Réseaux sublambda pour l'imagerie et la caractérisation de systèmes planétaires extrasolaires

Dimitri Mawet

UNIVERSITÉ de Liège

Faculté des Sciences **Département d'Astrophysique, Géophysique et Océanographie**

Subwavelength gratings for extrasolar planetary system imaging and characterization

THÈSE

présentée pour l'obtention du diplôme de

Docteur en Sciences

par

Dimitri Mawet

Soutenue publiquement le 15 septembre 2006 devant le Jury composé de :

Président :	Pr. Jean-Pierre Swings
Directeur de thèse :	Pr. Jean Surdej
Examinateurs :	Pr. Serge Habraken
	Pr. Claude Jamar
	Dr. Alain Léger
	Dr. Bertrand Mennesson
	Dr. Daniel Rouan

Institut d'Astrophysique et de Géophysique de Liège

Mis en page avec la classe thloria.

Acknowledgments

The completion of this work would not have been possible without the assistance of a number of people. First and foremost, I would like to thank my advisor, Professor Jean Surdej, for his guidance throughout its duration. His presence as a mentor back from my undergraduate student years and my diploma thesis has been a invaluable inspiration. I would also like to thank him for introducing me to the numerous international scientific collaborators that allowed me to develop the subject and the ramifications of this thesis.

My gratitude also goes to the "old" HOLOLAB team, Vincent Moreau, Bernard Tilkens and Yvon Renotte for introducing me to the practical aspects and applications of subwavelength gratings, dating back to the very beginning of my diploma thesis.

I am also indebted to the Centre Spatial de Liège "ZOG team". Their logistical support, competence and encouragements are at the heart of a great deal of this work. My warmest thanks go in particular to Serge Habraken, Denis Vandormael, Jérôme Loicq, David Verstraeten, Jean-Yves Plesseria and Karl Fleury.

I would like to express my gratitude to the LESIA "four-quadrant team" who hosted me at Observatoire de Paris-Meudon during my numerous stays and allowed me to work on this tremendous component. My thanks go in particular to Daniel Rouan, Anthony Boccaletti and Pierre Baudoz. I would like to address my very special thanks to Jacques Baudrand, who is one of the major actors in the success of the "four-quadrant multinational", for his continuous interest and support in my work.

I am very grateful to the "Darwin team" at Institut d'Astrophysique Spatiale, who hosted me in Orsay and gave me the opportunity to work on the fascinating but challenging subject of nulling interferometry. It was also the occasion for me to enrich my skills by participating to the studies of the very interesting and original topic of Ocean Planets. My thanks go in particular to Alain Léger and Alain Labèque who proposed me to participate to the studies on the achromatic phase shifters for Darwin, which became one of the major subjects of this thesis. I would also like to warmly thank Marc Ollivier, Franck Brachet and Bruno Chazelas.

These stays in Paris at LESIA and IAS would not have been possible without the intervention of Pierre Léna, whom I want to specially thank. I would like to also thank Jean Schneider for its stimulating encouragements from the beginning of my entry in the community of "exoplanet hunters".

I am grateful to Professors Jean-Pierre Swings, Serge Habraken, Claude Jamar, Dr Alain Léger, Dr Daniel Rouan and Dr Bertrand Mennesson, members of my thesis committee, for accepting to read and evaluate my work.

I would particularly like to thank three colleagues and friends. This work would not have been possible without their help and friendship. I don't know by whom to start because they all contributed to this work differently but with the same subjective importance to me. In alphabetical order then, I would like to warmly thank Olivier Absil for guiding my very first steps in the field and as being a roommate with whom I learnt and laughed a lot. I also want to address a very special thank to Cédric Lenaerts. In the lab, for mathematical simulations, his encouragements and the discussions we have had are innumerable. And, last but not least (he surely will recognize himself), for his "ideas of the day", Wednesday's chocolate bread, four-times-a-day coffeeeee, for his passion of Belgian beers and cheese, for the endless political and philosophical discussions we

have had, for his aversion of professional amateurism, for his incommensurable obligingness, I would like thank Pierre Riaud, alias Poulpe, the most Belgian French that I know. He largely contributed to this thesis.

I also wish to thank all my colleagues at the Institut d'Astrophysique et de Géophysique de Liège. Special thanks to Denise for taking care of my travels and refunds.

Finally, I take a particular pleasure in thanking my parents and my family for their support back from the beginning of my studies (especially when I needed it most).

I would not have finished these acknowledgements without thanking my little secretary Jessica, who helped me in the redaction of the weighty and essential bibliography of this work. More importantly though, she provided the encouragement, the reassurance, and the strength that I needed to see this through, in particular during the final stages of this work. I dedicate this thesis to her.

This research was supported by a fellowship from the Belgian National Science Foundation ("Boursier du FRIA"). During his stay at Observatoire de Paris-Meudon and at Institut d'Astrophysique Spatiale, the author was supported by a European Community Marie Curie Fellowship.

Contents

Notations and acronyms

Atmospheric windows

Y band	from 0.95 to 1.1 μm
J band	from 1.15 to 1.4 μm
H band	from 1.5 to 1.8 μm
K band	from 2.0 to 2.4 μm
N band	from 8 to 13 μm

Units

arcmin (or $'$)	minute of arc (2.90888×10^{-4} radian)
arcsec (or $''$)	second of arc (4.848137×10^{-6} radian)
mas	milli-arcsec (10^{-3} arcsec)
AU	Astronomical Unit (1.495978×10^{11} m)
pc	parsec (3.085678×10^{17} m)
kpc	kiloparsec (1000 parsec)
K	degree Kelvin
micron	abbreviation of micrometer (μm)
sccm	standard cubic centimeters per minute
Myr	Mega-years (one million years)
Gyr	Giga-years (one billion years)
zodi	density unit for dust disks, equivalent to the solar zodiacal disk

Notations

$M_\odot$	Mass of the Sun (1.98892×10^{30} kg)
$R_\odot$	Radius of the Sun (6.96×10^{8} m)
$L_\odot$	Luminosity of the Sun (3.846×10^{26} W)
M_J	Mass of Jupiter (1.8987×10^{27} kg)
R_J	Radius of Jupiter (7.1492×10^{7} m)
$M_\oplus$	Mass of the Earth (5.97370×10^{24} kg)
$R_\oplus$	Radius of the Earth (6.37814×10^{6} m)
$M_\star$, $T_\star$, $L_\star$	Mass, effective temperature and luminosity of a star
M_{pl}, T_{pl}, L_{pl}	Mass, effective temperature and luminosity of a planet
$\theta_\star$	Angular radius of a star
a	Semimajor orbital axis of a planet
b	Interferometric baseline length
R, N	Rejection ratio (or radius), null depth
R_λ	Spectral resolution

Acronyms

1D, 2D, 3D	One-Dimension(al), Two-Dimension(al), Three-Dimension(al)
4QZOG	Four-Quadrant Zero-Order Grating
ADONIS	ADaptive Optics Near Infrared System (ESO)
AGN	Active Galactic Nucleus
AGPM	Annular Groove Phase Mask
AIC	Achromatic Interfero-Coronagraph
ALADDIN	Antarctic L-band Astrophysics Discovery Demonstrator for Interferometric Nulling
AO	Adaptive Optics
APS	Achromatic Phase Shifter
AR	Anti-reflective
AT	Auxiliary Telescope (VLTI)
BBAR	BroadBand Anti-Reflective
BLINC	BracewelL Infrared Nulling Cryostat (for the MMT)
BLR	Broad emission Line Region
CCD	Charge Coupled Device
CEA	Commissariat à l'Energie Atomique (France)
CFHT	Canada France Hawaii Telescope
CGH	Computer Generated Hologram
CHARA	Center for High Angular Resolution Astronomy (Georgia State University)
CNES	Centre National d'Etudes Spatiales
CONICA	COronagraphic Near Infrared CAmera
CVD	Chemical Vapor Deposition
Darwin	Not an acronym—infrared space interferometer (ESA project)
DOE	Diffractive Optical Element
DPMC	Disk Phase Mask Coronagraph
DZPMC	Dual Zone Phase Mask Coronagraph
EGP	Extrasolar Giant Planet
ELT	Extremely Large Telescope
EMT	Effective Medium Theory
ESA	European Space Agency
ESO	European Southern Observatory
FOV	Field Of View
FQPM	Four-Quadrant Phase Mask
GENIE	Ground-based European Nulling Interferometry Experiment (for the VLTI)
HST	Hubble Space Telescope
HWP	HalfWave Plate
IAGL	Institut d'Astrophysique et de Géophysique de Liège
IAS	Institut d'Astrophysique Spatiale
ICP	Inductively Coupled Plasma
IONIC	Integrated Optics Near-infrared Interferometric Camera
IR	Infrared
IRAS	Infra-Red Astronomical Satellite
ISM	InterStellar Medium
ISO	Infrared Space Observatory
JPL	Jet Propulsion Laboratory
JWST	James Webb Space Telescope
KIN	Keck Interferometer Nuller
LAOG	Laboratoire d'Astrophysique de l'Observatoire de Grenoble
LBT	Large Binocular Telescope

Acronyms

LESIA	Laboratoire d'Etudes Spatiales et d'Instrumentations en Astrophysique
LETI	Laboratoire d'Electronique de Technologie de l'Information (Grenoble)
MIDI	Mid-InfrareD Instrument (VLTI)
MIRI	Mid-IR Instrument (JWST)
MMT	Multiple Mirror Telescope
NA	Numerical Aperture
NASA	National Aeronautic and Space Administration
NACO	NAOS-CONICA
NAOS	Nasmyth Adaptive Optics System
NIL	Nulling Interferometer for the LBT
NIREST	Nulling Infra-Red survey of Exo-Systems for TPF
NOMIC	Nulling Optimized Mid-Infrared Camera
OPD	Optical Path Difference (or Delay)
Pegase	Not an acronym—infrared space interferometer (CNES)
PKC	Phase Knife Coronagraph
PMMA	PolyMethylMethacrylate
PSD	Power Spectral Density
PSF	Point-Spread Function
QSO	Quasi-Stellar Object
QWP	QuarterWave Plate
RCWA	Rigorous Coupled-Wave Analysis
RIE	Reactive Ion Etching
rms	root mean square
RPBE	Reactive Plasma-Beam Etching
RV	Radial Velocity
SEE-COAST	Super Earth Explorer-Coronagraphic Off-Axis Space Telescope
SED	Spectral Energy Distribution
SEM	Scanning Electron Microscope
SDI	Simultaneous Differential Imaging
SNR	Signal-to-Noise Ratio
SPHERE	Spectro-Polarimetric High-contrast Exoplanet REsearch (formerly VLT-PF)
TE	Transverse Electric
TM	Transverse Magnetic
TIR	Total Internal Reflection
TIRG APS	Total Internal Reflection Grating APS
TIS	Total Integrated Scattering
TPF-C	Terrestrial Planet Finder-Coronagraph (NASA project)
TPF-I	Terrestrial Planet Finder-Interferometer (NASA project)
UT	VLT Unit Telescope
VINCI	VLT INterferometer Commissionning Instrument
VLTI	Very Large Telescope Interferometer (ESO)
VLT-PF	Very Large Telescope-Planet Finder (recently renamed SPHERE)
VSI	VLTI Spectro-Imager
WFE	WaveFront error
WGP	Wire-Grid Polarizer
XAO	eXtreme Adaptive Optics
YSO	Young Stellar Object
ZOG	Zero-Order Grating (subwavelength grating)

Introduction

Astronomy and optics have a long tradition of cooperation, dating back at least to the invention of the first refractor by Galileo. Significant breakthroughs in astronomy, and later in optical astrophysics, i.e., observational astrophysics from the ultraviolet to the thermal infrared, have been linked to some major advances in optics. Indeed, theoretical astrophysics has to rely on or be confronted with observations. For a long time, this strong link has been based mostly on the design of telescopes. With the advent of photographic plates and, more recently, of modern detectors (CCD), the field of focal instrumentation has grown enormously, thanks in particular to the emergence of new concepts in optics.

Astrophysics is now facing a quantum leap in the understanding of major problems including the evolution of the very early universe, the presence of extrasolar planets and exobiology. At the eleventh anniversary of the discovery of the first extrasolar planet around a solar-type star, at a time where about 200 planets have been discovered outside our own system, daunting questions about their formation, evolution, and for a few of them, their habitability, are more than ever posed. These interrogations about our origins have triggered the emergence of new technological concepts aiming at tackling the fantastic observational challenges and a strong will for pushing existing technologies to their limit. The most recent advances in optics have indeed allowed the frontier set by the atmospheric turbulence and the finite size of apertures to be broken. In fact, it is only very recently that technological breakthroughs in adaptive optics and interferometry have succeeded in shedding new light on many fields in high angular resolution astrophysics and their future is most promising.

Firstly, in the next decade, the tools of astrophysicists in optics will move for ground-based facilities from the 8-10 meter class telescopes to decametric or even hectometric optical telescopes. The design of these telescopes and their optical train up to the focal plane instruments represents an extraordinary challenge. With some very recent exceptions, current 8-meter class telescopes and their associated instrumentation are extrapolated from the generation of 4-meter telescopes, sometimes leading to cumbersome instruments. Further extrapolation of classical techniques will not work for next-generation extremely-large telescopes (ELTs). Old and known limitations will also be exacerbated, like mirror polishing or simply the influence of the atmosphere. Adaptive optics, which can provide diffraction limited images at the focus of current telescopes, sharper by one to two orders of magnitude than images limited by the turbulence, will not be easily extrapolated towards larger apertures since this technique relies on critical components such as wavefront sensors and deformable mirrors, revealing very bulky with traditional technologies.

Secondly, and parallel to the emergence of ELTs, next generation post-VLTI interferometers are being considered. Dating back to the beginning of the last century (Michelson 1920), interferometry is not a new topic in astrophysics but the big challenge for the next decade is to develop kilometric arrays of telescopes.

These ELTs and kilometric interferometers will absolutely necessitate radical changes in the techniques. For that, evolution is underway: optical fibers, integrated optics, MEMS, subwave-

length gratings, etc. The interest for micro-optics in astrophysics grows exponentially every day. Apart from the miniaturization aspects, such an attraction is comprehensible. Historically, one of the needs of micro-structured components comes from the space environment. Indeed, in order to be "spatializable", the components have to be lightweight, operate at cycling temperature when they are in orbit and in this context, monolithic micro-components provide an incomparable stability with respect to classical bulk optics. Alternative solutions are now emerging but at the same time involve a technological breakthrough. None of the specific techniques is unique but the combination or synergy makes it very complex.

Micro-optics manufacturing techniques are inherited from the micro-electronics industry and given the speed at which the fabrication technologies are currently developing in this prolific sector, more accurate control of more and more parameters will be possible every year, opening the door to rapid substantial performance improvements. Compared to micro-electronics industry, the processes are somehow similar or derived but the materials are generally different. A process optimization taking into account the knowledge of physical and chemical properties is required. A long learning phase of R&D with industrial supports is therefore also necessary.

The present thesis is dedicated to the study of a class of micro-components: the subwavelength gratings. Such gratings have the appealing property of being optically malleable. This ability is often referred to as "optical property engineering". Applications are numerous but the present work concentrates on astrophysical applications in the framework of the most demanding subject of exoplanet detection and characterization. This dissertation is organized as follows. The first part presents the scientific context and objectives of what is called "high-contrast imaging" (Chapter 1). Then, in Chapter 2, we discuss the importance of one key component in high dynamic range imaging: the achromatic phase shifter. The second part of this work concerns the presentation of subwavelength gratings from the theoretical point of view (Chapter 3) and from the practical one by showing their numerous applications (Chapter 4). The third part is then dedicated to the applications of subwavelength gratings to phase coronagraphy in Chapter 5, where we present the 4QZOG coronagraph, Chapter 6, where the AGPM coronagraph is described, and Chapter 7 where their fabrication using state-of-the-art technologies is studied. The fourth and last part of this work presents the application of subwavelength gratings in nulling interferometry, from the theoretical point of view in Chapter 8, from the practical one in Chapter 9.

Part I

Nulling the light to unveil hidden worlds

1

High contrast astrophysics

Contents

Abstract. *High contrast imaging opportunities abound in astronomy. Some of them carry such high intrinsic interest that they justify considerable effort to master the associated technical challenges. From this point of view, the last decade has seen major advances on multiple fronts, scientific as well as technical. From the scientific point of view, the discovery of the first extrasolar planets has triggered a renewal of the permanent question on the possible presence of life outside the solar system. From the technical point of view, large telescopes, space telescopes, interferometers, adaptive optics, micro-optics are all advancing, building the technological foundations for future high contrast imaging experiments and facilities that will certainly allow us to answer this ancestral question in scientific terms.*

1.1 The need for high dynamic range

Astronomy features large dynamic ranges. The ratio of the brightness of the Sun to the faintest detected galaxy is about 10^{22}, and current technology deals with that adequately. But astronomers insist on studying highly-contrasted sources which are close to one another. The need for a high dynamic range together with a high angular resolution is the ultimate challenge.

There are two main issues for high contrast imaging: one is the capacity of the detection scheme to accommodate the dynamic range, and the other is the disturbing tendency of light from a bright source to spread out because of physical limitations of the observing telescope (atmospheric turbulence, diffraction, etc.). One can distinguish three broad categories of targets with relevant high contrast imaging scenarios which, while not exhaustive, cover most of the topics of interest to astronomy: one is stars with nearby fainter stars or planets; the second is stars with fainter, diluted and extended circumstellar material; the third is active galactic nuclei (AGN) with adjacent fainter stars, disks, jets, etc.

Let us now review each type of targets in more details, with an emphasis on the very popular and exciting subject of extrasolar planets.

1.2 Extrasolar planets

The discovery of the first extrasolar planets by Wolszczan & Frail (1992) and Mayor & Queloz (1995) has engendered a dramatic explosion in the field of extrasolar planet research. Three main themes in this topic can be identified: the first one is the characterization and understanding of the planetary populations in our Galaxy (distribution, host-star properties, etc.); the second one concerns the understanding of the formation and evolution of planetary systems (e.g., accretion, migration, interaction, mass-radius relation, albedo, etc.); the third one, last but not least, consists in comparative planetology. This subject implies the characterization of the planetary atmospheres, the search for and study of biological markers, with resolved imaging and the search for intelligent life as the "ultimate" and much more distant goal.

1.2.1 Planetary population in our galaxy

The first exoplanet detected around a solar-type star (51 Peg) unveiled an unexpected new type of planets: hot Jupiters. With its projected mass of 0.468 M_J and orbital period of 4.23 days (semimajor axis of 0.052 AU), 51 Peg b has become the prototype of what is now named after it "Pegasides", i.e., planets that are heated up to about 1000 K due to the proximity to their parent star. Since 1995, 45 planets have been found with semimajor axes $a \leq 0.1$ AU, i.e., with orbital periods ≤ 10 days (see Fig. 1.1, left). Planets have also been found farther away from their host stars, with semimajor axes up to ~ 270 AU (AB Pic b). Most of them are assumed to be gaseous giants similar to the Jovian planets of our solar system, as 123 of the 200 planets found to date have a minimum mass larger than 1 M_J.

The majority of the exoplanets discovered so far, i.e., 188 out of 200, were indirectly detected by the so-called radial velocity (RV) technique in which the stellar wobble of the star moving about the center of mass of the star-planet system is measured with spectral Doppler information. Unfortunately, the RV technique only provides a measurement of $M \sin i$, where M is the mass of the planet and i the system inclination with respect to the plane of the sky. As the latter is generally unknown, the measure only gives the "projected mass" which is a lower limit. So far,

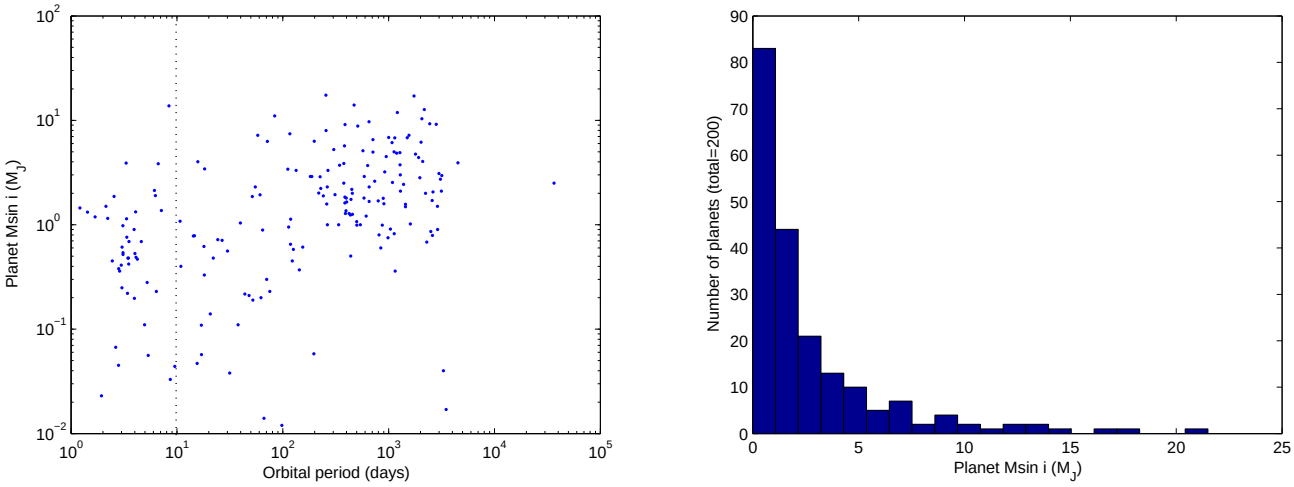

Figure 1.1: Left: exoplanet minimum mass versus orbital period. Pegasides are concentrated at the left part of this diagram. Right: number of detected exoplanet versus minimum mass. The dearth of companions with $M \sin i$ larger than 12 M_J confirms the presence of a "brown dwarf desert". Data from Schneider (2006).

about 5% of the stars monitored in RV surveys have been found to harbor at least one planet. This is a lower limit as low-mass and/or long-period objects cannot be detected with the current techniques. However, and despite the inevitable bias induced by the RV method towards high mass/short orbital period planets, remarkable statistical properties have emerged from the 200 planets detected to date:

