Universiteit Leiden
The Netherlands

It's just a phase: high-contrast imaging with patterned liquid-crystal phase plates to facilitate characterization of exoplanets

Doelman, D.S.

Citation

Doelman, D. S. (2021, June 22). *It's just a phase: high-contrast imaging with patterned liquid-crystal phase plates to facilitate characterization of exoplanets*. Retrieved from https://hdl.handle.net/1887/3191978

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3191978

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University Repository

The handle https://hdl.handle.net/1887/3191978 holds various files of this Leiden University dissertation.

Author: Doelman, D.S.
Title: It's just a phase: high-contrast imaging with patterned liquid-crystal phase plates to facilitate characterization of exoplanets
Issue Date: 2021-06-22

It's just a phase

High-contrast imaging with patterned liquid-crystal phase plates to facilitate characterization of exoplanets

David Sebastiaan Doelman

It's just a phase

High-contrast imaging with patterned liquid-crystal phase plates to facilitate characterization of exoplanets

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 22 juni 2021
klokke 15.00 uur

door

David Sebastiaan Doelman

geboren te
Amsterdam, Nederland
in 1993

Promotor:	Prof. dr. C. Keller	
Co-promotor:	Dr.ir. F. Snik	
Promotiecommissie:	Prof. dr. Huub Röttgering	Universiteit Leiden
	Prof. dr. Ignas Snellen	Universiteit Leiden
	Dr.ir. S. Pereira	Technische Universiteit Delft
	Prof. dr. P. Tuthill	University of Sydney
	Dr. E. Huby	Observatoire de Paris

Dedicated to my grandfather,
Corstiaan Arie Drogendijk

Front cover:
Artist impression of the detection of a gas giant using the vector-Apodizing Phase Plate coronagraph at installed in the Subaru/SCExAO instrument. The phase pattern is encoded in the orientation of white lines.
Image of the gas giant by Danielle Futselaar and Franck Marchis.
Cover design by the author.
ISBN: 978-94-6419-239-1
An electronic copy of this thesis can be found at https://openaccess.leidenuniv.nl
© David S.Doelman, 2021

Contents

1 Introduction **1**
1.1 The era of exoplanet characterization . . . 3
1.1.1 Methods of exoplanet characterization . . . 3
1.1.2 Direct imaging of HR 8799 . . . 4
1.2 Direct imaging systems . . . 7
1.2.1 The challenge of direct imaging . . . 7
1.2.2 A modular view of high-contrast imaging . . . 8
1.3 Ground-based and space-based high-contrast imaging . . . 20
1.3.1 The compatibility of ground and space . . . 20
1.3.2 The current state and limitations . . . 22
1.4 Broadband enabling technologies . . . 24
1.4.1 Properties of the geometric phase . . . 25
1.4.2 Geometric phase in the Jones formalism . . . 28
1.4.3 A different look on geometric phase: a personal note . . . 31
1.4.4 Technical implementation of a GPH . . . 35
1.5 This thesis . . . 39
1.6 Future Outlook . . . 43
1.6.1 A ground-based perspective . . . 43
1.6.2 A space-based perspective . . . 48
1.6.3 New applications of liquid-crystal technology for exoplanet detection . . . 50

2 The world of vAPP **61**
2.1 Introduction . . . 62
2.2 Design of a gvAPP . . . 65
2.2.1 Dark zone considerations . . . 66
2.2.2 Optimization of the APP design . . . 69
2.2.3 Adding functionality with holograms . . . 70
2.2.4 Additional design choices . . . 76
2.3 Observing and data reduction . . . 81
2.3.1 Influence of the D-shaped dark zone . . . 81
2.3.2 Influence of wavelength smearing . . . 84
2.3.3 Data reduction . . . 84
2.4 The world of vAPP . . . 86
2.4.1 The vAPP: a global view . . . 86
2.4.2 The vAPP: individual properties . . . 91
2.5 Current status and future developments . . . 101
2.6 Conclusion . . . 103

3 L-band spectroscopy of HR 8799 c,d,e with a vAPP-ALES combination **109**
3.1 Introduction . . . 110
3.2 Characterization of the dgvAPP360. . . . 113
3.3 Observations and data reduction . . . 116

3.3.1 Spectral cube extraction 117
3.3.2 Data reduction 119
3.3.3 Spectral extraction 122
3.4 Results 124
3.5 Discussion 125
3.5.1 Data reduction and the increased flux of HR 8799 d 125
3.5.2 Implications of the common flat and featureless spectra . . 128
3.6 Conclusions 129

4 Minimization of polarization leakage 133
4.1 Introduction 134
4.1.1 Diffractive phase plate coronagraphs 134
4.1.2 The Vector Vortex Coronagraph 135
4.1.3 The vector Apodizing Phase Plate coronagraph 136
4.1.4 Leakage mitigation strategies 136
4.2 Double-grating diffraction theory 138
4.2.1 The geometric phase hologram 138
4.2.2 The double-grating element 140
4.2.3 Simulated broadband performance of double-grating elements 141
4.2.4 The effect of non-zero grating separation 143
4.2.5 Multi-grating coronagraph architectures 144
4.3 Experimental results of double-grating implementations 146
4.3.1 Lab setup and double-grating manufacturing 146
4.3.2 Characterization of the double-grating polarization leakage 148
4.3.3 Separation of the double-grating PSF 149
4.4 Experimental results for a double-grating Vector Vortex Coronagraph 151
4.5 The double-grating vector-Apodizing Phase Plate for the Large Binocular Telescope. 154
4.6 Multi-grating coronagraphs and system-level perspectives 156
4.6.1 Multiple grating combinations 156
4.6.2 Applications of multiple-grating focal-plane coronagraphs . 158
4.7 Conclusions 159

5 Lab demonstration of Holographic Aperture Masking 163
5.1 Introduction 164
5.2 Theory 164
5.2.1 Multiplexed holographic blazed gratings 166
5.2.2 Liquid-crystal technology 167
5.2.3 Polarimetry with HAM (HAMPol) 169
5.3 Design 169
5.3.1 Bandwidth 169
5.3.2 One dimensional combinations 170
5.3.3 Multiplexed gratings 171
5.4 Lab results 174
5.4.1 Manufactured design 174

5.4.2 Manufacturing and lab testing 175
5.5 Conclusion 177
5.6 Acknowledgements 177

6 Holographic aperture masking with the Keck OSIRIS Imager 179
6.1 Introduction 180
6.2 HAM design 183
6.2.1 Geometric phase holograms 184
6.2.2 General considerations for a HAM design 186
6.3 The influence of polarization leakage 193
6.4 Design of a prototype mask for Keck OSIRIS 199
6.5 Data reduction pipeline 200
6.5.1 Pipeline overview 200
6.5.2 Spectroscopic closure phase extraction 202
6.6 Manufacturing of prototypes HAM v1 and HAM v1.5 204
6.7 Lab tests 206
6.7.1 The Point-spread function 208
6.7.2 Closure phases 209
6.8 On-Sky Verification 212
6.8.1 Binary HD 90823 212
6.8.2 Observations 213
6.8.3 Data Reduction 215
6.8.4 Spectroscopic parameter retrieval 218
6.9 Conclusions and Outlook 223
6.10 Appendix: Spectral resolution of the holographic interferograms . . 224

7 The vector-Zernike wavefront sensor 229
7.1 Introduction 230
7.2 Theory 232
7.3 Simulation 233
7.4 Lab results 236
7.5 Chromaticity of the vZWFS 237
7.6 Conclusion 239

8 Multi-color holography 241
8.1 Introduction 242
8.2 Concept of multi-color liquid-crystal holography 244
8.3 Lab experiment and result 246
8.3.1 Design of the multi-color geometric phase holograms 246
8.3.2 Manufacturing process 247
8.3.3 Lab setup 248
8.3.4 Experimental results 248
8.4 Wide-gamut and true-color holography 250
8.4.1 Color selection 250

8.4.2 True-color holography 251
8.4.3 True-color vAPP coronagraphs 253
8.5 Conclusion 256

9 Nederlandstalige samenvatting 259

List of Publications 267

Curriculum Vitae 271

Acknowledgements 272

1 Introduction

"For then why may not every one of these Stars or Suns have as great a Retinue as our Sun, of Planets, with their Moons, to wait upon them?" - Christiaan Huygens, Cosmotheoros

Christiaan Huygens (1629-1695) was a brilliant Dutch scientist and is well known for his work in the field of mathematics, physics, and astronomy. He also had an aptitude for instrumentation, building telescopes together with his brother, which he used to discover Saturn's moon Titan, and showing with observations that Saturn?s rings were in fact, rings. His final work *Cosmotheoros* had a more philosophical nature, where he, like many before and after him, wondered about the existence of planets around other stars. The idea of planets existing around other stars was not unheard of after the publications of Giordano Bruno (1584), however, what sets him apart is the idea to use telescopes to detect these planets. *"For let us fancy our selves placed at an equal distance from the Sun and fix'd Stars; we should then perceive no difference between them. For, as for all the Planets that we now see attend the Sun, we should not have the least glimpse of them, either that their Light would be too weak to affect us, or that all the Orbs in which they move would make up one lucid point with the Sun."* Huygens (1698), translation by Timothy Childe. It is likely that from his work on the distance between the Earth and the Sun and by comparing the brightness of Sirius with the Sun, he knew that these planets would be extremely close in angular separation and that detecting them would be difficult, if not impossible. Many after Christiaan Huygens have sought to prove the theory that planets exist around other star, yet it was not until 1992 that the first extra-solar planets were discovered by Wolszczan & Frail (1992). These planets were approximately four times the mass of the Earth, and orbited a pulsar which is the remnant of a supernova. Wolszczan and Frail used arrival time variations of radio waves emitted by the pulsar to infer the existence of these two planets. We note that with Christiaan Huygens' prediction, it is perhaps not surprising that these first exoplanets have been discovered through indirect detection.

These two planets provided a first look into the diversity in planets, as they orbited a stellar remnant and with their masses they had no counterpart in the solar system. A second clue that our solar system is not a standard system was given by the detection of Pegasi 51 b, a gas giant orbiting extremely close to the host star (Mayor & Queloz, 1995). Such planets are massive enough to significantly to make the star 'wobble' around their common center of mass. Using the Doppler effect, the radial velocity (RV) of these stellar movements can be measured and the existence of exoplanets inferred. This method was used by Mayor and Queloz to determine the mass of Pegasi 51 b (0.47 M_J) and the orbital period of 4.2 days. Jovian planets orbiting so close to their host star absorb a lot of stellar flux, which heats them to extreme temperatures (> 1000 K), resulting in their classification as hot Jupiters. It was realized that this type of planet would have a significant chance of passing in front of their star, partially obscuring them. Measuring the flux decrease of a host star when a planet passes in front is a third indirect method[1]

[1] If this method is purely indirect can be debated. The atmosphere of the planet leaves a

of exoplanet detection and is called the transit method. The extinction ratio of the stellar flux gives a measure of the planetary radius, and the first planet detected with this method is HD 209458 b (Charbonneau et al., 2000; Henry et al., 2000) with a radius of 1.27 R_J and an orbital period of 3.52 days, similar to Pegasi 51 b. HD 209458 b was also detected with the RV method, resulting in a mass estimate of 0.62 M_J (Henry et al., 2000). Interestingly, the density is much lower than that of Jupiter, showing that it is inflated due to the extreme temperature of the planet.

Probing the full exoplanet population is difficult due to intrinsic biases of the aforementioned detection methods. The RV and the transit method are most sensitive to massive planets orbiting close to the star and in the line of sight of the observer. A fourth indirect detection method uses precise astrometry to measure the wobble of a star and it alleviates the line of sight bias. It is particularly sensitive to planets orbiting in a face-on orbit, making it complementary to the RV and transit method. The *Gaia* satellite is capable of precise astrometry and current estimates indicate that it will find 21,000 exoplanets in a 5 year period (Perryman et al., 2014). The *Gaia* results are expected soon. However, indirect measurements are still biased towards close-in planets and provide little more information than the orbit, mass and radius. Measuring the bulk composition does not constrain anything about the atmosphere, which makes up only a small fraction of the mass and radius of a planet. Understanding the atmosphere is important for distinguishing between a Venus-like and Earth-like exoplanet, and in the end determining the habitability of an exoplanet. *Cosmotheoros* is a famous example on how misconceptions about planetary atmospheres can lead to wrong conclusions on habitability. Christiaan Huygens philosophized that other planets in the solar system were inhabited like Earth, because they too would have the most important ingredient for life: liquid water. He based this assumption on the observation that Jupiter has dark spots which he identified as clouds, and, that if Earth and Jupiter had water and clouds, so would the other planets. Interestingly, Huygens identified the problem that water on Earth would freeze on the outer planets and evaporate on the inner planets, but rather than defining a habitable zone, he proposed a change in water properties on the other planets[2]. Today we recognize that even finding an Earth sized planet in the habitable zone of a star does not mean that it can or does harbor life. In order to answer the fundamental question if we are alone in the universe, we first must understand the many processes involved in planet formation and evolution, as well as have a good understanding of atmospherical and geological processes (Snellen et al., 2019).

spectral imprint and extremely hot planets also can be detected in emission. See section 1.1.

[2] "Every Planet therefore must have its Waters of such a temper, as to be proportion'd to its heat: Jupiter's and Saturn's must be of such a nature as not to be liable to Frost; and Venus's and Mercury's of such, as not to be easily evaporated by the Sun. " - Huygens (1698)

1.1 The era of exoplanet characterization

Exoplanet research aims to understand the place of Earth and the Solar System in the universe. During the last three decades the number of detected exoplanets has rapidly grown to over 4300 exoplanets [3]. For a significant fraction of these planets we know the their bulk-composition through a comparison between the measured densities and models for rocky planets or gaseous planets (Howard et al., 2013). From the bulk composition we can infer that there are different types of planets that were not present in our solar-system, e.g. mini-Neptunes or super-Earths (Petigura et al., 2013) This gave us already a more complete picture of the planet population, and statistical properties of the population as a whole become unique constraints on planet formation models (Benz et al., 2014). Moreover, we learned that rocky planets are very common, and the chances of an Earth analogue orbiting a nearby star are high. However, if we are to understand how unique Earth, and the life on Earth truly is, it is critical that we gain a more detailed understanding of the exoplanet population and especially their atmospheres.

1.1.1 Methods of exoplanet characterization

Two techniques have successfully demonstrated being capable of characterizing exoplanet atmospheres, transit spectroscopy and direct imaging. Low-resolution transit spectroscopy measures the transit depth variations as function of wavelength caused by the atmospheric constituents changing the opacity of the atmosphere. Hot Jupiters are prime targets for transit spectroscopy due to their size and distance to the star. Accurately measuring the transit depth as function of wavelength has revealed the existence of clouds and hazes in hot Jupiters (Sing et al., 2016). Extending the observations to the full phase curve with emission spectroscopy, the vertical and longitudinal, thermal structure of the atmosphere of WASP-43 was resolved (Stevenson et al., 2014). A similar effort for lower-mass planets revealed that finding molecular features using transit spectroscopy is hindered by the presence of clouds (Kreidberg et al., 2014). With high-resolution spectroscopy (HRS) it is possible to measure the Doppler shift of planetary atomic lines during the transit. Exploiting this shift revealed neutral iron and ionized atomic iron and titanium (Fe^+ and Ti^+) in the hot Jupiter Kelt-9b (Hoeijmakers et al., 2018b). High-resolution spectroscopy can also be used to detect Doppler-shifted emission and absorption features from hot Jupiters during the full phase curve (Snellen et al., 2010).
With direct imaging the planet is spatially resolved from the star, enabling the direct measurement of planet light that is reflected and/or thermal. HRS combined with direct imaging and optical stellar suppression techniques (coronagraphs) provides additional suppression of stellar light and can be used to characterize planetary atmospheres (Wang et al., 2017). Moreover, HRS can be used to measure the Doppler shift of the planet light, which can be used to constrain the orbit

[3] `https://exoplanets.nasa.gov/` - Feb 16, 2021

and the linewidth is a measure for rotational broadening (Snellen et al., 2014). Direct imaging combined with medium-resolution spectroscopy cannot resolve the Doppler shift, however it still resolves spectral lines of molecular species not present in the stellar spectrum. This has been used to detect molecular species like water and carbon monoxide and the accretion tracer Hα in young giant exoplanets at large separations (dit de la Roche et al., 2018; Hoeijmakers et al., 2018a; Haffert et al., 2019). Low-resolution spectroscopy is only sensitive to molecular bands, but is the most widely used form of spectroscopy for exoplanet characterization. So far this has lead to the detection of water, carbon monoxide and methane in young giant exoplanets (Konopacky et al., 2013a; Macintosh et al., 2015), and the combination is used to constrain the C/O ratio compared to their host star (Wang et al., 2020; Mollière et al., 2020). Moreover, polarimetric measurements of exoplanets could be used to determine oblateness or cloud coverage (Stolker et al., 2017), or even the formation history through the orientation of circumplanetary disks (van Holstein et al., 2021). Lastly, photometric monitoring of directly imaged exoplanets has been successful (Zhou et al., 2020), and could also provide more insight in the cloud coverage.

1.1.2 Direct imaging of HR 8799

To demonstrate what information can be extracted from directly imaged exoplanets, we summarize the results of the most studied system with direct imaging: HR 8799. HR 8799 A is a young (<100 Myr) A star at 39 parsec with a large debris disk (Su et al., 2009). The HR 8799 system hosts at least four giant planets which have been detected by Marois et al. (2008, 2010). The planets orbit at large separations, between 15 and 70 AU, and have masses around 7 $M_{\rm Jup}$, although this is age dependent. A selection of some of the results is shown in Fig. 9.2. Gas giants, like the HR 8799 planets, are thought to be similar to field brown dwarfs, as they have a similar radius, composition and slow cooling rate (Skemer et al., 2014). Field brown dwarfs are substellar objects with masses between 13 and 80 $M_{\rm Jup}$, which is not enough mass to trigger sustained nuclear fusion of hydrogen in their cores. Therefore, they slowly cool down after formation through thermally radiating the gravitational energy released during formation. The atmospheres of brown dwarfs (and gas giants) regulate this cooling and thus their evolution, and observing these atmospheres is essential for their characterization. Field brown dwarfs are more easily studied as they are not obscured by a much brighter host star, and many already have been found (Kirkpatrick et al., 2011) and characterized (Burgasser, 2014). As brown dwarfs are cooling they change spectral type due to changes in their atmospheres. One of these transitions is the L-T transition, where their atmospheres transition from CO-rich and cloudy (L-type), to CH_4-rich and relatively cloud free (T-type), e.g. Vos et al. (2019). For field brown dwarfs this happens for temperatures cooler than 1400K, over a narrow temperature range of $\sim$ 200K. During this period, the clouds are thought to sink below the photosphere and the near-IR color shifts toward the blue. Due to the similarities between gas giants and brown dwarfs, it is thought that gas giants also show this L-T transition. We

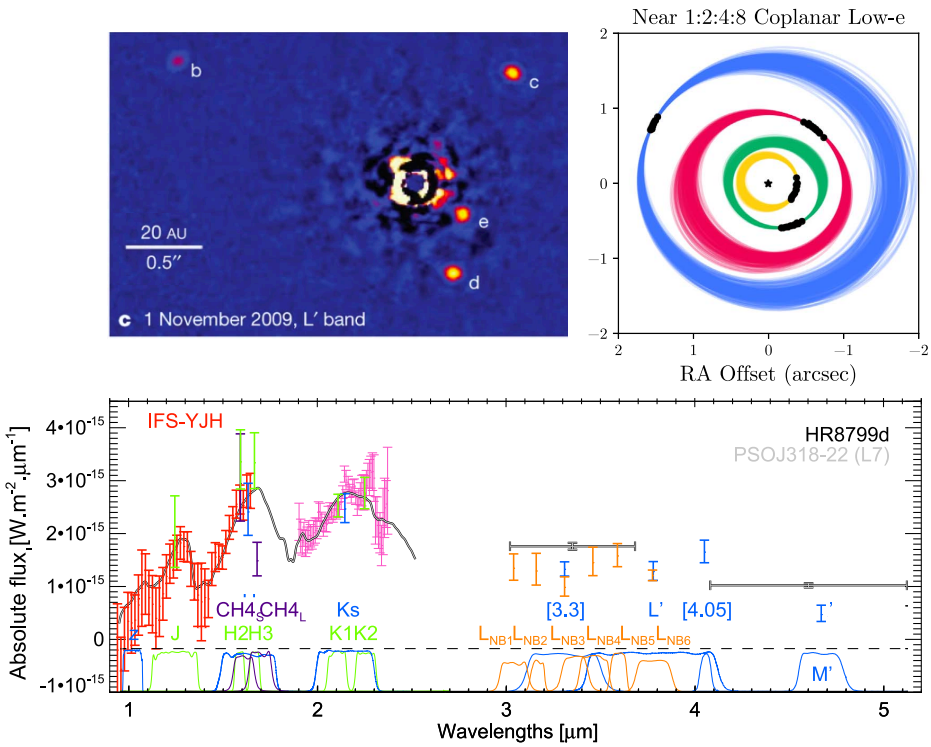


Figure 1.1: Selected results of studies on the HR 8799 system. *Top left:* Image of the HR 8799 planetary system, showing all four planetary companions. Adapted from Marois et al. (2010). *Top right:* Allowed orbits of the four planets, taking into account dynamical constraints. Adapted from Wang et al. (2018). *Bottom:* Comparison between the combined spectrum of planet d taken with multiple instruments and a field brown dwarf of type L7. Adapted from Bonnefoy et al. (2016).

note that the physical mechanism of the L-T transition in brown dwarfs is not yet fully understood (Vos et al., 2019). With the detection of the HR 8799 planets, it was shown that their location of the in a near-IR color-magnitude diagram was close to the L-T transition (Marois et al., 2008). Their temperatures (1100K) were closer to a T-type brown dwarf, while clouds were required to properly fit the photometric points, directly challenging the brown dwarf models. Medium resolution spectroscopy of HR 8799 b also showed no significant methane absorption, which would be expected for a T-type brown dwarf of similar absolute magnitude (Bowler et al., 2010). Using high-resolution spectra it was confirmed that HR 8799 b does contain water, methane and carbon monoxide (Konopacky et al., 2013b; Barman et al., 2015). New M band and L band photometry also suggested the presence of thick clouds and absence of CH_4 (Skemer et al., 2014; Currie et al.,

2014). Overall, the early results showed significant differences between field brown dwarfs and their models, possibly as a consequence of the lower surface gravity (Marley et al., 2012). In addition, adaptations to the models resulted in exploration of carbon disequilibrium chemistry and uniform or patchy clouds (Bowler et al., 2010; Barman et al., 2011; Skemer et al., 2012).
New direct imaging instruments with low-resolution spectroscopic capabilities were used to get the first spectra of the HR 8799 planets (Oppenheimer et al., 2013; Zurlo et al., 2016; Greenbaum et al., 2018). These spectra were not fitted well with spectra of field brown dwarfs, although a better fit is realized by comparing with low-gravity brown dwarfs (Bonnefoy et al., 2016). Modelling the spectra to retrieve atmospheric parameters was partially successful by including thicker clouds compared to brown dwarfs. However, many different models only properly fit 1.1-2.5 μm or the thermal infrared (3-5 μm), not both simultaneously, e.g. (Barman et al., 2015; Bonnefoy et al., 2016). Attempts at characterizing the atmospheres therefore lead to different conclusions on cloud coverage (uniform vs patchy) (Mollière et al., 2020; Currie et al., 2014) and equilibrium vs. disequilibrium chemistry (Lavie et al., 2017; Mollière et al., 2020), depending on which part of the spectrum was prioritized. Recently, updated self-consistent models started to produce better results by taking into account the lower gravity, producing an increase of cloud optical depth (Charnay et al., 2018). Moreover, these models predict a greenhouse effect that shifts the carbon chemical equilibrium, changing the abundances of species compared to similar temperature brown dwarfs with higher surface gravity. These models do not (yet) include non-uniformity in cloud layers. However, is to be expected that in the transition from uniform clouds (L-type) to no clouds (T-type), patchy clouds can be an intermediate state. Another way of probing cloud patchiness is through photometric monitoring of planetary companions, which is already observed for brown dwarfs (Apai et al., 2017). It is found that many brown dwarfs (L- and T-type) are variable, but they vary most strongly during the L-T transition (Metchev et al., 2015). Measuring this variability for HR 8799 planets was attempted but sensitivity not good enough (Biller et al., 2021). Yet another method to probe cloud patchiness is through polarimetric observations. For planets with uniform clouds the linear polarization signal is expected to be zero, while patchy clouds could generate a linear polarization signal of up to 0.03% and 1.33%(Stolker et al., 2017). Measurements of linear polarization of the HR 8799 planets resulted in upper limits of $\sim 1\%$ (van Holstein et al., 2017, 2021). While this is a stringent upper limit, it does not constrain the cloud patchiness.
Other information from direct imaging comes from continuous monitoring of the HR 8799 system to track the orbits of the planets. These studies show that the planets are in near 1:2:4:8 resonance and from dynamical constraints it can be shown that the orbits are almost co-planar (Konopacky et al., 2016; Lacour et al., 2019). The resonances have implications on the planet formation, suggesting that they were locked in resonance quickly after formation, before the gaseous protoplanetary disk disappeared (Wang et al., 2018). However, it can not be used to probe the formation mechanism of the planets, and the uncertainty about the for-

mation mechanism also results in different upper limits on the derived dynamical mass of the planets. A method that can be used to study the formation history of the HR 8799 planets requires a measurement of the planetary C/O ratio. Depending on the formation mechanism and formation location in the protoplanetary disk, the C/O ratio of the planet can be different to the stellar C/O ratio (**?**). The C/O ratios in the atmospheres of the HR 8799 planets are consistent with superstellar (Konopacky et al., 2013b; Barman et al., 2015; Mollière et al., 2020), although stellar C/O ratios are also not excluded due to degeneracies (Wang et al., 2020; Mollière et al., 2020). A superstellar C/O ratio could be indicative of planet formation beyond the CO snowline, the region in the disk with temperatures that make CO freeze onto grains, and planet migration (Mollière et al., 2020). Lastly, the interaction with the debris disk could also be used to find evidence of planet migration. The orbital fits of planet b at show that its location is consistent with sculpting the inner edge of the debris disk, favouring a more stable system (Wang et al., 2018). However, the debris disk has a large halo of small grains extending out to 1000 AU, possibly originating from unstable orbits or planet migration (Su et al., 2009).
To summarize, direct imaging has given a wealth of information about the HR 8799 system, which already provided tests for planet formation and evolution theories, as well as drive the exoplanet atmosphere model development. Up to now direct imaging has been limited to young, massive planets on wide orbits, yet it has the potential to provide the same revolution for a wider range of planets. In this thesis we will show the challenges of direct imaging and explore technological advancements that will help find lower mass exoplanets closer to their star.

1.2 Direct imaging systems

1.2.1 The challenge of direct imaging

The goal of direct imaging is to resolve the light of the companion from that of the star. This is challenging due to the extreme difference in brightness between the star and the planet and the small angular separation. Young massive self-luminous exoplanets and brown dwarfs have flux ratios in the near-infrared of $\sim 10^{-5}$. e.g. HR 8799 b,c,d,e (Marois et al., 2008, 2010). For a solar-system analogue at 10 parsec observed at visible wavelengths, looking at the stellar light reflected of the planet, an Earth-twin has a separation of 100 mas and a flux ratio of $\sim 10^{-10}$ (Traub & Oppenheimer, 2010). The flux ratio of the planet flux and the stellar flux is also called the *contrast* ratio, and direct imaging is therefore also called "high-contrast imaging". Moreover, while the planet and the star are physically separated, the properties of the imaging system (telescope) determine if the objects can be separated in the image. Imaging a point sources results in a finite size point-spread function (PSF) due to the diffraction of light. The resolving power of a telescope, the smallest angular separation it can resolve, is given by the Rayleigh criterion which equals $1.22\lambda/D$ for a circular aperture, with λ is the wavelength and

D is the diameter of the telescope. The Rayleigh criterion drives the need for ever larger telescopes or going to shorter wavelengths, resolving objects closer to the star or for looking at the same angular separation at larger distances. Being able to observing targets further from Earth increases the sample size and the additional area of the telescope helps dealing with the decreased flux. However, increasing the telescope diameter is only a partial solution for the extreme brightness ratio between the planet and star as the diffraction structure of the PSF is bright and extensive. For a telescope with a clear circular aperture, the diffraction structure has a flux $> 10^{-2}$ compared to the PSF core at the first Airy ring ($1.64\lambda/D$), and is $< 10^{-4}$ beyond $\sim 7\lambda/D$. A complicating factor is that aberrated optical systems have even more stellar flux present in these structures outside the PSF core. Ground-based telescopes look through Earth's atmosphere which creates large rapidly-changing aberrations, and therefore subtracting a reference PSFs does not work. Also, photon noise of the stellar PSF is much larger than the planet signal even if the system is diffraction limited. So, to directly image exoplanets at small separations the stellar flux needs to be suppressed by many orders of magnitude.

1.2.2 A modular view of high-contrast imaging

Directly imaging exoplanets requires instruments that have the capability to 1) correct for aberrations, 2) remove the diffraction structure at the location of the planet, 3) image the light of the planet in a way that can be used to enhance the contrast as well as enable characterization, e.g. imaging, spectroscopy, or polarimetry, 4) have a data processing infrastructure that allows for additional contrast enhancement and characterization. Such an instrument is called a high-contrast imaging system, and a general schematic of such a system is presented in Fig. 1.2. We note that contrast-enhancing techniques and characterization are closely related, but they are separate things. For example, techniques that use the spectral difference between a star and a planet enhance the contrast for a specific spectral feature, and the presence or absence of this feature within the enhanced detection limit characterizes the atmosphere of the planet. In this section we will focus on the contrast-enhancing techniques and we will shortly go over the challenges and current status of each component, although ultimately the components should not be viewed as separate.

Adaptive optics

The main problem for ground-based telescopes is Earth's atmosphere distorting the light from stars and planets. Atmospheric layers have temperature differences and potentially pressure differences that change the local refractive index of the air, and the difference in wind speed and direction generates turbulence between these layers. Turbulence in the atmosphere changes the optical path and thus arrival times of light differently for every location in the telescope pupil, distorting the wavefront temporally, and spatially and creating a PSF that is smeared-out

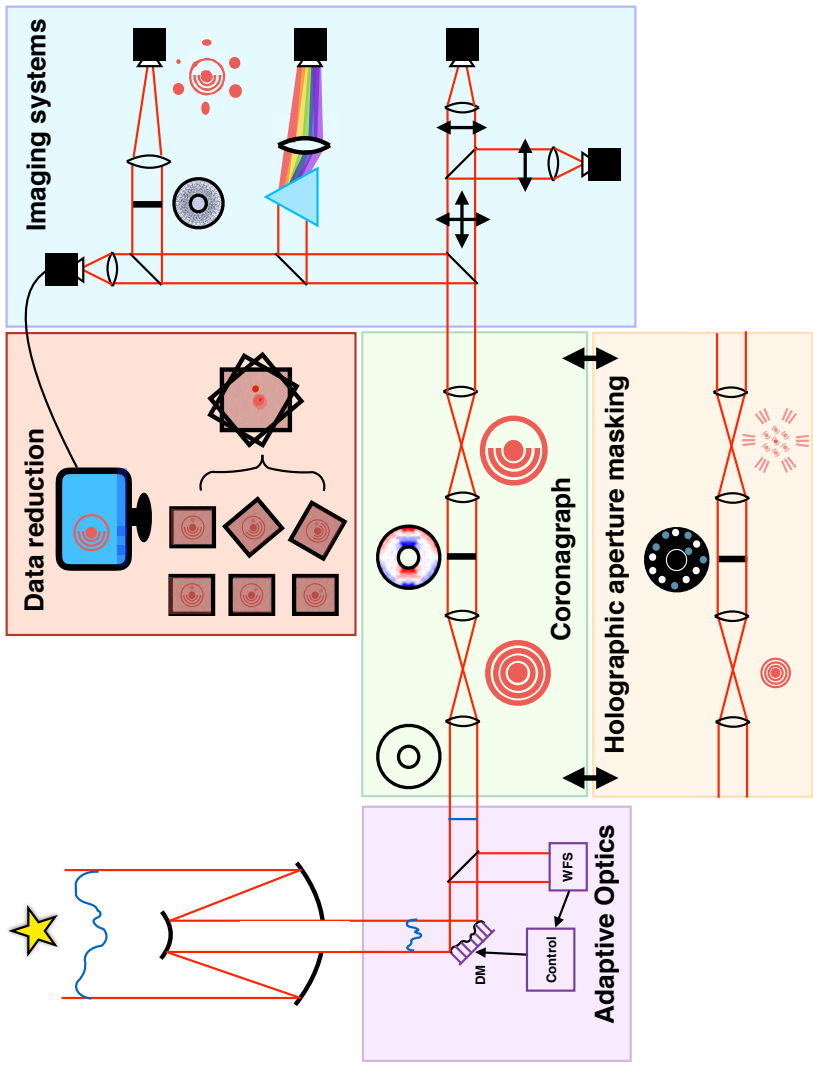


Figure 1.2: Schematic of an HCI system with five modules that combined enable the direct imaging of exoplanets. The modules are adaptive optics, coronagraph, (holographic) aperture mask, imaging system and data reduction.

compared to the diffraction limit. The turbulence is characterized by the Fried parameter r_0 (Fried, 1966), which is the coherence length of the turbulence and is $\sim$ 10 cm for a wavelength of 500 nm and average seeing. We note that r_0 scales with wavelength as $r_0 \propto \lambda^{\frac{6}{5}}$ such that observations at longer wavelengths are less affected by turbulence. For uncorrected turbulence, the average PSF size is directly related to r_0 as the averaged angular extent is given by λ/r_0. Therefore, any increase in the telescope size beyond the r_0 is only effective in terms of spatial resolution if the atmospheric turbulence is corrected. With adaptive optics (AO) it is possible to partially correct for the turbulence and return to the diffraction limit with the current 8-m class telescopes (Fusco et al., 2016). The principle of AO is to actively measure the shape of the wavefront and correct it. A wavefront sensor (WFS) measures the wavefront by modifying the light in a way that the wavefront shape becomes an intensity modulation that can be measured with a camera. measurable quantity on a camera, like spot shifts (Shack-Hartmann WFS) or intensity variations (pyramid WFS). The Shack-Hartmann WFS (Hartmann, 1900; Shack, 1971) uses a microlens array in the pupil to create spots that shift with local wavefront tilts, directly measuring the wavefront slope. A pyramid wavefront sensor (Ragazzoni, 1996) has a pyramidal-shaped prism in the image plane to create four pupil images with intensity variations that correlate with the wavefront slope. One important property of a WFS is its efficiency, i.e. how many photons need to be captured to provide an accurate measurement of the wavefront. An efficient wavefront sensor can be operated faster or have the same performance on fainter targets. The signal of a WFS is then converted by control algorithm to shape a deformable mirror (DM), which takes the inverted shape of the wavefront with half the amplitude and thereby flattening it. A flattened wavefront restores the PSF, making it resemble the theoretical unaberrated PSF for the optical system. We note that if the DM correction is never perfect and a flattened wavefront does not necessarily result in the best contrast (Radhakrishnan et al., 2018).
When the DM is placed in the optical path before the wavefront sensor, the system forms a closed loop and the wavefront can be kept stable. A measure for the correction quality of the AO system is the Strehl ratio, which is defined as peak intensity of the corrected PSF and the perfect theoretical PSF. Even for extreme AO systems, with DMs that have many actuators controlling the mirror surface, efficient wavefront sensors and fast loop speeds (>1 kHz), the Strehl ratio is not 1. In this case, the PSF will exhibit modulations of the intensity with a size of λ/D that mimic planet signals called speckles. Speckles are coherent with the diffraction structure and are generated by many effects, but currently there are two main causes that prevent a perfect correction for a PSF at the science focal plane. First, the wavefront sensor does not measure the aberrations from optical elements between the WFS and the science focal plane. These aberrations, called non-common path aberrations (NCPA) arise from chromaticity of optics, alignment errors or surface errors, and are slowly changing due to changing gravity vectors, humidity, and temperature. Speckles generated from these aberrations are changing on timescales from minutes to hours and therefore are difficult to distinguish from planet signals (Martinez et al., 2013). Measuring NCPA is difficult as the location

where you would ideally measure them is occupied by the coronagraph. Many solutions exist and most of them are summarized in Jovanovic et al. (2018). They can be separated in three categories, having a separate wavefront sensor temporarily replacing the coronagraph, using light reflected from a coronagraphic mask, or applying some known diversity to obtain the wavefront information directly from the science focal plane. In practice, all NCPA wavefront sensing techniques only gain a factor of 2-5 when tested on-sky for ground-based telescopes (Vigan et al., 2018a; Bos et al., 2019; Galicher et al., 2019). The second limiting effect for optimal AO correction is the servo-lag error, created by the finite time it takes to update the shape of the DM. Turbulence is not static as the atmosphere moves with a wind speed V_{wind} and the typical time scale for the evolving atmosphere is therefore given by $\tau_0 \propto \frac{r_0}{V_{wind}}$ (Greenwood, 1977). With wind speeds up to tens of meters per second with multiple directions in the atmospheric column a telescope looks through, the loop speeds need to be significantly higher than 1 kHz. When the loop speed is lower, the residual errors at high temporal frequency create a halo of light in the wind speed direction, the wind-driven halo (Cantalloube et al., 2018, 2020). Upgrading the loop speed is an extreme technological challenge. Another solution is to change the control loop to predict the wavefront by extrapolating the wind speed and the previous evolution (Hardy, 1998; van Kooten et al., 2019). Without predictive control the AO is fundamentally limited by photon noise at the 10^{-4}-10^{-6} level (Guyon, 2005; Males & Guyon, 2018).
For space-based high-contrast imaging systems the performance is also limited by wavefront control, albeit at a very different contrast level. The thermal load variation caused by solar radiation results in slow drifts in the wavefront and vibration from the reaction wheels result in high-frequency wavefront aberrations (Shi et al., 2017). Also, it is possible that the deformable mirror itself can drift (Seo et al., 2019). Low-order coronagraphic wavefront sensors are used in combination with focal-plane wavefront-sensing techniques to achieve the 10 picometer RMS wavefront that is required for 10^{-10} contrast (Steeves et al., 2020). An overview of current work on wavefront sensing for space-based missions was given by Pueyo et al. (2019). Adaptive optics, including both hardware and software, is the most critical system that largely sets the contrast limit for a high-contrast imaging instrument.

The coronagraph

After correction of the wavefront, the diffraction structure close to the PSF core is still much brighter than the signal of planets. A coronagraph is an optical device that suppresses this diffraction structure. The advantage of a coronagraph is two-fold. By suppressing the diffraction structure, the photon noise of the stellar light is reduced, and second, the influence of speckles is suppressed. Therefore, a coronagraph significantly enhances the signal-to-noise ratio of the planet signal. The first coronagraph was developed by Bernard Lyot to image the solar corona by masking the disk of the sun and blocking the diffracted light in a consecutive pupil plane (Lyot, 1939). Since then, many different coronagraphs have

been designed and tested as is summarized in Mawet et al. (2012) and Ruane et al. (2018). These coronagraphs can be compared in performance by looking at five properties: inner working angle, contrast, sensitivity to low-order aberrations, throughput, and chromaticity. The *inner working angle* (IWA) is defined as the smallest angular separation at which the total energy throughput of an off-axis source reaches 50% of the maximum throughput, or where the contrast reaches a threshold value (Ruane et al., 2018). The IWA is not necessarily the same as the diffraction limit, but can always be expressed as a factor of the diffraction limit set by the telescope diameter. We note that coronagraphs also have an outer working angle, which has the same definition as the IWA but for the outer edge of the dark zone. The *contrast* is defined as the flux ratio with respect to the star where a point source can significantly (5σ) be detected (Jensen-Clem et al., 2017). For a coronagraph design, the contrast is then the (spatially variant) suppression of the starlight divided by the coronagraphic throughput for a clear telescope aperture, measured in a circle with a diameter of approximately the diffraction limit. In telescopes the optical aberrations reduce the contrast but this is not the same for any coronagraph. Some coronagraphs are more resistant to low-order modes at the cost of a larger (theoretical) inner working angle, yet the measured contrast with these aberrations might be better at the smallest separations (Mawet et al., 2010). Lastly, exoplanet characterization requires multi-wavelength observations or spectroscopy. Coronagraphs that deliver good contrast over large bandwidths have an improved the observation efficiency and higher signal-to-noise per spectral bin.

These coronagraphic properties are easily visualized for a Lyot coronagraph, which has an opaque mask in the focal plane that blocks the starlight, see Fig. 1.3. Another amplitude mask (or Lyot stop) in the following pupil plane blocks the stellar light that is diffracted around the opaque focal-plane mask. The inner working angle is then determined by the focal-plane mask radius, the throughput of planets far beyond the IWA is given by the Lyot stop area compared to the actual pupil size, and the chromaticity is largely determined by the wavelength scaling of the PSF. The Lyot coronagraph is a focal-plane coronagraph, as the opaque mask that blocks the star is in the focal plane. Other focal-plane coronagraphs have different focal-plane masks, but operate on the same principle with a Lyot stop, e.g. the vector vortex coronagraph (VVC) (Mawet et al., 2005a), phase-induced amplitude apodization complex mask coronagraph (PIAACMC) (Guyon et al., 2005). If a phase mask is used in the focal plane (VVC, PIAACMC) the phase pattern results in destructive interference in the pupil. Focal-plane coronagraphs have a theoretical performance in both inner working angle and throughput is better than the other type of coronagraphs that operate in the pupil plane. Moreover, it has been shown that the use of focal-plane phase masks result in an improved performance over amplitude masks (Mawet et al., 2009a; Por, 2020) . Yet, because focal-plane coronagraphs are sensitive and have small inner working angles, they are highly sensitive to tip-tilt aberrations that can be caused by vibrations.

Pupil-plane coronagraphs have a different philosophy compared to focal-plane coronagraphs, as they do not block the starlight but rather change the shape

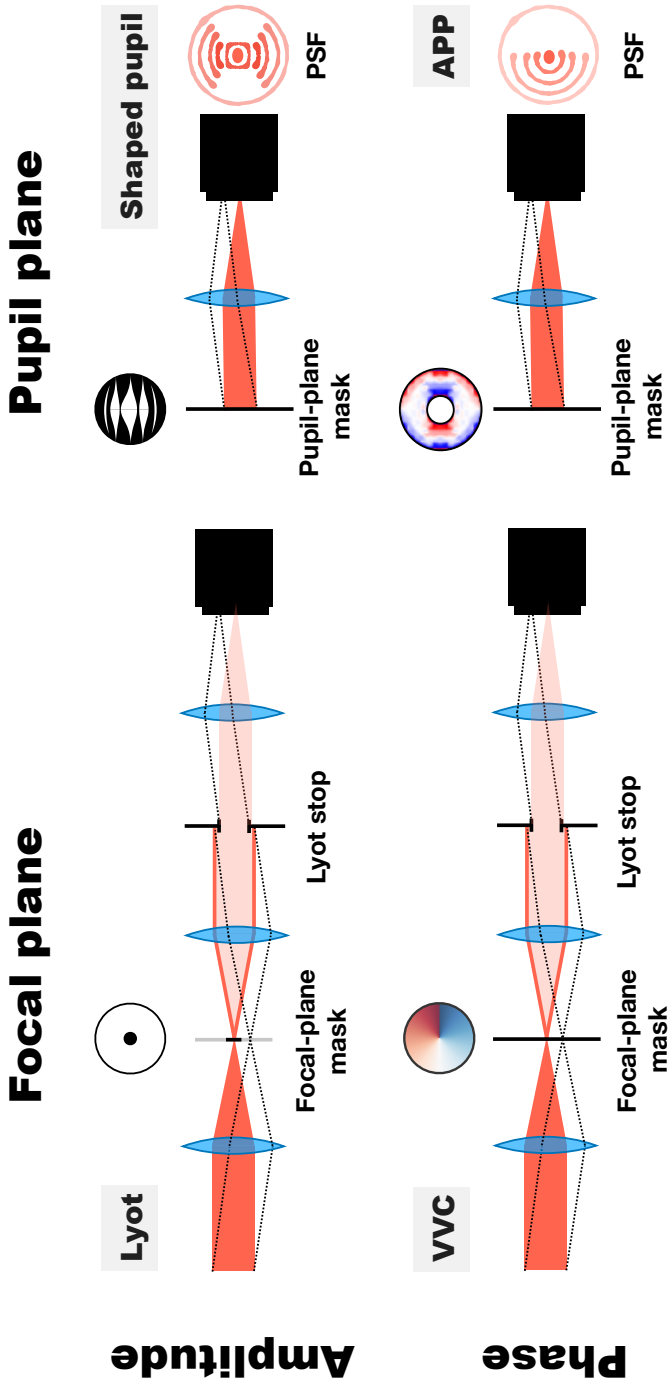


Figure 1.3: Schematic of four types of coronagraphs, having phase or amplitude masks in the focal plane or pupil plane.

of the PSF with a region where the flux is reduced. This so called dark zone is created by nulling the starlight with either a phase or amplitude mask, e.g. the apodizing phase plate (APP) (Codona et al., 2006; Kenworthy et al., 2007; Snik et al., 2012) or the shaped pupil (Kasdin et al., 2003) respectively. Both are shown in Fig. 1.3. Pupil-plane coronagraphs change the shape of all objects in the field-of-view in the same way, making the coronagraphs tip-tilt insensitive. In addition, they occupy a single pupil plane, making it easy to install them in existing high-contrast imaging systems. The advantage of phase patterns over amplitude masks is that it enables the creation of one-sided dark zones with smaller inner working angles. Moreover, it can be shown that for annular dark zones the phase-only solution always provides a better performance compared to amplitude masks (Por, 2017). A detailed description of the APP coronagraphs can be found in chapter 2.
A third type of coronagraph are the hybrid coronagraphs, having a phase or amplitude mask in both the pupil plane and focal plane. With two masks the design freedom is larger, making it easier to adapt to any pupil shape (Soummer, 2004), or increase the bandwidth (Soummer et al., 2011b). For the Nancy Grace Roman Space Telescope (previously named WFIRST), two hybrid coronagraphs have been selected, the (hybrid) shaped pupil coronagraph and the hybrid Lyot coronagraph (HLC). The shaped pupil coronagraph has an amplitude pupil-plane and focal-plane mask in addition to a Lyot stop (Balasubramanian et al., 2015). The HLC uses a static offset on two deformable mirrors in combination with a standard Lyot setup (Trauger et al., 2016). A different version of the Lyot coronagraph, the apodized pupil Lyot coronagraph (APLC), is commonly used for ground-based observations (Macintosh et al., 2006; Beuzit et al., 2019). The APLC has an amplitude mask in the pupil plane with a standard Lyot setup (Soummer, 2004). Replacing the amplitude mask in the pupil plane with a phase mask, the phase apodized pupil Lyot coronagraph (PAPLC), results in an improved performance (Por, 2020), and enables one-sided dark zones.
Over the last decade, coronagraph designs have improved for arbitrary apertures both terms of contrast and inner working angle (Ruane et al., 2016; Por, 2017; N'Diaye et al., 2018; Por, 2020). Due to rapid technological advancements, the coronagraphic performance has also been demonstrated in the lab. For example, the Lyot coronagraph has delivered extreme stellar suppression ratios down to 4×10^{-10} between 3-8 λ/D for 10% bandwidth (Seo et al., 2019). These coronagraphs are soon ready to use for the next generation of space-based missions, where the aberrations are small. On the other hand, ground-based instruments are limited by the AO system to a contrast of $\sim 10^{-5}$. It can be shown that switching between coronagraphs in AO-limited systems does not result in a significantly improved performance (Carlomagno et al., 2016). The parameter space for coronagraph design has therefore shifted to chromaticity (ground- and space-based) and robustness against low-order aberrations (space-based). Comparing coronagraphs with the aforementioned metrics is relatively simple but does not consider all the differences between coronagraphs. For example, some coronagraphs are better for focal-plane wavefront sensing or less sensitive to the shape of a telescope pupil or have different dark-zone shapes. These properties also impact the operation and

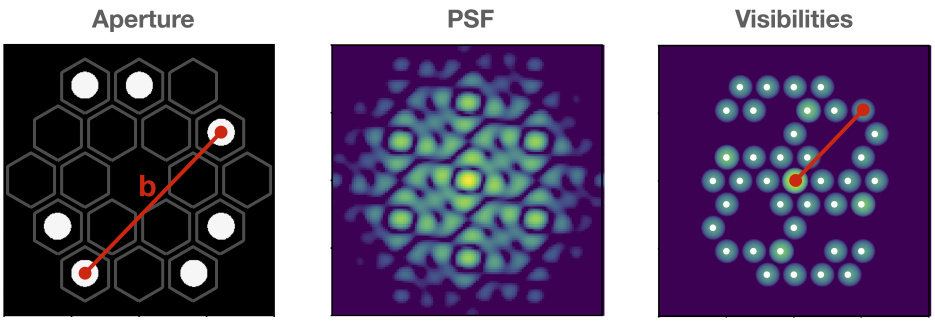


Figure 1.4: *Left:* Layout of a 7-hole aperture mask where every hole pair has a unique vector called a baseline (b). *Middle:* Simulated PSF of the aperture mask, which is the combination of all fringes generated by the 21 hole pairs. *Right:* The Fourier transform of the PSF (visibilities), decomposing the PSF in the individual frequencies corresponding to the baselines.

performance on-sky. Only in recent years, a true comparison was attempted by considering a full observing sequence that would be optimal for each coronagraph and looking at the exoplanet yield assuming a planet population (Stark et al., 2019).

Aperture masking

A technique complementary to coronagraphy in terms of detecting objects at small separations is sparse aperture masking (SAM). SAM is an interferometric technique and works by turning a telescope aperture into an interferometric array using an opaque mask with small holes in the pupil plane (Haniff et al., 1987; Tuthill et al., 2000). A commonly used version of SAM uses a mask with holes that are placed in a non-redundant fashion, which means that each baseline (the vector that connects two apertures) appears only once. The PSF of such a mask is an interferogram which contains many fringes with unique spatial frequencies corresponding to these baselines. By recovering the phase and amplitude of all fringes in the interferogram, it is possible to reconstruct an image of the observed object. The fringe phase and amplitude recovery is usually done by directly fitting all fringes in the interferogram at the same time (Lacour et al., 2011a; Greenbaum et al., 2014) or by Fourier transforming the image to create the visibility plane (Tuthill et al., 1999; Sallum & Eisner, 2017). The conversion to an image is done using the CLEAN algorithm (Högbom, 1974) or maximum entropy method (Gull & Skilling, 1984; Sivia, 1987). SAM has two distinct advantages over direct imaging. First, the fringes of the PSF have a first null at $0.5\lambda/B$ instead of $1.22\lambda/D$, where B is the longest baseline, resulting in a reduction of the full width at half maximum of the PSF by a factor ≈ 2. The diffraction limit of SAM is therefore

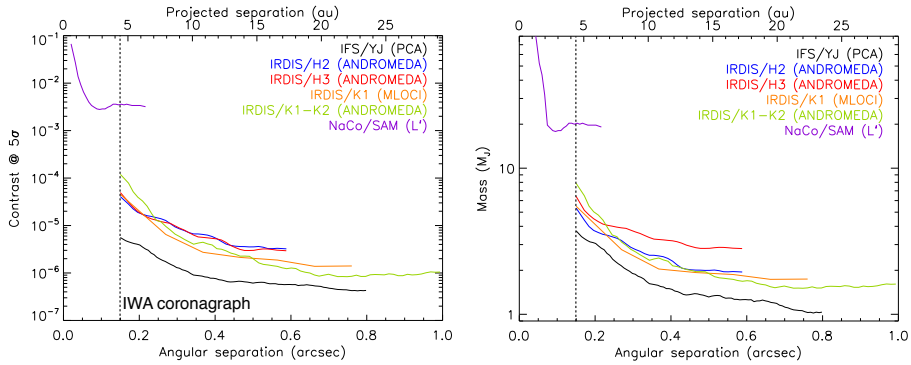


Figure 1.5: *Left:* 5σ-contrast of observations of 51 Eridani using NACO/SAM (L'), SPHERE/IRDIS, and SPHERE/IFS. *Right:* Same as left, but converted to mass using models from (Baraffe et al., 2003, 2015) and BT-Settl models (Allard, 2013). Adapted from (Samland, M. et al., 2017).

twice as small for the same telescope for a baseline that spans across the pupil. Second, the non-redundancy can be exploited to reject the effect of atmospheric or instrumental aberrations on the fringe phases and amplitudes. In particular, there are robust observables called closure phases that are independent of these aberrations but depend on the source structure. Closure phases are formed by taking the sum of phases around baselines forming a closed triangle of subapertures in the pupil. It was this combination of properties of SAM that allowed for imaging the regions closest to stars before adaptive optics became a standard practice in astronomy (Tuthill et al., 1999).

The downside of sparse aperture masks, such as those in VLT/SPHERE (Lacour et al., 2011b) and Keck/NIRC2 (Tuthill et al., 2006), is that they block $80 - 90\%$ of the incident light. Moreover, the closure phases are only stable against static phase aberrations and fast imaging is required to combat the influence of turbulence. Thus, sparse aperture masking was limited to bright targets, however with the invention of AO the requirement on the imaging speed could significantly be relaxed. When SAM is used in concert with AO, SAM provides extremely stable closure phases that resulted in a leap in sensitivity and contrast (Tuthill et al., 2006). The superior calibration particularly of closure phases makes SAM more sensitive than coronagraphs, probing separations that are otherwise dominated by aberrations ($< 1 - 2\lambda/D$), e.g. Gauchet et al. (2016); Cheetham et al. (2016); Samland, M. et al. (2017). The parameter space for exoplanet detection is also shown in Fig. 1.5. Most recent results show that SAM at VLT/SPHERE can reach a 5σ contrast of $\sim 10^{-3}$ down to 30 mas (0.6 λ/D in K-band) (Claudi et al., 2019; Pérez et al., 2020).

Imaging systems and data reduction

The imaging system is closely connected to the data reduction process and together they also represent additional methods of enhancing the contrast. Moreover, what information can be recovered from the planet signal is directly determined by the imaging system. We reiterate that contrast-enhancing techniques and characterization, e.g. through spectroscopy and polarimetry, are separate things. The physical differences between the star and planet generate differences in the stellar flux or planetary flux that can be exploited to enhance the contrast. There are four methods that we will briefly discuss: reference star differential imaging (RDI), angular differential imaging (ADI), spectral differential imaging (SDI), and polarimatric differential imaging (PDI). The first two methods use the fact that the planet is spatially separated from the star.

ADI: Angular differential imaging is currently the most widely used differential technique. The technique makes use of the sky rotation with respect to the telescope pupil for an alt-azimuth telescope (Marois et al., 2006a). If the pupil is stabilized the sky rotates with respect to the PSF structures, e.g. from spiders or speckles, and the PSF structures remain in a mean image of all the data. However, off-axis objects rotate with the parallactic angle and the mean will average this to a smeared and reduced signal. Subtracting the mean image from every frame therefore removes the unwanted PSF structures and derotating and stacking the residuals leads to a great enhancement of the detection limit. More advanced methods of combining the images to create a better reference PSF produce better results (Lafreniere et al., 2007). Another method creates subtracts the principal components from the data before derotation, i.e. principal component analysis (PCA) (Soummer et al., 2012). ADI works best for larger separations, where the size of the diffraction structures is much smaller than the rotation path of the object. It takes much longer for an object at small separation to have rotated by a diffraction width than at larger separations. If the rotation is smaller than the object size then the object is partially subtracted from itself, i.e. self-subtraction. Self-subtraction can also happen for more face-on circumstellar disks, such that ADI can distort the retrieved disk shape (Sissa et al., 2018; Ruane et al., 2019).

RDI: With RDI an image of a reference star is subtracted from the data to remove the stellar PSF or the target, and was already used by Smith & Terrile (1984) to reveal the circumstellar disk around β Pictoris. RDI assumes that the reference star does not have companions and that the reference PSF is a good representation of the target PSF. Space-based observations have an extremely stable PSF, and by using a library of PSFs that are aberrated differently gathered over many observations it is possible to even remove the full non-coronagraphic diffraction structure. This was demonstrated by Soummer et al. (2011a), who detected HR 8799 b,c,d in data of the Hubble Space Telescope by using RDI to gain approximately two orders of magnitude in contrast. Interestingly, it was archival data of 1998, taken a decade before the first detection of HR8799 by Marois et al. (2008). Examples like this show the true power of data reduction and how far this field has progressed in the last decade and a half. RDI also works for ground-based

observations, where a stable AO performance minimizes the changes between the PSF of a reference star and the target. Observing with a 1.5-m subaperture on the Hale Telescope, Serabyn et al. (2010) were able to detect HR 8799 b,c,d using RDI, with HR8799 d at only twice the diffraction limit (Serabyn et al., 2010). RDI can have an improved performance up to a factor of 5 compared to ADI at the smallest separations (Ruane et al., 2019). In addition, it is less prone to self-subtraction which makes it also useful for circumstellar disks. A different implementation of RDI is possible with the vAPP coronagraph, which creates two copies of the coronagraphic PSF. One can be used as reference for the other PSF by rotating it 180 degrees (Otten et al., 2017).

SDI: Spectral differential imaging encompasses different methods that use simultaneous multi-wavelength observations to differentiate stellar diffraction structures from the planet PSF (Sparks & Ford, 2002). These simultaneous multi-wavelength observations are obtained with an integral field spectrographs (IFS), which combines imaging with spectroscopy. The classical version uses the similarity of the stellar PSF between the IFS observations to create a reference PSF. The reference PSF at a separated wavelength bin will have a different scale due to wavelength scaling. As the separation of the planet and the star is not wavelength dependent, scaling the reference PSF will move the planet in the scaled image. Similar to ADI, taking a mean of these scaled images gives a good PSF model while the planet signal is suppressed. Subtracting the scaled PSF model from the unscaled images for each wavelength reveals the planet. However, in essence this does not use the spectral information that the star has a different temperature and constituents than the planet. This is difficult with the current integral field spectrographs as they have a spectral resolution ($\lambda/(\Delta\lambda)$) between 20-100, e.g. SPHERE/IFS (Claudi et al., 2008), GPI/IFS (Chilcote et al., 2012), SCExAO/CHARIS (Groff et al., 2014), LBT/ALES (Skemer et al., 2018). A downside of SDI is that large spectral bandwidths are required to minimize self-subtraction close to the star. For example, observations of a single band ($\Delta\lambda/\lambda \sim 20\%$) have self-subtraction within $\sim 6\lambda/D$, which also influences the retrieved spectrum. Therefore, IFS data with low-resolution spectrographs is usually treated as completely separate for each wavelength bin.

Medium- and high-resolution IFSs offer a larger tool box due to resolving spectral lines of the stellar and planetary atmosphere. These new tools could be considered a branch of spectral differential imaging as they do use the spectral differences to reject stellar light. A first tool is cross-correlation with a molecular template (Snellen et al., 2010). After removing an averaged and scaled spectrum of a star from each pixel, their residual spectrum should be mostly free of stellar contamination. If a planet is present, the residual spectrum contains a low-signal to noise spectrum of the planet with many molecular features that are too faint to detect. Cross-correlation with a molecular template boosts the signal-to-noise by combining the signal of all molecules into a single signal, and this has been used to detect a few different molecular species in atmospheres of gas giants (Snellen et al., 2015; dit de la Roche et al., 2018; Hoeijmakers et al., 2018a). High-resolution spectrographs with a resolution of $\sim$ 100.000 can resolve the spectral shift (or Doppler

shift) from the movement of the planet around the star. Cross-correlation is again used to boost the signal and the radial velocity of the planet can be determined, which characterizes the orbit and mass of the planet (Snellen et al., 2014).
PDI: The fourth method of polarization differential imaging uses differences in polarization state of the light of the star and surrounding objects. Polarization refers to the preferred oscillation direction of a group of electromagnetic waves in a beam of light. If there is no preferred oscillation direction, the beam is called unpolarized. A beam can be composed of the combination of polarized and unpolarized light, and the degree of polarization is given by the fraction of polarized over unpolarized light. PDI uses that stellar light is highly unpolarized, while scattered light is linearly polarized. Circumstellar disks present large scattering surfaces and PDI has been widely used to discover them (de Boer, J. et al., 2016; Avenhaus et al., 2018). Exoplanets scatter stellar light resulting in a significant degree of polarization, and thus PDI could be used to enhance the contrast. Moreover, it is expected that young gas giants also emit partially polarized light in the thermal infrared, yet the degree of polarization is low ($< 1\%$) (Stolker, T. et al., 2017). Instrumental polarization, i.e. changes in polarization state due to (retarding) optics in the instrument, can result in incorrect measurements of the polarization. However, recent work on instrumental polarization in VLT/SPHERE showed that it can be calibrated using a Mueller matrix model to an absolute polarimetric accuracy of $< 0.1\%$ (De Boer et al., 2020; Van Holstein et al., 2020). Current issues with PDI are polarization aberrations and beam-shifts [4]. The true potential of PDI was shown by van Holstein et al. (2017) by reaching a 1σ contrast of 2×10^{-7} on HR 8799. This resulted in a non-detection, giving an upper limit on the degree of polarization of the HR 8799 planets $< 1\%$ (1σ).
The aforementioned differential techniques improve the contrast, however, this is not the full picture[5]. As discussed in section 1.1, spectroscopy enables the characterization of the atmospheres of exoplanets by measuring its constituents and bolometric flux. Polarimetry can reveal asymmetric structures like oblateness and patchiness clouds for gas giants (De Kok et al., 2011; Stolker, T. et al., 2017; van Holstein et al., 2017) and even determine the presence of liquid oceans (Stam, D. M., 2008; Klindžić et al., 2020) or even life (Patty et al., 2019). Temporal variations can reveal existence of patchy clouds and even more if combined with the other techniques (Apai et al., 2013).

To summarize, detecting and characterizing exoplanets with direct imaging is extremely challenging and requires a large-aperture telescope, high-precision wavefront sensing and correction, a coronagraph with a small inner working angle to suppress the stellar diffraction halo by masking or diffracting stellar light, advanced observational and data-reduction strategies and intricate optical measurement systems, such as polarimetry or integral field spectroscopy (Ruane et al., 2018; Jovanovic et al., 2018; Snik et al., 2018).

[4] van Holstein in prep.
[5] pun intended

1.3 Ground-based and space-based high-contrast imaging

The high-contrast imaging instrument layout is applicable both to ground-based and space-based high-contrast imaging. It is the atmosphere that makes that these systems operate in entirely different regimes in terms of contrast. However, we will show here that despite the atmosphere, they are *complementary* to each other.

1.3.1 The compatibility of ground and space

As explained in section 1.2.2 and 1.2.2, the turbulence in the atmosphere degrades the contrast levels that can be reached compared to lab results by orders of magnitude. While the impact of turbulence decreases for longer wavelengths, the thermal background from the atmosphere becomes the limiting factor beyond ~ 3.0 μm (L-band). Earth's atmosphere also has significant molecular absorption features resulting in an incomplete coverage of the full electromagnetic spectrum. We note that these molecules are also likely to be present in exoplanets similar to Earth and any other type of planets, e.g. super-Earths, mini-Neptunes and gas giants. Examples of these molecules are H_2O and CO_2 and detecting these molecules is critical for our understanding of their atmospheres. With space observatories the aforementioned atmospheric effects are no longer an issue and a full wavelength coverage can be achieved. Therefore, it seems that space-based observatories have a clear advantage over ground-based telescopes. Yet, this advantage was not yet fully exploited due to technology only having been matured over the last decade. No existing or decommissioned space-based observatory has had a dedicated high-contrast imaging mode with adaptive optics. Only lab demonstrations have reached the performance that would be good enough to image planets in reflected light, i.e. 10^{-10} (Traub & Oppenheimer, 2010; Seo et al., 2019). Moreover, it will take over a decade to launch a big space mission that works with the current technology. This time is used to create robust systems that do not need servicing or repairs, which is impossible for space missions that have orbits around the second Lagrangian point. It also means that the current technology is likely not launched until 15 years from now. Lastly, space-based systems are on average an order of magnitude more expensive and need to fit onto a rocket. A most important consequence is that the diameter of the primary mirror can not be very large compared their ground-based counterparts, limiting the spatial resolving power. By comparison, the next generation space telescope, the James Webb Space Telescope, will have a primary mirror diameter of 6.5 m, while the next generation ground-based telescope, the Extremely Large Telescope, will have a diameter of 39 meters. Besides mirror diameter, another advantage of ground-based systems is that they are great for rapidly prototyping technology and software. The difference in resolving power and contrast limit between ground-based and space-based systems is the crux of the harmony between the two, as is shown in Fig. 1.6. Both types probe a different parameter space and provide a comple-

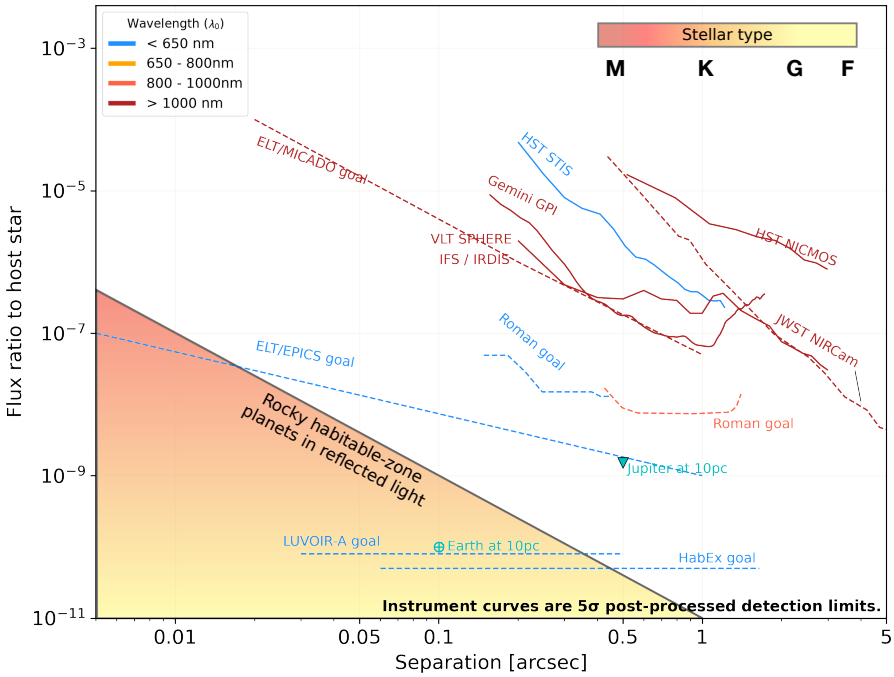


Figure 1.6: Measured or predicted contrast curves of current and future high-contrast imaging facilities. Created using `https://github.com/nasavbailey/DI-flux-ratio-plot`.

mentary look on the exoplanet population, as is most clear for the future space- and ground-based missions. Future large ground-based telescopes are able to look closer-in at a moderate contrast, which is ideal for looking for exoplanets in the habitable zone around M-dwarfs. On the other hand, space-based observatories are stable enough to reach extreme contrasts further out, which is necessary for observing the habitable zone around solar-type stars. Moreover, current telescopes are already oversubscribed and with the next generation this will be worse. It is unlikely that a significant fraction observing time can be spent on large exoplanet surveys, so next-generation telescopes are likely to spend the time characterizing exoplanets found by indirect measurements. These numbers will be large due to the fast amount of current and future missions and surveys with indirect detection methods that are underway and planned, e.g. *Gaia*, TESS, JWST, HARPS3 and ESPRESSO (Snellen et al., 2019). Following up on these targets can and should be divided between all available telescopes to maximize the scientific output and understanding of the planet population. A more in-depth analysis of the timeline and impact of these future missions can be found in section 1.6. Lastly, space-based and ground-based telescopes have similar architectures and the development of subsystems is intertwined. With space operations requiring technology to have reached the maximum technology readiness level, many systems have been further developed from a ground-based system. This is highlighted in balloon missions like HiCIBaS (Marchis et al., 2018) and PICTURE-C (Mendillo et al., 2019), providing a platform to test technology in a space-like environment.

1.3.2 The current state and limitations

With the absence of space-based high-contrast imagers, the scientific output of high-contrast imaging during the last decade has been dominated by ground-based facilities. Mostly driven by technological advances of hardware and new methods of data reduction, the field has transitioned from photometric detection at large separations to characterization of planets a few times the diffraction limit (Rajan et al., 2017; Lagrange et al., 2019). This has been accompanied with the creation of many high-contrast imaging systems at manyx‘ 8-m class telescopes around the world, spanning all astronomical observing bands between 500 nm and 10 μm. Especially the use of an integral field spectrograph enabled the spectral characterization of brown dwarf and exoplanet companions, and polarimetry has been widely used to detect and characterize circumstellar disks. We note that the impact of the development of ever more advanced models of exoplanet atmospheres should not be forgotten in the understanding of observational data. While the improvements lead to a great leap in knowledge of stellar surroundings, the amount of directly detected exoplanets has been low (< 10). The absence of detected exoplanets did put constraints on the demographics of young giant exoplanets by looking at the fraction of stars that have significant detections (Vigan et al., 2020; Nielsen et al., 2019). Two extensive surveys, SHINE and GPIES with SPHERE and GPI respectively, have been completed and observed for hundreds of hours and hundreds of stars. Their results are shown in Fig. 1.7. These surveys clearly demonstrated

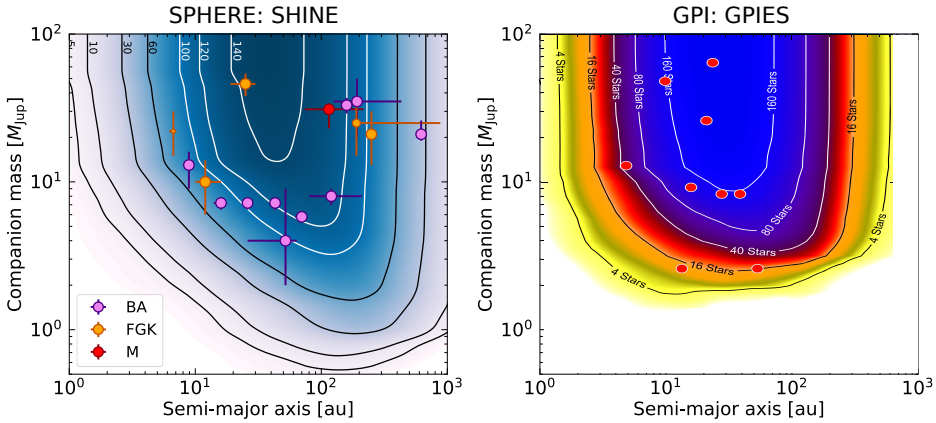


Figure 1.7: *Left*: Depth of search of the SHINE survey for the 150 stars in the sample, from Vigan et al. (2020). *Right*: Depth of search for the first 300 stars observed by GPIES, from Nielsen et al. (2019). The contour lines give the numbers of stars around which the survey is sensitive to substellar companions as a function of mass and semimajor axis. The circles represent the detected substellar companions in the sample.

that the current systems lack the sensitivity and the angular resolution to detect a significant fraction of the exoplanet population. Planned upgrades aim for higher contrasts closer at the closest separations (Chilcote et al., 2018; Boccaletti et al., 2020). Interestingly, some of these limits can be pushed by instruments not necessarily designed for high-contrast imaging. Examples are the detection of PDS 70 c with MUSE, an AO-assisted, non-coronagraphic IFS (Haffert et al., 2019) and the direct detection β Pictoris c with GRAVITY (Nowak, M. et al., 2020). The focus of high-contrast imaging has therefore shifted to developing improved systems, both by decreasing the inner working angle and increasing the contrast performance. A large focus is on the AO-system as this is the limiting factor for all systems, making them more sensitive, faster and adding predictive control. Non-common path aberrations are the next limiting factor, and many systems that apply focal-plane wavefront sensing or low-order wavefront sensing are being investigated (Jovanovic et al., 2018). Some of these solutions are integrated with the coronagraph, e.g. cWFS (Wilby et al., 2017) or SCC (Galicher et al., 2010), and show an integrated way of looking at the high-contrast imaging system. We note that post-processing techniques like ADI can also be seen as integration between the observing strategy and data reduction. However, recently there has been a large increase in new concepts that are being investigated. For example, a coronagraph design that is optimized for the injection of light in a single-mode fiber for high-resolution spectroscopy (Ruane et al., 2018; Por & Haffert, 2020; Haffert et al., 2020). Other examples include combining spectroscopy with polarimetry in Subaru/SCExAO

(Lozi et al., 2019), adapting the coronagraph or dark zone control real time depending on the atmospheric conditions (Carlotti et al., 2018; Radhakrishnan et al., 2018), the integration of photonics with interferometry to decrease the instrument size and increase stability (Jovanovic et al., 2016), or the creation of low-noise detectors with spectral resolving powers like MKIDs (Baselmans, 2012; Meeker et al., 2018). Many of these examples are focused at improving the characterization of exoplanets, which is the ultimate goal of high-contrast imaging, and their timing is fortunate. There will be room for testing these developments on current generation telescopes once the ELTs come online, and the lessons-learned can be implemented in the second-generation ELT instruments to fully exploit the large diameters.

1.4 Broadband enabling technologies

With spectroscopic characterization of exoplanets becoming the main focus of high-contrast imaging, each subsystem of the instrument needs to facilitate broadband technology. For ground-based systems the AO-system delivers good correction over large bandwidths and is common path with multiple imaging systems observing at bands spanning more than an octave in wavelength. Instruments like VLT/SPHERE (Beuzit et al., 2019) and Subaru/SCExAO (Lozi et al., 2018) even have a common coronagraph for these imaging systems. However, most coronagraphs have an optimal performance in a single band ($\Delta\lambda/\lambda \sim 20\%$) due to chromaticity of the phase (PIAACMC, Martinez et al. (2020)) or chromaticity of retardance (VVC, **?**), or the PSF scaling on the focal-plane mask (Lyot). Observing in a single band decreases the observing efficiency and flexibility. Capitalizing on the availability of all instruments is enabled by broadband coronagraphs. The broadband performance of space-based systems is inherently limited by the chromaticity of the wavefront control, reducing the possible instantaneous bandwidth to 20%. To facilitate characterization, the space-based systems are capable of observing over a much wider bandwidth, resulting in the use of many different coronagraphs (HabEx Study Team, 2019). All these coronagraphs require a different calibration due to manufacturing errors. A single broadband coronagraph would make the calibration easier and create flexibility in the observing modes and increases the efficiency of spectroscopy. So, a high-performance broadband coronagraph is a necessity.

Both for some focal-plane coronagraphs and pupil-plane coronagraphs it has been shown that solutions with phase masks instead of amplitude masks provide the most optimal combination of inner working angle and throughput, (Mawet et al., 2009a; Por, 2017, 2020). Here, we will discuss how to make complex phase patterns and how to make them achromatic.

There are different techniques to generate phase patterns. The most common technique generates differences in the local optical path length to offset the phase (Born & Wolf, 2013). Optical path differences (OPD) are created by travel time variations of light, either by changing the travel distance d, or the speed through

the differences in refractive index n. A phase shift caused by the OPD is also called the dynamic or classical phase shift, and is given by

$$\phi = \frac{2\pi}{\lambda} nd. \tag{1.1}$$

Changing either n, d, or both results in a phase shift and many different technical implementations exist. For example, a deformable mirror changes only d. Moreover, thickness variations Δx of materials with a high refractive index n_{plate} compared to the refractive index of the embedded medium n_{air} to create a phase delay $\Delta\phi$, given by

$$\Delta\phi = \frac{2\pi(n_{\text{plate}} - n_{\text{air}})\Delta x}{\lambda}. \tag{1.2}$$

This technique of phase generation is well known and has many applications, e.g. lenses. Another technique uses changes in refractive index of an optic axis of a liquid-crystal birefringent medium (= retarder) to introduce phase differences. These changes in refractive index are created by applying a voltage to the liquid crystals, modifying their orientation along the axis of propagation of the incoming light. By creating a device with different pixels, each with their own voltage regulator, any phase pattern can be created within the limits of the refractive index modulation. Such a device is called a spatial light modulator (SLM) (Love, 1997). All these implementations of dynamical phase have their own advantages, but they also have two common downsides. The generated phase is wavelength-dependent and deviating from the optimal relative phase degrades the performance of the coronagraph. Moreover, it is difficult to accurately create freeform phase patterns for these transmissive optics, limiting the parameter space for coronagraphs like the APP (Codona et al., 2006). An SLM is an exception to this last point, however it suffers from polarization effects (Yeh, 2007).
Ideally, the phase plates of coronagraphs would be built as a device that is capable of generating arbitrary achromatic phase differences. Such a device can be realized through geometric phase, which was discovered by S. Pancharatnam in 1956 (Pancharatnam, 1956), and later rediscovered for quantum mechanics by M. Berry in 1984 (Berry, 1984). Geometric phase, now also called Pancharatnam-Berry phase, arises when light travels through a retarder with a locally changing *orientation* of the fast-axis rather than a changing thickness or refractive index (Escuti et al., 2016). First, we will go into the properties of the geometric phase, then we will present a short derivation using Jones matrices, followed by an intuitive explanation of geometric phase that is connected to the history of geometric phase, and we conclude with the technical implementation of geometric-phase optics using liquid-crystal technology.

1.4.1 Properties of the geometric phase

Geometric phase, as the name implies, only depends on the local geometry of the optic it travels through. For an optic made out of glass, the refractive index is the same for any orientation of the light (or linear polarization state), and geometric

phase is not generated. On the other hand, birefringent materials are anisotropic, meaning that the refractive index changes depending on the polarization state of the incoming light. Therefore, their orientation with respect to the polarization state of light matters and different orientations lead to changes in polarization state and phase of the light that passes through the optic (Pancharatnam, 1956; Escuti et al., 2016). This orientation is defined as an angle between the oscillation direction of a ray of light and some axis of anisotropy in the birefringent material, i.e. the crystal axis. Optics from birefringent materials (= retarder) are commonly manufactured with the anisotropy in a single direction, i.e. the optic axis, that is parallel to the front surface of the optic. The refractive index is n_o for linearly polarized light that oscillates parallel to this optic axis, while the refractive index for the orthogonal polarization state is n_e. Either of the two axes with the lowest refractive index is called the fast axis, and the axis with the highest refractive index is called the slow axis. The (dynamic) phase delay between these two axes is the retardance and depends on the thickness and refractive indices of the optic, and the wavelength of the light. If the linear polarization state does not align with any of the two axes, the light-wave can be decomposed in two components that are parallel and orthogonal to the optic axis. The phase delay between these components leads to a change in the output polarization state. As an example, we will look at a retarder with a phase delay that equals π (= half-wave retarder) with an angle χ between the fast-axis and the linear polarization state. The output polarization state is also linear, but rotated by an angle $2\chi = \Phi$. Circularly polarized light travelling through the same half-wave retarder switches orientation such that left-circular polarization becomes right-circular polarization and vice versa, independent of fast-axis orientation. However, the phase of the output circular polarized light *is* dependent on the fast-axis orientation χ. The total phase delay of the circularly polarized light travelling through the half-wave plate is given by

$$\phi = \frac{2\pi n_{slow} d}{\lambda} \pm 2\chi. \tag{1.3}$$

The first part is the dynamic phase and is independent of the geometry of the optic axis, while the second part does depend on the orientation of the optic axis but is independent of wavelength. The second part is the geometric phase and also depends on the circular polarization state of the light coming in, where Φ is positive for left-circularly polarized light. If we compare the phase shift between circularly-polarized light with the same orientation travelling through the same retarder with a different orientation $\Delta\chi$, the phase *difference* is purely geometric phase.

$$\Delta\phi = \pm 2\Delta\chi = \pm\Delta\Phi. \tag{1.4}$$

A half-wave retarder with a fast-axis orientation pattern therefore is an optic that generates a phase pattern with a phase that is twice the angle of the fast-axis orientation pattern. Such an optic can also be called a geometric phase hologram (GPH), and has four important properties.

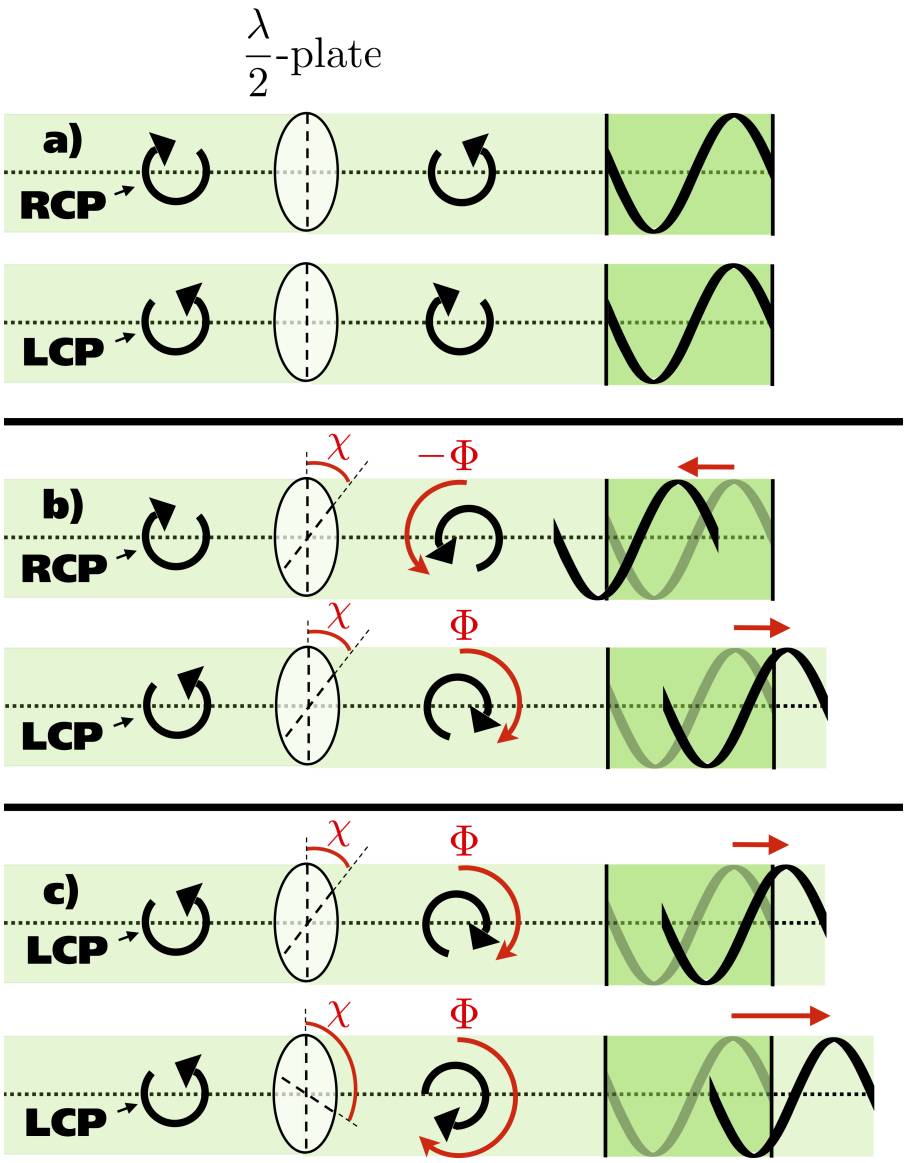


Figure 1.8: Properties of circularly polarized light passing through a half-wave retarder. **a)** Left-handed circularly polarized light (LCP) becomes right-handed circularly polarized light (RCP) and vice versa with a dynamical phase delay which we set as a reference. **b)** Rotating the half-wave retarder compared to the reference orientation results in an additional phase delay Φ called the geometric phase. The sign of the phase delay depends on the handedness of the incoming beam. **c)** The geometric phase shift Φ is plus or minus twice the rotation angle χ.

1. A GPH flips the circular polarization state.
2. The applied phase has an opposite sign for opposite handedness of the incoming circular polarization state.
3. A GPH applies geometric phase that only depends on the local fast-axis orientation and is therefore independent of wavelength.
4. When the retardance deviates from half-wave, the diffraction efficiency decreases and polarization leakage emerges.

A graphical representation of the first three properties is shown in Fig. 1.8 and Fig. 1.9. The fourth property can represent a major technical limitation, which is discussed in section 1.4.4

1.4.2 Geometric phase in the Jones formalism

To derive the properties of the geometric phase, we will have to describe polarized light in the Jones formalism. A beam of light that travels in the z-direction in Cartesian coordinates consists of an electric field that oscillates perpendicular to the z-direction. Decomposing this oscillation in the x-y plane results in the description of light in the Jones formalism (Born & Wolf, 2013), i.e.

$$\mathbf{E} = \begin{pmatrix} E_x \\ E_y \end{pmatrix} \tag{1.5}$$

If these components are in phase, the polarization state is linear, and the polarization state is circular if they are phase shifted by $\pi/2$ and have equal amplitude. Any other state is referred to as elliptical polarization, a combination of the linear and circular. Any optical element can then be written as a 2×2 Jones matrix and the Jones matrix of a linear retarder (LR) $\mathbf{J}_{\mathrm{LR}}$ is given by

$$\mathbf{J}_{\mathrm{LR}} = \begin{pmatrix} e^{i\phi_1} & 0 \\ 0 & e^{i\phi_2} \end{pmatrix}, \tag{1.6}$$

where $\phi_1 = \frac{2\pi}{\lambda} n_{\mathrm{o}} d$ and $\phi_2 = \frac{2\pi}{\lambda} n_{\mathrm{e}} d$. Here we assume that the optic axis is in the y-direction. The rotation matrix $\mathbf{J}_{\mathrm{rot}}(\chi)$ is given by

$$\mathbf{J}_{\mathrm{rot}}(\chi) = \begin{pmatrix} \cos\chi & \sin\chi \\ -\sin\chi & \cos\chi \end{pmatrix}. \tag{1.7}$$

The output Jones vector $\mathbf{E}_{\mathrm{out}}$ is then given by

$$\mathbf{E}_{\mathrm{out}} = \mathbf{J}_{\mathrm{rot}}(-\chi)\mathbf{J}_{\mathrm{LR}}\mathbf{J}_{\mathrm{rot}}(\chi)\mathbf{E}_{\mathrm{in},\pm} = \mathbf{J}\ \mathbf{E}_{\mathrm{in}} \tag{1.8}$$

For a half-wave plate $\phi_2 - \phi_1 = \pi(1 + 2 * n)$ with n an integer, and $\mathbf{J}$ becomes

$$\mathbf{J} = e^{i\phi_1} \begin{pmatrix} \cos\chi & -\sin\chi \\ \sin\chi & \cos\chi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{i(\phi_2-\phi_1)} \end{pmatrix} \begin{pmatrix} \cos\chi & \sin\chi \\ -\sin\chi & \cos\chi \end{pmatrix} \tag{1.9}$$

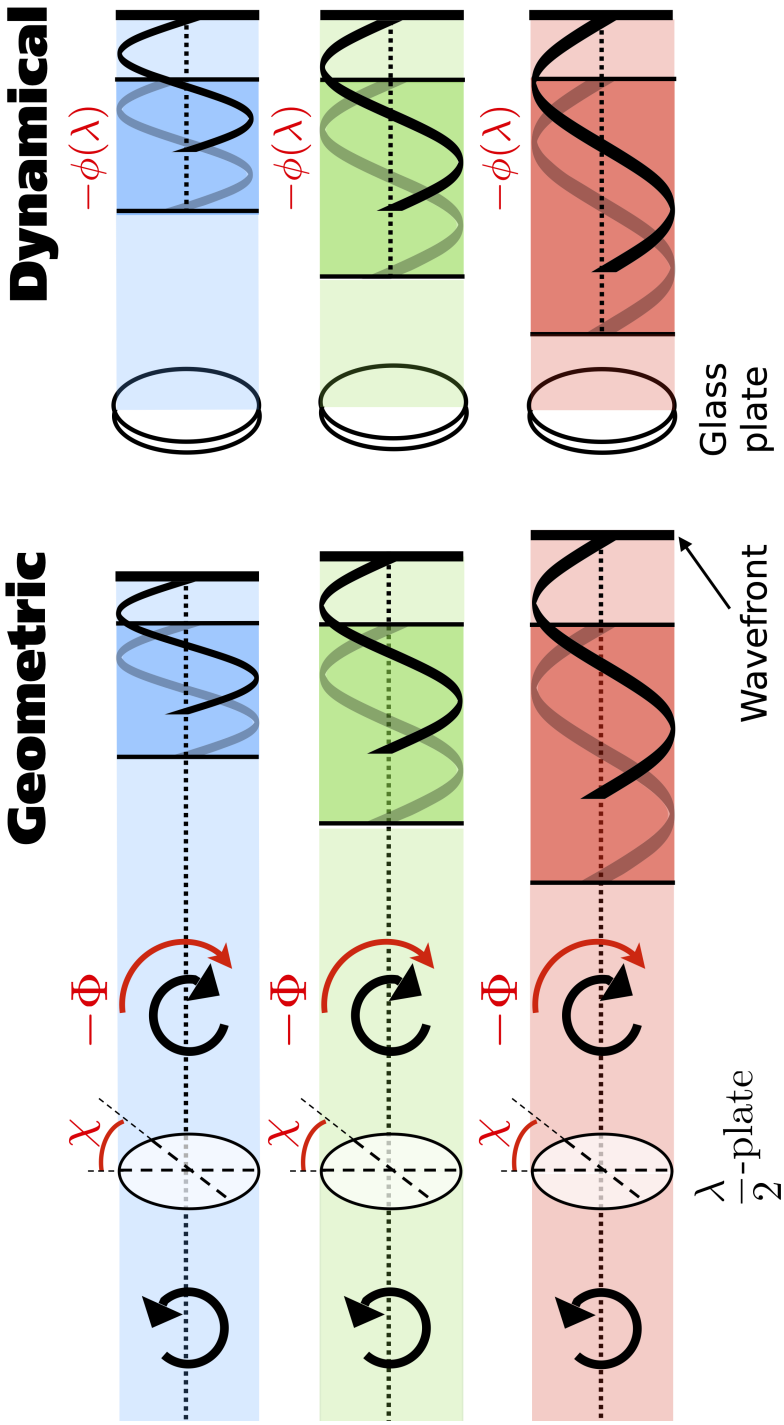


Figure 1.9: Comparison between the geometric phase and dynamical phase. The geometric phase shift is the same for each wavelength, while the dynamical phase shift is the same in terms of optical path length.

$$\mathbf{J} = e^{i\phi_1} \begin{pmatrix} \cos 2\chi & \sin 2\chi \\ \sin 2\chi & -\cos 2\chi \end{pmatrix} \tag{1.10}$$

Switching to the circular polarization basis reveals the beauty of the geometric phase for a rotating half-wave retarder (Mawet et al., 2009b):

$$\mathbf{J_{CP}} = e^{i\phi_1} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} \cos 2\chi & \sin 2\chi \\ \sin 2\chi & -\cos 2\chi \end{pmatrix} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}^{-1} = e^{i\phi_1} \begin{pmatrix} 0 & e^{2i\chi} \\ e^{-2i\chi} & 0 \end{pmatrix}. \tag{1.11}$$

For circularly polarized light the Jones vector in the circular polarization basis is given by

$$\mathbf{E}_{\mathrm{in},LH} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \mathbf{E}_{\mathrm{in},RH} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.12}$$

From Eq. 1.11 and Eq. 1.12 the first three properties of a GPH are revealed:

$$\mathbf{E}_{\mathrm{out}} = \mathbf{J_{CP}}\mathbf{E}_{\mathrm{in},-} = e^{i\phi_1} \begin{pmatrix} 0 & e^{2i\chi} \\ e^{-2i\chi} & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = e^{i(\phi_1+2\chi)} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{1.13}$$

So we see that the circular polarization state is flipped, the geometric phase is opposite for opposite polarization states and phase only depends on the orientation of the fast axis when comparing the phase of two different orientations.
To derive what happens if the retardance is not half-wave we introduce a deviation from half-wave retardance ϵ, such that $\mathbf{J}$ becomes

$$\mathbf{J} = e^{i\phi_1} \begin{pmatrix} \cos\chi & -\sin\chi \\ \sin\chi & \cos\chi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{i(\pi+\epsilon)} \end{pmatrix} \begin{pmatrix} \cos\chi & \sin\chi \\ -\sin\chi & \cos\chi \end{pmatrix}. \tag{1.14}$$

In the circular polarization basis, the Jones matrix of the system is given by

$$\mathbf{J_{CP}} = c_V e^{i\phi_1} \begin{pmatrix} 0 & e^{2i\chi} \\ e^{-2i\chi} & 0 \end{pmatrix} + c_L e^{i\phi_1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{1.15}$$

where $c_V = \cos(\epsilon/2)$ and $c_L = \sin(\epsilon/2)$. The efficiency, defined as the fraction of intensity that acquires the geometric phase, is given by $|c_V|^2$ and depends only on the retardance. The fraction of intensity that does not acquire the geometric phase is called the *leakage term*, and is given by $|c_L|^2$. Also important to note is that the leakage term does not flip polarization states, resulting in the leakage term having the orthogonal circular polarization state as the main beam. This can be used for filtering the leakage term with circular polarizers (Mawet et al., 2010) or with polarization gratings as described in Chapter 4.
Another way of presenting polarized light is the Stokes formalism, where the light is represented by the Stokes vector:

$$\mathbf{E} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix}, \tag{1.16}$$

where I is the total intensity of the beam, Q and U represent linearly polarized light with an oscillation in the 0° (x-direction) and 45° respectively, and V describes circularly polarized light. The total intensity $I \geq \sqrt{Q^2 + U^2 + V^2}$. For any degree of polarization, the polarization state can described by a vector in Q, U, V-space and a length of $\sqrt{Q^2 + U^2 + V^2}$. Objects that change the polarization state without depolarization, such as ideal retarders, change orientation of the vector by rotating it in the (Q,U,V) subspace of the 1+3 Hilbert space. All possible polarization states with the same length form a sphere in the Q, U, V-space, which is called the Poincaré sphere (Poincaré, 1892). This sphere is useful for a visual representation of the polarization state as the phase differences between the Jones vector components are represented by unique points on the sphere. The Poincaré sphere can also be used to visualize geometric phase.

