Shaping light beams with dielectric metasurfaces

Daniël Pieter Stellinga

Doctor of Philosophy

University of York
Physics

May 2016

Abstract.

With the advent of techniques such as Stimulated Emission Depletion Microscopy (STED) and light sheet microscopy, the generation of specialised light beams has become an exciting field. However, while very effective methods of generating such beams exist, the components necessary to do so are generally large and cumbersome. Metasurfaces promise the replacement of these traditional bulky optical elements with sub-wavelength thick and flat alternatives, paving the way for integration into microscale form factors.

Metasurfaces commonly use a distribution of nanoscale resonant elements to engineer a phase plate that shapes light through the Huygens' principle, allowing them to mimic and improve upon traditional optics. Initially, plasmonic resonant elements were explored by the community, but their dissipative losses have severely limited the efficiency of these devices.

Here I discuss my work on the development of dielectric sub-wavelength grating based metasurfaces. Four types of metasurface, each using a different manifestation of grating physics are explored: direct phase, polarisation conversion, geometric phase, and active metasurfaces. I show that these different types of metasurface together allow the shaping of a wide variety of beams under a large range of different conditions, while retaining efficiencies on the order of $80-90\%$. Examples of the beam shapes explored include focused beams, vortex beams, Bessel beams and cylindrical vector beams.

The development of high efficiency dielectric metasurfaces brings ultrathin optics closer to practical applications. Their materials and sizes facilitate the integration into previously unavailable form factors, including applications in microfluidics.

Contents

List of Figures

List of Tables

Acknowledgements.

No work worth doing can be accomplished by a person on their own, and the work I present here is no different. My sincerest gratitude goes out to all those who have aided me along the way, as this thesis would look very different without their contributions.

First and foremost, my supervisor Prof. Thomas F. Krauss, for providing me with the environment needed to pursue projects I am passionate about. Not just the physical environment, but more importantly the atmosphere he cultivates among the research group has been enjoyable and stimulating. Without this atmosphere and his patient guidance over the years I would have stranded many times over.

I would also like to thank the other members of the Photonics research group in York, both past and present. Every one of them has been instrumental in getting me past hurdles encountered along the way, and all of them have become good friends in the process: José Juan Colas, Donato Conteduca, Dr. Matthias Fischer, Giampaolo Pitruzello, Dr. Christopher Reardon, Lewis Reeves, Dr. Bryan O'Regan, Andrei Ruskuc, Dr. Christian Schuster, Dr. Mark Scullion, Dr. Yue Wang, and Davide Zecca.

Particular thanks go out to Dr Annett Fischer and Graham Triggs. Without Annett's early work on the fabrication and testing of the devices I designed this project would almost certainly have stranded before it began. Her expertise in the cleanroom and her ability to deal with my often chaotic explanations were very much appreciated. Graham and I have gone through our PhDs in parallel, and the entire way he has been the focus of my ramblings. Any problem I struggled with or random ideas that popped into my head, he would patiently listen and help me get my thoughts in order. For this I am extremely grateful.

During the course of my PhD I collaborated with a large number of people. Not all of this work came to fruition or has made it into this thesis, but all of them were a significant part in my academic growth.

At the university of St Andrews, I worked in close collaboration with the research groups of Prof. Kishan Dholakia and Prof. Ifor Samuel on a variety of projects. In particular I would like to thank James Glackin, Dr. Monika Pietrzyk, Dr. Mingzhou Chen, and Dr. Tom Vettenburg, for their continuing perseverance in helping test ideas that I can't test myself.

At the Sun Yat-sen University in Guangzhou, my thanks goes to Dr. Juntao

Li, Dr. Wang Jiahui and Quanquan Zhang, for being the most enthusiastic scientists I have had the pleasure to visit and work with.

Finally, at CUDOS in the university of Sydney, I would like to thank Neetesh Singh, Dr. Alvaro Casas, Dr. Darren Hudson, Tomonori Hu, and Fernando Diaz, for making me feel at home and helping to broaden my horizons in every sense of the word.

The support of my family through regular Skype calls and occasional flights was also genuinely appreciated. Their faith in me throughout the ten years I have now been at different universities has been amazing.

Last but far from least, my fiancée, Annick Meijboom. She has been the person who has most closely shared in the ups and downs, the good days and the bad, that have made up the last three and a half years. I can't properly express how grateful I am to have had her unwavering support at my side every step of the way.

Declaration of Authorship.

I, Daniël Pieter Stellinga, declare that this thesis titled, Shaping light beams with dielectric metasurfaces, and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Part of the work presented in chapter 4 of this thesis was previously presented in the PhD thesis of Dr. Annett B. Fischer, titled "Planar optics with wavelength-scale high contrast gratings", in March 2015 at the University of York.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

The following lists provide details of publications that have had contributions from the work undertaken for this thesis:

Journal articles.

- A. B. Klemm, **D. Stellinga**, E. R. Martins, L. Lewis, G. Huyet, L. OFaolain, T. F. Krauss. Experimental high numerical aperture focusing with high contrast gratings. *Optics letters*, 38(17), 3410-3413, 2013
- A. B. Klemm, **D. Stellinga** (shared first author), E. R. Martins, L. Lewis, L. O'Faolain, T. F. Krauss, *Optical Engineering*, 53(9), 095104, 2014
- Y. Wang, **D. Stellinga**, A. B. Klemm, C. P. Reardon, T. F. Krauss, *IEEE J Sel Top Quant*, 21(4), 2700706, 2015
- G. J. Triggs, M. Fischer, **D. Stellinga**, M. G. Scullion, G. J. O. Evans, T. F. Krauss, *IEEE Photonics Journal*, 7(3), 2015

Conference attendance.

- **D. Stellinga**, A. B. Klemm, E. R. Martins, L. Lewis, G. Huyet, L. OFaolain, and T. F. Krauss. 2D High contrast gratings as high numerical aperture focussing elements (poster presentation, 3rd place poster prize). Complex Nanophotonics Science Camp 2013, Windsor, England, August 2013.

- **D. Stellinga**, A. B. Klemm, L. Lewis, K. Dholakia, and T. F. Krauss. Planar Optics using 2D High Contrast Gratings (poster presentation). Photonic and Electromagnetic Crystal Structures (PECS XI), Fudan, China, April 2014.

- **D. Stellinga**, A. B. Fischer, Y. Wang, and T. F Krauss. Beam shaping using spatially varying subwavelength gratings (oral presentation). CLEO Europe, Munich, Germany, June 2015.

- **D. Stellinga**, A. B. Fischer, Y. Wang, and T. F. Krauss. Phase and polarisation control with high contrast grating metasurfaces (poster presentation). Frontiers in Nanophotonics 2015, Monte Verita, Switzerland, August 2015.

1 Introduction.

Light is constantly all around us. It is what allows us to see, receive energy from the sun and transmit data, to name but a few. But when light is controlled, shaped and directed into beams, things get really interesting. The harmless light we live with every moment of every day can then suddenly be used to cut through metals; to see things so small that we didn't even know they existed until recently[1, 2]; or even to grab hold of particles and move them around[3, 4].

Such feats have paved the way for our modern society, and without the contributions of light beams the world would look very different today. But only in the last few decades has the potential of some really interesting beams been unlocked. These beams range from vortex beams that can rotate trapped particles[5, 6] or enable microscopy down to the nanometre scale[1, 2], to the use of Bessel beams in light sheet microscopy[7, 8] and optical injections[9].

These beams have only started making headway since the means of creating them have become more accessible. Precision optics or diffractive elements, or the more versatile spatial light modulators, have allowed many of these beams to become much more common in research today. But all of these methods require large and costly beam conversion elements, severely limiting their integration into every day devices.

Metasurfaces promise one avenue to reduce both the footprint and the cost of beam shaping[10–13]. Relying on localised optical resonances, these structures manipulate light on scales close to the wavelength itself, allowing us to shape passing light into almost any shape we desire, all in a package that can be shrunk to less than the diameter of a human hair. The first metasurfaces were based on plasmonic resonances that confine light down to extremely small sizes[10, 14–16].

Of course, there is a downside to metasurfaces. Their use of highly localised plasmonic resonances relies on metallic nanostructures. These structures are inherently lossy, as they concentrate light into resonances on the surface of the highly absorbing metals. Initially, the efficiencies of most plasmonic metasurfaces were very low, typically in the range of 10% or less. For metasurfaces operating in reflection this has recently been increased to the $80 - 90\%$ range[17], but in transmission even the best ones do not exceed 30%[11]. This is still a long way from the almost negligible losses that can be achieved with high precision optics at the macro scale.

The primary aim of the work in this thesis is to offer an alternative: dielectric metasurfaces[18–22]. In contrast to metals, dielectrics are far more transparent in large ranges of the electromagnetic spectrum, significantly reducing absorption as

a loss channel. Their resonances are also fundamentally less concentrated, however, and creating a spatially varying resonant structure is not without challenges.

The field of metasurfaces is largely built on the older concept of transmit-arrays and reflect-arrays as used in the microwave community[15, 23–25]. These arrays rely on the coupling of microwaves to circulating currents in the conductive antennas they were built from, and adjusted the antenna geometries locally to shape the microwave output. Since the advent of mainstream access to microfabrication in the research community, a lot of this work has been brought down in scale to sizes were the same feats can be replicated with infrared and even visible light. At this new scale they are called metasurfaces.

As mentioned, metasurfaces were initially based on plasmonic resonant effects. Surface plasmon polaritons can act as the circulating currents that were used in their microwave counterparts, allowing an almost one-to-one correlation between the two fields to be developed[14]. These metallic nanostructures still rely on a variety of possible effects to achieve their goal, leading to a wide variety of antennas geometries having been used. From early work that simply tuned the length of rectangular metal bars to achieve a low efficiency π phase range, to later more complicated designs consisting of a layer of v-shaped antennas coupled to a mirror image of themselves that can achieve the full 2π phase range and efficiencies up to 80% in reflection.

In transmission, however, plasmonic metasurfaces are far more limited in their efficiencies. Largely due to the dissipative losses owing to their metallic nature, the best efficiencies shown in the literature are around 30%[10, 11, 26]. Other features that decrease their efficiency are imperfect polarisation conversion leading to the requirement of cross polarisers to achieve properly converted beams, or the use of higher diffraction orders for the actual beam shaping that make the lower diffraction orders loss channels. The combination of these effects leads to most work on metasurfaces not coming close to their peak efficiencies, and efficiencies in the range of 5% are far more common[27].

A lot of these limitations stem from their metallic nature. An obvious step to try is therefore to move to dielectrics[18–21, 28]. However, dielectrics can not support surface plasmon polaritons, and therefore different resonant effects were necessary than were used in the original reflect-array work in the microwave regime. On top of that, the refractive indices, or permittivities, of dielectrics are far lower than those of metals, and therefore the light can only be squeezed down to sizes several times smaller than the wavelength, in contrast to the size of plasmonic resonators that are sometimes down to $\sim \lambda/100$[10]. The latter is mostly a concern in that it can limit the accuracy with which beams can be

reproduced. For example, lenses based on plasmonic metasurfaces can achieve numerical apertures, a dimensionless measure of focusing power that will be discussed in section 2.1, of ~ 0.8[29], while dielectric metasurfaces have so far been limited to ~ 0.6[30].

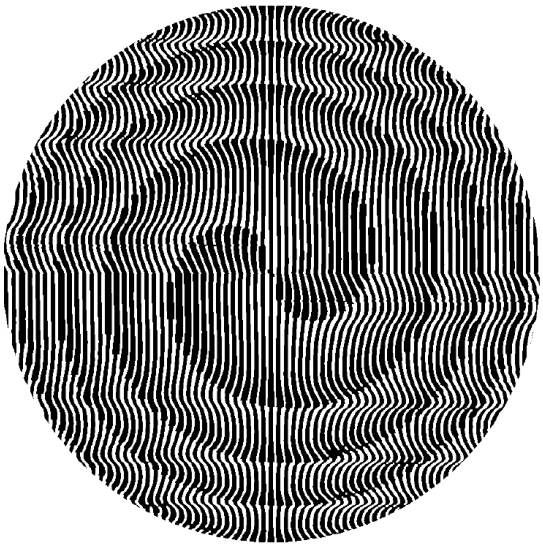
These two challenges have more recently been addressed by the dielectric photonics community. Pioneering steps in this direction were based on so-called high contrast gratings, periodic dielectric structures with a refractive index much higher than the surrounding media[28, 31, 32]. These high contrast gratings effectively fulfil the role of diffractive coupling element and waveguide simultaneously, and exploit that to give access to a class of resonances dubbed Guide Mode Resonances (GMRs)[33–36]. These GMRs were initially used for purposes such as high Q resonances, large area resonances, and very efficient broadband mirrors[36–40] but in 2010 several groups almost simultaneously showed how they could be used to locally manipulate the phase of reflected light through the full 2π phase range to create lenses and beam deflectors[21, 35], thereby opening the way to their use as dielectric metasurfaces.

Since then, several other methodologies of creating dielectric metasurfaces have been explored. Some of the more promising are arrays of dielectric disk resonators that use Mie resonances and can achieve full 2π phase coverage with transmission efficiencies of up to 55%[19]. Another type uses the different interactions of polarisations of incident light across and along silicon nano-wires to create a $\frac{\lambda}{2}$-plate effect and exploit a geometric phase in a similar approach to many plasmonic metasurfaces, though the use of silicon at visible wavelengths did lead to high absorption losses on the order of 30%[20].

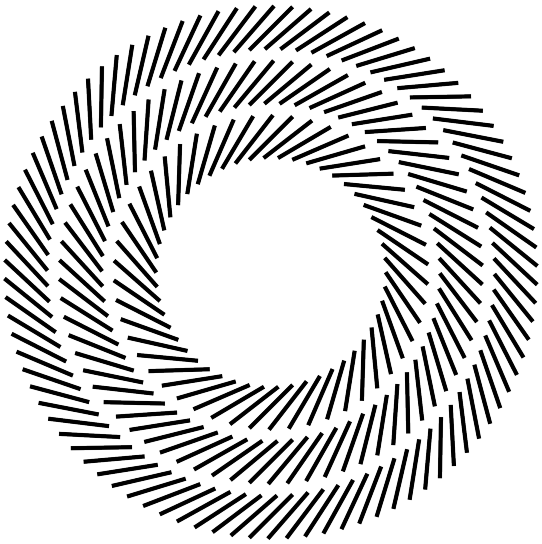
While all these approaches to dielectric metasurfaces are possible, my work focuses on those based on high contrast gratings. The inherent periodic nature of gratings introduces resonance types not seen in isolated structures. For periods smaller than the wavelength of the operating light the interaction between these grating resonances and a passing beam can be made to be 100% efficient, both in reflection and transmission[28]. This leads to metasurfaces potentially as efficient as traditional optics.

In the following chapters I will show how gratings can be used to design not just one, but four different types of metasurface, each with their own strengths and weaknesses, but relying on the same fundamental grating physics. An example design for each of these four types is shown in figure 1.1.

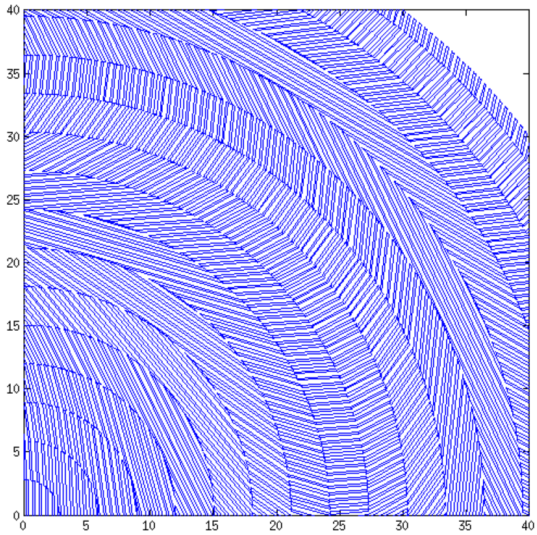
Chapter 2 is the foundation on which the rest of the work is built. In it, I explore the physics and mathematical techniques associated with light beams and their propagation, starting from their full vectorial descriptions and ending



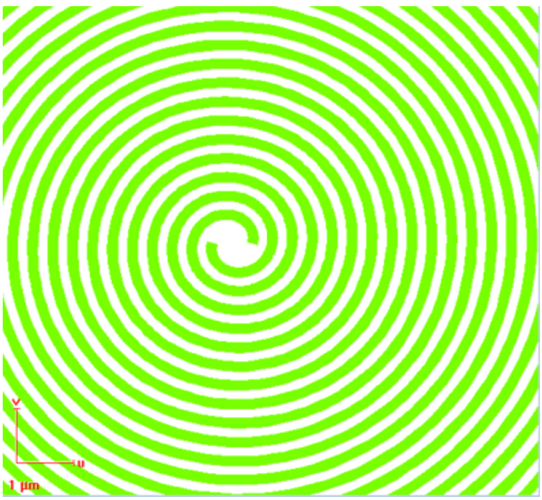
(a) Direct phase metasurface example.



(b) Polarisation conversion metasurface example.



(c) Geometric phase metasurface example.



(d) Active metasurface example.

Figure 1.1 – Example designs for the four types metasurface discussed in this thesis. (a) shows a design for a direct phase metasurface as discussed in chapter 4 that converts an incident plane wave into a focused vortex beam with azimuthal mode index $l = 2$, by varying the local period and dutycycle of the grating to shape the phase front. (b) shows a schematic design for a polarisation conversion metasurface (see chapter 5), consisting of gratings that act as $\frac{\lambda}{4}$-plates arranged around a ring at 45 deg to the radial, thereby converting an incident circularly polarised beam into a radial or azimuthally polarised version. (c) shows the design for a geometric phase metasurface that varies the orientation of $\frac{\lambda}{2}$-plate gratings to create a lens, also discussed in chapter 5. (d) shows an example of an active metasurface design that lets a DFB laser emit directly into an azimuthally polarised, $l = 2$ Laguerre-Gaussian beam, which is treated in chapter 6.

with more intuitive scalar approximations. I then describe the nature of the most important beams for this work, namely Gaussian beams and Laguerre-Gaussian, or vortex beams. This leads into how metasurfaces manage to shape such beams, and what sets metasurfaces apart from traditional optics.

The chapter concludes with an in depth study of sub-wavelength scale gratings and their associated resonances. I develop the theory of Guided Mode Resonances (GMRs) from the fundamental Bloch modes of the gratings, and discuss their applications to the design of metasurfaces in particular.

From there, I use chapter 3 to explain the techniques and methods I employed to perform my research and obtain my results. These fall into three clearly separated categories: numerical modelling, fabrication, and characterisation. The modelling section introduces Rigorous Coupled Wave Analysis (RCWA), a quasi-analytical technique that is both accurate and efficient at calculating the optical response of an arbitrary periodic structure. RCWA is the first step towards the design of every structure in this thesis, so is of paramount importance to my work. Also in this section I give a short overview of the commercial software COMSOL multiphysics, used to simulate how the fully designed metasurfaces function under a variety of conditions that would be challenging to perform experimentally.

The fabrication section shortly outlines the cleanroom processes used to create grating based metasurfaces. It gives an overview of the workings of electron beam lithography, reactive ion etching, and all the steps in between that are required to get to the finished product. In the characterisation section I then give a brief description of the two primary techniques for studying the fabricated structures, namely scanning electron microscopy and optical characterisation.

In chapter 4 I then show the first of the four types of metasurfaces I will discuss: direct phase metasurfaces. These grating based structures work by adiabatically varying the grating parameters of period and duty cycle in the plane, preserving the validity of grating thickness but altering the phase response of the excited resonances. Using this approach I show how we designed a variety of different metasurfaces including lenses, vortex beam generating phase plates, and flat analogues to axicons.

Direct phase metasurfaces were subjected to extensive tolerance tests in simulations, including a study of their spectral and angular bandwidths. In parallel, with the aid of Dr Annett Fischer, they were also fabricated and experimentally validated, and has been previously published[30, 41].

The subsequent chapter, 5, introduces a different way of manipulating waves: resonant form birefringence. In it I show that it is possible to tune the phase delay between resonant modes excited along and across the grating ridges while retaining their separate transmission efficiencies. This leads to a class of gratings

that, for all intents and purposes, are perfect $\frac{\lambda}{4}$- and $\frac{\lambda}{2}$-plates. I then show a study of such polarisation converting gratings performed using COMSOL, to show their ideal functionality and their spectral response.

These polarisation converting gratings are then shown to be useful for the design of two types of metasurface: polarisation conversion metasurfaces and geometric phase metasurfaces. The former type directly uses the polarisation conversion qualities of the gratings to spatially alter the polarisation state of a passing beam, turning an incident scalar beam into an output vector beam. Geometric phase metasurfaces on the other hand make use of the so-called Pancharatnam-Berry phase that arises from the relative orientation between waveplates. This effect gives a direct relation between the angle of the gratings in the plane of the metasurface and the output phase of a passing beam, thereby providing an alternative way of shaping the wavefront. While both these types of metasurface are conceptually explored and design methodologies developed, I have so far only experimentally verified the polarisation converting metasurfaces.

In chapter 6 I talk about the fourth and last type of metasurface I have explored. This last type of metasurface is not a metasurface in the traditional sense, in that it does not convert one beam type into another. Instead, active metasurfaces are feedback elements of a thin film laser (here an organic semiconductor laser) that allow direct shaping of the emitted beam. This concept relies on the tailoring of the spatial composition of the gratings that serve as the resonant feedback elements for organic gain-material based distributed feedback lasers. I show some preliminary results on this, as well as on a related concept that allows for spatial control of the spectral qualities of the emitted lasers.

Together the four metasurface types discussed in chapters 4, 5, and 6, provide a toolbox for efficient beam shaping under a variety of circumstances. In contrast to established, plasmonic metasurfaces the potential efficiencies in both transmission and reflection are far higher, due to absorption losses being minimised while retaining a full 2π phase coverage. The direct phase metasurfaces discussed in chapter 4 in particular are experimentally shown to achieve greater than 80% efficiency, and the geometric phase metasurfaces discussed in chapter 5 promise even higher efficiencies in the visible wavelength range.

2 Theory.

The title of this thesis is "Shaping light beams with dielectric metasurfaces." In this chapter I explain what this really means and how it works, and place it in the context of existing work in the field. I start with a very fundamental description of the relevant aspects of light beams and the applications of both traditional and novel beam types. The second section describes the concepts that govern the tools which create these beams, from traditional optics and spatial light modulators to modern metasurfaces, as well as describing the strenghts and weaknesses of these existing technologies. The chapter ends with the physics that lies at the base of the specific type of metasurface discussed in this thesis: sub-wavelength gratings.

By using this top-down approach, I hope to provide an understanding of the fundamentals on which the later chapters are built and paint a clear picture of the motivation for this research and how it fits in the field.

2.1 Beams.

Merriam-Webster defines a light beam as "a line of light coming from a source."[42] The Oxford English Dictionary refers to "a directional flow of particles or radiation."[43] Personally, I think of a beam as a light field that has a common optical axis along which it propagates. To understand the complicated forms such beams can take, a basic understanding of what light *is*, and how it propagates, is necessary. To this end, it is generally more insightful to start from a fully vectorial description; especially for metasurfaces the vectorial description is absolutely necessary and the scalar approach breaks down.

2.1.1 Fundamental vectorial equations.

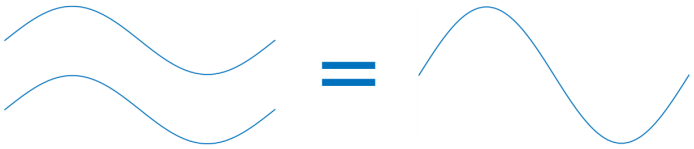
Light is a propagating excitation of the electromagnetic field[44–46]. When an electric field changes with time, it induces a magnetic field, and vice versa. The standard solution to this problem is a wave that travels at the speed of light with oscillating electric and magnetic fields. Mathematically, this arises straightforwardly from Maxwell's equations of electrodynamics[47]:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \tag{2.1}$$

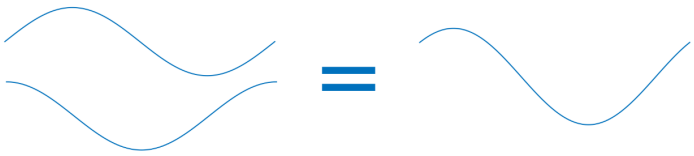
$$\nabla \cdot \mathbf{B} = 0, \tag{2.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{2.3}$$

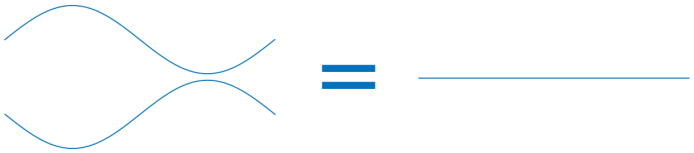
$$\nabla \times \mathbf{B} = \mu_0(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}). \tag{2.4}$$



(a) Constructive



(b) Partial



(c) Destructive

Figure 2.1 – Diagram to illustrate interference. If two waveforms coincide, their respective phases determines the resultant fields. a) Equal phase leads to constructive interference and an increased amplitude. b) Partially out of phase leads to a change in both amplitude an peak position. c) π out of phase leads to a complete cancellation of the fields.

Here $\mathbf{E}$ is the electric field, $\mathbf{B}$ the magnetic field, ϵ_0 and μ_0 the permittivity and permeability of free space, ρ and $\mathbf{J}$ the charge and current densities. Setting the charge and current densities to zero, as they are in vacuum, leaves the divergence equations both equal zero, while the curl equations equal the time-derivatives of the other field. These can then be easily rewritten to

$$\mu_0\epsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2} - \nabla^2\mathbf{E} = 0, \tag{2.5}$$

$$\mu_0\epsilon_0\frac{\partial^2\mathbf{B}}{\partial t^2} - \nabla^2\mathbf{B} = 0, \tag{2.6}$$

which are well known wave equations. (Note that $\mu_0\epsilon_0 = \frac{1}{c^2}$, with c the speed of light in a vacuum.) These two equations govern everything in this thesis, and they lead to a description of light in the form of purely a wave, which is almost exact on the length scales used here. The general form is $\mathbf{E} = \mathbf{E_0}\cdot e^{j(\mathbf{k}\cdot\mathbf{r}+\phi-\omega t)}$ (and equivalent for $\mathbf{B}$), with the amplitudes in $\mathbf{E_0}$ always perpendicular to both those of the magnetic field and the k-vector, or wavevector. $\mathbf{k}\cdot\mathbf{r}$ together describe the phase evolution of the wave as it travels through space, and ϕ is the instantaneous phase at the origin. As I am interested only in the steady state solutions, the temporal part ωt is set to zero in the rest of this work.

Moreover, the full solution to the wave equations is the sum, or superposition, of arbitrarily many of these so called plane waves. Where for a single plane wave the phase ϕ is essentially arbitrary, when dealing with multiples they become crucial to the shape of the total field. For example, as is shown schematically in figure 2.1, for two co-propagating plane waves with the same k-vector and amplitudes, $\mathbf{E_1}$ and $\mathbf{E_2}$, if both waves have the same phase (so the same value of ϕ) their amplitudes simply add up. However, if the phase differs by π, say with $\phi_1 = 0$ and $\phi_2 = \pi$, then we get $\mathbf{E_1}e^{j\times 0} = \mathbf{E_1} \times 1$ and $\mathbf{E_2}e^{j\times\pi} = \mathbf{E_2} \times -1$. In other words, the sum vanishes and we are left with no fields at all. This is the principle of interference which lies at the core of all beams.

A second consequence of the Maxwell equations, when combined with the conservation of energy, is the Poynting theorem[48]:

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{S} - \mathbf{J_f} \cdot \mathbf{E}, \tag{2.7}$$

where u is the energy density of the fields, $\mathbf{S}$ is the Poynting vector and J_f is the free current density. The energy density can also be described by

$$u = \epsilon_0 \mathbf{E} \cdot \mathbf{E} + \frac{1}{\mu_0}\mathbf{B} \cdot \mathbf{B}, \tag{2.8}$$

where I assume free space and ignore the polarisation field $\mathbf{P}$ and magnetisation field M for convenience. Combining these two and the Maxwell equations, it is straightforward to show[44] that the Poynting vector takes the form

$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B}, \tag{2.9}$$

which is a measure for the instantaneous power flow in an electromagnetic field, giving both the direction and the amplitude. Usually, the time averaged version is used, which can be shown to be $< \mathbf{S} >= \frac{1}{2}Re(\mathbf{S})$. The Poynting vector is therefore an immediate measure of how energy flows in a beam.

2.1.2 Scalar approximations.

Before giving an accurate mathematical model for beam propagation built on the wave equations given above, it is informative to look at a more intuitive and often sufficient model: the Huygens-Fresnel principle. This idea was first conceived of by Christiaan Huygens in 1678 and developed further by Augustin-Jean Fresnel in 1816, both long before James Clerk Maxwell first wrote down his equations, and as such lacks some of the mathematical rigour[49, 50]. While

it makes some assumptions and seemingly arbitrary choices, it still accurately predicts almost everything needed to understand beams and how they develop upon propagation; it is also very intuitive.

The premise of the Huygens-Fresnel principle is to consider each point in a beam as a spherical point source[45]. The amplitude and phase of the point source are given by its position in space and time in the beam. To find the fields at another position, it is then sufficient to sum up the contributions of these secondary point sources to that new position, using the principle of interference described before. This is especially easy in the case of free space, as each point source individually can be described as

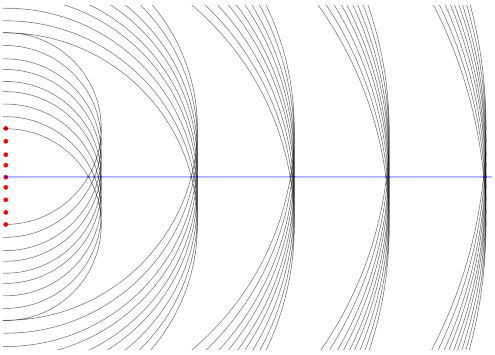
$$E(r) = E_0 \frac{e^{j(kr+\phi_0)}}{r}, \tag{2.10}$$

with r the distance from the point source, k the corresponding wave-vector, E_0 the amplitude at $r = 0$ and ϕ_0 the same for the phase. The phase, amplitude and distance all depend on the position in the beam. Note that this is a scalar formula, not vectorial as Maxwell's equations, so some information on the light is lost in the process.

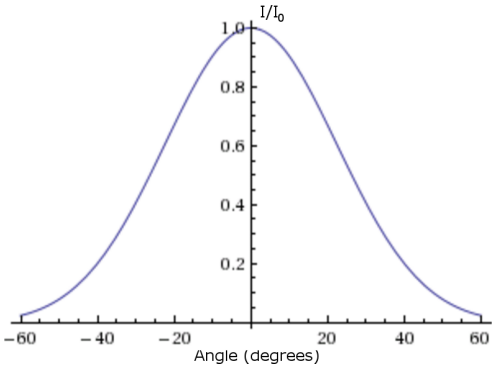
To make this method work, one needs to know the amplitude, phase and positions of a sufficient number of these point sources. To get exact results, one needs to know all of these on a closed surface around the target point, though usually we assume they are zero everywhere except in a region of interest, such as a beam cross section or wavefront[45, 51].

To get correct values of the fields out of this process, Huygens and later Fresnel had to make further assumptions, such as that a wave mainly propagates forward so the individual sources are not quite isotropic spherical sources but have a directional factor asssociated to them. However, even without these, the Huygens-Fresnel principle already allows one to qualitatively describe the propagation of a beam. For example, if a simple square wave with flat phase profile is given as a wavefront, so $E(x, z = 0) = \Pi(x)$, equivalent to the field resulting from a plane wave impinging on an aperture, it is immediately obvious from the Huygens-Fresnel principle that as it propagates in the z direction, the beam will spread and smoothen, developing diffraction side lobes. This is shown schematically in figure 2.2, by drawing lines of equal phase radiating from a distribution of point sources to illustrate the interference as a result of Huygens-Fresnel description of a wavefront.

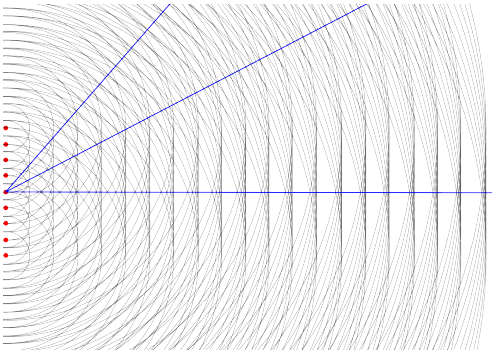
Other features that can be seen immediately include, for example, that an input Gaussian beam with a flat phase front will diverge but stay Gaussian, while a Gaussian beam with a curved phase front will converge or diverge depending



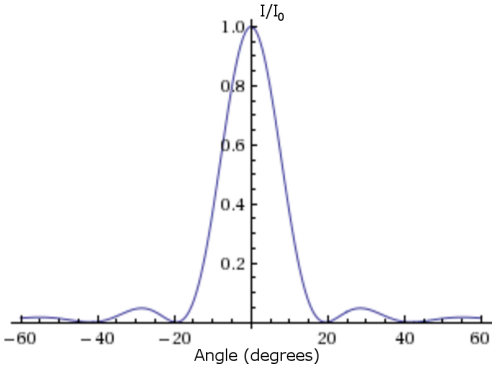
(a) Aperture width equal to wavelength, Huygens-Fresnel propagation



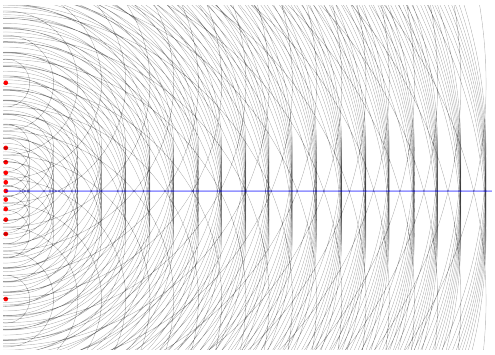
(b) Aperture width equal to wavelength, far field pattern



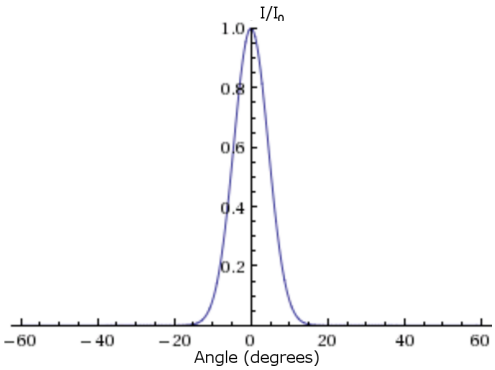
(c) Aperture width three times the wavelength, Huygens-Fresnel propagation



(d) Aperture width three times the wavelength, far field pattern



(e) Gaussian, respresented by unequally spread sources, with w_0 twice the wavelength, Huygens-Fresnel propagation



(f) Gaussian with w_0 twice the wavelength, far field pattern

Figure 2.2 – Demonstration of how the Huygens-Fresnel principle gives insight into the propagation of beams. A) Diffraction of a plane wave from an aperture with a width equal to the wavelength. Red points are the sources on the original wavefront, the grey circles are lines of equal phase due to those sources, and the blue line indicates a maximum in the intensity, i.e. a line along which more sources interfere constructively than destructively. B) Calculated far field diffraction pattern (normalised intensity) for an aperture of this type, as a function of angle (in degrees) from the centre of the aperture. C and D are the same for an aperture width three times larger than the wavelength, E and F for a Gaussian beam waist with w_0 twice the wavelength.

on the sign of the curvature. If it converges, it is also obvious that the beam will reach a point where the sign of the curvature inverts: a focal point or beam waist.

Many of the more traditional problems in optics can be explained with the Huygens-Fresnel principle as well, from refraction to double slit interference patterns and the Arago spot. The principle is not, however, immediately useful for quantifiable results. Gustav Robert Kirchhoff later derived a more mathematically useful formulation of the principle directly from the scalar wave equation[51]. Using Green's theorem, he was able to give a full integral formulation for the value of a field at a point if at each point on a closed boundary around it, the value of that field and its normal spatial derivative are known[45, 52]. As before, the trick to getting a finite result is choosing the integration boundary such that the contribution from anywhere other than the known region can be shown to be zero. The final result is:

$$E(r) = \frac{1}{4\pi} \int dS(\frac{\partial E}{\partial n}\frac{e^{jkr}}{r} - E\frac{\partial}{\partial n}(\frac{e^{jkr}}{r})), \tag{2.11}$$

where S is now the known part of the integration boundary, e.g. the wavefront or the spatial cross-section of the beam.

The Huygens-Fresnel principle and Kirchhoff's theorem are both very useful in describing many of the properties of propagating beams, however both are fundamentally scalar in nature and completely neglect the relations between the different field components. In most cases, at least in a homogeneous and isotropic medium, this is a good approximation and the separate field components can be treated independently without much loss in accuracy. Even the power flow, normally given by the Poynting vector and an inherently vectorial quantity, can be approximated quite well with this scalar theory. The direction of power flow is almost always normal to the wavefront and the magnitude is effectively $\frac{\epsilon_0}{2}E^2$. To describe beams, it is therefore rarely necessary to solve the full vectorial problem, unless the beam shape depends explicitly on the polarisation.

2.1.3 Example beams.

In this thesis I will discuss a number of different beam types, each with their own peculiarities. In order these are: (focused) Gaussian, Laguerre-Gaussian (vortex), Bessel, and Cylindrical Vector (CV) beams. Here, I will briefly describe the definition of each type of beam, discuss what makes them special and show how they are commonly used.

Let's start with the Gaussian beam, one of the most common of beams

in optics. Figure 2.2e shows an example of how a Gaussian beam propagates, while figure 2.3a shows how it acts near the narrowest point in the beam. The full (scalar) equation to describe a Gaussian beam is

$$E(r,z) = E_0 \left(1 + \left(\frac{z\lambda}{\pi w_0^2}\right)^2\right)^{-1/2} G_{Re}(r,z) G_{Im}(r,z) e^{-j\left(kz - \tan^{-1}\left(\frac{z\lambda}{\pi w_0^2}\right)\right)}, \quad (2.12)$$

where w_0 is the radius at which the amplitude of the field drops to $1/e$ of its maximum at the beam waist, or $z = 0$. The functions G_{Re} and G_{Im} are both Gaussian functions in r with a z dependent width that described the real amplitude decay (the width of the beam envelope) and the phase dependent amplitude modulation (the curvature of the phase front) of the beam respectively:

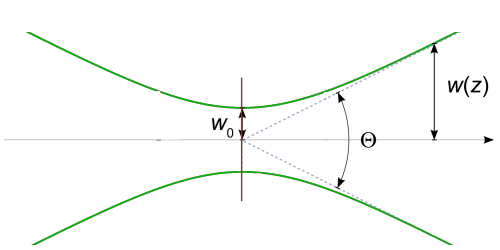
$$G_{Re}(r,z) = e^{-r^2/w_0^2\left(1+\left(\frac{z\lambda}{\pi w_0^2}\right)^2\right)} \quad (2.13)$$

$$G_{Im}(r,z) = e^{-jkr^2/2z\left(1+\left(\frac{\pi w_0^2}{z\lambda}\right)^2\right)}. \quad (2.14)$$

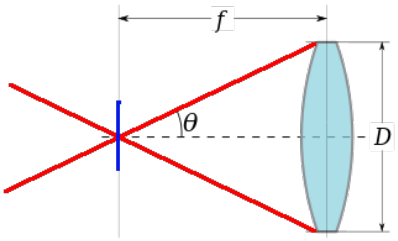
These Gaussian functions are what give the Gaussian beam its name. The last part of equation 2.12, $\tan^{-1}\left(\frac{z\lambda}{\pi w_0^2}\right)$, describes the Gouy phase of the beam, an additional phase term that changes most prominently near the focus, but is not relevant for any of the work discussed here. The rest of equation 2.12 is constant in r and describes the amplitude decay and modulation along the optical axis of the beam.

From this formulation, all key characteristics of the Gaussian beam can be seen immediately. G_{Re} is symmetric in $z = 0$, so the beam goes through a minimum width at $z = 0$. G_{Im} is antisymmetric in the same plane, so the phase curvature changes sign going through this plane. The size of the envelope, or beam width, $w(z) = w_0\sqrt{\left(1 + \left(\frac{z\lambda}{\pi w_0^2}\right)^2\right)}$ tends to $\frac{\lambda z}{\pi w_0}$ for sufficiently large z. Reversing this, an effective collection half-angle for the beam can be defined as $\theta = \tan^{-1}\left(\frac{\lambda}{\pi w_0}\right)$ which in free space is approximately equal to the numerical aperture (NA) of the beam $NA = n\sin(\theta)$.

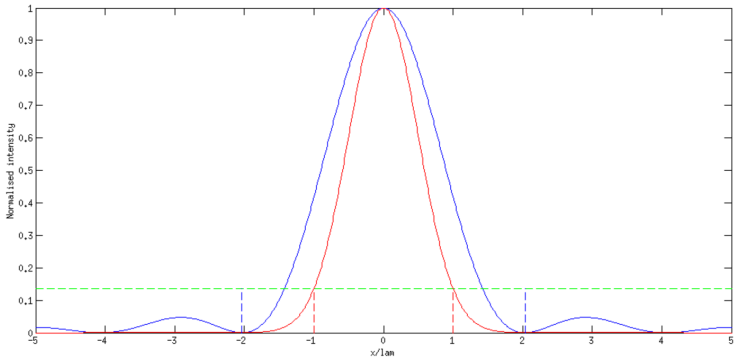
The NA is completely valid as a measure of beam divergence for Gaussian beams, although it is more commonly associated with intentionally focused beams. In that case the collection half-angle is instead defined through the radius of the first minimum of the Airy disk, so $\theta \approx \sin^{-1}\left(\frac{0.61\lambda}{r_0}\right)$. Alternatively the same value can be found geometrically as $\theta = \tan^{-1}\left(\frac{r}{f}\right)$, with r the radius of the focusing element and f the focal length. This is schematically shown in figure 2.3.



(a) Gaussian beam half angle



(b) Lens half angle



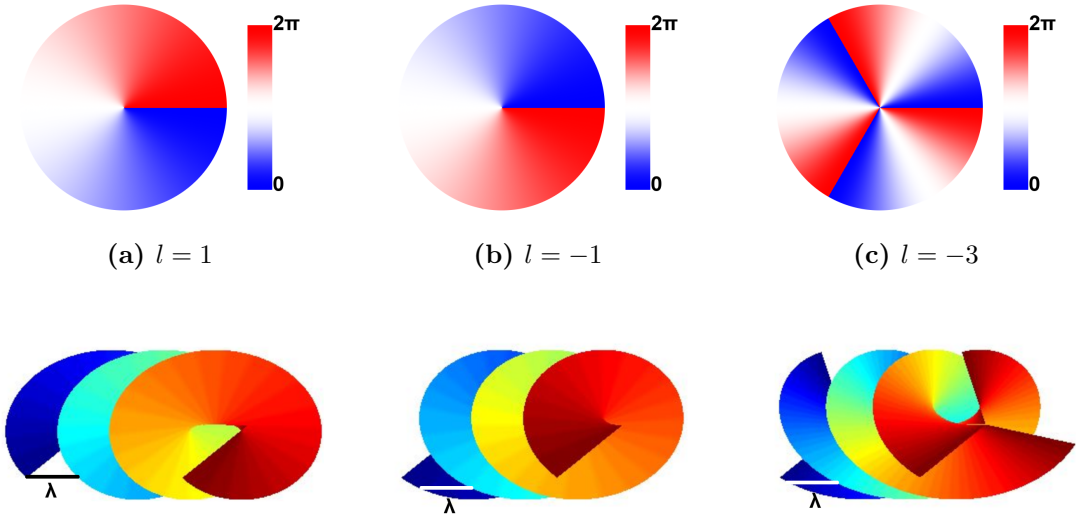
(c) Intensity profiles at the focal planes

Figure 2.3 – Demonstration of the differences between numerical aperture of a Gaussian beam and a plane wave focused by a lens. A) Gaussian beam showing the half angle used to determine its NA. B) Lens system showing the half angle used to determine its NA. C) Normalised intensities at the focal planes for both systems along the cross sections of the same colours indicated in A and B. For proper comparison the NA of both systems is approximately 0.3. The dashed green line indicates an intensity of $1/e^2$, while the dashed blue and red lines show the beam width used in the calculation of NA for lens and Gaussian respectively.

Gaussian beams are, in fact, only the lowest complexity example of a larger family of beams: the Laguerre-Gaussians[53, 54]. Usually written as $\mathrm{LG_{lp}}$, the Gaussian beam is technically a $\mathrm{LG_{00}}$ beam. Here l and p are an azimuthal and radial mode index, respectively. In my work I am only interested in the azimuthal mode index and will assume p = 0, which sets the most complicated factor in the beam description, the generalised Laguerre polynomial, to 1. Ignoring the Gouy phase, the field of such a beam can then be written as

$$E_{LG_{l0}}(r,z) = E_{LG_{00}}(r,z)e^{jl\phi}, \tag{2.15}$$

with $E_{LG_{00}}$ the function given in equation 2.12 and ϕ the azimuthal angle around the z-axis. l can take any integer value and effectively counts the direction and number of "twists" given to the original Gaussian beam, leading to a helical wavefront (see figure 2.4) with a steeper incline the higher the magnitude of l.



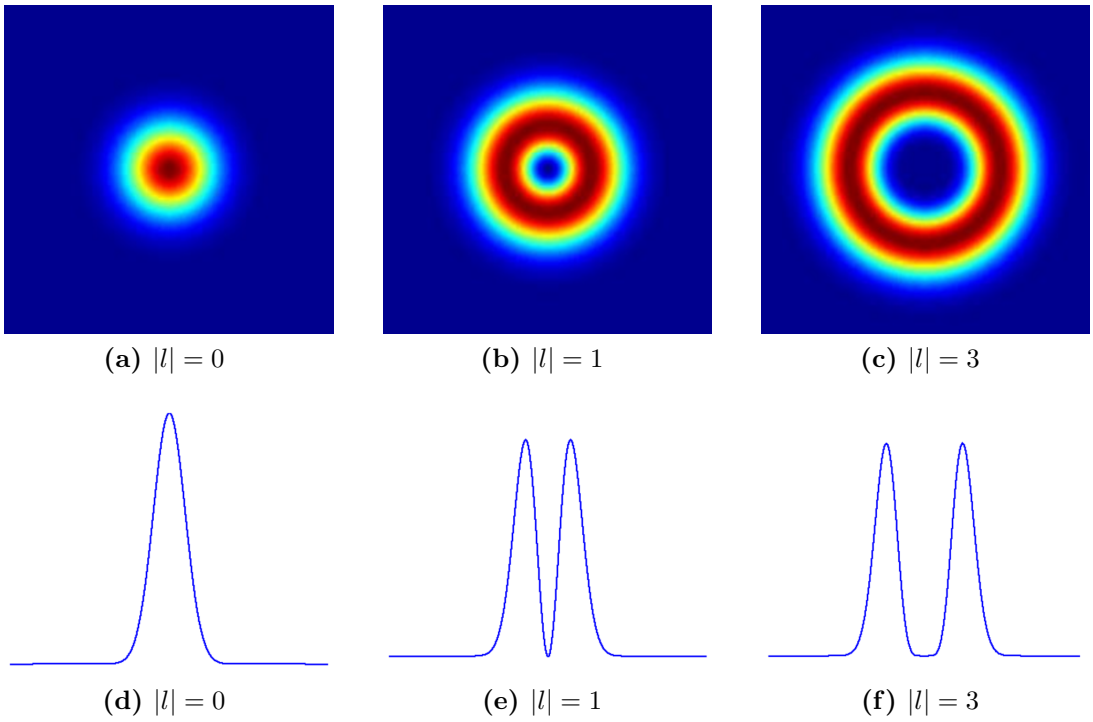
(a) $l = 1$ **(b)** $l = -1$ **(c)** $l = -3$

(d) $l = 1$ **(e)** $l = -1$ **(f)** $l = -3$

Figure 2.4 – Several orders of Laguerre-Gaussian beams. The top row, A-C, shows the phase of the beam at a spatial cross-section normal to the propagation direction. The bottom row, D-F, shows the wavefront of a beam, i.e. the surface of equal phase. A and D show an $l = 1$ beam, B and E an $l = -1$, and C and F an $l = -3$.