- even with the bias mentioned here above, 40% of the detected planets have a projected mass smaller than 1 M_J and only 16% have a mass larger than 5 M_J. Within ~ 0.3 AU from the central star, there is a deficit of massive planets ($M \sin i > 4\, M_J$);
- only 6 objects have a semimajor axis smaller than 0.035 AU, which suggests that there is a pileup at this rather well-defined separation;
- according to Lineweaver & Grether (2003), at least ∼9% of Sun-like stars should have planets in the mass and orbital period ranges $M \sin i > 0.3\,\mathrm{M_J}$ and $P < 13$ years and at least ∼22% should have planets in the larger range $M \sin i > 0.1\,\mathrm{M_J}$ and $P < 60$ years. Even this larger area of the log(mass)-log(period) plane is less than 20% that occupied by our solar system, suggesting that the hypothesis that about 100% of stars have planets is consistent with both the observed exoplanet data that probe only the high-mass, close-orbiting exoplanets and with the observed frequency of circumstellar disks in both single and binary stars. If the fraction of Sun-like stars that possess planets is representative of all stars, this means that out of the $\sim$ 300 billion stars in our Galaxy, there are between $\sim$ 75 and $\sim$ 300 billion planetary systems;
- the planet-mass distribution follows a power law $dN/dM \propto M^{-1}$ (see Fig. 1.1, right), affected very little by the unknown $\sin i$. The dearth of companions with $M \sin i$ larger than 12 M_J confirms the presence of a "brown dwarf desert" at least for companions with orbital periods up to about 10 years;
- multiple planets are common (20 systems discovered so far), often in resonant orbits, and show a tendency for the inner planets to be less massive, which can be interpreted either as suppressed accretion from the outer disk or as a mere selection effect (Marcy et al. 2005);
- planet occurrence rises rapidly with stellar metallicity (see Fig. 1.2, left), i.e., the abundance

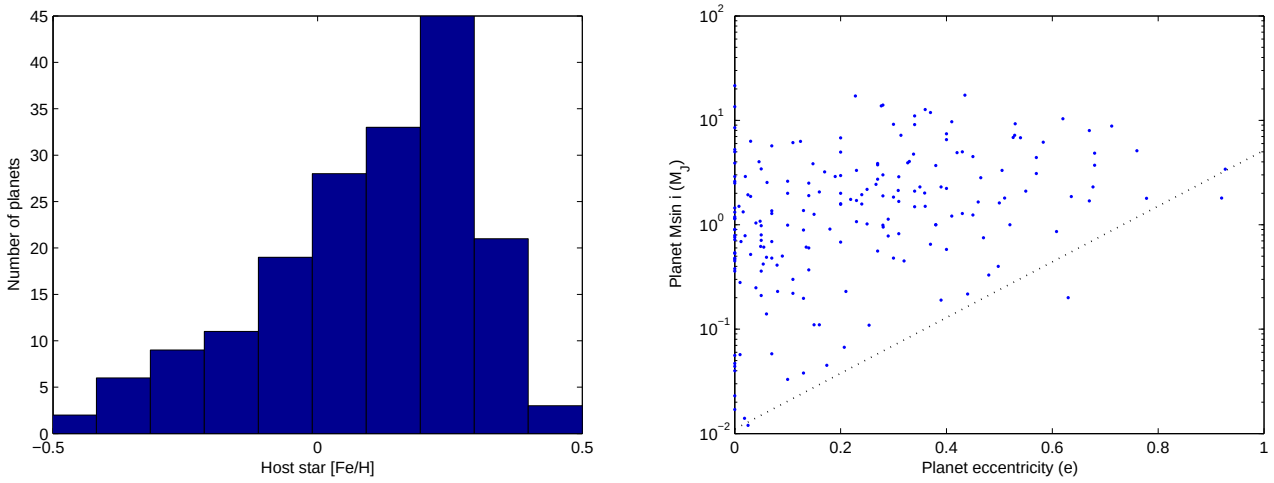

Figure 1.2: Left: number of exoplanets versus host-star metallicity. Planet occurrence rises rapidly with stellar metallicity. Right: exoplanet minimum mass versus eccentricity. Planets of highest mass tend to have higher orbital eccentricities than those of lower mass which is a puzzling trend. Data from Schneider (2006).

of heavy elements in the host star. According to Santos et al. (2005) and Marcy et al. (2005), $\sim$ 25% of the stars with twice the solar metallicity harbor a planetary mass companion, whereas this percentage decreases to below 5% for stars with the same metal content as our Sun. The favored explanation for the high metallicity of planet host stars is of course that planets form more easily in a metal-rich environment;

- apart from planets with $a < 0.05$ AU that are submitted to tidal circularization, eccentric orbits are common (see Fig. 1.2, right). This is in sharp contrast with the planets of our solar system. Fig. 1.2 (right) also shows that higher eccentricities are obtained for more massive planets, which is consistent with the trend of higher eccentricities at larger separations together with the fact that more massive planets are at larger separations, or equivalently, at longer periods (see Fig. 1.1, left) . It is to be noted that the period-eccentricity relations of exoplanets is significantly different from those of binary stars (Halbwachs et al. 2005).

This set of unexpected properties provides strong reasons to believe that exoplanets are formed by a different mechanism than low-mass companion stars and are thus putting strong constraints on the planetary formation theories as we will discuss below.

1.2.2 Planetary formation and evolution

Terrestrial planets are believed to be formed via solid body accretion of kilometer-sized objects, which themselves are produced as a result of the sedimentation and collisional growth of dust grains in the protoplanetary disk (Lissauer 1993). This theory comes from the idea already proposed in the 19th century that the Earth and the other terrestrial planets were formed from meteoritic material. But it is only in the 1960s that the theory was developed in a quantitative way by Safronov (1969), who calculated in details the different stages of terrestrial planet formation.

As far as giant planets are concerned, two theories have been proposed. According to the first theory, giant planets are formed through the collapse and fragmentation of protostellar disks. This scenario is related to that proposed by Laplace in 1796. The second theory, proposed by Cameron in 1973, is called the core accretion model (Pollack et al. 1996). A solid core is first assembled

in the same way as terrestrial planets, i.e., by accumulation of solid particles in the outer part of protoplanetary disks, beyond the "snow line" (Sasselov & Lecar 2000), where they can later on capture substantial amounts of gas from the disk in a runaway accretion process once the core has become massive enough, i.e., by reaching the so-called cross-over mass (typically at around a tenth of an Earth mass).

However, the large spread of orbital periods observed for giant extrasolar planets was not expected in any of the classical planet formation theories, shattering the understanding of the formation of planetary systems that had been patiently constructed based upon the study of a single example: our own solar system. The diversity amongst the newly-discovered systems has taught us how dangerous it can be to build theories on too small a data set. At a time when searching for life as we know it becomes possible, this comes as a serious reminder to keep our approaches as unbiased and open as possible.

Indeed, the discovery of giant planets orbiting close to their parent star completely invalidated the hypothesis that planets were formed in situ like in our planetary system. The temperature in the inner regions is actually far too high to allow the condensation of solid particles and an insufficient amount of gas is available there to form such big planets. This contradiction is now understood in the frame of migration processes (Goldreich & Tremaine 1979; Lin et al. 1996; Ward 1997). Migration involves the tidal interaction between the protoplanet and the gas in the surrounding protoplanetary disk by means of angular momentum exchange. Three types of migration mechanisms can be distinguished:

- type I migration, which applies to an embedded protoplanet of small mass for which the disk response can be modelled using linear analysis. In this case, the protoplanet is assumed to excite density waves at the Lindblad resonances[1] that propagate on both sides of its orbit. The torque exerted by these waves causes the protoplanet to migrate inwards the disk;
- type II migration, which applies when the protoplanet is massive enough to open a gap;
- runaway or type III migration, which is a new form of potentially fast migration applicable to massive disks that could be driven by coorbital torques.

Apart from planet location and migration issues, the only formation model allowing quantitative comparisons with observations is the core accretion scenario. The timescale (about 10 Myr) to form Jupiter at its present location is however uncomfortably close to the typical lifetime of protoplanetary disks, which is believed to be of the order of 1-10 Myr. It is to be noted that the first theory, based on local gravitational collapse of the protoplanetary disk, was proposed to allow for a more rapid formation of giant planets (Boss 1998). Unfortunately, this model is unable to account for the period-eccentricity relations of the newly-observed systems. A new model of giant planet formation that extends the classical core accretion model to include migration, disk evolution and gap formation can lead to much more rapid formation of giant planets, making it compatible with the typical disk lifetimes inferred from observations of young circumstellar disks (Alibert et al. 2005). This speed up is due to the fact that migration prevents the severe depletion of the feeding zone as observed in classical core accretion simulations. Hence, the growing planet is never isolated and it can reach the cross-over mass on a much shorter timescale.

[1] An external (e.g., induced by a planet) forcing gravitational potential $\psi_m(r,\phi)$ with azimuthal mode number m that rotates with a pattern frequency ω_p in a disk with angular velocity $\Omega(r)$ triggers a response that exchanges angular momentum with the orbit whenever, neglecting effects due to pressure, $m(\Omega(r)-\omega_p)$ is equal either 0 or $\pm\kappa$, with, for a Keplerian disk to adequate acuracy, $\kappa \equiv \Omega$ being the epicyclic frequency. The first possibility occurs when $\Omega = \omega_p$ and thus corresponds to a corotation resonance. The second possibility corresponds to an inner Lindblad resonance located inside the orbit for $\Omega = \omega_p + \kappa/m$ and an outer Lindblad resonance outside the orbit for $\Omega = \omega_p - \kappa/m$.

The bestiary of current detected exoplanets contains mainly Pegasides and Jupiter- or Saturn-like planets which are supposed to be hydrogen/helium ones, but thanks to the continuous refinements of indirect detection methods, Neptune-mass planets have also been detected (Santos et al. 2004). Over the past two years, the search for low-mass extrasolar planets has indeed led to the detection of so-called "hot Neptunes" or "super-Earths". These planets have masses 5-20 times larger than the Earth and are mainly found on close-in orbits with periods of 2-15 days. Rivera et al. (2005) detected a planet as light as 7.5 $M_\oplus$ but the palm goes to the 5.5 $M_\oplus$ planet, OGLE-2005-BLG-390Lb (Beaulieu et al. 2006), which has been revealed by a microlensing event[2]. The presence of this super-Earth rotating at about 2.6 AU from an M star suggests that such cool, intermediate-mass planets may be more common than gas giant planets, as predicted by the core accretion theory. Very recently, a remarkable system of three Neptune-mass planets with periods of 8.67, 31.6 and 197 days, has been detected orbiting the nearby star HD 69830 (Lovis et al. 2006). Theoretical calculations favor a mainly rocky composition for both inner planets, while the outer planet probably has a significant gaseous envelope surrounding its rocky/icy core; the outer planet orbits within the habitable zone of this star.

These detected intermediate-mass planets pose the question of their possible internal composition. In this respect, the diversity of the discovered systems so far was fully unexpected although the theoretical tools for predicting migration, for example, were available for a long time but not exploited. In order to avoid being in the same uncomfortable situation of complete surprise, why not imagine new kinds of planets provided that their existence is physically plausible ?

Ocean Planets (see Annex F). "Ocean planets" (Léger et al. 2004) or "Volatile-rich planets" (Kuchner & Spergel 2003) are putative planets in between the rocky terrestrial planets and the gaseous giants. It seems reasonable to assume that planets resembling our Uranus and Neptune, or slightly less massive ones, may have formed in cold regions of a protoplanetary disk and migrated inwards, possibly into the habitable zone where liquid water can be present at their surface. These planets would be extremely interesting as their large radius (for a given mass, they have a density significantly lower than rocky planets) makes them rather easily detectable by transit space missions (CoRoT, Kepler) and analysable by Darwin/TPF (see Sect. 1.5.4, here below). Imaging a planet with twice the Earth radius requires an integration time 16 times shorter than an Earth analogue, for the same distance and signal-to-noise conditions.

There are many other kinds of objects one may think of. Here is another example.

Carbon Planets. Kuchner & Seager (2005) suggest that some extrasolar planets will form substantially from silicon carbide and other carbon compounds. Pulsar planets and low-mass white dwarf planets are especially good candidate members of this new class for planets, but these objects could also conceivably form around stars like the Sun. This planet formation pathway requires only a factor of two of local enhancement of the protoplanetary disk's C/O ratio above solar's, a condition that pileups of carbonaceous grains may create in ordinary protoplanetary disks. Hot, Neptune-mass carbon planets should show a significant paucity of water vapor in their spectra compared to hot planets with solar abundances. Cooler, less massive carbon planets may show hydrocarbon-rich spectra and tar-covered surfaces. The high sublimation temperatures of diamond, SiC, and other carbon compounds could protect these planets from carbon depletion at high temperatures.

[2]Gravitational microlensing consists in the light amplification of a background star as a foreground compact mass, such as a planet or star, passes very close to the line of sight of the more distant source.

1.2.3 Exoplanet characterization by imaging and the search for life

Although extremely difficult, direct imaging is the most promising method as we will now discuss. Recently, Chauvin et al. (2005a) presented deep VLT[3]/NACO[4] infrared imaging and spectroscopic observations of the brown dwarf 2MASSWJ 1207334.393254. This $\sim$ 25-M_J brown dwarf, located $\sim$ 70 pc from Earth, has been recently identified as a member of the TW Hydrae association (age $\sim$ 8 Myr). Using adaptive optics infrared wavefront sensing to acquire sharp images of the circumstellar environment, Chauvin et al. (2005a) discovered a very faint and very red object at an angular separation of 778 mas (55 AU). This discovery is considered as the first image of an exoplanet (see Fig. 1.3), obtained ten years after the detection of Peg 51b, which demonstrates the difficulty of the task. The discovery was made easy by the large distance of the companion and the relatively small contrast $\Delta m \sim 5$. The characteristics of the planet according to evolutionary models are a mass $M = 5 \pm 2 M_J$ and an effective temperature $T_{eff} = 1250 \pm 200$ K.

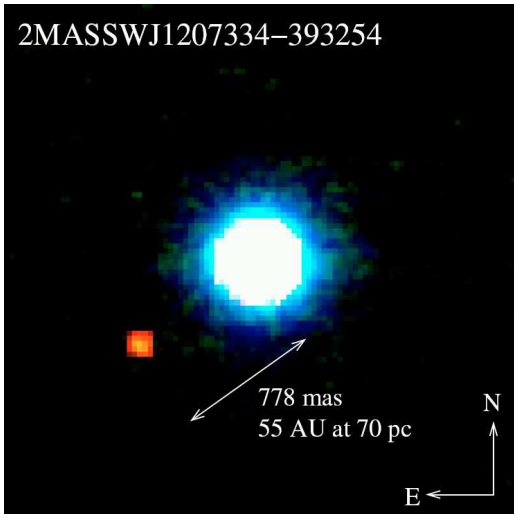

Figure 1.3: Composite image of the brown dwarf 2M1207 (Chauvin et al. 2005a). The companion appears clearly distinguishable in comparison to the color of the brown dwarf 2M1207.

Since the discovery of Chauvin et al. (2005a), three other low-mass objects have been imaged: GQ Lup b (Neuhäuser et al. 2005), AB Pic b (Chauvin et al. 2005b) and SCR 1845 b (Biller et al. 2006). The imaging of these objects was made easier since young planets and sub-stellar companions of a few Myr are generally much brighter than Gyr old sub-stellar objects because of on-going contraction and possibly accretion. It is worth noting that the planetary status of the discovered objects is not clear yet. Indeed, their mass is derived from spectral and flux observations that are injected into evolutionary models like the Baraffe et al. (2003) "COND" or the Burrows et al. (1997) "Tucson" models. Unfortunately, these models are age-dependent and unapplicable to ages below 10 Myr, which reveals critical[5] in the previous observations since the objects are supposed to be very young, with a large uncertainty about their respective ages.

[3]VLT stands for "very large telescope". This observatory of the European southern observatory (ESO) is installed in Cerro Paranal (Chile).

[4]NACO stands for NAOS-CONICA, i.e., Nasmyth adaptive optics system coronagraphic near-infrared camera. It is one of the best performing adaptive optics system in the world and it is installed on the VLT's fourth unit telescope (UT4).

[5]The initial conditions imposed in the mentioned evolutionary models generally possess a great variability that does not significantly impact the subsequent convergence of the model. The 10-Myr age is too close to the initial state and is therefore affected by its variability.

Imaging and characterization. From the theoretical point of view, when considering direct imaging of extrasolar planets, one must have in mind the three types of emission that can be detected (Chabrier et al. 2004):

- emission due to the reflection of starlight by the atmosphere or surface of the planet, which strongly depends on the star-planet distance and is modulated by the phase angle. The reflected-light luminosity L_{pl}^{refl} is given by

$$L_{pl}^{refl}(t) = \frac{A_{pl}}{4} \times \left(\frac{R_{pl}}{a}\right)^2 \times L_* \times \phi(t) \tag{1.1}$$

  with L_* the luminosity of the star, A_{pl} the planet Bond albedo, R_{pl} the planet radius, a the distance of the planet to the star, and $\phi(P, i, e, \omega, t)$ the orbital phase factor. More light is reflected at wavelengths where the star is bright, i.e., in the visible. It depends on the planet radius rather than its mass, and it may be strongly polarized (Stam et al. 2004). Barman et al. (2001), Sudarsky et al. (2000, 2003) and Burrows et al. (2004), for example, describe in details the physical processes taking place in the atmosphere of irradiated giant planets (see Fig. 1.4);

- the thermal emission of the planet, heated by the star at an equilibrium temperature

$$T_{eq} = T_* \times \sqrt{\frac{R_*}{2a}}(1 - A_{pl})^{1/4}, \tag{1.2}$$

  which is given by

$$L_{pl}^{eq} = 4\pi R_{pl}^2 \times \sigma T_{eq}^4 \tag{1.3}$$

  with T_* the temperature of the star, and R_* its radius. This thermal emission strongly depends on the star spectral type. It is also characterized by spectral features, mostly due to CH_4 and H_2O absorption bands (Fig. 1.4), commonly found in spectra of Saturn mass objects up to brown dwarfs ($0.3 - 30\,M_J$). Synthetic spectra in the thermal infrared regime can be found in Allard et al. (2001) and Baraffe et al. (2003), for instance;

- the intrinsic emission of the planets, due to residual cooling after their formation

$$L_{pl}^{eff} = 4\pi R_p^2 \times \sigma T_{eff}^4 \tag{1.4}$$

  where T_{eff} is the intrinsic effective temperature of the planet. It decreases with decreasing planet mass and system age. This contribution to emission is independent of the orbital distance since the planet is self-luminous. Due to the planet effective temperature (100−1000 K), the intrinsic emission increases towards the mid-infrared.

In Fig. 1.5, we show the comparison of the black body flux ratio for the Earth and Jupiter with respect to the Sun, from the visible to the thermal infrared wavelengths. The difficulty of high contrast imaging of exoplanets clearly appears on these diagrams. Indeed, the flux reflected by Earth-like planets is not expected to represent more than a few 10^{-10} of the stellar flux in the visible range, while its thermal emission should amount to a few 10^{-7} of the stellar flux around 10 μm, where it peaks.

Let us now review the planet characteristics that would be accessible by imaging, assuming that the contrast problem is solved. One of the most obvious observable accessible by imaging is the orbit of the detected planet. Two orbital positions are in principle sufficient to determine the orbital parameters P, i, ω, Ω, T_o and e. However, a third observation is necessary to unambiguously link the object gravitationally to its host star. It is to be noted that the number of observations

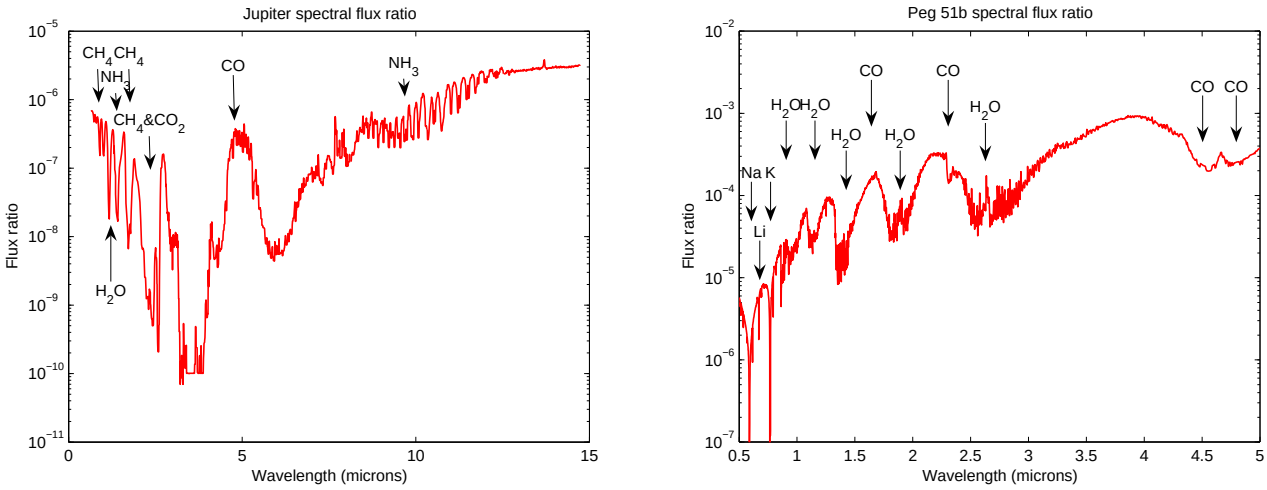

Figure 1.4: Left: Jupiter spectral contrast, dominated by methane features notably at 1.65 μm. Right: Peg 51b (Hot Jupiter) expected spectral contrast, very different from that of Jupiter.

can be reduced from three to two in the reflected light case because of the possible determination of the orbital phase factor $\phi(P, i, e, \omega, t)$. A second observable is the planet radius that can be constrained to a lower limit in the reflected light case thanks to Eq. 1.1 and from the fact that the albedo has an upper limit of 1. It is worth noting that the thermal emission, through Eq. 1.3, gives a more precise value of the planet radius. The temperature of the planet T_{pl} can be determined both from reflected light through Eq. 1.2 and from thermal emission through a simple black body fit. However, the thermal infrared approach is more robust since it does not depend on the independent knowledge of the albedo. The mass cannot be determined by direct imaging, in principle. Indeed, only the dynamical perturbation of the star motion by the planet is likely to reveal its mass. Nevertheless, the amount of Rayleigh scattering measured by low-resolution spectroscopy in the visible is representative of the density of the atmosphere. From the latter, one can approximately deduce the mass of the planet.

