The handle http://hdl.handle.net/1887/44483 holds various files of this Leiden University dissertation

**Author:** Otten, Gilles  
**Title:** Suppressing a sea of starlight: enabling technology for the direct imaging of exoplanets  
**Issue Date:** 2016-11-29
Mankind has often philosophized about the existence of planets around other stars. It has been recorded as early as 300 BC when Epicurus argued, in a letter to his disciple Herodotus, that innumerable worlds must exist, both similar and dissimilar to ours (Laertius, 300). Later in the 17th century, Christiaan Huygens wrote in his philosophical treatise *Cosmotheoros* on the existence of planets around other stars, as well as seas, mountains, plants and animals on those planets\(^1\). Using our modern view we can tackle this problem more scientifically. We can formulate more structured questions such as; how do planets form?; is our solar system normal or special?; how unique is our earth? and finally, the ultimate question; is there life elsewhere in the universe? To answer these questions we need to discover and study planets around other stars.

Planets around other stars (extrasolar/exoplanets) were first discovered in 1992 when Wolszczan and Frail discovered two low mass exoplanets around the pulsar PSR1257+12 based on periodic variations of the pulsar’s pulse arrival times (Wolszczan and Frail, 1992). This long-expected confirmation proved that planets existed around other (albeit dead) stars. The hunt was on to find a planet around a solar-like star.

In 1995, Mayor and Queloz managed to detect radial velocity variations of the solar-type star 51 Pegasi that implied the presence of a planet of at least 0.47 Jupiter masses \(M_{\text{Jup}}\) (Mayor and Queloz, 1995), where there is a lower limit depending on the inclination of the orbit. In the following years more planets were discovered using the radial velocity (RV) technique (e.g. Marcy and Butler, 1996; Butler and Marcy, 1996) and it was evident that the first detected objects were mainly roughly Jupiter-mass planets at unexpectedly small Mercury-like orbits. The big question was how planets could be present at these locations. Quickly the realization came that the probability of such an object passing directly in front of its star as seen from Earth was about 10%. This transit would be seen as a small periodic decrease in the amount of light received from the parent star as the planet physically blocked part of the starlight.

The first planet detected using this *transit* method was HD 209458 b by Charbonneau et al. (2000); Henry et al. (2000), which was originally discovered with the radial velocity technique. This method allows for a relatively easy way of searching for exoplanets by monitoring the brightness of thousands of stars simultaneously. Within several years it has quickly become the most prolific planet detection method, especially with dedicated projects such as Kepler and SuperWASP. The detections of exoplanets through both RV and transit studies have raised questions on planet formation and migration (Udry et al., 2007), and has given us constraints on the occurrence of different types of planets (down to Earth radii) at different orbits around stars (Fressin et al., 2013).

Two other important techniques to discover exoplanets are microlensing (Gaudi, 2012) and astrometry. *Astrometry* is related to the RV method as it also uses the motion of the parent star to infer the presence of companions. The reflex motion of the star caused by the presence of a com-

\(^1\)http://www.staff.science.uu.nl/~gent0113/huygens/huygens_ct_en.htm
panion also has a transverse component that can be measured by accurately recording the position of the star in the plane of the sky. Previously, this technique was used to support several claimed detections of extrasolar planets. Most notoriously, the detection of two sub-Jupiter-mass planets around Barnard's star was presented by van de Kamp (1982) but later studies showed that these signals were spurious as planets have been excluded down to several Neptune masses (M. Kürster et al., 2003). A candidate was finally found with astrometry by Muterspaugh et al. (2010) orbiting in a binary system. This planet still has to be confirmed and it is unknown around which of the binary stars it orbits, thereby its properties are still ambiguous. A previously known exoplanet Gl876 b has also been confirmed with the Hubble Space Telescope (HST) through astrometry (Benedict et al., 2002). While this technique still has to discover its first official confirmed planet it is expected that the Gaia satellite, with its high astrometric precision, will detect thousands of Jupiter-mass objects within its lifetime (Perryman et al., 2014).

Microleising uses the effect that a massive object acts as a lens and distorts space-time around it. This lensing effect increases the amount of light that is received from a star behind the foreground star + planet combination. This brightening follows from Einstein's theory of general relativity and the magnification factor was consequently derived by Einstein (1936). Deviations from a simple lens can indicate the presence of a companion. The paper by Paczynski (1986) predicted that dark matter candidates could be found using this method by observing millions of stars for several years. Later, Mao and Paczynski (1991) wrote that exoplanets located at approximately the Einstein radius away from the star would be detectable in a similarly sized sample of stars. The first detection using the microlensing method was the 3 $M_{\text{Jup}}$ planet OGLE 2003-BLG-235 b/MOA 2003-BLG-53 b by Bond et al. (2004).

The current discoveries with the major detection methods, with known mass and separation, are shown in Fig. 1.1. This plot includes discoveries made by the direct imaging technique that will be described in Section 1.1. Each of the methods has their advantages and disadvantages (see Table 1.1). Pulsar timing has led to the discovery of planets in a handful of cases, as the planets need to survive stellar death or otherwise form around a dead star. Microlensing candidates can only be studied during a single statistically unlikely event. Astrometry is expected to become routine with the astrometric performance of Gaia. The radial velocity and transit methods are currently the most prolific methods but are mainly sensitive to planets close to the star. The direct imaging method described in this paper overcomes many of these biases.

The previously mentioned detection methods are indirect ways of determining properties of the exoplanets such as period, radius and mass. Periods are easy to determine and can be translated into orbital radii with an estimate of the mass of the star. The radius is determined by measuring the amount of blocked light during a transit. A lower limit to the mass can be derived through radial velocity measurements. When combined with the radius of transiting planets this yields the density and therefore can give an indication of the bulk composition of the planet. Several methods are also sensitive to the properties of the planetary atmosphere, like wind speed (Snellen et al., 2010) and chemical abundances (Seager and Deming, 2010). For instance, the transit method can detect atmospheric features imprinted on the stellar light (Charbonneau et al., 2002) when the planet passes directly in front of its star. The presence of a spectral feature will appear as an apparent increase in the amount of blocked light and therefore derived radius of the planet. In the case of relatively hot planets, the thermal emission of the day-side of the planet can even be recorded during the secondary eclipse when the planet hides behind the star. The secondary eclipse probes the thermal radiation of the planet at the height in the atmosphere where it becomes
Figure 1.1: Mass versus separation plot for currently known extrasolar planets. This plot also includes some candidates that span the transition between planets and brown dwarfs. The official IAU definition of a planet has a cut-off at 13 $M_{\text{Jup}}$. Source: www.exoplanet.eu (October 2015) and known solar-system values.

optically thin. Different wavelengths therefore probe different heights in the atmosphere and can be used to constrain atmospheric temperature/pressure profiles. Using the radial velocity method at high spectral resolution ($R \sim 100\,000$) it is also possible to probe the atmosphere by stacking the resolved spectral features of the planet using an atmospheric template (Snellen et al., 2010). The rotation rate of the known planet Beta Pictoris b was also determined with this technique (Snellen et al., 2014).

1.1 The direct imaging of exoplanets

The previously mentioned detection methods mostly depend on information derived from the light of the star and are therefore indirect ways of probing exoplanets. With direct imaging, the light of the planet itself is detected spatially resolved from the star, yielding a direct measurement of the thermal and/or reflected light of a planet. With enough spectral resolution and signal-to-noise, determination of atmospheric composition is possible. This makes it a very powerful technique for directly studying the atmospheres of planets. Furthermore, direct imaging does not require a full orbit of a planet around its star to confirm it is real.

