

Universiteit
Leiden
The Netherlands

Novel approaches for direct exoplanet imaging: Theory, simulations and experiments

Por, E.H.

Citation
Por, E. H. (2020, December 11). *Novel approaches for direct exoplanet imaging: Theory, simulations and experiments*. Retrieved from https://hdl.handle.net/1887/138516

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/1887/138516

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

Cover Page



Universiteit Leiden



The handle http://hdl.handle.net/1887/138516 holds various files of this Leiden University dissertation.

Author: Por, E.H.
Title: Novel approaches for direct exoplanet imaging: Theory, simulations and experiments
Issue date: 2020-12-11

Novel approaches for direct exoplanet imaging

theory, simulations and experiments

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op vrijdag 11 december 2020
klokke 12:30 uur

door

Emiel Hugo Por

geboren te Zoetermeer, Nederland
in 1992

Promotor: prof.dr. Christoph Keller
Co-promotor: dr. Matthew Kenworthy

Promotiecommissie:

prof. dr. Huub Röttgering	Universiteit Leiden
prof. dr. Ignas Snellen	Universiteit Leiden
prof. dr. Simon Portegies Zwart	Universiteit Leiden
dr. Rebecca Jensen-Clem	University of California, Santa Cruz
dr. Olivier Guyon	University of Arizona

Cover design: the designed pupil-plane masks for (left) an apodizing phase plate coronagraph with an annular dark zone for the SCExAO/Subaru instrument, and (right) an apodized-pupil Lyot coronagraph for the LUVOIR-A telescope. Both masks were optimized using algorithms developed during my PhD.

Keywords: coronagraph, wavefront sensing, optimization, Python
ISBN: 978-94-6361-500-6
An electronic copy of this thesis can be found at https://openaccess.leidenuniv.nl
© Emiel H. Por, 2020

To Mom and Dad,

for without your advice,
your patience,
and your love,
I would not be who I am today.

Contents

1 Introduction **1**
- 1.1 Star and planet formation . . . 3
  - 1.1.1 From molecular cloud to young stellar system . . . 3
  - 1.1.2 Pathways for planet formation . . . 5
  - 1.1.3 Atmospheric composition and biomarkers . . . 6
- 1.2 Observational techniques for finding planets . . . 6
  - 1.2.1 Transit photometry . . . 7
  - 1.2.2 Radial velocity . . . 9
  - 1.2.3 Astrometry . . . 9
  - 1.2.4 Direct imaging . . . 10
- 1.3 Anatomy of a high-contrast imaging instrument . . . 10
  - 1.3.1 The coronagraph . . . 14
  - 1.3.2 The wavefront control system . . . 31
  - 1.3.3 Image post processing . . . 34
- 1.4 This thesis . . . 35
- 1.5 Future outlook . . . 37

2 Optimal design of apodizing phase plate coronagraphs **49**
- 2.1 Introduction . . . 50
- 2.2 Linearization, discretization and correction . . . 51
  - 2.2.1 Linearization . . . 51
  - 2.2.2 Discretization . . . 53
  - 2.2.3 Speed improvements . . . 53
  - 2.2.4 Tilt correction . . . 55
- 2.3 Case studies . . . 56
  - 2.3.1 D-shaped dark zones . . . 56
  - 2.3.2 Annular dark zones . . . 59
- 2.4 Conclusions . . . 65

3 The SCAR coronagraph I **67**
- 3.1 Introduction . . . 68
- 3.2 Modal filtering using single-mode fibers . . . 71
  - 3.2.1 Nulling in single-mode fibers . . . 71
  - 3.2.2 Single-mode fiber arrays using microlenses . . . 73

3.3 Coronagraphy with a single-mode fiber array 76
3.3.1 Conventional coronagraphy 76
3.3.2 Direct pupil-plane phase mask optimization 77
3.4 Single-mode fiber coronagraph properties 84
3.4.1 Fiber mode field diameter 84
3.4.2 Throughput and inner working angle 84
3.4.3 Spectral bandwidth 88
3.4.4 Tip-tilt sensitivity and stellar diameter 88
3.4.5 Sensitivity to other aberrations 88
3.5 Comparison to the vortex coronagraph 91
3.6 Conclusion 97

4 The SCAR coronagraph II 101
4.1 Introduction 102
4.2 Optical setup details and first results 104
4.2.1 Lab setup description 104
4.2.2 Fiber alignment procedure 106
4.2.3 Apodizing phase plate designs 106
4.2.4 Liquid crystal plate 108
4.2.5 Lab setup results 108
4.3 Tolerance simulation analysis 116
4.3.1 Fiber alignment tolerance 117
4.3.2 MLA surface 118
4.3.3 Fiber mode shape 119
4.3.4 FIU Monte Carlo analysis 120
4.4 Conclusions 122

5 High Contrast Imaging for Python (HCIPy) 125
5.1 Introduction 126
5.2 Core functionality 127
5.2.1 Coords, Grids and Fields 128
5.2.2 Field generators and visualization 129
5.2.3 Fourier transforms 129
5.2.4 Mode bases 130
5.3 Optical systems 131
5.4 Adaptive optics 132
5.4.1 Atmospheric modeling 132
5.4.2 Wavefront sensing 133
5.4.3 Wavefront control 134
5.5 Coronagraphy 136

5.6 Miscellaneous . . . 137
5.6.1 Polarization . . . 137
5.6.2 Performance . . . 137
5.7 Conclusions . . . 141
5.7.1 Overview . . . 141
5.7.2 Future plans . . . 141

6 Origin of the asymmetry of the wind-driven halo 145
6.1 Introduction . . . 146
6.2 Description of the observed asymmetry . . . 146
6.3 Interference between scintillation and temporal error . . . 149
6.4 Simulations of the effect . . . 153
6.5 Conclusions . . . 155

7 Phase-apodized-pupil Lyot coronagraphs 161
7.1 Introduction . . . 162
7.2 Overview of the numerical optimization problem . . . 164
7.2.1 Problem definition . . . 164
7.2.2 Simplification and convexification . . . 168
7.2.3 Symmetry considerations . . . 170
7.2.4 Tip-tilt correction for one-sided dark zones . . . 170
7.3 Parameter study for point-symmetric dark zones . . . 171
7.4 Parameter study for one-sided dark zones . . . 174
7.4.1 Contrast, inner working angle and central obscuration ratio . . . 177
7.4.2 Achromatization and residual atmospheric dispersion 178
7.5 Case studies for VLT/SPHERE and LUVOIR-A . . . 179
7.5.1 VLT/SPHERE . . . 179
7.5.2 LUVOIR-A . . . 183
7.5.3 Performance . . . 183
7.6 Conclusions . . . 187
7.7 Appendix: The full optimization problem . . . 188

8 First laboratory demonstration of the PAPLC 191
8.1 Introduction . . . 192
8.2 PAPLC with deformable mirror . . . 194
8.2.1 Monochromatic performance . . . 194
8.2.2 Broadband performance . . . 195
8.3 Simultaneous high-order wavefront sensing . . . 197
8.3.1 Principle . . . 197

8.3.2 Empirical modal response and reconstruction 201
8.3.3 Sensitivity to photon noise 203
8.4 Laboratory demonstration 204
8.4.1 The THD2 bench 204
8.4.2 Coronagraphic performance 207
8.4.3 Phase-retrieval wavefront sensor 212
8.5 Conclusions 222

English summary **225**

Nederlandse samenvatting **229**

Curriculum Vitae **233**

Acknowledgments **235**

Chapter 1

Introduction

> *"It is as unlikely that a single world should arise in the infinite as that one single ear of corn should grow on a large plain."*
> – Metrodorus of Chios, fourth century BC

The question of the existence of life outside of Earth has long fascinated humanity. The ancient Greeks, notably Democritus and Epicurus, entertained the idea of cosmic plurality, the existence other life-bearing "worlds". The term "world" here refers more to the modern-day notion of parallel universes than that of other planetary systems. With the strong opposition of Plato and Aristotle, who argued for a perfect, unique and stationary world, the debate was largely decided against cosmic plurality in the Western world.

This stance was challenged in the Renaissance after Nicolaus Copernicus proposed the heliocentric model. This discovery demonstrated that the Earth was not the center of the universe and that the planets were not that much different from Earth. This led to the hypothesis that these other celestial bodies could be inhabited. In fact, there were two possibilities: either the Earth was somehow unique, a hypothesis that evolved into the modern-day *Rare Earth hypothesis*, or that life was common and that every other planet was inhabited. Nowadays it may be considered amusing that Sir William Herschel was a proponent of the latter theory, arguing that, even though the Moon seemed to have little to no atmosphere and no large seas on its surface, *"we can certainly not object to the conveniences afforded by the moon, if those that are to inhabit its regions are fitted to their conditions, as well as we on this globe are to ours"* and that *"I believe the analogies that have been mentioned are fully sufficient to establish the high probability of the moon's being inhabited like the earth"*.

While extra-solar planets, planets that orbit stars other than our own Sun, had long been hypothesised, the first discovery of an exoplanet puzzled scientists (Wolszczan & Frail, 1992). The two exoplanets orbit a pulsar, the rapidly rotating collapsed core of a giant star after a supernova. At the time, there was no known pathway for planets to survive a supernova. A third object orbiting the same pulsar was found two years later (Wolszczan, 1994). Only a few other *pulsar planets* have been found since (Rodin, 2019; Suleymanova & Rodin, 2014).

Only three years after this first discovery, an exoplanet was found around a Sun-like star (Mayor & Queloz, 1995). The mass of this planet, 51 Pegasi b, is about half of that of Jupiter, and it orbits its star in just over four days, making it about seven times closer to its host star than Mercury to the Sun. This small orbital distance makes the planets temperature a scorching 1000 °C. Additionally, the planet is likely *tidally locked* with its star, a phenomenon that makes one side of the planet always face its host star, similar to how the Moon always shows the same hemisphere to Earth. At that time, there was no known formation process that could produce such a massive planet that close to its host star. Since this discovery, many other exoplanets like 51 Pegasi b have been found, proving that they are quite common (e.g. see Fischer et al., 2016, for a review). This class of exoplanet is now collectively known as *hot Jupiters*.

Having demonstrated that finding exoplanets was possible with the technology at the time, astronomers embarked on a journey to find more. In the last two decades we have found out that most stellar systems contain multiple exoplanets, and on average every system has one planet orbiting in the habitable zone, where water can be liquid on the planet's surface (Dawson & Johnson, 2018a; Udry & Santos, 2007; Winn & Fabrycky, 2015). Most striking is the diversity of discovered planets. We have seen super-Earths, planets that have masses a few times that of our Earth composed of iron and silicates. We have seen sub-Neptunes, planets that have radii smaller than Neptune yet more massive (Luque et al., 2019). We have seen super-puff planets, which have radii larger than Neptune yet weigh only a few times the mass of our Earth (Libby-Roberts et al., 2020), though these might actually be smaller planets with giant ring systems that make them appear larger than they actually are (Piro & Vissapragada, 2020). We even have seen planets that seem to be disintegrating or evaporating, shedding large amounts of dust in their wake (Rappaport et al., 2012; van Lieshout et al., 2016).

The topic of this thesis is focused on the technology behind these discoveries, in particular for the method of *direct imaging*. This method attempts to image the exoplanet directly rather than infer its existence indirectly. This has the advantage of enabling spectroscopic characterization of the planets atmosphere, allowing us to infer its chemical composition and search for biomarkers, features that indicate the presence of extraterrestrial life on its surface. Direct imaging requires extensive technical knowledge of the behaviour of light, complex dynamic optical systems to mold it, and advanced computational tools to design and operate this optical system,

and to analyze the data produced by it.

In this introductory chapter, I give a brief outline of our current understanding the formation of planetary systems, then lead into the techniques using for detecting exoplanets. Of particular importance to this thesis is *direct imaging*, which deserves a more lengthy introduction. I conclude with an outline of how I contributed to the techniques used in direct imaging, what future directions are likely to progress this field and answer our current questions for exoplanet formation.

1.1 Star and planet formation

Star and planet formation is a complex process that covers a wide range of length and time scales. As star formation is inherently linked to planet formation, I will first discuss star formation before transitioning to planet formation. Figure 1.1 shows schematically an overview of the different stages of the entire process.

1.1.1 From molecular cloud to young stellar system

Molecular clouds consist of mostly hydrogen, helium, carbon monoxide, nitrogen gas and small silicate dust particles. Due to small density fluctuations, these clouds will become gravitationally unstable and contract into multiple cores, each with a mass of a few solar masses. These cores will slowly collapse, and form into proto-stars (Shu, 1977). Because of conservation of angular momentum, the rotation of the core will gradually speed up during its collapse. Small random motions cause a deviation from a perfect spherical collapse, resulting in a non-zero amount of angular momentum within the cloud. The cloud collapses and as the gas kinematically redistributes the momentum within the cloud, a preferred rotational axis is defined, forming a flattened disk structure. The pseudodisk will gradually flatten out and turn into a *proto-planetary disk*, mostly supported by the rotation of the disk (Bate, 1998; Matsumoto & Hanawa, 2003).

At this point, the star accretes most of its mass via the disk, with viscosity being the main driver for accretion (Shakura & Sunyaev, 1973). However, the disk is replenished faster from the core envelope, leading to the disk growing in mass. If the disk is heavy enough, it may itself become gravitationally unstable, and start to form spiral density waves and clumps (Kratter & Lodato, 2016). When the envelope runs out of material, the disk will start to deplete its gas, which is either accreted by the protostar or driven away by energetic processes such as stellar winds, photoevaporation

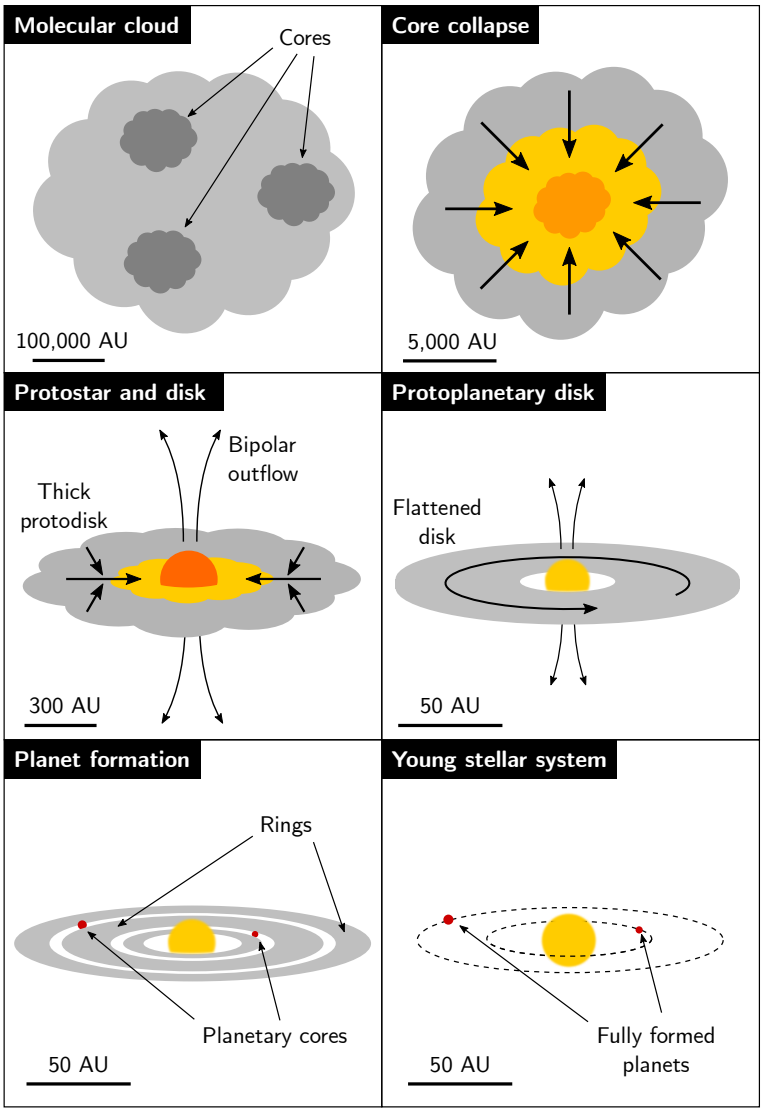


Figure 1.1: A schematic overview of the stages of star and planet formation, from a molecular cloud to a young stellar system. The scales indicate typical length scales for that stage.

or tidal encounters (Hollenbach et al., 2000). When all gas is depleted, the proto-planetary disk has turned into a *debris disk*. This transition from proto-planetary disk to debris disk takes a few million to ten million years and as there is no gas anymore after the disk becomes a debris disk, at least all giant planets must be formed during the proto-planetary disk stage.

1.1.2 Pathways for planet formation

One pathway of planet formation is gravitational instability within the proto-planetary disk. Especially in the outer regions of the disk, where collapse is easier, clumps may start to contract to form giant proto-planets (Boss, 1997). These grow by accreting gas until its orbit is cleared out. Their composition will be the same as the disk out of which they are born, and will therefore be close to that of the star. These giant planets may migrate through the disk (Paardekooper & Mellema, 2006), and provide a possible explanation for hot Jupiters (see also the review by Dawson & Johnson, 2018b).

Rocky exoplanets are formed via another pathway called core accretion. In the protoplanetary disk, small dust grains stick together if they hit each other (Dominik et al., 2007). Once the grains become heavy enough, they decouple from the gas and settle in the midplane of the disk in a process called *vertical settling*. This concentrates grains and promotes further grain growth. As the gas in the disk is partially supported by its pressure gradient, it rotates slightly slower than the local Keplerian velocity. The decoupled grains however do not feel this pressure gradient and orbit at Keplerian velocity. They feel a slight headwind which slows them down and causes them to drift radially inward and spiral towards the star. When particle sizes become on the order of a few millimeters, we start to call them *pebbles*. When pebbles collide, they tend to fragment instead of stick together, and pebbles will therefore not grow to larger sizes (Windmark et al., 2012). Additionally, pebbles will drift too quickly towards the star to aggregate into larger particles that are less susceptible to the radial drift. This problem is known as the *meter size barrier*.

Several mechanisms have been proposed to overcome this barrier, all of which try to concentrate the pebbles to promote growth. For example, if there is a gap in the disk, pebbles will accumulate at the outer edge of the gap as there is no headwind in the gap itself. Another possibility are *streaming instabilities*, where the interaction between the stream of solids and gas in the disk produce dense filaments which can provide densities of over 1000 times the local gas density (Johansen et al., 2011, 2015). Finally, *snowlines*

can create local pressure bumps which trap pebbles. Frozen ice particles drift radially inward and start sublimating when they get hot enough. The resulting gas does not drift inwards, leading to local enrichment of the gas with that element (Ciesla & Cuzzi, 2006; Öberg et al., 2011). Furthermore, the ices on the pebbles may make them stickier outside the iceline, which can further promote the grain growth rate. If these planetesimals become heavier than about ten Earth-masses, they start to accrete the gas and dust in their orbit and finally become a gaseous giant planet (Ikoma et al., 2000). Lower-mass planetesimals may never accrete any gas and become rocky exoplanets. Both of these types of planets may migrate through the disk.

1.1.3 Atmospheric composition and biomarkers

As the accreted gas has a different chemical composition as function of the distance from the star, we can probe the birth site, method of formation and migration history of an exoplanet by looking at this composition. Spectroscopy of exoplanets can tell us this composition, and also about the existence of life on these exoplanets through the detection of *biomarkers* (Schwieterman et al., 2018). Life (as we know it) produces disequilibrium chemistry in an exoplanet atmosphere (Krissansen-Totton et al., 2016), for example molecular oxygen as the byproduct of photosynthesis. Oxygen can be observed either directly with the A-band at $\sim$ 760nm, or indirectly by looking for ozone, which is created in the upper stratosphere when oxygen is exposed to ultra-violet light. Another biomarker is the vegetation red edge, the rapid change in reflectance of vegetation in the infrared resulting from the absorption spectrum of chlorophyll (Tucker, 1979). An interesting recent example is the detection of phosphine (PH_3) gas in the cloud deck of Venus with no currently known abiotic production routes (Greaves et al., 2020). Any single biomarker on their own could be the result natural processes, so we would require the detection of multiple separate biomarkers for an unambiguous statement of life on other planets.

1.2 Observational techniques for finding planets

Many techniques for detecting exoplanets have been developed in the last few decades. Figure 1.2 shows all discovered exoplanets coloured by their primary detection technique. For planets with unknown mass, its mass was estimated from their radius using the mass-radius relations in Bashi et al. (2017). Clearly, each technique has its own biases and strengths, making

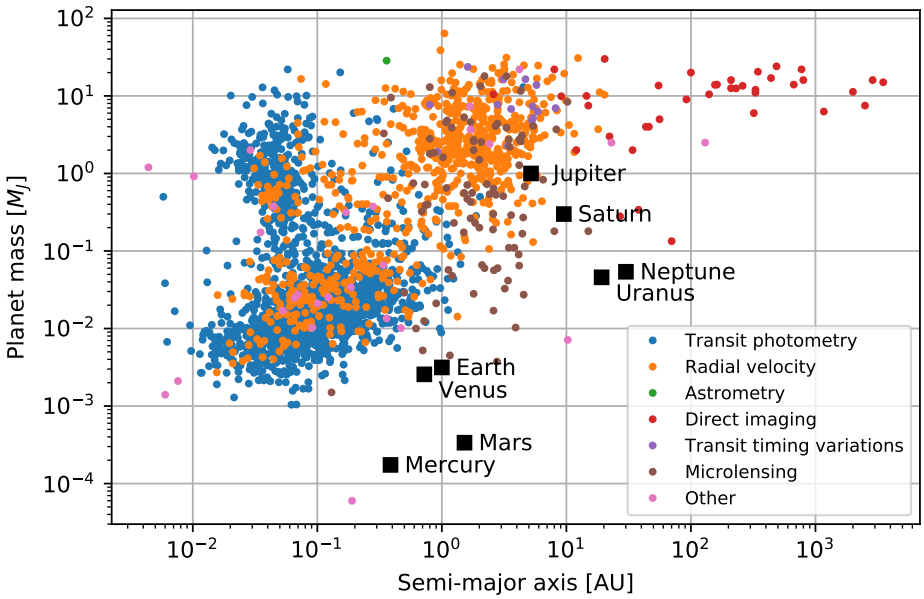


Figure 1.2: The mass and semi-major axis of all discovered exoplanets as of August 2020, coloured by their discovery method. Planets in our solar system are also shown for comparison. Data is taken from the NASA Exoplanet Archive.[1]

them more sensitive in different parts of the parameter space. Therefore, it is necessary to use multiple techniques simultaneously to obtain a more unbiased and complete view of exoplanets. It is outside of the scope of this introduction to discuss all methods in detail; I will briefly discuss the four main methods shown in Figure 1.3.

1.2.1 Transit photometry

Transit photometry has been the most prolific detection method over the years, especially with the launch of the dedicated survey telescope Kepler (Borucki et al., 2010). This method looks for the characteristic periodic dimming which occurs as the planet passes in front of their host star. During the transit, the amount of light received will be reduced proportional to

[1]The NASA Exoplanet Archive is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program (`https://exoplanetarchive.ipac.caltech.edu/`).

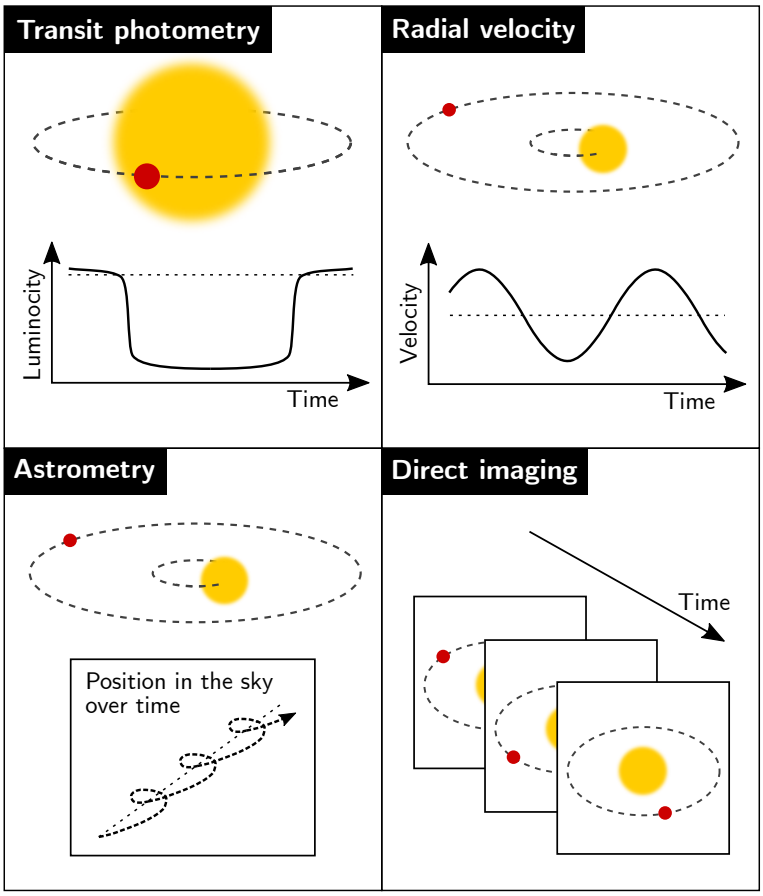


Figure 1.3: Four of the main exoplanet detection methods: transit photometry, radial velocity, astrometry and direct imaging.

the ratio of the areas of the planet and star. Transit depths can vary with Jupiter-sized planets producing a $\sim 1\%$ and Earth-sized planets a $\sim 0.1\%$ drop in flux. Additionally, spectroscopy during the transit can provide us with information on the atmospheric composition as light passing through the atmosphere is partially transmitted depending on wavelength, making the planet appear slightly larger or smaller at different wavelengths. Transits however relies on the fortunate alignment of the planet and star with respect to our point of view. Close-in planets have a higher chance of transiting, with hot Jupiters having a $\sim 10\%$ and Earth-analogues a mere $\sim 0.5\%$. Furthermore, we are biased towards small-period planets as these produce more transits in a given time interval, making it easier to extract

their signal from background noise. Larger planets produce a larger transit depth and are therefore easier to detect as well. This makes transits especially sensitive for hot Jupiters and super-Earths. Over 3000 transiting planets have been detected by *Kepler* alone, and its success has inspired several more satellites, notably TESS (Ricker et al., 2014), PLATO (Rauer et al., 2014) and CHEOPS (Rando et al., 2018).

1.2.2 Radial velocity

The *radial velocity* method is often used in conjunction with the transit method. Radial velocity looks for the small periodic reflex motion of the star when a planet is orbiting it. The size of the reflex motion is directly proportional to the mass of the planet, making heavier planets easier to detect. Typical reflex motions are $\sim$ 30 m/s for a hot Jupiter and $\sim$ 3 cm/s for an Earth-analogue. Radial velocity measures the mass of the planet, or more precisely, a lower bound of the mass in the form of the mass function $M_p \sin i$: if the orbit is inclined only part of the reflex motion will be seen. In case a planet is seen both with the transit and radial velocity method, we know the inclination as the orbit must be close to edge-on to be able to transit, and we can derive an estimate for the bulk density of the planet. Similarly to the transit method, there is a bias for small-period exoplanets as the multiple orbits can be stacked to decrease the noise over a given observation period. Detection is often limited by calibration of the spectrograph or astrophysical noise caused by stellar activity, which makes radial velocity worse for young stars.

1.2.3 Astrometry

Astrometry has so far produced fewer planet detections than the transit and radial velocity methods, but this is set to change with the GAIA mission (Gaia Collaboration et al., 2016). Astrometry looks at the reflex motion of the star as a planet is orbiting it by accurately measuring the centroid of the star's image with respect to a fixed reference frame. Astrometry is biased towards close-in planets where the orbital period of the planet is about twice the duration of the GAIA mission (Sahlmann et al., 2016, 2018). It can detect planets at any orbital inclination, removing one significant bias from the other two earlier methods, is far less limited by astrophysical noise and can be applied stars over a wide range of stellar ages. The GAIA mission will discover and measure several thousands of giant planets out to 3-4 AUs from stars within 200 pc at the end of the full mission (Casertano et al.,

2008) and will discover on the order of 10 planets around nearby bright stars suitable for observation with the Roman Space Telescope (Beichman et al., 2018).

1.2.4 Direct imaging

Direct imaging attempts to take an image of the exoplanet spatially separated from its host star. Direct imaging looks directly at the light of the planet, unlike the indirect methods described above, and can take spectra of the light emerging from the exoplanet and its environs. The main challenge of direct imaging stems from the high contrast ratio between the light from the planet and that of the star. This contrast ratio is a strong function of the wavelength of the observations. At optical wavelengths we are looking at reflected light off of the planet, where this contrast ratio is typically $\sim 10^8$ for Jupiter-analogues and $\sim 10^{10}$ for Earth-analogues around a Solar type star. At infrared wavelengths, however, for younger planetary systems (ages less than 500 Myr or so) we instead see the thermal radiation from the planet, which is still cooling down from the formation process. This improves the contrast ratio to $\sim 10^5$ for Jupiter analogues. Additionally, the small angular separation between the planet and the star makes it difficult for an instrument to distinguish between stellar and planet light.

Figure 1.4 shows the contrast (in reflected light) and angular separation for all discovered exoplanets to date. The planet-to-star flux ratio is estimated from the planet radius and physical separation, and assumes that the planet is observed at quadrature. If the planet radius is not known, as is the case for most radial velocity planets, it is estimated from the planet mass analogous to Lovis et al. (2017). It is clear that observing exoplanets in reflected light is extremely challenging, requiring intricately designed instruments and advanced data processing techniques to produce high-quality data.

1.3 Anatomy of a high-contrast imaging instrument

Diffraction plays an integral role in high-contrast imaging. Even for a perfect optical system, diffraction provides the fundamental resolution of the images that we achieve. For an unobstructed circular telescope pupil, the smallest angular size $\Delta\theta$ that we can resolve is

$$\Delta\theta = 1.22\lambda/D, \tag{1.1}$$

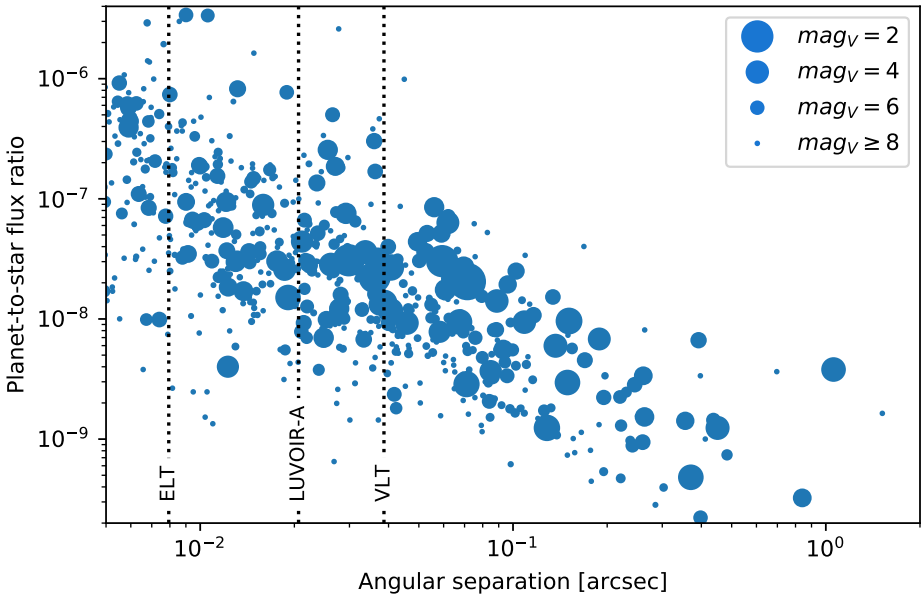


Figure 1.4: A contrast-separation plot for all discovered exoplanets as of August 2020 in the visible. The marker size indicates the magnitude of their host star in the optical (V-band). Approximate inner working angles are shown for the VLT, ELT and LUVOIR-A telescopes at $2\lambda/D$ for $\lambda = 750$nm. Data is taken from the NASA Exoplanet Archive.

where λ is the wavelength of the light and D is the diameter of the telescope mirror. Any object smaller than this size will look like the point spread function (PSF) of the telescope, which is fully determined by the shape of the telescope pupil, as seen in Figure 1.5 panel (a). Most stars are unresolved by current telescopes, and will therefore look like the telescope PSF. In Fig. 1.5, I show a few example of PSFs for various telescopes around the world. Generally, less than 80% of the star light will be contained within the central core of the PSF. The rest of the light is spread out into a large number of successively fainter *diffraction rings* around the core. This means that at the location of the planet, typically only a few to a few tens of diffraction widths away from the star, most of the received photons originate from the diffraction halo of the star and not the planet, even when we observe with a idealized perfect optical system. In order to filter out this unwanted light and reveal the dim planet buried underneath, we use a *coronagraph*. The coronagraph is an optical device – an angular filter –

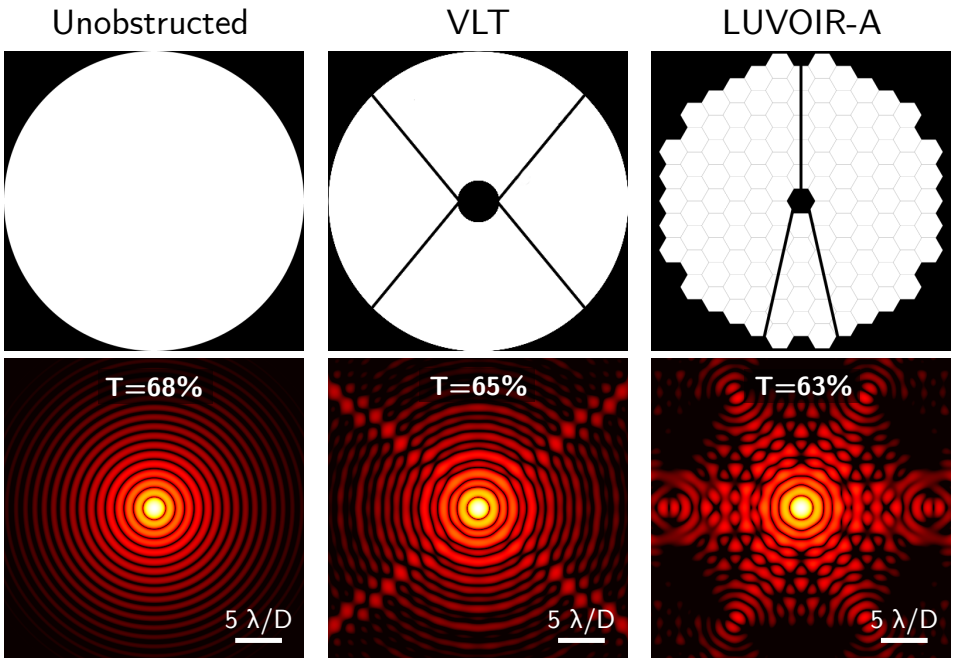


Figure 1.5: Simulated point spread functions for (a) an unobstructed circular pupil, (b) the VLT, and (c) the LUVOIR-A telescope. The throughput T is indicated on each of the PSFs and is calculated as the fraction of the power inside a $1.4\lambda/D$ circular aperture centered on the peak.

that is able to suppress the on-axis star light, while letting through most of the off-axis light from planets, the disk, or background objects.

In a perfect world, a coronagraph by itself will always produce high quality coronagraphic images, in practice this is not the case. As starlight propagates through space, the originally spherical wavefronts centered on the star gradually become flatter, like ripples in a pond expanding out towards the edges. When the light wave arrives at Earth, its wavefront will be almost completely flat. However, when passing through the atmosphere, this flat wavefront becomes distorted as it propagates through the Earth's turbulent atmosphere. Changes in the temperature of the air create a time and position dependent index of refraction across the telescope pupil (Fried, 1966). Imperfections in the manufactured optics, or the alignment of these optics in our optical system, provide an additional source of wavefront distortion. Both these sources of wavefront errors are dynamic: atmospheric wavefront errors will change on on millisecond timescales as the moving tur-

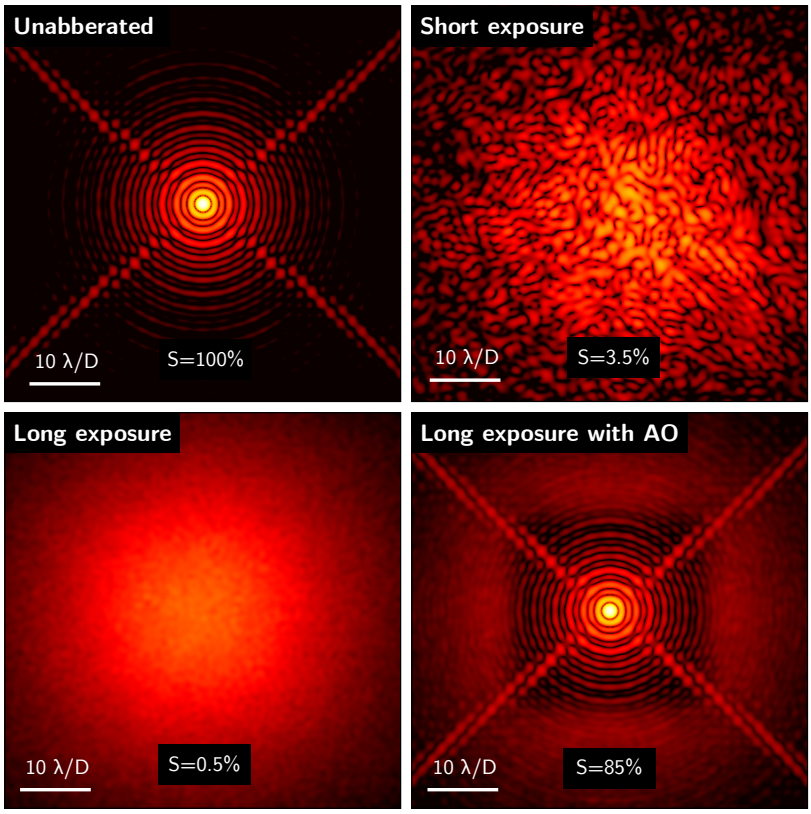


Figure 1.6: Simulated focal-plane images for the VLT in various circumstances. (a) No aberrations; (b) Short exposure with atmospheric aberrations; (c) Long exposure through atmospheric turbulence; (d) Long exposure through atmospheric turbulence with AO.

bulence in the atmosphere is blown across the telescope pupil (Greenwood, 1977) and the alignment of the optical train will change on minute to hour timescales, as the telescope moves to track the star across the sky over the course of the night (Hinkley et al., 2007; Martinez et al., 2013; Milli et al., 2016). Temperature fluctuations and gradients within the instrument may move optics due to thermal expansion of the mechanics to which the optics are mounted.

As our coronagraph is only designed to suppress light with a flat wavefront, it will transmit any deviations from this ideal wavefront, resulting in a cloud of speckles around the optical axis. If the wavefront aberrations are strong enough, typically more than 1/4 of a wave across the pupil, the

core of the PSF is also broken up, spreading the light over a larger area which makes it much more difficult to detect a planet. A rough metric for the wavefront quality is called the Strehl ratio S, which is approximated by $e^{-\sigma^2}$ where σ is the root mean square deviation of the wavefront phase in radians over the pupil. For an ideal optical system, $S = 1$ and for most ground based high contrast imaging cameras, $S \sim 0.6-0.9$ for 8 metre class telescopes at wavelengths of $1-3$ microns. We use *wavefront control system*, also called *adaptive optics* (AO) system for ground-based observatories, to reflatten the wavefront before the light passes into the coronagraph. This system measures the wavefront aberration and attempts to apply the opposite and equal aberration with a wavefront modulator. Figure 1.6 shows typical examples of focal-plane images for the VLT pupil with and without AO system.

Finally, we need to capture the planet light on a camera. This is the job of the *science instrument*. This part of the instrument provides both scientific data for the astronomer and calibration data for the coronagraph or wavefront control system. The calibration data allows us to further correct the scientific images for errors made by the coronagraph, the wavefront control system or any other optics in our instrument. In general, the science instrument can be simply a detector or a more complicated optical system, detecting the spectrum of light at each pixel with an integral field spectrograph, eg. SPHERE/IFS (Claudi et al., 2008) or GPI/IFS (Larkin et al., 2014), or detecting the polarization state at each pixel with a polarization beam splitter, eg. SPHERE/ZIMPOL (de Boer et al., 2020; Schmid et al., 2018), or both (Rodenhuis et al., 2014; Tinyanont et al., 2019).

1.3.1 The coronagraph

Many different types of coronagraphs have been developed over the last two decades. A comprehensive review of the current field of coronagraphy are in the excellent review articles of Guyon et al. (2006), Mawet et al. (2012) and Ruane et al. (2018). The focus of this section is to provide an overview of the considerations that need to be taken into account when designing a coronagraph, and the theory behind how a typical coronagraph suppresses star light.

End-to-end simulations

Any specific coronagraph design will be a compromise between a large number of different metrics. The environment that the coronagraph will be op-

erating in determines to what degree each of these metrics count towards the performance of the overall system. This makes it difficult to tell in advance which type of coronagraph will perform best in a specific system, and often extensive end-to-end simulations or even on-sky testing of actual implementations are required to reach a conclusion.

An example of the output of current state-of-the-art end-to-end simulations can be found in Figure 1.7 as performed for the proposed LUVOIR-A mission by AYO, the Altruistic Yield Optimizer (Stark et al., 2014; The LUVOIR Team, 2019). AYO operates in a simulated universe complete with simulated stars. A hypothetical mission is defined with coronagraph models, limited mission lifetime and estimated operating costs, and then attempts to detect and characterize exoplanets around these stars. AYO decides which stars too look at, for how long and with which coronagraph or observing mode, in order to maximize the exoplanet candidate yield and the number of characterized exoplanets at the end of the mission. Several simulated missions are usually performed to average out statistical errors and provide confidence intervals.

Figure 1.7 shows preliminary results for two different mission configurations, the first being the baseline LUVOIR-A mission and the second where one of the coronagraphs was replaced by a phase-apodized-pupil Lyot coronagraph (PAPLC), developed in Chapter 7 and 8 of this thesis. This shows that the PAPLC provides $\sim 25\%$ more exoplanet candidates, a significant improvement.

Terminology and design metrics

While we would prefer to perform complete end-to-end simulations to guide the design process, in most cases we do not have the time or the resources. Often the design is chosen subjectively by the coronagraph designer based on a number of simpler metrics. In this section, I give an overview of the definition of these metrics and in what way they impact the exoplanet candidate yield.

- One of the most obvious metrics is *raw contrast*. This describes to what level the starlight is suppressed. The raw contrast is the amount of light within a given aperture (usually defined as a small multiple of the FWHM of the central core PSF) at a given angular separation from the central star, and normalised to the total enclosed flux within the same aperture centered on the star. External factors, such as the performance of the wavefront control system and the post-processing

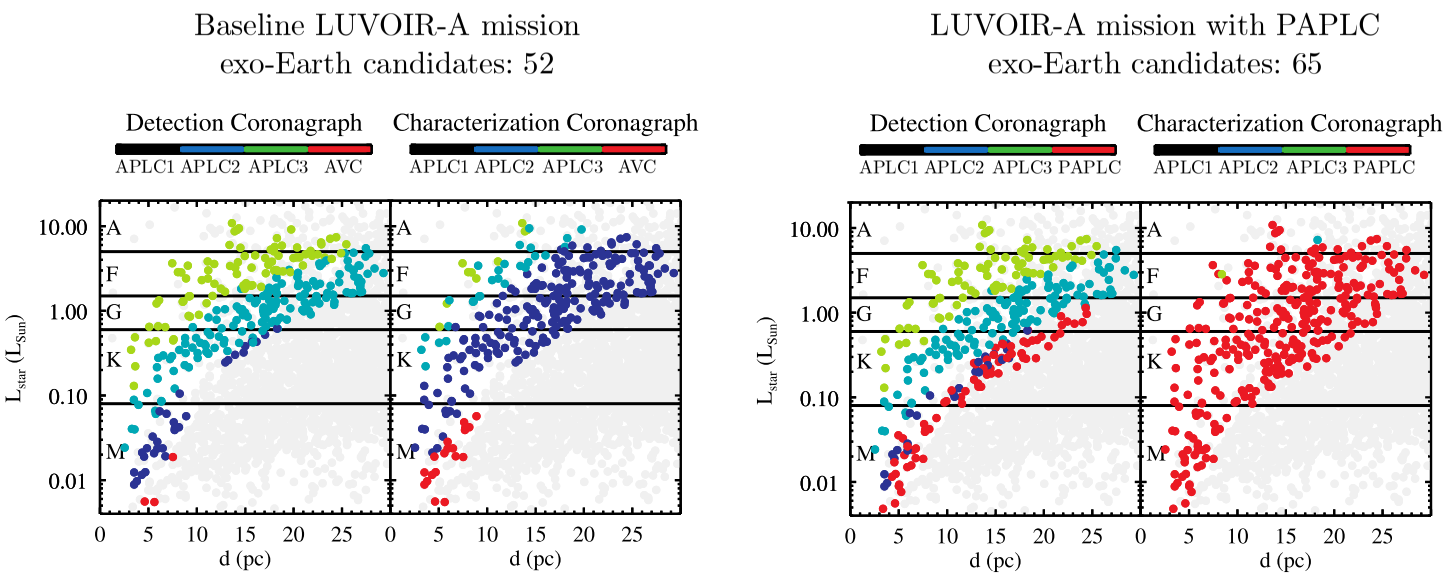


Figure 1.7: The luminosity-distance plots showing exo-Earth candidates for a simulated LUVOIR-A mission with the baseline coronagraph configuration (left) and a configuration where the apodized vortex coronagraph is replaced by the PAPLC (right). Simulations were performed by Chris Stark using the PAPLC design presented in Chapter 7. The PAPLC is chosen by AYO for almost all characterization observations.

techniques, may play a role in defining this quantity. Designing a coronagraph with a raw contrast much deeper than the wavefront control system is able to provide will not improve the quality of our images. Typically, we require raw contrasts of 10^{-5} – 10^{-8} for Earth-based observatories, and 10^{-8} – 10^{-10} (where the telescope diameter is smaller and therefore the diffraction limit is larger) for space-based observatories.

- Another metric is *planet throughput*. This describes the fraction of light that we receive from an exoplanet compared to an observation without the coronagraph inserted into the optical train. The apparent magnitude of Earth-analogues are expected to be $\gtrsim 26$ mag, even for those orbiting our closest stars, so we do not want to unnecessarily block/remove too many precious photons for our science instrument. Simulations show that a planet throughput of $< 25\%$ is deemed insufficient for exoplanet imaging, and planet throughputs of $> 50\%$ are to be preferred.

- Planet throughput is rarely independent of the relative location of the planet with respect to the star. Planet throughput starts at zero for a zero angular separation, as it will be located at the same position as the star and its light will be indistinguishable from star light and therefore suppressed in the same way. For planets at larger nonzero angular separations, planet throughput rises up to a certain maximum theoretical throughput. A convenient quantity is the *inner working angle*, which defines the angular separation where the planet throughput is half of the maximum planet throughput for that coronagraph. Mimimizing inner working angles allows us to look closer to the star at the expense of other metrics.

- A related metric is *field of view*. Some coronagraphs only obtain the contrast in a certain part of the focal plane, the region of interest. When performing blind surveys, we do not know where or even if there is an exoplanet. Coronagraphs with a larger region of interest are able to search more efficiently for exoplanets compared to coronagraphs with a smaller region of interest. One interesting case here are coronagraphs with a region of interest on only one side of the star, rather than in an annulus surrounding the star.

- *Robustness against low-order aberrations* is a useful property of a coronagraph. Dynamic misalignment of optics tend to induce mostly

low-order aberrations. While the wavefront control system tries to correct these continuously, some portion remains and we view these as drifting low-order aberrations during the observation. When a coronagraph is robust against these, we can handle lower photon fluxes as we do not need to control the aberrations as tightly as we otherwise would. However, as the throughput at small angular separations is mostly determined by the throughput of low-order modes, especially the inner working angle is most affected when making a coronagraph more robust.

- Robustness against specifically tip-tilt aberrations plays a big role in the *robustness against stellar diameter*. Future telescopes are becoming so large that they are on the verge of resolving the stellar surface for nearby stars. Stellar angular diameters can be as large as $0.01 - 0.1\lambda/D$ for 8-meter class telescopes observing nearby stars at wavelengths of scientific interest. Coronagraphs must therefore be designed to suppress an extended rather than a point source. The problem of stellar diameters primarily affects space telescopes compared to ground-based telescopes due to their deeper raw contrast requirement.

- An often overlooked metric is the *optical complexity* of the coronagraph design. Coronagraphs with a more complicated optical layout are more susceptible to misalignment. The more complex an optical element is, the more costly it will be to manufacture and the higher the chance of manufacturing errors or damage. On the other hand, the more complex a coronagraph is, the higher the design freedom, making it have higher performance compared to simpler coronagraphs.

- Finally, an *integration with wavefront sensing* can be extremely useful to have. Coronagraphs with a built-in wavefront sensor can provide continuous wavefront telemetry to the wavefront control system. As the wavefront sensing occurs at the plane of the coronagraph, this eliminates most non-common-path aberrations while also allowing simultaneous wavefront control and coronagraphic imaging.

There is a complex interplay of trade offs between the most significant factors in optical and scientific design of a given instrument, and there are several more that might be added to this initital list, depending on the specific science case and the engineering constraints.

An idealized coronagraph

Before discussing physical coronagraph implementations, it is instructive to first look at an idealized or "perfect" coronagraph. While these perfect coronagraphs cannot be manufactured, they can serve as the fundamental limit for coronagraphy: they allow us to see the optimal performance of an idealized coronagraph with which to compare actual coronagraph designs against.

The perfect coronagraph was introduced by Cavarroc et al. (2006) and projects the incoming electric field on an unaberrated pupil-plane electric field that represents an ideal point source incident on the telescope optics. This projection is then subtracted from the incoming light, providing the output electric field. The second-order perfect coronagraph can be represented by the operator

$$\mathcal{P}^{(2)}\{\Psi\} = \Psi - \mathrm{proj}_{V_2}\{\Psi\}, \tag{1.2}$$

where proj_{V_2} denotes the projection operator onto the vector space $V_2 = \mathrm{span}\,(\Pi)$, and Π denotes an unaberrated pupil-plane electric field. Any coronagraph can be ultimately represented by an operator and this operator has to be linear as the light levels in a telescope are not strong enough to excite non-linear optical phenomena. The coronagraph in Equation 1.2 is "perfect" in the sense that it maximally retains all light that is not star light, while fully suppressing all starlight and still being a linear operator.

In Figure 1.8, I show the image of a point source through a second-order perfect coronagraph at different angular separations from the optical axis, along with the core throughput, i.e. the amount of light transmitted in a $1.4\lambda/D$ diameter circle centered on the point source. As expected, the coronagraph suppresses all light when the object is on-axis and has almost no effect when the object is far from the optical axis. However, even at extremely small angular separations a significant fraction of light is transmitted: the coronagraph is very sensitive to low-order aberrations.

We can describe this mathematically as follows. For a small tip-tilt aberration, the input electric field is

$$\Psi_{\mathrm{tiptilt}} = \Pi \exp(\alpha x) \tag{1.3}$$

$$= \Pi(1 + \alpha x + \mathcal{O}(\alpha^2)), \tag{1.4}$$

which yields for the light through the coronagraph

$$\mathcal{P}^{(2)}(\Psi_{\mathrm{tiptilt}}) = \alpha \cdot (x\Pi - \mathrm{proj}_{V_2}\{x\Pi\}) + \mathcal{O}(\alpha^2)), \tag{1.5}$$

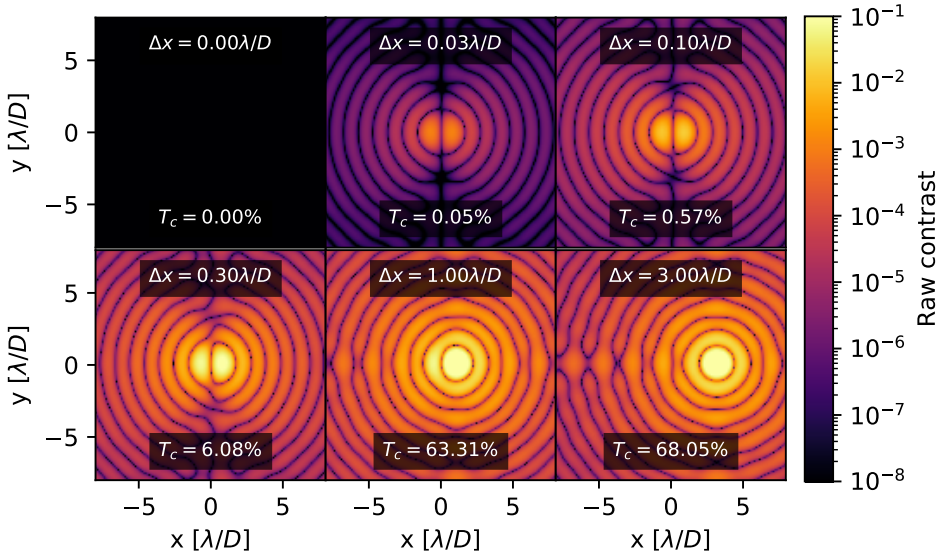


Figure 1.8: Images for on- and off-axis point sources through a second-order perfect coronagraph. For each image, the core throughput, the fraction of light inside a $1.4\lambda/D$ diameter circle centered on the object, is listed.

where x is the x-coordinate in the pupil. This shows a quadratic response in the intensity leaking through the coronagraph, $I_{\text{leak}} = |\mathcal{P}^{(2)}\{\Psi_{\text{tiptilt}}\}|^2 \propto \alpha^2$ as function of α, the strength of the aberration. This quadratic behaviour is the reason why this type of perfect coronagraph is known as a second-order perfect coronagraph.

Guyon et al. (2006) extends the perfect coronagraph to higher orders to add robustness to low-order aberrations. This is done by extending the vector space V_2 to include more than just the telescope pupil. For example, a fourth-order and sixth-order perfect coronagraph can be described by the vector spaces

$$V_4 = \text{span}\left(\Pi, x\ \Pi, y\ \Pi\right), \tag{1.6}$$

$$V_6 = \text{span}\left(\Pi, x\ \Pi, y\ \Pi, x^2\ \Pi, xy\ \Pi, y^2\ \Pi\right), \tag{1.7}$$

and the general n-th order perfect coronagraph as

$$\mathcal{P}^{(n)}\{\Psi\} = \Psi - \text{proj}_{V_n}\{\Psi\}. \tag{1.8}$$

Looking at the leakage light for a small tip aberration for a fourth-order

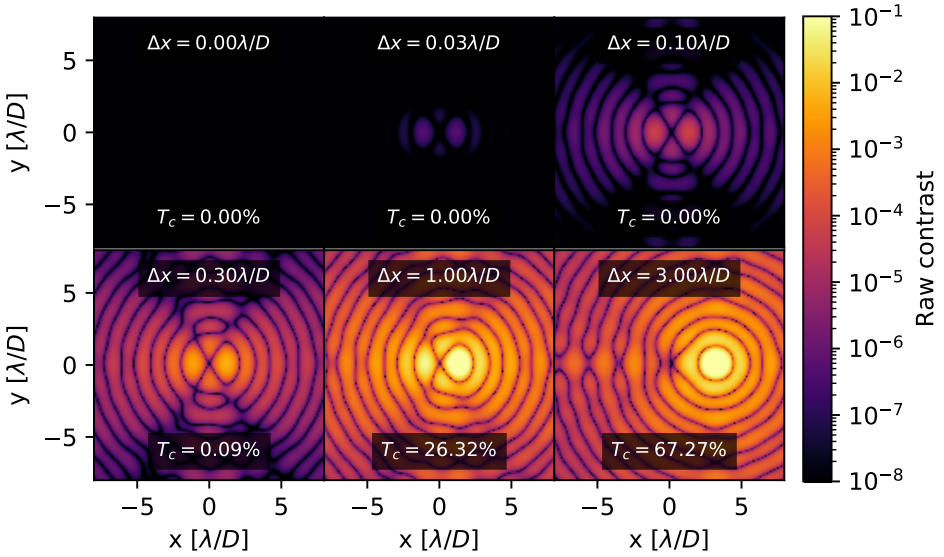


Figure 1.9: Images for on- and off-axis point sources through a fourth-order perfect coronagraph. For each image we list (i) the core throughput, and (ii) the fraction of light inside a $1.4\lambda/D$ diameter circle centered on the object. Compared to the same images for a second-order coronagraph, the amount of leakage light is strongly reduced, see Figure 1.8.

perfect coronagraph

$$\mathcal{P}^{(4)}(\Psi_{\text{tiptilt}}) = \alpha^2 \cdot (x^2\Pi - \text{proj}_{V_4}\{x^2\Pi\}), \tag{1.9}$$

we see a quartic response in intensity as function of α, the strength of the aberration. Likewise, for the sixth-order perfect coronagraph this yields a sixth-order polynomial response.

Looking at the same on- and off-axis images for the fourth-order perfect coronagraph in Figure 1.9, we can visually see this increased robustness to low-order aberrations. However, we can also see the increase in the inner working angle of the coronagraph. For close in planets, the incoming light contains predominantly low-order electric field modes which are now filtered out by the coronagraph and therefore do not count towards the planet throughput. Higher-order perfect coronagraphs have necessarily larger inner working angles.

The classical Lyot coronagraph

We now transition from the idealized perfect coronagraphs to physical implementations of coronagraphs. Physical coronagraphs typically consist of a series of pupil and focal planes, in each of which there may be a mask that modifies the complex electric field (i.e. both amplitude and phase of the wavefront). These masks are designed in such a way as to suppress the light from the star, while transmitting the light from the planet.

The first coronagraph was developed by Lyot (1939) and was originally intended for imaging the solar corona. It was later adapted for use with stars, and remains a standard layout for many coronagraphs. Its optical layout is shown schematically in Fig 1.10 along with the electric field in all relevant planes.

We will now trace the light through the Lyot coronagraph starting with the electric field just before plane A, $\Psi_{A-} = \Pi$. As there is no mask in this pupil plane, the light right after plane A is the same:

$$\Psi_{A,+} = \Psi_{A,-} \tag{1.10}$$
$$= \Pi. \tag{1.11}$$

Propagating to plane B, a focal plane, we get

$$\Psi_{B,-} = \mathcal{C}\{\Psi_{A,+}\} \tag{1.12}$$
$$= \mathcal{C}\{\Pi\}, \tag{1.13}$$

where $\mathcal{C}\{\cdot\}$ is a linear operator that propagates a field from the pupil to the focal plane, typically approximated with a Fraunhofer propagator. In this focal plane, we put a small opaque disk centered on the star. We call this mask $M(\mathbf{k})$. Just after this focal plane mask, the electric field is

$$\Psi_{B,+} = M(\mathbf{k}) \cdot \Psi_{B,-} \tag{1.14}$$
$$= M(\mathbf{k}) \cdot \mathcal{C}\{\Pi\}. \tag{1.15}$$

Note that this mask by itself only blocks the center of the field, and that outside this opaque mask nothing has changed. Bernhard Lyot realized that if we look at this field in the next pupil plane, plane C, most of the light is outside of the original telescope pupil. This can be seen in Fig. 1.10, but it is instructive to view this from a mathematical perspective as well. The electric field just in front of the Lyot stop mask is

$$\Psi_{C,-} = \mathcal{C}^{-1}\{\Psi_{B,+}\} \tag{1.16}$$
$$= \mathcal{C}^{-1}\{M(\mathbf{k}) \cdot \mathcal{C}\{\Pi\}\}, \tag{1.17}$$

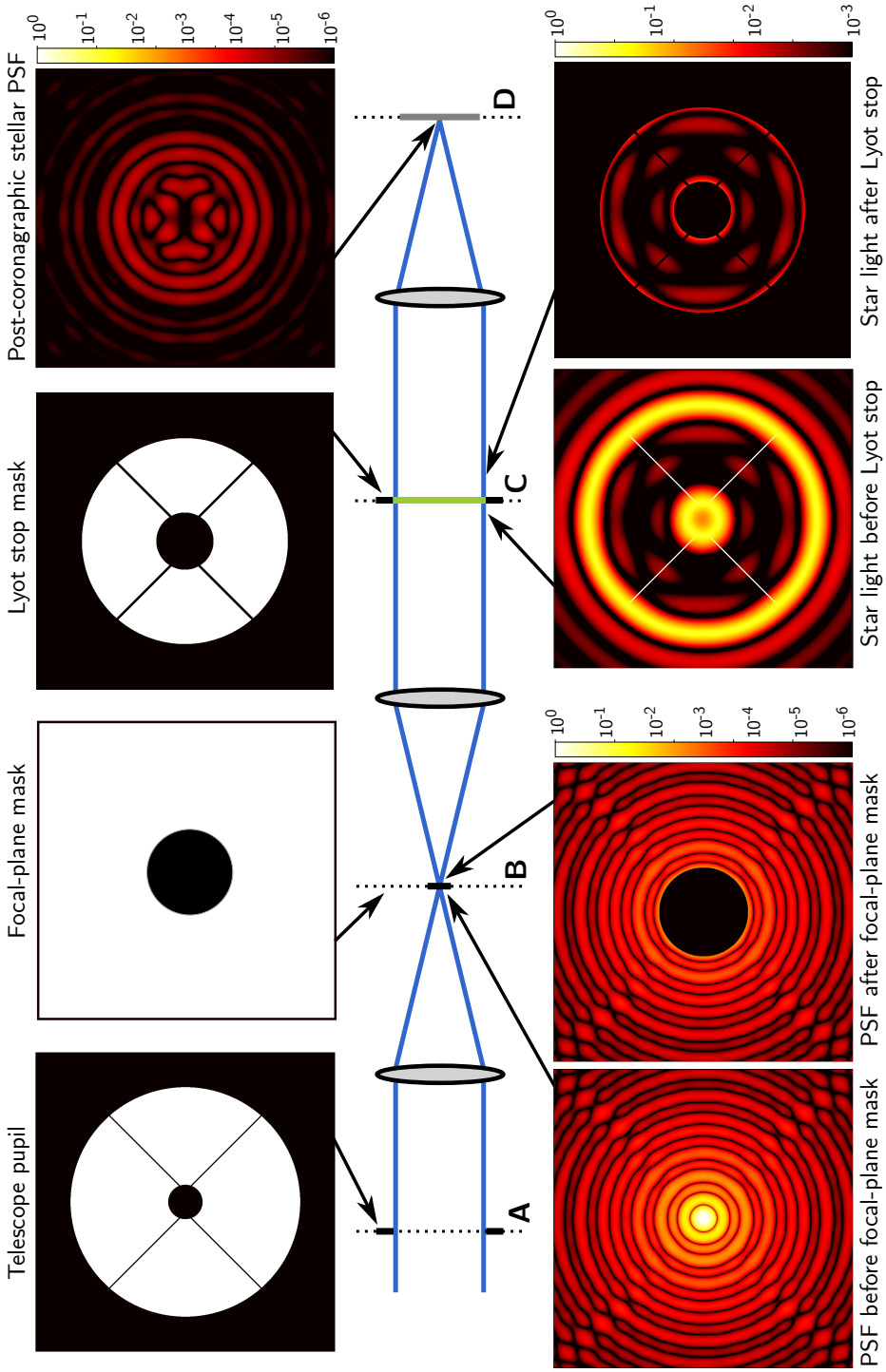


Figure 1.10: The propagation of light through a Lyot coronagraph.

which can be rewritten using Babinet's principle as

$$\Psi_{C,-} = \Pi - \mathcal{C}^{-1}\{(1 - M(\mathbf{k})) \cdot \mathcal{C}\{\Pi\}\}. \tag{1.18}$$

Comparing this to Eq. 1.2 for the perfect coronagraph, we can see a similar structure. Instead of subtracting a scaled version of the telescope pupil, we are subtracting a filtered version of the incoming field. As $(1 - M(\mathbf{k}))$ is a small circular aperture and the operator $\mathcal{C}\{\cdot\}$ is analogous to a Fourier transform, the second term is actually a low-pass-filtered version of the telescope pupil. Therefore the the subtraction will be good in the middle of the telescope pupil, but the hard edge of the telescope pupil is not well subtracted, yielding a bright ring on the edge of the pupil. The Lyot coronagraph blocks this starlight with a mask, nowadays called Lyot stop after its inventor Bernard Lyot. Planet light on the other hand will pass through mostly unhindered, as the majority of its light will be inside the telescope pupil.

Blocking with the Lyot stop mask $L(\mathbf{x})$ yields

$$\Psi_{C,+} = L \cdot \Psi_{C,-} \tag{1.19}$$

$$= L \cdot \left[\Pi - \mathcal{C}^{-1}\{(1 - M(\mathbf{k})) \cdot \mathcal{C}\{\Pi\}\}\right], \tag{1.20}$$

so that we can write the Lyot coronagraph operator $\mathcal{L}\{\cdot\}$ as

$$\mathcal{L}\{\Psi\} = L \cdot \left[\Psi - \mathcal{C}^{-1}\{(1 - M(\mathbf{k})) \cdot \mathcal{C}\{\Psi\}\}\right]. \tag{1.21}$$

While the Lyot coronagraph is very simple, it typically only suppresses the star by one or two decades in raw contrast. In fact, there is not enough design freedom to completely suppress the star while still having a reasonable planet throughput: no combination of focal-plane mask size and Lyot-stop diameter will make the two terms in Eq. 1.21 equal to each other. However, many modifications have been proposed to improve upon the traditional Lyot coronagraph.

Lyot-style coronagraphs with pupil-plane apodizers

The first of these was attempted as early as the late 1980's, where Ftaclas et al. (1988) notes that using a graded transmission mask in in the pupil plane before the focal-plane mask lets the Lyot stop suppress more light. We can see intuitively that by smoothing out any edges of the pupil makes it look closer to its low-pass-filtered version, which provides better stellar suppression. This technique was further developed by Soummer

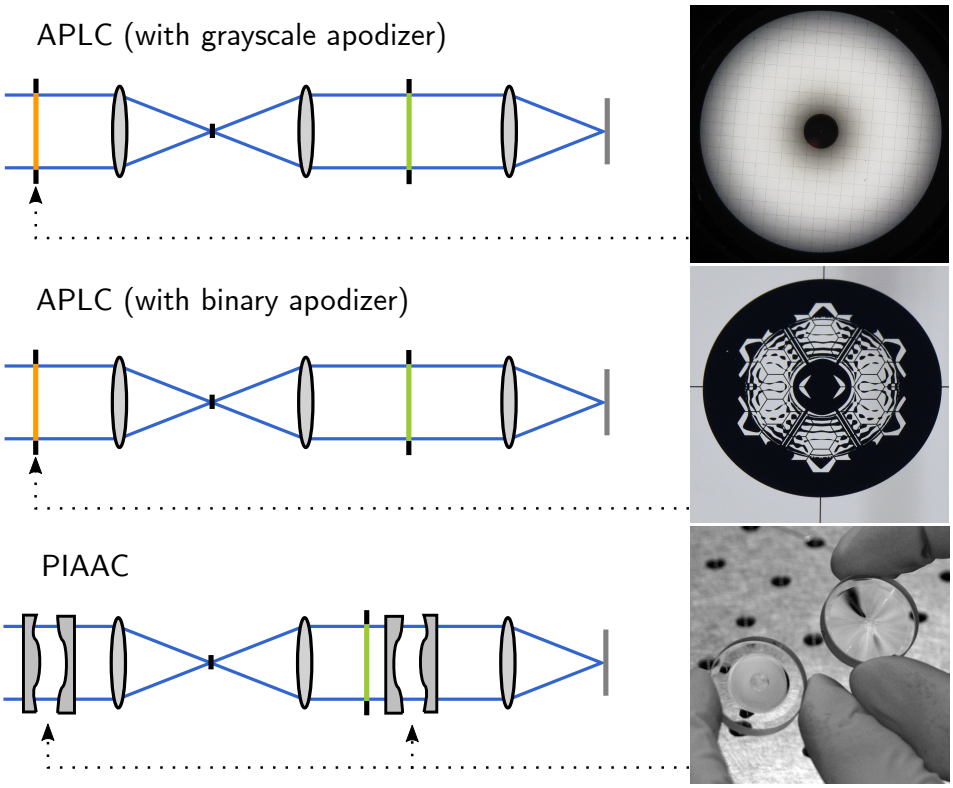


Figure 1.11: Schematic optical layouts of three types of Lyot coronagraphs with pupil-plane apodizers. The microscopy and photographic images are adapted from Sivaramakrishnan et al. (2010), Soummer et al. (2018), and Lozi et al. (2009).

et al. (2003a) and Soummer (2005), showing that a prolate apodization profile can fully suppress the star. Later on, the grayscale apodizers were replaced with binary masks as globally-optimal optimization methods were developed (N'Diaye et al., 2016; Zimmerman et al., 2016a). This collection of coronagraphs is now known as the apodized-pupil Lyot coronagraph (APLC) family. APLCs have seen widespread usage within the current generation of ground-based instruments, such as VLT/SPHERE (Guerri et al., 2011) and Gemini/GPI (Sivaramakrishnan et al., 2010). For a number of future space telescopes, the APLC is the baseline coronagraph, for example for the Roman Space Telescope (Zimmerman et al., 2016a) or the LUVOIR-A telescope (The LUVOIR Team, 2019; Zimmerman et al., 2016b).

