Diss ETH No. 17832

Applications of High-Precision Polarimetry to Extrasolar Planet Search and Solar Physics

A dissertation submitted to
ETH ZÜRICH

for the degree of
DOCTOR OF SCIENCES

presented by

CHRISTIAN THALMANN

Dipl. Phys., ETH Zürich

born
29.09.1979

citizen of
Wiezikon TG, Switzerland

accepted on the recommendation of

Prof. Dr. Simon Lilly
Prof. Dr. Jan Olof Stenflo
PD Dr. Hans Martin Schmid
Dr. Markus Feldt

2008

L I A M F

I N J I U

U F I N A

I I I I S

A M N A H

Contents

Contents

Contents

Symbols

The following list denotes the symbols used for certain physical quantities or concepts in this work. They may be complemented with subscript or superscript indices to specify particular instances or components of the quantity. A symbol used for different quantities in different contexts is listed multiple times. The number in the right-hand column indicates the section in which the symbol is first used.

$''$	decimal fraction of arcseconds	5.2.2
$a(\phi)$	phase albedo at phase angle ϕ	5.2.3
a_{g}	geometric albedo	5.2.3
α	azimuth angle	4.2
$\vec{B}$	magnetic field	A.1
c	vacuum speed of light	A.1
c_w	circular aperture of width w	5.6.2
C	contrast	5.2.3
d	distance (Earth to target system)	5.2.3
D	telescope aperture diameter	2.2
δ	phase shift	A.1
$\hat{e}$	unit vector	A.1
$\vec{E}$	electric field	A.1
ε_0	electric permittivity	A.1
$\vec{\mathcal{E}}$	electric complex amplitude	A.1
g_w	gaussian aperture of width w	5.6.2
I	intensity Stokes component	A.2
ι	incidence angle; inclination angle	4.2
$\vec{k}$	wave propagation vector	A.1
λ	wavelength	2.2
μ_0	permittivity of free space	A.1
n	index of refraction	4.2
n	a natural number	5.3
N	number of iterations	5.2.2
N	a natural number	5.3
ω	angular frequency	A.1
p	degree of polarization	A.1
ϕ	phase angle (of a planet)	5.2.3
ϕ	field rotation angle	5.3
Q	rectilinearly polarized Stokes component	A.2

r	radius	5.2.3
R	orbital radius	5.2.3
ρ	refracted beam angle	4.2
$\vec{S}$	Stokes vector	A.2
σ	standard deviation	5.2.3
t	time	A.1
T	transmittance	4.2
θ	angular separation (planet to parent star)	5.2.3
τ_0	time scale of speckle evolution	2.3
U	diagonally polarized Stokes component	A.2
V	circularly polarized Stokes component	A.2
w	width (of a sampling aperture)	5.6.2

Abbreviations

4QPM	Four-Quadrant Phase Mask coronagraph
AAA	Active [field rotation, and] Angular Averaging
ADC	Atmospheric Dispersion Corrector
AO	Adaptive Optics
ASTRON	Stichting ASTRonomisch Onderzoek in Nederland
BB	BroadBand
BS	Beam Splitter
CAOS	Code for Adaptive Optics Simulation
CCD	Charge-Coupled Device
CHEOPS	CHaracterizing Exoplanets by Opto-infrared Polarimetry and Spectroscopy
CPU	Central Processing Unit (computer processor)
CTE	Charge Transfer Efficiency
CTI	Charge Transfer Inefficiency
E-ELT	European Extremely Large Telescope
ESO	European Southern Observatory
ETH	Eidgenössische Technische Hochschule (Swiss Federal Institute of Technology)
FF	FlatField
FITS	Flexible Image Transport System (image format)
FLC	Ferro-electric Liquid Crystals
GUI	Graphical User Interface
HST	Hubble Space Telescope
HWP2	Half-Wave Plate 2, used for signal switching
IDL	Interactive Data Language
IRDIS	InfraRed Dual-Imaging Spectrograph
IRSOL	Istituto Ricerche SOlari Locarno
IWA	Inner Working Angle
JWST	James Webb Space Telescope
LAOG	Laboratoire d'Astrophysique de l'Observatoire de Grenoble
mas	milli-arcseconds
MDI	Michelson Doppler Imager
NAA	Natural [field rotation, and] Angular Averaging
NASA	National Aeronautics and Space Administration
NIR	Near Infra-Red
OWL	OverWhelmingly Large telescope

Abbreviations

PanSTARRS	Panoramic Survey Telescope & Rapid Response System
P1	Polarimetric observation mode without field derotation
P2	Field-stabilized polarimetric observation mode
P3	Quasi-field-stabilized polarimetric observation mode
PEM	Piezo-Elastic Modulator
PLANET	Probing Lensing Anomalies NETwork
PRIMA	Phase-Referenced Imaging and Micro-arcsecond Astrometry
PS	PostScript (document format)
PSD	Power Spectrum Density
PSF	Point-Spread Function
QE	Quantum Efficiency
RMS	Root of Mean Square ($\sqrt{\langle x^2 \rangle}$)
RON	Read-Out Noise
SAXO	SPHERE Adaptive optics for eXoplanet Observation
SC3	Standard Case 3 (default set of simulation parameters)
SC4	Standard Case 4
S/N	Signal-to-Noise ratio
SOHO	SOlar and Heliospheric Observatory
SPHERE	Spectro-Polarimetric High-contrast Exoplanet REsearch
SPW	Solar Polarization Workshop
TPF	Terrestrial Planet Finder
USM	UnSharp Masking
VLT	Very Large Telescope
WFS	Wave-Front Sensor
ZIMPOL	Zurich Imaging POLarimeter

Abstract

The thesis at hand is divided into two distinct parts, each of which concerns itself with a particular occurrence of scattering polarization in nature.

THE FIRST PART is dedicated to determining the feasibility of finding extrasolar planets by imaging polarimetry.

Direct detection of a mature extrasolar planet has not yet been achieved due to the prohibitive brightness contrast between the planet and its parent star (typically of the order 10^{-8}) and their small angular separation. However, the starlight reflected off such a planet's atmosphere is polarized due to Rayleigh scattering, providing a basic possibility of distinguishing it from the unpolarized direct starlight. The SPHERE instrument (Spectro-Polarimetric High-contrast Exoplanet REsearch) currently being developed by ESO for the Very Large Telescope in Chile meets this challenge with an 8 m mirror, a high-end adaptive optics system, a stellar coronagraph, and a detector based on the Zurich Imaging Polarimeter (ZIMPOL) design operating at 600–900 nm.

ZIMPOL relies on polarization modulation at kHz frequencies and specialized demodulating CCD sensors, and has been proven to achieve sensitivities of 10^{-5}. Since even small levels of background polarization ($\sim 1\%$) significantly raise the level of detector noise in ZIMPOL, we developed and successfully tested a compensator device able to counteract a variable polarization background in real time, keeping it well below 0.1%. This work is described in Chapter 4.

In order to quantify and characterize the feasibility of planet detection with SPHERE, all components of the telescope and instrument must be taken into account. To this end, we carried out a detailed and comprehensive end-to-end computer simulation of such an observation. A specifically adapted diffraction code produces realistic broadband point-spread functions for both the coronagraphed target star and the off-axis planets, accounting for a plethora of parameters such as adaptive optics performance, atmospheric seeing, static aberrations, temporal and polarized differential aberrations, and coronagraph properties. A dedicated photometry and reduction code then implements photon noise and various properties of the detector, star and planets, applies calibration and signal enhancement methods, and computes the S/N for planet detection.

Using this code, we elaborated a comprehensive analysis of the expected system performance and its dependence on the various input parameters. We conclude that Jupiter-sized planets can indeed be detected with SPHERE around the eight nearest bright stars, using signal-switching calibration and angular differential imaging to suppress the spurious polarized background structures caused by aberrations.

Abstract

Around α Cen A/B and Sirius, even Neptune-sized planets and super-Earths can be discerned. This performance can be further improved by pushing the differential wedge behavior of the modulator and the surface roughness of the half-wave plate used for signal switching below the current specifications. The inclusion of a four-quadrant phase mask coronagraph is found to expand the target list for planet detection to over 20 stars. Finally, experimentation with different signal extraction methods yielded insights on desirable observation procedures. This work is presented in Chapter 5.

THE SECOND PART of this thesis describes a laboratory experiment intended to shed light on the causes of the enigmatic scattering polarization signature of the D_1 lines in the solar spectrum.

Precise measurement of the scattering polarization at the solar limb as a function of wavelength reveals a spectrum as richly structured as the well-known intensity spectrum. Each spectral line leaves a characteristic signature in this "Second Solar Spectrum", betraying the particular quantum nature of the underlying atomic or molecular transition as well as providing insight on the local temperature, density and magnetic field conditions. While most signatures have been understood by theory, a particular feature—the positive peak in the center of Na I D_1 and Ba II D_1—has thus far obstinately resisted such attempts.

To determine whether the solution lies in solar or atomic physics, we isolated the pure D_1 scattering event from its complex solar circumstances and investigated it under simple, well-defined laboratory conditions. We illuminated a sample of potassium vapor (whose K I D_1 transition shares the exact same quantum properties of Na I D_1) with a tunable solid-state laser, whose wavelength could be precisely scanned across the spectral line, and studied the polarized light scattered off the atoms at a right angle. Six different polarization states could be selected for the ingoing light, and a magnetic field could be applied to the vapor along any one of the three coordinate axes. This work is presented in Chapters 8 and 9.

While the measurements do not yet offer a solution to the solar D_1 enigma, they unequivocally show that current theory is missing a crucial puzzle piece, and also imply that solar magnetic fields are likely involved. A promising new theoretical approach is now being developed on the basis of these measurements.

Zusammenfassung

Die vorliegende Dissertation gliedert sich in zwei Teile, die sich mit je einem spezifischen Auftreten von Streupolarisation in der Natur befassen.

DER ERSTE TEIL widmet sich einer Machbarkeitsstudie zum direkten Nachweis von extrasolaren Planeten durch bildgebende Polarimetrie.

Die direkte Detektion eines "kühlen" Planeten (der seine Entstehungswärme bereits abgestrahlt hat) ist auf Grund des extremen Helligkeitskontrasts zwischen Planet und Stern (typische Grössenordnung 10^{-8}) sowie des kleinen Winkelabstands noch nicht gelungen. Das von der Planetenatmosphäre reflektierte Sternlicht ist jedoch auf Grund der Rayleigh-Streuung polarisiert, was eine Unterscheidung vom unpolarisierten direkten Sternlicht ermöglicht. Das Instrument SPHERE (Spectro-Polarimetric High-contrast Exoplanet REsearch), das zur Zeit von ESO für das Very Large Telescope in Chile entwickelt wird, stellt sich dieser Herausforderung mit einem 8 m grossen Spiegel, einer hochgezüchteten adaptiven Optik, einem Sternkoronographen und einem Detektor, der auf dem Konzept des Zurich Imaging Polarimeter (ZIMPOL) basiert und den Bereich von 600–900 nm abdeckt.

ZIMPOL basiert auf Polarisationsmodulation im kHz-Bereich und speziellen demodulierenden CCD-Detektoren. Es erzielt nachweislich eine Empfindlichkeit von 10^{-5}. Da in ZIMPOL bereits kleine Mengen von Hintergrundpolarisation ($\sim$1%) das Detektorrauschen merklich verstärken, wurde ein Kompensationsgerät entwickelt, das einem veränderlichen Polarisationshintergrund in Echtzeit entgegenwirkt und ihn zuverlässig unterhalb 0.1% stabilisiert. Diese Arbeit wird in Kapitel 4 beschrieben.

Um die Machbarkeit der Planetendetektion mit SPHERE quantitativ und qualitativ abzuschätzen, müssen alle Komponenten von Teleskop und Instrument einbezogen werden. Dafür erstellten wir eine detaillierte und umfassende Computersimulation der gesamten Beobachtungssequenz. In einem ersten Schritt berechnet ein speziell dafür angepasster Diffraktionscode realistische, breitbandige Teleskopbilder ("Punktantworten") sowohl für den koronographierten Zielstern wie auch für die weiter aussen liegenden Planeten. Dabei wird eine Vielzahl von Parametern wie die Leistung der adaptiven Optik, atmosphärisches Seeing, statische Aberrationen, zeitliche und polarisierte differentielle Aberrationen und Eigenschaften des Koronographen berücksichtigt. Ein eigener Code für Photometrie und Reduktion fügt dann Photonenrauschen und verschiedene Eigenschaften von Detektor, Stern und Planeten hinzu, wendet Kalibrations- und Bildbearbeitungsmethoden an, und errechnet die Detektierbarkeit der Planeten.

Mit Hilfe dieses Programms erarbeiteten wir eine umfassende Analyse der erwarteten Systemleistung und ihrer Abhängigkeit von den Eingabeparametern. Wir kommen zum Schluss, dass jupitergrosse Planeten der acht nächsten hellen Sterne tatsächlich mit SPHERE nachgewiesen werden können, indem die durch Aberrationen verursachten polarisierten Störsignale im Hintergrund mittels Signalaustausch-Kalibration und Ausnützung der Feldrotation unterdrückt werden. Bei α Cen A/B und Sirius sind sogar neptungrosse Planeten und "Super-Erden" nachweisbar. Diese Leistung kann noch verbessert werden, indem man das differentielle Keilverhalten des Modulators und die Oberflächenrauheit der Halbwellenplatte, die für den Signalaustausch verwendet wird, stärker minimiert. Die Verfügbarkeit eines Vierquadranten-Phasenplatten-Koronographen erweitert die Kandidatenliste für die Planetensuche auf über 20 Sterne. Zudem führte das Experimentieren mit Signalextraktionsmethoden zu Erkenntnissen, wie die Beobachtungsabläufe gestaltet werden sollten. Diese Arbeit wird in Kapitel 5 vorgestellt.

DER ZWEITE TEIL dieser Dissertation beschreibt ein Laborexperiment, dessen Ziel es war, die Hintergründe einer rätselhaften Signatur in der Streupolarisation der D_1-Linien im Sonnenspektrum aufzudecken.

Hochpräzise Messungen der Streupolarisation am Sonnenrand als Funktion der Wellenlänge enthüllen eine Fülle an Informationen, die in ihrer reichhaltigen Struktur dem Intensitätsspektrum um nichts nachsteht. Jede Spektrallinie hinterlässt in diesem "Zweiten Sonnenspektrum" eine charakteristische Signatur, die die quantenmechanische Natur des verursachenden atomaren oder molekularen Übergangs offenbart und darüber hinaus Einblicke in die lokalen Temperatur-, Dichte- und Magnetfeldverhältnisse gewährt. Während die meisten Signaturen durch theoretische Überlegungen verstanden worden sind, trotzt ein gewisses Signal – der positive Ausschlag im Linienzentrum von Na I D_1 und Ba II D_1 – bislang jedem Erklärungsversuch.

Um zu verstehen, ob die Ursache dafür in der Sonnen- oder der Atomphysik liegt, isolierten wir den D_1 Streuprozess von der komplexen Umgebung der Sonnenatmosphäre und untersuchten ihn unter einfachen, nachvollziehbaren Laborbedingungen. Wir bestrahlten Kaliumdampf (dessen Übergang K I D_1 exakt dieselben Quanteneigenschaften wie Na I D_1 aufweist) mit einem abstimmbaren Festkörperlaser, dessen Wellenlänge präzise durch den Bereich der Spektrallinie gescannt werden konnte, und untersuchten das polarisierte Licht, das von den Atomen im rechten Winkel weggestreut wurde. Sechs verschiedene Polarisationszustände konnten für den eingehenden Strahl eingestellt werden, und ein Magnetfeld konnte entlang jeder der drei Koordinatenachsen angelegt werden. Diese Arbeit wird in den Kapiteln 8 und 9 vorgestellt.

Zwar haben die Messungen noch keine Lösung zum solaren D_1-Rätsel geliefert; fest steht jedoch, dass sie mit der gängigen Theorie nicht vereinbar sind und somit auf ein fehlendes Puzzleteil im heutigen Verständnis der Streuphysik hindeuten. Zudem lassen sie vermuten, dass Magnetfelder in der Sonnenatmosphäre am Phänomen beteiligt sind. Auf der Grundlage dieser Messungen wird nun ein vielversprechender neuer theoretischer Ansatz entwickelt.

16

Part I

The Polarimetric Search for Exoplanets

Chapter 1

Introduction

1.1   Extrasolar Planets

Few ideas hold such sway over the imagination of astronomers as the thought of other worlds, other planets not much unlike ours, and the unknown riches of biological diversity they may hold. It is only natural, then, that the discovery and characterization of planets, as well as the investigation of their abundance and distribution in the universe, be among the hottest topics of astronomical research.

A mere two decades ago, planetary scientists suffered from the same fundamental limitation as cosmologists still do—they only had a single specimen of their object of interest available. The solar system had to suffice as the lone representative of the planet genus. But while the universe remained unique, the Milky Way offered many billion stars as potential parents to planetary systems waiting to be discovered.

Wolszczan & Frail (1992) reported the first detection of an extrasolar planet, a veritable "freak system" of Earth-mass bodies orbiting a millisecond pulsar that betrayed its existence through periodic modulation of the pulsar's stroboscopic timing. The fact that the planets must have formed after the supernova event producing the pulsar implied that planet formation was not as rare an occurrence as had been feared.

Mayor & Queloz (1995) surprised the science community with the detection of a jovian planet in an extremely close orbit around 51 Peg; a configuration previously considered unphysical, but which lent itself well to Doppler spectroscopy. The possible range of planet properties appeared to be much more diverse than anticipated. Alerted to this particularly detectable demographic of planets, more detections soon followed. The discovery of transiting planets further increased and complemented the rapidly expanding corpus of extrasolar planet data.

At the time of writing, 288 extrasolar planets are known around nearby stars (website[1]), including stars in binary star systems (e.g. Hatzes et al., 2003). At least two dozen systems have been found to bear several detectable planets. The star 55 Cancri is even known to harbor five planets (Fischer et al., 2008).

Due to the limitations of the detection methods, most of the known exoplanets are gas giants akin to Jupiter. In a few rare instances, so-called "super-Earths" with a projected mass of less than 10 Earth masses which might be of rocky nature have been found (Rivera et al., 2005). The pulsar planets of Wolszczan & Frail
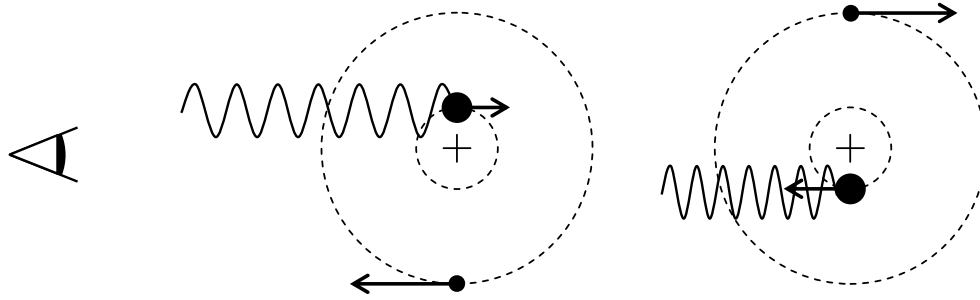
---

[1] `http://exoplanet.eu`

Fig. 1.1: Principle of Doppler spectroscopy. As the star (big dot) orbits around the mass center of the star-planet system (cross), the light sent towards the observer (to the left in this sketch) is red- or blueshifted when the velocity vector points away from or towards the observer, respectively. This frequency modulation of the star's spectrum can be measured. The planet remains unseen.

(1992), whose masses reach down to the level of Earth's moon, remain a unique exception. These findings implicate the likely existence of a vast population of low-mass planets to which our current instruments are blind, but which might be probed and quantified by future projects.

1.2   Detection Methods

1.2.1   Doppler Spectroscopy (Radial Velocity)

Planets do not revolve around static stars; rather, both the planets and their parent stars orbit their common center of gravity. Due to the large mass ratio between star and planet, the star's orbital radius is typically very small compared to the planet's. For a massive planet in close orbit, however, this reflex motion of the star (often referred to as a "wobble") can reach velocities of $10\,\mathrm{m/s}$ or even $100\,\mathrm{m/s}$, causing a measurable redshift modulation of the stellar spectrum by means of the Doppler effect. The vast majority of the currently known exoplanets have been detected in this way, using highly accurate spectrometers such as ESO's HARPS (Mayor et al., 2003).

The ideal configuration for Doppler spectroscopy (also commonly called the "radial velocity method") is a massive planet in a very close orbit seen edge-on, i.e. with an inclination angle $\iota \approx 90°$. Only the line-of-sight component of the star's motion produces a Doppler shift, therefore an observation does not measure the planet mass m, but the projected mass $m \sin \iota$. Since the inclination of the system in generally unknown, this method only provides a lower limit for the planet mass. However, edge-on systems are statistically more common than face-on system, thus the real mass often lies close to the projected value. Transiting planets are known to be well-aligned with the line of sight, thus the exact planet mass can be derived from the Doppler measurement.

Very active or variable stars can obscure the presence of planets with the inherent Doppler signatures of their evolving surfaces (O'Toole et al., 2008).
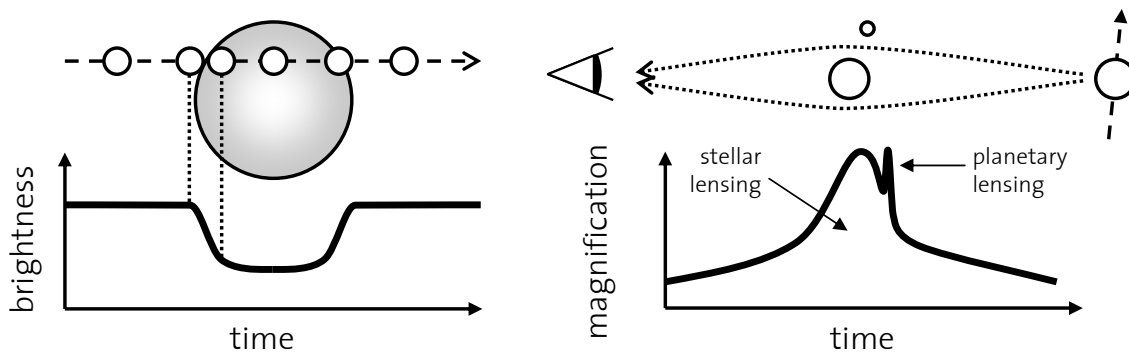
Fig. 1.2: Left: Principle of transit photometry. The obscuration of the stellar disk causes a dip in the star's light curve. The planet's radius can be inferred from the duration of the ingress and egress phases. **Right:** Principle of microlensing. The gravity well of a foreground star acts a lens for a passing background star, magnifying its light. The presence of a planet in the foreground system introduces a distortion to the lens that causes a sharp signature in the light curve.

1.2.2 Transit Photometry

Planets whose orbital planes are so parallel to the line of sight that they pass across the projected disk of their parent star can betray their presence by blotting out a part of the starlight, causing a periodic dip in the stellar light curve (e.g. Charbonneau et al., 2000).

Aside from the obvious orbital period, the planet's radius can be gleaned from the depth and shape of the light curve depression. During primary eclipses, the starlight passing through the upper planetary atmosphere can pick up spectral absorption lines, such as of atmospheric sodium (Charbonneau et al., 2002). If the planet's photometric signal is strong enough, it can be directly measured during secondary eclipses by subtracting the eclipse signal (star only) from the two-body signal (Charbonneau et al., 2005).

While transit photometry provides very valuable data about extrasolar planets, it requires a very specific geometric setup. For planets in close orbits, the chance of a transiting alignment is about 10%, vanishing very quickly for wider orbits. Furthermore, the rate of false positives is high, requiring a follow-up observation with a different detection method (ideally Doppler spectroscopy) for confirmation. On the positive side, wide-angle surveys such as the upcoming PanSTARRS project in Hawai'i (Afonso & Henning, 2007) can track the photometric behavior of a great number of stars simultaneously and are expected to find many candidates in a short time.

1.2.3 Microlensing

The gravity well of a star can act as a lens to magnify the light of another star further in the background. Since this requires the stars to be exceedingly well aligned and all participating bodies are moving with their proper motion, such events appear unpredictably and merely last weeks or days. The presence of a planet near the

foreground star distorts the lens shape, resulting in caustic patterns in the projected image of the background star, which can be measured as a brief distortion of the light curve of the lensing event.

Such events are always unique and therefore unavailable for follow-up observations. However, the detection is unambiguous, and provides information about the mass and orbital separation of the planet. Due to the extreme range of this detection methods, statistical information about the galactic planet population can be accumulated.

The method is notably sensitive to low-mass planets in wide orbits, as exemplified by the ~ 5 Earth mass planet OGLE-2005-BLG-390Lb (Beaulieu et al., 2006). Lowering the mass of the planet does not decrease the amplitude of the light curve disturbance, but merely its duration.

World-wide networks like PLANET (Dominik et al., 2002) are employed to keep continuous watch over a large number of target stars, aiming to notice any microlensing events as early as possible and covering their entire duration with observations so as not to miss any planet signatures.

1.2.4   Astrometry

Like Doppler spectroscopy, astrometry relies on the parent star's reflex motion to infer the presence of a planet. The projected orbit of the star around the center of gravity of the star-planet system can in principle be tracked and measured by a sufficiently precise instrument. The most favorable configuration for this detection method is a massive planet in a wide face-on orbit around its star, thus complementing the target parameter space of transit and Doppler methods, which require edge-on orbits. On the downside, the orbital period of such a system can easily exceed a decade, requiring long-running observation programs.

A planet around Gliese 876, previously detected through radial velocity, has been confirmed by astrometry using the Hubble Space Telescope (Benedict et al., 2002). ESO's upcoming PRIMA instrument for the VLT interferometer (Quirrenbach et al., 1998) is specialized for this purpose.

1.2.5   Pulsar Timing

Pulsars are extremely fast-rotating neutron stars emitting radio waves in highly periodic pulses. The reflex motion caused by orbiting planets modulates its distance from Earth, causing the pulses to arrive slightly too early or delayed. By measuring the timing of the radio pulses to a high precision, the masses and orbits of the planets can be inferred.

The method is extremely sensitive, as the two Earth-sized and one Moon-sized planets detected around the pulsar PSR 1257+12 have proven (Wolszczan & Frail, 1992). However, since pulsars planets are exotic objects and likely very rare, the chances of finding more planets with this method are extremely slim.
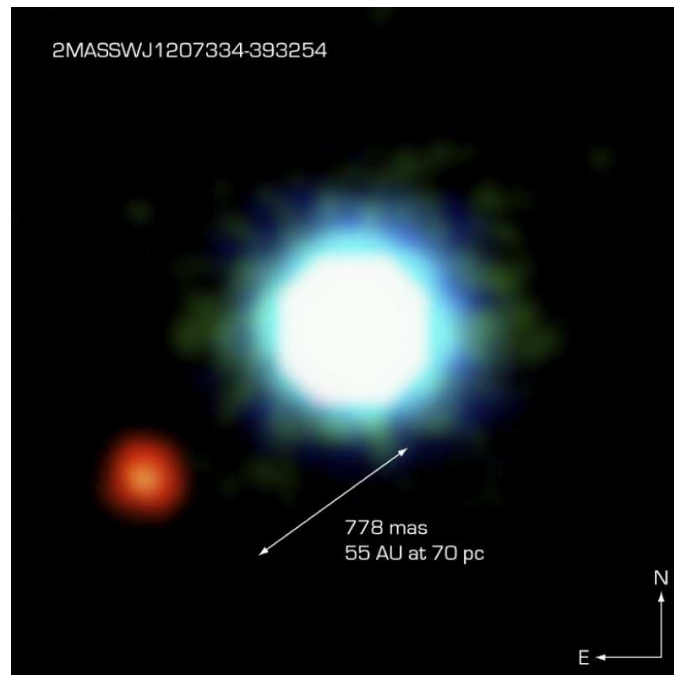
Fig. 1.3: The nearby brown dwarf 2MASSWJ1207334-393254 and its supposed companion, the only planetary object that has been directly imaged to date. Source: ESO PR Photo 14a/05.

1.2.6 Direct Imaging

This most classical of astronomical detection methods is particularly challenging to apply to extrasolar planets. The planets are typically many orders of magnitude fainter than their parent stars, yet very close to them in angular separation. The overwhelming level of background light from the star's PSF halo drowns out the planet signal.

Nevertheless, efforts are being made to harness the most powerful telescopes currently available for this purpose. Extreme adaptive optics and several differential imaging methods are envisioned to peel away the starlight and discern the planet's signal from spurious ones.

Direct imaging provides a particularly interesting data set about the planet, including its exact orbital parameters and surface properties. Spectral features and temporal variability of the light could yield information about the chemical composition of the atmosphere, point to the presence of cloud layers or continents, or even biological activity.

A particularly favorable configuration for direct imaging is a young gas giant still self-luminous from gravitational contraction located in a wide orbit around a dim red dwarf. Old planets, on the other hand, are visible only in starlight reflected from their atmospheres or surfaces, and are therefore yet harder to detect. The typical brightness contrast between an old gas planet and its star lies at 10^{-7}–10^{-9}.

Figure 1.3 shows the current frontier of direct imaging, an object likely to be a young, hot planet-mass companion to a brown dwarf (Chauvin et al., 2005). It is

often considered the first directly imaged exoplanet. No planets of main-sequence stars have yet been imaged. This work, however, will focus on a particular effort to find more exoplanets with the direct imaging technique: The SPHERE project (Section 3).

Chapter 2

Basic Concepts

This thesis uses the Stokes formalism to discuss polarized light. For the reader unfamiliar with these concepts, an introduction is given in Appendix A.

2.1 Telescopes

Within the scope of this work, a star can be regarded as a point-like source of light. While every photon emitted from this source propagates as an expanding spherical wave, its origin is so far away from us that the wavefronts incident on the Earth are no longer measurably curved; we can consider them to be plane.

Reflector telescopes exploit the fact that a paraboloid mirror returns every incoming plane wave as an outgoing spherical wave converging on a point in its *focal plane*, its lateral location corresponding to the angular position of the source on the sky. A detector placed in the focal plane will therefore register a sharp image of the sky field (Figure 2.1). In actual telescopes, additional mirrors are employed to fold the light path to achieve a more compact and practical design.

A detailed introduction to the topic of telescope optics can be found e.g. in Born & Wolf (1980).

2.2 Point-Spread Function

From a mathematical point of view, the plane wave incident on the telescope aperture (the *pupil*) is the Fourier transform of the point-source on the sky, performed by the transfer function of empty space. The function of a telescope, then, is simply to apply the reverse transformation.

In practice, however, this reconstruction of the original image is flawed due to the wave nature of light. The telescope mirror has a finite diameter D, causing diffraction on its edges and acting as a low-pass filter on the information encoded in the light wave. The resulting star image is therefore not a sharp point (the Fourier transform of a homogeneous plane), but an *Airy pattern*, the Fourier transform of a circular aperture. This pattern consists of a central core and an infinite sequence of concentric rings containing a progressively lesser share of the total intensity (see Figure 2.2a). The angular radius of the core is λ/D, where λ is the wavelength of the light. This imposes a hard limit on the resolution of a telescope with a given size observing at a given wavelength.

para-bolic mirror

images:

of star A

of star B

directions:

from star B

from star A

focal plane
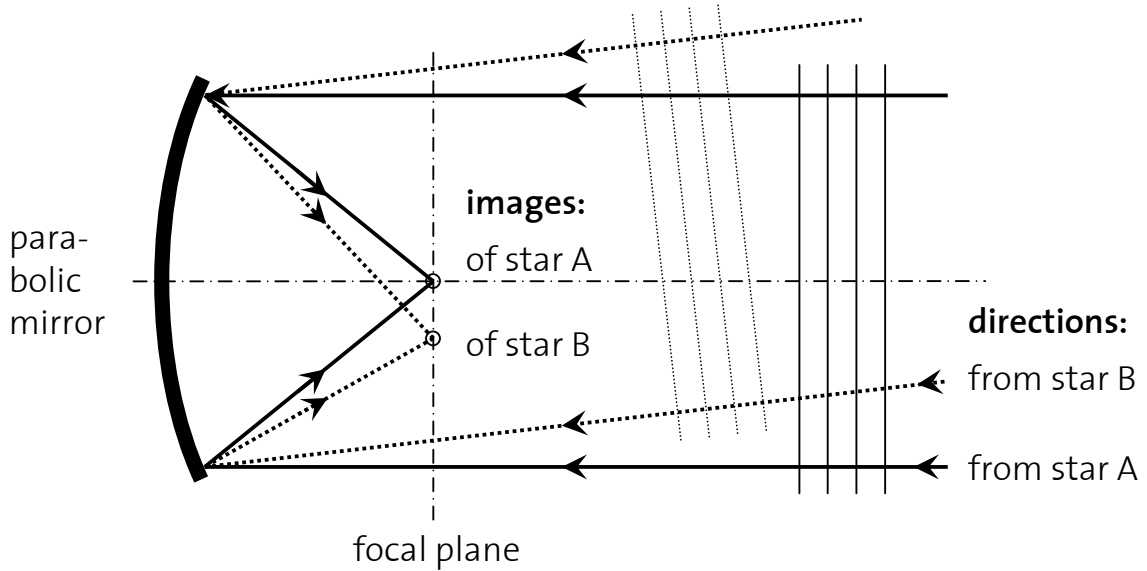
Fig. 2.1: Basic principle of a telescope. The path of two representative light rays are shown for each of two stars. The lines arranged perpendicular to the rays represent the plane wavefronts.
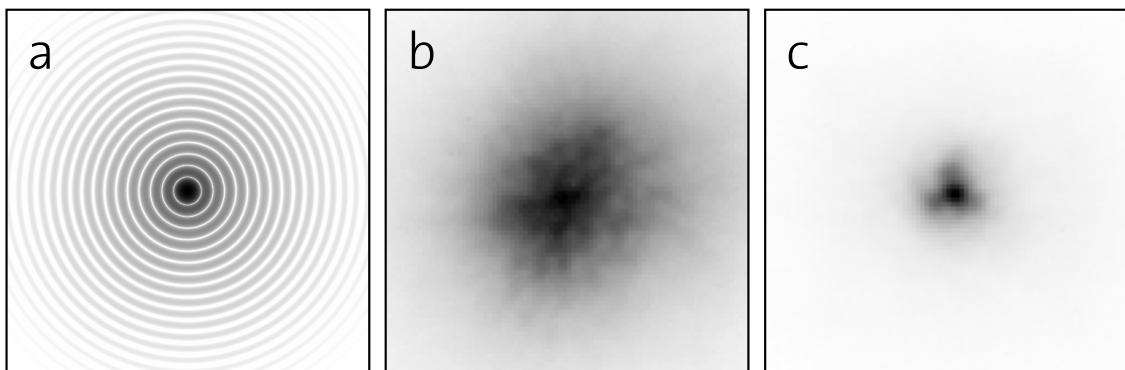
Fig. 2.2: a: The intensity distribution of the Airy pattern, i.e. the PSF of a perfect, diffraction-limited telescope. The greyscale is logarithmic in order to show the high-order diffraction rings. **b, c:** A point-spread function smeared out by atmospheric seeing, before (**b**) and after (**c**) correction by adaptive optics. The greyscale is linear. These two images are taken from the first light observations of the laser guide star AO at Palomar on June 13, 2006.

The intensity distribution on the detector produced by a point-like source on the sky is referred to as the *point-spread function* (PSF) of the telescope (e.g. Born & Wolf, 1980).

2.3 Adaptive Optics

While the geometric constraint of diffraction cannot be circumvented for a given telescope size, the resolution of actual telescopes usually suffers from even graver limitations. Imperfections on optical surfaces buckle and distort the incoming wavefronts, degrading the sharpness of the PSF and giving rise to an inhomogeneous halo of stray light in the final image, strewn with *speckles* on the λ/D length scale. For earthbound telescopes, these static or quasi-static aberrations are drowned out by an even stronger speckle pattern, the so-called *seeing*. These constantly shifting aberrations are caused by turbulence and density fluctuations in the layers of Earth's atmosphere, resulting in a characteristic "boiling" of the speckle pattern on a time scale of $\tau_0 \sim 10\,\mathrm{ms}$.

Unlike the diffraction limit, the effect of aberrations can be mitigated by *adaptive optics* (AO). First, a wave-front sensor (WFS) must measure the momentary shape of the aberrations. A deformable mirror powered by an array of piezo-elastic actuators can then reshape itself in such a way to cancel out the wavefront errors to some degree. The feedback cycle between WFS and mirror must be significantly faster than the time scale of the speckle evolution. The more actuators are available, the more spatial Fourier modes of the aberrations can be corrected.

An adaptive optics system improves a speckle-degraded image by restoring a part of the stray light back into the Airy pattern. The reduction of the stray light halo is effective only within an area around the central peak described by the so-called control radius, a function of the number of actuators used by the AO system. The reconstructive power of AO is illustrated in Figure 2.2b,c.

More information about adaptive optics can be found e.g. in Roddier (1999).

2.4 Differential Imaging

For high-contrast applications, the PSF of a star after AO correction may still be too "bumpy" to reveal a faint signal in its vicinity, such as a planet or the surface details of a circumstellar disk (Racine et al., 1999). Differential imaging offers a way to calibrate such background structures away by exploiting a property of the target object's light that sets it apart from the residual PSF of its parent star. The principle is to obtain two separate images featuring the same background landscape, only one of which however containing the desired object. The two images can then be subtracted from one another, canceling away the identical backgrounds and revealing the unique object. Figure 2.3 illustrates this principle.

The discriminating properties that have been successfully employed for differential imaging in astronomy include spectral features (Marois et al., 2000), polarization
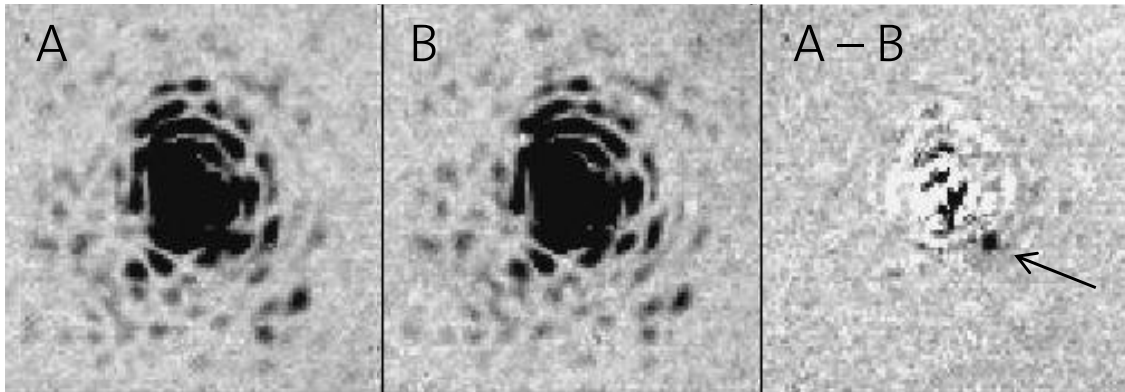
Fig. 2.3: A simulated observation with differential imaging. Two images of the same target star are obtained in two filter bands A and B. The signal of a faint companion star is present in A but not in B, since the spectral range of B is chosen to lie in an absorption band of such a companion's spectrum. In the difference image $A - B$, the companion signal becomes visible (arrow). The residual background structures in $A - B$ are due to differential aberrations between the two image channels. These images are taken from Racine et al. (1999).

(e.g. Kuhn et al., 2001), and frame of reference (angular differential imaging, Marois et al., 2006). Section 3.5 describes the methods used in this work in more detail.

The effectiveness of differential techniques is limited by the fidelity with which the background can be exactly reproduced in the two images. Any aberrations or impurities that affect one image differently than the other, so-called *differential aberrations*, will produce a spurious signal in the difference image, and must therefore be avoided as far as possible.

2.5   Coronagraphs

The French astronomer Bernard Lyot invented the coronagraph in 1930 for the purpose of imaging the solar corona. This had previously only been possible during eclipses, where the moon would occlude the bright disk of the photosphere and allow a glimpse on the solar atmosphere, which is fainter by many orders of magnitude. The intensity of direct photospheric light would otherwise burn out the telescope's optics. A coronagraph, then, is a device intended to replace the moon in its role of blocking the sunlight.

The concept of the solar coronagraph has been adapted for use in night-time astronomy (stellar coronagraph), allowing circumstellar objects much fainter than their parent star to be observed at long exposure times while avoiding saturation of the detector. Such objects include circumstellar disks, cool companion stars, brown dwarfs and—of particular interest in the scope of this work—extrasolar planets.

2.5.1 Classical Lyot coronagraph

The working principle of a Lyot coronagraph is sketched in Figure 2.4. The target star is assumed to be aligned with the optical axis of the telescope.

The primary component is a circular, opaque disk, called the Lyot mask, which is placed into the primary focus. The core of the stellar PSF is imaged onto the center of the mask, where its light is absorbed and thus removed from the optical path. This technique appears quite straightforward; however, the diffraction rings lying outside of the masked area still convey a significant fraction of the total starlight, hindering the detection of faint circumstellar objects.

In a second step, the light transmitted beyond the primary focus is relayed into a second pupil plane. The pupil image, a Fourier transform of the masked focus image, is now no longer flat as it was at the primary pupil. The remaining starlight is distributed according to a high-pass filtered image of the pupil, with most of the light concentrated in a ring at the edge of the relayed primary pupil aperture. Another opaque mask, the so-called Lyot stop, is placed in the pupil. Unlike the disk-shaped focal mask, the pupil stop takes the shape of a circular aperture in an opaque plane. Its radius is chosen to absorb the main annulus of the diffracted starlight. Only the small fraction of starlight within the aperture is transmitted and finally focused onto the detector.

On the other hand, any circumstellar object separated far enough from the central star will survive the passage through the coronagraph mostly intact. Its PSF core, containing most of its light, will miss the focal mask. The resulting pupil image will be flat rather than annular, thus again the greatest part of the light will not be absorbed by the pupil stop.

If the telescope features a central obscuration in the pupil (a) from its secondary mirror, the residual light in the post-mask pupil (d) will gather along the inner edge of the imaged pupil aperture as well as the outer one. The pupil stop must therefore also cover the imaged secondary mirror in order to attenuate the starlight effectively.

The radius of the Lyot mask is often chosen to correspond to a few resolution elements λ/D in the projected image. A larger mask has the advantage of rejecting more starlight, and the disadvantage of blocking a larger area on the field of view. A compromise must be made according to the expected brightness and angular separation of the target object.

2.5.2 Apodized Lyot coronagraph

Although Lyot's original design is still used in the present day, there have been a number of efforts to improve its performance. One idea (Soummer et al., 2003) is to place an apodizer, a partially transparent plate with a very specific radial transmissivity profile, into the entrance pupil. The pupil image is thereby shaped into a so-called truncated Prolate function, which approximates the Fourier transform of the focal mask as well as possible. In the relayed pupil after the application of the focal mask, the two functions will interfere destructively, leaving very little residual starlight inside the imaged telescope aperture.

pupil-plane images

focal mask

pupil stop
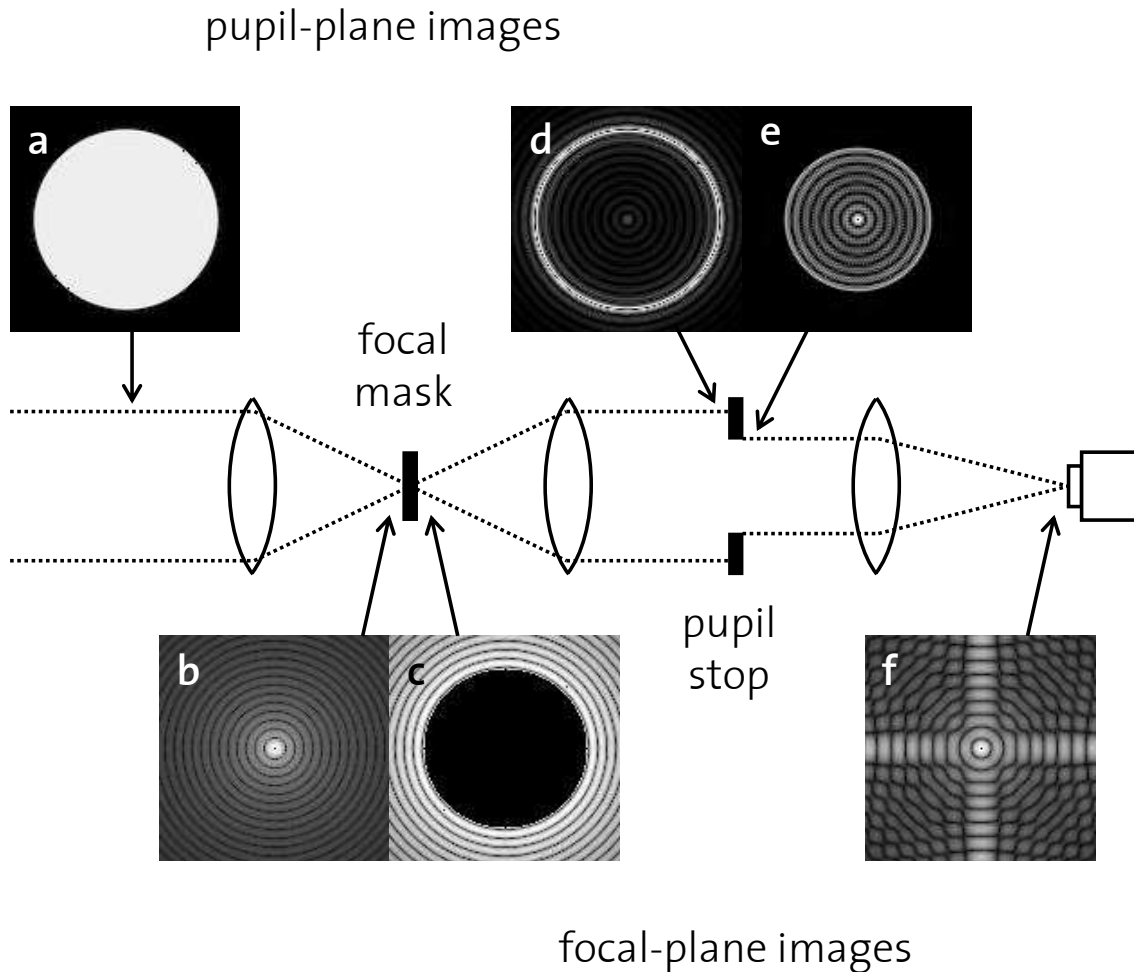
focal-plane images

Fig. 2.4: Concept of the classical Lyot coronagraph. The light of the central star arrives at the telescope pupil as a flat plane wave (a), which is then imaged into the focal plane, yielding a diffraction pattern (b). The Lyot mask absorbs core of this PSF, leaving the outer diffraction rings (c). The light is transformed back into a pupil plane (d), where it now concentrates along the outer edge of the imaged telescope aperture. The Lyot stop blocks that area, rejecting most of the remaining light (e). The final image in the focal plane contains only a small fraction of the original starlight (f). This example assumes a telescope without a central obstruction in the pupil.

Fig. 2.5: Left: A sketch of the distribution of retardances δ of a four-quadrant phase mask, and the star's PSF centered on it. **Right:** The light distribution at the pupil plane after passage through the 4QPM. Destructive interference has diverted all light outside the imaged telescope aperture (black disk), where the pupil stop can absorb it. This example assumes a telescope without a central obstruction in the pupil.

2.5.3 Four-quadrant phase mask

An even more exotic alternative to the classical Lyot coronagraph pioneered by Roddier & Roddier (1997) and Rouan et al. (2000) is to replac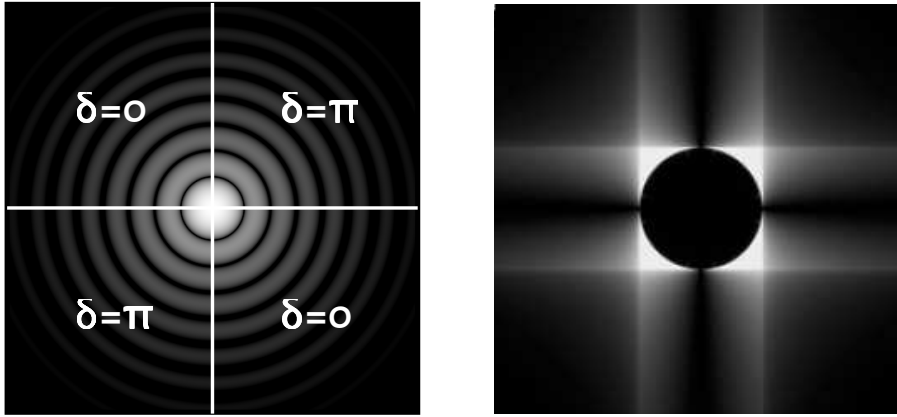e the opaque focal mask by a phase plate which manipulates the starlight in such a way as to cause destructive interference. Several different phase plate geometries have since been proposed. In one particular design, known as the monochromatic four-quadrant phase mask (4QPM) coronagraph, the phase mask is an assembly of four plates of unequal thickness arranged as quadrants, whose induced phase delays alternate between zero and half-wave across each boundary. The target star's PSF is centered on the assembly, such that its complex amplitude is reversed in two of the four quadrants. In the subsequent pupil plane, the two phase-shifted halves of the stellar flux assume the same distribution, cancelling each other out. The entire starlight is diverted outside of the imaged telescope aperture, where it is absorbed by the pupil stop. This is illustrated in Figure 2.5.