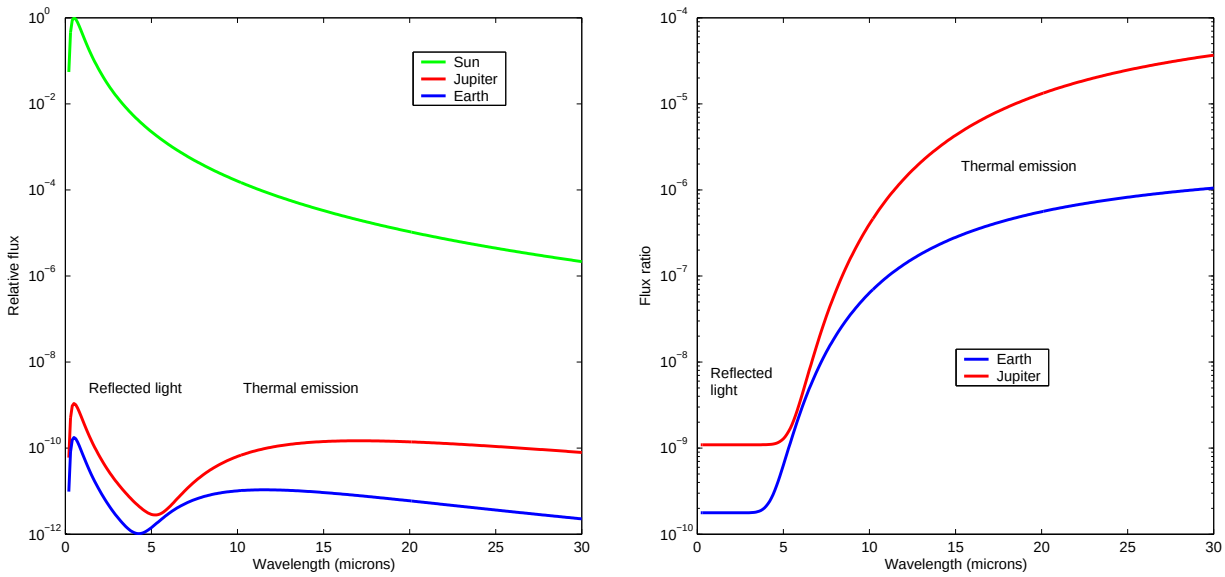

Figure 1.5: Left: comparison between Sun, Jupiter and Earth spectral flux from visible to thermal infrared wavelengths. Right: flux ratio for Jupiter and Earth with respect to the Sun.

One of the most interesting observable is the planet spectral contrast. Both visible and thermal infrared domains possess their own features (see Fig. 1.4 and Fig. 1.6). They provide complementary information about the nature of the planet surface, the chemical composition of its atmosphere, the presence of life forms, etc. Through the temporal variations of the reflected flux, after correction of orbital effects, several contributors could be disentangled among the seasonal effects, the cloud coverage variations or even the planet rotation through the succession of lands, sees, icecaps. Another advantage of visible wavelengths is that they provide a richer palette of spectral signatures than the infrared; at least when observations over a large wavelength range, and with high spectral resolution and signal-to-noise are available. Therefore we can obtain information about a planet in the visible that is not accessible in the infrared. For a planet like the present Earth, we can measure spectral features of water (H_2O), Chappuis ozone band (O_3), oxygen (O_2), Rayleigh scattering (column abundance of all gases above the surface and clouds), "red edge" (indicating plant leaves on dry land, Arnold et al. 2002), color (blue, green, red, infrared bands indicating whether the planet is similar to Venus, Earth, Mars, or Jupiter), brightness (indicating whether terrestrial or Jovian in size), and polarization (characteristic of a molecular atmosphere and of Venus-like cloud droplets). Furthermore, spectral information can be used to infer properties including the planet temperature, diameter, mass, surface gravity, and atmospheric pressure.

Life detection. Let us now come back to the most exciting theme of life detection. The first thing to consider is the definition of life. Quite apart from the philosophical point of view which is beyond the scope of the present work, we will adopt the traditional definition that life signifies "far from equilibrium organic systems, transforming stellar light into complex organisms". An admitted prerequisite in an anthropomorphic view based on our knowledge of life on Earth is the suitability of a planet for supporting life, or habitability. With the general consensus among biologists that carbon-based life requires liquid water for its self-sustaining chemical reactions, the search for habitable planets has therefore focused on identifying environments in which liquid water is stable over billions of years. This environmental constraint is primarily controlled by the star-planet separation, but is affected by factors such as planet rotation combined with atmospheric convection. For Earth-like planets orbiting main-sequence stars, the inner edge is bounded by water loss and the runaway greenhouse effect, as exemplified by the CO_2-rich atmosphere and resulting temperature of Venus. The outer boundary is determined by CO_2 condensation and runaway glaciation. These considerations result, for a Sun-like star, in an inner habitability boundary at about 0.7 AU and an outer boundary at around 1.5 AU or beyond (Kasting & Catling 2003).

Within the 1-AU habitability zone, Earth-like planets can be considered as those with masses between about 0.5 and 10 Earth masses, or equivalently, assuming Earth density, radii between 0.8 and 2.2 Earth radii. Planets below this mass in the habitable zone are likely to lose their life-supporting atmospheres because of their low gravity and lack of plate tectonics, while more massive systems are unlikely to be habitable because they are expected to build an H-He atmosphere and become gas giants. Habitability is also likely to be governed by the range of stellar types for which life has enough time to evolve, i.e., stars not more massive than spectral type A. However, even F stars have narrower continuously habitable zones because they evolve more strongly (and rapidly), while planets orbiting in the habitable zones of late K and M stars become trapped in synchronous rotation due to tidal damping, which may preclude life apart from close to the light-shadow line. Mid- to early-K and G stars may therefore be optimal for the development of life.

Some new ideas about the possibility of life developing around subgiant and red giant stars have recently been suggested (Lopez et al. 2005). Indeed, the habitable zone is expected to evolve

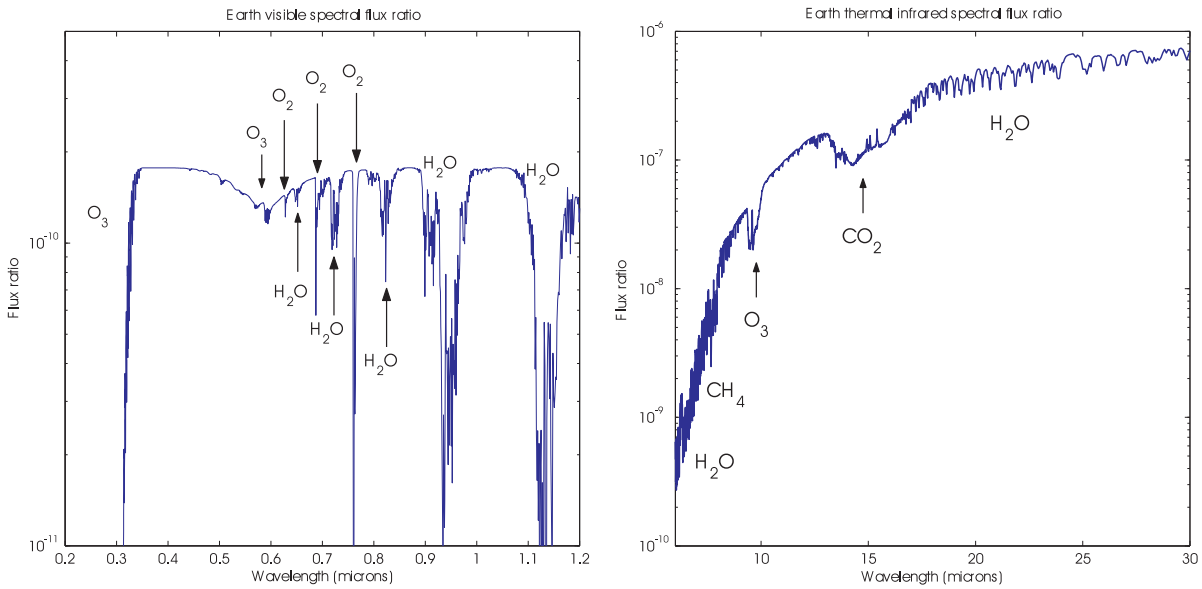

Figure 1.6: Left: Earth spectral contrast in the visible. Spectral feature of biosignatures like O_2, O_3, H_2O clearly appears. Note that the CO_2 feature is absent in the visible. Right: Earth spectral contrast in the thermal infrared where the CO_2 feature appears.

with time as the star gets older. In other words, the distance between the star and the habitable zone, as well as its width, increases with time as a consequence of stellar evolution. The habitable zone moves outwards after the star leaves the main sequence, sweeping a wider range of distances from the star until the star reaches the tip of the asymptotic giant branch. Currently there is no clear evidence as to when life actually formed on the Earth, but recent isotopic data suggest life existed at least as early as 700 Myr years after the Earth was formed. Thus, if life could form and evolve over time intervals from 500 Myr to one Gyr, there could be habitable planets with life around red giant stars. For a solar-mass star at the first stages of its post main-sequence evolution, the temporal transit of the habitable zone is estimated by Lopez et al. (2005) to be several Gyr at 2 AU and around 100 Myr at 9 AU. Under these circumstances life could develop at distances in the range $2 - 9$ AU in the environment of subgiant or giant stars, and in the far distant future in the environment of our own solar system. After a star completes its first ascent along the red giant branch and the He flash takes place, there is an additional stable period of quiescent He-core burning during which there is another opportunity for life to develop. For a solar-mass star there is an additional Gyr with a stable habitable zone in the region from 7 to 22 AU.

Detecting and characterizing an exoplanet in the habitable zone put strong constraints on the angular separation between the star and its potential Earth-like companion. For instance, it would be about 100 mas for a planet at 1 AU from a solar-type star located at 10 parsecs. Since the only known way to remotely assess the presence of life on a rocky extrasolar planet is to investigate the chemical composition of its atmosphere by means of spectroscopy, the planetary and stellar spectra have to be distinguished. This is possible only by spatially resolving the two components of the system. Indirect methods are not appropriate, except the transit method which can provide spectral information on the planetary emission without actually resolving the two components when the contrast is not too large (see, e.g. Ehrenreich et al. 2006).

Owen (1980) argued that large-scale biological activity on a telluric planet necessarily produces a large quantity of O_2. Photosynthesis builds organic molecules from CO_2 and H_2O with the help

of H^+ ions which can be provided from different sources. In the case of oxygenic bacteria on Earth, H^+ ions are provided by the photodissociation of H_2O, in which case oxygen is produced as a by-product. However, this is not the case for anoxygenic bacteria, and thus O_2 is to be considered as a possible but not a necessary by-product of life (for this signature of biological activity, as well as for any other, a key issue is that of false positives, i.e., cases where the signature is detected but there is no actual life on the planet, while the case of false negatives, when there is some life on the planet but the signature is absent, is significantly less "serious"). Indeed, the Earth's atmosphere was O_2-free until about 2 billion years ago, i.e., absent for more than 1.5 billion years after life originated. Owen (1980) also noted the possibility, quantified by Schneider (1994) based on transit measurements, of using the 760-nm band of oxygen as a spectroscopic tracer of life on another planet since, being highly reactive with reducing rocks and volcanic gases, it would disappear in a short time in the absence of a continuous production mechanism. Plate tectonics and volcanic activity provide a sink for free O_2, and are the result of internal planet heating by radioactive uranium and of silicate fluidity, both of which are expected to be generic whenever the mass of the planet is sufficient and when liquid water is present. For sufficiently small planet masses, volcanic activity disappears some time after planet formation, as do the associated oxygen sinks. O_3 is itself a tracer of O_2, with a prominent spectral signature at 9.6 μm in the infrared. These considerations are motivating the development of infrared space interferometers for the study of molecular species such as H_2O at 6-8 μm, CH_4 at 7.7 μm, O_3 at 9.6 μm, CO_2 at 15 μm and H_2O at 18 μm (Angel et al. 1986). Higher resolution studies might reveal the presence of CH_4, its presence on Earth resulting from a balance between anaerobic decomposition of organic matter and its interaction with atmospheric oxygen; its co-existence with O_2 being highly outside equilibrium, it could be a strong evidence for the existence of life.

The possibility that O_2 and O_3 are not unambiguous identifications of Earth-like biology, but rather a result of abiotic processes, has been considered in detail by Léger et al. (1999) and Selsis et al. (2002). They considered various production processes such as abiotic photodissociation of CO_2 and H_2O followed by the preferential escape of hydrogen from the atmosphere. In addition, cometary bombardment could bring O_2 and O_3 sputtered from H_2O by energetic particles, depending on the temperature, greenhouse blanketing, and presence of volcanic activity. They concluded that a simultaneous detection of significant amounts of H_2O and O_3 in the atmosphere of a planet in the habitable zone presently stands as a criterion for large-scale photosynthetic activity on the planet. Such an activity on a planet illuminated by a star similar to the Sun, or cooler, is likely to be a significant indication that there is local biological activity, because this synthesis requires the storage of the energy of at least 2 photons (8 in the case of Earth) prior to the synthesis of organic molecules from H_2O and CO_2. This is likely to require delicate systems that have developed during a biological evolutionary process. The biosignature based on O_3 seems to be robust because no counter example has been identified. It is not the case for the biosignature based on O_2 (Selsis et al. 2002), where false positives can be encountered. This puts a hierarchy between observations that can detect O_2 and those that can detect O_3.

1.3 Circumstellar disks

Let us now come back to the planetary formation question in more details. First of all, planetary and stellar formation and evolution are closely linked. The scenario of star formation has been divided, according to different stages of evolution, in four classes from class 0 to class III. Stars are formed from a cloud of dust and gas where turbulent processes lead dense enough regions to collapse, leading to an embedded pre-stellar class 0 core. A significant portion of the left-over dust and gas spirals into a class I protostar adding to its mass (this is known as accretion). This produces a flattened disk that is rotating around the central protostar. The grains which remains in the disks are subjected to many forces and frequently collide with each other. Generally, at the same time, bipolar outflows from forming star-disk systems appear. Class II objects, or young stellar objects (YSO) comprising T Tauri and Herbig AeBe stars, see their spherical envelope dissipate which makes the central star and the accretion disk observable over the whole spectrum. The last division of this classification corresponds to the moment where the gas reservoir has dissipated and the accretion stops, leading to a class III star with planetesimals and forming planets in a more tenuous disk (e.g. β Pic, AU Mic). This disk is replenished by collisions of planetesimals and evolves into a debris disk (e.g. Fomalhaut, Vega). In the transition between class II and III objects, we find the TW Hydrae-like objects (e.g. PDS70, see here below), which see their disk currently being cleared.

1.3.1 Young stellar objects

The study of young and intermediate class II-class III objects, i.e., between 1 and 30 Myr, is of special importance to constrain the planet formation scenario. This age range indeed corresponds to the timescale of disk dissipation and planetesimal building. The most natural explanation for the presence of planets is that the growth from micron-sized dust to planetesimals is extremely efficient. Possible mechanisms for such an efficient growth are the gravitational instability of the solids themselves or due to turbulence induced by shear, or the collisional aggregation of particles. Once planetesimals grow beyond the km-size, runaway growth is thought to drive those which are far enough from the central star to planetary size. This scenario of planet formation is now currently approved for low and intermediate mass stars but is still to be proven for massive stars. An accurate study of the inner (at a few AU) and outer (up to several hundreds of AU) disk regions searching for small-scale structures is required to answer questions related to planet formation such as:

- evolution of the dust phase (size, chemistry);
- evolution of dust and gas distribution in the radial and vertical direction;
- signs for planet formation and for already formed planets (local dust concentrations revealing accretion regions around the planet, grain growth, inner clearings or gaps, large-scale spiral structures, etc.).

For that, very good spatial resolutions and dynamic range are needed both in visible and thermal infrared in order to observe both scattered light and dust re-emission and to study the geometry, chemical composition and structure of the disk. In the visible and near-IR (reflected light) for example, emission line regions and forbidden emission lines in jets, the most prominent being $H\alpha$, are accessible. For resolving dust absorption-emission, crystalline (silicate), amorphous phases and Polycyclic Aromatic Hydrocarbons (PAH) features and taking into account the kinematics involved (e.g., velocity distribution in forbidden line emission in the jets of the order of a few hundreds km/s), a spectral analysis in the infrared domain (thermal emission) is necessary (6-12 μm and 18-

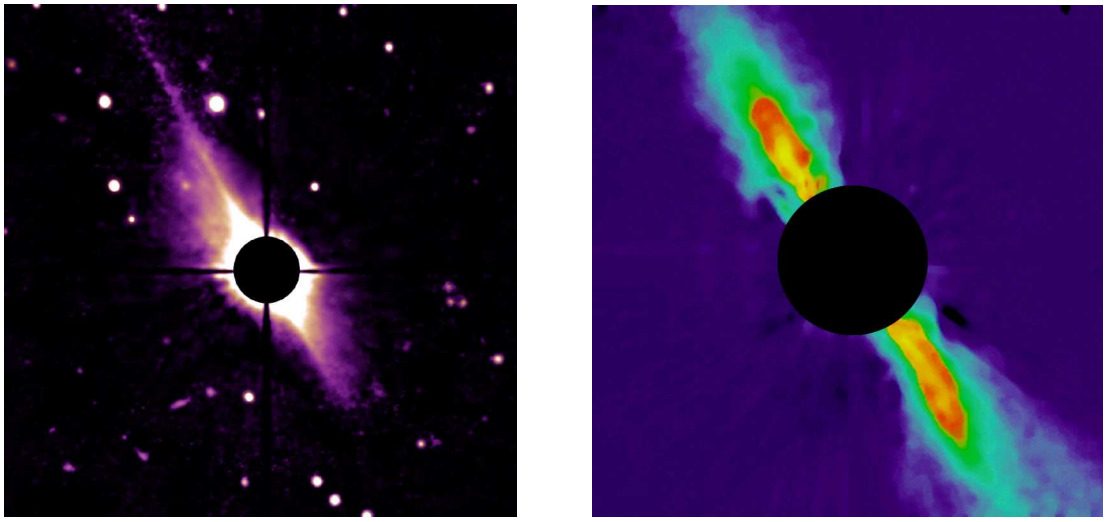

Figure 1.7: Left: β Pic observed at 0.5 micron from Mauna Kea with University of Hawaii 2.2-meter telescope (Kalas & Jewitt 1995). Light from the star is blocked by a 6.5″ diameter opaque coronagraph inserted at the focal plane of the telescope, revealing the much fainter dust disk. Right: β Pic observed at 1.2 microns (J band) with ADONIS (ADaptive Optics Near Infrared System) at the ESO 3.6-meter telescope (Mouillet et al. 1997).

30 μm). Again, the complementarity between the two observing domains is indubitable. However, contrasts ranging between 10^{-1} and 10^{-4} are expected in the near-IR depending on the age and geometry of the disk. Imaging of large regions with a high dynamic range is therefore essential for features related to planet formation but not only. High contrast imaging is also mandatory for studying features related to stellar physics such as outflows, jets, expected structures like outbursts due to increased accretion rate, etc.

As an illustrative example, the paper presented in Sect. 1.3.3 concentrates on the high contrast imaging of a circumstellar disk around a young T Tauri star (PDS70) thanks to a modern coronagraphic tool. Note that an overview of high dynamic range detection methods will be the subject of a following section (Sect. 1.5).

1.3.2 Debris disks

The first discovery of circumstellar dust around a main-sequence star dates back to the launch of the Infra-Red Astronomical Satellite (IRAS), with which Aumann et al. (1984) quickly found Vega's far-infrared flux to be in excess with respect to the expected photospheric flux. Similar excesses were discovered soon after around a large number of main-sequence stars using photometric observations with infrared telescopes such as IRAS (e.g. Aumann 1988; Mannings & Barlow 1998) or ISO (e.g. Fajardo-Acosta et al. 1999; Laureijs et al. 2002). These studies have shown that about 10% of main-sequence stars have excess emission in the 25 μm region, with a larger occurrence around early-type stars, and that this proportion increases to about 17% at 60 μm. Infrared excesses around main-sequence stars are now commonly understood as the signature of second-generation dust grains originating from collisions between small bodies (asteroids) or from the evaporation of comets (Backman & Paresce 1993), like the zodiacal disk in our solar system. These disks are supposed to be continuously replenished by these processes as dust grains have a limited lifetime of a few Myr at most, due to the effects of collisions, radiation pressure and Poynting-Roberston drag (Dominik & Decin 2003). They have to be distinguished from the protoplanetary disks of gas and dust detected around pre main-sequence stars, which originate

from primordial interstellar medium, are optically thick and dissipate on timescales of 10 Myr, comparable to the cessation of accretion (Mamajek et al. 2004).

However, unlike the solar zodiacal disk, these *debris disks* are much more massive and composed of cold dust grains, assumed to be arranged in structures similar to the solar Kuiper Belt at several tens of AU from their parent star. This assumption was confirmed by imaging some of these debris disks, first in the visible regime where the stellar light scattered by dust particles has been evidenced around young main-sequence stars such as β Pic (Smith & Terrile 1984, see Fig. 1.7) or AU Mic (Kalas et al. 2004), and then in the sub-millimetric and millimetric regimes, where ring-like structures of cold dust were first discovered around Vega, β Pic, Fomalhaut and ϵ Eri (Holland et al. 1998; Greaves et al. 1998; Koerner et al. 2001). These structures suggest that planets may be forming or may have already formed in these systems, because they generally show planet-related features such as central clearings inside the cold dust ring or the presence of warps, clumps or rings. However, the lack of angular resolution and/or dynamic range has prevented these techniques from imaging more than a dozen of disks, and have mainly limited the investigations to the outer parts of these disks, which are not relevant in the context of future planet-finding missions like Darwin/TPF (see Sect. 1.5.4).

Knowledge of dust distribution in the first few AUs around solar-type stars is indeed currently mostly limited to the observations of the solar zodiacal cloud. The zodiacal cloud, a sparse disk of 10-100 μm diameter silicate grains, is in fact the most luminous component of the solar system after the Sun. Its optical depth is only $\sim 10^{-7}$, but its integrated emission at 10 μm is about 300 times larger than the flux of an Earth-sized planet. The presence of circumstellar disks (exozodiacal disks, or *exozodi*) around the Darwin/TPF targets may present a severe limitation to their Earth-like planet detection capabilities, as such clouds would become the main source of noise if they are more than 20 times as dense as the solar zodiacal cloud. The prevalence of exozodiacal disks around nearby solar-type stars must therefore be assessed before finalizing the design of the Darwin/TPF missions.

