\right)e^{i\varphi}\hat{e}_L \qquad (2.90)$$

Where $\hat{e}_i$ corresponds to the basis vector (Right or Left), and θ, φ are the polar coordinates of a sphere. We can see that this is true if we choose some angles and investigate the corresponding polarisation state: if we use $\theta = 0$, then equation 2.90 amounts to $\boldsymbol{E} = \hat{e}_R$, which is simply the state of the north pole being RCP; for $\theta = \pi$ equation 2.90 amounts to $\boldsymbol{E} = \hat{e}_L$ (where we ignore the term in φ as this is only important for comparing the phase difference between the right and left states); $\theta = \pi/2$ conforms to the equator of the Poincaré sphere, corresponding to linear polarisations, where $\varphi = 0, \pi$ correspond to *x*- and *y*- polarised light, respectively,

and confirms the fact that linearly polarised light is simply a superposition of orthogonally circularly polarised light states with different phases (and vice versa); and lastly, for arbitrary θ, φ (not on the poles or equator) it corresponds to elliptically polarised light.

A question we must ask ourselves, as did Pancharatnam, was "how do we define if two beams of light in different polarisation states are in phase with one another?" We know, of course, that when two beams are in phase in the same polarisation state they result in an interference maximum. Similarly, Pancharatnam defined that two beams in different states should be interfered and are said to be in phase when the resultant signal is the maximum. This definition is termed the "Pancharatnam Connection" (by Berry in [54]), and is represented by the equation:

$$(\boldsymbol{E_1} + \boldsymbol{E_2})^* \cdot (\boldsymbol{E_1} + \boldsymbol{E_2}) = 2 + 2Re(\boldsymbol{E_1}^* \cdot \boldsymbol{E_2}) \tag{2.91}$$

where the terms in $\boldsymbol{E_i}$ are identical to the equation given in equation 2.90, except that now we must introduce a phase difference between the states *1* and *2*, given by $e^{-i\xi}$ (with ξ simply being a phase value between $-\pi$ and $+\pi$); we then express these states as:

$$\boldsymbol{E_1} = \cos\left(\frac{\theta_1}{2}\right)\hat{e}_R + \sin\left(\frac{\theta_1}{2}\right)e^{i\varphi_1}\hat{e}_L \tag{2.92a}$$

$$\boldsymbol{E_2} = \left[\cos\left(\frac{\theta_2}{2}\right)\hat{e}_R + \sin\left(\frac{\theta_2}{2}\right)e^{i\varphi_2}\hat{e}_L\right]e^{-i\xi} \tag{2.92b}$$

We must introduce orthonormality conditions for the basis vectors $\hat{e}_R$, $\hat{e}_L$, to satisfy

the following:

$$\hat{e}_R{}^* \cdot \hat{e}_R = \hat{e}_L{}^* \cdot \hat{e}_L = 1 \tag{2.93a}$$

$$\hat{e}_R{}^* \cdot \hat{e}_L = \hat{e}_L{}^* \cdot \hat{e}_R = 0 \tag{2.93b}$$

Upon examination of equation 2.91, it is clear that the dot product between $\boldsymbol{E_1}^*$ and

$\boldsymbol{E_2}$ may have terms which are imaginary (for an arbitrary state) and may also have

the special case of being equal to zero. Due to this, we must also impose the following

lemma such that the 'in-phase' explanation by Pancharatnam holds true, where we

require:

$$Re(\boldsymbol{E_1}^* \cdot \boldsymbol{E_2}) > 0 \tag{2.94a}$$

$$Im(\boldsymbol{E_1}^* \cdot \boldsymbol{E_2}) = 0 \tag{2.94b}$$

Upon substitution of equations 2.92 into the above constraint equations, we obtain

two unique solutions for the phase difference ξ.

We show how this is beneficial, and indeed the key result of Pancharatnam,

by presenting three polarisation states A, B, C. We impose that state B is in phase

with A, state C is in phase with B, and a state, identical in polarisation to A, labelled

A' is in phase with C; however, this argument is dependent upon the fact that state

C need not be in phase with A, but instead A' – therefore, state A and A', although

occupying the same 'point' on the Poincaré sphere (and hence being identical states),

are not in phase with each other. These three states trace the outline of a triangle on

the Poincaré sphere surface (also called a geodesic triangle). For simplicity, without

loss of generality, we can choose that state A corresponds to the north-pole of the

Poincaré sphere, namely RCP. We choose state B to coincide with a state on the

prime meridian geodesic line of longitude (where the prime meridian is for azimuthal

angle $\varphi = 0$, and we set this coinciding with *x-polarised* light on the equator) with

some angle $\theta = \theta_B$, and a phase retardation compared to state A being $e^{-i\xi_B}$. State

C is then an arbitrary state, given by angles $\theta = \theta_C$, $\varphi = \varphi_C$, and phase retardation

compared to state B (which itself is with respect to state A) as being $e^{-i(\xi_B + \xi_C)}$. As

was said before, state C is in phase with state A' and not state A, where we assign

the phase difference of state A' compared to state A as being $e^{-i\xi_{A'}}$. These three

equations are represented in **Figure 2.7** and are then given as the following:

$$E_A = \hat{e}_R \tag{2.95a}$$

$$E_B = \left[\cos\left(\tfrac{\theta_B}{2}\right)\hat{e}_R + \sin\left(\tfrac{\theta_B}{2}\right)e^{i(\varphi_B = 0)}\hat{e}_L\right]e^{-i\xi_B} \tag{2.95b}$$

$$E_C = \left[\cos\left(\tfrac{\theta_C}{2}\right)\hat{e}_R + \sin\left(\tfrac{\theta_C}{2}\right)e^{i\varphi_C}\hat{e}_L\right]e^{-i(\xi_B + \xi_C)} \tag{2.95c}$$

$$E_{A'} = e^{-i\xi_{A'}}\hat{e}_R \tag{2.95d}$$

We now substitute concurrent pairs of these (AB, BC, CA') into the constraint

equations given in equations 2.94a,b, and we obtain the following final result (for

BC) as being:

$$\tan\xi_C = \tan\xi_{A'} = \frac{\sin\varphi_C \sin\left(\frac{\theta_B}{2}\right)\sin\left(\frac{\theta_C}{2}\right)}{\cos\left(\frac{\theta_B}{2}\right)\cos\left(\frac{\theta_C}{2}\right)+\cos\varphi_C \sin\left(\frac{\theta_B}{2}\right)\sin\left(\frac{\theta_C}{2}\right)} \tag{2.96}$$
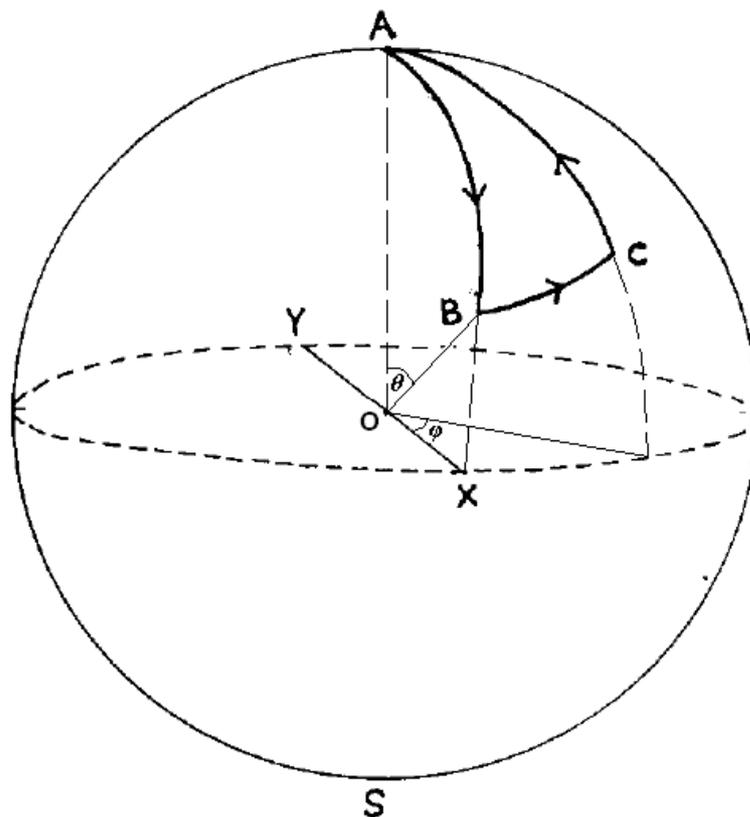
Figure 2.7: Three distinct polarisation states A,B,C on the Poincaré sphere (image from [57])

where we have labelled the phase term of $\xi_C = \xi_{A'}$ from the previous explanation of

state C being in phase with state A'. This result tells us that the phase redardation

$\xi_{A'}$, which is the phase accumulated from the cyclic polarisation state evolution from

A to A', is directly proportional to the angles subtended by the three polarisation

states A,B, and C. However, we cannot directly assume this is correct and equal to

the area of this triangle. There is a fundamentally derived equation to obtain the

area of such a spherical triangle in terms of unit vectors centred at the centre of a

sphere of unit radius, given by [58]:

$$\tan(\Omega/2) = \frac{|\boldsymbol{a} \cdot \boldsymbol{b} \times \boldsymbol{c}|}{1 + \boldsymbol{b} \cdot \boldsymbol{c} + \boldsymbol{c} \cdot \boldsymbol{a} + \boldsymbol{a} \cdot \boldsymbol{b}} \tag{2.97}$$

Where Ω is the area of the spherical triangle (or the "spherical excess"), and $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are the unit vectors from the centre of the sphere to the triangle vertices, given by

$$\boldsymbol{a} = (0,0,1), \qquad \boldsymbol{b} = (\sin\theta_B, 0, \cos\theta_B), \qquad \boldsymbol{c} = (\sin\theta_C \cos\varphi_C, \sin\theta_C \sin\varphi_C, \cos\theta_C).$$

Substitution of these vectors into equation 2.97 yields:

$$\tan(\Omega/2) = \frac{\sin\varphi_C \sin\theta_B \sin\theta_C}{(1+\cos\theta_B)(1+\cos\theta_C)+\cos\varphi_C \sin\theta_B \sin\theta_C} \qquad (2.98)$$

It can then be proved that we then have that $\Omega/2 = \xi_{A'}$, by some mathematical trickery and trigonometric/algebraic juggling (as advised in ref [57]) by multiplying both the numerator and denominator of equation 2.96 by $4\cos\left(\frac{\theta_B}{2}\right)\cos\left(\frac{\theta_C}{2}\right)$ to convert the half-angles to whole angles. Many experimental works on this have proven it to be true [53,59,60] and is accounted for in setups where time-variant systems make use of many polarisers.

One small disadvantage of these optical setups to achieve Pancharatnam's geometric phase is that they (typically) involve the use of polarisers to produce such a cyclic path on the Poincaré sphere, and therefore involve large experimental setups. These setups are space-invariant, and depend upon the propagation of the light beams through optical elements to invoke a polarisation state change. In the works by Hasman [37–39,60–63], though, a *planar* approach is utilised to control the local polarisation state of a beam when encountering a transversely inhomogeneous surface composed of polarising elements. These works make use of the fact that such a device, which is spatially-varying in-plane, induces local polarisation elements to perturb the

wavefront of the light. Such an effect occurs when an incident CP wave encounters

the optical elements, rotated at some angle with respect to the linear basis, that

behave as linear polarisers and thus have corresponding polarisation states lying on

the equator (linear polarisation). These beams can then either be transmitted with

the same CP handedness, or be flipped to the opposite handedness. It is those that

are flipped which obtain a geometric phase of the form:

$$\varphi = \pm 2\theta \tag{2.99}$$

Where the θ in this equation is simply the difference in the rotation angle between

the linearly polarising elements of the inhomogeneous planar device (metasurface)

with respect to a chosen linear basis (normally the x-axis), and the $+/-$ signs are due

to the fact that the incident light can be either Right or Left polarised, respectively;

one handedness will result in an opposite sign of the acquired phase compared to the

orthogonal handedness - This can be shown in **Figure 2.8** (from [64]). The geometric

phase arises from the fact that the incident CP light, upon encountering such angled

linearly polarising elements, will traverse different paths on the Poincaré sphere (as

shown by the paths $\hat{A}$ and $\hat{B}$ in Figure 2.9) and arrive at the opposite handedness of

light, thus creating a closed path and producing a phase difference equivalent to

twice the angular difference between the elements. It still holds, as with the classical

description of the Pancharatnam phase, that the phase is equal to half of the area

enclosed on the Poincaré sphere. As a trivial example, if we have two optical
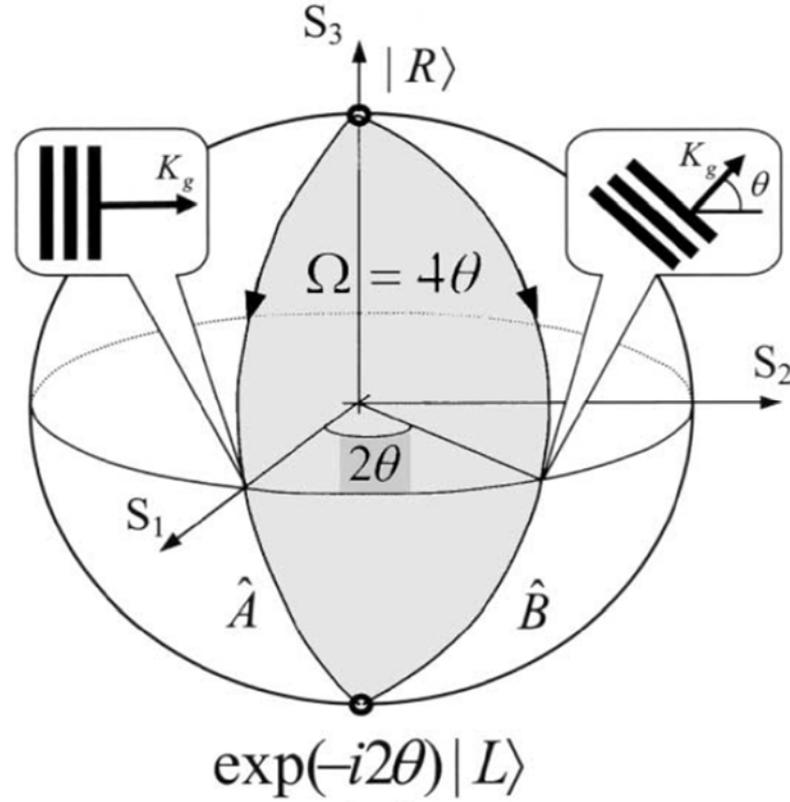
Figure 2.8: Representation of Geometric (Pancharatnam) phase introduced for space-varying angled elements [64]

The optical elements have differing angles, and so act as mini linear polarising elements, thus providing inhomogeneity in the transverse (in-plane) directions. These elements result in a geometric phase acquisition due to the fact that CP light will traverse different paths on the Poincaré sphere, resulting in a geometric phase equal to twice the angular difference between the polarising elements. Equivalently, the phase is still exactly half of the area of the closed loop on the Poincaré sphere.

elements, with an angle of $\theta = \pi/4$ between them, then for an incident CP light beam we will have a phase difference of $\varphi = \pm 2\theta = \pm 2(\pi/4) = \pm \pi/2$; as a simple affirmation of this result, we can fix one of these elements as x-polarised on the equator, and the other will be shifted along the equator by $\pi/2$ (corresponding to 45° polarised). We know that the area of a sphere (of unit radius) is simply 4π, so a quarter of this is simply π, which, when using the result of Pancharatnam, gives us

a phase of half of this area i.e. $\varphi = \pm\,\Omega/2 = \pm\,\pi/2$, as calculated above. This

surprisingly simple, yet powerful method, of acquiring phase has been utilised in

many devices, especially in the developing field of metasurfaces [41,65–69], as all that

is required to modify the phase of a wavefront is a simple rotation of the optical

elements (e.g. rod antennas). This method of acquiring phase is dispersionless, and

does not rely on the need for complex fabrication techniques.

2.4.3   Generalised Law of Refraction

One of the first optical problems encountered in the development of electromagnetism was that of the refraction of light at the boundary from one medium to another. The earliest recorded quantitative description of this phenomenon dates back to the 2ⁿᵈ century A.D. by the mathematician Claudius Ptolemy [70]. However, it wasn't until the 17ᵗʰ century that the well-established and renowned "*Snell's Law*" was officially discovered by Willebrord van Roijen Snell, which relates the angles of incidence and transmission of a beam of light with the refractive indices of the medium, and is given by:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{2.100}$$

where subscripts 1,2 correspond to the incident and transmitted media. Although this equation was attributed credit to Snell, it was independently discovered by Descartes at around the same time (and the equation is thus alternatively known as Snell-Descartes Law). However, the physical interpretations of Descartes were not correct, and it was this that made Pierre de Fermat snub his reasoning and to derive this from first principles, using his idea that light is a wave and travels from one point to another by *taking the least optical path* [40,71]. By using the relationship that the speed of light in a vacuum compared with in a medium is related to the refractive index of the medium by $c = n_i v_i$, it can be derived using Fermat's principle to obtain Snell's Law from geometries of triangles (shown below in **Figure 2.9**).

However, it was mentioned in Chapter 1 that recently seminal work carried out by

Capasso [40] yielded anomalous refraction and reflection of light by the application

of abrupt phase changes on the interface between two dielectric media. This was

coined as the *Generalised Snell's Law*, due to the fact that it included a term to

describe the linear phase progression along an interface and that the traditional

Snell's law could be retrieved by simply setting this phase term to zero.

Fermat's principle additionally states that two (or any arbitrary number of)

paths traversed by light, which lie an infinitesimal distance away from the actual

path taken by light, between two media are equal in phase and constructively

interfere. Shown in **Figure 2.9** is a schematic which illustrates two paths taken by

light travelling from point A to point B when crossing an interface between mediums

1 and 2 (with refractive indices of n_i and n_t, respectively). If we treat the light as

having a wavevector (in vacuum) of k_0 then the wavevectors within medium 1 and

2 are $k_0 n_i$ and $k_0 n_t$, respectively. Traditionally, Snell's law deals with isotropic media

where the crossing point on the boundary between two media is identical in phase

wherever a light ray crosses. However, we now have the capability to engineer abrupt

phase changes on a boundary using metasurfaces which involve ultrathin antennas.

This allows us to tailor the phase accumulated by light as a function of position, as

is shown in Figure 2.9 by paths 1 and 2, which pass through points on the interface

and acquiring phases of ϕ and $\phi + d\phi$, respectively. These points are infinitesimally
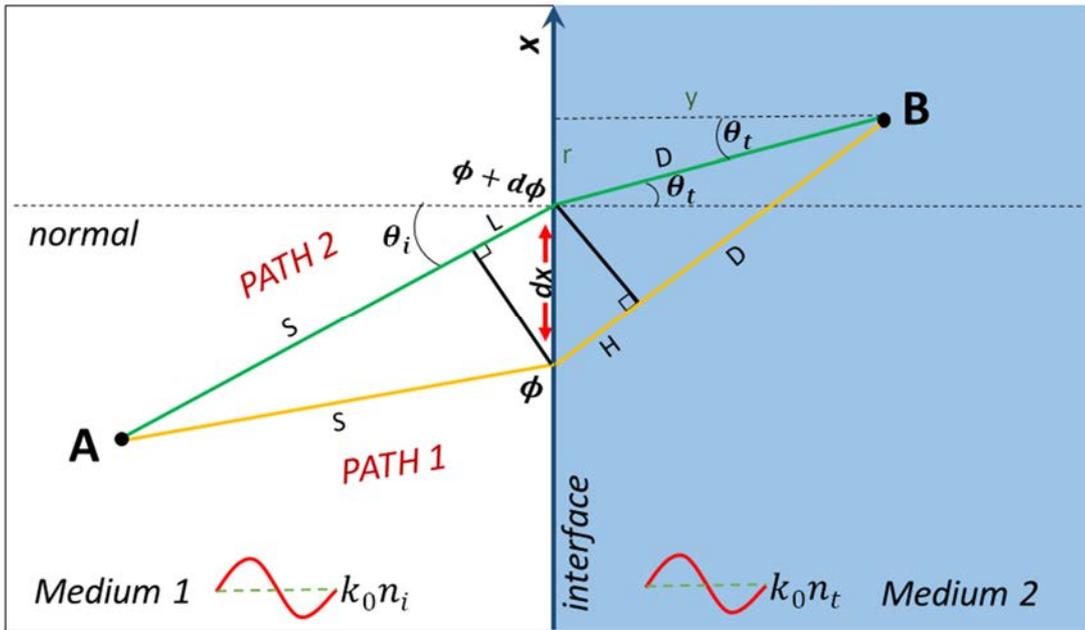
***Figure 2.9: Geometric representation of Fermat's 'Principle
of Least Optical Path' for deriving Generalised Snell's Law***

Light is proposed to travel the least optical path between the points **A** and
B, where the . The beam of light must traverse the boundary between these
two points at some crossing location **O**, which is allowed to move, such that
the minimised time is least when utilising the differing velocities and distances
travelled by the light rays.

separated by a distance of *dx.* We see that Path 1 differs to Path 2 by the extra

distances H and L travelled in media 1 and 2, respectively, where S and D correspond

to a fixed distance travelled by the wave in media 1 and 2, respectively. Therefore,

the only terms which contribute to differences in phase travelled by light along Paths

1 and 2 are *H, L,* ϕ and $\phi + d\phi$.

We can calculate *L* trivially from using *dx* and the incident angle θ_i to give:

$$L = \sin \theta_i \ dx \tag{2.101}$$

Calculating H is a little more cumbersome – we see that the distances y and r are simply given as:

$$y = \mathrm{D} \cos \theta_t \tag{2.102}$$

$$r = D \sin \theta_t \tag{2.103}$$

And therefore by using Pythagoras' Theorem we have the following equation:

$$(H + D)^2 = (dx + r)^2 + (y)^2 \tag{2.104}$$

Where by substitution of y and r from equations 2.102-3 into the above gives us:

$$(H + D)^2 = (dx + D \sin \theta_t)^2 + (\mathrm{D} \cos \theta_t)^2 \tag{2.105}$$

$$\rightarrow \quad H^2 + D^2 + 2HD = dx^2 + D^2 \sin^2 \theta_t + 2D \sin \theta_t \, dx + \mathrm{D}^2 \cos^2 \theta_t \tag{2.106}$$

We know that dx is infinitesimal, and therefore so is H (from the figure) so we can remove the squared terms in these two. Collecting like terms then gives us:

$$D^2 + 2HD = D^2 + 2D \sin \theta_t \, dx \tag{2.107}$$

$$\rightarrow \quad H = \sin \theta_t \, dx \tag{2.108}$$

It is known that the phase of a wave can be obtained by multiplying it's wavevector by the distance traversed, which we now carry out (including the abrupt phase accumulations dependent on the crossing points of Path 1 and 2), and setting the total phase of each path as being equal due to the constraints of Fermat's principle:

Path 1 $\qquad\qquad Sk_0 n_i + \phi + k_0 n_t \sin \theta_t \, dx + Dk_0 n_t$ $\qquad\qquad$ **(2.109)**

Path 2 $\qquad = Sk_0 n_i + k_0 n_i \sin \theta_i \, dx + \phi + d\phi + Dk_0 n_t$

We can cancel like terms on both sides of the equation to give:

$$k_0 n_t \sin\theta_t \, dx = k_0 n_i \sin\theta_i \, dx + d\phi \qquad (2.110)$$

And with some shuffling, we then obtain the ***Generalised Snell's Law*** as being:

$$n_t \sin\theta_t - n_i \sin\theta_i = \frac{1}{k_0}\frac{d\phi}{dx} \qquad (2.111a)$$

$$n_t \sin\theta_t - n_i \sin\theta_i = \frac{\lambda}{dx}\frac{d\phi}{2\pi} \qquad (2.111b)$$

Where the left hand side is identical to the Classical Snell's law, whilst the right hand

side tells us that if the abrupt phase changes linearly with position along a boundary

then an anomalously refracted beam will be produced. As a simple example of this, let

us assume that light with a wavelength of λ is normally incident onto a phase gradient

interface with a spacing of $dx = \lambda/8$ and an abrupt phase shift of $d\phi = \pi/4$, then equation

2.111b will simply reduce to:

$$\sin\theta_t = \frac{1}{n_t}\frac{\lambda}{(\lambda/8)}\frac{(\pi/4)}{2\pi} = \frac{1}{n_t}\cdot 8\cdot\frac{1}{8} = \frac{1}{n_t} \qquad (2.112)$$

If we assume that the substrate with refractive index of n_t is simply glass with an index

of 1.5, then we find that the refracted angle of the normally incident light due to a phase

gradient along the boundary is:

$$\theta_t = \sin^{-1}\left(\frac{1}{1.5}\right) \cong 42° \qquad (2.113)$$

This result is very profound, due to the fact that simply by controlling the abrupt phase

discontinuities on the interface between two media we can tailor the angle of the refracted

beam which differs from that of the classical law of refraction (Snell's law). We can

then generalise this to control the abrupt phase in a non-linear way, which allows

applications such as lensing, holography, and beam-shaping to be carried out. As a final

note, if we look back at the Generalised Snell's law (or Generalised Law of Refraction) in

equation 2.111 and assume that there does not exist any abrupt phase changes in a

system, such that $\frac{d\phi}{dx} = 0$, then the equation simply reduces back to the form of:

$$n_t \sin\theta_t - n_i \sin\theta_i = 0 \qquad\qquad (2.114)$$

which is just the classical Snell's law of refraction between two isotropic media of

refractive indices n_i and n_t.

Chapter 3

Experimental Techniques

In this chapter, I shall discuss the methods undertaken for fabricating and characterising the metasurfaces developed in later chapters of this thesis. The first part of this chapter (3.1) covers Fabrication, where topics such as photolithography, plasma etching, and metal evaporation will be explained. The second part of the chapter (3.2) touches on the methods of characterising the fabricated samples, such as surface structure and topography, through methods including optical microscopy and surface profilometry. The optical properties of Terahertz devices are probed using a Terahertz Time Domain Spectroscopy (THz-TDS) system, able to angularly resolve the refracted beams metamaterial samples and characterise the linearly polarised Jones Matrix components, which in turn can be used to retrieve the circularly polarised Jones Matrix information.

3.1   Fabrication

Fabricating metasurfaces involves a variety of techniques and methods. This section aims to explain some of these techniques, which were used to fabricate the metasurfaces detailed in the proceeding chapters of this thesis.

3.1.1   Sample Preparation

An important step in the fabrication of samples is correct preparation of the substrate, and making sure that it is as clean as possible before anything else is done with it. A dirty substrate will not give the necessary accuracy of feature sizes and dimensions due to the sensitive nature of processes that take place; An example is the spinning of a photoresist on a silicon wafer, in which debris or impurities on the surface will cause unwanted 'streaks' and thickness variations of the resist leading to issues when patterning and developing. The following paragraph outlines the typical cleaning routine carried out for a silicon wafer substrate, although this can be applied to glass substrates also.

The first step in cleaning a silicon wafer is using a nitrogen gun to remove visible dust and stray debris or particles on the surface, for obvious reasons. Once this is done, the wafer is placed into a glass beaker containing acetone, which is widely used for laboratory cleaning due to it being a good solvent and relatively safe, and then placed into an Ultrasonic Bath, shown in Figure 3.1, below. Acetone is used

as a solvent to dissolve and lift organic residues present on the surface, e.g.

fingerprints, whilst the ultrasonic bath helps to speed up this process by *agitation;*

this is typically carried out for 10 minutes. The wafer is then removed from the

***Figure 3.1: Overview of Hot-plate, Spin-coater and
Ultrasonic Bath.***

The photo shows from left-to-right, respectively, the Hot-plate, the Spin-
Coater, and the Ultrasonic Bath, used in the preparation of samples for
fabrication.

acetone and cleaned with *De-Ionised* (DI) water for at least 1 minute, and is placed

into a glass beaker with *Isopropyl Alcohol* (IPA) to remove any chemical residues

left on the wafer surface from the acetone. Again, the beaker with IPA is placed into

an ultrasonic bath for approximately 10 minutes, and then rinsed off with DI water

for at least 1 minute. The wafer is then blown dry of water droplets with the nitrogen

gun and placed onto a hotplate (see Figure 3.1, above) at a temperature of 120°C,

which aids to completely evaporate any adsorbed water on the wafer surface. Finally,

the wafer is removed from the hotplate, and then allowed to cool to room temperature

before storing in a suitably sized wafer-carrier, and is then ready for fabrication.

The prior routine is the typical process used when the quality of the sample surface is necessary for micron sized feature fabrication. However, if one wishes to achieve nanometre resolution features, or removing a particularly resilient resist, we must perform an additional cleaning process with so-called *Piranha* solution. This solution is a mixture of *Sulphuric Acid* (H_2SO_4) and *Hydrogen Peroxide* (H_2O_2), which undergoes an exothermic reaction when combined. This mixture is very strongly oxidising, and in conjunction with the exothermic nature, in which temperatures can exceed 100°C, it is extremely efficient at removing organic matter from substrate surfaces. Equal parts of H_2SO_4 and H_2O_2 are normally used, although different percentages are sometimes required for tough-to-remove organic matter (such as the photoresist SU8, which is covered later). Firstly, a small measure of sulphuric acid is added to a glass beaker (large enough to accommodate the aforementioned clean wafer), which for a 4" wafer is typically approximately 50ml. Then, making sure to use a separate measuring beaker, an equal amount of hydrogen peroxide is *slowly* added to the sulphuric acid. If this is done too quickly, there is a possible risk of explosion, due to the volatility of the (exothermic) chemical process and the sharp increase in temperature; the mixture is always made by adding the H_2O_2 to the H_2SO_4 and never the other way around. Once all of the H_2O_2 is added to the H_2SO_4, a stopwatch is set to a time of 10 minutes, as a benchmark for the cleaning process, and the clean wafer carefully submerged. During the cleaning process, fuming may occur and so is carried out in a suitably well-ventilated fume-cupboard to prevent

hazardous gases being inhaled. Additionally, the process is carried out whilst wearing

protective gear, including a double-layer of acrylic gloves and protective eye goggles.

Once the wafer has undergone the 10 minutes of cleaning it is rinsed under DI water

for at least 2 minutes to remove any piranha solution which may still be present.

The wafer is then blown dry using a nitrogen gun, and baked on the hot plate to

remove adsorbed water.

3.1.2   Photolithography

Once we have our cleaned substrate, we can then perform the next step of

fabrication, which for many cases is that of *photolithography.* Photolithography is a

very useful fabrication technique whereby a *'top-down'* approach to achieving

micrometre (and larger) feature sizes on a substrate can be carried out quickly and

in parallel. This is done by: coating a substrate with a light-sensitive polymeric

solution called *Photoresist*; covering this with a patterned hard-mask; then exposing

all of this with a light source (typically UV light) which causes a chemical reaction

in the photoresist to take place. Upon developing, the photoresist is then

characterised for the mask-pattern with only the desired patterned areas remaining

and the rest of the photoresist being dissolved in the solution. This is carried out to

either have the photoresist itself as the desired structure, or to partially protect the

substrate surface to then lead on to (dry) etching or for lift off processes. A schematic

diagram of this process is shown in Figure 3.2a, below.
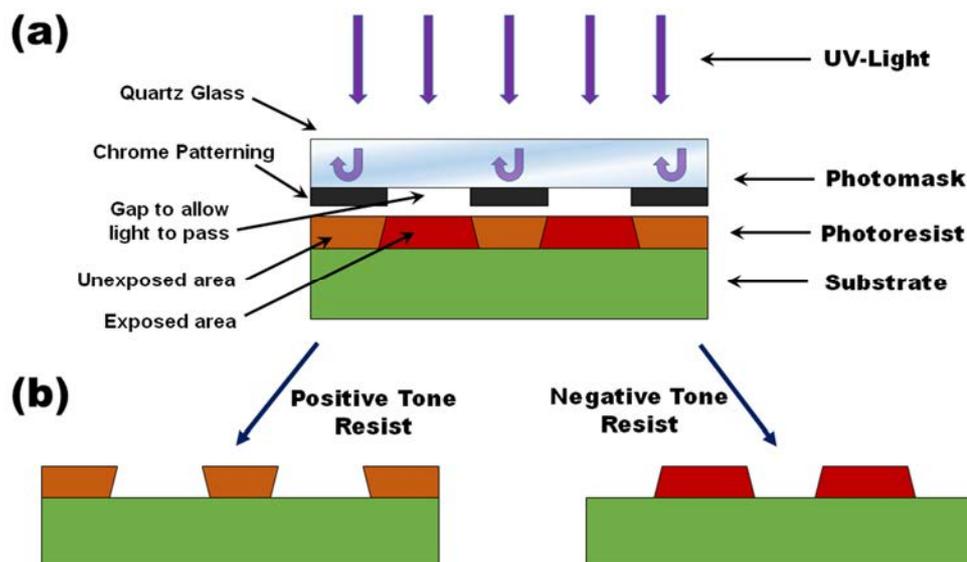
Figure 3.2: Schematic of Photolithographic process

(a) A substrate which has been spun-coated with photoresist is exposed to UV light with a patterned-mask defining the desired exposure selectivity. Upon developing, a direct image of the mask is uncovered in the photoresist. **(b)** Dependent upon the type of resist used (either positive or negative tone) the resultant developed pattern can either exhibit a direct or inverse image of the mask.

Photoresist is a light-sensitive polymeric solution, which undergoes a chemical reaction upon illumination of UV light. It consists of a mixture of photoactive monomers/polymers, solvent, sensitiser, and possible other additives. The polymers either polymerise, to form larger cross-linked molecules, or undergo chain-scission upon illumination of light; the solvent is used to allow spin-coating; and the sensitiser is to control the photoactivity of the polymer. Photoresists can either be *Positive* or *Negative* tone, where positive implies that the exposed areas become more soluble in a developer, whilst negative tone resists have their exposed areas become less soluble in a developer and remain after development (an example of both types of resist is

shown in Figure 3.2b). The photoresist is applied to the surface of the substrate (typically a clean, flat wafer) through the use of *spin-coating.* This is where the sample is affixed to a spin-coating machine, (as shown below in Figure 3.3) which is programmed to rotate at very high speeds (of the order of 100-4000 revolutions per minute), and pouring on a small amount of photoresist which then spreads evenly across the sample to create a flat uniform area. This is necessary to allow good contact to be made with a hard-mask. The spin-coating typically takes only a few minutes or less to perform, with cycles in the speed and acceleration taking place (increasing in speed, from low to high) to assure uniform spreading. The fastest of the spin speeds in these cycles determines the thickness of the resist, with higher speeds resulting in a thinner resist, and slower speeds resulting in a thicker resist; this is due to the centripetal force of the rotating substrate causing the liquid resist
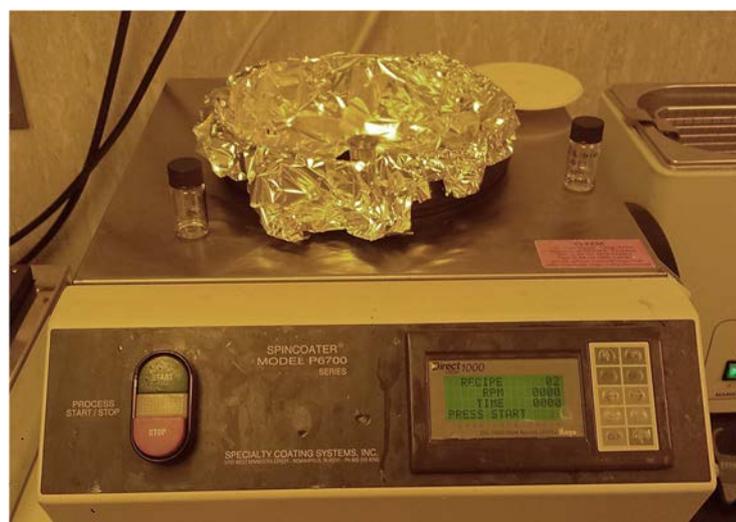
Figure 3.3: Close-up view of the Spincoater^{TM} P6700 system

The spin-coating system accommodates both 4" wafers and smaller substrates, with a built in vacuum suction feature to keep the substrate affixed whilst spinning.

on top to be forced outwards. Once the resist has been spun onto the substrate, it is then transferred to a hotplate to encourage evaporation of the solvent and a stronger adhesion to the substrate. This causes a solidification of the resist, which would otherwise be problematic for contact with the photomask if it were still viscous.