1.4.3 A different look on geometric phase: a personal note

Whereas the creation of classical phase delays by optical path length differences is easily explained in text and figures, providing an intuitive explanation of geometric (Pancharatnam-Berry) phase is much harder. In literature, abstract descriptions are often presented, e.g. "A geometric-phase shift, by contrast, arises as a kind of 'memory' of the evolution of a lightwave through an anisotropic parameter space" (Escuti et al., 2016). While accurate, I found these explanations did not give me a good idea *how* geometric phase arises from geometry differences, just that it does. In the Jones formalism, mathematically the geometric phase emerges from the coordinate transformations of the Jones matrix given by the orientation of the half-wave retarder, but I did not find that much enlightening. As this thesis is centered around geometric phase, it is clear that this lack of a good explanation would be unacceptable. A clear description was formulated by Michael Berry himself (Berry, 1988):
"Take a pencil, lay it on the north pole of a globe and point it in the direction of any of the meridians: the lines of longitude that radiate from the pole. Move the pencil down along the line to the equator and, keeping it perpendicular to the equator, slide it to another line of longitude. Move the pencil back to the north pole along the new meridian, and you will find that although the pencil has been returned to its starting point and at no time was rotated, it no longer points along the original line of longitude."
This example clearly shows that the orientation of the pen depends only on the geometry of the circuit, not on the energy or mass of the pencil, or even its initial direction. However, it is seemingly disconnected to the geometric phase of half-wave retarders. This could not be further from the truth. It was the pioneering work of Shivaramakrishnan Pancharatnam that showed that geometric phase, polarization, and the area of closed curves on the Poincaré sphere are strictly connected (Pancharatnam, 1956). In fact, his work marks the start of the theory of geometric phase, which he measured almost 30 years before Michael Berry formalized the same theory for quantum mechanics. In his work, Pancharatnam used birefringent wave plates (= retarder) and a Mach-Zender interferometer to

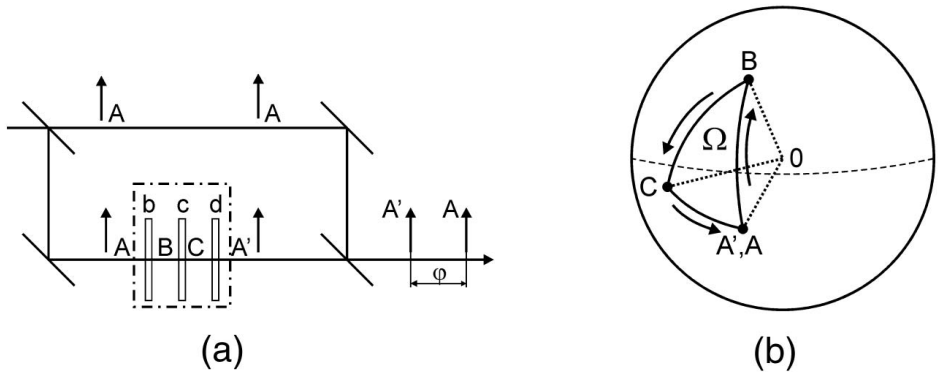


Figure 1.10: a) Set of birefringent media in a Mach-Zehnder setup and b) proper construction of the Pancharatnam's triangle to calculate the geometrical phase. From Kurzynowski et al. (2011).

investigate the interference between polarization states. His setup is described in Fig. 1.10 a), where he changed the polarization states of a second arm from input polarization **A**, to **B**, **C** and back to **A**', using the retarders *b*, *c* and *d*. Here, **A**' is identical to **A**, which is the unchanged polarization state of the first arm and the input. Pancharatnam showed that the second arm experiences a phase shift ϕ with respect to the first arm, and that the phase shift depends only the intermediate polarization states **B** and **C**, i.e. the history of **A**'. More importantly, he proved that this phase shift ϕ is equal to half the area (or steradian) of the geodesic triangle on the Poincaré sphere, Ω with the corners defined by the Stokes vectors of the polarization states **A**, to **B**, **C**. This geodesic triangle is shown in Fig. 1.10 b). The retarders are moving the polarization state by parallel transport or rotation on the surface of the Poincaré sphere, equivalent to the movement of the pen in the example of Berry.

The theory of Pancharatnam was later rediscovered for the rotation of the spin in quantum mechanics by Berry, who investigated (adiabatic) cyclic changes of quantum systems (Berry, 1984). For cyclic changes, by definition the output state is returned to the original state, like the polarization state in the experiment of Pancharatnam. This is different for half-wave retarders, where the output state is the opposite circular polarization state. So, how does this reconcile with the geometric phase arising from the cyclic change? For a single beam and a retarder, one can theoretically reconstruct a triangle using the wave plate's eigenwaves (Courtial, 1999) and reconstruct the dynamical phase and geometric phase from this triangle (Kurzynowski et al., 2011). However, conceptually easier is the interference between two beams, where a closed area on the Poincaré sphere can be formed if the two input and output polarization states are the same and their propagation distance is equal. This area is non-zero if each path experiences a different variation. For example, if the retarder *d* is flipped and moved from the second beam

to the first, the output polarization states are **C** and **C**', with the same relative phase delay $\phi = \Omega/2$. The area does not necessarily have to form a closed triangle, other shapes are also possible if they form a closed area on the Poincaré sphere. This connects the half-wave retarders – with differently oriented fast-axis -flipping one circular polarization state to the other – to geometric phase, as is shown in Fig. 1.11 a). For a relative rotation of a half-wave axis, $\Delta\chi$, of a perfect half-wave retarder, the phase shift is $\Delta\Phi = 2\Delta\chi = 4\Omega$.
To summarize, the Poincaré sphere is a simple and effective visualization tool for the geometric (Pancharatnam) phase of light that is acquired when changing polarization states after being transmitted through an anisotropic medium. An identical visualization of the geometric (Berry) phase of adiabatically changing quantum systems can be made using the Bloch sphere (Bloch et al., 1946), which mathematically has the same properties. Another example is a sphere (unnamed) existing of Laguerre-Gaussian modes and Hermite-Gaussian modes of light (Padgett & Courtial, 1999; Alonso & Dennis, 2017), where also geometric phase accumulates for closed-loop transformations between these modes. The commonality of these examples is the angular momentum of the particles, that is circular polarization (= spin) and orbital angular momentum for photons and spin for the quantum systems. Photons have three different kinds of angular momentum, which are fundamentally connected through spin-orbit interactions and geometric phase (Bliokh et al., 2015). These spin-orbit interactions provide a theoretical frame work for many phenomenon in optics that are potentially interesting for high-contrast imaging systems. Examples are the spin-Hall effects in inhomogeneous media and at optical interfaces (e.g. the Imbert-Federov effect) (Bliokh & Aiello, 2013), and the interaction with subwavelength structures used for spin-controlled shaping of light, e.g. metasurfaces. Metasurfaces are a type of diffractive optics that are created with subwavelength features and enable pixelated control over the electric field and thus dynamic phase, polarization state, and geometric phase (Kildishev et al., 2013; Yu & Capasso, 2014). The term metasurface encompasses multiple types of subwavelength structures which modify the light based on different optical phenomenon (Huang et al., 2018). For example, dielectric metasurfaces realize dynamical phase differences with form birefringence, creating a local change in the refractive index with anisotropic subwavelength structures. Dielectric metasurfaces can also be used to generate geometric phase by rotating identical subwavelength structures (equivalent to liquid-crystal optics), or control both dynamical and geometric phase by changing both the structure shape and orientation (Rubin, 2020). Metallic metasurfaces are based on plasmonics, the interaction of light with a metallic interface, and they provide superior control of the light through spin-orbit interactions, e.g. surfaces with both polarization and wavelength dependent diffraction for simple polarimeters (Chen et al., 2016). Metasurfaces have a greater design freedom which could lead to new devices that can improve the performance of high-contrast imaging systems, much like the liquid-crystal devices in this thesis. Currently, metasurfaces with high diffraction efficiencies over large bandwidths can not be realized, and could face the same difficulties as stacks of independent patterned retarders (Mawet et al., 2006).

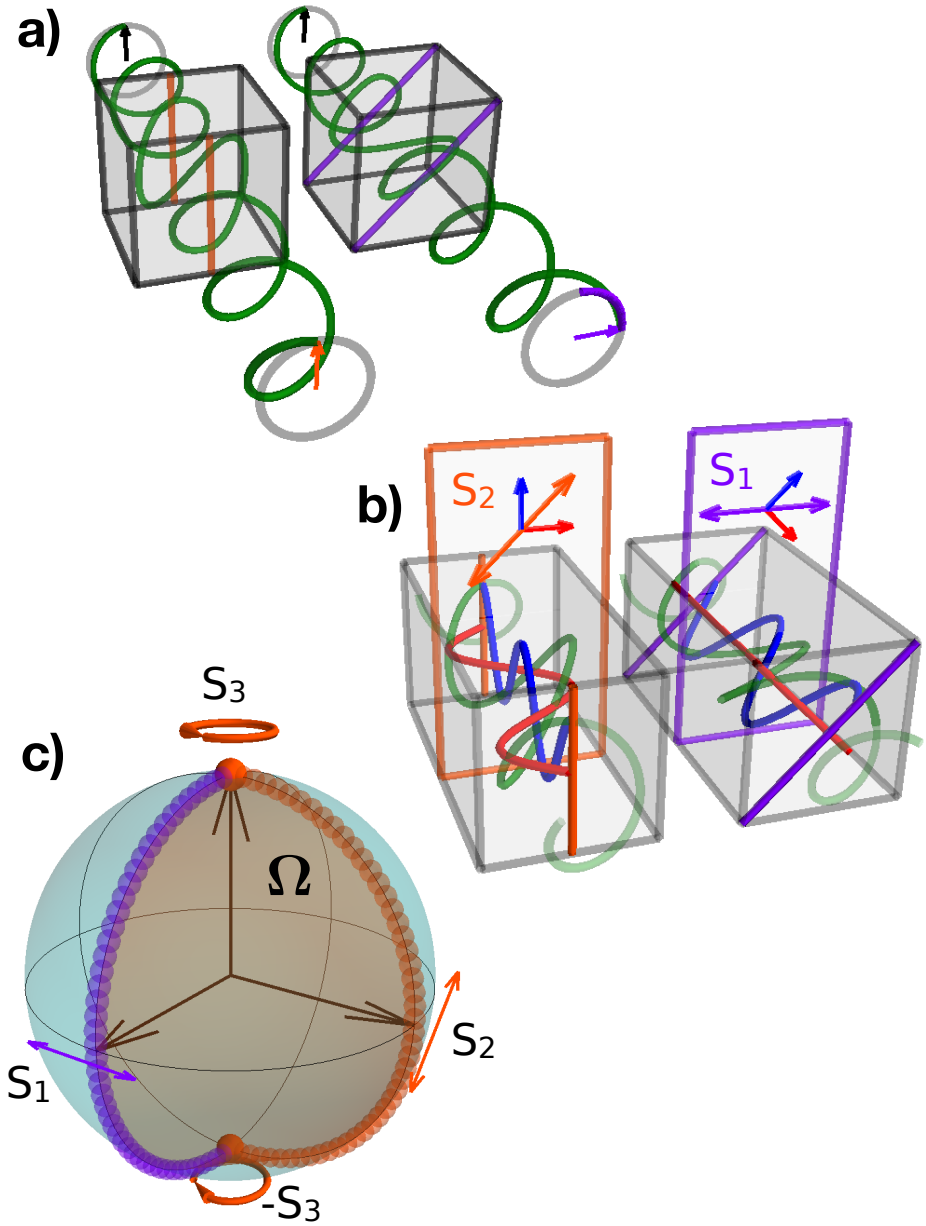


Figure 1.11: The effect of two neighbouring cells of a half-wave retarder with varying fast-axis orientations (0, 45°) on the incoming circularly polarized light. **a)** The resulting geometric phase represented by a phasor. **b)** The polarization states half-way through the same two cells, showing that the state depends on the fast-axis orientation. **c)** The closed curves on the Poincaré sphere with the area $\mathbf{\Omega}$ that is twice geometric phase, i.e. the area of a sphere is 4π and a quarter is π.

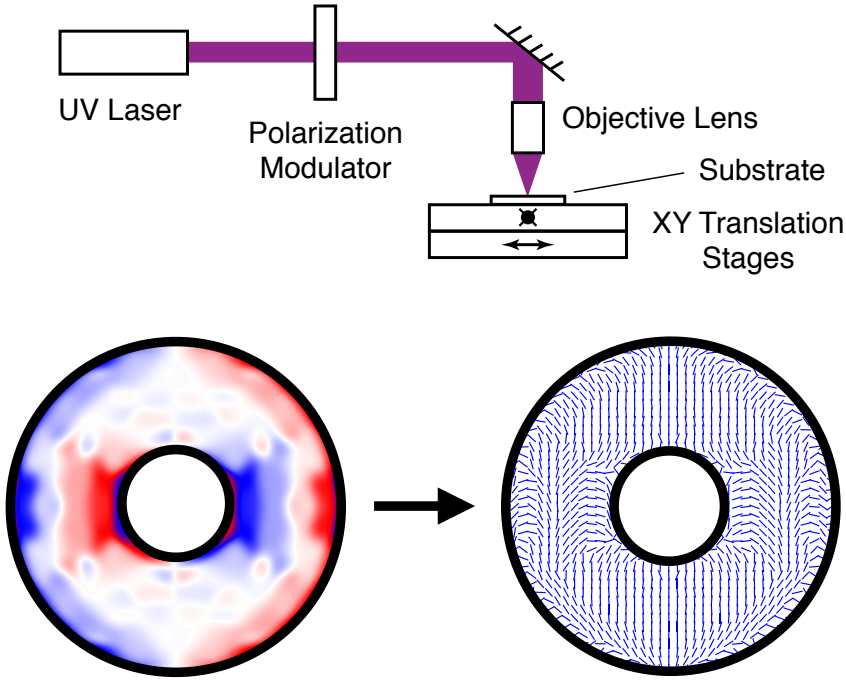


Figure 1.12: *Top:* Schematic of the direct-write system used to pattern a photo-alignment layer (PAL). The PAL records the orientation of incoming linearly-polarized UV radiation which is controlled by the polarization modulator. The substrate is translated to record arbitrary patterns. Adapted from Miskiewicz & Escuti (2014). *Bottom:* The phase pattern of a vAPP is encoded in the orientation of the fast axis of a half-wave retarder, and is written by the direct-write system. The phase pattern was used for the MagAO vAPP (Otten et al., 2017).

1.4.4 Technical implementation of a GPH

Theoretically, a geometric-phase hologram can be described as a half-wave retarder with a spatially varying fast-axis orientation. The technical implementation therefore will have to realize two important features. First, it needs to be capable of changing the fast-axis orientation as function of position. Second, it needs to have a half-wave retardance over a large bandwidth to ensure a small leakage term at all relevant wavelengths, especially to capitalize on the achromatic nature of the geometric phase. Liquid crystals are birefringent, so a thin layer can be used to create a zero-order half-wave plate. Patterning liquid crystals and achromatization are enabled by liquid-crystal technology. Creating arbitrary patterns is possible with a photo-alignment layer (PAL) (Chigrinov et al., 2008). Polymers in the PAL record the orientation of incoming linearly-polarized UV radiation, and direct-write laser scanning (Miskiewicz & Escuti, 2014; Kim et al., 2015) allows

for accurately writing arbitrary patterns, An example of the direct-write system is shown in Fig. 1.12. As the geometric phase is twice the fast-axis orientation angle, the written pattern is half of the desired phase. With the high-precision translation stages and a small beam size, it is currently possible to accurately write phase patterns with a resolution down to ~ 1 μm or cover an area up to 30×30 cm with a lower resolution. Therefore, the direct-write system enables the manufacturing of extreme phase patterns that can be required for vAPP coronagraphs or create vortex coronagraphs with extremely small central defects.
It was again Pancharatnam who first realized that achromatic retarders could be created with combinations of birefringent wave plates (Pancharatnam, 1955). Specifically, he showed with geometric arguments on the Poincaré sphere that a combination of three retarders, where the first and the last one have the same retardance and orientation, introduces a combined retardance that is similar over a larger wavelength range. The achromatization of phase plate coronagraphs started similarly, with the use of a two-stage stack of Quartz and MgF_2 for the four-quadrant phase-mask coronagraph (Mawet et al., 2006). Moreover, a three-stage stack was tried for the eight-octant phase-mask coronagraph manufactured with photonic crystals, which have periodic nanostructures with high and low refractive indices (Murakami et al., 2010, 2016). The form-birefringence of subwavelength structures can also be achromatized over 20% bandwidth, which has been used to create a vector-vortex coronagraph, i.e. the annular groove phase mask coronagraph (AGPM) (Mawet et al., 2005b; Catalán et al., 2016). A problem with a multi-stack is the alignment of the masks with respect to each-other, while the bandwidth of an AGPM is limited to 20%. Liquid crystals have a property that can be used to overcome these issues. The initial hologram orientation pattern in the PAL is passed on to the first and subsequent layers by *spontaneous self-alignment*. This way, liquid-crystal layers with different thickness and chiral twists may be arranged into a monolithic film (Komanduri et al., 2013). Another option is to combine materials with different refractive index (Schirmer & Schmidt-Kaler, 2000). The many degrees of freedom of these multi-twist retarder (MTR) films – given by the thickness and twist of each layer – may be used to tune the retardation of the liquid-crystal layer to half-wave retardance, achieving high efficiency (Kim et al., 2015; Escuti et al., 2016) in GPHs over large bandwidths into a single diffracted order. The tuning is an optimization problem that takes into account the physical constraints on the maximum twist to minimize the difference between the MTR retardance and half-wave retardance for a given wavelength range. An example of the monolithic structure of an MTR is shown in Fig. 1.13 a). Liquid-crystal MTRs can also be described on the Poincaré sphere, although their path is not a simple straight line on the surface. As MTRs are a combination of multiple birefringent liquid-crystal layers that act as separate retarders, each layer rotates the polarization on the Poincaré sphere. The combination of the layers can be manufactured to behave as a half-wave retarder, where the combined rotation is from one pole to the other for circularly polarized light, see Fig. 1.11 b). While the circuit on the surface of the Poincaré sphere is different for MTRs, the total area is the same as for a single half-wave retarder, resulting in the same geometric

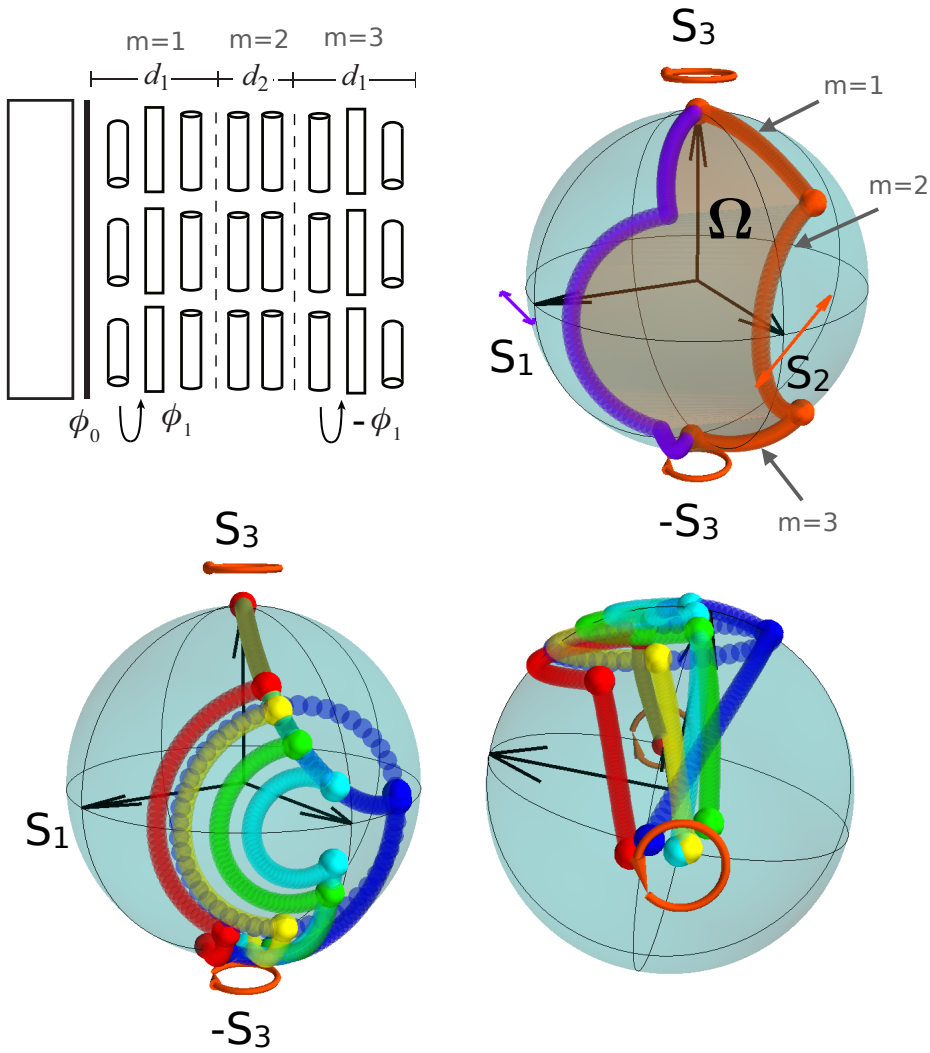


Figure 1.13: **a)** Illustration of the multi-twist retarder (MTR) monolithic structure, with one substrate and alignment layer, and three layers following a Pancharatnam design. The MTR design has a high efficiency (> 97%) between 2-5 μm. Adapted from Komanduri et al. (2013). **b)** The closed curves on the Poincaré sphere of two MTRs with a fast-axis orientation of (0,30°). **c)** A side view and **d)** a bottom view of the curves of five different wavelengths on the Poincaré sphere, equally spaced between 2 μm (blue) and 5 μm (red).

phase. A single retarder has a wavelength dependent retardance, resulting in a different rotation on the Poincaré sphere for each wavelength. This principle holds for MTRs and can be used to optimize the recipe for close-to half-wave retardance over a large wavelength range. A Pancharatnam design for MTRs results in a first layer and third layer with the same thickness and an opposite twist, and a second layer with a different thickness and no twist. The chromatic behavior of such a design for a three-layer twist retarder is shown in Fig. 1.13.
With the direct-write and multi-twist retarder technologies it is possible to write any phase pattern with a high efficiency ($> 96\%$) over large bandwidths ($\delta\lambda/\lambda = 100\%$). Another advantage is the capability to write unbounded continuous phase while being completely flat. This advantage enables the increase of diffraction efficiencies for gratings, with up to 100% for a single order depending on the retardance. Such a liquid-crystal diffraction grating is called a polarization grating (PG), and has a continuously rotating fast-axis. A PG can also be described as a geometric phase prism because it constitutes a continuous phase ramp. Polarization gratings are diffractive polarization beam splitters due to the sign of the geometric phase changing with circular polarization state. The high efficiency, diffraction (= chromatic beam steering), and the polarization splitting property of PGs make them a versatile tool in optics. With the direct-write it is also possible to created shaped PGs, adding them to a phase pattern to separate polarization states or steering parts of a beam to different locations or indeed total freeform holograms. In this thesis, many applications of shaped PGs are discussed and how they can be used to overcome the challenges that come with the use of liquid-crystal technology. So far, we have not found an indication that the writing accuracy is not good enough for the applications in this thesis, so the main technological challenges are the minimization of the polarization leakage, the diffractive nature and for some applications the phase having the opposite sign for opposite circular polarization states. Polarization leakage limits the achieved contrast for many coronagraphs, e.g. vAPP and VVC, while the diffractive nature can be a problem for separating circular polarization states with PGs. The opposite sign of the phase reduces the planet throughput by a factor two for the vAPP and results in a diversity in aberrations between the two polarization states for the VVC (Riaud et al., 2012). One solution for the polarization leakage for the vAPP is adding a polarization grating pattern to the vAPP phase, i.e. the grating-vAPP (gvAPP) (Otten et al., 2014). However, this makes the vAPP diffractive and the planet light and stellar speckles are also diffracted, reducing the contrast. These examples clearly demonstrate some of the issues that occur when using liquid-crystal technology. In this thesis we present similar issues and, more importantly, how these can be solved by using more than one polarization grating.

1.5 This thesis

The goal of this thesis is to use the strengths of liquid-crystal technology to improve and integrate components of high-contrast imaging systems to facilitate detailed characterization of exoplanets. We use the extreme patterning and the inherently achromatic geometric phase to create optics that operate over large bandwidths, are more sensitive, or have additional capabilities. To this end, Chapter 2 of this work shows that liquid-crystal coronagraphs, such as the vAPP coronagraph, can be easily adapted to any coronagraphic system with good coronagraphic performance and summarize how it can be adapted for wavefront sensing or improving exoplanet photometry. Moreover, in Chapter 3 and 4 we demonstrate that improving the broadband performance of liquid-crystal coronagraphs is necessary to realize the full potential of integral field spectrographs and efficient exoplanet characterization. In Chapter 5 and 6 we enhance sparse aperture masking, capable of detecting companions beyond the diffraction limit, by using liquid-crystal phase masks to enable low-resolution spectroscopy and using most of the pupil. Both the liquid-crystal aperture mask and a vAPP with multiple gratings have gone through the complete development cycle, starting with the idea of a novel concept, to design, simulation, lab testing, on-sky verification, to science observations. In Chapter 7 we demonstrate that a broadband liquid-crystal version of the Zernike wavefront sensor can accurately and efficiently measure phase and amplitude aberrations simultaneously, facilitating extreme contrasts. The last Chapter, 8, describes a general method to efficiently modulate broadband light with independent phase control of orthogonal circular polarization states and wavelength bins.

Chapter 2: The vAPP coronagraph: creating dark holes in PSFs all over the world

The vector-Apodizing Phase Plate (vAPP) coronagraph is a versatile single-optic pupil-plane coronagraph that is easily adapted to any telescope aperture. Manufacturing vAPPs with the direct-write method for liquid-crystal technology enables extreme and accurate patterning and the multi-twist retarder technology provides excellent efficiencies ($> 96\%$) for bandwidths up to 100% between 0.55 μm and 5 μm. Furthermore, the vAPP is insensitive to tip-tilt errors, can easily be adapted to include focal-plane wavefront sensing capabilities, and can be designed with holographic off-axis PSFs that can be used as an astrometric and photometric reference. In this chapter, we report on the design, manufacturing, installation, verification, and the first on-sky results of seven different vAPPs in six different instruments. We give an overview of the different vAPPs and their instruments and explain their individual design choices. For all vAPPs we find that they are operating as expected by measuring PSF quality with the internal source of their respective instruments. We show how the data reduction of vAPPs and their D-shaped dark zone make observing and data reduction different from other coronagraphs and present preliminary software solutions. Lastly, we describe the

designs of vAPPs baselined in future instruments, i.e. VLT/ERIS, ELT/METIS, and ELT/MICADO. Together, the vAPPs in these future instruments operate between 1.1 μm and 5 μm, and will provide high-resolution spectra ($R > 20,000$) of many already known, RV-detected planets, in addition to detecting and characterizing new planets.

Chapter 3: First low-resolution L-band spectroscopy of HR 8799 c,d,e with a LBT/ALES-vAPP combination.

In recent studies, both atmospheric modelling and IFS observations of directly imaged gas giants have focused on y-K band (0.9-2.5μm). However, observations of gas giants in the mid-infrared (3-5 μm) are critical to remove model degeneracies, determine their cloud properties, and give insight in their L-T transition. In this chapter, we report on the first on-sky results of the upgraded mid-infrared ALES spectrograph in combination with the newly installed vAPP coronagraph. This combination provides a unique window for exoplanet atmosphere studies, capable of providing low-resolution spectra between 2-5 μm using multiple instrument modes. We present HR 8799 observations with the ALES-vAPP combination using the L-band mode, spanning 2.8-4.2 μm with a spectral resolution of $\sim$ 35 representing the first mid-IR spectroscopy of HR8799 e. We employed a new observation strategy without nodding and only a limited amount of thermal background observations before and after, yielding an on-target observation efficiency of 93%. We detect all three HR 8799 planets in the field of view (HR 8799 c,d,e), becoming the first planets characterized with a vAPP coronagraph. In addition, we retrieve low-resolution spectra for all three planets between 3.5-4.1 μm, which are completely featureless and consistent with constant flux. The spectra of HR 8799 c, e are consistent with all previously reported photometric points, while HR 8799 d shows an unexplained increase in flux of 60%. However, all three spectra are consistent with previously reported Brα photometry (4.05 μm) of these planets. This is specifically interesting because most atmospheric models that use uniform clouds do not reproduce the flux at this wavelength, while rudimentary models of patchy clouds are more successful. Not only does this ALES-vAPP combination open a new spectroscopic window for exoplanets, it is also capable of time-variability studies enabled by the high observation efficiency and the presence of the stellar PSF. Based on the presented HR 8799 observations, the sensitivity would be $<\sim$ 5% per 1.5 hours of observing for all three planets, limited by the thermal background.

Chapter 4: Minimizing the effects of polarization leakage for liquid-crystal coronagraphs with multi-grating pattern combinations.

Currently, most ground-based coronagraphs have optimal performance in a single spectral band, while multi-band observations with integral field spectrographs (IFSs) are key to efficiently characterize exoplanets. In addition, the vector vortex coronagraph (VVC) is baselined for the HabEx mission, however, it does not

reach the contrast requirement for the 20% bandwidth without extreme filtering of polarization leakage. In this chapter, we demonstrate that multi-grating pattern combinations can be used for liquid-crystal coronagraphs to suppress the polarization leakage by multiple orders of magnitude over a large bandwidth. This concept is different to other leakage minimization strategies because we use the patterning of separate liquid-crystal optics placed in succession to suppress leakage, rather than optimizing the retardance of the liquid-crystal layer itself. Using two polarization gratings composed of a single-layer liquid crystal structure in the lab, we demonstrate a leakage suppression of more than an order of magnitude over a bandwidth of 133 nm centered around 532 nm, compared to a single liquid-crystal optic. We also demonstrate a leakage suppression of three orders of magnitude at the center wavelength. Moreover, we apply the double-grating concept successfully to a VVC in the lab and a vAPP with an annular dark zone that is currently installed in LBT/LMIRcam. Simulations indicate that a combination of two multi-layered liquid-crystal PGs can suppress leakage to $< 10^{-5}$ for 1-2.5 μm. Such a double-grating VVC or vAPP would have optimal coronagraphic performance for the full bandwidth of a ground-based IFS, e.g. Subaru/CHARIS. In addition, we present simulations of a triple-grating solution that has $< 10^{-10}$ leakage on the first Airy ring from 450 to 800 nm for space-based applications. This triple-grating VVC would significantly relax the requirements of polarization leakage filtering, and can have a major impact on the system-level design and trade-offs, including the option to add spectroscopy and polarimetry.

Chapter 5&6: Multiplexed Holographic Aperture Masking with liquid-crystal geometric phase masks.

Sparse aperture masking (SAM) is an important technique to study processes like planet formation at solar-system scales. SAM is capable of characterizing structures down to half the diffraction limit and it produces observables that are resistant against wavefront aberrations, e.g. closure phases. However, it comes at the cost of a reduction in throughput, typically by 80-90%. In these two chapters we present holographic aperture masking (HAM), which aims at increasing the throughput of SAM by selectively combining all subapertures across a telescope pupil in multiple interferograms using a liquid-crystal phase mask, in addition to adding low-resolution spectroscopic capabilities. In Chapter 5 we show the principle of multiplexing, i.e. combining multiple phase ramps into a single phase pattern, in addition to designs and lab results of a technology demonstrator. In Chapter 6 we further develop the design with a complementary central (SAM) component and report on the design, construction and commissioning of a HAM prototype at the Keck OSIRIS Imager. We create a pipeline that extracts monochromatic closure phases from the central component as well as multi-wavelength closure phases from the holographic component. We test the performance of the HAM mask in the lab and on-sky, where we observed the binary star HDS 1507. We retrieve a constant separation of 120.9 $\pm$0.5 mas for the independent wavelength bins, which is excellent agreement with literature values.

Combining the low-resolution spectra from observations in J, H, and K-band enables spectral characterization through spectral slopes which help distinguishing between scattered light and thermal signals.

Chapter 7: Simultaneous phase and amplitude aberration sensing with a liquid-crystal vector-Zernike phase mask.

Imaging Earth-like exoplanets in reflected light ($\sim 10^{-10}$ contrast), requires extremely sensitive, photon efficient phase and amplitude control with high spatial-frequency resolution. Similarly, ground-based instruments also require sensitive measurements of the wavefront to combat non-common path aberrations, which are currently limiting the achieved contrast at the smallest separations. The Zernike wavefront sensor (ZWFS) has these properties and therefore has recently been selected for the HabEx and LUVOIR missions. In this chapter we present an enhanced version of the ZWFS, the "vector-Zernike" wavefront sensor (vZWFS), which consists of a patterned liquid-crystal phase mask and optics to split circular polarization. The vZWFS extends the capabilities of the ZWFS, adding the capability to measure phase *and* amplitude aberrations *simultaneously*. We demonstrate reconstruction of the complex wavefront with monochromatic lab measurements and show in simulation the high accuracy and sensitivity over a bandwidth up to 100%, twice the bandwidth of a ZWFS. The advantages for future exoplanet missions are clear. Because the vZWFS measures the complex wavefront, it does not require amplitude calibration measurements, resulting in a more accurate reconstruction and a higher duty cycle. Moreover, the sensitivity over a large bandwidth increases the number of photons and consequently the reconstruction accuracy.

Chapter 8: Multi-color holography with a two-stage patterned liquid-crystal element.

The gvAPPs that are installed in many different instruments have two properties that are suboptimal for exoplanet imaging. Their grating is necessary to separate the left- and right-circular polarization states from each other and the leakage term, yet the dispersion smears planet light over many pixels, introducing more speckle and read noise. Secondly, the separation of circular polarization states directs half of the light in the bright-side of vAPP PSF, reducing the effective throughput by a factor of 2. More generally, the computer-generated geometric-phase holograms have wavelength-dependent output and cannot be used for the production of color imagery. In this chapter, we introduce a two-stage patterned liquid-crystal geometric phase hologram capable of producing white-light holograms. We implement a two-stage approach that first uses the wavelength-dependent diffraction to separate colors and circular polarization states, and second, directs these colors through separate holographic patterns. Moreover, by utilizing the geometric phase, we obtain diffraction efficiencies close to 100%. We successfully create a white light hologram from RGB input in the lab. In addition, we show in simu-

lation how this two-stage element could be used for of true-color holograms. The two-stage GPH is directly portable for truly broadband vAPPs, free of polarization leakage and can be designed to have inner working angle that changes inversely wavelength. Likewise, the two-stage GPH provides direct control of both circular polarization states such that the D-shaped dark zone can be created on the same side.

1.6 Future Outlook

This thesis focuses on applications of liquid-crystal technology in high-contrast imaging (HCI) systems to implement any desired phase pattern in an achromatic and polarization-controlled fashion these systems. The fine resolution, achromatic nature and excellent efficiency of liquid-crystal geometric phase plates provides many opportunities to integrate multiple subsystems of HCI systems. As most applications have been developed during the last few years and have only been installed recently, they have not been used to their full potential.

1.6.1 A ground-based perspective

Ground-based telescopes have been the drivers of high-contrast imaging due to their large primary mirrors and the possibility to upgrade instruments. With the absence of competitive space-based high-contrast imaging systems until the launch of the Nancy Grace Roman Space Telescope ($\sim$2025) (Roman, 2019), ground-based HCI is likely to continue to dominate the scientific discoveries from direct imaging the coming years. Large surveys from SPHERE and GPI are mostly completed, and have discovered less exoplanets than earlier models had predicted. Therefore, most efforts related to ground-based HCI are currently oriented towards developing and implementing new technology to improve the performance of existing HCI systems, i.e. achieving a better contrast at small inner working angles. Moreover, the lessons learned are implemented in first-generation instruments of the extremely large telescopes, which will open up a whole new parameter space in terms of IWA once these instruments are completed. Here we will describe how the work in this thesis will impact future observations and instruments.

Impact of broadband coronagraphs

Studies on coronagraph design have shown that phase-only masks have a better performance than amplitude masks (Mawet et al., 2009a; Por, 2017, 2020), and liquid crystals can be used to create arbitrary geometric phase patterns with high efficiencies. In addition, liquid-crystal optics can be tuned to operate from the visible to the infrared (0.55-5 μm), with a possible extension to from the UV to mid-infrared ($\sim$ 300 nm - 40 μm) (Packham et al., 2010). Applying this diversity to the vAPP in combination with the instantaneous wavelength coverage of more than an octave has an impact on the capabilities of multiple instruments. This is most relevant for instruments with an integral field spectrograph, observing with

the full bandwidth instantaneously. A large advantage of this observing mode is that the coronagraphic performance is stable over the full wavelength range. In addition, spectral differential methods will benefit from the increased bandwidth. The first attempts of observing with a low-resolution (R=30) spectroscopic mode, using the full bandwidth of an integral field spectrograph, have been successful. Observations of HD 91312 with Subaru/SCExAO showed a constant coronagraphic performance for the full wavelength coverage of the CHARIS IFS. Moreover, the vAPPs installed in LBT/LMIRcam have provided the first mid-infrared spectrum of HR 8799 planets, see Chapter 3. While the spectrum is flat for gas giants with patchy clouds and a temperature above 1300K, cooler planets are expected to have transitioned to cloud free, CH_4-rich T-type atmospheres. Observations of these cooler planets, like 51 Eridani b, using the ALES spectrograph and the double-grating vAPP can constrain CH_4 content from the spectral slope between 3.5-4.1 μm. In addition, the contrast ratio for even cooler planets ($\sim$ 500 K) is most favorable in L or M (3-5 μm) band, providing good hunting grounds for LMIRcam imaging using the vAPP.

Another instrument that will benefit from the vAPP is the Magellan/MagAO-X system (Males et al., 2018), which provides an excellent AO performance in the visible. Compared to SPHERE/ZIMPOL, the AO system of MagAO-X will operate much faster (>3 kHz) and with more actuators on the deformable mirror (2K), resulting in a Strehl of 70% at Hα. Residual atmospheric tip-tilt aberrations do not result in a degradation of contrast because the vAPP is a pupil-plane coronagraph. Moreover, with the implementation of focal-plane wavefront sensing it is expected that the vAPP coronagraph will deliver unprecedented raw contrast ($< 10^{-3}$) at separation down to 50 mas. Accreting exoplanets like PDS 70 b and c will be prime targets for the Hα imaging mode of MagAO-X.

The VLT/ERIS instrument (Amico et al., 2012) will also become operational in the coming years, re-opening the mid-infrared window for the VLT after VLT/NACO was decommissioned. The imaging part of ERIS, NIX (Pearson et al., 2016), is equipped with a vAPP coronagraph (Boehle et al., 2018) that has been improved with respect the Magellan/MagAO vAPP (Otten et al., 2017) that is also operating in the Southern hemisphere. The ERIS instrument has two features that will help with the detection and characterization of new exoplanets. ERIS has a long-slit spectroscopic mode with a spectral resolution of $\sim$ 450 in L-band for non-coronagraphic imaging and the instrument has a visible wavefront sensor that can be fed by a laser guide star (Amico et al., 2012). Faint targets that were inaccessible with NACO can now be studied in great detail.

Components of the VLT/SPHERE instrument (Beuzit et al., 2019) will also be gradually upgraded in the coming years (Boccaletti et al., 2020), both for enhanced science as well as a roadmap towards a future dedicated high-contrast imaging system on the European Extremely Large Telescope (ELT), ELT/EPICS (Kasper et al., 2010). Equipped with a second-stage AO-system operating in the infrared and a medium-resolution integral field spectrograph, SPHERE will hunt for new exoplanets around fainter (lower mass) stars and characterize exoplanets in more detail. In addition, SPHERE will be able to perform high-resolution spectroscopy

of gas giants with the HiRISE project (Vigan et al., 2018b), coupling SPHERE and CRIRES+ with a fiber. Both spectroscopic modes will benefit from an upgrade of the coronagraph under good observing conditions. While the liquid-crystal recipes for 100% bandwidth have only been applied for the vAPP, the vector vortex coronagraph can also benefit from the extended spectral bandwidth. A double-grating VVC of charge 2 with a pupil apodizer would provide excellent performance for the full SPHERE bandwidth ($0.9-2.5$ μm), with a theoretical inner working angle that is three times smaller than the most-used coronagraph.
Another coronagraph that is considered for SPHERE is an upgraded version of the current Apodized-Pupil Lyot Coronagraph (APLC) (Carbillet et al., 2011). A by-product of the creation of the vAPP optimizer was the proof that phase apodization always result in solutions with higher Strehl than amplitude apodization. Similarly, changing the pre-apodizer of the APLC to a phase apodizer, resulted in an improved performance for D-shaped dark zones. The Phase-Apodized-Pupil Lyot Coronagraph (PAPLC) with a knife-edge focal-plane mask yields inner working angles as close as 1.0 λ/D at contrasts of 10^{-10} with post-coronagraphic throughput up to 75% (Por, 2020). This coronagraph can be adapted to any pupil shape, which makes it ideal for segmented telescopes or telescopes with a non-zero central obscuration. A liquid-crystal phase plate implementation would extend the bandwidth of the PAPLC significantly, although this would require splitting of the opposite circular polarization states, similar to a gvAPP. The disadvantages are equivalent to a gvAPP, e.g. spectral smearing, D-shaped dark zone and losing factor two in planet throughput, see Chapter 2. Even so, the PAPLC had the best performance of all considered coronagraphs in simulations of a future SPHERE instrument and is therefore a prime candidate to replace the coronagraph (Boccaletti et al., 2020). Operating it in SPHERE will also be a first step to test these kinds of new technologies for ELT/EPICS.
With the five times larger primary mirror of ELT, it will be possible to detect fainter planets much closer in, down to solar system scales (<10 AU) for the nearest star forming regions. A first-generation ELT instrument is the AO assisted integral field spectrograph ELT/HARMONI (Thatte et al., 2016), which will provide R$\sim 500-20000$ resolution spectra for between 32000-8000 spaxels between 0.47-2.45 μm. Similar to VLT/MUSE and VLT/SINFONI, ELT/HARMONI can be used for direct imaging of exoplanets using the molecule-mapping technique. The other two first-generation instruments, ELT/METIS (Brandl et al., 2016) and ELT/MICADO (Davies et al., 2016), are capable of high-contrast imaging, have two and one vAPPs baselined respectively, and also have large bandwidths and spectroscopic capabilities (Kenworthy et al., 2018). Long-slit spectroscopy will provide modest spectral resolution for METIS (R~ 1500) in L and M band (3-5 μm), and good spectral resolution for MICADO (R $\sim 20,000$) in J, H, and K band (1-2.5 μm). METIS is also equipped with a high-resolution (R$\sim 100,000$) integral field spectrograph. The broadband capabilities of the vAPP are a good match with these modes, and the design of the vAPPs can easily be adapted to deliver good contrasts at small inner working angles despite the thick spiders and large central obscuration. Moreover, during the first years of operation both

instruments and the telescope will face unexpected difficulties that limit their performance. Especially vibrations will provide a challenge as the imaging resolution will increase a factor 5, while the mechanical structures have become much larger. Another challenge is the atmospheric dispersion, which will significantly impact MICADO where the focal-plane masks are in front of the atmospheric dispersion corrector (ADC). Because the coronagraphic performance of vAPPs are not sensitive to tip-tilt errors, the vAPP in MICADO will likely have an edge compared to other focal-plane coronagraphs. METIS has tip-tilt correction and an ADC, but residual tip-tilt and pointing drift will still significantly impact the performance (Carlomagno et al., 2020).
With the increased resolution, improved sensitivity and spectrographic capabilities of the first-generation of ELT instruments, it is expected that many new exoplanets will be directly imaged and characterized. The limitation of these instruments will be the performance of the adaptive-optics systems, which will not be good enough to directly image rocky exoplanets. ELT/EPICS, the second-generation instrument dedicated to high-contrast imaging, plans to capitalize on improved technology to open up this new parameter space (Kasper et al., 2010). The design of this instrument will depend on the lessons learned of technology demonstration projects like the SPHERE upgrades and the SCExAO and MagAO-X instruments. The instrument will have to be fully optimized on a system level, with hybrid coronagraphs, integrated focal-plane wavefront sensing, high-resolution spectroscopy, and polarimetry. Once completed, EPICS is expected to deliver 10^{-7}-10^{-8} post-processed contrast thanks to spectroscopy and/or polarimetry, which enables the detection and characterization of rocky planets in the habitable zone around the nearest M stars. EPICS will mainly observe planets previously found by indirect methods, characterizing the planet population by their atmospheric properties. Moreover, EPICS will have the spatial resolution necessary for imaging structures of a few AU at the nearest star-forming regions. EPICS will therefore be able to resolve snow lines in these regions and provide insight in their role in planet formation. In addition, EPICS enables the determination of the mass-luminosity relation for young gas giants, which also provides a rigorous test for planet formation models. Together with high-contrast imaging systems of the other ELTs, EPICS will provide fundamental insight in planet formation, disk-planet interaction and perhaps even habitability.

Benefits of a system approach

The freedom of the direct-write and multi-twist retarder technology for liquid-crystals also have enabled the integration between different subsystems in a HCI instrument. A first example is the possibility to add multiple PSF copies to the vAPP PSF, enabled only by the fine writing resolution. These holographic PSFs can be used as reference for the flux of companion PSFs, in addition to the coronagraphic PSF if it is not saturated, equivalent to Marois et al. (2006b) Flux variations that do not originate from companion variability, like a changing AO-performance with airmass, can be divided out. So far, the limits of this technique

have not been investigated, but variations on the level of 1% could reveal the presence of clouds (Apai et al., 2013). These causes for companion variability are strongest in the near-infrared, and in the visible variability in H_α emission could provide insight in accretion mechanisms. At these wavelengths the AO-performance is more variable and using a vAPP is more beneficial.
Second, writing detailed patterns also enables holographic aperture masking (HAM), which provides spectrally-resolved closure phases and squared visibilities. Spectrally-resolved closure phases present a new observable for aperture masks used in imaging mode, and can be used to spectrally characterize targets down to 0.5 λ/D. Simultaneous spectral and spatial measurements can help constrain infrared spectral slopes, similar to the wavelength-dependent brightness ratio that is recovered for the binary HD 90823 in Chapter 6. The sensitivity that will be achieved by HAM is very dependent on the calibration of the closure phases and the polarization leakage. Currently, HAM in OSIRIS can determine wavelength-dependent brightness ratios with spectral resolutions of R=$6-15$ in J, H, and K band for binaries with contrast rations $\sim 10^{-1}$. With a future upgrade of the HAM optic, adding a double-grating, and an improved calibration strategy it will be possible to provide low-resolution spectra for objects higher contrasts ($< 10^{-2}$) down to $\sim$ 30 mas in H-band. For the ELTs the relative throughput of sparse aperture masks will be lower compared to an 8-m class telescope. The data reduction of SAM requires that the wavefront variations inside the individual holes should be minimal, resulting in a maximum hole size that depends on the coherence length of the atmosphere, not the mirror diameter. Segment gaps and phase steps from segmented telescopes are also a problem, and therefore holes in the aperture mask will be limited to a size on the order of a segment diameter. Moreover, many smaller non-redundant holes have a lower throughput than fewer larger ones. So, HAM would be an interesting way of improving the throughput. Unfortunately, the increase in the PSF size due to the smaller holes also reduces the spectral resolution, such that HAM on the ELT will likely be unable to spectrally resolve closure phases.
Combining high-contrast imaging systems with low-resolution spectroscopy has resulted in new insights in the atmospheres of gas giants, yet the gain in contrast by using the additional spectral information has been limited. This is different for medium- and high-resolution spectroscopy, where the raw-contrast of non-coronagraphic systems has been enhanced up to a factor 1000. The spectral differences between a star and its companion can be used to filter speckles and simultaneously characterize planets. With cross-correlation techniques of spectral templates it is possible to detect molecular features in the atmospheres of exoplanets (Snellen et al., 2010, 2015; Hoeijmakers et al., 2018a). Noise from the stellar PSF is limiting these retrievals, which suggests that using high-contrast imaging techniques could be beneficial. Adding a fiber in the focal plane to connect high-contrast imaging systems to high-resolution spectrographs is relatively straightforward, but it has an inner working angle that is limited by conventional coronagraph designs ($> 2\lambda/D$). A third integrated method resulted in two new concepts that have been developed with smaller inner working angles ($< 1\lambda/D$).

This integrated method works by creating a new type of coronagraph that uses the mode-filtering property of the single-mode fibers to suppress the starlight, i.e. Single-mode Complex Amplitude Refinement (SCAR) (Por & Haffert, 2020; Haffert et al., 2020) and Vortex Fiber Nulling (VFN) (Ruane et al., 2018). Both concepts change the shape the PSF with liquid-crystal phase masks in the pupil plane, such that locally the starlight has one or multiple nulls at the location of the fibers. These nulls do not contain the fundamental mode of the fibers, and high rejection ratios of 10^4 are achieved. SCAR is a direct descendant of the vAPP, optimizing a phase plate for different focal-plane constraints given by the fiber. The writing precision and the achromatic phase resulted in deep nulls over a 20% bandwidth, limited by the PSF scaling with wavelength. In addition, the first test version already benefited from the polarization grating to separate the leakage and the grating mask to create a pupil, two concepts developed for the vAPP. Both SCAR and the VFN would greatly benefit from the double-grating method to reduce the influence of polarization leakage. A double-grating version would remove the need for polarization filtering, increasing the efficiency of both by a factor of 2. The combination of these extreme inner working angles and high-spectral resolution will likely result in the detection and characterization of many planets, from rocky worlds to gas giants.

1.6.2 A space-based perspective

A long-term goal of high-contrast imaging is the detection and characterization of Earth-like exoplanets around solar-type stars. It is expected that ELTs will have the resolution and contrast to image rocky planets in the habitable zone around M-stars (Snellen et al., 2019). However, the extreme contrasts of such Earth-like planets in reflected light around solar-type stars, i.e. 10^{-10}, requires stabilization of the wavefront that can only achieved with space-based observatories. The technology demonstration for space-based systems have reached these contrast levels in lab conditions for small bandwidths, limited by chromatic bandwidths. It will take until the mid 2020s before the first technology demonstration mission is launched, i.e. the Nancy Grace Roman Space Telescope[6], and likely a decade thereafter before the current state of the art technology is operating in space.
As of yet, only the Hubble Space Telescope has some coronagraphic capabilities, although they are not competitive with ground-based results. Similarly, the James Webb Space Telescope will likely not outperform ground-based HCI systems in the visible and near-infrared, having an estimated instrumental wavefront error of 131 nm (Stahl, 2007). The performance of both NIRCam and MIRI will be limited by this wavefront aberration (Beichman et al., 2010; Boccaletti et al., 2015). It is likely that the post-processed performances will improve over time, as the extreme stability of space and low thermal background will help with improving the data reduction like with HST. We note that the stability is also beneficial for the non-redundant mask of NIRISS (2.7-5 mum), predicted to have a contrast of 10^{-5} between 60 and 500 mas (Sivaramakrishnan et al., 2010). While the contrast

[6]The Nancy Grace Roman Space Telescope was previously known as WFIRST

levels will be similar to ground-based, the unique wavelength coverage will provide new insights on disks through direct imaging and spectroscopy and close-in planets through transmission and secondary eclipse spectroscopy. Moreover, while the Roman Space Telescope will likely reach extreme contrasts of 10^{-8} and do amazing science, it is only a technology demonstrator for high-contrast imaging (Mennesson et al., 2020). These two aforementioned telescopes are part of the next-generation space telescopes, and are the only two with high-contrast imaging capabilities. Therefore, directly imaging Earth analogues will be out of reach for the coming decade. To fill this gap and to take advantage of the recent technology developments, two competing space missions have been proposed that have a large focus on the direct characterization of rocky Earth-like planets. These missions are HabEx (HabEx Study Team, 2019), a 4-m off-axis telescope and LUVOIR (LUVOIR Team et al., 2019), an 15-m (A) or 8-m (B) segmented telescope, both with a predicted launch date in 2035.

In preparation for these missions, the focus of space-based high-contrast imaging for the coming years is technology development and demonstration. Two key aspects that will ultimately determine the performance of these systems are wavefront sensing and coronagraphy. The challenges are to correct the wavefront with extreme precision, keeping it stable afterwards, and to manufacture coronagraphs that provide 10^{-10} contrast over a 20% bandwidth. The wavefront stability that is necessary for 10^{-10} contrast is 10 picometer RMS, putting stringent requirements on the wavefront sensor (Steeves et al., 2020). The wavefront sensor that has been selected for both HabEx and LUVOIR is a Zernike wavefront sensor (ZWFS), which is extremely sensitive, photon efficient, and has high spatial-frequency resolution. The geometric-phase version of the ZWFS, the vector-Zernike wavefront sensor (vZWFS), has several advantages compared of the ZWFS. The vZWFS is capable of reconstructing both phase and amplitude, does not require an amplitude calibration measurement and provides good reconstruction for larger bandwidths. These advantages would benefit the HabEx and LUVOIR missions and the vZWFS is therefore now being tested at the Jet Propulsion Laboratory, where they are investigating both a liquid-crystal version and a meta-material version. Moreover, the vZWFS is under study for implementation at the W. M. Keck Observatory.

The proposed coronagraphic systems of both HabEx and LUVOIR-B will make use of a vector-vortex coronagraph, where LUVOIR A and B will also be equipped with an apodized pupil Lyot coronagraph and a phase-induced amplitude apodization complex mask coronagraph. Liquid-crystal vector-vortex coronagraphs, however, do not reach the contrast requirement of $< 10^{-10}$ without extreme filtering of polarization leakage. Even with filtering, multiple VVCs will have to be installed to cover the full wavelength range of the coronagraphic instruments. A triple-grating vector-vortex coronagraph (tgVVC) can be manufactured to suppress the on-axis leakage by a factor of 10^{-8} over 100% bandwidth, resulting in $< 10^{-10}$ contrast at the inner working angle of a charge 6 vortex. For smaller bandwidths of $\sim 20\%$ a double-grating VVC will also provide enough leakage suppression. A charge 6 tgVVC with 1TR layers has been manufactured and will be tested in the near future in the test benches of the Jet Propulsion Laboratory. If these tests show

that the tgVVC works as in simulation, a 100% bandwidth version can have a major impact on the system-level design. HabEx is now baselined with two arms in the visible, each with 40% bandwidth in two separate channels in the visible, spanning 0.45-0.67 μm (Blue channel) and 0.67-1.0 μm (Red channel). Another option that is considered uses a polarizing beam splitter to separate the two arms, enabling polarimetry. Both channels have a separate coronagraph, optimized for their respective wavelengths. A tgVVC would enable the creation of two identical arms which could be equipped with both blue and red filters, removing single-point failures. Moreover, a larger instantaneous bandwidth up to 50% is ideal for spectroscopy using single-mode fibers (Coker et al., 2019b,a). Lastly, the deformable mirrors of the two identical arms could be used to create complementary dark-hole shapes. So, a broadband coronagraph could significantly impact the instrument design.
The LUVOIR-B aperture with a central obscuration does not work well with the vortex coronagraph without pre-apodization. Another solution is the PAPLC, which was demonstrated to have the highest Earth-like exoplanet yield of all coronagraphs considered for LUVOIR-B. For small spectral bandwidths the weak phase apodization can easily be offloaded to the deformable mirrors of high-contrast imaging systems. Larger spectral bandwidths would require a liquid-crystal phase mask, like for SPHERE. Independent of the final implementation, the PAPLC and the tgVVC6 are prime candidates for next-generation space-based high-contrast imaging systems.

1.6.3 New applications of liquid-crystal technology for exoplanet detection

Besides the applications of liquid-crystal technology presented in this thesis and mentioned above, other new ideas or applications have been developed by our group and described in literature, some of which have been directly inspired by the work in this thesis. Here we will summarize these applications and discuss their impact.
The first application is the vector-speckle grid (Bos, 2020), which is closely related to the coronagraphic holograms used for photometric monitoring. The vector-speckle grid creates off-axis PSF copies that have both circular polarization states, rather than one as for the vAPP. Because the opposite polarization states have the opposite phase or amplitude modulation, they are instantaneously incoherent with the underlying halo and thus more stable. With liquid-crystal technology a vector-speckle grid is easily manufactured with high efficiency over large bandwidths, making them also ideal for spectroscopy.
Besides the vZWFS, liquid-crystal technology been used for improving upon other existing wavefront sensing methods. The liquid-crystal generalized optical differentiation wavefront sensor (Haffert, 2016) is more efficient than the other optical differentiation wavefront sensors, and it is possible to change the focal-plane amplitude modulation. A step function makes the wavefront sensor akin a pyramid

wavefront sensor with extremely sharp edges (< 2 μm) for a lower price. Another type of wavefront sensor is the three-wave lateral shearing interferometer (TWLSI), which has extreme dynamic range and exquisite sensitivity. Commercial implementations of the TWLSI have low throughput due to inefficient gratings, which can be solved with liquid-crystal TWLSI. Combined with the wavelength ranges that liquid-crystal technology can provide, the TWLSI would prove to be an easy and efficient method for quality control of optical elements, or even serve as primary wavefront sensor in adaptive optics systems.

A completely different application of liquid-crystal technology has to do with precise astrometry for exoplanet detection. Detecting exoplanets with astrometry requires micro-arcsecond stability of an instrument. Effects like mechanical drifts or creep of optical surfaces create astrometric signals that overwhelm the signal of exoplanets. For instruments like ELT/MICADO, the instrument design is optimized to reduce any of these effects, increasing the cost and complexity of the instrument. The TOLIMAN space telescope (Bendek et al., 2018; Tuthill et al., 2018) has a radically different approach, using diffraction structures as astrometric ruler for differential astrometry of Alpha Centauri. Diffraction structures are interesting rulers as they are distorted the same way the objects are distorted. So rather than maximizing instrument stability, TOLIMAN creates a ruler that is immune to many of the errors that would drown the signal. Creating diffraction structures that form an astrometric grid is an optimization problem similar to the vAPP optimization, albeit with a less clear merit function. The grid is created using 0-π phase structures, and the pattern is optimized to create a regular grid of speckles with a homogeneous intensity. By using a liquid-crystal phase plate, it is also possible to add photometric reference holograms at the corners of the detector. These reference holograms create a small spectrum for spectral monitoring (R $\sim$ 500). A technology demonstrator mission called TinyTol employs such a liquid-crystal TOLIMAN optic and will be launched in a cubesat early 2021. The TOLIMAN project aims to launch a 30 cm telescope in the coming decade for the astrometric monitoring of Alpha Centauri A and B to look for Earth-like exoplanets in the habitable zone, which requires sub-microarcsecond astrometric metrology.

Finding an Earth-like exoplanet, or even finding biomarkers on these planets, does not equal finding life. Many proposed biomarkers can be generated by inorganic pathways, resulting in false positives when not carefully studied. Remote sensing of the only place known to harbour life, Earth, presents an opportunity to study what signals could present unbiased biomarkers. Instruments that observe an unresolved or partially resolved Earth will provide information that is similar to what future instruments will record for Earth-like exoplanets and will provide valuable benchmark spectra for radiative transfer codes used to model exoplanet spectra. Spectro-polarimetric measurements of Earth at different scattering angles can be used to infer the presence of liquid water oceans, clouds, and derive atmospheric properties. Moreover, they can give key insights in biomarkers, like O_2, the Chlorophyll Green Bump and the Vegetation Red Edge, and continuous monitoring can be used to detect continents and their specific features, like rain forests,

deserts and ice caps (Stam, D. M., 2008; Klindžić et al., 2020). Perhaps even more interesting is spectro-polarimetry of circularly polarized light, which is likely an indicator of homo-chirality of life (Patty et al., 2019). Such spectro-polarimetric instruments are necessarily space-based, and could be placed on satellites or even on other celestial bodies. Without the need to resolve features using large optics, these instruments can be made very compact and without any moving parts. Two instruments in particular, LOUPE (Karalidi et al., 2012; Klindžić et al., 2020) and LSDpol (Snik et al., 2019), provide large spectral ranges and precision polarimetry in small volumes. One of the enabling technologies is the patterned liquid-crystal half-wave retarder, that is used to create the passive polarimetric modulation used for full-Stokes measurements with minimal diffraction. The work in this thesis on liquid-crystal recipes enables high broadband polarimetric efficiency for these instruments. It is likely that LOUPE will be installed on a Lunar lander, rover or orbiting mission, while LSDpol aims for installation on the ISS. We predict that their measurements are key to further our understanding the requirements for a remote detection of life.

Bibliography

Allard, F. 2013, Proceedings of the International Astronomical Union, 8, 271

Alonso, M. A., & Dennis, M. R. 2017, Optica, 4, 476

Amico, P., Marchetti, E., Pedichini, F., et al. 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, Vol. 8446, International Society for Optics and Photonics, 844620

Apai, D., Radigan, J., Buenzli, E., et al. 2013, The Astrophysical Journal, 768, 121

Apai, D., Karalidi, T., Marley, M., et al. 2017, Science, 357, 683

Avenhaus, H., Quanz, S. P., Garufi, A., et al. 2018, The Astrophysical Journal, 863, 44

Balasubramanian, K., White, V. E., Yee, K. Y., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011005

Baraffe, I., Chabrier, G., Barman, T. S., Allard, F., & Hauschildt, P. 2003, Astronomy & Astrophysics, 402, 701

Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, Astronomy & Astrophysics, 577, A42

Barman, T. S., Konopacky, Q. M., Macintosh, B., & Marois, C. 2015, The Astrophysical Journal, 804, 61

Barman, T. S., Macintosh, B., Konopacky, Q. M., & Marois, C. 2011, The Astrophysical Journal, 733, 65

Baselmans, J. 2012, Journal of Low Temperature Physics, 167, 292

Beichman, C. A., Krist, J., Trauger, J. T., et al. 2010, Publications of the Astronomical Society of the Pacific, 122, 162

Bendek, E., Tuthill, P., Guyon, O., et al. 2018, in Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, Vol. 10698, International Society for Optics and Photonics, 106980G

Benz, W., Ida, S., Alibert, Y., Lin, D., & Mordasini, C. 2014, arXiv preprint arXiv:1402.7086

Berry, M. 1988, Scientific American, 259, 46

Berry, M. V. 1984, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 392, 45
Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, arXiv preprint arXiv:1902.04080
Biller, B., Apai, D., Bonnefoy, M., et al. 2021, Monthly Notices of the Royal Astronomical Society
Bliokh, K. Y., & Aiello, A. 2013, Journal of Optics, 15, 014001
Bliokh, K. Y., Rodríguez-Fortuño, F. J., Nori, F., & Zayats, A. V. 2015, Nature Photonics, 9, 796
Bloch, F., Hansen, W., & Packard, M. 1946, Phys Rev, 70, 460
Boccaletti, A., Lagage, P.-O., Baudoz, P., et al. 2015, Publications of the Astronomical Society of the Pacific, 127, 633
Boccaletti, A., Chauvin, G., Mouillet, D., et al. 2020, arXiv preprint arXiv:2003.05714
Boehle, A., Glauser, A. M., Kenworthy, M. A., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 107023Y
Bonnefoy, M., Zurlo, A., Baudino, J.-L., et al. 2016, Astronomy & Astrophysics, 587, A58
Born, M., & Wolf, E. 2013, Principles of optics: electromagnetic theory of propagation, interference and diffraction of light (Elsevier)
Bos, S. P. 2020, arXiv preprint arXiv:2005.08751
Bos, S. P., Doelman, D. S., Lozi, J., et al. 2019, A&A, 632, A48, doi: 10.1051/0004-6361/201936062
Bowler, B. P., Liu, M. C., Dupuy, T. J., & Cushing, M. C. 2010, The Astrophysical Journal, 723, 850
Brandl, B. R., Agócs, T., Aitink-Kroes, G., et al. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 990820
Burgasser, A. J. 2014, arXiv preprint arXiv:1406.4887
Cantalloube, F., Por, E., Dohlen, K., et al. 2018, Astronomy & Astrophysics, 620, L10
Cantalloube, F., Farley, O., Milli, J., et al. 2020, Astronomy and Astrophysics-A&A
Carbillet, M., Bendjoya, P., Abe, L., et al. 2011, Experimental Astronomy, 30, 39
Carlomagno, B., Absil, O., Kenworthy, M., et al. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 990973
Carlomagno, B., Delacroix, C., Absil, O., et al. 2020, Journal of Astronomical Telescopes, Instruments, and Systems, 6, 035005
Carlotti, A., Mouillet, D., Correia, J.-J., et al. 2018, in Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III, Vol. 10706, International Society for Optics and Photonics, 107062M
Catalán, E. V., Huby, E., Forsberg, P., et al. 2016, Astronomy & Astrophysics, 595, A127
Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45, doi: 10.1086/312457
Charnay, B., Bézard, B., Baudino, J.-L., et al. 2018, The Astrophysical Journal, 854, 172
Cheetham, A. C., Girard, J., Lacour, S., et al. 2016, in Optical and Infrared Interferometry and Imaging V, Vol. 9907, International Society for Optics and Photonics, 99072T
Chen, W. T., Török, P., Foreman, M. R., et al. 2016, Nanotechnology, 27, 224002
Chigrinov, V. G., Kozenkov, V. M., & Kwok, H.-S. 2008, Photoalignment of liquid crystalline materials: physics and applications, Vol. 17 (John Wiley & Sons)
Chilcote, J. K., Larkin, J. E., Maire, J., et al. 2012, in Ground-based and Airborne

Instrumentation for Astronomy IV, Vol. 8446, International Society for Optics and Photonics, 84468W

Chilcote, J. K., Bailey, V. P., De Rosa, R., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 1070244

Claudi, R., Maire, A.-L., Mesa, D., et al. 2019, Astronomy & Astrophysics, 622, A96

Claudi, R. U., Turatto, M., Gratton, R. G., et al. 2008, in Ground-based and Airborne Instrumentation for Astronomy II, Vol. 7014, International Society for Optics and Photonics, 70143E

Codona, J. L., Kenworthy, M. A., Hinz, P. M., Angel, J. R. P., & Woolf, N. J. 2006, 6269, 62691N, doi: 10.1117/12.672727

Coker, C. T., Sayson, J. L., Shaklan, S., et al. 2019a, in Techniques and Instrumentation for Detection of Exoplanets IX, Vol. 11117, International Society for Optics and Photonics, 111171A

Coker, C. T., Shaklan, S. B., Riggs, A. E., & Ruane, G. 2019b, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 045003

Courtial, J. 1999, Optics communications, 171, 179

Currie, T., Burrows, A., Girard, J. H., et al. 2014, The Astrophysical Journal, 795, 133

Davies, R., Schubert, J., Hartl, M., et al. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 99081Z

De Boer, J., Langlois, M., Van Holstein, R. G., et al. 2020, Astronomy & Astrophysics, 633, A63

de Boer, J., Salter, G., Benisty, M., et al. 2016, A&A, 595, A114, doi: 10.1051/0004-6361/201629267

De Kok, R., Stam, D., & Karalidi, T. 2011, The Astrophysical Journal, 741, 59

dit de la Roche, D. P., Hoeijmakers, H., & Snellen, I. 2018, Astronomy & Astrophysics, 616, A146

Escuti, M. J., Kim, J., & Kudenov, M. W. 2016, Optics & Photonics News, 27, 22

Fried, D. L. 1966, J. Opt. Soc. Am., 56, 1380, doi: 10.1364/JOSA.56.001380

Fusco, T., Sauvage, J.-F., Mouillet, D., et al. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 99090U

Galicher, R., Baudoz, P., Rousset, G., Totems, J., & Mas, M. 2010, Astronomy & Astrophysics, 509, A31

Galicher, R., Baudoz, P., Delorme, J.-R., et al. 2019, A&A, 631, A143, doi: 10.1051/0004-6361/201936282

Gauchet, L., Lacour, S., Lagrange, A.-M., et al. 2016, Astronomy & Astrophysics, 595, A31, doi: 10.1051/0004-6361/201526404

Greenbaum, A. Z., Pueyo, L., Sivaramakrishnan, A., & Lacour, S. 2014, The Astrophysical Journal, 798, 68, doi: 10.1088/0004-637X/798/2/68

Greenbaum, A. Z., Pueyo, L., Ruffio, J.-B., et al. 2018, The Astronomical Journal, 155, 226

Greenwood, D. P. 1977, JOSA, 67, 390

Groff, T. D., Kasdin, N. J., Limbach, M. A., et al. 2014, in Ground-based and Airborne Instrumentation for Astronomy V, Vol. 9147, International Society for Optics and Photonics, 91471W

Gull, S. F., & Skilling, J. 1984in , IET, 646–659

Guyon, O. 2005, ApJ, 629, 592, doi: 10.1086/431209

Guyon, O., Pluzhnik, E. A., Galicher, R., et al. 2005, The Astrophysical Journal, 622,

744
HabEx Study Team. 2019, HabEx, Habitable Exoplanet Observatory, Tech. rep., Jet Propulsion Laboratory. https://www.jpl.nasa.gov/habex/pdf/HabEx-Final-Report-Public-Release.pdf
Haffert, S. 2016, Optics express, 24, 18986
Haffert, S., Bohn, A., de Boer, J., et al. 2019, Nature Astronomy, 3, 749
Haffert, S., Por, E., Keller, C., et al. 2020, Astronomy & Astrophysics, 635, A56
Haniff, C. A., Mackay, C., Titterington, D. J., et al. 1987, Nature, 328, 694
Hardy, J. W. 1998, Adaptive optics for astronomical telescopes, Vol. 16 (Oxford University Press on Demand)
Hartmann, J. 1900, Zeitschrift für Instrumentenkunde, 20, 47
Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, The Astrophysical Journal Letters, 529, L41
Hoeijmakers, H., Schwarz, H., Snellen, I., et al. 2018a, Astronomy & Astrophysics, 617, A144
Hoeijmakers, H. J., Ehrenreich, D., Heng, K., et al. 2018b, Nature, 560, 453
Högbom, J. 1974, Astronomy and Astrophysics Supplement Series, 15, 417
Howard, A. W., Sanchis-Ojeda, R., Marcy, G. W., et al. 2013, Nature, 503, 381
Huang, L., Zhang, S., & Zentgraf, T. 2018, Nanophotonics, 7, 1169
Huygens, C. 1698
Jensen-Clem, R., Mawet, D., Gonzalez, C. A. G., et al. 2017, The Astronomical Journal, 155, 19
Jovanovic, N., Schwab, C., Cvetojevic, N., Guyon, O., & Martinache, F. 2016, Publications of the Astronomical Society of the Pacific, 128, 121001
Jovanovic, N., Absil, O., Baudoz, P., et al. 2018, Review of high-contrast imaging systems for current and future ground-based and space-based telescopes: Part II. Common path wavefront sensing/control and coherent differential imaging, doi: 10.1117/12.2314260
Karalidi, T., Stam, D., Snik, F., et al. 2012, Planetary and Space Science, 74, 202
Kasdin, N. J., Vanderbei, R. J., Spergel, D. N., & Littman, M. G. 2003, The Astrophysical Journal, 582, 1147
Kasper, M., Beuzit, J.-L., Verinaud, C., et al. 2010, in 1st AO4ELT conference-Adaptive Optics for Extremely Large Telescopes, EDP Sciences, 02009
Kenworthy, M. A., Absil, O., Carlomagno, B., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 10702A3
Kenworthy, M. A., Codona, J. L., Hinz, P. M., et al. 2007, \Apj, 660, 762, doi: 10.1086/513596
Kildishev, A. V., Boltasseva, A., & Shalaev, V. M. 2013, Science, 339
Kim, J., Li, Y., Miskiewicz, M. N., et al. 2015, Optica, 2, 958, doi: 10.1364/OPTICA.2.000958
Kirkpatrick, J. D., Cushing, M. C., Gelino, C. R., et al. 2011, The Astrophysical Journal Supplement Series, 197, 19
Klindžić, D., Stam, D. M., Snik, F., et al. 2020, arXiv preprint arXiv:2007.16078
Komanduri, R. K., Lawler, K. F., & Escuti, M. J. 2013, Optics Express, 21, 404, doi: 10.1364/OE.21.000404
Konopacky, Q., Marois, C., Macintosh, B., et al. 2016, The Astronomical Journal, 152, 28
Konopacky, Q. M., Barman, T. S., Macintosh, B. A., & Marois, C. 2013a, Science, 339, 1398, doi: 10.1126/science.1232003

—. 2013b, Science, 339, 1398
Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014, Nature, 505, 69
Kurzynowski, P., Woźniak, W. A., & Szarycz, M. 2011, JOSA A, 28, 475
Lacour, S., Tuthill, P., Amico, P., et al. 2011a, Astronomy & Astrophysics, 532, A72, doi: 10.1051/0004-6361/201116712
Lacour, S., Tuthill, P., Ireland, M., Amico, P., & Girard, J. 2011b, The Messenger, 146, 18
Lacour, S., Nowak, M., Wang, J., et al. 2019, Astronomy & Astrophysics, 623, L11
Lafreniere, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É. 2007, The Astrophysical Journal, 660, 770
Lagrange, A.-M., Boccaletti, A., Langlois, M., et al. 2019, Astronomy & Astrophysics, 621, L8
Lavie, B., Mendonça, J. M., Mordasini, C., et al. 2017, The Astronomical Journal, 154, 91
Love, G. D. 1997, Applied optics, 36, 1517
Lozi, J., Guyon, O., Jovanovic, N., et al. 2018, in Adaptive Optics Systems VI, Vol. 10703, International Society for Optics and Photonics, 1070359
Lozi, J., Guyon, O., Jovanovic, N., et al. 2019, in 6th International Conference on Adaptive Optics for Extremely Large Telescopes, AO4ELT 2019
LUVOIR Team, et al. 2019, arXiv preprint arXiv:1912.06219
Lyot, B. 1939, Monthly Notices of the Royal Astronomical Society, 99, 580
Macintosh, B., Graham, J., Palmer, D., et al. 2006, in Advances in Adaptive Optics II, Vol. 6272, International Society for Optics and Photonics, 62720L
Macintosh, B., Graham, J. R., Barman, T., et al. 2015, Science, 350, 64, doi: 10.1126/science.aac5891
Males, J. R., & Guyon, O. 2018, Journal of Astronomical Telescopes, Instruments, and Systems, 4, 019001
Males, J. R., Close, L. M., Miller, K., et al. 2018, Proc. SPIE
Marchis, F., Thibault, S., Côté, O., et al. 2018, AGUFM, 2018, P41C
Marley, M. S., Saumon, D., Cushing, M., et al. 2012, The Astrophysical Journal, 754, 135
Marois, C., Lafreniere, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006a, ApJ, 641, 556
Marois, C., Lafreniere, D., Macintosh, B., & Doyon, R. 2006b, The Astrophysical Journal, 647, 612
Marois, C., Macintosh, B., Barman, T., et al. 2008, Science, 322, 1348
Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010, Nature, 468, 1080
Martinez, P., Kasper, M., Costille, A., et al. 2013, Astronomy & Astrophysics, 554, A41
Martinez, P., Beaulieu, M., Barjot, K., et al. 2020, Astronomy & Astrophysics, 635, A126
Mawet, D., Pueyo, L., Moody, D., Krist, J., & Serabyn, E. 2010, Proc. SPIE, 7739, 773914, doi: 10.1117/12.858240
Mawet, D., Pueyo, L., Moody, D., Krist, J., & Serabyn, E. 2010, in Modern Technologies in Space-and Ground-based Telescopes and Instrumentation, Vol. 7739, International Society for Optics and Photonics, 773914
Mawet, D., Riaud, P., Absil, O., & Surdej, J. 2005a, ApJ, 633, 1191, doi: 10.1086/462409
—. 2005b, The Astrophysical Journal, 633, 1191
Mawet, D., Riaud, P., Baudrand, J., et al. 2006, Astronomy & Astrophysics, 448, 801
Mawet, D., Serabyn, E., Liewer, K., et al. 2009a, Optics Express, 17, 1902

—. 2009b, Optics express, 17, 1902, doi: 10.1364/OE.17.001902
Mawet, D., Pueyo, L., Lawson, P., et al. 2012, ArXiv e-prints, 8442, 844204, doi: 10.1117/12.927245
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
Meeker, S. R., Mazin, B. A., Walter, A. B., et al. 2018, Publications of the Astronomical Society of the Pacific, 130, 065001
Mendillo, C. B., Hewawasam, K., Howe, G. A., et al. 2019, in Techniques and Instrumentation for Detection of Exoplanets IX, Vol. 11117, International Society for Optics and Photonics, 1111707
Mennesson, B., Juanola-Parramon, R., Nemati, B., et al. 2020, arXiv preprint arXiv:2008.05624
Metchev, S. A., Heinze, A., Apai, D., et al. 2015, The Astrophysical Journal, 799, 154
Miskiewicz, M. N., & Escuti, M. J. 2014, Optics Express, 22, 12691, doi: 10.1364/OE.22.012691
Mollière, P., Stolker, T., Lacour, S., et al. 2020, Astronomy & Astrophysics, 640, A131
Murakami, N., Nishikawa, J., Akaiwa, N., et al. 2016, in Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II, Vol. 9912, International Society for Optics and Photonics, 99126G
Murakami, N., Nishikawa, J., Yokochi, K., et al. 2010, The Astrophysical Journal, 714, 772
N'Diaye, M., Fogarty, K., Soummer, R., et al. 2018, in Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, Vol. 10698, International Society for Optics and Photonics, 106986A
Nielsen, E. L., De Rosa, R. J., Macintosh, B., et al. 2019, The Astronomical Journal, 158, 13
Nowak, M., Lacour, S., Lagrange, A.-M., et al. 2020, A&A, 642, L2, doi: 10.1051/0004-6361/202039039
Oppenheimer, B., Baranec, C., Beichman, C., et al. 2013, The Astrophysical Journal, 768, 24
Otten, G. P., Snik, F., Kenworthy, M. A., et al. 2014, in Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation, Vol. 9151, International Society for Optics and Photonics, 91511R
Otten, G. P. P. L., Snik, F., Kenworthy, M. A., et al. 2017, ApJ, 834, 175, doi: 10.3847/1538-4357/834/2/175
Packham, C., Escuti, M., Ginn, J., et al. 2010, PASP, 122, 1471, doi: 10.1086/657904
Padgett, M. J., & Courtial, J. 1999, Optics letters, 24, 430
Pancharatnam, S. 1955, Proceedings of the Indian Academy of Sciences - Section A, XLI
—. 1956, Proceedings of the Indian Academy of Sciences - Section A, 44, 247, doi: 10.1007/BF03046095
Patty, C. L., Ten Kate, I. L., Buma, W. J., et al. 2019, Astrobiology, 19, 1221
Pearson, D., Taylor, W., Davies, R., et al. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 99083F
Pérez, S., Casassus, S., Hales, A., et al. 2020, The Astrophysical Journal Letters, 889, L24
Perryman, M., Hartman, J., Bakos, G. Á., & Lindegren, L. 2014, The Astrophysical Journal, 797, 14
Petigura, E. A., Marcy, G. W., & Howard, A. W. 2013, The Astrophysical Journal, 770, 69

Poincaré, H. 1892, Théorie mathématique de la lumière II.: Nouvelles études sur la diffraction.–Théorie de la dispersion de Helmholtz. Leçons professées pendant le premier semestre 1891-1892, Vol. 1 (G. Carré)

Por, E., & Haffert, S. 2020, Astronomy & Astrophysics, 635, A55

Por, E. H. 2017, Proc. SPIE, 10400, 104000V

—. 2020, The Astrophysical Journal, 888, 127

Pueyo, L., Bailey, V., Bolcar, M., et al. 2019, BAAS, 51, 215

Radhakrishnan, V. M., Keller, C. U., & Doelman, N. 2018, in Adaptive Optics Systems VI, Vol. 10703, International Society for Optics and Photonics, 107034N

Ragazzoni, R. 1996, Journal of modern optics, 43, 289

Rajan, A., Rameau, J., De Rosa, R. J., et al. 2017, Astronomical Journal, doi: 10.3847/1538-3881/aa74db

Riaud, P., Mawet, D., & Magette, A. 2012, Astronomy & Astrophysics, 545, A150

Roman, N. G. 2019, Annual Review of Astronomy and Astrophysics, 57, 1

Ruane, G., Jewell, J., Mawet, D., Pueyo, L., & Shaklan, S. 2016, in Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II, Vol. 9912, International Society for Optics and Photonics, 99122L

Ruane, G., Wang, J., Mawet, D., et al. 2018, The Astrophysical Journal, 867, 143

Ruane, G., Riggs, A., Mazoyer, J., et al. 2018, in Proc. Spie, Vol. 10698, doi: 10.1117/12.2312948

Ruane, G., Ngo, H., Mawet, D., et al. 2019, The Astronomical Journal, 157, 118

Rubin, N. A. 2020, PhD thesis

Sallum, S., & Eisner, J. 2017, The Astrophysical Journal Supplement Series, 233, 9, doi: 10.3847/1538-4365/aa90bb

Samland, M., Mollière, P., Bonnefoy, M., et al. 2017, A&A, 603, A57, doi: 10.1051/0004-6361/201629767

Schirmer, J., & Schmidt-Kaler, T. 2000, Optics Communications, 176, 313

Seo, B.-J., Patterson, K., Balasubramanian, B., et al. 2019, in Techniques and Instrumentation for Detection of Exoplanets IX, Vol. 11117, International Society for Optics and Photonics, 111171V

Serabyn, E., Mawet, D., & Burruss, R. 2010, Nature, 464, 1018

Shack, R. V. 1971, J. Opt. Soc. Am., 61, 656

Shi, F., Cady, E., Seo, B.-J., et al. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, Vol. 10400, International Society for Optics and Photonics, 104000D

Sing, D. K., Fortney, J. J., Nikolov, N., et al. 2016, Nature, 529, 59

Sissa, E., Gratton, R., Garufi, A., et al. 2018, Astronomy & Astrophysics, 619, A160

Sivaramakrishnan, A., Lafrenière, D., Tuthill, P. G., et al. 2010, in Space Telescopes and Instrumentation 2010: Optical, Infrared, and Millimeter Wave, Vol. 7731, International Society for Optics and Photonics, 77313W

Sivia, D. S. 1987, PhD thesis, University of Cambridge

Skemer, A. J., Hinz, P., Stone, J., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 107020C

Skemer, A. J., Hinz, P. M., Esposito, S., et al. 2012, The Astrophysical Journal, 753, 14

Skemer, A. J., Marley, M. S., Hinz, P. M., et al. 2014, The Astrophysical Journal, 792, 17

Smith, B. A., & Terrile, R. J. 1984, Science, 226, 1421

Snellen, I., de Kok, R., Birkby, J., et al. 2015, Astronomy & Astrophysics, 576, A59

Snellen, I., Albrecht, S., Anglada-Escudé, G., et al. 2019, arXiv preprint arXiv:1908.01803
Snellen, I. A., Brandl, B. R., de Kok, R. J., et al. 2014, Nature, 509, 63
Snellen, I. A., De Kok, R. J., De Mooij, E. J., & Albrecht, S. 2010, Nature, 465, 1049
Snik, F., Otten, G., Kenworthy, M., et al. 2012, Proc. SPIE, 8450, 84500M, doi: 10.1117/12.926222
Snik, F., Absil, O., Baudoz, P., et al. 2018, in , doi: 10.1117/12.2313957
Snik, F., Keller, C. U., Doelman, D. S., et al. 2019, in Polarization Science and Remote Sensing IX, Vol. 11132, International Society for Optics and Photonics, 111320A
Soummer, R. 2004, The Astrophysical Journal Letters, 618, L161
Soummer, R., Hagan, J. B., Pueyo, L., et al. 2011a, The Astrophysical Journal, 741, 55
Soummer, R., Pueyo, L., & Larkin, J. 2012, The Astrophysical Journal Letters, 755, L28
Soummer, R., Sivaramakrishnan, A., Pueyo, L., Macintosh, B., & Oppenheimer, B. R. 2011b, The Astrophysical Journal, 729, 144
Sparks, W. B., & Ford, H. C. 2002, ApJ, 578, 543, doi: 10.1086/342401
Stahl, H. P. 2007, in Optical Manufacturing and Testing VII, Vol. 6671, International Society for Optics and Photonics, 667102
Stam, D. M. 2008, A&A, 482, 989, doi: 10.1051/0004-6361:20078358
Stark, C. C., Belikov, R., Bolcar, M. R., et al. 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 024009
Steeves, J., Wallace, J. K., Kettenbeil, C., & Jewell, J. 2020, Optica, 7, 1267
Stevenson, K. B., Désert, J.-M., Line, M. R., et al. 2014, Science, 346, 838
Stolker, T., Min, M., Stam, D. M., et al. 2017, Astronomy & Astrophysics, 607, A42
Stolker, T., Min, M., Stam, D. M., et al. 2017, A&A, 607, A42, doi: 10.1051/0004-6361/201730780
Su, K., Rieke, G., Stapelfeldt, K., et al. 2009, The Astrophysical Journal, 705, 314
Thatte, N. A., Clarke, F., Bryson, I., et al. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 99081X
Traub, W. A., & Oppenheimer, B. R. 2010, in Exoplanets (University of Arizona Press, Phoenix), 111–156, doi: 10.1098/rsta.2013.0090
Trauger, J. T., Moody, D. C., Krist, J. E., & Gordon, B. L. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011013
Tuthill, P., Monnier, J., Danchi, W., Wishnow, E., & Haniff, C. 2000, Publications of the Astronomical Society of the Pacific, 112, 555
Tuthill, P., Lloyd, J., Ireland, M., et al. 2006, in Advances in Adaptive Optics II, Vol. 6272, International Society for Optics and Photonics, 62723A
Tuthill, P., Bendek, E., Guyon, O., et al. 2018, in Optical and Infrared Interferometry and Imaging VI, Vol. 10701, International Society for Optics and Photonics, 107011J
Tuthill, P. G., Monnier, J. D., & Danchi, W. C. 1999, Nature, 398, 487, doi: 10.1038/19033
van Holstein, R., Stolker, T., Jensen-Clem, R., et al. 2021, Astronomy & Astrophysics
van Holstein, R. G., Snik, F., Girard, J. H., et al. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, Vol. 10400, International Society for Optics and Photonics, 1040015
Van Holstein, R. G., Girard, J. H., De Boer, J., et al. 2020, Astronomy & Astrophysics, 633, A64
van Kooten, M., Doelman, N., & Kenworthy, M. 2019, JOSA A, 36, 731
Vigan, A., N'Diaye, M., Dohlen, K., et al. 2018a, in Adaptive Optics Systems VI, Vol. 10703, International Society for Optics and Photonics, 107035O

Vigan, A., Otten, G., Muslimov, E., et al. 2018b, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 1070236

Vigan, A., Fontanive, C., Meyer, M., et al. 2020, arXiv preprint arXiv:2007.06573

Vos, J. M., Allers, K., Apai, D., et al. 2019, arXiv preprint arXiv:1903.06691

Wang, J., Mawet, D., Ruane, G., Hu, R., & Benneke, B. 2017, The Astronomical Journal, 153, 183

Wang, J., Wang, J. J., Ma, B., et al. 2020, The Astronomical Journal, 160, 150

Wang, J. J., Graham, J. R., Dawson, R., et al. 2018, The Astronomical Journal, 156, 192

Wilby, M. J., Keller, C. U., Snik, F., Korkiakoski, V., & Pietrow, A. G. M. 2017, A&A, 597, A112, doi: 10.1051/0004-6361/201628628

Wolszczan, A., & Frail, D. A. 1992, Nature, 355, 145

Yeh, P. 2007, in 2007 Conference on Lasers and Electro-Optics-Pacific Rim, IEEE, 1–1

Yu, N., & Capasso, F. 2014, Nature materials, 13, 139

Zhou, Y., Apai, D., Bedin, L. R., et al. 2020, The Astronomical Journal, 159, 140

Zurlo, A., Vigan, A., Galicher, R., et al. 2016, Astronomy and Astrophysics, 587, A57, doi: 10.1051/0004-6361/201526835

2 The vAPP coronagraph: creating dark holes in PSFs all over the world

D. S. Doelman, E.H. Por, S.P. Bos, G.P.P.L. Otten, S. Haffert, M. Wilby, K. Miller, M. Ouellet, A. Bohn, J. de Boer, B. Sutlieff, J. L. Birkby, M. Kenworthy, C.U. Keller, F. Snik

Instrument teams:
NCSU: M. J. Escuti, S. Shi, N.Z. Warriner, K. Hornburg
MagAO: J. Males, K. Morzinski, L.M. Close, J. Codona
MagAO-X: J. Males, L.M. Close, J. Long, L. Schatz, J. Lumbres, A. Rodack, K. Van Gorkom, A Hedglen
SCExAO/CHARIS: O. Guyon, J. Lozi, T. Groff, J. Chilcote, N. Jovanovic
HiCIBaS: S. Thibault, C. de Jonge, G. Allain, C. Vallée, D. Pate, O. Côté, C. Marois
LMIRcam/ALES: P. Hinz, J. Stone, A. Skemer, Z. Briesemeister
ERIS: A. Boehle, A. Glausser, William Taylor
METIS: O. Absil, B. Carlomagno
MICADO: P. Baudoz, E. Huby

Accepted by Applied Optics

Over the last decade, the vector-apodizing phase plate (vAPP) coronagraph has been developed from concept to on-sky application in many high-contrast imaging systems on 8-m class telescopes. The vAPP is an geometric-phase patterned coronagraph that is inherently broadband, and its manufacturing is enabled only by direct-write technology for liquid-crystal patterns. The vAPP generates two coronagraphic PSFs that cancel starlight on opposite sides of the point spread function (PSF) and have opposite circular polarization states. The efficiency, that is the amount of light in these PSFs, depends on the retardance offset from half-wave of the liquid-crystal retarder. Using different liquid-crystal recipes to tune the retardance, different vAPPs operate with high efficiencies ($> 96\%$) in the visible and thermal infrared (0.55 µm to 5 µm). Since 2015, seven vAPPs have been installed in a total of six different instruments, including Magellan/MagAO, Magellan/MagAO-X, Subaru/SCExAO, and LBT/LMIRcam. Using two integral field spectrographs installed on the latter two instruments, these vAPPs can provide low-resolution spectra (R$\sim$30) between 1 μm and 5 μm. We review the design process, development, commissioning, on-sky performance, and first scientific results of all commissioned vAPPs. We report on the lessons learned and conclude with perspectives for future developments and applications.