Partially blocking light in the telescope pupil is wasteful: planet light passes through the apodizer mask in the same way as star light, so planet photons get absorbed as well, decreasing planet throughput. Guyon (2003) overcomes this problem by replacing the grayscale apodizer with two free form mirrors that redistribute the light in the pupil to achieve the same apodization as the apodizer mask would. This coronagraph is known as the phase-induced amplitude apodization (PIAA) coronagraph and yields more efficient coronagraphs as the apodization is achieved without any light loss. This comes at the cost of a higher optical complexity as freeform mirrors can be hard to manufacture with sufficient accuracy. Additionally, light at larger angular separations is significantly distorted by the pupil remapper, in some cases requiring the use of an inverse pupil remapper at the Lyot stop plane. This counteracts the distortion at the cost of further optical complexity.

Lyot-style coronagraphs with complex focal-plane masks

Another option that was explored was using a graded focal-plane mask (Kuchner & Traub, 2002). This type of coronagraph was later extended towards focal-plane masks that apodize both in amplitude and phase, forming the hybrid Lyot coronagraph (HLC) (Moody & Trauger, 2007), and has been chosen as one of the two coronagraphs on board of the Roman Space Telescope (Trauger et al., 2016). The focal-plane masks for the HLC are manufactured with metallic and dielectric layers deposited on a glass substrate (Trauger et al., 2016).

Roddier & Roddier (1997) proposes to replace the small opaque dot in the focal plane mask of the Lyot coronagraph by a small dot that induces a phase of π onto the light passing through. This corresponds to a mask transmissivity of -1 instead of 0, effectively subtracting twice the low-pass-filtered version from the original electric field. This allows for shrinking of the size of the focal-plane mask, making the phase-mask coronagraph more efficient at small angular separations.

The focal-plane mask is often manufactured using photolithographic techniques where the phase profile is etched directly into the substrate. This means that the applied phase shift will be dependent on the wavelength of the light, making the phase-mask coronagraph only capable of operating efficiently only close to its design wavelength. Soummer et al. (2003b) generalizes the phase-mask coronagraph by adding an additional ring around the central dot. This enables achromatization of the coronagraph, correcting for both size chromatism (the growth of the PSF as

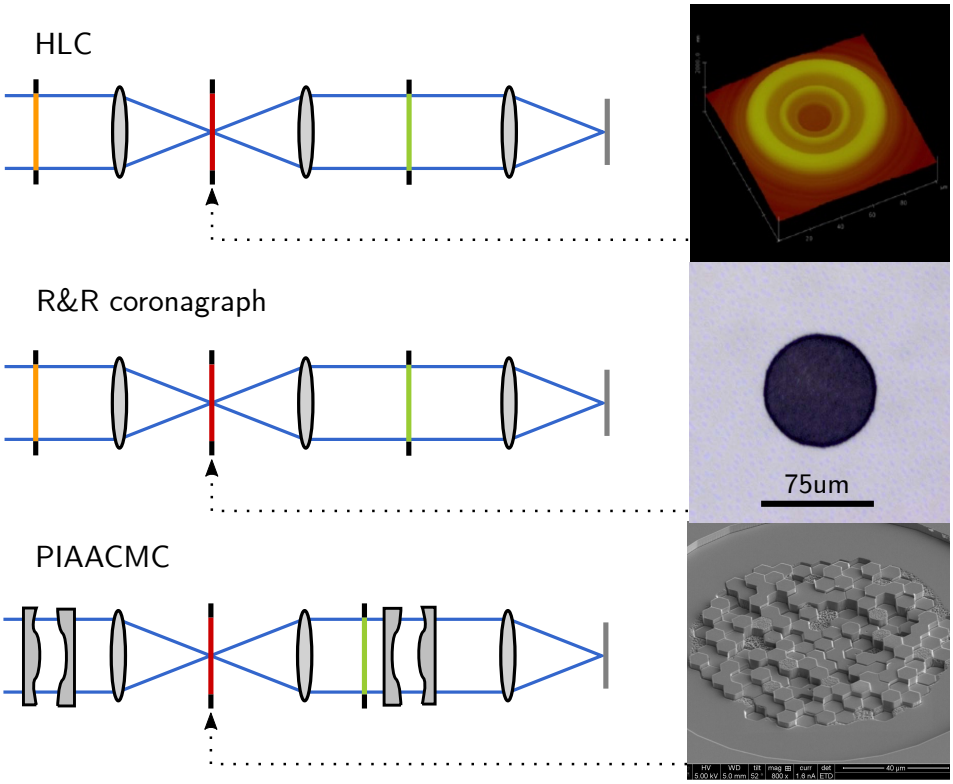


Figure 1.12: Schematic optical layouts of three types of Lyot coronagraphs with focal-plane apodizers. The scanning electron and optical microscope images are adapted from Trauger et al. (2016), Doelman et al. (2019), and Knight et al. (2018). While the image for the Rodier & Rodier (R&R) coronagraph depicts a microscopic image of a Zernike wavefront sensor focal-plane mask, the analogous image for the R&R coronagraph, although not manufactured, would look very similar.

function of wavelength), and phase chromatism (the chromatic response of the substrate material).

This phase-mask coronagraph can be combined with a pupil-plane apodizer (Soummer et al., 2003b) and the pupil remapper can be used to achieve higher throughput (Guyon et al., 2010). The focal-plane mask can be additionally divided into multiple zones to correct for chromatism, either in rings (Newman et al., 2016) or with hexagonal tiling (Knight et al., 2018).

Phase-mask coronagraphs

There are several types of coronagraphs that retain the optical layout of the Lyot coronagraph but use a different approach for the focal-plane mask. Rather than having a localized effect, their focal-plane masks extend out to infinity. The earliest example of these coronagraphs is the four-quadrant phase-mask (FQPM) coronagraph (Rouan et al., 2000). Its focal-plane mask consists of four equally-sized regions, alternating in phase between 0 and π and extending radially from the center. Higher-order versions have been developed as well, providing increased robustness against low-order aberrations (Murakami et al., 2008) in a similar fashion as higher-order perfect coronagraphs.

While these coronagraphs tend to have excellent planet throughput at large angular separations, the planet light is scattered when it hits the edges between neighbouring regions, so the planet throughput drops at those locations. This problem is solved by the vortex coronagraph, which replaces the discrete regions with a smoothly varying azimuthal phase gradient (Foo et al., 2005; Mawet et al., 2005). The number of times the phase hits 0 is known as the charge of the vortex, which acts in a similar way to the order in perfect coronagraphs (Ruane et al., 2017), again at the expense of increased inner working angles.

For unobstructed telescope pupils, all starlight will be blocked by both the FQPM and vortex coronagraphs. However, for obstructed pupils there will be star light leaking through. One of the solutions that has been investigated is adding an apodizer in the pupil plane upstream from the focal-plane mask. For simple circularly-symmetric pupils, analytical solutions can be derived for this apodizer (Fogarty et al., 2017; Mawet et al., 2013a), and numerical apodizers can be found for non-symmetric telescope pupils as well (Carlotti, 2013; Ruane et al., 2016).

The vortex coronagraph has seen successful deployment on many ground-based instruments (Absil et al., 2016; Mawet et al., 2013b; Serabyn et al., 2010, 2017) and is the basis for the future HabEx space telescope (Gaudi

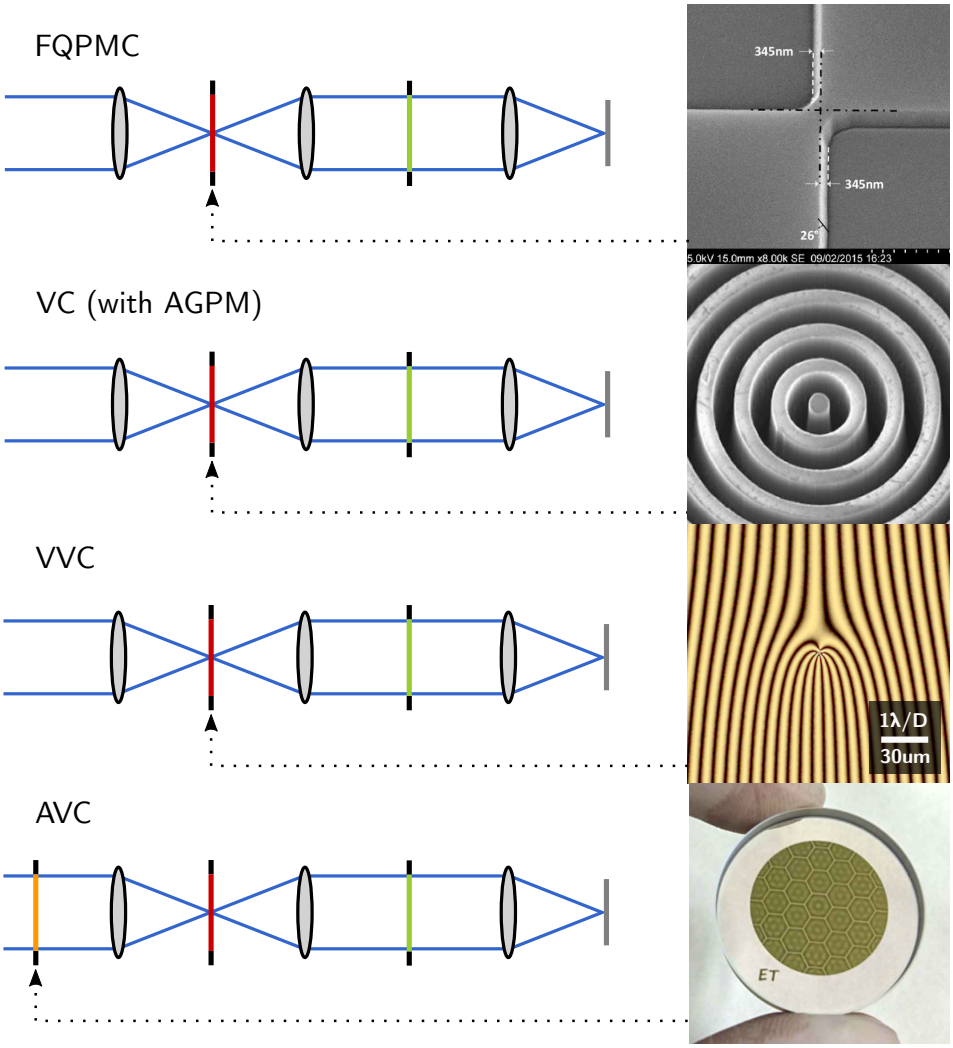


Figure 1.13: Schematic optical layouts of four types of phase-mask coronagraphs. The scanning electron microscope, optical microscope and photographic images are adapted from Bonafous et al. (2016), Delacroix et al. (2013), Doelman et al. (2020), and Llop-Sayson et al. (2020).

et al., 2020), which is an off-axis telescope built specifically to accommodate a vortex coronagraph.

Pupil-plane coronagraphs

One final category of coronagraphs are the pupil-plane coronagraphs. These do not share the same optical layout as the Lyot coronagraph, but instead consist of only a single pupil-plane apodizer and no focal-plane mask or Lyot stop. This mask is designed in such a way as to suppress the light in a region of interest around the star. As the planet PSF is the same as the stellar PSF, the mask additionally needs to maximize the amount of light in the core of the PSF to retain planet throughput. While this may seem a step back from the Lyot coronagraph, their significantly reduced optical complexity and ability to work for arbitrary telescope pupil geometries lets them see widespread adoption on current ground based telescopes, for example on Magellan/MagAO (Otten et al., 2017), Subaru/SCExAO (Doelman et al., 2017), VLT/ERIS (Boehle et al., 2018; Kenworthy et al., 2018b), LBT/LMIRcam (Doelman et al., 2020), HiCIBaS (Marchis et al., 2018), and on the future ELT instrument METIS (Kenworthy et al., 2018a).

Shaped pupil coronagraphs attempt to do the apodization by only modifying the amplitude of the light. Early apodizers were based on analytical expressions (Kasdin et al., 2003), while later global optimization methods were developed to derive the best possible apodizers (Carlotti et al., 2011). Inherent to introducing only amplitude apodization in the pupil, the region of interest necessarily has to be point-symmetric around the star. While never deployed on any major high-contrast imaging instruments, this coronagraph was instrumental to the development of the APLC described above. Additionally, it provided the coronagraph for early wavefront control experiments.

The apodizing phase plate (APP) coronagraph performs the apodization by only modifying the phase of the light. Early apodizers were designed using iterated Fourier transforms (Codona et al., 2006), while later global optimization were developed to derive the best possible apodizers (Por, 2017), described in Chapter 2 of this thesis. Phase only apodizers are now proven to always perform better than amplitude apodizers, albeit often only by a small margin. Contrary to the shaped pupil coronagraphs, the APP coronagraph can produce PSFs with a region of interest on only one side of the star, which provides a higher planet throughput at the expense of a smaller field of view. Early implementations were manufactured by using diamond turning to directly write the phase pattern of the apodizer

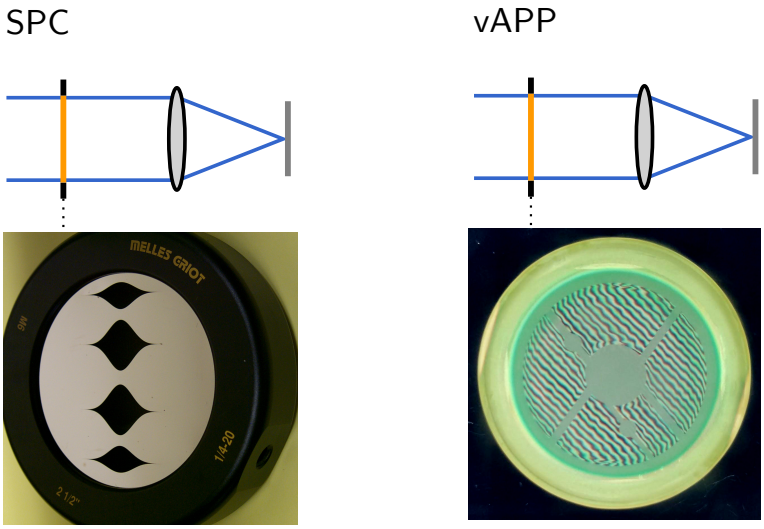


Figure 1.14: Schematic optical layouts of the two types of pupil-plane coronagraphs. The photographic images are adapted from Kasdin et al. (2004) and Doelman et al. (2017).

in a glass substrate Kenworthy et al. (2007) and discovered HD100546b (Quanz et al., 2013) and produced the first thermal light images of Beta Pictoris b (Quanz et al., 2010). The development of vectorial phase plates by Otten et al. (2017); Snik et al. (2012) made these obsolete, providing efficient achromatic phase apodization over a full photometric band.

1.3.2 The wavefront control system

The wavefront control system, or adaptive optics system as it is usually called on ground-based telescopes, corrects for any wavefront distortion in real time, whether originating from the turbulence in the atmosphere or the instrument optics (Babcock, 1953). It uses a wavefront sensor (WFS) to measure the incoming aberration, and a wavefront corrector to apply the opposite aberration, effectively flattening the wavefront. Due to the limited number of actuators on the deformable mirror – typically a few thousand on current generation devices – low spatial frequencies will be well corrected, while high spatial frequencies cannot be controlled. This results in a circular or square *control region* on the image where the PSF is well corrected, while outside this region no correction can be discerned.

The wavefront sensor typically uses shorter wavelength light split off by

a beam splitter from the longer wavelength light in the science beam, which yields a wavefront measurement that is close to, but not exactly equal to, the wavefront at the science instrument. The difference between the wavefront at the wavefront sensor and at the science instrument is called *non-common-path aberration* (NCPA). NCPAs are dynamic, and can limit the performance of the instrument when not removed well enough (Martinez et al., 2013). In addition to the fast wavefront sensor upstream, current solutions favour additional focal-plane wavefront sensing at the science instrument (Jovanovic et al., 2018) or a secondary wavefront sensor situated as close as possible to, or preferably even integrated into, the coronagraph (N'Diaye et al., 2013; Shi et al., 2018).

There are many different options for the wavefront sensor. One of the most commonly-used is the *Shack-Hartmann WFS* (SH-WFS; Hartmann 1900; Shack 1971), shown schematically in Figure 1.15. This WFS uses a microlens array in a conjugate pupil plane, creating a small image of the star for each lenslet. By measuring the shift of each spot from the on-axis position, we measure the local tilt of the wavefront at that lenslet and therefore are able to recover the wavefront in the pupil by use of an algorithmic transformation. Due to its simplicity and dynamic range, the SH-WFS has been implemented in a large number of adaptive optics systems (Beuzit et al., 2019; Lenzen et al., 2003; Macintosh et al., 2014; Rousset et al., 2003; Wizinowich et al., 2000).

Two other WFSs deserve mentioning. The *Pyramid WFS* (PyWFS; Ragazzoni 1996) uses a prism in the focal plane to split up the image into multiple (typically four) pupils. The wavefront can be reconstructed from these pupil images. The PyWFS has a much higher sensitivity, but requires sophisticated prisms (Lardière et al., 2017; Tozzi et al., 2008) and trades in dynamic range (Burvall et al., 2006; Ragazzoni, 1996). Better reconstruction algorithms seem to alleviate the latter problem at the expense of computational power (Frazin, 2018; Landman & Haffert, 2020). The PyWFS has been implemented on several HCI instruments, eg. Subaru/SCExAO (Jovanovic et al., 2015) and LBT/FLAO (Esposito et al., 2010)), and is planned to be implemented on most HCI instruments under development, eg. Magellan/MagAO-X (Males et al., 2018) and ELT/METIS (Brandl et al., 2016).

Finally, the *Zernike WFS* (Zernike, 1935) uses a phase dimple in the focal plane to produce a single pupil image in which the phase is encoded in the amplitude. This WFS has an optimal sensitivity (Guyon, 2005), but a tiny dynamic range. Amplitude aberrations need to be calibrated

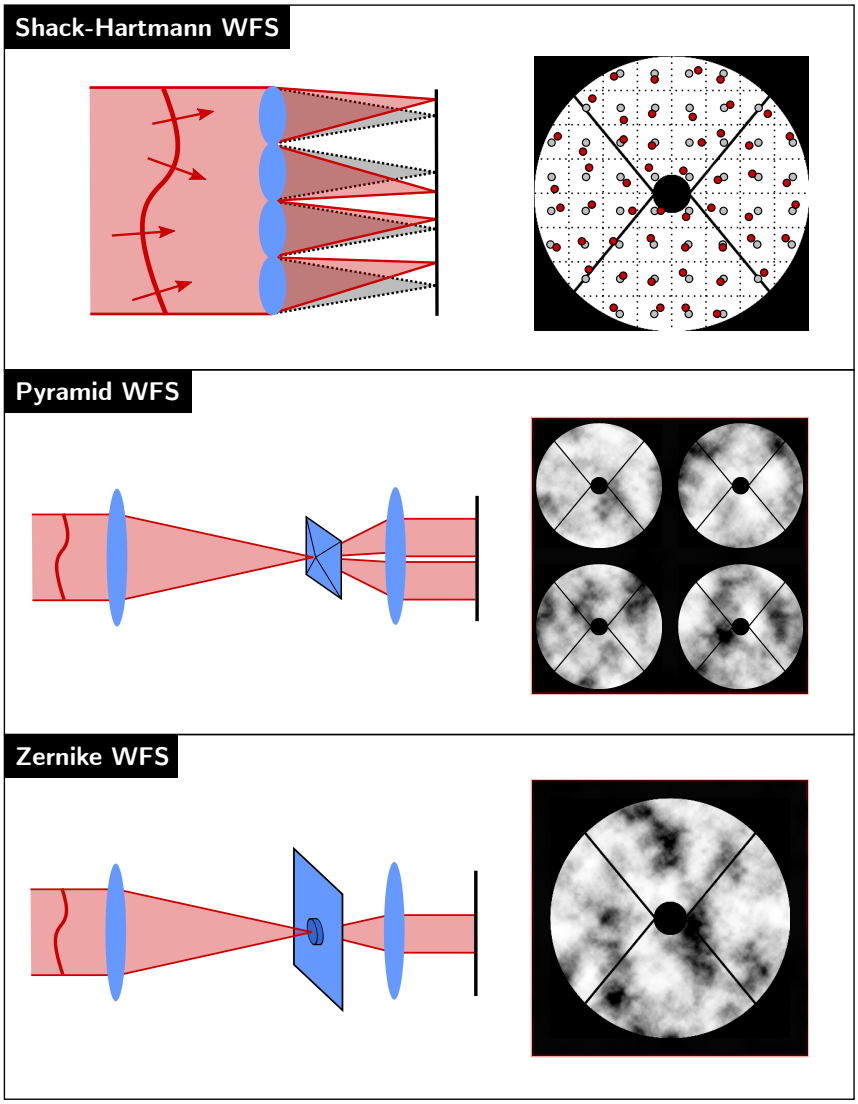


Figure 1.15: Three common wavefront sensors: the Shack-Hartmann, Pyramid and Zernike wavefront sensor. A schematic optical layout is shown on the left, and an example output for an aberrated wavefront on the right, for each of the wavefront sensors.

out, which is usually done sequentially (N'Diaye et al., 2013). Solutions to this problem have been proposed (Doelman et al., 2019), though not widely implemented as of yet. Despite these hurdles, the Zernike WFS is an attractive candidate for removing slowly-evolving aberrations, such as those due to NCPAs (N'Diaye et al., 2013; Vigan et al., 2019) or at space-based observatories (Pueyo et al., 2017).

1.3.3 Image post processing

From a photon noise perspective, separating planet light from star light should be done optically and before detection of the individual photons. However it is still possible to improve upon the raw contrast in our images with post-facto image processing methods, essentially for free provided that we observe in a manner amendable to the chosen post-processing method. The main theme in most of these methods comes back to the fact that the stellar light properties differ from the planet light properties (coherence, spectral energy distributions, position on the sky with respect to each other).

The stellar image is assumed to be stationary between images, or only vary in a limited number of degrees of freedom, presumably related to the degrees of freedom in the optical system and long-term atmospheric optical transfer function. This allows us to build a model of the stellar image without contributions from planet or disk photons, and subtract that from our observed image(s). This however requires some type of *diversity*, where the planet light behaves differently than the star light in some way.

- Reference Differential Imaging (RDI; Choquet et al. 2015; Lafrenière et al. 2009) uses a series of reference images of known single stars with similar spectral energy distributions, taken with the same instrument, to provide the stellar reference image.
- Angular Differential Imaging (ADI; Marois et al. 2006) uses the rotation of the sky, and therefore also the planet and disk, with respect to the telescope.
- Spectral Differential Imaging (SDI; Smith 1987) uses the difference in spectral signature between star and planet light.
- Polarization Differential Imaging (PDI; Kuhn et al. 2001) uses the different in polarization signature between star and planet light.

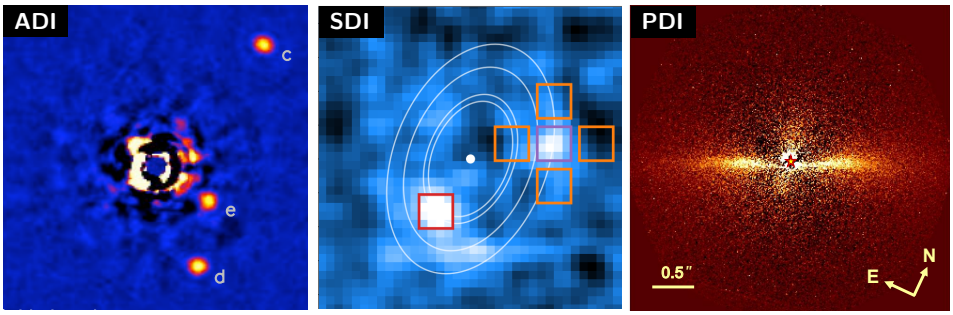


Figure 1.16: Example output images for different image post-processing techniques. (a) The discovery image of HR8799e, the fourth planet in this system using ADI (Marois et al., 2010). (b) The discovery of PDS 70c using its H-alpha emission with SDI (Haffert et al., 2019). (c) The HIP 79977 debris disk in polarized light using PDI (Engler et al., 2017).

- Coherent Differential Imaging (CDI; Guyon 2004) uses the fact that star light can interfere constructively and destructively with the stellar core, while planet and disk light can not.
- Finally, Orbital Differential Imaging (ODI; Males et al. 2015) uses the movement of the planet during its orbit over time as diversity signal.

Of course, multiple diversities can be combined to further improve the retrieved stellar reference image (eg. Christiaens et al. 2019 and Flasseur et al. 2020). In Figure 1.16, I show a few examples of images retrieved with these data processing techniques, typically gaining one or two decades in contrast on top of the raw contrast of the raw images.

1.4 This thesis

This thesis explores possible novel approaches to high-contrast imaging instruments, concerning coronagraphy, wavefront sensing and the tools that we use to study the complex interplay between these systems.

Chapter 2: optimal design of apodizing phase plate

Chapter 2 presents the development of an optimal optimization method for apodizing phase plate coronagraphs. As this optimization method can be proven to be globally optimal, that is, it will always find the best possible

solution to the problem, it supersedes previous methods that do not possess this property. I derive the fundamental limits for the APP coronagraph.

Chapter 3 & 4: the single-mode complex amplitude refinement coronagraph

Chapter 3 presents the theoretical development of the single-mode complex amplitude refinement (SCAR) coronagraph. This novel coronagraph concept consists of a pupil-plane phase plate and a microlens-fed single-mode fiber array in the focal plane. The phase plate is designed such that the contrast is achieved after the single-mode fibers instead of before, as is the convention for most coronagraphs. This additional filtering significantly improves performance, allowing us to search for exoplanets closer to the diffraction limit compared to the APP coronagraph.

Chapter 4 presents the first laboratory demonstration of the SCAR coronagraph. We demonstrate a 1×10^{-4} contrast at $1\lambda/D$ using manufactured SCAR liquid-crystal phase plates. We also perform a Monte-Carlo tolerance analysis for observing the exoplanet Proxima b using the VLT/SPHERE instrument.

Chapter 5: high-contrast imaging for Python

Chapter 5 presents the development of the software package High-Contrast Imaging for Python (HCIPy). HCIPy is an open-source modular Python package for high-contrast imaging on current and future telescopes. By integrating atmospheric simulation, wavefront sensing and coronagraphy in a single library, HCIPy enables both rapid prototyping of individual system elements and the full high-contrast imaging system.

Chapter 6: the asymmetric wind-driven halo

Chapter 6 describes an explanation for the asymmetry in the wind-driven halo observed in high-contrast imaging systems. The delay between the measurement of wavefront aberrations in the atmosphere and the correction of those aberrations with the deformable mirror in the adaptive optics system causes a butterfly-like halo to appear in high-contrast images. Interference with scintillation speckles caused by high-altitude turbulence causes one wing of the butterfly to brighten and the other to dim, making the image asymmetric.

Chapter 7 & 8: the phase-apodized-pupil Lyot coronagraph

Chapter 7 presents the theoretical development of the phase-apodized-pupil Lyot coronagraph. This novel coronagraph consists of a pupil-plane phase plate, a focal-plane mask and a Lyot stop. Using a knife-edge focal-plane mask and a one-sided dark zone, one can achieve high-throughput designs for space-based telescopes at small inner working angles. We explore the design parameters and present several designs for realistic telescope pupils.

Chapter 8 presents the first laboratory demonstration of the PAPLC on the *Très Haute Dynamique 2* (THD2) testbed at the Observatoire de Paris. It also describes and tests an integrated high-order wavefront sensor using only light reflected by the focal-plane mask. We demonstrate raw contrasts comparable to other coronagraphs on the THD2, and wavefront sensing performance comparable with the fundamental photon noise limit.

1.5 Future outlook

The wide variety of developed coronagraphs, wavefront sensors and post-processing methods is a testament to the enthusiasm of instrumentalists and the discovery of thousands of exoplanets. The first generation of high-contrast imaging instruments are currently being upgraded to use new coronagraphs and AO systems. GPI 2.0 (Chilcote et al., 2018), SPHERE+ (Boccaletti et al., 2020) and the continually refined SCExAO will be seeing more on-sky operation in the following years, both in testing and in scientific capacity. Additionally, preparations are well underway for the Roman Space Telescope, which should be the first major test of high contrast imaging from space with a wavefront control system. While only a technology demonstrator, a successful deployment and operation should pave the way towards more ambitious projects, such as HabEx, which could provide around ten spectra of an exo-Earths (Gaudi et al., 2020), or LUVOIR, which could provide several tens (The LUVOIR Team, 2019).

As always, we must take care in comparing the resulting performance characteristics between different instruments and even subsystems, for two reasons. First, performance is highly dependent on environmental conditions – seeing conditions, phenomena such as the low-wind effect (Milli et al., 2018), and vibrational or thermal environment – can significantly influence the results, and the limited time available on world-class facilities makes it hard to obtain data under comparable conditions. Secondly, the interplay between different subsystems is crucial, so comparison needs to be done at the system level, not the sub-system level, which is not trivial.

Science output, that is, exoplanet candidate yield, should be our ultimate metric, but we have not found enough to yet place useful constraints. We are limited to comparing contrast curves of sensitivity versus angular separations,and sometimes even this is condensed into a single quantity. Unless standardized, the resulting curves are meaningless by themselves and require extensive knowledge of both systems to be comparable. First attempts at standardization of contrast have been made (Marois et al., 2008; Pairet et al., 2019; Ruane et al., 2018; Ruffio et al., 2018), which highlight the complications of providing a sufficiently good quantity for comparison. Unless these are adopted by the community, we risk the pitfall of adding yet another metric to the stack.

Simulations might be the only way to fairly compare coronagraph and wavefront control systems. The current mission of the Segmented Coronagraphs Design and Analysis (SCDA) study is to link exoplanet candidate yields to aberration sensitivity, especially those for giant space-based segmented telescopes. In doing so, it is developing a standardized way of sharing coronagraph designs, and (re)formalizing the interfaces to software packages capable of calculating exoplanet candidate yields (such as AYO, described above). This effort should streamline collaboration between research groups and allow for simpler comparison between coronagraph solutions.

While SCDA will provide excellent empirical knowledge on how aberrations impact exoplanet candidate yield, it does not provide an explanatory knowledge. Belikov et al. (2019) provides an initial step for perfect coronagraphs, which will become increasingly relevant as coronagraph designs progress towards this level of perfection.

An open question is still to what degree the coronagraph, wavefront sensor and wavefront control system can help each other. One option could be for the coronagraph to extract photons for high-order wavefront sensing in a more smart way than the conventional beam splitter, without affecting the coronagraphic image; another option is to tune the wavefront control system to allow more power in modes that are better suppressed by the coronagraph. Both may be combined for further improvements.

For both ground- and space-based instruments, predictive control will be a groundbreaking advancement, as it allows for a significantly more efficient use of the available photons. By using spatial-temporal correlations instead of just spatial correlations, it can provide a controlled wavefront beyond the sensing limit of the individual wavefront sensor frames. This improves raw contrast and also, perhaps more importantly, secondary effects

such as the asymmetry in the wind-driven halo (see Chapter 6) by temporally decorrelating the residual atmospheric speckles. Predictive control requires us to consistently nail the changing spatial-temporal correlations, either using brute-force based on wavefront sensor telemetry (Guyon & Males, 2017) of even better, based on this telemetry and a system model, where the model acts as prior knowledge to reduce the number of parameters that need to be estimated in real time. Additionally, this improved modeling can help us pinpoint which parts of our instrument need to be improved. While attempted for many years, recent advances in machine learning and computational performance bring this goal closer than ever.

These technological advances will enable future ground-based telescopes, such as the ELT, GMT and TMT, to characterize rocky exoplanets around nearby lower-mass stars in the near and mid-infrared. In addition to their orbital properties, photometric monitoring campaigns will enable the determination of their rotational periods (including giant gas planets) and multi-year campaigns of rocky exoplanets may reveal the presence of seasons and continents (Robinson et al., 2011). In combination with high resolution spectroscopy, biomarker gases may be detectable in their atmospheres. Smaller-aperture space-based telescopes (e.g. the Roman Space Telescope) will search for and characterise rocky planets around solar-type stars. Looking to the next few decades, four large space telescope missions are being proposed that will have the angular resolution to image and characterize dozens of exoplanets, enabling us to maybe answer whether there is life elsewhere in the Universe.

References

Absil, O., Mawet, D., Karlsson, M., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99080Q

Babcock, H. W. 1953, PASP, 65, 229

Bashi, D., Helled, R., Zucker, S., & Mordasini, C. 2017, A&A, 604, A83, doi: 10.1051/0004-6361/201629922

Bate, M. R. 1998, ApJ, 508, L95, doi: 10.1086/311719

Beichman, C., Stapelfeldt, K., Patel, R., Meshkat, T., & ygouf, m. 2018, in American Astronomical Society Meeting Abstracts, Vol. 231, American Astronomical Society Meeting Abstracts #231, 246.46

Belikov, R., Sirbu, D., Bendek, E., et al. 2019, in American Astronomical Society Meeting Abstracts, Vol. 233, American Astronomical Society Meeting Abstracts #233, 237.02

Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155, doi: 10.1051/0004-6361/201935251

Boccaletti, A., Chauvin, G., Mouillet, D., et al. 2020, arXiv e-prints, arXiv:2003.05714. https://arxiv.org/abs/2003.05714

Boehle, A., Glauser, A. M., Kenworthy, M. A., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII, 107023Y

Bonafous, M., Galicher, R., Baudoz, P., Firminy, J., & Boussaha, F. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9912, Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II, 99126J

Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977

Boss, A. P. 1997, Science, 276, 1836

Brandl, B. R., Agócs, T., Aitink-Kroes, G., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 990820

Burvall, A., Daly, E., Chamot, S. R., & Dainty, C. 2006, Optics Express, 14, 11925, doi: 10.1364/OE.14.011925

Carlotti, A. 2013, A&A, 551, A10, doi: 10.1051/0004-6361/201220661

Carlotti, A., Vanderbei, R., & Kasdin, N. J. 2011, Optics Express, 19, 26796, doi: 10.1364/OE.19.026796

Casertano, S., Lattanzi, M. G., Sozzetti, A., et al. 2008, A&A, 482, 699, doi: 10.1051/0004-6361:20078997

Cavarroc, C., Boccaletti, A., Baudoz, P., Fusco, T., & Rouan, D. 2006, A&A, 447, 397, doi: 10.1051/0004-6361:20053916

Chilcote, J. K., Bailey, V. P., De Rosa, R., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII, 1070244

Choquet, É., Pueyo, L., Soummer, R., et al. 2015, in Society of Photo-Optical

Instrumentation Engineers (SPIE) Conference Series, Vol. 9605, Techniques and Instrumentation for Detection of Exoplanets VII, 96051P

Christiaens, V., Casassus, S., Absil, O., et al. 2019, MNRAS, 486, 5819, doi: 10.1093/mnras/stz1232

Ciesla, F. J., & Cuzzi, J. N. 2006, Icarus, 181, 178, doi: 10.1016/j.icarus.2005.11.009

Claudi, R. U., Turatto, M., Gratton, R. G., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, 70143E

Codona, J. L., Kenworthy, M. A., Hinz, P. M., Angel, J. R. P., & Woolf, N. J. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6269, 62691N

Dawson, R. I., & Johnson, J. A. 2018a, ARA&A, 56, 175, doi: 10.1146/annurev-astro-081817-051853

—. 2018b, ARA&A, 56, 175, doi: 10.1146/annurev-astro-081817-051853

de Boer, J., Langlois, M., van Holstein, R. G., et al. 2020, A&A, 633, A63, doi: 10.1051/0004-6361/201834989

Delacroix, C., Absil, O., Forsberg, P., et al. 2013, A&A, 553, A98, doi: 10.1051/0004-6361/201321126

Doelman, D. S., Fagginger Auer, F., Escuti, M. J., & Snik, F. 2019, Optics Letters, 44, 17, doi: 10.1364/OL.44.000017

Doelman, D. S., Por, E. H., Ruane, G., Escuti, M. J., & Snik, F. 2020, PASP, 132, 045002, doi: 10.1088/1538-3873/ab755f

Doelman, D. S., Snik, F., Warriner, N. Z., & Escuti, M. J. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10400, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 104000U

Dominik, C., Blum, J., Cuzzi, J. N., & Wurm, G. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil, 783

Engler, N., Schmid, H. M., Thalmann, C., et al. 2017, A&A, 607, A90, doi: 10.1051/0004-6361/201730846

Esposito, S., Riccardi, A., Fini, L., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7736, Adaptive Optics Systems II, 773609

Fischer, D. A., Anglada-Escude, G., Arriagada, P., et al. 2016, PASP, 128, 066001, doi: 10.1088/1538-3873/128/964/066001

Flasseur, O., Denis, L., Thiébaut, É., & Langlois, M. 2020, A&A, 637, A9, doi: 10.1051/0004-6361/201937239

Fogarty, K., Pueyo, L., Mazoyer, J., & N'Diaye, M. 2017, AJ, 154, 240, doi: 10.3847/1538-3881/aa9063

Foo, G., Palacios, D. M., & Swartzland er, Grover A., J. 2005, Optics Letters, 30, 3308

Frazin, R. A. 2018, Journal of the Optical Society of America A, 35, 594, doi: 10.1364/JOSAA.35.000594

Fried, D. L. 1966, Journal of the Optical Society of America (1917-1983), 56, 1372

Ftaclas, C., Siebert, E. T., & Terrile, R. J. 1988, in Space Optics for Astrophysics and Earth and Planetary Remote Sensing, Vol. 10, 62–62
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1, doi: 10.1051/0004-6361/201629272
Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, arXiv e-prints, arXiv:2001.06683. https://arxiv.org/abs/2001.06683
Greaves, J. S., Richards, A. M. S., Bains, W., et al. 2020, Nature Astronomy, doi: 10.1038/s41550-020-1174-4
Greenwood, D. P. 1977, Journal of the Optical Society of America (1917-1983), 67, 390
Guerri, G., Daban, J.-B., Robbe-Dubois, S., et al. 2011, Experimental Astronomy, 30, 59, doi: 10.1007/s10686-011-9220-y
Guyon, O. 2003, A&A, 404, 379
—. 2004, ApJ, 615, 562, doi: 10.1086/423980
—. 2005, ApJ, 629, 592, doi: 10.1086/431209
Guyon, O., & Males, J. 2017, arXiv e-prints, arXiv:1707.00570. https://arxiv.org/abs/1707.00570
Guyon, O., Martinache, F., Belikov, R., & Soummer, R. 2010, ApJS, 190, 220, doi: 10.1088/0067-0049/190/2/220
Guyon, O., Pluzhnik, E. A., Kuchner, M. J., Collins, B., & Ridgway, S. T. 2006, ApJS, 167, 81, doi: 10.1086/507630
Haffert, S. Y., Bohn, A. J., de Boer, J., et al. 2019, Nature Astronomy, 3, 749, doi: 10.1038/s41550-019-0780-5
Hartmann, J. 1900, Zeitschrift für Instrumentenkunde, 20, 47
Hinkley, S., Oppenheimer, B. R., Soummer, R., et al. 2007, ApJ, 654, 633, doi: 10.1086/509063
Hollenbach, D. J., Yorke, H. W., & Johnstone, D. 2000, in Protostars and Planets IV, ed. V. Mannings, A. P. Boss, & S. S. Russell, 401–428
Ikoma, M., Nakazawa, K., & Emori, H. 2000, ApJ, 537, 1013, doi: 10.1086/309050
Johansen, A., Klahr, H., & Henning, T. 2011, A&A, 529, A62, doi: 10.1051/0004-6361/201015979
Johansen, A., Mac Low, M.-M., Lacerda, P., & Bizzarro, M. 2015, Science Advances, 1, 1500109, doi: 10.1126/sciadv.1500109
Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, PASP, 127, 890, doi: 10.1086/682989
Jovanovic, N., Absil, O., Baudoz, P., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, 107031U
Kasdin, N. J., Vanderbei, R. J., Littman, M. G., Carr, M., & Spergel, D. N. 2004, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5487, Optical, Infrared, and Millimeter Space Telescopes, ed. J. C. Mather, 1312–1321
Kasdin, N. J., Vanderbei, R. J., Spergel, D. N., & Littman, M. G. 2003, ApJ, 582, 1147
Kenworthy, M. A., Absil, O., Carlomagno, B., et al. 2018a, in Society of

Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII, 10702A3
Kenworthy, M. A., Codona, J. L., Hinz, P. M., et al. 2007, ApJ, 660, 762, doi: 10.1086/513596
Kenworthy, M. A., Snik, F., Keller, C. U., et al. 2018b, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII, 1070246
Knight, J. M., Guyon, O., Lozi, J., Jovanovic, N., & Males, J. R. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10706, Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III, 107065O
Kratter, K., & Lodato, G. 2016, ARA&A, 54, 271, doi: 10.1146/annurev-astro-081915-023307
Krissansen-Totton, J., Bergsman, D. S., & Catling, D. C. 2016, Astrobiology, 16, 39, doi: 10.1089/ast.2015.1327
Kuchner, M. J., & Traub, W. A. 2002, ApJ, 570, 900, doi: 10.1086/339625
Kuhn, J. R., Potter, D., & Parise, B. 2001, ApJ, 553, L189, doi: 10.1086/320686
Lafrenière, D., Marois, C., Doyon, R., & Barman, T. 2009, The Astrophysical Journal, 694, L148, doi: 10.1088/0004-637x/694/2/l148
Landman, R., & Haffert, S. Y. 2020, Optics Express, 28, 16644, doi: 10.1364/OE.389465
Lardière, O., Pazder, J., Véran, J.-P., & van Kooten, M. 2017, Adaptive Optics for Extremely Large Telescopes V (AO4ELT5)
Larkin, J. E., Chilcote, J. K., Aliado, T., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 91471K
Lenzen, R., Hartung, M., Brandner, W., et al. 2003, in Proc. SPIE, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 944–952
Libby-Roberts, J. E., Berta-Thompson, Z. K., Désert, J.-M., et al. 2020, AJ, 159, 57, doi: 10.3847/1538-3881/ab5d36
Llop-Sayson, J., Ruane, G., Mawet, D., et al. 2020, AJ, 159, 79, doi: 10.3847/1538-3881/ab6329
Lovis, C., Snellen, I., Mouillet, D., et al. 2017, A&A, 599, A16
Lozi, J., Martinache, F., & Guyon, O. 2009, PASP, 121, 1232, doi: 10.1086/648392
Luque, R., Nowak, G., Pallé, E., et al. 2019, A&A, 623, A114, doi: 10.1051/0004-6361/201834952
Lyot, B. 1939, MNRAS, 99, 580
Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proceedings of the National Academy of Science, 111, 12661
Males, J. R., Belikov, R., & Bendek, E. 2015, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9605, Techniques and Instrumentation for Detection of Exoplanets VII, 960518
Males, J. R., Close, L. M., Miller, K., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics

Systems VI, 1070309
Marchis, F., Thibault, S., Côté, O., et al. 2018, in AGU Fall Meeting Abstracts, Vol. 2018, P41C–3747
Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, ApJ, 641, 556
Marois, C., Lafrenière, D., Macintosh, B., & Doyon, R. 2008, ApJ, 673, 647, doi: 10.1086/523839
Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010, Nature, 468, 1080, doi: 10.1038/nature09684
Martinez, P., Kasper, M., Costille, A., et al. 2013, A&A, 554, A41
Matsumoto, T., & Hanawa, T. 2003, ApJ, 595, 913, doi: 10.1086/377367
Mawet, D., Pueyo, L., Carlotti, A., et al. 2013a, ApJS, 209, 7, doi: 10.1088/0067-0049/209/1/7
Mawet, D., Riaud, P., Absil, O., & Surdej, J. 2005, ApJ, 633, 1191, doi: 10.1086/462409
Mawet, D., Pueyo, L., Lawson, P., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8442, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, 844204
Mawet, D., Absil, O., Delacroix, C., et al. 2013b, A&A, 552, L13, doi: 10.1051/0004-6361/201321315
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
Milli, J., Banas, T., Mouillet, D., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9909, Adaptive Optics Systems V, 99094Z
Milli, J., Kasper, M., Bourget, P., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, 107032A
Moody, D. C., & Trauger, J. T. 2007, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6693, Proc. SPIE, 66931I
Murakami, N., Uemura, R., Baba, N., et al. 2008, PASP, 120, 1112, doi: 10.1086/592285
N'Diaye, M., Dohlen, K., Fusco, T., & Paul, B. 2013, A&A, 555, A94, doi: 10.1051/0004-6361/201219797
N'Diaye, M., Soummer, R., Pueyo, L., et al. 2016, ApJ, 818, 163, doi: 10.3847/0004-637X/818/2/163
Newman, K., Conway, J., Belikov, R., & Guyon, O. 2016, PASP, 128, 055003, doi: 10.1088/1538-3873/128/963/055003
Öberg, K. I., Murray-Clay, R., & Bergin, E. A. 2011, ApJ, 743, L16, doi: 10.1088/2041-8205/743/1/L16
Otten, G. P. P. L., Snik, F., Kenworthy, M. A., et al. 2017, ApJ, 834, 175
Paardekooper, S. J., & Mellema, G. 2006, A&A, 459, L17, doi: 10.1051/0004-6361:20066304
Pairet, B., Cantalloube, F., Gomez Gonzalez, C. A., Absil, O., & Jacques, L. 2019, MNRAS, 487, 2262, doi: 10.1093/mnras/stz1350
Piro, A. L., & Vissapragada, S. 2020, AJ, 159, 131, doi: 10.3847/1538-3881/

ab7192
Por, E. H. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10400, Proc. SPIE, 104000V
Pueyo, L., Zimmerman, N., Bolcar, M., et al. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10398, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 103980F
Quanz, S. P., Amara, A., Meyer, M. R., et al. 2013, ApJ, 766, L1, doi: 10.1088/2041-8205/766/1/L1
Quanz, S. P., Meyer, M. R., Kenworthy, M. A., et al. 2010, ApJ, 722, L49, doi: 10.1088/2041-8205/722/1/L49
Ragazzoni, R. 1996, Journal of Modern Optics, 43, 289, doi: 10.1080/09500349608232742
Rando, N., Asquier, J., Corral Van Damme, C., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, 106980K
Rappaport, S., Levine, A., Chiang, E., et al. 2012, ApJ, 752, 1, doi: 10.1088/0004-637X/752/1/1
Rauer, H., Catala, C., Aerts, C., et al. 2014, Experimental Astronomy, 38, 249, doi: 10.1007/s10686-014-9383-4
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, in Proc. SPIE, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, 914320
Robinson, T. D., Meadows, V. S., Crisp, D., et al. 2011, Astrobiology, 11, 393, doi: 10.1089/ast.2011.0642
Roddier, F., & Roddier, C. 1997, PASP, 109, 815, doi: 10.1086/133949
Rodenhuis, M., Snik, F., van Harten, G., Hoeijmakers, J., & Keller, C. U. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9099, Polarization: Measurement, Analysis, and Remote Sensing XI, 90990L
Rodin, A. E. 2019, Astronomy Reports, 63, 224, doi: 10.1134/S1063772919030090
Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., & Labeyrie, A. 2000, PASP, 112, 1479
Rousset, G., Lacombe, F., Puget, P., et al. 2003, in Proc. SPIE, Vol. 4839, Adaptive Optical System Technologies II, ed. P. L. Wizinowich & D. Bonaccini, 140–149
Ruane, G., Jewell, J., Mawet, D., Pueyo, L., & Shaklan, S. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9912, Proc. SPIE, 99122L
Ruane, G., Mawet, D., Jewell, J., & Shaklan, S. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10400, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 104000J
Ruane, G., Riggs, A., Mazoyer, J., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, 106982S
Ruffio, J.-B., Mawet, D., Czekala, I., et al. 2018, AJ, 156, 196, doi: 10.3847/

1538-3881/aade95
Sahlmann, J., Martín-Fleitas, J., Mora, A., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9904, Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, 99042E
Sahlmann, J., Mora, A., Martín-Fleitas, J. M., et al. 2018, in IAU Symposium, Vol. 330, Astrometry and Astrophysics in the Gaia Sky, ed. A. Recio-Blanco, P. de Laverny, A. G. A. Brown, & T. Prusti, 343–344
Schmid, H. M., Bazzon, A., Roelfsema, R., et al. 2018, A&A, 619, A9, doi: 10.1051/0004-6361/201833620
Schwieterman, E. W., Kiang, N. Y., Parenteau, M. N., et al. 2018, Astrobiology, 18, 663, doi: 10.1089/ast.2017.1729
Serabyn, E., Mawet, D., & Burruss, R. 2010, Nature, 464, 1018, doi: 10.1038/nature09007
Serabyn, E., Huby, E., Matthews, K., et al. 2017, AJ, 153, 43, doi: 10.3847/1538-3881/153/1/43
Shack, R. V. 1971, J. Opt. Soc. Am., 61, 656
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 500, 33
Shi, F., Seo, B.-J., Cady, E., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Proc. SPIE, 106982O
Shu, F. H. 1977, ApJ, 214, 488, doi: 10.1086/155274
Sivaramakrishnan, A., Soummer, R., Oppenheimer, B. R., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, 773586
Smith, W. H. 1987, PASP, 99, 1344, doi: 10.1086/132124
Snik, F., Otten, G., Kenworthy, M., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8450, Modern Technologies in Space- and Ground-based Telescopes and Instrumentation II, 84500M
Soummer, R. 2005, ApJ, 618, L161
Soummer, R., Aime, C., & Falloon, P. E. 2003a, A&A, 397, 1161
Soummer, R., Dohlen, K., & Aime, C. 2003b, A&A, 403, 369, doi: 10.1051/0004-6361:20030278
Soummer, R., Brady, G. R., Brooks, K., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, 106981O
Stark, C. C., Roberge, A., Mandell, A., & Robinson, T. D. 2014, ApJ, 795, 122, doi: 10.1088/0004-637X/795/2/122
Suleymanova, S. A., & Rodin, A. E. 2014, Astronomy Reports, 58, 796, doi: 10.1134/S1063772914110067
The LUVOIR Team. 2019, arXiv e-prints, arXiv:1912.06219. https://arxiv.org/abs/1912.06219
Tinyanont, S., Millar-Blanchaer, M. A., Nilsson, R., et al. 2019, PASP, 131, 025001, doi: 10.1088/1538-3873/aaef0f
Tozzi, A., Stefanini, P., Pinna, E., & Esposito, S. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7015, Adaptive Op-

tics Systems, 701558
Trauger, J., Moody, D., Krist, J., & Gordon, B. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011013, doi: 10.1117/1.JATIS.2.1.011013
Tucker, C. J. 1979, Remote Sensing of Environment, 8, 127, doi: 10.1016/0034-4257(79)90013-0
Udry, S., & Santos, N. C. 2007, ARA&A, 45, 397, doi: 10.1146/annurev.astro.45.051806.110529
van Lieshout, R., Min, M., Dominik, C., et al. 2016, A&A, 596, A32, doi: 10.1051/0004-6361/201629250
Vigan, A., N'Diaye, M., Dohlen, K., et al. 2019, A&A, 629, A11, doi: 10.1051/0004-6361/201935889
Windmark, F., Birnstiel, T., Güttler, C., et al. 2012, A&A, 540, A73, doi: 10.1051/0004-6361/201118475
Winn, J. N., & Fabrycky, D. C. 2015, ARA&A, 53, 409, doi: 10.1146/annurev-astro-082214-122246
Wizinowich, P., Acton, D. S., Shelton, C., et al. 2000, PASP, 112, 315
Wolszczan, A. 1994, Science, 264, 538, doi: 10.1126/science.264.5158.538
Wolszczan, A., & Frail, D. A. 1992, Nature, 355, 145
Zernike, F. 1935, Tech. Phys, 16, 454
Zimmerman, N. T., Eldorado Riggs, A. J., Jeremy Kasdin, N., Carlotti, A., & Vanderbei, R. J. 2016a, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011012, doi: 10.1117/1.JATIS.2.1.011012
Zimmerman, N. T., N'Diaye, M., St. Laurent, K. E., et al. 2016b, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9904, Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, 99041Y

Chapter 2

Optimal design of apodizing phase plate coronagraphs

Adapted from
E. H. Por (2017), Proc. SPIE 10400

Abstract
Direct observations of exoplanets require a stellar coronagraph to suppress the diffracted starlight. An Apodizing Phase Plate (APP) coronagraph consists of a carefully designed phase-only mask in the pupil plane of the telescope. This mask alters the point spread function in such a way that it contains a dark zone at some off-axis region of interest, while still retaining a high Strehl ratio (and therefore high planet throughput).

Although many methods for designing such a phase mask exist, none of them provide a guarantee of global optimality. Here we present a method based on generalization of the phase-only mask to a complex amplitude mask. Maximizing the Strehl ratio while simultaneously constraining the stellar intensity in the dark zone turns out to be a quadratically constrained linear algorithm, for which a global optimum can be found using large-scale numerical optimizers. This generalized problem yields phase-only solutions. These solutions are therefore also solutions of the original problem.

Using this optimizer we perform parameter studies on the inner and outer working angle, the contrast and the size of the secondary obscuration of the telescope aperture for both one-sided and annular dark zones. We reach Strehl ratios of $> 65\%$ for a 10^{-5} contrast from 1.8 to $10\lambda/D$ with a one-sided dark zone for a VLT-like secondary obscuration. This study provides guidelines for designing APPs for more realistic apertures.

2.1 Introduction

For detecting Earth-like exoplanets, we need to overcome the huge contrast between planet and its host star. A coronagraph allows for separation of starlight and the light from its companion by suppressing the former and transmitting the latter. This suppression can be achieved by various types of optical systems (Guyon et al., 2006; Mawet et al., 2012). One of these coronagraphic systems is known as the apodizing phase plate (APP), which consists of a single phase-only pupil-plane optic (Codona et al., 2006). In this type of coronagraph, the starlight itself is not absorbed, but rather suppressed in a certain region of interest in the focal-plane known as the dark zone. The phase pattern on the pupil-plane element must therefore be designed to yield a point-spread function (PSF) that has extremely little light inside this dark zone.

Of course, the phase plate reduces the Strehl ratio of the star, as light is scattered out of the Airy core. Although we do not care about the Strehl ratio of the star, as the planet light is also incident on the same phase plate, its PSF is altered in the same way and its Strehl ratio is diminished by the same factor. We therefore need to simultaneously have a high Strehl ratio of the phase plate, while having the transmission inside the dark zone be extremely small.

We can therefore conclude that finding the phase pattern requires solving the following optimization problem:

$$\underset{\phi(x)}{\text{maximize}} \quad |E(0)|^2 \tag{2.1}$$

$$\text{subject to} \quad |E(k)|^2 \leq |E(0)|^2 \cdot 10^{-c} \;\forall\; k \in D \tag{2.2}$$

where $E(k)$ is the electric field in the focal plane, defined by

$$E(k) = \mathcal{F}\{A(x) \exp i\phi(x)\} \tag{2.3}$$

and c is the contrast in the dark zone that we want to reach, $A(x)$ the telescope aperture, $\phi(x)$ the phase pattern in the pupil plane, x a position in the pupil plane, k a position in the focal plane, and D the dark zone in the focal plane. Note that there are many phase patterns that satisfy the constraint of contrast in the dark zone, however we want to select the phase pattern that simultaneously yields the highest planet throughput.

Solving this optimization problem is quite hard due to the non-linearity in the complex phase exponential. Previous methods have therefore not attempted to solve the full optimization problem, but rather to find a phase

solution that is close to the global optimum. These algorithms often linearize the problem around a certain phase pattern and take small steps towards maximization of some merit function that includes both the planet throughput and contrast in the dark zone. This includes phase iteration techniques (Codona et al., 2006; Codona & Angel, 2004), which were one of the first proposed methods for APP optimization, and modified Gerchberg-Saxton algorithms (Ruane et al., 2015). Due to the linearization involved in these optimization techniques, they will often get stuck in local optima and give no guarantee that they converge to a solution close to the optimal one. Additionally these algorithms do not constrain the intensity in the dark zone, but rather minimize that intensity without any regard for the set contrast limit.

Finding the global optimum has been previously attempted using general non-linear optimization methods such as simulated annealing. While these methods in theory provide the global optimum, they require vast amounts of computation time and only provide a guarantee of global optimality if given an infinite amount of computation time. Global optimization of shaped pupil coronagraphs (Carlotti et al., 2011) has been modified to optimize APPs (Carlotti et al., 2013), however the resulting APPs contained only several discrete phase transitions, and were not globally optimal, as will be shown in this work. In this paper we slightly modify the approach of Carlotti et al. (2013) and provide several improvements, both in speed and correctness at small inner working angles. We then present the fundamental limits of pupil-plane-only coronagraph designs, using parameter studies on simplified apertures using both one-sided and two-sided dark zones.

2.2 Linearization, discretization and correction

2.2.1 Linearization

The linearization of the general optimization problem follows closely the work by Carlotti et al. (2013), but differs in a few key points. These will be indicated in the following derivation.

We first change the complex phase exponential to a complex function $X(x)+iY(x)$ where we require $X^2(x)+Y^2(x) = 1$ and change optimization variables to finding the functions $X(x)$ and $Y(x)$. Note that this linearizes the calculation of $E(k)$ but does not change anything in the optimization problem: the non-linearity is now hidden in the constraints. The optimiza-

tion problem now reads

$$\begin{aligned} \underset{X(x),Y(x)}{\text{maximize}} \quad & |E(0)|^2 && (2.4)\\ \text{subject to} \quad & |E(k)|^2 \leq |E(0)|^2 \cdot 10^{-c} \; \forall \; k \in D && (2.5)\\ & E(k) = \mathcal{F}\{A(x)(X(x) + iY(x))\} && (2.6)\\ & X^2(x) + Y^2(x) = 1 \forall x. && (2.7) \end{aligned}$$

To make the objective function linear, we can simply maximize the real value of $E(0)$ instead of its absolute value. While this may seem like an oversimplification, it only fixes the average phase of the pupil-plane phase pattern. Any other chosen phase (ie. optimizing for the imaginary part of $E(0)$ instead) can be obtained by multiplying the pupil-plane complex amplitude pattern by a constant complex phase exponential (ie. $\exp(i\pi/2)$ for the case mentioned above). This is possible as all constraints depend on the amplitude of some complex amplitude in the pupil or focal plane only and have no dependence on the phase of that quantity. The optimization problem now reads

$$\begin{aligned} \underset{X(x),Y(x)}{\text{maximize}} \quad & \Re\{E(0)\} && (2.8)\\ \text{subject to} \quad & |E(k)|^2 \leq |E(0)|^2 \cdot 10^{-c} \; \forall \; k \in D && (2.9)\\ & E(k) = \mathcal{F}\{A(x)(X(x) + iY(x))\} && (2.10)\\ & X^2(x) + Y^2(x) = 1 \forall x. && (2.11) \end{aligned}$$

The objective is now linear, and all constraints are convex and quadratic except for the constraint on the absolute value of the complex amplitude in the pupil-plane. To make this convex, we generalize the optimization problem to finding an optimal complex amplitude mask, instead of a phase-only mask. The new optimization problem does not guarantee phase-only solutions, however empirically this generalized optimization problem does yield phase-only solutions for all dark zone shapes, contrast limits and aperture geometries. Various examples of APP solutions for one-sided and two-sided dark zones, both for simple and complex aperture geometries, can be found in Figures 2.3 and 2.6. The optimization problem now reads

$$\begin{aligned} \underset{X(x),Y(x)}{\text{maximize}} \quad & \Re\{E(0)\} && (2.12)\\ \text{subject to} \quad & |E(k)|^2 \leq |E(0)|^2 \cdot 10^{-c} \; \forall \; k \in D && (2.13)\\ & E(k) = \mathcal{F}\{A(x)(X(x) + iY(x))\} && (2.14)\\ & X^2(x) + Y^2(x) \leq 1 \forall x. && (2.15) \end{aligned}$$

This optimization problem is now convex and linear, and therefore has a unique global optimum which can be found using standard optimization techniques.

2.2.2 Discretization

To numerically optimize this optimization problem we have to discretize all functions. For simplicity, we discretize the focal and pupil plane into discrete set of points, $\{k_i\}$ and $\{x_i\}$ respectively. In this way calculation of the focal-plane electric field at those points can be done by a vector-matrix multiplication as

$$\mathbf{E}_{\text{foc}} = M\mathbf{E}_{\text{pup}}, \tag{2.16}$$

where $\mathbf{E}_{\text{foc}}$ is the focal-plane electric field at points $\{k_i\}$, $\mathbf{E}_{\text{pup}}$ the pupil-plane electric field at points $\{x_i\}$, and M the transformation matrix between pupil and focal plane.

The choice of the positions of discretization points in the focal plane is critical for correct performance. We always have discretization error: in between two points in the focal plane the PSF might actually exceed the contrast limit, while at the same time the PSF at the discretization points is within the limit. Due to computational time limitations we always want as little points in the focal plane as possible, while still covering the entire dark zone. In practice this works out to > 2 points per λ/D, corresponding to the Nyquist limit for the used aperture. For APPs with a high contrast (typically deeper than 10^{-6}) this is not enough, and we need ~ 3 points per λ/D. This is due to superoscillations (Aharonov et al., 1990; Ferreira & Kempf, 2006) that occur in this regime: the electric field oscillates locally faster than the Nyquist rate of the aperture. This means that even if we constraint the electric field at > 2 points per Nyquist-limited period, the electric field will still vary within those points at an apparent spatial frequency higher than the Nyquist frequency itself. Constraining the electric field on more points bounds the strength of these superoscillations (Ferreira & Kempf, 2006).

2.2.3 Speed improvements

While a direct implementation of the above discretized optimization problem yields correct APPs, there are some changes that can be made to improve runtime and performance. In practice excluding the Strehl ratio from the dark zone constraint yields more than an order of magnitude shorter computation times. The reason for this can be attributed to the

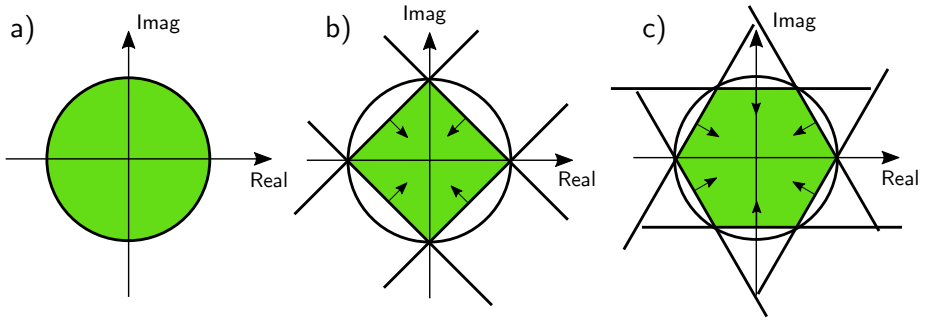


Figure 2.1: Rewriting square constraints to linear constraints can help computation times, while only having a small influence on the final Strehl ratio. *a)* A quadratic constraint contains all phase space but is computationally expensive. *b)* A box-shaped constraint consists of four linear constraints and balances coverage of the phase space, while keeping computation times in check. *c)* Higher-order regular polygons, such as this hexagon, require more linear constraints and therefore more computational resources, while yielding no significant improvement in the resulting Strehl ratios.

inner workings of the optimizer that was used (Gurobi Optimization, 2016). Other optimizers may react differently. We replace actual Strehl ratio with the expected Strehl ratio of the optimization S_{expected}. The optimization problem now reads

$$\underset{X(x),Y(x)}{\text{maximize}} \quad \Re\{E(0)\} \tag{2.17}$$

$$\text{subject to} \quad |E(k)|^2 \leq S_{\text{expected}} \cdot 10^{-c} \; \forall \; k \in D \tag{2.18}$$

$$E(k) = \mathcal{F}\{A(x)(X(x) + iY(x))\} \tag{2.19}$$

$$X^2(x) + Y^2(x) \leq 1 \forall x. \tag{2.20}$$

Note that to get a correct value for the Strehl ratio, we need to iterate this optimization multiple times, starting with $S^{(0)}_{\text{expected}} = 1$ and update the expected Strehl ratio after each optimization. In practice the Strehl ratio converges after $\sim 3-4$ iterations, except for extremely low Strehl ratios, in which we are not interested anyway.

Another improvement is the removal of quadratic constraints. These type of constraints are notoriously difficult for numerical optimizers and linear constraints are preferred. We can approximate the circle by an inscribed box as shown in Figure 2.1. Although this reduces the phase space

of the optimization, in practice this doesn't reduce the Strehl ratio of the final APPs by much. Orienting the diagonal of the box in the direction of the unaltered electric field at that point in the focal plane increases the Strehl slightly and tends to reduce artifacts. The result of using a box-shaped constraint instead of a circular one is that the optimizer now prefers to put the final electric field in the direction of the diagonals. This also means that we cannot replace the pupil-plane quadratic constraints with box-shaped linear constraints: the optimized APP would consist of discrete phases (namely $0, \pi/2, \pi, 3\pi/2$). The discretized phases are precisely what Carlotti et al. (2013) found. In the focal plane we do not care as much about the phase of the residual speckles and we can safely apply the box-shaped linear constraints. Using higher order regular convex polygons, such as the hexagon or octagon was found to yield a negligible improvement while significantly increasing computation time. The optimization problem now reads

$$\underset{X(x),Y(x)}{\text{maximize}} \quad \Re\{E(0)\} \tag{2.21}$$

$$\text{subject to} \quad \Re\{E(k)\} + \Im\{E(k)\} < \sqrt{S_{\text{expected}} \cdot 10^{-c}} \tag{2.22}$$

$$\Re\{E(k)\} - \Im\{E(k)\} < \sqrt{S_{\text{expected}} \cdot 10^{-c}} \tag{2.23}$$

$$-\Re\{E(k)\} + \Im\{E(k)\} < \sqrt{S_{\text{expected}} \cdot 10^{-c}} \tag{2.24}$$

$$-\Re\{E(k)\} - \Im\{E(k)\} < \sqrt{S_{\text{expected}} \cdot 10^{-c}} \tag{2.25}$$

$$E(k) = \mathcal{F}\{A(x)(X(x) + iY(x))\} \tag{2.26}$$

$$X^2(x) + Y^2(x) \leq 1 \forall x. \tag{2.27}$$

2.2.4 Tilt correction

When optimizing for small inner-working angle dark zones, we can often see that the core of the PSF is pushed away from the dark zone, effectively increasing the inner-working angle by that same amount: the core of the planet PSF is moved towards the star. This effect is commonly found in other APP optimization algorithms: introducing a tilt in the pupil-plane phase pattern and dealing with the reduced throughput of the core at $(0, 0)$ is advantageous compared to suppressing the starlight that close to the PSF core. This is of course not desirable and we have to suppress this behavior in some way. Other optimizers deal with this by removing the introduced tilt each iterations (as in the case of the modified Gerchberg-Saxton algorithm) or by removing tip-tilt from the mode basis used for optimization.

As we optimize electric field and have no access to the optimization algorithm directly, we instead have to introduce a constraint in the optimization problem to counteract this behavior.

The simplest and most effective way is to enforce that the PSF reaches its maximum at $(0, 0)$. In this way moving the PSF is not allowed. This constraint can be approximated as a linear constraint by enforcing that the absolute value of both the real and imaginary part of the electric field may not be larger than its value at $(0, 0)$. A quadratic constraint on the PSF itself would be best, but takes more computation time. Typically the PSF moves perpendicular to the dark zone shape, so we only need to enforce this in the direction opposite to the dark zone, starting at $(0, 0)$ itself. The optimization problem now reads

$$\underset{X(x),Y(x)}{\text{maximize}} \quad \Re\{E(0)\} \tag{2.28}$$

$$\text{subject to} \quad \Re\{E(k)\} + \Im\{E(k)\} < \sqrt{S_{\text{expected}} \cdot 10^{-c}} \tag{2.29}$$

$$\Re\{E(k)\} - \Im\{E(k)\} < \sqrt{S_{\text{expected}} \cdot 10^{-c}} \tag{2.30}$$

$$-\Re\{E(k)\} + \Im\{E(k)\} < \sqrt{S_{\text{expected}} \cdot 10^{-c}} \tag{2.31}$$

$$-\Re\{E(k)\} - \Im\{E(k)\} < \sqrt{S_{\text{expected}} \cdot 10^{-c}} \tag{2.32}$$

$$E(k) = \mathcal{F}\{A(x)(X(x) + iY(x))\} \tag{2.33}$$

$$X^2(x) + Y^2(x) \leq 1 \forall x \tag{2.34}$$

$$\Re\{E(k)\} \leq \Re\{E(0)\} \forall k \tag{2.35}$$

$$-\Re\{E(k)\} \leq \Re\{E(0)\} \forall k \tag{2.36}$$

$$\Im\{E(k)\} \leq \Re\{E(0)\} \forall k \tag{2.37}$$

$$-\Im\{E(k)\} \leq \Re\{E(0)\} \forall k. \tag{2.38}$$

We use this final optimization problem for the rest of this paper.