As derived in the previous section, the direction of power flow in a beam in the scalar description can be found by taking the normal to the wavefront. In the case of Laguerre-Gaussians this means the power flows helically: partially in the direction of the optical axis and partially in the direction normal to a radial from the optical axis. Effectively the light coils around the optical axis much like wind coils around the eye of a storm. This spiralling behaviour results in the phase front of the beam being tilted with respect to the propagation axis, and therefore the Poynting vector of the beam has an azimuthal component as well as an axial one. This gives rise to an orbital angular momentum to the beam which breaks the symmetry between left- and righthanded Laguerre-Gaussian beams[6, 53]. This effect is similar to the spin angular momentum associated with circularly polarised light, resulting from the phase delay between two orthogonal polarisations, which can similarly be left- and righthanded in nature. Notably, the orbital angular momentum is quantised as $l\hbar$, with l the azimuthal mode number of the Laguerre-Gaussian, while spin angular momentum can take only the values $\pm\hbar$, which results in orbital angular momentum having a much higher potential for exercising torque forces on objects, for example dielectric particles trapped in an optical tweezer[6]. This is also where the colloquial name for Laguerre-Gaussians finds its origin: vortex beams[4, 53].

The most obvious and striking feature of the Laguerre-Gaussians, however, is



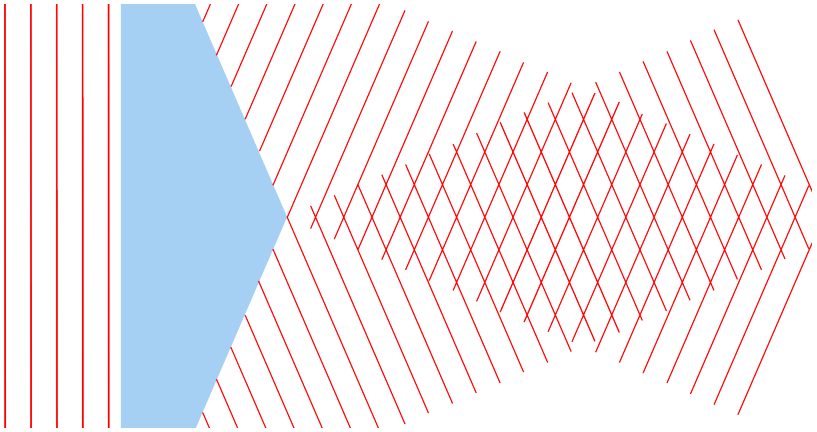
(a) $|l| = 0$ **(b)** $|l| = 1$ **(c)** $|l| = 3$

(d) $|l| = 0$ **(e)** $|l| = 1$ **(f)** $|l| = 3$

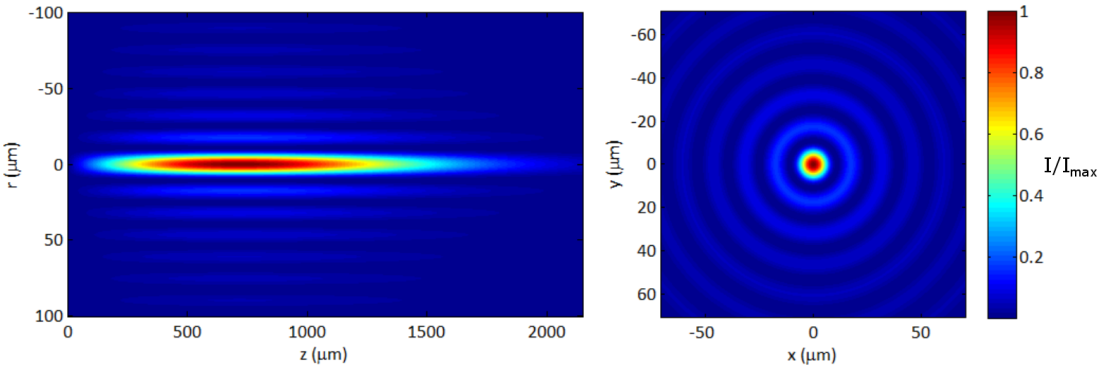
Figure 2.5 – Several orders of Laguerre-Gaussian beams. The top row, A-C, shows the intensity distribution of a beam at the beam waist in two dimensions. The bottom row, D-F, shows a cross section through the centre of those plots. A and D show an $l = 0$ or regular Gaussian beam, B and E an $|l| = 1$, and C and F an $|l| = 3$.

that, other than the plain Gaussian, all of them have a singularity on the optical axis where no field can exist. Mathematically, this follows immediately from ϕ being undefined at that point. The continuity of phase then forces the amplitude to be zero to compensate. More insightful is to notice that for any integer l and radius r_0, there are always equally many points with a given phase ϕ_0 as points with a phase of $\phi_0 + \pi$. From the Huygens-Fresnel principle, it is then immediately obvious that the contributions from these points have to cancel out at $r = 0$. The magnitude of l determines the exact size of the ring formed when the beams are focused down (see figure 2.5.) This central "hole" in the beam has given rise to the second nickname for this class of beams: donut beams.

Where Gaussian beams are traditionally used in applications such as imaging, optical trapping, laser lithography, etc., vortex beams have now been applied to most of these areas to complement them. In imaging, they form an integral part of the recent Nobel prize in Chemistry 2014[55], for super resolution imaging down to several tens of nanometres. The method, called Stimulated Emission Depletion Microscopy (STED), uses the donut shape of the vortex beam to over stimulate fluorophores everywhere except in the dark centre, allowing a normal Gaussian based fluorescence setup to capture only the signals from the small cen-



(a) Schematic of axicon operation.



(b) Quasi Bessel beam, axial cross cut.

(c) Quasi Bessel beam, radial cross cut.

Figure 2.6 – Example of a quasi Bessel beam and its generation with an axicon. The intensity plots are normalised to the maximum intensity. (a) shows a schematic diagram of an axicon and how the plane waves refract and interfere beyond it. (b) shows a cross section of an example quasi Bessel beam along its propagating direction, with $z = 0, r = 0$ the location of the axicon tip. (c) is a cross cut of (b) through its maximum.

tral part[1, 56]. In optical trapping, the helical wavefront is used to apply an orbital angular momentum to the trapped particles and spin them[5, 6].

All Laguerre-Gaussian type beams share the common property that they diverge or diffract as they propagate. Inherent in the definition is that this divergence can only be reduced by widening the beam. There are, however, several solutions to the wave equations that describe non-diffracting beams, i.e. beams that don't diverge as they propagate. These beams have been gaining traction in the microscopy community for their ability to provide very even illumination over areas or volumes much larger than typically possible with traditional Gaussians. They have also been used in more exotic applications such as optical injections, where a single beam is used to puncture a cell membrane and serve as a guiding gradient force trap to transport particles into the cell[9].

Besides the true plane wave, the easiest to understand non-diffracting beam is the Bessel beam. In two dimensions, a Bessel beam is formed by interfering two infinite plane waves with an angle between their optical axes[57–60]. The resulting interference pattern has a new optical axis that is the average of those of the plane waves, and has bright fringes parallel to that optical axis, the width and modulation amplitude of which depend only on the wavelength and angle between the plane waves. By rotating this pattern around the optical axis it creates a three dimensional beam with a cylindrical symmetry and a central fringe that is brighter than the surrounding ones, but otherwise retains the same properties (see figure 2.6). As with Gaussian beams, an angular term can be added to create higher order vortex-Bessel beams without destroying any of its other features.

If the infinite plane waves that form the ideal Bessel beam are replaced by real Gaussian beams, the result is a Bessel beam propagating over a limited distance depending on the Gaussian beam width and the angle between them, a so-called Quasi-Bessel beam. An effective way of generating a Quasi-Bessel beam is the axicon, which in its simplest form is just a conical dielectric structure. The cone's tip-angle and refractive index determine the deflection of an incident Gaussian beam[59].

All non-diffracting beams share a common feature that leads to them also being called "self-reconstructing" beams. If an obstruction, for example a trapped dielectric particle, is placed in the path of the beam then for normal diffracting beams the area behind that obstruction lies in the shadow of that particle and the beam is irrevocably disturbed[57–59]. Non-diffracting beams, however, will eventually return to the same state as before the obstacle as if it weren't there at all. This seems surprising, but looking at how a Bessel beam is constructed from angled plane waves it can be explained as the light that makes up any point in the beam not actually being transported to that point in the beam along the optical axis, but instead along the optical axes of the constituent plane waves, which some distance behind the obstacle never interact with it at all.

All the beams discussed so far have had one important element in common: they can be defined completely independently from their polarisation. In principle, this makes it possible for several of them to exist in the same space and be separated using only polarising elements. For example, a Gaussian y-polarised vortex beam and an x-polarised Bessel beam can share an optical axis, and appear as a complicated superposition of the two. When a polariser is put in the path of this beam, the transmitted beam can be made to be purely one or the

other. Mathematically, the two polarisation states are completely orthogonal.

There is, however, a class of beams that relies on polarisation to exist. This class of beams is commonly known as vector beams since they can't be described accurately as a scalar wave function. The easiest example are Cylindrical Vector beams (CV-beams)[61, 62]. In phase terms, these beams look a lot like the Laguerre-Gaussians or Bessel beams above, but they have an additional position-dependent polarisation. Mathematically, while all the previously discussed terms can be written as $E_0 e^{j\phi}$, with E_0 a scalar, for cylindrical vector beams the phase component stays the same but E_0 becomes a vector quantity, leading to a field of the form

$$\begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} e^{j\phi}. \tag{2.16}$$

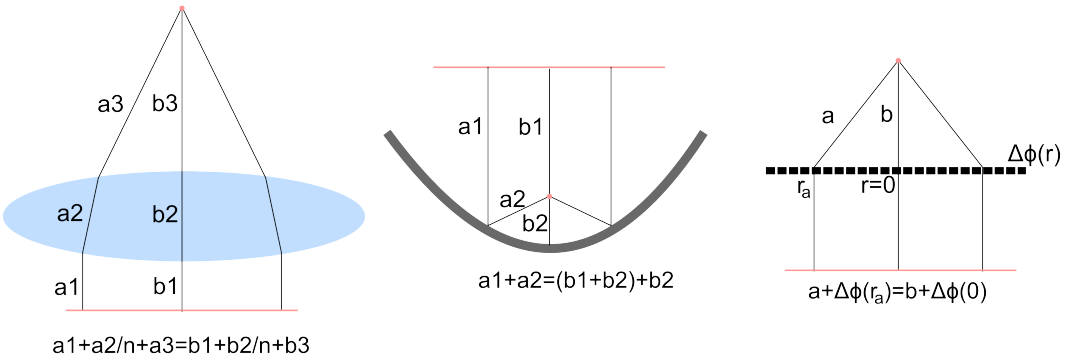
The lowest order CV-beams are azimuthally or radially polarised Bessel beams. The former have the polarisation always pointed normal to the radial from the beam's centre and can be described as $E_x = \sin(\theta)$, $E_y = -\cos(\theta)$ and $E_z = 0$, while the polarisation of the latter always points along the radial and is given by $E_x = \cos(\theta)$, $E_y = \sin(\theta)$ and $E_z = 0$.

2.2 Metasurfaces.

To create the beams discussed in section 2.1, a variety of methods is available. The traditional method involves the manipulation of optical pathlengths. For transmissive elements, this usually takes the form of changing the thickness of a transparent dielectric such that the phase delay of passing rays matches the difference between the desired input and output beams. If the result is a focusing effect, this is also known as a lens. Reflective elements achieve the same by varying the mirror surface topography. In either case, the result is generally a bulky curved structure. (See figure 2.7.)

Fabrication technologies have greatly improved in the last few decades allowing the creation of micro-lenses and mirrors with well-defined shapes down to only several wavelengths in size[63]. However, their inherently curved nature still means that they are hard to integrate into other structures. The primary goal of metasurfaces is to overcome that drawback.

The term metasurface derives from metamaterials, their 3D counterpart[10, 15, 64]. Where metamaterials are structured such that their apparent permittivity and permeability are customisable, rather than material dependent, metasurfaces attempt a similar feat in a 2D form factor. Succinctly put, metasurfaces attempt to convert an input beam into a desired output beam within a subwavelength



(a) Transmission **(b)** Reflection **(c)** Metasurface

Figure 2.7 – Diagrams of the operation of traditional transmissive and reflective lenses vs a metasurface. The equations included in the subfigures are all in terms of the optical path lengths, and therefore in units of phase, which means an integer multiple of 2π can be added or subtracted from any side of the equations if necessary. In each image different letters are used to denote different paths to the focal point, while the numbers indicate separate sections along the same path. (a) shows the operation of a traditional lens, where focusing is achieved by differing path lengths through a transparent and curved dielectric element. (b) shows how focusing is achieved by a traditional parabolic mirror, where the optical path lengths in air are restricted by the curvature of the mirror. (c) shows the operation of a transmissive metasurface, where the necessary phase difference to achieve focusing is folded into the different unit cells that make up the surface.

volume, creating apparent discontinuities in the phase evolution of a passing light beam.

Given their origin in the field of metamaterials and the prevalence of plasmonic, metallic, structures in that field, it is not very surprising that most of the research in the field of metasurfaces is focused on plasmonic metasurfaces[10, 11, 14, 29]. They have a number of obvious advantages. For example, the size of a plasmonic resonator can be far below the wavelength of the light, and they generally only weakly couple to their neighbours allowing for very local switching from one type to another. However, their shortcomings are equally obvious. Being metallic, they are lossy and the more effectively they interact with a beam, the more energy will be absorbed. Still, even with these severe limitations on their efficiency, most of the existing work on metasurfaces is in this area and much of it translates to the work in this thesis.

The crucial difference between metasurfaces and traditional optical elements is the subwavelength form factor. To achieve this small size, light needs to be manipulated much more strongly than in traditional optics. Take, for example, a lens. From the Huygens-Fresnel principle it is straightforward to deduce the phase front required to focus a beam, as the requirement is simply for all

rays to be in phase with each other at the focal point. In a lens, this is achieved by changing the refractive index such that light accumulates varying phases as it travels through different thicknesses of the lens. Therefore the absolute minimum thickness of a lens is found by solving:

$$\frac{2\pi}{\lambda}t + 2\pi = \frac{2\pi n}{\lambda}t \Rightarrow t = \frac{\lambda}{n-1}, \tag{2.17}$$

for a lens with refractive index n surrounded by air on all sides. This is the minimum thickness required to get a 2π phase modulation. Without the full 2π phase range it is still possible to achieve focusing to an extent, as in for example Fresnel zone plates, but in that case not every part of the structure can contribute fully to the focusing effect and the efficiency is restricted.

A metasurface needs to achieve the same effect in a sub-wavelength thickness. The only way to do so is through resonances. In a resonant element light gets trapped, bouncing around a number of times before escaping. The simplest resonant configuration is that of a Fabry-Perot cavity which in its most basic form consists of two parallel mirrors. If light is shone onto this cavity at normal incidence, some amount of the light will pass through one of the mirrors into the space between them. Most of that light reflects of the second mirror, returns to the first, and reflects again. Critically, when the roundtrip distance between the mirrors is an integer-half multiple of the wavelength, the returning light is out of phase with the originally reflected light of the first mirror. The returning light would therefore destructively interfere with the initially reflected light, meaning that more light will transmit into the cavity than would if the second mirror didn't exist, culminating in unity transmission through the cavity when exactly on resonance.

The reflectivity of the mirrors does not influence the resonance conditions. It does, however, determine how sharp the resonance is, with a higher reflectivity requiring a closer adherence to the resonance condition to overcome the natural reflection of the first surface.

In plasmonic resonators, the coupling is usually between the oscillating electric field of the input light beam and free electrons in the metallic nanoantennas, exciting Surface Plasmon Polaritons (SPPs) that form the resonating standing wave[14]. These resonances are tightly bound to the surface of the resonator, allowing the individual resonators to be relatively close together without strong coupling between them.

The most obvious way of modulating the phase of resonances is by slightly detuning from the centre wavelength, thereby changing the time light is trapped and therefore the phase with which it leaves the resonator. However, this has the

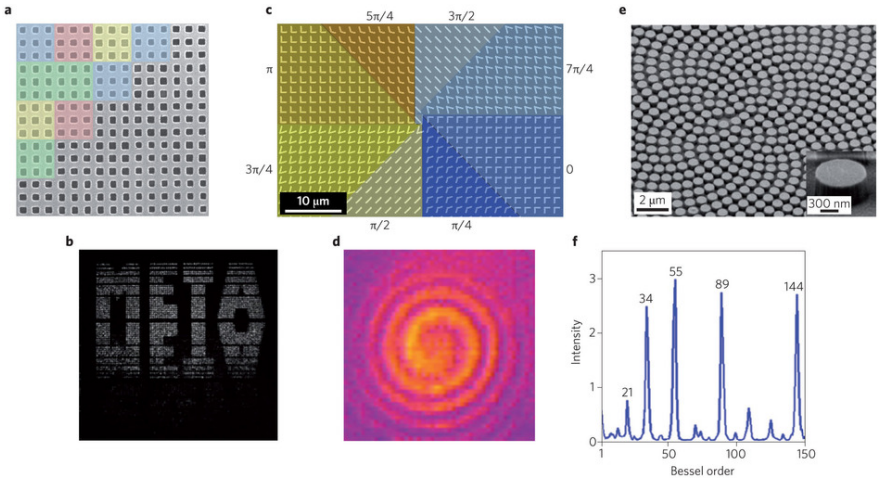


Figure 2.8 – Examples of different types of plasmonic metasurfaces as can be found in the literature[10]

downside that for stronger resonances, the coupling efficiency to the incoming light drops off quickly as one detunes, so different phases have different coupling efficiencies. More complicated metasurfaces will use several different resonator geometries to retain the conversion efficiency. The key point is that all metasurfaces use spatially variant resonators to encode a phase profile in discrete steps. (See figure 2.8 for examples). Therefore, all metasurfaces need to follow Nyquist's sampling theorem.

The Nyquist sampling theorem states that to accurately reproduce a cyclical valued function with a discrete number of points, the function needs to be sampled at least twice per cycle[65, 66]. For example, if a sine wave is probed more than twice per period, the lowest frequency sine that can be fit to those points has the same frequency as the original wave. In the case of beams, this means that for any 2π phase change at least two discrete points on the function need to exist for the desired beam cross-section to be the lowest order approximation to those points. The more frequently the function is sampled, so the slower the phase changes from unit cell to unit cell, the better.

For metasurfaces, this puts a fundamental limit on what wavefronts can be produced accurately with a given type of resonator surface. For example, a parabolic mirror has a phase function that is given by

$$\phi(r) = \frac{2\pi}{\lambda}(\sqrt{r^2 + f^2} - f), \tag{2.18}$$

with f the focal length, r the radius from the optical axis and λ the wavelength.

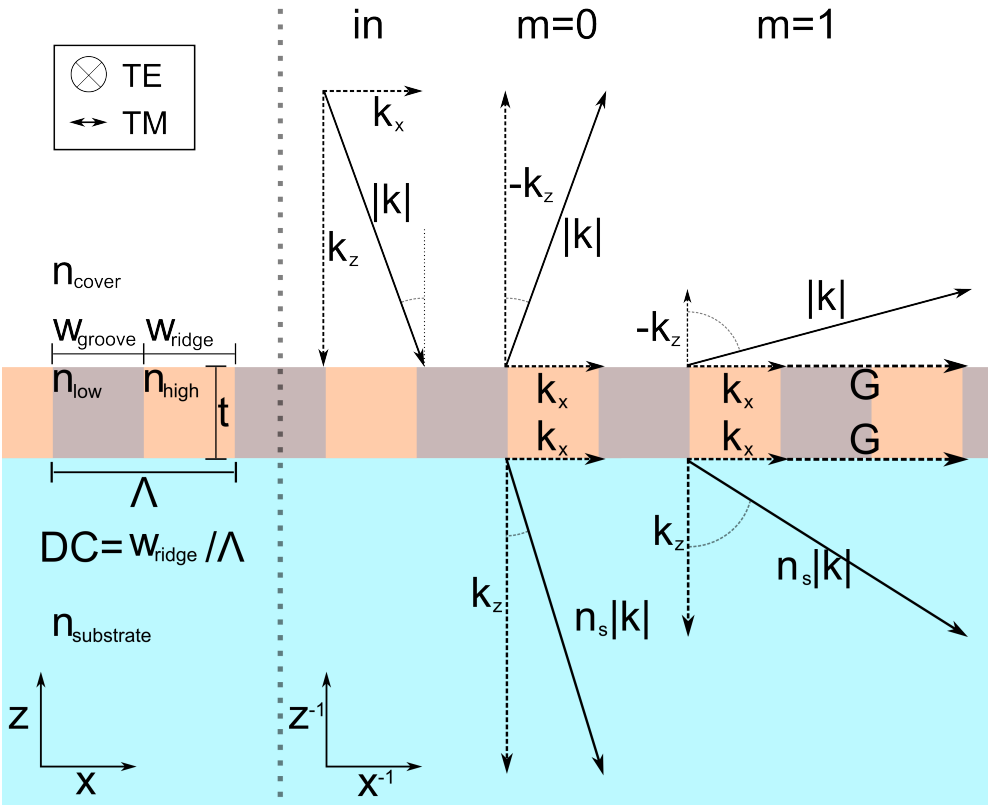


Figure 2.9 – Diagram of a typical grating geometry. Left of the dashed line labels the different grating parts in real space. Right of the dashed line illustrates the reciprocal space vectors used in the text (for simplicity n_{cover} is set to 1.)

To find the size of a 2π phase interval this can be rewritten to

$$\Delta r(\Delta\phi = 2\pi) = \sqrt{\left(\frac{\lambda\left(\phi' + 2\pi\right)}{2\pi} + f\right)^2 - f^2} - \sqrt{\left(\frac{\lambda\phi'}{2\pi} + f\right)^2 - f^2}, \qquad (2.19)$$

which for $\lambda\phi' >> f$ simplifies to just $\Delta r(\Delta\phi = 2\pi) \approx \lambda$. In other words, to create an arbitrarily large metasurface that acts as a parabolic mirror, the unit cells of that metasurface need to be smaller than half the wavelength. If the unit cells are larger there is a maximum numerical aperture that can be reached with that particular metasurface.

2.3 Sub-wavelength gratings.

So far, metasurfaces have been based on individual resonant structures with subwavelength size in all three dimensions. Those discussed in the rest of this thesis were designed with a slightly different starting point: dielectric gratings.

Gratings are structures that are periodic in one or more dimensions. Those discussed in this work are periodic in one dimension, continuous in a second and have a finite thickness in a third (see figure 2.9). While fabricated gratings are of course finite in all dimensions, calculations are usually done assuming infinity

in both the in-plane dimensions. With those assumptions in place, the geometry can be studied in only two spatial dimensions, and solved quasi-analytically using rigorous computational techniques discussed in the next chapter. However, purely analytically a number of useful qualitative observations can already be made. Except where explicitly mentioned, I will be assuming normal incidence light in the following discussion.

The periodic nature of gratings create certain restrictions on the stable modes that can exist within them[46]. Specifically, the electric field anywhere in the geometry has to fulfil the condition:

$$E(x) = E(x + \Lambda), \tag{2.20}$$

where Λ is the period and x the coordinate along the periodic axis. If the electric field is in the form of a wave, so $E(x) = E_0 e^{j(kx+\phi)}$, then this implies a stricter relation that uses the grating momentum $G = q\frac{2\pi}{\Lambda}$. This choice of momentum sets $e^{jG\Lambda} = 1$ for all integer q, and therefore:

$$E(x) = E_0 e^{j(Gx+\phi)} = E_0 e^{j(Gx+\phi)} e^{jG\Lambda} = E_0 e^{j(G(x+\Lambda)+\phi)} = E(x + \Lambda). \tag{2.21}$$

E_0 still depends on the out of plane dimension z. The grating modes that are described by this transformation are called Bloch modes[28, 67]. In general, all modes that exist in a grating can be decomposed into out of plane waves and in plane Bloch modes, resulting in a full field of the form:

$$E(x, z) = E_0 \sum_{q=-\infty}^{\infty} u(q) e^{jk_z z} e^{jGx} e^{j\phi}, \tag{2.22}$$

with $u(q)$ the amplitude corresponding to that mode and

$$k = \frac{2\pi}{\lambda} = \sqrt{G^2 + k_z^2} = \sqrt{\left(q\tfrac{2\pi}{\Lambda}\right)^2 + k_z^2}, \text{ or} \tag{2.23}$$

$$k_z = 2\pi\sqrt{\left(\tfrac{1}{\lambda}\right)^2 - \left(\tfrac{q}{\Lambda}\right)^2}, \tag{2.24}$$

for all integer q.

This last relation gives a lot of insight into how different gratings act. For example, take $\Lambda/q > \lambda$ with $q \neq 0$. Here k_z has a positive and real value, which implies a wave propagating at least partially in the z direction. The angle with

which it does so (with respect to the z-axis) is given by

$$\theta = \tan^{-1}\left(\frac{G}{k_z}\right) = \tan^{-1}\left(\frac{q}{\sqrt{\left(\frac{\Lambda}{\lambda}\right)^2 - q^2}}\right), \tag{2.25}$$

which corresponds to the usual far field diffraction formula $\theta = \sin^{-1}\left(\frac{q\lambda}{\Lambda}\right)$ wherever the angle is purely real.

When $\Lambda/q = \lambda$, k_z goes to zero. These modes therefore only propagate in the grating plane and therefore don't propagate into the far field. At this point the energy that would be radiated into that diffraction order suddenly gets redistributed over the other orders, leading to a sharp change in intensity in the other orders. This is called Wood's anomaly[33, 68].

When $\Lambda/q < \lambda$ the mode becomes evanescent, as k_z becomes purely imaginary. When $\Lambda < \lambda$, even the first order mode is of this type, and only the solutions for $q = 0$ retain a real valued k_z and therefore couple to propagating waves away from the grating. This mode has equal total k and k_x values as the incident mode, with only k_z changing sign on reflection. In other words, it propagates away from the grating with the same angle as the incident angle. This last type of grating is commonly known as sub-wavelength grating[31, 69, 70].

So far this section has focused purely on the implications of a periodic geometry, disregarding what makes up that geometry. While this approach gives plenty of information about the nature of diffraction and the possible far field patterns, it neglects the contents of the unit cell, i.e. the refractive index distribution of the problem. Even without knowing anything about the actual shape, size and refractive indices of the grating, some observations can be made readily to gain a further level of insight, and to connect grating physics with resonance theory. The important term for this discussion is the effective refractive index, n_{eff}[46, 71, 72].

The effective refractive index of a layer is an approximation of the real permittivity distribution. In simplest terms, it's a series expansion of the analytical description of the refractive index of a unit cell, as experienced by a propagating wave. The full problem depends on the direction of propagation, the wavelength, the material dispersion, the polarisation, the thickness of the layer, etc. However, a simple zero order approximation split in two polarisations leads to surprisingly accurate results under the right circumstances, i.e. when the period is sufficiently

small with respect to the wavelength of the light inside the grating[46]:

$$n_{\text{eff,TM}} = 1/\sqrt{\frac{\text{DC}}{n_{high}^2} + \frac{(1-\text{DC})}{n_{low}^2}}, \tag{2.26}$$

$$n_{\text{eff,TE}} = \sqrt{\text{DC}n_{high}^2 + (1-\text{DC})\, n_{low}^2}. \tag{2.27}$$

The terms Transverse Magnetic (TM) and Transverse Electric (TE) are used to indicate the polarisations of light incident on the grating that are perpendicular and parallel to the ridges, respectively.

With this effective index it becomes apparent that while the grating momentum G is completely unaffected, the light's momentum k scales with n_{eff} as $k = \frac{2\pi}{\lambda} = \frac{2\pi\,\text{n}_{\text{eff}}}{\lambda_0}$. This results in equation 2.24 becoming

$$k_{z,q} = 2\pi\sqrt{\left(\frac{\text{n}_{\text{eff}}}{\lambda_0}\right)^2 - \left(\frac{q}{\Lambda}\right)^2}. \tag{2.28}$$

So if n_{eff} is higher in the grating than the incidence region, k_z will also be higher, but the allowed values of k_x are still fixed by the grating period, so the angle with which the diffraction orders propagate within the grating are smaller than in the surrounding material. In the extreme case, this means that the Wood's anomaly for a particular order of diffraction is reached outside the grating first as well, leading to differing numbers of diffraction orders in the different regions. This is schematically shown in figure 2.10, which illustrates the propagation paths of light in a grating that supports values of q up to 2 while remaining sub-wavelength in the surrounding materials.

Based on these observations gratings can be roughly divided into several categories[28]: diffraction gratings ($\Lambda > \lambda_0$), sub-wavelength gratings ($\lambda_0 > \Lambda > \lambda_0/\,\text{n}_{\text{eff}}$) and deep sub-wavelength gratings ($\lambda_0/\,\text{n}_{\text{eff}} > \Lambda$). In the first category the grating modes can mostly be neglected and classical diffraction theory leads to good results. In the last category there is only one mode possible per region, allowing diffraction to be neglected and effective index theory to be a good first order approximation with equations 2.26 and 2.27 good approximations. The intermediate region is where the interesting things happen.

The first of the requirements for a sub-wavelength grating means there is only one angle with which light propagates away from the grating into the far field. (I have and will continue to neglect incidence angles and only consider normal incidence for convenience, but the angle is fixed by k_x staying the same and k_z correcting for the material.) The intensity distribution between transmission and reflection is, however, controlled by a large number of competing effects mak-

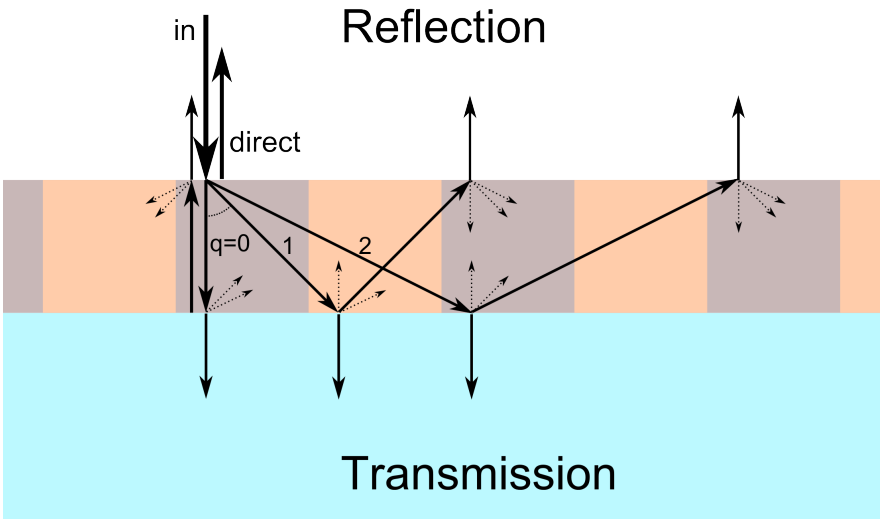


Figure 2.10 – Schematic of the first few grating resonance propagation paths. The numbers in the figure indicate the diffraction orders inside the grating necessary to achieve a given path. For a resonance condition to be satisfied the phase accumulation along at least one of these paths needs to be a multiple of 2π.

ing the exact solution in this region very hard to find analytically. In general a computational method is used, for example Rigorous Coupled Wave Analysis[73] which will be discussed in the next chapter, but it is informative to look at the effects that lead to these complicated pictures.

The first step in getting an intuitive insight into the type of resonances supported by sub-wavelength gratings is to make the assumption that the entire unit cell can be described as a single uniform piece with a given effective index. Equations 2.26 and 2.27 give an approximation of such an effective index as a function of only duty cycle and polarisation. While, as mentioned, these are a good approximation for $\lambda_0 > \mathrm{n_{eff}}\,\Lambda$ where the wavelength can not correctly resolve the grating, when λ_0 decreases below this condition, it quickly becomes overly simplistic and a true effective refractive index would need terms dependent on λ_0/Λ[71, 72]. However, to develop the resonant theory the effective index is still a useful tool.

The second step is to consider the phase evolution of light as it propagates through the grating layer. This porpoagation can be decoupled into an out-of-plane and an in-plane component, k_z and k_x. The out-of-plane component is then effectively a Fabry-Perot cavity as discussed in the previous section, with the standard k-vector being replaced by the k_z component. The resonance condition can then be written as:

$$k_{z,q} \cdot 2t = m \cdot 2\pi. \tag{2.29}$$

Filling in equation 2.28 for $k_{z,q}$ and solving for λ_0 leads to the full out of plane resonance condition for sub-wavelength gratings:

$$\lambda_0 = \frac{\mathrm{n_{eff}}}{\sqrt{\left(\frac{m}{2t}\right)^2 + \left(\frac{q}{\Lambda}\right)^2}}. \tag{2.30}$$

For $q = 0$, this reduces to the standard thin film interference equation $\lambda_0 = 2\,\mathrm{n_{eff}}\,t/m$, but for higher orders, the resonance wavelength also becomes a function of the period.

For the design of metasurfaces, it is more useful to look at a fixed wavelength and thickness and rewrite the equation to give period as a function of the effective index, and by extension duty cycle. The result has the form

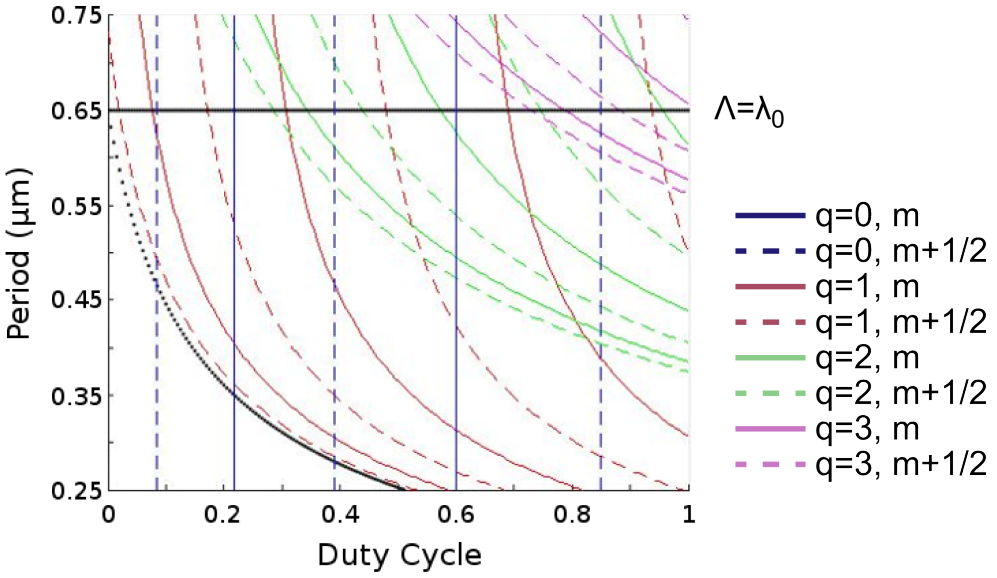
$$\Lambda = \frac{q}{\sqrt{\left(\frac{\mathrm{n_{eff}}}{\lambda_0}\right)^2 - \left(\frac{m}{2t}\right)^2}}. \tag{2.31}$$

An example of this representation is plotted in figure 2.11 next to a numerically calculated reflectivity map for the same grating parameters. The general behaviour of the predictions match the numerical results quite well, showing the mode spectrum getting more complicated towards the top right, the general curving behaviour of the modes, and the sharp cutoff at the $\lambda = \Lambda$. The exact positions of the resonances do not match, however, most likely due to the inaccuracy off the effective indices used for figure 2.11a.

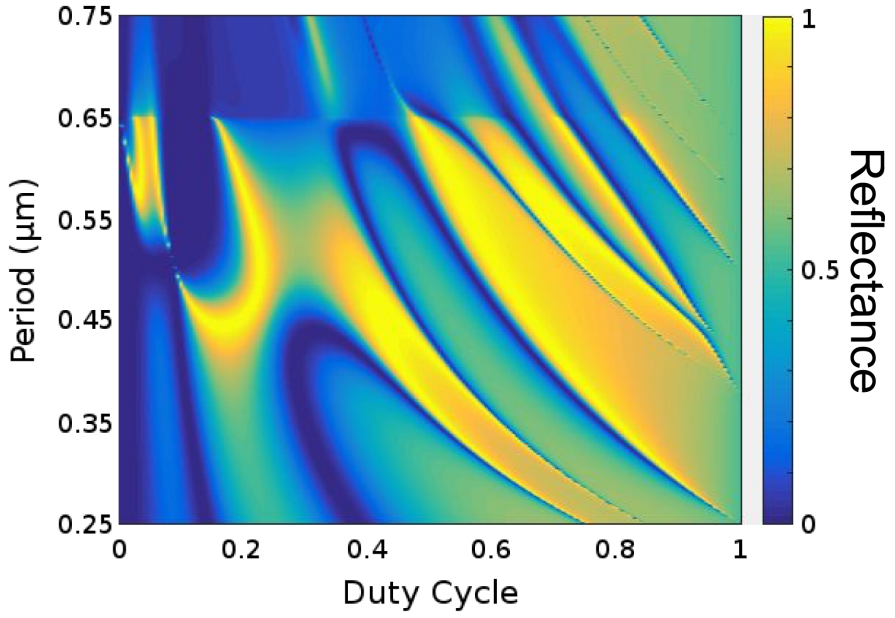
While this method of solving the resonance conditions is very straightforward, it does not make it clear what the different orders of resonance look like. To make this more obvious, equation 2.25 provides the necessary insight by relating q to the diffraction angle through $\theta = \tan^{-1}\left(\frac{G}{k_{z,q}}\right)$. The results (schematically depicted in figure 2.10) show that for $q = 0$ the diffraction angle inside the grating is 0 degrees and therefore a simple Fabry-Perot. For higher q however, the angle is nonzero and the mode propagates laterally through the grating. The grating effectively becomes a waveguide. For this reason this type of resonance is also known as Guided Mode Resonance (GMR).

The consequence is that the $q = 0$ modes are very short range, low propagation length resonances. The GMRs however can under the right circumstances travel quite far along the grating before scattering out. This can lead to very high Q resonances, although their long propagation lengths are not conducive to rapidly changing grating structures.

The allowed resonance orders for GMRs are given by the thickness, which puts a limit on the mode index m or the number of wavelengths that fit into



(a) Resonance conditions



(b) Reflectivity

Figure 2.11 – Graphs to compare the predictions of equation 2.31 (a) with numerically calculated reflectivities (b). The latter was calculated using Rigorous Coupled Wave Analysis (RCWA). Both figures were calculated for TE polarised, normal incidence light with a wavelength of 650nm and a grating thickness of 350nm, setting the refractive index of the ridges to 3.5 and all other regions to 1. The effective index used in the predicted resonance conditions was calculated using equation 2.27.
The thick black line in (a) indicates the relation $\Lambda = \lambda_0$, i.e. the sub-wavelength criterion, while the black dotted curve represents $\Lambda = \lambda_0 / n_{\text{eff}}$, or approximately where the deep-subwavelength region should start. Blue, red, green and magenta coloured lines represent $q = 0$, 1, 2 and 3 respectively, with continuous lines the integer m relations and dashed lines the integer plus half m.

the optical path, as well as the period, which determines the non-evanescent diffraction orders q. The effective index influences both equally. However, while the thickness can be increased to allow for higher order modes without penalty, the period is limited by the sub-wavelength condition, i.e. $\Lambda < \lambda_0$. The majority of sub-wavelength grating resonances are therefore only of the $q = 0$ and $q = 1$ types, with no other diffraction angles allowed in the grating structure.

This leads to a major feature of sub-wavelength gratings. Because there are only two modes that allow for energy transfer through the grating, the transmission of the grating is entirely determined by the sum of the transmittivities of these two modes. If they happen to be of equal amplitude but opposite phase, the total transmission of the grating drops to zero, and as a consequence the reflection becomes 100%[28]. Moreover, if two of such conditions happen relatively close together, the region in between tends to have extremely high reflectivity as well, leading to highly efficient broadband mirrors[74]. The opposite condition relies on the balancing of the two grating mode reflectivities and the direct reflection, making actual 100% transmission less common, though high transmission regions are still plentiful.

The full picture needs to take into account that on every interaction with either grating surface light diffracts and redistributes over the different diffraction orders (shown in figure 2.10 by the short dashed arrows.) This leads to a highly complicated cross-coupling between the grating modes, adding an extra layer of complication.

A similar analysis can be performed on the k_x component of the phase, this time matching it to the period rather than the thickness. However, this turns out to be rather trivial given that for normal incidence we have $k_x = G$, and therefore:

$$k_x \cdot \Lambda = q\frac{2\pi}{\Lambda}\Lambda = q \cdot 2\pi = m \cdot 2\pi, \tag{2.32}$$

with both q and m integers. Clearly the m condition is always satisfied and the $m + \frac{1}{2}$ can never be fulfilled. In other words, the very nature of Bloch waves enforces consecutive periods to be coherent for any guided wave, and therefore the only relevant resonances are those controlled by the out of plane propagation of the light, the Fabry-Perots and GMRs.

In the metasurfaces work presented here, the goal is to get sets of grating parameters that allow control of a wide range of phases, maintaining the amplitudes of either transmission or reflection, while being able to change be-

tween them rapidly. In other words, short propagation lengths are desired and the high Q conditions of GMRs need to be avoided. We note that, both the Fabry-Perot modes and the GMR modes are usually relatively local, allowing for adiabatic changing of the grating parameters without overly altering the local grating response.

As with all resonant effects, the phase of the outcoupled light inherently goes through a 2π shift as we go from one resonance order to the next, given that the optical pathlengths of the resonances increase by one wavelength. Typically, the intensity between the resonant conditions falls off sharply. In sub-wavelength gratings, the interplay between the two types of mode allows for regions where the drop in transmission or reflection is mitigated. In effect, one mode-type bridges the gap for the other. By choosing the right combination of grating parameters, it becomes possible to use this effect to find 2π phase paths with almost constant high reflectivity or transmittivity.

As I hope to have illustrated in this section, finding the exact combination of grating parameters that lead to such regions is exceedingly difficult analytically. In practice, the problem is tackled through numerical calculations as will de discussed in the next chapter.

3 Methods.

The methods used for the work described in this thesis can be divided into three distinct categories: modelling, fabrication and characterisation. As each of these is a necessary prerequisite for the next, this order naturally reflects the research process. In this chapter I will explain the fundamental processes for each of these three categories that are used for all work in this thesis. This should give a complete picture of the steps involved in the experimental research of dielectric grating-based metasurfaces. The methods unique to each metasurface will be treated in their respective chapters.

3.1 Modelling.

Modelling is the computational workhorse of the design process. At its core it uses a computer to solve complex systems of equations that reflect physical processes of interest. More specifically it can provide a virtual experimental process that allows for quick approximations to how a physical system would work.

For this work, modelling is primarily used for two purposes: to calculate the response of linear gratings to incident light, which provides the building blocks on which the metasurfaces are based, and to test metasurface designs before fabrication. The former is accomplished through a procedure called Rigorous Coupled Wave Analysis, while for the latter I have made use of the commercial finite element modelling software COMSOL Multiphysics.

3.1.1 Rigorous Coupled Wave Analysis.

Rigorous Coupled Wave Analysis (RCWA) is a pseudo-analytic method of calculating the optical response of periodic multilayer systems. It is one of the most efficient calculation methods for determining the reflection and transmission properties of an arbitrary grating type structure, developed in the early 1980s by Moharam and Gaylord[73, 75]. Given that all structures in this work are based on dielectric gratings, RCWA is ideally suited as a fast and flexible method for investigating the fundamental gratings that act as building blocks for the eventual metasurfaces. RCWA is therefore the first and most critical step towards designing most of the structures in this thesis (chapter 6 being the exception).

RCWA can be summarised in a handful of separate steps with the end goal of finding the optical response of a unit cell repeated with a fixed period. The first step involves splitting the geometry into parallel layers which are each individually homogeneous in the direction perpendicular to the periodicity (see figure 3.1). The wave equation is then solved for each layer separately, before requiring

continuity in the tangential electric and magnetic field components between the layers. The periodicity of the geometry allows the solutions to the wave equation to be found as a Fourier series that can be truncated to a limited number of orders. This truncation reduces the problem to a manageable system of coupled differential equations, which can be solved analytically. The end result is the complex diffraction efficiency into the various reflected and transmitted diffraction orders of the structure, or in the case of a sub-wavelength grating effectively the phase and amplitude of reflectivity and transmittivity.

In order to describe a structure by RCWA, it needs to fulfil three criteria. First, it needs to have one dimension along which it is bounded by semi-infinite and homogeneous regions in which the solutions to the wave equation can be written as simple plane waves. Second, all perpendicular dimensions used in the calculation need to be infinitely periodic with the same period for the entire geometry. Finally, the geometry has to be able to be split into layers that are each individually constant in the non-periodic direction. This last requirement means that any structure not consisting of vertical steps can only be approximately solved for by using multiple layers as a staircase approximation to the actual geometry.

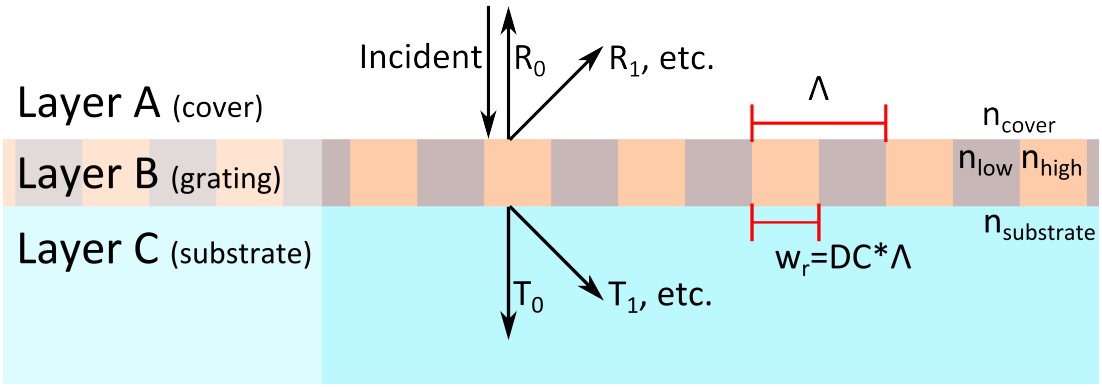


Figure 3.1 – Diagram to illustrate the Rigorous Coupled Wave Analysis procedure.

Once a geometry is chosen, the next step is to find a general solution to the wave equation for each layer. The allowed electric field solutions can always be written as a sum of plane waves:

$$E = \sum_{m=-\infty}^{\infty} A_m e^{i\mathbf{k_m}\cdot\mathbf{r}}, \tag{3.1}$$

where A_m is the amplitude of the m^{th} order, $\mathbf{k_m} = \frac{2\pi n}{\lambda}\hat{\mathbf{k}}_\mathbf{m}$ is the corresponding wavevector, and $\mathbf{r}$ is the position vector. For the different layers, this simplifies in several ways. For convenience I will label the semi-infinite region of incidence

and reflection as A, the intermediary layers as B, and the transmission layer as C.