In a perfect system, a source placed on the center of such a phase plate is completely eliminated, whereas other sources whose PSF cores lie completely within one quadrant remain unscathed (Abe et al., 2003). However, the design is very sensitive to asymmetries in the PSF shape (as caused by aberrations and atmospheric speckles), and to ever so slight misalignments of the stellar PSF on the plate center. Finally, due to the wavelength-dependent phase delay of the plates, a monochromatic 4QPM will perform suboptimally in a broad-band observation.

Despite those vulnerabilities, the four-quadrant phase mask (4QPM) offers the unique possibility to image faint companions essentially as close as a single resolution element from the star. A Lyot coronagraph would require a focal mask radius as small as $1\,\lambda/D$ to achieve such a small inner working angle, and would perform very poorly due to the large amount of starlight surviving the focal mask stage.

Chapter 3

The SPHERE Project

3.1 History

As one of the second-generation instruments for the Very Large Telescope (VLT) in Paranal, Chile, the European Southern Observatory (ESO) commissioned the development of a "planet finder", a high-contrast differential imager capable of detecting exoplanets by direct imaging.

Two Phase A studies were conducted by independent consortia to come up with an innovative yet feasible and robust design that would use the scientific potential of the telescope to its fullest. One of them, named CHEOPS (CHaracterizing Exoplanets by Opto-infrared Polarimetry and Spectroscopy), was led by the Max Planck Institute of Astronomy in Heidelberg (MPIA) and included German, Italian, Dutch, Portuguese and Swiss scientists, including the institute of astronomy of ETH Zürich. The other consortium united several institutes under the leadership of the Laboratoire d'Astrophysique de Grenoble (LAOG).

At the conclusion of this phase, ESO saw elements of merit in both design studies, and decided to fuse them both into a single synthesized project, named SPHERE (Spectro-Polarimetric High-contrast Exoplanet REsearch), combining the superior adaptive optics, coronagraphs and the near-infrared spectral differential imager of the consortium led by LAOG with the unique visible-light imaging polarimeter and integral field spectrograph of CHEOPS.

In Phase B, the preliminary design of the instrument was developed. After the successful conclusion of the preliminary design review at ESO in September 2007, the project is now in the final design phase (Phase C). The instrument is expected to be installed at the telescope towards the end of 2010, and to see first light in early 2011. A good overview of the project is given in Beuzit et al. (2006).

3.2 Purpose

Having been conceived as a planet finder, SPHERE's primary goal is indeed the discovery and study of new extrasolar planets by direct imaging of their circumstellar environment. Since the main obstacle to that goal is the high brightness contrast between planet and star in conjunction with their small angular separation, the instrument is designed to reach the maximum attainable imaging contrast in the close vicinity of a target star. An extreme adaptive optics system, coronagraphs,

and several differential imaging methods are employed to reject enough of the light in the stellar seeing halo to render the faint signature of a planetary companion visible.

SPHERE will probe the planet population of the stellar neighborhood in two complementary ways:

The spectral differential imagers IRDIS and IFS will survey several hundred stars for young self-luminous planets aglow with the internal heat of gravitational contraction. Due to the large number of targets, a significant sample of planets in various stages of evolution will be discovered, providing the highly active scientific fields of planet formation and migration with much-needed observational data. Since the direct imaging method aims at a different planet population than Doppler spectroscopy and transit photometry, this sample will circumvent some of the observational biases of the current exoplanet catalogue. In particular, SPHERE will be able to determine the abundance of massive planets in wide orbits, a currently unknown demographic of the exoplanet population, and shed light on the distribution of planets in the mass/orbital period parameter space.

This new window on planet formation and migration has the potential to clear away many of the open issues and uncertainties of current theories concerning the birth and interaction of planets, and to answer the fundamental questions of how common planets are in the universe and how unique our own Solar System is.

The differential imaging polarimeter ZIMPOL will search the closest bright stars for signs of old, cooled planets. Since those no longer emit a measurable intrinsic radiation in the visual range, they must be detected by the starlight reflected off their atmospheres and surfaces. This light is much fainter than the glow of young planets: The brightness contrast to their parent stars is typically on the order of 10^{-8} rather than 10^{-5}. Furthermore, the contrast deepens with increasing physical distance between star and planet.

However, light reflected off a gas giant's atmosphere acquires a certain degree of polarization through Rayleigh scattering (the same effect that causes the blue color and polarization of the sky on Earth). Planetary atmosphere models (Stam et al., 2004) and observations of the gas planets of our Solar System (Joos, 2007) suggest a range of 10–50% of polarization for scattering at a right angle. This provides the fundamental possibility to discover the presence of an old planet by direct imaging of its polarized flux amidst the unpolarized scattered-light halo of its parent star with a sufficiently sensitive imaging polarimeter. The work presented in Thalmann (2004) confirms that SPHERE's ZIMPOL is suited to this task.

While the target sample available for polarimetric detection is tiny compared to the hundreds of targets intended for the spectral survey, such a discovery would be unique in astrophysics and provide direct information about the planet's visible surface properties.

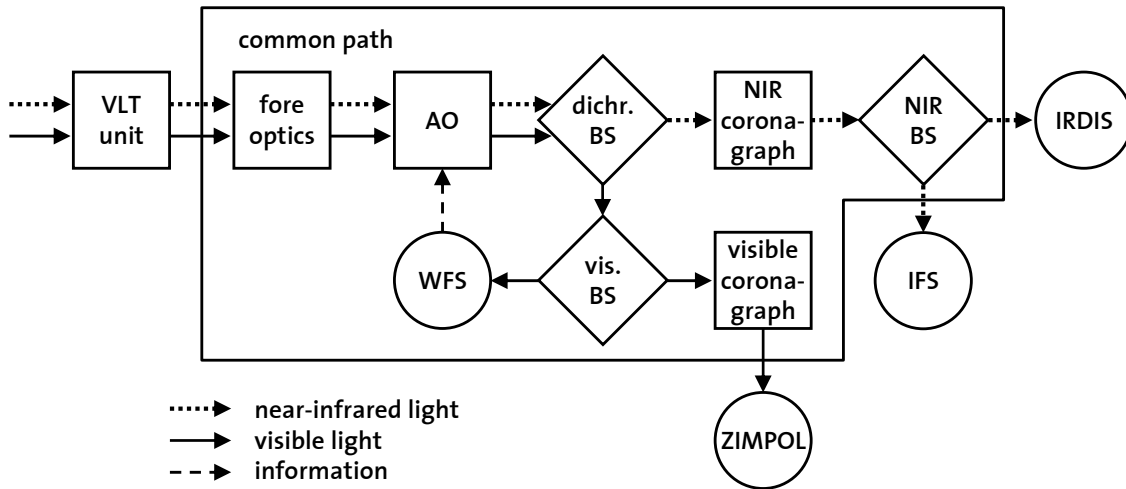This work concentrates on the ZIMPOL branch of the SPHERE instrument.

Fig. 3.1: Overall concept of the SPHERE instrument design. The common path comprises all optics upstream of the three scientific detectors IRDIS, IFS and ZIMPOL. A dichroic beamsplitter (BS) separates the near-infrared light from the visible; these beams are further distributed among the detectors by two more beamsplitters. The two spectral bands each have their own coronagraph.

Aside from the primary mission of planet finding through direct imaging, SPHERE's high contrast performance will enable a plethora of observational possibilities to other fields of astrophysical research, including debris disks and protoplanetary disks—whose edges, gaps and waves can point to the presence of unseen planets—, brown dwarfs, evolved massive stars (such as red giants), and even Solar System science. While these secondary applications add much weight to the importance of the instrument in the astrophysical research effort of the near future, they do not drive the system requirements of the project, which are dictated entirely by the challenge of planet finding.

3.3 Instrument Design

The SPHERE instrument is divided into four basic components: The three scientific detectors IRDIS, IFS and ZIMPOL, and the common path, an umbrella term for all other optics between the telescope interface and the scientific detectors. Figure 3.1 shows a schematic of the instrument design and its data flow.

The working principle and implementation of the ZIMPOL detector, being central to this work, merit their proper section (3.4). The other three basic components are briefly described below, on the basis of Beuzit et al. (2006).

3.3.1 Common Path

Directing the flow of light onto the science detectors is but the simplest of functions fulfilled by the common path optics. Among the fore optics upstream of the AO system are a folding mirror designed to counteract the instrumental polarization caused

by the Nasmyth folding mirror (M3), and two elements particularly important to ZIMPOL observations for calibration purposes.

A half-wave plate (HWP2) mounted on a rotation stage can be used to reverse the sign of the polarization of the light beam, allowing a form of beam-exchange calibration (subsequently referred to as "signal switching").

Further downstream, a three-mirror field derotator can physically reorient the coordinate system of the sky field image by a given angle. If this component is deactivated, the motion of the altitude-azimuth mounting of the VLT unit telescope while tracking a target on the continuously spinning globe of the sky will result in a rotation of the field image over time. This mode of operation is called "P1" for ZIMPOL. The derotator can be activated to counteract the natural field rotation, resulting in a constant field orientation over the whole observation. This is called field stabilization or P2 mode. Finally, to avoid smearing during long exposures in P1 mode, the derotator can be set to compensate the differential field rotation during each single exposure of the science detector, but to reset to its starting position between exposures. This is known as P3 mode. The derotator allows the field orientation to be arbitrarily changed for the purpose of angular differential imaging.

Then follow the various corrective components of the adaptive optics system, called SAXO (SPHERE Adaptive optics for eXoplanet Observation). It uses a deformable mirror with 41×41 actuators of $180\,\text{mm}$ diameter with inter-actuator stroke $> \pm 3.5\,\mu\text{m}$, and a two-axis tip-tilt mirror with $\pm 0.5\,\text{mas}$ resolution (Fusco et al., 2006). Further downstream in the visible branch of the optical path, a Shack-Hartman wavefront sensor (WFS) with 40×40 lenslets operating in the 0.45–$0.95\,\mu\text{m}$ band at a temporal sampling frequency of $1.2\,\text{kHz}$ using a 240×240 pixel electron-multiplying CCD detector feeds the corrective optics with real-time information, enabling a global AO loop delay below $1\,\text{ms}$.

A dichroic beam splitter separates the visible from the near-infrared (NIR) part of the light, allowing the wavefront sensor to use the full visible flux during NIR observations. Further beamsplitters divide the beams between the WFS and ZIMPOL, and IRDIS and IFS, respectively.

Finally, the common path includes one full coronagraph assembly for each of the two spectral bands. Each features selection wheels that allow the focal mask and pupil stop to be chosen from a set of Lyot or four-quadrant phase mask (4QPM) components. The coronagraphs attenuate the central peak of the target star's image, greatly reducing the total brightness contrast in the image and thereby allowing longer frame exposure times, while at the same time purging the diffraction features caused by the propagated pupil edges from the image.

3.3.2   IRDIS

In its default observing mode, the InfraRed Dual-beam Imager and Spectrograph (IRDIS) simultaneously records a picture of the target object in two neighboring spectral bands. The light is divided into two beams by a beamsplitter, which are then passed through a matching pair of spectral filters and registered by two sections

of an infrared image detector. By rescaling the images according to wavelength and subtracting them from each other, the background of residual starlight is effectively suppressed. The filter bands are chosen such that one lies within an expected absorption band of an exoplanet's atmosphere, which is absent from the parent star's spectrum. As a result, the planet is visible only in one of the two filter bands, surviving the subtraction unscathed. The two pictures are recorded by separate detector areas which each receive half of the total light from a beamsplitter. The speckles affect both pictures in the same way and therefore cancel out in the subtraction. The differential aberrations between the two channels are kept below 10 nm RMS by using high-quality polishing.

IRDIS features 10 complementary filter pairs for differential imaging and 12 identical filter pairs for simple imaging, covering the range of 950–2320 nm. The field spans more than $11''$ at a pixel scale of 12.25 mas. Furthermore, two long-slit spectroscopy modes are foreseen with resolving powers of 50 and 500 (Dohlen et al., 2006). A simple imaging polarimetry mode is also available.

3.3.3 IFS

While the Integral Field Spectrograph (IFS) is also based on simultaneous spectral differential imaging, its technical implementation is quite different from IRDIS. Instead of the dual-beam approach, a single beam is imaged onto an array of microlenses with a grid scale of one resolution element. The microlenses are shaped such as to focus the light of their resolution elements into small, parallel spectra on the subsequent near-infrared detector. From the picture, a set of diffraction-limited images in various wavelength intervals can be extracted. Comparing these images allows efficient background estimation and subtraction, revealing any localized spectral features (such as planetary absorption bands) that do not match the stellar spectrum. The differential aberrations are negligible; however, the technical implementation of this concept is challenging.

The IFS design for SPHERE is modified from the classical TIGRE concept for the purpose of high-contrast, diffraction-limited observations (Lee et al., 2001). It will operate in the J and H bands (0.95–1.70 μm), allowing parallel usage of IRDIS and IFS. The planned spectral resolution is 50 for 2 pixels in the J mode, and 30 in $J + H$. The field of view will span $1''\!.77$.

3.4 ZIMPOL

To understand the significance of the ZIMPOL instrument concept, the inherent problems of imaging polarimetry must be considered. Astronomical cameras (CCDs) excel at measuring intensity and are often able to count individual photons, but cannot discern their polarization state. Therefore, polarization information must be converted into an intensity information in order to be measurable. This can be achieved in two ways.

3.4.1 Dual-beam polarimetry

The most straightforward design for an imaging polarimeter is the dual-beam setup. The light is passed through a beamsplitter, dividing it into two beams carrying complementary linear polarization components. The beamsplitter acts like a matching pair of Stokes filters, such as $\pm Q$ (see Section A.2). The two beams are projected onto separate CCD sensor areas, and their intensity images $I_+ = \frac{1}{2}(I + Q)$, $I_- = \frac{1}{2}(I - Q)$ are recorded. The Q/I polarization image can then be reconstructed as the difference between the two images renormalized by the total intensity (given by the sum of the two images):

$$P(x,y) = \frac{Q(x,y)}{I(x,y)} = \frac{I_+(x,y) - I_-(x,y)}{I_+(x,y) + I_-(x,y)}. \tag{3.1}$$

The pixel values $P(x,y)$ will then be restricted to the range $[-1, 1]$. In particular, a pixel with unpolarized light will have $P(x,y) = 0$, whereas light fully polarized along Stokes $\pm Q$ will result in $P(x,y) = \pm 1$. A similar image of Stokes U/I or V/I polarization can be obtained by rotating the beamsplitter by $45°$ or placing a quarter-wave retarder plate before the beamsplitter, respectively.

Any optical components that produce polarization signatures of their own, such as filters or field derotators, can be safely placed after the beamsplitter, since the scientific information is encoded as an intensity signal at this point.

The inherent flaw of dual-beam imaging polarimetry is the fact that the two component images are relayed through non-common optical paths and measured by different CCD sensor areas, giving rise to differential instrumental effects. The individual pixels of a CCD are not perfectly equal in performance and behavior. These defects can depend non-linearly on various factors like temperature, intensity or gradients in the local intensity distribution, and are therefore difficult to correct through flatfielding. This limits the attainable polarimetric accuracy.

3.4.2 Single-beam polarimetry

These differential instrumental effects can be avoided with a single-beam setup. Rather than spatially separating the two polarization components into beams, they are encoded as a time series. The light is passed through a variable retarder— either a rotating retarder plate or an electro-optical device whose retardance can be varied—, analyzed by a fixed polarizer, and focused onto a CCD camera. The intensity image transmitted to the camera then contains the polarization information as a temporal modulation, oscillating between I_+ and I_-. By recording a time series of the image and disentangling the two component images, the polarization image can be reconstructed as in Equation 3.1.

Since both component images are registered by the same CCD pixels, the differential instrumental effects are greatly minimized. However, since the images are not acquired simultaneously but serially, the atmospheric speckle pattern changes from frame to frame, leaving differential structures in the polarization image.

These temporal differential effects can be minimized by increasing the modulation frequency. In order to demodulate the encoded information, the exposure time
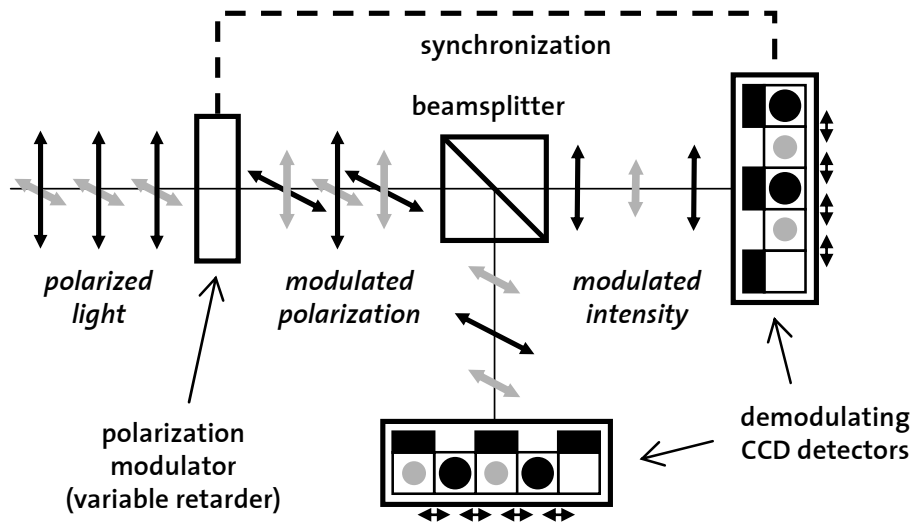
Fig. 3.2: A schematic of the ZIMPOL detector concept. Note that a single CCD camera would suffice for a fully functional imaging polarimeter; the second camera merely ensures that no photons are wasted.

of the CCD sensor must be reduced accordingly. However, the read-out of a CCD frame takes a finite amount of time, providing a hard upper limit to the modulation frequency. Furthermore, for very short exposures, the photon count per pixel drops to single-digit numbers, comparable to the electronic read-out noise and dark current variations of the sensor. Therefore, such a polarimeter design cannot be built to modulate faster than the atmospheric seeing time scale of $\sim 10\,\mathrm{ms}$.

3.4.3 The ZIMPOL Principle

The Zürich Imaging Polarimeter (ZIMPOL) is a technique invented by H.P. Povel at ETH (Povel et al., 1990, 1994; Povel, 1995) to bypass the modulation frequency problem of the single-beam imaging polarimeter as described above. Using *fast polarization modulation* and real-time on-sensor demodulation with a special CCD camera, the two polarization images are actually recorded with the same pixels "at the same time", i.e. at modulation frequencies in excess of the rate at which the speckle pattern changes. Implementations of the ZIMPOL concept have been applied in scientific research to great success, particularly in the field of solar physics (e.g. Stenflo, 1996). Examples of applications in night-time astronomy are found e.g. in Gisler & Schmid (2003) and Joos (2007).

ZIMPOL is based on a classical single-beam setup with a fast electro-optical polarization modulator. In order to defeat speckle noise, the modulation must be faster than the time scale on which the speckles evolve, i.e. in the kHz range. As mentioned before, CCD sensors cannot be read out at this high rate.

The ZIMPOL concept solves this problem with a special sensor design that performs real-time demodulation on sensor itself. Every other pixel row of the CCD sensor is covered by an opaque mask that effectively "blinds" those pixels, pre-

venting them from generating photoelectrons. These pixels can be used to safely buffer the photoelectrons gathered from one polarization component image while the neighboring exposed pixels are recording the other component image. The two images are effectively stored in interlaced sets of pixel rows, only one of which is sensitive to the incoming light. As the modulator alternatingly sends the $+Q$ and $-Q$ image towards the detector, the global charge map is alternately shifted upwards or downwards by one row, placing the previously active image data in the inert masked rows and instead exposing the previously passive image data to the open pixels. Note that the hardware pixels and the opaque mask remain fixed; only the photoelectrons representing the accumulated image data are transfered from pixel to pixel.

This way, the sensor can integrate both component images quasi-simultaneously over normal frame exposure times (e.g. one second), using the *same detection pixels* for both images, until the exposure is stopped with a shutter and the charges are read out of the CCD sensor in due time. The vertical shifting of the charge map by one line is quick enough to allow for modulation frequencies in the kHz range. The speckle noise issue of the conventional single-beam setup is effectively defeated. Figure 3.2 illustrates this working principle.

Using a polarizing beamsplitter in the role of the polarizer allows the rejected component beam to be projected onto a second demodulating CCD sensor in order not to waste half of the incoming light. This does not turn the ZIMPOL system into a dual-beam design; rather, each sensor acts as a fully functional single-beam imaging polarimeter. This redundancy can be exploited to calibrate residual instrumental effects, or to zoom in on two different areas of interest in the field, for instance.

By employing two serial polarization modulators and a CCD sensor with three masked pixel rows for every exposed row, the full Stokes vector can be measured simultaneously.

While ZIMPOL is remarkably free of differential flatfielding and calibration issues, laboratory tests for high-contrast duty in SPHERE (Roelfsema & Gisler, 2007) have revealed a more subtle weakness of the system: Polarized differential aberrations. Switching the FLC modulator between its two retardance states also slightly changes its other optical properties. A slight state-dependent wedge effect that affects the pointing of the transmitted beam ("differential pointing") as well as some higher-order phase-like differential aberrations have been found. Although these effects are very small (of the order 1 nm RMS), they constitute the primary limiting factor to the polarimetric precision of ZIMPOL in high-contrast applications.

3.4.4 Implementation in SPHERE

The ZIMPOL detector of SPHERE is designed to work in the V, R and I bands and is optimized for 600–900 nm. The role of the fast polarization modulator is played by a ferro-electric liquid crystal (FLC) whose crystal elements can be globally reoriented by applying a directed voltage, thereby changing the refractive ellipsoid of the material in a well-defined fashion. In effect, it acts as a half-wave plate whose fast and slow axes can be rotated by 45°. The time scale of this retardance switching
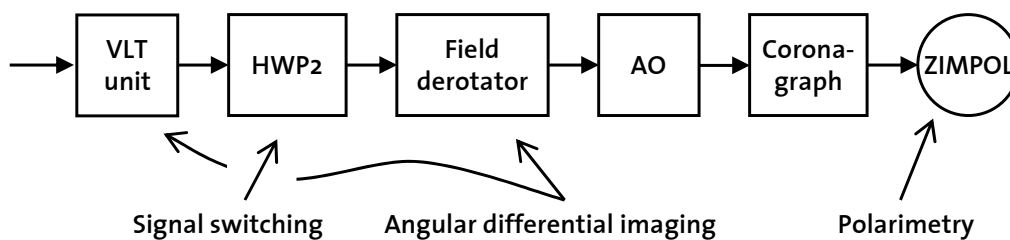
Fig. 3.3: A simplified scheme of the SPHERE/ZIMPOL instrument pointing out the key components for the three differential imaging techniques. Angular differential imaging can be based on the "natural" field rotation caused by the telescope tracking, or on artificial "active" field rotation performed by the derotator optics.

is short enough to allow a modulation frequency in excess of $1\,\mathrm{kHz}$.

The SPHERE/ZIMPOL system is not designed for full Stokes polarimetry; only one linear polarization (IQ) is measured at once. A single modulator is used, and the CCD features one masked pixel row for every exposed pixel row.

The field of view will be a square of $1.5''$ that can be shifted around the target star within a field of $4''$ radius. A polarimetric accuracy of at least 10^{-5} has been attested to the system in laboratory tests (Thalmann, 2004), and 10^{-6} appears feasible.

Aside from the differential imaging polarimetry, the ZIMPOL camera will also provide high-resolution direct imaging in a variety of broad- and narrow-band visible filter ranges, filling the important niche of imaging circumstellar regions at high contrast and high angular resolution.

3.5 Differential Imaging Methods with SPHERE/ZIMPOL

In order to meet the daunting constrast requirements of direct planet imaging, the ZIMPOL branch of SPHERE employs several layers of differential imaging. The raw intensity image of the coronagraphically masked target star will be dominated by speckle-like noise caused by the static and quasi-static aberrations of the instrument, about four orders of magnitude above the level where a planet signal could be detected.

The requirements that instrumental polarization must be kept low and that measurements of the absolute degree of polarization must be possible impose restrictions on the optical design and therefore on the differential methods available.

Figure 3.3 shows the components of the instrument responsible for each differential method. Section 5 will treat these quantitative aspects in greater detail.

Polarimetry

The most obvious method employed in ZIMPOL/SPHERE is imaging polarimetry: By taking an image of two complementary polarization states and subtracting them from each other, the unpolarized component of the image, including all the starlight,

should be removed. In practice, polarized differential aberrations between the two polarization channels will limit the efficiency of this subtraction, leaving behind an artificial background landscape in the polarization image. Nevertheless, the S/N gain brought about by polarimetry is on the order of 10^2.

Signal switching

To calibrate this spurious polarization background away, a second level of differential imaging is employed. By rotating a half-wave plate (HWP2) far upstream in the optical path by 45°, the sign of the incoming Stokes Q polarization is reversed. The instrumental aberrations, on the other hand, remain unchanged, resulting in the same background landscape as before. If the polarization images before and after the signal switching are subtracted from one another, the real polarization signals of the astronomical target add up constructively while the static background is canceled out.

Again, this method is only effective down to the level at which the background can be reproduced in the second image; any change to the optical path on the time scale of the signal switching (i.e. temporal differential aberrations) will lead to residual structures surviving in the double-difference image, limiting the detectivity of faint signals. The half-wave plate is foreseen to be switched about every 5 minutes during observation.

This calibration method is essentially equivalent to the conventional subtraction of a reference PSF (taken from another star) from the target image; the half-wave plate switching is simply an elegant solution to obtaining a reference PSF with a minimum of differential effects.

Angular differential imaging

A third level of noise suppression is achieved through angular differential imaging. Over the course of the observation, the VLT unit telescope's tracking will cause the observed sky field to rotate slowly with respect to the detector, whereas the optical elements responsible for aberrations will remain fixed in the instrument's frame of reference. This property can be exploited to discern the real signals originating in the sky from spurious instrumental effects through so-called angular differential imaging (Marois et al., 2006).

A median average of all exposures of the observation run is calculated and subtracted from each individual exposure, removing all static parts of the image. Since the planet moves across the field in an arc during the observation, it does not leave a strong footprint in the median background and therefore survives the background subtraction. The exposures are then rotated into the sky's frame of reference such that the planet signals coincide and add up constructively, whereas any residual background noise is averaged down. In cases where the natural field rotation is insufficient, it can be enhanced by artificially changing the image orientation by means of the derotator optics in discrete steps.

With all three differential methods applied, the noise can likely be suppressed down to the statistical photon noise for a four-hour observation.

Chapter 4

Real-Time Polarization Compensator

4.1 Motivation

The depth of the footprint of an exoplanet's reflected light in the polarization image is typically on the order of 10^{-4}. It is embedded on a thick layer of background polarization of up to several 10^{-2}, caused by reflections on telescope mirrors and the polarization of the sky field. Since this background is quite flat over the field of view, it can be subtracted and therefore in itself poses no direct threat to planet detection. However, the effects of some noise sources like detector non-linearities or calibration errors scale in proportion with the background polarization level. Using software to subtract the flat polarization offset later at the reduction stage does not undo these noise effects; they can only be avoided by removing the background polarization from the light *before* it reaches the polarimeter. This is a problem of high-contrast polarimetry in general.

A particular sensitivity of the ZIMPOL system to background polarization is related to the detector's fixed-pattern noise. As mentioned in Section 3.4, the instrument is able to correct most of the fixed-pattern noise introduced in the charge-shifting CCD detectors by operating in "two-phase mode". While this method greatly improves the polarimetric accuracy of the instrument under ideal conditions, its performance quickly degrades if those conditions are not met.

In two-phase mode, the fixed-pattern noise in one particular image pixel is eliminated by reversing the role of the two adjacent buffer pixels between each exposure, and then subtracting the resulting polarization images from one another. The assumption is that the $+Q$ photocharges buffered in the upper masked pixel during one exposure will suffer the exact same noise effects as the $-Q$ photocharges buffered in the same masked pixel during the next exposure. However, these noise effects (such as charge trapping or leaking) often depend on the number of stored photocharges in a highly non-linear fashion. Therefore, the correction will only function as intended if the $+Q$ and $-Q$ intensities are equal, i.e. the incident light is unpolarized. Laboratory experiments (Thalmann, 2004) have shown that even a background polarization on the order of 1% causes a distinct increase in residual noise in the polarization image (see Figure 4.1).

Experience with ZIMPOL observations on various telescopes suggests that an unpolarized field is an exception rather than the rule. Even if the astronomical target itself is unpolarized, as is the case for the integrated light of the parent star
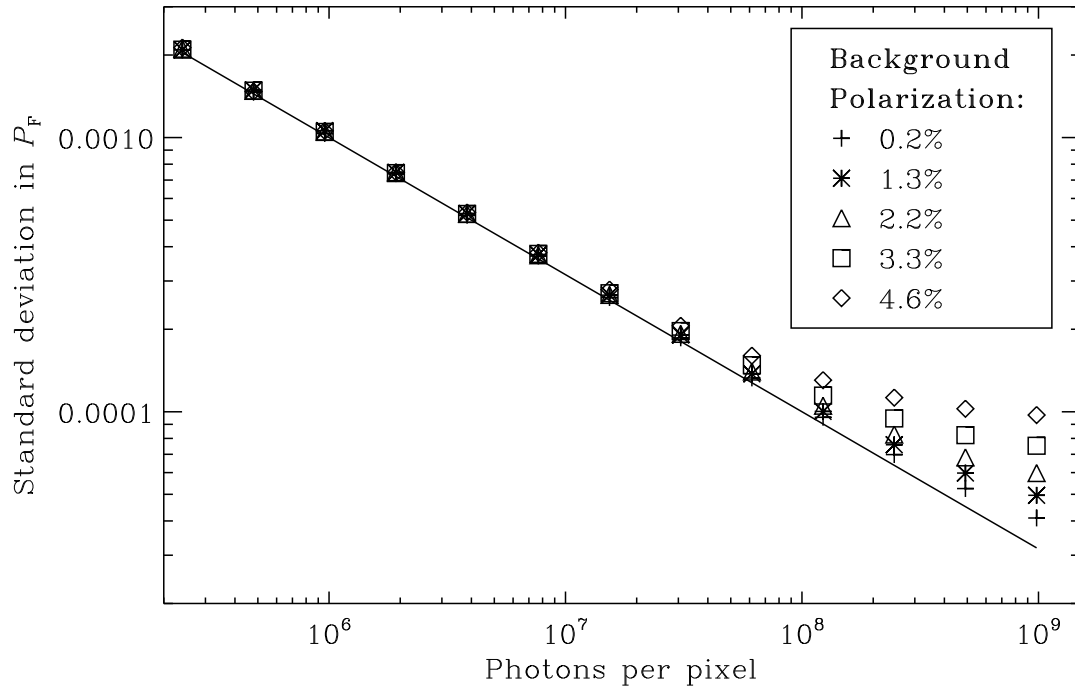
Fig. 4.1: Dependency of the ZIMPOL detector noise level on background polarization. As expected, a non-zero background polarization reduces the effectiveness of the two-phase noise suppression method.

when searching for exoplanets, the reflections at the telescope mirrors can impart a significant degree of linear polarization (a few percent) on the light. Depending on the telescope geometry, this instrumental polarization may evolve non-trivially over the course of the observation as the telescope tracks the target across the sky.

Due to these circumstances, we have constructed and programmed a device that can induce a linear polarization offset in the beam to compensate for an instrumental linear polarization of arbitrary orientation and strength (up to $\sim 15\%$), provided it is reasonably flat over the field of view. During the acquisition of science images, a background process constantly adjusts the compensation to adapt to fluctuations of the background polarization. Since only the zero-order term of the polarization image is removed, none of the spatially structured information is lost. In case the absolute polarization of the incoming light is scientifically relevant, it can easily be reconstructed from logged compensator parameters.

Section 4.2 describes the hardware setup, Section 4.3 the functionality of the runtime code, and and Section 4.4 the very satisfactory performance of the device in practical use. A polarization compensator of this design is foreseen as a component of the ZIMPOL instrument in SPHERE. It is also a desirable addition to other high-contrast polarimeters which suffer from non-zero background polarizations.
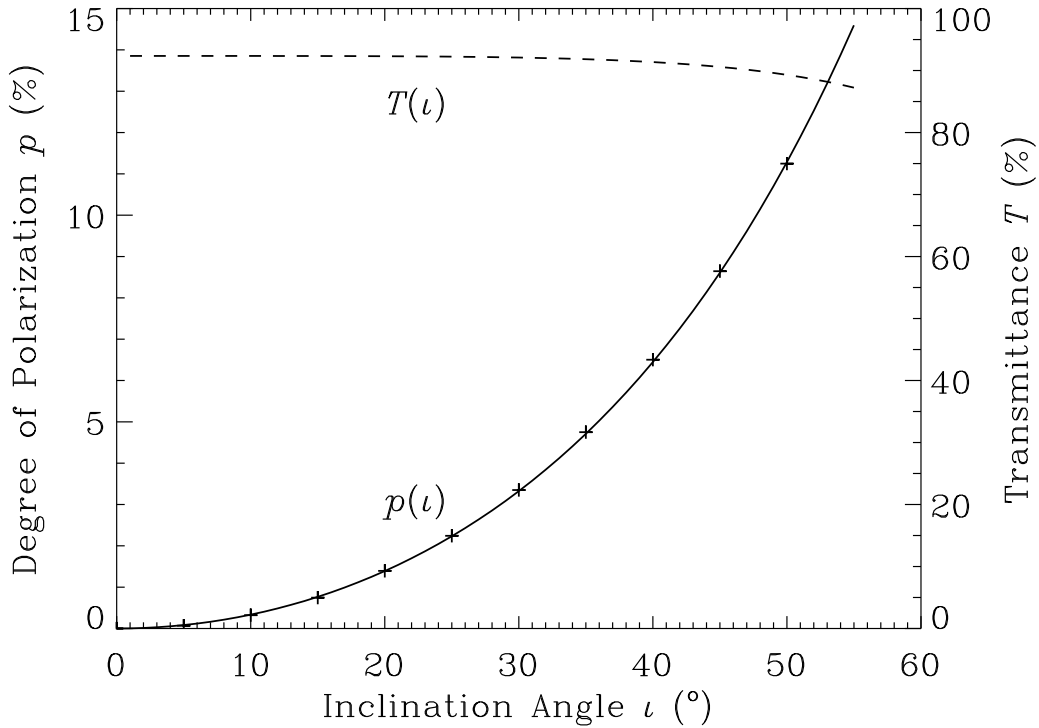
Fig. 4.2: The degree of polarization $p(\iota)$ and transmittance $T(\iota)$ as a function of the angle of incidence ι for a glass plate with an index of refraction $n = 1.49$. The crosses represent actual gauge measurements taken at IRSOL.

4.2 Hardware

4.2.1 Glass Plate

The compensator is a very inexpensive but robust low-tech solution based on a simple uncoated glass plate. Held into a light beam, the plate reflects a part of the incident light. If the incidence is not normal to the plate surface, the polarization component parallel to the tilt axis will preferentially be reflected. This polarized light is then "missing" from the transmitted light, leaving a polarization signature perpendicular to the tilt axis in the beam.

According to Collett (1993), an unpolarized beam incident on a glass plate at the angle ι will emerge on the other side with a degree of polarization $p(\iota)$ given by

$$p(\iota) = \left| \frac{1 - \cos^4(\iota - \rho)}{1 + \cos^4(\iota - \rho)} \right|, \tag{4.1}$$

where ρ is the angle of the refracted beam within the glass corresponding to the angle of incidence ι, which in turn follows from Snell's law, $\sin \iota = n \sin \rho$, where n designates the index of refraction of the glass. Comparison with a measured gauge curve (see Figure 4.2) of the actual glass plate used in this experiment yields a perfect fit for $n = 1.49$.

The polarization imparted by the transmission through the plate can be considered to stack additively with the pre-existing beam polarization if all involved degrees of polarization are small ($p \ll 1$).

Due to the necessary reflection, a part of the beam's intensity is lost. The transmittance $T(\iota)$ of the plate is given by

$$T(\iota) = \frac{1}{2} \left[\frac{\sin 2\iota \, \sin 2\rho}{(\sin(\iota + \rho) \cos(\iota - \rho))^2} \right]^2 \cdot (1 + \cos^4(\iota - \rho)), \qquad (4.2)$$

which yields $T(\iota) \approx 92\%$ for most of the relevant range of ι. In general, this loss of photons is preferable over the degradation of polarimetric accuracy due to background polarization.

4.2.2 Mounting

Our construction involves two motorized rotation stages that can be controlled by the computer running the ZIMPOL software. The primary stage is fixed vertically on the instrument bench such that its rotation axis coincides with the optical axis of the experiment. The secondary stage is mounted on the mobile ring of the primary stage with its rotation axis intersecting the optical axis at a right angle for all orientations of the primary stage. Finally, the glass plate is attached on the mobile ring of the secondary frame at a right angle to its rotation plane. The setup is sketched in Figure 4.3.

We use the convention of $+Q$ polarization as vertical and $+U$ as $\pi/4$ counterclockwise as seen from the detector's point of view. We refer to the orientation angle of the primary stage as the *azimuth angle* α, and that of the secondary stage as the *inclination angle* ι. Let $\alpha = 0$ designate the configuration where the foot of the secondary stage is at the lowest point of the primary's rotating ring, and let α increase when the primary stage is rotated counterclockwise as seen from the detector. Let the glass plate be perpendicular to the optical axis for $\iota = 0$.

The polarization generated by the plate in the configuration (α, ι) is then given by

$$\frac{Q}{I}(\alpha, \iota) = -p(\iota) \cdot \cos 2\alpha, \qquad \frac{U}{I}(\alpha, \iota) = -p(\iota) \cdot \sin 2\alpha. \qquad (4.3)$$

The inversion $(Q, U) \rightarrow (\alpha, \iota)$ is degenerate. For a given solution (α, ι), there are seven other configurations which produce the same polarization: $(\alpha, -\iota)$, $(\alpha + \pi, \pm\iota)$, $(\alpha, \pm\iota + \pi)$, and $(\alpha + \pi, \pm\iota + \pi)$. The reasons lie in the spatial symmetry of the glass plate, the π-periodicity of polarization (see Section A.2) and the even symmetry of the $p(\iota)$ function (Equation 4.1).

4.3 Compensation Strategy

The two motorized rotation stages of the compensator are driven by a software package written in the proprietary script language of the ZIMPOL system. The package is loaded into the ZIMPOL control and image acquisition software `zimpol2`,
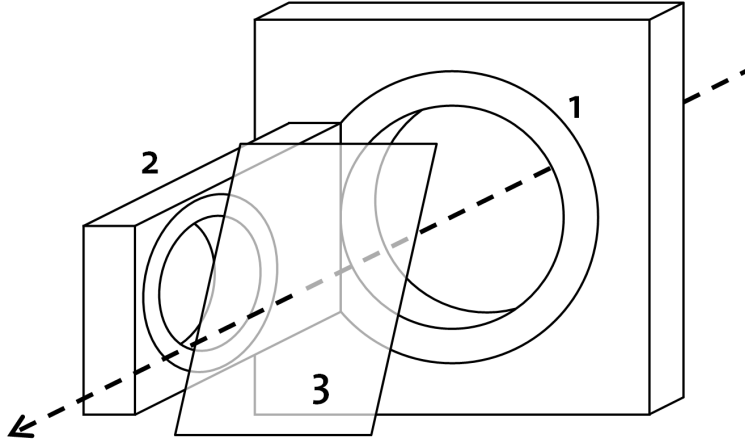
Fig. 4.3: Schematic of the compensator hardware. 1: primary stage, 2: secondary stage, 3: glass plate. The optical axis is marked with a dashed line. In the depicted configuration ($\alpha = -\pi/2$, $\iota \approx \pi/6$), the plate is contributing some vertical linear polarization ($+Q$) to the beam, i.e. it can be used to compensate an existing $-Q$ background polarization.

where the programs it offers can be launched with console commands. A detailed account of the internal and user-oriented routines provided by the package is given in Appendix B.

While some ZIMPOL configurations allow to measure the full Stokes vector at once (requiring two modulators in series, and a CCD detector with three masked buffer pixel rows for each exposed pixel row), simpler setups are more commonly employed. In particular, the SPHERE instrument design includes only a single FLC modulator and a single buffer pixel row for each exposed row, allowing only a single Stokes parameter aside from I to be measured at a given time. In the Kitt Peak observation run of 2005 (see Section 4.4), a piezo-elastic modulator (PEM) was used instead of the FLC, allowing the measurement of circular polarization together with one linear polarization component. Thus, in order to obtain a map of the full linear polarization, a typical observation series is split into a number of subseries, alternatingly measuring $IQ[V]$ and $IU[V]$.

The first instance of the compensator software tackled this problem by treating Q and U data acquisition separately. During Q measurements, the azimuth angle was set to either $\alpha = 0$ (to compensate a $+Q$ background) or $\alpha = \pi/2$ (to compensate $-Q$), disregarding the unmeasured U background; likewise, during U measurements, only the two configurations $\alpha = \pm\pi/4$ were used.

This "one-dimensional" compensation mode was successfully employed in the Kitt Peak observation campaign mentioned in Section 4.4. However, using different compensator settings in Q and U measurements bears an inherent risk of introducing differential effects. For instance, the unmeasured and uncompensated Stokes component in a can influence the other, measured components via crosstalks, complicating the procedure to remove spurious polarization from the data. Furthermore, a tilted glass plate imparts a spatial offset to the traversing beam, resulting in a slight misalignment between the Q and U images taken at different plate orientations.

For this reason, a second version of the software was developed. This "two-dimensional" compensation mode maintains an internal model of the Q and U background polarizations, and chooses its plate orientation such as to eliminate both components at once. Since only one component of the model can be updated in real-time during a given measurement subseries, the other component is assumed to remain constant during that time. This mode of operation has also been successfully tested on a telescope.

Two issues remain that need to be addressed in future versions of this code. Firstly, the range of α is currently enforced to the interval $-67.5° < \alpha \leq 112.5°$ in order to keep the conversion between angles and polarization states unambiguous, and to avoid tangling up the motor cables through unchecked rotation of the stages. However, this can cause an abrupt jump of $90°$ in α as the background polarization drifts across that threshold. It would be preferable to be able to freely select the desired arc for α prior to an observation. Secondly, the one-dimensional code entered a passive state without real-time correction if the background polarization dropped below the threshold of 0.1% in order to avoid an erratic and time-consuming behavior of α due to small fluctuations of p around zero. This feature has not yet been implemented in the two-dimensional code. Nevertheless, we consider the one-dimensional code obsolete and will henceforth concentrate on the two-dimensional code.

In some telescope designs, such as the one installed at IRSOL, the instrumental polarization evolves very slowly over a typical observation time. In this case, real-time adjustment of the compensator plate is not needed, and would in fact slow down the image acquisition process. It is therefore preferable to run the compensation software for a few iterations prior to the actual observation until a stable, optimal plate position has been found, and then to terminate the program, leaving the plate fixed during the acquisition of science images.

4.4   Performance

The one-dimensional compensation mode was successfully tested in the laboratory and at IRSOL, and put to active use during our observation run at the McMath-Pierce telescope on Kitt Peak, AZ, in March 2005 (Stenflo, 2006; Holzreuter & Stenflo, 2007a,b), where it was instrumental in attaining a high polarimetric accuracy. The residual background polarization was continuously kept well below 0.1%.

The two-dimensional mode has likewise performed well in laboratory tests, and is now being employed for solar spectropolarimetry observations at IRSOL, such as the synoptic survey observations by Lucia Kleint (publication in preparation).

Chapter 5

Performance Simulation and Analysis for SPHERE/ZIMPOL

5.1   Motivation

The detection of a polarization signal as faint as 10^{-5} in night-time astronomy is no simple task; its requirements in design, engineering, manufacturing and operation push the envelope of the currently feasible. For this reason, it is imperative to know in advance what performance can be expected from the instrument, which factors will have the most critical impact on the detectivity, which methods bring along the best leverage for improvements.

In Phase A of the planetfinder project, a feasibility study we made within the framework of the CHEOPS proposal concluded that the instrumental, astrophysical and conceptual givens of the undertaking were promising enough to pursue it further. As Phase A turned into B and SPHERE emerged as the fusion of its two predecessor proposals, a more concrete and comprehensive study became necessary.

Having been alerted to the potentially fatal threat of static and differential aberrations (Cavarroc et al., 2006), we decided to harness for ZIMPOL the power of the problem-solving environment CAOS[1] (Carbillet et al., 2004; Fini et al., 2001) created by our French collaborators at LAOG within the framework of the Interactive Data Language (IDL)[2]. We co-developed the SPHERE software package with them, including a code module dedicated to the specific needs of the ZIMPOL instrument. Section 5.2 describes the code in detail.

This software allowed us to run end-to-end simulations of ZIMPOL observations including aberration effects, to gauge the expected performance of the instrument as well as to determine how dependent the results are on a given input parameter. A description of these simulation runs and their results is given in Sections 5.4 and 5.5.

This work has been presented to ESO in the CHEOPS internal document 35 (Thalmann & Schmid, 2004) for Phase A and the SPHERE internal document 191 (Thalmann, 2007) for Phase B.

---

[1] `http://www-luan.unice.fr/caos/`
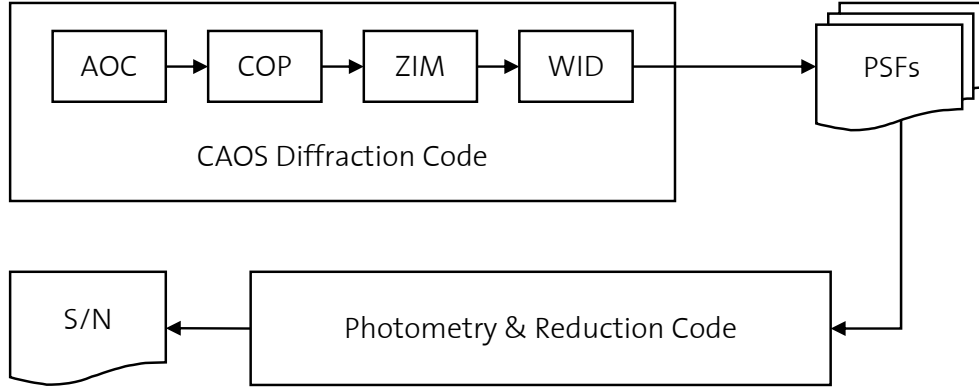[2] Research Systems, Inc., `http://www.rsinc.com`

Fig. 5.1: Architecture of the ZIMPOL/SPHERE performance simulation software.

5.2   The Simulation Code

5.2.1   Architecture

As shown in Figure 5.1, the ZIMPOL/SPHERE performance simulation code is composed of two distinct programs:

- The *diffraction code* produces point-spread functions (PSFs) both for the central star occluded by the coronagraph, and for the off-center planets. It is implemented in the IDL-based problem-solving environment CAOS as part of the SPHERE software package. The code simulates the effects of AO-corrected atmospheric turbulence and various static and differential aberrations. The output is saved to file.

- The *photometry and reduction code* combines a set of PSFs provided by the diffraction code with numerical parameters describing properties of the target star, the planets, the observation procedure, the optical path and the detector, to simulate the final output pictures of an observation run. Differential imaging and optional signal enhancement methods are applied, and the resulting signal-to-noise curves are plotted to a PS file. This code is implemented as an IDL routine named `pfb_zim_red`.

The required CPU time for a full-fledged diffraction code run is typically on the order of one day, although "quick shots" without turbulent atmosphere simulation can be completed in a quarter of an hour. The photometry and reduction code, on the other hand, terminates within a few seconds in its usual configuration. Complicated reduction options such as angular differential imaging can increase this to about a minute.

The following subsections describe the two programs in greater detail. The choice of input parameters is presented and discussed in Section 5.4.