2.3 Case studies

2.3.1 D-shaped dark zones

In this section we consider a D-shaped dark zone. This dark zone is parameterized by their inner-working angle, outer-working angle and contrast. For the aperture we take a circular aperture with a central obscuration. This geometry is shown graphically in Figure 2.2. In the following sections we

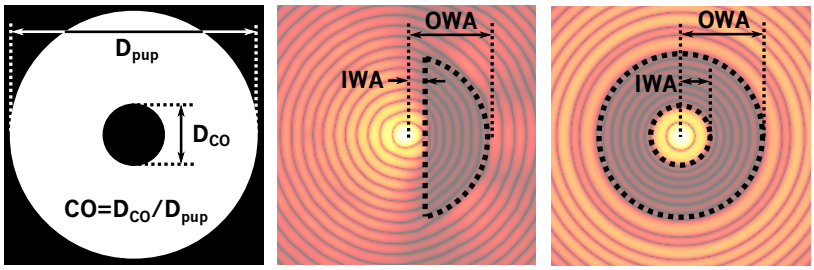


Figure 2.2: The pupil-plane and focal-plane geometry considered in this work. Any spiders holding up the central obscuration are neglected. *Left* The pupil of the telescope. *Middle* A D-shaped dark zone. *Right* An annular dark zone.

vary the dark-zone parameters and use the Strehl ratio as a metric. In Figure 2.3 we show some solutions with their parameters along with the resulting PSF.

Outer-working angle

We fix the inner-working angle at $IWA = 2\lambda/D$, the contrast at 10^{-5} and vary the outer-working angle and central obscuration. The Strehl ratio for these optimizations are shown in Figure 2.4. We can see that the Strehl ratio asymptotically converges to a fixed Strehl ratio for increasing outer-working angles. This suggests that designing infinite outer-working angle solutions might be feasible, and other optimization algorithms yield similar results.

Inner-working angle and contrast

The relation between the Strehl ratio S, inner-working angle and contrast paints a more complicated picture. We fix the outer-working angle at $8\lambda/D$ and vary the contrast and inner-working angle for various values of the central obscuration. The results of these optimization are shown in Figure 2.5. For an unobstructed aperture, the iso-Strehl lines generally have a constant gradient: if an APP is desired with a contrast 10 times deeper, the inner-working angle must be increased by $\sim 0.2\lambda/D$ to yield the same Strehl ratio. For the $S = 0.8$ line however, the gradient changes at $IWA = 1.6\lambda/D$ and a contrast of 10^{-4} to $\sim 0.5\lambda/D$ per decade in contrast. Introducing a 10% central obscuration, we can see that this turn-off point moves towards shal-

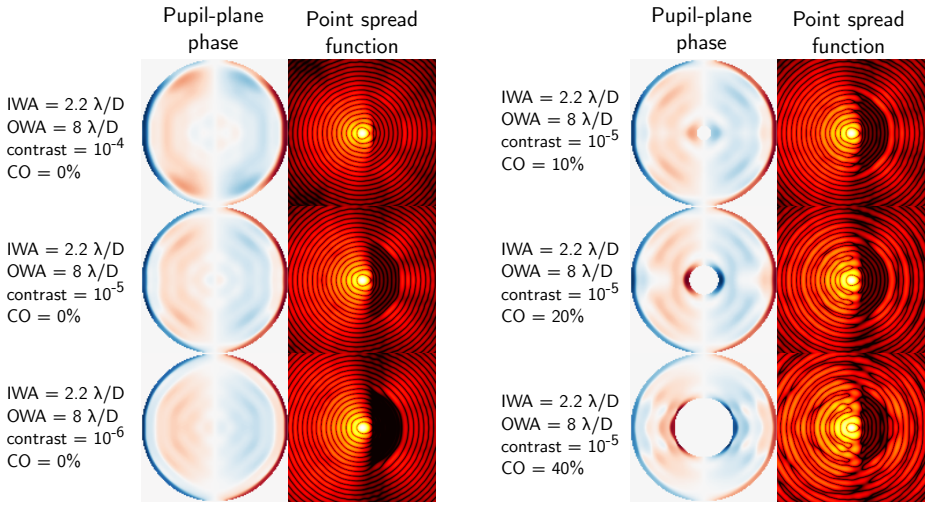


Figure 2.3: Some examples for optimization with D-shaped dark zones. For each set of input parameters, we show the final phase pattern and its corresponding point spread function.

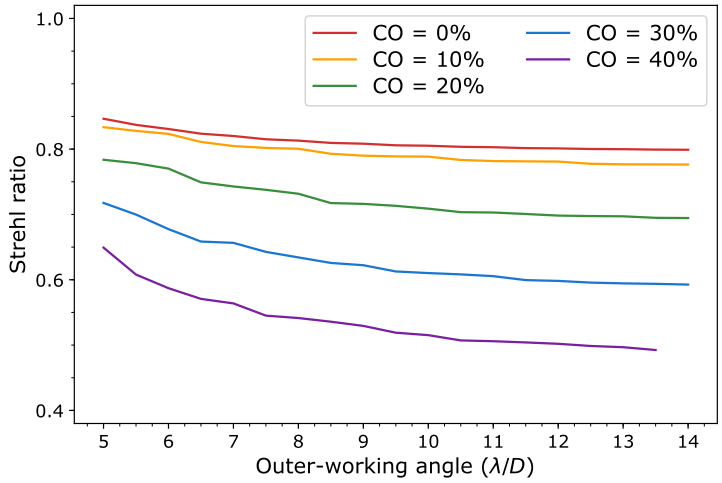


Figure 2.4: The Strehl ratio for a fixed inner-working angle ($2\lambda/D$) and contrast (10^{-5}) but varying outer-working angle for various values of the central obscuration CO. As the outer-working angle becomes larger, the Strehl ratio changes less and less. Outer-working angle is in these simulations limited by computation resources, both in memory and time.

lower contrasts. If we increase the central obscuration further, this trend continues and other iso-Strehls start to follow the same behavior.

Note that in these graphs, the increase in Strehl ratios at extremely low inner-working angles and high contrasts should not be trusted. These solutions tend to circumvent the tilt correction method described in Section 2.2.4 by producing a second peak beyond the range the tilt correction can protect against. This second peak is then the major contribution to Strehl ratio, while the peak at the origin $(0, 0)$ still decreases.

2.3.2 Annular dark zones

Annular dark zones are generally easier to optimize, due to the symmetry that they provide. While symmetric optimization problems do not necessarily have symmetric solutions, see for example the excellent paper by Waterhouse (Waterhouse, 1983), in this case they do. The reasoning for this is simple. If a global optimum x has been found, the transformed solution $x' = T(x)$, where T is the transformation corresponding to a symmetry in the optimization problem, must also be a global optimum. In the case of a convex optimization problem where the objective function is strictly convex, the global optimum is unique. Therefore $x' = x$ and x must be a symmetric solution. As our optimization problem is convex with a strictly convex objective function, all symmetries in the system provide a symmetry of the solution.

This result can be applied to many symmetries. For example: when all focal-plane constraints are situated in points-symmetric positions around the origin (ie. a two-sided dark zone), the transformation $Y(x) \to -Y(x)$ is a symmetry of the optimization problem, as its corresponding focal-plane transformation is $E(k) \to \overline{E}(-k)$. Therefore, a global optimal solution for such a dark zone geometry must satisfy $Y(x) = -Y(x) = 0$, so the pupil-plane phase is real. Empirically we again find that the absolute value of the pupil-plane apodizer is maximized, so that $X(x) \in \{1, -1\}$. This simplification reduces all quadratic constraints to linear constraints, which makes the optimization problem a simple linear problem. This means that the APPs found by Carlotti et al. (2013) were indeed globally optimal. An example of such an APP is shown in Figure 2.6.

Additionally, if the optimization problem is rotationally symmetric (ie. a rotationally symmetric pupil and dark zone), the solution must consist of rings of 0 or π in phase. Exploiting this symmetry leads to a significantly reduced computation time, and allows for extensive parameter studies.

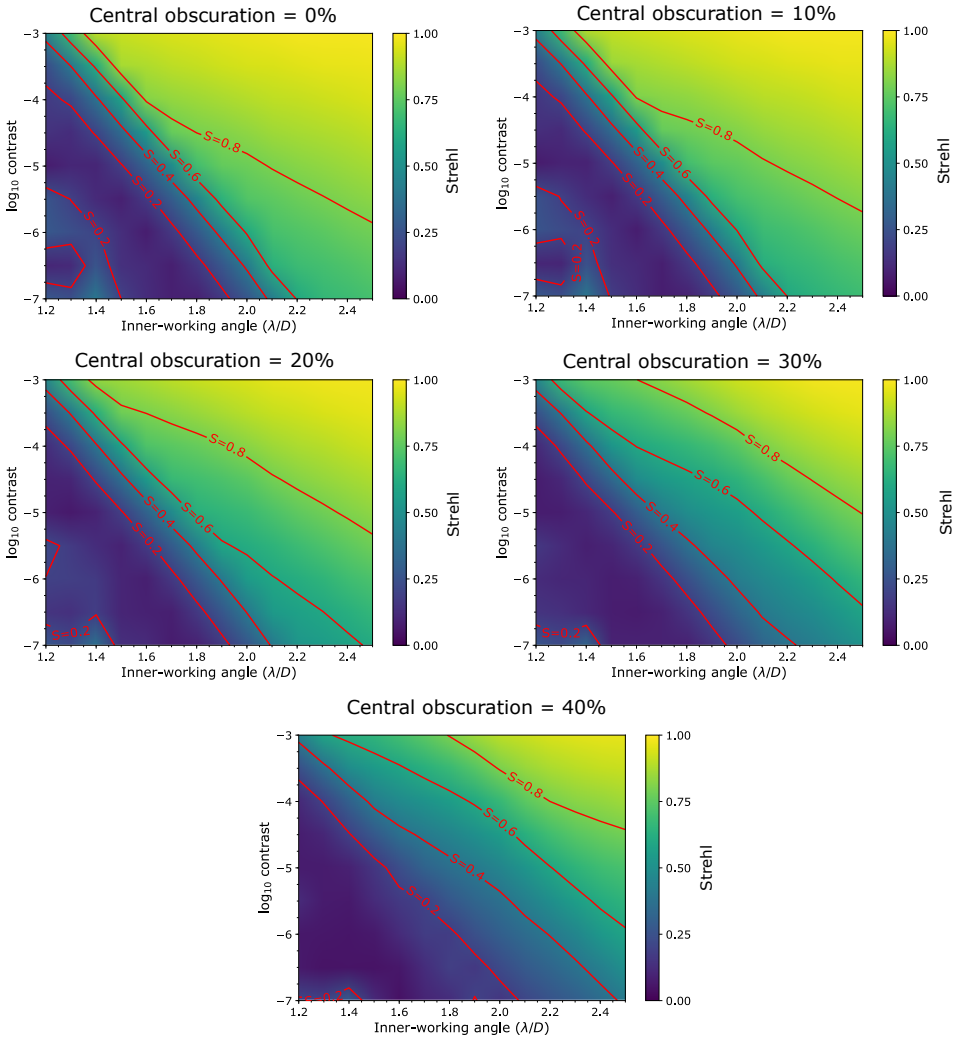


Figure 2.5: The Strehl ratio as a function of inner working angle, contrast and central obscuration size for a D-shaped dark zone. The outer working angle was fixed at $8\lambda/D$. See the text for a qualitative description of all features visible in this figure. These graphs can be used as a starting point for designing an APP for more complicated apertures.

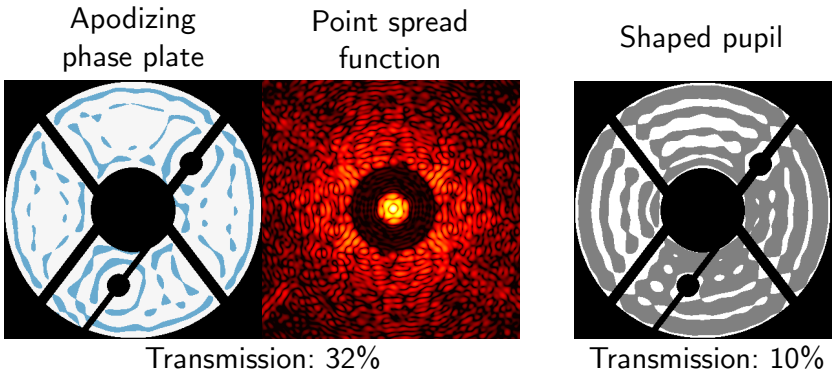


Figure 2.6: An example of a 360 APP for a complicated aperture. This aperture, based on the Subaru pupil, contains a large central obscuration, spiders, and masked dead deformable mirror actuators. A shaped pupil design for the same dark-zone geometry is also shown. As the phase space for shaped pupils is completely contained in the phase space allows in the APP optimizations, the transmission of the shaped pupil will always be lower than the optimized APP.

Inner-working angle and contrast

We consider the case of a annular dark zone, parameterized by their inner working angle, outer working angle and contrast. As the aperture shape, we again use a circular aperture with a varying central obscuration size. This geometry is shown graphically in Figure 2.2. In Figure 2.7 we fix the contrast at 10^{-4} and vary the inner working angle, outer working angle and central obscuration. We show the Strehl ratio of the corresponding APP pattern. The most prominent feature of these graphs are the plateaus of almost equal Strehl ratio: for a wide range ($\sim 1\lambda/D$) in both inner and outer working angles, the Strehl does not vary significantly. This is not an artifact of the optimization method but rather is related to the structure of the Airy rings of the unaltered telescope PSF. The Strehl depends on the number of Airy rings that need to be suppressed, rather than the total area of the dark zone.

The transitions between the plateaus do not correspond directly to the center of an Airy ring, but the plateaus are slightly broader. This can be attributed to the Airy ring being 'compressed' more easily than being suppressed entirely. Also note that the transition from good Strehl ($\gtrsim$ 50%) to bad Strehl ($\lesssim$ 10%) is extremely steep in inner working angle, and

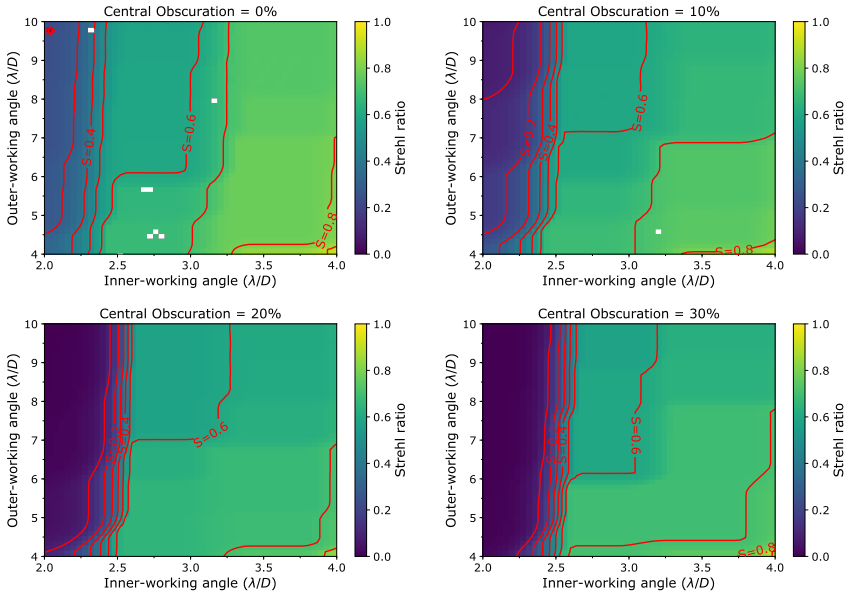


Figure 2.7: A parameter study on inner and outer working angle and central obscuration for annular dark zones. The contrast was fixed at 10^{-4}. The most prominent features are the existence of plateaus of equal Strehl ratio. These are determined by the positions of the original position of the Airy rings. A qualitative discussion can be found in the text.

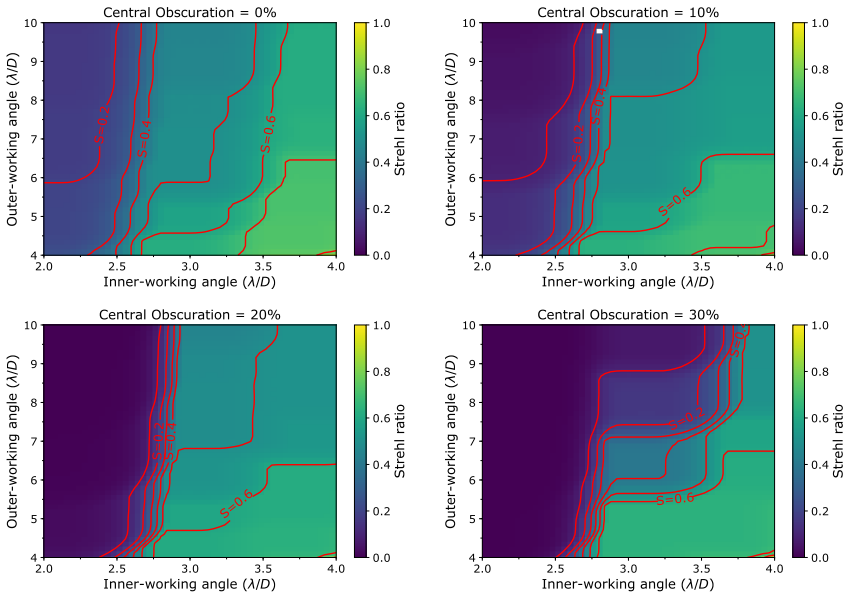


Figure 2.8: Identical to Figure 2.7 except for the changed contrast to 10^{-5}.

becomes steeper when the central obscuration is larger. The reason for this is the relative strength of the first Airy ring compared to the Airy core, which rises when the central obscuration is enlarged. In addition, the Airy ring also moves outwards with larger central obscurations. This means that more Strehl needs to be expended to suppress that Airy ring, as the Strehl is correlated with the amount of light that needs to be suppressed, rather than the suppressed area.

In Figure 2.8 we show the same results as in Figure 2.7, but fix the contrast to 10^{-5}. The general behavior is similar to Figure 2.7. All effects mentioned above are however much more visible due to the deeper contrast requirements. For example, for a 30% central obscuration, the transition between good and bad Strehl even extends towards the third Airy ring.

Large outer-working angles

While the case for large outer working angles for D-shaped APPs was limited to $\sim 14\lambda/D$ due to computer memory limitations, for the circularly symmetric case we can go much further than that. In Figure 2.9 we show an example solution for a $100\lambda/D$ outer working angle. In this example, the inner-working angle was $3.5\lambda/D$, the contrast 10^{-5} and the central obscura-

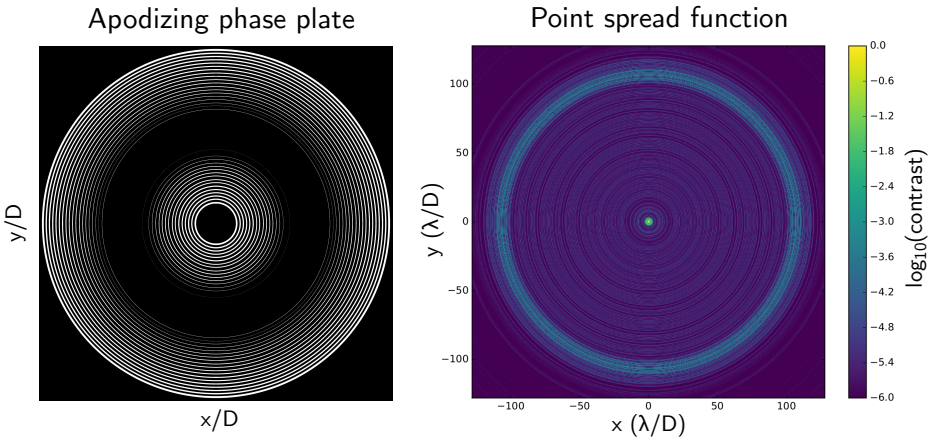


Figure 2.9: An example of an APP with a annular dark zone and an extremely large outer working angle of $100\lambda/D$. Other parameters were an inner working angle of $3.5\lambda/D$, a contrast of 10^{-5} and a central obscuration of 10%. The Strehl ratio was 40%.

tion 10%. The optimized APP had a Strehl ratio of $S = 40\%$. The solution again consists of many rings of 0 or π phase. We can see that the width of the rings changes smoothly over the aperture, being wide at the edge and at the center of the aperture, while vanishing in between. This suggests, and this is confirmed by other 360 APP optimizations, that the size of the rings, and therefore their spacing, is dependent on the outer working angle, and their width is modulated by a spatial frequency determined by the inner working angle.

In the PSF we can see that the APP pushes the light into a ring just outside of the outer working angle. In the dark zone, many rings are visible, each with a maximum contrast of 10^{-5}. We can therefore see that if we keep increasing the outer working angle, this ring of light will be evenly distributed over the dark zone and eventually we will run out of light. At that point the contrast outside of the outer working angle will also be smaller than the required contrast and we will have obtained an infinite outer working angle APP with an annular dark zone.

2.4 Conclusions

We have shown a new method for optimizing apodizing phase plate coronagraphs. This method, based on the work by Carlotti et al. (2013), uses a convex linearized version of the full optimization problem, which can be globally optimized using large-scale linear optimizers. While this optimization problem in theory allows for amplitude apodization, in practice all solutions turn out to have unit amplitude, making them phase-only pupil-plane apodizers. We performed parameter studies for one-sided D-shaped and two-sided annular dark zones. These simulations serve as a starting point for a parameter study for realistic, more complicated apertures. Additionally, we showed that globally optimized two-sided APPs contain only 0 or π phase.

Future research will apply this optimizer to the optimization of pre-apodizers for phase-mask coronagraphs, such as the vortex (Carlotti et al., 2014; Ruane et al., 2015) or the four-quadrant phase-mask coronagraph (Carlotti et al., 2014). Similar pre-apodizers can most likely be found for apodized Lyot coronagraphs (N'Diaye et al., 2016; Zimmerman et al., 2016).

References

Aharonov, Y., Anandan, J., Popescu, S., & Vaidman, L. 1990, Physical review letters, 64, 2965

Carlotti, A., Kasdin, N. J., Vanderbei, R. J., & Riggs, A. E. 2013, in SPIE Optical Engineering+ Applications, International Society for Optics and Photonics, 88641Q–88641Q

Carlotti, A., Pueyo, L., & Mawet, D. 2014, Astronomy & Astrophysics, 566, A31

Carlotti, A., Vanderbei, R., & Kasdin, N. 2011, Optics Express, 19, 26796

Codona, J., Kenworthy, M., Hinz, P., Angel, J., & Woolf, N. 2006, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 62691N–62691N

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

Ferreira, P. J., & Kempf, A. 2006, IEEE transactions on signal processing, 54, 3732

Gurobi Optimization, I. 2016, Gurobi Optimizer Reference Manual. `http://www.gurobi.com`

Guyon, O., Pluzhnik, E., Kuchner, M., Collins, B., & Ridgway, S. 2006, The Astrophysical Journal Supplement Series, 167, 81

Mawet, D., Pueyo, L., Lawson, P., et al. 2012, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 844204–844204

N'Diaye, M., Soummer, R., Pueyo, L., et al. 2016, The Astrophysical Journal, 818, 163

Ruane, G. J., Huby, E., Absil, O., et al. 2015, Astronomy & Astrophysics, 583, A81

Waterhouse, W. C. 1983, The American Mathematical Monthly, 90, 378

Zimmerman, N. T., Riggs, A. E., Kasdin, N. J., Carlotti, A., & Vanderbei, R. J. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011012

Chapter 3

The Single-mode Complex Amplitude Refinement (SCAR) coronagraph

I. Concept, theory and design

Adapted from
E. H. Por and S. Y. Haffert (2020), A&A 635, A55

Context The recent discovery of an Earth-mass exoplanet around the nearby star Proxima Centauri provides a prime target for the search for life on planets outside our solar system. Atmospheric characterization of these planets has been proposed by blocking the starlight with a stellar coronagraph and using a high-resolution spectrograph to search for reflected starlight off the planet.
Aims Due to the large flux ratio and small angular separation between Proxima b and its host star ($\lesssim 10^{-7}$ and $\lesssim 2.2\lambda/D$ respectively; at 750nm for an 8m-class telescope) the coronagraph requires high starlight suppression at extremely-low inner working angles. Additionally, it must operate over a broad spectral bandwidth and under residual telescope vibrations. This allows for efficient use of spectroscopic post-processing techniques. We aim to find the global optimum of an integrated coronagraphic integral-field spectrograph.
Method We present the Single-mode Complex Amplitude Refinement (SCAR) coronagraph that uses a microlens-fed single-mode fiber array in the focal plane downstream from a pupil-plane phase plate. The mode-filtering property of the single-mode fibers allows for the nulling of starlight on the fibers. The phase pattern in the pupil plane is specifically designed to take advantage of this mode-filtering capability. Second-order nulling on the fibers expands the spectral bandwidth and decreases the tip-tilt sensitivity of the coronagraph.
Results The SCAR coronagraph has a low inner working angle ($\sim 1\lambda/D$) at a contrast of $< 3 \times 10^{-5}$ for the six fibers surrounding the star using a sufficiently-good adaptive optics system. It can operate over broad spectral bandwidths ($\sim 20\%$) and delivers high throughput ($> 50\%$ including fiber injection losses). Additionally, it is robust against tip-tilt errors ($\sim 0.1\lambda/D$ rms). We present SCAR designs for both an unobstructed and a VLT-like pupil.
Conclusions The SCAR coronagraph is a promising candidate for exoplanet detection and characterization around nearby stars using current high-resolution imaging instruments.

3.1 Introduction

The discovery of many rocky exoplanets around stars (Borucki et al., 2011) has prompted the radial velocity search of the closest and brightest ones. This led to the discovery of a terrestrial exoplanet in the habitable zone around Proxima Centauri (Anglada-Escudé et al., 2016). This planet does not transit its host star (Kipping et al., 2017), making transit spectroscopy impossible. Proxima b however has an angular separation in quadrature of $\sim 2.2\lambda/D$ at 750nm for an 8m-class telescope, making a spatially resolved imaging approach feasible. Lovis et al. (2017) explores the possibility of coupling the high-contrast imager SPHERE (Beuzit et al., 2008) with the high-resolution spectrograph ESPRESSO (Pepe et al., 2010). The implementation of the coronagraph was left as an open question. Here we show a new type of coronagraph that will enable a combination of SPHERE and a high-resolution spectrograph to successfully observe Proxima b.

With the advent of extreme adaptive optics systems (xAO), such as SPHERE (Beuzit et al., 2008), GPI (Macintosh et al., 2014) and SCExAO (Jovanovic et al., 2015), direct detection has taken some major steps forward. These systems create a diffraction-limited point spread function (PSF), which allows for the use of coronagraphy to enhance the contrast of observations. Coronagraphs need to suppress stellar light at the location of the planet, while having high throughput for the planet itself. Additionally, they need to operate over a broad wavelength range and in the presence of residual telescope aberrations, both static and dynamic. Telescope vibrations in particular turned out to be a major concern for these high contrast imaging instruments (Fusco et al., 2016). The performance of pupil-plane-only coronagraphs is independent of telescope pointing, making telescope vibrations less of a concern. An inherent disadvantage of pupil-plane coronagraphs is that the coronagraphic throughput loss is the same for the star and the planet. Therefore designs of pupil-plane coronagraphs that reach high contrasts or extremely-low inner working angles have an intrinsically low throughput. As a result, coronagraphs combining focal-plane and pupil-plane optics often outperform pupil-plane-only coronagraphs at extremely-low inner working angles (Mawet et al., 2012). For example, an apodizing phase plate (APP) with an annular dark zone optimized from $1.2\lambda/D$ to $2.3\lambda/D$ for a contrast of 10^{-4} has a Strehl ratio (ie. planetary throughput) of 0.13%. The APP optimization was performed following Por (2017).

Even with the best adaptive optics systems, residual aberrations will

always limit the raw contrast of ground-based observations to $\sim 10^{-6}$ to 10^{-7} (Guyon, 2005). Currently however, observations are limited by the non-common-path errors between the wavefront sensor and science camera, creating quasi-static speckles in the focal-plane. These speckles amplify atmospheric residuals (Aime & Soummer, 2004) and are notoriously hard to predict. The raw contrast in the intensity image provided by the coronagraphic system is often enhanced by post-processing techniques to form the final contrast curve of the observation. Diversity of some kind is often used to calibrate the instrument itself. Angular diversity (Marois et al., 2006) uses the rotation of the sky with respect to the instrument and has provided excellent results. A recent development in this field uses the diversity in radial velocities of the star and the planet: stellar-light speckles still retain the radial velocity of the star, while the planet appears at a different velocity altogether. Cross-correlation techniques on high-resolution spectroscopy (Konopacky et al., 2013; Riaud & Schneider, 2007; Sparks & Ford, 2002) combined with coronagraphy (Kawahara et al., 2014; Snellen et al., 2015; Wang et al., 2017) show great promise. They provide dayside spectroscopy of τ Boötis b (Brogi et al., 2012) and even the first measurement of planetary spin (Snellen et al., 2014).

Integration of both coronagraphy and high-resolution spectroscopy into a single concept has only recently been attempted. Mawet et al. (2017) use a fiber injection unit in the focal-plane downstream from a conventional vortex coronagraph. A single-mode fiber was centered around the planet and the stellar light speckles were removed using active speckle control algorithms. While this setup does allow for transportation of the light to a dedicated high-resolution spectrograph, it does not optimally combine both methods. A system for which all components are optimized simultaneously is always better than, or at least as good as, a system for which each component is optimized separately. In this paper we allow the coronagraph design to depend on the single-mode fibers in the focal plane. This allows for more freedom in the design process and provides better planetary coronagraphic throughput as those modes filtered out by the fiber injection unit do not need to be suppressed by the upstream coronagraph. The increase in throughput can be quite significant. For example, optimizing a SCAR coronagraph with an operating angular separations from $1\lambda/D$ to $2.5\lambda/D$ yields a throughput of $\sim 54\%$, compared to the $\sim 0.13\%$ of the corresponding APP mentioned above.

The schematic layout of the proposed coronagraph is shown in Figure 3.1. The SCAR coronagraph uses a phase plate in the pupil plane to

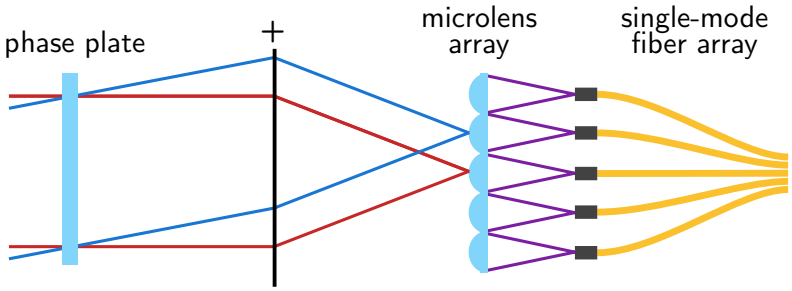


Figure 3.1: Schematic layout of the proposed system. The phase plate located in the pupil-plane alters the PSF that is imaged on the microlens array. Each microlens focus is imaged on a single-mode fiber. An off-axis source will be spatially separated in the focal-plane and its Airy core will fall on a different microlens.

alter the PSF. The light is then focused on the microlens array which focuses the light into the single-mode fiber array. This provides additional filtering of the incoming starlight. The starlight is nulled on off-axis fibers, while light is let through in the center fiber. This means that the light from an off-axis companion, of which the PSF is the same as that of the star, will be let through by the off-axis fiber on which it is centered. In this paper, we aim to provide an overview of the principle and optimization behind the SCAR coronagraph.

Interestingly, the use of single-mode fibers for coronagraphy is not new. Haguenauer & Serabyn (2006) already propose using a single-mode fiber to null the star by using a π phase shift on part of the pupil. Martin et al. (2008) develop this further and Hanot et al. (2011) finally put this system on sky. These applications of single-mode fibers for coronagraphy were based on interferometry. Mawet et al. (2017) is the first to put single-mode fibers behind a conventional coronagraph.

In Sect. 3.2 we describe nulling on single-mode fibers, extend the fiber injection unit to use multiple single-mode fibers and show the coronagraphic capabilities and throughput of such a system. In Sect. 5.5 we use an apodizing phase plate coronagraph to expand the spectral bandwidth and decrease the tip-tilt sensitivity. In Sect. 3.4 we describe the throughput, inner working angle, chromaticity and sensitivity to aberrations of this new system. We conclude with Sect. 3.6.

3.2 Modal filtering using single-mode fibers

3.2.1 Nulling in single-mode fibers

The coupling efficiency $\eta_{\rm sm}$ of light into a single-mode fiber can be calculated by the projection of the input electric field $E_{\rm in}$ onto the mode of the fiber $E_{\rm sm}$ as

$$\eta_{\rm sm} = \frac{\left|\int E_{\rm in}^* E_{\rm sm} \mathrm{d}A\right|^2}{\int |E_{\rm in}|^2 \mathrm{d}A \int |E_{\rm sm}|^2 \mathrm{d}A}, \tag{3.1}$$

where the integration is done over all space. The fiber mode $E_{\rm m}$ can be calculated using waveguide theory and the geometry of the fiber in question, but in this paper we use the Gaussian approximation (Marcuse, 1978)

$$E_{\rm sm}(r) = \exp\left[-\frac{r^2}{w^2}\right] \tag{3.2}$$

where r is the distance from the center, and $2w$ is the mode field diameter of the fiber. We see that the coupling efficiency $\eta_{\rm sm} \leq 1$ for all input fields and that maximum coupling is only attained when $E_{\rm in}$ matches the fiber mode.

Suppose now that we put a single-mode fiber in the focal plane of a telescope, with its mode field diameter matched to that of the Airy core of the PSF. Using Equation 3.1 we can calculate the coupling efficiency $\eta_s(\mathbf{x})$ of the star as a function of focal-plane position x. We can do the same thing for the planet, yielding $\eta_p(\mathbf{x}, \mathbf{x}_0)$ where $\mathbf{x}_0$ is the location of the planet. The raw contrast at the fiber output can be written as

$$C_{\rm raw}(\mathbf{x}, \mathbf{x}_0) = \frac{\eta_{\rm s}(\mathbf{x})}{\eta_{\rm p}(\mathbf{x}, \mathbf{x}_0)}. \tag{3.3}$$

When the fiber is centered around the planet, ie. $\mathbf{x} = \mathbf{x}_0$, the electric field of the planet will couple efficiently into the fiber, as the Airy core is closely matched to the Gaussian fiber mode. The electric field of the star at this position will however consist of Airy rings. These will be smaller, not only in intensity but also spatially so that around two Airy rings will be visible on the fiber face. This is possible as the Airy core itself has a size of $\sim 1\lambda/D$ full-width half-maximum (FWHM), while the Airy rings are sized $\sim 0.5\lambda/D$ FWHM. As neighboring Airy rings have opposite phase, the light from the two Airy rings will (partially) cancel each other in the projection integral of Eq. 3.1, resulting in a lower stellar throughput. This nulling

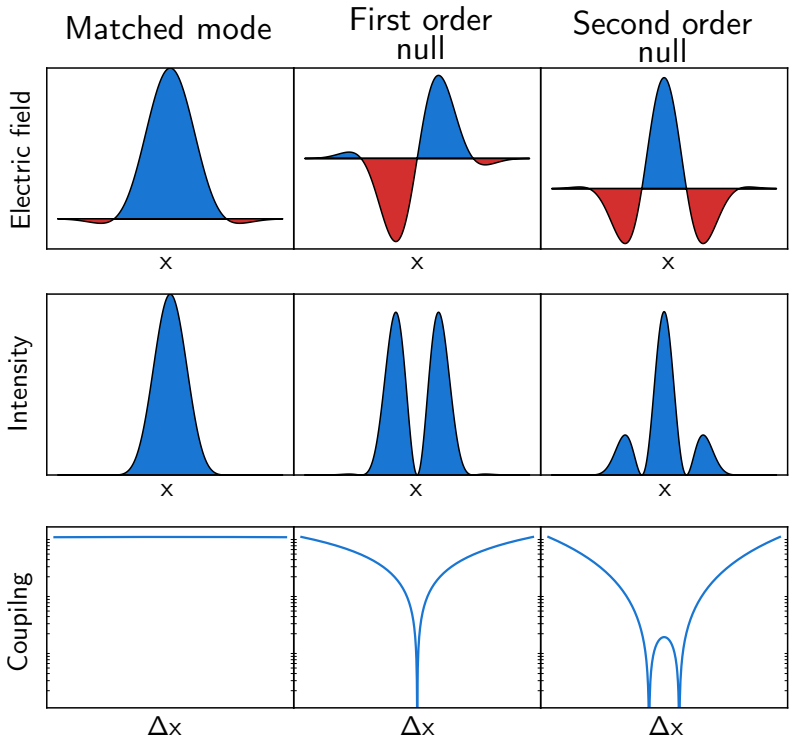


Figure 3.2: Coupling into a single-mode fiber with a (a) matched electric field, (b) first order null and (c) second order null. The first row shows the electric field projected onto the fiber. The second row shows the intensity on the fiber face. The third row shows the coupling efficiency for off-center fibers. The matched mode couples well into the fiber, even for small off-center positions. The first-order null has no throughput: its odd electric field ensures a zero overlap integral in Eq. 3.1. Off-center positions however still transmit because the odd structure is lost. The second-order employs an even electric field where the contribution of the central peak is canceled by the two sidebands. This creates a much broader null when decentering the fiber.

provides an additional contrast enhancement not possible with multimode fibers. Fig. 3.2 illustrates this graphically in columns one and two.

Column three in Fig. 3.2 shows an alternative null structure. This second-order null balances the contribution of the core with its two sidebands. This has the effect of broadening the null for decenter as the loss in overlap with one of the sidebands is compensated by the increase in overlap with the other. In this case we have essentially split the second-order null into two first-order null by subtracting a tiny fraction of the matched mode. This gives a characteristic double dip in the coupling curve and broadens the null even more by raising the coupling between the two first-order nulls. This coupling at the center must be kept below the design coupling.

This second-order null is the basis for the SCAR coronagraph. For comparison with existing coronagraph implementations, we define the "relative integration time" as the ratio between the integration time for unresolved imaging and coronagraphic imaging to reach a predefined signal-to-noise ratio. This can be expressed in the star and planet throughput as

$$\frac{\Delta T_{\text{coronagraphic}}}{\Delta T_{\text{unresolved}}} = \frac{\eta_s(\mathbf{x})}{\eta_p^2(\mathbf{x}, \mathbf{x}_0)}. \tag{3.4}$$

This metric takes into account both the raw contrast and planet throughput of the coronagraph. Noise sources other than photon noise were ignored in this respect. As these only become important for small planet throughputs, we will show both the relative integration time η_s/η_p^2 and the planet throughput η_p.

3.2.2 Single-mode fiber arrays using microlenses

To cover the field of view around a star, we need to fill the focal plane with single-mode fibers. This means that the fibers are impractically close together. A more reasonable solution is to use a microlens array with a single-mode fiber in each focus, as shown by Corbett (2009). Each fiber face now contains a strongly spatially-filtered telescope pupil. The corresponding focal-plane mode for each fiber can be recovered by back-propagating the fiber mode to the focal plane. An example of such a mode is shown in Fig. 3.3. The amplitude of this back-propagated mode is still Gaussian in amplitude. In phase however, it is flat within the central lenslet, but picks up a phase gradient on off-axis lenslets: light hitting off-axis lenslets need to have a huge tilt to still couple into the central fiber. We denote this off-axis contribution as lenslet crosstalk, and it is taken into account in every optimization and calculation done in this paper.

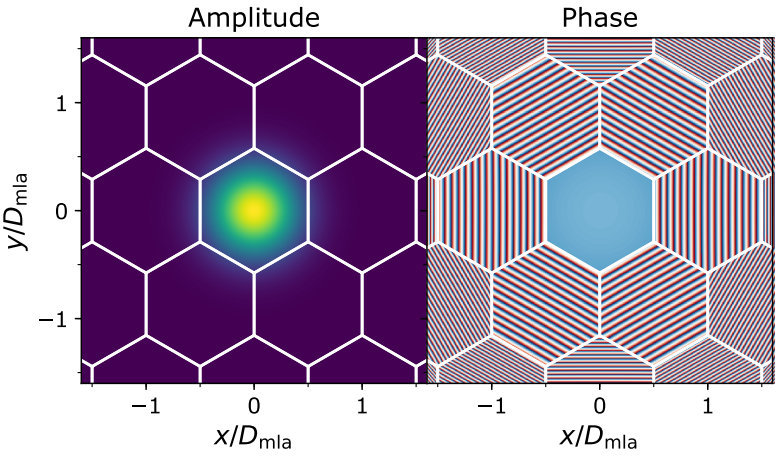


Figure 3.3: Backpropagated mode of a single-mode fiber to the microlens array surface. Conceptually our microlens array and single-mode fiber can still be thought of as focal-plane electric field filtering using this modified mode. The mode is still Gaussian on the central microlens, but picks up an additional tilt on off-axis microlenses: on those microlenses the light needs to have a huge tilt to be propagated into the central fiber.

The throughput of the single-mode fiber array depends on the position of the object and the mode field diameter of the fibers. Figure 3.4 shows the throughput for a clear aperture with slices through the best and worst-case position angles. The throughput is dominated by the lenslet closest to the center of the PSF and is only weakly dependent on the mode field diameter around the optimal value. Additionally, at each position in the focal plane, the optimal value of the mode field diameter is the same, simplifying implementation.

Figure 3.5 shows the throughput of an off-axis lenslet as a function of microlens diameter, while keeping the PSF centered around the on-axis lenslet. We can clearly see that at a diameter of $\sim 1.28\lambda/D$ no stellar light is transmitted by the fiber. This contrast is solely the result of the mode-filtering property of the single-mode fiber: if we were to use multimode fibers instead, the contrast would still be $\sim 3 \times 10^{-2}$ at this point. The nulling can be classified as first order: only where the electric field of the Airy core and the first Airy ring exactly cancel do we see the contrast reduction. Moving the PSF only slightly already destroys this nulling.

Since the PSF changes in size with wavelength, the throughput of an

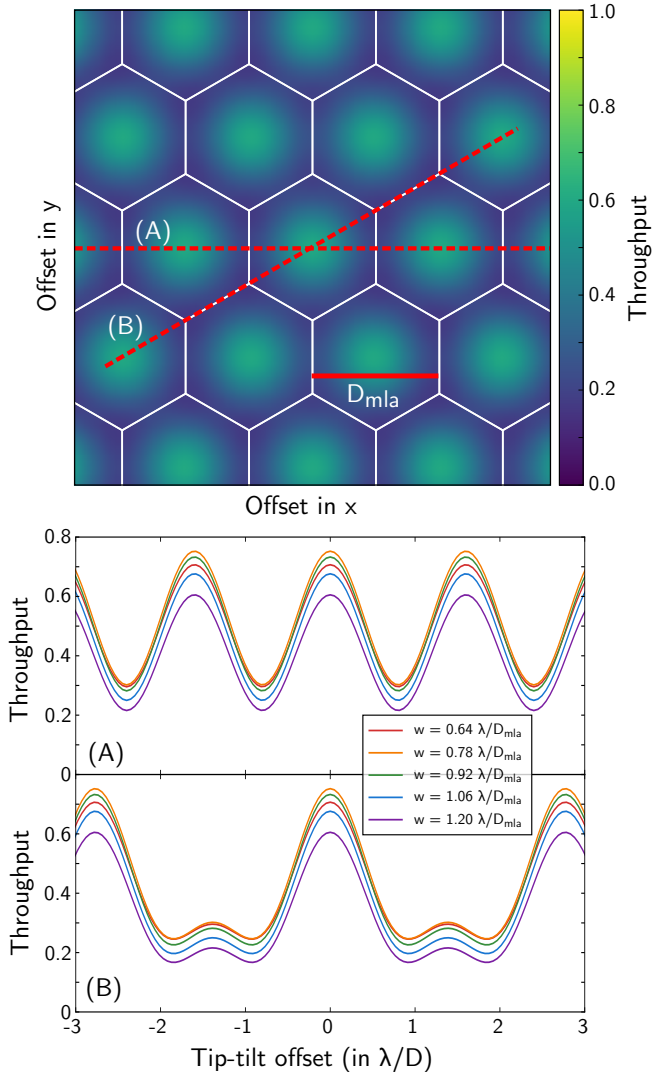


Figure 3.4: Throughput of the single-mode fibers as a function of tip-tilt offset of the source. On the top a two-dimensional throughput map is shown. On the bottom two slices through this map are plotted for several values for the mode field diameter of the single-mode fibers. Maximum throughput of $\sim 72\%$ is reached at the center of a lenslet. On the edge of two microlens the throughput of two fibers have to be added to reach $\sim 30\%$ throughput. The worst case is the triple-point at which the maximum throughput drops to $\sim 25\%$. A fiber mode field radius of $w = 0.78\lambda/D$ achieves the highest throughput for all tip-tilt offsets.

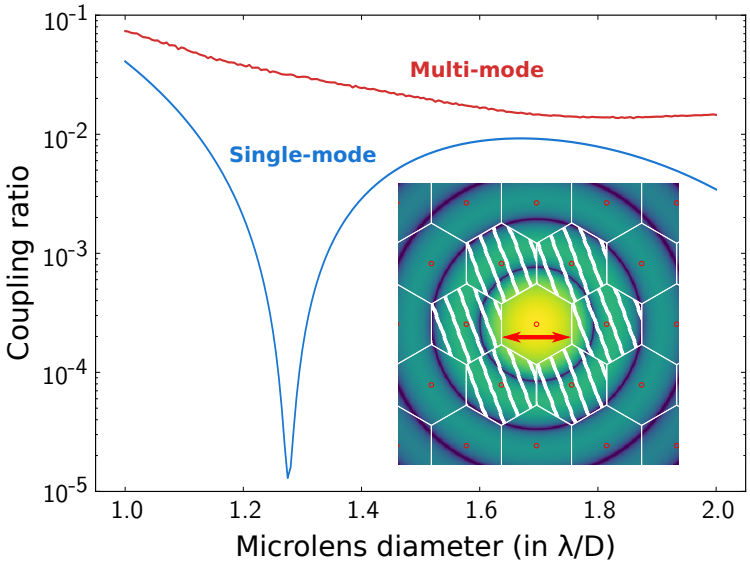


Figure 3.5: Coupling ratio between an on-axis and off-axis source through an off-axis microlens as a function of microlens diameter for a multimode and single-mode fiber. The gain in contrast by using a single-mode fiber can readily be seen in the ratio of these two functions. For most microlens diameters this gain amounts to several orders of magnitude and reaches infinity at $\sim 1.28\lambda/D$ where the light is perfectly nulled on the fiber face by the single-mode fiber. This nulling is first order and is therefore very sensitive to wavelength and centering of the star around the central lenslet.

off-axis fiber is inherently chromatic. We can read off the spectral bandwidth from Fig. 3.5 directly. A contrast of 10^{-4} is reached for $1.26\lambda/D < D_{\mathrm{mla}} < 1.30\lambda/D$, corresponding to a spectral bandwidth of just 3%. Nevertheless this demonstrates that significant gains can be obtained by using single-mode fibers instead of multimode fibers or even conventional intensity detectors.

3.3 Coronagraphy with a single-mode fiber array

3.3.1 Conventional coronagraphy

We can use conventional coronagraphy techniques to reduce the spot intensity and ignore the mode-filtering property in the design process. As

an example we use the apodizing phase plate (APP) coronagraph (Codona et al., 2006; Otten et al., 2017; Snik et al., 2012). This coronagraph consists of a single phase-only optic in the pupil plane, making it impervious to tip-tilt vibrations of the telescope or adaptive optics system. The phase pattern is designed to yield a dark zone in a certain region of interest in the focal plane. This region of interest can be both one- and two-sided, and can have arbitrary shapes. Most often the one-sided regions of interest are D-shaped and the two-sided are annular. See Por (2017) for a recent description of APP design. As both the PSF of the star and the planet are altered, the Strehl ratio is maximized to retain planet transmission.

Figure 3.6 shows the contrast through a fiber-fed single-mode fiber array using an APP designed for a contrast of 10^{-5} in a D-shaped region with an inner working angle of $2\lambda/D$ and outer working angle of $10\lambda/D$. While the use of single-mode fibers does enhance the contrast by $\sim 3\times$ on average, this enhancement is not consistent: in some fibers the contrast is enhanced by $> 10\times$ while in others we barely see any improvement at all. This shows that the factor of 3 enhancement that Mawet et al. (2017) finds for a dynamic random speckle field holds true for a single-mode fiber in a static structured speckle field, such as a residual coronagraphic electric field, even when the mode shape is modified from a Gaussian to a constricted Gaussian profile.

3.3.2 Direct pupil-plane phase mask optimization

This improvement brings up the question: can we make use of this mode-filtering in the coronagraph design? As the single-mode fiber array already filters out some electric field modes, the coronagraph does not have to suppress those modes; only modes that are transmitted by the single-mode fiber array need to be suppressed by the coronagraph. The coronagraph needs to minimize the coupling through the single-mode fibers, not the intensity at those positions in the focal plane. Designing a coronagraph specifically for the fiber array therefore allows for more design freedom compared to conventional coronagraph design. In principle, any coronagraph can be designed to take the fiber coupling into account. As a case study, we use a pupil-plane phase plate to alter the PSF in the focal plane. A schematic layout of the proposed system is shown in Fig. 3.1.

To find the phase pattern, we use the novel optimizer from Por (2017), based on the work by Carlotti et al. (2013), that maximizes the throughput (ie. Strehl ratio) for a complex pupil mask, while constraining the stellar intensity in the dark zone to be below the desired contrast. Since the trans-

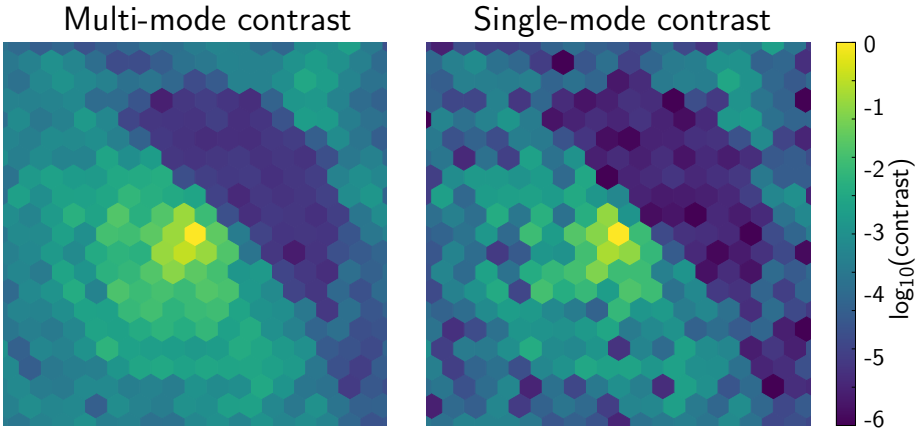


Figure 3.6: Coupling through a microlens array using a conventional APP coronagraph with multimode fibers (left) and single-mode fibers (right). Using a conventional coronagraph such as an APP contrast is still enhanced by the single-mode fibers. The contrast in the dark zone is still enhanced by $\sim 3\times$ on average when using single-mode fibers compared to multimode fibers.

formation between the pupil and focal plane is linear in electric field, this optimization problem is linear, and its global optimum can be easily found using large-scale numerical optimizers such as Gurobi (Gurobi Optimization, 2016). In practice the optimization produces phase-only solutions, which is surprising as non-phase-only solutions are still feasible solutions. As the phase-only optimization problem is simply a more constrained version of the linear one, the phase-only solution must therefore be a global optimum of both problems.

The fiber coupling integral in Eq.3.1, or rather the amplitude of the coupled electric field

$$E_{\text{coupled}} = \frac{\int E_{\text{in}}^{*} E_{\text{sm}} \mathrm{d}A}{\left| \int E_{\text{sm}} \mathrm{d}A \right|} \tag{3.5}$$

is still linear in the input electric field E_{in}, so we can apply the same method here. We maximize the throughput of the central fiber, while constraining the coupling through the specified off-axis fibers. To counter the chromaticity mentioned in Sect. 3.2.2, we constrain the off-axis stellar intensity at several wavelengths simultaneously, which ensures that the contrast is attained over a broad spectral bandwidth. Jitter resistance is kept in check in a similar manner: the desired raw contrast must be attained for several

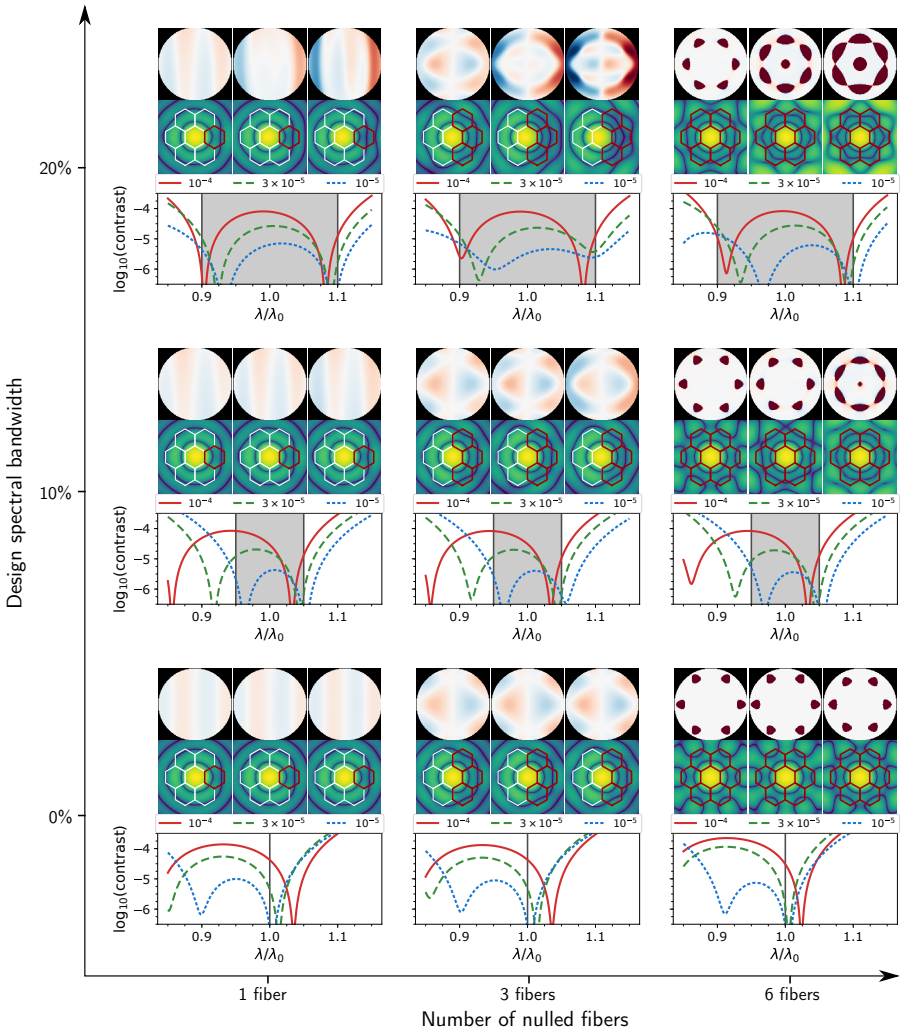


Figure 3.7: A series of optimizations for one, three and six fibers on the first ring of microlenses for a clear aperture. The design spectral bandwidth were 0%, 10% and 20% and the contrasts 1×10^{-4}, 3×10^{-5} and 1×10^{-5}. A $0.06\lambda/D$ peak-to-peak telescope tip-tilt jitter was also taken into account. Each microlens has a circum-diameter of $1.8\lambda/D$. For each SCAR design we show the pupil-plane phase pattern, its corresponding point spread function and its raw contrast η_s/η_p as a function of wavelength averaged over the marked fibers. The chromatic response shows the raw contrast after the single-mode fiber. The second-order nulling on the fiber face is clearly visible in every design. In Table 3.1 we list the fixed and varied parameter in this figure. The throughput of all SCAR designs shown in this figure can be found in Figure 3.9.

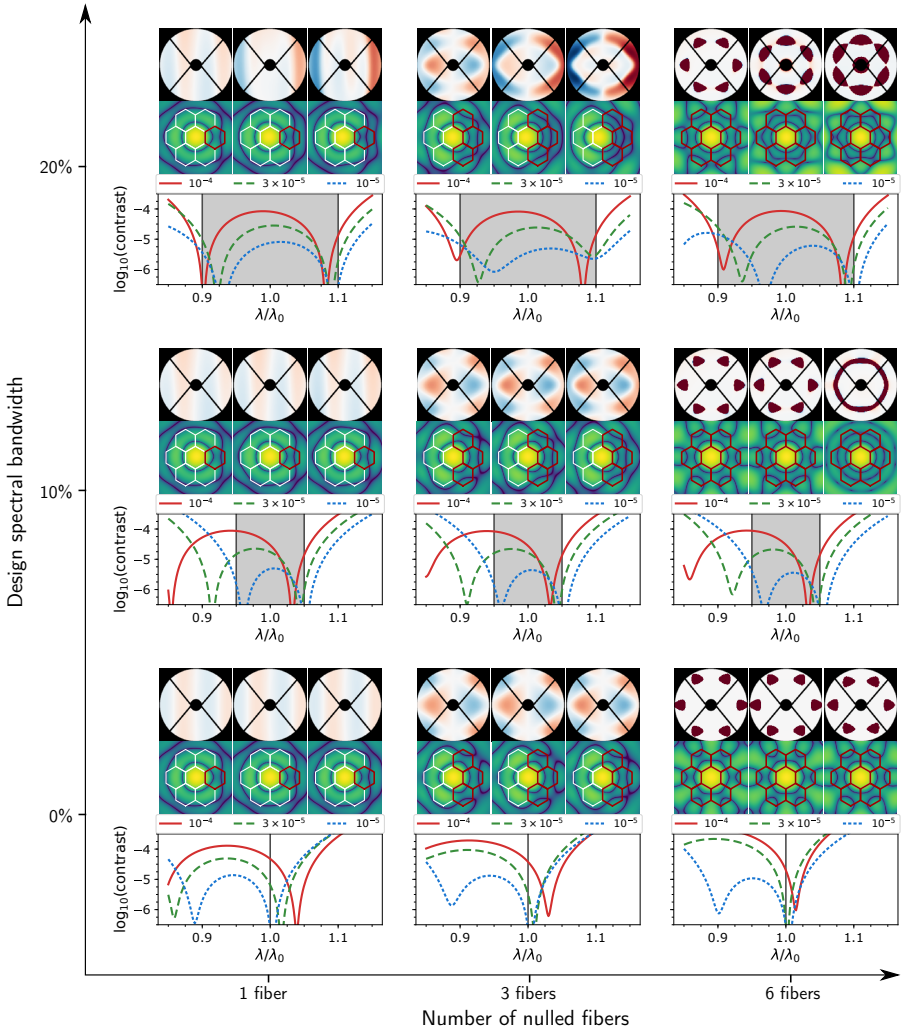


Figure 3.8: The same as Fig. 3.7 for a VLT aperture. The aperture was subject to a 1% binary erosion, ie. undersizing the pupil and oversizing central obscuration, spiders and other pupil features by 1% of the aperture size, to accommodate a pupil misalignment. The general structure of the solutions is similar to the case of a clear aperture. The central obscuration increases the strength of the first Airy ring, thereby decreasing the throughput of these SCAR designs slightly. The relatively thin spiders have no influence on the throughput at these angular separations. In Table 3.1 we list the fixed and varied parameter in this figure. A summary of the throughput of all SCAR designs shown in this figure can be found in Fig. 3.9.

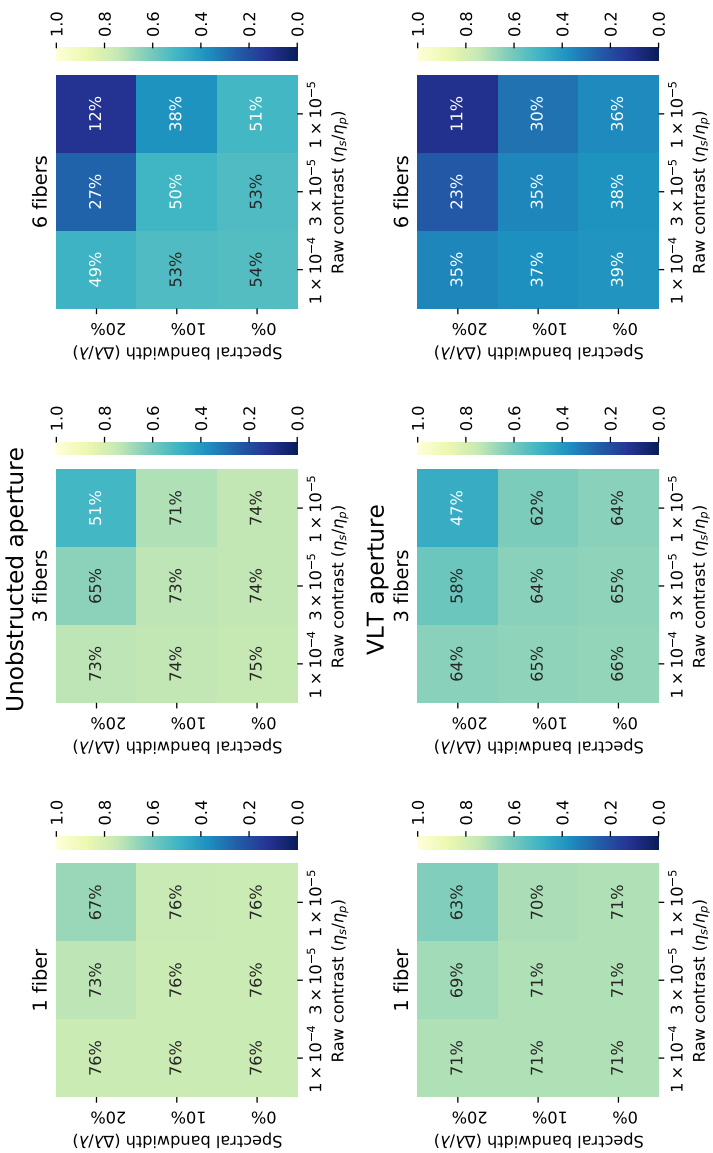


Figure 3.9: The throughput (η_p) for the SCAR designs shown Figs. 3.7 and 3.8. The shown throughput is the total throughput of the coronagraphic system. It includes the Strehl reduction of the phase plate, the spatial filtering of the microlenses and the mode filtering of the single-mode fibers. This throughput is achieved through a single single-mode fiber when the planet is located directly on top of an off-axis lenslet. The throughput at other positions varies generally according to Figure 3.4, but varies slightly depending on the specific SCAR design.

Table 3.1: The design parameters used for all SCAR designs throughout this work. All SCAR designs generated with these parameters can be found in Figures 3.7 and 3.8.

Parameter name	Value
Raw contrast limit	$\{1 \times 10^{-4}, 3 \times 10^{-5}, 1 \times 10^{-5}\}$
Spectral bandwidths	$\{0\%, 10\%, 20\%\}$
Tip-tilt errors	$0.06\lambda/D$ peak-to-peak
Microlens circum-diameter	$1.8\lambda/D$
Microlens shape	Hexagonal
Fiber mode shape	Gaussian
Fiber mode field diameter	$1.7\lambda/D_{\text{mla}}$
Pupil mask	{unobstructed, VLT}

tip-tilt positions simultaneously.

In Fig. 3.7 we show a few examples of optimizations for one, three and six fibers for a contrast of 1×10^{-4}, 3×10^{-5} and 1×10^{-5} using a $0\%, 10\%$ and 20% spectral bandwidth, along with their corresponding PSF and chromatic response. The design parameters are shown in Table 3.1. These design parameters were chosen to show a variety of SCAR designs for realistic implementations. At the shown contrast limits, the residual atmospheric speckles will limited the on-sky contrast, even after an extreme AO system. The spectral bandwidths were chosen as wide as possible, without compromising on planet throughput. The resulting spectral bandwidths are large enough to apply spectral cross-correlation techniques.

In each case the optimizer prefers a second order null. This second order null is much more stable against bandwidth and tip-tilt jitter. The reason for this is explained graphically in Fig. 3.2. Furthermore, this second order null is even present in monochromatic optimizations and in optimizations without accommodation for tip-tilt errors. This means that the second-order null requires less phase stroke to achieve and therefore provides higher Strehl ratios.

As the optimizer can handle arbitrary apertures, optimizations for other aperture shapes are also possible. Figure 3.8 shows optimizations for a VLT aperture for the same parameters as for the clear apertures. The aperture was subject to a 1% binary erosion, ie. undersizing the pupil and oversizing central obscuration, spiders and other pupil features by 1% of the aperture size, to accommodate for a misalignment in the pupil mask. Although the overall structure is quite similar, there is one key difference compared to

Table 3.2: The different throughput terms that are important for the SCAR coronagraph. A distinction is made between theoretical and experimental terms. Experimental throughput terms will be non-existent with perfect manufacturing, while theoretical throughput terms are unavoidable. Typical values in the visible are shown for each term. The values for the experimental throughput terms are indicative of current technology performance and may increase as technology matures.

Throughput term	Typical values
Theoretical	
Geometric lenslet throughput	$\sim 80\%$
Fiber injection losses	$90\% - 95\%$
Planet location	$50\% - 100\%$
Phase plate Strehl reduction	$60\% - 80\%$
Total	$\sim 41\%$
Experimental	
Phase plate transmission	$> 85\%$
Fresnel losses on the fiber	$\sim 90\%$
Microlens transmission	$> 95\%$
Strehl ratio of the AO system	$\sim 50\%$
Total	36%
Total	14%

a clear aperture: the relative transmission T is lower for all phase plate designs. This means that the relative transmission depends strongly on the size of the central obscuration. This is obvious as larger central obscurations strengthen the first Airy ring and brighter features typically cost more stroke, and therefore relative transmission, to change, similar to conventional APP design. Effectively, this means that each feature in the phase pattern becomes larger to compensate for larger central obscurations.

We summarize the multitude of SCAR phase pattern designs in Figure 3.9. This figure shows the total planet throughput η_p, provided that the planet is located in the center of the off-axis microlens. This throughput includes all theoretically unavoidable terms, but excludes all experimental terms. A summary of important throughput terms are listed in Table 3.2.

In the rest of this paper, we consider the outlined SCAR design in Fig. 3.7 and 3.8 using a 10% spectral bandwidth for a contrast of 3×10^{-5}. Although optimized for only 10%, this specific design performs exceptionally well and a contrast of $< 10^{-4}$ is obtained for a spectral bandwidth of

18% centered around the design wavelength.

3.4 Single-mode fiber coronagraph properties

In this section we show the properties of this new coronagraph and perform parameter studies on the fixed parameters in Table 3.1. We discuss the mode field diameter of the single-mode fiber in Sect. 3.4.1, throughput and inner working angle in Sect. 3.4.2, the chromatic response in Sect. 3.4.3, the tip-tilt sensitivity of the SCAR designs in Sect. 3.4.4 and the sensitivity of other modes in Sect. 3.4.5.

3.4.1 Fiber mode field diameter

The phase plate reduces the throughput of planet light. This reduction will also affect the optimal value of the mode field diameter. Smaller mode field diameters result in larger back-propagated fiber modes in the focal plane, which makes it easier to squeeze the three rings necessary for the second-order nulling into this mode. Therefore, we expect higher Strehl ratios (the throughput relative to the unaltered PSF throughput) as the mode field diameter becomes smaller. This is superimposed on the actual throughput of the unaltered PSF. Both curves are shown in Fig. 3.10 for both a clear and the VLT aperture.

3.4.2 Throughput and inner working angle

The throughput shown here is the fractional transmission of light from the entire pupil into the central single-mode fiber: it includes all theoretical terms as listed in Table 3.1. It however excludes all experimental terms. It is clear that larger microlenses generally give a better throughput, as is expected. Additionally, we can see that the optimal mode field diameter as a function of microlens diameter for the unaltered PSF moves to larger mode field diameters, as it is essentially matching the Airy-core width rather than the size of the microlens itself. The optimal mode field diameter for the SCAR however stay the same as smaller mode field diameter have an advantage in their Strehl ratio.