First, the incident layer is generally chosen as a single plane wave, so the sum can be dropped and the entire system of equations normalised to unit incidence, making it:

$$E_{inc} = e^{i\mathbf{k}_{\mathrm{A,inc}}\cdot\mathbf{r}}. \tag{3.2}$$

In the same layer the only other possible existing waves are those propagating away from the structure, the reflected waves, so the total electric field in this layer becomes:

$$E_{\mathrm{A}} = e^{i\mathbf{k}_{\mathrm{A,inc}}\cdot\mathbf{r}} + \sum_{m=-\infty}^{\infty} r_m e^{i\mathbf{k}_{\mathrm{A,m}}\cdot\mathbf{r}}, \tag{3.3}$$

with r_m being the amplitudes of the m^{th} reflected diffraction order (so the reflectivity is $R_m = |r_m|^2$), and $\mathrm{sgn}(\hat{\mathbf{k}}_{\mathrm{A,inc}} \cdot \hat{\mathbf{z}}) = -\mathrm{sgn}(\hat{\mathbf{k}}_{\mathrm{A,m}} \cdot \hat{\mathbf{z}})\ \forall\ m$ (in other words, they need to be travelling in the opposite z directions.)

The bottom, transmission, layer is very similar except it only needs the sum and the waves travel in the same direction as the original incident wave:

$$E_{\mathrm{C}} = \sum_{m=-\infty}^{\infty} t_m e^{i\mathbf{k}_{\mathrm{C,m}}\cdot(\mathbf{r}-d\hat{\mathbf{z}})}, \tag{3.4}$$

where t_m are the amplitudes of the transmitted diffraction orders ($T_m = |t_m|^2$) and d is the total thickness of the structure.

The intermediate layers are more complicated due to the inhomogeneity of the refractive index. However, we can make several observations based on the restrictions placed on the geometry that simplify the problem: each layer is homogeneous in the $\hat{z}$ direction and periodic in all perpendicular directions. (To illustrate I will be treating the problem from here on as 2D with the only periodic dimension being $\hat{x}$, but it holds completely for 3D as well.) In other words $n(x, z) = n(x + \Lambda, z)$, with n the refractive index and Λ the period, so the refractive index can be Fourier expanded as:

$$n(x, z) = \sum_{m=-\infty}^{\infty} n_m(z) e^{im\mathbf{G}\cdot\mathbf{x}} = \sum_{m=-\infty}^{\infty} n_m(z) e^{imGx}, \tag{3.5}$$

with $\mathbf{G} = G\hat{\mathbf{x}} = \frac{2\pi}{\Lambda}\hat{\mathbf{x}}$ the grating vector. In an analogous fashion, we can do the same for the electric fields:

$$E_{\mathrm{B}} = \sum_{m=-\infty}^{\infty} S_m e^{i\sigma_{\mathbf{m}}\cdot\mathbf{r}}, \tag{3.6}$$

where $\sigma_{\mathbf{m}} = \mathbf{k}_{\mathrm{B},\mathbf{0}} - m\mathbf{G}$ (here σ is also technically a function of the layer number.) This last equation can be further simplified by phase matching the different layers, as $\mathbf{k}_{\mathrm{A},\mathbf{m}} \cdot \hat{\mathbf{x}} = \sigma_{\mathbf{m}} \cdot \hat{\mathbf{x}} = \mathbf{k}_{\mathrm{C},\mathbf{m}} \cdot \hat{\mathbf{x}}$.

These electric field and refractive index formulations can then be plugged into Maxwell's equations (see equations 2.1) to form an infinite system of second order differential equations that can rarely be solved exactly. By truncating the number of retained orders in the infinite sums, the system becomes solvable, at the cost of no longer being strictly analytical. Finding out how many orders need to be retained requires a convergence test, essentially running the same problem with increasing number of orders until the result remains stable. The more complex the unit cell, the more orders are necessary to reach this point. All the work in this thesis is based on binary single layer gratings, the simplest unit cell possible, so the number of required orders is relatively small. For all the results obtained in this thesis I chose to retain ± 21 orders. (It is possible that this is not enough for some parts of the RCWA results shown in this thesis, however the gratings used for devices have all been verified either using COMSOL or experimentally and therefore I have never had any reason to doubt this choice.)

The system of equations can be written in matrix form and solved using a variety of standard mathematical techniques to give the field distributions corresponding to each order in each layer separately. Finally, we need to introduce the boundary conditions between the layers in the z dimension to solve for the amplitudes $S_m(z)$, R_m and T_m uniquely. The relevant boundary conditions are the requirement of the tangential components of the electric and magnetic fields to be continuous on any boundary. This means that we have two equations on every boundary: between regions A and B, B and C, and between the separate layers that make up region B.

The complete calculation can then be written as a single matrix equation that solves for all amplitudes. The matrix size is $2(N+1)M \times 2(N+1)M$, where N is the number of intermediary layers in the geometry (always just 1 in this work) and M the number of diffraction orders retained (here 21, so the total matrix size used here is 84×84).

The implementation of RCWA we use was originally written in MATLAB by Kevin Harper[76], and adjusted by members of our group (most notably Dr. Emiliano Martins and myself) in various stages to better suit our purposes. The code is a purely two dimensional implementation of the procedure and accepts all the relevant inputs: wavelength, angle of incidence and polarisation of the incident wave, as well as sizes and refractive indices of the different regions in the geometry. The outputs are the complex reflection and transmission coefficients

of the geometry, the magnitudes of which give the diffraction efficiencies into each of the diffracted orders while the arguments are the corresponding phases. The code does not output the intermediary amplitudes S_m or wavevectors σ_m and does not directly give information on the propagation length, Q-factors or mode shapes of the intermediary grating modes. However, as we generally do not look at any one RCWA calculation in isolation but rather at multidimensional arrays of results it is generally straightforward to indirectly identify the regions of interest for metasurface design, i.e. the low propagation length regions.

3.1.2 COMSOL.

While Rigorous Coupled Wave Analysis is essentially a design tool, COMSOL Multiphysics[77] is instead used to test the finished designs. COMSOL is a commercially available Finite Element Modelling (FEM) program in which light propagation through arbitrary geometries can be numerically simulated. This approach makes COMSOL ideally suited for studying the response of metasurfaces under different types of illumination. With increasing size of the geometry, the required system resources grow significantly though, severely limiting the size of the geometries it can be used on. This scaling is mostly a problem for three dimensional simulations, though even for two dimensional simulations the limits of our available computers are easily reached. Still, within these limits, a large range of useful test cases can be explored in detail.

A thorough explanation of FEM simulation methods is beyond the scale of this thesis, but the essence is straightforward. The geometry is split into a discrete set of individual elements, each of which is assigned a single value for each relevant property of the problem, i.e. material constants and electromagnetic fields. The Maxwell equations are then iteratively solved on boundaries between adjacent elements until a stable solution is reached for the whole system.

From this description, several of the most important features of FEM can be extracted. First, the element size needs to be small enough to resolve the physical effects of interest. In the case of light this usually means about 5 or 6 elements per wavelength in large media, or the same per decay length near boundaries and subwavelength features, which follows from Nyquist's sampling theorem[66]. Second, as the ratio of size and complexity to wavelength increases, the number of elements necessary also grows. As for each step in the calculation process the Maxwell equations need to be solved on all boundaries simultaneously, this large number of boundaries impacts on the required computing power. Balancing these two principles results in the dividing of a model into elements, or meshing, being one of the most crucial steps. Third, the formulation of boundary conditions crucially determines if and how a given simulation converges. The wrong boundary

conditions can lead to no converging solution at best and misleading results at worst.

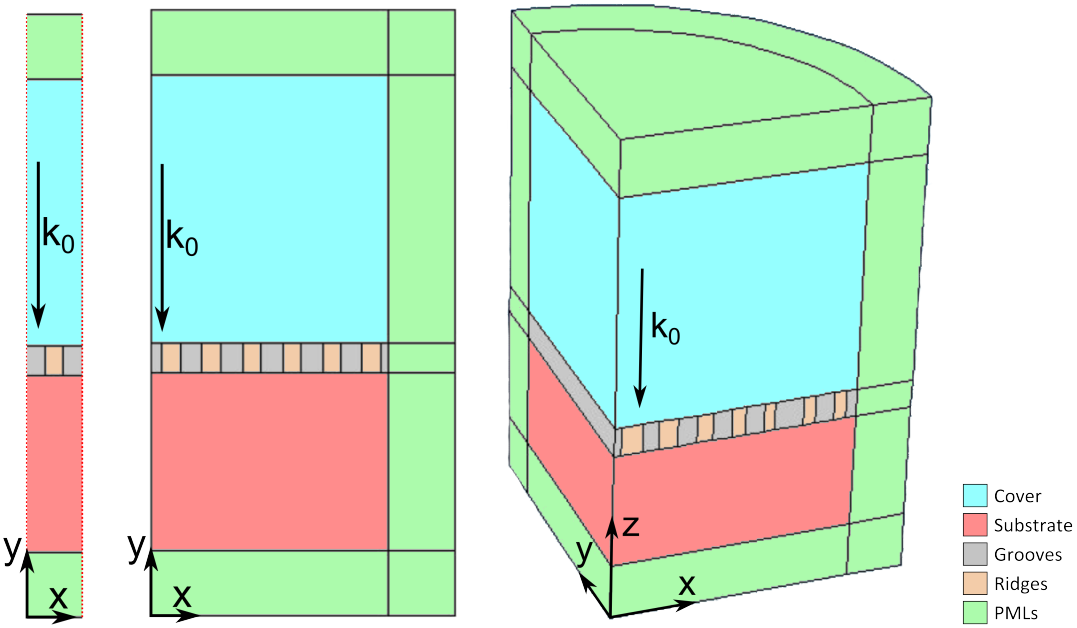
Other than the peculiarities of the user interface the rest of FEM is actually very robust. Once the above are chosen properly, simulating metasurfaces with varying size and functionality, under a large range of conditions such as wavelength or angle of incidence, is relatively straightforward. The results offer a wealth of information to test any number of cases that would be difficult or time consuming in an experimental setting. Efficiencies, bandwidth or tolerance to fabrication error are just a few examples of how COMSOL is used in this work.

In the course of this project I have used COMSOL extensively for a multitude of such tests, each requiring slight adjustments to the actual simulation geometry or setup. Almost all of them can be put into three categories, which are illustrated in figure 3.2: 2D periodic, 2D full device and 3D full device. The latter two would usually be used in the fashion shown in the figures, i.e. with symmetry boundaries on the optical axis, but in some cases the full metasurface has been simulated.

In most cases, a plane wave excitation is defined at the top or bottom boundary of the geometry, acting as a collimated beam illumination that overfills the aperture of the metasurface. This is the most common illumination method, as the metasurfaces are all designed for collimated, normal incidence, monochromatic light. For the non-periodic simulations I have also used point dipole sources to test the reverse response of for example focussing metasurfaces. The boundary conditions are some combination of Perfectly Matched Layers (PMLs) that act as nonreflecting boundaries, ports that allow excitation and absorption of light with specific k-vectors, Perfect Electric/Magnetic Conductors that work as absorbing boundary for incident light with parallel/perpendicular polarisation and symmetry boundary for the other, and periodic boundary conditions. (For precise mathematical definition of these boundary conditions please see the COMSOL manual, [78].)

3.2 Fabrication.

Fabricating dielectric photonic microstructures in general and metasurfaces in particular requires several steps: pattern a masking layer that covers the dielectric, transfer that mask into the dielectric through an etching process, then strip the remaining mask to leave only the desired structure. Each of these steps can be accomplished in numerous ways, but for the metasurfaces described in this thesis the methods used are basically the same.



(a) 2D (periodic). **(b)** 2D. **(c)** 3D. **(d)** Legend.

Figure 3.2 – Typical simulation geometries in COMSOL. (a) shows a 2D simulation of a 1D periodic grating, where the full calculation is done for a single unit cell only with periodic (Floquet) boundary conditions on the sides, effectively solving for a grating with infinitely many of such unit cells side by side. (b) shows a 2D simulation geometry employed when the grating structure is no longer perfectly periodic and periodic boundary conditions can not be used. The structure shown has a symmetry boundary condition on the left side, but depending on the symmetry of the simulation full gratings were also simulated. (c) shows the simulation geometry employed when performing 3D simulations. Due to limitations in computing power, these were only ever performed in the manner shown, with symmetry planes in both the xz-plane as well as the yz-plane.

1. Sample preparation:

   (a) Cleaning: a sample of the right material and size is chosen and cleaned using a combination of alcohols, ultra-sonic baths and acids.

   (b) Spinning: a polymer resist layer is spun onto the clean sample creating a thin uniform masking layer.

   (c) Baking: the sample with polymer is then put on a hotplate at moderately high temperatures to solidify the mask.

   (d) (Optional) Metal evaporation: if the sample material is an insulator, a thin charge dissipation layer is deposited onto the polymer using thermal evaporation.

2. Mask patterning:

(a) Exposure: the sample is then placed in the Electron Beam Lithography system and the desired pattern is exposed in the resist.

(b) (Optional) Metal removal: if a charge dissipation layer was deposited, it now gets removed with an acid etch.

(c) Development: the exposed areas of the mask are removed by a brief submersion in developer, leaving the desired pattern. (If a positive resist is used, the un-exposed parts are removed instead.)

(d) Inspection: the mask is then inspected in an optical microscope (and occasionally an SEM) for obvious flaws, before continuing.

3. Pattern transfer:

(a) Dry etching: the masked sample is placed in a Reactive Ion Etcher which etches the exposed areas into the sample material.

(b) Mask stripping: whatever polymer remains of the mask is now removed in much the same way as the first cleaning step.

(c) Inspection: samples are inspected optically and with electron microscopy before the devices are tested.

Most of these steps are fairly straightforward both in purpose and execution. The most involved steps by far are the Electron Beam Lithography (EBL) and Reactive Ion Etching.

Apart from the EBL, every step is executed in a cleanroom.

3.2.1 Cleaning procedures and resists.

In order for the lithography process to produce accurate patterns, the resist layer needs to be thin, uniform and smooth. This requires samples to be thoroughly cleaned beforehand, as any contaminants can interfere with the adhesion of the resist to the dielectric. The most common contaminants are either organic films, for example residues left by fingerprints or leftover resist, or granular particles such as dust. The general cleaning procedure adopted in our research group deals with these using a combination of organic solvents and ultrasonic agitation.

The primary cleaning step involves putting a sample into a beaker filled with acetone, which in turn is placed inside an ultrasonic bath. The combination of the alcohol that binds to organics and the cavitation in the liquid caused by the high frequency agitation scrubs most of the organics from the surface, as well as dislodging any relatively large particles. The sample is then rinsed shortly in Iso-Propanol which evaporates more cleanly from the sample surface on drying

with a pressurised nitrogen flow. For most samples this procedure is sufficient to deliver good results.

In some cases, more effort is required. A contaminant might be resistant to the acetone, the sample could be too fragile, as with, for example, membraned structures, or the sample itself could be affected by the Acetone in an adverse way. For these cases, several other options are available that can be combined as appropriate:

- For polymer resists a specialised mixture of alcohols that specifically attacks the polymer called 1165 can be used as a replacement for acetone. 1165 is commonly used as a first step if a significant amount of a previously applied resist mask is left on a sample, for example after etching.

- Stubborn organic contaminants can be aggressively attacked with an acid commonly known as piranha. This acid, a 2:1 mixture of sulfuric acid and hydrogen peroxide, is extremely aggressive towards anything organic, essentially chemically burning the sample surface. It also slightly oxidises the surface of most dielectrics in the process.

- Repeated application and pealing off of a specialised low adhesion tape with no adhesive transfer to the surface can be surprisingly effective in cases where the sample is negatively effected by alcohols. This mostly deals with dust and is less effective with organics.

- In rare cases, it may be necessary to scour a sample surface clean using a O_2 plasma etch. This procedure effectively burns off any organic matter and has the advantage of being somewhat directional, making it suitable for some samples that cannot be treated with piranha.

Once a sample is clean, the resist mask can be applied, which is generally accomplished through a process called spin coating. The sample is placed on a sample holder and held in place through suction. A liquid solution of polymer dissolved in a solvent is dropcast onto the sample. The holder is then rotated around its centre axis at high speeds, generally several thousand rpm (rotations per minute). The balance between forces keeping the solution in place, such as surface tension, Van der Waal's forces and gravity, balances the centrifugal force created by the spinning, thus producing a thin and uniform layer over the sample. The thickness is controlled by viscosity and spin speed. To a lesser extend, factors such as surface hydrophilicity, humidity, temperature, spin duration, etc. influence the thickness as well, although these more obviously influence the uniformity of the layer. The same factors determine the size of the edge beads, which are

thicker areas of resist around the edges of a sample due to the increased surface tension.

After spinning, the sample is baked on a hotplate (an oven can also be used, but generally takes longer to achieve very similar results) at temperatures above the solvent evaporation point, yet not high enough that the polymer make-up is influenced. In this way the solvent is mostly boiled off, leaving an amorphous polymer layer of resists that can be used in the further lithography steps.

Resists come in many varieties. Their primary functionality is to change solubility in a developer when exposed to an external form of energy, generally in the form of light or charged particle beams. In this way a pattern is formed. A commonly known example is the analogue film used in old cameras that changes colour when exposed to light producing a negative of an image projected onto it. This analogy is the origin of a naming convention for resists, where those with the exposed parts becoming more susceptible to development being called positive resists, while those for which the opposite is true are called negative resists. The resists used in this thesis are all positive in nature.

The resists used in semiconductor fabrication are generally polymers. They are designed to be susceptible to specific high energy (UV) wavelengths of light (photoresists) or high energy electron beams (ebeam-resists.) In the work described here, the patterns are created almost exclusively using electron beam lithography, but the further process is completely analogous. After exposure of a positive resist the exposed areas become more susceptible to be dissolved by the developer. The increased solubility allows a short submersion into this chemical to remove the exposed areas, whilst leaving the rest intact. The ratio between the solubility of the pre- and post-exposure resist to the developer is one of the primary characteristics that distinguishes high resolution resists from other polymers, and is critical if very small or high aspect ratio features are needed.

The second feature that sets lithography resists apart is its resistance to etchants. A successful mask used for pattern transfer needs to resist the etching procedure for the time needed to etch into the underlying layer to the required depth. Depending on the etchant used, e.g. chemical wet etches or plasma dry etches, different polymers are used. For the partially directional Reactive Ion Etching technique used mostly in this thesis, the mask resistance can be controlled to an extent by increasing the thickness of the resist mask. For high accuracy or small features, this approach does not work very well, either because the resist's structural integrity cannot handle the high aspect ratio between height and feature width or because the chemical part of the etching erodes the feature too quickly from the sides.

The resists used for the structures in this thesis are all proprietary polymers. The early work on direct phase metasurfaces (chapter 4) was done using ZEP 520A[79], later work mostly with Allresist AR-P 6200 with varying solid-contents[80] to control the viscosity depending on the required layer thickness. These resists are marketed as high resolution ebeam resists with a high etch resistance to common fluorine based dry etches. Specific recipes vary from sample to sample and I will leave those for the relevant chapters.

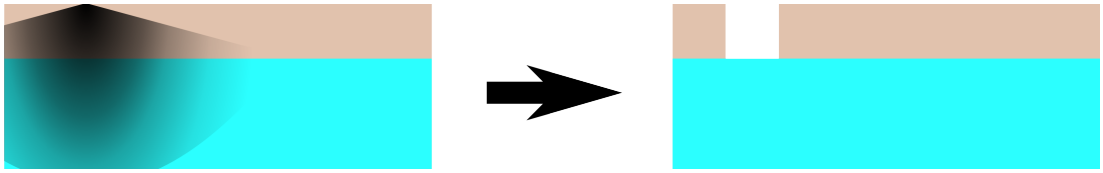
3.2.2 Electron Beam Lithography.

The process of creating a pattern in a surface is commonly known as lithography. Depending on the size and type of surface, many techniques fall under this heading, from book printing to deep UV photolithography as used in transistor fabrication. For research in dielectric microstructures, where a relatively low quantity but high variety of structures is desired, the technique of choice is Electron Beam Lithography (EBL), a system which uses a focused high energy beam of electrons to expose a susceptible masking layer in a predefined pattern. With EBL, it is possible to create nanometre scale designs with high accuracy. Moreover, in contrast to photolithography which projects a pattern onto a resist, EBL traces out the design as if drawing it with a pen, making arbitrary patterns that change from run to run easier to accomplish.

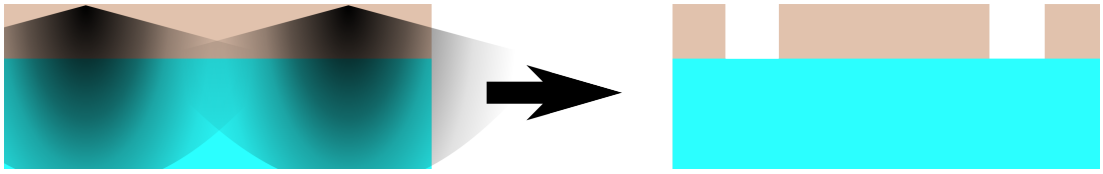
In an EBL system, a focused beam of high energy electrons is pointed at a substrate on a sample holder or stage that is located in a vacuum chamber. The beam can be blanked and accurately deflected within a set area, called the write field, allowing it to trace out preset patterns. To make larger designs, the stage is moved such that the region accessible by the electron beam is adjacent to the previous region. For accurate long range patterns, it is therefore crucial to have a highly precise stage positioning system as well as beam deflection control.

The lithography system used by our research group and for all structures discussed in this thesis is a 50kV Raith Voyager. This system uses electrostatic beam deflection for faster writes than the more common electromagnetic systems. It has a $500\mu\text{m} \times 500\mu\text{m}$ write field area, which is big enough for most of the designs shown in my work. The software divides the write field into a square grid of 1nm^2 pixels, which is the limiting factor for the accuracy of the beam positioning. Most importantly, this means that relative positions in structures, such as the periods of gratings, tend to be accurate to about a nanometre.

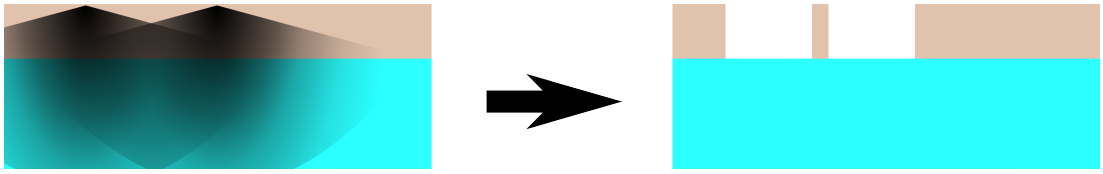
The resolution limit of individual features is much less accurate, however. The feature size depends on an interplay between the direct electron beam, as well



(a) Isolated exposure.



(b) Sufficiently separated exposure points.



(c) Exposures too close together.

Figure 3.3 – Illustration of densities in electron beam lithography, and the resulting structures in the developed resists. Blue represents a substrate, brown the resist and black the electron density. (a) shows the result from exposing an isolated feature, with the electron density from the exposure on the left and the resulting pattern in the resist after developing on the right. (b) shows how two features spaced far enough apart produce two copies of that same feature, while (c) shows how bringing the exposure areas too close together increases the electron density between them and results in larger features.

as forward and backward scattered electrons, with the resist and the influence this has on the polymer when developing. In practice, only isolated spots or lines in very thin films of resists can achieve few nanometre sizes. For thicker resists to expose the full depth of the mask requires exposure to a degree that the scattered electrons expose the material isotropically around it. For similar reasons, exposing multiple areas close together leads to the sum of the scattered electrons of both areas influencing the area in between (see figure 3.3.) The latter is called the proximity error.

To account for these effects, a dose test is initially performed. A dose test is a trial run in which a set of easily identifiable patterns of similar size to the eventual structure is designed. Generally, the exposed areas are reduced in size by a percentage factor to account for the spreading of the scattered electrons, so these trial samples usually consist of an array of repeats of the same structure, with a varying dose on one axis and size reduction factor on the other. With a fixed development process, the results can be inspected optically and by Scanning Electron Microscope (see the next section, 3.3) to find the combination that most closely matches the desired structure. These values are then used to write the

actual grating structures for further testing.

This method works fairly well for any structure with a uniform local density over the area, such as simple periodic structures or isolated features. In more complicated structures, such as metasurfaces, the density can change significantly across the structure. The local density also changes as one gets closer to the edge of an exposed area. In both cases, if accuracy is required everywhere, a single size reduction and dose combination usually does not suffice and one needs to use proximity error correction, which takes the local density into account to adjust the dose by location[81]. For the best results this should be applied to all the structures in this thesis, but generally we are only interested in a proof of concept for which this turns out to be unnecessary in most cases.

Due to the constant stream of high energy electrons in the write field, samples need to be sufficiently conducting to ensure charge dissipation and for an accurate pattern to be formed. If the sample is an insulator, the electrons do not flow away from the exposed areas fast enough and a negative charge builds up on the sample surface that deflects the beam away from its intended target. For silicon this is not generally an issue, but for other materials such as glass it is necessary to deposit a thin (10-20nm) layer of aluminium on top of the resist. This is done with a thermal evaporator system, effectively a vacuum chamber in which a metal is evaporated through resistive heating with the sample held above it. A very thin layer solves the charge build up issue by allowing low energy electrons to escape while causing very little disruption to the high energy electrons in the writing beam.

The metal does prevent the developer from interacting with the resist, however, so it needs to be removed before development. Most samples can be submerged safely in a beaker of phosphoric acid for a few minutes without doing any damage to the important parts of the chip or the resist, but easily dissolving the Aluminium charge dissipation layer. Following the wet etch, the mask can then be developed as normal.

3.2.3 Reactive Ion Etching.

To transfer the pattern into the dielectric substrate, I have used Reactive Ion Etching (RIE.) In RIE, a chemically reactive plasma is accelerated towards the sample surface. The combination of chemical bonding and physical ion bombardment then removes material from the areas exposed to the plasma. With the right parameters this process can produce very accurate patterns with smooth vertical side-walls. In contrast, most wet etching techniques will either follow the crystal axes, or etch isotropically in the case of an amorphous substrate.

In its simplest form, RIE consists of a parallel plate capacitor and RF coils in a vacuum chamber. Gases are introduced into the chamber at well defined rates and ionised by the RF field. The positively charged ions are then accelerated towards the sample by the electric field between the capacitor plates.

The physical impact of the ions determines whether the etching is anisotropic. The choice of gas also gives control over the chemical aspect of the etch. The reactive ions that make up the plasma can bind to the substrate surface, creating a surface layer of compound materials that are less tightly bound to the rest of the substrate and therefore more susceptible to the physical etching. Other ions can bind to the surface of the resist and make it instead more resistant to further etching, allowing it to retain its shape longer. The interplay of etchant gases and pressure, substrate and resist materials, as well as the physical ionisation energy and acceleration voltages is what determines the etching speed, anisotropy and uniformity.

The RIE process in our cleanroom is based on fluorine chemistry optimised for silicon etching. The most used gas recipe is a combination of fluoroform (CHF_3) and sulfur hexafluoride (SF_6). The highly reactive fluorine ions do the majority of the etching, while the hydrogen ions bond to the resist sidewalls to form a passivation layer. Other materials used etch with similar chemistries, but the exact recipe is different for each material. The other controlling parameters, pressure, RF power and acceleration voltage, all need to be adjusted to produce the best results. The specific recipes used for each structure will be given in the relevant chapters.

3.3 Characterisation.

The characterisation of photonic devices falls roughly in two categories: physical inspection and optical response characterisation. Of the former, the most relevant technique is Scanning Electron Microscopy (SEM). For the latter, the exact experimental methods vary depending on the specific structure, but the fundamental principle is very similar. As characterisation has been a relatively minor part of my work this section will be slightly shorter than the previous two.

3.3.1 Scanning Electron Microscopy.

Optical microscopy is limited to resolving features around the size of the wavelength of the light used, so a visible light microscope can see features down to about 500nm. The structures discussed in this thesis consist of features of approximately that size, which means that while the structures as a whole can generally easily be found, the individual grating ridges are only just visible, if at

all. In order to inspect the quality and measure the sizes of fabricated metasurfaces to the required level we need to go several orders of magnitude smaller in resolution.

Scanning Electron Microscopy (SEM) has the capacity to resolve features down to the nanometre level. Very similar to Electron Beam Lithography (see 3.2.2) a SEM uses a beam of electrons that is scanned across a sample. (In fact, many older EBL systems are modified SEMs.) This beam interacts with a sample surface in a variety of ways, many of which can be used to gather information about it. The most commonly used method is to detect secondary electrons that get re-emitted from surface atoms ionised by the electron beam. This method gives a very detailed measurement of the surface topology as the ionisation of atoms decreases very rapidly into the material. For the best surface resolution relatively low energy electron beams are used to further limit the penetration depth.

As with EBLs, the surface needs to be sufficiently conducting to prevent charge build up. While there are advanced SEM techniques that mitigate this problem ("gentle beam" for example applies a negative bias to the sample stage to reduce the problem), the most common way to make SEM useful for nonconducting samples is to sputter a thin (several nms) layer of metal (usually PtPd) onto the surface. At this thickness the metal coating is sufficient to provide the necessary conductivity while almost exactly following the contours of the underlying substrate.

SEMs are used extensively in this thesis to check fabrication results. They are the only accurate method at our disposal to determine the geometrical results of the lithography process. Both developed resists and etched dielectrics have been inspected, the former of which always with a charge dissipation layer while the latter depends on the specific material. The most common goals of this inspection are to measure period and duty cycle of fabricated gratings, mainly in dose tests; to inspect finished structures for unexpected flaws such as broken ridges or distorted features; to inspect the surface roughness of an etch; or when used on cleaved samples at angle to check the etch or development depth and profile.

3.3.2 Optical response.

To experimentally verify the function of fabricated metasurfaces, it is of course necessary to use them in the intended way and accurately characterise the results. In general, this means illuminating the metasurfaces at normal incidence with a polarised, collimated and mono-chromatic beam, then imaging the reflected or

transmitted beam. While the various metasurface types require slightly different combinations of components to properly evaluate beam conversion efficiencies, the general setup used is much the same in all cases.

For all metasurfaces we need monochromatic, collimated and normal incidence light, with the aim to image either the reflected or the transmitted light. The scheme used to accomplish this is a microscope setup, shown in figure 3.4. In the case of transmission measurements, a laser is focused into the backfocal plane of an objective, thus collimating the beam onto the sample. On the other side of the sample is a second objective used for imaging. The sample can be moved in the z direction, effectively scanning the volume above the sample to build up a beam profile while keeping incidence the same (assuming proper collimation of the incident beam.) For reflection measurements, a beam splitter is inserted in the incident light path before the first objective. This single objective is then used to both collimate the incidence light and to image the reflected beam.

In either case the incident light needs to be polarised, therefore a polariser is placed between source and objective. Generally monochromatic sources are used, so no wavelength filtering is necessary. Depending on the type of metasurface and the beam generated, other elements can be inserted before and after the sample, from quarter wave plates to analysing polarisers. The specific combination of these optional elements used for a measurement will be specified in the relevant chapters.

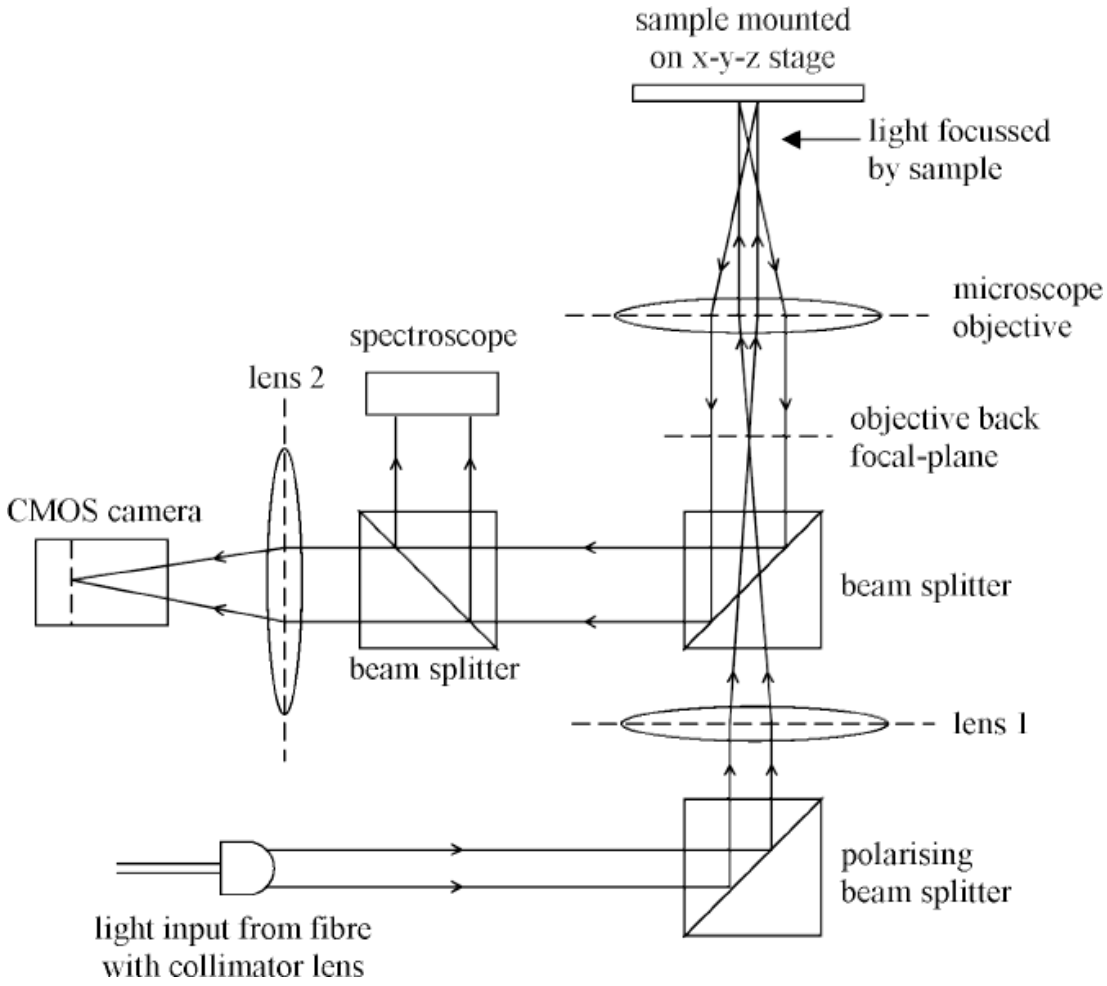


Figure 3.4 – Schematic diagram of the optical setup used to characterise reflective metasurfaces. For transmission measurements the imaging arm is moved to the other side of the sample. (Image courtesy of Andrei Ruskuc.)

4 Direct phase control metasurfaces.

The most straightforward type of metasurface I have studied is based on high contrast grating resonances[21, 22, 28, 69]. The essence of this type of metasurface is the ability to locally tune the phase response of the grating by adiabatically changing the unit cell of a grating. This means that light coupling into these modes will experience a slightly different phase delay depending on the local geometry of the grating. Effectively, it takes the propagation length differences that make up many wavelengths in a regular optical component and folds it into a wavelength-scale volume.

The theory necessary to understand these metasurfaces has already been discussed in chapter 2. The first section of this chapter will expand on the most relevant parts and how these are used to get to a design, as well as expanding on the fundamental methods from chapter 3. The latter section shows results of both simulation tests and experimental characterisation.

A significant portion of the fabrication and testing discussed in this chapter is work by Dr. Annett Fischer[82] and to a lesser extend Andrei Ruskuc. They exclusively worked on focusing mirror gratings, but their groundwork was crucial for the further work done on the other types of metasurfaces presented here as well.

Sections 4.1 and 4.2 are adjusted from work we have previously published[30, 41].

4.1 Design.

A metasurface in its simplest form is a sub-wavelength thick structure that changes an input wavefront into a desired output wavefront[10]. In other words, at any point on the surface, a pre-defined phase shift is added to an incident wave. The metasurfaces in this chapter achieve this by tuning local grating resonances such that the phase difference between input and output wave changes in a well controlled fashion.

The design of such metasurfaces requires a basis set of grating parameters that maps onto a phase range of at least 0 to 2π. Once such a set is found, it can be mapped directly onto a desired wavefront. The actual resulting metasurface is a discrete approximation to this ideal map discretised by the size of the unit cells, which in this case is the period. Given the cyclic nature of phases, the exact phase delays are unimportant, as long as the difference between maximum and minimum covers the full circle.

This section details the steps involved in designing and fabricating metasurfaces based on this principle.

4.1.1 Materials and thickness.

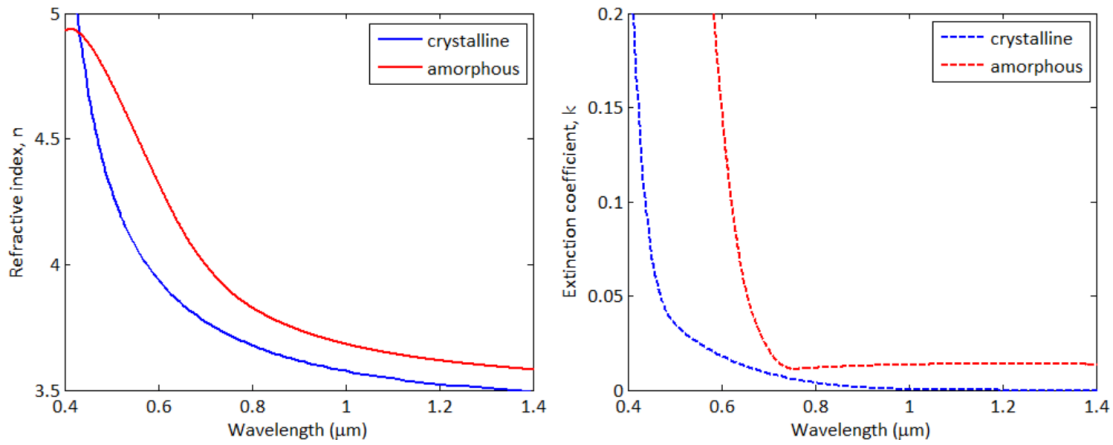
The first step in any metasurface design is to find a suitable material combination. This choice, of course, depends on the purpose and constraints the metasurface needs to fulfil, but ultimately it boils down to three primary factors: operating wavelength, dielectric environment and whether the metasurface is to be used in transmission or reflection.

Both the real and imaginary parts of the refractive index of a material are important for the design of a well functioning metasurface. The real part, n, greatly influences the available resonant modes (see chapter 2.3) and by extension the possibility of finding a good set of grating parameters to cover the full phase range. It can be somewhat compensated by adjusting the thickness of the grating layer, but this is limited by the fabrication techniques. More accurately, the difference between the refractive index of the ridges and the highest index among surrounding materials needs to be high enough to support at least one Guided Mode Resonance (GMR) in the sub-wavelength region[28, 35, 69].

The imaginary part of the refractive index, also known as the extinction coefficient k, effectively determines the absorptivity of a material[83, 84]. In other words, it puts an upper limit on the efficiency that can be achieved. The absorptivity of a material is described by $\alpha = \frac{4\pi k}{\lambda}$, with total absorption in a layer being given by $A = 1 - e^{-\alpha\Delta}$ where Δ is the optical path length. Through $T + R = 1 - A = e^{-\alpha\Delta}$, and realising that Δ and λ are usually in the same order of magnitude in our structures, this reduces to $T + R \approx e^{-4\pi k}$. Therefore even a k of 0.01 already limits the efficiency of the structure to 90%!

In general it is therefore desirable to have a material with an n as high as possible and a k as low as possible. Both of these values are generally decreasing with wavelength in dielectrics[85]. An example of this can be seen in figure 4.1, where the refractive index and extinction coefficient for crystalline and amorphous silicon are shown. The extinction coefficient in particular tends to drop off as $k \propto \frac{1}{|\lambda - \lambda_R|}$, where λ_R is a material resonance wavelength, usually in the UV. As a result, there tends to be a fairly distinct range of wavelengths below which a material is no longer useful. Problematically, dielectrics with a higher real part of the refractive index also tend to have a higher value of λ_R.

This is partially what has led to the rise of silicon as a dielectric platform for nanophotonics in general with wavelengths in the near IR, or above approximately 850nm, where it is cheap and relatively easy to structure while still being transparent with a high refractive index of ~ 3.5 or higher, as shown in figure



(a) Refractive index n **(b)** Extinction coefficient k

Figure 4.1 – Real and imaginary parts of the refractive index of crystalline[85] and typical amorphous silicon (measured by Dr. Yue Wang).

4.1. The differences between amorphous and crystalline silicon are mostly due to impurities embedded in the amorphous material during fabrication, which tend to increase both the refractive index and the extinction coefficient. The high refractive index and relatively low refractive index are the main reason why the work in this chapter has been done entirely in silicon. The refractive index of silicon is just about high enough to get a reasonable range of grating parameters between the sub-wavelength criteria on a substrate and the deep-subwavelength region.

If the refractive index is much lower, it becomes significantly harder to find an efficient full phase set of grating parameters while still on a substrate. Materials such as titanium dioxide (TiO_2, $n \approx 2.7$) or silicon nitride (Si_3N_4, $n \approx 2.1$)[85] that are mostly transparent in the visible wavelength range and can be fabricated with the same techniques as silicon might allow for reasonable solutions at lower wavelengths, but only in a membraned fashion which highly limits the feasability.

Even using silicon on glass, changing just the cladding material from air to water, a refractive index increase of just 0.33, turns out to make it far more challenging to achieve the full phase range with a high efficiency.

Our choice of materials is silicon on a glass substrate. Using silicon necessitates a near IR wavelength as design wavelength, but otherwise leaves us free to explore the limits of direct phase modulation metasurfaces based on dielectric gratings. Once this is set, the next step in the design process is to determine the thickness of the silicon that can be used.

There are several considerations that need to be made for this choice, both of a practical and of a fundamental nature. The former limits the aspect ratios

between groove or ridge width and etch depth that can be achieved in silicon. While a 1 : 2 ratio between these is relatively easy, a 1 : 5 ratio is very difficult to do properly. If the aspect ratio is too high and the ridges get too thin for their height the resist tends to fall over before the etch is completed, leading to asymmetric and chaotic structures. Conversely, if the grooves are too narrow, the etching process is slowed down and therefore a thicker resist is required. The former can be slightly compensated for by reducing resist thickness, the latter by increasing it, but the margins are fairly tight. The most favourable aspect ratio is found with a duty cycle of 0.5, and the period is limited by the wavelength in the output region, so in short, as a rule of thumb:

$$t \lesssim 5 \cdot \min\left(DC_{min}, 1 - DC_{max}\right) \cdot \Lambda < 2.5\frac{\lambda_0}{n_0}, \tag{4.1}$$

where n_0 is the refractive index of the output region. That upper limit applies for every parameter being ideal, and realistically anything above approximately $t \approx \frac{\lambda_0}{2n_0}$ can be considered ambitious. Most of the work in this chapter is done with wavelengths of approximately $1\,\mu$m, in either air ($n_0 = 1$) or water ($n_0 = 1.33$), so the thicknesses we want to work with are below $\sim 500\,$nm and $\sim 380\,$nm, respectively.

The lower limit comes from a more fundamental consideration. The grating needs to be thick enough to support at least the lowest order GMR in order to achieve an efficient 2π phase range. Rewriting equation 2.30 to bring the thickness to the left, and plugging in the sub-wavelength condition ($\Lambda < \frac{\lambda_0}{n_0}$)for the period to simplify gives

$$t = \frac{m}{2}\frac{1}{\sqrt{\left(\frac{\mathrm{n_{eff}}}{\lambda_0}\right)^2 - \left(\frac{q}{\Lambda}\right)^2}} > \frac{m\lambda_0}{2}\frac{1}{\sqrt{\mathrm{n_{eff}}^2 - \left(n_0 q\right)^2}}. \tag{4.2}$$

This goes up with m as well as q, with $m \geq 2$ and $q \geq 1$ for a metasurface to be possible, so this simplifies further to

$$t > \frac{\lambda_0}{\sqrt{\mathrm{n_{eff}}^2 - n_0^2}}. \tag{4.3}$$

The lower limit goes up with decreasing effective index, exceeding the practical upper limit $t < \frac{\lambda_0}{2n_0}$ when $n_{eff} < \sqrt{5}n_0$. Moreover, this limit only needs that mode to exist for the maximum value of n_{eff}, so we can set the duty cycle to 1 and replace n_{eff} with n_{high}, or the refractive index of the ridge. Filling this in with, for example, silicon in air at a wavelength of $1\,\mu$m gives a minimum thickness of $\sim 300\,$nm. Trying in Si_3N_4 on the other hand would require a thickness of at

least 540 nm, which is greater than the 500 nm upper limit found before, and this in fact turns out to be very difficult.

Both "limits" are only guidelines, of course. The upper limit can, if necessary, be increased because the possible aspect ratios can be extended by finetuning the fabrication. Moreover, they are naturally different from one material to the next. The lower limit depends on the ill-defined effective refractive index and therefore the thickness limit has an equally large margin of error. Still, it is clear from the above that most suitable thicknesses will fall between $\lambda_0/10n_0$ and $\lambda_0/2n_0$.

An example of the thickness limits can be seen in figure 4.2. The figure shows the best average reflectivities possible for a 2π phase path for an amorphous silicon metasurface with an illumination wavelength of 980 nm. A clear drop in efficiency can be seen at the lower thickness end. The insets show the increased aspect ratios necessary at the higher thicknesses that make them difficult to fabricate.

4.1.2 Pathfinding.

The limits on the thickness, refractive index and period only indicate the region outside of which searching is not useful. They do not, however, state that within the limits a good basis set for a metasurface can be found. It still depends on a complicated interplay of the different modes that vary with all the different parameters as discussed in chapter 2.3.

In order to find such a basis set, we are forced to resort to numerical simulations using Rigorous Coupled Wave Analysis (RCWA, see section 3.1.1.) Using RCWA, we can simulate the grating response for a large number of different grating parameters and search the results for a "path" through them that connects 0 to 2π in phase while retaining a high reflection or transmission throughout.

The challenge in finding a path is two-fold: first, given a map of the grating response as a function of period and duty cycle, find the most suitable path to base a metasurface on. Second, find the fixed grating parameters of wavelength, refractive index, thickness and polarisation that lead to the best path.

Finding a path through a 2D set of parameters is a problem that occurs in many disciplines. There are a multitude of algorithms available for very similar problems[86, 87]. However, most of them require knowledge of the starting and ending points. In the case of these metasurfaces these two points are unimportant, as long as the "length" of the path is 2π.