Following the spin-coating, the resist is then required to be exposed to UV light in order to make it selectively-soluble, and have a desired pattern. This is done with the use of a *photomask,* which is inserted into a *photomask aligner*, as shown in Figure 3.4. The photomask was designed through the use of Autocad™ software, where the exact dimensions and features can be accurately controlled. This design is then saved and sent for fabrication at an external company. A photomask typically consists of a thick, flat quartz square with the desired pattern printed onto one side using chromium. Feature sizes can be as small as 2μm, which nears the limit

Table 3.1: The Procedure for Processing of SPR220-7.0 Photoresist

Process Type	RPM	Spin Time (seconds)	Ramp Time (seconds)
Cycle 1	500	15	4
Cycle 2	2500	45	5
Prebake	105°C for 90 seconds		
Exposure	20 seconds @ 365nm (~300mJ/cm²)		
Development	Agitation in MF-26A solution for ~ 2 minutes		

of optical focussing where smaller feature sizes would cause diffraction effects. The

mask aligner used was a Karl Suss MJB-3 Mask-Aligner, which operates at a

wavelength of 365-405nm. The mask is affixed to the maskholder with the use of a

vacuum seal (underside of the black platform circled in Figure 3.4 and inset), which

surrounds the aperture used for exposing the sample beneath. Since the exposure

aperture of the maskholder is of a finite size (the aperture is circular with a diameter

of 40mm, as shown in the inset of Figure 3.4), it is required that the mask be designed

in such a way that the desired pattern be contained within the area of this aperture.
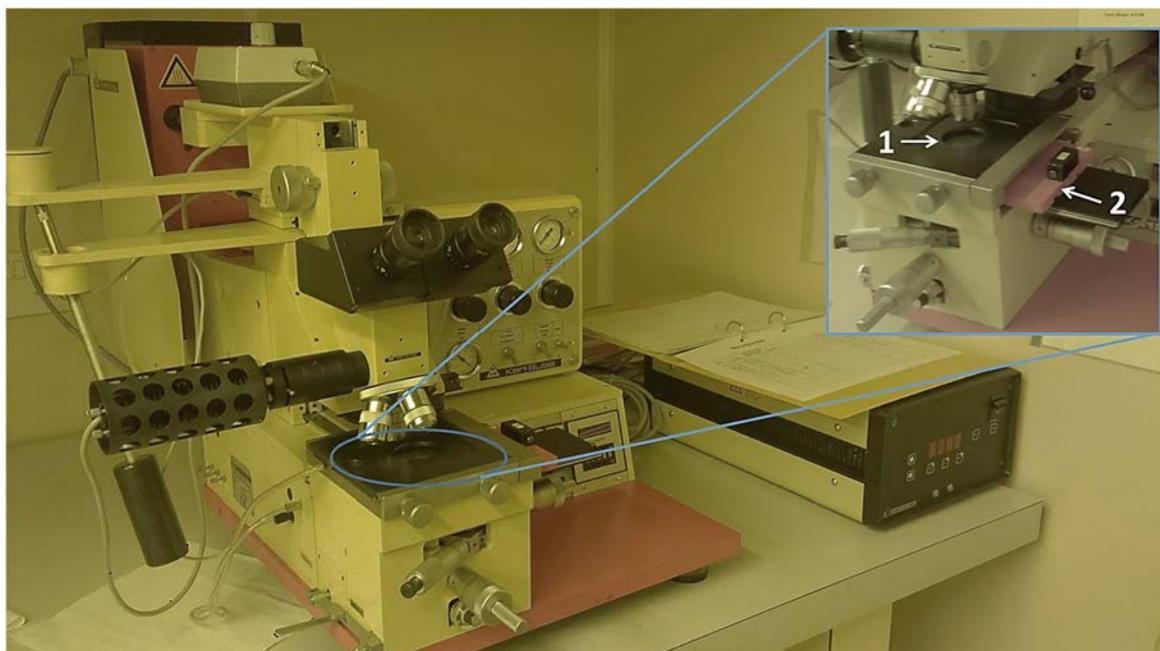
Figure 3.4: Photograph of the Karl Suss MJB-3 Mask Aligner

The mask-aligner accommodates 100mm (4") masks and has an exposure
aperture of 40mm (labelled 1, inset), where the desired patterned region on
the mask can be positioned. A sample of 30mm or smaller is placed beneath
this aperture (on the red platform, inset 2), in contact with the mask, and is
exposed to 365nm UV light. The inset shows where the mask and sample are
placed, along with adjustment knobs.

Due to this, and the limited movement of the mask whilst fitted to the maskholder,

it was required that the mask have 5 accessible regions – an example of a photomask

with 5 regions, of different designs, is shown in Figure 3.5. These regions were all

accessible through careful positioning of the mask to make sure the selected region

was centred over the aperture, which involved rotating and translating the mask on

the maskholder stage. When a region was selected and centred over the aperture the

vacuum was activated such that the vacuum seal affixed the mask to the maskholder

and rendered it unmovable. This is a necessity for photolithography, when even a
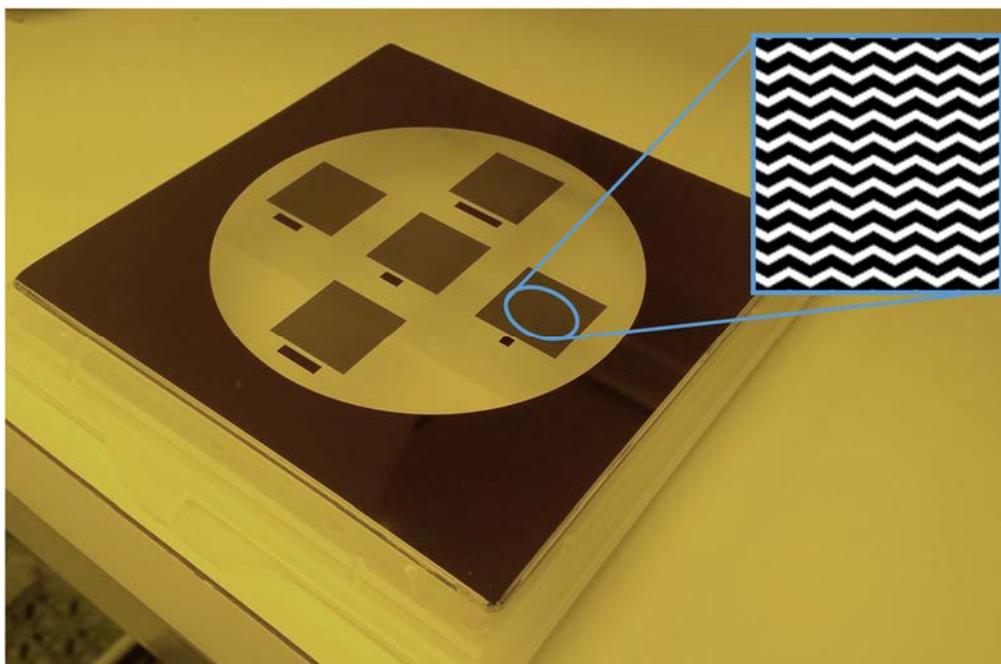
**Figure 3.5: Photograph of a photomask designed for use
with the Karl Suss MJB-3 Mask Aligner**

This is a typical chrome-plated quartz mask used for photolithography in the
MJB-3 mask-aligner. The mask size is 4", and the 5 regions are approximately
1 square-inch each. Only one side is coated with chromium to define the
pattern. Inset: a close-up of patterned-region of the mask.

slight movement may cause misaligning the sample with the mask pattern, especially when dealing with the smallest feature sizes of a few micrometres. The maskholder could then be inverted and fixed back into the stage.

Samples to be exposed were placed on the slide-able red stage situated beneath the maskholder (and mask), as shown by the inset of Figure 3.4, and then raised to provide hard-contact with the chrome-side of the mask. Exposure of the samples takes place using contact mode (where the surface of the photoresist is brought very close and touching the mask), as this is the best means of achieving small feature sizes due to less optical scattering distance. Exposure times can vary for each resist, and indeed for different mask aligners, but for the MJB-3 system it was of the order of 30 seconds or less. Following exposure, it was crucial to leave the sample to 'rest' for an hour or so prior to developing to allow the photoactive chemistry to take place.

Once the sample had been given adequate resting time after exposure, it could then be developed. This took place in a fume cupboard, due to the presence of toxic chemicals used for development and to prevent the inhalation of any gases that may have been produced. A glass beaker large enough to accommodate the sample was partly filled with the requisite developer, of at least 5mm depth, and the sample submerged. A steady agitation was carried out to encourage solubility of the exposed resist, by swirling the developer in a circular manner. Utilising this method normally resulted in complete solubility within 2 minutes, where the resist was seen to dissolve

by a clouding effect and the absence of clouding signalling the end of development.

The sample was then removed from the solution and rinsed under DI water for 1

minute, before being gently dried with a nitrogen gun.

3.1.3   RIE (Plasma) Etching

Etching of a substrate typically follows photolithography, whereby the

previously exposed photoresist has uncovered areas allowing access to the substrate

beneath; this in turn can allow a direct transferral of the photoresist pattern into the

actual substrate. One such method of etching is Reactive Ion Etching (RIE), or

simply *plasma* etching (for ease of narrative), where chemically-reactive gaseous ions

are used to bombard the substrate surface and carry material away. This process is

highly anisotropic, and is used to produce very high-aspect ratio features in a

substrate, with steep sidewalls. The mechanism behind RIE can either be due to

physical bombardment of the high-energy plasma ions; chemical reactivity of the ions

at the substrate surface; or a combination of these two, which is often the case; an

overview of these processes are shown in **Figure 3.7**.

A typical RIE system (*for explanatory purposes, here is referenced an STS ICP*

etcher [72]) is composed of a vacuum chamber, into which a sample can be placed on

an electrically-isolated platform near the bottom. Inlets near the top of the system

allow entry of the gases used to produce the plasma, whilst an outlet situated near

the bottom allows evacuation of the heavier volatile reactants by a vacuum pump

(seen in **Figure 3.6**). The plasma is generated by Inductive Coupling of radio-

frequency magnetic fields generated from electrodes, referred to as the RF Matching

Unit (shown in Figure 3.6) which operate at a frequency of 13.56MHz, and the strong

oscillating magnetic field generated from the current through the wires is used to

provide the energy for ionisation of the gases. In the case of ICP there are two RF

matching units used; one of the units is solenoidal and situated close to the
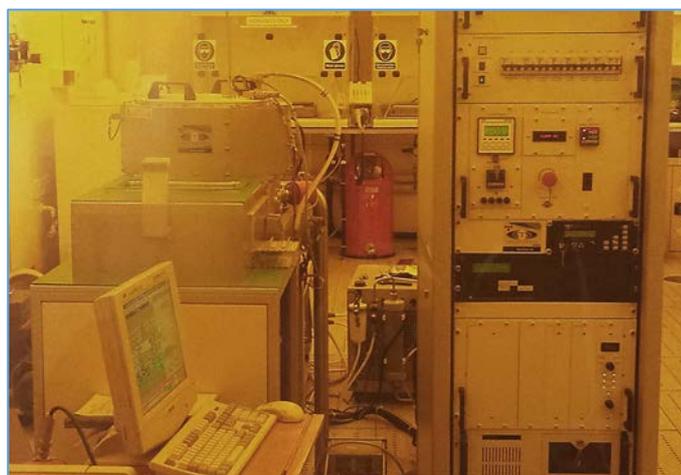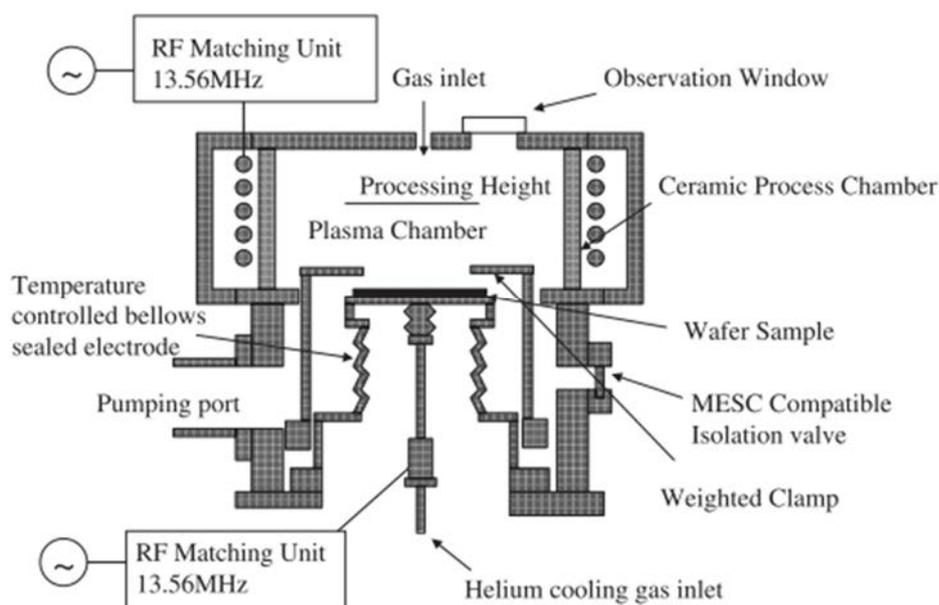
Figure 3.6: Schematic and photograph of an STS ICP DRIE etcher [72]

gas inlets to induce a plasma, whilst the second is situated orthogonally to this and is utilised to drag the plasma ions towards the sample stage by providing a negative bias on the sample platform. This is achieved from the freely-moving electrons which strike the sample platter causing a build-up of negative charge and results in a large negative bias voltage. Because the plasma is then deficient of electrons, the overall charge of the bulk plasma becomes slightly positive (due to the population of positive ions exceeding those of the electrons), whereby a drifting of these positive ions towards the negatively charged platter occurs, and bombard the sample. As a note, because the force of the magnetic field on the charged particles depends on the charge and mass, the larger positive ions move relatively little compared to the much lighter electrons which is why these electrons are drawn out of the main body of the plasma to be carried away by the biased platter. Additionally, the sample platter can be cooled if necessary to provide extra control over etching rates.

One of the most beneficial aspects of RIE is the means to achieve a very highly anisotropic etch process due to the acceleration of the positive ions arising from the DC bias of the sample platter. This allows very high aspect ratio structures to be fabricated on a sample, otherwise unachievable with chemical etch processes which typically produce isotropic etch profiles. The amount of gas, and which gas (or gases, commonly) to use, depends on the etch process needed, of which there are two commonly used processes: mixed etch, or Bosch Process [73] etching. Whilst both of

these processes can achieve high aspect ratio structures, it is the Bosch process, also termed Deep Reactive Ion Etching (DRIE), which yields the best results. The majority of work carried out on etching is done so on silicon, as it is a vital material used in the semiconductor industry where feature sizes and controllability of fabrication processes are essential. Of the gases used to etch silicon, the two most commonly used are Sulphur Hexafluoride (SF_6) and Octafluorocyclobutane (C_4F_8), along with Oxygen (O_2) as an additive gas. Because the work outlined in this thesis has research relating to the fabrication of (and due to the expansive research and technological applications of) silicon, the chemical processes described for these particular gases will be in the case for etching of silicon.

It is noticeable that the gases used are fluorine based with large numbers of fluorine atoms forming the molecules, due to the fact that there exists a chemical reactivity between silicon and fluorine radicals. The typical result of creating a plasma from a fluorinated gas is that there will exist a distribution of ions, consisting of free fluorine radicals, electrons, and heavier fluorine-carrier ions (e.g. SF_x, where $x \leq 5$). It is the free fluorine radicals which cause the chemical etching process, by *adsorbing* to the silicon and undergoing the following dominant reaction:

$$Si + 4F = SiF_4$$
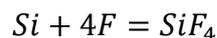
where the SiF_4 reactant then *desorbs* from the surface and is carried away via the vacuum pump in a process similar to that shown in **Figure 3.7**. Additionally, the

heavier fluorine-carrier ions may bombard the silicon substrate to cause physical

sputtering of silicon atoms by means of kinetic energy transfer (shown in Figure 3.7

also). An increase in the density, or amount, of the SF_6 present in the chamber will

yield a higher concentration of free fluorine radicals (and also positively charged

carrier ions) which would result in an increased etch rate of the silicon substrate. In

a similar fashion, altering the power of the RF matching units and/or the bias voltage

applied to the sample platter would result in the etch rate being altered.

One issue that may arise when using an etchant gas is the undesirable lateral
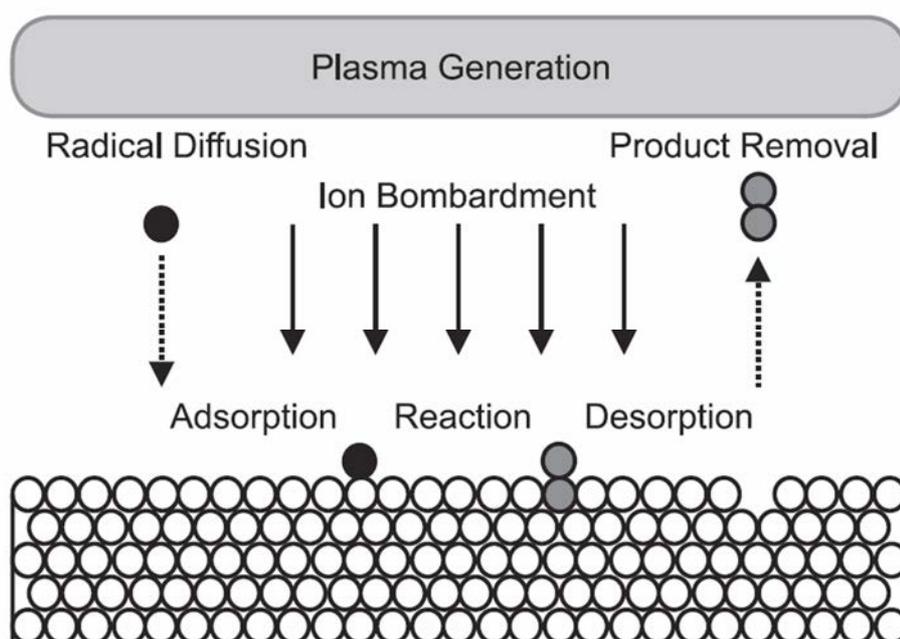
***Figure 3.7: Schematic representation of the etching
methods present in a RIE process*** [74]

The ions present in the main chamber of a RIE system can etch the substrate
through two mechanism: the first is by adsorption of the light radicals
(typically fluorine), which then undergo a chemical reaction with the
substrate, and the reactant product is desorbed from the surface; the second
is by physical bombardment of the heavy ions (which have been stripped of
the lighter radical) which transfer some of their kinetic energy to result in
sputtering of the surface atoms. *Image from* [74].

etching of the substrate, due to the isotropic component of the etch process from the spontaneous etching of fluorine radicals with silicon. Even with the processing parameters at our disposal, there may not exist a combination which allows full controllability over the anisotropy of the etch process. This is where we can utilise other gases to *inhibit* the effect of the SF6 etchant gas – the two most commonly used gases for this are Oxygen (O_2) [74] and Octafluorocyclobutane (C_4F_8) [75] (or both combined) as mentioned previously. Firstly, oxygen is used as an inhibitor due to the fact that it binds strongly with silicon to form silicon dioxide (SiO_2) – this silicon dioxide forms a protective layer preventing a chemical reaction with the free fluoride radicals, and only by physical sputtering of these surface SiO_2 adsorbents will the bulk silicon then be exposed. This inhibition is useful because it prevents the spontaneous etching of the silicon. Only the physical sputtering by ion bombardment can remove the oxygen adsorbents, and because this bombardment is highly direction (in the vertical direction because of the negative platter bias) the chemical reaction process for etching only occurs where physical bombardment has occurred – thus, a level of anisotropy is achieved. With an increasing etch depth, the oxygen *passivates* the sidewalls, which again can only be removed by ion sputtering, and so the anisotropy is self-sustaining, where an example of the high-level of anisotropy of oxygen as a passivation gas is shown in **Figure 3.8**. However, it should be noted that this process of oxygen passivation works best at lower temperatures (well below 0°C), which is beyond the scope of the work carried out in this thesis. As a note,
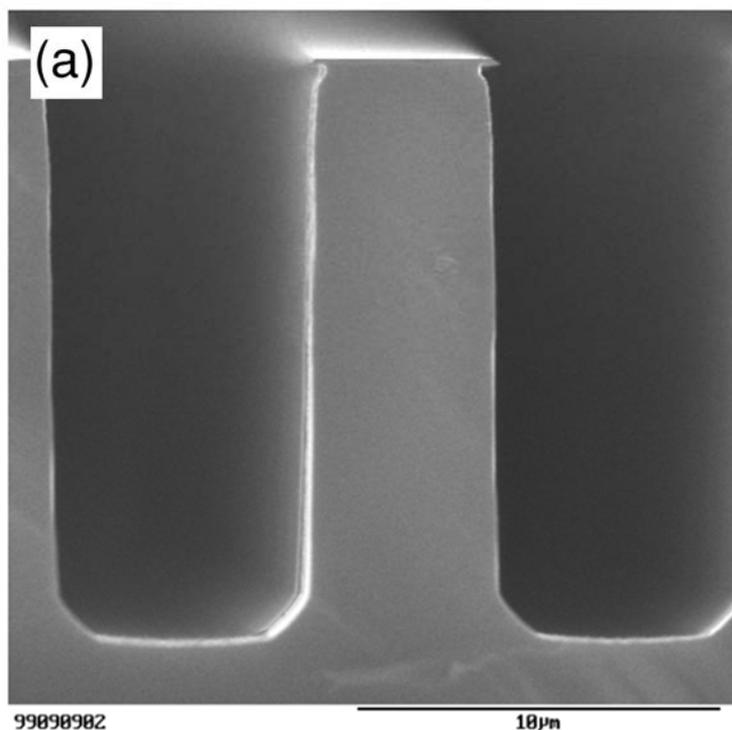
***Figure 3.8: DRIE etch profile of silicon with SF6 when using Oxygen
as a passivation/inhibitor gas to obtain vertical sidewalls*** [74]

oxygen plasma is also used to clean samples which may have organic compounds

present as the typical reactions yield carbon monoxide (CO_2) and thus desorbs from

the surface. This procedure is typically used following the development of a

photoresist and prior to metal-deposition (as described in the proceeding sub-section

'*3.1.4 Metal-Deposition*') so that any remaining resist after development and cleaning

process is removed and the deposited metal can then have the best surface adhesion.

Because oxygen does not readily etch silicon it is safe to do this process, provided

the ion energy and power

is low enough not to cause substantial physical sputtering but high enough to remove

any carbon based residues, and without the worry of the exposed silicon areas being

etched to a different depth to the resist-covered areas.

Due to the necessity for low temperatures when using oxygen as a passivation

gas, it will not be the primary means of achieving anisotropic etch profiles and instead

will only be used in small amounts alongside the fluorocarbon Octafluorocyclobutane

(C_4F_8). This gas is used to passivate the surface of the substrate with a 'Teflon'-like

polymer layer [75–78], which performs in a similar manner to the oxygen passivation.

This prevents the free fluorine radicals from spontaneously reacting with the silicon

surface and only physical sputtering of the C_4F_8 molecules allow access for reactive

etching. The benefit of using C_4F_8 is that it can be used at room temperature and

the fact that it develops a thin film over the substrate, which would yield better

Basic Bosch Process

Passivation   **Etching**
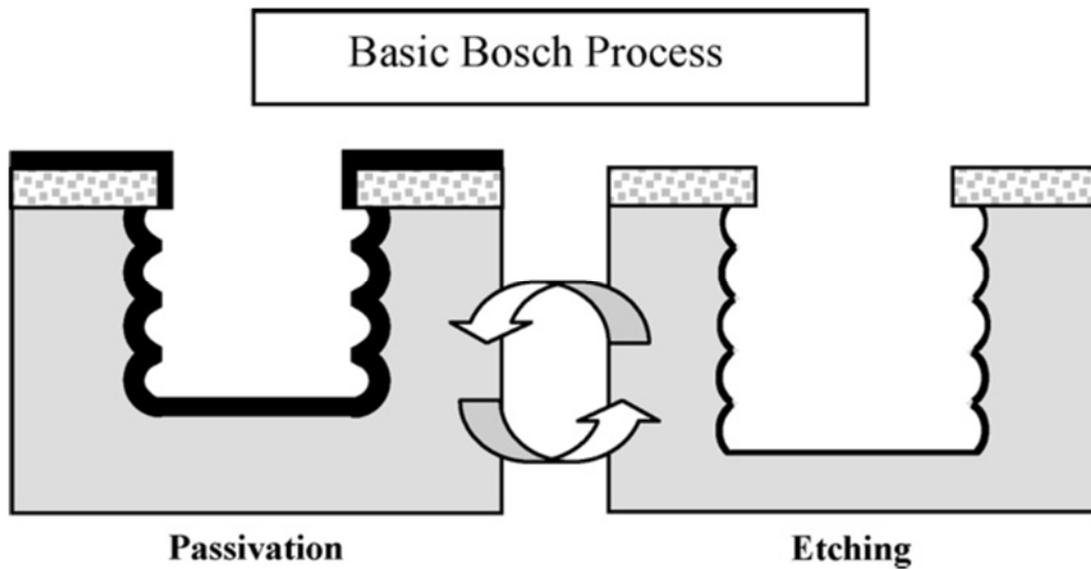
Figure 3.9: Schematic representation of the Bosch Process for achieving DRIE [76]

The well-known Bosch process [73] works by a cyclic etch-passivation process, where the first step deposits a thin film of C_4F_8 followed by an etch step using SF_6. The undulations of the sidewall are due to the isotropic nature of SF_6 when etching, and the fact that C_4F_8 is primarily directional and so the base of the etched trench will have a thicker layer of C_4F_8 than the sidewalls.

uniformity compared to oxygen (which doesn't form a film but is scattered around).

As mentioned previously, when using SF_6 and C_4F_8, there are two primary etch processes that can be used, which are either mixed etch or Bosch Process. Mixed etch processing is where both SF_6 and C_4F_8 are added to the RIE chamber at the same time and a subtle balance between the amounts of these two is required to obtain anisotropic etching – if too much C_4F_8 is used then the SF_6 won't be able to etch the silicon at a fast enough rate, whereas if too little C_4F_8 is used then the etching will be much less anisotropic. Unwanted interplay between the plasma products from these two compounds may also occur. A more preferable method is the Bosch process, which is a well-established patented two-step process developed by Robert Bosch GmbH [73] as shown in **Figure 3.9**, where passivation by C_4F_8 is carried out for the first step, followed by etching using SF_6 for the second step; this is then cycled for a set number of times to obtain a desired etch depth. Each step can be controlled separately, and so provides a good controllability, where step times, gas flow-rates, platter bias, and RF matching power can all be altered. Firstly, a step for C_4F_8 passivation is carried out, where it forms a thin layer across the surface of the patterned photoresist and substrate – some will coat the side of the photoresist where voids are present. Next, the C_4F_8 input is stopped and SF_6 allowed into the chamber where physical sputtering of the base of the patterned voids takes place, leading to the substrate being exposed to fluorine radicals which can chemically react and lead

Table 3.2: Typical Bosch Process used for DRIE of Silicon

Process Type	Cycle time	SF$_6$	C$_4$F$_8$	O$_2$
Passivate	5s	0 sccm	100 sccm	0 sccm
Etch	8s	130 sccm	0 sccm	13 sccm
Etch depth	~30µm			
Total time	9min 58 secs (46 cycles)			
Etch rate	~50 nm s^{-1}			

to etching. It should be noted that the physical sputtering may also remove the sidewall coverage of the C$_4$F$_8$ and hence isotropic etching by the fluorine radicals occurs, which gives a *scalloped* profile; this is especially apparent when these two steps (passivation/etch) steps are cycled. This scalloping can be seen below in **Figure 3.10a**) and undulates down the sidewall with each concurrent two-step cycle. Even though the platter is biased, and the majority of ions vertically accelerated, some of these will be incident upon the sample at some angle. However, with increasing etch-depth, and therefore aspect ratio, these ions can only reach the bottom of the trench if they have smaller and smaller incident angles as otherwise they would strike the sidewalls before reaching the base. This is apparent in **Figure 3.10b** where a narrowing of the trench occurs as it gets deeper. Additionally, the fluorine radicals which make it to the bottom of the trench may undergo a reaction, but the desorbed etch product may not be able to leave the trench with the incoming flux of ions.

There will therefore exist a maximum etch depth that can occur before the continual

narrowing leads to a finite trench width and etching will no longer occur. A typical

recipe and process for deep etching of silicon (30μm) using the Bosch Process is shown

in Table 3.2 (to produce the herringbone structure shown in the next chapter for).

One issue that can occur is when the photoresist is not fully removed and

small amounts remain on the sample needing to be etched: these remnants can then

act like 'mini-etch-masks' causing unwanted and spurious spikes or 'grassing' to occur

which will degrade the performance of devices (as occurred for the removal of SU8-
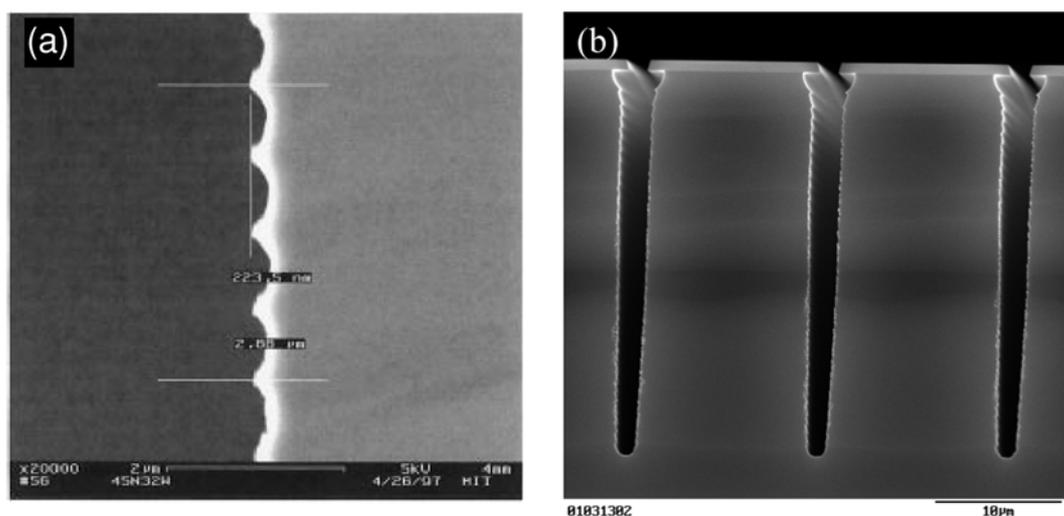
Figure 3.10: Scalloping effect of the sidewalls in the Bosch Process
[79] ***and the narrowing of the trench with increasing depth*** [74]

(a) The 'undulating' scalloping effect due to step-cycling between passivating
and etching for the sidewall profile when using the Bosch Process [79]. **(b)**
When very high aspect ratio structures are required the etch mechanics of the
ions and desorbed products come into play, due to the fact that those entering
or leaving at smaller projectile angles are successful at reaching (or leaving)
the base of the trench, whilst those with larger angles will strike the sides.
This is more pronounced near the top of the trench, as witnessed by the much
larger scalloping effect [74].

2000 resist in Chapter 4 for the fabricated herringbone structure in). The same thing

can also occur for when very deep trenches are etched and the volatile products can

no longer escape the vertical ascent which will result in re-deposition and act like

miniature masks. Care should be taken to avoid these issues from occurring, and for

etching parameters to be carefully chosen.

3.1.4   Metal-Deposition and Lift-off Process

An alternative route which can occur after photolithography is that of Metal-

Deposition, rather than etching of the substrate. This infers a direct hard-mask of

metal to fill the voids present in the developed photoresist. Depending upon the tone

of the photoresist, the metal patterning can be either the inverse pattern (for positive

resist) or the replica pattern (for negative resist) of the photomask used in the mask

aligner. Such a procedure is typically used for performing lift-off process, where the

metallic structures that remain are the desired outcome or device, such as for

plasmonic antennae or Split Ring Resonators [80]. However, the metal structure can

also be used as a hard-mask to protect areas from DRIE. The metallic 'caps' may

also be functional or simply provide greater resistance to the DRIE process to obtain

massively-deep etches. Two types of metal deposition procedures are regularly

available – either through ***Thermal Evaporation***, with a small amount of low

melting-point metal, or ***Sputter Coating***, from a metal source.

Thermal Evaporation of a metal is carried out through means of locally heating

a small amount of metal to temperatures high enough to cause melting, whereby energetic atoms of this metal then escape and condense on cold surfaces undergoing solidification. Metals used for evaporating in this process are those which usually have low melting points, such as Gold or Silver – this is due to the method in which thermal evaporation is carried out. A small '*boat' (or crucible),* consisting of a strip or wire of high melting point metal e.g. Tungsten, is used to contain the evaporant metal; the boat is shaped to have a small hemispherical indent in the middle into which small amounts of the evaporant metal can be placed. This boat is connected into an electrical circuit, via terminals at each end, which allows a current to pass through the metal. It is known that for a thinner wire the resistance increases, and in turn the Joule Heating increases. The heating of the boat can thus be controlled by altering the current supplied through the terminals into which it is connected, whereby a larger current will result in a larger heating effect. So as not to damage the boat used to contain the evaporant metals the current should not be increased too greatly; this infers the necessity of using a high-melting point metal for the boat such that it too doesn't evaporate with the softer low-melting point metal. The sample needing to be coated is typically placed directly above the evaporating metal source. A schematic diagram of a thermal evaporator system is shown below in **Figure 3.11a** and a photograph of a BOC Edwards Auto 500 evaporator system is shown in **Figure 3.11b**.
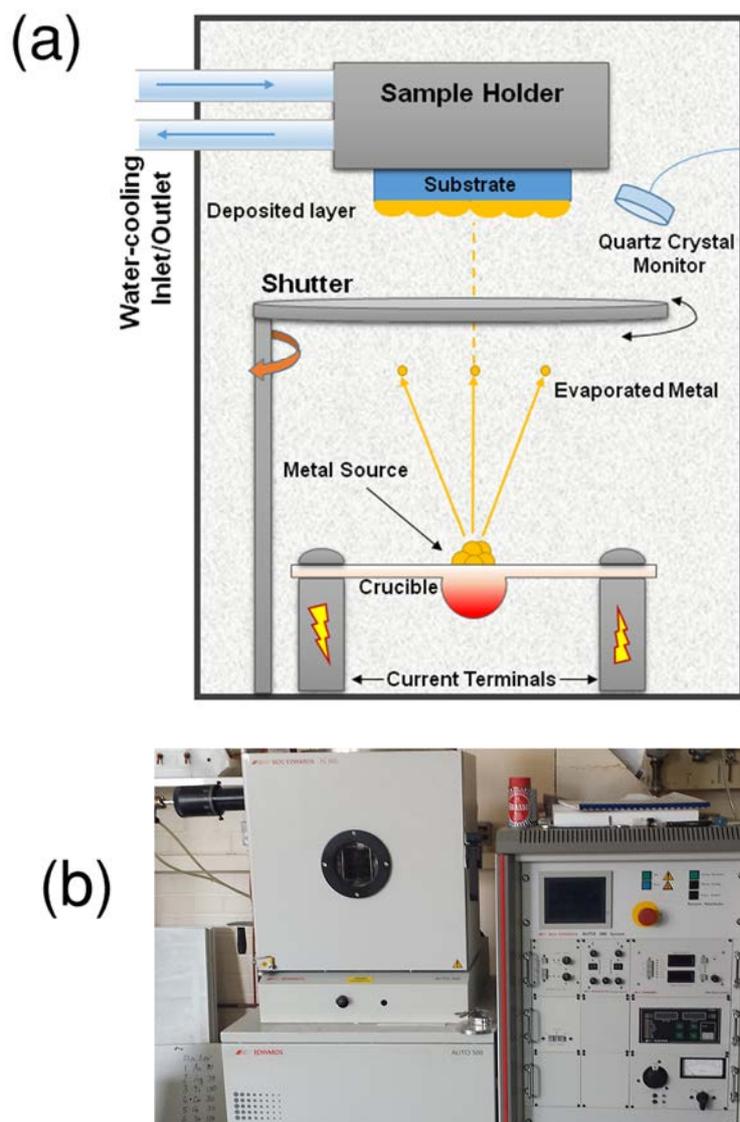
Figure 3.11: (a) Schematic Diagram of a Thermal Evaporator System and the (b) BOC Edwards Auto 500 Evaporator system
(a) A sample is affixed to the water-cooled sample holder. Current is applied to terminals which are connected by the crucible containing the evaporant metal. The crucible is heated, and a shutter used to allow the evaporated metal to strike the sample. The thickness of the metal layer is monitored by a quartz crystal monitor close to the sample. **(b)** BOC Edwards evaporator system, with the vacuum-chamber on the top-left (with the circular window).

The vacuum chamber is normally pumped down to as best a pressure as possible. This is typically of the order of ~10^{-6} mbar, but it is common practice to leave the system pumping down overnight (after inserting the sample and source

metal) in which the pressure can then reach its best value, which is preferably ~10^{-7} mbar. When heating the source metal, it is best to maintain the current at low values first in order to desorb any contaminants which may be present in the crucible and also adsorbed to the metal; these may include gases and water vapour. When doing this, it is crucial that the shutter is closed (as shown in Figure 3.11) which is used to prevent thermal evaporants from striking the sample. It can be assumed that contaminants are being desorbed by keeping an eye on the chamber pressure – an increase in the pressure from the inactive pressure indicates that gases are being desorbed from the crucible. Once the pressure then starts decreasing, it can be inferred that most of the contaminants have been successfully removed from the crucible and metal source, although it is good practice to continue slowly increasing the current and monitoring the increase/decrease pattern.

Once the metal is ready to be evaporated, it will be heated up to the point at which melting occurs and a bright glow observed through the window aperture (seen in **Figure 3.11b).** The shutter can then be opened and the evaporating metal allowed to travel towards the sample. As the temperature will be very high, the sample holder is water-cooled to maintain the sample at a constant temperature; high temperatures may damage or warp the photoresist present on the substrate so cooling is necessary. In order to know the thickness of the layer deposited on the sample, there is a quartz-oscillator monitor near to where the sample-holder is situated which

measures the deposition rate and can therefore deduce the total thickness of metal

deposited on the sample as being approximately equal to the amount deposited on

the quartz. From actual experimental verification, the quartz monitor can be

calibrated to give the best accuracy for layer thickness. To produce a nice uniform

layer of gold, the current is maintained low enough that the evaporation rate is

between 0.04-0.10nm s^{-1}, as if the rate is too high then clustering may occur and the

deposited metal layer will have an increased roughness. However, different metals

(e.g. silver) may require different evaporation rates to this, for various reasons. Once

the desired thickness is achieved, the shutter is closed to prevent any more deposition

and the current source slowly decreased to zero. The vacuum chamber is then purged

of any remnant gases present for approximately 15 minutes (and for the sample and

sample holder to be cooled) before the sample is removed.