Figure 3.11 shows the throughput (η_p) of the coronagraph for different values for the microlens diameter. The mode field radius of the fiber was fixed at $w = 0.85\lambda/D_{\mathrm{mla}}$ and the contrast at 3×10^{-5}. We adopted a $0.1\lambda/D$ rms telescope tip-tilt jitter with a normal distribution, corresponding to a 2 mas rms tip-tilt jitter at a wavelength of $\lambda = 750$nm. This level of tip-tilt

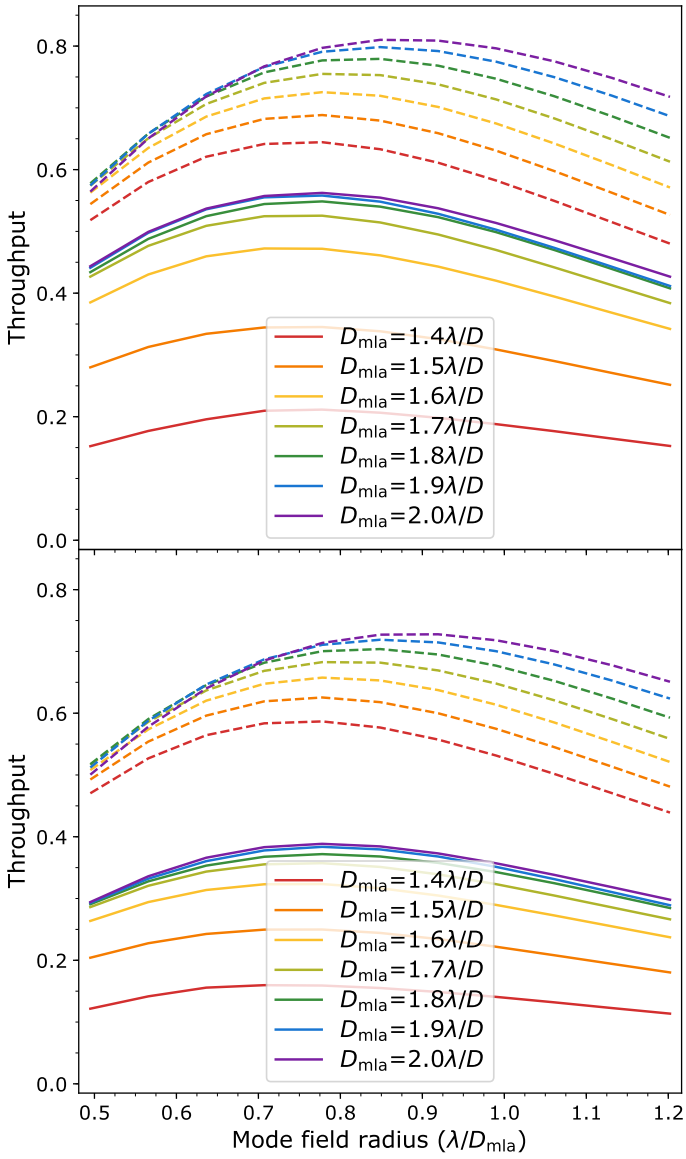


Figure 3.10: Throughput of the central lenslet as a function of mode field radius for various values of the microlens diameter. Each data point represents a different SCAR design. Solid lines indicate the SCAR PSFs, dashed lines an unaltered PSF. The top panel shows the throughput for an clear aperture, the bottom panel for a VLT aperture.

jitter was chosen to mimic a SPHERE-like adaptive optics system, according to Fusco et al. (2016). The throughput is averaged over all pointing positions and over the full 10% spectral bandwidth. The throughput of off-center fibers is negligible to that of the central fiber: all throughput is concentrated in only one single-mode fiber.

As the stellar throughput is only minimized for the first ring of microlenses, we do not have much throughput beyond the edge of the first off-axis microlens. Designs can be made for more than one ring of microlenses, although this complicates the design procedure and will be discussed in future work. Extremely close to the star, we have no throughput, as the Airy core is still mostly on the central lenslet. At $\sim 0.5\lambda/D$ the throughput starts to rise, reaching a maximum at the center of the first microlens. A throughput of $\sim 50\%$ of the maximum SCAR throughput is already reached at $\sim 1\lambda/D$, which is the usual definition of inner working angle. Also, up to microlens diameters of $\sim 1.8\lambda/D$ the throughput at small angular separations ($< 1\lambda/D$) does not change much, but the maximum throughput still increases. For larger microlens diameter we still gain in throughput at the center of the microlens, however the throughput at these small angular separations starts to suffer, which is especially visible in the $D_{\mathrm{mla}} = 2.0\lambda/D$ throughput curve.

Lines for the theoretical throughput of other coronagraphs are overplotted in Fig. 3.11. Perfect coronagraphs refer to the notion introduced by Cavarroc et al. (2006) and Guyon et al. (2006). A second-order perfect coronagraph removes a constant term from the pupil-plane electric field. A fourth-order additionally removes the x and y components from the electric field. A sixth-order perfect coronagraph furthermore removes the x^2, xy and y^2 modes from the electric field. For the theoretical coronagraphs, the throughput is calculated using a circular aperture of $0.7\lambda/D$ centered around the off-axis planet. We can see that the SCAR throughput lags behind the theoretical second-order coronagraph, but stays close to the fourth-order and beats the sixth-order at angular separations $< 1.7\lambda/D$.

Figure 3.12 shows the relative integration time (η_s/η_p^2) under the same conditions as in Fig. 3.11. We now see that, even though the throughput of the theoretical second-order coronagraph is good, its integration time is minor because it does not outweigh the loss in starlight suppression. SCAR however performs similar to theoretical fourth-order coronagraph for angular separations $< 1.8\lambda/D$. A sixth-order coronagraphs does even better, but suffers from a lack of throughput which becomes noticeable in cases where the raw contrast (η_s/η_p) is limited, which is the case in any

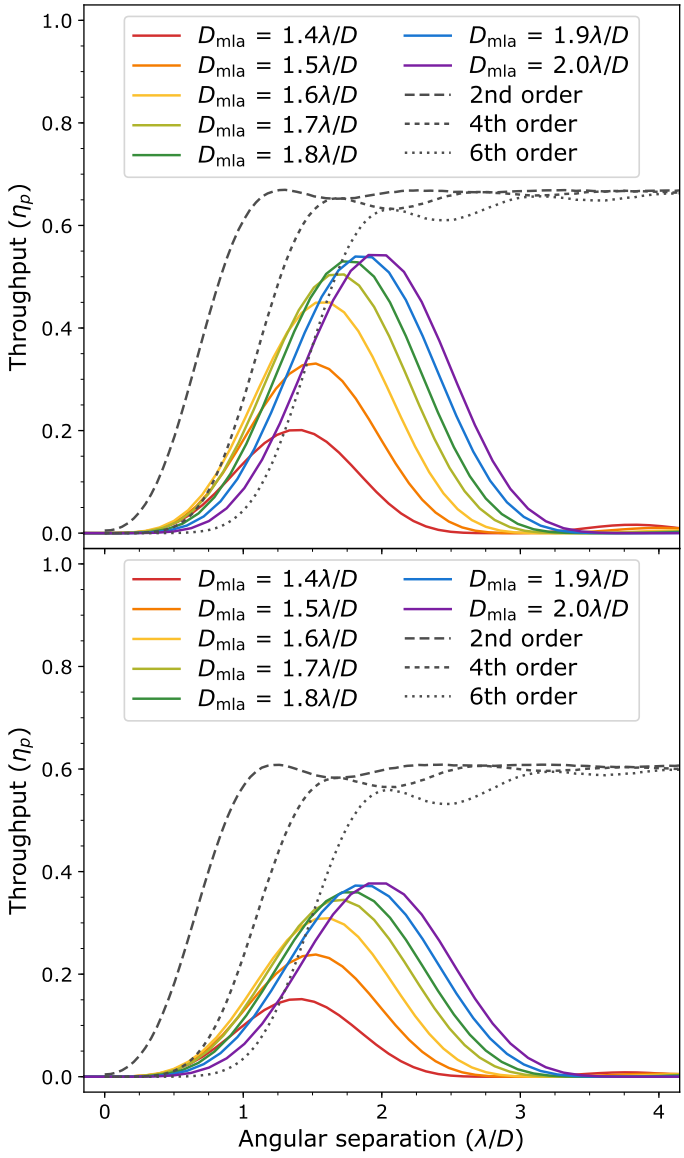


Figure 3.11: Throughput as function of off-axis angle for various values of the microlens diameter. The throughput of a perfect second, fourth and sixth-order coronagraph is also plotted. The top panel shows the throughput for an unobstructed aperture, the bottom panel for the VLT aperture. The throughput is calculated for the optimal position angle of the off-axis source (ie. directly across an off-axis microlens). The throughput for the theoretical coronagraphs is taken to be the fractional flux within an aperture of radius $0.7\lambda/D$ centered around the planet.

ground-based telescope. This suggests that SCAR is a close-to-optimal coronagraph.

3.4.3 Spectral bandwidth

Figures 3.7 and 3.8 show the chromatic response for all designs. Every design exhibits the double-dipped structure of the second-order null on the fiber. For all designs with a non-zero spectral bandwidth, we can also see that the contrast is hard to achieve on the long wavelength side. At these longer wavelengths the bright Airy core starts to grow into the microlens array. This means that the second Airy ring needs to be made much brighter to compensate, which requires substantial deviations in the phase pattern. Qualitatively, the location of the second null is chosen by the optimizer such that the spectral bandwidth requirement is reached.

3.4.4 Tip-tilt sensitivity and stellar diameter

Figure 3.13 shows the average contrast (η_s/η_p) over the full 10% bandwidth as a function of tip-tilt error upstream of the fiber injection unit. The double-dip structure is again clearly visible, which greatly improves the tip-tilt response. Both coronagraph designs achieve a tip-tilt stability of $\sim 0.1\lambda/D$ rms. This figure also shows that the SCAR coronagraph is insensitive for stars with an angular diameter up to $0.1\lambda/D$.

3.4.5 Sensitivity to other aberrations

To show the sensitivity to other aberrations, we perform a sensitivity analysis on the SCAR coronagraph: we aim to find the mode basis of orthogonal modes ordered by their sensitivity. These principal modes can be found by taking the first-order Taylor expansion of the phase in the pupil-plane around the nominal position. In this way a linear transformation G_λ can be constructed from a phase deformation $\delta\varphi$ to the resulting change in electric field in the fibers δE_λ:

$$\delta E_\lambda = G_\lambda \delta\varphi. \tag{3.6}$$

G_λ and δE_λ both depend on wavelength as the response of the coronagraph is inherently chromatic. A singular value decomposition of the matrix G_λ yields the monochromatic principal phase modes of the coronagraph. The corresponding singular values denote the importance of those modes. This expansion is similar to the one used in electric field conjugation (Give'on et al., 2007).

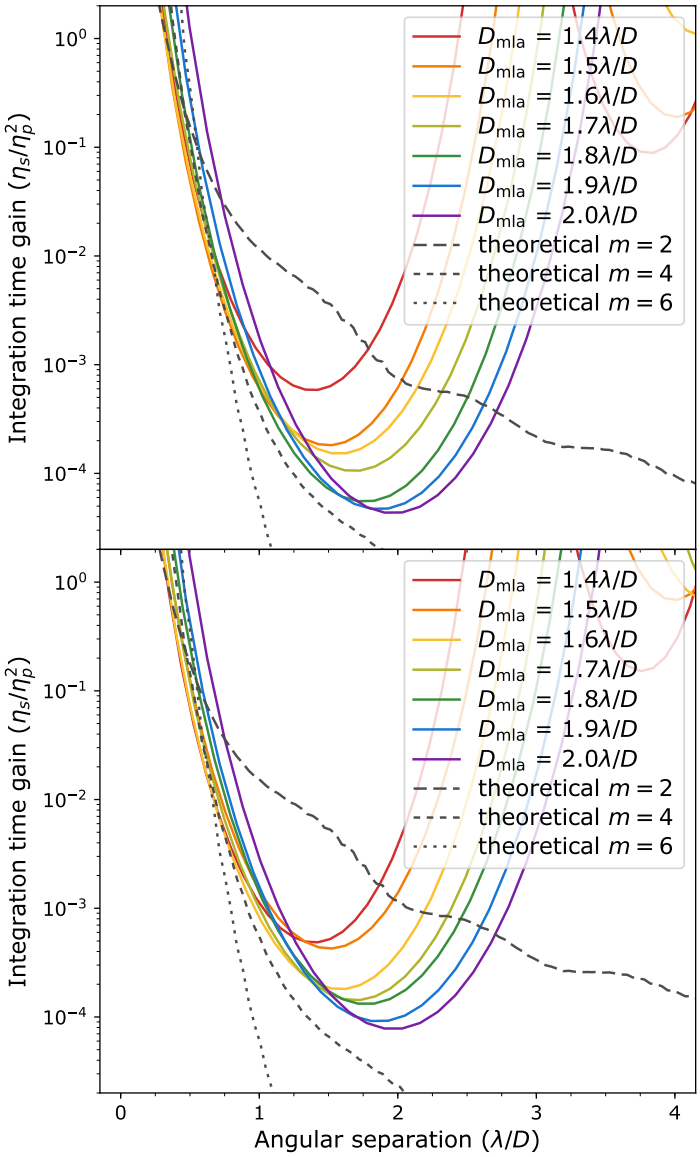


Figure 3.12: Relative integration time (η_s/η_p^2) as function of off-axis angle for various values of the microlens diameter. The top panel shows the relative integration time for an unobstructed aperture, the bottom panel for the VLT aperture. The integration time for the theoretical coronagraphs is calculated on the flux within an aperture of radius $0.7\lambda/D$ centered around the planet.

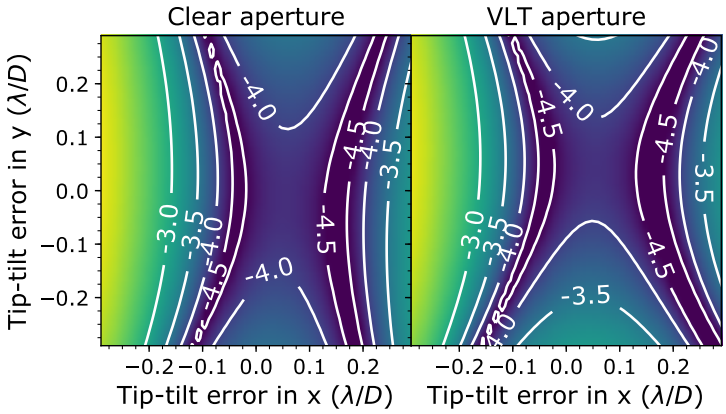


Figure 3.13: Map of the worst contrast through an off-axis fiber over a 10% bandwidth as a function of tip-tilt error upstream of the microlens array. This shows that the coronagraph is reasonably stable against tip-tilt, allowing for a $\sim 0.15\lambda/D$ tip-tilt error until the contrast drops to 10^{-4}. The contour labels indicate $\log_{10}(\eta_s/\eta_p)$.

Broadband principal modes can be obtained by stacking several G_λ matrices for wavelengths within the spectral bandwidth into a single matrix G as

$$\begin{bmatrix} \delta E_{\lambda_1} \\ \delta E_{\lambda_2} \\ \vdots \\ \delta E_{\lambda_N} \end{bmatrix} = \begin{bmatrix} G_{\lambda_1} \\ G_{\lambda_2} \\ \vdots \\ G_{\lambda_N} \end{bmatrix} \delta\varphi. \tag{3.7}$$

A singular value decomposition on the matrix G now yields the broadband principal modes. The singular values are now indicative of the amount of electric field each phase mode induces in the fibers as a function of wavelength. This method is again similar to the one used in broadband electric field conjugation (Give'on et al., 2007).

Figure 3.14 shows the broadband principal modes for the SCAR design for the VLT aperture, along with their singular values. Only six modes are important for the final contrast. Naively we would expect two modes per fiber, so 12 modes in total, as we need to control both the real and imaginary part of the electric field. However in our case the system shows an antihermitian symmetry: the transmitted electric field on fibers in opposite points in the focal plane are not independent if only small phase aberrations are present. This means that one phase mode determines the electric field

for both fibers so that only two modes per two fibers are needed. Only half of the original 12 modes determine the contrast in a monochromatic system, meaning that six modes are left. The omitted six modes correspond to amplitude errors. For the broadband principal modes we of course expect some additional modes with low importance, corresponding to the spectral bandwidth increase. The first six of these are shown as modes 7 to 12 in Fig. 3.14.

The first principal mode is trefoil, which is the result of the six-fold symmetry in the fiber locations. The second trefoil is missing: its corresponding electric field change is completely filtered out by the single-mode fibers: it creates radial first-order nulls directly on the center of every lenslet. Other important modes include secondary-astigmatism-like modes for modes 2 and 3, coma-like modes for modes 4 and 5, and a perturbed spherical mode for mode 6.

In Fig. 3.15 we show the response of the SCAR coronagraph to each of these modes. The average normalized irradiance is the average raw contrast over the full spectral bandwidth and all six fibers in the first ring of microlenses. It can be seen that the response for the first six modes is generally a lot stronger than that of the last six. The average raw contrast is limited by the residual leakage from the SCAR coronagraph at the low aberration end, while the contrast increases when more aberration is added. Additionally, we can see different behaviour, depending on the mode that we are looking at. Some modes (modes 1 to 6, 9, 11 and 12) show quadratic behaviour, while others (modes 7, 8 and 10) show a fourth order behaviour. Also notice that mode 7 actually increase the average raw contrast. While this may seem an impossibility, as the SCAR coronagraph is optimized for raw contrast, in these cases the fibers on one side become brighter than the contrast constraint, while the others compensate by becoming dimmer. This would actually decrease the effectiveness of the coronagraph while still increasing the average raw contrast.

3.5 Comparison to the vortex coronagraph

The described performance of the SCAR coronagraph begs the question on how it compares to other coronagraphs using single-mode fibers. In this section we will provide a comparison of the SCAR coronagraph with the single-mode fiber injection unit with vortex coronagraph proposed by Mawet et al. (2017), and a comparison with a multiplexed fiber injection unit as shown in Sect. 5.5 behind a vortex coronagraph. Figure 3.16 shows

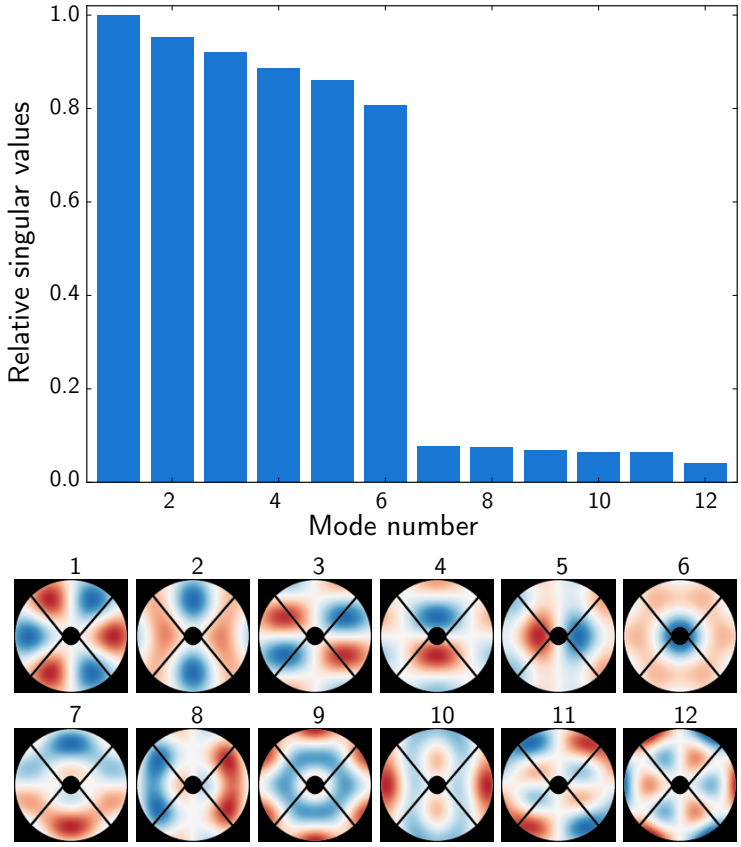


Figure 3.14: Principal phase modes for the SCAR coronagraph using the design described in the text. The top panel shows the singular value of each mode, indicating its significance for the obtained contrast after phase correction. The bottom panel shows the pupil-plane phase for each mode.

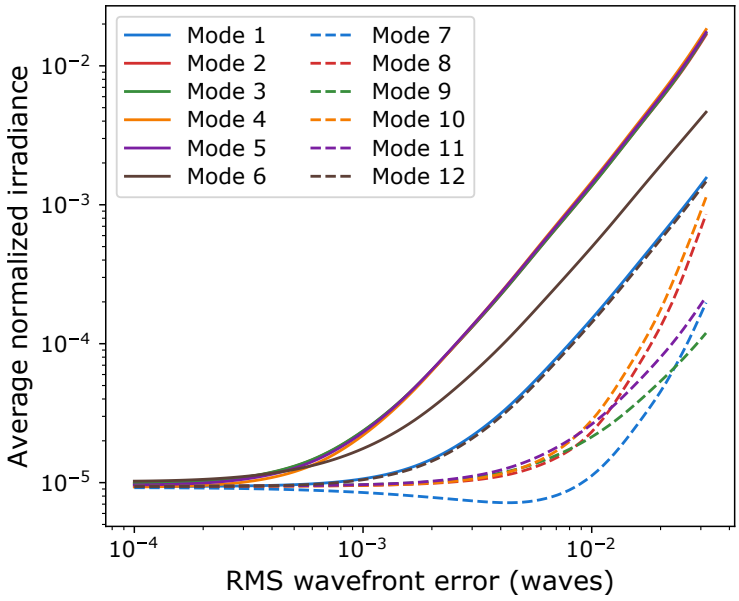


Figure 3.15: Average normalized irradiance for each of the modes in Fig. 3.14. The normalized irradiance is averaged over all fibers in the first ring of microlenses, and over the full 10% spectral bandwidth. These curves can yield constraints on the post-AO residual variance for each of the shown modes. Higher-order modes in this expansion generally have less response and do not have to be controller as well as low-order modes.

the throughput and relative integration time of a conventional vortex coronagraph using a clear aperture. A conventional intensity detector (ie. sum of all flux in an aperture centered around the star) is compared with a single-mode fiber centered around the planet. The mode field diameter was optimized for maximum throughput of the PSF without a coronagraph. To better capture the best and worst case performance for comparison with the vortex coronagraph, we take a telescope tip-tilt jitters of $0.05\lambda/D$ for the best case performance and $0.15\lambda/D$ for the worst case performance Similarly to Fig. 3.11, a telescope tip-tilt jitter of $0.1\lambda/D$ rms was chosen.

For charges $m = 4$ and $m = 6$ a decrease in throughput, compared to the vortex coronagraph with multimode fiber, can be seen, approximately corresponding with the maximum coupling of an Airy pattern through a Gaussian single-mode fiber. Additionally the vortex in the focal plane imprints a phase ramp on off-axis sources, which degrades throughput even further. This effect is more pronounced with smaller angular separations and higher charge vortices. Most of this phase ramp can be easily negated by tilting the fiber slightly, depending on the focal-plane position of the planet. We can see that the SCAR coronagraph wins in throughput compared to vortex coronagraphs with charger $m > 2$. However the charge 2 vortex coronagraph does not suppress the star very well, resulting in a moderate relative integration time compared to the SCAR coronagraph.

Figure 3.17 shows the throughput and relative integration time of a vortex coronagraph charge $m = 2$ through a microlens-fed single-mode fiber-array as the fiber injection unit. The diameter of the microlenses is varied from 1.4 to $2.0\lambda/D$, and the mode field diameter is optimized for maximum throughput of an Airy pattern. For large angular separations the throughput oscillates due to the transmission of the microlens array as shown in Fig. 3.4. For the first ring of microlenses the throughput rises quickly, again reaching its maximum at the center of the microlens. Even though the coupling efficiency for smaller microlenses is higher, the geometric throughput decreases more rapidly. A trade-off between these two throughput terms leads to an optimal microlens diameter of $\sim 1.8\lambda/D$ for the vortex coronagraph as well. The throughput at this microlens diameter is comparable to the performance of the SCAR coronagraph even though the vortex coronagraph has a more complicated optical setup. This multiplexed single-mode fiber vortex coronagraph however does not suppress starlight as well as the SCAR, leading to a worse relative integration time. The vortex coronagraph however does have the advantage of an infinite outer working angle.

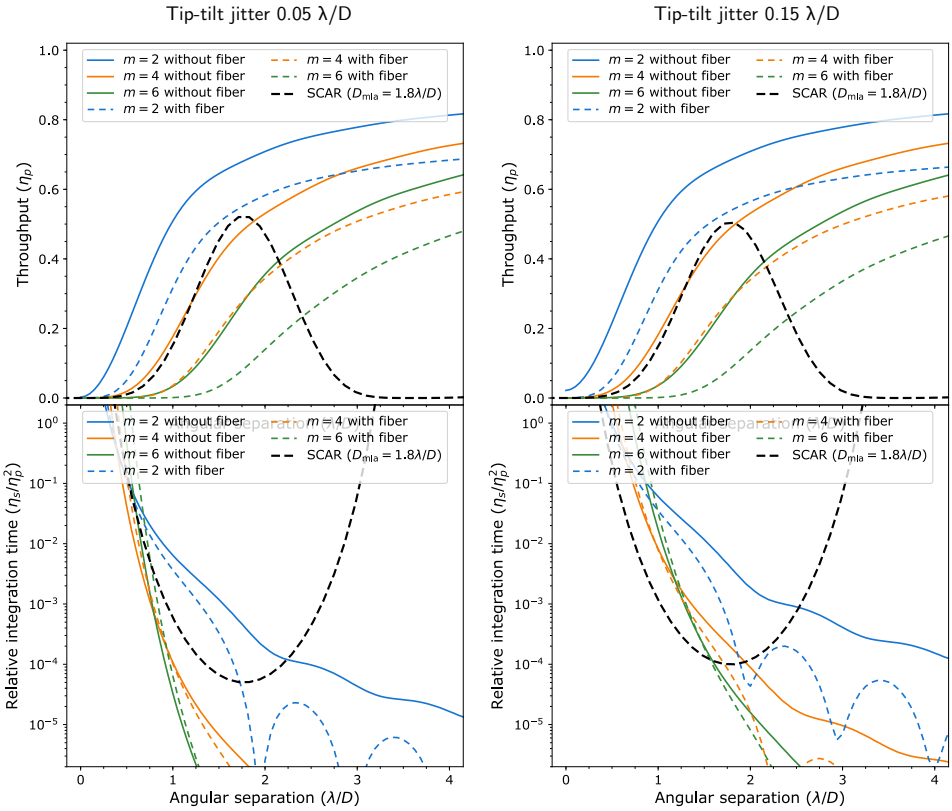


Figure 3.16: The throughput (η_p) and relative integration time (η_s/η_p^2) for a vortex coronagraph using a clear aperture for charges $m = 2, 4, 6$ using a conventional intensity detector or a single-mode fiber centered around the planet. A telescope tip-tilt jitter of (left column) $0.05\lambda/D$ or (right column) $0.15\lambda/D$ rms was taken into account. These values correspond to a best and worst case performance of the adaptive optics system. The throughput is calculated for the optimal position angle of the off-axis source (ie. directly across an off-axis microlens). For the vortex coronagraph without fiber, the throughput and relative integration time is calculated on an aperture of radius $0.7\lambda/D$ centered around the planet.

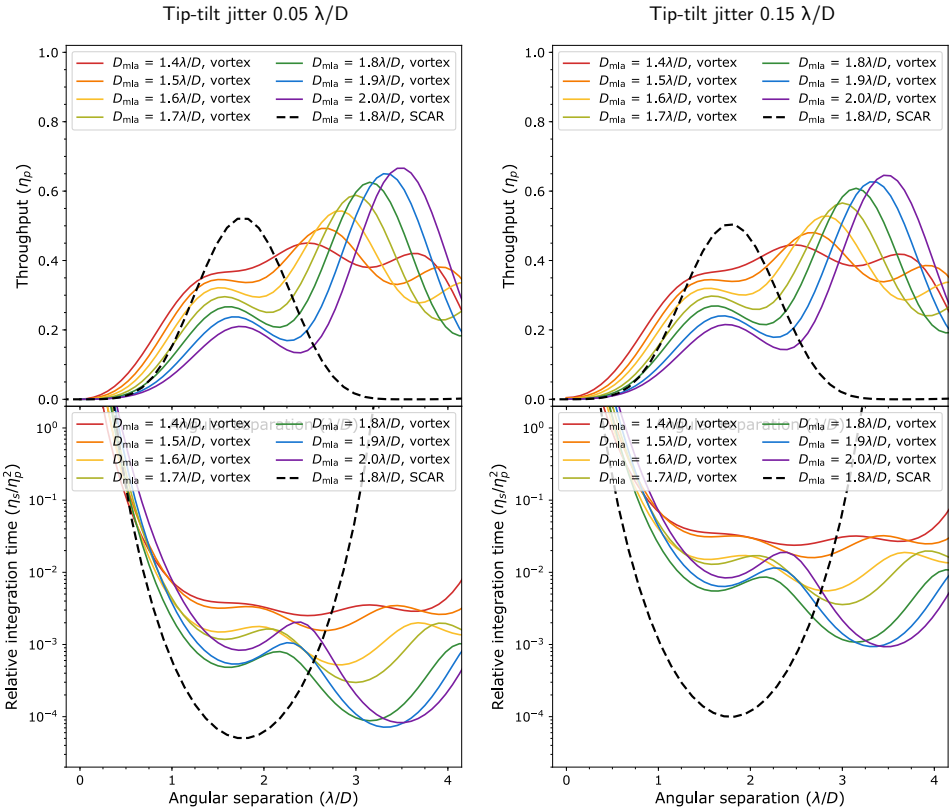


Figure 3.17: The throughput (η_p) and relative integration time (η_s/η_p^2) of a charge $m = 2$ vortex coronagraph on a microlens-fed single-mode fiber array. The diameter of the microlenses is varied from 1.4 to 2.0λ/D, and the mode field diameter is optimized for maximum throughput of an Airy pattern. The throughput of the SCAR coronagraph designed for 1.8λ/D microlens diameter was added for comparison. The assumed tip-tilt jitter is (left column) 0.05λ/D or (right column) 0.15λ/D rms. These values correspond to a best and worst case performance of the adaptive optics system. The throughput is calculated for the optimal position angle of the off-axis source (ie. directly across an off-axis microlens).

3.6 Conclusion

In this paper we described the principle behind coronagraphs leveraging the design freedom offered by the use of single-mode fibers as a mode filter. We have shown the properties of a microlens-array fed single-mode fiber-array, making it possible to perform exoplanet searches. We combined this with a pupil-plane phase plate, yielding the SCAR coronagraph and presented the following advantages of this new coronagraph:

1. *Low inner working angles.* Inner working angles as low as $1\lambda/D$ can be reached using current designs.

2. *High throughput.* These designs reach a maximum throughput of 50% and 30% for a clear and the VLT aperture, respectively.

3. *High contrast.* Starlight can be suppressed to $< 3 \times 10^{-5}$ for the six fibers surrounding the star over the full spectral bandwidth until the throughput starts to drop.

4. *Broad spectral bandwidth.* This suppression is achieved over the full 20% spectral bandwidth.

5. *Robust against tip-tilt errors.* The SCAR coronagraph is stable against $\sim 0.1\lambda/D$ rms tip-tilt errors upstream due to the use of second order nulling on the fibers.

6. *Residual speckle suppression.* Residual speckles are are reduced by $\sim 3\times$ in intensity, due to the coupling efficiency of a random electric field into the single-mode fibers.

All advantages can be obtained into a single SCAR design. All these advantages make this coronagraph a prime candidate for future upgrades of extreme AO systems. In particular, the SCAR coronagraph is perfectly suited for spectral characterization of Proxima b: it satisfies all coronagraphic requirements set by Lovis et al. (2017). A companion paper (Haffert et al., 2018) provides a tolerancing study for this specific application.

Future research will explore active control of the fiber throughput of the SCAR coronagraph. Application of the SCAR methodology to other coronagraphs is also left for future research. An interesting example in this case is the design of a Lyot-plane mask for a conventional Lyot or vortex coronagraph, akin to Ruane et al. (2015a). Even optimizing the focal-plane mask itself might be realizable for the fiber array in these coronagraphs (Ruane et al., 2015b).

Another avenue is the use of photonic technologies to further process the light in the fibers. A simple example is the use of fiber Bragg gratings for filtering the atmospheric OH lines (Trinh et al., 2013). Another example is building a phase-shifting interferometer of the six fibers. This will provide information about the coherence of the light in each of the fibers with respect to the star, and would allow for synchronous interferometric speckle subtraction (Guyon, 2004).

References

Aime, C., & Soummer, R. 2004, The Astrophysical Journal Letters, 612, L85

Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, Nature, 536, 437

Beuzit, J.-L., Feldt, M., Dohlen, K., et al. 2008, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 701418–701418

Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, The Astrophysical Journal, 736, 19

Brogi, M., Snellen, I. A., de Kok, R. J., et al. 2012, Nature, 486, 502

Carlotti, A., Kasdin, N. J., Vanderbei, R. J., & Riggs, A. E. 2013, in SPIE Optical Engineering+ Applications, International Society for Optics and Photonics, 88641Q–88641Q

Cavarroc, C., Boccaletti, A., Baudoz, P., Fusco, T., & Rouan, D. 2006, Astronomy & Astrophysics, 447, 397

Codona, J., Kenworthy, M., Hinz, P., Angel, J., & Woolf, N. 2006, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 62691N–62691N

Corbett, J. C. 2009, Optics express, 17, 1885

Fusco, T., Sauvage, J.-F., Mouillet, D., et al. 2016, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 99090U–99090U

Give'on, A., Kern, B., Shaklan, S., Moody, D. C., & Pueyo, L. 2007, in Optical Engineering+ Applications, International Society for Optics and Photonics, 66910A–66910A

Gurobi Optimization, I. 2016, Gurobi Optimizer Reference Manual. `http://www.gurobi.com`

Guyon, O. 2004, The Astrophysical Journal, 615, 562

—. 2005, The Astrophysical Journal, 629, 592

Guyon, O., Pluzhnik, E., Kuchner, M., Collins, B., & Ridgway, S. 2006, The Astrophysical Journal Supplement Series, 167, 81

Haffert, S. Y., Por, E. H., Keller, C. U., et al. 2018, arXiv e-prints, arXiv:1803.10693. `https://arxiv.org/abs/1803.10693`

Haguenauer, P., & Serabyn, E. 2006, Applied optics, 45, 2749

Hanot, C., Mennesson, B., Martin, S., et al. 2011, The Astrophysical Journal, 729, 110

Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, Publications of the Astronomical Society of the Pacific, 127, 890

Kawahara, H., Murakami, N., Matsuo, T., & Kotani, T. 2014, The Astrophysical Journal Supplement Series, 212, 27

Kipping, D. M., Cameron, C., Hartman, J. D., et al. 2017, The Astronomical Journal, 153, 93

Konopacky, Q. M., Barman, T. S., Macintosh, B. A., & Marois, C. 2013, Science, 339, 1398

Lovis, C., Snellen, I., Mouillet, D., et al. 2017, Astronomy & Astrophysics, 599, A16
Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proceedings of the National Academy of Sciences, 111, 12661
Marcuse, D. 1978, JOSA, 68, 103
Marois, C., Lafreniere, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, The Astrophysical Journal, 641, 556
Martin, S., Serabyn, E., Liewer, K., et al. 2008, in Proc. SPIE, Vol. 7013. https://doi.org/10.1117/12.789484
Mawet, D., Pueyo, L., Lawson, P., et al. 2012, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 844204–844204
Mawet, D., Ruane, G., Xuan, W., et al. 2017, The Astrophysical Journal, 838, 92
Otten, G. P., Snik, F., Kenworthy, M. A., et al. 2017, The Astrophysical Journal, 834, 175
Pepe, F. A., Cristiani, S., Lopez, R. R., et al. 2010, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 77350F–77350F
Por, E. H. 2017, in Techniques and Instrumentation for Detection of Exoplanets VIII, Vol. 10400, International Society for Optics and Photonics, 104000V
Riaud, P., & Schneider, J. 2007, Astronomy & Astrophysics, 469, 355
Ruane, G. J., Huby, E., Absil, O., et al. 2015a, Astronomy & Astrophysics, 583, A81
Ruane, G. J., Swartzlander, G. A., Slussarenko, S., Marrucci, L., & Dennis, M. R. 2015b, Optica, 2, 147
Snellen, I., de Kok, R., Birkby, J., et al. 2015, Astronomy & Astrophysics, 576, A59
Snellen, I. A., Brandl, B. R., de Kok, R. J., et al. 2014, Nature, 509, 63
Snik, F., Otten, G., Kenworthy, M., et al. 2012, in SPIE Astronomical Telescopes+ Instrumentation, 84500M–84500M
Sparks, W. B., & Ford, H. C. 2002, The Astrophysical Journal, 578, 543
Trinh, C. Q., Ellis, S. C., Bland-Hawthorn, J., et al. 2013, The Astronomical Journal, 145, 51
Wang, J., Mawet, D., Ruane, G., Hu, R., & Benneke, B. 2017, The Astronomical Journal, 153, 183

Chapter 4

The Single-mode Complex Amplitude Refinement (SCAR) coronagraph

II. Lab verification, and toward the characterization of Proxima b

Adapted from
S. Y. Haffert, **E. H. Por**, C. U. Keller, M. A. Kenworthy,
D. S. Doelman, F. Snik and M. J. Escuti (2020), A&A 635, A56

Abstract
We present the monochromatic lab verification of the newly developed SCAR coronagraph that combines a phase plate (PP) in the pupil with a microlens-fed single-mode fiber array in the focal plane. The two SCAR designs that have been measured, create respectively a 360 degree and 180 degree dark region from $0.8 - 2.4\lambda/D$ around the star. The 360 SCAR has been designed for a clear aperture and the 180 SCAR has been designed for a realistic aperture with central obscuration and spiders. The 360 SCAR creates a measured stellar null of $2 - 3 \times 10^{-4}$, and the 180 SCAR reaches a null of 1×10^{-4}. Their monochromatic contrast is maintained within a range of $\pm 0.16\lambda/D$ peak-to-valley tip-tilt, which shows the robustness against tip-tilt errors. The small inner working angle and tip-tilt stability makes the SCAR coronagraph a very promising technique for an upgrade of current high-contrast instruments to characterize and detect exoplanets in the solar neighborhood.

4.1 Introduction

We are currently at a breakthrough moment where more and more Earth-like exoplanets are being discovered. Every detection of Earth-sized planets brings us closer to finding life on another planet. The recent discovery of Proxima Centauri b (Anglada-Escudé et al., 2016) confirms that the solar neighborhood has many planets waiting to be discovered. From current surveys it is also clear that most of the planets in the habitable zone will have a separation close to the diffraction limit of current and future large telescopes. Characterization and detection of these planets can be done through high-contrast imaging, which overcomes the huge contrast between planet and star.

While the indirect methods have been very successful in discovering planets, direct imaging of exoplanets is lagging behind in the number of planets. This is mainly due to the difficulties involved in direct imaging. The largest problem is the close angular position of the planet to the star which is at best a few λ/D according to the current statistics of exoplanet orbits (Galicher et al., 2016). Here λ is the wavelength used to observe the system and D the telescope diameter.

The influence of the photon noise on the planet can be reduced by spatially separating the planet signal from the stellar signal. Current and future large optical telescopes have the resolution to resolve planets from their host star. This is done by using extreme adaptive optics (XAO) systems to enable imaging at the diffraction limit on ground based telescopes. The spatial separation also enables the use of coronagraphs to suppress the stellar light. This is the common approach on high contrast imaging (HCI) instruments like SPHERE (Beuzit et al., 2008), GPI (Macintosh et al., 2014) and SCExAO (Jovanovic et al., 2015).

Combining HCI with high-resolution spectroscopy (HRS) over a broad wavelength range gains further orders of magnitude in contrast close to the star (Riaud & Schneider, 2007; Sparks & Ford, 2002), because high-resolution spectra are able to exploit the difference in spectral lines between the star and planet. This difference can be due to a different Doppler velocity for reflected light and/or due to the presence of different molecular species. This technique has been successfully applied to characterize the atmosphere of several giant exoplanets (Brogi et al., 2012; Konopacky et al., 2013; Snellen et al., 2014). Recent papers (Kawahara et al., 2014; Snellen et al., 2015; Wang et al., 2017) show that this can be used as a robust post processing technique to remove residual stellar speckles which limit current HCI instruments (Aime & Soummer, 2004; Martinez et al., 2012). Snellen

et al. simulated a hypothetical Earth-like twin around Proxima Centauri for the European Extremely Large Telescope(E-ELT). The combination of HCI and HRS was able to detect and even characterize the Earth twin. The discovery of an actual Earth-like planet around Proxima Centauri makes this technique even more relevant as Lovis et al. show that an upgraded SPHERE (SPHERE+) can be used to characterize Proxima b if it is coupled to a high-resolution spectrograph. In this approach the focal plane of SPHERE+ would be coupled through a fiber link to the high-resolution spectrograph.

Mawet et al. (2017) argue that using a single-mode fiber (SMF) link between HCI and HRS instruments has an advantage over multi-mode fibers (MMF). A single-mode fiber is more robust against speckle noise due to the mode filtering capabilities. This property has been appreciated by the interferometry community, where single-mode fibers or waveguides are used to combine and filter multiple beams. Mawet et al. considers a system where the coronagraph and fiber injection unit (FIU) act separately on the stellar light. In the companion paper(Paper I) Por & Haffert (2017) we demonstrate the concept of the SCAR coronagraph, where am pupil plane phase plate is designed that uses the properties of the single-mode fiber to reach a deep null close to the star. This system is related to the Apodizing Phase Plate (APP) (Codona et al., 2006; Otten et al., 2017), which also uses an pupil plane phase optics to create dark holes in the PSF. The main difference is that the APP creates a dark hole by reducing the intensity, while SCAR changes the electric field such that the light can be rejected by a single-mode fiber. The SCAR coronagraph can work over a broad spectral bandwidth with high throughput and is tip-tilt insensitive to a large extent. It is well suited to be used as the interface between high-contrast imaging instruments and high-resolution spectrographs. This is the first system that combines the FIU and coronagraph in a single unit.

The SCAR coronagraph works on the basis of electric field filtering by electric field sensitive photonics. The implementation of this work uses single-mode fibers. With the advances of AO and especially extreme AO it is possible to achieve high coupling efficiency into SMFs (Bechter et al., 2016; Jovanovic et al., 2017). The amount of light that couples into a single-mode fiber is defined by the coupling efficiency

$$\eta = \frac{\left|\int E_{\text{in}}^{*} E_{\text{SMF}} dA\right|^2}{\int |E_{\text{in}}|^2 dA \int |E_{\text{SMF}}|^2 dA}. \tag{4.1}$$

Here η is the relative amount of light from E_{in} that is coupled into a single-mode fiber with mode profile E_{SMF}. For a SMF the mode profile

is Gaussian. The light that couples into a SMF is effectively averaged by a Gaussian weighing function. Due to this property if the electric field is zero on average it will not couple into the fiber and is rejected. A remarkable property here is that the intensity does not have to be zero, while the electric field can be. This is a much less stringent requirement than the zero intensity for normal high-contrast imaging. The phase plate is used to modify the point spread function (PSF) in such a way that the stellar light couples very badly into the fiber, but the planet still couples well. This coherent imaging approach is very reminiscent of interferometry where these kind of approaches have been used (Angel et al., 1986; Labadie et al., 2007). Because of the benefits of coherent imaging such approaches are now also starting to be exploited for conventional direct imaging (Mawet et al., 2017).

In this paper we show the first lab verification of the SCAR coronagraph and look at the manufacturing feasibility with Monte Carlo simulations. Section 2 describes the optical setup for the measurements and the lab results. And in Section 3 the manufacturing requirements are derived and a Monte Carlo simulation is done to estimate the expected performance. Section 4 summarizes the results of our study.

4.2 Optical setup details and first results

4.2.1 Lab setup description

The SCAR coronagraph uses a single-mode fiber array fed by a microlens array. A detailed theoretical description can be found in Paper I. To measure the SCAR performance we created a setup that can emulate the measurements of a microlens fed fiber array. The setup was built in Leiden on a vibration damped optical table in air without any active controlled components. The lab setup used for the measurements can be seen in Figure 4.1. The input source to our setup was a single-mode fiber with a 2.8-4.1 μm Mode Field Diameter(MFD) at 488 nm, that is fed by a helium neon laser. The fiber was mounted on a XYZ-translation stage. A Thorlabs AC508-1000-B achromatic doublet with a focal length of 1000 mm collimates the fiber. And just before the conjugated plane we placed a pupil stop with a diameter of 3.8 mm. This pupil diameter combined with the 1000 mm focal length ensures that the input source's MFD $\approx 0.02\lambda/D$ which is much smaller than λ/D and therefore creates a good clean point source.

In the conjugated plane slightly behind the pupil stop we placed the

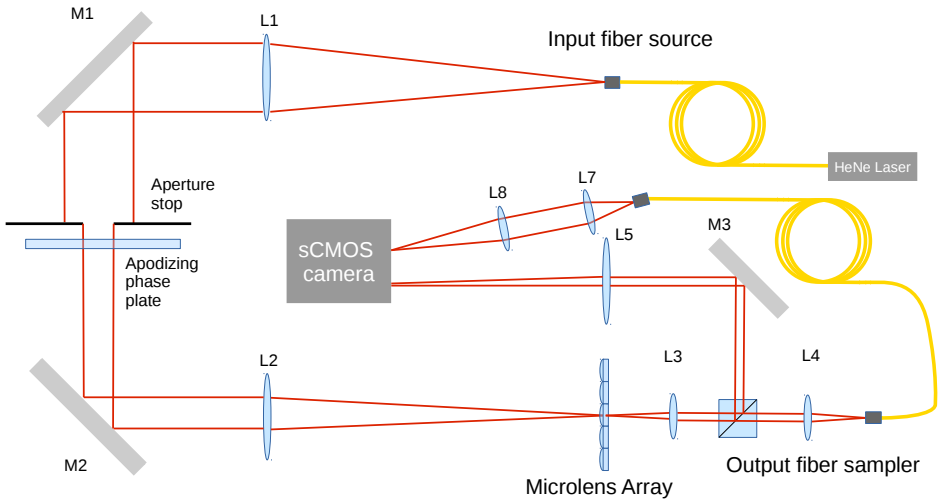


Figure 4.1: Schematic of the lab setup. The gray colored surfaces labeled M1 to M3 are mirrors. The blue shaded surfaces L1 to L8 are lenses. The yellow curves represent the single-mode fibers. The central wavelength of the setup is 0.633 nm.

phase plate which modifies the PSF. Another Thorlabs AC508-1000-B achromatic double focusses the light. The resulting PSF is sampled by Okotech APH-Q-P250-F2 hexagonal Micro Lens Array (MLA) with a pitch and diameter of 250 μm and a 2.18 mm focal length. The scale of the PSF is 166 μm per λ/D, this means that each microlens samples 1.5 λ/D. The MLA sampled PSF is relayed by a set of achromatic doublet lenses which have focal lengths of respectively 100 mm and 75 mm. In the intermediate collimated beam we placed a 90/10 beam splitter which splits 90 percent of the light toward a single-mode fiber and ten percent toward a camera. The transmitted beam was sampled by a LMA-8 photonic crystal fiber of NKT Photonics. This fiber has a mode field diameter of 8.4 μm $\pm$ 1.0 μm that is constant as a function of wavelength. The output of this fiber is reimaged on a Andor Zyla sCMOS camera by two positive achromatic doublets. The reflected beam is focused by a 500 mm lens to create an image of the MLA spots onto the same Andor camera. Both the microlens spots and the fiber coupling can be monitored at the same time in this way.

This setup was only able to measure the light coupling through the on-axis central lenslet because there is only a single fiber. So to emulate the

measurements through an off-axis lenslet the input source was shifted. This relaxes the alignment requirements because the fiber alignment is critical. The fiber has to be aligned within 1/8th of the MFD to create a deep null. Moving the fiber every time to a new lenslet to sample its throughput would be very time consuming due to this alignment requirement. Shifting the PSF is easier because $1\lambda/D$ is 166 μm. With a set of digitial micrometer actuators from Thorlabs we were able to shift the PSF with micron accuracy and precision. With this scanning strategy we were able to perform repeatable sub-λ/D shifts and measure the throughput as function of shift with respect to the central microlens. The actuator range is two inch which allowed us to scan a range of $\pm 150\lambda/D$.

4.2.2 Fiber alignment procedure

A misalignment of the output fiber can lead to a reduction in throughput and off-axis nulling. A good alignment is therefore critical. The fiber alignment is done in multiple steps to ensure good nulling. First the input source is aligned on the central microlens without a phase plate in the beam. The microlens spots, which can be viewed on the camera, should be radially symmetric in intensity after this first step because the PSF is radially symmetric. During the second step the fiber is coarsely aligned to find the brightest spot in the field which is done by moving the fiber such that the brightest spot can be seen by eye on the fiber face. The last step takes care of the fine alignment by moving the fiber such that the fiber throughput is maximized.

4.2.3 Apodizing phase plate designs

Two different SCAR designs have been measured in our setup. The first design is a phase pattern that generates a 360 degree dark region in the first ring of lenslets around the PSF with a contrast of 5×10^{-5} and a throughput of 30 percent. The throughput includes the coupling efficiency into the SMF. The second design is a 180 degree phase pattern with a central obscuration and spiders. This creates a one-sided dark region next to the PSF with a contrast of 1×10^{-5} and a 60 percent throughput. Both patterns are designed for a spectral bandwidth of 20 percent. The phase patterns with their corresponding PSFs can be seen in Figure 4.2.

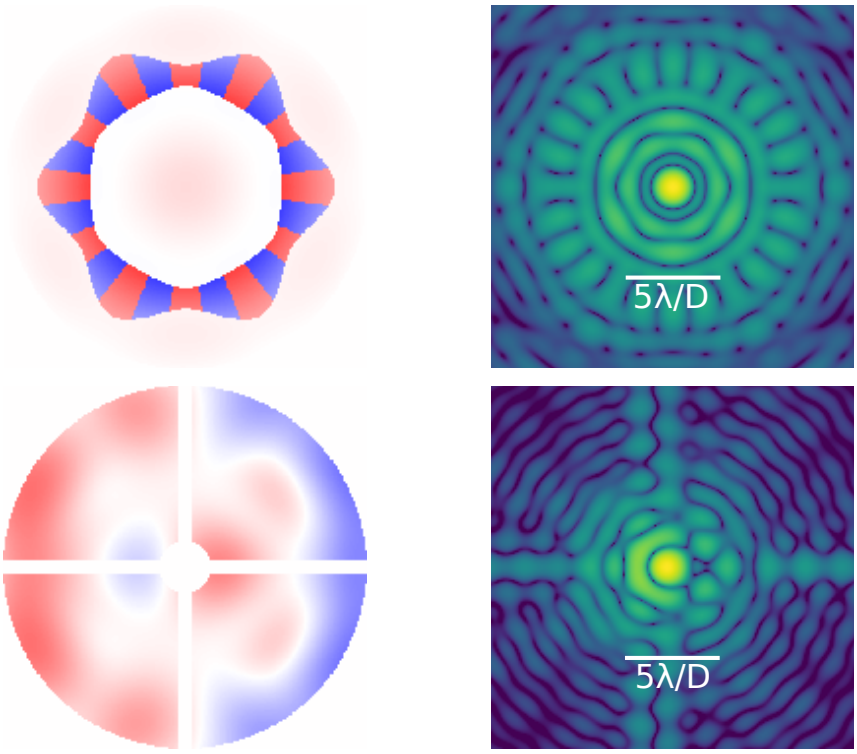


Figure 4.2: Left column shows theoretical phase pattern. Here blue is π phase, white is zero phase and red is $-\pi$ phase. The corresponding PSF is shown on the right with a 5 λ/D scale bar. Both images of the PSF are on log scale with the color scale shown on the right. The first phase pattern creates a 360 dark region for a clear aperture. The second pattern creates a one-sided dark region for a pupil that has a central obscuration and spiders.

4.2.4 Liquid crystal plate

The phase plates are manufactured by using a direct write approach where the fast axis of the liquid crystals is written with a laser (Miskiewicz & Escuti, 2014; Snik et al., 2012). The liquid crystals add a geometric phase to the incoming light which only depends on the angle of the fast axis. This then acts as an achromatic phase pattern with a chromatic piston term which can be ignored. An important aspect is that it acts on circular polarized light. Left circular and right circular polarizer light both get the same phase pattern but with a change of sign. Because of this it is important to separate the two polarizations. The separation is done by adding a tilt to the phase pattern, and because both polarization get opposite phase they split into different directions. This is also done in the grating-vAPP (Otten et al., 2014). If the liquid crystal plate is not perfectly half-wave there will be a leakage term that does not see the phase pattern. The leakage creates a normal Airy pattern. This can be seen in Figure 4.3.The leakage term can limit the contrast if the retarder substantially deviates from half wave. One way to reduced the effect of the leakage is to separate it like the grating-vAPP or use a technique like the double-grating vAPP where the leakage is scattered away (Doelman et al., 2017). For the results in this paper we used the grating-vAPP approach. The additional pair of spots in Fig. 4.3 are created by the grating mask (Doelman et al., 2017). The grating mask is a phase tilt that is applied outside the aperture. All the light that falls outside of the defined aperture diffracts due to the phase tilt. With the grating mask we can create a well defined aperture shape and size. Our phase plates were made with 8.75 μm LC pixels and the pupil itself is 3.8 mm allowing for 434 pixels across the pupil. The aperture spots fall off quickly enough that they do not influence the modified PSFs.

4.2.5 Lab setup results

The throughput as function of PSF position for the 360 SCAR can be seen in Fig. 4.4. The measured throughput is overlayed on the model curve. Given that the variables of our model are not fitted but taken as is from the manufacturing specifications, the measurements and the model agree very well. As said before this is the relative throughput as function of distance from a microlens center. This is not a contrast curve. For this system with microlenses the contrast curve is discrete and it only changes when going to another microlens as shown in Figure 4.5. The contrast is defined as the ratio between the coupling of the on-axis object and the

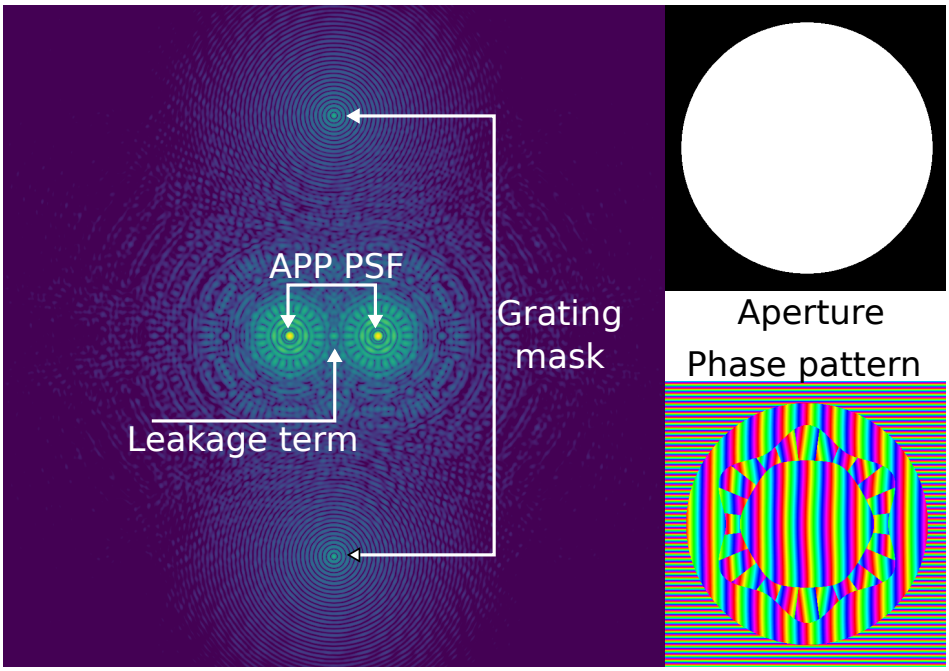


Figure 4.3: Left: Simulated focal plane of the phase pattern. Right: Phase and aperture. On the focal plane several PSF's can be seen. The left and right PSFs are the PSFs that contain the phase pattern. The top and bottom correspond to the light that scatters away due to the grating mask. And in the center there is a 1 percent leakage term. The horizontal PSFs are at $\pm 10\lambda/D$ and the aperture PSFs are at $\pm 50\lambda/D$

off-axis object. Figure 4.5 shows that the contrast changes as the source moves over the microlens due to a change in throughput.

On the linear scale it is easier to read off the throughput loss as the source is shifted. The 50 % throughput is at an offset of 0.5 λ/D. We define the inner working angle as the smallest angular separation where the throughput of the companion is equal to 50% of the maximum, and the outer working angle as the largest angular separation where the throughput is equal to 50% of the maximum. A source has a maximum throughput when it is in the middle of a microlens and needs to shift by 0.5 λ/D for it to reach half the maximum throughput. Therefore we can define the effective inner and outer working angles as $1.1\lambda/D$ and $2.1\lambda/D$ for the first ring of lenslets.

At $0.75\lambda/D$ the PSF is precisely on the edge between two microlenses. At this position 25% throughput remains in each of the two fibers. So a total throughput of 50% is achieved by combining multiple fiber outputs. If a binary system is observed with field rotation then the relative throughput of the source fluctuates between 50% and 100% as it rotates over the lenslets. For an off-axis source that is on the edge between the central lenslet and an off-axis lenslet then we can only capture 25% of the off-axis source, but with an impressive source separation of $0.75\lambda/D$.

The contrast changes as a source moves over the micro-lens array because the throughput changes. This is shown in Figure 4.5. The MMF shows the intrinsic contrast for normal imaging. Compared to a MMF a SMF already increases the contrast by a factor of 10 due to the rejection of the nonGaussian modes that are in the PSF. The 360 SCAR can in theory reach an average contrast of 5×10^{-5} but we are limited by static aberrations in the system, which limit the contrast to 3×10^{-4} at $-1.5\lambda/D$ and 2×10^{-4} at $1.5\lambda/D$. The decrease of contrast can be attributed to 10 nm rms low order wavefront aberration, but this is a very rough estimate. The change in contrast is symmetric around $1.5\lambda/D$. The contrast of the right fiber stays below 1×10^{-3} and the left fiber stays below 2×10^{-3}. Figure 4.5 also shows that if the source is on a different microlens it can still couple into the other microlenses and reach an acceptable contrast.

The broad gap around $1.5\lambda/D$ in Figure 4.4 gives insight into the wavelength scaling and the effects of jitter. When the wavelength changes then λ/D changes and we have to check the stellar throughput at a different part of the curve. So the width of the gap is a measure for the bandwidth. The spectral bandwidth is then roughly $2 * \Delta\theta/\theta_0 \approx 0.2$ with $\Delta\theta$ the gap width and θ_0 the gap center. Next to the wavelength response it also says

something about the monochromatic jitter resistance. If the star jitters a bit then the off-axis throughput is still low because of the gap. The gap where we still achieve the contrast for the 360 SCAR is between $-1.8\lambda/D$ and $-1.3\lambda/D$. The gap on the right is from $1.36\lambda/D$ to $1.7\lambda/D$. This demonstrates that the coronagraph should be able to handle $\pm 0.15\lambda/D$ monochromatic tip-tilt residuals. The SMF without any pupil phase optic in contrast has a very narrow rejection area. There is only one position at which it nulls the star and if there is a small amount of jitter the contrast quickly deteriorates to about 10^{-2}. This shows the advantage of SCAR, which creates a broad dark area. Because the system is completely passive it is also very robust. The dark region remained dark over several weeks.

The filtering effect of the -mode fiber can be seen in Figure 4.6, where the measured microlens spots around the deepest null are shown together with the fiber throughput. The spot has a triple peak structure that is created by the phase plate. The triple peak structure is the feature that increases the bandwidth and tip-tilt stability. This is characteristic of a second-order null. The triple peak structure can be seen in the middle frame of Figure 4.6. While the change in total power from frame to frame is small, the fiber throughput changes drastically. This shows the modal filtering capability of single mode fibers.

The throughput results for the 180 SCAR can be seen in Figure 4.7. The measurements reach a contrast of 1.15×10^{-4}. There is a mismatch between the measured and simulated throughput curves, but we can see that compared to the 360 design this design reaches a deeper contrast. The deepest part is also relatively flat between -1.7 and -1.4 λ/D. Because of the flat response the design can handle tip-tilt errors of $\pm 0.15\lambda/D$. The corresponding constrast curves are shown in Figure 4.8. The contrast curves are asymmetrical due to the asymmetric PSF that is created by the 180 SCAR.

Apart from maximizing the contrast it is important to optimize the absolute throughput of the planet light through the fiber. For the proposed system we have defined the coupling not as the total amount of light that couples into the fiber, but as the amount of light that falls within the micro-lens aperture that couples into the fiber. The total amount of light that couples into the fiber depends on how much of the PSF the micro-lens captures. Therefore there is a trade-off between spatial resolution and throughput (Por & Haffert, 2017). The 360 SCAR has a maximum coupling of 87% percent for a single micro-lens. Due to the hexagonal shape of the micro-lens the maximum coupling efficiency is slightly higher than the 82%

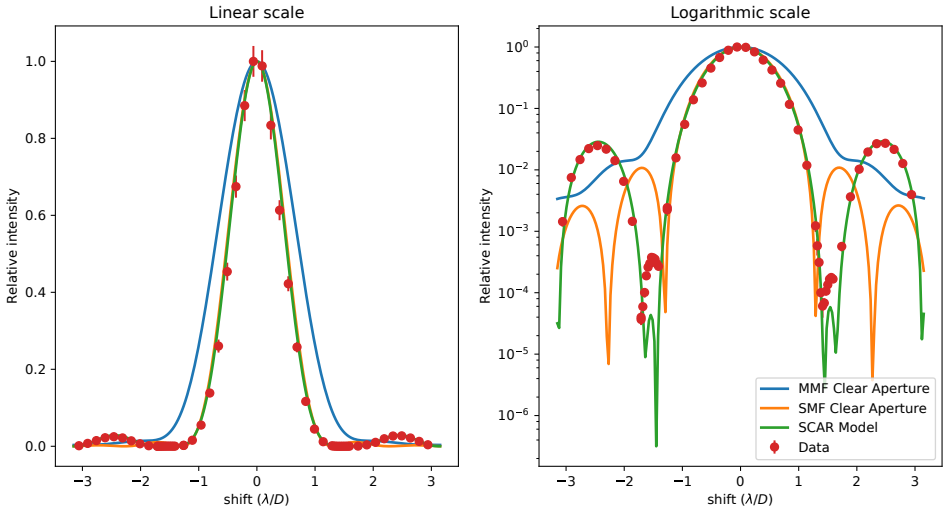


Figure 4.4: Left: Relative throughput on a linear scale as function of relative shift with respect to the microlens center. The red points are the measurements with errorbars due to random errors. The green line shows the model of this SCAR design. Right: Throughput on a logarithmic scale. The null is uneven between left and right and not as deep as designed. This is suspected to be caused by 10 nm rms residual low-order aberrations. The blue line shows the amount of light that falls on the microlens with a clear aperture and without SCAR. This is comparable to the contrast curve for normal imaging or using a multi-mode fiber. The orange line shows the normalized throughput with a unobstructed aperture and a SMF. The SMF shows a gain compared to the MMF.

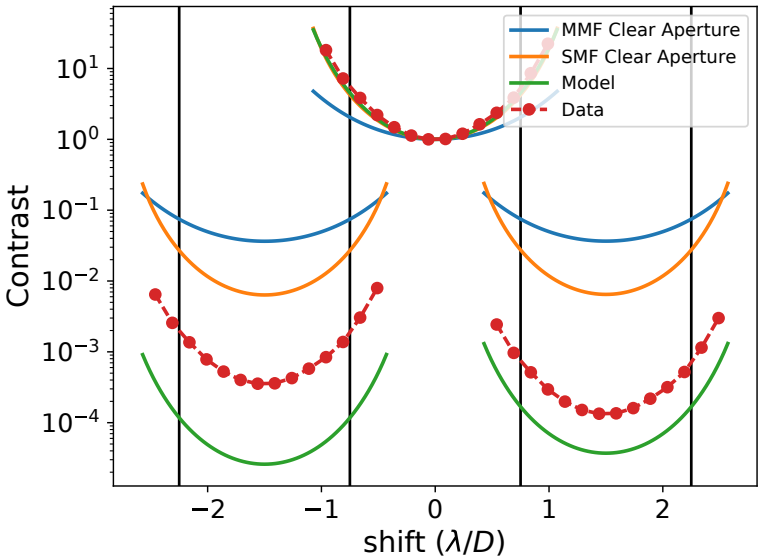


Figure 4.5: Contrast as a function of position on the microlens array. The black lines show the borders of the microlenses. The blue lines show the contrast for a multi-mode fiber. The orange lines show the contrast for a single-mode fiber. A single-mode fiber already provides extra contrast compared to a multi-mode fiber. The green lines show the model of the SCAR coronagraph, and the red dots are the measurements.

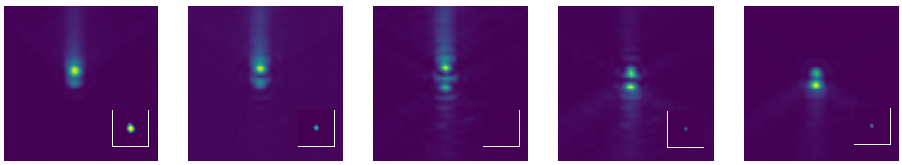


Figure 4.6: Measured microlens spot structure for the 360 SCAR. The inset shows the throughput of the single mode fiber. The deepest null occurs with a triple spot structure, which is a second order null due to the two zero crossings. The white circle shows the position and MFD of the fiber.

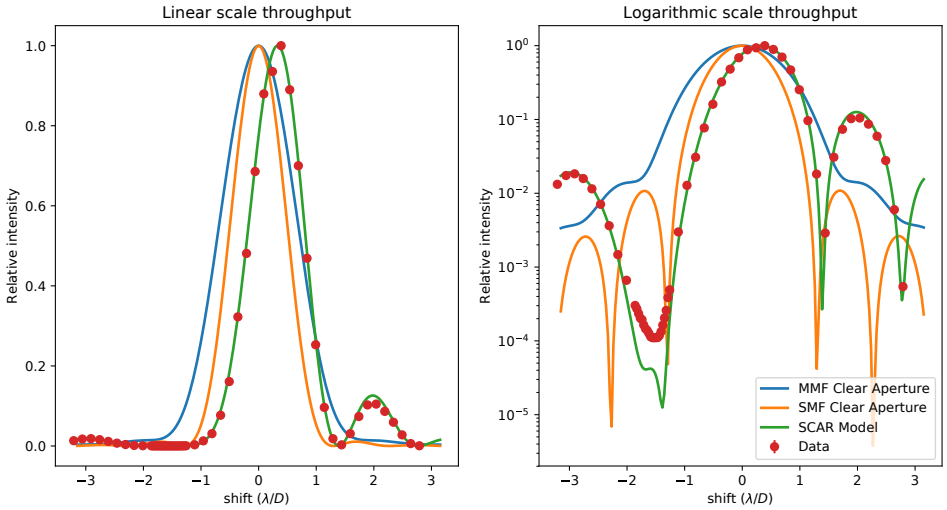


Figure 4.7: Left: Relative throughput on a linear scale as function of relative shift with respect to the microlens center. The green points are the measurements. Errorbars are included but are smaller than the size of the plotting symbols. The errorbars are very small and show that random errors can not explain the difference between the model and the measurements. The orange line shows the nonfitted model of this SCAR design. Right: Throughput on a logarithmic scale. The deepest contrast that we reach is 1.15×10^{-4}. The throughput curve is also flat between -1.6 and -1.4 λ/D. The measured null does not reach the design null due to residual low order aberrations aberrations on the order of 10 nm rms. The blue curve shows the amount of light that falls in a microlens of a clear aperture, which shows the contrast without SCAR. This is comparable to the raw contrast curve for normal imaging or using a multi-mode fiber. The orange line shows the normalized throughput with an unobstructed aperture and a SMF. The SMF shows a gain compared to the MMF.