One solution to the problem is to use a variation on Dijkstra's path finding algorithm[88]. This algorithm takes a starting point in a map, then searches outward until it reaches its destination. By choosing the next point in parameter

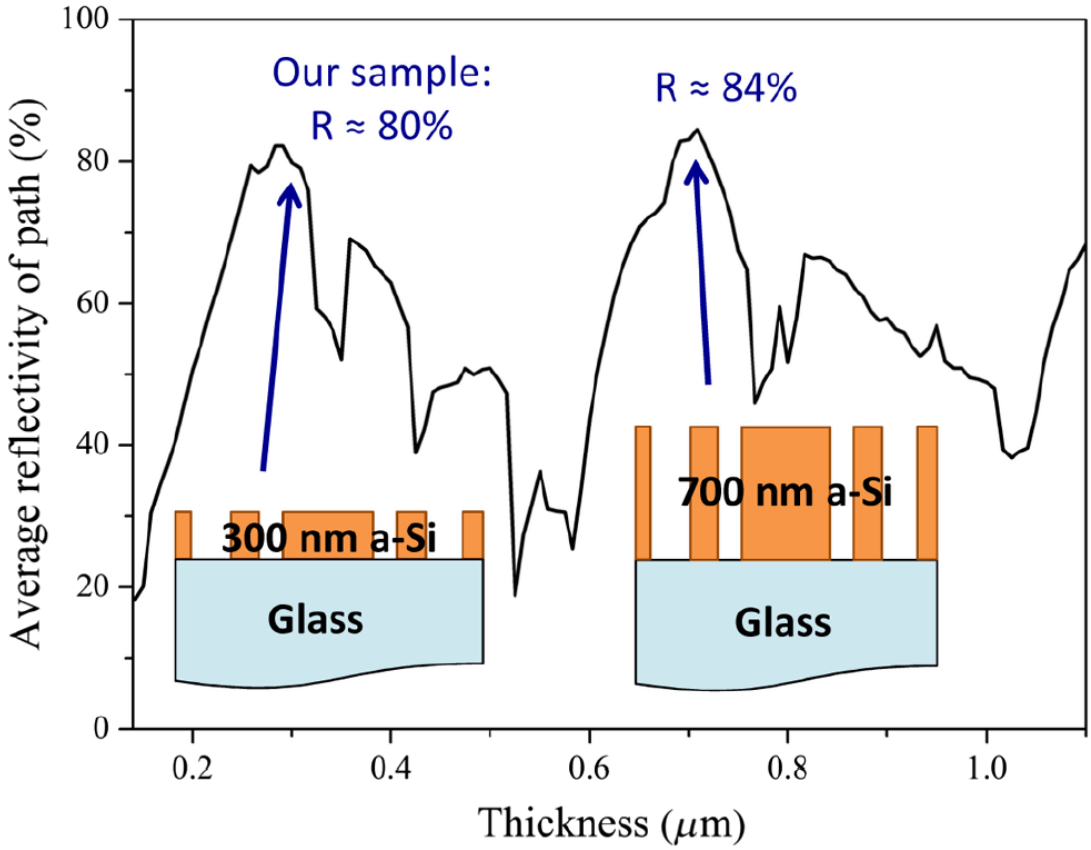


Figure 4.2 – Best average reflectivity of paths found for different thicknesses of amorphous silicon on glass by the modified Dijkstra algorithm discussed in section 4.1.2. The reflectivity shown is the average reflectivity of a set of grating parameters that together are used to design metasurfaces in amorphous silicon of the given thicknesses, thereby giving a measure of the expected efficiency that can be achieved at those thicknesses. This graph therefore gives an indication of the most promising thicknesses to work with when trying to design reflective amorphous silicon metasurfaces as discussed in this chapter, for a TM polarised illumination with 980 nm light. The insets show example diagrams of gratings based on the paths found near the two maxima, highlighting the increased aspect ratios necessary at higher thicknesses that make them less suitable for fabrication.[41, 82].

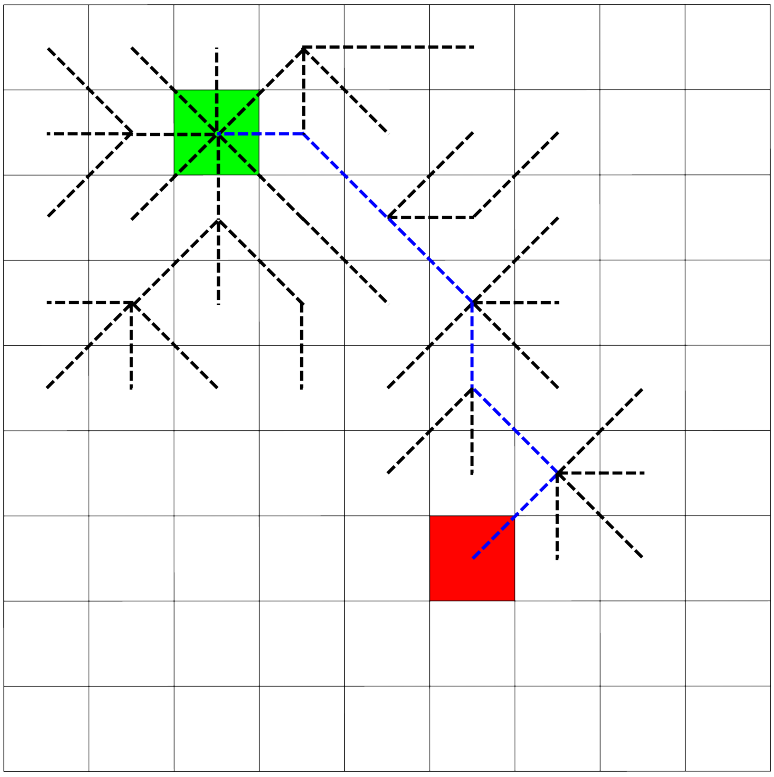


Figure 4.3 – Diagram to illustrate the pathfinding procedure. The process starts from a chosen source coordinate in a 2D matrix of results, indicated in green in the figure. It then explores the distances to each of the eight nearest neighbours in the matrix. From the closest one it repeats this process, without doubling back onto previously explored coordinates. This process is repeated from the end of shortest path explored up to that point, until a path reached the destination coordinate, shown in red. The successful path is shown in blue.

space to probe as the one that increases the total path length up to that point by the least, the first time the destination is probed is automatically guaranteed to be the "shortest" path. The procedure is illustrated in figure 4.3, and works as follows:

- Choose a starting coordinate.
- Store the distances to the eight neighbouring points in the array.
- Go to coordinate at the end of the shortest path currently stored.
- Add the distances to the seven neighbouring points to the current path length and store them.
- Remove any dead ends from the stored list.
- Repeat the previous three steps until a path of length 2π is found.

An example result of this process is shown in figure 4.4, where this pathfinding algorithm was ran on a map of reflectances and phases as calculated using RCWA.

The path found is drawn on the maps, and the resulting set of grating parameters is plotted against phase in the graph below them.

The trick lies in how to define the term "shortest", for which there is no best answer. A figure of merit needs to be assigned to each step that makes a trade off between factors such as reflectivity/transmittivity, gradient of phase, favourability for fabrication, etc. These factors in turn depend on the exact purpose of the metasurface. For example, a lens that is only used for collecting power is mostly concerned with the highest average amplitude. The figure of merit per step chosen for this was[30]

$$fom/step = \frac{|\phi_{new} - \phi_{old}|}{(R_{new} + R_{old})^g}, \tag{4.4}$$

with g a weighting of how important reflectivity is compared to the length in the space spanned by period and duty cycle, which correlates to fabrication difficulty. This was usually chosen to be 2, though if reflectivity is of paramount importance it can be set higher to increase its importance for determining the shortest path. A figure of merit used for imaging applications on the other hand would also require that the amplitude is similar for all positions, for example by including the standard deviation of reflectivities on each path.

Given a suitable figure of merit the algorithm still requires a starting point to be chosen. The obvious solution is to simply start from all possible points in a map. But this solution scales badly against the number of data points on the map.

An alternative method fixes the period and only allows for variation in duty cycle. This approach has merit, because the variation of period is effectively a change in the size of the unit cells. This in turn influences the apparent orientation of the grating ridge and leads to a small amount of polarisation conversion, as well as producing discontinuities in the final grating designs that form weak points during fabrication. If, in contrast, the period is kept fixed, it becomes feasible to compare all possible 0 to 2π phase ranges in a map and compute the figure of merit for the paths as a whole. This method is faster and more robust in fabrication, but limits the absolute efficiency of the path.

A straightforward continuation from the path finding method would be to run it for every one of a range of the fixed grating parameters of thickness and polarisation. While this is certainly possible, it is also extremely inefficient and time consuming.

Many of the maps produced by RCWA can be ruled out quite easily at a glance. For example, if there is no 2π range of phases in a map at all or the high

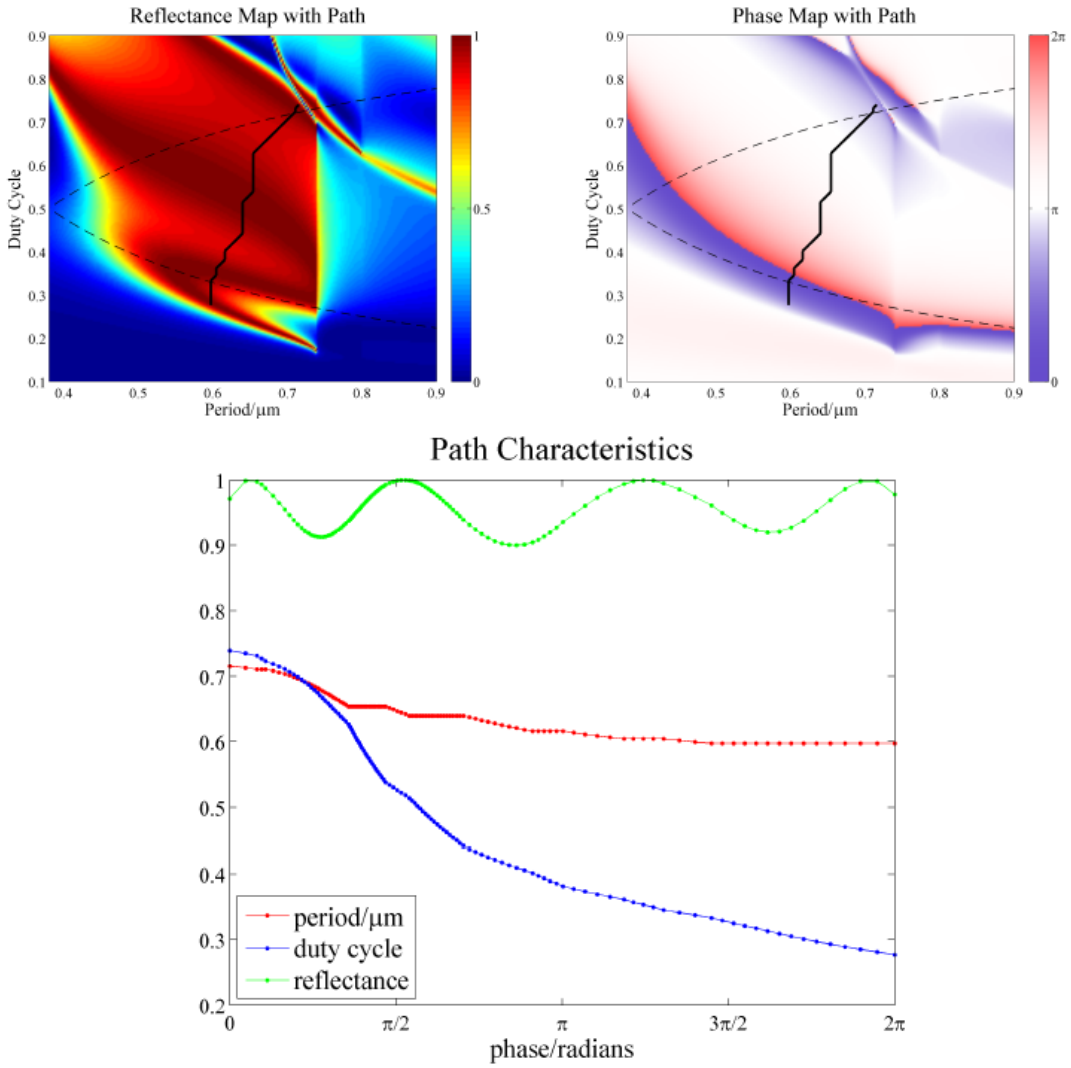


Figure 4.4 – Example of RCWA reflectivity amplitude (a) and phase (b) plots as a function of period and duty cycle. The dashed black lines indicate a 1:2 aspect ratio between narrowest feature size and etch depth, and the thick black line the path found by the Dijkstra algorithm. (c) shows that path as a function of phase, plotting its reflectivity, period and duty cycle.

amplitude regions are completely disconnected, it is unlikely that a good path will be found. Excluding these beforehand and only running the pathfinding on likely candidates is more practical.

The final result is a thickness, polarisation and set of periods and duty cycles that together form the basis set for the design of a metasurface.

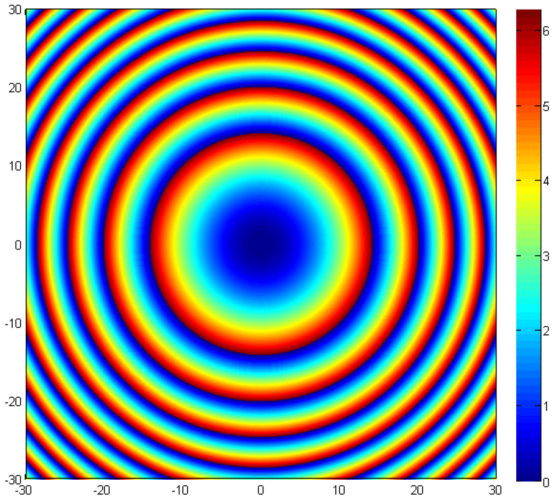
4.1.3 Phase profiles and grating designs.

Once a suitable basis set of grating parameters is found, the next step is to determine the phase profile that needs to be encoded into the metasurface. As discussed in chapter 2.2, this involves finding the difference between the phase profile of a beam incident onto the surface and then subtracting it from the desired phase profile of the output beam. Time reversal is still true, so the output phase profile can be found by propagating the resulting beam in the reverse direction using, for example, Huygen's principle (see chapter 2.1) and again determining its phase distribution where the metasurface intersects it.

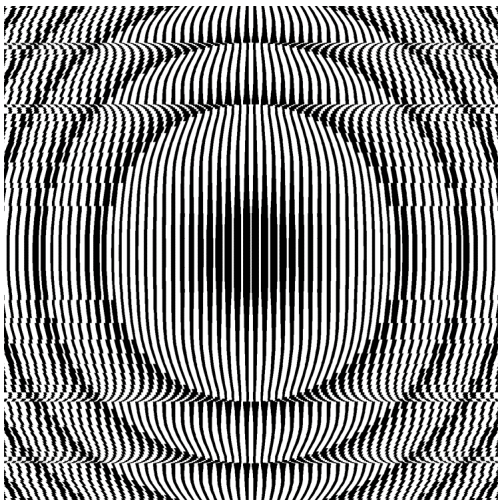
In principle this holds for any combination of phase profiles, limited only by the unit cell size and Nyquist's theorem. For simplicity, assuming a plane wave and normal incidence makes the phase profile on the incident side a constant, and therefore permits only encoding the desired phase profile onto the metasurface.

Crucially, a phase profile is independent of transmission or reflection. In other words, a parabolic mirror has the same phase profile as a lens. Moreover, while the phases of transmission and reflection are not equal and the resonance conditions are independent, they do both increase with optical path length and therefore tend to show similar behaviours along a path. This behaviour leads to a metasurface designed for reflection generally also creating a similar beam profile in transmission, albeit less intense.

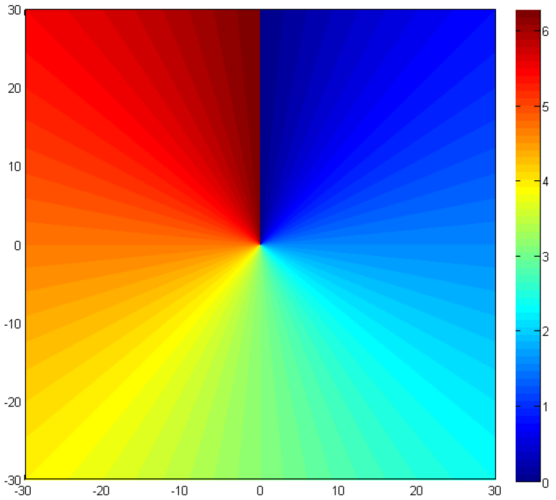
Two particular beam types tend to be interesting for exploring metasurface functionalities: lenses (or parabolic mirrors) and vortex beam phase plates. The former has a purely radial phase profile which is completely independent of the angle around the origin, while the latter creates a vortex beam with a purely azimuthal phase distribution that is independent of phase. Their orthogonal nature means they can also be combined into a single metasurface that combines both behaviours without loss of functionality. Phase profiles and example metasurfaces based on them are shown in figure 4.5. In the phase profiles (a), (c) and (e), the symmetry of the different types of beam is reflected very clearly, with (a) changing only radially, (c) only azimuthally and (e) in both those dimensions. This symmetry is still present, though less obvious, in the resulting metasurface designs of (b), (d) and (f).



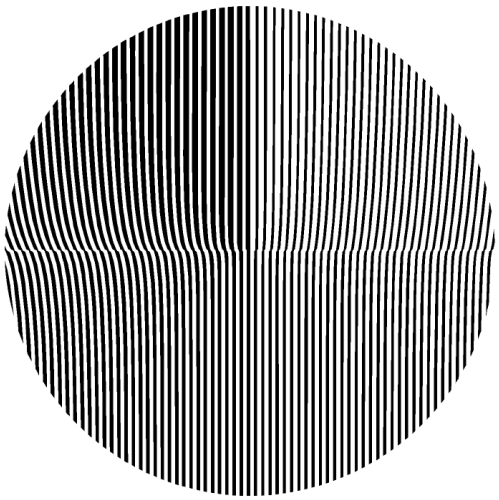
(a) Parabolic mirror phase profile



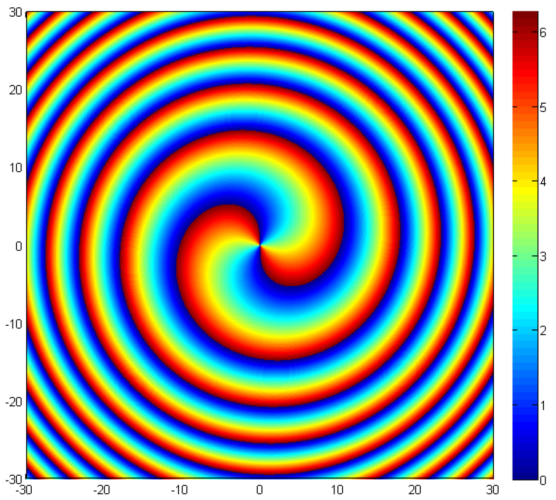
(b) Parabolic mirror grating design



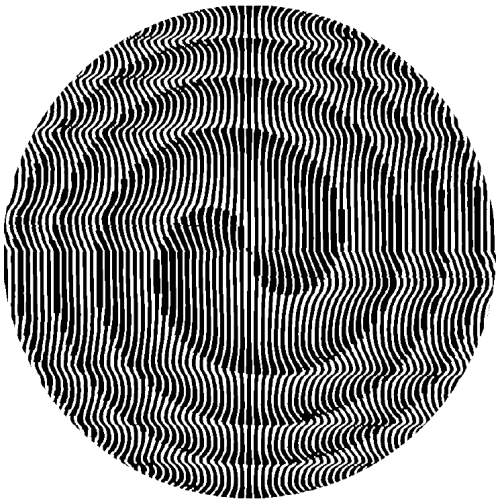
(c) Phase plate phase profile



(d) Phase plate grating design



(e) Combined phase profile



(f) Combined grating design

Figure 4.5 – On the left are shown the phase profiles of a focusing mirror (a), LG generating phase plate with $l = 1$ (c), and a combination of lens and phase plate with $l = 2$ (e). On the right corresponding grating designs.

To find the phase profile of a lens, Huygen's principle and trigonometry is sufficient, as we can visualise the phase propagation of light from the metasurface to the focal point of the lens as the hypotenuse of a triangle formed by the radius and the focal length. The only necessary information is therefore the focal length and the wavelength. At the focal point, assumed to be at $r = 0$ and $z = f$, or the focal length, every wave originating from the metasurface needs to be in phase to constructively interfere. The phase profile on the metasurface itself is therefore completely determined by the optical path length between that focal point and the structure.

This optical path length is given by $\Delta\Phi = k \cdot \Delta R$, where $k = \frac{2\pi}{\lambda}$ as usual and ΔR is the distance from the surface to the focal point. The latter can be found simply through application of Pythagoras's theorem, as $\Delta R = \sqrt{r^2 + f^2}$. As any phase front can be changed by an arbitrary constant without changing the results, making it such that the origin has a phase of zero is an allowed transformation that helps further down the line. The resulting phase profile can therefore be written as

$$\phi(r) = \Delta\Phi(r) - \Delta\Phi(0) = \frac{2\pi}{\lambda}\left(\sqrt{r^2 + f^2} - f\right). \tag{4.5}$$

This phase front is shown in figure 4.5a.

A similar consideration can be made for the phase plates that generate vortex beams. In this case, the consideration is to have complete destructive interference everywhere along the optical axis. This condition requires an equal number of points at any given r with a phase ϕ_0 and $\phi_0 + \pi$. The only way to retain a circularly symmetric interference pattern is then a continuous phase change with azimuthal angle such that going around 2π angle corresponds to an integer multiple of 2π phase, in other words a Laguerre-Gaussian (see chapter 2.1.) This makes the phase profile

$$\phi(\theta) = l\theta, \tag{4.6}$$

with l the azimuthal mode number of the vortex beam. This has to be true for any value of z, so also at the metasurface itself. This phase profile with $l = 1$ is shown in figure 4.5c.

With a phase profile and basis set of grating parameters determined, a lens or phase plate can be encoded into a metasurface. It is straightforward to assign the ideal duty cycle and period at each point in the plane, but that would not take into account the finite size of the unit cells. Along the grating ridges, this is not an issue as a continual change is in principle possible, but across the ridges the period fixes the size of the unit cells and by extension determines the accuracy with which a phase distribution can be reproduced.

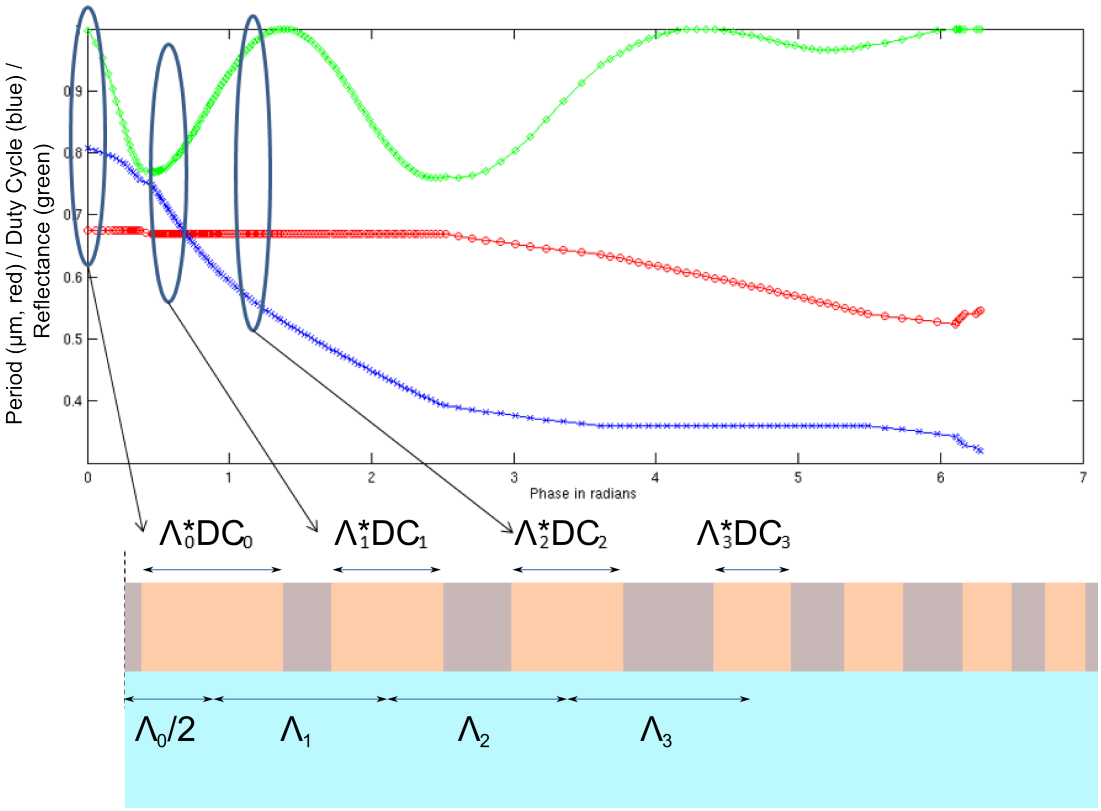


Figure 4.6 – Schematic diagram of the design process. A ridge is created with the grating parameters corresponding to a phase of 0. The period that leads to the phase best matching the phase profile is then found, and a ridge constructed with the corresponding duty cycle. This repeated until the desired grating size is reached.

To account for the finite size of the unit cells an iterative design process is used that starts at the origin and works outward across the grating[30]. The phase profiles are all adjusted to be zero at the origin, which allows for the largest radial range before a discontinuity in the basis set, i.e. a phase jump. The period corresponding to this 0 phase is used to determine the position of the next unit cell. The phase that most closely matches that next position is determined, and the corresponding period is used to determine the position of the third ridge. This procedure is repeated until a given radius is reached. See figure 4.6 for a schematic illustration of this process.

This procedure is repeated in discrete steps in the direction along the ridges, y, choosing a stepsize sufficiently small to reproduce the phase evolution accurately. Finally all the slices in y are connected to form full ridges.

Owing to the circular symmetry of the beams, the program that does this, written in MATLAB, starts from the origin and works outwards. Due to symmetry designing a quarter lens is sufficient.

The results for a parabolic mirror, reflecting phase plate with $l = 1$, and

Material	Etch depth	Resist	Spin speed 1	Bake 1 temp (time)	Spin speed 2	Bake 2 temp (time)
amorphous silicon	$\sim$ 300 nm-350 nm	ZEP 520A	3200rpm	180 deg (10min)	N/A	N/A
amorphous silicon	$\sim$ 280 nm	ARP 6200.13	1500rpm	110 deg (10min)	N/A	N/A
crystalline silicon	$\sim$ 380 nm	ARP 6200.13	1500rpm	100 deg (1min)	3000rpm	110 deg (10min)

Table 4.1 – Resist spinning recipes used in this chapter. All spins take 1 minute.

combination of the two with $l = 2$ are shown in figure 4.5, alongside colourmaps of their ideal phase profiles. These patterns are saved as polygons in file formats that can be read into either the design software of the lithography machine or into COMSOL for further modelling.

4.1.4 Fabrication specifics.

All metasurfaces of the type discussed in this chapter were designed and fabricated in a silicon-on-glass platform. In the early stages of the work, sputtered amorphous silicon was used, but in later stages we used crystalline silicon for more reproducible fabrication results. Amorphous silicon has a higher refractive index than its crystalline version, but also a higher extinction coefficient (see figure 4.1.) Moreover, the exact refractive index of the amorphous variant depends on the conditions during deposition, and the films are often granular in composition leading to flaws in the fabricated gratings.

Either way, the fabrication procedures are based on existing recipes present within the group for the fabrication of photonic crystals in Silicon-On-Insulator (SOI). The major difference in fabrication between the metasurfaces described in this work and photonic crystals is the thickness of the layer that needs to be etched. For the SOI commonly used in our group, this is standardised to 220 nm, while for metasurfaces it is usually quite a bit thicker. This requires longer etching times, and therefore thicker resist layers. For the exact resists and spinning recipes, see table 4.1. The RIE recipe used takes a flow 14.5sccm of CHF_3 and 12.5sccm of SF_6 at a pressure of 27mTorr. The DC voltage is set to approximately 185V. This recipe gives a bulk silicon etching speed of just over 3 nm s^{-1}.

The full fabrication procedure follows the same pattern as described in chapter 3.2.

The crystalline silicon samples are prepared for us by the research group of Li Juntao in Sun Yat-Sen University in Guangzhou, China. The crystalline silicon from an SOI donor chip with a 500 nm thick device layer is transferred onto a glass

substrate. Essentially, the device side of the donor chip is glued onto the new glass substrate with an optical glue. Most of the silicon handle of the SOI is then wet etched away, with the last few microns more slowly etched away using RIE. Finally, the buried oxide layer is removed through a short Hydrogen Fluoride (HF) wet etch. Occasionally the buried oxide is left in place for transport, in which case the final step is performed here.

At this point, a pristine, almost exactly 500 nm thick layer of crystalline silicon has been transferred onto a glass substrate. There is, however, an optical adhesion layer, of a few micrometres thickness, between the actual glass handle layer and the silicon. While this layer has the same refractive index as the glass and therefore doesn't affect the optical or photonic functionality of the devices, it does dissolve in some solvents. The layers are also very sensitive to ultrasonic agitation. Therefore, only short cleans using the standard chemicals are possible, and most of the cleaning is done through exfoliation using wafer tape and low power oxygen plasma etches using RIE.

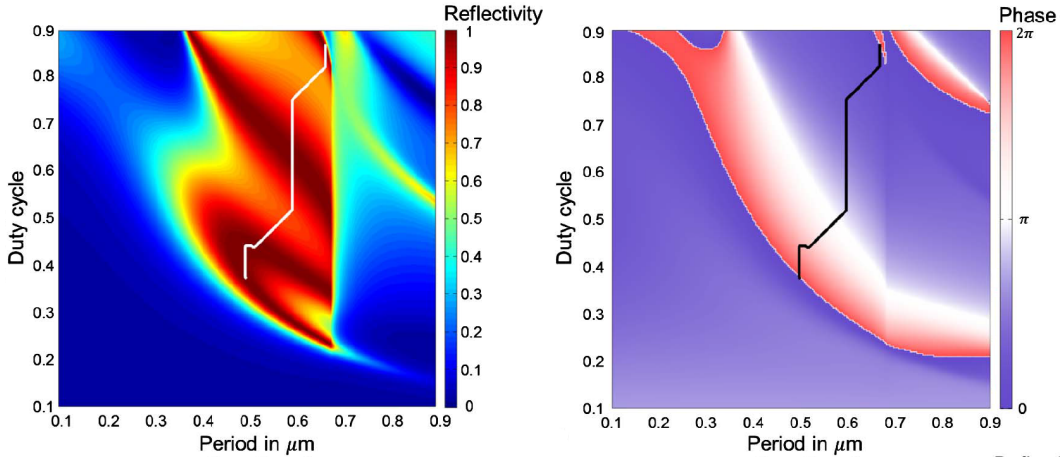
The amorphous silicon used in this work was deposited through Plasma-Enhanced Chemical Vapour Deposition (PECVD) directly onto glass microscope slides by collaborators at Tyndall National Institute Cork, Ireland. These layers are much more robust to standard fabrication techniques than the crystalline silicon samples, but the thickness and structural integrity of the silicon layer varies between each deposition run.

Most of the amorphous silicon samples were used with the thickness that we received them, as they were generally already relatively thin compared to the optimal thickness. The crystalline silicon samples were thinned down to appropriate thicknesses using the same RIE recipe as the etching of the structures themselves, but without mask. The thicknesses of all samples were measured using a home built reflectometry setup with the final metasurface designs being adjusted to the measured thickness.

After the thinning all subsequent steps are identical to the procedure described in chapter 3.2.

4.2 Results and discussion.

The following section has been divided into different metasurface functionalities. The first of these contains the bulk of the work, as the majority proof of concept and further exploration towards applications was done with reflective lens designs. In this subsection, extensive tests of efficiency and tolerances of direct



(a) Reflectivity **(b)** Phase of the reflectivity

Figure 4.7 – Paths used for the design of the parabolic mirrors used through most of section 4.2.1. (a) shows the reflectivity R as a function of duty cycle and period, (b) the corresponding phase. The lines (white in (a), black in (b)) through the maps indicate the optimal path found. The horizontal drop in intensity around a period of 676 nm is due to the subwavelength condition in the substrate (glass, $n \approx 1.45$) being reached. The maps were calculated for amorphous silicon on glass in air, with a thickness of 325 nm at a wavelength of 980 nm and TM polarisation.

phase metasurfaces were performed both through modelling and experiment.

The latter two sections include designs and modelling results on other possible beam shaping applications, including vortex beams.

4.2.1 Parabolic mirrors and lenses.

The original purpose of our metasurface work was to create flat parabolic mirrors. Earlier work had established cylindrical grating mirrors where the grating parameters were only varied in the direction of periodicity[21, 22, 28]. The goal here was to extend that to a three dimensional focusing system based on the same principles. Only later did we extend the concept past focusing mirrors to generalised metasurfaces.

These parabolic mirrors, using the phase profile given in equation 4.5, were intended to have a high collection efficiency in the focal point. The figure of merit for the pathfinding was therefore set as the highest average reflectivity. The gratings were designed, modelled and fabricated for a wavelength of 980 nm in 325 nm ±10 nm thick amorphous silicon with TM polarised incident light. (The fabrication and experimental testing of these parabolic mirrors was performed by Dr. Annett Fischer[82].) The RCWA maps and corresponding paths for these gratings can be seen in figure 4.7.

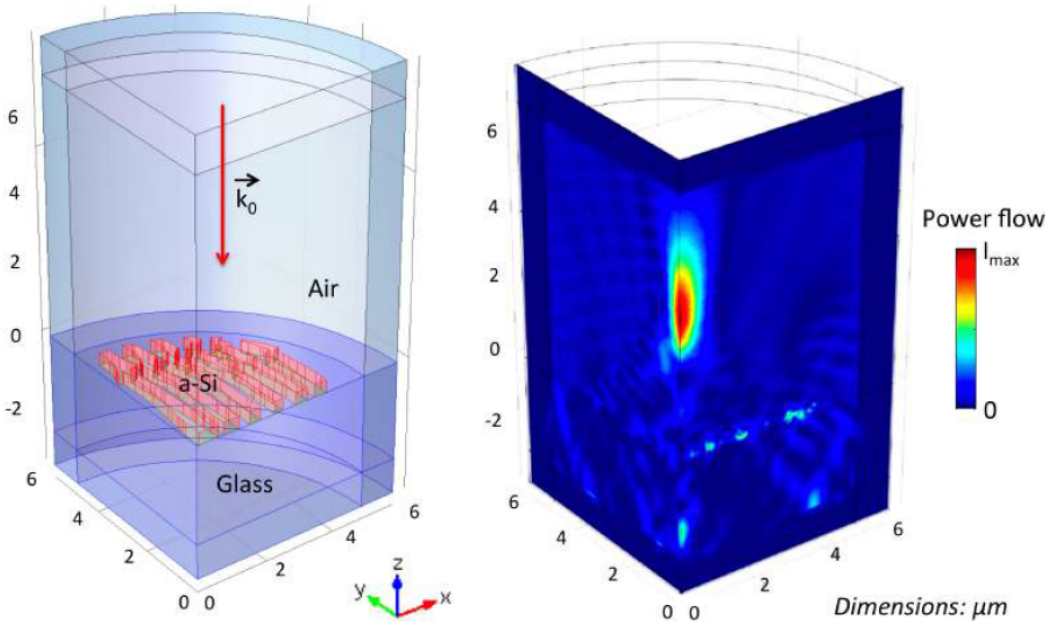


Figure 4.8 – 3D simulation of a reflective focusing grating using COMSOL Multiphysics. (a) shows the geometry used in the simulation, with (b) the resulting power flow. The grating used for the model is based on the paths shown in figure 4.7 for a focal length of 5 μm.

3D model

Proof of concept of the working of gratings based on these paths was performed using a 3D COMSOL model (see figure 4.8) as described in chapter 3.1.2. Our computational resources restricted the calculation domain to a much smaller than optimal design, with only a 5 μm focal length and radius. This implies a spatial extend of only a handful of wavelengths and, more importantly, periods. In fact, as can be seen in the figure, only 12 grating periods were used. This size limited the usefulness for an extensive exploration of the grating physics. Nevertheless, the model clearly displayed a focal spot at the intended location which collected the majority of the light incident on the grating region.

For further testing, and allowing the simulation of larger gratings, 2D simulations were employed. A 2D model sacrifices the effects of the grating modulation in the third dimension, but should give a good approximation of how the actual gratings respond to various situations.

2D model

Most reflective systems in this section use the paths shown in figure 4.7, and are therefore based on 325 nm thick amorphous silicon on glass in air, with 980 nm TM polarised incident plane waves. The simulations are all performed with incident light from the top, or positive y direction, in the fashion described in section 3.1.2. The gratings are all designed according to equation 4.5, but the

focal length and aperture size vary slightly.

Under ideal conditions, a grating parabolic mirror based on these parameters performs almost exactly as intended. Figure 4.9a shows the power flow above such a grating with a design focal length of 100 μm and an aperture diameter of 80 μm, giving it a numerical aperture of 0.37. A very clear, uniform focal spot can be observed at the desired distance above the grating. In fact, if a cross section is taken through the focal plane (see figure 4.10) a sinc-type intensity distribution is obtained, as expected from diffraction theory. The first minima of this distribution are located approximately 1.3 μm either side of the optical axis, which is only slightly larger than the expected value of 1.225 μm for an ideal cylindrical lens of these dimensions.

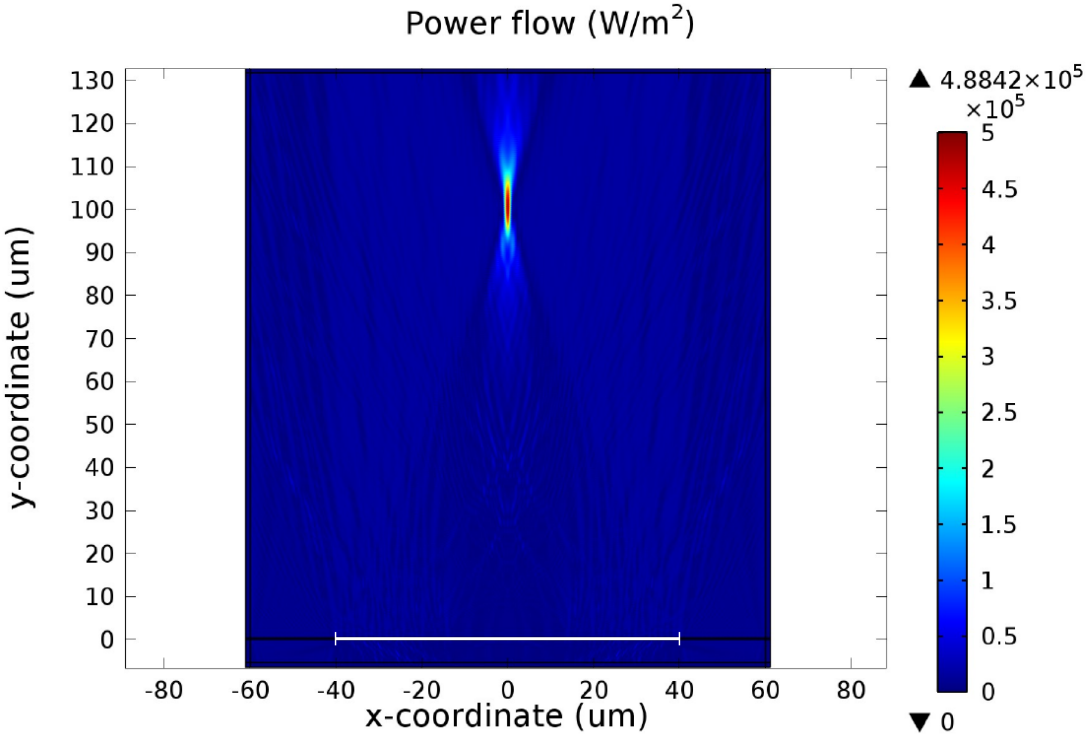
The path used for this design has an average reflectivity of approximately 80%, so it is informative to also look at the transmitted side (see figure 4.9b). This side shows a focal point, shifted to approximately $f = f_0/n = 100\,\mu\text{m}\,/1.45 = 69\,\mu\text{m}$ and with a narrower sinc function, both due to the refractive index of the glass substrate (see also figure 4.10). The intensity is much lower, as expected, with the peak power being approximately five times lower than that of the reflected spot. Given that the transmitted beam is sampled through the grating, with corresponding scattering losses, this value agrees reasonably well with the expected 80% reflectivity.

Several useful tolerance tests can also be performed with COMSOL simulations, such as spectral or angular bandwidths.

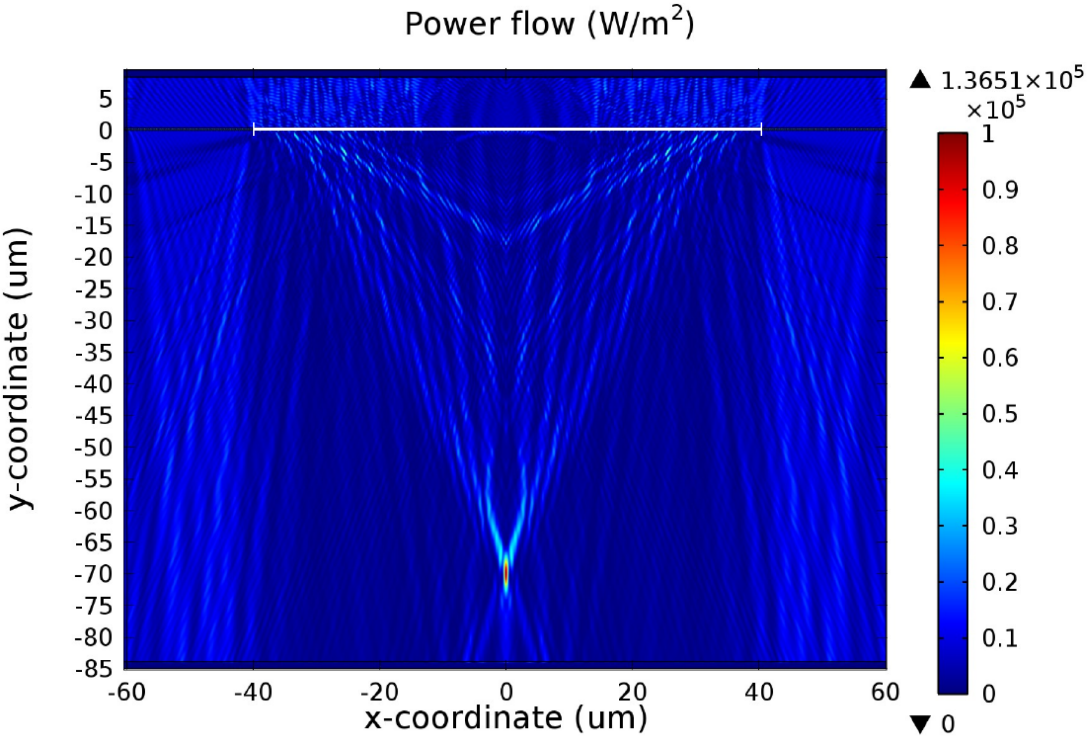
Spectral bandwidth

For most applications of metasurfaces, it is relevant to know the response to less than ideal illumination, for example illumination at an angle or at deviating wavelengths. We therefore ran several sets of COMSOL simulations to investigate the spectral and angular bandwidth of parabolic mirror gratings. An example grating based on the paths in figure 4.7 with a focal length of 60 μm and an aperture diameter of 80 μm was used to investigate both of these, with only the illumination being swept in the calculations.

The results for the wavelength simulations can be seen in figure 4.11. The illumination wavelength was swept over a bandwidth of approximately $\pm$15% around the centre wavelength of 980 nm. The most striking feature of these results is the fact that the efficiency stays high with relatively large wavelength offsets. In fact, for up to about a 10% deviation in wavelength, or 100 nm in either direction, the power flow through the focal spot only drops by 10%. Beyond these values the focus seems to split for the higher wavelengths, while the reflectivity



(a) Reflection



(b) Transmission

Figure 4.9 – 2D COMSOL simulations of a reflective focusing metasurface based on the paths shown in figure 4.7, illuminated under ideal conditions with a normal incidence plane wave. In both images the grating is located between $y = 0\,\mu$m and $0.325\,\mu$m, with the area above consisting of air and below of glass. (a) shows the power flow through the reflection region, while (b) shows the transmission region. (Note that the scales for both images are chosen to best show the relevant features and are therefore not the same.)

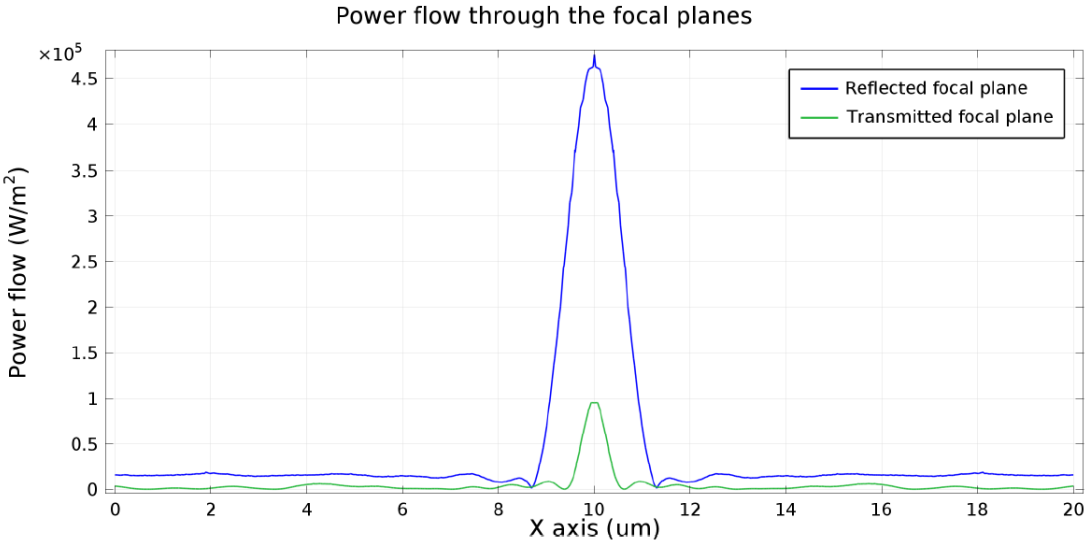


Figure 4.10 – Cross section through the focal planes of the simulations show in figure 4.9, highlighting the power flow through the focal planes.

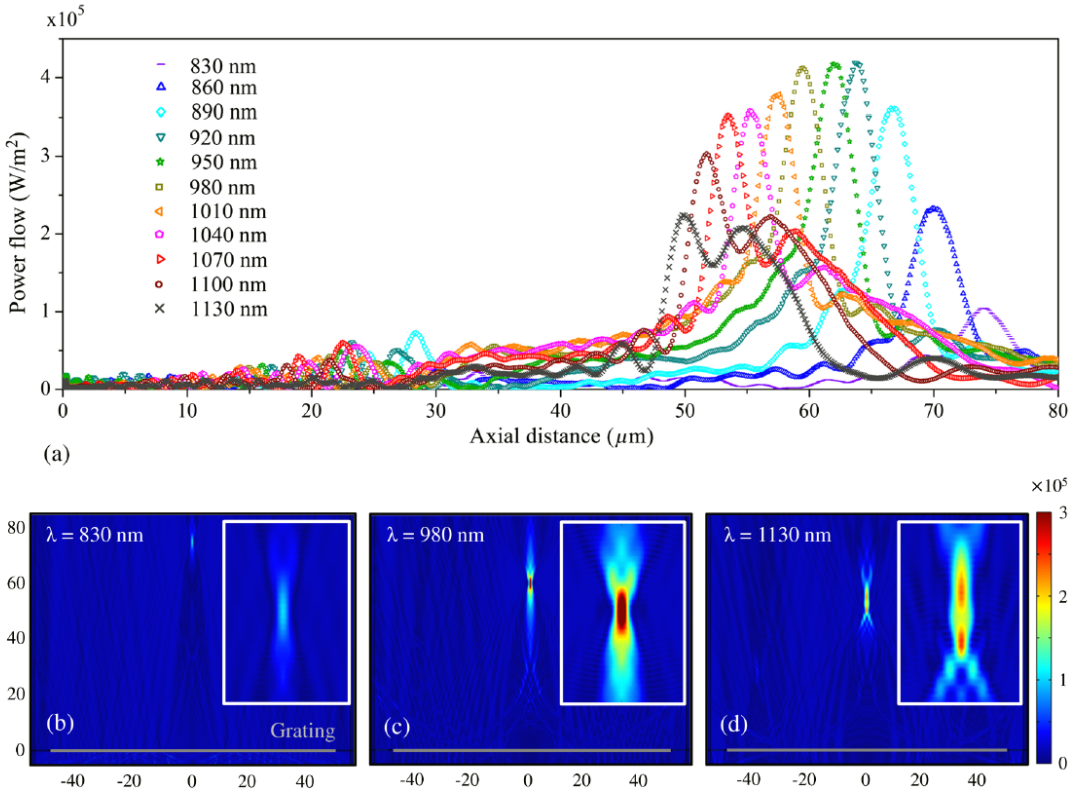


Figure 4.11 – Results of COMSOl simulations with varying incidence wavelength on a parabolic mirror grating based on the paths from figure 4.7. The design wavelength is 980 nm with a range of ±15%, or ±150 nm, swept as the incident wavelength. (a) shows cross sections of the powerflow along the optical axis of the results. (b), (c) and (d) give the 2D powerflow plots for 830 nm, 980 nm and 1130 nm, respectively. These highlight that good lateral focusing is retained throughout the 30% bandwidth range.

drops off quickly for the lower wavelengths. Nevertheless, fairly stable operation over a 20% bandwidth is obtained, which is much better than initially expected based on the resonant nature of these structures.