The ZIM module, the photometry and reduction code, and some alterations to the COP module were programmed by the author.
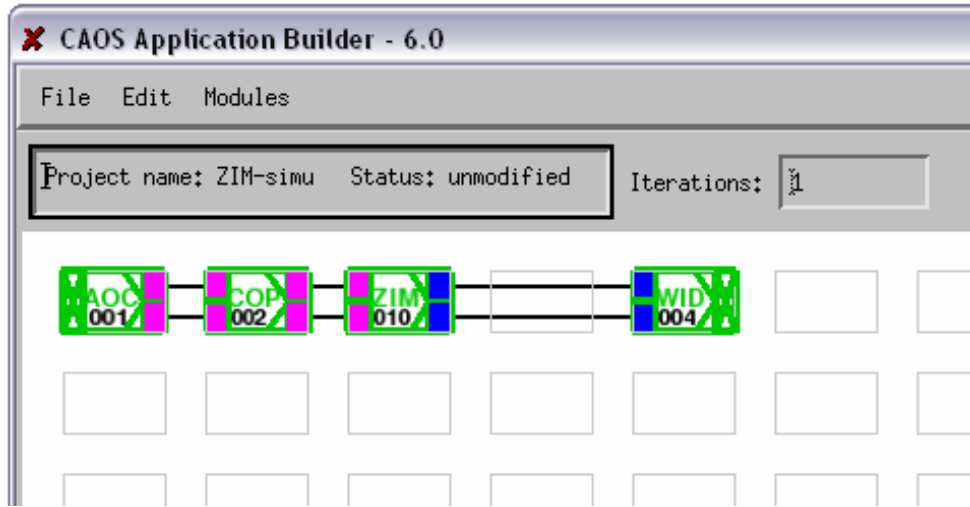
Fig. 5.2: A ZIMPOL simulation project built in the graphical user interface of the CAOS environment. The black connectors signify data flow (from left to right), while the colored boxes encode the data type of the inputs and outputs of each module.

5.2.2 The Diffraction Code

The CAOS environment provides a graphical application builder interface in which pre-defined code modules can be arranged and connected into a sequence to form a program. The SPHERE software package comprises seven such modules, four of which are required to build a ZIMPOL simulation project: AOC (Adaptive Optics Code), COP (COmmon Path), ZIM (ZIMPOL instrument) and WID (Write Image Data). Figure 5.2 shows the finished project on the CAOS worksheet.

Clicking on a module in the project worksheet opens a graphical user interface (GUI) that allows the user to choose its input parameters. Finally, the number of iterations N which the program will execute can be set in the worksheet.

When the project is compiled and launched, each module first runs through its initialization routine, loading the user-defined parameters and preparing data structures. Then, the modules are executed in sequence N times, each passing on its output to the next module in line. In the case of the ZIMPOL simulation, each iteration produces a set of speckled PSFs from a single uncorrelated phase screen representing an instantaneous snapshot of the AO-corrected atmospheric turbulence. The speckle patterns from all iterations are added up incoherently in order to simulate a finite exposure time. The atmospheric speckle variations average down asymptotically with an increasing number of iterations, stabilizing into a smooth halo for $N \gtrsim 100$. The roughness of the PSF is then completely dominated by the instrumental static and quasi-static aberrations, as it is expected for a long observation.

Due to the significant computer time required for a hundred iterations, it is often preferable to run a single iteration only, with the turbulent atmosphere code deactivated. The resulting PSF will lack the smooth seeing halo and therefore be unrealistic; however, the aberration-dominated "bump map" will be identical to the one obtained from a turbulent simulation at the $N \to \infty$ limit. The planet detec-
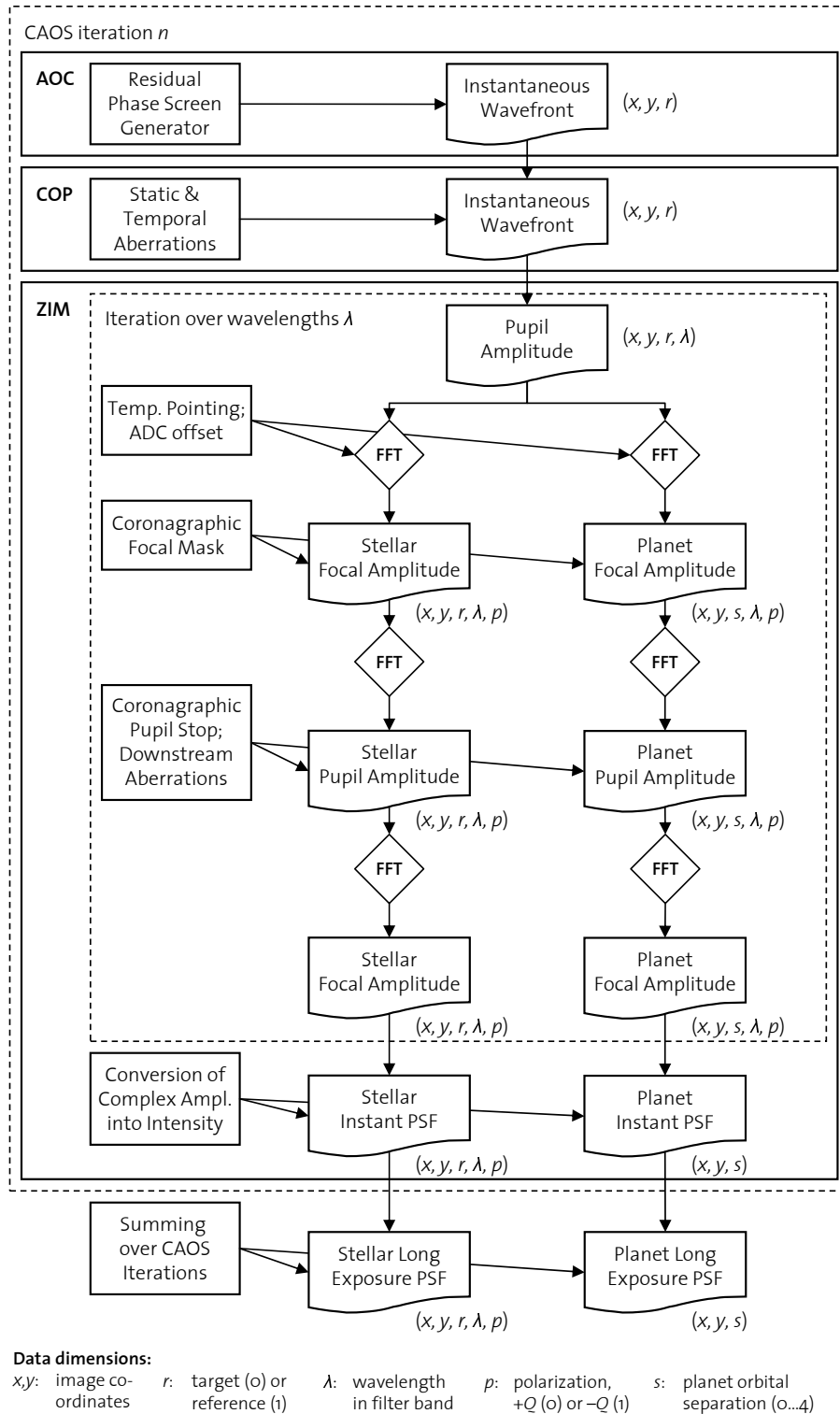
Fig. 5.3: A schematic overview of the data flow in the CAOS modules AOC, COP and ZIM. The final products are a five-dimensional array for the stellar (coronagraphic) PSFs and a three-dimensional array for the planet (off-axis) PSFs.

tivities calculated from such non-turbulent "quick shot" simulations are accurate as long as the noise level remains aberration-dominated. As soon as the statistical photon noise becomes relevant, a turbulent simulation is necessary.

The CAOS code tracks two separate realizations of the coronagraphic star PSF at all times. One represents the target star around which the supposed exoplanet orbits, the other a reference star used for calibration. A number of temporal differential effects between the two observations are accounted for by the simulation, limiting the efficiency of the calibration. In the case of the ZIMPOL instrument, however, no reference star is necessary; instead, this infrastructure is used to represent the two channels of signal switching (see Section 3.5). Although both of these channels observe the planet-bearing target star, they will be referred to as the "target" and "reference" channels for simplicity.

The ZIM module adds further dimensions to the data cube. Each on-axis PSF is stored in the form of two orthogonal polarization components, allowing to introduce polarized differential aberrations between them. Also, several monochromatic versions of each PSF are calculated and saved in order to compose a pseudo-polychromatic broadband image later in the photometry code.

Finally, apart from those on-axis PSFs for the coronagraphic star image, the ZIM module also produces off-axis PSFs for five planets at angular separations ranging from $0.''05$ to $0.''5$. Since only the core of these PSFs is going to be observable, the images for each wavelength and polarization component are not stored separately, but added up into a single PSF for each planet.

The following paragraphs describe the function of each module. Figure 5.3 provides a schematic overview of the data flow in the diffraction code. A detailed account of the input parameters will be given in Section 5.4.

The AOC Module

This module generates random wavefront aberrations that mimic the residuals from atmospheric seeing and instrumental jitter after AO correction has been applied. The wavefront errors are composed of several terms described by power spectrum densities (PSD). The GUI provides access to parameters defining the properties of the turbulent atmosphere, the target star, the telescope dimensions and the AO system.

The output is a set of wavefront maps in nanometers.

The COP Module

This module adds a number of wavefront errors to the wavefronts passed down from the AOC module, accounting for all static aberrations upstream of the coronagraph, including the three telescope mirrors, as well some temporal differential aberrations. The GUI provides access to the parameters describing those aberrations.

The output is a set of wavefront maps in nanometers.

The ZIM Module

This module handles the whole optical path downstream of the coronagraph, ending with the ZIMPOL cameras.

First, the wavefronts must be converted from spatial "bump maps" into complex amplitude maps, i.e. from optical path difference to phase. Since this calculation is wavelength-dependent, it is done for each considered wavelength separately.

Next, the complex amplitude must be converted from the pupil to the focus plane, where the coronagraphic mask can be applied. For this purpose, each monochromatic complex amplitude map is placed in the center of a larger empty array, the size of which is chosen proportional to the wavelength (1024 pixels squared for the shortest considered wavelength). The array is then Fourier-transformed and re-centered, yielding the monochromatic PSF. A wavelength-specific offset from the true center is applied to represent the residual chromatic effects from the atmosphere not fully compensated by the atmospheric dispersion corrector (ADC). The PSF of the reference channel is shifted by an additional small offset to represent the imperfect alignment stability of the telescope and coronagraph (this is a temporal differential aberration). At this point, the complex transmission map of the coronagraphic mask is multiplied onto the PSF. A variety of coronagraph designs, including classical Lyot and four-quadrant phase mask coronagraphs (introduced in Section 2.5), can be selected.

In the next step, the PSFs are transformed back into the pupil plane, where the transmission map of the pupil stop is multiplied onto the complex amplitude. All static and differential aberrations downstream of the coronagraph are added here.

Finally, the complex amplitudes are transformed into the focus plane once more, and squared to yield intensity PSFs. They are cropped down to the size of the smallest array (i.e. 1024 pixels squared), or a fraction thereof if the user is interested only in the central region. In the case of exoplanet search, a window of 512 pixels squared is sufficient, since the expected signal-to-noise ratios are unusably low in the outer regions of the field.

The output is a five-dimensional array $[x, y, w, r, p]$ storing the on-axis coronagraphic star PSFs, and a three-dimensional array $[x, y, s]$ storing the off-axis planet PSFs, where x, y are the image coordinates, w identifies the wavelength, r discriminates between the target and reference channels, p distinguishes the two orthogonal polarization components, and s chooses among the five different star/planet angular separations.

The WID Module

This module writes the data arrays to file for further use by the photometry and reduction code. The output is a `*.sav` file readable by IDL.

5.2.3 The Photometry and Reduction Code

Once the CAOS code has produced its PSFs, the routine `pfb_zim_red` is responsible for the entire rest of the simulation. It is written in IDL and does not require a

CAOS installation to run. A host of keywords can be set in order to modify the simulation parameters.

Photometry

Running `pfb_zim_red` first invokes the auxiliary routine `pfb_zim_flux`, which calculates the total number of photons from the target star that pass through the telescope aperture, survive the passage through the various components, and end up being registered by a detector, not taking into account the rejection of photons by the coronagraph (this rejection is already encoded in the relative scaling of the on-axis and off-axis PSFs provided by the diffraction code). The stellar spectrum is multiplied with the transmission spectra of all components of the optical path, starting with the reflectivity of the primary mirror and ending with the quantum efficiency of the CCD chips. Furthermore, the brightness contrasts C between the star and the planets is derived from the planet radius r, phase albedo $a(\phi)$ for the phase angle ϕ, and orbital radius $R(\theta)$ of the planets:

$$C(\theta, \phi) = a(\phi) \cdot \frac{r^2}{R^2(\theta)}, \tag{5.1}$$

whereby the angular separation θ, the phase angle ϕ and the distance d from the Earth to the target star are used to obtain the orbital radius $R(\theta) = d \cdot \theta / \sin \phi$.

Next, `pfb_zim_red` loads a set of PSFs from a designated `*.sav` file and re-bins them according to Nyquist sampling for the ZIMPOL spectral range (2 pixels per resolution unit λ/D at $600\,\mathrm{nm}$), scaling them down by a factor of ~ 3 in the process. In the actual hardware, each of these Nyquist pixels is realized by 8×8 CCD pixels, half of which are masked to serve as buffer memory (see Section 3.4), on each of the two CCD cameras. The code does not concern itself with these engineering details, directly simulating the reduced, combined and re-binned final images rather than the raw data.

The individual monochromatic star PSFs are weighted according to the flux in their spectral range, and added up into a pseudo-polychromatic PSF. The planet PSFs, which are already provided in a pseudo-polychromatic form by the diffraction code, are simply scaled with the total stellar photon count times the star/planet contrasts.

Subsequently, the dimension of polarization is added to the data. For the planets, two different atmosphere chemistries are considered, one with a pessimistic 10% and one with a very optimistic 50% of scattering polarization. For each case, two instances of each planet PSF are saved to represent two orthogonal polarization components. The polarization of the residual star PSFs has already been accounted for in the diffraction code.

Reduction

At this stage, Poissonian photon noise, read-out noise, sky glow and systematic errors (due to the finite polarimetric accuracy) are applied to the data. The polarization components are combined to form the Stokes I (intensity) and Q (linear

polarization) images. Image processing, such as the simulation of field rotation and spatial filtering, is optionally performed. This is described in detail in Section 5.3. Finally, Q images of the star PSF from the target and reference channels are subtracted from one another, yielding the final double-difference background image.

While the images themselves can be obtained as output, the code provides a built-in evaluation tool that extracts the relevant information into easy-to-read detectivity curves. Since the coronagraphic star PSF is roughly circularly symmetric in general structure, it can be characterized by a plot of the 5σ-detectivity as a function of angular separation from the star. The detectivity at a given separation is calculated as five times the standard deviation within an annulus of that angular radius in the background image. Plotting the planet signals, which are simply defined as the values of the brightest pixel of the respective planet PSF, onto the same logarithmic graph reveals whether or not a given planet has been detected in the simulated observation: If its signal lies exactly on the detectivity curve, it can be distinguished from the background noise with a confidence of 5σ. The plots in Sections 5.4.5 and 5.5 illustrate the use of this output format.

5.3   Signal Enhancement Techniques

This section describes software image processing methods that can be applied to the measured images to enhance the detectivity of the desired planet signals. The quantitative effect of these methods are discussed in Section 5.6.

5.3.1   Angular Differential Imaging

Even after calibrating a polarization image by signal switching, a significant amount of spurious structures caused by aberrations are expected to remain in the image, dominating above the statistical photon shot noise. While the latter is simply Poissonian white noise at the pixel scale, the the former takes the shape of a larger-scale landscape featuring bumps and ridges of the size of a resolution element, competing with a planet signal and hindering its detection.

Angular differential imaging offers a powerful tool to suppress this background. The concept is rather intuitive in nature: During the observation, the target star moves across the sky and requires the telescope to continuously adjust its bearing to keep the star centered in the field of view. For most telescope designs, including the altitude-azimuth mounting of the VLT units, this motion causes the coordinate system of the sky field to rotate with respect to the fixed frame of reference of the instrument optics and the detector. This relative motion provides a means to discriminate between real signals that originate on the sky and spurious signals introduced by the optics.

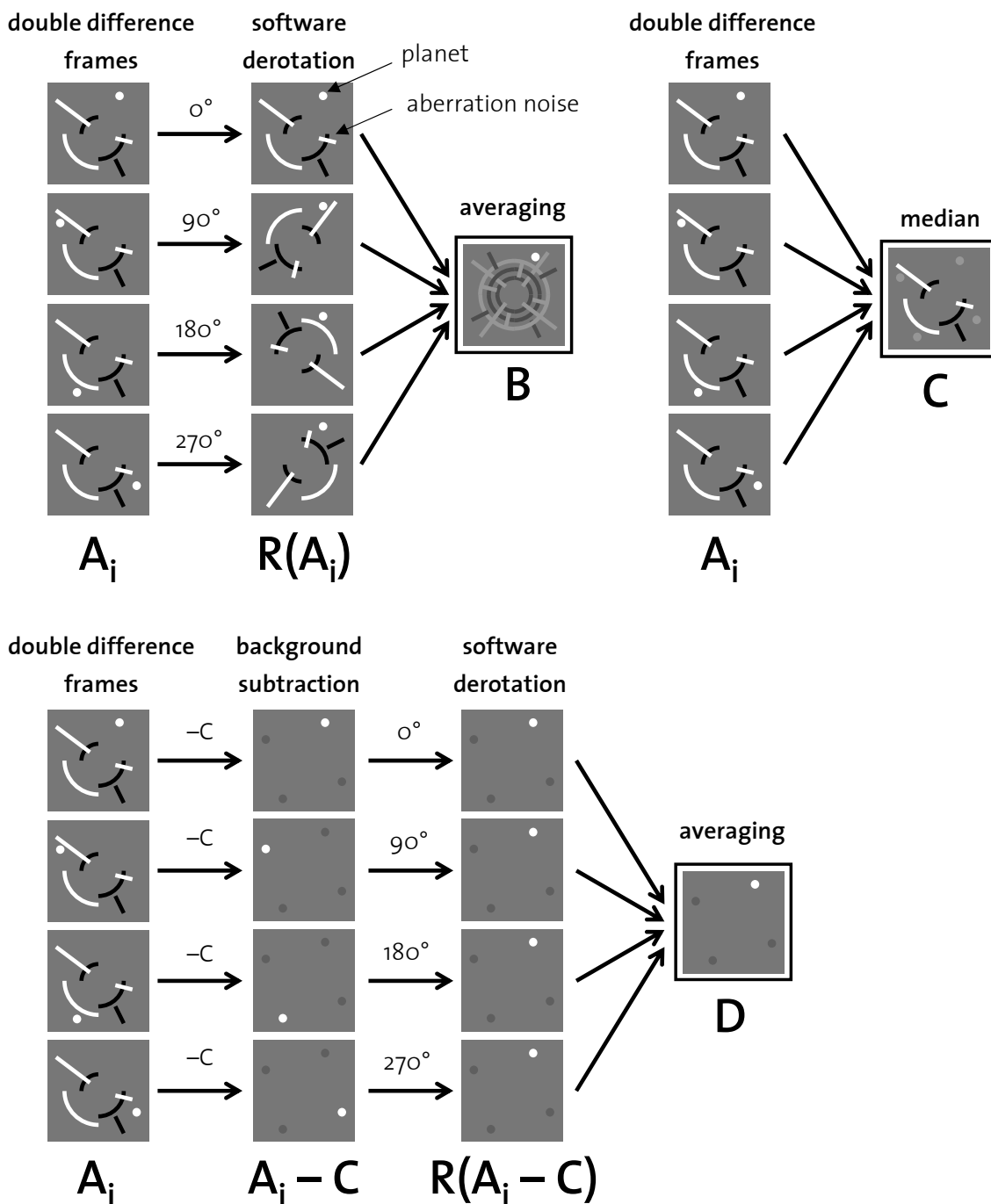Figure 5.4 illustrates the image processing methods described below.

Fig. 5.4: Schematic illustration of angular differential methods on the example of active field rotation with four positions. For *angular averaging*, the individual exposures (A_i) are simply derotated into the coordinate system of the sky ($R(A_i)$) and added up (B), whereby the background structures are averaged down. For proper angular differential imaging with *background subtraction*, a median background of the raw exposures is first calculated (C) and subtracted from each exposure ($A_i - C$) before derotation ($R(A_i - C)$) and final co-addition (D).

Angular Averaging

Simply adding up all raw exposures taken during an observation will result in a diluted arc-shaped planet signal, smeared out from its momentary dot shape by field rotation. Since the evolution of the field orientation is well known, one can rotate each exposure into the rest frame of the sky field using image processing software. The planet signals will then coincide in the same location on all exposures while the background landscape is spread out along an arc. Summing the exposures up restores the planet signals to full strength and averages the background down.

The resulting gain at a given angular separation is proportional to the square root of the number of individual planet-sized (λ/D) background samples that are averaged to produce a given planet-sized spot on the co-added image. Thus, a large total field rotation angle is desirable. Furthermore, the method is more effective at large separations from the center, where a given rotation angle corresponds to a greater arc length than in the central area. The maximum number of independent samples at an angular distance of $n\,\lambda/D$ with a field rotation angle of ϕ (in radian) is $N = \phi \cdot n$, thus the maximum gain achievable by averaging is $\sqrt{\phi \cdot n}$.

Background Subtraction

The optical components responsible for the bulk of the aberrations are fixed to the telescope and the optical bench, maintaining a constant position and orientation in the beam. While there are moving parts and slowly drifting contributions (quasi-static aberrations), one can nevertheless assume that the background landscape retains most of its likeness over the course of the observation.

This can be exploited to identify and remove this static background rather than just averaging it down. The static background, calculated as the pixel-wise median of all exposures, is calculated and then subtracted from each exposure. This process can be repeated in the rest frames of other coordinate systems, such as the derotator hardware, as long as it behaves different from the sky frame. When the images are finally co-added in the rest frame of the sky, a planet signal at almost full strength and an arc of very faint negative "ghost planets" remain (see Figure 5.4).

In principle, this technique has the potential to reduce the background down to the photon noise limit. In practice, the shape of the stellar PSF depends on the different aberration contributions in non-linear ways; therefore the method will not work perfectly when several strong noise sources with different rotational rest frames or strong quasi-static aberrations are present.

Marois et al. (2006) provide a contemporary evaluation of this method, which they name *angular differential imaging*.

Natural vs. Active Field Rotation

The arc provided by natural field rotation during a typical 4-hour observation is strongly dependent on the position in the sky. For instance, a star near the celestial pole will experience no appreciable field rotation. For this reason, it will be desirable

to repeatedly change the field orientation in discrete steps over the course of an observation, using the hardware field derotator and half-wave plates.

Such "active" field rotation provides access to the full 2π arc for angular differential imaging, which is particularly beneficial for in the detector area close to the star, where the arc length for a given rotation angle is very small. On the other hand, the use of the derotator hardware introduces its own share of temporally variable aberrations. Test observations will be necessary to determine the most advantageous method. A combination of the two, i.e. periods of natural field rotation interrupted by artificial "jumps", might be desirable.

5.3.2   Spatial Filtering

Given the knowledge of the shape and width of a planet signal, the image can be treated such as to suppress all signals on size scales larger or smaller than that, enhancing the detectability of planets.

Aperture Convolution

The simplest way to obtain a signal-to-noise ratio is to compare the brightest pixel of the supposed planet signal with the pixel-to-pixel noise in a ring-shaped area centered on the star. However, in our Nyquist-sampled case, the actual shape of the planet signal is a point spread function with a half-power beam width of two pixels at $600\,\mathrm{nm}$ (and three pixels at $900\,\mathrm{nm}$), thus the pixel-wise signal extraction wastes a lot of photons.

A potentially better way is to convolve the image with a sampling aperture of the size of a PSF core before applying the brightest-pixel method. Each pixel value in the convolution image represents the integral over an aperture-shaped sample of the original image. This way, the pixels that form the core of the planet's PSF add up to form a strong signal, whereas the smaller-scale (pixel to pixel) noise of the background is dampened by the integration. Any spurious background structures on the spatial scale of a planet signal (such as speckles) or on larger scales (such as gradients) survive the process unscathed and are therefore not suppressed.

Under our standard simulation conditions (see Section 5.4), the double-difference image turns out to be dominated by a landscape of aberration residuals, so the S/N gain from aperture convolution is negligible. However, if that landscape can be suppressed below the level of the photon noise limit by angular differential imaging, the dominant noise factor will be pixel-sized, and a S/N gain of more than a factor of 2 can be achieved by aperture convolution. The top row of Figure 5.5 illustrates this enhancement.

Unsharp Masking

While aperture convolution filters out the pixel-to-pixel noise, unsharp masking does the same thing for the spatial scales well above the size of a planet signal.

A map of the large-scale background landscape is obtained by convolving the detection image with a coarse smoothing function, such as a 7×7 pixel square. This
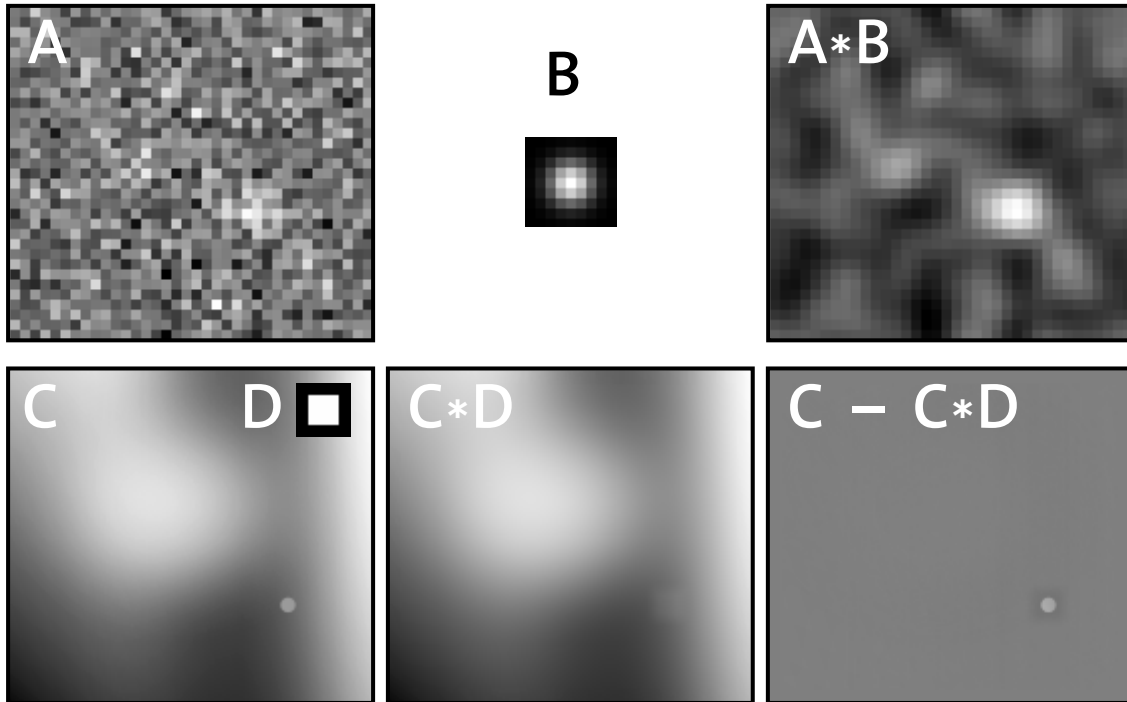
Fig. 5.5: Illustration of spatial filtering techniques. **Top row:** Aperture convolution. Image A is dominated by statistical noise on the pixel-to-pixel scale, representing photon noise. A gaussian signal with a width of $3\,\mathrm{px}$ and a peak height of 3σ is buried in the lower right quadrant. Convolving A with a sampling aperture B, which shares the shape and size of the signal, results in the image $A * B$, averaging down the pixel-scale noise and enhancing the signal. **Bottom row:** Unsharp masking. Image C is dominated by a large-scale landscape and features a small-scale signal in the lower right quadrant. Convolution with an aperture D, which is significantly larger than the signal, results in the background estimation $C * D$, in which the signal is smeared out. The subtraction $C - C * D$ removes the background landscape while preserving the small-scale signal.

landscape is then subtracted from the image, leaving only the small-scale structures behind. This process is known as unsharp masking and corresponds to the application of a high-pass filter. The planet signal is small enough (3×3 pixels) to be flattened out by the smoothing; it will not be noticeable in the calculated background map and therefore survive the subtraction unscathed. The bottom row of Figure 5.5 illustrates the process.

This technique is particularly effective against very large, smooth structures in the background, such as the result of angular averaging with natural field rotation.

5.4 The Standard Case SC3

During the evolution of the SPHERE project, the accepted set of standard input parameters for the performance simulation has been updated regularly. The current version, Standard Case 3 (SC3) is defined and discussed in this section.

The entries are sorted by the routines in which they are accepted as input. Each numerical parameter is given in the unit expected by the routine. Wavefront aberrations are measured in nm RMS. Values deviating from the default settings of the routines are marked with a bullet ($\bullet$).

5.4.1 AOC Parameters

The following input parameters are given to the AOC module. They are based on the SPHERE internal document 248 (Fusco, 2007).

Atmospheric parameters		*Telescope parameters*	
Seeing at 500 nm	0″85	Diameter [m]:	8
Turbulent Layers:	3	Obscuration ratio:	0.14
– altitude z [m]:	0; 1000; 10000	Zero point [e$^-$/m^2 s]	$2.1 \cdot 10^{10}$
– C_n^2 profile [%]:	20; 60; 20		
– wind speed v [m/s]:	12.5; 12.5; 12.5	*AO system parameters*	
– direction [°]:	0; 45; 90	Imaging wavelength [m]	$\bullet$ $7.5 \cdot 10^{-7}$
		Linear # sub-apertures:	40
AO guide star parameters		Integration time [s]:	0.000833
Magnitude:	$\bullet$ 4	Loop delay [s]:	0.001
Zenith angle [°]:	30	Loop gain:	0.465999
Azimuth angle [°]:	0	WFS wavelength [m]:	$6.5 \cdot 10^{-7}$
		Read-out noise [e$^-$]:	0.5
		Dark current [e$^-$/s]:	2
		Spatial filter efficiency:	0.526298

The atmosphere parameters describe median conditions at the Paranal site. The AO performance is insensitive to the guide star brightness down to ~ 8 mag, where it starts to degrade due to photon noise in the WFS. For this reason, an arbitrary bright magnitude of 4 is chosen. The imaging wavelength is by default set to the IRDIS band; this must be adjusted to the ZIMPOL band (750 nm central) for SC3.

5.4.2 COP Parameters

The following input parameters are given to the COP module. They are based on the SPHERE internal document 161 (Dohlen, 2007).

Static Aberrations		*Temporal Aberrations*	
Telescope Mirror M1 [nm]:	11.9	Defocus [nm]:	• 0.1
Telescope Mirror M2 [nm]:	11.9	Pupil shear [D unit]:	0.002
Telescope Mirror M3 [nm]:	16.6	Pupil rotation [°]:	• 0.3
Instrument [nm]:	34.5	Differential rotation [°]:	• 0
AO calibration [nm]:	7.4	Diff. beam shift [nm]:	• 0
Fresnel propagation [nm]:	• 0	Diff. high freq. [nm]:	• 11
Beam shift [nm]:	• 0		

The telescope mirror aberrations are provided as pre-made FITS files containing a two-dimensional wavefront map; the RMS values given in the table above are not directly accessible. The Fresnel propagation and beam shift effects can be neglected in the ZIMPOL case, because the science observation uses the same spectral range as the AO's WFS (this is not the case for IRDIS observations).

The temporal aberrations describe the changes and drifts in the common path on the time scale of the signal switching, which therefore cause differential effects between the target and reference channel, and limit the effectiveness of this calibration method. The half-wave plate responsible for signal switching (HWP2) is foreseen to toggle its state every five minutes in order to minimize these effects. For this reason, most parameters can be set to small values or zero. The imperfect surface of the half-wave plate itself, however, introduces its own temporal aberration whenever it rotates. While the low-frequency components of these aberrations are corrected by the AO, a high-frequency term of 11 nm RMS remains.

The values assume observation mode P1 (see Section 3.3), i.e. no field derotation. A constantly moving derotator would increase some of the temporal differences, such as pupil rotation.

5.4.3 ZIM Parameters

The following input parameters are given to the ZIM module. They are based on the SPHERE internal documents 179 (Schmid, 2007), 180 (Roelfsema & Schmid, 2007), 185 (Roelfsema & Gisler, 2007) and 229 (Boccaletti et al., 2007).

Coronagraph parameters		*ZIMPOL Aberrations*	
Mask radius [λ/D at 600 nm]:	5	ADC residuals [mas RMS]:	3.97
Pupil stop transmission [%]:	80	Temp. diff. pointing [mas]:	0.1
Filter bandwidth [nm]:	300	Phase diff. aberr. [nm]:	1
Filter central wavelength [nm]:	750	Pointing diff. aberr. [nm]:	1.5
# wavelengths simulated:	6	Downstream static aberr. [nm]:	50
		Beam shift [nm]:	1.1

The default coronagraph type foreseen for the visual branch of SPHERE is a

classical Lyot design, consisting of an opaque disk in the focal plane to absorb the stellar PSF core, and an opaque stop in the pupil plane to cover the edges of the projection of the primary and secondary mirrors, which is where the stray light is most concentrated in the pupil image. The stop is chosen to be rotationally symmetric, although the symmetry of the actual pupil image is broken by the shadow of the support spider of the secondary mirror. Since the pupil image will rotate with respect to the pupil stop during observations and it is technically impractical to keep the stop actively aligned, the diffraction pattern of the spider remains in the image.

Other classical Lyot coronagraphs selectable from the GUI include focal mask radii of 3 or $4\,\lambda/D$ at $\lambda = 600\,\mathrm{nm}$ and corresponding pupil stops with transmissions of 65% and 74%, respectively. The focal mask radius can be set to zero to simulate noncoronagraphic observations. Apodized Lyot or four-quadrant phase mask (4QPM) coronagraphs are also implemented. An "ideal coronagraph" that simply subtracts a stellar Airy pattern from the data is provided as a point of reference against which to judge the performance of the more realistic designs.

The polychromatic PSF of an actual ZIMPOL observation is only roughly approximated by the 6 monochromatic PSFs summed up by the simulation under standard settings. Such approximations are acceptable for comparative studies as in Section 5.5. If the absolute performance is important, a more realistic radial smearing of speckles can be achieved with 30 simulation wavelengths. Apart from the ZIMPOL broadband spectral range, medium- and narrowband filters such as for CH_4 or $H\alpha$ are available.

The light of a star does not generally reach an earthbound telescope in a straight line; the passage through the Earth's atmosphere and its density gradient refracts the direction of the ray towards the vertical. Due to dispersion, the resulting deflection angle depends on wavelength, causing a star image to be smeared out into a spectrum on the detector. In SPHERE, an atmospheric dispersion corrector (ADC) mostly counteracts this effect by exploiting the dispersion characteristics of certain optical materials. The residual offsets of each monochromatic PSF with respect to the center are provided to the ZIM module as an ASCII table. Figure 5.6 shows a plot of these values. In the ZIMPOL spectral range, the RMS is 3.97 mas.

The pointing differential aberrations conceptually belong to the temporal effects treated in the COP section. They appear in the ZIM interface instead due to their intimate connection to the coronagraph. The term describes the imperfect alignment stability of the star PSF on the center of the coronagraphic mask on the time scale of signal switching, i.e. 5 min.

The FLC polarization modulator of ZIMPOL is the only component that changes its state on the time scale of (and in synchronization with) the demodulation on the CCD detectors. As such, it is the prime source of polarized differential aberrations. Any change in optical behavior between the two modulator states apart from the retardation will cause spurious effects to appear in the polarization image.

Laboratory tests conducted at ASTRON and ETH Zürich (Roelfsema & Gisler, 2007) have shown that the beams emerging from the FLC in its two states are not exactly parallel. This differential pointing corresponds to a gradient-shaped aberration of 1.5 nm RMS. Furthermore, an interferometer test revealed "phase-like"
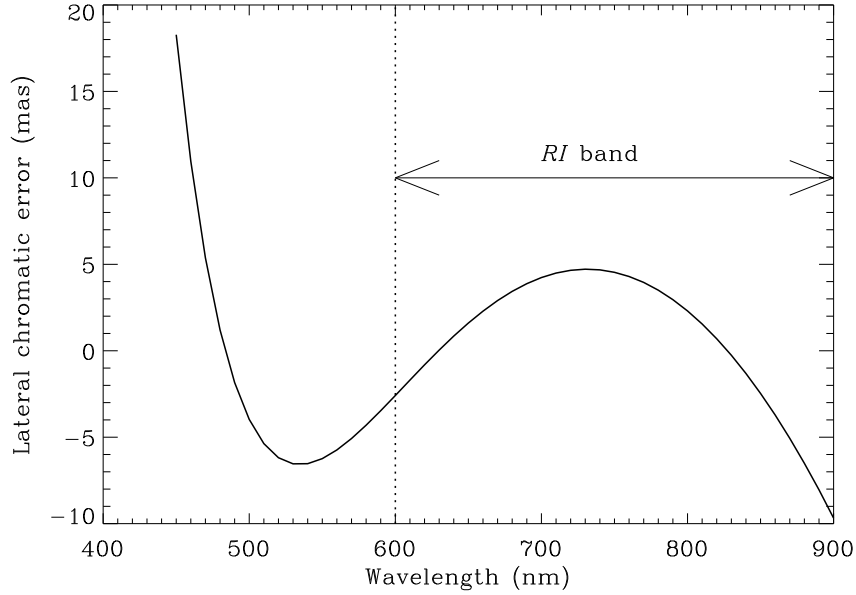
Fig. 5.6: Plot of the residual chromatic offsets after ADC correction as a function of wavelength. The standard *RI* band for ZIMPOL observations is marked.

(higher-order) differential aberrations of 1 nm RMS. Finally, the differential pointing causes the outgoing beams to pass through slightly different crosssections of the subsequent optical surfaces, thus the static aberrations induced by the imperfections of those surfaces are not identical between the two polarizations. This so-called beam shift effect adds another 1.1 nm of phase-like differential aberrations.

5.4.4 Photometry and Reduction Parameters

The following input parameters are given to the photometry and reduction code. They are based on the SPHERE internal documents 180 (Roelfsema & Schmid, 2007) and 234 (Gratton & Mouillet, 2007).

Photometric parameters		*Planet parameters*	
Star type:	G0V	Angular separations ['']:	0.05–0.5
Star distance [pc]:	3	Size [Jupiter radii]:	1
Exposure time [h]:	4	Phase angle [°]:	80
Total transmission:	0.067	Phase-dependent albedo:	0.13
Full-well cap. [photons/px]:	10^5	Degree of polarization [%]:	10; 50
Read-out noise [photons/px]:	10		

Here, [px] refers to a CCD hardware pixel. A pixel in the final reduced image, on the other hand, is implemented with 32 such hardware pixels (8 exposed and 8 masked pixels on each of 2 cameras).

The default five planet separations are 0″.05, 0″.1, 0″.2, 0″.3 and 0″.5. The innermost planet lies within the threshold of the inner working angle (IWA) of the

$5\,\lambda/D$ Lyot coronagraph, thus most of its PSF is absorbed by the focal mask. The outermost planet, on the other hand, is located beyond the control radius of the adaptive optics and must therefore contend with a significantly higher noise level.

The phase-dependent albedo $a(\phi)$ is the phase curve of the planet, scaled such that the face-on value is equal to the geometric albedo, $a(0°) = a_{\mathrm{g}}$. It is used to calculate the planet–star brightness contrast as detailed in Equation 5.1. The phase angle is chosen at an advantageous $\phi = 80°$. An angle around $\phi \approx \pm 90°$ is needed for scattering polarization; for $\phi \approx 0°, 180°$, the configuration is too symmetric for one polarization orientation to prevail over the others.

The scattering polarization is perpendicular to the star–planet axis. The code assumes that the polarimeter's Stokes Q direction is aligned with this orientation, i.e. that the plane of orbit is known a priori. If this is not the case, both a Q and a U observation have to be made in order to measure the full linear polarization vector, doubling the necessary observation time.

The maximum degree of scattering polarization is highly dependent on the chemistry, temperature, and weather of the planet. Polarizations of up to $\sim 60\%$ are expected for pure Rayleigh-scattering gas atmospheres, whereas haze and especially clouds can strongly suppress the effect through diffuse scattering (Stam et al., 2004; Joos, 2007). In this work, we choose a pessimistic ($p = 10\%$) and a very optimistic case ($p = 50\%$) to delimit the extremes, expecting a real exoplanet to lie somewhere in between.

5.4.5 Standard Case Performance

The results of the performance simulation under SC3 conditions are presented in Figure 5.7. The curves represent the 5σ detectivity threshold for various stages of image processing. All images were convoluted with a 3×3 pixel aperture as described in Section 5.3.2. For this plot, 30 different wavelengths were simulated rather than the usual 6 in order to provide a particularly realistic PSF texture. All signals are normalized by the signal that a non-coronagraphic image of the target star would yield. The locations of the planet intensity signals on the vertical axis therefore directly correspond to the brightness contrasts between the star and the planets, ranging from $3 \cdot 10^{-7}$ down to 10^{-8} in the investigated interval of angular separations.

It is immediately obvious that one cannot hope to detect an exoplanet with mere intensity imaging (curve A in Figure 5.7). The detectivity level in an intensity image lies three orders of magnitude above the planet signals, dominated by the static aberrations. The subtraction of a reference intensity image (B) can only lower the noise by up to a factor of ten, due to the temporal aberrations between target and reference images.

Going from intensity to polarization (C), on the other hand, cuts the noise level down by two orders of magnitude, whereas the planet signals are diminished only by their degree of polarization. This gain in detectivity is primarily limited by the polarized aberrations caused by the FLC modulator, which are significantly smaller than the static and temporal aberrations.
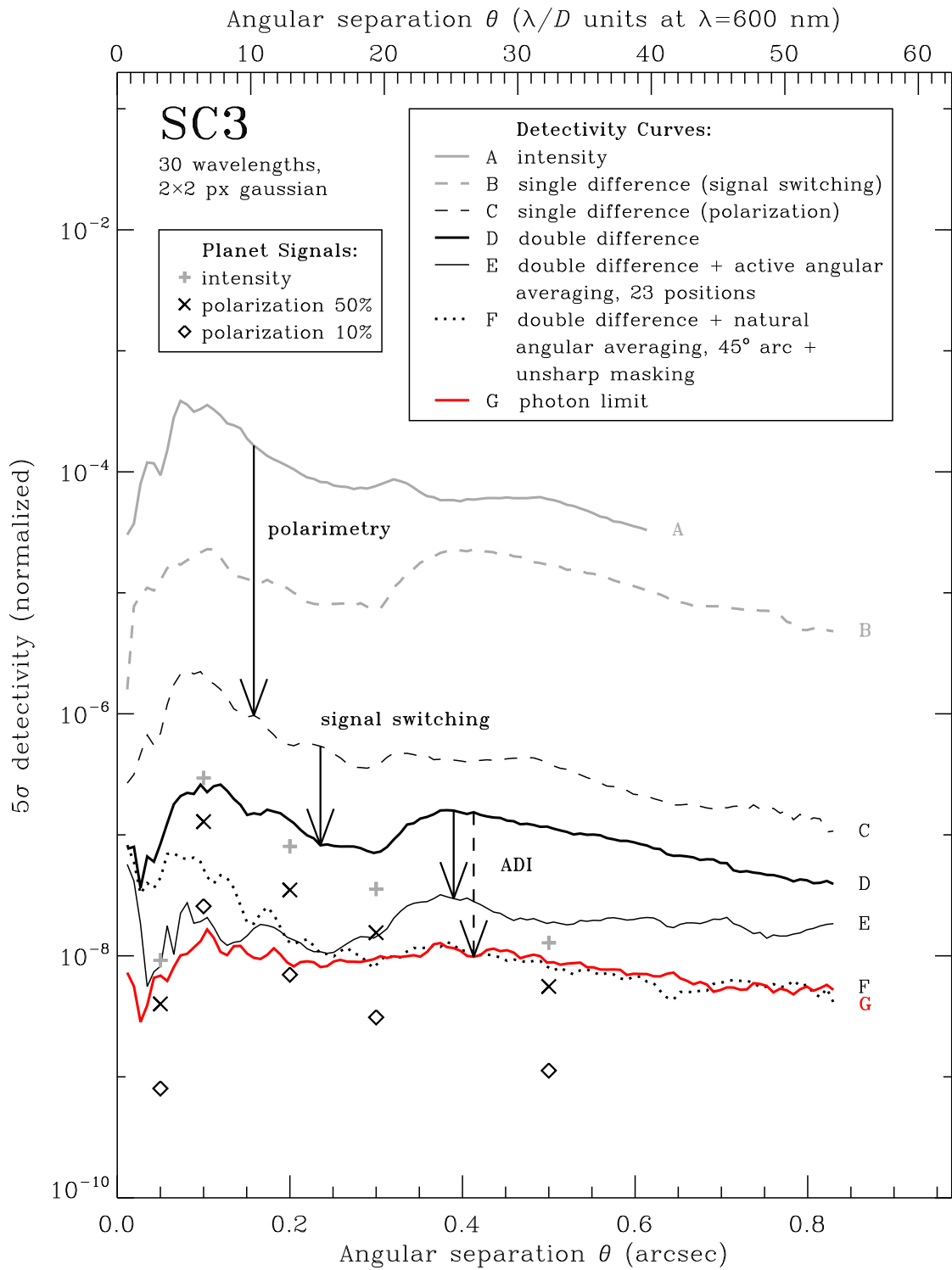
Fig. 5.7: Performance plot for the Standard Case 3 simulation. The planet at $\theta = 0\rlap{.}{''}05$ is obscured by the coronagraphic mask.

The two aforementioned differential methods can be combined by subtracting two polarization images obtained through signal switching from each other, yielding a *double-difference* image (D). The benefits of the two methods are additive, resulting in a noise level roughly three magnitudes below that of a simple intensity image. Under SC3 conditions, this noise level is still completely determined by aberrations rather than statistical photon noise. None of the sample planets are detectable at this level, though the strongest planet signal (50% polarized at $\theta = 0\rlap{.}''1$) constitutes a 3σ event.

A third level of differential imaging is necessary to achieve a detection. Angular differential imaging as described in Section 5.3.1 is well suited to this role. For this evaluation, we consider two different forms of angular averaging:

Active field rotation with 23 discrete orientations of the field distributed evenly over the full circle (E). Assuming that the half-wave plate HWP2 toggles its state every 5 minutes and that the field orientation is changed after each full signal-switching period (10 minutes), there is room for 24 different orientations in a 4-hour observation. The number 23 was chosen for its prime nature such as not to reinforce any symmetries in the background.

Natural field rotation over an arc of 45° (F). The rate of rotation was assumed to be uniform over the observation time for simplicity's sake. Furthermore, unsharp masking with a 7×7 pixel aperture was applied to the resulting image to remove the large-scale landscape.

It is evident that the active method performs better in the innermost part of the field ($\theta < 0\rlap{.}''2$), where the natural method suffers from a very restricted arc length. The natural method appears superior beyond this threshold, where radial residual structures appear in the active angular averaging image. However, the performances of both methods can likely be significantly improved by background subtraction. The exact gain from this final step depends on how stable the double-difference image proves to be, i.e. on the evolution time scale of the quasi-static speckles and their weight with respect to the static part. For a highly static background, the noise curve can reach the statistical photon noise limit (G), as demonstrated in Section 5.3.1.

Note that a prolongation of the observation time by the factor N will lower the photon limit by a factor of $1/\sqrt{N}$. However, this gain can only be harnessed if all other noise sources are successfully suppressed below the photon noise level, which presents an increasingly challenging task.