2.1 Introduction

Many different coronagraphs have been proposed since the creation of the first solar coronagraph in 1939 by Bernard Lyot (Lyot, 1939). Originally, the existing coronagraph concepts could be organized in a family tree (Mawet et al., 2012). However, this tree was cut down in the Lorentz Center workshop paper (Ruane et al., 2018), because new developments in coronagraph design lead to the merging of various branches. Due to its simplicity, an adapted version of the Lyot coronagraph is still the most used coronagraph for ground-based high-contrast imaging systems. The use of Lyot coronagraphs resulted in many scientific breakthroughs in understanding exoplanets and circumstellar disks. Preliminary results of the two largest exoplanet surveys are presented by Nielsen et al. (2019) (GPIES) and Vigan et al. (2020) (SHINE), and a summary paper on circumstellar disks was presented by Avenhaus et al. (2018).
The Lyot coronagraph and other focal-plane coronagraphs require accurate centering of the star on their focal-plane masks. Pupil-plane coronagraphs do not have this disadvantage, as the coronagraphic PSF does not change significantly with position in the field of view. The Apodizing Phase Plate (APP) coronagraph is a single-optic pupil-plane coronagraph (Codona et al., 2006; Kenworthy et al., 2007). An APP modifies the phase in the pupil plane to create regions in the PSF where the star light is suppressed, so-called dark zones. A pupil-plane coronagraph has some distinct advantages over a focal-plane coronagraph. First, the coronagraph is simple because it consists of only a single optic, making it easy to install in any high-contrast imaging instrument. Second, the coronagraphic performance is insensitive to tip-tilt errors caused by vibrations or residual wavefront from the adaptive optics system. Moreover, tip-tilt insensitivity is good for the near-infrared (3-5 μm) where nodding is required to remove the background. With nodding the PSF moves over a large distance ($\gg \lambda/D$) on the detector and alignment with a focal plane mask afterwards takes time and the final position is not always the same, resulting in PSF differences. For a pupil-plane coronagraph no realignment is required after nodding, increasing the on-axis time. Third, the coronagraph design can easily be adapted to include complex pupil shapes (e.g. segments/spiders). Fourth, all objects have the coronagraphic PSF, enabling high-contrast imaging of binary systems. These advantages contributed to success of the APP, having imaged multiple substellar companions (e.g. β Pictoris b, discovering HD100546 b and HD 984, (Quanz et al., 2010; Quanz et al., 2013; Meshkat et al., 2015)). Disadvantages of pupil-plane coronagraphs are the high intensity of the stellar PSF often resulting in saturation, the lower planet throughput due to the reshaping of the PSF, and the larger inner working angle with respect to the current best focal-plane coronagraphs. More specifically for the APP, the properties of the diamond-turned phase plate limited the exoplanet yield. As the APP applies dynamic (chromatic) phase, the APP performance is only optimal for a single wavelength, and the dark zone contrast is degraded for broadband light. More importantly, the diamond turning requires smooth phase transitions, which limited the APP designs to an outer working angle of 9 λ/D and restricted dark

zones to a single side of the PSF.
The vector-Apodizing Phase Plate (vAPP) coronagraph is an upgraded version of the APP coronagraph and induces geometric (or Pancharatnam-Berry) phase (Pancharatnam, 1956; Berry, 1987) for circularly polarized light (Snik et al., 2012; Otten et al., 2014a). The vAPP generates two coronagraphic PSFs that have dark zones on opposite sides of the point spread function and have opposite circular polarization states, see Fig. 2.1. The vAPP is a patterned half-wave retarder where the fast-axis orientation changes as a function of position. The induced phase, ϕ, is equal to plus/minus twice the fast-axis orientation θ: $\phi = \pm 2\theta$, with opposite sign for the opposite circular polarization states. Geometric phase is by definition achromatic, and the efficiency, the percentage of light that acquires this phase, depends on the retardance offset from half-wave. The broadband performance of the vAPP is therefore determined by the retardance as a function of wavelength.
The vAPP is manufactured with liquid-crystal technology. A direct-write system is used to print the desired fast-axis orientation pattern in a liquid-crystal photo-alignment layer that has been deposited on a substrate (Miskiewicz & Escuti, 2014). The induced orientation depends on the linear polarization of the incoming light of the direct-write system. Multiple layers of self-aligning birefringent liquid-crystals are deposited on top with varying thickness and twist, carefully designed to generate the required half-wave retardance (Komanduri et al., 2012, 2013). The stack is referred to as a multi-layered twisted retarder (MTR). By tuning the twist and thickness of layers in the MTR, very high efficiencies ($> 96\%$) can be achieved for large wavelength ranges, e.g. $2 - 5\mu$m (Otten et al., 2014a).
As opposite circular polarization states create a PSF with a dark hole on opposite sides, the vAPP implements circular polarization splitting to separate two complementary dark holes. The relative intensity of the two coronagraphic PSFs depends on the circular polarization state of the incoming light. A vAPP for broadband imaging is obtained in combination with a polarizing beamsplitter (e.g. a Wollaston prism) and a quarter-wave plate, but the non-coronagraphic PSFs corresponding to various polarization leakage terms degrade the contrast in the dark holes (Snik et al., 2012; Otten et al., 2014a; Bos et al., 2018). The polarization leakage can be separated from the coronagraphic PSFs by adding a grating pattern (= phase tilt) to the phase pattern, i.e. the "grating-vAPP" (gvAPP) (Otten et al., 2014b). The gvAPP has the same advantages as all pupil-plane coronagraphs, and also overcomes two limitations of the vAPP. First, no polarization splitting or filtering optics are required due to the grating. Second, the gvAPP has two dark holes on opposite side, increasing the search space by a factor two. This comes at the cost of a factor of two in exoplanet throughput, with half of the light being imaged at the bright side of one of the two coronagraphic PSFs. We note that for unknown companions this factor is the same when using an APP, having to rotate the APP to image both sides. Advantages of the two vAPP dark holes are that they have the same AO performance, such that one PSF can be used as a reference for post processing (Otten et al., 2017), and that the antisymmetric PSF is beneficiary for wavefront sensing (Bos et al., 2019). Moreover, the direct-write system is capable of writing much finer structures ($\sim 1\mu$m) compared to diamond

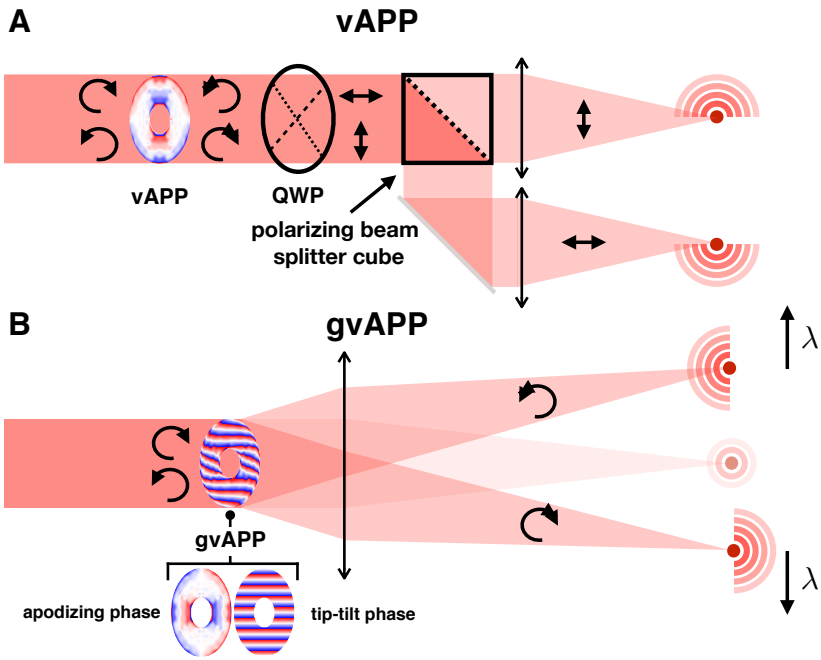


Figure 2.1: Schematic of two different vAPP implementations. *Top:* The vAPP, where PSFs with dark zones on opposite sides and opposite polarization states are separated using a quarter-wave plate and a polarizing beam-splitter. *Bottom:* The gvAPP where polarization states are separated by adding a phase ramp (= polarization grating) to the vAPP phase pattern, which generates opposite tilt for light with left- and right-circular polarization, adapted from (Bos et al., 2018).

turning ($\sim 50\mu$m) and has the ability to write discontinuous phase steps. These properties enable a tremendously increased phase pattern complexity that can be used for holography (Haffert et al., 2018) and manufacture more optimal phase patterns (Doelman et al., 2020).
A third kind of vAPP is the double-grating vAPP (dgvAPP). The dgvAPP combines a gvAPP with a second polarization grating on a separate substrate (Doelman et al., 2020). This second grating is identical to the gvAPP grating and has the opposite effect, diffracting both the main beams back on axis. Polarization leakage of the liquid-crystal film on the first substrate is diffracted outside the dark zone by the second grating, reducing the total on-axis polarization leakage by multiple orders of magnitude. Because both beams are recombined, the phase pattern needs to produce a 360° dark zone, which requires more extreme patterns, increases the inner boundary of the dark zone, and generally reduces the PSF-core throughput. However, the planet light is also recombined in the dark zone, so the throughput is a factor 2 higher.
A common property is that all vAPP implementations is that they operate over 100% bandwidth with extremely high efficiency ($> 96\%$).

2.2 Design of a gvAPP

The design of a gvAPP starts with the optimization of an APP phase pattern. The goal of the optimization is to null the stellar PSF in the dark zone to a very high degree with respect to the PSF core, e.g 10^{-5} . We define the contrast as this ratio of flux in the dark zone divided by the flux in the PSF core and it depends on the focal plane coordinates, see Fig. 2.2. Light has to be diffracted from the stellar PSF core to null the dark zone, reducing the intrinsic Strehl ratio of the star. We define the Strehl ratio as the summed flux of the non-coronagraphic stellar PSF core in an aperture divided by the summed flux of energy in the coronagraphic stellar PSF core using the same aperture. In simulation without noise this aperture has a diameter of one pixel, and for all other applications it is common to use a diameter of $\sim 1.4\lambda/D$(Ruane et al., 2018). Destructively interfering the star light in the dark zone results in constructive interference of the light on the opposite side of the PSF. A reduction of the stellar Strehl ratio is unwanted, as the Strehl ratio of the companion PSF is affected in the same way. Therefore, the optimization simultaneously maximizes the Strehl ratio of the stellar PSF, while minimizing the flux inside the dark zone. This optimization problem is highly non-linear in the complex phase exponential. Consequently, first attempts at calculating APP patterns did not aim to find the optimal solution, their aim was to find solutions that are close. These attempts used phase iteration techniques (Codona & Angel, 2004; Codona et al., 2006) or a modified Gerchberg-Saxton (GS) algorithm (Kostinski & Yang, 2005) and were moderately successful. They have shown that these methods can produce a dark zone of any shape with extreme contrasts ($< 10^{-10}$) (Keller, 2016), yet their Strehl ratios are low for small inner working angles. Another set of solutions were generated by adapting

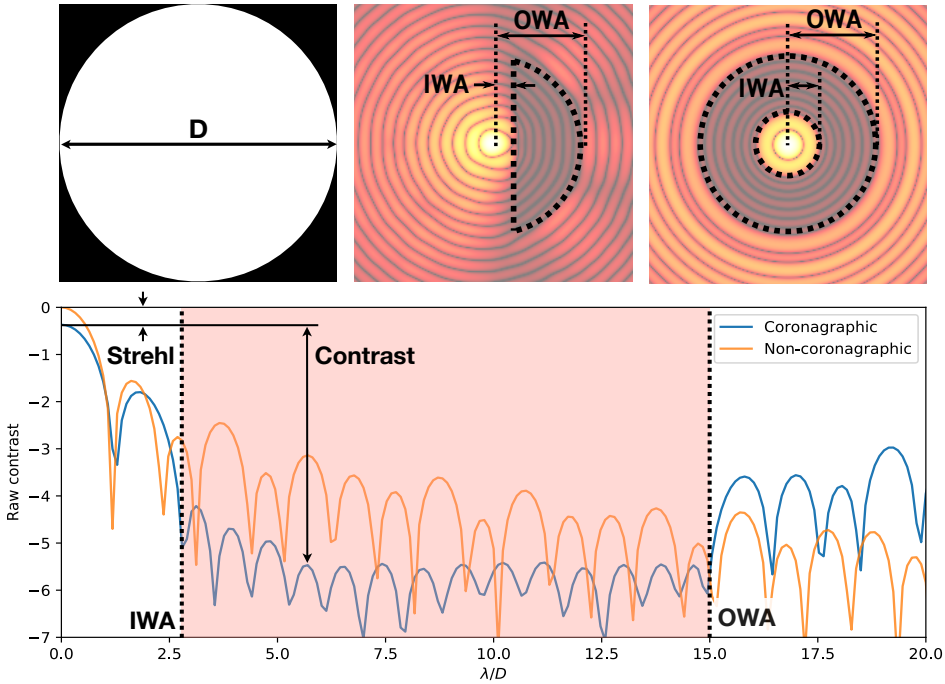


Figure 2.2: Definitions of the dark zone properties. *Top left:* The pupil of the telescope. *Top middle:* A D-shaped dark zone ($\log_{10}$ scale). *Top right:* An annular dark zone ($\log_{10}$ scale). *Bottom:* PSF slice of a non-coronagraphic and coronagraphic design (annular). Adapted from Por (2017).

the global optimization algorithms for shaped-pupil coronagraphs to include phase (Carlotti, 2013). This algorithm produces APPs with regions containing several discrete phases and sharp transitions in between these regions. This algorithm was improved to yield *globally* optimal solutions with unity amplitude across the pupil, smooth phase patterns for 180° and D-shaped dark zones, and 0-π solutions for 360° dark zones (Por, 2017). As this last algorithm allows both phase and amplitude modulation in the pupil, this proves empirically that APPs are the globally optimal solutions for pupil-plane coronagraphs.

2.2.1 Dark zone considerations

The globally optimal solutions are critical to ensure a high exoplanet yield. However, it is not the full story. As shown in (Por, 2017), there is a trade-off between inner working angle (IWA), outer working angle (OWA), contrast, and Strehl. The definitions of the four properties can be found in Fig. 2.2. For focal-plane coronagraphs, the IWA is defined as the smallest angular separation at which the total energy throughput of an off-axis source reaches 50% of the maximum through-

put (Ruane et al., 2018). For pupil-plane coronagraphs this can be adapted to the smallest angular separation where the planet throughput reaches 50% of the maximum throughput in the region where the contrast reaches the target contrast. Equivalently, the OWA is defined as the largest angular separation where the planet throughput still reaches 50% of the maximum throughput in the region where the contrast reaches the target contrast. A good choice of these parameters can be different for every planetary system, observing conditions, wavelength, telescope design and instrument performance. This becomes clear by using a metric that defines a planet detection given instrument limits. In Ruane et al. (2018), the integration time Δt that is required for a 1σ detection is given by

$$\Delta t \propto \left[\frac{\eta_s + \sum_n a_n}{\eta_p^2 - b^2/\epsilon^2} \right], \tag{2.1}$$

where η_s fraction of available star light detected at the planet location, η_p fraction of available planet light detected, and ϵ is the planet-to-star flux ratio. The factor a_n represents noise factors like the background or detector noise such that $\sigma_n^2 = a_n N_\star$, with $N_\star$ the total signal from star in photo-electrons. In the same way $\sigma_{\text{speck}} = bN_\star$, which represents the speckle noise. In this framework, η_p is directly proportional to the vAPP Strehl and the contrast is given by η_s/η_p. We do not minimize Δt with vAPP optimization due to several complicating factors. However, we can use this equation to explore the impact of design choices. Here we outline some of the considerations when choosing these parameters.

1. Strehl: The Strehl ratio of a vAPP design impacts the planet throughput for the full field of view. In the case of the photon noise limit, Eq. 2.1 becomes $\Delta t = \eta_s/\eta_p^2$, demonstrating that Strehl greatly impacts the integration time that results in a 1σ detection. The gvAPP already has the disadvantage of reducing the planet throughput by a factor of 2, so keeping the Strehl high is crucial for good performance. The definition of a high Strehl is somewhat arbitrary given the impact of the shape of the telescope pupil, AO performance and presence of other noise sources. Therefore, we use the Strehl mostly for a trade-off between the other dark zone properties.

2. Contrast: For ground-based high-contrast imaging systems, the intrinsic contrast of a coronagraph is almost never reached. AO residuals and non-common path errors result in quasi-static speckles that limit the performance. In addition, atmospheric jet streams create a wind-driven halo, as the temporal lag between the application of the wavefront correction and the evolving turbulence (Cantalloube, F. et al., 2020). Designing a gvAPP with an intrinsic contrast much lower than the expected raw contrast does not yield the optimal performance. This is clear from Eq. 2.1, where planet is not detected in the presence of speckle noise if $\eta_p < b/\epsilon$ (negative integration time), independent of η_s. Yet, removing bright PSF structures by creating a dark zone reduces speckle pinning, i.e. speckles that are spatially confined to secondary maxima in the diffraction limited PSF (Bloemhof et al., 2001). Taking into account speckle pinning, it can be argued that a design contrast lower than the expected AO-limited contrast does result in an improved performance, although that has not been studied in detail. For inner working

angles $> 2.0\lambda/D$ and a central obscuration lower than 30%, a design contrast of 10^{-4} results in Strehl ratios higher than 70% (Por, 2017). This design contrast is already significantly lower than the raw contrast of extreme AO systems. Another consideration is that constant contrast in the dark zone will yield the highest Strehl, although the changes in Strehl are minor for lower contrasts further out. These considerations lead to the conclusion that a more optimal gvAPP design has a contrast that decreases gradually with the estimated AO residuals and becomes constant at a point where other noise factors take over, such as detector noise or background noise ($\sim 10^{-6}$). An example of such a design contrast is shown in Fig. 2.2. In this case, the vAPP design contrast can be determined by AO simulations and verified with end-to-end simualations of the performance.
3. IWA: All indirect detection methods for exoplanets (e.g. transit, radial-velocity, astrometry), show that there is a huge fraction of planets that is currently out of reach for direct imaging. While these methods are biased to find these close-in planets, it shows that a small IWA is a critical to find new worlds with direct imaging. However, the Strehl ratio is also very dependent on the IWA. For 10^{-6} contrast, an OWA of 8 λ/D, and a central obscuration of 10% changing the IWA from 2.05 to 1.75 λ/D reduces the Strehl from 60% to 20% (Por, 2017). Furthermore, the gain of a smaller inner working angle is usually limited. Uncorrected low-order aberrations, both atmospheric or non-common path, reduce the performance at the smallest separations for ground-based telescopes.
4. OWA: The outer working angle has limited impact on the Strehl of APP designs, while the dark zone area increases with the OWA squared. The non-coronagraphic PSF contains only a small fraction of the energy outside of 10 λ/D, so increasing the contrast in this region results in a small decrease in Strehl as well. In these regions the SNR of the exoplanets is not dominated by the contrast, however. Further out, techniques like ADI are more effective in removing speckles, the wind-driven halo and PSF structures. Thermal background in the near- and mid-infrared ($\lambda > 2$ µm) or detector noise start to dominate the SNR. It is therefore unnecessary for most applications to create a gvAPP with a OWA $> 20\lambda/D$.
There is no general design that works for any telescope or wavelength range. All gvAPP designs are different because the optimal phase pattern depends on the telescope aperture and properties like the AO performance and background noise. It is therefore advisable to perform a grid search to find the trade-offs between IWA, OWA, contrast at the IWA and the slope of the contrast as function of radius. An example grid could be the IWA between 1.8 and 2.3 λ/D in steps of 0.1, the OWA between 12 and 20 λ/D in steps of 1, the log(contrast) between $10^{-3.5}$ and 10^{-5} in steps of 0.5, and a slope between 0.25 and 0.5 log(contrast) per λ/D in steps of 0.05.
Another consideration is the dark zone shape itself. A gvAPP with D-shaped dark zones provides phase solutions with higher Strehl ratios compared to 180 degree dark zones with identical inner working angles. This difference in Strehl increases for smaller inner working angles ($< 2.5\lambda/D$), which is why the D-shape is used in most gvAPPs. The impact of the D-shape on observing is discussed in section 2.3.1. An interesting trade-off exists for planet detection at larger IWAs

($> 3\lambda/D$). Beyond this IWA threshold, the designs with an annular dark zone will have a Strehl ratio that can be competitive with D-shaped dark zones. While the Strehl ratio of designs with annular dark zones are still significantly lower, dgvAPPs with annular dark zones have twice the planet throughput compared to a gvAPP. The origin of this factor two is that half of the planet light is imaged on the bright side of the coronagraphic PSF for a gvAPP. The trade-off between D-shaped and annular dark zones is highly dependent on the telescope pupil, as a central obscuration size and spider thickness greatly impact the Strehl of annular dark zone designs. We note that comparing Strehl ratios with a factor two correction factor for planet throughput and the same contrast levels does not necessarily select the best of the two. There are other factors that are more difficult to add to this trade-off. For example, a dgvAPP will not have wavelength smearing due to grating diffraction, companions will be inside the dark zone for all parallactic angles, post-processing is different, and a gvAPP requires a larger field of view. Because factors like these are difficult to quantify, we did not attempt a general trade-off study. So far, only one dgvAPP has been installed on the large binocular telescope (LBT). The pupil of the LBT is favourable with a central obscuration ratio of $\sim 11\%$ and thin spiders that were not included in the design. Moreover, a vAPP with an annular dark zone was installed on the William Herschel Telescope (WHT), which was designed for an off-axis 1-m pupil without central obscuration and spiders.

2.2.2 Optimization of the APP design

The optimization of the APP starts with a pupil definition. Existing HCI instruments usually have a pupil camera where the pupil can be measured or have their own mask to define the pupil. Extreme caution is warranted when defining the pupil, as the maxima and minima of the phase pattern will be located near the pupil edges. An error in pupil definition leads to a stark reduction in vAPP performance. For this reason we define an APP pupil, which is the undersized version of the true pupil to accommodate alignment errors, definition errors, and pupil movement. Drastic undersizing increases the IWA, negatively impacts Strehl with thicker spiders and a larger central obscuration ratio, and reduces throughput by removing effective telescope area. Therefore, this balance results in an undersizing of the instrument pupil by 2-5%, depending on the amplitude of the expected pupil movement and the alignment tolerances, and in coordination with the instrument team. In addition, it can be beneficial to change the orientation of the dark zones depending on the pupil shape. Spiders add narrow diffraction structures that locally enhance the PSF intensity. A D-shaped dark zone can, depending on the orientation with respect to the pupil, overlap with one or more spider diffraction structures. Removing these structures leads to a decrease in Strehl, so if there is freedom to choose the dark zone orientation the straight edge should be oriented parallel to two spiders. Residual atmospheric dispersion, detector ghosting effects, the FOV of an IFS or the orientation of a slit or image slicer can limit this design freedom.

For a given set of APP parameters and a pupil definition, we use the optimization solver, Gurobi Optimizer (Gurobi Optimization, 2020), to calculate the optimal solution. Computer memory and run time limitations limit the APP phase pattern to 100×100 pixels and a dark zone up to 14 λ/D. For N pixels in the pupil plane, the optimization problem scales as N^2 in computer memory and $N^{3.5}$ in run time. These scaling laws also hold true for pixels in the focal plane for the OWA. Hence, it is not possible to calculate the APP design at the full resolution that is used in the direct-write system, which is on the order of a 1000×1000 pixels. We therefore adopt a two-stage approach for finding the vAPP phase pattern. First we find the optimal APP phase pattern for a downscaled version of the APP pupil. If necessary we decrease the OWA to the 14 λ/D limit set by the computer memory. This downscaled version also does not contain additional dark zones for the coronagraphic PSF with opposite circular polarization or dark zones that minimize crosstalk with the leakage term or additional holograms. In section 2.2.3 the use of these additional dark zones and holograms is explained. Secondly, we upscale this low-resolution APP phase pattern design with Fourier upsampling (Soummer et al., 2007) and use this solution as a prior for the second stage. This second stage adds all dark zones and corrects any errors made during the upsampling of the low-resolution phase pattern. We use a modified Gerchberg-Saxton (GS) algorithm (Gerchberg, 1972) for this stage. This algorithm does not guarantee an optimal solution like the global optimizer does. However, with the starting point already quite close the optimal solution, the GS algorithm is likely to converge before unacceptable Strehl losses occur.
An example for resulting phase patterns after the first and second stage are shown in Fig. 2.3. In practice, we find that limiting the design contrast during the first stage of optimization often increases the Strehl of the final design. Typically, for a final design contrast $< 10^{-4}$ the design contrast during the first stage should be between 10^{-2} and $10^{-3.5}$ to recover a new solution with the highest Strehl. Both optimization stages are implemented in `HCIPy` (Por et al., 2018) and an example of vAPP optimization using only GS can be found in its documentation[1].

2.2.3 Adding functionality with holograms

Enabled only by the accurate and high-resolution of the direct-write process and the liquid-crystal properties, it is possible to add capabilities to all types of vAPPs. Here we will discuss two capabilities, focal-plane wavefront sensing (FPWFS) and reference spots for astrometry and photometry. Both capabilities have solutions based on holograms. We define holograms as PSF copies imaged off-axis, which can be biased with a wavefront aberration. In essence, the vAPP PSFs are also holograms. Combining multiple holograms into a single phase screen is possible through multiplexing wavefronts into a single phase screen, as described in (Dong et al., 2012; Wilby et al., 2017; Doelman et al., 2018). A multiplexed phase screen directs light to multiple holograms, each with a different PSF. These holograms can be added anywhere in the focal plane up to the Nyquist limit, with any bias, and

[1] `https://docs.hcipy.org`

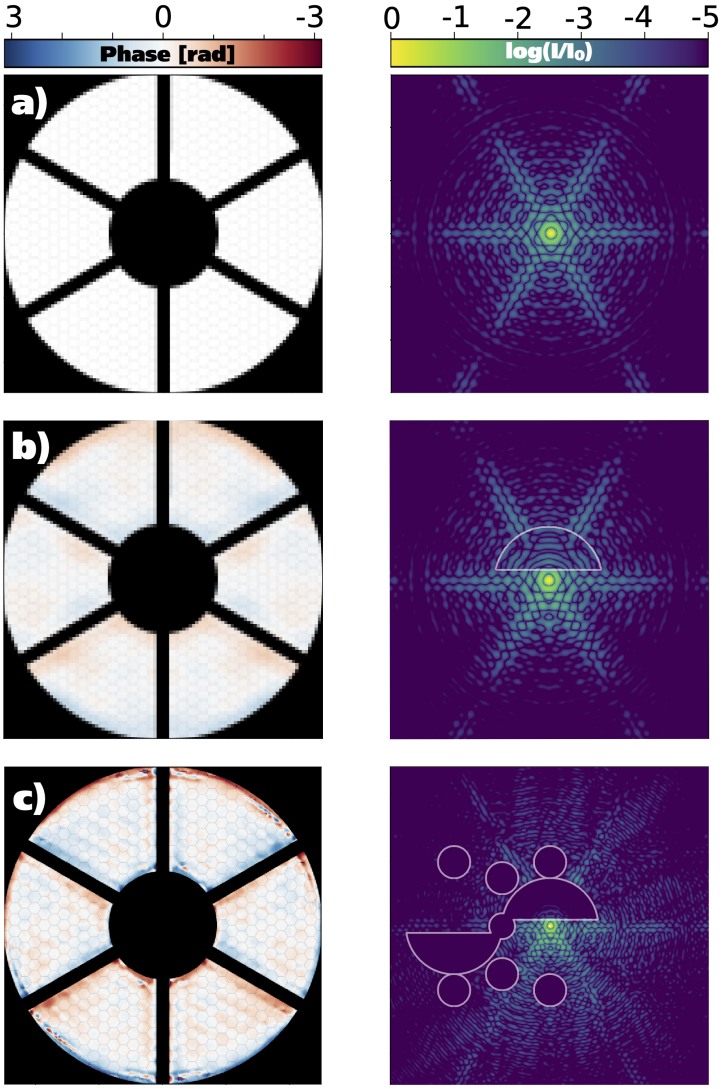


Figure 2.3: Optimization of an APP pattern, showing the phase pattern (left) and corresponding PSF (right). **a)** The low-resolution input pupil for the global optimizer and the corresponding PSF. **b)** Output of the global optimizer for a lower contrast (10^{-3}) and a smaller dark zone, indicated by the white lines. **c)** High-resolution APP design created using the Gerchberg-Saxton algorithm on the Fourier-scaled global optimal APP design. Additional dark zones have been added for the opposite coronagraphic PSF, the leakage PSF and additional holograms.

with any amplitude mixing ratio. Due to circular-polarization splitting, adding a holographic copy at a certain position in the focal plane will necessarily also add a hologram with opposite bias at the location mirrored in the optical axis. These properties of holograms make them very diverse and easily adapted for FPWFS and reference spots. A natural extension is to generate multiple coronagraphic PSFs per polarization state. For example, it is possible to position two coronagraphic PSFs for one polarization such that they precisely overlap with the two coronagraphic PSFs of the opposite polarization state. The advantage is that the otherwise polarized coronagraphic PSFs now become unpolarized (Bos et al., 2020). This has benefits for including polarimetry and/or focal-plane wavefront sensing with the vAPP. One important downside is that multiplexing many holograms results in significant crosstalk. Crosstalk adds PSFs at locations that are the vector-addition of the individual gratings and multiples thereof.

Photometric and astrometric reference spots

For accurate photometry of a companion the vAPP already has an advantage compared to focal-plane coronagraphs. The coronagraphic PSFs themselves provide a photometric reference, given that the two PSFs are not saturated. Otherwise the leakage PSF can serve as photometric reference if it is not too affected by speckles. However, it does have a different spectrum due to the wavelength dependence of the diffraction efficiency. A solution is an unbiased reference spot, which is a copy of the non-coronagraphic PSF. Multiplexing unbiased wavefronts with the vAPP results in reference PSFs that can be used both for photometry and astrometry. For other coronagraphs, such spots have been generated using the deformable mirror or static phase screen. A full summary of these efforts can be found in Bos (2020) and references therein. Moreover, the vector speckle grid that is proposed in Bos (2020) produces speckles that are effectively incoherent with the underlying halo, which greatly improve the photometric and astrometric accuracy. By multiplexing the vAPP with two holograms with opposite phase modulation on the opposite polarization states, this vector speckle grid is easily implemented in the vAPP coronagraph. Astrometry with a vector speckle grid could provide astrometric solutions with a precision of $< 0.01\lambda/D$ (Bos, 2020), however this has not yet been tested on-sky.

Focal-plane wavefront sensing

Adding a bias wavefront to holograms changes their sensitivity to the incoming wavefront. Holographic modal focal-plane wavefront sensing is possible because of these variations (Keller, 2016). FPWFS is critical to remove non-common path aberrations (NCPA), aberrations from optics after the AO system, which are therefore unseen by the AO wavefront sensor. An example of a focal-plane wavefront sensor is the coronagraphic modal wavefront sensor (cMWS) (Wilby et al., 2017; Haffert et al., 2018), which provides simultaneous coronagraphic imaging and focal-plane wavefront sensing with the science point-spread function. The cMWS creates

multiple holograms, each biased with a different wavefront mode drawn from a suitable basis set. The normalized difference between the PSF copies with opposite circular-polarization state responds linearly to the corresponding aberration mode present in the input wavefront. The main advantage of this wavefront sensor is that the monitoring the Strehl ratios of the holograms and the wavefront reconstruction are straightforward and can be done at high speeds and with the actual science camera, independent from the dark hole(s). Moreover, no modulation of the wavefront is required. Multiplexing holograms removes light from the coronagraphic PSF, and for efficient wavefront sensing each mode takes away $\sim 1\%$ of Strehl, on top of the already lower vAPP Strehl. The trade-off between modal coverage, sensitivity and Strehl is difficult to optimize, as it is a priori unclear how much the selected wavefront modes contribute to speckles. The cMWS has successfully been tested with the Leiden exoplanet instrument (LEXI) (Haffert et al., 2018), and in lab demonstrations of Magellan/MagAO-X (Miller et al., 2019).
Another method for FPWFS is an adaptation of phase diversity (PD) (Gonsalves, 1982). Where classical PD requires an in-focus and out-of-focus image, phase diversity holograms are biased with defocus of opposite amplitude, similar to the cMWS with a larger bias. This larger bias results in an increase in modes that can be reconstructed, while decreasing the average intensity of the hologram. So compared to the cMWS, a larger fraction of the incoming light is diffracted to the PD holograms, e.g. 10% instead of 1%. This FPWFS method was implemented in the Subaru/SCExAO(Doelman et al., 2017), Magellan/MagAO-X(Miller, 2018) and HiCIBaS instruments. A different implementation of PD is described in Por & Keller (2016), directly reconstructing the focal-plane electric field in the dark zone. Coronagraphic PSF copies are created with each a unique probe in the dark zone, making use of the pairwise nature of the holograms (Keller, 2016). The unaltered science PSF is still available, providing the same advantages as the cMWS, but adding the capability to sense all modes corresponding to the dark zone.
Interestingly, the coronagraphic PSFs themselves can also be designed to encode wavefront information in the bright field by including a pupil-plane amplitude asymmetry. The added advantage of the vAPP is that there are two coronagraphic PSFs, with opposite bright fields that cover all of the spatial frequencies. This reduces the number of modes that can otherwise not be measured by one coronagraphic PSF (Sun, 2019). We refer to such vAPP designs as Asymmetric Pupil vAPPs (APvAPPs), and was successfully implemented by Bos et al. (2019). They show that if the pupil is asymmetric, both odd and even modes can be recovered from the coronagraphic PSFs only. Earlier versions of an asymmetric pupil wavefront sensor (Martinache, 2013) could not be combined with a coronagraph. A non-linear model-based wavefront sensing algorithm was proposed by (Bos et al., 2019) and is limited, due to computational considerations, to the first ~ 100 Zernike modes, yet it showed nearly an order of magnitude improvement in contrast between 2 and 4 λ/D with the internal source, and on-sky the gain was a factor of 2. Spatial linear dark field control (LDFC) (Miller et al., 2017, 2019) is also a very powerful combination with the APvAPP. LDFC empirically derives a linear relationship between DM actuation and focal-plane intensity re-

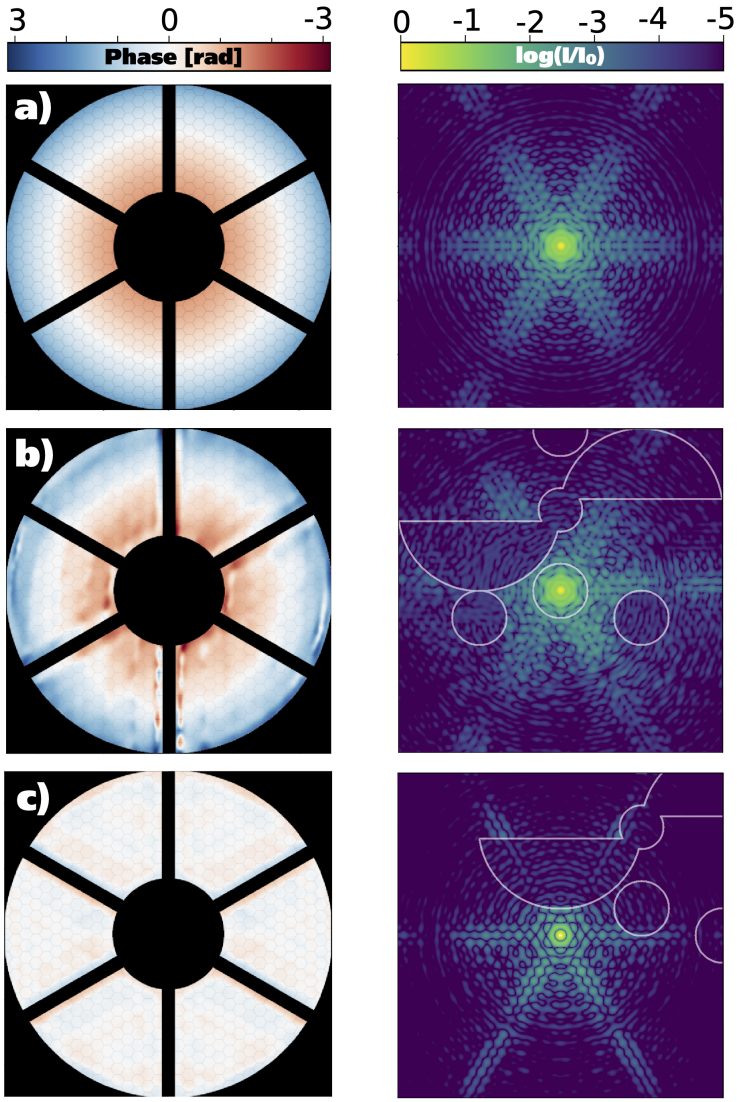


Figure 2.4: Updates of the holographic phase patterns to minimize crosstalk and light scattering into the dark zones. **a)** An unaltered phase diversity hologram with a defocus of 0.87 radian RMS. **b)** The updated PD hologram with suppressed diffraction structures in all dark zones, calculated using the Gerchberg-Saxton (GS) algorithm. **c)** Same as b), only without defocus to create a photometric or astrometric reference hologram. Each hologram has displaced dark zones. This displacement corresponds to the relative location of the hologram with respect to the dark zones of the coronagraphic PSF after multiplexing the phase patterns.

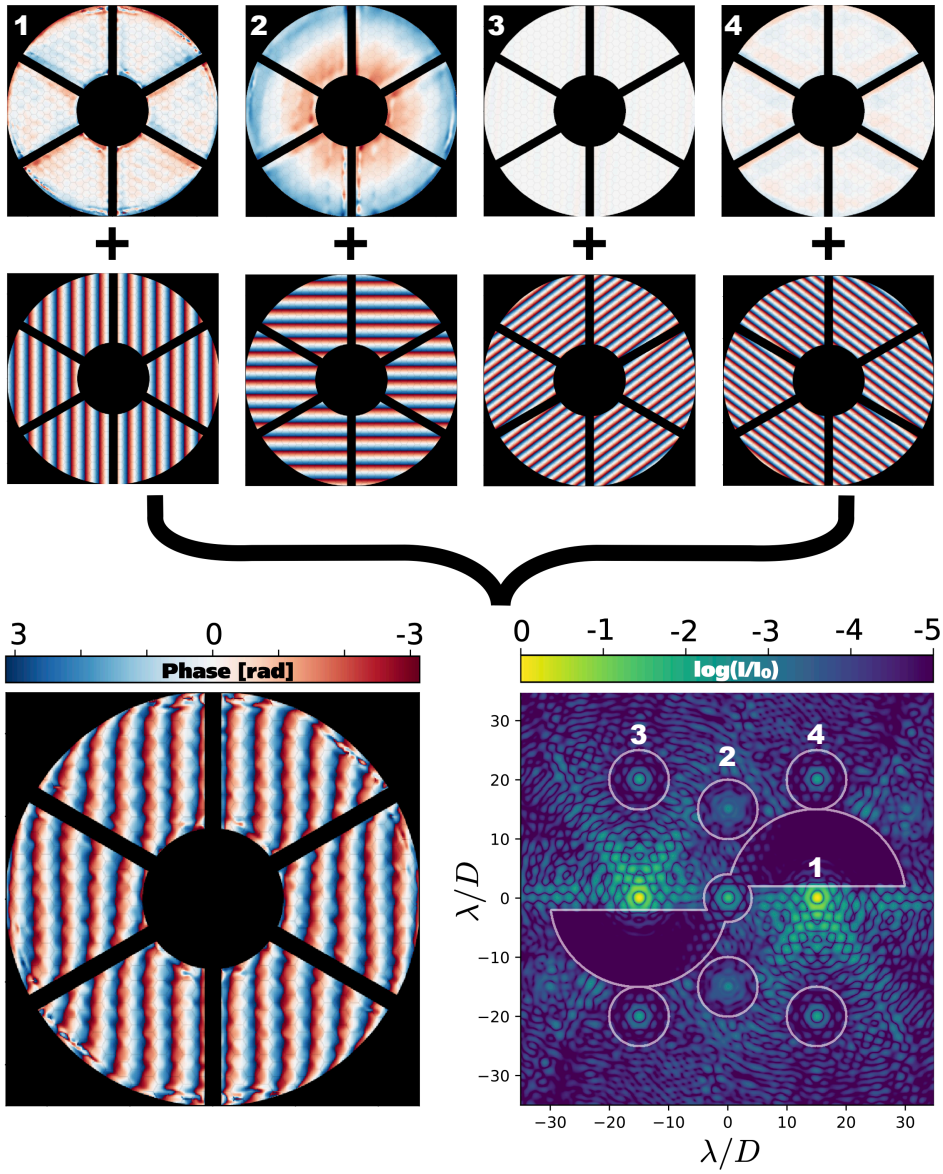


Figure 2.5: Creation of a gvAPP phase pattern by multiplexing the APP design (1) with the three holograms. Hologram 2 is defocused and could be used for focal-plane wavefront sensing. Hologram 3 and 4 are unaberrated and are located in a square grid at the grating frequency for astrometric referencing. Additional iterations with GS have cleared the dark zone of second order effects caused by multiplexing. The PSF is calculated assuming unpolarized light and a polarization leakage of 1%.

sponse in the bright field, allowing it to control a larger number of modes. As LDFC can only drive back the wavefront to a reference state, it requires an initial calibration by non-linear model-based wavefront sensing algorithm, making both methods complimentary. LDFC uses the linear response to wavefront aberrations in the bright side of the vAPP PSF to reconstruct the wavefront. For Magellan/MagAO-X, simulations have shown that LDFC is capable of correcting the full dark zone using 400 modes (Miller et al., 2019). Recently, LDFC in combination with an APvAPP was installed and tested using Subaru/SCExAO (Miller et al., 2020). An on-sky demonstration followed soon after (Bos et al., Submitted for publication). By multiplexing two coronagraphic PSFs per polarization state, each of these PSFs designed with a different pupil-plane amplitude asymmetry, it is also possible to design APvAPPs that can measure both pupil-plane phase and amplitude modes simultaneously (Bos et al., 2020). The possibility of adding photometric, astrometric, and wavefront sensing capabilities to the vAPP make it a unique and diverse coronagraph. Exploring these techniques for a more integrated approach to building high-contrast imaging systems shall lead to a better performance at the systems level.

Implementation into vAPP designs

Adding holograms by multiplexing phase patterns does not result in an optimal combination. Crosstalk between modes and the vAPP PSF will decrease the contrast in the vAPP coronagraphic dark zones. Moreover, the crosstalk will change the response of the holograms. Minimizing this crosstalk is implemented by changing the vAPP PSF, the modes and optimizing after multiplexing. First, the coronagraphic PSF can be suppressed at the locations of the holograms, such that interference is minimized. Additional dark zones can be added during the upscaling of the APP design. Second, the modes can be adapted to also have a reduced intensity at the location of the APP dark zone. These dark zones can be added using the GS algorithm, as is shown in Fig. 2.4. All modified holograms, including the coronagraphic vAPP hologram, can then be combined by multiplexing. Third, the crosstalk in the vAPP dark zones caused by the multiplexing can be removed by applying an additional round of GS. The combined result is shown in Fig. 2.5. Every time GS is applied, it has a minor impact on the Strehl of the coronagraphic PSF. The final cleaning of the dark zone could impact the performance of holographic wavefront sensors, though we estimate that this effect is minimal, as no significant deviation from expected performance was found in (Miller et al., 2019)

2.2.4 Additional design choices

So far we have focused on the design of the vAPP phase pattern. However, there are additional choices regarding the vAPP optic that are independent of the phase pattern. We will highlight two design choices that impact the manufacturing complexity and the wavelength dependent performance of the vAPP.

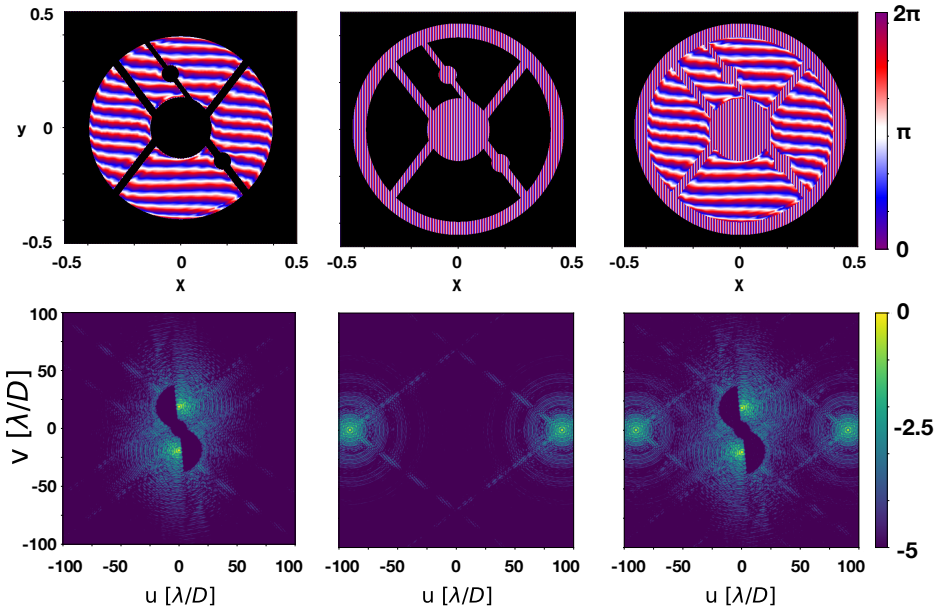


Figure 2.6: *Top left:* Phase of a grating-vAPP for an undersized pupil of the Subaru Telescope. An amplitude mask for the pupil is shown in black. *Bottom left:* Logarithmic plot of the PSF of the grating-vAPP, showing two coronagraphic PSFs for opposite handedness of circularly polarized light. The grating-vAPP was simulated to be a perfect half-wave retarder, so no leakage is present. *Center:* The grating mask (top) and resulting PSF (bottom) that will be used as an amplitude mask. The grating has 90 periods across the diameter of the pupil. *Right:* The phase and PSF of the grating-vAPP combined with the grating mask. The incoming light has a pupil of radius 0.45 (black) and the light is separated in a grating mask component and a grating-vAPP component. A small misalignment of the pupil does not affect the coronagraphic PSF, relaxing alignment tolerances. Adapted from Doelman et al. (2017).

The grating mask

As discussed in section 2.2.2, the pupil definition is critical for the performance of the vAPP. During manufacturing it is therefore standard to add a physical binary amplitude mask to the assembly that is aligned with sub-micron accuracy with the phase pattern. These amplitude masks and their alignment add significant complexity and cost to the vAPP optic. Especially for lab demonstration projects that often require multiple phase patterns, it is not cost effective to use these amplitude masks. Separate metal laser-cut masks can provide a cheap solution at the cost of reduced alignment accuracy. We note that it is not necessary to absorb or reflect the light at the pupil plane. A different solution is to diffract the light outside of the field of view of the coronagraphic PSFs. Diffracting the light that is not falling onto the coronagraphic phase pattern to outside the FOV requires a patterned high-frequency polarization grating outside of the pupil, e.g. beyond the outer diameter, inside the spiders, and within the central obscuration. This grating should have a different orientation and frequency than the grating of the gvAPP. A simple spatial filter in an intermediate focal plane can remove the light outside the pupil. We call this concept the "grating mask". This concept is described in detail in Doelman et al. (2017), and a reworded version is added here for clarity. An example of the gvAPP with a grating mask is shown in Fig. 2.6. Critical for the workings of a PG as amplitude mask is that the tilt is applied *locally*, as the phase is accumulated inside the liquid-crystal layer (Escuti et al., 2016). This is unlike a regular grating where the interference between the *full* grating results in the diffraction orders. Adding a mask only outside the pupil does not significantly alter the coronagraphic PSF. It does, however, change the leakage term from the PSF of the undersized vAPP pupil to the PSF of the telescope or instrument pupil.

How the grating mask reduces the complexity of the vAPP optic can be seen in Fig. 2.7. With a grating mask the first substrate is no longer necessary, removing a glue layer as well. Moreover, the lab demo vAPPs can be manufactured without any additional substrates. While their phase patterns and pupils can be extremely complex, only homogeneous illumination is necessary for them to work as expected. We note that an end cap would make them more durable and less easy to damage. For lab testing a 1 cm grating mask was produced with 8 pixels of 8.75 micron per period. The pupil is a simplified version of the Nancy Grace Roman Space Telescope and was chosen for the thin spiders and complex pupil design. The full mask between crossed polarizers is shown on the left in Fig. 2.8 in addition to a microscopic image which shows a spider with the grating pattern. The width of the spider is a few times the grating period, yet this does not affect the performance of the grating mask. The right panel of Fig. 2.8 shows the reimaged pupil when a field stop is added to the focal plane. The light outside the simplified RST pupil, which is diffracted by the grating, is blocked and does not appear in the reimaged pupil.

When applied in a real high-contrast imaging system, two things need to be considered. First, one has to be careful where the diffracted light from outside the

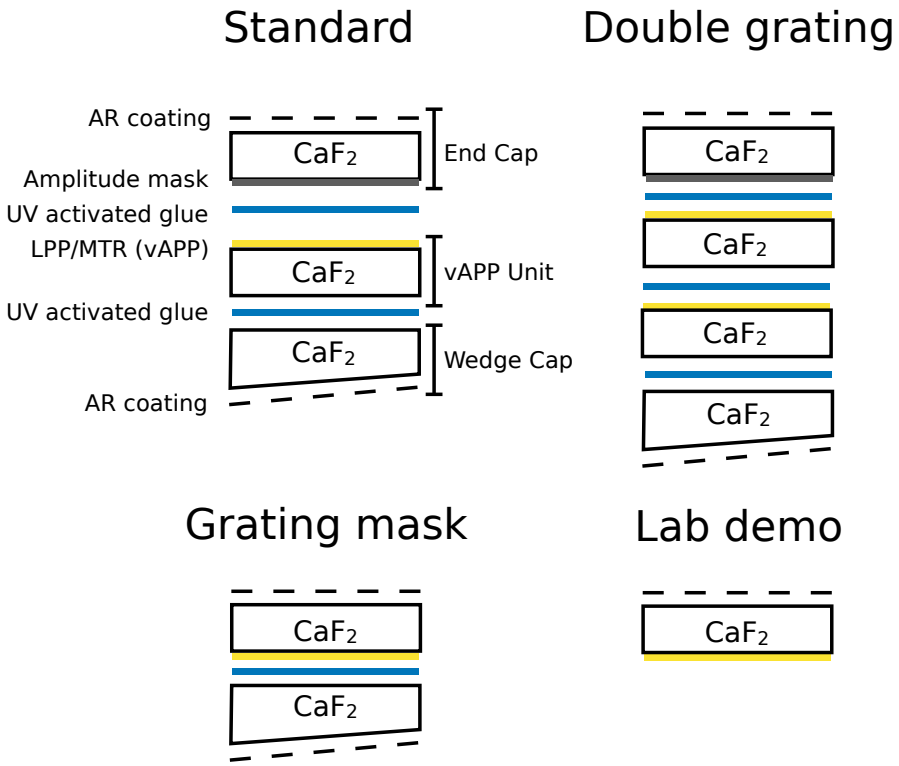


Figure 2.7: Various substrate designs for vAPP manufacturing.

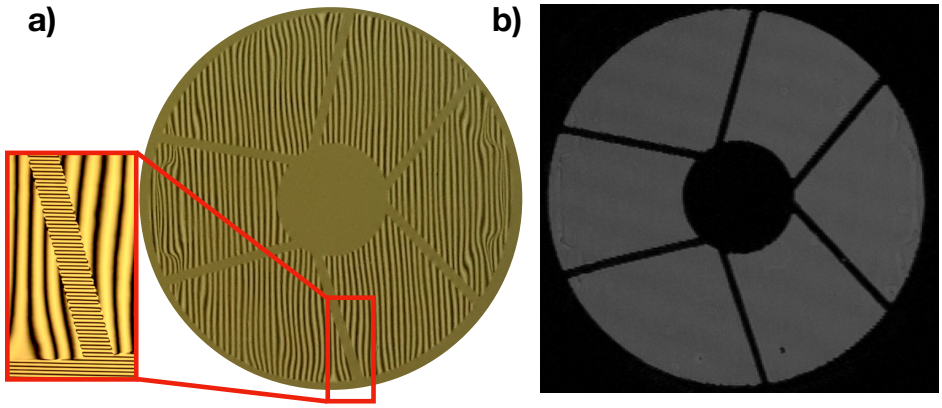


Figure 2.8: **a)** Photograph of the vAPP with a simplified RST pupil between crossed polarizers, manufactured for testing the grating mask. The zoomed in microscope image shows the high-frequency grating used for the grating mask that is not captured by the normal camera. **b)** Reimaged pupil with a field stop inserted in the focal plane. The field stop removes the diffracted light from outside the pupil defined by the grating mask. Even spiders with a width of a few times the grating period are removed extremely well.

pupil is directed to. The diffracted light can scatter back onto the detector by reflections inside the instrument if not blocked. Because the direction of the PG is a free design parameter, the reflections can be directed to non-reflective parts with prior knowledge of the instrument layout. Second, a few percent of the light that would be normally blocked by an amplitude mask would end up on the science camera due to polarization leakage. This polarization leakage is always non-zero for a deviation in retardance from half-wave. The influence of the polarization leakage will be small when the coronagraphic PSFs are have enough spatial separation from the leakage. This separation is also a free parameter and is set by the frequency of the grating of the coronagraphic phase pattern. In addition, the pupil defined by the grating mask will be undersized a few % compared to the telescope pupil, so the amount of light that is diffracted by the grating mask is limited.
Overall, the grating mask is a cheap alternative for amplitude masks and alleviates the problems that come with aligning an amplitude mask with the phase pattern.

Wavelength selective multi-twist retarders

Another option for the vAPP involves the tuning of the retardance in a non-standard way. For most applications the retardance is tuned to be close to half-wave over the full bandwidth of an instrument (Komanduri et al., 2013; Otten et al., 2017) to maximize the efficiency. However, by changing the liquid-crystal recipe one can, in principle, make the retardance follow any continuous curve as

function of wavelength (Hornburg et al., 2014). We note that this is possible because the fast-axis orientation and retardance of the devices are decoupled. One specific application of a liquid-crystal recipe is the "wavelength-selective coronagraph" (Doelman et al., 2017). Here we include a short summary of this work. A wavelength-selective coronagraph has rapidly changing efficiencies, where the retardance switches between intervals of 180 degrees as function of wavelength. A retardance of zero or a full wave results in a zero efficiency, such that the light traveling through the coronagraph does not acquire geometric phase. The coronagraph is effectively switched off at those wavelengths, while it is switched on in the range where the retardance is close to half-wave. This technology has been demonstrated for RGB lenses (Hornburg et al., 2019). An example for the vAPP coronagraph is shown in Fig. 2.9. The retardance is close to 180 degrees for Hα and close to zero for smaller than 600 nm and larger than 750 nm. The light outside the Hα spectral band can be used by a wavefront sensor, enabling wavefront sensing close to the science bandwidth with optimal efficiency. In addition, the wavefront sensor can be placed after the coronagraph to minimize non-common path aberrations. This requires a simple focal plane mask that separates the central term from the coronagraphic PSFs, aided by some spectral filters to minimize spectral cross-talk.

2.3 Observing and data reduction

The gvAPP can be described as a simple coronagraph, as it is only a single pupil-plane optic that suppresses the diffraction halo. However, there are two differences between other coronagraphs that make both observing and data reduction much less straightforward, i.e. the D-shaped dark zone and the grating. For the D-shape dark zone of a gvAPP the dark zones do not cover 360 degrees, which takes away a larger fraction for closer-in orbits caused by sky-rotation. This has to taken into account for planning observations and in the data reduction process. In addition, the grating can smear the planet PSF over multiple pixels in the grating direction, reducing signal-to-noise and the effectiveness of observing techniques like angular differential imaging (ADI) (Marois et al., 2006).

2.3.1 Influence of the D-shaped dark zone

The D-shaped dark zone provides phase solutions that have much higher Strehl ratios for similar smaller inner working angles than a 180 degree dark zone, while providing a similar total area. We note that the term inner working angle is somewhat misleading here. It is correct that the D-shaped and 180 degree dark zone designs can detect a planet down to the same separation, yet the *amount of time* it is in the dark zone for pupil tracking mode is very different. For a 180 degree dark zone the coverage is a full 360 degrees, while a bar is missing for the D-shaped dark zone, which takes away a larger fraction for closer-in orbits. The fraction of a circle, F_{DH}, with radius r that is inside a dark zone with an inner

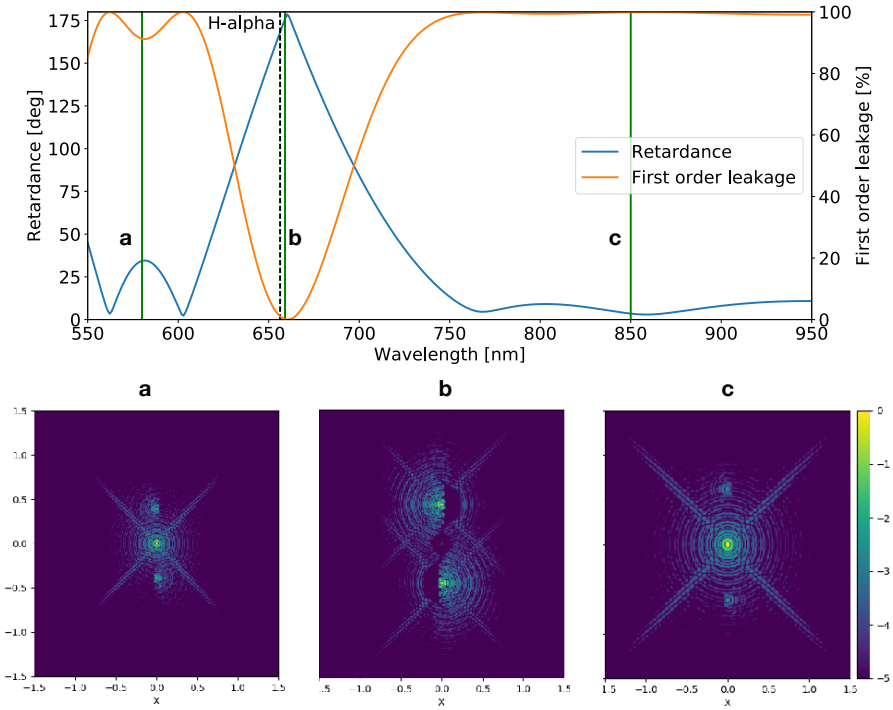


Figure 2.9: *Top:* Simulated retardance (blue) and first order leakage (orange) as a function of wavelength for a 3-layered MTR. The total thickness is 20.7 micron and it was optimized to have 0 retardance everywhere except for a 10 nm band around Hα wavelength (656.28 nm), where the retardance is close to 180 degrees. *Bottom:* The simulated PSFs of a grating-vAPP with the retardance profile shown above. The green lines in the top figure correspond to the wavelengths of the bottom panel. At wavelengths shorter than Hα the leakage term dominates and almost no coronagraphic PSF can be seen (a). This is the same for wavelengths larger than 750 nm (c). Around Hα the leakage term disappears and the gvAPP operates as normal (b).

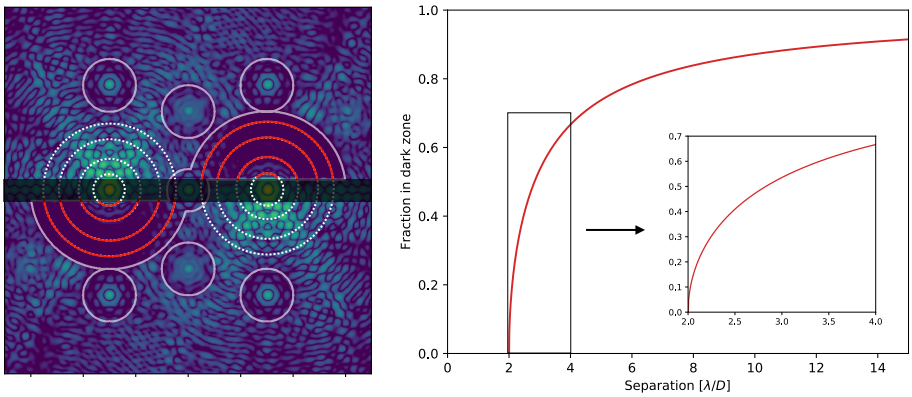


Figure 2.10: Planets rotating in the dark zone in pupil-stabilized mode are not always in the dark zone. *Left:* Fraction of a circle (white) that is in the dark zone (red). *Right:* The same fraction as function of separation for an inner working angle of $2\lambda/D$.

working angle IWA is given by

$$F_{DH} = \frac{\pi - 2\arcsin(\mathrm{IWA}/r)}{\pi}. \tag{2.2}$$

This fraction is also shown in Fig. 2.10 for a typical gvAPP design. On-sky the rotation speed can change dramatically near Zenith, and the position in the dark zone depends on the local sidereal time at the observatory. Therefore, observations with the gvAPP should be carefully planned if a known target is observed. Planning does not help with detecting unknown objects where certain areas around the star will have more noise due to the sky-rotation placing objects in the bar between dark zones.

For known objects the blind spots created by the incomplete coverage of the D-shaped dark zones can be predicted. To this end, we created a preliminary version of an observation preparation tool in `python`, which can be found on `Github` [2]. The goal of this tool is to predict the locations of known companions and provide information on when it is in the dark zone to help determine observation strategies. The current version has an internal library that contains the information of a few operational gvAPPs, such that a simple query returns the predicted locations of companions in a PSF image. The code needs the instrument name, target name, the companion position angle and separation, and the observation times. Wavelength scaling has not been implemented yet. The core of the observation tool is based on `Astropy` (Price-Whelan et al., 2018), which is used to retrieve object information and coordinate transformations.

[2] `https://github.com/18alex96/vAPP_preparation`

2.3.2 Influence of wavelength smearing

The grating of the gvAPP is an elegant solution to separate the polarization leakage from the coronagraphic PSFs. It does, however, represent a challenge for the data reduction. For narrow-band data it displaces the two coronagraphic PSFs which need to be recombined after centering them separately. This is similar for data taken with an integral field spectrograph (IFS), except the instrument provides simultaneous multi-wavelength observations. For each wavelength, the stellar PSF and companion PSF will have shifted and also needs to be recombined. This can be characterized, especially if there are no optical distortions in the IFS such that the displacement from the grating is linear. Wavelength smearing is likely to reduce the astrometric accuracy in addition to make object spectra more sensitive to the lenslet flat. For larger bandwidths the wavelength smearing introduces more problems. The light of the planet is smeared over many pixels, resulting in an increase in camera noise and a larger sensitivity to speckle noise. In addition, the planet PSF rotates around the stellar PSF for every wavelength, which is a different location on the detector for every wavelength. In the derotated rest frame of the stellar PSF the orientation of the elongation changes in time, such that ADI techniques are less effective. Collapsing the PSF in the grating direction before derotation might solve this problem, but this does not remove the sensitivity to speckle and camera noise.
The classical vAPP, which separates the two polarization states with a Wollaston prism, would not have these issues caused by the grating (Snik et al., 2012). However, they would be limited by polarization leakage at the smallest separations for large bandwidths. A double-grating vAPP is also not affected by the grating, it is not limited by polarization leakage, and does have full 360 degree FOV at the cost of a larger inner working angle. Combining these two techniques creates a solution for the wavelength smearing and large inner working angle. This combination would consist of a Wollaston prism to separate the two polarization states of a double-grating vAPP with D-shaped dark zones (Bos et al., 2018).

2.3.3 Data reduction

From the complications of data reduction caused by the D-shape of the dark zone and the grating, it is clear that gvAPP data can not simply be reduced with commonly-used data reduction pipelines like the vortex image processing package (`VIP`) (Gonzalez et al., 2017) or `PynPoint` (Amara & Quanz, 2012; Stolker et al., 2019). These pipelines include different data reduction algorithms that have been developed for data taken without coronagraphs or with focal-plane coronagraphs. In both situations the stellar PSF is not separated and the search space is 360 degrees. However, with a relatively simple adaptation it is possible reshape gvAPP data in a way that these pipelines can be used. We create a window for each coronagraphic PSF and combine them into a single new data frame that is centered on both PSF cores. An example is shown in Fig. 2.11. This method also works for IFS data if the windows change location with wavelength,

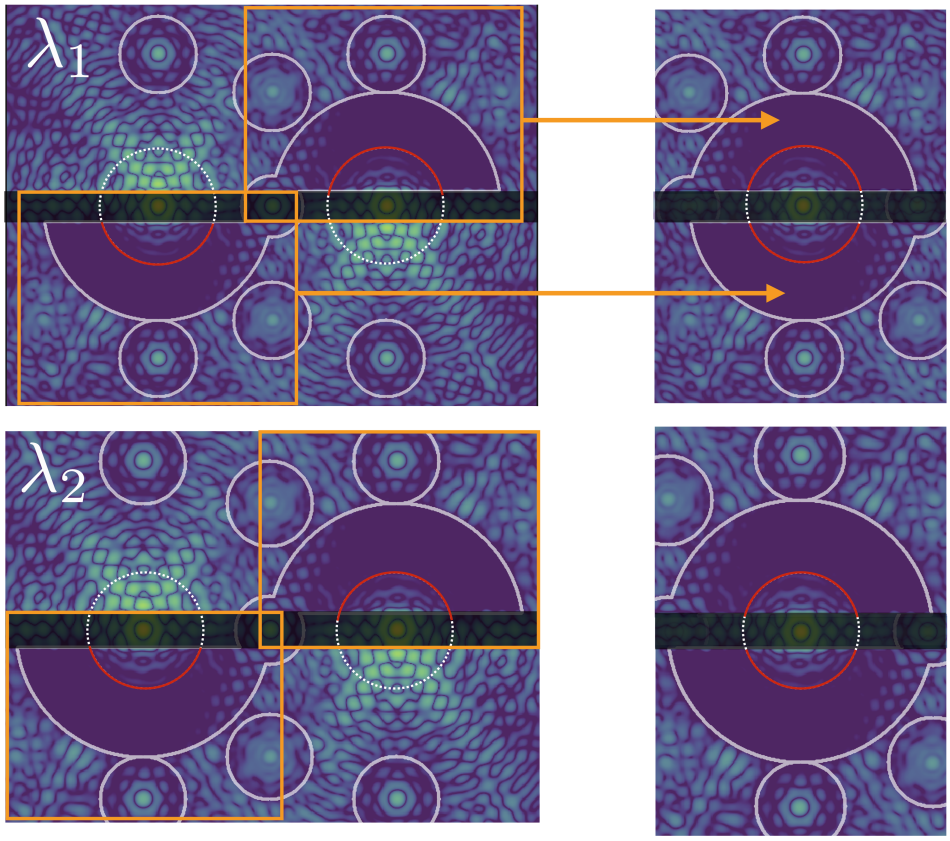


Figure 2.11: Method of recombining both coronagraphic PSFs such that the data can be processed using conventional data reduction techniques.

but does not solve the wavelength smearing for large bandwidths. In addition, it does not make use of the fact that both coronagraphic PSFs are aberrated by the same wavefront aberration. The rotate subtract algorithm, suggested by (Otten et al., 2017), provides an additional contrast gain by using this symmetry. An improved version of this algorithm does not simply subtract the average PSF but uses principle component analysis (Wold et al., 1987) to create a PSF library using one coronagraphic PSF and optimally removing these components from the other PSF. Another interesting technique that seems well suited for gvAPP data is the Temporal Reference Analysis for Exoplanets (TRAP) algorithm (Samland, 2019). TRAP creates a data driven model of the temporal behavior of the systematics using reference pixels, which can also be the pixels at the other coronagraphic PSF. This algorithm would also work for large bandwidths.

2.4 The world of vAPP

Over the last decade the vAPP has been installed in 7 instruments on 6 telescopes, and will be installed in two instruments on the Extremely Large Telescope (ELT). The success of the vAPP can be explained by two of its properties: simplicity and adaptability. Each vAPP is unique and fully optimized for the system it is implemented in, both in wavelength coverage as phase design, while only a single optic needs to be installed. The vAPPs cover almost an order of magnitude in bandwidth, i.e. 0.55-5 μm, provide multiple ways of focal-plane wavefront sensing, and provide photometric/astrometric capabilities. First, we will provide an overview of the different properties and locations of the vAPPs. Then, we will shortly describe the unique capabilities of each vAPP.

2.4.1 The vAPP: a global view

High-contrast imaging systems are now part of most 8m-class telescopes. The vAPP has benefited from the enormous growth in the number of high-contrast imaging systems. A global view of the vAPP coronagraphs is shown in Fig. 2.12. The vAPPs are operational in both the Northern and Southern hemisphere. For each vAPP, we show the phase pattern and the simulated PSF. In addition, we display the project status, indicated by the color of the marker on the map. As of 2020, the vAPP has been tested in two former instruments, is operational in four instruments and will be installed on three more instruments. A summary of most properties of all vAPPs are presented in Table 2.1.
Another important property of the vAPP is the polarization leakage as function of wavelength. The design of the liquid-crystal recipes aims to minimize polarization leakage over the full bandwidth and are given in Table 2.1. For small bandwidths ($< 15\%$), the liquid-crystal design can be kept simple with only single layer with zero twist. Larger bandwidths require multiple layers, each with different thickness and twist, to form a multi-twist retarder (MTR).

Table 2.1: Properties of the vAPP coronagraphs. FPWFS techniques: 1 = AP-WFS (Bos et al., 2019), 2 = cMWS (Keller, 2016; Wilby et al., 2017; Miller et al., 2019), 3 = PD (Gonsalves, 1982; Miller, 2018).

Telescope	Instrument	Type	Inst.	λ [μm]	MTR	Contrast	IWA [λ/D]	OWA [λ/D]	Strehl [%]	FPWFS
Magellan	MagAO/Clio	gvAPP 2x180	2015	2-5	3	10^{-5}	2.0	7	39.8	-
WHT	LEXI	vAPP 360	2017	0.6-0.8	1	10^{-4}	2.5	6	57.2	2
Subaru	SCExAO/ CHARIS	gvAPP 2x180	2018	1-2.5	3	10^{-5}	2.1	11	49.2	1+3
FC	HiCIBaS	gvAPP 2x180	2018	0.83-0.88	1	10^{-6}	2.1	8.5	28.8	2+3
Magellan	MagAO-X	gvAPP 2x180	2018	0.55-0.9	3	10^{-5}	2.1	15	40.7	1+2+3
LBT	LMIRcam	gvAPP 2x180	2018	2-5	3	10^{-5}	1.8	15	52.6	3
LBT	LMIRcam/ ALES	dgvAPP 360	2018	2-5	3	10^{-5}	2.7	15	42.0	-
VLT	ERIS	gvAPP 2x180	2022	2-5	3	10^{-5}	2.2	15	50.9	1
ELT	METIS	gvAPP 2x180	2028	2.9-5.3	3	10^{-5}	2.5	20	63.8	1
ELT	MICADO	gvAPP 2x180	2026	1-2.5	3	10^{-5}	2.6	20	68.8	-

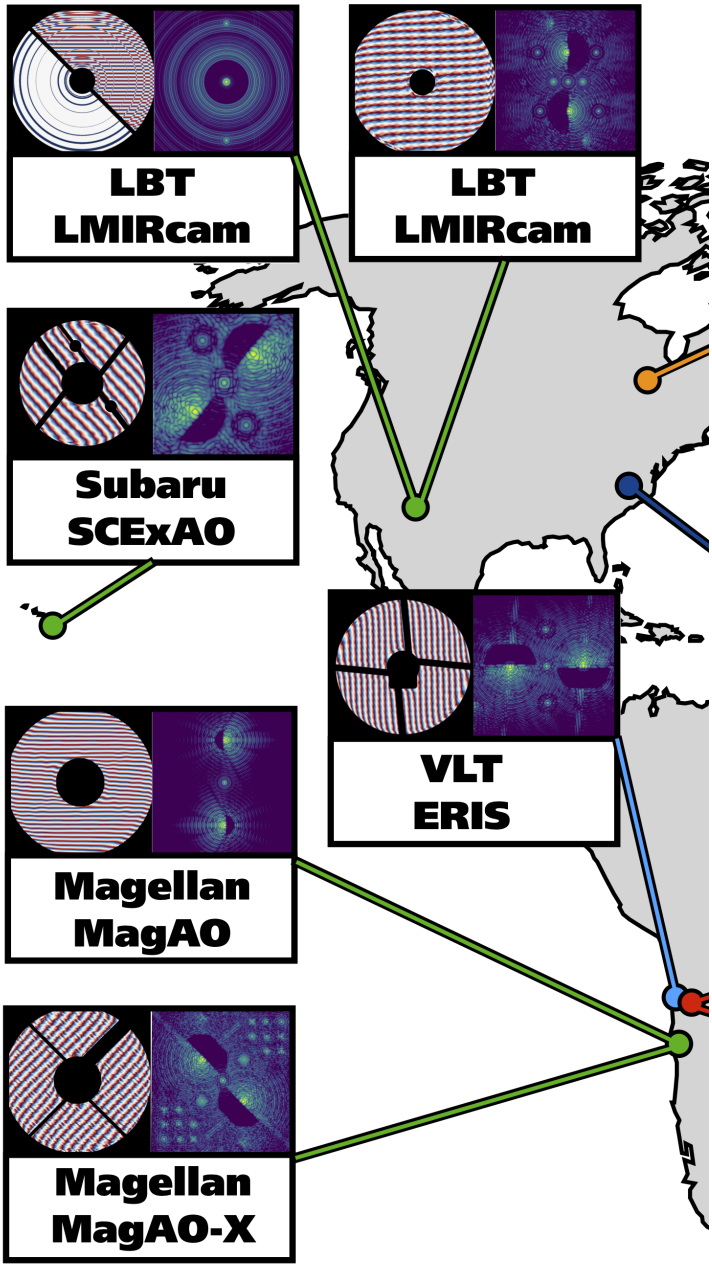
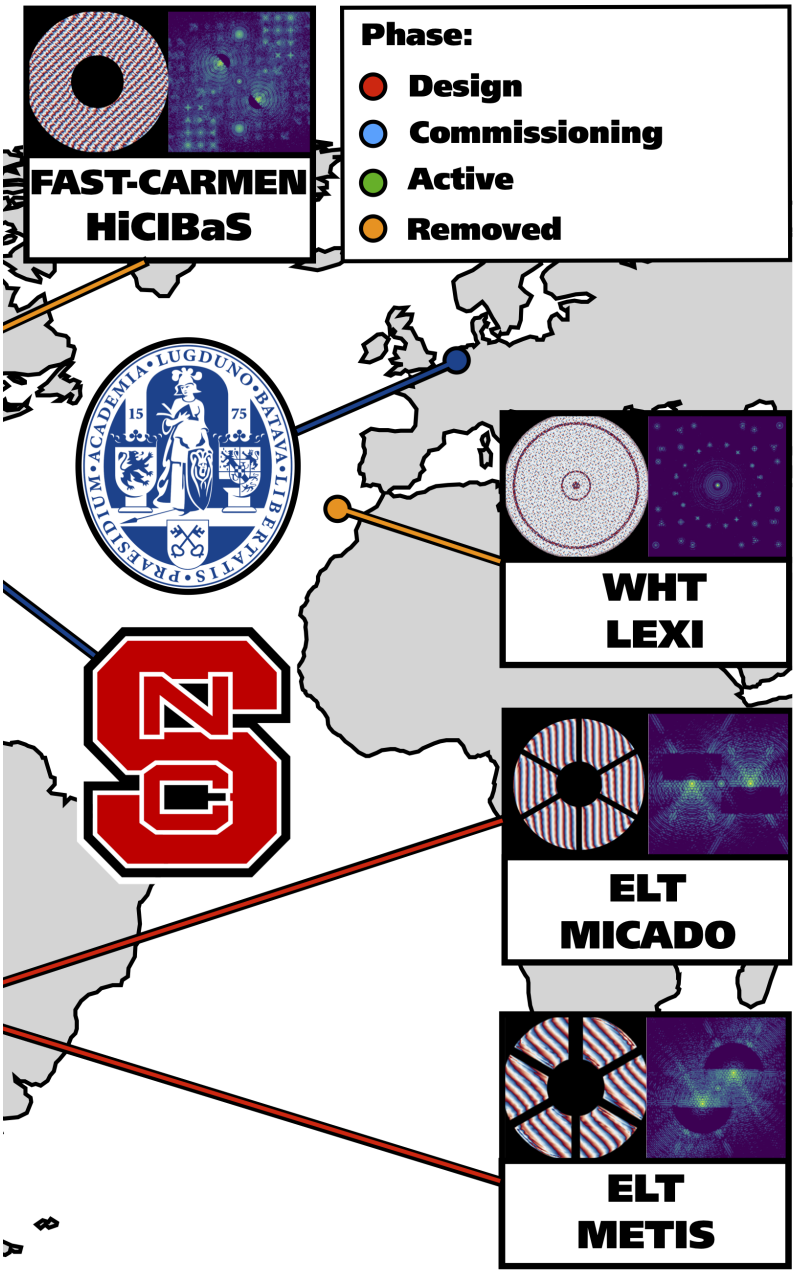


Figure 2.12: Overview and status of the vAPP coronagraphs around the world.


Phase:
Design
Commissioning
Active
Removed
FAST-CARMEN
HiCIBaS
WHT
LEXI
ELT
MICADO
ELT
METIS

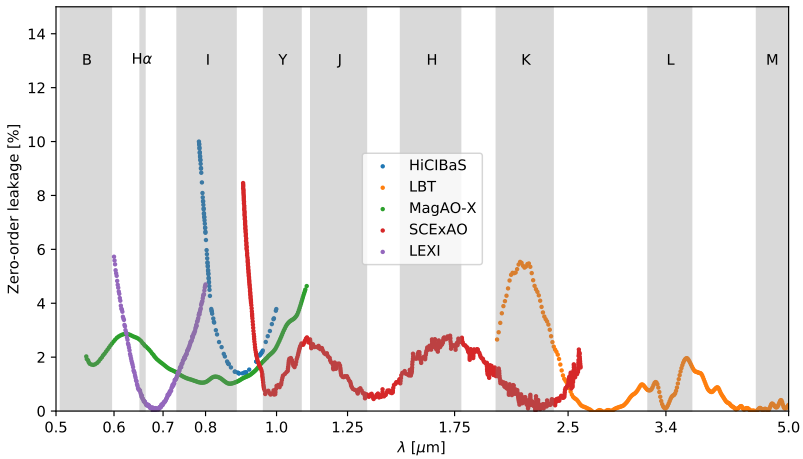


Figure 2.13: Zero-order leakage as a function of wavelength for different liquid-crystal recipes, colored by instrument.

In general, if the bandwidth is between $15 - 40\%$, a double-layered recipe, i.e. a 2TR, can have a polarization leakage $< 3\%$ for the full bandwidth. Similar performance can be reached with a 3TR for bandwidths up to 100%.
When a specific design is chosen, the manufacturing process is tuned to generate a liquid-crystal film with a measured retardance and polarization leakage that mimics the theoretical as closely as possible. The manufactured test films are polarization gratings (PGs), such that it is easy to measure the zero-order leakage using transmission measurements. When a recipe is perfected, the vAPP is manufactured using this recipe. Measuring the zero-order leakage of the final optic can be cumbersome with the small angular separations between the main beams and the leakage, caused by the large grating periods of a gvAPP. Therefore, the leakage measurement of a representative PG is used. These polarization leakages of five different recipes, representative of 5 different instruments and wavelength ranges, is shown in Fig. 2.13. All presented recipes have $< 6\%$ leakage over their required bandwidth. Together, they span the almost the full visible and near-infrared (0.55-5 μm). Interestingly, we currently cover this range with three standardized recipes that have good performance in their respective bandwidths. The same data is presented in log-scale in Fig. 2.14. We added the estimated performance of the double-grating vAPP that is installed in the LBT, by multiplying the single-grating curve with itself. We note that no direct measurement of its performance exists.
While the leakage measurements are representative, the actual vAPPs can have different performances. Lab conditions during manufacturing, like temperature and humidity, can generate differences between liquid-crystal films for identical

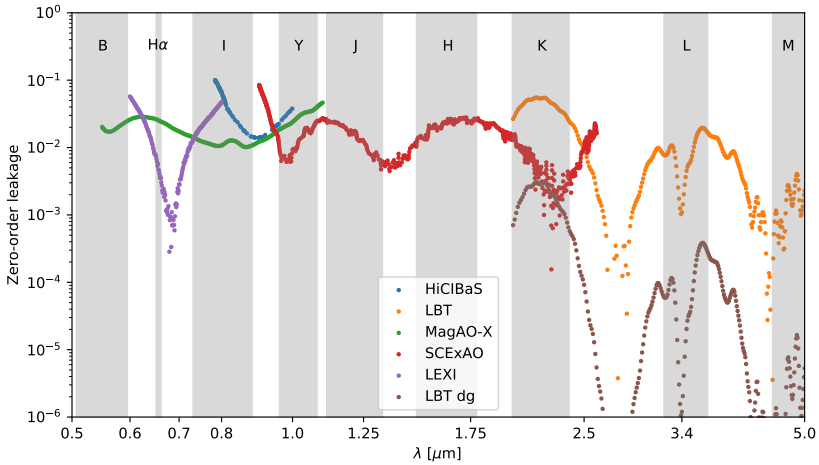


Figure 2.14: Same as Fig. 2.13, but on log-scale and an additional estimate of the LBT double-grating vAPP.

recipes. Characterizing the actual polarization leakage of a vAPP can then be done after installation or in a high-contrast imaging setup. Moreover, all vAPPs with identical wavelength coverage will have different polarization leakages as function of wavelength. These differences can be up to 2% for a specific wavelength, but the impact on the average is minimal. Most likely, it is merely a shift of the maxima and minima as function of wavelength.

2.4.2 The vAPP: individual properties

A large diversity exists between vAPPs, in designs, design goals, additional functionality, and the instruments that use them. We will provide a summary of design trade-offs, unique properties, historical details, notable results, and references to more detailed papers of each vAPP.

Magellan/MagAO/Clio

The first operational broadband gvAPP was installed in 2015 in the MagAO/Clio instrument (Close et al., 2010; Morzinski et al., 2014) at the 6.5-m Magellan Clay telescope. The gvAPP has a dark zone between 2-7 λ/D and operates between 2 to 5 micron, with a polarization leakage less than 3% over this bandwidth. A detailed description of the design, manufacturing and on-sky performance can be found in Otten et al. (2017). Being the first gvAPP to be commissioned, it was a technology demonstrator with the additional goal to use it as a science-grade component. To demonstrate the performance, Otten et al. observed a bright A-

type star with an L'-band magnitude between 0 and 1. Characterization of the leakage term and the PSF showed that the gvAPP was performing as expected. A new method of data reduction, subtracting the opposite rotated coronagraphic PSF, yielded an improvement in contrast of 1.46 magnitudes, i.e. a factor 3.8, at 3.5λ/D. The performance of the gvAPP for this observation was better than the best reported performance in literature of coronagraphs at other systems at this small angle. We note that the method of calculating the noise is necessarily different between the contrast curves reported in literature, as noted in Jensen-Clem et al. (2017). The method of noise calculation in Otten et al. (2017) does not take into account azimuthally correlated speckle noise caused by non-common path aberrations, pointing jitter, and thermal variations. For a gvAPP it is especially difficult to properly estimate the speckle noise at the smallest separations due to the limited amount of photometric apertures that fit in the D-shaped dark zones. A better comparison between this gvAPP and other coronagraphs would be to use the same detection limit methods that are commonly used, e.g. fake planet injection.
So far, the scientific yield of the MagAO/Clio gvAPP is limited, with only a marginal detection of a brown dwarf companion of HR 2562 (Sutlieff et al., 2019, Submitted for publication). This can be explained by the total awarded observing time (<10 nights), which is insignificant in comparison to the SPHERE and GPI surveys (>300 nights) that detected on the order of 10 sub-stellar companions (Nielsen et al., 2019; Vigan et al., 2020). Moreover, observing with a gvAPP can be challenging, especially when using broadband filters, see section 2.3.

WHT/LEXI

The Leiden Exoplanet Instrument (LEXI) was a visiting instrument at the William Herschel Telescope (WHT) (Haffert et al., 2016, 2018). The goal of the LEXI instrument was to develop and test new technology for new high-contrast imaging techniques, e.g. the cMWS and high-contrast, high-dispersion integral field spectroscopy (Wilby et al., 2016, 2017; Haffert et al., 2018). The instrument changed between different visits, changing the pupil (4m on-axis, 1m off-axis), the wavefront sensor, imaging modes, spectrographs and vAPP designs. The first observation run in 2016 employed a spatial light modulator with an APP + cMWS design. This test succeeded, making it the first demonstration the focal plane wavefront sensing capabilities of the cMWS on-sky. A second run in 2017 used a simplified instrument, replacing the SLM for a vAPP. Two instrument properties were taken into account for the design of the vAPP. First, the ALPAO deformable mirror with 97 actuators, which has a limited control radius. Second, the AO-loop speed was limited at 800 Hz by the camera frame rate. Thus, the vAPP designs were optimized for 10^{-4} contrast between 2.5-6 λ/D, resulting in a phase pattern 0-π concentric rings. A grating mask used to define the aperture with the vAPP. With a narrow science band between 600 nm and 800 nm, the vAPP is a 1TR, resulting in a polarization leakage less than 6%. The intensity of the leakage at the IWA

is lower than the design contrast, which enabled the use of a 360 vAPP without gratings. The vAPP was multiplexed with 20 Disk Harmonic holograms, resulting in an improved wavefront correction (Haffert et al., 2018).

Subaru/SCExAO/CHARIS

The Subaru Coronagraphic Extreme Adaptive Optics (SCExAO) (Jovanovic et al., 2015) instrument is an ever-evolving high-contrast imaging system at the 8.2-m Subaru Telescope, with the aim of providing a platform for technology demonstration for next generation telescopes. A more recent, and already outdated, overview of the instrument can be found in (Lozi et al., 2020). The gvAPP in SCExAO was installed in 2017 in the near-infrared arm (0.9-2.5 μm) (Doelman et al., 2017). The design of the gvAPP has a reduced grating frequency compared to previous designs to make the PSF fit in the SCExAO/CHARIS field of view (FOV) for the full bandwidth (Doelman et al., 2017). CHARIS is an integral field spectrograph with a resolution of $R \sim 18.4$ for a bandwidth spanning the J, H, and K bands (1.13-2.39 μm) and a FOV of 2" $\times$ 2" (Groff et al., 2017). To fit these instrument specifications, the gvAPP was manufactured with a new liquid-crystal recipe with a diffraction efficiency greater than 97% between 1-2.5 μm. Two additional features of this gvAPP are the grating mask that is undersized with respect to the amplitude mask of SCExAO and defines the gvAPP pupil, and the first implementation of phase diversity spots. At the inner working angle of 2.1 λ/D, the design contrast is 10^{-5}, and the outer working angle is 11 λ/D, which is limited by the CHARIS FOV in K-band.
The gvAPP was installed together with a patterned ND-filter in the focal plane, which reduces the coronagraphic and leakage PSF intensity by a factor 10-50, depending on wavelength. An image of the PSF with the patterned ND-filter is shown in Fig. 2.15. Such a focal plane mask enables longer integration times and reduces the risk of saturation and resulting persistence of the sensitive NIR cameras. The PD holograms are not blocked and can also be used as a photometric reference. After installation, the first-light images of the gvAPP with CHARIS showed a PSF shift by ~ 1 arcsecond caused by the wedge of the optic. A second wedged optic that is identical to the one used in the gvAPP was installed in a filter wheel downstream to successfully correct this shift.
Testing with the internal source showed promising results with a raw contrast $< 10^{-3}$ at the inner working angle, and $< 10^{-4}$ beyond 4.5 λ/D (Bos et al., 2019), demonstrating that the delivered gvAPP is within specifications and commensurate with the performance of the system and other coronagraphs. However, on-sky testing resulted in raw-contrasts that were a factor 5-10 worse, caused by prominent diffraction structures present in the dark zones. These structures are visible in Fig. 2.15 as bright lines of speckles orthogonal to the flat side of the dark zone. These structures are not present when using any other coronagraph present in SCExAO. Inspection of pupil images with and without the gvAPP showed that these structures are not caused by pupil misalignment. In addition, the absence of these structures in the internal source data suggests that they are caused by optics

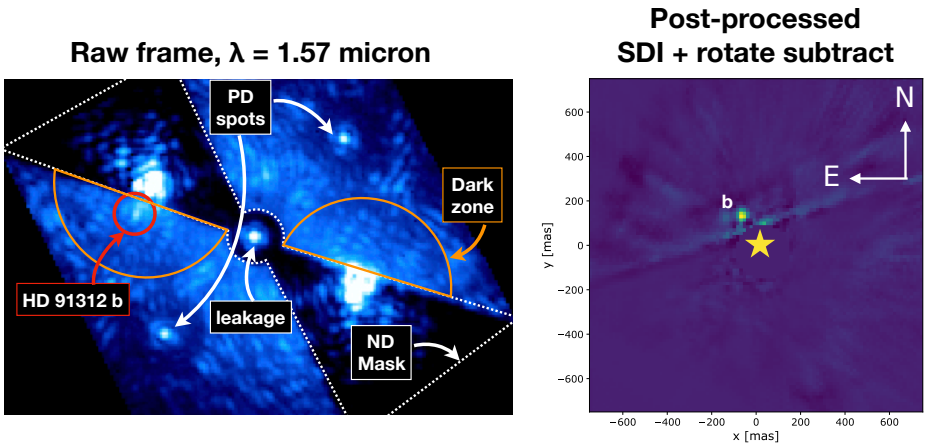


Figure 2.15: Observations of HD 91312 with the gvAPP with CHARIS. *Left:* Image slice from the extracted data cube showing the raw PSF. Right: Final image of HD 91312, showing the companion clearly. The bright streak is a result of post-processing. The scale of both images is different.

upstream of SCExAO, e.g. the telescope mirrors or AO188, an upstream adaptive optics instrument. Simulations show that annular phase offsets at the pupil edges can reproduce similar dark zone structures, which would not be noticed by other coronagraphs when a Lyot stop is used. An annular phase structure can be seen in the on-sky wavefront reconstruction in (Bos et al., 2019) and it also has been observed with the pyramid wavefront sensor, but so far any attempt to correct these phase offsets did not result in a significant improvement of the dark zone contrast.

Despite the lower-than-expected raw-contrast, on-sky observations have been carried out with both the imager SCExAO/Chuckcam and the IFU SCExAO/CHARIS. One of the first tests of the gvAPP was the observation of HD 91312 with CHARIS on 7 February 2018, for a total of integration time of 5 minutes 29 seconds. HD 91312 is a good candidate as it is a binary system, with a companion at a separation of 150 mas ($\sim$ 5 AU), which has a minimal mass of 80 M_{Jup} (Borgniet et al., 2019), and therefore is most likely a M-dwarf. Because HD 91312 A is an A-star, the relative contrast to HD 91312 b in the near infrared is already favourable and becomes lower towards longer wavelengths. The companion was detected in the raw frames, even with the diffraction structures present. Post-processing with spectral differential imaging (SDI) (Racine et al., 1999) and rotate subtract (Otten et al., 2017) removed the diffraction structures. The extracted PSF at $\lambda = 1.57\mu$m and the final post-processed image are shown in Fig. 2.15. While the on-sky raw contrast is limited by the uncorrected diffraction structures, the SCExAO gvAPP is still valuable as a technology demonstration of focal-plane wavefront sensing. The phase diversity between both coronagraphic PSFs, combined with the asym-

metric pupil of SCExAO, resulted in the first demonstration of FPWFS with only the coronagraphic PSFs (Bos et al., 2019). The improvement of a factor two in raw contrast between 2-4 λ/D is similar to other FPWFS techniques (Galicher et al., 2019). An additional improvement can be realised by implementing LDFC (Miller et al., 2019, 2020; Bos et al., Submitted for publication). However, it is unclear if the combination of both FPWFS techniques will resolve the issue with diffraction structures in the on-sky PSF. A future upgrade of AO188 could resolve the issue and enhance the gvAPP performance.

FAST-CARMEN/HiCIBaS

As is mentioned in the name, the High-Contrast Imaging Balloon System (HiCIBaS) is a balloon-borne telescope and is installed on the FAST-CARMEN gondola (Côté et al., 2018). The main goals of this mission were to test new high-contrast imaging technology in a space-like environment and the training of five master students of Laval university (Thibault et al., 2019). The scientific goal was to measure and characterize the wavefront distortions from high-altitude (>40km) turbulence, using low-order wavefront sensors (LOWFS) (Allain et al., 2018). The balloon launched August 25 2018 at 11:18 pm local time from the Timmins Stratospheric Balloon Base in Ontario, Canada. The flight plan was to ascend to 40 km, and stay there for the duration of $\sim$ 12 hours. A separate arm was created for high-contrast imaging. This arm would benefit from the pointing stabilization necessary for the LOWFS measurements. The 37-element IRIS AO MEMS deformable mirror was installed to correct wavefront aberrations, however, it was not connected to the LOWFS to reduce complexity. The gvAPP was selected because the coronagraph is stable against residual pointing inaccuracies. The design has a small dark zone from 2.1 to 8.5 λ/D and a contrast of 10^{-6}, delivering more extreme performance in the low-turbulance regime. With a limited bandwidth between 0.83 and 0.88 μm, a 1TR gvAPP was sufficient. Also, the gvAPP was designed to provide complementary wavefront sensor information using a 12-mode Zernike basis cMWS or PD holograms. The performance of the LOWFS and the gvAPP FPWFS were successfully tested close-loop in the system before the flight. The PSF before the flight is shown in Fig. 2.16. An internal laser source was used to show that both the cMWS and the PD holograms worked.
While the launch and flight happened according to the flight plan, an unfortunate mechanical failure resulted in the inability to operate the telescope. As a result, the telescope never pointed at a celestial source. An internal source was present in the flight instrument case of such a failure, which was used to record the gvAPP PSFs. It was also discovered that the deformable mirror did not respond to commands in flight and was stuck in a random, non-flattened state. Extreme noise from electronic crosstalk resulted in low-quality images of the PSF. The resulting images are not useful to characterize the gvAPP performance during flight, see Fig. 2.16. The balloon and payload were recovered after the flight. The high-contrast imaging system, including the gvAPP, were still operational. Post-launch images of the gvAPP PSF do not show any significant deviation from pre-launch

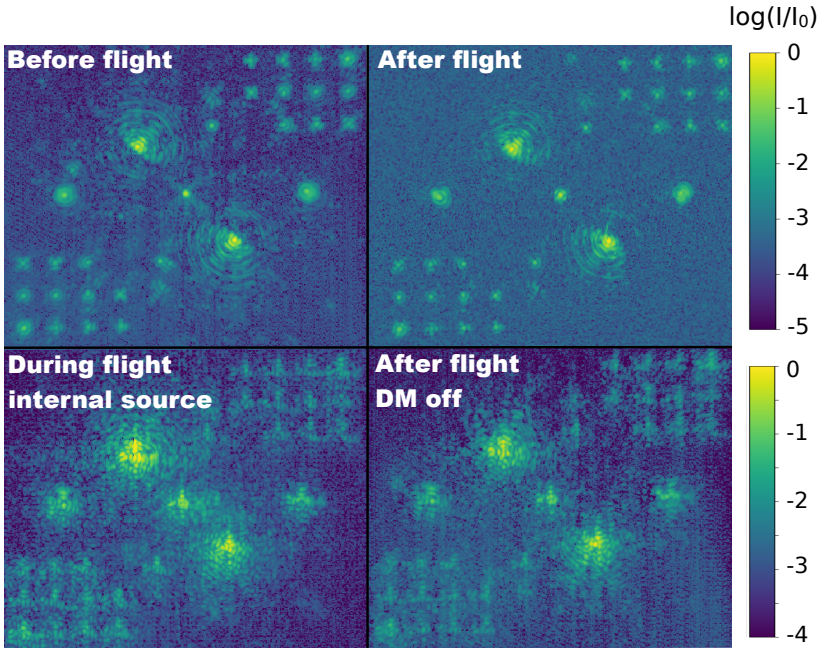


Figure 2.16: Measurements of the HiCIBaS gvAPP PSF ($\log_{10}$ scale) before, during, and after the flight using the internal source of the HiCIBaS instrument. No optical component of HiCIBaS sustained major damage during the landing. The system was realigned after the flight.

performance, and the flight PSF is similar to the after-flight PSF with the DM off. We note that both comparisons are qualitative, and have not been fully characterized. We conclude that the gvAPP or similar LC coronagraph can operate in a space-like environment.