In fact, the same conclusion can already be drawn from the path that goes into the design of the metasurface. To do this rigorously requires extending the map of the path into the wavelength dimension to see the influence of a shift, but an indication can already be found in the two dimensional period-duty cycle maps given in figure 4.7. From chapter 2.3, we know that for most grating effects, it holds that $\lambda \propto \Lambda$. Therefore, observing the shift in period gives a reasonable idea of the effect of a small shift in wavelength. Doing so with the path in figure 4.7 then makes it immediately obvious that while phase stays relatively unchanged for quite a broad range, the reflectivity quickly drops off when sufficient shift is reached, while staying relatively untouched beforehand.

The second feature that immediately demands attention is the significant chromatic aberration that can be observed with the moving of the focal spot towards the grating for increasing wavelengths. However, from geometric considerations we would expect the focal length to increase with wavelength, which is counter to what we see here. We did not have access to a tuneable source to test this behaviour experimentally, but given the accuracy of the lens when modelled under the design conditions the simulations seem likely to be correct. However, we were not able to satisfactorily explain this counterintuitive chromatic aberration.

Angular bandwidth

The next parameter of interest is the angular tolerance. Simulating the angular tolerance breaks the symmetry of the optical axis and therefore requires the entire structure to be included in the simulation. In order to stay within computational limits, only half the aperture was used when compared to the spectral bandwidth simulations and only a $40\,\mu$m aperture was used, but otherwise the grating and settings are the same.

The incidence angle is varied from 0 deg to 12 deg in steps of 2 deg. The results are shown in a similar fashion as before in figure 4.12. It appears that the focusing behaviour is preserved quite well for angles of incidence up to approximately 10 deg, but the efficiency drops by two thirds in the process. For angles beyond 10 deg the focus effectively disappears.

Once again, under the assumption that most effects contributing to the working of the metasurface depend on the periodicity, equation 2.23 can be used to relate this behaviour such that it becomes predictable from the path through the

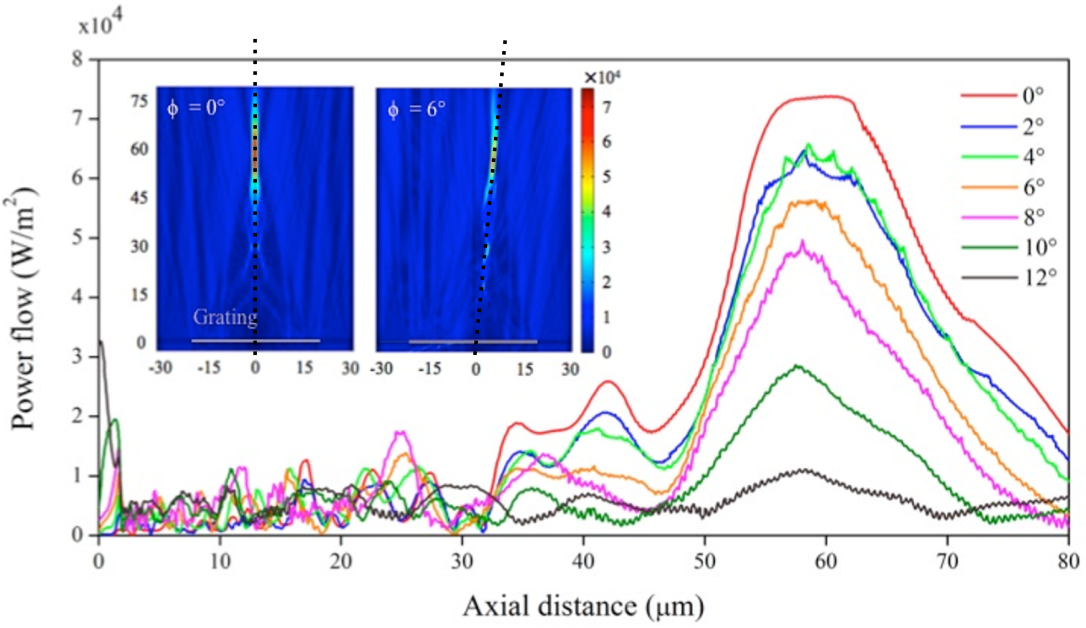


Figure 4.12 – Results of COMSOl simulations with varying angle of incidence on a parabolic mirror grating based on the paths from figure 4.7. The main figure shows the power flow along the optical axes of the simulations, i.e. from the centre of the metasurfaces in the reflected direction at the same angle as the angle of incidence. The insets show 2D plots of the power flow at normal and 6 deg incidence.

RCWA maps. The full version of that equation including incidence angle reads:

$$k_z^2 = k^2 - (k_x + G)^2 \,. \tag{4.7}$$

To compare with the RCWA maps we need to equate this equation to its normal incidence version where $k_x = 0$, which can be done by defining an effective wavelength. In other words, we can write:

$$k_z^2 \lambda^2 = k_{z,eff}^2 \lambda_{eff}^2, \text{ or} \tag{4.8}$$

$$\left(k^2 - (k_x + G)^2\right) \lambda^2 = \left(k_{eff}^2 - G^2\right) \lambda_{eff}^2. \tag{4.9}$$

Solving equation 4.9 for the effective wavelength leads to an asymmetric result:

$$\lambda_{eff} = \lambda + \frac{\Lambda}{q} \sin(\theta) \,, \tag{4.10}$$

where q is ± 1 for the modes of interest. This splitting in θ is to be expected as the off-normal incidence breaks the symmetry of the grating modes. Crucially, this relates the angular tolerance to the spectral bandwidth as the expected maximum

angle of incidence can be defined as:

$$\theta_{max} = \sin^{-1}\left(q\frac{\lambda_0}{\Lambda}\frac{\Delta\lambda}{\lambda_0}\right). \tag{4.11}$$

The path used for the angular tolerance simulations is the same as was used for the spectral bandwidth model, which showed a working bandwidth of approximately 10%. Combined with the central wavelength of 908 nm and an average period of approximately 575 nm, equation 4.11 predicts an angular tolerance of about 10 deg, which agrees reasonably well with the simulated results.

Experimental characterisation

Gratings based on the same principle were fabricated and tested experimentally. This work was almost entirely performed by Dr Annett Klemm and Andrei Ruskuc, but is based on the designs and simulations done by me.

The optical response of the gratings were tested using the setup described in chapter 3.3.2. The setup was used in reflection mode, so with a beam splitter in place before the objective. This arrangement allows the source to be focused into the back focal plane of the objective through the beam splitter, while the same objective is used to image through the other arm of the beam splitter. The incident beam is linearly polarised to ensure TM incidence on the grating. While this method is intended to get optimal measurement conditions for the reflected beam, the thin silicon samples with transparent substrate also allow for (partial) imaging through the grating to identify the transmitted beam scattering from imperfections in the substrate.

Metasurfaces were designed for a wide range of focal length and aperture diameters in order to test the size limits and accuracy of the designs. Both gratings for operation in air and in water were designed with different purposes in mind. The former was based on an amorphous silicon on glass platform identical to that used in the simulations and was intended primarily as a proof of concept with an aim towards possible imaging applications. The latter were fabricated in crystalline silicon on glass with the aim of being integrated into microfluidic channels for possible optical trapping applications[3, 89, 90].

A typical measurement result is shown in figure 4.13. The plane of imaging is parallel to the plane of the grating and scanned along the perpendicular axis to build up an array of slices of the intensity. The resulting slices together provide a beam profile of the reflected light and provide information about the focal spot position, collected intensity, and axial and lateral extents. The focal length measured in this particular example is significantly longer than the designed focal length. In the early experiments, there was a significant spread in the focal

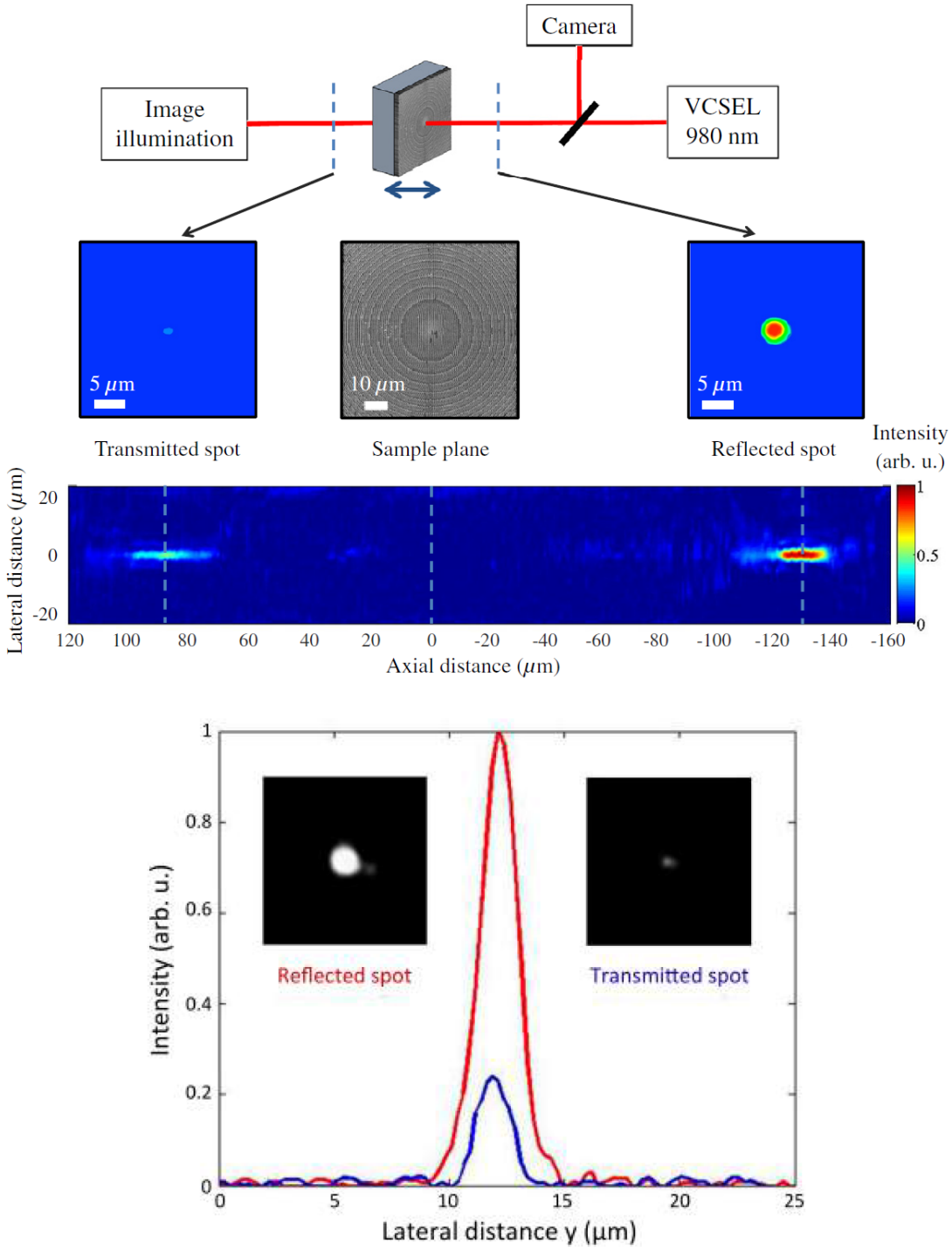
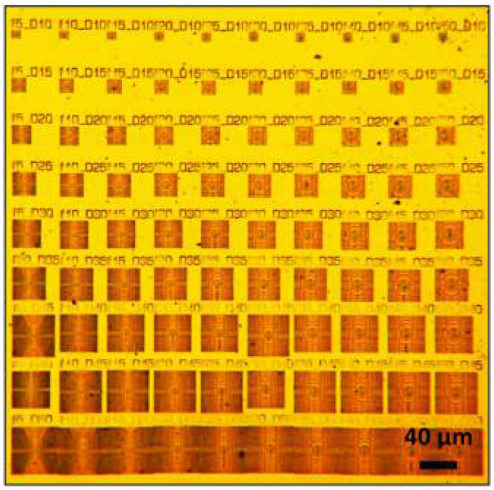
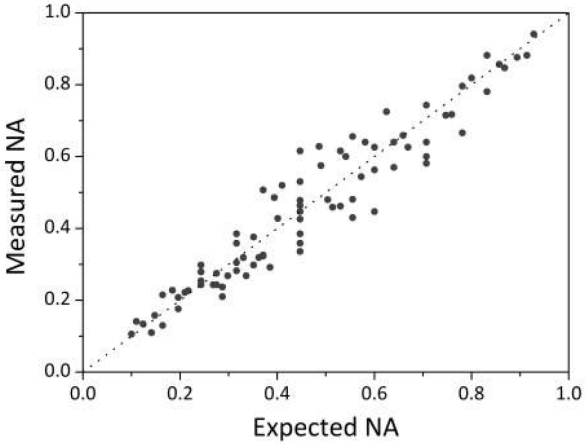


Figure 4.13 – Example of experimental setup and measurement results of a grating metasurface based on the paths in figure 4.7. The designed focal length was 100 μm with an aperture diameter of 80 μm, giving an ideal numerical aperture of 0.37. The top figure shows a simplified schematic of the measurement setup. The second row shows cross sections through the focal planes and the grating surface itself. The third row shows the intensity measured at different distances from the grating. Finally the bottom image shows a cross section of the intensities in the focal spots to show the airy-like distribution.



(a) Optical image of the parabolic mirror array.



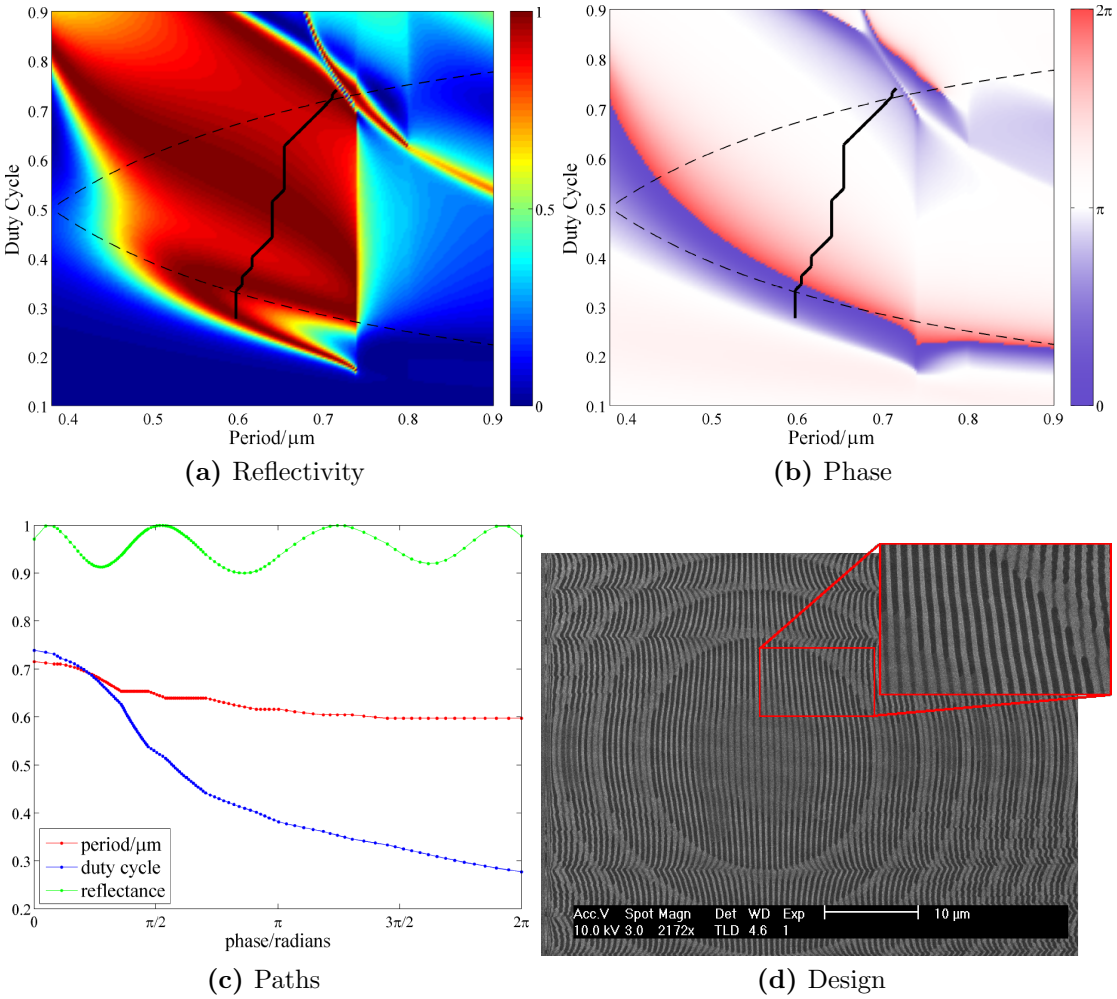
(b) Results expressed in terms of Numerical Aperture.

Figure 4.14 – A large range of metasurfaces were fabricated with differing aperture diameter and focal lengths. (a) shows an optical image of such an array of gratings. (b) shows the associated measurement results. Numerical aperture was chosen as a good measure for unifying the results from all the differing gratings into a single easy to grasp graph. Note that the numerical aperture is determined here as $\sin\left(\tan^{-1}\left(\frac{D}{2f}\right)\right)$, with f and D the actual measured values. It seems likely that the higher "NA"s in this graph did not actually achieve diffraction limited focusing, but this was not investigated thoroughly.

lengths of the fabricated lenses that we could never adequately explain. However, the average fabrication result was still close to the design value (see figure 4.14b). We therefore felt confident attributing the variation to fabrication error. In later fabrication runs, when we were more aware of the important variables in the grating design, the results are generally much closer to the designed specifications.

Parabolic mirror gratings were also designed for use in water, with the intent of employing them for optical trapping of dielectric particles, and ultimately cells. Due their flat microscale nature, metasurfaces are straightforward to integrate into microfluidics for this purpose. Trapping requires high gradients of intensity to create a strong enough force on a particle to overcome other forces, such as Brownian motion or liquid flow pressure. For the lateral trapping force, almost any numerical aperture can be used by just increasing the incident power. However, the axial trapping lies along the same dimension as the scattering force on the particle (i.e. the beam pushing the particle in the direction of propagation)[4, 91].

For this reason, most optical trapping is done with numerical apertures greater than 0.65, which ensures a suitable ratio between the axial gradient and the scattering forces. However, this is an empirically found "safe" number, and not a



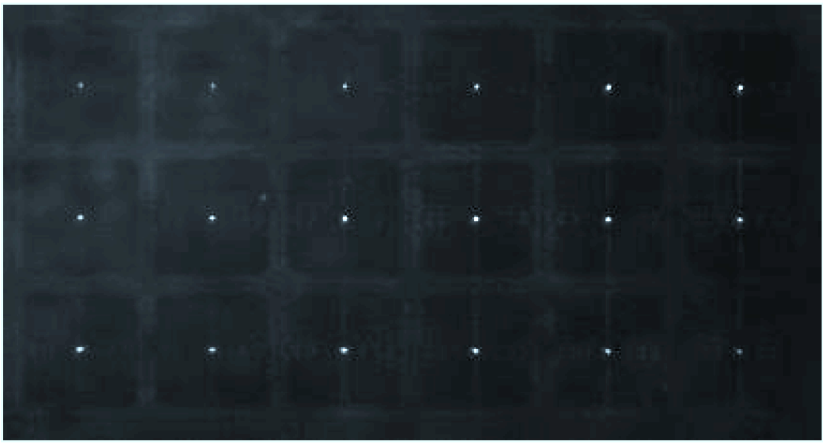
(a) Reflectivity **(b)** Phase

(c) Paths **(d)** Design

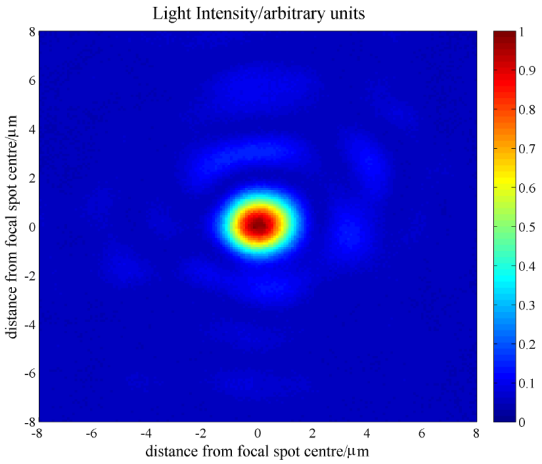
Figure 4.15 – Maps of the reflectivity (a) and phase (b) as function of period and duty cycle for a crystalline silicon grating on a glass substrate covered by water. The silicon thickness used is 380 nm for a wavelength of 1064 nm with TM polarised incident light. Solid lines through the maps indicate the chosen path, while the dashed lines highlight the 1:2 aspect ratio between narrowest grating feature and etch depth. (c) shows the resulting paths period, duty cycle and reflectivity as a function of phase. (d) shows an SEM image of a parabolic mirror grating based on this path. The inset is a higher magnification scan of the same grating.

hard limit. Given favourable other conditions, trapping may be achieved with much lower numerical apertures down to 0.3[91, 92]. As focusing grating metasurfaces are inherently limited by Nyquist's sampling theorem, they can only reach a numerical aperture of ~ 0.6, and more realistically 0.5. To this end the gratings designed for use in water tend to be designed with a numerical aperture of approximately 0.6.

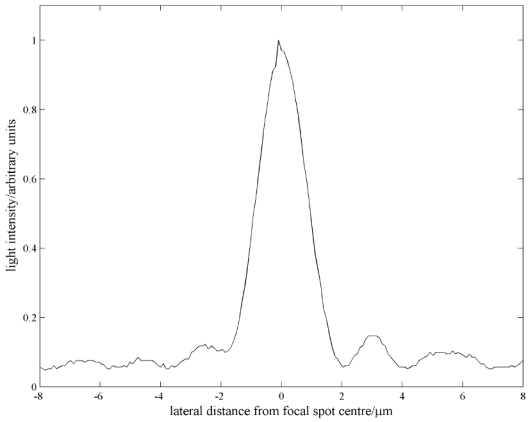
These structures were designed in crystalline silicon and use the paths shown in figure 4.15. The results so far were strongly determined by available materials, where in this case the high starting thickness (500 nm) allowed thinning of the sample to the ideal thickness of 380 nm. The path used here is therefore the best



(a) Optical image of several gratings in an array.



(b) 2D intensity in the focal plane.



(c) Cross section of intensity through maximum of focal spot.

Figure 4.16 – Optical response of the operation of parabolic mirror gratings in water based on the paths shown in figure 4.15c. (a) is an image of the focal plane above an array of gratings. (b) and (c) show the intensity in the focal plane of one such grating as imaged with a high numerical aperture objective.

continuous path for crystalline silicon on glass we could find, with an average reflectivity of approximately $95 \pm 5\%$. This is both a higher and more constant reflectivity than what was used before. Moreover the grating parameters change smoothly with phase and almost entirely fall within easily fabricatable regions allowing for accurate reproduction of designs.

The results are shown in figure 4.16. The top image shows the operation of an array of gratings with the same design but slightly varying writing doses and sizes in the lithography step of fabrication. The focusing behaviour seems to be retained throughout the resulting range of different duty cycles, with primarily the efficiency dropping.

Figures 4.16b and 4.16c show the intensity at the focal point of the brightest of these spots. A very obvious airy disk pattern can be observed, with the first

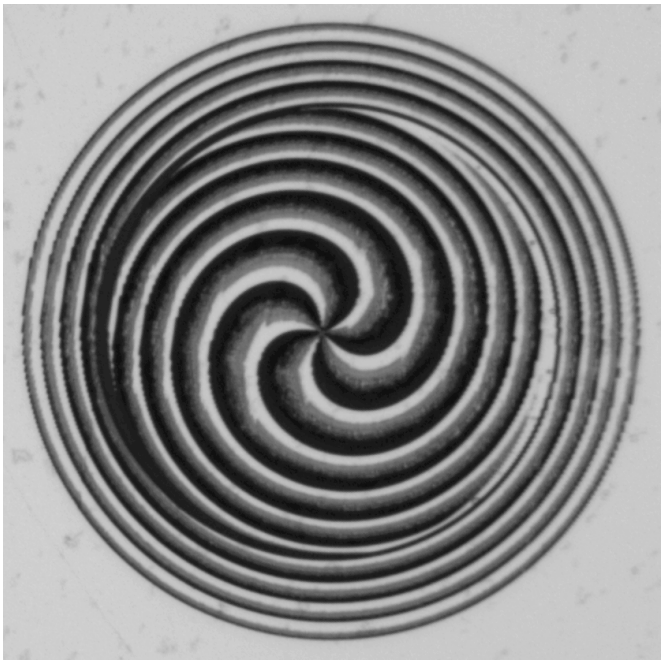


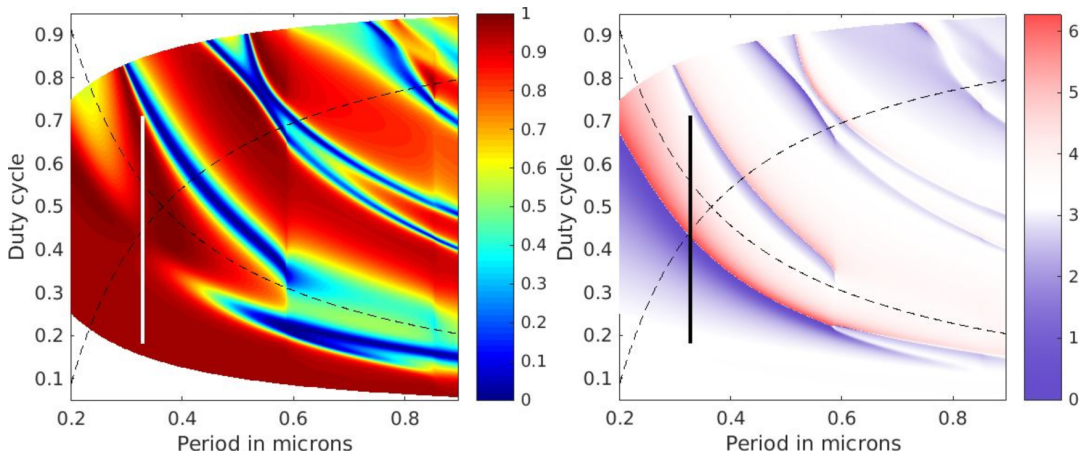
Figure 4.17 – Optical image of an attempt at fabricating a transmissive metasurface in crystalline silicon based on the paths shown in figure 4.18. The black areas are entirely etched away, the lightest areas entirely unetched, with only the grey parts in between approximating the intended grating design. This failed fabrication is due to the overly demanding range of duty cycles included in the design.

minimum at a radius of approximately $2\,\mu$m. This radius corresponds to an NA of 0.43, which is not quite the designed value of 0.62. This deviation implies that only the centre two thirds of the grating are operating as desired. Nevertheless, the quality of the focal spot is better than before and even with the lower NA, it might be possible to achieve optical trapping.

Similar gratings were integrated into microfluidic channels to attempt trapping. Preliminary tests performed in St Andrews by Andrei Ruskuc and Dr Mingzhou Chen are promising, clearly showing lateral trapping but not quite succeeding at trapping in the axial direction. More extensive tests are in progress.

Transmission lenses

All experimental work and therefore most modelling was performed on reflective parabolic mirrors. This was partially influenced by applications, but also by limitations of the fabrication. For the available materials and wavelengths, the fabrication turns out to be more challenging, requiring some sort of proximity error correction to be in place. This is due to the fact that suitable paths tend to require a wide range of aspect ratios (see figure 4.18), to the point that a single dose in the lithography process can not accurately reproduce both ends of the path. Figure 4.17 shows an optical image of a metasurface based on this path, fabricated without employing proximity correction. The black parts here



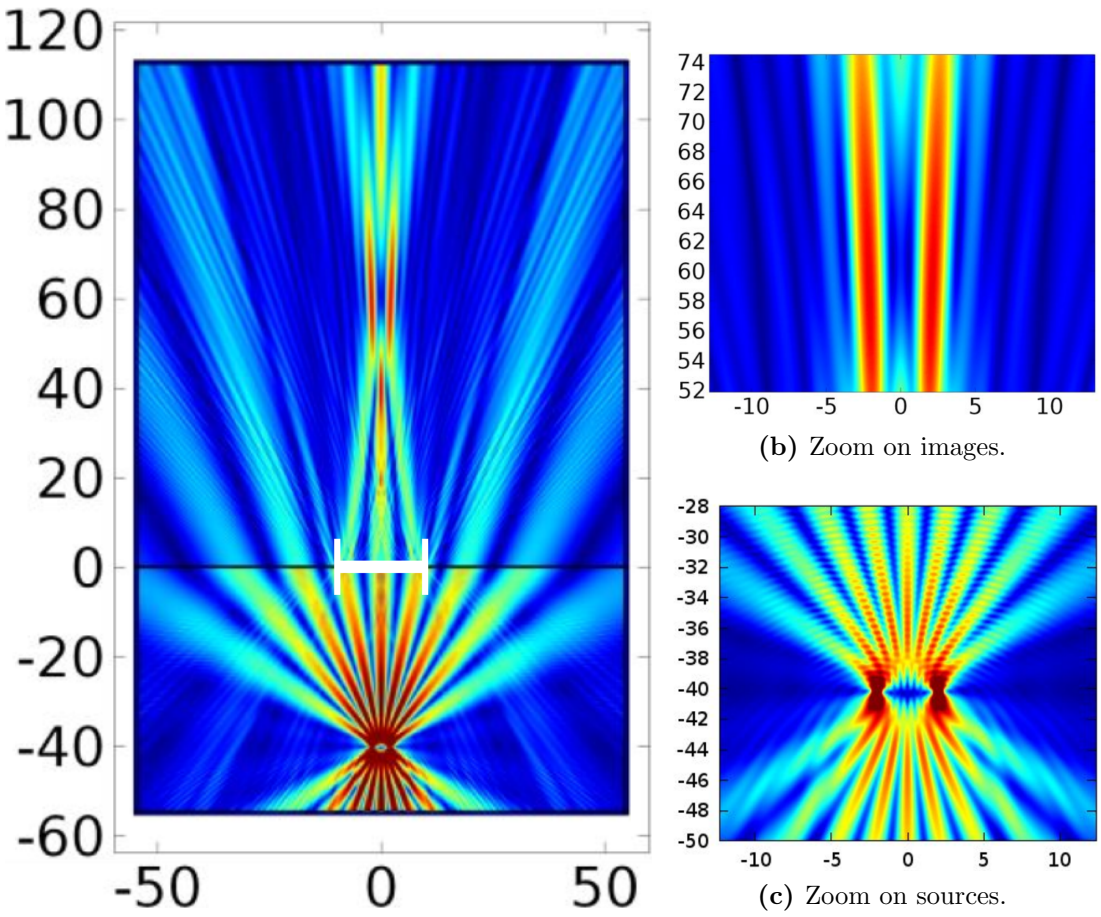
(a) Reflectivity. **(b)** Phase.

Figure 4.18 – RCWA period-duty cycle maps for a 365 nm thick crystalline silicon grating on glass, operating in transmission for incidence from the substrate direction. The operating wavelength is 853 nm for TM polarised incident light. The solid lines in both figures indicate the chosen path, while the dashed lines indicate a 1 : 2 aspect ratio.

are completely etched away, while the light parts are not etched at all.

Modelling the behaviour of such lenses is only of limited interest as it can not be matched to experiment. Still, for imaging purposes transmissive lenses are far more useful than reflective ones, so several simulations were run to examine the imaging potential. An example is shown in figure 4.19, which involves the 2D simulation of a transmissive grating lens with point dipole sources for excitation. The two point dipoles were moved apart at the $2f$ plane of the lower (air) side of the model and the resulting images in the cover (glass) region checked for accuracy.

The grating in question had a numerical aperture of only 0.45, with a focal length of 20 μm and an aperture diameter of 20 μm. The rest of the lateral extend of the device layer is unpatterned silicon. While clear focal spots can be seen and the resolution is close to the expected value, the distance from the grating is much larger than $2f/1.45$. Moreover, if the dipole sources get moved too far apart, the focal spots are entirely lost. Both of these behaviours can be readily explained by the behaviour of the gratings under off-normal angles of incidence as discussed earlier. While it is entirely valid to put a point source in the designed focal point to invert the operation of the metasurface, putting it at $2f$ effectively changes the angle of incidence on each point of the grating as $\Delta\theta = \tan^{-1}\left(\frac{r}{f}\right) - \tan^{-1}\left(\frac{r}{2f}\right)$. Moving the sources laterally further increases the angle of incidence. The resulting angles grow quickly away from the design, and therefore from what the grating can tolerate.



(a) Overview.

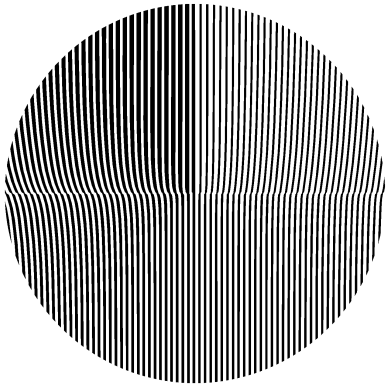
(b) Zoom on images.

(c) Zoom on sources.

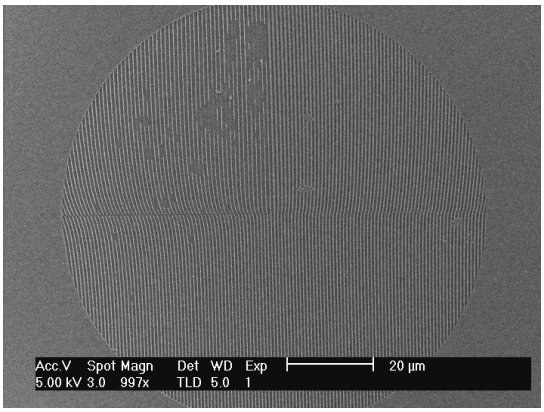
Figure 4.19 – Modelling results for a transmissive silicon grating lens based on the paths shown in figure 4.18. Two dipole point sources in the lower domain (air) positioned at the $2f$ plane of the design serve as the excitation, and are then moved apart to test the resolution of the system. Only the centre 20 μm of silicon is patterned as a lens, as indicated in the figure in white, with the rest left unpatterned. (a) shows an example resulting 2D intensity plot for a source separation of approximately five times the wavelength. (b) is a zoom on the images in a, (c) while (c) zooms i on the source dipoles.

The main problem here is not the feasability of imaging lenses, as there are clear images in the results and the resolution is as expected given the distance from the grating and therefore the collecting angle. The main problem is that the lens design needs to take into account the angle of incidence from the imaging plane in the paths. While entirely possible, this is sufficiently far away from the current implementation of the pathfinding that I did not attempt it.

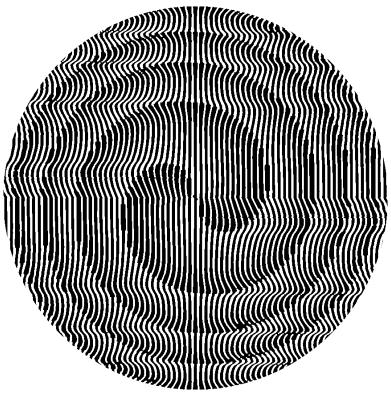
We are currently also exploring the use of cylindrical, 2D, lenses where the grating is varied in the dimension along the ridges rather than across. Varying in this dimension has several advantages, most notably the lack of a strict limit based on Nyquist and a decreased sensitivity to angle of incidence. The combination should give more freedom in employing such lenses for imaging purposes. This



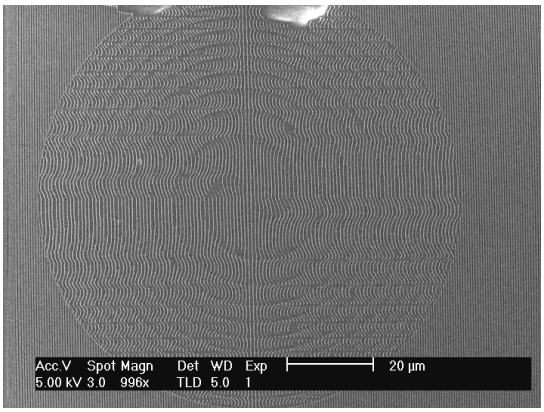
(a) Design $l = 1$ phase plate.



(b) SEM of fabricated $l = 1$ phase plate.



(c) Design $l = 2$ focusing phase plate.



(d) SEM of fabricated $l = 2$ focusing phase plate.

Figure 4.20 – Designs and fabricated results for reflective metasurfaces intended to generate vortex beams. The structures were all designed and fabricated in 375 nm amorphous silicon on glass with an intended wavelength of 1064 nm. (a) shows the design for a phase plate analogue that generates an $l = 1$ vortex beam. (c) shows the design for a focusing phase plate with $l = 2$ and an focal length of 100 μm. (b) and (d) are SEM images of the fabricated results corresponding to (a) and (c) respectively.

particular direction of research is still in its early stages, however.

4.2.2 Further beams.

Most of the work presented here was done on focusing metasurfaces. To prove the general beam shaping capabilities of our metasurfaces we also attempted to construct metasurfaces for the generation of several other beam types, namely vortex beams and Bessel beams (see chapter 2.1.)

Vortex beams

Vortex beams have a phase profile defined as $\phi = l\theta$, where l is the azimuthal mode index of the vortex beam describing the number of twists in the beam, and θ the azimuthal angle around the origin. This phase profile is entirely orthogonal

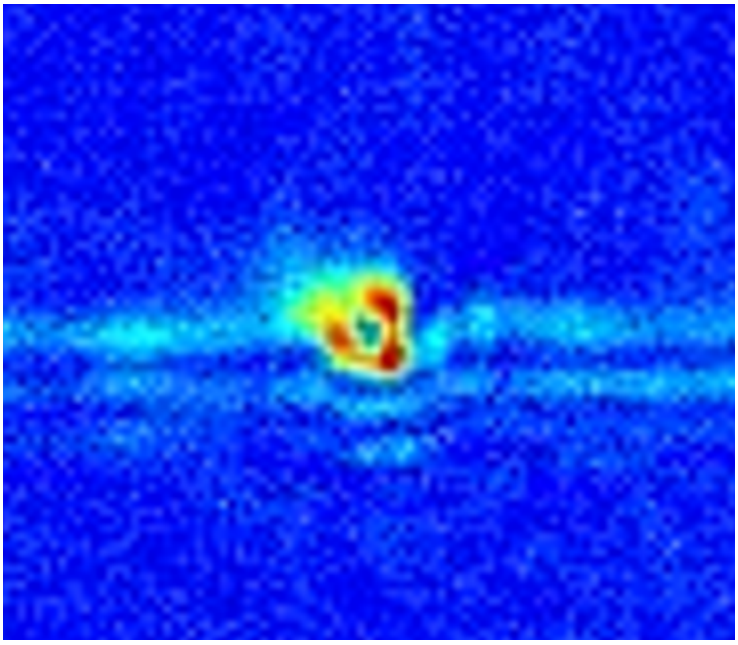


Figure 4.21 – Measured intensity in the focal plane of a reflective metasurface with both a focusing and an azimuthal phase profile. The grating, shown in figure 4.20d, was designed in 375 nm amorphous silicon on glass with an azimuthal mode number $l = 2$ and a focal length of 100 μm, for a wavelength of 1064 nm.

to that of a lens, which makes it possible to combine them into a phase profile that generates a focused vortex beam, $\phi = \frac{2\pi}{\lambda}\left(\sqrt{r^2 + f^2} - f\right) + l\theta$.

The basic vortex beam is easier to fabricate, with the phase only changing quickly close to the centre of the grating and the number of phase jumps limited to l. The focused vortex beam, on the other hand, is far easier to experimentally investigate as the contrast between the inner dark spot and the ring around it is much higher. Therefore, we chose to make both types.

Designs and SEM images of fabricated gratings for both a regular $l = 1$ and a focusing $l = 2$ metasurface are shown in figure 4.20. The metasurfaces were designed and fabricated in 375 nm thick amorphous silicon on glass, for an operation wavelength of 1064 nm. For testing, the fabricated gratings were sent to a collaborator in St Andrews, Dr Mingzhou Chen. The setup used was identical to the one shown in figure 3.3.2.

The non-focusing metasurfaces did not produce any promising results in the preliminary tests, but the focusing metasurfaces showed clearly recognisable donut shapes near the expected focal distance. An example is shown in figure 4.21.

To definitively prove the vortex nature of the measured beams an interference measurement would be necessary, as a central dark spot can in principle be caused by any number of effects. By interfering the beam with a reference beam unique intensity distributions can be measured. For example, interfering with a plane wave at an angle produces a forked interference pattern, while interfering with a

colinear plane wave leads to a spiral shape where the number of arms match the azimuthal mode number l of the beam. (An alternative method is described in chapter 6.2.)

To facilitate such a measurement, we realised that the interference setup could be build into a single metasurface design. First, an inner metasurface is designed as a focusing vortex beam generating grating with a circular aperture with radius r_1. In a ring around this first metasurface we then place a second focusing metasurface with the same focal length, but without an azimuthal phase component. With the total radius of the design $r_1 + r_2$ and $r_2 = \left(\sqrt{2} - 1\right) r_1$, the area of both segments is equal, and therefore the expected intensity contribution in the focal plane as well. The result should be an interference pattern in the focal plane between two colinear beams, one with and one without an azimuthal phase.

These latter measurements were designed, but before progressing onto them we decided that these metasurfaces would be more useful if done in transmission. As explained in the previous section, however, the transmission metasurfaces have not yet been successfully fabricated. An example at an attempt at fabricating such a grating can be seen in 4.17, which was designed as a focusing $l = 5$ centre grating surrounded by a lens with no azimuthal phase component.

Bessel beams

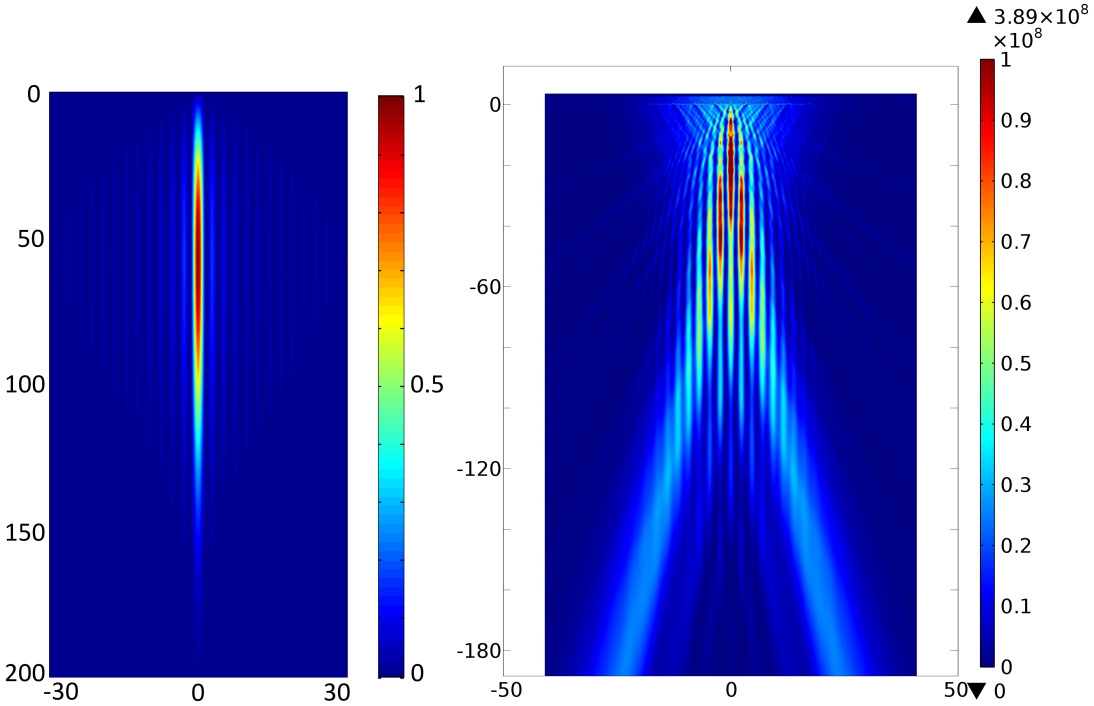
The last beam type investigated with direct phase metasurfaces is the Bessel beam. As discussed in chapter 2.1, Bessel beams are formed by two plane waves propagating at a slight angle with respect to each other. The usual way of generating these is through an axicon (see figure 2.6a), a dielectric material shaped as a cone. This is just a way of refracting light with a linear slant along each radial, or in phase terms:

$$\phi = \alpha_0 r, \tag{4.12}$$

where α_0 is the slope of phase, and therefore it is also the output angle of a plane wave incident on a metasurface with this phase profile.

This should be straightforward to imprint on a metasurface as it is a much simpler phase function than that of lenses or even vortex beams. I never fabricated any Bessel beam generating gratings, instead studying simulated transmissive versions in two dimensions. As it turns out, they show a feature of direct phase metasurfaces that is less obvious for the other beam types discussed.

Figure 4.22 shows results of 2D COMSOL simulations of transmissive metasurface with a linear radial phase profile. Along side it is shown an analytical calculation of the same incident Gaussian beam with $w_0 = 20\,\mu$m after passing through an axicon with the same angle α_0. The biggest difference between the two calculations a priori is that the analytical calculation is for a conical axicon,



(a) Analytical calculations for an axicon.

(b) COMSOL simulations for an analogous metasurface.

Figure 4.22 – Comparison between a quasi Bessel beam generated by an axicon and a metasurface. The incident beam is in both cases a Gaussian with width of 20 μm and wavelength 800 nm, and both optical elements are designed for an angle of 10 deg between the two interfering plane waves. (a) was analytically calculated for a perfect axicon. (b) was numerically simulated in COMSOL in 2D, for a metasurface with a linear radial phase profile in a crystalline silicon on glass platform.

while the metasurface is an analogue to a wedge. This is why the intensity distribution of the metasurface diverges into two Gaussian beams far away from the grating, while the axicon symmetrically dissipates. Apart from that difference though, both results clearly show the expected pattern for a quasi Bessel beam, with long, parallel and equidistant lines of high intensity propagating away from the grating.

A more concerning effect unique to the metasurface becomes apparent in figure 4.23, which gives a closer look at the intensities close to the grating. As can be observed, the intensity and width of the beams as they propagate is not smooth, but rather uneven. This results from a higher order diffraction effect that is present to some degree in all direct phase metasurfaces.