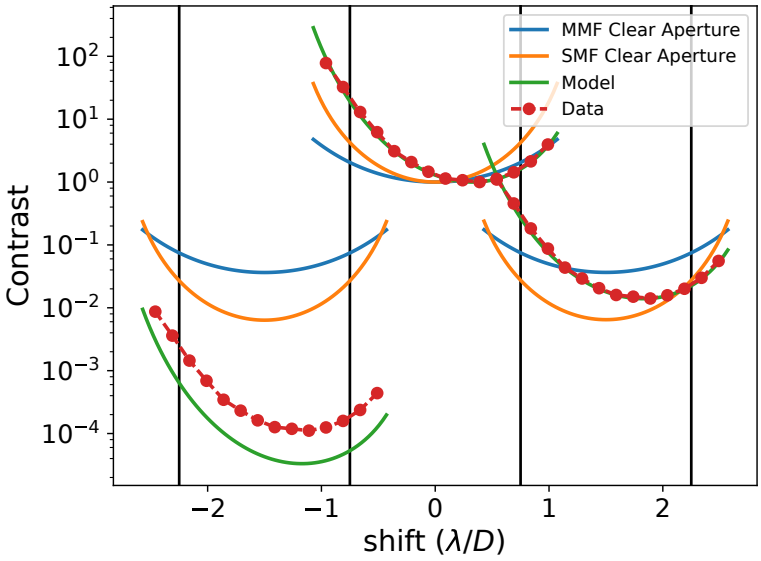


Figure 4.8: Contrast as a function of position on the microlens array. The black lines show the borders of the microlenses. The blue lines show the contrast for a multi-mode fiber. The orange lines show the contrast for a single-mode fiber. A single-mode fiber already provides extra contrast compared to a multi-mode fiber. The green lines show the model of the SCAR coronagraph, and the red dots are the measurements.

theoretical maximum for circular apertures (Shaklan & Roddier, 1988). Our lab measurement of the coupling was defined as the ratio between the fiber spot intensity and the MLA spot intensity. After correcting for the uneven beam-splitter between the two spots our absolute coupling is 76 ± 3 %. This includes Fresnel losses at the interface of the fiber, fiber propagation losses and roughness due to the polishing of the fiber. The measured coupling is very close to the theoretical maximum if we consider these losses. The absolute throughput is the product of the coupling, a correction for the spatial sampling and the Strehl ratio of the phase plate. The measured absolute throughput is 26 ± 1%.

4.3 Tolerance simulation analysis

Lovis et al. (2017) show that within certain assumptions about Proxima b that it can be characterized by combining SPHERE+ with a high-resolution spectrograph (ESPRESSO in this case). Several challenges are to be solved. Current coronagraphs on SPHERE are not able to suppress the stellar diffraction halo at the position of Proxima b with the required contrast, and the AO system is not good enough at the angular separation of Proxima b. Lovis et al. (2017) propose SPHERE+ where both the coronagraph and the AO system are upgraded. Switching from a Shack-Hartmann wavefront sensor to a Pyramid wavefront sensor would greatly improve performance at small inner working angles (Fusco et al., 2006; Vérinaud et al., 2005). For SPHERE+ Lovis et al. assumed a hypothetical coronagraph. The listed requirements of this coronagraph are:

- A contrast of at least 5000.
- A relatively broad wavelength range of at least 20 percent.
- The stellar rejection region should encompass the orbit of Proxima Centauri b, which has a maximum estimated separation of 36 milliarcseconds (2.2 λ/D at 0.7 μm).
- An inner working angle of 1 λ/D to reach the full resolving power of the telescope.
- Either circular or asymmetric dark holes.
- The coronagraph should be able to handle tip-tilt errors within 3 mas.

We designed a phase plate for the SCAR coronagraph that would fit these requirements in Paper I. This SCAR design is able to null a circular area from 0.8 to 2.4 λ/D, which is large enough as Proxima b has a maximal separation of 2.2 λ/D at 0.7 μm. The design bandwidth is 10%. Within this bandwidth the raw contrast is 3×10^{-5}, which is 10 times higher than required. Outside of the design bandwidth it still works well as the contrast stays below 1×10^{-4} up to 20% bandwidth. The bandwidth of the coronagraph is slightly smaller than required. With this design we would be able to reach almost all requirements of the coronagraph for the characterization of Proxima b. The final requirement is necessary as the tip-tilt jitter of SPHERE is 3 mas (Fusco et al., 2016).

It is important to estimate the effects of manufacturing errors which change the reachable contrast. Our target contrast including manufacturing errors is 10^{-4}. There are three major parts that can influence the final performance. The SCAR phase plate is manufactured with the same techniques as the APP. Because the APP has demonstrated on sky that it can achieve a high contrast (Otten et al. (2017)). For the required contrast levels we can safely assume that the manufacturing errors in the phase plate are negligible. The other two aspects are the manufacturing tolerances on the fiber array and the residual wavefront error after the SPHERE AO system. We focussed on the manufacturing tolerances of the fiber injection unit because residual wavefront errors can not be solved by the coronagraph.

4.3.1 Fiber alignment tolerance

To couple well into a single mode fiber it is necessary to have a good alignment of the fiber with respect to the center of the microlens surface. For normal operations of a single mode fiber, which is getting in as much light as possible, the alignment tolerance is already strict. For the fiber coronagraph the alignment tolerance becomes even stricter. In Figure 4.9 the throughput as function of the fiber shift is shown. The figure shows the throughput of the central lenslet (which is the planet coupling) and the throughput on an off-axis lenslet (which is the nulling of the star) as function of fiber offset. The white circle shows the largest fitting circular region where the contrast is still below 10^{-4}. In the same region the on axis relative throughput is above 95 percent. This shows that injecting light into a fiber is easier than using a fiber to cancel the light as the off-axis throughput surface is a much steeper function of misalignment. The white circle has a radius of 1/8 of the mode field diameter of the fiber. The fiber alignment should be within this diameter to reach the required contrast.

Step index single mode fibers are the most used single mode fiber and they have a mode field diameter with a size around 5 μm at a wavelength of 700 nm. Given this size the alignment tolerance would be roughly ± 0.6 μm. This is very strict and most manufacturing procedures have an alignment tolerance of 1 μm. The submicron tolerance can be circumvented with the use of fibers with a larger mode field diameter. Fibers with a large mode field diameter can be made with Photonic Crystal Fibers(PCF). Photonic crystal fibers do not use internal reflection to guide the light as step index fiber do, but use the geometry of the fiber structure. Specific geometric configurations create bandgaps which allows certain optical modes to propagate (Corbett (2006)). Large mode area photonic crystal (LMA) fibers are fibers with a large mode field diameters which can be up to 50 μm (Jansen et al. (2012); Stutzki et al. (2014)) and only allow propagation of the fundamental mode. Another advantage of the PCFs is their endlessly single mode property that allows for a very large wavelength range to be propagated through the fiber, which is convenient for spectrographs. The manufacturing tolerances of 1 μm would require a fiber with a MFD of at least 8 μm. Fibers with a mode field diameter of 12.5 μm are readily available and would fulfill the alignment tolerance requirement. These fibers have a less strict tolerance which is $\approx 1.5\mu$m. With the PCFs the alignment of the fibers should not be an issue, and a manufacturer has been identified.

The LMA fibers have a constant MFD but the numerical aperture (NA) depends on wavelength because of the conservation of etendue. The common step-index fibers have the opposite behaviour: the NA is fairly constant but the MFD changes with wavelength. The microlens launches the light with a constant NA into the fiber, which can lead to a mode-mismatch in the case of LMA fibers. The mode-mismatch reduces the coupling as a function of wavelength and is optimal only for a single wavelength. In Paper I we simulated and designed the phase plates with the LMA fibers in mind. There we saw that the throughput as a function of bandwidth changes very slowly. If the wavelength changes by 50%, the throughput drops from 55% to 40%. Within the design bandwidth of 20% the throughput varies between 50-55%. For our bandwidth this will lead to a slightly lower efficiency compared to step-index fibers.

4.3.2 MLA surface

The microlens array will have surface errors. Therefore it is important to know how much the errors influence the final contrast. The effects of

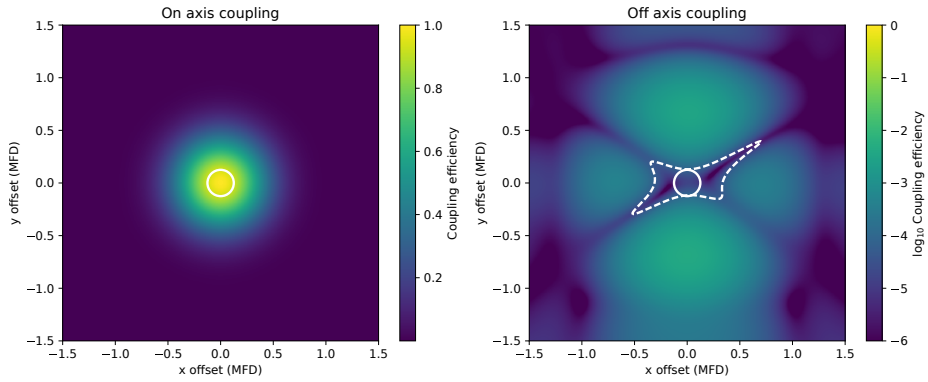


Figure 4.9: Throughput of an on-axis source as function of fiber misalignment is shown on the left. The right shows the off-axis contrast as function of misalignment. The white contour encircles the area where the contrast is below 1×10^{-4}. The white circle is the largest circle that fits within the contour with a radius of 1/8th of the mode field diameter

Zernike wavefront errors on the microlens array are shown in Figure 4.10. For the low order Zernike modes there are two curves per figure which show the contrast for two different fibers. Due to symmetry the other four fibers behave in the same way as one of these two. The most important surface errors are defocus and astigmatism. Both rapidly degrade the contrast. The higher order Zernike modes have almost no influence on the contrast this can be seen in the fourth panel in Figure 4.10. The defocus can be partially compensated by moving the fiber in the axial direction. The astigmatism can not be compensated by changing the alignment of the fibers, so it puts a requirement on the surface deviation of the microlens array. The astigmatism of the microlenses should be smaller than $\lambda/6$ peak to valley to reach the required 10^{-4} contrast.

4.3.3 Fiber mode shape

From our investigations we noticed that the mode profile is not very important. In this analysis we changed the mode radius, eccentricity and the orientation of the resulting multivariate Gaussian mode. The mode field radius can change by $\pm 10\%$ while still create a contrast below 1×10^{-4}. Usual manufacturing constrains on the mode field radius are also within $\pm 10\%$, therefore the mode field radius is not an issue. Eccentricities up to 0.5 and the orientation of the ellipse had no significant impact on the

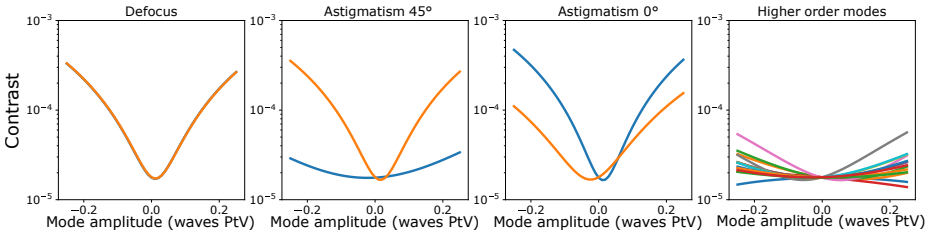


Figure 4.10: Effect of wavefront errors in the microlenses on the contrast due to a certain Zernike mode. The first three panels show the effects on two different fibers. Due to symmetry the other four fibers behave in the same way as these two. The radial symmetry of the defocus mode causes the same effect in all fibers, therefore the curves of the two fibers overlap. In the rightmost panel the effects of Zernike modes 6 to 10 are shown. Focus and astigmatism are the most dominant wavefront errors for lowering the contrast.

throughput and contrast. The average contrast went up from 2.7×10^{-5} to $\approx 3 \times 10^{-5}$ due to these two parameters. From this we conclude that the tolerances on the mode profile of the fiber will be easily met.

4.3.4 FIU Monte Carlo analysis

In the previous section several manufacturing errors have been looked at independently of other errors. A Monte Carlo analysis is performed to estimate the degradation of the coronagraph due to all manufacturing errors. The Monte Carlo analysis generated 3000 realizations of the system within the parameter space given in Table 4.1. The results of this analysis can be seen in Figure 4.11, where the probability density function as function of wavelength and contrast is plotted. The expected performance of the coronagraph plotted on the figure is well under 1×10^{-4}. The three sigma threshold shows that within the manufacturing specifications we will reach the required contrast with very high certainty. The expected stellar nulling within 15% bandwidth is below 3×10^{-5}, within 20% below 1×10^{-4} and within 25% below 2×10^{4}. After correction for throughput variations this meets the required specifications for characterizing Proxima b with SPHERE+.

Table 4.1: The Monte Carlo parameters for the tolerance analysis. The parameter column shows which parameters are varied and the distribution column shows what distribution is assumed for a parameter.

Parameter	Distribution parameter	Distribution
Mode field diameter	±5 percent	Uniform
Fiber misalignment	±1/12 MFD	Uniform
Microlens focus	$\sigma = 0.01$ percent	Gaussian
Microlens astigmatism	$\sigma = \lambda/4$	Gaussian

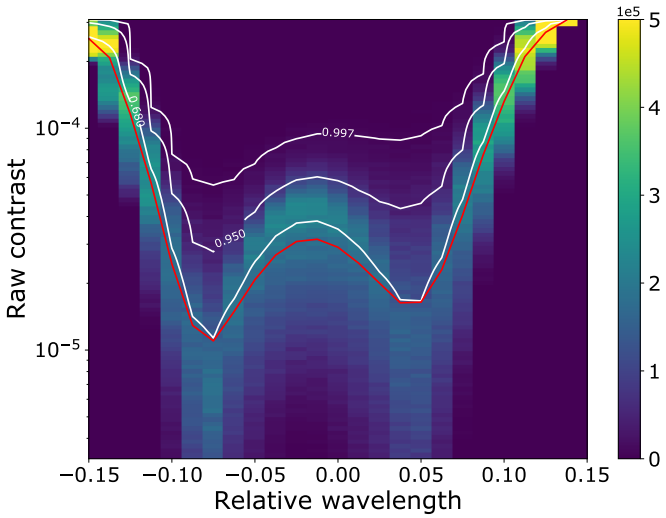


Figure 4.11: Results of the Monte Carlo tolerance analysis as function of wavelength. The red curve shows the expectation value of the contrast. The white curves are the 0.68, 0.95 and 0.997 percentile confidence limits. The analysis indicates that within specified tolerances a contrast below 1×10^{-4} can be reached over a 15 percent bandwidth. Within the full 20 percent bandwidth the contrast is below 2×10^{-4}.

4.4 Conclusions

Combining a pupil plane phase plate with a single mode fiber array creates a new coronagraph that can be used to search for planets very close to their host star. Adding the high-resolution spectroscopy post-processing makes this a very robust system for exoplanet characterization. From the lab measurements and tolerance simulations we can conclude that:

1. We have shown a proof of principle for the SCAR coronagraph in the lab and reached a contrast 1×10^{-4} for the 180 SCAR and $2 - 3 \times 10^{-4}$ for the 360 SCAR. These contrasts limit were due to residual wavefront errors, which requires an active system to remove.

2. The monochromatic tip-tilt stability of the coronagraph has been measured and is estimated to be on the order of $\pm 0.15\lambda/D$ for both the 360 and 180 design. This agrees with the designed tip-tilt stability and falls within the expected jitter of SPHERE.

3. Within expected manufacturing tolerances the coronagraph will be able to meet the requirements with a high degree of confidence (more than 3σ).

4. The most important aspect of the FIU is the alignment of the fiber with respect to the microlens center according to the tolerance simulations. The tolerance requirement can be achieved by using photonic crystal fibers with large mode field diameters.

With the SCAR coronagraph we meet the requirements for the characterization of Proxima b with SPHERE+ and high-resolution spectroscopy. Due to the simplicity of the optical setup only minor modifications are necessary to accommodate the SCAR coronagraph. Furthermore the single mode fibers simplify the design and decrease the size of the high-resolution integral field spectrograph as the input becomes diffraction limited. Therefore adding the new coronagraphic system as an upgrade to existing HCI instruments at current-generation telescopes will allow characterization of exoplanets in the solar neighborhood. An on-sky prototype is being built as the next step in the development process.

References

Aime, C., & Soummer, R. 2004, The Astrophysical Journal Letters, 612, L85

Angel, J. R. P., Cheng, A. Y. S., & Woolf, N. J. 1986, Nature, 322, 341, doi: 10.1038/322341a0

Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, Nature, 536, 437

Bechter, A., Crass, J., Ketterer, R., et al. 2016, in Proc. SPIE, Vol. 9909, Adaptive Optics Systems V, 99092X

Beuzit, J.-L., Feldt, M., Dohlen, K., et al. 2008, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 701418–701418

Brogi, M., Snellen, I. A., de Kok, R. J., et al. 2012, Nature, 486, 502

Codona, J., Kenworthy, M., Hinz, P., Angel, J., & Woolf, N. 2006, in SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 62691N–62691N

Corbett, J. C. W. 2006, New A Rev., 50, 305, doi: 10.1016/j.newar.2006.02.028

Doelman, D. S., Snik, F., Warriner, N. Z., & Escuti, M. J. 2017, in Proc. SPIE, Vol. 10400, Techniques and Instrumentation for Detection of Exoplanets VIII, 10400 – 10400 – 12. http://dx.doi.org/10.1117/12.2273406

Fusco, T., Rousset, G., Sauvage, J.-F., et al. 2006, Opt. Express, 14, 7515, doi: 10.1364/OE.14.007515

Fusco, T., Sauvage, J.-F., Mouillet, D., et al. 2016, in Proc. SPIE, Vol. 9909, Adaptive Optics Systems V, 99090U

Galicher, R., Marois, C., Macintosh, B., et al. 2016, A&A, 594, A63, doi: 10.1051/0004-6361/201527828

Jansen, F., Stutzki, F., Jauregui, C., Limpert, J., & Tünnermann, A. 2012, Opt. Lett., 37, 4546, doi: 10.1364/OL.37.004546

Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, Publications of the Astronomical Society of the Pacific, 127, 890

Jovanovic, N., Schwab, C., Guyon, O., et al. 2017, A&A, 604, A122, doi: 10.1051/0004-6361/201630351

Kawahara, H., Murakami, N., Matsuo, T., & Kotani, T. 2014, The Astrophysical Journal Supplement Series, 212, 27

Konopacky, Q. M., Barman, T. S., Macintosh, B. A., & Marois, C. 2013, Science, 339, 1398, doi: 10.1126/science.1232003

Labadie, L., Le Coarer, E., Maurand, R., et al. 2007, A&A, 471, 355, doi: 10.1051/0004-6361:20067005

Lovis, C., Snellen, I., Mouillet, D., et al. 2017, A&A, 599, A16, doi: 10.1051/0004-6361/201629682

Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proceedings of the National Academy of Sciences, 111, 12661

Martinez, P., Loose, C., Aller Carpentier, E., & Kasper, M. 2012, A&A, 541, A136, doi: 10.1051/0004-6361/201118459

Mawet, D., Ruane, G., Xuan, W., et al. 2017, The Astrophysical Journal, 838, 92

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

Otten, G. P., Snik, F., Kenworthy, M. A., Miskiewicz, M. N., & and, Johanan L. Codona, M. J. E. 2014, in Proc. SPIE, Vol. 9151, 9151 – 9151 – 10. https://doi.org/10.1117/12.2056096

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

Por, E. H., & Haffert, S. Y. 2017, Astronomy & Astrophysics, submitted (not yet accepted)

Riaud, P., & Schneider, J. 2007, A&A, 469, 355, doi: 10.1051/0004-6361:20077085

Shaklan, S., & Roddier, F. 1988, Appl. Opt., 27, 2334, doi: 10.1364/AO.27.002334

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

Snellen, I. A., Brandl, B. R., de Kok, R. J., et al. 2014, Nature, 509, 63

Snik, F., Otten, G., Kenworthy, M., et al. 2012, in Proc. SPIE, Vol. 8450, Modern Technologies in Space- and Ground-based Telescopes and Instrumentation II, 84500M

Sparks, W. B., & Ford, H. C. 2002, The Astrophysical Journal, 578, 543

Stutzki, F., Jansen, F., Otto, H.-J., et al. 2014, Optica, 1, 233, doi: 10.1364/OPTICA.1.000233

Vérinaud, C., Le Louarn, M., Korkiakoski, V., & Carbillet, M. 2005, MNRAS, 357, L26, doi: 10.1111/j.1745-3933.2005.08638.x

Wang, J., Mawet, D., Ruane, G., Hu, R., & Benneke, B. 2017, AJ, 153, 183, doi: 10.3847/1538-3881/aa6474

Chapter 5

High Contrast Imaging for Python (HCIPy)

an open-source adaptive optics and coronagraph simulator

Adapted from
E. H. Por, S. Y. Haffert, V. M. Radhakrishnan, D. S. Doelman, M. A. M van Kooten and S. P. Bos (2018), Proc. SPIE 10703

Abstract
HCIPy is a package written in Python for simulating the interplay between wavefront control and coronagraphic systems. By defining an element which merges values/coefficients with its sampling grid/modal basis into a single object called Field, this minimizes errors in writing the code and makes it clearer to read. HCIPy provides a monochromatic Wavefront and defines a Propagator that acts as the transformation between two wavefronts. In this way a Propagator acts as any physical part of the optical system, be it a piece of free space, a thin complex apodizer or a microlens array.

HCIPy contains Fraunhofer and Fresnel propagators through free space. It includes an implementation of a thin complex apodizer, which can modify the phase and/or amplitude of a wavefront, and forms the basis for more complicated optical elements. Included in HCIPy are wavefront errors (modal, power spectra), complex apertures (VLT, Keck or Subaru pupil), coronagraphs (Lyot, vortex or apodizing phase plate coronagraph), deformable mirrors, wavefront sensors (Shack-Hartmann, Pyramid, Zernike or phase-diversity wavefront sensor) and multi-layer atmospheric models including scintillation).

HCIPy aims to provide an easy-to-use, modular framework for wavefront control and coronagraphy on current and future telescopes, enabling rapid prototyping of the full high-contrast imaging system. Adaptive optics and coronagraphic systems can be easily extended to include more realistic physics. The package includes a complete documentation of all classes and functions, and is available as open-source software.

5.1 Introduction

During the design process of an astronomical instrument, numerical simulations of the optical system play an integral role. All components are tested separately for their functionality, and then often integrated into a complete end-to-end simulation to verify that all subsystems are able to operate seamlessly together. The rapid prototyping provided by computer simulations has sped up the development process of high contrast imaging instruments tremendously. Some examples of simulation packages for adaptive optics (AO) are: YAO (Rigaut, 2002), OCTOPUS (Le Louarn et al., 2006), DASP (Basden et al., 2010), COMPASS (Gratadour et al., 2014) and most recently SOAPY (Reeves, 2016). Some examples of packages that simulate coronagraphs are PROPER (Krist, 2007) and POPPY (Perrin et al., 2012). In each of these packages, either the adaptive optics system or the coronagraph is simulated and and subsystems can only be optimized one at a time.

More recently, people have started looking into optimizing the full system, rather than each element separately. For example, one can optimize the coronagraph to suppress the aberrations that are common for a specific AO system. In this method there is little interaction between coronagraph and the AO system: the AO system can be simulated first, and the resulting data can be used for optimizing the coronagraph. In other cases the interaction can be stronger. For instance, in the case of optimizing the AO system to keep a specific region of interest in the focal plane dark, rather than flattening the wavefront (Radhakrishnan et al., 2018). Another example in this category is end-to-end simulations of post-coronagraphic focal-plane wavefront sensing methods including feedback to the AO system. For examples, see this excellent review (Jovanovic et al., 2018) and references therein. These use cases require both accurate and simultaneous simulation of both the AO system and the downstream coronagraph.

These aforementioned type of simulations place a particular set of requirements on the simulation software. The software needs to give baseline performance, to allow the user to simulate part of the system that he/she is not interested in with minimal effort, while still being modular to allow for inclusion of completely new components. The following guiding principles were used during the design process of HCIPy:

- *Modularity.* Components should be written to work independently of each other. For example, in this way, wavefront sensors can be exchanged, coronagraphs replaced and optical elements be made more

realistic, without having any influence on the operation of the other systems. This allows for rapid prototyping of new components, such as coronagraphs or wavefront sensors, and to compare different components as equally as possible.

- *User friendliness.* Components should be written to have extensive and sensible default parameters. This provides out-of-the-box baseline performance, meaning that even inexperienced users in one of the fields can code a working system.

- *Error avoidance.* Common user errors should be hard to make. That is, HCIPy is designed to handle many mathematical details, such as sampling requirements, automatically and in the background, providing clean and readable code. This allows the user to focus on system architecture rather than the details. HCIPy allows for access to these mathematical details if necessary, but their explicit nature makes mistakes easier to catch.

- *Pythonic.* HCIPy is written in the Python language, an interpreted high-level programming language. This programming language emphasizes code readability, allows for object-oriented programming, and is in use by many astrophysical projects (The Astropy Collaboration et al., 2018).

In Section 5.2 we discuss the core functionality of HCIPy. In Sections 5.3, 5.4, 5.5 and 5.6 we explore the functionality of HCIPy further, discussing optical systems, adaptive optics, coronagraph and more. We conclude with Section 8.5, and look towards future work.

5.2 Core functionality

The main mechanism in HCIPy for following these design principles is the use of `Field`s throughout the code base. These objects behave like a sampled physical field, sampling the value of a physical quantity in space. A `Field` contains a `Grid`, which is used to define the positions of the points in space, on which a `Field` is sampled. A `Coords` object in turn defines the values of the coordinates in a `Grid` object, while not containing any information on how to interpret these coordinates.

While the concepts of `Coords`, `Grid`s and `Field`s might seem esoteric and cumbersome, in most cases the user doesn't have to interact with these classes directly. Instead, the user interacts with functions to create and

modify them, and only uses the classes explicitly when he/she needs more control over the sampling for a specific part of their code. The following subsections present the implementation details of these three objects and explore the flexibility when programming with these objects.

5.2.1 Coords, Grids and Fields

An object of a `Coords` class can yield values for each dimension for each point in a `Grid`. Indexing is done using a single value, rather than one for each dimension. This is done to support `UnstructuredCoords`, which is a set of points with no internal structure. A `SeparatedCoords` object constrains coordinates to be a tensor product of the discretization along each axis. Some mathematical operations, such as Fourier transforms, can be performed much quicker when coordinate axes can be separated. Finally, a `RegularCoords` object constrains regularly-spaced coordinates on all axes. Fourier transforms on such a grid can use a Fast Fourier Transform (FFT) which greatly reduces computation times. Each of these `Coords` classes calculates their values on demand, rather than storing them explicitly. This greatly reduces the amount of required computer memory, especially for larger regularly-spaced grids.

A `Grid` object can be thought of as a discretization of some vector space. It contains a `Coords` class and applies physical meaning to the coordinate values by introducing a coordinate system. Additionally it can provide weights to each of the points: the interval, area, volume or hypervolume that a point covers in its vector space. These weights simplify calculation of integrals and derivatives. HCIPy supports a `CartesianGrid` for Cartesian spaces of any dimension, `PolarGrid` for polar coordinates and `SphericalGrid` for spherical coordinates. Conversion between coordinate systems can be performed by calling the `Grid.as_()` function with the required coordinate system. In addition, `Grid`s can be scaled and shifted by calling their respective functions.

A `Field` object is the discretized version of a physical field. It contains a `Grid` and an array of values. HCIPy supports scalar fields (e.g. an intensity field), vector fields (e.g. a vector electric field) and tensor fields of any dimension (e.g. a Jones matrix field, or a Mueller matrix field). Note that `Field`s in HCIPy do not have to be particle field, but simply a quantity that has a value for each point in space. For a scalar `Field` there is only a single spatial axis in the value array, independent of the dimension of the vector space, as is usually done. The dimensionality is hidden by the grid, rather than the value array. This may seem impractical and unjustified;

nevertheless this is the only way to consistently handle all types of `Coords`, in particular the `UnstructuredGrid`. Moreover, this strengthens the idea that a `Field` can be viewed as a covariant vector. For a vector `Field` the value array is two dimensional, the last axis being the spatial axis, and for tensor `Field`s the value array is n-dimensional for an $(n-1)$-dimensional tensor, again the last axis being the spatial axis. Special functions exist to make handing of vector and tensor fields easier, including but not limited to a `field_dot()` function, which multiplexes a dot product over the spatial axis, a `field_inv`, which takes the inverse of a two-dimensional tensor field for each point separately, and a `field_einsum`, which can calculate all Einstein-summation-convention formulas multiplexed over the spatial axis.

5.2.2 Field generators and visualization

We also use the concept of "field generators", which is a function that accepts a `Grid` as its sole argument and returns a `Field` on that `Grid`. This is in places where an analytical function needs to be evaluated on a `Grid` at another position in the code. A prime example is telescope apertures, which can be evaluated on any `Grid`, and can be user defined.

To make visualization of `Field`s easier, HCIPy includes several functions mimicking, and working in conjunction with, the "Matplotlib" plotting library (Hunter, 2007). The most versatile function is `field_imshow`, which can draw two-dimensional scalar fields. `Field`s with `UnstructuredCoords` are automatically interpolated to be able to draw them. Complex scalar fields are drawn using a custom two-dimensional color scale, which is extremely useful to be able to see both phase and amplitude in electric field plots. Masking of unused parts of the image is supported, and useful for phase plots on a pupil. Additional functions for contour plots for `Field`s are also implemented.

An example script in how to construct and transform `Grid`s, evaluate Field generators, and display `Field`s can be found in Listing 1, along with its output in Figure 5.1.

5.2.3 Fourier transforms

Fourier transform objects act on a `Field`, and return its Fourier transformed version, on the correct `Grid` corresponding to frequency space. HCIPy implements a `NaiveFourierTransform`, which naively implements the full Fourier integral, a `FastFourierTransform`, which uses the FFT algorithm and acts on regularly-spaced grids, and a `MatrixFourierTransform`, which

```python
# Import packages
from hcipy import *
import matplotlib.pyplot as plt
import numpy as np

# Make a separated polar grid
r = np.logspace(-1, 1, 11)
theta = np.linspace(0, 2*np.pi, 11)
coords = SeparatedCoords((r, theta))
polar_grid = PolarGrid(coords)
cartesian_grid = polar_grid.as_('cartesian')
plt.plot(cartesian_grid.x, cartesian_grid.y, '.')
plt.show()

# Create and plot Field
pupil_grid = make_pupil_grid(1024)
aperture = make_magellan_aperture(True)
evaluated_aperture = evaluate_supersampled(aperture, pupil_grid, 8)
imshow_field(evaluated_aperture)
plt.show()
```

Listing 1: An example code showing how to create, manipulate and show fields. A more detailed guide can be found in the online documentation for HCIPy.

uses a discrete-time Fourier transform, also known as a matrix Fourier transform (Soummer et al., 2007), and acts on any separated coordinates, including regularly-spaced grids. These three classes can be used directly or HCIPy can automatically choose which type of Fourier transform is fastest on the required input and output `Grid`s, and use that one.

5.2.4 Mode bases

Similar to field generators, HCIPy implements a wide range of mode bases, among others the Zernike modes, Fourier modes and disk harmonics. A `ModeBasis` is a list of objects, that form a linear mode basis. The modes of a `ModeBasis` can be easily put into a matrix, providing easy access to projection and deprojection of a function onto the mode basis. A code example for constructing several mode bases can be found in Listing 2 along with its output in Figure 5.2.

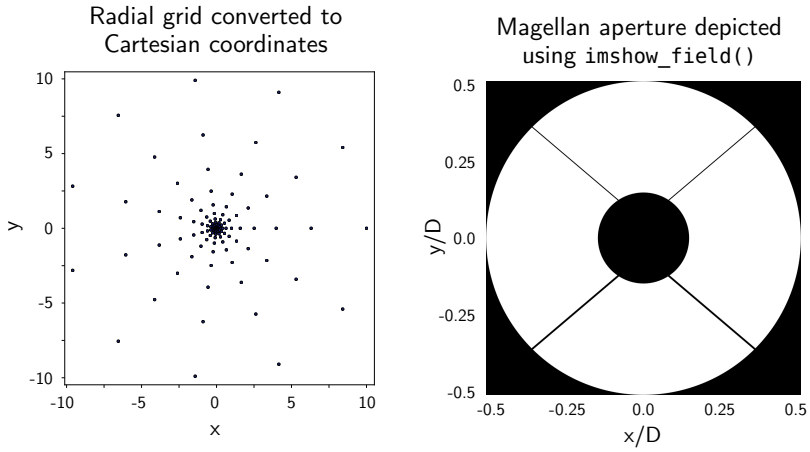


Figure 5.1: The resulting plots from the code in Listing 1. (*left*) The points for a separated grid using polar coordinates. (*right*) The Magellan aperture evaluated using supersampling on a pupil grid.

5.3 Optical systems

`Wavefront`s represent a monochromatic optical wavefront and combine an electric field with a wavelength. To support polarization, the electric field can be both scalar and vectorial, being a two-dimensional Jones vector or a full three-dimensional electric field. `Wavefront`s support many methods to make visualization easier, such as direct access to the phase, amplitude, intensity and power of the electric field.

These `Wavefront`s can be propagated through an optical system using `OpticalElement`s. An `OpticalSystem` can represent any physical optical element, such as `Apodizer`s, `SurfaceAberration`s, `MicrolensArray`s and even `DeformableMirror`s. All `OpticalElement`s support both forward and inverse propagation, as well as calculation of their transformation matrix, ie. the matrix of the linear transformation from input to output planes. The latter is extremely useful for post-coronagraphic wavefront control calculations and coronagraph optimization.

Propagations through space are also `OpticalElement`s, specifically `Propagator`s. Currently implemented are a `FresnelPropagator`, which is a near-field, paraxial propagator using an angular-spectrum method, and a `FraunhoferPropagator`, which is a far-field, paraxial propagator, which can be used to simulate lenses of specific focal lengths.

```
grid = make_pupil_grid(256)
zernike_modes = make_zernike_basis(num_modes=16, D=1, grid=grid)

for i, m in enumerate(zernike_modes):
    plt.subplot(4, 4, i + 1)
    imshow_field(m, cmap='RdBu')
plt.show()

gaussian_hermite_modes = make_gaussian_hermite_modes(grid, num_modes=16,
↪  mode_field_diameter=0.3)
gaussian_laguerre_modes = make_gaussian_laguerre_modes(grid, pmax=5, lmax=2,
↪  mode_field_diameter=0.3)
```

Listing 2: An example code demonstrate the implementation of mode bases in HCIPy. A more detailed guide can be found in the online documentation.

Light can be detected by `Detector`s, which can simulate camera defects and noise. A `NoisyDetector` simulates simple, empirical noise on the detector images, consisting of a flat field, dark current, photon noise and read noise. More complicated noise models are under consideration. An image from a detector can be processed using a `FrameCorrector`. This object performs dark and flat field corrections to any incoming image.

To support changing optics, a `DynamicOpticalSystem` class is available. This class allows for sub-sampled temporal integrations to allow for a fractional number of frames lag in extreme adaptive optics simulations. Updates to the optical system are scheduled and the light is internally propagated through the system. These updates are for example a changing atmosphere, a change in actuator positions for a deformable mirror, or readout of a detector. This class forms the basis for an AO system class.

5.4 Adaptive optics

5.4.1 Atmospheric modeling

To simulate the atmosphere above the telescope, HCIPy uses infinitely-thin phase screens, implemented as `AtmosphericLayer` objects. These phase screens can move over the telescope aperture, according to the "Frozen Flow" approximation (Taylor, 1938). The phase pattern can be generated in two different ways, extendable by the user. Subharmonics (Lane et al., ????) can be added by using a multi-scale Fourier transform method

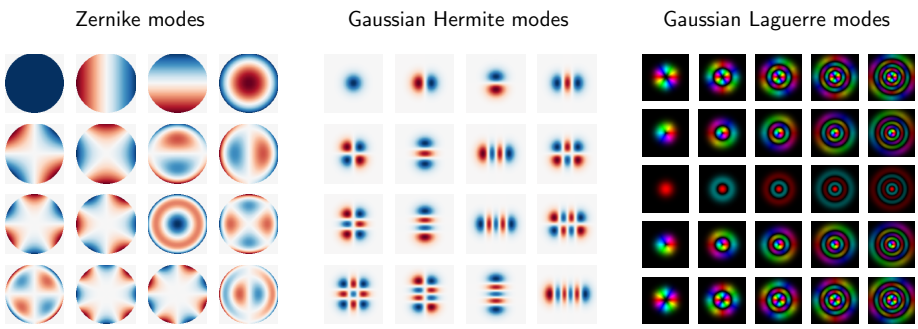


Figure 5.2: Some of the supported mode bases in HCIPy. This is the result of Listing 2.

by `FiniteAtmosphericLayer`. This yields excellent results for short-time simulations, where the phase screen doesn't move that much. For longer simulations, a `InfiniteAtmosphericLayer` is used, which extrudes a long ribbon of phase screen by adding rows sequentially, based on previous samples (Assémat et al., 2006). This method can also simulate non-stationary turbulence.

These single layers can be combined into a `MultiLayerAtmosphere` object. This object simulated the propagation between individual layers using Fresnel propagation, which can simulate scintillation. An example script introducing how to construct and use atmospheric layers and models can be found in Listing 3, along with an example output in Figure 5.3.

5.4.2 Wavefront sensing

Wavefront sensing is done using four operations from different classes. A `WavefrontSensorOptics` class simulates the optics in front of the detector for a certain wavefront sensor. Afterwards, a `Detector` class is used to get the wavefront sensor image. This image can be corrected for flat-field and dark effects using a `FrameCorrector` object. Finally, a `WavefrontSensorEstimator` is used to reduce the corrected detector image into an estimate of the wavefront. This can be either slopes, or actual wavefront, or anything that varies with wavefront. A `WavefrontSensor` object is available for folding all the different steps into a single object. This separation into optics and estimator makes it possible to play back and estimate wavefronts from real wavefront sensor images. An example script on how to construct and use a Pyramid wavefront sensor is presented in

```
# Create an atmospheric layer
layer = InfiniteAtmosphericLayer(pupil_grid, Cn_squared, L0, velocity,
↪  height)

# Propagate a wavefront through the layer
wf = layer(wf)

imshow_field(wf.phase, cmap='RdBu')
plt.show()

# Make a multi-layer atmosphere
layer = make_standard_atmospheric_layers(pupil_grid, L0)
atmos = MultiLayerAtmosphere(layers, scintillation=True)

wf = atmos(wf)
imshow_field(wf.intensity)
plt.show()
```

Listing 3: An example code showing the creation of an atmospheric model including scintillation. A more detailed guide can be found in the online documentation for HCIPy.

Listing 4.

HCIpy supports several wavefront sensors: among others, a Shack-Hartmann, Pyramid (Ragazzoni, 1996; Ragazzoni et al., 2002) (only unmodulated), Zernike (Bloemhof & Wallace, 2003; Zernike, 1935) and (generalized) optical-differentiation wavefront sensor (Haffert, 2016; Sprague & Thompson, 1972). Example detector images are shown for each of these wavefront sensors in Figure 5.4.

5.4.3 Wavefront control

All wavefront controllers are derived from a base `Controller` class. This class can transform measured wavefront sensor estimates from one or more `WavefrontSensorEstimator`s into actuator voltages for one or more `DeformableMirror`s. We will implement an integral controller, which uses a (leaky) integrator, and a full PID controller.

The wavefront sensor estimates can be optionally filtered by an `Observer`, which allows for separation of estimation and control. We will implement a modal reconstructor, which reconstructs modes using a linear transformation matrix, a Kalman filter and a linear minimum mean squared error

```
# Create the optics for the WFS
wfso = PyramidWavefrontSensorOptics(pupil_grid, pupil_separation=1.5,
↪  num_pupil_pixels=64, refractive_index=1.5)

# Create detector model
detector = NoisyDetector(input_grid=wfso.output_grid, flat_field=0.01)

# Create the frame corrector
frame_corrector = BasicFrameCorrector(flat_field=detector.flat_field)

# Create the WFS estimator
wfse = PyramidWavefrontSensorEstimator(aperture=circular_aperture(1),
↪  output_grid=wfso.output_grid)

# Combine all into a single object
wfs = WavefrontSensor(wfso, detector, frame_corrector, wfse)

# Measure an incoming wavefront using a 1sec exposure
wfs.integrate(incoming_wavefront, 1)
slopes = wfs.read_out()
```

Listing 4: An example code showing the setup and reading out of an unmodulated Pyramid wavefront sensor. A more detailed guide can be found in the online documentation for HCIPy.

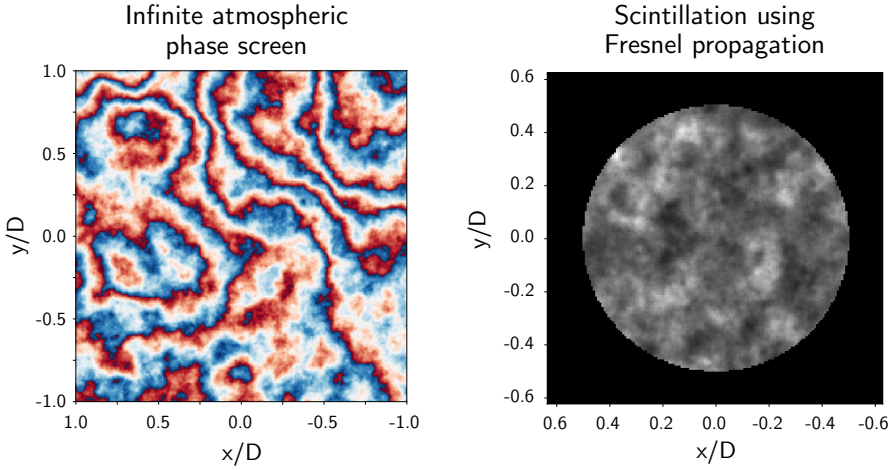


Figure 5.3: The simulated images from Listing 3. These images show a simulated phase screen that can be extruded infinitely in any direction (*left*), and the scintillation from a series of phase screens with Fresnel propagation between layers (*right*).

(LMMSE) estimator. All observers are designed to be used outside of wavefront control as well.

5.5 Coronagraphy

Coronagraphs are implemented as optical elements, most of them taking a pupil-plane `Wavefront` as input and outputting a post-coronagraphic pupil-plane `Wavefront`. A `LyotCoronagraph` and `OccultedLyotCoronagraph` implement both cases of Lyot coronagraphs. The first implements a Lyot coronagraph with a small occulting mask size, and assumes that the part outside of the focal-plane mask is transmitted. The second case assumes that the part outside of the focal-plane mask is occulted. Both of these are in use by, for example, the shaped-pupil Lyot coronagraph. The focal-plane and Lyot-stop masks can be replaced by realistic `OpticalElements` to more accurately describe a specific implementation of a Lyot-style coronagraph.

The `VortexCoronagraph` class implements a propagation through an optical vortex (Foo et al., 2005; Mawet et al., 2005). It uses a multi-scale optical propagation scheme to resolve the vortex singularity. Functions for

generating pupil and Lyot-stop masks for the ring-apodized vortex coronagraph (Mawet et al., 2013) are implemented.

The `ApodizingPhasePlateCoronagraph` implements an APP coronagraph (Codona & Angel, 2004; Kenworthy et al., 2007; Snik et al., 2012). Several functions for optimizing this coronagraph, ie. calculating a phase pattern that creates a dark zone in a region of interest in the focal plane, are implemented. These methods include a quick-and-versatile method based on a modified Gerchberg-Saxton algorithm by Christoph Keller (in prep.), and a globally optimal linear optimization method (Por, 2017). The latter of these requires the installation of Gurobi (Gurobi Optimization, 2016) including its Python interface.

A `ShapedPupilCoronagraph` is also available, including methods for optimizing those, based on global linear optimization (Carlotti et al., 2011). The same function can optimize shaped-pupil Lyot coronagraphs (Zimmerman et al., 2016) as well, for fixed focal-plane and Lyot-stop masks.

A `PerfectCoronagraph` is also implemented. The implementation is based on fitting and subsequent subtraction of the electric field of an unaberrated PSF (Cavarroc et al., 2006). Higher-order perfect coronagraphs are implemented as well using a similar method (Guyon et al., 2006). These coronagraphs are highly computationally efficient and can be used for quick comparisons with other coronagraphs, or if a specific coronagraph model is not needed.

5.6 Miscellaneous

5.6.1 Polarization

Polarization is supported in HCIPy using Jones calculus. This includes the implementation of polarizers and waveplates with (spatially) varying fast-axis orientation, retardance and circularity. This has allowed for research into broadband vector Apodizing Phase Plates for polarimetry and polarization aberrations, see Figure 5.6.

5.6.2 Performance

While computational performance is not the main goal of HCIPy, care has been taken to retain as much performance as possible. Algorithmic improvements are always preferred in cases where they don't compromise on readability and/or generality. This can be seen in the judicious usage of the Matrix Fourier transform, wherever possible. When performing

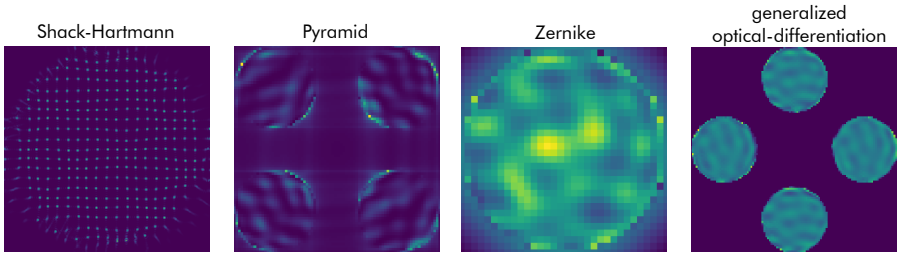


Figure 5.4: Simulated images for different wavefront sensors using HCIPy. All wavefront sensors see the same phase aberration. The amplitude of the aberration was scaled to approximately the linear range of the wavefront sensor, for visualization purposes.


```
# Create the pupil
pupil_grid = make_pupil_grid(2048, 4)
pup = circular_aperture(1)(pupil_grid)

# Create a vortex coronagraph
coro = VortexCoronagraph(pupil_grid, charge=4)

# Create some aberrations
aber = SurfaceAberration(pupil_grid, ptp=1/8, diameter=1)

# Create a Lyot stop mask
lyot = lambda grid: circular_aperture(0.95)(grid) +
    circular_aperture(0.05, [1.8, 0])(grid)
lyot = Apodizer(lyot(pupil_grid))

# Create post-coronagraphic image
img = prop(lyot(coro(aber(Wavefront(pup)))))
```

Listing 5: An example code showing the simulation of an image for a self-coherent camera (SCC; Baudoz et al., 2005) behind a charge 4 vortex coronagraph. A more detailed guide can be found in the online documentation.

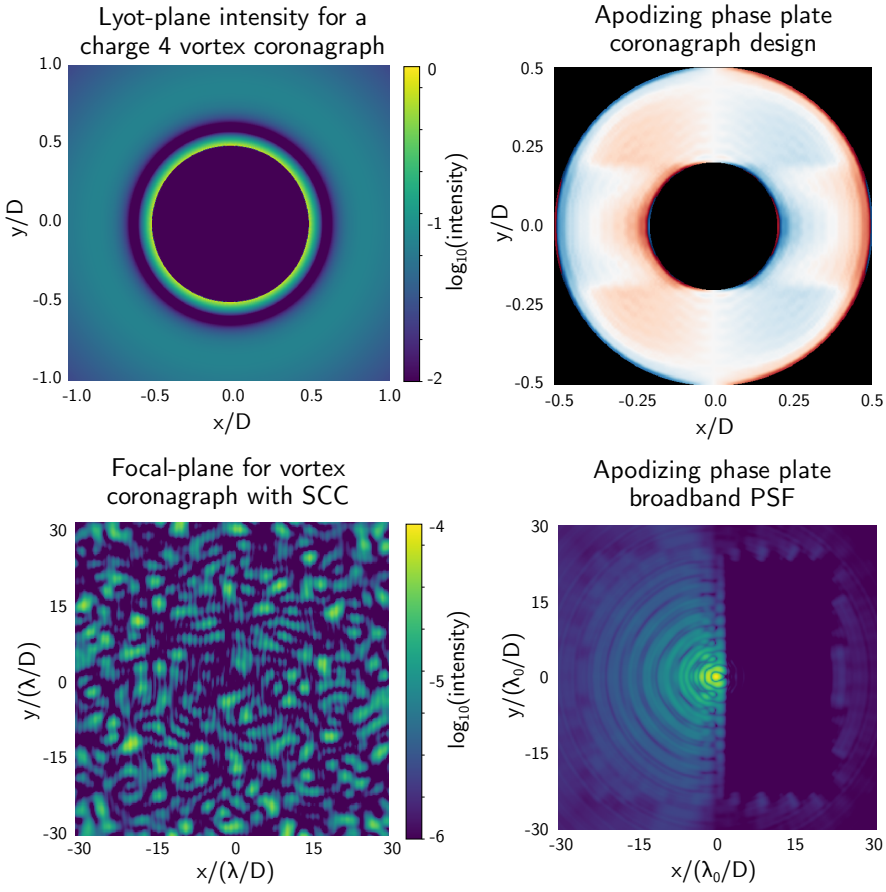


Figure 5.5: A series of simulated coronagraphic images with HCIPy. (*top left*) The Lyot-plane field for a charge 4 vortex coronagraph. (*bottom left*) The focal-plane image of the vortex coronagraph, with a self-coherent camera (Baudoz et al., 2005) Lyot-stop mask. The code for generating this image can be found in Listing 5. (*top right*) An example apodizing phase plate design optimized using HCIPy. (*bottom right*) The broadband point spread function for this APP design.

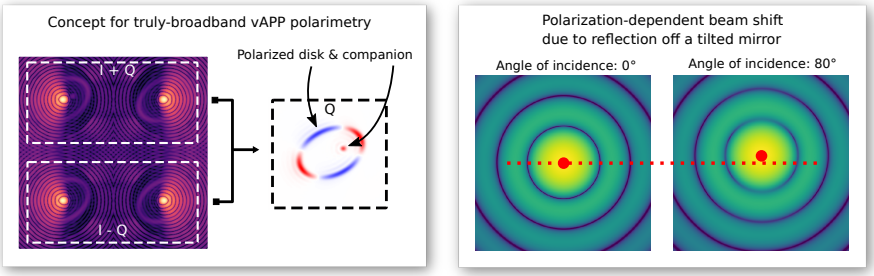


Figure 5.6: Some examples of polarization effects that can be simulated. (*left*) A truly-broadband vAPP implementation for polarimetry. A double difference detection scheme is simulated. This figure is adapted from Bos et al. (2018). (*right*) The effect of polarization due to reflection off a tilted aluminum-coated mirror. This figure is courtesy of Rob van Holstein, work to appear in Van Holstein et al. in prep..

a Fourier transform, HCIPy automatically determines whether a Matrix Fourier transform or a zero-padded Fast Fourier transform would yield the shortest computation time.

Care has been taken with the numerical efficiency as well. The library "NumPy" is used for arrays and simple linear algebra operations (Walt et al., 2011). The library "SciPy" is used for more intricate linear algebra (Jones et al., 2014). The library "PyFFTW" is used to calculate FFTs using the well-known FFTW library (Frigo & Johnson, 2005). HCIPy does not internally use multiprocessing or multithreading yet, apart from the parallelization by linear algebra packages and FFTW. This seemingly strange design decision has its basis in the level at which to parallelize. It is much easier and computationally efficient to parallelize at the highest level possible, as in most cases the workload is embarrassingly parallel. Think for instance at multiplexing different wavelengths for a broadband simulation, or the different parameter sets in a parameter study. It is much harder to efficiently parallelize the Fast Fourier transform. The task of parallelizing is therefore left to the user: he/she now has to write the code for parallelization. GPU support is under development.

5.7 Conclusions

5.7.1 Overview

HCIPy is an object-oriented framework written in Python for performing end-to-end simulations of high-contrast imaging instruments. It is built around the concept of Fields, a sampled version of physical fields. This union allows the user to focus on the high-level structure of their own code, rather than worry about sampling details.

HCIPy defines wavefronts, optical elements and optical systems. Propagators are optical elements that are used for propagation through free space; HCIPy implements both Fraunhofer and Fresnel diffraction propagators. Jones calculus is used for polarization calculations, with polarizers and waveplates supported out of the box. Atmospheric turbulence is simulated using thin infinitely-long phase screens. Scintillation is modeled using Fresnel propagation between individual layers. Implemented wavefront sensors include the Shack-Hartmann and Pyramid wavefront sensors. For coronagraphy the vortex, Lyot and APP coronagraphs are implemented, with methods for globally optimizing pupil-planes, both in phase and amplitude, based on linear optimization.

As the AO and coronagraphy can be used in series in the same software package, HCIPy allows for end-to-end simulations of post-coronagraphic focal-plane wavefront sensing including feedback and implementation of AO control algorithms specifically designed for high-contrast imaging applications. HCIPy is available as open-source software on GitHub, licensed under the MIT license, at https://github.com/ehpor/hcipy.

5.7.2 Future plans

The following subjects are actively considered or in development:

- *GPGPU support.* With the advances in the processing power contained in Graphics Processing Units (GPUs) and the almost universal support for General Purpose computations on GPUs (Nickolls et al., 2008; Stone et al., 2010), support for GPGPU cannot be neglected in any major simulation package. Especially the GPUs affinity for linear algebra operations cannot be ignored. GPGPU support is currently under development for HCIPy. As most heavy numerical calculations are performed on Fields, it makes sense to seamlessly do calculations on Fields in the background on the GPU.

- *Data reduction.* With ground-based observations post-processing techniques play an important role. This cannot be neglected in simulations: we would want to optimize our AO system and coronagraph for post-processed contrast rather than pre-processed contrast. Data-reduction techniques will preferably be included in HCIPy with wrappers to more advanced and mature data-reduction packages such as pyKLIP (Wang et al., 2015) and VIP (Gonzalez et al., 2017).

- *Improved detector modeling.* Noise behavior of detectors have a significant influence on the AO system performance, especially at low flux levels. Improved detector models, such as an (empirical) CCD and CMOS (Konnik & Welsh, 2014), or EMCCD (Hirsch et al., 2013) noise model would be helpful in characterizing these effects.

- *Non-paraxial vector diffraction.* Taking into account the vector nature of light goes further than just supporting polarization-sensitive optical elements: the propagation of light needs to be changed as well, especially in low F-number beams where non-paraxial effects play a much larger role (Bos et al., 2017).

- *More wavefront sensors.* Support for more wavefront sensors is always valued. Currently the coronagraphic Modal Wavefront Sensor (Wilby et al., 2017) and phase diversity (Gonsalves, 1982) are planned.

Community input and extensions are warmly appreciated.

References

Assémat, F., Wilson, R. W., & Gendron, E. 2006, Optics express, 14, 988
Basden, A., Geng, D., Myers, R., & Younger, E. 2010, Applied optics, 49, 6354
Baudoz, P., Boccaletti, A., Baudrand, J., & Rouan, D. 2005, Proceedings of the International Astronomical Union, 1, 553
Bloemhof, E. E., & Wallace, J. K. 2003, in Astronomical Adaptive Optics Systems and Applications, Vol. 5169, International Society for Optics and Photonics, 309–321
Bos, S. P., Doelman, D. S., De Boer, J., et al. 2018, in Proc. SPIE, Vol. 10706, Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III
Bos, S. P., Haffert, S. Y., & Keller, C. U. 2017, in Polarization Science and Remote Sensing VIII, Vol. 10407, International Society for Optics and Photonics, 1040709
Carlotti, A., Vanderbei, R., & Kasdin, N. 2011, Optics Express, 19, 26796
Cavarroc, C., Boccaletti, A., Baudoz, P., Fusco, T., & Rouan, D. 2006, Astronomy & Astrophysics, 447, 397
Codona, J. L., & Angel, R. 2004, The Astrophysical Journal Letters, 604, L117
Foo, G., Palacios, D. M., & Swartzlander, G. A. 2005, Optics letters, 30, 3308
Frigo, M., & Johnson, S. G. 2005, Proceedings of the IEEE, 93, 216
Gonsalves, R. A. 1982, Optical Engineering, 21, 215829
Gonzalez, C. A. G., Wertz, O., Absil, O., et al. 2017, The Astronomical Journal, 154, 7
Gratadour, D., Puech, M., Vérinaud, C., et al. 2014, in Adaptive Optics Systems IV, Vol. 9148, International Society for Optics and Photonics, 91486O
Gurobi Optimization, I. 2016, Gurobi Optimizer Reference Manual. http://www.gurobi.com
Guyon, O., Pluzhnik, E., Kuchner, M., Collins, B., & Ridgway, S. 2006, The Astrophysical Journal Supplement Series, 167, 81
Haffert, S. 2016, Optics express, 24, 18986
Hirsch, M., Wareham, R. J., Martin-Fernandez, M. L., Hobson, M. P., & Rolfe, D. J. 2013, PLoS One, 8, e53671
Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55
Jones, E., Oliphant, T., & Peterson, P. 2014
Jovanovic, N., Absil, O., Baudoz, P., et al. 2018, in Proc. SPIE, Vol. 10703, Adaptive Optics Systems VI
Kenworthy, M. A., Codona, J. L., Hinz, P. M., et al. 2007, The Astrophysical Journal, 660, 762
Konnik, M., & Welsh, J. 2014, arXiv preprint arXiv:1412.4031
Krist, J. E. 2007, in Optical Modeling and Performance Predictions III, Vol. 6675, International Society for Optics and Photonics, 66750P
Lane, R., Glindemann, A., Dainty, J., et al. ????

Le Louarn, M., Vérinaud, C., Korkiakoski, V., Hubin, N., & Marchetti, E. 2006, in Advances in Adaptive Optics II, Vol. 6272, International Society for Optics and Photonics, 627234

Mawet, D., Pueyo, L., Carlotti, A., et al. 2013, The Astrophysical Journal Supplement Series, 209, 7

Mawet, D., Riaud, P., Absil, O., & Surdej, J. 2005, The Astrophysical Journal, 633, 1191

Nickolls, J., Buck, I., Garland, M., & Skadron, K. 2008, in ACM SIGGRAPH 2008 classes, ACM, 16

Perrin, M. D., Soummer, R., Elliott, E. M., Lallo, M. D., & Sivaramakrishnan, A. 2012, in Proc. SPIE, Vol. 8442, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, 84423D

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

Radhakrishnan, V. M., Doelman, N., & Keller, C. U. 2018, in Proc. SPIE, Vol. 10703, Adaptive Optics Systems VI

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

Ragazzoni, R., Diolaiti, E., & Vernet, E. 2002, Optics communications, 208, 51

Reeves, A. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 99097F

Rigaut, F. 2002, YAO Adaptive Optics Simulation Package. `https://github.com/frigaut/yao`

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

Sprague, R. A., & Thompson, B. J. 1972, Applied optics, 11, 1469

Stone, J. E., Gohara, D., & Shi, G. 2010, Computing in science & engineering, 12, 66

Taylor, G. I. 1938, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 476

The Astropy Collaboration, et al. 2018, ArXiv e-prints. `https://arxiv.org/abs/1801.02634`

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science & Engineering, 13, 22

Wang, J. J., Ruffio, J.-B., De Rosa, R. J., et al. 2015, Astrophysics Source Code Library

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

Zernike, F. 1935, Tech. Phys, 16, 454

Zimmerman, N. T., Riggs, A. E., Kasdin, N. J., Carlotti, A., & Vanderbei, R. J. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011012

Chapter 6

Origin of the asymmetry of the wind-driven halo observed in high-contrast images

Adapted from
F. Cantalloube, **E. H. Por**, K. Dohlen, J.-F. Sauvage, A. Vigan, M. Kasper, N. Bharmal, Th. Henning, W. Brandner, J. Milli, C. Correia and T. Fusco (2018), A&A 620, L10

Abstract
The latest generation of high-contrast instruments dedicated to exoplanets and circumstellar disk imaging are equipped with extreme adaptive optics and coronagraphs to reach contrasts of up to 10^{-4} at a few tenths of arcseconds in the near-infrared. The resulting image shows faint features, only revealed with this combination, such as the wind driven halo. The wind driven halo is due to the lag between the adaptive optics correction and the turbulence speed over the telescope pupil. However, we observe an asymmetry of this wind driven halo that was not expected when the instrument was designed. In this letter, we describe and demonstrate the physical origin of this asymmetry and support our explanation by simulating the asymmetry with an end-to-end approach. From this work, we find that the observed asymmetry is explained by the interference between the AO-lag error and scintillation effects, mainly originating from the fast jet stream layer located at about 12 km in altitude. Now identified and interpreted, this effect can be taken into account for further design of high-contrast imaging simulators, next generation or upgrade of high-contrast instruments, predictive control algorithms for adaptive optics, or image post-processing techniques.

6.1 Introduction

With the arrival of the new generation of high-contrast imaging (HCI) instruments equipped with extreme adaptive optics (XAO) and advanced coronagraphs, dedicated to exoplanet and circumstellar disk imaging, we can now visualize optical effects that were expected but never before revealed. On 8m class telescopes, instruments such as VLT/SPHERE (Beuzit et al., 2008), Gemini/GPI (Macintosh et al., 2008), Clay/MagAO-X (Close et al., 2012; Males et al., 2014), and Subaru/SCExAO (Jovanovic et al., 2015) are equipped with XAO, providing a Strehl ratio of up to 95% in the near-infrared, and coronagraphs, providing a raw contrast of up to 10^{-4} at a few hundred milliarcseconds (mas). Images obtained with these instruments show features such as the correction radius of the XAO, the deformable mirror actuator grid print-through, the bright central spot due to diffraction effects in the Lyot coronagraph (Poisson spot or Arago spot), and the wind driven halo due the temporal lag between the application of the XAO correction and the evolving turbulence. All these features were expected and taken into account when designing and simulating the instrument.

However, some unexpected features are also visible within HCI images: the wind driven halo often shows an asymmetry, one wing being brighter and broader than the other, and the point-spread function (PSF) sometimes breaks up, leading to catastrophic loss of performance. While the latter, known as the low wind effect, is described elsewhere (Milli et al., 2018), describing and understanding the asymmetric wind driven halo, which also limits the high-contrast capabilities of the instrument, is the object of this letter.

We first describe qualitatively the observed asymmetry of the wind driven halo (Sect. 6.2). Based on these observations, we propose an explanation and derive its mathematical demonstration (Sect. 6.3). To prove our interpretation, we perform end-to-end simulations taking into account the optical effect that generates the asymmetry and checked that the asymmetry indeed varies as expected with the parameters upon which it depends (Sect. 6.4).

6.2 Description of the observed asymmetry

The wind driven halo (WDH) is the focal plane expression of the AO servolag error (also often referred to as temporal bandwidth error). The AO-

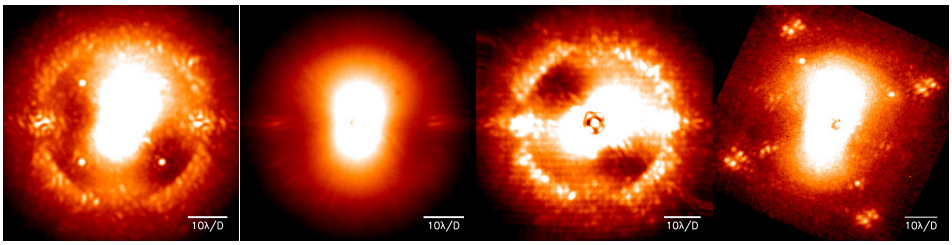


Figure 6.1: Coronagraphic focal plane images showing the asymmetry of the wind driven halo. Left: One exposure obtained with SPHERE-IRDIS (H2 band, 1.593 μm, $\Delta\lambda \approx 53$ nm). Middle left: One exposure obtained with SPHERE-IFS (second channel of YH mode, 0.991 μm, $\Delta\lambda \approx 30$ nm). Middle right: One exposure obtained with SPHERE-IRDIS in broadband (H band, 1.625 μm, $\Delta\lambda \approx 291$ nm). Right: One exposure obtained with GPI (second channel of YH mode, 1.503 μm, $\Delta\lambda \approx 45$ nm). The images are purposely stretched in intensity to highlight the asymmetry (log scale).

lag temporal error appears when the turbulence equivalent velocity above the telescope pupil (defined via the coherence time τ_0, up to a few tens of milliseconds under good conditions) is faster than the adaptive optics correction loop frequency (being about 1.4 kHz for SAXO, the XAO of SPHERE, Petit et al., 2014). Using a coronagraph and a sufficiently long detector integration time (DIT) reveals, in the focal plane, the starlight diffracted by this specific error. As a consequence, the PSF is elongated along the projected wind direction, making a butterfly-shaped halo appear on the images. By definition, this aberration being a phase shift in the pupil plane, it must be symmetric in the focal plane. In practice, however, we observe an asymmetry of the WDH along its axis: one wing being smaller and fainter than the other.

The images obtained with SPHERE and GPI[1] (see Figure 6.1) show the asymmetry of the WDH. To highlight the asymmetry, Fig. 6.2 shows the radial profile along the wind direction and the azimuthal profile at 6 λ/D of the SPHERE-IRDIS image presented in Fig. 6.1 (left).

By definition, the WDH is produced by high wind speed turbulent layers. It has been confirmed that it is mainly triggered by the high-altitude jet stream layer, located in a narrow region of the upper troposphere, at about 12 km above sea level (200 mbar) and with a wind speed from 20 m/s to 50 m/s (Osborn & Sarazin, 2018; Tokovinin et al., 2003). Madurowicz

[1]SPHERE images have been published in respectively Bonnefoy et al. (2018); Samland et al. (2017); Wahhaj et al. (2015), and GPI data in Rameau et al. (2016).

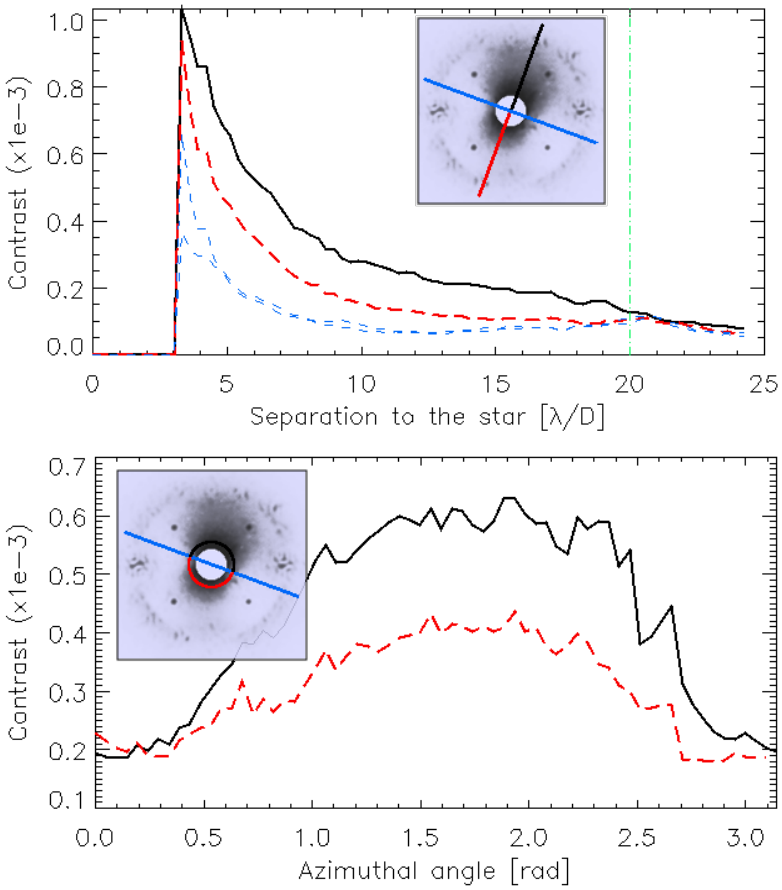


Figure 6.2: Profiles of the wind driven halo showing the asymmetry in a SPHERE-IRDIS image. The solid line is along the brighter and bigger wing; the dashed line is along the fainter and smaller wing. Top: Radial profile along the WDH direction (black solid and red long-dashed lines) and its perpendicular direction (blue dashed lines). The DM cutoff frequency is at 20 λ/D (green dot-dashed line). Bottom: Azimuthal profile at 6 λ/D from the star.

et al. (2018) demonstrated it by correlating the WDH direction with the wind direction at different altitude given by turbulence profiling data for the whole GPIES survey data (Macintosh, 2013). A forthcoming paper, which draws the same conclusion, will similarly analyse the WDH within SPHERE data.

Focal plane asymmetries can only be created by combining phase and amplitude aberrations. As we observed that the asymmetry is pinned to the servolag signature (butterfly shape), we considered that it may be caused by the interaction between servolag errors and amplitude errors created by scintillation, where the phase errors generated by high atmospheric layers propagate into amplitude errors following Fresnel's propagation laws.

6.3 Interference between scintillation and temporal error

In the following we provide an analytical demonstration that the combination of two well-known effects, the AO loop delay (servolag error) and scintillation (amplitude error), which indeed create the asymmetric starlight distribution observed in the high-contrast images.

In the pupil plane, the electric field can be written as

$$E = (1 - \epsilon).e^{i\phi}, \tag{6.1}$$

where ϵ is the amplitude aberration and ϕ the phase aberration. An adaptive optic system measures the phase $\phi(t)$ at a given time t via the wavefront sensor (WFS) and corrects it using a deformable mirror (DM). However, between the analysis of the WFS information taken at an instant t and the command sent to the DM at an instant $t + \Delta t$, if the incoming turbulent phase has varied during Δt, a temporal phase error will remain (the AO servolag error). As a general rule, this absolute time delay Δt varies with both the AO-loop gain and the AO-loop speed and is intrinsic to any AO system. The remaining phase error $\Delta\phi$ can be written as a function of this absolute time delay Δt following (in a closed loop system)

$$\Delta\phi = \phi(t) - \phi(t - \Delta t) \sim \Delta t\, \phi', \tag{6.2}$$

where ϕ' is the time derivative of the phase. This approximation is valid for spatial frequencies affected by the servolag error, that is to say much lower than $1/(v_{wind}.\Delta t)$ under the frozen flow hypothesis (i.e. only the wind speed is responsible for the turbulent phase variation).