The technique of direct imaging is biased towards planets at larger separations from the star (5-1000 AU). This complements the other methods that are more sensitive to planets that are close
Table 1.1: Main exoplanet detection methods and their strengths and weaknesses.

<table>
<thead>
<tr>
<th>Method</th>
<th>Strength</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radial velocity</td>
<td>gives lower limit of mass, confirm transiting planets, atmospheric studies using high spectral resolution, sensitive to planets close to stars</td>
<td>sensitive to stellar oscillations, stellar type, magnetic activity and multiplicity, long periods require extensive monitoring</td>
</tr>
<tr>
<td>Transits</td>
<td>gives radius estimate, density of planets when combined with RV mass limits, atmospheric studies possible through spectroscopy or spectrophotometry, can probe Earth-like planets in habitable zone around select targets</td>
<td>inclination has to be close to 90 degrees, only measures relative radius, long periods require extensive monitoring, relatively short events that contain most information</td>
</tr>
<tr>
<td>Pulsar timing</td>
<td>sensitive down to small masses</td>
<td>dead systems, very rare occurrence</td>
</tr>
<tr>
<td>Microlensing</td>
<td>sensitive to Earth-like planets at several AU separation</td>
<td>one-time events, statistically unlikely, degeneracies</td>
</tr>
<tr>
<td>Astrometry</td>
<td>hardly depends on stellar type</td>
<td>requires well controlled systematics</td>
</tr>
<tr>
<td>Direct imaging</td>
<td>directly measures thermal and/or reflected planet light, does not require long baselines to confirm planet, allows direct atmospheric characterization</td>
<td>biased towards large, young planets at large separations, probes interesting (solar-system scale) separations only for nearest stars</td>
</tr>
</tbody>
</table>

...to their star. Improvements in all detection methods will lead to overlapping areas of sensitivity that can be used to scrutinize the existing planet formation models and to further constrain the properties of the planet.

The first claimed directly imaged planet was 2M1207 b, a giant planet orbiting a brown dwarf (Chauvin et al., 2004); see Fig. 1.2a. Direct detections of other low mass companions have since followed (e.g., 1RXS J1609 b (Lafrenière et al., 2008), Fomalhaut b (Kalas et al., 2008), Beta Pictoris b (Lagrange et al., 2009), GJ 504 b (Kuzuhara et al., 2013), GU Piscium b (Naud et al., 2014), HD 106906 b (Bailey et al., 2014)). The list of detections even includes a planetary system of giant planets around HR8799, discovered by Marois et al. (2008, 2010); see Fig. 1.2b. Recently, the detection of 51 Eridani b was presented by Macintosh et al. (2015) that, depending on the formation history, may have a mass as low as $2 M_{\text{Jup}}$. 
1.1 The direct imaging of exoplanets

(a) 2M1207 b was the first directly imaged planet and is located around a brown dwarf. Discovered by Chauvin et al. (2004).

(b) HR 8799 is a directly imaged planetary system containing 4 giant planets. The ‘b’, ‘c’, ‘d’ planets were discovered by Marois et al. (2008) and the ‘e’ planet by Marois et al. (2010).

(c) 51 Eridani b is the least massive directly imaged planet to date. Assuming a hot-start formation model it may be as massive as $2M_{Jup}$. Discovered by Macintosh et al. (2015).

Figure 1.2: A selection of famous directly imaged planets.

A major challenge in the direct imaging of planets is that the star is many orders of magnitude brighter than the planet. A planet is mainly detectable either through reflected or scattered starlight (in visible to near-infrared range), or by thermal radiation from the planet itself (in near-infrared to mid-infrared range). To show the order of magnitude of this problem; the ratio between the surface area of our Earth to that of the Sun is approximately $10^4$. Combined with the average surface temperature of the Sun (5777K) and the Earth (287K) this means that the bolometric flux ratio between the Earth and the Sun is about $10^{-8}$ to $10^{-9}$, although at select wavelengths the ratio becomes more favorable. The reflected light in the visible wavelength range from an Earth-like planet around a sunlike star is $10^{-10}$. A Jupiter-like planet around a Sun-like star will give a bolometric flux ratio of $10^{-7}$. The reflected light of Jupiter-like planets in the visible range is even fainter at $10^{-8}$ (Traub and Oppenheimer, 2010).

This immense difference in brightness is often compared to trying to see a firefly next to a lighthouse (resp., 1/500 Candela and 2 million Candela or a factor of $10^{-9}$). Without tackling the light of star, the planet will be an indistinguishably small bump next to the star, sitting in a sea of starlight. The origin of this sea of starlight will be explained in more detail later.

There are more factors that limit the detection of planets in direct imaging (e.g., atmospheric turbulence, instrumental wavefront changes, these will be explained in Section 1.2) and so current detections have been mostly limited to self-luminescent giant planets orbiting around young stars at relatively large separations. As the thermal emission of self-luminescent young planets peaks in the near-infrared it is logical that most of the previous planet detections were initially made in that wavelength regime. A spectrum can be taken of these directly imaged planets to detect atmospheric signals (e.g., Mohanty et al. (2007); Janson et al. (2010); Barman et al. (2011); Konopacky et al. (2013)). In the spectroscopic study of HR8799 c Konopacky et al. (2013) found signatures of both CO and H$_2$O and a C/O ratio higher than that of the parent star suggesting this planet formed using core accretion, in which a giant planet forms around a rocky core by rapidly accreting gas.

Even in the case of a non-detection, limits can be placed on the absence of companions down
to a certain magnitude of contrast. *Contrast* is defined as the brightness of a target with respect to its parent star at which it can detected with a certain signal to noise ratio (typically $S/N = 5$ or equivalently a $5\sigma$ detection). Sometimes the term contrast is used when talking about normalized intensity profiles. The recent publication by Vigan et al. (2015) produced one of the highest contrasts measured to date by imaging Sirius with the newest generation of planet-finding instruments (VLT/SPHERE); see Fig. 1.3. While it did not find a new companion the study excludes their presence down to several $M_{\text{Jup}}$ (equivalent to a $5\sigma$ contrast of $10^{-6.5}$ at 0.5 arcseconds in $Y J H$-bands).

![Contrast curve of the detection limit reached by direct imaging observations of Sirius. From Vigan et al. (2015).](image)

**Figure 1.3:** Contrast curve of the detection limit reached by direct imaging observations of Sirius. From Vigan et al. (2015).

To interpret the spectra and images many models exist, such as AMES-COND (Allard et al., 2001; Baraffe et al., 2003), AMES-DUSTY (Allard et al., 2001; Chabrier et al., 2000), BT-SETTL (Allard et al., 2012) and models by Burrows et al. (2004); Marley et al. (2007); Fortney et al. (2008); Morley et al. (2012); Morley et al. (2014a,b) amongst others. These models produce spectra dependent on the cooling history, mass, age, and metallicity, amongst other variables. By comparison with these models, multi-wavelength studies and spectra through direct imaging can constrain the planet’s temperature, mass and atmospheric properties without depending on the completion of several orbits around the star as is the case with radial velocity studies. Furthermore it is possible to test the formation mechanism of the planet (disk instability versus core accretion).

In Fig. 1.4 we show the expected contrast ratio based on the thermal radiation between an F-type star and a planet with varying temperature. It is easily seen that observing in the near-infrared and especially at $4 - 5$ micron is easier in terms of contrast and can probe lower temperatures and
less massive planets for a given contrast ratio. From the presence of spectral features we can infer atmospheric composition and constrain the formation history of the planet.