As the paths that make up the designs for these structures do not have a completely uniform transmission, there is a modulation to the intensity of the light transmitted through different parts of the metasurface. This modulation can be imagined as an extra, unintended, diffraction grating with a much larger period,

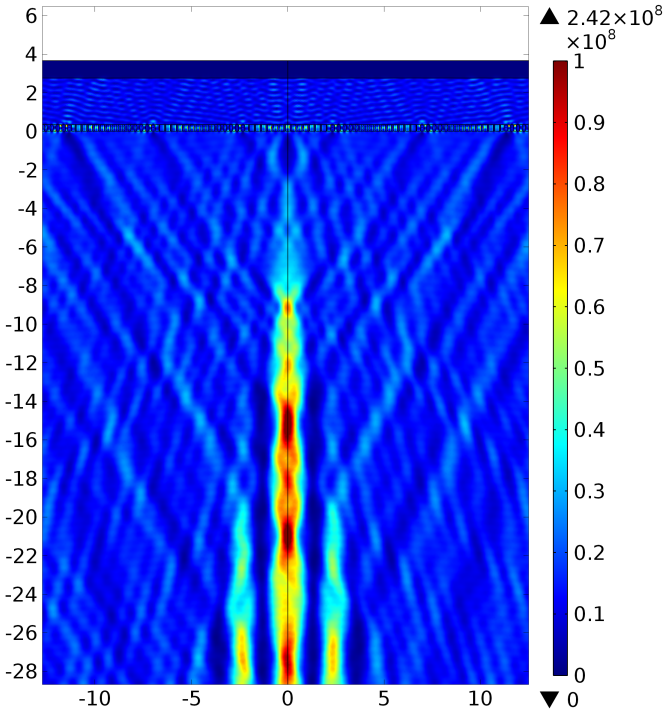


Figure 4.23 – Zoom in on the area close to the grating for the simulation shown in figure 4.22. The slight asymmetry that can be observed in the fields in some places is most likely due to insufficiently small meshing.

superimposed upon the actual grating. As such, it causes a small but noticeable amount of diffraction. This serves a loss channel that limits the efficiency, or, as is apparent in the intensities of the Bessel beam, imperfections in the generated beams.

Upon close inspection these imperfections can also be found in the earlier results, but they become much more apparent for the relatively low intensities of the Bessel beam cores.

5 Polarisation control and geometric phase metasurfaces.

While the metasurfaces in the previous chapter add a specific phase as a function of the position on the surface, the devices discussed in this chapter use a principle that is a bit more complicated in conception. These metasurfaces consist of a mosaic of $\lambda/2$ plates, oriented slightly differently in the plane such that their relative orientations create a modulated phase front.

The specifics of how this method works will be treated in the following section, but importantly, it means that only one grating unit cell is used everywhere, rather than a different one for each phase. This feature turns out to make highly efficient metasurfaces possible at lower refractive indices than are required for those discussed before. Consequently, they can be made out of materials that are of lower refractive index and fully transparent in the visible.

The first section of this chapter will work through the theory of polarisation control through gratings and geometric phase metasurfaces, the design process of such structures and an in depth study of several devices uniquely suited to them. The section will conclude with a description of the fabrication specifics for the materials and designs used. The second section will deal with simulation and experimental results of wave plates, ring gratings and general geometric phase metasurfaces.

5.1 Theory and methods.

5.1.1 Form birefringence and beyond.

The key concept that this chapter is built on is the ability to control the polarisation of light. By this I mean being able to take an incident beam with a given polarisation and changing it such that the polarisation of the reflected or transmitted beam is being converted into a specified different polarisation. The analogue in traditional optics would be a $\frac{\lambda}{2}$- or a $\frac{\lambda}{4}$-plate.

Waveplates work by tailoring the phase delay between two orthogonal axes[44]. Depending on the polarisation of the incident light and the orientation of the waveplate, the transmitted light will have a different polarisation. While depending on the exact phase retardation and orientation, almost any transformation is possible, there are two variants of particular interest: $\frac{\lambda}{2}$-plates and $\frac{\lambda}{4}$-plates. $\frac{\lambda}{2}$-plates delay the phase of light polarised along one axis by π relative to the other axis. This effectively flips the component of the polarisation along that axis, resulting in linear polarisation being "turned" to point in a new direction or

circular polarisation changing handedness. $\frac{\lambda}{4}$-plates, in contrast, delay the light along one axis by $\pi/2$. This allows the conversion from linear polarised light into circular and vice versa.

The primary method of achieving this phase delay is birefringence. Essentially, birefringence describes a material that has different refractive indices along different axes. While the difference is generally small, it accumulates as light propagates through the material. The two polarisations then collect phase at slightly different rates and, given sufficient propagation length, get out of phase by the required amounts.

Linear gratings, by their very nature, have a different dielectric environment for TM and TE polarised light. For very small periods, this is immediately obvious from the effective indices given in equations 2.26 and 2.27. Simply taking the difference between them gives rise to

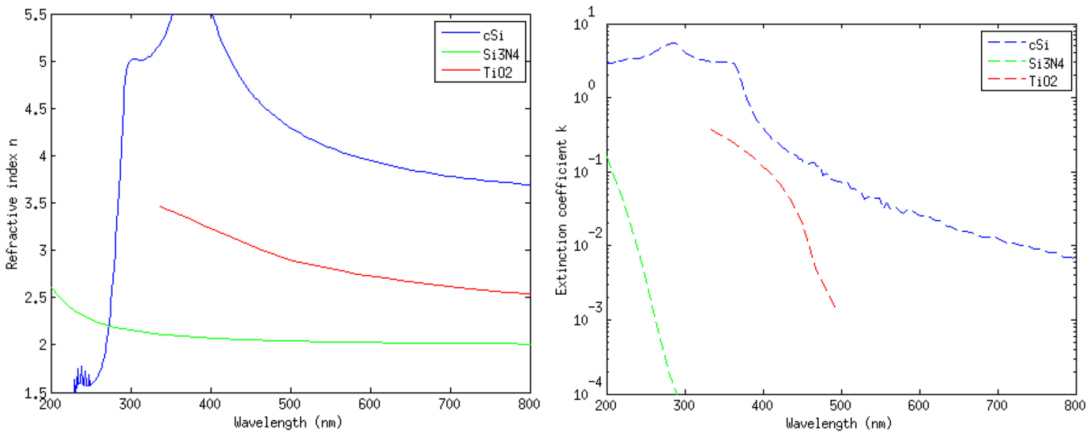
$$\Delta\phi = \frac{2\pi t}{\lambda_0}\left(n_{eff,TM} - n_{eff,TE}\right) = \frac{2\pi t}{\lambda_0}\Delta n, \text{ or} \tag{5.1}$$

$$t = \frac{\lambda_0 \Delta\phi}{2\pi\Delta n}. \tag{5.2}$$

Here Δn is a function of the ridge refractive index (n_{high}), groove index (n_{low}), and duty cycle. The maximum value Δn can take for any combination of dielectrics lies between 0 and $n_{high} - n_{low}$. While the exact solution for the maximum of Δn is a bit unwieldy to show here, a good approximation for most combinations of materials commonly used is $\Delta n \approx \frac{\left(n_{high}-n_{low}\right)^{\frac{3}{2}}}{3}$. That value is usually in the range of 0.3 to 1.3, which far exceeds the Δn of naturally occurring birefringent materials which tends to be on the order of 0.01 or less. This birefringent behaviour of periodic structures is also known as form birefringence[62, 93–95].

However, in this particular version, the practical applications of form birefringence are rather limited. The reason for this is that with different refractive indices also come different reflectivities. For example, if using silicon ($n \approx 3.5$) in air ($n = 1$) at the duty cycle with the maximum $\Delta n \approx 1.3$, then there is an almost 20% difference in reflectivities just of the first surface of the grating. As in the ideal case, only the phase gets influenced and transmittivity of both polarisations is simply 1, or at the very least equal, this lowered efficiency severely limits the use of form birefringence. If, instead we choose a material and duty cycle such that the Δn is that of a traditional waveplate, e.g. 0.01, then the grating would have to be 25 wavelengths thick for a $\frac{\lambda}{4}$-waveplate while still retaining a deep sub-wavelength period. This is both undesirable and difficult to fabricate.

Using resonances allows a good compromise. Exploiting the resonances of sub-



(a) Refractive index. **(b)** Extinction coefficient.

Figure 5.1 – Real (a) and imaginary (b) parts of the refractive indices of the three materials discussed in this chapter: crystalline silicon (cSi), silicon nitride (Si_3N_4) and amorphous titanium dioxide (TiO_2). The data for the cSi and Si_3N_4 were taken from filmetrics.com[85], while for the TiO_2 data from Kischkat *et al.*[96] was used. The latter's measurements of k did not extend below 10^{-3}.

wavelength gratings provides the opportunity to combine resonances for one or both polarisations, which could lead to a phase difference of any desired amount between the polarisations. Doing so allows actual 100% transmittance or reflectance for both polarisations to be obtained while retaining a given desired phase retardation.

Importantly, such resonant form birefringent waveplates can be designed in materials with relatively low refractive indices. This lower index allows a change in material from silicon to a material transparent in the visible parts of the spectrum such as Si_3N_4 or TiO_2. Figure 5.1 shows the optical constants of the respective materials.

There are two distinct types of metasurface that can be based on polarisation converting unit cells. The first directly uses the polarisation conversion of the gratings, as a surface can convert an input beam's polarisation into a desired, spatially varying, output polarisation. This property can be used to design metasurfaces that convert conventional beams into vector beams[62].

The goal of the second type is a change in the phase profile of a passing beam. Here, the change in the phase profile is mediated by polarisation conversion, specifically by flipping the handedness of circularly polarised light. To do so, a concept called the Pancharatnam-Berry phase, or geometric phase, needs to be introduced[20, 97].

5.1.2 The Pancharatnam-Berry phase.

The Pancharatnam-Berry phase, also known as the geometric phase, is a phase term that is more commonly found in quantum mechanics[10, 97–100]. It arises when a system can take several different paths to get from a given input to a given output state, with the different paths leading to a phase difference between the otherwise identical results. The implementation for geometric phases in this work provides a nice example for the use of geometric phases in Photonics.

The geometric phase as used here arises from the combination of circularly polarised light with $\frac{\lambda}{2}$-plates. Circularly polarised light consists of equal amplitudes of light polarised in two orthogonal directions, but with a $\frac{\pi}{2}$ phase difference between them. When this light passes through a $\frac{\lambda}{2}$-plate, one polarisation gets retarded by π with respect to the other. Due to the cyclical nature of phase, this effectively changes the handedness of the light but otherwise keeps it untouched.

However, there is another phase component that is added to the light when this happens, which is generally ignored. This extra phase arises naturally if we consider the complete set of phase transformations that are occurring when light passes through a $\frac{\lambda}{2}$-plate.

First, we can describe the electric field of the circularly polarised light as

$$E = e^{j\phi}\hat{x} + e^{j\left(\phi+\frac{\pi}{2}\right)}\hat{y}. \tag{5.3}$$

The waveplate can generally be described by the transformation matrix

$$T = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{5.4}$$

which describes a reflection with respect to the x-axis. However, as the waveplate has a well defined fast and slow axis, this matrix is associated with its own coordinate set $\hat{x'}$ and $\hat{y'}$. The actual transformation therefore also requires a rotation matrix with the angle between the two coordinate sets, $R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, making the full transformation matrix:

$$\begin{aligned} T' &= R^{-1}TR = R(-\theta)\,TR(\theta) \\ &= \begin{pmatrix} \cos(-\theta) & \sin(-\theta) \\ -\sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \\ &= \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}, \end{aligned} \tag{5.5}$$

which now describes reflection in a line at angle θ to the original x-axis. If we

apply this to the circular polarised input light, we get

$$E_{out} = T' E_{in} = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix} \begin{pmatrix} e^{j\phi} \\ e^{j\left(\phi+\frac{\pi}{2}\right)} \end{pmatrix} = \begin{pmatrix} e^{j(\phi+2\theta)} \\ e^{j\left(\phi-\frac{\pi}{2}+2\theta\right)} \end{pmatrix}. \quad (5.6)$$

In effect, twice the angle θ directly becomes an added phase. As this addition happens equally for both phases, it can normally be ignored. The added phase component does have an influence, however, when a second waveplate is added to the first at an angle. Both waveplates now flip the handedness of the light, but the extra phase of $\phi = 2\theta$ due to the orientation is different. This extra phase is the geometric phase.

The correlation between the orientation of a polarisation converting element and the output phase is what most metasurfaces rely on. By covering a surface with birefringent elements with relative angles derived from the relation in equation 5.6, an incident circularly polarised wave is converted into a wave with the opposite handedness as well as the desired phase profile.

Metasurfaces based on this concept have several advantages over the type of metasurface described in the previous chapter, all of which stem from the fact that every unit cell is identical to the next. All unit cells therefore have the exact same amplitude of transmission or reflection, and by extension so does the metasurface as a whole. Where other types of metasurfaces often have losses and imperfections due to diffraction of the varying amplitude over the surface, these do not have that issue. (For an example of these effects, see figure 4.9b.) Moreover, only a single unit cell with the desired properties needs to exist for a platform to be viable, rather than a full 2π phase coverage, which makes the structure easier to fabricate and reduces the required refractive index contrast.

5.1.3 Design.

For the design of the metasurfaces discussed here, many of the previous considerations remain true. It is still important that the structures can be fabricated, so the aspect ratios can not be too high, and the gratings still need to be thick enough and of sufficiently high refractive index to support at least one Guided Mode Resonance for both polarisations.

However, all of these requirements are less strict in the case of a geometric phase metasurface. First, it is easier to get high aspect ratios fabricated in the relevant materials. For example, in Si_3N_4, aspect ratios of $1:10$ can be achieved and $1:5$ is straightforward. Second, while in the previous cases the resonant modes needed to exist over a range of duty cycles to allow for a 2π phase coverage,

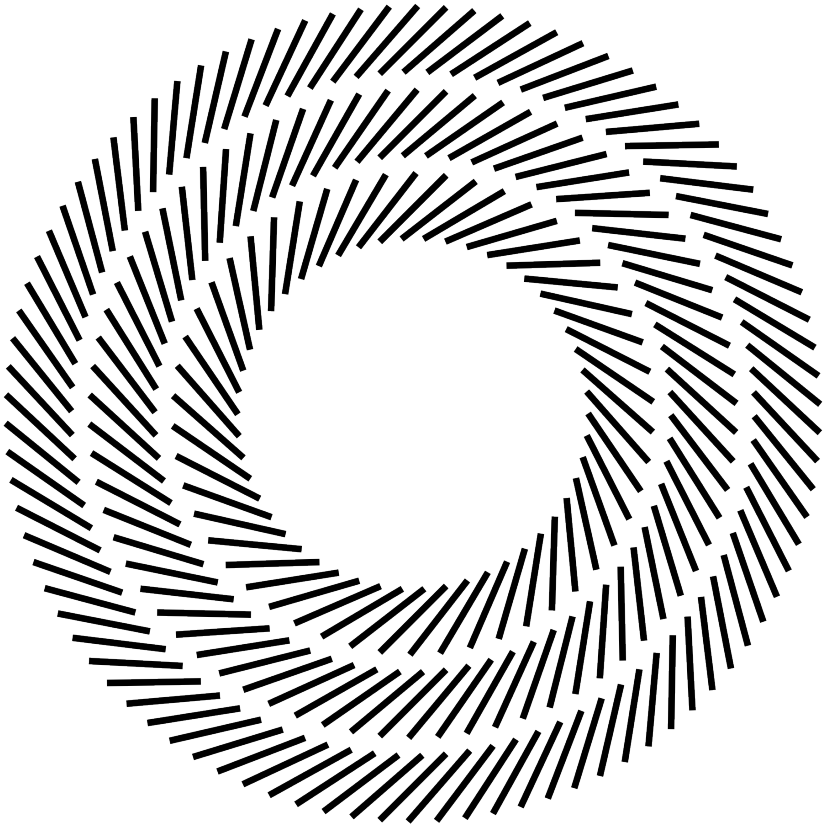


Figure 5.2 – Example design for a concentric ring grating with $\frac{\lambda}{4}$-plate grating unit cells. Each ridge is oriented at 45 deg to the local radial, thereby converting circularly polarised incident light into either radially or azimuthally polarised light on transmission.

now only a single point in period and duty cycle needs to be found to satisfy the required characteristics.

The optical characteristics can be summarised in three different figures of merit:

1. $\delta\phi$ - The phase difference accumulated between the TE and TM polarisations. As it is not important which of the two is the fast or slow axis, the phase difference only spans a range of 0 to pi. Depending on the application, the phase difference usually needs to be as close as possible to either π for a $\frac{\lambda}{2}$-plate, or $\frac{\pi}{2}$ for a $\frac{\lambda}{4}$-plate.

2. δT (or δR) - The difference between the transmission (or reflection) amplitudes between the TE and TM polarisations. Any significant difference in this value results in the output being eliptical in polarisation, rather than the desired circular or linear, so this number needs to be minimised.

3. $< T >$ (or $< R >$) - The average of the transmission (or reflection) amplitudes for TE and TM polarisations. While not crucial for the correct output

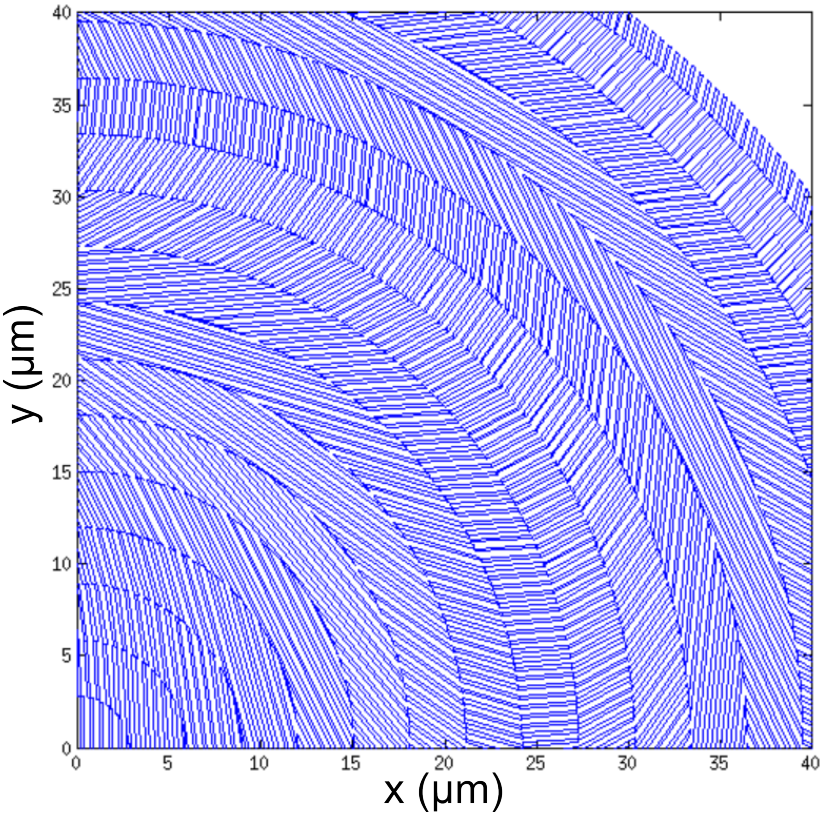


Figure 5.3 – Example design for a geometric phase metasurface for a lens with a focal length of 1mm at a wavelength of 633 nm, based on TiO_2 $\frac{\lambda}{2}$-plate gratings. Only the top right quarter is shown for symmetry reasons. The design shown makes use of the radial symmetry and tiles the surface with concentric circles with a constant width that allows for at least six periods of the grating everywhere in the ring.

beam shape, the average transmission does control the total efficiency. (As for a sufficiently high $< T >$ the δT naturally drops to 0, this is in some ways just a more strict version of number two.)

All three of these conditions are fairly easy to meet, and depending on the purposes, it is not difficult to come up with suitable weightings between them. Finding the best point in a given RCWA map from there is easy (see figure 5.4). However, fabrication tolerance is far from trivial for actual structures, which in practice means there are actually six figures of merit: the previous three as well as their derivatives with respect to duty cycle.

As it turns out, the best points obtained from the figure of merit considerations tend to be at crossings of rather narrow resonance curves where only a slight error in duty cycle can completely change the optical response, so the derivatives perform badly in these points. However, it is often possible to get a much more reasonable fabrication tolerance by making a slight compromise in one or more of the original figures of merit, usually $< T >$.

Once a set of grating parameters has been found that optimises the figures of merit, the metasurfaces can be designed. As mentioned, there are two types of metasurfaces that can be based on them. An example of the first kind can be seen in figure 5.2. This figure shows a grating designed for a Si_3N_4 membrane with $\frac{\lambda}{4}$-plate unit cells around a ring. Each ridge, and therefore the fast or slow axis of the $\frac{\lambda}{4}$-plate, is oriented at 45 deg to a radial from the centre of the rings. The goal of this design is to convert circularly polarised incident light into a radially or azimuthally polarised cylindrical vector beam. Similar designs for $\frac{\lambda}{2}$-plate gratings that convert linearly polarised light into vector beams are less regular and therefore not as suitable for membranes.

An example for a geometric phase metasurface design is shown in figure 5.3. This particular grating consists of concentric ring shaped regions with a single grating orientation per ring. This ideally makes use of the circular symmetry of the lens phase profile it aims for, while ensuring that each ring has the minimum number of period necessary for the gratings to function. For less symmetric phase profiles, for example a focusing vortex phase like the one shown in figure 4.5e, the designs consist of tiling the surface with square or hexagonal unit cells instead.

As most of these structures are not feasible in membrane form, these designs played a large part in our decision to move from Si_3N_4 to TiO_2. However, due to supply issues with the TiO_2 we have not yet been able to experimentally verify metasurfaces of this type.

5.1.4 Fabrication specifics.

Two distinctly different material platforms were used for this work: silicon nitride and amorphous titanium dioxide (crystalline TiO_2 is itself birefringent, which would greatly complicate the current work.) Both materials retain a reasonably high refractive index while being mostly transparent in the visible part of the spectrum (see figure 5.1). Si_3N_4's refractive index turns out to be only high enough when membraned however, as the Δn on a glass substrate is too low, while TiO_2 becomes lossy below approximately 500 nm and is therefore most suitable for green light and wavelengths above.

Much of the following has been limited by what materials we originally had available. These were silicon nitride membranes, made of not quite stochiometric Si_3N_4, supplied by Norcada[101]. The actual material, $Si_{3+x}N_{4-x}$ had slightly more silicon in the structure and as a result also a slightly higher refractive index. However, the extinction coefficient was also significantly higher, making it more lossy than TiO_2 while not quite reaching its refractive index. Still, it did serve for proof of principle experiments. Only near the end did we get access to

Material	DC bias (V)	Pressure (mT)	Gas 1	flow rate (sccm)	Gas 2	flow rate (sccm)	Rate (nm/min)
silicon nitride	355 ± 5	94	CHF_3	58	O_2	2	~ 25
titanium dioxide	270 ± 3	70	SF_6	30	CHF_3	10	~ 195

Table 5.1 – RIE etching recipes used for silicon nitride (both Si_3N_4 and $Si_{3+x}N_{4-x}$) and titanium dioxide (TiO_2).

TiO_2 on glass wafers as well.

Membranes require quite a bit more care than samples with substrates. Notably they do not generally survive ultrasonic agitation and the structures need to be sufficiently connected to the outer supports everywhere to support their own weight. Nitride is a rather strong material and these issues are not as limiting as they would be in most membranes, but they still need to be taken into account; ultimately these membranes led to a very low yield of working samples. To preserve the maximum integrity, the membranes were all used at the thickness we received them, i.e. 500 nm.

The TiO_2 samples, on the other hand are very robust and can stand up to all cleaning methods discussed in chapter 3.2. The wafers, supplied by Silson[102], came with a TiO_2 layer with a thickness ranging from about 450 nm to 550 nm, therefore any samples need to be thinned down to the appropriate thickness before use. While the recipes for this were all developed, no actual working gratings in TiO_2 were fabricated before the end of my PhD.

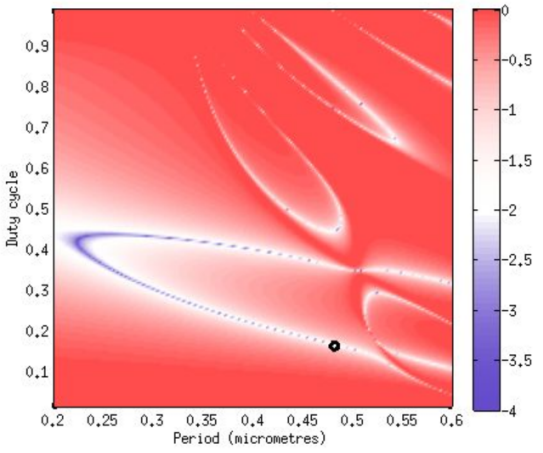
Both materials have a very low electrical conductance and therefore a charge dissipation layer is required during the lithography step. For both materials the resist AR-P 6200.13 was used with the same details as for crystalline silicon as given in table 4.1. The etching recipe for the silicon nitride was developed by Dr Yue Wang[38], while for TiO_2 I developed a recipe myself, which has led to good preliminary results. Both recipes can be found in table 5.1.

5.2 Results

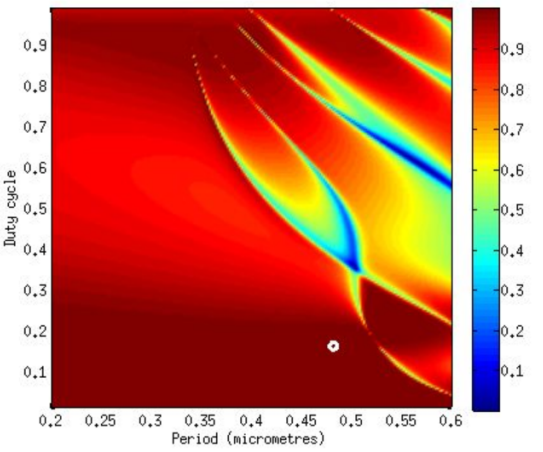
5.2.1 $\lambda/4$- and $\lambda/2$-plates.

The first step towards metasurfaces based on polarisation conversion gratings is to study the waveplates in isolation. To do so, I designed both a $\frac{\lambda}{2}$- and a $\frac{\lambda}{4}$-plate for a 500 nm thick Si_3N_4 membrane using the method described earlier. This thickness limits the wavelengths at which a very high efficiency grating can be found, with the best $\frac{\lambda}{4}$-plate being found for a wavelength of 635 nm while for a $\frac{\lambda}{2}$-plate it was necessary to go down to 355 nm.

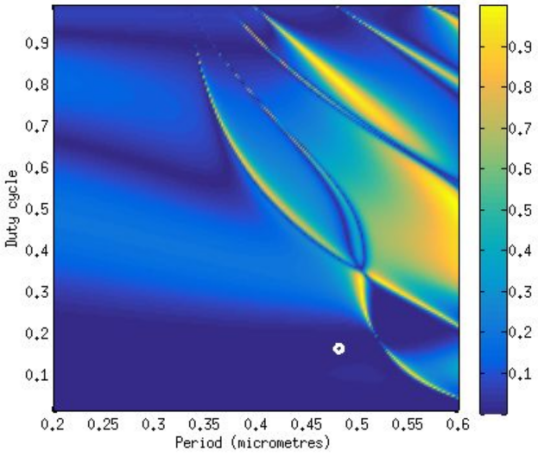
RCWA maps showing the three figures of merit for a 500 nm thick Si_3N_4 mem-



(a) Log of $\delta\phi$.



(b) Average transmittivity.



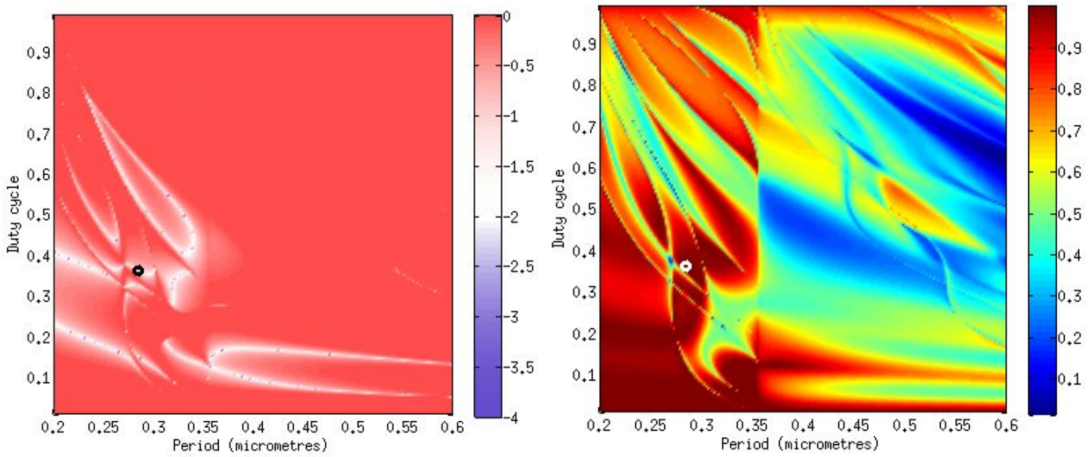
(c) Difference in transmittivities.

Figure 5.4 – RCWA period vs duty cycle maps used for the design of $\frac{\lambda}{4}$-plate gratings. The results shown are: (a) $^{10}\log\left(\left||\phi_{TE}-\phi_{TM}|-\frac{\pi}{2}\right|\right)$, (b) $\left(T_{TM}+T_{TE}\right)/2$ and (c) $|T_{TM}-T_{TE}|$. Indicated in each is the chosen point that serves as the design for *quarterwav*-plates in this chapter. The results shown are for a 500 nm thick membrane of Si_3N_4 illuminated at normal incidence with a wavelength of 635 nm.

brane at a wavelength of 635 nm can be seen in figure 5.4, with the parameters chosen for the simulations indicated by a circle. The $\delta\phi$ plot shows the log of the difference with the ideal of $\frac{\pi}{2}$, which highlights how narrow the best points in the maps generally are. The same plots for a wavelength of 355 nm are given in figure 5.5 with the chosen $\frac{\lambda}{2}$-plate parameters indicated.

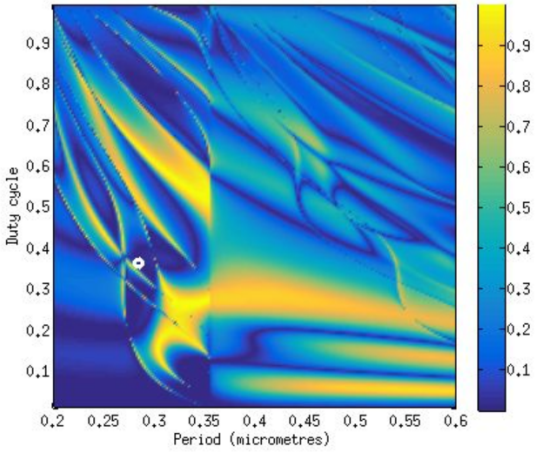
The results of COMSOL simulations for the $\frac{\lambda}{4}$- and $\frac{\lambda}{2}$-plate designs are shown in figures 5.6a and 5.7a, respectively. As these simulations only look at uniform gratings, they make use of periodic boundary conditions and only simulate a single period of the grating, with repeats of that period added for visualisation purposes only.

The illumination consists of a plane wave incident from the top with a phase



(a) Log of $\delta\phi$.

(b) Average transmittivity.

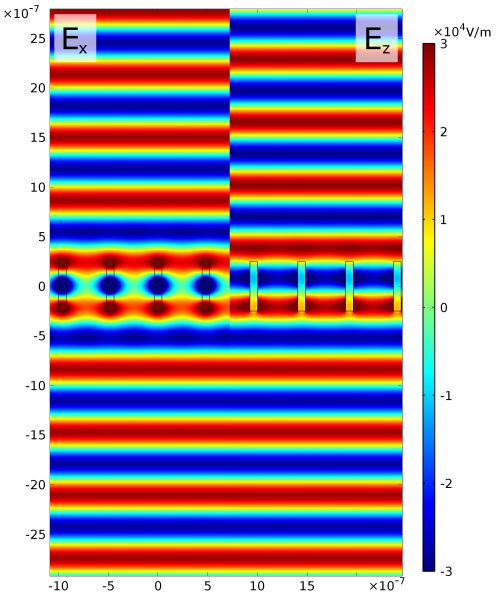


(c) Difference in transmittivities.

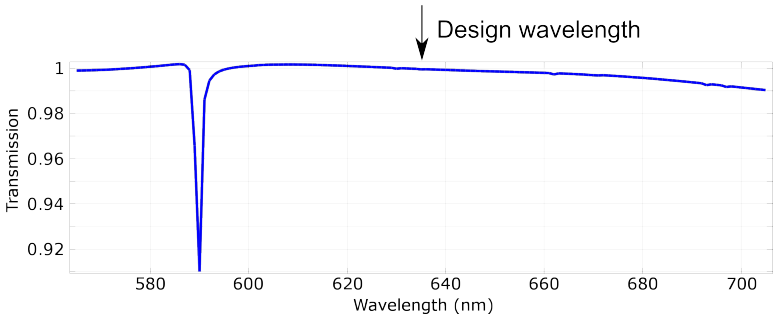
Figure 5.5 – RCWA period vs duty cycle maps used for the design of $\frac{\lambda}{2}$-plate gratings. The results shown are: (a) $^{10}\log\left(||\phi_{TE} - \phi_{TM}| - \pi|\right)$, (b) $\left(T_{TM} + T_{TE}\right)/2$ and (c) $|T_{TM} - T_{TE}|$. Indicated in each is the chosen point that serves as the design for $\frac{\lambda}{2}$-plates in this chapter. The results shown are for a 500 nm thick membrane of Si_3N_4 illuminated at normal incidence with a wavelength of 355 nm.

delay of $\frac{\pi}{2}$ between the in- and out-of-plane electric field components for the $\frac{\lambda}{4}$ design and π for the $\frac{\lambda}{2}$ design. These choices for the excitation simulate circularly polarised incidence for the former and -45 deg polarised for the latter design, which should in both cases lead to linear, 45 deg polarised output if they work as intended.

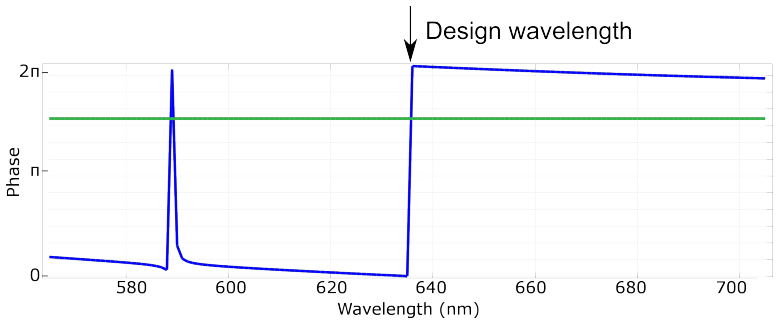
The results are split into the in-plane part of the electric field, E_x, on the left and the out-of-plane component, E_z on the right. The phase difference between the two components is then immediately obvious from the delay between the maxima on both sides, clearly showing a perfect conversion from the illumination polarisation above the grating to completely in phase polarisation on the transmitted side. Moreover, as the amplitude of the fields is identical above and below



(a) Electric field amplitudes.

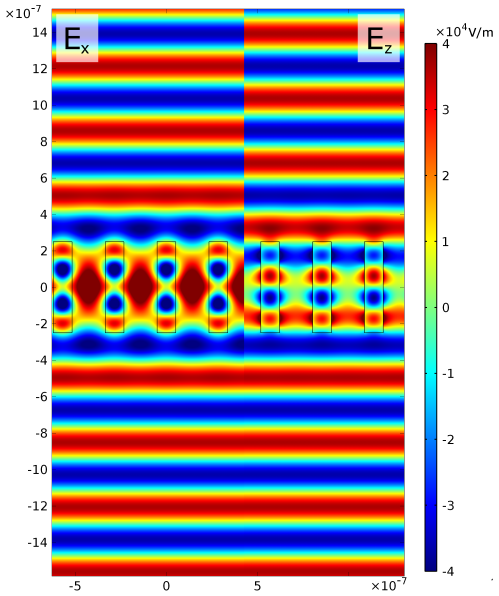


(b) Transmission.

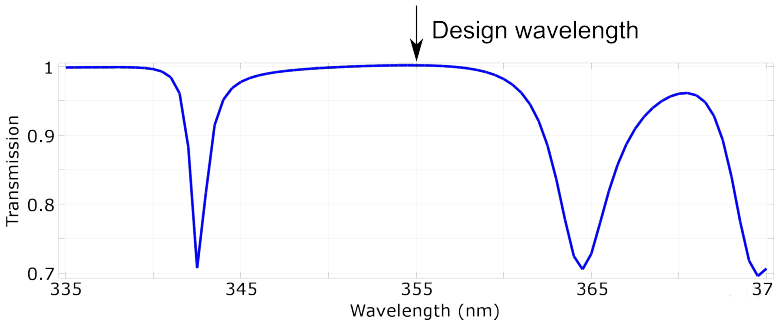


(c) Phase difference.

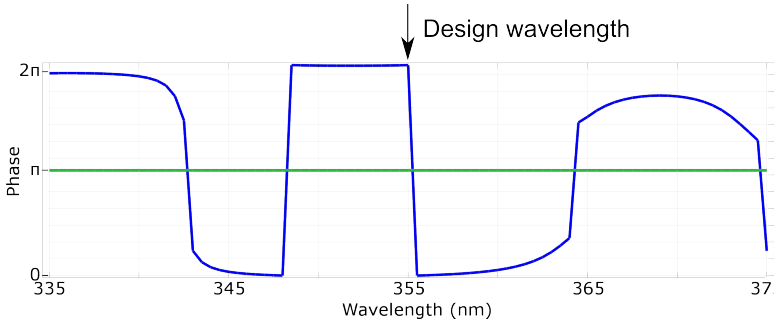
Figure 5.6 – Results of COMSOL simulations of a Si_3N_4 grating designed to work as a $\frac{\lambda}{4}$-plate based on the RCWA maps shown in figure 5.4, with a period of 481 nm, DC 0.16, thickness 500 nm and design wavelength of 635 nm. (a) Shows the amplitudes of the electric fields as a circularly polarised incident wave at the design wavelength of 635 nm passes through the grating from the top to become linearly polarised on the transmitted side, with the in-plane component of the electric field on the left and the out-of-plane on the right. (b) and (c) show the spectral behaviour of the total transmission of the grating and the phase difference between the polarisations, respectively. Indicated in green in the latter is the phase difference of the illuminated wave, for comparison. The sharp peak at 590 nm is a crossing TE resonance. The desired values are for the phase to be close to 0 or 2π and the transmission to be as high as possible. Note that the apparent jump at 635 nm is due to the cyclical nature of phase.



(a) Electric field amplitudes.

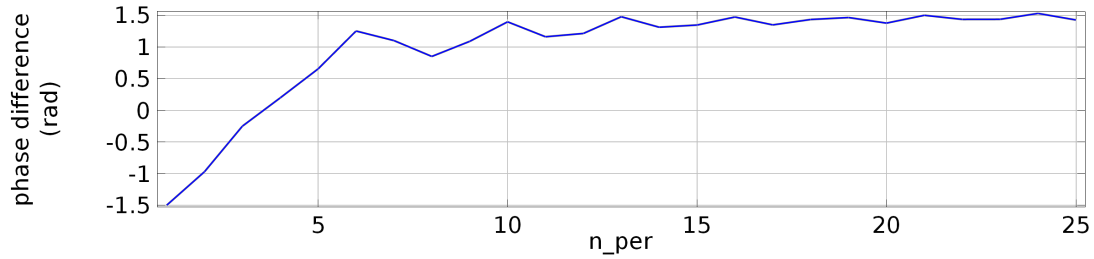


(b) Transmission.

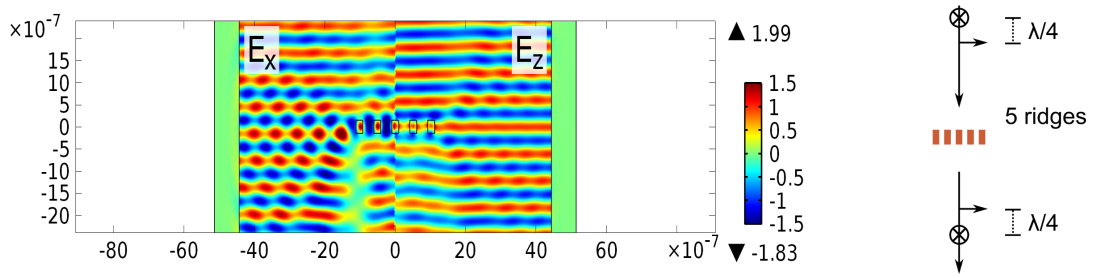


(c) Phase difference.

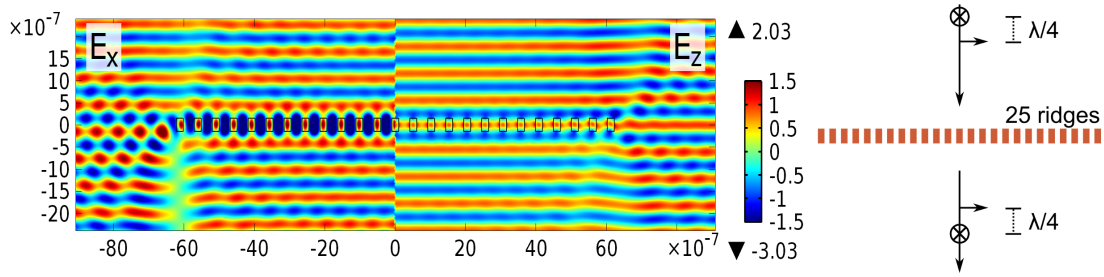
Figure 5.7 – Results of COMSOL simulations of a Si_3N_4 grating designed to work as a $\frac{\lambda}{2}$-plate based on the RCWA maps shown in figure 5.5, with a period of 285 nm, DC 0.36, thickness 500 nm and design wavelength of 355 nm. (a) Shows the amplitudes of the electric fields as a −45 deg linear polarised incident wave at the design wavelength of 355 nm passes through the grating from the top to become +45 deg linear polarised on the transmitted side, with the in-plane component of the electric field on the left and the out-of-plane on the right. (b) and (c) show the spectral behaviour of the total transmission of the grating and the phase difference between the polarisations, respectively, with the useful range from approximately 345 nm to 360 nm. Indicated in green in the latter is the phase difference of the illuminated wave, for comparison. The desired values are for the phase to be close to 0 or 2π and the transmission to be as high as possible. Note that the apparent jumps at 347 nm and 355 nm are due to the cyclical nature of phase.



(a) Phase difference as function of number of periods.



(b) Five ridges.

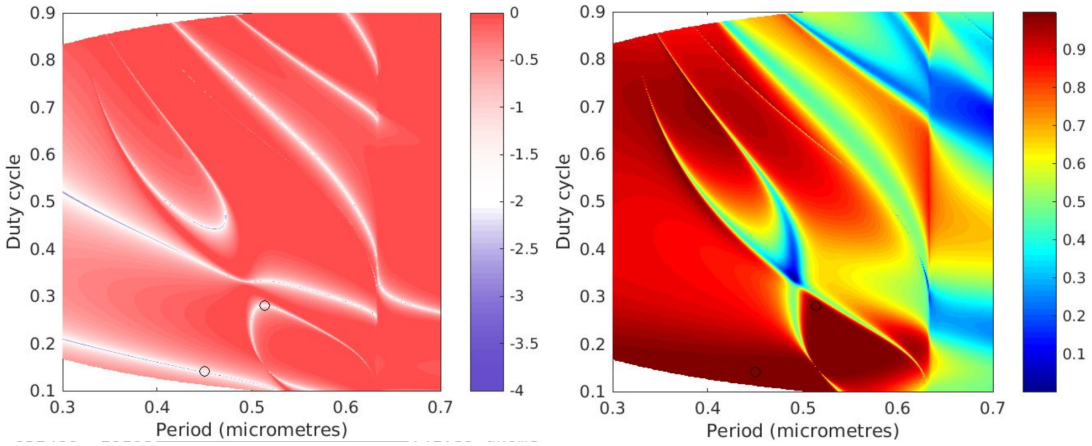


(c) 25 ridges.

Figure 5.8 – Results of 2D simulations of a Si_3N_4 membrane grating with varying numbers of ridges. The results shown are for a 300 nm thick membrane of Si_3N_4 illuminated at normal incidence with circularly polarised light at a wavelength of 610 nm. (a) gives the phase difference between E_x and E_z at the point $x = 0\,\mu\text{m}, y = -2.5\,\mu\text{m}$ as a function of the number of periods included. (b) and (c) are the electric fields for five and 25 ridges, respectively, with E_x on the left and E_z on the right.

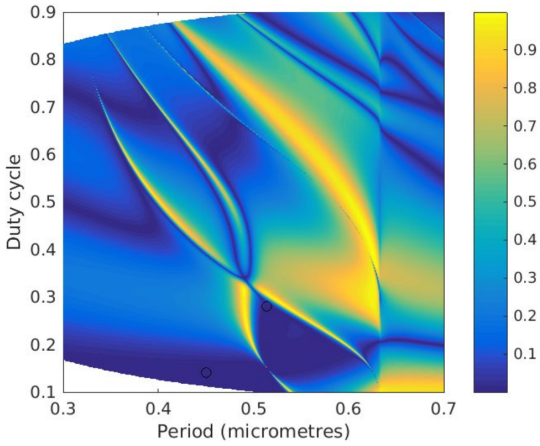
the grating for both E_x and E_z the transmission, and therefore the efficiency, is close to 100% in both simulations.

To show the spectral behaviour of these gratings, a wavelength sweep was performed for both designs. The results can be seen in figures 5.6 and 5.7. This sweep once again shows the direct relationship between wavelength and period, as the behaviours correspond almost exactly to what would be expected from the RCWA in figures 5.4 and 5.5. In the case of the $\frac{\lambda}{4}$-plate grating, the amplitude and phase both change only slowly in period, with a sharp undesired resonance crossing at slightly higher period, or lower wavelength (~ 590 nm) in the COMSOL results. The $\frac{\lambda}{2}$ design on the other hand was found in the centre of a large number of resonances, which is clearly reflected in both the transmission



(a) Log of $\delta\phi$.

(b) Average transmittivity.



(c) Difference in transmittivities.

Figure 5.9 – RCWA period vs duty cycle maps used for the design of $\frac{\lambda}{4}$-plate gratings. The results shown are: (a) $^{10}\log\left(\left||\phi_{TE}-\phi_{TM}|-\frac{\pi}{2}\right|\right)$, (b) $\left(T_{TM}+T_{TE}\right)/2$ and (c) $|T_{TM}-T_{TE}|$. Indicated in each are two chosen points that serve as the design for *quarterwav*-plates utilised in fabricated ring gratings. The results shown are for a 500 nm thick membrane of $Si_{3+x}N_{4-x}$ illuminated at normal incidence with a wavelength of 633 nm.

and phase plots.

As a result, the $\frac{\lambda}{4}$ design is much more robust with good results over a bandwidth of ± 20 nm where the $\frac{\lambda}{2}$ design breaks down after only 5 nm. Similarly, from the duty cycle axis in the RCWA plots, it is to be expected that the $\frac{\lambda}{4}$ grating is less susceptible to fabrication error, though in both cases the phase changes relatively quickly with duty cycle.

These simulations show that, while under ideal conditions, the polarisation converting gratings work extremely well, their spectral bandwidth is very small, ranging from $\sim 1\%$ for the $\frac{\lambda}{2}$-plate design to $\sim 3\%$ for the $\frac{\lambda}{4}$-plate. This is in direct contrast to the broadband nature of the direct phase metasurfaces, and indicates that an experimental realisation will be difficult, but not impossible.

For the design of geometric phase metasurfaces it is important to know how many ridges are necessary for the grating physics to remain valid. To that end, I performed a series of simulations with an example $\frac{\lambda}{2}$-plate grating design where I vary the number of ridges in the system. These simulations are in 2D and simulate the full Si_3N_4 grating surrounded by air on all sides, illuminated with circular polarised incident light at the design wavelength of 610 nm.