For visual reference, the two-dimensional detection images at each stage are shown in Figure 5.8. The aperture convolution has not been applied here in order to illustrate the exact nature of the background noise. Human eyesight can reproduce the signal-enhancing effect of aperture convolution on its own.
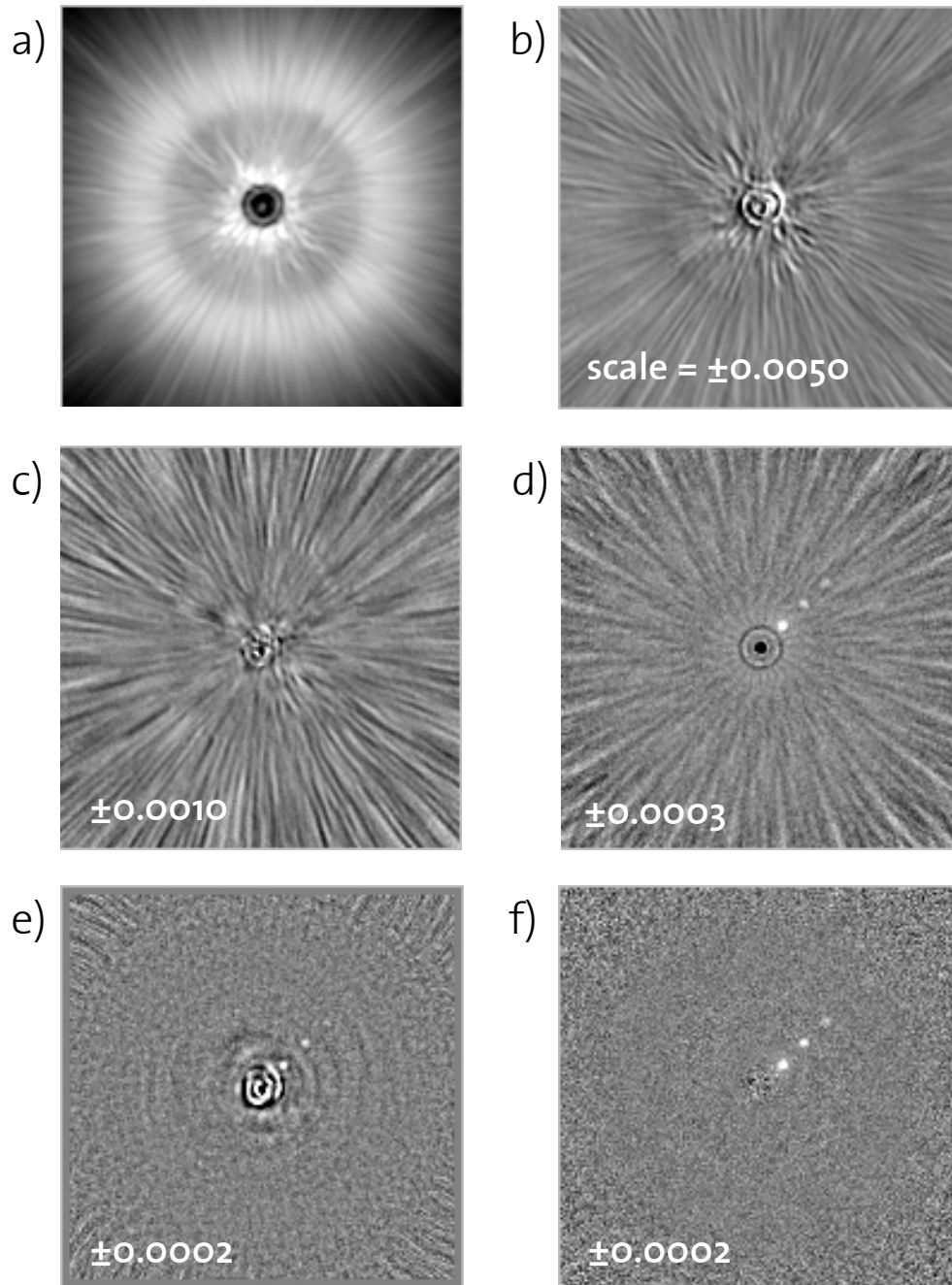
Fig. 5.8: Simulated images for SC3. **a:** The intensity PSF. **b:** The polarization image (single difference). **c:** The reference-subtracted polarization image (double difference). **d:** The double-difference image after angular averaging over 23 positions of active field rotation. **e:** The double-difference image after angular averaging over 45° of natural field rotation and unsharp masking. **f:** The photon limit. The numbers indicate the greyscale range of the Q/I images. The planets are 50% polarized.
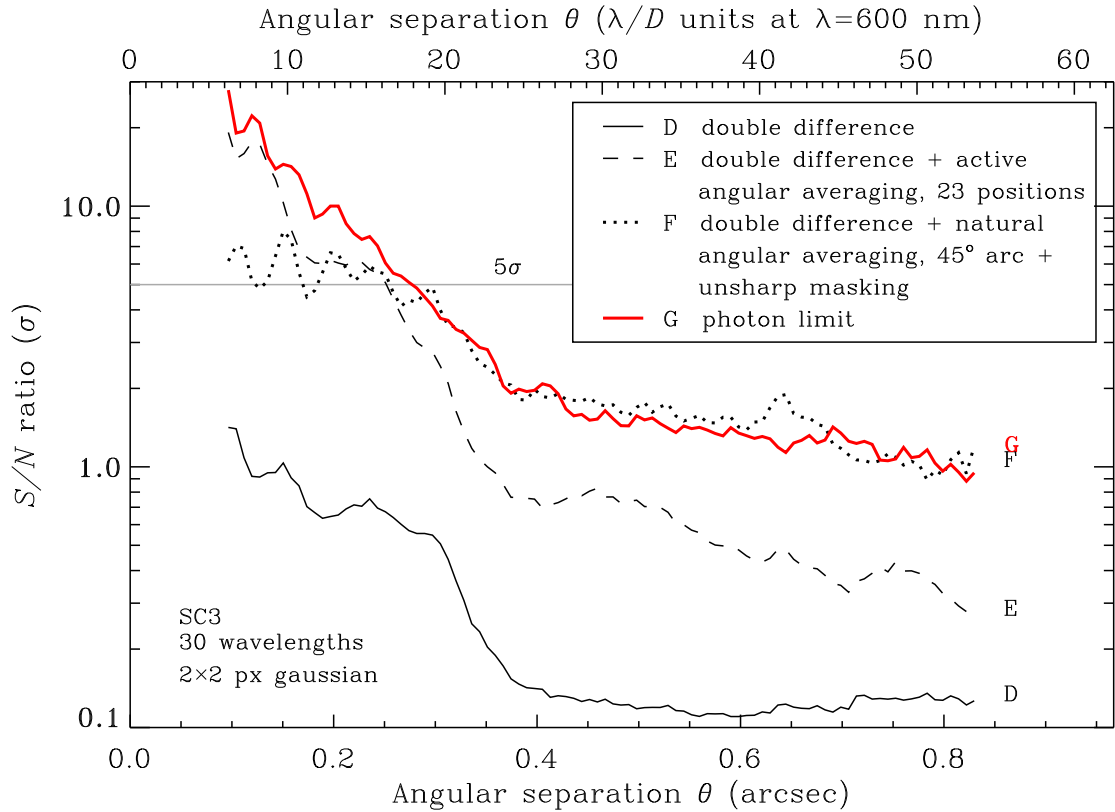
Fig. 5.9: Signal-to-noise ratios for a 25% polarized planet under SC3 conditions as a function of the angular separation θ. The curves nomenclature corresponds to Figure 5.8.

Figure 5.9 shows the signal-to-noise ratios for a 25% polarized planet in the images for the detectivity curves D–G. We come to the following conclusions regarding the simulated performance of SPHERE/ZIMPOL under SC3 conditions:

- A double-difference image resulting from polarimetry and signal switching is too noisy for planet detection.

- Angular differential imaging can lower the noise level in a double-difference image significantly, potentially down to the photon limit.

- In pure statistical photon noise, planet detection is achievable with confidence levels ranging up to $S/N = 25$.

- Even in the photon-limited ideal case, the possibility of planet detection is restricted to the "hot zone" of $0\rlap{.}''1 \lesssim \theta \lesssim 0\rlap{.}''35$. The inner boundary is given by the coronagraphic mask, the outer by the decreasing planet brightness and the control radius of the AO, within which the stray light suppression is efficient.

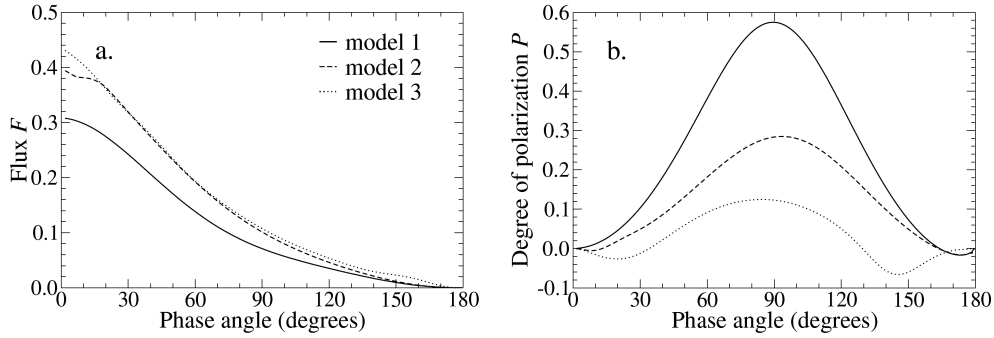- Within that range, planets with smaller angular separations from their star are easier to detect.

Fig. 5.10: Simulated phase and polarization curves of an exoplanet as a function of the phase angle ϕ for three different model atmospheres. Model 1 is a pure Rayleigh-scattering gas atmosphere, Model 2 adds a haze layer in the troposphere, and Model 3 features an additional cloud layer in the stratosphere. These graphs are taken directly from Stam et al. (2004).

5.5 Parameter Dependences

While the Standard Case is chosen to be as representative as possible, an actual observation is likely to deviate from it in many ways. This section describes the way the performance responds to changes in the astronomical, instrumental and computational simulation parameters, allowing us both to determine the technical specifications for the instrumental components required to achieve the best performance possible, and to estimate the population of viable astronomical targets available.

In many cases, the aberration-dominated noise curves are found to be sensitive to a given simulation parameter, whereas the photon noise level remains unaffected. An image treated with angular differential imaging is on the threshold between the two regimes of aberrations and photon statistics, tending towards the latter. The exact effect on the final performance is strongly dependent on the ADI strategy and the efficiency with which it can be applied, both of which have yet to be determined through experience.

For this reason, we separate the two regimes, choosing the untreated *double-difference* level (curve D in Figure 5.7) as the representative benchmark for parameter dependence in the aberration-dominated regime. In cases where the photon-dominated regime is also noticeably affected by the given parameter variations, we furthermore provide the pure *photon-noise level* (curve G in Figure 5.7) as the benchmark for this regime.

5.5.1 Dependence on Planet Characteristics

In the context of this simulation, a planetary body is characterized by the albedo $a(\phi)$, polarization $p(\phi)$ and angular separation $\theta(\phi)$ at the fixed phase angle $\phi = 80°$, its physical radius r. The measurable signal from such a planet is directly proportional to a, p, and the planet's cross-section πr^2, whereas the noise level is unaffected.

$$S/N \propto p,\ a,\ r^2$$

The albedo is highest for "full moon" phase angles around $\phi = 0°$, peaking in $a(0°) = a_{\mathrm{g}}$, and vanishes for "new moon" phase angles around $a(180°) = 0$. The degree of polarization p, on the other hand, vanishes in both of these cases ($p(0°) = p(180°) = 0$), rising to its maximum around the "half moon" phase angles $\phi = \pm 90°$. The resulting polarized planet signal, being proportional to $a \cdot p$, is maximal for $\phi = \pm 80°$ and zero for $\phi = 0°$ and $\phi = 180°$. For a detailed treatment of phase curves and planetary atmospheres, refer to Stam et al. (2004) and Sudarsky et al. (2005).

In a planetary system seen edge-on from Earth, the planet is therefore only observable near the outer extremes of its projected orbit. In a face-on system, on the other hand, the phase angle is a constant $\phi \cong 90°$, providing a continuously high polarized flux. However, assuming an isotropic population of planetary systems, the statistical occurrence of systems with an inclination angle ι between the angular momentum and the line of sight is proportional to $\sin \iota$, favoring edge-on systems.

For a given parent star, a planet's brightness and therefore its signal strength is proportional to the fraction of the star's total light flux it intercepts, which is proportional to the inverse square of its orbital radius, R^{-2}. Since $R = d \cdot \theta(\phi) \cdot \sin \phi$, the signal strength for a fixed $\phi = 80°$ is also proportional to θ^{-2}. The noise curves, on the other hand, are non-trivial functions of θ, as illustrated by detectivity plots throughout this chapter. In general, though, the noise curves are less steep than the decline in planet signal strength. As a result, S/N increases with decreasing angular separation, favoring planets in close orbits.

5.5.2 Dependence on Target Distance

The distance from Earth to the target system is the most limiting external factor for the ZIMPOL planetfinder. While the detectivity of exoplanets around α Cen is exceedingly favorable, it decays very rapidly with distance, dropping below the 5σ threshold at $d = 5\,\mathrm{pc}$ under SC3 conditions.

With increasing distance d from the observer, a given angular separation θ between planet and guide star corresponds to a proportionally increasing physical distance $R = d \cdot \theta \cdot \sin \phi$. For an observable planet, the phase angle ϕ is close to $\pm 90°$. The fraction of the star's total light flux incident on the planet is proportional to R^{-2}, thus the relative brightness contrast between planet and star at a fixed angular separation diminishes with the distance modulus d^{-2}.

In the aberration-limited case, the absolute brightness and the corresponding level of photon noise is largely inconsequential. The dependence of performance on target distance in the double-difference image is therefore

$$S/N \propto d^{-2}.$$

In the photon-limited case, the relative noise level is locked to the inverse square root of the photon flux received from the star. This brightness in turn scales with a distance modulus of d^{-2}. As a result, the noise level is proportional to d. The dependence of performance on target distance in the photon-noise image is therefore

$$S/N \propto d^{-3}.$$

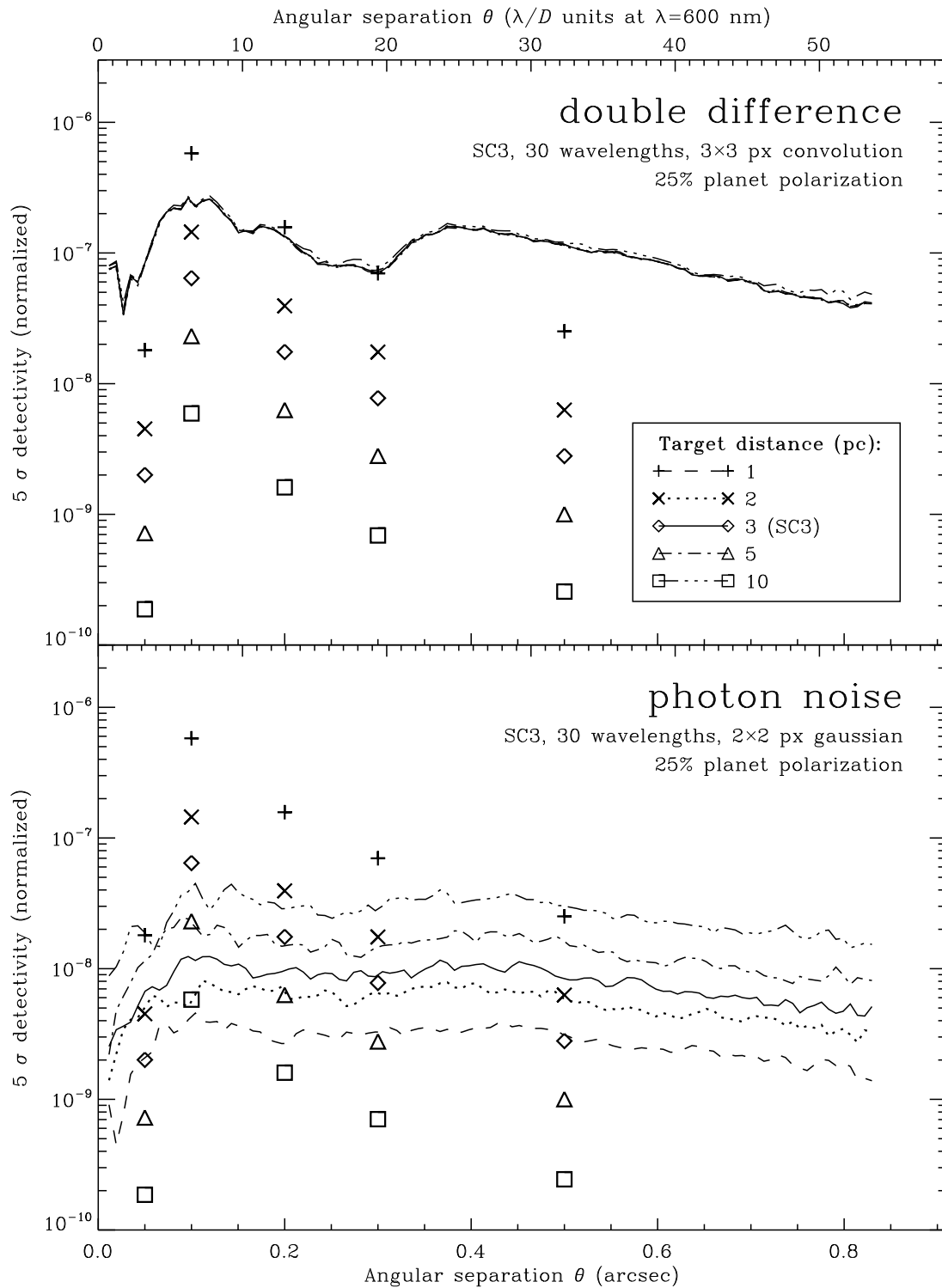Figure 5.11 illustrates these dependences.

Fig. 5.11: Dependence of detection performance on the distance d to the target system. The S/N ratios decrease with d^{-2} in the double-difference image, and with d^{-3} in the photon-noise image.
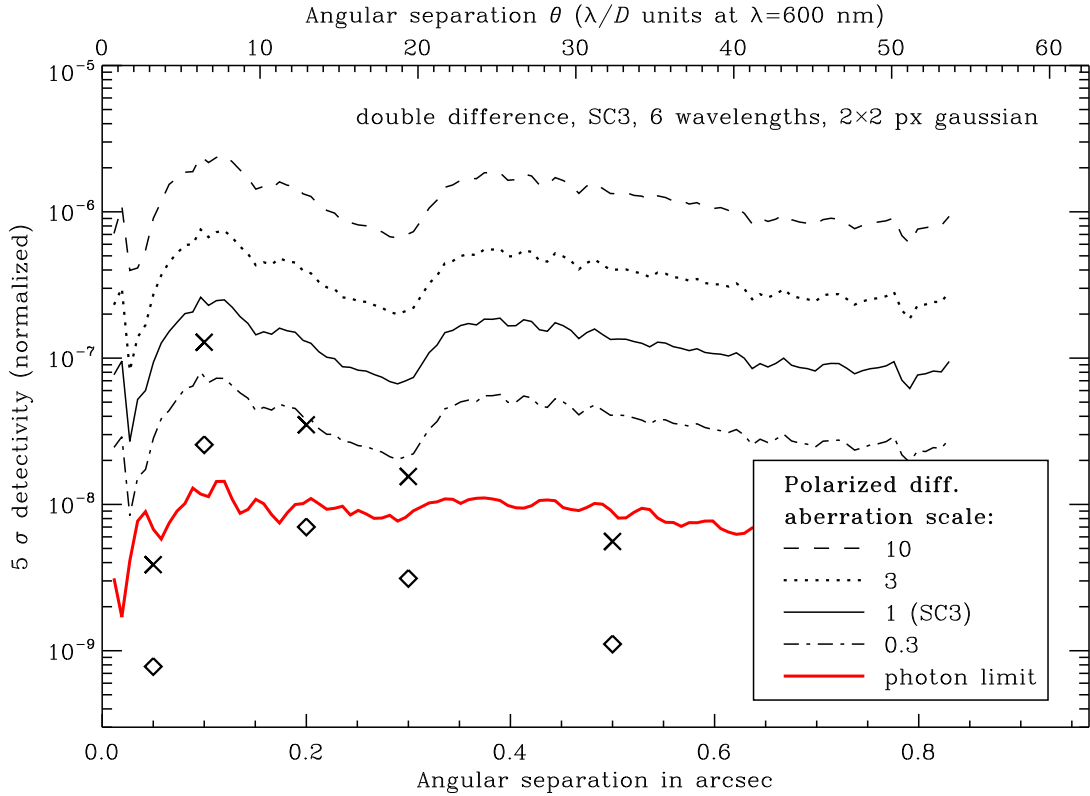
Fig. 5.12: Dependence of double-difference performance on the polarized differential aberrations. The S/N ratio in the aberration-dominated regime is inversely proportional to that factor.

5.5.3 Dependence on Polarized Differential Aberrations

The differential aberrations between the two polarization channels, caused by the polarization modulator (FLC), are a highly critical internal factor of the ZIMPOL planetfinder. They directly limit the gain achieved by polarimetry, the primary differential method of the instrument.

The effect is described by three separate simulation parameters: Differential pointing (a gradient in the wavefront), phase aberrations (higher-order wavefront errors), and the beamshift effect on subsequent optical surfaces due to the differential pointing.

The noise level in the double-difference image is directly proportional to the level of these aberrations, thus the S/N ratio is inversely proportional to that level. Once the aberration noise drops below the photon noise level, the S/N decouples and stabilizes at the photon limit, which is insensitive to the polarized aberrations.

Figure 5.12 illustrates this dependence of double-difference performance on the polarized differential aberration level. The default SC3 values for the three simulation parameters have been scaled by factors of 10, 3, 1 and 0.3, respectively. The photon limit is given as a reference.

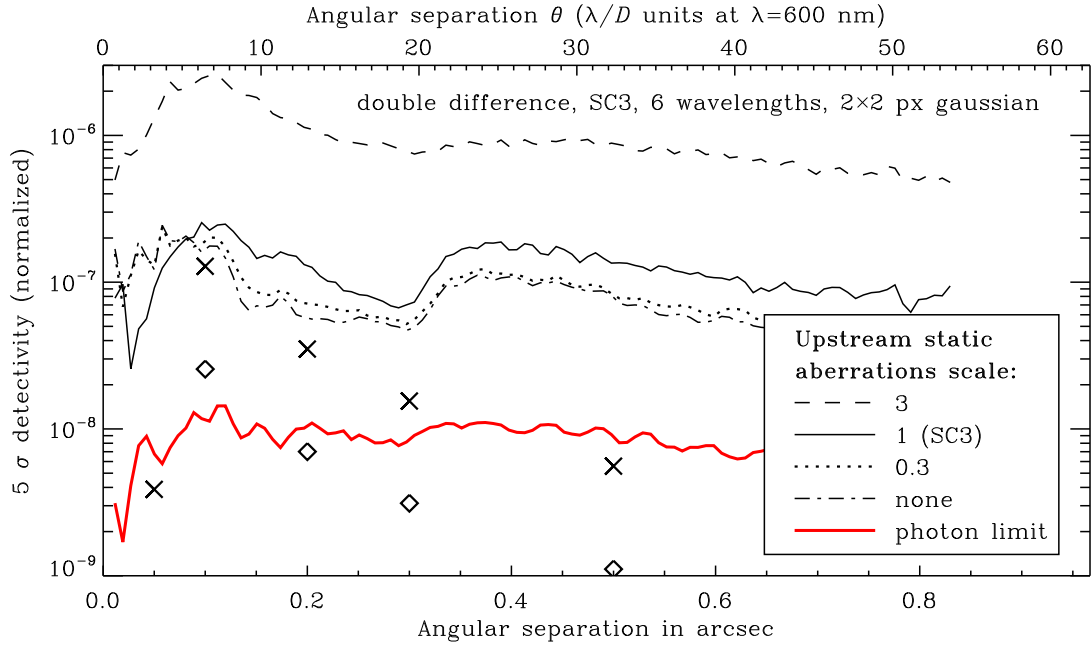Being on the order of $1\,\mathrm{nm}$ RMS, the differential phase and pointing aberrations

Fig. 5.13: Dependence of double-difference performance on the static aberrations upstream of the coronagraph.

of the polarization modulator are difficult to measure and to correct. The manufacturer of our modulators was not aware of their existence prior to our inquiry. While it is questionable that the aberrations can be avoided with technical improvements of the manufacturing process, we will commission the production of several modulators and select the specimen with the minimal differential pointing.

5.5.4   Dependence on Static Aberrations

Another very sensitive internal factor is the amount of static aberrations that affects the wavefront before it reaches the coronograph. These are referred to as *upstream* static aberrations. If the PSF is distorted when it falls onto the coronographic mask, it will spill more starlight past the mask into the final image, and amplify the effect of subsequent aberrations (including the particularly detrimental differential ones).

The upstream static aberrations comprise four input parameters in the COP module: Instrument aberrations, AO calibration, Fresnel propagation, and beam shift. The latter two are chromatic effects that can be set to zero for the ZIMPOL instrument, since its science detector operates in the same spectral band as the AO wavefront sensor.

Figure 5.13 illustrates the dependence of double-difference performance on the general level of those upstream static aberrations. The SC3 values for instrument and AO calibration aberrations have been scaled by a factor of 3, 1, 0.3 and 0, respectively. The effect is dramatic: Increasing the aberrations by a factor of 3 from the SC3 values raises the noise level by a factor of ~ 10. On the other hand, a factor
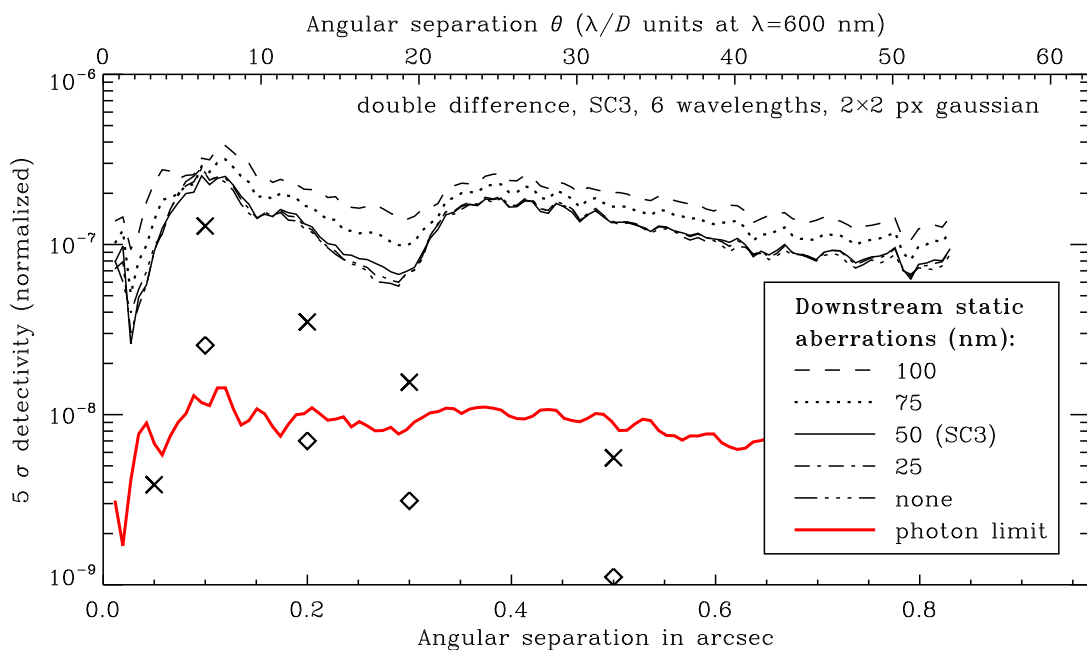
Fig. 5.14: Dependence of double-difference performance on the static aberrations downstream of the coronagraph.

of 2 in double-difference S/N could be gained by reducing the aberrations below the SC3 level before a plateau is reached.

In contrast, static aberrations *downstream* of the coronograph can no longer disperse light from the target star's bright PSF core, and therefore are much less critical.

Figure 5.14 illustrates the performance with downstream static aberrations of 25 nm, 50 nm (the SC3 value), 75 nm, and 100 nm. For reference, the unphysical case of no static aberrations (0 nm) is also included. As expected, the impact on performance is less dramatic than with the static aberrations upstream of the coronagraph. The SC3 noise level is barely distinguishable from the level achieved with zero downstream static aberrations.

5.5.5 Dependence on the Coronagraph Type

Focal Masks

The baseline coronagraph design for SPHERE/ZIMPOL is a classical Lyot coronagraph with a focal mask radius of $5\,\lambda/D$ at $\lambda = 600\,$nm. It blocks a large amount of starlight and yields a good performance in the AO control area of $0.''1 < \theta < 0.''3$. However, it is also foreseen to have other coronagraphs available for specialized applications such as the imaging of circumstellar disks or the search for exoplanets at very small angular separations.

Figure 5.15 illustrates and compares the simulated performances at the double-difference and photon-limit stage with four different coronagraph types. An intro-
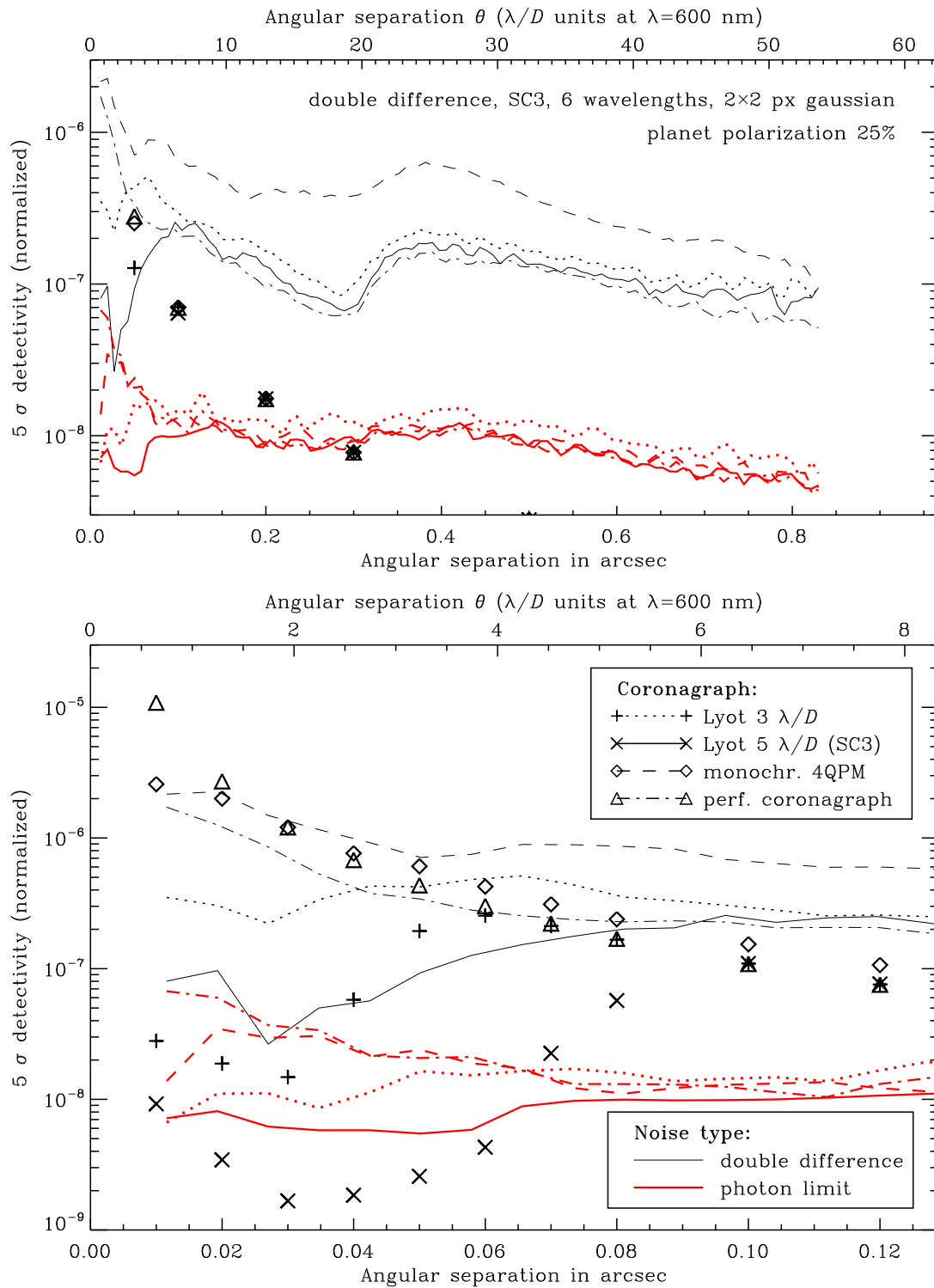
Fig. 5.15: Dependence of performance on the coronagraph type. The upper panel shows the behavior in the outer field while the lower panel focuses on the inner region of $\theta \lesssim 0\rlap{.}''1$. Both the double-difference (thin lines) and photon-limit (thick line) detectivity curves are given.

duction to coronagraph designs has been given in Section 2.5.

The first two are classical Lyot coronagraphs, consisting of a circular opaque mask in the focal plane with radius $3\,\lambda/D$ and $5\,\lambda/D$ at $\lambda = 600\,\mathrm{nm}$, and a pupil stop with circular aperture and circular inner obscuration in the pupil plane. The radii of the pupil stop are chosen such that 63% and 80% of the original pupil area remain clear. These numbers are the result of an optimization calculation by Anthony Boccaletti (Boccaletti et al., 2007). The second of these coronagraphs is the SC3 baseline design.

The third design is a monochromatic four-quadrant phase mask (4QPM). Instead of an opaque mask, this design features a composite retarder plate in the focal plane. It is made of four separate retarder plates arranged in a square such that their edges form a cross centered on the target star. The retardance of the component plates alternates between zero and half-wave as one moves from quadrant to quadrant. As a result, the sign of the complex amplitude of the stellar PSF is reversed in half of the focal plane area. In the subsequent pupil plane, these two components interfere destructively, rejecting the light into a pattern outside of the pupil aperture, where it is absorbed by a pupil stop as in the Lyot case.

The fourth coronagraph considered in this comparison is a perfect coronagraph, defined as a device that precisely removes the original star PSF from the image without disturbing anything else. This is a purely theoretical construct without a physical counterpart, intended as a guideline to the maximal performance that can be expected from coronagraphy. It is implemented in the simulation as a simple subtraction of a flat pupil image from the wavefront at the stage where the pupil stop is applied for the other coronagraph designs.

In a perfect system, such a 4QPM coronagraph completely eliminates the stellar PSF. However, the asymmetries in a realistic PSF caused by aberrations will result in residual light in the final image. Furthermore, the 4QPM design foreseen for SPHERE/ZIMPOL is monochromatic, i.e. the retardance values are only accurate for a particular design wavelength, further reducing the effectivity of the coronagraph.

Figure 5.15 shows that the double-difference noise curves of the Lyot coronagraphs follow the perfect coronagraph rather closely across most of the angular separation range. Among the Lyot designs, the wider mask ($5\,\lambda/D$) yields a slightly lower noise level, since it swallows more of the primary starlight that is scattered into spurious background structures by subsequent aberrations. Finally, the 4QPM double-difference noise level lies well above those of the Lyot designs by a varying factor of 2–4. From these considerations, the $5\,\lambda/D$ Lyot is the obvious choice.

However, the situation changes as soon as the planet signals are taken into account. While they are insensitive to the coronagraph type over most of the angular separation range (by definition of the y-axis normalization), their signal quickly vanishes when the focal mask of the Lyot coronagraphs covers their PSF core. This defines an inner working angle $\theta_{\mathrm{IWA}} = 0\rlap{.}''07$ for the $5\,\lambda/D$ mask and $\theta_{\mathrm{IWA}} = 0\rlap{.}''04$ for $3\,\lambda/D$, within which a planet is obscured. In the case of the monochromatic 4QPM, on the other hand, the planet signals continue to rise with θ^{-2}, detaching only in the innermost $0\rlap{.}''01$. Even there, the signal strength is extremely strong, and the

S/N reaches its maximum.

The photon-limited detectivity curves adhere closely to that of the perfect coronagraph over the entire usable domain of each coronagraph. The $3\,\lambda/D$ Lyot design is again slightly noisier than the rest, this time due to the decrease in photon statistics caused by the low transmission of its pupil stop (63%).

In the photon-limited case under SC3 conditions, the 4QPM is the superior design, achieving the best possible performance in the whole field and suffering from no inner working angle restrictions. However, the double-difference noise level is higher than in the $5\,\lambda/D$ Lyot case, rendering the task of suppressing the aberration noise with ADI to a achieve a photon-limited image more challenging. Furthermore, the 4QPM is much more sensitive to alignment stability (see Section 5.5.6).

For a 25% polarized planet, the photon-limited S/N ratio after four hours of observation reaches 5 at $\theta = 0\farcs25$ and rises to ~ 20 at $\theta = 0\farcs1$ for both candidate coronagraphs. Further in, the Lyot coronagraph turns blind, while the S/N continues to rise for the 4QPM design, reaching 100 at $\theta = 0\farcs01$.

Summary: The standard Lyot coronagraph with a focal mask radius of $5\,\lambda/D$ is the best option for angular separations down to $\theta_{\mathrm{IWA}} \approx 0\farcs1$. A four-quadrant phase mask is required to unlock smaller angular separations down to $\theta_{\mathrm{IWA}} \approx 0\farcs01$, where particularly high S/N ratios could be achieved. The $3\,\lambda/D$ Lyot coronagraph can be useful to narrow down the inner working angle in case the temporal pointing stability turns out to be prohibitive for the use of the 4QPM coronagraph.

Pupil Stops

The SC3 evaluations of coronagraph types assume the use of a particular set of Lyot pupil stop in conjunction with the Lyot focal masks. These inner and outer radii of those stops have been optimized for the monochromatic case by Boccaletti et al. (2007). They transmit 80% (for the $5\,\lambda/D$ mask) and 65% (for the $3\,\lambda/D$ mask) of the flux of non-coronagraphic pupil image.

In the extreme broadband case of ZIMPOL observations, however, different pupil stop proportions might achieve a better result than these defaults. A focal mask of radius $5\,\lambda/D$ at $\lambda = 600\,\mathrm{nm}$ will appear as a $3.3\,\lambda/D$ mask at $\lambda = 900\,\mathrm{nm}$, resulting in a wider pupil diffraction pattern in the infrared than in the visible light.

In Figure 5.17, we illustrate the flux density as a function of radius in the broadband relayed pupil image of a star occulted by a $5\,\lambda/D$ and $3\,\lambda/D$ Lyot mask, respectively. One can clearly discern a diffraction peak at either edge of the pupil aperture (the primary mirror edge at $r = D/2$ and the central obstruction from the secondary mirror at $r = 0.14\,D/2$), and a plateau of noise in between caused by the upstream static aberrations. Due to the wide wavelength window, however, the transitions between these domains are not well delimited.

For each case, we choose three different inner and outer radii for the pupil stop aperture, ranging from a wide cut that covers only the most obvious diffraction peaks to a tight cut that transmits only the flat section. The transmission values, i.e. the fractions of the flux in a flat pupil image that are transmitted through the open area of the pupil stop, range from 70% to 82% in the $5\,\lambda/D$ case, and from 56%
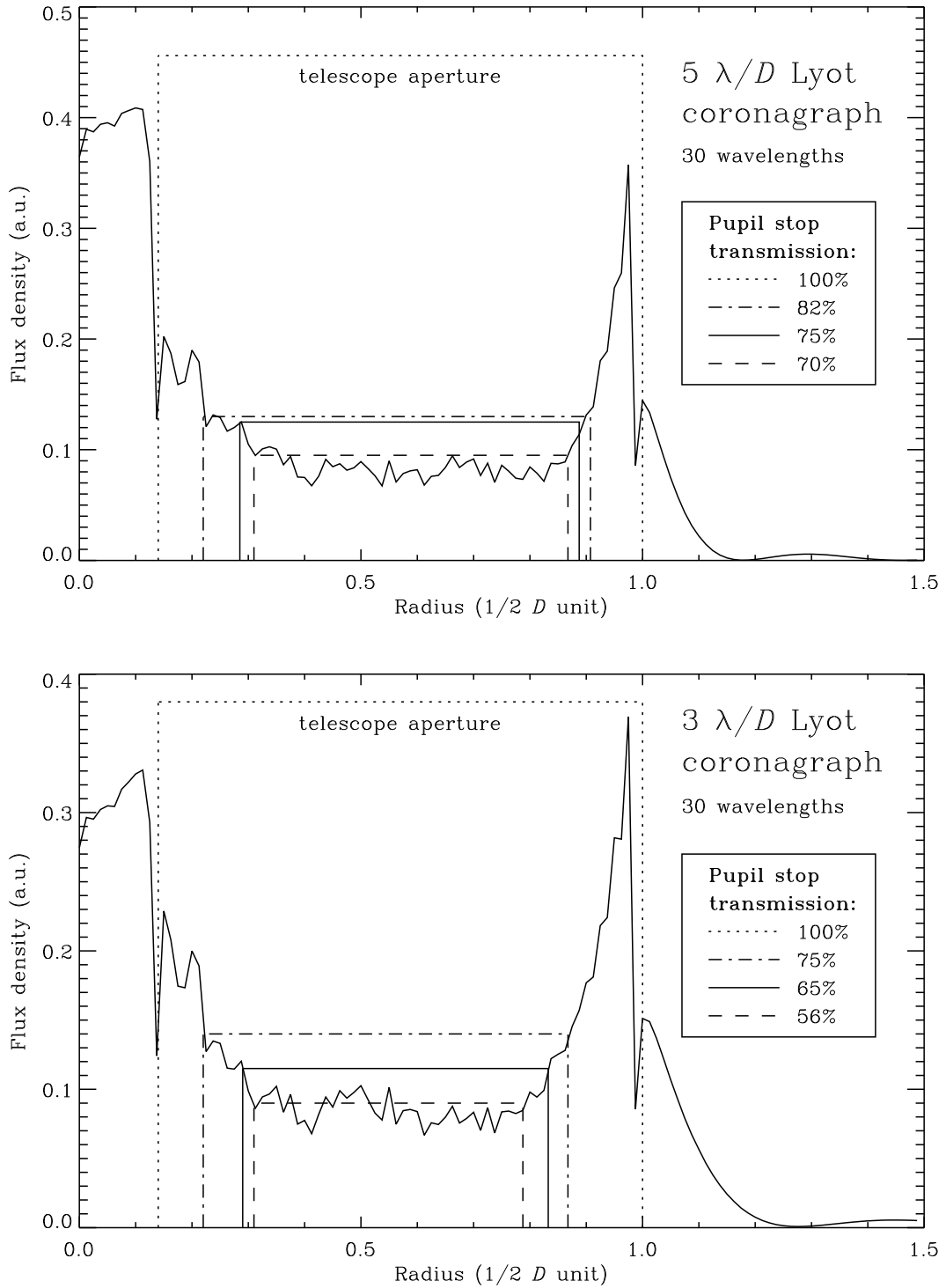
Fig. 5.16: Broadband radial intensity profiles of the relayed pupil for an on-axis source after a classical Lyot focal mask of $5\,\lambda/D$ (top) and $3\,\lambda/D$ (bottom) radius. The dotted bracket marks the inner and outer edges of the imaged telescope aperture, delimited by the secondary and primary mirror, respectively. The other brackets show the inner and outer edges of the chosen Lyot pupil stops.
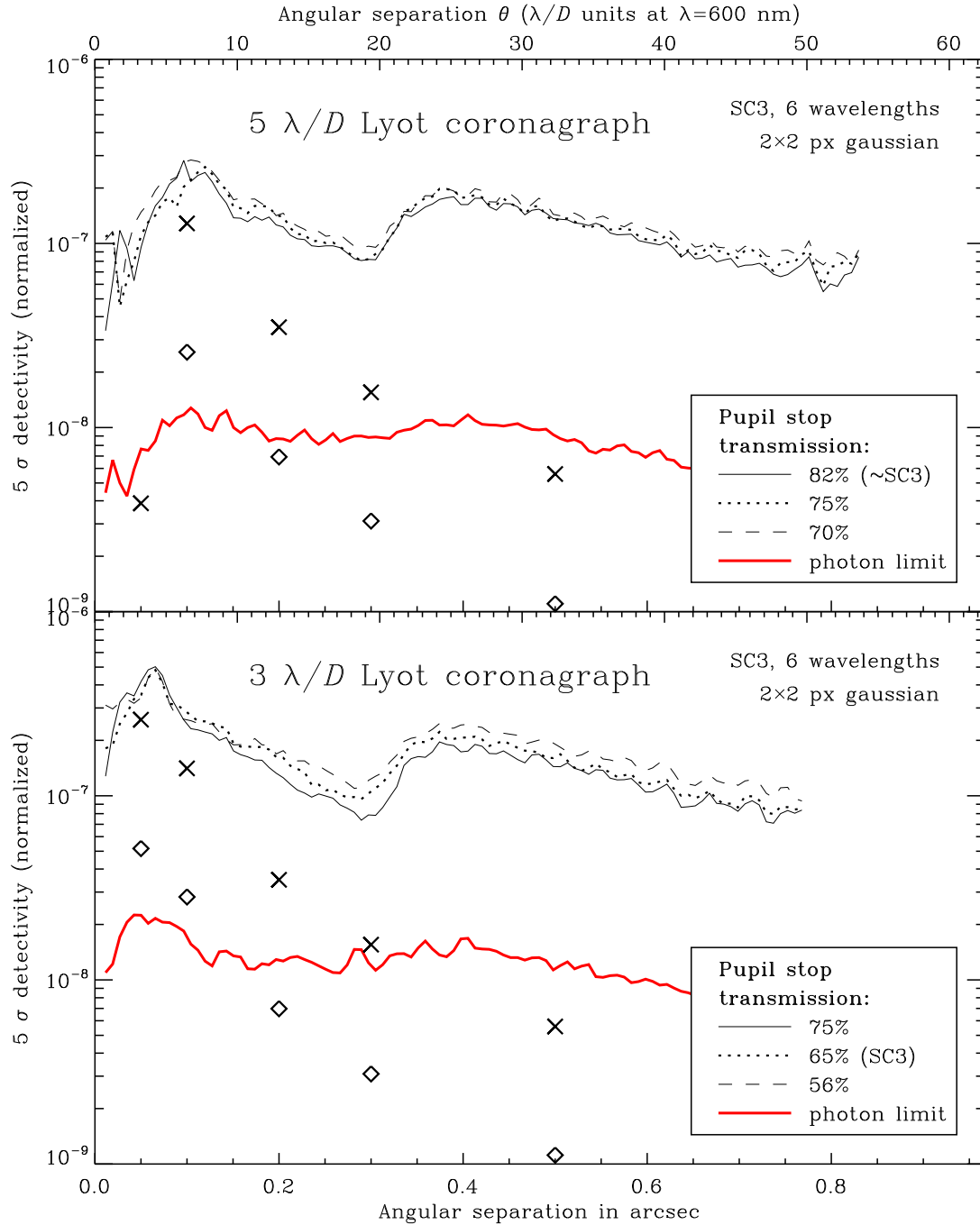
Fig. 5.17: Dependence of double-difference performance on the choice of pupil stop. The inner and outer radii of the pupil stops are defined in Figure 5.16.

to 75% in the $3\,\lambda/D$ case. For comparison, the transmissions of the stops optimized for the monochromatic case (which are used in SC3 simulations) are 80% and 65%, respectively.

Figure 5.17 presents the double-difference performance achieved with each of these pupil stops under SC3 conditions. The differences are rather subtle; however, the widest stops (with the highest transmission) appear to yield a consistently higher performance than their alternatives over the whole usable span of angular separations. Although the noise in these simulations is dominated by aberrations, the slight gain in photon statistics provided by the high-transmission stops is noticeable. On the other hand, the better reduction of residual starlight around the coronagraphic mask image by the tight-cut pupil stops has a vanishingly faint effect on the noise curves.

These evaluations conclude that a pupil stop of $\sim 82\%$ [75%] transmission is optimal for broad-band use of the 5 [3] λ/D classical Lyot coronagraph.

5.5.6 Dependence on Temporal Aberrations

Since the time scale of signal switching is on the order of five minutes, the target and reference images cannot be considered simultaneous. This gives rise to differential aberrations caused by temporally variable conditions between the two image channels. We refer to such effects as *temporal aberrations*.

The most direct temporal effect is due to the mechanism of signal switching itself. When the half-wave plate HWP2 is rotated by $45°$ in order to invert the incoming signal, the aberrations caused by its optical surfaces also rotate. Although the AO will compensate this to some degree, the high-frequency terms will survive as a temporal aberration.

The quantitative effect is expected to be 11 nm RMS for P1 mode, and 17.8 nm for P2/P3. Figure 5.18 illustrates the performance for high-frequency temporal aberrations of 110, 33, 11 (= SC3), 3.3 and 1.1 nm RMS. No improvement is gained by lowering the value below 3.3 nm, likely because the other temporal aberrations (in particular the temporal pointing stability) begin to dominate at this level.

Technical possibilities to lower the high-frequency temporal aberrations include a very careful alignment and fixation of the rotator stage on which HWP2 is mounted, and ensuring the best attainable surface quality for the half-wave plate itself.

Another source of temporal aberration is the fact that the alignment of the stellar PSF on the coronographic mask is not perfectly stable. This temporal pointing stability is given by the mechanical accuracy of both the telescope pointing and the coronagraphic mask mounting. Since the reference observation for the second level of differential imaging is obtained from frequent signal switching rather than from a previous observation of another star, the relevant time scale is the rate of signal switching (~ 5 minutes) rather than on the order of hours. Therefore, the pointing stability has been assumed to be an optimistic 0.1 mas in SC3.