![Flux ratio between planets and F-type stars based on the BT-SETTL models.](image)

**Figure 1.4:** Flux ratio between planets and F-type stars based on the BT-SETTL models. Favorable contrast can be found in the near-infrared, especially in $L$ and $M$ band. Outside of the grey filter bands the Earth’s atmosphere is generally opaque due to water bands. Visible band filters are not shown but the atmosphere is mostly translucent between 350 nm and 1 micron.

## 1.2 Challenges in direct imaging

### 1.2.1 The diffraction limit & Adaptive Optics

Imaging point sources with a telescope with finite size results in images in which the point sources look like a central broad spot with rings surrounding it (*Point Spread Function* or PSF). This is due to a fundamental limit in spatial resolution due the wave-like nature of light and the aperture of the telescope. A circular telescope with diameter $D$, observing at wavelength $\lambda$ can only separate objects that are more than $1.22 \frac{\lambda}{D}$ apart (related to the width of the central spot); this is the Rayleigh Criterion. In the case of larger sources the finite aperture of the telescope produces images of the object plane that are blurred with this scale. The smallest spatial details are therefore lost. Improving the spatial resolution by increasing $D$ is one of the main driving forces for building increasingly larger telescopes. The increase in diameter allows us to take sharper images. A nearly perfect PSF is shown in the top center frame of Fig. 1.5.

The diffraction limit is easily reached from space as there are fewer sources of optical aberrations. Observing from the ground is a different story, as our atmosphere is turbulent and produces a wavefront that varies in both time and space. Light at different parts of the pupil arrive at different times. This distorted wavefront creates an extended and time varying PSF that, when observed over time scales of a second, averages to a seeing disk with a width of approximately $\lambda/r_0$, where $r_0$ (Fried parameter) is the typical size of the turbulence ($\sim 15 \text{ – } 30 \text{ cm}$ in the optical range at...
Introduction

Figure 1.5: Top row shows simulated instantaneous monochromatic PSFs with and without Adaptive Optics (AO) and the typical spatial scales associated with them. The rightmost image shows the post-AO PSF with the ideal PSF subtracted (the residuals). The intensity for all PSFs is normalized to the peak of the ideal PSF. The residuals are atmospheric speckles, shown with a range of $\pm10^{-2}$. The speckles are dominant along the horizontal wind direction. The pupil wavefront maps associated with the pre-AO and post-AO PSFs are seen below their respective PSFs in units of radians.

While space-based direct imaging has many advantages, the biggest disadvantages are that it is expensive ($\sim 1$ billion dollars) and the diameter of the telescope that can be launched is limited. On the ground, segmented mirrors can be built that are almost an order of magnitude larger in diameter and instruments can be easily serviced and upgraded. This comparison between ground-based and space-based direct imaging is also made in Table 1.2. In this thesis we focus on improvements in the ground-based direct detection of exoplanets to optimally use the high spatial resolution and light collecting area of the largest telescopes.

Adaptive Optics (AO) is the enabling factor that allows us to reach the diffraction limit from the ground for telescopes larger than $r_0$. Current $6 - 10$ meter class telescopes can reach solar-system scales around nearby stars when operating at the diffraction limit. Adaptive Optics actively measures the shape of the wavefront as it arrives at the wavefront sensor (WFS) and flattens the wavefront using a deformable mirror (DM) on timescales of milliseconds. It is important to note that a perfect PSF can not be fundamentally reached from the ground. An important indicator used for the post-AO image quality is Strehl Ratio (SR). The Strehl ratio is defined as the intensity of the PSF peak with respect to a theoretically perfect PSF. A simple closed-loop AO setup is seen in Fig. 1.6 with simulated PSFs with and without AO in the two left panels of Fig. 1.5.

When comparing the signal to noise ratio $S/N$ (defined as the integrated flux within a certain
1.2 Challenges in direct imaging

Table 1.2: Comparison between direct imaging from space and ground.

<table>
<thead>
<tr>
<th></th>
<th>advantage</th>
<th>disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>space-based</td>
<td>$10^{-5}$ raw contrast, no atmospheric aberrations, no transmission windows, relatively static diffraction limited PSF</td>
<td>limited diameter, hard to service/repair/upgrade, order of magnitude more expensive, limited lifetime</td>
</tr>
<tr>
<td>ground-based</td>
<td>easy to upgrade/repair/service, largest diameter technology can support</td>
<td>$10^{-5}$ raw contrast, can only image in atmospheric transmission windows, seeing limited unless AO-capable</td>
</tr>
</tbody>
</table>

Figure 1.6: Schematic overview of typical closed-loop adaptive optics setup. The wavefront sensor measures the wavefront at a high rate (on the order of kHz) and sends commands to the deformable mirror to flatten the incoming wavefront. Note that the WFS can not see the wavefront errors in the branch towards the science camera. Image courtesy of Claire Max.

radius from the star and the variation of this flux measured across realizations) of a point source for diffraction limited telescopes versus seeing-limited telescopes we see that the $S/N$ scales as respectively $D^2$ versus $D$. The amount of collected light remains the same but it is concentrated into an area that is smaller. Merely switching on the AO system significantly increases the $S/N$ when imaging point sources from the ground. In the case of direct imaging, where faint targets sit in a variable halo of starlight with total intensity proportional to $1 - \text{Strehl Ratio}$, the improvement is even greater.
1.2.2 Seeing and speckles

An important source of noise in direct imaging is the presence of variable speckles (Racine et al., 1999; Hinkley et al., 2007). Speckles are planet-mimicking features that arise from uncorrected wavefront aberrations in the atmosphere and/or optical system. Speckles appear as modulations of intensity in the static diffraction structure of the PSF. These modulations are an important source of noise in direct imaging. Any planet is literally sitting in a sea of starlight. The rightmost PSF in Fig. 1.5 shows the residuals of a post-AO PSF after subtracting the ideal PSF from it. The point-like features at a level of $10^{-2}$ and lower are the speckles with a typical size on the order of $\lambda/D$.

Atmospheric speckles

The Adaptive Optics system needs a finite time to respond and can only correct down to certain spatial scales dependent on the amount of actuators across the telescope pupil. This correction needs to be done on time scales of the order of $\tau_0 \sim \frac{r_0}{V_{\text{wind}}}$, where $r_0$ is coherence length or Fried parameter of the turbulence and $V_{\text{wind}}$ is the wind speed. This means that the correction is not perfectly matched to the incoming distorted wavefront. This mismatch creates a wavefront that shows small residual variations, on small spatial scales. When looking at the PSF this means that a time varying cloud of $\lambda/D$-sized speckles remain and modulate the PSF. The variation of these speckles is a source of noise and, in many cases, larger than the photon noise contribution and consequently the most important limiting factor in detecting planets close to the star. The amount of variation is larger on top of the Airy rings which can be seen as pinned speckles. The limited amount of actuators also creates a control region with a radius $\frac{NA}{2D}$, where $N$ is the amount of actuators across the pupil. Outside of this radius AO correction will not be effective (see Fig. 1.7). An added effect of the finite response time combined with a dominant wind speed and direction creates a large uncorrected wavefront at the edge of the telescope. When looking at the PSF the unsensed edge projects a pronounced halo of speckles in the direction towards and away from the
Another important limiting factor is the speckles created by uncorrected instrumental wavefront aberrations that slowly evolve in time. The cause of these quasi-static aberrations can be found in mechanical changes in the telescope and instrument due to factors such as flexing and thermal, pressure and gravity vector variations (Soummer et al., 2007). The wavefront sensor cannot correct for these aberrations as they occur in the (non-common path) science arm of the instrument. This non-commonality can be seen in Fig. 1.6. These speckles also create a noise floor unless they are taken care of.