Magellan/MagAO-X

The MagAO-X instrument is a completely different instrument than MagAO, operating in the visible and near-infrared with an extreme AO (ExAO) system (Males et al., 2018). The key science goal of MagAO-X is high-contrast imaging of accreting protoplanets at Hα (656.28 nm). To achieve this goal, MagAO-X will deliver high Strehls (>70% at Hα), and high contrasts ($< 10^{-4}$) from ~ 1 to 10 λ/D. In addition, MagAO-X implements many of the lessons-learned from SCExAO, including the real-time controller CACAO[3]. The first phase (Phase I) aims to fully characterize and calibrate the instrument while obtaining the first science results. With non-optimal wavefront correction in phase I, the gvAPP was chosen as a robust coronagraph. To accommodate wavefront sensing and good coronagraphic performance, the Strehl ratio of the design is lower than most other coronagraphs (40.7%). The coronagraph design creates a dark zone from 2.1-15 λ/D, with 10^{-5} contrast at the IWA, and the dark zones are elongated in the grating direction. Therefore, wavelength smearing up to 20% does not compromise the dark zone towards the edges.
The gvAPP pupil was specifically altered with a thicker spider to be asymmetric with the goal of enabling focal-plane WFS using the coronagraphic PSFs. As a result, the gvAPP design employs three ways of wavefront sensing, i.e. PD, cMWS, and AP-WFS, in addition to LDFC for wavefront control. Each of these wavefront sensing modes have been tested separately using a test plate with different gvAPP designs (Miller, 2018; Miller et al., 2019). The wavefront sensing capabilities enable better instrument calibration by accurately measuring non-common path aberrations (NCPA), which are more severe in the visible. MagAO-X employs a separate DM that is dedicated to correct NCPA, which will be fed directly by the focal-plane wavefront sensors of the gvAPP. FPWFS can be done at much greater speeds with MagAO-X because a laser-cut mask in the intermediate focal plane separates the dark zones from the bright field. This mask transmits the dark holes to the EMCCD science camera and reflects the stellar bright field back to a second EMCCD that will act as the dedicated FPWFS. The cMWS was especially useful for alignment of the off-axis paraboloids, providing direct visual feedback on the first 9 low-order Zernike modes. For on-sky operations, a likely strategy would be to use PD, cMWS and AP-WFS to minimize the non-common path aberrations before observing and use LDFC to remove any drifts.
MagAO-X had first light in 2019, where it demonstrated a Strehl ratio of 46%, and a 5σ post-processed gvAPP contrast of 10^{-4} at 3 λ/D at 900 nm (z') (Males et al., 2020). These results are promising for discovering exoplanets during formation through direct detection of $H\alpha$ from accretion of hydrogen onto these protoplan-

[3] `https://github.com/cacao-org`

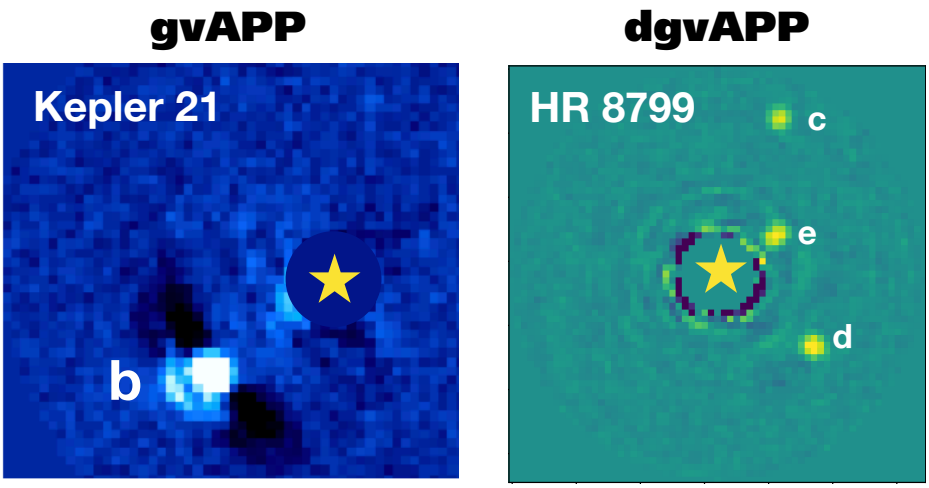


Figure 2.17: Detection of Kepler 21B using the gvAPP at LBT/LMIRcam ($\Delta m =$ 3.9 and the inner three HR 8799 planets ($\Delta m =\sim$ 10) using the dgvAPP at LMIRcam/ALES.

ets. Applying the Massive Accreting Gap (MAG) protoplanet model(Close, 2020) to the observed first light MagAO-X contrasts results in a predicted maximum yield of 46±7 planets from 19 stars(Close et al., 2020).

LBT/LMIRcam/ALES

The Large Binocular Telescope Mid-Infrared camera (LMIRcam) is an imager/spectrograph that operates behind the Large Binocular Telescope Interferometer (LBTI) (Skrutskie et al., 2010). The LBT is a unique telescope with two 8.4-m primary mirrors mounted next to each other. Both apertures can be combined with LBTI and create fringes on LMIRcam, resulting in a 22.8-m baseline. The LBT and LMIRcam are ideal for L and M band observations with a minimal number of warm optics and a deformable secondary mirror. This makes LBTI/LMIRCAM the most sensitive infrared instrument in the Northern hemisphere. We take advantage of this opportunity with two different vAPPs, a grating vAPP (1.8-15 λ/D) and a double-grating vAPP (2.7-15 λ/D). The designs are pushing the limits in inner working angle, and the gvAPP also incorporates focal-plane wavefront sensing with PD holograms and photometric/astrometric reference spots. Their more extreme designs are enabled by the small secondary and the extremely thin secondary support structure. The optic contains two vAPPs, one for each aperture. Their beams can be imaged separately, partially combined with overlapping fringing holograms, or completely overlapped to produce fringes for nulling.
LMIRCam was augmented in 2015 with the Arizona Lenslets for Exoplanet Spectroscopy (ALES) (Skemer et al., 2015), an integral field spectrometer (IFS). After

a recent upgrade, ALES facilitates 2 to 5 μm low-resolution (R$\sim$35) spectroscopy of directly imaged gas-giant exoplanets (Skemer et al., 2018). Currently, it is the only IFS operating in the L and M bands, containing broad molecular features (CO, CH_4), and also covers the 3.1 μm ice feature, 3.3 μm PAH feature, and 4.1 μm Br-alpha emission line. Therefore, low-resolution spectroscopy can provide a unique window for gas giants in the L-T transition, where their atmospheres transition from CO-rich and cloudy (L-type), to CH_4-rich and relatively cloud free (T-type).
Both vAPPs installed in LMIRcam are available for ALES, and, together with the annular groove phase mask (AGPM) coronagraph, they are the only coronagraphs installed in LMIRcam/ALES that cover the full wavelength range. Both vAPPs do have significant absorption (up to 100%) between 3.15 and 3.45 μm from glue and the liquid-crystals. The outer working angle of both coronagraphs is larger than the FOV of ALES for wavelengths larger than $\sim$ 3 μm. For the gvAPP only one of the two coronagraphic PSFs can be fitted inside the FOV, and the wavelength smearing moves it across the detector. While this is not ideal, the gvAPP provides a much smaller inner working angle compared to the dgvAPP. The dgvAPP provides a stable central PSF and is more sensitive than the gvAPP due to the recombination of both coronagraphic PSFs. Moreover, as it provides a 360 degree dark zone it is much more user friendly and intuitive.
Tests with the internal source show that both vAPPs are operating as expected. Only a few observations have been carried out, but already show the great promise of these vAPPs, see Fig. 2.17. The results on the HR 8799 system will be discussed in more detail in a future paper. The unique feature of the vAPPs is the presence of the stellar PSF on the detector. This is particularly interesting for photometric monitoring of exoplanets. For example, monitoring the stellar PSF can be used to reduce the impact of variability in AO performance during the observing run. In addition, an upgraded data reduction technique has been developed to remove the thermal background with great accuracy using the science frames themselves. Nodding the telescope to characterize the thermal background with off-source measurements is still necessary, however at much lower cadence, increasing the on-source time. To summarize, there are still many opportunities with the vAPPs in LMIRcam/ALES.

VLT/ERIS

The Enhanced Resolution Imager and Spectrograph (ERIS) will be installed in the Cassegrain focus of UT4 at the Very Large Telescope in the near future (Amico et al., 2012; Davies et al., 2018). ERIS is a 1-5 μm instrument with an imager (NIX) operating from J-M_p band and an IFS (SPIFFIER) in J-K band with R$\sim$8000. ERIS/NIX will have a vortex coronagraph and a gvAPP, while ERIS/SPIFFIER is non-coronagraphic. The ERIS gvAPP has a dark zone from 2.2-15 λ/D, which has been widened to accommodate wavelength smearing, and the APP phase is multiplexed with two photometric/astrometric reference spots (Kenworthy et al., 2018b). The pupil has an asymmetric central obscuration because the (warm)

tertiary mirror (M3) will be moved aside to illuminate the Cassegrain focus where ERIS is residing. While the asymmetry is small, it is still possible to estimate NCPA with FPWFS (Bos et al., 2019) if the software is implemented. The gvAPP has been manufactured in 2018, and was later tested in the near-infrared test bench of the group for Exoplanet & Habitability at ETH Zurich (Boehle et al., 2018). The setup imaged the gvAPP PSF at 3.8 μm and was limited by background noise to $\sim 10^{-3}$. Since then an updated setup has been realized and the gvAPP was tested to the $\sim 10^{-5}$ level (Boehle et al., Submitted for publication). These measurements demonstrate the ERIS/gvAPP operates as expected, within the limits of the test bench.
While ERIS/NIX is an imager, it will have a long-slit spectroscopic mode with R$\sim$850, covering the full L band. With the gvAPP absorption feature between 3.15-3.45 μm, direct detection of methane will be more difficult. For atmospheres with elevated methane levels, e.g. T-type brown dwarfs, the spectral slope beyond 3.5 μm can be used to measure methane levels. Another unique opportunity of ERIS/NIX is enabled by the combination of a sensitive wavefront sensor and the powerful (20W) laser guide star. The laser guide star allows observations of much fainter science targets, which will greatly increase the number of potential targets accessible to this new instrument.

ELT/METIS and ELT/MICADO

The Extremely Large Telescope (ELT) (Gilmozzi & Spyromilio, 2007) will revolutionize direct imaging of exoplanets because of its primary mirror diameter of 39 meters. The supreme resolution will be fully exploited by first generation instruments. With five times the resolution of the VLT and 25 times the sensitivity, the ELT will provide new insights in planet formation, protoplanetary disks, planet-disk interaction, planetary atmospheres and planet evolution. Both the Mid-infrared ELT Imager and Spectrograph (METIS) (Brandl et al., 2016) and the Multi-Adaptive Optics Imaging CamerA for Deep Observations (MICADO) (Davies et al., 2016) have exoplanet detection and characterization as one of the main science goals. To this end, both instruments have a gvAPP baselined.
METIS covers the science bands between 3-14 μm, and has a high-resolution IFS mode with $R \sim 100,000$ in L and M band, in addition to an imaging mode from L to N band. This high spectral resolution is achieved over a large FOV of 0.5 by 1.0 arcseconds, which is made possible by an image slicer that divides the field of view into multiple strips on a slicer mirror. Simulations show that METIS will detect and characterize multiple planets found with radial-velocity studies (Quanz et al., 2015). Two gvAPPs with different dark zone designs will be installed in the IFS and the imager in L and M band (Kenworthy et al., 2018a). A specific downside of the gvAPP is that the star light is not blocked, leading to high intensities in the IFS focal plane. This can be prevented by changing the position of the star in the field of view using the chopper. Due to this offset only the dark zone of the gvAPP PSF is imaged onto the image slicer. The ELT pupil is not favourable for coronagraph design, with the large central obscuration. The METIS pupil is

a cold stop and has additional thick spiders, complicating the design process further. The METIS gvAPP design therefore has a larger inner working angle of 2.5 λ/D to keep the Strehl above 65%. Moreover, the gvAPP has a more moderate contrast towards the inner working angle, going down from a few times 10^{-4} at 3 λ/D to 10^{-6} beyond 6 λ/D. Simulations show that this does not affect the performance of the coronagraph compared to other baselined coronagraphs like the AGPM, as NCPA, residual wavefront aberrations, and missing mirror segments of the primary will limit the performance at the smallest separations. In addition, the AGPM is quite sensitive to tip-tilt which can be significant for nodding and high frequency structural vibrations (Glück et al., 2016).
MICADO, the European ELT first-light imager (Clénet et al., 2018a), covers the science bands between 0.8-2.4 μm and will benefit from both the M4 single-conjugate AO system (Clénet et al., 2018b) and the dedicated multi-conjugate AO system instrument (MAORY) (Diolaiti et al., 2016). With diffraction limited performance, MICADO will deliver extremely high-resolution imagery to study the inner regions of planetary systems (Baudoz et al., 2019). The high contrast imaging modes will comprise two classical Lyot coronagraphs (Perrot et al., 2018), one gvAPP and two non redundant masks (Lacour et al., 2014). One challenge in designing the high contrast imaging modes of MICADO comes from the fact that they have to be limited to one focal plane upstream the atmospheric dispersion corrector, and one downstream pupil plane. While the atmospheric dispersion might have an impact on the performance of the smallest Lyot coronagraph in broadband, gvAPP is not sensitive to this effect and will thus offer high performance even at high airmass. The MICADO gvAPP design is optimized between 2.5 λ/D and 20 λ/D with a square shape. On-sky expected contrast performance reaches 2×10^{-5} at 3.5 λ/D in K band with a Strehl ratio of 70%. For bright nearby stars, the starlight could saturate the detector for the shortest exposure time, and the use of a neutral density filter might be necessary.

2.5 Current status and future developments

With the vAPP successfully installed in many different telescopes, the focus for the vAPP developments is now transitioning from testing and commissioning the technology, to science observations. Much of the time spent on vAPPs so far has been oriented towards technology development and design optimization. However, the on-sky technology demonstration, observation planning, and data reduction are critical to achieve competitive science results.
The majority of the current on-sky data with a vAPP have been taken with the MagAO, LEXI and SCExAO instruments. Bright targets were selected to provide optimal AO correction for technology demonstration of the vAPP and FPWFS. While the technology demonstration was often successful and benefited from test observations and consequent understanding and iterations, time awarded to science observations has been limited, and conditions for these observations have been sub-optimal. Moreover, attempts on more challenging targets, e.g. HD

206893 with SCExAO, resulted in non-detections. We tried bridging this gap by pursuing targets with intermediate contrast by looking into stellar multiplicity and to understand the detection limits. This science case was awarded a few nights at the LBT and Subaru, most of which were lost due to bad weather. Currently the vAPPs of the LBT have yielded the best results, recovering HR 8799 c,d,e.

To fully exploit the opportunities of the ERIS, MICADO and METIS gvAPPs, getting more experience with on-sky observations is paramount. We note that these instruments have had a gvAPP baselined in the instrument design, and have dedicated engineering time planned to resolve any issues. Moreover, of all the vAPPs that are currently installed, only the SCExAO gvAPP did not reach the expected performance on-sky. The diffraction structures in the SCExAO dark zones are not caused by the gvAPP itself because the lab results do not show these structures. So, it is likely caused by uncorrected aberrations from the optics upstream of the internal source injection, which is unlikely for the ERIS instrument that is installed in the Cassegrain focus. For METIS and MICADO the vAPPs will be tested extensively and problems should be identified during this process. We conclude that future efforts should focus on observation planning and a standardized vAPP data reduction pipeline.

For both observation planning and a data reduction pipeline a preliminary version already exists, as discussed in Section 2.3. The vAPP observation preparation tool does include most vAPPs that are currently available, however, it does not provide a proper indication whether a companion will be in the dark zone. In future upgrades we will add this functionality, in addition to showing the field rotation of the companion in the dark zones of on-sky PSFs. Ultimately, this could be used to calculate which days of a semester are optimal for observations of known objects to help proposal writing.

The data reduction pipelines are currently all custom for each vAPP. A unification of these pipelines would centralize all efforts on vAPP data reduction and provide a standardized pipeline that can be used by anyone with vAPP data. In addition, it would increase efficiency of implementing new algorithms like rotate-subtract or TRAP. To minimize duplication of previous efforts, it is essential that the pipeline can also port to existing post-processing pipelines like the vortex image processing package (`VIP`) (Gonzalez et al., 2017) or `PynPoint` (Stolker et al., 2019). Lastly, these pipelines should include methods to use capitalize on the astrometric and photometric holographic PSFs.

Future technical developments should be towards enabling broadband imaging with the vAPP. The main challenge to overcome is a (super)achromatic QWP that adds minimal wavefront aberration and ghosts to the system (Bos et al., 2018). By adding another QWP before the vAPP, it is also possible to combine the vAPP with imaging polarimetry (Snik et al., 2014; Bos et al., 2018), another powerful tool to distinguish astronomical signals from speckle noise (Kuhn et al., 2001). Adding multiple coronagraphic PSFs per polarization state can then simplify more advanced polarimetric vAPP designs (Bos et al., 2020). Another potential interesting avenue to explore is the combination of the vAPP with multi-color holography (Doelman et al., 2019), as that would completely eliminate leakage

problems (Bos et al., 2020). However, this implementation is more difficult to implement in an instrument and puts strong requirements on the shape of the pupil.

2.6 Conclusion

The vector-apodizing phase plate coronagraph is a versatile single-optic pupil-plane coronagraph that is easily adapted to any telescope aperture. It delivers good contrast (¡10^{-4}, limited by AO performance) at small inner working angles ($\sim 2\lambda/D$), it can host many different focal-plane wavefront sensing techniques, and can include astrometric and photometric reference PSFs. The manufacturing using direct-write for liquid-crystal technology enables extreme and accurate patterning and the multi-twist retarder technology provides excellent efficiencies ($> 96\%$) for bandwidths up to 100% between 0.55 μm and 5 μm. Because of these properties, the vAPP has now been installed in 6 different instruments. Measurements with the internal source of these instruments demonstrate that the vAPPs are operating as expected. The grating of gvAPPs and their D-shaped dark zone make observing and data reduction more complicated compared to focal-plane coronagraphs. Future efforts should focus on software for observation planning and data reduction to fully exploit the opportunities provided by future instruments that have the vAPP baselined, i.e. VLT/ERIS, ELT/METIS, and ELT/MICADO.

Bibliography

Allain, G., Brousseau, D., Thibault, S., et al. 2018, in Adaptive Optics Systems VI, Vol. 10703, International Society for Optics and Photonics, 107035T

Amara, A., & Quanz, S. P. 2012, Monthly Notices of the Royal Astronomical Society, 427, 948

Amico, P., Marchetti, E., Pedichini, F., et al. 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, Vol. 8446, International Society for Optics and Photonics, 844620

Avenhaus, H., Quanz, S. P., Garufi, A., et al. 2018, The Astrophysical Journal, 863, 44

Baudoz, P., Huby, E., & Clénet, Y. 2019, in SF2A-2019: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, ed. P. Di Matteo, O. Creevey, A. Crida, G. Kordopatis, J. Malzac, J. B. Marquette, M. N'Diaye, & O. Venot, Di

Berry, M. V. 1987, Journal of Modern Optics, 34, 1401

Bloemhof, E., Dekany, R., Troy, M., & Oppenheimer, B. 2001, The Astrophysical Journal Letters, 558, L71

Boehle, A., Doelman, D., et al. Submitted for publication, Journal of Astronomical Telescopes, Instruments, and Systems

Boehle, A., Glauser, A. M., Kenworthy, M. A., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 107023Y

Borgniet, S., Lagrange, A.-M., Meunier, N., et al. 2019, Astronomy & Astrophysics, 621, A87

Bos, S., Miller, K., Lozi, J., et al. Submitted for publication, Astronomy & Astrophysics

Bos, S. P. 2020, arXiv preprint arXiv:2005.08751

Bos, S. P., Doelman, D. S., de Boer, J., et al. 2018, in Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III, Vol. 10706, International Society for Optics and Photonics, 107065M

Bos, S. P., Doelman, D. S., Miller, K. L., & Snik, F. 2020, in Adaptive Optics Systems VII, Vol. 11448, International Society for Optics and Photonics, 114483W

Bos, S. P., Doelman, D. S., Lozi, J., et al. 2019, Astronomy & Astrophysics, 632, A48

Brandl, B. R., Agócs, T., Aitink-Kroes, G., et al. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 990820

Cantalloube, F., Farley, O. J. D., Milli, J., et al. 2020, A&A, 638, A98, doi: 10.1051/0004-6361/201937397

Carlotti, A. 2013, Astronomy & Astrophysics, 551, A10

Clénet, Y., Buey, T., Gendron, E., et al. 2018a, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, ed. L. M. Close, L. Schreiber, & D. Schmidt, 1070313, doi: 10.1117/12.2311481

Clénet, Y., Buey, T., Gendron, E., et al. 2018b, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, ed. L. M. Close, L. Schreiber, & D. Schmidt, 1070313, doi: 10.1117/12.2311481

Close, L. M. 2020, The Astronomical Journal, 160, 221

Close, L. M., Gasho, V., Kopon, D., et al. 2010, in Adaptive Optics Systems II, Vol. 7736, International Society for Optics and Photonics, 773605

Close, L. M., Males, J., Long, J. D., et al. 2020, in Adaptive Optics Systems VII, Vol. 11448, International Society for Optics and Photonics, 114480U

Codona, J., Kenworthy, M., Hinz, P. M., Angel, J. R. P., & Woolf, N. 2006, in Ground-based and Airborne Instrumentation for Astronomy, Vol. 6269, International Society for Optics and Photonics, 62691N

Codona, J. L., & Angel, R. 2004, The Astrophysical Journal Letters, 604, L117

Côté, O., Allain, G., Brousseau, D., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 1070248

Davies, R., Schubert, J., Hartl, M., et al. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 99081Z

Davies, R., Esposito, S., Schmid, H.-M., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 1070209

Diolaiti, E., Ciliegi, P., Abicca, R., et al. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 99092D

Doelman, D. S., Escuti, M. J., & Snik, F. 2019, Optical Materials Express, 9, 1246

Doelman, D. S., Por, E. H., Ruane, G., Escuti, M. J., & Snik, F. 2020, Publications of the Astronomical Society of the Pacific, 132, 045002

Doelman, D. S., Snik, F., Warriner, N. Z., & Escuti, M. J. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, ed. S. Shaklan, Vol. 10400, International Society for Optics and Photonics (SPIE), 224 – 235, doi: 10.1117/12.2273406

Doelman, D. S., Tuthill, P., Norris, B., et al. 2018, in Optical and Infrared Interferometry

and Imaging VI, Vol. 10701, International Society for Optics and Photonics, 107010T
Dong, S., Haist, T., & Osten, W. 2012, Applied optics, 51, 6268
Escuti, M. J., Kim, J., & Kudenov, M. W. 2016, Optics and Photonics News, 27, 22
Galicher, R., Baudoz, P., Delorme, J.-R., et al. 2019, Astronomy & Astrophysics, 631, A143
Gerchberg, R. W. 1972, Optik, 35, 237
Gilmozzi, R., & Spyromilio, J. 2007, The Messenger, 127, 3
Glück, M., Pott, J.-U., & Sawodny, O. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 99093N
Gonsalves, R. A. 1982, Optical Engineering, 21, 215829
Gonzalez, C. A. G., Wertz, O., Absil, O., et al. 2017, The Astronomical Journal, 154, 7
Groff, T., Chilcote, J., Brandt, T., et al. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, Vol. 10400, International Society for Optics and Photonics, 1040016
Gurobi Optimization, L. 2020, Gurobi Optimizer Reference Manual. `http://www.gurobi.com`
Haffert, S., Wilby, M., Keller, C., & Snellen, I. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 990867
Haffert, S., Wilby, M., Keller, C., et al. 2018, in Adaptive Optics Systems VI, Vol. 10703, International Society for Optics and Photonics, 1070323
Hornburg, K. J., Komanduri, R. K., & Escuti, M. J. 2014, in Polarization: Measurement, Analysis, and Remote Sensing XI, Vol. 9099, International Society for Optics and Photonics, 90990Z
Hornburg, K. J., Komanduri, R. K., & Escuti, M. J. 2019, JOSA B, 36, D28
Jensen-Clem, R., Mawet, D., Gonzalez, C. A. G., et al. 2017, The Astronomical Journal, 155, 19
Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, Publications of the Astronomical Society of the Pacific, 127, 890
Keller, C. U. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, Vol. 9908, International Society for Optics and Photonics, 99089V
Kenworthy, M. A., Absil, O., Carlomagno, B., et al. 2018a, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 10702A3
Kenworthy, M. A., Codona, J. L., Hinz, P. M., et al. 2007, The Astrophysical Journal, 660, 762
Kenworthy, M. A., Snik, F., Keller, C. U., et al. 2018b, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 1070246
Komanduri, R. K., Kim, J., Lawler, K. F., & Escuti, M. J. 2012, in Emerging Liquid Crystal Technologies VII, Vol. 8279, International Society for Optics and Photonics, 82790E
Komanduri, R. K., Lawler, K. F., & Escuti, M. J. 2013, Optics Express, 21, 404
Kostinski, A. B., & Yang, W. 2005, Journal of Modern Optics, 52, 2467
Kuhn, J., Potter, D., & Parise, B. 2001, The Astrophysical Journal Letters, 553, L189
Lacour, S., Baudoz, P., Gendron, E., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, ed. S. K. Ramsay, I. S. McLean, & H. Takami, 91479F, doi: `10.1117/12.2057070`

Lozi, J., Guyon, O., Vievard, S., et al. 2020, in Adaptive Optics Systems VII, Vol. 11448, International Society for Optics and Photonics, 114480N

Lyot, B. 1939, Monthly Notices of the Royal Astronomical Society, 99, 580

Males, J. R., Close, L. M., Miller, K., et al. 2018, in Adaptive Optics Systems VI, Vol. 10703, International Society for Optics and Photonics, 1070309

Males, J. R., Close, L. M., Guyon, O., et al. 2020, in Adaptive Optics Systems VII, Vol. 11448, International Society for Optics and Photonics, 114484L

Marois, C., Lafreniere, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, The Astrophysical Journal, 641, 556

Martinache, F. 2013, Publications of the Astronomical Society of the Pacific, 125, 422

Mawet, D., Pueyo, L., Lawson, P., et al. 2012, in Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, Vol. 8442, International Society for Optics and Photonics, 844204

Meshkat, T., Bailey, V. P., Su, K. Y., et al. 2015, The Astrophysical Journal, 800, 5

Miller, K., Guyon, O., & Males, J. 2017, Journal of Astronomical Telescopes, Instruments, and Systems, 3, 049002

Miller, K., Bos, S., Lozi, J., et al. 2020, Astronomy & Astrophysics

Miller, K. L. 2018, PhD thesis, The University of Arizona

Miller, K. L., Males, J. R., Guyon, O., et al. 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 049004

Miskiewicz, M. N., & Escuti, M. J. 2014, Optics Express, 22, 12691

Morzinski, K. M., Close, L. M., Males, J. R., et al. 2014, in Adaptive Optics Systems IV, Vol. 9148, International Society for Optics and Photonics, 914804

Nielsen, E. L., De Rosa, R. J., Macintosh, B., et al. 2019, The Astronomical Journal, 158, 13

Otten, G. P., Snik, F., Kenworthy, M. A., et al. 2014a, in Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation, Vol. 9151, International Society for Optics and Photonics, 91511R

Otten, G. P., Snik, F., Kenworthy, M. A., et al. 2014b, in Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation, Vol. 9151, International Society for Optics and Photonics, 91511R

Otten, G. P., Snik, F., Kenworthy, M. A., et al. 2017, The Astrophysical Journal, 834, 175

Pancharatnam, S. 1956, Proceedings of the Indian Academy of Sciences - Section A, 44, 398, doi: 10.1007/BF03046095

Perrot, C., Baudoz, P., Boccaletti, A., et al. 2018, arXiv e-prints, arXiv:1804.01371. https://arxiv.org/abs/1804.01371

Por, E. H. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, Vol. 10400, International Society for Optics and Photonics, 104000V

Por, E. H., Haffert, S. Y., Radhakrishnan, V. M., et al. 2018, in Adaptive Optics Systems VI, Vol. 10703, International Society for Optics and Photonics, 1070342

Por, E. H., & Keller, C. U. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 990959

Price-Whelan, A. M., Sipőcz, B., Günther, H., et al. 2018, The Astronomical Journal, 156, 123

Quanz, S. P., Amara, A., Meyer, M. R., et al. 2013, The Astrophysical Journal Letters, 766, L1, doi: 10.1088/2041-8205/766/1/L1

Quanz, S. P., Crossfield, I., Meyer, M. R., Schmalzl, E., & Held, J. 2015, International Journal of Astrobiology, 14, 279

Quanz, S. P., Meyer, M. R., Kenworthy, M. A., et al. 2010, The Astrophysical Journal Letters, 722, L49

Racine, R., Walker, G. A., Nadeau, D., Doyon, R., & Marois, C. 1999, Publications of the Astronomical Society of the Pacific, 111, 587

Ruane, G., Riggs, A., Mazoyer, J., et al. 2018, in Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, Vol. 10698, International Society for Optics and Photonics, 106982S

Samland, M. S. 2019, PhD thesis

Skemer, A. J., Hinz, P., Stone, J., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, ed. C. J. Evans, L. Simard, & H. Takami, Vol. 10702, International Society for Optics and Photonics (SPIE), 78 – 85, doi: 10.1117/12.2314091

Skemer, A. J., Hinz, P., Montoya, M., et al. 2015, in Techniques and Instrumentation for Detection of Exoplanets VII, Vol. 9605, International Society for Optics and Photonics, 96051D

Skrutskie, M., Jones, T., Hinz, P., et al. 2010, in Ground-based and Airborne Instrumentation for Astronomy III, Vol. 7735, International Society for Optics and Photonics, 77353H

Snik, F., Otten, G., Kenworthy, M., Mawet, D., & Escuti, M. 2014, in Ground-based and Airborne Instrumentation for Astronomy V, Vol. 9147, International Society for Optics and Photonics, 91477U

Snik, F., Otten, G., Kenworthy, M., et al. 2012, in Modern Technologies in Space-and Ground-based Telescopes and Instrumentation II, Vol. 8450, International Society for Optics and Photonics, 84500M

Soummer, R., Pueyo, L., Sivaramakrishnan, A., & Vanderbei, R. J. 2007, Optics Express, 15, 15935

Stolker, T., Bonse, M. J., Quanz, S. P., et al. 2019, Astronomy & Astrophysics, 621, A59

Sun, H. 2019, PhD thesis, Princeton University

Sutlieff, B. J., Birkby, J. L., Kenworthy, M. A., et al. 2019, in Poster presented at Spirit of Lyot, Tokyo, Japan

Sutlieff, B. J., Bohn, A. J., Birkby, J. L., et al. Submitted for publication, MNRAS

Thibault, S., Allain, G., Côté, O., et al. 2019, in Education and Training in Optics and Photonics, Optical Society of America, 11143_130

Vigan, A., Fontanive, C., Meyer, M., et al. 2020, arXiv preprint arXiv:2007.06573

Wilby, M., Keller, C. U., Haffert, S., et al. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 990921

Wilby, M. J., Keller, C. U., Snik, F., Korkiakoski, V., & Pietrow, A. G. 2017, Astronomy & Astrophysics, 597, A112

Wold, S., Esbensen, K., & Geladi, P. 1987, Chemometrics and intelligent laboratory systems, 2, 37

3 First low-resolution L-band spectroscopy of HR 8799 c,d,e with a vAPP-ALES combination

D. S. Doelman, J. Stone, Z. Briesemeister, A. Skemer, F. Snik, A. Bohn, M. Kenworthy

Context In recent studies, both atmospheric modelling and IFS observations of directly imaged gas giants have focused on y-K band (0.9-2.5μm). However, observations of gas giants in the mid-infrared (3-5 μm) are critical to remove model degeneracies, determine their cloud properties, and give insight in their L-T transition.
Aims We aim to use mid-infrared observations of HR 8799 c, d, and e to provide additional spectral constraints for atmospheric models that are used to infer atmospheric properties.
Methods Using the upgraded mid-infrared ALES spectrograph in combination with the newly installed vAPP coronagraph, we retrieve low-resolution 3.0-4.2 μm spectra (R$\sim$35) of HR 8799 c, d, and e. We compare the obtained spectra with photometric literature values and models reported in literature.
Results The retrieved spectra are consistent with a flat spectrum for all three planets in the observed bandwidth, indicating a low abundance of CH_4. Moreover, the spectra are consistent with previously reported Brα photometry (4.05 μm), strengthening the case for patchy clouds.
Conclusions The retrieved spectra of all three planets are consistent with models with non-equilibrium carbon chemistry and patchy clouds, rather than uniform clouds. This suggests that the HR 8799 planets might be beginning their L-T transition.

3.1 Introduction

Directly imaged exoplanets offer great opportunities to study exoplanet atmospheres. Due to technical limitations, direct imaging has a bias towards young, self-luminous, massive planets at large separations, e.g. HR 8799 bcde (Marois et al., 2008, 2010), β Pic b (Lagrange et al., 2009), 51 Eridani b (Macintosh et al., 2015), PDS 70 bc (Keppler et al., 2018; Haffert et al., 2019) and TYC 8998-760-1 bc (Bohn et al., 2020a,b). These planets have the most favourable contrast with their host star in the thermal infrared (3-5 μm), where the performance of the adaptive optics is also optimal and the thermal background does not yet dominate. The warm gas giants are often compared to field brown dwarfs, as these objects have a similar radius, composition, and also exhibit slow cooling because of the lack of internal fusion (Skemer et al., 2014). Earlier comparisons of photometry of directly imaged planets with measured spectra or modelled spectra of older brown dwarfs showed significant differences (Marois et al., 2008; Morzinski et al., 2015). These differences could only be explained by the presence of disequilibrium carbon chemistry and thicker (patchy) clouds (Bowler et al., 2010; Barman et al., 2011a; Currie et al., 2011; Skemer et al., 2012, 2014). Clouds are patchy when they are non-uniformly distributed and are modelled in literature both as a mix between optically thick and thin clouds, or optically thick clouds combined with cloud free areas. In Barman et al. (2011b) and Marley et al. (2012), the authors suggest that the differences in cloud opacity could be explained by the lower surface gravity of these young exoplanets compared to field brown dwarfs. This was confirmed by comparing young gas giants with low-gravity field brown dwarfs, yielding better fits, e.g. HR 8799 d,e (Bonnefoy et al., 2016), β Pic b (Chilcote et al., 2017) or 51 Eridani b Rajan et al. (2017). These more detailed comparisons are enabled only by the advancements in direct imaging instruments, especially VLT/SPHERE (Beuzit et al., 2019), and GEMINI/GPI (Macintosh et al., 2014) and their integral field spectographs (IFSs) with a combined spectral range spanning the $y-K$ bands (1-2.5 μm). Updated self-consistent models in these wavelength ranges also started to produce better results by taking into account the lower gravity, producing an increase of cloud optical depth (Charnay et al., 2018). Moreover, these models predict a greenhouse effect that shifts the carbon chemical equilibrium, changing the abundances of species compared to similar temperature brown dwarfs with higher surface gravity.

So far, the description of young, self-luminous, massive planets as low-gravity and more cloudy field brown dwarfs is compelling. However, due to inconsistencies in spectra between instruments and different physics in models, it can be difficult to create a true coherent story. For example, 51 Eridani b, a giant planet with many similarities to T-type brown dwarf with strong methane absorption, provides an interesting challenge. When comparing SPHERE and GPI data from Samland et al. (2017) and Rajan et al. (2017), the extracted spectra differ in J and K band, leading to different cloud properties (uniform vs patchy, respectively) and different metallicity (highly super-solar vs solar). Moreover, for the HR 8799 planets both disequilibrium chemistry and the cloud coverage have been disputed. In Currie

et al. (2014), the authors show that for HR 8799 b and c, non-equilibrium carbon chemistry is required, yet only weak evidence for non-equilibrium chemistry is found for HR 8799 d and e. Similarly, both Bonnefoy et al. (2016) and Lavie et al. (2017) find that HR 8799 d,e can be fit well with chemical equilibrium models, in sharp contrast with Skemer et al. (2014) for HR 8799 d,e and Mollière et al. (2020) for HR 8799 e. The same papers disagree differently on the existence of patchy clouds (Skemer et al., 2012; Currie et al., 2014) versus uniform clouds (Bonnefoy et al., 2016; Mollière et al., 2020). We note that patchy clouds or banded cloud structures can generate a linear polarization signal of up to 0.03% and 1.33% respectively, assuming an oblateness of 0 (Stolker et al., 2017). SPHERE/IRDIS measurements of the HR 8799 planets did not recover the planets in linear polarization, despite the unprecedented sensitivity of $\sim 1\%$ (van Holstein et al., 2017). Differences in literature on the retrieved atmospheric properties of the HR 8799 planets can partially be explained by added physics in updated models and the availability of new data. For example, constraining the C/O ratio of HR 8799 e in Mollière et al. (2020) was only possible with the high quality K-band spectrum of the interferometer GRAVITY (GRAVITY Collaboration et al., 2019) and updates of petitRADTRANS (Mollière et al., 2019) to include opacity tables of individual atmospheric absorbers. The update was not included in Wang et al. (2020), where difference in priors significantly changed the retrieved C/O abundance for HR 8799c. However, the exclusion of L and M band data is another factor for explaining in the differences of the retrieved atmospheric properties in literature. In Wang et al. (2020), the retrieved C/O ratio was different when L- and M-band data - compiled by Bonnefoy et al. (2016)- was included, showing that these bands provide additional constraints on the abundances and cloud condition. The focus of modelling has gone toward the $y-K$ band, where VLT/SPHERE, Gemini/GPI, SUBARU/CHARIS (Groff et al., 2017), VLT/GRAVITY and Keck/OSIRIS (Larkin et al., 2006) provided excellent data. Yet, in order to simultaneously pin down T_{eff}, surface gravity, chemistry, and cloudiness, a broad wavelength coverage through the thermal-IR is required (Skemer et al., 2014). Moreover, the L and M-band data presented in Skemer et al. (2012, 2014); Currie et al. (2014) (LMIRcam, non-coronagraphic) can only be explained by modelling patchy clouds for the HR8799 planets, where uniform-cloud models fail to reproduce the 4.05 μm flux, i.e. Bonnefoy et al. (2016); Greenbaum et al. (2018); Mollière et al. (2020). As gas-giants cool and fade, their atmospheres transition from CO-rich and cloudy (L-type), to CH_4-rich and relatively cloud free (T-type). Therefore, it can be expected that gas giants have patchy clouds during the L-T transition, caused by cloud dissipation, and/or clouds sinking below the photosphere of a planet/brown dwarf (Skemer et al., 2014). In addition, the absorption cross-section per molecule of CH_4 peaks between 3-4 μm, while CO peaks between $4.1-5\mu$m (Currie et al., 2014), although the precise shape depends on the pressure-temperature profile. Hence, gaining insight in the cloud properties and C/O ratio during the L-T transition requires thermal-IR observations. Another reason for thermal-IR observations is that planets with lower masses, elderly planets, and cold-start planets will all be cooler than the planets currently detected with direct imaging (Fortney et al.,

2008). These cooler planets emit most of their light in the thermal-IR, where the contrast is also most favourable with respect to the host-star. To summarize, thermal-IR observations are necessary to remove model degeneracies in gas giants, to improve the understanding in their L-T transition, and to detect cooler planets. The Arizona Lenslets for Exoplanet Spectroscopy (ALES) is an adaptive optics-fed integral field spectrograph with sensitivity out to 5 microns (Skemer et al., 2015, 2018) and is installed at the Large Binocular Telescope (LBT). ALES was specifically designed to provide insight in the L-T transition of gas giants. The updated ALES has a main mode with a square field of view of $\sim$2.2 arcseconds with a spectral resolution of $\sim$35, spanning the $2.8 - 4.2$ μm range. Other modes have different spectral resolutions (R = 20-150) and spectral coverage spanning $2 - 5$ μm. Additionally, LBT/LMIRCAM (Wilson et al., 2008) has been upgraded with a double-grating 360-degree vAPP (dgvAPP360) coronagraph (Snik et al., 2012; Otten et al., 2017; Doelman et al., 2020), suppressing the stellar diffraction halo by multiple orders of magnitude over the full $2 - 5\mu$m bandwidth. The vAPP coronagraph in LMIRCAM can also be used for the ALES spectrograph. This combination of ALES+vAPP has several advantages. First, using the same coronagraph over the full wavelength range makes it easier to compare data taken in different instrument modes. Second, the diffraction structure of the vAPP point-spread function (PSF) is more stable as function of wavelength than that of focal-plane coronagraphs. Focal-plane coronagraphs have a changing transmission of the focal-plane masks near their inner working angle, which, combined with the wavelength dependence of atmospheric dispersion correction, changes the performance with wavelength. The vAPP PSF is only shifted while keeping the same shape, allowing for improved spectral differential imaging (SDI) (Racine et al., 1999). Third, being a pupil plane coronagraph, the vAPP generates a PSF that is independent on the location in the field of view (FOV). Therefore, all stars and companions have the identical PSFs, in contrast to focal-plane coronagraphs. Identical PSFs respond the same to variations in Strehl, making the stellar PSF a good photometric reference, especially in the thermal-IR where fast imaging is standard due to the high thermal background. Fourth, with the absence of a focal-plane mask and the PSF being independent on the location in the FOV, it is no longer necessary to accurately align the PSF on focal-plane masks. This reduces overhead at the beginning of an observing sequence, as well as reducing the complexity for chopping and nodding.

In this paper, we present the first low-resolution spectra of HR 8799 cde in the 3.5-4.1 μm range using the ALES/vAPP combination. First, we characterize the vAPP performance in section 3.2. Section 3.3 summarizes the observational data and data reduction. We show the retrieved spectra in section 3.4 and discuss them in section 3.5.

3.2 Characterization of the dgvAPP360.

The dgvAPP360 in LMIRCAM was installed early September 2018. The design and first-light results are presented in Doelman et al. (2020). In this paper we show that the double-grating vAPP is operating as expected, and that it produces a single coronagarphic PSF in contrast to standard gvAPPs (Otten et al., 2017). In addition, observations of the disk PDS 201 in L'-band (3.41 - 4.00 μm) using this vAPP show that the vAPP has an improved sensitivity of a factor two compared to non-coronagraphic imaging in the regions closest to the star (450-800 mas) (Wagner et al., 2020). Yet, the performance and vAPP properties are not fully characterized because of the contribution of the disk. In this section, we characterize the vAPP by studying the PSF using the internal source.
First, we image the PSF using LMIRCAM to explore the star light suppression in the dark zone of the vAPP. We use the internal source of LMIRCAM and the installed L-band filter (2.82 - 4.21 μm) [1]. We take 30 images of 0.5 seconds integration time and subtract the average dark of 10 images with the same integration time. We median-combine the all images, and the final PSF is shown in the top left of Fig. 3.1. The almost vertical stripes present in the image are created by the internal source of LMIRCAM and are caused by light leaking around a pin hole mask. These stripes are not present in on-sky data, and are ignored in the further analysis of the PSF. We select a D-shaped region on the left side of the PSF to estimate the raw contrast. Moreover, we mask two ghosts, in this dark zone and two features that are aligned with the stripes to obtain an accurate estimate of the raw contrast of the vAPP. The masks are shown on the top right of Fig. 3.1. The flux is averaged azimuthally in half annuli and normalized on the peak flux of the PSF. In addition, we simulate a PSF with the same plate scale and bandwidth, assuming a uniform spectral profile, using `HCIPy` (Por et al., 2018). The measured raw contrast and the raw contrast of the simulated PSF are shown in the bottom panel of Fig. 3.1. The raw contrast reaches the noise floor of 5.8×10^{-5} ($\Delta m = 10.6$) at a separation of 470 mas, calculated at 3.9 μm. Closer in, the contrast is reduced by 1-2 magnitudes compared to the theoretical curve, likely due to some low-order aberrations. The theoretical inner working angles and outer working angles for the vAPP are presented in table 3.1. The noise floor, dominated by the thermal background, is given by the average flux between the two bright rings of the PSF (3.0"-3.4"), using half-annuli. We conclude that, within the limitations of the measurement, the coronagraphic performance of the vAPP is as expected.
The double-grating vAPP has two separate liquid-crystal layers, an additional glue layer and an extra substrate compared to a standard gvAPP (Doelman et al., 2017, 2020). These additional layers lead to extra absorption, specifically in the 3.24-3.5 μm range, where both the liquid-crystal molecules and the glue molecules have an absorption feature from carbon-carbon bonds (Otten et al., 2017). We note that

[1] `http://svo2.cab.inta-csic.es/svo/theory/fps3/index.php?id=LBT/LMIRCam.L_77K&&mode=browse&gname=LBT&gname2=LMIRCam#filter`

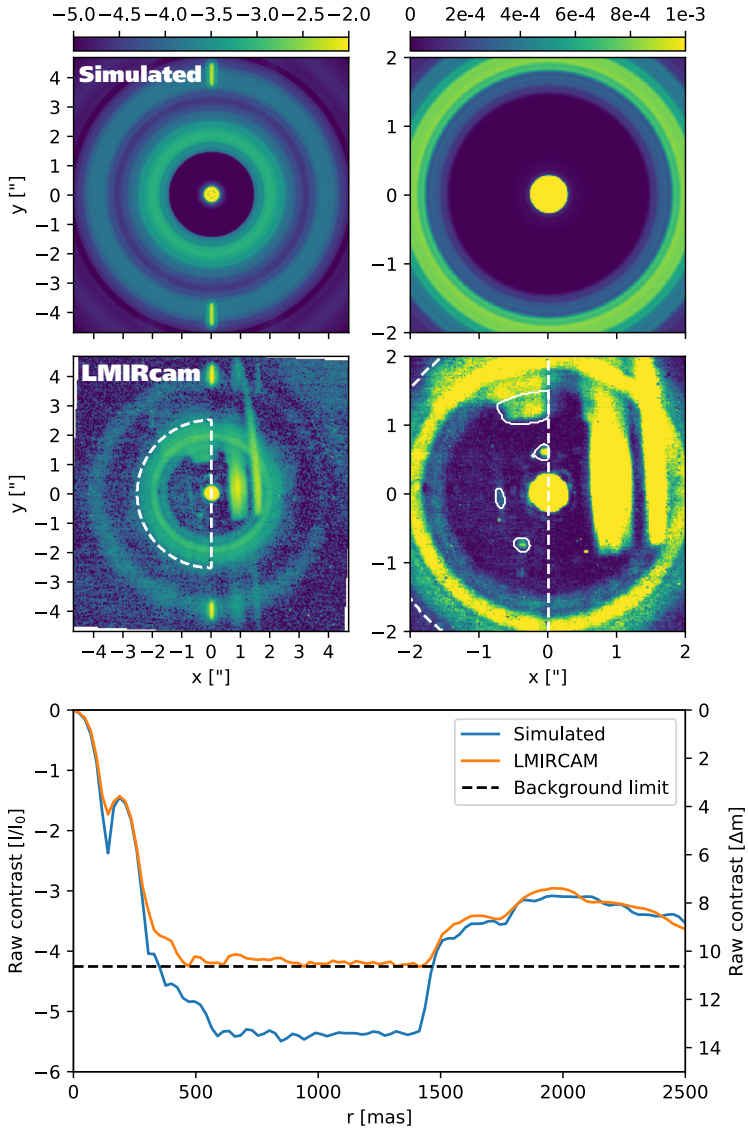


Figure 3.1: Images and the raw-contrast curve of the L-band point-spread function of the vAPP in LMIRCAM using the internal source. The stripes are created by the internal source and are not present in on-sky data. The raw contrast was calculated by averaging half-annuli in the marked D-shaped region, omitting the regions marked by the solid line in the top right figure. These regions contain ghosts and stripes from the internal source.

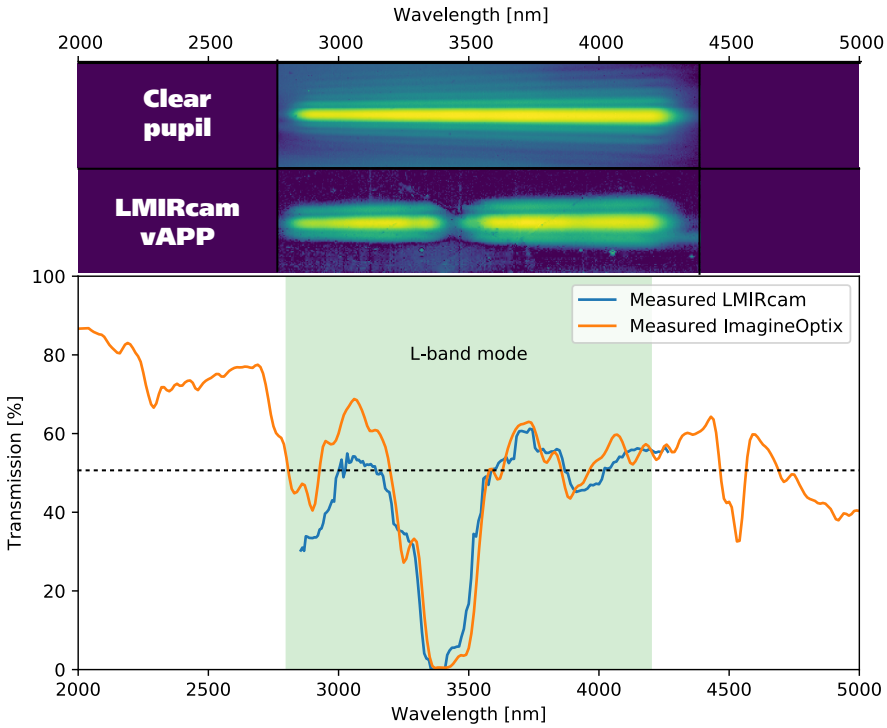


Figure 3.2: Transmission measurements of the vAPP in the 2.8-4.2 μm range using a grism to disperse the vAPP and non-coronagraphic PSF.

Table 3.1: Theoretical inner working angles (IWAs) and outer working angles (OWAs) as function of wavelength, λ. The OWAs larger than the ALES field of view are in italics.

λ (μm)	2.0	2.3	2.8	3.3	3.9	4.2
IWA (mas)	143	200	250	278	300	357
OWA (mas)	793	1110	*1388*	*1547*	*1666*	*1983*

the Annular Groove Phase Mask (AGPM) coronagraph also has this absorption feature because they are made from diamond (Defrere et al., 2014). Because these features are in the spectral range of ALES, we conduct an experiment to mesaure the impact. We disperse the coronagraphic (vAPP) PSF and non-coronagraphic (clear pupil) PSF using a grism in tandem with the L-Spec filter, using only the left aperture (SX) of the internal pupil mask. The dispersed PSFs are displayed in Fig. 3.2. In addition, we obtain narrow band images at 2.9 μm, 3.3 μm, 3.5 μm, and 3.9 μm that are used for wavelength calibration. Because the pupil selection mask is in the same filter wheel as the narrow band filters, the narrow band images are the coherent sum of both pupils, creating Fizeau fringes. We remove this effect for our wavelength calibration by summing the flux in 100 pixels in the fringe direction. We fit a Gaussian to the one-dimensional sum of each wavelength to retrieve accurate centroids. We then use the centroids to calculate the wavelength solution of the dispersion.
Calculating the transmission of the vAPP is complicated by the difference in intrinsic Strehl between the two PSFs. The vAPP has an intrinsic Strehl of 46% and the PSF core is slightly broadened due to the apodization by the vAPP in the pupil plane. We retrieve the actual throughput by forward modelling both PSFs for the full bandwidth. For continuous wavelength coverage, we generate a wavelength-scaled PSF every 2 pixels in the image, corresponding to an average spectral resolution of 10 nm. The model PSF is the incoherent sum of all individual PSFs with an individual scale factor. We minimize the difference between the simulated and measured PSF by changing these scale factors, thereby retrieving the true input spectrum. The final transmission is the optimized vAPP spectrum divided by the non-coronagraphic spectrum, and is shown in Fig. 3.2. We compare the results with the transmission that is measured by the manufacturer. The curves are in good agreement except for the transmission between 2.9 and 3.2 micron. It is unclear what causes the difference of more than 10% in this spectral bin, compared to the LMIRcam average transmission of 47%. A significant fraction of the spectral band is unavailable when using the vAPP because of the absorption feature. Specifically, the absorption is centered on a spectral feature of CH_4. While this inhibits the detection of methane in the atmospheres of L-type gas giants, cooler gas giants with more CH_4 will have a measurable spectral slope between 3.5-4.1 μm, where the average transmission of the vAPP is 51%.

3.3 Observations and data reduction

We observed HR 8799 on 2019 September 18 for 1 hour 53 minutes with the LBTI/ALES instrument in combination with the vAPP coronagraph. Only the left-side telescope fed the ALES instrument at the time of observation, the other side was feeding a different instrument. We used the L-band mode of ALES, spanning $2.8 - 4.2$ μm, with a spectral resolution of ~ 35 (Skemer et al., 2018). We acquired 1300 frames on-target with an integration time of 3900 milliseconds per frame, for a total of 1 hour 24 minutes. The conditions were stable with a seeing

between 0.8 and 1.1 arcseconds, and the derotator was switched off, yielding a total field rotation of 85°.64 through meridian crossing. In addition, we take a total of 99 background frames, where the first 13 were taken after 100 science frames and the other 86 directly after the science sequence. This unconventional observing sequence without beam switching increases the time on-source and minimizes overhead time, yielding an on-target observation efficiency of 93%. The downside of this method is a higher photon noise in the background estimates.

3.3.1 Spectral cube extraction

A lenslet-based integral field spectrograph images the focal-plane onto a microlens array, followed by a dispersing element. Raw images from this instrument contain a spectrum for every focal-plane location that corresponds to individual lenslets. To extract the images as function of wavelength (= spectral cube), we use a data reduction pipeline called Methods for Extracting ALES Data (`MEAD`) (Briesemeister et al., 2018). This pipeline cleans up the raw images (e.g. dark subtraction, pixel flat-fielding), calculates the focal-plane geometry to correct displacements between the prism and the lenslets, implements the wavelength calibration, and maps every light-sensitive pixel to a weight in the created data cube. We inspect the extracted cubes and find two unusual properties of the data. First, we see that the background is highly variable as function of wavelength, specifically at the lower end of the wavelength range. This indicates that there is a light leak present in the ALES instrument that contributes to the thermal background. While the light leak is additive, and can thus be removed through standard background subtraction, it contributes to a structured photon noise pattern on the detector. Second, we find that during the observation sequence the PSF moves compared to the background in a u-shape. This u-shape is $\sim$ 35 mas (=1 spaxel) in the x-direction and $\sim$ 70 mas in the y-direction, while the frame-to-frame jitter is $\sim$ 4 mas. A known source of PSF movement in ALES is the lenslet array, which moves due to flexure of the instrument with telescope pointings. The movement of the PSF on the detector is correlated with airmass, suggesting that flexing is indeed a contributor. However, the movement of the lenslets has not been known to exceed one spaxel, and perhaps other factors contribute as well. We note that any PSF movement prior to the lenslets is not affecting the coronagraphs performance because the vAPP is a pupil-plane coronagraph. PSF movement on the focal-plane mask of coronagraph's like the AGPM would reduce their performance. As a final step, we bin the data per four frames in time by averaging, reducing the number of cubes from 1300 to 325.

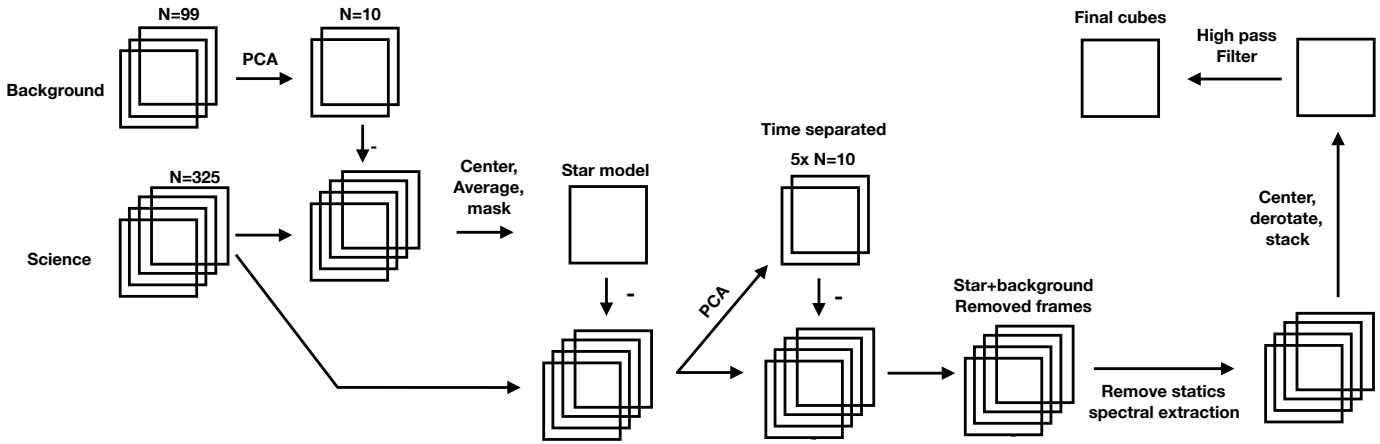


Figure 3.3: Flowchart of the data-reduction method.

3.3.2 Data reduction

Most conventional data reduction algorithms remove the background from the frames at this point. However, removing the thermal background from the science data using only the 99 background frames is suboptimal. Any residuals in the background-subtracted frames that are static, will not be static compared to the shifting frame of reference of the PSF. Techniques like angular differential imaging (ADI) (Marois et al., 2006) will not capture these structures and they will carry over to the final frames. Yet, the background is present in every science image, of which there are 1300 before binning. Besides this background, the science frames also contain the stellar and planetary PSFs. Because the planets move across the detector with the parallactic angle, they can easily be removed from the background estimate. We can mask their predicted locations, such that their flux is not influencing the background estimates. The star constitutes a larger problem, as possible quasi-static speckles or ghosts will contaminate the estimated background. In addition, the background on the stellar PSF location is inaccessible, which makes accurate stellar photometry impossible.
We introduce a combined method where we model the star and background separately and subtract them from all frames for each wavelength. A flowchart of the method is shown in Fig. 3.3. We start by subtracting the 99 background frames from the science data to create a stellar PSF model. While these background frames introduce noise at the planet locations, the background noise is much lower than the stellar photon noise. Therefore, an accurate model of the stellar PSF can be extracted by centering and co-adding the background subtracted frames. We create a mask to select features that are significant and co-moving with the star, i.e. a ghost and ringing from the spectral cube extraction. The ringing could be a combined effect from electrical ghosts by inter-channel capacitive coupling (Finger et al., 2008) and the cube extraction. The data is also clipped to remove all features that are less than 2σ compared to the background outside of the mask. An example of the PSF model for a single wavelength is shown in Fig. 3.4. Removing the stellar PSF from the science data using this PSF model is now straightforward. We fit a decentered PSF model to background subtracted data frames for each wavelength, retrieving the stellar PSF intensity and the science frames without the stellar PSF.
After masking the planets, we can model the background from the star-subtracted science frames. To minimize self-subtraction, we select for every frame the frames separated in time by 30 minutes and calculate 10 PCA components from this selection. We optimally subtract these components, such that the residuals in the frame are dominated by the photon noise of the star and background. This method of background subtracting results in cubes with a subtracted star and background, where the background is calculated from the science frames themselves. Now, we can check if this method indeed reduces the background noise compared to simply using 10 PCA components of the 99 background frames. By subtracting 10 PCA components from the raw data for both methods, we compare the standard deviation of the residuals after masking the star. Between 2.9 μm and 4.2 μm,

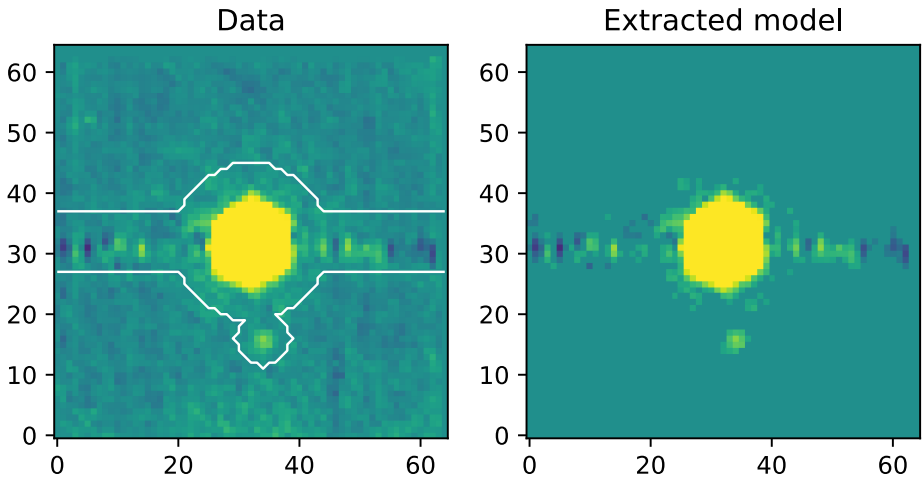


Figure 3.4: *Left:* Background-subtracted PSF at 3.55 μm, using 10 PCA components of the 99 background frames. Residual structures indicate the presence of a ghost, some speckles, and a horizontal oscillating structure. *Rigth:* Same as left, but clipped at 2σ inside the area of a mask, indicated by the white lines left image.

we obtain a reduction of 10% in the standard deviation of the residuals using the data as a background estimate. While this would be less than what could expected from photon noise alone, we proceed with the science data background estimate. We inspect the star- and background-subtracted frames for residual structure by averaging them in time and wavelength. We find that the residuals are not well described by purely Gaussian noise, but contain structures that are column and row specific and vary in time, see Fig. 3.5. The structures are faint (1-10 counts) and originate from the spectral cube extraction. Each pixel in the spectral cubes correspond to a spaxel in the raw data, which are separated by multiple pixels on the detector. Discontinuities in rows and columns are therefore expected. We characterize these discontinuities by fitting polynomials of the third order to each row and column, which are shown in Fig. 3.5. A third order polynomial is low-order enough compared to the 65 pixels that it is minimally affected by planet signal. Yet, we mask the star with an 18 pixel circular mask and the planets with a 5 pixel circular mask and remove those pixels from the fit. Using a Kolmogorov-Smirnov test we verify that both the before and after distributions are non-Gaussian, however we find that the average of the noise distribution is now consistent with zero and the standard deviation is reduced by 10%. We note that the polynomial background fit has a large number of variables for the full image, but it was the only method that captured the behavior of this phenomenon. Combined with the stellar PSF removal and the background removal, the row and column fits remove most structures present in the data in a way that minimizes self-subtraction of planets. The authors would like to *reiterate* that this unconventional data reduc-

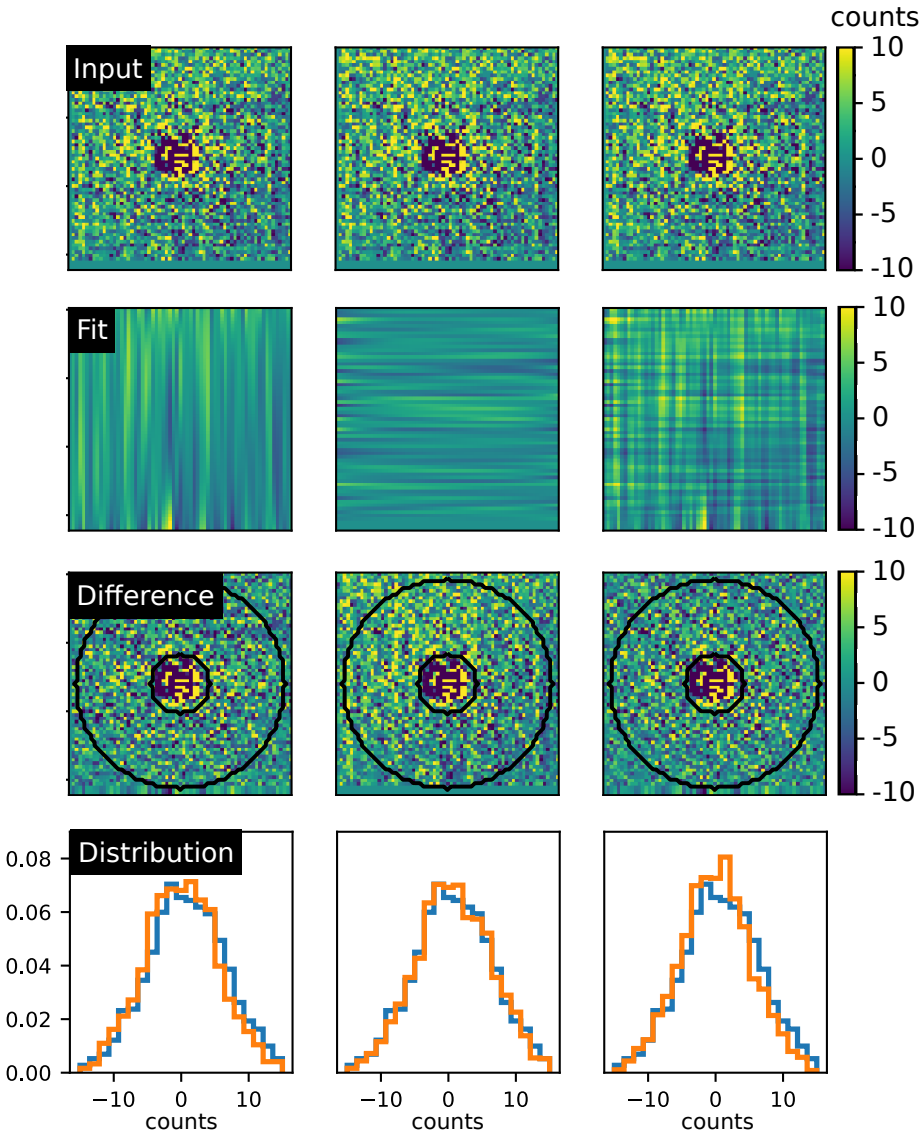


Figure 3.5: Removal of row and column discontinuities by fitting a third-order polynomial to every row and column.

tion method is required because of the low number of available background frames and the movement of the stellar PSF. ADI centered on the star would not be good at characterizing the static structures of the background, including the row and column discontinuities.
We center and derotate the background subtracted cubes and median stacking them in time. A final step is a high-pass filter, where we remove global structures on the background by removing a Gaussian smoothed frame for each wavelength in the final cube. The Gaussian has a standard deviation of 5 pixels (FWHM = 11.8 pixels) and we mask the locations of the HR 8799 planets and the star.

3.3.3 Spectral extraction

The data reduction aims to minimize self-subtraction. However, self-subtraction is still non-zero and it is wavelength-dependent. To extract the final spectra, we negatively inject planets at the locations of the HR 8799 planets in the extracted data cubes. The injected planets are copies of the stellar PSF at each wavelength. The raw cubes are reduced using the method described in Fig. 3.3. We optimize their location and amplitude by evaluating the Hessian at the planet location in a circular aperture with a diameter of 7 pixels. The Hessian is a measure of the curvature of the image surface, which is minimal when the planet is completely removed (Stolker et al., 2019). To evaluate confidence of the retrieved spectra we apply bootstrapping to the data reduction method, selecting 325 frames at random from the data for 50 times in total. The fluxes of the planets for each wavelength are retrieved using aperture photometry, rather than fake planet injection, as the data reduction method is rather computationally expensive. Assuming that the self-subtraction is similar for all iterations, the distribution of retrieved amplitudes should be a good approximation of the distribution with negative planet injection. From this posterior distribution of planet fluxes, we take the standard deviation as the 1σ error bars. Moreover, we bin the data between 2.99 and 3.17 μm and 3.17 and 3.36 μm in single wavelength bins to retrieve two photometric points for wavelengths short of the vAPP aborption feature. The negative planet is injected for all spectral bins separately, however, the final evaluation of the Hessian is done on the median combined images in the wavelength bins. Bootstrapping with the same parameters is applied to find the error bars on these spectra as well.
We perform the true flux calibration of the spectra using the SED analyser VOSA (Bayo et al., 2008). We fit a BT-Settle model to the SED of the host star and retrieve a temperature of 7200 K, log(g) = 4 log(cm/s^2), a metallicity of 0.5, an α = 0 and and a multiplicative dilution factor of 6.416e-19. This BT-Settle model is used to calculate the true flux of the host star in every wavelength bin. We multiply the relative flux of HR 8799c,d,e with the stellar flux to retrieve the final spectra.

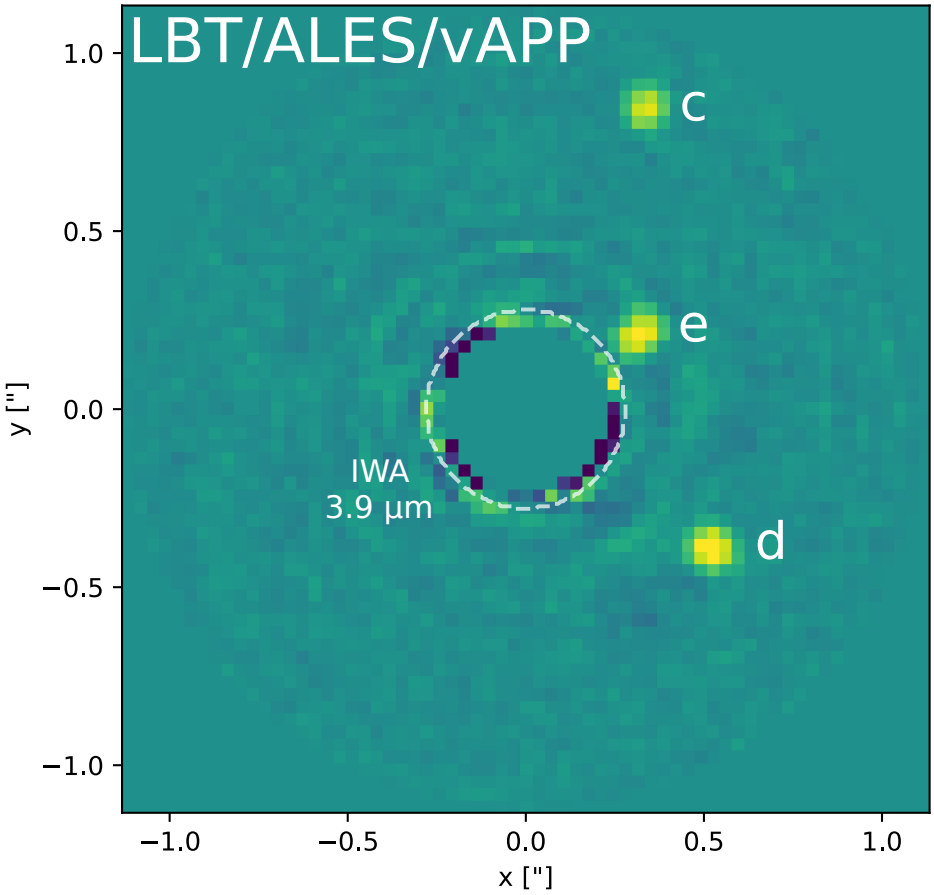


Figure 3.6: LBT/ALES image of HR 8799 cde using the vAPP coronagraph. The final image is the combination of the individual wavelength slices between 3.55-4.25 μm. North is up, East is left. The images are 2.2" across.

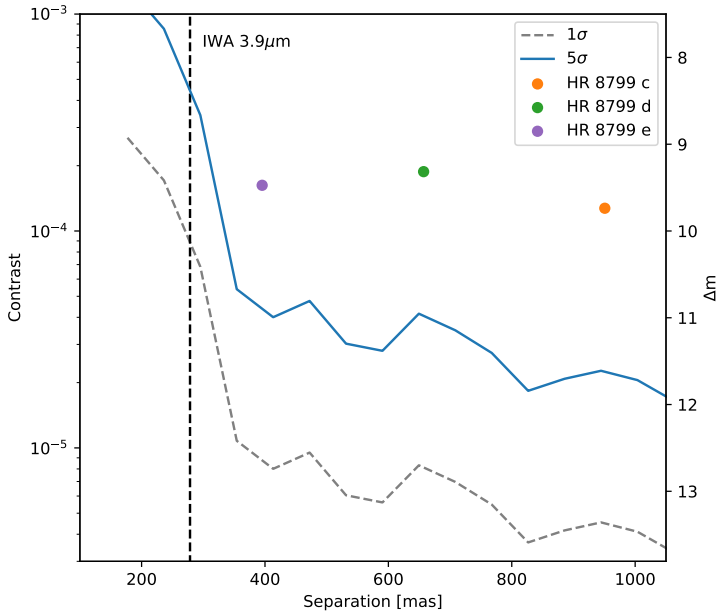


Figure 3.7: Planet-to-star contrast for the LBT/ALES as a function of angular separation.

3.4 Results

Using the combination of ALES with the dgvAPP360, we recover HR 8799 c, d, and e. We note that HR 8799 b is outside of the field of view of ALES. We optimally combine the data between 3.5 μm and 4.2 μm to maximize the signal-to-noise ratio, where the throughput of the vAPP is highest. The final image is shown in Fig. 3.6, where all three planets are clearly visible. Only at the location of HR 8799 e there are some residuals from speckles, while the residuals at the locations of HR 8799 c and d are dominated by the thermal background. We calculate the post-processed contrast using aperture photometry. We compare the standard deviation in the sums of the flux in subapertures with a diameter of $1.7\lambda/D$ for a given angular separation. The resulting contrast curve is shown in Fig. 3.7, and quickly reaches a noise floor beyond the inner working angle. HR 8799 c, d, and e are detected with a signal-to-noise ratio of 29, 25 and 19 respectively.
The retrieved spectra of the inner three HR 8799 planets and literature values for these wavelengths are compiled in Fig 3.8. Interestingly, the retrieved spectra of HR 8799 c, d, and e are almost equivalent, given an absolute offset. They are almost completely flat and featureless over the used bandwidth of the ALES

spectrograph. HR 8799 d is the brightest, followed by HR 8799 e, and HR 8799 c. We significantly detect HR 8799 c and d between 2.99 and 3.35 μm, with fluxes that are comparable with the recovered fluxes for wavelengths longer than 3.5 μm. No significant detection was made between 3.35 μm and 3.5 μm due to the absorption of the vAPP coronagraph. The retrieved spectra of HR 8799 c and e are consistent with the reported literature values. On the other hand, HR 8799 d shows significant deviations of compared to the values reported in Skemer et al. (2014), up to a factor of 3. We will discuss this further in section 3.5.1.

3.5 Discussion

3.5.1 Data reduction and the increased flux of HR 8799 d

We find that the retrieved spectra of HR 8799 c and e are consistent with literature values, and deliver a complementary view of the system. The spectra show a channel-to-channel variation that is often larger than the bootstrapped error bar. This is curious as the spectral extraction is done at a higher spectral resolution ($R \sim 200$) than the spectral resolution of the spectrograph ($R \sim 35$) (Skemer et al., 2018). Systematic errors from the cube extraction are the likely cause of the channel offsets. More importantly, the recovered spectrum of HR 8799 d shows a larger flux than the reported L-band fluxes. As the recovered fluxes of the other two planets are consistent with literature, it is not likely that it is merely a scaling issue. Moreover, being the middle planet of the three, it is implausible that algorithmic mechanisms like self-subtraction would cause this difference. It is more likely that offsets are introduced either during the cube extraction, the star subtraction, or that the increase in flux is astronomical of origin. These three possibilities would have a different time dependency, meaning that the flux of the planets would change in time in a different way. We explore the time variability of the three planets by combining 22 minutes of data and the wavelength channels between 3.6 and 4.1 μm into a single photometric point. The resulting five photometric points for each planet are shown in Fig. 3.9. The error bars are determined by bootstrapping the 22 minutes of data 50 times and taking the standard deviation of the planet flux. Flux variations from varying AO performance of sky transmission are corrected by photometry of the stellar PSF. It is clear that the behavior of HR 8799 c and e is again different from the behavior of HR 8799 d, which is not relatively constant but decreases by a factor of 2 over the observing time.

It is worthwhile to note that HR 8799 d is on a different quadrant on the detector than HR 8799 c and e. The current extraction does not use a lenslet flat, accounting for non-uniformity in the illumination of the lenslets. Moreover, field dependent aberrations, such as astigmatism, can change the shape of a PSFlet to align with or be orthogonal to the dispersion direction of the grism. Not only would this lead to a change in the effective spectral resolution, but depending on the width of the mask used for spectral extraction, could lead to flux variations.

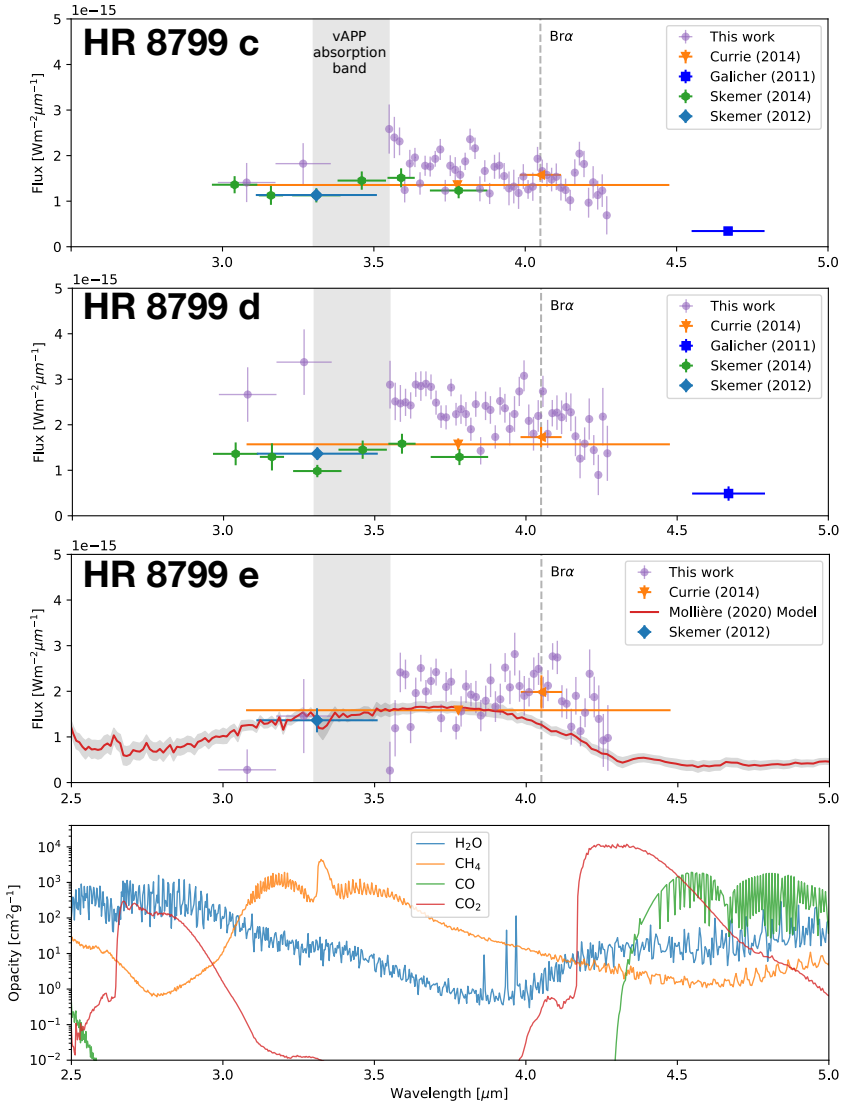


Figure 3.8: Spectra recovered with negative planet injection for HR 8799 cde, showing flux at 10 pc. We use the data 3.3 μm magnitudes from Skemer et al. (2012), the different narrow-band magnitudes from Skemer et al. (2014), the broad L'-band and Brα magnitudes from Currie et al. (2014), and the M-band magnitudes from (Galicher et al., 2011). In addition, we add the best-fit model for HR 8799 e from Mollière et al. (2020). The bottom panel shows the opacities of H_2O, CH_4, CO, and CO_2 in the same wavelength range, calculated with PetitRADTRANS for a temperature of 1200 K and a pressure of 1 bar. (Mollière et al., 2019).

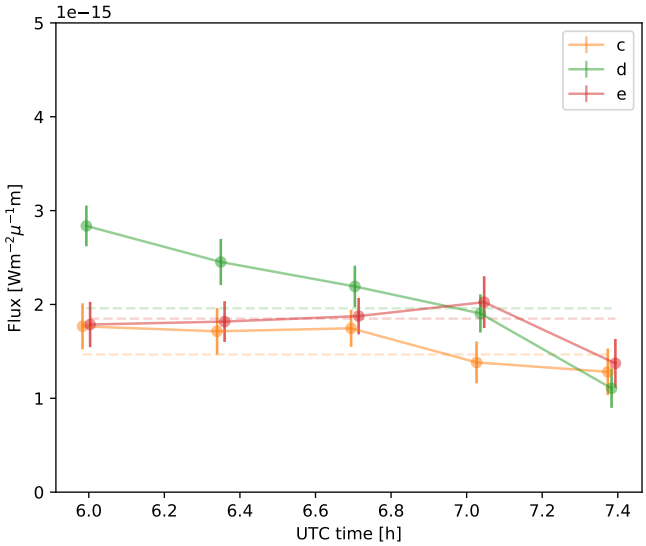


Figure 3.9: Retrieved flux of HR 8799 c, d and e between 3.6 and 4.1 μm for five different time bins spanning 22 minutes.

Such effects have not been characterized, and are currently not captured in the `MEAD` pipeline. However, this is unlikely to cause a factor of 2 in flux variation for a single quadrant of the detector.
Another possibility is the star subtraction, where we subtract average PSF structures. The horizontal oscillating structure has a similar magnitude as the HR 8799 planets and HR 8799 d passes through this structure in the first 22 minutes of the observation run. HR 8799 e passes through some of the PSF structures that look like speckles, yet the photometry is the most stable. It is likely that these structures affected the photometry of HR 8799 d and e, although this can not be responsible for the total decline in the flux of HR 8799 d after the first 22 minutes. HR 8799 c is not affected by any stellar subtraction, yet it also shows a decline during the last 44 minutes. The star subtraction can therefore not be solely responsible for the flux offset of HR 8799 d. We note that not subtracting these PSF structures would also lead to flux variations of the planets. An improved PSF characterization and subtraction would be beneficial to the spectrum and the variability measurements. We reject the possibility that the increased flux of HR 8799 d with respect to literature values is caused by intrinsic variability of the planet. While some variability is expected for a cloudy atmosphere, this would not be very strong in the L-band (Lew et al., 2020a). Therefore, it can not explain a variation of a factor 2 in flux between our measurements and the Skemer et al. (2014) data.

3.5.2 Implications of the common flat and featureless spectra

The ALES spectrograph covers 2.8-4.2 μm, which contains only a few features from the most abundant molecules, shown in the bottom panel of Fig. 3.8. Most prominent is the methane band (3.0-4.0 μm), which has a peak opacity at 3.3 μm. At one edge is an H_2O band (2.5-3.0 μm), while just beyond the other edge there is a CO absorption band (4.3-5.1 μm). It is therefore expected that for T-type gas giants their spectrum in the ALES band is dominated by CH_4 absorption, while the CO dominated atmospheres of L-type gas giants do not show a lot of structure. As noted before in Skemer et al. (2014), the spectra of HR 8799 c and d relatively flat at between 3 and 3.8 μm wavelengths, indicating that the atmospheres are still CO dominated. Photometry of HR 8799 e at 3.3 μm could only be explained by carbon disequilibrium chemistry, with lower CH_4 abundances (Skemer et al., 2012) compared to field brown dwarfs, which was later confirmed by Mollière et al. (2020) using the radiative transfer code PetitRADTRANS. It is therefore expected that HR 8799 c, d, and e only have a weak methane absorption feature, which is in agreement with our data. A common feature of our retrieved spectra of the three HR 8799 planets is that they are almost constant in flux between 3.0 and 4.0 μm. Due to the absorption band of the vAPP coronagraph it is not possible to further constrain the methane absorption.
For wavelengths longer than 4.0 micron we detect planet fluxes that are consistent with their 3.5-4.0 micron flux. Moreover, our data at 4.05 micron (Brα) agrees with the values reported by Currie et al. (2014) for all three planets, showing that the flatness of the spectrum beyond 4 μm is likely real. This flatness of the spectrum beyond 4 μm is interesting because our data for HR 8799e clearly shows a discrepancy with the uniform-cloud PetitRADTRANS model (Mollière et al., 2020) for these wavelengths. This discrepancy between models and the fluxes at these longer wavelengths has been reported before (Madhusudhan et al., 2011; Bonnefoy et al., 2016). A solution for the disagreement between the flux in the Brα filter and simulated spectra has been used as a key argument in favour of patchy clouds (Skemer et al., 2012, 2014; Currie et al., 2014). Patchy cloud models have a flatter spectrum beyond 4 μm, extending to the CO absorption band at 4.3 μm similar to our data. The slope and total flux in this wavelength range does depend on the implementation of patchy clouds, which can be a partial cloud cover with cloud-free patches or patches with clouds at different temperatures (cloud mixing). Cloud mixing gives better results (Skemer et al., 2014), but is not self-consistent. The cloud models are not good enough to retrieve atmospheric properties, but they are promising for explaining our HR 8799 c, d, and e spectra.
The inconsistency of modelled and measured flux beyond 4 μm is not the only indication for the existence of patchy clouds. Brown dwarfs show photometric variability that could be explained by a heterogeneous cloud coverage (Apai et al., 2013; Lew et al., 2020b,a). In addition, the luminosity of gas giants and their

colors are in tension with each other, resulting in models of gas giants with planet radii that are not physical, i.e. < 1 M_J (Bonnefoy et al., 2016; Greenbaum et al., 2018). Electron degeneracy pressure prevents the radii from decreasing further than this limit. A solution is again given by patchy clouds, such that the planet emits non-isotropically with brighter and darker patches, like the bands of Jupiter. The bright patches dominate the shape of the spectral energy distribution, while the total emitting area is smaller, allowing for a physical planet radius (Skemer et al., 2012) that matches the luminosity.
New observing methods or new instrumentation have provided multiple indications of patchy clouds for gas giants in their L-T transition, and highlighted the discrepancy between models and and the spectra of gas giants. In the near future the James Webb Space Telescope will deliver more stable photometry at longer wavelengths, and high-resolution spectroscopy combined with coronagraphy will provide more insight the presence of molecular species in atmospheres of gas giants (Mawet et al., 2016; Otten et al., 2020). To fully understand the data and its implications for the atmospheres of young massive exoplanets or field brown dwarfs, a next step would be to develop atmospheric models that include patchy clouds.

3.6 Conclusions

We report on the first science results of the upgraded ALES spectrograph, with a spectral higher resolution(R$\sim$35) and a new double-grating vAPP coronagraph with a 360 degree dark zone. We characterize the coronagraphic performance, reaching a raw contrast of 3×10^{-4} at 300 mas at 3.9 μm and the background limit (5.8×10^{-5}) at 470 mas. Using the upgraded ALES in combination with the broadband double-grating vAPP, we recover these three planets between 3.0 and 4.25 μm and obtain the first low-resolution L-band spectra of HR 8799 c, d, and e. We find that the spectra of all three planets are similar and flat in these wavelength ranges. The recovered spectra show no significant flux decrease beyond 4 μm and find that the planet fluxes at Brα (4.05 μ) are consistent with literature values. So far, only models implementing a heterogeneous cloud coverage can explain this excess flux, which strengthens the case that the atmospheres of HR 8799 c, d, and e do not contain uniform clouds, and perhaps have started their L-T transition. We detect no significant methane absorption, although this is not well constrained by our data as an absorption band of the vAPP coronagraph obscures the 3.3 micron feature. The lower abundance of methane compared to field brown dwarfs of similar temperature is consistent with previous findings, indicating that disequilibrium chemistry plays a role for all three planets.

Bibliography

Apai, D., Radigan, J., Buenzli, E., et al. 2013, The Astrophysical Journal, 768, 121

Barman, T. S., Macintosh, B., Konopacky, Q. M., & Marois, C. 2011a, The Astrophysical Journal, 733, 65

—. 2011b, The Astrophysical Journal Letters, 735, L39

Bayo, A., Rodrigo, C., Barrado Y Navascués, D., et al. 2008, A&A, 492, 277, doi: 10.1051/0004-6361:200810395

Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, Astronomy & Astrophysics, 631, A155

Bohn, A., Kenworthy, M., Ginski, C., et al. 2020a, Monthly Notices of the Royal Astronomical Society, 492, 431

Bohn, A. J., Kenworthy, M. A., Ginski, C., et al. 2020b, The Astrophysical Journal Letters, 898, L16

Bonnefoy, M., Zurlo, A., Baudino, J.-L., et al. 2016, Astronomy & Astrophysics, 587, A58

Bowler, B. P., Liu, M. C., Dupuy, T. J., & Cushing, M. C. 2010, The Astrophysical Journal, 723, 850

Briesemeister, Z., Skemer, A. J., Stone, J. M., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 107022Q

Charnay, B., Bézard, B., Baudino, J.-L., et al. 2018, The Astrophysical Journal, 854, 172

Chilcote, J., Pueyo, L., De Rosa, R. J., et al. 2017, The Astronomical Journal, 153, 182

Currie, T., Burrows, A., Itoh, Y., et al. 2011, The Astrophysical Journal, 729, 128

Currie, T., Burrows, A., Girard, J. H., et al. 2014, The Astrophysical Journal, 795, 133

Defrere, D., Absil, O., Hinz, P., et al. 2014, in Adaptive Optics Systems IV, Vol. 9148, International Society for Optics and Photonics, 91483X

Doelman, D. S., Por, E. H., Ruane, G., Escuti, M. J., & Snik, F. 2020, Publications of the Astronomical Society of the Pacific, 132, 045002

Doelman, D. S., Snik, F., Warriner, N. Z., & Escuti, M. J. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, ed. S. Shaklan, Vol. 10400, International Society for Optics and Photonics (SPIE), 224 – 235, doi: 10.1117/12.2273406

Finger, G., Dorn, R. J., Eschbaumer, S., et al. 2008, in High Energy, Optical, and Infrared Detectors for Astronomy III, Vol. 7021, International Society for Optics and Photonics, 70210P

Fortney, J. J., Lodders, K., Marley, M. S., & Freedman, R. S. 2008, The Astrophysical Journal, 678, 1419

Galicher, R., Marois, C., Macintosh, B., Barman, T., & KONOPACkY, Q. 2011, The Astrophysical Journal Letters, 739, L41

GRAVITY Collaboration, Lacour, S., Nowak, M., et al. 2019, A&A, 623, L11, doi: 10.1051/0004-6361/201935253

Greenbaum, A. Z., Pueyo, L., Ruffio, J.-B., et al. 2018, The Astronomical Journal, 155, 226

Groff, T., Chilcote, J., Brandt, T., et al. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, Vol. 10400, International Society for Optics and Photonics, 1040016

Haffert, S., Bohn, A., de Boer, J., et al. 2019, Nature Astronomy, 3, 749

Keppler, M., Benisty, M., Müller, A., et al. 2018, Astronomy & Astrophysics, 617, A44

Lagrange, A.-M., Gratadour, D., Chauvin, G., et al. 2009, Astronomy & Astrophysics,

493, L21
Larkin, J., Barczys, M., Krabbe, A., et al. 2006, New Astronomy Reviews, 50, 362
Lavie, B., Mendonça, J. M., Mordasini, C., et al. 2017, The Astronomical Journal, 154, 91
Lew, B. W., Apai, D., Marley, M., et al. 2020a, arXiv preprint arXiv:2009.10714
Lew, B. W., Apai, D., Zhou, Y., et al. 2020b, The Astronomical Journal, 159, 125
Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, proceedings of the National Academy of Sciences, 111, 12661
Macintosh, B., Graham, J., Barman, T., et al. 2015, Science, 350, 64
Madhusudhan, N., Burrows, A., & Currie, T. 2011, The Astrophysical Journal, 737, 34
Marley, M. S., Saumon, D., Cushing, M., et al. 2012, The Astrophysical Journal, 754, 135
Marois, C., Lafreniere, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, The Astrophysical Journal, 641, 556
Marois, C., Macintosh, B., Barman, T., et al. 2008, science, 322, 1348
Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010, Nature, 468, 1080
Mawet, D., Wizinowich, P., Dekany, R., et al. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 99090D
Mollière, P., Wardenier, J., van Boekel, R., et al. 2019, Astronomy & Astrophysics, 627, A67
Mollière, P., Stolker, T., Lacour, S., et al. 2020, Astronomy & Astrophysics, 640, A131
Morzinski, K. M., Males, J. R., Skemer, A. J., et al. 2015, The Astrophysical Journal, 815, 108
Otten, G., Vigan, A., Muslimov, E., et al. 2020, arXiv preprint arXiv:2009.01841
Otten, G. P., Snik, F., Kenworthy, M. A., et al. 2017, The Astrophysical Journal, 834, 175
Por, E. H., Haffert, S. Y., Radhakrishnan, V. M., et al. 2018, in Adaptive Optics Systems VI, Vol. 10703, International Society for Optics and Photonics, 1070342
Racine, R., Walker, G. A., Nadeau, D., Doyon, R., & Marois, C. 1999, Publications of the Astronomical Society of the Pacific, 111, 587
Rajan, A., Rameau, J., De Rosa, R. J., et al. 2017, The Astronomical Journal, 154, 10
Samland, M., Mollière, P., Bonnefoy, M., et al. 2017, Astronomy & Astrophysics, 603, A57
Skemer, A. J., Hinz, P., Stone, J., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, ed. C. J. Evans, L. Simard, & H. Takami, Vol. 10702, International Society for Optics and Photonics (SPIE), 78 – 85, doi: 10.1117/12.2314091
Skemer, A. J., Hinz, P. M., Esposito, S., et al. 2012, The Astrophysical Journal, 753, 14
Skemer, A. J., Marley, M. S., Hinz, P. M., et al. 2014, The Astrophysical Journal, 792, 17
Skemer, A. J., Hinz, P., Montoya, M., et al. 2015, in Techniques and Instrumentation for Detection of Exoplanets VII, Vol. 9605, International Society for Optics and Photonics, 96051D
Snik, F., Otten, G., Kenworthy, M., et al. 2012, in Modern Technologies in Space-and Ground-based Telescopes and Instrumentation II, Vol. 8450, International Society for Optics and Photonics, 84500M
Stolker, Bonse, M. J., Quanz, S. P., et al. 2019, A&A, 621, A59, doi: 10.1051/0004-6361/201834136
Stolker, T., Min, M., Stam, D. M., et al. 2017, Astronomy & Astrophysics, 607, A42

van Holstein, R. G., Snik, F., Girard, J. H., et al. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, Vol. 10400, International Society for Optics and Photonics, 1040015

Wagner, K., Stone, J., Dong, R., et al. 2020, The Astronomical Journal, 159, 252

Wang, J., Wang, J., Ma, B., et al. 2020, arXiv preprint arXiv:2007.02810

Wilson, J. C., Hinz, P. M., Skrutskie, M., et al. 2008, in Optical and Infrared Interferometry, Vol. 7013, International Society for Optics and Photonics, 70133A

4 Minimizing the polarization leakage of geometric-phase coronagraphs with multiple grating pattern combinations

Adapted from
D. S. Doelman, E.H. Por, G. Ruane, M. J. Escuti, F. Snik
Accepted by PASP (2020)

The design of liquid-crystal diffractive phase plate coronagraphs for ground-based and space-based high-contrast imaging systems is limited by the trade-off between spectral bandwidth and polarization leakage. We demonstrate that by combining phase patterns with a polarization grating (PG) pattern directly followed by one or several separate PGs, we can suppress the polarization leakage terms by additional orders of magnitude by diffracting them out of the beam. Using two PGs composed of a single-layer liquid crystal structure in the lab, we demonstrate a leakage suppression of more than an order of magnitude over a bandwidth of 133 nm centered around 532 nm. At this center wavelength we measure a leakage suppression of three orders of magnitude. Furthermore, simulations indicate that a combination of two multi-layered liquid-crystal PGs can suppress leakage to $< 10^{-5}$ for 1-2.5 μm and $< 10^{-10}$ for 650-800 nm. We introduce multi-grating solutions with three or more gratings that can be designed to have no separation of the two circular polarization states, and offer even deeper suppression of polarization leakage. We present simulations of a triple-grating solution that has $< 10^{-10}$ leakage on the first Airy ring from 450 nm to 800 nm. We apply the double-grating concept to the Vector-Vortex coronagraph of charge 4, and demonstrate in the lab that polarization leakage no longer limits the on-axis suppression for ground-based contrast levels. Lastly, we report on the successful installation and first-light results of a double-grating vector Apodizing Phase Plate pupil-plane coronagraph installed at the Large Binocular Telescope. We discuss the implications of these new coronagraph architectures for high-contrast imaging systems on the ground and in space.