1.3.3 Article: *Coronagraphic imaging of three Weak-line T Tauri Stars: evidence of planetary formation around PDS70*

The following paper, accepted for publication in *Astronomy & Astrophysics*, deals with coronagraphic observations of young stars using the NAOS-CONICA adaptive optics system of the VLT and the FQPM coronagraph installed at its focus. Results are very interesting since a disk has been imaged around PDS70, showing evidences of planetary formation taking place right now. It proves the importance and utility of new generation coronagraphic tools in this context.

Astronomy & Astrophysics manuscript no. ms5232˙m
July 7, 2006

Coronagraphic imaging of three weak-line T Tauri stars: evidence of planetary formation around PDS 70 ⋆

P. Riaud[1], D. Mawet[1], O. Absil[1], A. Boccaletti[2], P. Baudoz[2], E. Herwats[1,3], and J. Surdej[1]

[1] IAGL, Université de Liège, 17 Allée du 6 Août, B-4000 Sart-Tilman, Belgium
e-mail: riaud@astro.ulg.ac.be
[2] LESIA, Observatoire de Paris-Meudon, 5 pl. Jules Janssen, 92195 Meudon, France
[3] Laboratoire d'Astrophysique de l'Observatoire de Grenoble, BP 53, 38041 Grenoble Cedex 9, France

Received 20 March 2006 / Accepted 20 June 2006

ABSTRACT

Context. High angular resolution imaging of nearby pre-main sequence stars with ages between 1 and 30 Myr can give valuable information on planet formation mechanisms. This range of ages is thought to correspond to the dissipation of the optically thick dust disk surrounding young stars and to the end of the planet formation.
Aims. This paper presents new observations of three weak-line T Tauri Stars (WTTS) of intermediate ages ranging from 7 to 16 Myr. It aims at increasing the knowledge and sample of circumstellar disks around "old" WTTS.
Methods. We observed three stars with the VLT's NAOS-CONICA adaptive optics system in coronagraphic mode. The four-quadrant phase mask coronagraph was used to improve the dynamic range (by a factor of ∼ 100) while preserving the high angular resolution (inner working angle of 0.″15).
Results. One object of our sample (PDS 70), a K5 star, exhibits a brown dwarf companion and a disk in scattered light with a surface brightness power law of $r^{-2.8}$, extending from a distance of 14 to 140 AU (assuming a stellar distance of 140 pc) and an integrated luminosity of $16.7 mJy$ in the K_s-band. The mass of the companion can be estimated to be within a range between 27 and 50 Jupiter masses with an effective temperature of $2750 \pm 100 K$. This object also shows a resolved outflow stretching up to ∼ 550 AU.
Conclusions. This newly detected circumstellar disk shows strong similarities with the disk around TW Hya, and adds to the observed population of "old" TTS surrounded by circumstellar material. Moreover, three clues of planetary formation are brought to light by this study.

Key words. Stars: individual: PDS 70, PDS 81, PDS 99 – Stars: planetary systems: protoplanetary disks – Stars: circumstellar matter – Instrumentation: adaptive optics – Methods: observational

1. Introduction

Three different periods of star formation are generally distinguished. Young stars like those in Taurus and Chamaeleon (1-3 Myr) are embedded in their cocoon emitting only in the far-infrared and millimeter wavelengths. Older stars like Vega (350 Myr) or Fomalhaut (200 Myr) show large dissipated debris disks residing between 50 and 150 AU. In between, we find objects of intermediate age like β Pic or Au Mic (20 Myr), which are surrounded by disks of gas and dust still in the process of forming planets.

The study of objects between 1 and 30 Myr is therefore of special importance to constrain the planet formation scenario. Moreover, previous works indicating that the 1 to 10 Myr period is likely to be the timescale for disk dissipation (e.g., Haisch et al. 2001; Mosqueira & Estrada 2006) have to be confirmed by further observations in this age range. When the gas reservoir is dissipated and the accretion stops, class III stars are left with planetesimals, potentially continuing planet formation in a more tenuous disk. The most natural explanation for the presence of planets is that the growth from micron-sized dust to planetesimals is extremely efficient. Once planetesimals grow beyond the km-size, runaway accretion is thought to drive those that are far enough from the central star to planetary size. The alternative scenario of gravitational instability, proposed by Boss (2002), could also be at the origin of planet formation around low and intermediate mass stars.

The main observable feature at these early formation stages is the general morphology of the disk (brightness profile, asymmetries, etc.) from which timescales for disk accretion, dissipation, and planet building can be inferred, as we will discuss. To better understand these phenomena, one needs to increase the number of observations of young and intermediate objects. A

Send offprint requests to: P. Riaud, email riaud@astro.ulg.ac.be
⋆ Based on observations obtained with NACO/FQPM at the Paranal Observatory, Chile, in ESO programs 075.C-0730(A).

classical reservoir for young stars is the TW Hydra association, which contains various PMS stars. Only three optically thick disks have been detected so far in young associations: firstly around TW Hya (Weinberger et al. 2002; Krist et al. 2000), GM Aurigae (Schneider et al. 2003), and recently around PDS 144 (Perrin et al. 2006). Other young stellar associations exist in our neighbourhood (< 150 pc), such as Centaurus, Lupus, or Ophiuchus, and their observation is actually far from being completed.

In this paper, we present the results of coronagraphic observations of three PMS stars (T Tauri) with ages comprised between 7 and 16 Myr (in Scorpius, Corona Australis, and Centaurus). These observations were performed in the near-infrared K_s-band with the VLT's NACO adaptive optics system during 3 nights from the 22th to the 24th of June 2005. One star of our sample (PDS 70, a K5 star) presents a large disk and a jet-like structure. The detection of the disk in scattered light was possible thanks to the four-quadrant phase-mask (FQPM) coronagraph (Rouan et al. 2000; Riaud et al. 2001). This new generation coronagraphic device allows both good stellar extinction and high angular resolution imaging.

In Sect. 2, we describe the targets, the data analysis procedure, and the associated observational artefacts, taking the effect of the FQPM coronagraph into account. In Sect. 3, we present the K_s-band observation results. Section 4 is then dedicated to the discussion of the general properties of the PDS 70 disk. Two complementary numerical models are introduced for that purpose. They show that the presence of a young cold dust disk under dissipation reproduces the observed disk characteristics and sheds new light on former thermal infrared data.

2. Observations and data analysis

2.1. Observations

Observations were performed with the VLT's NAOS-CONICA adaptive optics system (NACO) in the coronagraphic mode, using the FQPM coronagraph. This phase mask coronagraph uses a four quadrant π phase-shift distribution in the focal plane to provide an efficient destructive interference of the on-axis star. The FQPM coronagraph has been validated on a test bench in monochromatic light (Riaud et al. 2003) showing peak attenuation of $\sim 10^5$, and recently, in polychromatic light between 500 to 900 nm, (Mawet et al. 2006) with a peak attenuation of ~ 750. A monochromatic device manufactured on an infrasil substrate has been installed and commissionned on the NACO instrument (Boccaletti et al. 2004). Under good seeing conditions (< 0.″8), a peak attenuation of about 10-30 is routinely obtained, for an inner working angle of 0.″15.

A sample of young stellar objects were imaged during three nights (June 22 to 24, 2005), using the visible wavefront sensor of NACO (Rouan et al. 2000). All FQPM coronagraphic images were taken with the K_s filter and the S13 camera (13.27 mas/pixel). This relatively high image sampling (4 pixels per λ/d) allows a precise centering on the phase mask coronagraph, and therefore a good reference subtraction for the data analysis. The observing conditions are reported in Table 1. Two objects, PDS 70 and its associated reference star HBC 609, were observed under poor seeing conditions during the first night. For this reason, PDS 70 was re-observed during the second night, but with another reference star (TTS18). The two other WTTS were also observed with their proper reference star: PDS 81 with HIC 89529 and PDS 99 with SS300. It is to be noted that, to close the loop, AO wavefront sensing is directly performed in the visible on the target under acquisition, be it a reference star or the scientific object. It is necessary to have the same atmospheric turbulence corrections between the target and the reference star. For that, all reference stars were chosen for their similar magnitudes in the V band (for similar AO correction) and K_s-band (for similar a signal-to-noise ratio in the CONICA camera). Due to the need for similar colours, almost all of the observed objects are young stars (T Tauri). This could lead to some issues in circumstellar material detections, as the references are also likely to possess a disk. This eventuality has been carefully checked by looking at their thermal infrared excesses in the IRAS catalog (see Table 2).

The total integration time ranged from 900 to 2200s depending on the target. To calibrate time-dependent PSF variations, we acquired reference coronagraphic images ninety minutes before or after the scientific exposures at almost the same parallactic angle. To reduce drift and pupil rotation, the target star centering was checked and corrected every 60s.

2.2. Data analysis

The sum of individual short coronagraphic images is processed in the following way. A normalized "super-flat" is created by taking the median of five lamp-flats with appropriate dark frame subtractions. The NACO coronagraphic mode requires the calibration of each target star with a corresponding median sky exposure, allowing the subtraction of the background and dark contributions. The subsequent normalization of each subtracted image with the "super-flat" provides a first stage of data processing. However, as we will see, this simple treatment will not completely prevent the presence of an electronic noise due to the readout process, as well as two electronic ghosts of the star, appearing on both sides of the detector center.

Next, all images are co-added with a sub-pixel centering process. Indeed, because of the camera sampling of 13.27 mas per pixel, it has been empirically demonstrated that a sub-pixel precision of 2 mas rms is necessary to achieve an efficient coronagraphic image addition. For this operation, a two-dimensional Gaussian fit of the dark center of the FQPM (FWHM of 60 mas) coronagraph provides the needed sub-pixel precision for the centering in the Fourier domain. Doing this, we obtained about $\sigma = 0.05$ pixel or 0.7 mas rms of centering error, which is 3 times better than the specification mentioned here above. In the reduced and co-added images, the diffracted starlight appears in fact much brighter than the flux of a putative disk. Fig. 1 shows the preprocessed coronagraphic image of the PDS 70 source. One can notice the classical structure of FQPM coronagraphic images: four peaks at the center surrounded by a large smooth starlight halo.

A reference star is then subtracted to increase the contrast of the coronagraphic image by minimizing the effect of the

Table 1. VLT/NACO observing log of 7 stars in the K_s-band with the FQPM coronagraph.

Target	α (J2000)	δ (J2000)	UT date	Exposure time	DIT	Seeing	$\tau_0^*(ms)$	Airmass	Strehl	Astigmatism
PDS 70	14:08:10	-41:23:53	22/07/2005	1500 s	5s	1.″4	1.1	1.1	20% ± 3%	134° ± 2°
HBC609[ref]	15:59:16	-41:57:10	22/07/2005	900 s	5s	1.″7 – 2.″7	1-0.8	1.05	15% ± 6%	131° ± 2°
PDS 70	14:08:10	-41:23:53	23/07/2005	1140 s	5s	0.″9	1.5	1.1	32% ± 3%	130° ± 2°
TTS18[ref]	15:14:47	-42:20:14	23/07/2005	1080 s	5s	0.″9	1.6	1.1	35% ± 2%	128° ± 2°
PDS 81	16:14:08	-19:38:28	24/07/2005	2200 s	5s	1″	1.8	1.05	30% ± 9%	130° ± 2°
HIC89529[ref]	18:16:07	-18:37:03	24/07/2005	2200 s	3s	1″	1.8	1.1	35% ± 8%	131° ± 2°
PDS 99	19:09:46	-37:04:26	24/07/2005	960 s	4s	1″	1.8	1.3	28% ± 10%	122° ± 2°
SS300[ref]	17:18:08	-38:08:27	24/07/2005	960 s	4s	1.″1	1.4	1.4	30% ± 7%	128° ± 2°

* τ_0 corresponds to the atmospheric correlation time at 0.5 μm recorded during the observation.
ref Reference stars.

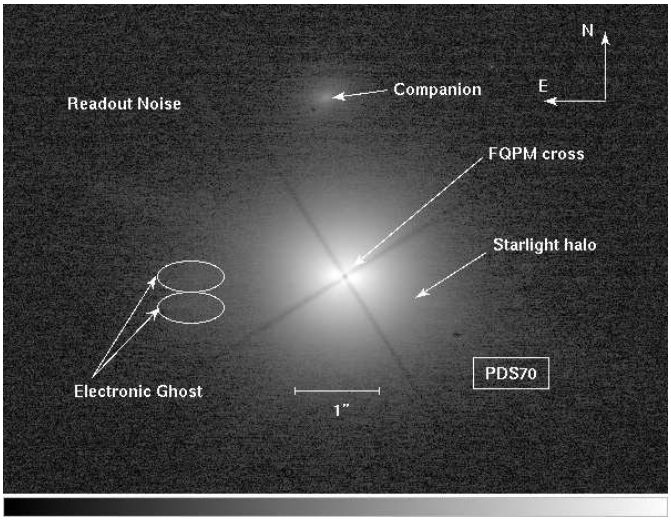

Fig. 1. Preprocessed coronagraphic image of PDS 70 obtained by the sum of 19 images (19 minutes of total exposure time). The FQPM cross and the dark hole in the center of the image are clearly visible. Four bright peaks are surrounded by a residual starlight halo (3.″3 in diameter). We also show two electronic ghosts and the noise due to the readout. The logarithmic brightness scale ranges between 0 and 2000 ADU.

diffracted starlight smoothed halo in a field of view (FOV) of 3.″5 in diameter centered on the FQPM mask. Let $I_{\rm obj}$, $I_{\rm ref}$ be the flux for the target and the reference star, respectively, in the coronagraphic image. For a proper subtraction, a classical least square procedure is applied. The corresponding figure-of-merit function ϵ corresponds to the residue of the subtraction between the target and the reference:

$$\epsilon = \sum_{i,j} \frac{|I_{\rm obj}(i,j) - \alpha.I_{\rm ref}(i,j) - \beta|^2 \times S(i,j)}{\sqrt{I_{\rm obj}(i,j)}} \quad (1)$$

The minimization parameter α is the scale intensity factor for the reference image. β corresponds to the background offset between the reference and target images. The latter is determined by taking the median value of the image parts where no significant signal is detected. The S function is a pixel-varing function used to mitigate the image over-subtraction issue by increasing the weight of pixels with large deviation relative to the shot noise. For example, this parameter is set to 1 if the subtraction in the pixel (i, j) is lower than three time the shot noise $\sqrt{I_{\rm obj}}$, otherwise it is set to 10:

$$S = \begin{cases} 10 & \text{if } |(I_{\rm obj} - \alpha.I_{\rm ref} - \beta)| > 3\sqrt{I_{\rm obj}} \\ 1 & \text{if } |(I_{\rm obj} - \alpha.I_{\rm ref} - \beta)| < 3\sqrt{I_{\rm obj}} \end{cases} \quad (2)$$

In fact, tests have been carried out with the parameter S ranging from 1 to 100, and 10 is the retained trade-off value. It is to be noted that if the reference star possesses circumstellar features, it would create a false over-subtraction issue. However, we have noted that the calculus of the optimal α that minimizes ϵ is a robust procedure little sensitive to the exact value of S and to the presence of circumstellar features around the reference.

Finally, it can be convenient to know the flux level in the pixel (i, j) relative to the azimuthally median value in a crown at the distance $r = \sqrt{(i^2 + j^2)}$. For that, an azimuthally averaged profile is removed from the reduced data to further enhance the contrast of imperfectly circular circumstellar features. Before this operation, it is necessary to check the image to detect the presence of perfectly circular circumstellar features. However, it is to be noted that this operation can give a negative value for data below the median profile.

2.3. Residual noise

The dominant source of residuals after subtraction is the speckle noise due to the variation of the turbulence conditions between the science object and the reference star. The main effect is that the resulting Strehl ratio after adaptive optics correction is different for both stars. Moreover, due to the relatively faint V magnitude of our targets, the performance of the adaptive optics system is limited by the shot noise and therefore provides incomplete correction of low-order aberrations like astigmatism (Z_5, Z_6), coma (Z_7, Z_8), or trifoil (Z_9, Z_{10}). Indeed, all coronagraphic images present a strong residual astigmatism that is oriented in the East direction at about 130° (see Table 1), leading, after subtraction, to an important speckle pattern oriented according to the astigmatism mismatch between the target and the reference stars. Therefore, a calibration of

the astigmatism for all stars has been performed on the image taken without the coronagraph. For that, we fit a 2-D elliptical Gaussian profile on the PSF image using the IDL task GAUSS2DFIT. For PDS 70, the presence of a companion allows us to calibrate this astigmatism directly in the coronagraphic image using the same procedure.

Table 2. Spectral and photometric characteristics of the stars.

Target	Sp	m_V	m_K	$F^*_{12\mu m}$	$3\sigma^\dagger$	$F^*_{25\mu m}$	$3\sigma^\dagger$
PDS 70	K5	12.0	8.5	270	24	430	51.6
HBC609	K8	12.0	8.6	*nd*	-	*nd*	-
TTS18	K1	11.3	9.0	*nd*	-	*nd*	-
PDS 81	M0	11.8	7.7	610	48.8	1320	158.4
HIC89529	M1?	11.3	7.5	*nd*	-	*nd*	-
PDS 99	M2	13.1	8.3	580	40.6	1430	85.8
SS300	M2?	12.1	7.4	*nd*	-	*nd*	-

* The flux in 12μm and 25μm in mJy is provided by the IRAS catalog (Neugebauer et al. 1988).
† Photometric error in mJy.
nd Not detected with IRAS.

In addition to the speckle noise, all subtracted images also exhibit various residual features:
• two electronic ghosts, which can only be subtracted at the level of the shot noise (3σ);
• some residuals due to the spider diffraction pattern;
• a negligible dither (electronic) noise due to the read-out noise of the InSb Aladdin 3 camera of CONICA.

3. Observations in the K_s-band

This section is devoted to the presentation of the observation results for our three targets. First of all, PDS 81 and PDS 99 present some bright residual speckles near the center ($r < 0\farcs5$). These residuals are in general oriented as the residual astigmatism. This speckle structure in the final image partly prevents the detection of faint disk-like structures with angular separations smaller than $0\farcs5 - 1''$.

3.1. PDS 81 (IRAS 16112-1930) in Scorpius

For PDS 81, the coronagraphic image after all data analysis presents residual circular features near the center (see Fig. 2). Their orientations strongly depend on the reference star used (astigmatism mismatch). No significant extended structure is detected.

3.2. PDS 99 (IRAS 19063-3709) in Corona Australis

For PDS 99, images were obtained with a worse seeing ($1\farcs1$) than PDS 70. The final coronagraphic image presents a residual speckle pattern within $0\farcs5$ and oriented according to reference star astigmatism (see Fig. 2 where the reference is SS300). Again, no large-scale circumstellar feature is detected.

3.3. PDS 70 (IRAS 14050-4109) in Centaurus

Owing to the better adaptive optics corrections for the PDS 70 source (see Table 1), the final image presents a smaller speckle noise than for the previous targets. Table 3 summarizes the main characteristics of the star in the near- and mid-infrared. The presence of an optically thick circumstellar disk has already been suggested by the detection of a mid-infrared excess and of a strong emission in the millimeter regime (Metchev et al. 2004). In our K_s-band coronagraphic data, a large disk feature in scattered light is detected as well as a companion. The disk orientation is $155° \pm 2.5°$ (i.e., not related to the residual astigmatism, orientated at $130° \pm 2°$).

Table 3. Main characteristics of the PDS 70 star.

Age (Myr)	$< 10^\dagger$
Temperature (K)	$4406^\dagger$
Stellar mass (solar mass)	0.82 (K5 estimation)
Stellar radius (solar radius)	$1.39^\ddagger$
Stellar luminosity (solar unit)	$0.64^\ddagger$
Stellar luminosity (bolometric in solar unit)	$0.78^\ddagger$
Distance (pc)	140
Temperature of the dust	$270^\ddagger$ / $45^\ddagger$
Dust luminosity fraction $f_d = L_{IR}/L_*$	$0.29^\dagger$ / $0.24^\ddagger$
A_v	$0.74^\dagger$ / $0.81^\ddagger$
U / B / V (mag)	14.32/13.15/12.15 †
R / I (mag)	11.35/10.58 †
J / H / K (mag)	9.55/8.82/8.54 **
N (mag)	5.49*
60 / 100 μm	0.915/2.11 Jy +

† (Gregorio-Hetem & Hetem 2002)
‡ (Metchev et al. 2004)
* (Kessler-Silacci et al. 2005)
** (Skrutskie et al. 2006)
+ (Neugebauer et al. 1988)

3.3.1. Circumstellar feature analysis

The residual speckles from the coronagraphic image of the star are particularly sensitive to instrumental drifts. To increase the robustness of the disk detection, all reference stars (HBC 609, TTS 18, HIC 89529, SS300) were subtracted from the PDS 70 image following the data analysis procedure presented here above. Then, a median image between the four subtracted images was derived. It must be noted that the speckle level of this final image can be determined by an analysis of the six cross subtractions between the four references. The result of this procedure is that the speckle noise can be estimated and compared with the median image of PDS 70. The last data reduction step, as mentioned in Sect. 2.2, consists of subtracting an azimuthally averaged profile. In the case of PDS 70, this leaves an important over-subtraction residual in the direction perpendicular to the disk. Fig. 3 shows the median frame obtained.

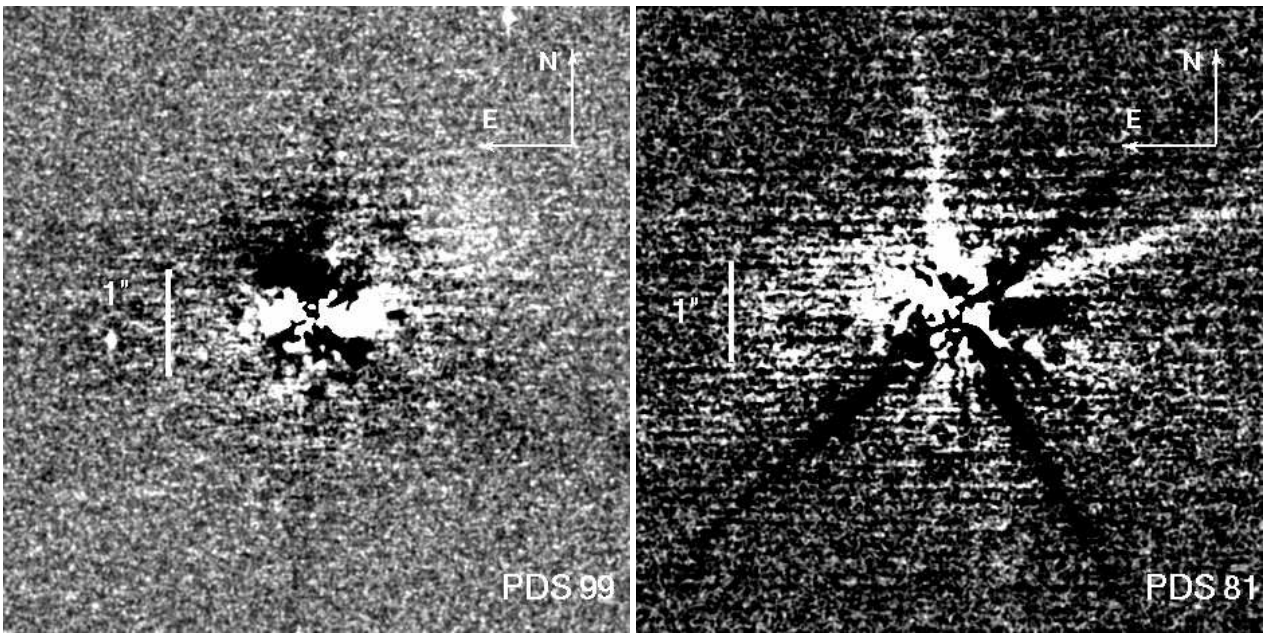

Fig. 2. Coronagraphic images: left, PDS 99 (SS300 reference star) and right, PDS 81 (HIC89529 reference star). The reference subtraction reveals bright residual speckles near the center but no large extension. For PDS 81, we can see residuals due to the differential spider rotation with respect to the reference star. The brightness scale is logarithmic.

