Development of Metasurface Implementations for high-performance Vortex Phase Masks

PhD thesis submitted by
Lorenzo König

Advisor: Olivier Absil
Co-advisor: Michaël Lobet

Université de Liège
Faculté des Sciences
Département d'Astrophysique, Géophysique et Océanographie

12 September 2023

Members of the jury:

Prof. Serge Habraken (President), University of Liège
Dr. Olivier Absil (Advisor), University of Liège
Dr. Michaël Lobet (Co-advisor), University of Namur
Dr. Christian Delacroix, University of Liège
Prof. Mikael Karlsson, Uppsala University
Prof. Jérôme Loicq, Technical University of Delft
Prof. Dimitri Mawet, California Institute of Technology

This project was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 819155).

Acknowledgments

This work would not have been possible without the support of many people, and I'm thankful for all interactions which directly or indirectly supported me during the past four years. In particular, I would like to acknowledge:

Olivier, for being a great advisor. Your guidance in the past four years has been extraordinary, from countless face-to-face discussions shaping this thesis to detailed feedback about my publications. But not only: meeting you for football after work was a real pleasure, where you showed that a respectful relationship with your students is not limited to office hours.

Michaël, for the great support when I took my first steps in the field of metasurfaces, and for agreeing to be my co-advisor. The enthusiasm and engagement transmitted to your students make a real difference, promoting this exciting field of research.

The jury, who followed the progress of my work in the past years, and kindly agreed to review this manuscript. Thank you for diving into this exciting subject with me, and providing valuable input to my research.

The PSILab team, for introducing me to the field of high contrast imaging and providing a friendly and supportive work environment. In particular, thank you Christian and Gilles for the frequent discussions, and for mentoring me in the lab.

All my collaborators I had the pleasure to work with in the past years. Thank you Mikael and Pontus for a valuable time in Uppsala, giving me the opportunity to work in a clean room environment and getting my hands on in the microfabrication lab (I was looking forward to visiting you since the first day). Thank you Michaël and Nicolas for working on a complex project with us, and choosing us to be the first to benefit from your new optimizer. And thank you Skyler and Niyati, for giving birth to a great collaboration, and especially for the countless genuine discussions (it's true that you learn the best from your peers).

My friends in Liège, making this place my home for the past years, and to which I hope to come back frequently. Impossible to name everybody, but Jahanvi, Maxime, Lionel, Charly and Rakesh, for sure you had a special impact on who I became in the past four years.

My family, for creating the environment that allowed me to follow my passion for astronomy. For being excited of my choices, and for being proud of me as much as I am of them.

Niyati, for being by my side in this past year, for inspiring me, for making me rediscover my passion for science and for establishing a flourishing collaboration. Thank you for being my best friend and so much more.

Abstract in English

Direct imaging is one of the key techniques to study exoplanets and is an especially exciting field of research in light of the next generation of ground- and space-based telescopes. The direct imaging of exoplanets enables their detailed study and spectral characterization, which is crucial to understand their atmospheres and potential habitability, and also to better constrain their physical properties and orbital parameters. Especially for the study of terrestrial planets it is important to advance high-contrast imaging technologies. The vortex coronagraph is a promising approach to push high-contrast imaging instruments towards the required performance levels for directly imaging Earth-like planets. This thesis addresses the technology development of subwavelength grating and metasurface implementations for vortex phase masks. In particular, this work develops the necessary simulation tools to design new coronagraphic masks building on the rigorous coupled-wave analysis method to optimize periodic patterns and using the finite-difference time-domain method to describe aperiodic features. This enabled the optimization of subwavelength gratings including the annular groove phase masks for the METIS instrument and for the annular groove phase mask central region. In addition, it allowed the study of subwavelength grating vortex phase masks of topological charge 4 using segmented designs. Furthermore, the simulation framework was used to optimize metasurface patterns for the annular groove phase mask central region using a machine learning approach. The development of these tools allowed for the investigation of vector and scalar vortex phase mask designs based on metasurface patterns, enabling new microstructured phase mask designs of arbitrary phase patterns. In addition, this thesis also focused on the manufacturing and testing of the first annular groove phase masks for METIS, and a charge-4 subwavelength grating vortex phase mask. Overall, this work pushed the development of new vortex phase masks which will ultimately be promising candidates for installation in future high-contrast imaging instruments. A detailed outline of the manuscript structure is presented in Chapter 3.

Résumé en Français

L'imagerie directe est l'une des techniques clés pour étudier les exoplanètes et un domaine de recherche spécialement prometteur à l'approche des nouveaux télescopes terrestres et spatiaux. L'imagerie directe d'exoplanètes permet leur étude détaillée et leur caractérisation spectrale, ce qui est crucial afin de comprendre leurs atmosphères et leur habitabilité, et aussi pour mieux déterminer leurs propriétés physiques, ainsi que leurs orbites. En particulier, il est important de faire avancer les technologies d'imagerie haut contraste pour étudier les planètes terrestres. Le coronographe vortex est une approche prometteuse pour pousser les instruments dédiés à l'imagerie haut contraste vers les performances nécessaires afin de détecter directement des planètes terrestres. Cette thèse adresse le développement de réseaux sub-lambda et de métasurfaces pour des masques vortex. En particulier, le travail présenté développe les outils de simulation nécessaires pour la conception de nouveaux masques coronographiques exploitant la méthode RCWA pour des motifs périodiques, et en utilisant la méthode FDTD pour des motifs apériodiques. Ceci permet d'optimiser des réseaux sub-lambda, ce qui inclut les AGPM pour METIS et la partie centrale de l'AGPM en général. De plus, ceci permet d'étudier des masques vortex de charge topologique 4 exploitant des motifs segmentés. Les outils développés ont aussi été utilisés pour optimiser des métasurfaces dans la partie centrale de l'AGPM exploitant une approche d'apprentissage automatique. En plus, ces outils ont permis d'explorer la conception de vortex vectoriels et scalaires basés sur des métasurfaces, pouvant conduire à la conception de nouveaux masques de phase avec des profils de phase arbitraires. Cette thèse a aussi exploré la fabrication des premiers masques AGPM pour METIS, et d'un nouveau masque avec charge topologique 4. En résumé, cette thèse a poussé le développement de nouveaux masques de phase vortex qui seront des candidats prometteurs pour être installés sur de futurs instruments haut contraste.

Contents

List of Figures

List of Tables

Part I

Introduction

1 High-contrast imaging

Since the most ancient cultures humans have looked up to the sky, fascinated by the infinite space that could harbor distant worlds. While even the existence of planets outside our solar system was the subject of vivid debate in the past, exoplanets have now become an active area of research gathering the scientific community around the goal of finding planets compatible with life as known on Earth.

Since the first exoplanet discovery three decades ago, the exoplanet community has grown exponentially leading to huge advances in their detection and characterization. While the technologies so far have allowed for the detection of a large number of exoplanets via indirect detection methods, only a handful of planets have been observed directly thanks to high-contrast imaging techniques. This chapter briefly introduces the direct imaging of exoplanets as one of the main science motivations for the development of high-contrast imaging instrumentation, which is an important tool paving the road towards the direct detection of an Earth-like planet around a Sun-like star.

1.1 The hunt for exoplanets

Since the first theories of cosmic pluralism placing the Earth as one planet among many others, scientists have debated about the existence of extrasolar planets and many astronomers of the 20th century expected planetary systems around most main sequence stars. The first exoplanets however were discovered in a rather harsh environment. Observations of the millisecond radio pulsar PSR1257+12 revealed the presence of two planets of a few Earth masses (Wolszczan & Frail, 1992). Pulsars are old, rapidly rotating neutron stars exhibiting extremely periodic radio pulses due to the emission of radiation along the magnetic poles of the star, generally not aligned with their rotation axis. The rotation period of the star determines the period of the radio pulses, which is on the order of seconds to milliseconds. A companion to the pulsar can be detected by measuring the radial motion of the host star with respect to the observer, caused by the gravitational interaction of the planets and the star, both orbiting the common center of mass. Unless the orbital plane is orthogonal to the line of sight, the orbital motion of the star has a radial component leading to a Doppler shift of the pulsar frequency. The planetary system around PSR1257+12 was detected by measuring the periodic varying pulsation frequency.

It was only a few years after this discovery, that the first exoplanet around a main-sequence star was found using the radial velocity method (Mayor & Queloz, 1995).

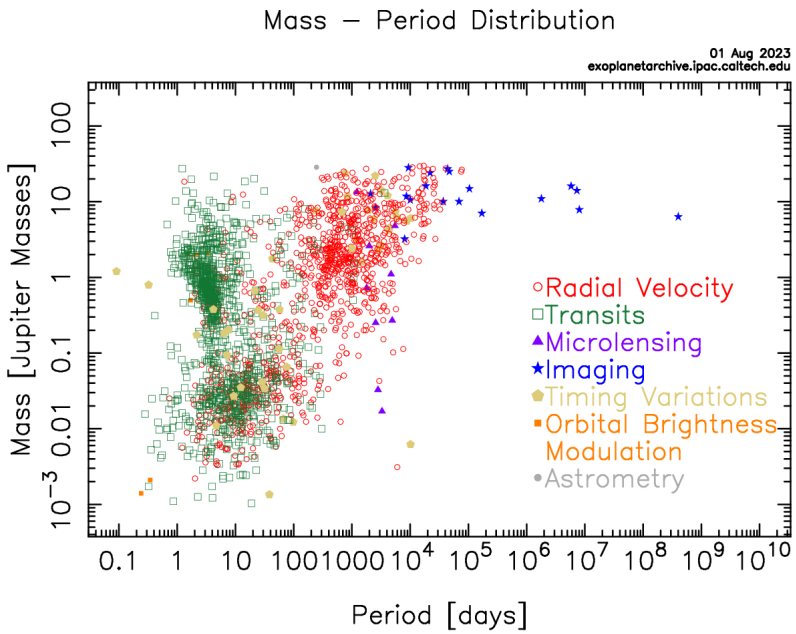

Figure 1.1: Characteristics of detected exoplanets. The parameter space spanned by the mass of the exoplanets and their orbital period is shown. Different detection techniques are shown as different colors, illustrating the discovery space of each technique. Retrieved from NASA exoplanet archive on the 1st of August 2023.

This method uses the same physical effect of Doppler shifted signals, but applied to the stellar spectrum. In this way, a periodic shift of the spectrum can be used to infer the mass of a companion. In practice, the variation of radial velocity of the star is larger for companions of higher mass and shorter periods, and is therefore suited to detect massive planets close to their host star. The detection of many of these massive planets close to their host star (hot Jupiters) has considerably impacted the theories of planet formation and planetary system architectures. However, the observations are biased towards massive planets in short orbits, covering only a small region of the parameter space of orbital period and planetary mass, as shown in Figure 1.1.

A few years later another method was used to confirm the planetary nature of a few previously detected planets, and infer the presence of more exoplanets around large samples of stars. The transit method monitors the brightness of the host star and searches for periodic drops in its light curve. When the orbital plane of a planet is aligned with the observer, the planet will block part of the starlight during its orbit, resulting in periodic drops of the star's brightness. The first discovery of a transiting planet (Charbonneau et al., 1999; Henry et al., 1999) opened a new region of the parameter space shown in Figure 1.1. As of today, the transit method has become the most successful exoplanet detection method in terms of discoveries, thanks to photometric surveys of large portions of the sky.

A series of other indirect exoplanet detection methods besides the two above exist, and have also been used to detect exoplanets. The presence of additional planets can be detected in a system of transiting planets, by observing the exact time of transit. Variations in the transit timing allow conclusions about further planets influencing the orbit of the transiting planets and host star. Another method using the photometric variability of the star-planet system is to detect microlensing events. Microlensing events happen when a star-planet system crosses the line of sight to a distant faint source. The foreground star can then act as a gravitational lens magnifying the intensity of the background source by several orders of magnitude, creating a characteristic light curve. The light curve in the presence of a planet around the foreground star will have a characteristic asymmetric shape with potentially sharp features, revealing the presence of the planet. Finally, astrometric measurements of the position of a star can reveal a companion by the periodic proper motion of the star caused by the planet orbiting its star. This method has the same cause as the radial velocity method, but now measures displacement normal to the line of sight. A detailed description of these exoplanet detection methods can be found in Fischer et al. (2014).

1.2 Direct imaging of exoplanets

Thousands of exoplanets have been discovered to date with the previously discussed methods. The characterization of these planets is however challenging due to the nature of indirect detection techniques. For example, to measure the spectrum of a transiting planet, the stellar flux has to be calibrated extremely precisely, but more critically the planet orbit has to be precisely aligned with the observer, favouring planets at closer separations from their host star. With direct imaging instead, the planet light is detected directly, and spectral characterization is mainly limited by the planet flux and speckles. In addition, direct imaging gives access to the planet orbit and mass, without fundamentally excluding any orientation of the orbital plane. It was therefore necessary to develop the techniques to detect an exoplanet directly, resulting in the first direct detection of a giant planet around a brown dwarf (Chauvin et al., 2004).

Direct imaging of exoplanets faces two main challenges. First, the apparent distance of a planet to its host star is very small. For the closest star α Cen, a planet on an Earth-like orbit would appear at an angular separation of approximately 1 arcsecond, which is the limit imposed by atmospheric seeing for ground-based applications. However, an efficient adaptive optics system can correct for the atmospheric turbulence, leading to diffraction limited imaging on most major ground-based telescopes. Still, the apparent separation of planet and star is small and decreases for farther away targets, which requires large apertures, or interferometric techniques to obtain the necessary angular resolution.

The second challenge is the extreme contrast between a planet and its host star. For example, the contrast between the Sun and the Earth is on the order of 10^{10} in visible

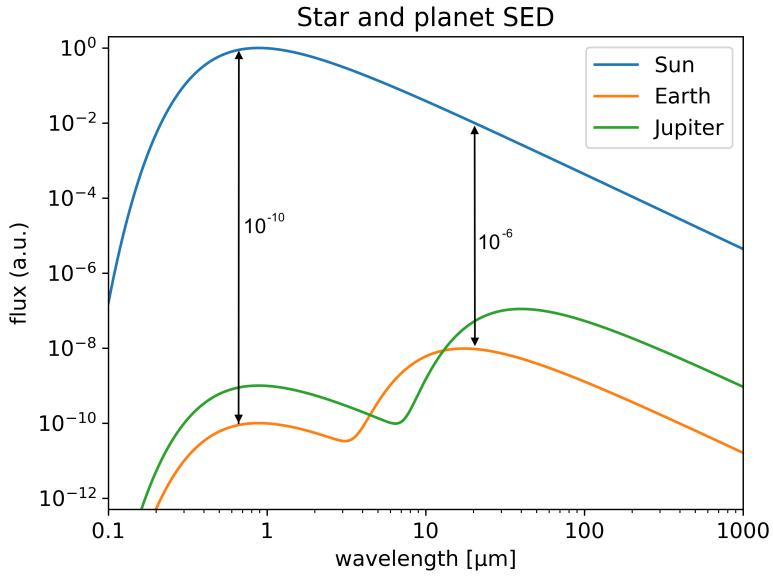

Figure 1.2: Planet and stellar spectral energy distribution (SED) modeled for Jupiter- and Earth-like planets (10^{-9} and 10^{-10} contrast in reflected light, respectively). The contrast is relaxed for longer wavelengths (from 10^{-10} in the visible to 10^{-6} in the mid-infrared for the Earth-Sun system) due to the thermal emission of the planets.

light. One way to reduce the contrast is to observe in different wavelength ranges. While a star will emit most of its light in the visible or near-infrared wavelength range, a cooler planet will emit a blackbody radiation peaking in the mid- to far-infrared, where the star is less bright, as shown in Figure 1.2. However, to detect and study planets in the visible and infrared wavelength range, it is mandatory to reduce the contrast of the system, which can be achieved with a coronagraph. A coronagraph centered on the star can attenuate its light by several orders of magnitude, while transmitting the light of a faint off-axis companion nearly unaffected. Different coronagraphic strategies to suppress the starlight are discussed in Chapter 2.

Along with a powerful coronagraph, a clever observing strategy is necessary to remove the starlight residuals (speckles) in the coronagraphic image caused by wavefront residuals from the adaptive optics system and instrumental errors. Quasi-static speckles can be particularly problematic because they are stable over the time of an observation run and are therefore hard to distinguish from a true companion.

Several observing strategies and image processing techniques can be used to mitigate the speckle pattern. PSF subtraction is a technique using a reference star to measure the quasi-static speckle pattern, which is then subtracted from the coronagraphic image. Spectral differential imaging (Marois et al., 2000) uses a sequence of observations of the same target in several bands, at which the contrast is expected to change consid-

erably. Spectroscopic differential imaging (Sparks & Ford, 2002) extends this method by using integral field spectroscopy. The reference image is obtained by scaling the speckle pattern with the wavelength. Polarization differential imaging (Kuhn et al., 2001) is a technique exploiting the partly polarized nature of the light reflected off an exoplanet compared to the unpolarized light of the star. Finally, angular differential imaging (Marois et al., 2006) is a technique exploiting the rotation of the field of view during an observing run to distinguish between a companion moving in the field of view, while speckles remain fixed with respect to the field of view.

The first hints of exoplanets were observed through the observation of circumstellar disks for which the requirements on the coronagraph are relaxed because their extended size results in more reflected stellar light. The existence of planets was therefore inferred through their signature in circumstellar disks, such as gaps, spirals and asymmetries. An asymmetry in the shape of the disk of βPic led to the prediction of the planet βPic b, which was confirmed a decade later through direct imaging (Lagrange et al., 2009). Another interesting discovery was the direct imaging of forming protoplanets, such as PDS70b (Keppler et al., 2018), which typically emit in the infrared, where the contrast requirements are relaxed. These observations animated the discussion about planet formation mechanisms.

1.3 Current and future instrumentation for high-contrast imaging

Many telescopes have been equipped with instruments optimized for high-contrast imaging in the past years. Here, current instruments optimized for high-contrast imaging are briefly presented, paving the way for the development of the next generation of ground- and space-based instruments for high-contrast imaging.

1.3.1 Current ground-based instruments

The first adaptive optics systems were developed in the 1990s, resulting in the first precursor instruments COME-ON (Rigaut et al., 1991) and ADONIS (Beuzit et al., 1997). Adaptive optics systems have since been installed on major ground-based telescopes (VLT, Rousset et al., 2003; Keck, Wizinowich et al., 2006; Gemini, Hodapp et al., 2003). These instruments were combined with advanced data processing and observing strategies leading to the detection of the first directly imaged exoplanet in 2004.

Since then, a second generation of adaptive optics systems has replaced the first instruments, providing extreme adaptive optics corrections optimized for exoplanet imaging. In particular, the vortex coronagraph was installed on VLT/SPHERE (Beuzit et al., 2019) and Gemini/GPI (Macintosh et al., 2014), which both also feature an apodized Lyot coronagraph. In addition, the vortex coronagraph has been tested at

Subaru/SCExAO (Lozi et al., 2018), which is an instrument used to test new technologies (coronagraphs, wavefront sensing and control, and detectors). Combined with advanced coronagraphs, this generation of instruments has led to the direct detection of a series of exoplanets to date.

The next generation of ground-based instruments targeted at high-contrast imaging is on their way on the next generation of extremely large telescopes.

1.3.2 Future ground-based instruments on the ELT

The Extremely Large Telescope (ELT, Gilmozzi & Spyromilio, 2007) features three first-light instruments. The Multi-AO Imaging Camera for Deep Observations (MICADO, Davies et al., 2021) is a near-infrared high-resolution imaging instrument enabling unprecedented image resolution thanks to the ELT's large primary mirror. It will enable many science applications such as the study of the galactic center or the study of individual stars in other galaxies. The High Angular Resolution Monolithic Optical and Near-infrared Integral field spectrograph (HARMONI, Thatte et al., 2021) is an instrument made for a wide range of science cases, ranging from the study of quasars and distant galaxies to the characterization of exoplanets. The third first-light instrument of the ELT is the Mid-infrared ELT Imager and Spectrograph (METIS, Brandl et al., 2021). One of its main science goals is the detection and characterization of exoplanets, enabled through high-contrast imaging and spectroscopy. METIS will cover the atmospheric windows in the mid-infrared range of the spectrum (3-13 µm) with a series of filters, providing different high-contrast imaging modes.

METIS provides a classical vortex coronagraph mode in all bands, featuring focal plane vortex phase masks optimized for the L ($2.9 - 4.1\,\mu m$), M ($3.9 - 5.3\,\mu m$) and N-bands ($8.1 - 13.1\,\mu m$). In addition, METIS will feature a ring-apodized vortex coronagraph mode in L and M-band, while for the N-band this mode is not provided because the sky background is very high, and the ring apodizer would reduce the throughput considerably. Finally, METIS will feature an Apodizing Phase Plate covering the L and M-band.

The ELT is planning for other future instrument which are currently developed, such as the dispersion echelle spectrograph ANDES (Marconi et al., 2022) and the multi-object spectrometer MOSAIC (Sánchez-Janssen et al., 2020). In addition, the ambitious goal of detecting and characterizing Earth-like planets is pushing the development of a future Planetary Camera and Spectrograph (ELT-PCS, Kasper et al., 2021).

1.3.3 Space-based instruments

While ground-based instruments have been equipped with advanced coronagraphs, space-based instruments require mature and established technologies due to limited maintenance possibilities. Lyot-type coronagraphs are available in space since the

launch of the Hubble space telescope (Grady et al., 2003), and more recently of the James Webb space telescope, which in addition provides for the first time the four-quadrant phase mask on a space telescope (Boccaletti et al., 2022).

The next large space observatory to be launched in the coming years is the Roman space telescope, which is a widefield instrument targeted at surveying large portions of the sky. It will thereby study essential questions of cosmology, but also detect exoplanets via the microlensing effect. In addition to the widefield instrument, Roman features a coronagraphic instrument which aims to demonstrate the technology needed for future space based instruments tailored for high-contrast imaging applications. In this context, Roman will for the first time demonstrate active wavefront control in a space-based instrument (Kasdin et al., 2020), and serve as a laboratory to identify and study the technological challenges of the future great observatories targeting high-contrast imaging.

NASA's latest decadal survey for astronomy and astrophysics (Astro20, 2021) identifies the search for terrestrial exoplanets as a main science goal. This goal is addressed by the two originally proposed mission concepts HabEx (Gaudi et al., 2020) and LUVOIR (LUVOIR final report, 2019), which converged into the Habitable Worlds Observatory. Both HabEx and LUVOIR feature a high-performance vortex coronagraph that will provide the 10^{-10} contrast necessary to image an Earth-like planet around a Sun-like star. A concept based on a scalar vortex phase mask to achieve this ambitious goal is discussed in Chapter 11.

2 Focal-plane coronagraphs

2.1 Efforts towards the perfect coronagraph

In 1939 Bernard Lyot first combined an occulting mask with a downstream pupil stop (aka Lyot stop) to image the solar corona, giving birth to the first coronagraph. Since then, the general architecture of a coronagraph has remained essentially the same. A coronagraphic mask diffracts the light of an on-axis source outside a certain region of the pupil where it is blocked by an undersized diaphragm, the Lyot stop. In this way, faint objects surrounding a bright source can be observed without saturating the science detector.

Since the first coronagraph concept of Lyot, substantial efforts have been made towards the perfect coronagraph. New applications needed higher contrasts, and with larger telescopes the complexity of the optical setup increased needing advanced coronagraphic systems. Today's telescopes generally feature a large central obscuration with supporting struts which cause a non-negligible amount of stellar light to leak through the coronagraph, and the latest generation of telescopes featuring segmented mirrors have complicated shapes which have to be accounted for. Various techniques have been developed to tackle these issues and are nowadays often combined in coronagraph systems to obtain the most out of a modern telescope. It is the aim of this chapter to briefly retrace the path that led to one of today's most successful coronagraph systems, the vortex coronagraph.

2.1.1 Lyot coronagraph

The first and most intuitive design of a coronagraph was proposed by Bernard Lyot to observe the solar corona (Lyot, 1939). An opaque mask is introduced in the focal plane at the position of the sun to block out its blinding light, while the surrounding light of the faint corona is transmitted. The key invention of Lyot was to insert an undersized diaphragm (Lyot stop) in the relayed pupil plane to remove light diffracted from the hard edges of the mask. Figure 2.1 shows the optical setup of a Lyot coronagraph, which consists of a combination of focal-plane and pupil-plane masks. The light of the telescope is focused on an opaque mask covering the PSF core. Light diffracted by the sharp edges is then blocked by a slightly undersized Lyot stop in a relayed pupil plane, before the residual light is focused to compose the final coronagraphic image.

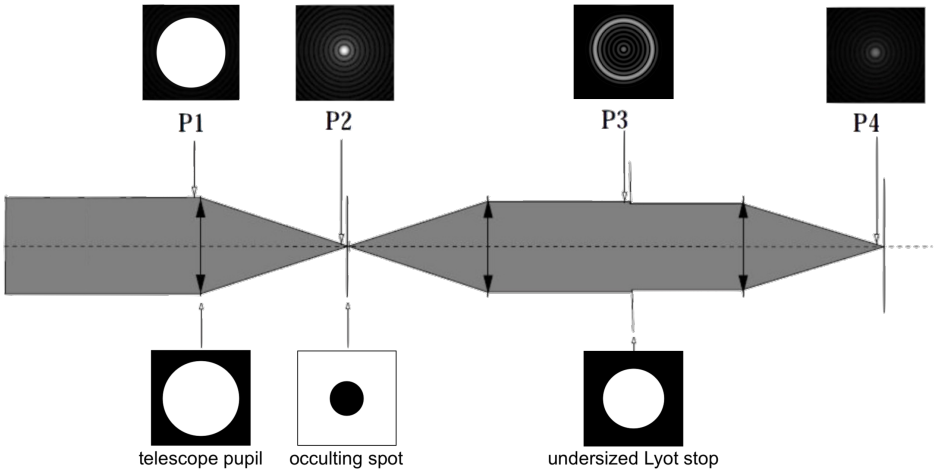

Figure 2.1: Configuration of the Lyot coronagraph along the optical axis. The distribution of the intensity is shown in the four planes P1-P4 (top), together with the occulting masks in P1-P3 (bottom). P1: Telescope pupil, P2: focal plane PSF before applying the Lyot occulting spot, P3: relayed pupil plane (Lyot plane) before the undersized diaphragm (Lyot stop), P4: Image plane with attenuated PSF.

The simple concept of Lyot is still used today, for example in the James Webb space telescope (Boccaletti et al., 2022).

The Lyot coronagraph suffers from two effects limiting its performance. First, the opaque mask blocks only the central part of the Airy pattern, so the outer rings not covered by the mask propagate to the Lyot plane resulting in residual starlight inside the pupil, with increasing intensity toward the pupil edges. This can be partially accounted for by using a strongly undersized Lyot stop, at the cost of reducing the throughput of the system. Moreover, complex aperture shapes with central obscurations and spiders create a similar effect, which can be addressed with a Lyot stop adapted to the pupil shape, but reducing the throughput even more. Second, the occulting spot has a finite size and therefore also blocks faint sources that are close to the bright star. Reducing the size of the opaque mask increases the discovery space, but reduces the achievable contrast. Since the discovery of the first exoplanet in 1995, the prospect of directly imaging such planets has driven the technology development of coronagraph concepts capable of achieving higher contrasts at smaller angular separations.

2.1.2 Roddier and dual zone phase mask

An alternative coronagraph design is based on a phase mask in the focal plane instead of an occulting spot (Roddier & Roddier, 1997). The phase mask features a small

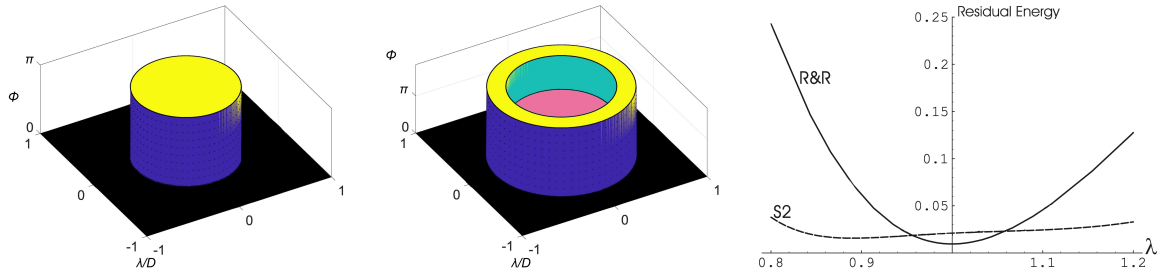

Figure 2.2: Design of the Roddier (left) and dual zone phase mask (center) together with their broadband performance (right, adapted from Soummer et al., 2003b). The Roddier phase mask (R&R) has high performance only at the design wavelength, while the dual zone phase mask (S2) provides decent performance across 40 % bandwidth. The radius and phase shift are $r_{R\&R} = 0.52\,\lambda/\mathrm{D}$ and $\phi_{R\&R} = \pi$ for the Roddier dimple, and $r_{S2,1} = 0.52\,\lambda/\mathrm{D}$, $\phi_{S2,1} = 0.94\pi$ and $r_{S2,2} = 0.70\,\lambda/\mathrm{D}$, $\phi_{S2,2} = 1.84\pi$ for the inner and outer phase step of the dual zone phase mask.

circular region at the center of the mask inducing a π phase shift. The size of this region is chosen such that half the light of an on-axis source acquires the π phase shift and interferes destructively with the other half not acquiring the phase shift. This design increases the discovery space for planets at smaller angular separation from the host star. The metric used to quantify the discovery space is the inner working angle of a coronagraph, which is the smallest separation from the on-axis source at which a companion can be detected (more precisely it is the distance at which 50 % of the light of the companion is transmitted). Though the Roddier phase mask works well in monochromatic light, the chromatic nature of the phase mask limits its performance in realistic conditions, since astronomical observations of faint planets are typically done in broadband filters. The chromatic behavior of the mask arises from its concept using a phase dimple covering a well-specified region of the Airy disk. While the physical size of the dimple is fixed, the size of the PSF varies with wavelength, which causes the phase mask to perform optimally only at the design wavelength. Moreover, the phase dimple is typically implemented as a physical step in a dielectric substrate and the phase shift itself is therefore chromatic as well.

An attempt to correct for the chromatic dependence has been done by Soummer et al. (2003b) introducing a second phase shifted zone. The resulting dual zone phase mask introduces a second phase step in the phase dimple, which provides an opposite phase shift to relax the chromaticity. Instead of having very good performance at the design wavelength and degrading rapidly when moving away from it, the dual zone phase mask provides a decent performance across a larger bandwidth, and is therefore more suited for astronomical observations in broadband filters. The design of a Roddier phase mask, dual zone phase mask, and their performance are shown in Figure 2.2. Nevertheless the correction is far from being perfect and other designs have been proposed to avoid a chromatic phase dimple.

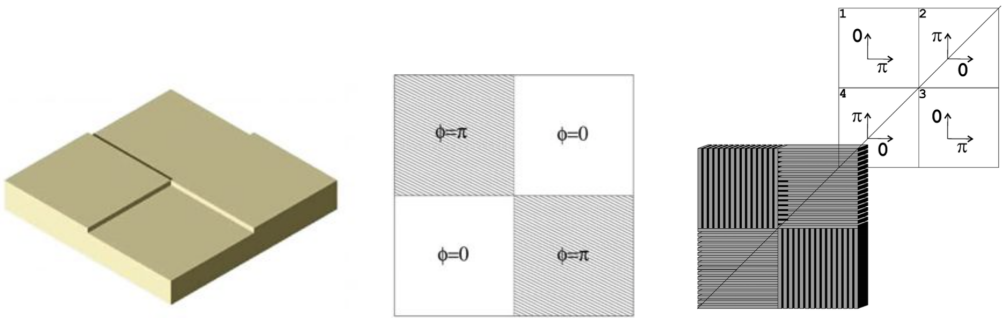

Figure 2.3: Four quadrant phase mask working principle. Physical shape of a four quadrant phase mask in a dielectric material (left), corresponding phase profile (center) and implementation using half-wave plates with orthogonal preferred axes (right). Half of the light (two of the four quadrants) acquires a π phase shift and interferes destructively with the other half not acquiring the phase shift. The half-wave plate implementation can be understood by considering the two polarization components independently. Each polarization component acquires a π phase shift in two of the four quadrants and interferes destructively with the light of the other two quadrants. Adapted from Rouan et al. (2007), Rouan et al. (2000).

2.1.3 Four quadrant phase mask

Instead of dividing the focal plane image in an inner and outer ring with a phase shift of π, Rouan et al. (2000) divided the focal plane image in four quadrants, of which two adjacent segments have a relative phase difference of π. This design is achromatic in the sense that its performance does not depend on the size of the PSF. The on-axis light in two diagonal quadrants interferes destructively with the light in the other two quadrants with π phase creating the coronagraphic effect. The general implementation using a physical step in a dielectric substrate is however still chromatic. One solution is to use an approach based on the geometric phase instead of the longitudinal phase delay. The geometric phase (Pancharatnam, 1956; Berry, 1987) can be imparted by a rotation of the polarization as commonly exploited in wave plates.

An achromatic way of implementing the geometric phase for the four quadrant phase mask is to use an achromatic half-wave plate providing a π phase shift between orthogonal polarizations. Placing half-wave plates with orthogonal fast axes relative to each other in neighboring quadrants then creates a π phase shift between the two segments. Figure 2.3 shows the design of a four quadrant phase mask and the working principle of the geometric phase.

Typically half-wave plates are made of a birefringent material with anisotropic dielectric permittivity. The two polarization states aligned with the fast and slow axis experience a different refractive index and therefore have different optical path lengths.

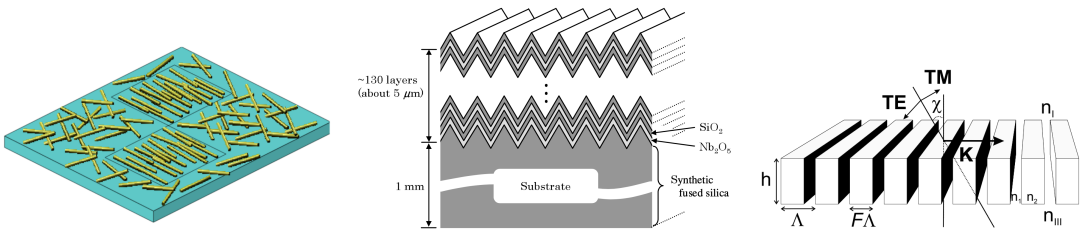

Figure 2.4: Achromatic half-wave plate concepts. Implementation based on liquid crystal polymers (left), photonic crystals (center) and subwavelength gratings (right). Adapted from (Mawet et al., 2011; Murakami et al., 2010; Mawet et al., 2005a).

This causes both polarizations to acquire a phase shift, which can be engineered to be exactly π by controlling the height of the half-wave plate. However, the π phase shift needs to be accurately controlled in a certain bandwidth, which can be challenging to a level required for coronagraphic applications. Achromatic half-wave plates can be implemented in several ways, including liquid crystal polymers, photonic crystals, and subwavelength gratings (Figure 2.4).

Liquid crystal polymers are a state of matter that have properties of liquid and crystalline materials. The polymers can be aligned using polarized light and then cured to provide a stable optical component. Liquid crystal polymers have proven to be suited to manufacture four quadrant phase masks (Mawet et al., 2009) and vortex phase masks (Tabirian et al., 2017).

Photonic crystals are artificial materials consisting of 3-dimensional periodic structures creating an artificial birefringence in the material. Both eight-octant (Murakami et al., 2010) and vortex phase masks (Murakami et al., 2012) have been manufactured using photonic crystals made of thin alternating layers of two dielectric materials. These masks provide good results in limited bandwidths (Murakami et al., 2012).

Subwavelength gratings are gratings with a period smaller than the wavelength. They are also known as zero-order gratings (ZOGs) since only the zeroth order is propagating, while higher diffraction orders are evanescent. Subwavelength gratings have the advantage that the difference in phase shift between the polarizations parallel and orthogonal to the grating can be made relatively achromatic by optimizing the grating parameters. This effect has been applied to the four quadrant phase mask (Mawet et al., 2005b), and the vortex phase mask (Mawet et al., 2005a), leading to the development of the annular groove phase mask described in the following sections.

2.2 Vortex phase masks

A problem encountered with the four quadrant phase mask is that a planet lying on the transition between two quadrants experiences destructive interference and is canceled

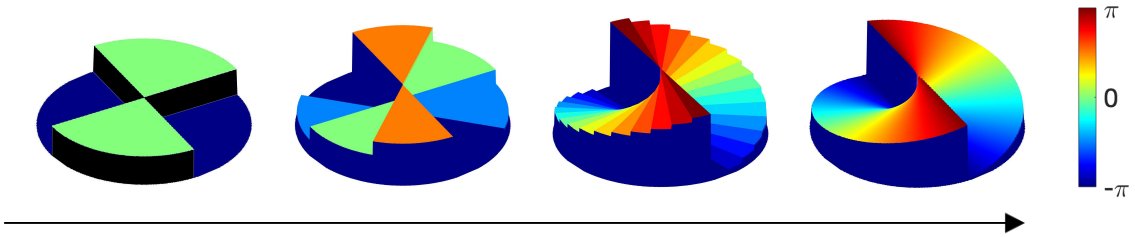

Figure 2.5: Phase mask with increasing number of segments. From left to right: 4, 8, 32 segments, which finally turn into the continuous helical phase ramp.

together with the stellar light. This effect can be mitigated by increasing the number of segments and therefore reducing the phase shift between neighboring segments. For eight segments the phase step is $\pi/2$, for 32 only $\pi/8$, and finally the phase approaches a continuous helical ramp, as illustrated in Figure 2.5.

The result of adding more segments is the vortex phase mask providing a continuous helical phase ramp with a phase singularity at the center. The helical phase ramp diffracts all the light of an on-axis source outside the pupil, where it is intercepted by a slightly undersized Lyot stop. Due to the absence of sharp edges, there is virtually no light leaking into the pupil, and the performance is only limited by the implementation of the phase ramp.

The main advantage of using an optical vortex instead of segmented designs is the continuous helical phase ramp. While an on-axis source is centered on the phase singularity and experiences the vortex effect, the PSF of an off-axis companion falls on a region where the acquired phase is nearly flat, as can be seen in Figure 2.6. Unlike the four quadrant phase mask which has dead zones along the transition regions between the quadrants, the vortex has a clear 360° discovery space.

A vortex phase mask placed in the focal plane of the telescope provides theoretically perfect cancellation of an on-axis source only for even topological charge. The topological charge ℓ_p of the vortex represents the steepness of the phase ramp and is defined as the number of times the geometric phase accumulates 2π around the optical axis: $\phi = \ell_\mathrm{p} \cdot \theta$, where θ is the azimuthal coordinate.

The attenuation of a point-like source as a function of the separation from the vortex singularity is shown in Figure 2.7. The characteristic metric to define the inner detection limit of a coronagraph is the inner working angle, which is defined as the separation from the center at which a companion is transmitted by 50 %. The inner working angle increases with the topological charge of the vortex. With a topological charge 2 vortex planets down to $0.9\,\lambda/\mathrm{D}$ can be detected, but the system is sensitive to low-order aberrations. Vortices of higher topological charge are more resilient to low-order aberrations. In fact, increasing the topological charge gradually removes the sensitivity to the lowest order modes tip/tilt, focus, astigmatism, and so on (Jolivet et al., 2019; Mawet et al., 2010; Ruane et al., 2018). In addition, the charge-2 vortex

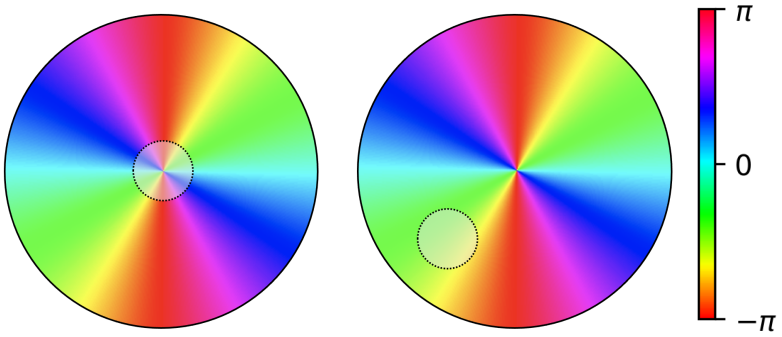

Figure 2.6: Position of the PSF of an on-axis (left) and an off-axis (right) source on the helical phase ramp in the focal plane. For clarity only the Airy-core is shown as dashed circle. The on-axis PSF is centered on the phase singularity and therefore sees a symmetric phase ramp resulting in destructive interference, while the off-axis PSF is centered on a part of the phase ramp which is relatively far from the singularity, and therefore only sees a small phase gradient resulting in high transmission.

with its small IWA is also sensitive to resolved stars. Figure 2.7 shows how any light at a certain separation from the center is partly transmitted by the coronagraph, which means that part of the PSF of an extended source is transmitted by a certain amount. The off-axis transmission curve scales with x^{ℓ_p} at small separations, which means that a vortex of higher topological charge cancels the light in a larger region, so any extended source (but also any companion) extending into this region are better canceled. In the end the choice of topological charge is a compromise between the inner working angle to detect planets at close angular separations to their host star, and the sensitivity of the coronagraph to low order aberrations and to partially resolved stars.

Vortex coronagraphs can be divided in two categories, based on the way the helical phase ramp is implemented. The vector vortex coronagraph exploits the geometric phase which is inherently achromatic, but polarization-sensitive. The scalar vortex coronagraph is based on longitudinal phase delays providing a chromatic phase ramp, but is polarization-insensitive.

2.2.1 Vector vortex phase masks

The concept of the vortex coronagraph relies on the helical phase ramp imprinted on the PSF in the focal plane of the telescope. Generally for astronomical applications it is important to observe in large bandwidths, especially when observing faint sources. It is therefore important that the vortex coronagraph provides the required contrast in a sufficiently large bandwidth.

The vector vortex coronagraph is a good candidate to achieve achromatic performance, since the design based on the geometric phase is inherently achromatic. The only

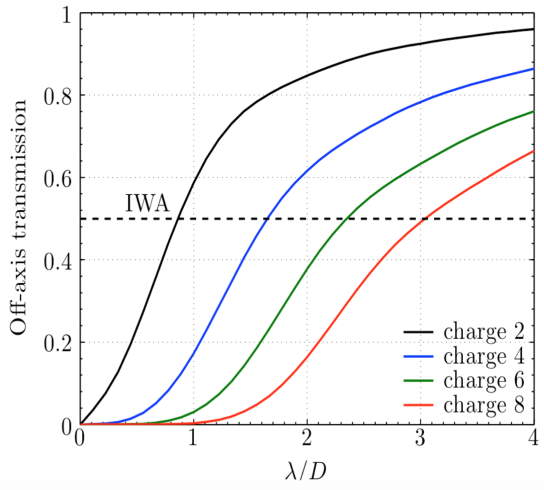

Figure 2.7: Attenuation of an off-axis source depending on the distance to the central singularity. A charge-2 vortex has an inner working angle (IWA) of $0.9\,\lambda/D$, which means it can detect companions down to this separation. Vortex phase masks of higher topological charge have larger inner working angles, but are more resilient to low order aberrations. Additionally, they can cancel the light from spatially resolved sources with a PSF spread out in the focal plane.

chromaticity arises from the actual implementation of the phase ramp using half-wave plates. One of the most successful approaches to implement a vector vortex coronagraph is based on subwavelength gratings. Since their geometric parameters can be engineered to provide an achromatic half-wave plate, subwavelength gratings are an excellent framework to implement a vector vortex phase mask.

Starting from the four quadrant phase mask design using subwavelength gratings (Mawet et al., 2005b) one can increase the number of segments to reduce the phase step between segments. The overall design then features segments with grating lines orthogonal to one segment edge, and when using a large number of segments this translates into a design featuring annular grooves providing the continuous helical phase ramp of an optical vortex.

2.2.2 Annular groove phase mask

The annular groove phase mask (AGPM, Mawet et al., 2005a) is the result of using subwavelength gratings to create the vortex phase ramp. The circular grating provides a vortex of topological charge 2. The working principle of the AGPM can be understood by considering the two circular polarizations independently. A circular polarization hitting a half-wave plate is flipped in helicity and acquires a geometric phase that is twice the angle of the preferred axis along the grating lines. By varying the grating

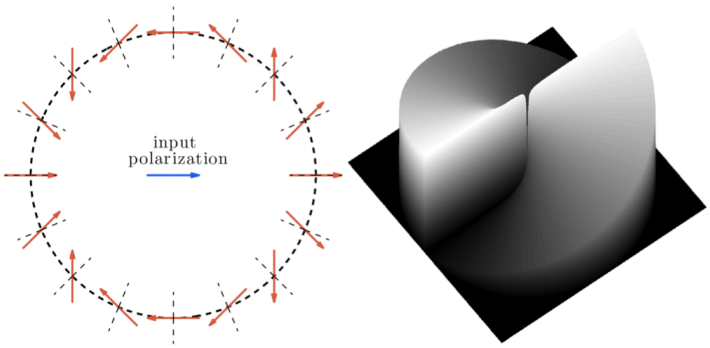

Figure 2.8: Left: Implementation of the geometric phase via rotation of the polarization. An incoming linear polarization is rotated by an angle of twice the azimuthal coordinate. Right: Resulting charge-2 phase ramp. From Mawet et al. (2011).

vector spatially as function of the azimuth angle θ one can therefore imprint a desired phase pattern ϕ in the form $\phi(\theta) = 2 \cdot \psi(\theta)$, where $\psi(\theta)$ is the direction of the grating vector.

The difference between both circular polarizations is that they acquire phase ramps of opposite helicity. In the ideal case both polarizations acquire a perfect phase ramp causing the vortex effect for each polarization independently. But in realistic conditions, when applying wavefront corrections with a deformable mirror to compensate for an imperfect wavefront, both polarizations need different wavefront correction and wavefront control can be challenging even with multiple deformable mirrors. Typically one would filter out one polarization to perform wavefront control for high-contrast applications. Nevertheless, for applications on ground based telescopes with lower contrast requirements the AGPM is an excellent coronagraphic mask without polarization filtering. Figure 2.8 shows the working principle of the geometric phase, along with the imparted phase ramp.

AGPMs have been demonstrated on sky on some of the most advanced telescopes to date, such as VLT/VISIR (Delacroix et al., 2012), VLT/NACO (Mawet et al., 2013a), LBT/LMIRCam (Defrère et al., 2014) and Keck/NIRC2 (Serabyn et al., 2017), and are being developed for the next generation of instruments (VLT/ERIS, Davies et al., 2023 and ELT/METIS, Carlomagno et al., 2020).

2.2.3 Manufacturing process of AGPMs

Several manufacturing approaches have been explored to implement the AGPM (Delacroix, 2013), leading to the development of a fabrication process based on multi-step plasma etching of polycrystalline diamond substrates (Forsberg & Karlsson, 2013a). More expensive monocrystalline diamond substrates have been investigated as well, but they do not provide any relevant advantage over the polycrystalline masks.

A schematic of the manufacturing process used for previous AGPMs is shown in Figure 2.9. The diamond substrate grown by chemical vapour deposition is cleaned with acetone and isopropanol, followed by piranha solution. The clean samples are then sputtered with a thick aluminum layer ($\sim 1\,\mu m$), followed by an intermediate silicon layer ($\sim 300\,nm$) and finally a thin aluminum layer ($\sim 300\,nm$). The layer thickness depends on the desired grating parameters, a deeper grating needing thicker mask layers. Finally a layer of photoresist is spin coated on top of the mask layers.

The grating pattern is not written directly to the stack of masking layers since it is based on expensive E-beam lithography. Instead, one master E-beam pattern is written and then transferred by molding a polydimethylsiloxane (PDMS) mask on top. The thick liquid PDMS is poured on the E-beam master and left resting for 20 min to allow uniform spreading and a roughly uniform thickness of the mask. Then the stack is baked at 75 °C for 20 min. Finally the soft PDMS mask featuring the negative grating pattern is peeled off the E-beam master, which can be re-utilized to mold further PDMS copies.

The pattern is transferred using a solvent-assisted micromolding technique (SAMIM, Kim et al., 1997). The PDMS mask is gently placed on the layered substrate right after spin coating with the photoresist layer to avoid any particle inclusion between the photoresist and the PDMS. This stack is then placed in a large covered petri dish together with a small amount of ethanol. The ethanol vapour slowly permeates the PDMS mask allowing the photoresist layer to fill the grooves of the PDMS mask by capillary effects. After completion of the pattern transfer ($\sim 30\,min$) the stack is baked to harden the photoresist, and the PDMS is peeled off the substrate leaving its pattern in the photoresist.

The substrate with the various mask layers is now ready for etching. The grating pattern in the photoresist is transferred to the substrate layer by layer using selective etching with different etch recipes. First, the thin Al layer is etched using a chlorine plasma (Cl_2/BCl_3), then the etched Al layer is used as a mask to etch the Si layer (Ar/SF_6 cycled with O_2), and finally the Si mask is used to etch the thick Al layer. The last step is to etch the diamond with an Ar/O_2 plasma with strong bias (ion bombardment) using the thick Al layer as a mask. During this long last etching step ($\sim 1\,h$), the etching is done in short intervals ($\sim 10\,min$) followed by breaks of similar duration to allow for the substrate to cool down.

The grating depth is difficult to infer from non-invasive methods. The AGPMs are therefore always etched for a slightly lower-than-optimal etch depth along with a reference sample. The latter is cracked to check the exact grating profile in the scanning electron microscope. From the geometry of the reference sample, the remaining etch time for the AGPM is calculated, enabling precise control of the etch depth.

Finally the AGPM is cleaned in acetone, isopropanol and piranha solution. More details about the manufacturing including the exact etch recipes for previous AGPM manufacturing can be found in Forsberg & Karlsson (2013a). The same process is used

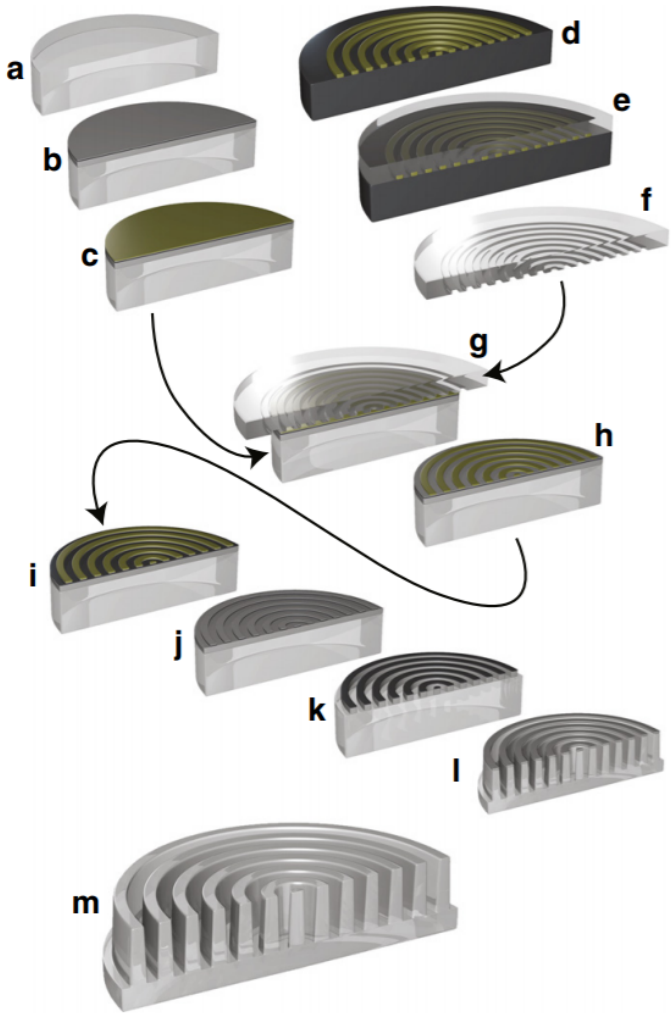

Figure 2.9: Manufacturing process schematics for fabrication of an AGPM. The substrate is sputtered with several metal layers (a-b), and then spin coated with a thin photoresist layer (c). The AGPM pattern is written only once on a master wafer using E-beam lithography (d) and a PDMS mask is molded on top of the E-beam master pattern (e). This soft mask (f) is then used to transfer the pattern to the substrate by using a solvent-assisted micromolding technique (g). The substrate is then baked to harden the photoresist and the soft mask is removed (h). Finally, selective etching in multiple steps is used to successively etch the metal layers (i-k) and finally the diamond substrate (l). The resulting diamond mask is cleaned in acid baths to remove residual mask material (m). From (Forsberg & Karlsson, 2013a).