Thus, after the AO correction, the electric field becomes

$$\Delta E = (1 - \epsilon).e^{i(\Delta t\, \phi')}, \tag{6.3}$$

which, under the approximation of the small phase and small amplitude errors, simplifies to

$$\Delta E \simeq (1 - \epsilon).(1 + i(\Delta t.\phi')) \sim 1 - \epsilon + i\,(\Delta t\, \phi'). \tag{6.4}$$

Seen through a perfect coronagraph (the patterns exclusively due to diffraction effects of a plane wavefront by the entrance pupil are entirely removed), the post-AO electric field ΔE_c is transformed into

$$\Delta E_c \sim -\epsilon + i\,\Delta t\, \phi'. \tag{6.5}$$

The Earth's turbulent atmosphere is present to different degrees throughout the three dimensions of the atmosphere. Fresnel propagation translates phase variations in the upper atmosphere into amplitude variations via the Talbot effect, creating the so-called scintillation. By the formalism of Zhou & Burge (2010), the phase variations in an atmospheric layer located at altitude z produces an amplitude distribution at the telescope pupil of

$$\epsilon = sin(2\pi \frac{z}{z_T})\, \phi, \tag{6.6}$$

where z_T is the Talbot length, defined as $z_T \doteq 2/(f^2\lambda)$, where f is the spatial frequency and λ the wavelength. The distance from which a pure phase error is fully converted into a pure amplitude error is at one quarter of the Talbot length[2]. For SPHERE, the highest imaging wavelength is 2.2 μm (K band) and the highest corrected spatial frequency is $2.5m^{-1}$, given by the DM inter-actuator spacing (40×40 actuators over the 8 m diameter telescope pupil), yielding a minimum distance of about 36 km altitude, which is above the highest turbulence layers. This explains why, for both GPI and SPHERE, this effect was neglected when designing the instrument.

Adding the scintillation into the coronagraphic post-AO electric field of Eq. (6.5) gives

$$\Delta E_c \sim -sin(2\pi \frac{z}{z_T})\, \phi + i\,\Delta t\, \phi'. \tag{6.7}$$

[2]Under the hypothesis of a monochromatic propagation within infinite pupil extent and small phase approximation.

The resulting intensity observed at the focal plane (I_c) is, within the Fraunhofer framework, the squared modulus of the Fourier transform of the electric field ΔE_c:

$$I_c = |FT[\Delta E_c]|^2 = |-sin(2\pi\frac{z}{z_T})\,FT[\phi] + i\Delta t\,FT[\phi']|^2 \tag{6.8}$$

with $FT[\phi'] = \frac{\partial FT[\phi]}{\partial t} = FT'[\phi]$ being the time derivative of the Fourier transform of the phase. If we assume an arbitrary phase whose general expression can be written $\phi = \exp(i2\pi f.r)$, f being the spatial frequency and r the position, then by making the change of variable $r \leftarrow r + \Delta r$ where we define the beam shift factor $\Delta r = (v_{wind}\,.\,\Delta t)$ to account for the servolag shift (under the frozen flow hypothesis), Eq. (6.8) becomes

$$\begin{aligned} I_c &= |-sin(2\pi\frac{z}{z_T})\,FT[\phi] - 2\pi\,f\,v_{wind}\,\Delta t\,FT[\phi]|^2 \\ &= |FT[\phi]|^2\left(sin(2\pi\frac{z}{z_T}) + 2\pi\,f\,\Delta r\right)^2, \end{aligned} \tag{6.9}$$

where $|FT[\phi]|^2$ is by definition the power spectral density of the turbulent phase and Δr represents the physical spatial shift between the turbulent layer and the AO correction. Developing Eq. (6.9) leads to an asymmetric function of the spatial frequency f: I_c indeed shows an asymmetric distribution of light in the high-contrast images with respect to the centre, originating from interference. Therefore, the intensity of each wing of the WDH can be written, respectively for constructive and destructive interference I_+ and I_-, as follows:

$$I_+ = |FT[\phi]|^2\left(sin(2\pi\frac{z}{z_T}) + 2\pi\,f\,\Delta r\right)^2; \tag{6.10}$$

$$I_- = |FT[\phi]|^2\left(sin(2\pi\frac{z}{z_T}) - 2\pi\,f\,\Delta r\right)^2. \tag{6.11}$$

We thus demonstrate that a temporal phase shift (from temporal delay of the AO loop) between phase error (from the atmospheric turbulence) and amplitude error (from the scintillation effect) creates an asymmetry pinned to the wind driven halo in the focal plane image.

We can define the relative asymmetry factor, $F_{asymmetry}$, as the nor-

malized difference between these two intensities:

$$F_{asymmetry} \doteq \frac{I_+ - I_-}{I_+ + I_-} \tag{6.12}$$

$$= \frac{2\, sin(2\pi \frac{z}{z_T}) f \Delta r}{\left(sin(2\pi \frac{z}{z_T})\right)^2 + (f\, \Delta r)^2}. \tag{6.13}$$

This factor is thus between 0 (no asymmetry) and 1 (all the light is spread in only one wing). [3]

With current HCI instruments, the WDH has a typical contrast of 10^{-4} (see Fig. 6.2), whereas the scintillation has a typical contrast of 10^{-6} (Tatarski, 2016) so we can ignore the scintillation term in the denominator and simplify to $sin(\frac{z}{z_T}) \sim \frac{z}{z_T}$, which yields, after replacing the Talbot length by its expression, the following approximation:

$$F_{asymmetry} = \frac{z f \lambda}{v_{wind}\, \Delta t} + \mathcal{O}\left(\left(\frac{z f \lambda}{v_{wind}\, \Delta t}\right)^2\right). \tag{6.14}$$

We consequently expect the asymmetry factor to grow linearly with the spatial frequency, and therefore with the angular separation to the star. From this demonstration we can already infer a few effects. First, as the interference is taking place between the turbulence residuals and the AO correction lag, any type of coronagraph will reveal the asymmetry of the WDH. Second, even though the Talbot length is 36 km while the jet stream layer is at an altitude of 12 km, the propagation distance is sufficient to convert a small fraction of the phase error into amplitude error and therefore produce the observed asymmetry. Consequently, the higher the altitude of the fast layer, the more asymmetry is produced. On the contrary, the ground layer does not produce this asymmetry. Third, knowing that the amplitude errors are only due to the turbulence, whereas the delayed phase error is due to both the wind speed and the AO loop correction speed, the asymmetry varies with temporal parameters as follows: (i) if the AO loop delay Δt increases (e.g. the AO loop is slower) we lose the correlation between the amplitude errors and the delayed phase errors, making the asymmetry smaller; (ii) if the wind speed v_{wind} is higher, the correlation between the amplitude error and the delayed phase error decreases,

[3] The asymmetry factor is maximum ($F_{asymmetry} = 1$) when the numerator is equal to 1/2 (i.e. $sin(2\pi \frac{z}{z_T}).f\, \Delta r = 1$: the amplitude error is fully correlated with the phase error), and is minimum ($F_{asymmetry} = 0$) when the numerator is null (null wind speed, no temporal lag: there is no wind driven halo) or equal to infinity (there is no correlation at all between amplitude error and phase error).

making the asymmetry smaller. In other word, if the beam shift Δr between the turbulent layer and the AO correction increases, the correlation decreases and so does the asymmetry. Finally, we expect this asymmetry to increase with wavelength as the Talbot length is inversely proportional to wavelength, while the other parameters are independent of wavelength.

As a consequence, an observation site such as Mauna Kea, which suffers less from jet stream compared to observatories located at Paranal in Chile (e.g. Sarazin et al., 2003), would be beneficial to avoid the wind driven halo[4] in the high-contrast images, and the subsequent asymmetry which may arise depending on the AO correction setting and the speed of the high-altitude turbulent layers.

6.4 Simulations of the effect

In the following, we describe a numerical simulation of an idealized AO system reacting to a simplified atmosphere with a single, high-altitude turbulence layer. The goal is to explore the connection between servolag, scintillation, and the occurrence (or absence) of an asymmetric WDH. The simulations are conducted using the HCIPy package (Por et al., 2018), which is available as open-source software on GitHub[5].

We simulated a single atmospheric layer at the altitude of the jet stream, which is then moved across the telescope aperture according to the frozen-flow hypothesis. The light is propagated from the layer to the ground using an angular-spectrum Fresnel propagation code. This light is sensed using a noiseless WFS, which in turn is used to drive a DM. An integral controller with a gain of 0.5 is assumed. The flattened wavefront is then propagated through a perfect coronagraph (Cavarroc et al., 2006) before being focused onto the science camera. We carry out 500 independent short-exposure simulations which are then stacked to form the final long-exposure image. A list of the nominal simulation parameters can be found in Table 8.1.

Figure 6.3 shows the coronagraphic simulated images obtained with or without AO lag and with or without scintillation. As expected, only the combination of both amplitude error and AO servolag error leads to an asymmetric WDH.

[4]The Subaru/SCExAO high-contrast images do not show the wind driven halo and its asymmetry. This might also be explained by the use of predictive control algorithm based on machine-learning techniques, which aims to eliminate the servolag error (Guyon & Males, 2017; Males & Guyon, 2018).

[5]`https://github.com/ehpor/hcipy`

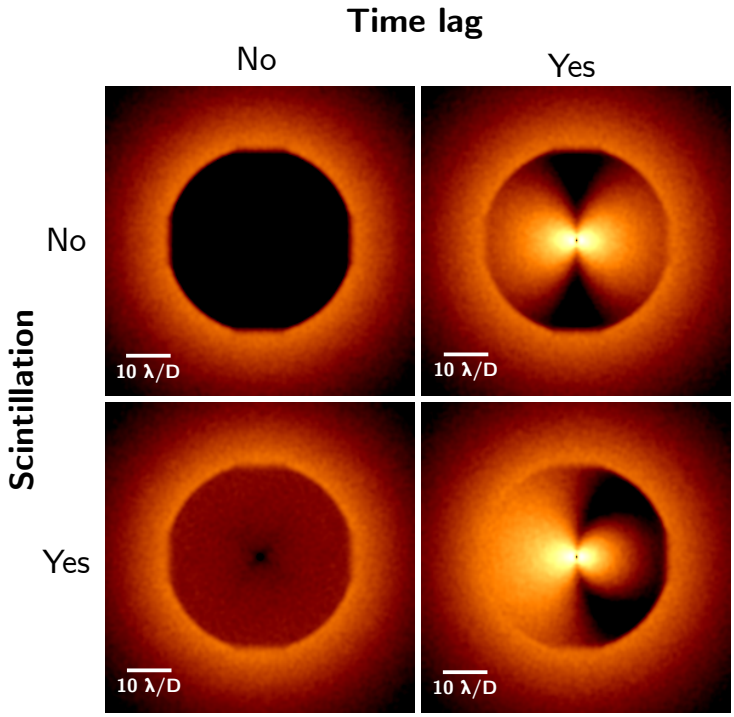


Figure 6.3: Simulated images using HCIPy for the parameters gathered in Table 8.1. The images with no time lag were produced with an infinite AO loop speed. Only a time-lagged WDH and scintillation yields an asymmetric coronagraphic PSF. The images have been stretched in intensity to highlight the asymmetry and scintillation (log scale).

Table 6.1: Nominal set of parameters used for our simulations.

Parameter name	**Value**
Wavelength	2.2 μm (K band)
Pupil diameter	8 m
Seeing	$r_0 = 20$ cm at 500 nm
Outer scale	22 m
Jet stream height	12 km
Jet stream velocity	30 m/s
AO system loop speed	1380 Hz
AO system controller	Integral control
AO system loop gain	0.5 for all modes
Corrected modes	1000 modes
Number of actuators	40×40 rectangular grid
Influence functions	Gaussian with $\sigma = 22$ cm projected
Coronagraph	Perfect (Cavarroc et al., 2006)
Wavefront sensor	Noiseless

Figure 6.4 shows the radial profile of the simulated images along the wind direction (top) and the corresponding asymmetry factor as defined at Eq. (6.12) (bottom), as a function of the separation to the star, where we observe that the asymmetry grows linearly with the separation. We also demonstrate that the scintillation from the jet stream layer at 12 km altitude is enough to create the asymmetry of the wind driven halo and that lower altitude layers create less asymmetry. As expected from the approximation of Eq. (6.14), our simulations also show that the asymmetry is stronger when the wind speed decreases or when the AO loop frequency decreases (for a fixed AO loop gain). We also checked that the asymmetry factor is indeed higher at longer wavelengths.

In a forthcoming paper we will compare this analysis to on-sky images obtained with SPHERE, which involves isolating the contribution of the WDH in the image since other error terms are hiding these trends.

6.5 Conclusions

In this letter we pointed out the presence of an asymmetry of the wind driven halo that is revealed in high-contrast images. We described and demonstrated its origin as being due to interference between AO correction lag (delayed phase error) and scintillation (amplitude errors). We

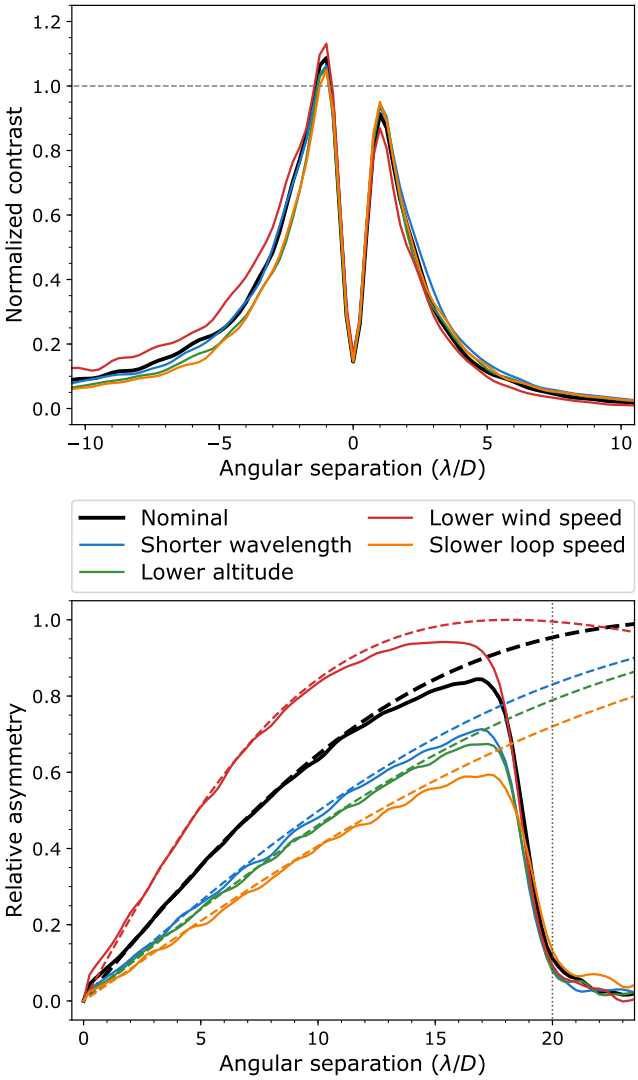


Figure 6.4: Radial profiles along the wind direction (top) and asymmetry factor as defined in Eq. (6.12) (bottom) for the simulated data sets. The solid black line uses the nominal parameters from Table 8.1 (image shown in the bottom right of Fig. 6.3). Other lines differ in one parameter playing a role in the asymmetry: shorter wavelength (H-band, $\lambda = 1.6$ μm, blue line), lower altitude ($z = 8$ km, green line), lower wind speed ($v_{wind} = 20$ m/s, red line), and slower AO loop frequency ($f_{AO} = 800$ Hz, yellow line). The DM cutoff frequency is at 20 λ/D (dotted grey line). In the bottom plot, solid lines indicate the asymmetry from the simulated images and dashed lines show the prediction from Eq. (6.14).

supported our demonstration by simulating this effect using an end-to-end simulator. From these simulations we confirmed the expected behaviour of the asymmetry with different atmospheric turbulence conditions, XAO correction, and imaging wavelength. We further demonstrated that the jet stream layer is the main culprit for this aberration since it is responsible for both servolag error (being a fast layer) and scintillation (being a high-altitude layer). Therefore, an observing site with weak or no jet stream would get around this aberration.

While the current letter focuses on exploring the origin of the wind driven halo asymmetry so as to better understand our current observations and AO systems for future designs, a more quantitative analysis of its implication on high-contrast imaging capabilities and potential mitigation strategies will be detailed in a separate paper. Indeed, the servolag error, when present, is now one of the major effects limiting the high-contrast capabilities of the current instruments (along with the low wind effect, the non-common path aberrations, and residual tip-tilt errors). Knowing that this wind driven halo shows an asymmetry makes it more difficult to deal with in post-processing (as using for instance the residual phase structure functions yields a symmetric phase error or that most filters have a symmetric effect).

Now that this effect has been acknowledged and demonstrated, the next step is to take it into account within end-to-end XAO simulators (e.g. COMPASS or SOAPY, Gratadour et al., 2014; Reeves, 2016) or analytical simulators (e.g. PAOLA, Jolissaint, 2010) and more generally in XAO error budgets, when used in the HCI framework. This study gives insights into the instrument operations, essential to designing optimal post-processing techniques or AO predictive control tools, which both aim to eliminate the servolag error signature (e.g. Correia, 2018; Males & Guyon, 2018). This effect is also important in order to design the next generation of high-contrast instruments (e.g. MagAOX Close et al., 2018, or giant segmented mirror telescopes instruments dedicated to HCI) or to lead the upgrades of existing high-contrast instruments (e.g. GPI or SPHERE, Beuzit et al., 2018; Chilcote et al., 2018), for instance by adding a second DM to correct for the scintillation.

References

Beuzit, J., Mouillet, D., Fusco, T., et al. 2018, A possible VLT-SPHERE XAO upgrade: going faster, going fainter, going deeper (Conference Presentation), doi: 10.1117/12.2313542. https://doi.org/10.1117/12.2313542

Beuzit, J.-L., Feldt, M., Dohlen, K., et al. 2008, in Proc. SPIE, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, 701418

Bonnefoy, M., Perraut, K., Lagrange, A.-M., et al. 2018, ArXiv e-prints. https://arxiv.org/abs/1807.00657

Cavarroc, C., Boccaletti, A., Baudoz, P., Fusco, T., & Rouan, D. 2006, Astronomy & Astrophysics, 447, 397

Chilcote, J., Bailey, V., De Rosa, R., et al. 2018, Upgrading the Gemini planet imager: GPI 2.0, doi: 10.1117/12.2313771. https://doi.org/10.1117/12.2313771

Close, L. M., Males, J. R., Kopon, D. A., et al. 2012, in Adaptive Optics Systems III, Vol. 8447, 84470X

Close, L. M., Males, J. R., Durney, O., et al. 2018, Optical and mechanical design of the extreme AO coronagraphic instrument MagAO-X, doi: 10.1117/12.2312280. https://doi.org/10.1117/12.2312280

Correia, C. M. 2018, Advanced control laws for the new generation of AO systems, doi: 10.1117/12.2313131. https://doi.org/10.1117/12.2313131

Gratadour, D., Puech, M., Vérinaud, C., et al. 2014, in Adaptive Optics Systems IV, Vol. 9148, International Society for Optics and Photonics, 91486O

Guyon, O., & Males, J. 2017, ArXiv e-prints. https://arxiv.org/abs/1707.00570

Jolissaint, L. 2010, Journal of the European Optical Society, 5, 10055, doi: 10.2971/jeos.2010.10055

Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, PASP, 127, 890, doi: 10.1086/682989

Macintosh, B. 2013, The Gemini Planet Imager Exoplanet Survey, NASA OSS Proposal

Macintosh, B. A., Graham, J. R., Palmer, D. W., et al. 2008, in Proc. SPIE, Vol. 7015, Adaptive Optics Systems, 701518

Madurowicz, A., Macintosh, B., Ruffio, J.-B., et al. 2018, Characterization of lemniscate atmospheric aberrations in Gemini Planet Imager data, doi: 10.1117/12.2314267. https://doi.org/10.1117/12.2314267

Males, J. R., & Guyon, O. 2018, Journal of Astronomical Telescopes, Instruments, and Systems, 4, 019001, doi: 10.1117/1.JATIS.4.1.019001

Males, J. R., Close, L. M., Morzinski, K. M., et al. 2014, The Astrophysical Journal, 786, 32

Milli, J., Kasper, M., Bourget, P., et al. 2018, ArXiv e-prints. https://arxiv.org/abs/1806.05370

Osborn, J., & Sarazin, M. 2018, Monthly Notices of the Royal Astronomical Society, 480, 1278, doi: 10.1093/mnras/sty1898

Petit, C., Sauvage, J.-F., Fusco, T., et al. 2014, in Proc. SPIE, Vol. 9148, Adaptive Optics Systems IV, 91480O

Por, E. H., Haffert, S. Y., Radhakrishnan, V. M., et al. 2018, in Proc. SPIE, Vol. 10703, Adaptive Optics Systems VI. https://doi.org/10.1117/12.2314407

Rameau, J., Nielsen, E. L., De Rosa, R. J., et al. 2016, ApJ, 822, L29, doi: 10.3847/2041-8205/822/2/L29

Reeves, A. 2016, in Adaptive Optics Systems V, Vol. 9909, International Society for Optics and Photonics, 99097F

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

Sarazin, M. S., Graham, E., Beniston, M., & Riemer, M. 2003, in Future Giant Telescopes, Vol. 4840, International Society for Optics and Photonics, 291–299

Tatarski, V. I. 2016, Wave propagation in a turbulent medium (Courier Dover Publications)

Tokovinin, A., Baumont, S., & Vasquez, J. 2003, Monthly Notices of the Royal Astronomical Society, 340, 52, doi: 10.1046/j.1365-8711.2003.06231.x

Wahhaj, Z., Cieza, L. A., Mawet, D., et al. 2015, A&A, 581, A24, doi: 10.1051/0004-6361/201525837

Zhou, P., & Burge, J. H. 2010, Applied optics, 49, 5351

Chapter 7

Phase-apodized-pupil Lyot coronagraphs for arbitrary telescope pupils

Adapted from
E. H. Por (2020), ApJ 888, 127

Abstract
The phase-apodized-pupil Lyot coronagraph (PAPLC) is a pairing of the apodized-pupil Lyot coronagraph (APLC) and the apodizing phase plate (APP) coronagraph. We describe a numerical optimization method to obtain globally-optimal solutions for the phase apodizers for arbitrary telescope pupils, based on the linear map between complex-amplitude transmission of the apodizer and the electric field in the post-coronagraphic focal plane. PAPLCs with annular focal-plane masks and point-symmetric dark zones perform analogous to their corresponding APLCs. However with a knife-edge focal-plane mask and one-sided dark zones, the PAPLC yields inner working angles as close as $1.4\lambda/D$ at contrasts of 10^{-10} and maximum post-coronagraphic throughput of $> 75\%$ for telescope apertures with central obscurations of up to 30%. We present knife-edge PAPLC designs optimized for the VLT/SPHERE instrument and the LUVOIR-A aperture. These designs show that the knife-edge PAPLC retains its performance, even for realistic telescope pupils with struts, segments and non-circular outer edges.

7.1 Introduction

In the last few decades, we have seen tremendous advances in the field of exoplanets. Initiated by the discovery of the first planet orbiting another main-sequence star by Mayor & Queloz (1995), we now know that most stars harbor a companion in the habitable zone (Borucki et al., 2011). The majority of planets are detected using indirect methods, such as radial velocity (Mayor & Queloz, 1995) and transits (Charbonneau et al., 2000; Henry et al., 2000). For the brightest stars with transiting planets, spectral characterisation is possible during the transit itself. Longer period planetary transits require precise ephemerides and are limited by the decreasing frequency of observed transits. Direct imaging of these planetary systems provides a way for the detection and characterization of the atmospheres, including variability induced by the rotational modulation of cloud and weather systems and the discovery of liquid water surfaces through glints off liquid surface detectable with polarization.

With the advent of extreme adaptive optics systems, such as VLT/SPHERE (Beuzit et al., 2019), Gemini/GPI (Macintosh et al., 2008), Clay/MagAO-X (Close et al., 2012; Males et al., 2014), and Subaru/SCExAO (Jovanovic et al., 2015), and dedicated space-based instrumentation, such as WFIRST/CGI (Spergel et al., 2013) and HabEx (Mennesson et al., 2016), spatially-resolved imaging of exoplanets has started to become a reality. An optical system known as a coronagraph filters out the light from the on-axis star, while letting through the light from off-axis sources, such as that from faint companions or debris disks. This permits analysis of the off-axis light directly, without being overwhelmed by the on-axis star, and therefore easier chemical characterization of the material orbiting the star. Coronagraphs are both currently used and planned for both future and current space- and ground-based systems.

Many families of coronagraphs have been developed over the years. Among the simplest are the pupil-plane coronagraphs. These coronagraphs apodize the light only in a single pupil plane. The pattern of apodization is designed in such a way as to generate a dark region in the focal plane. Note that, as both on- and off-axis light is apodized in the same way, the apodization pattern must be as minor as possible as to not block too much of the light from the companion or disk. Generally during the design process of such a coronagraph, the throughput is maximized while simultaneously constraining the stellar intensity in the dark zone. Pupil-plane coronagraphs can be separated into two types:

- *Shaped pupil coronagraphs (SPC).* These coronagraphs apodize the pupil with a binary amplitude pattern. Amplitude apodization initially started off as grey-scale (Slepian, 1965), but has since changed to binary (Kasdin et al., 2003), as Carlotti et al. (2011) showed that convex optimization of a gray-scale apodizer yields a globally-optimal binary amplitude mask. SPCs can only create dark zones with point symmetry: as the Fourier transform of a real function is Hermitian, any amplitude-apodized pupil, either binary or gray-scale, inherently has a point-symmetric point spread function (PSF).
- *Apodizing phase plate coronagraphs (APP).* These coronagraphs apodize the pupil with a phase-only mask (Codona et al., 2006; Otten et al., 2017; Snik et al., 2012). Early designs used Fourier iteration techniques (Codona et al., 2006) to find a valid phase pattern. Currently globally-optimal phase patterns can be found using direct convex optimization (Por, 2017). APPs can create dark zones with or without point symmetry.

While it may seem that combining both phase and amplitude apodizing in a pupil-plane coronagraph might yield coronagraphs with higher throughput than either SPCs and APPs, this is not the case. Por (2017) shows that global optimization of a complex-amplitude pupil-plane apodizer will always yield a phase-only apodizer. A corollary is that an APP coronagraph will always outperform a SPC, barring implementation details, as the solution space for SPCs is a subset of the solution space for pupil-plane coronagraphs with a complex-amplitude apodizer. That is, for a fixed telescope pupil shape, dark zone geometry and contrast requirement, the optimal APP will have the same or a higher throughput compared to the optimal SPC. In practice however, for point-symmetric dark zones the gain in throughput is usually minimal, except when the design requirements are so demanding that the throughput is already low for both the SPC and APP coronagraphs (Por, 2017).

The sheer simplicity of the optical layout of pupil-plane coronagraphs has led to their widespread use in high-contrast imaging instruments (Currie et al., 2018; Doelman et al., 2017; Otten et al., 2017). However this simple optical layout also implies worse performance compared to coronagraphs with a more complicated optical layout, due to their more limited design freedom. Because of this, the SPC is often combined with a Lyot stage downstream of the apodizer (Soummer, 2005; Zimmerman et al., 2016). A Lyot stage consists of a focal-plane mask, which apodizes part of the point spread function, and a pupil-plane mask, called a Lyot-stop mask, that

further filters out the residual stellar light. An SPC combined with a Lyot stage is called an Apodized Pupil Lyot Coronagraph (APLC). The added Lyot stage has the effect of reducing the inner working angle and allowing deeper design contrasts. The APLC is able to achieve space-based contrasts at reasonable inner working angles and throughput, making it a baseline coronagraph to which other, more complicated coronagraph designs are compared (Eldorado Riggs et al., 2017; Pueyo et al., 2017).

The success of the APLC leads us to the question: what is the performance of a phase-apodized-pupil Lyot coronagraph (PAPLC)? In Section 7.2 we will outline the numerical optimization method for designing a PAPLC. We will distinguish two types of PAPLCs: one with an annular focal-plane mask and point-symmetric dark zones, and one with a knife-edge focal-plane mask and one-sided dark zones. We will perform a study for the parameter space for simplified telescope pupils for each type in Section 7.3 and 7.4 respectively. To demonstrate the PAPLC for realistic telescope pupils we show designs for the VLT/SPHERE instrument and LUVOIR-A telescope in Section 7.5. Finally, we will conclude with Section 8.5.

7.2 Overview of the numerical optimization problem

In this section we will outline the optimization procedure for PAPLCs. This procedure is based on convex optimization and modifies that of Por (2017), where convex optimization is used for optimizing APPs. We will start by formally defining the optimization problem. Then we will convexify this problem to make global optimization more efficient. Furthermore, we will study how symmetries can be included in the optimization and how these affect the optimal phase pattern. Finally, we discuss how to constrain the tip-tilt of the apodizer in a way that keeps the optimization problem convex.

7.2.1 Problem definition

The optical layout of the PAPLC is shown schematically in Figure 7.1. While joint optimization of the focal-plane mask and Lyot stop is in principle possible, we will restrict ourselves in this study to parameterized focal-plane masks and Lyot stops only. Their parameters will be viewed as hyperparameters on the optimization problem for finding the optimal apodizer. In this study, the number of hyperparameters is limited, and brute-force

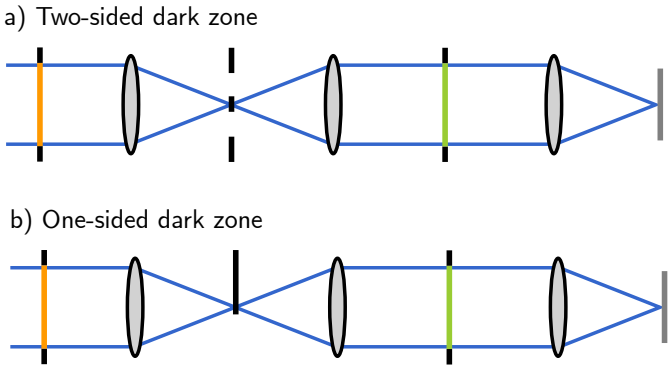


Figure 7.1: The optical layout of the PAPLC with *a)* point-symmetric dark zones, and *b)* one-sided dark zones follows a standard Lyot-style optical setup. The focal-plane mask for point-symmetric dark zones is annular, while it is a knife edge for the one-sided dark zone. In this study we optimize the pre-apodizer (in orange), viewing the parameters of the focal-plane mask and the Lyot stop (in green) as hyperparameters.

optimization is used to optimize them at an acceptable performance cost. More advanced black-box global optimizers, such as Bayesian optimization approaches (Kushner, 1964; Snoek et al., 2012) or Monte-Carlo techniques (Fogarty et al., 2018), can be used if more hyperparameters are required.

Additionally, while many types of focal-plane mask designs are possible, we restrict ourselves in this study to either annular focal-plane masks for point-symmetric dark zones, or an offset knife-edge focal-plane mask for one-sided dark zones. For our parameter studies in Sections 7.3 and 7.4 we will use simplified apertures. There we will use a circularly-obscured telescope pupil and an annular Lyot stop. Furthermore, we will solely use annular dark zones for the point-symmetric dark zones, and D-shaped dark zones as one-sided dark zones. All parameters for the telescope pupil, focal-plane mask, Lyot stop and dark zone geometry are shown schematically in Figure 7.2.

We will use aperture photometry as the main metric for coronagraph performance, and follow Ruane et al. (2018) for our definitions. Here we give a short summary of these definitions for completeness.

We define η_0 as the encircled energy within a circle with a radius of $0.7\lambda/D$ of a normalized PSF generated by the optical system without any coronagraphic masks, so with no apodizer mask, focal-plane mask or Lyot stop mask. This PSF is normalized such that the total power equals one.

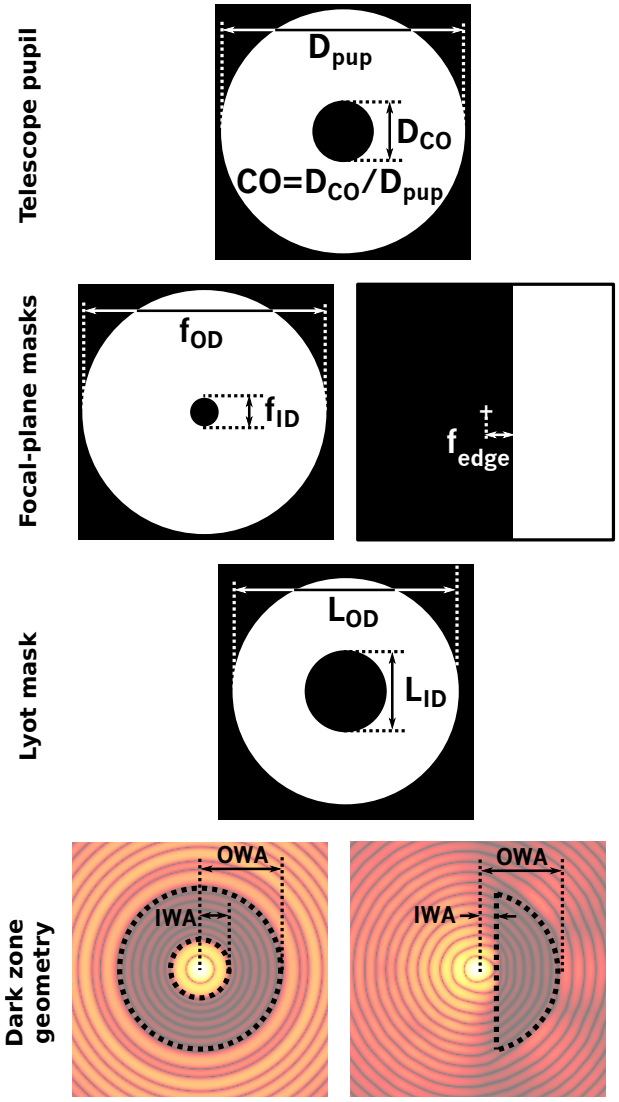


Figure 7.2: The definition of all masks used in this work. These masks are used for the parameter study in Sections 7.3 and 7.4. Centered masks are used for both point-symmetric and one-sided dark zones. The left-justified masks are for two-sided dark zones, while the right-justified masks are used for one-sided dark zones. In general though, arbitrary telescope pupils, Lyot masks, focal-plane masks and dark zone geometries can be used with a PAPLC.

We define $\eta_p(\mathbf{k}, \lambda)$ as the encircled energy within a circle with a radius of $0.7\lambda/D$ centered around $\mathbf{k}$, of the planetary, off-axis PSF, where the planet is located at $\mathbf{k}$, through the coronagraphic optical system. We define $\eta_s(\mathbf{k}, \lambda)$ as the encircled energy within a circle with a radius of $0.7\lambda/D$ centered around $\mathbf{k}$, of the stellar, on-axis image through the coronagraphic optical system. We can now define the throughput $T(\mathbf{k}, \lambda)$ as the ratio between encircled energies of the non-coronagraphic PSF and the off-axis coronagraphic PSF:

$$T(\mathbf{k}, \lambda) = \eta_p(\mathbf{k}, \lambda)/\eta_0. \tag{7.1}$$

The raw contrast $C(\mathbf{k}, \lambda)$ is defined as the ratio between stellar and planetary encircled energies:

$$C(\mathbf{k}, \lambda) = \eta_s(\mathbf{k}, \lambda)/\eta_p(\mathbf{k}, \lambda). \tag{7.2}$$

The design raw contrast C_{design} is defined as the maximum raw contrast in the dark zone D over the whole spectral band:

$$C_{\text{design}} = \max_{\mathbf{k} \in D, \lambda \in [\lambda_-, \lambda_+]} C(\mathbf{k}), \tag{7.3}$$

where λ_-, λ_+ are the minimum and maximum wavelength in the spectral band. Finally we define the inner working angle *IWA* as the smallest angular separation for which the throughput is larger than half of its maximum value for the whole spectral band:

$$IWA = \min_{\{\mathbf{k}: T(\mathbf{k}, \lambda) > \frac{1}{2} \max_{\mathbf{k}} . T(\mathbf{k}, \lambda)\}, \lambda \in [\lambda_-, \lambda_+]} |\mathbf{k}| \tag{7.4}$$

We can now define the optimization problem for the PAPLC. We try to maximize the throughput of the planet while simultaneously constraining the raw contrast in the dark zone. The phase pattern $\phi(\mathbf{x})$ can vary across the aperture. As the throughput $T(\mathbf{k}, \lambda)$ varies across the field of view and as function of wavelength across the spectral band, we take the maximum attained throughput at the center wavelength λ_0 as a measure for the overall throughput. The optimization problem is given by:

$$\underset{\phi(\mathbf{x})}{\text{maximize}} \quad \max_{\mathbf{k}} T(\mathbf{k}, \lambda_0) \tag{7.5a}$$

$$\text{subject to} \quad \eta_s(\mathbf{k}, \lambda) < \eta_p(\mathbf{k}, \lambda) \cdot 10^{-c(\mathbf{k})} \quad \forall\, \mathbf{k} \in D \;\forall\; \lambda \in [\lambda_-, \lambda_+], \tag{7.5b}$$

where $10^{-c(\mathbf{k})}$ is the design contrast in the dark zone, $\mathbf{x}$ is a position in the pre-apodizer, $\mathbf{k}$ is a position in the post-coronagraphic focal plane, D is the dark zone, λ is the wavelength of the light, and $[\lambda_-, \lambda_+]$ is the spectral bandwidth for which we want to optimize.

7.2.2 Simplification and convexification

This optimization problem is non-convex. This means that there could be many local optima and ensuring that the found solution is globally optimal requires a full search of the parameter space. We often prefer convex optimization problems, as they only permit only a single local optimum (which is then also globally optimal). This makes solving convex optimization problems much easier than non-convex problems. In order to convexify our non-convex optimization problem, we need to simplify it quite a bit.

We will discard the aperture photometry methodology in the optimization procedure. This will help us to convexify the objective function later on and will simplify the notation. We will still evaluate all designs using aperture photometry. This yields for the optimization problem:

$$\begin{aligned} \underset{\phi(\mathbf{x})}{\text{maximize}} \quad & |E_{\text{noncoro},\lambda_0}(\mathbf{0})|^2 && \text{(7.6a)} \\ \text{subject to} \quad & |E_{\text{coro},\lambda}(\mathbf{k})|^2 < 10^{-c(\mathbf{k})}|E_{\text{noncoro},\lambda}(\mathbf{k})|^2 && \text{(7.6b)} \\ & \qquad \forall\ \mathbf{k} \in D\ \forall\ \lambda \in [\lambda_-, \lambda_+], \end{aligned}$$

where $E_{\text{coro},\lambda}(\mathbf{k})$ is the on-axis PSF at wavelength λ and $E_{\text{noncoro},\lambda}(\mathbf{k})$ is the on-axis PSF without the focal-plane mask but with the apodizer and Lyot stop mask in the optical system:

$$E_{\text{coro},\lambda}(\mathbf{k}) = \mathcal{P}_\lambda\{L(\mathbf{x})\mathcal{P}_\lambda^{-1}\{M(\mathbf{k})\mathcal{P}_\lambda\{E_{\text{pup}}(\mathbf{x})\}\}\}, \tag{7.7a}$$

$$E_{\text{noncoro},\lambda}(\mathbf{k}) = \mathcal{P}_\lambda\{L(\mathbf{x})E_{\text{pup}}(\mathbf{x})\}, \tag{7.7b}$$

$$E_{\text{pup}}(\mathbf{x}) = A(\mathbf{x}) \exp i\phi(\mathbf{x}), \tag{7.7c}$$

where $A(\mathbf{x})$ is the telescope pupil, $M(\mathbf{k})$ is the focal-plane mask, $L(\mathbf{x})$ is the Lyot stop, $\mathcal{P}_\lambda\{\cdot\}$ is the propagation operator that propagates an electric field from a pupil plane to a focal plane given a wavelength of λ and $\mathcal{P}_\lambda^{-1}\{\cdot\}$ is the inverse of this operator, propagating an electric field from a focal plane to a pupil plane.

This simplification makes the optimization more tractable, but not yet convex. We change the complex phase exponential $\exp i\phi(\mathbf{x})$ into the complex amplitude $X(\mathbf{x}) + iY(\mathbf{x})$, so that

$$E_{\text{pup}}(\mathbf{x}) = A(\mathbf{x})(X(\mathbf{x}) + iY(\mathbf{y})), \tag{7.8}$$

and add the phase-only constraint

$$X^2(\mathbf{x}) + Y^2(\mathbf{x}) = 1 \tag{7.9}$$

to the optimization problem. This additional constraint requires the amplitude of the now complex-amplitude apodizer transmission to be one.

Furthermore we can remove the piston symmetry from the optimization problem: the problem is invariant under the transformation $S : \phi(\mathbf{x}) \rightarrow \phi(\mathbf{x}) + \alpha$, where α is any arbitrary constant. So when we have found a solution $\hat{\phi}(\mathbf{x})$, we know that $S\hat{\phi}(\mathbf{x}) = \hat{\phi}(\mathbf{x}) + \alpha$ is also a solution of the problem. This means that the solution to the problem is non-unique and the problem therefore non-convex. We remove this symmetry by maximizing the real part of the non-coronagraphic electric field, rather than its absolute value. The choice of maximizing the real part, instead of any other linear combination of real and imaginary part is arbitrary. The removal of this symmetry alone does not guarantee a unique solution in general; it only removes a source of non-convexity from the problem. The optimization problem now reads:

$$\underset{X(\mathbf{x}),Y(\mathbf{x})}{\text{maximize}} \quad \Re\left\{E_{\text{noncoro},\lambda_0}(\mathbf{0})\right\} \tag{7.10a}$$

$$\text{subject to} \quad |E_{\text{coro},\lambda}(\mathbf{k})|^2 < 10^{-c(\mathbf{k})}|E_{\text{noncoro},\lambda}(\mathbf{k})|^2 \quad \forall\, \mathbf{k} \in D\ \forall\, \lambda \in [\lambda_-, \lambda_+] \tag{7.10b}$$

$$X^2(\mathbf{x}) + Y^2(\mathbf{x}) = 1 \forall \mathbf{x}. \tag{7.10c}$$

At this point the objective function is fully linear and therefore convex, and the first constraint is quadratic but convex as well. The only remaining source of non-convexity stems from the phase-only constraint on the complex-amplitude apodizer transmission. Similar to Por (2017) we allow the apodizer to vary not only in phase, but also in amplitude. This convexifies the last constraint and yields the following convex optimization problem:

$$\underset{X(\mathbf{x}),Y(\mathbf{x})}{\text{maximize}} \quad \Re\left\{E_{\text{noncoro},\lambda_0}(\mathbf{0})\right\} \tag{7.11a}$$

$$\text{subject to} \quad |E_{\text{coro},\lambda}(\mathbf{k})|^2 < 10^{-c(\mathbf{k})}|E_{\text{noncoro},\lambda}(\mathbf{k})|^2 \quad \forall\, \mathbf{k} \in D\ \forall\, \lambda \in [\lambda_-, \lambda_+] \tag{7.11b}$$

$$X^2(\mathbf{x}) + Y^2(\mathbf{x}) \leq 1 \forall \mathbf{x}. \tag{7.11c}$$

This problem can easily be solved using standard large-scale optimization algorithms, such as those implemented in Gurobi (Gurobi Optimization, 2016). This convexified problem does not guarantee a phase-only solution, but we will see that in practice all solutions turn out to be phase only. Furthermore, similarly to SPCs and APPs as mentioned above, the solutions space for APLCs is a subspace of this complex-amplitude apodizer

optimization. As the latter produces PAPLCs in practice, a PAPLC will always perform the same or better than an APLC for a given telescope pupil, dark zone geometry and design contrast.

7.2.3 Symmetry considerations

In general symmetric optimization problems are guaranteed to yield symmetric globally-optimal solutions if the optimization problem has multiple solutions (Waterhouse, 1983). Applying the symmetry transformation to one globally-optimal solution can yield a different, but also globally-optimal solution. In our case, the final optimization problem is convex, and as such has only a single, unique solution, so any symmetry in the optimization problem must also be satisfied by the unique solution.

Making use of these symmetries can significantly reduce the computational complexity of the optimization. For example, for a point-symmetric focal-plane mask $M(\mathbf{k}) = M(-\mathbf{k})$ and a point-symmetric dark zone ($-\mathbf{x} \in D \; \forall \; \mathbf{x} \in D$), the transformation $Y(\mathbf{x}) \rightarrow -Y(\mathbf{x})$ is a symmetry of the problem. Therefore $Y(\mathbf{x}) = -Y(\mathbf{x}) = 0 \; \forall \; \mathbf{x}$ and the complex transmission of the apodizer is real-valued. The optimization problem is now significantly simplified. The only remaining non-linear (in this case quadratic) constraint in Equation 7.11c can be replaced by two linear constraints. This yields a linear program, which is extremely easy to solve, even for a large number of variables.

Another interesting example is that of circular symmetry. If the telescope aperture, focal-plane mask, Lyot stop and dark zone are circularly symmetric, then the apodizer must consist of rings and must be completely real-valued (as circular symmetry implies point symmetry). This yields in practice an apodizer consisting of rings of zero and π phase. This simplification significantly reduces the dimensionality of the solution space, thereby substantially reducing the computational complexity, which enables more extensive parameter studies, as shown in Section 7.3.

7.2.4 Tip-tilt correction for one-sided dark zones

For one-sided dark zones, the contrast is constrained only on one side of the PSF. In this case the optimizer tends to add a small tilt on the phase solution. The reason for this is that the optimizer maximizes the real part of the non-coronagraphic PSF at the optical axis, not at its peak. This seemingly tiny difference allows the optimizer to shift the peak of the non-coronagraphic PSF slightly in cases where the decrease in flux at the optical

axis due to the shifted PSF is compensated by the increase in coronagraph throughput due to a less aggressive phase plate design. This centroid shift is unwanted as it effectively increases the inner working angle of the coronagraph. This effect is particularly prevalent for aggressive designs with small inner working angles, as a lot of throughput can be gained from shifting the PSF by a small amount. In these cases, the optimizer will produce a design with a larger inner working angle than what was asked.

The same effect is also commonly seen when optimizing one-sided APPs (Por, 2017), and we deal with it here in the same way. We constrain the intensity of the non-coronagraphic PSF to be smaller or equal to the intensity at the center of the non-coronagraphic PSF. This ensures that the maximum of the non-coronagraphic PSF is always attained at the optical axis so that any movement of the centroid of the planet is not allowed. Mathematically, this constraint is expressed as

$$|E_{\text{noncoro},\lambda_0}(\mathbf{k})|^2 \leq |E_{\text{noncoro},\lambda_0}(\mathbf{0})|^2 \quad \forall\, \mathbf{k}. \tag{7.12}$$

This constraint is convex, and does therefore not affect convexity of the optimization problem. Despite this, the resulting optimization problem is in practice extremely slow to solve, due to the quadratic nature of the added constraint. Adopting a linearized version of this constraint, akin to Por (2017), yields an order of magnitude improvement in run time. A complete version of the optimization problem can be found in Appendix 7.7, including all approximations and modifications necessary to create an efficient numerical optimization problem.

7.3 Parameter study for point-symmetric dark zones

First we discuss point-symmetric dark zones. As this case is extremely similar to APLCs, we compare the PAPLC directly to the equivalent APLC. These APLCs are obtained using a similar optimization procedure. This can be derived starting from Equation 7.11, setting $Y(\mathbf{x}) = 0$ and additionally constraining $X(\mathbf{x} \geq 0$. This optimization problem for APLCs is equivalent to that used by Zimmerman et al. (2016).

To show the performance of a PAPLC, we use simplified telescope pupils. We use a circular telescope pupil with a circular central obscuration with a fractional size of $CO = D_{CO}/D_{tel}$. We use an annular Lyot mask parameterized by an inner and outer diameter, L_{ID} and L_{OD} respectively. These masks are shown schematically in Figure 7.2. We will use an annular focal-plane mask, parameterized by an inner and outer diameter,

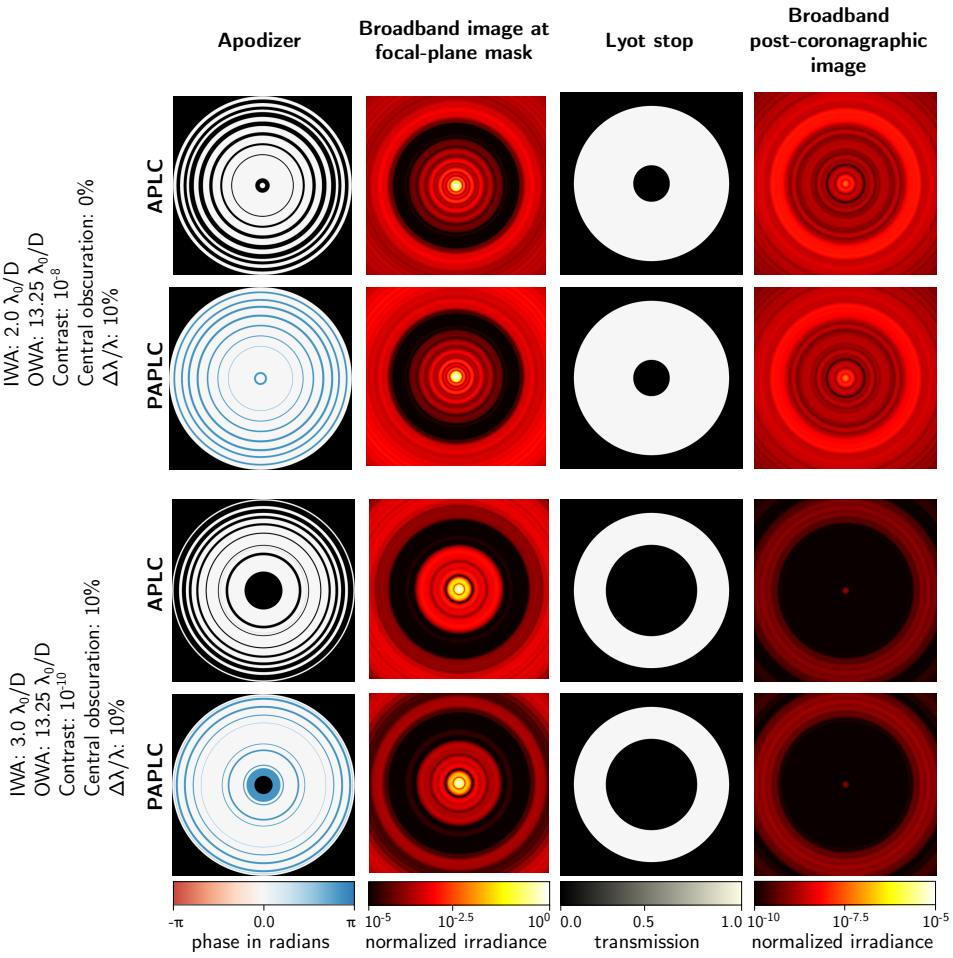


Figure 7.3: Some examples of PAPLC designs with point-symmetric dark zones. For two sets of parameters, we show both the APLC design and the PAPLC design. The phase patterns for the PAPLC consist of regions of 0 or π radians in phase, while the APLC designs consist of regions of 0 and 1 transmission. We show a 10% broadband image just in front of the focal-plane mask in log-scale from 10^{-5} to 10^{0}, and the post-coronagraphic image in log-scale from 10^{-10} to 10^{-5}. The Lyot stop and focal-plane mask are optimized as hyper parameters.

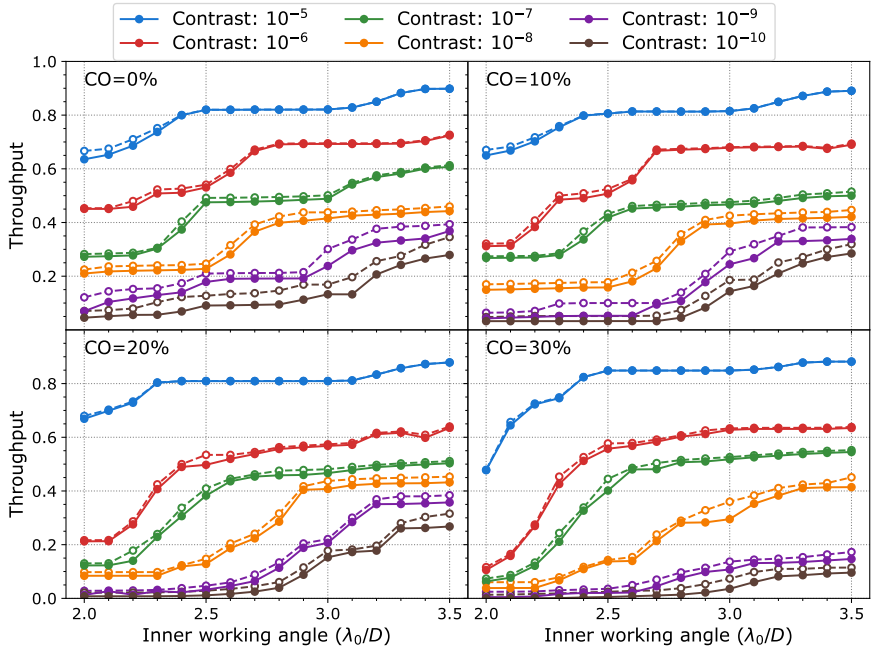


Figure 7.4: Throughput vs inner working angle for various contrasts for an annular dark zone. Solid lines and solid points are APLC designs, dashed lines and open points are PAPLC designs. The design contrast ranges from 10^{-5} to 10^{-10}. Each point is a coronagraph design for which all hyperparameters (focal-plane mask size, and Lyot stop inner and outer diameters) have been optimized.

f_{ID} and f_{OD} respectively. The dark zone is also annular, parameterized by an inner and outer radius $DZ_{min} \geq f_{ID}/2$ and $DZ_{max} \leq f_{OD}/2$. These masks are shown schematically in Figure 7.2.

In Figure 7.3 we show some example PAPLCs along with equivalent APLC designs. Overall we can see that the ring structure in the PAPLCs is very similar to that of the APLCs. The rings are smaller by about a factor of two, which is to be expected as the apodization in phase has twice the effect of a zero transmission ring, however the rings are at the same position.

We perform a full parameter study on the PAPLC and compared it to the similar APLC parameter study. We let the dark zone inner diameter change from $DZ_{min} = 2.0\lambda_0/D$ to $DZ_{min} = 3.5\lambda_0/D$, and fix the dark zone outer diameter at $DZ_{max} = 13.25\lambda_0/D$. We vary the focal-plane mask inner

diameter from $f_{ID} = 2DZ_{min} - 5\lambda_0/D$ to $f_{ID} = 2DZ_{min}$. The focal-plane mask outer diameter is fixed at $f_{OD} = 2DZ_{max}$, as it was found to have no influence on the throughput of both the PAPLC and the APLC. We vary the Lyot mask inner diameter from $L_{ID} = CO$ to $L_{ID} = CO{+}0.4$, and the outer diameter from $L_{OD} = 0.85$ to $L_{OD} = 1$. The relative spectral bandwidth was 10%. We performed the parameter study for design contrasts from 10^{-5} to 10^{-10} with central obscuration ratios varying from 0% to 30%, to represent a full range of potential ground-based and space-based instrument parameters.

In Figure 7.4 we show the maximum throughput for a combination of dark zone inner diameter, central obscuration ratio and design contrast, where all other hyperparameters have been optimized out using the brute-force optimization procedure in Section 7.2.1. APLCs are denoted by filled points and solid lines, while the PAPLC has open points and dashed lines. It is clear that PAPLCs for point-symmetric dark zones do not hold a big advantage over APLCs. Only when throughput is already compromised, the PAPLC can gain a significant advantage, at most $\sim 50\%$ in this parameter space.

Also clear is the plateau behaviour of the throughput: at some points the throughput can be almost insensitive to dark zone inner diameter, while at other points the throughput can drop rapidly for even a small change in dark zone inner diameter. This drop in throughput occurs every 0.5 to $1\lambda_0/D$. The drops change their center position as function of central obscuration ratio and contrast, and can sometimes merge. This behaviour is similar to that of APPs and shaped pupils with annular dark zones Por (2017).

In conclusion: the PAPLC is marginally better than the APLC, but the difference between them is extremely minor, easily overshadowed by the ease of manufacturing of binary amplitude masks. Only where the throughput is low, the PAPLC offers a large relative, but small absolute, performance gain.

7.4 Parameter study for one-sided dark zones

As phase-only apodizers can bring about one-sided dark zones, it is interesting to look at a Lyot-style coronagraph based on a one-sided dark zone. We use a focal-plane mask that blocks all the light on one side of the focal-plane. This mask is offset from the center of the PSF by f_{edge}. We again use an annular Lyot stop. The dark zone is D-shaped on the side of the

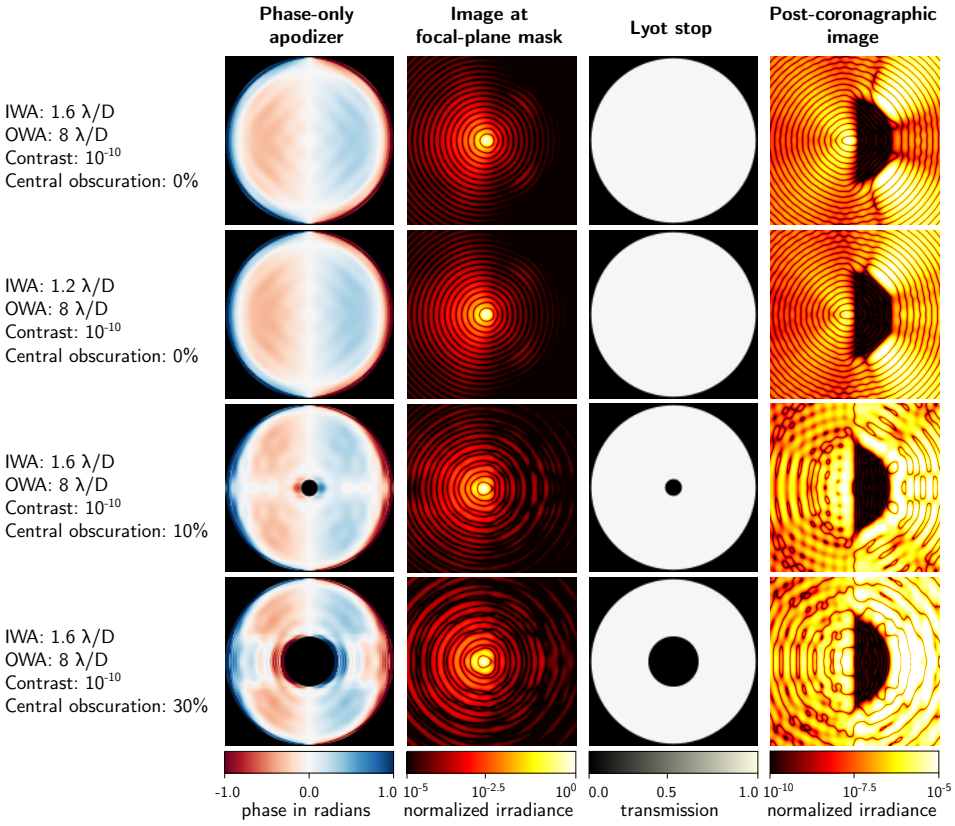


Figure 7.5: Some examples of PAPLC designs with one-sided dark zones. The color scale for phase is from -1 rad to 1 rad but typically the phase pattern rms is ~ 0.4 rad. We show the image at the focal-plane mask with a translucent focal-plane mask to show the positioning of the focal-plane mask relative to the peak of the PSF. In the coronagraph the focal-plane mask is completely opaque. The image at the focal-plane mask is in log-scale from 10^{-5} to 10^{0}. The post-coronagraphic is also in log scale from 10^{-10} to 10^{-5}. The focal-plane mask offset, and Lyot-stop inner and outer diameters were optimized to maximize post-coronagraphic throughput.

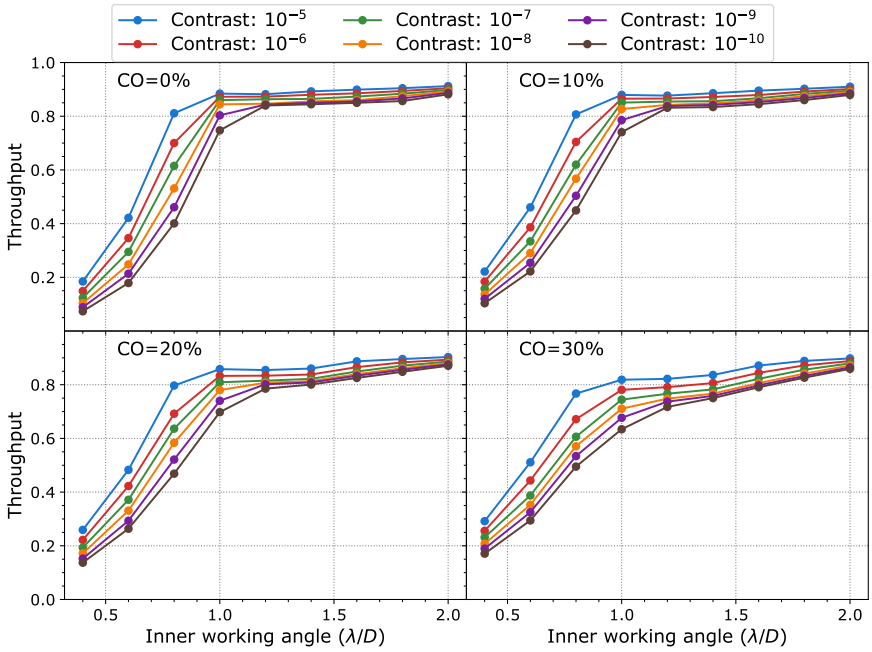


Figure 7.6: Throughput vs inner working angle for various contrasts for a one-sided dark zone. All designs are PAPLC designs. The design contrast ranges from 10^{-5} to 10^{-10}. Each point is a coronagraph design for which all hyperparameters (focal-plane mask offset, and Lyot stop inner and outer diameters) have been optimized. Each of the example designs in Figure 7.5 correspond to a point in this figure.

PSF that is not blocked by the focal-plane mask. These masks are shown schematically in Figure 7.2.

The propagation through the focal-plane mask is performed using standard forward and backward FFTs on a zero-padded pupil. As the knife-edge is invariant across the y-axis, we can view all rows of the pupil as independent and avoid performing an FFT across the y-axis, as well as all FFTs across the x-axis on the zero-padded rows. This makes for a much faster propagation and reduced memory usage. An implementation of this method is available in the open-source package HCIPy (Por et al., 2018).

We show some examples in Figure 7.5. We can see that the phase apodizer acts as an APP, in that it creates a one-sided dark zone with a deepening raw contrast as function of angular separation. At no point however does the stellar PSF at the focal-plane mask reach the required design contrast. The design raw contrast is produced by the focal-plane mask and the Lyot-stop mask, deepening the contrast by more that three decades.

7.4.1 Contrast, inner working angle and central obscuration ratio

We perform a full parameter study on the PAPLC for one-sided dark zones. We let the dark zone inner radius change from $DZ_{min} = 0.4\lambda_0/D$ to $DZ_{min} = 2.0\lambda_0/D$, and fixed the outer radius at $DZ_{max} = 8\lambda_0/D$, mainly limited by the computational run time for the full parameter study. We varied the focal-plane mask offset from $f_{edge} = DZ_{min}$ to $f_{edge} = DZ_{min} - 1.0\lambda_0/D$. The Lyot-mask parameters are varied in the same way as for the point-symmetric dark zone. All masks were calculated for a single wavelength only: we presume monochromatic light. We performed the parameter study for design contrasts from 10^{-5} to 10^{-10} with central obscuration ratios varying from 0% to 30%, to represent a full range of potential ground-based and space-based instrument requirements.

In Figure 7.6 we show the maximum throughput for a combination of dark zone inner diameter, central obscuration ratio and design contrast, where all other parameters have been optimized out. Shrinking the Lyot stop had no positive effects on the throughputs: having the Lyot stop the same as the telescope pupil yielded the best throughput. Also clear is that for dark zone inner radii of $\gtrapprox 1.2\lambda_0/D$ the throughput is relatively independent of design contrast. This is a useful property for coronagraphs destined for space-based instruments. We also see that throughput at a fixed dark zone inner radius is relatively insensitive to central obscuration

ratio of the telescope pupil.

7.4.2 Achromatization and residual atmospheric dispersion

We can produce an achromatic design from any monochromatic design by centering the focal-plane mask (ie. using $f_{\text{edge}} = 0$) and introducing a wavelength-dependent shift using a phase tilt at the phase-only apodizer. This phase tilt acts in the same way as the phase pattern, so we can simply modify the apodizer pattern by adding a tilt on it. In this way, as the PSF grows with wavelength, it will offset the PSF by the same amount, leaving the edge of the focal-plane mask in the same position relative to the rescaled PSF. This makes the one-sided PAPLC completely achromatic in theory (barring experimental effects). One possible downside to this practice is that the planetary PSF inherits this phase tilt, which acts as a grating smearing out its light across the detector. For small focal-plane mask offsets however, this effect can be quite small. For example, for a relative spectral bandwidth of $\Delta\lambda/\lambda_0 = 20\%$, and a focal-plane offset of $f_{edge} = 1.6\lambda/D$, the planet is smeared out across $\Delta\lambda/\lambda_0 \cdot f_{edge} = 0.32\lambda_0/D$, well within the size of the Airy core of the planet. This smearing is independent of field position.

The focal-plane mask is translation invariant in one direction. This means that any tip-tilt errors in that direction will have no influence on the coronagraphic performance other than movement of the coronagraphic PSF. We will explore the tip-tilt senstivity of the PAPLC further in Section 7.5.3. Here we focus on the application of this insensitivity for residual atmospheric dispersion for ground-based telescopes. As telescopes get larger, atmospheric dispersion will become stronger relative to the size of the Airy core, making the performance of the atmospheric dispersion corrector even more critical for future large ground-based telescopes (Pathak et al., 2016).

As the PAPLC is insensitive to tip-tilt along one axis, we can align the residual atmospheric dispersion along the knife edge. In this case, the atmospheric dispersion doesn't degrade the coronagraph performance, and we would only require $\lesssim 1\ \lambda_0/D$ of residual atmospheric dispersion, instead of less than a few tenths to hundredths of λ_0/D for other focal-plane coronagraphs. This significantly relaxes the constraints on the atmospheric dispersion correctors and simplifies their implementation and complexity. Of course, this is only possible on telescopes where the orientation of the pupil is fixed with respect to the zenith, which is the case for all alt-azimuth-mounted telescopes, the majority of current large telescopes.

In Figure 7.7 we show each of these effects for an example PAPLC design. We show the design PAPLC post-coronagraphic PSF, a post-coronagraphic PSF with (isotropic) tip-tilt jitter and a broadband light source, a post-coronagraphic PSF with broadband light and a $0.5\lambda/D$ residual atmospheric dispersion pointed along the focal-plane mask edge, and finally a post-coronagraphic PSF with (isotropic) tip-tilt jitter, residual atmospheric dispersion, broadband light and an injected planet.

7.5 Case studies for VLT/SPHERE and LUVOIR-A

To show that the PAPLC can handle more complicated apertures as well, we present two case studies. The first is a design for VLT/SPHERE, showing that the design method can deal with a complex telescope pupil consisting of spiders and dead deformable mirror actuators. The second is a design for LUVOIR-A, showing that designs with space-based contrasts are possible, and showing that the PAPLC can handle the segmented telescope pupil with spiders and central obscuration seen in future large space telescopes.

7.5.1 VLT/SPHERE

As VLT/SPHERE is a ground-based instrument, it contains an AO system that will limit the raw contrast of resulting images to a level of $\sim 10^{-4}$ to $\sim 10^{-6}$. We fix the design raw contrast at 10^{-7} to avoid having the coronagraph limit the raw contrast of observations. The outer working angle was fixed at $30\lambda/D$. For the Lyot mask we used that of the existing ALC2 Lyot mask in VLT/SPHERE (Guerri et al., 2011) to simplify integration in the VLT/SPHERE instrument. We performed a small parameter study on the inner working angle, of which we present here only one of the solutions. This solution has an inner working angle of $1.4\lambda/D$ and a focal-plane mask offset of $f_{edge} = 1.0\lambda/D$. We show the phase solution, PSF on the focal-plane mask, intensity at the Lyot stop and post-coronagraphic PSF in Figure 7.8.