4.1 Introduction

Direct imaging and spectral characterization of exoplanets constitutes an exciting yet immense challenge, set by the small angular separation of the exoplanet to the star and their extreme difference in brightness. For a solar-system analogue at 10 parsec observed at visible wavelengths, an earth-twin has a separation of 100 mas and a contrast ratio of $\sim 10^{-10}$. More massive self-luminous exoplanets and brown dwarfs in young systems have contrast ratios in the near-infrared that are less extreme, $\sim 10^{-5}$ for HR 8799b,c,d,e (Marois et al., 2008). Detecting and characterizing these young planets with direct imaging, already requires a large-aperture telescope, high-precision wavefront sensing and correction, a coronagraph with a small inner working angle to suppress the stellar diffraction halo by masking or diffracting stellar light, advanced observational and data-reduction strategies (e.g. angular differential imaging (ADI) and spectral differential imaging (SDI)), and intricate optical measurement systems, such as integral field spectrographs (IFSs) and polarimetry (Marois et al., 2006; Racine et al., 1999; Ruane et al., 2018; Jovanovic et al., 2018; Snik et al., 2018).

The unprecedented spatial resolution that the Extremely Large Telescopes (ELTs) currently under construction will offer will not only allow us to detect more exoplanets, but also characterize exoplanets found either by direct imaging or by other methods. Time-resolved, spectral and polarimetric measurements will provide crucial constraints on planet formation, atmospheric properties and eventually habitability and even habitation. From the ground the focus will be on habitable planets orbiting M-dwarfs as their contrast is more favorable, but a large telescope aperture is necessary to resolve their angular separation. A next generation of space telescopes with extreme contrast performance is necessary to observe rocky planets in the habitable zones around sun-like stars. Direct exoplanet characterization requires broadband observations to capture many photons and enable spectroscopy, and, therefore, coronagraph designs have to accommodate large spectral bandwidths. In addition to being able to spectrally characterize exoplanets, other advantages that broadband coronagraphs offer include increased observing flexibility and efficiency, diagnostics for wavelength-dependent instrument calibration, and spectral differential methods (Sparks & Ford, 2002; Snellen et al., 2015). Even current ground-based high-contrast imaging systems could benefit from an enhanced broadband coronagraph implementation (provided sufficient wavefront control), as they all have a coronagraph in the common path and split up to different wavelength channels afterwards (Macintosh et al., 2014; Lozi et al., 2018; Hinz et al., 2016; Males et al., ????; Beuzit et al., 2019). Many coronagraphs installed in these systems are optimized for a single observing band, and perform suboptimal when used in a broad spectral band or at different wavelengths.

4.1.1 Diffractive phase plate coronagraphs

A promising type of coronagraph for broadband observations with small inner working angles is the diffractive phase plate coronagraph. Phase-only solutions of-

ten provide the most optimal combination of inner working angle and throughput, both for focal-plane coronagraphs and pupil-plane coronagraphs (Mawet et al., 2009b; Por, 2017). Classical phase plates are chromatic by nature, but the so-called geometric phase (Escuti et al., 2016a) allows for the design and manufacturing of strictly achromatic devices, as the phase it applies only depends on the orientation pattern of the fast axis of a (optically flat) retarder. A geometric-phase coronagraph has two important properties: the coronagraphic mask applies a phase pattern that is independent of wavelength and the fraction of light that acquires this phase depends on the retardance of the mask. The first property demonstrates why geometric-phase coronagraphs are excellent candidates for broadband observations, while the second property can represent a major technical limitation. When the retardance of a mask deviates from half-wave, a fraction of light does not acquire the geometric phase, and thus produces a regular non-coronagraphic PSF. In this paper, we will refer to this fraction of light that does not acquire the geometric phase as "polarization leakage". Polarization leakage is one of the factors limiting the performance (Mawet et al., 2009a; Ruane et al., 2019). Most geometric phase elements (including coronagraphs) are manufactured with patterned birefringent liquid crystals. With a direct-write system (Miskiewicz & Escuti, 2014), any geometric-phase pattern can be applied with $\sim$1 μm resolution through an alignment layer that is deposited on any flat substrate. This orientation pattern imposed on the fast axis of consecutive liquid-crystal layers, which can be tuned to have half-wave retardance at a single wavelength (single-layer zero-order retardance), or to have close-to-half-wave retardance over a large spectral bandwidth by combining two or more liquid-crystal layers with different thicknesses and material properties like birefringence and twist (Komanduri et al., 2013).
Two types of geometric-phase coronagraphs have become popular: the Vector Vortex Coronagraph (VVC) and the vector Apodizing Phase Plate (vAPP).

4.1.2 The Vector Vortex Coronagraph

The VVC is one realization of the optical vortex coronagraph, a focal-plane coronagraph (Foo et al., 2005) that diffracts all on-axis light out of the beam onto the Lyot stop for an unobstructed circular telescope aperture. The VVC has a phase mask in the focal plane and an amplitude mask (Lyot stop) in the downstream pupil plane to block the diffracted on-axis light. The phase mask has an azimuthal phase ramp of $2\pi n$, where n is an integer indicating the "charge" of the coronagraph. A higher charge offer larger resilience against optical aberrations including tip/tilt errors, but also increases the inner working angle Ruane et al. (2017). Note that the VVC is very suitable for broadband implementation as its phase pattern is scale-invariant and therefore wavelength scaling of the PSF theoretically does not influence performance. The VVC has been installed on many telescopes and has been successful in detecting exoplanets and brown dwarfs (Serabyn et al., 2010; Absil et al., 2016; Serabyn et al., 2017). The polarization leakage of a VVC mask does not limit the reached contrast for these ground-based observations in single spectral bands for which the masks were optimized. Instead, the systems are lim-

ited by non-common path aberrations and adaptive optics residuals. However, for ground-based simultaneous multi-band observations at $\sim 10^{-5}$ raw contrast, limited by the Adaptive Optics residuals, and space-based single-band observations at $\sim 10^{-10}$ raw contrast, the leakage is still a limiting the attainable contrast.

4.1.3 The vector Apodizing Phase Plate coronagraph

The vector-Apodizing Phase Plate (vAPP) is a version of the Apodizing Phase Plate (APP) coronagraph, a single-optic pupil-plane coronagraph (Codona et al., 2006; Kenworthy et al., 2007). An APP modifies the phase in the pupil plane to create regions in the PSF where the star light is suppressed, so-called dark zones. As opposite circular polarization states create a PSF with a dark hole on opposite sides, the vAPP implements circular polarization splitting that produces two complementary dark holes. A truly broadband vAPP is obtained in combination with a Wollaston prism and a quarter-wave plate, but the non-coronagraphic PSFs corresponding to various polarization leakage terms degrade the contrast in the dark holes (Snik et al., 2012; Otten et al., 2014b). The polarization leakage can be separated from the coronagraphic PSFs by adding a grating pattern (= phase tilt) to the phase pattern, i.e. the "grating-vAPP" (gvAPP) (Otten et al., 2014a). At present, one grating-vAPP is installed in the MagAO/Clio2 instrument at the Magellan Clay telescope (2–5 μm) (Otten et al., 2017), one at the MagAO-X instrument (Miller et al., 2019) (550-1100 nm) at the same telescope, one in the SCExAO instrument at the Subaru telescope (1–2.5 μm) (Doelman et al., 2017), and several other gvAPPs are being designed/manufactured/commissioned at the moment. The grating introduces significant wavelength smearing for the two coronagraphic PSFs, reducing the simultaneous bandwidth of the gvAPP for standard imaging. This implementation is still applicable to integral-field spectroscopy, but the overall PSF structure is generally large in comparison with the typically limited field-of-view of an IFS. In addition, half of the exoplanet light observed with a gvAPP with a D-shaped dark zone ends up on the bright-side of PSF, reducing the effective planet throughput. This can be negated by using phase solutions for 360° donut-shaped dark holes. However, these solutions are still limited by wavelength smearing when using a grating or on-axis leakage when used without a grating. Furthermore, the inner working angle is larger than for 2×180° solutions.

4.1.4 Leakage mitigation strategies

Current leakage mitigation strategies have been applied mostly for the VVC, as the grating-vAPP is very resistant against leakage. For the VVC, previous efforts to reduce or mitigate the leakage for the VVC mask can be split into four approaches. The first approach involves the engineering of the device such that its retardance is half-wave to within the tolerances set within the operational spectral band. For liquid-crystal devices this can be achieved by adding more layers. These additional can be self-aligning (Komanduri et al., 2013), or separate structures that are aligned and fixed manually (Mawet et al., 2011; Roberts et al., 2019). The lat-

ter approach has recently been shown to reduce the leakage to less than $< 0.02\%$ for 10% bandwidth (Roberts et al., 2019) and around 0.1% for 20% bandwidth (Serabyn et al., 2019). Geometric phase patterns can also be imposed through form birefringence through sub-wavelength grating patterns, which for the charge-2 VVC leads to the implementation of the Annular Groove Phase Mask (AGPM) (Mawet et al., 2005). The retardance of such a device can be achromatized by tuning the depth and shape of the sub-wavelength grooves, and by combining different materials (Vargas Catalán et al., 2016). This technology can deliver polarization leakage of a few 10^{-3} over mid-infrared bands where this technology is readily applied (Jolivet et al., 2019).

Even after applying these advanced manufacturing methods the polarization leakage is filtered for space-equivalent extreme contrast experiments, by means of sandwiching the VVC between "circular polarizers" consisting of a linear polarizer and a quarter-wave plate (Mawet et al., 2010; Snik et al., 2014). This therefore constitutes a second method, although this implementation does complicate the optical configuration. Moreover, the polarizers and quarter-wave plates need to be of high quality to offer additional leakage suppression of $\sim 10^{-8}$ to reach the contrast necessary to directly detect Earth-twins at 10^{-10} contrast in the visible.

A third approach that has recently been revived in Ruane et al. (2019), where they use a scalar vortex instead of a vector vortex. A scalar vortex obviously has no polarization leakage and incident light acquires the same phase shift regardless of polarization state. But the major challenge is to achromatizing such a mask by using two different glass types, and optimizing the Lyot stop diameter and two deformable mirror shapes.

The fourth approach is most similar to the multi-grating concept presented in this paper, and adds a grating to the vector vortex phase pattern to yield a "forked grating". The forked grating separates the leakage spatially from the vortex beam. Only one orthogonal circular polarization state is selected by blocking the leakage and the opposite state. The diffraction of the grating is compensated in a different plane. This approach uses either a spatial light modulator with a computer generated hologram (Leach & Padgett, 2003), binary amplitude mask (Ruane et al., 2014; Kanburapa & Swartzlander, 2012) or two volume phase holograms (Mariyenko et al., 2005). These approaches reach high diffraction efficiency but filter $> 50\%$ of the light and require complex setups that are difficult to integrate in current setups.

In this paper, we present the multi-grating concept to minimize the effects of polarization leakage for both the VVC and vAPP coronagraphs. The fundamental principle of the multi-grating concept is that the main polarization leakage terms are removed through diffraction, while the coronagraphic PSFs are recombined on-axis in a single optic. The diffracted leakage terms are diffracted outside the beam or the useful field-of-view. It is a simple, yet powerful upgrade that is easily implemented in current and future high-contrast imaging systems, both on the ground and in space.

In Sect 4.2 we explain the multi-grating principle and demonstrate in simulation a leakage suppression of many orders of magnitude. We present our lab validation

of a double-grating element in Sect. 4.3. In Sect. 4.2.5 we introduce the coronagraph architectures of both a multi-grating focal-plane coronagraph (VVC) and a double-grating pupil-plane coronagraph (vAPP). We demonstrate the performance of a double-grating VVC in Sect. 4.4, and of a double-grating vAPP at the Large Binocular Telescope in Sect. 4.5. We discuss the implications on systems level of these novel coronagraph architectures in Sect. 4.6.

4.2 Double-grating diffraction theory

Both the VVC and the vAPP coronagraph are commonly manufactured as geometric phase holograms (GPHs) (Kim et al., 2015; Escuti et al., 2016a). In this section we will look into the properties of GPHs and how the geometry of phase patterns can be used to remove leakage. We simulate the performance for several double-grating combinations consisting of multi-layer liquid-crystal GPHs, and show drastically reduced polarization leakage over large spectral bands. Lastly, we characterize the influence of grating separation.

4.2.1 The geometric phase hologram

A geometric phase hologram is a half-wave retarder with a spatially varying fast-axis orientation. Circularly polarized light propagating through a GPH acquires geometric phase (or Pancharatnam (Pancharatnam, 1955, 1956)-Berry (Berry, 1984, 1988) phase), that depends on the fast-axis orientation. The space-variant Jones matrix of such a retarder in the circular polarization basis is given by

$$\mathbf{M} = c_V \begin{bmatrix} 0 & \mathrm{e}^{i2\chi(x,y)} \\ \mathrm{e}^{-i2\chi(x,y)} & 0 \end{bmatrix} + c_L \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{4.1}$$

Here $\chi(x, y)$ is the spatially varying fast-axis orientation, and both c_V and c_L are parameters that depend on the retardance $\Delta\phi$ (Mawet et al., 2009b; Ruane et al., 2019) and are given by

$$c_V = \sin\frac{\Delta\phi}{2},\ c_L = -i\cos\frac{\Delta\phi}{2}. \tag{4.2}$$

The first term in Eq. 6.4 describes the fraction of the light that acquires a geometric phase of

$$\Phi(x, y) = \pm 2\chi(x, y), \tag{4.3}$$

where the sign of the phase depends on the handedness of the incoming circular polarization. The second term describes the polarization beam, and is unaffected by the fast-axis orientation pattern. When the retardance is perfectly half-wave, i.e. $c_V = 1$ and $c_L = 0$, the two basis circular polarization states are converted to

$$\mathbf{RC_{out}} = \mathbf{MRC_{in}} = \mathbf{M}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \mathrm{e}^{-i\Phi(x,y)} \end{bmatrix}, \tag{4.4}$$

and,

$$\mathbf{LC_{out}} = \mathbf{MLC_{in}} = \mathbf{M}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}e^{i\Phi(x,y)}\\0\end{bmatrix}. \tag{4.5}$$

From these equations we derive four properties of the geometric phase hologram:

1. A GPH applies geometric phase that only depends on the local fast-axis orientation and is therefore independent of wavelength.

2. The applied phase has an opposite sign for opposite handedness of the incoming circular polarization state.

3. A GPH flips the circular polarization state.

4. When the retardance deviates from half-wave, the diffraction efficiency decreases and polarization leakage emerges.

Unpolarized light has no preferred state of polarization and contains on average equal amounts of left and right circular polarization. When unpolarized light travels through a GPH, we can define three distinct waves emerging from the GPH. They are called the primary (+), conjugate (-) and leakage (0) wave (Hasman et al., 2002; Kim et al., 2015; Escuti et al., 2016a). They correspond to the two basis vectors in the circular polarization basis, **LC** and **RC**, that acquire phase with opposite sign and the polarization leakage. We define the diffraction efficiencies for these waves as follows

$$\eta_+ = |\langle\mathbf{E_{in}}|\mathbf{LC}\rangle|^2 \sin^2(\Delta\phi/2), = |\langle\mathbf{E_{in}}|\mathbf{LC}\rangle|^2 |c_V|^2, \tag{4.6}$$

$$\eta_- = |\langle\mathbf{E_{in}}|\mathbf{RC}\rangle|^2 \sin^2(\Delta\phi/2), = |\langle\mathbf{E_{in}}|\mathbf{RC}\rangle|^2 |c_V|^2, \tag{4.7}$$

$$\eta_0 = \cos^2(\Delta\phi/2) = |c_L|^2. \tag{4.8}$$

Here $\langle\mathbf{A}|\mathbf{B}\rangle$ is the dot product of **A** and **B**. While the individual diffraction efficiency of the primary and conjugate wave depends on the input polarization state, the overall diffraction efficiency of a GPH, as used in this paper, is given by $|c_V|^2 = \eta_+ + \eta_-$. The diffraction efficiency is therefore independent of polarization state. Note that it is equivalent to define $|c_V|^2 = \eta_\pm$ when assuming fully circularly polarized light. The retardance can be tuned to be close to half-wave over large bandwidths by adding multiple layers with different thickness and twist (Komanduri et al., 2013; Escuti et al., 2016b). Efficiencies are typically ∼99% over spectral bandwidths as large as an octave, which implies a polarization leakage of ∼1%.

Classically, the fast-axis orientation pattern and the retardance of a GPH can be treated as two separate properties, i.e. the polarization leakage can be minimized independent of the fast-axis orientation pattern. The multi-grating concept is different from this classical idea to minimize polarization leakage because it uses the geometry of the phase pattern instead of tuning the retardance more precisely by adding more liquid-crystal layers.

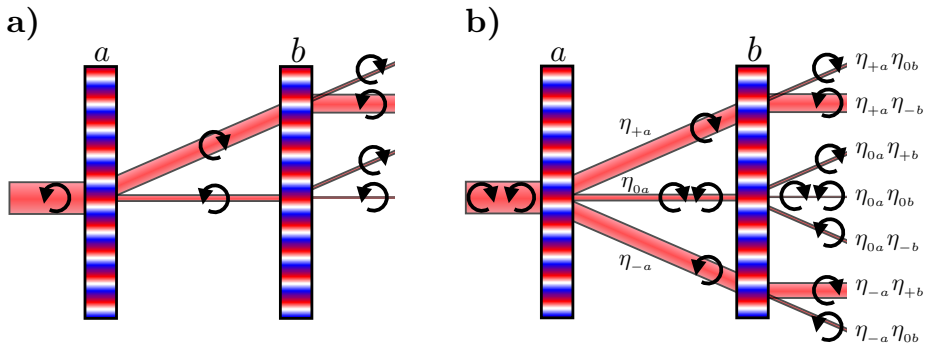


Figure 4.1: Schematic of the propagation of 100% circularly polarized light (left) unpolarized light (right) through a double-grating element. The thickness of the beams correspond to the amount of light in each beam, assuming close-to $\lambda/2$-retardance for both polarization gratings. The main beams are diffracted back on-axis because of the polarization flip that occurs when propagating through a GPH.

4.2.2 The double-grating element

A polarization grating (PG) is a GPH with a uniformly rotating fast-axis orientation in a defined direction and a constant orientation in the orthogonal direction (Oh & Escuti, 2008; Packham et al., 2010). The PG is therefore a GPH that applies a phase ramp in this defined direction. A PG splits two circular polarization states, diffracting them in opposite direction. One special property of a PG is that it has "polarization memory": contrary to regular gratings, a fully circularly polarized beam will in its entirety be diffracting in one particular order, barring polarization leakage. This is a direct consequence of Eq. 4.6 and 4.7.
A double-grating element introduced in this paper, consists of two individual GPHs. The first GPH has an orientation pattern corresponding to a coronagraphic pattern, e.g a VVC or vAPP, with a coherently added polarization grating pattern. We call this GPH the coronagraphic polarization grating (CPG). The second GPH is a normal polarization grating that has the same pattern as the polarization grating that was added to the CPG. A schematic of how circularly polarized light propagates through a double-grating element is depicted in Fig. 4.1. Here, the CPG is GPH a, and the PG is GPH b.
When a beam with a single circular polarization state propagates through the CPG, the main beam is diffracted and a minor leakage term propagates on-axis. The PG diffracts the main beam back on-axis, as the PG patterns are equivalent, but the circular polarization state has been flipped by the CPG. The real benefit of the double-grating concept is demonstrated by what happens with the leakage term of the CPG. This leakage term is mostly diffracted out of the beam, such that the remaining on-axis leakage is further suppressed by a factor $\eta_{0a}\eta_{0b}$.

For unpolarized light, both the primary and the conjugate beam of the CPG are diffracted back on-axis by the PG, given the polarization memory of PGs. In conclusion, a double-grating element suppresses leakage by diffracting most of the leakage off-axis while diffracting the main beam twice in opposite direction with high efficiency, keeping the main beam on-axis.
If the retardance of both the CPG and the PG are close to half-wave, the diffraction efficiencies $(\eta_{+a,b} + \eta_{-a,b}) \approx 1$ and $\eta_{0a,b} \ll 1$. As the leakage is always the product of the two leakage factors $\eta_{0a}\eta_{0b}$, the leakage for two identical liquid-crystal films is suppressed by orders of magnitude. As an example, two polarization gratings with each 1% leakage will have a on-axis polarization leakage as low as 10^{-4}. We emphasize that this is achieved with phase pattern geometry and not by optimizing the manufacturing process to optimize the retardance of individual optics beyond the ∼1% leakage performance offered by standard liquid-crystal techniques.
Combining gratings this way is only possible with polarization gratings, and no other type of grating, because PGs operate on circular polarization states and diffract into a single order. This polarization memory is key to make this setup highly efficient.
Without polarization splitting, the double-grating concept constrains the design of the phase pattern. This constraint comes from the property that a double-grating element introduces phase with opposite sign for both circular polarization states. Phases with opposite sign in the pupil plane correspond to PSFs in the focal plane that are point-symmetric mirror images. If the PSFs are not point-symmetric, both circular polarization states will contaminate each other. This limits the use of the double-grating concept to point-symmetric patterns like the vortex phase, or $0 - \pi$ phase patterns, e.g. in the pupil plane to produce PSFs with 360° dark holes.

4.2.3 Simulated broadband performance of double-grating elements

We simulate the on-axis terms for different liquid-crystal recipes to demonstrate the broadband performance of a double-grating element. Analogous to the first approach of minimizing polarization leakage in Sect. 4.1, we optimize the liquid-crystal recipes to minimize leakage for both the CPG and the PG in the double-grating element, assuming a flat spectrum and fully circularly polarized light. These recipes consist of multiple individual liquid-crystal layers, each with a thickness and chiral twist, that are arranged into a monolithic film, i.e. a multi-twist retarder (MTR) (Komanduri et al., 2013; Kim et al., 2015; Escuti et al., 2016b). We tune the retardance by adjusting the thickness and chiral twist for each layer in a two-layered retarder (2TR) or three-layered retarder (3TR). The recipes of the CPG and the PG are optimized simultaneously using simulated annealing. In these simulations, the coronagraphic pattern is set to zero, such that only two PGs remain. We investigate solutions for three different regimes, each with different applications. First, we optimize for extremely large bandwidths (> 90%) with moderate leakage suppression of $< 10^{-3}$ that are capable of feeding multi-

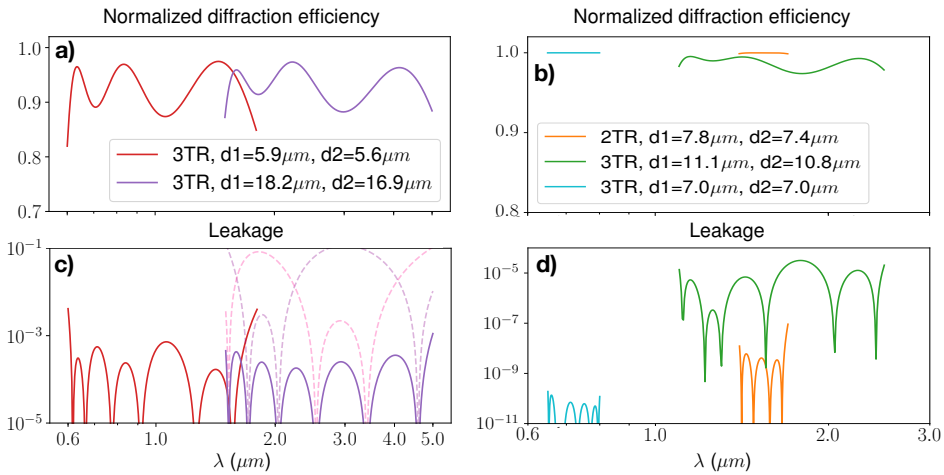


Figure 4.2: **Top:** Simulated diffraction efficiencies for five double-grating elements, with the average diffraction efficiency being higher than 90%. The combined thickness of the layers are shown with d1 and d2, for the CPG and PG respectively. **Bottom:** Leakage as function of wavelength for the same double-grating elements. The dotted lines indicate the leakage of both PGs making up the double-grating element. These single PGs have different liquid-crystal recipes to minimize the total leakage.

ple ground-based instrument arms. We show two solutions in the left panels of Fig. 4.2. Here we define the bandwidth as $\Delta\lambda/\lambda_0$, where $\Delta\lambda$ is the total bandwidth in nm and λ_0 is the central wavelength. Second, we develop a recipe for good leakage suppression ($< 10^{-5}$) for a bandwidth up to (90%). Third, we explore possible recipes for small bandwidths (20%) with extremely low leakage ($< 10^{-8}$) that could be used for a single band in a space-based instrument. Solutions of the second and third regimes are shown in the right panels of Fig. 4.2.

For the largest bandwidths, we optimize two double-grating recipes that are shown in Fig. 4.2(a) and Fig. 4.2(c). These recipes have bandwidths of 100% or more, the first one ranging from 0.6 μm to 1.8 μm (red) and the second one ranging from 1.5 μm to 5 μm (purple) with a leakage $< 10^{-3}$. Optimizing the recipes of the GPH and PG simultaneously yields non-trivial solutions where the two elements have different recipes. This is illustrated by the dashed lines that show the individual leakage terms for the PG and GPH. While the combined leakage is $< 10^{-3}$ at all wavelengths, the individual recipe for the GPH or the PG can have as much as 10% leakage for certain wavelengths, which is then compensated by the other element that has less 1% leakage at that wavelength. This compensation is critical for extending the bandwidth beyond the broadband range that single-element 3TRs reach, which is typically up to 90% for a leakage lower than 3% (Doelman et al., 2017). A downside is that the increased leakage for single el-

ements of the double-grating reduces diffraction efficiency. Therefore, the average combined diffraction efficiency is reduced to 93% for these two designs.
For the second regime with wavelength ranges covered by a single ground-based instrument arm like an IFS, the goal is to suppress the leakage PSF below the speckle limit for extreme adaptive optics systems, which is around 10^{-5} in good circumstances. We optimize the double-grating recipes for the SCExAO (Lozi et al., 2018) instrument, operating between 1.1 μm and 2.5 μm combined with the integral-field spectrograph CHARIS (Groff et al., 2017). The grating-vAPP that is currently installed in this instrument has an average leakage of 2% over this bandwidth (Doelman et al., 2017). For a vector vortex coronagraph this would limit the performance. However, using the double-grating concept reduces the leakage by three additional orders of magnitude, going from 2% to $< 0.001\%$, while having a similar diffraction efficiency ($> 97\%$), as shown with the 3TR recipe (green) in Fig. 4.2(b) and Fig. 4.2(d).
For space-based instruments, extremely low leakage is required. One advantage is that the bandwidth is limited, as the instantaneous wavefront correction with two deformable mirrors at these extreme contrast levels is limited to $\sim$20% (HabEx Study Team, 2019). As a first example, using two 2TRs for less complexity, it is possible to suppress the leakage from 1.5 μm to 1.8 μm (H-band, orange) to 10^{-8}. In addition, we optimize for a wavelength range of 650-800 nm, as visible bands such as this one are planned for future space missions (HabEx Study Team, 2019). The transmission and leakage are shown in Fig. 4.2(b) and Fig. 4.2(d), indicating that two three-layer stacks have an average polarization leakage of less than 10^{-10} and an average diffraction efficiency higher than 99.9%. This shows that a double-grating element can suppress the leakage to the required levels. Nevertheless, we introduce triple-grating solutions in Sect.4.6.1 to cover the entire visible spectral range with one coronagraph and enable extremely high contrast performance at any smaller band within that range. Note that it may be sufficient to suppress the first Airy ring of the leakage PSF to 10^{-10}, not the core itself. As the first Airy ring is typically at $\sim 10^{-2}$ of the core intensity, this could reduce the complexity of liquid-crystal recipes that are used.

4.2.4 The effect of non-zero grating separation

From Fig. 4.1 it is apparent that the polarization gratings split the beam in two separated main coronagraphic beams with opposite circular polarization. For the pupil-plane double-grating vAPP, a small shift of the two pupil on the order of a few micron is generally inconsequential, whereas for the focal-plane double-grating VVC this broadens or even splits up the off-axis PSF. This effect is further elaborated in Fig. 4.3.
We derive the beam shift by using the grating equation with $m = 1$ to calculate the total lateral beam shift, Δ, to be

$$\Delta = 2d \tan\left(\arcsin\left(\frac{\lambda}{P}\right)\right). \tag{4.9}$$

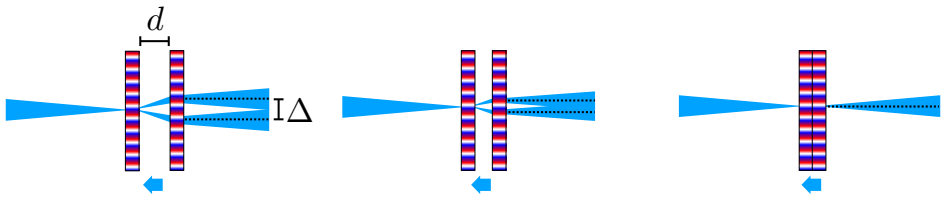


Figure 4.3: Any finite distance, d, between two polarization gratings generates a shift, Δ, between the primary and the conjugate beam. This separation decreases linearly with distance between the two gratings.

Here, d is the separation between the plates, P the period of the polarization grating, and the factor two comes from the two circular polarization states being diffracted in opposite direction. For a pupil plane the effect is insignificant as Δ is orders of magnitude smaller than the pupil diameter. However, the grating separation in the focal plane quickly becomes an issue, i.e. when Δ is a significant fraction of λ/D. We can minimize Δ by maximizing polarization grating period and minimizing the distance between the gratings. The optical system determines the maximum grating period, as the leakage needs to be diffracted out of the beam by a full pupil distance in the next pupil plane. Because the geometric phase is applied very locally (Escuti et al., 2016a), the gratings can be very close together and still operate independently. Moreover, the minimal distance of the gratings is set by the thickness of an optical adhesive that is used to glue both gratings together. Therefore, a finite shift is unavoidable. Even so, most high-contrast imaging systems have large F-numbers and in practice this shift is much less than the width of the point-spread function. As an example, we take an optical system with $\lambda = 1500nm$, a grating period of $30\mu m$, a separation given by a glue layer of 50 μm and an F-number of 50, which gives a separation of 0.067 λ/D. For smaller F-numbers, significant wavelength-dependent splitting will occur for all the objects in the field, decreasing point-source detectability and astrometric and photometric accuracy. In Sect. 4.6.1 we introduce multi-grating solutions that solve this PSF splitting issue by design.

4.2.5 Multi-grating coronagraph architectures

In this section, we discuss where the leakage terms are blocked for a double-grating implementation of both the VVC and the vAPP. Schematic drawings of coronagraphic systems with a focal-plane phase coronagraph like the VVC, and with a pupil-plane phase coronagraph like the vAPP are presented in Fig. 4.4 for a classical and double-grating configuration. The simulated point-spread functions include no aberration, a planet at a level of 10^{-4}, and we assume a leakage of 1% for each element.

For the VVC in Fig. 4.4(a), most of the leakage in the classic configuration passes through the Lyot stop and is imaged onto the detector. The double-grating

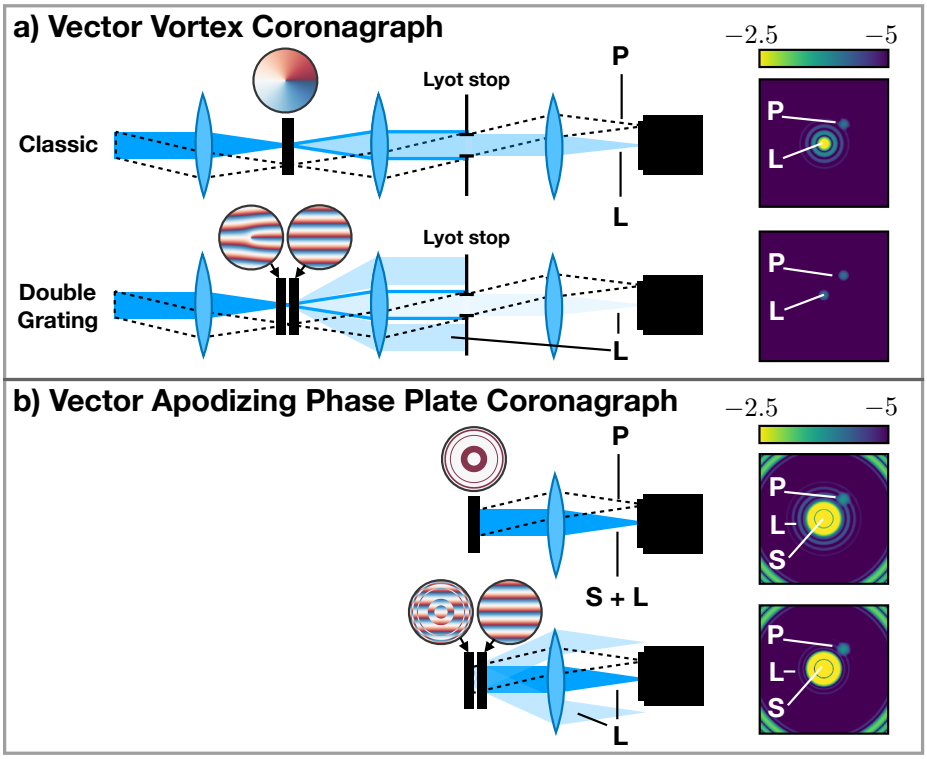


Figure 4.4: Comparison between the Vector Vortex coronagraph, a), and the vector Apodizing Phase Plate coronagraph, b), and their double-grating counterparts. The schematics show the optical paths of the diffracted leakage terms on a system level. The simulated point-spread functions demonstrate the effect of removing additional leakage when the double-grating coronagraphs are used. We annotate the planet PSF, P, the coronagraphic leakage, L, and the stellar PSF, S.

VVC however, diffracts ∼99% of the leakage term outside of the beam such that it is blocked by the Lyot stop, reducing the on-axis leakage to the level of 10^{-4}. To diffract the non-coronagraphic leakage terms out of the beam, we require a grating period of $P < f\lambda/D$, where λ is the wavelength, and f and D are the focal length and the diameter of the lens before the double-grating element, respectively. The leakage terms are non-coronagraphic pupils, and have minimal intensity outside of their diameter. For leakage terms that have the coronagraphic pattern imprinted and therefore have light diffracted outside of the pupil plane, the separation needs to be larger, based on how much light is tolerable in the actual pupil of the main beam traveling through the Lyot stop.
The regular (no-grating) vAPP with an annular dark hole has a leakage PSF that is imaged onto the same location as the apodized PSF. This leakage term introduces extra light in the dark hole at the 10^{-4} level close to the inner edge of the dark zone at $3\lambda/D$, as shown by the Airy rings present in the PSF of Fig. 4.4(b). The double-grating vAPP diffracts the leakage terms beyond the outer edge of the dark zone when the grating period $P < D/(2\Omega)$, where D is the diameter of the pupil of the vAPP and Ω is the outer working angle in λ/D. Here we take into account that one of the two diffracted leakage terms is also apodized, see Fig. 4.1. The double-grating version of both coronagraphs clearly reduce the signal of the leakage at the location of close-in companions.

4.3 Experimental results of double-grating implementations

In the previous sections we derived two properties of the double-grating concept. First, we mathematically showed that the on-axis polarization leakage is suppressed by additional orders of magnitude, and second, we derived that small grating separations do not generate significant polarization splitting for large F-numbers. Here, we demonstrate these two properties with lab measurements. We characterize the double-grating performance with two polarization gratings, by measuring all leakage terms individually. Furthermore, we investigate the impact of grating separation on the polarization splitting by measuring the splitting as function of separation.

4.3.1 Lab setup and double-grating manufacturing

We characterize the performance of the double-grating elements by placing two PGs in an imaging setup that is fed with a single-mode fiber. There are two configurations with different locations of the double-grating element. These two configurations of the setup are shown in Fig. 4.5. We reimage the fiber onto the camera with an in-between focal plane. The camera is a FLIR Chameleon3 2.8 MP mono camera operating at 12 bits. We define a pupil plane before the double-grating element with an iris. A second pupil plane contains the Lyot stop if the double-grating VVC is used. For both configurations, the single mode fiber has an

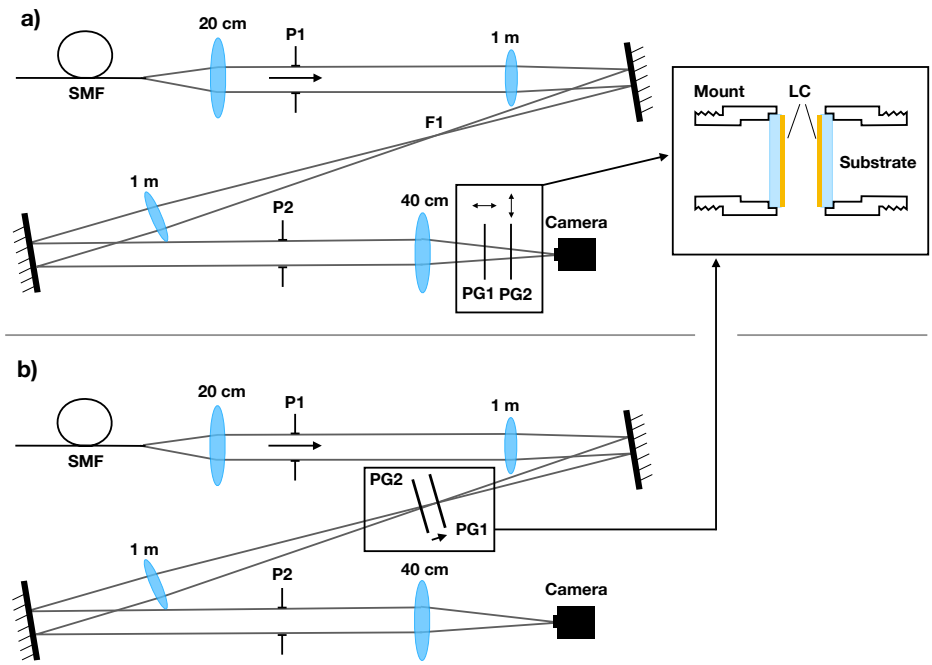


Figure 4.5: The two configurations of the lab setup used to measure the leakage as function of wavelength and PSF separation as function of grating separation respectively.

injection unit for broadband light from an Ocean Optics HL 2000 halogen lamp and two Thorlabs laser diodes (CPS532a and CPS635S), operating at 532 nm and 635 nm. For the halogen lamp we use Thorlabs FKB-VIS-10 narrowband filters (FWHM = 10 nm) spaced at 50 nm intervals between 500 nm and 800 nm. We use a polarizer before the first pupil plane (P1) to equalize the intensity in the two circular polarization states. The intensity is equal because linear polarization can be decomposed in equal amounts of left and right circular polarization. In the first configuration we measure the polarization leakage as function of wavelength. The second configuration allows us to determine the PSF splitting for non-zero distance between the PGs and to evaluate the performance of the double-grating vector vortex coronagraph.

We manufactured two identical PGs on two different substrates. The PGs are printed on 2 inch glass (D263) windows with a thickness of 0.7 mm, have a period of 17 micron, a pixel size of 1 micron and are 10 mm by 10 mm in diameter. The PG pattern is written in a photo-alignment layer using the direct-write laser scanning system (Miskiewicz & Escuti, 2014; Kim et al., 2015) using a solid-state 355 nm laser. The retardance of the liquid-crystal pattern is optimized using three

sublayers of birefringent non-twisting self-aligning liquid-crystals to form a single thin film (1TR) that is cured with a 365 nm LED (Komanduri et al., 2013). We measured the layer thickness to be approximately 1.1μm using an ellipsometer, corresponding to a half-wave retardance for 532 nm. A more detailed description of the manufacturing of these PGs can be found in Doelman et al. (2019). The two substrates are glued into a mount with a small groove, facing outward. The substrate extends beyond the edge of the mount. This way, two liquid-crystal surfaces can touch when gently pressed together. The separation between the mounts can be regulated using translation stages.

4.3.2 Characterization of the double-grating polarization leakage

To measure the double-grating polarization leakage, we use configuration a, with crossed polarization gratings. When the PGs are crossed, all leakage terms are imaged onto different locations in the focal plane. We separate the PGs by a small amount ($\sim$ 10 mm), and increase the distance to the camera until the leakage PSFs are well separated in the focal plane. The resulting images at 532 nm and 800 nm are shown in Fig. 4.6.

These images are used to reconstruct the individual diffraction efficiencies for both PGs using aperture photometry. To remove transmission effects, we normalize the intensities on the total flux measured in all 8 spots for each wavelength. The PGs are measured simultaneously in a stack, which means that we measure combinations, such as $\eta_{-,-} = \eta_{-,1}\eta_{-,2}$. Here 1 and 2 are used to distinguish between PG1 and PG2 respectively. To recover the diffraction efficiency of the first grating, $\eta_{-,1}$, from the image, we add three terms

$$\eta_{-,-} + \eta_{+,+} + \eta_{-,0} = \eta_{-,1}\eta_{-,2} + \eta_{+,1}\eta_{+,2} + \eta_{-,1}\eta_{0,2}. \tag{4.10}$$

We use that $\eta_{+,a} = \eta_{-,a}$ for linear polarization and normalize the flux of the PSFs such that

$$\eta_{-,-} + \eta_{+,+} + \eta_{-,0} = \eta_{-,1}\left(\eta_{-,2} + \eta_{+,2} + \eta_{0,2}\right) = \eta_{-,1}. \tag{4.11}$$

For high-contrast imaging we are interested in the power of the leakage terms of the gratings compared to the first-order diffraction efficiency. This relative intensity sets the contrast of the leakage term compared to a stellar PSF. We calculate these relative leakage intensities for both polarization gratings and the double-grating and these are plotted in linear scale (left) and logarithmic scale (right) in Fig. 4.7. We fit these relative leakage intensities for both polarization gratings using a Mueller matrix model of a twisted nematic cell with zero twist and varying thickness and fit the dispersion of the birefringence using a Cauchy expansion. The model for the double-grating leakage is the multiplication of the models for both PGs. The fitted models are in good agreement with both the measured individual leakage intensities and the combined double-grating leakage.

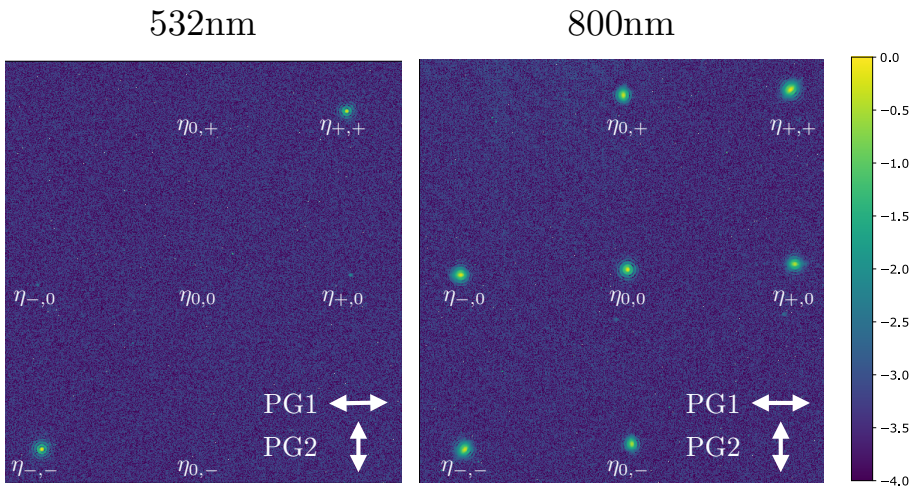


Figure 4.6: Measured lab PSFs of a single source imaged through crossed polarization gratings that were optimized for 532 nm. The PGs are crossed to separate all leakage terms onto individual locations in the focal plane. The spot identifiers, e.g. $\eta_{+,+}$, correspond to the primary beam with efficiency $\eta_{+a}\eta_{+b}$. **Left:** The diffraction efficiency of each grating is optimized for 532 nm such that 99% of the light is imaged on the combined first order of these two gratings. **Right:** The diffraction efficiency at 800 nm is less than 50%, creating all zero and first order diffraction spots of roughly equal brightness. Note that the absence of $\eta_{-,+}$ and $\eta_{+,-}$ is due to the polarization splitting of the first grating.

The fitted layer thickness are 1.07 μm for PG1 and 1.14 μm for PG2. The double-grating element reduces the leakage significantly over the full wavelength range and even by more than an order of magnitude between 500 nm and 633 nm. Similarly, we infer from the models the bandwidth where the leakage is lower than 1%. The individual PGs reach a leakage suppression lower than 1% over 46 nm bandwidth, compared to 150 nm for the double-grating. In addition, at 532 nm the double-grating leakage is lower than 1×10^{-5}, compared to 6×10^{-3} or 7×10^{-4} for PG1 and PG2 respectively. These results show that by using two polarization gratings in succession it is possible to significantly reduce the polarization leakage in the on-axis beam.

4.3.3 Separation of the double-grating PSF

We measure the splitting from non-zero distance between the two PGs using configuration b shown in Fig. 4.5. We inject the 532 nm laser in a single mode fiber that is reimaged on the two polarization gratings with an F-number of 100.

The distance between the gratings is controlled with a translation stage. We start with a separation of zero mm, with the two gratings gently pressed together.

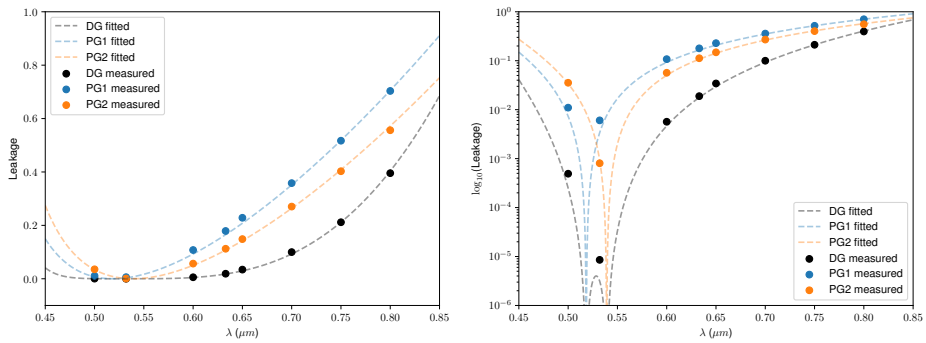


Figure 4.7: **Left:** Lab data of the relative leakage intensity with fitted models as function of wavelength in linear scale. **Right:** The same data and models in logarithmic scale. The relative double-grating leakage is less than 1% over a bandwidth of 133 nm and less than 10^{-5} at 532 nm.

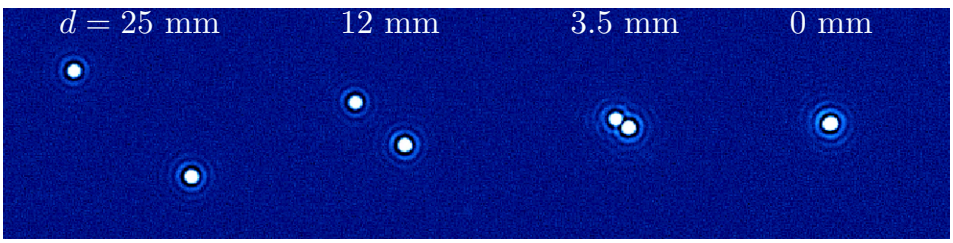


Figure 4.8: Lab images of the separation of the left- and right-circularly polarized beams when two polarization gratings are separated by a distance of 25, 12, 3.5 and 0 mm.

We take 10 images that are background-subtracted and median-combined. The sampling between 0 mm and 25 mm separation changes and has three different step sizes. Between 0 and 2 mm, 2-5 mm, 5-25 mm, the step sizes are 50 micron, 100 micron, and 500 micron respectively. The resulting images are shown in Fig. 4.8 for separations of 25, 12, 3.5 and 0 mm. When pressed together, the two PSFs overlap to a very high degree. To quantify the PSF separations Δ as a function of plate separation, we model both PSFs using a non-aberrated Airy pattern and fit these to the data simultaneously. We plot the measured Δ as function of separation in λ/D in Fig. 4.9. The 1σ error bars of the data are calculated from the jitter of the center of the PSFs. Using the plate scale, we convert the calculated separation from Eq. 4.9 to λ/D.

The calculated and measured separation are in good agreement with each other beyond 0.5 mm separation. However, there are hints of a modulation on the order of $0.5\lambda/D$, which might be caused by interference between the two PSFs. To confirm the slope, we fit a line through the data and recover the slope of

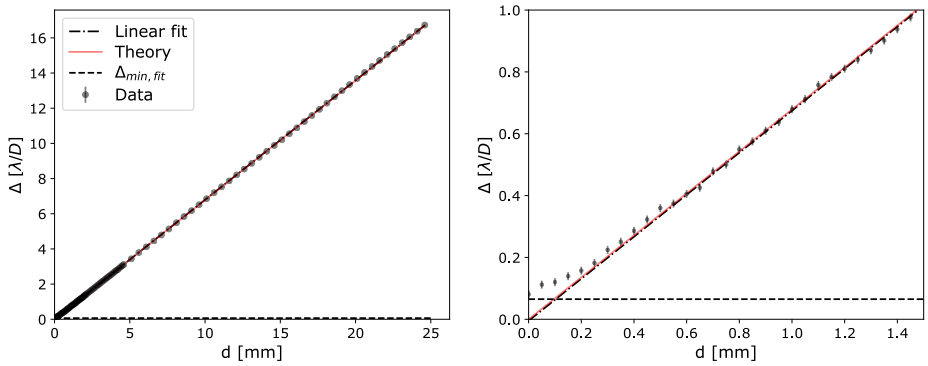


Figure 4.9: **Left:** PSF separation measured as function of grating separation, with 1σ error bars. **Right:** The measured separations at the closest distances between the gratings. The data and the linear fit to the data follow the theoretical to a very high degree. The dashed line shows the minimal separation the two-PSF model can distinguish when fitted to a single PSF.

0.6795 ± 0.0003 (λ/D)/mm (1σ) and an intercept of -0.005 ± 0.003 λ/D. The fitted slope is not within 1σ of the theoretical slope of 0.6788 λ/D/mm. One explanation for the inconsistency between the data and the theory could be that the plates were not perfectly parallel, or that the grating constants of the two PGs are not exactly the same. When the gratings are closer, the determined PSF separation Δ asymptotically reaches $0.1\lambda/D$. This is most likely an artifact of the model fitting that cannot distinguish the two separate PSFs anymore for small separations. Furthermore, the model fitting is affected by interference between the two PSFs. We test the model fitting sensitivity by fitting the two PSFs to a single PSF of one circular polarization state for the 30 largest separations. The separations are larger than 0 and the median value, $0.065\lambda/D$, is shown in Fig. 4.9 as a dashed line. It is therefore unclear if the the theoretical separation of exactly zero is actually achieved.

4.4 Experimental results for a double-grating Vector Vortex Coronagraph

We have shown that a combination of two polarization gratings can be used to reduce on-axis polarization leakage. In this section we show that the double-grating concept can be used to reduce the impact of polarization leakage on a Vector Vortex Coronagraph (VVC). The first polarization grating used in previous experiments is on the same substrate as a charge-2 and charge-4 vortex pattern with an added grating, i.e. a forked grating. Therefore, these coronagraphs have the same spectral retardance profile as the first grating in the experiments in Fig. 4.7. We image the two vortices and the second grating under a microscope between crossed

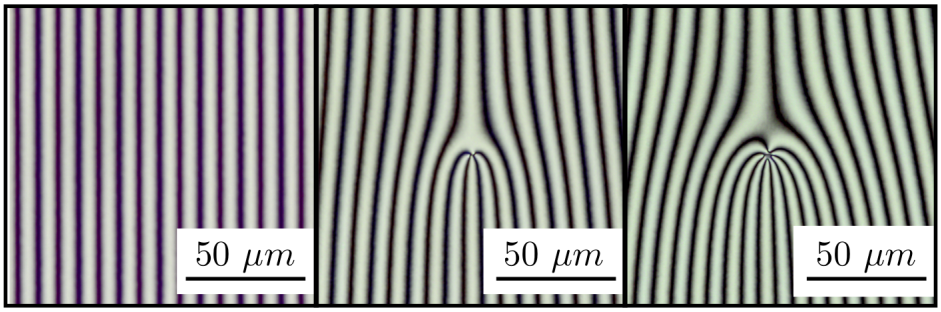


Figure 4.10: Microscopic images through crossed polarizers of the polarization grating and VVC + grating pattern charge 2 and 4.

polarizers. The results are shown in Fig. 4.10. Both the charge-2 and charge-4 vortex have a small central defect of < 1 μm and 2 μm, respectively, and no other defects are observed. We conclude that the produced VVCs are of high quality.
We test the performance of the double-grating VVC charge 4 in the setup using configuration b shown in Fig. 4.5, with a Lyot stop at P2. From the leakage intensity measurements it is expected that the polarization leakage for 532 nm at the 10^{-5} level will not limit the coronagraphic performance, given the limited optical performance of our setup. We measure the on-axis suppression, i.e. the coronagraphic PSF, and off-axis point spread function with a grating separation of zero. The off-axis PSF is shifted by $> 10\lambda/D$, so it is a good approximation for a non-coronagraphic PSF. We conduct these on-axis and off-axis measurements at 532 nm and 633 nm. Note that this is a passive setup, there is no active element implemented that measures and corrects the wavefront and the performance is limited by aberrations in the system.
The results are shown in Fig. 4.11. For 532 nm, we reach a contrast of 10^{-5} at $2.5\lambda/D$, which is in good agreement with other VVC experiments with a passive setup using polarization filtering (Mawet et al., 2009a; Delacroix et al., 2013). To compare this result with a standard VVC, we use the measured leakage of PG1 of 6×10^{-3} at 532 nm, which has been manufactured on the same substrate as the coronagraphic polarization grating. From this comparison it seems that the gain of using a double-grating is minimal. However, the contrast is no longer limited by leakage, which can not be corrected with adaptive optics, so adding active wavefront control would improve the contrast. Furthermore, there is a factor 2 gain in throughput as we do not filter one of the two circular polarization states before or after the VVC mask. Lastly, the leakage term that is used for comparison is a factor 2-4 lower than reported broadband liquid-crystal recipes (Otten et al., 2017; Doelman et al., 2017).
At 633 nm, the dgVVC coronagraphic PSF is dominated by the leakage term at the predicted 10^{-2} level. However, the this is an order of magnitude lower than the estimated leakage of the VVC. This clearly demonstrates that the double-grating

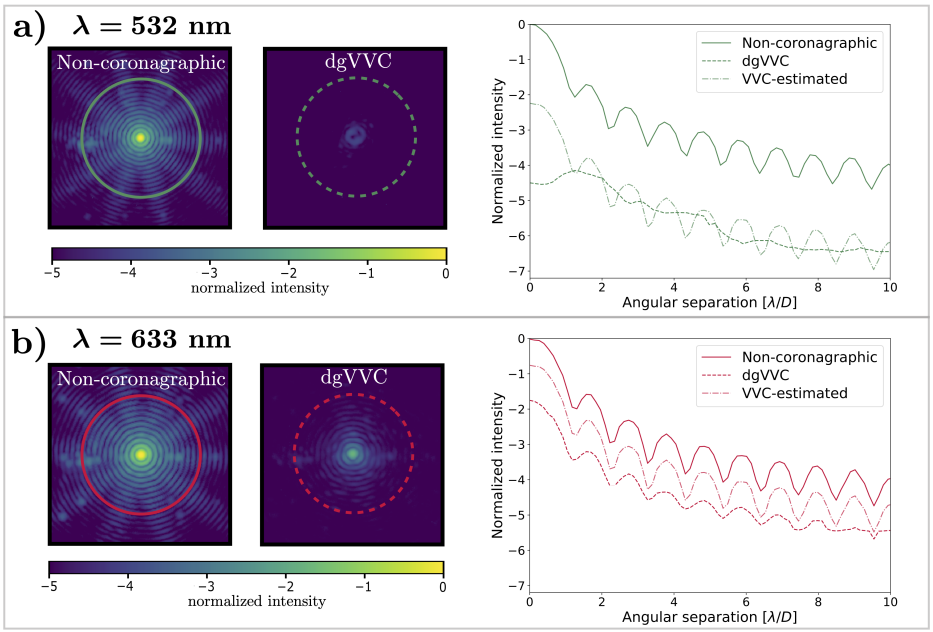


Figure 4.11: Measured coronagraphic and non-coronagraphic point-spread functions for the charge 4 double-grating VVC (left) and their azimuthally-averaged normalized intensities (right). We added an estimated leakage for a standard VVC (no gratings) as a reference. This VVC leakage is the non-coronagraphic PSF scaled to the level of the leakage of PG1, measured in Sect. 4.3.2.

element suppresses leakage. In addition, it demonstrates that the gain of using a double-grating is only as good as the leakage of the individual liquid-crystal recipes. In this case, the single layer recipe provides only good leakage suppression around 532 nm.

To demonstrate that improving the liquid-crystal recipe would improve the contrast, we separate the gratings by a distance of 20 mm. Equivalent to Sect. 4.3.2, this separates the leakage PSF from the coronagraphic PSFs. The resulting PSF and azimuthally averaged intensities are shown in Fig. 4.12. For comparison, we added the results from in Fig. 4.11b) in gray. The comparison shows that the measured on-axis coronagraphic PSF in the non-separated case, dgVVC $\Delta = 0$ has a similar intensity as the leakage term for $\Delta > 0$. The difference could be explained by the interference between the vortex PSF and the leakage PSF. By looking at the separated coronagraphic PSFs in Fig. 4.12, we find that the performance of the coronagraph at 633 nm without leakage is comparable to the suppression at 532 nm. This shows that currently the contrast of the dgVVC at 633 nm is limited by the leakage, which is already an order of magnitude lower than for a single-element. We summarize that a further improved leakage suppression at 633nm

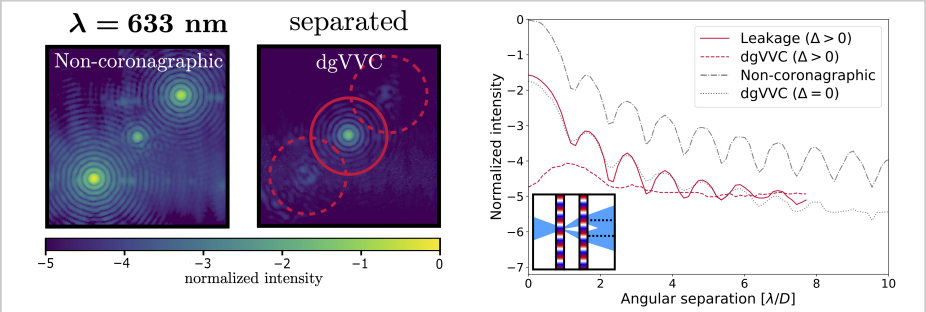


Figure 4.12: Measured coronagraphic and non-coronagraphic point-spread functions for the charge 4 double-grating VVC (left), with the substrates separated by 20 mm, and the azimuthally-averaged normalized intensities (right). In gray are the non-separated measurements, as shown before in the previous figure. The non-zero distance between the substrates separates the coronagraphic and leakage PSFs on the detector. The two off-axis (coronagraphic) dgVVC radial profiles are averaged.

would enhance coronagraphic performance. A double-grating vector-vortex with two multi-layered liquid-crystal recipes can achieve good performance for a larger bandwidth. Assuming the $1-2.5\mu m$ 3TR recipe shown in Fig.4.2 (green) a double-grating VVC could suppress this leakage to the measured level of 10^{-5} for $\sim$90% bandwidth. In conclusion, these lab results demonstrate that the double-grating concept works for the VVC and may be key to manufacturing a high-performance broadband VVC mask.

4.5 The double-grating vector-Apodizing Phase Plate for the Large Binocular Telescope.

In this section, we present the design and the first on-sky images of the double-grating vector-Apodizing Phase Plate coronagraph for the L/M-band (3-5 μm) InfraRed Camera (LMIRcam) (Skrutskie et al., 2010) at the Large Binocular Telescope (LBT). The design of the phase pattern of the double-grating vAPP for the LBT aims to minimize the stellar diffraction halo by generating an annular dark zone that extends from close to the core of the stellar PSF, to beyond the control radius of the deformable secondary mirror ($\sim 13\lambda/D$). This double-grating vAPP is compatible with the ALES integral-field spectrograph (Skemer et al., 2015; Skemer et al., 2018), as its single combined PSF well matches its small field-of-view, and it offers stable coronagraphic performance over the entire wavelength range, even extending down to K-band at 2 μm. This is enabled by the liquid-crystal recipe described in (Doelman et al., 2017). The design of the vAPP phase pattern is computed using the global optimizer described in Por (2017) (Por, 2017)

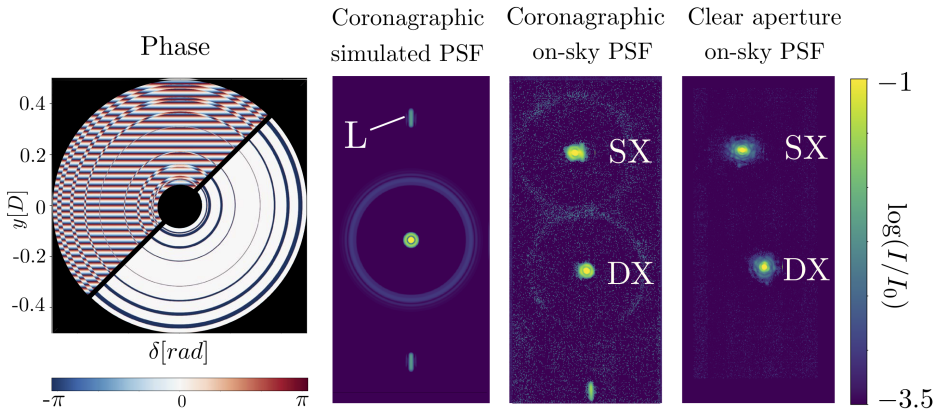


Figure 4.13: **Left:** vAPP phase pattern with and without grating. **Center:** Simulated broadband PSF of the vAPP, showing the diffracted leakage (L). **Right:** First light of the double-grating vAPP at LBT, showing the PSFs of both telescope apertures, SX and DX in L-band. Note that the PSFs can be positioned such that the leakage terms are not imaged in the dark zone of the other PSF.

Table 4.1: Fitted intensity and leakage values for L-band for both vAPPs and a clear aperture. The intensity and leakage are fitted using least-squares fitting with unaberrated broadband PSF models to the clear aperture data and vAPP PSF models to the vAPP data separately.

Telescope aperture	Clear DX	Clear SX	dgvAPP DX	dgvAPP SX
I_{Norm}	0.81	1.0	0.44	0.45
Diffracted leakage	-	-	0.04	0.01
Estimated zero-order leakage	-	-	1.6×10^{-3}	1×10^{-4}

and consists of annular rings with 0 or π phase with an added grating with 40 cycles over the pupil. Both the phase pattern with and without grating are shown in Fig. 4.13. This design has an annular dark zone from 2.7 to 15 λ/D with a raw contrast of 10^{-4} close to the inner working angle and 10^{-5} further out. The inherent Strehl ratio for the coronagraphic PSF core is 46%.

The presented design for the double-grating vAPP has been installed inside LMIRCam on the Large Binocular Telescope early September 2018. The first on-sky results have been obtained on December 24th 2018, with the aim of demonstrating the double-grating concept on-sky. The star HIP 75097 was observed in L-band (std-L, 3.41 μm - 3.99μm) with and without the double-grating vAPP coronagraph with 1.3 arcseconds seeing at an airmass of 1.36. The total integration time of with the vAPP is 5.5 seconds and the total integration time with a clear aperture is 25 seconds. The reduced image, containing PSFs for both telescope

apertures are shown in Fig. 4.13. For comparison, we simulate the PSF of a single aperture over 12% bandwidth using HCIPy (Por et al., 2018). The simulations visually match the on-sky data apart from the AO residuals (mostly wind-driven halo) and low-order aberrations. Because only one PSF (including diffracted leakage term PSFs) is visible for each telescope aperture, it can be concluded that the PSFs of both polarization states overlap and the double-grating principle works, and the manufacturing and installation was successful.
We investigate the on-sky transmission and leakage from the L-band data by fitting PSF models to both the coronagraphic and non-coronagraphic data using least-squares fitting. The bandwidth of the models is 3.55 μm to 3.95 μm, where the throughput is modelled as a tophat profile. The fit parameters are the position on the detector, the plate scale, the PSF intensity and the residual background. For the vAPPs we also fit phase pattern rotation, leakage and the circular polarization fraction (derived from the flux ratio of the leakage PSFs) for each telescope aperture seperately. The fitted intensities and leakage values can be found in Table 4.1. Here I_{norm} is the transmission normalized on one of the two PSFs ("SX") for a clear aperture, the diffracted leakage is the intensity of the leakage terms obtained from the off-axis PSFs, and the on-axis leakage is estimated from these off-axis leakage terms. Note that the inherent Strehl is not included in the transmission. The transmission below 50% can be explained by a strong absorption feature of the glue (Norland NOA 61) and liquid crystals between 3.55 μm to 3.7 μm. Furthermore, both apertures ("SX" and "DX") do not have the same transmission, where DX has a transmission that is 20% lower than the SX aperture and the PSF is severely aberrated. For the vAPP no such effect is seen, where I_{norm} is the same for SX and DX. This could be explained with a slight pupil misalignment for the clear aperture. Because the vAPP apertures are undersized, the misalignment has a smaller impact on the vAPP.

4.6 Multi-grating coronagraphs and system-level perspectives

We can extend the multi-grating principle from the double-grating VVC in Sect. 4.4 to a triple-grating or generally N-grating implementation. The additional layer(s) of leakage-term diffraction enable better intrinsic contrast performance and/or achieving that contrast over larger spectral bandwidths. Moreover, these techniques can avoid PSF splitting (Komanduri et al., 2011). First we will show the triple-grating concept and expand this to N gratings.

4.6.1 Multiple grating combinations

The schematic of the triple grating element is shown in Fig. 4.14. The triple-grating element consists of three gratings with different periods in a 1:3:2 ratio. The primary and the conjugate beams of the first grating are diffracted in the opposite direction by the second grating and combined by the third grating. Only

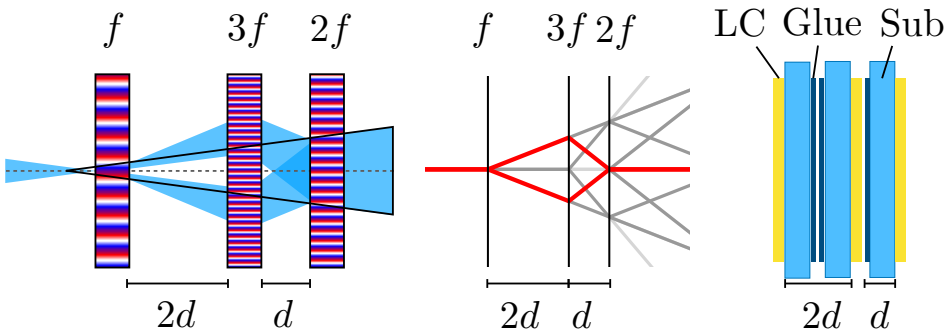


Figure 4.14: **Left:** Schematic of the triple grating concept. By combining three polarization gratings with different grating frequencies, f, in a 1:3:2 ratio, it is possible to overlap the beams with opposite circular polarization. This additional path length results in a shifted virtual focus. **Center:** Schematic of the paths of all diffracted terms for the triple grating. The gray scale indicates the intensity of the brightest leakage term. The red lines indicate the path of the beams that are diffracted by all gratings. **Right:** Implementation of the triple-grating concept with single thickness substrates.

one leakage wave and the primary and the conjugate beams can end up on-axis, i.e. the triple leakage and the triply diffracted beams. No other combination of the grating phases adds up to zero. To recombine both circular polarization states at the optical axis, it is required that the distance ratio between the gratings is exactly 1:2. This can be achieved with high accuracy if the liquid-crystal films and their substrates are positioned as shown in the right in Fig. 4.14. Any deviation from the 1:2 ratio will separate the PSFs according to Eq. 4.9. Note that for the pure grating patterns, only their orientation is relevant in the mutual alignment. With an average polarization leakage of 1% for a single grating, the combined leakage would be 10^{-6} at the center, and 10^{-8} at the first Airy ring.. Such a triple-grating (or any other N-grating) configuration in the focal plane does lead to a minor shift in the (virtual) focus.

For multi-element combinations, i.e. $N > 2$, there are two constraints. First, the total sum of the grating frequencies should add up to zero, taking into account the switch of the sign of the polarization state with each element, i.e. $\sum_i^N (-1)^i f_i = 0$. Only then the beam that is diffracted by all elements is parallel to the optical axis. Second, all the grating frequencies should be unique. This ensures that no other combination of frequencies is zero. Note that these frequencies can be negative, such that the grating direction is inverted. These constraints produce solutions that are scale invariant, i.e. they only constrain the ratios of the grating frequencies and distances. Hence, we are free to normalize the grating frequencies on the minimum frequency. There are multiple solutions for a single N-grating combination for $N > 3$. To ensure manufacturability the solutions should ensure the lowest grating frequency.

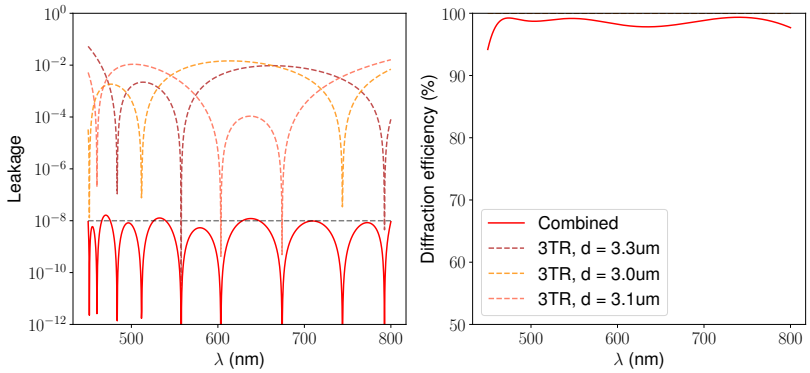


Figure 4.15: **Left:** Simulated leakage as function of wavelength for a triple-grating VVC in the visible wavelength range (450-800 nm). Each element is a 3TR with a slightly different recipe. **Right:** Simulated diffraction efficiencies for the same triple-grating VVC. The average diffraction efficiency is 98%.

4.6.2 Applications of multiple-grating focal-plane coronagraphs

For space applications, the double-grating and the triple-grating VVC implementations offer exciting opportunities to deliver extremely high intrinsic contrasts over large spectral bandwidths. We already showed that in simulation a multi-layered double-grating elements is able to suppress leakage to 10^{-10} for bandwidths up to 20%, see Fig. 4.2(d). We expand these simulations to a triple-grating element, optimizing the liquid-crystal recipes for three PGs simultaneously. These simulations show that a triple-grating VVC implementation has the potential to reach 10^{-10} leakage suppression on the first Airy ring. The optimized triple-grating element suppresses the leakage-term by eight orders of magnitude over the entire visible wavelength range, without any external polarization filtering, see Fig. 4.15. Moreover, a tgVVC has in theory no split-up of the off-axis PSF, as a dgVVC may have to a small extent. In future work, we will perform detailed simulations of the performance of a tgVVC6 (including full internal ray-tracing), and validate these simulations with a prototype element at a high-contrast test-bed.
In future space missions dedicated to high-contrast imaging, like the HabEx concept (HabEx Study Team, 2019) that features a 4-m off-axis telescope that is ideal for a charge-6 Vortex coronagraph, a tgVVC can have a major impact on the system-level design and trade-offs:
A single tgVVC has the potential to suppress the leakage to 10^{-10} on the first Airy ring in any 20% band over the entire visible range, meaning that a dichroic split-up in e.g. a blue arm and a red arm would be no longer necessary. Two identical arms can then implement filter wheels that contain all relevant filters, to still perform simultaneous observations in complementary spectral bands, or in the same

spectral band with complementary dark-hole shapes. This architecture provides full redundancy in case of a single-point failure in either arm. Moreover, it can be considered to perform spectroscopic observations over a larger spectral band than 20%, and rely on a single-mode fiber feed and spectral differential techniques to compensate for the intrinsically worse contrast performance of the broadband system (Por & Haffert, 2019; Coker et al., 2019).

In theory, it is not required to perform any polarization filtering when using the tgVVC6, and all light could go through. The charge-6 VVC has been selected for its robustness to low-order aberrations like tip/tilt and astigmatism, and particularly their polarization-dependent versions that invariably emerge for converging beams on fold mirrors and telescopic systems, respectively (Davis et al., 2018). In practice, however, it may still be necessary to install polarization optics. This would then mostly serve to mitigate the effects of polarization aberrations (Breckinridge et al., 2015) and facilitate polarization-dependent dark-hole control. Filtering linear polarization can then be combined with converting the linear polarization state to circular polarization, ensuring that only of the of two vector vortex twist modes, as induced by the geometric phase, is controlled by the AO system. Note that no further polarization filtering optics are then required after the tgVVC6 mask, which alleviates the overall instrument complexity. For now, we assume that a single layer of linear polarization filtering plus conversion to circular polarization is part of a realistic systems implementation.

We can combine the aforementioned strategies. With a two-arm tgVVC6 implementation, linear polarization filtering can be implemented with a polarizing beam-splitter that feeds the two identical arms, instead of a dichroic splitter (HabEx Study Team, 2019). This configuration now allows the implementation of dual-beam polarimetry with identical spectral bands and dark holes in both arms. This system can be upgraded to a complete linear beam-exchange polarimeter by including a rotating half-wave plate in front of the polarizing beam-splitter, polarization aberrations permitting (Snik & Keller, 2013; Snik et al., 2014).

4.7 Conclusions

We conclude that multi-grating geometric phase hologram elements, consisting of two or more polarization gratings, can be used to generically suppress polarization leakage of diffractive phase plate coronagraphs by orders of magnitudes compared to a single element. We demonstrate that by adding a coronagraphic phase pattern to the polarization grating of the first element, it is possible to make low-leakage geometric phase coronagraphs, i.e. a Vector Vortex Coronagraph and a vector-Apodizing Phase Plate coronagraph. We show in the lab that the double-grating Vector Vortex Coronagraph has much reduced on-axis leakage and has similar performance as other VVC coronagraphs in literature. We have demonstrated the double-grating concept on-sky with a newly installed double-grating vector-apodizing phase plate (vAPP) coronagraph, operating in LMIRCam at the Large Binocular Telescope. For ground-based telescopes, double-grating coron-

agraphs like the double-grating VVC, preferably of charge 2 (dgVVC2), and the dgvAPP360 are ideally suited to work in tandem with integral-field spectrometers, as they offer a single PSF that easily fits within the field-of-view, and, moreover, delivers stable, high-contrast performance over spectral bandwidths as large as a full octave. We aim to use our double-grating vAPP over the full bandwidth of the ALES integral field spectrograph in LMIRCam to detect and characterize exoplanets in K, L & M band. Furthermore, we will install a 1-2.5 μm double-grating VVC in the SCExAO instrument at the Subaru telescope to feed the CHARIS instrument integral field spectrograph. Finally, we will further develop and analyze double-grating and triple-grating VVCs that are optimized for $\sim 10^{-10}$ raw contrast of future space instrumentation, and demonstrate contrast performance over large spectral bandwidths.

Funding

ERC Starting Grant 678194 (FALCONER);

Acknowledgments

Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). We thank our all our collaborators from the LBT for helping us acquire the first on-sky results with the double-grating vAPP, especially Phil Hinz, Steve Ertel, Jordan Stone, and Eckhart Arthur Spalding. This research made use of HCIPy (Por et al., 2018), an open-source object-oriented framework written in Python for performing end-to-end simulations of high-contrast imaging instruments. We thank Christoph Keller for the many fruitful discussions, which helped improve the results presented in this work. We thank Matthew Kenworthy for doubting that this principle would ever work, but being a good sport when indeed it did.

Bibliography

Absil, O., Mawet, D., Karlsson, M., et al. 2016, Proc. Spie, 9908, 99080Q

Berry, M. V. 1984, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 392, 45

—. 1988, Scientific American, 259, 46. `http://www.scientificamerican.com/article/the-geometric-phase/`

Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, arXiv preprint arXiv:1902.04080

Breckinridge, J. B., Lam, W. S. T., & Chipman, R. A. 2015, PASP, 127, 445, doi: `10.1086/681280`

Codona, J. L., Kenworthy, M. A., Hinz, P. M., Angel, J. R. P., & Woolf, N. J. 2006, 6269, 62691N, doi: `10.1117/12.672727`

Coker, C. T., Shaklan, S. B., Riggs, A. J. E., & Ruane, G. 2019, arXiv e-prints, arXiv:1907.03921. https://arxiv.org/abs/1907.03921

Davis, J., Kupinski, M. K., Chipman, R. A., & Breckinridge, J. B. 2018, Proc. SPIE, 10698, 106983H, doi: 10.1117/12.2313670

Delacroix, C., Absil, O., Forsberg, P., et al. 2013, A&A, 553, A98

Doelman, D. S., Escuti, M. J., & Snik, F. 2019, Opt. Mater. Express, 9, 1246, doi: 10.1364/OME.9.001246

Doelman, D. S., Snik, F., Warriner, N. Z., & Escuti, M. J. 2017, Proc. SPIE, 104000, 10400U

Escuti, M. J., Kim, J., & Kudenov, M. W. 2016a, Opt. Photon. News, 27, 22, doi: 10.1364/OPN.27.2.000022

—. 2016b, Optics & Photonics News, 27, 22

Foo, G., Palacios, D. M., Swartzlander Jr, G. A., & Swartzlander, G. A. 2005, Optics letters, 30, 3308

Groff, T., Chilcote, J., Brandt, T., et al. 2017, Proc. SPIE, 10400, 1040016

HabEx Study Team. 2019, HabEx, Habitable Exoplanet Observatory, Tech. rep., Jet Propulsion Laboratory. https://www.jpl.nasa.gov/habex/pdf/HabEx-Final-Report-Public-Release.pdf

Hasman, E., Bomzon, Z., Niv, A., Biener, G., & Kleiner, V. 2002, Optics Communications, 209, 45, doi: 10.1016/S0030-4018(02)01598-5

Hinz, P. M., Defrère, D., Skemer, A., et al. 2016, Proc. SPIE, 9907, 990704, doi: 10.1117/12.2233795

Jolivet, A., de Xivry, G. O., Huby, E., et al. 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 025001

Jovanovic, N., Absil, O., Baudoz, P., et al. 2018, Review of high-contrast imaging systems for current and future ground-based and space-based telescopes: Part II. Common path wavefront sensing/control and coherent differential imaging, doi: 10.1117/12.2314260

Kanburapa, P., & Swartzlander, G. A. 2012, in Frontiers in Optics, Optical Society of America, FTh4E–2

Kenworthy, M. A., Codona, J. L., Hinz, P. M., et al. 2007, \Apj, 660, 762, doi: 10.1086/513596

Kim, J., Li, Y., Miskiewicz, M. N., et al. 2015, Optica, 2, 958, doi: 10.1364/OPTICA.2.000958

Komanduri, R. K., Lawler, K. F., & Escuti, M. J. 2011, Proc. SPIE, 8052, doi: 10.1117/12.887140

Komanduri, R. K., Lawler, K. F., & Escuti, M. J. 2013, Optics Express, 21, 404, doi: 10.1364/OE.21.000404

Leach, J., & Padgett, M. 2003, New Journal of Physics, 5, 154

Lozi, J., Guyon, O., Jovanovic, N., et al. 2018, 10703, 1070359, doi: 10.1117/12.2314282

Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proceedings of the National Academy of Science, 111, 12661, doi: 10.1073/pnas.1304215111

Males, J. R., Close, L. M., Miller, K., et al. ????

Mariyenko, I. G., Strohaber, J., & Uiterwaal, C. J. G. J. 2005, Opt. Express, 13, 7599, doi: 10.1364/OPEX.13.007599

Marois, C., Lafreniere, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, ApJ, 641, 556

Marois, C., Macintosh, B., Barman, T., et al. 2008, Science, 322, 1348

Mawet, D., Pueyo, L., Moody, D., Krist, J., & Serabyn, E. 2010, Proc. SPIE, 7739, 773914, doi: 10.1117/12.858240

Mawet, D., Riaud, P., Absil, O., & Surdej, J. 2005, ApJ, 633, 1191, doi: 10.1086/462409

Mawet, D., Serabyn, E., Liewer, K., et al. 2009a, ApJ, 709, 53
—. 2009b, Optics Express, 17, 1902
Mawet, D., Murakami, N., Delacroix, C., et al. 2011, Proc. SPIE, 8151, 815108
Miller, K., Males, J. R., Guyon, O., et al. 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 1 , doi: 10.1117/1.JATIS.5.4.049004
Miskiewicz, M. N., & Escuti, M. J. 2014, Optics Express, 22, 12691, doi: 10.1364/OE.22.012691
Oh, C., & Escuti, M. J. 2008, Optics Letters, 33, 2287, doi: 10.1364/OL.34.003637
Otten, G. P. L., Snik, F., Kenworthy, M. A., et al. 2014a, Proc. SPIE, 9151, 91511R, doi: 10.1117/12.2056096
Otten, G. P. P. L., Snik, F., Kenworthy, M. A., Miskiewicz, M. N., & Escuti, M. J. 2014b, Opt. Express, 22, 30287, doi: 10.1364/OE.22.030287
Otten, G. P. P. L., Snik, F., Kenworthy, M. A., et al. 2017, ApJ, 834, 175, doi: 10.3847/1538-4357/834/2/175
Packham, C., Escuti, M., Ginn, J., et al. 2010, PASP, 122, 1471, doi: 10.1086/657904
Pancharatnam, S. 1955, Proceedings of the Indian Academy of Sciences - Section A, XLI
Pancharatnam, S. 1956in , Springer, 247–262
Por, E. H. 2017, Proc. SPIE, 10400, 104000V
Por, E. H., & Haffert, S. Y. 2019, A&A, in press. https://arxiv.org/abs/1803.10691
Por, E. H., Haffert, S. Y., Radhakrishnan, V. M., et al. 2018, Proc. SPIE, 10703, doi: 10.1117/12.2314407
Racine, R., Walker, G. A., Nadeau, D., Doyon, R., & Marois, C. 1999, Publications of the Astronomical Society of the Pacific, 111, 587
Roberts, D., Xianyu, H., Nersisyan, S., Tabiryan, N. V., & Serabyn, E. 2019, in 2019 IEEE Aerospace Conference, IEEE, 1–15
Ruane, G., Mawet, D., Jewell, J., & Shaklan, S. 2017, Proc. SPIE, 10400, 104000J
Ruane, G., Mawet, D., Riggs, A., & Serabyn, E. 2019, arXiv preprint arXiv:1908.09786
Ruane, G., Riggs, A., Mazoyer, J., et al. 2018, in Proc. Spie, Vol. 10698, doi: 10.1117/12.2312948
Ruane, G. J., Kanburapa, P., Han, J., & Swartzlander, G. A. 2014, Appl. Opt., 53, 4503, doi: 10.1364/AO.53.004503
Serabyn, E., Mawet, D., & Burruss, R. 2010, Nature, 464, 1018
Serabyn, E., Prada, C. M., Chen, P., & Mawet, D. 2019, JOSA B, 36, D13
Serabyn, E., Huby, E., Matthews, K., et al. 2017, The Astronomical Journal, 153, 43
Skemer, A. J., Hinz, P., Stone, J., et al. 2018, Proc. SPIE, 10702, 107020C
Skemer, A. J., Hinz, P., Montoya, M., et al. 2015, Proc. SPIE, 9605, 96051D
Skrutskie, M. F., Jones, T., Hinz, P., et al. 2010, Proc. SPIE, 7735, 77353H
Snellen, I., de Kok, R., Birkby, J. L., et al. 2015, A&A, 576, A59, doi: 10.1051/0004-6361/201425018
Snik, F., & Keller, C. U. 2013, in Planets, Stars and Stellar Systems, ed. T. D. Oswalt & H. E. Bond (Springer), 175, doi: 10.1007/978-94-007-5618-2{_}4
Snik, F., Otten, G., Kenworthy, M., Mawet, D., & Escuti, M. 2014, Proc. SPIE, 9147, 91477U, doi: 10.1117/12.2055452
Snik, F., Otten, G., Kenworthy, M., et al. 2012, Proc. SPIE, 8450, 84500M, doi: 10.1117/12.926222
Snik, F., Absil, O., Baudoz, P., et al. 2018, in , doi: 10.1117/12.2313957
Sparks, W. B., & Ford, H. C. 2002, ApJ, 578, 543, doi: 10.1086/342401
Vargas Catalán, E., Huby, E., Forsberg, P., et al. 2016, Astronomy and Astrophysics, 595, A127, doi: 10.1051/0004-6361/201628739

5 Multiplexed Holographic Aperture Masking with liquid-crystal geometric phase masks

Adapted from
D. S. Doelman, P. B. Tuthill, B. Norris, M. J. Wilby, C. U. Keller, M. J. Escuti, F. Snik
Proceedings of the SPIE, Volume 10701, 107010T (2018)

Sparse Aperture Masking (SAM) allows for high-contrast imaging at small inner working angles, however the performance is limited by the small throughput and the number of baselines. We present the concept and first lab results of Holographic Aperture Masking (HAM) with extreme liquid-crystal geometric phase patterns. We multiplex subapertures using holographic techniques to combine the same subaperture in multiple non-redundant PSFs in combination with a non-interferometric reference spot. This way arbitrary subaperture combinations and PSF configurations can be realized, giving HAM more uv-coverage, better throughput and improved calibration as compared to SAM, at the cost of detector space.

5.1 Introduction

Sparse aperture masking (SAM) is a technique that turns a single dish telescope into an interferometer by masking out a large fraction of the pupil. The opaque mask consists of a sparse combination of holes such that the point-spread function (PSF) is a combination of interferometric fringes from each baseline. Because SAM is an interferometric method it is possible to measure down to half the diffraction limit of a single dish telescope. More importantly, with SAM it is possible to measure closure phases, an observable that is independent of the incoming wavefront aberration. Both advantages make SAM a good option for high contrast imaging and with the improved calibration that came with adaptive optics systems, sparse aperture masking has been extremely successful in imaging asymmetric structures around stars at small separations unreachable by other techniques like coronagraphy. SAM has been used to measure the shape and grain sizes of dust shells around stars (Norris et al., 2012), discovering substellar companions around young stars (Kraus & Ireland, 2012; Huelamo et al., 2011) and measure stellar multiplicity in star-forming regions (Ireland et al., 2008; Martinache et al., 2009; Cheetham et al., 2015).
Non-redundant masking is a subtechnique of SAM and requires the holes to be places in non-redundant patterns. Only a limited amount of holes can be combined in a non-redundant way and non-redundant masks therefore have a low throughput. Techniques like segment tilting and pupil remapping are ways to improve the throughput by making different non-redundant combinations of the aperture. Both techniques are complex to implement for a given telescope compared to the simplicity of sparse aperture masking where only one mask with holes is needed. Holographic aperture masking (HAM) uses a single phase plate to combine non-redundant subapertures at different focal plane locations, maintaining the simplicity of SAM. The possibilities with HAM go beyond the segment tilting as holographic techniques can be used to make multiple copies of each subaperture, enabled by liquid-crystal technology. A comparison between SAM and HAM is given in Fig. 5.1. HAM can incorporate SAM while adding more baselines by interfering additional subapertures with PSF copies at a separate location in the focal plane. While more detector space is required, HAM has more baselines and closure phases, uses more subapertures, allows for broadband operation and can be used to generate amplitude reference spots.
In this paper, we present the theory for blazed gratings and liquid-crystal technology in section 5.2, the design of holographic aperture masks in section 5.3, lab results in section 5.4 and conclusions are presented in section 5.5.

5.2 Theory

Holographic aperture masking (HAM) uses the freedom that any subaperture of the pupil can be imaged on any place in the focal plane. These subaperture spots are interfered with other subapertures by placing copies of the point-spread

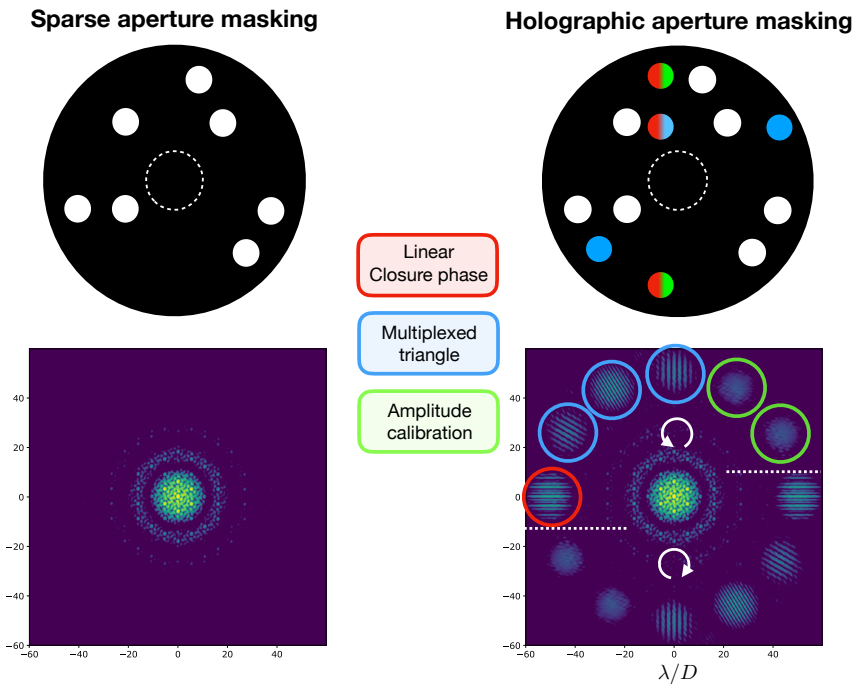


Figure 5.1: Comparison of sparse aperture masking (left) and holographic aperture masking (right). Holographic aperture masking uses a phase plate to combine otherwise redundant subapertures at separate locations in the pupil plane. A non-redundant mask is shown on top left with the resulting PSF on the bottom left. Top right is a schematic of the HAM combinations, where each subaperture with the same color is combined at the PSF encircled in that color. Subapertures are multiplexed to make multiple PSF copies. Combinations can be one dimensional at the same location (red) or in triangles at a different location for each baseline (blue). In green are the two subapertures that have non-interferometric PSFs used for amplitude monitoring. Two copies of each PSF are created with opposite circular polarization state.