Once the sample has been coated with the desired metal, and the patterned

voids in the photoresist filled with this metal, we can then perform lift-off process of

the resist to leave behind only the metallic structures on the substrate. An image of

a gold SRR array, for application in the work carried out in Chapter 5, which was

performed by lift-off, is shown in **Figure 3.12**. Lift-off process takes place by

submerging the metal-coated sample in a solvent, which is typically acetone or IPA
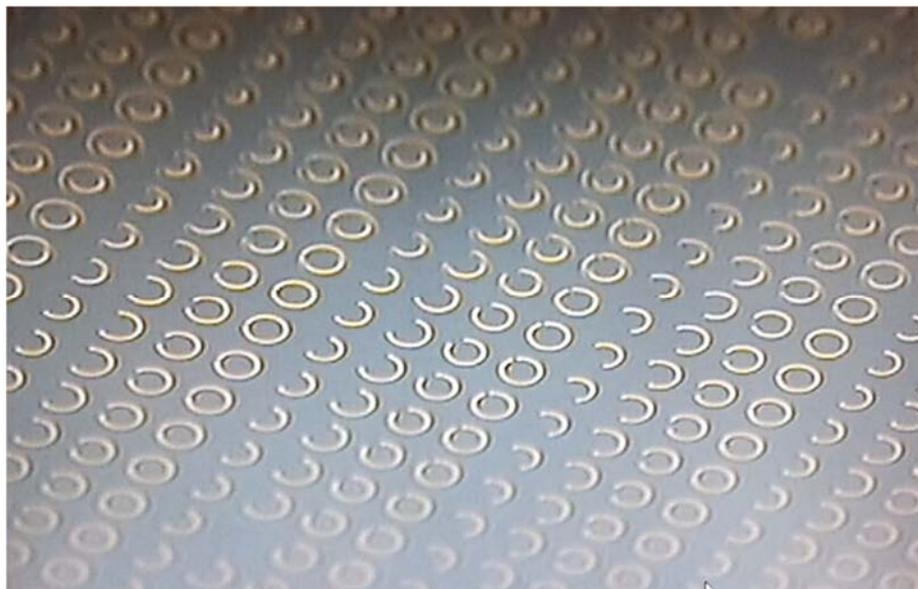
***Figure 3.12: Optical Microscopy image of gold SRR's
fabricated using lift-off process***

Split-ring Resonator structures deposited on a silicon substrate performed by
deposition of gold (200nm thick) onto a pre-patterned, developed photoresist,
and subsequent lift-off processing in acetone solution.

as most photoresists are soluble in either of these cleaning solvents. In much the same

way as development, the samples are agitated in the solvent, sometimes with the aid

of an ultrasonic bath, to aid in the removal of the photoresist and metal which sits

atop the resist. However, because the metal films are sometimes quite thick, it may

be necessary to leave the samples submerged in a solvent overnight to minimise

damage to the structures by too much agitation. When making structures using soft

metals, such as gold, it is sometimes preferable to first deposit a very thin layer (~10-

20nm) of a stronger metal, such as titanium or chromium, which helps to aid in the

adhesion of the gold structures to the surface and minimise damage or removal of

these when performing agitation.

3.2   Characterisation

In order to examine the samples which are fabricated for both quality control and optical data acquisition, we must characterise them with a number of experimental methods. This section aims to explain some of these systems and how they help to investigate the samples which are fabricated. As most of the fabrication, which was carried out by myself, in this thesis is using photolithography – and thus micron scale structures – only a very brief coverage of Scanning Electron Microscopy and optical bench setups (where data acquisition and optical characterisation was obtained by more specialised colleagues) will be given, but are included in this section for completeness.

3.2.1   Optical Microscopy

Arguably the earliest, and still one of the most useful, methods of characterising small scale objects is optical microscopy. It is a corner-stone of fabrication, especially so for micron-sized structures, which are too small to see with the eye alone. In the case of photolithography, optical microscopy helps to position the samples relative to the mask. This is crucial in the case for alignment between a pre-existing pattern on the sample and a secondary pattern on the mask, as was the case for the work carried out in Chapter 4.

Seen in **Figure 3.13** is a photograph taken through the eyepiece of an optical

microscope whilst performing a photolithography alignment – the image shown is of the herringbone pattern on the photomask. The smallest feature sizes on the mask were of the order of ~10µm and so diffraction wasn't an issue for visible light, therefore all of the features could be resolved with great detail. Microscopy was also very useful for examining the final production of lift-off processed samples, and to check for any damaged or 'loose' structures that would have occurred in process (as seen in Figure 3.12).
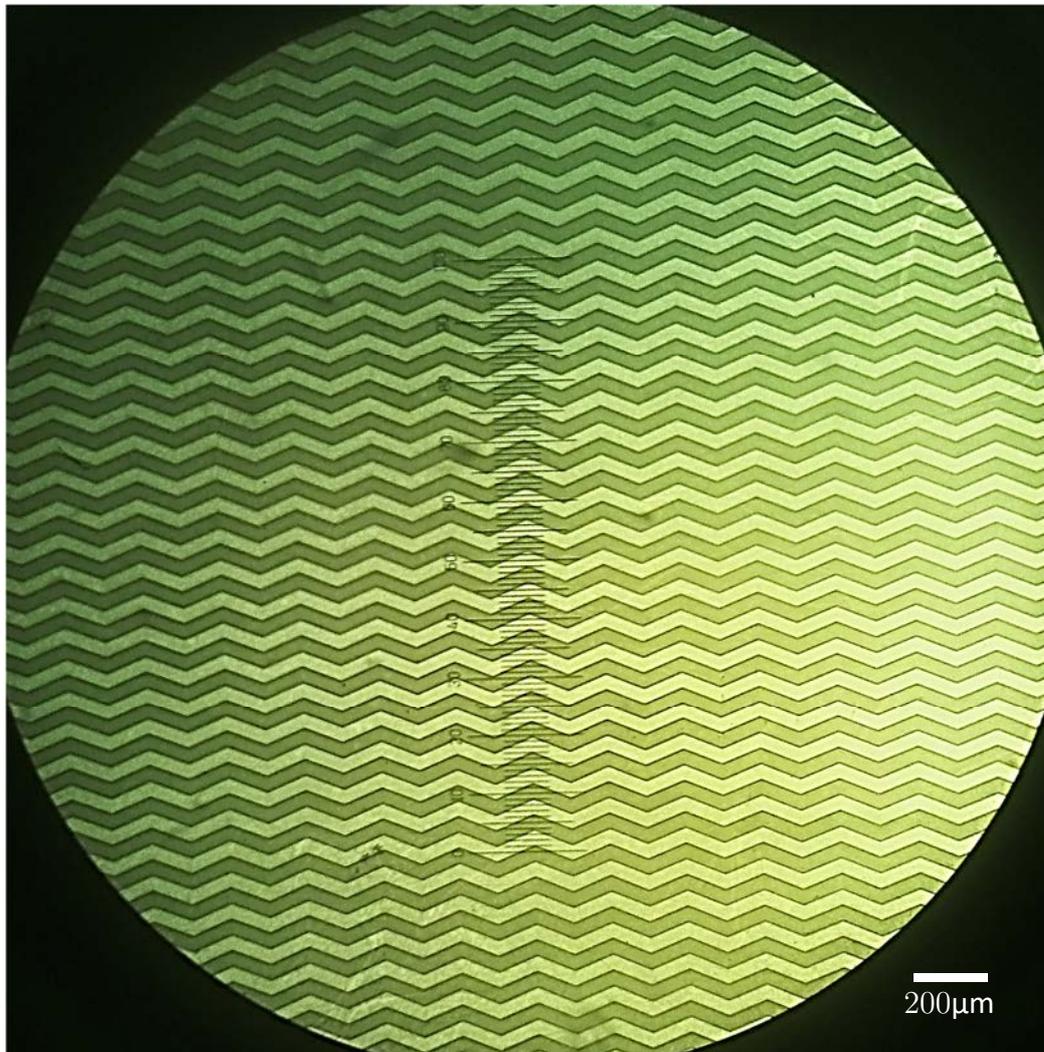
Figure 3.13: Photograph taken performing Optical microscopy of the herringbone pattern on a photomask

3.2.2   Scanning Electron Microscopy

When dealing with structures much smaller than the wavelength of visible light, we can no longer rely on optical microscopy due to severe diffraction effects occurring which causes a loss of resolution. Hence, we can no longer resolve the structure or quality of these objects whilst using optical methods. A preferable means of viewing such small scale structures is by using a *Scanning Electron Microscope* (SEM). Typically used to image conductive surfaces (due to the fact they consist of free electrons) a beam of focussed electrons, incident onto the sample, are scanned across and the deflected or excitation-ejected electrons are examined by a detector to provide details of the surface morphology. A schematic diagram of an SEM is shown below in **Figure 3.14.**

An SEM consists of a vacuum chamber into which samples to be examined are loaded. The electron source is typically a Tungsten filament which has a very sharp tip, through which a high voltage – and ultimately a potential difference great enough to overcome the free space impedance – is supplied. Vacuum pressures are preferably as low as possible to prevent oxidation of the tungsten tip, and to allow a mean-free-path as large as possible for the emitted electrons. Through various lenses, the electrons are focussed to point on the surface of the sample. These electrons, upon passing through these magnetic lenses, are accelerated to thousands of volts, whereby the impact with the surface will cause loss of kinetic energy and the re-emission of

secondary electrons, which are collected by a detector. Due to the fact that the

surface of the sample is probed through electronic interactions, and not light-

Field Emission Gun

Anodes

1st Condenser Lens

2nd Condenser Lens

Objective lens

Secondary Electron Detector

Scanning Coils

Sample
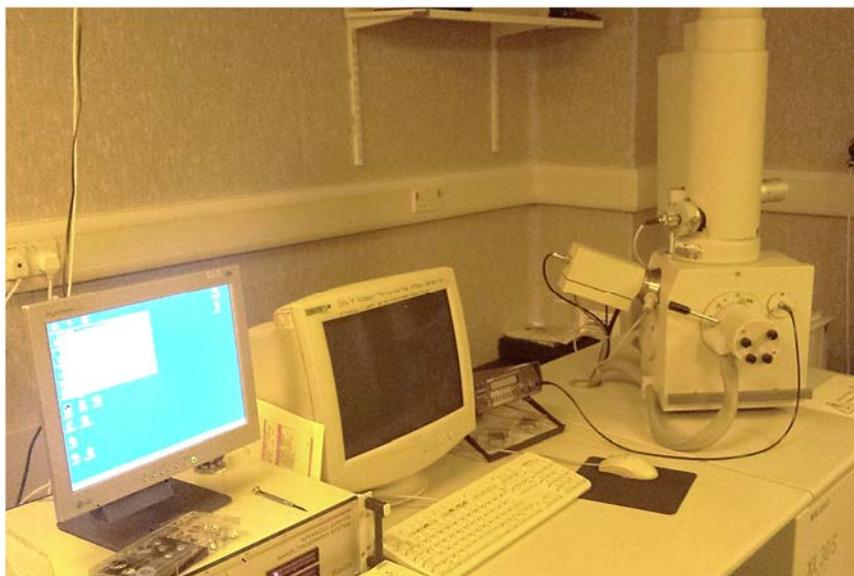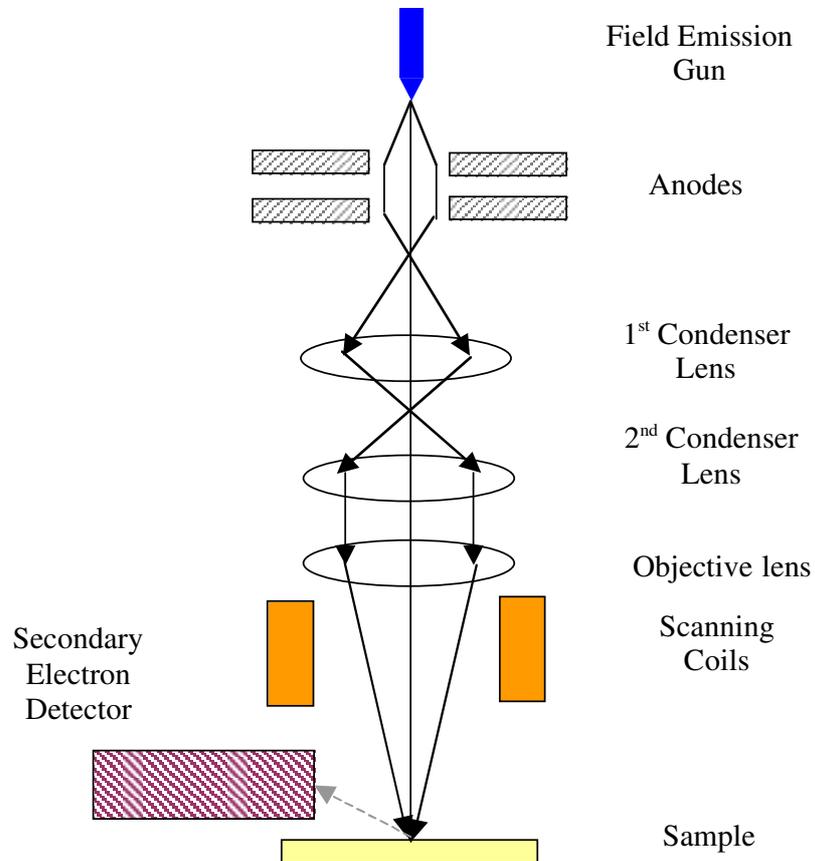
***Figure 3.14: Schematic diagram and photograph of a
Scanning Electron Microscope***

focussing, the resolution is much greater than that possible for visible light and can

be used to image features as small as 10nm.

In addition to being used to image surfaces, an SEM system can also be used

to fabricate structures through *Electron-Beam Lithography* (EBL), which acts in a

similar manner to that of photolithography where an e-beam sensitive resist is

patterned. However, this is a slow process compared to photolithography (which

allows a parallel exposure of the whole resist at once when covered with a suitable

mask) as each structure has to be exposed step-by-step as a *bottom-up* approach,

rather than top-down. Regardless, EBL still remains as one of the most useful and

widely used means of fabricating nano-sized patterns to date.

3.2.3   Surface Profilometry

In order to measure the heights and widths of fabricated structures, it was

necessary to make use of a surface profilometer. The system used was a Dektak³ ST

profilometer as seen in **Figure 3.15.** A surface profilometer works by gently

'scratching' the surface of the sample with a stylus, and variations in the height of

the stylus whilst scanning the sample are converted to signals allowing the user to

graphically see the surface profile. In order to locate the specific topography or feature

on the sample that is wished to be examined, a microscopic camera contained within

the system relays the image to a screen. An example of this imaging and the resultant

graphical profile is shown in **Figure 3.16,** where the sample scanned is of a RIE

Figure 3.15: Dektak³ ST Surface Profilometer

herringbone pattern as used in Chapter 4. The inset of the lower image of Figure 3.16 (in the yellow circle) shows the details of the parameters used to obtain the surface profile. The length of the scan (indicated by the red arrow on the top image) can be chosen by the user, as well as the scan time, stylus force, data resolution, desired height range, and the predicted 'step' profile (as shown by "hills and valleys"). These make the surface profilometer a very useful piece of equipment for characterising samples, and is essential for measuring, for example, the thickness of a photoresist, the lateral distance and height of etched features on a silicon substrate, or the thickness of deposited metal structures – which is very helpful for calibrating the equipment used for metal evaporation (as explained in Section 3.1.4). The vertical resolution of the stylus can be as low as 1nm (realistically) whilst the lateral resolution is somewhat lower due to the finite sharpness of the stylus tip – if the lateral distance between features is too small, the stylus tip cannot get to the bottom of the gap before the increasing stylus-radius contacts the corners of the structures.
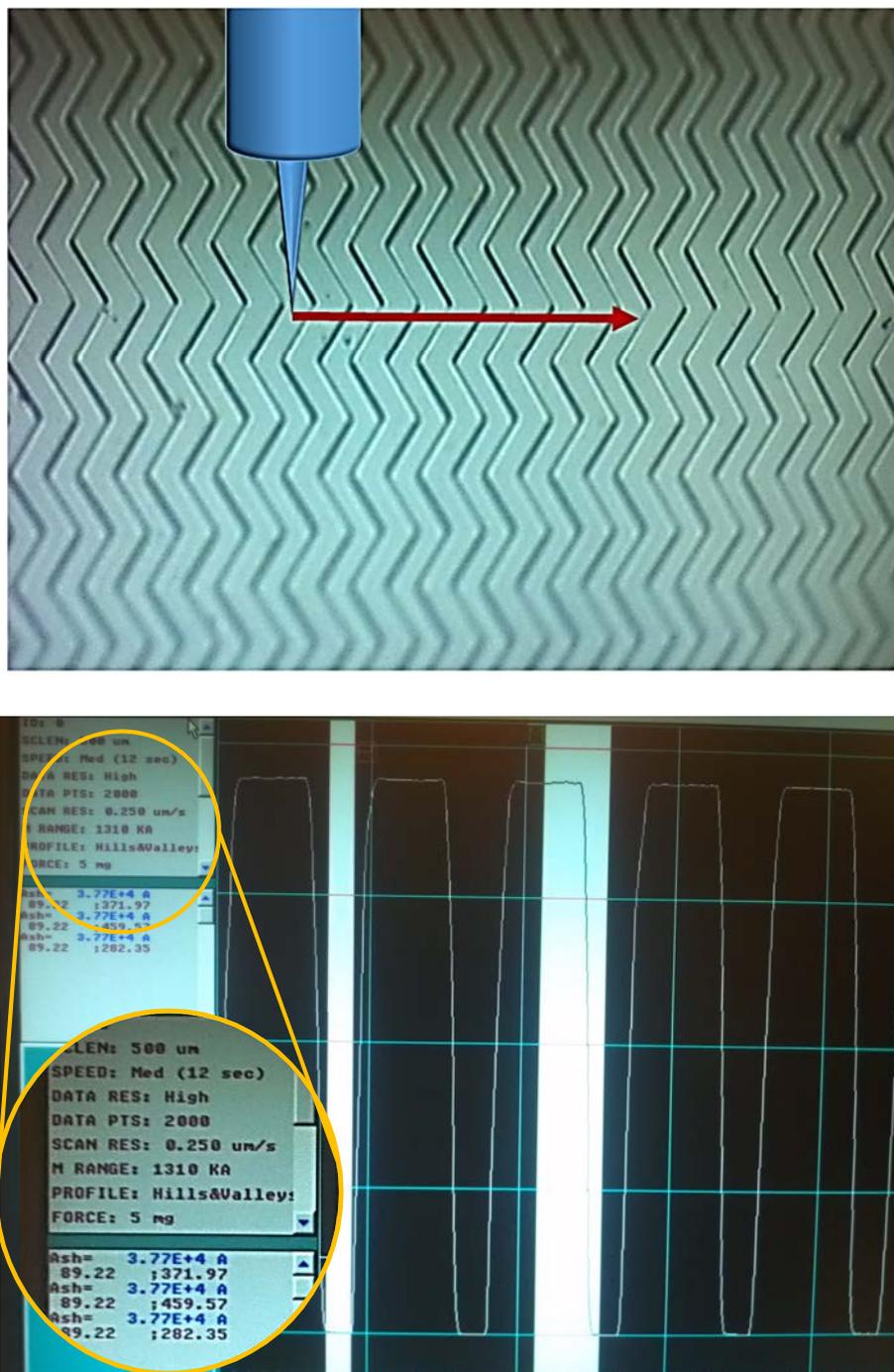
Figure 3.16: Camera image and corresponding surface profile of a section of etched herringbone pattern

Top: the stylus makes contact with the surface and follows the red arrow.
Bottom: the vertical deviations of the stylus when moving across the sample
are relayed to a computer and the graphical profile is plotted, where two
moveable selection regions allow normalisation of the height and ultimately
the differences in height between high and low regions of the profile.

This is especially noticeable when the aspect ratio of the features are very high (i.e. the vertical height is much greater than the lateral width). The dimensions of features are determined by the control of two variable width cursor regions. These regions are used to take averages of the profiles contained within their edges, and so can be used to normalise the base flatness (by having both of these regions only selecting distant 'low' areas), and then the difference in heights between the low and high regions can be deduced based on this normalisation.

3.2.4   Terahertz Time Domain Spectroscopy (THz-TDS)

Terahertz technologies are much sought after during recent years due to the lack of natural (and also engineered) means of generating and detecting this frequency range of light. Many natural phenomena have subtle terahertz resonances occurring, and so it is difficult to investigate these processes without the necessary equipment to probe. The so-called "Terahertz Gap" [81] has inspired the aim to rectify this issue and to develop devices which can perform equally as well as the counterpart devices in the various other frequency regimes of light – such as the plethora of developed technologies in optical, microwave, and X-ray frequencies.

Many of the metasurfaces designed to work at Terahertz frequencies rely on characterisation with Terahertz Time-Domain Spectroscopy (THz-TDS) systems, where a MenloSystems THz-TDS system [82] was used for obtaining the data for the work carried out in this thesis – a schematic diagram of the experimental setup is
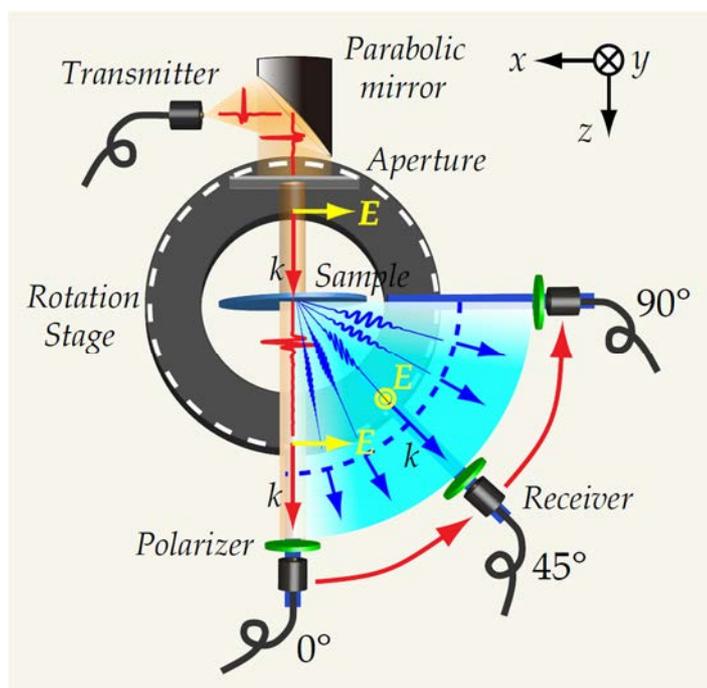
***Figure 3.17: Schematic diagram of a THz-TDS
experimental setup (Image from** [83]**)***

A typical Terahertz Time-domain Spectrometer (THz-TDS) experimental
setup for measuring transmitted THz signals through a metasurface sample,
where the THz beam is refracted according to the Generalised Snell's Law.

shown in **Figure 3.17**. The basic operation of a THz-TDS system is given for the

case of a THz wave being transmitted through a phase gradient metasurface which

refracts the incident beam into anomalous orders [83] which is related to the work

carried out in Chapter 5. A Femtosecond laser is pulsed onto a pair of

Photoconductive Antennas (PCAs), sat atop a semiconductor, whereby a bias voltage

is continuously applied to these PCAs to provide a potential difference in the gap

between them. When the laser pulse hits this gap the energy provided allows the

combination of an electron and a hole, and a Terahertz wave is emitted upon

relaxation of this electron-hole pair. This initial femtosecond pulse is split and the

probe beam passed along a variable path. The PCA emitter is placed at the focal

point of a parabolic mirror, such that the THz beam emitted is then reflected as a parallel beam (seen in Figure 3.17) towards the sample. Following the mirror, an aperture is used to define the beam size for improving the spatial frequency resolution (in the case for Figure 3.17 the aperture size was 1.0cm in diameter). This beam is then incident onto the sample, where the desired optical effect occurs. After interaction with the sample, the THz beam will be deflected into some arbitrary angle: to detect this, another PCA device, termed the receiver, is fixed to an arm attached to a rotatable stage capable of angularly resolving the refracted beam. In addition, the sample itself is mounted on a smaller rotatable stage atop this larger stage so that non-normal incident beam angles can be investigated. Prior to this PCA receiver is a wire-grid polariser, capable of filtering the linear polarisation component of the refracted beam. Finally, both this THz beam and the initial femtosecond probe-pulse arrive at the receiver. The probe-beam will charge the surface of the semiconductor under-layer of the PCA, whilst the electric field of the THz beam will induce a current – the strength of this current will be proportional to the amplitude of the time dependent THz beam, and thus we can deduce the THz beam signal after it has interacted with the sample by taking a Fourier Transform with respect to a reference sample (usually bare substrate material). This Fourier Transform will give us the frequency response of such a THz metasurface. Additionally, we can insert wire-grid polarisers before the sample, and after the parabolic mirror, to excite the metasurface sample with a required linear polarisation. Another wire-grid polariser
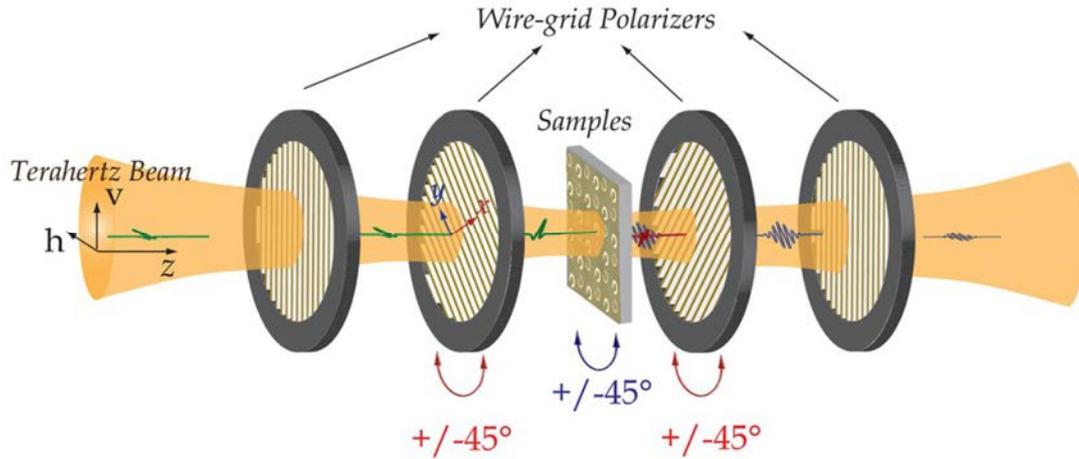
Figure 3.18: Schematic diagram of a THz-TDS experimental setup (Image from [84])

The four wire-grid polariser configuration to provide capability of measuring the Jones Matrix components for linearly polarised light. These components can then be extrapolated to obtain the Jones matrix components for circularly polarised light.

is placed with the original polariser after the sample so that we have the capability

to measure the linear Jones Matrix components of the metasurface, as shown in

Figure 3.18, which in turns enables us to investigate the circular polarisation basis

Jones matrix components. The 1st and 4th polarisers are oriented in the same

alignment to provide signal normalisation, whilst the 2nd and 3rd polarisers are freely

rotated, as shown, to enable the independent incidence and transmitted x and y

polarisations to be measured. In conclusion, the THz-TDS system, in conjunction

with the four polariser arrangement, is a useful tool for investigating the frequency

responses of metamaterials whilst being able to distinguish the linear polarisation

components necessary to formulate the Jones Matrix of the metamaterial device.

Chapter 4

Silicon Herringbone Metasurface for Giant Chiroptical Response of Terahertz Light

Manipulating the polarisation of light is crucial for holographic and imaging applications. One such aspect in particular is controlling the orthogonal circular polarisations of light separately, such that there is an asymmetry between these when transmitted through a device. However, all present methods of achieving this rely on lossy and complex multilayer structures to achieve a high selectability of cross-polarisation conversion [85,86]. This work proposes a monolithic silicon herringbone metasurface to realise a strong cross-polarisation conversion for circularly polarised light, with T_{RL} exceeding 80%, whilst prohibiting the inverse cross-polarisation for T_{LR}. The device is purely dielectric, so has negligible losses, and relies on the combination of geometric-phase in conjunction with subwavelength gratings to invoke both angular disparity and form-birefringence, which leads to a preferential cross-polarisation transmission arising from interference effects. The design is robust

and not overly sensitive to fabrication errors, whilst also exhibiting a broadband operation in the Terahertz regime.

In this chapter, I shall discuss the development of this dielectric silicon-based metasurface used to achieve an asymmetry in the conversion efficiencies between incident and transmitted circularly polarised (CP) light in the frequency regime of 1THz. Form-birefringent subwavelength gratings (SWGs) were proposed to act as half-wave plates, for CP handedness switching, and then arranged in a 45° *herringbone* pattern to provide the necessary geometric phase of $\pm\pi/2$ to induce an asymmetry in the transmitted efficiencies of the converted CPs. A final addition of a dielectric spacer for one of the two angled SWGs was incorporated to provide a dynamic phase of $+\pi/2$ such that the final total phase values are either 0 or π for the flipped transmissions from left or right incident CPs, respectively. The result obtained surpasses previously reported efficiencies for Circular Conversion Dichroism, where a potentially obtainable conversion of 80% from left to right CP light is achieved.

Theoretical framework, analytical modelling, numerical simulations, sample fabrication/imaging/integrity, and data analysis were performed and carried out entirely by myself, with assistance and guidance from Prof. Shuang Zhang for the theoretical and analytical work, whilst optical characterisation and data acquisition were carried out independently by Dr. Teun-Teun Kim and Xueqian Zhang.

4.1   Motivation

The need for on-chip and integrated optical devices is ever growing due to the wide range of technologies now utilising light for various applications, where more efficient and faster data capabilities are necessary to progress. Typical methods of utilising light is achieved with rather dated and bulky technologies, which, although very capable and suitable, aren't desirable due to a number of reasons, such as high cost, large and bulky size, low efficiency, and the magnitude to which the phenomena can be achieved, due to natural materials normally only exhibiting small optical effects.

Metamaterials have been a promising and effective contribution to the fields of optics and electromagnetism, where they provide a robust and tailorable means of controlling optical phenomena to a much greater degree than naturally occurring materials [24–36]. However, the fact that Metamaterials are still a *bulk* arrangement of periodic structures is an issue for scalability and compactness, meaning that such devices have not been adapted for practical applications. More recently has been the investigation into monolayer (2D) metamaterials, namely Metasurfaces [40–42,65–69]. These have the benefit of being more easily fabricated, have smaller associated losses, and, more importantly, can be better incorporated into real-world applications. This is, of course, very beneficial to the aforementioned necessity for technologies requiring integrated optical devices.

High-efficiency has been realised for a variety of metasurface applications [69,87–90], but most of these devices are typically *symmetric* in operation such that the desired optical effect yields the same transmitted efficiency for orthogonal circular polarisations of incident light. In certain cases this is not desirable, especially when a particular handedness of light is preferable over another; The cuticles of beetles have been shown to reflect mostly Left circularly polarised light [91,92], and many sensitive drugs or molecules which have different enantiomeric forms may provide different biological functions (as is the case for Thalidomide, mentioned in Chapter 1). These enantiomers exhibit different transmission efficiencies depending upon the handedness of the incident light and therefore it is important to be able to distinguish between them. One well established technique of characterising such *chiral* samples is through the use of Circular Dichroism, which looks at the difference between the amount of Left and Right circularly polarised light transmitted through the sample. This is a well-established technique and used in many fields to examine the chirality of samples. However, it usually only examines the total transmittances for opposite CP incidences without caring about conversion between the two.

When examining the Jones matrix for circularly polarised light (Chapter 2, equation 2.33), we see that the off-diagonal terms correspond to cross-polarisation conversion. Until only recently, it was not realised how these components of the Jones matrix could be accessed – 3D chiral structures, typically the case for many of

the naturally occurring handed materials (sugar, proteins, etc.) exhibit circular dichroism and optical activity yet give equal contributions of cross-polarisation, corresponding to equal off-diagonal terms $T_{LR} = T_{RL}$. It was the seminal works on 2D planar chiral structures [93,94] which realised that these structures are phenomenologically and symmetrically different to 3D chiral structures. In the case for 3D chirality, the response for polarised light is the same from forward and backward directions – this can be understood when looking at a wound spring, and realising that it will have an intrinsic 'twist' (either clockwise or anti-clockwise) regardless of whether it is looked at from the front or back. However, for 2D chiral structures (like a spiral) the image is reversed when viewed from different directions, and so will result in a different polarisation response compared to 3D structures. These structures can access the off diagonal terms of the circular Jones matrix, for which a disparity is apparent between cross-polarisation terms such that $T_{LR} \neq T_{RL}$, yet the diagonal terms are equal. In an analogous labelling for 3D chiral structures exhibiting Circular Dichroism, the effect exhibited by 2D chiral structures is labelled *Circular Conversion Dichroism (CCD)* [86,95–98] or (*Circular*) *Asymmetric Transmission (CAT or AT)* [85,86,96,97,99–111]. There are mixed views on this terminology, with the former being more accurate a description but with the latter having more uses in the literature; the issue lies with the fact that Asymmetric Transmission has meaning in different areas of science and implies that forward and backward propagation are not reciprocal.

In many cases, both terms are used by the same author; a great deal of research upon this effect has been carried out by Eric Plum [97,100,103,112,113], where his thesis referred to it as Circular Conversion Dichroism whilst most of his publications referred to it as Asymmetric Transmission. For the sake of confusion, I shall refer to the effect as Circular Conversion Dichroism (CCD). In the early work by Zheludev [95] an anisotropic lossy planar-chiral 'fish-scale' structure was investigated and shown to exhibit CCD in the microwave region; this was attributed to the 'twist' of the fish-scale, and given a twist vector **W** which followed the well-known 'cork-screw' law (*as explained in* [95]). This work was then scaled down to work in the visible spectrum [99] and exhibited the same effect. Many such works on 2D chiral structures have taken place since this work [85,96–114] at many different frequency ranges, including Infrared (IR) [85,96,108], Terahertz [112], and microwave [86,95]. However, for most of these realisations of achieving CCD, the responses are usually very small with a cross-polarisation *difference* (termed as being equal to the difference between the modulus-squared off-diagonal Jones matrix components) of only 0.25 or less.

Recent methods have aimed to improve on this low conversion difference by utilising layered metasurfaces [85,86] in which CCD differences of 0.5 and upwards were achieved. However, these devices have very complex designs, involving time-consuming optimisations of layer-to-layer distance and impedance matching, not to mention fabrication complexity. In addition, these devices are all composed of metal

and rely on the process of impedance effects due to current flows within the structures owing to the incident handedness of CP light. This means of achieving CCD is complex and cannot be easily derived analytically, and the fact that they are all composed of metals leads to significant and unavoidable losses (of the order of 37% in [85]). To this end, it is proposed to achieve CCD using dielectric materials, such that losses are negligible, and that the device relies on interference effects between phases rather than impedance matching.

4.2   Theoretical Framework

4.2.1   Achieving Circular Conversion Dichroism

It is a well-established process of using half-wave plates to fully convert the handedness of CP light to the opposite handedness or to rotate the polarisation angle of a linearly polarised wave (with respect to the optical axis of the wave plate). These devices rely on the optical effect of *birefringence* to induce a phase delay between light polarisations travelling along its principle axes (more information on birefringence is given in Chapter 2). High quality half-wave plates (and indeed the majority of other wave plates) are typically manufactured using naturally occurring birefringent crystals, such as calcite. However, as was explained in Chapter 2, these naturally occurring crystals are typically expensive, requiring high-quality crystal structuring, and the refractive index difference requires that the thickness of these crystals are much larger than a single wavelength of light.

It is well studied that Subwavelength Gratings (SWGs) can be used to exhibit birefringence, through the application of effective medium theory. These Subwavelength Gratings are typically fabricated from a dielectric or semi-conductor substrate which would normally allow light of the desired wavelength to be transmitted without any manipulation (or absorption). However, when deep periodic gratings are etched into the substrate, the light experiences different refractive indices depending on whether the polarisation is parallel (TE) or perpendicular (TM) to the grating stripes and is essentially equivalent to the case for natural birefringent media. A schematic diagram of a subwavelength grating is given in Chapter 2, Figure 2.3). If we refer back to equation 2.53, we have that

$$\Lambda \leq \frac{\lambda}{n_{Diel}} \tag{4.1}$$

which governs the periodicity of the SWG dependent upon the wavelength of interest and the refractive index of the substrate at this wavelength. For silicon (intrinsic, Ω = 10kOhm), at a wavelength of 300μm (which corresponds to 1THz) and a corresponding refractive index of 3.418, we obtain a grating periodicity of $\Lambda = 86\mu m$. This periodicity is also ideally suitable for fabrication using photolithographic methods. The frequency of 1THz is chosen due to the demand for devices in this frequency regime, and the fact that no high-efficiency CCD devices have been understood at this frequency.

To calculate the depth of the gratings required such that there exists a half-

wave plate functionality, providing a phase-difference of π, we refer back to equations

2.54-2.67, which gives us:

$$h = \frac{\lambda}{2\pi} \frac{\Delta\Phi_{TE-TM}}{\Delta n_{form}} = \frac{\lambda\pi}{2\pi \times 1.16} \approx 0.4\lambda \qquad (4.2)$$

and as we have chosen a wavelength of 300μm we obtain $h = 129\mu m$. Such a grating

works to convert incident circularly polarised light into the opposite handedness,

with an equal response for both handedness'. However, this cannot achieve CCD

alone as the conversion between opposite handedness' of CP light is equal whereas

we require a disparity in this conversion. We propose that the disparity be invoked

through the use of geometric (Pancharatnam-Berry) phase, from the angular

orientation of space-variant SWGs, similar to the work in [38,115]. If we choose a

periodically repeating angular-orientation of $\theta = 45°$ between subsequent gratings,

such that a zig-zag or *herringbone* pattern is produced, then from [38,41,68,115] we

have that the phase is equivalent to $\Phi = \pm 2\theta = \pm 90° = \pm \pi/2$, where the + sign

corresponds to T_{LR} (RCP incidence to LCP transmission) and the – sign corresponds

to T_{RL} (LCP incidence to RCP transmission). Now, if a dielectric step of a specific

thickness is added beneath one of these paired, angled SWGs, we can supply an

additional dynamic phase term of $+\pi/2$. This will result in an interference of the

phases supplied by the geometric and dynamic phases, given by:

$$\Phi = \varphi_{Dynamic} \pm \varphi_{Geometric} = \frac{\pi}{2} \pm \frac{\pi}{2} \qquad (4.3)$$

Thus, the resultant phases can only be $\Phi = \pi, 0,$ corresponding to destructive interference (for RCP incidence), which is akin to a reflective dielectric mirror, or complete transmission (for LCP incidence), which is akin to an anti-reflection (AR) coating, respectively. A schematic diagram of the proposed Silicon herringbone structure is shown in **Figure 4.2a**, and the functionality of the interference is shown in **Figure 4.2c,d**. To calculate the thickness of the step beneath one of the SWGs, we simply use a modified version of equation 2.54:

$$\Delta\varphi_{Dynamic} = \Delta n_{Si-Air}(2\pi d/\lambda) \tag{4.4}$$

where $\Delta\varphi_{Dynamic}$ is set to $\pi/2$ and $\Delta n_{Si-Air} = n_{Si} - n_{Air} = 2.418.$ This gives us a step thickness, $d = 31\mu m$. We can assume that the SWGs are simply birefringent crystals, and so the analogous representation is shown in **Figure 4.2b**.