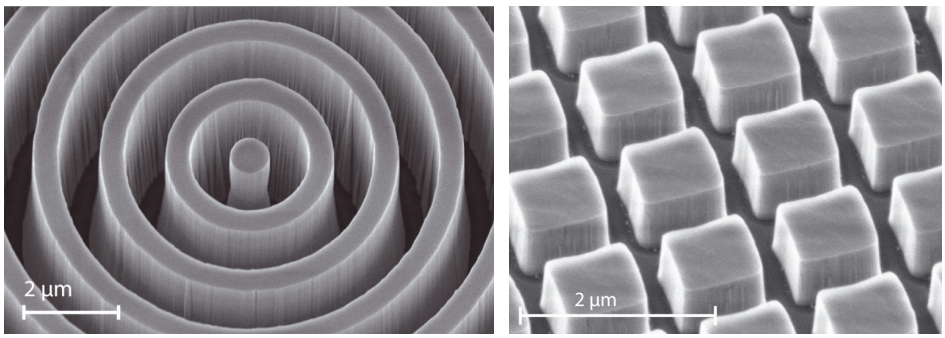

Figure 2.10: Left: Scanning electron microscope (SEM) image of the central portion of an AGPM. Right: SEM image of an anti-reflective grating for the L-band. From Delacroix et al. (2013).

to etch an anti-reflective grating on the back of the diamond substrate to mitigate internal reflections in the substrate, as described in more detail in Section 6.1. Figure 2.10 shows an AGPM and anti-reflective grating manufactured with this process.

This manufacturing process was partly modified during the writing of this thesis. While the overall process has remained essentially the same, the relevant changes and exact etch recipes are presented later on along with the results.

Due to the etching process using high bias, the cross section of the grooves is not rectangular but features a trapezoidal profile with a sidewall angle of a few degrees, as shown in Figure 2.11. The relevant grating parameters are the grating period Λ, the filling factor F (or equivalently the line width $F\Lambda$), the height h and the sidewall angle α. For $\alpha \neq 0$, the grating behaves as if it had an effective filling factor F_{equiv}. Depending on the etching recipe the sidewall angle can be partly controlled, but usually a sidewall-angle between 1 and 3° is achieved. The depth of the grooves is therefore limited by the geometry, as shown in the right graph of Figure 2.11. So far, gratings with aspect ratios of up to 13.5:1 (h:($\Lambda - F\Lambda$)) have been demonstrated with this method (Vargas Catalán et al., 2016a), but recent improvements of the manufacturing process and a new plasma etcher might reach even higher aspect ratios, as explained in Chapter 6. The geometry of the grating needs to be very well controlled since small changes in the grating parameters have a large influence on the effective filling factor and therefore on the performance of an AGPM. A variation of the effective filling factor can typically be compensated by adjusting the etch depth.

2.2.4 Subwavelength grating vortex phase masks of topological charge 4 and higher

The characteristic property of the AGPM is its simple design featuring a constant grating period everywhere on the mask. This is advantageous for manufacturing, but also presents a fundamental advantage compared to subwavelength grating vortex

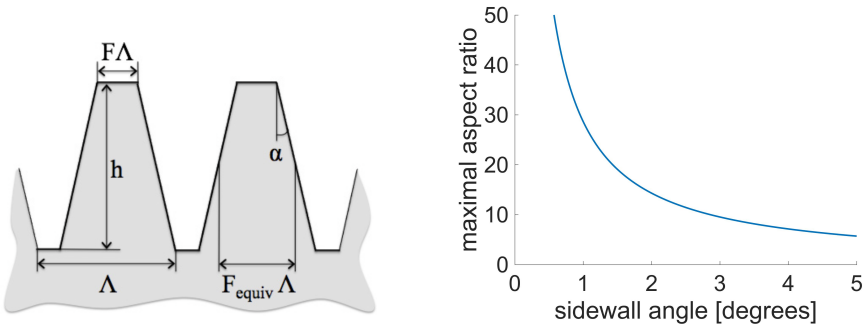

Figure 2.11: Left: Grating geometry with slanted sidewalls. A cross section of the grooves of the AGPM is shown. The relevant parameters are the grating period Λ, the filling factor F (or equivalently the line width $F\Lambda$), the height h and the sidewall angle α. The grating then behaves as if it had an effective filling factor F_{equiv}. From Delacroix et al. (2013). Right: Maximal achievable aspect ratio (defined as the grating height divided by the gap size) depending on the sidewall angle.

phase mask (SGVPM) designs of topological charge $\ell_p \neq 2$. Figure 2.12 shows that only the charge-2 design can be implemented without locally breaking the subwavelength requirement, which is needed to suppress higher diffraction orders.

A possible way to implement SGVPMs of higher topological charge is to divide the phase ramp back into segments for which the grating period satisfies the subwavelength condition. Two different ways of segmenting the subwavelength grating vortex phase mask design have been proposed (Delacroix et al., 2014): a design using segments with straight grating lines of constant period, and a design using segments with curved lines of decreasing period.

The design using straight grating lines divides the mask into wedges with the same grating parameters in all wedges. The grating lines are oriented according to the desired angle of the center of the wedge, which results in a staircase phase ramp (as shown earlier in Figure 2.5). This means, that there are zones on the phase mask where a companion is partially attenuated. On the other hand, this design features the same grating parameters everywhere on the mask which is advantageous for manufacturing due to the absence of the microloading effect (Hedlund et al., 1994). The microloading effect results in different etch depth for different filling factors and is described in more detail later in Section 11.7. A design featuring straight grating lines is shown in Figure 2.13 for a charge-4 vortex with 32 segments. It can be seen that this design features an unnecessarily large number of segments along the x-axis, where the grating lines can be approximated by nearly concentric curved lines.

A second design therefore features segments of varying size with curved grating lines. The advantage of this design is that the phase ramp is continuous, since the orientation of the grating lines follows exactly the nominal helical phase ramp, and no additional

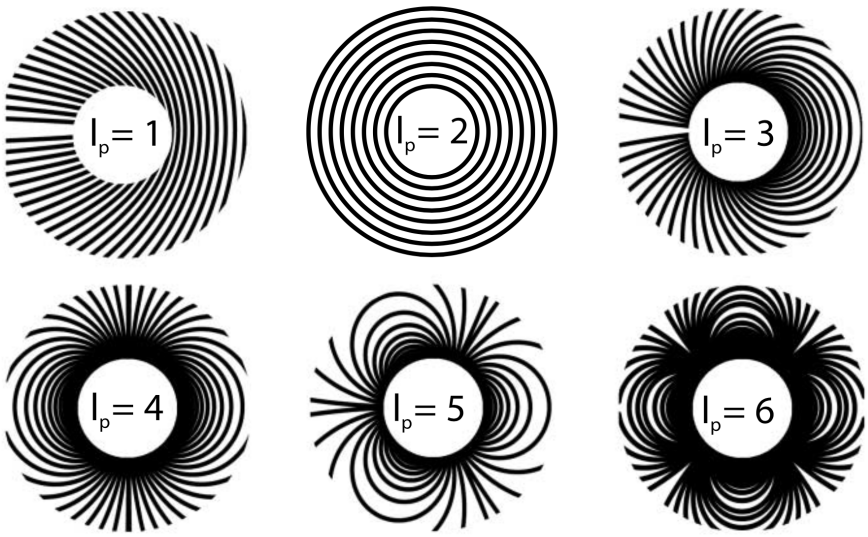

Figure 2.12: Theoretical design for subwavelength grating vortex phase masks of various topological charges l_p. From Mawet et al. (2005a).

attenuation of an off-axis source occurs. However, this implies that the grating period varies across the segments. Therefore the segments have to be chosen small enough to keep the variation of the grating period within reasonable limits. In particular, the period has to be small enough to be in the subwavelength domain, and large enough to be within reach for manufacturing. The design shown in Figure 2.13 features grating segments with a period that never drops below 80 % of the grating period Λ. However, this design leads to a large number of segments approaching the y-axis.

A promising approach is to combine the advantages of both designs in a hybrid design featuring curved lines in regions along the x-axis, and straight lines along the y-axis. This design is discussed more in detail in Chapter 8.

2.2.5 Scalar vortex phase masks

An alternative to the well-established vector vortex phase mask to implement the helical phase ramp is based on longitudinal phase delays. This scalar concept is naturally chromatic, but clever design choices can make scalar vortex phase masks more achromatic.

The simplest design for a scalar vortex is a helical step in a dielectric material, which provides the helical phase ramp at a specific wavelength. The height of the step depends on the desired topological charge and the material. The height variations in the material will therefore be higher for higher topological charge and lower refractive index. However when considering a certain bandwidth the performance quickly degrades. Combining two or more materials partly solves the chromaticity issue, but results in

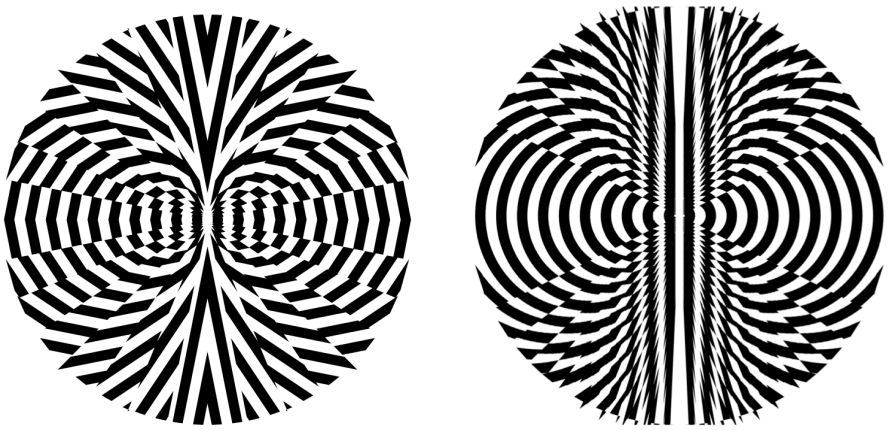

Figure 2.13: SGVPM designs of topological charge 4 based on (left) segments with straight grating lines and (right) segments with curved grating lines.

thick designs and an increased manufacturing complexity (Swartzlander, 2006; Ruane et al., 2019).

Another way to achromatize the design of the scalar vortex is clever phase wrapping. For example, a charge-6 vortex phase ramp spanning 12π can be divided into six segments spanning 2π each, which relaxes the manufacturing needs by reducing the overall height of the structure. The chromatic performance can further be improved by optimizing the topology of the phase ramp (Galicher et al., 2020; Desai et al., 2022). In this way, the broadband performance can be increased by up to one order of magnitude. Figure 2.14 shows several scalar vortex phase mask designs. The phase ramps are shown as 3d ramps and as azimuthal phase profiles.

2.2.6 Metasurfaces

Another way of implementing a scalar vortex phase mask is to use a full metasurface design. Metasurfaces are thin structures consisting of subwavelength building blocks in a single layer (Yu & Capasso, 2014). The subwavelength gratings used to create achromatic half-wave plates are a type of metasurface of increased symmetry. However they are polarization sensitive and not suited for a metasurface implementation of the scalar vortex. Polarization-independent 2-dimensional metasurfaces instead can provide the longitudinal phase delays needed for the scalar vortex in a single layer of constant thickness. Moreover, the 2d design of metasurfaces increases the degree of freedom and can therefore help to achromatize the scalar vortex design.

Metasurfaces can drastically reduce the thickness of optical elements. In particular, they enable flat designs where classical optics require bulky components. This is because classical optical components rely on phase shifts gradually accumulated when propagating through a dielectric, while metasurfaces provide an abrupt phase shift in a layer with subwavelength thickness. Tailoring this phase response spatially across

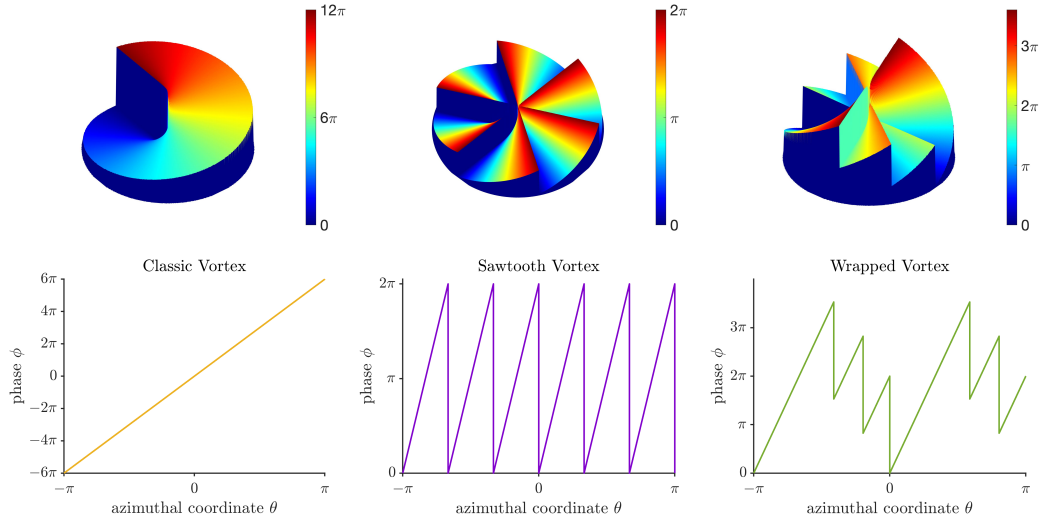

Figure 2.14: Different scalar vortex designs of topological charge 6 using phase wrapping. Phase ramp (top) and azimuthal phase profile (bottom) for classical vortex (left), sawtooth vortex (middle) and wrapped vortex design (right) presented in Desai et al. (2022).

the metasurface enables the implementation of arbitrary phase profiles for various applications. For instance, metasurface-based lenses have been fabricated with thickness $< 1\,\mu m$ (Khorasaninejad et al., 2016). The increased design freedom using abrupt phase shifts can further be used to implement generalized laws of reflection and refraction (Yu et al., 2011). Other applications include beam-steering (Shalaev et al., 2015), holography (Arbabi et al., 2015), wave plates (Yu et al., 2012) and polarimetry (Rubin et al., 2022). These applications rely on the engineering of the phase and transmission response of the metasurface building blocks.

Several implementations of polarization-independent metasurfaces have been reported in the past years (Jiang et al., 2013; Arbabi et al., 2015) and substantial efforts have been made to make metasurface-based optics achromatic by design (Shrestha et al., 2018; Chen et al., 2020). However, the achromatic behavior of metasurfaces for vortex phase masks for astronomical applications has not been studied in detail, and no metasurface implementation of vortex phase masks has been manufactured so far.

Metasurfaces can be divided into plasmonic (metallic) and all-dielectric metasurfaces, depending on their material platform. For high-contrast imaging applications, an all-dielectric framework is mandatory, since plasmonic metasurfaces suffer from high losses due to their metallic components, and high transmission efficiency is crucial in observing faint planets. Several applications of all-dielectric vortex beam generators have been reported, exploiting both the geometric phase (Shalaev et al., 2015; Devlin et al., 2017) and longitudinal phase delays (Chong et al., 2015). It is part of this thesis to investigate these metasurface designs for their ability to implement an achromatic scalar vortex phase mask.

2.3 Dealing with obstructions and struts

The vortex coronagraph provides perfect cancellation of an on-axis source for unobstructed circular pupils. While this concept works well in theory, virtually all large telescopes have central obstructions and support structures due to their secondary mirror. In addition, most recent large telescopes have segmented primary mirrors, that result in complex entrance pupils of the telescope. The performance of the vortex coronagraph is considerably degraded in these instruments, since the complex pupil shape causes part of the light which is normally rejected by the vortex to be diffracted back into the coronagraphic system. To provide high contrast, the vortex coronagraph has to therefore be adapted to these specific circumstances. Amplitude and phase apodizers can be used together with the coronagraphic mask to mitigate these errors.

2.3.1 Apodizers

Apodizers are pupil plane masks that are introduced in the telescope system to modulate the incoming wavefront in the pupil plane and redirect the light to specific regions in the focal plane. In this way it is possible to create a dark region in the focal plane without the need for a focal plane mask. Combined with focal plane masks, apodizers can also correct for the error introduced by complex pupil shapes. Two types of apodizers can be used to enhance the contrast: amplitude and phase masks. Figure 2.15 illustrates the two different apodization techniques.

The apodizing phase plate (APP) is a phase apodizer in the pupil plane. It produces an asymmetric PSF in the focal plane without the need for a coronagraphic mask in the focal plane. The APP has been demonstrated on sky at the MMT observatory (Kenworthy et al., 2007) and is part of the coronagraphic design of the METIS instrument at the ELT (Brandl et al., 2021).

Shaped pupils are a type of binary amplitude apodization which have shown to provide dark regions in the focal plane without the need for a coronagraphic focal plane mask. These designs have shown to provide extended dark regions in the focal plane (Kasdin et al., 2003), while maintaining a simple binary design.

Amplitude apodization is also used in the apodized Lyot coronagraph (Aime et al., 2002; Soummer et al., 2003a). In this case, an amplitude apodizer similar to a Gaussian profile is introduced in the entrance pupil to smoothen the sharp edges of the pupil, thereby increasing the contrast by factors of typically 10^4 to 10^6.

2.3.2 Ring-apodized vortex coronagraph

Another application of an amplitude apodizer is the ring apodized vortex coronagraph (RAVC, Mawet et al., 2013b). A classical vortex coronagraph provides perfect starlight cancellation for an unobstructed circular pupil. However, large telescopes have more

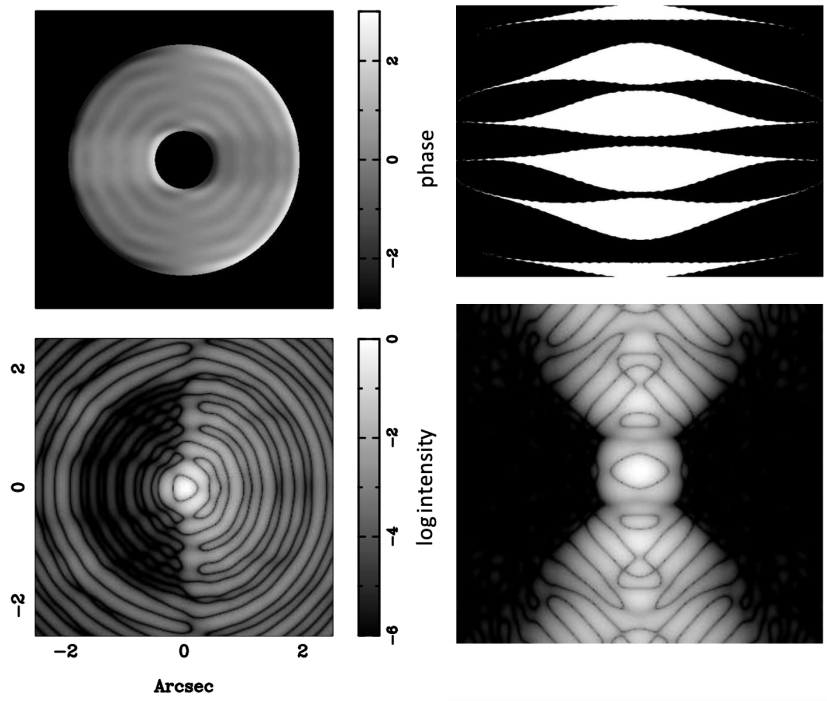

Figure 2.15: Working principle of different pupil plane apodizers. Left: Apodizing phase plate creating a dark region in the focal plane via phase modulation of the entrance pupil. Top: phase modulated pupil, bottom: focal plane intensity. From Kenworthy et al. (2007). Right: Binary shaped pupil creating two dark regions in the focal plane. Top: binary pupil shape, bottom: focal plane intensity. The dark regions theoretically reach contrasts below 10^{-10}. From Kasdin et al. (2003).

complex pupil shapes, like segmented mirrors, a central obscuration caused by the secondary mirror and support structures. These elements create diffraction causing a non-negligible fraction of the starlight to leak through the coronagraph. An apodizing mask in the pupil plane of the telescope can partially correct for this effect. In the case of a circular pupil with a circular central obscuration a relatively simple solution is available, which introduces an amplitude apodizer in the pupil plane featuring a series of discrete grayscale rings (Mawet et al., 2013b). Such a ring-apodizer restores the perfect behavior of the classical vortex coronagraph at the expense of a grayscale region in the pupil and an increased Lyot stop inner diameter, resulting in reduced throughput. Figure 2.16 shows the effect of the ring apodizer on the coronagraphic system.

In the case of a charge-2 vortex, the pupil plane apodizer of the RAVC consists of a grayscale mask featuring a single grayscale ring of a certain radius and transmission. For higher topological charge the number of discrete rings needed to restore the perfect behavior of the vortex scales with the topological charge, more precisely the number of rings required equals half the topological charge.

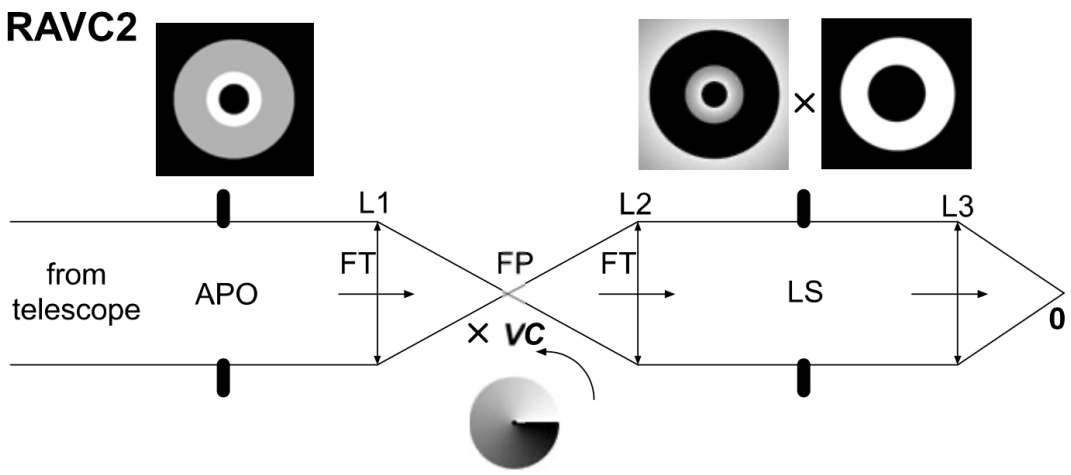

Figure 2.16: Effect of a ring apodizer in the coronagraphic system. The ring apodizer in the entrance pupil creates a dark region in the Lyot plane. A tailored undersized Lyot stop removes the residual light restoring the perfect null in the coronagraphic image. The three insets show (left) the apodized pupil, and (right) the intensity distribution in the Lyot plane, as well as the tailored Lyot stop blocking the residual light. From Mawet et al. (2013b).

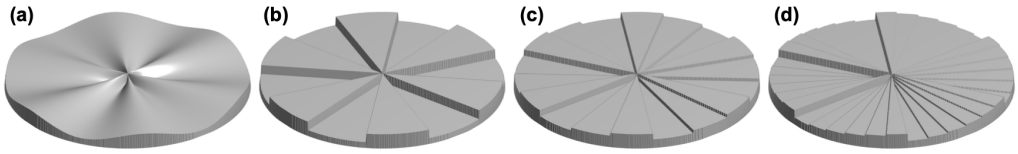

Figure 2.17: Alternative phase mask topologies providing the coronagraphic effect. All masks have a charge-6 like behavior. From Ruane et al. (2019).

2.4 Other coronagraphic focal-plane masks

The vortex phase mask is an established coronagraphic concept providing optimal starlight cancellation. However, it is part of a larger family of phase masks providing on-axis starlight cancellation applicable to coronagraphy. In particular, any purely azimuthal phase profile that can be decomposed into even vortex modes (topological charge $\ell_p = \pm 2, \pm 4, \text{etc.}$) provides the coronagraphic effect (Ruane et al., 2019). Figure 2.17 shows a series of phase masks obtained from this principle. A promising design is the wrapped vortex discussed in Section 2.2.5, increasing the chromatic performance of the vortex by one order of magnitude. Another promising design is the cosine phase mask, consisting of an azimuthally smooth cosine-like phase profile.

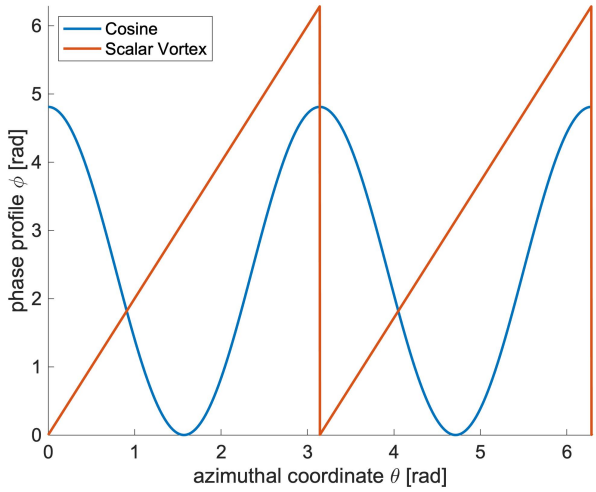

Figure 2.18: Azimuthal phase profile of cosine and scalar vortex phase mask. The phase coverage needed to implement the cosine phase mask is only $\Delta\phi = 4.80966$, while the scalar vortex needs at least $\Delta\phi = 2\pi = 6.28319$.

2.4.1 Cosine phase mask

The cosine phase mask is an example of a purely azimuthal phase mask providing perfect starlight cancellation similar to the vortex phase mask (Hénault, 2018). However, the advantage of a cosine phase profile is the absence of 2π phase jumps. For vector vortex designs this is not an issue, since the phase is continuously modulated by imparting a geometric phase and there is no physical region where a 2π phase jump occurs (for the AGPM the grating design is continuous, and for higher topological charge the only phase error arises from the segmented designs). But for scalar vortex designs relying on longitudinal phase delays the 2π phase jump translates into a physical region on the mask with sharp design features. This can be implemented as a sharp physical step in a dielectric substrate, or the abrupt change in metasurface block size breaking the local periodicity of the design (see Section 11.4).

The cosine phase mask design also uses a smaller phase coverage across the mask ($\Delta\phi = 2 \cdot z_0 = 2 \cdot 2.40483 < 2\pi$, where z_0 is the first zero of the zeroth order Bessel function, Hénault, 2015), as shown in Figure 2.18. This design reduces the overall thickness of a dielectric mask or the design metasurface building block size variation across the mask, relaxing manufacturing constraints. The cosine phase mask is therefore a promising candidate for a first metasurface phase mask prototype and its design is therefore considered more in detail in Section 11.5.

3 Scope and goal of this thesis

The upcoming generation of ground- and space-based telescopes will push the technology limits in the field of high-contrast imaging towards the direct detection of Earth-like planets around Sun-like stars. Driven by this science goal, the performance of current coronagraphs needs to be improved. This includes the optimization of current vortex phase masks such as the AGPM, which are a prime candidate for implementation in future flagship projects like the Planetary Camera and Spectrograph (PCS) of the ELT, but also the exploration of completely new designs based on two-dimensional metasurfaces which could provide the necessary requirements for the Habitable Worlds Observatory (HWO). These metasurface building blocks introduce additional degrees of freedom, which can be leveraged to achromatize the vortex performance and increase its topological charge. The goal of this thesis is to develop such high-performance metasurface-based vortex phase masks pushing the limits of current microstructured vortex phase mask technologies.

3.1 Goals of this thesis

AGPMs have been installed on many of the most advanced ground-based telescopes in the past years. Due to their proven capability of providing efficient starlight cancellation, they have been chosen as the technology to implement the vortex coronagraph modes of METIS, one of the first light instruments of the Extremely Large Telescope. METIS will feature several AGPMs optimized for different wavelength bands in the mid-infrared spectral range, enabling high-contrast imaging of terrestrial exoplanets. In addition, it will feature the ring-apodized vortex coronagraph mode for the first time on-sky in the L and M-band. The design optimization of the new METIS AGPMs and ring apodizer, their manufacturing, and characterization are therefore the first goal of this thesis.

The grating parameters of the AGPM have so far been optimized using RCWA, which assumes a periodic pattern of straight grating lines. This method has successfully been used to optimize and manufacture AGPMs reaching null depths[1] below 10^{-3}. However, the assumption of straight grating lines does not hold for the region of the AGPM close to the phase singularity, where the grating curvature becomes significant. Other aperiodic design features in microstructured vortex phase masks can not be described

[1]The null depth is a measure of the attenuation of an on-axis source provided by the coronagraph. It is introduced in more detail in Section 5.3.

accurately with this method. This is the case for segmented designs of higher topological charge, or aperiodic nanopost designs for scalar vortex phase masks. Another goal of this thesis is therefore the implementation of a simulation framework using the FDTD method capable of properly describing aperiodic features of microstructured vortex phase masks.

Null depths down to 3×10^{-4} have been measured in lab for the best AGPMs manufactured so far, which is well above the noise floor of the coronagraphic test setup. However, RCWA simulations show that in principle a better null depth can be reached with the nominal grating parameters of the best AGPMs. So far it is unclear whether the in-lab performance limitation arises from the manufacturing uncertainty on the grating parameters, or from the influence of the AGPM central region, for which RCWA has been shown to break down. Therefore, the third goal of this thesis is to study the effect of the imperfect performance of the AGPM at its center on the null depth measured in a realistic coronagraphic setup, which can then be compared with the thorough simulation framework described above.

In addition to reaching higher contrasts, higher topological charge vortices will also benefit the future generation of extremely large telescopes by reducing the sensitivity to low order aberrations and to partially resolved stars. However at higher topological charges, the simple circular groove design of the AGPM becomes more complex, leading to segmented designs in order to avoid the spatially variant grating parameters across the mask. These designs have characteristic regions of aperiodic features causing an additional leakage term, which has to be accurately described with the simulation framework described above. A further goal of this thesis is therefore to investigate vortex phase mask designs of topological charge 4 featuring segmented subwavelength grating patterns.

In addition to the optimization of subwavelength grating designs, it is useful to explore other implementations of vortex phase masks based on more complex microstructured designs. The large spread of two-dimensional metasurfaces for various applications has shown a lot of potential in making microstructured optical components achromatic. Furthermore, phase masks based on two-dimensional metasurface building blocks (nanoposts) can potentially mitigate the aperiodic features encountered with designs of subwavelength gratings of topological charge 4. The exploration and assessment of two-dimensional metasurface patterns and their application to the design of vortex phase masks is therefore the last goal of this thesis.

The presented goals can be achieved by dividing the work performed during this thesis into three main steps. First, a thorough simulation framework is developed capable of properly simulating the performance of vortex phase masks. Second, the simulation framework is used to optimize and improve current subwavelength grating vortex phase mask designs, which combined with an updated manufacturing process leads to the fabrication of charge-2 and charge-4 vortex phase masks. Finally, the same simulation tools are used to explore new designs based on two-dimensional metasurface building

blocks, which require specific manufacturing tests to be performed before they can be produced.

3.2 Structure of this manuscript

The present manuscript covers the goals presented in this chapter in the following structure.

Part I of this thesis introduces the reader to the necessary background to understand the results presented in this manuscript. After a brief introduction to the scientific background of high-contrast imaging applications in Chapter 1, the technical concepts leading to the development of current vortex coronagraph technologies are presented in Chapter 2. These two sections aim to put this work into context before describing the goals of this thesis in Chapter 3 and provide the foundation of the results presented in the following chapters.

Part II describes the methods used within this thesis. Chapter 4 describes the thorough simulation framework developed during the first year of this thesis project. The optimal use of the different simulation tools for coronagraphic applications as well as the challenges encountered and the strategies to solve them are discussed. Chapter 5 describes the performance metrics associated with these simulations, and how they are used to assess the performance of vortex phase masks in different contexts. It also covers the experimental test setup used to characterize AGPMs in-lab.

After introducing the scientific background and presenting the methodology used in this thesis, Part III discusses their application to the design of vortex phase masks based on subwavelength gratings. Chapter 6 presents the typical workflow when developing AGPMs, applied to the development of the AGPMs for the METIS instrument. This covers the motivation of the contrast requirements and the design optimization of AGPMs in the mid-infrared, as well as the manufacturing of the first N-band AGPMs using an updated manufacturing process and their preliminary characterization. Chapter 7 then discusses a possible limitation of the RCWA simulation framework used so far for high-performance AGPMs arising from the central phase singularity and presents quantitative results showing the effect of the AGPM central region, which are confirmed experimentally. Chapter 8 finally discusses the application of subwavelength gratings to vortex phase mask designs of higher topological charge, and proposes a promising design combining the advantages of two patterns as well as first manufacturing steps. This chapter concludes the discussion of subwavelength grating vortex phase masks.

Part IV discusses the potential of two-dimensional metasurface patterns for vortex phase masks. Chapter 9 reconsiders the optimization of the AGPM center and extends it allowing for more design freedom at the AGPM center using a pattern of rectangular nanofins. This chapter applies a machine learning approach to the complex design

problem involving many free parameters. Chapter 10 extends the metasurface framework to the design of the full phase mask and discusses the limitations encountered with rectangular metasurface building blocks when optimizing the design of vector vortex phase masks. Chapter 11 finally presents a scalar metasurface framework avoiding the issues encountered with rectangular metasurface blocks, and applies it to the design of scalar vortex phase masks.

Finally, Part V summarizes the results, putting them into context of the goals of this thesis. In addition, it describes the perspectives for future research, building on the results obtained so far within this thesis.

Part VI includes an appendix presenting additional work done during this thesis which does not fit within its direct scope. Chapter A discusses the application of the simulation framework developed within this thesis to the optimization of the METIS ring apodizer design using a semi-random microdot pattern to create a pupil mask featuring a grayscale region. Even though the ring apodizer is an important concept to increase the performance of the METIS coronagraphic modes, this chapter is placed in the appendix because it does not pertain to the development of vortex phase masks themselves.

Part II

Methods

4 Simulation framework

The microstructured metasurface components introduced in the previous chapters rely on the interaction between light and matter, which is governed by Maxwell's equations. It is therefore mandatory to use thorough simulation tools to accurately describe and predict the performance of new components. This chapter presents the different simulation tools used to describe the behavior of these components.

Several simulation tools can be used to describe microstructured optical components depending on their characteristic feature size. If the feature size is well above the wavelength, Fourier optics describes the light propagation sufficiently well. On the other hand, if the features are much smaller than the wavelength, effective medium theory (Zhang & Wu, 2015) is a good approximation and the description of complex microstructured materials can be simplified by assuming effective material properties. However, for building blocks of intermediate size just below the subwavelength limit, both theories are inappropriate, and full rigorous simulation tools considering the vectorial nature of light are needed to properly describe all the phenomena arising on these scales. Most optical components studied in this thesis rely on designs with characteristic feature size in the regime close to the subwavelength limit, and therefore rigorous simulation tools are needed to accurately describe their behavior.

Several rigorous methods for solving Maxwell's equations can be used to properly describe structures in the subwavelength regime. The following sections present the tools used in this thesis and describe their applicability and limitations for different scenarios. Two scenarios have to be simulated to determine the performance of a coronagraph system. First, the precise electric field response of each component has to be calculated in a near-field configuration. The results are then used to simulate the full coronagraphic system in far field configuration, which is governed by Fourier optics.

4.1 Simulating the electric field response

Phase and amplitude masks consisting of subwavelength structures have to be simulated with rigorous methods to properly describe the vectorial nature of light. Several tools are available to solve Maxwell's equations either in time-domain, or in reciprocal space in frequency domain. Carrying out the simulations in either domain has advantages and drawbacks.

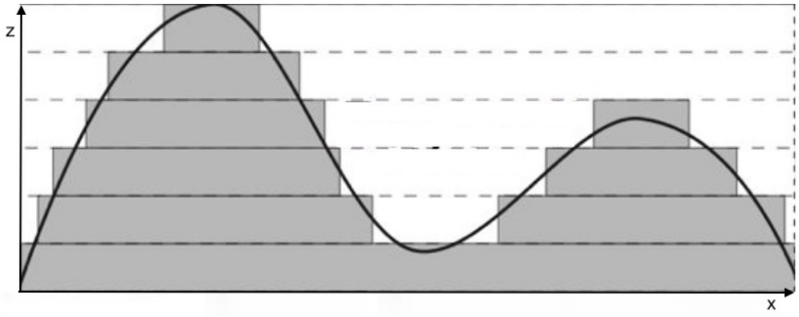

Figure 4.1: Discretization of a complex grating profile for the use with RCWA. The grating profile is divided into a number of layers and approximated by a staircase profile. RCWA is used to solve each layer independently and then combine the results to obtain the overall response of the structure. Adapted from Mawet (2006).

4.1.1 Rigorous coupled-wave analysis

Rigorous coupled-wave analysis (RCWA) is a method for solving Maxwell's equations in frequency space (Moharam & Gaylord, 1981). This method is particularly suited to describe periodic grating structures, since the geometry is translated into frequency space and the solver assumes periodic boundary conditions. A periodic repetition of the grating structure is required in order to expand the dielectric permittivity and the fields in a Fourier series. A home-made RCWA code developed at the University of Liège (Mawet, 2006) implemented in Matlab is used to describe the behavior of one-dimensional periodic gratings as used for the AGPM. RCWA has the advantage that the calculation only needs relatively little computational resources and is therefore suited to do brute-force exploration of the parameter space spanned by the grating parameters.

The simulation of subwavelength gratings featuring slanted sidewalls is possible with RCWA by approximating the grating geometry with a staircase profile, as shown in Figure 4.1. RCWA is then used to solve each layer independently, and the overall response of the grating is calculated by combining the results of the individual layers. More details about the RCWA description can be found in Mawet (2006) and Delacroix (2013).

Depending on the desired application, it is important to use sufficiently high resolution to compute an accurate phase response. The resolution of an RCWA simulation is represented by the number of plane waves used to compute the field response, including higher (evanescent) diffraction orders. For example, to obtain an accurate simulation of the null depth of an AGPM design, the phase response of the grating has to be known very precisely and the number of plane waves chosen has to be sufficiently large (typically > 40). On the other hand, less plane waves are sufficient (typically > 8) to simulate the behavior of anti-reflective gratings, where the important metric is the transmission. While the transmission is obtained accurately at relatively low

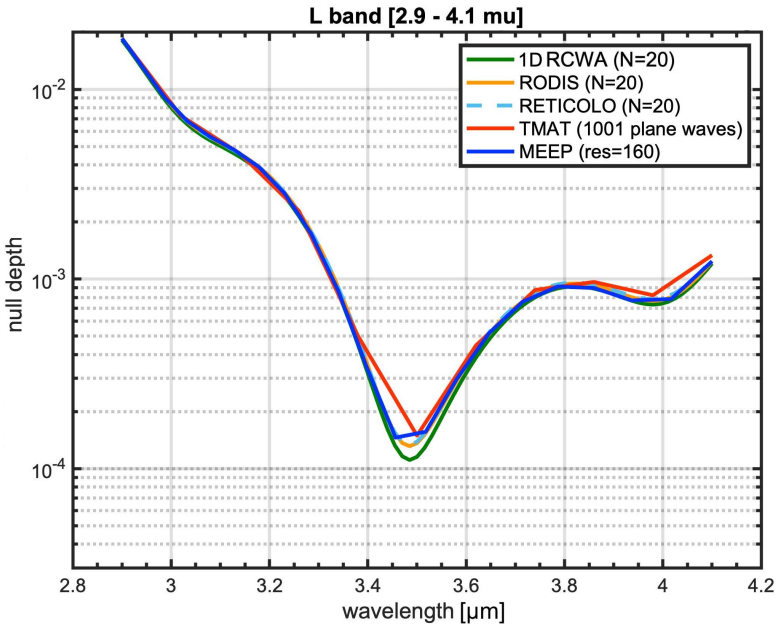

Figure 4.2: Simulated null depth of a one-dimensional straight grating obtained with different RCWA implementations. While all implementations converge to similar results, RETICOLO converges faster providing a modular and user-friendly interface. N is the number of the orders considered in the simulation, and is related to the number of plane waves $2N + 1$. The curve obtained with MEEP for periodic boundaries is shown as comparison (res is the resolution in pxl(λ/D)).

resolution, it necessitates an RCWA software capable of simulating crossed grating geometries.

4.1.2 RCWA for two-dimensional gratings

For the simulation of anti-reflective gratings and metasurface building blocks the two-dimensional RCWA code RETICOLO is used (Hugonin & Lalanne, 2005). This software implements RCWA in Matlab for crossed grating geometries including arbitrary periodic two-dimensional structures. Other implementations of RCWA exist, such as T-MAT (Deparis et al., 2009; Vigneron & Lousse, 2006), S4 (Liu & Fan, 2012) or RODIS (Dhoedt, 1995; Delbeke, 2002), all optimized for different applications. These tools lead to similar simulation results, as shown in Figure 4.2. The choice of RETICOLO for this work is based on its optimization for the precise phase retrieval necessary for accurately simulating vortex phase masks, enabling faster convergence than the other tools, and at the same time providing a modular and user-friendly interface.

Complex geometries simulated with RETICOLO need to be decomposed into rectangular shapes, and are approximated as a pixelized version of the nominal geometry.

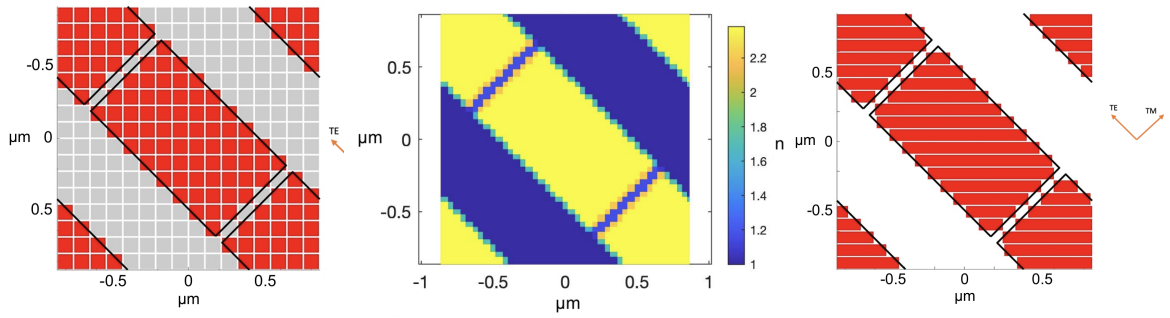

Figure 4.3: Discretization of rectangular metasurface building blocks (nanofins) that are not aligned with the simulation cell. Different approaches to discretize the nanofins are shown, starting with the projection of the nominal geometry onto a binary cartesian pixelization grid (left), adding intermediate gray scale values to smoothen the sharp features (center), and using rectangular blocks to approach the exact geometry (right).

To increase the accuracy of the simulation, a finer pixelization grid has to be used at the expense of higher computational cost. Rectangular shapes aligned with the pixelization axes are treated very efficiently, while circular blocks and rotated rectangular blocks (nanofins) need more computational resources for the results to converge (up to several orders of magnitude). This can be an issue especially when retrieving the precise phase response of nanofins that are not aligned with the simulation cell. Different approaches can be used to reduce the computational need of these simulations, as shown in Figure 4.3. One way to reduce the computational cost when simulating non-cartesian geometries is to use intermediate values for the refractive index of pixels at sharp transitions of the dielectric permittivity. In practice, using intermediate grayscale pixels is more representative of the effective material distribution and therefore describes the geometry more accurately. However, the convergence is still slow since the number of pixels remains high. Another way to improve the efficiency of the simulations is to divide the nominal geometry in slices, effectively consisting of long rectangles. In this way, the overall number of blocks used to discretize the nominal geometry is reduced considerably.

Even still, the discretization approaches mentioned above are too computationally heavy and it is difficult to achieve convergence with reasonable computational resources, especially for the exact phase response of complex metasurface shapes. For rectangular shapes, it is possible to exploit the advantage of aligning the block axes with the axes of the simulation cell. Instead of discretizing the nominal geometry in rectangular blocks, a supercell containing several nanofins aligned with the nanofin axes can be used (Figure 4.4). This approach does not make any approximation regarding the geometry. For every rotation angle of the nanofin that can be expressed as the arctangent of a fraction of two integer numbers ($\tan(\psi) = P/Q$), a supercell containing several nanofins aligned with them exists. For fractions of small integers (such as 1:1, 1:2, 2:3, 1:3) the supercell is sufficiently small to be simulated with reasonable

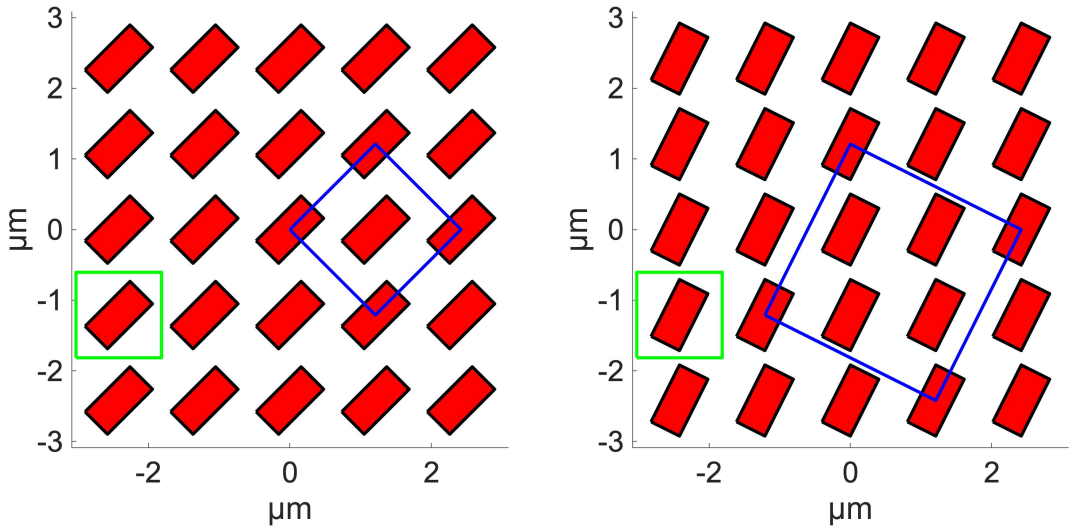

Figure 4.4: Supercell approach for the simulation of rotated rectangular nanofins. Instead of rotating the nanofin in its unit cell, a supercell containing several nanofins can be chosen to be aligned with them. This is possible for rotation angles ψ that can be expressed as $\tan\psi = P/Q$, where P and Q are integers. The supercell (blue) is shown for a rotation angle of $\psi = \tan^{-1}(1:1) = 45°$ (left) and $\psi = \tan^{-1}(1:2) = 26.6°$ (right) together with the unit cell (green).

computational resources (notably on a regular laptop within a few hours). The drawback of this method is that only discrete rotation angles can be simulated, and that a larger simulation cell results in more plane waves to properly describe the geometry. This method is used to simulate the nanofin designs discussed in Chapter 10.

4.1.3 The finite-difference time-domain method

The RCWA method presented in the previous sections is an appropriate tool to simulate infinite periodic structures. While this is a good approximation for metasurface patterns where each metasurface building block is surrounded by blocks of similar geometric parameters, this assumption does not hold at the central singularity of a vortex phase mask. It is therefore important to use a full 3d simulation tool capable of simulating a finite patch of the vortex phase mask.

The finite-difference time-domain method (FDTD, Taflove & Hagness, 2005) is such a 3d simulation tool. FDTD is a brute-force approach for solving Maxwell's equations in time domain, and is therefore computationally very expensive. Simulating a full AGPM with FDTD is unrealistic, and not really needed, since the behavior of the grating far from the center is well described with RCWA assuming a periodic infinite grating. However, at the center of the AGPM close to the phase singularity, the grating

curvature becomes non-negligible and a full 3d finite simulation tool such as FDTD is needed to describe the behavior of the grating correctly.

Similarly, the center of a metasurface mask is not well described with RCWA due to the phase singularity. More precisely, RCWA is an appropriate tool for simulating regions of a metasurface mask that are locally periodic, that is where neighboring unit cells are sufficiently similar to each other. This is the case for most regions of metasurface scalar vortex phase masks, since a continuously varying phase profile is expected to translate into a continuously varying geometry. However, at the center of the vortex the phase gradient becomes very high and the design is not locally periodic anymore. The same effect happens at the characteristic phase discontinuity of a scalar vortex. Therefore, FDTD is needed to properly simulate the response of these regions of the mask featuring a high phase gradient.

The FDTD implementation used within this thesis is MEEP (Oskooi et al., 2010), an open source software available as a python package featuring parallel computing compatibility via the message passing interface protocol (MPI). MEEP is used on the high performance computing facilities at the University of Liège (CECI[1] and Alan cluster) to perform embarrassingly parallel FDTD simulations, providing the necessary computational resources to simulate critical parts of the phase masks studied in this thesis. A typical simulation setup is shown in Figure 4.5. A plane wave source is excited a few µm before the grating structure. The wave propagates through the grating and the fields stabilize after a certain simulation time. The resulting electric field is then monitored at a certain distance after the grating inside the substrate. The simulation cell is surrounded by perfectly matched layers (PMLs), which are artificial regions with matched impedance resulting in perfect absorption. In practice, the PMLs are discretized by the finite resolution (pxl/µm) of the simulation cell. Therefore their thickness has to be large enough to reduce reflection artifacts at the cell boundaries. For high-resolution simulations of the AGPM the PMLs can take up to 90 % of the simulation volume, but this is needed to provide the necessary accuracy to assess the AGPM performance.

The choice of MEEP over other FDTD implementations such as Lumerical (Lumerical Inc.) and XFdtd (RemCom Inc.) is based on its implementation as a python package which makes it easy to integrate in a simulation workflow. Moreover, MEEP features parallel computing compatibility which is necessary for running simulations of relevant parts of the phase masks considered in this thesis. Finally, MEEP's source code is openly accessible in the spirit of Open Science (UNESCO, 2021).

[1]Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

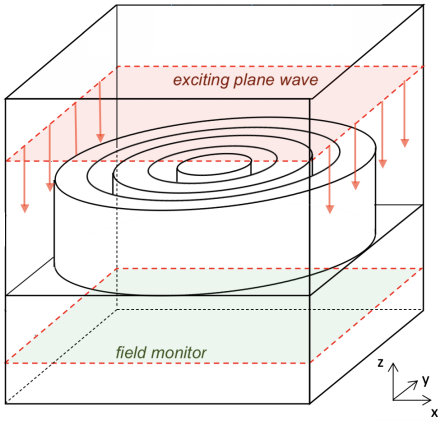

Figure 4.5: FDTD simulation cell setup as implemented in MEEP for the simulation of microstructured vortex phase masks. Here, the center of the AGPM is simulated. The simulation cell contains an exciting source (red) before the AGPM on the air side, and a field monitor (green) inside the substrate to record the resulting fields. The simulation cell is surrounded on all sides by perfectly matched layers (PMLs) not shown in this sketch.

4.1.4 Benchmarking FDTD against RCWA simulations

The described MEEP simulation framework is validated against RCWA simulations in a simple case of a straight grating, for which RCWA is known to be valid, by considering a finite patch of straight grating lines in MEEP. Figure 4.6 shows the grating design for the chosen patch size of $33\,\Lambda \times 33\,\Lambda$ (Λ is the grating period as defined in Figure 2.11), and the leakage distribution after the grating for an illuminating wavelength $\lambda = 3.5\,\mu\text{m}$. The leakage is the part of the light which is transmitted unaffected through the grating, as described later in Chapter 5. The leakage is measured directly after the grating at the focal plane, using an incident circularly polarized plane wave, which is a good description of the center of the Airy disk at the focal plane. The beam is propagated only just before and after the grating, which corresponds to the inset in Figure 5.2, where the handedness of the input polarization is flipped by the half-wave plate character of the subwavelength grating. The outgoing circular polarizations represent respectively the leakage (with same handedness as the incident wave) and the converted beam (with opposite handedness), which are both normalized by the total transmitted flux. Because the AGPM grating line orientation changes with azimuth, it must be ensured that all grating orientations are sufficiently well described with MEEP. Here, a grating oriented at 45° from the cartesian cell axes is used because, as illustrated below, the convergence of the leakage in MEEP simulations is slower for diagonal grating lines than for grating lines parallel to the cartesian axes. Due to the non-zero side wall angle described earlier in Figure 2.11, it must be noted that the filling factor F varies along the optical axis (z-axis). Here and in all following figures, the filling factor at half grating height is shown, because it is more representative of

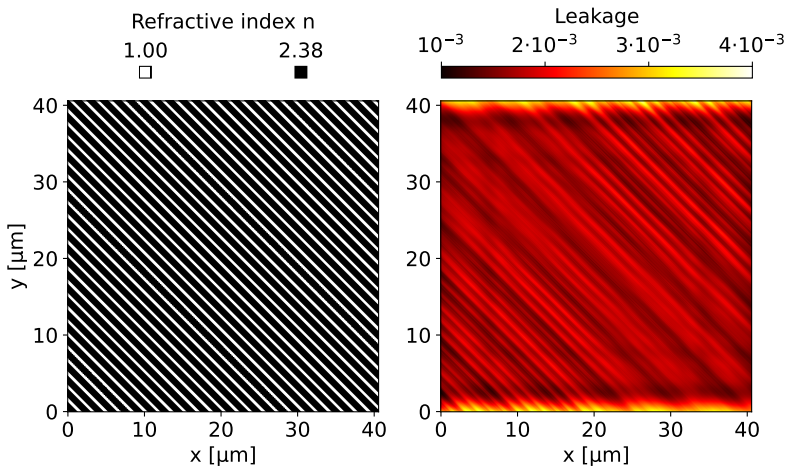

Figure 4.6: Grating design (left) and leakage map (right) simulated with FDTD for a finite patch of straight grating lines at $\lambda = 3.5\,\mu\text{m}$. The leakage is normalized by the total transmitted flux. Optimal grating parameters for the L-band ($2.9-4.1\,\mu\text{m}$) are used: $\Lambda = 1.21\,\mu\text{m}, F = 0.5373, h = 5.5343\,\mu\text{m}, \alpha = 2.45°$. The grating is etched into a diamond substrate with refractive index $n \approx 2.38$ in this wavelength range.

the effective material distribution inside the grating than the nominal filling factor F at the top of the grating. The simulations still take into account the trapezoidal profile of the grating.