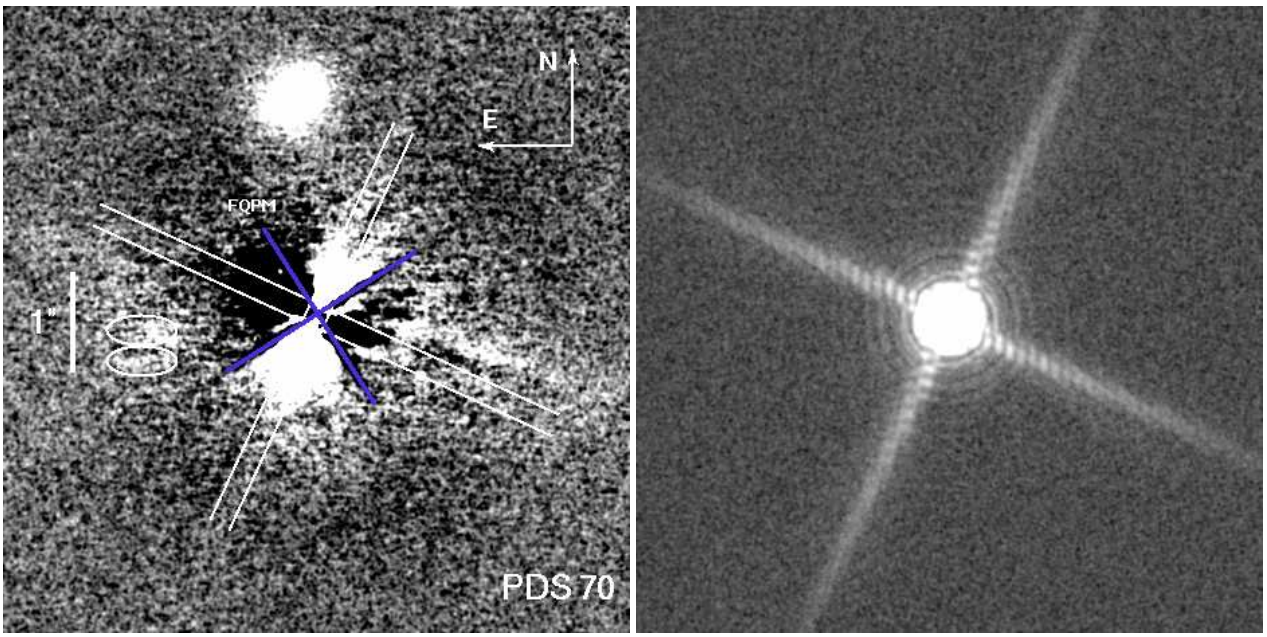

Fig. 3. Left: final image of PDS 70 star after the data analysis procedure explained in Sect. 2. The two white ellipses show the read-out ghost residuals, the white cross shows the location of the spider diffraction pattern, and the FQPM orientation is displayed in blue. Right: numerical coronagraphic model of the optical train response to calibrate the contamination due to the spider rotation (both images have the same orientation and pixel scale).

The knowledge of the spider diffracted light in the final image is mandatory to calibrate the subsequent contamination. For that, we simulated adaptive optics snapshot coronagraphic images under a low Strehl ratio ($\approx$ 30%), taking into account pupil rotation during the exposure time. Between the first and the last exposure on PDS 70, the spider rotates by 22°. The corrected atmospheric turbulence was simulated with 750 phase screens (corresponding to a 1 second exposure time). All amplitude images were then passed through by a monochromatic FQPM coronagraph with a working wavelength of 2.15μm.

The frames produced with the 750 phase screens were co-added, rescaled (bilinear approximation), shifted, and rotated to match image size, position, and inclination. Thanks to this simulation, the contribution of the spider to the diffracted light

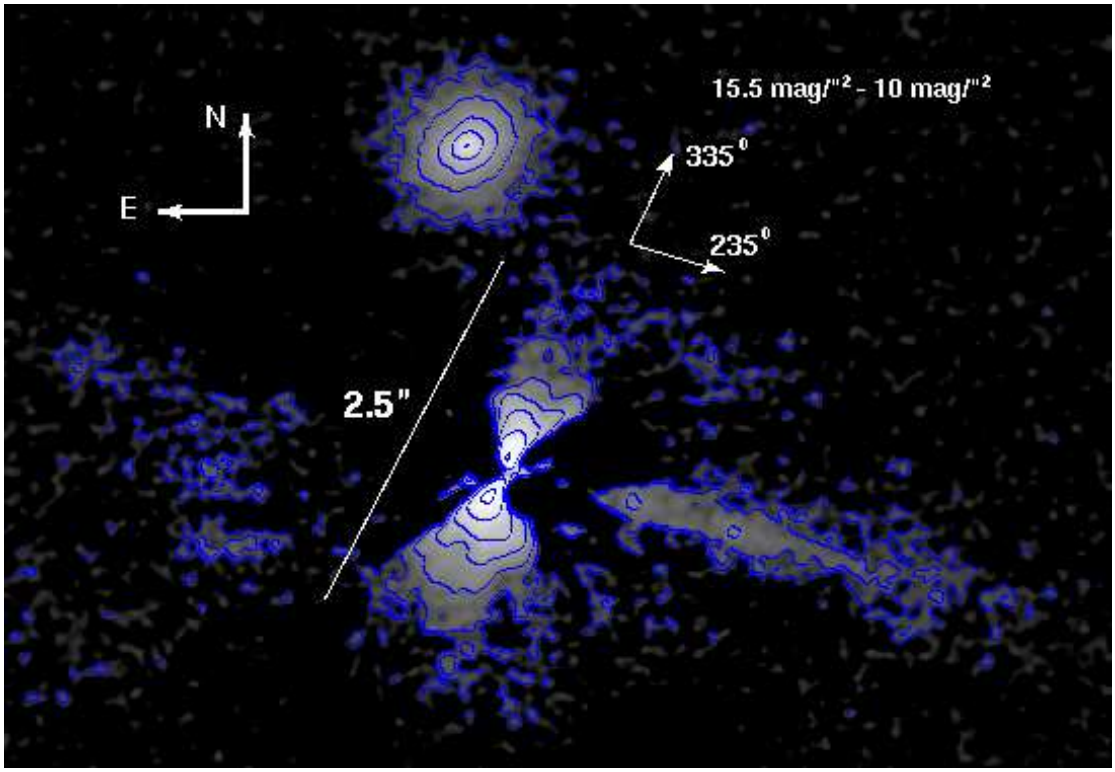

Fig. 4. Final image of PDS 70 after the complete data processing discussed in Sects. 2 and 3. To improve the detection, the final image was smoothed using a Gaussian beam with FWHM=4 pixels = λ/d. The circular disk is seen with an inclination of 62.2° ± 1.6° and a position angle of 155° ± 2.5°. Perpendicular to the disk, a jet cone is detected in the foreground with a total opening angle of 12°. In the opposite direction of the jet only some residuals are detected. This difference can be explained by the strong absorption of the disk ahead. The scattered image shows only the thin layer of the disk corresponding to $\tau < 1$. The brightness scale is logarithmic between levels of 1.125 and 282.72 photons after a proper photometric calibration on the companion. The contour plots show the disk brightness intensity with the values of 10, 11, 12, 13, 14, 15, and 15.5 $mag/arcsec^2$

was calibrated. This contribution comprises two main features:
• the large scale classical four arm spider diffraction pattern (Fig. 3, right);
• and an inner halo ($r < 0\farcs6$) slightly brighter, but still 20 to 50 times fainter than the residual flux detected around PDS 70 (Fig. 3).
The final result (Fig. 4) shows only a weak contamination due to the spider contribution (as compared with Fig. 3, where it is not removed). For example, some insignificant artifacts at angular separations greater than 1″ were removed.

3.3.2. Companion analysis

The coronagraphic image of PDS 70 shows the presence of a possible companion (PDS 70B) to the North. We do not have astrometric data to confirm its bounded character. Its K magnitude is 13.29 ± 0.02. The K_s-band photometry has been performed with the DAOPHOT package included in the IDL astrolib library. We checked the photometry accuracy with various aperture radii ranging from 10 to 80 pixels. Best results were obtained with a radius of 30 pixels for the companion and 70 pixels for the PDS 70 star. If this previously unknown companion was gravitationally linked with PDS 70 (140 pc), it would correspond to a M8 stellar type located at 301.75 ± 0.06 AU from PDS 70A. The companion is very similar to 2M1207 A (Chabrier et al. 2000) as far as luminosity and age (also 5 to 10 Myr) are concerned. We have thus used the dusty models of Chabrier et al. (2000) and Baraffe et al. (2002, 2003) to model PDS 70B. The mass of the companion can then be estimated to be in a range between 27 and 50 Jupiter masses with an effective temperature of 2750 ± 100 K, both values in favor of a brown dwarf type. The confirmation that the companion is bounded would be interesting in that it would increase the population of the ”brown dwarf desert” (see Matzner & Levin 2005).

4. Discussion

In the following, we discuss the properties of the PDS 70 disk by interpreting our near-infrared observations (Sect. 4.1). Thermal infrared photometry data and the geometrical shape knowledge acquired provide a robust Spectral Energy Distribution (SED) fitting for this object (Sect. 4.2). Finally, the disk modeling will allow us to constrain the disk mass and extension (Sect. 4.3).

4.1. Near infrared observations

Disk fitting. The high dynamic range provided by the FQPM on the NACO instrument allowed the detection of the faint disk of PDS 70. The classical method for measuring the ellipticity of circumstellar disks is to calculate the radii of isophotal contours as a function of the azimuthal angle. Unfortunately, the presence of the FQPM cross imposes that we only fit isophots in the non-attenuated zones of the disk coronagraphic image (see Fig. 4). The resulting best fit for four isophots gave $e = 0.466 \pm 0.025$ where e is the disk ellipticity. This fit takes the previously determined disk orientation of 155° ± 2.5° into account. If this ellipticity is interpreted as the result of the inclination of a circular disk ($\cos i$), the measured inclination is $i = 62.2° \pm 1.6°$. The total disk flux, measured in an annulus between 0.″05 and 1″ (7.4 to 140 AU assuming 140 pc for the distance of PDS 70) is $16.7 \pm 0.8\,mJy$ (11.49 ± 0.05 mag) in the K_s-band. The error bar is related to the photometric error, but ground-based observations under medium seeing conditions always lead to systematic uncertainties, making this value a lower bound. Indeed, the various sources of residuals add to the measured flux.

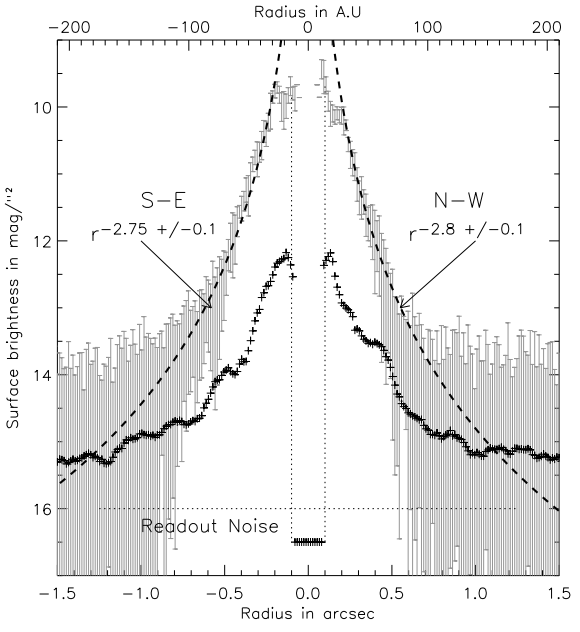

Fig. 5. Surface brightness power law analysis: the figure shows an averaged profile of the disk in a cone of ±10° around 155°. The dashed line shows the power law fit for the two opposite directions of the disk. The results are $r^{-2.8}$ (S-E) and $r^{-2.75}$ (N-W) for angular separations between 0.″21 to 0.″55 in radius. The disk profile near the center $r < 0.''21$ is flat with an oscillation around 0.″15 corresponding to distance of 20 AU. The error bars represent the minimum and the maximum values encountered in the disk. Profiles for distance $r > 1.''25$ are readout noise limited. For comparison, we show the averaged profile perpendicular to the main disk orientation, consisting mostly of star light residuals, with plus signs.

The averaged disk surface brightness in the K_s-band is presented in Fig. 5. The profile is calculated between 30 and 70 AU in a cone of ±10° around 155° and gives a fitted power law of $r^{-2.8\pm0.1}$ for the S-E direction and of $r^{-2.75\pm0.1}$ for the N-W direction. The quoted uncertainties are not due to the photon noise, but rather to the speckle noise near the mask center. The surface brightness peaks at 0.″1 (N-W) and 0.″19 (S-W) (14 and 26 AU, respectively) around 10 mag $arcsec^{-2}$ or 65.7 mJy $arcsec^{-2}$. The disk brightness falls off more sharply in the N-W direction, while presenting a weaker surface brightness. The difference of the disk extension between the S-E and N-W directions is clearly visible in Fig. 4. We also note that the disk brightness seems to flatten with an oscillation inside 30 AU. A large jet in the E-W direction (PA=235° ± 1°) is also detected with angular distances between 0.″29 and 4″ (≈ 41 to 550 AU) with a brightness of 15 mag $arcsec^{-2}$. Note that the jet orientation and the spider diffraction pattern are angularly separated by only 9°. We also notice that in the opposite direction (E), just a few faint features (15.5 to 16 mag $arcsec^{-2}$) are detected. This observation is compatible with the strong absorption by the thick disk ahead.

Potential planet detectivity. It would be interesting to estimate the detection limit of the NACO/FQPM imaging for young planets as a function of the angular separation in this case . For that, we calculated an azimuthally averaged profile perpendicular to the main disk orientation (see Fig. 5). We then estimated the 5σ contrast with respect to PDS 70 (Fig. 6). The luminosity of the putative young planet is calculated with the mass-dependent evolutionary model of Baraffe et al. (2003).

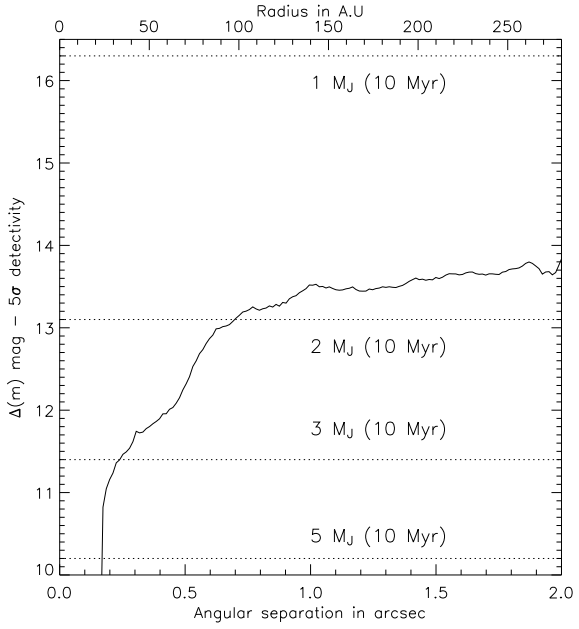

Fig. 6. 5σ contrast in Δm vs. angular separation in arcsec (bottom) and AU (top). The solid line represents the detection limit at 5σ given by the averaged profile perpendicular to the main disk orientation. The four dotted lines refers to the expected ΔM of young giant planets for 5 to 1 M_J. The detection limit is $2M_J$ at 0.″7 and $3M_J$ at 0.″3.

The K_s magnitude of the 10 Myr planet is then computed for masses ranging from 1 to 5 M_J. Finally, these planet fluxes are compared to the 5σ detectivity curve (Fig. 6). The detection limit is $2M_J$ at 0.″7 and $3M_J$ at 0.″3. These results are to be compared with the detectivity of the spectral differential imaging (SDI) method presented in Masciadri et al. (2005).

4.2. Thermal infrared data revisited

The interpretation of the dust emission in young disks relies on models describing how radiation is transferred through them. However, these models are not well constrained because of the lack of observation. Let us then derive some basic characteristics of the disk by revisiting previous mid-infrared SEDs with the new constraints provided by the coronagraphic image. Indeed, the observed disk morphology and orientation help remove the degeneracy of SED modeling.

It is known from previous observations (Metchev et al. 2004; Kessler-Silacci et al. 2005) that the PDS 70 disk presents a large infrared luminosity ($L_{IR}/L_* = 0.24 - 0.34$) similar to the TW Hydra stars. This excess corresponds to the thermal reemission of an optically thick disk in response to the central star heating.

SED fitting. Metchev and collaborators came to the conclusion that the PDS 70 circumstellar environment radiates at two different temperatures: 270 ± 10 K corresponding to an inner disk of warm dust and 45 ± 5 K for an outer component (Metchev et al. 2004) corresponding to a young cold dissipating debris disk. Unfortunately, their SED fit is not precise enough to give further valuable information concerning PDS 70. Therefore, we performed a new fit with the same data (photometry between the B band and 100 μm) plus one data point at 10.7μm obtained with the Long Wavelength Spectrometer at the W.M. Keck Observatory. The B to I photometry was taken from Gregorio-Hetem & Hetem (2002), while J to K was taken from the 2MASS catalog (Skrutskie et al. 2006). The 12/25/60/100 μm data points were taken from the IRAS catalog (Neugebauer et al. 1988). All photometric data are summarized in Table 3.

A NextGen stellar atmosphere model (Claret & Hauschildt 2003; Hauschildt et al. 1999) with $T_{eff} = 4400\,K$, $log(g) = 4$ was used to fit the stellar flux up to 7 μm. Then, up to 12 μm, a stellar spectrum and blackbody with an effective temperature of 270 K was used to model the inner optically thick disk. A simple cold component at 45K gave a strong mismatch with the far infrared IRAS photometry, although the IRAS data is quite accurate and therefore relevant for this target (see Table 2).

To remedy this mismatch, we performed a simulation of the outer disk using the DISKPIC program included in the GENIEsim package (Absil et al. 2006), which simulates the thermal emission from an optically thin debris disk. We assumed that the disk contributing to the far-infrared emission was optically thin and that the temperature varies as $r^{-0.5}$, with r the distance to the star. We then fixed the transition between the inner (optically thick) and outer disk (optically thin) at ≈ 30 AU, according to the K_s-band observation showing a discontinuity at this particular distance (≈ 0.″21, see Fig 5). Indeed, the surface brightness profile of the disk shows a different power law at short distances, where it seems to be flatter. It is to be noted that only high resolution imaging in the N band could provide the precise temperature variation of the disk.

The main result of this observationally constrained model is that a cold disk with a temperature gradient from 85 K at 30 AU to 35 K at 180 AU fits the 60 and 100 μm IRAS photometry well. However, concerning the 25 μm photometry, the simulated flux remained 2 times lower. This difference can possibly be explained by the presence of a strong amorphous silicate emission feature around $20 - 25\mu m$. This emission originates from the Mie scattering of amorphous Olivine and Pyroxene dust grains with sizes between 2 μm and 5 μm (Kessler-Silacci et al. 2006). The final result of this fit is displayed in Fig. 7. It is to be noted that the stellar atmosphere is affected by a strong reddening ($A_V = 0.78$).

These new insights on the PDS 70 SED analysis seem to confirm the beginning of the disk clearing process.

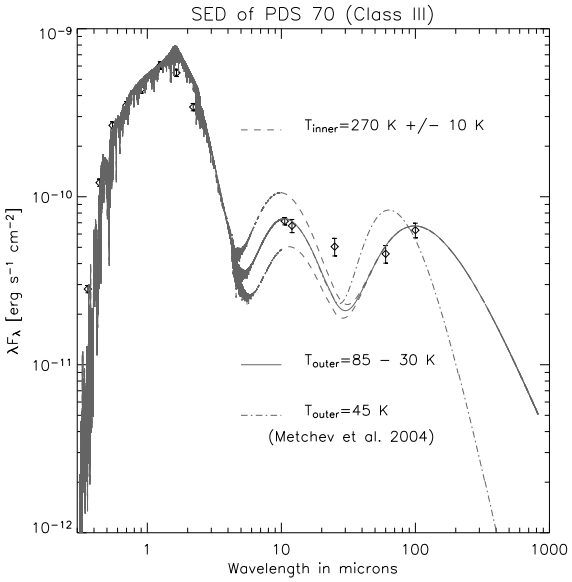

Fig. 7. Measured spectral energy distribution (SED) for PDS 70 compared to that of a reddened NextGen stellar atmosphere model with $T_{eff} = 4400\,K$, $log(g) = 4$. The dashed line includes a simple 270 K blackbody in addition; the dash-dotted line, a 270 K and a 45 K blackbody as suggested by Metchev. The continuous line includes the 270 K component plus a temperature gradient ($T \propto r^{-}0.5$) from 85 K (30 AU) to 35 K (180 AU).