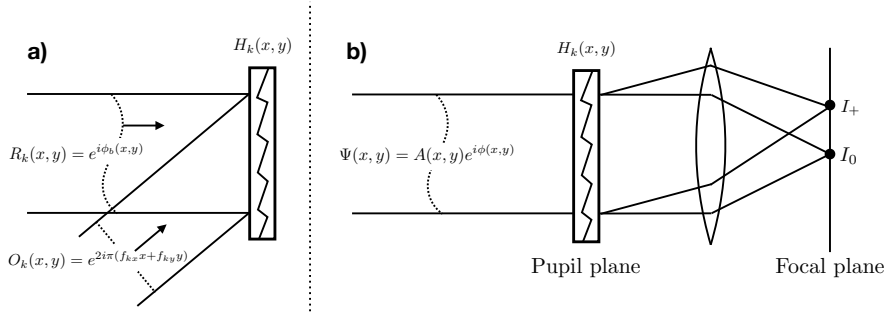


Figure 5.2: Diagram of holographic blazed gratings. **a)** Visualized generation of the holographic phase pattern. $H_k(x, y)$ is blazed by using only one interference term. **b)** A selected amount of the incoming wavefront is imaged at a separate location in the focal plane. Image adapted from Dong et al. (2012) (Dong et al., 2012) and Wilby et al. (2016)(Wilby et al., 2016).

function (PSF) of both apertures on the same location. This allows for arbitrary combinations of subapertures at any location of the focal plane. We generate these copies of a subaperture PSF using holographic blazed gratings.

5.2.1 Multiplexed holographic blazed gratings

Blazed gratings have an optimal diffraction efficiency to a single order. The generation of these holograms is depicted in Fig. 5.2, where we adapt the notation of Dong et al. (Dong et al., 2012).

We generate an interferogram between a reference wavefront $R_k(x, y)$ with a biased phase $\phi_b(x, y)$ and a object wavefront $O_k(x, y)$. The reference wavefront is given by

$$R_k(x, y) = e^{i\phi_b(x,y)} \tag{5.1}$$

and the object wavefront by

$$O_k(x, y) = e^{2i\pi(f_{kx}x+f_{ky}y)}, \tag{5.2}$$

where f_{kx} and f_{ky} are the spatial frequencies the hologram is placed in the focal plane, $f_{kx} = x'_k/f\lambda$. Here the focal plane coordinates are given by (x'_k, y'_k). HAM does not require a biased reference wavefront other than a piston term $(\phi_b(x, y) = c_k)$ that is used to phase scramble interferometric PSFs. The interferogram $H_k(x, y)$ between the two waves is then given by

$$H_k(x, y) = |O_k(x, y) + R_k(x, y)|^2 \tag{5.3}$$

$$H_k(x, y) = |O_k|^2 + |R_k|^2 + O_k^* R_k + O_k R_k^* \tag{5.4}$$

$$H_k(x, y) = 2 + O_k^* R_k + O_k R_k^*, \tag{5.5}$$

where * stands for the complex conjugate operator. The interferogram now generates two PSF copies, the ± 1 orders of the grating. Having only one of the two copies is preferred for HAM, as having two would increase the necessary detector space by a factor of two. We therefore blaze the grating by selecting one interference term,

$$H_k(x, y) = O_k R_k^*, \tag{5.6}$$

and only one PSF copy is created. Selecting one term is possible because we numerically generate the hologram.
The complex electric field in the pupil entering the holographic blazed grating can be described by

$$\Psi(x, y) = A(x, y) e^{i\phi(x,y)}, \tag{5.7}$$

where $A(x, y)$ is the pupil function of a subaperture of the telescope and $\phi(x, y)$ is the incoming wavefront. The consecutive focal plane intensity is given by $I = |\mathscr{F}[H_k \Psi]|^2$. Assuming a blazed grating, the intensity is given by

$$I_k(x'_k, y'_k) = \delta(x' - x'_k)\delta(y' - y'_k) * |\mathscr{F}[A(x, y)]|^2 * \left|\mathscr{F}\left[e^{i(\phi(x,y)+c_k)}\right]\right|^2, \tag{5.8}$$

using $*$ is the convolution operator. Eq. 5.8 is the subaperture PSF at (x'_k, y'_k) with a phase offset c_k. The blazed grating therefore only generates one PSF copy. Changing the PSF locations of subapertures is also possible with mirrors (Tuthill, 2012). Holography, however, allows for much more freedom than beam tilting only. An example of the freedom is phase scrambling and was already mentioned above. More importantly, holography enables multiplexing of holograms, creating more than one copy for each subaperture. HAM does benefit greatly from this property, we can choose to interfere subapertures arbitrarily with selected intensities anywhere in the focal plane. We multiplex blazed gratings by taking the sum over all complex interferograms scaled by a factor s_k. Both the reference and object wavefront have unity amplitude and therefore we create a phase-only hologram by taking the argument of the multiplexed hologram,

$$\phi_h(x, y) = \frac{1}{\pi} \arg\left[\sum_k^N s_k H_k(x, y)\right]. \tag{5.9}$$

An example of a multiplexed holographic grating is shown in Fig. 5.3.

5.2.2 Liquid-crystal technology

Phase patterns with the complexity required for HAM can be generated by applying geometric phase (or Pancharatnam-Berry) phase to circularly polarized light. When light that is left- or right-handed circularly polarized propagates through a half-wave retarder, the light acquires not only phase from the optical path difference between the fast-and slow-axis of the retarder, but also an extra phase shift

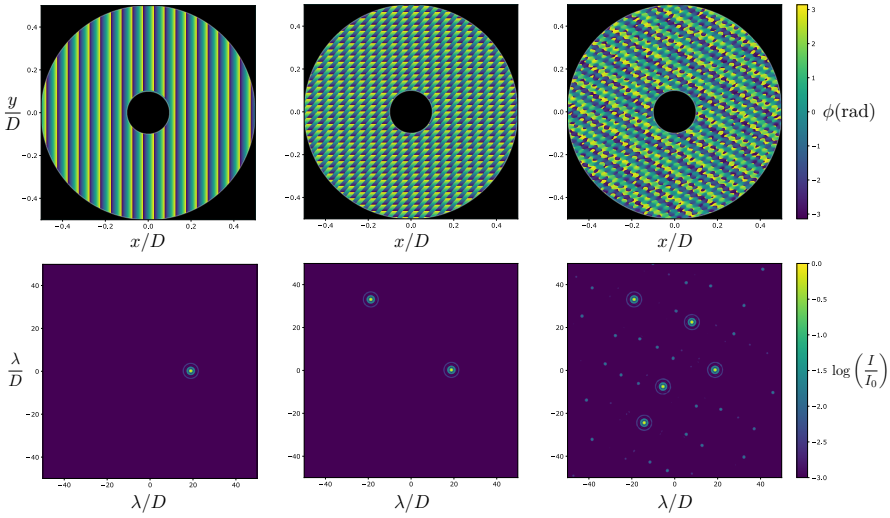


Figure 5.3: Holographic multiplexing of an aperture. The light is multiplexed in one, two and five holograms from left to right respectively.

is introduced. This extra phase shift ϕ only depends on the orientation of the fast-axis θ of the half-wave retarder and the circular polarization state (Pancharatnam, 1956; Berry, 1987; Escuti et al., 2016)

$$\phi = \pm 2\theta. \tag{5.10}$$

Note that the geometric phase is independent of wavelength but requires a retardance that is half-wave. When the retardance deviates from half-wave, only the amount of light that acquires the geometric phase (=efficiency), decreases and a leakage term emerges. The leakage term does not acquire any geometric phase.
Generating the complex phase pattern for HAM requires an optic with locally varying fast-axis orientation and a half-wave retardance. For broadband operation the retardance needs to be tuned to the desired wavelength range. Both requirements can be satisfied with liquid-crystal technology. A fast-axis orientation pattern can be written in a liquid-crystal polymerizable polymer (LCP) layer using a direct-write method (Miskiewicz & Escuti, 2014). This photo-alignment layer orients itself to the varying angle of linear polarization of the incoming uv-laser that scans the substrate. The retardance is tuned using birefringent self-aligning liquid-crystal layers. By changing the thickness and twist in each layer it is possible to achieve a retardance close to half-wave for very large bandwidths ($> 100\%$)(Komanduri et al., 2013). The layers are cured with uv-radiation, such that the device passively retains it broadband performance. Many complex patterns have been manufactured with this technology already (Escuti et al., 2016; Kim et al., 2015; Doelman et al., 2017).

5.2.3 Polarimetry with HAM (HAMPol)

Sparse aperture masking can be combined with polarimetry (SAMPol), making optimal use of the very small inner working angles with the ability to measure polarization signals (Tuthill et al., 2010). These polarization signals usually arise from scattering or reflection of light and SAMPol was used to spatially resolve circumstellar dust shells at less than 2 stellar radii of three stars on the asymptotic giant branch (Norris et al., 2012). SAMPol is complementary to other techniques because SAMPol probes the most inner stellar regions not resolvable by conventional imaging polarimetry. The VAMPIRES instrument installed at the Subaru telescope takes advantage of the small diffraction limit in the visible to eventually scales as small as 10mas (Norris et al., 2015).
With holographic aperture masking it is also possible to measure polarization, similar to how the vector-apodizing phase plate can be used for polarimetry (Snik et al., 2012). The thin liquid-crystal film applies geometric phase to circular polarization states. The applied phase has a different sign for both polarization states (Eq. 5.10). Combined with the ability to make single holographic copies, the detector will have two holographic PSFs at opposite locations, one for each circular polarization state. When the liquid-crystal device is placed between quarter-wave plates, HAM separates linear polarization. HAM is therefore a natural implementation polarimetric imaging with aperture masking.

5.3 Design

There is a lot more freedom when designing a holographic aperture mask compared with a sparse aperture mask. Designing a sparse aperture mask involves optimizing the number of apertures, aperture sizes and their location to maximize uv-coverage and throughput under the constraint that they are non-redundant (Carlotti & Groff, 2010). The advantage of HAM is that it allows to multiplex any combination of subapertures at any location in the focal plane. The throughput of HAM is then easily increased compared to SAM by imaging multiple non-redundant combinations at different locations. The non-redundancy constraint still applies to combinations at the same focal plane location but not to combinations at different locations. In addition, subapertures can be multiplexed and can therefore be used more than once with any relative intensity. The signal to noise in the focal plane is optimal when combinations of subapertures have the same intensity, but that is not constrained by the technology. In the end, HAM is limited to the number of non-redundant combination PSFs that fit the detector space, as every PSF has the size of λ/D_s, where D_s is the diameter of the subaperture.

5.3.1 Bandwidth

HAM manufactured with liquid-crystal technology applies geometric phase that is independent of wavelength. While this would allow for broadband operation, the element is diffractive and wavelength smearing washes out fringes quickly for

off-axis PSFs. Mirror tilting does not have this disadvantage because it applies optical path difference (OPD), such that the center of the off-axis PSFs does not change with wavelength (Tuthill, 2012). A holographic grating that places a PSF copy at $30\lambda/D$ smears the PSF over $0.3\lambda/D$ for a 1% bandwidth. Any gain in throughput from HAM is then nulled by limiting the bandwidth to 1%. There are two solutions to this problem. The first one is to use the design freedom to combine subapertures in only one direction such that the fringes are one dimensional and can be placed orthogonal to the smearing direction. The second solution is to use a Wynne lens that corrects the dispersion by introducing a chromatic difference of magnification with opposite sign. This is not entirely similar to segment tilting as off-axis sources are still dispersed and the dispersion is a function of the distance to the star. Using a Wynne lens has the advantage that it does increase the signal per pixel, however it consists of a set of custom lenses and the dispersed planet changes the fringe pattern.

5.3.2 One dimensional combinations

Combinations of subapertures in one dimension will create fringes in the same direction. We place these fringes orthogonal to the wavelength smearing direction such that we can increase the bandwidth arbitrarily without decorrelating the fringes. This is seen in Fig. 5.4. The fringes appear on both sides for opposite polarization states, no leakage was simulated. Using achromatic phase enabled by liquid-crystal technology, together with the optimal fringe placement, HAM is able to operate on bandwidths much larger than a few percent that is typical for sparse aperture masking. HAM design is now a trade-off between bandwidth and detector real estate.

Only combining apertures in one dimension at one point in the focal plane greatly reduces the number of baselines and closure triangles. The larger bandwidth and the ability to use more subapertures can in certain cases make up for the loss in the number of closure phases, especially when the detector is large and the observed object has a simple geometry. However, closure phases are the most used way of acquiring unambiguous measurements of an observed object. Another way to increase the number of baselines and closure phases is to add holograms to a SAM design, or some light can be multiplexed from a SAM design to locations on the detector for better amplitude calibrations, as shown in Fig. 5.1. The bandwidth is then limited by the SAM design.
Having a large bandwidth is not only useful for more photons on the detector. The phase offsets of each subaperture is wavelength dependent and therefore the fringes will change as function of wavelength. Fitting the broadband fringes could enable the extraction of closure phases, increasing the amount of information stored in a single hologram. This was not simulated and is outside the scope of this paper.

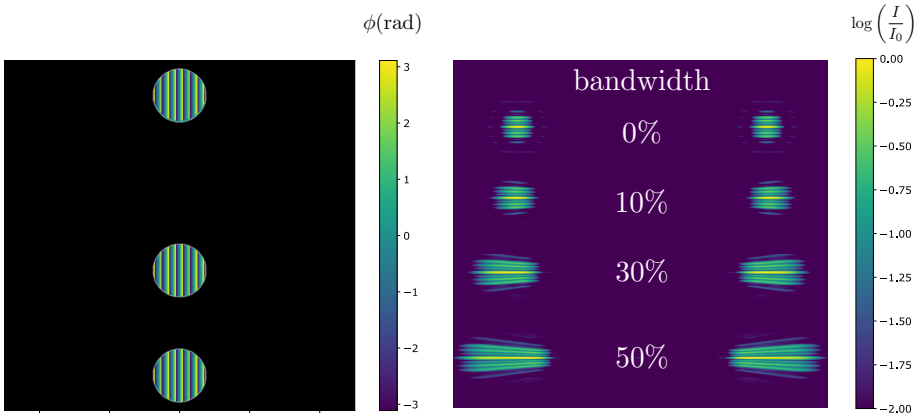


Figure 5.4: Bandwidth smearing of fringes (right) from a one dimensional combination of subapertures (left).

5.3.3 Multiplexed gratings

When the number of combined subapertures in one dimension is larger than two, it is possible to reconstruct a closure phase and there is no need for multiplexing the subapertures. Multiplexing is useful for combining subapertures that do not line up. A closure triangle can be obtained by combining the individual combinations per two subapertures at the location orthogonal to that specific baseline. This way, the bandwidth smearing is radial and a much larger bandwidth can be used compared to combining the three subapertures on one location of the focal plane. Multiplexing can also be used to link two separate subaperture combinations. However, to keep the intensities of all individual subapertures combinations the same, the percentage of light per holographic PSF needs to be scaled. As an example we assume that subapertures A, B and C are interfered at three different locations in the focal plane and A is linked for 40% to another combination D. Now A only has 60% of the light to make the combinations with B and C. Then the A-B combination with equal intensity uses only 30% from subaperture B. This can be compensated by making the B-C combination use 70% of subaperture B and C. From this example it is clear that optimizing the combinations with multiplexing can be very complex.

Moreover, multiplexing individual subapertures can be used for amplitude calibration by making a copy that is not interfered with any of the other subapertures. By measuring the encircled energy of this hologram, assuming the subaperture size is on the size of the Fried parameter r_0, the amplitude is directly obtained and no longer a fit parameter in fringe fitting. This could improve the precision of the calculated closure phases.

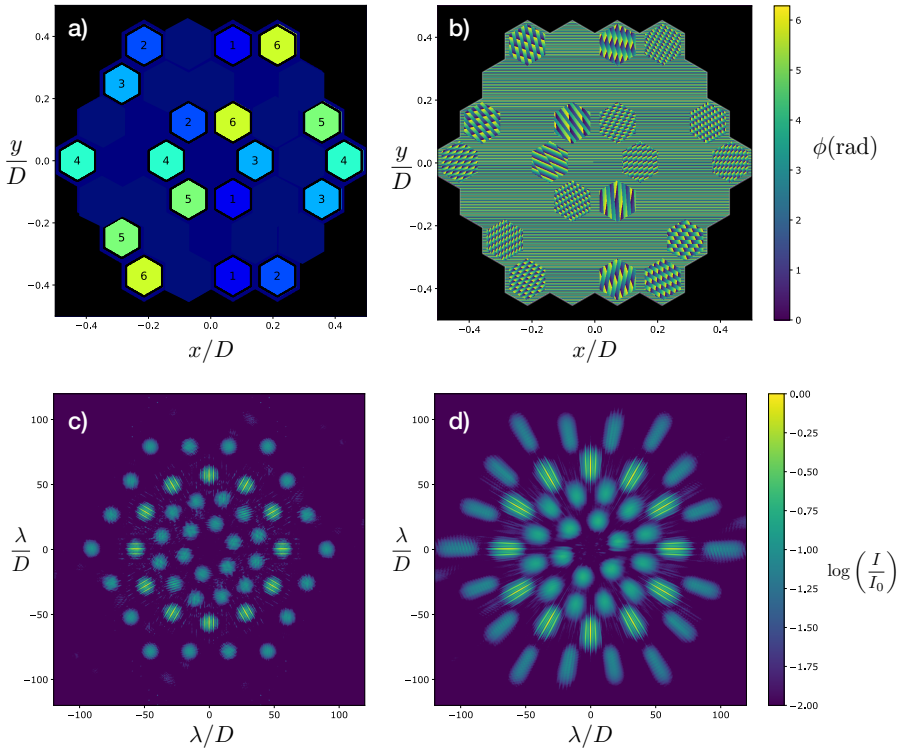


Figure 5.5: **a)** Diagram with the combinations of subapertures. Every subaperture with the same number is combined in one PSF on the focal plane and every subaperture is multiplexed to have a non-interferometric reference PSF. **b)** Phase pattern that corresponds to the design. The subapertures are cut out using a grating mask. **c)** Simulated PSF of the design, where all interferometric PSFs are placed in a ring. The reference PSFs within this ring are placed in smaller concentric rings, while the reference PSFs outside the ring are placed on a hexagonal grid. **d)** Simulated PSF for 30% broadband light. No indiviual PSFs overlap.

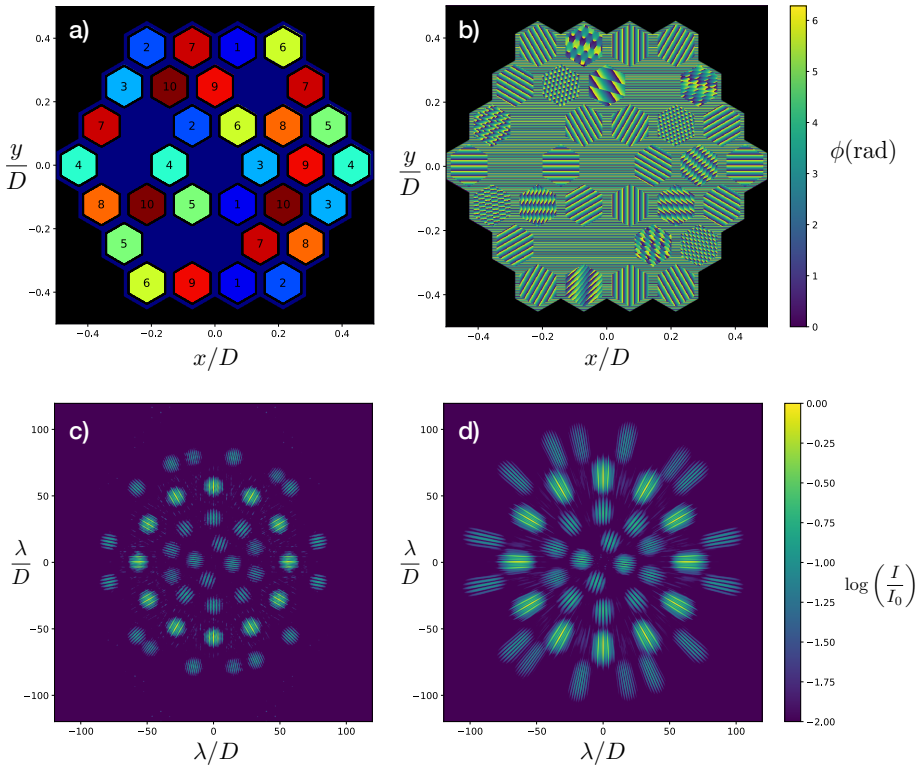


Figure 5.6: **a)** Diagram with the combinations of subapertures. Every subaperture with the same number is combined in one PSF on the focal plane. The pattern is optimized for the number of subapertures and closure phases. **b)** Phase pattern that corresponds to the design. The subapertures are cut out using a grating mask. **c)** Simulated PSF of the design, where all interferometric PSFs are placed in concentric rings. **d)** Simulated PSF for 30% broadband light. No individual PSFs overlap.

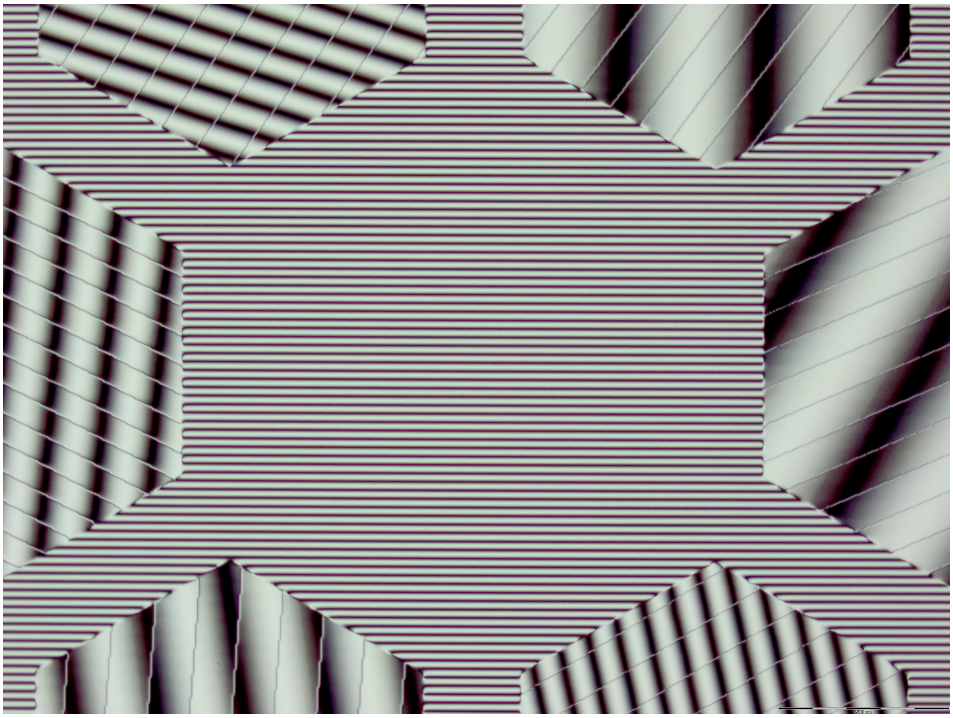


Figure 5.7: Microscopic image of the liquid-crystal pattern for HAM design one through crossed polarizers. The image is magnified 20 times and only contains the central part. The pattern matches the design to a very high degree.

5.4 Lab results

We designed two holographic aperture masks to be tested in the lab. We used a hexagonal aperture from Keck and the subapertures are similar to undersized single mirror segments. Using such a regular grid and relatively large subapertures greatly reduces the degrees of freedom. The designs are optimized by hand as a global optimization process is still being researched. By printing a grating mask, a high frequency grating outside of the used subapertures, there is no need for a real amplitude mask for the specific designs (Doelman et al., 2017).

5.4.1 Manufactured design

The first design combines all possible one dimensional combinations of three subapertures. The numbering and color in panel **a)** indicate the which subapertures are combined at one location of the focal plane. For this design 18/36 subapertures are used to get 18 baselines and 6 closure phases. We multiplexed every subaperture such that it has a copy that can be used for amplitude calibration.

The resulting phase pattern is shown in panel **b)**. The monochromatic PSF and the PSF for 30% bandwidth are shown in panel **c)** and **d)** respectively. The total amount of spots is 48. The six interference point-spread functions are combined at $60\lambda/D$ to minimize crosstalk and to keep the necessary field of view as small as possible. The less important individual PSFs have been added to form the most compact configuration, allowing for some crosstalk.
The second design has the same optimal basis as the first design with all combinations of three subapertures. We optimized the amount of used subapertures, baselines and closure phases by hand, starting from the interferometric PSFs of design one. The second design has an extra combination of four subapertures, number 7 in Fig. 5.6 **a)**, and extra combinations of three subapertures, number 8-10. In total 31/36 possible subapertures have been used for 33 baselines and 12 closure phases with 42 spots. The combinations, the phase and the PSFs are displayed in Fig. 5.6.

5.4.2 Manufacturing and lab testing

The designs were manufactured in December 2017 and both were printed on the same two inch substrate with a single layer of liquid-crystals. The single layer has an optimized thickness to tune the retardance to be half-wave at a wavelength of 532nm. The pixel size of the pattern is 5 micron and the maximum width of the hexagonal aperture is 11mm. The subapertures are undersized by a factor of **x** and the grating period of the grating mask is **35 micron**. No anti-reflection coating or wedges have been added. The pattern was inspected between crossed polarizers under a microscope with a magnification of 20. The resulting image is shown in Fig. 5.7. No significant differences between the design and the manufactured pattern were found. We tested the lab plates on a preliminary version of the Leiden HCI testbed. A laser diode (CPS532) from Thorlabs operating at 532nm is fed into a single mode fiber for a diffraction limited input. The output of the fiber is collimated and an iris stops the pupil down to a size of 12mm. The holographic aperture mask is conjugated to this pupil plane with two intermediate focii and one reflective pupil where no active element is installed. We image the beam on a SBIG detector with a field of view of more than 200x200 λ/D. We compare the result with a simulated PSF where the ratio between left and right circular polarization was matched to the measurement. The simulated PSF is shown in Fig. 5.8 and the measured PSF is shown in Fig. 5.9. The measured PSF has only very small deviations from the simulated PSF, showing that liquid-crystal technology is capable of accurately making the phase patterns required for HAM.

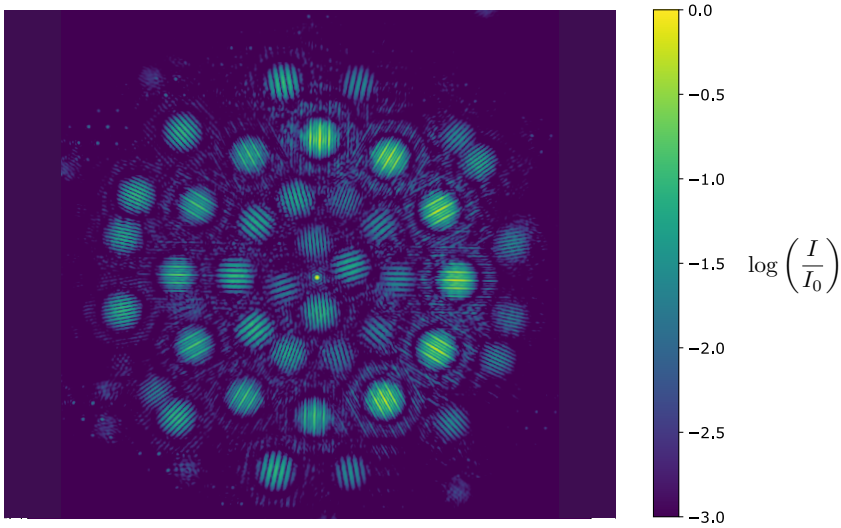


Figure 5.8: Simulated PSF for HAM design 2 for monochromatic light and a 1:3 ratio between left and right circular polarization. The simulated amount of leakage is 0.2%.

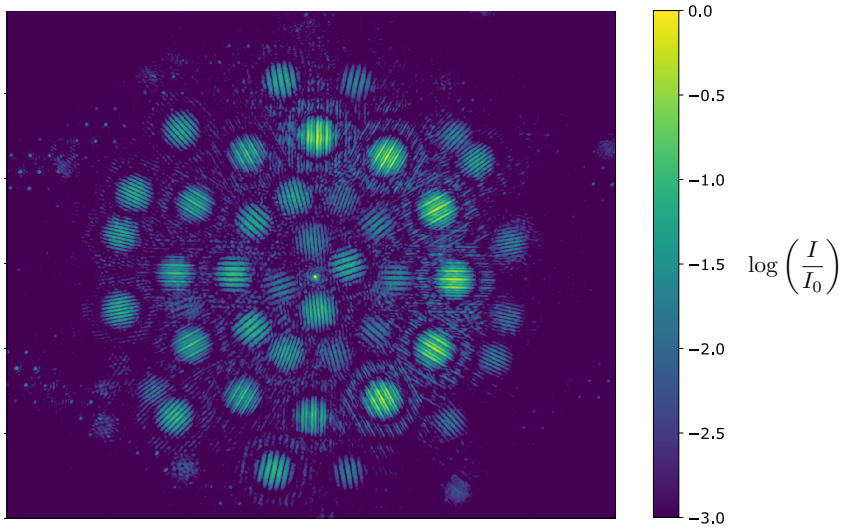


Figure 5.9: Measured PSF for HAM design 2 at a wavelength of 532nm.

5.5 Conclusion

- Liquid-crystal technology enables the use of holographic techniques for masking interferometry.
- A standalone holographic aperture mask (HAM) with one dimensional subaperture combinations can operate up to more than 30% bandwidth, however the number of closure phases is limited.
- We spectrally resolve fringes with HAM using these one dimensional subaperture combinations with large bandwidths.
- Multiplexing subapertures not used by sparse aperture masking (SAM) increases the number of baselines, closure phases and throughput at the cost of detector space.
- Leakage from deviations of half-wave retardance interfere with the SAM point-spread function but can be controlled using a double grating technique.
- Holographic aperture masking works in the lab and will be on sky soon.
- Liquid-crystal technology also can be used for broadband achromatic nullers.

5.6 Acknowledgements

The research of David S. Doelman and Frans Snik leading to these results has received funding from the European Research Council under ERC Starting Grant agreement 678194 (FALCONER)

Bibliography

Berry, M. 1987, Journal of Modern Optics, 34, 1401, doi: 10.1080/09500348714551321

Carlotti, A., & Groff, T. 2010, in Ground-based and Airborne Telescopes III, Vol. 7733, International Society for Optics and Photonics, 77333D

Cheetham, A. C., Kraus, A. L., Ireland, M. J., et al. 2015, Astrophysical Journal, 813, 1, doi: 10.1088/0004-637X/813/2/83

Doelman, D., Snik, F., Warriner, N., & Escuti, M. 2017, in Proceedings of SPIE - The International Society for Optical Engineering, Vol. 10400, doi: 10.1117/12.2273406

Dong, S., Haist, T., Osten, W., Ruppel, T., & Sawodny, O. 2012, Applied Optics, 51, 1318, doi: 10.1364/AO.51.001318

Escuti, M. J., Kim, J., & Kudenov, M. W. 2016, Opt. Photon. News, 27, 22, doi: 10.1364/OPN.27.2.000022

Huelamo, N., Lacour, S., Tuthill, P., et al. 2011, 7, 2009, doi: 10.1051/0004-6361/201016395

Ireland, M. J., Kraus, A., Martinache, F., Lloyd, J. P., & Tuthill, P. G. 2008, The Astrophysical Journal, 678, 463, doi: 10.1086/529578

Kim, J., Li, Y., Miskiewicz, M. N., et al. 2015, Optica, 2, 958, doi: 10.1364/OPTICA.2.000958

Komanduri, R. K., Lawler, K. F., & Escuti, M. J. 2013, Optics Express, 21, 404, doi: 10.1364/OE.21.000404

Kraus, A. L., & Ireland, M. J. 2012, Astrophysical Journal, 745, doi: 10.1088/0004-637X/745/1/5

Martinache, F., Guyon, O., Lozi, J., et al. 2009, Exoplanets and Disks: Their Formation and Diversity, 9-12 March 2009, 1158, 329, doi: 10.1117/12.894310

Miskiewicz, M. N., & Escuti, M. J. 2014, Optics Express, 22, 12691, doi: 10.1364/OE.22.012691

Norris, B., Schworer, G., Tuthill, P., et al. 2015, Monthly Notices of the Royal Astronomical Society, 447, 2894, doi: 10.1093/mnras/stu2529

Norris, B. R., Tuthill, P. G., Ireland, M. J., et al. 2012, Nature, 484, 220, doi: 10.1038/nature10935

Pancharatnam, S. 1956in , Springer, 247–262

Snik, F., Otten, G., Kenworthy, M., et al. 2012, Proc. SPIE, 8450, 84500M, doi: 10.1117/12.926222

Tuthill, P., Lacour, S., Amico, P., et al. 2010, in Ground-based and Airborne Instrumentation for Astronomy III, Vol. 7735, International Society for Optics and Photonics, 77351O, doi: 10.1117/12.856806

Tuthill, P. G. 2012, in Optical and Infrared Interferometry III, Vol. 8445, International Society for Optics and Photonics, 844502, doi: 10.1117/12.925794

Wilby, M. J., Keller, C. U., Snik, F., Korkiakoski, V., & Pietrow, A. G. M. 2016, 112, 1, doi: 10.1051/0004-6361/201628628

6 Design and Performance Analysis of a prototype Holographic Aperture Mask for the Keck OSIRIS Imager

Adapted from
D. S. Doelman , J. P. Wardenier, P. B. Tuthill, M. P. Fitzgerald, J. Lyke, S. Sallum, B. Norris, N. Z. Warriner, P. Wizinowich, C.U. Keller, M. J. Escuti, F. Snik
Accepted by Astronomy and Astrophysics

Context As an interferometric technique, sparse aperture masking (SAM) is capable of imaging beyond the diffraction limit of single telescopes. This makes SAM an important technique to study processes like planet formation at solar-system scales. However, it comes at the cost of a reduction in throughput, typically by 80-90%.
Aims We report on the design, construction and commissioning of a prototype aperture masking technology implemented at the Keck OSIRIS Imager: the holographic aperture mask. Holographic aperture masking (HAM) aims at increasing the throughput of sparse aperture masking by selectively combining all subapertures across a telescope pupil in multiple interferograms using a phase mask, in addition to adding low-resolution spectroscopic capabilities.
Methods Using liquid-crystal geometric phase patterns we manufacture a HAM mask that uses an 11-hole SAM design as central component and a holographic component that compromises 19 different subapertures. Thanks to a multi-layer liquid-crystal implementation, the mask has a diffraction efficiency higher than 96% from 1.1 to 2.5 micron. We create a pipeline that extracts monochromatic closure phases from the central component as well as multi-wavelength closure phases from the holographic component. We test the performance of the HAM mask in the lab and on-sky.
Results The holographic component yields 26 closure phases with a spectral resolutions between R$\sim$6.5 to R$\sim$15, depending on the interferogram positions. On April 19 2019, we observed the binary star HDS 1507 in the H_{bb} filter ($\lambda_0 = 1638$ nm and $\Delta\lambda = 330$ nm) and retrieve a constant separation of 120.9 $\pm$0.5 mas for the independent wavelength bins, which is excellent agreement with literature values. For both the lab measurements and the observations of unresolved reference stars we record non-zero closure phases; a potential source of systematic error that we traced to polarization leakage of the HAM optic. We propose a future upgrade that improves the performance, reducing this effect to an acceptable level.
Conclusions Holographic aperture masking is a simple upgrade of SAM with increased throughput and a new capability of simultaneous low-resolution spectroscopy which provides new differential observables (e.g. differential phase with wavelength).

6.1 Introduction

Many of the most critical aspects of stellar physics play out in a theatre at solar-system scales. These include star and planet formation, mass loss and debris disks to name only a few. High-fidelity imaging of circumstellar environments can provide key insights in these processes. High-contrast imaging instruments with adaptive optics (AO) provide high resolution imagery with great sensitivity, resolving many protoplanetary disks and sub-stellar companions (Chilcote et al., 2018; Beuzit, J.-L. et al., 2019). However, the performance of high-contrast imaging systems is still limited by residual phase and non-common path aberrations, reducing the sensitivity (Macintosh et al., 2019; Beuzit, J.-L. et al., 2019). Nonetheless, extreme AO facilities have been able to reach high contrasts (>14 magnitudes) down to ∼200 mas: a few times the diffraction limit in the near-infrared (Vigan et al., 2015; Nielsen et al., 2019). For nearby populous star forming regions such as Taurus, 200 mas corresponds to ~ 30 AU (larger than the orbit of Jupiter and Saturn), which leaves a blind spot for critical scales of disk evolution and planet formation. An additional technique called sparse aperture masking (SAM), often used in concert with AO, has been able to resolve finer structures beyond the diffraction limit, e.g. 20 mas at 1.65 μm (Tuthill et al., 1999).

SAM works by turning a telescope aperture into an interferometric array using an opaque mask with small holes (Haniff et al., 1987; Tuthill et al., 2000). For most applications, the holes are placed in a non-redundant fashion, which means that each baseline (the vector that connects two apertures) appears only once. Imaging with such a mask results in an interferogram that contains many fringes with unique spatial frequencies in the image plane. The first null of these fringes is at $0.5\lambda/B$ instead of $1.22\lambda/D$, where λ is the wavelength, B the longest baseline and D the telescope aperture diameter. A second profound advantage is the rejection of phase noise. Non-redundancy acts to remove noise in both visibility amplitudes and phase measurements, and in particular robust observables known as closure phases have been exploited with great success. Closure phases are formed by taking the sum of phases around baselines forming a closed triangle of subapertures in the pupil. Even before adaptive optics (AO) became well established, the robust observables delivered by SAM allowed for imaging the regions closest to stars (Tuthill et al., 1999).

Sparse aperture masking is even more powerful when used in concert with AO, providing extremely stable closure phases that resulted in a leap in sensitivity and contrast (Tuthill et al., 2006). The superior calibration particularly of closure phases makes SAM more sensitive than coronagraphs for probing the smallest separations ($< 1 - 2\lambda/D$), e.g. Gauchet et al. (2016); Cheetham et al. (2016); Samland, M. et al. (2017). Furthermore, SAM has established productivity in directly resolving stellar environments, recovering dust shells (Haubois, X. et al., 2019) and structures in protoplanetary disks (Kraus et al., 2008; Willson et al., 2019) at solar system scales.

Current SAMs, such as those in Keck/NIRC2 and VLT/SPHERE, block $80 - 90\%$ of the incident light, improving resolution at the expense of throughput and Fourier

coverage. Several different approaches to improve this throughput have been proposed. One approach is the kernel phase method introduced by Martinache (2010). Here, the concept of closure phase is generalised to the so-called kernel phase, which are also phase-noise resistant observable quantities. The advantage of this formalism over traditional closure phases is the fact that it can also be applied to redundant pupils (placing it within the lineage of development from Speckle Interferometry), provided that phase aberrations are small. In the high-Strehl regime, the Kernel phase method presents a huge increase in throughput compared to sparse aperture masking, although the achieved contrast is comparable to non-redundant masking (Pope et al., 2015; Sallum & Skemer, 2019). Another advantage of Kernel phase is that a full aperture yields more independent Kernel phases, allowing for a better characterization of the target. Lastly, it easy to implement in any extreme-AO instrument, as it requires only regular imaging and a different post-processing algorithm.

A different approach that aims to increase the throughput of aperture masks is segment tilting interferometry (Monnier et al., 2009). With this approach subsets of mirrors are imaged onto separate image plane locations, resulting in a few completely independent non-redundant fringe patterns. While an on-sky survey using this approach with the Keck 1 telescope proved successful, determining the sizes of many young-stellar objects, the experiment was stopped shortly after with the decommissioning of the Long Wavelength Spectrometer instrument. The simple concept is appealing, it has a huge increase in throughput, superior uv-coverage and the possibility of a photometric channel. However, the technical aspect is challenging with strict requirements on keeping the tip-tilt alignment and piston offset of individual mirror segments (Tuthill, 2012). Even so, there are many opportunities to revive this approach, as many future telescopes having a segmented primary mirror, e.g. JWST (Gardner et al., 2006), the TMT (Sanders, 2013), and the ELT (Gilmozzi & Spyromilio, 2007).

A third approach is pupil remapping interferometry, which uses photonic techniques to remap the pupil onto a non-redundant output pupil, such that the full pupil can be used (Perrin et al., 2006). Different photonic techniques have been implemented to remap the pupil, e.g. single-mode fibers (Huby et al., 2012, 2013) and integrated photonic chips (Jovanovic et al., 2012). The photonic chips are especially powerful, as a single integrated chip can contain multiple components that split single subaperture waveguides into multiple channels, control the phase delay of every channel and recombine all subaperture waveguides, interfering every subaperture with all others (Cvetojevic et al., 2018; Norris et al., 2019). Both photonic techniques perform spatial filtering, are compact and their monolithic design promises superior stability. Moreover, pupil remapping lends itself for spectroscopy, as the output can be arranged as 1D arrays of subapertures, which are easily dispersed. Overall, pupil remapping has some unique advantages compared to sparse aperture masking at the cost of increased complexity, as pupil remapping requires a separate instrument.

Sparse aperture masking is an elegant technique due to its simplicity, i.e. a single mask in the pupil plane adds the capability of imaging beyond the diffraction limit.

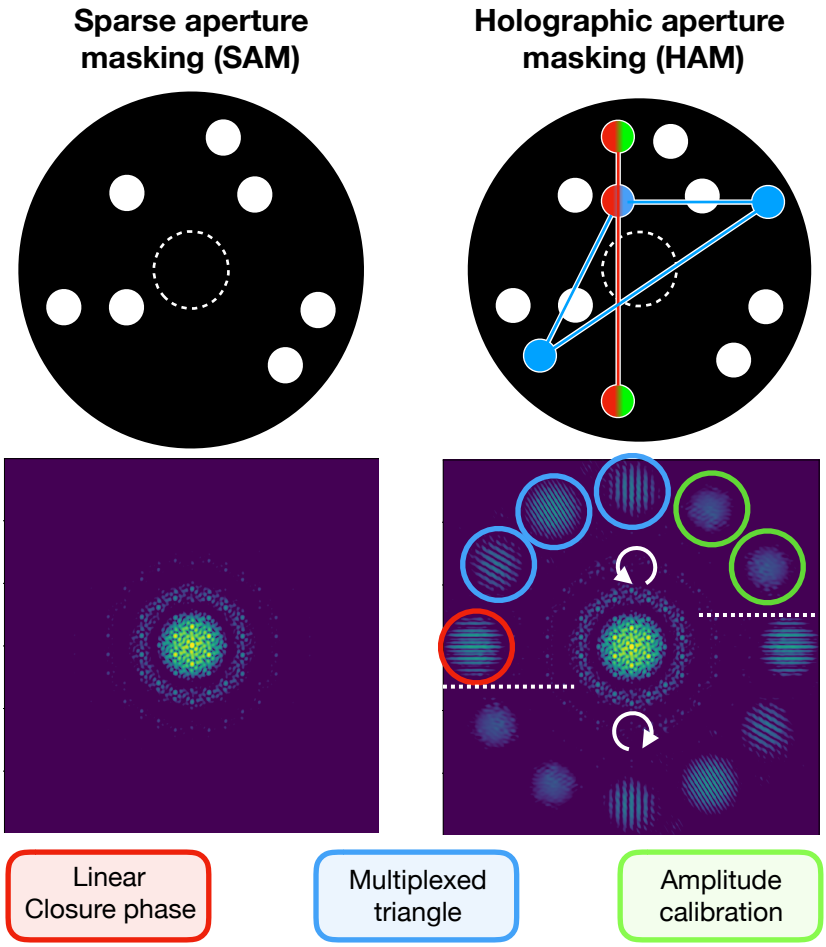


Figure 6.1: Cartoon of sparse aperture masking (left) and holographic aperture masking (right). The masks are shown on top with their resulting PSFs at the bottom. Holographic aperture masking (HAM) combines an aperture mask with a phase plate. HAM adds baselines by imaging otherwise redundant subapertures at separate locations in the focal plane. The SAM mask design is incorporated as the central component and the off-axis interferograms are part of the holographic component. Subapertures can be imaged onto multiple PSF locations to make different combinations. The different colors shown for HAM indicate different ways of combining subapertures and correspond to the same color coded off-axis interferograms. Multiple subapertures in one dimension can be combined at the same location (red) or in two dimensions by imaging single baselines at different locations (blue). It is also possible to make a non-interferometric PSF of a single subaperture for amplitude monitoring (green). Two copies of each PSF are created with opposite circular polarization state.

Yet, the approaches that have an increased throughput compared to SAM present demanding system requirements, requiring high Strehl (Kernel phase), a complete overhaul in the primary mirror alignment (segment tilting) or even a completely new instrument (pupil remapping).
In this paper we present the implementation of a novel approach, holographic aperture masking (HAM), that improves the throughput, number of baselines and closure phases, adds the capability of low-resolution spectroscopy, while also maintaining the simplicity of a single pupil-plane optic. The fundamental idea of HAM is equivalent to segment tilting, where interfering different subapertures on separate locations in the pupil allows for increasing the throughput without creating redundant baselines (Doelman et al., 2018). Instead of tilted mirrors, the phase mask introduces achromatic phase tilts to shift the location where subapertures are imaged onto the detector to form an interferogram. As developed in this paper, HAM is implemented as an addition to a non-redundant SAM mask, with the distinction between the two components of the hybrid experiment discussed as the SAM *central component*, and the off-axis HAM interferograms, the *holographic component*. Both components provide complementary information and can be used independently of each other. The difference between SAM and HAM is shown in Fig. 6.1. One unique advantage of the phase masks is that single subapertures can be imaged onto multiple locations. This is done by holographically combining these phase tilts through multiplexing. For example, the three blue subapertures in Fig. 6.1 are all imaged onto two holograms, such that all baselines are measured and a closure phase can be extracted. Moreover, the achromatic phase tilts make the HAM mask diffractive, i.e. the diffraction angle depends on wavelength. When the one dimensional subaperture combinations are deflected to a point in the image plane orthogonal to their baseline, the wavelength smearing does not act to blur out the fringes. Therefore, the holographic component can be designed to operate as a low-resolution spectrograph without an additional dispersing element.
We address the implementation and design of a HAM mask in Sect. 6.2, the influence of polarization leakage in Sect. 6.3, the HAM mask design for the Keck OSIRIS Imager in Sect. 6.4, the data reduction pipeline in Sect. 6.5, the manufacturing of two HAM masks for the Keck OSIRIS Imager in Sect. 6.6, the lab verification of one mask in Sect. 6.7, and the on-sky verification with a binary in Sect. 6.8.

6.2 HAM design

The HAM concept shows promise to improve upon SAM mask designs adding throughput, Fourier coverage and wavelength diversity. Here we explore the implementation of the HAM phase mask, and how the properties of the phase mask influence the design. A critical property of the phase mask is that it needs to be able to image subapertures in off-axis interferograms. The off-axis interferograms are rather large with size scaling with λ/D_{sub}, where λ is the wavelength and

D_{sub} is the diameter of the subaperture. Therefore, imaging multiple interferograms onto separate locations (so as to avoid overlap) on the detector requires large phase tilts. This makes it difficult to manufacture classical phase implementations of a HAM phase mask for transmissive pupil planes. A solution is offered by liquid-crystal geometric phase holograms (GPHs), as they have unbounded continuous phase over a large wavelength range. Next, we will explore the properties of these GPHs.

6.2.1 Geometric phase holograms

A geometric phase hologram (Escuti et al., 2016) is a phase plate that applies geometric phase (or Pancharatnam-Berry phase), rather than classical phase that arises from optical path differences. Geometric phase arises when light travels through an anisotropic medium, and only depends on the geometry of the light path through the medium. One type of GPHs are half-wave retarders with varying fast-axis orientation, and they apply geometric phase to circularly polarized light. The geometric phase, $\phi(x, y)$, of such a device only depends on the geometry of the fast-axis orientation $\theta(x, y)$, i.e.

$$\phi(x, y) = \pm 2\theta(x, y), \tag{6.1}$$

where the sign is determined by the handedness of the circular polarization state of the incoming light and x, y indicate the pupil-plane coordinates (Escuti et al., 2016). Unpolarized light is defined as having no preferred state of polarization and contains on average equal amounts of left and right circular polarization. The fraction of light that acquires geometric phase depends on the retardance, where half-wave retardance yields close to 100% diffraction efficiency. The fraction of light that does not acquire geometric phase is called the leakage term and is apart from a global piston term unaffected by the optic. The impact of this leakage term for HAM is explored in Section 6.3.

A GPH phase mask for HAM requires control of the fast-axis orientation and tuning of the retardance for optimal efficiency over the desired bandwidth. Both of these properties can be controlled to a very high degree with liquid-crystal technology. With a direct-write method, almost arbitrary phase patterns can be written in a photo-alignment layer (PAL) (Miskiewicz & Escuti, 2014). Birefringent liquid-crystal layers deposited on the PAL keep this orientation pattern due to spontaneous self-alignment. Changing the retardance is possible by stacking these layers with different thickness and twist into a monolithic film (Komanduri et al., 2013). By tuning these parameters, multi-twist retarders are capable of achieving high diffraction effiencies over large bandwidths. The layers are cured with UV radiation, and the liquid-crystal film therefore constitutes a static phase pattern. Moreover, the optic is completely flat and can easily be combined with an amplitude mask.

An example of a GPH is a polarization grating (PG). It has a continuously rotating fast-axis orientation, which can also be interpreted as a phase ramp in accordance

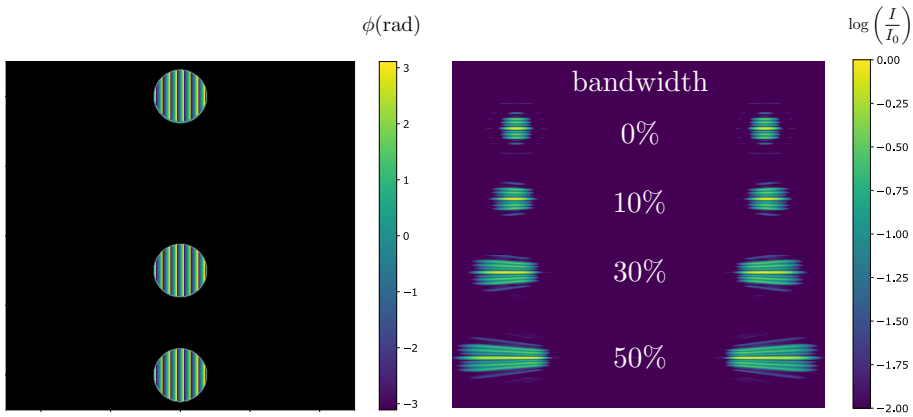


Figure 6.2: A linear arrangement of subapertures diffracted by a phase ramp or polarization grating. Such a design can be used over a large simultaneous bandwidth because the fringes are orthogonal to the bandwidth smearing.

to Eq. 6.1. A simple version of a HAM phase mask could therefore be manufactured as a collection of PGs for different subapertures. Yet, it is important to consider the diffractive nature of a GPH, which rules out a direct copy of mask designs for segment tilting. As geometric phase is independent of wavelength, the diffraction angle does change with wavelength. Thus, a continuous phase ramp will disperse the light, hence the term polarization grating. Large phase tilts required for HAM smear out fringes in the direction of the phase ramp for any non-zero bandwidth, while a major advantage of liquid-crystal technology is the high efficiency over large bandwidths. However, like segment tilting, there is total freedom to choose where baselines are imaged off-axis. If the fringes are orthogonal to the phase ramp, no amount of wavelength dispersion decreases the fringe visibility. We demonstrate this is Fig. 6.2. In this case, the broadband capabilities of the liquid-crystal technology can be fully exploited for broadband observations. A second advantage of using liquid-crystal GPHs, is that it is possible to write any phase pattern. We exploit this by combining polarization gratings into a single phase pattern that images a single subaperture onto multiple locations in the focal plane. The mathematical description of multiplexing can be found in Doelman et al. (2018). An example of multiplexing blazed gratings is shown in Fig. 6.3. Multiplexing creates great freedom to image any subaperture to multiple locations on the focal plane with high efficiency. However, multiplexing also gives rise to higher-order diffraction effects, that could potentially create interference between unwanted subapertures, see Fig. 6.3.

To summarize, the holographic component of HAM is created using geometric phase holograms, and the properties that will affect the HAM design are as follows.

1. Subapertures can be imaged onto any location in the focal plane using phase ramps.

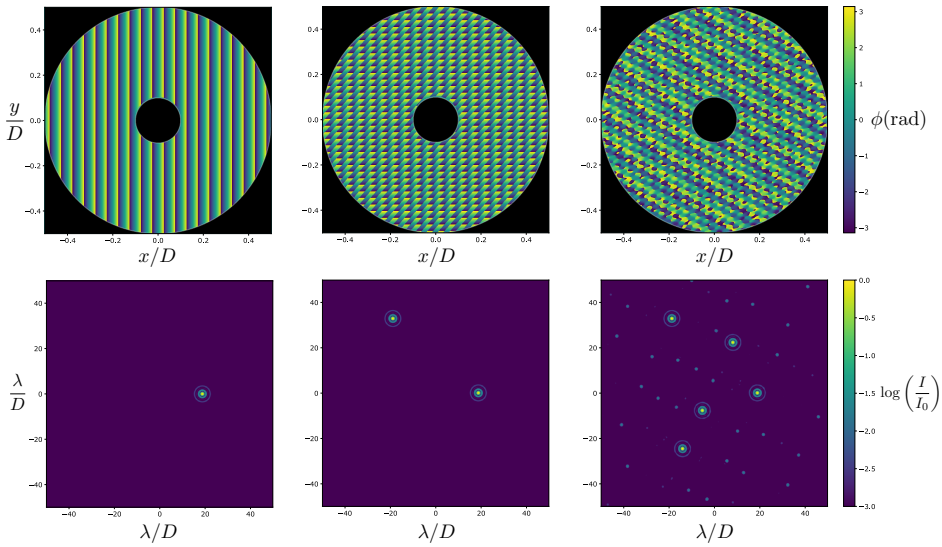


Figure 6.3: Simulations of combining multiple blazed gratings in a single hologram with multiplexing. The aperture multiplexes one, two and five holograms. Multiplexing multiple gratings increases crosstalk, i.e. the fainter PSFs seen easily in the 3rd panel.

2. Phase ramps can be combined for a single aperture with multiplexing.

3. A GPH is a diffractive phase element and the location of the PSF will change with wavelength.

4. A phase ramp will produce two PSFs at opposite location, one for each circular polarization state.

6.2.2 General considerations for a HAM design

Now that we have established how the liquid-crystal technology can be used to create off-axis interferograms, we can explore the design space it offers. The design of a HAM mask can be separated into several steps.

1. *Central component.* Select a SAM design that becomes the central component.

2. *Holographic component.* Select combinations depending on the number of closure phases versus the number of holograms and S/N per baseline.

3. *Focal plane.* Select the distances of holograms from the central component, which determines the spectral resolution of each baseline.

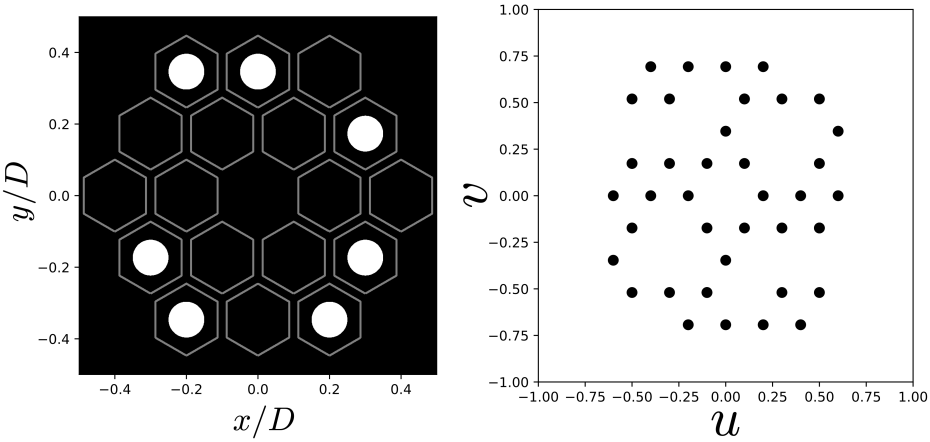


Figure 6.4: Example of a semi-redundant SAM design and the uv-coverage.

4. *Dynamic range.* Change piston phases of subapertures in the central component to reduce peak values.

We will elaborate on each of these points, adding design considerations.

The Central component

The goal of HAM is to increase the number of closure phases, throughput, simultaneous bandwidth and spectral resolution compared to a sparse aperture mask. A defining feature of aperture masking is that the detector information density, i.e. the number of baselines and closure phases, rapidly increase with the number of non-redundant subapertures that are imaged onto the same location. The diffractive nature of a liquid-crystal HAM mask limits the number of apertures that can be combined into a single hologram. As such, the highest information density on the detector can be reached by including a SAM mask that is imaged on-axis. Therefore, the design of a HAM mask starts with optimizing a SAM mask. Details on this optimization can be found in Carlotti & Groff (2010) and Tuthill (2018). We select a 7-aperture semi-redundant SAM mask for the demonstration of HAM design, as shown in Fig. 6.4.

The holographic component

With the central component defined, there is still $\sim 80\% - 90\%$ of the mask that can be used for the holographic component. Combined with the possibility to multiplex each subaperture, there are an extremely large number of combinations possible. Here we will explore the parameter space and list design rules and some of the trade-offs that can be made.
The design rules provide handles for a HAM design, yet they do not limit the

design freedom to a single solution, as this is also dependent on the science case. The design rules come from the properties of the geometric phase holograms listed in Section 6.2.1. From the diffractive nature (GPH property 3), we derive the first design rule, which also has the largest implication on the design.

1. Make 1D combinations of subapertures and image them onto a single interferogram that is placed orthogonally to their baselines.

If the baselines of subapertures are not coaligned, the fringes will smear out due to the dispersion not being orthogonal to the fringes. Only a limited number of subapertures can be arranged in a non-redundant way in a single dimension. Therefore, the number of closure triangles per interferogram is inherently restricted. Adding more independent closure triangles therefore requires GPH property 2: multiplexing. Multiplexing provides great design freedom but also has drawbacks, which is why design rule 2 states:

2. Minimize multiplexing.

The first two design rules that are in tension with each other. First we will explore the reasons multiplexing should be minimized before we look at the trade-offs that can be made.
Multiplexing splits up the light of a single aperture and directs it to multiple focal-plane locations. Adding one baseline, or multiplexing one additional grating, reduces the light in all other baselines of this subaperture. To demonstrate this, we look at a holographic component with n subapertures where every baseline has a separate interferogram. The number of holograms grows as $n(n-1)/2$, every subaperture is multiplexed $(n-1)$ times, and the number of closure phases grows as $(n-1)(n-2)/2$. So the number of holograms divided by the number of closure phases, i.e. the detector density, goes as $n/(n-2)$. However, the intensity of the interferograms decreases with $(n-1)$ due to the multiplexing. Overall, multiplexing does decrease the signal to noise ratio of all interferograms rather quickly. This is unlike the central component, where adding more subapertures does not decrease the visibility amplitude of baselines. This fundamental principle of the holographic component demonstrates why creating many closure triangles comes at a large cost.
A way to optimize the holographic component is to maximize the number of closure phases per holographic interferogram divided by the number of gratings multiplexed in a subaperture. That way we maximize the intensity of the holograms in combination with the number of baselines and closure phases, i.e the information density. This information density is highest when all supabertures combinations that contain coaligned baselines are imaged onto a single focal-plane location. For example, three subapertures imaged onto one location reduces the number of interferograms from 3 to 1. In conclusion, in order to minimize multiplexing while keeping a sufficient amount of closure triangles, it is key that a subset of subapertures is chosen that can be combined in multiple one-dimensional arrays. We define a subset as a collection of subapertures that are interfered with all other

Table 6.1: Specification of the holographic components of the three designs presented in Fig. 6.5. The central component (CC) is added as reference. $N_{subapertures}$ is the total amount of subapertures that are used in the holographic component, $N_{baselines}$ is the number of unique baselines, and $N_{holograms}$ is the total number of holograms. Moreover, N_{CLPH} is the number of closure phases, and $N_{Multi,max}$ and $N_{Multi,mean}$ show the maximum and mean amount of gratings multiplexed in the phases of the subapertures in the holographic component, where g and r indicate the green and red subset.

Design	1	2	3	CC
$N_{subapertures}$	9	10	11	7
$N_{baselines}$	9	15	19	19
$N_{holograms}$	6	28	46	-
N_{CLPH}	3	12	18	15
$N_{Multi,max}$	1	4	g: 6, r: 3	-
$N_{Multi,mean}$	1	3.1	g: 5.14, r: 3	-

subapertures in that collection and no other subapertures. A subset can have multiple interferograms, or only one like the central component.

3. Combine subapertures in multiple subsets.

Combining many subapertures leads to many baselines. The number of holograms that can be imaged onto the detector depend on the detector size, the subaperture diameter and the bandwidth of the filter. Different subsets of subapertures can be created to add baselines and closure triangles when an increase of the number of holograms for a single subset goes beyond the amount that would fit on the detector. Another advantage is that making multiple subsets reduces the amount of gratings that are multiplexed in single apertures. Again, this shows the trade-off between adding one subaperture that gives many new baselines and closure triangles but reduces the SNR of all holograms, versus adding a few new subapertures with higher SNR but less unique closure traingles. We will explore this trade-off in more detail using the designs in Fig. 6.5.

4. A subset is non-redundant if all interferograms, not baselines, are non-redundant.

Rule 4 adds more freedom for designing subsets, which becomes more useful when most of the mask is already filled with subapertures. A limit is set by the available detector space in the direction orthogonal to the redundant baselines, in accordance to rule 1, which automatically leads to a more homogeneous spread of baselines in all directions for square detectors.
With these design rules in mind, we can explore some of the possibilities of HAM designs. This is a similar trade-off that is present for SAM design, trading subaperture size with the number of non-redundant baselines that fit onto a mask (= SNR vs. *uv*-coverage). A mask with large subapertures and limited *uv*-coverage

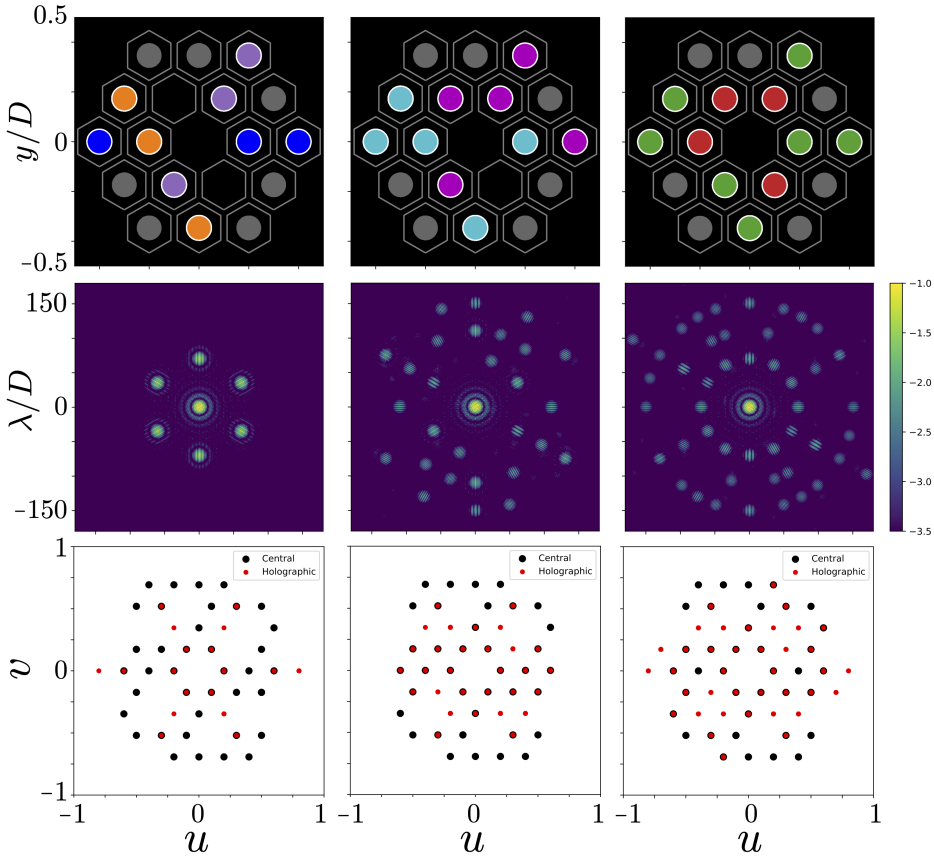


Figure 6.5: Three different designs of the holographic component. *Top:* pupil designs, where the colors indicate the subapertures that are combined in single or multiple holographic interferograms. *Middle:* focal plane image of the mask on logarithmic scale. *Bottom:* uv-coverage of the masks, with the central component in black and the holographic component in red.

provides requires less observation time to characterize targets with low complexity, e.g. a binary system, than a mask with smaller holes and fine sampling in the uv-plane. However, the finer sampling is key to retrieve complex structures of a source. For HAM the trade-off is also influenced by the available detector space and available bandpass.
We explore three designs, each with a different SNR and uv-coverage. The designs, their PSF and the uv-coverage are presented in Fig. 6.5, and the specifications are shown in Table 6.1. The first design optimizes the SNR, consists of three subsets of 1D combinations of subapertures without multiplexing. With only nine unique baselines and three closure phases, the holographic component is not optimal for operating in stand-alone mode. However, it is a good design for extreme broadband observations, e.g. $\Delta\lambda/\lambda_0 \sim 100\%$. Moreover, with enough sky-rotation it can still provide sufficient uv-coverage. Design 3 is the complete opposite of the first design, with extreme uv-coverage but low SNR. The number of unique baselines is equal to the SAM mask, and the number of closure phases even exceeds the SAM mask design. Each subaperture is multiplexed 4.3 times on average, some even six times. The focal plane is crowded, filling most of three rings with holograms that have less than 1% peak intensity compared to the central component peak intensity. Together, both components almost fill all possible locations in the uv-plane that correspond to the hexagonal grid of the mirror segments. On average the holograms are quite close together, limiting the simultaneous spectral bandwidth. A compromise of these two extremes is presented in design 2. Two components of five subapertures create 12 closure phases and 15 baselines, which comes close to the central component design. This design only adds four new baseline with respect to the design of the central component. Compared to the green subapertures in design 3 the holograms have increased a factor 1.7 in brightness. This trade-off makes the second design suitable for stand-alone mode for a more diverse range of targets.
It is worth pointing out that this specific example assumes a hexagonal grid of subapertures and an equal size of the subapertures in both components. For example, a good design will more look like design 1 if the subapertures are small in comparison to the full aperture. Not only are there more non-redundant subapertures available in a line, the off-axis holograms are also larger. As such, it is difficult to generalize the example. A complete review on design trade-offs is outside the scope of the paper.

Focal plane design

We switch to the focal plane to explore the effects of hologram placement in more detail. If the subsets of apertures are chosen, the focal plane has a set of holograms that need to be given a location. Each hologram can be imaged onto a line, where the separation with respect to the central component is a design freedom. This impacts spectral resolution, spectral bandwidth, and interference between higher-order terms. First, we will explore the spectral resolution of holograms.
The spectral resolution of the holographic component is actually defined by the

maximum of two terms. Both terms are independent and are defined in different planes, i.e the focal plane and the uv-plane, and are derived in Appendix 6.10. The first term has to do with the diffraction by the gratings of the HAM phase mask. For two subapertures with a subaperture diameter D_{sub}, a phase grating with period P, the spectral resolution in the focal plane is given by

$$R_{fp} = \frac{\lambda}{\Delta\lambda} = \frac{D_{sub}}{1.22P}. \tag{6.2}$$

The second term is a fundamental property of interferometry, where the uv-points, or baselines, are wavelength dependent due to diffraction. If the two subapertures form a baseline $\mathbf{b}$, their spectral resolution in the uv-plane is given by

$$R_{uv} = \frac{\lambda}{\Delta\lambda} = \frac{|\mathbf{b}|}{D_{sub}}. \tag{6.3}$$

With different methods we are able to fit fringes directly in the focal plane or retrieve complex visibilities from the uv-plane independently. We can choose which method to apply to obtain the highest spectral resolution. For most holograms $\frac{D_{sub}}{1.22P} > \frac{|\mathbf{b}|}{D_{sub}}$, yet it can be worthwhile to fill the image plane closest to the central component with the longest baselines. A maximum spectral resolution is obtained when the holographic interferograms are placed near the edges of the detector. Doing so introduces a larger sensitivity to the effects of non-perfect optics. This is particularly problematic for closure triangles where the baselines are imaged at different parts of the detector. For example, image distortion can change the shape and location of the holographic interferograms. Variable image quality across the focal plane can add different phase offsets to each baseline, resulting in a non-zero closure phase. A full analysis of these effects is outside of the scope of the paper. Another consideration is the crosstalk between multiplexed gratings, as shown in Fig. 6.3. By minimizing multiplexing this effect is already reduced, however, it is also possible to reduce the impact by switching circular polarization states of interferograms. Changing the sign of the grating frequency of gratings changes the crosstalk locations. Moreover, changing the sign of the grating frequency changes the circular polarization state, potentially making the interferogram incoherent with the crosstalk term. This is useful if an interferogram is near a crosstalk PSF, or if two interferograms have similar baseline directions.

Dynamic range

A last design step is to change the peak fluxes of the central component to reduce the brightness ratio to the holographic components. Due to multiplexing, the light from any HAM subaperture is redirected to multiple image plane locations. HAM interferogram spots therefore have a fraction of the intensity compared to the central component. For example, a central component can be the interferogram of nine holes, compared to a holographic interferogram that may only combine one third of the intensity of two holes: an issue made still more pronounced when

off-axis light is spread by bandwidth smearing. This can lead to more than an order of magnitude difference between the peak intensities of the holographic and central components.
All subapertures of the central component constructively interfere on the optical axis, leading to a highly peaked PSF. Nulling these peaks of the central component reduces the required dynamic range significantly. The nulls are generated by changing the piston phase offsets of individual subapertures of the central component. It is not necessary to reach a perfect null, the goal is only to homogenize the intensity of the central component. To this end, the piston phase offsets could be random, as long as the piston was corrected for in the closure phase calculations. However, liquid-crystal technology limits us to 0 or $\pm\pi$ phase offsets due to applying opposite phase to the opposite circular polarization state, i.e. Eq. 6.1. Other piston phases would make the baseline appear as redundant with two piston offsets. As a phase piston of 0 or $\pm\pi$ is the optimal way of nulling, this is not a problem. Another advantage of geometric phase is that the phase offsets are achromatic, such that the interferogram does not change with wavelength. Selecting which subapertures apply the phase offset depends on the central component design. To null the central peak, half of the subapertures should acquire the phase piston. This is not possible for designs with an odd number of subapertures. When the central component design contains doubly redundant baselines, there is an additional constraint. The identical baselines need to have equal piston offsets or they are nulled. An example of the nulling of the central component is presented in Fig. 6.6. The peak brightness in the three designs, normalized on the first and from left to right, are [1,0.41, 0.53]. So nulling reduces the peak intensity by a factor two.

6.3 The influence of polarization leakage

Both the central component and a polarization leakage term are imaged onto the optical axis of the system. Here we wile the influence of polarization leakage on the central component, and we will explain its effects on HAM observations. Polarization leakage emerges when the retardance of a geometric phase hologram is not exactly half-wave. As mentioned before, the geometric phase hologram is a half-wave retarder with a spatially varying fast axis. The space-variant Jones matrix of such a retarder in the circular polarization basis is given by

$$\mathbf{M} = c_V \begin{bmatrix} 0 & \mathrm{e}^{i2\chi(x,y)} \\ \mathrm{e}^{-i2\chi(x,y)} & 0 \end{bmatrix} + c_L \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{6.4}$$

Here $\chi(x, y)$ is the spatially varying fast-axis orientation, and both c_V and c_L are parameters that depend on the retardance $\Delta\phi$ (Mawet et al., 2009; Ruane et al., 2019) and are given by

$$c_V = \sin\frac{\Delta\phi}{2},\ c_L = -i\cos\frac{\Delta\phi}{2}. \tag{6.5}$$

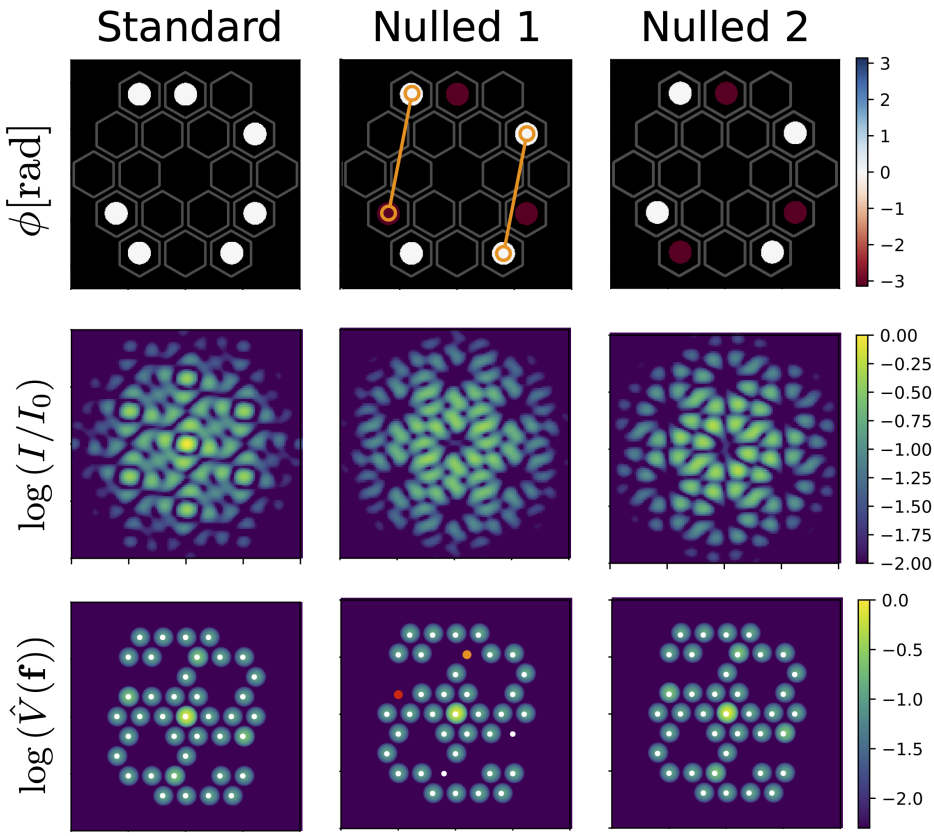


Figure 6.6: Phase, PSF and the visibilities of the central component with and without nulling. Adding π phase offsets to different holes shapes the PSF and reduces peak intensity. If doubly redundant baselines (orange) have different phase offsets, the baseline is nulled achromatically.

The first term in Eq. 6.4 describes that a fraction C_V of the light acquires a geometric phase of

$$\Phi(x, y) = \pm 2\chi(x, y), \tag{6.6}$$

where the sign of the phase depends on the handedness of the incoming circular polarization. The second term describes the polarization leakage beam, and is unaffected by the fast-axis orientation pattern. When the retardance is perfectly half-wave, i.e. $c_V = 1$ and $c_L = 0$, the right-handed circular polarization states are converted to

$$\mathbf{RC_{out}} = \mathbf{MRC_{in}} = \mathbf{M}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\\ \mathrm{e}^{-i\Phi(x,y)}\end{bmatrix}, \tag{6.7}$$

With non-perfect retardance, the output electric field for incoming right-circular polarization is given by

$$\mathbf{RC_{out}} = \mathbf{MRC_{in}} = \mathbf{M}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}c_L\\ c_V\mathrm{e}^{-i\Phi(x,y)}\end{bmatrix}. \tag{6.8}$$

The leakage term and main beam have an orthogonal polarization state and are therefore incoherent, assuming no polarization cross-talk due to the optical system. In addition, unpolarized light contains on average equal amounts of left- and right-circular polarization and these states are incoherent, see Hecht & Zajac (1974). Therefore, we can describe the impact of leakage, typically on the order of 1%, for 1 circular polarization state without loss of generality. While objects do not appear fully unpolarized due to instrumental polarization or interstellar polarization, we will first explore this simplification because it demonstrates how closure phases and complex visibilities are less resistant against wavefront aberrations.
From Eq. 6.8, it is clear that the point-spread function (PSF) is actually an incoherent sum of the HAM PSF and the leakage PSF, where the leakage PSF is the unaltered PSF from the HAM amplitude mask. Moreover, the Fourier transform is a linear operator. Therefore, we can calculate the visibilities of the HAM PSF and the leakage PSF separately and co-add them. The combined measured visibility is then given by

$$V(\mathbf{f}) = c_V^2 V_{SAM}(\mathbf{f}) + c_L^2 V_{holo}(\mathbf{f}). \tag{6.9}$$

Here, $V_{SAM}(\mathbf{f})$ is different from $V_{holo}(\mathbf{f})$, as their amplitude masks are different. The leakage term contains many redundant baselines for realistic HAM designs, as it is the combination of the subapertures of the central and holographic component. Therefore, $V_{holo}(\mathbf{f}$ is a sum over all redundant baselines, each with a different phase. While $C_V^2 \gg C_L^2$ for a retardance close to half-wave, $c_L^2 V_{holo}(\mathbf{f})$ can be a significant fraction of $V(\mathbf{f})$ if their phases are non-random. Any phase aberration has a non-zero sum of phase for the redundant baselines in the HAM mask, and adds an offset to the visibility phase that is unique per baseline. This offset is affects both the closure phase retrieval as well as the squared visibilities. The impact depends on aberration strength and time dependent wavefront errors change the closure phases in a way that can not be calibrated by observing a calibrator.

For any linearly polarized light fraction it is not correct to assume that the leakage PSF is incoherent. Linearly polarized light can also be written as the sum of the two circular polarization states, however, they are still coherent. That means that the polarization leakage of left-circular polarization interferes with the main beam of the right-circular polarization state that becomes left-circularly polarized after going through the half-wave retarder, i.e.

$$\mathbf{RC_{out}} = \mathbf{MLP_{in}} = \frac{1}{\sqrt{2}}\mathbf{M}\begin{bmatrix}1\\-i\end{bmatrix} = \begin{bmatrix}c_L - ic_V \mathrm{e}^{i\Phi(x,y)}\\ c_V \mathrm{e}^{-i\Phi(x,y)} - ic_L\end{bmatrix}. \tag{6.10}$$

The PSF will contain an interference term with a relative intensity $\sim C_V C_L$, which is smaller than C_V^2 when the retardance is close to half-wave, but much larger than C_L^2. The effect of this interference depends on many factors, including the retardance of the HAM mask, the linear polarization fraction, the wavefront aberrations and the nulling of baselines using $\pm\pi$ offsets. This effect is strongest for 100% linearly polarized light coming in on the HAM mask, but can be significant even for small fractions of linearly polarized light at the level of 1%. Moreover, the equations are the same for polarization crosstalk due to instrumental polarization. This means that time-variable polarization states, e.g. induced by a derotator at different orientations (van Holstein, R. G. et al., 2020), cannot be calibrated with a single calibrator at the beginning of the observing sequence. Similarly, differences in linear polarization state between a calibrator and a science target induces a non-zero signal in closure phase that can also not be calibrated. Simulations of a mask design in Appendix **??** show that the maximum deviation of closure phases is linear with the degree of linear polarization. We find that a linear polarization fraction of 1% corresponds to a 1 degree offset for 2.5% percent leakage. In the same fashion, the $\pm\pi$ phase offsets of some subapertures are also impacting the retrieved complex visibilities when leakage is present. A unique baseline from the central component with this phase offset is nulled by multiple non-shifted baselines from the polarization leakage PSF. This leads to a reduction of squared visibility of phase-shifted baselines. Wavefront aberrations further complicate the picture, and calibrating the squared visibilities is not possible with a calibrator. Overall, extreme caution is warranted when the linear polarization fraction is non-zero and the polarization leakage is above 1%.

There are two solutions that will limit the effect of polarization leakage. First, one can filter circular polarization states before and after the HAM optic, similarly to what is done for the vector-vortex coronagraph (Foo et al., 2005; Mawet et al., 2009). For an existing instrument this is usually not possible to implement, although a HAM optic could be sandwiched between circular polarizers to make it into a single optic. The downside of such a solution is the 50% throughput reduction. The second solution is a double-grating HAM optic (Doelman et al., 2020). Adding a polarization grating pattern to the full phase pattern, followed by a second identical polarization grating, reduces the polarization leakage by at least one order of magnitude. This solution is explored further in Section 6.9.

Table 6.2: Specifications of the full HAM mask (right), decomposed into contributions from the central component (CC, left) and the holographic component (HC, middle).

	CC	**HC**	**HAM**
Subapertures	11	19	30
Throughput	10.1%	21.9%	32.0%
Baselines	55	30	85
Unique baselines	50	23	51
Single baselines	45	17	27
Repeated baselines	10	13	58
• Pairs	5	5	16
• Triplets	0	1	7
• Quadruplets	0	0	0
• Quintuplets	0	0	1
Closure Triangles			
• w/ repeated SAM baselines	165	26	191
• w/o repeated SAM baselines	88	26	114
Unique Closure Triangles			
• w/ repeated SAM baselines	165	26	190
• w/o repeated SAM baselines	88	26	114
***uv*-points**			
• w/ repeated SAM baselines	100	46	102
• w/o repeated SAM baselines	90	46	98

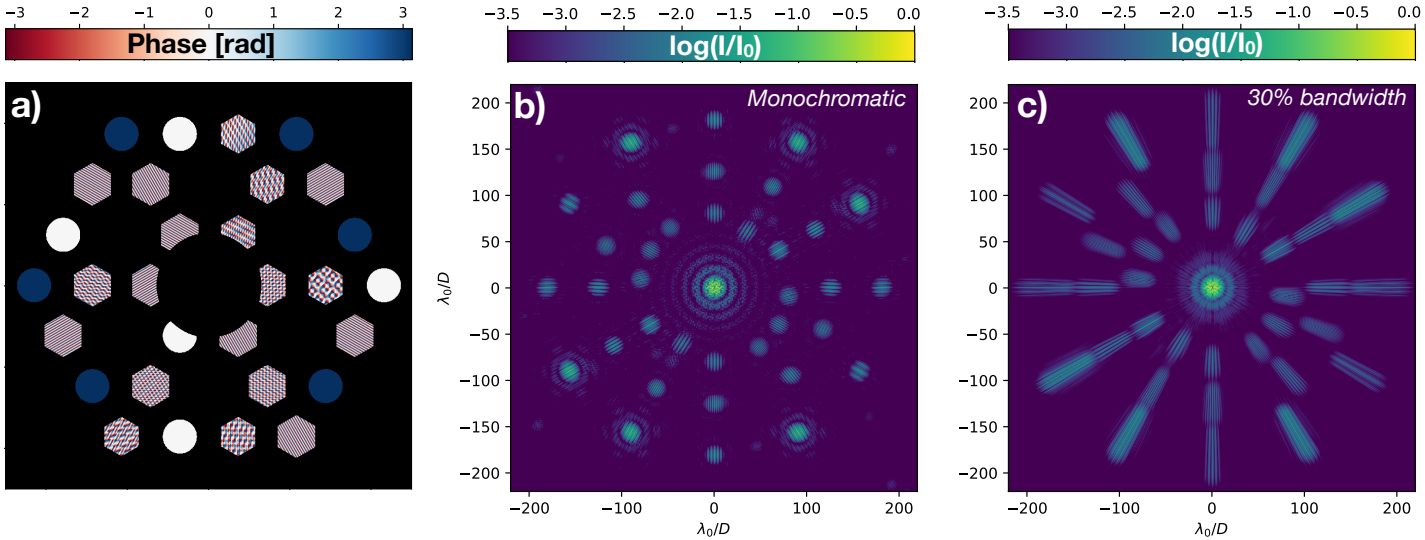


Figure 6.7: The design of the HAM mask for OSIRIS. (*a*) Phase pattern of the HAM optic, masked by the amplitude mask. (*b*) The simulated monochromatic PSF. (*c*) The simulated PSF with 30% bandwidth.

6.4 Design of a prototype mask for Keck OSIRIS

In this section, we discuss the design of a prototype HAM mask for the OH-Suppressing Infrared Integral Field Spectrograph (OSIRIS) (Larkin et al., 2006). The design follows the considerations mentioned in Section 6.2.2. The OSIRIS detector, a Teledyne Hawaii-2RG HgCdTe detector with a size of 2048×2048 pixels, provides a unique opportunity. With a plate scale of 10 mas/pixel, the detector is large enough to add many baselines to the holographic component. We place the subapertures in alignment with the Keck primary mirror segmentation, such that every subaperture is centered on one segment. The segments that are crossed by a spider are not used in the design. Moreover, the subapertures of the holographic component are hexagonal to increase their throughput, and we increase their diameter if the smallest baselines are larger than one the distance between neighboring subapertures. The subapertures of the central component are circular as this keeps the central component PSF circularly symmetric and the outer Airy rings have equal strength for all holograms at equal radius. The full design and both a monochromatic and broadband PSF are shown in Fig. 6.7.

The central component consists of an 11-hole SAM mask. It is optimized to maximize throughput, the number of baselines, and the number of closure triangles at the cost of a few redundant baselines. Tuthill (2018) showed that adding redundancy allows for a boost the S/N ratio with respect to non-redundant masking by $\gtrsim$50%. In addition, we offset five holes of the central component by π to reduce the required dynamic range of the detector and the influence of polarization leakage, as explained in Sect. 6.2.2. This specific combination of phase-shifted subapertures minimized the PSF of the central component at the location of the polarization leakage PSF maxima and does not null redundant baselines.

The holographic component uses the remaining apertures, for a total of 19. Nine subapertures are used for three linear combinations of three holes each, yielding nine baselines and three closure phases for three holograms. Because these holograms are not multiplexed, they yield the interferograms with the highest signal-to-noise ratio. The other ten subapertures are devided into a highly-multiplexed combination of six holes and a combination of four holes. These designs are optimized to maximize the number of baselines and closure triangles per hologram. Overall, the holographic components consists of 18 holographic interferograms per polarization, yielding 26 closure triangles.

The subaperture locations and their mapping to the focal plane can be found in Appendix **??**. The full overview of HAM properties is provided in Table 6.2. Additionally, it makes a distinction between the individual contributions from the central component and holographic component. Several numbers are worth pointing out. Apart from transmission and diffraction efficiency, the total throughput of the mask is $> 30\%$, which is a factor of 3 higher than the throughput of the central component. Whereas the central component features 10 repeated baselines (i.e. five pairs of "redundant" baselines), 13 out of the 30 HAM baselines have at least one identical counterpart. However, because the light is mapped onto different spots in the focal plane, their complex visibility can be computed independently.

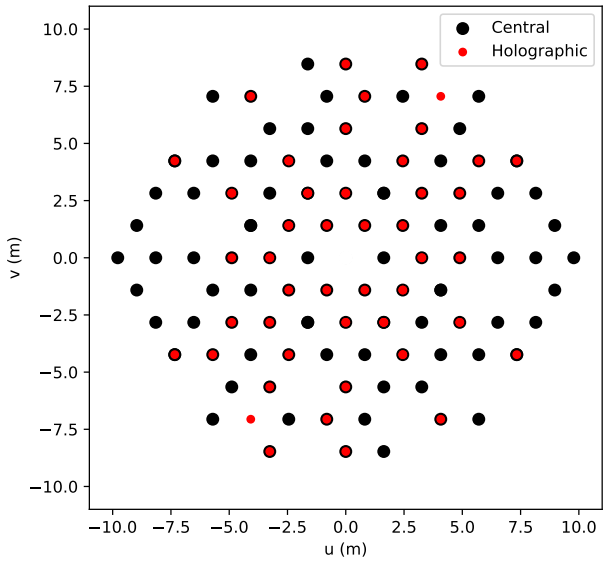


Figure 6.8: The *uv*-coverage of both the central and holographic component.

The HAM mask has a total of 85 baselines, 51 of which are unique. The unique baselines can be divided into two groups: (i) *single* baselines, which occur only once, and (ii) *repeated* baselines, which occur either 2, 3 or 5 times (see table 6.2). With the 50 unique baselines accommodated by the central component, the *uv*-plane is uniformly sampled and provides great coverage. The 23 unique baselines of the holographic component provide sufficient *uv*-coverage that allow it to be used on its own. Moreover, the baselines present in both components can be used to improve calibration of the data.

6.5 Data reduction pipeline

In this section, we will briefly go over the new pipeline that was developed for HAM data reduction.

6.5.1 Pipeline overview

Generally, pipelines for aperture masking consist of the same building blocks.

1. *Pre-reduction.* This includes flat-field corrections, bad-pixel corrections and background subtraction;

2. *Complex visibility.* The phase and amplitude of the complex visibility are measured on each baseline. To this end, there exist two approaches: (i) fringe-fitting methods, which rely on measuring the fringe phases and amplitudes directly from the focal plane, e.g. Lacour et al. (2011), Greenbaum et al. (2014), and (ii) Fourier methods, which extract information from the Fourier Transform of the PSF, as shown in Figure 1.4, e.g. Tuthill et al. (1999), Sallum & Eisner (2017);

3. *Closure phases.* The visibility phases are combined to yield robust observables, such as closure phases and squared visibilities;

4. *Calibration.* Closure phases of a point source are subtracted from the closure phases of the science target, to correct for systematic effects;

5. *Model fitting.* By fitting the data with an analytical source model, one can estimate the source parameters.

The HAM pipeline is written in Python using the `HCIPy` package (Por et al., 2018). All operations are carried out separately for the central and the holographic component. The pipeline uses both methods to retrieve the complex visibilities. For the fringe-fitting method, we apodize the PSF of the central component using a power 2 super-Gaussian window function with a FWHM of 110 λ/D in both axes. This separates the central component from the holographic component, in addition to suppressing high-frequency noise in the uv-plane. Similarly, for the holographic component we apodize the PSFs with a power 2 super-Gaussian with a FWHM of 33 λ/D, placed at the location of the interferogram. This location is pre-computed according to the plate scale of the data. The pipeline builds a fringe-library for every interferogram at their own location, such that a direct fit can be made without shifting any interferograms. The fit is a least-squares optimization with a model matrix that contains the flattened versions of all fringes in the fringe-library.
The Fourier method uses the same super-Gaussian masks for both components. Masking causes information from neighbouring uv-pixels to blend, such that the central pixel value is representative for the entire splodge. Visibilities are then extracted at the central locations of splodges in the uv-plane. For the holographic component we mask individual interferograms before doing the Fourier transform. It is not necessary to center on the interferograms. While this does introduce a large phase slope in the uv-plane, we know the location of the PSF and can subtract a pre-calculated phase slope. Moreover, if we assume that there is no distortion or only symmetric distortion in the image plane, we can average the visibilities of the interferograms with opposite circular polarization. As they are on exactly opposite sides of the image center, the phase slopes cancel.
The closure phases are calculated for each closure triangle from the individual visibility phases. After subrracting the closure phases of a calibration target, the pipeline can fit an analytical model to the residual closure phases. For now the only model is a binary object, and the fit is performed using least-squares minimization.