Since the same grating parameters as for the RCWA simulation are used, similar results for the leakage are expected, provided that a patch of $33\,\Lambda \times 33\,\Lambda$ is sufficiently representative of the infinite straight grating assumed by RCWA. Figure 4.7 shows the spectral leakage measured with MEEP, compared to the RCWA results for an optimally designed grating (optimal in the sense of RCWA). The results for both diagonal and horizontal grating lines are shown for different values of the MEEP resolution, ranging from 14 pxl/µm to 23 pxl/µm. A monochromatic MEEP simulation of this size at a resolution of 23 pxl/µm needs 2700 CPUhours using 170 GBofRAM, while the equivalent RCWA simulation only takes 40 seconds on a recent laptop, which illustrates the high computational load of MEEP. For diagonal grating lines, the results of the MEEP simulations differ by an absolute leakage of the order of 10^{-3} from the RCWA results for a resolution higher than 20 pxl/µm. This discrepancy is acceptable, considering that only a finite patch is simulated instead of an infinite structure, and considering the simulation artifacts visible in Figure 4.6, which could account for deviations in the results. Also, MEEP will be used mainly to simulate the AGPM center, where the leakage values are higher than for the off-axis patch presented in this section: typically the leakage at the center is on the order of 10^{-1}, which is large compared to the simulation accuracy of 10^{-3} for straight grating lines. For horizontal lines, the results converge at 17 pxl/µm, while for diagonal lines the results do not converge even

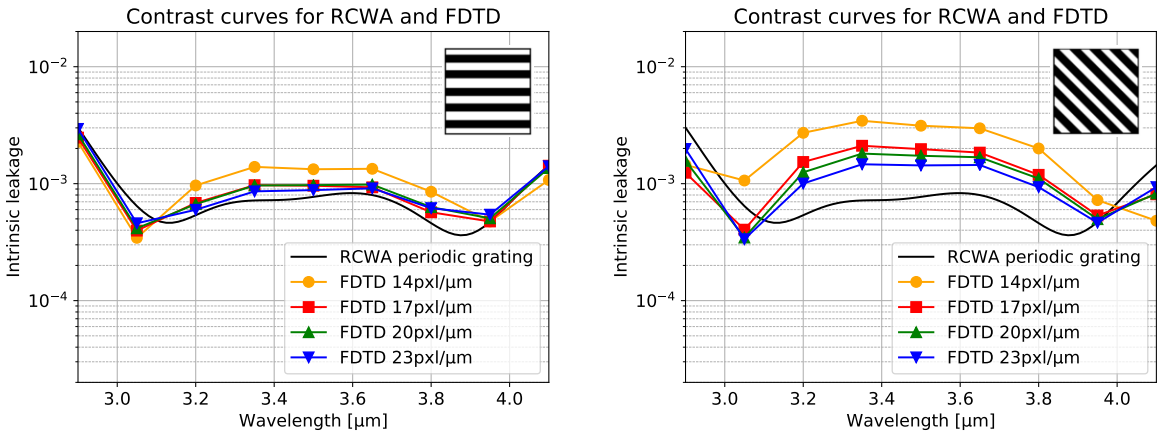

Figure 4.7: Spectral leakage for a $33\,\Lambda \times 33\,\Lambda$ patch of straight grating lines for a given set of grating parameters calculated with RCWA (black) and FDTD (colored). Left: horizontal grating lines, and right: diagonal grating lines. The low number of simulation points is due to the high computational cost of FDTD. Optimal grating parameters for the L-band ($2.9 - 4.1\,\mu\text{m}$) are used: $\Lambda = 1.21\,\mu\text{m}, F = 0.5373, h = 5.5343\,\mu\text{m}, \alpha = 2.45°$.

at 23 pxl/µm, which is the limit of what can be simulated in terms of computational resources available within this project. To keep computational time reasonable, a resolution of 20 pxl/µm is chosen, which is sufficient to reach the needed accuracy. The difference in the matching accuracy for both grating orientations is due to the sub-pixel averaging used in MEEP to define the permittivity function. This feature is crucial to define a material distribution representative of the actual geometry. For horizontal grating lines parallel to the cartesian cell axes, the sub-pixel averaging is done in one dimension (normal to the grating lines) while for diagonal lines the averaging has to be done in two dimensions. Simulations with other grating parameters yield similar results. Despite the non-negligible discrepancy of the results for diagonal grating lines, it is concluded that MEEP at a resolution of 20 pxl/µm is an appropriate tool to simulate a finite patch of the AGPM, and well suited to simulate the center.

4.2 Simulating the coronagraphic setup

Once the response of the coronagraphic phase mask itself is simulated using RCWA and FDTD, its behavior in a full coronagraphic setup has to be evaluated. In a telescope, pupil and focal planes are linked through a Fourier transform, and the propagation of the electric field is therefore best simulated with software packages based on this tool. Two implementations of the Fourier transform are introduced here, which combined with clever handling of numerical errors provide the necessary simulation accuracy to quantify the performance of new vortex phase mask designs.

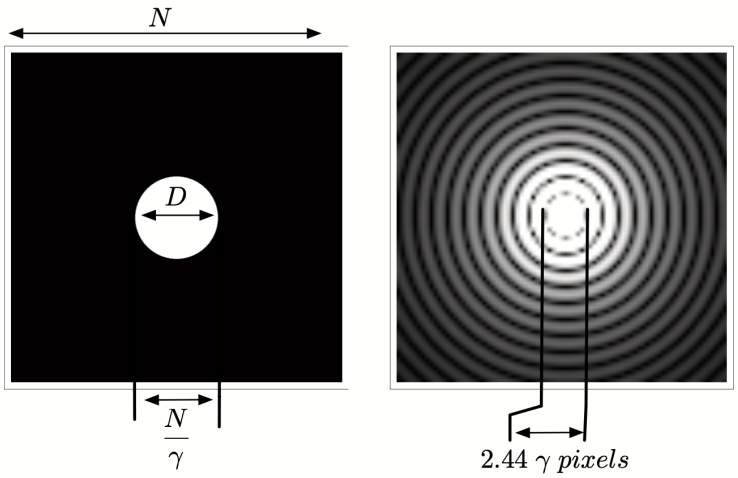

Figure 4.8: Resolution in the focal and pupil plane for FFTs. Left: Pupil plane array of size $N \times N$, pupil diameter D and zero padding factor $\gamma = N/D$ (here $\gamma = 4$). Right: Focal plane image. The resolution in the focal plane is $\text{res}_{\text{FP}} = \gamma \frac{\text{pxl}}{\lambda/D}$. From Soummer et al. (2007).

4.2.1 Fast Fourier transform

An efficient way to perform numerical Fourier transforms is through the fast Fourier transform (FFT) algorithm. The resolutions in the pupil and focal plane are important parameters for these simulations. The parameter defining the focal plane resolution is the zero padding $\gamma = \frac{N}{D}$ of the pupil, where N is the array size and D is the pupil diameter in pixels, i.e. the pupil plane resolution. The resolution in the focal plane then becomes: $\text{res}_{\text{FP}} = \gamma \frac{\text{pxl}}{\lambda/D}$, as illustrated in Figure 4.8. To obtain a good resolution in the focal plane, the zero padding γ has to be sufficiently high. On the other hand, a reasonable sampling of the circular pupil D is mandatory, which leads to huge array sizes N, ultimately limiting the resolution by computational limitations.

The pixelization of the pupil and focal plane creates numerical artifacts that ultimately limit the accuracy of FFTs. One way to reduce the numerical artifacts is to implement the pupil edges as grayscale pixels instead of a binary pupil. In this way, and using the highest array size possible on a typical laptop computer ($N = 8192 \times 8192$), contrasts on the order of 10^{-9} can be simulated.

However, for space-based applications the target contrast is on the order of 10^{-10}, and the numerical residuals have to be further reduced. One possible way to achieve this is to exploit the perfect starlight cancellation of the vortex. Propagating the perfect vortex to the image plane leads to residual light due to numerical artifacts. By artificially removing the residuals in the image plane and propagating the fields back to the entrance pupil, one can obtain a modified entrance pupil providing perfect cancellation by construction. This modified pupil features small variations of the fields compensating for the numerical artifacts arising from the limited resolution of the

vortex singularity. For small errors in the vortex phase ramp, the numerical corrections can be assumed to be still valid and can decrease the numerical background in the image plane.

4.2.2 Matrix Fourier transform

Another way to increase the accuracy of the Fourier transform algorithms is to use varying resolution in the focal plane. The matrix Fourier transform (MFT) can be used to describe the vortex in the focal plane with two regions of different resolutions (Soummer et al., 2007). The resolution of the region around the vortex singularity is higher than the resolution of the rest of the mask where the phase gradient is smaller. Both regions are connected with a Tukey window. The fast linearized coronagraph optimizer (FALCO, Riggs et al., 2018) and hcipy (Por et al., 2018) are codes implementing this method in Matlab and python respectively. Both methods can quantitatively assess vortex phase masks down to contrasts $< 10^{-12}$, which is sufficient for space-based applications.

4.2.3 Dark hole digging and wavefront control

The simulation tools describing the performance of vortex phase masks rely on the propagation of the fields to the final image plane. However, in realistic conditions instrumental aberrations dominate the performance of the vortex phase mask. These wavefront errors can be corrected by closed-loop wavefront control methods implemented with a deformable mirror. Wavefront control techniques like electric field conjugation (EFC, Give'on et al., 2007) can also be used to create a dark region in the final focal plane image (Malbet et al., 1995) by improving the starlight suppression of the vortex. The process of optimizing the shape of the deformable mirror is called dark hole digging and can improve the contrast of a vortex phase mask by several orders of magnitude. In particular, the deformable mirror can be used to correct for the chromatic errors of the vortex phase mask, making dark hole digging with wavefront control algorithms an important tool to achieve very high contrast levels. Simulations of EFC for dark hole digging is used in Section 11.6 to push the contrast of scalar vortex coronagraphs towards the 10^{-10} limit.

5 Metrics

To assess the performance of vortex phase masks it is important to use a consistent metric. However, different metrics are suited for different applications. While the general performance criterion for a vortex phase mask is its ability to suppress an on-axis source, for more specific tasks like the imaging of exoplanets the residual starlight in a certain region in the coronagraphic image is more important. This chapter therefore aims to introduce the different metrics used in the following chapters to characterize the performance of vortex phase masks.

5.1 The helical phase ramp

The concept of the vortex coronagraph relies on the helical phase ramp imprinted on the beam by a focal plane mask. This setup sends the light of an on-axis source outside the pupil, leaving a dark region in the Lyot plane. The perfect vortex phase mask features a linear phase increase ϕ with azimuth θ: $\phi = \ell_p \cdot \theta$, where ℓ_p is the topological charge of the vortex. However, the shape of the actual phase ramp depends on the implementation of the vortex phase mask, and deviations from the perfect phase ramp influence the achievable contrast.

For scalar vortex coronagraphs, the phase ramp is chromatic, which means each wavelength sees a slightly different phase ramp. In the case of a scalar vortex phase mask made of a helical shaped dielectric, adding a factor λ/λ_0 to the phase ramp is sufficient to describe the chromaticity, which results in a phase ramp of non-integer topological charge $\phi = \ell_p \cdot \theta \cdot \lambda/\lambda_0 = \ell_{p,\text{eff}} \cdot \theta$. Since the beam now acquires phase ramps of different non-even topological charge, the coronagraphic performance in the final coronagraphic image is reduced (since only even topological charges provide perfect starlight cancellation, as described in Section 2.2).

The use of a metasurface framework as discussed in Chapter 11 introduces more complex phase errors in the phase ramp. Similarly to the helical shaped dielectric masks, the phase ramp of a nanopost phase mask is perfect for the central wavelength. However, for other wavelengths, the phase response of the nanopost design is not just a scaled version of the phase ramp, but introduces deviations from the linear azimuthal profile, as shown in Figure 5.1.

For vector vortex phase masks, the phase ramp is based on the geometric phase which is inherently achromatic. In the case of the AGPM using a circular subwavelength

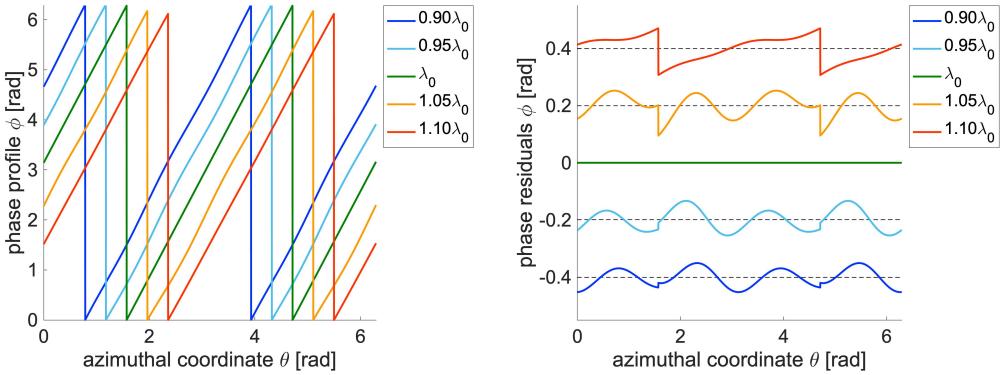

Figure 5.1: Azimuthal phase profile showing slight deviations from the linear profile (left) and corresponding residuals (right) for a square nanopost implementation of a scalar vortex phase mask for the design discussed in Chapter 11. Five different wavelengths in a bandwidth of 20 % are shown. A constant offset has been applied to the different curves for clarity.

grating, the resulting phase ramp is perfect wherever the geometric phase is defined by the grating lines orientation. This is not the case at the center of the AGPM where the increased grating curvature becomes the limiting factor, resulting in an imperfect phase ramp.

The use of other technologies exploiting the geometric phase can additionally influence the phase ramp. For liquid crystal polymers, the orientation of the single polymers determines the fast axis, and therefore errors in the alignment translate into errors of the phase ramp. At the center of the mask, this results in a region in which the phase ramp is poorly defined. When using photonic crystals, the geometric phase depends on the geometry in a similar way as for the AGPM, and therefore errors due to the manufacturing and due to the central singularity have similar effects on the phase ramp.

However, for the vector vortex case the main limitation is not the imperfect phase ramp, but the polarization leakage arising from the chromatic imperfection of the half-wave plate character of the mask.

5.2 Polarization leakage

For vector vortex coronagraphs, the phase ramp is implemented using the geometric phase, which is inherently achromatic. The main factor limiting the performance is the imperfect half-wave plate character across the design bandwidth, which leads to a part of the light not acquiring the phase shift. Considering an incoming circularly polarized plane wave, most of the incoming light experiences the vortex effect and acquires the phase ramp while being converted to the opposite helicity. A small part

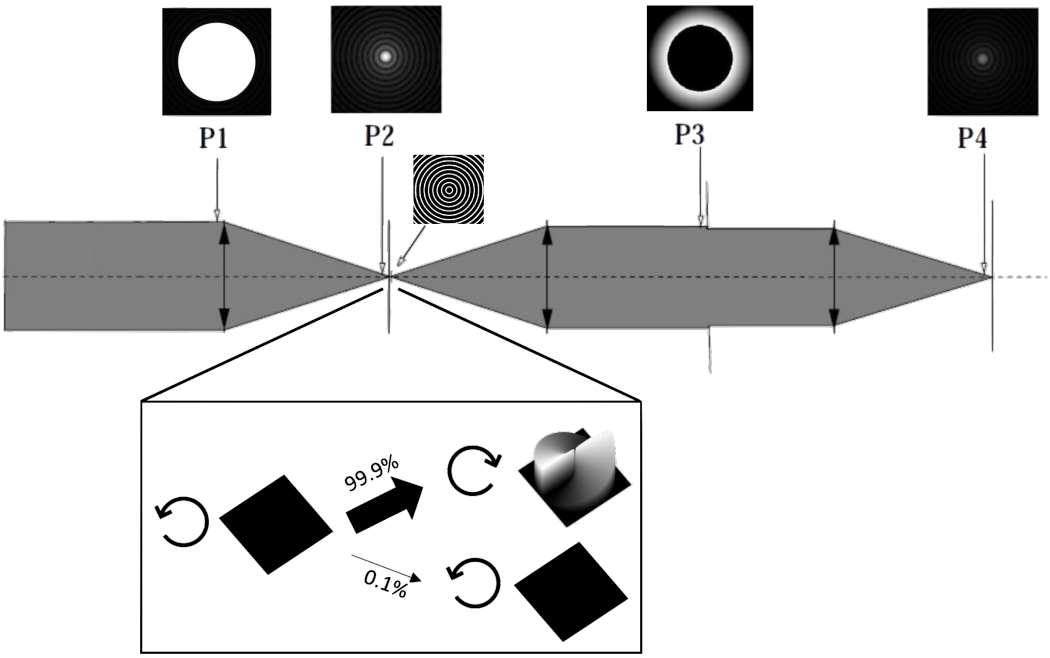

Figure 5.2: Configuration of the vector vortex coronagraph along the optical axis. The distribution of the intensity is shown in the four planes P1-P4 (top). The effect of the vector vortex phase mask symbolized as AGPM is illustrated as inset. An incoming circular polarization acquires the phase ramp, while a small fraction (typically on the order of 10^{-3}) remains unaffected, referred to as polarization leakage. P1: Telescope pupil, P2: focal plane PSF, P3: relayed pupil plane (Lyot plane) before the undersized diaphragm (Lyot stop), P4: image plane with attenuated PSF.

of the light is however left unaffected due to the chromaticity of the half-wave plate and does not acquire the phase ramp, maintaining its polarization state, as illustrated in Figure 5.2. This part of the light that does not acquire the phase ramp is referred to as polarization leakage. Since the polarization leakage is transmitted unaffected through the phase mask, it propagates to the coronagraphic image where it creates a scaled-down version of the PSF. The contrast in the coronagraphic image is in this case defined by the polarization leakage. This means, that the polarization leakage can be used as the relevant metric for the vector vortex coronagraph without the need to propagate the leakage to the coronagraphic image.

5.3 Null depth

To understand the polarization leakage in case of the AGPM, consider a plane wave incident on a subwavelength grating. As described in Mawet et al. (2005a), the vortex effect relies on the destructive interference of different parts of the phase mask.

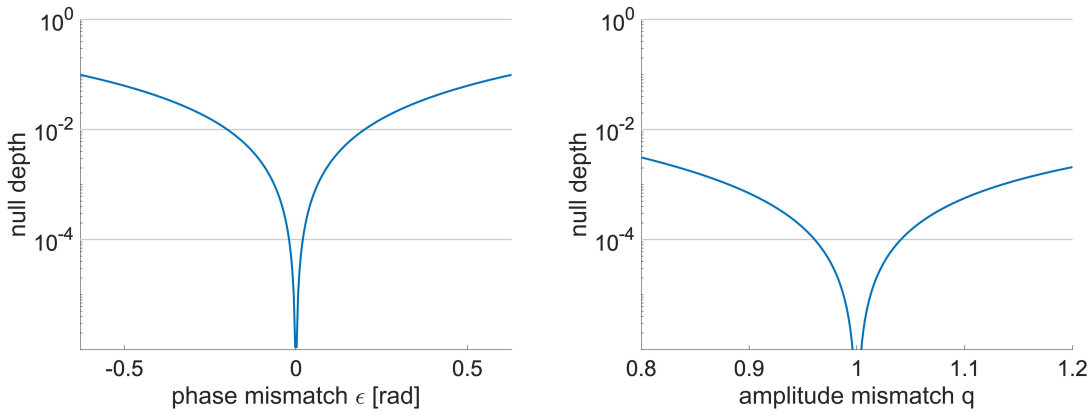

Figure 5.3: Effect of phase (left) and amplitude mismatch (right) on the null depth. The null depth is plotted for a variation of ±20 % of the phase and amplitude mismatch.

Notably, regions with grating lines orthogonal to each other interfere destructively. This means, that the transverse electric (TE) and transverse magnetic (TM) modes of the grating (with their E-field vector parallel and orthogonal to the grating lines) have to interfere destructively with each other, which is guaranteed for a π phase shift between them. However, the transmission of both modes is different due to the grating geometry. Additionally, the phase shift is not exactly π across the band. Both factors influence the leakage term, which becomes:

$$N(\lambda) = \frac{[1 - \sqrt{q(\lambda)}]^2 + \epsilon(\lambda)^2 \sqrt{q(\lambda)}}{[1 + \sqrt{q(\lambda)}]^2} \tag{5.1}$$

where q is the amplitude mismatch (the ratio of TE and TM transmitted amplitude) and $\epsilon = \pi - \Delta\Phi$ the phase mismatch, both dependent on the wavelength. The quantity N is referred to as null depth, which translates into the attenuation of the stellar PSF in the coronagraphic image.

Equation 5.1 can be visualized to illustrate the influence of both mismatch parameters on the null depth. This is shown in Figure 5.3, showing that the phase mismatch has a more important effect on the null depth than the amplitude mismatch.

Two equivalent metrics are used in the literature to assess the performance of the AGPM, as described in König et al. (2022): the null depth and its inverse, the rejection ratio. For consistency, the null depth as described in Section 5.3 will be used in this thesis. However, two terms related to the null depth will be distinguished for clarity: the null depth will refer to the global performance of an AGPM, and the intrinsic leakage of the grating will refer to the fractional amount of light leaking through a periodic infinite straight grating calculated with RCWA. This means, that the null depth also captures the effect of the AGPM center, where the intrinsic leakage can not be used as valid performance metric, as described in Chapter 7.

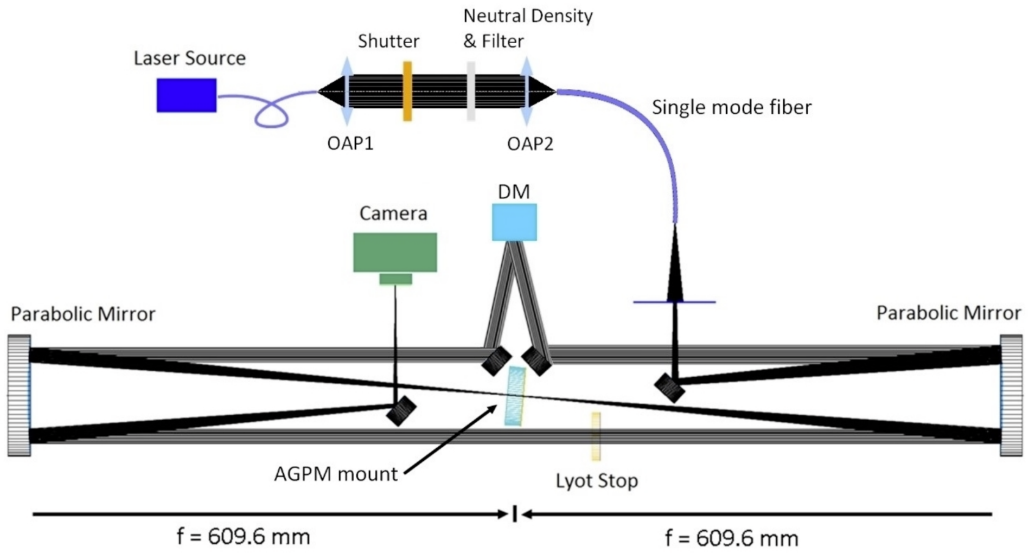

Figure 5.4: Schematics of the VODCA test bench in the configuration used in this thesis.

5.4 Performance metric on VODCA

VODCA (Vortex Optical Demonstrator for Coronagraphic Applications) is the coronagraphic test bench in our lab used to measure the performance of AGPMs. VODCA consists of a setup described in detail in Jolivet et al. (2019), Jolivet (2019). The setup allows for the performance measurement of AGPMs in the infrared from H to M-band ($1.44 - 5.0\,\mu m$) using a supercontinuum laser operating up to $4.1\,\mu m$ (H to L-band) combined with broadband and narrowband filters, and two quantum cascade lasers covering the M-band (4.5 and $5.0\,\mu m$). A deformable mirror is also available to correct for static aberrations of the setup. Figure 5.4 shows a schematic of VODCA as used within this thesis. However other applications are currently envisaged, like focal plane wavefront sensing (Quesnel et al., 2022).

In principle the null depth can be measured by dividing the flux in the coronagraphic image by the flux of a reference image without vortex. In practice, the reference image is obtained by laterally displacing the phase mask in the focal plane, so that the PSF center is sufficiently far from the phase singularity to not experience the vortex effect. The null depth is then calculated as the ratio of the flux of the coronagraphic and the reference PSF in a certain region of interest. Typically, for the null depth measurement on our coronagraphic test bench VODCA, the full width at half maximum (FWHM) of the Airy disk is used (circular region of radius $0.514\,\lambda/D$), as described in Jolivet et al. (2019). This method guarantees a robust metric in the case of a scaled down version of the PSF. In some cases it is however still useful to consider the radial profile of the coronagraphic PSF. In case of aberrations on the bench, the PSF shape can be

altered considerably, questioning the validity of the FWHM as representative region to calculate the null depth. For example, the PSF shows a donut-like shape in the presence of astigmatism, which artificially reduces the null depth within the FWHM. Therefore, a larger region of up to several λ/D can be used to calculate the null depth, notably in the case of more spread out PSFs (Jolivet, 2019).

5.5 Scoring regions when using wavefront control

To assess the general performance of a vortex phase mask, the null depth is an appropriate metric describing the attenuation of the stellar PSF. However, in high-contrast imaging typically the capability of detecting a companion at a certain separation is of interest. It is therefore useful to consider the contrast defined as the ratio of the flux at a certain separation from the center (for example the inner working angle) in the coronagraphic image and the peak of the non-coronagraphic reference PSF.

This metric is of particular use when applying wavefront control algorithms to improve the contrast in a scoring region. The contrast values typically used when performing wavefront control refer to the flux in a region beyond the inner working angle of the coronagraph normalized by the peak of the non-coronagraphic PSF. For a charge-6 vortex coronagraph as considered in Chapter 11 the region of interest is beyond the inner working angle of 2.5 λ/D, and in practice typically a region between 3 and 10 λ/D is used to compute the contrast. In Chapter 11, a region from 3 to 10 λ/D on one half of the final focal plane image is used to dig a dark hole using an electric field conjugation algorithm.

Figure 5.5 illustrates how the contrast requirement in the scoring region is related to the null depth, provided that the coronagraphic PSF is a scaled down version of the off-axis PSF. The 10^{-10} contrast requirement motivated in Section 1.3.3 corresponds to a null depth of 5×10^{-7}. However, using wavefront control algorithms, 10^{-10} contrast from 3 to 10 λ/D could be achieved without the need for such a null depth, since wavefront control algorithms have been used to successfully dig and maintain a dark hole gaining up to two orders of magnitude in contrast. This can potentially relax the null depth requirements for a vortex phase mask while remaining compatible with the goals of future space telescope missions.

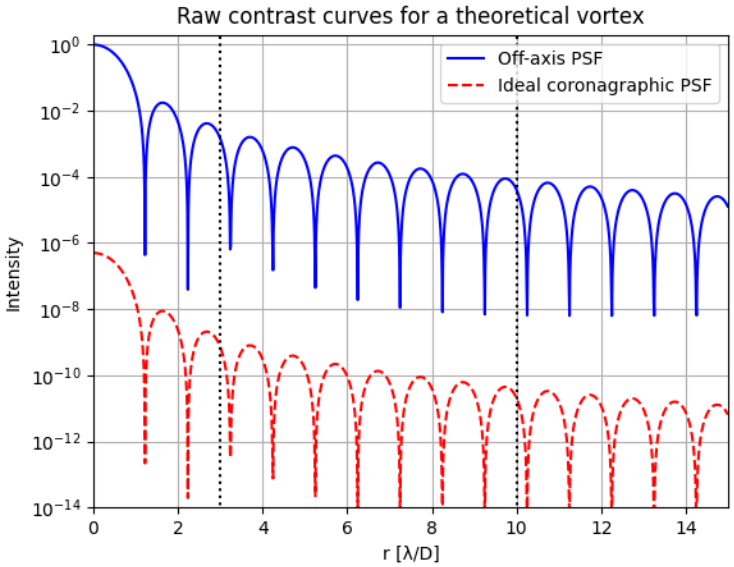

Figure 5.5: Theoretical contrast curve of an ideal vortex coronagraph for future space telescope missions illustrating the connection of the null depth and the contrast in an annular region between 3 and 10 λ/D. The mean contrast of the ideal coronagraphic PSF in this region is 10^{-10} for a null depth of 5×10^{-7}.

Part III

Subwavelength gratings

6 AGPM development for METIS

The METIS high-contrast imaging modes rely on a series of phase and amplitude masks in the focal- and pupil planes. In particular, the vector vortex coronagraph modes feature AGPMs in the focal plane covering the wavelength range from 3 to 13 µm. The high-contrast imaging goals of METIS impose certain requirements on the performance of the vortex coronagraph, that translate into requirements on the AGPMs themselves. Notably, the AGPMs should not limit the achievable contrast after postprocessing when considering all instrumental errors, which is verified using end-to-end simulations (Delacroix et al., 2022). The resulting requirements are defined in the METIS final design report documentation. An average null depth below 2×10^{-3} is required at L and M-band, with a peak null depth below 1×10^{-2} at each individual wavelength. At N-band, the requirements are relaxed, as explained later, and an average null depth below 5×10^{-2} is required, with a peak null depth below 1×10^{-2} at each individual wavelength.

Another effect to be considered is the bright star limit. This requirement is defined as the capability of observing faint sources around α Cen without using additional neutral density filters and translates into a contrast requirement of 10^{-3} at L and M-band, if a saturation by a factor 10 of the detector is tolerated. This requirement is relaxed at N-band, making the stellar leakage described above the limiting factor of the AGPM performance. However, the achievable N-band contrast is also limited by the dominant effect of water vapor seeing on the error budget, as explained in the following sections.

The AGPMs for METIS are made of polycrystalline diamond (as previous AGPMs) and have a diameter of 20 mm providing a sufficiently large field of view. The AGPMs for the N-band are smaller (15 mm) because three of them have to be placed next to each other in the focal plane to implement chopping.

It is planned to publish the results presented in this chapter as soon as the L and M-band AGPMs are tested end of 2023.

6.1 Design of AGPMs for METIS

The AGPMs for the METIS bands are optimized with the simulation framework described in Section 4.1. Notably, the grating parameters of the AGPMs for L ($2.9 - 4.1$ µm), M ($3.9 - 5.3$ µm), and N-band ($8.1 - 13.1$ µm) are optimized using RCWA, which is a valid tool to simulate the performance of AGPMs at the contrast levels

required for METIS. The design of the anti-reflective gratings etched on the back of the diamond substrates will also be optimized with RCWA.

6.1.1 Anti-reflective gratings

The AGPMs feature an anti-reflective grating etched on the backside of the diamond substrate to minimize unwanted reflections at the diamond-air interface. A flat interface would result in the reflection of a significant part of the light described by the Fresnel reflection coefficient $R = (\frac{n_1-n_2}{n_1+n_2})^2$. For a diamond-air interface in the infrared ($n = 2.38$) this means, that $R = 17\,\%$ of the light is reflected at the backside of the AGPM. This reflection reduces the throughput, but also introduces a ghost signal caused by the reflected light interacting with the AGPM grating, where part of the light is reflected again. The ghost signal is particularly problematic because it dominates the residual light transmitted by the AGPM. To mitigate this effect, an anti-reflective crossed grating (Forsberg & Karlsson, 2013b) is etched on the backside of the AGPM (as illustrated in Figure 2.10), providing transmission values close to 100 %.

The anti-reflective grating has to be optimized for a certain bandwidth, which is done separately for each band using RETICOLO, as described in Section 4.1. With these simulations, the grating parameters are optimized for highest average transmission across the band. Figure 6.1 shows the result of the optimization for L, M and N-band, showing transmission above 98 % for all parts of the L and M-band, and above 96 % for N-band. The average transmission across the bands is higher (above 99 % for L and M-band, and above 98 % for N-band).

6.1.2 RCWA design optimization

METIS features two AGPMs optimized for L and M band. The grating parameters were optimized with RCWA as described in Section 4.1. Figure 6.2 (top) shows the null depth obtained for the optimal design of an L band AGPM assuming a sidewall angle of 2.45°. The null depth is below 10^{-3} across the band, except for the two ends of the wavelength range. The null depth as described in Section 5.3, as well as the influence of the ghost signal described in the previous section are shown. The performance decrease at the upper end of the band can be tolerated, since this region of the band is also covered by the M band (which provides slightly better performance due to the smaller bandwidth). The performance decrease at the lower end of the band can be tolerated since it is of lower scientific interest.

The tolerances on the optimal parameters are illustrated in Figure 6.2 (bottom). The mean null depth is shown for a region of the parameter space spanned by the grating parameters around their optimal values. Each of the three parameters is fixed in one of the maps, while the remaining two parameters vary around the optimum, which illustrates the effect of imperfect grating parameters on the null depth. In particular, the tolerances on the grating parameters are small when targeting a null depth below 10^{-3},

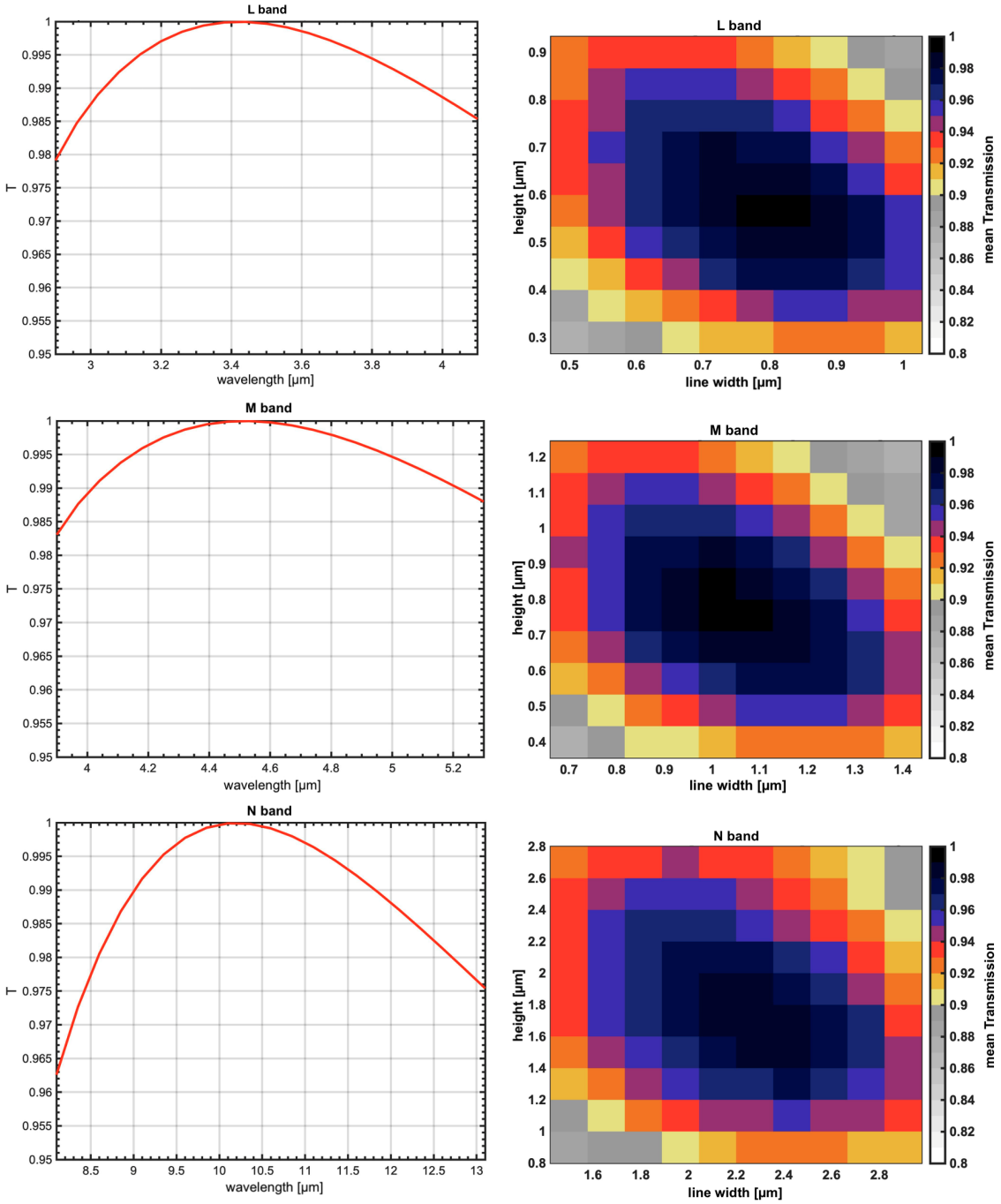

Figure 6.1: Transmission of the anti-reflective gratings for the METIS AGPMs optimized with RCWA. The spectral transmission (left) is shown for the optimal grating parameters, along with the mean transmission for a region of the parameter space spanned by the height and the line width of the crossed grating (right) for L (top), M (middle) and N-band (bottom).

as required for METIS. However, the plots also show that small deviations in one parameter can be compensated by tuning the remaining grating parameters. A deviation from the optimal sidewall angle (which is the parameter with the largest manufacturing uncertainty) can be compensated by tuning the grating depth accordingly, which is typically done when manufacturing an AGPM.

For the M-band, the optimization leads to the results shown in Figure 6.3. Note the slightly improved null depth due to the smaller bandwidth of the M-band (30 % compared to 34 % for the L-band). This not only leads to a better overall performance of the design, but also relaxes the tolerances of the manufacturing process, as can be seen by the larger dark regions in the mean null depth maps. In addition, the grating parameters scale with the wavelength, relaxing the tolerances on the depth and the line width even more.

The N-band for METIS spans a 47 % bandwidth, which is too much to guarantee the required contrast with a single AGPM. Therefore, the N-band was initially divided into two subbands *N1* (8.0-10.5 µm) and *N2* (10.0-13.5 µm), which would be covered each with its own AGPM optimized for the narrower subbands. The optimized designs for these subbands provide a null depth below 10^{-3}, as required by the original requirements.

However, the performance of METIS in the N-band will be limited by water vapor seeing (Absil et al., 2022). Since this effect is chromatic, the adaptive optics corrections at K-band do not fully account for the water vapor seeing at longer wavelengths. This effect is stronger for longer wavelengths and dominates the wavefront error budget at N-band. Therefore, the METIS requirements for the N-band AGPMs were relaxed and now require a mean null depth below 10^{-2}, with a peak null depth below 5×10^{-2} at all individual wavelengths within the band.

Following the updated requirement, an optimal grating design was obtained covering the full N-band. The results of this optimization are shown in Figure 6.4 along with the null depth curves for the two subbands. The design covering the full N-band with one single AGPM satisfies the updated null depth requirements and can therefore be used as cost-efficient alternative to the two separate AGPMs proposed initially. This is even more important since the N-band AGPMs are used for chopping and therefore three AGPMs are needed for each subband, reducing the number of N-band AGPMs from 6 to 3 with this super broadband solution.

A summary of the optimization results for all the METIS AGPMs is presented in Table 6.1.

6.2 Manufacturing of N-band AGPMs

The manufacturing of the METIS AGPMs was started with the N-band, because their testing is expected to take more time than for the L and M-band due to the need for a

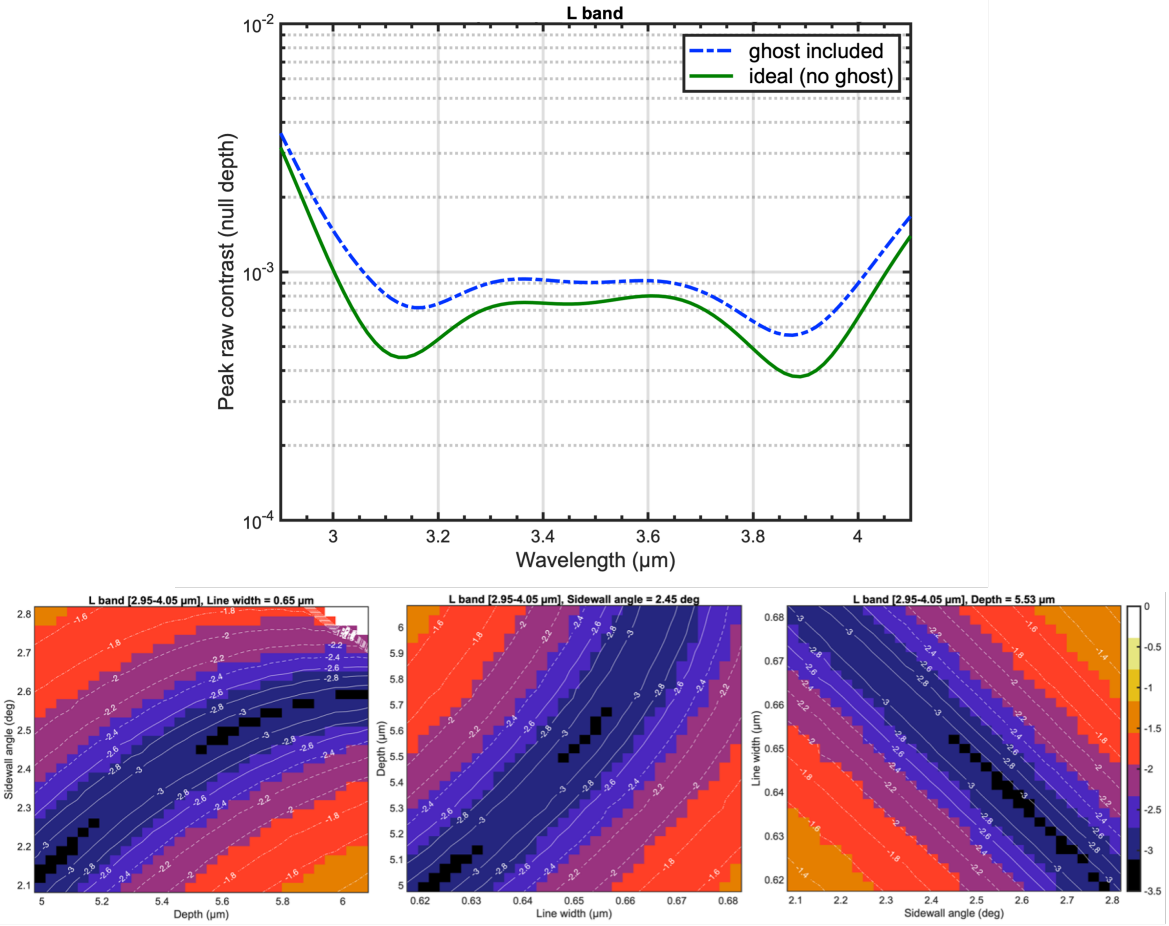

Figure 6.2: Null depth of the AGPM optimized for the METIS L-band. Top: theoretical null depth without internal reflections (solid green line) and realistic null depth including the faint ghost signal (almost canceled by the anti-reflective grating, dashed blue line). The optimal grating parameters for the L-band are: $\Lambda = 1.21\,\mu\text{m}$, $F\Lambda = 0.65\,\mu\text{m}$, $h = 5.53\,\mu\text{m}$, $\alpha = 2.45°$. Bottom: Mean null depth for a region of the parameter space around the optimal values of the three adjustable grating parameters shown as color coded maps. The three maps show the null depth for fixed line width (left), sidewall angle (middle) and groove depth (right), while varying the remaining two parameters around their optimal values. The white region can not be accessed due to the merging sidewalls.

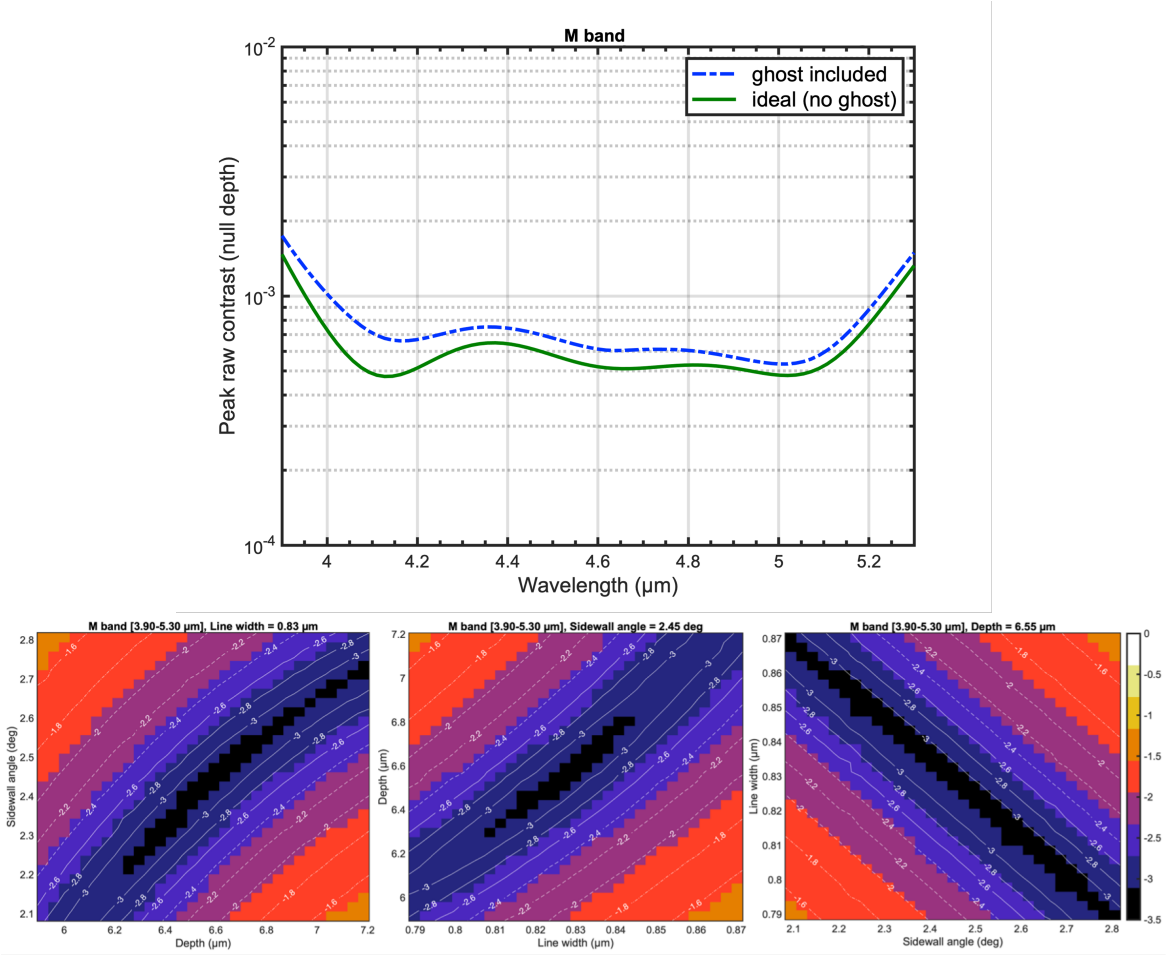

Figure 6.3: Null depth of the AGPM optimized for the METIS M-band. Top: theoretical null depth without internal reflections (solid green line) and realistic null depth including the faint ghost signal (almost canceled by the anti-reflective grating, dashed blue line). The optimal grating parameters for the M-band are: $\Lambda = 1.63\,\mu\text{m}$, $F\Lambda = 0.83\,\mu\text{m}$, $h = 6.55\,\mu\text{m}$, $\alpha = 2.45°$. Bottom: Mean null depth for a region of the parameter space around the optimal values of the three adjustable grating parameters shown as color coded maps. The three maps show the null depth for fixed line width (left), sidewall angle (middle) and groove depth (right), while varying the remaining two parameters around their optimal values.

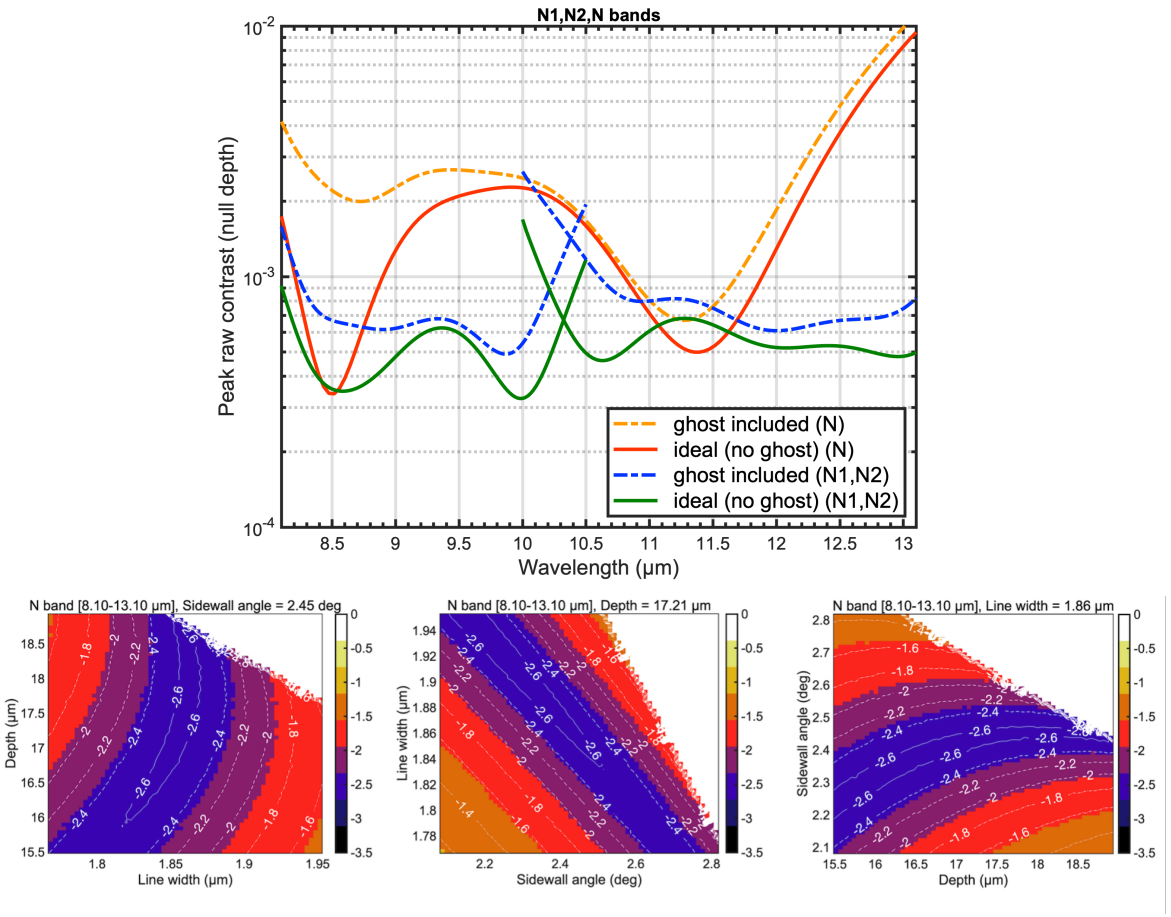

Figure 6.4: Null depth of the AGPMs optimized with RCWA for the METIS N-band. Top: theoretical null depth without internal reflections (solid red line) and realistic null depth including the faint ghost signal (almost canceled by the anti-reflective grating, dashed orange line). The graph also shows the null depth curves obtained for the initially proposed *N1* and *N2*-subbands (green and blue). The optimal grating parameters for the N-band are: $\Lambda = 3.40\,\mu\text{m}$, $F\Lambda = 1.86\,\mu\text{m}$, $h = 17.21\,\mu\text{m}$, $\alpha = 2.45°$. Bottom: Mean null depth for a region of the parameter space around the optimal values of the three adjustable grating parameters shown as color coded maps. The three maps show the null depth for fixed sidewall angle (left), groove depth (middle) and line width (right), while varying the remaining two parameters around their optimal values. The white region can not be accessed due to the merging sidewalls.

band	Λ	$F\Lambda$	h
L $(2.9 - 4.1\,\mu m)$	$1.21\,\mu m$	$0.65\,\mu m$	$5.53\,\mu m$
M $(3.9 - 5.3\,\mu m)$	$1.63\,\mu m$	$0.83\,\mu m$	$6.55\,\mu m$
N $(8.1 - 13.1\,\mu m)$	$3.40\,\mu m$	$1.86\,\mu m$	$17.21\,\mu m$

Table 6.1: Optimal design parameters of the METIS AGPMs. All designs assume a sidewall angle $\alpha = 2.45°$.

cryogenic test bench. Three AGPMs for the N-band have to be provided to METIS, as they will be used together in a single mount, thereby enabling chopping. As of the writing of this document, three AGPMs have been manufactured, and their performance is currently assessed. Here I describe my contribution to the updated manufacturing process, and the subsequent manufacturing performed by our collaborators in Uppsala.