The light at the positions of the dead actuators on the deformable mirror in VLT/SPHERE are blocked at the apodizer. This provides greater resilience against the unknown positions of the dead actuators. For traditional Lyot coronagraphs and also APLCs, dead deformable mirror actuators are usually blocked in the Lyot stop. This however requires a small blocking element in the focal-plane mask, as in this case the local pertur-

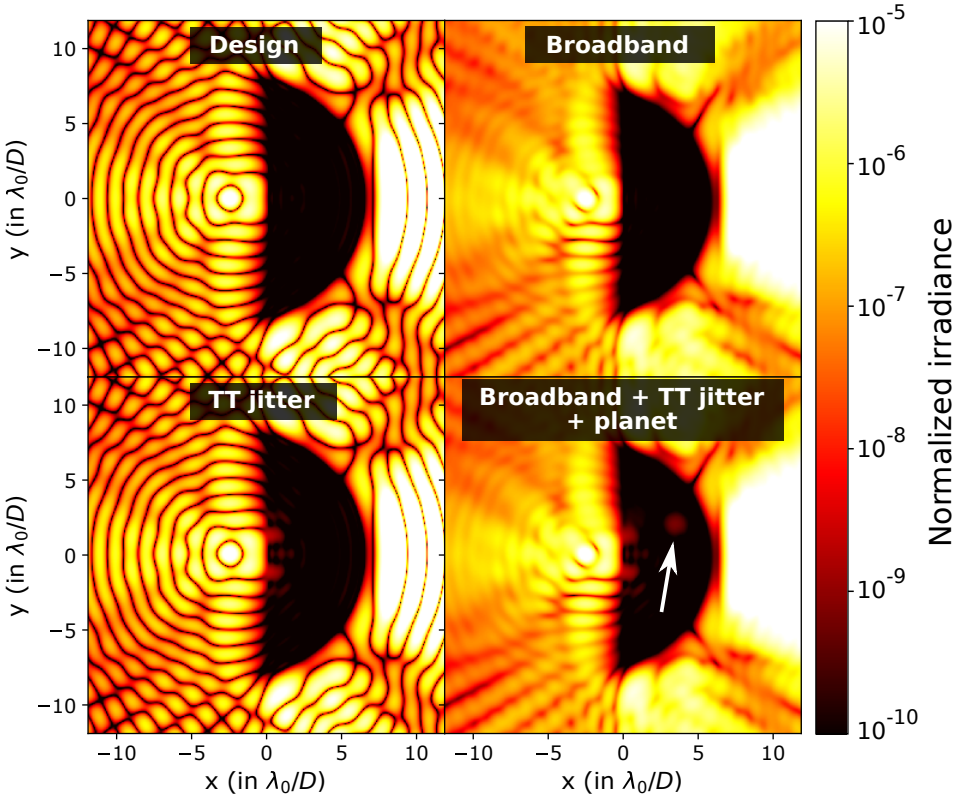


Figure 7.7: Raw post-coronagraphic images for a one-sided dark zone with an inner working angle of $1.6\lambda/D$ with increasing imperfections. *Top left:* Only tip-tilt jitter with $0.003\lambda/D$ rms. *Top right:* tip-tilt jitter and 20% broadband light. *Bottom left:* tip-tilt jitter, broadband light and $0.5\lambda_0/D$ residual dispersion from the ADC. *Bottom right:* tip-tilt jitter, broadband light, residual dispersion and a planet, indicated with an arrow, with a raw contrast of 10^{-9} relative to the host star.

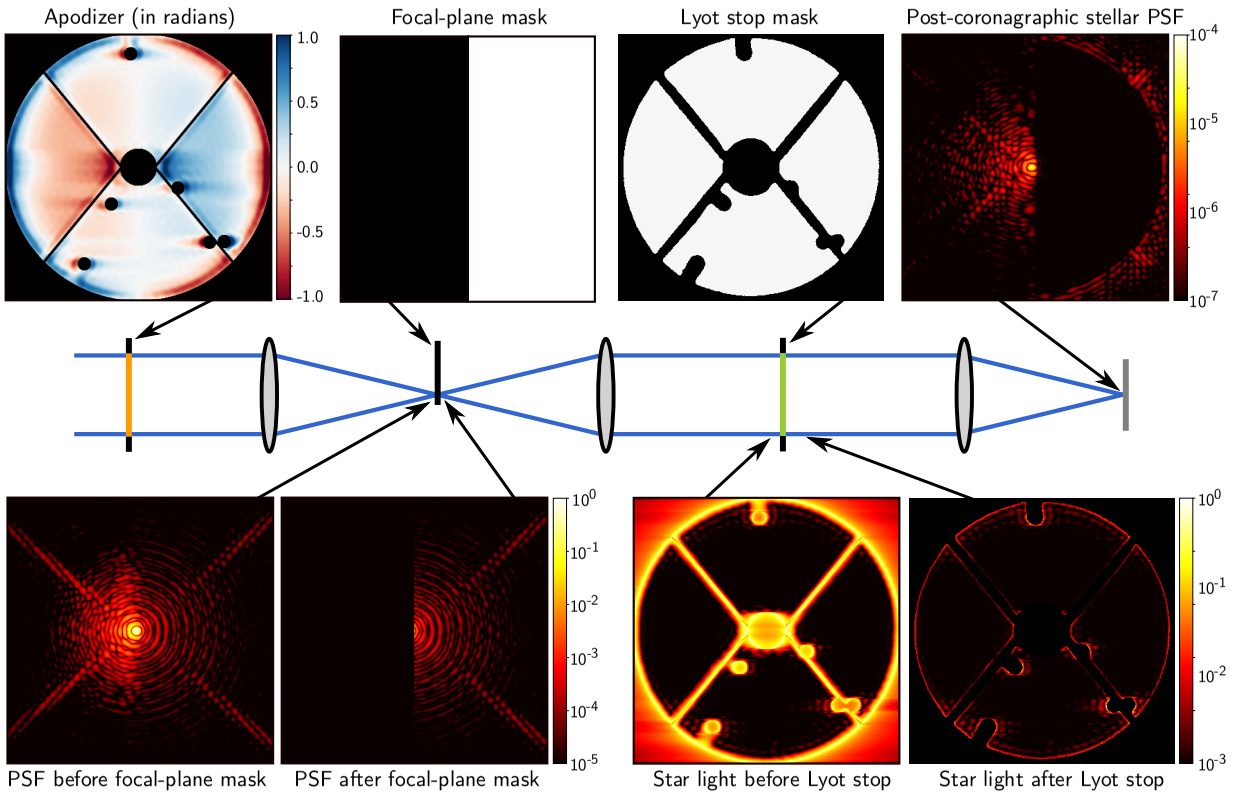


Figure 7.8: The case study design for VLT/SPHERE. We show the apodizer phase pattern, focal-plane mask and Lyot stop. Additionally, we show the light in each of the coronagraphic planes: before and after the focal-plane mask (on a logarithmic scale), and before and after the Lyot stop (on a logarithmic scale, normalized to the peak intensity). Finally, the normalized irradiance of the post-coronagraphic stellar PSF is shown (on a logarithmic scale). Note that the peak in the post-coronagraphic stellar PSF is not the Airy core, but rather a stellar leakage at a relative intensity of $\sim 2 \times 10^{-4}$ that of the star PSF.

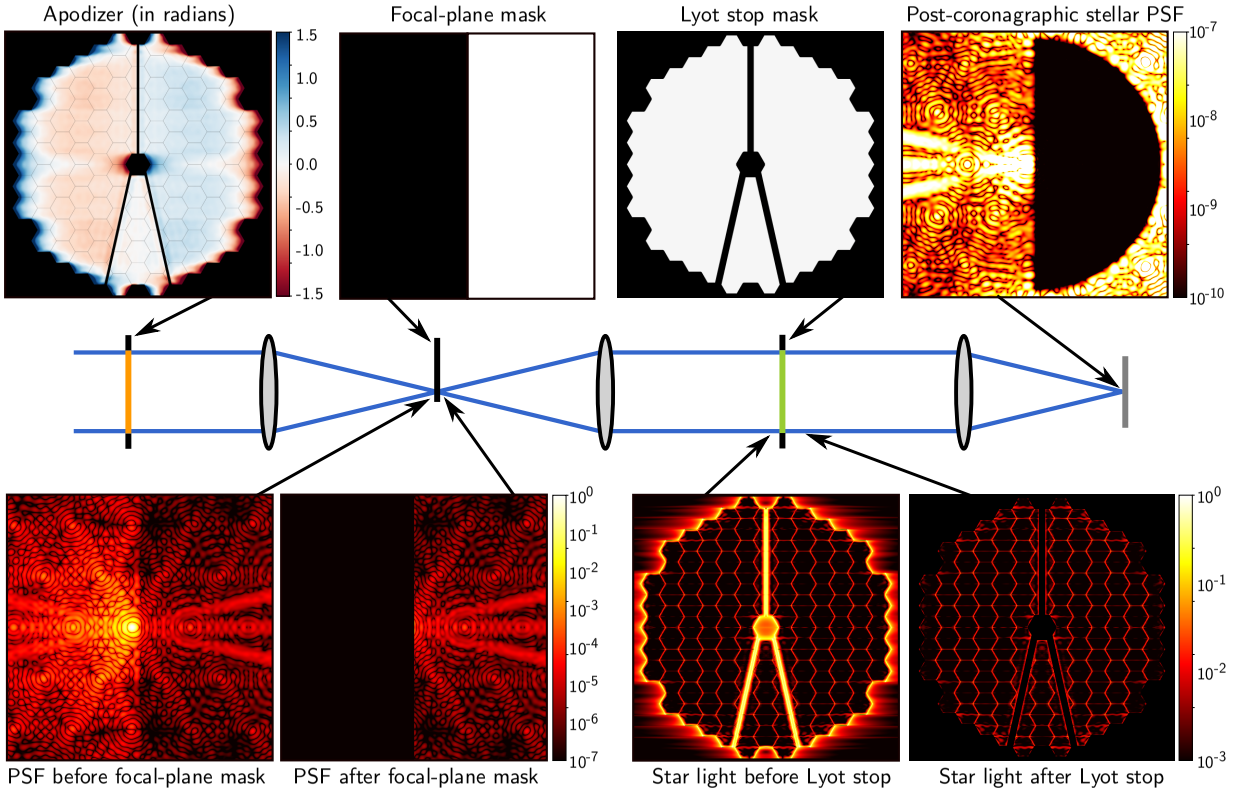


Figure 7.9: The case study design for the LUVOIR-A telescope. We show the apodizer phase pattern, focal-plane mask and Lyot stop. Additionally, we show the light in each of the coronagraphic planes: before and after the focal-plane mask (on a logarithmic scale), and before and after the Lyot stop (on a logarithmic scale, normalized to the peak intensity). Finally, the normalized irradiance of the post-coronagraphic stellar PSF is shown (on a logarithmic scale). Note that the peak in the post-coronagraphic stellar PSF is not the Airy core, but rather a stellar leakage at a relative intensity of $\sim 2 \times 10^{-5}$ that of the star PSF.

bation caused by the dead deformable mirror actuator is kept local by the focal-plane mask making it possible to efficiently block its resulting speckles in the Lyot stop. In our case however, the focal-plane mask blocks over half of the field of view, making it necessary for the light impinging on dead actuators on the deformable mirror to be blocked upstream at the apodizer, as speckles caused by a dead actuator are now spread out in the Lyot stop. Also the support structure of the secondary mirror has been thickened, the secondary obscuration broadened and the outer diameter of the pupil shrunk to accommodate a misalignment in translation of the apodizer of up to 0.5% of the diameter of the re-imaged telescope pupil.

7.5.2 LUVOIR-A

As LUVOIR-A is a space telescope, we fix the design raw contrast at 10^{-10}. The outer working angle was also fixed at $30\lambda/D$. For the Lyot mask we used a thickened version of the LUVOIR-A pupil, where segment gaps, spiders and central obscuration were broadened and the outer diameter was shrunk by $\sim 1.5\%$. No attempt was made to optimize this percentage as a hyperparameter. We performed a small parameter study on the inner working angle, of which we present here only one of the solutions. This solution has an inner working angle of $2.2\lambda/D$ and a focal-plane mask offset of $f_{edge} = 1.8\lambda/D$. We show the phase solution, PSF on the focal-plane mask, intensity at the Lyot stop and post-coronagraphic PSF in Figure 7.9.

7.5.3 Performance

We show the throughput and contrast for both case studies in Figure 7.10. We see that the inner working angles for the two coronagraph designs is $1.4\lambda/D$ for VLT/SPHERE and $2.2\lambda/D$ for LUVOIR-A. At larger angular separations the throughput rises quickly, reaching 90% of its maximum throughput at $4\lambda/D$ and $4.2\lambda/D$ for the VLT/SPHERE and LUVOIR-A design respectively.

The maximum throughput is 66% and 78% for the VLT/SPHERE and LUVOIR-A design respectively. For the VLT/SPHERE design this maximum throughput is primarily limited by the Lyot mask. The throughput without phase-apodizer is $\sim 69\%$, and the addition of any phase pattern on top can only reduce the throughput from there on. The throughput for the LUVOIR-A design however is shared between the phase apodization and the Lyot stop: without the Lyot-stop the throughput is $\sim 87\%$.

We also show the throughput for novel APLC designs for the VLT/SPHERE

	VLT/SPHERE		LUVOIR-A	
Quantity	PAPLC	APLC	PAPLC	APLC
IWA	$1.4\lambda_0/D$	$2.4\lambda_0/D$	$2.2\lambda_0/D$	$3.7\lambda_0/D$
$T_{\max}$	66%	38%	78%	28%

Table 7.1: The inner working angle and throughput for all coronagraph designs shown in Figure 7.10. Care must be taken when directly comparing maximum throughput between PAPLC and APLC designs, due to their different field of view. A discussion of theses quantities can be found in the text.

instrument and LUVOIR-A telescope. The VLT/SPHERE APLC design is a preliminary solution for a possible future upgrade of VLT/SPHERE (courtesy Mamadou N'Diaye). The LUVOIR-A APLC design is a part of a coronagraph design study for the LUVOIR-A aperture (courtesy Rémi Soummer). Their design procedure for both is based on the hybrid shaped pupil/APLC designs by N'Diaye et al. (2016). The inner working angle and maximum throughput of the PAPLC and APLC designs are summarized in Table 7.1. Care must be taken when directly comparing throughput between APLC and PAPLC designs, due to their different field of views. During survey mode, one needs to observe at several sky-rotation angles or roll angles to retrieve a complete image for the full field of view, effectively reducing the throughput by a factor corresponding to the number of observations. During characterization mode however, field of view is irrelevant, and a direct comparison can be made. The PAPLC designs yield almost double or triple the maximum throughput, for the VLT/SPHERE and LUVOIR-A design respectively, mostly or completely neutralizing the disadvantage in field of view. Furthermore, it provides a significantly reduced inner working angle by $1.0\lambda_0/D$ and $1.5\lambda_0/D$ for the VLT/SPHERE and LUVOIR-A designs respectively.

To test the coronagraph as function of tip-tilt jitter of the on-axis source, we show slices of the normalized intensity at various values for tip-tilt errors in Figure 7.11. We assume a normal, isotropic distribution of the tip-tilt offset with a standard deviation of σ. For the VLT/SPHERE design a $< 3 \times 10^{-6}$ contrast for angular separations $> 2.1\lambda/D$ is still achieved with a tip-tilt rms of $\sigma < 0.1\lambda/D$. This tip-tilt performance is (almost) achieved with current high-contrast imagers from the ground at infrared wavelengths (Escárate et al., 2018; Fusco et al., 2014). For the LUVOIR-A design, a contrast of $< 5 \times 10^{-9}$ for angular separations $> 2.5\lambda/D$ is achieved at a

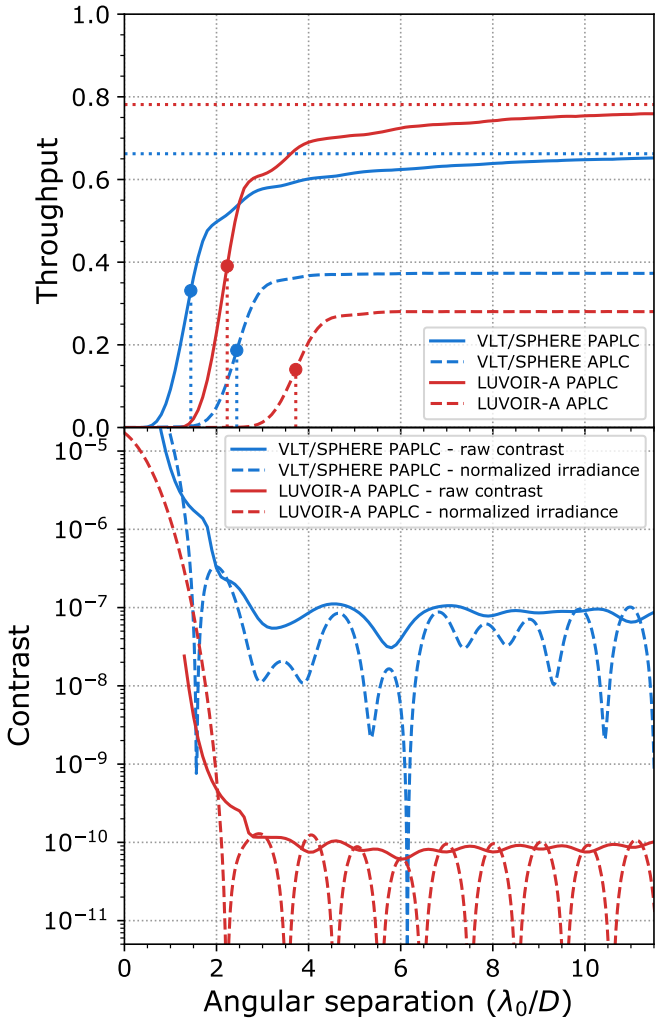


Figure 7.10: The throughput, raw contrast and normalized irradiance for both the VLT/SPHERE and LUVOIR-A designs. Also shown are the throughput for APLC designs for each telescope. Note that the PAPLC has a smaller field of view compared to the APLC designs, which should be taken into account during survey mode but is irrelevant in characterization mode. The inner working angles and maximum throughput for each of the coronagraph designs are listed in Table 7.1.

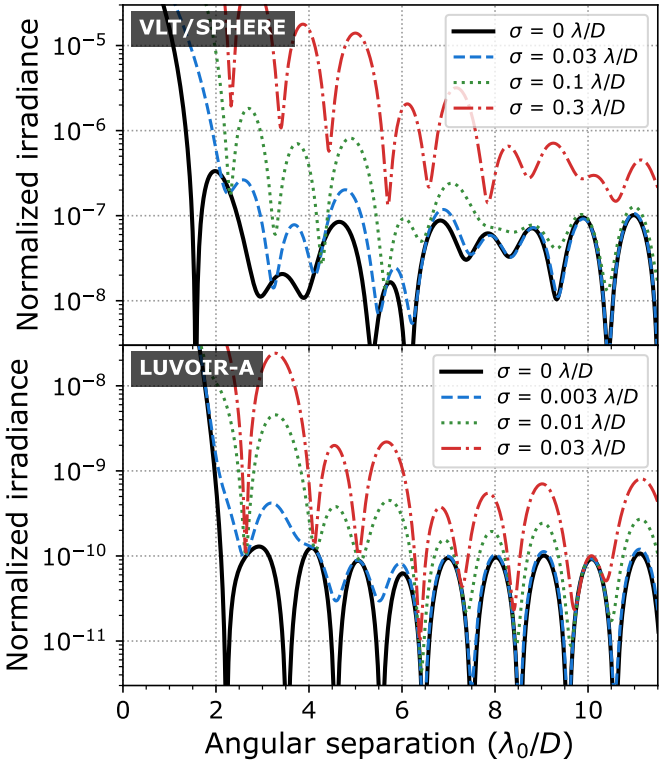


Figure 7.11: Slices of the normalized irradiance for varying values of the RMS tip-tilt error on the star for both the VLT/SPHERE and LUVOIR-A design. The different RMS values were chosen to show the transition from no effect to a significant effect on the normalized irradiance. A normal, isotropic distribution was assumed for tip-tilt.

tip-tilt rms of $\sigma < 0.01\lambda/D$. This tip-tilt sensitivity is significantly worse than the APLC for LUVOIR-A, and has to be improved for the PAPLC to be considered a viable option for giant space telescopes.

Both designs presented in this section, in fact all designs presented in this work, are not made robust against aberrations or misalignment of the Lyot stop. As APLCs can be made robust to aberrations by including these aberrations in the optimization problem (N'Diaye et al., 2015), one can postulate that PAPLCs might be able to be made robust as well. The design of robust PAPLCs and an analysis of the corresponding hit in coronagraphic throughput is left for future work.

7.6 Conclusions

In this work we presented the phase-apodized-pupil Lyot coronagraph. This coronagraph uses a standard Lyot-style architecture and its design procedure is a mix between that for the APLC and the APP coronagraph. Starting from an aperture-photometric methodology, we derive a tractable optimization problem to obtain a globally-optimal solution for the phase pattern in the PAPLC. This shows that an PAPLC will always perform equally or better an APLC by design, given a certain focal-plane mask and Lyot-stop, barring experimental or manufacturing errors.

We distinguished two cases for a PAPLC. The first uses a conventional annular focal-plane mask and produces point-symmetric dark zones. This case provides performance analogous to the APLC, showing similar structure in the apodizer design. Apodizers consist of regions of 0 or π radians in phase, rather than 0 or 1 in amplitude for the APLC.

The second case uses a knife-edge focal-plane mask and is optimized to produce a one-sided dark zone. This case yields apodizers similar to APPs, but use the Lyot stop to gain in contrast. These designs show inner working angles as close as $1.4\lambda/D$ and can be made entirely achromatic. Additionally the coronagraph can reach space-based contrasts ($< 10^{-10}$) at these inner working angles at a throughput of around $60\% - 80\%$ for central obscurations up to 30%. Furthermore, as the knife edge is invariant to translation along one axis, the coronagraph can handle tilt along that axis as well. We can use this to make the coronagraph invariant to residual atmospheric dispersion.

We presented two designs for realistic telescope pupils: one for VLT/SPHERE as an example of a ground-based telescope, and one for LUVOIR-A as an example of a space-based telescope. This shows that the PAPLC can deal with blocking dead deformable mirror actuators, secondary support structure and the segmentation in these telescope pupils.

Future research will focus on testing PAPLC in a lab setting and finally on sky. Additionally, making the PAPLC robust against low-order aberrations is certainly intriguing from a design perspective. Another interesting avenue for future research is integrating the PAPLC with wavefront sensing. As the light from the bright side of the PSF is blocked by the focal plane mask, one can envision using a reflective focal-plane mask instead, and reimaging the bright side on a separate, fast detector. Adding a defocus to this reimaged PSF allows reconstruction of the phase of the incoming wavefront using phase diversity (Gonsalves, 1982) or spatial linear dark field control (Miller et al., 2017).

7.7 Appendix: The full optimization problem

Here we state the full optimization problem, as solved by the large-scale optimization software. This includes linearized constraints on the contrast, and a linearized version of the tip-tilt correction algorithm as presented in Section 7.2.4.

$$\underset{X(\mathbf{x}),Y(\mathbf{x})}{\text{maximize}} \quad \Re\{E_{\text{noncoro},\lambda_0}(\mathbf{0})\} \tag{7.13a}$$

$$\text{subject to} \quad X^2(\mathbf{x}) + Y^2(\mathbf{x}) \leq 1 \quad \forall\, \mathbf{x} \tag{7.13b}$$

$$\Re\{E_{\text{coro},\lambda}\} + \Im\{E_{\text{coro},\lambda}\} \leq \sqrt{10^{-c(\mathbf{k})}S} \quad \forall\, \mathbf{k} \in D \;\forall\, \lambda \in [\lambda_-, \lambda_+] \tag{7.13c}$$

$$\Re\{E_{\text{coro},\lambda}\} - \Im\{E_{\text{coro},\lambda}\} \leq \sqrt{10^{-c(\mathbf{k})}S} \quad \forall\, \mathbf{k} \in D \;\forall\, \lambda \in [\lambda_-, \lambda_+] \tag{7.13d}$$

$$-\Re\{E_{\text{coro},\lambda}\} + \Im\{E_{\text{coro},\lambda}\} \leq \sqrt{10^{-c(\mathbf{k})}S} \quad \forall\, \mathbf{k} \in D \;\forall\, \lambda \in [\lambda_-, \lambda_+] \tag{7.13e}$$

$$-\Re\{E_{\text{coro},\lambda}\} - \Im\{E_{\text{coro},\lambda}\} \leq \sqrt{10^{-c(\mathbf{k})}S} \quad \forall\, \mathbf{k} \in D \;\forall\, \lambda \in [\lambda_-, \lambda_+] \tag{7.13f}$$

$$\Re\{E_{\text{noncoro},\lambda_0}(\mathbf{k})\} \leq \Re\{E_{\text{noncoro},\lambda_0}(\mathbf{0})\} \quad \forall\, \mathbf{k} \tag{7.13g}$$

$$-\Re\{E_{\text{noncoro},\lambda_0}(\mathbf{k})\} \leq \Re\{E_{\text{noncoro},\lambda_0}(\mathbf{0})\} \quad \forall\, \mathbf{k} \tag{7.13h}$$

$$\Im\{E_{\text{noncoro},\lambda_0}(\mathbf{k})\} \leq \Re\{E_{\text{noncoro},\lambda_0}(\mathbf{0})\} \quad \forall\, \mathbf{k} \tag{7.13i}$$

$$-\Im\{E_{\text{noncoro},\lambda_0}(\mathbf{k})\} \leq \Re\{E_{\text{noncoro},\lambda_0}(\mathbf{0})\} \quad \forall\, \mathbf{k} \tag{7.13j}$$

Here S is the expected transmission of the coronagraphic design. After optimization, this expected Strehl ratio can be updated by:

$$S = (\Re\{E_{\text{noncoro},\lambda_0}(\mathbf{0})\})^2. \tag{7.14}$$

The above optimization problem is then restarted with the updated expected Strehl ratio. This process is repeated until the expected Strehl ratio converges.

References

Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155, doi: 10.1051/0004-6361/201935251

Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, ApJ, 736, 19, doi: 10.1088/0004-637X/736/1/19

Carlotti, A., Vanderbei, R., & Kasdin, N. J. 2011, Optics Express, 19, 26796, doi: 10.1364/OE.19.026796

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45, doi: 10.1086/312457

Close, L. M., Males, J. R., Kopon, D. A., et al. 2012, in Adaptive Optics Systems III, Vol. 8447, 84470X

Codona, J. L., Kenworthy, M. A., Hinz, P. M., Angel, J. R. P., & Woolf, N. J. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy. Edited by McLean, Ian S.; Iye, Masanori. Proceedings of the SPIE, Volume 6269, id. 62691N (2006)., 62691N

Currie, T., Kasdin, N. J., Groff, T. D., et al. 2018, PASP, 130, 044505, doi: 10.1088/1538-3873/aaab41

Doelman, D. S., Snik, F., Warriner, N. Z., & Escuti, M. J. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10400, Proc. SPIE, 104000U

Eldorado Riggs, A. J., Zimmerman, N. T., Nemati, B., & Krist, J. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10400, Proc. SPIE, 104000O

Escárate, P., Christou, J. C., Rahmer, G., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Proc. SPIE, 107034F

Fogarty, K., Mazoyer, J., St. Laurent, K., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Proc. SPIE, 106981J

Fusco, T., Sauvage, J. F., Petit, C., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9148, Proceedings of the SPIE, Volume 9148, id. 91481U 15 pp. (2014)., 91481U

Gonsalves, R. A. 1982, Optical Engineering, 21, 829

Guerri, G., Daban, J.-B., Robbe-Dubois, S., et al. 2011, Experimental Astronomy, 30, 59, doi: 10.1007/s10686-011-9220-y

Gurobi Optimization, I. 2016, Gurobi Optimizer Reference Manual, Gurobi Optimization, Inc. http://www.gurobi.com

Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, ApJ, 529, L41, doi: 10.1086/312458

Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, PASP, 127, 890, doi: 10.1086/682989

Kasdin, N. J., Vanderbei, R. J., Spergel, D. N., & Littman, M. G. 2003, ApJ, 582,

1147, doi: 10.1086/344751
Kushner, H. J. 1964, Journal of Basic Engineering, 86, 97, doi: 10.1115/1.3653121
Macintosh, B. A., Graham, J. R., Palmer, D. W., et al. 2008, in Proc. SPIE, Vol. 7015, Adaptive Optics Systems, 701518
Males, J. R., Close, L. M., Morzinski, K. M., et al. 2014, ApJ, 786, 32, doi: 10.1088/0004-637X/786/1/32
Mayor, M., & Queloz, D. 1995, Nature, 378, 355, doi: 10.1038/378355a0
Mennesson, B., Gaudi, S., Seager, S., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9904, Proceedings of the SPIE, Volume 9904, id. 99040L 10 pp. (2016)., 99040L
Miller, K., Guyon, O., & Males, J. 2017, Journal of Astronomical Telescopes, Instruments, and Systems, 3, 049002, doi: 10.1117/1.JATIS.3.4.049002
N'Diaye, M., Pueyo, L., & Soummer, R. 2015, ApJ, 799, 225, doi: 10.1088/0004-637X/799/2/225
N'Diaye, M., Soummer, R., Pueyo, L., et al. 2016, ApJ, 818, 163, doi: 10.3847/0004-637X/818/2/163
Otten, G. P. P. L., Snik, F., Kenworthy, M. A., et al. 2017, ApJ, 834, 175, doi: 10.3847/1538-4357/834/2/175
Pathak, P., Guyon, O., Jovanovic, N., et al. 2016, PASP, 128, 124404, doi: 10.1088/1538-3873/128/970/124404
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 Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Proc. SPIE, 1070342
Pueyo, L., Zimmerman, N., Bolcar, M., et al. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10398, Proc. SPIE, 103980F
Ruane, G., Riggs, A., Mazoyer, J., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Proc. SPIE, 106982S
Slepian, D. 1965, Journal of the Optical Society of America (1917-1983), 55, 1110
Snik, F., Otten, G., Kenworthy, M., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8450, Modern Technologies in Space- and Ground-based Telescopes and Instrumentation II. Proceedings of the SPIE, Volume 8450, article id. 84500M, 11 pp. (2012)., 84500M
Snoek, J., Larochelle, H., & Adams, R. P. 2012, arXiv e-prints, arXiv:1206.2944. https://arxiv.org/abs/1206.2944
Soummer, R. 2005, ApJ, 618, L161, doi: 10.1086/427923
Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, arXiv e-prints, arXiv:1305.5422. https://arxiv.org/abs/1305.5422
Waterhouse, W. C. 1983, The American Mathematical Monthly, 90, 378, doi: 10.2307/2975573
Zimmerman, N. T., Eldorado Riggs, A. J., Jeremy Kasdin, N., Carlotti, A., & Vanderbei, R. J. 2016, Journal of Astronomical Telescopes, Instruments, and Systems, 2, 011012, doi: 10.1117/1.JATIS.2.1.011012

Chapter 8

First laboratory demonstration of the phase-apodized-pupil Lyot coronagraph with integrated high-order wavefront sensor

Adapted from
E. H. Por, A. Potier, P. Baudoz, R. Galicher,
M. A. Kenworthy & C. U. Keller
To be submitted to A&A

Context The next generation of high-contrast imaging instruments on space-based observatories requires sophisticated wavefront sensing and control in addition to a high-performance coronagraph.
Aims We provide a first laboratory demonstration of the phase-apodized-pupil Lyot coronagraph (PAPLC). We show that a single deformable mirror (DM) can serve as the phase-apodizer in monochromatic light. Additionally, we present the integration of a phase-retrieval wavefront sensor to measure high-order wavefront errors simultaneously with coronagraphic images.
Method We installed both a non-reflective and a reflective knife-edge in the focal plane of the *Très Haute Dynamique 2* (THD2) testbed at the Observatoire de Paris. We used electric field conjugation using pairwise DM diversity to minimize light in the dark zone. The light reflected by the focal-plane mask is reimaged with a slight defocus onto the phase-retrieval camera. The resulting image allows us to reconstruct the wavefront using weighted least squares and an empirical interaction matrix.
Results We demonstrate a mean raw contrast of 1.9×10^{-8} in monochromatic light, and 6.7×10^{-8} in 7.5% broadband light for a dark zone between $2\lambda/D$ and $9\lambda/D$ using a coronagraph with an inner working angle of $1.2\lambda/D$. Furthermore, we demonstrated open-loop reconstruction of the wavefront with an integrated phase-retrieval wavefront sensor. The reconstruction error was 30pm per mode for the first 32 Zernike modes for small wavefront aberrations, demonstrating a performance within $3\times$ the fundamental photon-noise limit.
Conclusions These laboratory tests confirm our earlier simulated results in Por (2020) and pave the way for an optically simple approach to broadband high-contrast imaging from space that also features unprecedented wavefront sensing capabilities.

8.1 Introduction

In the last few decades, we have started to unravel the mystery that has captivated humanity since antiquity: is there extraterrestrial life? We have indirectly detected many rocky exoplanets (Borucki et al., 2011), and have started to directly detect Jupiter-sized planets with the latest generation of extreme adaptive optics (ExAO) systems, such as VLT/SPHERE (Beuzit et al., 2019), Gemini/GPI (Macintosh et al., 2008), Clay/MagAO-X (Close et al., 2012; Males et al., 2014), and Subaru/SCExAO (Jovanovic et al., 2015). Technology developments for future space-based observatories with dedicated high-contrast imaging (HCI) instruments such as Roman/CGI (Spergel et al., 2013) and LUVOIR/ECLIPS (Pueyo et al., 2019) are underway.

A typical HCI instrument consists of a coronagraph, which minimizes the stellar light, and a wavefront control system, which corrects for static and dynamic disturbances in the optical system. These disturbances can be categorized as either low-order or high-order aberrations, based on the number of cycles across the pupil. Low-order aberrations result from large-scale flexure and vibrations across the telescope mirrors, caused by thermal and mechanical load, and the relative movement of the optics in the telescope and the instrument. Low-order aberrations typically evolve rapidly, and therefore are controlled by a wavefront control system using the telemetry from a low-order wavefront sensor measuring the light rejected by the coronagraph (Shi et al., 2017, 2018). Additionally, the coronagraph can be designed to be robust to low-order aberrations, which relaxes the constraints of the wavefront control system (eg. N'Diaye et al., 2015; Ruane et al., 2017).

High-order aberrations are caused by polishing errors of the primary mirror, and, for a segmented telescope, by segment misalignments. For a monolithic mirror in space, high-order aberrations tend to evolve slowly, and are therefore easily controlled by the speckle control system, which uses telemetry from the science camera. However, for segmented telescopes vibrations are expected to dynamically misalign segments (Coyle et al., 2019), requiring active control of high-order aberrations in a similar way to low-order aberrations. However, coronagraphs cannot reject high-order aberrations as it would reduce the throughput for the planet. Additionally, most coronagraphs do not permit high-order wavefront sensing without impacting the science image.

The phase-apodized-pupil Lyot coronagraph (PAPLC; Por 2020) is an exception. It uses a standard Lyot-style optical layout, shown schematically

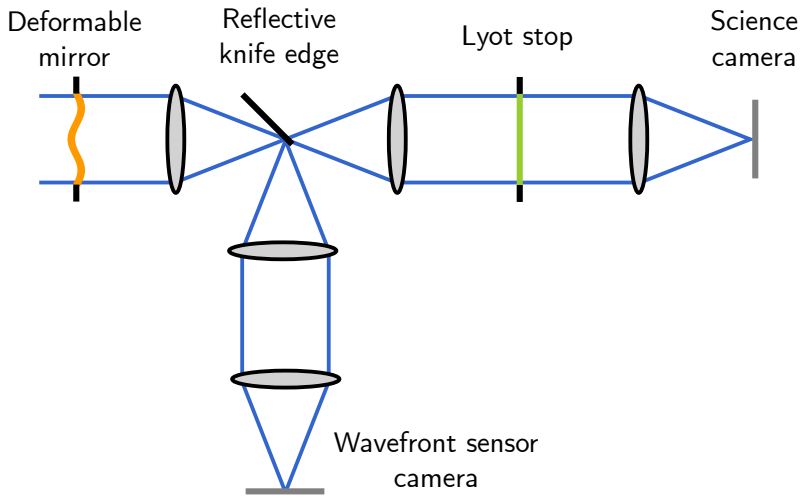


Figure 8.1: The schematic layout for the PAPLC with a deformable mirror and a phase-retrieval wavefront sensor. The PAPLC as presented by Por (2020) uses an achromatic phase-only mask instead of the deformable mirror.

in Figure 8.1. It consists of a pupil-plane phase-only apodizer that creates a one-sided dark zone. A knife-edge focal-plane mask blocks the bright side of the resulting PSF, and the dark side is further filtered using a Lyot-stop mask. Por (2020) demonstrates that PAPLC designs for the segmented LUVOIR-A telescope can achieve inner working angles (IWAs) as small as $2.4\lambda/D$ at a 10^{-10} contrast and high coronagraphic throughput. Por (2020) also eluded to the possibility of adding high-order wavefront sensing by using the light reflected by the focal-plane mask.

In this paper, we present the first laboratory demonstration of the PAPLC and its integrated high-order wavefront sensing capabilities on the *Très Haute Dynamique 2* (THD2) testbed at the Observatoire de Paris. Sect. 8.2 investigates the theoretical limitations of our implementation using optical simulations. Sect. 8.3 presents simulations of the integrated high-order wavefront sensor and its noise characteristics. Sect. 8.4 presents our experimental results for both the coronagraph and wavefront sensor. We summarize our conclusions in Sect. 8.5.

8.2 PAPLC with deformable mirror

The PAPLC, as presented in Por (2020), uses an achromatic phase-only apodizer to modify the PSF falling onto a knife-edge focal-plane mask. The PSF itself is offset by a grating superimposed on the phase-only apodizer, leading to an offset of the PSF relative to the knife-edge that grows linearly with wavelength. This results in a coronagraph that is inherently achromatic, as the changes in size of the PSF as a function of wavelength are compensated by the chromatic offset of the PSF relative to the knife-edge focal-plane mask.

In this paper, we instead use a DM as the apodizer, which introduces a chromatic phase pattern, and we physically offset the focal-plane mask instead of adding tip-tilt on the DM. Additionally, the DM has a much smaller number of degrees of freedom compared to the freeform phase mask of the PAPLC. We investigate two consequences of these changes in this section. In Sect. 8.2.1 we present the monochromatic performance of this version of the PAPLC; Sect. 8.2.2 presents the broadband performance.

8.2.1 Monochromatic performance

A globally-optimal design algorithm is presented in Por (2020) to obtain the phase pattern for the phase apodizer in a PAPLC. We now want to apply this phase pattern on the DM. A simple projection of the optimal phase pattern onto the surface of the DM is insufficient to reach the required contrast due to fitting errors on the DM surface. Additionally, directly enforcing the phase pattern to be a linear combination of DM modes in the globally-optimal design algorithm presented in Por (2020) is not feasible because the optimization is performed on complex electric field amplitudes rather than phase only. Hence, the constraints for enforcing the phase pattern to be separable into DM modes is non-linear, which increases the optimization time by many orders of magnitude.

Instead, we use electric field conjugation (EFC) as presented by Give'On (2009) to iteratively optimize the DM voltages to obtain the phase pattern for the PAPLC, starting from a flat DM. This may not produce the DM pattern with the highest-possible throughput as it does not explicitly maximize throughput, but rather minimizes stellar intensity in the region of interest. In practice, however, the EFC routine yields DM solutions with a sufficiently high throughput, as long as non-aggressive PAPLC design parameters are used. In these cases the DM solution will be driven by the dark zone contrast rather than the throughput of the coronagraph, and

the EFC routine and a global optimizer will find approximately the same solution. A further advantage is that this approach can readily be used on a high-contrast imaging testbed, as most testbeds already include an implementation for performing EFC in combination with some electric field estimation method.

Similar to Mazoyer et al. (2018), the stroke on DM solutions for the PAPLC can be quite large compared to conventional stroke minimization procedures (Pueyo et al., 2009). This means that the local Jacobian can be quite different from the initial Jacobian. We follow Mazoyer et al. (2018) and recompute the Jacobian every ten iterations to ensure that the correction is computed using the local Jacobian of the non-linear system. In practice, simulations without periodic recalculations of the Jacobian tend to converge slower and to a worse contrast level compared to DM solutions that are obtained with periodic recalculations. Again, these differences are small when non-aggressive PAPLC solutions are used.

In Fig. 8.2 we show the result of an EFC simulation with an optical system that mimics the laboratory tests presented in Sect. 8.4. We used an offset of $0.9\lambda_0/D$ between the PSF center and the knife edge of the focal-plane mask; the region of interest starts at $1.8\lambda_0/D$. The other important parameters are shown in Table 8.1. Even with the reduced number of degrees of freedom compared to the apodizer, the DM is capable of producing a pattern that suppresses the stellar light in the dark zone. The total peak-to-valley stroke of the DM surface is $\sim$ 250nm, the vast majority of which is used on actuators that are blocked by the Lyot stop. For actuators inside the back-projected Lyot stop, the peak-to-valley stroke is $\sim$ 40nm, giving rise to the high throughput of the PAPLC. The coronagraphic core throughput is 82% of the core throughput without coronagraph. The achieved contrast is 4.0×10^{-10} averaged over the dark zone from $1.8\lambda_0/D$ until $9\lambda_0/D$.

8.2.2 Broadband performance

We show the wavelength dependence of the PAPLC with a DM by calculating the EFC solution for monochromatic light and then calculating the coronagraphic image for a number of wavelengths across a 7.5% band centered on this wavelength. Fig. 8.3 shows these simulated coronagraphic images. Contrast curves for simulated images with different bandwidths are shown in Fig. 8.4.

As expected, the raw contrast decreases with increasing distance form the center wavelength, which implies that this version of the PAPLC is fairly

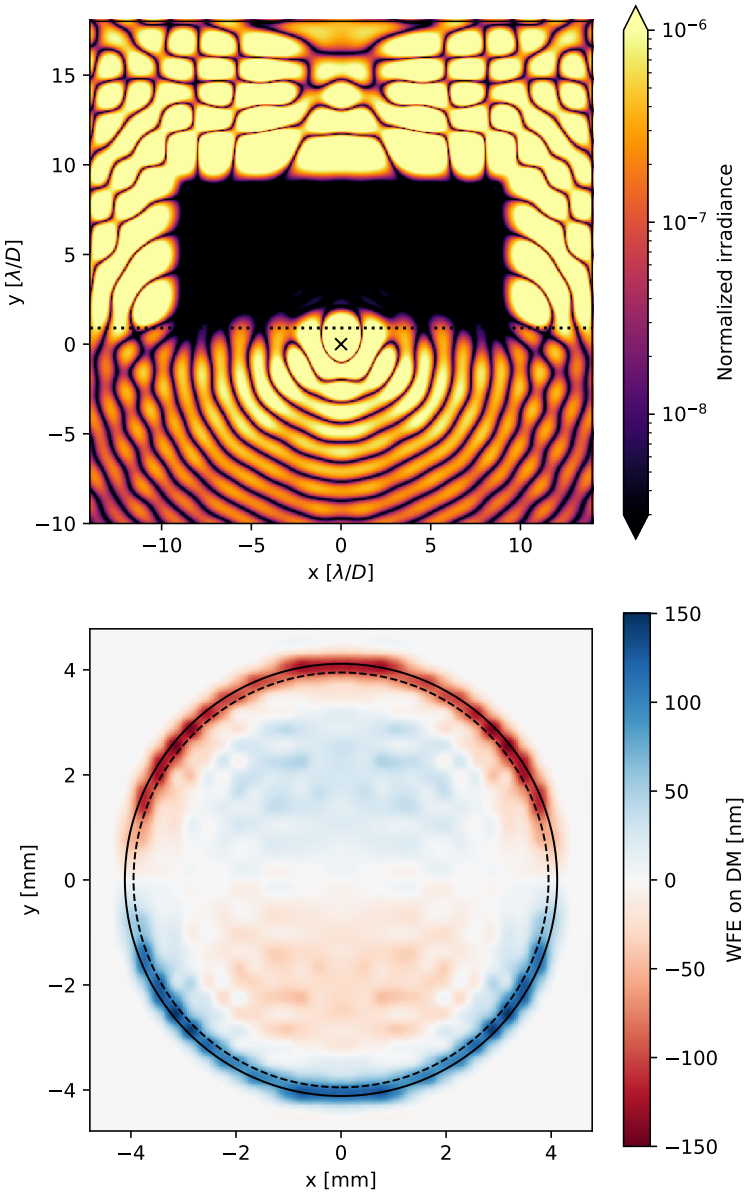


Figure 8.2: The simulated performance of a PAPLC with a DM for monochromatic light, along with the DM surface map. The knife-edge position is indicated by the dotted line. In the DM surface map, the solid and dashed lines indicate the pupil and Lyot stops, respectively.

Parameter	Value
Knife-edge offset	$0.9\lambda_0/D$
Inner working angle	$1.8\lambda_0/D$
Outer working angle	$9\lambda_0/D$
Central wavelength	785nm
Pupil diameter	8.23mm
Lyot mask diameter	7.9mm
DM actuator pitch	300μm
Size of the control region	$\sim 27 \times 27\lambda_0/D$
Number of actuators inside the pupil	591
EFC gain factor	0.9
Jacobian recalculation	Every ten iterations
Pixels across pupil	220pix
Oversampling of the PSF at the knife edge	$32\text{pix}/\lambda/D$

Table 8.1: The simulation parameters listed here were chosen to be analogous to, but not a complete end-to-end simulation of, the THD2 bench.

chromatic. However, the DM solution is not optimized for this bandwidth, which makes this the worst-case scenario. Broadband EFC can, in principle, be used to find a better broadband DM solution, but this was not attempted in our lab experiments on the THD2, and therefore is not performed in this simulation.

8.3 Simultaneous high-order wavefront sensing

8.3.1 Principle

Phase retrieval, in the field of astronomy, is a procedure in which the phase in the exit pupil is determined using one or more intensity images (Gonsalves, 1982). It relies heavily on the mathematical relationship between the phase pattern and the intensity in the measured images, and inverts this non-linear model to recover the unknown phase pattern (eg. Fienup, 1982; Gerchberg, 1972). Here, we will only use the light rejected by the PAPLC by making the knife edge in the focal plane reflective and reimaging the reflected light. Furthermore we will restrict ourselves to a linearized approach, which is appropriate given the small wavefront errors expected in a space-based system, and will only use a single measurement plane (Gonsalves, 2001; Meimon et al., 2010; Polo et al., 2013; Smith et al., 2013).

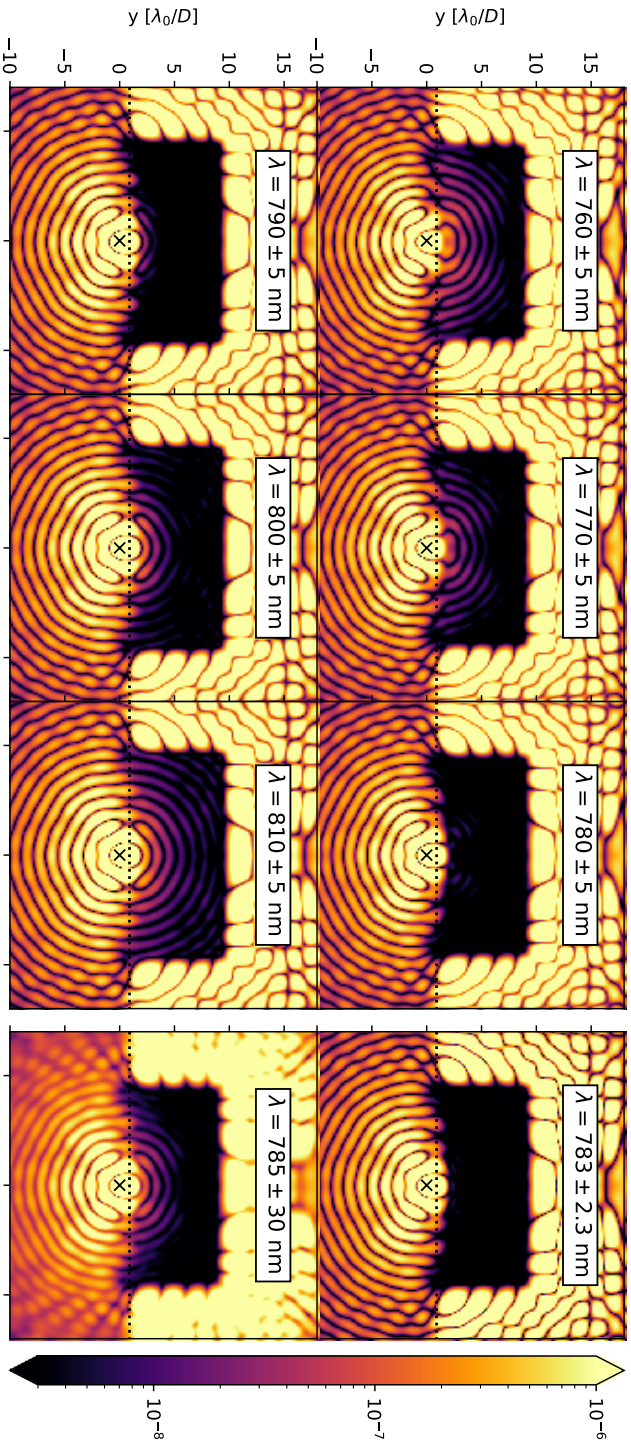


Figure 8.3: The simulated performance of a PAPLC with a single DM in broadband light.

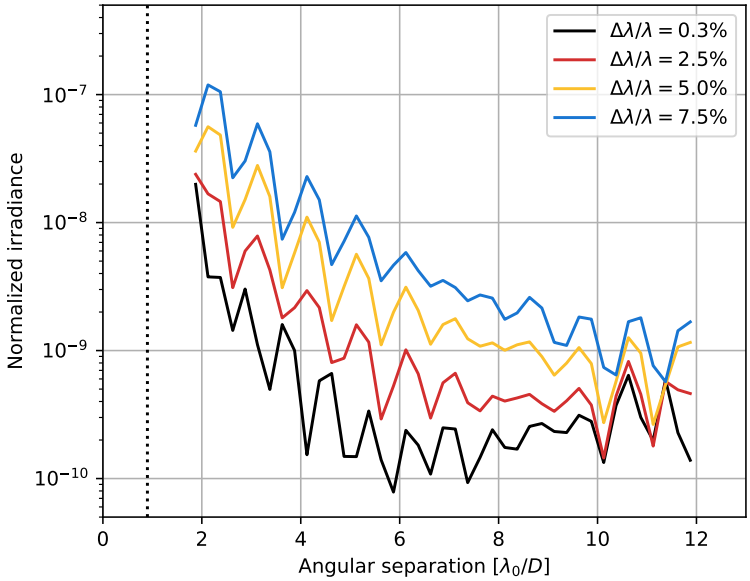


Figure 8.4: Normalized, radial irradiance profiles of simulated images with varying spectral bandwidths.

One well-known problem with single-image phase retrieval is the sign-ambiguity for even modes (Gonsalves, 2001). For odd modes, such as tip, tilt and coma, the focal-plane image will look different depending on the sign of the aberration. Therefore, even from a single image it is immediately clear what sign the aberration has, so we can uniquely reconstruct the phase. However, for even aberrations things are not that easy. When an even aberration, such as defocus, is present, we cannot determine the sign of this defocus since the focal-plane image will look the same regardless of the sign of the aberration. There are two solutions to this problem.

1. *Use an asymmetric telescope pupil*, either by blocking part of the pupil with an opaque mask, or by using the natural asymmetry of the telescope pupil itself (Bos et al., 2019; Martinache, 2013). While this is an extremely simple modification, the partial blocking of the telescope pupil, which increases the sensitivity to even modes, simultaneously reduces the number of photons at the camera, reducing the overall efficiency of the wavefront sensor.

2. *Adding a defocus on the phase-retrieval camera.* This adds the necessary diversity for reconstructing the even modes without any loss in

photon flux (eg. Tokovinin & Heathcote, 2006). However, its application in a coronagraphic instrument is limited due to the simultaneous constraint of a high contrast in the dark zone. Therefore, the defocus must be applied sequentially to coronagraphic observations (Paul et al., 2013), or require reflection off a fold mirror close to the focal-plane mask of the coronagraph (Brady et al., 2018). Both methods reduce the duty cycle of the whole system, but can serve as a calibration method for other wavefront sensors, or initial calibration of the coronagraphic system.

Here we use the light reflected by the focal-plane mask of the coronagraph itself. This has the advantage of relaxing the constraint between contrast and wavefront sensitivity to even modes, by allowing an arbitrary amount of defocus on the reimaging arm for the reflected light while leaving the coronagraphic arm unaffected. This is similar to the current approach for low-order wavefront sensing for Roman/CGI (Shi et al., 2017, 2018), where a Zernike phase-dimple is imprinted onto the light reflected off the coronagraphic focal-plane mask. As this dimple only affects the reflected beam, it has no influence on the transmitted, coronagraphic beam, allowing for simultaneous low-order wavefront sensing and coronagraphic observations. In the next section, we will perform numerical simulations showing the trade-off between sensitivity and the amount of defocus on the phase-retrieval camera.

There is an important difference between our PAPLC and the low-order Zernike wavefront sensor mentioned above: the focal-plane of the PAPLC has an infinite extent, albeit in a single direction, instead of a finite extent as is the case for the focal-plane mask for the hybrid Lyot and Shaped Pupil Coronagraphs in the Roman Space Telescope. This means that the reflected light of the PAPLC also contains information on high-order wavefront aberrations. A Zernike wavefront sensor implementation, similar to the one for the Roman Space Telescope, can also be used for the PAPLC and would most likely perform better than a phase-retrieval-based solution. However, it requires a more sophisticated focal-plane mask implementation, making it more complicated, and such an implementation will be left for future work. For the proof-of-concept high-order simultaneous coronagraphic wavefront sensing presented in this paper, we opted to use the simpler approach of phase retrieval at the cost of a slightly-worse theoretical photon-limited sensitivity.

8.3.2 Empirical modal response and reconstruction

Phase retrieval requires an accurate optical model of the system, which is inverted for wavefront reconstruction. In the small-aberration regime, the optical system can be well approximated using a linearized version of the original model (Gonsalves, 2001). This linearized model can easily be inverted to retrieve the phase aberration. Essentially, the image is modelled as the linear deviation from a reference image a function of of the aberration coefficients. The reference image can be taken either from the original model or from the testbed, at a point when a high contrast has been achieved. The former case is generally limited by the accuracy of the model, while the latter requires another wavefront control system to achieve this high contrast by itself before acquisition of the reference image. In this paper, we used the latter approach.

The response of the system to a specific aberration mode is called a *response function*. The response function for each mode can be obtained in two ways:

1. *From a model.* We can create an accurate model of the optical system, and calibrate all relevant parameters of this model with measurements from the physical optical system. In this case, the accuracy of the reconstruction is generally limited by the accuracy of the model.

2. *From the testbed.* We can simply poke each of the modes on the deformable mirror in the optical setup and empirically measure the response on the phase-retrieval camera. In this case, the reconstruction is typically limited either by the noise floor of the phase-retrieval camera, or by small changes in the optical setup between calibration time and run time.

Here we will use the second method due to its simplicity and accuracy. The reconstruction is performed using a weighted least-squares fit on the phase-retrieval images. Each pixel in the image is weighted using a pixel-by-pixel noise model, which is a combination of the expected read noise and photon noise.

Figure 8.5 shows an example of a phase retrieval using the PAPLC described in Sect. 8.2. Again, simulation parameters are chosen to be comparable to the laboratory demonstration described in Sect. 8.4. A defocus of 1 radian RMS was applied on the reimaged PSF reflected off the focal-plane mask. We used a photon count of 1.8×10^8 per image in the input light beam. The PSF sampling was matched to that of the camera in the

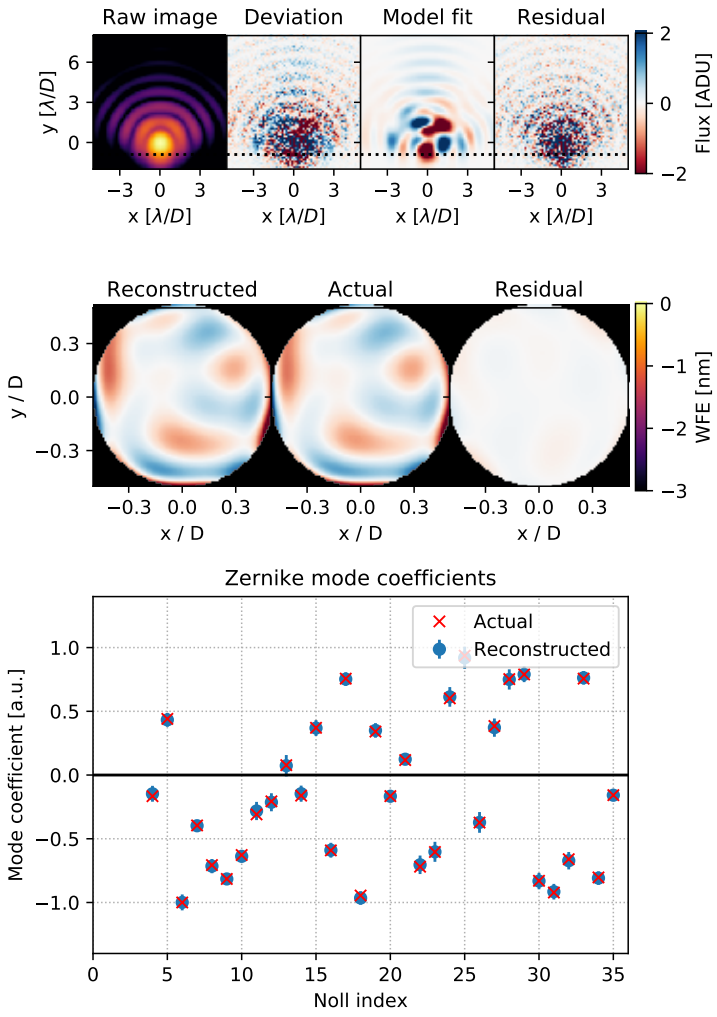


Figure 8.5: An example simulated reconstruction for a phase-retrieval wavefront sensor for a PAPLC coronagraph. The top row shows the reduction of a single frame in the simulated data set with, from left to right, the raw image on a logarithmic scale, the deviation from a reference image, the model fit, and the fit residuals. The middle row depicts the reconstructed wavefront, the actual wavefront, and the residuals of the reconstruction. The bottom plot shows the average reconstructed and actual Zernike mode coefficients. The error bar indicates the 1σ reconstruction error for each frame. The added aberration is 543pm RMS, which is reconstructed with 56pm RMS.

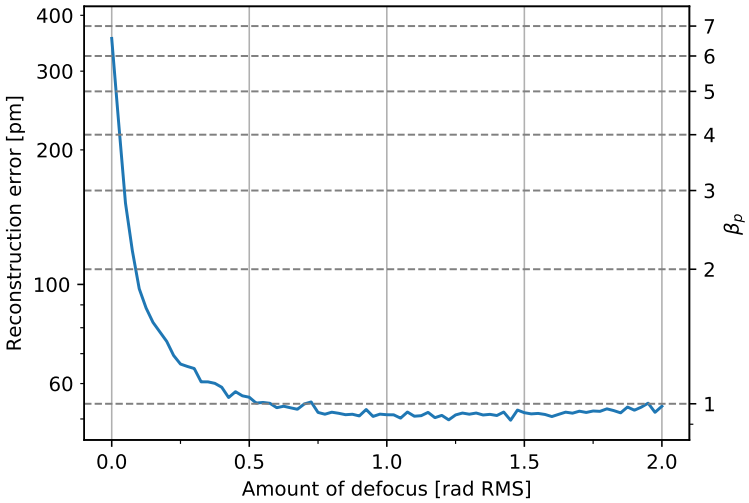


Figure 8.6: The simulated photon noise sensitivity for the phase retrieval. A random aberration of 543pm was applied. When no defocus is applied, even modes do not have a linear response and can therefore not be reconstructed. This yields a reconstruction error of $\sim \frac{1}{\sqrt{2}}$ times the original aberration.

THD2 at 12 px/(λ_0/D). We obtain an empirical interaction matrix with noiseless images. The model fits the simulated image well with the residual containing only photon noise. The reconstructed phase aberration matches the actual aberration perfectly, only deviating by stochastic noise with no observable bias. The same can be seen in the Zernike mode decomposition.

8.3.3 Sensitivity to photon noise

To choose the optimal amount of defocus to apply in our laboratory experiments, we varied the defocus in our simulations and retrieved the reconstruction error due to photon noise. For each amount of defocus, the same random aberration of 543pm RMS was applied. No other noise was present in the simulation. The results are shown in Fig. 8.6. We can see that without any defocus on the reimaged PSF, we retrieve a reconstruction accuracy of 384pm RMS. In this case, even modes cannot be estimated as they have no linear response. The tiny regularization applied during the inversion of the model yields a zero coefficient for these modes. Odd modes do have a linear response and are estimated well, which yields a reconstruction error of $\sim 1/\sqrt{2}$ of the applied aberration. When more defocus is applied, the reconstruction error gradually drops to the fundamental photon noise limit

$\beta_p = 1$ at around 1 to 1.5 radians RMS of added defocus.

8.4 Laboratory demonstration

8.4.1 The THD2 bench

We performed our laboratory demonstration on the THD2 bench that has already been described extensively in the literature (Baudoz et al., 2018a,b; Galicher et al., 2020; Patru et al., 2018; Potier et al., 2018; Singh et al., 2019). A schematic layout of the bench including our additions for the phase-retrieval wavefront sensor is shown in Fig. 8.7. The THD2 contains several light sources; in this work, we used a laser diode (783 ± 2.3nm) and the supercontinuum source with spectral filters with a $\sim$ 10nm spectral bandwidth and center wavelengths of 760, 770, 780, 790, 800 and 810nm. The flux and spectrum that enters the THD2 bench are continuously monitored for calibration purposes.

A single-mode fiber transports the light to the main THD2 bench. The light is collimated, reflects off several deformable mirrors (DMs) and ends up at the knife edge in the focal plane. Light transmitted by this mask is collimated, filtered by a Lyot-stop mask, and focused onto the science camera. The phase apodization is implemented using DM3, which is conjugated to the pupil. DM1, which is 269mm away from the pupil, was left in its flat state. DM2 was replaced with a flat mirror.

We performed experiments with two different knife edges, one opaque and one reflective. Photos of both focal-plane masks are shown in Fig. 8.8. The opaque knife edge was based on a razor blade, mounted in an optical mount. The reflective knife edge was a standard D-shaped flat mirror with a broadband dielectric coating from Thorlabs (BBD1-E02). We obtained slightly better results with the opaque knife edge, but also spent more time optimizing the speckle control algorithm and alignment with this focal-plane mask. Therefore, the difference is presumably not caused by the quality of the knife edge itself. Another design, based on a coated right-angle knife-edge prism was also considered.

Light reflected off the knife edge is reimaged by a single lens onto the phase-retrieval camera (PR-Cam). The lens, with a focal length of 200mm is 600mm from the knife edge. This yields an unused pupil plane image at 200mm and a focal plane image at 300mm from the lens. The camera (Allied Vision Manta G-319B) was offset by a conservative $\sim$ 4mm from the optimal focus, corresponding to a $\sim$ 20% reduction in peak flux at the

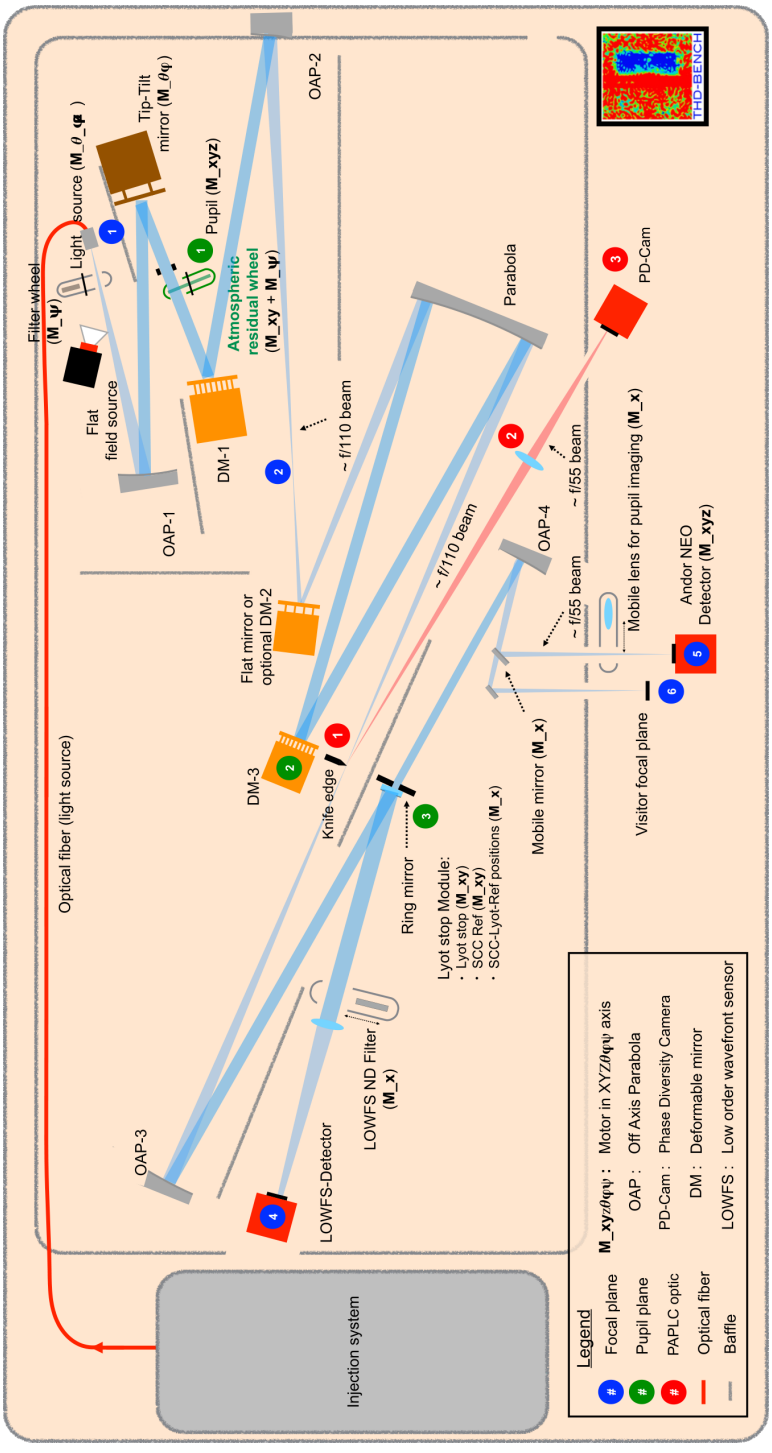


Figure 8.7: Schematic layout of the THD2 bench. Active optics are shown in orange. Numbered green and blue points are pupil and focal planes, respectively. The location of PAPLC optics, including those used by the phase-retrieval wavefront sensor are indicated by numbered red points. The light reflected off the knife edge focal-plane mask is colored red for clarity.