### 1.3 Contrast enhancing techniques

In order to get rid of the light of the star and the corresponding speckles while still minimizing the impact on the light of the planet, both optical and data-reduction techniques have been developed. We start with the data-reduction techniques.

#### 1.3.1 PSF subtraction

One of the simplest ways to get rid of the starlight and some of the slowly varying speckles is by imaging a second star without a known companion (Reference Differential Imaging; RDI, Smith and Terrile (1984)). This second star can be used as a reference to subtract the starlight of the first object while keeping the planet intact. The reference object is not imaged simultaneously but before, after, or interleaved with the observations of the primary target. Due to the difference in time and telescope motions between the targets, wavefront changes build up that limit how well the starlight is subtracted. By using a local reference in both time and space this effect can be mitigated.

**Simultaneous / Spectral Differential Imaging** (SDI) is another way of removing the stellar light. The spectrum of a star and a planet have different absorption features and this can be leveraged to remove the starlight (Rosenthal et al., 1996; Racine et al., 1999; Marois et al., 2003; Lenzen et al., 2004). A large planet survey of 45 young and nearby stars (Biller et al., 2007) used this technique and specifically looked for Methane (CH$_4$) absorption on giant exoplanets following the predictions of Baraffe et al. (2003) for field brown dwarfs. The surprising lack of detections at this band showed that these massive planets do not follow the behavior of field brown dwarfs at cooler temperatures. More recently, successful SDI detections have been made at the Hydrogen $\alpha$ line (Close et al., 2014). This is a sign of the accretion of Hydrogen gas onto a forming planet.

With the **Spectral Deconvolution** technique (SD, Sparks and Ford, 2002) the spectral signature can also be used to effectively remove speckles as their angular distance from the star scales with wavelength while the distance of the planet to the star is fixed. This approach is taken as an extra correction step by the newest generation of dedicated planet finders (SPHERE, GPI), that contain Integral Field Units (IFUs) that carve up the image into tiny spectra (e.g., Mesa et al., 2015).

A third spectral diversity technique that can be used to improve the contrast while directly imaging planets is by Snellen et al. (2014) and uses the radial velocity shift of a planetary signature...
at high spectral resolution ($R > 100,000$) in order to resolve individual spectral lines. The signature of the planetary atmosphere cannot be seen in an individual exposure. The data is therefore cross-correlated with template spectra. In the stacked data the atmospheric feature is above the noise floor and can be measured.

**Polarization Differential Imaging (PDI)** uses the fact that starlight is unpolarized while light reflecting off a planet is polarized up to a large fraction (Stam, 2003). By splitting the light according to their linear polarizations and taking the difference, the unpolarized starlight is canceled out while the light of the planet only partially cancels. This approach is often used by disk scientists to study debris disks around stars (Kuhn et al., 2001; Avenhaus et al., 2014). In the future it is expected that PDI will improve the detection of reflected light planets (Keller et al., 2010; Roelfsema et al., 2014).

**Angular Differential Imaging (ADI)** is a powerful technique that can be used to remove starlight and speckles without making modifications to a telescope (Marois et al., 2006). The angle of the instrument with respect to the telescope is fixed as much as possible and the sky (and therefore the planet) is let to rotate on the detector. As quasi-static speckles are associated with the telescope they remain at the same angles on the detector. This creates diversity that allows us to discriminate between (quasi-)static features and planets. By removing the median of the observations the starlight, and diffraction features and speckles associated with it are canceled while the rotating planet is minimally impacted. By derotating the images and stacking them the signal of a potential planet becomes even more significant. A schematic demonstration of ADI can be seen in Fig. 1.8.

**Figure 1.8:** Schematic representation of ADI adapted from an image by Christian Thalmann (Roelfsema et al., 2014). A cube of images is taken at fixed parallactic angle. This means that the planet will seem to orbit around the star while features caused by the telescope and instrument are static (indicated above with white and black stripes/arc). A median “C” can be subtracted to reduce the static features and afterwards the images are individually derotated and stacked to reveal the planet.
1.3 Contrast enhancing techniques

Variations of ADI that take time variations into account have been developed (e.g., LOCI, PCA). LOCI stands for \textit{Locally Optimized Combination of Images} and was developed by Lafrenière et al. (2007). Variants of LOCI also exist (Marois et al., 2014; Wahhaj et al., 2015). The technique subdivides all frames into annular wedges and for every subregion optimally produces a reference by linearly combining the corresponding subregions from a PSF library. This library can also contain frames of the observations themselves. The optimal combination is made by performing a least-square fit to minimize the residual after subtraction. Note that in case the PSF library contains the target frames themselves, frames very close in time need to be excluded from the fit in order to minimize self-subtraction of the planet signal. ADI and LOCI made it possible to recover the planet HR8799 b in archival data taken with the HST in 1998 (Lafrenière et al., 2009); ten years before it was eventually discovered by Marois et al. (2008), also using ADI.

\textit{Principal Component Analysis} (PCA) has also been recently used to improve PSF subtraction (Amara and Quanz, 2012; Soummer et al., 2012). PCA extracts principal components out of a PSF reference library using singular value decomposition or by projecting on a Karhunen-Loève basis (KLIP algorithm). A linear combination of a limited amount of orthogonal basis components is then fitted and subtracted from each of the frames. The PCA basis construction, fitting and subtraction can also be performed on subsections of each of the frames. This was shown by Meshkat et al. (2014) to improve the contrast with respect to LOCI and a global optimization with PCA. While self-subtraction is a concern when the principal components are constructed from the target dataset, Meshkat et al. (2014) also showed that, in their case, including all frames actually improved the $S/N$ the most.

1.3.2 Coronagraphs

An additional approach to improve the contrast can be taken by suppressing the starlight before it is captured by the detector. The reduction of starlight reduces the strength of the $S/N$-degrading speckles. This can be done using a \textit{coronagraph}. These were originally developed by Bernard Lyot in 1939 for masking the disk of the sun in order to image the faint solar corona (hence the name) without having to wait for solar eclipses (Lyot, 1939). Coronagraphs not only improve the contrast between the star and the planet, by suppressing the intensity of the diffraction structure the noise contribution of the previously mentioned speckles is also reduced. At least four important quantities are used when talking about coronagraph performance: throughput, inner working angle, contrast and chromaticity. The \textit{inner working angle} is the smallest angle at which a planet can be detected with $> 50\%$ throughput. A small inner working angle helps with optimally using the diffraction limit offered by the telescope. The \textit{contrast} is the flux ratio with respect to the star for a significant point source companion detection (typically used is $5 \sigma$ threshold). Higher contrast allows to image smaller, cooler and older planets. Finally the wavelength dependence or \textit{chromaticity} of the coronagraph is important for broad-band imaging or spectroscopy. Due to factors outside of our control such as scintillation and atmospheric chromaticity\footnote{http://www.naoj.org/staff/guyon/04research.web/14hzplanetsELTs.web/content.html} a raw contrast limit exists for 30 meter class telescopes that is on the order of $10^{-6}$. This means that the raw contrast of a coronagraph operating from the ground does not have to be better than $10^{-6}$. Another important aspect to consider about coronagraphs is whether they can handle complicated/obstructed pupils.
Focal-plane coronagraphs

The conceptually easiest type of coronagraph is the one that Bernard Lyot himself devised and uses an additional pupil and focal plane in the instrument. The Lyot (or focal-plane) coronagraph physically blocks the star in the additional focal plane with a small amplitude mask to get rid of the starlight. A Lyot stop sitting in the additional pupil plane cuts off the light at the edge of the pupil that diffracted around the focal-plane mask that would otherwise impact the contrast. Figure 1.9 shows a schematic representation of an optical setup with a focal-plane coronagraph.