6.2.1 Updated manufacturing process using a single thick silicon mask

Previous AGPMs have been manufactured following the procedure described in Section 2.2.3. This process uses a stack of sputtered metal layers of decreasing thickness on top of the diamond substrate, using each layer as mask for the next thicker layer. However, this results in many etching steps, during each of which the pattern might be altered when transferred from layer to layer. Therefore, an updated manufacturing process has been investigated using a single thick silicon mask between the diamond substrate and the photoresist mask, enabled by the high selectivity of the recipe used to etch diamond.

In practice, the manufacturing process for the METIS N-band AGPMs can be summarized as follows. The diamond substrate is sputtered with a thick layer of silicon ($\sim 2\,\mu m$) and spin coated with a thin photoresist layer. The pattern is then transferred to the photoresist with the SAMIM technique described in Section 2.2.3. The etching of the stack now consists of only two steps, first etching the silicon layer using the photoresist as mask, second etching the diamond using the silicon layer as mask.

The first etching step consists of etching a thick silicon layer ($\sim 2\,\mu m$) using a thin photoresist layer ($\sim 200\,nm$) as mask. This is somehow challenging with the process used so far because the photoresist layer is worn down before reaching the required etch depth. Therefore, a different method based on a Bosch process is used to etch the silicon layer (Lärmer & Schilp, 1996; Rickard & McNie, 2001). This process consists of two steps. First, the silicon is etched isotropically with a SF_6 plasma. Second, C_8F_4 gas is used as polymerization reactant to deposit a protective polymer layer on the etched surfaces. In the following etching step the polymer layer at the bottom of the grooves is quickly removed by ion bombardment, exposing the silicon to the etchant. The protective layer on the sidewalls is removed slower because of the anisotropy induced by the applied bias during etching, which accelerates the ions towards the bottom

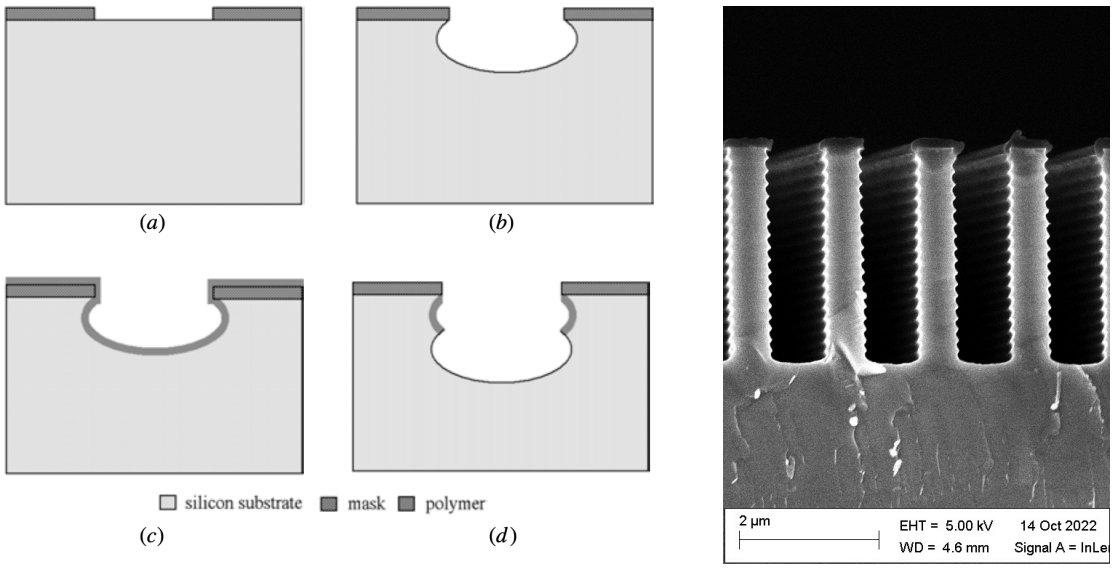

Figure 6.5: Illustration of the Bosch process used for etching of the thick silicon mask. Left: illustration of the steps of the Bosch process. The patterned silicon substrate (a), the etching step with a SF_6 plasma (b), the polymerized protection layer (c), and the next etching step (d) are shown. From Zhou et al. (2004). Right: cross section (SEM-image) of a silicon grating etched with the Bosch process, used when tuning the parameters of the process in order to obtain vertical sidewalls.

of the grooves. Repeating these two steps leads to the characteristic stepped profile shown in Figure 6.5.

The manufacturing process described here has been influenced by technical issues encountered with the inductively coupled plasma (ICP) etcher used so far (PlasmaTherm). Therefore, another plasma etcher was tested (Tegal), which turned out to provide steeper sidewalls than the ICP etcher. This means, that higher aspect ratios can be achieved using this etcher, increasing the range of possible grating designs.

6.2.2 VPM-MET-N4

Using this updated manufacturing process, a first *N*-band AGPM was manufactured in February 2023. The AGPM is referred to as VPM-MET-N4, since it is etched on substrate number 4 out of the 5 substrates that were ordered initially. However, most of the substrates showed an unusually high number of defects, and therefore only substrate number 4 was used to manufacture an AGPM. Further substrates have been acquired later, notably used to etch VPM-MET-N8 and VPM-MET-N9 in June 2023.

Figure 6.6 shows VPM-MET-N4 at different stages of the manufacturing process, as described above. The mask shows a well pronounced grating pattern, with parameters

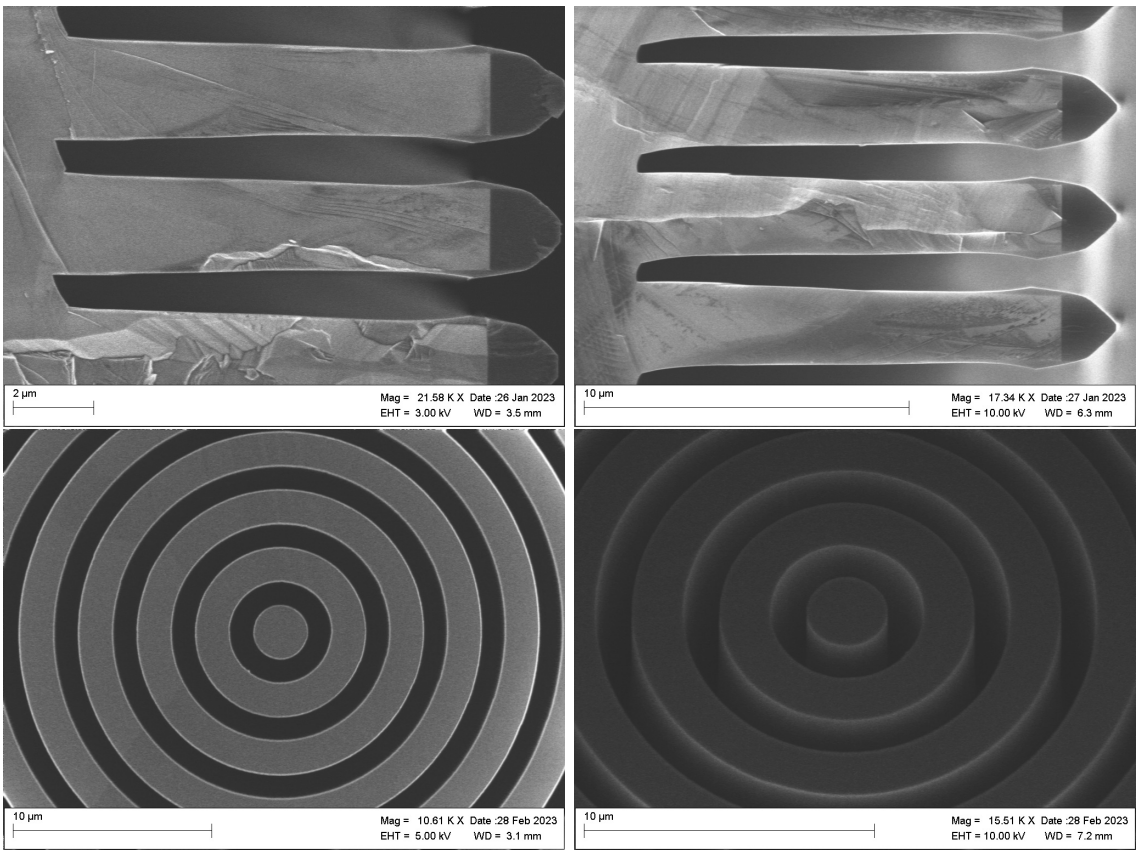

Figure 6.6: VPM-MET-N4 at different stages of the manufacturing process. Top: grating profile of the twin sample after 30 min (left) and 37.2 min (right) of etching. Bottom: top (left) and slanted (right) view of the AGPM center.

that are close to the optimal values. Its performance has therefore been measured on a dedicated coronagraphic test bench (see next section) and its parameters are summarized in Table 6.2. The two other AGPMs (VPM-MET-N8 and VPM-MET-N9) that have been manufactured more recently for the N-band also show promising results so far, and their performance is currently assessed.

6.3 Inverse retrieval of grating parameters

The performance of VPM-MET-N4 has been tested on the dedicated coronagraphic test bench at CEA Saclay (Ronayette et al., 2020). Its setup has been optimized to measure the performance of the N-band AGPMs for METIS. The METIS N-band is covered by 5 monochromatic sources. The null depth of VPM-MET-N4 has been measured in a first test campaign earlier this year and the results are shown in Table 6.3.

The null depth measurements presented here can be used to retrieve the actual grating parameters of VPM-MET-N4 by fitting a grating model to the data points. The same

	Λ	$F\Lambda$	h	α
VPM-MET-N4 (nominal)	3.40 µm	1.95 – 2.00 µm	13.9 µm	1.4°
VPM-MET-N4 (retrieved)	3.40 µm	1.97 µm	12.79 µm	1.4°

Table 6.2: Nominal and retrieved grating parameters of VPM-MET-N4. The sidewall angle is considerably smaller than described earlier because the Tegal etcher was used. The line width and height of the grating have been re-optimized to account for this sidewall angle, resulting in a smaller optimal height.

band	null depth
8.0 µm	2.0×10^{-2}
9.0 µm	3.6×10^{-3}
10.5 µm	4.5×10^{-3}
11.5 µm	1.4×10^{-2}
12.5 µm	5.0×10^{-2}

Table 6.3: Measured null depth of VPM-MET-N4 on the dedicated coronagraphic test bench at CEA Saclay.

RCWA optimization code as used for the optimization of the grating parameters can be used for this purpose, while changing the metric of performance (using the model mismatch as metric).

Using this method, a tentative grating design has been retrieved, as shown in Figure 6.7. The design features a considerably shallower groove depth than measured with SEM images on the twin pattern. The twin pattern is a smaller substrate etched at the same time as the AGPM featuring the same pattern. It is generally cracked after etching to assess the grating profile, since it is difficult to measure the grating profile with non-destructive methods. Since both patterns are etched simultaneously, the grating parameters should be identical. However, the AGPM is positioned at the center of the etcher to guarantee the most uniform etching possible, while the twin pattern is positioned next to it. Therefore, there might be a slight difference in the etch rate resulting in slightly different grating parameters. Since the difference between the two samples is expected to be small, it is unlikely that the actual groove depth of VPM-MET-N4 is below 13.4 µm. Further measurements are therefore planned to confirm these results, which could lead to updated retrieved grating parameters. If these results are confirmed, re-etching as described in Vargas Catalán et al. (2016b) will be considered in order to increase the performance.

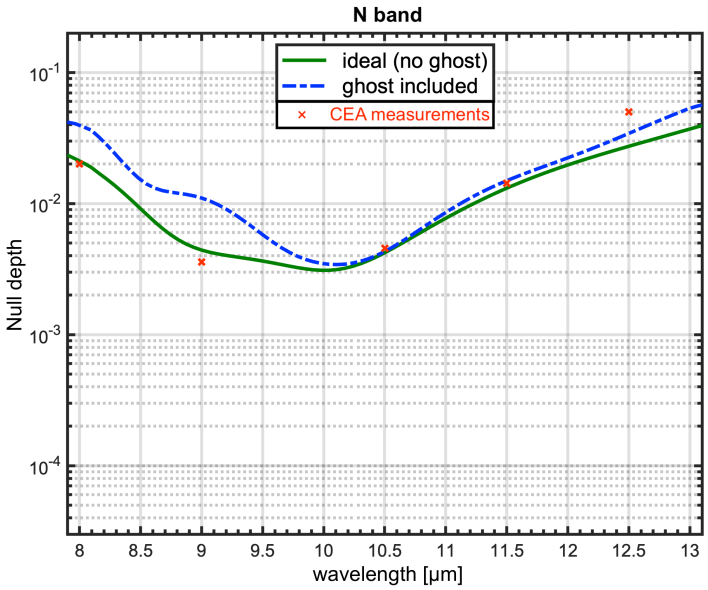

Figure 6.7: Retrieval of the grating parameters of VPM-MET-N4 from the null depth measurements (red crosses). The null depth for the most probable grating parameters is shown (green and blue lines, without and with ghost, respectively). Grating parameters of the model which best fits the data: $\Lambda = 3.40\,\mu\text{m}$, $F\Lambda = 1.97\,\mu\text{m}$, $h = 12.79\,\mu\text{m}$, $\alpha = 1.4°$.

7 Optimization of the AGPM central region

The null depths of the AGPMs developed so far for ground based applications are on the order of 10^{-3}, reaching 3×10^{-4} for the best performing AGPM (Jolivet et al., 2019). At low contrasts, the performance of the AGPM is predicted accurately with RCWA. However, at higher performance levels, the validity of the RCWA simulations is questioned by the curved grating close to the phase singularity. High-performance AGPMs might therefore be limited by the imperfect central region, creating an additional leakage term. Since RCWA simulations show that the performance of the best AGPM designs is not limited to 3×10^{-4}, it is not clear whether the current performance limit is due to the manufacturing uncertainty resulting in imperfect grating parameters, or whether the leakage of the central region of the AGPM is limiting the null depth. The aim of this chapter is therefore to study the influence of the curved grating at the AGPM center on the null depth achievable in a realistic coronagraphic setup. The results presented in this chapter have been published in König et al. (2022).

7.1 The limitation of current AGPMs

The AGPM imprints a vortex phase ramp on an incoming beam using the geometric phase defined by its subwavelength grating design. In principle, this results in a perfect achromatic phase ramp, since the geometric phase is intrinsically achromatic. However, the half-wave plate character of the subwavelength grating is not perfectly achromatic as described in Section 2.1.3. Therefore, although the phase ramp is achromatic, the imperfect half-wave plate character of the subwavelength grating results in a leakage term creating a scaled-down version of the PSF in the coronagraphic image. This leakage term typically dominates the chromatic performance of the AGPM. However, for high-performance AGPMs, a second effect becomes significant.

The second leakage term arises from the curvature of the grating lines at the center of the AGPM, close to the phase singularity. Initially, the grating parameters are optimized to obtain a half-wave plate for an infinite straight parallel grating. However, the grating curvature at the AGPM center becomes too high for RCWA to describe its behavior accurately. Therefore, FDTD needs to be used to describe the effect of the curved grating lines at the AGPM center. Here, a finite patch of the AGPM center including the first 15 grating lines is used, with a size of $30\,\Lambda \times 30\,\Lambda$. This

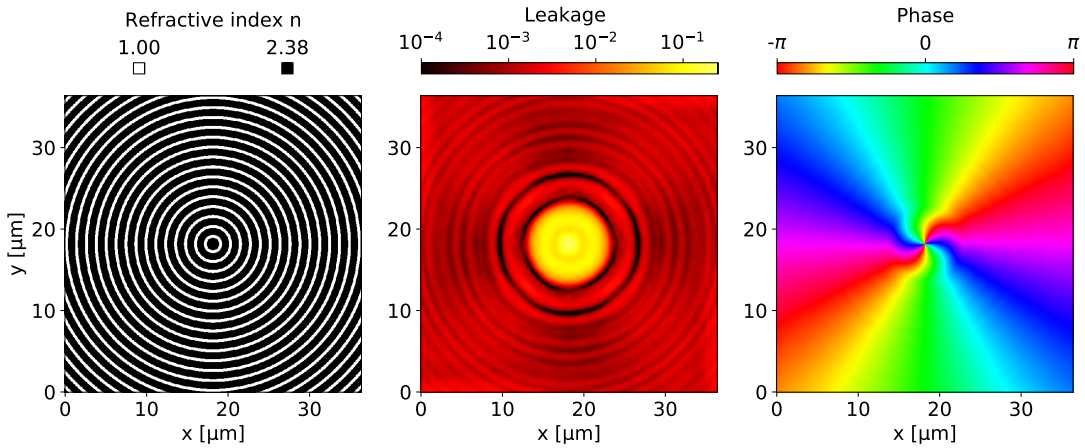

Figure 7.1: Design (left), leakage (middle) and phase ramp (right) at the AGPM center simulated with FDTD at $\lambda = 3.5\,\mu\text{m}$ for the optimal grating parameters for the L-band ($2.9 - 4.1\,\mu\text{m}$): $\Lambda = 1.21\,\mu\text{m}, F = 0.5373, h = 5.5343\,\mu\text{m}, \alpha = 2.45°$.

region corresponds to a patch of radius $0.2\,\lambda/\text{D}$ for an AGPM illuminated at $f/47$ with $\lambda = 3.5\,\mu\text{m}$. Figure 7.1 shows the design, the resulting leakage, and the phase ramp for an AGPM with optimal design parameters (in the sense of RCWA) at an illuminating wavelength $\lambda = 3.5\,\mu\text{m}$. The leakage map clearly shows the central region of increased leakage, which is higher than expected from RCWA. This can be used to define the limit of validity of RCWA in terms of grating line curvature, in the case of the AGPM. The size of the increased leakage region actually depends on the illuminating wavelength and on the details of the central grating geometry, notably the size of the central pillar or hole (see Section 7.4 for more details). For a central pillar with a diameter similar to the grating period, as used here, the central leakage is mostly localized in a circle with radius $5\,\Lambda$. This result means that we can rely on RCWA to describe the AGPM beyond $5\,\Lambda$ from the center, and that we only need to simulate the first five grating lines with FDTD. However because the central leakage region depends on the wavelength, and on the grating design to some level, we choose to use a larger simulation region of radius $15\,\Lambda$ when optimizing the central design, to ensure that all the central leakage is included (Section 7.4).

7.2 Influence of the AGPM central defect

For a realistic estimate of the null depth of an AGPM, it is important to consider the contribution of the central leakage term to the null depth of the full mask. This can be done by calculating the central leakage with FDTD, and by assuming that RCWA describes well the AGPM far from the center. In practice, the AGPM is usually illuminated by an Airy pattern, assuming a circular telescope aperture and an efficient

adaptive optics correction. The central region of the AGPM, where the curvature of the grating is high, generally covers a small fraction of the central Airy disk. By dividing the AGPM into a region of approximately periodic (i.e., locally straight) lines, and another region close to the center with curved lines, the amount of flux incident on both regions can be calculated, depending on the f-number (i.e., the ratio of the focal length to the pupil diameter). The null depth of the full AGPM is then calculated by weighting the null depth of both regions by the incident flux on them. Using the AGPM with different f-numbers changes the size of the Airy disk (a smaller f-number resulting in a smaller Airy pattern), and therefore the influence of the central leakage region, which also changes the overall null depth.

The relative importance of the central leakage term may vary for different grating parameters of the AGPM. It is important to understand how these parameters influence the amount of central leakage, and therefore the overall null depth, in order to understand how deviations from optimal design parameters due to manufacturing uncertainties affect the AGPM performance. Several simulations with non-optimal grating parameters were performed, showing that the relative contribution of the central leakage depends on the intrinsic leakage of the parallel grating (i.e., the null depth calculated with RCWA). This is due to the fact that the central leakage changes more slowly than the intrinsic leakage when changing the grating parameters. Figure 7.2 shows the influence of a higher intrinsic leakage of the parallel grating on the contribution of the central leakage to the null depth. The importance of the central leakage term increases when the intrinsic leakage of the straight grating geometry decreases, and when the f-number on the AGPM decreases. This means that, for high-performance AGPMs with optimal manufactured grating parameters, the central leakage term will be more important than for AGPMs with lower intrinsic performance. The central leakage term will also be more important when using the AGPM at small f-numbers.

7.3 Experimental verification of the central leakage

From the results of the previous section, the influence of the central leakage on the null depth of the full AGPM can be calculated. This result can be compared to the null depth measurements on our infrared coronagraphic test bench VODCA. VODCA uses a setup described earlier in Section 5.4, with a Lyot stop 80 % of the pupil size. Note however that the interpretation of possible discrepancies between predictions and measurements is not straightforward since the manufacturing uncertainty on the grating parameters can have an important influence on the null depth. That means that a poor performance of an AGPM can be both due to manufacturing errors, or to the central leakage term. A possible way to circumvent this problem is to operate the bench with varying f-number. Figure 7.3 shows the simulated null depth for the most probable grating parameters of AGPM-L5 (Vargas Catalán et al., 2016b), currently available in our lab, for different f-numbers. The grating parameters are difficult to infer from non-destructive methods. This uncertainty on the grating parameters

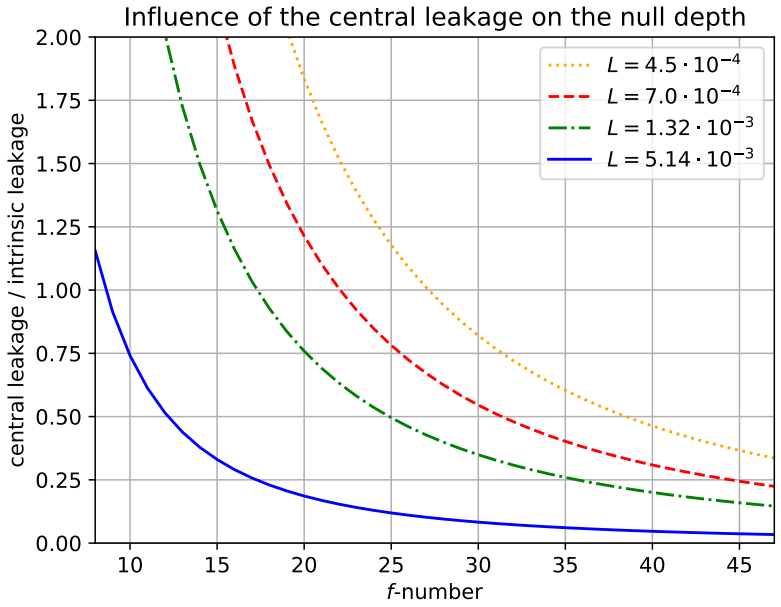

Figure 7.2: Relative contribution of the central leakage term to the null depth, displayed as the ratio of the central leakage term to the intrinsic leakage as a function of the f-number. The curves for four different values of the intrinsic leakage L are shown.

translates into an error on the simulated null depth of up to around 20 %, which is not shown in Figure 7.3 for the sake of clarity. For large f-numbers, the region of central leakage is small compared to the size of the illuminating Airy pattern, and the central leakage has little influence on the null depth. For smaller f-numbers, the size of the Airy pattern decreases, and the central leakage becomes more important, resulting in a significant increase of the null depth. For AGPM-L5, with a central pillar of diameter 1.62 FΛ, the central leakage would contribute 17 % to the null depth of the AGPM at $f/47$, the default f-number of VODCA. At $f/17$, which corresponds to the case of the future METIS instrument at the Extremely Large Telescope (Brandl et al., 2021), the central leakage would contribute a more significant 60 % to the null depth. For smaller f-numbers, the central leakage term would become even more important.

These predictions were tested on VODCA using two different pupil diameters (12.9 mm and 21.9 mm), which correspond to effective f-numbers of $f/47$ and $f/28$. The lower limit of the f-number is defined by the maximal diameter of the beam on VODCA. The upper limit is chosen small enough to show a significant change in performance while keeping a reasonable signal-to-noise ratio on the detector. The null depth for AGPM-L5 was measured with VODCA as described in Section 5.4. The results are shown in Table 7.1 and the resulting performance ratio is illustrated in Figure 7.4. The error bars on the wavelength (horizontal) represent the bandwidth of the filters used on VODCA, while the error bars on the performance increase (vertical) represent

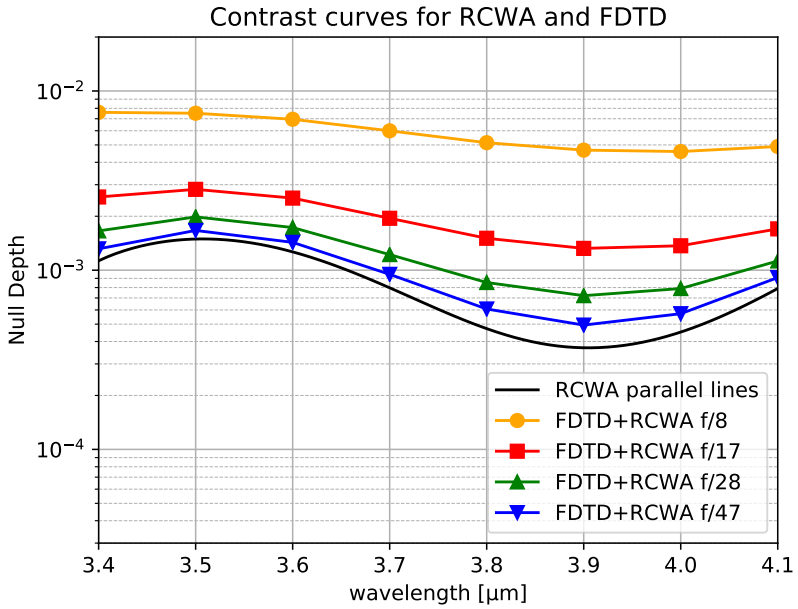

Figure 7.3: Estimated spectral null depth for AGPM-L5 (grating parameters: $\Lambda = 1.21\,\mu\text{m}, F = 0.4320, h = 4.6820\,\mu\text{m}, \alpha = 2.70°$) calculated with RCWA (black line) and combining FDTD and RCWA assuming different f-numbers (colored). The low number of simulation points is due to the high computational cost of FDTD.

	Filter L1 (3425-3525 nm)	Filter L2 (3710-3890 nm)	Filter L3 (3960-4120 nm)
$f/28$	$2.62(15) \times 10^{-3}$	$7.02(40) \times 10^{-4}$	$1.32(8) \times 10^{-3}$
$f/47$	$2.39(14) \times 10^{-3}$	$4.62(26) \times 10^{-4}$	$8.94(51) \times 10^{-4}$

Table 7.1: Null depth measured on VODCA for AGPM-L5 at different f-numbers.

the uncertainty typically observed when measuring the null depths on VODCA. The measurement error is much smaller than the measured difference, except for filter L1 where it is in the same order. In any case the measurements are comfortably above the null depth floor introduced by wavefront errors on VODCA, which is below 10^{-4}, as demonstrated in Jolivet et al. (2019). These results confirm the predicted influence of the central leakage term on the null depth, which is more pronounced for small f-numbers. It is therefore crucial to minimize the central leakage contribution by optimizing the AGPM design at the center.

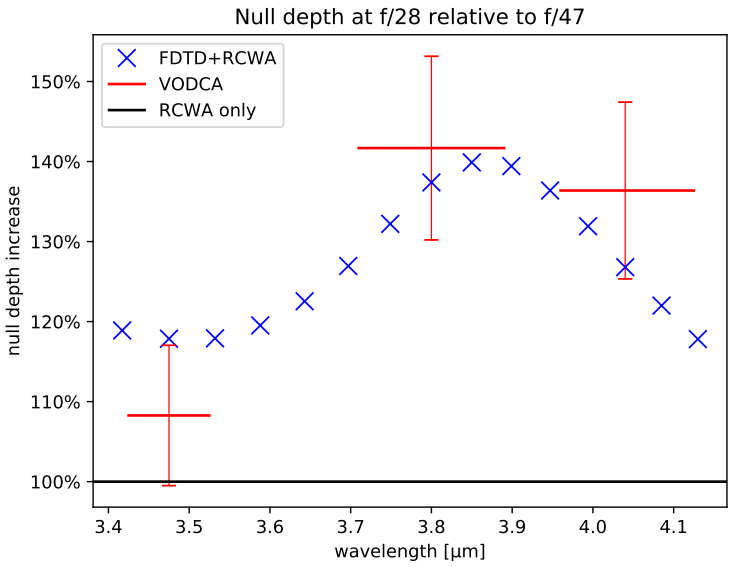

Figure 7.4: Ratio of the null depth at $f/28$ to the null depth at $f/47$ as measured on the VODCA bench for AGPM-L5 (red bars), compared with FDTD+RCWA simulations (blue crosses) along with RCWA-only simulations (black line). The horizontal width of the red bars corresponds to the spectral bandwidth of the VODCA filters.

7.4 Design optimization of the central pillar

The central part of the AGPM may contribute a considerable amount of leakage to the null depth when operated in configurations with small f-numbers. One possible solution is to adapt the central design to minimize the leakage term arising from the curvature of the grating. For this purpose, several values for the central pillar diameter are explored, and whether a pillar or a hole at the center yields the best performance. A graphical table of the simulation results for a set of central designs and various operating wavelengths is shown in Figures 7.6 and 7.7, where the size of the pillar or hole is given at the top of the grating. The central leakage of the different designs is summarized in Figure 7.5. Simulations for a hole of diameter smaller than 1 µm are not performed because of the finite side wall angle of the grating: when the hole size becomes too small, the side walls merge, which makes the manufacturing process unpredictable. The optimal performance is reached with a central pillar of size $1.0\,\mu\text{m} = 1.6\,\text{F}\Lambda$, which is (by pure coincidence) close to the range of pillar sizes that was used for the AGPMs manufactured so far ($1 - 2\,\text{F}\Lambda$).

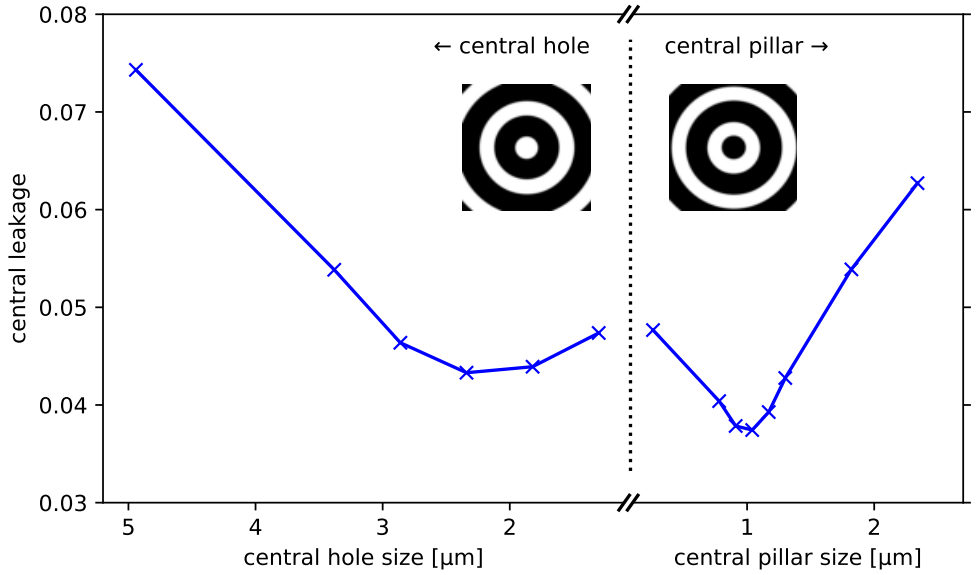

Figure 7.5: Simulated central leakage of an optimal AGPM for the L-band for different size of the central pillar or hole. The best performance is reached at a central pillar of diameter 1.0 µm corresponding to 1.6 FΛ. The grating parameters of the AGPM used here are $\Lambda = 1.21\,\mu\text{m}, F = 0.5373, h = 5.5343\,\mu\text{m}, \alpha = 2.45°$.

7.5 Design optimization of the AGPM center with gradient parameters

Another way to optimize the central design is to introduce a gradient in the period and filling factor in the central region ($r < r_0$) (Yu et al., 2011; Yu & Capasso, 2014), using the following form:

$$\Lambda(r < r_0) = \Lambda_0 + \nabla\Lambda \cdot (r_0 - r)$$
$$F(r < r_0) = F_0 + \nabla F \cdot (r_0 - r)$$

with Λ_0 and F_0 being the optimal grating period and filling factor for a straight grating (denoted Λ and F in the previous sections), r the distance from the AGPM center, $r_0 = 5\Lambda_0$ the radius of the central region, and where $\nabla\Lambda$ and ∇F are the gradient parameters to be optimized. An exploration of the parameter space spanned by the two gradients $\nabla\Lambda$ and ∇F was performed here. The results are shown in Figure 7.8. For clarity, the grating design is shown as insets for selected points of the parameter space. Here, the simulations are limited to the relevant region of the parameter space, where both the subwavelength condition and manufacturing limits (minimal feature size) are fulfilled. Also, the finite side wall angle leads to a region of the parameter space with merging sidewalls at the bottom of the grooves (green region), for which

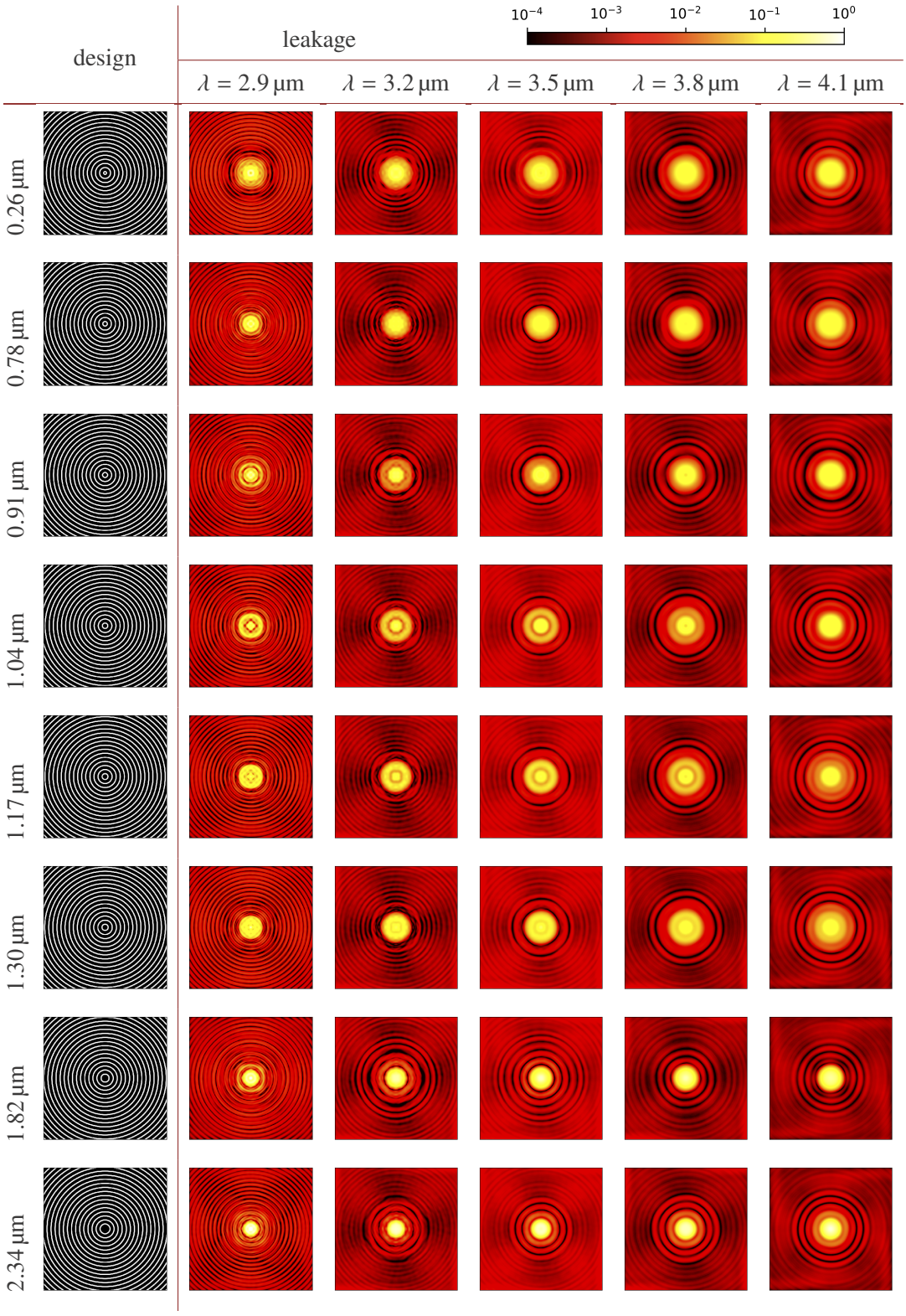

Figure 7.6: Leakage for several designs with different central pillar size. From left to right: nominal diameter of the central pillar, design, and leakage maps for 5 wavelengths in the L-band ($2.9 - 4.1\,\mu m$).

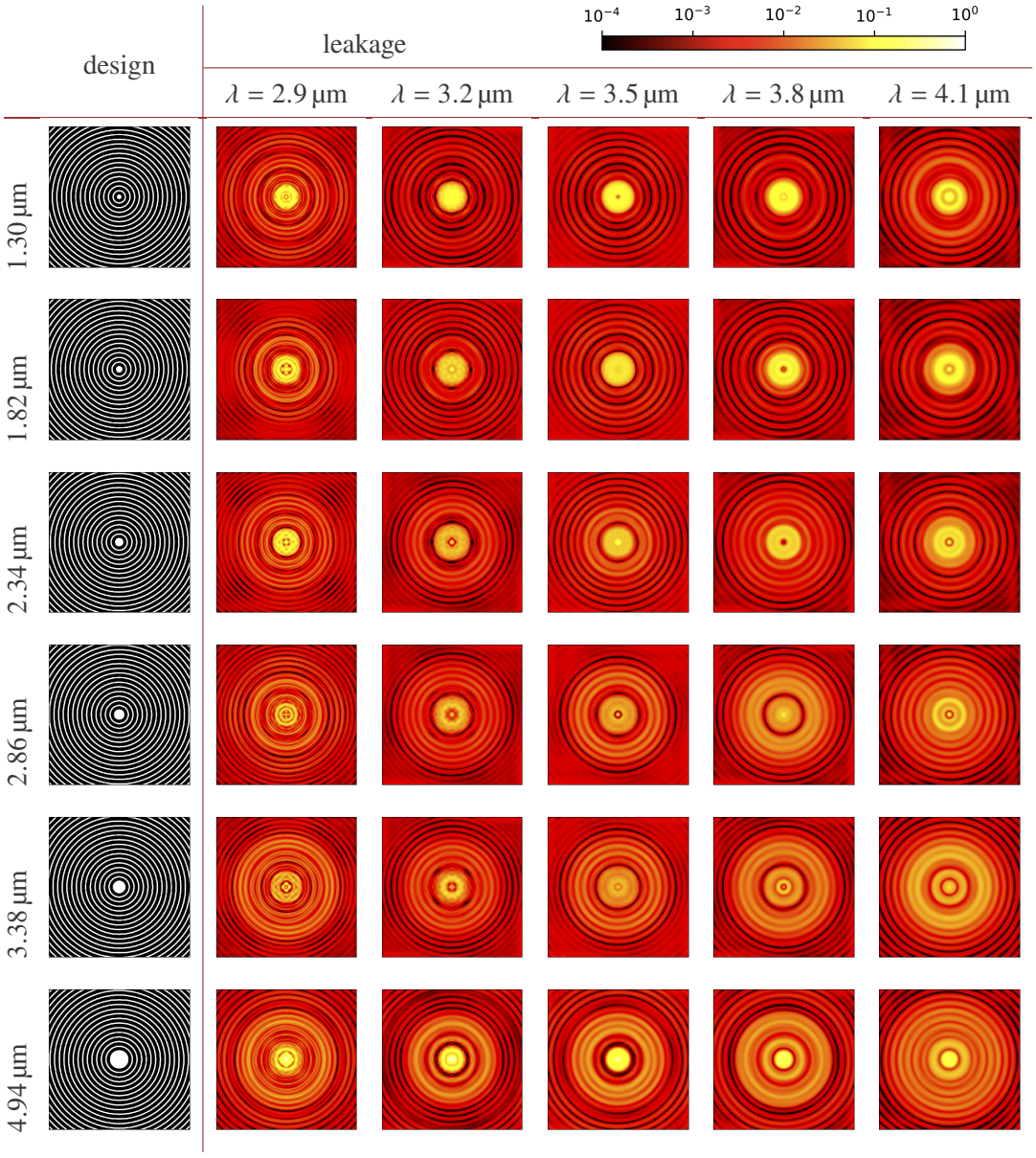

Figure 7.7: Leakage for several designs with different central hole size. From left to right: nominal diameter of the central hole, design, and leakage maps for 5 wavelengths in the L-band ($2.9 - 4.1\,\mu m$).

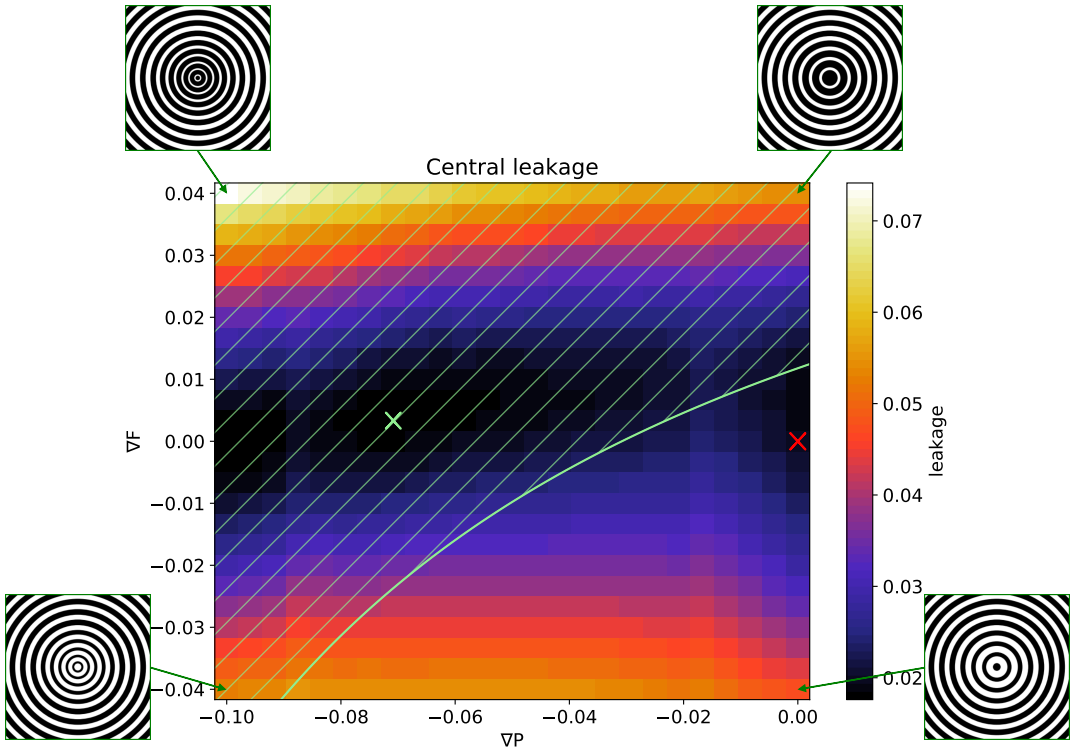

Figure 7.8: Leakage map of the parameter space spanned by the gradient of the period $\nabla\Lambda$ and of the filling factor ∇F in the innermost 5Λ. Several insets of the central designs are shown for selected points of the parameter space. The optimum is marked with a green cross, the design with constant period and filling factor is marked as red cross. The region above the green line is challenging to manufacture because of the finite side wall angle of the grating resulting in designs with merging sidewalls, for which the manufacturing process has not been tested. The optimal grating parameters for the L-band ($2.9-4.1\,\mu m$) are used: $\Lambda = 1.21\,\mu m, F = 0.5373, h = 5.5343\,\mu m, \alpha = 2.45°$. For different grating parameters, the resulting optimal designs might change slightly, but the conclusions are expected to be similar.

the manufacturing process has not been tested. The central leakage at the optimal parameters $\nabla F = 0.0033$ and $\nabla\Lambda = -0.0708$ (green cross) is reduced by 8 % compared to the design with $\nabla F = 0$ and $\nabla\Lambda = 0$ used so far (red cross). This would lead to a limited improvement in the total null depth of $< 1\,\%$ on VODCA and $< 5\,\%$ for METIS, at the expense of a more challenging manufacturing process.

8 Subwavelength grating vortex phase masks of topological charge 4

The AGPM, a charge-2 implementation of a subwavelength grating vortex phase mask (SGVPM), has been installed on some of the most advanced ground-based telescopes in the past years providing excellent results. However, it seems mandatory to increase the topological charge of the vortex to make the coronagraph more resilient to low-order aberrations and to partly resolved stars, which will be the case for future ELT-class telescopes. Though a theoretical grating design with continuous grating lines exists for all even topological charges, the implementation of SGVPMs of charge $\ell_p \neq 2$ is not readily possible, as explained in Section 2.2.4. This is due to the varying grating period necessary to provide the helical phase ramp, leading to designs for which the grating period becomes larger than the subwavelength limit in certain regions of the mask, and smaller than the manufacturing limit in other regions. Segmented designs are a promising approach to tackle these issues and implement a vortex of topological charge 4 and higher. This chapter quickly reviews the two charge-4 designs developed so far (Delacroix et al., 2014) and presents a new hybrid design combining the two approaches, along with its implementation using the updated diamond microfabrication process described in the previous chapter. This chapter is largely based on results published in König et al. (2020).

8.1 Current charge-4 designs

For the AGPM, a continuous pattern of concentric grating lines is used. For higher topological charges, the pattern becomes more complicated and impossible to manufacture. To mitigate this problem, Delacroix et al. (2014) suggested the segmentation of the grating pattern. Two main patterns have been proposed, one using equally spaced segments with straight lines and one using curved lines with segments of variable size, as shown earlier in Figure 2.13.

Three subwavelength grating vortex phase masks of topological charge 4 (SGVPM4) for the *L*-band have been manufactured in the past years: one eight-segment mask (SGVC4-S08), and two identical 32-segment masks (SGVC4-S32A and SGVC4-S32B), all featuring straight grating lines. The manufacturing of these masks is discussed in Vargas Catalán et al. (2018) and their performance was tested on the VODCA test bench. Only one of the masks (SGVC4-S32B) reached an acceptable null depth of

	Filter *L* (3575-4125 nm)	Filter *L1* (3425-3525 nm)	Filter *L2* (3710-3890 nm)	Filter *L3* (3960-4120 nm)
SGVC4-S08	1.01×10^{-2}	1.33×10^{-2}	9.01×10^{-3}	6.06×10^{-3}
SGVC4-S32A	1.69×10^{-2}	1.75×10^{-2}	1.64×10^{-2}	1.92×10^{-2}
SGVC4-S32B	1.65×10^{-3}	1.04×10^{-3}	1.59×10^{-3}	3.46×10^{-3}

Table 8.1: Null depth of the three charge-4 masks featuring straight grating lines. The performance is measured in one broadband (*L*) and three narrowband filters (*L1*, *L2*, *L3*) covering the astronomical *L*-band.

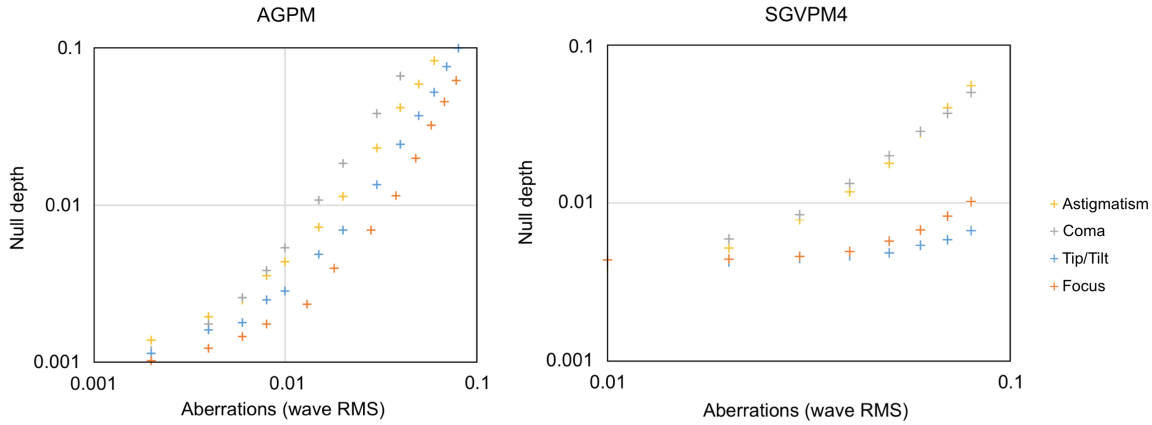

Figure 8.1: Influence of low-order aberrations on the null depth for the AGPM (left) and SGVC4-S32B (right). Adapted from Jolivet (2019).

1.65×10^{-3} over the full *L*-band, which is comparable to the results obtained with an AGPM. SGVC4-S32A performs significantly worse although having the same pattern suggesting an anomaly in the fabrication process. It is not possible to conclude on the performance of the eight-segment mask SGVC4-S08 since only one mask has been manufactured and the poor performance can be due to manufacturing errors. The null depths of the three masks measured on VODCA are shown in Table 8.1. The validity of the 32-segment straight lines pattern is confirmed with these results. Note that strictly speaking the original acronym SGVC4 (subwavelength grating vortex coronagraph of charge 4) is misleading since it should refer to the phase mask only and not to the full coronagraph system. The notation when referring to the three phase masks presented in (Vargas Catalán et al., 2018) is however kept for consistency with previous results.

It is interesting to assess the performance of the SGVPM4 in the presence of low order aberrations, which can be added on VODCA using a deformable mirror. Tests on the sensitivity of the null depth in presence of aberrations have been performed in Jolivet (2019) for an AGPM and the SGVC4-S32B. The results are shown in Figure 8.1. The presence of tip/tilt and focus proves to have less influence on the performance of the SGVPM4, which is expected from theory (Huby et al., 2015).

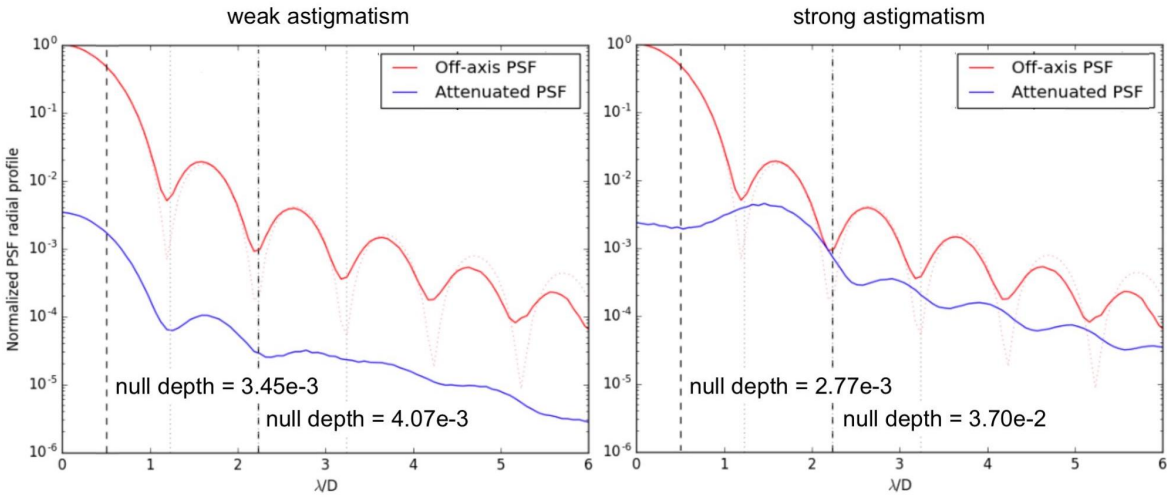

Figure 8.2: Influence of astigmatism on the null depth. The coronagraphic PSF (blue) is shown along with the reference PSF (red) for weak (left) and strong (right) astigmatism. A sufficiently large region needs to be included in the calculation of the null depth to capture the effect of astigmatism. Adapted from Jolivet (2019).