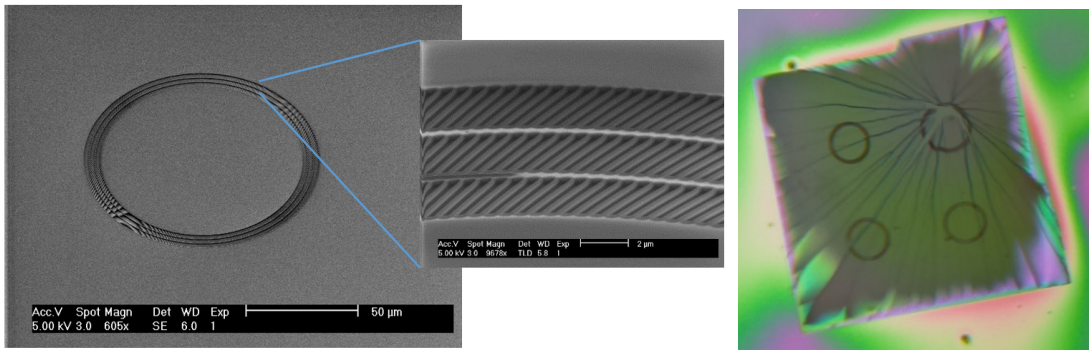
Two example results are shown in figure 5.8, for 25 ridges and 5 ridges. For 25 ridges, the simulation shows the grating working almost exactly the same as with periodic boundary conditions, with only a slight modulation due to diffraction at the edges. At five ridges, the behaviour of a $\frac{\lambda}{2}$-plate can still be recognised, but there is significant distortion in the transmitted fields. For less than five ridges the grating stops working altogether. Therefore, to construct a working metasurface, each unit cell needs to be a the very least five ridges wide, preferably more, which severely limits the gradients of the phase profile they can encode.

The difference between the E_x and E_z behaviours seems to be due to the different natures of the TM and TE resonances. In the example, it can be seen that one of the two (in this case E_x) has a field distribution indicative of a Guided Mode Resonance, i.e. with significantly different field distributions in the ridges and the grooves, while the other (E_z) is laterally uniform. This would suggest that the E_x, or TM, component has a larger lateral propagation length and is more perturbed by edge effects, which also agrees with the examples. The TM mode therefore stops being available when a minimum number ridges is exceeded and starts acting as a single scatterer, while the TE mode is relatively unchanged down to a single ridge. This results in them producing almost the same phase delay for numbers of ridges lower than the GMR propagation length.

While this behaviour is specifically true for this particular system, all other simulations on $\frac{\lambda}{2}$-plate gratings show a similar dichotomy between the TM and TE modes (see for example figure 5.7). That universal feature of a GMR and a Fabry-Perot resonance working together suggests that the behaviour in these simulations is likely to be representative for all $\frac{\lambda}{2}$-plate gratings.

5.2.2 Si_3N_4 ring gratings.

To experimentally test these concepts, we decided to start with polarisation control metasurfaces that convert an incident beam into a Cylindrical Vector beam, specifically a radially or azimuthally polarised beam. At visible wavelengths the 500 nm thick membranes of $Si_{3+x}N_{4-x}$ turned out to be better suited to $\frac{\lambda}{4}$-plate designs, with two reasonable candidates (see figure 5.9). I first attempted to fabricate gratings with a period of 450 nm and a duty cycle of 0.14



(a) SEM image of ring grating.

(b) Optical image of membrane with ring gratings.

Figure 5.10 – Fabrication results of ring gratings in $Si_{3+x}N_{4+x}$ membranes. (a) shows an SEM image of the successfully fabricated ring grating used for figure 5.12. (b) shows an optical image that illustrates a common issue with fabricating gratings in membranes.

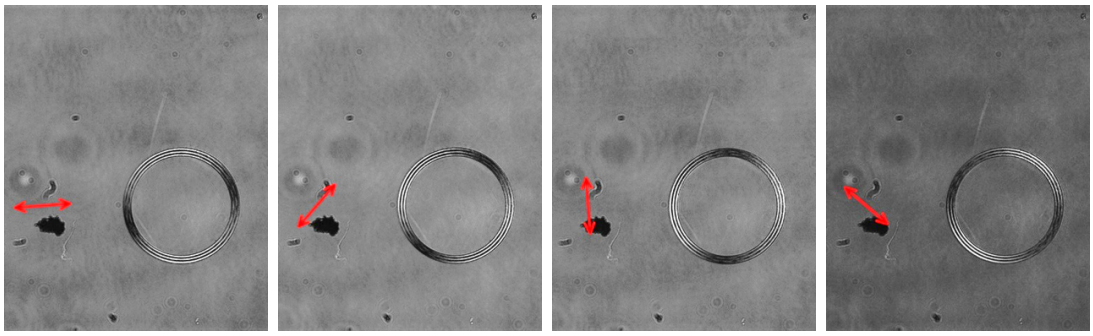
(the lower left point indicated in the maps), but these turned out to weaken the membranes too much and almost always broke (see figure 5.10b.) The results in this section are therefore based on gratings with a design period of 514 nm and duty cycle of 0.28.

$\frac{\lambda}{4}$-plates convert circularly polarised light to linear polarised light at ± 45 deg to both the slow and fast axes of the grating. (The handedness of the circular polarised incident light determines the sign.) To use these gratings for generation of radial or azimuthal polarised light, they need to be arranged around the circle in such a way that the fast axes, i.e. the grating ridges, are always oriented at 45 deg to a line from the centre of that circle. The handedness of the incident light then determines whether the output beam is radially or azimuthally polarised.

There are two primary concerns that complicate these designs. First, the circumference and period together determine the number of ridges around the ring, but the circumference changes with radius while the period ideally doesn't. To compensate, extra ridges need to be added for increasing radius. Second, as the designs are intended to be fabricated in membranes, the structural integrity of the membranes needs to be preserved, so ridges can not just appear out of nothing. To solve both of these issues in one go, instead of making one continuous grating we chose to make three concentric ring gratings with a thin ring of unpatterned silicon nitride in between them.

Examples of the fabricated gratings can be seen in figure 5.10. The optical image on the right shows how a slight overdose in only one of the gratings can compromise the structural integrity of an entire membrane. The SEM image on the left shows a successfully fabricated ring grating with a high aspect ratio.

The fabricated gratings where then experimentally tested in a setup much



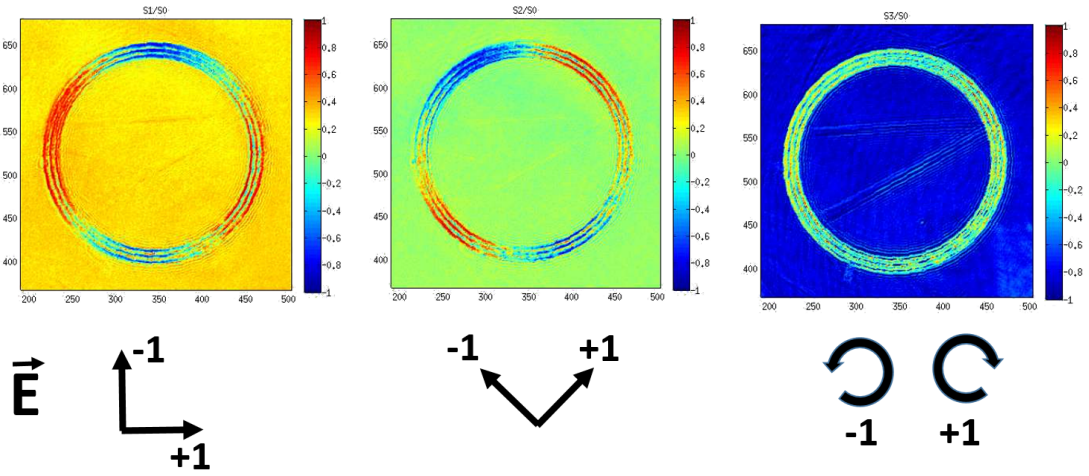
(a) 0 deg **(b)** 45 deg **(c)** 90 deg **(d)** 135 deg

Figure 5.11 – Transmission measurements on rings of $\frac{\lambda}{4}$-plate gratings oriented everywhere at 45 deg to the radial. The rings were illuminated with right-handed circularly polarised light at the design wavelength of 633 nm and imaged through a linear analysing polariser oriented at different angles as indicated in the figures in red.

like before, but including several combinations of polarising analyser and quarter wave plate. The incident, linearly polarised light is sent through a quarter wave plate at 45 deg to convert it to circular polarised light before reaching the grating. On the transmitted side, sixteen permutations of the measurement setup are used per grating. First, the transmitted light is imaged through an analysing polariser stepped in 45 deg steps through a full circle. Results of this for right-handed incident light are shown in figure 5.11, which shows a clear azimuthal polarisation on the ring. Second, the same is done again with a $\frac{\lambda}{4}$-plate between the gratings and the analyser oriented at 90 deg to its counterpart on the incident side.

The sixteen images are then processed and combined to extract the four Stokes' parameters of the light transmitted through the gratings[62, 95, 103]: S_0 is defined as the sum of the intensities measured for two orthogonal polarisations, or, for decreased noise, the sum of all eight positions of the analysing polariser without the quarter wave plate divided by four; S_1 is a measure for the linearity of the polarisation in an $x - y$ coordinate system, calculated through the subtraction of the linear polarisation intensities measured along one axis from those measured at 90 deg to that axis; S_2 is the same for the 45 deg rotated coordinate system; and S_3 finally is a measure for the circularity of the light's polarisation, by passing the transmitted light through the quarter wave plate before the analyser, then subtracting the intensities measured with the analyser at 45 deg to the quarter wave plate's fast axis from those measured with the analyser at -45 deg, which is effectively subtracting the transmitted left-handed from right-handed.

The Stokes' parameters can be reduced one step further by dividing S_1, S_2 and S_3 by S_0, which normalises the results. The resulting three images now show the degree of polarisation at any point in the image on a scale of -1 to 1. Results



(a) $S1/S0$ **(b)** $S2/S0$ **(c)** $S3/S0$

Figure 5.12 – Results of experimental polarisation state measurements on a fabricated $Si_{3+x}N_{4-x}$ ring grating. The pictures show the normalised spatially defined Stokes parameters. (a) indicates the degree of polarisation along the x and y axes of the image, with 1 indicating entirely x-polarised, -1 entirely y-polarised, and 0 equal amounts of both. (b) is the same for 45 deg rotated axes. (c) Shows the degree of circularity of the measured polarisation.

for a fabricated ring (the same one as in figure 5.11) with left-handed incident light are shown in figure 5.12. In this case, the transmitted light through the ring has been almost entirely converted to radially polarised light. The high degree of conversion is especially clear from the third image, S_3/S_0, where everywhere outside the gratings the transmitted light is entirely left-handed polarised (-1), i.e. unchanged from the incidence, but on the grating it is linear everywhere.

6 Active metasurfaces.

The structures I discuss in this chapter do not fall exactly under the traditional definition of a metasurface, as they do not convert an incident beam profile into a desired outgoing beam profile. An appropriate term for them might be *active* metasurface. These structures are intended to simultaneously act as a feedback resonator for a distributed feedback (DFB) laser, as well as controlling the beam shape of the emitted laser. A schematic of the type of organic gain material-based DFB lasers used here is shown in figure 6.1[104, 105].

The principle behind these active metasurfaces is still grating based. However, rather than tuning a phase delay between different periods or unit cells as in the previous chapters, the key here is that in a sufficiently coherent DFB laser, all unit cells can be considered as emitting in the same phase[106, 107]. By slightly shifting sections of the grating with respect to others, it is then possible to alter the far field emission pattern that results from the structure. As we do not have the facilities to fabricate and test the full DFB structure in York, the work in this chapter was performed in close collaboration with the University of St. Andrews, mainly James Glackin.

This chapter consists of three sections. The first discusses the concept and the design method associated with active metasurfaces. The second shows the results achieved for this project so far. The final section in this chapter is related work, where we spatially vary the spectral properties of the laser with the intent of creating a coherent broadband light source.

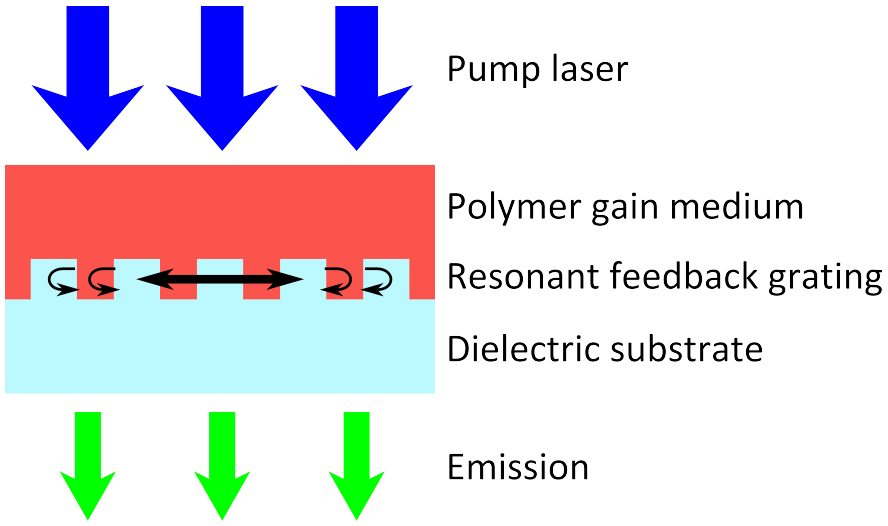


Figure 6.1 – Diagram of the operations of an organic distributed feedback (DFB) laser. A pump laser excites resonant modes in the grating region, which partially reflect back upon themselves to create a resonant feedback effect that induces lasing in the polymer gain medium.

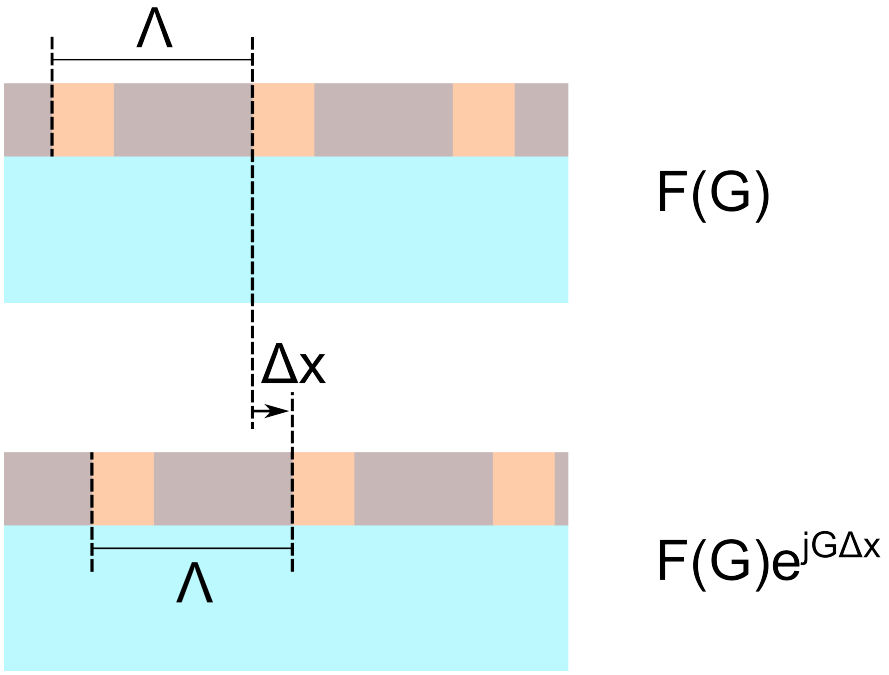


Figure 6.2 – Schematic illustration of the core concept of modulating the far field output by shifting gratings. A small shift Δx of the grating elements adds a phase factor $e^{jG\Delta x}$ to the Fourier transform, which expresses itself as a constant phase factor added to the beam. This phase factor has no relevance for an isolated grating, but allows us to exploit interference effect in more complex grating geometries.

6.1 Theory and design.

The concept used here is based on earlier work in the field of DFB lasers that used Fourier analysis to design substructured gratings that would lower the lasing threshold[107]. The core theory behind both that work and active metasurfaces can be found in the relation of phase and position within the unit cell. Specifically, I make use of the fact that a shift of the ridge within the unit cell leads to a proportional phase shift in the far field.

To illustrate this effect, take a single unit cell of a linear binary grating and Fourier transform it. Then do the same for a unit cell with the same period and duty cycle but the centre of the ridge shifted by a short distance, Δx. The exact form of the Fourier transform is unimportant, but the two now only differ by a factor $e^{jG\Delta x}$, with G the grating vector $\frac{2\pi}{\Lambda}$. The Fourier transform of the gratings therefore includes a constant phase factor that cycles through 0 to 2π as the ridge moves through a period. (See figure 6.2.)

This phase has very little influence under normal circumstances, but as with the geometric phase metasurfaces in the previous chapter it becomes relevant when several shifted gratings are superimposed. Under those circumstances, the phases between the different gratings interfere and allow us to shape the wavefront. To demonstrate the concept of active metasurfaces we chose to initially

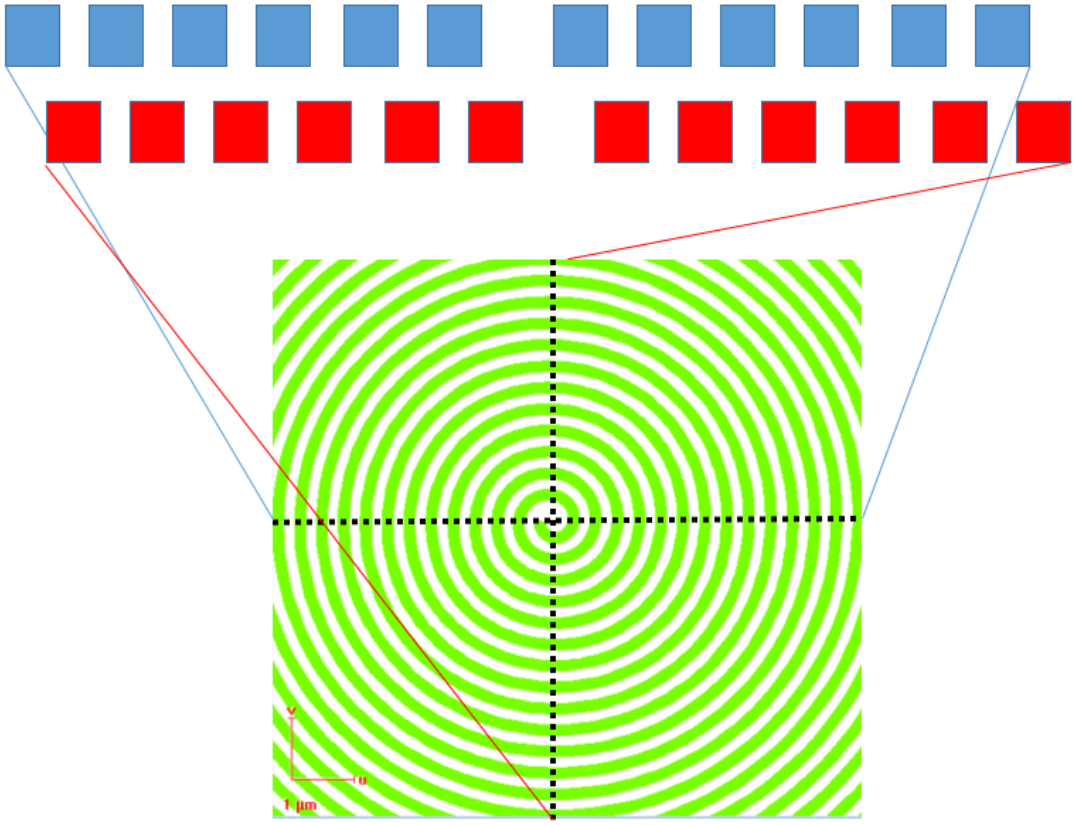


Figure 6.3 – Example of the design process based on the concept shown in figure 6.2. The total grating consists of gratings that are shifted with respect of the origin such that any two opposite sections are $\Lambda/2$ apart.

focus on the direct generation of vortex beams from an organic DFB laser.

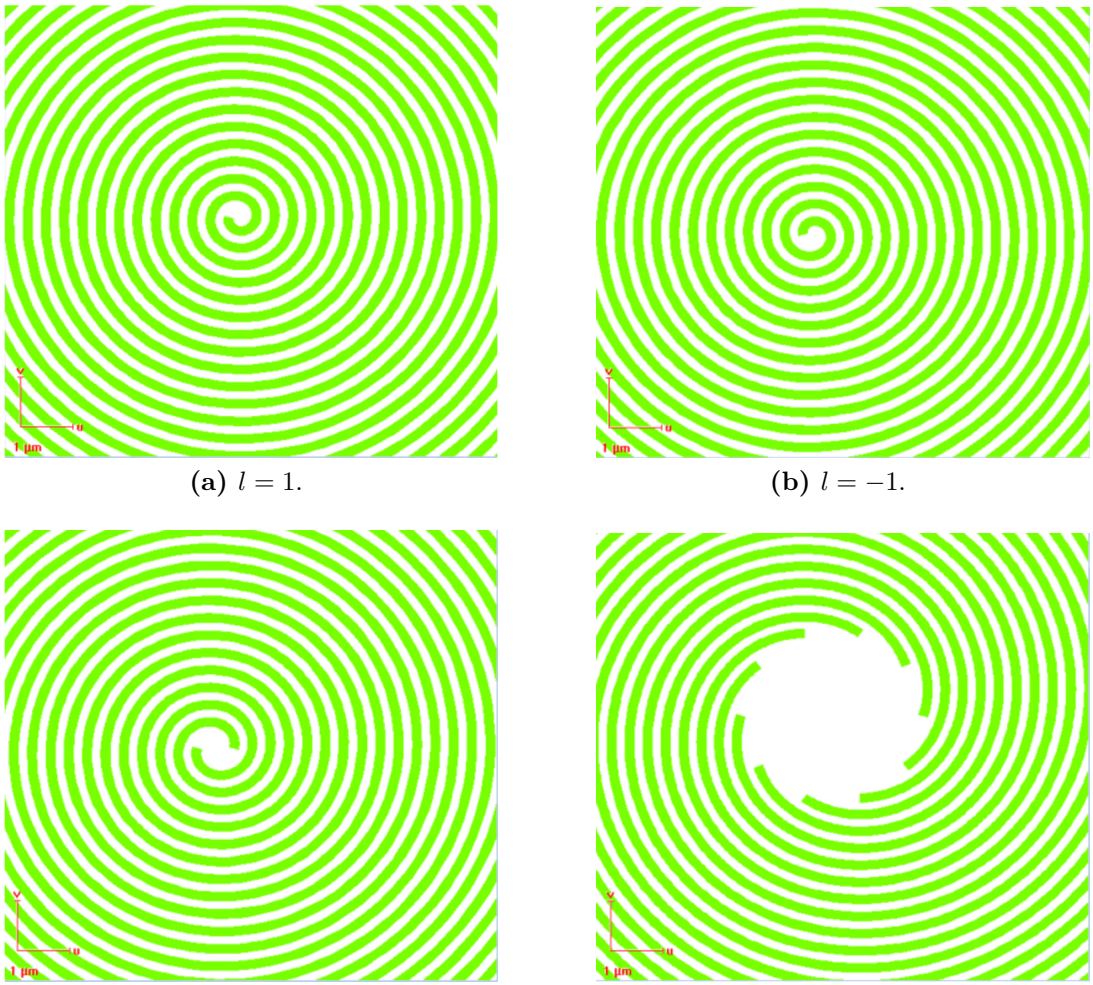
To generate a vortex beam directly from a DFB laser based on this concept, the design process is straightforward. As the phase of any given grating is given by $\phi = G\Delta x = 2\pi\frac{\Delta x}{\Lambda}$, and this only works if the period is constant, we can directly replace ϕ in the phase profile of the vortex beam:

$$\phi = 2\pi\frac{\Delta x}{\Lambda} = l\theta, \text{ or} \tag{6.1}$$

$$x_0 = \frac{l\theta\Lambda}{2\pi}, \tag{6.2}$$

with l the azimuthal mode number of the vortex beam and θ the azimuthal angle in the plane of the grating. Consequently, the phase profile is simply a shift of the grating with respect to the centre that is linearly proportional to the azimuthal angle.

This concept is shown schematically in figure 6.3. A discontinuity in the grating forms at the origin, with either side of the origin along any single dimension being shifted by exactly $\Lambda/2$ in the case of $l = \pm 1$. As the change is continuous



(a) $l = 1$. **(b)** $l = -1$. **(c)** $l = 2$. **(d)** $l = 10$.

Figure 6.4 – Example grating designs to directly generate vortex beams of various azimuthal mode orders from a DFB laser. The different spiral designs (a) through (d) correspond to $l = 1$, -1, 2 and 10, respectively.

and for any integer l moves $\Delta x \to \Delta x + |l|\Lambda$, the result is a perfect spiral design. At any particular section of the grating this is almost indistinguishable from the layout of a concentric circle grating, but taken with the diametrically opposite section, they are exactly out of phase.

For higher order spirals, a given ridge going around the circle will have shifted by multiple periods. By starting a number of ridges equal to the mode number equally spaced around the circle, the spirals stay exactly parallel and the period everywhere is preserved. As an example several designs for $l = 1$, -1, 2 and 10 can be seen in figure 6.4.

So far this section has only talked about gratings in general, without specifying the actual grating parameters. The reason for this is that they simply do not matter for the phase modulation of this type. The only important fea-

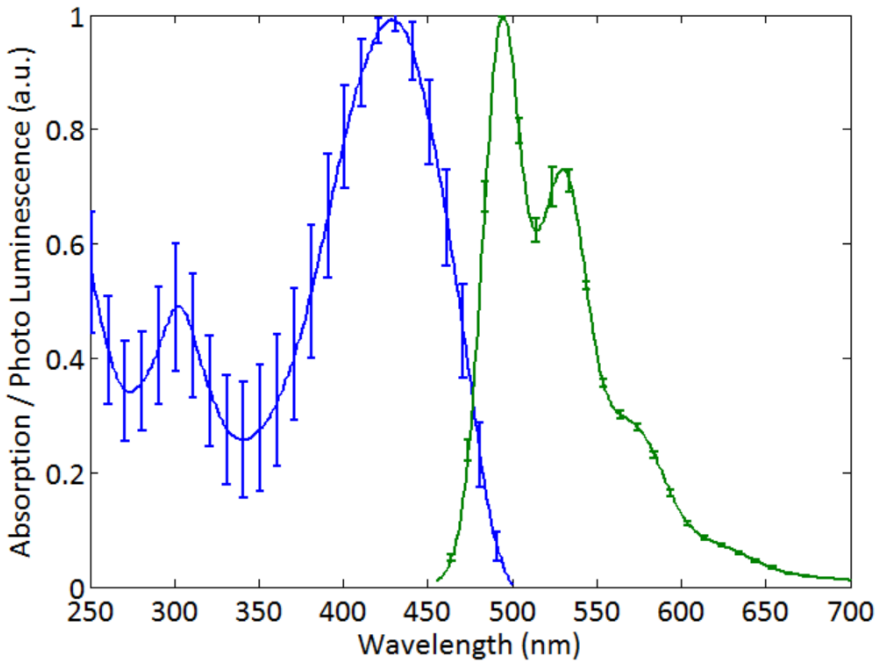
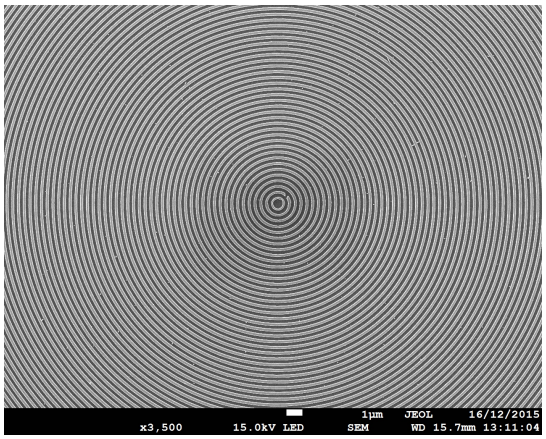


Figure 6.5 – Typical absorption (blue) and photo luminescence (green) curves for the type of polymer (BBEHP-PPV) used in the DFB lasers in this chapter. Each batch of polymer is slightly different. The curves shown here are the normalised average of five such batches[104].

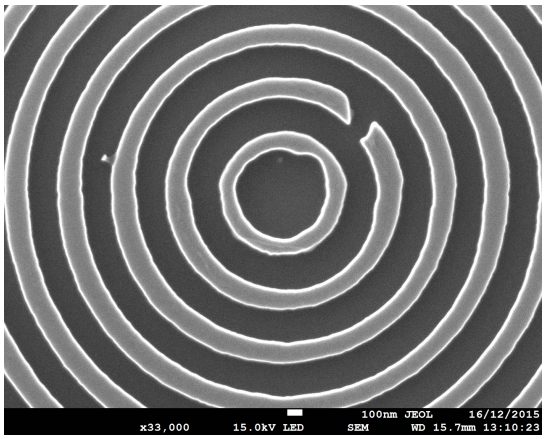
ture is the periodicity and the shift relative to the period. The specific grating parameters are then obtained by the need to generate feedback for an organic laser.

The grating parameters can be separated into several categories: those that influence the resonance quality and those that set the resonant wavelength. In the former category the duty cycle and etch depth are the most important. A full discussion of DFB resonators is beyond the scope of this thesis, but in general these two parameters can be adjusted to find a balance between the resonance Q-factor and the feedback, which greatly influence the threshold and the coherence length. The resonant wavelength on the other hand is fixed by the period and the effective index[104, 105, 108, 109].

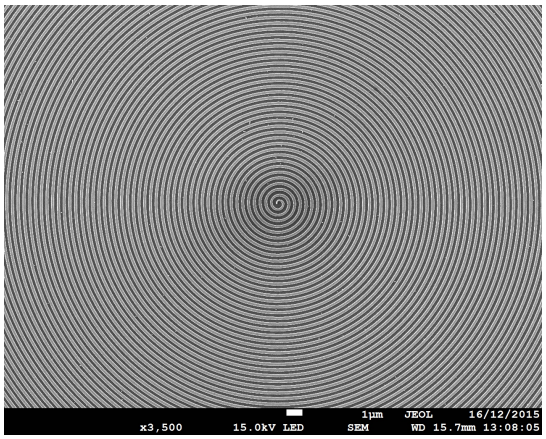
The optimal gratings for active metasurfaces would have long propagation lengths, and therefore high spatial coherence. However, with the chosen grating material of glass and a polymer gain medium the refractive index modulation is generally very small, so this condition is easily satisfied. The grating parameters also need to be chosen such that the resonant wavelength overlaps the high gain region of the polymer gain medium, but again this tends to be a fairly broad region so no fine-tuning is necessary.



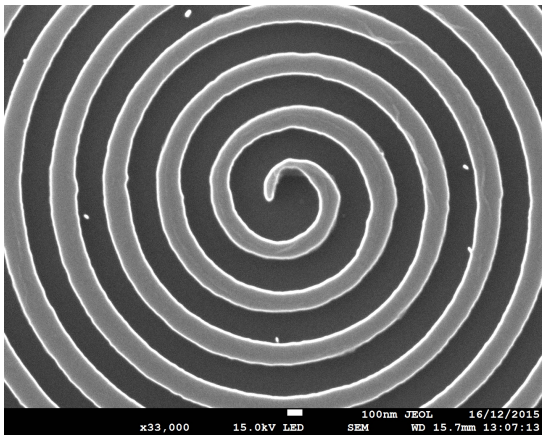
(a) Concentric circle grating ($l = 0$).



(b) Zoom on $l = 0$.



(c) Spiral grating ($l = 1$).



(d) Zoom on $l = 1$.

Figure 6.6 – SEM images of fabricated DFB gratings in silicon. The gratings shown here were used as masters for the nano-imprint of the actual DFB resonance gratings in glass. The gratings have a period of 350 nm and a designed duty cycle of 0.5. (a) shows a concentric circle ($l = 0$) grating, with (b) a zoom on the centre. (c) and (d) are the same for a spiral $l = 1$ grating. The apparent distortions along the horizontal and vertical radii in (a) and (c) are due to a Moiré pattern occurring between the grating periodicity and the SEM scan frequency.

6.2 Fabrication and results.

The gratings in this project are surface relief gratings, etched approximately 80 nm deep into a bulk glass substrate with a refractive index of approximately 1.5. The gratings were designed and fabricated by myself in York. I made several attempts at etching the gratings into glass, but was unsuccessful for reasons I was not able to resolve. Therefore, we decided to instead fabricate the gratings into bulk silicon samples. These silicon gratings were then used in St. Andrews as masters in a further nano-imprint fabrication step to create the eventual glass gratings. (Nano-imprint is a low resolution lithography technique that involves a combination of mechanical shaping and photolithography.) On these glass

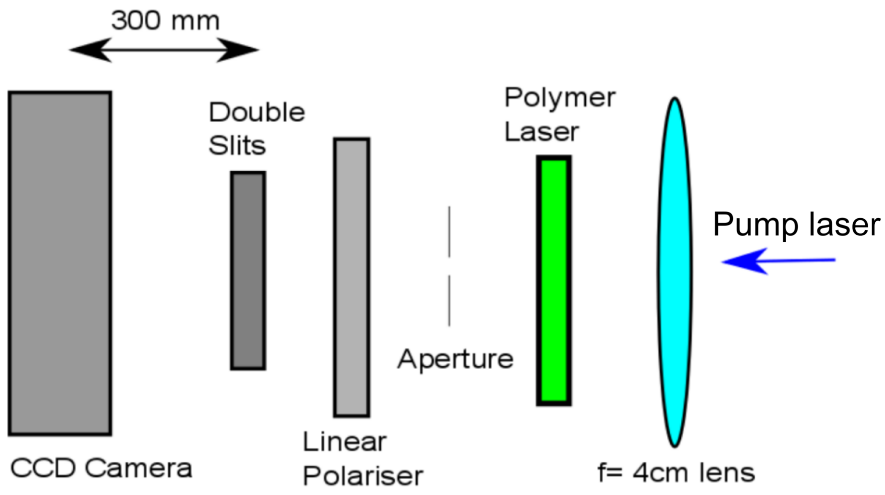


Figure 6.7 – Diagram of the experimental setup used in the current chapter. A pump laser (OPO) is focused onto the back of a sample DFB laser. The emitted beam is then imaged by a CCD camera, with various combinations of a linear polariser, pin hole and double slit aperture added for some of the measurements.

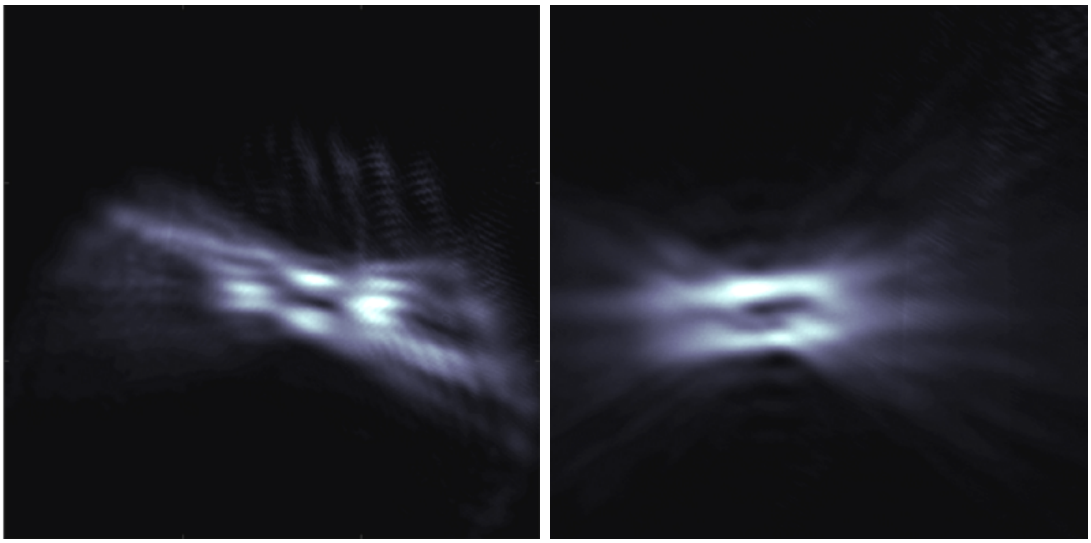
substrates a layer of polymer that acts as the gain medium was then spun using a spin coater.

The polymer used in this work is called BBEHP-PPV. This is a custom made polymer of which the precise make-up changes from batch to batch, but typical absorption and photo luminescence (PL) curves are shown in figure 6.5. The notable features in these two curves are the absorption peak near 430 nm and the broad peaks in the (PL) at 490 nm and 540 nm. The former sets the optimal wavelength for the pump laser, a tunable OPO. The latter suggests that for efficient lasing the grating period should be approximately $\frac{490\,\mathrm{nm}+540\,\mathrm{nm}}{2n_{eff}} \approx 345\,\mathrm{nm}$. To allow for a bit of spread in effective index the designed periods where chosen to be 340 nm, 350 nm and 360 nm.

Since the duty cycle and etch depth are relatively unimportant for the phase modulation, these were chosen as standard sizes used for DFBs with no further optimisation. The designed duty cycles were 0.3, 0.5 and 0.7, with the target groove depth approximately 80 nm.

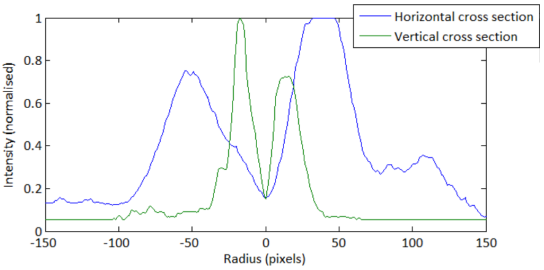
The full sample included $l = 0$ (or circular) and $l = 1$, 2 and 3 (spiral) gratings. SEM images of a circular and $l = 1$ spiral grating on the final fabricated silicon master are shown in figure 6.6. While the quality of the grating is not perfect, especially near the centre, the low magnification images on the left seem to show a high degree of uniformity over most of the gratings.

The complete DFB lasers were tested in St Andrews by James Glackin, using the experimental setup shown schematically in figure 6.7. The polariser, pin hole and double slits where added in various combinations for later measure-

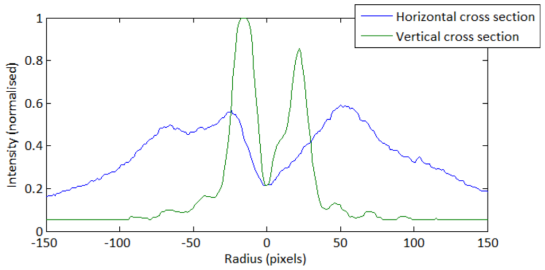


(a) Concentric circle grating ($l = 0$)

(b) Spiral grating ($l = 1$)



(c) Concentric circle grating ($l = 0$), cross sections.



(d) Spiral grating ($l = 1$), cross sections.

Figure 6.8 – Images of the emitted beam from DFB lasers using two different resonant grating designs. (a) is designed as a grating of concentric circles, while (b) is a spiral grating with $l = 1$. Both gratings have a period of 350 nm and a designed duty cycle of 0.7. (c) and (d) show horizontal and vertical cross sections through the centre of each beam.

ments. The first step consisted of imaging the emitted beam from several gratings to study the intensity profiles. The results for a circular and $l = 1$ spiral grating are shown in figure 6.8. While the intensity profile of the spiral grating does, as intended, show a dark centre spot, so does the circular grating. Moreover, both beam profiles show a large, unexpected and undesired asymmetry.

To find the source of the asymmetry the gratings were rotated within the setup. As the beam profiles rotated along with the samples, the asymmetry seems to be inherent to the fabricated DFB lasers, so not an issue with the alignment of the experimental setup. This would indicate that in contrast to what the SEM images in figure 6.6 seem to show, the grating is in fact not uniform in all dimensions. As we were not expecting any significant asymmetry when fabricating the silicon masters this was not extensively investigated at the time. However, later results when attempting to fabricate similar gratings in

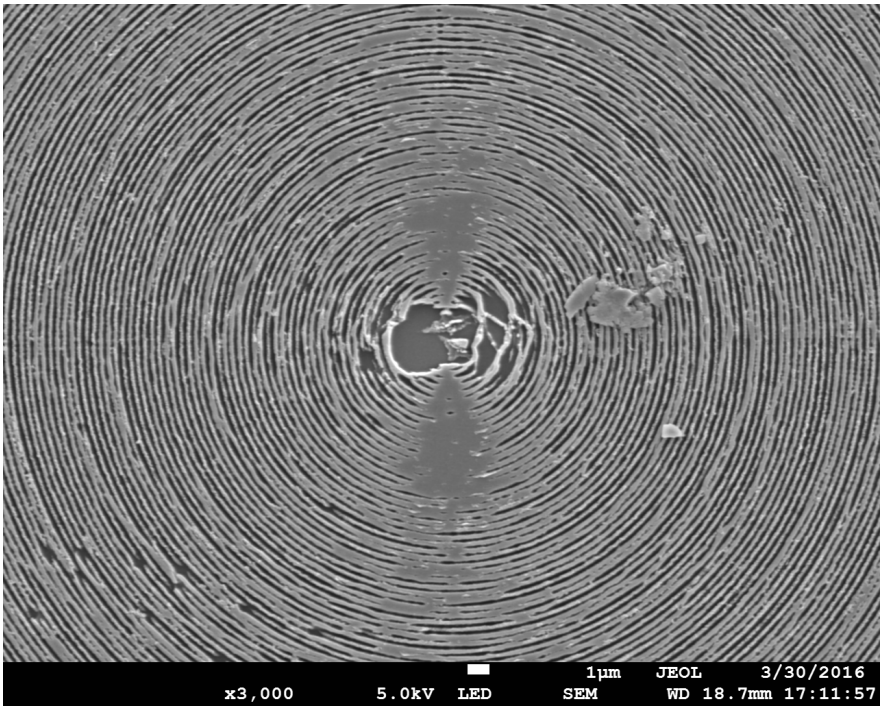


Figure 6.9 – SEM image of a spiral grating etched into glass through an underdosed mask. Note the different duty cycle between the horizontal radials and the vertical ones.

glass by Dr Yue Wang show that different radials of the grating have different clearing doses (see figure 6.9.) It is very possible that this was also the case for the gratings used in these experiments, resulting in small but significantly different duty cycles along different directions, and therefore different lasing efficiencies.

The presence of the central dark spot in both the spiral as well as the circular grating designs is more fundamental, however. It stems from the polarisation symmetry of the emitted beam. As the lasing resonator axes follow the grating period, the polarisation of the resonating light when lasing is determined entirely by the local direction of the grating periods, as illustrated schematically in figure 6.10. The result is that for both circular and spiral grating designs the actual emitted beam is a cylindrical vector beam, azimuthally polarised to be exact[110]. As with a vortex beam, this results in a central dark spot in the emitted beam.

To separate the effect of the polarisation and the phase on the beam profile and thereby determine whether the gratings function as intended, a further experiment using diffraction from a double slit was conducted[111, 112]. When a beam with an azimuthally varying phase impinges onto two slits, there is a significant phase difference between the top and bottom of each slit, and to a lesser degree between the two slits as well. The interference pattern should therefore be different than for a beam without an azimuthal phase component.

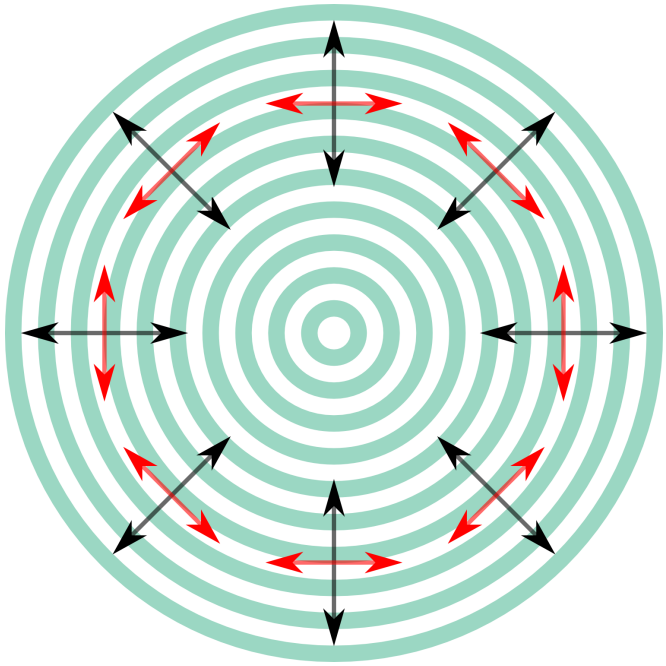


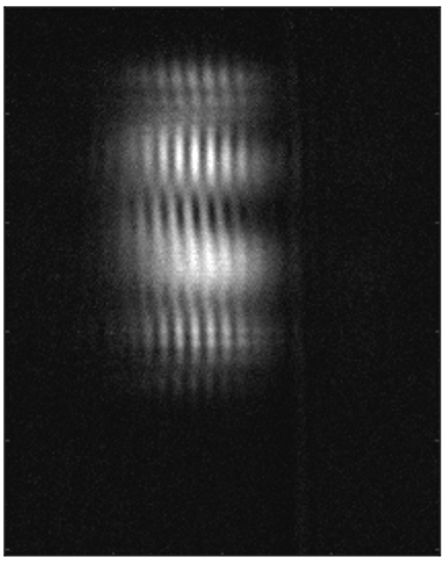
Figure 6.10 – Diagram of the resonant behaviour of a concentric circle grating. The black arrows indicate the propagation direction of the resonant modes, or k, perpendicular to the grating ridges. The red arrows indicate the electric field orientation of those resonances, which correspond to the polarisation distribution of the emitted beam. Here we consider the resulting vector beam as azimuthally polarised.

To simulate the expected interference patterns rigorously requires the full Kirchoff diffraction integral as described in chapter 2.1. However, as long as we limit ourselves to the far field, it suffices to only take a Fourier transform of the aperture plane, which is much easier to implement. This is a scalar approximation, however, while the beams are vectorial in nature. In order to obtain a reasonable approximation of the expected fields we therefore first need to separate the fields on the apertures into two orthogonal polarisation components, do a Fourier transform on each separately, and then recombine the resultant intensities. An azimuthally polarised plane wave with azimuthal phase component can be decomposed as

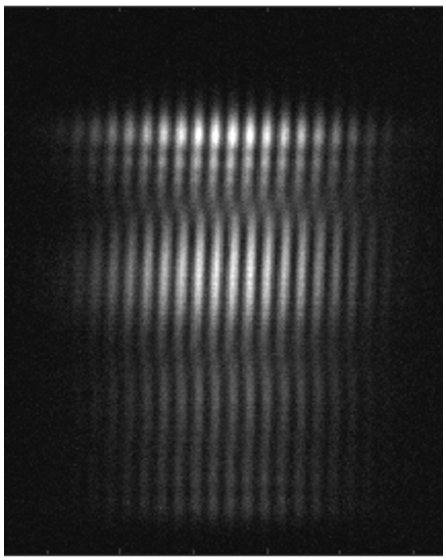
$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \sin(\theta) \\ \cos(\theta) \end{pmatrix} e^{jl\theta}. \tag{6.3}$$

So to get an approximation of the far field interference pattern, E_x and E_y from this description were separately multiplied by an aperture transmission function and Fourier transformed, before being recombined into the total intensity.

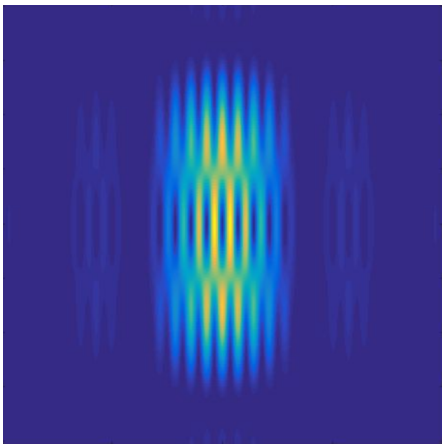
Both preliminary measurement and simulation results are shown in figure 6.11, for a circular grating and a spiral $l = 1$ grating. The results so far are not conclusive either way, with neither the circular nor the spiral grating measurement results matching either of the simulated interference patterns particularly well.