Due to the complexity of such a device, where there are multiple layers and multiple mechanisms of phase accumulations (both dynamic/propagative and geometric), it is non-trivial to derive the theoretical foundation of what occurs for the reflected light. It would be very intuitive to derive the equations for anti-reflection for such a device, as the form-birefringent gratings can essentially be viewed as an anti-reflection coating between the air and bulk silicon. However, anti-reflection is typically best understood when dealing with linearly polarised light incident upon an isotropic AR layer – in this system, we are dealing with CP light incident upon birefringent gratings (and so switch the handedness of CP light passing
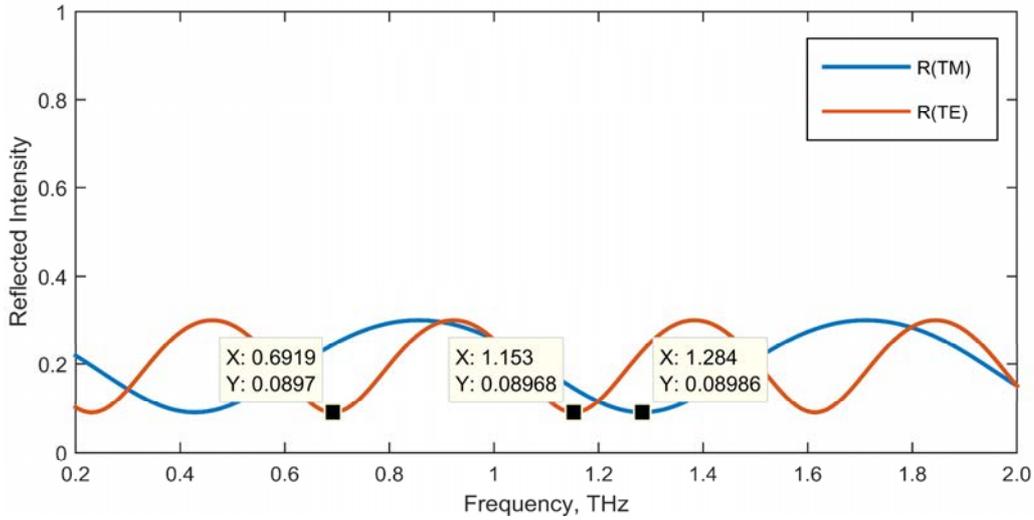
**Figure 4.1: Reflectivity of Form-Birefringent Gratings
without angular disparity and without silicon step**

through) that do not have an easily defined refractive index, which are angled at **45°**

to provide a handedness dependent phase, as well as having them staggered by a

specific thickness of silicon to provide a phase delay between these angled gratings

(the boundary between these steps and the gratings would cause a phase change of

π for any CP wave that is reflected, which is well known in itself to switch CP

handedness [116], but would also then undergo yet another switching of handedness

once passing back through the half-wave plate grating). Therefore, trying to calculate

the anti-reflection response of CP light passing through birefringent, staggered,

geometrically angled gratings, which all contribute phase terms to the light upon

both forward and backward wave propagation, and also including CP handedness

switching due to reflections from the many layers, is indeed beyond the capability of

this investigation. However, we can imagine a system for where the gratings are not

angled at **45°** (so no $\pm\frac{\pi}{2}$ geometric phase) and are not staggered (so no $+\frac{\pi}{2}$ dynamic

phase). Then we are dealing with the relatively simple case where we can investigate

the anti-reflection responses of light which is linearly polarised either parallel ($n_{TE} =$

2.52) or perpendicular ($n_{TM} = 1.36$) to the gratings. Using the Fresnel equations

2.37-2.43 for reflectivity from a three layer system, and for form-birefringent gratings

of depth $h = 129\mu m$ in a frequency range of 0-2THz, we obtain the reflection

responses for both TE and TM polarised light as shown in **Figure 4.1**. We see that

there are three prominent anti-reflection responses around the operation wavelength

of 1THz, occurring at 0.69THz (TE), 1.15THz (TE), and 1.28THz (TM). Because

the refractive indices of the form birefringent grating, namely n_{TE} and n_{TM}, are

dispersionless due to only being dependent upon the filling factor and refractive

indices of the air and silicon substrate (given by equations 2.66a,b), we can ascertain

that the device has broadband capabilities. We know that a circularly polarised wave

is simply the superposition of two linearly polarised waves (with a quarter-wave

delay) and so we can therefore deduce that a CP wave incident onto such a grating

will experience an averaging of the anti-reflection responses, where the average of the

above three listed frequencies are:

$$\frac{0.69THz + 1.15THz + 1.28THz}{3} = 1.04THz \tag{4.5}$$

Although this is an improvised calculation, CP light in conjunction with the

broadband half-wave plate grating will undoubtedly introduce a non-negligible anti-

reflection response which can contribute to the high CCD efficiency of this device.
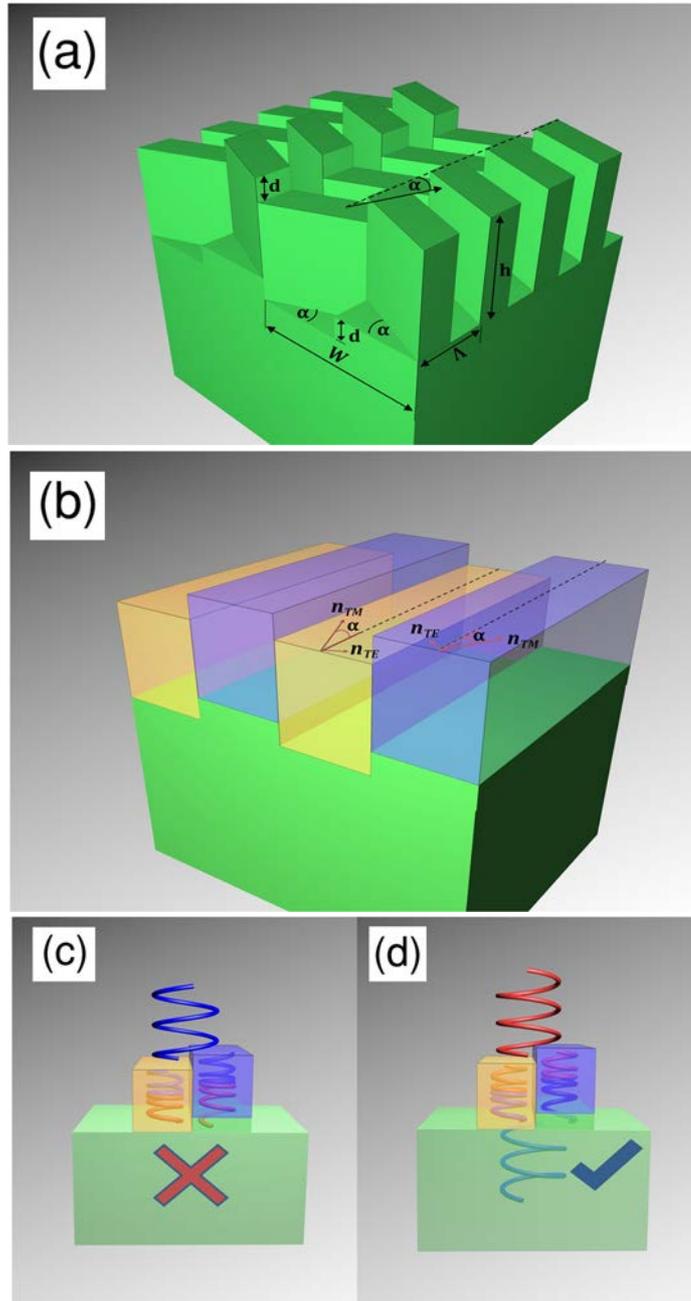
Figure 4.2: Computer generated visualisations of the Herringbone device

(a) Graphical model of the herringbone device, where α is half of the angle between the gratings (α=θ/2=22.5°), d is the step invoking the +π/2 dynamic phase, h is the grating depth, Λ is the grating periodicity, and W is the unit cell width. (b) Model showing the analogous structure for achieving CCD with birefringent crystals, rather than SWGs. (c) The case for destructive interference between the phases, with an incident polarisation of RCP (blue helix). (d) The case for complete transmission due to cancelling out of the phases, with an incident polarisation of LCP (red spiral).

4.2.2   Analytical Modelling using Fresnel's Equations

To further support our theoretical predictions, a simplified analytical model based on Fresnel's equations for transmittance was employed. A linear formulation of circularly polarised light was used in the form of $\hat{e}_R = \frac{1}{\sqrt{2}}(\hat{e}_x + i\hat{e}_y)$ and $\hat{e}_L = \frac{1}{\sqrt{2}}(\hat{e}_x - i\hat{e}_y)$, where $\hat{e}_R$, $\hat{e}_L$ correspond to the Right and Left circular polarisation unit-vectors, respectively, and $\hat{e}_x$, $\hat{e}_y$ correspond to the x and y linear polarisation unit-vectors, respectively. The system was considered to have three-layers, as shown in **Figure 4.3**, with layer 1 being air, layer 2 being an SWG, and layer 3 being bulk silicon. Each layer had corresponding values of the refractive index n, with layer 1 having $n_1 = n_{air} = 1$, layer 3 having $n_3 = n_{Si} = 3.418$, and layer 2 having two refractive indices, due to the anisotropy of the gratings as described by equations 2.55a (TE) and 2.55b (TM) (in chapter 2), with $n_{2x} = n_{TE} = 2.55$ and $n_{2y} = n_{TM} = 1.36$. From this, we used the Fresnel equation for a three-layer system as:

$$t_i^{fres} = \frac{t_{12_i}t_{23_i}e^{-i\phi}}{1+r_{12_i}r_{23_i}e^{-2i\phi}} \tag{4.6}$$
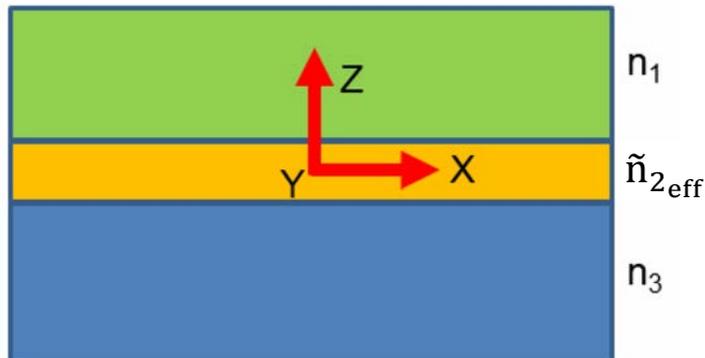
Figure 4.3: Simple schematic of the three-layer system

which is simply the same as equation 2.43b, except now the subscript 'i' corresponds

to either x or y unit vectors, due to the anisotropy of the refractive index n_{2i}. From

this, we have:

$$t_{12_i} = 2n_1/(n_1 + n_{2i}) \tag{4.7a}$$

$$t_{23_i} = 2n_{2i}/(n_{2i} + n_3) \tag{4.7b}$$

$$r_{12_i} = (n_1 - n_{2i})/(n_1 + n_{2i}) \tag{4.7c}$$

$$r_{23_i} = (n_{2i} - n_3)/(n_{2i} + n_3) \tag{4.7d}$$

$$\phi_i = \frac{2\pi d}{\lambda} n_{2i} \tag{4.7e}$$

where d is the thickness of the SWG (129μm). A detailed derivation is given in

Appendix A, where we now show that – absent of any dynamic or geometric phases

– the values of the transmission coefficients for a single SWG in terms of the Fresnel

equation in equation 4.3 are (and omitting the ^{fres} superscript):

$$t_{RR} = \frac{1}{2}(t_x + t_y) \tag{4.8a}$$

$$t_{RL} = \frac{1}{2}(t_x - t_y) \tag{4.8b}$$

$$t_{LL} = \frac{1}{2}(t_x + t_y) \tag{4.8c}$$

$$t_{lR} = \frac{1}{2}(t_x - t_y) \tag{4.8d}$$

Now, if we assume that the second SWG is simply identical to the first, albeit with

multiplicative terms incorporating the phase information, where the dynamic phase

is $e^{i\phi_{Dyn}} = e^{i\Delta n_{Si-Air}(2\pi d/\lambda)} = e^{i(n_3 - n_1)\frac{2\pi d}{\lambda}}$ and   the   geometric   phase   is   simply

$e^{i\phi_{Geom}} = e^{\pm i\frac{\pi}{2}}$, where the $\pm$ corresponds to the incident light being RCP or LCP,

respectively. Now, if we have the second SWG transmission coefficients as:

$$\tilde{t}_{RR} = t_{RR} \times e^{i\phi_{Dyn}} = t_{RR}e^{i(n_3 - n_1)\frac{2\pi d}{\lambda}} \tag{4.9a}$$

$$\tilde{t}_{RL} = t_{RL} \times e^{i\phi_{Dyn}} \times e^{i\phi_{Geom}} = t_{RL}e^{i[(n_3 - n_1)\frac{2\pi d}{\lambda} - \frac{\pi}{2}]} \tag{4.9b}$$

$$\tilde{t}_{LL} = t_{LL} \times e^{i\phi_{Dyn}} = t_{LL}e^{i(n_3 - n_1)\frac{2\pi d}{\lambda}} \tag{4.9c}$$

$$\tilde{t}_{LR} = t_{LR} \times e^{i\phi_{Dyn}} \times e^{i\phi_{Geom}} = t_{LR}e^{i[(n_3 - n_1)\frac{2\pi d}{\lambda} + \frac{\pi}{2}]} \tag{4.9d}$$

where only the terms in RL and LR have the geometric phase factor due to the

constraint of Pancharatnam-Berry phase only providing a geometric phase

contribution for conversion between CP handedness'. Now, we calculate the total

contribution of each SWG by summing together equations 4.4 and 4.5, whilst keeping

the corresponding components together, where we have:

$$t_{RR}^{TOTAL} = \frac{1}{2}(\tilde{t}_{RR} + t_{RR}) \tag{4.10a}$$

$$t_{RL}^{TOTAL} = \frac{1}{2}(\tilde{t}_{RL} + t_{RL}) \tag{4.10b}$$

$$t_{LL}^{TOTAL} = \frac{1}{2}(\tilde{t}_{LL} + t_{LL}) \tag{4.10c}$$

$$t_{LR}^{TOTAL} = \frac{1}{2}(\tilde{t}_{LR} + t_{LR}) \tag{4.10d}$$

Finally, we convert these Jones matrix components into transmittances (energy

fractions) by utilising equation 2.41 from Chapter 2, giving us the general form:

$$T = \frac{n_3}{n_1}\left|t_{ij}^{TOT}\right|^2 \tag{4.11}$$

Because $n_1 = 1$ for air, and substituting the expressions from equations 4.9 and 4.10 into equation 4.11, we obtain the generalised analytical equation providing us with the transmission coefficients, or Transmittances, for any combination of incident and transmitted CP light as:

$$T_{ij} = n_3 \left| \frac{1}{2} t_{ij} \left[1 + e^{i(\phi_{Dyn} + \phi_{Geom})} \right] \right|^2 \tag{4.12}$$

Using this equation, a frequency dependent response was calculated using Matlab, with the transmitted intensities shown below in **Figure 4.4a**. An obvious difference can be seen between the four curves, with the most pronounced occurring between the cross-polarisation curves. As expected, when light is incident with left circular polarisation almost all of the light is converted into right circularly polarised transmitted (T_{RL}, red solid curve) whilst for right incidence the converted output for left transmission is negligible. A maximum transmittance of over 80% occurs at 1.1THz for T_{RL} whilst a minimum of ~0% at 1.1THz occurs for T_{LR}, which is very close to the designed frequency of 1.0THz.

The slight mismatch can be attributed to the fact that because the herringbone metasurface layer is approximated as an SWG from the effective medium theory, the intuitive 'single-pass' theoretical foundation does not take interfacial aspects of the complete structure into consideration (and so impedance matching between layers of differing refractive index should be considered); therefore, Fabry-

Pérot resonance effects resulting from the reflectance terms in the denominator of

equation 4.3 lead to the analytical transmittances differing from the simple 'phase-

only' predictions of equation 4.1, where this reasoning can also be applied to explain

why the maximum value of T_{RL} is not 100%. Additionally, the basis of the structure

is using form-birefringence to design half-wave plate gratings; the effective medium

approach used is simplified and only first order. Works on higher order effective

medium theory have been produced [117], which would improve the functionality

and accuracy of the device. However, it is beyond the capability of this project to
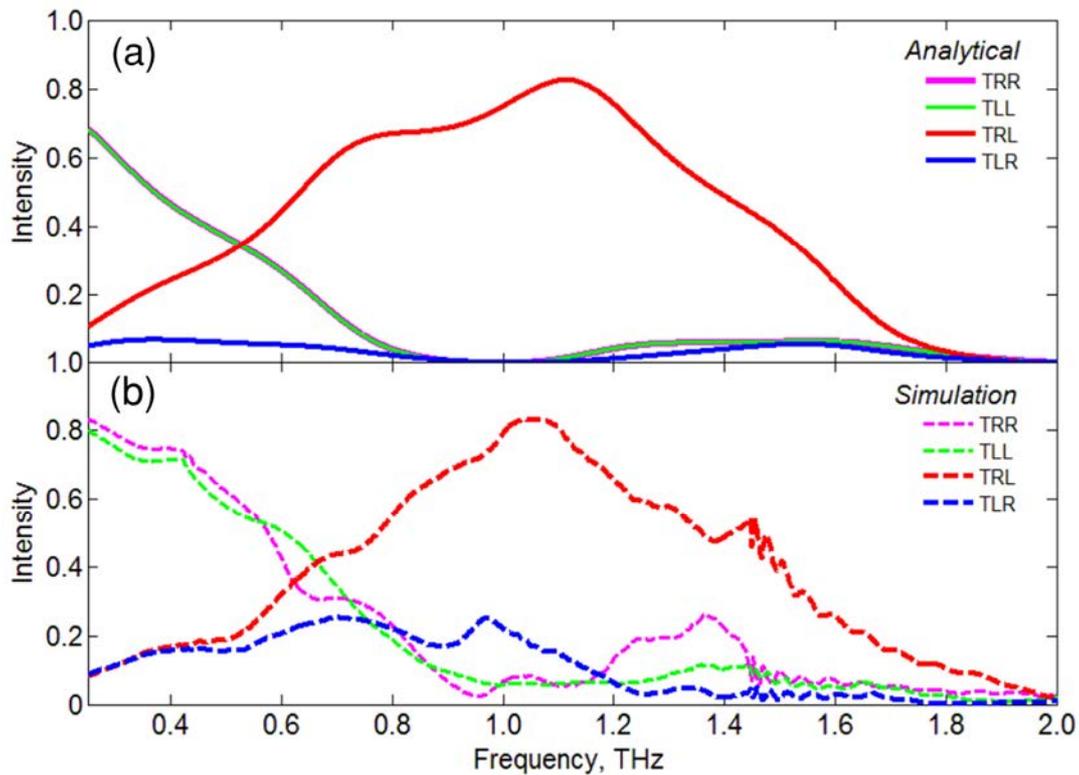
Figure 4.4: Analytical and Simulated frequency dependent
transmittances of the silicon herringbone structure

(a) Analytically modelled transmittance curves using a simplified Fresnel
based treatment of a 3-layer system. **(b)** Numerical simulation modelled
transmittance curves using CST Microwave Studio.

incorporate such a complex analytical representation of the effective medium theory, and the fact that the first order effective medium theory is widely accepted and acceptable for use in developing form-birefringent gratings. Irrespective of this oversight, the curves displayed in a show a good intuitive representation of the theoretical soundness, where the terms in T_{RR} and T_{LL} are also negligible at the resonance frequency of interest, as expected.

An important note is that the total energy of the system must be conserved, and so we must examine the separate intensity pathways from Left or Right incident CP light, separately. We have that:

$$t_R = t_{LR} + t_{RR} \tag{4.13a}$$

$$t_L = t_{RL} + t_{LL} \tag{4.13b}$$

where the left hand side subscript of equations 4.13a,b correspond to the incident CP handedness. Using the above equations to evaluate Figure 4.4a, we have that $t_R \approx 0.05$ and $t_L \approx 0.85$. For true energy conservation, and assuming that all of the light is transmitted (with no reflections or absorption), we should ideally have that $t_R = t_L = 1.0$. This is clearly not the case, and because the analytical calculations use lossless refractive index parameters, we must assume that 0.95 (95%) of the light from an RCP wave (t_R) is reflected, which corresponds to equation 4.3 yielding a π total phase, namely destructive interference, whilst only 0.15 (15%) of the light for LCP incidence is reflected, which corresponds to equation 4.3 yielding a total phase

of 0, namely an anti-reflection response. It would be beneficial to analytically derive

the reflection response to compare to the curves for transmission shown in

Figure 4.4a. However, for practicality, reflection measurements are difficult due to

not being able to both excite and detect the reflected light at normal incidence – the

incident and reflected light will have to be at some oblique angle. This will degrade

the performance of such a device and will not operate as expected, as the boundary

conditions defining the form birefringence are now dealing with non-normal TE and

TM modes for circularly polarised light, and that the propagation distance through

the active structures will not correspond to the calculated and required phases. For

this reason, reflection will not be investigated in this project, either analytically,

through simulation, or experimentally. Additionally, the majority of works carried

out in this field are only interested in the transmission response, and not the

reflection, (asymmetric *transmission* rather than asymmetric reflection) most

probably due to the aforementioned experimental issues.

4.3   Computer Simulations

In order to reinforce the theoretical reasoning and functionality of our device,

3D Finite-Difference Time Domain (FDTD) simulations were carried out using the

commercially available *CST Microwave Studio* software package. The structure was

modelled using periodic boundary conditions, and a linearly polarised plane-wave

incident from the substrate side. The results of the circular transmission components

are shown in **Figure 4.4b.** All linear transmittance results exceeded unity in the simulation; this arose from the herringbone metasurfaces having a higher transmissivity compared to bulk silicon (which was automatically taken as reference in the model from the plane wave being incident at the vacuum-silicon boundary). Because of this, all of the simulated results were multiplied by a normalisation factor (calculated from the Fresnel equations for 2-layer transmittance) of $C = \sqrt{4n_1n_2/(n_1+n_2)^2} = 0.84$, where $n_1 = 1$ is the refractive index of air and $n_2 = 3.418$ is the refractive index of silicon; this reduced all transmittances to below unity as necessary. These linear results were then converted into circular transmittances using the equations given in 2.34, Chapter 2.

As can be seen, the results show a very good correspondence to those for the analytical model, especially apparent for T_{RL} exceeding a transmittance of 80%, and also show a clear difference between the cross-polarisation components of T_{RL} and T_{LR}. To date, this is the highest efficiency achievable for a handedness-sensitive circular polarisation converter for transmission which simultaneously exhibits CCD effects in a broadband frequency range. One discrepancy that is worth noting is that the device can no longer be considered as subwavelength for frequencies much larger than the operational frequency of ~1THz, and would result in diffraction occurring causing spurious interference effects. Similarly, for frequencies much lower than the operational frequency, corresponding to wavelengths larger than 300μm, it can no

longer be assumed that the light is confined and localised within an individual SWG

and hence may not provide the necessary dynamic phase as portrayed in

Figure 4.2c,d.

As mentioned in the previous section, we must consider the total energy

contributed to the system. From the simulated curves, we have (from equations

4.13a,b) at the highest transmission of T_{RL} (which occurs at ~1.05THz) that

$t_R \approx 0.30$, implying 70% of the light is reflected at this frequency when the incident

light is RCP, whilst $t_L \approx 0.90$, implying that only 10% of the light is reflected when

the incident light is LCP. It is worth noting that the broadband nature of this device

is due to the dispersionless operation of the refractive indices of the SWG's, in

conjunction with the dispersionless nature of the Geometric phase imparted from the

45° angular disparity between the gratings which always provides the same phase

additions regardless of the frequency. The phase provided by the silicon step is

periodic with wavelength, and so can be treated as quasi-broadband. The limiting

factors in the design which degrade the broadband operation are that the gratings

are a finite depth and width, which means that for frequencies much different to

1THz the half waveplate functionality is no longer apparent, whilst the

subwavelength approximation also breaks down for large frequencies. Even so, the

device is very broadband, giving a T_{RL} full-width at half-maximum (FWHM) of

approximately 0.87THz with a central frequency of 1.05THz.

4.4   Fabrication and Experimental Results

To experimentally verify our theoretical reasoning, the structure shown in Figure 4.2a was fabricated by conventional photolithography and plasma etching using a two-step pattern process as outlined in Chapter 3 of this thesis. Firstly, Intrinsic silicon wafers (525μm thickness, 100mm diameter) were spin-coated with ~8μm thickness of SPR-220 7.0 and then exposed to UV-light (Karl Suss MJB-3 Mask Aligner) with a photomask stripe pattern of period $W = 208$μm (shown in Figure 4.2a). After developing in a TMAH (Tetramethylammonium Hydroxide) based solution the sample was then Deep Reactive Ion Etched (DRIE) to a depth of ~31μm, using an alternating etch/passivate (CF_4/C_4F_8) Bosch-process, in an STS Multiplex ICP DRIE Etcher. Next, the sample was then cleaned to remove the SPR220-7.0 and a ~50μm layer of SU8-2050 was spun-coat on top of the etched silicon stripe pattern. Transparency of the SU8 allowed visible alignment of the second 'herringbone' mask-pattern to be exposed and overlaid on the stripes below. The sample was developed in a PGMEA (Propylene glycol methyl ether acetate) based solution and then etched to a depth of ~129μm; the SU8 is removed from the sample by use of Piranha solution (H_20_2:H_2SO_4) due to its very high durability and resistance to chemicals. The complete fabricated device is shown in **Figure 4.5**, imaged using a Scanning Electron Microscope (SEM). It is apparent that grassing on the lower step edges, to which the incomplete removal of the SU8 resist is accredited.
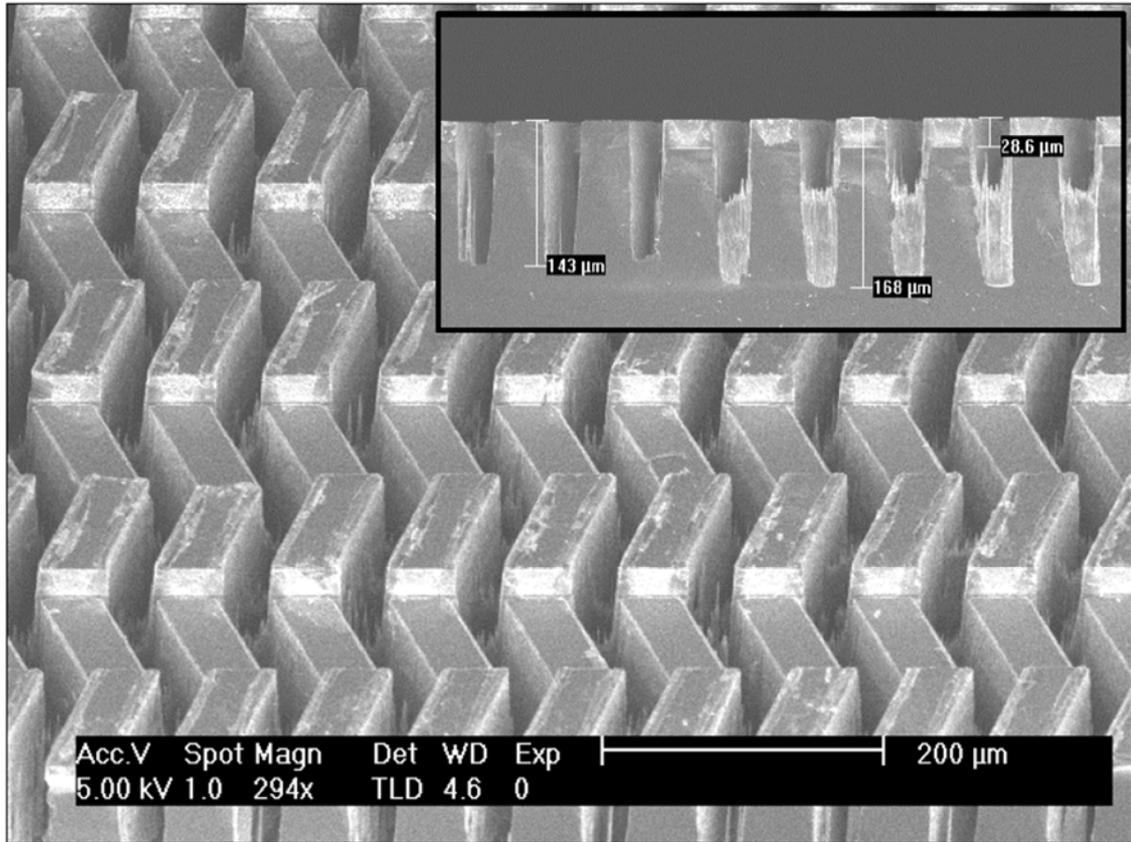
***Figure 4.5: SEM Image of etched Silicon Herringbone
metasurface***

The intrinsic silicon wafer was fabricated according to the processes governed
in Chapter 2, involving a 2-step fabrication process to combine two structures
together as given in Figure 4.2. Significant grassing and roughness occurred,
to which the SU8 photoresist is accredited the cause, arising from incomplete
development whereby small remnants of this SU8 resist will act as micromasks
causing long, thin structures to the be etched beneath.

If small remnants of SU8 remain prior to plasma etching, these will act as small

micromasks causing long, thin, grass-like features to be formed. However, due to the

high-durability of SU8, it is very difficult to fully remove all traces of it. For future

samples, it is advisable to perform oxygen ashing (oxygen plasma etching) of the

sample prior to silicon etching to fully remove all organic compounds present on the

surface.

To characterise and obtain the transmission data for our device, a fiber-based Terahertz Time-Domain Spectroscopy (THz-TDS) system is used to measure linear Jones matrix components ($T_{xx}, T_{yy}, T_{xy}, T_{yx}$) of the herringbone structure at normal incidence for a frequency range of 0.2 - 2.0 THz. The results for the experimental data characterisation are shown in **Figure 4.6**, where the reference of the data was taken as a blank piece of intrinsic silicon. It is clear from Figure 4.6 that there is indeed an asymmetry between the T_{RL} (red) and T_{LR} (blue) components as expected. At the operational frequency of 1THz there is an extinction ratio of nearly 4:1 between T_{RL} and T_{LR}, with the maximum value of ~60% transmittance occurring at f_{max}=0.9THz for T_{RL}. Furthermore, T_{RL} shows a broadband operation, spanning a frequency range of Δf = 0.47THz (centred at f_{max}=0.9THz) which exceeds a transmittance of 40%, and a FWHM of 0.75THz (again, centred at 0.9THz).
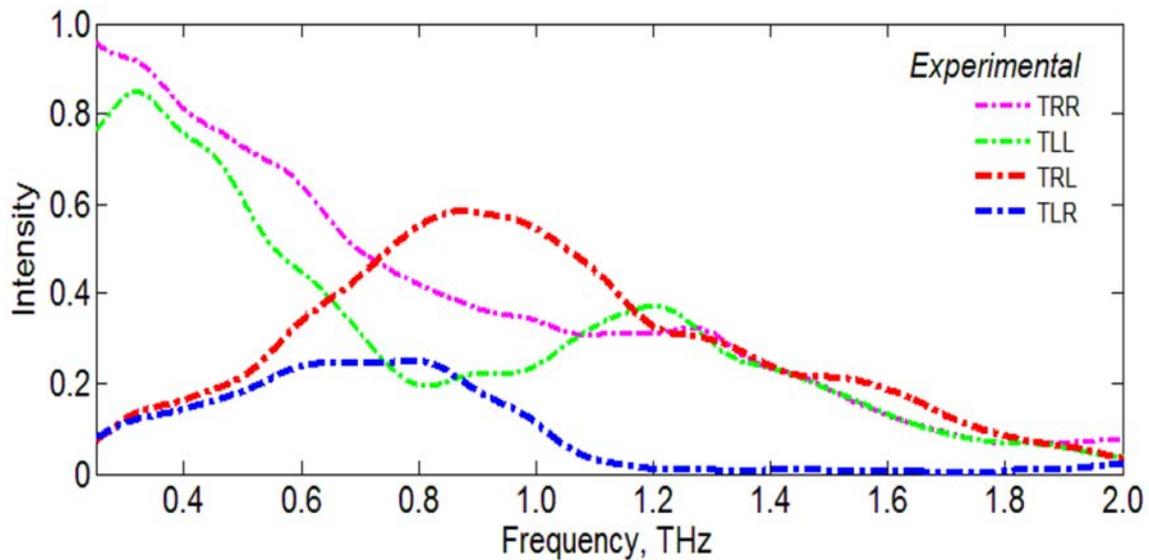

Figure 4.6: The Experimentally obtained transmittance data for the fabricated silicon herringbone device

Experimentally obtained transmittance curves using THz-TDS system.

In addition, T_{RL} has a consistently higher response than that of T_{LR} which typically only shows a transmittance of ~25% or less over the frequency spectrum of interest. Although the device shows strong chiral response, namely CCD effect, it does not perfectly compare to the theoretical and simulated results. This is in main part due to fabrication errors; alignment between the stripe and herringbone mask patterns will undoubtedly result in the degree of structure-chirality to lessen, affecting the geometric phase contribution. Also, a bottleneck could occur for both UV resist exposure of the SU8 at the bottom of the 30μm trenches, and also DRIE due to the lack of radicals able to escape from the very deep trenches. This is evident from the 'grassing' effect seen in Fig. 4.5 inset, whereby etch products re-settle where they act as 'micro masks' and protect the silicon below from being etched, although this grassing can be primarily attributed to the difficulty in removing the SU8 resist situated at the bottom of the trenches, which could similarly act as 'micro masks' in the DRIE process. Regardless of these difficulties, the device still exhibits the same responses as the theoretical results, showing that this method of achieving a chiral response using dielectric materials is robust and not overly sensitive to fabrication errors. Due to the limited availability of the fabrication facilities used by myself, it was not possible to fabricate a perfect sample corresponding to Figure 4.2a – it is hoped that this can be carried out in the near future and to compare results for a more well-matched device.

Examining the total energy of the system, we have from equations 4.13a,b at

0.9THz that $t_R \approx 0.60$ (with $t_{LR} \approx 0.20$ and $t_{RR} \approx 0.40$), meaning 40% of the light

is reflected for an incident handedness of RCP, and $t_L \approx 0.83$ (with $t_{RL} \approx 0.60$ and

$t_{LL} \approx 0.23$), meaning that only 17% of the LCP light is reflected. It is clear that

$t_L > t_R$ but not as large as the difference experienced for both the theoretical and

simulated results. However, it is the cross polarisation difference between t_{RL} and t_{LR}

which is much larger (t_{RL} is three times larger than t_{LR}) and of more importance in

the operation of this device.

4.5   Conclusion

In conclusion, we have demonstrated and fabricated a functional monolithic dielectric device to achieve a strong asymmetry between the orthogonal circular polarisations of transmitted light, which provides a very high broadband capability and transmittance of one cross-polarisation whilst prohibiting the opposite one. The broadband capability stems from the dispersionless nature of both the Geometric phase and the effective medium dependent refractive indices, perpendicular and parallel to the gratings. The limiting factors to the broadband capability are due to the dynamic phases due to the specific depth of the gratings and silicon step, which are designed to work at 1THz, and the periodicity of the gratings no longer being sub-wavelength for frequencies much higher than 1THz. Impressively, the herringbone metasurface not only provides a simulated cross-polarisation transmittance of 0.8 for T_{RL}, and 0.6 for the fabricated device, but also has a greater transmittance than pure silicon alone or any other similar works carried out using planar or multi-layered metasurfaces to achieve a chiral response, namely the Circular Conversion Dichroism. The high efficiency of the light transmitted for LCP is attributed to an anti-reflection type effect, whilst the inhibition of RCP light is believed to be due to destructive interference and behaviour akin to a dielectric mirror as deduced by the small amount of transmitted light – the 'missing' light is considered to be reflected, as material losses are believed to be negligibly small.

However, such reflected light is not directly observed due to the complexity in the experimental setup which would be required, and the fact that the light would need to be incident at oblique angles which would affect the desired functionality of the device. Improvements to the design of the device can be realised by using the higher-order effective medium theory (compared to the first order used here) and by a systematic sweeping of the structure parameters used in the simulation to optimise the CCD effect. This can pass onto the fabricated device, where sample fidelity and quality could be improved by using better photolithography equipment, for alignment issues, and a tailoring of the DRIE process to make sure the sidewalls of the gratings are parallel and to prevent the bottleneck which may introduce the grassing seen at the base of the trenches.

Due to the lack of metallic structures, losses are negligible and the application of subwavelength gratings in conjunction with a geometric phase provides a robust and facile means of achieving such Circular Conversion Dichroism functionality, which may provide a preferable route for cheaper and more efficient applications for optical computing, image processing, or biological characterisation, where the demand for high-efficiency circular polarisation disparities is crucial.

Chapter 5

Broadband Metasurface with Simultaneous Control of Phase and Amplitude

Electromagnetic waves have a range of controllable degrees of freedom, such as polarisation, phase, amplitude, and frequency. These degrees of freedom have been utilised to achieve a multitude of applications, from simple lenses to 3D-holography. However, it can be argued that the ultimate aspects for control are those of amplitude and phase (whilst operating for a range of frequencies) as polarisation essentially yields a change in either amplitude or phase. Here, this work aims to *simultaneously* achieve control of both phase and amplitude, whilst operating under a broadband frequency range in the Terahertz (THz) regime. The working principle is dependent upon two types of antenna – *Split Ring resonators* (SRR's) can control the phase of an incident linearly polarised (LP) wave by adjustment of its geometry, whilst *rod antennas* can control amplitude by a simple rotation of the angle between its long

axis and the incident LP wave. The combination of these two types of antenna yield SRR's that utilise geometry to control phase, and the rotation angle between the symmetry line of the SRR and the incident LP wave to control the amplitude. These two degrees of freedom can be controlled in a smooth fashion, and the working principle is applied to create diffraction gratings capable of displaying, one, two, or three diffraction order configurations, although any arbitrary number of orders can be chosen. Even though the SRR's are designed to operate at a specific design frequency, the results clearly display broadband activity owing to the interplay between the C-antenna symmetric and anti-symmetric modes - this allows a robustness and tolerance to fabrication errors.

~~~

This chapter includes passages from the publication "L. Liu, X. Zhang, **M. Kenney** et al, *Broadband Metasurfaces with Simultaneous Control of Phase and Amplitude,* Adv. Mater. **26**, 5031-5036 (2014)" [118], which was a collaborative effort, involving myself as a main contributing author. *My primary contribution to this project was fabrication of the metasurface samples, with assistance from L. Liu and N. Xu. Simulations and Theory were performed by L. Liu, experimental design and measurements were carried out by X. Zhang and X. Su, and Prof. S. Zhang helped to devise and oversee all aspects of the project.*

~~~

5.1   Motivation

Metamaterials, an artificial arrangement of periodic subwavelength optical elements, have continuously attracted a great deal of attention due to their unusual and controllable properties. These metamaterials can be utilised for controlling the propagation of light, with notable examples including negative and zero refraction [24–30], sub-diffraction imaging [34–36], and invisibility cloaking [31–33]. Regardless of the fact that metamaterials have been successful in paving the way for an array of novel potential applications and fundamental physics, it is still a big challenge devising metamaterials for real-world applications ─ this is primarily due to their bulky stature, materials losses, and issues with fabrication. Seeing as metamaterials are bulk arrangements, it is difficult to overcome these problems. Most metamaterials are composed of metals, which have significant ohmic losses attributed at optical frequencies. The three-dimensional arrangement also requires precise alignment between layers, which is challenging even with the most state-of-the-art fabrication facilities.