When adding astigmatism, the figure of merit to calculate the null depth needs to be adapted to the radial intensity profile of the attenuated PSF. Integrating the FWHM of the Airy disk yields a counter-intuitive increase of the rejection ratio when adding astigmatism. The true null depth is calculated including the first Airy ring, that is up to $2.23\,\lambda/\mathrm{D}$ (see Figure 8.2). It is therefore crucial to know the radial profile of the PSF to correctly interpret the attenuation.

SGVC-S32B features a pattern using straight grating lines segments, resulting in a staircase phase ramp as shown earlier in Figure 2.5. Though the mask has shown comparable performance to the AGPM, the transition regions between the segments create a phase step ($\pi/4$ in case of a charge-4 32-segments mask) partially attenuating an off-axis source. Figure 8.3 shows the transmission map of one quadrant of SGVC4-S32B. The transmission decreases by 20 % on the transitions between segments, resulting in an attenuation of an already faint off-axis companion.

To avoid this effect, a second pattern featuring curved grating lines was proposed, as illustrated in the top right part of Figure 8.4. This pattern has no phase steps between segments since the orientation of the grating lines in each segment follows the theoretical pattern needed to obtain a perfect vortex phase ramp. However, this results in segments with locally varying grating parameters. In particular, for segments of fixed size the grating period could become larger than the subwavelength limit, or smaller than the limits imposed by the manufacturing process. In addition, the varying grating parameters across the mask result in reduced performance, since the grating parameters are not optimal in every part of the mask. In practice, the height and filling

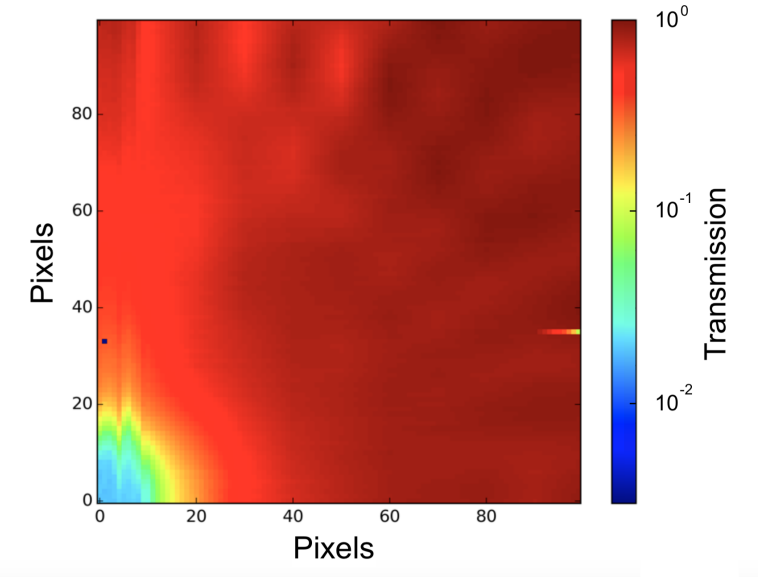

Figure 8.3: Off-axis transmission map for SGVC4-S32B. The transmission of one quadrant is shown, with the mask center located at the bottom left of the image.

factor of the grating are maintained constant, while the grating period is reduced by up to 20 %, which is chosen as practical limit for which the grating still provides an acceptable null depth. This condition results in few large segments along the horizontal axis of the mask, and many small segments along the vertical axis.

8.2 Hybrid charge-4 design

The two concepts capable of implementing the vortex phase ramp with segmented patterns presented so far have the drawback of using an unnecessary high number of segments along one axis (the straight lines pattern has a lot of segments along the horizontal axis, which could be approximated with a curved lines segment, while the curved lines design has many small segments along the vertical axis). A third hybrid pattern combining the advantages of both concepts is therefore proposed, as illustrated in Figure 8.4. The regions along the vertical axis of the straight lines pattern are combined with the curved lines pattern in regions along the horizontal axis, avoiding the unnecessary large number of segments along the other axis, respectively.

The three different patterns were simulated with FDTD, as shown in Figure 8.5. The design as well as the polarization leakage and the phase ramp are shown. In the leakage maps, the effect of the discontinuities of the grating pattern between segments is clearly visible as regions of increased polarization leakage in all three patterns. However, the curved lines pattern shows lower leakage along the horizontal axis than the straight lines pattern, because this region is covered by one single segment without grating discontinuities. The phase ramp is also best for large segments of the mask, justifying

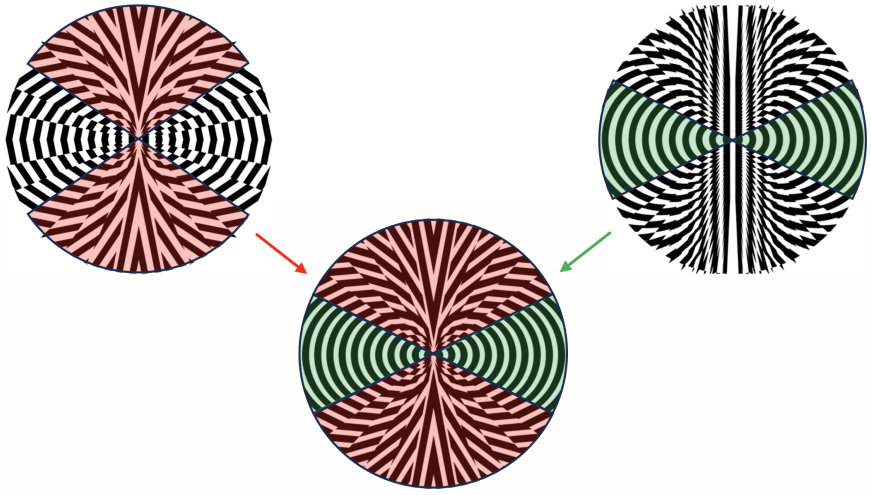

Figure 8.4: Hybrid design of charge-4 SGVPMs. The hybrid pattern (middle) combines the advantages (red and green regions) of both straight (left) and curved (right) lines patterns described in Figure 2.13.

the use of segments with curved grating lines along the horizontal axis of the combined hybrid pattern. However, the curved grating pattern features a lot of very small segments along the vertical axis, leading to a poor phase ramp and higher leakage. For this region, the straight lines pattern provides the better phase ramp, which justifies its choice along the vertical axis of the hybrid pattern. Consequently, the hybrid design shows the best phase ramp and also the lowest leakage (though all three patterns have similar leakage), as expected since it combines the advantages of both straight and curved lines designs.

8.3 Manufacturing of hybrid charge-4 design

Three SGVPM4 featuring straight grating lines have been manufactured in the past years. Since then, the manufacturing process has evolved, reducing the number of steps in the process, as described in Section 6.2. Using the updated process, the promising hybrid charge-4 SGVPM has been recently prepared for manufacturing. The pattern was transferred to the substrate sputtered with a thick silicon layer using the SAMIM technique. The first image in Figure 8.6 shows the central part of the pattern transferred to the photoresist layer on top of the silicon layer. The pattern was transferred successfully, and even sharp features of the pattern were reproduced accurately.

The photoresist mask was then used to etch the silicon layer. Figure 8.6 also shows the central part of the mask after etching the silicon layer. A top view of the pattern is included, as well as a slanted view showing the characteristic profile of the sidewalls due to the Bosch process that is used for the silicon etching (see Section 6.2). The

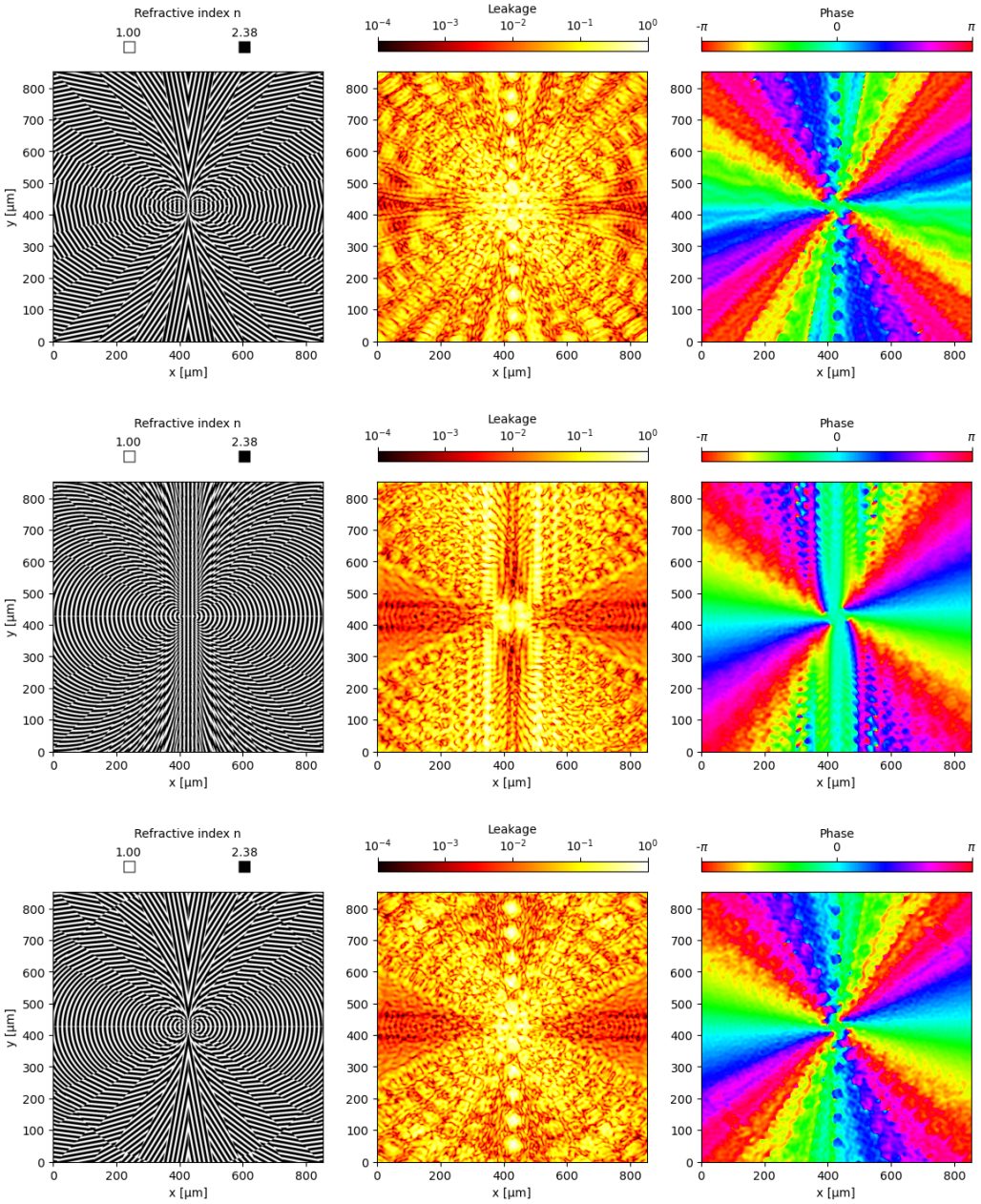

Figure 8.5: Design (left), polarization leakage (middle), and phase ramp (right) of charge-4 vortex phase mask designs simulated with FDTD. The three patterns featuring straight lines (top), curved lines (middle) and the hybrid design (bottom) are shown. The mean polarization leakage of the designs are $l = 6.64 \times 10^{-2}$ (straight lines), $l = 7.14 \times 10^{-2}$ (curved lines) and $l = 6.28 \times 10^{-2}$ (hybrid). The designs are optimized for the L-band ($3.4 - 4.1\,\mu\text{m}$), and the results are shown for $\lambda = 3.8\,\mu\text{m}$.

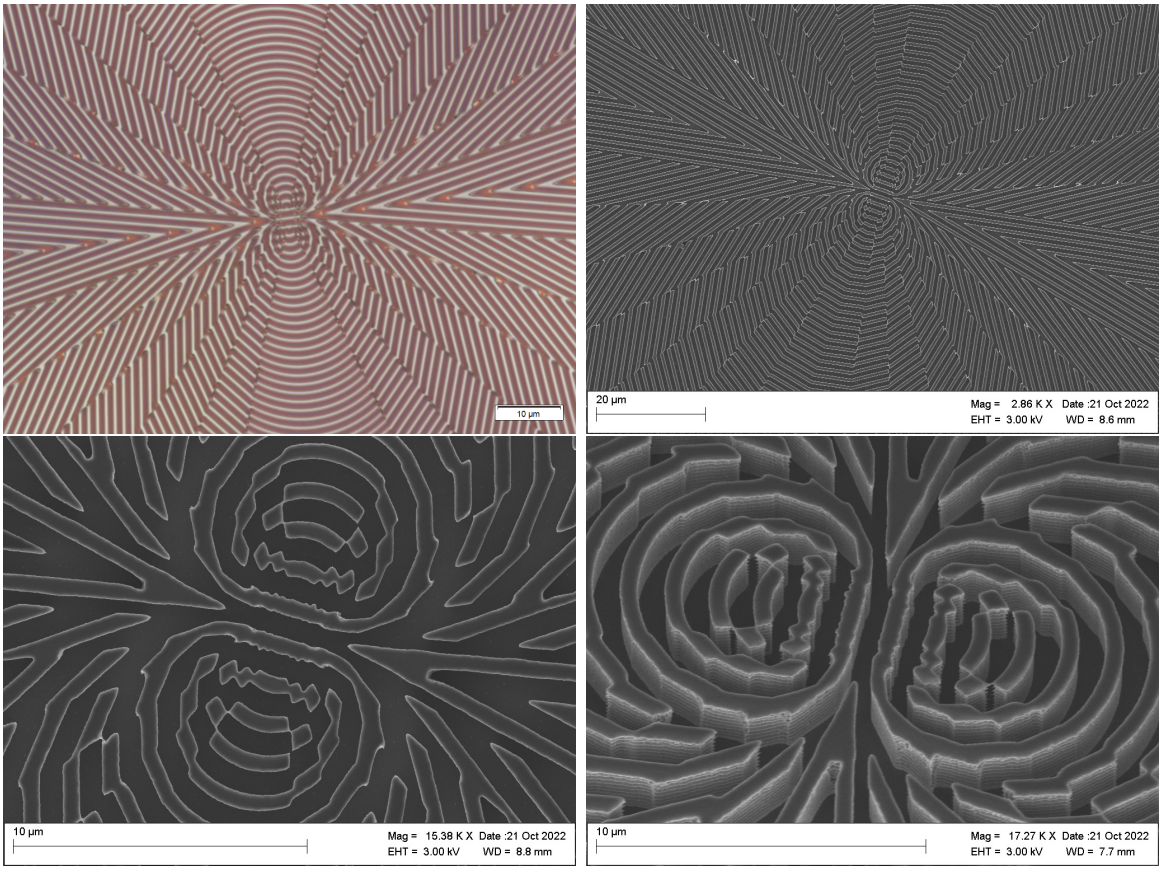

Figure 8.6: Optical microscope (top left) and SEM (top right and bottom) images of the hybrid charge-4 SGVPM. Top: the pattern imprinted on the mask is shown before (left, optical microscope image) and after (right, SEM image) the silicon etching. Bottom: top view (left) of the central part after the silicon etching showing that even sharp features are transferred well to the silicon mask, as well as slanted view (right) showing the characteristic sidewalls due to the Bosch process that was used for etching the silicon layer.

pattern is well transferred into the silicon layer, including the sharp features close to the center. The final etching of the diamond substrate has not been done so far due to technical issues encountered with the inductively coupled plasma etcher (PlasmaTherm ICP). While another plasma etcher is available (referred to as Tegal), which was used successfully to manufacture the N-band AGPMs for METIS (see Chapter 6), it is not possible to use this etcher for the hybrid charge-4 pattern, because it provides a sidewall angle ($\approx 1.5°$) for which the pattern was not optimized (the pattern was optimized for a sidewall angle of 2.1°). While small variations of the sidewall angle can be compensated to a certain extent by adjusting the etch depth, the difference between the two etchers is too high to be compensated, and using the different etcher would lead to a degraded optical performance of the mask. The mask will therefore be finalized when the adequate etcher is available.

Part IV

Metasurface designs

9 A machine learning approach to optimize the AGPM central region

The performance of the subwavelength grating designs presented in the previous chapters relies on the optimization of the grating parameters and the orientation of the grating lines across the phase mask. The free parameters that are optimized are the height, period and filling factor of the grating, as well as the sidewall angle to a certain extent (limited by manufacturing constraints). These designs have a limited number of parameters to be optimized and achieve good performance across large bandwidths. However, the periodicity exploited in these designs breaks down at the center of the vortex as described in Chapter 7, or when simulating segmented designs with aperiodic pattern as described in Chapter 8. One way to increase the performance capability of these regions is to leverage the increased freedom introduced with a two-dimensional metasurface design framework. This design framework introduces new degrees of freedom for the design of broadband vortex phase masks. However, optimizing these metasurface structures results in an optimization problem featuring a large number of free parameters. The standard algorithms used for subwavelength grating structures are not suited for this more complex task and necessitate a clever optimization strategy, presented in this chapter. The optimizer described in this chapter was developed by Nicolas Roy in the framework of his dissertation. The results in this chapter were obtained with this optimization framework and are largely based on a paper submitted for publication (Roy et al., 2023).

9.1 The need for more efficient optimizers

Generally, exploiting the flexibility of metasurface designs results in a large number of free parameters, proportional to the number of metasurface building blocks. For metasurface designs featuring thousands of building blocks this results in a prohibitively large number of free parameters. A full circular metasurface phase mask of 10 mm diameter for application in the L-band would contain 54 million metasurface blocks, each with two positional coordinates, two lateral block sizes, and a rotation angle, resulting in 270 million free parameters. Clearly this kind of optimization problem must be tackled by an adequate approach, since standard optimization algorithms are not suited to handle the design freedom introduced by metasurfaces.

A full optimization of all free parameters of the metasurface design is not always necessary and the problem can be simplified considerably by exploiting symmetries. For example, in the case of a vector vortex phase mask with rectangular metasurface building blocks, the rotation of the blocks can be fixed according to the vectorial nature of the design, imparting the desired geometric phase. The number of free parameters can also be reduced by one order of magnitude by exploiting the symmetry axes of the pixelization grid. For example, for a cartesian pixelization grid the number of free parameters is reduced by a factor of 8 in the case of a charge-2 design, since the mask can be divided in 8 equivalent segments.

Another way to reduce the number of free parameters is to assume that similar metasurface blocks have a similar response to an incident wave. In this case the response of the metasurface blocks can be inferred by calculating only a small subset of block sizes and interpolating the results for the missing intermediate block sizes.

Still, the number of simulation parameters remains considerably high, and a lot of simulations would be needed to sample the parameter space sufficiently well. This is even more of an issue, since a metasurface pattern is non-periodic in the case of a vortex phase mask. As discussed in Section 4.1, this requires a rigorous tool such as FDTD to take into account the finite aperiodic features of the vortex, increasing the necessary computational power considerably. This requires an optimization framework converging to a solution faster than conventional optimization algorithms, thereby reducing the number of computationally heavy FDTD simulations to be performed.

9.2 The surrogate model using U-Net

One solution to reduce the number of FDTD simulations is to use a surrogate model. Instead of simulating the metasurface response of each queried parameters with FDTD, the response of the metasurface is inferred via a surrogate model previously trained with a set of FDTD simulations. This allows for performance prediction of a certain metasurface design without rigorously simulating it. The surrogate model can be understood as an interpolation of a set of known true simulation results for the query parameters. With a classical interpolation framework, the number of true values scales with the number of dimensions of the problem, which is the number of free parameters. With a surrogate model, only a fraction of these designs is needed, since the surrogate model reduces the dimensionality of the problem, breaking it down to its principal components.

Initially, the surrogate model is trained with a random set of metasurface patterns and their respective responses. These patterns are simulated with FDTD and saved to a database which is used to train the model. The design can then be optimized using the model as a surrogate to query specific points in the parameter space that have not been covered with FDTD simulations. In a further step, promising designs resulting

from this optimization are evaluated with FDTD and used to update the surrogate model.

Surrogate models have the additional advantage of reducing the noise of a problem of high dimensionality by providing a smoothened performance metric to the optimizer. Several surrogate models have been proposed in the past, such as radial basis functions (Powell, 1992) and Kriging (Bouhlel et al., 2016), or classic machine learning methods such as regression trees (Chen & Guestrin, 2016). Deep neural networks (Sun & Wang, 2019; Kochkov et al., 2021) have gained particular interest in photonics (Chen et al., 2022), as their architecture can be adapted to the specific problem.

Here, the U-Net (Ronneberger et al., 2015) is used as the surrogate model. The U-Net is similar to a convolutional auto encoder and has proven to perform well in image translation tasks. In this particular case, the dielectric permittivity of a metasurface design is used as a two-dimensional input and translated to a leakage map at same resolution. Figure 9.1 shows a two-dimensional cut of the dielectric structure image going through the different convolution operations until reaching a representation of lower dimensionality before being transformed back to match its original size. Unlike a convolutional auto encoder, the U-Net allows for connections between the layers skipping the central representation of lowest dimensionality. The advantage of allowing for shortcuts between the layer architecture is the ability to link specific features early to a certain response without having to go through the full network architecture with all layers. This allows low-level localized features to pass through and merge with the features going through all convolution layers. The U-Net has recently proven its strengths for the application to diffracting systems (Chen et al., 2022) and is therefore a promising approach for the application to metasurface designs.

9.3 Performance of the U-Net

The U-Net has been applied to the problem of predicting the leakage distribution of metasurface designs and compared to various other techniques showing superior performance to the other considered approaches. Figure 9.2 shows the correlation of the predicted leakage using different surrogate models compared to the actual FDTD simulation results. The same 2500 random designs have been used for each model. Each of the four plots shows the correlation of the mean predicted leakage map of the model and the actual simulated mean leakage. The training dataset used to train the respective models is shown along with the validation dataset used to assess the predictive performance of the model. For a perfect model, the expected behavior is a perfect diagonal line in the graph. The metric to quantify the quality of the models is the Pearson coefficient (Freedman et al., 2007), which describes how well the model reproduces the true value. The U-Net outperforms all three considered concurrent approaches reaching a correlation coefficient of 98 %, while a classical convolutional auto encoder (CAE) reaches 82 %. Although both methods use similar architectures

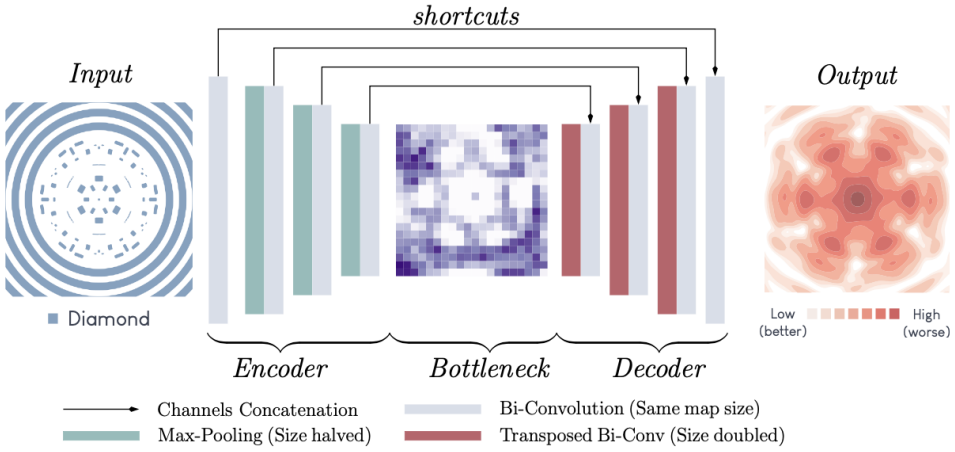

Figure 9.1: Structure of the U-Net showing the dielectric structure of a sample design going through the different convolution operations until reaching the representation of lowest dimensionality (bottleneck), before being transformed back to the original size. The shortcuts between layers allow characteristic low-level features to skip the bottleneck.

and reach high performance, the key difference between them is the presence of skip-connections in the U-Net used to link finer features of the design directly to a specific localized leakage distribution. Clearly the U-Net describes more accurately fine features of the leakage maps, as shown in the right part of Figure 9.2, where a set of leakage maps is reconstructed from the model parameters of both CAE and U-Net, and compared with the exact FDTD leakage maps. The other two models considered are deep neural networks (DNN) and partial least square Kriging (KPLS). Both methods reach correlation factors of only around 50 %. This poor result is mainly due to the complexity of the problem and the way DNNs and KPLS deal with the data, since both methods directly infer the mean leakage from the input geometric parameters of the designs. These results show that the U-Net is optimally suited for the presented problem, particularly because it allows the use of characteristic two-dimensional slices that are ready to be processed.

Generally, a machine learning algorithm requires a training dataset consisting of thousands of simulations. This can be problematic for computationally heavy simulations like FDTD. It is therefore important to study the behavior of the U-Net with limited training data. Figure 9.3 shows the performance of the U-Net compared to a DNN for training datasets of different size. The Pearson correlation coefficient is used as performance metric, showing values above 0.5 for the U-Net with a training dataset of as few as 100 simulations. The U-Net performs significantly better than the DNN, which is shown as a black line for comparison. To reduce the number of FDTD simulations further, data augmentation is used by cropping and rotating the metasurface patterns and respective leakage maps randomly (Shorten & Khoshgoftaar, 2019). In this way,

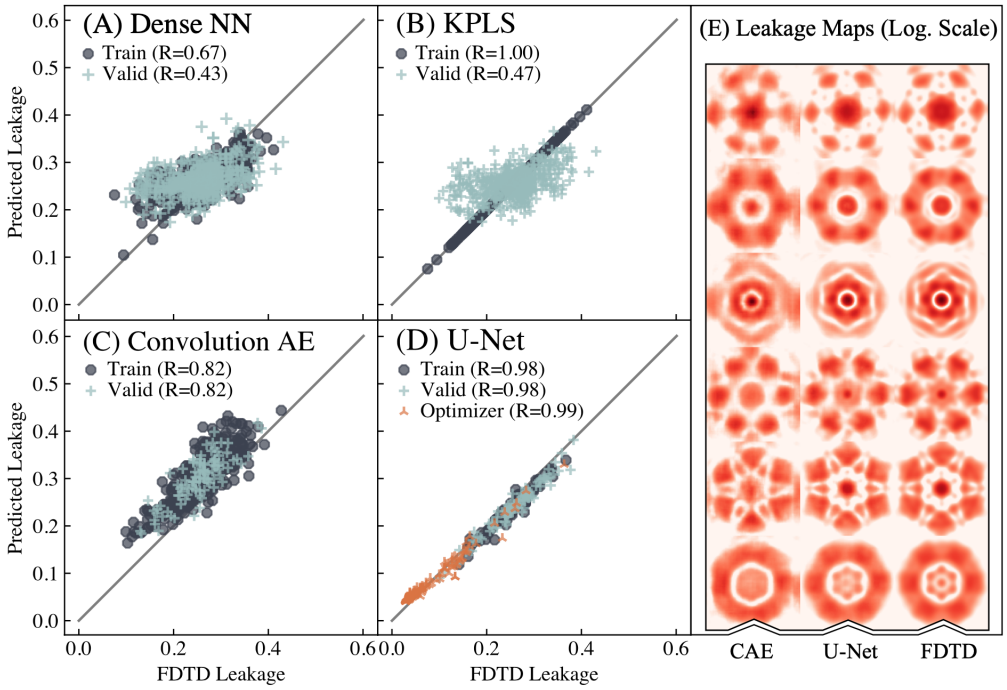

Figure 9.2: Surrogate models tested on a metasurface design comprising rectangular building blocks. Four methods are compared: A trilayer dense neural network (A), partial least square kriging (B), a convolutional auto encoder (C) and the U-Net (D). The plots show the predicted mean leakage against the simulated mean leakage obtained with FDTD. The pearson correlation coefficients R are shown as insets showing that the U-Net performs best. The difference between the two best methods is illustrated in inset (E) which shows a few example leakage maps predicted by both methods and the corresponding simulated leakage maps obtained with FDTD.

several configurations can be extracted from a single FDTD simulation. Data augmentation generally helps to increase the robustness of a machine learning model since the model is exposed to more training data. With data augmentation, the correlation factor of 0.5 is reached with as few as 50 FDTD simulations. Using data augmentation can therefore reduce significantly the computational load needed to achieve optimal performance of the U-Net.

9.4 Particle swarm optimization

Defining a surrogate model is an important step in reducing the computational load of the metasurface optimization problem. The U-Net as a surrogate model has shown to be an efficient and accurate model to connect a metasurface design to its leakage map

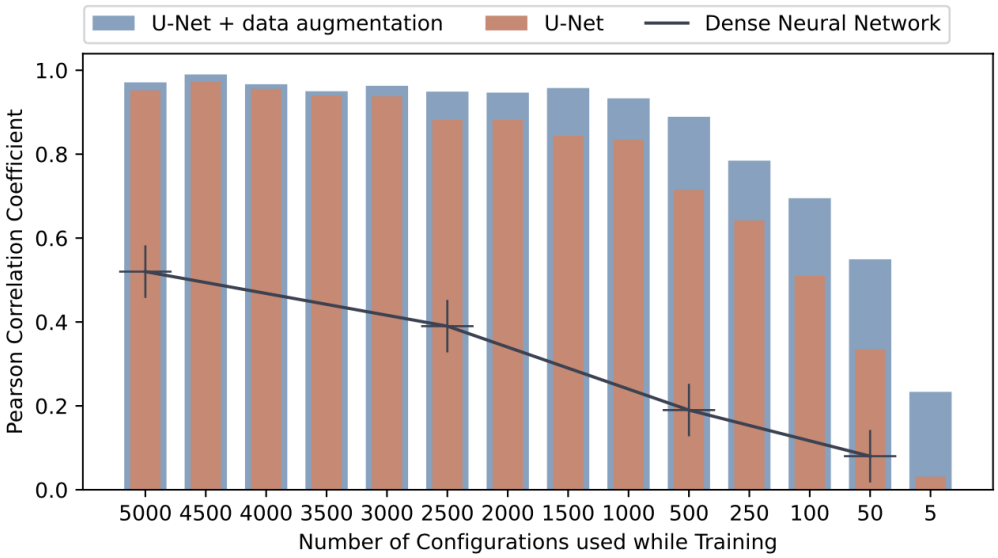

Figure 9.3: Influence of the dataset size on the performance of different surrogate models. The U-Net is clearly outperforming the dense neural network for the given training sets. The figure also shows that data augmentation can efficiently increase the performance of the U-Net, especially for small training sets providing a pearson correlation coefficient above 0.5 for as few as 50 datasets.

reducing drastically the number of FDTD simulations needed to sample the parameter space. The U-Net can now be used in a global optimization algorithm. Here, particle swarm optimization (PSO) (Kennedy & Eberhart, 1995) is used since it is particularly suitable for application to smooth optimization problems with continuous parameters, which is the case for the U-Net as surrogate model. PSO draws inspiration from swarm intelligence and uses cluster classification to find a global optimum.

The workflow of the PSO algorithm is illustrated in Figure 9.4. Starting with a random set of 50 designs representative of the parameter space, their performance is evaluated using FDTD. The results constitute the first training set of the U-Net, which is then used as a surrogate model for the particle swarm optimization algorithm. PSO queries a large number of designs using the U-Net. However, these designs rely on the surrogate model and can not be added directly to the training set since they are based on the model itself. Instead, the most representative designs are evaluated using FDTD and added to the training set. These designs are the ones that give additional information on the search space and the designs performing the best. It is important to balance designs exploring the search space and designs giving the best performance in order to converge to a global optimum. Refining only the best performing designs might end up in a local minimum, while trying too many less promising designs might not converge at all. To allow for both options, k-means clustering with four clusters is used. One sample

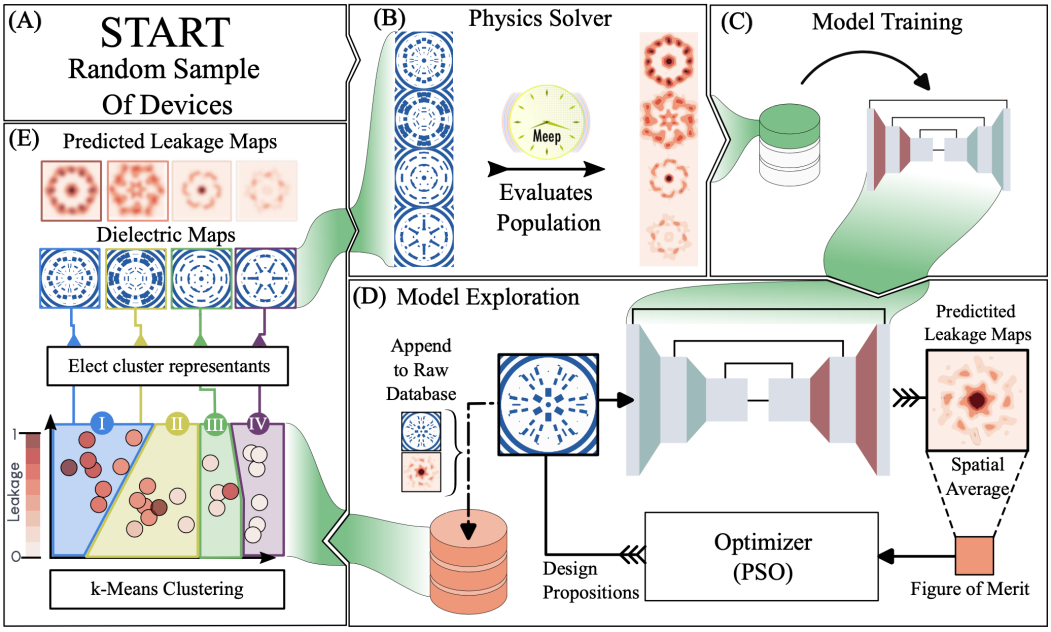

Figure 9.4: Workflow of the particle swarm optimization algorithm combined with the U-Net as the surrogate model. The optimizer starts with a random set of 50 designs (A) which are simulated using FDTD (B). The resulting leakage maps are used to train the surrogate model (C) and then optimized using particle swarm optimization (D). The resulting designs are then divided in four clusters using k-means clustering, and the most representative design of each cluster is selected (E). Finally, these designs are simulated with FDTD and incorporated in the model (B, C).

from each cluster is then selected using a probability inversely proportional to the mean leakage. The selected designs are then evaluated using FDTD and incorporated in the U-Net model, closing the optimization loop.

9.5 Results of the optimization framework

The presented optimization framework is now applied to two scenarios. Both scenarios are based on the circular AGPM pattern and include an optimized diamond metasurface pattern close to the center in a region with a radius of 5 grating periods. Figure 9.5 illustrates the two scenarios showing a two-dimensional cut of the designs. The physical simulation cell is surrounded by perfectly matched layers, as described in Section 4.1.

The first scenario features grating lines of varying line width and position inside the optimizable area. The design is defined by the inner and outer radius of the 5 circular grating lines, resulting in 10 free parameters.

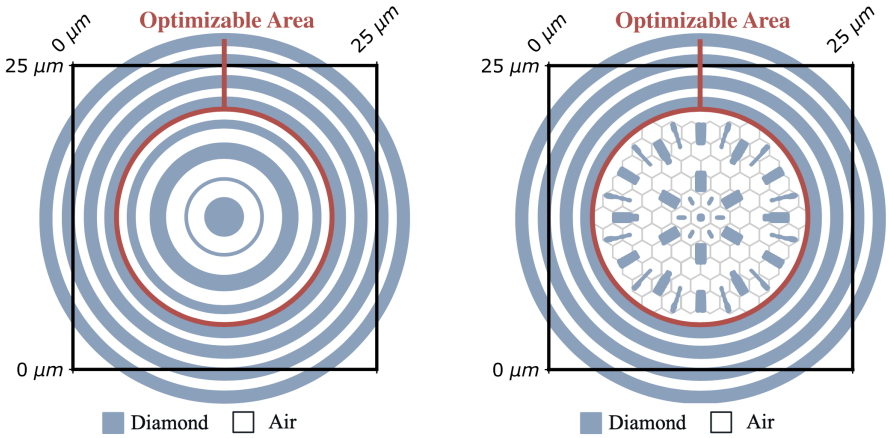

Figure 9.5: Two scenarios for the AGPM center optimization using machine learning. Left: Circular grating lines of varying size and position defined by their inner and outer radii. Right: Rectangular nanofin pattern featuring up to 60 rectangular building blocks and a central pillar.

The design freedom is drastically increased in the second scenario with a metasurface framework using rectangular nanofins placed in a hexagonal grid. This leads to an optimization problem of 61 rectangular blocks with 5 parameters each (size and position in x and y, as well as rotation angle) resulting in 455 free parameters. However, the number of free parameters can be reduced by imposing the rotation angle as defined by the geometric phase, and by exploiting the symmetry axes of the hexagonal pixelization grid, as described at the beginning of this chapter. This results in the design constraints shown in Figure 9.6, where the region of independent blocks is highlighted in green. This area comprises 11 building blocks and one central pillar ($11 \cdot 4 + 1 = 45$ independent parameters). The symmetry constraints remove further parameters (notably there is no azimuthal displacement of the blocks highlighted in gray) resulting in $45 - 7 = 38$ free parameters. During the optimization some nanofins of the pattern might become very small and it can therefore be advantageous to remove them from the design.

The optimization is carried out for both designs featuring circular grooves and rectangular nanofins. Figure 9.7 shows the resulting mean polarization leakage of the optimization against the number of evaluations needed to converge for different optimization runs. The annular groove scenario (AG) is shown as circular markers, while the nanofin scenario (NF) is shown as diamond-shaped markers. In addition, the figure compares the performance of a classical direct optimization (D) as green markers and the surrogate model optimization (S) as orange markers. It can be clearly seen that the surrogate model converges after less evaluations than the classical direct optimization. At the same time, both models lead to designs of similar polarization leakage on the order of 0.02. The surrogate-based optimization leads to designs of low leakage more often than the direct optimization.

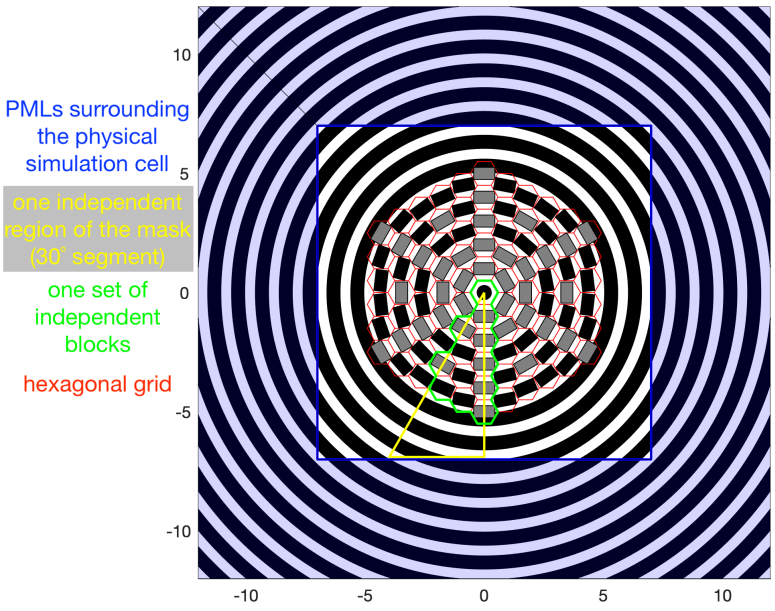

Figure 9.6: Hexagonal pixelization grid of the rectangular nanofin design illustrating the number of free parameters. The independent green highlighted region is replicated twelve times to fill the full pattern. The blocks highlighted in gray have fixed position in azimuthal direction, reducing the number of free parameters of this design to 38.

Figure 9.7 also shows the best performance of the optimization of the AGPM central region as discussed in Chapter 7 obtained from a brute-force parameter search with 625 evaluations. Both scenarios presented in this chapter lead to designs with slightly lower leakage. Overall, the results presented in this chapter show that a design featuring circular grating lines at the center performs better than a nanofin design when applied to the AGPM central region.

This chapter focused on an optimization framework developed for the AGPM central region. However the surrogate optimization approach may be applied to other designs of vector vortex phase masks, like the designs of higher topological charge presented in Chapter 8. The optimization of other designs might be carried out in a future project utilizing this framework.

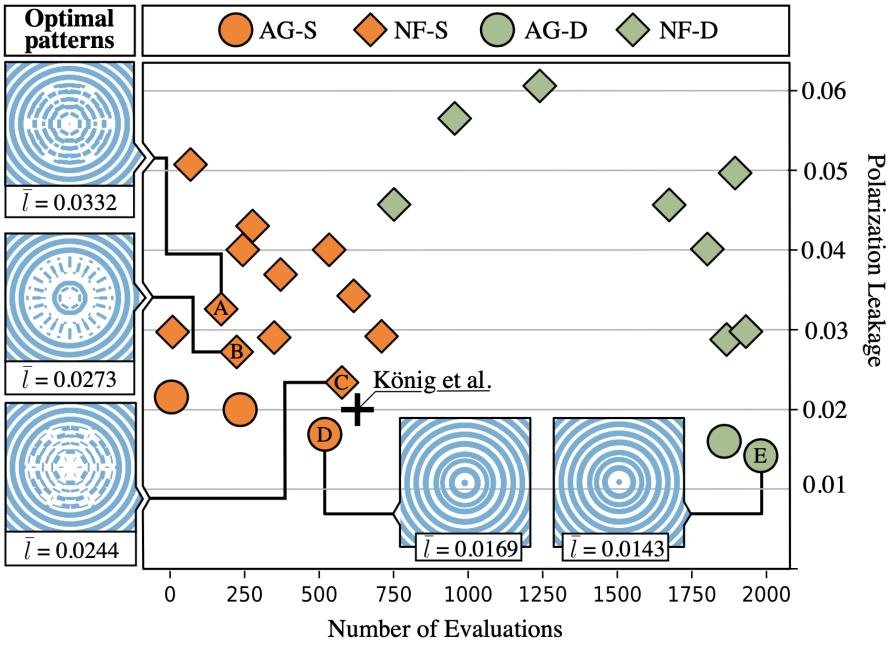

Figure 9.7: Results of the optimization process for both annular grooves and rectangular nanofin scenarios. The polarization leakage is shown against the number of evaluations needed for convergence using (orange) the surrogate optimizer presented in this chapter and (green) a classical direct optimizer. Both annular groove (AG) and nanofin (NF) designs are shown, as well as the results of the AGPM center optimization of Chapter 7 of this thesis (black cross marker). Finally, four promising designs are shown as insets along with their mean polarization leakage $\bar{l}$.

10 Metasurfaces for vector vortex phase masks

Subwavelength grating vortex phase masks of topological charge other than 2 lead to patterns with spatially varying grating parameters. These designs can only be implemented using segmented patterns, as described in Chapter 8. However, the pattern transitions between segments degrade the null depth performance, as FDTD simulations revealed. To mitigate this issue, a metasurface framework consisting of rectangular building blocks (nanofins) is explored in this chapter. The nanofin orientation is expected to vary gradually across the mask in order to exploit the geometric phase, making the pattern locally periodic and therefore avoiding the transition regions encountered with segmented designs. This framework is therefore a promising approach for imprinting a quasi-continuous phase pattern, which is in principle only limited by the size of the metasurface blocks, and more specifically the size of a metasurface unit cell.

10.1 Local birefringence with rectangular nanofins

The subwavelength grating vortex phase masks presented in the previous chapters rely on the control of the geometric phase imposed by the spatially variant subwavelength grating acting as a half-wave plate. The local orientation of the birefringence axis of the half-wave plate determines the geometric phase imprinted onto the beam. In the case of subwavelength gratings, the birefringence axis is aligned with the grating lines. The geometric parameters of the subwavelength grating are optimized using RCWA assuming periodic boundaries, that is for continuous parallel grating lines. For a framework consisting of rectangular nanofins, the geometric phase is expected to be defined by the birefringence axis, which itself depends on the local material anisotropy. The material anisotropy is defined by the orientation of the nanofins acting as half-wave plate. The geometric parameters of the nanofins can be optimized with RCWA because the pattern is locally periodic (except at the center, where FDTD is needed).

Similar to the AGPM, a vector vortex phase mask can be implemented with rectangular nanofins arranged in a circularly symmetric pattern. This pattern imitates the circular grooves of the AGPM, and introduces an additional degree of freedom (length of the nanofins in addition to their width). This additional freedom can be used to improve the chromatic half-wave plate performance of the phase mask. The optimization of

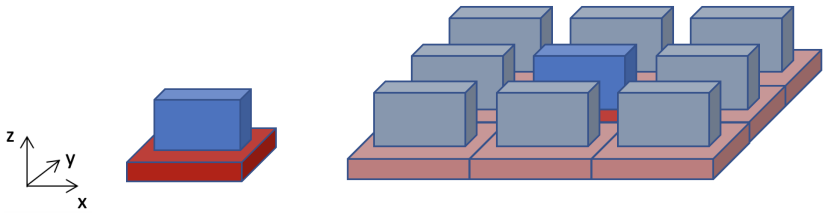

Figure 10.1: Metasurface framework based on rectangular nanofins. Nanofin aligned with the unit cell axes (left) and periodic pattern used for RCWA simulations (right). The unit cell at the center is highlighted, and the neighboring metasurface pixels are shown with pale colors. The local periodicity is ensured for designs where the block parameters vary slowly across the phase mask. Light is incident along the z-axis.

the nanofin geometric parameters is performed with the two-dimensional RCWA code RETICOLO presented in Chapter 4. In a first step, the nanofins are assumed to be oriented parallel to the unit cell axes, as shown in Figure 10.1. This choice is reasonable since the nanofins pattern for a charge-2 vortex imitates the circular groove pattern of the AGPM, and the nanofins are arranged in a polar grid, with their axes along the radial and azimuthal direction.

The nanofin parameters were optimized with RETICOLO and the results are shown in Figure 10.2. The optimal nanofin parameters are close to the parameters of a continuous grating. This is somewhat expected, since the half-wave plate behavior is based on the anisotropy of the nanofins, which is maximized for long rectangles merging into continuous grating lines in one direction. However, the simulations show that a slightly better performance can be achieved when a small gap between the nanofins is present in the azimuthal direction of the AGPM pattern. The physical cause of this behavior has not been investigated further within this thesis and could be the subject of future work.

10.2 Design pixelization

The design optimization carried out with RCWA assumes nanofins that are aligned with the unit cell axes. This is a good description of the charge-2 pattern imitating the circular grating of the AGPM. However, the main advantage of a nanofin framework over subwavelength gratings is that virtually any continuous phase pattern can be imprinted on a mask, especially the helical phase ramps of vortices of topological charge other than 2. The polar pixelization grid assumed in the previous section is a good choice for a charge-2 vortex, but is inconvenient for charge-4. Ideally, a pixelization grid with uniform pixel size following the theoretical charge-4 design shown earlier in Figure 2.12 could be chosen, but no such pixelization grid exists, as shown in Figure 10.3.

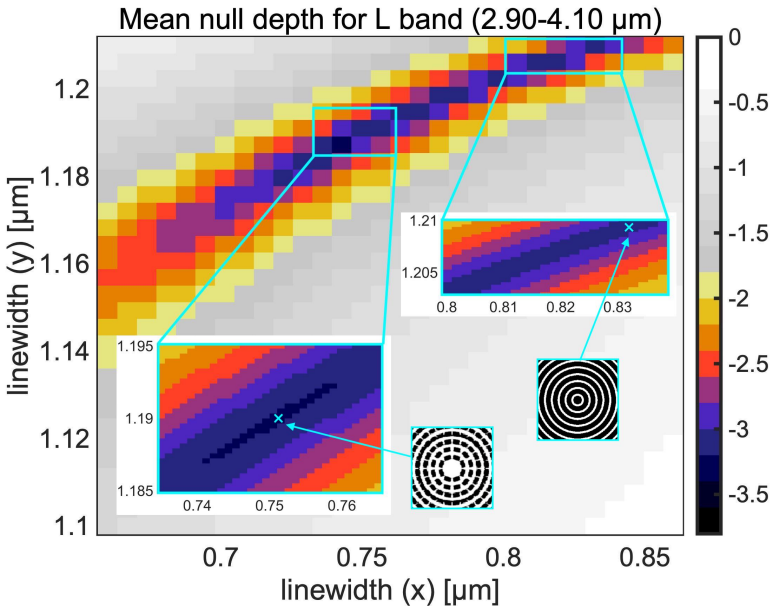

Figure 10.2: Averaged null depth N of a metasurface framework based on rectangular nanofins. The null depth is averaged across the L-band ($2.9 - 4.1\,\mu m$) and shown for the parameter space spanned by the nanofin width and length for the optimal nanofin height of $5.03\,\mu m$. The two insets show the optimal nanofin parameters (left, $N = 9.8 \times 10^{-4}$) and the point of the parameter space corresponding to continuous grating lines (right, $N = 1.3 \times 10^{-3}$), showing that a small gap of $\sim 100\,nm$ could reach slightly better performance than continuous grating lines. The simulations assume diamond nanofins with straight sidewalls ($\alpha = 0°$).

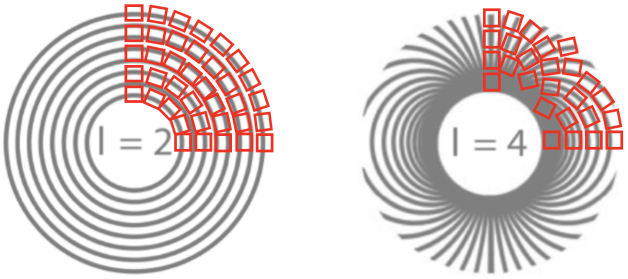

Figure 10.3: Ideal pixelization grids for charge-2 (left) and charge-4 (right) vortex. The charge-2 vortex pattern can be pixelized following a polar pixelization grid featuring nearly square pixels, while a pixelization grid following the charge-4 pattern results in a pixelization grid with large gaps.

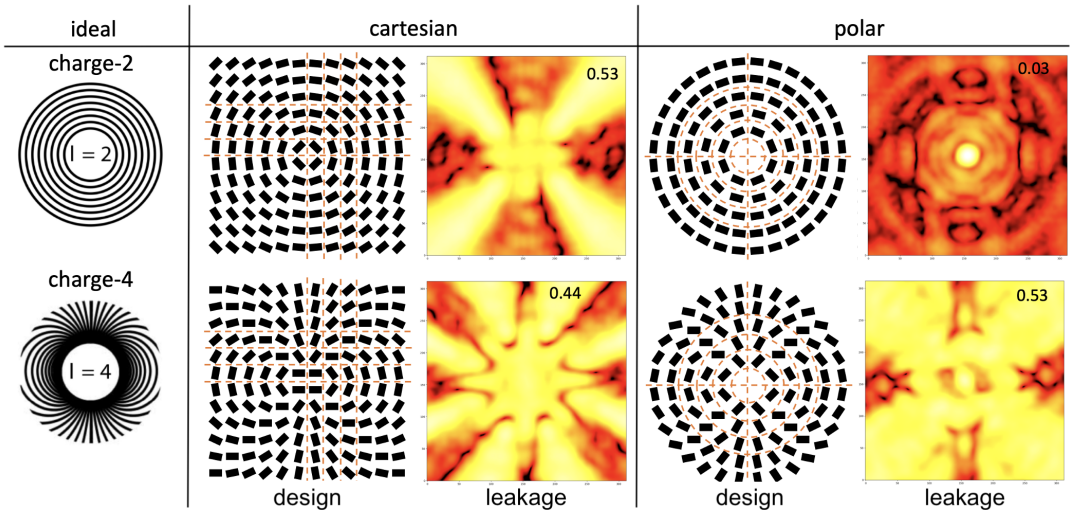

Figure 10.4: Pixelization grids for a nanofin metasurface framework. The theoretical design for a charge-2 (top) and a charge-4 (bottom) vector vortex phase mask is shown in the left column, followed by the design and simulated leakage maps of the corresponding nanofin pattern for a cartesian (left) and polar (right) pixelization grid. The number on the leakage maps is the average leakage in the considered region showing that only the polar pixelization of the charge-2 vortex (0.03) reaches an acceptable performance.

To exploit the flexibility of a nanofins metasurface framework, the nanofins have to be arranged conveniently in a general pixelization grid which provides the local periodicity needed for RCWA simulations. Using an optimizer based on RCWA is preferable because of the greatly reduced computational resources needed compared to FDTD. This also makes it more suited to perform large sweeps of the parameter space. However, designs with varying spacing could benefit from the more accurate simulation using FDTD, but are unrealistic for large regions of the mask.