Since the first Lyot coronagraph, many types of focal-plane coronagraphs have been developed with either amplitude, phase or hybrid masks and stops with improved performance in terms of contrast, inner working angle and bandwidth. The original Lyot coronagraph could only improve in contrast by sacrificing inner working angle by making a larger blocking mask. By adjusting the optical elements in different planes of the instrument improvements can be made. For instance, the amplitude focal-plane mask can be altered with the Band-limited Lyot Coronagraph (Kuchner and Traub, 2002). The first pupil plane can also be used to preshape the PSF for higher contrast. This is the principle behind the Apodized Lyot Coronagraph (Soummer et al., 2003a) that uses an amplitude mask with an increasingly less transmissive edge to reduce the strength of the Airy rings. The Phase Induced Amplitude Apodization (PIAA) design achieves the same amplitude apodization but with no transmission loss (Guyon, 2003) and can be used with and without focal-plane mask.

By using a focal-plane phase mask with phase discontinuities to produce an amplitude singularity the on-axis stellar light can be diffracted to the edge or even outside of the Lyot pupil. In combination with a Lyot stop this removes the starlight from the optical system. The phase-based focal-plane mask improves the throughput and inner working angle with respect to the amplitude-based focal-plane masks. Beside radial phase designs like the Disk Phase Mask (Roddier and Roddier, 1997) and Dual Zone Phase Mask (Soummer et al., 2003b) azimuthal models like the Four-Quadrant Phase Mask (FQPM, Rouan et al., 2000) and 8 Octant Phase Mask (8OPM, Murakami et al., 2008) were developed to improve the chromatic behavior, contrast and IWA even further. Both the FQPM and 8OPM suffer from the problem that when a planet is crossing the phase discontinuities, the flux of the planet is also suppressed, leading to dead zones. The Annular Groove Phase Mask and the optical Vector Vortex Coronagraph use a smooth azimuthally varying phase gradient to overcome this effect and has been analytically proven to
(in theory) produce perfect rejection of on-axis light with a critically sized Lyot mask (Mawet et al., 2005). The Hybrid Lyot coronagraph uses a complex (both amplitude and phase) focal-plane mask to produce raw contrast in excess of $10^{-9}$ for space missions (Moody et al., 2008). An extension of the PIAA coronagraph (PIAACMC) also has a complex focal-plane mask to push the performance even closer to the theoretical limit of coronagraphs (Guyon et al., 2010).

The biggest issue that still remains with even the most advanced focal-plane coronagraphs is that they are sensitive to tip/tilt misalignments and vibrations. Vibrations are practically unavoidable as they are caused by crucial components related to the functioning of the telescope and instruments. Instruments need to be cooled using vacuum pumps to cryogenic temperatures, and wind pushing against the structure of the telescope can excite resonances in the telescope that can be seen as tip/tilt motion. Only with significant investments (such as a high accuracy multiple kHz tip/tilt correction stage) these vibrations no longer limit the performance of the focal-plane coronagraphs. These problems will only get worse when we move to even larger telescopes such as the European Extremely Large Telescope (E-ELT).

**Pupil-plane coronagraphs**

It is also possible to leave the core of the starlight largely intact and focus on suppressing starlight close to the core at a location where potential planets may occur. A phase and/or amplitude modification can be made to the light at the pupil of the telescope with a *pupil-plane coronagraph*. As the pupil-plane and focal-plane are connected through a Fourier transform, the pupil modification can be chosen in such a way that it alters the PSF of the telescope by highly suppressing the diffraction rings in a predefined region. The amount of starlight at the location of a potential planet will be reduced. See Fig. 1.10 for a schematic representation of the optical setup for such a type of coronagraph. Most importantly, by being located in the pupil of the telescope, the PSF is identical across the whole image plane and the contrast is therefore inherently insensitive to tip/tilt offsets and vibrations. All sources will be affected equally, including any planets. This means that pupil-plane coronagraphs also work for multiple stars. By choosing a modification that largely conserves the peak flux, the impact to the planet signal is minimized. The pupil modification can be either amplitude variations, phase variations or both. One example of pupil-plane coronagraphs are the shaped pupils of Kasdin et al. (2003); Kasdin et al. (2004) that use black and white regions of transmission that yields point-symmetric PSFs. Concepts using only phase were pioneered by (Yang and Kostinski, 2004; Kostinski and Yang, 2005) and perfected with the Apodizing Phase Plate (APP) (Codona and Angel, 2004; Codona et al., 2006; Kenworthy et al., 2007). Other important developments in phase designs have been made by Carlotti (2013); Ruane et al. (2015). These phase-only solutions allow one-sided dark holes near the PSFs. An important tradeoff for pupil-plane coronagraphs is the size of the suppressed PSF region, the required contrast and the effective throughput of the plate.

Coronagraphs have different advantages and disadvantages depending on their design. A comparison between focal-plane and pupil-plane coronagraphs is made in Table 1.3. The main advantage of the focal-plane coronagraphs is their theoretically high contrast performance. They are limited by vibrations and occupy multiple planes in an instrument which complicates the installation. Pupil-plane coronagraphs are insensitive to vibrations and only occupy one plane in a telescope. This greatly facilitates the installation into instruments.

A careful review of the then-known coronagraph designs by Guyon et al. (2006) showed
that many designs of coronographs are sensitive to the finite size of the star. The impact of this effect was generally ignored in coronagraphic simulations to simplify the calculations. Part of the coronagraph designs have to be excluded when considering this effect. The impact will only become more significant as the diffraction limit of the future ground-based telescopes continues to approach the angular sizes of nearby stars.

Many designs of coronographs also have an inner working angle that is relatively large \((3 - 4 \frac{\lambda}{D})\) and therefore do not optimally use the capabilities of the telescope in terms of angular resolution. A subset of coronagraph designs with inner working angles of \(1 - 2 \frac{\lambda}{D}\) was summarized by Mawet et al. (2012) and distilled into multiple families of coronagraphs as seen in Fig. 1.11. The smaller inner working angle of these designs can be used to image planets closer to the star or to expand the study of planets towards stars that are farther away. Small inner working angle coronographs enable us to use the angular resolution of the largest telescopes to the fullest extend.

### Apodizing Phase Plate

The **Apodizing Phase Plate** (APP) coronagraph is designed as a flux conserving, vibration insensitive pupil-plane coronagraph with small inner working angle (Codona et al., 2006; Kenworthy et al., 2007). Therefore, it combines the qualities needed to create a robust but powerful coronagraph. It uses a phase-only modification to the light in the pupil of the telescope. In the original APP the phase delay is created by physically varying the thickness \(\Delta x\) of a piece of high refractive index \((n_{\text{plate}})\) glass across the pupil, embedded in a medium with index \(n_{\text{air}}\). The physical difference in distance traveled through the plate gives a varying phase delay \(\Delta \phi\) to the light with wavelength \(\lambda\) (see Fig. 1.12 and Eq. 1.1).

\[
\Delta \phi = 2\pi \frac{\Delta \lambda}{\lambda} = 2\pi(n_{\text{plate}} - n_{\text{air}}) \frac{\Delta x}{\lambda}
\]  

Unfortunately this phase offset is chromatic, as light of different wavelengths will see a different phase offset. Deviating from the optimal wavelength generally degrades the contrast and limits the wavelength range at which the coronagraph can be used. The phase delay is typically diamond-turned into a substrate with high refractive index. The diamond-turning limits the spatial frequencies of the phase pattern. Discontinuous phase transitions are nearly impossible to
Table 1.3: types of coronagraphs and their (dis)advantages with examples. ALC: Apodized Lyot Coronagraph, DPM: Disk Phase Mask, DZPM: Dual Zone Phase Mask, FQPM: Four-Quadrant Phase Mask, 8OPM: Eight-Octant Phase Mask, OVC: Optical Vortex Coronagraph, VVC: Vector Vortex Coronagraph APP: Apodizing Phase Plate, PPA: Pupil Phase Apodization.