During the past decade or so, metasurfaces, which are the two-dimensional counterpart of metamaterials and consisting of a monolayer of resonant structures, are capable of controlling the wavefront of light; they can therefore be used to serve as an alternative approach to overcome the issues associated with volumetric metamaterials, due to the much more straightforward fabrication procedures and

smaller losses, and can lead the way to bridging the gap between fundamental research and useful, practical applications [40,41,65,67–69,83,87–89,119–137]. The complex and time-consuming alignment and fabrication procedures necessary for bulk metamaterials is not required, whilst the optical losses associated with these are negligible for metasurfaces as they consist of structures typically only a fraction of the thickness of the wavelength of light being investigated. Despite being at its infancy, metasurfaces have shown great promise for novel applications, as shown by numerous devices, such as high resolution three-dimensional metasurface holograms [65,126–129], high efficiency [69,87–89] and switchable surface plasmon couplers [67,125], ultrathin flat lens [41,123,124], and various other functional interfaces [83,130–138].

Of the majority of metasurfaces realised thus far, most only seem able to control the phase profile of the transmitted light – to completely manipulate the propagation of light, however, requires both phase *and* amplitude to be controlled *simultaneously.* This is especially crucial in applications such as holography, laser beam shaping, and generation of complex wave fields, where the manipulation of both of these degrees of freedom is required to produce high quality holographic or far-field images. Due to this, previous attempts have already been made to achieve simultaneity of phase and amplitude control with metasurfaces which use the antenna geometry to do so [128–130]. In [130], a detour phase scheme was used to realise complete control of

the phase and amplitude, where the pixel size of the diffractive surface is much

greater than the wavelength of the light in question; this, however, is undesirable for

device applications gearing towards small form-factors. More applicable work was

carried out in [128], where a single layer metasurface, using subwavelength pixels,

was used to engineer both phase and amplitude. However, the design of this work is

complicated due to the fact that each pixel requires a different geometry to achieve

arbitrary phase and amplitude, which limited the selection of phase and amplitude

to only a few discrete levels. This is due to the fact that each combination of phase

and amplitude corresponds to a different geometric antenna design, which in turn

reduces the applicability of this design scheme as a suitable device as both fabrication

and computation of arbitrary pixels is too complex to be readily utilised.

Additionally, because it is necessary to manufacture each pixel individually,

corresponding to a set value amplitude and phase, it would be difficult to make such

a device have broadband frequency operation.

Here, a robust and facile approach is employed for achieving simultaneous

phase and amplitude manipulation in a single layer metasurface over a broadband

frequency range in the terahertz regime. Compared to previous approaches for similar

works, this metasurface design uniquely combines a number of important merits –

subwavelength pixel size for continuous wavefront manipulation, easy fabrication,

robust and broadband control of both the phase and amplitude, and finally that the

amplitude, as well as the phase, can be engineered precisely and continuously with very little modification to the pixel geometry.

5.2   Background Theory

The design of the broadband metasurface relies on the combination of two types of metasurfaces, with different functionalities: a metasurface for phase control, determined by the geometrical configuration of each antenna; and a metasurface for amplitude control, which is realised by manipulation of the angular orientation. The first type of metasurface (for phase control) consists of an array of C-shaped Split Ring Resonators, referred to as C-antennas, where each of these have a carefully designed geometry. This metasurface operates for linear incident polarisation and under a cross-polarisation scheme, where the phase is robustly controlled dependent upon the antenna geometry – namely the radius, arm width and the open angle – via a transmitted beam which is orthogonal to the incident beam. A schematic representation of this first type of metasurface is shown in **Figure 5.1a**. These C-shaped antennas work similarly to the widely adopted V-shaped antennas [40,119] but are smaller for the equivalent resonance frequency, and so can be more subwavelength. *The line of symmetry of each antenna lies along either +45° or -45° in order to maximise the conversion between incident and transmitted orthogonal polarisations in the horizontal and vertical directions.*
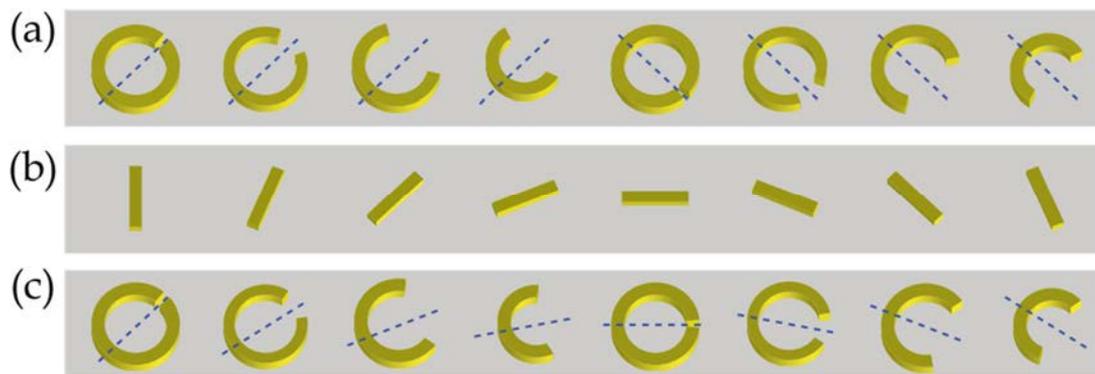
Figure 5.1: Design concept of metasurfaces with simultaneous phase and amplitude control [118]

(a) C-shaped antenna array with differing geometrical parameters for the phase control of linear polarisation conversion. **(b)** Rod antenna array with differing orientations for amplitude control of linear polarisation conversion. This type of metasurface has a duality between phase and amplitude, where phase is controlled for an incident circularly polarised beam. **(c)** C-shaped antenna array, using the same geometrical parameters as those in '(a)' but combined with the differing angular orientations used in '(b)' to give control of both phase and amplitude for linear polarisation conversion.

The metasurface of the second type, for amplitude control, consists of an array of rod-antennas, all with the same dimensions, but varying orientation angles with respect to the horizontal axis. Interestingly, there exists a duality between the phase and amplitude control in this type of metasurface — namely, this metasurface operates as a *phase* plate for a circularly polarised incident beam, or as an *amplitude* plate for incident linearly polarised light (which we utilise in this work). This metasurface is shown in **Figure 5.1b**. Under linearly polarised illumination, the scattering amplitude of the cross-polarisation is controlled smoothly by the orientation angle of each rod-antenna [40]. This same configuration — utilising the orientation angle of the rod-antenna array — can also be employed for controlling

the phase of the scattering wave for incident circularly polarised light, as demonstrated in the work carried out by Huang et al [68].

Combining the design concepts of these two types of metasurfaces provides a metasurface consisting of C-shaped antennas, which can simultaneously control phase and amplitude depending upon their orientation angle and geometry, respectively, as shown in **Figure 5.1c**. This allows us to construct an almost arbitrary complex transmission, or reflection, coefficient distribution at the interface simply by arranging the previously well studied C- and/or V-shaped antennas in varying orientation angles, without having to specially design a new antenna for each combination of phase and amplitude. Thus, this approach provides a facile and robust way to obtain a metasurface which allows simultaneous phase control and a continuously tuneable amplitude profile, greatly facilitating the step towards complete control of light propagation.

To implement this approach for simultaneous control of phase and amplitude, we design a metasurface consisting of C-shaped antennas, as shown in **Figure 5.2a**. When a linearly polarised plane wave is incident onto a C-shaped antenna the symmetric $\tilde{E}_s$ and anti-symmetric $\tilde{E}_{as}$ modes contribute to an orthogonally polarised output wave, whose phase and amplitude can be engineered by adjusting the geometric parameters of the antenna [40,83] (see **Figure 5.2** for details of the parameters).
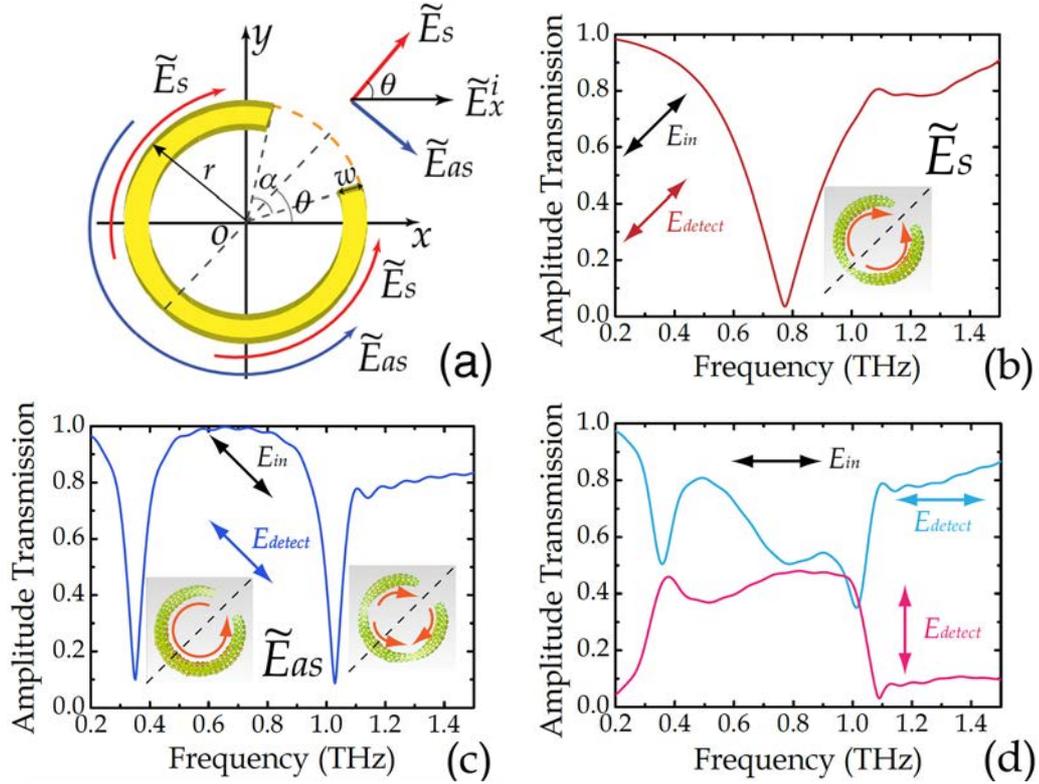
Figure 5.2: Angular dependence of phase and geometrical dependence of amplitude of a C-Shaped antenna [83,118]

(a) A schematic of the C-shaped antenna, with opening angle α, radius r, width ω, and orientation angle θ with respect to the x-axis. The red and blue curved arrows around the C-antenna represent the symmetric mode $\tilde{E}_s$ or the anti-symmetric mode $\tilde{E}_{as}$ of the antenna, respectively. The vector arrows, inset, represent the polarisation direction required to excite the **(b)** symmetric $\tilde{E}_s$ (red) or **(c)** anti-symmetric $\tilde{E}_{as}$ (blue) modes of the antenna, and the **(d)** black arrow $\tilde{E}_x^i$ is the actual incident polarisation for the desired cross-polarisation effect, which corresponds to a vector sum of the symmetric and anti-symmetric modes $\tilde{E}_{as}$ giving rise to the cross polarisation transmission. The broadband nature of a C-antenna is due to the excitation of all three resonances (both $\tilde{E}_s$ and $\tilde{E}_{as}$ modes) which combine to yield a broad response as shown in (d) which spans from 0.4-1.0THz. Additionally, it is seen that the amplitude is more or less constant for this broadband response.

When an x-polarised wave $\tilde{E}_x^i$ is incident onto a C-shaped antenna with its symmetry

axis oriented along an arbitrary direction forming an angle θ with the x-axis (as

shown in the inset of Figure 5.2a, above), the resultant y-polarised scattered field $\tilde{E}_y^s$

can be written as [83]:

$$\tilde{E}_y^s = \frac{1}{2}\tilde{E}_x^i \sin(2\theta)\left(A_s e^{i\phi_s} + A_{as} e^{i\phi_{as}}\right) = \tilde{E}_x^i A e^{i\phi} \tag{5.1}$$

where A_s, A_{as} and ϕ_s, ϕ_{as} denote the scattered amplitude and phase from the

symmetric and anti-symmetric modes (respectively) when the symmetry axis of the

structure is along $\theta = 45°$, and A and ϕ denote the overall scattered amplitude and

phase, respectively. It is seen in **Figure 5.2b** that the symmetric mode $\tilde{E}_s$ has a

resonance at 0.8THz when the polarisation is along the symmetry axis of the C-

antenna. Similarly, when the polarisation is orthogonal to the symmetry axis, shown

in **Figure 5.2c**, two anti-symmetric modes $\tilde{E}_{as}$ are excited with resonances occurring

at 0.35THz and 1.02THz for the dipole and multipole responses, respectively.

According to Equation 5.1, for a fixed antenna design, the amplitude of $\tilde{E}_y^s$ is solely

determined by the orientation angle θ, and when the symmetry axis of the antenna

is along $\theta = 45°$ the amplitude of $\tilde{E}_y^s$ reaches the maximum value. The accumulation

of all of these responses for $\tilde{E}_s$ and $\tilde{E}_{as}$ –at 0.35THz, 0.8THz and 1.02THz – yields a

highly broadband response when the incident polarisation $\tilde{E}_x^i$ is along the x-axis, and

can be shown in **Figure 5.2d**. This is owing to the fact that a polarisation incident

in the x-axis causes an excitation of all three modes simultaneously, leading to a

broad overlap between them. The cross-polarised output wave $\tilde{E}_y^s$ can be seen to

operate at a relatively constant amplitude between 0.4-1.0THz, *which is the physical*

interpretation behind the broadband capability of this device, as realised later in the

chapter. Conversely, when altering the orientation angle of an antenna between 0°

and +90° the phase of the scattered wave is not affected at all, whereas shifting the

angle to go below the *x*-axis, where the angle θ is between -90° and 0°, simply yields

an additional phase shift of π.
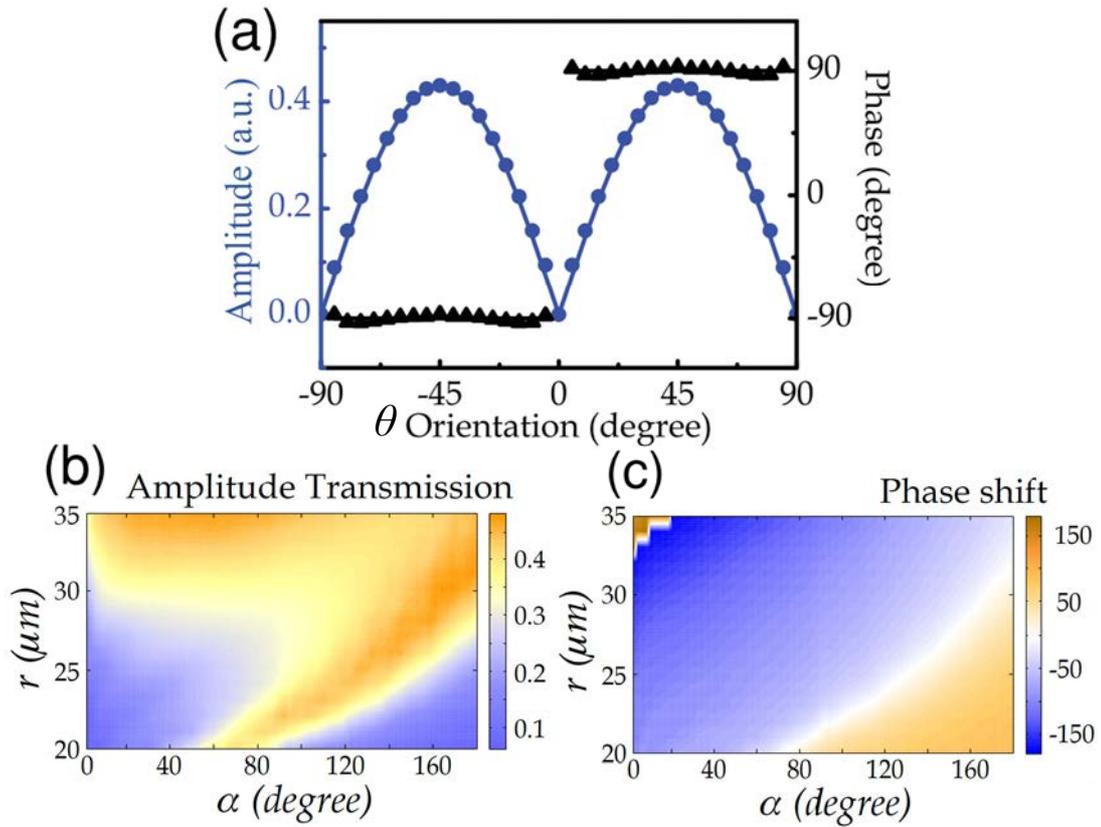
Figure 5.3: Simulated transmission amplitude and phase response of a fixed-geometry C-shaped antenna array [118] **and varying-geometry C-antennas** [83]

(a) The transmitted cross-polarised amplitude (blue solid circles) and phase (black solid triangles) profiles of an array (unit cell 80μm in both *x*- and *y*-directions) of C-shaped antennas with fixed geometry $(r, \alpha) = (34\mu m, 11°)$ and θ varying from -90° to +90° at 0.63THz. The amplitude profile follows a $|\sin(2\theta)|$ dependence (blue solid line), whereas the phase profile remains constant apart from a phase jump of π when the angle increases past $\theta = 0°$. **(b)** Simulated variation of the cross-polarised transmission amplitude and **(c)** cross-polarised transmission phase, when the C-antenna radius, *r*, and opening-angle, α, are varied from 20-35μm and 0-180°, respectively.

The simulated amplitude and phase variations of a C-shaped antenna at different orientation angles from -90° to 90° at a frequency of 0.63THz are shown in **Figure 5.3a**, where the amplitude varies as $|\sin(2\theta)|$ whilst the phase remains constant (due to the fixing of the geometric parameters to r = 34µm and α = 11°) in two separate angular ranges of -90° to 0° and 0° to +90° — only an abrupt change of π at $\theta = 0$° occurs. This abrupt phase change of π can be viewed as us flipping the C-antenna about the *x*-axis, which then causes the $\widetilde{E}_s$ and $\widetilde{E}_{as}$ excitation polarisations to switch places (from inset of Fig.2a). In essence, $y \rightarrow$ -y, and so the output cross-polarised signal in the *y*-axis will be a half-wavelength out of phase compared to the original case. Hence, the orientation angle θ serves as an important parameter used to control the amplitude of a scattered wave without the need to resort to designing new antenna geometry. **Figure 5.3b** shows the cross-polarised transmission amplitude when the C-antenna radius, *r*, is varied from 20µm to 35µm and the opening-angle, α, is varied from 0° to 180°, respectively, whilst **Figure 5.3c** shows the cross-polarised transmission phase when the C-antenna has *r* and α varied by the same values as for the amplitude. *The variation of the amplitude and phase responses due to the variation of the C-antenna geometric parameters are explained by the near-field interactions between the two arm 'ends' (either side of the gap)* [133] *– by altering the distance between the arm 'ends', and similarly the length/circumference of each C-antenna, these interactions will in turn affect the scattering and absorption*

cross sections (namely, extinction) of each C-antenna, which results in a change in the amplitude and phase responses at resonance.

This approach for realising metasurfaces with simultaneous control of phase and amplitude is applied to the design of terahertz metasurface gratings which can arbitrarily control the diffraction orders for a linearly polarised incident beam. In general, to generate a grating having desired diffraction order amplitudes A_m, the transmission through the grating is expressed as,

$$t(x) = A(x) \exp[i\phi(x)] = \sum_m A_m \exp(-i2m\pi x/d) \tag{5.2}$$

where d is the grating periodicity and m is an integer denoting the diffraction order. Achieving a single diffraction order ($m = -1$) requires that the transmission function is simply a linear phase gradient along the x direction whilst maintaining a uniform amplitude. This is referred to as *anomalous refraction* and is the outcome of the result for the Generalised Snell's Law as derived in Chapter 2, which has been realised previously using metasurface phase gradients [40,119]. However, because the amplitude is a function of x, the amplitude and phase need to be carefully controlled when wishing to generate multiple diffraction orders as the transmission also varies with position.

5.3   Sample Design

Three metasurface designs were proposed to generate different numbers of diffraction

orders, and are shown in **Figure 5.4** below. For the sake of this project, three

diffraction order combinations were chosen: *m=-1*; *m=-1,-3*; and *m=-1,-2,-3*. For the

first design, where we choose just one diffraction order, we have (from equation 5.2):

$$t_1(x) = A_1 \exp(i2\pi x/d) \tag{5.3}$$

We can choose that $A_1 = 1$, and indeed for all three designed diffraction gratings, to

normalise all of the amplitudes such that they are equal. The grating periodicity **d**

is simply the number of unit cells in one grating period multiplied by the size of one

unit cell. For all of the figures shown in Figure 5.4, we have a unit cell size of **80μm**.

In **Figure 5.4a**, we have a single diffraction order and we choose 8 pixels for one

grating period. This therefore give us a value of $d = 8 \times 80μm = 640μm$. We then

substitute both A_1 and **d** into equation 5.3 to give us:

$$t_1(x) = \exp(i2\pi x/640μm) \tag{5.4}$$

This equation was then used to plot both the amplitude (*abs[t(x)]*) and phase

(*arg[t(x)]*) using the commercial software *Matlab*, where the solid lines in Figure 5.4a

(and indeed all of Figure 5.4) are the output plots from Matlab. We can see for two

diffraction orders in **Figure 5.4b** that the analytical equation plots (solid lines) are

a little more complex. Due to this we chose to use 16 unit cells, rather than 8, so

that we could have a greater granular control over the phase increments along the

grating. This in turn gives a value of $d = 16 \times 80\mu m = 1280\mu m$. For the sake of showcasing the flexibility of these gratings, we chose to use the *m=-1* and *m=-3* orders, rather than *m=-1* and *m=-2*. As was previously stated, all of the amplitude coefficients were normalised to unity and so we have that $A_1 = A_3 = 1$. This then gives us an overall transmission response of:

$$t_2(x) = \exp(i2\pi x/1280\mu m) + \exp(i6\pi x/1280\mu m) \qquad (5.5)$$

and similarly the amplitude and phase were plotted using Matlab, as shown in Figure 5.4b (solid lines). Lastly, the grating for generating three diffraction orders, m=-1, m=-2, m=-3, was again generated using a value for the periodicity $d = 16 \times 80\mu m = 1280\mu m$, where we chose to have 16 unit cells as the plots shown in **Figure 5.4c** are relatively complex and require a greater deal of control over the phase than is available when only using 8 unit cells. The equation used for plotting three diffraction orders is given as:

$$t_3(x) = \exp(i2\pi x/1280\mu m) + \exp(i4\pi x/1280\mu m) + \exp(i6\pi x/1280\mu m) \quad (5.6)$$

where we again normalise the amplitudes to unity, with $A_1 = A_2 = A_3 = 1$. All of the diffracting metasurfaces contain four different antenna geometries [83], whose phase evenly span from 0 to π with a phase-step of $\pi/4$, and with a flip of each antenna along the *x*-axis to introduce an additional phase shift of π (as explained above). These four C-antennas were carefully chosen from selecting the required phase values from the phase map shown in Figure 5.3c. The arm width of all four
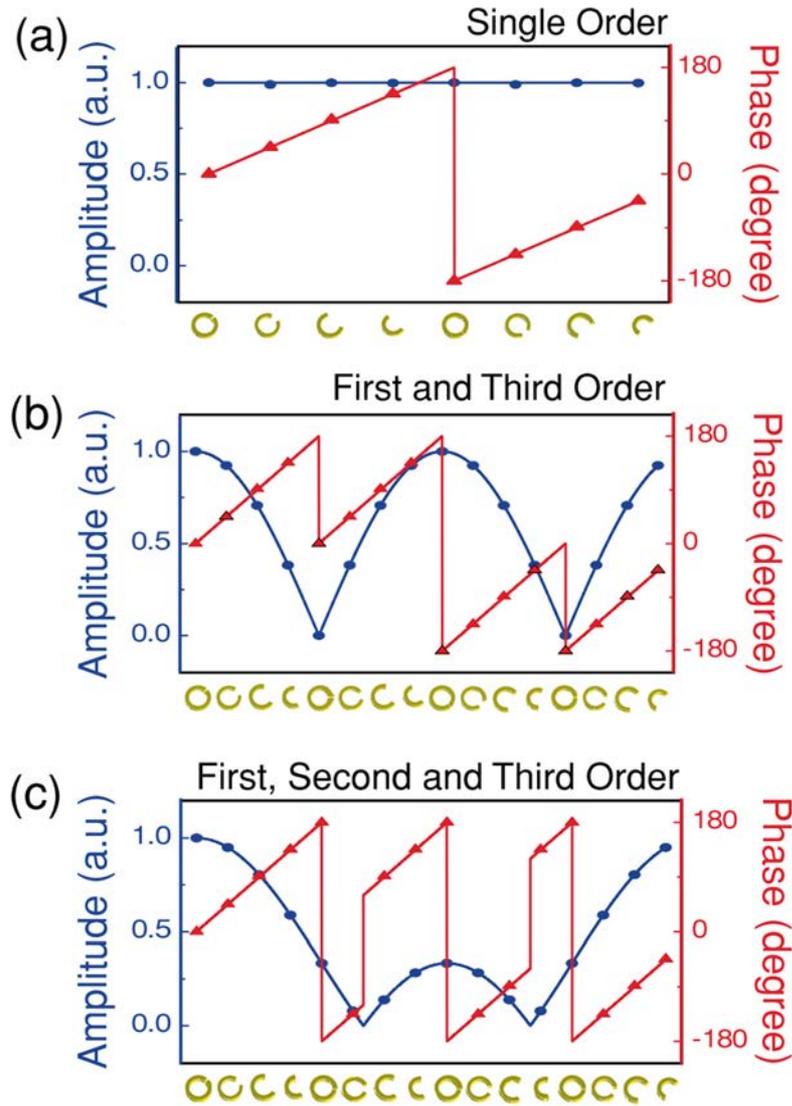
Figure 5.4: Amplitude and phase profiles of the three diffracting metasurfaces [118]

For each metasurface, four different geometries of C-shaped antenna are designed with the same w=5μm but different (r,α) = (34μm, 11°), (32.3μm, 47°), (34.4μm, 117°), and (29.8μm, 140°). Simulated amplitude (normalised) and phase profiles the x-axis for sample of **(a)** a single diffraction order, **(b)** two diffraction orders, **(c)** three diffraction orders; all at 0.63THz for y-polarised transmission from an x-polarised incidence. The symbols (circles/triangles) represent simulated results, whilst the solid lines are obtained from Equation 5.2-5.6 for the designed profiles. In (a), the orientations of each individual antenna are either +45° or -45°, whereas in (b) and (c) the orientations vary *between* these two angles due to the varying amplitude requirement. The antenna spacing (pixel size) was fixed as 80μm in the simulation. Inset: super-cell arrangements of antennae for each design.

C-antennas was fixed as $w = 5\mu m$, whilst the radius and open-angle were varied and were chosen as (r,α) = (34μm, 11°), (32.3μm, 47°), (34.4μm, 117°), and (29.8μm, 140°). As shown in Figure 5.4a, the super-cell of the metasurface (defined as the total length of the antennas which span a full phase from 0 to 2π) for generating a single diffraction order (m=−1st order) consists of eight antennas (seen underneath each figure of curves, Figure 5.4), whose symmetry axis are either along +45° or −45° to realise a linear phase gradient between 0 and 2π whilst maintaining a constant amplitude profile. The super-cells of the metasurfaces for generating two (m=−1st and m=−3rd) and three (m=−1st, −2nd and −3rd) diffraction orders consist of sixteen antennas, with the orientation of the symmetry axis of each antennas continuously varying between +45° or −45° to achieve the desired phase and amplitude profiles (Figure 5.4b, c). In all of these plots, the phase (red triangles) and amplitude (blue circles) profiles arising from the simulated responses of the individual C-antennas (with the above listed geometries) are overlaid on the analytically calculated Matlab curves. There appears to be a very good correspondence between both simulated and analytically calculated phase and amplitude responses, and so we can be confident that the diffraction grating arrangements shown beneath all of the figures in Figure 5.4 are well representative of the physically occurring diffraction orders from such devices.

5.4   Fabrication and Experimental Results

In order to experimentally verify our theoretical analysis, four samples (A-D) were fabricated. This involved the use of photolithography along with metal deposition and lift-off process methods, as explained in Section 3.1 of Chapter 3. Sample A, with an antenna spacing of 80μm (of which the super-cell size is 640μm × 80μm), is designed to provide only a single diffraction order. Sample B and C are both designed for achieving two diffraction orders but using slightly different antenna spacings (100μm for sample B and 80μm for sample C), and sample D is designed for achieving three diffraction orders (using an antenna spacing of 100μm, super-cell size = 1600μm x 100μm). The antennas were fabricated by patterning the C-shaped antennas as voids into photoresist on an intrinsic silicon substrate, evaporating 200nm of aluminium, and then performing lift-process to leave behind metallic
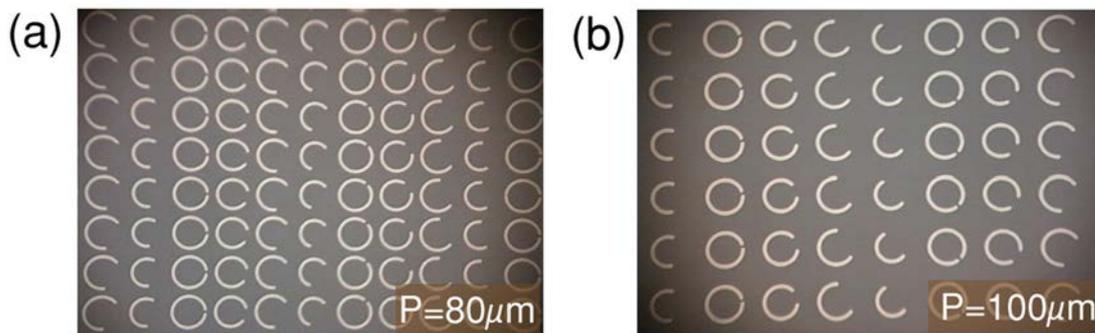

Figure 5.5: Optical Images of diffractive metasurfaces C and D [118]

(a,b) Top view of the fabricated aluminium metasurfaces C and D, respectively, where C is used for achieving two diffraction orders and D is used for achieving three diffraction orders. P is the pixel size, i.e. the spacing between adjacent antennas.

C-shaped antennas on the silicon substrate. Optical images of sample C and D are shown above in **Figure 5.5 (a)** and **(b)**, respectively.

To characterise the optical response of the metasurfaces, namely the diffraction orders in the range of 0.4 and 1.0THz, a fiber-based angular resolved terahertz time-domain spectroscopy (THz-TDS) system is used. The measurements took place at normal incidence, and the transmitted beam was angularly resolved. (More information on the operation of a THz-TDS system is given in Section 3.2 of Chapter 3). The Terahertz transmitter and receiver were a pair of commercially available fiber-based terahertz photoconductive switches (Menlon System). The wave emitted from the transmitter was first focussed onto the sample by a dielectric lens, before being refocussed by an identical lens onto the receiver. Samples were placed in the centre of a rotational stage with the super-cell (phase gradient) direction of the metasurface aligned with the horizontal axis (normal to the beam) with the receiver attached to a swivelling metal arm attached to the centre of the rotational stage. This enabled the receiver to gather the signals successfully from the sample, regardless of the angle of the transmitted beam. The diffracted beam from the samples were measured by rotating the stage at a fixed incremental angle of 3° spanning from $-90°$ to $+90°$. In order to eliminate any non-orthogonal transmission (seeing as the operation of the metasurface is cross-polarisation) two polarisers were

employed – one before the sample, and one after – with horizontal alignment for the

first (incidence) and vertical alignment for the second (transmission).

The results obtained experimentally from the THz-TDS setup are shown in

Figure 5.6a-d, where the diffraction spectra are shown in a broad angular range

from −70° to +70° over a frequency range of 0.4 to 1.0THz. It is clearly visible in

Sample A (Figure 5.6a) that a single diffraction order exists, which agrees with

previous work using this same metasurface carried out by Zhang et al [83], as
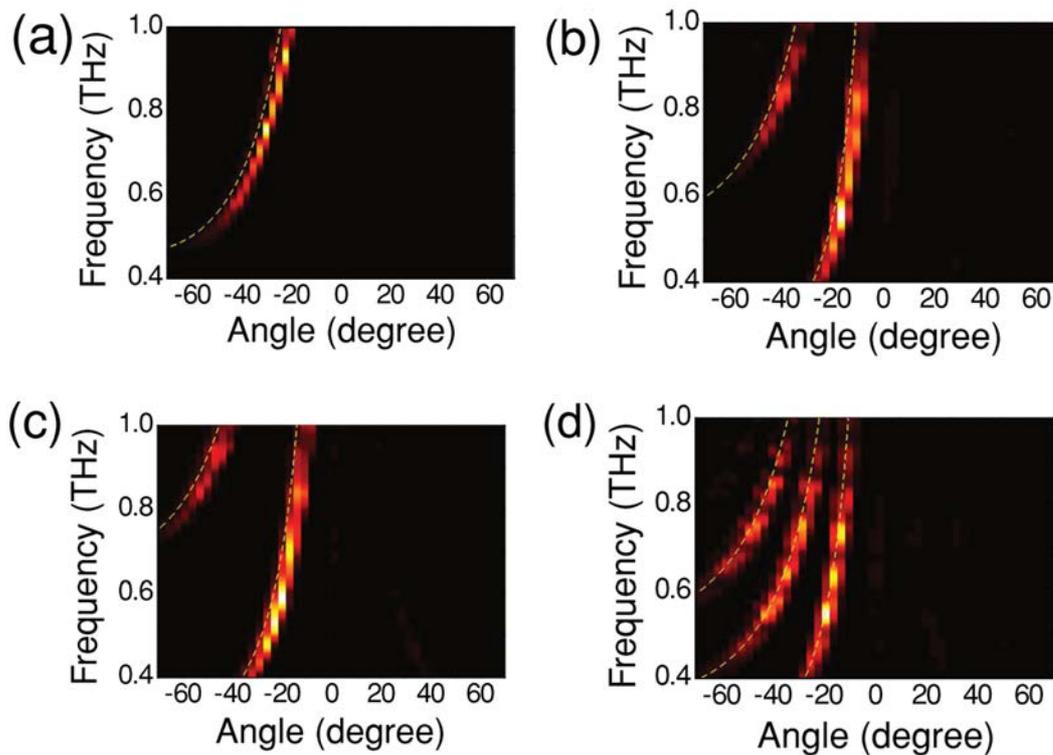
Figure 5.6: Experimental characterisation of the metasurfaces at normal incidence. [118]

(a-d) Measured diffraction signals of samples A-D, respectively.
(a) Sample A, amplitude and phase profile from Figure 4.4a, P=80μm.
(b) Sample B, amplitude and phase profile from Figure 4.4b, P=100μm.
(c) Sample C, amplitude and phase profile from Figure 4.4c, P=80μm.
(d) Sample D, amplitude and phase profile from Figure 4.4d, P=100μm. The dashed yellow lines in the spectra represent the calculated diffraction angles.

required. Even though the metasurfaces were designed to operate at a frequency of

0.63THz, it is evident from the spectra that a broadband operation exists from 0.5

to 1.0THz, due to the excitation of both symmetric and anti-symmetric modes as

explained previously. As predicted by the design configuration, samples B and C

both exhibit two diffraction orders, but due to their differing pixel sizes they result

in different diffraction angles, as seen by the mismatch between Figure 5.6 (b) and

(c) where the curves are shifted more to lower angles in (c). Again as expected, the

diffraction spectra of sample D show three distinct diffraction orders, as designed.

For all of the diffractive metasurfaces, the measured diffraction angles agree well

with the calculated values over for the whole frequency range of 0.4 to 1.0THz – the

observed multiple diffraction orders in sample B-D confirm that the metasurfaces

performed as required and exhibit well controlled phase and amplitude profiles. As a

note, for samples B-D, which have more than one diffraction order, it is found that

the intensity of the lowest diffraction order is always the strongest, even though the

metasurfaces were designed to exhibit equal electric field amplitudes for all of the

orders. An explanation for this is that the projection area of the antenna array along

the propagation direction of a diffraction order differs for each; for larger diffraction

angles, corresponding to higher diffraction orders, the area projected is smaller and

in turn carries less energy. Additionally, because the antennae are modelled using

normal incidence, it is less understood what occurs for the resonance behaviour at

oblique angles of incidence, and due to reciprocity it can be inferred that this is also

true for oblique angles of transmission – if the electric field polarisation is not planar to the antenna, it would lead to a reduction of the quality of resonance compared to that for normal incidence and excitation from in-plane polarisation.