The simplest approach to discretize a phase pattern using metasurface building blocks is to arrange them in a cartesian pixelization grid. The geometric phase is imparted pixel by pixel by changing the orientation of the nanofins, resulting in a quasi-continuous phase pattern only limited by the metasurface pixel size. However, for specific designs, other pixelization approaches are more favorable, such as the polar pixelization grid used in the previous section. Figure 10.4 shows the nanofin designs for a charge-2 and a charge-4 vortex phase mask for these three different pixelization schemes, providing very different results. Only the polar pixelization grids for a charge-2 vortex provide reasonable performance. More precisely, the leakage maps show that the half-wave plate character is only fulfilled for nanofins parallel to the cell axes. This is due to the periodic environment used to optimize the nanofin parameters with RCWA to provide a half-wave plate. For nanofins that are not aligned with the

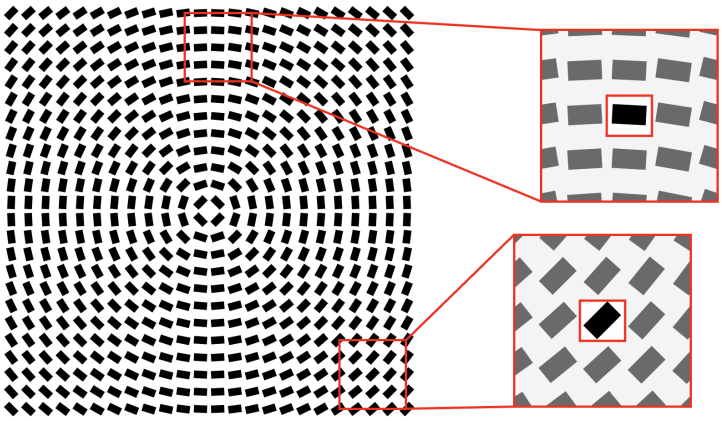

Figure 10.5: Locally varying effective periodicity of the cartesian nanofins patterns. A charge-2 vortex phase mask pattern is shown along with two insets showing different local effective periodicity of the pattern. While the geometric parameters were optimized for an orientation of the nanofins along the unit cell axes (top right inset), the same parameters are not optimal for a nanofin rotated by 45° (bottom right inset). Therefore, the size of the nanofins has to be re-optimized for each rotation angle of the nanofin.

pixelization axes the local periodicity changes, and the nanofins do not behave as a half-wave plate anymore.

In order to restore the half-wave plate character, the parameters of the rotated nanofins have to be re-optimized. Figure 10.5 shows the different arrangement of the nanofins at different parts of the cartesian charge-2 vortex phase mask design. Two regions of the mask behave differently due to the different local periodicity. For example, the pattern with nanofins rotated by 45° imitates straight grating lines with an effective period of 71 % of the unit cell size in the direction of the nanofin axes.

Optimizing the nanofins locally can restore their half-wave plate character at different rotation angles. However, the accurate simulation of the phase and amplitude response of nanofins that are not aligned with the unit cell axes is challenging and can only be done for certain rotation angles, as discussed in Section 4.1. Re-optimizing the nanofin parameters leads to the pattern of a charge-2 vortex phase mask shown in Figure 10.6 with different sectors of locally constant nanofin size. This pattern has been simulated with FDTD, showing that re-optimizing the size of the nanofins indeed restores the half-wave plate character of the mask, providing low leakage for all orientations. However, the results also show that the phase does not follow the desired characteristic helical shape of a charge-2 vortex. The reason for this result is that re-optimizing the nanofin parameters results in a different absolute phase response of the nanofins in different sectors of the mask. This arises from the different effective refractive index of the nanofins, which depends on their geometric parameters.

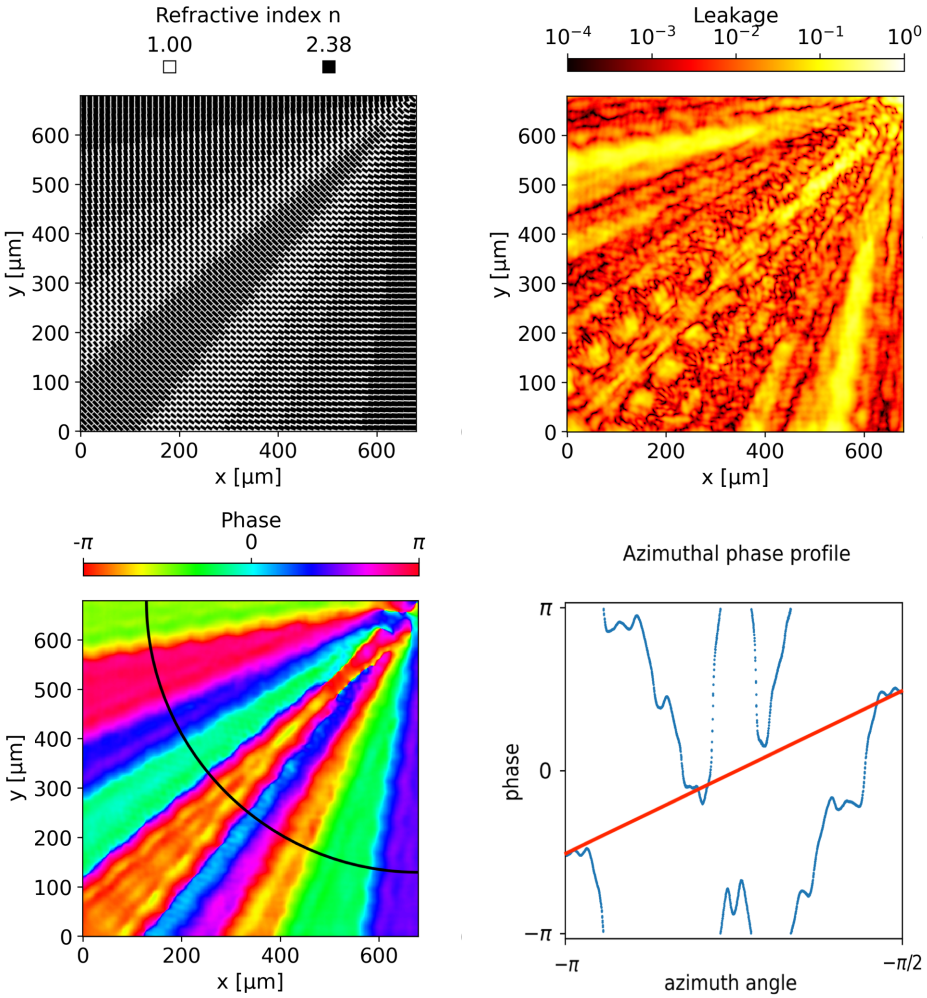

Figure 10.6: Design, leakage, phase, and azimuthal phase profile (from left to right) of one quadrant of the re-optimized charge-2 nanofin pattern simulated with FDTD. The phase is not fully defined by the orientation of the nanofins and acquires an additional phase depending on the local nanofin parameters. The phase response therefore does not show the expected helical phase ramp (shown as red curve in the right most graph). The azimuthal phase profile is obtained from a cut along the black line shown in the phase response.

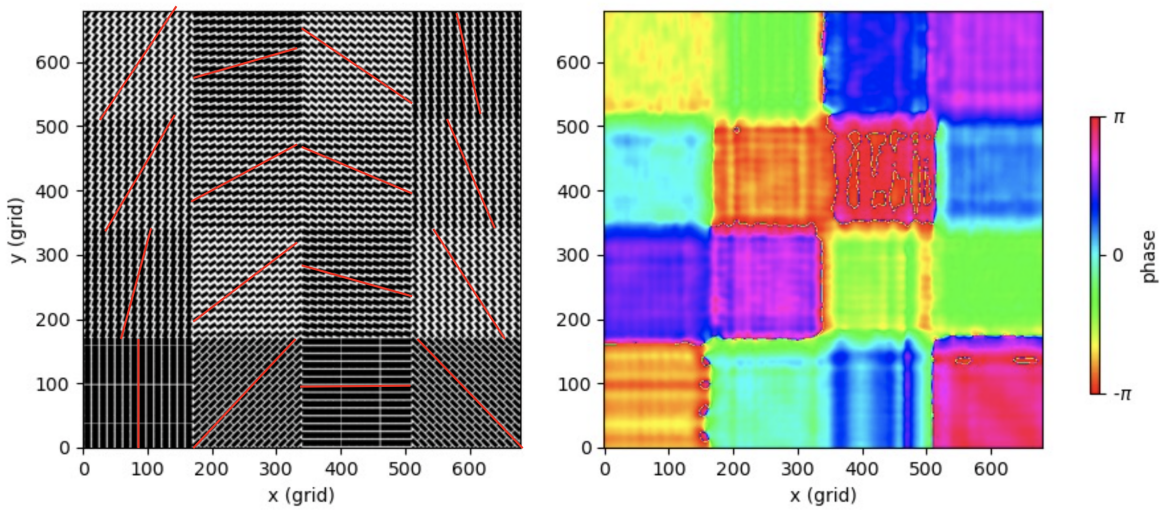

Figure 10.7: Design (left) and phase (right) of different nanofin patterns with varying orientation simulated with FDTD. The patterns are arranged in a 4×4 grid with gradually increasing rotation angle. The orientation of the nanofins is illustrated as a red line in each segment. The simulated phase does not follow the expected relation $\phi_g = 2\theta$, where ϕ_g is the geometric phase and θ is the angle between the birefringence axis and the vertical axis.

In addition, the birefringence axis of the nanofin patterns is not fully defined by the orientation of the nanofins, but rather by the pixelization scheme. Figure 10.7 shows the phase response of a series of patterns featuring gradually rotated nanofins and their phase response simulated with FDTD. The orientation of the nanofins is shown as red line in each pattern. For rotation angles other than 45° the nanofins imitate a grating along the cartesian pixelization axes, resulting effectively in a slanted grating. The effective birefringence axis is therefore oriented in the same direction. For a rotation angle of 45° the nanofins imitate a continuous grating along the diagonal of the pixelization axes with their effective birefringence axis oriented in the same direction. The resulting birefringence axis of the pattern is therefore aligned with the horizontal, vertical or diagonal direction of the pixelization axes. The spatially varying rotation of the nanofins is therefore not suited to imprint arbitrary phase patterns purely based on the geometric phase.

10.3 Application of metasurfaces for vector vortex coronagraphs

The nanofins framework presented in this chapter has been investigated as a promising implementation of arbitrary phase patterns based on the geometric phase. However,

the results obtained with FDTD simulations show that this approach can not readily be implemented for two reasons. First, the half-wave plate character of a periodic nanofin pattern is optimized for a certain orientation of the nanofins compared to the pixelization axes, and therefore the geometric parameters of the nanofins have to be re-optimized for each nanofin orientation. However, re-optimizing the nanofins locally results in an additional phase due to the resulting varying effective filling factor. Second, the pixelization scheme influences the effective birefringence axis of the nanofins pattern, which is no longer defined by their orientation. Both effects make it challenging to implement a nanofin metasurface phase mask based on the geometric phase.

Though the results presented in this chapter suggest that a nanofin vector vortex phase mask design can not be readily implemented, a few designs based on rectangular nanofins have been demonstrated in the literature (Liu et al., 2020; Khorasaninejad et al., 2016; Devlin et al., 2017). These designs are often paired with polarization filtering which removes the polarization leakage, but results in significant throughput loss, and are therefore not suited for coronagraphic applications. Even though the application of this nanofin metasurface framework to the design of vortex phase masks is not readily possible, it is important to note that other applications based on rectangular nanofin patterns can be envisaged. However, these applications are beyond the scope of this thesis.

11 Metasurfaces for scalar vortex phase masks

The previous chapters have focused on the development and optimization of microstructured vector vortex phase masks. While these masks have successfully been used on major ground-based telescopes, future space-based telescopes require contrast values currently challenging to achieve with a vector vortex phase mask due to its polarization sensitivity. This chapter focuses on the implementation of a polarization-independent (scalar) vortex coronagraph using a metasurface-based phase mask, which could reach the contrast values required for space-based telescopes to directly image an Earth-like planet around a Sun-like star.

11.1 Scalar vortex coronagraph for terrestrial planet imaging

There have been several recent technological advancements in coronagraphic techniques for the pursuit of high contrast imaging, but no methods have so far demonstrated the capability of directly capturing light from an Earth-like planet around a Sun-like star. So far, the vortex coronagraph has proven to be one of the most promising approaches for directly characterizing exoplanets and achieving this goal. Efforts toward advancing vortex coronagraph technologies are motivated by the science driven requirements of the 10^{-10} contrast[1] needed to detect Earth-like planets, as proposed in the mission concepts HabEx (Gaudi et al., 2020) and LUVOIR (LUVOIR final report, 2019), which converged into NASA's Habitable Worlds Observatory.

The direct imaging goals for the future flagship mission require a coronagraph that achieves a contrast of 10^{-10} at its inner working angle for a 20 % bandwidth. Both HabEx and LUVOIR-B feature a vector vortex coronagraph of topological charge 6 as the most suitable technology for starlight suppression, providing an inner working angle of 2.5 λ/D. Using wavefront control, contrast values on the order of 2×10^{-9} have been demonstrated in lab for a bandwidth of 10 % using a vortex phase mask based on liquid crystal polymers (Ruane et al., 2022). However, differential polarization

[1] The contrast values used in this chapter refer to the contrast in a certain region in the focal plane, typically from 3 to 10 λ/D, as described in Section 5.5.

aberrations become difficult to control at such high contrasts, requiring careful wavefront control in both polarizations independently. This is challenging to achieve with a single or even multiple deformable mirrors because the vector vortex coronagraph is polarization-sensitive and imprints phase ramps of opposite signs on the two polarization states, requiring the polarizations to be split. This results in considerably more complex instrumental designs basically doubling the number of optics or eliminating half of the total throughput per channel. A polarization-independent (scalar) vortex coronagraph circumvents this issue since it imprints the same phase ramp regardless of polarization and is therefore an interesting alternative to the well established concept of the vector vortex coronagraph. Understanding the chromatic behavior of the scalar vortex coronagraph is crucial in achieving the contrast goals required for future space telescope missions. The problem of chromaticity is fundamentally different for vector and scalar vortex.

The vector vortex coronagraph exploits the geometric phase of a birefringent material, which results in a perfect phase ramp imprinted on the beam for all wavelengths. Both circular polarizations acquire perfect phase ramps, but of opposite topological charge. The primary limitation of the vector vortex is the polarization leakage resulting from the imperfect half-wave plate character of the phase mask across a certain bandwidth. The performance of vector vortex phase mask is therefore primarily characterized by the polarization leakage propagating unaffected to the image plane, where it creates a scaled-down version of the stellar PSF. The achievable contrast is therefore directly linked to the amount of light leaking through the coronagraph and can readily be used as performance metric without explicitly propagating the fields to the final coronagraphic image. A second effect becomes limiting in the presence of aberrations at very high contrasts, where wavefront control is used to dig a dark hole in the image plane. Since both polarizations acquire phase ramps of opposite helicity, wavefront control is required in both polarizations independently, which is challenging to achieve even with multiple deformable mirrors. To tackle this issue, the polarizations can be split in two separate channels increasing the complexity of the instrument; alternatively one polarization can be filtered using a polarizer-analyzer setup, at the cost of losing half of the light.

The chromaticity of the scalar vortex coronagraph arises from a different effect. Scalar vortex phase masks use spatially variant longitudinal phase delays to create a certain phase response. This concept leads to the same phase ramp regardless of polarization, which is advantageous for the use with wavefront control algorithms, but suffers from chromaticity. In fact, the phase ramp for typical scalar vortex implementations based on a helical shaped dielectric mask is only perfect for the design wavelength because it is based on a physical path length rather than a fixed phase delay. The phase response of such a component depends linearly on the wavelength (provided the refractive index is constant across the considered bandwidth). This effect results in phase ramps of different steepness and therefore different (non-integer) topological charge, which leads to a polarization-independent leakage component. While these chromatic errors are more important than in the case of the vector vortex coronagraph, wavefront control

algorithms can be used to dig a dark hole in a limited scoring region in the image plane, increasing the contrast by several orders of magnitude. This technique is a promising approach towards achieving very high contrasts in large bandwidths.

Scalar vortex coronagraphs using longitudinal phase delays have been shown to yield performance on the order of 10^{-8} in narrowband, and 2×10^{-7} across 10 % bandwidth, showing encouraging results so far (Desai et al., 2023). Stacking several layers is one possible solution to increase broadband performance but usually results in bulky components which are not suitable for use in the focal plane (Swartzlander, 2006; Ruane et al., 2019).

The achromatic performance of scalar vortex phase mask designs remains a challenging goal. Compared to existing scalar vortex phase mask technologies based on the longitudinal phase delay of different dielectrics, a metasurface framework benefits from a better broadband performance within a single layer. This is due to the different metasurface topologies which introduce new degrees of freedom by changing the shape of their constituent building blocks. As a result, metasurfaces are a promising approach to increase the broadband performance of the scalar vortex coronagraph, which can be implemented with relatively simple and well established fabrication techniques. However, to date no study of a metasurface-based scalar vortex coronagraph for high contrast imaging applications has been reported.

11.2 A polarization-independent metasurface framework

Metasurfaces are thin structures consisting of subwavelength building blocks conveniently arranged to create a specific phase and amplitude response to an incoming beam. Recent developments have shown that metasurfaces can be engineered to provide precise 2π phase coverage and high throughput across large bandwidths (Yu & Capasso, 2014; Kruk et al., 2018; Khorasaninejad et al., 2016). This makes them a prime candidate for achromatic high-performance coronagraphs. Several applications of metasurfaces generating vortex beams have been reported (Devlin et al., 2017; Chong et al., 2015; Shalaev et al., 2015) and recently significant progress has been made in making metasurface components achromatic across large bandwidths (Heiden & Jang, 2022; Sun et al., 2021; Li & Fan, 2022). Polarization-independent metasurfaces consisting of circular or square nanoposts are of particular interest because they do not create polarization leakage, which is limiting the current vector vortex coronagraphs using subwavelength gratings or liquid crystal polymers. Compared to existing scalar vortex phase mask technologies based on the longitudinal phase delay of different types of glasses, a metasurface framework has the potential to provide better achromatic performance within a single layer due to the additional freedom introduced by different metasurface geometries. This avoids stacked designs and greatly increases

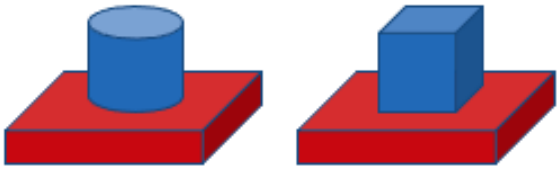

Figure 11.1: Metasurface framework based on circular nanopillars (left) and square nanoposts (right). Both frameworks are polarization-independent and therefore suited for the design of scalar vortex phase masks.

our ability to make the scalar vortex coronagraph achromatic and ready for the needs of future space telescope missions.

The design process of scalar vortex phase masks starts from identifying a promising metasurface framework. Here, two similar frameworks are considered: circular nanopillars and square nanoposts, as shown in Figure 11.1. These are the simplest shapes for a polarization-independent metasurface framework, and both can easily be arranged in a cartesian pixelization grid to sample the full phase mask. More complicated shapes can be envisaged, like hexagonal nanoposts on a hexagonal grid (Arbabi et al., 2015), designs with varying spacing between the blocks (Chong et al., 2015) or shapes with holes or crosses (Shrestha et al., 2018; Chen et al., 2023), introducing more degrees of freedom which could further improve broadband performance in the future.

The key property of metasurfaces is their phase and transmission response, which is controlled by the size and height of the blocks, as well as their distribution across the mask. For the metasurface masks considered here, the height of the blocks will be constant across the phase mask, and has to be chosen high enough to enable 2π phase coverage. At the same time, the height of the metasurface should be small enough to envision manufacturing of a prototype in the near future. Typically, the manufacturing of diamond nanostructures with aspect ratios of up to 1:13.5 is well mastered, as discussed in Section 2.2.3, which corresponds to an etch depth of 10 µm for a grating period of 1.42 µm and 50 % filling factor. Therefore, this limit on the block height will be used for the present design analysis, although the sidewall angle induced by current manufacturing technologies will ultimately limit the achievable etch depth, depending on the opening size between the metasurface blocks. As a first step, only straight sidewalls will be considered here, to demonstrate the capability of metasurfaces of achieving the ambitious contrast goals.

A scalar vortex phase mask design needs a metasurface framework providing optimal phase coverage to be able to imprint a helical phase ramp on the beam. At the same time it has to maintain relatively constant transmission across the mask, because the vortex effect relies on destructive interference of light at different positions of the mask, which requires destructively interfering beams to have the same intensity. This is ideally achieved by keeping the intensity constant over the full mask. In addition, high throughput is crucial in high contrast imaging of exoplanets, which is an additional

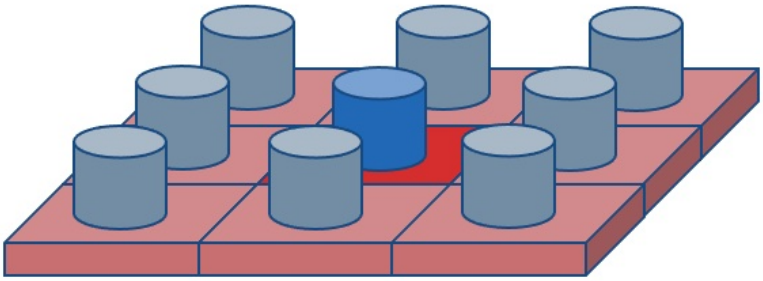

Figure 11.2: Circular nanopillar unit cell used in an RCWA simulation. The unit cell at the center is highlighted, and the neighboring metasurface pixels are shown with pale colors. The local periodicity is ensured for designs where the block parameters vary slowly across the phase mask.

motivation to have the highest possible transmission everywhere on the mask. To fulfill these requirements, the parameter space spanned by block size and metasurface height has to be explored for the different metasurface frameworks to provide both optimal phase coverage and relatively constant high transmission.

To assess the phase and transmission properties of a metasurface framework, a series of RCWA simulations are performed. Using RCWA implies using a periodic and infinite pattern of metasurface building blocks, as illustrated in Figure 11.2. Locally, these conditions are fulfilled when a metasurface block is surrounded by other blocks of similar geometrical parameters. This is the case for most parts of the phase masks considered in this chapter, except at the 2π phase jumps and the central singularity. These regions of the phase mask that do not fulfill the local periodicity condition will be assessed in a second step with FDTD to characterize the effect introduced by the phase discontinuity.

The phase and transmission response of a periodic array of circular diamond nanopillars is shown in Figure 11.3 for an exemplary metasurface height of 9.5 µm and a periodicity of 1.4 µm (for application in the astronomical *L*-band, $3.4 - 4.2$ µm). The RCWA simulation is carried out for different pillar sizes (x-axis) at a wavelength of 3.8 µm. The most important quantity to assess the metasurface framework is the phase coverage of 2π needed to imprint a helical phase ramp. The varying phase pattern of a vortex phase mask is achieved by spatially varying the optical path length across the mask. This is achieved by locally changing the effective refractive index, which itself depends on the local filling factor of the metasurface design. The achievable phase coverage increases with increasing metasurface height and therefore a certain minimum height is needed to provide full 2π phase coverage.

To explain the achievable phase coverage, consider a single metasurface pixel. A circular nanopillar can not fill the pixel unless it is extended and truncated beyond the edge of the pixel, as shown in Figure 11.4. Varying the radius of the nanopillar from 0 to $0.5\,\Lambda$ therefore does not result in an effective filling factor coverage of 0 to 1, but

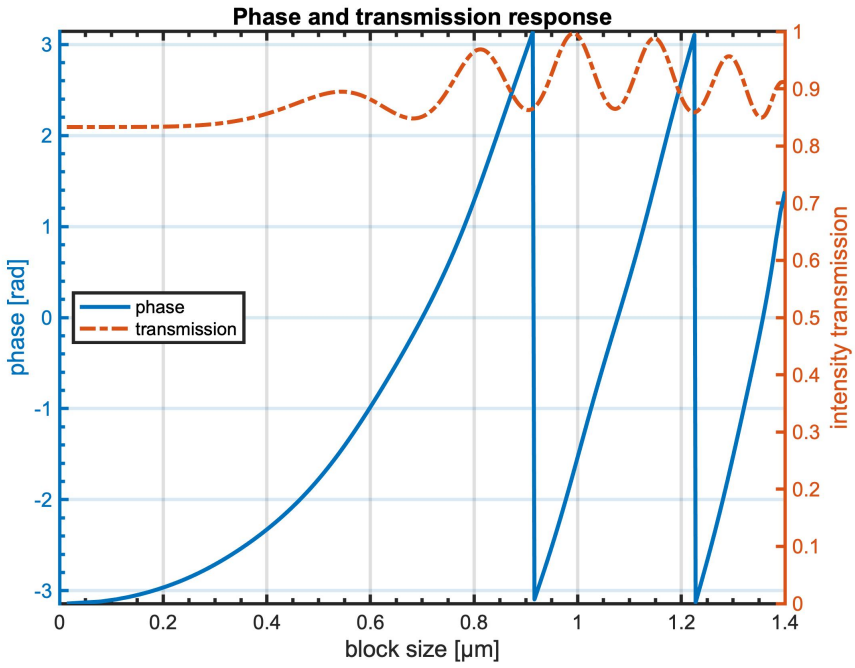

Figure 11.3: Phase (blue solid line) and transmission (orange dashed line) response of a circular nanopillars metasurface at fixed height of 9.5 µm calculated with RCWA.

0 to $\pi/4 \approx 79\,\%$. This means that the effective filling factor is varying less rapidly, which reduces the range of effective filling factors that can be achieved with a circular nanopillar metasurface framework of fixed height compared to a square nanopost framework, and therefore results in a smaller phase coverage at similar manufacturing complexity. Figure 11.5 shows that a metasurface framework consisting of square nanoposts is more adequate to provide 2π phase coverage than circular nanopillars because 2π phase coverage is reached at a smaller height. Therefore for the following optimization a metasurface framework using square nanoposts is considered.

The phase and transmission response of a square nanopost periodic array is shown in Figure 11.6. Compared to the circular nanopillar results shown earlier in Figure 11.3, the phase curves for square nanoposts are steeper, meaning that 2π phase coverage is achieved with a smaller range of block sizes. This is favorable for manufacturing of these components, since large variations of the filling factor make manufacturing processes based on plasma etching challenging because of the microloading effect, resulting in uneven etch depth. The microloading effect is explained in Section 11.7 at the end of this chapter. Furthermore, a metasurface design using a smaller range of block sizes results in neighboring metasurface blocks being more similar to each other, which reduces the errors due to the pixelization grid in regions of fast spatial phase variations, like at the central singularity of the vortex phase ramp. Due to its advantages over the circular nanopillar framework, the square nanopost framework is chosen for the first

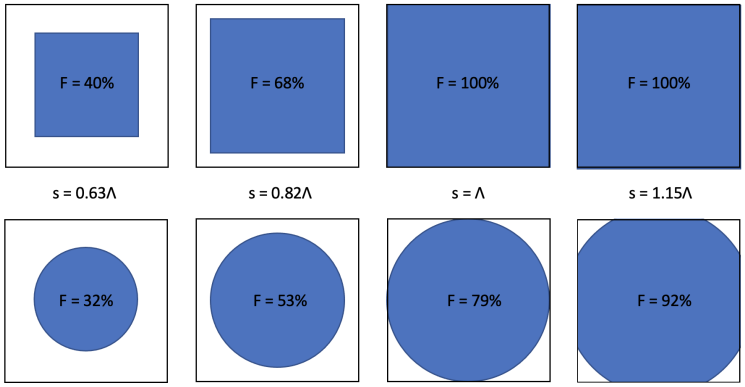

Figure 11.4: Effective filling factor for a metasurface framework consisting of square nanoposts (top) and circular nanopillars (bottom) for different block diameters s (left to right). The square nanoposts fill the pixel fully while the circular nanopillars have to be truncated to reach the same range of effective filling factor.

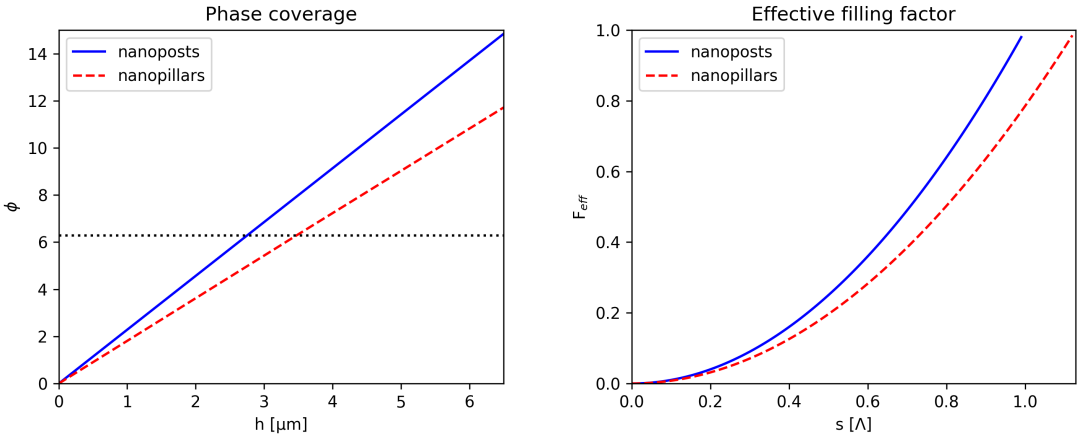

Figure 11.5: Left: Phase coverage of the metasurface framework for different block height. 2π phase coverage (black dotted line) is achieved starting from a certain block height. This characteristic height is higher for circular nanopillars (red dashed curve) than it is for square nanoposts (blue solid curve). Right: Effective filling factor achieved with a circular nanopillar (red dashed curve) and a square nanopost metasurface framework (blue solid curve). The effective filling factor is shown against the diameter of the blocks. A larger range of effective filling factors can be achieved with a square nanopost metasurface framework.

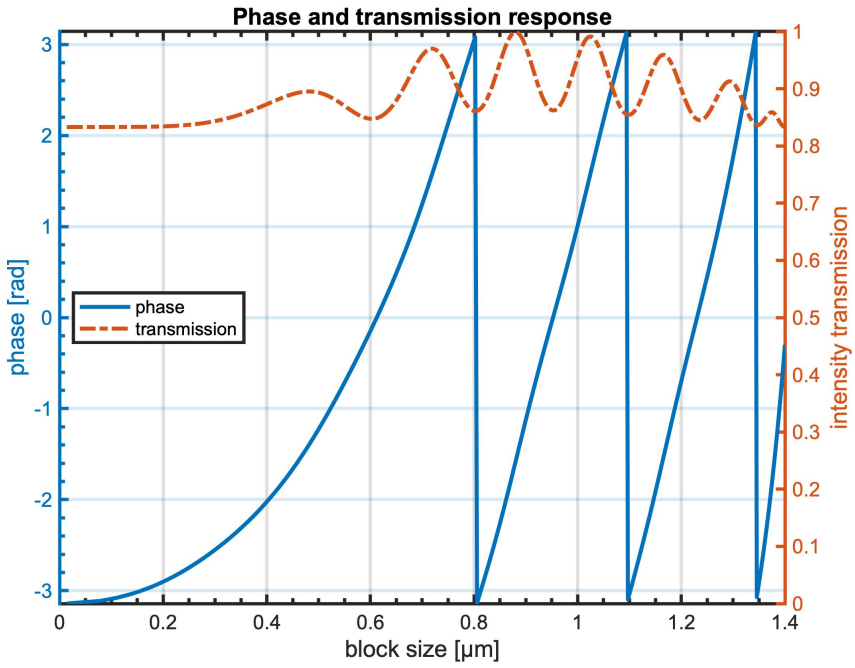

Figure 11.6: Phase (blue solid curve) and transmission (orange dashed curve) response of a square nanopost metasurface framework at fixed height of 9.5 µm calculated with RCWA.

scalar phase mask designs presented here. However, a circular nanopillar metasurface framework might still be a valid option for other applications of phase masks, which depend in a different way on the phase and transmission than the vortex.

Another important quality of a metasurface framework for vortex phase masks is the uniformity of the transmission. A perfect vortex phase mask is a pure phase mask with uniform transmission. Therefore, a non-uniform transmission pattern arising from a non-uniform metasurface pattern has to be reasonably small. Typically, the 20 % transmission variation encountered in the metasurface frameworks considered in this chapter results in a peak-to-peak leakage term of the order of 10^{-6}, which is acceptable for the scalar vortex phase mask applications discussed here. It is interesting to note that a non-zero sidewall angle has an anti-reflective effect (Forsberg & Karlsson, 2013b) and therefore result in smaller transmission variations. A design featuring slightly slanted sidewalls is therefore more favorable in terms of transmission uniformity. Figure 11.7 shows that the transmission variation for a sidewall angle of 2° is smaller, at the expense of a shallower phase curve requiring a larger range of block sizes to reach 2π phase coverage. On the other side, slanted sidewalls reduce the range of achievable filling factors for a design, since the sidewalls might touch at the bottom of small gaps in the design, which leads to merging blocks in those regions of the mask. In total, the full 2π phase coverage is found to be more important to reach high contrasts than a uniform transmission, and therefore designs featuring vertical sidewalls will be pre-

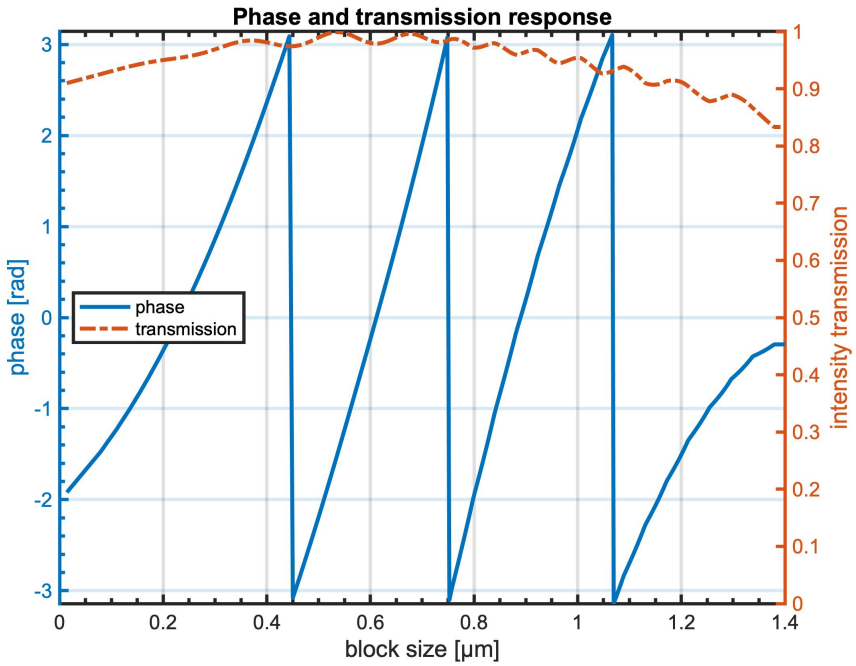

Figure 11.7: Phase (blue solid curve) and transmission (orange dashed curve) response of a square nanopost metasurface framework with trapezoidal side walls of 2° at fixed height of 9.5 µm calculated with RCWA. The slanted sidewalls act as an anti-reflective layer making the transmission more uniform (3 % variation for the optimal post size range). Due to the sidewall angle, the best performance is achieved for smaller blocks. The block size refers to the size at the top of the metasurface layer, and increases linearly when moving towards the substrate.

ferred in this chapter. For manufacturing perspectives however, vertical sidewalls are challenging to achieve and the performance of designs featuring slanted sidewalls needs to be estimated when envisaging manufacturing of a prototype.

11.3 Optimizing for broadband performance

While achromatic metasurface implementations in the literature focus mostly on the absolute phase response as used in metalenses (Chung & Miller, 2020; Chen et al., 2020), there is no need to do so for vortex phase masks. The important metric for a vortex phase mask is to obtain the phase ramp corresponding to the topological charge of the vortex. For scalar vortices made of a helical shaped dielectric mask, the phase is highly chromatic. For a metasurface framework, the absolute phase response of a certain element is chromatic as well, but the relative phase delay across the mask is much less chromatic.

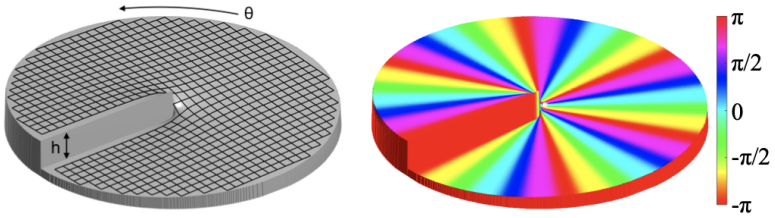

Figure 11.8: Helical shaped dielectric mask (left) and corresponding ideal phase ramp (right) for a charge-6 scalar vortex phase mask design. Adapted from (Ruane et al., 2019).

To understand this effect, consider a helical shaped dielectric mask, as shown in Figure 11.8. The phase response at every part of the mask is:

$$\begin{aligned}\phi(\theta) &= \frac{d(\theta)}{\lambda_n} + \frac{h - d(\theta)}{\lambda} \\ &= \frac{\theta h/2\pi}{\lambda/n} + \frac{h - \theta h/2\pi}{\lambda} \\ &= \frac{h}{\lambda} \cdot \left(1 + \frac{\theta \cdot (n-1)}{2\pi}\right)\end{aligned}$$

where $d(\theta)$ is the height profile of the helical shaped mask, λ the vacuum wavelength, $\lambda_n = \lambda/n$ the wavelength at the refractive index of the mask n, h the physical height of the helical ramp, and θ the azimuthal coordinate. The phase has a reciprocal λ-dependence causing the chromatic behavior. For the vortex effect, the phase accumulated on a closed path around the optical axis needs to be constant: $\Delta\phi = \ell_p \cdot 2\pi$. For the scalar vortex this translates to:

$$\begin{aligned}\Delta\phi &= \phi(\theta = 2\pi) - \phi(\theta = 0) \\ &= \frac{h}{\lambda} \cdot (n-1).\end{aligned}$$

The accumulated phase has therefore a reciprocal dependence on the wavelength.

For a metasurface-based scalar vortex phase mask, the relevant phase shift is acquired in a layer of constant thickness across the mask. The phase shift does not rely on the physical path difference in regions of different refractive index (air and dielectric), but on the phase shift acquired in this single layer, which is achieved by locally varying the effective refractive index:

$$\phi(\theta) = \frac{h}{\lambda/n_{\text{eff}}(\lambda, \theta)}$$

where h is the constant height of the metasurface layer and λ/n_{eff} the effective wavelength, dependent on the effective refractive index n_{eff} of the metasurface pattern at

the wavelength of interest λ. Assuming that the phase ramp scales sufficiently linear with θ, the relevant metric for the vortex effect is:

$$\begin{aligned}\Delta\phi &= \phi(\theta = 2\pi) - \phi(\theta = 0) \\ &= 2\pi \cdot h \cdot \left(\frac{1}{\lambda/n_{\text{eff}}(\lambda, \theta = 2\pi)} - \frac{1}{\lambda/n_{\text{eff}}(\lambda, \theta = 0)} \right) \\ &= \frac{2\pi h}{\lambda} \cdot (n_{\text{eff}}(\lambda, \theta = 2\pi) - n_{\text{eff}}(\lambda, \theta = 0)) \\ &= \frac{2\pi h}{\lambda} \cdot \Delta n_{\text{eff}}(\lambda)\end{aligned}$$

For blocks much smaller than the wavelength, the relation between filling factor and effective refractive index n_{eff} can be described semi-analytically with effective medium theory (Zhang & Wu, 2015). But when the wavelength approaches the feature size of the metasurface, non-linear effects become non-negligible and full wave propagation codes such as RCWA and FDTD have to be used, as described in Chapter 4.

The key property of metasurfaces is that while the phase induced by the metasurface itself might be chromatic, the phase difference $\Delta\phi$ acquired along a closed path around the optical axis of the vortex phase mask can be engineered to be less chromatic across a certain bandwidth by optimizing the size of the metasurface building blocks. This is illustrated in Figure 11.9, which shows the chromatic phase response for a square nanopost framework along with the effective refractive index n_{eff} at several wavelengths. While n_{eff} is chromatic, the phase difference $\Delta\phi$ is relatively achromatic across the 20 % bandwidth considered here.

11.4 Metasurface scalar vortex phase mask design

After finding a metasurface framework providing 2π phase coverage across 20 % bandwidth, a scalar vortex design is produced using this framework. Here, we will focus on a charge-6 scalar vortex phase mask, as specified by the HabEx and LUVOIR mission concepts. This topological charge is small enough to provide relatively small inner working angle ($2.5\,\lambda/\text{D}$) and large enough to be resilient to the lowest order aberrations, as explained in Section 1.3.3.

Different scalar vortex topologies described in Section 2.2.5 have been developed to increase the broadband performance of the scalar vortex coronagraphs (classical, sawtooth and wrapped vortex). Figure 11.10 compares the achievable performance of the different design topologies. Clever phase wrapping can increase the broadband performance by up to one order of magnitude. Scalar vortex phase masks for application at visible wavelengths have been fabricated and tested recently (Desai et al., 2021). Combined with wavefront control, contrasts on the order of 10^{-7} have been reached in 10 % bandwidth.

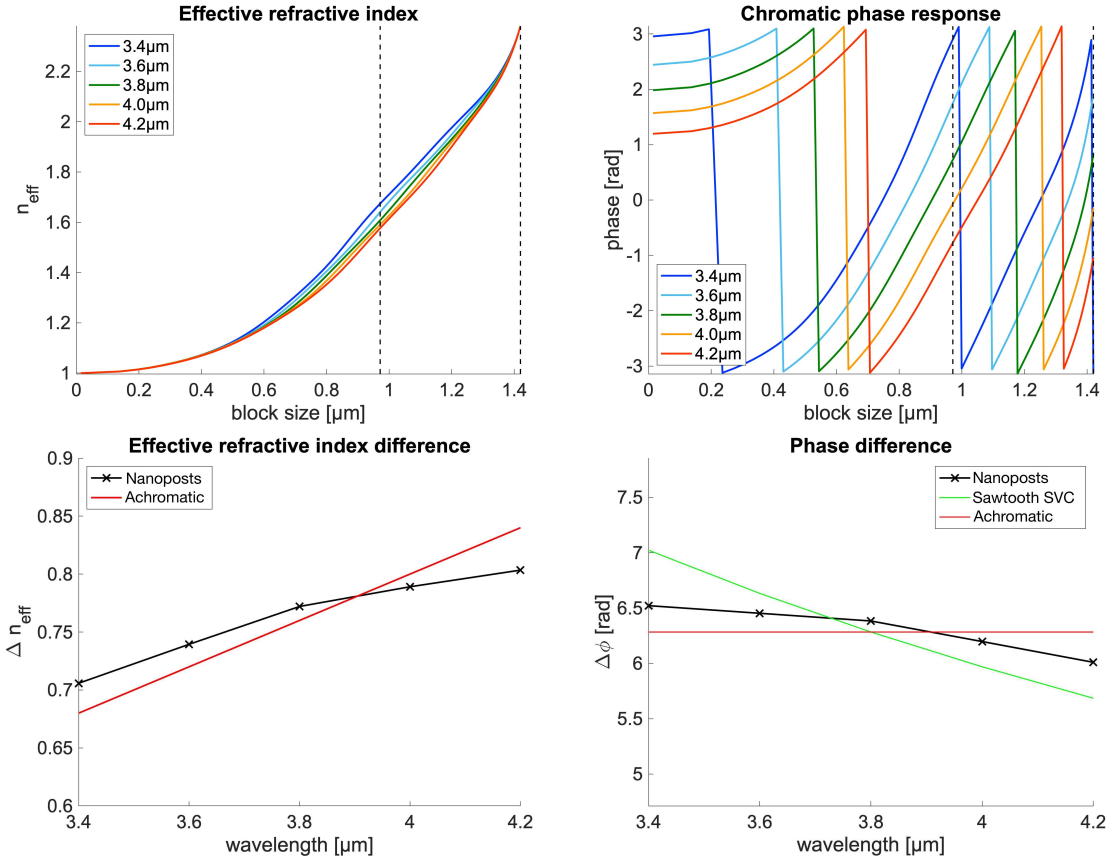

Figure 11.9: Chromaticity of a metasurface framework consisting of square nanoposts. Top: Effective refractive index (left) and phase response (right). A metasurface framework consisting of square nanoposts of relatively shallow height (5 µm) is chosen to illustrate the chromatic behavior of the effective refractive index n_{eff} (a larger height would feature steeper phase curves reducing the readability of the right graph). n_{eff} is obtained from the simulated phase response at fixed height ($\phi = d/(\lambda/n_{\text{eff}})$). The optimal region is highlighted with dashed black lines. Bottom: Effective refractive index difference Δn_{eff} (left) and phase difference $\Delta\phi$ (right) at the 2π phase discontinuities. While n_{eff} itself is chromatic, the acquired phase difference $\Delta\phi$ is relatively achromatic because the linear dependency of Δn_{eff} compensates for the $1/\lambda$ dependency of the phase difference. The achromatic vortex case is shown as red line, and the curve for a sawtooth scalar vortex (as described in Section 2.2.5) made from a helical shaped dielectric mask is shown as green line.

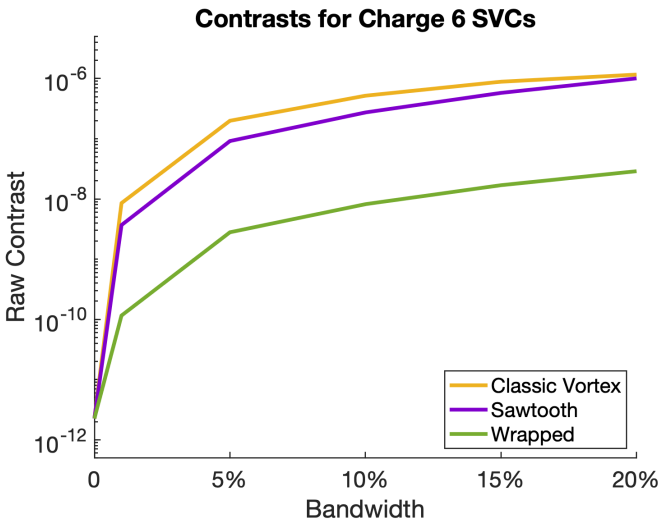

Figure 11.10: Performance of different scalar vortex designs of topological charge 6 using phase wrapping. The raw contrast in a region between 3 and 10 λ/D is shown. From Desai et al. (2022).

For the implementation of a scalar vortex with the present metasurface framework, a sawtooth vortex design is chosen. This choice is motivated by the limited phase coverage that can be achieved with a nanopost framework. The sawtooth design needs only 2π phase coverage, while a wrapped vortex needs 3.6π and a classical vortex needs 12π, as illustrated earlier in Figure 2.14. Since the maximal height coverage of a nanopost framework is limited by the height of the nanoposts (Figure 11.5), the design topology with the smallest phase coverage is chosen for a first nanopost scalar vortex phase mask design. Furthermore, the height and the variation of the block size across the mask has to be kept reasonably small to keep the design within reach for current manufacturing techniques.

Figure 11.11 shows a design for a sawtooth vortex of topological charge 6 using a nanopost framework. The design shown is a closeup of the central part of the mask including the phase singularity. It is clearly visible that at increasing distance from the center the pattern becomes locally periodic, except at the 2π phase discontinuities. The design shown here is optimized for the astronomical L-band ($3.4-4.1\,\mu$m) but due to the scale invariance of the problem the results can be applied to other wavelength bands, such as the visible range. However, due to the subwavelength condition, designs optimized for shorter wavelengths feature smaller building blocks and are therefore more challenging to manufacture.

The present sawtooth design features six phase discontinuities along with the central phase singularity. Since the optimization done with RCWA relies on the local periodicity of the pattern (which is not fulfilled in these regions), the effect of these discontinuities has to be considered. A full 3d simulation tool like FDTD is therefore used to simulate the response of these regions properly.

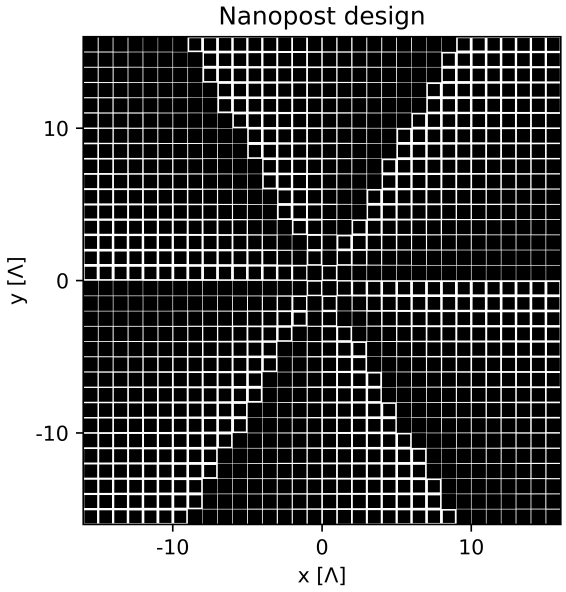

Figure 11.11: Square nanopost design for a charge-6 scalar vortex phase mask. This design uses square nanoposts of fixed height for application in the L-band (3.4-4.1 µm, $\Lambda = 1.4$ µm).

To describe the regions of the mask where the local periodicity condition is not fulfilled, a full 3d tool like FDTD is needed. Additionally, an FDTD simulation of the mask center is also useful to confirm that the phase mask design actually provides the desired phase ramp.

Figure 11.12 shows the response of a square nanopost design to a plane wave simulated with FDTD. The resulting phase ramp shows artifacts at the central singularity and at the 2π phase jumps, where the design can not be considered to be locally periodic. The resulting central defect is localized within a circular region of 10 unit cells, and the phase jump regions have a lateral extent of about 4 unit cells. These defects will propagate through the coronagraphic system and potentially reduce the performance of the phase mask. However, the defects are small and localized in small regions of the mask (compared to the full size of a nanopost mask) and therefore their effect on the contrast is expected to be small. A quantitative analysis will be needed in the future to confirm this assumption for different metasurface frameworks.

11.5 Metasurface cosine phase mask design

The nanopost implementation of a sawtooth vortex naturally features a series of phase discontinuities at the 2π phase jumps, degrading the performance of the mask due to the finite size of the transition region in the phase ramp. The number of phase discontinuities can be reduced by choosing another design topology, but at the expense

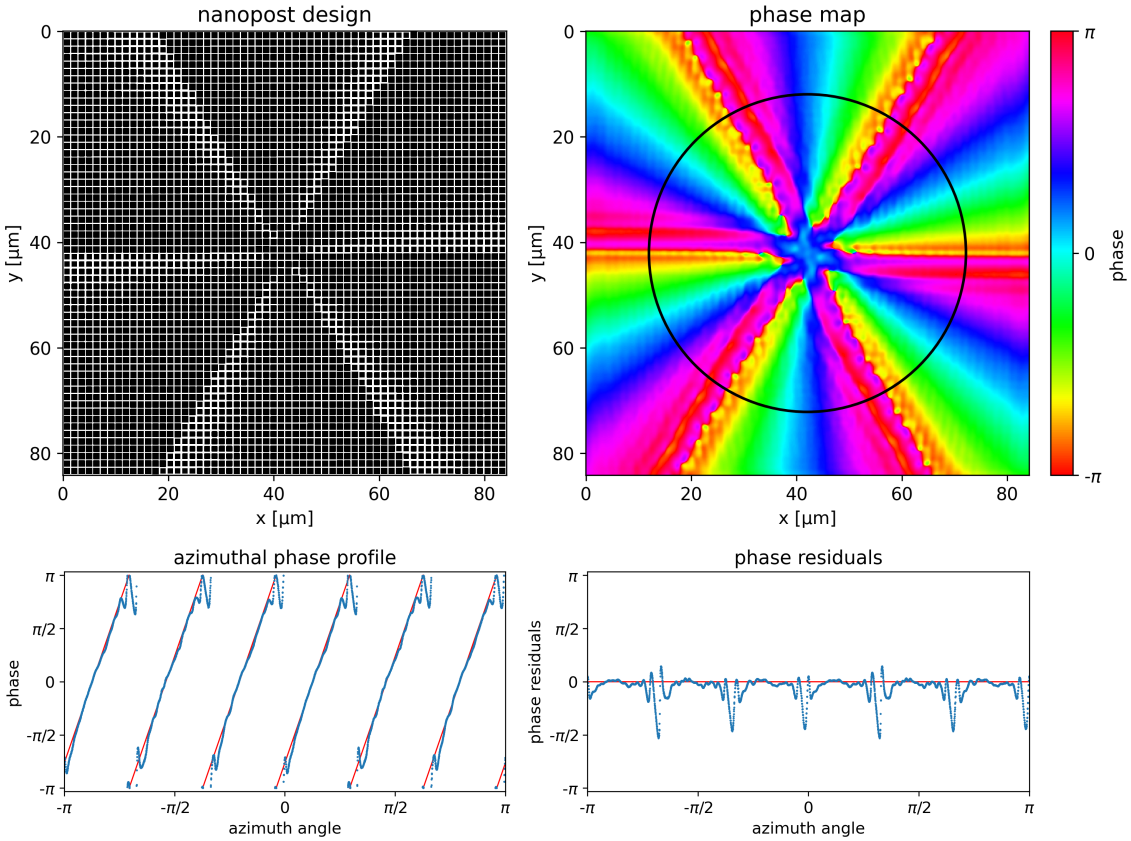

Figure 11.12: Top: Nanopost design for a sawtooth vortex phase mask of topological charge $\ell_p = 6$ (left) and FDTD simulation (right). This design uses square nanoposts of 9.5 µm height and is optimized for the L-band (3.4 – 4.2 µm). Bottom: The azimuthal phase profile (left) is obtained along the cut shown as black circle in the phase map. Its residuals (right) are small for regions with a local periodic pattern, but higher close to the characteristic 2π phase discontinuities.

of a higher overall phase coverage, or smaller topological charge. However, at least one 2π phase jump is needed to implement the scalar vortex.