The standard case with its large coronagraphic mask is quite insensitive to this parameter. The performance starts to degrade at values of 3 mas and above. However, for the 4QPM coronagraph, the exact placement of the stellar PSF on the
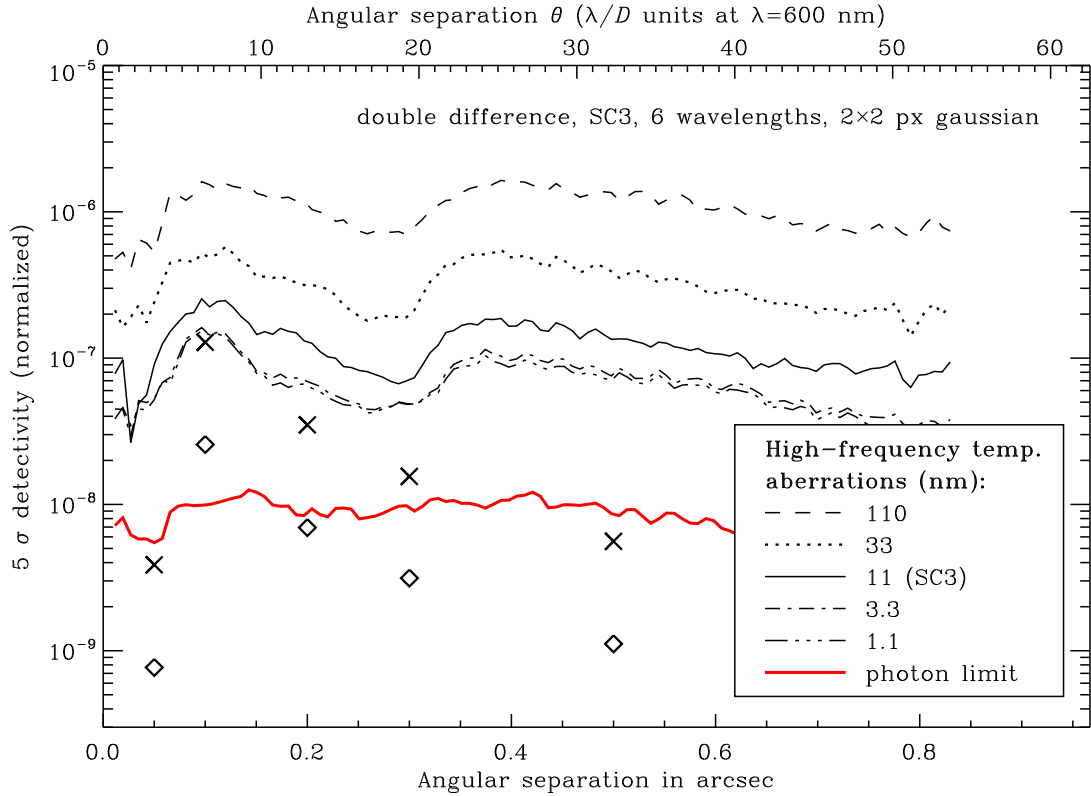
Fig. 5.18: Dependence of double-difference performance on the high-frequency temporal aberrations introduced by the half-wave plate HWP2 during signal switching.

center of the retarder assembly is crucial to its functionality. A pointing offset unbalances the ratio of the two phase-shifted components of the complex amplitude, allowing for light from the PSF core to survive in the pupil area and to reach the detector. For this reason, the performance of the 4QPM already suffers significantly from any increase of the temporal pointing stability over the SC3 value, particularly in the innermost $0\rlap{.}{''}1$ of the field, where the potential detectivity is at its highest. However, it is worth noting that even for the pessimistic case of 1 mas pointing stability, the degraded performance of the 4QPM in this area is still preferable over the Lyot coronagraphs, which obscure the planet's PSF core with their focal masks.

We do not foresee a possibility to improve the temporal pointing stability past the SC3 value, since recent evaluations have shown it to be very optimistic. We are meanwhile aiming at goal of 0.3 mas, and even that cannot be guaranteed.

Figure 5.19 illustrates the dependence of double-difference performance on the temporal pointing offset for two Lyot coronagraphs with focal mask radii of $3\,\lambda/D$ and $5\,\lambda/D$, as well as a monochromatic 4QPM coronagraph.

Pupil rotation constitutes the third and final contribution to the temporal aberrations in our simulation. Since ZIMPOL observations are not pupil-stabilized, the diffraction pattern of the mechanical support structure ("spider") slowly rotate during the observation. On the 5 minute time scale of signal switching, we have cal-
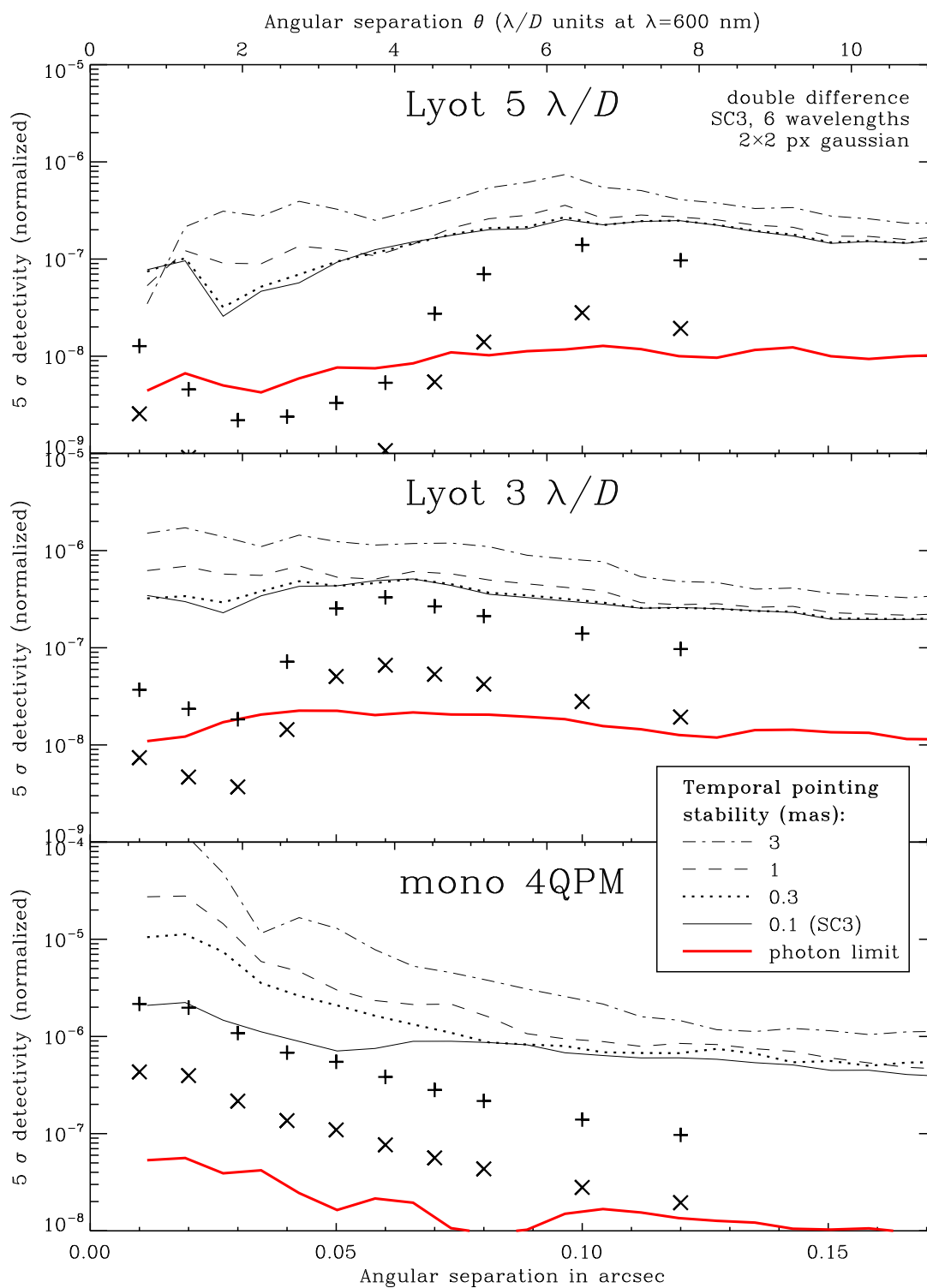
Fig. 5.19: Dependence of double-difference performance on the temporal stability of the telescope pointing, i.e. of the centering of the target star on the coronagraph, for three different coronagraph designs. Only the inner part of the field is shown.
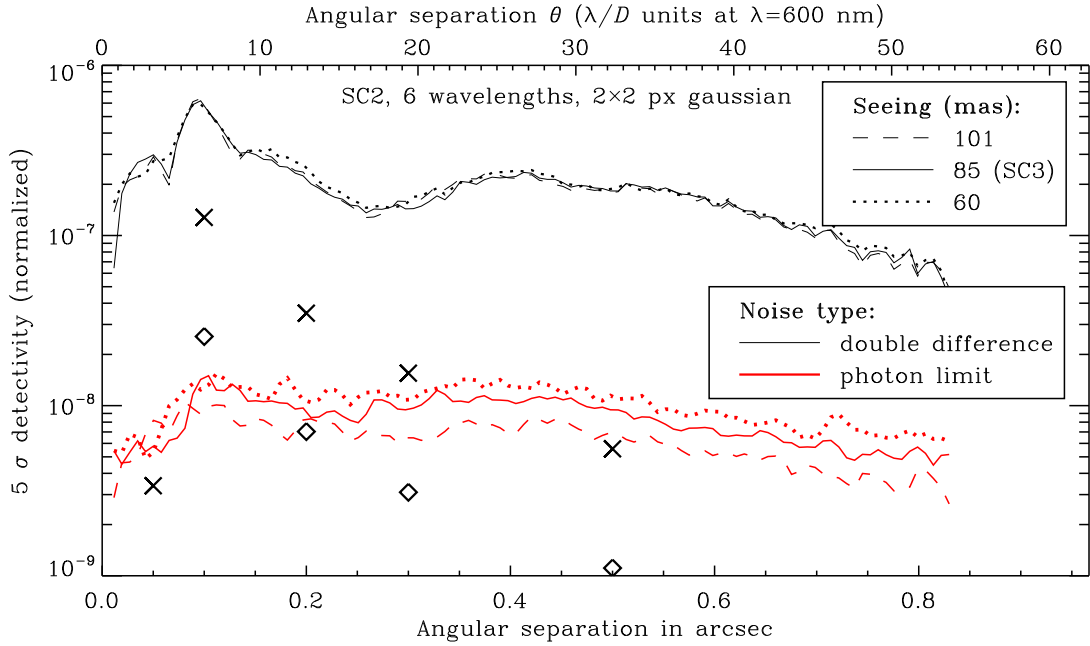
Fig. 5.20: Dependence of double-difference and photon-limit performance on the seeing conditions.

culated an expected pupil rotation of 0.3° for P1 mode (where the derotator remains passive) and 1.2° for P2/P3 mode (field-stabilized). The performances for the SC3 case of 0.3° as well as for 1° turn out to be indistinguishable from the case of zero pupil rotation. Even for a high 3° and 10°, the degradation in S/N is less than a factor of 2.

5.5.7 Dependence on Weather Conditions (SC2)

This test was run under the outdated SC2 conditions and has not been recalculated for SC3. However, the parameter dependences are expected to behave identically. The change from SC2 to SC3 consisted of minor technical details.

A technical memo by Jay A. Frogel[3] shows a table on page 5 that contains statistics on the seeing measured on Paranal between 1989 and 1999. The figures are given for an infinite outer scale, as well as for one of 8 m. In our AOC module, the outer scale is assumed to be 25 m, so the mean value of those two data rows may serve as a good assumption for the CAOS simulation. The 25, 50 (median) and 75 percentile values for the seeing are then $0.''60$, $0.''87$ and $1.''01$. The median is close to the AOC default assumption of $0.''85$.

Figure 5.20 shows SC2 performance for a seeing of $0.''60$, $0.''85$ (= SC2, SC3) and $1.''01$, while keeping the rest of the atmosphere parameters unchanged (this is likely a simplification). The double-difference curves appear identical. This changes in the photon-limited case, where smaller seeing values result in slightly lower noise levels

---

[3] http://snap.lbl.gov/pubdocs/Seeing_at_Observatories_v3.0.doc

. A website by ESO[4] suggests that seeing conditions have worsened since 1999, thus the values assumed above might prove to be slightly optimistic.

5.5.8 Dependence on Computational Limitations

Over the hours of a real telescope observation, countless (on the order of 10^6) uncorrelated instantaneous PSFs are integrated up, whereby the speckle pattern caused by atmospheric aberrations left uncorrected by the AO are averaged out into a smooth halo. Any remaining speckle structure will be due to static aberrations.

It is impractical to calculate that many AO phase screens in the simulation code. However, one finds that the sum of 100 uncorrelated phase screens already provides a very good approximation to the long-exposure limit. The resulting speckle pattern is already completely dominated by the static aberrations at that stage.

The AOC module provides the option to switch off the simulation of AO-corrected turbulent phase screens. In such a simulation, the halo of stray light caused by the atmosphere is missing from the final PSF, thus significantly underestimating the photon noise level. Its speckle pattern, however, being determined exclusively by the aberrations, is identical to that of the long-integration limit of a simulation with an infinite number of turbulent phase screens. Therefore, such a non-turbulent simulation yields perfectly accurate noise levels in the differential images as long as they remain aberration-dominated.

The top panel in Figure 5.21 demonstrates these relations. In the Standard Case, the asymptotic long-exposure double-difference noise curve lies well above the photon noise limit due to the dominating polarized differential aberrations.

A non-turbulent simulation requires only a single iteration of the CAOS code, as opposed to 100 iterations for a realistic turbulent simulation (which takes about 20 hours of computation on our machines). For this reason, most plots in this chapter are based on non-turbulent simulation runs. Turbulent simulations were necessary only for the photon-limited curves.

The default spectral range for ZIMPOL observations is 600–900 nm, a broadband filter spanning the R and I bands. The broadband PSF can be described as the integral of the the monochromatic PSFs over the wavelength range:

$$I_{\mathrm{BB}}(x,y) = \int_{\lambda_{\min}}^{\lambda_{\max}} I_{\mathrm{mono}}(x,y,\lambda) \cdot \mathrm{d}\lambda. \tag{5.2}$$

Since the spatial scale of a monochromatic PSF and its speckle pattern depends linearly on the wavelength, the speckles will be smeared out along a radial trace whose length is proportional to the angular separation from the center and to the relative bandwidth, $(\lambda_{\max} - \lambda_{\min})/\frac{1}{2}(\lambda_{\max} + \lambda_{\min}) = 0.4$ in the SC3 case.

In the simulation code, the integral in Equation 5.2 must be replaced with a summation over a discrete number n of wavelengths. In order to attain a satisfactory approximation of the continuous radial smoothing, the number must be chosen high enough that different instances of a given speckle overlap in the summation.
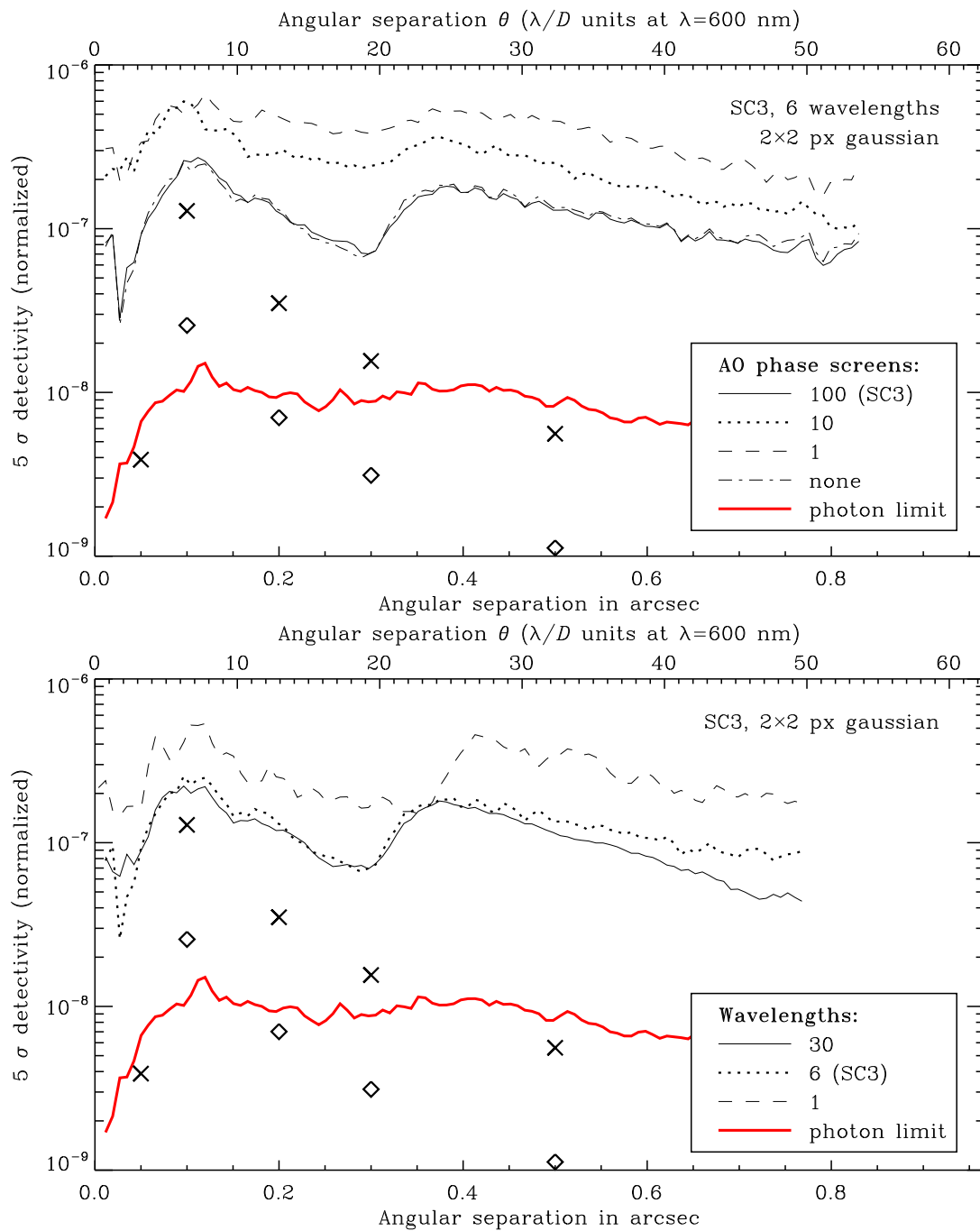
---

[4] http://www.eso.org/gen-fac/pubs/astclim/paranal/seeing/singstory.html

Fig. 5.21: Dependence of double-difference and photon-noise performance on the number of turbulent atmospheric AO phase-screens simulated (top) and the number of wavelengths simulated (bottom). In reality, both numbers are several orders of magnitude larger. The SC3 parameters are simplifications compatible with the available CPU power that approximate the realistic case quite well.

The bottom panel of Figure 5.21 illustrates the dependence of double-difference performance on the number of simulated wavelengths. A monochromatic simulation ($n = 1$) is clearly insufficient. For $n = 6$, the smoothing appears to be well reproduced for $\theta \leq 0''.4$, while the radial speckle traces start to decay into individual speckles further out. The "hot zone" interesting for the purpose of exoplanet detection ends at the AO control radius around $\theta = 0''.3$, thus this kind of simulation is good enough for most of our purposes. For $n = 30$, finally, the PSF is smooth in the whole considered field (up to $\theta \leq 0''.8$).

In this work, most evaluations are based on simulations with 6 wavelengths in order to keep the necessary CPU time short. Only for the detailed presentation of the Standard Case (Figure 5.7), and for tests with natural field rotation and unsharp masking where the smoothness of the PSF is essential, elaborate simulations with 30 wavelengths were calculated.

5.5.9   Dependence on the Filter Band

Since the ideal PSF shape, the AO performance and many instrumental noise effects are wavelength-dependent, it is conceivable that narrowing the ZIMPOL broadband filter range down to a window with a particularly favorable PSF could improve overall performance.

Figure 5.22 shows the double-difference and photon-limit performances for three different filter bands: the default broadband RI (600–900 nm) as well as the two medium-band filters R (590–700 nm) and I (750–900 nm).

For the standard $5\,\lambda/D$ Lyot coronagraph, the double difference noise level in the long-wavelength band I is slightly preferable over that of the SC3 broadband, owing to the higher efficiency of the AO and the larger control radius at those wavelengths. However, the photon limit is less favorable, due to the smaller number of photons collected.

With the 4QPM coronagraph, the broadband filter also appears to be the best choice. Only in the innermost $0''.03$ do the more limited bands win over the broadband filter, likely due to the fact that the wavelength-dependent residual atmospheric dispersion slightly de-centers each monochromatic stellar PSFs on the 4QPM, degrading its nulling. The broadband filter sums over a wider range of such misaligned (and therefore incompletely nulled) stellar PSFs.

In conclusion, restricting the bandwidth of ZIMPOL to the I band is generally not beneficial. However, it can have its uses where the signal extraction methods fail to push the noise down to the photon limit, as it may be the case for very bright targets such as Sirius and α Centauri, or when looking for planets at extremely close angular separations with the 4QPM coronagraph.

5.5.10   Dependence on the Observation Time

Figure 5.23 illustrates the dependence of the detectivity on total integration time. In SC3, the double difference noise level decouples from the photon limit after ~ 10 minutes of observation, when the noise level of the instrument effects is reached.
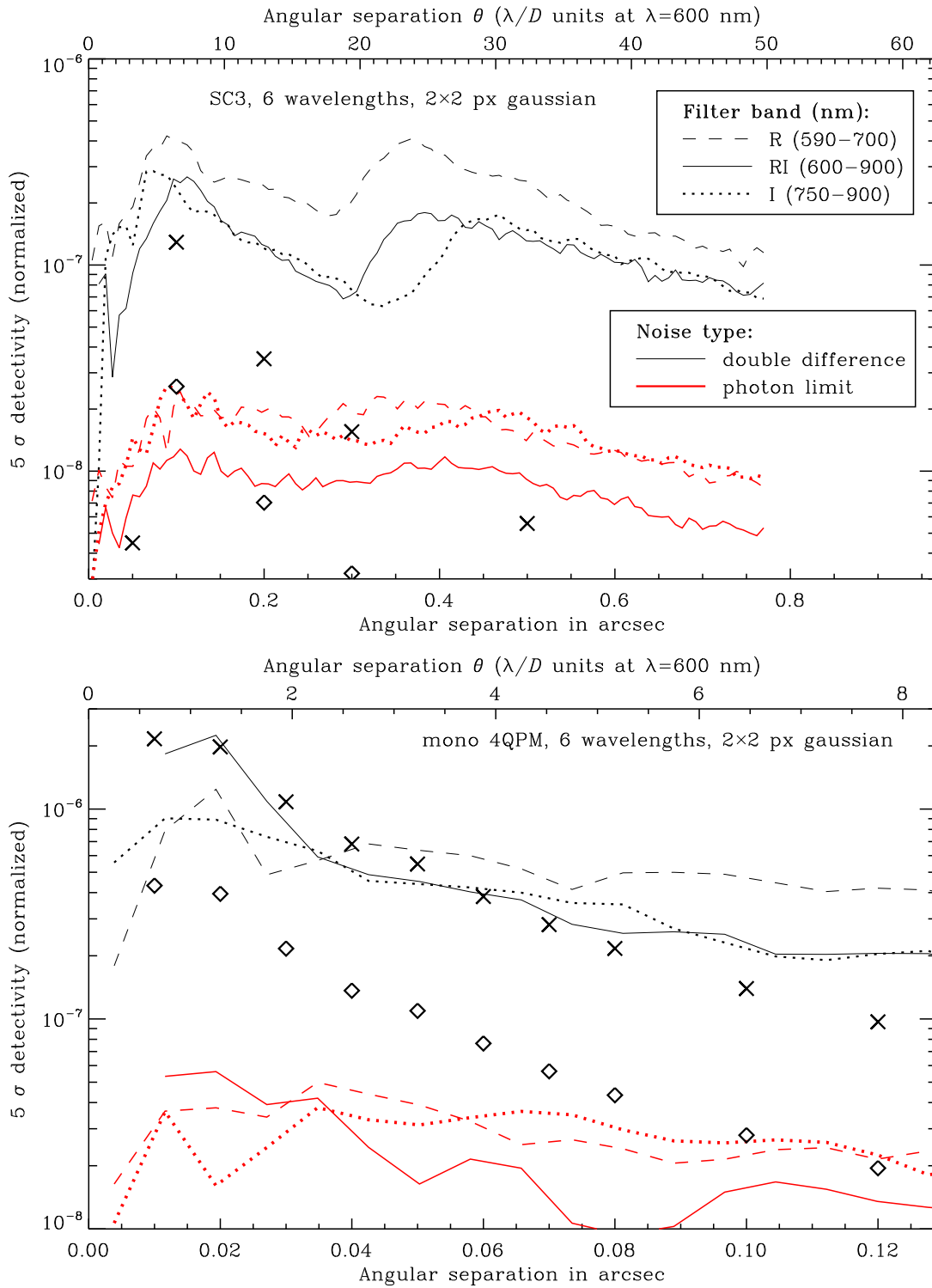
Fig. 5.22: Dependence of double-difference and photon-noise performance on the filter band, for the standard $5\,\lambda/D$ Lyot coronagraph (top panel) and the monochromatic 4QPM (bottom panel). The default broadband filter RI displays the best overall performance.
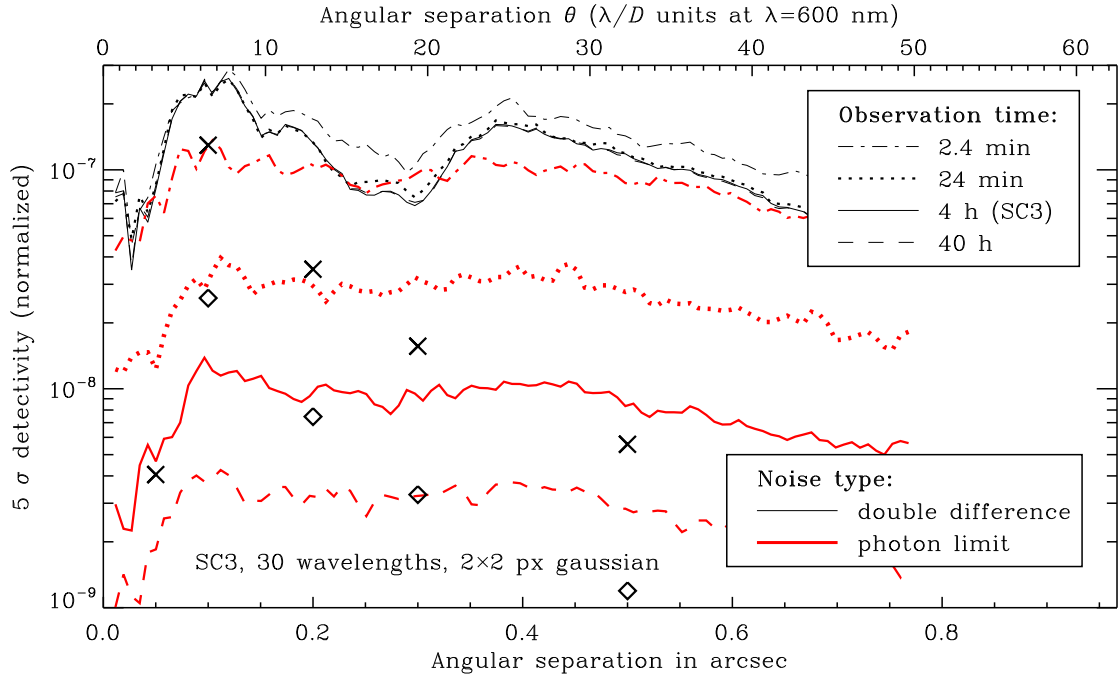
Fig. 5.23: Dependence of double-difference and photon-noise performance on the observation time.

The photon limit continues to decrease with the square root of the integration time whereas the double difference level remains constant. Increasing the absolute magnitude of the target star is equivalent to increasing the observation time by the same factor.

In practice, the fundamental limit to detectivity will be the degree to which data reduction and signal extraction methods will be able to eliminate the non-photonic noise sources. Once a photon-limited detection can no longer be achieved, additional integration time cannot improve detectivity any further.

5.5.11 Dependence on Detector Properties

The two detector-related parameters that have the potential to affect the performance are read-out noise (RON) and the flatfielding (FF) residuals, i.e. the fundamental non-statistical noise limit imposed by detector imperfection and non-linearity.

The SC3 value for read-out noise is 10 photons per hardware pixel per readout. Simulations show that this level is three orders of magnitude too small to significantly raise the SC3 double-difference noise level.

The SC3 assumption for flatfielding residuals is $2 \cdot 10^5$, which has been achieved in lab experiments even before dithering (Thalmann, 2004). Tests show that this level would have to be more than 10 times worse to affect the SC3 double-difference noise curve. Therefore, the detector properties can be considered entirely uncritical to performance.
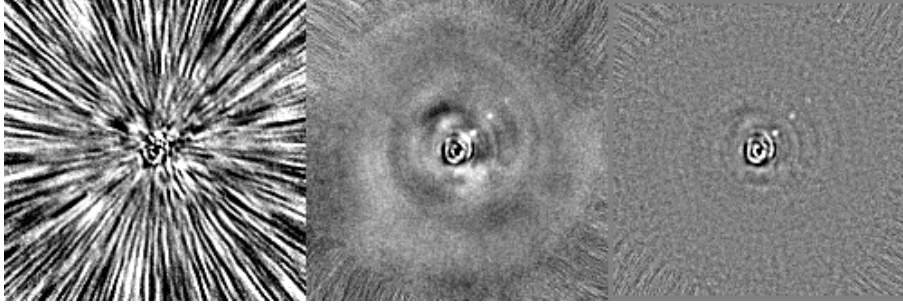
Fig. 5.24: The SC3 double-difference image (left) after angular averaging with 45° of *natural* field rotation (center) and unsharp masking (right). The grayscale is $\pm 2.5 \cdot 10^{-4}$ in all images.

5.6 Effectivity of Signal Enhancement

This section discusses the effectivity of signal enhancement methods as a function of their chosen parameters. The performance in an actual observation will depend on additional circumstances, such as the degree to which the double-difference background remains static over the course of the observation, and on which time scale the non-static component varies.

5.6.1 Effect of Angular Averaging and Angular Differential Imaging

The performance gain provided by image processing methods exploiting field rotation depends on the behavior of the field orientation over course of the observation. We distinguish natural field rotation caused by the telescope's tracking of the target on the rotating sky sphere, usually a slow and continuous drift of the orientation across a limited angular span, from active field rotation, where the orientation is artificially reset in discrete steps in order to cover the full circle of angular space.

Natural Field Rotation

Our code simulates a simplified form of natural field rotation by assuming a uniform angular drift rate of the field orientation. Furthermore, the CAOS code provides a single, final PSF for the whole observation and therefore does not account for the possibility that the PSF shape evolves with time due to quasi-static aberrations.

In an actual observation, changing the field rotation relocates the planet to a different position on the detector, while the instrumental aberration noise pattern remains fixed. The reduction software then derotates each exposure into the rest frame of the planet before summing, thus smearing out the background along an arc while the planet signal stays in place.

To simulate the result of angular averaging with natural field rotation in our code, we directly calculate the final, angularly smeared background image. We split the static stellar PSF into a number of identical copies, rotate each of those by an angle that slowly increases from copy to copy, and then add them back up. Each
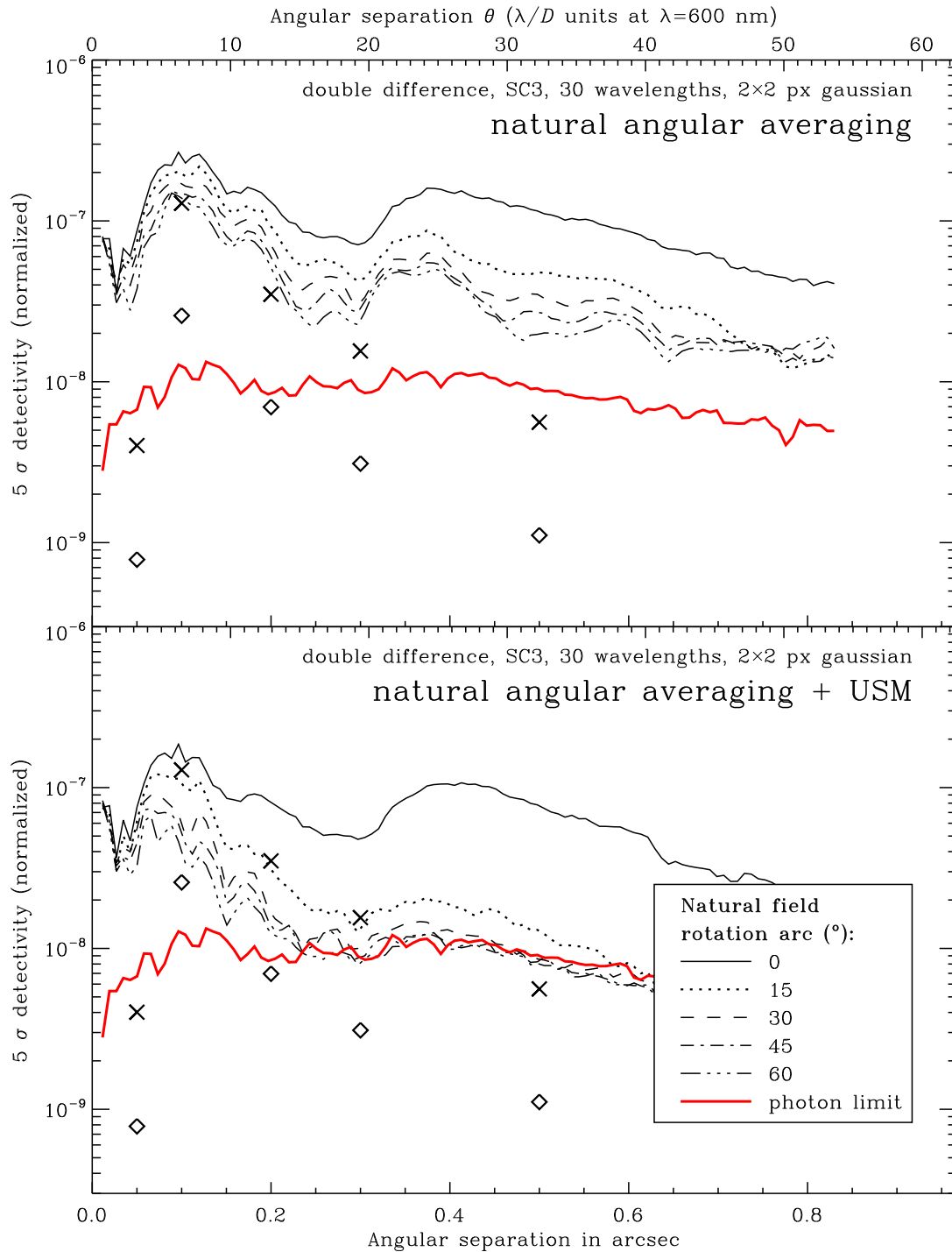
Fig. 5.25: Effect of *natural* field rotation coupled with angular averaging (top panel) and with additional unsharp masking (USM, bottom panel) on double-difference performance. USM is very efficient where the local speckle smearing scale exceeds the smoothing scale, but fails for small separations. Angular differential imaging (ADI) is also possible with natural field rotation; the performance is expected to be similar to the one presented in Figure 5.27.

static speckle in the PSF is smeared out along an arc. The speckle-like noise in the image is thereby averaged down. The S/N gain expected from this process at a given angular separation θ is equal to the square root of the number of resolution elements covered by the field rotation arc at that radius:

$$S/N_{\text{natural}}\,(\theta, \delta) = S/N\,(\theta) \cdot \sqrt{\frac{\delta \cdot \theta}{\lambda/D}},$$

where δ is the total angle by which the field rotated during the observation.

If unsharp masking (described in Section 5.3.2) is applied to the averaged images, the performance improves significantly, reaching the photon limit for $\theta \gtrsim 0\rlap{.}{''}2$. In this area, the speckle pattern is smeared both radially (due to the large relative bandwidth of the ZIMPOL broadband filter) and azimuthally (due to angular averaging), resulting in a smooth background without small-scale structures, which can be efficiently removed with unsharp masking. In the center of the field, where the arc of angular averaging is very small, the background retains some small-scale features able to survive the unsharp masking, thus the gain in S/N is less pronounced there.

Figure 5.24 visualizes the images resulting from this process for a natural field rotation angle of 45°, whereas Figure 5.25 illustrates the effect on performance achieved with different rotation angles.

The CAOS simulation does not account for slowly evolving quasi-static speckles; the background remains completely static during the observation time. A strongly evolving background could slightly reduce the efficiency of USM, since strong and relatively short-lived speckles could leave an azimuthally small-scaled mark on the final image.

Natural field rotation can also be exploited for angular differential imaging (described in Section 5.3.1), as long as the planet signal moves a distance of at least λ/D over the course of the observation. For the $5\,\lambda/D$ Lyot coronagraph, the innermost observable planet lies at $\theta = 6\,\lambda/D$, requiring a minimal field rotation angle of $\delta \geq (\lambda/D)/\theta \approx 10°$. The 4QPM coronagraph, on the other hand, can in principle detect a planet at $\theta = 1\,\lambda/D$, necessitating a field rotation angle of $\delta \geq (\lambda/D)/\theta \approx 60°$. This is a hard constraint for natural field rotation; it is clearly preferable to use active field rotation in this case.

Note that two independent planet images are the bare minimum for ADI, as the planet signal is effectively halved in the background subtraction. In order to keep the planet signal from leaking into the median background, it is desirable to obtain as many non-overlapping planet images as possible. In the separation ranges where planet detection is most likely, active field rotation is better suited to this task than natural field rotation is.

Since our code simulates natural field rotation a posteriori on the basis of a static long-exposure PSF, it is not trivial to reproduce the results of natural ADI in a realistic manner. One can assume that the performance will be similar to ADI with very few positions of active field rotation (Figure 5.26).
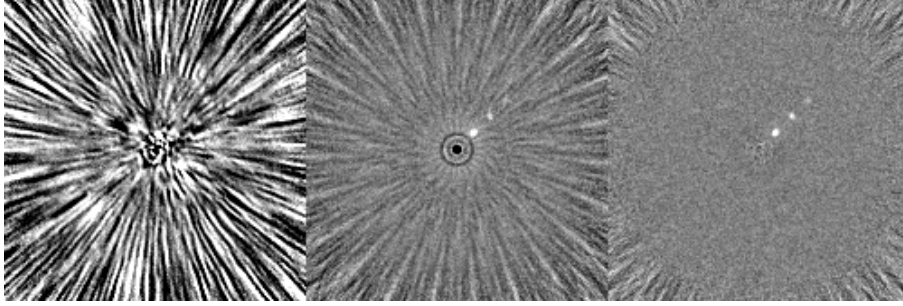
Fig. 5.26: The SC3 double-difference image (left) after angular averaging with 23 positions of *active* field rotation (center) and perfect ADI background subtraction (right). The grayscale is $\pm 2.5 \cdot 10^{-4}$ in all images.

Active Field Rotation

Unlike natural field rotation, the amount of averaging achieved with active field rotation is limited by the number of artificially selected field orientations n rather than by the arc length (which is $360°$ here). Only for high n and small angular separations θ does the available arc space impose a lower limit on the number of non-overlapping samples to be averaged:

$$
S/N_{\text{active}}(\theta, n) = S/N(\theta) \cdot \min\left\{ \sqrt{n}, \sqrt{\frac{2\pi\theta}{\lambda/D}} \right\}.
$$

For the realistic value $n = 23$, the gain is generally larger than the one achieved with natural field rotation. Applying USM to natural field rotation evens the field somewhat (though active field rotation remains better for $\theta \lesssim 0''25$). Due to the low amount of signal overlap between orientations, active field rotation is generally better suited to ADI.

Our implementation of angular averaging with active field rotation makes use of the same simplifications as with natural field rotation. The static stellar PSF provided by CAOS is split into n identical copies, which are individually rotated by a certain angle and finally added back up. In order to make the most of the freedom provided by the active choice of field orientation, the orientation angles are spread evenly across the full $360°$ of angular space, thus each copy is rotated by $360°/n$ with respect to the previous copy. The resulting image features a characteristic n-fold rotational symmetry. The planet signal, on the other hand, remains fixed in its place.

We exploit this symmetry to estimate the effect of ADI retroactively on the basis of the angularly averaged image (whereas in reality, ADI is applied to the individual exposures *before* they are derotated and summed). We add the planet signal into the angularly averaged stellar PSF, then split the resulting images into n copies, which we rotate individually by all multiples of $360°/n$. We calculate the median background of these copies and subtract it from the original angularly averaged image. This removes all the rotationally symmetric parts from the image, leaving
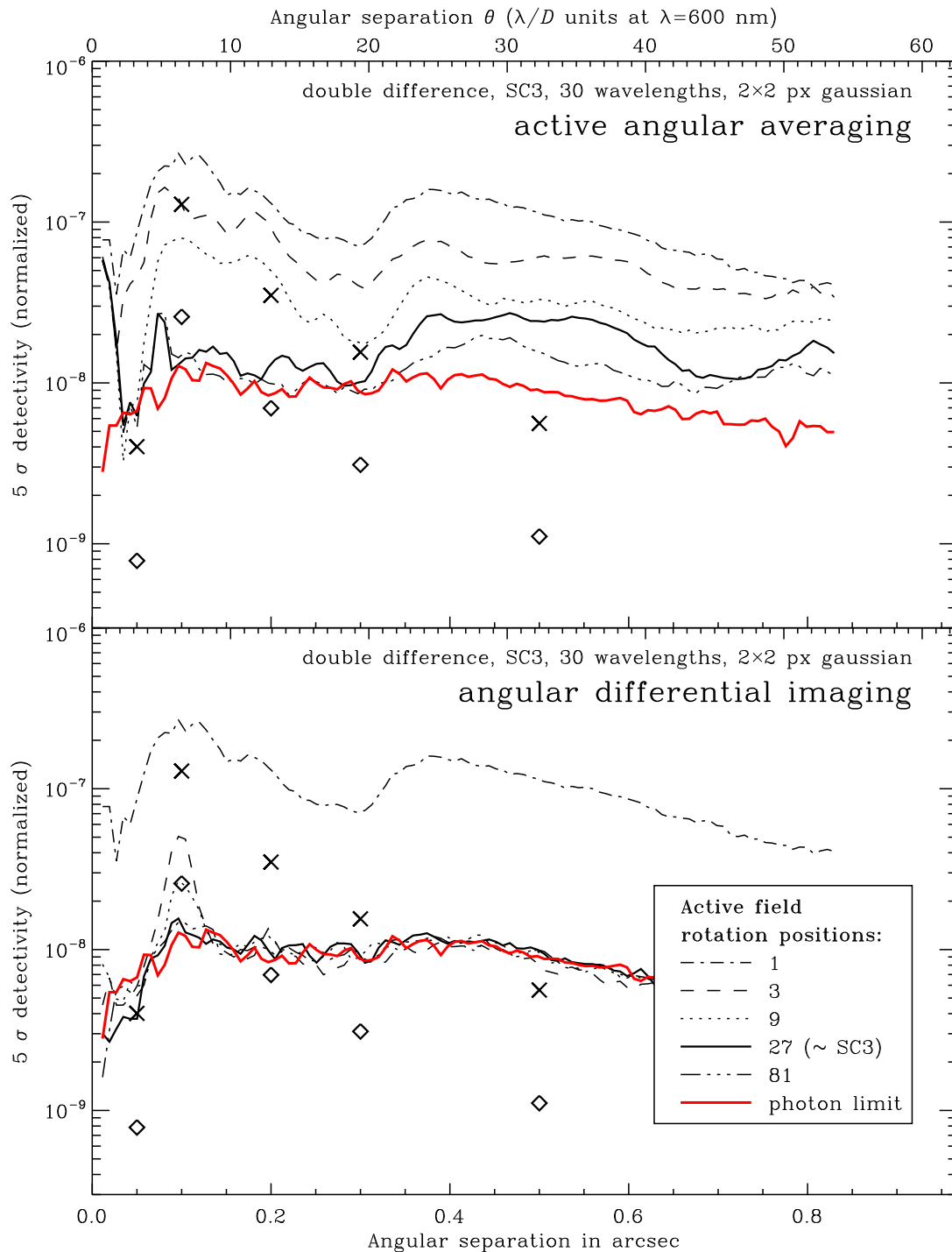
Fig. 5.27: Effect of *active* field rotation coupled with angular averaging (top panel) and angular differential imaging (ADI, bottom panel) on double-difference performance. The ADI case is idealized, featuring a perfectly static background (no quasi-static speckles), and therefore illustrates the upper limit of performance that can be expected from the technique. For a low number of field positions, the brightest planet signals leave "ghosts" behind in the background that produce peaks in the noise curve.

behind only the unique planet signal, the photon noise, and possibly a trail of faint negative planet "ghosts" where the planet signal leaked into the median background.

Figure 5.26 visualizes the images resulting from this process for 23 positions of active field rotation, whereas Figure 5.27 illustrates the effect on performance achieved with numbers of field orientations.

Since the background is completely static in our simulation, the ADI principle works perfectly, reducing the aberration noise down to the photon limit under SC3 conditions. A departure from the photon limit is only observed for significantly higher levels of aberration noise, in which case numerical errors in the software rotation and derotation process start to become relevant. At any rate, the performance yielded by our ADI simulations should be considered the upper limit of expected performance rather than a true prediction.

If the stellar PSF shape should evolve significantly over the course of the observation, e.g. due to quasi-static speckles, the effectivity of ADI would suffer. Although the background patterns caused by different aberrations do not add up linearly, the static aberrations are expected to dominate significantly over the quasi-static ones, keeping the background self-similiar enough for ADI to achieve an advantage over simple angular averaging. Lafrenière et al. (2007) propose to calculate the background landscape as a "moving average" over the observation, calibrating each exposure with a combination of a local selection of other exposures rather than with the median of the whole observation series. This technique could counteract the negative effects of quasi-static speckles to some degree.

5.6.2   Effect of Aperture Convolution

In order to determine the quantitative effects of aperture convolution, we choose the SC3 simulation with high spectral resolution (30 wavelengths) as the benchmark test. We convolve the double-difference and photon-limited images with circular and gaussian sampling apertures ranging from 1 to 6 pixels in full width. In order to improve the reliability of the results, we rerun each test 100 times, averaging out any bias from the individual realizations of the photon noise. Note, however, that the same PSFs (and therefore the same aberration settings) have been used for all tests, thus the double-difference curves, which do not react noticeably to variations of the photon noise, still include a bias from the particular realization of the aberration noise used to generate the PSFs.

Figure 5.28 illustrates the dependence of performance on the width of the sampling aperture for both circular and gaussian apertures. The performance is measured as the S/N ratio of the three simulated planets at $\theta = 0\rlap{.}''1$, $0\rlap{.}''2$ and $0\rlap{.}''3$. The data points in each plot panel are normalized by the S/N ratio achieved with the single-pixel circular aperture, which is essentially the discrete delta function, leaving the original picture unchanged. The value of a data point is therefore the S/N gain factor achieved by application of the aperture convolution.