5.5   Numerical Simulations

To further reinforce the results from both the experimental and theoretical calculations, and to gain a more clear insight, 3D Finite-Difference Time-Domain (FDTD) simulations were performed using the commercially available software *CST Microwave Studio*. The parameters of the four fabricated samples A-D were modelled individually, where a Terahertz beam is incident from the substrate side onto the
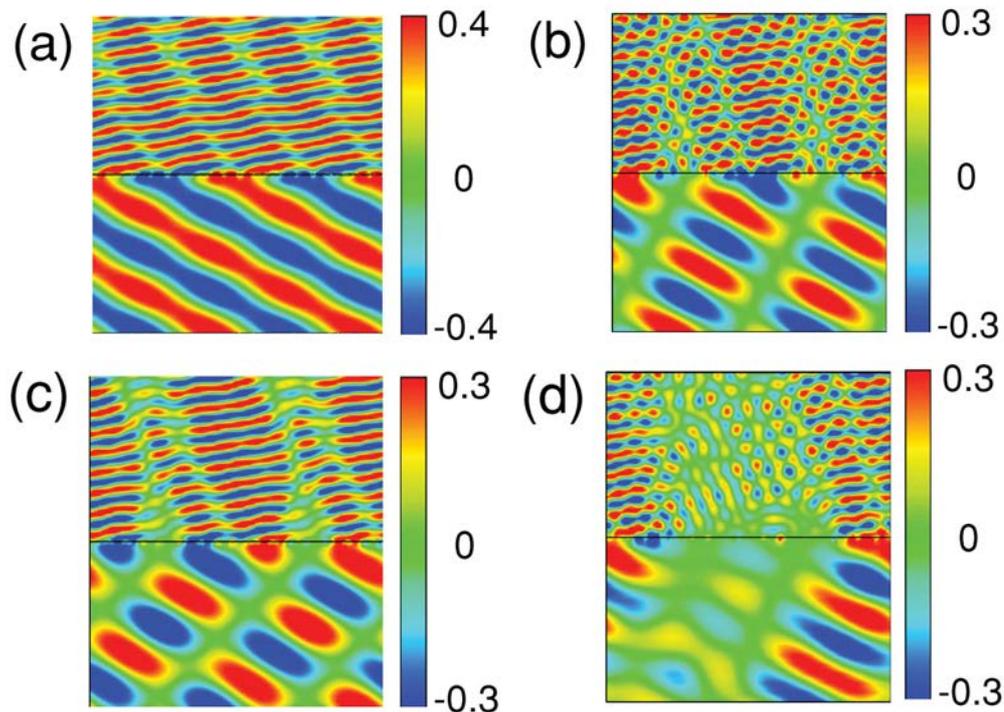
Figure 5.7: Simulated electric field distributions at 0.8THz [118]

The simulated y-polarised electric field distributions of **(a-d)** for Samples A-D, respectively, at 0.8THz. The terahertz beam is incident from the silicon substrate side (top-half) at normal incidence.

metasurfaces. The simulated electric field distributions at 0.8THz, for designated

diffraction orders, are shown in **Figure 5.7**, where the displayed data is for cross-

polarisation with respect to the incident wave.

As seen in **Figure 5.7a**, the simulation results shown are for fabricated

Sample A, and a uniform transmitted anomalous wave is observed which agrees well

with the experimental and calculated designs for a metasurface exhibiting a single

diffraction order. **Figure 5.7b**, **c**, and **d** show, respectively, the simulation results
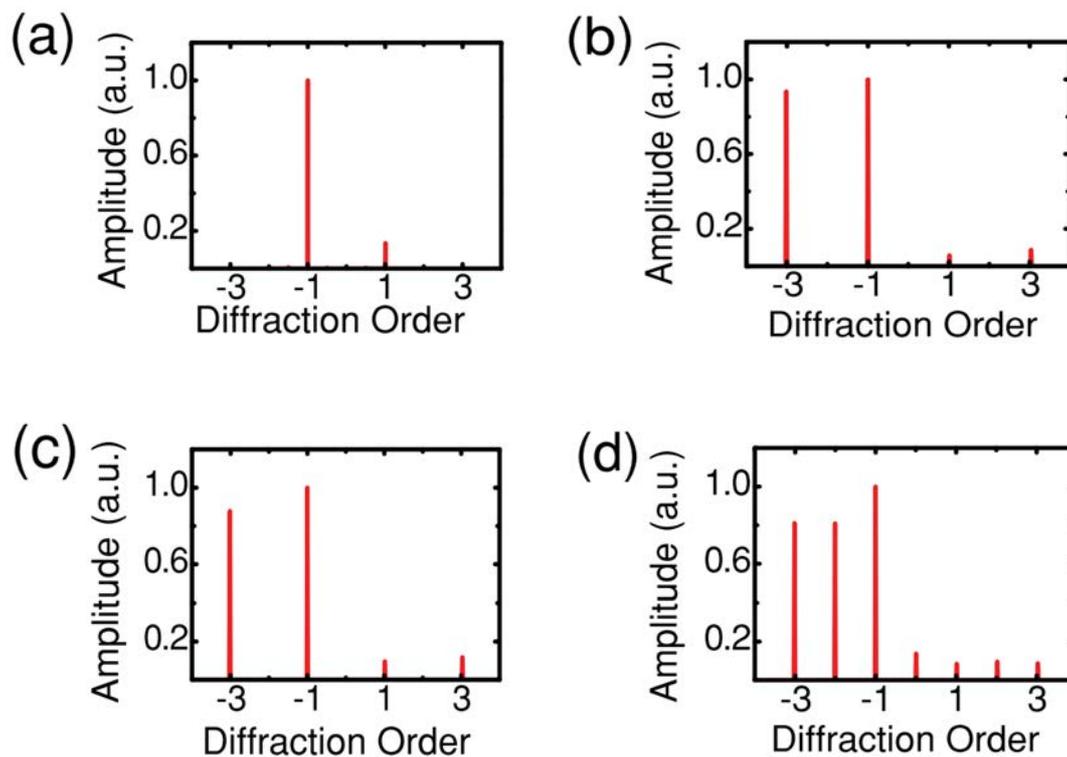
Figure 5.8: Numerically retrieved diffraction orders from
simulated electric field profiles. [118]

(a-d) The diffraction order distributions calculated by Fourier Transform of
the *y*-polarised electric field distribution from the FDTD simulations. The
data was acquired from the *x*-axis at a distance of 900μm from the
metasurfaces in the air side. The antenna spacing is 80μm in **(a)**, **(c)** and
100μm in **(b)**, **(d)**, so as to be consistent with the experimental measurement.

for samples B and C – metasurfaces exhibiting two diffraction orders – and sample D, exhibiting three diffraction orders. However, it is obvious that these plots no longer display an electric field profile which is comprehendible, due to the near-field interference of the multiple diffraction orders. In order to ascertain whether these metasurfaces do indeed operate correctly and exhibit multiple and well-defined diffraction orders, we must see the far-field distribution, which requires a Fourier Transform of the near-field distribution. A Fourier Transform is applied to the data obtained from the simulations at a distance of 900μm from the metasurfaces in the air-side, and these results are shown in **Figure 5.8a-d**. It can be seen that a single dominant diffraction peak occurs at $m = -1$ (**Figure 5.8a**), two dominant diffraction orders at $m = -1$ and -3 (**Figure 5.8b,c**) and three dominant diffraction orders at $m = -1$, -2, and -3 (**Figure 5.8d**) are observed for the three metasurfaces, respectively (where Sample B and C, corresponding to Fig's 5.8b and c, have two diffraction orders, but at different diffraction angles due to pixel size differences). Therefore, it is clear that the simulations do indeed provide a good agreement with the experimental measurements and calculated results.

5.6   Conclusion

To summarise, a design scheme for realising metasurfaces capable of controlling both the phase and amplitude profile has been proposed. The design of the metasurface is simple, robust and broadband, whilst benefitting from the freedom of simultaneously engineering the geometry and orientation angle of the C-shaped antennas comprising the metasurface. As a proof of concept, it is shown that the design can be applied to realise a metasurface grating capable of arbitrarily controlling the intensity of multiple diffraction orders. This approach can be utilised to engineer complex holograms, for example, with simultaneous phase and amplitude control, which paves the way towards high quality computer generated holography and the generation of arbitrarily complex optical patterns.

Chapter 6

High Efficiency Broadband Reflective Metasurface for Holography Applications

Surfaces covered by ultrathin plasmonic structures – so-called metasurfaces [40,88,119,121] – have recently been shown to have the capability for complete control of the phase of light, representing a new standard for the design and construction of innovative optical devices such as ultrathin flat lenses [41,123,124], directional couplers for surface plasmon polaritons [67,121,122,125], and wave plate vortex beam generation [40,139]. Of all of these types of metasurfaces, geometric metasurfaces – which consist of an array of metallic rod antennas with spatially varying orientations – have shown superior phase control due to the geometric nature of their phase profile [68,140]. Metasurfaces have recently been utilised for computer-generated holograms [65,127–129,141,142], but with too low a holographic efficiency at visible wavelengths for practical purposes. Here, we report the design and

realisation of a geometric metasurface hologram reaching diffraction efficiencies of 80% at 825nm and a broad bandwidth between 630nm and 1,050nm. A 16-level-phase computer-generated hologram is demonstrated, and combines the advantages of a geometric metasurface for controlling the phase profile with reflectarrays for achieving high polarisation conversion efficiency. The design of the hologram presented here is achieved from the integration of a metal ground-plane with the geometric metasurface to exhibit an enhancement between the conversion efficiency of two orthogonal circular polarisation states – this leads to the reported high diffraction efficiency without the necessary complications required in other fabrication processes for holography production. Due to the advantages of this work, various practical holographic applications can be realised.

~~~

This chapter includes passages from the publication "G. Zheng, H. Mühlenbernd, **M. Kenney** et al, "*Metasurface Holograms reaching 80% Efficiency*", Nat. Nanotechnol. **10**, 308-312 (2015)" [69], which was a collaborative effort, involving myself as a main contributing author. *My primary contributions to this project were design, modelling, and optimisation of the multilayer reflectarray structure, with assistance from G. Zheng. Fabrication was performed by H. Mühlenbernd, and measurements were carried out by both G. Zheng and G. Li. Data analysis was performed by G. Zheng, G. Li, T. Zentgraf, and S. Zhang.*

~~~

6.1   Motivation

In traditional phase-only computer-generated hologram (CGH) designs, the phase profile is controlled by etching different depths into a transparent dielectric substrate, for *dynamic* phase control. Due to the ease of fabrication, two-level binary CGHs have been widely adopted. These CGHs have a theoretical diffraction efficiency of only 40.5% and an unavoidable issue of *twin-image* generation occurs. One way to bypass this issue is to use multi-level-phase CGHs, which overcome the low efficiency and twin-image problems, but this requires expensive and complicated means of fabrication such as greyscale, variable-dose or multi-step lithography [143]. Consequently, such techniques for fabrication will incur etching errors, alignment errors, and resolution errors, which will undoubtedly negatively affect the performance, and lead to a lessening of the signal-to-noise ratio, strong zero-order intensity, and poor uniformity. To alleviate these issues, an effective medium approach has been proposed in [143], where two-level-depth subwavelength structures with varying cell compositions behave as effective media. In this work, each unit-cell consists of a tall thin pillar of dielectric media on top of a thick reflecting metal surface, and can be controlled (via lithography) to have differing lateral dimensions – the whole unit-cell can then be modelled as an effective medium, with thinner pillars (which are effectively devoid of dielectric) tending towards 0-phase and thicker pillars (which introduces a larger overall effective index) to have

larger phase values. Because of the geometry-dependent (namely, effective index-dependent) phase, the system can be used as a multi-level CGH. However, due to the small, high-aspect ratio features, the observed efficiency of this CGH when using three phase levels is limited to only 29% – much less than the theoretical value of 48.5%.

Geometric metasurfaces (GEMS) provide an alternative approach to achieving high-efficiency holograms without the need for complicated fabrication procedures. These GEMS rely on inversion of the helicity of polarisation (either transmission or reflection) with respect to that of the incident circularly polarised one [38,144] – this is equivalent to flipping the circular polarisation in transmission or maintaining the same circular polarisation in reflection. In contrast, a typical polished metal mirror will reverse the handedness of (normal incidence) circularly polarised light upon reflection. A geometric phase, or Pancharatnam-Berry phase, as detailed in Chapter 2, is acquired through the inversion of the handedness, which lead to an antenna-orientation controlled phase; this phase does not depend on any specific antenna design or resonance frequency, which makes this phase acquisition technique highly robust against fabrication issues, material property variations, and also results in a broadband performance.

6.2   Theory and Sample Design

Thus far, the performance of such GEMS devices has been severely limited in converting one circular polarisation into the opposite, especially so in the visible and near-infrared wavelengths, where the efficiency of conversion for such planar metasurfaces is typically less than 20% [65]. To increase the efficiency of GEMS, a multilayer design is utilised, which works to achieve a high polarisation conversion [87,145–147]. Our reflective metasurface element consists of three layers: a metal
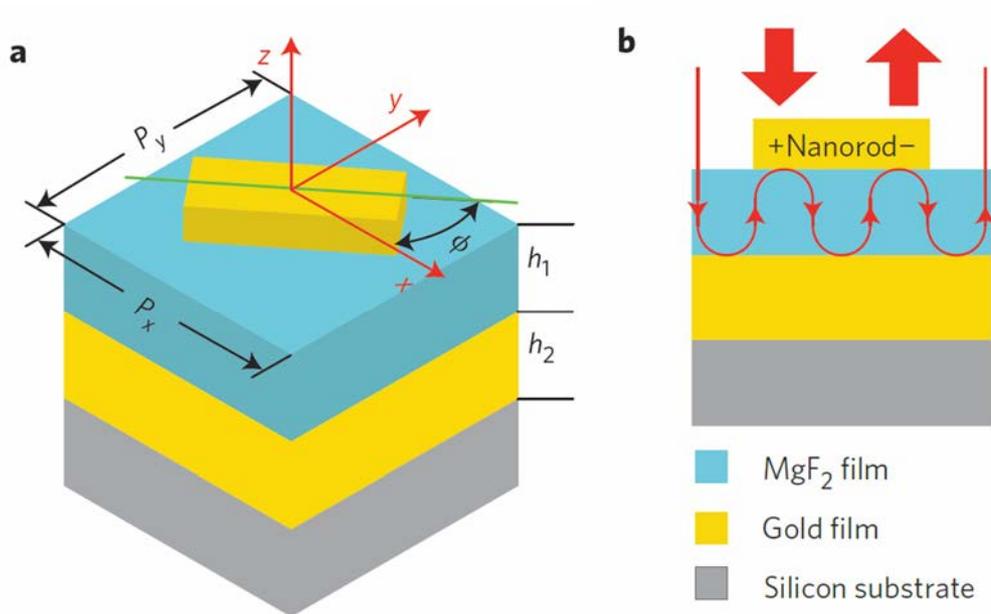
Figure 6.1: Schematic diagram of the unit-cell element used for high polarisation conversion [69]

(a) A single unit-cell element of the high efficiency polarisation conversion metasurface. The nanorod can rotate in the *x-y* plane with an orientation angle of Φ to create differing phase delays. The pixel are arranged with periods P_x = 300nm and P_y = 300nm. The nanorods have length L = 200nm, width W = 80nm, and height H = 30nm. The MgF_2 and gold films have thickness of h_1 = 90nm and h_2 = 130nm, respectively. **(b)** Cross-section of the unit-cell. The gold films acts as a mirror to reflect the incident light, whilst the MgF_2 layer acts as a Fabry-Pérot cavity to continue to re-excite the nanorod and generate a phase delayed output beam.

ground plane, a dielectric spacer later, and on top of this the layer of antennas

forming the metasurface. A schematic diagram of this is shown in **Figure 6.1**; this

element operates in such a way that the transmitted light, which has interacted with

the nanorod antenna, is reflected back from the metal ground plane and re-excites

the antenna in a Fabry-Pérot type mechanism with the dielectric spacer-layer acting

as the cavity (**Figure 6.1b**). In terms of optical devices, it is well known that a half

wave plate can fully convert a circularly polarised beam into the opposite handedness

by imparting a phase delay of π between the two plane waves composing the beam,

arising from the fast and slow axis of the wave plate. This type of device is termed

birefringent, as detailed in Chapter 2, which is where the refractive index (or relative

permittivity) is different along the orthogonal crystal axes. In a similar light, to

achieve a high conversion between the orthogonal circular polarisation states, it is

desirable that the phase difference between the *reflected* light, with polarisation along

the long axis (r_l) and the short axis (r_s) of the nanorod antenna, is also equal to π.

Simulated results of incident linear polarisation upon a metasurface using the high

efficiency design, and a fixed rotation angle Φ, are shown in **Figure 6.2**. The

proposed three-layer element shows excellent reflection coefficient responses

exceeding amplitudes of 0.8 for both linear polarisations. More importantly, the phase

plots shows that the difference between the phase responses along the long and short

axes is to a good approximation equivalent to π, over a broad bandwidth between
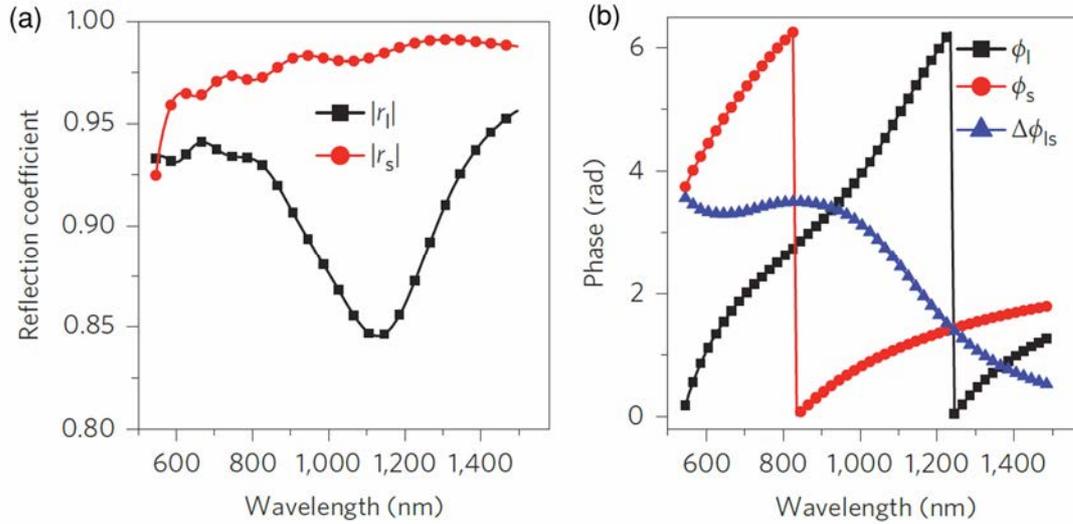
***Figure 6.2: Simulated Amplitude and Phase responses of a
metasurface array with fixed rotation angle*** [69]

(a) Simulated amplitude $|r_l|$, $|r_l|$ and **(b)** phase Φ_l, Φ_s (with phase difference
$\Delta\Phi_{ls}$) of the reflection coefficients r_l and r_s, where l and s denote the long
and short axis directions of the nanorods, respectively.

600nm to 1000nm. Consequently, because these responses are those acquired from

linear polarisations incident along the long and short axes of the nanorods, and the

fact that circular polarisation is essentially a rotating electric field vector (thereby

encapsulating all possible linear polarisation directions), it can be assumed that the

proposed method of obtaining such a response using the three-layer high efficiency

element is independent of the rotation angles Φ of the nanorods, and the phase

difference of π is expected to completely flip the handedness of the incident circular

polarisation upon reflection (thus preserving its handedness given that the

wavevector is also reversed). This means of obtaining a complete flipping of the

circular polarisation forms the basis of our high-efficiency GEMS, and ultimately the

ability to obtain high-efficiency holography. A simplified model is shown in Appendix

B, which gives the theoretical framework and more detail on explaining the high efficiency and broadband response of the nanorod metasurface.

To verify the high-efficiency of maintaining the same handedness of circular polarisation upon reflection, numerical simulations were performed using the commercially available software CST Microwave Studio and are shown in **Figure 6.3**. A uniform metasurface was modelled, where all of the rotation angles were fixed at 45° (to minimise coupling between adjacent nanorods as much as possible). The reflected wave contains both circular polarisation states, where one is the opposite handedness to that of the incident wave, and is unperturbed by the metasurface, whilst the other has the same handedness as the incident wave but with an additional phase delay of 2Φ, where Φ is the orientation angle of the nanorod antenna (in this case the angle is equal to 45°, thus the phase delay is equal to 90° or $\boldsymbol{\pi/2}$). For the specific geometry as shown in Figure 6.1, for a circularly polarised wave incident at normal incidence, the simulation results in **Figure 6.3** show that the reflectivity of light with the same circular polarisation (cross-polarisation in terms of the receiver) exceeds 80% reflectivity in a broad wavelength range between 550nm and 1,000nm, which covers nearly a full optical octave. Fascinatingly, even with the unavoidable Ohmic losses of the metals used (especially at visible and near-infrared frequencies), this efficiency is still astoundingly high.
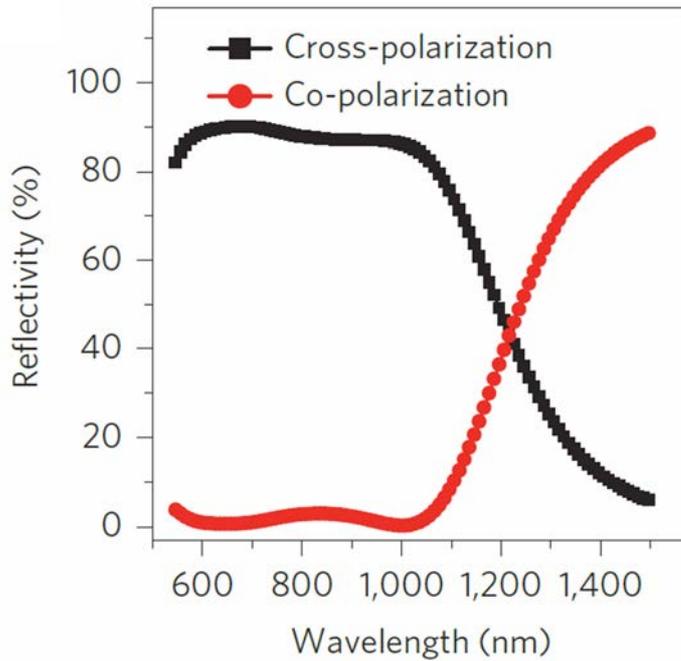
***Figure 6.3: Simulated cross-polarisation and co-polarisation
with normal light incidence for design in Fig. 5.1.*** [69]

Interestingly, the ohmic loss in this configuration is very close to that of light
transmitted through a single metasurface layer (without the metal ground plane)
around the resonance wavelength (800-850nm) of the antenna. Conversely, and
desirably, the efficiency of the unwanted handedness of reflected light (co-
polarisation with respect to the receiver) is extremely low (<3%) over the same
broadband wavelength range from 550nm to 1,000nm. *As a note, optimisation of the
high efficiency elements (including nanorod antenna and unit cell dimensions, and
the spacer layer thickness) was carried out to obtain the best conversion efficiency
whilst exhibiting the broadband response.*

To confirm the high efficiency of our numerical simulations we designed a
GEMS-based CGH, as shown in **Figure 6.4**. The CGH was designed for circularly

polarised light at normal incidence. A design where the holographic image appears

off axis was used to avoid overlapping of the holographic image and the zero-order

spot. A digital image of Einstein's portrait measuring 550×300 pixels and 256

greyscale levels (see inset of Figure 6.4b) was used for the holographic target image.

The CGH was designed to create a wide image angle of approximately $60° \times 30°$.

Because of this large angular range, the Rayleigh-Sommerfeld diffraction method was

used for simulating the holographic image [148], and the hologram was pre-

compensated to avoid pattern distortion. Additionally, a 2×2 periodic array of the

complete holographic phase pattern was used (**Figure 6.4c**) to avoid the formation

of laser speckles in the holographic image, where the concept of Dammann gratings

was utilised [149] – more information as to the advantage of the 2×2 periodic

arrangement over a single hologram phase pattern is provided in Appendix B, Figure

B.3. To create a holographic image with an image pixel array number measuring m

$\times$ n within the angular range $\alpha_x \times \alpha_y$ in the far-field, the period of the CGH

(metasurface array) in the x and y directions can be calculated according to

$$d_x = m\lambda/(2\,tan(\alpha_x/2)) \tag{6.1}$$

$$d_y = n\lambda/\big(2\,tan\big(\alpha_y/2\big)\big) \tag{6.2}$$

where the m and n correspond to the far-field (holographic) image pixel numbers.

The number of pixels in the CGH (metasurface array) is determined by $M = d_x/\Delta p$

and $N = d_y/\Delta p$, where Δp is the pixel size of the CGH in both x and y directions

(this can be a fixed value for a square unit cell, which is true for our case). Using the

above structural parameters, we developed a phase-only CGH using a metasurface

pixel dimension Δp of 300nm $\times$ 300nm and periods d_x, d_y of 333.3μm $\times$ 333.3μm (for

M $\times$ N = 1111 $\times$ 1111 pixels), which was designed according to the classical

Gerchberg-Saxton algorithm [150]. For the wavelength range in question, from
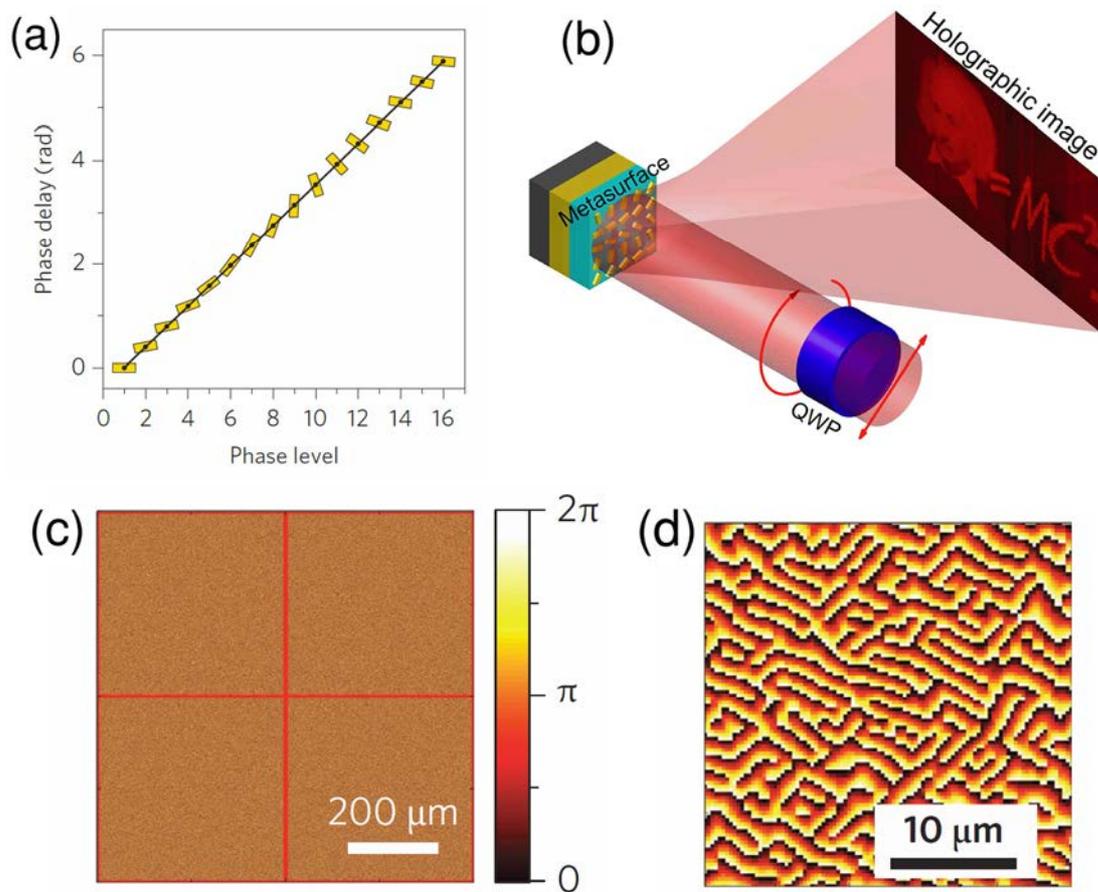
Figure 6.4: Working principle and phase distribution of the metasurface hologram [69]

(a) Phase delay for 16-level phase steps. On each data point is the corresponding nanorod orientation. **(b)** Illustration of the reflective nanorod-based CGH under a circularly polarised incident beam, which is converted from a linearly polarised beam upon passing through a quarter wave plate (QWP). The reflective beam forms the holographic image in the far field. **(c)** The 16-level phase distribution with 2 $\times$ 2 periods designed to generate the target holographic image in the far field. **(d)** Enlarged phase distribution (100 $\times$ 100 pixels) of the upper left corner of (c).

500nm-1,000nm, we can see that the pixel (unit cell) size is at least less than half of the wavelength and the hologram pattern is sampled at least twice the maximum spatial frequency in either direction, thereby satisfying the Shannon-Nyquist sampling theorem [151]. Because the geometric phase delay of the metasurface is determined solely by the orientation of the nanorod antennas, 16 phase-levels (**Figure 6.4a**) were used to obtain the high performance from the CGH. The phase distribution obtained for the hologram is shown in **Figure 6.4c**, with a zoomed in section in **Figure 6.4d**. Simulation shows that in our optimised design with an ideal hologram, and neglecting optical losses, the window efficiency, which is defined as the ratio between the optical power projected into the image region and the input power, reaches 94%.

6.3   Fabrication and Experimental Results

The metasurface CGH was fabricated on top of a silicon substrate following the design described above (**Figure 6.5a**). A 130nm thick gold metal ground layer was first evaporated onto the silicon substrate, and a 90nm thick layer of MgF_2 was deposited on top. Standard electron beam lithography was then carried out on a resist to produce the CGH pattern, and a 30nm deposition of gold, followed by lift-off processing gave us our gold metasurface CGH. To determine the conversion efficiency of our CGH, a linearly polarised beam from a Supercontinuum light source (Fianium Supercontinuum) was converted to circularly polarised light by the use of

a linear polariser followed by a quarter wave plate. A lens with focal length of 300mm was used to focus the incident beam onto the metasurface, and the reflected holographic image was collected by two identical condenser lenses with high numerical aperture (NA = 0.6) and focussed onto a power meter in the wavelength range: 400nm-1,1100nm. A schematic diagram of the optical setup is shown in **Figure 6.5b**. The incident beam was focussed onto the sample with a spot size of ~300μm in diameter, which is less than the size of the 2 × 2 CGH metasurface pattern (666.6 × 666.6μm^2). Additionally, colour filters were used to remove unwanted light generated by the Supercontinuum laser, and an iris was used to block the scattered beam from the multi-reflections between the optical interfaces. The incident laser was increased in wavelength from 600nm-1,100nm in steps of 25nm, and the corresponding holographic image output was measured. The optical efficiency (holographic window efficiency) was determined by subtracting the zeroth-order signal from the image intensity (**Figure 6.5c**). It is found that the sample exhibits a high window efficiency exceeding 50% over a broad spectral range from 630nm to 1,050nm. This reaches a maximum efficiency of 80% at a wavelength of 825nm, whilst simultaneously giving a zeroth-order efficiency of only ~2%, which is the unwanted co-polarised signal that does not undergo a phase delay. As well as these results, we did not observe the twin image effect which occurs for traditional binary holograms and negatively affects the efficiency.
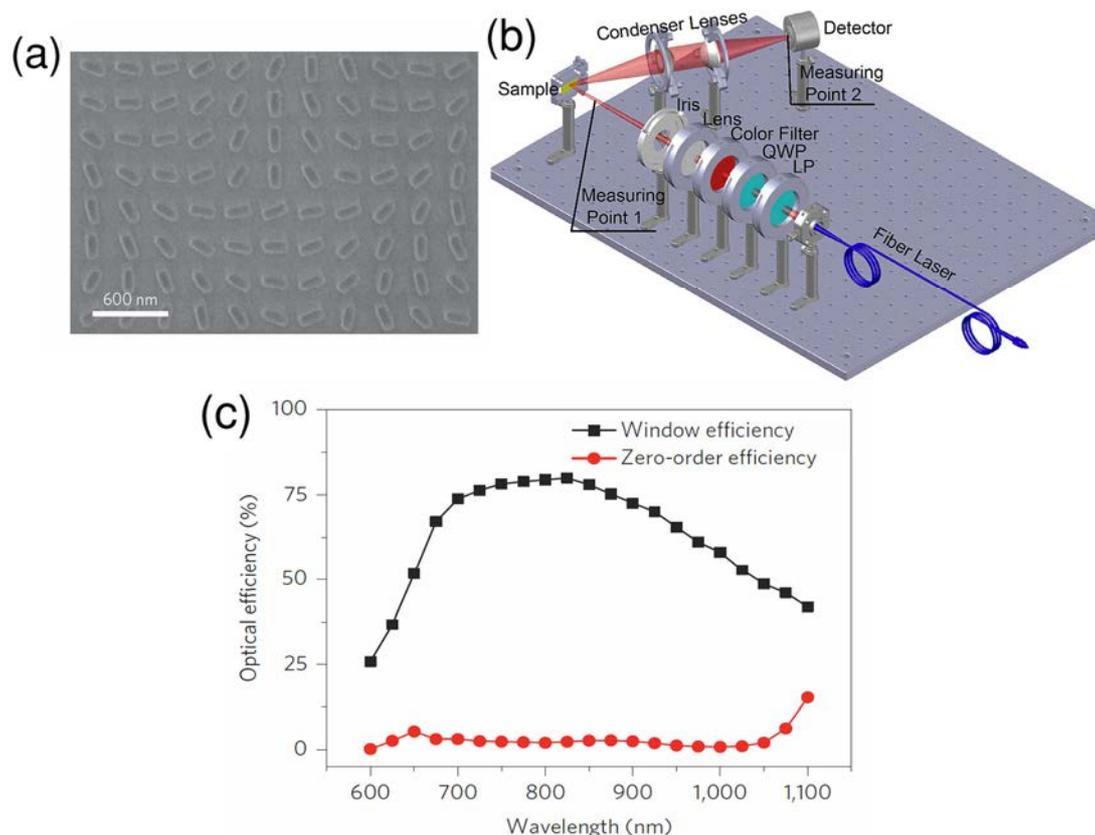
Figure 6.5: Scanning electron microscopy image, experimental optical setup and results for optical efficiency [69]

(a) Scanning electron microscopy image of the fabricated nanorod array (partial view). **(b)** Illustration of the optical efficiency measurement setup. The incident circularly polarised beam is focussed on the sample and the diffracted light collected by a power meter. **(c)** Experimentally obtained optical efficiency for both the image and zeroth-order beam.

Theoretically, the simulated metasurface hologram has broader spectral response (Figure 6.3) than that which occurs for the measured sample. This lessening of the broadband efficiency may possibly arise from the fact that the simulated conversion efficiency is obtained for a metasurface illuminated with normally incident light, whereas in the experiment the holographic image from the metasurface CGH was projected into a broad angular range (approximately 60° × 30°). It can be expected that this broad-angle scattering will result in a narrower bandwidth and

lessening of the reflected window efficiency when compared to the simulated results,

and that the holographic metasurface contains many neighbouring antenna which

are not of the same orientation angle – this may induce more apparent plasmonic

coupling effects, lessening the conversion efficiency, when compared to the fixed angle

simulated response in Figure 6.3.
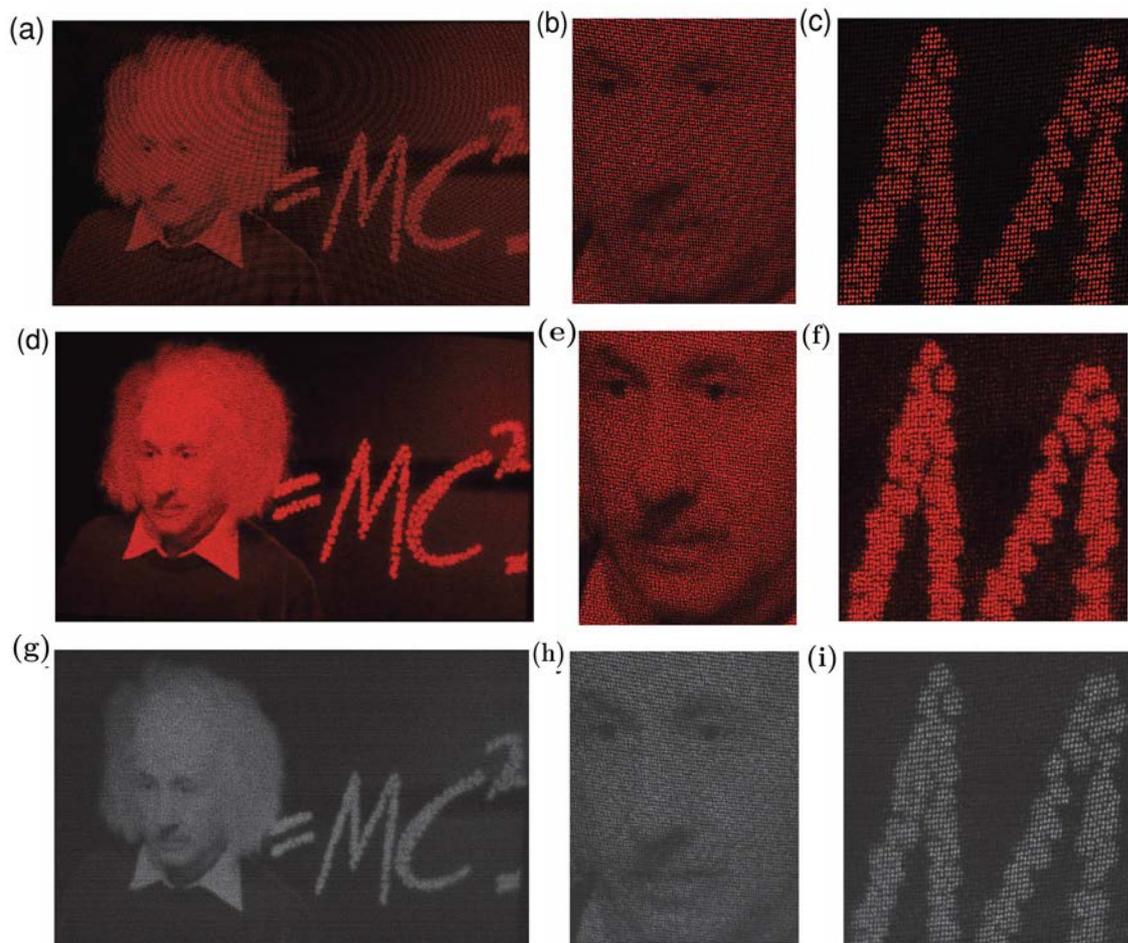
Figure 6.6: Simulated and optical verification of the holographic image at different wavelengths [69]

(a-c) Simulated holographic image of Einstein's portrait with an enlarged zoom of his face and the letter 'M'. **(d-f)** Experimentally obtained images captured by a 'visible' camera in the far-field. The operating wavelength is 633nm. **(g-i)** Experimentally obtained images captured by an infrared camera in the far-field, with an operating wavelength of 780nm.