6.5.2 Spectroscopic closure phase extraction

The spectroscopic capabilities of HAM are also exploited in the pipeline, which is capable of extracting multi-wavelength closure phases from a single image. For the Fourier method this is an almost trivial addition, as the baseline length in the uv-plane can be scaled according to the wavelength. The sampling in the uv-plane can be changed accordingly with the use of a matrix Fourier transform (Soummer et al., 2007). In addition, we change the mask for the interferogram to be an elliptical super-Gaussian, with the major axis in the smearing direction. Extracting the fringe visibilities and closure phases is done at a wavelength sampling that is higher than the intrinsic spectral resolution. This spectral resolution is given by the dispersion in the uv-plane which depends on the wavelength and baseline length, see Eq. 6.3. Therefore, the spectral resolution between baselines that are combined in a single closure triangle might differ. We oversample the uv-plane compared to the highest spectral resolution and recover the wavelength-dependent closure phases using this sampling . As such, the recovered closure phases are not completely independent as function of wavelength.
Spectroscopic extraction of closure phases is different for the fringe-fitting method. Using the fact that a broadband PSF is the incoherent sum of multiple monochromatic PSFs, it is possible to fit a fringe to multiple wavelength-scaled locations of the holograpic interferograms. The spectral resolution is given by Eq. 6.2. If the monochromatic PSFs have a separation of $1.22\lambda/D_{sub}$, they can be used as individual fringes in the fringe-library and fitted simultaneously. However, extracting the fringes at a higher spectral resolution with overlapping PSFs requires regularization of the design matrix, as the columns are no longer linearly independent. We tried two methods of regularization, i.e. Tikhonov regularization and regularization using polynomial coefficients. For Tikhonov regularization, we construct the Tikhohnov matrix Γ such that it acts as a difference operator. As a result, the regularisation punishes large values of the finite differences of the fringe phase between two adjacent wavelength bins. While this enforces a smooth transition with wavelength, it could affect the calculated closure phases. In addition, Tikhonov regularization requires tuning which is computationally expensive and might change between targets. As this is not ideal, a different regularization method was explored. For simple systems like a binary system, the fringe phases behave smoothly with wavelength. A low-order polynomial could be a good estimate of such behavior, and, can therefore be combined with the fringe-fitting method as regularization. Instead of directly fitting fringe phases, we fit polynomial coefficients to the wavelength dependency of the fringe phases. The full equations are shown in Appendix **??**.
We simulate a binary with a separation of 1.08 λ/D, a brightness ratio of r = 0.5 and use all three methods described in this section to retrieve the broadband closure phases of the holographic component.The results for two closure triangles are shown in Fig. 6.9. We explore a different sampling in wavelength for every method, i.e. uv-plane sampling (Fourier), number of wavelength bins (Tikhonov) and polynomial order (Polynomial). For the Fourier method, the solutions barely

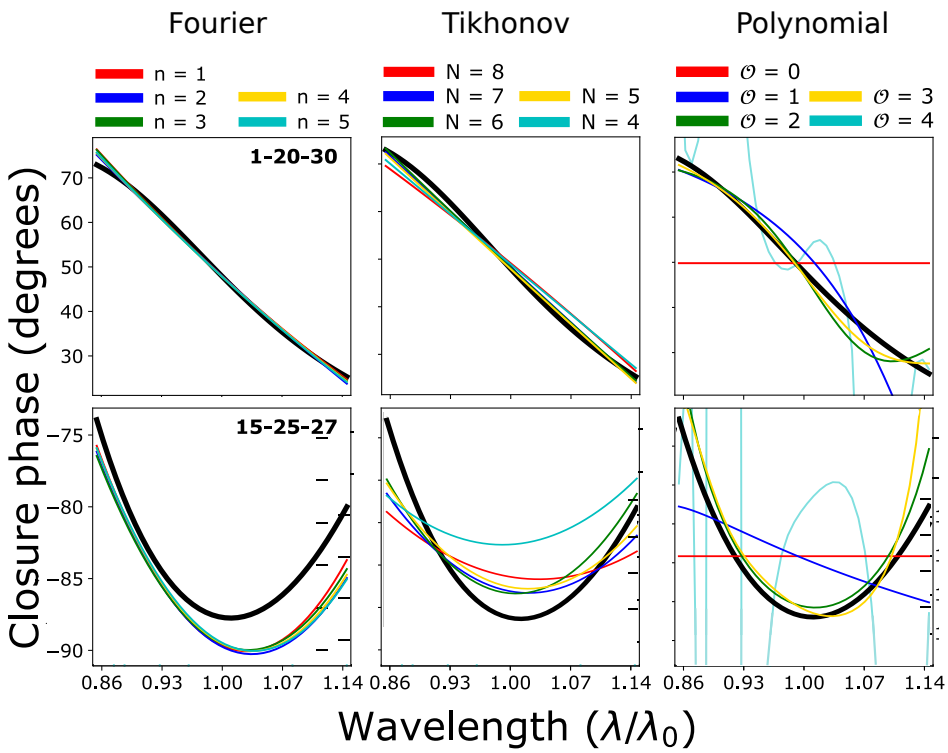


Figure 6.9: Analytical closure phases (black) and extracted closure phases (color) of a binary system applying different closure phase extraction algorithms to the holographic component. Two representative closure triangles were selected. The binary has a separation of 1.08 λ/D, a brightness ratio of r = 0.5. The simulated bandwidth is 30% (i.e. $\Delta\lambda/\lambda_0 = 0.3$). For the Fourier method we change the uv-plane sampling from n = 1 to 5, meaning $2D/1000$ to $2D/5000$. Closure phases extracted using fringe-fitting with Tikhonov regularization is carried out for a different number of wavelength bins N, and we vary the polynomial order $\mathcal{O}$.

depend on the uv-plane sampling, as long as it is high enough. The polynomial fit best captures the behavior of the analytical model at order two and three. The extreme behavior of the fourth order polynomial suggests over-fitting, implicating that the spectral resolution is not high enough to constrain more than four components. The Tikhonov extraction has similar performance with $N > 4$, and has decreased performance toward the edge of the band. This could be a direct result of the choice of regularization.
To summarize, all three methods have similar performance for the optimal sampling of the bandwidth. They are capable of reproducing the change of closure phases with wavelength within a five degrees over 30% bandwidth. The limited spectral resolution of the holographic spots is just enough to capture the change of baseline units with wavelength for this binary system. More rapidly changing closure phases, e.g. from a more structured object, are not sampled well and will be recovered as a smoothed function. This limits the detail that can be added to the model.

6.6 Manufacturing of prototypes HAM v1 and HAM v1.5

Two prototype HAM devices were manfactured by ImagineOptix in August 2018, labelled part A and part B. The first HAM device, part A, was manufactured using a 1 inch flat CaF2 substrate with a thickness of 5 mm, while the second HAM device, part B, was fabricated using a 1 inch wedged CaF2 substrate with a thickness of 1 mm. The front-sides have the same 3-layered liquid-crystal multi-twist retarder film, aimed at minimizing polarization leakage between 1 and 2.5μm. Both devices have an anti-reflection coating for the this bandpass on the backside of the substrates. The phase pattern with a diameter of 25.4 mm was generated with 5 micron pixels. The polarization leakage measured by the manufacturer is less than 3% between 1 and 2.5μm, see Fig. 6.14. Additional alignment markings have been added to the pattern, outside of the pupil diameter of 13.5 mm. An image of the optic between polarizers is shown in Fig. 6.10a. In addition, the optic was inspected under a microscope between polarizers. The four microscope images presented in Fig. 6.10b show the high quality of the manufacturing process. Amplitude masks were laser cut in 100um brass and/or 304 stainless steel foils with a diameter of 20.83 mm, using a picosecond laser machining facility (OptoFab node of ANFF, Macquarie University, Sydney). It is screwed in place in a holder in the filter wheel assembly. The first version (HAM v1) was installed in the imaging arm of OSIRIS at the Keck I telescope in September 2018. Each position of its first filter wheel contains separate pupil mask and filter mount assemblies. The amplitude mask was installed in the pupil assembly facing the incoming beam, while the HAM optic was installed in the opposite 1-inch filter side of the wheel assembly. Therefore, a gap of several mm was present between the amplitude mask and the HAM optic. This version has been tested with an internal source in OSIRIS in April 2019. The results are presented in section 6.8.

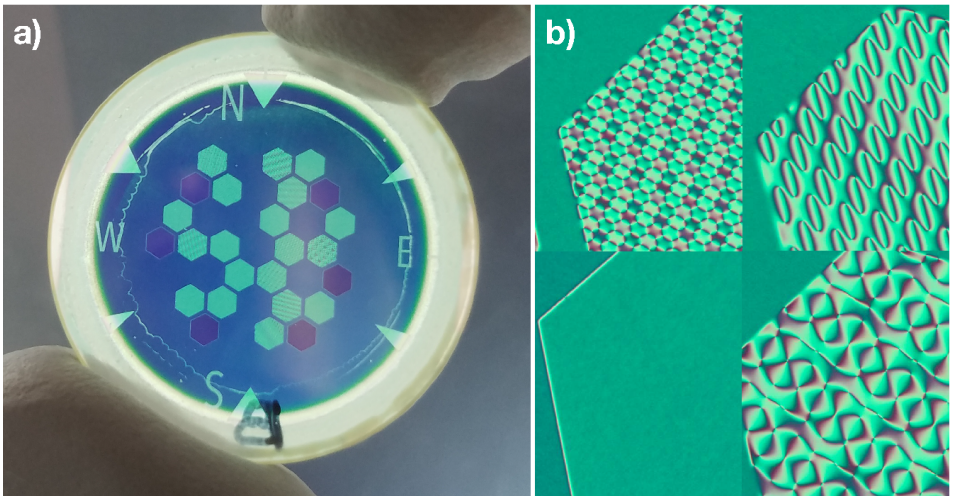


Figure 6.10: Images of the HAM optic between polarizers. Image credit: ImagineOptix

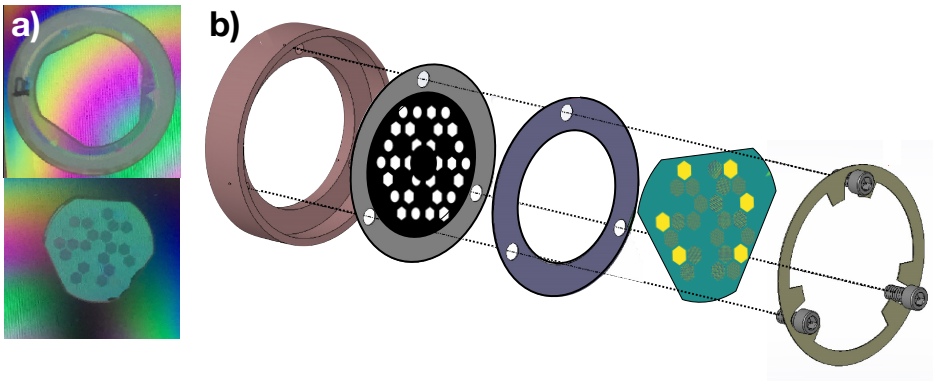


Figure 6.11: Manufacturing of HAM v1.5. (*a*) Images of the diced HAM v1.5 phase optic. (*b*) The assembly of HAM v1.5 in the OSIRIS pupil mount.

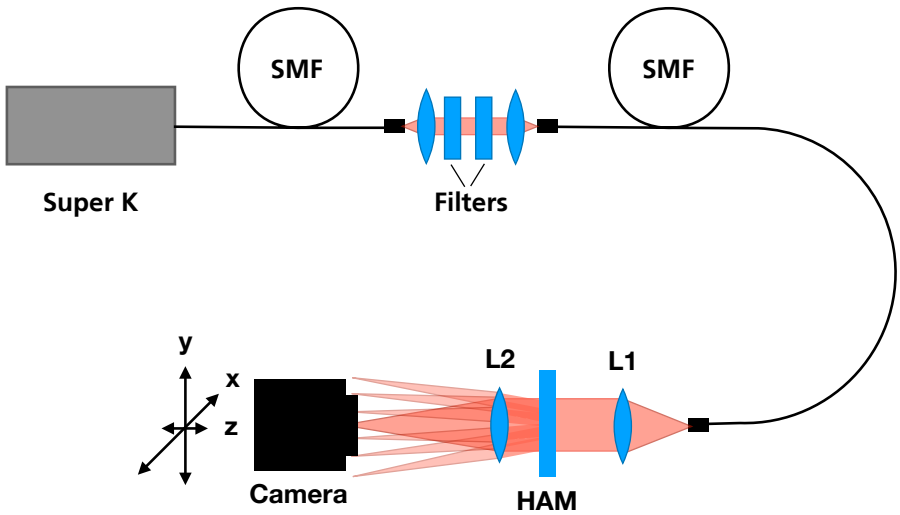


Figure 6.12: Lab setup used to characterize the HAM v1.5 optic.

An updated version (HAM v1.5) has been created to allow the HAM optic and the amplitude mask to be installed in the same pupil mask holder, reducing the separation between the optics to almost zero. The HAM optic of HAM v1.5 is a cut-out version of the spare HAM v1 phase mask (part B). The cut-out mask and assembly of the optic in the mount are shown in Fig. 6.11. Initial lab tests were conducted in Sydney in the Fall of 2019. HAM v1.5 has been installed in OSIRIS in February 2020, replacing HAM v1 in OSIRIS. We present the results of the lab tests in the next section.

6.7 Lab tests

We tested the HAM v1.5 optic in the lab using the setup described in Fig. 6.12. The light source is a SuperK COMPACT from NKT photonics and is connected to a custom reimaging system through a single-mode fiber (SMF). The reimaging system allows us to insert spectral filters and neutral density filters in a collimated beam before injection into the optical setup with a second SMF. We used filters from the Thorlabs IR Bandpass Filter Kit from 1000 nm to 1600 nm. Light from the second SMF is collimated with a Thorlabs 1 inch doublet with a focal length of 150 mm (AC254-150-C-ML). A second 300 mm (AC254-300-C-ML) doublet is placed close to the HAM optic to prevent vignetting of individual HAM apertures. The camera is a CRED2 and is mounted on motorized X,Y,Z translation stages to fully capture the HAM PSF. We used this motorized camera because a 2K pixel science grade detector, like a Teledyne Hawaii-2RG HgCdTe detector, did not fit

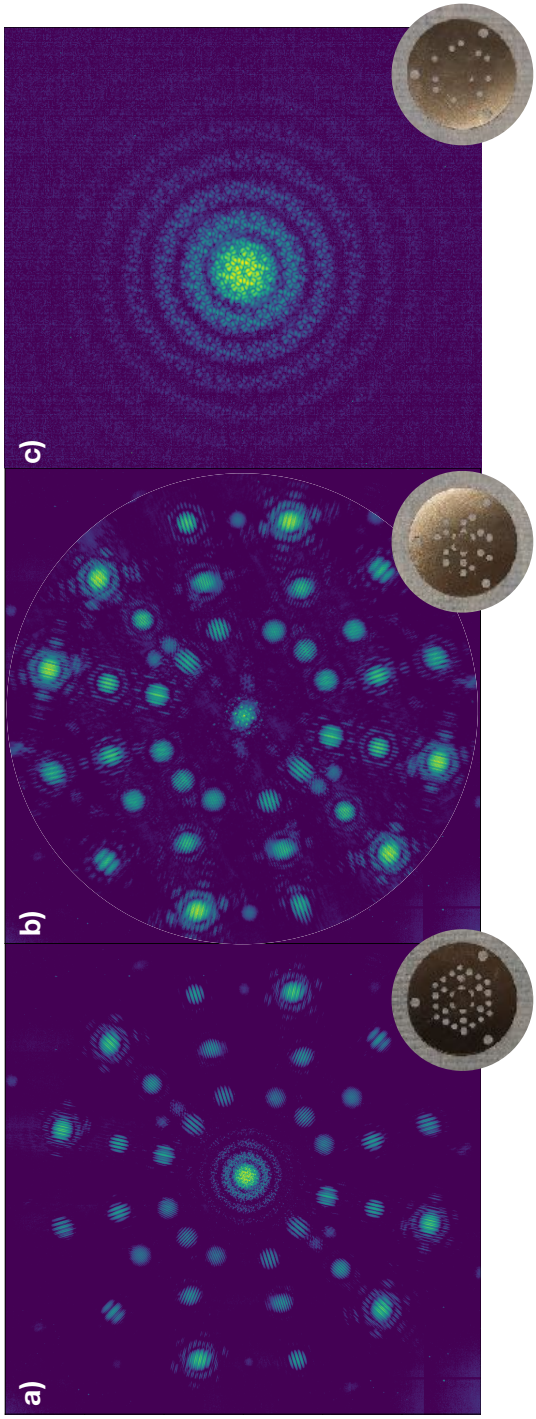


Figure 6.13: Images of the normalized HAM PSF with different amplitude masks at 1400 nm. (*a*) Full HAM PSF showing that all holograms are present and contain interferograms. (*b*) HAM PSF of the holographic component where the central term is zero order diffraction term of the holographic gratings. (*c*) Close-up of the central component PSF.

in the budget of the lab experiment. The full PSF is captured by recording images in a 5x5 grid. With this sampling neighbouring images now overlap significantly, and this overlap is use for more precise image registration between images. Both the CRED2 and the translation stages are controlled by a Matlab script that can capture images of the PSF. The background is estimated from the median of 100 images where the source is turned off. The background is captured before the image sequence, where for every position 10 images are averaged, and background subtracted.

6.7.1 The Point-spread function

The translation between neighbouring images are extracted using the scikit-learn `feature.register_translation` function on a masked PSF. Individual images are shifted accordingly and stored in a 3D array, where the first axis corresponds to the number of images and the other two to the x and y positions in the combined image. The final mosaic is the median along the first axis, which can be the single pixel value if a PSF region is imaged only once, or the median of multiple values when there is overlap. We remark that this method heavily relies on PSF stability, especially as fringes are the features used to align images with respect to each other. Any change to the fringe phase could lead to minor misalignment of images with respect to each other. We assume this effect is small as the overlapping region between neighbouring images contains multiple interferograms with different orientations. Any shift of the fringe phase in a single spot would not throw off the alignment. A passive setup can only generate changes in many closure phases simultaneously when optics move, which does not occur on timescales of two consecutive images. However, it demonstrates the sensitivity of fringe phases to the alignment of images with respect to each other. In addition, the PSFs are Nyquist sampled for the shortest wavelengths, meaning that even small subpixel shifts lead to large phase offsets. This is only a limitation of the lab setup, and will not affect the performance of HAM in the OSIRIS instrument where the PSF is fully captured by the camera.
We recorded images of the HAM PSF using three different amplitude masks. he different amplitude masks are the full HAM amplitude mask and two masks to isolate only the central and holographic components of the PSF. This allows us to analyze their individual PSFs and characterize the zero-order leakage of the HAM phase optic. The three masks and their corresponding PSFs at 1400 nm are presented in Fig. 6.13. The HAM PSFs closely resemble the simulated PSFs, and they confirm that the HAM optic was not damaged during the dicing process. The holographic component directly shows the zero-order leakage. Using `HCIPy`, we forward-model the HAM PSF with variable retardance. Changing the retardance changes the intensity of the central leakage term with respect to the holographic interferograms. By minimizing the normalized difference of the modelled and measured PSFs, we extract the retardance at different wavelengths. The fitting results are shown in Fig. 6.14 in addition to measurements of the efficiency of a polarization grating with the same liquid-crystal recipe. There is an offset of

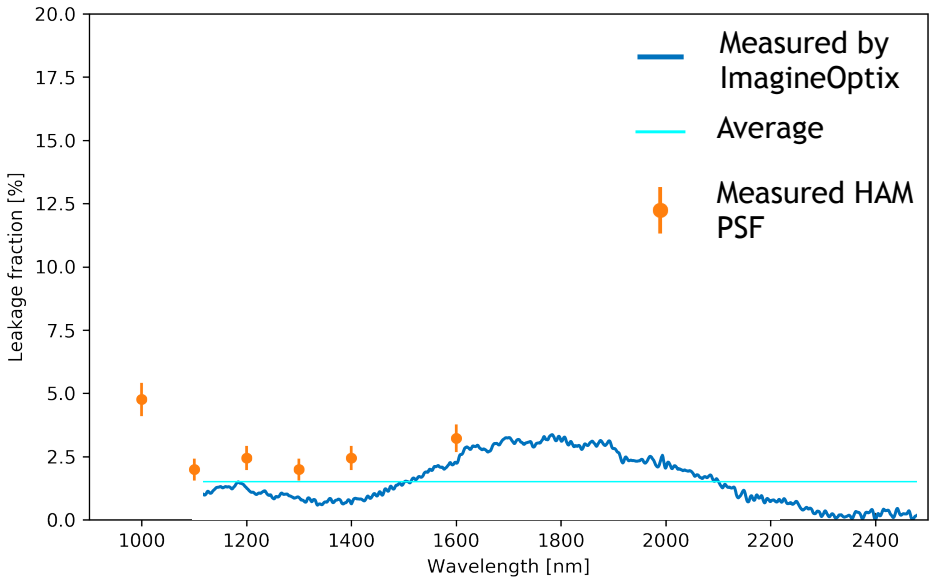


Figure 6.14: Measured zero-order leakage fraction of the HAM v1.5 optic.

roughly 1% between these two methods of measuring the zero-order leakage. A reduced diffraction efficiency for multiplexed gratings as compared to single gratings could explain this difference in zero-order leakage intensity. Multiplexing gratings with high frequencies leads to local phase patterns with extreme phase gradients, which might not be fully captured by the direct-write method. If that is the case, the diffraction efficiency is reduced. However, the presented measurements are unable to distinguish a change in diffraction efficiency due to multiplexing from a change in retardance. A consequence of the increased zero-order leakage is an increased sensitivity of the closure phases to polarized light, as discussed in Sect. 6.3.

6.7.2 Closure phases

We extract closure phases of both the central component and the holographic component of the full HAM PSF using our pipeline. In Fig. 6.15 we show the closure phases of the central component for 1400 nm, some with large deviations from zero. As explained in section 6.3 and in appendix **??**, these closure phases are non-zero due to the sensitivity to wavefront aberrations and the polarization state of the incoming light. Therefore, we fit a simple model that includes some low-order aberrations, a linear polarization fraction and polarization leakage to the closure phases. The low-order aberrations are represented with 10 Zernike modes starting with defocus. A monochromatic PSF is calculated using `HCIPy` with matching plate scale to the lab PSF, and our pipeline extracts the closure phases from the

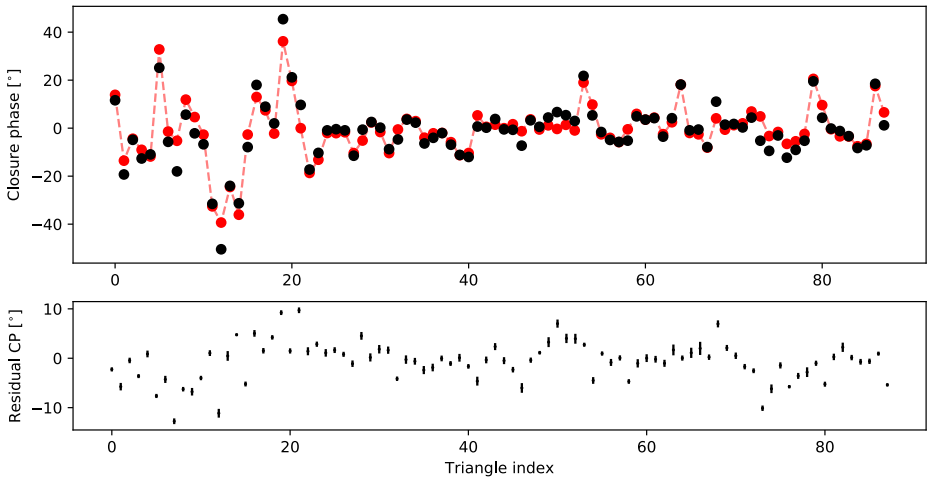


Figure 6.15: *Top:* closure phase of each closure triangle of the central component of the full HAM PSF at 1400nm (black), in addition to closure phases of a forward model (red). The forward model includes a linear polarization fraction of the source, polarization leakage and low-order aberrations. *Bottom:* The residuals between the model and observed closure phases. The error bars indicate the standard deviation of the residual closure phases during 22 hours of measurements at 15 minute intervals.

PSF. The best fit is shown in red in Fig. 6.15, and is capable of explaining almost all features of the data. The model has a linear polarization fraction of 7%, a polarization leakage of 3% and 47 nm RMS wavefront error, consisting mostly of defocus (38 nm RMS), astigmatism (16 nm RMS) and coma (20 nm RMS). Significant residuals remain, which could be explained by effects not present in our simple model, such as higher-order aberrations, inaccurate sampling of the model PSF or fast-axis deviations from $\pi/2$ in the phase-shifted subapertures.
The non-zero closure phases again showcase the sensitivity of HAM to aberrations and the polarization state of the incomming light. When the polarization state and wavefront aberrations are stable in time, the closure phases should stay constant. We address the stability in the lab by imaging the PSF 90 times during 22 hours, roughly 15 minutes apart. The camera will not return to exactly the same location in the 5x5 grid for every mosaic. Shifts in these positions bias the stability measurement significantly, inflating the error bars. To be less dependent on the registration of translation between images in the mosaic, we first register the translation between all images of a single camera position. We minimize the difference between these 90 images by aligning them all to the first frame in time. After all images of a single camera position are aligned with respect to each other, we repeat the process for all other camera positions. Then, we calculate the translation of the 25 camera positions with the first frames of the aligned images. Because all images are aligned, this offset is the same for all 90 mosaics. We generate the 90 mosaics by translating the aligned images with this general offset, and stitch them together like before.
From the PSF in each mosaic we extract the closure phases of the central component and the standard deviation of each closure phase is indicated by the error bars in the bottom plot of Fig. 6.15. Almost all closure phases have a standard deviation of less than one degree, indicating that under laboratory conditions they are stable. This is also true for the underlying cause of the offsets, i.e. the polarization fraction and the wavefront aberrations. Additionally, we extract the closure phases of the holographic component, and calculate the time variability. The results are presented in Fig. 6.16, showing two interesting features. First, there is a deviation from zero for these closure phases as well, and it is independent of the number of interferograms per closure phase. These closure phases are not affected by polarization leakage, as the holographic component is imaged far away from the central component. Second, there is a difference in time variability, e.g. the largest variation is 6 degrees compared to less than a degree for the first three closure triangles. The most stable closure triangles are the ones imaged in a single interferogram, i.e. closure triangle 0, 1, 2, 10, 14 and 18 in Fig. 6.16. This suggests that the time variability does depend on the number of interferograms per closure phase. Moreover, it suggests that the cause of the closure phase offset is different than the cause of the variability differences. Most likely, the difference in variability is not caused by effects of the optical setup as this would have impacted the closure phases of the central component as well. A possible explanation for the variability difference could be that there are still remaining translation registration errors for different interferograms in time.

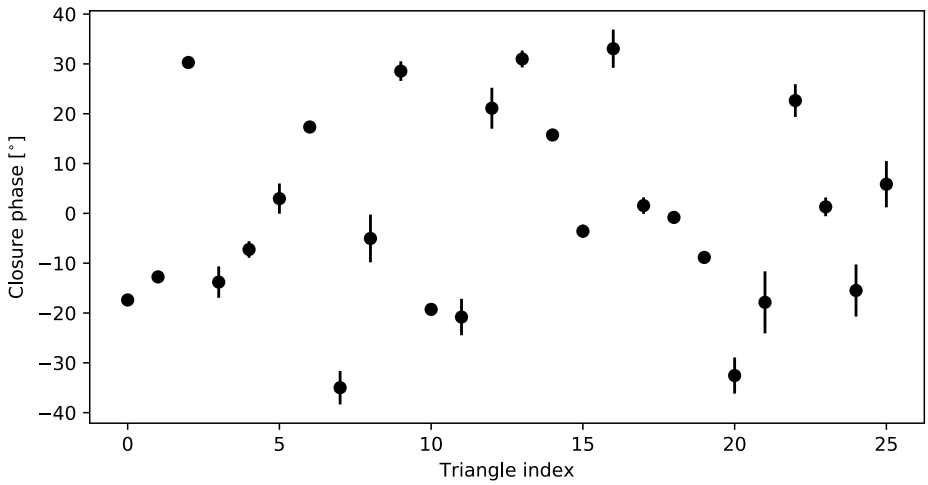


Figure 6.16: Measured average closure phases of the holographic component during 22 hours of measurements with 15 minute intervals.

On the other hand, the offsets from zero for the closure phases of the holographic component could be generated by the optical system. Light from each subaperture has a different optical path between the HAM optic and the focal plane. When two subapertures are imaged onto the same location in the focal plane, differences in optical path length introduces shifts in the fringe phase. These aberrations can differ for each baseline, such that combining the phases into a closure phase does not add to zero. Yet, as the optical setup itself is static, they are stable in time.

6.8 On-Sky Verification

We observed the binary HD 90823 (also known as HDS 1507 or WDS 10294+1211) and an unresolved reference star in two different filters. The primary goal is to verify whether the correct system parameters of HD 90823 (i.e. contrast ratio and separation) can be inferred from the data. Moreover, we aim to assess HAM's broadband performance by extracting wavelength-dependent closure phases from the holographic spots in the focal plane. In the simple case of a binary, the apparent angular separation between the two stars should not change as a function of wavelength.

6.8.1 Binary HD 90823

HD 90823 is an ideal verification target because the binary has a relatively low contrast $\Delta m \approx 1.2$ in the V and I bands, making the companion easy to detect, and, two sets of orbital elements have been published in the literature (Cvetković et al. 2016, Tokovinin 2017), allowing us to compute the predicted on-sky separation

vector at any point in time. In 2016, Cvetković et al. published a result that was based on four measurements spread over more than a decade. The authors found an orbital period of 23 years, yet they warn that their result is *"highly-tentative"* and that *"new observations are very desirable"*. Based on three new data points, Tokovinin *"radically revised"* the orbital elements in 2017 and lowered the binary's period to just over 15 years. Table 6.3 provides a further overview of the relevant HD 90823 parameters that are reported in both papers.

Table 6.3: An overview of different HD 90823 parameters as reported in Cvetković et al. (2016) and Tokovinin (2017). A and B refer to the binary components. For further details, please refer to the tables in the papers.

	Cvetković et al. (2016)	Tokovinin (2017)
Period (years)	23.361	15.59 ± 0.11
M_A ($M_\odot$)	1.66	1.53
M_B ($M_\odot$)	1.30	1.18
Contrast (Δm)	1.19 ± 0.17 (V)	1.21 ± 0.15 (I)
Spectral type	F0 (A) + F7 (B)	F2

6.8.2 Observations

The on-sky verification of the prototype HAM v1.0 test took place in the evening of 16 April 2019 (Hawaiian time). Observations were carried out in H band and time was divided equally among the binary HD 90823 and the calibrator source HD 90700. According to the SIMBAD database, the difference between the apparent magnitudes of both objects in H band is small: $m_H = 6.2$ for HD 90823 versus $m_H = 5.6$ for HD 90700. Both objects were observed in a narrowband filter (henceforth Hn5) and a broadband filter with a 20% bandwidth (henceforth Hbb). An observing log and the details of the filters are provided in Tables 6.4 and 6.5, respectively.

The OSIRIS Imager comprises a Teledyne Hawaii-2RG HgCdTe detector with a size of 2048×2048 pixels and a minimum integration time of 1.476 seconds (see Arriaga et al. 2018 for further specifications). Because this number exceeds the typical time associated with atmospheric seeing and the subapertures are larger than r_0 at the filter bandwidths, aperture masks in OSIRIS can only operate in conjunction with adaptive optics. Tests with an internal source showed that the imaging quality changed considerably as function of the input source location in the field of view. The separation of the HAM phase and amplitude mask of HAM v1.0 was considered as the source of this changing imaging quality. Light that passes through the holes of the amplitude mask at an angle to the optic axis then intercepts the phase mask off-axis, possibly illuminating the phase mask beyond the edges of the phase pattern of individual holes. If present, such a leakage term would affect the central component PSF. However, images of some sections the

[1] `https://www2.keck.hawaii.edu/inst/osiris/scale_filter.html`

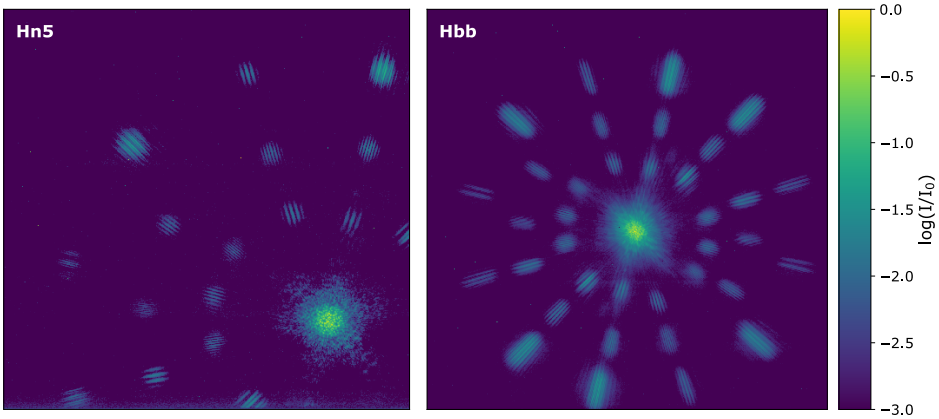


Figure 6.17: The star HD 90700 imaged with HAM in the Hn5 filter (left panel) and the Hbb filter (right panel), respectively. These figures display only a sub-window of the full sensor area. Data recovered for the Hn5 PSF occupied the bottom-right corner of the detector, so not all holographic spots are visible.

Table 6.4: A summary of the observations taken with HAM at Keck during the evening of 16 April 2019 (Hawaiian time).

Start (UTC)	End (UTC)	Target	Filter	N_f	t_{exp} (s)
06:28:52	06:36:30	90700	Hn5	50	1.476
06:39:05	06:46:42	90823	Hn5	49[a]	1.476
06:49:48	06:57:26	90823	Hbb	48[b]	1.476
06:58:56	07:06:47	90700	Hbb	50	1.476

[a] One corrupted frame was discarded.
[b] Two frames with a much lower signal-to-noise were discarded.

Table 6.5: The two filters that were used during the observations. Values were copied from the OSIRIS filter table on the Keck website [1].

Filter	PSF position	λ_0 (nm)	$\Delta\lambda/\lambda_0$ (%)
Hn5	Bottom-right	1765	4.9%
Hbb	Center	1638	20.1%

HAM mask pupil did not show partial illumination of holes. As the whole optic could not be imaged in pupil viewing mode, and we can not be certain that all holes were fully illuminated. Another explanation could be a differential focus between the OSIRIS imager and the OSIRIS spectrograph, which was present at the time of observation. Reconstruction of the visibility amplitudes also showed a gradient in the pupil illumination. The OSIRIS imager has been realigned and HAM v1.5 has been installed after the presented observations. New internal source

measurements with the HAM v1.5 mask show little variation of the PSF quality as function position in the FOV.
Because the image quality of the central component looked highest in the bottom-right corner of the detector, we decided to locate it there during the Hn5 observations (sacrificing the majority of holographic spots – left panel of Figure 6.17). However, with the holographic spots being of primary interest in broadband, the PSF was shifted to the middle of the detector for the Hbb observations (right panel of Figure 6.17). This allowed us to compute HAM's full set of closure phases, albeit with lower image quality.

6.8.3 Data Reduction

We apply a dark correction using 100 dark frames and calculate the power spectrum of the holographic component and the central component separately on a *uv*-grid of 3000×3000 pixels. For the Hn5 filter we mask everything but the central component, while for the Hbb filter we apply the same data reduction as mentioned in Sect. 6.5. Before the fringe phases can be extracted, we need to determine exactly *where* in the *uv*-plane the visibility has to be sampled. This requires two pieces of information: (i) the rotation angle of the PSF, which is set by the mask's orientation with respect to the detector, and (ii) the radial scaling of the PSF, which depends on the wavelength and the magnification of the instrument. To find the optimal parameter values that describe the scaling and orientation of the power spectrum (and thus of the PSF), we use a model of the *uv*-plane, that is cross-correlated with the observed power spectra as a function of radial scaling and rotation angle. We find that the angular orientations of power spectra of the holographic and central components differ by $\sim 2.1°$. This suggests that there is an angular offset between the amplitude mask and the phase mask in both filter wheels. The orientation of the central component's power spectrum is determined by the holes in the amplitude mask, while the orientation of the holographic component's power spectrum is determined by the slopes on the phase mask, which has to be taken into account when fitting a model to the observed closure phases.
We extract closure phases from the observed PSFs using the Fourier methods for both monochromatic and broadband observations.

Closure Phases

Once all visibility phases are sampled in the *uv*-plane, we compute the closure phases associated with each triangle on the mask. This gives rise to 197 sets of closure phases, each of which corresponds to one of the science frames listed in Table 6.4. In addition, we average the closure phases of the calibrator HD 90700 per filter, and we consider the standard deviation as measure of the stability of a closure triangle over time. We then calibrate the binary closure phases of each frame by subtracting the average closure phases of the calibrator.
Some representative results are displayed in Figures 6.18 and 6.19. Fig. 6.18 shows

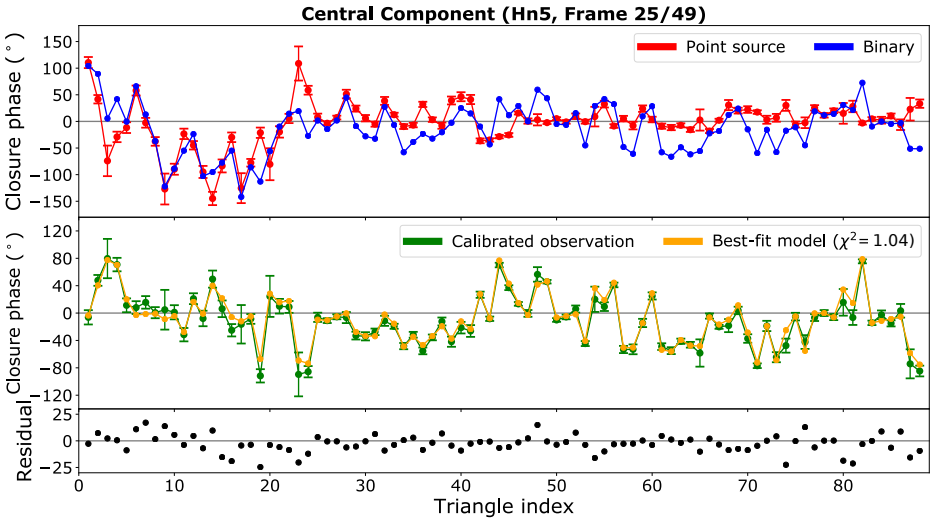


Figure 6.18: *Top:* closure phases of HD 90823 (binary) inferred from the central component in frame 25/49 in the Hn5 filter. The closure phases of HD 90700 (point source) are averaged over all frames. *Middle:* the calibrated observation, obtained by subtracting the closure phases of the point source from the closure phases of the binary. The best-fit model that was found is overplotted, with the corresponding χ^2 value reported in the legend. *Bottom:* residuals left after subtracting the model from the observation.

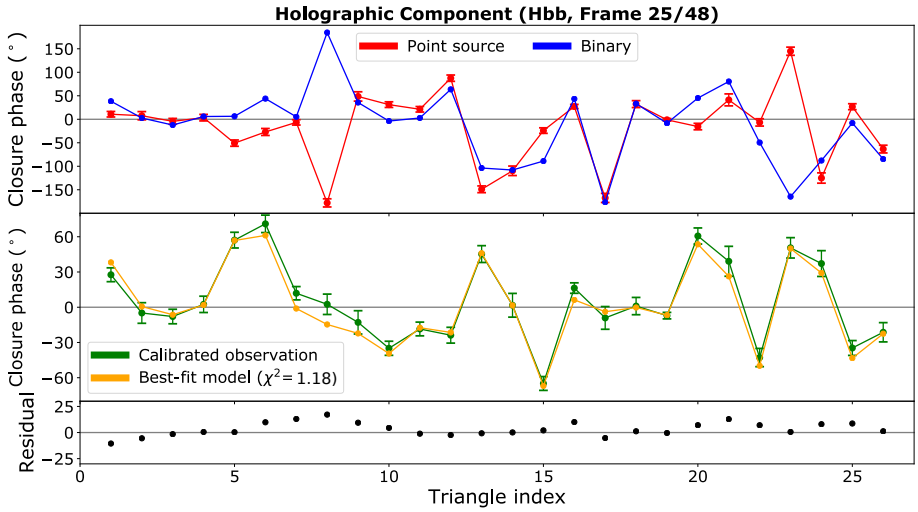


Figure 6.19: The same as Figure 6.18, but now for the **Holographic spots** in frame 25/48 in the Hbb filter. The uv-plane was sampled at 1638 nm.

the both the non-calibrated and calibrated closure phases of HD 90823 obtained from the PSF core in frame 25/49 in the Hn5 filter, as well as the best-fit model that was found. The average closure phases of the point source are plotted in the background. From Fig. 6.18 it is clear that the closure phases of HD 90700 (point source) exhibit a remarkably strong deviation from zero. The deviations are much stronger for HAM v1.0 on-sky than the offsets measured in-lab for HAM v1.5, as presented in the previous section. It is unclear if this is purely from the differences between HAM v1.0 and HAM v1.5. However, the overall structure of the closure phases agree with the lab measurements, with the first 24 closure phases (with triangle indices ≤ 24) of the central mask component having the largest deviations. Moreover, Fig. 6.18 suggests that the errors on these 24 triangles are also greater than the other errors, as they are more sensitive to changes in defocus. Even the PSF at the corner of the detector in the Hn5 filter has a poor quality, indicating that the PSF is strongly aberrated. Even with small fractions of linearly polarized light present, it is expected that the closure phase offsets are much larger under these conditions. This implies that calibration is essential to extract physical information from the data.

Fig. 6.19 shows the measured closure phases for frame 25/48 in the Hbb filter for the holographic spots at the central wavelength λ_0. Again, the closure phases for the single star deviate strongly from zero, and again much more than the lab measurements. We did not find a good explanation for this offset, but we can calibrate it for the binary system using the calibrator.

Table 6.6: The estimated separation and contrast ratio for HD 90823. A different set of closure phases applies to each row. The closure phases in the broadband Hbb filter were sampled at the central wavelength λ_0.

Subset	N_f	Separation ρ (mas)	Contrast ratio r
Central (Hn5)	38	121.9 ± 0.5	0.45 ± 0.01
Central (Hbb)	47	121.1 ± 0.8	0.45 ± 0.02
Holographic(Hbb)	46	120.9 ± 0.5	0.44 ± 0.02
Full HAM (Hbb)	93	121.0 ± 0.7	0.44 ± 0.02

Parameter Estimation

Given the measured closure phases and calibration of the zero points, the separation ($\boldsymbol{\rho} = (\rho_x, \rho_y)$) and the contrast ratio of HD 90823 can be estimated. We do this by finding the parameter combination $(r, \boldsymbol{\rho})$ that minimizes the chi-squared difference χ^2 between the observation and an analytical model of a binary system:

$$\chi^2(r, \boldsymbol{\rho}) = \frac{1}{N-m} \sum_{i=1}^{N} \left(\frac{[\phi_{i,\mathrm{obs}} - \phi_{i,\mathrm{mod}}(r, \boldsymbol{\rho}) + \pi]\%(2\pi) - \pi}{\sigma_i} \right)^2 . \tag{6.11}$$

Here, N is the number of closure phases, m the number of free parameters, $\phi_{i,\mathrm{obs}}$ is the i-th observed closure phase (in radians, with error σ_i) and $\phi_{i,\mathrm{mod}}$ is the i-th modelled closure phase (in radians). Adding π, applying the modulo operator % and subtracting π makes sure that all differences are mapped onto the interval $[-\pi, +\pi]$. We perform a total of 145 fits: 49 in the Hn5 filter and 2×48 in the Hbb filter, whereby the closure phases obtained from the PSF core and the holographic spots are treated separately. Table 6.6 provides an overview of the HD 90823 parameters that were found after evaluating Equation 6.11 on a high-resolution grid. The reported values for $\rho = \sqrt{\rho_x^2 + \rho_y^2}$ and r result from averaging the best-fit parameters over different subsets of frames. Only fits for which $\chi^2 < 2N_{CP}$ (with N_{CP} the number of closure phases) are included to reduce the effect of outliers. We find the same contrast ratio $r = 0.45 \pm 0.02$ for each of the different subsets. The value inferred for the separation ρ is ca. 1 mas larger in the Hn5 filter as compared to the Hbb filter, which can be explained by random errors only.
As far as the performance of the mask's holographic component is concerned, it is reassuring that the retrieved parameter values in the Hbb filter are consistent with each other. The observables obtained from the central and holographic components of the mask are within 1σ.

6.8.4 Spectroscopic parameter retrieval

Holographic aperture masking has the unique capability to extract low-resolution spectroscopic closure phases using the holographic component. Here, we will look

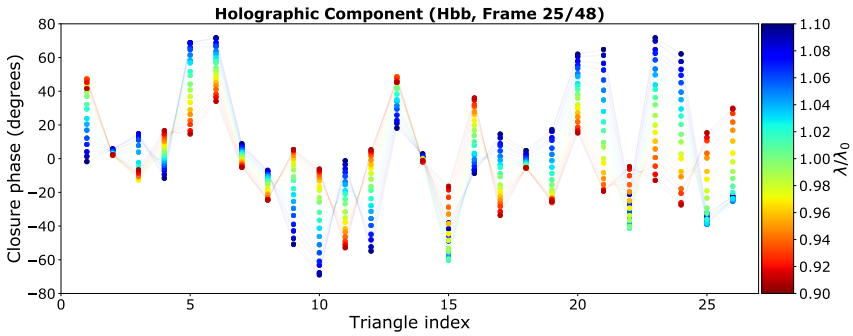


Figure 6.20: Calibrated closure phases obtained from the **Holographic spots** in frame 25/48 in the Hbb filter, plotted for different wavelengths within the 20% bandwidth (see colour bar).

at the multi-wavelength extraction of the closure phases, which allows us to extract the separation and magnitude from multiple wavelength channels in the Hbb band, spanning from 1473 nm to 1803 nm. We illustrate the wavelength-dependence of the measured closure phases in the Hbb filter in Fig. 6.20. The closure phases from the holographic spots are sampled at different coordinates in the uv-plane (corresponding to different wavelengths). The number of sampling points far exceeds the number of *independent* measurements. The latter, the spectral resolution, will be 2 or 3, depending on the triangle index, according to Eq. 6.2 and Eq. 6.3. The closure phases behave well and there seems to be a smooth transition from one wavelength to another.

We extract the binary parameter values by minimizing the χ^2 (Equation 6.11) for closure phases sampled at different points in the uv-plane. Fig. 6.21 and Fig. 6.22 display the parameters of HD 90823 that were inferred from the holographic spots in the Hbb filter as a function of wavelength. Fig. 6.21 shows the separation ρ between the binary components. Expressed in units of λ/D, the separation exhibits a $1/\lambda$ drop-off, which implies that ρ must be constant throughout the bandwidth. This is the expected result for a binary system, and is a powerful method to distinguish astronomical observables from instrument artefacts. As mentioned in Table 6.6, we find a separation of $121-122$ mas based on the closure phases at λ_0. This is also the value that follows from averaging over all wavelengths in the bandwidth, as shown by the horizontal line. Some points deviate significantly from the average. The error bars are determined using a Jackknife method (Roff & Preziosi, 1994), which does not take into account systematic errors, e.g. wavelength dependent errors in closure phase retrieval as shown in Fig. 6.9. The measured brightness ratio r as a function of wavelength is plotted in the right panel of Fig. 6.22.

The spectral types of the HD 90823 components are uncertain. According to the SIMBAD database, both stars are of type F2, but the corresponding quality labels suggest little reliability. On the other hand, Cvetković et al. (2016) state that the

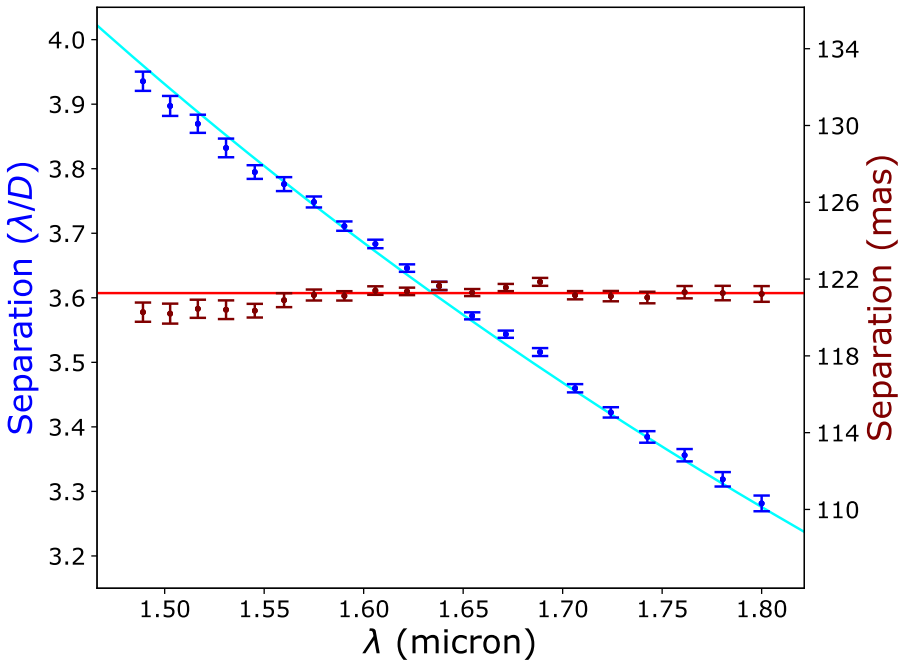


Figure 6.21: The separation between the components of HD 90823 inferred with HAM as a function of wavelength, expressed in units of λ/D (blue points) and expressed in milli-arcseconds (red points). The blue line is proportional to $1/\lambda$. The red line is a weighted average of the measured separations in mas.

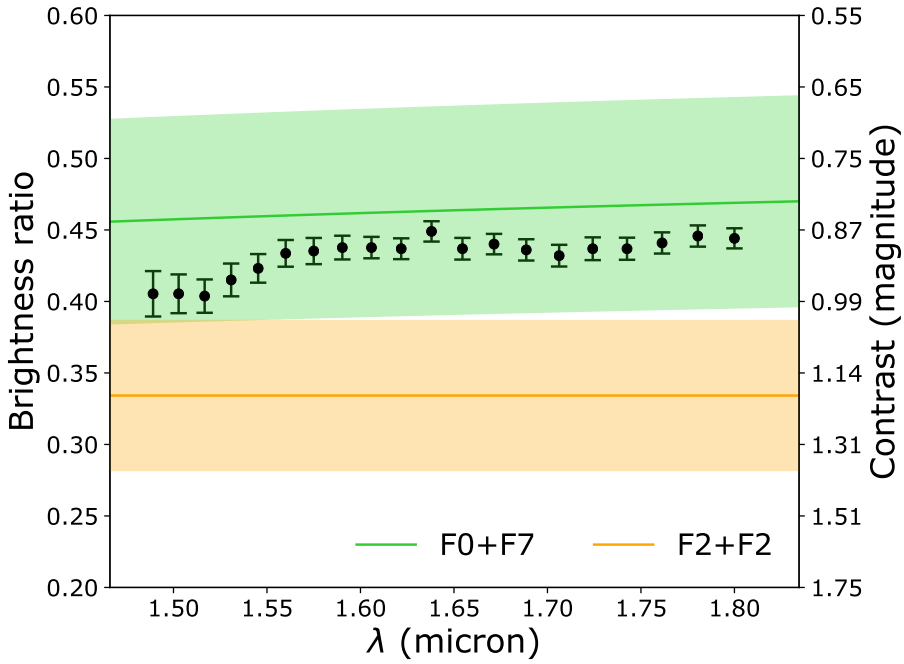


Figure 6.22: The recovered brightness ratio of the binary as a function of wavelength. The green line shows the expected brightness ratio for the spectral types (F0 and F7) reported by Cvetković et al. (2016). The orange line is the expected brightness ratio when both components are of type F2. The envelopes represent the 1σ error from the V-band and I-band measurements presented in table 6.3. The number of data points is much larger than the number of independent measurements.

bright component is hotter (type F0) than the faint one (type F7) with the difference in surface temperatures being roughly 1000 K. In order to find out which claim is most likely based on the HAM data, we computed the Planck spectra of the components in both scenarios, scaled them according to the reported contrast in V band (see Table 6.3) and divided them in the Hbb filter. Fig. 6.22 illustrates that the actual measurements lie in between the theoretical curves, but fully within the 1σ envelope of the F0+F7 scenario. Moreover, the gradient as function of wavelength is consistent with the F0+F7 scenario. This means that our estimate of the brightness ratio also suggests a difference between the components' temperatures and spectral types, in line with Cvetković et al. (2016).

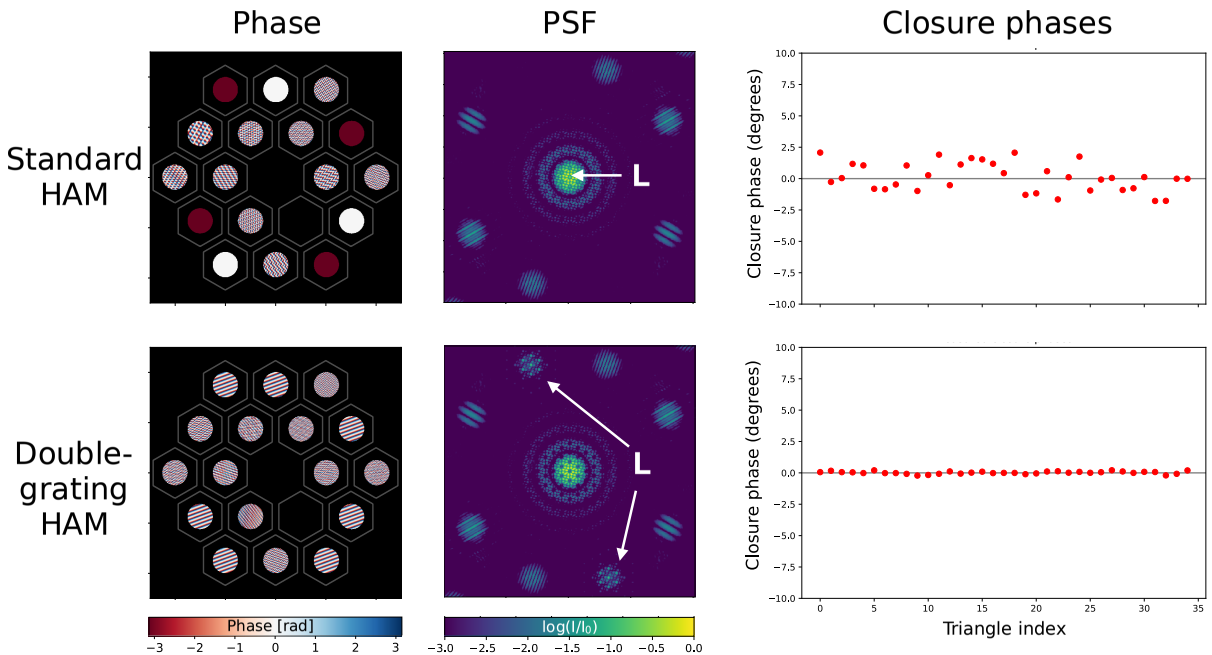


Figure 6.23: The difference between a standard HAM and a double-grating HAM. The first element of the double-grating is the standard HAM pattern with an added polarization grating pattern (= phase ramp) with 70 periods across the full pupil. The second grating is not shown. The resulting PSFs show that the polarization leakage is directed from the center into two off-axis PSFs. Both elements have 2% leakage and we assume a linear polarization fraction of 10%. Closure phase offsets due to polarization leakage are greatly reduced with the double-grating method.

6.9 Conclusions and Outlook

In the paper we show that holographic aperture mask show promise to empower a new generation of sparse aperture masking experiments: retaining the instrumental simplicity while adding spectroscopic capabilities, higher throughput and added Fourier coverage. We discuss the trade-offs in mask design, and how this was implemented for a prototype HAM mask in OSIRIS. A first version of this mask was installed at Keck OSIRIS in 2018. By observing the binary HD90823 with this mask, we obtain low resolution spectra of closure phases, confirming the broadband capabilities of the HAM mask. We investigated two limitations of this version. The first is the spatial separation between the phase and amplitude mask, resulting in spatially varying PSF quality. This is solved with an upgraded version, HAM v1.5, with a diced version of the same phase mask combined in the same holder as the amplitude mask, installed in OSIRIS in early 2020. Secondly, we find non-zero closure phases for a single star. We show in simulation that polarization leakage can produce these non-zero offsets in closure phases, and confirm this with lab demonstrations.

A future upgrade of the HAM mask can remove the effects of polarization leakage by several orders of magnitude suppression of the unwanted light using the double-grating method (Doelman et al., 2020). The double-grating method adds a phase ramp to the phase pattern, so that any polarization leakage travels in a different direction than the main beams. A second grating with the same phase ramp (= polarization grating) is installed directly after the first phase pattern, folding the two main beams back on axis. The polarization leakage term of the first grating, on the other hand, is diffracted away by this second grating, reducing the on-axis leakage by at least an order of magnitude. The location of the polarization leakage can be controlled with the phase ramp slope and direction, similar to the holographic interferograms. Placing the polarization leakage on an empty spot on the detector reduces the phase slope. Adapting the focal plane design of a HAM device to leave room for this leakage term would be beneficial. We simulate the performance of a HAM device with the second design presented in Fig. 6.7, assuming 2% leakage for both GPHs and a linear polarization fraction of 10%. The results are shown in Fig. 6.23. Using the double-grating method reduces the standard deviation of the closure phases from 1.1 degree to 0.1 degree. The residual closure phase pattern of the double grating HAM is not correlated with the standard HAM, which suggests that the deviation from zero is caused by different effect, e.g. inaccuracies in the data reduction. These simulations prove that a double-grating version of HAM would greatly reduce the impact of polarization leakage on the performance of HAM.

The potential of a double-grating HAM is exciting.

We outline a few scientific prospects that are enabled only by HAM. Studies of thermal emission from protoplanets in protoplanetary disks is complicated by disk features that can emulate exoplanet signals, e.g. light scattered by dust, that these systems can display (Kraus & Ireland, 2012; Sallum et al., 2015; Currie et al., 2019). However, spectral information can help with discriminating between disk

and planet signal (Haffert et al., 2019). HAM's simultaneous spectral and spatial measurements help constrain infrared spectral slopes, distinguishing between scattered light and thermal signals. Other opportunities are enabled by the higher throughput of a HAM mask compared to a SAM mask. Follow-up on the brightest *TESS* targets can help with ruling out background or binary contaminators. Monitoring brown dwarf binaries with HAM increases efficiency on determining orbits and dynamical masses, directly testing predictions for lithium burning, the stellar/substellar mass boundary, and substellar cooling rates (e.g., Dupuy & Liu (2017)). Suppressing the polarization leakage will be critical for improving closure phase stability, resulting in better contrast.

Acknowledgements

The research of David Doelman and Frans Snik leading to these results has received funding from the European Research Council under ERC Starting Grant agreement 678194 (FALCONER). Laser cutting of aperture masks and planar optics was carried out with the assistance of the OptoFab node of the Australian National Fabrication Facility, utilising NCRIS and NSW Gov. funding. We specifically thank Benjamin Johnston from OptoFab for the quick turn-around of the many amplitude masks and for laser cutting the HAM v1.5 mask. Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. We thank John Canfield and Peter Wizinowich for the technical support. We also thank Michael Liu for the fruitful discussions, which helped improve the results presented in this work.

6.10 Appendix: Spectral resolution of the holographic interferograms

Here we will derive the spectral resolution of hologrpahic interferograms. We start by looking at a single baseline, $\mathbf{b}$, between two subapertures with a phase ramp with a period P and a direction $\hat{\mathbf{a}}$. Assuming no piston phase offset between the subapertures, the electric field is given by

$$M(\mathbf{r}) = \Pi(\mathbf{r}) \otimes \left[(\delta(\mathbf{r} - \mathbf{b}/2) + \delta(\mathbf{r} + \mathbf{b}/2))\, \mathrm{e}^{2\pi i \mathbf{a}\mathbf{r}} \right], \tag{6.12}$$

where $\otimes$ is the convolution operator and $\delta(\mathbf{r})$ is the Dirac delta function, $\mathbf{a} = P/D_{sub}\hat{\mathbf{a}}$, and $\Pi(\mathbf{r})$ defines the subaperture. For a circular subaperture with di-

ameter D_{sub},

$$\Pi(\mathbf{x}) = \begin{cases} 1 & \text{if } |\mathbf{x}| \leq \frac{D_{sub}}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{6.13}$$

The PSF is then described by

$$p(\theta) = P(\boldsymbol{\theta}, \lambda) \left[2 + 2\cos\left(\frac{2\pi \mathbf{b}\boldsymbol{\theta}}{\lambda} \right) \right] \otimes \delta(\boldsymbol{\theta} - \lambda \mathbf{a}). \tag{6.14}$$

Here $\boldsymbol{\theta}$ and $\frac{\mathbf{r}}{\lambda}$ are the Fourier plane coordinates, and $P(\boldsymbol{\theta}, \lambda)$ is the PSF of the aperture function

$$P(\boldsymbol{\theta}, \lambda) = \text{Airy}\left(\frac{\pi D_{sub} |\boldsymbol{\theta}|}{\lambda} \right) \tag{6.15}$$

As the PSF location is directly proportional to $\lambda\mathbf{a}$ due to the grating, we have an independent measurement of a baseline phase when the shift is $1.22\lambda/D_{sub}$, i.e. the Raleigh criterion. Therefore we can define the spectral resolution as

$$R_{fp} = \frac{\lambda}{\Delta\lambda} = \frac{D_{sub}}{1.22P}. \tag{6.16}$$

Increasing the subaperture size and grating frequency yields a higher spectral resolution. For a subaperture with $D_{sub} = \frac{1}{10}D$ imaged at $100\lambda/D$, the period is $D_{sub}/10$ and $R \sim 8$.
The second spectral resolution of the holographic component is defined in the uv-plane. We calculate the uv-plane distribution $\hat{V}(\mathbf{f})$ with the Fourier transform of the PSF,

$$\hat{V}(\mathbf{f}) = (\Pi(\mathbf{r}) \star \Pi(\mathbf{r}))^{\frac{1}{2}} \otimes [\delta(\mathbf{f} - \mathbf{b}/\lambda) + \delta(\mathbf{f} + 2\delta(\mathbf{f}) + \mathbf{b}/\lambda)]\, \mathrm{e}^{2\pi i \mathbf{af}}. \tag{6.17}$$

We define the cross-correlation between the subaperture function $Pi(\mathbf{r})$, shifted by $\mathbf{b}/\lambda$ as a splodge. The location of the splodges change with wavelength and the shift depends on the length of the baseline. The phases of two splogdes of different wavelengths for the same baseline can be uniquely extracted when they are separated by D_{sub}. In principle this is possible for every SAM mask, however, to increase throughput many of these masks have large subaperture diameters and many baselines, such that even for a small bandwidth the splodges start to overlap. In the case of HAM, the holographic interferograms only contain a limited amount of baselines and the effect of overlapping splodges can be reduced by design. The spectral resolution in the uv-plane is given by

$$R_{uv} = \frac{\lambda}{\Delta\lambda} = \frac{|\mathbf{b}|}{D_{sub}}. \tag{6.18}$$

Bibliography

Arriaga, P., Fitzgerald, M. P., Johnson, C., Weiss, J. L., & Lyke, J. E. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, ed. H. Takami, C. J. Evans, & L. Simard, Vol. 10702 (SPIE), 103, doi: 10.1117/12.2313101

Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155, doi: 10.1051/0004-6361/201935251

Carlotti, A., & Groff, T. 2010, in Ground-based and Airborne Telescopes III, Vol. 7733, International Society for Optics and Photonics, 77333D

Cheetham, A. C., Girard, J., Lacour, S., et al. 2016, in Optical and Infrared Interferometry and Imaging V, Vol. 9907, International Society for Optics and Photonics, 99072T

Chilcote, J. K., Bailey, V. P., De Rosa, R., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, International Society for Optics and Photonics, 1070244

Currie, T., Marois, C., Cieza, L., et al. 2019, The Astrophysical Journal Letters, 877, L3

Cvetković, Z., Pavlović, R., & Ninković, S. 2016, The Astronomical Journal, 151, 83, doi: 10.3847/0004-6256/151/3/83

Cvetojevic, N., Huby, E., Martin, G., et al. 2018, in Optical and Infrared Interferometry and Imaging VI, ed. M. J. Creech-Eakman, P. G. Tuthill, & A. MÃ©rand, Vol. 10701, International Society for Optics and Photonics (SPIE), doi: 10.1117/12.2313262

Doelman, D. S., Por, E. H., Ruane, G., Escuti, M. J., & Snik, F. 2020, Publications of the Astronomical Society of the Pacific, 132, 045002

Doelman, D. S., Tuthill, P., Norris, B., et al. 2018, in Optical and Infrared Interferometry and Imaging VI, ed. A. Mérand, M. J. Creech-Eakman, & P. G. Tuthill, Vol. 10701 (SPIE), 12, doi: 10.1117/12.2313547

Dupuy, T. J., & Liu, M. C. 2017, The Astrophysical Journal Supplement Series, 231, 15

Escuti, M. J., Kim, J., & Kudenov, M. W. 2016, Optics and Photonics News, 27, 22, doi: 10.1364/OPN.27.2.000022

Foo, G., Palacios, D. M., & Swartzlander, G. A. 2005, Opt. Lett., 30, 3308, doi: 10.1364/OL.30.003308

Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Science Reviews, 123, 485

Gauchet, L., Lacour, S., Lagrange, A.-M., et al. 2016, Astronomy & Astrophysics, 595, A31, doi: 10.1051/0004-6361/201526404

Gilmozzi, R., & Spyromilio, J. 2007, The Messenger, 127, 11

Greenbaum, A. Z., Pueyo, L., Sivaramakrishnan, A., & Lacour, S. 2014, The Astrophysical Journal, 798, 68, doi: 10.1088/0004-637X/798/2/68

Haffert, S., Bohn, A., de Boer, J., et al. 2019, Nature Astronomy, 3, 749

Haniff, C. A., Mackay, C. D., Titterington, D. J., et al. 1987, Nature, 328, 694, doi: 10.1038/328694a0

Haubois, X., Norris, B., Tuthill, P. G., et al. 2019, A&A, 628, A101, doi: 10.1051/0004-6361/201833258

Hecht, E., & Zajac, A. 1974, Reading, Mass, 19872, 350

Huby, E., Perrin, G., Marchis, F., et al. 2012, A&A, 541, A55, doi: 10.1051/0004-6361/201118517

Huby, E., Duchêne, G., Marchis, F., et al. 2013, A&A, 560, A113, doi: 10.1051/0004-6361/201321894

Jovanovic, N., Tuthill, P. G., Norris, B., et al. 2012, Monthly Notices of the Royal Astronomical Society, 427, 806, doi: 10.1111/j.1365-2966.2012.21997.x

Komanduri, R. K., Lawler, K. F., & Escuti, M. J. 2013, Optics Express, 21, 404

Kraus, A. L., & Ireland, M. J. 2012, The Astrophysical Journal, 745, 5, doi: 10.1088/0004-637X/745/1/5

Kraus, A. L., Ireland, M. J., Martinache, F., & Lloyd, J. P. 2008, ApJ, 679, 762, doi: 10.1086/587435

Lacour, S., Tuthill, P., Amico, P., et al. 2011, Astronomy & Astrophysics, 532, A72, doi: 10.1051/0004-6361/201116712

Larkin, J., Barczys, M., Krabbe, A., et al. 2006, New Astronomy Reviews, 50, 362

Macintosh, B., Nielsen, E., & De Rosa, R. 2019, AAS, 233, 104

Martinache, F. 2010, The Astrophysical Journal, 724, 464, doi: 10.1088/0004-637X/724/1/464

Mawet, D., Serabyn, E., Liewer, K., et al. 2009, Optics Express, 17, 1902

Miskiewicz, M. N., & Escuti, M. J. 2014, Optics Express, 22, 12691, doi: 10.1364/OE.22.012691

Monnier, J., Tuthill, P. G., Ireland, M., et al. 2009, The Astrophysical Journal, 700, 491

Nielsen, E. L., De Rosa, R. J., Macintosh, B., et al. 2019, The Astronomical Journal, 158, 13

Norris, B. R. M., Cvetojevic, N., Lagadec, T., et al. 2019, Monthly Notices of the Royal Astronomical Society, 491, 4180, doi: 10.1093/mnras/stz3277

Perrin, G., Lacour, S., Woillez, J., & Thiébaut, E. 2006, Monthly Notices of the Royal Astronomical Society, 373, 747, doi: 10.1111/j.1365-2966.2006.11063.x

Pope, B., Tuthill, P., Hinkley, S., et al. 2015, Monthly Notices of the Royal Astronomical Society, 455, 1647, doi: 10.1093/mnras/stv2442

Por, E. H., Haffert, S. Y., Radhakrishnan, V. M., et al. 2018, in Adaptive Optics Systems VI, ed. D. Schmidt, L. Schreiber, & L. M. Close, Vol. 10703 (SPIE), 152, doi: 10.1117/12.2314407

Roff, D. A., & Preziosi, R. 1994, Heredity, 73, 544

Ruane, G., Mawet, D., Riggs, A. E., & Serabyn, E. 2019, in Techniques and Instrumentation for Detection of Exoplanets IX, Vol. 11117, International Society for Optics and Photonics, 111171F

Sallum, S., & Eisner, J. 2017, The Astrophysical Journal Supplement Series, 233, 9, doi: 10.3847/1538-4365/aa90bb

Sallum, S., & Skemer, A. 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 1, doi: 10.1117/1.JATIS.5.1.018001

Sallum, S., Follette, K. B., Eisner, J. A., et al. 2015, Nature, 527, 342, doi: 10.1038/nature15761

Samland, M., Mollière, P., Bonnefoy, M., et al. 2017, A&A, 603, A57, doi: 10.1051/0004-6361/201629767

Sanders, G. H. 2013, Journal of Astrophysics and Astronomy, 34, 81

Soummer, R., Pueyo, L., Sivaramakrishnan, A., & Vanderbei, R. J. 2007, Opt. Express, 15, 15935, doi: 10.1364/OE.15.015935

Tokovinin, A. 2017, The Astronomical Journal, 154, 110, doi: 10.3847/1538-3881/aa8459

Tuthill, P., Monnier, J., Danchi, W., Wishnow, E., & Haniff, C. 2000, Publications of the Astronomical Society of the Pacific, 112, 555

Tuthill, P., Lloyd, J., Ireland, M., et al. 2006, in Advances in Adaptive Optics II, ed. B. L. Ellerbroek & D. B. Calia, Vol. 6272, International Society for Optics and Photonics (SPIE), 1064 – 1073, doi: 10.1117/12.672342

Tuthill, P. G. 2012, in Optical and Infrared Interferometry III, ed. F. Delplancke, J. K. Rajagopal, & F. Malbet, Vol. 8445 (International Society for Optics and Photonics), 844502, doi: 10.1117/12.925794

Tuthill, P. G. 2018, in Optical and Infrared Interferometry and Imaging VI, ed. A. Mérand, M. J. Creech-Eakman, & P. G. Tuthill (SPIE), 11, doi: 10.1117/12.2313145

Tuthill, P. G., Monnier, J. D., & Danchi, W. C. 1999, Nature, 398, 487, doi: 10.1038/19033

van Holstein, R. G., Girard, J. H., de Boer, J., et al. 2020, A&A, 633, A64, doi: 10.1051/0004-6361/201834996

Vigan, A., Gry, C., Salter, G., et al. 2015, Monthly Notices of the Royal Astronomical Society, 454, 129, doi: 10.1093/mnras/stv1928

Willson, M., Kraus, S., Kluska, J., et al. 2019, A&A, 621, A7, doi: 10.1051/0004-6361/201630215

7 Simultaneous phase and amplitude aberration sensing with a liquid-crystal vector-Zernike phase mask

Adapted from
D. S. Doelman, F. Fagginger Auer, M. J. Escuti, F. Snik
Published in Optics Letters (2018)

We present an enhanced version of the Zernike wavefront sensor, that simultaneously measures phase and amplitude aberrations. The "vector-Zernike" wavefront sensor consists of a patterned liquid-crystal mask, which imposes $\pm\pi/2$ phase on the point spread function core through the achromatic geometric phase acting with opposite sign on opposite circular polarizations. After splitting circular polarization, the ensuing pupil intensity images are used to reconstruct the phase and the amplitude of the incoming wavefront. We demonstrate reconstruction of the complex wavefront with monochromatic lab measurements and show in simulation the high accuracy and sensitivity over a bandwidth up to 100%.

7.1 Introduction

The introduction of phase-contrast microscopy by Frits Zernike (Zernike, 1934, 1942a,b; Goodman, 1996) was a revolution for the field of biological sciences.While a standard microscope measures intensity variations due to absorption, a phase-contrast microscope measures phase differences from variations in index of refraction by coupling the phase differences into intensity variations with a phase mask inside the microscope. Otherwise unseen transparent structures could now be imaged in great detail. Another application of the phase-contrast method is the Zernike wavefront sensor (ZWFS). The ZWFS has a similar phase mask with a dot that offsets the phase of the core of the point spread function (PSF) by $\pi/2$. The optimal size of the dot is 1.06 λ/D, where D is the pupil diameter and λ the wavelength. The phase-shifted PSF core interferes with the rest of the PSF and phase aberrations are converted to electric field amplitude variations in the subsequent pupil plane (Goodman, 1996; N'Diaye et al., 2013; Guyon, 2005). Thus, a phase-only aberration can be measured directly with one intensity measurement. Extremely sensitive systems like direct-imaging instruments for exoplanet detection use the ZWFS to minimize instrumental aberrations. Non-common path correction with the ZWFS has been successfully demonstrated on-sky (Vigan et al., 2018) and picometer precision has been achieved in the lab for the Wide Field Infrared Survey Telescope (WFIRST) (Shi et al., 2018).
One major advantage of the Zernike wavefront sensor is that it is the most photon-efficient WFS with a sensitivity factor for photon noise of unity (Guyon, 2005; Paterson, 2008). Also, the ZWFS constitutes a simple implementation for any optical system that only requires a focal-plane mask and the ability to image the pupil plane. However, the dynamic range is limited, and the mask applies (wavelength-dependent) scalar phase shifts. In addition, a reference pupil image is needed for static amplitude aberration correction. Therefore, time-variable amplitude aberrations are reconstructed as phase aberrations. Static and dynamic amplitude variations can come from contamination of optics, degradation in optical performance, Fresnel propagation effects, and scintillation. Several systems that measure both amplitude and phase have been proposed, however these systems are complicated (Yamamoto et al., 2015; Cuche et al., 1999) or require diversity in time, like the phase-shifting ZWFS, where a variable-phase Zernike mask is used to reconstruct the complex wavefront with four consecutive measurements (Wallace et al., 2011). In this paper we present a liquid-crystal version of the ZWFS, the "vector-Zernike" wavefront sensor (vZWFS) that enables the simultaneous measurement of both phase and amplitude aberrations. This simple upgrade only requires the replacement of the Zernike mask in addition to splitting circular polarization. The property of optimal photon efficiency is maintained by the vZWFS and a liquid-crystal focal plane mask applies achromatic geometric phase, yielding an improved broadband performance (Bloemhof, 2014).

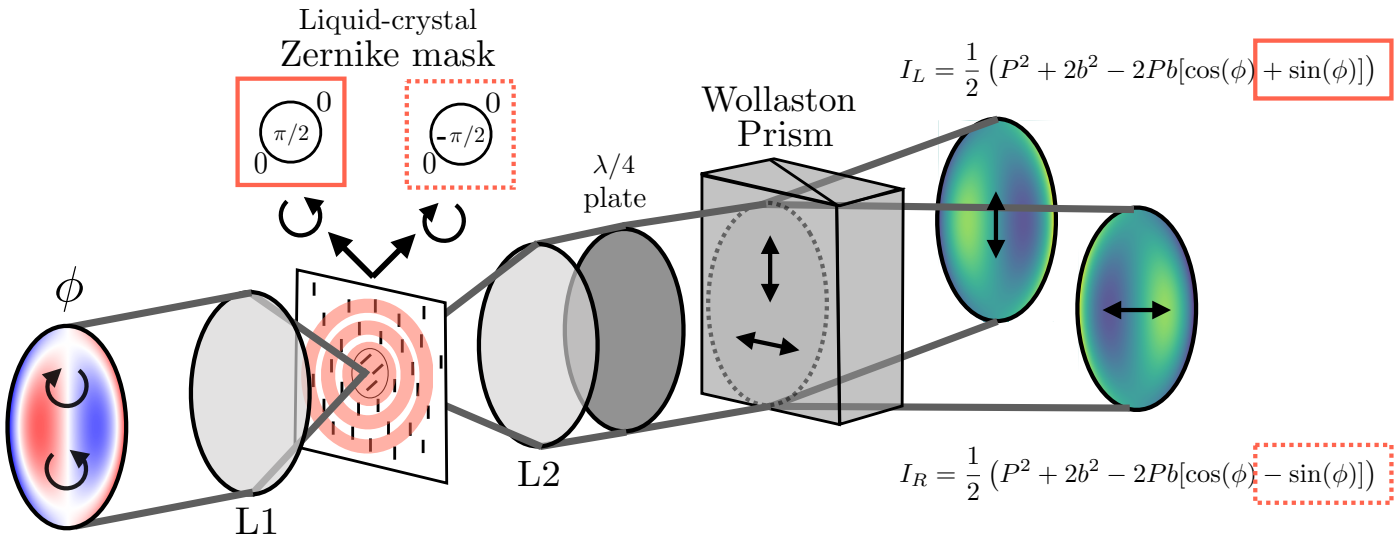


Figure 7.1: Layout of the vector Zernike wavefront sensor. The intensity distribution of the pupils depends on the incoming phase and sign of the polarization state.

7.2 Theory

The geometric phase applied by the mask, $\Delta\theta$, is independent of wavelength and only depends on the fast-axis orientation, α, of the mask and the handedness of the circular polarization state of the incoming light (Escuti et al., 2016)

$$\Delta\theta(x, y) = \pm 2\alpha(x, y). \tag{7.1}$$

Writing complex fast-axis orientation patterns is enabled by liquid-crystal direct-write technology (Miskiewicz & Escuti, 2014) and achromatizing the half-wave retardance is enabled by stacking self-aligning liquid-crystal layers to form a multi-twist retarder (Komanduri et al., 2013). The precise writing capability and broad-band performance of this technology has been verified in the lab for other optical elements (Oh & Escuti, 2009; Escuti et al., 2016). A Zernike mask manufactured with liquid-crystal technology simultaneously applies $+\pi/2$ phase to the right-handed circularly polarized fraction of the PSF core and $-\pi/2$ to the left-handed circularly polarized fraction of the PSF core. We split the light with opposite handedness with a quarter-wave plate and a polarizing beam-splitter as shown in Fig. 7.1.
To describe the resulting intensity in the two pupils, we adopt the formalism from N'Diaye et al. (N'Diaye et al., 2013). The electric field at the entrance pupil, Ψ_A, is defined as

$$\Psi_A = Pe^{i\phi} = P_0(1 - \epsilon)e^{i\phi}. \tag{7.2}$$

Here P is the normalized amplitude such that P_0 is the pupil function. The variables $\epsilon = \epsilon(u, v)$ and $\phi = \phi(u, v)$ are amplitude aberration and the phase aberration for each position in the pupil plane. Following the derivation of N'Diaye et al. (N'Diaye et al., 2013) and assuming $\theta = \pm\pi/2$, the intensity of the two pupils (I_L and I_R) can be written as

$$I_L = \frac{1}{2}\left(P^2 + 2b^2 - 2Pb[\cos(\phi) + \sin(\phi)]\right) \tag{7.3}$$

$$I_R = \frac{1}{2}\left(P^2 + 2b^2 - 2Pb[\cos(\phi) - \sin(\phi)]\right), \tag{7.4}$$

where b is the convolution of the Fourier transform of the binary Zernike mask and the pupil plane electric field. If the aberrations are small, b can be approximated by b_0, using the pupil function instead of the pupil plane electric field ($\Psi_A = P_0$). Splitting opposite circular polarization states introduces the factor of a half and creates two pupils with different intensities depending on the sign of $\sin(\phi)$. This is used to measure both the amplitude and the phase aberrations in a similar fashion as Wallace et al. (Wallace et al., 2011). They use the sum and the difference of four intensity measurements to directly calculate the amplitude and phase aberration. Here the phase is measured in the same way, only the amplitude is measured as the square root of the pupil intensity. The sum and the difference of the pupil

intensities are given by

$$I_R + I_L = P^2 + 2b^2 - 2Pb\cos(\phi) \tag{7.5}$$

$$I_R - I_L = 2Pb\sin(\phi). \tag{7.6}$$

Using these equations and the identity $\cos^2(\phi) + \sin^2(\phi) = 1$, we solve for the amplitude P and the phase ϕ. The exact reconstruction is given by

$$P = \sqrt{I_R + I_L + \sqrt{4b^2(I_R + I_L) - (I_R - I_L)^2 - 4b^4}} \tag{7.7}$$

$$\phi = \arcsin\left(\frac{I_R - I_L}{2Pb}\right). \tag{7.8}$$

Solving these equations can become numerically unstable for real wavefront sensor applications. Similar to N'Diaye et al. (N'Diaye et al., 2013), we assume that the vector-Zernike wavefront sensor operates in the low-aberration regime, where ϕ dominates the reconstruction. Therefore, $b \approx b_0$ and for the phase reconstruction P = 1. Then we calculate P using the approximated values for ϕ,

$$\phi \approx \frac{(I_R - I_L)}{2b_0} \tag{7.9}$$

$$P \approx \sqrt{I_R + I_L - b_0^2(2 - \cos^2(\phi))} - b_0\cos(\phi). \tag{7.10}$$

Measuring both P and ϕ allows us to estimate the aberrated electric field and approximate b. This process is iterated and converges in the linear regime to the right b, P and ϕ. Note that the reconstruction algorithms do not require many complex calculations such as matrix multiplications to recover the complex wavefront. Therefore the reconstruction is suitable for very fast wavefront reconstruction.

7.3 Simulation

To compare the effect of amplitude aberrations in the reconstruction for both the ZWFS and the vector-ZWFS, we perform numerical simulations of both systems with `HCIPy` (Por et al., 2018). We start with an unobstructed pupil with both phase and amplitude aberrations as shown in the left column of Fig. 7.2. The input phase aberration has a $\sigma_\phi = 0.22$ rad rms and 1.17 rad peak-to-valley (PtV). The normalized amplitude aberration has a $\sigma_\epsilon = 0.6\%$ rms ($\sigma_I = 1.3\%$) and 3.7% PtV (I: 7.2% PtV), where I is the intensity. We reconstruct the phase for the ZWFS using Eq. 14 in N'Diaye et al. (N'Diaye et al., 2013). For the vector-ZWFS we use three different methods to calculate the phase and the amplitude. The first method (*Linear*), uses modal-based wavefront reconstruction to reconstruct 200 Zernike mode coefficients of the phase and the amplitude uses Eq. 7.10. The second method (2nd *order*) uses Eq. 7.9-7.10 to reconstruct the aberrations directly from the intensity measurements with 5 iterations. The third method (*Exact*) uses Eq. 7.7-7.8 using $b = b_0$ and 5 iterations. We compare the standard deviation of

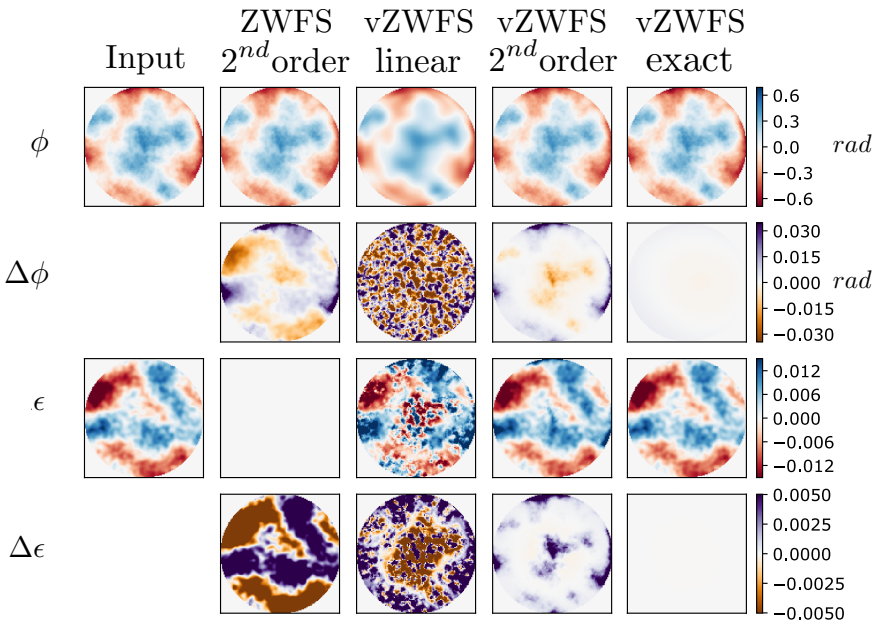


Figure 7.2: Comparison in the simulated reconstruction of both phase and amplitude between the Zernike wavefront sensor (ZWFS) and the vector-ZWFS. The phase aberration is shown in the top row, followed by the residual phase, i.e. the difference between the reconstructed and the input phase. The two bottom rows contain the same with amplitude aberrations. The different columns show the results for different measurement and reconstruction methods.

the residual phase between the ZWFS and the vZWFS and find that the non-linear algorithms of the vZWFS outperform the ZWFS. For the second row in Fig. 7.2, from left to right starting from the input, $\sigma_{\Delta\phi} = [0.221, 0.013, 0.029, 0, 007, 0.001]$ rad rms. The phase reconstruction of the classical ZWFS in the second column of Fig. 7.2 shows that the residual phase aberration in the pupil is caused by the amplitude aberration that it can not directly correct for. This is different from all three vZWFS reconstruction methods, where the residual phase is dominated by the inability to reconstruct high-frequency aberrations with the first 200 Zernike modes (*Linear*), approximations used in the phase calculation (2nd *order*), and the dynamic range of the vZWFS (*Exact*). No significant influence of the amplitude aberration is seen in the residual phase. The amplitude aberration measurements of the vZWFS are dominated by phase estimation errors. The residual amplitude aberrations from left to right, starting from the input, are $\sigma_{\Delta\epsilon} = [1.3, 1.3, 1.4, 0.4, 0.02]\%$ rms.

For practical implementations we look into the influence of imperfect optics shown in Fig. 7.1 and noise on the performance of the vZWFS. Measuring two pupils increases read noise with a factor of 2 and combining the two pupil measurements suffers from flat-fielding effects. In addition, the vZWFS setup requires the use of a liquid-crystal Zernike mask, a quarter-wave plate (QWP) and a polarizing beam splitter (PBS). The vector-Zernike mask has an efficiency (the amount of light that acquires the geometrical phase) that depends on the retardance of the liquid-crystal film (Komanduri et al., 2013). When the phase mask has an offset of the half-wave retardance, $\Delta\delta_{HW}$, the incoming unpolarized light that does not acquire the phase is split by the PBS and adds a background proportional to the pupil intensity. Similar contaminations happen with retardance deviations of the QWP and the rotation of the PBS. The consequent left pupil intensity distributions (identical to the right pupil) from an imperfect HWP, QWP and PBS respectively are given by

$$I_{L,\Delta 1} = I_L \cos^2(\frac{1}{2}\Delta\delta_{HW}) + \frac{1}{2}P^2 \sin^2(\frac{1}{2}\Delta\delta_{HW}), \tag{7.11}$$

$$I_{L,\Delta 2} = I_L \cos^2(\frac{1}{2}\Delta\delta_{QW}) + \frac{1}{2}(I_L + I_R) \sin^2(\frac{1}{2}\Delta\delta_{QW}), \tag{7.12}$$

$$I_{L,\Delta 3} = I_L \cos^2(\Delta\theta) + I_R \sin^2(\Delta\theta). \tag{7.13}$$

To first order, the difference of the pupil intensities is most affected by the extra terms. All terms proportional to $\sin^2()$ drop out and the difference now is proportional to a $\cos^2()$ term that reduces the response. The reduced response can be corrected with a gain factor from calibration. The sum of the pupil intensities is only affected by the deviation from half-wave retardance of the vector-Zernike mask, all other terms sum to the total intensity $I_L + I_R$. The chromatic retardance can be controlled with liquid-crystal technology such that $\Delta\delta_{HW} < 16°$ (Komanduri et al., 2013) for a large bandwidth, i.e. the leakage is $< 2\%$ of the intensity.

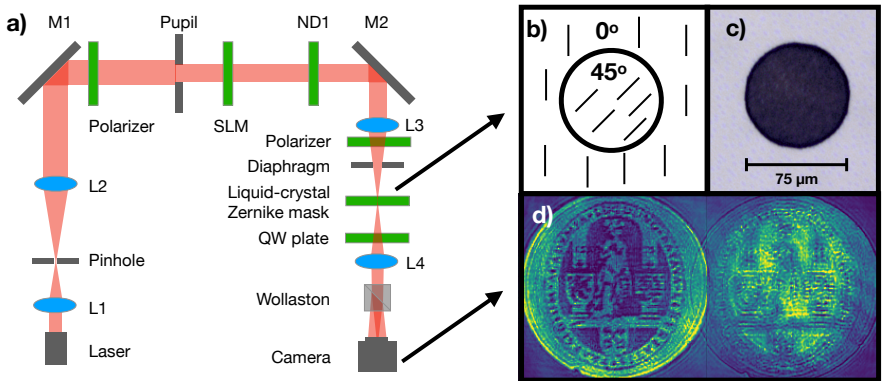


Figure 7.3: **a)** Layout of the vector-Zernike wavefront sensor setup. We generate a clean beam with a laser (633 nm) and a pinhole. A spatial light modulator is operated in phase-mostly configuration with two polarizers and is used to generate phase aberrations with a complicated pattern (i.e. the Leiden University logo). This configuration also generates some amplitude aberrations. The light is focused on the vector-Zernike mask with a spot diameter of 75 μm, corresponding to $\sim 1\lambda/D$. The liquid-crystal orientation is shown in panel **b)** and a parallel polarizer microscopic image is shown in panel **c)**. The detector image with aberrated pupils is shown panel **d)**, showing some pupil overlap because of the splitting angle of the Wollaston.

7.4 Lab results

As a proof of principle, we demonstrate the vector-Zernike wavefront sensor with a monochromatic setup. We manufactured a liquid-crystal vector-Zernike mask with a central spot diameter of 75μm, printed on a glass substrate (1 inch diameter, 1 mm thick, BK7) with an effective pixel size of 2 μm. The retardance was tuned to be $\lambda/2$ at 633 nm with a single liquid-crystal layer. Fig. 7.3 b) shows the fabricated mask that does not contain any significant defects. The edges have a smooth fast-axis transition region from 0° to 45° of 2-3 micron. This can be mitigated by modifying b for calculating P and ϕ, although the effect is negligible. We test the performance of the vector-Zernike using the setup shown in panel a) of Fig. 7.3. We use a LC2002 transmissive spatial light modulator (SLM) from HoloEye in phase-mostly configuration to generate phase aberrations. In this configuration the rotation of the liquid-crystals in combination with the polarizers also generates amplitude aberrations. The Wollaston prism is a WPQ10 prism from Thorlabs and the QWP is an achromat (AQWP05M-600) from Thorlabs. Panel d) in Fig. 7.3 shows a raw image after applying a binary logo of Leiden University on the SLM. For the characterization of phase and amplitude reconstruction we minimize the influence of the system by combining measurements with positive and negative phase applied by the SLM. We mean-combine them to remove the

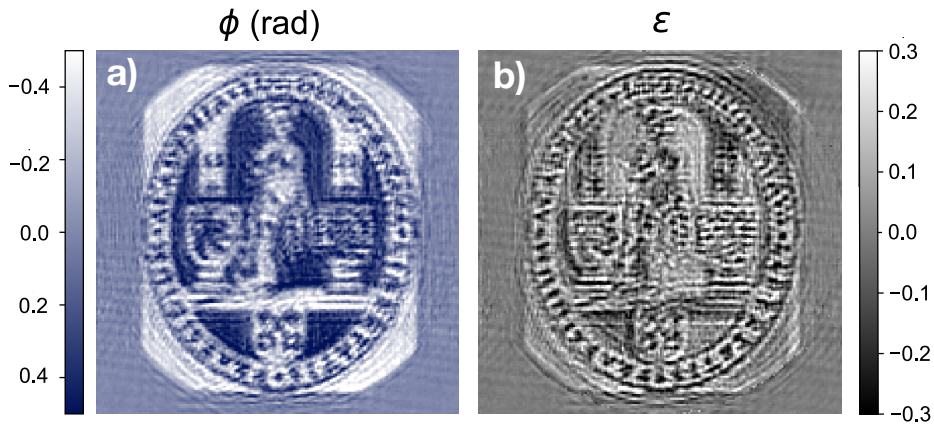


Figure 7.4: **a)** Reconstructed phase and **b)** reconstructed amplitude aberration using the vector-Zernike wavefront sensor.

non-SLM aberrations and we bin the image per 2 pixels to increase the signal to noise. We also take images without any phase applied on the SLM and use them for calculating the pupil intensity necessary for normalizing the aberrated images without taking out the vector-Zernike mask. Note that taking the average pupil intensity does not normalize the measurements correctly. Assuming $\phi \ll 1$, we solve Eq. 7.5 with $\cos\phi = 1$, such that P is given by

$$P = \sqrt{(I_L + I_R)/(1 + 2b_0^2/P^2 - 2b_0/P)}. \tag{7.14}$$

When $\epsilon \ll 1$, the normalized quantity $b_0/P \approx b_0/P_0$ and we calculate the normalization factor of the pupil images by estimating b_0 and averaging P over the pupil. We calculate the phase and amplitude from the two aberrated normalized pupils with Eqs. 7.9 - 7.10. Both measurements contain Fresnel diffraction effects from the out-of-plane diaphragm used as field stop. Determining the accuracy of the reconstruction is not possible as we do not know the details of this SLM.