Evidently, the gain achievable in the double-difference image is negligible. In the aberration-dominated regime, the noise level is limited by spurious structures at the size scale of the planet signal, which survive a planet-sized sampling aperture
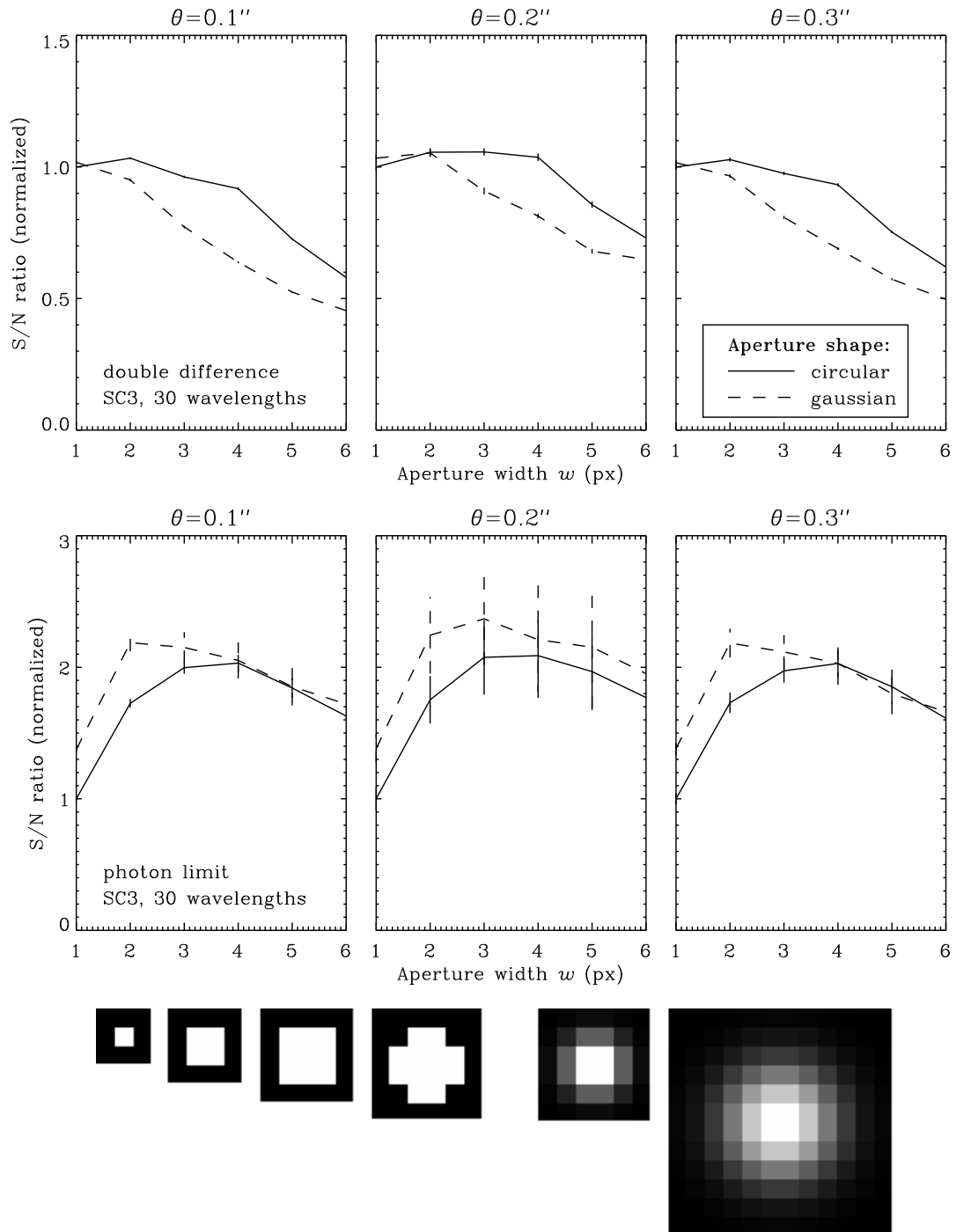
Fig. 5.28: Effect of aperture convolution on double-difference (top row) and photon-limit (middle row) performance. The S/N curves are normalized at the single-pixel circular aperture, which is equivalent to using no convolution. The bottom row shows a graphical representation of the circular apertures with width $w = 1, 2, 3, 4$ and the gaussian apertures with width $w = 2, 4$.

unscathed.

In the photon-limit image, on the other hand, the dominant noise is uncorrelated from pixel to pixel, and thus can be averaged down by summing over a multi-pixel sampling aperture. The S/N ratio is consistently doubled by the application of a circular convolution aperture with a 3 or 4 pixel diameter, and a gaussian aperture with a width of 2 or 3 pixels even slightly surpasses this performance.

We use the following definition of a circular aperture c_w and a gaussian aperture g_w with the aperture scale w:

$$
\begin{aligned}
c_w(x, y) &= \left\{ \begin{array}{ll} 1, & r(x, y) \leq w/2; \\ 0, & \text{otherwise} \end{array} \right. , \\
g_w(x, y) &= \exp\left(\frac{1}{2} \frac{r^2(x, y)}{(w/2)^2} \right),
\end{aligned}
$$

where $r(x, y)$ is the distance of the center of pixel (x, y) from the aperture center, and x, y are integer pixel coordinates. The aperture center lies in the center of a pixel for odd w and on a vertex between four pixels for even w. A selection of apertures is displayed in Figure 5.28.

The performance curves closely resemble each other at different angular separations, suggesting that the above-mentioned bias of using a single realization of aberration noise is minor.

In the light of this evaluation, the gaussian aperture of width 2 appears to be the best overall choice. It has therefore been used in all performance simulations shown in this chapter.

5.7 Target List

As discussed in Sections 5.5.2, the S/N ratio for the detection of an exoplanet at a given angular separation in the ideal photon-limited case plummets with d^{-3} as the distance d to the target star increases, due to the deepening planet/star contrast and the overall reduced photon statistics. Only the nearest stars therefore make good candidates for planet detection.

Furthermore, most stars in our 5 pc neighborhood belong to the faint, red end of the main sequence, providing bad photon statistics. As mentioned in Section 5.5.10, the photon limit can be lowered by extending the observation time on the target; however, the time investment quickly becomes prohibitive if more than a factor of 2 in S/N needs to be gained. Therefore, a star need also be reasonably bright to allow direct planet detection.

Table 5.1 provides a list of target stars conforming to these criteria. For each star, the S/N ratio for the detection of a 25% polarized Jupiter-sized planet under SC3 conditions at the angular separation $\theta = 0\rlap{.}''2$ is given as a representative of the AO-controlled "hot zone" at $0\rlap{.}''1 \lesssim \theta \lesssim 0\rlap{.}''3$. The value can also be taken as a good estimate of the S/N of a 50% polarized planet at $\theta = 0\rlap{.}''3$, or a 10% polarized planet at $\theta = 0\rlap{.}''1$.

Table 5.1: The most favorable target stars for SPHERE/ZIMPOL in terms of detectivity. For each star, the absolute V-band magnitude and distance in parsecs is given, as well as the S/N ratio for the detection of a 25% polarized Jupiter-sized planet at two angular separations ($0\rlap{.}''03$, $0\rlap{.}''2$) and two different noise level assumptions (active angular averaging over 23 field orientations, photon limit), and the frame exposure time t_{exp} needed to reach 80% of full well capacity. The $5\,\lambda/D$ classical Lyot coronagraph is assumed for the $\theta = 0\rlap{.}''2$ case, and the monochromatic 4QPM coronagraph for $\theta = 0\rlap{.}''03$. Otherwise, SC3 conditions are used. Integration time is 4 hours.

| | absolute | | S/N ratio | | | | frame |
| | magnitude | distance | act. ang. avg. | | photon limit | | exp. time |
star	M_{V}	d (pc)	$0\rlap{.}''03$	$0\rlap{.}''2$	$0\rlap{.}''03$	$0\rlap{.}''2$	t_{exp}(s)
G0V at 3 pc	4.24	3.00	12.2	5.56	159	7.78	5.1
M0V at 3 pc	8.62	3.00	11.4	1.54	9.66	1.13	293
α Cen A	4.36	1.34	83.6	39.9	1475	100	1.2
α Cen B	5.72	1.34	56.7	31.8	822	52.2	4.0
Sirius A	1.40	2.67	20.8	9.44	486	48.8	0.3
Procyon	2.68	3.50	12.6	5.45	163	12.2	1.7
Altair	2.20	5.14	4.88	2.21	46.1	4.85	2.3
ε Eri	6.18	3.22	13.4	2.99	34.9	3.08	36
τ Cet	5.68	3.65	7.44	2.57	27.4	2.57	29
ε Ind	6.89	3.63	8.53	1.59	24.9	1.49	87
AX Mic	8.71	3.95	5.65	0.61	12.4	0.47	551
82 Eri	5.35	6.06	2.92	0.70	12.0	0.67	59
δ Pav	4.62	6.11	2.92	0.87	10.6	0.88	31
$o2$ Eri	5.92	5.04	4.99	0.97	9.54	0.95	68
70 Oph A	5.67	5.09	4.37	1.09	9.58	0.92	305
Barnard's	13.2	1.82	6.04	0.38	9.47	0.72	7574
Proxima	15.5	1.29	5.00	0.30	8.39	0.67	29159
Lacaille 9352	9.76	3.29	5.96	0.55	8.09	0.62	1009
GJ 783	6.41	6.05	2.00	0.45	7.96	0.43	155
36 Oph A	6.41	6.10	2.75	0.35	7.57	0.41	151
ξ Boo A	5.57	6.70	2.32	0.47	7.50	0.46	88
70 Oph B	7.51	5.11	2.66	0.42	6.72	0.44	305
Sirius B	11.3	2.64	4.92	0.39	5.25	0.52	2759
36 Oph B	6.45	5.87	4.39	0.52	4.76	0.40	151
GJ 570 A	6.86	5.90	2.92	0.41	4.19	0.43	225
ξ Boo B	7.84	6.70	1.62	0.13	2.54	0.16	711
Formalhaut	2.00	6.76	2.56	1.35	35.5	2.33	3.3
Arcturus	-0.30	11.3	1.25	0.53	16.5	1.28	1.1
Pollux	1.00	10.8	1.08	0.50	15.3	0.89	3.4
β Hyi	3.45	7.47	2.81	0.76	8.32	0.83	16
δ Eri	3.74	9.04	1.31	0.40	8.04	0.45	22
1 Ori	3.67	8.02	1.83	0.51	6.44	0.60	30
Canopus	-3.10	29.9	0.15	0.08	3.47	0.30	0.6

The use of a 4QPM coronagraph extends the detection domain even further inwards, down to $\theta \approx 0.''01$, where the planet/star contrasts are the most favorable. To represent this regime, the S/N ratio for a 25% polarized planet at $\theta = 0.''02$ is given in the table.

Since the exact effectivity of angular differential imaging cannot be pinpointed in advance, each S/N value is given both for the conservative case of angular averaging (with active field rotation over 23 positions) and for the ideal case of the photon limit. For faint stars such as Sirius B, photon noise dominates even in the angular-averaging case after the SC3 exposure time of 4 hours. For a particularly powerful star such as α Centauri, though, the photon limit lies so far below the angular-averaging noise level that it is unlikely to be reached through angular differential imaging. In such special cases, it will be beneficial to sacrifice photons to improve the image quality in other ways, such as observing in a narrow long-wave filter band or selecting a subset of all individual exposures with particularly good instantaneous seeing conditions (frame selection). In general, a nearby star is a more reliable candidate than a brighter but more remote star.

If the orientation of the system's orbital plane is not known, the full linear Stokes vector (IQU) must be measured, doubling the observation time required to reach a given photon noise level.

Among the main-sequence stars in the solar neighborhood, 7 candidates offer a photon-limited S/N ratio of at least 2.5 for the 25% polarized sample planet at $\theta = 0.''2$, which could be improved to a 5σ detection in 16 hours of observation. On the other hand, a sample planet at $\theta = 0.''03$ observed with the 4QPM coronagraph fulfills this criterion around at least 24 nearby stars. Furthermore, the nearest few evolved stars such as Formalhaut are bright enough to achieve attractively low photon noise levels for the close-in sample planet. However, it will be difficult to reach down to the photon limit in these cases.

Discussion of individual targets

An explicit performance plot for each of these target stars can be found in Appendix C. The following paragraphs briefly discuss the specific situation for each target.

A polarization of 25% is assumed for a Jupiter-sized planet. The polarization signature of a Neptune-sized planet is presumed to be 8 times weaker than that of the standard Jovian, mostly due to the smaller radius (the polarization is higher due to the absence of clouds, but the albedo in the RI bands lower due to the blue color). For an Earth-like planet, the S/N is at least 300 times lower than for the Jovian. A 5σ confidence level is required for detection. If the location of the planet is known a priori, a 3σ event is already significant for confirmation.

Unless mentioned otherwise, the integration time is taken to be 4 hours.

α Centauri A and B: These are by far the best targets for ZIMPOL/SPHERE for planet search.

- The standard Jupiter-sized planet is detectable in the double-difference image inwards of $\theta \approx 0\rlap{.}''15$. With angular averaging, the 5σ boundary lies at $\theta \approx 0\rlap{.}''5$. In the photon-limited case, even the area outside the AO control radius is accessible, and Jovians can be observed out to at least $\theta = 1''$. With the 4QPM, close-in planets can achieve S/N ratios of ~ 15 in the double difference, ~ 300 with angular avearging, and ~ 1000 in the photon limit.

- Smaller gas planets of the size of Uranus and Neptune should be detectable inwards of $\theta \approx 0\rlap{.}''25$ with angular averaging, and out to $\theta \approx 0\rlap{.}''4$ at the photon limit.

- At the photon limit, even a terrestrial planet might reach a few σ of detectivity for α Cen A and B in the innermost $0\rlap{.}''05$. Of course, this terrestrial planet must have favorable properties for detection, in particular a thick Rayleigh scattering atmosphere. An only slightly larger "super-Earth" or "water giant" planet could likely break the 5σ threshold.

- Due to the high intensity of this star, the noise level given by aberrations lies far above the photon limit. Thus, reducing the noise level down to the photon limit is more challenging than at fainter stars. On the other hand, the performances given for the double-difference can already be achieved with a fraction of the SC3 exposure time of $4\,\mathrm{h}$, e.g. 20 minutes. Alternately, narrow filter bands or frame selection can be employed to improve the Strehl ratio.

Sirius A: While the planet/star contrasts are less favorable at Sirius than at closer stars, it is exceptionally bright, providing a low photon noise level. Reduction of the systematic noise down to the photon limit is challenging but particularly beneficial here.

- Jupiter-sized planed can be detected in the AO-controlled area of $\theta \lesssim 0\rlap{.}''3$ with angular averaging, and out to $\theta \approx 0\rlap{.}''8$ at the photon limit. S/N ratios reach ~ 50 in the former and ~ 500 in the latter case.

- Neptune-sized planets become detectable at $\theta \approx 0\rlap{.}''15$ with angular averaging, and at $\theta \approx 0\rlap{.}''3$ at the photon limit. The S/N ratios peak at ~ 6 and ~ 60, respectively.

- The high flux values make Sirius A a prime candidate for Strehl-improvement methods as discussed for α Cen.

Procyon and Altair: Procyon is a close approximation to the fictional SC3 standard star, while Altair is slightly less favorable due to its greater distance from the Sun. The detection threshold is $\theta \approx 0\rlap{.}''15$ for both stars with angular averaging, and $\theta \approx 0\rlap{.}''3$ and $0\rlap{.}''2$ in the photon limit, respectively. Since these stars are significantly less bright than the above-mentioned ones, a photon-limited performance is easier to achieve.

ε **Eridani:** This star is already faint enough that the performance of angular averaging coincides with the photon limit for all but the innermost $0.''02$. This opens up the option of increasing the performance by extending observation time, e.g. to 10–20 h. Since ε Eri is proven to harbor at least one planet and the orientation of the system plane is known, an investment of time is warranted.

The standard Jovian is detectable inwards of $\theta \approx 0.''15$, reaching a S/N of ~ 35 (more in the photon limit), whereas Neptunians are barely visible under ideal conditions in the innermost $0.''05$. The mean orbital separation of the known planet is around $1''$ and therefore outside of the detection zone; however, its excentricity is as high as 0.7, bringing it much closer to the star during periastron.

τ **Ceti and ε Indi:** These stars mark the limit of the planetfinding capabilities of the $5\,\lambda/D$ Lyot coronagraph. Only at the very edge of the Lyot mask, $\theta = 0.''1$, does the standard Jovian reach the detection threshold. With the 4QPM, however, S/N ratios of up to ~ 50 in the photon limit are achievable within this radius. Again, extending the observation time would be improve these limits.

5.8 Examples of Secondary Science

Although this work is focused on the use of SPHERE/ZIMPOL to detect extrasolar planets, the results can also be used to estimate the instrument's treatment of other science targets. Two sample applications are presented below.

5.8.1 Circumstellar Disks

The polarized flux from protoplanetary disks or debris disks is many orders of magnitude stronger than that of an extrasolar planet, and therefore less challenging to detect. Extending to over 100 AU from their star, they can be potentially be imaged at much larger distances than planets can. Albeit less prestigious than exoplanets themselves, these objects nevertheless harbor a wealth of information about planet formation and migration, and might even betray the presence of planets through gaps and swirls in their surfaces.

As a sample case, we adopt the debris disk of the star HD 169142 (SAO 186777, IRAS 18213–2948) described in Kuhn et al. (2001). It appears approximately face-on, with an angular diameter of some $3''$ despite its distance of 145 pc. The total polarized flux is $7.5 \cdot 10^{-3}$ of the total stellar flux.

The radial intensity distribution is not explicitly given, so we adopt the profile of the similar disk around TW Hya. According to Apai et al. (2004), the surface brightness drops with $\sim \theta^{-1}$ in the main part of the disk, and with $\sim \theta^{-3}$ further out. The power-law shift occurs around $\theta = 0.''8$.

Figure 5.29 presents the results of a simulated observation of the model disk after 1 min, 10 min and 1 h of exposure time. Note that the field of view ($3'' \times 3''$) is larger
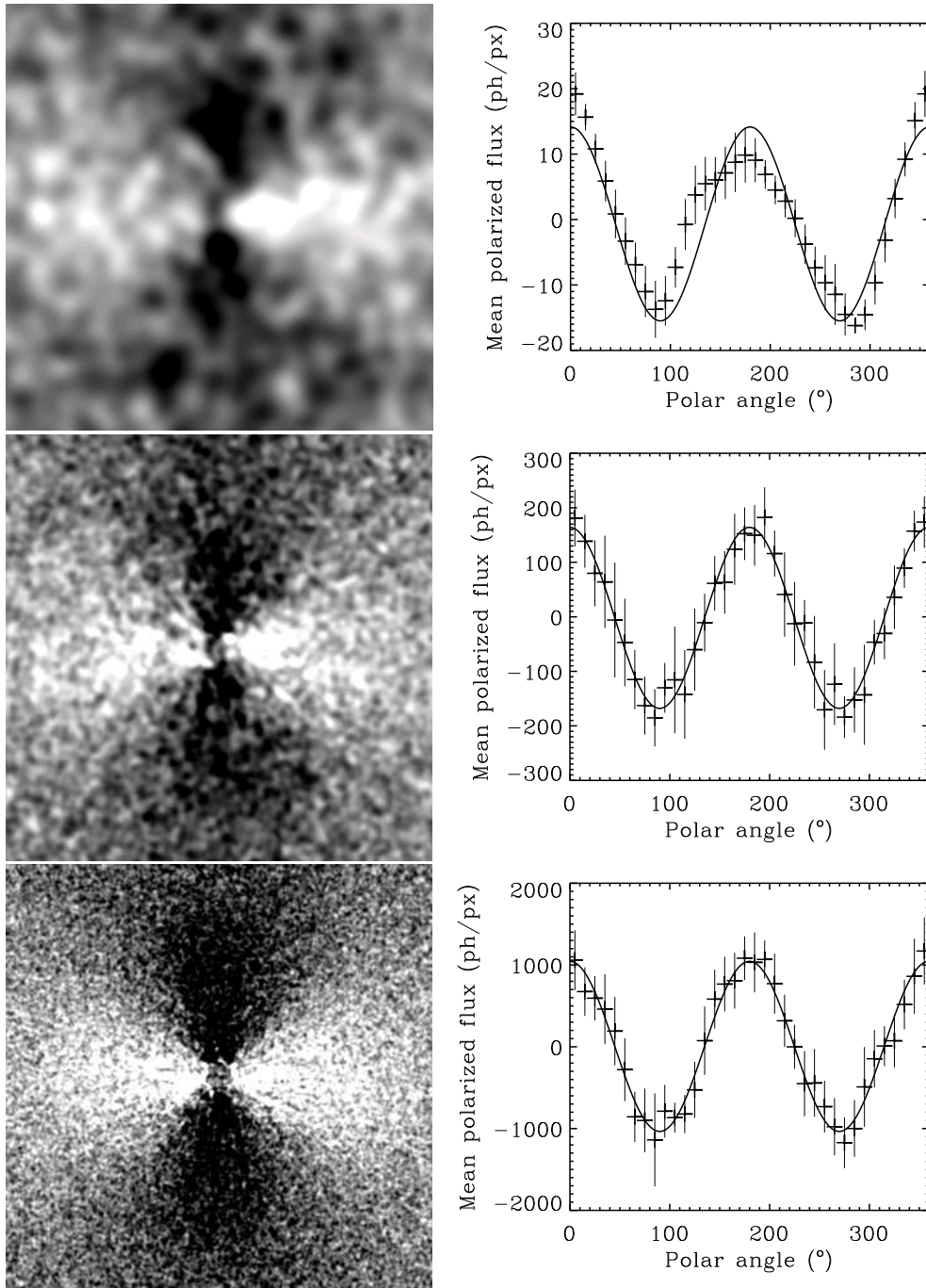
Fig. 5.29: Left: Simulated imaging polarimetry of the circumstellar disk of HD 169142 with a $5\,\lambda/D$ Lyot coronagraph, 100 turbulent phase screens and 1 simulated wavelength. The exposure times are 1 min, 10 min and 1 h. Gaussian sampling apertures of 15, 5 and 2 pixels width have been applied, respectively. The last image includes angular averaging over 7 active positions. The field scale is $3'' \times 3''$. **Right:** Mean polarization in the annulus between 6 and $20\,\lambda/D$ as a function of the polar angle, extracted from the images to the left. The curve is a double-frequency cosine fit to the data points.
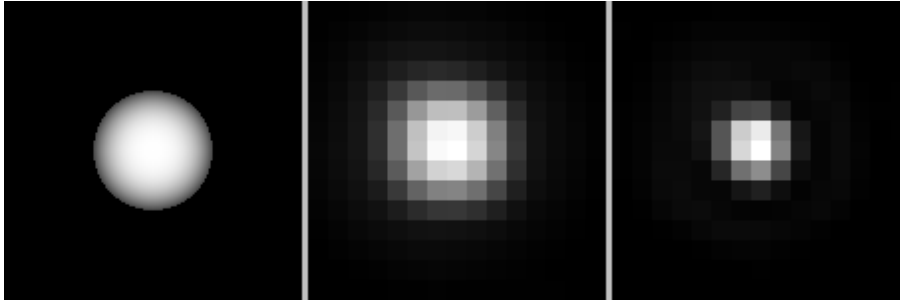
Fig. 5.30: Left: The assumed appearance of Betelgeuze. **Center:** A simulated image of Betelgeuze on the ZIMPOL camera. **Right:** The PSF core for comparison.

than in the previously presented images. The characteristic "butterfly" pattern of an extended object with tangential linear polarization is apparent already after one minute of observation. The 1 h image includes angular averaging over 7 positions of active field rotation in order to reduce the non-photonic noise.

Both passive and active field rotation may be harnessed to improve the image quality of an extended target. While imaging the disk as a whole, they can only be used for simple angular averaging. Angular differential imaging or unsharp masking would erase the large-scale structure, since these techniques exploit the strong spatial localization of a planet signal. However, they may be used to locate small-scale structures within the disk.

Conversely, the extended nature of a disk allows aperture convolution to be performed with much larger apertures than the 2×2 pixel gaussians used for exoplanets, facilitating the detection of the presence of a disk even against an overwhelming level of photon noise.

The bottom panel of Figure 5.29 demonstrates that variations in the polarized surface brightness of the disk could be discerned at quite small spatial scales. The SPHERE/ZIMPOL instrument can potentially detect clumps, swirls and gaps in such a disk, pointing to the presence of planets and their impact on the disk morphology. In combination with the 4QPM coronagraph, inner working angles of a few tens of milliarcseconds are available, greatly outperforming the present-day instruments used for such purposes.

5.8.2 Giant Stars

With an angular resolution in the range of 15–20 mas, the ZIMPOL camera of SPHERE is capable of resolving the disk of giant stars. Figure 5.30 demonstrates that α Orionis (Betelgeuze), having an angular diameter of ~ 45 mas, extends across about 3×3 resolution elements. Therefore, surface pulsations and starspots could be observed in direct imaging.

The exact radius and limb darkening of the star is not known; the image presented in the figure is an assumption made for the sake of illustration.
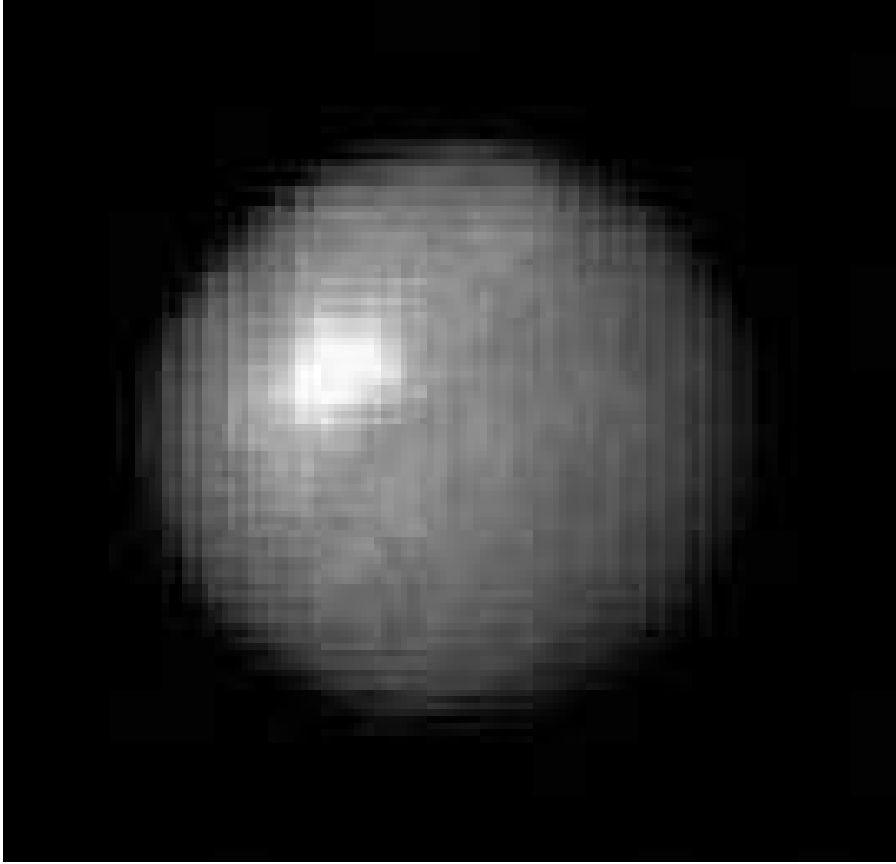
Fig. 5.31: The asteroid Ceres as observed with the Hubble Space Telescope. SPHERE will offer an angular resolution more than three times higher.

5.8.3 Asteroids

Similarly, ZIMPOL will allow SPHERE to image objects in our own solar system with image quality, asteroids in particular. The resolution offered by SPHERE will exceed that of the Hubble Space Telescope (HST) by a factor of more than three. This will allow investigations of surface structures and properties, as well as the detection of tiny satellites orbiting asteroids.

For comparison, Figure 5.31 provides an HST image of Ceres, the largest asteroid in the solar system.

Chapter 6

Conclusion and Outlook

6.1   Simulation

In this work, we have presented laboratory tests, detailed simulations and evaluations of the detection of extrasolar planets with SPHERE/ZIMPOL. Many effects have been taken into account as far as they can be predicted. These include the Earth's atmosphere, the static and slowly evolving (quasi-static) aberrations of the telescope and the common path optics, the performance of the adaptive optics system and the coronagraph, the static and differential aberrations of ZIMPOL, and the effect of software image processing.

The faint polarization signal of an exoplanet must be extracted from the overwhelming speckle halo of its parent star's coronagraphic PSF. This requires a polarimetric sensitivity of 10^{-5}, which has been demonstrated in laboratory experiments to be feasible with a ZIMPOL system.

However, a major obstacle to such a detection is the landscape of residual speckle noise introduced into the polarization image by aberrations in the instrument. These spurious patterns can be further reduced by signal switching using the half-wave plate HWP2. Rotating the half-wave plate by 45° inverts the true polarization signals originating on the sky, while instrumental effects (e.g. due to the differential aberrations from the polarization modulator) remain largely unchanged. Subtracting the images before and after switching will theoretically cancel out all static noise pattern—in practice, they will merely be strongly diminished—whereas any true signals will remain unharmed. The remaining noise level after this step will be due to temporal drifts and changes on the time scale of the signal switching that limit the effectivity of the background subtraction, such as time-dependent instrument parameters (in particular those pertaining to the AO performance) which are hard to quantify in advance, and effects caused by the rotating HWP2 itself. Finally, angular differential imaging provides an additional means to average down or even eliminate the remaining background structures.

The evaluations presented in Section 5.5 illustrate the dependence of the final detection performance on a host of parameters, and help identify the most critical elements where a technical improvement might bring the greatest returns in detectivity. In particular, the spurious polarization background caused by a combination of upstream static aberrations, differential aberrations from the modulator, and temporal aberrations from the imperfect surface of the rotating HWP2 have been found

to limit the overall performance with the greatest leverage. This insight has been fed back into the SPHERE optical design process; in particular, efforts are being made to minimize the differential wedge of the modulator and the surface roughness of HWP2.

While signal-switching and angular differential imaging are both proven concepts, it is difficult to foretell their exact effect on the performance of SPHERE/ZIMPOL. The presented models are educated guesses to some degree; it must be expected that a real instrument does not behave as predicted. We believe our assumptions not to be overly optimistic and that a careful monitoring of the instrument and analysis of data will allow the calibration of many instrumental effects, and reduce the systematic noise. In particular, our simulations do not exploit the property of ZIMPOL that each of its two cameras record the full polarization image at any given time—this redundancy could be harnessed as an additional tool to discern real from spurious signals.

6.2   Performance

The primary goal of detecting extrasolar planets with SPHERE/ZIMPOL is found to be feasible.

With the $5\,\lambda/D$ classical Lyot coronagraph, Jupiter-sized planets can principally be detected within the AO-controlled inner field (angular separation $\theta \lesssim 0''\!.3$) for the eight most favorable target stars. For the brightest three (α Cen A and B, Sirius), Neptune-sized planets or even "super-Earths" could be discerned, while Jupiter-sized planets are detectable even outside the AO control radius. For more remote stars, the planet/star contrast deepens and the photon noise increases, dropping the chances of detection dramatically.

A four-quadrant phase mask coronagraph, on the other hand, allows the observation of companions as close as $1\,\lambda/D$ to the star, where the brightness contrast is at its most favorable. Simulations suggest that Jupiter-sized planets in close orbits ($\theta < 0''\!.1$) could be detected around at least 20 nearby stars.

A more detailed account of the target stars and their expected planet detectivity is provided in Section 5.7.

6.3   Improvement

Based on the insight provided by this work, we see the following ways to optimize the performance of SPHERE/ZIMPOL.

Hardware: Reducing the differential aberrations of the polarization modulator directly translates into better performance in aberration-dominated observations. While the manufacturer does not foresee a method of suppressing these effects during production, one can select the best specimen from a batch of products. Aberrations introduced by other optical components (HWP2 in particular) should also be kept as low as feasible.

Data reduction: Frame selection can be employed to reduce the systematic effects. This procedure may require more observing time (e.g. a factor 2) but may help to push the systematic noise limit further down, particularly in the case of bright target stars with good photon statistics.

Observing strategy: It might be beneficial to fit more field rotation positions or HWP2 switching cycles into a given observation time. This will improve the efficiency of the related differential noise reduction methods, as can be shown with simulations. Shorter periods would entail some additional (but not substantial) overhead time during the observations.

Calibration: One can attempt to calibrate the pattern in the polarization image introduced by the FLC modulator by using other science observations. This requires that high precision observations be taken in exactly the same instrument configuration.

Telescope time: Especially for the not-so-bright targets (like ε Eri, τ Cet, etc.), an increase of telescope time has a significant impact on the detection limit. The SC3 integration time of 4 hours is not a hard limit, it is well possible to add up the observations of several consecutive nights and push the total integration time to 10 or even 20 hours.

The detection of reflected light from planets is challenging due to the huge brightness contrast, as illustrated by the CAOS simulations presented here for Standard Case 3. However, with the supplemental methods listed above, the instrument can potentially be pushed substantially beyond the performance calculated by the rather conservative SC3 model simulations.

In fact, the improvement of the high-order temporal aberrations of HWP2 and of the FLC modulator's polarized differential aberrations by a factor of 2 over the SC3 assumptions is considered feasible enough to have been defined as the next Standard Case, SC4. Figure 6.1 shows the performance expected under these new conditions.

6.4 Future

SPHERE is currently well underway in the Final Design Phase and foreseen to achieve first light in early 2011. This should provide it with a head start of a few years over its space-based competitors such as ESA's Darwin mission and NASA's James Webb Space Telescope (JWST) and Terrestrial Planet Finder (TPF), which also aim for the direct detection of extrasolar planets.

While ESO's original intention of building a next-generation optical telescope with a 100 m mirror diameter (the OverWhelmingly Large telescope, OWL) proved too ambitious for the near future, a scaled-down successor project called the European Extremely Large Telescope (E-ELT; Gilmozzi & Spyromilio, 2007) featuring a mirror diameter of 42 m is now being pursued. The design study suggests that construction could be completed as early as 2017.
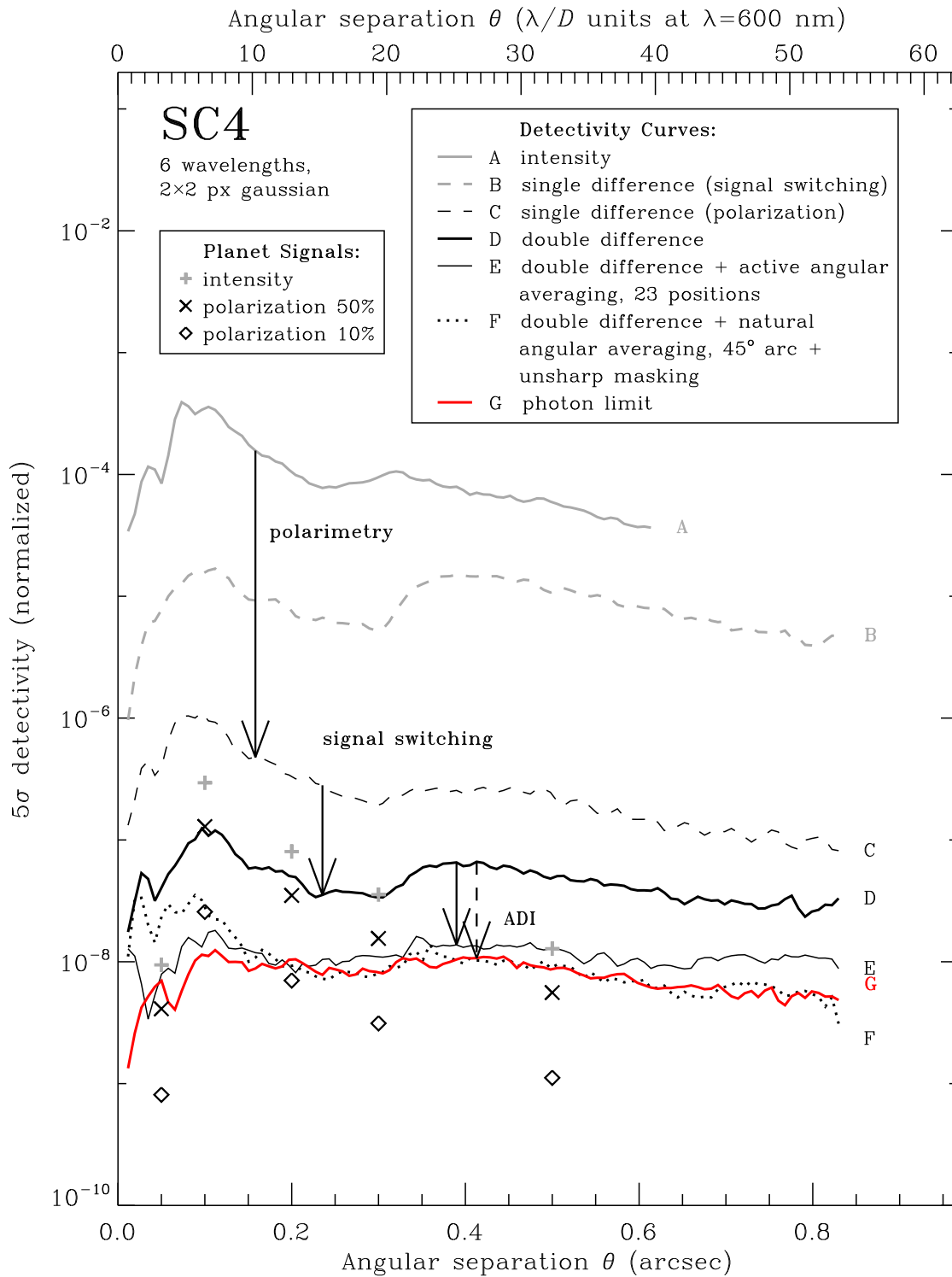
Fig. 6.1: Performance plot for the Standard Case 4 simulation.

As a part of this project, a planet finder instrument named EPICS (Exo-Planets Imaging Camera and Spectrograph) is being developed (Kasper et al., 2007). A big brother to the SPHERE instrument, it comprises an imaging polarimeter of the ZIMPOL type intended to continue the task of SPHERE/ZIMPOL on the much more powerful platform of E-ELT. The direct detection of Earth-like planets around nearby stars will be well within its capabilities.

Part II

The Solar D_1 Enigma

Chapter 7

Introduction

7.1 Solar Spectropolarimetry

Thanks to the discoveries of Isaac Newton and Joseph von Fraunhofer, it has been known since the Age of Enlightenment that the "white" light of the Sun and other stars can be unraveled into a spectrum of its constituent colors by passing it through a glass prism. A wealth of information about the chemical and thermal make-up of the source is betrayed by the numerous dark lines and bands in the stellar spectrum, where absorption by a specific atomic or molecular species in the star's atmosphere has attenuated the light of a particular wavelength. Since the density, temperature and composition of the solar atmosphere strongly depend on the altitude, different spectral lines yield information about different outer layers of the sun. The opacity of a given layer varies with wavelength, allowing observations at different wavelengths to penetrate the sun's "surface" down to different depths, providing three-dimensional insight.

Yet there is more information to be gleaned from the solar spectrum. It is revealed when measuring the polarization along with the intensity as a function of wavelength.

7.1.1 Zeeman Effect

An atomic state of a given energy has several degenerate magnetic sublevels that differ in the orbital angular momentum along the quantization axis and are distinguished by the quantum number m. In the presence of a magnetic field, however, the atom's magnetic moment adds a term proportional to m to the energy level, thus splitting the sublevels and breaking the degeneracy. Transitions involving that atomic state then produce several distinct spectral lines. The selection rules for photon emission and absorption link these shifted lines to different polarization states, resulting in a characteristic antisymmetric signature in the circular polarization spectrum and a symmetric signature in linear polarization. These are known as the longitudinal and transversal Zeeman effect, respectively, named after Zeeman (1897).

Hale (1908) first observed this effect in the solar spectrum, allowing him to infer the strength and orientation of the magnetic fields threading the sun's surface. This technique was refined into the "magnetograph" (Babcock, 1953; Kiepenheuer,

1953), and has been employed in monitoring programs such as SOLIS (Keller et al., 2003) and MDI aboard the SOHO satellite (Scherrer et al., 1995, website[1]) that have continuously mapped the solar magnetic field over the last decades. The full 3D magnetic field vector can be inferred from a measurement of the full Stokes polarization vector (Solanki et al., 2003, and references therein).

A weakness of Zeeman diagnostics is the fact that it is limited by the telescope resolution. A magnetic field that changes its orientation on spatial scales smaller than a resolution element does not show up on a Zeeman magnetogram because contributions with opposite signs cancel each other out. Thus the method is typically insensitive to fields below $\sim 100\,\mathrm{G}$.

7.1.2 Scattering Polarization and Hanle Effect

Another source of polarization on the sun is coherent scattering. If an atom absorbs a photon and subsequently re-emits the energy as another photon before it is disturbed by a collision, an electric dipole oscillation of the atom is induced by the primary photon's electric field. This atomic antenna then radiates its energy away as a dipole wave, whose intensity is maximum in the plane perpendicular to the dipole and zero along the dipole axis. For the special case of 90° scattering, the outgoing light is therefore in the classical case 100% polarized perpendicular to the scattering plane. In the general quantum case, the amount of polarization depends on the polarizability of the transition, a property of its quantum-number structure.

If an atom is illuminated from all sides isotropically, the resulting scattered light is unpolarized. However, this condition does not apply to the solar atmosphere. Strong gradients in temperature and density near the Sun's surface introduce an anisotropy into the incoming radiation field, resulting in a bias for a specific linear polarization direction (Trujillo Bueno, 2001; Holzreuter et al., 2005). Scattering polarization is at its strongest at the limb of the solar disk, and vanishes towards the center. It is usually oriented parallel to the limb (Stokes $+Q$) but can also be perpendicular to it ($-Q$) under special conditions.

This basic polarization can further be modified, in particular by the presence of magnetic fields. An atomic dipole excited by an incoming photon will precess around the magnetic field lines, altering the direction of polarization of the outgoing photon. These phenomena are called the Hanle effect. The expected linear polarization is thereby damped (by turbulent fields) or partially rotated (by directed fields). They can be exploited to measure solar magnetic fields ranging from a few milli-Gauss to about 100 Gauss. The presence of turbulent magnetic fields on spatial scales below the image resolution can be detected by Hanle depolarization, since the effect is insensitive to the polarity of the magnetic field and therefore cannot cancel itself out.

Spectral lines exhibiting scattering polarization were first identified around the middle of the last century (e.g. Brückner, 1963). A comprehensive atlas of the entire visible polarized spectrum was compiled by Stenflo et al. (1983a,b), revealing a rich landscape of scattering polarization.

---

[1] `sohowww.nascom.nasa.gov`

It was not until the 1990s that the true complexity and information density of the polarized solar spectrum were realized. The invention of the Zürich Imaging Polarimeter (ZIMPOL; Povel et al., 1990, 1994; Povel, 1995) allowed spectropolarimetry at an unprecedented sensitivity and precision, outperforming other contemporary instruments by 1–2 orders of magnitude. The polarization spectrum was found to exhibit a structure just as rich and complex as the intensity spectrum, bearing a wealth of new scientific information (Stenflo, 1996; Stenflo & Keller, 1997). For this reason, it is commonly referred to as the "Second Solar Spectrum". A comprehensive atlas of the Second Solar Spectrum in the visible and ultraviolet range down to the atmospheric cut-off has been assembled and published by (Gandorfer, 2000, 2002, 2005).

An extensive discussion of scattering polarization on the sun can be found in the book of Stenflo (1994). A description of the ZIMPOL instrument is provided in Section 3.4 of this dissertation.

7.2 The Sodium D_1 Enigma

The detailed inspection of the Second Solar Spectrum brought to light many unexpected and puzzling features, spurning an extensive theoretical effort to explain the underlying physics. Most of these riddles have since been solved, considerably refining our understanding of scattering physics.

One particular feature, however, remains in stark contrast with the theoretical prediction and has so far defied all new explanation attempts. The problem was first identified and presented in Stenflo (1996). A pair of spectral lines formed by neutral sodium, Na I D_1 and D_2, exhibit the non-trivial behavior shown in Figure 7.1.

The large-scale linear polarization landscape connecting the two lines, including the change of sign at the midpoint, was explained in Stenflo et al. (1980) to be a result of intermediate-state interference between the D_1 and the D_2 transition—in other words, the state of the scattering atom between absorption and emission is a quantum superposition of the two excited atomic states that form the upper levels of the D_1 and the D_2 transitions.

Furthermore, Fluri et al. (2003) were able to reproduce the striking triplet peak structure of the D_2 line polarization by taking partial frequency redistribution (PRD) into account when solving the polarized radiative-transfer problem in the optically thick solar atmosphere. However, the same method is utterly incapable of explaining the similar central peak seen in D_1.

As presented e.g. in Stenflo (1994) and further elaborated in Stenflo (1997), the phase matrix of an atom in a coherent scattering process in the limit of weak magnetic fields can be described as the sum of two terms. The first term represents classical dipole scattering that produces linear polarization as discussed in the previous section, while the second term describes isotropic and therefore unpolarized scattering. The weights of the two terms in the sum are W_2 and $(1 - W_2)$, respectively. The intrinsic polarizability W_2 therefore describes to what degree a given spectral line is able to produce scattering polarization.
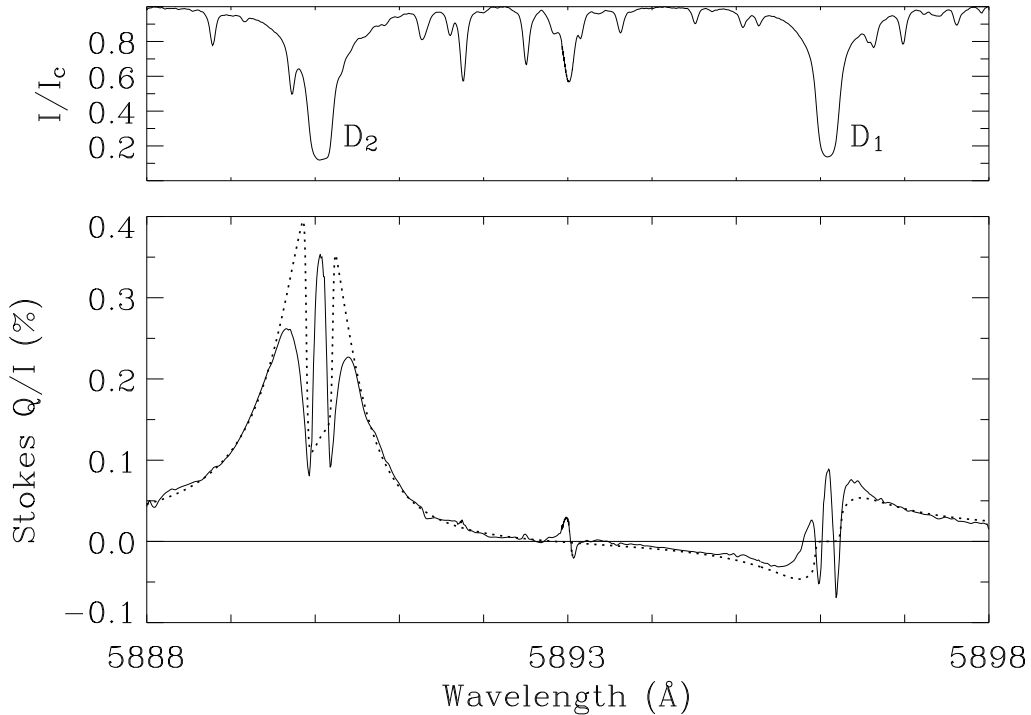
Fig. 7.1: Intensity (top panel) and fractional linear polarization (bottom panel, solid line) across the Na I D$_2$ 5889.97 and D$_1$ 5895.94 Å lines recorded with ZIMPOL at the National Solar Observatory (Kitt Peak) at 5″ inside the solar limb near the heliographic south pole (Stenflo et al., 2000b). The dotted curve in the lower panel represents a fit based on a simplified model including hyperfine structure as described in Section 11.4 of Feller (2007).

In the case of the Na I D$_1$ transition, a number of exact cancellations lead to a polarizability of $W_2 = 0$. In the total absence of dipole scattering, the line should therefore be incapable of producing polarization. Yet observations such as the one shown in Figure 7.1 have repeatedly demonstrated the presence of a distinct peak with even symmetry of $+Q$ polarization in the line center. The validity of this observation is corroborated by a similar peak found in the line center of Ba II D$_1$, which shares the exact same quantum structure including the hyperfine splitting. A detailed account of the underlying theory of this issue is presented in Chapter 11 of Feller (2007).

While many attempts at solving the Sodium D$_1$ Enigma through novel theoretical approaches have been undertaken (cf. Landi degl'Innocenti, 1998, 1999; Trujillo Bueno et al., 2002; Kerkeni & Bommier, 2002; Casini et al., 2002; Klement & Stenflo, 2003; Casini & Manso Sainz, 2005), none have succeeded thus far. Notably, Landi degl'Innocenti (1998) proposed that atomic polarization in D$_2$ could bleed off into D$_1$, but while his theory was indeed able to produce non-zero polarization in D$_1$, the effect was too faint by 1–2 orders of magnitude and antisymmetric in contrast to the symmetric signature found in observations (Stenflo et al., 2000a).

7.3 Laboratory Experiments

While this dichotomy between theory and observation clearly exists, it is not clear where the problem lies. The observations are reproducible and clean enough to rule out a persistent instrumental effect as the cause. Within the theory, however, there are several places where a puzzle piece could be missing. Is it a problem of atomic physics, an incomplete quantum-mechanical treatment of the scattering problem on the scale of single atoms and photons—or is it a problem of solar physics, an unknown or neglected aspect of the highly complex and turbulent plasma in which the scattering events take place?