Thermal infrared spectroscopy. Kessler-Silacci and collaborators (Kessler-Silacci et al. 2005) have recently presented photometry and spectra in the N band (9 to 12 μm). They observed a wide feature for PDS 70 in the 10.2-11.3 μm band and explained it by the emission of either various crystalline silicates or amorphous grains of increased sizes. However, the flux in the amorphous Olivine band ($9.8\mu m/F_{9.8}$) is 2.75 ± 0.3

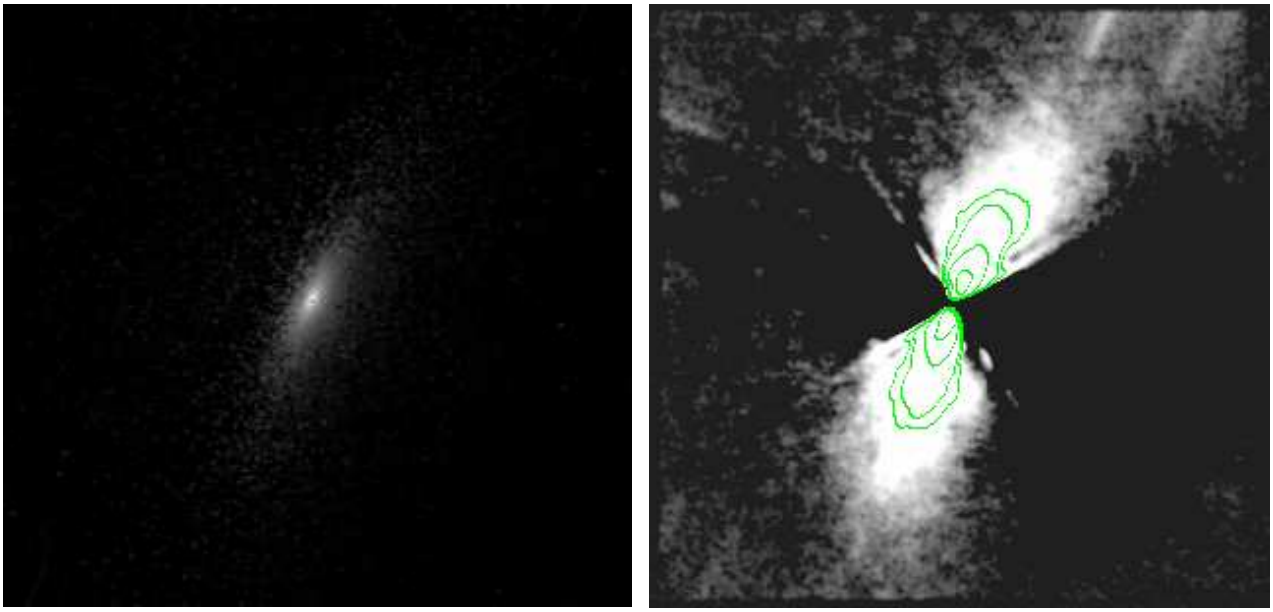

Fig. 8. Left: Numerical simulation in K_s-band (scattered light) of the PDS 70 disk without PSF convolution. Right: Same with PSF convolution and FQPM cross transmission map. For that, we created a synthetic disk and added it to the TTS18 reference star. The image was obtained using the second part of our data reduction algorithm (see Sect. 2.1). The disk orientation was set to 155° and the inclination to 62°. We also show the contour plots corresponding to 10 to 15.5 $mag/arcsec^2$ brightness scales. We choose a disk mass of $0.001 M_\odot$ without the envelope. All images are display in the logarithmic scale with the same cuts as the Fig. 4.

times that of the continuum, whereas the cristalline Forsterite one ($11.3\mu m/F_{11.3}$) is only 1.68 ± 0.18 (ratio $F_{11.3}/F_{9.8}$ = 0.61 ± 0.14). The absence of the 11.3 μm silicate cristalline feature indicates that the observed broadening is more likely related to grain growth and removal, indicating a possible planetary formation. Indeed, the formation of a cold debris disk is dynamically alleviated by the presence or the formation of giant planets that are ejecting small objects at the periphery of the disk (Moro-Martín & Malhotra 2005).

4.3. Modeling the scattered light image

To model the scattered light image, we used a 3-dimensional scattering code developed by Whitney & Hartmann (1992, 1993) to deduce some information about the disk structure. The three free parameters of the model are the full disk mass M_D, the maximum radial extension r_D, and the flare angle α_f. In addition, the model uses mixed grains with a dust size distribution $n(a) \propto a^{-p}$, where the grain size a ranges from 0.1 μm (ISM) to 1 cm (large grain in a dust disk). Since our observations do not allow us to constrain the exponent p of the power law, we have used a classical dust size distribution where $p = 3.5$.

A warm optically thick disk composed of mixed dust, gas, and large grains will naturally flare as a consequence of vertical hydrostatic equilibrium. However, the flare angle parameter does not appear to be critical for the PDS 70 disk since the coronagraphic observation shows only large disk distances ($r > 20\,AU$). In fact, our data gives only constraints for the disk mass and its radial extension. The remaining free parameters M_D and r_D are coupled by the disk density, i.e., a larger disk could reproduce the scattered light observation with a lower density. Moreover, in the visible and the near-infrared contributions, stellar radiation penetrates the disk only down to an optical depth of 1. Due to this limitation, the inferred mass will only be a lower bound (far-infrared or millimeter observations could accuratly give the actual disk mass).

Adding the geometrical constraints given by the coronagaphic image, the disk inclination is fixed to 62° and the image orientation with respect to the North is 335°. We must also consider image characteristics like sampling (1.858 AU/pixel), convolution by the PSF of PDS 70 without a coronagraphic device, the FQPM orientation, and attenuation in the NACO field of view. Taking all this into account to reproduce the observed geometry of the disk, the model requires a disk radius r_D larger than 500 AU with a total mass M_D between 0.001 to 0.002 $M_\odot$. We can compare our results for the circumstellar disk around PDS 70 star (a K5 star) with the case of the TW Hydra face-on disk (a K7 star). The same model gives an expected mass of the TW Hya disk of 0.0014 $M_\odot$ (Weinberger et al. 2002) compatible with our study, whereas the thermal infrared emission model of Calvet et al. (2002) gives a mass of 0.06 $M_\odot$. This mass discrepancy is representative of the scattered thin disk layer versus the far-infrared / millimeter deeper disk probing as explained before. The disk radius $r_D = 500\,AU$ is the minimum disk size that does not induce a large asymmetry of the disk extension, which is not observed in the coronagraphic images.

4.4. Dynamical considerations.

The transition of the disk profile around $a_L \approx 28 - 30$ AU , is likely due to a young giant planet orbiting at $a_p \approx 9.9 - 10.6$

AU in the inner disk. Indeed, circumstellar material exterior to the planet is expected to gain angular momentum from the system and to move away, while material interior to the planet loses angular momentum and migrates toward the star. A gap is expected to form if the tidal force overcomes the viscosity. The two main parameters of the tidal interaction are the ratio of the planet and the star masses, $q = M_p/M_s$ and the Reynolds number within the inner disk Re. If $q > 1/Re$ and $q^2 < 1/Re$, a gap should form with the gap edge at the outer Lindblad resonance. The Lindblad resonance can be determined easily from $\Omega_L = \omega_p m/(1+m)$, where $\omega_p = M_s^{1/2}(1+q)^{1/2}G^{1/2}a_p^{-3/2}$ is the angular frequency of the orbiting planet, G is the gravitational constant, and $m = 1$ for the first Lindblad resonance (Rice et al. 2003a,b). The ratio between the outer Lindblad resonance and planet semi-major axis becomes $a_L/a_p = 2^{3/2}$ for $m = 1$, giving $a_p \approx 10$ AU in our case.

5. Conclusion

We have discovered an optically thick dust disk around PDS 70, a young star in the Centaurus association (140 pc). The maximum disk extension observed in the coronagraphic image is around 180 AU and presents an important anisotropy in the S-E extension. Provided that it is gravitationally linked, a brown dwarf companion (27-50 M_J) with an effective temperature around 2750 K has also been detected at about 300 AU from its host star.

The disk anisotropy (S-E) can be explained either by thermal instabilities or by the dynamical effects of the brown dwarf companion. The inner disk (radial distances less than 30 AU) seems to be flatter. The shape of the SED, the dust grain properties in the thermal infrared, and dynamical considerations are in strong agreement with the formation of a young giant planet at $\sim 10AU$, which is currently clearing the inner disk by the gravitational tidal force. Indeed, the thermal infrared SED data indicate the prevalence of large amorphous features (all in favor of grain growth and planetary formation) rather than cristalline ones. Finally, our model of the optically thick disk at 270 K reproduces the scattered light observations. However, the large thermal emission in the far infrared can only be explained by the presence of a young cold debris disk with temperatures between 35 K and 85 K at large distances ($> 30AU$) with respect to the central star. This double disk model fits the PDS 70 SED very well and is corroborated by the numerical model.

Further analysis should be undergone using the J and H bands, polarimetry and SDI observations (Hartung et al. 2004) to infer the inner disk properties. It would also be interesting to image the CO map in millimeter wavelengths with an interferometric array to assess gas dissipation in the disk.

Acknowledgements. P.R. acknowledges the financial support of the University of Liège. D.M. acknowledges the financial support of the Belgian "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture". O.A. acknowledges the financial support of the "Fond National de la Recherche Scientifique". We thank J.-C. Augereau for the helpful discussions concerning infrared modeling. We thank N. Ageorge, T. Szeifert, and D. Nürnberger for their excellent assistance during the coronagraphic observations on the NACO instrument. The authors are also grateful to the anonymous referee who help improve the manuscript.

References

Absil, O., den Hartog, R., Gondoin, P., et al. 2006, A&A, 448, 787
Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 2002, A&A, 382, 563
Baraffe, I., Chabrier, G., Barman, T. S., Allard, F., & Hauschildt, P. H. 2003, A&A, 402, 701
Boccaletti, A., Riaud, P., Baudoz, P., et al. 2004, PASP, 116, 1061
Boss, A. P. 2002, ApJ, 567, L149
Calvet, N., D'Alessio, P., Hartmann, L., et al. 2002, ApJ, 568, 1008
Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. 2000, ApJ, 542, 464
Claret, A. & Hauschildt, P. H. 2003, A&A, 412, 241
Gregorio-Hetem, J. & Hetem, A. 2002, MNRAS, 336, 197
Haisch, K. E., Lada, E. A., & Lada, C. J. 2001, ApJ, 553, L153
Hartung, M., Herbst, T. M., Close, L. M., et al. 2004, A&A, 421, L17
Hauschildt, P. H., Allard, F., & Baron, E. 1999, ApJ, 512, 377
Kessler-Silacci, J., Augereau, J.-C., Dullemond, C. P., et al. 2006, ApJ, 639, 275
Kessler-Silacci, J. E., Hillenbrand, L. A., Blake, G. A., & Meyer, M. R. 2005, ApJ, 622, 404
Krist, J. E., Stapelfeldt, K. R., Ménard, F., Padgett, D. L., & Burrows, C. J. 2000, ApJ, 538, 793
Masciadri, E., Mundt, R., Henning, T., Alvarez, C., & Barrado y Navascués, D. 2005, ApJ, 625, 1004
Mawet, D., Riaud, P., Baudrand, J., et al. 2006, A&A, 448, 801
Metchev, S. A., Hillenbrand, L. A., & Meyer, M. R. 2004, ApJ, 600, 435
Moro-Martín, A. & Malhotra, R. 2005, ApJ, 633, 1150
Mosqueira, I. & Estrada, P. R. 2006, Icarus, 180, 93
Neugebauer, G., Habing, H., & Chester, T. 1988, IRAS Catalogs and Atlases Explanatory Supplement, Vol. 1 (Beichman C.)
Perrin, M. D., Duchene, G., Kalas, P., & R., G. J. 2006, ApJ, in press
Riaud, P., Boccaletti, A., Baudrand, J., & Rouan, D. 2003, PASP, 115, 712
Riaud, P., Boccaletti, A., Rouan, D., Lemarquis, F., & Labeyrie, A. 2001, PASP, 113, 1145
Rice, W. K. M., Armitage, P. J., Bonnell, I. A., et al. 2003a, MNRAS, 346, L36
Rice, W. K. M., Wood, K., Armitage, P. J., Whitney, B. A., & Bjorkman, J. E. 2003b, MNRAS, 342, 79
Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., & Labeyrie, A. 2000, PASP, 112, 1479
Schneider, G., Wood, K., Silverstone, M. D., et al. 2003, AJ, 125, 1467
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Weinberger, A. J., Becklin, E. E., Schneider, G., et al. 2002, ApJ, 566, 409
Whitney, B. A. & Hartmann, L. 1992, ApJ, 395, 529
Whitney, B. A. & Hartmann, L. 1993, ApJ, 402, 605

1.4 Extragalactic astrophysics

Until a few years ago, extragalactic astrophysics was not a predilected science topic for high contrast/high angular resolution imaging due to the limitations in sensitivity. Galaxies are rather spread out with relatively low surface brightness, hence high contrast cases do not seem to be prevalent. Typical galaxy K-band surface brightness is indeed of the order of 15-25 magnitudes per square arcsec (Brown et al. 2003). An exception is the case of active galaxies: Seyfert types and quasi-stellar objects (QSO) for example. In active galaxies, the nuclear region contains a very energetic and bright source. These so-called active galactic nuclei (AGN) produce non-stellar energy in terms of luminosity ($L = 10^8 - 10^{13} L_\odot$) and in terms of spectrum (flat and extending from radio to gamma-rays). Other remarkable characteristics are their high efficiencies of matter-energy conversion (about 0.1), the apparent compactness of the source (less than a few light days), the rapid time variabilities (a few hours) and the presence of relativistic jets. The most widely-accepted model comprises a central black hole in the supermassive ($10^6\, M_\odot$) or hypermassive ($10^9\, M_\odot$) range surrounded by an accretion disk responsible for the radiative energy and matter (outflows) release.

Observationally, the orientation of the AGN with respect to the line of sight is supposed to account for the two main classes, Seyfert 1 (broad emission lines) and Seyfert 2 (narrow lines). The dominating contribution to the Seyfert 1 signal comes from the so-called broad emission line regions (BLR), supposedly composed of small high-density gas clouds orbiting around the nucleus (ranging from a few light days to a few light years in the most luminous ones). At larger radii (1-100 pc), the nucleus is surrounded by an obscuring torus of cold gas and dust, the main contributor to the Seyfert 2 signal. Bright Seyfert nucleus magnitudes start around $K = 8$, and bright QSOs around $K = 10$, giving moderate contrast ratios which often allow imaging of the underlying galaxy from the ground or space with relatively conventional equipment, but which would surely benefit from advances in high contrast imaging techniques.

As a matter of fact, the Keck-interferometer (Keck-I), the very large telescope-interferometer (VLTI) and adaptive optics system NAOS-CONICA (NACO) both recently made major breakthroughs in the field of AGN studies. For example, Gratadour et al. (2005) observed the nucleus of NGC 1068 in the K band with NACO (Fig. 1.8), using the four-quadrant phase-mask coronagraph (FQPM, see Sect. 1.5.3). As the K-band emission is dominated by an intense compact central source, previous K-band adaptive optics images were severely affected by the limited exposure time and by scattered light in the immediate vicinity of the source. Thanks to the use of this new type of coronagraph coupled with adaptive optics, the complex dust structure near the central core is becoming observable at a resolution of 0.07 arcsec.

Wittkowski et al. (2004) used the VINCI instrument of the VLTI on the same AGN (NGC 1068) but at a smaller scale, thanks to the enhanced resolution provided by the interferometer. Their measurements support a multi-component model for the intensity distribution, where a part of the flux originates from scales smaller than 5 mas (0.4 pc), and another part from larger scales. Additional observations of NGC 1068 were obtained in the mid-infrared with MIDI at VLTI. Jaffe et al. (2004) detected warm (320 K) dust in a structure 2.1-pc thick and 3.4 pc in diameter, surrounding a smaller hot structure. This thick dust structure cannot be supported for the length of the active phase of the AGN and requires a continual replenishment from a nearby source. With the Keck interferometer (Keck-I), Swain et al. (2003) have reported an expectedly compact nucleus for NGC 4151, another nearby AGN, suggesting that the emission mainly originates from the central accretion disk.

One can therefore definitely argue that the recent results obtained at the VLT and the Keck

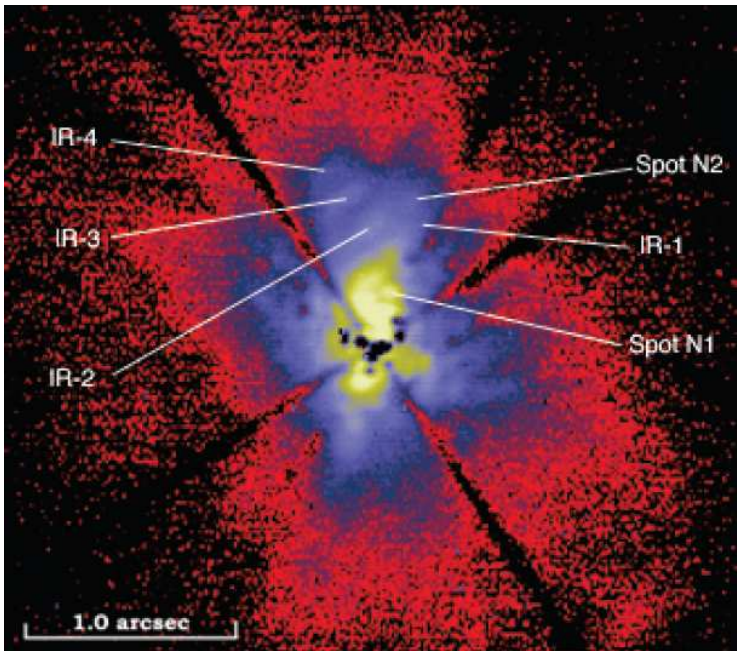

Figure 1.8: Reference subtracted K_s band coronagraphic image of NGC 1068 in a log-scale representation taken with the VLT's adaptive optics system NACO in the four-quadrant phase-mask coronagraphic mode (Gratadour et al. 2005). North of the nucleus, an elongated, bending structure, and a series of four elongated and well aligned knots can be well identified. These are thought to be tracers of shocks induced in the ISM by the passage of the jet, very close to its origin. Moreover, precise relative photometry supports the interpretation of very small dust grains transiently heated by UV photons of the central source. To the South, a new group of filamentary structures, distributed in a cone at about 150 pc from the core, is detected. They might trace the redshifted southern narrow line region, seen through the dust. Finally, on larger scale (within a radius of 300 pc) the source shows to have an overall biconical shape whose angle matches well with the bicone observed in the UV-visible.

observatories on NGC 1068 and NGC 4151 should definitely convince the community of the great potential of high dynamic range/high angular resolution imaging to make new and revolutionary science in this field. Let us then describe here below an example of extragalactic science case that would benefit from the recent advances in high dynamic range imaging.

Studying high redshift quasar host galaxies is fundamental to understand the link between quasar activity and galaxy formation in the early universe as well as their subsequent evolution up to present days. In the local universe ($z \leq 0.3$), ground- (McLeod & Rieke 1994; Taylor et al. 1996; Percival et al. 2001) and space-based (Disney et al. 1995; Schade et al. 2000; Dunlop et al. 2003; Pagani et al. 2003) imaging indicates that the majority of quasar hosts are massive galaxies dominated by the spheroidal component. These observations, together with the fact that local massive spheroids have an inactive supermassive black hole in their centers (Ferrarese 2002) tend to show that quasar activity is linked to the total mass of the system. In other words, the formation of massive galaxies and the fuelling of the inner black hole is fundamentally connected. Moreover, it is known that the cosmological evolution of the quasar population (Boyle 2001) is similar to the evolution of the star formation history (Franceschini et al. 1999).

The characterization of quasar host galaxies in the quasar peak activity redshift range ($z \sim 2 - 3$) is therefore mandatory to understand these links with the underlying question of the existence or not of the process of hierarchical merging. This model indeed predicts a significant drop in the host mass at these early stages.

In this context, the work of Falomo from the ground (Falomo et al. 2001, 2004) and Kukula from space (Kukula et al. 2001) for instance, indicate that the luminosity of host galaxies is consistent with the passive evolution of massive spheroids. However, in these studies, the main limitation of host detection is the light contamination by the extreme nucleus activity. The techniques used so far to characterize the host galaxies have not been optimal because the direct PSF (point-spread function) subtraction involves a major observational difficulty, all the more from the ground because of the uncorrected atmospheric turbulence and quasi-static aberrations (see Sect. 2.1.3). From space with the Hubble space telescope (HST), unfortunately, the quasar peak activity epoch is not attainable because of the modest aperture. The problem of PSF evaluation and subsequent subtraction have already been studied (Falomo et al. 2005; Kuhlbrodt et al. 2005). The best way to overcome or at least minimize the subtraction errors is to remove or attenuate the coherent contribution of the QSO nucleus by proper high contrast imaging techniques such as coronagraphy.

A first coronagraphic image of the nearest quasar 3C 273 has been taken in January 2003 with the ACS instrument onboard HST in the visible (Martel et al. 2003). Unfortunately, the HST/ACS is strongly limited by the large size of the Lyot mask (1.8″ in diameter) preventing detection closer than 0.9″ from the central source, and yielding a relatively low rejection factor (~ 2). This limitation of existing state-of-the-art facilities is an additional motivation for considering recent advances in high dynamic range imaging techniques.

1.5 Detection and characterization techniques

Before going into the details of high contrast direct imaging techniques, let us briefly review the detection methods that can be and that have been used in the most representative case of extrasolar planet detection (for more details, see Mawet (2002), for instance).

1.5.1 Overview

Detection methods for extrasolar planets can be broadly classified into those based on (see Fig. 1.9):

- dynamical effects (radial velocity, astrometry, or timing in the case of the pulsar planets);
- microlensing effects (astrometric or photometric);
- photometric signals (transits and direct imaging in the visible or infrared);
- miscellaneous effects (such as magnetic superflares, or radio emission).

Each method has its own strengths, and advances in each field will bring specific and often complementary discovery and diagnostic capabilities. Detection is a pre-requisite for the subsequent steps of detailed physico-chemical characterization demanded by the emerging disciplines of exoplanetology and exobiology.