7.5 Chromaticity of the vZWFS

One additional advantage of the liquid-crystal vector-Zernike wavefront sensor is that the phase is applied achromatically cf. Eq. 7.1. The classical ZWFS has decreased performance with increasing spectral bandwidth because of the chromatic phase of the focal plane mask in combination with the PSF scaling with wavelength. Both contribute to a suboptimal coupling of phase aberrations in the system to intensity variations in the ZWFS pupil plane (N'Diaye et al., 2013). We study the effect of the achromatic phase and the different reconstruction on the broadband performance of the vZWFS in simulation, taking into account the

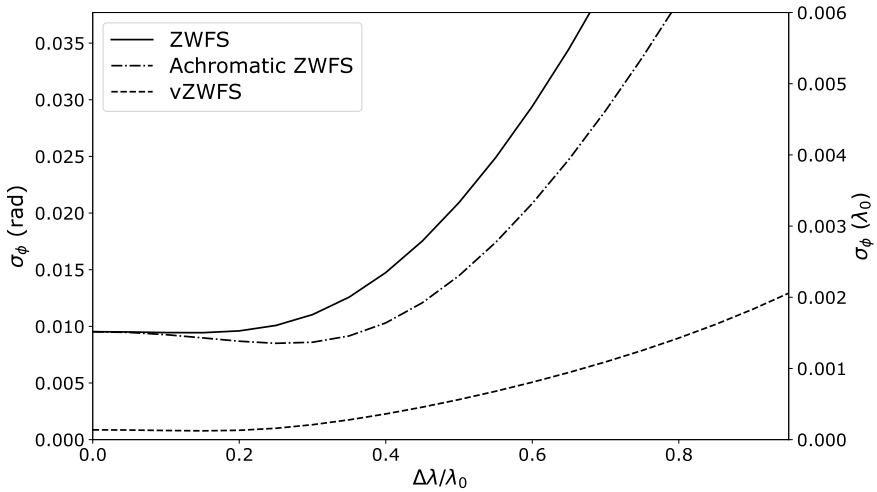


Figure 7.5: Residual wavefront aberrations as function of bandwidth for both the classical ZWFS and the vZWFS. The "achromatic ZWFS" uses the classical ZWFS reconstruction on one pupil of the vZWFS.

spatial scaling of the PSF with wavelength. For this, we exclude amplitude aberrations. Fig. 7.5 shows the results for both wavefront sensors up to 100% bandwidth, i.e. $\lambda/\lambda_0 = 0.5 - 1.5$. We assume a $1/\lambda$ scaling for the applied phase of the classical Zernike mask centered around λ_0 and b_0 is calculated at λ_0 as well. We investigate the influence of only the achromatic phase by using the classical ZWFS reconstruction on one of the vZWFS pupils, the "achromatic ZWFS". Note that all other chromatic effects, including leakage from polarization optics, are not taken into account in this simulation. Fig. 7.5 shows that the vZWFS with exact reconstruction outperforms the classical ZWFS for all bandwidths. Similarly, the classical reconstruction with the "achromatic ZWFS" improves the broadband performance, in agreement with Bloemhof (Bloemhof, 2014). The gain of using only achromatic phase is not as significant as the gain from the reconstruction using two pupils. The reason is that the classical ZWFS uses a second-order reconstruction that is sensitive to the chromatic b, while the vZWFS allows for an exact solution. Iterating the estimate of b with the updated phase and amplitude estimates removes the bias introduced by the chromaticity. Overall we show in simulation that the vZWFS can handle an increased bandwidth, up to 100%.

7.6 Conclusion

We conclude that the liquid-crystal vector Zernike wavefront sensor (vZWFS) is a simple yet powerful upgrade of the classical ZWFS. We show that both amplitude and phase aberrations can be measured simultaneously by replacing the Zernike mask with a liquid-crystal version and splitting circular polarizations in two pupils. We demonstrate both in simulations and with a monochromatic lab setup an improved reconstruction of the complex wavefront with the vZWFS. Like the ZWFS, the vZWFS is the most photon-efficient wavefront sensor (Guyon, 2005) and we demonstrate with simulations the accurate reconstruction for bandwidths up to 100%. For direct-imaging of exoplanets the vZWFS can be installed a parallel WFS beam path, alternated with a coronagraph (Vigan et al., 2018), or built into a coronagraph (Shi et al., 2018).

Funding. The research of DSD and FS leading to these results has received funding from the European Research Council under ERC Starting Grant agreement 678194 (FALCONER)

Bibliography

Bloemhof, E. 2014, Proc. SPIE, 8999, 89991k
Cuche, E., Marquet, P., & Depeursinge, C. 1999, Applied optics, 38, 6994
Escuti, M. J., Kim, J., & Kudenov, M. W. 2016, Optics and Photonics News, 27, 22
Goodman, J. 1996, Introduction to Fourier Optics, Electrical Engineering Series (McGraw-Hill). https://books.google.nl/books?id=jMcsmwEACAAJ
Guyon, O. 2005, The Astrophysical Journal, 629, 592
Komanduri, R. K., Lawler, K. F., & Escuti, M. J. 2013, Optics Express, 21, 404
Miskiewicz, M. N., & Escuti, M. J. 2014, Optics Express, 22, 12691
N'Diaye, M., Dohlen, K., Fusco, T., & Paul, B. 2013, Astronomy & Astrophysics, 555, A94
Oh, C., & Escuti, M. J. 2009, Optics Letters, 34, 3637
Paterson, C. 2008, Journal of Physics: Conf. Series., 139, 012021
Por, E. H., Haffert, S. Y., Radhakrishnan, V. M., et al. 2018, Proc. SPIE, 10703, 1070342
Shi, F., Seo, B.-J., Cady, E., et al. 2018, Proc. SPIE, 10698, 106982O
Vigan, A., N'Diaye, M., Dohlen, K., et al. 2018, arXiv preprint arXiv:1806.06158
Wallace, J. K., Rao, S., Jensen-Clem, R. M., & Serabyn, G. 2011, Proc. SPIE, 8126, 81260F, doi: 10.1117/12.892843
Yamamoto, K., Matsuo, T., Imada, H., & Kino, M. 2015, Applied optics, 54, 7895
Zernike, F. 1934, Monthly Notices of the Royal Astronomical Society, 94, 377
—. 1942a, Physica, 9, 686
—. 1942b, Physica, 9, 974

8 Multi-color holography with a two-stage patterned liquid-crystal element

Adapted from
D. S. Doelman, M. J. Escuti, F. Snik
Published in Optical Materials Express (2019)

Computer-generated Geometric Phase Holograms (GPHs) can be manufactured with high efficiency and high fidelity using photo-aligned liquid crystals. GPHs are diffractive elements, which therefore have a wavelength-dependent output and can generally not be used for the production of color imagery. We implement a two-stage approach that first uses the wavelength-dependent diffraction to separate colors, and second, directs these colors through separate holographic patterns. Moreover, by utilizing the geometric phase, we obtain diffraction efficiencies close to 100% for all wavelengths. We successfully create a white light hologram from RGB input in the lab. We demonstrate that this schematic allows for full control over individual (RGB) channels and can be used for wide-gamut holography by selecting any combination of wavelengths. In addition, we show with simulations how this two-stage element could be used for of true-color holograms with applications in high contrast imaging.

8.1 Introduction

Computer generated holography is the art of shaping light into custom 2D and 3D intensity profiles, often used to recreate the image of an object or a scene (Brown & Lohmann, 1966, 1969; Slinger et al., 2005; Lohmann, 2008; Yaraş et al., 2010). An optical element that encodes a computer generated hologram (CGH) must be able to change the phase or amplitude of light and must be programmable at least once for passive holography or multiple times for adaptive holography. In general, liquid-crystal (LC) layers are able to change phase, polarization and amplitude of light, depending on the orientation of the optic axis and surrounding films. A specific type of phase element that can be manufactured with LCs is a geometric-phase hologram (GPH) (Escuti et al., 2016; Kim et al., 2015; Hasman et al., 2003; De Sio et al., 2016). GPHs generate geometric phase shifts, wherein a phase change arises from the geometry of the evolution the light through the anisotropic medium. When LCs are used to form GPHs, this anisotropy occurs due to the material's birefringence, and this is also sometimes called the Pancharatnam-Berry phase (Pancharatnam, 1955; Berry, 1984; Anandan, 1992).

Besides liquid crystals, there are other materials capable of generating geometric phase holograms. These materials are made of metasurfaces or nano-structures and allow for full control of phase, amplitude and polarization simultaneously (Huang et al., 2013; Ni et al., 2013; Mueller et al., 2017). However, these materials still suffer from high scattering or low efficiencies, especially for broadband light ($\Delta\lambda/\lambda > 30\%$) (**?**Huang et al., 2016). Therefore we focus on liquid-crystal GPHs.

A LC GPH operates on circular polarization, *i.e.,* an incoming lightwave is decomposed into the two circular polarization states and the sign of the geometric phase depends on the handedness of each state. Its wavefront may be spatially varied by locally varying the orientation, Φ, of in-plane optic axis of the LC. The resulting geometric phase, δ, depends on the orientation and the handedness of the circular polarization according to

$$\delta(x, y) = \pm 2\Phi(x, y). \tag{8.1}$$

Creating arbitrary patterns is possible with a photo-alignment layer (PAL) (Chigrinov et al., 2008). Polymers in the PAL record the orientation of incoming linearly-polarized UV radiation and direct-write laser scanning (Miskiewicz & Escuti, 2014; Kim et al., 2015) allows for writing arbitrary patterns with high precision. The initial orientation pattern in the PAL and first LC layer is passed on to subsequent layers by spontaneous self-alignment. In this way, LC layers with different thickness and chiral twists may be arranged into a monolithic film (Komanduri et al., 2013). The many degrees of freedom of these multi-twist retarder (MTR) films may be used to tune the retardation of the LC layer to half-wave retardance, achieving high efficiency (Kim et al., 2015; Escuti et al., 2016) in GPHs over large bandwidths into a single diffracted order. Furthermore, GPHs are ideal to record CGHs as they can be manufactured with unbounded continuous phase, wide angular bandwidths, and very high spatial frequencies (Xiang et al., 2017a, 2018, 2017b).

The simplest example of a GPH is a polarization grating (PG). A PG has a linear ramp in phase in one direction, similar to a refractive prism. The two circular-polarization states are diffracted in opposite directions and the leakage is directly transmitted, corresponding to the first- and zero-diffraction orders, respectively. The broadband diffraction efficiencies for PGs can be manufactured to be close to 100% for the full visible range (Oh & Escuti, 2008). Similar to any other grating, the diffraction angle of a PG is dependent on wavelength. This is common for all GPHs. A computer generated Fraunhofer CGH will therefore have the same intensity distribution for all wavelengths, only wavelength-scaled (Xiang et al., 2017a). Note that CGHs created with dynamic phase elements, *i.e.,* produced by locally changing the optical path length, have decreased fidelity for wavelengths than other than the one the pattern was optimized for. For Fresnel CGHs the size of the hologram is the same for each wavelength, but the imaging position is now wavelength-dependent (Jesacher et al., 2014; Huang et al., 2016).

Generating multi-color holograms requires encoding different phase patterns for individual colors. Not only is the intensity distribution of the hologram usually different for each color, even generating the same intensity distribution for each color also requires scaling the phase to match the wavelength. For clarity, we enumerate here the different methods that generate multi-color holograms by the number of used elements. A first method uses volume holograms that stores the color information in a single thick optical element (Leith et al., 1966). A second set of methods use spatial light modulator (SLM). One uses shutters to illuminate the SLM with only one color at the time (Shimobaba & Ito, 2003). Fast switching of the color and phase pattern generates the illusion of a color hologram. In another, the SLM is combined with color filters for pixel columns, such that the RGB channels each acquire the desired phase (Tsuchiyama & Matsushima, 2017). Finally, it is also possible to use an SLM with a large dynamic range ($0 - 10\pi$) to encode a single multi-color hologram (Jesacher et al., 2014). A third method optimizes the phase pattern of a SLM to generate the RGB holograms spatially separated in the propagation direction (Makowski et al., 2008). Each color generates three consecutive holograms calculated for a specific distance using Fresnel propagation. At one specific distance all three colors combine to a single multi-color hologram. Similarly, it is also possible to separate the input spatially in the same image plane (Ito & Okano, 2004). When the RGB inputs originate from a different point on the image plane, all elements could have the same phase and combine in the center to a multi-color hologram.

Multi-element systems for color holography usually separate the light. Analogous to standard projection systems, most multi-element holography systems separate white light in RGB arms, each with its own SLM, using dichroic mirrors and beamsplitter cubes and recombining them in a similar fashion (Shiraki et al., 2009). This complex optical system uses three phase elements (SLMs) for the RGB channels.

In this paper we propose a simpler two-stage system capable of producing multi-color holograms with high efficiency.

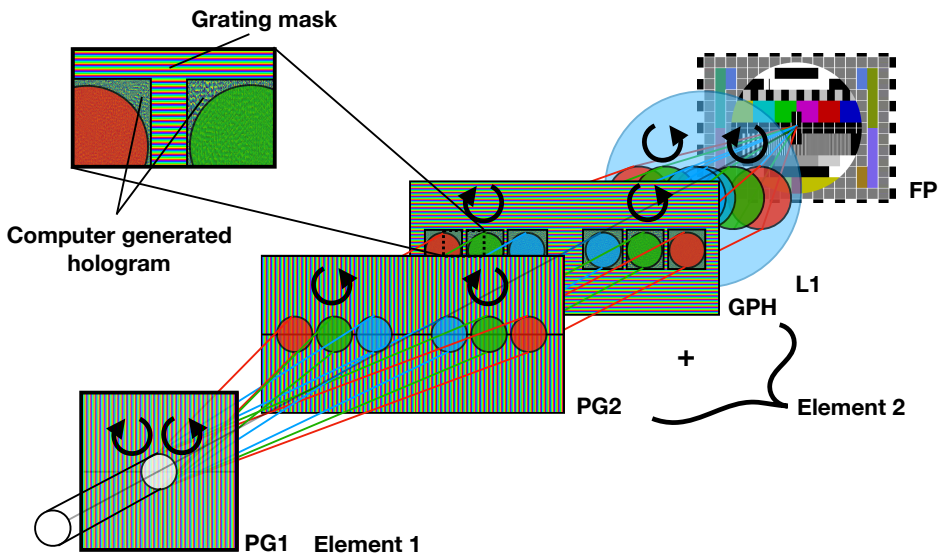


Figure 8.1: Schematic of the multi-color geometric phase hologram. A polarization grating (PG1) separates the multi-color input. A second stage collimates the beams with a polarization grating (PG2) and add the individual computer generated geometric phase hologram (GPH) to the pupils with different colors. A lens combines both circular polarizations to form one multi-color hologram. The second stage can be composed of two separate elements or comined into one single element.

8.2 Concept of multi-color liquid-crystal holography

As discussed above, GPHs are efficient over a large bandwidth and apply wavelength-independent geometric phase shifts (Xiang et al., 2017a). However, they have so far not been used much for multi-color holography as they are diffractive, generating wavelength scaled holograms. We present a two-stage element that produces a multi-color hologram by separating different colors with a PG and applying individually tuned CGHs, embodied as GPHs, to each color. The concept is shown in Fig. 8.1. A white-light pupil, containing equal amounts of red, blue and green light, is diffracted by the polarization grating. A polarization grating (PG) separates the pupil with efficiencies close to 100% in the ± 1-orders, where the intensity distribution between the $+1$ and -1 order depends on the amount of circular polarization in the source. The second stage does two things simultaneously. First, it collimates the light with a polarization grating pattern that is exactly the same as the first element. The beam is collimated with high efficiency because the polarization gratings have polarization memory. In addition, the second stage adds

the phase delay corresponding to the CGH for each individual color. Because the colors are separated, the CGH can be adjusted to make the correct intensity distribution for each wavelength. Combining the computer generated hologram with the polarization grating is possible because a polarization grating effectively adds a geometric phase tilt to the beam and can therefore be added in phase to the CGH pattern (Doelman et al., 2017). Finally, a high-frequency grating pattern can be added outside of the CGH to diffract light outside of the pupil away from the hologram. This grating mask (Doelman et al., 2017), removes the need for individual amplitude masks for individual colors. For clarity, the two phase patterns (PG and GPH) are shown in a consecutive order in Fig. 8.1. When the phase delays are applied to the collimated beams, all colors can be recombined in the image plane by imaging the beam with a lens. It is also possible to propagate to the far field, however the different colors will not overlap.

For the best result, individual colors need to be separated by the multi-color geometric phase hologram. This means that the distance between the two elements, d, and the grating frequency, n, need to be tuned to the hologram properties. Here n is the number of periods per meter. To make sure two pupils of different color do not overlap, the distance between the centers of the phase patterns, ΔL, is given by

$$\Delta L = d(\tan(n\lambda_2) - \tan(n\lambda_1)) > D, \tag{8.2}$$

Where D is the physical size of the computer generated holographic phase pattern and λ_1 and λ_2 are the two selected wavelengths where $\lambda_2 > \lambda_1$. When pupils overlap, two colors start mixing and this generates a scaled hologram similar to classical geometric phase holograms. The distance between the two elements is a free parameter and can be used to accommodate the correct separation.

Because the colors are physically separated, there is no requirement on the hologram shape or phase from other wavelengths. Therefore, the phase patterns that generate the hologram can be calculated individually. Individual color channels can have different shapes or sizes as long as it produces the desired holographic intensity distribution. For example, when generating a white-light hologram (the RGB channels produce the same hologram), the phase pattern can be scaled to match the wavelength dependent diffraction limit. To generate the phase patterns we use the Gerchberg-Saxton (GS) algorithm (Gerchberg & Saxton, 1972) implemented in HCIPy (Por et al., 2018). GS iterates between the focal and pupil plane and applies the pupil shape and the desired focal plane intensity as constraint, keeping the calculated phase. This way, arbitrary intensity distributions can be calculated.

Because the multi-color GPH combines liquid-crystal elements with high efficiencies and polarization memory, the multi-color GPH achieves close to 100% efficiency when using both circular polarization states. For larger holograms where the distance between left- and right-circular polarization becomes significant, imaging both with the same lens becomes impractical. Selecting one polarization reduces the efficiency by a factor of 2, but allows for a more compact system. Both circular-polarization states can also be used to create different holograms for re-

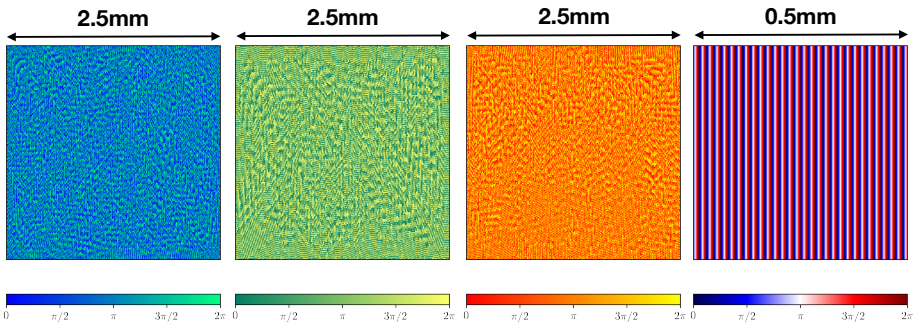


Figure 8.2: The phase designs for the computer generated holograms tested in the lab. The left panel shows the phase design for the blue laser ($450nm$), as indicated by the colormap. Similarly, the two center panels show the green and red phase design. The right panel shows the zoomed pattern corresponding to the polarization grating.

ducing speckles in the image with more homogeneous illumination and accessing different spatial frequencies (Kowalczyk et al., 2018). Taking this further, the two circular polarization states can create individual holograms for stereoscopic 3D holography. This is fully compatible with circularly polarizing eyewear for viewing projected stereoscopic motion pictures.

8.3 Lab experiment and result

8.3.1 Design of the multi-color geometric phase holograms

Next we test the concept of the multi-color geometric phase patterns. Using the Gerchberg-Saxton algorithm we generate three phase patterns for a focal plane hologram containing two separate neighbouring structures. The first structure is only wavelength scaled such that all three colors add to one white light hologram. The second structure was generated to test color blending necessary for wide gamut holography. We vary the intensity distribution in one direction only, adapted to combine the three colors at different intensity values. We place the three holograms next to each other as the ratio between the red (440 nm), green (510 nm), and blue (635 nm) is approximately the same. The three holograms can be operated in stand-alone mode, *i.e.* they do not contain the polarization grating. Figure 8.2 shows the fabricated phase patterns calculated for a square aperture. For clarity, the panels have different color maps that match the three colors belonging to the wavelengths the holograms are calculated for. A segment of the PG pattern is also shown in Fig. 8.2, the total size of both PGs is $10mm$, the three holograms have 5μm pixels, totalling 500 pixels. The period of the two polarization gratings is set to 17 μm to generate a more compact setup. The grating pattern was calculated

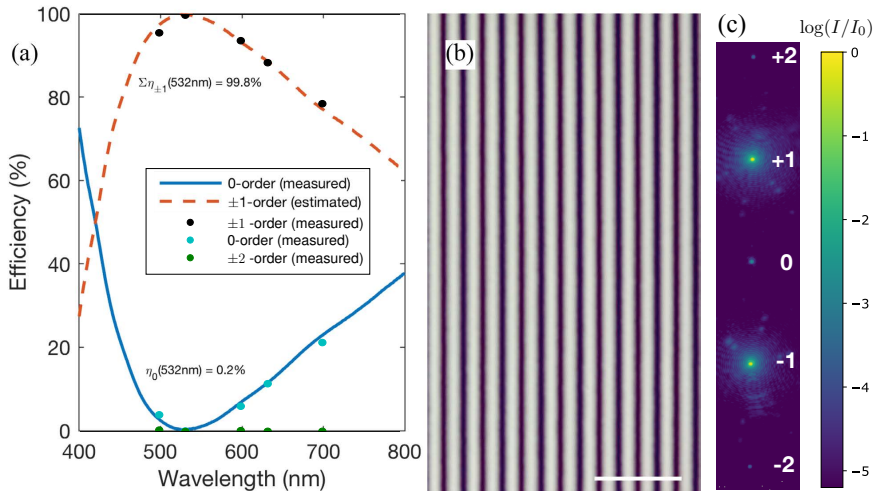


Figure 8.3: Measured properties of the fabricated PG: (a) diffraction efficiencies of the zero-, total first- and second-orders; (b) polarizing optical micrograph, where the scale bar indicates 50 μm; and (c) the PSF of the PG at 532nm that includes the first two diffraction orders, showing that the second-order diffraction efficiency is $<< 1\%$.

with a pixel size of 1μm. The patterns are checked for high gradients between single pixels to prevent disclination lines from appearing. No changes were made to the phase pattern as the gradients are mostly well behaved.

8.3.2 Manufacturing process

Fabrication begins with coating the azo-based photo-alignment layer (Chigrinov et al., 2008) LIA-CO01 (DIC Corp) on clean glass (D263) substrates (spin: 30 s @ 1500 rpm, bake: 60 s @ 130°C), about 30 nm thick. For exposure, we employ the direct-write laser scanning system described in Refs. (Miskiewicz & Escuti, 2014; Kim et al., 2015), using a solid-state 355 nm laser (Coherent Inc) delivering an exposure energy of about 1 J/cm^2. The PG pattern was simply a linear phase profile, while the three hologram phase patterns were discussed in the prior section, all sampled with 5 μm resolution. Both plates were processed and scanned in the same way, except for the scanned orientation profile. For the first liquid crystal polymer (LCP) network sublayer, we use a solution comprising solids RMM-A (Merck KGaA, $\Delta n = 0.16$ @ 532 nm) in solvent propylene-glycol-methyl-ether-acetate (PGMEA from Sigma-Aldrich), with a 5% solids concentration. This was processed (spin: 55 s @ 1200 rpm, cure: 30 s @ 190 mW of UV illumination from a 365 nm LED in dry nitrogen environment). For the second and third sublayers, we use a second solution, comprising 20% RMM-B (Merck KGaA, $\Delta n = 0.25$ @ 532 nm) in solvent PGMEA. This is processed (spin and cure steps identical to

the first LCP sublayer except for spin at 700 rpm). This lead to a total thickness of approximately 1.1 μm, as measured by an ellipsometer, which is a half-wave retardation around 532 nm.

In Fig. 8.3, we show measured properties of the fabricated PG. We were not able to directly measure the holograms in the same way due to their much more complex phase pattern. All samples had no observable haze and nearly all the output power appeared in the zero- and first-orders. Fig. 8.3(a) shows the measured zero-order efficiency spectrum of the PG, where $\eta_0 = 0.2\%$ at 532 nm and was below 50% for the red and blue wavelengths. The total first-order efficiency was estimated as $\Sigma\eta_{\pm 1} = 100\% - \eta_0$, also shown. We can therefore estimate $\Sigma\eta_{\pm 1} = 99.8\%$ at 532 nm. Both of these efficiency measurement should also apply to the three holograms, because the LCP coating was the same. A polarizing optical microscope was used to capture the LC texture of the PG, showing a smoothly varying sinusoidal variation in intensity, which indicates a very linear phase (*i.e.*, LC optic axis orientation).

8.3.3 Lab setup

We test the liquid-crystal geometric phase holograms in the lab using three laser diode modules as input. The laser diodes are the CPS450, CPS532a and CPS635S from Thorlabs, operating at 450 nm, 532 nm, and 645 nm respectively. We use the equal power of the lasers to match the input color intensity with the RGB response of the camera. We combine the beams in a single mode fiber using two 50/50 beamsplitters and a lens. The single mode fiber (SMF) creates a diffraction limited beam as input for the multi-color GPH. The light from the SMF is collimated and stopped down to the size of 2 mm. The first polarization grating diffracts the light off-axis. We use a second aperture to filter out the leakage light and the opposite circular polarization state. The second PG collimates the light and the GPH applies the phase of the computer generated hologram. A circular polarizer is used to filter out the GPH leakage term and the three pupils are imaged on a Canon EOS 6D CMOS camera. The setup is shown in Fig. 8.4.

Both structures in the hologram are imaged simultaneously with the camera. The Canon EOS 6D CMOS is a color camera, the color space is set to sRGB and we extract the raw frames directly.

8.3.4 Experimental results

We image the holographic structures with the three different colors and all combinations. The results are displayed in Fig. 8.5. The first structure, the polygon, has all three colors overlapping in both dimensions. Each combination of the colors, shown in the second row, produces the expected colors, being yellow, cyan and magenta from left to right. The colors are slightly saturated. However, the RGB combination in the bottom row clearly shows a white light hologram, with matching scales for the individual colors. This demonstrates that each pupil acquired the phase from the pattern designed for that wavelength and that the pupils do

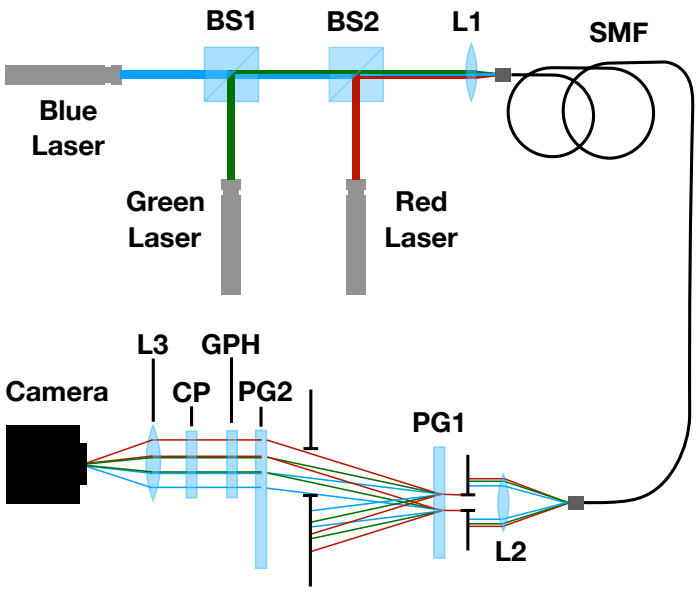


Figure 8.4: Lab setup used to test the multi-color GPH. The three lasers are combined with two beam splitters (BS) and imaged with a lens (L1) on a single mode fiber (SMF). The light from the SMF is collimated with a second lens (L2). Two irises create the pupil and select a single circular polarization from the two polarization gratings (PGs). A circular polarizer (CP) is used to filter leakage and third lens (L3) images the hologram on the detector.

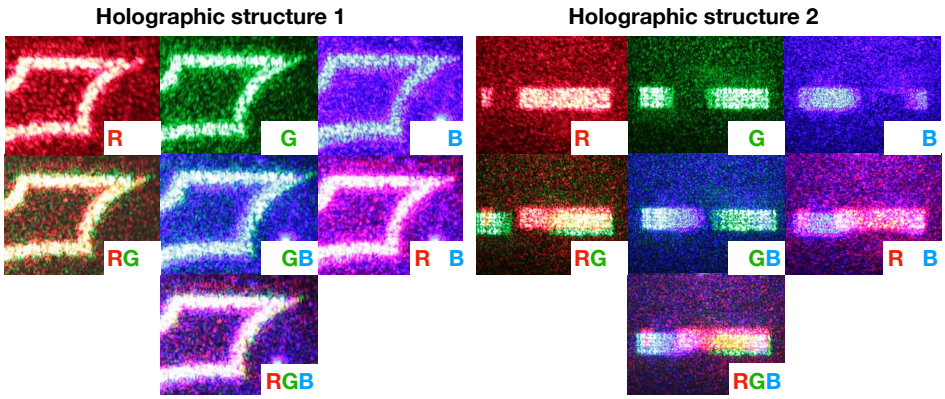


Figure 8.5: Measured intensity of the holograms for each combination of the RGB input colors. The multi-panel image on the left shows the first structure of the hologram, a 2D polygon. On the right is the one dimensional second structure with varying intensities, producing multiple colors.

not overlap. The second structure contains intensity variations in the x-direction and this allows to access different colors in the color space than the first structure. Both structures show that the multi-color GPH is capable of producing a multi-color hologram.

We selected the three laser diodes because of their equal power of 4.5 mW. The green laser has some offset compared to the designed wavelength, 532 nm instead of 510 nm. This can be seen both structures of the hologram, the green one is slightly larger and offset compared to the red and blue images. Moreover, both structures contain speckles instead of a more homogeneous intensity distribution. This is a well known problem for the Gerchberg-Saxton (GS) algorithm (Kuratomi et al., 2010; Pan & Shih, 2014) and can be solved by implementing the double-constraint Gerchberg-Saxton algorithm (Chang et al., 2015). In addition, there is a non-zero background with speckles that is not part of the structures. This background should not be attributed to the scattering of the liquid-crystal film (Xiang et al., 2017a), but rather, to the hologram phase design. In this design the hologram has a background that is only suppressed by a factor of 100, and can still be seen by the camera.

8.4 Wide-gamut and true-color holography

8.4.1 Color selection

Multi-color geometric phase holograms are not limited to RGB holography. The polarization grating physically separates all colors in a spectrum and the GPH can modify them individually when the differently colored pupils are separated.

Selecting more or other colors than RGB can create a more natural hologram, *i.e.*, the spectrum is closer to the reflectance spectrum of the object the hologram represents (Peercy & Hesselink, 1994). Note that the difference in color rendition of most objects with more than seven different wavelengths is almost indistinguishable for the human eye (Bjelkhagen & Mirlis, 2008). The optimal wavelength selection for 3-5 different colors is written in Bazargan et al. (Bazargan, 1992). Any number of separate holograms can be added as long as the separation follows Eq. 8.2 and the combined width is still within the manufacturing range ($<$ 30 cm (Kim et al., 2015)). Using more colors with a constant pupil size requires an increase in d, the distance between the elements, or an increase in the grating frequency n. The latter is limited by the writing resolution of the direct-write system, the former is limited by available space.

8.4.2 True-color holography

The multi-color geometric phase hologram also works with a white light source. With the white light source, the second element is illuminated with a continuum of dispersed pupils. This requires a different method of optimization of the phase pattern, where the phase pattern is calculated for all wavelengths simultaneously. However, using a white-light source gives access to the full gamut space and allows for generating the actual reflectance spectrum for even the most exotic objects.

To evaluate this concept, we investigate true-color geometric phase holograms in simulation. We divide the phase pattern with a width of 4 pupil diameters in 36 different colors, or wavelength bins, with overlapping pupils. Each color has its own focal plane constraint, *i.e.*, the desired intensity in the focal plane is a function of wavelength. The focal plane design is shown in the top right of Fig. 8.6. Note that the locations where the RGB colors normally overlap have been purposefully replaced by the wrong colors. They are not formed by the combination of RGB light but the light that is directed in these regions has a wavelength corresponding to the indicated color, *i.e.*, a true-color hologram. In addition, this hologram design is specifically good for checking color blending from overlapping pupils. Therefore, we replace magenta by violet, as magenta, being a combination of blue and red colors, is already blended. The constraints as function of wavelength are shown with dotted lines in bottom left of the same figure. We start with a uniform pupil and a random phase for each color and perform one iteration of GS. After applying the pupil plane mask for each wavelength, we mean-combine all phases to a single phase pattern, masking the regions that do not overlap and would add zeros to the mean. Because the pupil plane sampling is high and the focal plane constraint is smooth in wavelength, the individual phase patterns are similar for neighbouring colors. Therefore, averaging does not remove the calculated phase entirely and the algorithm converges when the averaged phase is used as input for the next iteration.

The combined phase pattern is shown on the top left of Fig. 8.6 and has 500x100 pixels. The color of the light the phase pattern is illuminated with is indicated by the color bar on top of the phase pattern. The phase pattern shows roughly six

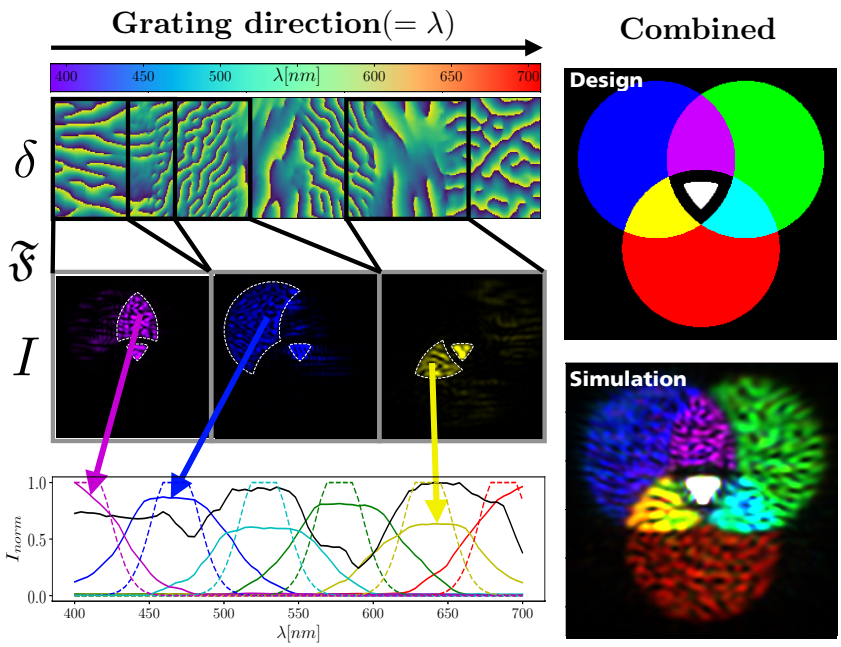


Figure 8.6: Simulation of a true-color hologram with a two-stage geometric phase hologram (GPH). A square pupil is dispersed by the first grating over a width of four pupil diameters when arriving at the GPH. *Top left:* the central color of the pupils is indicated by the color bar and the phase pattern is shown below. *Top right:* the design of the hologram, were the intersection of the RGB circles are purposely altered. *Center left:* Three pupils are selected and with Fourier propagation the intensities are calculated. *Bottom right:* The summed intensities multiplied in the six regions indicated for the six different colors as function of wavelength (solid line). The normalized spectrum of the center (white light) is shown in black. The images show reasonable agreement with the input spectra (dotted line). *Bottom right:* the combined hologram, integrated over all colors, converted to RGB. The simulation shows that using a two-stage GPH is capable of producing true-color holograms.

different regimes in the grating direction, corresponding to the six colors in the designed hologram. The bottom left in Fig. 8.6 show the hologram for a single wavelength and in solid lines the summed and normalized intensity in the differently colored regions. It is clear that the spectra of the colors are not perfectly constrained, the drop-off is much slower than designed and the intensities are not equal but deviate by a factor 2. The drop-off is mostly constrained by the pupil overlap, *i.e.*, the ratio of the phase plate width and the pupil diameter, and the gradient in the desired holographic spectrum. For this simulated full-color hologram the dominating colors in the overlapping regions are still the same as in the design. We calculate the colors by summing the focal plane intensities as function of wavelength multiplied by the design spectrum. This gives the holographic intensities for six colors. We calculate the RGB colors by spitting the cyan, violet and yellow colors evenly over the RGB channels. We normalize the intensities on the central white light region and saturate it to make the colors more visible. Overall the simulated hologram is in good agreement with the design, showing that in simulation true-color holography is possible with a two-stage liquid-crystal element.

The advantage of this two-stage liquid-crystal element for holography are the high efficiency and high-fidelity simultaneous beam shaping for multiple wavelengths and opposite circular polarization states. In the presented state with passive liquid-crystal films it is only possible to generate static holograms. Therefore, this technology is easily applicable for projection, either on a reflective surface or directly into the eye of an observer. In addition, when using the two opposite circular polarization states it is possible to make 3D stereoscopic holograms. These properties make the two-stage system interesting for the field of visual arts.
Making an active two-stage element is possible by replacing the geometric phase hologram with a separate polarization grating and a spatial light modulator (SLM). When the wavelengths are separated by two polarization gratings, the phase errors from using classical phase are relatively small. Such an active two-element system has a wider range of applications. Most notably, scientific applications like optical tweezers, where multi-wavelength intensity field manipulation could be useful for optical manipulation of microparticles with different absorption spectra (Hayasaki et al., 1999; Grier, 2012). Compared to the phase contrast image synthesis method (Glückstad, 1996; Palima & Glückstad, 2008), the two-stage element allows for different beam shapes for multiple wavelengths, giving more freedom for particle manipulation.

8.4.3 True-color vAPP coronagraphs

Another possible implementation of the true-color holograms is in the field of direct imaging of extra-solar planets (exoplanets). The light of exoplanets is 10^4- 10^{10} times fainter than the stellar light, and exoplanets orbit close to the star compared to the distance relative to Earth. The angular separation is often less than an arcsecond, 1/3600th of a degree, which is only a few diffraction widths away for the largest telescopes. The diffraction structure of a stellar PSF is many

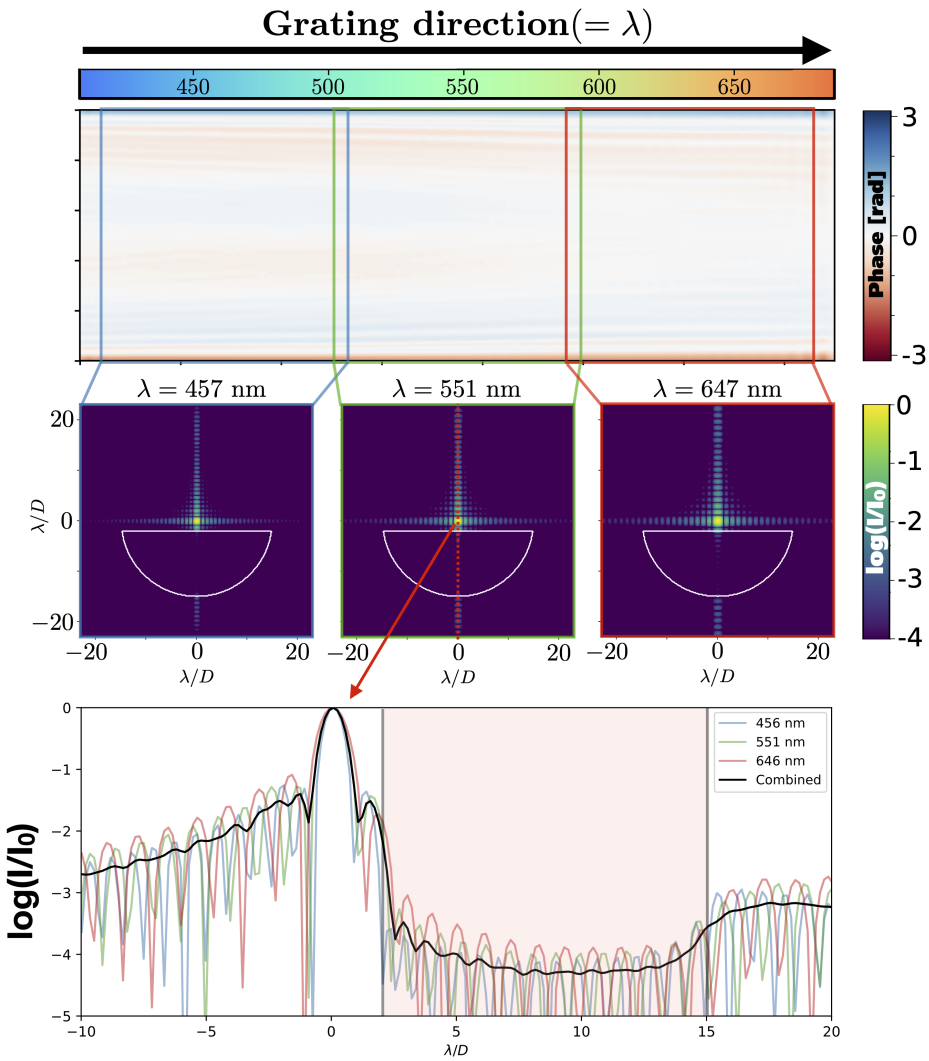


Figure 8.7: Simulation of a true-color vAPP hologram with a two-stage geometric phase hologram (GPH). The vAPP design is calculated using the same algorithm as the design presented in Fig. 8.6. The dark zone extends from 2-15 λ/D.

orders of magnitude brighter than the signal of the exoplanets. Optical components that remove the diffraction structure of on-axis sources are called coronagraphs, named after the first use of a such a device by Bernard Lyot to observe the solar corona (Lyot, 1939). A specific type of coronagraph is a phase only optic that is installed in the pupil plane called the grating vector-Apodizing Phase Plate (gvAPP) coronagraph (Snik et al., 2012; Otten et al., 2017), which is explained in more detail in chapter 2. The gvAPP creates two PSFs with a region called the dark zone, where the diffraction structure is nulled and exoplanets can be detected (Otten et al., 2014). These two PSFs are created by separating the two circular polarization states with the grating and have dark zones on opposite side, as follows from Eq. 8.1. The gvAPP coronagraph is also manufactured using the liquid-crystal technology described in this chapter. The gvAPP coronagraphs are installed in many different direct imaging instruments and have two properties that are suboptimal for exoplanet imaging. Their grating is necessary to separate the two circular polarization states from each other and the leakage term, yet the dispersion smears planet light over many pixels, introducing more noise. Secondly, the separation of circular polarization states directs half of the light in the bright-side of vAPP PSF, reducing the throughput of the planet light by a factor of 2. These two problems can be solved with a two-stage vAPP GPH.

The first grating of the GPH is unchanged and physically separates the two circular polarization states and the leakage term, just like the grating of the gvAPP. With the second stage, the phase pattern can be designed to generate two coronagraphic PSFs that have the dark zone on the same. Also, the phase pattern can be changed as function of wavelength to have the dark zone at the same physical distance from the PSF core. This increases the throughput of the star and exoplanet for the shortest wavelengths. Lastly, the polarization leakage can be blocked by a two-stage vAPP GPH coronagraph as it is physically separated from the main beams. We generated a true-color vAPP phase pattern using the GS algorithm, which is shown in Fig. 8.7. The dark zone extends from 2-15 λ/D, and has an average stellar suppression ratio of $> 10^4$ in this region. The slices of the PSF intensity through the strongest diffraction structure from a square pupil are also shown for three colors. These slices show how the diffraction structure is different as function of wavelength. Important is that the inner edge of the dark zone still shifts with wavelength, which is most likely caused by the constraint that the phase pattern must be smooth. A square pupil is also not ideal for telescopes with circular apertures, reducing the area by a factor $2/\pi \sim 0.64$. Optimizing the phase pattern for a more realistic telescope aperture is left as a future exercise.

8.5 Conclusion

We propose and demonstrate a two-stage patterned liquid-crystal element for multi-color holography. We separate colors by using the diffractive properties of a polarization grating (PG). A second element collimates the light and applies a computer generated geometric phase hologram, calculated for the separated colors individually. We show that the polarization memory of the PGs allows for efficiencies close to 100% for all visible wavelengths. In addition, we demonstrate full control over the individual colors and show that this concept can be adapted beyond RGB only to allow for wide-gamut holography. We validate with lab measurements that the multi-color liquid-crystal element produces color holograms unaffected by wavelength scaling. Lastly, we present a modified Gerchberg-Saxton algorithm for calculating white-light geometric phase holograms with dispersed pupils as input and apply it to design an improved vAPP coronagraph.

Acknowledgements

We thank Shuojia Shi for assistance in fabrication and Merck KGaA for customized materials.

Funding

The research of David S. Doelman and Frans Snik leading to these results has received funding from the European Research Council under ERC Starting Grant agreement 678194 (FALCONER).

Bibliography

Anandan, J. 1992, Nature, 360, 307

Bazargan, K. 1992, in Intl Symp on Display Holography, Vol. 1600, International Society for Optics and Photonics, 178–182

Berry, M. V. 1984, Proc. R. Soc. Lond. A, 392, 45

Bjelkhagen, H. I., & Mirlis, E. 2008, Applied optics, 47, A123

Brown, B., & Lohmann, A. 1969, IBM Journal of research and Development, 13, 160

Brown, B. R., & Lohmann, A. W. 1966, Applied Optics, 5, 967

Chang, C., Xia, J., Yang, L., et al. 2015, Applied optics, 54, 6994

Chigrinov, V. G., Kozenkov, V. M., & Kwok, H.-S. 2008, Photoalignment of liquid crystalline materials: physics and applications, Vol. 17 (John Wiley & Sons)

De Sio, L., Roberts, D. E., Liao, Z., et al. 2016, Optics express, 24, 18297

Doelman, D. S., Snik, F., Warriner, N. Z., & Escuti, M. J. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, Vol. 10400, International Society for Optics and Photonics, 104000U

Escuti, M. J., Kim, J., & Kudenov, M. W. 2016, Optics and Photonics News, 27, 22

Gerchberg, R. W., & Saxton, W. O. 1972, Optik, 35, 237

Glückstad, J. 1996, Optics Communications, 130, 225
Grier, D. G. 2012, Multi-color holographic optical trapping, Google Patents
Hasman, E., Kleiner, V., Biener, G., & Niv, A. 2003, Applied physics letters, 82, 328
Hayasaki, Y., Itoh, M., Yatagai, T., & Nishida, N. 1999, Optical review, 6, 24
Huang, K., Dong, Z., Mei, S., et al. 2016, Laser & Photonics Reviews, 10, 500
Huang, L., Chen, X., Mühlenbernd, H., et al. 2013, Nature communications, 4, 2808
Ito, T., & Okano, K. 2004, Optics Express, 12, 4320
Jesacher, A., Bernet, S., & Ritsch-Marte, M. 2014, Optics express, 22, 20530
Kim, J., Li, Y., Miskiewicz, M. N., et al. 2015, Optica, 2, 958
Komanduri, R. K., Lawler, K. F., & Escuti, M. J. 2013, Optics Express, 21, 404
Kowalczyk, A. P., Makowski, M., Ducin, I., Sypek, M., & Kolodziejczyk, A. 2018, Optics Express, 26, 17158
Kuratomi, Y., Sekiya, K., Satoh, H., et al. 2010, JOSA A, 27, 1812
Leith, E., Kozma, A., Upatnieks, J., Marks, J., & Massey, N. 1966, Applied optics, 5, 1303
Lohmann, A. W. 2008, Optics and Photonics News, 19, 36
Lyot, B. 1939, Monthly Notices of the Royal Astronomical Society, 99, 580
Makowski, M., Sypek, M., & Kolodziejczyk, A. 2008, Optics express, 16, 11618
Miskiewicz, M. N., & Escuti, M. J. 2014, Optics Express, 22, 12691
Mueller, J. B., Rubin, N. A., Devlin, R. C., Groever, B., & Capasso, F. 2017, Physical review letters, 118, 113901
Ni, X., Kildishev, A. V., & Shalaev, V. M. 2013, Nature communications, 4, 2807
Oh, C., & Escuti, M. J. 2008, Optics letters, 33, 2287
Otten, G. P., Snik, F., Kenworthy, M. A., et al. 2014, in Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation, Vol. 9151, International Society for Optics and Photonics, 91511R
Otten, G. P., Snik, F., Kenworthy, M. A., et al. 2017, The Astrophysical Journal, 834, 175
Palima, D., & Glückstad, J. 2008, Optics express, 16, 1331
Pan, J.-W., & Shih, C.-H. 2014, Optics express, 22, 6464
Pancharatnam, S. 1955in , Springer, 137–144
Peercy, M. S., & Hesselink, L. 1994, Applied optics, 33, 6811
Por, E. H., Haffert, S. Y., Radhakrishnan, V. M., et al. 2018, in Proc. SPIE, Vol. 10703, Adaptive Optics Systems VI
Shimobaba, T., & Ito, T. 2003, Optical review, 10, 339
Shiraki, A., Takada, N., Niwa, M., et al. 2009, Optics express, 17, 16038
Slinger, C., Cameron, C., & Stanley, M. 2005, Computer, 38, 46
Snik, F., Otten, G., Kenworthy, M., et al. 2012, in Modern Technologies in Space-and Ground-based Telescopes and Instrumentation II, Vol. 8450, International Society for Optics and Photonics, 84500M
Tsuchiyama, Y., & Matsushima, K. 2017, Optics express, 25, 2016
Xiang, X., Kim, J., & Escuti, M. J. 2017a, Crystals, 7, 383
—. 2018, Scientific Reports, 8, 7202
Xiang, X., Kim, J., Komanduri, R., & Escuti, M. J. 2017b, Optics Express, 25, 19298
Yaraş, F., Kang, H., & Onural, L. 2010, Journal of display technology, 6, 443

9 Nederlandstalige samenvatting

In ons zonnestelsel draaien acht planeten en vijf dwergplaneten rond de zon. Hoewel ze vanaf het aardoppervlak niet allemaal zichtbaar zijn met het blote oog, hebben ze sinds de oudheid al vele mensen geïnspireerd en hebben velen ze geobserveerd. Een van die mensen is de briljante Nederlandse wetenschapper Christiaan Huygens. Naast grote ontdekkingen in de wiskunde en natuurkunde, ontdekte hij met zijn zelfgemaakte telescoop Titan, de maan van Saturnus, en liet hij zien dat de ringen van Saturnus niet vastzaten aan Saturnus zelf en dus echte ringen waren. Zijn observaties van de andere planeten hebben hem in de laatste jaren van zijn leven aangezet om te werken aan een boekje van een meer filosofische aard genaamd *Cosmotheoros*. Hierin betoogt Christiaan Huygens, net als velen voor hem, dat er ook planeten, met manen, rond andere sterren draaien. Maar wat hem anders maakt dan vele voorgangers, is de gedachte dat wij deze planeten niet zouden kunnen zien omdat ze niet te onderscheiden zijn van hun ster. In feite dacht hij dus al na over een methode voor het ontdekken van exoplaneten, en zelfs over het karakteriseren van deze planeten. In *Cosmotheoros* betoogt hij namelijk ook dat er leven is op de andere planeten omdat er ook vloeibaar water zou zijn (met aangepaste eigenschappen om niet te bevriezen of verdampen), gebaseerd op het zien van donkere vlekken op het oppervlak van de planeten in ons zonnestelsel. We zouden Christiaan Huygens kunnen beschouwen als een pionier op het gebied van exoplaneetwetenschap.
Nu, meer dan 300 jaar na zijn dood, is het criterium dat vloeibaar water nodig is voor leven nog steeds belangrijk bij onze zoektocht naar buitenaards leven. Al is de redenatie voor de locatie van vloeibaar water omgedraaid, dus dat alleen op een bepaalde afstand van de zon water vloeibaar is en dus alleen daar leven zou kunnen ontstaan, i.e. de bewoonbare zone. Bovendien is het tegenwoordig wel mogelijk exoplaneten kunnen vinden, in tegenstelling tot de bewering van Huygens, door de extreme technologische vooruitgang en het ontstaan van andere methodes om exoplaneten te detecteren. Deze methodes zijn indirecte methodes, waarbij door precieze metingen aan de ster kan worden afgeleid dat er exoplaneten omheen moeten draaien. Sinds het vinden van de eerste twee exoplaneten in 1992 door Wolszczan and Frail en de eerste planeet rond een hoofdreeks ster in 1995 door Mayor and Queloz, zijn er al meer dan 4300 planeten gevonden. Doordat verreweg de meeste zijn ontdekt met indirecte methodes weten we vaak niet heel veel van deze planeten. Veel meer dan de omlooptijd, de diameter, de massa (en dus dichtheid) zijn vaak niet te achterhalen. Om te ontdekken of er leven is op deze planeten zullen we ze direct moeten bekijken, dus hun licht ontwarren uit het sterlicht, en dat analyseren. We zullen dus de uitdagingen die Christiaan Huygens al voorzag moeten overkomen.

Het fotograferen van een planeet

De planeet aarde is ongeveer tien miljard keer minder helder dan de zon in zichtbaar licht als je van een afstand kijkt. Ook staat de planeet dichtbij de ster,

zodat de hoekscheiding heel klein is. Als je vanaf het dichtstbijzijnde gebied waar stervorming plaatsvindt zou kijken naar de zon en de aarde, is de hoekscheiding tussen hen vier keer kleiner dan de grootte van een sesamzaadje op een broodje in Den Haag, gezien vanaf het dak van het Huygens gebouw in Leiden. Het is helaas niet zomaar mogelijk om het licht van een planeet te ontwarren van het licht van de ster. Telescopen hebben een eindige resolutie door het golfkarakter van licht, waardoor het licht van een puntbron wordt verspreid over een oppervlak. Het oplossend vermogen van een telescoop, de minimale hoekscheiding die kan worden waargenomen, hangt af van de diameter van de spiegel en de golflengte waarop je meet. Hierbij hebben grotere telescopen een groter oplossend vermogen en kan je dichterbij de ster kijken in het visueel dan in het infrarood. Voor de huidige generatie telescopen met een spiegel diameter van ongeveer 8 meter, is het waarnemen van een aardachtige planeet nog buiten bereik. Bovendien worden deze telescopen gelimiteerd door de turbulentie in de atmosfeer, die het sterlicht uitsmeert over de camera.

Om deze moeilijkheden te overkomen heeft een instrument dat afbeeldingen maakt van planeten, een *high-contrast imaging system*, meerdere complexe onderdelen. Deze onderdelen staan afgebeeld in Fig. 9.1. Het eerste onderdeel is een *adaptief optisch systeem*. Het doel van dit systeem is het opheffen van optische aberraties die worden gecreëerd door bijvoorbeeld de atmosfeer, trillingen, of polijstfouten van lenzen en spiegels. Het systeem meet de aberraties met een sensor die vervolgens, na berekeningen met de computer, een spiegel aanstuurt die kan vervormen. Voor telescopen op Aarde moet het meten en vervormen van de spiegel heel snel om de turbulentie op te heffen, tot wel 3000 keer per seconde. Voor telescopen in de ruimte hoeft de correctie niet zo snel, maar het is wel nodig voor het corrigeren van aberraties van het instrument zelf. Het nadeel van telescopen in de ruimte is dat ze door de kosten kleiner zijn en dus een kleiner oplossend vermogen hebben.

Het tweede onderdeel van het instrument is de *coronagraaf*. Het doel van de coronagraaf is het wegfilteren van het sterlicht en om zoveel mogelijk planeetlicht door te laten. Dit kan bijvoorbeeld door een kleine ronde zwarte schijf te plaatsen op de plek waar de ster staat, terwijl het licht van de planeet ernaast wordt afgebeeld. Ook zijn er coronagrafen die juist gebruik maken van het feit dat licht een golf is en het zo aanpassen dat het sterlicht zichzelf opheft (destructief interfereert). De optimale keuze voor een coronagraaf hangt af van vele factoren, zoals de hoeveel sterlicht ze wegfilteren, hoeveel planeetlicht ze doorlaten, de minimale hoekafstand waarop nog planeetlicht doorlaten en bij welke golflengtes (kleuren) ze werken.

De coronagraaf kan ook vervangen worden door een masker die de telescoop tot een interferometer maakt door een groot deel van het licht tegen te houden, behalve op de plekken waar kleine gaten zitten. Het licht van verschillende plekken op de telescoopspiegel dat wordt doorgelaten door het masker, wordt gecombineerd op dezelfde plek in een afbeelding, waarbij de samen- of tegenwerking van verscheidene golven een patroon maakt. Deze techniek heet *Apertuursynthese*. Uit het patroon van golven kan informatie worden gehaald over het object dat wordt bekeken door de telescoop. Het nadeel van deze techniek is dat je het overgrote gedeelte van het licht weggooit, maar daar krijg je twee belangrijke voordelen voor

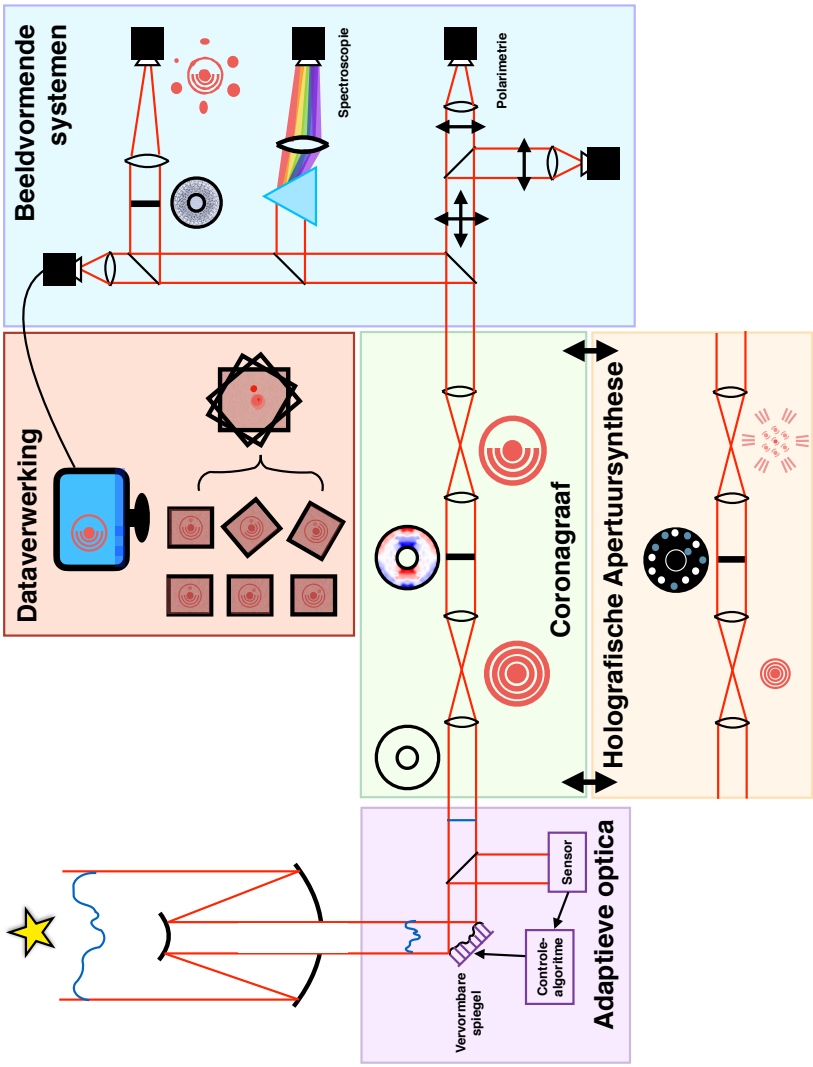


Figure 9.1: Schema van een high-contrast imaging system, meet vijf modules die gecombineerd worden om het direct fotograferen van exoplaneten mogelijk te maken.

terug. Ten eerste is de gevoeligheid voor optische aberraties en turbulentie een stuk kleiner, omdat je de faseverschillen tussen gaten kan kalibreren. Bovendien is het oplossend vermogen van de telescoop door deze techniek twee keer zo groot. Dit komt omdat je de verschuiving van de golven met een halve periode al goed kan waarnemen. Het grotere oplossend vermogen en de verminderde gevoeligheid voor optische aberraties zorgen ervoor dat deze techniek beter werkt dan een coronagraaf bij de kleinste hoekscheiding.
Het derde onderdeel bestaat uit *beeldvormende systemen*, waarmee verschillende eigenschappen van het licht van de planeten kunnen worden gemeten. Zo kunnen we aan het *spectrum*, de intensiteit van het licht per golflengte (kleur), de temperatuur afleiden van de planeet, iets zeggen over welke moleculen voorkomen in de atmosfeer en of er wolken zijn. Uit de voorkeursrichting van de trilling van het licht, *polarisatie*, kunnen we meer te weten komen over stof in de atmosfeer, of een planeet zelf snel ronddraait, of er een niet uniform wolkendek is, of dat er een grote stofschijf om de planeet draait. Deze kunnen zelfs gecombineerd worden (spectropolarimetrie), om oceanen te detecteren of misschien zelfs leven. Tot slot kunnen we met gewone afbeeldingen ook kijken naar de verandering in de tijd.
De correctie met adaptieve optica is niet perfect, waardoor de coronagraaf niet optimaal werkt en er toch sterlicht op de camera valt. Met het laatste onderdeel, *digitale dataverwerking*, kunnen we toch een deel van van dit sterlicht nog weghalen. Hierbij combineren we de afbeeldingen op een manier waarbij het overgebleven sterlicht en ruis van de camera zo veel mogelijk wordt weggehaald, en het licht van de planeet onaangetast blijft. Zo kun je bijvoorbeeld een andere gelijkaardige ster waarnemen die geen planeet heeft en deze als referentie gebruiken. Ook kun je de ster als referentie voor zichzelf gebruiken omdat de oriëntatie van een planeet ten opzichte van de ster verandert door de rotatie van de hemel door de draaiing van de aarde. Bovendien helpen spectroscopie als polarimetrie ook met het weghalen van overgebleven sterlicht. Maar dit sterlicht heeft een ander spectrum en het trilt niet in een voorkeursrichting (het is ongepolariseerd). Dus kunnen de verschillen worden gebruikt om extra te sterlicht te filteren.

Planeten karakteriseren

Met een combinatie van de voorgenoemde technieken, is het mogelijk om exoplaneten direct waar te nemen. Met de huidige technieken en telescopen zijn we gelimiteerd door de turbulentie en het oplossend vermogen, waardoor we alleen net gevormde (en dus warme) gasreuzen in de buitenste gebieden van zonnestelsels kunnen afbeelden. Maar door de snel verbeterende techniek is er al heel veel bekend over deze planeten. Een mooi voorbeeld is het 30 miljoen jaar oude systeem HR 8799, bestaande uit een ster met (minimaal) vier gasreuzen. In Fig. 9.2 staan voorbeelden van metingen uit de literatuur. De planeten staan tussen de 15 en 70 keer zo ver weg als de aarde van de zon staat, en de planeten zijn ongeveer zeven keer zo zwaar als Jupiter. De planeten zijn nog warm van de formatie, rond de 1200 graden Kelvin. Sinds de ontdekking van de buitenste drie planeten in

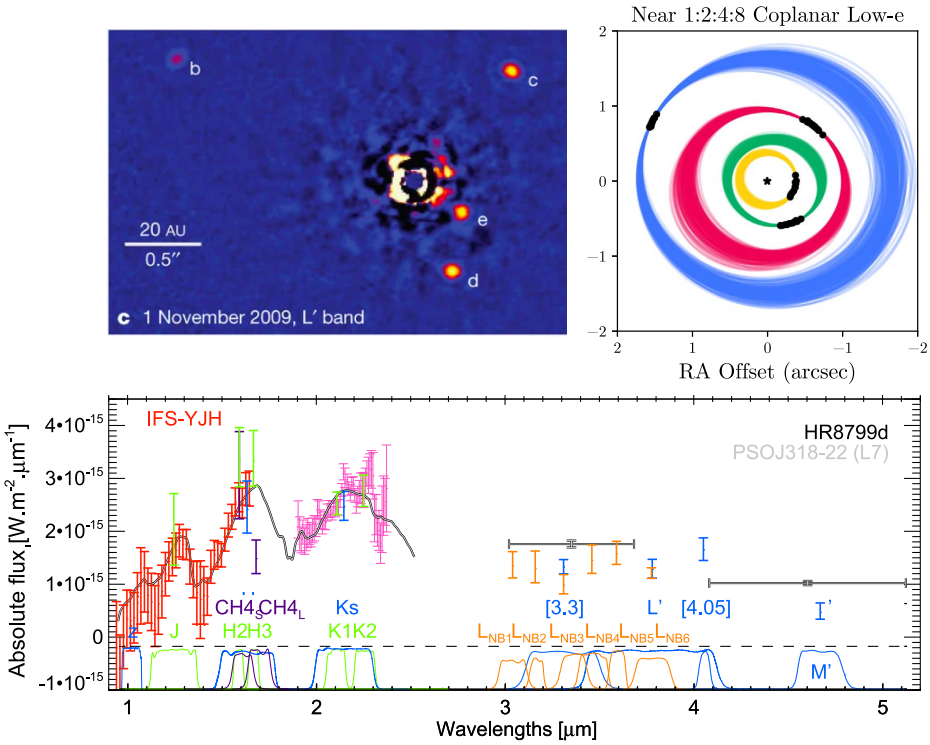


Figure 9.2: Geselecteerde resultaten van verschillende studies van het HR 8799 systeem. *Linksboven:* Afbeelding van het systeem waarbij alle vier de planeten zichtbaar zijn (Marois et al. 2010). *Rechtsboven:* Mogelijke banen van de vier planeten, gebaseerd op de zwarte meetpunten en de voorwaarde dat het systeem stabiel moet zijn (Wang et al. 2018). *Beneden:* Het spectrum van de 2e planeet van binnen, HR 8799 d, gemeten met verschillende instrumenten. In het grijs is het gemeten spectrum van een vergelijkbare vrijstaande bruine dwerg (Bonnefoy et al. 2016).

2008 en de binnenste in 2010 zijn ze heel veel geobserveerd. Daardoor is hun baan rond de ster met grote precisie bepaald. Door te kijken naar het spectrum van de planeten, is er ontdekt dat er water, methaan en koolstofmonoxide in de atmosfeer zit, en dat de atmosfeer bestaat uit een dik wolkendek. We weten uit modellen dat het wolkendek verdwijnt naarmate de planeet afkoelt, en dus is er ook veel onderzoek gaande naar het karakteriseren van het wolkendek. Uit precisie metingen van polarisatie en metingen van variabiliteit in de hoeveelheid licht die ze uitstralen zijn nog geen conclusies gekomen, behalve wat limieten voor extreme structuren. Tot slot weten we uit de grootte van de stofschijf om de ster en de verhoudingen van koolstof en zuurstof in de atmosfeer van de planeten, dat de planeten zeer waarschijnlijk niet zijn ontstaan op de plek waar ze nu bewegen, maar juist verder naar buiten.

Vloeibare-kristaltechnologie

Om nog meer te leren over planeten moeten de huidige systemen worden verbeterd. Een grote rol kan hierin worden vertaald door de integratie van de eerdergenoemde subsystemen van een high-contrast imaging system. Dit vraagt om optica waarmee je het sterlicht kan met grote precisie kan manipuleren om functionaliteit toe te voegen. Verder leren we veel over waargenomen planeten door naar hun spectrum te kijken. Dus is het belangrijk om optica te hebben die voor veel kleuren tegelijkertijd goed werken. Beide eigenschappen komen samen in optische componenten gemaakt met vloeibare-kristaltechnologie. Met deze technologie is het mogelijk om een heel gedetailleerd patroon te schrijven om het licht te manipuleren. Deze manipulatie gebeurt met de *fase* van lichtgolven, waarbij je de golven kort vertraagd ten opzichte van elkaar. Bij vloeibare-kristaltechnologie is deze vertraging, in tegenstelling to bijvoorbeeld optica van glas, niet afhankelijk van de kleur. De vertraging hangt namelijk alleen af van de lokale oriëntatie, of *geometrie*, van de vloeibare kristallen. Deze vertraging heet daarom ook wel *geometrische fase*.
En optische component met vloeibare-kristaltechnologie wordt gemaakt in twee stappen. Ten eerste wordt er een dunne laag fotopolymeriseerbaar polymeer aangebracht op een plat substraat die gevoelig zijn voor de trilrichting (polarisatie) van inkomend ultraviolette straling. Door met een uv-laser het oppervlak te scannen en tegelijkertijd de polarisatie aan te passen kan bijna elk patroon worden gemaakt. Dit patroon kan na het vastleggen niet meer worden veranderd. Bij de tweede stap worden er andere vloeibare kristallen aangebracht die zich oriënteren aan de hand van de laag eronder, die ook hard worden gemaakt met uv-straling. Deze extra laag of lagen van vloeibare kristallen bepalen de fractie van het licht dat de fase verschuiving krijgt. Deze fractie is met de huidige technologie erg hoog ($> 96\%$) voor grote spectrale bandbreedtes (al het zichtbaar licht of nabijinfrarood). Het licht wat niet deze fase verschuiving krijgt, *polarisatielekkage*, gaat door de optische component bijna alsof het er niet was. Dit licht is meestal ongewild en kan verhinderen dat we planeten vinden. Een andere eigenschap van deze optische componenten van vloeibare-kristaltechnologie is dat sterlicht zowel

een positieve fase verschuiving als een negatieve fase verschuiving krijgt. Voor veel van de toepassingen van vloeibare-kristaltechnologie voor het waarnemen van planeten die worden besproken in dit proefschrift zijn deze twee eigenschappen een limiterende factor of juist een voordeel.

Dit proefschrift

Het doel van dit proefschrift is het gebruiken van de sterke punten van vloeibare-kristaltechnologie voor het integreren van subsystemen in een high-contrast imaging system, om zo een betere karakterisatie van exoplaneten te faciliteren. Hiervoor maken we gebruik van de mogelijkheid om hoge-resolutie patronen te schrijven die kleuronafhankelijke fase aanbrengen met een hoge efficiëntie, om optische componenten te maken met verbeterde of nieuwe eigenschappen.
In hoofdstuk 2 bespreken we het ontwerpen, maken, installeren, verifiëren en eerste proefwaarnemingen met de vector-Apodizing Phase Plate (vAPP) coronagraaf voor zes verschillende telescopen. Vijf daarvan zijn geproduceerd in de laatste vier jaar. Deze vAPP coronagrafen zijn geoptimaliseerd voor verschillende golflengtes en het totale bereik van de instrumenten beslaat 550 nanometer (groen) tot 5 micrometer (infrarood). Een deel van deze coronagrafen is aangepast om optische aberraties te meten die worden geïntroduceerd door het instrument zelf. Ook zijn er soms zwakkere kopieën van het sterlicht toegevoegd om betere calibratie te doen van het licht van de planeet. Tot slot beschrijven we het ontwerp van de vAPP coronagrafen voor twee instrumenten voor de nieuwe generatie Extremely Large Telescope die momenteel wordt gebouwd en een spiegeldiameter zal krijgen van 39 meter.
In hoofdstuk 3 gebruiken wij een van deze coronagrafen, geïnstalleerd in het LMIRcam instrument in de Large Binocular Telescope, voor het waarnemen van de binnenste drie planeten van het systeem HR 8799. Dit instrument heeft een speciale spectrograaf die afbeeldingen maakt, en voor elk punt in de afbeelding ook het spectrum meet. De combinatie van het grote spectrale bereik van deze spectrograaf en de coronagraaf, die onafhankelijk van golflengte werkt, geeft een unieke kijk op het HR 8799 systeem. Uit de observaties halen we namelijk de eerste lage-resolutie middeninfrarode spectra (3.5 tot 4.1 micrometer) van deze planeten. Deze spectra zijn volledig vlak en consistent met photometrische metingen van Brα (4.05 micrometer) die worden vermeld in de literatuur. Een mogelijke verklaring voor onze metingen is dat de wolken in de atmosfeer van de drie planeten opgebroken zijn.
In hoofdstuk 4 introduceren we een nieuwe methode om de technische beperkingen die polarisatielekkage veroorzaken te overkomen. Coronagrafen die worden gemaakt met vloeibare-kristaltechnologie worden gelimiteerd door deze lekkage omdat deze helderder is dan het licht van planeten. De nieuwe methode combineert het patroon van een coronagraaf met een extra traliepatroon, bijeengevoegd met nog één of meerdere substraten met ook een traliepatroon. Met deze methode kan de polarisatielekkage met meerdere orders van grootte verkleind worden

over een groot spectraal bereik. In het lab laten we zien dat voor een simpele twee-tralie combinatie de lekkage kunnen verkleinen met drie orders van grootte voor een golflengte van 532 nm en een factor 10 over een spectrale bandbreedte van 133 nm. Verder laten we met simulaties zien dat de lekkage drie ordes van grootte kleiner kan worden met een complexere twee-tralie coronagraaf voor een spectraal bereik tussen de 1-2.5 micrometer, of zelfs zes orders van grootte voor een drie-tralie ontwerp.
In hoofdstuk 5 en 6 laten we zien dat de vloeibare-kristaltechnologie ook gebruikt kan worden voor het verbeteren van de maskers die worden gebruikt voor apertuursynthese. Doordat gaten in het masker op unieke plekken moeten zitten, passen er maar een gelimiteerd aantal gaten op het masker die daardoor weinig licht doorlaat. In hoofdstuk 5 presenteren wij het nieuwe idee van holografische apertuursynthese. Door een masker te maken dat meer gaten heeft die worden afgebeeld op een andere plek op de camera met een vloeibaar kristaloptiek, blijven de combinaties uniek terwijl je wel meer licht doorlaat. Deze nieuwe combinaties van gaten worden afgebeeld in holografische interferogrammen waarvan de plek van afbeelden verandert als functie van golflengte. Hierdoor kunnen we lage-resolutie spectroscopie doen zonder een aparte spectrograaf nodig te hebben. In hoofdstuk 6 bespreken we het ontwerp en fabricage van twee holografische apertuursynthese maskers voor de OSIRIS imager van de Keck telescoop. Bovendien beschrijven we de resultaten van de kalibratie in het lab en de eerste metingen van de dubbelster HDS 1507. Uit deze metingen halen we een hoekscheiding van 120.9 $\pm$ 0.5 milliboogseconde, die constant is als functie van golflengte.
In hoofdstuk 7 presenteren we de verbeterde versie van de Zernike wavefront sensor (ZWFS). Deze sensor is de meest efficiënte sensor voor het meten van optische aberraties met grote precisie. Onze versie, de vector-Zernike wavefront sensor (vZWFS) die wordt gemaakt met vloeibare-kristaltechnologie, is in staat om niet alleen aberraties in het golffront (fase) te meten, maar tegelijkertijd ook amplitudeverschillen. Door het tegelijkertijd meten van zowel de fase als amplitude heeft de sensor geen last van veranderingen in de amplitude, iets wat bij de gewone ZWFS tot foute metingen kan leiden. Bovendien werkt de vZWFS sensor ook voor een groter spectraal bereik, waardoor er meer licht kan worden gebruikt voor de meting.
In hoofdstuk 8 introduceren we een nieuwe methode om met vloeibare-kristaltechnologie meerkleurige hologrammen te maken. Dit is mogelijk door het licht te scheiden per kleur met een polarisatie tralie en vervolgens per kleur het licht aan te passen met een patroon dat is gecombineerd met dezelfde tralie. Omdat het patroon anders is per kleur kunnen we corrigeren voor de diffractie, die patronen groter maakt voor grotere golflengten, en zelfs elke kleur een compleet andere afbeelding laten maken. We laten tot slot ook zien dat deze methode kan worden toegepast om vAPP coronagrafen te maken die twee keer zo efficiënt zijn.

List of Publications

Refereed Publications

1. *Simultaneous phase and amplitude aberration sensing with a liquid-crystal vector-Zernike phase mask*
   Doelman, D. S.; Fagginger Auer, F.; Escuti, M. J.; Snik, F.
   Optics letters, 44(1), 17-20 (2019)

2. *Multi-color holography with a two-stage patterned liquid-crystal element*
   Doelman, D. S.; Escuti, M. J.; Snik, F.
   Optical Materials Express 9 (3), 1246-1256 (2019)

3. *Minimizing the polarization leakage of geometric-phase coronagraphs with multiple grating pattern combinations*
   Doelman, D. S.; Por, E. H.; Ruane, G.; Escuti, M. J.; Snik, F.
   PASP, 132(1010), 045002 (2020)

4. *First light of a holographic aperture mask: Observation at the Keck OSIRIS Imager*
   Doelman, D. S.; Wardenier, J. P.; Tuthill, P.; Fitzgerald, M. P.; Lyke, J.; Sallum, S.; Norris, B.; Warriner, N. Z.; Keller, C. U.; Escuti, M. J.; Snik, F.
   Accepted for publication in A&A (2021)

5. *The vector-apodizing phase plate coronagraph: design, current performance, and future development*
   Doelman, D. S.; Snik, F.; Por, E. H.; Bos, S. P.; Otten, G. P. P. L.; Kenworthy, M.; Haffert, S. Y.; Wilby, M.; Bohn, A. J.; Sutlieff, B. J.; Miller, K.; Ouellet, M.; de Boer, J.; Birkby, J. L.; Keller, C. U.; GPL team; MagAO team; MagAO-X team; SCExAO team; HiCIBaS team; LMIRcam/ALES team; ERIS team; MICADO team; METIS team
   Accepted for publication in Applied Optics (2021)

6. *Focal-plane wavefront sensing with the vector-Apodizing Phase Plate*
   Bos, S. P.; **Doelman, D. S.**; Lozi, J.; Guyon, O.; Keller, C. U.; Miller, K. L.; Jovanovic, N.; Martinache, F.; Snik, F.
   A&A, 632, A48 (2019)

7. *Detection of polarization neutral points in observations of the combined corona and sky during the 21 August 2017 total solar eclipse*
   Snik, F.; Bos, S. P.; Brackenhoff, S. A.; **Doelman, D. S.**; Por, E. H.; Bettonvil, F.; Rodenhuis, M.; Vorobiev, D.; Eshelman, L. M.; Shaw, J. A.
   Applied Optics, 59 (21), F71-F77

8. *Spatial linear dark field control on Subaru/SCExAO-Maintaining high contrast with a vAPP coronagraph*
   Miller, K. L.; Bos, S. P.; Lozi, J.; Guyon, O.; **Doelman, D. S.**; Vievard, S.; Sahoo, A.; Deo, V.; Jovanovic, N.; Martinache, F.; Snik, F.; Currie, T.
   A&A, 646, A145

9. *The Single-mode Complex Amplitude Refinement (SCAR) coronagraph II. Lab verification, and toward the characterization of Proxima b*
   Haffert, S. Y.; Por, E. H.; Keller, C. U.; Kenworthy, M. A.; **Doelman, D. S.**; Snik, F.; Escuti, M. J.
   A&A, 635, A56 (2020)

10. *Spatial linear dark field control and holographic modal wavefront sensing with a vAPP coronagraph on MagAO-X*
    Miller, K. ; Males, J. R.; Guyon, O.; Close, L. M.; **Doelman, D. S.**; Snik, F.; Por, E. H.; Wilby, M. J.; Keller, C.; Bohlman, C.; Van Gorkom, K.; Rodack, A.; Knight, J.; Lumbres, J.; Bos, S. P.; Jovanovic, N.
    JATIS, V5 (4), 049004 (2019)

11. *First Images of the Protoplanetary Disk around PDS 201*
    Wagner, K.; Stone, J.; Dong, R.; Ertel, S.; Apai, D.; **Doelman, D.S.**; Bohn, A.; Najita, J.; Brittain, S.; Kenworthy, M.; Keppler, M.; Webster, R.; Mailhot, E.; Snik, F.
    The Astronomical Journal 159 (6), 252

12. *METIS high-contrast imaging: design and expected performance*
    Carlomagno, B.; Delacroix, C.; Absil, O.; Cantalloube, F.; de Xivry, G. O. ; Pathak, P.; Agocs, T.; Bertram, T.; Brandl, B.; Burtscher, L.; **Doelman, D. S.**; Feldt, M.; Glauser, A.; Hippler, S.; Kenworthy, M.; Por, E. H.; Snik, F.; Stuik, R.; van Boekel, R.
    JATIS, V6 (4), 049801 (2020)

13. *Cryogenic characterization of the grating vector Apodizing Phase Plate coronagraph for the Enhanced Resolution Imager and Spectrograph at the Very Large Telescope*
    Boehle, A.; **Doelman, D. S.**; Konrad, B. S.; Snik, F.; Glauser, A. M.; Warriner, N. Z.; Shi, S.; Escuti, M. J.; Kenworthy, M. A.; Quanz, S.
    Submitted to JATIS

Selected Conference Proceedings

1. *Patterned liquid-crystal optics for broadband coronagraphy and wavefront sensing*
   Doelman, D. S.; Snik, F.; Warriner, N. Z.; Escuti, M. J.
   SPIE 10400, 104000U (2017)

2. *Multiplexed holographic aperture masking with liquid-crystal geometric phase masks*
   Doelman, D. S.; Tuthill, P.; Norris, B.; Wilby, M. J.; Por, E. H.; Keller, C. U.; Escuti, M. J.; Snik, F.
   SPIE 10701, 107010T (2018)

3. *New concepts in vector-apodizing phase plate coronagraphy*
   Bos, S. P.; **Doelman, D. S.**; Miller, Kelsey L.; Snik, Frans
   SPIE 11448, 114483W (2020)

4. *Fully broadband vAPP coronagraphs enabling polarimetric high contrast imaging* Bos, S. P.; **Doelman, D. S.**; de Boer, J.s; Por, E. H.; Norris, B.; Escuti, M. J.; Snik, F.
   SPIE 10706, 107065M (2018)

5. *A snapshot full-Stokes spectropolarimeter for detecting life on Earth*
   Snik, F.; Keller, C. U.; **Doelman, D. S.**; Kühn, J.; Patty, C. H. L.; Hoeijmakers, H. J.; Pallichadath, V.; Stam, D. M.; Pommerol, A.; Poch, O.; Demory, B.
   SPIE 11132, 111320A (2019)

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

7. *MagAO-X first light*
   Males, J. R.; Close, L. M.; Guyon, O; Hedglen, A. D.; Van Gorkom, K.; Long, J. D.; Kautz, M.; Lumbres, J. Schatz, L.; Rodack, A.; Miller, K.; **Doelman, D. S.**; Snik, F.; Bos, S.; Knight, J. M.; Morzinski, K.; Gasho, V.; Keller, C.U.; Haffert, S. Y.; Pearce, L.
   SPIE 11448, 114484L (2020)

8. *On-sky results of the Leiden EXoplanet Instrument (LEXI)*
   Haffert, S. Y.; Wilby, M. J.; Keller, C. U.; Snellen, I. A. G.; **Doelman, D. S.**; Por, E. H.; van Kooten, M.; Bos, S. P.; Wardenier, J.
   SPIE 10703, 1070323 (2018)

9. *HiCIBaS: A precursor mission for high contrast imaging balloon systems*
   Marchis, F.; Thibault, S.; Côté, O.; Brousseau, D.; Allain, G.; Lord, M. P.; Ouellet, M.; Patel, D.; Vallée, C.; Belikov, R.; Bendek, E.; Blain, C.; Bradley, C.; **Doelman, D.S.**; Daigle, O.; Doyon, R.; Grandmont, F. J.; Helmbrecht, M.; Kenworthy, M.; Lafrenière, D.; Marois, C.; Montminy, S.; Snik, F.; de Jonge, C.; Vasisht, G.; Veran, J. P.; Vincent, P.
   AGU Fall Meeting, P41C-3747 (2018)

10. *Cryogenic characterization of the grating vector APP coronagraph for the upcoming ERIS instrument at the VLT*
    Boehle, A.; Glauser, A. M.; Kenworthy, M. A.; Snik, F.; **Doelman, D.S.**; Quanz, S. P.; Meyer, M. R.
    SPIE 10702, 107023Y (2018)

11. *SCExAO, an instrument with a dual purpose: perform cutting-edge science and develop new technologies*
    Lozi, J.; Guyon, O.; Jovanovic, N.; Goebel, S.; Pathak, P.; Skaf, N.; Sahoo, A.; Norris, B.; Martinache, F.; N'Diaye, M.; Mazin, B.; Walter, A. B.; Tuthill, P.; Kudo, T.; Kawahara, H.; Kotani, T.; Ireland, M.; Cvetojevic, N.; Huby, E.; Lacour, S.; Vievard, S.; Groff, T. D.; Chilcote, J. K.; Kasdin, J.; Knight, J.; Snik, F.; **Doelman, D.S.**; Minowa, Y.; Clergeon, C.; Takato, N.; Tamura, M.; Currie, T.; Takami, H.; Hayashi, M.

List of Selected Presentations

1. *Patterned liquid-crystal optics for broadband coronagraphy and wavefront sensing*
   SPIE Optical Engineering + Applications, August 2017, San Diego, United States (oral)

2. *The world of vAPP: creating dark holes in PSFs all over the world*
   Netherlands Astronomy Conference, May 2018, Nijmegen, The Netherlands (poster, 2nd prize)

3. *Advanced wavefront sensing techniques enabled by patterned liquid-crystal technology*
   SPHERE upgrade workshop, May 2018, Grenoble, France (oral)

4. *Multiplexed holographic aperture masking with liquid-crystal geometric phase masks*
   SPIE Astronomical Telescopes & Instrumentation, June 2018, Austin, United States (oral)

5. *First light for the vAPP on SCExAO/CHARIS*
   SPIE Astronomical Telescopes & Instrumentation, June 2018, Austin, United States (oral)

6. *The world of vAPP: creating dark holes in PSFs all over the world*
   Exoplanets II, July 2018 , Cambridge, United Kingdom (poster)

7. *Liquid crystals for high-contrast imaging*
   NYRIA, Oct 2018, Leiden, The Netherlands (oral, LOC)

8. *The vector-Apodizing Phase Plate coronagraph*
   Spirit of Lyot, Oct 2019. Tokyo, Japan (oral)

Curriculum Vitae

David Sebastiaan Doelman was born on July 30th, 1993 in Amsterdam (The Netherlands) to Jolette Drogendijk and Gerard Doelman, only a few minutes after his sister (Eline) and brother (Matthijs). Together with his other brothers (Christiaan and Michiel, 1995), he grew up in Haarlem and later Heemstede. He came into contact with natural sciences and space in part through his grandfather, who intrigued him with his large collection of model spacecraft and later the science magazines he would bring. This interest expressed itself at College Hageveld, where he graduated with a specialization in the natural sciences in 2011.

In that same year he started a double Bachelor in Physics and Astronomy at Leiden University. He specifically enjoyed the practica of physics, the topics of astronomy, and observing with the Isaac Newton Telescope at La Palma. His interest in optics was sparked by his Bachelor research project, where he worked on shaping an optical beam profile to create ring-shaped plasma under the supervision of Dirk Bouwmeester and Chris Smiet. He graduated both Bachelor's in 2014.

He continued with a Master in Astronomy specializing in astronomical instrumentation, taking courses both in Leiden and Delft. During the Master's study he first worked on determining dust and gas distributions in two protoplanetary disks to look for radial drift using cycle-0 data of the *Atacama Large Millimeter Array* (ALMA), under supervision of Michiel Hogerheijde. For his second research project he optimized vector-Apodizing Phase Plate (vAPP) coronagraph designs using simulated annealing under supervision of Christoph Keller and Michael Wilby. One design for the *Large Binocular Telescope* (LBT) was selected for fabrication and installed in the LMIRcam instrument. He received his Master's degree *Cum Laude* in 2016.

In September 2016, he started his PhD under supervision of Frans Snik and Christoph Keller, where he developed novel concepts with liquid-crystal technology for diffraction limited optical and infrared high-contrast imaging instruments, to find and characterize extra-solar planets. He designed liquid-crystal optics currently used, or soon implemented, in 8 different instruments at ground-based telescopes. For most of these projects he was heavily involved in manufacturing, installation, the calibration, and on-sky testing. After observations with the vAPP at the LBT, he also led the data reduction efforts and retrieved the first-ever spectra of the HR 8799 planets in the thermal infrared. Finally, he worked on a liquid-crystal wavefront sensor and coronagraph for the next generation space-based high-contrast imaging instruments which soon will be tested at the Jet Propulsion Laboratory. He presented his work at many workshops and international conferences and during public talks.

He started the next phase as a postdoctoral researcher in the group of Frans Snik to further develop the liquid-crystal technology, focussing on space-based applications in collaboration with SRON.

Acknowledgements

It is just a phase. And now it is over. The result is this thesis, which can be seen as a token of a transformation into an independent researcher. However, if this thesis shows anything, it is that research requires collaboration. I have had the pleasure to work with many amazing people on diverse projects, in addition to enjoying the full support of friends, colleagues, and family. I would like to thank you personally.

First, I would like to thank Frans Snik. Your endless enthusiasm and the many opportunities you provided always encouraged me to make the most of my PhD. I would also like to thank Christoph Keller. You always made time to discuss issues and put me back on track. Both of you have also given me a lot of freedom to pursue projects and my own ideas, which helped me becoming an independent researcher.

Luckily, the PhD experience is a shared one, and I could not have done it without all my colleagues from the instrumentation group. Steven, Maaike, Rob, Sebastiaan, Emiel, Alex, Mireille, Kelsey, Olivier, Patrick and many others, you have made my PhD a wonderful adventure. I am grateful for all the knowledge, support, conversation, and inspiration you have given me during these years. I particularly enjoyed our many trips together, where, after the serious part, there was always plenty of time for exploration, relaxation, and enjoyment. To the students I supervised, Fedde and Joost, your energy was highly infectious and your work was excellent. I am happy to see you both pursuing an academic career, no doubt with great success. During the pandemic I also got to know some of the new members of our research group. Willeke, Floor, and Elina, it was a delight to have you as (safely distanced) company during the last few months of my PhD.

I find the high-contrast imaging community to be a welcoming one and I have been fortunate enough to collaborate with many incredible people from many institutes all around the globe. I can't name you all personally, so I want to thank the SCExAO, MagAO-X, HiCIBaS, LMIRcam, OSIRIS, ERIS, MICADO, and METIS teams. You were extremely helpful and accommodating, and even though the Zoom meetings were at dreadful times some days, I thoroughly enjoyed working together. In particular, I want to thank the people from the Geometric-Phase Photonics Lab, Michael Escuti, Shuojia Shi, Zane Warriner, and, Kathryn Hornburg. Our team-up has been a fruitful one, not in the least due to your generosity and hard work. I have fond memories of the two longer visits to Raleigh, testing all kind of crazy ideas with the liquid-crystal plates, experiencing Halloween, getting an electric shock from the lab door, and getting beaten at airhockey in the arcade hall. Here, I also want to thank Lucas and Steven, who also kept me company during these visits to Raleigh. The nights playing bingo at Ba-Da Wings with cajun fries and PBR were legendary.

I also want to thank my friends from outside of the research group, some of whom I have known for a very long time. Stijn, Leindert, and Leon, we have travelled a long road together. Since our first lecture as first year students, we have shared notes, offices, beers and friendship. Bavo, Kasper, and David, you inspire me ever since we met in office 101. Kirsty, you get a special thanks from my cats, yet I want

to thank you for enduring the technobabble at the morning coffees and forcing me to watch sports and drink beer at 6am in the morning. During my PhD I also helped organize a NYRIA workshop, and four of the observatory staff members went out of their way to help us. Monica, Marjan, Hafize, and Els, without your help we never would have been able to make it work.
Tot slot wil ik nog een aantal mensen bedanken die mijn leven buiten mijn promotie kleur hebben gegeven. Lieve leden van het dispuut Terra F., ik heb een geweldige tijd met jullie beleefd. Mede door jullie gezelligheid, alle activiteiten en schelvispekel heb ik mij geen moment verveeld. Ook nu ik geen PhD-student meer ben zal er nog minder stuko in de stukoburgers zitten. Maar, Kevin, Dionne, Lisanne en Martijn, ons gouden leventje is nog niet voorbij. Graag wil ik ook mijn schoonfamilie bedanken. Eric, Hanja en Roy, vanaf het eerste moment ben ik altijd welkom bij jullie. Ik heb jullie leren kennen als gedreven mensen met een hechte band, en ik voel me bevoorrecht om een onderdeel te zijn van jullie gezin.
Aan het eind van dit avontuur wil ik ook de ruimte maken om mijn ouders, broers en zus te bedanken. Mam, Pap, jullie hebben ons altijd gestimuleerd en geholpen ons onze dromen na te jagen, wat met vijf kinderen vast niet altijd makkelijk was. Jullie hebben mij altijd gesteund en waren er altijd voor me als ik het moeilijk had. Ook hebben jullie mij geleerd dat je door hardwerken en vastberadenheid alles kan bereiken. Jullie zijn een inspiratie.
Als onderdeel van een drieling is het moeilijk te omschrijven hoeveel jullie voor mij hebben betekend. Eline, ik wil je bedanken voor je zorgzaamheid maar ook alle gekkigheid die we hebben beleefd in bij SSR en daarbuiten. Matthijs, onze eeuwige competitie haalt het beste in mij naar boven. Nu ben jij weer aan zet. Christiaan en Michiel, we hebben het elkaar niet altijd even makkelijk gemaakt, maar weet dat ik ontzettend trots op jullie ben.
Als laatst richt ik het woord tot jou, lieve Nathalie. Dank je voor het geven van de ruimte die alle projecten van mij vroegen of leken te vragen. Dank je voor de rust die jij hebt gebracht als mijn hoofd overliep met ideeën, deadlines en stress. En dank je voor je eeuwige steun en liefde, die samen het fundament vormen van dit boekje.