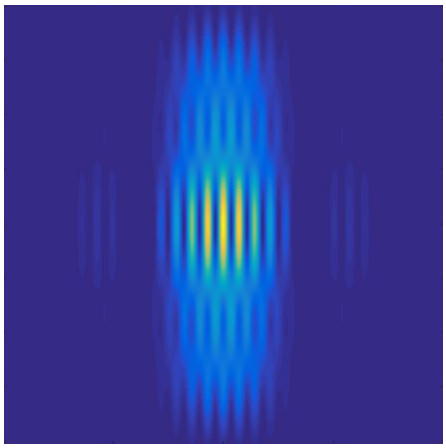
(a) Measured diffraction circular ($l = 0$) grating.



(b) Measured diffraction spiral ($l = 1$) grating.



(c) Simulated diffraction $l = 0$ plane wave.

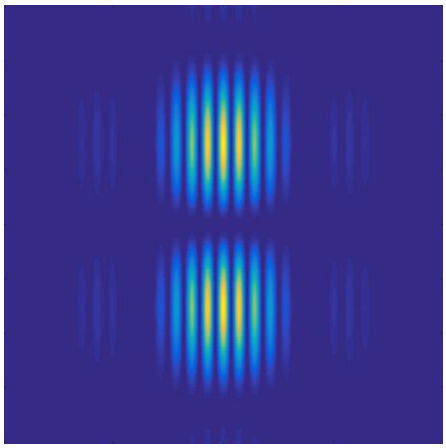


(d) Simulated diffraction $l = 1$ plane wave.

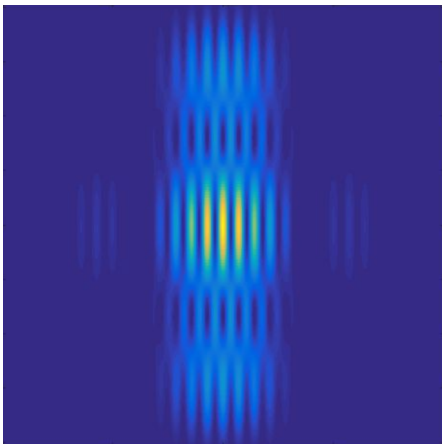
Figure 6.11 – (a) shows the measured diffraction pattern resulting from a beam emitted by a circular grating DFB laser after passing it through a double slit setup. (b) shows the same measurement for an $l = 1$ spiral grating. (c) and (d) show approximations of the expected diffraction patterns of azimuthally polarised plane waves with an azimuthal phase profile of $l = 0$ and $l = 1$, respectively.

However, even in the simulated patterns the differences between circular and spiral are less pronounced than we originally expected.

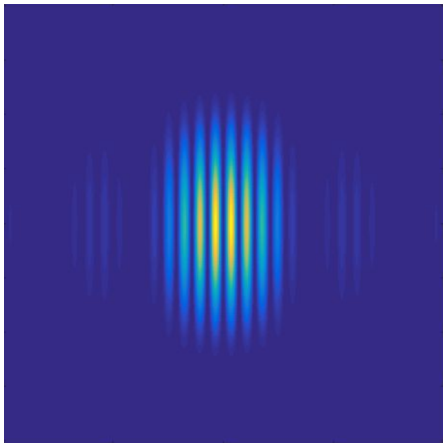
This same approach of approximating the full interference pattern with two orthogonal polarisations leads to a useful observation, however. To separate the polarisation and phase influences on the interference patterns, one straightforward method is to pass the emitted beams through a linear polariser before the double slits. With the natural orientations of the polariser along or across the slits, this gives two distinct interference patterns per beam that change much more



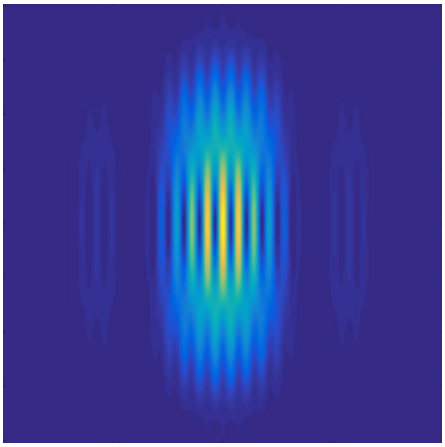
(a) $l = 0$ beam, $|E_x|^2$.



(b) $l = 1$ beam, $|E_x|^2$.



(c) $l = 0$ beam, $|E_y|^2$.

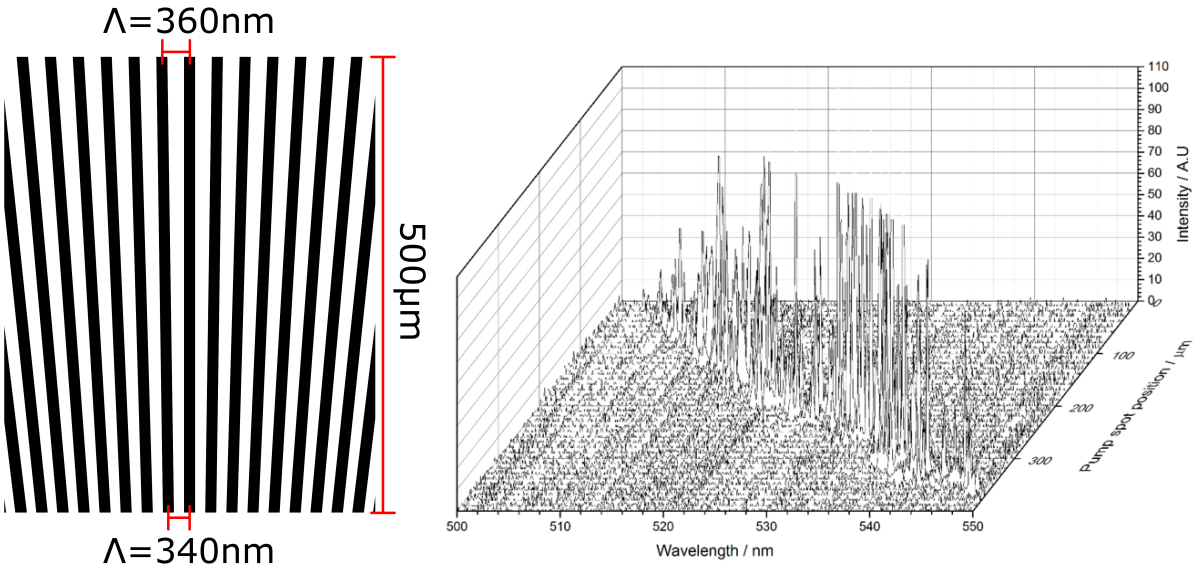


(d) $l = 1$ beam, $|E_y|^2$.

Figure 6.12 – Simulated intensities of the interference pattern of a double slit aperture. In all cases the original beam is an azimuthally polarised beam as described in equation 6.3, but filtered to only include the fields polarised across the slits (E_x) or along the slits (E_y). (a) shows the diffraction of the E_x component for $l = 0$, (b) the same for $l = 1$. (c) and (d) show the corresponding E_y components. (Figures 6.11c and 6.11d show the sums of (a)+(c) and (b)+(d), respectively.

distinctly with phase than their sum does. This concept is illustrated in figure 6.12 for $l = 0$ and $l = 1$ azimuthally polarised plane waves. Here the difference between the diffraction patterns expected for a circular and a spiral grating vary much more obviously.

These latter measurements are currently being pursued and will hopefully provide definitive results one way or another. If not there are still several alternative measurement methodologies that have been used in the literature for similar problems, such as diffraction of a knife edge[113] or a triangular aperture[114]. Both have the advantage that they also give information on the handedness of the vortex beam, which double slits only do indirectly if misaligned.



(a) Fanned grating design schematic.

(b) Measurement.

Figure 6.13 – Design and measurements for a fanned DFB grating resonator. (a) shows a schematic of the grating design employed, with a linearly changing period in the direction along the ridges. The period of the fabricated gratings changes by 20 nm over a 500 μm distance, while the duty cycle is kept constant. (b) shows the measured spectra of the emitted beam as a function of pump laser position on the grating.

6.3 Fanned grating DFB lasers.

While working with spatially varying gratings for DFB resonators, we realised that phase might not be the only interesting part of an emitted beam we could manipulate with such techniques. In particular, we realised that varying the grating period spatially, as long as it is done adiabatically, should result in different resonant emission wavelengths from different locations on the grating, while potentially retaining the spatial coherence.

To explore this, alongside the circular and spiral gratings discussed in the previous two sections, I also fabricated several fanned gratings (see figure 6.13a.) These gratings were made in the same fabrication run as the previous ones and therefore share many of the specifics, such as materials, duty cycles and etch depths. The spatial design however involves the continuously changing of period from 340 nm to 360 nm over a 500 μm distance. In this case we chose to change the period in the direction along the grating ridges. While this does potentially cause some polarisation impurity in the resulting emission which varying the period across the ridges would not have, as a first step it serves as a proof of concept.

To test these fanned gratings, the pump spot was narrowed down to several tens of micrometres in diameter and scanned along the centre of the grating from

the low period end to the high period, with the emitted beam captured by a spectrometer to measure spectra at a large number of positions. The results are shown in figure 6.13b. The position of the pump spot with respect to the grating limits is given on one axis, wavelength on another and intensity on the vertical. The emitted wavelength very obviously tracks the pump spot across the expected wavelength range.

These measurements are still very preliminary and further investigations need to be done. Features of interest that we want to explore include the spatial coherence and the allowed gradients in period. Our collaborators in St Andrews are also exploring the possibility of fabricating such gratings on a flexible substrate.

7 Conclusions.

7.1 Summary and conclusions.

This project began with the design of a flat focusing mirror based on grating physics. Over time, it grew into an exploration of the potential of dielectric gratings to form arbitrary beam shapes. From there, it was only a short step to study sub wavelength gratings in the context of metasurfaces.

Over the past three and a half years I have developed the theory and design aspects of four different types of metasurface. In this thesis, they have been labelled as direct phase, polarisation conversion, geometric phase, and active. Each of these types of metasurface fundamentally relies on the physics of high refractive index, sub wavelength scale gratings, and the resonant modes unique to such structures. At the same time, each type of metasurface uses a different expression of these resonances to achieve the goal of shaping light beams.

My work started with an investigation of so-called high contrast gratings. High contrast gratings are periodic dielectric structures with a sub-wavelength period that consist of a material with a higher refractive index than their surroundings. These high contrast gratings are primarily used as alternatives to distributed Bragg reflectors with a wavelength scale thickness, as they allow for near unity reflectivity over a broad bandwidth. Shortly before the start of my PhD, several research groups had shown their potential for phase manipulation[21, 22, 32], and building on their work I designed the first grating lenses described in this thesis. These high contrast grating lenses were the start of my exploration of what I would later label as direct phase metasurfaces, when it became necessary to distinguish them from the other types of metasurfaces I developed.

The study of direct phase metasurfaces required an in depth investigation into the propagation of beams and their manipulation through phase profiles. I identified several beams of interest, chosen both because of their ability to demonstrate the limits of metasurfaces and because of related interest in micro scale optics. These beams were focused Gaussians, focused and unfocused Laguerre-Gaussians, and Bessel beams. The phase profiles to generate each of these beam types are complementary, in that together they represent radial and azimuthal symmetries, as well as linear and quadratic phase gradients, allowing us to study a range of effects that could not quite be found by using any one of them in isolation. A final beam type that complements the others is that of Cylindrical Vector beams, where the beam has a spatially dependent polarisation component on top of the phase distribution.

As the physics of sub-wavelength gratings lies at the core of all four types of metasurfaces I discussed, a significant part of my work involved developing a suitable theoretical frame work for understanding the grating resonance effects that go into the design of metasurfaces. Based on an interplay between the nature of gratings as diffractive elements to couple light into them and the high refractive index that allows the grating to act as a waveguide, the concept of Guided Mode Resonances was developed. I then showed that the combination of such Guided Mode Resonances with the zero order Fabry-Perot, or thin film, resonances gives rise to a rich resonance environment on which a variety of useful effects can be based[28].

The designs of all metasurfaces required an accurate knowledge of the optical response of gratings with varying parameters, such as thickness, period, duty cycle, etc. The chosen method to calculate these optical responses was a numerical technique called Rigorous Coupled Wave Analysis (RCWA)[73]. Using RCWA I was able to accurately predict the reflectivities, transmittivities and associated phases of periodic structures with any desired parameters accurately and efficiently. I could then use the resulting data sets spanning ranges of grating parameters to find appropriate structure designs for the various metasurfaces.

In the case of direct phase metasurfaces, we tailor the phase of a passing light beam by spatially varying the period and duty cycle of a grating. The full structures required finding a 2π phase path through an RCWA data set spanned by the period and duty cycle. I developed a modified Dijkstra path finding algorithm to extract the most suitable paths for a given metasurface design, where the suitability depends on a compromise between fabrication tolerances, average reflectivity or transmittivity, and the modulation in the latter. The fabrication tolerance is the least fundamental of these considerations, but it also puts the most severe restrictions on the areas we can search. To streamline the design process, I developed several guidelines for the grating parameters that are most promising for this application, specifically giving restrictions on grating thickness, refractive index and period as a function of the wavelength.

With a basis set of grating parameters that give access to a full 2π phase range as a function of period and duty cycle, I was then able to design a variety of metasurfaces, from parabolic mirrors and lenses to ones that generate vortex beams or Bessel beams. These metasurfaces were extensively tested in simulations using COMSOL multiphysics. Among other things, a high spectral bandwidth of operation was shown, with 30% achieved for a typical reflective lens. The angular tolerance, on the other hand, turned out to be severely limited. Also using COMSOL, I showed that direct phase metasurfaces can perform equally

well in reflection and transmission.

The experimental work in this section was primarily executed by my colleagues Dr Annett Klemm and Andrei Ruskuc. They fabricated a large range of reflective lenses covering many focal lengths and diameters. In this way it became clear that while the focal length could be controlled fairly well and we achieved a high quality of focal spots, the numerical aperture of such gratings is restricted to about 0.6. I explained this with a fundamental limit of metasurfaces related to Nyquist's sampling theorem, in that reproducing a phase profile requires at least two points, or periods, per 2π phase change. This implies that for a phase profile of which the gradient gets too steep with respect to the period, the resulting behaviour is no longer as expected.

A final issue discussed for direct phase metasurfaces that impacts some, but not all, applications is the effect of the modulation of the reflectivity or transmittivity on the output beams. For many purposes this modulation only results in a slight irregularity of the beam and a small amount of loss, due to higher order diffraction effects. When the intensities in the output beam are close to that of the incident beam, however, the effects can become significant and lead to large aberrations in the output. For the beams we investigated, this was most noticeable in our simulations of Bessel beams.

Direct phase metasurfaces require a high refractive index to support a full 2π phase path. Materials with a refractive index high enough but no significant absorption losses are not available in the visible spectrum, which led to us using silicon and limiting our selves to the near infrared. However, for many applications, shaping beams only becomes truly interesting when it can be done in the visible. To accommodate these lower wavelengths, we began working with the concept of polarisation conversion and geometric phase metasurfaces.

These two types of metasurface are intimately related as they rely on the same basic concept, i.e. resonant form birefringence. Form birefringence is the result of differing effective refractive indices between the directions along or across a grating, leading to a change in polarisation state of passing light by changing the phase delay between TE and TM polarisations. Resonant form birefringence combines this idea with resonant modes to greatly increase the efficiency, giving 100% transmission for exactly $\frac{\pi}{2}$ and π phase differences, or perfect $\frac{\lambda}{4}$-plates and $\frac{\lambda}{2}$-plates.

In the polarisation conversion metasurfaces I use these effects to control the polarisation distribution on a beam, but otherwise leave it unchanged. This transforms a regular scalar beam into a vector beam counterpart. Specifically I designed, fabricated and tested ring shaped gratings out of Si_3N_4 $\frac{\lambda}{4}$-plate unit cells

that converted a passing beam into a radially or azimuthally polarised output. A proof of concept measurement on them showed an almost complete conversion from the incident circularly polarised light to the intended output polarisation states.

The geometric phase metasurfaces are based on gratings that act as $\frac{\lambda}{2}$-plates. By changing the orientation of these gratings in the plane of the metasurface it is possible to simultaneously flip the handedness of incident circularly polarised light as well as add an extra phase component proportional to the orientation angle of the grating. These metasurfaces can be made with much lower refractive indices than their direct phase counterparts, bringing phase profile manipulation into the visible wavelength range. Moreover, they can be constructed with only one unit cell design that has an almost unity transmission, removing the detrimental effects of amplitude modulation we saw with the Bessel beams in direct phase metasurfaces and promising very high efficiencies.

The downside of geometric phase metasurfaces when compared with their direct phase counterparts is the size of the unit cells. Where the phase changes on a per period basis in the direct phase structures, for geometric phase metasurfaces at least five periods appear to be necessary. This puts a significantly more severe restriction on the phase gradients that can be encoded into geometric phase metasurfaces, thereby limiting their potential for applications that require fast phase changes.

Geometric phase metasurfaces have been conceptually developed and designed, but the experimental verification is still work in progress.

The final type of metasurface I worked on was inspired by the desire to not just convert passing beams, but directly generate beams in the shape we desire. These active metasurfaces are based on organic Distributed Feedback (DFB) lasers, where I redesigned the resonant feedback grating to control the phase of the emitted lasers. The concept rests on the phase shift we expect when we introduce a spatial shift between two otherwise identical gratings.

To demonstrate this idea, I designed a variety of spiral shaped gratings in order to directly create azimuthally polarised vortex beams. The fabrication and experimentation of the full DFB structures was performed in close collaboration with James Glackin at the University of St Andrews. The initial results are presented in this thesis, but are as of yet inconclusive.

In parallel, we also explored the idea of spatial control of the spectral qualities of the emitted beams. This idea makes use of the close correlation between the period of the resonant feedback grating and the lasing wavelength. Initial experiments with a fanned grating where the period changes linearly over a relatively

large distance show a high connection between spatial position on the grating and emitted wavelength.

What all these metasurfaces have in common, and what sets them apart from their plasmonic counterparts, is their high efficiencies. Beam conversion efficiencies above 90% are not fantasy for any of the metasurface concepts I have explored, and with direct phase metasurfaces in particular we have experimentally shown working lenses with efficiencies exceeding 80%. This stands in sharp contrast with plasmonic metasurfaces which do not surpass the 30% range in transmission[10, 11, 26]. While the plasmonic resonators win out on the steepness of phase gradients they can encode, owing to their more compact size, the dielectric grating based metasurfaces I have presented here are clearly superior for a wide variety of applications.

7.2 Outlook.

While I have made significant progress in the development of the core concepts and design processes associated with the four types of metasurfaces discussed here, further steps can still be taken to improve them. Primarily this includes the experimental verification of several structures that have up to now only been studied in simulations, such as transmissive direct phase metasurfaces and geometric phase metasurfaces. On top of that an exploration of their applications to wavefronts not accessible by means of ordinary optics, e.g. holography, is an interesting further step.

However, even at this point it is clear that together, these different metasurface types provide a platform for a wide array of applications. For example, we are currently working on applying direct phase metasurfaces to optical trapping integrated into microfluidics. This application requires lenses operating at the limit of the numerical aperture possible with such metasurfaces, but potentially provides a platform for integrating arrays of traps with inter trap spacings that could traps cells while still allowing mutual interaction.

Another potential application we are exploring is the implementation of metasurfaces as projection lenses in 3D display technology. There are a multitude of challenges we face to make this application a reality, most notably the angular spread required, but the initial explorations are providing us with new ways of looking at our metasurfaces and seem promising. If successful, dielectric metasurfaces could reduce the thickness of displays based on this technology by a significant amount, which is paramount for a marketable product.

For active metasurfaces we are aiming to enable Stimulated Emission Deple-

tion (STED) microscopy integrated into a microfluidic platform. To accomplish STED effectively a donut shaped intensity pattern with a high contrast between the ring and the centre dark spot is necessary, so the current results are not quite ready for this, but we believe that improvements in the fabrication process should lead to better results.

Bibliography

[1] Stefan W Hell. Microscopy and its focal switch. *Nat. Methods*, 6(1):24–32, 2009. doi: 10.1038/nmeth.1291.

[2] Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(2006):1642–1645, 2006. ISSN 0036-8075. doi: 10.1126/science.1127344.

[3] A Ashkin, J M Dziedzic, J E Bjorkholm, and S Chu. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.*, 11(5):288, 1986. ISSN 0146-9592. doi: 10.1364/OL.11.000288.

[4] Justin E. Molloy and Miles J. Padgett. Lights , action : optical tweezers. *Contemp. Phys.*, 43(4):241–258, 2002.

[5] Mingzhou Chen, Michael Mazilu, Yoshihiko Arita, Ewan M Wright, and Kishan Dholakia. Dynamics of microparticles trapped in a perfect vortex beam. *Opt. Lett.*, 38(22):4919–22, November 2013. ISSN 1539-4794.

[6] Miles Padgett and Richard Bowman. Tweezers with a twist. *Nat. Photonics*, 5(6):343–348, June 2011. ISSN 1749-4885. doi: 10.1038/nphoton.2011.81.

[7] B.-C. Chen, Wesley R Legant, K. Wang, Lin Shao, Daniel E Milkie, Michael W Davidson, Chris Janetopoulos, Xufeng S Wu, J. A. Hammer, Zhe Liu, Brian P English, Y. Mimori-Kiyosue, Daniel P Romero, Alex T Ritter, J. Lippincott-Schwartz, L. Fritz-Laylin, R Dyche Mullins, Diana M Mitchell, A.-C. Reymann, Joshua N Bembenek, R. Bohme, Stephan W Grill, Jennifer T Wang, Geraldine Seydoux, U Serdar Tulu, Daniel P Kiehart, and Eric Betzig. Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science*, 2014. ISSN 0036-8075. doi: 10.1126/science.1257998.

[8] J. Nylk, C. Mitchell, T. Vettenburg, F. J. Gunn-Moore, and Kishan Dholakia. Wavefront shaping of a Bessel light field enhances light sheet microscopy with scattered light. In *Proc. SPIE 8949, Three-Dimensional Multidimens. Microsc. Image Acquis. Process. XXI, 89490V*, 2014.

[9] Helen A. Rendall, Robert F Marchington, Bavishna B Praveen, Gerald Bergmann, Yoshihiko Arita, Alexander Heisterkamp, Frank J Gunn-Moore, and Kishan Dholakia. High-throughput optical injection of mammalian cells

using a Bessel light beam. *Lab Chip*, 12(22):4816–20, 2012. ISSN 1473-0189. doi: 10.1039/c2lc40708f.

[10] Libo Yu, Elsie Barakat, Tristan Sfez, Lubos Hvozdara, Joab Di Francesco, and Hans Peter Herzig. Manipulating Bloch surface waves in 2D: a platform concept-based flat lens. *Light Sci. Appl.*, 3(1): e124, January 2014. ISSN 2047-7538. doi: 10.1038/lsa.2014.5. URL `http://www.nature.com/doifinder/10.1038/lsa.2014.5`.

[11] Patrice Genevet and Federico Capasso. Holographic optical metasurfaces: a review of current progress. *Rep. Prog. Phys.*, 78(2):024401, 2015. ISSN 1361-6633. doi: 10.1088/0034-4885/78/2/024401.

[12] N. M. Litchinitser. Structured Light Meets Structured Matter. *Science*, 337 (6098):1054–1055, 2012. ISSN 0036-8075. doi: 10.1126/science.1226204.

[13] Francesco Aieta and Mikhail A Kats. Multiwavelength achromatic metasurfaces. *Science*, 347(February):1342–1346, 2015. doi: 10.1126/science.aaa2494.

[14] F. Falcone, T. Lopetegi, M. A. G. Laso, J. D. Baena, J. Bonache, M. Beruete, R. Marques, F. Martin, and M. Sorolla. Babinet principle applied to the design of metasurfaces and metamaterials. *Phys. Rev. Lett.*, 93(19):2–5, 2004. ISSN 00319007. doi: 10.1103/PhysRevLett.93.197401.

[15] Stanislav B. Glybovski, Sergei A. Tretyakov, Pavel A. Belov, Yuri S. Kivshar, and Constantin R. Simovski. Metasurfaces: From microwaves to visible. *Phys. Rep.*, 2016. ISSN 03701573. doi: 10.1016/j.physrep.2016.04.004.

[16] Nanfang Yu, Patrice Genevet, Mikhail A. Kats, Francesco Aieta, Jean-Philippe Tetienne, Federico Capasso, and Zeno Gaburro. Light Propagation with Phase Discontinuities Reflection and Refraction. *Science*, 334 (October):333–337, 2011. ISSN 1095-9203. doi: 10.1126/science.1210713.

[17] Guoxing Zheng, Holger Mühlenbernd, Mitchell Kenney, Guixin Li, Thomas Zentgraf, and Shuang Zhang. Metasurface holograms reaching 80% efficiency. *Nat. Nanotechnol.*, 10(4):308–12, 2015. ISSN 1748-3395. doi: 10.1038/nnano.2015.2.

[18] Mikhail I. Shalaev, Jingbo Sun, Alexander Tsukernik, Apra Pandey, Kirill Nikolskiy, and Natalia M. Litchinitser. High-Efficiency All-Dielectric Metasurfaces for Ultracompact Beam Manipulation in Trans-

mission Mode. *Nano Lett.*, 15(9):6261–6266, 2015. ISSN 15306992. doi: 10.1021/acs.nanolett.5b02926.

[19] Manuel Decker, Isabelle Staude, Matthias Falkner, Jason Dominguez, Dragomir N. Neshev, Igal Brener, Thomas Pertsch, and Yuri S. Kivshar. High-Efficiency Dielectric Huygens' Surfaces. *Adv. Opt. Mater.*, 3(6):813–820, 2015. ISSN 21951071. doi: 10.1002/adom.201400584.

[20] D. Lin, P. Fan, E. Hasman, and M. L. Brongersma. Dielectric gradient metasurface optical elements. *Science*, 345(6194):298–302, 2014. ISSN 0036-8075. doi: 10.1126/science.1253213.

[21] David Fattal, Jingjing Li, Zhen Peng, Marco Fiorentino, and Raymond G Beausoleil. Flat dielectric grating reflectors with focusing abilities. *Nat. Photonics*, 4:466–470, 2010. doi: 10.1038/NPHOTON.2010.116.

[22] Fanglu Lu, Forrest G. Sedgwick, Vadim Karagodsky, Christopher Chase, and Connie J. Chang-Hasnain. Planar high-numerical-aperture low-loss focusing reflectors and lenses using subwavelength high contrast gratings. *Opt. Express*, 18(12):12606–12614, May 2010. ISSN 1094-4087. doi: 10.1364/OE.18.012606.

[23] Francesco Monticone and Andrea Alu. Leaky-wave theory, techniques, and applications: From microwaves to visible frequencies. *Proc. IEEE*, 103(5): 793–821, 2015. ISSN 00189219. doi: 10.1109/JPROC.2015.2399419.

[24] D. Berry, R. Malech, and W. Kennedy. The reflectarray antenna. *IEEE Trans. Antennas Propag.*, 11(6):645–651, 1963. ISSN 0096-1973. doi: 10.1109/TAP.1963.1138112.

[25] Winston E. Kock. Metal-Lens Antennas. *Proc. IRE*, 34(11):828–836, 1946. ISSN 00968390. doi: 10.1109/JRPROC.1946.232264.

[26] Carl Pfeiffer, Naresh K Emani, Amr M Shaltout, Alexandra Boltasseva, Vladimir M Shalaev, and Anthony Grbic. Efficient Light Bending with Isotropic Metamaterial Huygens Surfaces. *Nano Lett.*, 14:2491–2497, 2014.

[27] Amir Arbabi and Andrei Faraon. Fundamental Limits of Ultrathin Metasurfaces. *arXiv:1411.2537*, pages 1–6, 2014.

[28] Ye Zhou, Michael C Y Huang, Christopher Chase, Vadim Karagodsky, Michael Moewe, Bala Pesala, Forrest G Sedgwick, and Connie J Chang-hasnain. High-Index-Contrast Grating (HCG) and Its Applications in

Optoelectronic Devices. *IEEE J. Sel. Top. Quantum Electron.*, pages 1–15, 2009. doi: 10.1109/JSTQE.2009.2021145.

[29] Francesco Aieta, Patrice Genevet, Mikhail A. Kats, Nanfang Yu, Romain Blanchard, Zeno Gaburro, and Federico Capasso. Aberration-free ultrathin flat lenses and axicons at telecom wavelengths based on plasmonic metasurfaces. *Nano Lett.*, 12(9):4932–4936, 2012. ISSN 15306984. doi: 10.1021/nl302516v.

[30] Annett B Klemm, Daan Stellinga, Emiliano R Martins, Liam Lewis, Guillaume Huyet, Liam O'Faolain, and Thomas F Krauss. Experimental high numerical aperture focusing with high contrast gratings. *Opt. Lett.*, 38(17): 3410–3, September 2013. ISSN 1539-4794.

[31] Vadim Karagodsky and Connie J. Chang-Hasnain. Physics of near-wavelength high contrast gratings. *Opt. Express*, 20(10):1–7, 2012.

[32] Luca Carletti, Radu Malureanu, Jesper Mø rk, and Il-Sug Chung. High-index-contrast grating reflector with beam steering ability for the transmitted beam. *Opt. Express*, 19(23):23567–72, November 2011. ISSN 1094-4087.

[33] A. Hessel and A. A. Oliner. A New Theory of Woods Anomalies on Optical Gratings. *Appl. Opt.*, 4(10):1275, 1965. ISSN 0003-6935. doi: 10.1364/AO.4.001275.

[34] S. S. Wang, R. Magnusson, J. S. Bagby, and M. G. Moharam. Guided-mode resonances in planar dielectric-layer diffraction gratings. *J. Opt. Soc. Am. A*, 7(8):1470, 1990. ISSN 1084-7529. doi: 10.1364/JOSAA.7.001470.

[35] Vadim Karagodsky, Forrest G Sedgwick, and Connie J Chang-Hasnain. Theoretical analysis of subwavelength high contrast grating reflectors. *Opt. Express*, 18(16):16973–88, August 2010. ISSN 1094-4087.

[36] Robert Magnusson and Mehrdad Shokooh-Saremi. Physical basis for wideband resonant reflectors. *Opt. Express*, 16(5):3456–62, March 2008. ISSN 1094-4087.

[37] Armando Ricciardi, Stefania Campopiano, Andrea Cusano, Thomas F. Krauss, and Liam O'Faolain. Broadband mirrors in the near-infrared based on subwavelength gratings in SOI. *IEEE Photonics J.*, 2(5):696–702, 2010. ISSN 19430655. doi: 10.1109/JPHOT.2010.2059003.

[38] Yue Wang, Daan Stellinga, and Thomas F Krauss. Tunable Optical Filters Based on Silicon Nitride Membranes with Subwavelength Gratings. *Int. Conf. Opt. MEMS Nanophotonics*, 21(4):13–14, 2014.

[39] Yongjin Wang, Jiajia Chen, Zheng Shi, Shumin He, Martin Lopez Garcia, Lifeng Chen, Nikolai a. Hueting, Martin Cryan, Miao Zhang, and Hongbo Zhu. Suspended membrane GaN gratings for refractive index sensing. *Appl. Phys. Express*, 7(5), 2014. ISSN 18820786. doi: 10.7567/APEX.7.052201.

[40] G. J. Triggs, M. Fischer, D. Stellinga, M. G. Scullion, G. J O Evans, and T. F. Krauss. Spatial resolution and refractive index contrast of resonant photonic crystal surfaces for biosensing. *IEEE Photonics J.*, 7(3), 2015. ISSN 19430655. doi: 10.1109/JPHOT.2015.2435699.

[41] Annett B. Klemm, Daan Stellinga, Emiliano R. Martins, Liam Lewis, Liam OFaolain, and Thomas F. Krauss. Characterization of planar microlenses made of high contrast gratings. *SPIE Photonics West 2014-OPTO Optoelectron. Devices Mater.*, 8995(1):89950L, 2014. ISSN 1996756X. doi: 10.1117/12.2045726.

[42] M-W. Definition of BEAM., 2016. URL `http://www.merriam-webster.com/dictionary/beam`.

[43] OED. Definition of beam, 2016. URL `https://www.oxforddictionaries.com/definition/english/beam`.

[44] David J. Griffiths. *Introduction to electrodynamics*. Pearson, third edition, 1998. ISBN 978-0139199608.

[45] Joseph W. Goodman. *Introduction to fourier optics.* Viva Books, 2007. ISBN 978-8130908205.

[46] M. Born and E. Wolf. *Principles of optics.* Cambridge University Press, seventh edition, 1999.

[47] J Clerk Maxwell. A Dynamical Theory of the Electromagnetic Field. *Phil. Trans. R. Soc. L.*, 155(January):459–512, 1865. ISSN 0261-0523. doi: 10.1098/rstl.1865.0008.

[48] John Henry Poynting. On the transfer of energy in the electromagnetic field. *Philos. Trans. R. Soc. London*, (January):343–361, 1884. ISSN 0261-0523.

[49] Christiaan Huygens. *Traite de la lumiere.* 1690.

[50] A. Fresnel. -. *AnnChim Phys*, 2(1):239, 1816.

[51] G. Kirchhoff. Zur theorie der lichtstrahlen. *Ann. Phys.*, 254(4):663–695, 1883. doi: 10.1002/andp.18832540409.

[52] K.F. Riley, M.P. Hobson, and S.J. Bence. *Mathematical methods for physics and engineering*. Cambridge University Press, 2010. ISBN 978-0-521-86153-3.

[53] L Allen, Marco W. Beijersbergen, R J C Spreeuw, and J P Woerdman. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys. Rev. A*, 45(11):8185–8189, 1992.

[54] S. Franke-Arnold, L. Allen, and M. Padgett. Advances in optical angular momentum. *Laser Photonics Rev.*, 2(4):299–313, August 2008. ISSN 18638880. doi: 10.1002/lpor.200810007.

[55] Nobelprize.org. The Nobel Prize in Chemistry 2014., 2016. URL `https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2014/press.html`.

[56] Stefan W Hell and Jan Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluoresence microscopy. *Opt. Lett.*, 19(11):780–782, 1994. doi: 10.1364/OL.19.000780.

[57] D McGloin and K Dholakia. Bessel beams: Diffraction in a new light. *Contemp. Phys.*, 46(1):15–28, 2005. ISSN 0010-7514. doi: 10.1080/0010751042000275259.

[58] J. Durnin, J. Miceli, and J. H. Eberly. Diffraction-free beams. *Phys. Rev. Lett.*, 58(15):1499–1501, 1987. ISSN 00319007. doi: 10.1103/PhysRevLett.58.1499.

[59] Oto Brzobohatý, Tomás Cizmár, and Pavel Zemánek. High quality quasi-Bessel beam generated by round-tip axicon. *Opt. Express*, 16(17):12688–12700, 2008. ISSN 1094-4087. doi: 10.1364/OE.16.012688.

[60] Michael R. LaPointe. Review of Nondiffracting Bessel Beams. *SPIE Vol. 1527 Curr. Dev. Opt. Des. Opt. Eng.*, pages 258–276, 1991.

[61] Qiwen Zhan. Cylindrical vector beams: from mathematical concepts to applications. *Adv. Opt. Photonics*, 1(1):1, January 2009. ISSN 1943-8206. doi: 10.1364/AOP.1.000001.

[62] Ze'ev Bomzon, Gabriel Biener, Vladimir Kleiner, and Erez Hasman. Radially and azimuthally polarized beams generated by space-variant dielectric subwavelength gratings. *Opt. Lett.*, 27(5):285–287, 2002. ISSN 0146-9592. doi: 10.1364/OL.27.000285.

[63] H Ottevaere, R Cox, H P Herzig, T Miyashita, K Naessens, M Taghizadeh, R Völkel, H J Woo, and H Thienpont. Comparing glass and plastic refractive microlenses fabricated with different technologies. *J. Opt. A Pure Appl. Opt.*, 8(7):S407–S429, 2006. ISSN 1464-4258. doi: 10.1088/1464-4258/8/7/S18.

[64] J. B. Pendry. Negative refraction makes a perfect lens. *Phys. Rev. Lett.*, 85 (18):3966–3969, 2000. ISSN 00319007. doi: 10.1103/PhysRevLett.85.3966.

[65] H. Nyquist. Certain Topics in Telegraph Transmission Theory. *Trans. Am. Inst. Electr. Eng.*, 47(2):617–644, 1928. ISSN 00963860. doi: 10.1109/T-AIEE.1928.5055024.

[66] C E Shannon. Communication in the Presence of Noise. *Proc. IEEE*, 86 (2):447–457, 1998.

[67] Charles Kittel. *Introduction to Solid State Physics.* Wiley, 1996. ISBN 0-471-14286-7.

[68] R W Wood. On a Remarkable Case of Uneven Distribution of Light in a Diffraction Grating Spectrum. *Proc. Phys. Soc. London*, 18(1):269–275, 1902. ISSN 1478-7814. doi: 10.1088/1478-7814/18/1/325.

[69] S S Wang and R Magnusson. Theory and applications of guided-mode resonance filters. *Appl. Opt.*, 32(14):2606–13, May 1993. ISSN 0003-6935.

[70] C. Sauvan, J. P. Hugonin, R. Carminati, and P. Lalanne. Modal representation of spatial coherence in dissipative and resonant photonic systems. *Phys. Rev. A*, 89(4):043825, April 2014. ISSN 1050-2947. doi: 10.1103/PhysRevA.89.043825.

[71] Philippe Lalanne and Dominique Lemercier-Lalanne. On the effective medium theory of subwavelength periodic structures. *J. Mod. Opt.*, 43 (10):2063–2086, 1996. ISSN 0950-0340. doi: 10.1080/095003496154716.

[72] Philippe Lalanne and Dominique Lemercier-Lalanne. Depth dependence of the effective properties of subwavelength gratings. *J. Opt. Soc. Am. A*, 14 (2):450, 1997. ISSN 1084-7529. doi: 10.1364/JOSAA.14.000450.

[73] M G Moharam and T K Gaylord. Coupled-wave analysis of reflection gratings. *Appl. Opt.*, 20(2):240–4, January 1981. ISSN 0003-6935.

[74] Armando Ricciardi, Stefania Campopiano, Andrea Cusano, Thomas F Krauss, and Liam O'Faolain. Broadband Mirrors in the Near-Infrared Based on Subwavelength Gratings in SOI. *IEEE Photonics J.*, 2(5):696–702, October 2010. ISSN 1943-0655. doi: 10.1109/JPHOT.2010.2059003.

[75] T.K. Gaylord and M.G. Moharam. Analysis and applications of optical diffraction by gratings. *Proc. IEEE*, 73(5):894–924, 1985. ISSN 0018-9219. doi: 10.1109/PROC.1985.13220.

[76] Kevin Randolph Harper. *Theory, design, and fabrication of diffractive grating coupler for slab waveguide*. Msc, Brigham Young University, 2003.

[77] COMSOL. COMSOL Multiphysics, 2016. URL `https://uk.comsol.com`.

[78] COMSOL. *Introduction to COMSOL Multiphysics*. v4.3a edition, 2013.

[79] ZeonChemicals. ZEP 520A, 2016. URL `http://www.zeonchemicals.com/ElectronicMaterials/`.

[80] Allresist. Allresist, 2016. URL `http://www.allresist.com`.

[81] Robert C Frye, Kevin D Cummings, and Edward A Rietman. Proximity Effect Corrections in Electron Beam Lithography Using a Neural Network. *NIPS Proc. Beta*.

[82] Annett Birgit Fischer. *Planar optics with wavelength-scale high contrast gratings*. Phd thesis, University of York, 2015.

[83] Jean-Jacques Greffet, Rémi Carminati, Karl Joulain, Jean-Philippe Mulet, Stéphane Mainguy, and Yong Chen. Coherent emission of light by thermal sources. *Nature*, 416(6876):61–4, March 2002. ISSN 0028-0836. doi: 10.1038/416061a.

[84] Mark L Brongersma, Yi Cui, and Shanhui Fan. Light management for photovoltaics using high-index nanostructures. *Nat. Mater.*, 13(5):451–60, April 2014. ISSN 1476-1122. doi: 10.1038/nmat3921.

[85] Filmetrics. Filmetrics refractive index database, 2016. URL `http://www.filmetrics.com/refractive-index-database`.

[86] P.E. Hart, N.J. Nilsson, and B. Raphael. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Trans. Syst. Sci. Cybern.*, 4 (2):100–107, 1968. ISSN 0536-1567. doi: 10.1109/TSSC.1968.300136.

[87] T. van Vuren and G. R. M. Jansen. Recent developments in path finding algorithms: a review. *Transp. Plan. Technol.*, 12(1):57–71, 1988. ISSN 0308-1060. doi: 10.1080/03081068808717360.

[88] E W Dijkstra. A Note on Two Problems in Connexion with Graphs. *Numer. Math.*, 1:269–271, 1959.

[89] Kishan Dholakia, Gabriel Spalding, and Michael MacDonald. Optical tweezers: The next generation. *Phys. World*, 15(10):31–35, 2002. ISSN 09538585. doi: 10.1080/09500349808231712.

[90] Fabrice Merenda, Johann Rohner, Jean-Marc Fournier, and René-Paul Salathé. Miniaturized high-NA focusing-mirror multiple optical tweezers. *Opt. Express*, 15(10):6075–6086, 2007. ISSN 1094-4087. doi: 10.1364/OE.15.006075.

[91] Richard Bowman, Graham Gibson, and Miles Padgett. Particle tracking stereomicroscopy in optical tweezers: control of trap shape. *Opt. Express*, 18(11):11785–11790, 2010. ISSN 1094-4087. doi: 10.1364/OE.18.011785.

[92] S Stallinga. Optical trapping at low numerical aperture. *J. Eur. Opt. Soc. - Rapid Publ.*, 6:1–8, 2011. doi: 10.2971/jeos.2011.11027.

[93] De-Er Yi, Ying-Bai Yan, Hai-Tao Liu, and Guo-Fan Jin. Broadband achromatic phase retarder by subwavelength grating. *Opt. Commun.*, 227(1-3): 49–55, November 2003. ISSN 00304018. doi: 10.1016/j.optcom.2003.09.026.

[94] Z Ghadyani, S Dmitriev, N Lindlein, G Leuchs, O Rusina, and I Harder. Discontinuous space variant sub-wavelength structures for generating radially polarized light in visible region. *J. Eur. Opt. Soc. - Rapid Publ.*, 6 (11041):1–7, 2011.

[95] Yachao Liu, Xiaohui Ling, Xunong Yi, Xinxing Zhou, Hailu Luo, and Shuangchun Wen. Realization of polarization evolution on higher-order Poincare sphere with metasurface. *Appl. Phys. Lett.*, 104(19), 2014. ISSN 00036951. doi: 10.1063/1.4878409.

[96] J. Kischkat, S. Peters, B. Gruska, M. Semtsiv, M. Chashnikova, M. Klinkmüller, O. Fedosenko, S. Machulik, A. Aleksandrova,

G. Monastyrskyi, Y. Flores, and W. T. Masselink. Mid-infrared optical properties of thin films of aluminum oxide, titanium dioxide, silicon dioxide, aluminum nitride, and silicon nitride. *Appl. Opt.*, 51(28):6789–6798, 2012. ISSN 0003-6935. doi: 10.1364/AO.51.006789.

[97] D. Lin, P. Fan, E. Hasman, and M. L. Brongersma. Dielectric gradient metasurface optical elements. *Science*, 345(6194):298–302, 2014. ISSN 0036-8075. doi: 10.1126/science.1253213.

[98] M. V. Berry. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. London A*, pages 45–57, 1984. ISSN 1364-5021. doi: 10.1098/rspa.1983.0054.

[99] Michael J Escuti, Jihwan Kim, and Michael W Kudenov. Geometric-Phase Holograms. *Opt. Photonics news*, (February):22–29, 2016.

[100] K.Y. Bliokh, F.J. Rodríguez-Fortuño, F. Nori, and A.V. Zayats. Spin-orbit interactions of light. *ArXiv*, 9(December):1505.02864, 2015. ISSN 1749-4885. doi: 10.1038/nphoton.2015.201.

[101] Norcada. Norcada, 2016. URL `www.norcada.com`.

[102] Silson. Silson, 2016. URL `www.silson.com`.

[103] H G Berry, G Gabrielse, and a E Livingston. Measurement of the Stokes parameters of light. *Appl. Opt.*, 16(12):3200–3205, 1977. ISSN 0003-6935. doi: 10.1364/AO.16.003200.

[104] Yue Wang. *Low Threshold Organic Semiconductor Lasers*. Phd thesis, University of St Andrews, 2014.

[105] Yue Wang, Georgios Tsiminis, and Al Kanibolotsky. Nanoimprinted polymer lasers with threshold below 100 W/cm 2 using mixed-order distributed feedback resonators. *Opt. Express*, 21(12):9211–9216, 2013. ISSN 1094-4087. doi: 10.1364/OE.21.014362.

[106] Shuyu Zhang, Graham a. Turnbull, and Ifor D W Samuel. Highly directional emission and beam steering from organic light-emitting diodes with a substrate diffractive optical element. *Adv. Opt. Mater.*, 2(4):343–347, 2014. ISSN 21951071. doi: 10.1002/adom.201300441.

[107] Emiliano R. Martins, Yue Wang, Alexander L. Kanibolotsky, Peter J. Skabara, Graham a. Turnbull, and Ifor D W Samuel. Low-Threshold Nanoimprinted Lasers Using Substructured Gratings for Control of Distributed

Feedback. *Adv. Opt. Mater.*, 1(8):563–566, 2013. ISSN 21951071. doi: 10.1002/adom.201300211.

[108] G. A. Turnbull, A. Carleton, G. F. Barlow, A. Tahraouhi, T. F. Krauss, K. A. Shore, and I. D W Samuel. Influence of grating characteristics on the operation of circular-grating distributed-feedback polymer lasers. *J. Appl. Phys.*, 98(2):1–7, 2005. ISSN 00218979. doi: 10.1063/1.1935131.

[109] G. A. Turnbull, A. Carleton, A. Tahraouhi, T. F. Krauss, I. D W Samuel, G. F. Barlow, and K. A. Shore. Effect of gain localization in circular-grating distributed feedback lasers. *Appl. Phys. Lett.*, 87(20):1–2, 2005. ISSN 00036951. doi: 10.1063/1.2120915.

[110] G. A. Turnbull, P. Andrew, W. L. Barnes, and I. D W Samuel. Operating characteristics of a semiconducting polymer laser pumped by a microchip laser. *Appl. Phys. Lett.*, 82(3):313–315, 2003. ISSN 00036951. doi: 10.1063/1.1536249.

[111] H. I. Sztul and R. R. Alfano. Double-slit interference with Laguerre-Gaussian beams. *Opt. Lett.*, 31(7):999–1001, 2006.

[112] Yongnan Li, Xi-Lin Wang, Hu Zhao, Ling-Jun Kong, Kai Lou, Bing Gu, Chenghou Tu, and Hui-Tian Wang. Youngs two-slit interference of vector light fields. *Opt. Lett.*, 37(11):1790, 2012. ISSN 0146-9592. doi: 10.1364/OL.37.001790.

[113] Jochen Arlt. Handedness and azimuthal energy flow of optical vortex beams. *J. Mod. Opt.*, 50(10):1573–1580, 2003. ISSN 0950-0340. doi: 10.1080/0950034031oooO64942.

[114] A Mourka, M Mazilu, C Shanor, E M Wright, and K Dholakia. Characterizing the azimuthal and radial mode indices of a Laguerre-Gaussian beam using diffraction from a triangular aperture. *Front. Opt. Laser Sci. Tech. Dig.*, pages 18–19, 2011. ISSN 21622701.