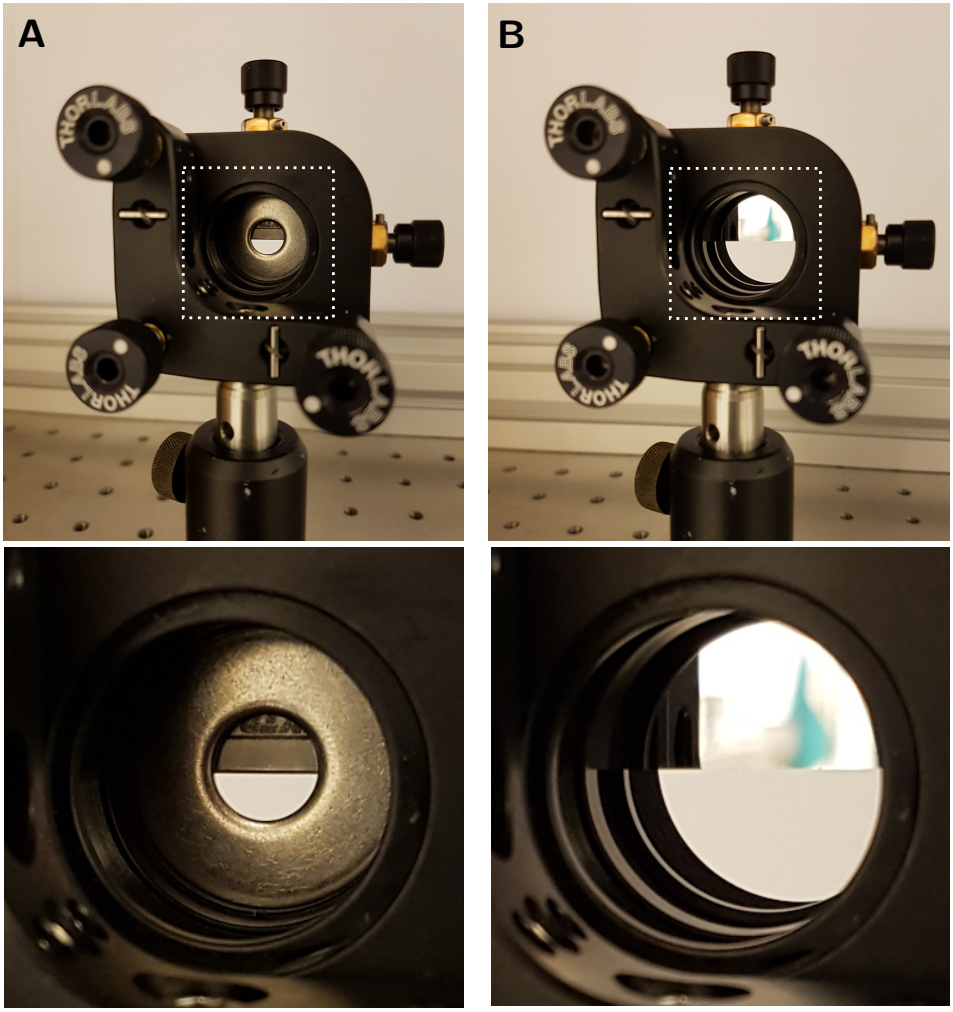


Figure 8.8: Photos of the knife-edge focal-plane masks used for the laboratory demonstration: a) an opaque knife edge based on a razor blade, and b) a reflective knife edge based on a D-shaped flat mirror.

laser diode wavelength and a $\sim$ 60nm rms defocus aberration. Furthermore, several neutral-density filters were placed in front of the camera to reduce the flux and allow the PR-Cam to take unsaturated images at reasonable integration times while simultaneously having sufficient flux on the science camera for speckle control. This, however, had the unintended effect of lengthening the focus, decreasing the amount of defocus to an estimated $\sim$ 0.8mm corresponding to $\sim$ 10nm rms defocus aberration. This decreased the expected wavefront-sensing performance due to photon noise based on simulations. Future experiments will increase the amount of defocus to improve the wavefront-sensing efficiency.

8.4.2 Coronagraphic performance

Monochromatic performance

We performed EFC as described by Potier et al. (2020). For electric field estimation, we used two DM probes, each consisting of a single poked actuator on DM3. Only DM3 was controlled; DM1 was left in its flat position for the whole run. Control voltages were based on a linearized mathematical model of the instrument with a knife-edge focal-plane mask with all DMs in their flat position.

No recalculation of the EFC matrix was performed during the EFC runs. This is not optimal, especially given the large strokes of the PAPLC solution, yielding slow convergence at the end of the speckle control loop as crosstalk between modes is handled by increasing the number of wavefront control iterations, and a higher chance of breaking of the loop. Simulations proved that while better results can be obtained with recomputation of the EFC matrix, it is not strictly necessary to achieve a sufficiently good result at the contrast levels achieved by the THD2 in previous experiments.

The EFC matrix was inverted with Tikhonov regularization. Its regularization parameter was gradually relaxed after convergence to gradually improve the raw contrast. Periodically, the run was stopped and the stellar PSF was recentered by hand using reference spots created by superimposing sinusoidal patterns on DM3. This periodic recentering of the PSF compensates for the tip-tilt induced by the EFC routine. The position of the focal-plane mask was calibrated at the start of the experiment using the flat-field light source, which uniformly illuminates the focal plane, clearly showing the shadow of the knife edge. The focal-plane mask is assumed to be static during the experiment.

Figure 8.9 shows the final monochromatic image in the science focal

plane, the corresponding DM surface deviation from the flat position and the normalized radial irradiance curves in the region of interest.

We partially correct for speckles outside of the final dark zone by first performing speckle control on a larger dark zone with an outer working angle of $13\lambda_0/D$. Halfway through the run, the outer working angle was reduced to $9\lambda_0/D$ to improve the contrast in the smaller dark zone from a better starting point. This is evident in the final image, seen as residual lines of speckles just outside of $13\lambda_0/D$.

Additionally, an incoherent ghost is seen, outlined by the dashed ellipse in Fig. 8.9. This ghost is likely an incoherent copy of the coronagraphic PSF positioned towards the bottom left at $(-2, -3)\lambda_0/D$, and we are seeing a copy of the extended edge of speckles at the top of the dark zone at $(0, 10)\lambda_0/D$. This ghost is not seen by the electric field sensing using DM probes, confirming its incoherent nature. We performed multiple experiments with different orientations of the focal-plane mask to avoid known ghosts in the THD2.

Additionally we see a line of speckles from the center towards the top, outlined by the straight, dashed lines. While these were coherent, as they were also seen by the DM probe measurements, they were not well controlled by the EFC routine. While similar features were seen in simulations, these were not as persistent, and usually disappeared after convergence. We do not have a clear explanation for their existence.

The DM surface map shows the general phase pattern of the apodizer in a PAPLC and the simulations performed in Sect. 8.2.1. However, there are a few peculiarities. The first are the visible vertical lines of actuators that deviate from their immediate surrounding. This can be explained by the assumed "flat" position of the DM. The starting DM flats were based on the last wavefront command of a wavefront control experiment with an FQPM coronagraph (Rouan et al., 2000). As the FQPM coronagraph inherently cancels all speckles on a horizontal and vertical line from the center, these flats still contain any aberrations that produce those speckles. Therefore, we still see an imprint of those aberrations on the DM surface map, as it shows the deviation from the FQPM flat.

Additionally, we see a pronounced structure at the edge of the pupil, the part that is blocked by the Lyot stop. The asymmetry of this edge points towards a small transverse translation of the Lyot stop on the THD2 bench as compared with the EFC model. Further experiments with a better centration of the Lyot stop should confirm this.

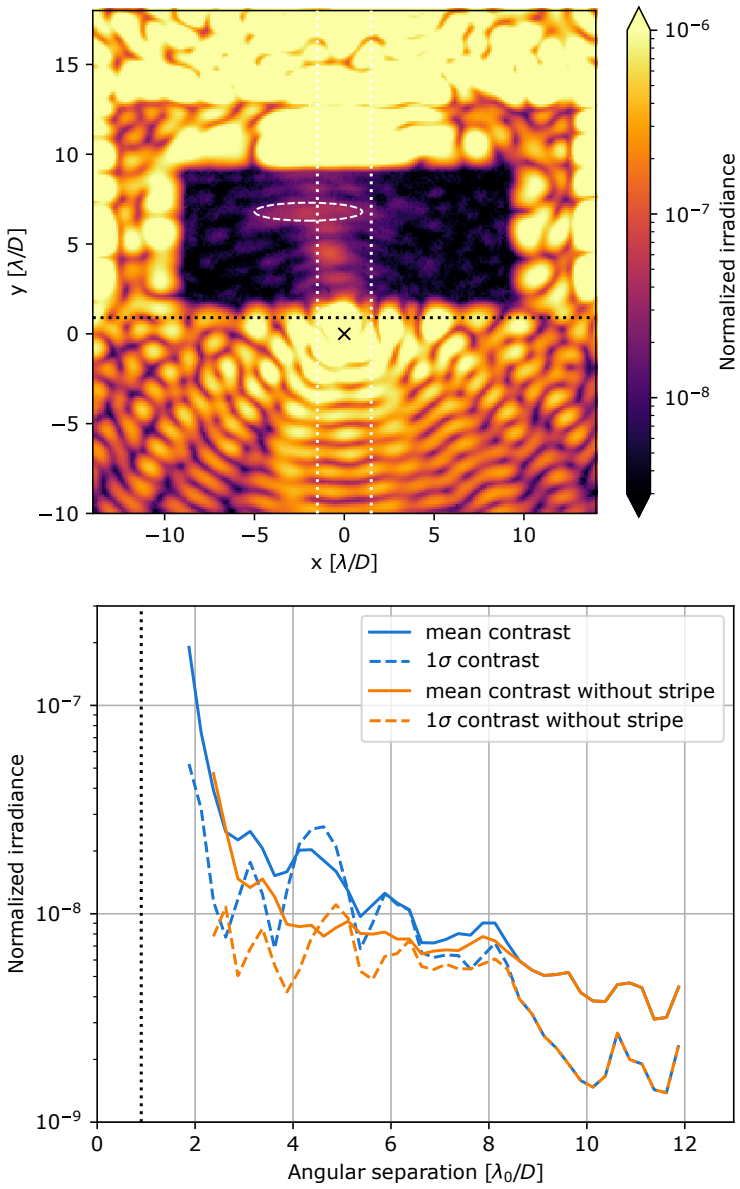


Figure 8.9: The coronagraphic normalized irradiance image and the normalized radial irradiance. The known incoherent ghost, outlined with the dotted white ellipse, has been excluded from the image before taking the radial profile. This curve has been corrected for the coronagraphic throughput of an off-axis source.

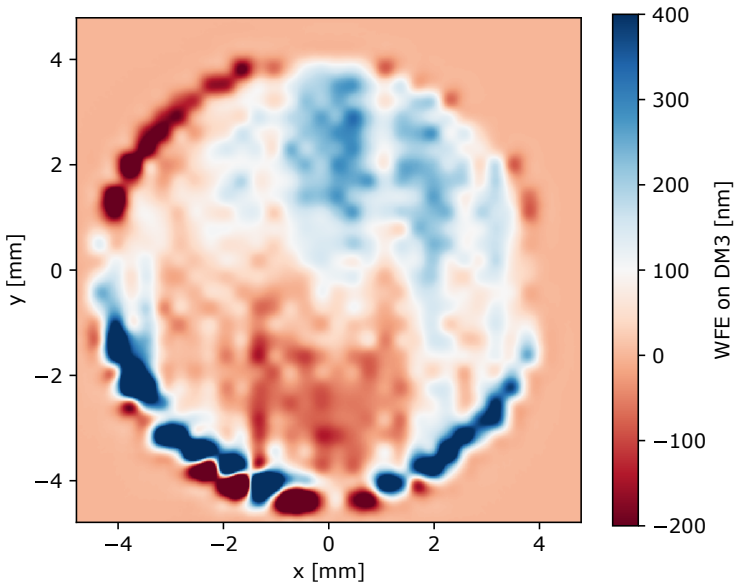


Figure 8.10: The DM3-induced wavefront corresponding to the image in Fig. 8.9 and the normalized radial irradiance profile. The large spikes on the bottom of the DM surface are likely due to a transverse translation of the Lyot-stop mask and pupil with respect to the used model.

8

Broadband performance

Due to the limited photon count in broadband light on the THD2, we cannot perform electric field sensing using a number of wavelengths in the spectral band, as is usually done. Instead, we perform wavefront control using a laser diode in the center of the band, and assess its performance using several broadband filters with the supercontinuum light source, without changing the DM shape.

Figure 8.11 shows the captured images for all filters. Clearly, the further away the wavelength is from the center wavelength of 783nm, the brighter the speckles in the region of interest become. Furthermore, speckles on opposite sides of the center wavelength have a striking similarity, except close to the center of the PSF. This indicates that the electric field is scaling linearly with wavelength, passing through zero at the center wavelength. Therefore, these are likely caused by amplitude aberrations that are corrected by a phase aberration on DM3, which is conjugated to the pupil (Pueyo & Kasdin, 2007). Furthermore, the more speckled and brighter

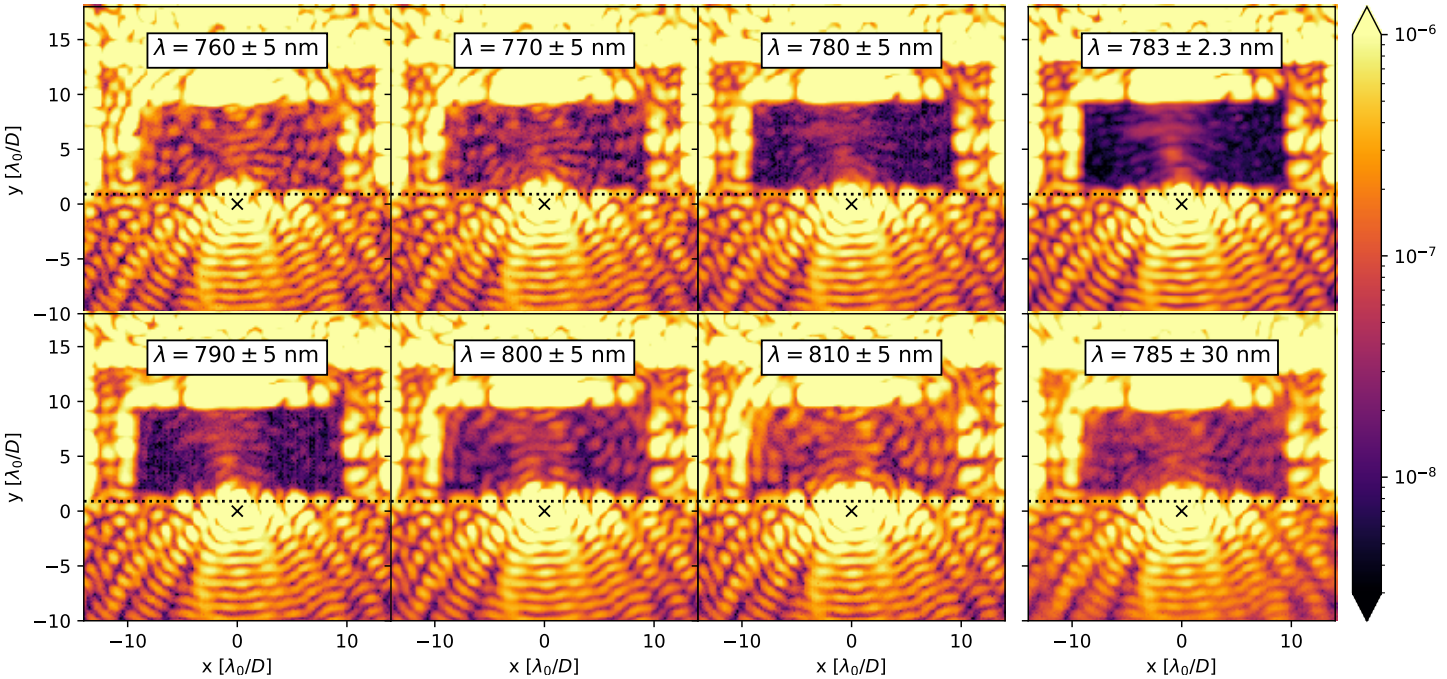


Figure 8.11: Coronagraphic images obtained at several wavelengths and spectral bandwidths. The DM command is optimized for the laser diode (783 ± 2.3nm, upper right). The images for the six narrowband filters between 760 and 810 ± 5nm (first, second and third columns) are combined into a single broadband image (785 ± 30nm, lower right). All images were taken with the same DM shape on January 16th, 2020.

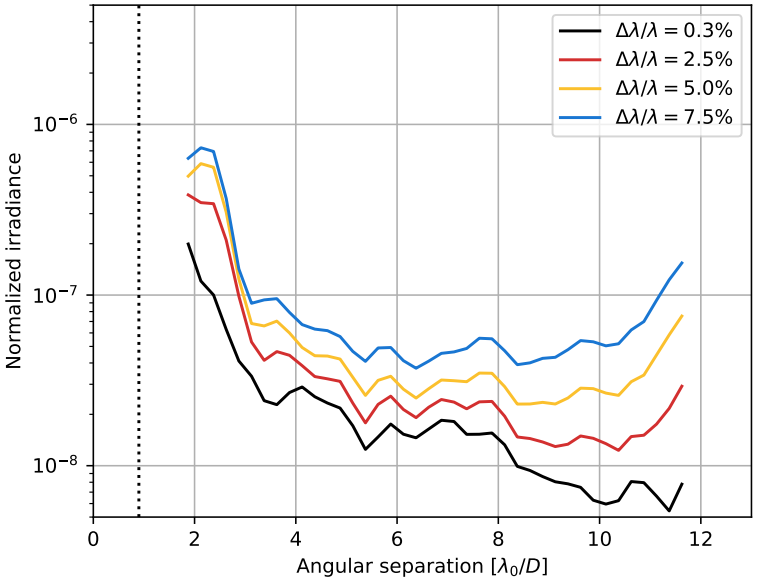


Figure 8.12: Coronagraphic normalized radial irradiance profiles for the images in Fig. 8.11, averaged to represent different spectral bandwidths. These curves are corrected for the coronagraphic transmission for an off-axis source, averaged over the annulli.

appearance compared to the broadband features seen in the simulations in Sect. 8.2.2 suggest that these originate from THD2 itself, rather than the coronagraph. Close to the PSF center, we see a different chromatic behaviour, which is likely caused by the chromaticity of the coronagraph itself.

Figure 8.12 shows the normalized irradiance radial profiles for synthetic broadband images based on the images shown in Fig. 8.11. Again, the radial profile excluded the known incoherent ghost, but included the stripe artefact. We can see that for a 7.5% bandpass, the contrast stays below 8×10^{-7} at $2\lambda_0/D$, and below 1×10^{-7} outside $3\lambda_0/D$.

8.4.3 Phase-retrieval wavefront sensor

Data acquisition and filtering

The PR-Cam integrated for 3ms and read out a 200x200px subregion to increase the frame rate. We performed standard dark correction and added up 80 images with a total effective integration time of 240ms. The images

were corrected for drifts in the power of the light source by normalizing the images to a constant integrated flux level over the detector. All wavefront-sensing experiments were performed with the narrowband 783nm laser diode. At this wavelength, the PR-Cam produces strongly-oversampled images with around 12 pixels per λ_0/D. We manually limit the maximum spatial frequency in the images to 4 cycles per λ_0/D by applying a circular aperture in the Fourier domain. This filters out most pixel-to-pixel read and photon noise, limiting their ability to contaminate measured modal response functions later on. This filtering process was found to significantly reduce crosstalk between modes, while having little effect on the noise level in the sensed modes.

Modal response

In analogy to the simulation in Sect. 8.3, we want to acquire an empirical response matrix on the THD2. However, after subtracting the reference image from our PR-Cam images, clear tip-tilt drifts are visible, even when keeping the THD2 static. Any tip-tilt drift between the positive and negative poke for a specific mode will be attributed to that mode, leading to crosstalk between tip-tilt and all measured modes. We therefore need to remove any tip-tilt drifts from our images before determining the interaction matrix for higher-order modes.

The DM is not capable of accurately reproducing tip-tilt without introducing other modes as well. This would yield noticeable crosstalk between tip-tilt and higher-order modes. As tip-tilt drift is much stronger than the aberrations we are trying to measure, this is not a sufficiently accurate way to calibrate the tip-tilt modes. We therefore opted to reconstruct the tip and tilt modes from the principal components of the reference data set. As the tip-tilt drift represents the vast majority of dynamic instability in the THD2, accounting for $> 98\%$ of all variance in the reference data set, the first two principal components will each be some linear combination of the tip and tilt response functions. The first two components are shown in Fig. 8.14. We manually fitted the direction of the tip-tilt to these images, which are shown in the inset. By comparing the images to the x and y-derivatives of the reference image, we determined the scaling to physical movement of the PSF. The tip-tilt drift during the reference data sets is shown in Fig. 8.15.

Having removed the tip-tilt drift from these images, we can now calibrate the higher-order modes. We chose to use just the first 32 Zernike modes starting from focus. In principle, nothing limits us from using more

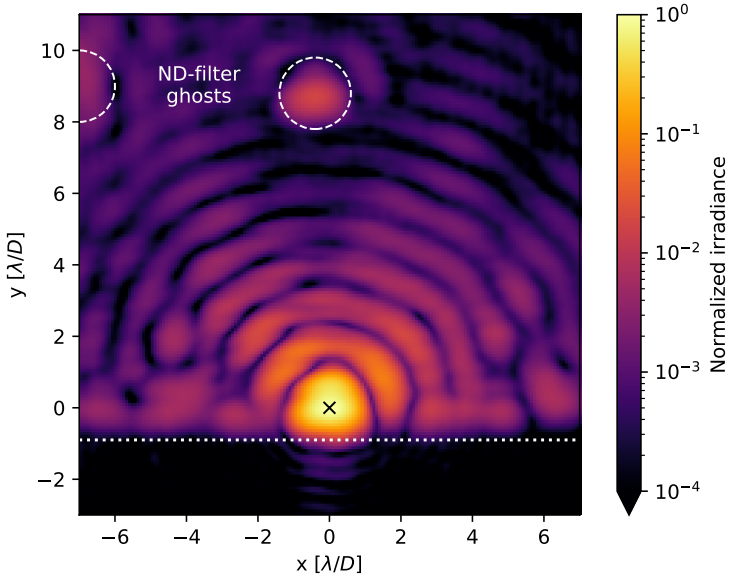


Figure 8.13: The reference image for PR-Cam. Labels identify some prominent ghosts from the ND-filters, which were added to avoid saturation. The cross indicates the location of the laboratory star.

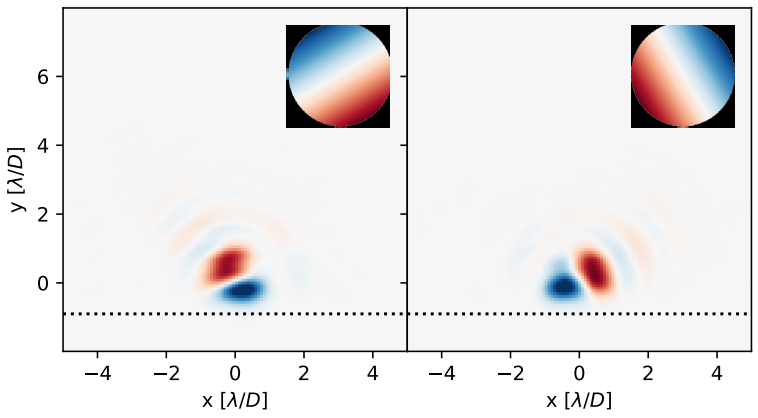


Figure 8.14: The reconstructed tip and tilt modes from the principal component analysis. The x and y tilt for each of these modes were fitted manually.

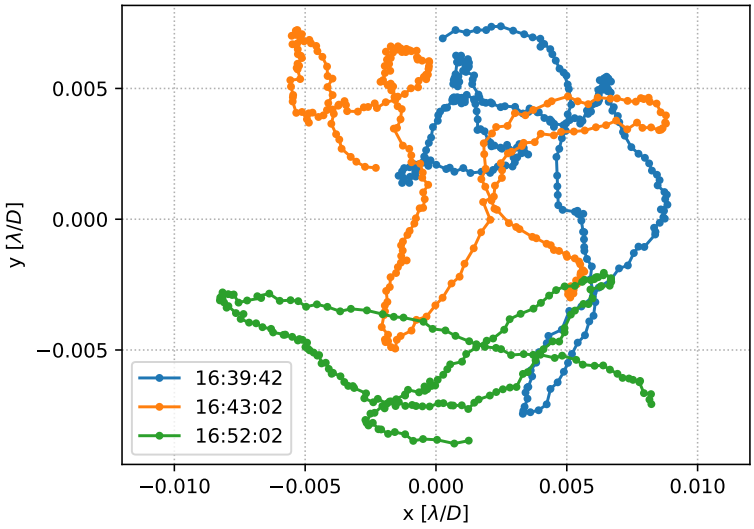


Figure 8.15: The reconstructed tip-tilt drifts during some of the data sets. The duration of each data set was about one minute, with the time between points being $\sim$ 277ms. While implemented for other coronagraphs at the THD2, we did not use any tip-tilt control loop. Future work will use this signal to drive the control loop. The data sets were taken on the 17th of January 2020; the label indicates the local time of the start of the measurement.

modes in the future. The reference diameter of the Zernike modes were slightly oversized by $\sim$ 1 actuator on each side compared to the actual pupil diameter to make sure that we do not suffer from edge effects due to the finite size of the DM influence functions. We ran PR-Cam continuously while sequentially applying our Zernike modes on the DM. Afterwards, we combined images for each mode to retrieve the response function for that mode. On average, we took $\sim$ 10 images per mode. The retrieved response functions are shown in Fig. 8.16 for all 32 modes.

We calibrated the stroke of the DM for all Zernike modes by capturing the coronagraphic science images during the acquisition of the response functions. We fitted a model PAPLC with aberrations to these images, varying the stroke of the Zernike modes, the diameter oversizing of the Zernike modes compared to the actual pupil diameter, the offset of the knife-edge focal-plane mask from the center of the PSF, and a transverse

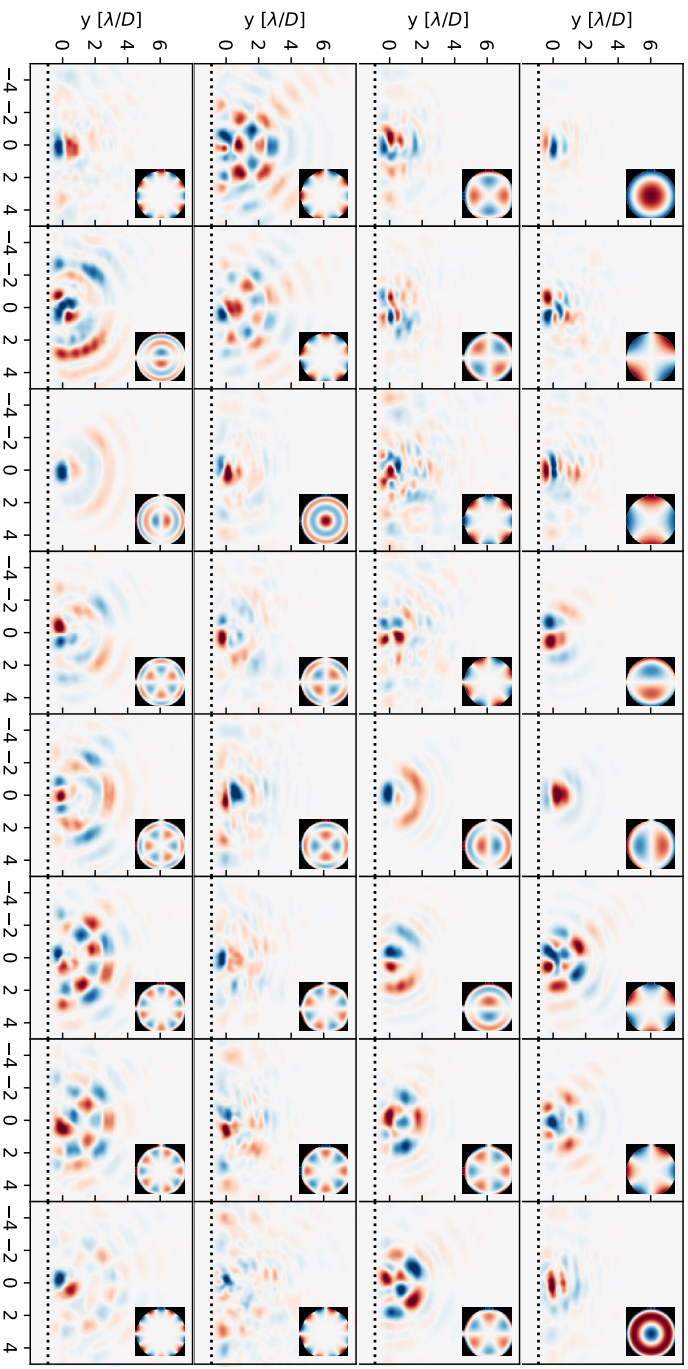


Figure 8.16: The measured PR-Cam response functions for all 32 Zernike modes. The colorscale for each mode is different in order to clearly show the structure in the response function.

	Case A	Case B
Aberration	578 pm rms	5780 pm rms
Systematic error	87 pm rms	521 pm rms
Stochastic error	147 pm rms	147 pm rms
Total error	170 pm rms	541 pm rms

Table 8.2: The approximate error terms. The values for photon noise are independently determined for case A and B.

shift on both the pupil and Lyot-stop masks. The measured coronagraphic images for all 32 modes are shown in Fig. 8.17.

Reconstruction

To show the accuracy of the reconstruction, we applied a known random aberration, composed of the first 32 Zernike modes, on DM3 and performed reconstruction using the image from PR-Cam. We first fit and remove the tip-tilt drift, and subsequently we fit the high-order modes using the retrieved empirical response matrix. This results in 32 coefficients for the Zernike modes for each PR-Cam image. To separate systematic and stochastic errors, we took several hundred images, each composed of 80 subimages. Figure 8.18 shows the analysis of the series of images taken with an aberration with a stroke of 578pm RMS. Figure 8.19 shows the same analysis for the same aberration scaled to a stroke of 5780pm RMS, to differentiate between additive and multiplicative systematic error sources. Additive systematic errors simply add a bias to the estimation, while multiplicative systematic errors scale with the added wavefront aberration.

Stochastic reconstruction errors: photon noise

We use the gain reported by the camera manufacturer ($2.6585e^{-}$/ADU) to perform a calibration of the number of photons incident on the camera. On average, we record 2.25×10^6 electrons per subframe, corresponding to 1.8×10^8 electrons per frame. To focus on the performance of the wavefront-sensing algorithm rather than the quality of the camera, we assume a quantum efficiency of 100% so that the number of electrons is the same as the number of incoming photons. In this framework, the quantum efficiency is integrated in the throughput factor for the optics. This makes sure that these results can be more easily converted to cameras with better quantum efficiencies.

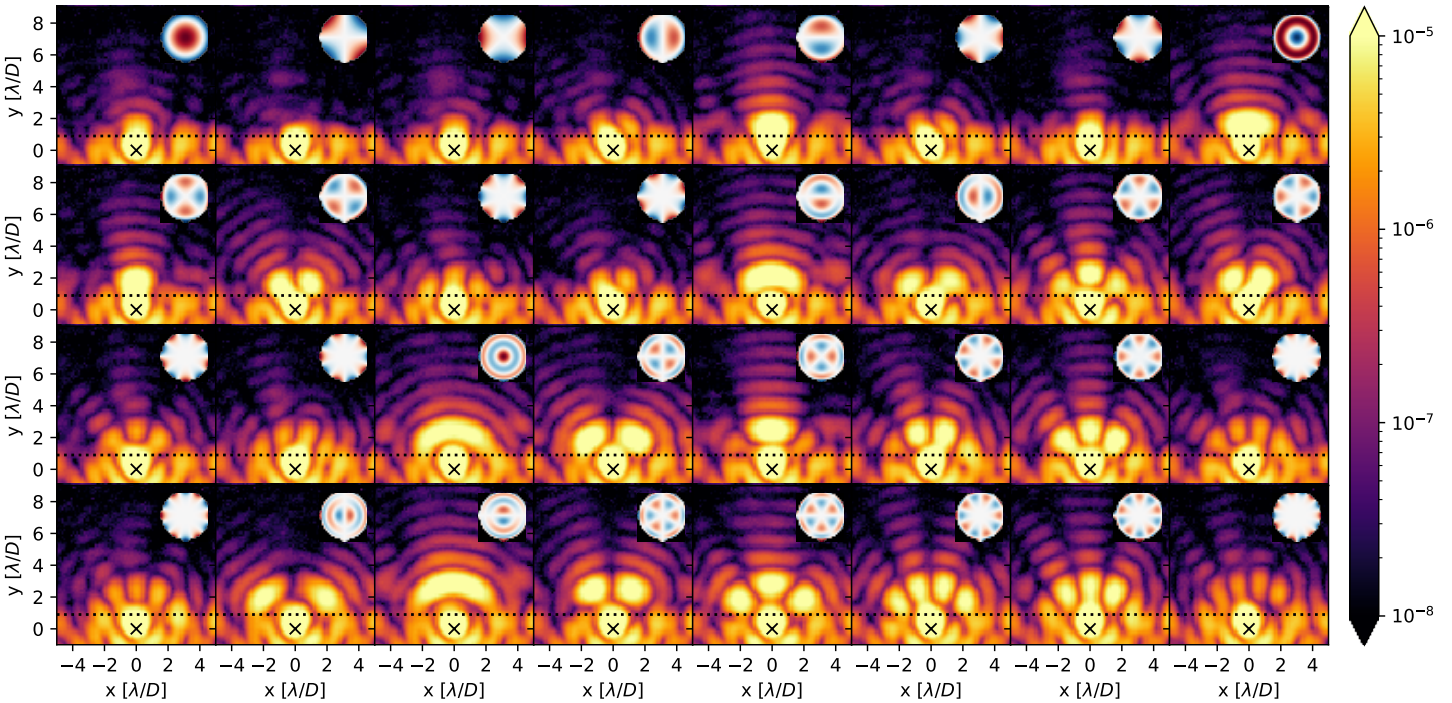


Figure 8.17: The measured coronagraphic images for all 32 Zernike modes. Each Zernike mode was applied with $\sim$ 1.7nm rms. The Zernike modes were applied on DM3 with an $\sim$ 7% larger diameter to avoid edge effects.

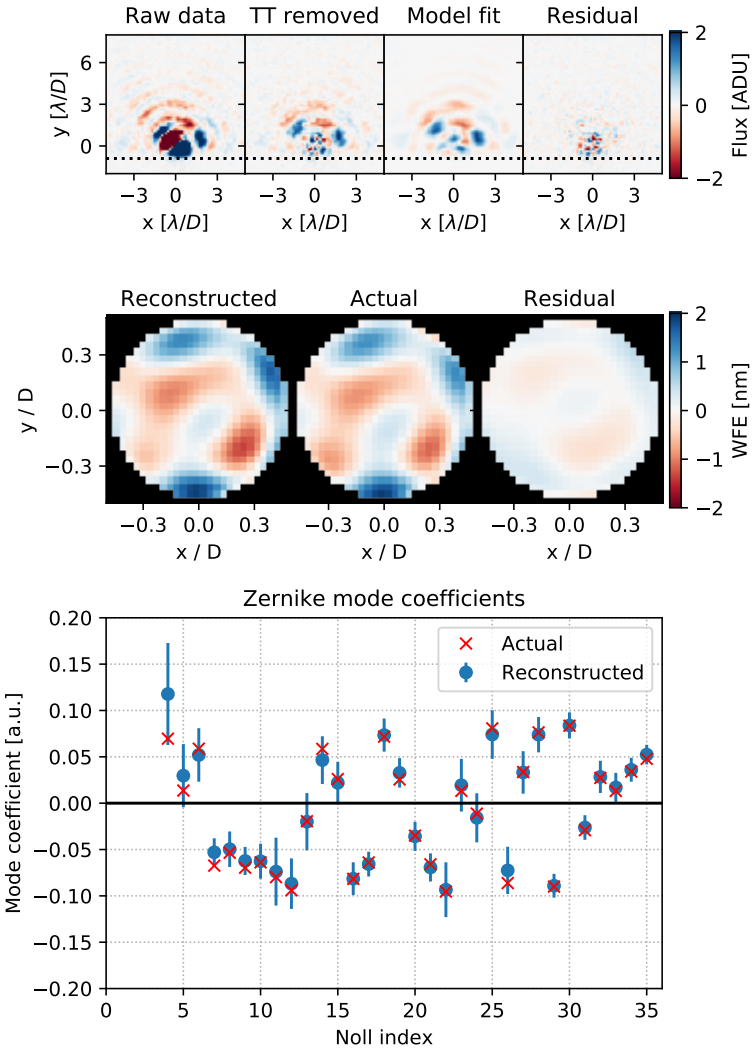


Figure 8.18: Retrieval of the wavefront for a small aberration. The top row shows the reduction of a single frame in the data set with, from left to right, the raw deviation from the reference image, having tip-tilt removed from the image, the fitted model image, and the fitting residuals. The middle row depicts the reconstructed wavefront from a single frame, the actual wavefront that was applied to the DM, and the reconstruction residuals. The bottom plot shows the average reconstructed and actual Zernike mode coefficients. The error bar indicates the 1σ reconstruction error in a single frame. Parameters for this retrieval are listed in Table 8.2 under case A.

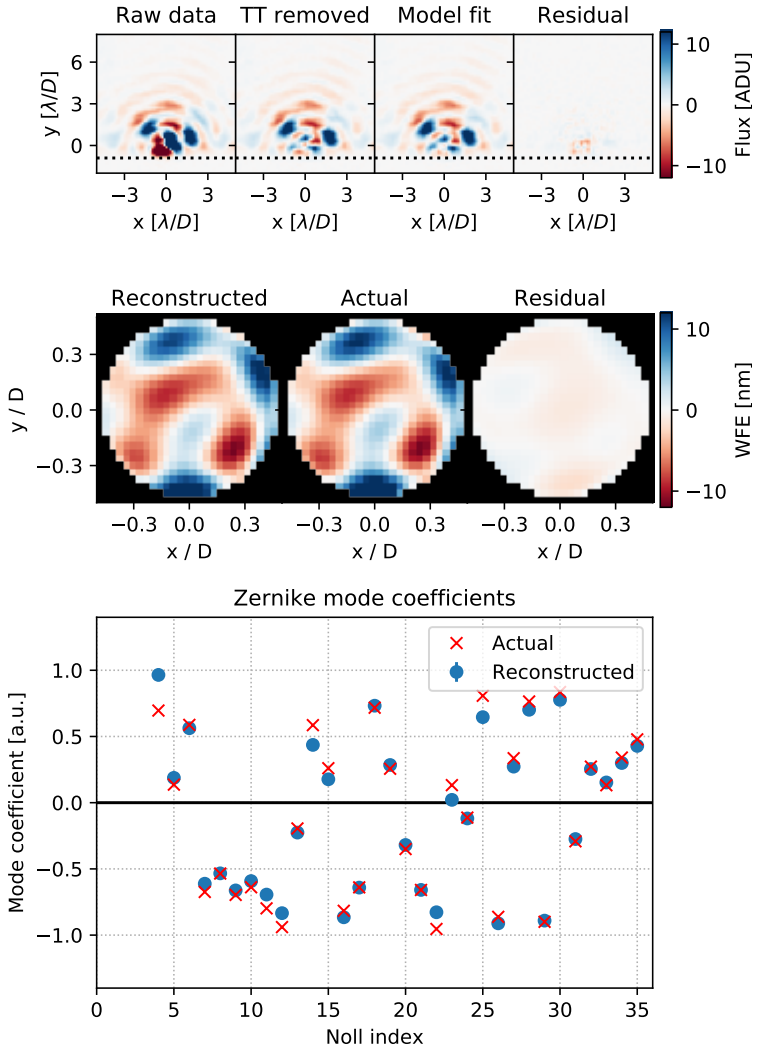


Figure 8.19: The same as Fig. 8.18, but with a larger aberration added. Parameters for this retrieval are listed in Table 8.2 under case B.

The gain yields a fundamental photon noise limit of $\sim$ 54pm for all 32 modes, and $\sim$ 9.5pm for each mode separately. This yields $\beta_p \simeq 3.1$ including systematic errors, and $\beta_p \simeq 2.7$ without systematic errors. This limited performance with respect to the photon-noise-limited performance in Sect. 8.3 is likely the result of the small defocus on the PR-Cam. Future experiments will increase the amount of defocus, yielding both increased wavefront sensor efficiency and enabling longer integration times due to the decreased flux level at the peak of the PSF. Both of these should improve the wavefront-sensing performance.

Systematic reconstruction errors: calibration error

During the acquisition of the empirical response matrix, we are still subject to photon noise. While it is tempting to ignore this noise source since multiple images are taken per mode to increase the signal-to-noise of the recorded response function, and each mode is poked with a larger amplitude to boost its signal, it can still have an effect on the reconstruction accuracy. As the response matrix is kept constant between images for the data set used for reconstruction, the photon noise during calibration will imprint a small crosstalk between modes onto the final reconstruction. In particular even modes, which have a weaker response compared to odd modes, have increased crosstalk with other even modes.

Repeating the simulations performed in Sect. 8.3 with photon noise during the acquisition of the response matrix exhibit similar levels of crosstalk as the observed levels of systematic error in our experiments. We therefore conclude that photon noise in the response matrix dominates other sources of systematic noise.

While this crosstalk impacts open-loop sensing, it should not strongly impact closed-loop performance of this wavefront sensor. After all, unless the crosstalk is extremely strong, a closed-loop experiment will still converge to zero albeit with a slight increase in the necessary number of iterations. In our case, the error caused by crosstalk is $< 10\%$, which has a negligible effect on closed-loop performance. If necessary, calibration can be improved by using stronger modes on the DM during acquisition of the response matrix, something that we opted against due to saturation of the science camera, or by integrating longer during calibration. The former is of course preferred, especially when extending this work to a larger number of modes.

Upon manual inspection of the data, we also observe a small correlation between retrieved tip-tilt and higher-order mode coefficients, indi-

cating crosstalk between tip-tilt and other modes. The strength of this crosstalk is still visible despite the substantial tip-tilt drifts compared to the strength of high-order modes that we are trying to measure. For example, the $\sim 0.005\lambda_0/D$ tip-tilt errors that we observe correspond to $\sim$ 1000pm RMS of wavefront error. We do not have enough data to provide a quantitative estimate for this crosstalk: there is not enough diversity in the tip-tilt signal to retrieve the correlation with sufficient accuracy to separate it from the systematic error due to calibration errors. With closed-loop tip-tilt control, which was turned off during our experiments, we expect this to be negligible.

8.5 Conclusions

This paper presents the first laboratory demonstration of the phase-apodized-pupil Lyot coronagraph with the in-air THD2 testbed. We have shown mean narrowband raw contrasts of 1.9×10^{-8} in a one-sided dark zone between $2\lambda_0/D$ and $9\lambda_0/D$ with an inner working angle of $1.2\lambda_0/D$. In 7.5% broadband light we have shown a mean raw contrast of 6.7×10^{-8} in the same dark zone. This broadband performance is likely the result of testbed limitations. This demonstrates that the original idea of a simple PAPLC is extremely powerful.

Additionally, we have shown a unique capability of the PAPLC: its integrated high-order wavefront sensor that only uses light rejected by the coronagraph. We showed in simulations that this wavefront sensor achieves the fundamental photon noise limit. Furthermore, we demonstrated this wavefront sensor with the PAPLC on the THD2 at a level of 3$\times$ the fundamental photon noise limit, retrieving the first 32 Zernike modes with an accuracy of 30pm per mode in a 240ms exposure. We identified a likely cause for the calibration errors, and showed that this is unlikely to influence closed-loop operation.

References

Baudoz, P., Galicher, R., Patru, F., Dupuis, O., & Thijs, S. 2018a, arXiv e-prints, arXiv:1801.06600. https://arxiv.org/abs/1801.06600

Baudoz, P., Galicher, R., Potier, A., et al. 2018b, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10706, Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III, 107062O

Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155, doi: 10.1051/0004-6361/201935251

Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, ApJ, 736, 19, doi: 10.1088/0004-637X/736/1/19

Bos, S. P., Doelman, D. S., Lozi, J., et al. 2019, A&A, 632, A48, doi: 10.1051/0004-6361/201936062

Brady, G. R., Moriarty, C., Petrone, P., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Proc. SPIE, 106986I

Close, L. M., Males, J. R., Kopon, D. A., et al. 2012, in Adaptive Optics Systems III, Vol. 8447, 84470X

Coyle, L. E., Knight, J. S., Pueyo, L., et al. 2019, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11115, UV/Optical/IR Space Telescopes and Instruments: Innovative Technologies and Concepts IX, 111150R

Fienup, J. R. 1982, Appl. Opt., 21, 2758, doi: 10.1364/AO.21.002758

Galicher, R., Huby, E., Baudoz, P., & Dupuis, O. 2020, A&A, 635, A11, doi: 10.1051/0004-6361/201936971

Gerchberg, R. W. 1972, Optik, 35, 237

Give'On, A. 2009, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7440, Proc. SPIE, 74400D

Gonsalves, R. A. 1982, Optical Engineering, 21, 829

—. 2001, Optics Letters, 26, 684, doi: 10.1364/OL.26.000684

Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, PASP, 127, 890, doi: 10.1086/682989

Macintosh, B. A., Graham, J. R., Palmer, D. W., et al. 2008, in Proc. SPIE, Vol. 7015, Adaptive Optics Systems, 701518

Males, J. R., Close, L. M., Morzinski, K. M., et al. 2014, ApJ, 786, 32, doi: 10.1088/0004-637X/786/1/32

Martinache, F. 2013, PASP, 125, 422, doi: 10.1086/670670

Mazoyer, J., Pueyo, L., N'Diaye, M., et al. 2018, AJ, 155, 7, doi: 10.3847/1538-3881/aa91cf

Meimon, S., Fusco, T., & Mugnier, L. M. 2010, Optics Letters, 35, 3036, doi: 10.1364/OL.35.003036

N'Diaye, M., Pueyo, L., & Soummer, R. 2015, ApJ, 799, 225, doi: 10.1088/0004-637X/799/2/225

Patru, F., Baudoz, P., Galicher, R., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, 107032L

Paul, B., Sauvage, J. F., & Mugnier, L. M. 2013, A&A, 552, A48, doi: 10.1051/0004-6361/201220940

Polo, A., Pereira, S. F., & Urbach, P. H. 2013, Optics Letters, 38, 812, doi: 10.1364/OL.38.000812

Por, E. H. 2020, ApJ, 888, 127, doi: 10.3847/1538-4357/ab3857

Potier, A., Baudoz, P., Galicher, R., Patru, F., & Thijs, S. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, 106986G

Potier, A., Baudoz, P., Galicher, R., Singh, G., & Boccaletti, A. 2020, A&A, 635, A192, doi: 10.1051/0004-6361/201937015

Pueyo, L., & Kasdin, N. J. 2007, ApJ, 666, 609, doi: 10.1086/518884

Pueyo, L., Kay, J., Kasdin, N. J., et al. 2009, Appl. Opt., 48, 6296, doi: 10.1364/AO.48.006296

Pueyo, L., Stark, C., Juanola-Parramon, R., et al. 2019, in Techniques and Instrumentation for Detection of Exoplanets IX, ed. S. B. Shaklan, Vol. 11117, International Society for Optics and Photonics (SPIE), 37 – 65. https://doi.org/10.1117/12.2530722

Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., & Labeyrie, A. 2000, PASP, 112, 1479

Ruane, G., Mawet, D., Jewell, J., & Shaklan, S. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10400, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 104000J

Shi, F., Cady, E., Seo, B.-J., et al. 2017, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10400, Proc. SPIE, 104000D

Shi, F., Seo, B.-J., Cady, E., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Proc. SPIE, 106982O

Singh, G., Galicher, R., Baudoz, P., et al. 2019, A&A, 631, A106, doi: 10.1051/0004-6361/201936206

Smith, C. S., Marinică, R., den Dekker, A. J., et al. 2013, Journal of the Optical Society of America A, 30, 2002, doi: 10.1364/JOSAA.30.002002

Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, arXiv e-prints, arXiv:1305.5422. https://arxiv.org/abs/1305.5422

Tokovinin, A., & Heathcote, S. 2006, PASP, 118, doi: 10.1086/506972

English summary

Since antiquity, humanity haas hypothesized about the plurality of worlds and the existance of extraterrestrial life. It is only in recent years that we have started to answer this question. The first exoplanet, a planet that orbits stars other than our own Sun, was discovered in 1992. Since then, we have found thousands of others, from large gas giants more massive than Jupiter that orbit their star every few days, to planets that look remarkably close to Earth.

Many different methods have been developed over the years to find these exoplanets. Most of these methods are indirect and infer the existance of the planet by looking at the light from their host star. For detecting the chemical composition of the atmospheres of exoplanets, we however need to detect the light from the planet itself. Direct exoplanet imaging can play a big role in this effort. By spatially separating the light from the star and the planet, it is able to provide high-quality spectra of the light passing through the planets atmosphere. This enables us to search for the spectral signatures of their constituent gases of their atmospheres.

Direct exoplanet imaging is however not as simple as pointing a telescope at a star and taking an image. First, exoplanets are located extremely close to their star. Their angular separations range from one arcsecond to several milli-arcseconds, corresponding to a few tens to a few tenths of the Rayleigh limit, the fundamental resolving limit due to the wave nature of light, for current-generation observatories at the relevant wavelengths. Secondly, exoplanets are extremely faint. At optical wavelengths we are looking for the light reflected off the planet surface. At these wavelengths, Earth-like exoplanets are expected to have contrasts ranging from 10^{-9} to 10^{-11} with respect to their host star. Even at near-infrared wavelengths, where we instead see the thermal radiation, especially for young exoplanets which are stil hot from their formation process, gas giants still have contrasts of only 10^{-5} to 10^{-6}.

These challenges of direct imaging can be overcome with advanced instrumentation. A typical high-contrast imaging instrument employs three vectors of attack, all of which have to work together to produce the highest quality images.

1. **A coronagraph.** Coronagraphs are intricately designed optical devices that filter out starlight in our images. At the same time, the coronagraph must transmit as much of the planet light as possible. Coronagraphs should be compared based on their optical complexity, their level of starlight suppression, their planet throughput and their minimum angular separation where they achieve sufficient throughput, their robustness against small wavefront aberrations and their ability to efficiently measure wavefront aberrations.

2. **An adaptive optics or wavefront control system.** Polishing errors in the mirrors and lenses, vibrations and deformations in the optical system and, on ground-based telescopes, the turbulence in the atmosphere, distort the incoming light and produce a cloud of speckles in our coronagraphic image. We employ an adaptive optics or wavefront control system to stabilize the wavefront, which lets the coronagraph filter out the starlight. Typically, an adapitve optics system measures the distortion of the incoming light a few hundred to a few thousand times per second and uses a deformable mirror to induce an opposite and equal distortion to cancel it out.

3. **Image post-processing algorithms.** After the adaptive optics system and coronagraph, we are still left with a cloud of speckles in our images. This is the result of imperfections in these two systems. Advanced image post-processing algorithms exploit any redundant information in our images to filter out the residual starlight. This finally leaves us with a calibrated image of the stellar environment.

This thesis aims to further our knowledge of coronagraphs and their integration into high-contrast imaging instruments.

Chapter 2: optimal design of apodizing phase plate coronagraphs
An apodizing phase plate (APP) coronagraph consists of a single phase-only pupil-plane apodizer mask. Contrary to most other coronagraphs, the APP coronagraph does not filter out starlight completely, but only inside a region of interest in the coronagraphic focal plane, also known as the "dark zone". This chapter presents a new way of optimizing the phase pattern of the apodizer. This new algorithm is guaranteed to find the phase pattern with highest possible planet throughput for a given contrast, and dark zone and pupil geometry. This result provides us with the fundamental limits of this type of coronagraph.

Chapters 3 & 4: the Single-mode Complex Amplitude Refinement (SCAR) coronagraph The SCAR coronagraph combines a microlens-fed single-mode fiber array in the focal plane and an upstream pupil-plane phase-only apodizer mask. The mode filtering capabilities of the single-mode fibers significantly relaxes the phase pattern on the pupil mask. This makes the SCAR coronagraph reach much smaller angular separations compared to the APP coronagraph for similar values of the planet throughput. These two chapters cover the theory, simulations and a prototype laboratory demonstration that reached a raw contrast of 10^{-4} at an angular separation of just $1\lambda/D$.

Chapter 5: High-Contrast Imaging for Python (HCIPy) HCIPy is an open-source software package written in Python for simulating the interplay between wavefront control and coronagraphic systems. It aims to provide a modular object-oriented framework to enable rapid prototyping of the full high-contrast imaging system. HCIPy is currently used at multiple institutes around the world, both for research and for education.

Chapter 6: the asymmetric wind-driven halo The high-altitude jet stream layer in the atmosphere is notorious for producing a butterfly-shaped halo in high-contrast images. This halo is caused by the lag between the measurement and correction by the adaptive optics system. However, we also commonly observe an asymmetry in the halo, where one wing of the butterfly becomes brighter than the other. In this chapter, we identify the origin of this asymmetry as the interference between lagged phase speckles and scintillation speckles caused when light propagates from the high-altitude jet stream layer to the ground. This asymmetry can now be incorporated into the design of future high-contrast imaging instruments and potentially be removed with image post-processing techniques.

Chapters 7 & 8: the phase-apodized-pupil Lyot coronagraph The PAPLC combines a phase-only pupil-plane apodizer mask with a standard Lyot-style coronagraph architecture. In particular with a one-sided dark zone and a knife-edge focal-plane mask, this coronagraph provides inner working angles as close as $1.4\lambda/D$ at contrasts of 10^{-10} with a maximum post-coronagraph throughput of $> 75\%$. Furthermore, the light reflected off the focal-plane mask can be used for high-order wavefront sensing. Chapter 7 covers the design process and theoretical performance of this new coronagraph. Chapter 8 presents the first laboratory demonstration at the

Très Haute Dynamique 2 (THD2) testbed at the Observatoire de Paris. This demonstration reached a mean raw contrast of 1.9×10^{-8} from $2\lambda/D$ to $9\lambda/D$ in monochromatic light. Furthermore, we demonstrated wavefront sensing using the light reflected off the focal-plane mask at $\sim 3\times$ the fundamental wavefront sensing limit due to photon noise.

Future outlook While the first generation of high-contrast imaging instruments, VLT/SPHERE and Gemini/GPI, were our first attempt at combining extreme adaptive optics and coronagraphy, these are currently being upgraded with new coronagraphs and more advanced control systems. Additionally, new imagers such as Magellan/MagAO-X and Keck/KPIC are being built from the ground up to take full advantage of the developed new technologies. High-contrast imaging from space will see its first major test with the launch of the Roman Space Telescope. This will demonstrate high-order wavefront control with a coronagraph to raw contrast levels not achievable with ground-based imagers.

Over the next few years, we will start to see the first on-sky results of these new imagers. Application of the same technologies on future ground-based observatories, such as the ELT, GMT and TMT, will allow us to characterize rocky exoplanets around nearby lower-mass stars in the near and mid-infrared. Photometric monitoring of these planets can reveal their rotational periods and perhaps even their seasons and continents on their surface. Smaller-aperture space-based observatories will search for and characterize rocky planets around solar-type stars. Finally in a few decades, future large space-based observatories will have the angular resolution and sensitivity to characterize dozens of Earth-like exoplanets, enabling us to maybe answer whether there is life anywere else in the Universe.

Nederlandse samenvatting

Sinds de oudheid heeft de mensheid gefilosofeerd over de veelvoud der werelden en het bestaan van buitenaards leven. Alleen in de laatste jaren zijn we begonnen deze vraag te beantwoorden. De eerste exoplaneet, een planeet die om een ster draait anders dan onze eigen Zon, werd ontdekt in 1992. Sindsdien hebben we duizenden anderen gevonden, van gasreuzen, zwaarder dan Jupiter, die in een paar dagen om hun ster draaien, tot planeten die verdacht veel lijken op onze Aarde.

Verschillende methodes zijn in de afgelopen jaren ontwikkeld om deze planet te kunnen vinden. Veel van deze methods zijn indirect en leiden het bestaan van een planeet af door te kijken naar het licht van de ster waar die omheen draait. Om de chemische compositie van de atmosfeer van de planeten waar te kunnen nemen, moeten we echter kijken naar het licht van de planeet zelf. Het direct waarnemen van exoplaneten speelt hier een grote rol in. Door het licht van de ster en de planeet ruimtelijk van elkaar te onderscheiden, zijn we in staat om hoge kwaliteit spectra te nemen van het licht wat door de atmosfeer van de planeet heenschijnt. Hierdoor kunnen we zoeken naar de spectrale vingerafdruk van de bestandsdelen van de atmosfeer.

Het direct waarnemen van exoplaneten is echter niet zo simpel als een telescoop op een ster richten en een plaatje schieten. Ten eerste, exoplaneten staan zeer dicht bij hun ster. De hoekscheiding is typisch tussen de een boogseconde en een aantal milli-boogesconden, wat overeenkomt met een aantal tientallen tot een aantal tienden van de Rayleigh limiet, de fundamentele oplossingslimiet door de golfeigenschappen van licht, voor huidige observatoria en de relevante golflengtes. Ten tweede is het licht van exoplaneten zeer zwak. Op de golflengtes die wij kunnen zien, kijken we naar het licht wat reflecteerd vanaf het planeet oppervlak. Op deze golflengtes verwachten we dat aardachtige planeten een contrast hebben van tussen de 10^{-9} en de 10^{-11} ten opzichte van hun ster. Zelfs in het nabij-infrarood, wanneer we het thermisch licht bekijken van de exoplaneten die nog steeds warm zijn van hun formatie, hebben gasreuzen nog steeds maar een contrast van 10^{-5} tot 10^{-6}.

Deze uitdagingen kunnen we aan door gebruik te maken van geavanceerde instrumentatie. Een typisch hoog-contrast cameras op een telescoop

maakt gebruikt van drie verschillende aanvalsvectoren, welke allemaal samen moeten werken om de beste kwaliteit afbeeldingen te kunnen nemen.

1. **Een coronagraaf.** Coronagraven zijn gecompliceerde optische apparaten die sterlicht wegfilteren uit onze afbeeldingen. Tegelijkertijd moet een coronagraaf zo veel mogelijk licht van de planeet doorlaten. Coronagraven worden vergeleken op basis van hun optische complexiteit, de hoeveelheid sterlicht die ze wegfilteren, de hoeveelheid planeetlicht die ze doorlaten en de minimale hoekafstand waarop ze genoeg planeetlicht doorlaten, de robustheid ten aanzien van kleine golffrontfouten, en hun vermogen om efficient golffront telemetrie te kunnen meten.

2. **Een adaptief optisch systeem of golffrontcontrole systeem.** Polijstfouten in de spiegels en lenzen, trillingen en vervormingen in het optisch systeem en, voor telescopen op Aarde, de turbulentie in te atmosfeer zorgen ervoor dat het binnenkomt licht vervormd en wordt verspreid in een wolk van licht in onze coronagrafishe afbeelding. We maken gebruik van een adaptieve optisch of golffrontcontrole systeem om het golffront te stabilizeren, wat de coronagraaf in staat stelt om het sterlicht weg te filteren. Typisch meet een adaptief optisch systeem de vervorming van het golffront een aantal honderd tot een aantal duizend keer per seconde en gebruikt een vervormbare spiegel om een gelijke en tegenovergestelde vervorming teweeg te brengen om zo de vervorming op te heffen.

3. **Beeldverwerkingstechnieken.** Na het adaptief optisch systeem en de coronagraaf blijkt er nog steeds een wolk van licht op onze afbeelding te zitten. Dit is het resultaat van imperfecties in de twee hiervoorbeschreven system. Geavanceerde Beeldverwerkingstechnieken benutten de overbodige informatie in onze afbeeldingen om zo het overige sterlicht weg te filteren. Dit geeft ons uiteindelijk een gecalibreerde afbeelding van de omgeving van onze ster.

Dit proefschrift tracht bij te dragen aan onze kennis van coronagraven en hun integratie in hoog-contrast cameras op een telescoop.

Hoofdstuk 2: het optimaal ontwerpen van geapodizeerde fasemasker coronagraven Een geapodizeerd fasemasker (APP) coronagraaf bestaat uit een enkele apodizerend fasemasker in een pupil vlak. In tegenstelling tot de meeste andere coronagraven filtert de APP coronagraaf niet

het sterlicht compleet weg, maar alleen in een gebied van interesse, ook wel "donkere zone" genoemd. Dit hoofdstuk presenteert een nieuwe manier om het fasepatroon te vinden, die het meeste planeetlicht doorlaat, voor een gegeven contrastniveau en vorm van de donkere zone en telescoop pupil. Dit geeft de fundamentele limieten van dit type coronagraaf.

Hoofdstukken 3 & 4: de enkele-modus complexe amplitude zuivering (SCAR) coronagraaf De SCAR coronagraaf combineert een rangschikking van glasvezels, elk gevoed door een microlens, in het brandspuntvlak en fasemasker in een pupilvlak hiervoor. Het vermogen de de enkele-modus glasvezels om modi te kunnen filteren relaxeert het fasepatroon op het fasemasker. Dit laat de SCAR coronagraaf veel kleinere hoekafstanden bereiken vergeleken met de APP coronagraaf voor vergelijkbare waardes voor de transmissie van planeetlicht. Deze twee hoofdstukken beslaan de theorie, simulaties en een prototype lab demonstratie dat een contrast bereikte van 10^{-4} op een hoekafstand van slechts $1\lambda/D$.

Hoofdstuk 5: Hoog-Contrast Cameras voor Python (HCIPy) HCIPy is een open-source software pakket geschreven in Python om de wisselwerking tussen golffrontcontrole en coronagrafische systemen te kunnen simuleren. Het beoogt om een modulair object-georienteert kader aan te bieden, om het mogelijk te maken om snel prototypes te kunnen ontwikkelen van een compleet hoog-contrast camera systeem. HCIPy wordt momenteel gebruikt op meerdere instituten ter wereld, zowel voor onderzoek als voor onderwijs.

Hoofdstuk 6: de asymmetrische wind-gevormde halo De straalstroom die zich op grote hoogte in onze atmosfeer bevind, staat algemeen bekent als oorzaak voor een vlinderachige halo in coronagrafische afbeeldingen. Deze halo wordt veroorzaakt door de vertraging tussen de meting en de correctie door het adaptieve optisch systeem. We zien echter vaak een asymetrie in de halo, waar een vleugel helderder is dan de andere. In dit hoofdstuk identificeren we de oorsprong van deze asymmetrie als de interferentie tussen de vertraagde fase-aberraties en de scintillatie veroorzaakt door de propagatie van licht van de straalstroom tot de grond. Deze asymmetrie kan nu worden opgenomen in het ontwerp voor toekomstige hoog-contrast camera instrumenten en mogelijkerwijs kunnen worden verwijderd met beeldverwerkingstechnieken.

Hoofdstukken 7 & 8: de fase-geapodizeerd-pupil Lyot coronagraaf De PAPLC combineert een fasemasker in de pupil met een standaard Lyot-achtige coronagraaf architectuur. In het bijzonder met een enkelzijdige donkere zone een een mesvormig brandspuntvlakmasker, kan de coronagraaf minimale hoekafstanden bereiken van slechts $1.4\lambda/D$ met contrasten van 10^{-10} en met een planeettransmissie van $> 75\%$. Bovendien kan het licht wat op het brandspuntvlakmasker valt gebruikt worden voor het meten van golffrontfouten met veel vrijheidsgraden. Hoofdstuk 7 beslaat het ontwerproces en de theoretische eigenschappen van deze nieuwe coronagraaf. Hoofdstuk 8 presenteert de eerste lab demonstratie op de *Très Haute Dynamique 2* (THD2) proefbank bij de Observatoire de Paris. Deze demonstratie bereikte een gemiddeld contrast van 1.9×10^{-8} van $2\lambda/D$ tot $9\lambda/D$ in monochromatisch licht. Bovendien laten we golffrontmetingen zien, gebruikmakend van het licht wat gereflecteerd is vanaf het brandspuntvlakmasker, op ~ 3 maal het fundamentele limiet door fotonruis.

Een blik op de toekomst Hoewel de eerste generatie hoog-contrast camera instrumenten, VLT/SPHERE en Gemini/GPI, onze eerste poging was om extreme adaptieve optische systemen te combineren met coronagrafie, worden deze systemen momenteel geupgrade met de nieuwste coronagraven en meer geavanceerde controle systemen. Bovendien worden nieuwe systemen zoals Magellan/MagAO-X en Keck/KPIC vanaf de grond opgebouwd om zo volledig te kunnen profiteren van de nieuwste technologieën. Hoog-contrast cameras op ruimtetelescopen zullen hun eerste grote test ondergaan met de lancering van de Roman Space Telescope. Dit zal golffrontcontrole met een coronagraaf demonstreren tot op contrasten die niet haalbaar zijn voor telescopen op Aarde.

De komende paar jaar zullen we de eerste afbeeldingen van de hemel gaan zien, genomen door deze nieuwe instrumenten. Dezelfde technologie zal worden toegepast op toekomstige observatoria op Aarde, zoals de ELT, de GMT en de TMT. Deze instrumenten zullen het mogelijk maken om rotsachtige exoplaneten om lichtere sterren in het mid- en nabij-infrarood te karakteriseren. Door fotometrie te doen op deze planeten kunnen we de rotatieperiode en misschien ook seizoenen en continenten zien. Kleine ruimtetelescopen zullen zoeken naar rotsachtige planeten rond Zonachtige sterren en deze karakteriseren. Over een aantal decennia, zullen toekomstige enorme ruimtetelescopen de hoekresolutie en de gevoeligheid hebben om een aantal dozijn Aardachtige exoplaneten vinden, zodat we misschien de vraag kunnen beantwoorden of er ergens anders leven is in dit universum.

Curriculum Vitae

I was born on the 19th of November 1992 in Zoetermeer, the Netherlands. With both my mom and dad having studied physics, both specializing in computational physics, I came into contact with the natural sciences and programming early on. Throughout my youth, my interest in those primarily focused on computer graphics with ray and path tracing. From 2004 to 2010 I attended secondary gymnasium education at the Stedelijk College Zoetermeer, later renamed to the Picasso Lyceum, with a specialization in the natural sciences.

I started a double Bachelor in Physics and Astronomy at Leiden University in 2010. My interest in astronomy grew over the years, and only solidified after starting my Bachelor thesis project on sparse aperture masking under supervision of Matthew Kenworthy. This project turned out to be the perfect combination of simulation, observing and data reduction. I graduated cum laude in 2013.

In September 2013, I continued with a Master in Astronomy, specializing in Astronomical Instrumentation. My first research project was on post-Newtonian N-body dynamics supervised by Simon Portegies Zwart and Adrian Hamers. I graduated cum laude in 2016 with a research project on holographic electric field sensing for high-contrast imaging supervised by Christoph Keller, and I got the opportunity to present this work at a conference in Edinburgh, UK.

Afterwards, I continued at Leiden University with my PhD, supervised by Matthew Kenworthy and Christoph Keller. While the original goal was to improve post-coronagraphic wavefront sensing techniques, my interest was drawn towards coronagraphy. I developed an algorithm for finding the global optimum solution for apodizing phase plate coronagraphs, for the first time providing fundamental limits of this coronagraph. This algorithm turned out to be the basis for the development of the SCAR and PAPLC coronagraphs later on, and many side projects that did not make it to published articles.

My love for programming led me to develop HCIPy, a software package for simulating high-contrast imaging systems. While initially this package was intended to simplify code sharing and unify definitions within our research group, we decided to release HCIPy as open-source software and

keep maintaining and developing over the next few years. In these years, the number of users of HCIPy grew steadily and at the time of writing at least 12 researchers across the world use HCIPy for, at least part of, their daily research. Furthermore, HCIPy has been used in education at the Master and PhD level.

I presented my work at many conferences across the world, from the USA to Europe and Japan. I also visited the Space Telescope Science Institute in Baltimore, USA, for a period of six weeks, where I worked on exploiting symmetries and progressive refinement for the optimization of apodized-pupil Lyot coronagraphs and implementing wavefront control algorithms on HiCAT, their high-contrast imaging testbed.

In January 2021, I will be starting as an NHFP Sagan fellow at the Russell B. Makidon Optics Laboratory at the Space Telescope Science Institute in Baltimore, USA.

Acknowledgments

During my ten years at Leiden University – from the start of my Bachelor to the end of my PhD – I have met a large number of amazing people. This single section is too short to thank them all to the degree that they deserve, but I am grateful to have met you all.

First of all, let me thank my direct supervisors during my PhD, Matthew and Christoph. You let me roam free and explore whatever piqued my interest at the time, which I highly enjoyed. At times, you definitely had to keep me grounded and on track; I hope that this job became easier as time progressed. Frans, your crazy ideas and neverending enthousiasm made my time at Leiden much more dynamic.

Many thanks to the current and former members of the instrumentation and high-contrast imaging group: Sebastiaan, David, Steven, Maaike, Kelsey, Vikram, Alex, Mike, Mireille, Patrick, Olivier, Christian, Jos and Schuyler. I've had the privilege to travel with you to many new countries for conferences and holidays, commissioning new instrumentation and coronagraphs, observing a solar eclipse, celebrating New Years Eve on a mountain top, visiting ancient temples and many other amazing experiences. Your company made these so much more memorable. Thanks for coping with my incessant questions. Also, many thanks to Kirsty, Stijn, Leon and Leindert for providing happy distractions from instrumentation, and most of the time research altogether, during the frequent coffee, tea and lunch breaks.

Teaching has been something I never thought I would enjoy but actually did. Besides teaching at the Astronomical Observing Techniques tutorial sessions, I had the pleasure of supervising several excellent Bachelor and Master students. Thijs, Christian and John: I learned a lot from supervising you; I wish you all a great future career.

My visit to the Space Telescope Science Institute in Baltimore, USA, was one of the highlights of my PhD. Thank you, Rémi, Marshall, Laurent, Iva, Jamie, Kathryn, Jules, Keira, Scott, Pete, Greg, Anand and Chris, for providing a welcoming atmosphere during the six weeks of my visit and afterwards during the many remote group meetings. Thank you Katie, for letting me stay in your spare room on such short notice.

I would be amiss to not mention NYRIA, the Network of Young Re-

searchers in Instrumentation for Astronomy, for connecting me to many new people, in particular to Faustine, who showed me how quickly an idea can be turned into a scientific paper, chapter six in this thesis.

And last, but definitely not least, I would like to thank my family. To Mom, thanks for showing me that curiosity never stops. To my late dad, our time together was too short. You still inspire me to this day. To my sister, Anne, thanks for helping me get through a difficult time in my life. I could not have done it without you.