Of major importance, in addition to the optical efficiency, is the quality of the actual holographic far-field image. Given in **Figure 6.6** are the simulated (**Figure 6.6a**) and experimentally (**Figure 6.6b,c**) obtained far-field images of the CGH encoded into the phase profile of the metasurface. These also include the zoomed-in views of Einstein's face and the letter 'M', which show a very good agreement between simulated and experimental results. To show that the metasurface design used was indeed broadband in operation, we captured the far-field holographic image at 2 wavelengths – the first was 633nm (red), using a commercial digital camera (Nikon D3200) for image-capture (**Figure 6.6b**) and a He-Ne laser for illumination, and the second wavelength was 780nm (near-infrared) using a fibre-coupled diode laser (New Focus) for illumination and an ELOP-Contour CMOS Infrared Digital Camera for capture (**Figure 6.6c**). The diameter of the laser beam is approximately 1.5mm, which fully covers the hologram, whilst the reflected holographic image was projected onto a white screen 300mm away from the metasurface sample. It is clear from the captured entire and zoomed-in images, especially paying close attention to the 'speckles' which form the holographic image, that the simulated and measured results are in very good agreement at both measured wavelengths. This demonstrates the high fidelity and applicability of the high-efficiency metasurface hologram design.

6.4   Conclusion

In summary, a reflective phase-only computer generated hologram based on geometric metasurfaces is realised with a diffraction efficiency as high as 80%, an extremely low zeroth-order efficiency less than 5%, and all occurring over a broad wavelength range in the visible to infrared. The design comprises of an ultrathin and uniform metasurface thickness of only 30nm, thus yielding negligible ohmic losses, and is compatible with scalar diffraction theory, even with subwavelength pixel sizes [152], which helps to greatly simplify the design of holograms. Due to the simple and robust phase control and its tolerance to variations in wavelength and fabrication errors, this geometric phase-based computer generated hologram design could overcome the current limitations of traditional depth-controlled holograms and could even find applications in field such as laser holographic keyboards, random spot generators for body motion, optical anti-counterfeiting and laser beam shaping. This approach can also be extended from phase-only to amplitude controlled holograms, simply by changing the size of the nanorods. One of the issues with this design, however, is that the mode of operation is reliant upon circularly incident polarisation, and thus is polarisation dependent, due to the geometric phase arising from the switching of the handedness between circular polarisation states – although this is beneficial in some sense as the initial conditions need not be so strict as to require special alignment of the polarisation with respect to the metasurface antenna.

Interestingly, though, this design can produce the twin-images which occur in traditional holograms simply by using linearly polarised incident light, which comes down to the fact that two perfectly interfering orthogonal circular polarisations result in a linearly polarised wave – in essence, when the sample is illuminated with linearly polarised light, it is then akin to being illuminated with both handedness' of circularly polarised light, which results in an image for each and occurring either side of the zeroth-order spot. Finally, we note that such nanorod metasurfaces could be fabricated on a large scale and at much lower costs by nano-imprinting, thus making them promising candidates for large-scale holographic technology.

Chapter 7

Summary and Future Research Applications

7.1 Thesis Summary

This thesis gives an investigation into the utilisation of metasurfaces for manipulating the wavefront of electromagnetic waves, with a particular focus on Terahertz devices, holographic applications, and high-efficiency circularly polarised light conversion, arising from geometric-phase. Metasurfaces are a strong candidate for future optical technologies, due to their small-size and ease of fabrication, pronounced wavefront control, and negligible losses – which is a strong factor in the current economic drive towards sustainability and better use of energy. Additionally, two of the works in this thesis take place at Terahertz frequencies, for which the useful components and devices in this frequency regime are in short supply due to the lack of naturally occurring or established semiconductor devices. The key results of this thesis can be summarised as follows:

- A monolithic silicon herringbone device was proposed and experimentally verified to achieve a high-efficiency of Circular Conversion Dichroism (CCD) at 1THz. By utilising Subwavelength Gratings (SWGs), we achieved a half-wave plate capable of fully converting the handedness of circularly polarised (CP) light into its orthogonal handedness; in conjunction with this, a geometric (Pancharatnam-Berry) phase was incorporated by a rotational alignment of 45° of the SWGs such that the phase of Left or Right circularly polarised light experienced equal and opposite phase accumulations of $\pi/2$ after traversing them. One of the two angled SWGs (per unit cell) was then raised with a substrate step such that a dynamic phase of $+\pi/2$ was applied to offset the $\pm\pi/2$ geometric phase, leading to an interference effect yielding destructive interference for one handedness conversion (Right-to-Left) and unaltered transmission for the other conversion (Left-to-Right). Analytical modelling using Fresnel's equations for a three-layer system gave good correspondence with FDTD simulation, and showed a conversion efficiency of ~80% at 1.1THz, whilst exhibiting a broadband performance, and prohibiting the opposite conversion pathway. Experimental verification of a fabricated sample gave a conversion performance exceeding ~60% at 0.9THz, where the mismatch between numerical and experimental results were attributed to the fabrication errors.

- A broadband Split-Ring Resonator (SRR) metasurface was proposed and experimentally verified to achieve the simultaneous control of both phase and amplitude of a wavefront at 1THz. Operating under a linear polarisation basis, C-shaped SRRs were proposed, utilising the behaviours of two types of antenna: where SRRs can be used to control the phase of scattered light through means of their geometries; whilst nanorod type antennas can control the amplitude of scattered linearly polarised light simply by varying the rotation angle. This resulted in SRRs which had varying angles between their symmetry axis and the incident light polarisation, to control the amplitude of scattered light; and varying geometries, to control the phase. As a proof of concept, these antennas were used to form a metasurface capable of controlling the intensities of arbitrary user-defined diffraction orders (up to three orders, m=-1,-2,-3), whilst exhibiting a broadband operation.

- Finally, the realisation and demonstration of a high-efficiency cross polarisation reflective metasurface was utilised to provide a broadband hologram in the visible-IR regime with an efficiency exceeding 80%. Gold nanorods separated from a metallic ground plane by a specific thickness of dielectric material gives rise to the high-efficiency for converting one circular polarisation into the opposite after reflection, due to a round-trip phase

difference between the long and short axes of the nanorods being equivalent to π, thus behaving like a half-wave plate. Geometric-phase control of the reflected wavefront is used to control the far-field interference by means of a simple rotation of the nanorods within their unit-cells and thus controlling the phase of the wavefront. This is then extended to a proof of concept design for a high-efficiency hologram, capable of converting 80% of the incident light into the holographic image at 825nm and obtaining a broadband response from 630nm to 1,050nm. Losses are negligible, and comparable to a single layer gold metasurface, where the reflected zero-order beam is less than 5%. This work demonstrates the highest efficiency visible-IR metasurface-induced holographic image to date (*at the time of writing*).

The work as laid out in this thesis has two common themes interlinking all of the projects – namely that they all have broadband operation, and that they also operate under "cross-polarisation" schemes: The first can be attributed to the dispersionless nature of the metasurfaces, whilst the second uses the careful interplay between the orthogonal polarisations and phase responses of the unit cells. The Silicon Herringbone structure in Chapter 4 has its dispersionless attribute stemming from both geometric phase and effective-medium derived refractive indices of the form-birefringent gratings. These properties allow the desired functionalities to exist independent of the wavelength of the light, where the broadband performance is only

diminished by the finite depths of the gratings and the silicon step leading to a variation of the imparted dynamic phases, and the periodicity of the gratings no longer being subwavelength (and hence no longer described by the effective medium approximation) for frequencies much higher than 1THZ. The cross-polarisation is due to the geometric phase dependent upon the handedness of the incident circularly polarised light. Opposite handednesses lead to equal but opposite phases imparted to the transmitted light, and an extra dynamic phase term leads to destructive or constructive handedness dependent energy pathways, which provides the desired functionality of the circular conversion dichroism effect. This work utilises both geometric and dynamic phases simultaneously to produce an effect typically only produced when using lossy metallic planar metasurfaces, except that this metasurface is not truly planar and is completely dielectric and has negligible losses. More impressively is that this novel approach to producing the CCD effect surpasses the cross-polarisation efficiency and asymmetry from all previous works from the past decade or so.

The C-antenna metasurface in Chapter 5 uses the interplay between symmetric and anti-symmetric resonance modes to produce the broadband capability. With the linearly incident polarisation parallel to the symmetry line of the C-antenna axis, a symmetric mode is produced at 0.8THz. When the linearly polarised incident wave is perpendicular to the symmetry line of the axis, two anti-symmetric modes are

produced at 0.35THz and 1.02THz, for the dipole and multipole resonances, respectively. By orienting the polarisation at 45° to either the symmetry axis or the 'anti-symmetry' axis (perpendicular to the symmetry axis) of the C-antenna, it is obvious that all three modes, symmetric and anti-symmetric, will be excited simultaneously to produce a broadband response between the lowest and highest resonance frequency. We know that a 45° linearly polarised wave is simply the superposition of two orthogonally polarised waves. Using this, we can instead keep the polarisation fixed along the x-axis, whilst we instead rotate the C-antenna to 45°. It is then clear that the C-antenna will produce both x- and y- polarised transmitted waves. However, we choose to measure only the y-polarised wave, orthogonal to the input wave, to avoid interferences and to additionally utilise the rotation angle of the C-antenna to control the amount of light which is converted into the y-axis, namely the transmission amplitude. The geometry of such an antenna is then altered to control the phase of the output wave, independent of the angularly resolved amplitude control. These two degrees of freedom allow us to utilise any combination of phase and amplitude to develop a metasurface capable of producing three diffraction orders. Such a mechanism is highly useful for holography and lensing applications, where no previous works were capable of smoothly controlling the phase and amplitude simultaneously in a trivial manner.

The high efficiency holographic metasurface in Chapter 6 has a complex interplay between the Fabry-Pérot resonance, reflection phases between the long and short axes of the nanorod antenna, and the geometric phase of the nanorod orientation to produce a broadband dispersionless cross-polarised reflected beam. It is realised that the dispersionless operation and cross-polarisation functionality are intertwined, where the phase differences of the round trip reflected light parallel to the long and short axes are π out of phase. Such a phase difference induces a half-wave plate functionality dependent upon the dielectric spacer between the nanorod and the ground-plane, which is dynamic in operation. This half-wave plate function produces a polarisation conversion for the incident circularly polarised light and therefore makes this device cross-polarisation in operation where the high-efficiency is accredited to the handedness switching capability. The broadband operation takes place between 630nm and 1,050nm due to the phase difference remaining at π between the long and short axis consistently. Such antennas are polarisation insensitive to circularly polarised light inasmuch as the in-plane orientation is concerned (as circularly polarised light is simply two linearly polarised waves superposed with a phase delay of a quarter-wave). This therefore allows us to produce a spatially varying geometric phase distribution which is applied to developing a high efficiency hologram. As it is only the cross-polarised light which undergoes the geometric phase accumulation, there is no interference between the incident and outgoing wave (the co-polarised wave has a low efficiency of 5% and is zero order)

and this allows us to encode nearly all of the incident energy into the holographic image. The only losses accounted for are ohmic due to the high plasmonic losses at optical frequencies. To date, at the time of writing, this method of reflective holography produces the highest efficiency of any metasurface hologram, whilst also providing a high resolution and needing very simple fabrication.

7.2   Future Work

The work in this this thesis covers the development of metasurfaces using dielectric, metallic, and plasmonic processes. All of these have the potential for applications in future research. Lately, work on dielectric metasurfaces has been of interest due to their negligible losses as illustrated in Chapter 4. Using this combination of both dynamic and geometric-phases can lead to very interesting applications, where either of these methods of phase-control alone is not sufficient or unable to produce the desired effect. In a similar fashion to the work carried out in Chapter 6, it may be of interest to see if a metallic reflective ground-plane in combination with the structure in Chapter 4 could yield similar and even better performance at achieving high-conversion efficiency between circular polarisations. Or instead, combine space-varying structures utilising varying angles of SWGs to perform multiple polarisation processes simultaneously.

The need for full colour devices has been crucial in the past few decades, ranging from TV's to smart phones. However, all of these devices use either LCD's or LED's,

which are typically larger than the wavelength of light used, whilst subwavelength pixels would have better applicability to developing these devices. An important step on the way to achieving this is to be able to have distinct colour and to have isolated resonances at RGB wavelengths. A full-colour hologram may be achieved through the combination of Chapters 5 and 6, except in a circular polarisation basis. This would give the opposite functionality of the SRRs in Chapter 5, as now the rotation angle induces a geometric control of phase, and the nanorod geometry controls the amplitude response. It is known from [153] that metasurfaces sat atop a very thin dielectric layer deposited on a metallic ground plane leads to strong absorption of light. The design in Chapter 6 works to provide a very broadband spectrum of operation, but for isolated colours a narrow resonance is required; therefore, by subtle control of the dielectric thickness, a balance between complete absorption and broadband operation can be used to produce single resonances for three differing antenna designs. With a careful lateral arrangement of the three unique nanorod antennas, it should be possible to encode three distinct holographic phase profiles such that the far-field holographic image is a combination of Red, Green, and Blue images, leading to a full-colour holographic image. Even though this efficiency will be much lower than for a broadband reflective hologram, due to the absorption effect, it could potentially still be higher than the holograms realised so far for transmission, which are typically below 20% efficiency.

Appendix A

Here, we show the derivation of the analytical model for the Silicon Herringbone device in Chapter 4. This modelling is based on traditional Fresnel equations for a three layer system, and includes multiplicative phase terms associated with the dynamic and geometric phases used to achieve Circular Conversion Dichroism.

We first begin by expressing the system in a Circularly Polarised basis, where we relate the unit vectors of right and left circularly polarised light in terms of the linear unit vectors as being:

$$\hat{e}_R = \frac{1}{\sqrt{2}}\left(\hat{e}_x + i\hat{e}_y\right) \tag{A.1}$$

$$\hat{e}_L = \frac{1}{\sqrt{2}}\left(\hat{e}_x - i\hat{e}_y\right) \tag{A.2}$$

where $\hat{e}_R$, $\hat{e}_L$ correspond to the Right and Left circular polarisation unit-vectors, respectively, and $\hat{e}_x$, $\hat{e}_y$ correspond to the x and y linear polarisation unit-vectors, respectively. Following this, it is useful to then define the *Transmission* coefficients for orthogonal circular polarisations, in terms of linear transmission, as being:

$$t_R = \frac{1}{\sqrt{2}}\left(t_x\hat{e}_x + it_y\hat{e}_y\right) \tag{A.3}$$

$$t_L = \frac{1}{\sqrt{2}}\left(t_x\hat{e}_x - it_y\hat{e}_y\right) \tag{A.4}$$

What we need to acquire, however, is the transmission coefficients of circularly polarised light in terms of circularly polarised unit vectors, not linear unit vectors. We can obtain the expression for both $\hat{e}_x$ and $\hat{e}_y$ in terms of the circular unit vectors as follows: firstly, by adding together equations A.1 and A.2, we get:

$$\hat{e}_R + \hat{e}_L = \frac{1}{\sqrt{2}}\left(\hat{e}_x + \hat{e}_x + i\hat{e}_y - i\hat{e}_y\right) = \sqrt{2}\hat{e}_x$$

and subtracting A.2 from A.1 gives us:

$$\hat{e}_R - \hat{e}_L = \frac{1}{\sqrt{2}}\left(\hat{e}_x - \hat{e}_x + i\hat{e}_y + i\hat{e}_y\right) = i\sqrt{2}\hat{e}_y$$

By rearranging the above, we then obtain the linear representations of circularly polarised light as:

$$\hat{e}_x = \frac{1}{\sqrt{2}}\left(\hat{e}_R + \hat{e}_L\right) \tag{A.5}$$

$$\hat{e}_y = \frac{1}{i\sqrt{2}}\left(\hat{e}_R - \hat{e}_L\right) \tag{A.6}$$

Because we now have these results, we can obtain the transmission coefficients for circularly polarised light in terms of circularly polarised unit vectors, which we obtain by inserting A.5 and A.6 into equations A.3 and A.4, respectively: after rearranging and collecting terms in $(\hat{e}_R, \hat{e}_L)$ the equations are as follows:

$$t_R \equiv t_R\hat{e}_R = \frac{1}{2}\hat{e}_R\left(t_x + t_y\right) + \frac{1}{2}\hat{e}_L\left(t_x - t_y\right) \tag{A.7}$$

$$t_L \equiv t_L\hat{e}_L = \frac{1}{2}\hat{e}_R\left(t_x - t_y\right) + \frac{1}{2}\hat{e}_L\left(t_x + t_y\right) \tag{A.8}$$

These equations are the key result of this model, where the transmission of circularly

polarised light is explained in terms of the linear transmission responses of a device.

We can easily calculate the linear response of a structure by using the Fresnel

equations; in the case for the silicon herringbone structure we assume a three-layer

system, where layer 1 is air, layer 2 is the SWG metasurface pattern, and layer 3 is

bulk silicon. Because the SWG has two distinct directions, labelled as TE (with the

polarisation of the normally incident wave being aligned with the gratings) and TM

(with the polarisation aligned perpendicular to the gratings), we correspondingly

have two Fresnel equations in the TE and TM directions – we label these, arbitrarily,

as the x and y directions for use in equations A.7 and A.8.

Referring back to Chapter 2 (and Chapter 4), we gave the Fresnel equation

for a three layer system as:

$$t_i^{fres} = \frac{t_{12_i} t_{23_i} e^{-i\phi}}{1 + r_{12_i} r_{23_i} e^{-2i\phi}} \tag{A.9}$$

where the subscript 'i' corresponds to the linear polarisation (x or y, relating to TE

or TM, respectively), and the terms in t, r, and ϕ are given as:

$$t_{12_i} = 2n_1 / (n_1 + n_{2i}) \tag{A.10a}$$

$$t_{23_i} = 2n_{2i} / (n_{2i} + n_3) \tag{A.10b}$$

$$r_{12_i} = (n_1 - n_{2i}) / (n_1 + n_{2i}) \tag{A.10c}$$

$$r_{23_i} = (n_{2i} - n_3) / (n_{2i} + n_3) \tag{A.10d}$$

$$\phi_i = \frac{2\pi d}{\lambda} n_{2i} \tag{A.10e}$$

The physical interpretation of, for example, t_{12_i} is the transmissivity between layers

1 and 2, along the i direction (TE or TM), which is governed by the refractive indices

of the corresponding layers. The term in n_{2i} is the refractive index along either the

TE or TM directions of the SWG, which is described using effective medium theory

from Chapter 2, using the equations:

$$n_{2x} = n_{TE} = \sqrt{Fn_I^2 + (1-F)n_{II}^2} \tag{A.11}$$

$$n_{2y} = n_{TM} = \frac{1}{\sqrt{\frac{F}{n_I^2} + \frac{(1-F)}{n_{II}^2}}} \tag{A.12}$$

where F is the duty cycle of the grating (set as 0.5) and (n_I, n_{II}) are the refractive

indices of air and bulk silicon (at 1THz), respectively. The calculated values for these

refractive indices (as shown in Chapter 4) are given as $n_{2x} = n_{TE} = 2.55$ and $n_{2y} = n_{TM} = 1.36$. Using equations A.9 – A.12, we are able to calculate the transmittances

of equations A.7 and A.8. It is clear that for either of these equations, they are

dependent upon the unit vector $\hat{e}_R$ or $\hat{e}_L$, even though the combined actual

transmission is solely Left or Right; therefore, we can deduce that each of these

equations corresponds to a co-polarisation and cross-polarisation term. This can

quantitatively given as:

$$t_R = t_{RR} + t_{RL} \quad \rightarrow \quad t_{RR} = \frac{1}{2}(t_x + t_y) \quad, \quad t_{RL} = \frac{1}{2}(t_x - t_y) \tag{A.13a,b}$$

$$t_L = t_{LL} + t_{LR} \quad \rightarrow \quad t_{LL} = \frac{1}{2}(t_x + t_y) \quad, \quad t_{LR} = \frac{1}{2}(t_x - t_y) \tag{A.13c,d}$$

where we have removed the unit vectors $(\hat{e}_R, \hat{e}_L)$ as the subscripts of t_{ij} infer this

information directly (e.g. LR corresponds to the conversion of incident Right to

transmitted Left, $\hat{e}_R \rightarrow \hat{e}_L$) due to the conversion properties of the device. The terms

in LR and RL are therefore the cross-polarisation terms of the Jones matrix, and the

terms in RR and LL are the co-polarisation terms.

We know that one unit cell of the silicon herringbone metasurface consists of

two mirrored SWGs. Therefore, we can summate the transmission of one of these

SWGs that is unaltered with another that has multiplicative phase terms to arrive

at the total response of such a unit cell; the transmission coefficient for the second

of the two SWGs is simply identical to the first, except having terms related to the

geometric and dynamic phases. The term in dynamic phase, corresponding to a $+\pi/2$

additional phase, is simply $e^{i\phi_{Dyn}} = e^{i\Delta n_{Si-Air}(2\pi d/\lambda)} = e^{i(n_3-n_1)\frac{2\pi d}{\lambda}}$, whilst the

geometric phase, which corresponds to either $\pm\pi/2$ (with $\pm$ depending upon the

incident handedness being RCP or LCP, respectively, and only for cross-polarisation)

or 0 (for co-polarisation), is simply $e^{i\phi_{Geom}} = e^{\pm i\frac{\pi}{2}}$ (or $e^{\pm i0} = 1$). We then express

the transmission coefficients for the second SWG as follows:

$$\tilde{t}_{RR} = t_{RR} \times e^{i\phi_{Dyn}} = t_{RR}e^{i(n_3-n_1)\frac{2\pi d}{\lambda}} \tag{A.14a}$$

$$\tilde{t}_{RL} = t_{RL} \times e^{i\phi_{Dyn}} \times e^{i\phi_{Geom}} = t_{RL}e^{i[(n_3-n_1)\frac{2\pi d}{\lambda}-\frac{\pi}{2}]} \tag{A.14b}$$

$$\tilde{t}_{LL} = t_{LL} \times e^{i\phi_{Dyn}} = t_{LL}e^{i(n_3-n_1)\frac{2\pi d}{\lambda}} \tag{A.14c}$$

$$\tilde{t}_{LR} = t_{LR} \times e^{i\phi_{Dyn}} \times e^{i\phi_{Geom}} = t_{LR}e^{i[(n_3-n_1)\frac{2\pi d}{\lambda}+\frac{\pi}{2}]} \tag{A.14d}$$

where the geometric phase is omitted for the co-polarisation terms (RR and LL) and

the terms in RL and LR have either – or + geometric phase, respectively.

We now have the both of the transmission responses for both SWGs in a unit

cell, and so we can add together the matching terms from A.13-A.14 and divide by

two:

$$t_{RR}^{TOTAL} = \frac{1}{2}(\tilde{t}_{RR} + t_{RR}) \tag{A.15a}$$

$$t_{RL}^{TOTAL} = \frac{1}{2}(\tilde{t}_{RL} + t_{RL}) \tag{A.15b}$$

$$t_{LL}^{TOTAL} = \frac{1}{2}(\tilde{t}_{LL} + t_{LL}) \tag{A.15c}$$

$$t_{LR}^{TOTAL} = \frac{1}{2}(\tilde{t}_{LR} + t_{LR}) \tag{A.15d}$$

The above equations, A.15a-d, give us the total transmitted amplitude response of

one unit cell of the silicon herringbone metasurface. A more interesting result is the

intensity, which is the fraction of energy being transmitted; to get this, we look back

to equation 2.41 from chapter 2, which is $T_{ij} = \frac{n_3}{n_1}\left|t_{ij}^{TOT}\right|^2$ for normal incidence. The

equations above, A.15a-d, are then modified to give T_{ij}, the intensity:

$$T_{ij} = n_3 \left|\frac{1}{2}t_{ij}\left[1 + e^{i(\phi_{Dyn}+\phi_{Geom})}\right]\right|^2 \tag{A.16}$$

where we have set $n_1 = 1$ for air. Using this equation, we can then numerically model

the transmitted response of the herringbone device over a frequency range of the

order of 1THz, shown in Chapter 4, a.

Appendix B

Physical Modelling of broadband capability

Here, a simple model based on Fresnel's equations for a three-layer system in conjunction with the Lorentzian polarisability for nano-antennas is shown to provide the dispersionless and high efficiency conversion of circularly polarised light by the reflectarray developed in Chapter 6. *This appendix is based on, and contains excerpts, from the supplementary information of reference* [69].

Firstly, we construct a three-layer model, with a plasmonic nano-antenna layer (layer 2), presumed to be very thin (thickness *d)* and homogeneous, being sandwiched between layers 1 and 3. The layers have corresponding refractive indices n_1, n_2, and n_3, respectively. We employ the use of the three-layer Fresnel equation for transmission, described in Chapter 2, equation 2.43b, as:

$$t_i^{fres} = \frac{t_{12_i}t_{23_i}e^{-i\phi}}{1+r_{12_i}r_{23_i}e^{-2i\phi}} \tag{B.1}$$

where the terms in in *t, r,* and ϕ are given in Appendix A, equations A.10a-e (taken from Chapter 2). We wish to express equation B.1 in a simpler form, which we can
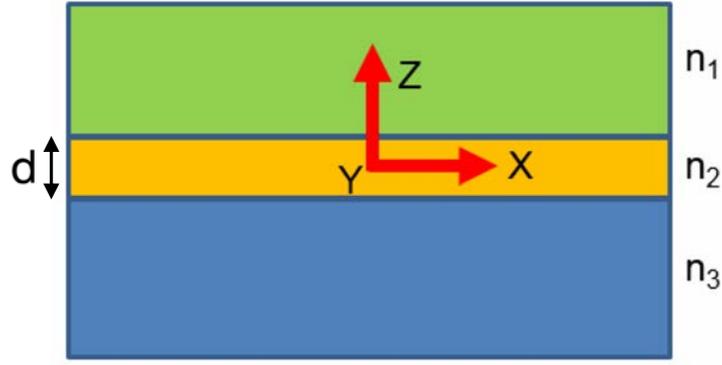
Figure B.1: Simple schematic of the three-layer system

achieve by taking the inverse of t, and upon expansion of the exponential terms to trigonometric sine and cosine terms (after some lengthy algebra), we obtain the following equivalent expression:

$$\frac{1}{t} = t^{-1} = \frac{1}{2}\cos\phi\left[1 + \frac{n_3}{n_1}\right] + \frac{i}{2}\sin\phi\left[\frac{n_2}{n_1} + \frac{n_3}{n_2}\right] \tag{B.2}$$

where $\phi = n_2 k_0 d$. Similarly, we can obtain for terms in r multiplied by inverse t as being:

$$rt^{-1} = \frac{1}{2}\cos\phi\left[1 - \frac{n_3}{n_1}\right] + \frac{i}{2}\sin\phi\left[\frac{n_3}{n_2} - \frac{n_2}{n_1}\right] \tag{B.3}$$

$$t'^{-1} = \frac{1}{2}\cos\phi\left[1 + \frac{n_1}{n_3}\right] + \frac{i}{2}\sin\phi\left[\frac{n_1}{n_2} + \frac{n_2}{n_3}\right] \tag{B.4}$$

$$r't'^{-1} = \frac{1}{2}\cos\phi\left[1 - \frac{n_1}{n_3}\right] + \frac{i}{2}\sin\phi\left[\frac{n_1}{n_2} - \frac{n_2}{n_3}\right] \tag{B.5}$$

where we have used that t and r are the transmission and reflections coefficients incident from medium 1, and t' and r' are those incident from medium 3. We can relate the susceptibility, χ, of medium 2 to the polarisability, α, of an individual antenna as:

$$\chi = \frac{\alpha}{a^2 d} \tag{B.6}$$

with a^2 being the unit cell size (squared) enveloping an antenna and d is the thickness

of this layer as shown in **Figure B.1**. The resonance of the plasmonic antennas leads

to the polarisability having the typical well-known Lorentzian form of:

$$\alpha \propto -\frac{1}{\omega - \omega_0 + i\gamma} \tag{B.7}$$

We saw in Chapter 2 that the relative permittivity, and thus the refractive index, of

a medium is related to its susceptibility by $\varepsilon_r = 1 + \chi \rightarrow n = \sqrt{\varepsilon_r} = \sqrt{1 + \chi} \approx \frac{1}{a}\sqrt{\frac{\alpha}{d}}$

where we have used the fact that if d is sufficiently thin, then the term in χ will be

much greater than 1 (due to the denominator containing d blowing up), and so we

approximate $1 + \chi \approx \chi$. If we now set the refractive indices of the three media as

$n_1 = 1$ (air), $n_3 = n_s$ (substrate), and $n_2 = n$, we then modify the term for the

inverse transmission incident from air (equation B.2) as being:

$$t^{-1} = \frac{1}{2}\cos(nk_0 d)\,[1 + n_s] + \frac{i}{2}\sin(nk_0 d)\left[n + \frac{n_s}{n}\right] \tag{B.8}$$

If we employ the small-angle approximations for $\sin\theta \approx \theta$ and $\cos\theta \approx 1 - \frac{\theta^2}{2}$, we

then acquire:

$$t^{-1} \approx \frac{1}{2}\left(1 - \frac{(nk_0 d)^2}{2}\right)[1 + n_s] + +\frac{i}{2}(nk_0 d)\left[n + \frac{n_s}{n}\right]$$

$$\Rightarrow \frac{1 + n_s}{2} - \frac{(nk_0 d)^2[1 + n_s]}{4} + \frac{i}{2}(k_0 d)(n^2 + n_s)$$

$$\Rightarrow \frac{1+n_s}{2} - \frac{\alpha k_0{}^2 d[1+n_s]}{4a^2} + \frac{i}{2}(k_0 d)\left(\frac{\alpha}{a^2 d} + n_s\right)$$

$$\Rightarrow \frac{1+n_s}{2} - \frac{\alpha k_0{}^2 d[1+n_s]}{4a^2} + \frac{ik_0\alpha}{2a^2} + \frac{ik_0 dn_s}{2}$$

where we can then set the terms containing d to zero (as d is very small), and we then obtain:

$$t = 1/\left(\frac{1+n_s}{2} + \frac{i}{2}\frac{k_0\alpha}{a^2}\right) \tag{B.9}$$

We can associate the term containing the polarisability as being

$$P = \frac{k_0}{\omega}\frac{\alpha}{2a^2} = \frac{\alpha}{2ca^2} = -\frac{g}{\omega-\omega_0+i\gamma} \tag{B.10}$$

which then modifies equation B.9 to give:

$$t = 1/\left(\frac{1+n_s}{2} + i\omega P\right) \tag{B.11a}$$

The term g is a coefficient to govern the coupling strength between the antenna and the incident light. We can do the same process as above for calculating the form of the reflection coefficient, and also the transmission and reflection coefficients from medium 3, to give:

$$r = \left(\frac{1-n_s}{2} - i\omega P\right)/\left(\frac{1+n_s}{2} + i\omega P\right) \tag{B.11b}$$

$$t' = n_s/\left(\frac{1+n_s}{2} + i\omega P\right) \tag{B.11c}$$

$$r' = \left(\frac{n_s-1}{2} - i\omega P\right)/\left(\frac{1+n_s}{2} + i\omega P\right) \tag{B.11d}$$

The issue remains that we have unknown variables within the equations B.11a-d, namely g, γ and ω_0. These unknown variables were obtained through the fitting of

the above equations to the simulated results for transmission of light incident from the air side onto a monolayer of antennas with a uniform orientation (fixed in-plane angle of rotation) sitting atop an MgF_2 dielectric substrate, using Comsol software (not performed by myself). The results of the simulation and fitting are shown in **Figure B.2** below, and fitting parameters resulted in $g = -0.19620$, $\gamma = 2\pi \times 6.35 \times 10^{12}$ rad/s and $\omega_0 = 2\pi \times 3.53 \times 10^{14}$ rad/s.

The previous was simply the modelling of a monolayer of antennas sandwiched
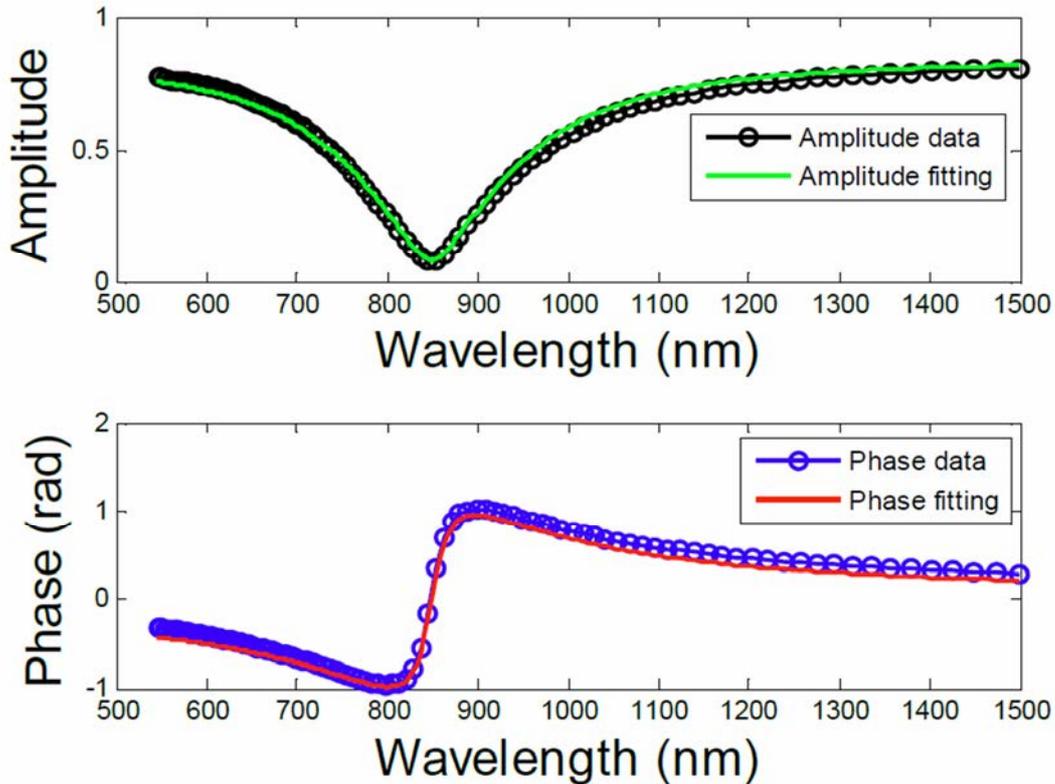
Figure B.2: Numerical simulations and fitted data of transmitted amplitude and phase [67]

(upper) Amplitude coefficients for light transmitted from the air side onto a monolayer of antennas with MgF_2 substrate, polarised parallel to the antennas long axis. (lower) the corresponding phase coefficients for the same system. The circles represent data points obtained numerically, whilst the solid line represent the fitting of equation B.11a.

between two bulk dielectrics (air and MgF$_2$) and a standard Lorentzian resonance of said antennas. Now, we shall look at the modelling of the full reflectarray structure, which includes the addition of a thick metal-ground plane beneath the MgF$_2$ layer which acts as a mirror for the impinging light. The reflection coefficient for this system (Antenna Sheet / MgF$_2$ / Metal Ground-plane) is then given by:

$$R = r + \frac{tt'r_m e^{i2n_3 k_0 d_3}}{1 - r'r_m e^{i2n_3 k_0 d_3}} \approx r + \frac{tt'e^{i\beta}}{1 - r'e^{i\beta}} \qquad (B.12)$$

with r_m being the complex reflection coefficient at the MgF$_2$/Ground-plane interface and $\beta = 2n_3 k_0 d_3 + \varphi(r_m)$ is the round trip phase of the MgF$_2$ layer plus that of the reflection at the ground plane (n_3 is the refractive index of MgF$_2$, d_3 is the thickness of the MgF$_2$ and $\varphi(r_m)$ the phase of the complex reflection coefficient r_m) where we have omitted the term in r_m from the approximated equation B.12 by including it in the term for β. If we substitute the equations for *t, r, t',* and *r'* from equations B.11a-d into equation B.12, we get:

$$R = -\frac{\frac{n_s-1}{2}(\omega-\omega_0+i\gamma)+i\omega g - \left[\frac{n_s+1}{2}(\omega-\omega_0+i\gamma)-i\omega g\right]e^{i\beta}}{\frac{n_s+1}{2}(\omega-\omega_0+i\gamma)+i\omega g - \left[\frac{n_s-1}{2}(\omega-\omega_0+i\gamma)-i\omega g\right]e^{i\beta}} \qquad (B.13)$$

If we choose that the total round trip phase $\beta = 2\pi$ and that the frequency of the light is equal to the resonance frequency $\omega = \omega_0$, the above equation then simplifies to:

$$R = -\frac{-i\gamma+i2\omega_0 g}{i\gamma+i2\omega_0 g} = -\frac{2\omega_0 g-\gamma}{2\omega_0 g+\gamma} \qquad (B.14)$$

This implies that at resonance, and ensuring a total round trip phase of 2π, the

complex reflection coefficient is negative, and implies a phase shift of π compared to the incident light. In turn, we only assume the light interacts with the long axis of the antenna at resonance and not with the short axis. For light polarised along the short axis, the parameter $g = 0$ and the phase of the reflection coefficient is approximately 2π which is the sum of the round trip phase (π) and the reflection phase at the metal ground plane (π). This is clear that if we set $g = 0$ in equation B.14, we simply obtain that R=1. Therefore, the phase difference between the long and short axes is equivalent to π, which is physically and phenomenologically similar to the case for a half wave plate. In essence, the rotation direction, namely the handedness, of circularly polarised light is completely reversed upon reflection from such a three-layer structure.