An alternative design avoiding the phase jump is based on a continuous cosine-like phase pattern. The cosine pattern is part of the same family of phase masks as the vortex and provides perfect cancellation of an on-axis source in the Lyot plane, as described in Hénault (2015). The phase ϕ of this design is:

$$\phi(\theta) = z_0 \cos(k \cdot \theta)$$

where $z_0 = 2.40483$ is the first zero of the type-J Bessel function $J_0(z)$, k is the azimuthal frequency analogous to the topological charge of the vortex, and θ is the azimuthal coordinate. Higher zeros of J_0 are allowed but increase the phase coverage of the design, which is disadvantageous for a nanopost implementation.

The theoretical concept of the cosine phase mask for coronagraphic application is described and discussed in more detail in Hénault et al. (2016), Hénault (2018), Ruane et al. (2019). In particular, it removes the regions of phase discontinuity except for the central singularity, featuring a smoothly varying phase everywhere else on the mask. Similarly to the vortex phase mask, a cosine phase mask provides theoretically perfect starlight rejection for even azimuthal frequencies ($k = 2, 4, ...$). A cosine phase mask design with $k = 6$ is therefore presented here as alternative to the scalar vortex phase mask design for implementation with the present nanopost framework.

Following the same methodology as for the scalar vortex phase mask, the nanopost framework is optimized to provide the required achromatic phase coverage. The difference of the cosine phase pattern is its phase coverage, which is $\Delta\phi = 2z_0 = 4.81\,\text{rad}$ and therefore smaller than the 2π required for a scalar vortex design. This is an advantage because one can choose a shallower height for the nanopost framework and a smaller range of nanopost sizes to create the design. Figure 11.13 shows the nanopost design of a cosine phase mask derived from the phase curve shown earlier in Figure 11.6 (the block size range has been adapted to provide the reduced phase coverage), and the phase response of the central part of the phase mask simulated with FDTD. It is clearly noticeable that the cosine design avoids the phase jumps in the sawtooth vortex pattern resulting in a smooth azimuthal phase profile.

The reduced phase coverage and the lack of phase discontinuities make the cosine phase mask an interesting alternative to the scalar vortex phase mask. Using a nanopost framework, the implementation using well-established microfabrication techniques is crucial and the present cosine phase mask design has several advantages over the sawtooth vortex design. First, the overall height of the nanoposts can be chosen smaller since the phase coverage requirement is relaxed compared to the vortex designs. Second, the range of block sizes needed to provide the relevant phase coverage is smaller, resulting in patterns with smaller variation of filling factor. Finally, the design topology is continuous without any phase jumps (except for the central singularity present in all designs) increasing the relative area of the mask for which the RCWA simulations used

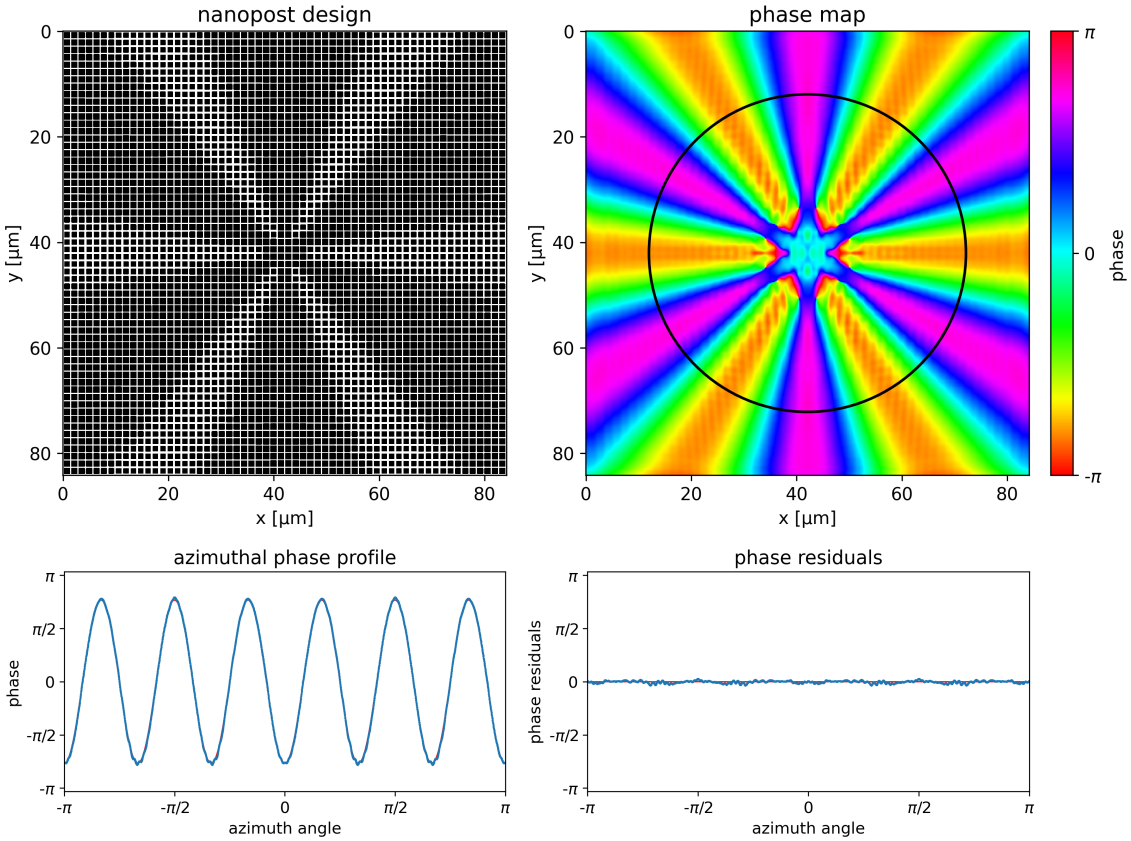

Figure 11.13: Top: Nanopost design for a cosine phase mask of azimuthal frequency $k = 6$ (left) and FDTD simulation (right). This design uses square nanoposts of 9.5 µm height and is optimized for the L-band (3.4−4.2 µm). Bottom: The azimuthal phase profile (left) is obtained along the cut shown as black circle in the phase map. Its residuals are smaller than for the sawtooth vortex design (right) due to the absence of the characteristic 2π phase discontinuities.

to optimize the nanopost framework are valid. However, the chromatic performance of both scalar vortex and cosine phase mask designs has still to be evaluated in the coronagraphic image in the final focal plane. This will ultimately allow a quantitative comparison of the metasurface implementation of both phase mask concepts.

11.6 Computing the contrast curves with wavefront propagation

After obtaining the near field response of the phase masks with RCWA and FDTD, the performance in a coronagraphic system has to be evaluated. FDTD can not be used to simulate the full phase mask as explained in Chapter 7, and it will be the subject of future work to implement a hybrid simulation framework using RCWA for locally periodic regions of the phase mask, and FDTD otherwise. Here, the near field response obtained with RCWA is propagated to the subsequent pupil plane where a slightly undersized Lyot stop (90 %) is used to block the light diffracted outside the pupil, and further to the final focal plane producing the coronagraphic image. The MFT algorithm implemented in FALCO is used to obtain the focal plane image, as described earlier in Section 4.2.2.

Figure 11.14 shows the contrast curves of the nanopost scalar vortex design at different wavelengths as well as the average contrast in a region from 3 to 10 λ/D. The nanopost design outperforms the theoretical vortex shown for comparison because of its optimization for a broad band. At the same time, at the central wavelength it performs significantly worse than the theoretical vortex. This is because while at the central wavelength both theoretical and nanopost vortex have the same perfect phase ramp, the nanopost framework suffers from varying transmission limiting the achievable contrast.

The same procedure can be applied to the cosine phase mask, leading to similar results shown in Figure 11.15. The resulting contrast is not as deep as for the scalar vortex, which could be due to the different design topology of the cosine compared to the scalar vortex.

After obtaining the raw contrast of both nanopost-based scalar vortex and cosine phase mask designs, the performance in the scoring region between 3 and 10 λ/D can be increased by using wavefront control algorithms and a deformable mirror to dig a dark hole in the coronagraphic image. Here, EFC is used to dig a dark hole with FALCO in a 180° dark hole between 3 and 10 λ/D. The resulting focal plane image is shown in Figure 11.16 along with the respective surface profile of the deformable mirror. Using EFC, the average contrast in the scoring region is 3.70×10^{-9} in 20 % bandwidth, which represents a moderate improvement by a factor of 3 compared to the raw contrast of 1.24×10^{-8}. The contrast values presented here are promising, but still far from the 10^{-10} goal. However, the use of more complex metasurface topologies

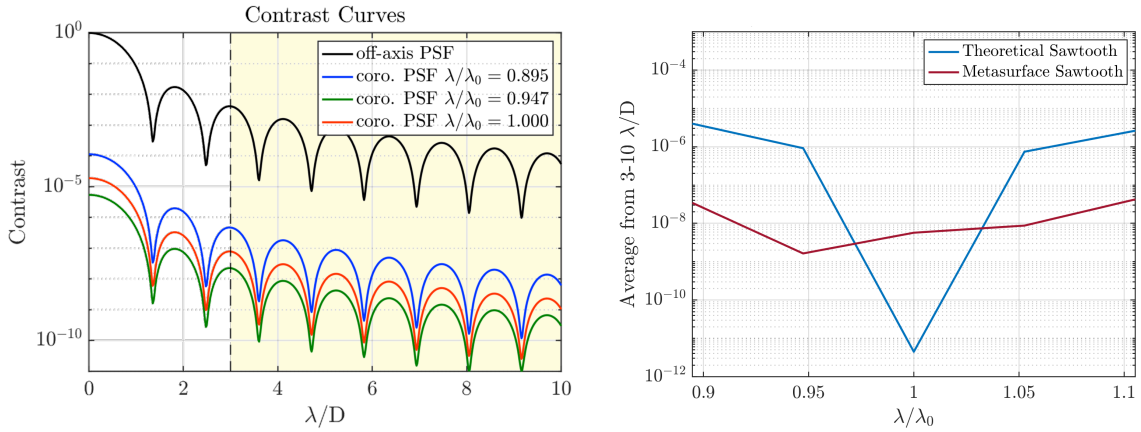

Figure 11.14: Contrast curves for a nanopost sawtooth vortex computed with FALCO at several wavelengths (left). Only three wavelengths are shown for the sake of clarity ($\lambda_0 = 3.8\,\mu\text{m}$). The region of interest ($3 - 10\,\lambda/\text{D}$) is highlighted (yellow), and used to obtain the averaged contrast shown in the second graph (right).

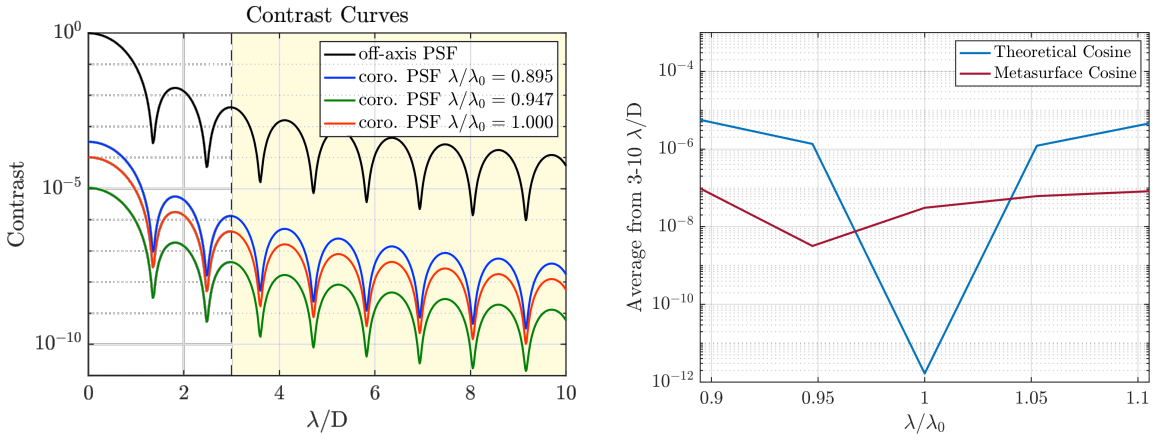

Figure 11.15: Contrast curves for a nanopost cosine phase mask computed with FALCO at several wavelengths (left). Only three wavelengths are shown for the sake of clarity ($\lambda_0 = 3.8\,\mu\text{m}$). The region of interest ($3 - 10\,\lambda/\text{D}$) is highlighted (yellow), and used to obtain the averaged contrast shown in the second graph (right).

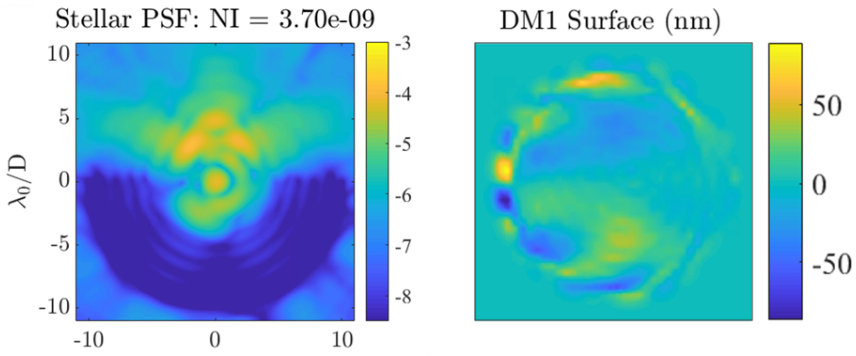

Figure 11.16: Dark hole obtained with EFC applied to a nanopost sawtooth vortex design (left) and respective surface profile of the deformable mirror (right). The scoring region is a circular region between 3 and 10 λ/D spanning 180° in the focal plane and optimized for 20 % bandwidth.

combined with an optimized material choice could reach the required contrast. This study is beyond the scope of this thesis and will be covered in a future project.

11.7 Perspectives for the manufacturing of nanopost phase masks

The nanopost framework discussed in this chapter has been optimized for the implementation of broadband scalar vortex and cosine phase masks. During the optimization, designs that are within reach of current manufacturing techniques were favored. In this way, the height of the metasurface blocks as well as their size variation across the mask were kept relatively small, so that manufacturing tests of these designs can be envisaged in the near future.

11.7.1 Influence of manufacturing uncertainties

When considering the manufacturing of a prototype, the manufacturing uncertainties have to be taken into account. To estimate the influence of manufacturing uncertainties, several designs with parameters deviating from the optimal parameters have been simulated. The performance of a nanopost sawtooth vortex design is shown in Figure 11.17 for an optimal design, and for the same design with undersized nanoposts. The performance of the design is sensitive to small variations of the post size, but still reaches a null depth of 3×10^{-3} for posts undersized by up to 20 %. The influence of a linearly varying nanopost height with the filling factor is found in a similar way to considerably degrade the design performance, reaching a null depth of only 2×10^{-2} for a height variation of ± 10 % across the mask. Therefore, the manufacturing process

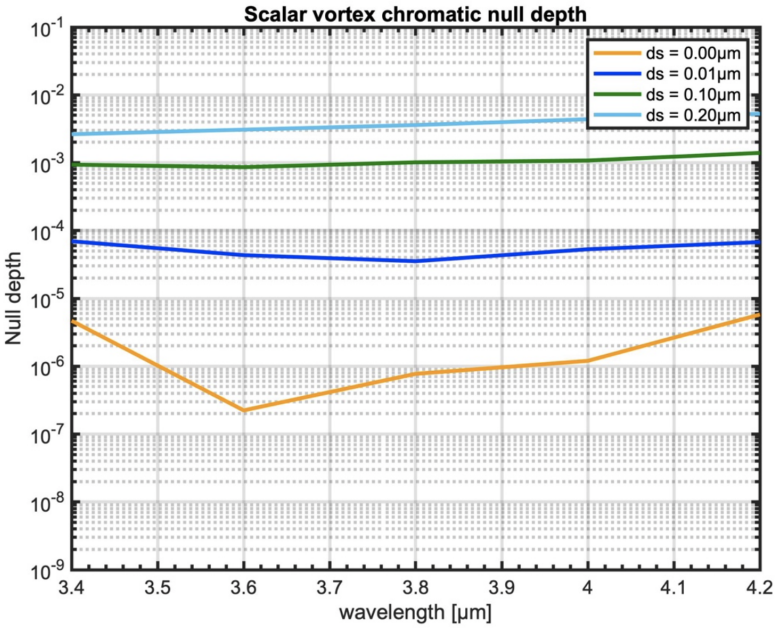

Figure 11.17: Performance of the nanopost mask for systematic errors on the post size. The null depth is shown for the optimal design, and for the same design with nanoposts undersized by 10, 100, and 200 nm.

for nanopost designs has to be controlled accurately to reduce the uncertainties as much as possible.

11.7.2 Adapted manufacturing process of nanopost phase masks

In addition to simulations of nanopost designs, consideration of different manufacturing processes to fabricate a nanopost prototype is crucial. The choice of the most suitable material can greatly increase the practical use of a prototype. Examples include silicon, titanium dioxide and diamond, each with its own advantages. Here, the manufacturing process used to manufacture the AGPMs described in Chapter 6 is considered. The main limitation in manufacturing is the non-zero sidewall angle, resulting in a trapezoidal cross section of the subwavelength structures. This effect is similar for a nanopost framework and can be accounted for by including the sidewall angle as fixed parameter in the optimization process.

An effect that has to be considered when etching nanopost designs is the microloading effect due to the varying filling factor across the mask. More precisely, the varying size of the gaps in the design (between the nanoposts) result in a varying etch rate (Hedlund et al., 1994), and therefore a varying etch depth when etching the design onto a substrate. While for small variations of nanopost size the microloading effect is expected to result only in small variations of the etch depth, the exact behavior has to be tested experimentally. For this purpose, the etch depth can be measured

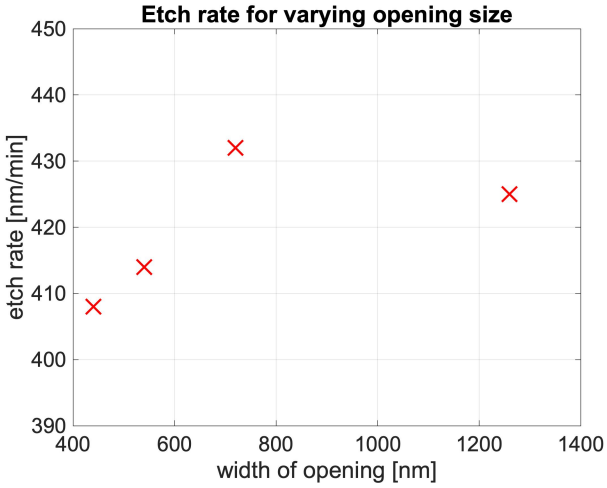

Figure 11.18: Etch rate for anti-reflective grating patterns featuring different gap size. The etch rate varies by up to 10 % across the gap sizes considered here, and can be considered as constant for the small variations of filling factor considered in this chapter.

period	width of opening	etched depth	etch rate
800 nm	440 nm	6120 nm	408 nm/min
1300 nm	540 nm	6210 nm	414 nm/min
1300 nm	720 nm	6480 nm	432 nm/min
3400 nm	1260 nm	6380 nm	425 nm/min

Table 11.1: Etch rate for anti-reflective grating patterns featuring different gap size.

for a pattern with varying filling factor across the same mask, or different patterns with different gap size can be etched simultaneously, if there is no single pattern with varying gap size available (as was the case for this study). The results are shown in Table 11.1 and compiled in Figure 11.18. The variation of the etch rate is on the order of 10 % for this large range of gap sizes, which is reasonably small. However, a manufacturing process based on selective etching involving a stopping layer might be considered if the etch rate variation caused by the microloading effect turns out to influence the performance of the mask too much. Figure 11.19 shows a sketch of such a manufacturing process using a SiO_2 stopping layer for a mask etched in silicon.

Another challenge when manufacturing patterns with varying filling factor is the transfer of the pattern to the substrate. The solvent-assisted micromolding process described in Section 2.2.3 relies on a constant filling factor across the mask. In case of varying filling factor across the mask, the micromolding process results in an uneven thickness of the photoresist mask and residual layer. If the variations in thickness are small, a quick oxygen plasma can be used to remove the residual layer without significantly affecting the mask itself. But if the thickness variations become a significant fraction

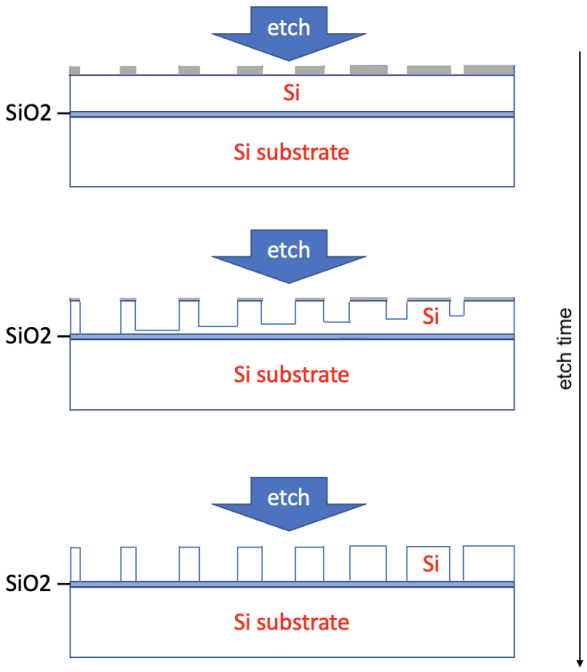

Figure 11.19: Possible manufacturing process involving a stopping layer. A silicon substrate sputtered with a thin SiO_2 layer and finally a Si layer whose thickness is determined by the nanopost height is etched selectively until the stopping layer is reached. Due to the selective etching, the SiO_2 stopping layer is etched much slower than the silicon layer and therefore the nominal etch depth can be achieved for a large range of nanopost sizes.

of the total mask thickness, the process results in strongly uneven thickness of the photoresist mask, which can lead to unwanted effects in the subsequent etching steps.

So far, the photoresist layer is spin coated as thin layer on top of the substrate. Then, the PDMS mask is placed on top, and the stack is exposed to ethanol vapor liquefying the photoresist, which then fills the PDMS mask by capillary effects. For patterns featuring a constant filling factor, the profile of the photoresist film is the same everywhere on the mask. But in the case of varying filling factor, the photoresist mask will be thinner in regions with large gaps of the PDMS mask (corresponding to larger nanoposts), leading to a varying photoresist thickness across the mask (as illustrated in Figure 11.20) because of the absence of lateral flow on the macroscopic scale.

To solve this issue, the way the pattern is transferred to the substrate can be adapted to account for varying filling factors. The idea is to fill the gaps in the PDMS pattern with photoresist prior to placing it on the substrate. In this way, the photoresist layer will be flat when placing it on the substrate, as shown in Figure 11.21. To provide optimal adhesion between the PDMS and the photoresist, the PDMS is treated with a quick

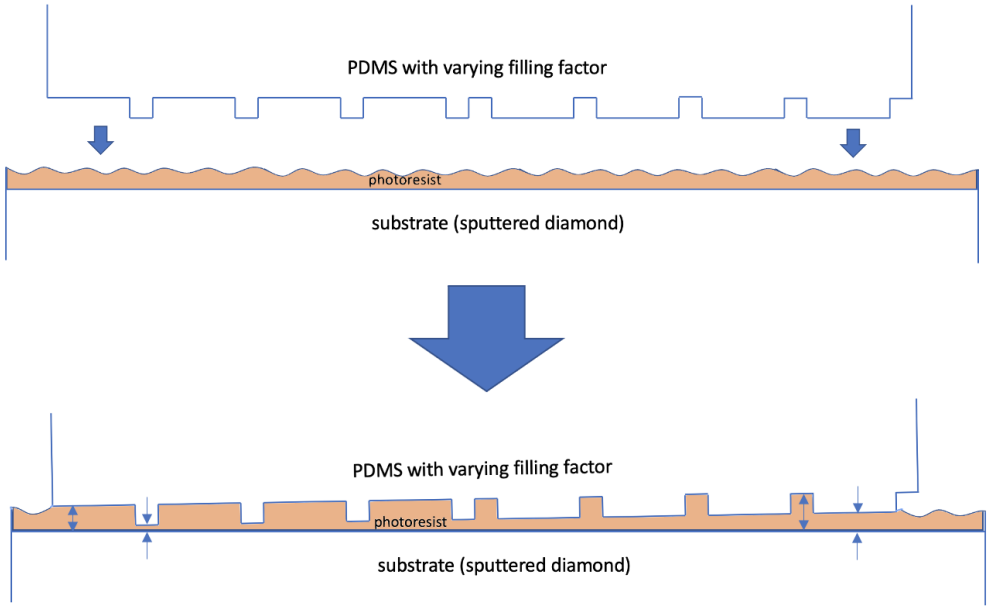

Figure 11.20: Solvent-assisted micromolding (SAMIM) technique applied to a nanopost pattern with varying filling factor. Due to the lack of lateral flow at the macroscopic scale, the process results in a varying thickness of the photoresist layer across the mask.

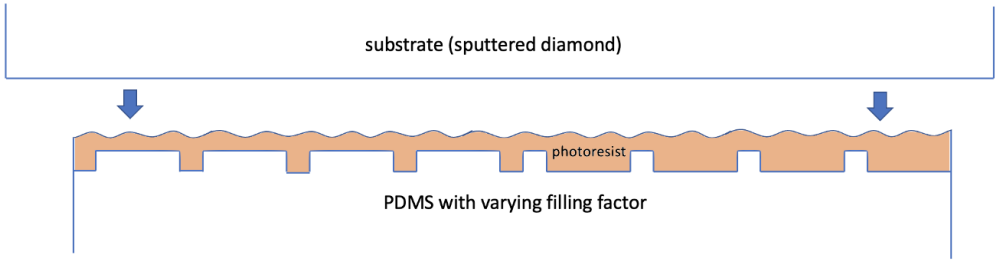

Figure 11.21: Solvent-assisted micromolding (SAMIM) technique optimized for a nanopost pattern with varying filling factor. By spin coating the PDMS, the photoresist fills the non-uniform gaps in the PDMS and the resulting thickness of the photoresist layer is constant across the mask.

oxygen plasma. The PDMS is then spin coated, and placed on the substrate (which is itself spin coated with a thin photoresist layer). Since now two layers of photoresist are used, the photoresist is diluted more to obtain a similar mask thickness. The photoresist is then baked as usual before removing the PDMS.

Due to time constraints, no etching tests have been performed with this new method yet. However, before manufacturing a nanopost phase mask prototype it will be necessary to perform preliminary tests and confirm that this method is suited to imprint the pattern on the photoresist. For this purpose, and to study the etching behavior for nanopost designs in detail, the first prototype pattern will feature not only the nanopost phase mask design, but also test patterns with uniformly increasing filling factor. The first prototype pattern will include three nanopost phase mask patterns and three test patterns featuring uniformly increasing filling factor, as illustrated in Figure 11.22. Two patterns will feature a charge-2 vortex phase mask and one pattern will feature a cosine phase mask with azimuthal frequency $k = 2$ for application in the L-band. The two scalar vortex patterns use different size ranges of the nanoposts to assess the influence of larger filling factor variations. All three designs are charge-2 designs optimized for the L-band, and for the typical sidewall angle achievable (1.5°). This will allow to compare the prototype with the previously manufactured AGPMs. The three test patterns featuring a linearly increasing filling factor will be optimized for the L, K, and H-bands. These patterns aim at exploring the current manufacturing limitations in terms of feature size and identify possible new challenges when moving towards applications for shorter wavelengths.

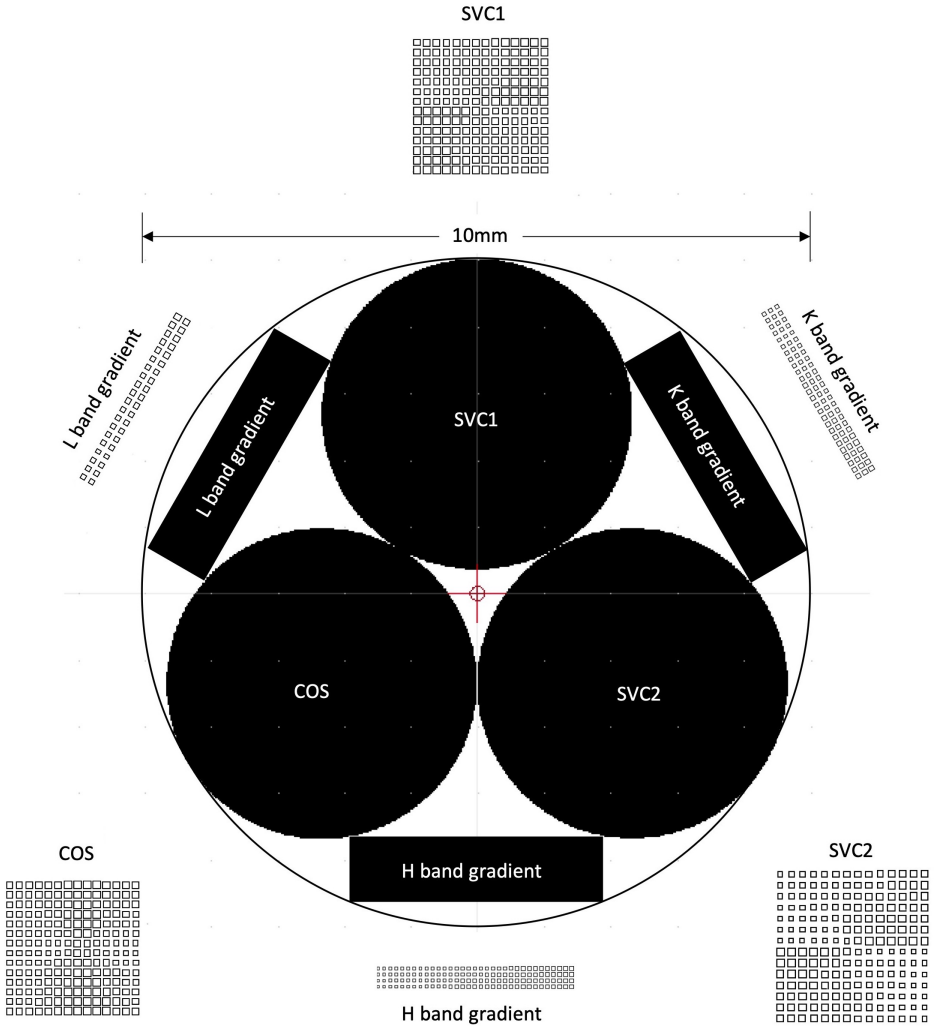

Figure 11.22: Design of the first nanopost phase mask prototype featuring two scalar vortex designs and one cosine phase mask design with azimuthal frequency $k = 2$. Three nanopost test patterns featuring a linearly increasing filling factor optimized for the L, K and H band are included to explore the current manufacturing limitations with this manufacturing approach. The unit cell size Λ of the patterns is 1.42 µm, 0.92 µm and 0.67 µm for the three bands, respectively. To account for the shrinking of the lateral size of the nanoposts during the manufacturing process, the nanoposts are oversized by 100 nm for the smallest, and 200 nm for the largest posts before writing the master pattern with E-beam lithography.

Part V

Conclusion

Conclusion

The main goal of this thesis is to optimize current vortex phase mask designs, and investigate new metasurface-based implementations of vortex phase masks of various topological charge. These new designs are motivated by the requirements of future ground- and space-based telescopes, for which the performance of the coronagraph needs to be improved.

The first step of the work presented in this thesis was to develop a simulation framework capable of simulating the AGPM including the central phase singularity, which was not possible with the RCWA simulation framework used before. This simulation framework can be applied to any microstructured coronagraphic mask, providing a rigorous description of both periodic and aperiodic features with RCWA and FDTD, respectively. The performance in a realistic coronagraphic system is then assessed with a wavefront propagation code. The simulation framework was applied to the simulation of the microstructured vortex phase mask designs in this thesis, as well as the design of the METIS ring apodizer.

The further work carried out within this thesis applies the new simulation framework to the design and optimization of microstructured vortex phase masks. The first application focused on the the design of the AGPMs for the METIS instrument, which includes the optimization of the grating parameters as well as the manufacturing of first AGPMs. The updated manufacturing process pushed the design towards smaller sidewall angles, enabling the manufacturing of higher aspect ratio structures. The same simulation tools were used to optimize the anti-reflective grating etched on the back of the AGPMs in order to avoid internal reflections. After manufacturing, first performance tests of the METIS N-band AGPMs were performed on a dedicated test bench. These performance measurements allowed the grating parameters of the first N-band AGPM to be inferred by comparing them with the predicted performance obtained from simulations. The changes made to the manufacturing process is also an important step towards the fabrication of more complex two-dimensional metasurface structures, enabling higher aspect ratios needed to provide optimal broadband performance of these masks. In addition, the lower sidewall angle increases the parameter space that can be accessed with diamond microfabrication.

The optimization of the AGPM central region using the simulation tools described above is the second application this thesis focused on. The central leakage term was identified as a limiting factor for future high-performance AGPMs, especially when operated at low f-numbers. These results were confirmed experimentally on the dedicated

coronagraphic test bench VODCA. The design of the AGPM center was optimized and it was shown that a moderate performance increase can be achieved by optimizing the central pattern. Two approaches have been proposed and evaluated. First, the optimization of the AGPM center using circular grooves of varying size and periodicity at the center, and second, a design based on rectangular building blocks optimized using a machine-learning approach to minimize the polarization leakage arising from the center. Both methods confirm that the circular grooves of the AGPM are close to the optimal design achievable for the AGPM center.

In addition to optimizing the AGPM design, vortex phase mask patterns of topological charge 4 were investigated using subwavelength gratings. The theoretical design for subwavelength grating vortex phase masks was discretized into segmented designs and their performance was simulated with the simulation framework introduced earlier. The behavior of the designs was quantified, revealing the regions of increased leakage, and showing that a new hybrid design combining the advantages of the two initially proposed segmented designs can improve the performance. The manufacturing of this hybrid design was started along with the manufacturing of the METIS AGPMs, but has not yet been finalized due to issues encountered with the plasma etcher. While an alternative plasma etcher is available, which is used for the METIS AGPMs, this etcher provides a sidewall angle for which the charge-4 pattern was not optimized and therefore would result in a considerably degraded contrast of the prototype.

The segmented designs investigated for the implementation of charge-4 vortex phase masks suffer from the transitions between segments. To avoid this issue, a metasurface framework consisting of rectangular nanofins was investigated and applied design of vortex phase masks of higher topological charge. In principle, designs exploiting the orientation of the nanofins imprint a discrete phase for each design pixel, leading to a quasi-continuous pattern (except at the central singularity). However, the work performed within this thesis revealed two issues arising from this concept. First, the nanofin parameters have to be re-optimized for each orientation in their pixelization grid to provide the characteristic half-wave plate behavior needed to minimize the polarization leakage. However, this process results in spatially varying nanofin parameters with varying effective refractive index, leading to an additional phase term which is not determined by the nanofin orientation. This effect makes the implementation of a vortex phase ramp using rectangular nanofins challenging, since the geometric parameters would have to be optimized for both half-wave plate performance and for the desired phase response in every part of the mask. Second, the design pixelization influences the orientation of the effective birefringence axis of the metasurface pattern, which results in a geometric phase not fully determined by the local orientation of the nanofins, but rather by the pixelization axes. These effects favor the study of polarization-independent (scalar) metasurface implementations over the polarization-dependent (vectorial) rectangular nanofins.

In addition to the nanofin metasurface framework, a scalar metasurface framework using polarization independent symmetric building blocks was investigated for the

implementation of scalar vortex phase masks. The study of square nanoposts led to the design of a scalar vortex phase mask increasing the contrast by more than one order of magnitude compared to a classical sawtooth scalar vortex phase mask. This technology is potentially applicable to future space telescope missions, which need contrasts on the order of 10^{-10} at the inner working angle of the coronagraph. Optimizing the two-dimensional metasurface framework for achromatic performance led to contrasts of 3×10^{-9} using a wavefront control algorithm, pushing the limits towards the requirements of future space telescopes tailored for imaging Earth-like planets around Sun-like stars.

Finally, the simulation tools were applied to the design of the METIS ring apodizer featuring a grayscale microdot pattern in order to implement the grayscale transmission required in an annular region of the apodizing pupil plane mask. The optimal design of the microdot pattern fulfilling the requirements for METIS led to a series of test patterns featuring different geometrical parameters of the microdots. These patterns have been manufactured and tested in lab, providing first results which confirm the transmission predictions of the RCWA simulations used for optimization. Further assessment of these patterns is still ongoing, notably the characterization of the phase response, which is another important requirement. The final pattern of the METIS ring apodizer will be fabricated after characterization of this effect, and tested for the nominal transmission and phase behavior.

Overall this thesis contributed to the advancement of vortex phase mask technology development. The simulation framework developed throughout this thesis specifically impacted the design optimization of current and future vortex phase masks, allowing for the optimization of complex two-dimensional metasurface patterns. Additionally, the experimental work done in this thesis extended the manufacturing techniques to improve the performance of AGPMs and also explored some of the expected limitations with two-dimensional metasurface patterns for scalar vortex phase mask designs.

Perspectives

The simulation tools presented in Part II of this thesis are the foundation of the results presented in the subsequent chapters. Several simulation methods have been combined, simulating the response of the microstructured masks described in this thesis. Combining the different methods into a robust simulation pipeline is a further step to make the design process of vortex phase masks more efficient, and will therefore be explored in a future project in the coming years.

The updated manufacturing process discussed in this thesis has been used to manufacture the first N-band AGPMs for METIS and their testing is currently underway. In case these masks do not meet the requirements of METIS, re-etching of these components will be considered, as well as the fabrication of further N-band AGPMs. Afterwards, the L and M-band AGPMs will be manufactured and tested. In addition, the remaining etching step for the charge-4 hybrid mask will be carried out to finalize its manufacturing.

Future AGPMs might feature the optimized design of the AGPM center presented in Chapter 7. This design might be considered for fabrication, provided that the effect of the varying grating parameters at the center of the mask on the manufacturing results is investigated more in detail, since varying grating parameters might result in varying etch depth. However, the manufacturing tests obtained so far with different anti-reflective grating patterns show that the difference in etch rate is small for the region of the parameter space considered for these designs.

The application of rectangular metasurface building blocks to the design of vortex phase masks has been shown to be challenging and needs further design investigation. The concept of scalar vortex phase masks using polarization-independent metasurfaces is however promising, introducing additional degrees of freedom which can be leveraged to make the scalar vortex more achromatic. A project building on the preliminary work performed in Chapter 11 of this thesis is therefore planned, extending the metasurface framework to more complex shapes featuring additional degrees of freedom, and pushing the contrast towards the 10^{-10} contrast level needed for future space telescope missions targeting the direct imaging of terrestrial exoplanets.

Building on the updated manufacturing process used in this thesis, the manufacturing of a first prototype is an important next step in the development of two-dimensional metasurface implementations of vortex phase masks. It is therefore mandatory to advance the manufacturing capabilities in order to overcome the challenges expected with these designs, notably the high aspect ratio and the varying grating parameters

across the mask. In addition, this involves the consideration of different materials for a prototype and choosing the most promising manufacturing approach, depending on the wavelength of the application, the desired chromatic performance as well as the feasibility of the manufacturing process to provide a high-performance component. Moreover, it is important to accurately characterize the phase response of a manufactured component to understand the limitations of the simulations and the manufacturing process, as well as the manufacturing uncertainties and how they impact the performance of the component. Several iterations adapting the design process to the manufacturing constraints will be needed before obtaining a high-performance component. The design and manufacturing will therefore be adapted continuously to increase the performance of the prototypes resulting in a design and manufacturing process which is mature enough to produce science-grade metasurface scalar vortex phase masks of different topological charges for different applications.

The advantage of an achromatic metasurface framework is the ability to imprint virtually any phase pattern on a phase mask, including other concepts such as the phase knife (Abe et al., 2001). This phase mask in particular is an interesting concept combined with the vortex fiber nuller (Serabyn et al., 2022) and relatively simple to manufacture, since it consists of only two discrete phase regions. A first prototype of a metasurface phase mask might therefore feature a phase knife. This will be a useful means to assess the metasurface framework due to its simple design and at the same time could be a candidate for a first metasurface mask installed on sky.

Part VI

Appendix

A METIS ring apodizer development

A.1 Ring apodizer design

METIS provides different coronagraphic modes covering the L, M and N-bands ($3 - 13\,\mu m$). The classical vortex coronagraph is available in all three bands. For the L and M-band the ring-apodized vortex coronagraph (RAVC) is provided as additional mode. This solution is not provided for the N-band because it implies a lower overall throughput which is critical in the N-band due to the higher background emission.

In the case of a charge-2 vortex the pupil plane apodizer of the RAVC consists of a greyscale mask featuring a single greyscale ring of a certain radius and transmission. The optimal size and transmission of the greyscale rings are defined by the size of the central obscuration. The charge-2 case corresponds to the case of METIS with a central obscuration of around 30 % of the pupil diameter. The design chosen for METIS features a greyscale ring with inner radius of 58.8 % of the pupil radius and 62.5 % intensity transmission. Here we refer to the intensity transmission (squared amplitude transmission) as opposed to Mawet et al. (2013b) who refers to the amplitude transmission of the greyscale region. Figure A.1 shows the ring apodizer design used for METIS.

A.2 The microdot technology

The property of the RAVC to restore the perfect starlight cancellation of the vortex relies on the careful control of transmission in the greyscale ring. It is therefore crucial to control this parameter across the desired wavelength range. One suited technology to create a greyscale apodizer is based on a microdot technology. This technology uses reflective chromium microdots with a size of a few times the wavelength (Zhang et al., 2018) to create an effective greyscale transmission. The microdot technology is well-mastered (Martinez et al., 2009) and apodizers based on microdot patterns have been used previously in the context of high contrast imaging (Kasper & Mawet, 2012; Macintosh et al., 2008). The microdots technology has the advantage of being achromatic in phase and transmission making it ideally suited for the fabrication of the ring apodizer. Additionally, the phase introduced by the microdots is minimal, which is important for the working principle of the RAVC based on a pure amplitude mask. Other potential technologies to create a ring apodizer exist, such as transmissive greyscale components relying on thin layers of metals and dielectrics, and have been

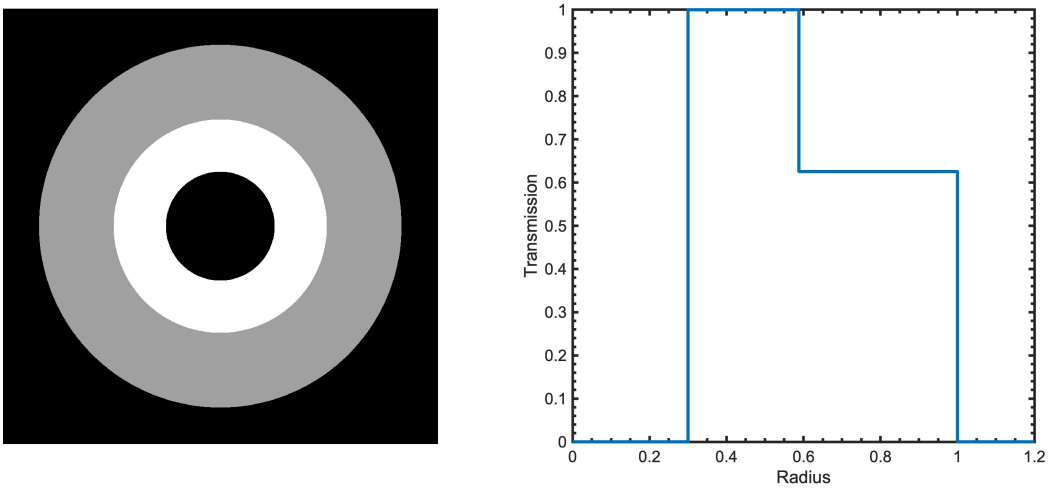

Figure A.1: Design of the ring apodizer for METIS. The greyscale ring provides 62.5 % intensity transmission. Left: Ring apodized pupil design, right: radial transmission profile.

successfully applied to coronagraphy (Trauger et al., 2011). For METIS, the microdot technology based on chromium microdots printed on a coated ZnSe substrate has been chosen to fabricate the ring apodizer.

A.3 Design and optimization of the microdot pattern

Here, the optimal conception of the microdot pattern is discussed. The goal of the microdot pattern is to provide the desired transmission across the pupil while limiting phase distortions due to the microdot technology. Rigorous simulation tools such as RCWA and FDTD are used to study the response of the microdot pattern to an incident beam, and find an optimal solution providing the desired transmission and maintaining a flat wavefront at the same time.

The algorithm to create the microdot pattern is based on an error-diffusion code (Floyd & Steinberg, 1976). This method is used to visualize greyscale patterns using binary pixel values, as done in inkjet printers or binary displays (Ulichney, 1988). Error diffusion takes a greyscale array as input and sweeps across all pixels deciding whether a pixel is closer to be considered opaque or transparent. The difference between the desired greyscale value of a pixel and its imposed binary value is then diffused to the adjacent pixels which have not yet been processed. The error is diffused according to a scheme defined by the kernel of the error diffusion algorithm, which determines how the error is spread across the adjacent pixels. In this way, the desired greyscale is achieved on a macroscopic scale while the pattern consists of binary pixels on a microscopic scale. Here, the microdots are implemented by depositing a thin chromium layer on a substrate in each opaque pixel, while no chromium is present in the transparent pixels.

For filling factors that are close to an integer fraction, the microdot pattern becomes periodic. This can create diffraction artifacts affecting the amplitude and phase response of the pattern and therefore the resulting effective transmission. The effective transmission of periodic patterns depends on the exact microdot pattern, and not only the filling factor, unlike for random patterns. This is because a specific filling factor can be obtained using different periodic patterns, each one with a specific phase and amplitude response, while for random patterns there is no discrete diffraction orders but rather a lot of faint orders creating a sort of halo around the 0th order. It is therefore necessary to avoid integer fractions of the filling factor and add a small offset if necessary. In the case of chromium microdots, integer fractions of the filling factor can be avoided by tuning the remaining parameters of the pattern, which are the dot size and the chromium layer thickness. Another issue to consider is the behavior of the error-diffusion algorithm close to sharp borders aligned with the pixel axes (horizontal or vertical), which tend to be more periodic than parts of the pattern that are far from such edges. When creating a pattern using error-diffusion it is therefore mandatory to check for periodic patterns close to straight edges, or to use a smaller patch than the full pattern. Figure A.2 illustrates this issue showing that the microdot pattern close to the left border of the full pattern is periodic, and that a smaller patch far from the left edge is a more appropriate choice.

The pattern resulting from error diffusion is quasi-random. Diffraction due to large-scale periodic features is therefore avoided. The resulting pattern of microdots is non-periodic and the best tool to simulate the full apodizer is FDTD. However, the size of the ring apodizer is too large to be considered for FDTD with reasonable computational resources, and even a small patch is computationally expensive since the size of the dots is several times the wavelength. Therefore, RCWA simulations with a large unit-cell are used to describe the field response of a quasi-periodic microdot pattern. Using large unit cells with 16×16 microdot pixels describes the local response of the microdots sufficiently well without introducing small scale periodic structures. The choice of 16×16 pixels in an RCWA simulation cell is a compromise between computational resources and accurate description of the semi-random pattern, resulting in approximately 100 CPUhours of computation time per simulation. The accuracy of the results is also limited by the fraction of opaque pixels in the simulation cell since the filling factor is fixed to fractions of $16^2 = 256$. However, the results of these simulations show that the transmission and phase can be interpolated between these discrete filling factors to obtain the response of intermediate greyscale values. Figure A.3 shows an RCWA simulation result for a 16×16 pixels microdot pattern. The resulting quantities from the simulation are the phase step between the clear and the greyscale part of the apodizer and the transmission of the 0th order. For each microdot simulation a reference simulation of a clear section with no microdots is used to normalize the transmission and phase. Since the microdots are larger than the wavelength, a certain amount of the transmitted light is diffracted into higher diffraction orders. For high-contrast imaging applications, the higher diffraction orders have to be at large enough

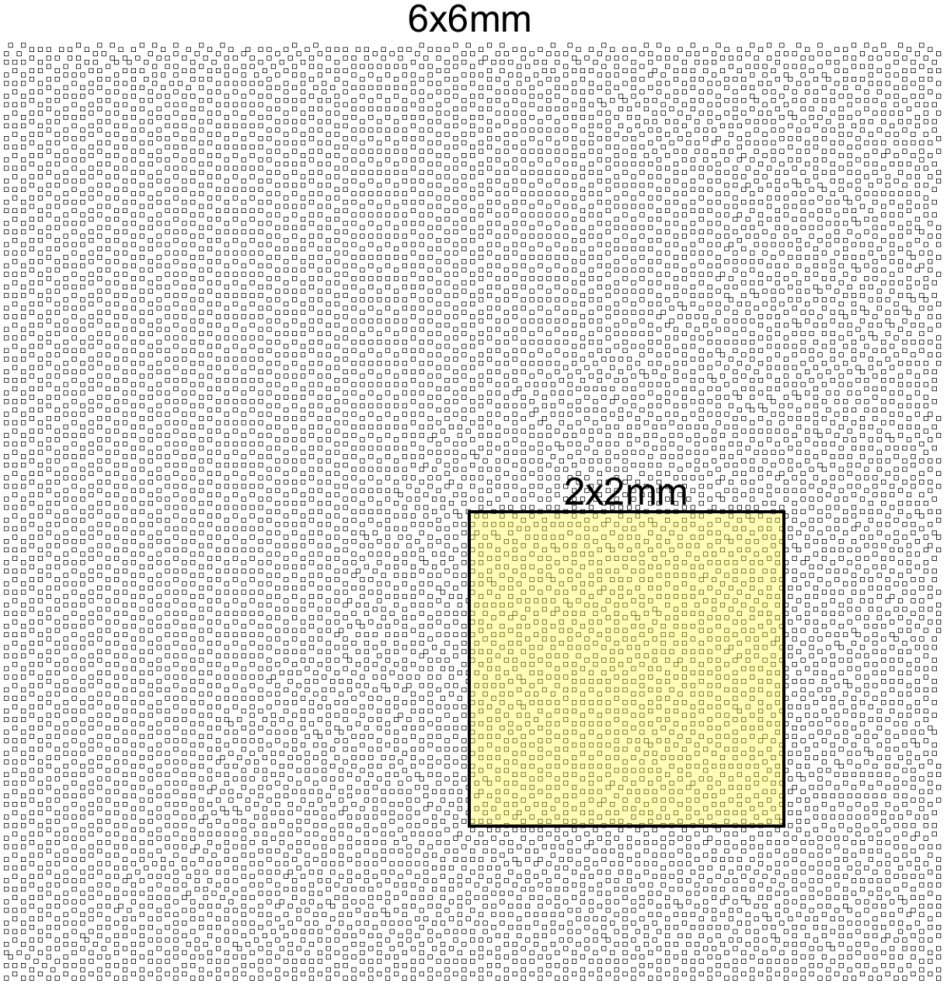

Figure A.2: 6 mm × 6 mm microdot pattern showing periodic behavior close to the left and bottom edge due to a filling factor close to an integer fraction (25 % = 1/4). Clever selection of the 2 mm × 2 mm inset position is needed to obtain a semi-random pattern highlighted in yellow.