<table>
<thead>
<tr>
<th></th>
<th>amplitude phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>focal plane</td>
<td>classical Lyot, ALC, band-limited, hybrid Lyot</td>
</tr>
<tr>
<td></td>
<td>theoretical high contrast</td>
</tr>
<tr>
<td></td>
<td>sensitive to vibrations,</td>
</tr>
<tr>
<td></td>
<td>occupies multiple telescope planes, hard to</td>
</tr>
<tr>
<td></td>
<td>align, stop blocks throughput and</td>
</tr>
<tr>
<td></td>
<td>limits diameter and IWA</td>
</tr>
<tr>
<td>pupil plane</td>
<td>shaped pupil</td>
</tr>
<tr>
<td></td>
<td>achromatic by nature,</td>
</tr>
<tr>
<td></td>
<td>insensitive to vibrations,</td>
</tr>
<tr>
<td></td>
<td>single plane, single optic,</td>
</tr>
<tr>
<td></td>
<td>easy to install and align,</td>
</tr>
<tr>
<td></td>
<td>works on all sources</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>reduced effective telescope area, reduced</td>
</tr>
<tr>
<td></td>
<td>peak flux (even less than APP),</td>
</tr>
<tr>
<td></td>
<td>larger inner working angle, light</td>
</tr>
<tr>
<td></td>
<td>remains on focal plane</td>
</tr>
</tbody>
</table>

manufacture in this way. Therefore, only low-order wavefront changes can be made and the amount of implementable phase designs is limited. Lastly, the spatially smooth antisymmetric coronagraphic phase designs only suppress one side of the star. This means that to image a full 360 degrees around the star a second dataset has to be taken after a rotation of the field by the derotator. See Fig. 1.13 for a typical phase pattern and PSF of the APP (Kenworthy et al., 2010).

Despite these caveats the APP has performed very well on-sky, both by characterizing known planets and discovering unknown companions. As an example, Quanz et al. (2010) imaged Beta Pictoris b in the NIR, thereby confirming the previous detections. They also combined their Narrow-band 4.05 micron photometry with previous $L'$ measurements to get a $[L' - \text{NB4.05}]$ color of $-0.05 \pm 0.38\text{mag}$. A comparison of this color with models of cool field dwarfs suggests that the planet is comparable to a field dwarf with a spectral type of L4 and an effective temperature of 1700 Kelvin. Meshkat et al. (2015b) discovered a low-mass companion around HD 984 with an estimated mass on the order of 90 $M_{\text{Jup}}$ at an angular separation of 0.19 arcseconds from the parent star. Furthermore, Meshkat et al. (2015a) used the APP extensively in a survey of stars with gapped debris disks (suggesting planet formation). The $5\sigma$ contrast reached with the APP by Meshkat et al. (2015c) is up to 10.6 magnitudes at 0.4 arcseconds in 2 hours of observation.
Figure 1.11: Figure with main types of coronagraphs with a small inner working angle, broken down into two families: phase-based focal-plane masks and pupil-plane apodizers. HBL: Hybrid Band-limited, PIAAC: Phase Induced Amplitude Apodization Coronagraph, PIAACMC: PIAA Complex Mask Coronagraph, APCMLC: Apodized Pupil Complex Mask Lyot Coronagraph. Adapted from Mawet et al. (2012).

Figure 1.12: Phase generating principle behind Apodizing Phase Plate. A wave with wavelength $\lambda$ picks up a phase delay $\Delta \phi$ with respect to another wave when it exits a high refractive index ($n_{\text{plate}}$) plate with a $\Delta x$ path length difference. The medium around the plate has a refractive index of $n_{\text{air}} \approx 1$. The dependence of the phase on the wavelength and refractive indices makes this method chromatic. All polarization states are affected equally.

time around a star with $L' = 5$. APP images of two imaged planet can be seen in Figure 1.14.

The limitations of the APP (its chromaticity, single-sided dark hole and phase pattern constraints) can be resolved by switching from classical phase to geometric phase methods. This will be detailed in the following section.
1.4 Enabling technology for achromatic geometric phase elements

1.4.1 Achromatization with geometric phase elements

A different way of generating phase is based on retarding optics where the refractive index is dependent on the angle between the incoming linear polarization and the optic axis of the retarder. A beam of light which is not parallel to the optic axis will experience a different refractive index
for orthogonal components when going through a birefringent optic. In practice, these optical elements are often made with the optic axis parallel to the front surface of the optic. The uniaxial materials used to construct retarders have a different index of refraction along the optic axis than in the plane perpendicular to the optic axis. The axis with the smallest refractive index is commonly referred to as the fast axis, and similarly, the axis with the highest refractive index is called the slow axis. An unpolarized beam of light can be decomposed into two components with orthogonal linear polarizations. These two components of a beam of light pick up a different phase delay depending on whether they travel along the fast or slow axis. The amount of phase delay between the components (or retardance) depends on the wavelength, and thickness and birefringent properties of the optic.

The effect of a half-wave retarder (where the delay between components is exactly half of a wavelength) on a linearly polarized ray with an angle $\theta$ with respect to the fast axis is a rotation by $2\theta$ with respect to a ray that travels along the fast axis. Similarly, a circular polarized ray of light receives a phase delay relative to a beam with the opposite circular polarization. This is the geometric or Pancharatnam-Berry phase, first used by Pancharatnam (1956) in the context of polarized light and later rediscovered by Berry (1984) in the context of quantum mechanics. As we are generally only interested in the relative phase, the phase equation becomes the one in Eq. 1.2.

$$\phi = \pm 2\theta$$

The retarder therefore flips the handedness of the circularly polarized beam and (depending on the handedness) the beam picks up either a positive or negative phase as can be seen in Fig. 1.15.

A phase pattern is no longer encoded in a physical path difference but encoded in the orientation of the fast axis of a half-wave retarder. This geometric phase is therefore not dependent on the wavelength but proportional to the orientation angle of the fast axis. However, there is a chromaticity associated with the retardance of the optic. In the case where the retardance is not exactly half-wave the emerging light also has a second component with a phase that is not dependent on the fast axis orientation. In extreme cases, a zero-wave plate consists of a beam with no added phase, while in the case of a perfect half-wave plate 100% of the beam has the required phase added. The intermediate cases produce a mixture of the two extremes. In the course of this thesis, in cases where the beam ideally has picked up a geometric phase, the component without a phase is commonly referred to as leakage term as it is generally unwanted.

While normal retarding optics have a $1/\lambda$ dependence on their retardance they can be relatively easily achromatized by stacking multiple retarding optics with slightly different orientations or birefringent properties following Pancharatnam (1955). In this way half-wave retardance can be approximated across a large wavelength range which minimizes the leakage term.

Achromatic linear retarders are easy to manufacture as the patterns are simple and therefore require only rotational alignment between layers to produce an approximately constant retardance. This becomes more difficult when the retarder is patterned as in coronagraphs and consequently the X-Y alignment between layers becomes critical, as was demonstrated by the work of Mawet (2010). A higher tolerance on the bandwidth and achromaticity requires more layers and further complicates the manufacturing.