Following this, the broadband properties, namely the dispersionless nature, needs to be accounted for – this is attributed to the phase difference between the long and short axis phase difference being dispersionless near the resonance frequency. Around this resonance frequency, we can express the phase as:

$$\beta \approx f_\varphi \Delta\omega \qquad\qquad (B.15)$$

where $\Delta\omega \approx \omega - \omega_0$ when near the resonance frequency, and the term in f_φ represents the frequency dispersion of the phase and contains the dispersion properties of both the round trip phase (in the MgF$_2$ dielectric layer) and the

reflection phase at the metal ground-plane. Substituting this approximation for β into equation B.13 and expanding around $\Delta\omega$ yields:

$$R = -\frac{\frac{n_s-1}{2}(\Delta\omega+i\gamma)+i\omega g-\left[\frac{n_s+1}{2}(\Delta\omega+i\gamma)-i\omega g\right](1+if\Delta\omega)}{\frac{n_s+1}{2}(\Delta\omega+i\gamma)+i\omega g-\left[\frac{n_s-1}{2}(\Delta\omega+i\gamma)-i\omega g\right](1+if\Delta\omega)} \qquad \text{(B.16)}$$

We can neglect the damping term in $i\gamma$ as it is much smaller than $i\omega g$, and also the terms which are quadratic in $\Delta\omega$ (as $\Delta\omega^2$ will be very small) which gives us:

$$R \approx -\frac{2\omega g+i\Delta\omega(1+f\omega g)}{2\omega g-i\Delta\omega(1-f\omega g)} \qquad \text{(B.17)}$$

We then approximate the phase of R as:

$$\varphi(R) \approx \pi + \frac{\Delta\omega(1+f\omega g)-\Delta\omega(1-f\omega g)}{2\omega g} = \pi + \frac{2\Delta\omega f\omega g}{2\omega g} = \pi + f\Delta\omega = \pi + \beta \qquad \text{(B.18)}$$

Along the short axis, the reflection phase is simply very close to β with a slight deviation due to the refractive index mismatch between air and MgF_2 (which are included in the original expressions using the Fresnel equations and giving rise to weak Fabry-Pérot effects). Therefore, the phase difference between the long and short axes, at resonance, is simply π, and therefore dispersionless (as it is absent of terms in β) which leads to the broadband ability of the reflectarray design.

Dammann Gratings concept for Hologram design

This work involved the design of a far field hologram which used a 2×2 pattern array, as opposed to a 1×1 array. This approach is based on the concept of Dammann gratings [149,154], and is particularly beneficial in the reduction of laser
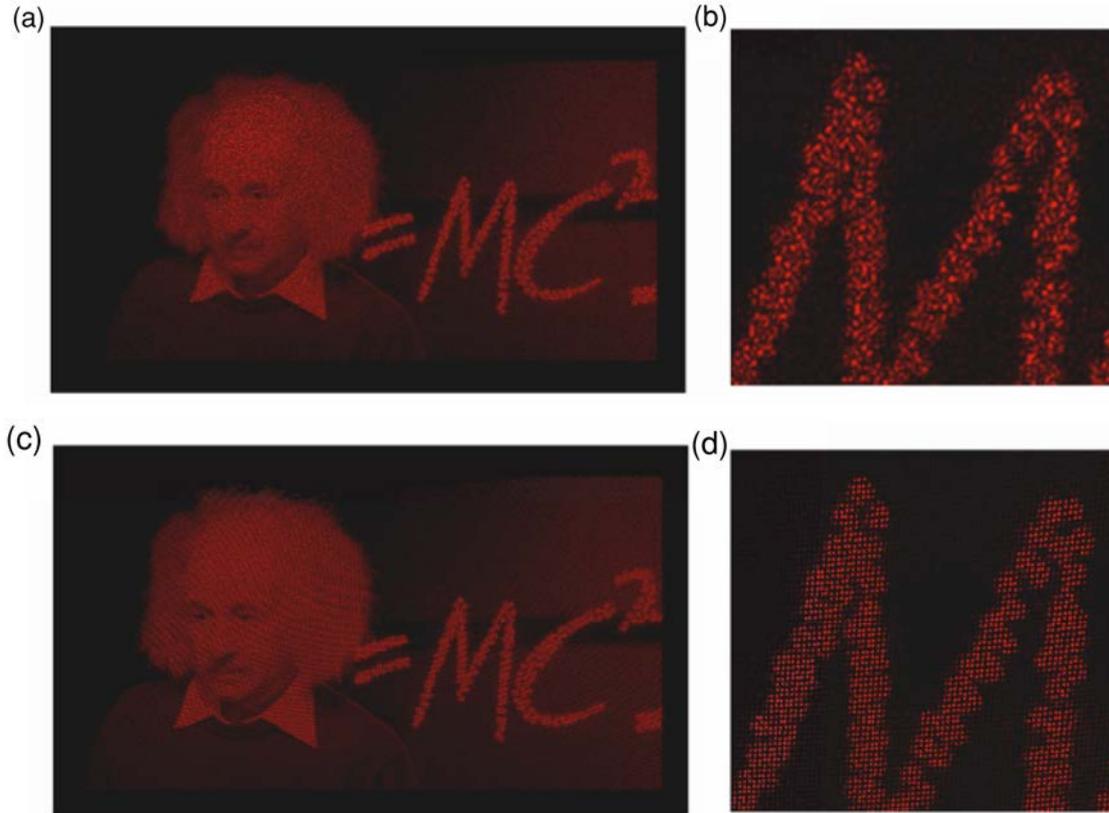
***Figure B.3: Simulated Far-field images when using either
a single or 2×2 hologram array* [67]**

(a) and (b) The simulated far-field image of Einstein and a zoomed in letter
'M' whilst using single hologram pattern array. **(c) and (d)** The simulated
far-field image and a zoomed in letter 'M' using a 2 × 2 hologram array. It is
clear that the 2 × 2 provides a much more uniform contrast and defined
image, with less laser speckles.

'speckles' [155] which form the pixels of the far-field holographic image, and therefore

improves the image quality. To this end, the Einstein image as portrayed in Chapter

6 was used to simulate far-field images for both a single and a 2 × 2 hologram arrays,

and are shown in **Figure B.3**. Figs 6.3a,b correspond to the simulated far-field image

when using only a single patterned array, whilst Figs 6.3c,d is the same simulated

far-field image but with a 2 × 2 pattern array. It is obvious that for the 2 × 2 array

a much better contrast of the pixels and definition of the image is produced, with

much more discrete spots, whereby for a single hologram the image, especially when

zoomed in, has obvious laser speckling and pixel brightness issues. This method of

using 2×2 patterned arrays can be scaled up to $N \times N$ (N being an integer)

patterned arrays to provide an even greater image fidelity, however this will increase

fabrication time and require a larger laser beam spot size to cover all of these

holograms.

References

1. Marchant, J. "In search of lost time". *Nature* **444,** 534–538 (2006).

2. Steele, J. M. "Celestial Measurement in Babylonian Astronomy". *Ann. Sci.* **64,** 293–325 (2007).

3. Gopher, a., Tsuk, T., Shalev, S. & Gophna, R. "Earliest Gold Artifacts in the Levant". *Curr. Anthropol.* **31,** 436–443 (1990).

4. Leonhardt, U. "Optical metamaterials: Invisibility cup". *Nat. Photonics* **1,** 207–208 (2007).

5. Freestone, I., Meeks, N., Sax, M. & Higgitt, C. "The Lycurgus Cup — A Roman nanotechnology". *Gold Bull.* **40,** 270–277 (2007).

6. Odom, T. W. "Colours at the nanoscale: Printable stained glass". *Nat. Nanotechnol.* **7,** 550–551 (2012).

7. Padovani, S., Puzzovio, D., Sada, C., Mazzoldi, P., Borgia, I., Sgamellotti, A., Brunetti, B. G. G., Cartechini, L., D'Acapito, F., Maurizio, C., Shokoui, F., Oliaiy, P., Rahighi, J., Lamehi-Rachti, M., Pantos, E., D'acapito, F., Maurizio, C., Shokoui, F., Oliaiy, P., Rahighi, J., Lamehi-Rachi, M. & Pantos, E. "XAFS study of copper and silver nanoparticles in glazes of medieval middle-east lustreware (10th–13th century)". *Appl. Phys. A Mater. Sci. Process.* **83,** 521–528 (2006).

8. "British Museum - The Lycurgus Cup". at "*http://www.britishmuseum.org/explore/highlights/highlight_objects/pe_mla/t/the_lycurgus_cup.aspx*" on 8 September 2015

9. "Christ of Wissembourg". at "*https://upload.wikimedia.org/wikipedia/commons/3/3e/Christ_de_Wissembourg.JPG*" on 8 September 2015

10. MSFC, S. M.-. "NASA - Telescope History". at "*http://www.nasa.gov/audience/forstudents/9-*

12/features/telescope_feature_912.html" on 24 September 2015

11. Drake, S. "*Galileo at Work: His Scientific Biography*". (University of Chicago Press, 1978).

12. Einstein, A. "Zur Elektrodynamik bewegter Körper". *Ann. Phys.* **322,** 891–921 (1905).

13. Weinstein, G. "Einstein Chases a Light Beam". **2,** 411–462 (2005).

14. Darwin, C. "*On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*". (J. Murray, 1859).

15. Maxwell, J. C. "A Dynamical Theory of the Electromagnetic Field". *Philos. Trans. R. Soc. London* **155,** 459–512 (1865).

16. Bose, J. C. "On the selective conductivity exhibited by some polarising substances". *Roy. Proc. Soc.* **60,** 433–436 (1896).

17. Franks, M. E., Macpherson, G. R. & Figg, W. D. "Thalidomide". *Lancet* **363,** 1802–1811 (2004).

18. Sengupta, D. L. & Sarkar, T. K. "Maxwell, Hertz, the Maxwellians, and the early history of electromagnetic waves". *IEEE Antennas Propag. Mag.* **45,** 13–19 (2003).

19. Hertz, H. "*Electric Waves: Being Researches on the Propagation of Electric Action with Finite Velocity Through Space*". (Macmillan and Company, 1893).

20. Munk, B., Kouyoumjian, R. & Peters, L. "Reflection properties of periodic surfaces of loaded dipoles". *IEEE Trans. Antennas Propag.* **19,** 612–617 (1971).

21. Chen, C. "Scattering by a two-dimensional periodic array of conducting plates". *IEEE Trans. Antennas Propag.* **18,** 660–665 (1970).

22. Ott, R. H., Kouyoumijian, R. G. & Peters, L., J. "Scattering by a two-dimensional periodic array of narrow plates". *Radio Sci.* **2,** 1347–1359 (1967).

23. Munk, B. A. "*Frequency Selective Surfaces: Theory and Design*". (John Wiley & Sons, 2005).

24. Pendry, J. B. "Negative Refraction Makes a Perfect Lens". *Phys. Rev. Lett.* **85,** 3966–3969 (2000).

25. Shelby, R. A., Smith, D. R. & Schultz, S. "Experimental verification of a negative index of refraction". *Science* **292,** 77–79 (2001).

26. Zhang, S., Fan, W., Panoiu, N. C., Malloy, K. J., Osgood, R. M. & Brueck, S. R. J. "Experimental Demonstration of Near-Infrared Negative-Index

Metamaterials". *Phys. Rev. Lett.* **95,** 137404 (2005).

27. Shalaev, V. M., Cai, W., Chettiar, U. K., Yuan, H.-K., Sarychev, A. K., Drachev, V. P. & Kildishev, A. V. "Negative index of refraction in optical metamaterials". *Opt. Lett.* **30,** 3356–3358 (2005).

28. Silveirinha, M. & Engheta, N. "Tunneling of Electromagnetic Energy through Subwavelength Channels and Bends using ε -Near-Zero Materials". *Phys. Rev. Lett.* **97,** 157403 (2006).

29. Alù, A., Silveirinha, M. G., Salandrino, A. & Engheta, N. "Epsilon-near-zero metamaterials and electromagnetic sources: Tailoring the radiation phase pattern". *Phys. Rev. B* **75,** 155410 (2007).

30. Huang, X., Lai, Y., Hang, Z. H., Zheng, H. & Chan, C. T. "Dirac cones induced by accidental degeneracy in photonic crystals and zero-refractive-index materials". *Nat. Mater.* **10,** 582–586 (2011).

31. Schurig, D., Mock, J. J., Justice, B. J., Cummer, S. A., Pendry, J. B., Starr, A. F. & Smith, D. R. "Metamaterial electromagnetic cloak at microwave frequencies". *Science* **314,** 977–980 (2006).

32. Valentine, J., Li, J., Zentgraf, T., Bartal, G. & Zhang, X. "An optical cloak made of dielectrics". *Nat. Mater.* **8,** 568–571 (2009).

33. Ergin, T., Stenger, N., Brenner, P., Pendry, J. B. & Wegener, M. "Three-dimensional invisibility cloak at optical wavelengths". *Science* **328,** 337–339 (2010).

34. Fang, N., Lee, H., Sun, C. & Zhang, X. "Sub-diffraction-limited optical imaging with a silver superlens". *Science* **308,** 534–537 (2005).

35. Taubner, T., Korobkin, D., Urzhumov, Y., Shvets, G. & Hillenbrand, R. "Near-field microscopy through a SiC superlens". *Science* **313,** 1595. (2006).

36. Liu, Z., Lee, H., Xiong, Y., Sun, C. & Zhang, X. "Far-field optical hyperlens magnifying sub-diffraction-limited objects". *Science* **315,** 1686. (2007).

37. Hasman, E., Kleiner, V., Biener, G. & Niv, A. "Polarization dependent focusing lens by use of quantized Pancharatnam-Berry phase diffractive optics". *Appl. Phys. Lett.* **82,** 328–330 (2003).

38. Bomzon, Z., Biener, G., Kleiner, V. & Hasman, E. "Space-variant Pancharatnam-Berry phase optical elements with computer-generated subwavelength gratings". *Opt. Lett.* **27,** 1141–1143 (2002).

39. Hasman, E., Bomzon, Z., Niv, A., Biener, G. & Kleiner, V. "Polarization beam-splitters and optical switches based on space-variant computer-generated

subwavelength quasi-periodic structures". *Opt. Commun.* **209,** 45–54 (2002).

40. Yu, N., Genevet, P., Kats, M. a., Aieta, F., Tetienne, J.-P., Capasso, F. & Gaburro, Z. "Light propagation with phase discontinuities: generalized laws of reflection and refraction". *Science* **334,** 333–337 (2011).

41. Chen, X., Huang, L., Mühlenbernd, H., Li, G., Bai, B., Tan, Q., Jin, G., Qiu, C.-W., Zhang, S. & Zentgraf, T. "Dual-polarity plasmonic metalens for visible light". *Nat. Commun.* **3,** 1198. (2012).

42. Khorasaninejad, M. & Crozier, K. B. "Silicon nanofin grating as a miniature chirality-distinguishing beam-splitter". *Nat. Commun.* **5,** 5386. (2014).

43. Fowles, G. R. "*Introduction to Modern Optics*". (Courier Corporation, 1975).

44. Goldstein, D. & Goldstein, D. H. "*Polarized Light, Revised and Expanded*". (CRC Press, 2003).

45. Menzel, C., Rockstuhl, C. & Lederer, F. "Advanced Jones calculus for the classification of periodic metamaterials". *Phys. Rev. A* **82,** 053811 (2010).

46. Heavens, O. S. "*Optical Properties of Thin Solid Films*". (Courier Corporation, 1991).

47. Maier, S. A. "*Plasmonics: Fundamentals and applications*". (Springer US, 2007).

48. Kretschmann, E. & Raether, H. "Radiative decay of non-radiative surface plasmons excited by light". *Z. Naturforsch.* **23,** 2135–2136 (1968).

49. Raether, H. "*Surface Plasmons on Smooth and Rough Surfaces and on Gratings*". **111,** (Springer Berlin Heidelberg, 1988).

50. Barnes, W. L., Dereux, A. & Ebbesen, T. W. "Surface plasmon subwavelength optics". *Nature* **424,** 824–830 (2003).

51. Bohren, C. F. & Huffman, D. M. "*Absorption And Scattering of Light By Small Particles*". (John Wiley & Sons, 2008).

52. "Poincare Sphere". at "***http://people.seas.harvard.edu/~jones/ap216/lectures/ls_1/ls1_u6/p_sphere_1.gif***" on 25 September 2015

53. Pancharatnam, S. "Generalized theory of interference, and its applications". *Proc. Indian Acad. Sci. - Sect. A* **44,** 247–262 (1956).

54. Berry, M. V. "The Adiabatic Phase and Pancharatnam's Phase for Polarized Light". *J. Mod. Opt.* **34,** 1401–1407 (1987).

55. Aharonov, Y. & Bohm, D. "Significance of Electromagnetic Potentials in the

Quantum Theory". *Phys. Rev.* **115,** 485–491 (1959).

56. Berry, M. "Pancharatnam, virtuoso of the Poincaré sphere: an appreciation". *Curr. Sci.* **67,** 220–223 (1994).

57. Aravind, P. K. "A simple proof of Pancharatnam's theorem". *Opt. Commun.* **94,** 191–196 (1992).

58. Eriksson, F. "On the measure of solid angles". *Math. Mag.* **63,** 184–187 (1990).

59. van Dijk, T., Schouten, H. F., Ubachs, W. & Visser, T. D. "The Pancharatnam-Berry phase for non-cyclic polarization changes". *Opt. Express* **18,** 10796–10804 (2010).

60. Marrucci, L., Manzo, C. & Paparo, D. "Pancharatnam-Berry phase optical elements for wave front shaping in the visible domain: Switchable helical mode generation". *Appl. Phys. Lett.* **88,** 221102. (2006).

61. Bomzon, Z., Biener, G., Kleiner, V. & Hasman, E. "Radially and azimuthally polarized beams generated by space-variant dielectric subwavelength gratings". *Opt. Lett.* **27,** 285–287 (2002).

62. Bomzon, Z., Kleiner, V. & Hasman, E. "Computer-generated space-variant polarization elements with subwavelength metal stripes". *Opt. Lett.* **26,** 33–35 (2001).

63. Niv, A., Biener, G., Kleiner, V. & Hasman, E. "Manipulation of the Pancharatnam phase in vectorial vortices". *Opt. Express* **14,** 4208–4220 (2006).

64. Biener, G., Niv, A., Kleiner, V. & Hasman, E. "Formation of helical beams by use of Pancharatnam-Berry phase optical elements". *Opt. Lett.* **27,** 1875–1877 (2002).

65. Huang, L., Chen, X., Mühlenbernd, H., Zhang, H., Chen, S., Bai, B., Tan, Q., Jin, G., Cheah, K.-W., Qiu, C.-W., Li, J., Zentgraf, T. & Zhang, S. "Three-dimensional optical holography using a plasmonic metasurface". *Nat. Commun.* **4,** 2808. (2013).

66. Chen, X., Zhang, Y., Huang, L. & Zhang, S. "Ultrathin Metasurface Laser Beam Shaper". *Adv. Opt. Mater.* **2,** 978–982 (2014).

67. Huang, L., Chen, X., Bai, B., Tan, Q., Jin, G., Zentgraf, T. & Zhang, S. "Helicity dependent directional surface plasmon polariton excitation using a metasurface with interfacial phase discontinuity". *Light Sci. Appl.* **2,** e70. (2013).

68. Huang, L., Chen, X., Mühlenbernd, H., Li, G., Bai, B., Tan, Q., Jin, G., Zentgraf, T. & Zhang, S. "Dispersionless phase discontinuities for controlling

light propagation". *Nano Lett.* **12,** 5750–5755 (2012).

69. Zheng, G., Mühlenbernd, H., Kenney, M., Li, G., Zentgraf, T. & Zhang, S. "Metasurface holograms reaching 80% efficiency". *Nat. Nanotechnol.* **10,** 308–312 (2015).

70. Wilk, S. R. "Claudius Ptolemy's Law of Refraction". *Opt. Photonics News* **14,** 14–17 (2004).

71. Cajori, F. "*A history of physics in its elementary branches: including the evolution of physical laboratories*". (Macmillan, 1922).

72. Docker, P. T., Kinnell, P. & Ward, M. C. L. "A dry single-step process for the manufacture of released MEMS structures". *J. Micromechanics Microengineering* **13,** 790–794 (2003).

73. Laermer, F. & Schilp, A. "Method of anisotropically etching silicon". *US Pat. 6284148* 7 (1996).

74. Blauw, M. A. "*PhD Thesis: Deep anisotropic dry etching of silicon microstructures by high-density plasmas*". (Delft University of Technology, 2004).

75. Cheneler, D. "*PhD Thesis: The Design and Analysis of a Micro Squeeze Flow Rheometer*". (University of Birmingham, 2009).

76. Abdolvand, R. & Ayazi, F. "An advanced reactive ion etching process for very high aspect-ratio sub-micron wide trenches in silicon". *Sensors Actuators A Phys.* **144,** 109–116 (2008).

77. Wu, B., Kumar, A. & Pamarthy, S. "High aspect ratio silicon etch: A review". *J. Appl. Phys.* **108,** 051101. (2010).

78. Chen, H. & Fu, C. "An investigation into the characteristics of deep reactive ion etching of quartz using SU-8 as a mask". *J. Micromechanics Microengineering* **18,** 105001. (2008).

79. Ayón, A. A. "Characterization of a Time Multiplexed Inductively Coupled Plasma Etcher". *J. Electrochem. Soc.* **146,** 339–349 (1999).

80. Pendry, J. B. B., Holden, A. J. J., Robbins, D. J. J. & Stewart, W. J. J. "Magnetism from conductors and enhanced nonlinear phenomena". *IEEE Trans. Microw. Theory Tech.* **47,** 2075–2084 (1999).

81. Tonouchi, M. "Cutting-edge terahertz technology". *Nat. Photonics* **1,** 97–105 (2007).

82. "MenloSystems-Discover the world of Terahertz". at "***http://www.menlosystems.com/assets/applications/terahertz/ME***

NLO-World-of-THz-B-EN-2014-07-16-3w.pdf on 23 September 2015

83. Zhang, X., Tian, Z., Yue, W., Gu, J., Zhang, S., Han, J. & Zhang, W. "Broadband terahertz wave deflection based on C-shape complex metamaterials with phase discontinuities". *Adv. Mater.* **25,** 4567–4572 (2013).

84. Lawrence, M. "*PhD Thesis: Symmetry and Topology at the Metasurface*". (University of Birmingham, 2015).

85. Pfeiffer, C., Zhang, C., Ray, V., Guo, L. J. & Grbic, A. "High Performance Bianisotropic Metasurfaces: Asymmetric Transmission of Light". *Phys. Rev. Lett.* **113,** 023902. (2014).

86. Wu, L., Yang, Z., Cheng, Y., Zhao, M., Gong, R., Zheng, Y., Duan, J. & Yuan, X. "Giant asymmetric transmission of circular polarization in layer-by-layer chiral metamaterials". *Appl. Phys. Lett.* **103,** 021903. (2013).

87. Grady, N. K., Heyes, J. E., Chowdhury, D. R., Zeng, Y., Reiten, M. T., Azad, A. K., Taylor, A. J., Dalvit, D. a R. & Chen, H.-T. "Terahertz metamaterials for linear polarization conversion and anomalous refraction". *Science* **340,** 1304–1307 (2013).

88. Sun, S., He, Q., Xiao, S., Xu, Q., Li, X. & Zhou, L. "Gradient-index meta-surfaces as a bridge linking propagating waves and surface waves". *Nat. Mater.* **11,** 426–431 (2012).

89. Sun, S., Yang, K., Wang, C.-M., Juan, T., Chen, W. T., Liao, C. Y., He, Q., Xiao, S., Kung, W.-T., Guo, G., Zhou, L. & Tsai, D. P. "High-efficiency broadband anomalous reflection by gradient meta-surfaces". *Nano Lett.* **12,** 6223–6229 (2012).

90. Ile, Z. L. I., Heng, G. U. Z., Ing, P. A. N. H. E., Ong, S. L. I., Eng, Q. I. D., Hao, J. I. Z., Ong, Y. A. I., Li, Z., Zheng, G., He, P., Li, S., Deng, Q., Zhao, J. & Ai, Y. "All-silicon nanorod-based Dammann gratings". *Opt. Lett.* **40,** 4285–4288 (2015).

91. Warrant, E. J. "Polarisation vision: beetles see circularly polarised light". *Curr. Biol.* **20,** R610–R612 (2010).

92. Sharma, V., Crne, M., Park, J. O. & Srinivasarao, M. "Structural origin of circularly polarized iridescence in jeweled beetles". *Science* **325,** 449–451 (2009).

93. Hecht, L. & Barron, L. D. "Rayleigh and Raman optical activity from chiral surfaces". *Chem. Phys. Lett.* **225,** 525–530 (1994).

94. Arnaut, L. R. & Davis, L. E. "Dispersion Characteristics of Planar Chiral Structures". in *ICEAA - Int. Conf. Electromagn. Adv. Applic. 1995* (1995). at

"**http://www.researchgate.net/publication/257242534_Dispersion_ Characteristics_of_Planar_Chiral_Structures**" on <u>24</u> <u>September</u> <u>2015</u>

95.   Fedotov, V. A., Mladyonov, P. L., Prosvirnin, S. L., Rogacheva, A. V., Chen, Y. & Zheludev, N. I. "Asymmetric Propagation of Electromagnetic Waves through a Planar Chiral Structure". *Phys. Rev. Lett.* **97,** 167401. (2006).

96.   Schwanecke, A. S., Fedotov, V. A., Khardikov, V. V., Prosvirnin, S. L., Chen, Y. & Zheludev, N. I. "Nanostructured Metal Film with Asymmetric Optical Transmission". *Nano Lett.* **8,** 2940–2943 (2008).

97.   Plum, E. "*PhD Thesis: Chirality and metamaterials*". (University of Southampton, 2010).

98.   Cao, T., Wei, C., Mao, L. & Li, Y. "Extrinsic 2D chirality: giant circular conversion dichroism from a metal-dielectric-metal square array". *Sci. Rep.* **4,** 7442. (2014).

99.   Fedotov, V. A., Schwanecke, A. S., Zheludev, N. I., Khardikov, V. V. & Prosvirnin, S. L. "Asymmetric Transmission of Light and Enantiomerically Sensitive Plasmon Resonance in Planar Chiral Nanostructures". *Nano Lett.* **7,** 1996–1999 (2007).

100. Plum, E., Fedotov, V. A. & Zheludev, N. I. "Planar metamaterial with transmission and reflection that depend on the direction of incidence". *Appl. Phys. Lett.* **94,** 131901. (2009).

101. Kang, M., Chen, J., Cui, H.-X., Li, Y. & Wang, H.-T. "Asymmetric transmission for linearly polarized electromagnetic radiation". *Opt. Express* **19,** 8347–8356 (2011).

102. Menzel, C., Helgert, C., Rockstuhl, C., Kley, E.-B., Tünnermann, A., Pertsch, T. & Lederer, F. "Asymmetric Transmission of Linearly Polarized Light at Optical Metamaterials". *Phys. Rev. Lett.* **104,** 253902. (2010).

103. Plum, E., Fedotov, V. A. & Zheludev, N. I. "Asymmetric transmission: a generic property of two-dimensional periodic patterns". *J. Opt.* **13,** 024006. (2011).

104. Mutlu, M., Akosman, A. E., Serebryannikov, A. E. & Ozbay, E. "Diodelike Asymmetric Transmission of Linearly Polarized Waves Using Magnetoelectric Coupling and Electromagnetic Wave Tunneling". *Phys. Rev. Lett.* **108,** 213905. (2012).

105. Huang, C., Feng, Y., Zhao, J., Wang, Z. & Jiang, T. "Asymmetric electromagnetic wave transmission of linear polarization via polarization conversion through chiral metamaterial structures". *Phys. Rev. B* **85,** 195131

(2012).

106. Xu, Y., Gu, C., Hou, B., Lai, Y., Li, J. & Chen, H. "Broadband asymmetric waveguiding of light without polarization limitations". *Nat. Commun.* **4,** 2561. (2013).

107. Li, Z., Mutlu, M. & Ozbay, E. "Chiral metamaterials: from optical activity and negative refractive index to asymmetric transmission". *J. Opt.* **15,** 023001. (2013).

108. Pan, C., Ren, M., Li, Q., Fan, S. & Xu, J. "Broadband asymmetric transmission of optical waves from spiral plasmonic metamaterials". *Appl. Phys. Lett.* **104,** 121112. (2014).

109. Huang, Y., Yao, Z., Wang, Q., Hu, F. & Xu, X. "Coupling Tai Chi Chiral Metamaterials with Strong Optical Activity in Terahertz Region". *Plasmonics* **10,** 1005–1011 (2015).

110. Ma, H. F., Wang, G. Z., Kong, G. S. & Cui, T. J. "Independent Controls of Differently-Polarized Reflected Waves by Anisotropic Metasurfaces". *Sci. Rep.* **5,** 9605. (2015).

111. Minovich, A. E., Miroshnichenko, A. E., Bykov, A. Y., Murzina, T. V., Neshev, D. N. & Kivshar, Y. S. "Functional and nonlinear optical metasurfaces". *Laser Photonics Rev.* **9,** 195–213 (2015).

112. Singh, R., Plum, E., Menzel, C., Rockstuhl, C., Azad, A. K., Cheville, R. A., Lederer, F., Zhang, W. & Zheludev, N. I. "Terahertz metamaterial with asymmetric transmission". *Phys. Rev. B* **80,** 153104. (2009).

113. Plum, E., Fedotov, V. A. & Zheludev, N. I. "Optical activity in extrinsically chiral metamaterial". *Appl. Phys. Lett.* **93,** 191911. (2008).

114. Ye, W., Yuan, X., Guo, C., Zhang, J., Yang, B., Liu, K., Zhu, Z. & Zeng, C. "Circular Polarizer Realized by a Single Layer of Planar Chiral Metallic Nanostructure". (2014). at "***http://arxiv.org/abs/1410.2656***"

115. Bomzon, Z., Kleiner, V. & Hasman, E. "Pancharatnam–Berry phase in space-variant polarization-state manipulations with subwavelength gratings". *Opt. Lett.* **26,** 1424–1426 (2001).

116. Plum, E. & Zheludev, N. I. "Chiral mirrors". *Appl. Phys. Lett.* **106,** 221901. (2015).

117. Lalanne, P. & Hugonin, J.-P. "High-order effective-medium theory of subwavelength gratings in classical mounting: application to volume holograms". *J. Opt. Soc. Am. A* **15,** 1843–1851 (1998).

118. Liu, L., Zhang, X., Kenney, M., Su, X., Xu, N., Ouyang, C., Shi, Y., Han, J., Zhang, W. & Zhang, S. "Broadband metasurfaces with simultaneous control of phase and amplitude". *Adv. Mater.* **26,** 5031–5036 (2014).

119. Ni, X., Emani, N. K., Kildishev, A. V., Boltasseva, A. & Shalaev, V. M. "Broadband light bending with plasmonic nanoantennas". *Science* **335,** 427. (2012).

120. Kang, M., Feng, T., Wang, H.-T. & Li, J. "Wave front engineering from an array of thin aperture antennas". *Opt. Express* **20,** 15882–15890 (2012).

121. Yin, X., Ye, Z., Rho, J., Wang, Y. & Zhang, X. "Photonic spin Hall effect at metasurfaces". *Science* **339,** 1405–1407 (2013).

122. Shitrit, N., Yulevich, I., Maguid, E., Ozeri, D., Veksler, D., Kleiner, V. & Hasman, E. "Spin-optical metamaterial route to spin-controlled photonics". *Science* **340,** 724–726 (2013).

123. Aieta, F., Genevet, P., Kats, M. a., Yu, N., Blanchard, R., Gaburro, Z. & Capasso, F. "Aberration-free ultrathin flat lenses and axicons at telecom wavelengths based on plasmonic metasurfaces". *Nano Lett.* **12,** 4932–4936 (2012).

124. Ni, X., Ishii, S., Kildishev, A. V & Shalaev, V. M. "Ultra-thin, planar, Babinet-inverted plasmonic metalenses". *Light Sci. Appl.* **2,** e72. (2013).

125. Lin, J., Mueller, J. P. B., Wang, Q., Yuan, G., Antoniou, N., Yuan, X.-C. & Capasso, F. "Polarization-controlled tunable directional coupling of surface plasmon polaritons". *Science* **340,** 331–334 (2013).

126. Zhou, F., Liu, Y. & Cai, W. "Plasmonic holographic imaging with V-shaped nanoantenna array". *Opt. Express* **21,** 4348–4354 (2013).

127. Chen, W. T., Yang, K. Y., Wang, C. M., Huang, Y. W., Sun, G., Chiang, I. Da, Liao, C. Y., Hsu, W. L., Lin, H. T., Sun, S., Zhou, L., Liu, A. Q. & Tsai, D. P. "High-efficiency broadband meta-hologram with polarization-controlled dual images". *Nano Lett.* **14,** 225–230 (2014).

128. Ni, X., Kildishev, A. V. & Shalaev, V. M. "Metasurface holograms for visible light". *Nat. Commun.* **4,** 2807. (2013).

129. Lin, J., Genevet, P., Kats, M. A., Antoniou, N. & Capasso, F. "Nanostructured holograms for broadband manipulation of vector beams". *Nano Lett.* **13,** 4269–4274 (2013).

130. Farmahini-Farahani, M., Cheng, J. & Mosallaei, H. "Metasurfaces nanoantennas for light processing". *J. Opt. Soc. Am. B* **30,** 2365–2370 (2013).

131. Yu, N., Aieta, F., Genevet, P., Kats, M. A., Gaburro, Z. & Capasso, F. "A broadband, background-free quarter-wave plate based on plasmonic metasurfaces". *Nano Lett.* **12,** 6328–6333 (2012).

132. Genevet, P., Yu, N., Aieta, F., Lin, J., Kats, M. A., Blanchard, R., Scully, M. O., Gaburro, Z. & Capasso, F. "Ultra-thin plasmonic optical vortex plate based on phase discontinuities". *Appl. Phys. Lett.* **100,** 013101. (2012).

133. Kats, M. a., Genevet, P., Aoust, G., Yu, N., Blanchard, R., Aieta, F., Gaburro, Z. & Capasso, F. "Giant birefringence in optical antenna arrays with widely tailorable optical anisotropy". *Proc. Natl. Acad. Sci.* **109,** 12364–12368 (2012).

134. Monticone, F., Estakhri, N. M. & Alù, A. "Full Control of Nanoscale Optical Transmission with a Composite Metascreen". *Phys. Rev. Lett.* **110,** 203903. (2013).

135. Pfeiffer, C. & Grbic, A. "Metamaterial Huygens' Surfaces: Tailoring Wave Fronts with Reflectionless Sheets". *Phys. Rev. Lett.* **110,** 197401. (2013).

136. Yu, N. & Capasso, F. "Flat optics with designer metasurfaces". *Nat. Mater.* **13,** 139–150 (2014).

137. Yang, Y., Wang, W., Moitra, P., Kravchenko, I. I., Briggs, D. P. & Valentine, J. "Dielectric meta-reflectarray for broadband linear polarization conversion and optical vortex generation". *Nano Lett.* **14,** 1394–1399 (2014).

138. Yang, B., Ye, W.-M., Yuan, X.-D., Zhu, Z.-H. & Zeng, C. "Design of ultrathin plasmonic quarter-wave plate based on period coupling". *Opt. Lett.* **38,** 679–681 (2013).

139. Li, G., Kang, M., Chen, S., Zhang, S., Pun, E. Y. B., Cheah, K. W. & Li, J. "Spin-enabled plasmonic metasurfaces for manipulating orbital angular momentum of light". *Nano Lett.* **13,** 4148–4151 (2013).

140. Dahan, N., Gorodetski, Y., Frischwasser, K., Kleiner, V. & Hasman, E. "Geometric Doppler effect: spin-split dispersion of thermal radiation". *Phys. Rev. Lett.* **105,** 136402. (2010).

141. Larouche, S., Tsai, Y. J., Tyler, T., Jokerst, N. M. & Smith, D. R. "Infrared metamaterial phase holograms". *Nat. Mater.* **11,** 450–454 (2012).

142. Yifat, Y. "Highly efficient and broadband wide-angle holography using patch-dipole nanoantenna reflectarrays". *Nano Lett.* **14,** 2485–2490 (2014).

143. Freese, W., Kämpfe, T., Kley, E.-B. B. & Tünnermann, A. "Design of binary subwavelength multiphase level computer generated holograms". *Opt. Lett.* **35,** 676–678 (2010).

144. Gori, F. "Measuring Stokes parameters by means of a polarization grating". *Opt. Lett.* **24,** 584–586 (1999).

145. Hao, J., Yuan, Y., Ran, L., Jiang, T., Kong, J. A., Chan, C. T. & Zhou, L. "Manipulating electromagnetic wave polarizations by anisotropic metamaterials". *Phys. Rev. Lett.* **99,** 063908. (2007).

146. Pors, A., Nielsen, M. G. & Bozhevolnyi, S. I. "Broadband plasmonic half-wave plates in reflection". *Opt. Lett.* **38,** 513–515 (2013).

147. Jiang, S.-C., Xiong, X., Hu, Y.-S., Hu, Y.-H., Ma, G.-B., Peng, R.-W., Sun, C. & Wang, M. "Controlling the Polarization State of Light with a Dispersion-Free Metastructure". *Phys. Rev. X* **4,** 021026. (2014).

148. Shen, F. & Wang, A. "Fast-Fourier-transform based numerical integration method for the Rayleigh–Sommerfeld diffraction formula". *Appl. Opt.* **45,** 1102–1110 (2006).

149. Dammann, H. & Görtler, K. "High-efficiency in-line multiple imaging by means of multiple phase holograms". *Opt. Commun.* **3,** 312–315 (1971).

150. Gerchberg, R. W. & Saxton, W. O. "A practical algorithm for the determination of phase from image and diffraction plane pictures". *Optik (Stuttg).* **35,** 237–246 (1972).

151. Toal, V. "*Introduction to Holography*". (CRC Press, 2011).

152. Hasman, E., Davidson, N. & Friesem, A. A. "Efficient multilevel phase holograms for CO2 lasers". *Opt. Lett.* **16,** 423–425 (1991).

153. Liu, N., Mesch, M., Weiss, T., Hentschel, M. & Giessen, H. "Infrared perfect absorber and its application as plasmonic sensor". *Nano Lett.* **10,** 2342–2348 (2010).

154. Feldman, M. R. & Guest, C. C. "Iterative encoding of high-efficiency holograms for generation of spot arrays". *Opt. Lett.* **14,** 479–481 (1989).

155. Kress, B. C. & Meyrueis, P. "*Applied Digital Optics: From Micro-optics to Nanophotonics*". (John Wiley & Sons, 2009).