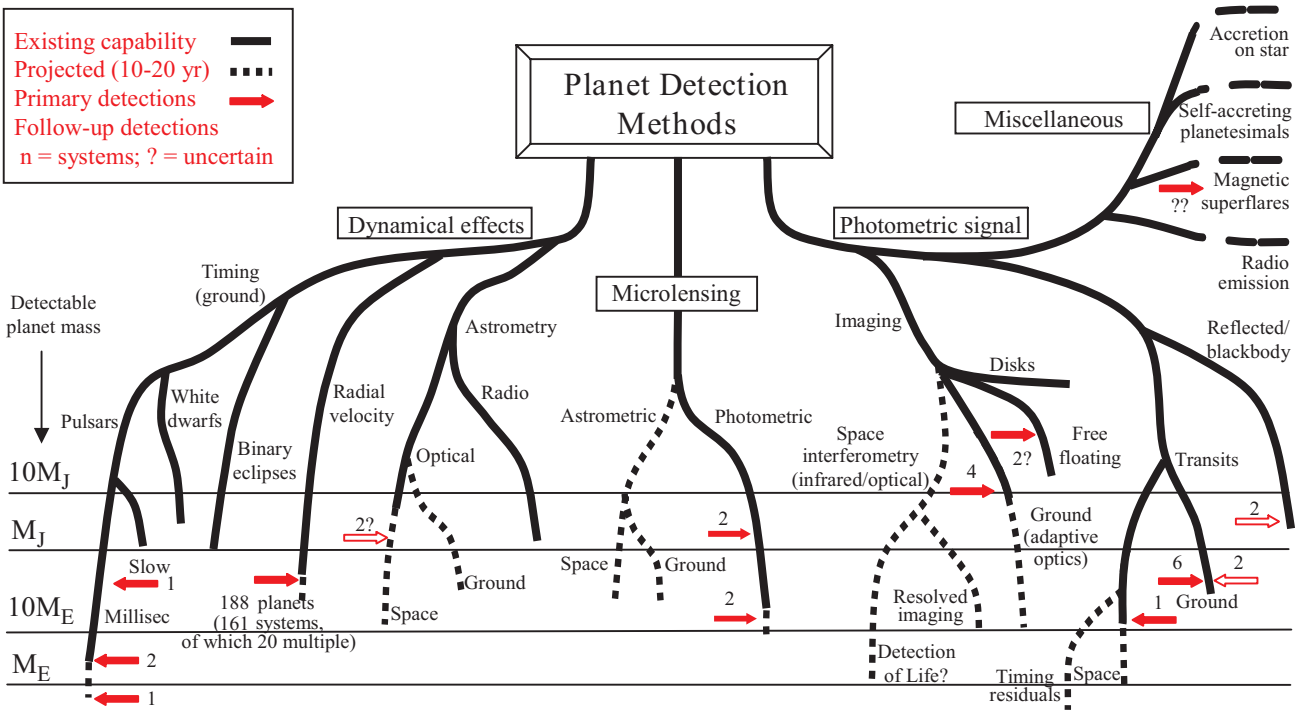

Figure 1.9: Detection methods for extrasolar planets. We have allowed ourselves to update this diagram from Perryman (2000). The lower extent of the lines indicates, roughly, the detectable masses that are in principle within reach of present measurements (solid lines), and those that might be expected within the next 10-20 years (dashed). The (logarithmic) mass scale is shown on the left. The miscellaneous signatures to the upper right are less well quantified in mass terms. Solid arrows indicate (original) detections according to approximate mass, while open arrows indicate further measurements of previously-detected systems. A question mark indicates uncertain or unconfirmed detections. The figure takes no account of the number of planets that may be detectable by each method.

1.5.2 Indirect detection methods

Almost all planets discovered so far have been evidenced with indirect methods, which rely on the effect of the planet on its parent star. Most of the discoveries up to now have been obtained thanks to precise measurements of the host star's radial velocity through spectroscopic observations, showing small shifts in the stellar spectral lines (typically a tens of meters per second) as the star moves back and forth due to the gravitational pull of its planet. Another successful technique relies on the dimming of the apparent stellar flux as the planet transits in front of the stellar disk. Even though the probability to detect such a transit is rather low ($p = R_*/a$ is about 10% for a hot Jupiter), ten hot Jupiters have been either found or confirmed using this technique.

The current technical and physical limitations of these methods have not allowed yet the detection of Earth-mass planets: the smallest planet discovered to date by radial velocity measurements has a minimum mass of $7.5\,M_\oplus$ (Rivera et al. 2005), while the smallest ever detected ($5.5\,M_\oplus$) has been revealed by a microlensing event (Beaulieu et al. 2006).

Although future space-based missions for transit searches (CoRoT, Kepler) or astrometric surveys (SIM, Gaia) are expected to push this limit down to one or a few Earth masses, they will still be restricted to the measurements of orbital parameters and will therefore provide limited information on the physics of these supposedly rocky bodies. Indirect methods still have many bright years ahead, but will gradually be complemented and replaced by direct methods, which aim at exoplanet imaging and could eventually lead to the detection of signposts of life outside our solar system.

1.5.3 Direct imaging methods: coronagraphy

Direct imaging of extrasolar systems is still in its first steps: it has only recently allowed imaging its first four extrasolar planets, in very favorable cases where the flux ratio between the star and its planet was only a few hundredths and the angular separation between them relatively large (see Sect. 1.2.3). These discoveries have been made possible only recently with the advent of adaptive optics (AO) systems on large ground-based telescopes, which correct for the distortion induced by the Earth's atmosphere in order to provide sharp images of relatively bright astronomical objects. The addition of a coronagraphic mask in the focal plane, optimized to dim the stellar light, should in a near future enhance the capability of large AO-corrected telescopes to distinguish giant planets around nearby stars with contrasts up to a few thousands. In space, where conditions are most favorable, coronagraphs are even envisioned to tackle the extremely high dynamics involved in Earth-like exoplanet imaging. But what does one exactly mean by "coronagraphy" ?

As initially conceived by Lyot (1939), a coronagraph is a device to suppress instrumental light diffraction by the use of a sequence of stops for the specific purpose of observing the solar corona. Since the Lyot coronagraph addresses light diffraction, a "coronagraph" has become a generic term for a system to suppress diffraction and scattered light in a telescope for astronomical purposes. In common usage, any system to achieve high contrasts with a single-aperture telescope is referred to as a coronagraph, even in cases where there is no physical or historical basis for the connection to the original problem of observing the solar corona. A practical definition that encompasses the current usage of the term is that a coronagraph is a device to suppress the noise associated with stellar light by rejecting it from an area of interest in the focal plane of a telescope. This light must be rejected because of the associated twofold noise issue: wavefront distortions that produce artifacts in the image that resemble planets (speckle noise), and photon shot noise.

Recent years have seen extensive developments of new concepts, all of which strive to search

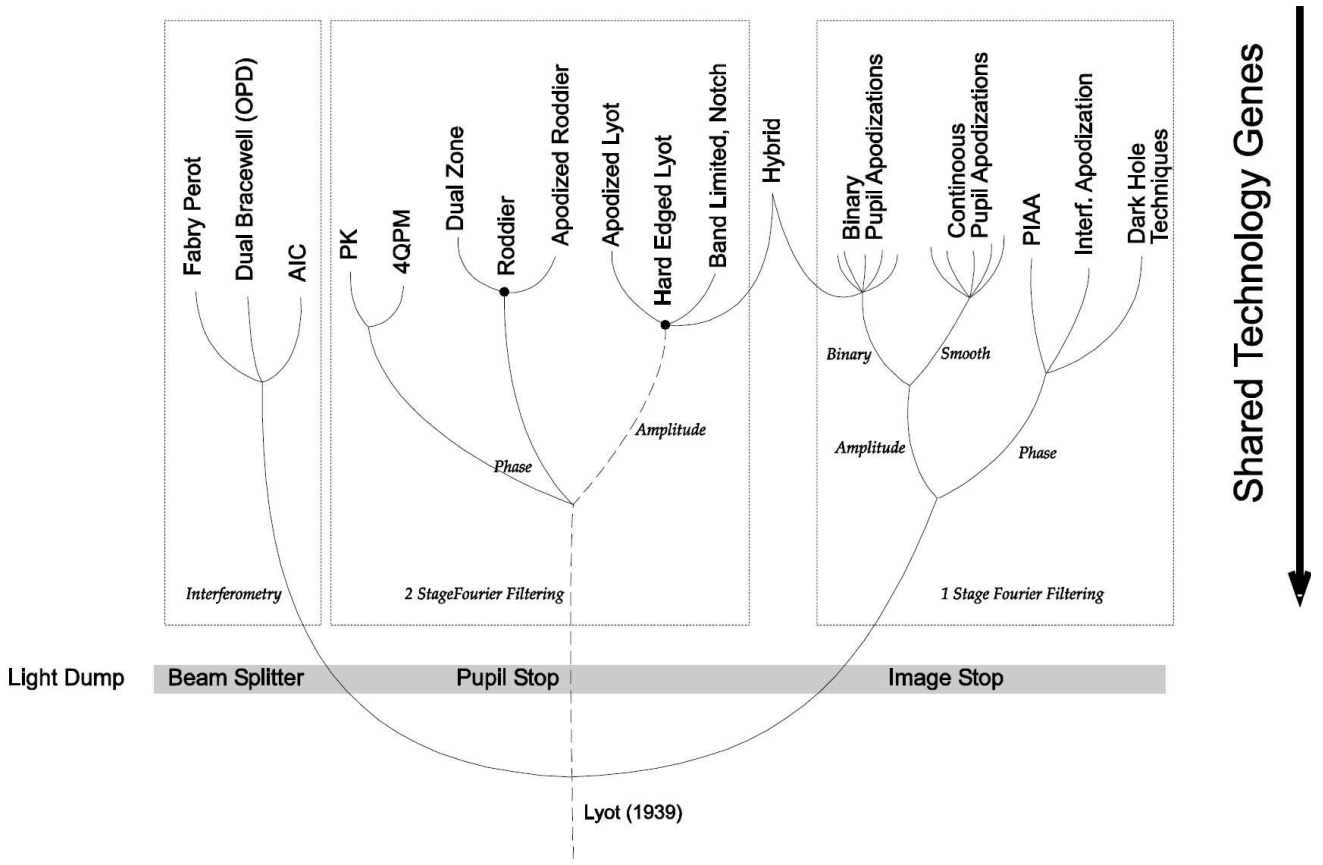

Figure 1.10: Existing coronagraphic concepts (non exhaustive). See text for details. Taken from Quirrenbach et al. (2004).

for the ideal coronagraph, a device that confines the stellar light as tightly as possible with the maximum efficiency to reveal images of extrasolar planets. A preliminary classification of the families of coronagraphs that have been developed is presented in Fig. 1.10. At first sight, the diversity of approaches is surprising. However, such a variety is representative of the dynamism of the community to always search for very efficient solutions at the frontier of scientific goals and technical constraints.

A primary division one could artificially define is between the mechanism the coronagraphs use as a "light dump", i.e., the method by which the stellar light is segregated: dumping the light in a sequence of stops in the image and exit pupil plane, as with a classical Lyot coronagraph (see Fig. 1.11); dumping the light in a confined region of the image plane or image stop; and one rejecting light by interference beam combination at a beam splitter, with a close parallel in nulling interferometry. The image-stop solutions are generally one-stage Fourier spatial filters: application of a method for controlling spatial frequencies in the pupil plane is used to confine the light in the image plane. The pupil-stop family of coronagraphs shares the property of having two (or in some cases more) stages of Fourier spatial filtering, one in the first image plane, and one in the exit pupil plane. Hybrid approaches can be generalized to an expanded number of filtering steps, and there is currently fertile development in mutually optimizing these designs.

A secondary division is between the occultation method implemented by the coronagraphic system. Derived from the original approach of Lyot, *amplitude coronagraphs* block the light directly either in the image or in the pupil plane, whereas *phase-mask coronagraphs* are transparent but induce spatially-distributed phase shifts with the goal of destructively reject the starlight out of the so-called *discovery zone*.

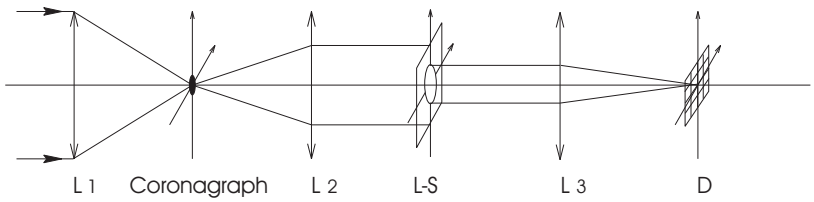

Figure 1.11: Traditional coronagraphic optical bench scheme. L1, L2, and L3 are three lenses in the optical system. L1 provides a large F/d ratio on the coronagraph to minimize spatial defects, L2 images the pupil in the second plane, the Lyot stop (L-S) suppresses the diffracted starlight, and finally L3 forms the coronagraphic image on the detector D.

Amplitude coronagraphs

Amplitude coronagraphs, derived from Lyot's original design (Fig. 1.11), have recently benefited from numerous refinements. One of the first evolution of the hard-edge Lyot coronagraph was the smoothing of its transmittance thanks to a Gaussian apodization intended at optimizing the relayed pupil-plane filtering by the Lyot stop (Watson et al. 1991). The same philosophy of optimized focal-plane amplitude filtering is applied by the so-called "band-limited/notch-filter" masks (Kuchner & Traub 2002; Kuchner & Spergel 2003; Debes et al. 2004; Kuchner et al. 2005; Shaklan & Green 2005; Crepp et al. 2005). These sophistications of the rejection capabilities of the Lyot coronagraph have unfortunately a price: a reduced optical throughput and discovery space.

Apodization also reveals to be an efficient solution if applied directly at the entrance pupil of the coronagraphic instrument (Aime et al. 2002; Soummer et al. 2003a; Aime 2005b,a; Soummer 2005). The cited schemes still require a classical hard-edge Lyot spot at the instrument focus whereas the so-called "shaped pupils" (nonuniform pupil transmittance) do not (Nisenson & Papaliolios 2001; Kasdin et al. 2003, 2005; Vanderbei et al. 2003b,a, 2004). In this respect, the first ones still belong to the pupil-stop family (2-stage Fourier filtering) while the last ones belong to the image-stop family (1-stage Fourier filtering). Let us also mention the very special case of the phase-induced amplitude apodization (PIAA, Guyon 2003; Guyon et al. 2005; Galicher et al. 2005; Pluzhnik et al. 2006; Martinache et al. 2006), which is an alternative to classical pupil-apodization techniques obtained by reflection of an unapodized flat wavefront on two well-shaped aspheric mirrors. This technique theoretically preserves both the angular resolution and light gathering capabilities of the unapodized pupil and is expected to provide excellent rejection capabilities. The drawback of the PIAA comes from the very difficult manufacturing of the aspheric mirrors with the required surface qualities.

Phase-mask coronagraphs

As already mentioned, amplitude coronagraphs (except the PIAA) possess a major inherent weakness: the physical extension of the opaque zone occults quite a significant region centered on the optical axis and thus all the sources behind it. This inevitably leads to a loss of global throughput and search efficiency. For this reason, new types of coronagraphs have been designed as alternative solutions to the amplitude coronagraphs. This new family of components acts on the phase instead of acting on the amplitude, they are therefore called phase-mask coronagraphs. The phase-mask coronagraphs are not affected by the inherent "dead zone" of the amplitude coronagraphs.

With a close parallel to nulling interferometry (Sect. 1.5.4), Gay & Rabbia (1996) proposed the concept of the "achromatic interfero-coronagraph" (AIC), which is a modified Michelson's

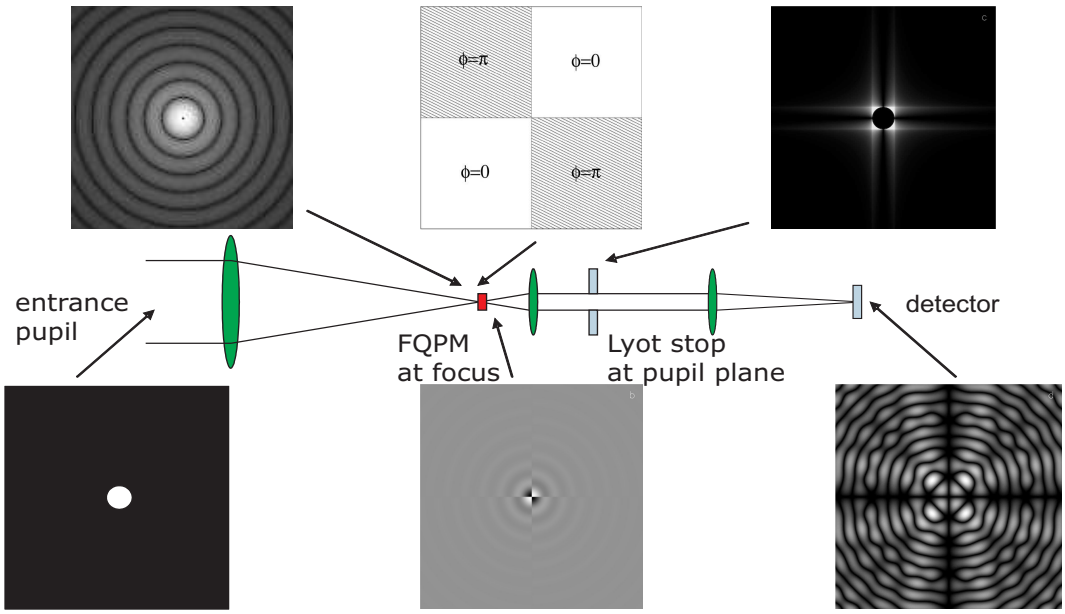

Figure 1.12: Principle of the FQPM. Top left: Airy pattern image. Top middle: FQPM focal plane phase distribution. Top right: relayed pupil plane where the redistribution of the light thanks to the FQPM takes place. Bottom left: entrance pupil. Bottom middle: image phase after transmission through the mask. Bottom right: final image where the numerical residuals are shown. In principle the attenuation of the FQPM is infinite in the perfect unobscured/unaberrated pupil case.

interferometer exploiting pupil rotation and therefore avoiding the use of physical masks (Gay et al. 1997; Baudoz et al. 2000a,b). The beam goes through a focus which provides an achromatic π-phase shift to produce a self-destructive interference for an on-axis source. It is also worth noting the existence of the so-called visible nulling coronagraph (VNC, Levine et al. 2006) which consists in a 4-beam nulling interferometer (dual Bracewell, see Fig. 1.15 in Sect. 1.5.4) synthesized from the telescope pupil, producing a very deep θ^4 null (Sect. 2.1.1) thanks to crossed lateral shears and phase-plate phase shifters, which is then filtered by a coherent array of single-mode fibers to suppress the residual scattered light. A totally symmetric alternative has been proposed by Ren & Serabyn (2005) in which the phase shift is performed thanks to orthogonal rooftop reflectors inducing an achromatic relative field flip (Sect. 2.2.3) between the two interferometer arms (rotational shearing interferometer).

Roddier & Roddier (1997) suggested to use a disk phase mask at the image plane with a size typically half the diameter of the Airy peak. The starlight self-cancellation in the relayed geometric pupil area is also provided by a destructive interference thanks to the mask π-phase shift (Guyon et al. 1999). The disk phase-mask coronagraph (DPMC), unlike the AIC, still requires a pupil-plane diaphragm (Lyot stop) to efficiently remove the diffracted starlight residual, and is moreover highly chromatic. Let us also mention the dual-zone phase mask (DZPM, Soummer et al. 2003b), which is an evolution of the DPMC, partly solving the chromaticity issue of the later, but requiring entrance pupil apodization to provide its full rejection potential.

Later on, Rouan et al. (2000) proposed a very performing design, the so-called four-quadrant phase-mask coronagraph (FQPM). We shall linger a little bit more on this coronagraph for it is the starting germ of this thesis. The principle of the FQPM is to divide the focal plane in four equal areas centered on the optical axis, with two of them providing a π-phase shift. This causes destructive interference to occur inside the relayed geometric pupil area (see Fig. 1.12). The final image is then formed after proper filtering through a classical pupil-plane Lyot stop. To be as

Figure 1.13: Panoramic view of the VLT's four UTs (Unitary Telescopes) and three ATs (Auxiliary Telescopes). NAOS-CONICA is installed at the UT4 (right). ©ESO.

exhaustive as possible, let us also mention the phase-knife coronagraph (PKC, Abe et al. 2001, 2003), a variant of the FQPM.

The FQPM has been extensively studied theoretically (Riaud et al. 2001), it has been validated on an optical bench both in monochromatic light (Riaud et al. 2003) and polychromatic white light (Mawet et al. 2006, see Sect. 2.3), and installed on the NAOS-CONICA adaptive optics at the VLT (see Fig. 1.13 and Boccaletti et al. 2004). Since then, it has given promising scientific results on circumstellar disks (Riaud et al. 2006, see Sect. 1.3.3) and close AGN environments (Gratadour et al. 2005), as well as perspectives for future instruments such as the mid-IR instrument (MIRI) of the James Webb space telescope (JWST, formerly the next-generation space telescope), the infrared successor of HST, and the VLT-planet finder (recently renamed SPHERE). Let us now present briefly these two important projects for the FQPM and its heirs.

JWST/MIRI. By 2013, JWST will be the largest telescope operating in space (Fig. 1.14, left), providing a wide and continuous spectral coverage from the visible ($0.6\,\mu$m) to the mid-IR ($28\,\mu$m) wavelengths, using its imaging and spectroscopic facilities. With a diameter of 6.57 m, JWST will provide an unprecedented sensitivity at all wavelengths. The JWST is a key mission in the NASA Origins program. The main objectives of this ambitious mission are the understanding of the universe, the birth and formation of stars and their planetary systems. To achieve this program, JWST will include four instruments one of which (MIRI) is dedicated to the study of faint red-shifted galaxies. This instrument is also particularly well suited to a coronagraphic search of extrasolar planets. The expected detection capability of the FQPM as calculated by Boccaletti et al. (2005) with inputs from successful infrared cryogenic breadboarding (Baudoz et al. 2004) is that planetary companions can be searched for around nearby stars. Up-to-date results presented in Baudoz et al. (2006) show that, from the pure detectivity point of view, the expected performance of the instrument should allow the imaging ($SNR \approx 10$, 1-hour integration time) of a giant exoplanet orbiting at 10 AU from its parent star located at 10 pc, and with a temperature ≥ 400 K, i.e., according to evolutionary models (see, e.g. Burrows et al. 1997), corresponding to a few Jupiter masses. Multi-wavelength coronagraphic observations will allow the characterization of giant exoplanets by measuring their effective temperatures as well as the ammonia abundance in their atmosphere. MIRI will also be capable of detecting circumstellar disks one order of magnitude brighter than the Kuiper belt in the solar system.