Self-aligning Multi-Twist Retarders (MTR) overcome this problem. These use layers of bire-
Figure 1.15: Phase generating principle of half-wave retarders. The panels show how a circularly polarized wave is decomposed into orthogonally polarized components. The component propagating along the slow axis picks up a phase delay of half a wave with respect to the other component. The two rows show different handedness of polarization. The leftmost half-wave retarder has a fast axis orientation $\theta$ of $+60$ degrees while the rightmost half-wave retarder has a fast axis orientation of $-60$ degrees. Depending on the angle of the fast axis (dashed line on retarder face) the emerging beam is delayed by twice that angle with respect to the case of where the angle is $0^\circ$ (middle column of retarders). The handedness flip naturally emerges as well as the sign flip on the phase. This effect is the same regardless of the wavelength as long as the retardation is exactly half-wave.
fringent liquid crystals as a retarding medium (the polarization modifying properties of liquid crystals have been known since the 19th century, Reinitzer (1888)). A first layer of liquid crystals is deposited on a substrate with a certain fast axis orientation. Multiple liquid crystal layers with varying birefringent properties (e.g., twists) and thicknesses are deposited on top of this initial layer. Due to the elongated shape of the liquid crystals every consecutive layer aligns itself to the pattern on the previous layer (see Fig. 1.16) (Komanduri et al., 2013). To generate 100% bandwidth about 3 layers are required.

1.4.2 Accurate phase patterns

The constraint on the smoothness of the phase pattern can be relaxed by locally setting the fast axis of the retarder. With the liquid crystal based retarders this can be done with a polarized UV source with a beamwidth of the order of 10 microns (see Fig. 1.17). This direct-write approach allows almost any phase pattern and can therefore be used for a wide range of applications, ranging from lenses and gratings to q-plates (Miskiewicz and Escuti, 2014; Kim et al., 2015; Gao et al., 2015).

Within astronomy simple linear phase ramps can be used to produce polarization gratings (PGs, Escuti et al., 2006; Oh and Escuti, 2008; Packham et al., 2010), which generate two spectrally dispersed polarized copies of the imaged target (see Fig. 1.18). Complex patterns like a Vector Vortex Coronagraph (Mawet et al., 2009) and Apodizing Phase Plate coronagraph can also be made using such a scanning UV beam.

1.5 Work presented in this thesis

1.5.1 The vector Apodizing Phase Plate

The vector Apodizing Phase Plate (vAPP) was developed to overcome the limitations of the Apodizing Phase Plate using the patterned half-wave retarder introduced in Section 1.4 to create an achromatic phase delay (Snik et al., 2012).

When circularly polarized light hits a half-wave retarder with a certain fast axis orientation, the polarization state flips and a phase delay is added to the light. The opposite handedness of polarization will also flip states but pick up a negative phase delay. Unpolarized light can be interpreted to be composed of equal amounts of orthogonal linear and circular states. Therefore unpolarized light (such as starlight) can be decomposed into two opposite circularly polarized beams of light. As the phase design of Fig. 1.13 is antisymmetric these opposite circular polarizations will produce PSFs with dark holes on either side. The phase of the light depends only on the fast axis orientation of the retarder and is therefore achromatic. This means that selecting a different wavelength will impose the same phase to the light. If the retarder is not half-wave at all wavelengths part of the light will travel through the plate unaltered, leaking the original PSF on top of the coronagraph PSFs. This reduces the efficiency but as mentioned before retarders can be achromatized easily with a Pancharatnam design (Pancharatnam, 1955; Komanduri et al., 2013) which can reduce the impact to an acceptable level.

The vAPP is designed as a patterned half-wave retarder where the fast axis orientation is chosen to be exactly half of the required phase. Due to the nature of the retarder this will impart exactly
1.5 Work presented in this thesis

Figure 1.16: Left: schematic representation of an achromatic multi-twist retarder. The substrate and initial alignment layer are seen on the left-hand side. Subsequent multiple self-aligning layers with different thickness $d$ and birefringent properties $\phi$ produce an achromatic half-wave retardance. Several representative retardance curves of MTRs with quasi-achromatic half-wave retardance are shown on the right. Adapted from Komanduri et al. (2013).

Figure 1.17: Schematic setup of direct writing tool used by Miskiewicz and Escuti (2014). A polarized UV laser is used to align the first liquid crystal layer on a substrate. The sample is translated underneath the polarized UV spot with an adjustable angle of polarization to vary the fast axis as a function of position.

Figure 1.18: Effect of a polarization grating on an unpolarized beam of white light. The beam is split into two main beams that have orthogonal circular polarization states and a third ‘leakage’ beam that is unaltered and depends on the retardance of the PG. The splitting angle of the two main beams is dependent on wavelength and therefore the two beams are spectrally dispersed. Image courtesy of Geometric-Phase Photonics Lab, NCSU.
Figure 1.19: The required APP phase is no longer encoded in physical path length differences but as the orientation of the fast axis of a retarder. This figure shows an example with the APP phase pattern. Image courtesy of Frans Snik.

the right phase delay to the light as it passes through the coronagraph. Using a laser writing tool this patterning can be done with high spatial resolution on the order of 10 microns (Miskiewicz and Escuti, 2014). The retarder is manufactured to be approximately half-wave (to within a certain tolerance) within the required wavelength range. After splitting according to the circular polarization state with a quarter-wave plate and a Wollaston prism, two beams with opposite phases are formed. By focusing these beams with a lens we get two images of the same source with PSFs with opposite phases. For the default APP PSF this means that we get two PSFs with dark holes on opposite sides of the star as seen in Fig. 1.20. The coverage is now 360 degree in one shot, the coronagraph is achromatic in phase (see Fig. 1.21) and the phase design itself can be more complicated, thereby overcoming the limitations of the APP coronagraph.

Figure 1.20: The operational principle of the vector Apodizing Phase Plate coronagraph. A patterned half-wave plate converts input circular polarizations into their opposite handedness with an extra phase delay added. One circular polarization accrues a positive phase pattern while the opposite polarization receives a negative phase. A QWP plate at 45 degrees and a Wollaston prism then splits the beam based on the circular polarization, while a camera lens images complementary PSFs on the camera (Snik et al., 2012).
1.5 Work presented in this thesis

Figure 1.21: Laboratory measurements of PSF of vAPP taken at three different wavelengths demonstrating the achromatic nature of the technology. The shape of the PSF and the darkness in the dark hole stays consistent while the size scales with $\lambda/D$ as expected. See Chapter 3 for a detailed characterization of the optic.

The aforementioned techniques in this introduction are stepwise improvements to seeing-limited imaging. It starts with Adaptive Optics to enable diffraction limited imaging from the ground. This massively improves the contrast that can be achieved by increasing the peak flux and stabilizing the PSF. Afterwards, the vAPP coronagraph creates a suppressed region close to the star. These improvements are shown in Fig. 1.22 as step 1 through 3. By getting rid of the stellar light in this region the residual noise after PSF subtraction will be reduced which is of course the crucial improvement needed to enhance the detection and characterization of exoplanets. To show the sensitivity of the coronagraph we show the $5\sigma$ point source sensitivity around a bright star in steps 4 and 5, which is based on data taken in this thesis with MagAO on Magellan at 3.94 microns. Step 4 only includes the APP + ADI while step 5 also has an extra PSF subtraction step explained in Chapter 6.