In order to track down the underlying cause of the D_1 enigma to its origin, we performed a series of laboratory experiments. The idea was to excise the atomic scattering problem from the three-dimensional, complex environment of the solar atmosphere dominated by collisions, radiative transfer and its tangled fractal-like magnetic field, and to investigate it under clean, quantifiable laboratory conditions and geometries.

The first generation of this experiment is described in detail in Feller (2007). A glass cell filled with argon buffer gas and an optically thin amount of hot sodium vapor was illuminated with a beam of light from a low-pressure sodium vapor lamp. Light scattered off the atoms at a 90° angle was collimated through a Fabry-Pérot etalon and imaged by a ZIMPOL system. Since the very narrow transmission lines of the etalon depend on the angle of incidence, the wavelength registered on the detector decreases with increasing distance from the center, providing in effect a spectrum in the shape of concentric circles. The Na I D_1 and D_2 emission lines were visualized as distinct bright circles, whose line-integrated polarization could be measured. Furthermore, the light incident on the atoms could be fully polarized perpendicular or parallel to the scattering plane (Stokes $\pm Q$).

While this experiment was able to provide some constraints on the nature of scattering on sodium atoms, it could not conclusively narrow down the open question of the D_1 enigma. The main obstacles were low photon statistics and low spectral resolution.

In order to bypass these problems, a second-generation experiment was launched, investigating the D_1 and D_2 lines of K I rather than Na I. This allowed the use of a tunable solid-state laser as the light source, providing ample photon statistics and obviating the need for a spectrograph, since the extremely narrow band of the laser could be scanned across the spectral line to obtain a line profile. The K I D_1 transition features the exact same quantum structure as the enigmatic Na I and Ba II lines, and should therefore exhibit the same scattering behavior. Furthermore, a Helmholtz coil was added to the vapor cell to allow the introduction of a magnetic field at the scattering location.

The following two chapters present the laboratory setup, the data acquisition and reduction methods, and the results obtained during the last three years. Chapter 8 represents the state of late 2005 and concentrates on the technical and computational parts of the experiment. Chapter 9 focuses on the more recently obtained results, in particular on the dependence of the scattering polarization on the magnetic field.

Chapter 8

Laboratory Experiment for Polarized Scattering at Potassium Vapor[*]

C. Thalmann[1], J.O. Stenflo[1], A. Feller[1], and A. Cacciani[2]

Abstract

The observed solar polarization peaks of the D_1 lines of sodium at 5896 Å and barium at 4934 Å still elude theoretical explanation, in spite of considerable theoretical efforts over the past decade. To clarify the underlying physics of D_1 scattering we have set up a laboratory experiment to explore the spectrally resolved profiles of the Mueller matrix for D_1 and D_2 scattering at potassium vapor. Here we present initial results of this experiment.

8.1 Introduction

The pronounced and symmetric polarization signatures that have been repeatedly observed in the Doppler cores of the Ba II D_1 line at 4934 Å and the Na I D_1 line at 5896 Å (Stenflo, 1996; Stenflo et al., 2000a; Stenflo et al., 2000b) appear to contradict predictions from quantum mechanics and have remained enigmatic for many years, in spite of numerous theoretical attempts to explain the observations (cf. Landi degl'Innocenti, 1998, 1999; Trujillo Bueno et al., 2002; Kerkeni & Bommier, 2002; Casini et al., 2002; Klement & Stenflo, 2003; Casini & Manso Sainz, 2005). The theory fails by 1–2 orders of magnitude in the polarization and predicts the wrong symmetry for the core profile. To answer the question whether this problem is one of solar physics or one of quantum/atomic physics, we have set up a laboratory experiment to study the properties of polarized scattering in the potassium transitions K I D_2 7664.9 and K I D_1 7699.0 Å. These transitions have the same quantum structure as the Ba II and Na I lines, including hyperfine structure splitting, and have the decisive advantage that a solid state tunable laser can be used as a light source in this spectral range. The setup consists of the laser, polarizing beam splitter, potassium vapor cell, polarization analyser based on piezoelastic

---

[1] Institute of Astronomy, ETH Zurich, 8092 Zurich
[2] University of "La Sapienza", Piazzale Aldo Moro 2, I–00185 Rome

modulation, transfer lenses, and photomultiplier with lock-in amplifier. As the laser allows monochromatic excitation of the potassium atoms, the experiment gives us the polarization for 90° scattering as a function of position within the line profile.

8.2 Experiment Setup

The present setup represents a successful second-generation version of our D_1 scattering experiment. In the first and unsuccessful generation of this experiment we attempted to mimic solar conditions by irradiating sodium gas with either polarized or unpolarized radiation from a moderately broad-band light source (a low-pressure sodium lamp), while separating the D_1 and D_2 lines with a narrow-band filter. This experiment failed for a number of reasons: insufficiently clean separation of D_1 and D_2, insufficient S/N ratio, and inacceptably high stray-light levels. In addition no information on the profile shape was obtained. We then realized that there is absolutely no reason to try to reproduce the Sun, since what we are after is to test the underlying quantum-mechanical scattering theory under the most optimum conditions that best constrain this theory, and the best setup for doing this has little if any resemblance to solar conditions.

All the problems of the previous experiment are completely eliminated if a tunable laser is used. It not only provides fully resolved polarized line profiles with superb S/N ratio and no spurious contaminations, but also brings other information, like on the optical pumping properties. Since solid-state tunable lasers are not available for the sodium wavelengths, we choose to do our experiment with potassium gas instead.

We thus irradiate a sample of potassium gas with a tunable narrow-band laser and measure the polarization of the light scattered at 90°. By tuning the laser wavelength across the spectral line, we can fully resolve the line profile without having to use a spectrograph (which would not have worked because the S/N ratio would have been reduced by orders of magnitude). Figure 8.1 provides a schematic overview of the laboratory setup.

The heart of the experiment is a cross-shaped glass cell containing potassium vapor, built for this purpose by A. Cacciani in Rome. Cells based on the Cacciani design have been used for decades in solar astronomy (cf. Cacciani & Fofi, 1978; Cacciani, 1981; Cacciani et al., 1997). Argon buffer gas in the cell prevents the potassium from condensing onto the entry and exit windows. We use a vapor temperature of 100° C.

The source is a 15 mW solid-state laser tunable across 2.5 Å, of which we use about 200 mÅ. The spectral width is less than 2 mÅ. We expand the beam and pass it through a set of polarizers, pumping the K vapor with a selection of six Stokes states: $100\% \pm Q/I$, $\pm U/I$, $\pm V/I$. We use two dedicated lasers, centered around the D_1 and D_2 lines, respectively.

In a second optical arm, light scattered at 90° passes through a photo-elastic modulator and a polarizer, which convert the polarization information of one Stokes component (Q, U, or V) into an intensity modulation (at 84 kHz for Q and U,
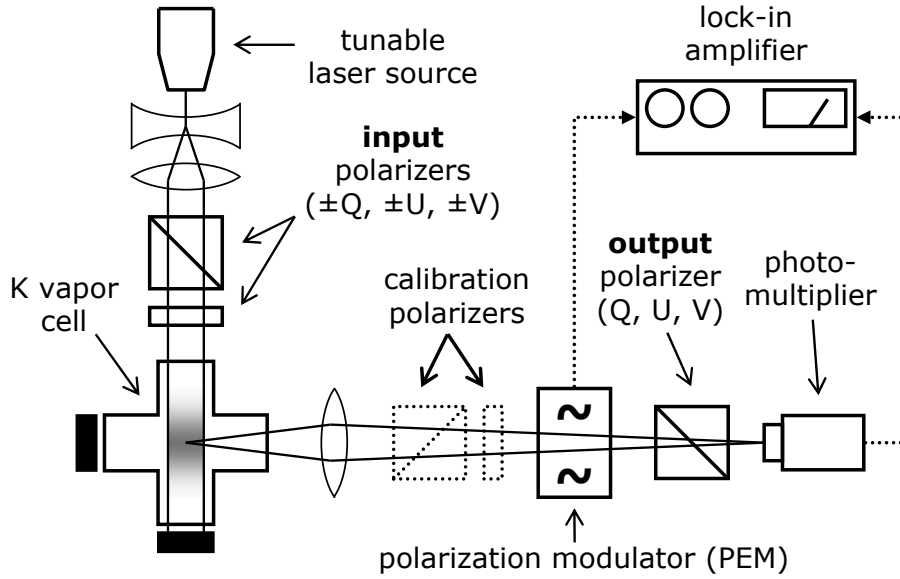
Fig. 8.1: Overview of the experimental setup.

at 42 kHz for V). A photomultiplier detects the signal and a lock-in amplifier demodulates it. Combined with the 6 input polarization states and the 3 output states, a full measurement set consists of 18 line profiles.

8.3 Data Acquisition and Reduction

8.3.1 Procedure to Obtain the Empirical Mueller Matrix

For each of the 6×3 combinations of input and output polarization states, we tune the laser through the spectral line and correct the recordings for dark current and calibration. Since absolute intensity cannot be measured reliably with our setup, we normalize all (wavelength-dependent) Stokes vectors by the maximum of the intensity curve. Furthermore, we subtract a constant polarization offset from each polarized flux curve to set the polarization level outside the line profile to zero, since Q, U, V must go to zero outside the line as I goes to zero there.

From this set of polarization profiles, the elements of the phase matrix P_{ij}, $i, j = 1, \ldots, 4$ can be extracted (renormalized to the maximum of P_{11}).

Figure 8.2 shows how the matrix elements of rows 2, 3, and 4 are determined as illustrated for the example of the $Q \to Q$ measurements. The element P_{11} simply describes the shape of the total intensity profile. Since the input Stokes vectors have the shape $I(1, \pm 1, 0, 0)^T$ (where T indicates the transpose operation), the measured Q profiles represent $P_{12} \pm P_{22}$, respectively. Thus, P_{12} is simply half the sum and P_{22} half the difference of the two $\pm Q \to Q$ measurements. By applying this method to all 18 combinations of input and output polarizations, we get 18 independent measurements of P_{11}, 3 for each P_{1j}, and one for each of P_{2j}, P_{3j}, P_{4j}, where $j = 2, 3, 4$.
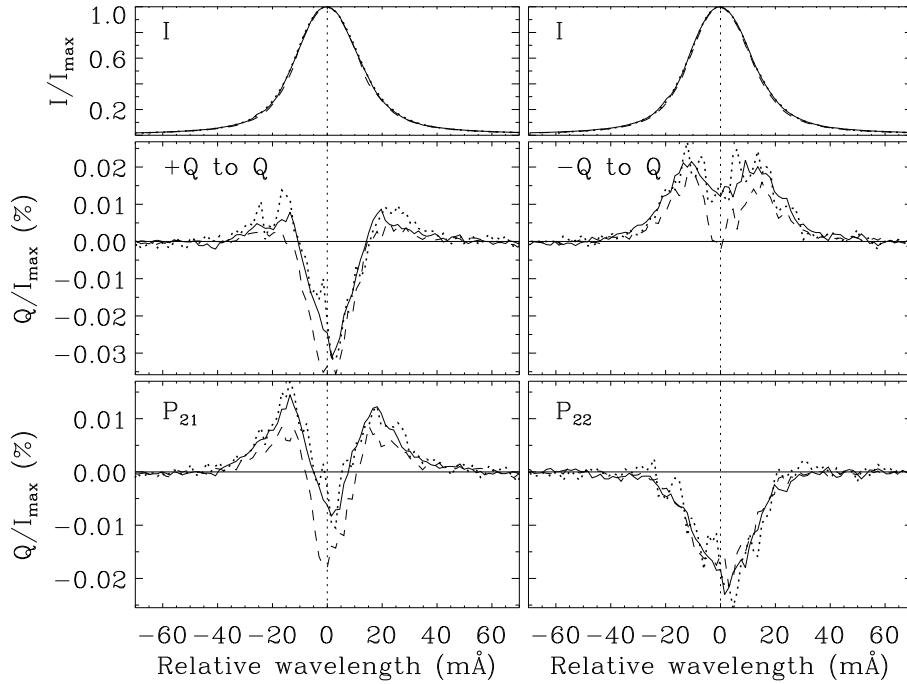
Fig. 8.2: Stokes $Q \rightarrow Q$ measurements and Mueller matrix elements for the K I D_1 line. The solid and dashed profiles represent measurements with the full laser power, the dotted lines with only a tenth of the full power. For the dashed profiles an artificial instrumental polarization was introduced by tilting an optical component. **Top:** Total intensity profile, i.e. P_{11}. **Center:** Stokes Q profiles for input states $100\% +Q/I$ and $-Q/I$, respectively. **Bottom:** Phase matrix elements P_{21} and P_{22} as calculated from these two measurements. Note how the artificial instrumental polarization vanishes for P_{22} (but not for P_{21}), as expected.

Note that since P_{2j}, P_{3j}, P_{4j} are formed as the difference between two measurements, instrumental polarization effects subtract out completely. This is not the case for P_{1j}, since it is formed as the sum of two measurements (see Fig. 8.2).

8.3.2 Test Measurements

When the laser power is reduced, the results remain the same, except in the case of $+V \rightarrow V$, for which the signal is diminished. This nonlinearity indicates that optical pumping and atomic polarization are involved. Reducing the temperature of the K gas simply reduces the scattered intensity but not the amount of polarization.

For another series of $Q \rightarrow Q$ measurements, an artificial instrumental polarization was introduced by tilting an optical component. This produced an $I \rightarrow Q$ cross talk, but left matrix columns 2–4 untouched, since for them the offset subtracts out (see previous section). P_{1j} may therefore contain spurious fractions of the P_{11} profile, while P_{ij}, $i, j = 2, 3, 4$, are robust and free from such effects.

Finally, an experiment was conducted with the PEM placed *before* rather than after the scattering cell, i.e., the gas was pumped with polarization modulated at 84 kHz. In this case the $Q \rightarrow +Q$ profile exhibited a strong intensity dependence:
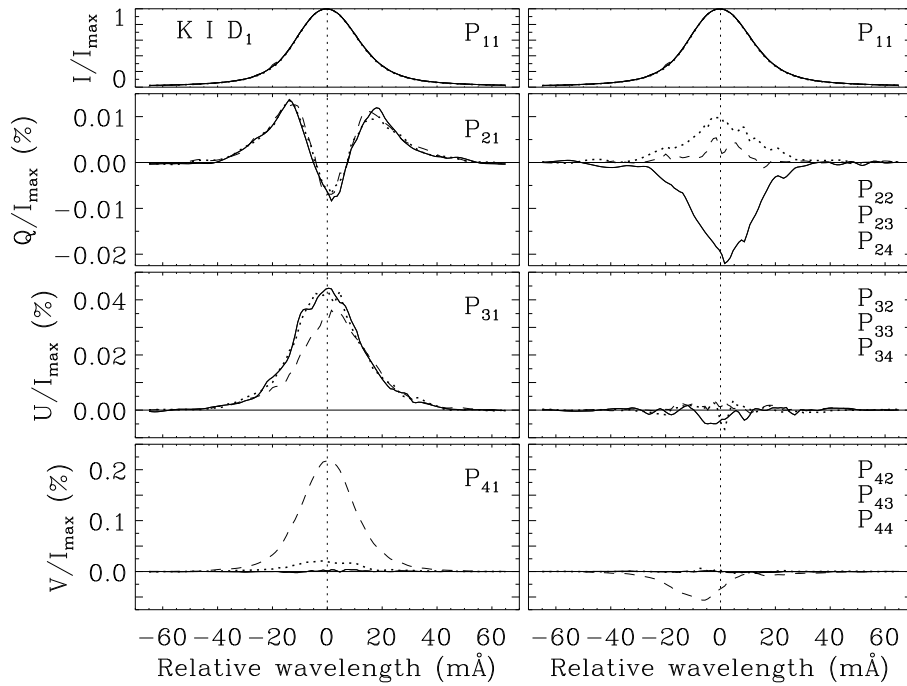
Fig. 8.3: Phase matrix elements for the K I D_1 line. In each frame, the solid curves come from measurements with input $100\% \pm Q/I$, the dotted ones from $100\% \pm U/I$, and the dashed ones from $100\% \pm V/I$. In the right panels these three types of curves represent P_{i2}, P_{i3}, and P_{i4}, respectively ($i = 2, 3, 4$).

For full laser power, a negative peak of $Q = -0.08\%$ was observed. For 1/2 and 1/5 power, the Q polarization was greatly reduced. For 1/10 of the full laser power, the signal vanished completely. This behavior can be understood if the optical pumping takes place on a time scale that is comparable to the modulation period, i.e. $\sim 10^{-5}$ s.

8.4   Results

Figures 8.3 and 8.4 show all the reconstructed phase matrix elements for the K I D_1 and D_2 lines, respectively. The lasers were run at full power, instrumental artefacts were minimized as far as possible, and no magnetic field beyond the Earth's field was applied. The noise level is of order 2×10^{-5}. A very conservative wavelet smoothing algorithm has been applied to the polarized profiles. Remember that the polarized matrix elements of the first column of the Mueller matrix (left panels in the figures) are likely to include some cross talk from the P_{11} profile, while those in the right panels do not.

Note how the independent measurements of P_{11} and P_{21} match each other perfectly, which confirms the consistency and repeatability of the measurements. In the case of P_{31} and P_{41}, it is mainly the curves with circular input polarization that deviate from the other two, apparently due to non-linear optical pumping effects.

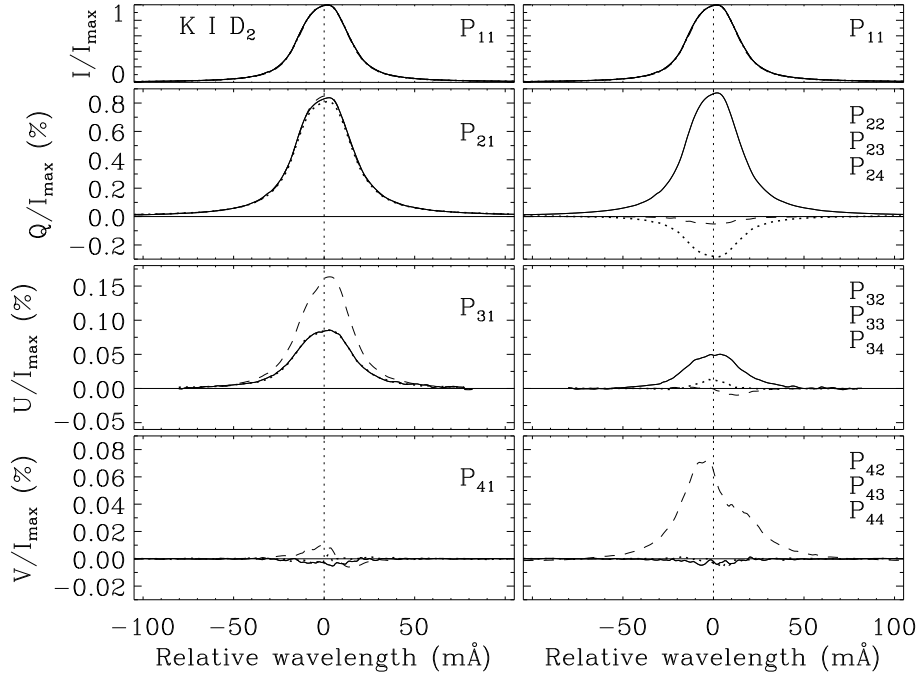Since the D_2 line is better understood than D_1, it can be used to estimate the

Fig. 8.4: Phase matrix elements for the K I D_2 line. The notation is the same as in Fig. 8.3.

collision depolarization factor for both experiments, i.e., the factor by which the polarizations should be scaled to represent the collision-free case. We estimate it to be ~ 34. Such a large value is almost unavoidable when using a buffer gas to avoid potassium deposits on the cell windows. However, since it may be regarded as a global scaling factor for both D_1 and D_2, and since our measuring system provides a very high S/N ratio even with the low polarization levels that we are dealing with, collisional depolarization is not a critical issue here.

8.5 Conclusion and Outlook

To shed light on the complex physics that underlies the enigmatic solar barium, sodium, and potassium D_1 polarization we have set up a laboratory experiment to explore the basic scattering physics under highly idealized and controlled conditions. The experiment is giving us a rich set of fully resolved polarized profiles with high S/N ratio for the scattering Mueller matrix elements. This empirical data set will provide powerful constraints on the underlying theory of quantum scattering and guide the theoretical efforts that we are currently undertaking.

In a second phase we are investigating the behavior of these matrix elements under the influence of external magnetic fields. We have Helmholtz coils capable of delivering any field strength within ± 40 G, which we can mount around the arms of the vapor cell along any of the three coordinate axes. The observed fact that the unexplained sodium D_1 polarization is found to vary substantially with location on the Sun (cf. also the potassium D_1 measurements presented in SPW4 by Stenflo (2006)) suggests that the solution of the solar enigma has to involve magnetic fields.

After the Boulder Workshop, measurements with a magnetic field perpendicular to the scattering plane have been conducted for K I D$_1$ with spectacular and quite unexpected results.

Acknowledgements

Chapter 9

Magnetic Field Dependence of Polarized Scattering on Potassium[*]

C. Thalmann[1]**, J.O. Stenflo**[1]**, A. Feller**[1]**, and A. Cacciani**[2]

Abstract

The polarization that is generated by scattering of light on atoms depends on the atomic quantum number structure. According to standard scattering theory the sodium D_1 (5896 Å) and potassium D_1 (7699 Å) lines should be intrinsically unpolarizable. Since however the existence of D_1 scattering polarization in the solar spectrum has been well documented but resisted theoretical explanations, we have carried out a laboratory experiment to find out if this is a problem of solar physics or of quantum physics. We find that the polarization matrix for scattering of the Stokes vector in the potassium D_1 line exhibits a rich structure that changes dramatically when the field strength is varied in the sub-gauss regime. In contrast, standard scattering theory predicts zero polarization, even when optical pumping of the magnetic sub-levels of the hyperfine structure multiplet is accounted for.

9.1 Introduction

At the previous Solar Polarization Workshop we presented initial results from a laboratory experiment (Thalmann et al., 2006) with the aim of shedding light on the enigmatic linear polarization peaks observed in the Doppler cores of the lines of Ba II D_1 4934 Å and Na I D_1 5896 Å in the Second Solar Spectrum (Stenflo, 1996; Stenflo & Keller, 1997; Landi degl'Innocenti, 1998; Stenflo et al., 2000a; Stenflo et al., 2000b; Trujillo Bueno et al., 2002; Casini et al., 2002; Casini & Manso Sainz, 2005). The recently observed spatial variability of the Na I D_1 peak (Stenflo et al., 2000b; Stenflo, 2008) suggests that local magnetic fields may be involved. Here we extend our previous laboratory work to explore the magnetic field dependence of the scattering polarization.

---

[*] This chapter will be published in in *Solar Polarization 5*, ASP Conference Series (2008).

[1] Institute of Astronomy, ETH Zurich, 8092 Zurich

[2] University of "La Sapienza", Piazzale Aldo Moro 2, I–00185 Rome

Our experiment allows us to isolate the scattering process from the complex environment of the solar atmosphere so that we can investigate it under clean, controlled conditions. A potassium vapor cell constructed by Alessandro Cacciani is illuminated with a laser that can be tuned across the chosen spectral line. The polarization state of the incident beam can be selected with a set of polarizers. Light scattered at 90° is passed through a piezoelastic modulator and a polarizer to a photomultiplier. A lock-in amplifier synchronized with the modulator demodulates the measured signal and extracts the polarization information with high sensitivity ($\sim 2 \times 10^{-5}$). A set of Helmholtz coils allows us to apply an external magnetic field anywhere within $\pm 30\,\mathrm{G}$ oriented along any one of the three spatial dimensions.

Since the Ba II and Na I lines observed on the Sun lie outside the spectral range in which tunable solid-state lasers are available, our experiment instead uses the potassium lines K I D_1 7699.0 Å and K I D_2 7664.9 Å lines. These line transitions have the same quantum structure as those of Ba II and Na I, including nuclear spin with hyperfine structure.

Note that our experiment is not intended to, and in fact does not, mimic sun-like conditions. The aim is rather to provide a stable observational test bench for scattering theory.

9.2 Experiment Design

The laboratory setup is shown in Fig. 9.1 and described in detail in our initial publication (Thalmann et al., 2006). The new feature here is the pair of Helmholtz coils, which can be mounted around the vapor cell in three different configurations, producing a magnetic field (1) perpendicular to the scattering plane, (2) parallel to the output beam, or (3) parallel to the input beam. The maximum attainable field strength is $\sim 20\,\mathrm{G}$ for the vertical configuration (1), and $\sim 30\,\mathrm{G}$ for the horizontal configurations (2) and (3).

We have one dedicated laser emitter for each of the K I D_1 and K I D_2 lines. Each has a tuning range of 2 Å, but our typical line profile scan uses only 200 mÅ. Since the laser allows monochromatic excitation of the potassium atoms, the experiment gives us the Stokes parameters for 90° scattering as functions of the position within the line profile without the need for a spectrograph.

The full parameter space of the experiment consists of one spectral profile for each combination of six input polarizations ($I\pm Q$, $I\pm U$, $I\pm V$), three output polarizations (Q, U, V), three magnetic field orientations, and a continuum of magnetic field strengths. We sample this vast parameter space by looking at cross-sections through the data cube for certain values of the magnetic field (e.g. $B = -20$, 0, and $+20\,\mathrm{G}$ perpendicular to the scattering plane), which guides us to promising areas that we investigate in greater detail by filling in measurements with a number of other field strength values.

In the data reduction process, the measured line profiles are corrected for dark current, calibrated, and spectrally aligned with the help of Gaussian fits of the intensity profile (the laser tuning range slowly drifts over time). A flat offset is
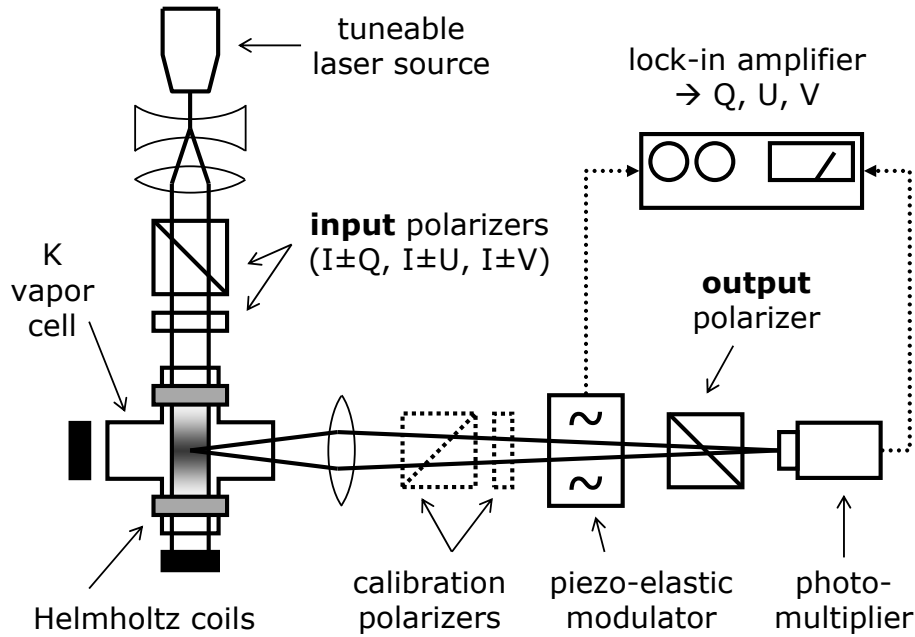
Fig. 9.1: Overview of the experimental setup with Helmholtz coils parallel to the ingoing beam.

subtracted from the Q, U, and V profiles, so that the values in the far wings converge to zero. Finally, since our setup does not provide a reliable absolute scale for the scattered intensity, the Q, U, and V profiles are divided by the maximum I_{max} of the intensity profile as normalization.

In general the scattered intensity is nearly independent of field strength. Only for one particular parameter combination (see Sect. 9.3) did we observe a strong dependence of the scattered intensity on the magnetic field. In this case we interlaced frequent reference measurements with $B = 0$ to determine how I_{max} varied with field strength. Note, however, that for the normalization of Q, U, and V measured for a given field strength, we use the value of I_{max} obtained for that same field strength.

9.3 Results

The K I D_2 polarization is found to behave as theoretically expected. Figure 9.2 illustrates this for the case of $I + Q$ input (linear polarization perpendicular to the scattering plane) and a magnetic field along the output beam. It shows Hanle depolarization, Hanle rotation, and longitudinal Zeeman effect as expected. We regard this as a proof-of-concept for the experimental setup.

The circumstance that we find a non-zero and non-trivial behavior in K I D_1 is in itself a contradiction with standard scattering theory, which predicts zero polarization for all geometries. While the theoretical approach of Landi Degl'Innocenti (1998) is able to produce some weak polarization through coherent pumping via the D_2 line, his mechanism does not apply to the present experiment, since our monochromatic laser light only excites the D_1 transition and does not involve D_2 at all in the pumping process.
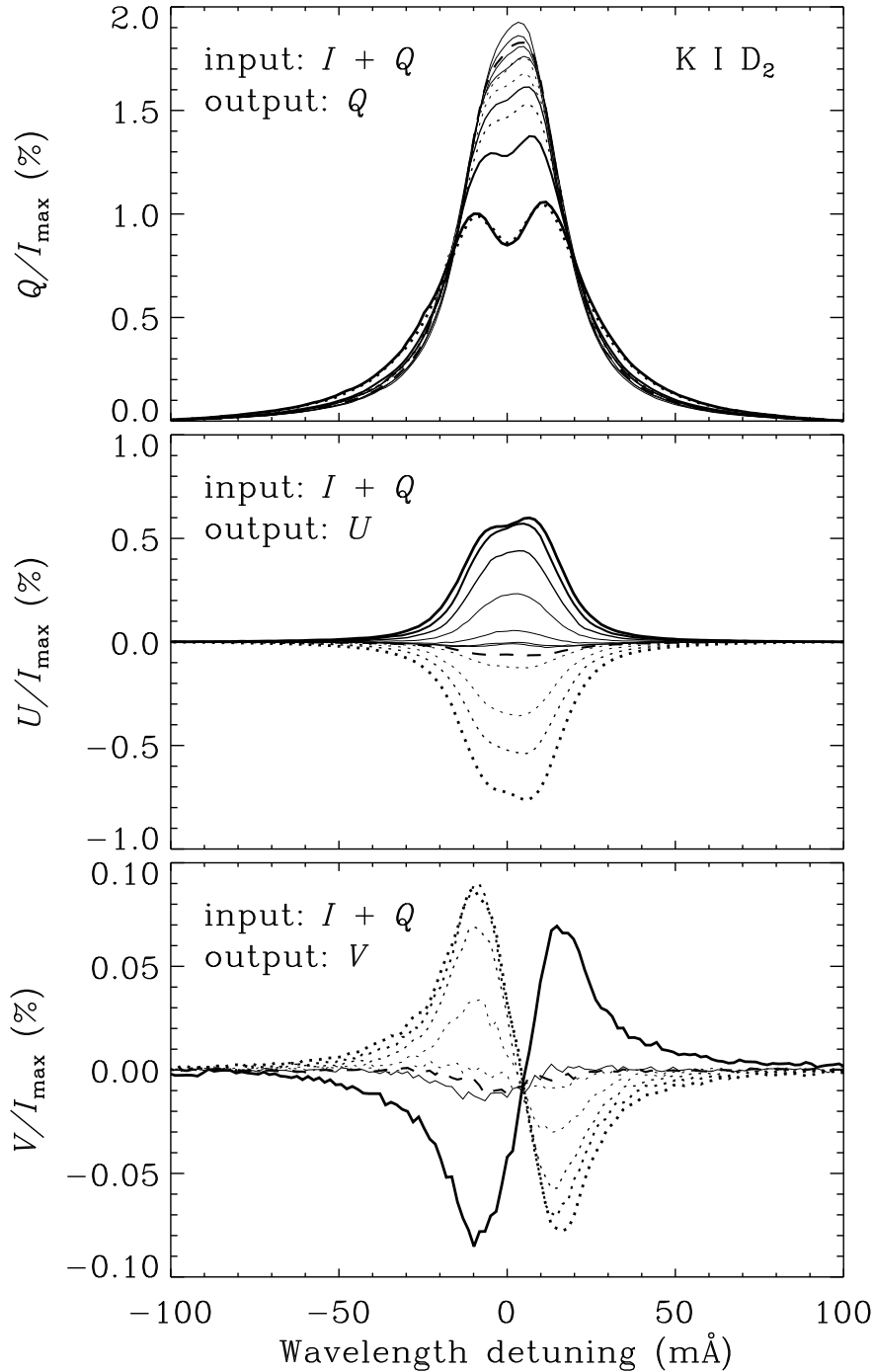
Fig. 9.2: Polarized line profiles of the well-behaved K I D$_2$ line. The ingoing pumping light is polarized as Stokes $I + Q$ (perpendicular to the scattering plane). The external magnetic field B is oriented along the *output beam* axis. Solid lines represent positive values of B (field pointing towards the detector), dotted lines represent negative values. The line thickness increases with field strength $|B|$, with $|B|_{max} \approx 30\,G$. The dashed curve in the top panel represents the case of zero field. We clearly observe the expected Hanle depolarization (top panel), Hanle rotation (middle panel), and longitudinal Zeeman effect (bottom panel).
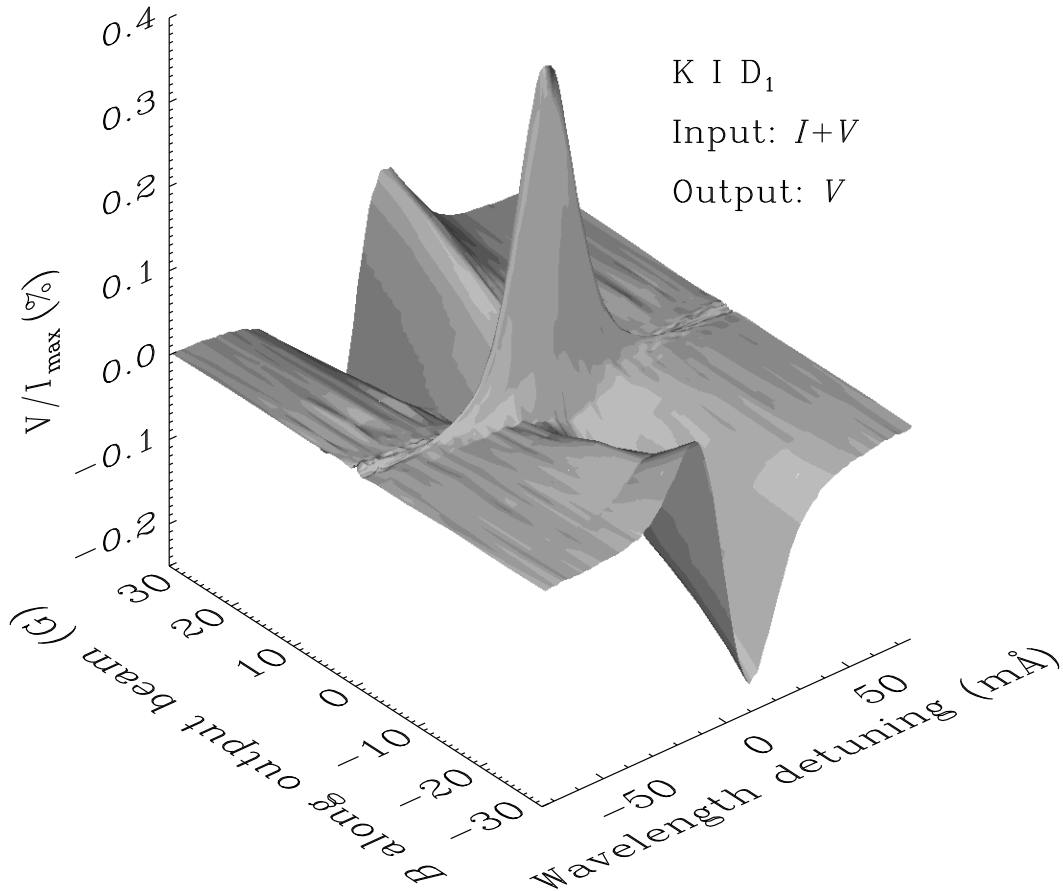
Fig. 9.3: Stokes V profiles of the K I D_1 line as functions of the magnetic field that is parallel to the *output beam*. The ingoing pumping light is polarized as Stokes $I + V$. Note that while the profiles are antisymmetric in the strong-field regime, a prominent symmetric signature appears around zero field. This may be an effect of Earth's "misaligned" magnetic field, which dominates in the sub-gauss regime and leads to a different net vector orientation of the external field.

Two examples of the rich structure of D_1 are shown in Figs. 9.3 and 9.4. In general, circular polarization ($I + V \rightarrow V$) produces the strongest signals, comparable in amplitude to the corresponding effects in D_2. The linear polarization effects tend to be weaker. For instance, the $I + Q \rightarrow Q$ polarization is an order of magnitude smaller than in D_2 and has the opposite sign (negative Q polarization is measured when pumping with $I + Q$, cf. Stenflo 2008).

Many measurements show a steep variation of the polarized line profile in the sub-gauss regime. These effects are influenced by the detailed vector orientation of Earth's magnetic field. While the applied magnetic field can compensate for it in one dimension, the remaining components slightly skew the geometry of the experiment and even dominate the direction of the total magnetic field in the sub-gauss regime. Figures 9.3 and 9.4 illustrate the extreme variability in this regime.

In most cases the magnetic field affects only the polarization but not the intensity of the scattered light. In one particular geometry, however (input $I \pm V$, magnetic
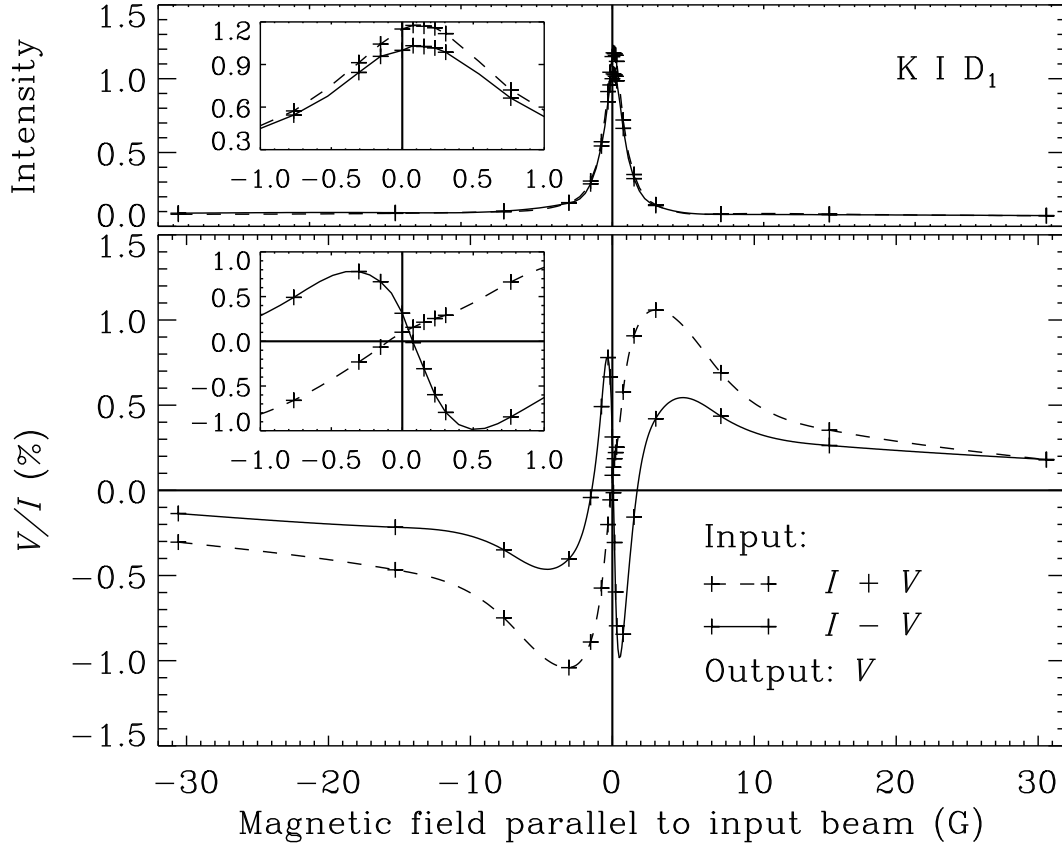
Fig. 9.4: Wavelength-integrated intensity (upper panel) and degree of circular polarization (integrated V divided by integrated I, lower panel), for the case when the incident radiation is circularly polarized (Stokes $I \pm V$) and directed along the magnetic field (while the scattered radiation is perpendicular to the field). The solid and dashed curves are spline fits to the data points (plus symbols) and refer to right- and left-handed circular polarization of the input beam, respectively. The inset panels provide a magnified view of the sub-gauss regime.

field parallel to the input beam), the scattered intensity is found to drop dramatically with increased field strength. Already a field of $|B| \approx 0.5\,\mathrm{G}$ cuts the intensity in half, and at $|B| \approx 10\,\mathrm{G}$ it levels off at 10% of its zero-field value. Although remarkable, this effect is not particularly enigmatic. In this geometry, the circularly polarized photons try to pump the atomic ground state into the sub-state with the highest or lowest m quantum number (depending on the handedness of the circular polarization), from which further excitation is suppressed. The dependence of the scattered intensity on magnetic field strength is illustrated in Figure 9.4. However, even with this extremely lopsided population of the ground state, standard scattering theory would predict zero integrated polarization in the scattered beam, in contrast to the highly structured behavior seen in the lower panel.

9.4 Conclusions and Outlook

Our laboratory experiment has revealed a richly structured behavior of the K $\textsc{i}$ D_1 scattering polarization, including an extreme sensitivity to the ambient magnetic field in the sub-gauss regime. These results appear inexplicable within the framework of standard scattering theory, which predicts zero results for all scattering geometries. An extension of quantum scattering theory that accounts for final-state interference in the scattering process (Stenflo, 2008) has recently produced some promising results that resemble the experimental data both qualitatively and quantitatively. While we cannot yet claim to have solved the sodium D_1 enigma, we have obtained clear evidence for deficencies in current scattering theory and provided a rich data set that will be of great value in constraining future attempts to find the correct theory.

Acknowledgements

Chapter 10

Conclusion and Outlook

Over the course of several years, our effort to find the telltale fingerprint of the D_1 Enigma in laboratory experiments has evolved from a casual idea to a custom-built scientific instrument. An extensive corpus of data on D_1 scattering has been acquired under much simpler and more well-defined conditions than the tangled fractal maelstrom of the solar atmosphere, providing valuable constraints and leverage to future theoretical approaches.

While the positive polarization peak in the center of Na I D_1 yet remains unexplained, the mere fact that consistent and reproducible scattering polarization signals have been measured in D_1 at all constitutes a glaring contradiction to the accepted theory, which predicts zero polarization.

The spatial variability of the Na I D_1 across the solar disk implies that localized conditions, most likely magnetic fields, are involved. The measured sensitivity of our polarized line profiles to weak magnetic fields supports this theory. The transversal Zeeman effect has been proposed as a possible source of positive linear polarization in D_1, but the vertical magnetic field strengths required to achieve such a strong signal would be prohibitively high for the lower chromosphere where the Na I D_1 line originates (Stenflo, 2006).

On the basis of the experimental data, Stenflo (2008) developed a novel theoretical approach involving quantum interference between the final states of the scattering process, rather than only in the intermediate states. His preliminary computational simulations succeed in generating scattering polarization in D_1 and even produce polarized line profiles that resemble the experimental data qualitatively and quantitatively. This promising theoretical effort is still ongoing and may yet hold the key to the solution of the D_1 enigma.

We conclude that the missing puzzle pieces are likely to be found in *both* atomic and solar physics. On one hand, standard scattering theory fails to deal with the fact that scattering polarization occurs in D_1 in our experiment; on the other hand, we expect the signature in the Second Solar Spectrum to be linked to local magnetic structures in the solar atmosphere in yet unknown ways.

The acquisition of data with our laboratory experiment is on a temporary hiatus at the time of writing, as the hardware is being transfered to the Istituto Ricerche Solari Locarno (IRSOL), where measurements are foreseen to be resumed by the local staff.

Appendix A

Polarized Light

A.1 Electromagnetic Waves

The four Maxwell equations provide the solid foundation for the entire theoretical edifice of electrodynamics. They describe the shape of electric and magnetic fields and their dependence on each other and on the sources that fuel them, the electric charges and currents. The two dynamic equations in particular state that a temporal change of the electric field gives rise to a magnetic field component, and vice versa. This enables electric and magnetic fields to sustain each other even in the absence of sources, in the form of electromagnetic radiation such as light.

In empty space devoid of charges, a three-dimensional wave equation for both the electric field $\vec{E}$ and the magnetic field $\vec{B}$ can be distilled from Maxwell's equations:

$$\nabla^2 \vec{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}, \qquad \nabla^2 \vec{B} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2}. \tag{A.1}$$

See Jackson (1975) for a detailed discussion. The solutions for the electromagnetic field under these conditions are *waves* that propagate at the speed $1/\sqrt{\mu_0 \varepsilon_0} =:$ c, i.e. the vacuum speed of light. Maxwell's equations further imply that at any given point in the wave, $\vec{E}$ and $\vec{B}$ must be perpendicular to each other and to the wave propagation vector $\vec{k}$, forming a right-handed trihedron $\{\vec{k}, \vec{E}, \vec{B}\}$. Finally, the energy density in both fields must be identical, i.e. $|\vec{E}| = |\vec{B}|$ in Gaussian units. Therefore, a description of $\vec{k}$ and $\vec{E}$ already fully characterizes an electromagnetic wave.

Let $\vec{k}$ point along the positive z axis. Then the general solution for $\vec{E}$ can be written as the sum of two orthogonal waves oscillating in the x and y directions, respectively:

$$\begin{aligned} E_x(\vec{r}, t) &= E_{0,x} \cdot \cos(\vec{k} \cdot \vec{r} - \omega \cdot t + \delta_x), \\ E_y(\vec{r}, t) &= E_{0,y} \cdot \cos(\vec{k} \cdot \vec{r} - \omega \cdot t + \delta_y), \end{aligned} \tag{A.2}$$

where $\omega = c \cdot |\vec{k}|$ is the angular frequency. The free parameters defining the wave are the real amplitudes $(E_{0,x}, E_{0,y})$ and the phases (δ_x, δ_y). Especially in quantum mechanics, it is popular to write these parameters as a complex amplitude $\vec{\mathcal{E}}$ given by $\mathcal{E}_{0,x} = E_{0,x} \cdot e^{-i\delta_x}$, $\mathcal{E}_{0,y} = E_{0,y} \cdot e^{-i\delta_y}$ and

$$\begin{aligned} \vec{E}(\vec{r}, t) &= \operatorname{Re}\left(\vec{\mathcal{E}}(\vec{r}, t)\right) = \operatorname{Re}\left(\vec{\mathcal{E}}_0 \cdot e^{i\vec{k}\vec{r} - i\omega t}\right) \\ &= \operatorname{Re}\left(\hat{e}_x \cdot E_{0,x} \cdot e^{i\vec{k}\vec{r} - i\omega t + \delta_x} + \hat{e}_y \cdot E_{0,y} \cdot e^{i\vec{k}\vec{r} - i\omega t + \delta_y}\right). \end{aligned} \tag{A.3}$$
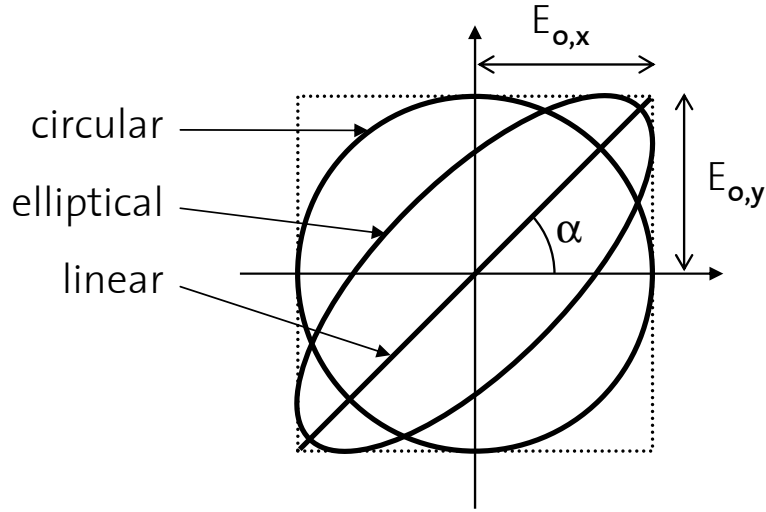
Fig. A.1: Three polarization states with the same $E_{0,\text{x}}$ and $E_{0,\text{y}}$, but different phases $\delta = \pi/2$, $\pi/2 > \delta > 0$, and $\delta = 0$. The tip of the $\vec{E}(t)$ vector then traces a circle, an ellipse, and a line, respectively (thick black curves). Circular and linear polarization are special cases of the general polarization ellipse.