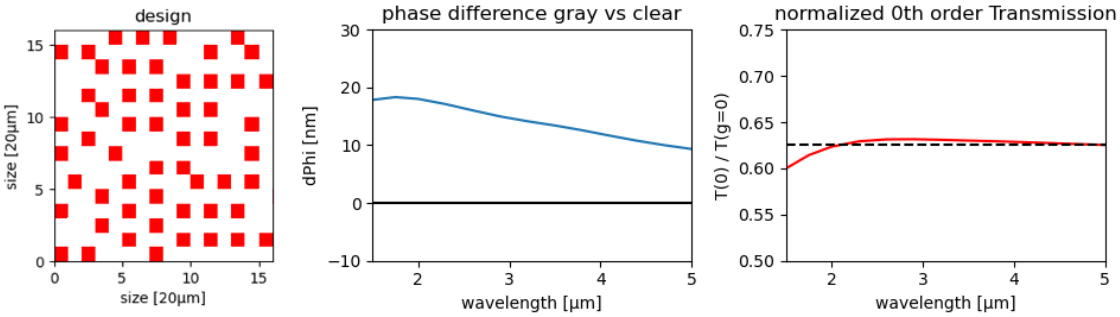

Figure A.3: Left: 16 × 16 microdot pattern design simulated with RCWA in the relevant wavelength range (1.5-5 µm). The design shows the unit cell used in the RCWA simulations which is replicated periodically due to the periodic boundaries necessary for RCWA. Center: chromatic phase step between the microdot pattern and a clear region of the apodizer without microdots. Right: chromatic 0th-order intensity transmission normalized by the transmission of the clear region of the apodizer with no microdots. The nominal intensity transmission of 62.5 % is shown as dashed line.

angles from the optical axis to be blocked by the pupil stops and mechanics of the instrument, which leaves only the 0th order propagating through the coronagraph.

The size of the microdots should be several times the wavelength to provide smooth results (Zhang et al., 2018). Smaller microdots produce undesired resonances resulting in features in the phase and transmission values. Figure A.4 shows the phase step at the clear-greyscale transition and the 0th-order transmission for different dot sizes. Both quantities show resonant features for dot sizes smaller than 3-4 times the wavelength.

In principle, the transmission of a microdot pattern is defined by the filling factor of the pattern. The fraction of light falling on the chromium microdots is reflected, while the fraction falling between them is transmitted. However, since the microdot pattern is not operating in the subwavelength domain, non-zero orders are propagating. This means, that a significant part of the light is diffracted at low angles. In the coronagraph, only the 0th order is used, and higher orders have to propagate at sufficiently high angles to be intercepted by the pupil stops and not affect the coronagraphic performance. When assessing the microdot pattern, only the transmission in the 0th order is considered as metric which should match the 62.5 % intensity transmission needed for METIS. At the same time, light propagating in low diffraction orders has to be given careful consideration to avoid straylight or back-reflections from any component downstream the ring apodizer. The RCWA simulations show that the typical amount of light diffracted in higher orders accounts for a total of around 10 % of the incoming light, and therefore could be a substantial source of straylight. Since the diffraction angles depend on the dot size, the microdots must be small enough to avoid light diffracted at small angles. For microdots of 20 µm size, the lowest significant diffrac-

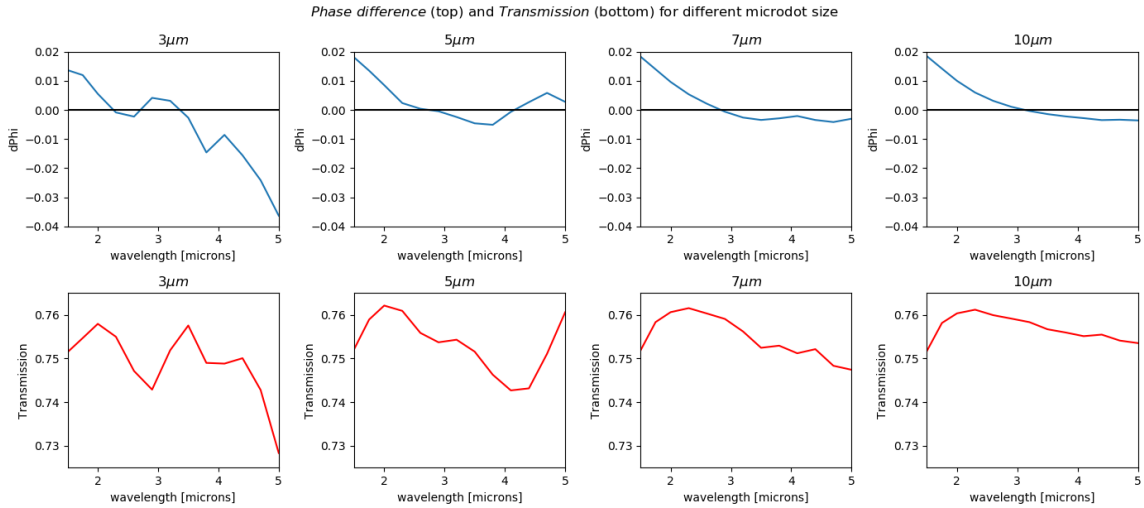

Figure A.4: Phase step (top) and transmission (bottom) for different dot sizes (3, 5, 7 and 10 µm from left to right) in the relevant wavelength range (1.5-5 µm). The filling factor used for these tests is 12.5 % and the chromium layer thickness is 113.5 nm. Both phase and transmission curves show resonant features for wavelengths smaller than 3-4 times the wavelength.

tion orders have intensities in the order of 10^{-5} and propagate at an angle of 0.5°, which is high enough for the METIS straylight requirements.

To optimize the performance of the ring apodizer, the following parameters can be tuned: dot size, filling factor and chromium layer thickness. The dot size has an influence on the subwavelength behavior of the mask and has to be chosen large enough to avoid resonances due to subwavelength effects. At the same time, the pixel size influences the diffraction behavior of the ring apodizer, which requires the dots to be small enough to have the lowest significant diffraction orders at large enough angles from the optical axis. The range of filling factors has to be chosen to avoid integer fractions such as 25 % = 1/4, resulting in periodic microdot patterns. The present simulations suggest that the optimal filling factor is close to this value for a nominal transmission of 62.5 % as required by METIS. The thickness of the chromium layer finally influences the optimal choice of filling factor, since thin chromium layer transmits a certain amount of light in the considered wavelength range. A thin chromium layer also results in a larger differential optical path and a transmission which is more sensitive to small variations in the layer thickness. On the other hand, thick chromium layers are more complicated to manufacture.

A series of simulations has been performed exploring the parameter space spanned by the dot size, filling factor and chromium layer thickness. The results are summarized in Figure A.5 showing the behavior of transmission and phase step for the different parameters.

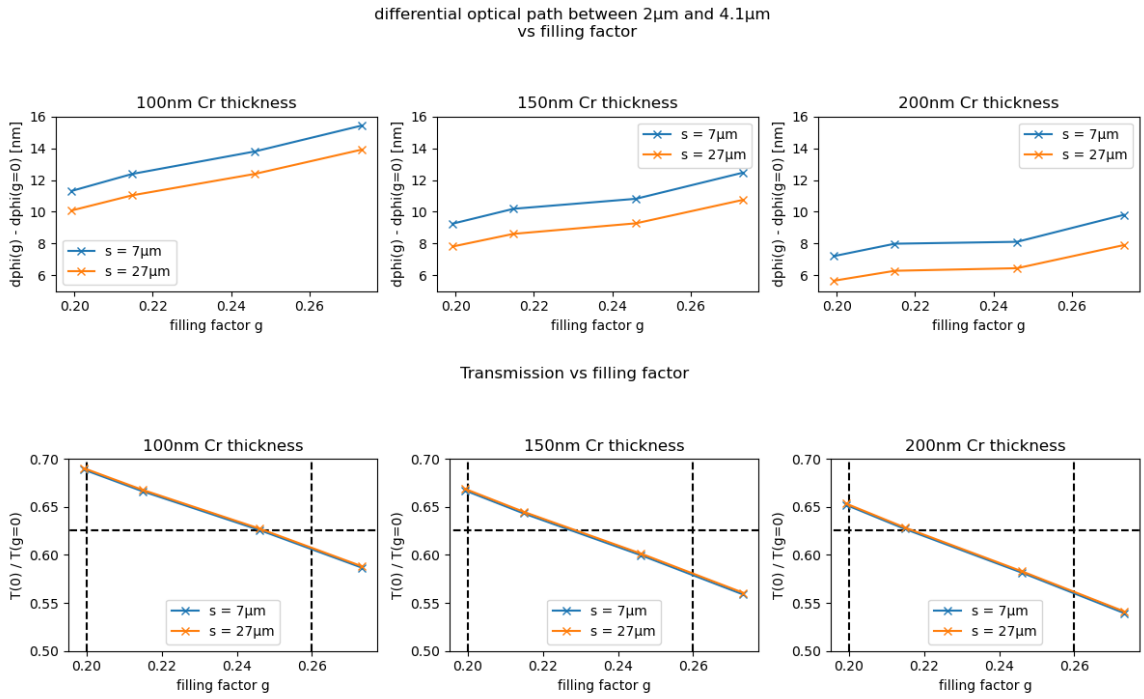

Figure A.5: Intensity transmission (top) and differential optical path between K and L/M-band (bottom) for microdot patterns with different parameters. The intensity transmission is averaged in the range of 3-5 µm corresponding to L and M-band, while the differential optical path is calculated between K (2 µm) and L/M-band (4.1 µm). Each subplot shows the results obtained for a certain chromium layer thickness. The differential optical path is always below the specification limit of 20 nm for METIS. For each chromium layer thickness, the nominal transmission of 62.5 % can be achieved by tuning the filling factor accordingly.

A critical effect of the ring apodizer to be quantified is the phase step between the clear and the greyscale region. The METIS specifications require that the differential optical path between K and LM-band should not be larger than 20 nm. This is due to the fact that the single conjugate adaptive optics (SCAO) system operates at K-band, while the science bands used with the ring apodizer are the L and M-bands (3-5 µm). Figure A.5 shows that the differential optical path increases with filling factor and decreases with chromium layer thickness. This means, a thick chromium layer and small filling factor result in a lower differential optical path between K and LM-bands. However, all parameters considered here are within the specifications of 20 nm.

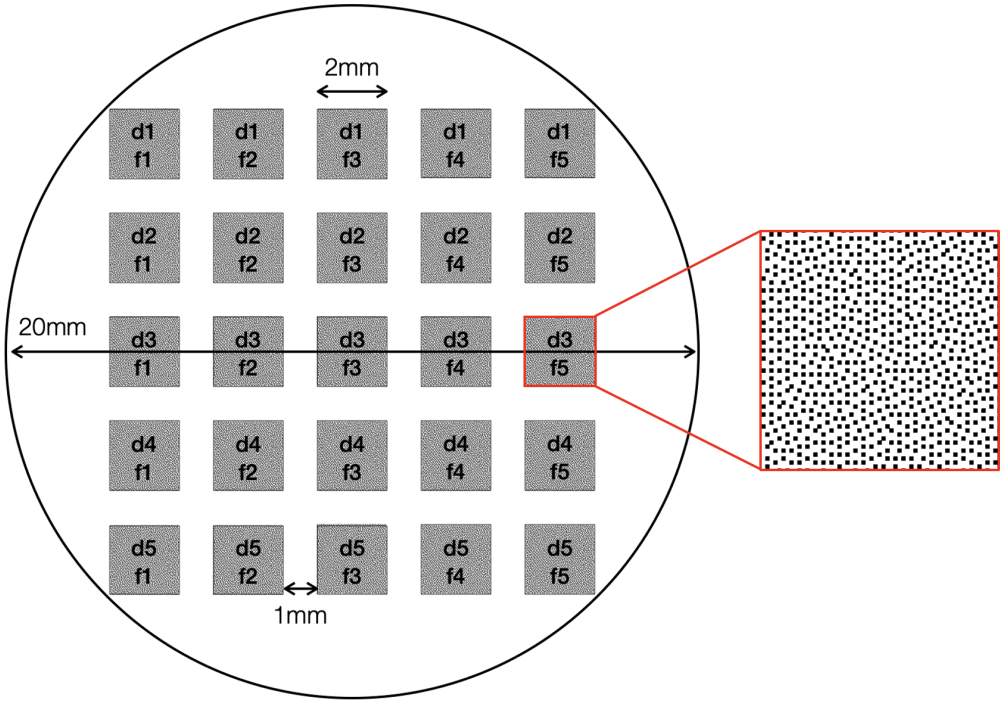

Figure A.6: Microdot test design to be manufactured on a 20 mm coated ZnSe substrate. The pattern consists of 25 patterns of size 2 mm × 2 mm with 1mm gaps used for reference measurements. Each column features a fixed filling factor f while each row features a fixed dot size d. Three substrates will be manufactured using different chromium layer thicknesses of 100, 150 and 200 nm.

A.4 A prototype featuring a series of test patterns

Though the RCWA simulation framework describes the general behavior of the microdot patterns sufficiently well, it is not fully suited to describe the exact transmission and phase response of the aperiodic semi-random pattern. It is therefore mandatory to experimentally test the simulation results before manufacturing the ring apodizer. For this purpose, a set of test patterns with varying dot size and filling factor are manufactured on 3 test substrates of varying chromium layer thickness. The parameters of the test samples are chosen to cover a certain range of the parameter space satisfying the requirements mentioned above and include a certain margin to compensate for imperfect simulation results and manufacturing tolerances. Each of the three test substrates has the same set of microdot test patterns imprinted on it. In this way, only one master pattern has to be produced, which is then transferred to the three test patterns with different chromium layer thickness. The smallest chromium thickness is chosen to be 100 nm, which is the thickness at which chromium becomes sufficiently opaque across the considered wavelength range (3-5 µm), and the highest value is chosen to be 200 nm, which is the maximum that can be provided by the manufacturer

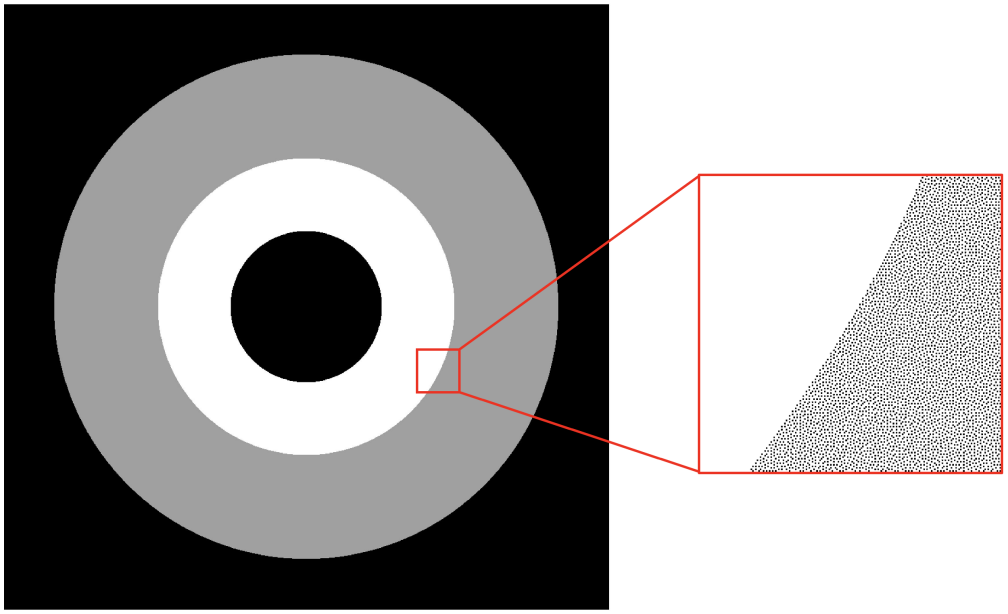

Figure A.7: METIS ring apodizer design. The microdots create the effect of a smooth greyscale pattern, while the inset shows the actual microdot pattern. The final microdot parameters will be chosen according to the experimental results obtained from the test patterns.

Optoline. Each of the substrates has the same design consisting of 25 test patterns arranged in a 5×5 grid featuring different combinations of dot size and filling factor. A schematic of the test pattern is shown in Figure A.6, including an inset showing the actual microdot pattern in one of the patterns. The parameter space covered in this test design ranges from 7 to 27 µm dot size and from 0.20 to 0.26 filling factor. Each pattern is 2 mm × 2 mm with a 1 mm gap between patterns used as clear reference to measure the normalized transmission and phase step.

The test substrates will be manufactured a few months before the ring apodizer to provide sufficient time to test their phase and transmission response. These tests will be carried out on the VODCA bench at ULiège and will allow to define an optimal set of parameters providing the desired transmission and fulfilling the phase step requirement defined by METIS. The final ring apodizer microdot pattern will then be created using these optimal parameters. The schematic design of the METIS ring apodizer is shown in Figure A.7 including an inset of the microdot pattern with the optimal simulation results. The final parameters will be defined after the test patterns have been characterized properly to account for simulation errors and manufacturing uncertainties.

Publications

Publications as first author

König, L., Absil, O., Delacroix, C., Lobet, M., Karlsson, M., Vargas Catalán, E., Orban de Xivry, G., Loicq, J. and Habraken, S.: "Vortex phase masks of topological charge 4 and higher with diamond subwavelength gratings", *Proceedings of the SPIE* **11451**, 1145144 (2020).

The work presented in Chapter 8. Main contributor to the publication.

König, L., Absil, O., Lobet, M., Delacroix, C., Karlsson, M., Orban de Xivry, G. and Loicq, J.: "Optimal design of the annular groove phase mask central region", *Optics Express* **30**(15), 27048–27063 (2022).

The work presented in Chapter 7. Main contributor to the publication.

König, L., Palatnick, S., Desai, N., Absil, O., Mawet, D., Millar-Blanchaer, M.: "Metasurface-based Scalar Vortex Phase Mask in pursuit of 1e-10 contrast", *Proceedings of the SPIE* **12680**, accepted (2023).

The work presented in Chapter 11. Main contributor to the publication.

Publications as co-author

Roy, N., König, L., Absil, O., Beauthier, C., Mayer, A. and Lobet, M.: "Photonics Structure Optimization Using Highly Data-Efficient Deep Learning: Application To Nanofins And Annular Groove Phase Masks", *ArXiv preprint*, arXiv:2309.01995 (2023).

The work presented in Chapter 9. Second author contributing the simulation framework of the AGPM and nanofins phase mask used by the optimization method developed by Nicolas Roy.

Palatnick, S., König, L., Millar-Blanchaer, M., Wallace, K., Absil, O., Mawet, D., Desai, N., Echeverri, D., John, D. and Schuller, J.: "Prospects for metasurfaces in exoplanet direct imaging systems: from principles to design", *Proceedings of the SPIE* **12680**, accepted (2023).

Based on the work presented in Chapter 11. Second author contributing significant ideas and discussions to the publication.

Bibliography

Abe, L., Vakili, F. and Boccaletti, A.: "The achromatic phase knife coronagraph", *Astronomy & Astrophysics* **374**(3), 1161–1168 (2001).

Absil, O., Delacroix, C., Orban de Xivry, G., Pathak, P., Willson, M., Berio, P., van Boekel, R., Matter, A., Defrère, D., Burtscher, L. et al.: "Impact of water vapor seeing on mid-infrared high-contrast imaging at ELT scale", *Proceedings of the SPIE* **12185**, 1218511 (2022).

Aime, C., Soummer, R. and Ferrari, A.: "Total coronagraphic extinction of rectangular apertures using linear prolate apodizations", *Astronomy & Astrophysics* **389**(1), 334–344 (2002).

Arbabi, A., Horie, Y., Bagheri, M. and Faraon, A.: "Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission", *Nature nanotechnology* **10**(11), 937–943 (2015).

Astro2020: The National Academies of Sciences, Engineering, and Medicine: "Pathways to Discovery in Astronomy and Astrophysics for the 2020s", *The National Academies Press, Washington, DC* (2021).

Berry, M.: "The adiabatic phase and Pancharatnam's phase for polarized light", *Journal of Modern Optics* **34**(11), 1401–1407 (1987).

Beuzit, J., Vigan, A., Mouillet, D., Dohlen, K., Gratton, R., Boccaletti, A., Sauvage, J., Schmid, H., Langlois, M., Petit, C. et al.: "SPHERE: the exoplanet imager for the Very Large Telescope", *Astronomy & Astrophysics* **631**, A155 (2019).

Beuzit, J., Demailly, L., Gendron, E., Gigan, P., Lacombe, F., Rouan, D., Hubin, N., Bonaccini, D., Prieto, E., Chazallet, F. et al.: "Adaptive Optics on a 3.6-Meter Telescope. The ADONIS System.", *Experimental Astronomy* **7**, 285–292 (1997).

Boccaletti, A., Cossou, C., Baudoz, P., Lagage, P., Dicken, D., Glasse, A., Hines, D., Aguilar, J., Detre, O., Nickson, B. et al.: "JWST/MIRI coronagraphic performances as measured on-sky", *Astronomy & Astrophysics* **667**, A165 (2022).

Bouhlel, M., Bartoli, N., Otsmane, A. and Morlier, J.: "Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction", *Structural and Multidisciplinary Optimization* **53**, 935–952 (2016).

Brandl, B., Bettonvil, F., van Boekel, R., Glauser, A., Quanz, S., Absil, O., Amorim, A., Feldt, M., Glasse, A., Güdel, M. et al.: "METIS: the mid-infrared ELT imager and spectrograph", *The Messenger* **182**, 22–26 (2021).

Carlomagno, B., Delacroix, C., Absil, O., Cantalloube, F., Orban de Xivry, G., Pathak, P., Agocs, T., Bertram, T., Brandl, B., Burtscher, L. et al.: "METIS high-contrast imaging: design and expected performance", *Journal of Astronomical Telescopes, Instruments, and Systems* **6**(3), 035005 (2020).

Charbonneau, D., Brown, T., Latham, D. and Mayor, M.: "Detection of planetary transits across a sun-like star", *The Astrophysical Journal* **529**(1), L45 (1999).

Chauvin, G., Lagrange, A., Dumas, C., Zuckerman, B., Mouillet, D., Song, I., Beuzit, J. and Lowrance, P.: "A giant planet candidate near a young brown dwarf-direct VLT/NACO observations using IR wavefront sensing", *Astronomy & Astrophysics* **425**(2), L29–L32 (2004).

Chen, T. and Guestrin, C.: "Xgboost: A scalable tree boosting system", *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* 785–794 (2016).

Chen, W., Zhu, A. and Capasso, F.: "Flat optics with dispersion-engineered metasurfaces", *Nature Reviews Materials* **5**(8), 604–620 (2020).

Chen, M., Lupoiu, R., Mao, C., Huang, D., Jiang, J., Lalanne, P. and Fan, J.: "WaveY-Net: physics-augmented deep-learning for high-speed electromagnetic simulation and optimization", *Proceedings of the SPIE* **12011**, 120110C (2022).

Chen, W., Park, J., Marchioni, J., Millay, S., Yousef, K. and Capasso, F.: "Dispersion-engineered metasurfaces reaching broadband 90% relative diffraction efficiency", *Nature Communications* **14**(1), 2544 (2023).

Chong, K., Staude, I., James, A., Dominguez, J., Liu, S., Campione, S., Subramania, G., Luk, T., Decker, M., Neshev, D. et al.: "Polarization-Independent Silicon Metadevices for Efficient Optical Wavefront Control", *Nano Letters* **15**(8), 5369–5374 (2015).

Chung, H. and Miller, O.: "High-NA achromatic metalenses by inverse design", *Optics Express* **28**(5), 6945–6965 (2020).

Davies, R., Hörmann, V., Rabien, S., Sturm, E., Alves, J., Clénet, Y., Kotilainen, J., Lang-Bardl, F., Nicklas, H., Pott, J. et al.: "MICADO: the multi-adaptive optics camera for deep observations", *The Messenger* **182**, 17–22 (2021).

Davies, R., Absil, O., Agapito, G., Berbel, A., Baruffolo, A., Biliotti, V., Black, M., Bonaglia, M., Bonse, M., Briguglio, R. et al.: "The Enhanced Resolution Imager and Spectrograph for the VLT", *Astronomy & Astrophysics* **674**, A207 (2023).

Defrère, D., Absil, O., Hinz, P., Kuhn, J., Mawet, D., Mennesson, B., Skemer, A., Wallace, K., Bailey, V., Downey, E. et al.: "L'-band AGPM vector vortex coronagraph's first light on LBTI/LMIRCam", *Proceedings of the SPIE* **9148**, 91483X (2014).

Delacroix, C., Absil, O., Mawet, D., Hanot, C., Karlsson, M., Forsberg, P., Pantin, E., Surdej, J. and Habraken, S.: "A diamond AGPM coronagraph for VISIR", *Proceedings of the SPIE* **8446**, 84468K (2012).

Delacroix, C., Absil, O., Forsberg, P., Mawet, D., Christiaens, V., Karlsson, M., Boccaletti, A., Baudoz, P., Kuittinen, M., Vartiainen, I. et al.: "Laboratory demonstration of a mid-infrared AGPM vector vortex coronagraph", *Astronomy & Astrophysics* **553**, A98 (2013).

Delacroix, C.: "Exoplanet imaging with mid-infrared vector vortex coronagraphs: design, manufacture, validation and first light of the annular groove phase mask", PhD thesis, University of Liège (2013).

Delacroix, C., Absil, O., Carlomagno, B., Piron, P., Forsberg, P., Karlsson, M., Mawet, D., Habraken, S. and Surdej, J.: "Development of a subwavelength grating vortex coronagraph of topological charge 4 (SGVC4)", *Proceedings of the SPIE* **9147**, 2901–2909 (2014).

Delacroix, C., Absil, O., Orban de Xivry, G., Shinde, M., Pathak, P., Cantalloube, F., Carlomagno, B., Christiaens, V., Boné, A., Dolkens, D. et al.: "The High-contrast End-to-End Performance Simulator (HEEPS): influence of ELT/METIS instrumental effects", *Proceedings of the SPIE* **12187**, 121870F (2022).

Delbeke, D.: "Design and fabrication of a highly efficient light-emitting diode: the Grating-Assisted Resonant-Cavity Light-Emitting Diode", PhD thesis, Ghent University (2002).

Deparis, O., Khuzayim, N., Parker, A. and Vigneron, J.: "Assessment of the antireflection property of moth wings by three-dimensional transfer-matrix optical simulations", *Physical Review E* **79**(4), 041910 (2009).

Desai, N., Llop-Sayson, J., Jovanovic, N., Ruane, G., Serabyn, E., Martin, S. and Mawet, D.: "High contrast demonstrations of novel scalar vortex coronagraph designs at the high contrast spectroscopy testbed", *Proceedings of the SPIE* **11823**, 118230Q (2021).

Desai, N., Llop-Sayson, J., Bertrou-Cantou, A., Ruane, G., Riggs, A., Serabyn, E. and Mawet, D.: "Topological designs for scalar vortex coronagraphs", *Proceedings of the SPIE* **12180**, 121805H (2022).

Desai, N., Ruane, G., Llop-Sayson, J., Betrou-Cantou, A., Potier, A., Riggs, A., Serabyn, E. and Mawet, D.: "Laboratory demonstration of the wrapped staircase scalar vortex coronagraph", *Journal of Astronomical Telescopes, Instruments, and Systems* **9**(2), 025001 (2023).

Devlin, R., Ambrosio, A., Wintz, D., Oscurato, S., Zhu, A., Khorasaninejad, M., Oh, J., Maddalena, P. and Capasso, F.: "Spin-to-orbital angular momentum conversion in dielectric metasurfaces", *Optics Express* **25**(1), 377–393 (2017).

Dhoedt, B.: "Theoretical and Experimental Study of Free Space Optical Interconnections Based on Diffractive Lens Arrays", PhD thesis, Ghent University (1995).

Fischer, D., Howard, A., P, L., Macintosh, B., Mahadevan, S., Sahlmann, J. and Yee, J.: "Exoplanet detection techniques", *Protostars and Planets* **VI**, 715–737 (2014).

Floyd, R. and Steinberg, L.: "An Adaptive Algorithm for Spatial Greyscale", *Proceedings of the Society for Information Display* **17**(2), 75–77 (1976).

Forsberg, P. and Karlsson, M.: "Inclined surfaces in diamond: broadband antireflective structures and coupling light through waveguides", *Optics Express* **21**(3), 2693–2700 (2013b).

Forsberg, P. and Karlsson, M.: "High aspect ratio optical gratings in diamond", *Diamond and Related Materials* **34**, 19–24 (2013a).

Freedman, D., Pisani, R. and Purves, R.: "Statistics: Fourth International Student Edition", *W.W. Norton & Company* (2007).

Galicher, R., Huby, E., Baudoz, P. and Dupuis, O.: "A family of phase masks for broadband coronagraphy example of the wrapped vortex phase mask theory and laboratory demonstration", *Astronomy & Astrophysics* **635**, A11 (2020).

Gaudi, B., Seager, S., Mennesson, B., Kiessling, A., Warfield, K., Cahoy, K., Clarke, J., Domagal-Goldman, S., Feinberg, L., Guyon, O. et al.: "The Habitable Exoplanet Observatory (HabEx) mission concept study final report", *arXiv preprint* arXiv:2001.06683 (2020).

Gilmozzi, R. and Spyromilio, J.: "The European extremely large telescope (E-ELT)", *The Messenger* **127**(11), 3 (2007).

Give'on, A., Kern, B., Shaklan, S., Moody, D. and Pueyo, L.: "Electric field conjugation-a broadband wavefront correction algorithm for high-contrast imaging systems", *American Astronomical Society Meeting Abstracts* **211**, 135–20 (2007).

Grady, C., Proffitt, C., Malumuth, E., Woodgate, B., Gull, T., Bowers, C., Heap, S., Kimble, R., Lindler, D., Plait, P. et al.: "Coronagraphic Imaging with the Hubble Space Telescope and the Space Telescope Imaging Spectrograph1", *Publications of the Astronomical Society of the Pacific* **115**(811), 1036 (2003).

Hedlund, C., Blom, H. and Berg, S.: "Microloading effect in reactive ion etching", *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* **12**(4), 1962–1965 (1994).

Heiden, J. and Jang, M.: "Design framework for polarization-insensitive multifunctional achromatic metalenses", *Nanophotonics* **11**(3), 583–591 (2022).

Hénault, F.: "Strehl ratio: a tool for optimizing optical nulls and singularities", *Journal of the Optical Society of America A* **32**(7), 1276–1287 (2015).

Hénault, F., Carlotti, A. and Vérinaud, C.: "Analysis of nulling phase functions suitable to image plane coronagraphy", *Proceedings of the SPIE* **9912**, 99126K (2016).

Hénault, F.: "Analysis of azimuthal phase mask coronagraphs", *Optics Communications* **423**, 186–199 (2018).

Henry, G., Marcy, G., Butler, R. and Vogt, S.: "A transiting "51 Peg-like" planet", *The Astrophysical Journal* **529**(1), L41 (1999).

Hodapp, K., Jensen, J., Irwin, E., Yamada, H., Chung, R., Fletcher, K., Robertson, L., Hora, J., Simons, D., Mays, W. et al.: "The Gemini Near-Infrared Imager (NIRI)", *Publications of the Astronomical Society of the Pacific* **115**(814), 1388 (2003).

Huby, E., Baudoz, P., Mawet, D. and Absil, O.: "Post-coronagraphic tip-tilt sensing for vortex phase masks: The QACITS technique", *Astronomy & Astrophysics* **584**, A74 (2015).

Hugonin, J. and Lalanne, P.: "Reticolo software for grating analysis", *arXiv preprint* arXiv:2101.00901 (2005).

Jiang, Z., Yun, S., Lin, L., Bossard, J., Werner, D. and Mayer, T.: "Tailoring dispersion for broadband low-loss optical metamaterials using deep-subwavelength inclusions", *Scientific reports* **3**(1), 1571 (2013).

Jolivet, A., Orban de Xivry, G., Huby, E., Piron, P., Vargas Catalán, E., Habraken, S., Surdej, J., Karlsson, M. and Absil, O.: "L- and M-band annular groove phase mask in lab performance assessment on the vortex optical demonstrator for coronagraphic applications", *Journal of Astronomical Telescopes, Instruments, and Systems* **5**(2), 025001 (2019).

Jolivet, A.: "Development and exploitation of an infrared coronagraphic test bench for vortex phase mask performance assessment", PhD thesis, University of Liège (2019).

Kasdin, N., Vanderbei, R., Spergel, D. and Littman, M.: "Extrasolar planet finding via optimal apodized-pupil and shaped-pupil coronagraphs", *The Astrophysical Journal* **582**(2), 1147 (2003).

Kasdin, N., Bailey, V., Mennesson, B., Zellem, R., Ygouf, M., Rhodes, J., Luchik, T., Zhao, F., Riggs, A., Seo, B. et al.: "The Nancy Grace Roman space telescope coronagraph instrument (CGI) technology demonstration", *Proceedings of the SPIE* **11443**, 114431U (2020).

Kasper, M. and Mawet, D.: "Gearing up the SPHERE", *The Messenger* **149**(9), 17–21 (2012).

Kasper, M., Urra, N., Pathak, P., Bonse, M., Nousiainen, J., Engler, B., Heritier, C., Kammerer, J., Leveratto, S., Rajani, C. et al.: "PCS–A Roadmap for Exoearth Imaging with the ELT", *The Messenger* **182**, 38–43 (2021).

Kennedy, J. and Eberhart, R.: "Particle swarm optimization", *Proceedings of ICNN'95-international conference on neural networks* **4**, 1942–1948 (1995).

Kenworthy, M., Codona, J., Hinz, P., Angel, J., Heinze, A. and Sivanandam, S.: "First on-sky high-contrast imaging with an apodizing phase plate", *The Astrophysical Journal* **660**(1), 762 (2007).

Keppler, M., Benisty, M., Müller, A., Henning, T., van Boekel, R., Cantalloube, F., Ginski, C., van Holstein, R., Maire, A., Pohl, A. et al.: "Discovery of a planetary-mass companion within the gap of the transition disk around PDS 70", *Astronomy & Astrophysics* **617**, A44 (2018).

Khorasaninejad, M., Chen, W., Devlin, R., Oh, J., Zhu, A. and Capasso, F.: "Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging", *Science* **352**(6290), 1190–1194 (2016).

Kim, E., Xia, Y., Zhao, X. and Whitesides, G.: "Solvent-Assisted Microcontact Molding: A Convenient Method for Fabricating Three-Dimensional Structures on Surfaces of Polymers", *Advanced Materials* **9**, 651–654 (1997).

Kochkov, D., Smith, J., Alieva, A., Wang, Q., Brenner, M. and Hoyer, S.: "Machine learning–accelerated computational fluid dynamics", *Proceedings of the National Academy of Sciences* **118**(21), e2101784118 (2021).

König, L., Absil, O., Delacroix, C., Lobet, M., Karlsson, M., Vargas Catalán, E., Orban de Xivry, G., Loicq, J. and Habraken, S.: "Vortex phase masks of topological charge 4 and higher with diamond subwavelength gratings", *Proceedings of the SPIE* **11451**, 1145144 (2020).

König, L., Absil, O., Lobet, M., Delacroix, C., Karlsson, M., Orban de Xivry, G. and Loicq, J.: "Optimal design of the annular groove phase mask central region", *Optics Express* **30**(15), 27048–27063 (2022).

Kruk, S., Ferreira, F., Mac Suibhne, N., Tsekrekos, C., Kravchenko, I., Ellis, A., Neshev, D., Turitsyn, S. and Kivshar, Y.: "Transparent dielectric metasurfaces for spatial mode multiplexing", *Laser & Photonics Reviews* **12**(8), 1800031 (2018).

Kuhn, J., Potter, D. and Parise, B.: "Imaging polarimetric observations of a new circumstellar disk system", *The Astrophysical Journal* **553**(2), L189 (2001).

Lagrange, A., Gratadour, D., Chauvin, G., Fusco, T., Ehrenreich, D., Mouillet, D., Rousset, G., Rouan, D., Allard, F., Gendron, ėt al.: "A probable giant planet imaged in the beta Pictoris disk - VLT/NaCo deep L'-band imaging", *Astronomy & Astrophysics* **493**(2), L21–L25 (2009).

Lärmer, F. and Schilp, A.: "Method of anisotropically etching silicon", German patent 4241045 (1996).

Li, X. and Fan, Z.: "Controlling Dispersion Characteristic of Focused Vortex Beam Generation", *Photonics* **9**(3), 179 (2022).

Liu, V. and Fan, S.: "S^4 : A free electromagnetic solver for layered periodic structures", *Computer Physics Communications* **183**(10), 2233–2244 (2012).

Liu, X., Deng, J., Li, K., Jin, M., Tang, Y., Zhang, X., Cheng, X., Wang, H., Liu, W. and Li, G.: "Optical telescope with Cassegrain metasurfaces", *Nanophotonics* **9**(10), 3263–3269 (2020).

Lozi, J., Guyon, O., Jovanovic, N., Goebel, S., Pathak, P., Skaf, N., Sahoo, A., Norris, B., Martinache, F., N'Diaye, M. et al.: "SCExAO, an instrument with a dual purpose: perform cutting-edge science and develop new technologies", *Proceedings of the SPIE* **10703**, 1266–1277 (2018).

The LUVOIR Team: "The LUVOIR mission concept study final report", *arXiv preprint* arXiv:1912.06219 (2019).

Lyot, B.: "A study of the solar corona and prominences without eclipses", *Monthly Notices of the Royal Astronomical Society* **99**, 580 (1939).

Macintosh, B., Graham, J., Palmer, D., Doyon, R., Dunn, J., Gavel, D., Larkin, J., Oppenheimer, B., Saddlemyer, L., Sivaramakrishnan, A. et al.: "The Gemini Planet Imager: from science to design to construction", *Proceedings of the SPIE* **7015**, 701518 (2008).

Macintosh, B., Graham, J., Ingraham, P., Konopacky, Q., Marois, C., Perrin, M., Poyneer, L., Bauman, B., Barman, T., Burrows, A. et al.: "First light of the gemini planet imager", *Proceedings of the National Academy of Sciences* **111**(35), 12661–12666 (2014).

Malbet, F., Yu, J. and Shao, M.: "High-dynamic-range imaging using a deformable mirror for space coronography", *Publications of the Astronomical Society of the Pacific* **107**(710), 386 (1995).

Marconi, A., Abreu, M., Adibekyan, V., Alberti, V., Albrecht, S., Alcaniz, J., Aliverti, M., Prieto, C., Gómez, J., Amado, P. et al.: "ANDES, the high resolution spectrograph for the ELT: science case, baseline design and path to construction", *Proceedings of the SPIE* **12184**, 1218424 (2022).

Marois, C., Doyon, R., Racine, R. and Nadeau, D.: "Efficient speckle noise attenuation in faint companion imaging", *Publications of the Astronomical Society of the Pacific* **112**(767), 91 (2000).

Marois, C., Lafreniere, D., Doyon, R., Macintosh, B. and Nadeau, D.: "Angular differential imaging: a powerful high-contrast imaging technique", *The Astrophysical Journal* **641**(1), 556 (2006).

Martinez, P., Dorrer, C., Carpentier, E., Kasper, M., Boccaletti, A., Dohlen, K. and Yaitskova, N.: "Design, analysis, and testing of a microdot apodizer for the Apodized Pupil Lyot Coronagraph", *Astronomy & Astrophysics* **495**(1), 363–370 (2009).

Mawet, D., Riaud, P., Absil, O. and Surdej, J.: "Annular Groove Phase Mask Coronagraph", *The Astrophysical Journal* **633**(2), 1191–1200 (2005a).

Mawet, D., Riaud, P., Surdej, J. and Baudrand, J.: "Subwavelength surface-relief gratings for stellar coronagraphy", *Applied Optics* **44**(34), 7313–7321 (2005b).

Mawet, D.: "Subwavelength gratings for extrasolar planetary system imaging and characterization", PhD thesis, University of Liège (2006).

Mawet, D., Serabyn, E., Liewer, K., Hanot, C., McEldowney, S., Shemo, D. and O'Brien, N.: "Optical vectorial vortex coronagraphs using liquid crystal polymers: theory, manufacturing and laboratory demonstration", *Optics Express* **17**(3), 1902–1918 (2009).

Mawet, D., Pueyo, L., Moody, D., Krist, J. and Serabyn, E.: "The Vector Vortex Coronagraph: sensitivity to central obscuration, low-order aberrations, chromaticism, and polarization", *Proceedings of the SPIE* **7739**, 773914 (2010).

Mawet, D., Murakami, N., Delacroix, C., Serabyn, E., Absil, O., Baba, N., Baudrand, J., Boccaletti, A., Burruss, R., Chipman, R. et al.: "Taking the vector vortex coronagraph to the next level for ground-and space-based exoplanet imaging instruments: review of technology developments in the USA, Japan, and Europe", *Proceedings of the SPIE* **8151**, 815108 (2011).

Mawet, D., Absil, O., Delacroix, C., Girard, J., Milli, J., O'Neal, J., Baudoz, P., Boccaletti, A., Bourget, P., Christiaens, V. et al.: "L'-band AGPM vector vortex coronagraph's first light on VLT/NACO. Discovery of a late-type companion at two beamwidths from an F0V star", *Astronomy & Astrophysics* **552**, L13 (2013a).

Mawet, D., Pueyo, L., Carlotti, A., Mennesson, B., Serabyn, E. and Wallace, J.: "Ring-apodized vortex coronagraphs for obscured telescopes. I. Transmissive ring apodizers", *The Astrophysical Journal Supplement Series* **209**(1), 7 (2013b).

Mayor, M. and Queloz, D.: "A Jupiter-mass companion to a solar-type star", *Nature* **378**(6555), 355–359 (1995).

Moharam, M. and Gaylord, T.: "Rigorous coupled-wave analysis of planar-grating diffraction", *Journal of the Optical Society of America* **71**(7), 811–818 (1981).

Murakami, N., Nishikawa, J., Yokochi, K., Tamura, M., Baba, N. and Abe, L.: "Achromatic eight-octant phase-mask coronagraph using photonic crystal", *The Astrophysical Journal* **714**(1), 772 (2010).

Murakami, N., Nishikawa, J., Traub, W., Mawet, D., Moody, D., Kern, B., Trauger, J., Serabyn, E., Hamaguchi, S., Oshiyama, F. et al.: "Coronagraph focal-plane

phase masks based on photonic crystal technology: recent progress and observational strategy", *Proceedings of the SPIE* **8442**, 844205 (2012).

Oskooi, A., Roundy, D., Ibanescu, M., Bermel, P., Joannopoulos, J. and Johnson, S.: "MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method", *Computer Physics Communications* **181**(3), 687–702 (2010).

Pancharatnam, S.: "Generalized theory of interference, and its applications: Part I. Coherent pencils", *Proceedings of the Indian Academy of Sciences-Section A* **44**(5), 247–262 (1956).

Por, E., Haffert, S., Radhakrishnan, V., Doelman, D., van Kooten, M. and Bos, S.: "High Contrast Imaging for Python (HCIPy): an open-source adaptive optics and coronagraph simulator", *Proceedings of the SPIE* **10703**, 1070342 (2018).

Powell, M.: "The theory of radial basis function approximation in 1990", *Advances in numerical analysis* **2**, 105–210 (1992).

Quesnel, M., Orban de Xivry, G., Louppe, G. and Absil, O.: "A deep learning approach for focal-plane wavefront sensing using vortex phase diversity", *Astronomy & Astrophysics* **668**, A36 (2022).

Rickard, A. and McNie, M.: "Characterization and optimization of deep dry etching for MEMS applications", *Proceedings of the SPIE* **4407**, 78–88 (2001).

Rigaut, F., Rousset, G., Kern, P., Fontanella, J., Gaffard, J., Merkle, F. and Léna, P.: "Adaptive optics on a 3.6-m telescope-Results and performance", *Astronomy and Astrophysics* **250**, 280–290 (1991).

Riggs, A., Ruane, G., Sidick, E., Coker, C., Kern, B. and Shaklan, S.: "Fast linearized coronagraph optimizer (FALCO) I: a software toolbox for rapid coronagraphic design and wavefront correction", *Proceedings of the SPIE* **10698**, 878–888 (2018).

Roddier, F. and Roddier, C.: "Stellar coronograph with phase mask", *Publications of the Astronomical Society of the Pacific* **109**(737), 815 (1997).

Ronayette, S., Mouzali, S., Barrière, J., Pantin, E., Orduna, T., Gallais, P., Lortholary, M., Dumaye, L., Absil, O., Delacroix, C. et al.: "An N-band test bench for the METIS coronagraphic masks", *Proceedings of the SPIE* **11447**, 597–611 (2020).

Ronneberger, O., Fischer, P. and Brox, T.: "U-net: Convolutional networks for biomedical image segmentation", *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015* **3**(18), 234–241 (2015).

Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y. and Labeyrie, A.: "The four-quadrant phase-mask coronagraph. I. Principle", *Publications of the Astronomical Society of the Pacific* **112**(777), 1479 (2000).

Rouan, D., Baudrand, J., Boccaletti, A., Baudoz, P., Mawet, D. and Riaud, P.: "The four quadrant phase mask coronagraph and its avatars", *Comptes Rendus Physique* **8**(3-4), 298–311 (2007).

Rousset, G., Lacombe, F., Puget, P., Hubin, N., Gendron, E., Fusco, T., Arsenault, R., Charton, J., Feautrier, P., Gigan, P. et al.: "NAOS–the first AO system of the VLT: on-sky performance", *Proceedings of the SPIE* **4839**, 140–149 (2003).

Roy, N., König, L., Absil, O., Beauthier, C., Mayer, A. and Lobet, M.: "Photonics Structure Optimization Using Highly Data-Efficient Deep Learning: Application To Nanofins And Annular Groove Phase Masks", *arXiv preprint* (2023).

Ruane, G., Mawet, D., Mennesson, B., Jewell, J. and Shaklan, S.: "Vortex coronagraphs for the Habitable Exoplanet Imaging Mission concept: theoretical performance and telescope requirements", *Journal of Astronomical Telescopes, Instruments, and Systems* **4**(1), 015004 (2018).

Ruane, G., Mawet, D., Riggs, A. and Serabyn, E.: "Scalar vortex coronagraph mask design and predicted performance", *Proceedings of the SPIE* **11117**, 454–469 (2019).

Ruane, G., Riggs, A., Serabyn, E., Baxter, W., Mejia Prada, C., Mawet, D., Noyes, M., Poon, P. and Tabiryan, N.: "Broadband vector vortex coronagraph testing at NASA's high contrast imaging testbed facility", *Proceedings of the SPIE* **12180**, 1218024 (2022).

Rubin, N., Chevalier, P., Juhl, M., Tamagnone, M., Chipman, R. and Capasso, F.: "Imaging polarimetry through metasurface polarization gratings", *Optics Express* **30**(6), 9389–9412 (2022).

Sánchez-Janssen, R., Hammer, F., Morris, S., Cuby, J., Kaper, L., Steinmetz, M., Afonso, J., Barbuy, B., Bergin, E., Finoguenov, A. et al.: "MOSAIC: the high multiplex and multi-IFU spectrograph for the ELT", *Proceedings of the SPIE* **11447**, 1144725 (2020).

Serabyn, E., Huby, E., Matthews, K., Mawet, D., Absil, O., Femenia, B., Wizinowich, P., Karlsson, M., Bottom, M., Campbell, R. et al.: "The WM Keck observatory infrared vortex coronagraph and a first image of HIP 79124 B", *The Astronomical Journal* **153**(1), 43 (2017).

Serabyn, E., Ruane, G. and Echeverri, D.: "Observing inside the coronagraphic regime with optimized single-mode nulling interferometry", *Proceedings of the SPIE* **12180**, 121800N (2022).

Shalaev, M., Sun, J., Tsukernik, A., Pandey, A., Nikolskiy, K. and Litchinitser, N.: "High-Efficiency All-Dielectric Metasurfaces for Ultracompact Beam Manipulation in Transmission Mode", *Nano Letters* **15**(9), 6261–6266 (2015).

Shorten, C. and Khoshgoftaar, T.: "A survey on image data augmentation for deep learning", *Journal of big data* **6**(1), 1–48 (2019).

Shrestha, S., Overvig, A., Lu, M., Stein, A. and Yu, N.: "Broadband achromatic dielectric metalenses", *Light: Science & Applications* **7**(1), 85 (2018).

Soummer, R., Aimé, C. and Falloon, P.: "Stellar coronagraphy with prolate apodized circular apertures", *Astronomy & Astrophysics* **397**(3), 1161–1172 (2003a).

Soummer, R., Dohlen, K. and Aimé, C.: "Achromatic dual-zone phase mask stellar coronagraph", *Astronomy & Astrophysics* **403**(1), 369–381 (2003b).

Soummer, R., Pueyo, L., Sivaramakrishnan, A. and Vanderbei, R.: "Fast computation of Lyot-style coronagraph propagation", *Optics Express* **15**(24), 15935–15951 (2007).

Sparks, W. and Ford, H.: "Imaging spectroscopy for extrasolar planet detection", *The Astrophysical Journal* **578**(1), 543 (2002).

Sun, G. and Wang, S.: "A review of the artificial neural network surrogate modeling in aerodynamic design", *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* **233**(16), 5863–5872 (2019).

Sun, T., Hu, J., Ma, S., Xu, F. and Wang, C.: "Polarization-insensitive achromatic metalens based on computational wavefront coding", *Optics Express* **29**(20), 31902–31914 (2021).

Swartzlander Jr, G.: "Achromatic optical vortex lens", *Optics Letters* **31**(13), 2042–2044 (2006).

Tabirian, N., Xianyu, H. and Serabyn, E.: "High efficiency broadband liquid crystal polymer vector vortex waveplates", *Proceedings of the 2017 IEEE Aerospace Conference* 1–7 (2017).

Taflove, A. and Hagness, S.: "Computational Electrodynamics: The Finite-Difference Time-Domain Method, 3rd edition.", *The Electrical Engineering Handbook* **2062** (2005).

Thatte, N., Tecza, M., Schnetler, H., Neichel, B., Melotte, D., Fusco, T., Ferraro-Wood, V., Clarke, F., Bryson, I., O'Brien, K. et al.: "HARMONI: the ELT's First-Light Near-infrared and Visible Integral Field Spectrograph", *The Messenger* **182**, 7–12 (2021).

Trauger, J., Moody, D., Gordon, B., Krist, J. and Mawet, D.: "A hybrid Lyot coronagraph for the direct imaging and spectroscopy of exoplanet systems: recent results and prospects", *Proceedings of the SPIE* **8151**, 81510G (2011).

Ulichney, R.: "Dithering with blue noise", *Proceedings of the IEEE* **76**(1), 56–79 (1988).

UNESCO: "UNESCO Recommendation on Open Science", doi:10.5281/zenodo.5834767 (2021).

Vargas Catalán, E., Huby, E., Forsberg, P., Jolivet, A., Baudoz, P., Carlomagno, B., Delacroix, C., Habraken, S., Mawet, D., Surdej, J. et al.: "Optimizing the

subwavelength grating of L-band annular groove phase masks for high coronagraphic performance", *Astronomy & Astrophysics* **595**, A127 (2016b).

Vargas Catalán, E., Forsberg, P., Absil, O. and Karlsson, M.: "Controlling the profile of high aspect ratio gratings in diamond", *Diamond and Related Materials* **63**, 60–68 (2016a).

Vargas Catalán, E., Piron, P., Jolivet, A., Forsberg, P., Delacroix, C., Huby, E., Absil, O., Vartiainen, I., Kuittinen, M. and Karlsson, M.: "Subwavelength diamond gratings for vortex coronagraphy: towards an annular groove phase mask for shorter wavelengths and topological charge 4 designs", *Optical Materials Express* **8**(7), 1976–1987 (2018).

Vigneron, J. and Lousse, V.: "Variation of a photonic crystal color with the Miller indices of the exposed surface", *Proceedings of the SPIE* **6128**, 325–334 (2006).

Wizinowich, P., Le Mignant, D., Bouchez, A., Campbell, R., Chin, J., Contos, A., van Dam, M., Hartman, S., Johansson, E., Lafon, R. et al.: "The WM Keck Observatory laser guide star adaptive optics system: overview", *Publications of the Astronomical Society of the Pacific* **118**(840), 297 (2006).

Wolszczan, A. and Frail, D.: "A planetary system around the millisecond pulsar PSR1257+ 12", *Nature* **355**(6356), 145–147 (1992).

Yu, N., Genevet, P., Kats, M., Aieta, F., Tetienne, J., Capasso, F. and Gaburro, Z.: "Light propagation with phase discontinuities: generalized laws of reflection and refraction", *Science* **334**(6054), 333–337 (2011).

Yu, N., Aieta, F., Genevet, P., Kats, M., Gaburro, Z. and Capasso, F.: "A Broadband, Background-Free Quarter-Wave Plate Based on Plasmonic Metasurfaces", *Nano Letters* **12**(12), 6328–6333 (2012).

Yu, N. and Capasso, F.: "Flat optics with designer metasurfaces", *Nature Materials* **13**(2), 139–150 (2014).

Zhang, X. and Wu, Y.: "Effective medium theory for anisotropic metamaterials", *Scientific reports* **5**(1), 7892 (2015).

Zhang, M., Ruane, G., Delorme, J., Mawet, D., Jovanovic, N., Jewell, J., Shaklan, S. and Wallace, J.: "Characterization of microdot apodizers for imaging exoplanets with next-generation space telescopes", *Proceedings of the SPIE* **10698**, 1688–1698 (2018).

Zhou, R., Zhang, H., Hao, Y. and Wang, Y.: "Simulation of the Bosch process with a string–cell hybrid method", *Journal of micromechanics and microengineering* **14**(7), 851 (2004).