1.5.2 Phase-Sorting Interferometry

Quasi-static aberrations are important contrast-limiting factors. While part of their influence can be reduced by using advanced PSF subtraction techniques like LOCI and PCA, and PSF suppressing coronagraphs it is better if they are corrected in real-time within the instrument. Phase Sorting Interferometry (PSI) is a technique developed by Codona et al. (2008) and expanded by Codona and Kenworthy (2013) to use the wavefront information on residual atmospheric speckles to interferometrically probe the complex amplitude of quasi-static speckles on the science detector. Because the quasi-static speckles have lifetimes that are considerably longer than the atmospheric speckles and the average integration time of the science camera they can be reconstructed and corrected based on the wavefront sensor telemetry and the science camera image cubes. After constraining the phase and amplitude for each speckle the deformable mirror can be adjusted to create a conjugated speckle to remove it. Alternatively, this information can be used in a post-processing step to create an improved PSF model for PSF subtraction.
Figure 1.22: Development of the normalized intensity and 5σ contrast after implementing different correction steps. Starting with 1: a 1 arcsecond FWHM seeing disk that is created after taking a long exposure with normal atmospheric turbulence in the optical band. For graphs 2 through 5 we assume a high performance AO system is operational. Graph 2 shows the intensity profile of a diffraction limited PSF. Graph 3 shows the same but after adding the APP coronagraph to the optical system. The 5σ point source detection sensitivity after ADI is shown in Graph 4 and the sensitivity after Rotation and Subtraction (RS) + ADI is shown in Graph 5. This reduction process will be explained in more detail in this thesis.
1.6 Thesis outline

Chapter 2

Part of instrumental aberrations are unsensed by the wavefront sensor. These unsensed aberrations create quasi-static speckles that limit the contrast that can be achieved. In Chapter 2 we present the AO laboratory testbed at Leiden Observatory and the techniques we propose to use to test the impact of the time synchronization between the wavefront sensor and science camera, the achromacity of the atmosphere and other limiting factors on the PSI focal-plane wavefront reconstruction technique. The AO setup has two separate laser sources to simulate starlight at two distinct frequencies. Depending on the wavelength the light is either sent to a wavefront sensor or a science camera using a dichroic mirror, as is commonly done in direct imaging instruments. This setup can be used to test the performance of the previously mentioned PSI method in conjunction with coronagraphs such as vector Apodizing Phase Plates, Lyot coronagraphs and VVCs. The phase solutions given by the PSI method can be compared with independent phase solutions coming from the differential Optical Transfer Function (dOTF) method and a phase-stepping interferometer. This setup enables us to rapidly prototype high-contrast imaging techniques and has since been used to characterize the first visible band vAPP prototypes. This chapter is based on Otten et al. (2012).

Chapter 3

The concept of the vAPP was first introduced by Snik et al. (2012) and an initial narrow-band prototype was shown to produce two complementary PSFs with dark holes. In Chapter 3 the first achromatic vAPP prototype based on a three-layer multi-twist retarder was tested. We have performed both pupil and PSF measurements of the achromatic vAPP at multiple wavelengths in the visual band (400-800 nm). We recorded the intensity of the coronagraph pupil between parallel and crossed polarizers at different position angles. We used pupil measurements to reconstruct the three important parameters of the coronagraph as a function of position: transmission, retardance and fast axis orientation. Based on a Mueller matrix representation of the coronagraph in its measurement setup these properties could be constrained. These three parameters were applied in a PSF model based on Jones calculus and compared to PSF measurements taken in the same bands. The intensity profile of the forward model and the reality were in agreement with an average normalized intensity ratio of $10^{-3.8}$ measured from 2 to 7 $\lambda/D$. While the coronagraph performed within specifications it was concluded after analyzing the different coronagraph properties that the most detrimental impact to the contrast was caused by the offset from half-wave retardance, which causes light of the original PSF to leak through. This can be resolved by either filtering the leakage with polarizing filters or producing a retarder with improved half-wave behavior, for instance by adding more retarding layers. This chapter is based on work presented in Otten et al. (2014b).

Chapter 4

The principle of the vector Apodizing Phase Plate coronagraph presented in Snik et al. (2012) and Otten et al. (2014b) works in the lab but is harder to implement in an instrument than the single
optic regular APP as it requires an achromatic beamsplitter and quarter-wave plate that need to be carefully aligned. Furthermore, one of the conclusions of Chapter 3 is that the vAPP’s performance is limited by the offset from half-wave retardance of the coronagraph, creating leakage terms that deteriorate the contrast of the coronagraphic PSFs. While the leakage can be suppressed by filtering polarization states, this complicates the installation into instruments even more. Chapter 4 shows that a phase ramp (i.e., a polarization grating phase pattern) added to the vAPPs phase pattern can reduce the complexity and increase the robustness of the coronagraph. The included phase ramp creates a splitting based on the circular polarization states, thereby eliminating the quarter-wave plate and Wollaston prism. The coronagraphic PSFs are moved away from the non-half-wave leakage term. This eliminates the impact that the retardance has on the contrast ruining leakage. This design is called the grating-vAPP (gvAPP). The robustness of the gvAPP comes at a cost because spectral smearing in one dimension limits the width of an individual filter to approximately 5%. However, thanks to liquid-crystal achromatization one gvAPP device can be used with narrow-band filters throughout a large wavelength range. We also present a manufactured vAPP design that can suppress a 360 degree annular region around the star. This phase pattern is so extreme that it can not be manufactured using conventional diamond turning manufacturing techniques. The validity of the phase design is tested in a laboratory setup and shows a rearrangement of the light in such an annular area. This chapter is based on work presented in Otten et al. (2014a).

Chapter 5

In Chapter 5 we present the first on-sky results with the narrow-band gvAPP coronagraph on the Large Binocular Telescope. These results show that the vAPP coronagraph can be taken within a few months from concept to installation and works in the near-infrared (4 microns) under cryogenic conditions as expected. The phase ramp that was added in the design works as predicted by keeping the leakage PSF away from the coronagraphic PSFs. Our analysis shows however that the first light measurements taken at 4.05 microns narrow-band are limited in performance by a pupil misalignment and a significant amount of wavefront error in the form of astigmatism. Our PSF simulation shows that the measured performance can be explained by the combined impact of both factors. The pupil alignment issue can be solved by realigning the pupil before taking the observations and the wavefront error can be dialed out using the deformable mirror of the instrument.

Chapter 6

In Chapter 6 we present the first on-sky results with the gvAPP coronagraph at MagAO/Clio2. This coronagraph is optimized to work at an extremely broad wavelength range from 2 to 5 microns and also includes a phase grating to separate the coronagraphic and leakage PSFs. We demonstrate that the two PSFs can be combined to perform an even better PSF subtraction, especially in the speckle halo close to the star. The apparent limitations seen at the LBT are not present in this coronagraph and instrument combination, and it achieves an unprecedented contrast close to the star. After rotating, optimally scaling one PSF, and subtracting it from the other PSF we see a contrast improvement by 1.46 magnitudes at $3.5 \frac{\lambda}{D}$. Applying regular Angular Differential Imaging as a final step, the MagAO gvAPP coronagraph delivers a $5\sigma$ $\Delta$mag contrast
of 8.3 ($= 10^{-3.3}$) at 2 $\lambda/D$, 10.8 ($= 10^{-4.3}$) at 2.5 $\lambda/D$, 12.2 ($= 10^{-4.8}$) at 3.5 $\lambda/D$, and 12.5 ($= 10^{-5}$) at 4.5 $\lambda/D$, all measured at 3.9 $\mu$m in 1.5 hours of observing time. Compared to other coronagraphs on current 6 – 10 meter telescopes the vAPP performs up to 2 magnitudes better inward of 5 $\lambda/D$. These results can also be seen at a glance in Fig 1.22. These numbers improve with more advanced PSF subtraction methods like PCA. This work is to be submitted to ApJ.

We conclude the scientific part of the thesis with an outlook on future work and scientific possibilities with the vAPP in Chapter 7.
References


REFERENCES


REFERENCES