The intensity of an electromagnetic wave, i.e. the time-averaged amount of transported energy per area, is given by

$$I = c\,\varepsilon_0 \left\langle \vec{E}^2 \right\rangle = \frac{c\,\varepsilon_0}{2}\,\vec{E}_0^2 = \frac{c\,\varepsilon_0}{2}\,(E_{0,\text{x}}^2 + E_{0,\text{y}}^2). \tag{A.4}$$

The relative strength of $E_{0,\text{x}}$ and $E_{0,\text{y}}$ and their phase difference $\delta = \delta_\text{x} - \delta_\text{y}$ describe the *polarization* of the wave, i.e. the way its $\vec{E}(\vec{r}, t)$ vector behaves on the xy plane for a fixed $\vec{r}$. The most general case is elliptic polarization; the tip of the vector $\vec{E}(\vec{r}, t)$ continually traces an ellipse on the xy plane (see Figure A.1). One distinguishes two special cases:

Linear polarization. For $\delta = 0$ or $\delta = \pi$, the two components are in phase, and the short axis of the ellipse vanishes. The resulting behavior of $\vec{E}$ is a simple oscillation along a line in the xy plane at an angle $\alpha = \arctan E_{0,\text{y}}/E_{0,\text{x}}$ from the x axis. Such light is called linearly polarized. The particular cases $E_{0,\text{x}} = 0$ or $E_{0,\text{y}} = 0$ describe light linearly polarized along the y or x axes, respectively.

Circular polarization. For $\delta = \pi/2$ or $3\pi/2$, and $E_{0,\text{x}} = E_{0,\text{y}}$, each component is at its full power when the other component passes through zero, and the amplitudes are equal. Consequently, $\vec{E}$ traces a circle on the xy plane. Such light is called circularly polarized. One distinguishes between right and left circular polarization according to the sense of rotation.

The general elliptical polarization state can be described as a sum of a linearly polarized and a circularly polarized component. Similarly, a linear polarization can be written as the sum of two circular polarizations, and vice versa.
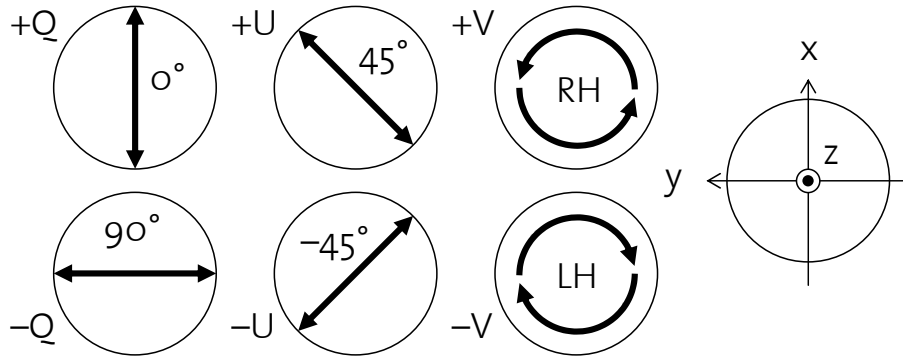
Fig. A.2: The definition of the six Stokes filters. The thick arrows designate the temporal behavior of the tip of the $\vec{E}$ vector. The propagation vector $\vec{k}$ points along the z axis, i.e. out of the page. In this work, we commonly choose the x axis (and thereby the $+Q$ direction) to be vertical.

A.2 Stokes Formalism

The description of polarization through the complex amplitude $\vec{\mathcal{E}}$ is known as Jones formalism. In astronomy, however, this notation is impractical. It can only describe *fully polarized* waves, in which one description of the $\vec{E}$ vector behavior holds true for the whole wave. On the other hand, all natural sources of light apart from exotic phenomena such as astrophysical masers emit incoherent radiation, in which the individual wave quanta (photons) do not share a common phase relation or polarization. While each individual photon is fully polarized, the uncorrelated ensemble might not exhibit any bias towards a linear direction or sense of rotation. Such light is referred to as *unpolarized*. Light which is not fully polarized but nevertheless biased towards a certain polarization state is called *partially polarized*. Any partially polarized state can be described as the superposition of one unpolarized and one fully polarized component.

Astronomical cameras measure intensity $I \propto \langle \vec{E}^2 \rangle$ rather than field strength $\vec{E}$, and are therefore blind to the phase properties of light. In 1852, Sir George Gabriel Stokes (1819–1903) published a formalism that characterizes polarization exclusively in terms of intensities. In comparison with the Jones formalism, it lacks a description of absolute phase (which is generally unobservable) but can instead express partial polarization.

The formalism is based on six polarizing filters, each of which projects the incoming light onto a pair of complementary polarization states, transmitting one of them and rejecting the other. The filter set comprises four linear polarizers oriented at $\alpha = 0°$, $90°$, $45°$ and $-45°$, as well as a right- and a left-handed circular polarizer (see Figure A.2). The light is passed through each filter, and the transmitted intensities are registered. The Stokes vector $\vec{S} = (I, Q, U, V)$ is then defined as

$$
\begin{aligned}
I &= I_{0°} + I_{90°} &&= |\mathcal{E}_\mathrm{x}|^2 + |\mathcal{E}_\mathrm{y}|^2 \\
Q &= I_{0°} - I_{90°} &&= |\mathcal{E}_\mathrm{x}|^2 - |\mathcal{E}_\mathrm{y}|^2 \\
U &= I_{45°} - I_{-45°} &&= 2\,\mathrm{Re}(\mathcal{E}_\mathrm{x}\mathcal{E}_\mathrm{y}^*) \\
V &= I_\mathrm{RH} - I_\mathrm{LH} &&= 2\,\mathrm{Im}(\mathcal{E}_\mathrm{x}\mathcal{E}_\mathrm{y}^*)
\end{aligned}
\tag{A.5}
$$

The three components Q, U, V form an orthonormal base of the possible range of polarizations. For example, fully vertically polarized light is described by the Stokes vector $\vec{S} = I(1, 1, 0, 0)$, fully left-hand circularly polarized light by $\vec{S} = I(1, 0, 0, -1)$. Due to the complementarity of the three filter pairs, the intensity can also be written as $I = I_{45°} + I_{-45°} = I_{\mathrm{RH}} + I_{\mathrm{LH}}$. Thus, only four intensity measurements are necessary to determine the full Stokes vector, e.g. I, $I_{0°}$, $I_{45°}$, and I_{RH}.

The total polarized flux I_{P} and the degree of polarization p are given by

$$I_{\mathrm{P}} = \sqrt{Q^2 + U^2 + V^2} \leq I; \qquad p = \frac{I_{\mathrm{P}}}{I}. \tag{A.6}$$

Fully polarized light is therefore characterized by $I_{\mathrm{P}} = I$, while the Stokes vector $I(1, 0, 0, 0)$ represents unpolarized light.

Since absolute phase is not measured in Stokes polarimetry, a linear polarization at the angle α is indistinguishable from linear polarization at angle $\alpha + \pi$. The angular periodicity of the system is π rather than 2π. The Stokes vector for fully linearly polarized light at the angle α is $I(1, \cos 2\alpha, \sin 2\alpha, 0)$.

For a more in-depth description of polarized light and the Stokes formalism, refer to Collett (1993).

Appendix B

Polarization Compensator Routines

The software package for the current two-dimensional compensation mode is built around the following programs (excluding minor helper routines). For simplicity, we set $I := 1$ such that Q, U refer to relative polarization rather than polarized flux.

Internal Routines

`Rismp` is the basic image-sequence acquisition routine with two-phase fixed-pattern correction, slightly adapted from the usual version to accomodate the needs of the compensator plate. Specifically, it enforces the execution of "quick look" image pre-processing for every exposure frame taken, in order to monitor the background polarization in real-time. This computation slightly slows down the image acquisition process, thus it is usually disabled during conventional ZIMPOL observations, where there is no need for instantaneous data feedback. The name `rismp` derives from the words "[instantaneous] reduction", "image sequence [acquisition]" and "multi-phase [fixed-pattern correction]".

`P_from_i` returns the polarization $p(\iota)$ caused by the glass plate at given inclination angle ι. It is implemented as a simple polynomial in ι^2 and ι^4 that fits the theoretical curve given in Equation 4.1 perfectly in the relevant range ($0° < \iota \lesssim 50°$).

`Q_from_ai, U_from_ai` return the $Q(\alpha, \iota)$ and $U(\alpha, \iota)$ polarizations caused by the glass plate at a given azimuth α and inclination ι as given by Equation 4.3, making use of `P_from_i`.

`I_from_qu` returns the plate inclination $\iota(Q, U)$ needed to compensate a given degree of polarization $p = \sqrt{Q^2 + U^2}$. It is implemented as a fourth-order polynomial fit in the variable $\sqrt[4]{p}$.

`A_from_qu` returns the plate azimuth $\alpha(Q, U)$ in the interval $-67.5° < \alpha \leq 112.5°$ required to compensate linear polarization in the direction given by Q, U. It is implemented with an arctangent of Q and U and some infrastructure to ensure a continuous mapping of polarization directions to azimuth angles.

Appendix B. Polarization Compensator Routines

User Routines

Setpolrec defines the rectangular area on the image in which to minimize the mean polarization. This is useful to exclude a "dirty" part of the image from the algorithm.

QUkill sets the plate to a neutral position ($\iota = 0°$), acquires a polarization image of both Q and U, calculates a plate configuration to compensate the measured background using A_from_qu and I_from_qu, and positions the plate accordingly.

QUcomp, when launched as a background process during an image acquisition series, constantly monitors the mean background polarization value provided by the image pre-processing functionality of the ZIMPOL software. As soon as the value changes (i.e. a new exposure has been taken and pre-processed), the routine reconstructs the current background polarization by adding this measured residual polarization to the polarization being compensated by the plate (which is derived from the plate angles using Q_from_ai and U_from_ai), then calculates new plate angles to compensate this background, and adjusts the plate accordingly. In order to prevent the plate from constantly attempting to compensate the small statistical fluctuations in measured polarization, one can define a number of exposures to be accumulated and averaged before each iterative compensation.

QUcomp is intended to make incremental adjustments to an already reasonably well-oriented compensator plate. It is therefore advisable to execute QUkill prior to launching it for the first time or after a change of the experimental setup.

Polcompstop toggles a boolean variable ordering a running QUcomp background process to terminate.

Compensate is an automated image series acquisition routine making use of real-time polarization compensation. First, QUcomp is launched as a background process. Then, a specified number of image subseries are taken, each consisting of 30 Q images followed by 30 U images. This ensures that both components of the internal model of the two-dimensional background polarization are updated regularly, and that any slow drifts in the properties of the science target affect the Q and U records equally. After each image, the current compensator parameters and measured residual polarizations are piped to the output console, where they can be picked up and written to a log file by an external application (z2csh). The routine terminates with a call to Polcompstop.

While Compensate relieves the user from micromanaging the QUcomp background process, it is still necessary to execute QUkill prior to the first observation in order to prepare the plate in a sensible starting position.

CompensateQ is a simpler variant of Compensate which does not take U measurements. It is useful if one is interested exclusively in Q polarization.

Appendix C

SPHERE/ZIMPOL Performance Plots for Specific Target Stars

Figures C.1–C.8 show the simulated performance of SPHERE/ZIMPOL for eight specific target stars under SC3 conditions. In each plot, the 5σ detectivity curves are given for the double-difference image, for the same after angular averaging over 23 positions of active field rotation, and for the photon-limit. The performance that can be expected from the real instrument lies somewhere between the latter two curves. The closer the two curves lie to each other, the more probable it is that a photon-limited detectivity can be achieved.

Planet signals are given for the optimistic case of 50% polarization (crosses) and the conservative case of 10% polarization (diamonds) at various angular separations.

These results are briefly discussed in Section 5.7.

Fig. C.1: Performance plot for α Centauri A under SC3 conditions.
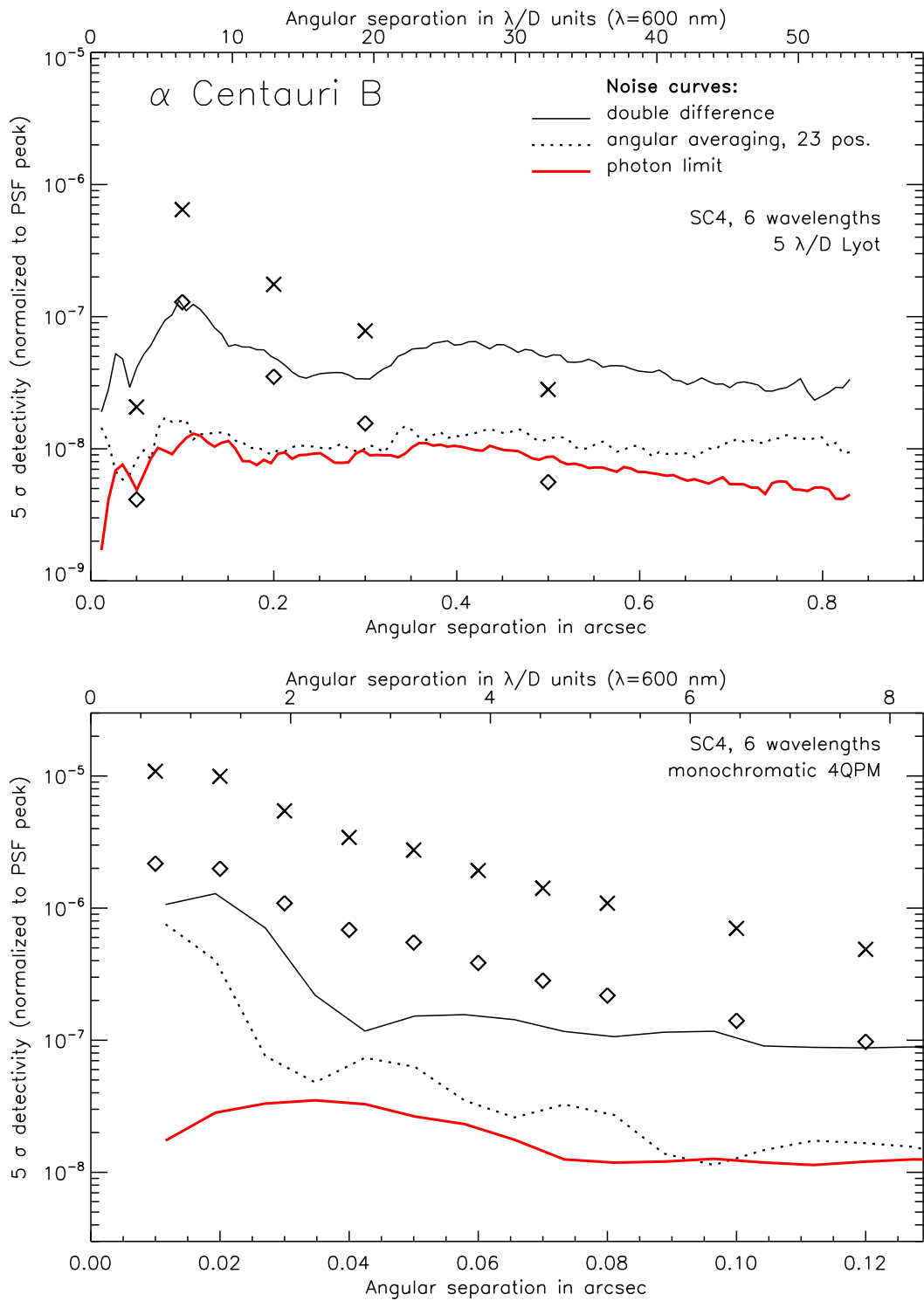
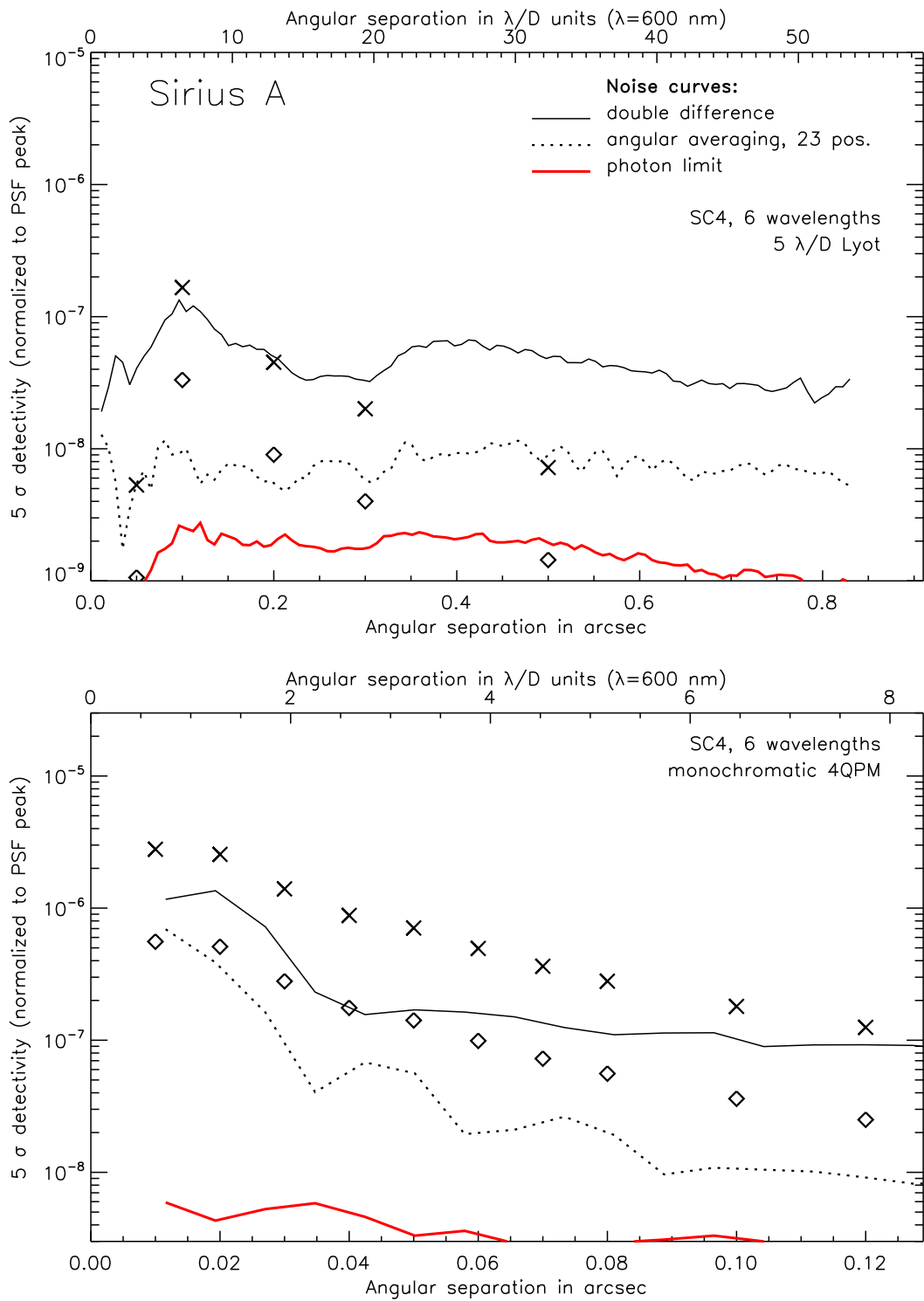**Fig. C.2:** Performance plot for α Centauri B under SC3 conditions.

Fig. C.3: Performance plot for Sirius A under SC3 conditions.
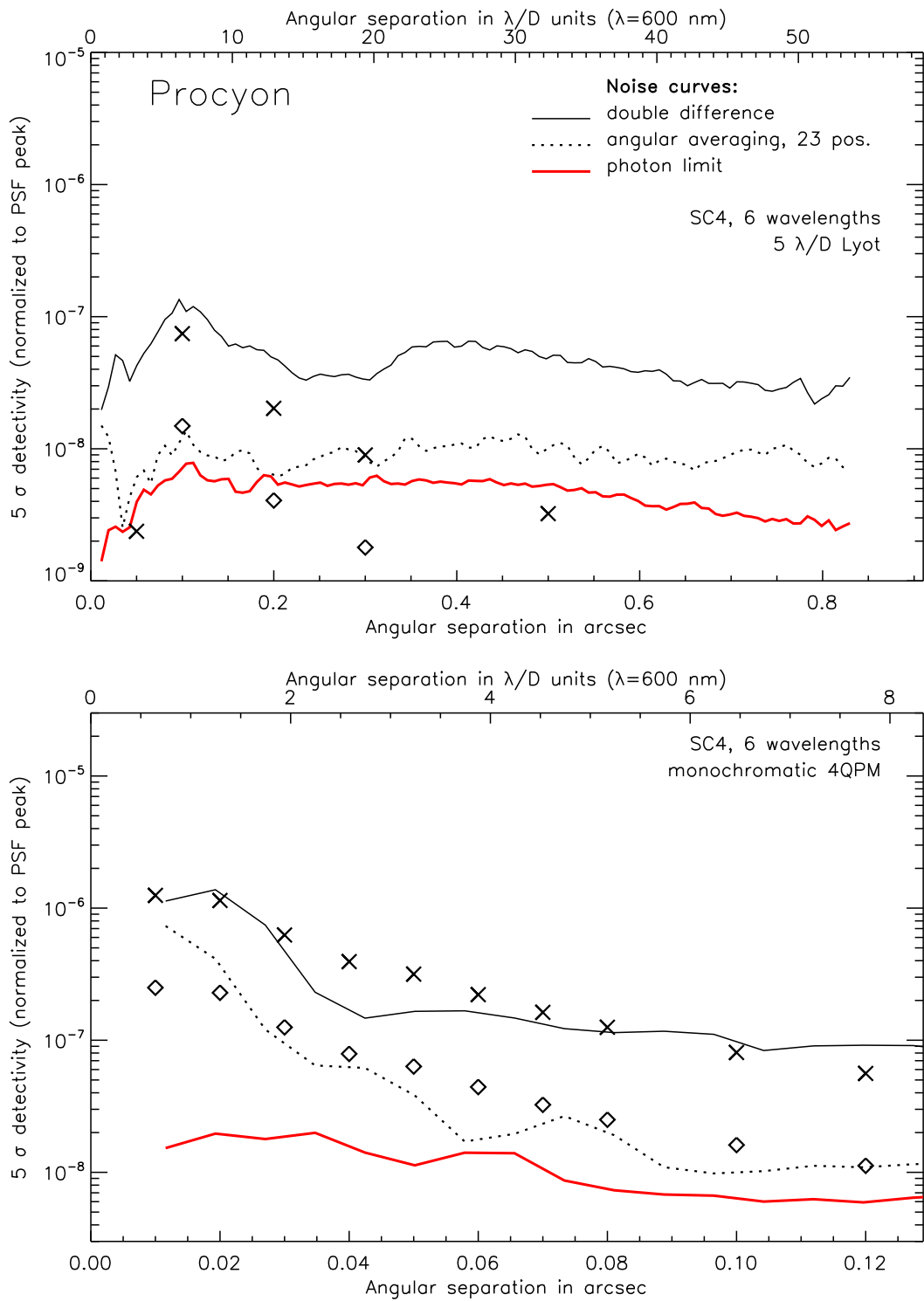
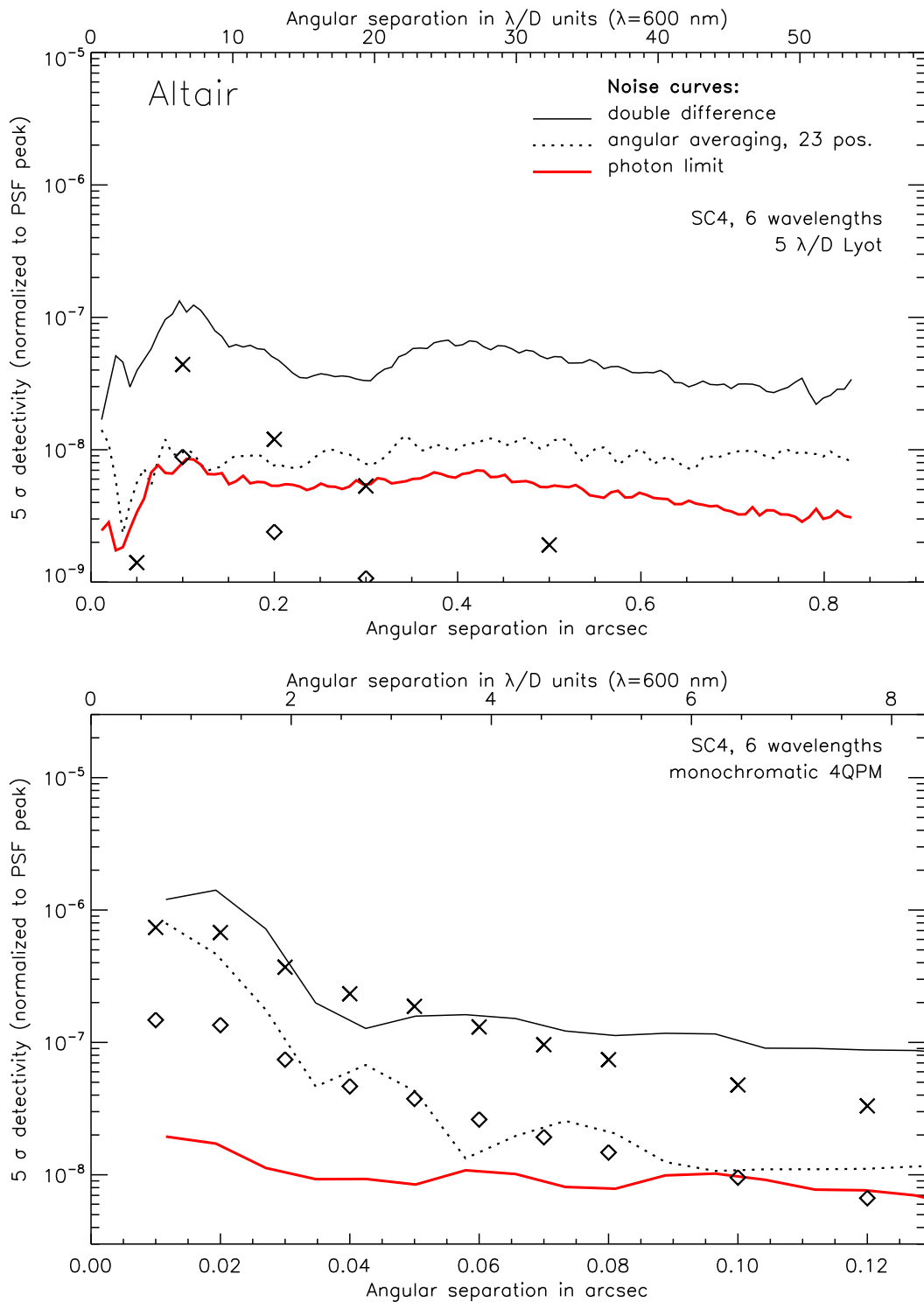**Fig. C.4:** Performance plot for Procyon under SC3 conditions.

Fig. C.5: Performance plot for Altair under SC3 conditions.
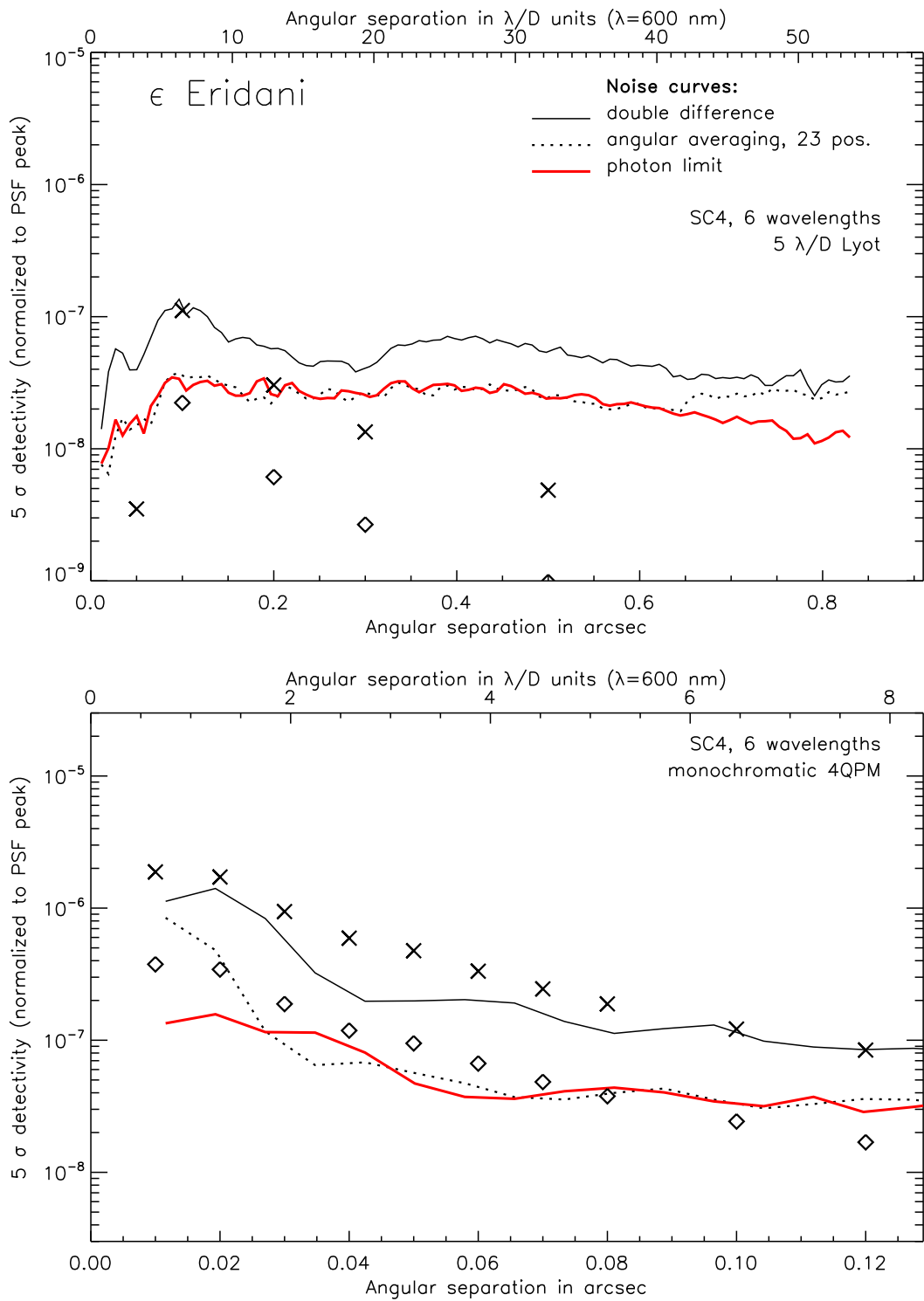
Fig. C.6: Performance plot for ε Eridani under SC3 conditions.

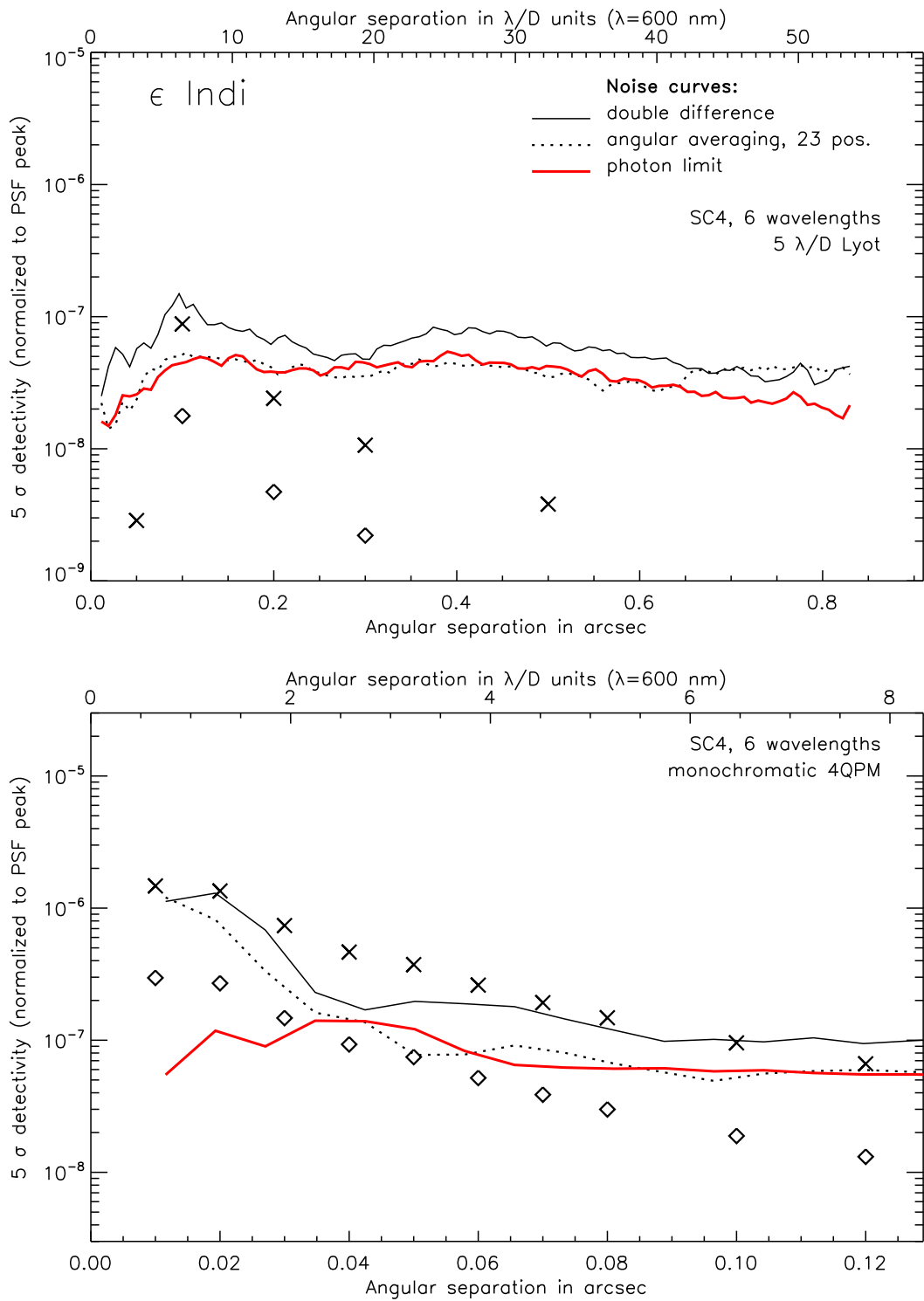**Fig. C.7:** Performance plot for τ Ceti under SC3 conditions.

Fig. C.8: Performance plot for ε Indi under SC3 conditions.

Bibliography

Abe, L., Domiciano de Souza, Jr., A., Vakili, F., & Gay, J. 2003, A&A, 400, 385

Afonso, C. & Henning, T. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 366, Transiting Extrapolar Planets Workshop, ed. C. Afonso, D. Weldrake, & T. Henning, 326–+

Apai, D., Pascucci, I., Brandner, W., et al. 2004, A&A, 415, 671

Babcock, H. W. 1953, ApJ, 118, 387

Beaulieu, J.-P., Bennett, D. P., Fouqué, P., et al. 2006, Nature, 439, 437

Benedict, G. F., McArthur, B. E., Forveille, T., et al. 2002, ApJ, 581, L115

Beuzit, J.-L., Feldt, M., Dohlen, K., et al. 2006, The Messenger, 125, 29

Boccaletti, A., Daban, J.-B., & Schmid, H. M. 2007, Coronagraph Design Report, ESO internal document VLT-TRE-SPH-14690-0229, LESIA, LUAN, ETHZ

Born, M. & Wolf, E. 1980, Principles of optics. Electromagnetic theory of propagation, interference and diffraction of light (Oxford: Pergamon Press, 1980, 6th corrected ed.)

Brückner, G. 1963, Zeitschrift fur Astrophysik, 58, 73

Cacciani, A. 1981, Space Science Reviews, 29, 403

Cacciani, A. & Fofi, M. 1978, Sol. Phys., 59, 179

Cacciani, A., Moretti, P. F., & Rodgers, W. E. 1997, Sol. Phys., 174, 115

Carbillet, M., Verinaud, C., Guarracino, M., et al. 2004, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 5490, Advancements in Adaptive Optics. Edited by Domenico B. Calia, Brent L. Ellerbroek, and Roberto Ragazzoni. Proceedings of the SPIE, Volume 5490, pp. 637-648 (2004)., ed. D. Bonaccini Calia, B. L. Ellerbroek, & R. Ragazzoni, 637–648

Casini, R., Landi Degl'Innocenti, E., Landolfi, M., & Trujillo Bueno, J. 2002, ApJ, 573, 864

Casini, R. & Manso Sainz, R. 2005, ApJ, 624, 1025

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

Charbonneau, D., Allen, L. E., Megeath, S. T., et al. 2005, ApJ, 626, 523

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45

Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, ApJ, 568, 377

Chauvin, G., Lagrange, A.-M., Dumas, C., et al. 2005, A&A, 438, L25

Collett, E. 1993, Polarized Light: Fundamentals and Applications (Marcel Decker)

Dohlen, K. 2007, System Analysis Report, ESO internal document VLT-SPE-SPH-14690-0161, LAM

Dohlen, K., Beuzit, J.-L., Feldt, M., et al. 2006, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy. Edited by McLean, Ian S.; Iye, Masanori. Proceedings of the SPIE, Volume 6269, pp. 62690Q (2006).

Dominik, M., Albrow, M. D., Beaulieu, J.-P., et al. 2002, Planet. Space Sci., 50, 299

Feller, A. 2007, PhD thesis, ETH Zürich (Switzerland)

Fini, L., Carbillet, M., & Riccardi, A. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne, 253–+

Fischer, D. A., Marcy, G. W., Butler, R. P., et al. 2008, ApJ, 675, 790

Fluri, D. M., Holzreuter, R., Klement, J., & Stenflo, J. O. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 307, Astronomical Society of the Pacific Conference Series, ed. J. Trujillo-Bueno & J. Sanchez Almeida, 263–+

Fusco, T. 2007, AO Analysis Report, ESO internal document VLT-TRE-SPH-14690-0248, ONERA

Fusco, T., Petit, C., Rousset, G., et al. 2006, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 6272, Advances in Adaptive Optics II. Edited by Ellerbroek, Brent L.; Bonaccini Calia, Domenico. Proceedings of the SPIE, Volume 6272, pp. 62720K (2006).

Gandorfer, A. 2000, Social Studies of Science

Gandorfer, A. 2002, Social Studies of Science

Gandorfer, A. 2005, Social Studies of Science

Gilmozzi, R. & Spyromilio, J. 2007, The Messenger, 127, 11

Gisler, D. & Schmid, H. M. 2003, in ASP Conference Series, Vol. 307, Solar Polarization 3, ed. J. Trujillo Bueno & J. Sánchez Almeida, 48

Gratton, R. & Mouillet, D. 2007, Science Case, ESO internal document VLT-SPE-SPH-14690-0234, INAF-OAPD, LAOG

Hale, G. E. 1908, PASP, 20, 287

Hatzes, A. P., Cochran, W. D., Endl, M., et al. 2003, ApJ, 599, 1383

Holzreuter, R., Fluri, D. M., & Stenflo, J. O. 2005, A&A, 434, 713

Holzreuter, R. & Stenflo, J. O. 2007a, A&A, 467, 695

Holzreuter, R. & Stenflo, J. O. 2007b, A&A, 472, 919

Jackson, J. D. 1975, Classical Electrodynamics, 2nd edn. (New York: John Wiley & Sons)

Joos, F. 2007, PhD thesis, ETH Zürich (Switzerland)

Kasper, M., Verinaud, C., Beuzit, J.-L., et al. 2007, in In the Spirit of Bernard Lyot: The Direct Detection of Planets and Circumstellar Disks in the 21st Century, ed. P. Kalas

Keller, C. U., Harvey, J. W., & The Solis Team. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 307, Astronomical Society of the Pacific Conference Series, ed. J. Trujillo-Bueno & J. Sanchez Almeida, 13–+

Kerkeni, B. & Bommier, V. 2002, A&A, 394, 707

Kiepenheuer, K. O. 1953, ApJ, 117, 447

Klement, J. & Stenflo, J. O. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 307, Astronomical Society of the Pacific Conference Series, ed. J. Trujillo-Bueno & J. Sanchez Almeida, 278–+

Kuhn, J. R., Potter, D., & Parise, B. 2001, ApJ, 553, L189

Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É. 2007, ApJ, 660, 770

Landi degl'Innocenti, E. 1998, Nature, 392, 256

Landi degl'Innocenti, E. 1999, in Astrophysics and Space Science Library, Vol. 243, Polarization, ed. K. N. Nagendra & J. O. Stenflo, 61–71

Lee, D., Haynes, R., Ren, D., & Allington-Smith, J. 2001, PASP, 113, 1406

Marois, C., Doyon, R., Racine, R., & Nadeau, D. 2000, PASP, 112, 91

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

Mayor, M., Pepe, F., Queloz, D., et al. 2003, The Messenger, 114, 20

Mayor, M. & Queloz, D. 1995, Nature, 378, 355

O'Toole, S. J., Tinney, C. G., & Jones, H. R. A. 2008, MNRAS, 336

Povel, H. 1995, Optical Engineering, 34, 1870

Povel, H. P., Aebersold, F., & Stenflo, J. 1990, Appl. Opt., 29, 1186

Povel, H. P., Keller, C., & Yadigaroglu, I.-A. 1994, Appl. Opt., 33, 4254

Quirrenbach, A., Coude Du Foresto, V., Daigne, G., et al. 1998, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 3350, Proc. SPIE Vol. 3350, p. 807-817, Astronomical Interferometry, Robert D. Reasenberg; Ed., ed. R. D. Reasenberg, 807–817

Racine, R., Walker, G., Nadeau, D., Doyon, R., & Marois, C. 1999, PASP, 23, 587

Rivera, E. J., Lissauer, J. J., Butler, R. P., et al. 2005, ApJ, 634, 625

Roddier, F. 1999, Adaptive optics in astronomy (Adaptive Optics in Astronomy)

Roddier, F. & Roddier, C. 1997, PASP, 109, 815

Roelfsema, R. & Gisler, D. 2007, ZIMPOL FLC Test Report, ESO internal document VLT-TRE-SPH-14690-0185, ASTRON

Roelfsema, R. & Schmid, H. M. 2007, ZIMPOL System Design & Analysis, ESO internal document VLT-TRE-SPH-14690-0180, ASTRON, ETHZ

Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., & Labeyrie, A. 2000, PASP, 112, 1479

Scherrer, P. H., Bogart, R. S., Bush, R. I., et al. 1995, Sol. Phys., 162, 129

Schmid, H. M. 2007, ZIMPOL System Overview, ESO internal document VLT-TRE-SPH-14690-0179, ETHZ

Solanki, S. K., Lagg, A., Woch, J., Krupp, N., & Collados, M. 2003, Nature, 425, 692

Soummer, R., Aime, C., & Falloon, P. E. 2003, A&A, 397, 1161

Stam, D. M., Hovenier, J. W., & Waters, L. B. F. M. 2004, A&A, 428, 663

Stenflo, J. O. 1994, Solar Magnetic Fields (Dordrecht: Kluwer)

Stenflo, J. O. 1996, Nature, 382, 588

Stenflo, J. O. 1997, A&A, 324, 344

Stenflo, J. O. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 358, Astronomical Society of the Pacific Conference Series, ed. R. Casini & B. W. Lites, 215–+

Stenflo, J. O. 2008, in Solar Polarization 5, Astronomical Society of the Pacific Conference Series

Stenflo, J. O., Baur, T. G., & Elmore, D. F. 1980, A&A, 84, 60

Stenflo, J. O., Gandorfer, A., & Keller, C. U. 2000a, A&A, 355, 781

Stenflo, J. O. & Keller, C. U. 1997, A&A, 321, 927

Stenflo, J. O., Keller, C. U., & Gandorfer, A. 2000b, A&A, 355, 789

Stenflo, J. O., Twerenbold, D., & Harvey, J. W. 1983a, A&AS, 52, 161

Stenflo, J. O., Twerenbold, D., Harvey, J. W., & Brault, J. W. 1983b, A&AS, 54, 505

Sudarsky, D., Burrows, A., Hubeny, I., & Li, A. 2005, ApJ, 627, 520

Thalmann, C. 2004, Master's thesis, ETH Zürich

Thalmann, C. 2007, ZIMPOL Simulations and Performance Analysis Report, ESO internal document VLT-TRE-SPH-14690-0191, ETHZ

Thalmann, C. & Schmid, H. M. 2004, ZIMPOL Detection Simulation, ESO internal document CHEOPS-TRE-ETH-00035, ETHZ

Thalmann, C., Stenflo, J. O., Feller, A., & Cacciani, A. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 358, Solar Polarization 4, ed. R. Casini & B. W. Lites, 323–+

Trujillo Bueno, J. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 236, Advanced Solar Polarimetry – Theory, Observation, and Instrumentation, ed. M. Sigwarth, 161–+

Trujillo Bueno, J., Casini, R., Landolfi, M., & Landi Degl'Innocenti, E. 2002, ApJ, 566, L53

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

Zeeman, P. 1897, ApJ, 5, 332

Acknowledgements

First of all, Prof. Jan Olof Stenflo and PD Dr. Hans Martin Schmid merit my gratitude for making possible and supervising both my Masters and PhD theses, for providing a stable, creative and productive working environment, for making time in their busy schedules whenever I needed their input, and in turn for spurning me into action before rapidly approaching deadlines... oh, and of course for allowing me to visit conferences in locations as beautiful as Santorini.

Daniel Gisler deserves special thanks for the inordinate amount of time he had to sacrifice helping me out in the laboratory, resolving problems in IDL coding, retrieving crucial equipment from obscure storage places, or just giving advice on whether the weather was going to be clear enough to run a telescope practicum in the evening. Similarly, Alex Feller helped me out above and beyond the call of duty with expertise on numerical recipes, insight into the laboratory experiment etc. I would also like to appreciate the aid and advice readily provides by René Holzreuter, Dominique Fluri, Marina Battaglia, Svetlana Berdyugina and many other colleagues on a plethora of matters ranging from science and publishing to LaTeX and UNIX. Peter Steiner deserves special mention for patiently explaining my numerous blunders in the use of the workstations and the intranet. David Mouillet volunteered to review the SPHERE part of this thesis, for which I am grateful.

A huge thank you goes to all members, past and present, of the institute's "geek squad", for the countless hours of social fun spent with science fiction DVD sessions, playing poker and, most importantly, the table-top fantasy roleplaying game Dungeons & Dragons. My unending gratitude goes to Stephen Marsden for introducing me to D&D in the first place, and for the engaging, entertaining and scary (but secretly merciful) mastering of the Geoff campaign. I still can't believe Coram Machira made it to 10th level with a d4 hit die and a Constitution score of 10 without ever dying. His daughter should be about a year old by now. ☺ Athman Boukhaoua deserves no less thanks, having mastered for us two entire campaigns with great attention to detail and internal consistency, and an even higher scare factor (and body count) than Stephen's. An honorable mention goes to René for sharing my degree of ~~pathological addiction~~ enthusiasm for the game, and annihilating innumerable hours of working time with chats and technical discussions about it.

Special thanks go to my parents for raising me in the scientific mindset, fostering my interest in science as a whole, and financing my education.

Finally, and most importantly, I am above all eternally grateful to the love of my life, Simone Weinmann, for supporting me in every way, taking an active interest in my work, granting a +20 morale bonus in hard times, and making me happier than anyone deserves to be.

Curriculum Vitae

of Christian Daniel Thalmann, born September 29th 1979
citizen of Wiezikon TG, Switzerland

1985–1992	Primary school in Uster, Switzerland
1992–1999	Matura type C (math & science) Kantonsschule Stägenbuck, Dübendorf, Switzerland
Spring 1999	Compulsory recruiting into the Swiss army Basic training: medic Specialist training: emergency surgical assistance
Summer 1999	3 months of English language school in Santa Barbara, CA
Dec. 1999	Cambridge Certificate of Proficiency in English, Zurich
1999–2004	Master degree in physics at ETH Zurich Specialization in astrophysics
2003–2004	Master thesis "Testing the ZIMPOL Concept for Exoplanet Search" Institute of Astronomy, ETH Zurich Supervisors: Prof. J.O. Stenflo, PD Dr. H.M. Schmid
2004–2008	PhD thesis "Applications of High-Precision Polarimetry to Extrasolar Planet Search and Solar Physics" Institute of Astronomy, ETH Zurich Supervisors: Prof. J.O. Stenflo, PD Dr. H.M. Schmid