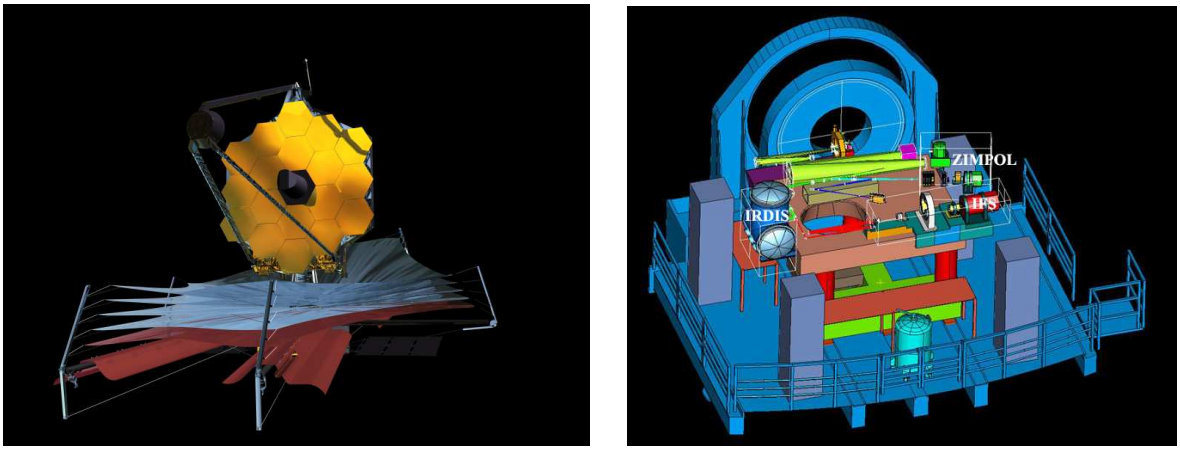

Figure 1.14: Left: James Webb Space Telescope (JWST) at L2 in 2013 (©NASA). Right: VLT-PF/SPHERE implementation on the VLT's Nasmyth platform showing the common path with the XAO system and the three science instruments IRDIS (Infra-Red Dual-beam Imager and Spectrograph), IFS (Infra-red Integral Field Spectrograph) and ZIMPOL (Zurich Imaging Polarimeter). ©ESO.

VLT-planet finder/SPHERE. The FQPM or one of its achromatic heir is also foreseen for the coronagraphs of the VLT-planet finder (a second-generation instrument for the VLT, see Fig. 1.14, right, and Mouillet et al. 2003; Beuzit et al. 2005; Boccaletti & Mouillet 2005). In order to pave the way towards extremely-large telescopes (ELTs) and space interferometers, one needs to progressively gain orders of magnitude in contrast in the close vicinity of bright stars. The VLT-planet finder, recently renamed SPHERE, is one of the mandatory steps aimed at detecting and characterizing hot extrasolar planets and circumstellar environments through the direct analysis of their emitted photons in the visible and at near-IR wavelengths. This near-future (2011), near-infrared extreme adaptive optics instrument (XAO) for the VLT (41x41 actuators leading to an expected Strehl ratio[6] of 90% in H band), will be equipped with high-performance tools for the removal of coherent and incoherent starlight: phase coronagraphs and differential imaging. The latter technique takes advantage of the physical properties of giant planets/circumstellar disks in scattered light: either the spectral features or the polarization of the planet atmosphere/circumstellar disks is used to increase the sensitivity through subtraction methods, defeating the so-called speckle noise (Racine et al. 1999; Marois et al. 2000, 2005; Lenzen et al. 2004, 2005, see also Sect. 2.1.3).

Consequently, the detection of planets with SPHERE will allow characterizing their atmosphere to some extent: effective temperature (depth of CH_4 and H_2O features with near-IR low-resolution spectro-photometry), dust grain size and cloud characteristics (degree of polarization), planet radius and geometric albedo. The contrast of 10^{-4} to 10^{-6} at angular separations of 0.1 to 1.5″ that will be achieved, with a gain of up to 8 magnitudes with respect to current instruments, should allow discovering and characterizing planets in the 1-10 M_J mass range and 1-1000 year period range, i.e., all the gas giants that were expected before the discovery of hot Jupiters.

It is to be noted that coronagraphs, thanks to their quite easy and straightforward implementation, are now available in almost every professional observatory[7]. They are also foreseen for numerous ground- and space-based projects. We have already mentioned the VLT-planet

[6]The modern definition of the Strehl ratio is the ratio of the observed peak intensity at the detection plane of a telescope or other imaging system from a point source compared to the theoretical maximum peak intensity of a perfect optical system working at the diffraction limit.

[7]In fact, Lyot coronagraphs are widespread but to our knowledge, the VLT is the only observatory equipped with an enhanced coronagraph like the FQPM, which is commissioned and thus officially available to the community.

finder (SPHERE) and JWST's MIRI coronagraphic subsystem, but we must also cite space-based projects like SEE-COAST (see Sect. 6.4), ECLIPSE (Trauger et al. 2003), EPIC (Clampin et al. 2004), which are 1.5-meter class telescopes working in the visible and the Japanese SPICA project (Matsumoto 2005), a 3.5-meter telescope for mid- and far-infrared coronagraphic observations. From the ground let us cite: at the Subaru telescope, HiCIAO, the extreme adaptive optics successor of the coronagraphic imager with adaptive optics (CIAO, Murakawa et al. 2004); at Gemini, the Gemini planet imager (GPI, Macintosh et al. 2006), etc. Of course, we must also cite the most ambitious terrestrial planet finder-coronagraph (TPF-C, Ford et al. 2004), a major NASA mission project, planned for launch around 2015, designed to detect Earth-like planets around nearby stars, to characterize them in terms of atmospheric and surface constituents, and to search for signs of life on them. TPF-C will try to detect extrasolar planets using visible- and near-infrared filter photometry. It will characterize those planets using low-resolution spectroscopy. Potentially detectable species include, as already mentioned, H_2O, O_2, O_3, CH_4, CO_2, Rayleigh scattering, and the "red-edge" signature of chlorophyll in the leaves of land plants. The mission is designed around a telescope with a large off-axis primary mirror (8 m by 3.5 m), and a focal-plane instrumentation comprising a coronagraph, a deformable mirror speckle nuller (see, e.g. Bordé & Traub 2006), and a spectrograph.

1.5.4 Nulling interferometry

The angular resolution necessary to achieve most-wanted science cases of high contrast astrophysics, such as Earth-like planet detection and characterization, can hardly be accomplished with current monolithic telescopes in the mid-infrared regime, where the angular resolution λ/D is limited to about 250 mas. Larger telescopes, up to 100 m in diameter, are currently being studied but, as such large apertures are not expected to be feasible in space, they will still be limited by the presence of the Earth's atmosphere which degrades the images and restricts the observations to the infrared transparency windows (e.g. the N band, between 8 and 13 μm). In order to benefit both from the full infrared electromagnetic spectrum and from a high angular resolution, space-based nulling interferometric missions have been proposed. The principle of interferometry is to synthesize the resolving power of a large monolithic telescope by combining the light collected by several smaller telescopes, separated by a distance equivalent to the required diameter for the large telescope.

In addition to the increased angular resolution, interferometry also provides a natural way to dim the stellar light, by using the wave properties of light: by inserting phase-shifting devices in the path of the light beams, one can adjust their respective phases in order to produce a totally destructive interference on the optical axis (see Fig. 1.15). This technique, referred to as "nulling interferometry" dates back to almost 30 years, when Bracewell (1978) proposed in a letter to Nature to "detect non-solar planets by spinning infrared interferometer". The rationale of this proposal was already based on the limited performances of indirect techniques (astrometry, radial velocity) and on the inability of "direct photography" to detect small and faint extrasolar planets around bright stars: even when a telescope point-spread function is very well calibrated and subtracted, the shot noise associated with stellar light remains, outshining any faint signal nearby. Bracewell thus proposed a way to enhance the planet over star flux ratio by placing a destructive interference on the star and a constructive one on the planet. The initial proposal was aiming at the detection of Jupiter-like planets in the far-infrared (40 μm) with a moderate interferometer baseline (7.7 m), but the key features of future missions for Earth-like planet detection were already included.

Since the first proposal of an infrared nulling interferometer to detect and characterize Earth-like planets, the technique has been studied and improved by many authors, leading to the current

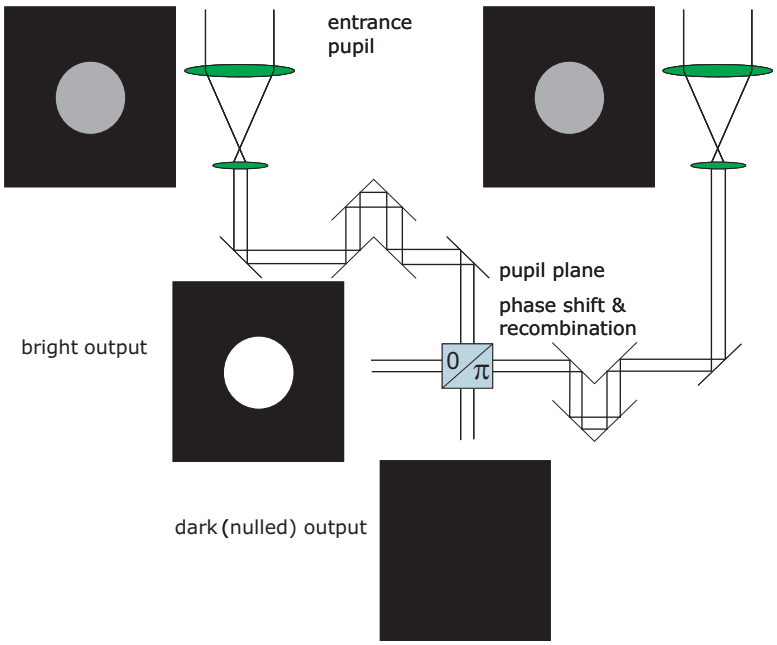

Figure 1.15: Principle of the Bracewell two-telescope interferometer. The light coming from the two telescopes is collimated while put in phase opposition thanks to a phase shifter. The light is recombined in the pupil plane, leading to a constructive interference on one output of the beam splitter and a destructive one on the other output. The detection is generally made on single-pixel detectors.

view of the future planet finding missions (Fridlund 2004a): a flotilla of 3 to 6 free-flying telescopes of moderate diameter ($\sim$ 3 m) capable of searching for planets around the 150 closest F, G and K-type stars in a 5 year mission. A road map of technological developments was set up in order to reach that goal in the 2015 time frame, both at the European space agency (ESA) with the Darwin mission (Fridlund 2004b), and at NASA with the terrestrial planet finder-interferometer (TPF-I, Beichman et al. 1999). More pragmatically, such challenging missions are not foreseen before ESA's 2020 Cosmic Vision horizon.

Before launching such an ambitious observatory, intermediate technological and scientific steps are mandatory. In this context, a mission has been proposed as an answer to the CNES call for ideas for a scientific payload on its formation-flying technological mission (Pegase, Ollivier et al. 2005). It consists of a Bracewell interferometer operating in the infrared (1.5-5 μm) and visible regimes. It has small telescopes (40 cm) but a substantial baseline (25 to 500 m). Its angular resolution reaches 1 mas at 4 microns. Its main scientific objectives are the spectroscopic study of low-mass companions including Pegasides and brown dwarfs bounded to other stars, with the goal of determining the composition of the atmospheres of these objects as well as their internal structure.

However, there remains two major unknown parameters driving the design of these missions: the incidence of Earth-like planets around nearby stars, and the amount of dust to be expected in their habitable zone. The first parameter will be addressed to a large extent by future space missions that will search for planets with the transit method (CoRoT and Kepler), as they are expected to be sensitive to planets down to one Earth radius (Bordé et al. 2003). The second parameter is more difficult to assess with current instruments, because it requires both high angular resolution and high dynamic range imaging capabilities. In the context of the Darwin programme, ESA and ESO have initiated a definition study for a ground-based technology and science demonstrator called GENIE (Absil et al. 2006, Ground-based European Nulling Interferometry Experiment).

Figure 1.16: Darwin space interferometer's six telescopes (original configuration), central view combining spacecraft, and communication satellite at L2. ©ESA.

This collaboration was initiated because ESO's VLT interferometer was soon identified to be the best place to install a European nulling instrument. A letter of intent, signed by the two agencies, was sent to the European astronomy community in mid-2001, identifying the objectives of the GENIE instrument:

1. To gain required technological experience and demonstrate the technique of nulling interferometry with a breadboard representative of the Darwin mission;
2. To carry out required precursor science:
   - detection and measurement of exozodiacal disks associated with the Darwin targets,
   - detection and characterization of large sub-stellar companions around the Darwin targets;
3. To allow European scientists to gain experience with a nulling interferometer in doing practical astronomy, and to train a new generation of "interferometrists" in preparation of the utilization of Darwin;
4. To carry out unique and valuable science.

While GENIE was unfortunately put on hold by ESA and ESO, other nulling experiments have been working for a few years or are about to be operational. The first and only nulling experiment to have published astrophysical results so far is the Bracewell infrared nulling cryostat (BLINC). This instrument was first installed on the segmented version of the multi-mirror telescope (MMT), using the beams from two of its 1.8-meter segments, and produced first scientific results on late-type giant stars around which dust outflows were detected (Hinz et al. 1998). After the refurbishment of the MMT in 1999, BLINC was re-installed on the new 6.5-meter monolithic telescope (Hinz et al. 2000). It was also used in the southern hemisphere, at the Magellan I telescope.

Because it operates in the mid-infrared (10-12 μm), where the background emission is huge (especially for a non-cryogenic instrument), the Keck interferometric nuller (KIN) implements an original way of removing the background emission in addition to the removal of the stellar light. This method is based on phase chopping, a technique currently considered for the Darwin

and TPF-I space missions. In order to modulate the signal from off-axis sources in front of the background emission, a dual-baseline nuller is synthesized by creating two distinct sub-apertures on each of the two Keck telescopes and then cross-combining the outputs of the two individual nulling beam combiners with a standard interferometric beam combiner (Serabyn et al. 2004). The beam combination in the two individual nullers is performed with modified Mach-Zehnder beam combiners (Serabyn & Colavita 2001). The KIN was designed to achieve exozodiacal disk detection at the 30-zodi[8] level around a G2V star located at 10 pc. This level of performance requires a deep and stable instrumental rejection ratio (see Sect. 2.1) of at least 10^3 at 10 μm. The integration and test laboratory of the Keck interferometer at JPL was used to demonstrate the capability of the instrument to reach such a performance: a broadband null depth better than 10^{-3} was obtained and stabilized for periods of many minutes (Serabyn et al. 2004). This led to the decision to deploy the hardware at the Mauna Kea observatory, where it is currently in commissioning phase (first N-band fringes obtained in August 2004, first stable null at 10^{-2} obtained in May 2005) and should soon provide its first scientific results.

The LBT interferometer, which combines the light from two 8.4-meter telescopes placed side by side on a single rigid alt-azimuth mount, is very well suited to develop and exploit the basic Bracewell nulling method. With its short baseline of 14.4 m, it provides an angular resolution of 70 mas in the mid-infrared, which corresponds to the typical angular size of habitable zones around nearby G-type stars. The LBTI instrument is composed, amongst other things, of the nulling interferometer for the LBT (NIL), and the nulling optimized mid-infrared camera (NOMIC), which forms an image of the field around the star and is capable of detecting infrared emission from surrounding dust disks and planets (Herbst & Hinz 2004). It is to be noted that the NIL instrument relies on an upgraded version of the BLINC instrument. The targeted null depth with the NIL instrument is 10^{-4}, with the objective to detect exozodiacal disks similar to the solar zodiacal disk (1-zodi level). Background subtraction, cophasing uncertainties and the absence of spatial filtering, even with the 670-actuator deformable mirror, could however prevent such a null depth with the current design of the nulling instrument. The first 8.4-meter mirror of the LBT was installed and achieved its first light in 2005. The second primary mirror has recently been transported from the University of Arizona to Mount Graham and has also been installed. The nulling instrument will be used to perform a survey called NIREST (nulling infra-red survey of exo-systems for TPF) of 80 TPF-I candidate stars. Its purpose is to search for and measure zodiacal dust emission strong enough to compromise TPF-I's performance. A by-product of the survey will be the detection of thermal emission from extrasolar giant planets.

1.5.5 Perspectives in Antarctica

One of the main limitations of coronagraphic observations as well as ground-based nulling interferometers is related to the influence of atmospheric turbulence. Active compensation of the harmful effects of turbulence requires real-time control systems to be designed with challenging performance. The choice of a good astronomical site with (s)low turbulence is therefore of critical importance. In this respect, recent studies suggest that the high Antarctic plateau might be the best place on Earth to perform high resolution observations in the infrared domain, thanks to its very stable atmospheric conditions (Agabi et al. 2006). Dome C is considered as "halfway to space". For this reason, many interferometer concepts have recently been proposed, all taking advantage of the expected unprecedented site quality.

The ALADDIN concept is an integrated Antarctic-based L-band nulling breadboard with rel-

[8] 1 zodi corresponds to the brightness of our solar system's zodiacal disk.

atively modest collectors (1 m) and baseline ($\leq$ 40 m). Because of its privileged location, this is sufficient to achieve a sensitivity (in terms of detectable exozodiacal dust levels) which is 1.6 to 3.5 times better than an equivalent nulling instrument on a large interferometer (such as GENIE at the VLTI), bringing it below the 20-zodi threshold value identified to carry out the Darwin precursor science (Absil 2006). An integrated design would enable top-level optimization and full access to the light collectors for the duration of the experiment, while reducing the complexity of the nulling breadboard. One can also mention the Antarctic plateau interferometer (API, Swain et al. 2004), an instrument concept capable of studying exoplanets in the habitable zone. API would use three 2-meter class telescopes, high dynamic range spectroscopy, and differential closure phase to achieve 10^{-5} contrast ratio measurements. Combining existing interferometer technology (adapted to the Antarctic environment) and containerized packaging would make it possible to begin operation at Dome C in 5 years. There is also the kiloparsec explorer for optical planet search (KEOPS, Vakili et al. 2004), an interferometer to be placed at Dome C plateau of Antarctica. It consists of an interferometric array of 39 telescopes 1 m to 2 m in diameter spread over kilometric baselines and operated in the thermal IR region. It could search and characterize all potential exoEarths within the 1 kpc diameter region observable from Dome C. Even in the very difficult operation conditions of Antarctica, such a facility could compete with future space missions but at a much lower cost, both for exoplanet studies and for sub-mas snap-shot imaging of galactic and extragalactic compact sources.

2

The need for achromatic phase shifters

Contents

Abstract. *The detection and characterization of extrasolar planets, as for every science case of high dynamic range imaging, require unprecedented levels of light suppression. The process of destructive interference is therefore submitted to very severe tolerances which translate into tight constraints on key components such as the achromatic phase shifters (APS). In this chapter, after exposing the basics of destructive interference processes, specificities proper to phase-mask coronagraphy and nulling interferometry implementations will be reviewed. Then, the existing concepts of achromatic phase shifters will be exposed.*

2.1 Interferometric and coronagraphic nulling

Nulling the light signifies creating a desctructive interference between different coherent wavefronts. As already mentioned in Sect 1.5.4, the idea of exploiting this phenomenon for high dynamic range imaging in astrophysics was first exposed by Bracewell (1978). He proposed a way to enhance the planet over star flux ratio by placing an interference pattern $\sin^2(\pi\theta/\Phi)$ with a maximum on a planet at $\theta = \Phi/2$ and a minimum on the on-axis star, i.e., centered at $\theta = 0$. Φ is defined as follows: $\Phi = \lambda/b$, with λ the wavelength, and b the interferometric baseline length. As already presented in Sect. 1.5.4 and Fig. 1.15, the interference pattern is generated by overlapping the light beams collected by two telescopes on a balanced beam splitter in the pupil plane while tuning their respective phases by means of phase-shifting devices so that all the on-axis stellar light is sent to only one of the two complementary outputs of the beam splitter. Such a configuration happens when a phase shift of π radians is maintained between the two input beams. This is in contrast with classical pupil-plane interferometry, where the optical path difference between the beams is modulated in order to record a complete interferogram.

It is worth emphasizing once again that the underlying principle common to all phase coronagraphs is quite the same as in nulling interferometry. Indeed, the single-telescope incoming wavefront is artificially divided into several parts which are phase shifted with respect to each other and then recombined in the following optical train to form a high contrast image. One of the major differences between nulling interferometry and phase-mask coronagraphy concerns the latter point: in nulling interferometry, no image is formed, the detection being generally made on a single pixel.

For interferometry as well as for coronagraphy, the quality of the nulling process is quantified by the so-called *rejection ratio* R, or its inverse, the *null depth* N

$$N(\underline{r}, \lambda, t) = R^{-1}(\underline{r}, \lambda, t) = \frac{I_{min}}{I_{max}} \tag{2.1}$$

where I_{min} is the total residual intensity of the destructive output (resp. with the coronagraph) and I_{max}, the total intensity of the constructive output (resp. without the coronagraph). In order to maximize the rejection ratio or equivalently minimize the null depth, several conditions must be satisfied by the interfering wavefronts. These conditions translate into tight spatial, spectral and temporal constraints in interferometry as well as in phase coronagraphy, because of the convolution process of the phase mask in the focal plane.

2.1.1 Spatial constraints: geometrical leakage

Nulling interferometer. Even with an ideal Bracewell interferometer, the stellar signal is perfectly cancelled only on the optical axis. Part of the starlight will therefore leak through the created transmission map due to the finite size of the stellar photosphere, an effect referred to as *geometric stellar leakage* because it is related to the physical arrangement of the telescopes and to the angular size of the stellar photosphere. The transmission close to the center of the field of view is proportional to the square of the off-axis angle θ, so that the Bracewell configuration is called a "θ^2 configuration". The rejection ratio in this case is then given by

$$R = \frac{4}{\pi^2}\left(\frac{\lambda/b}{\theta_\star}\right)^2 \tag{2.2}$$

It depends on two parameters: the ratio of wavelength to baseline, which sets the angular resolution of the interferometer, and the angular radius of the star $\theta_\star$. The rejection ratio decreases for longer

baselines (or shorter wavelengths) as the star gets more and more resolved. Although the residual stellar leakage can theoretically be removed by rotational modulation if the star is symmetric, the shot noise associated with the detected leakage still remains and generally represents one of the major noise contributions in the context of Earth-like planet detection. This limitation can be overcome by using more than two telescopes to produce the destructive interference. This idea was first considered by Angel (1990), who proposed to use a four-telescope array to reach a deeper null, proportional to θ^4 close to the optical axis instead of θ^2 for the two-telescope Bracewell interferometer.

The potential of multi-telescope arrays was then exploited in the context of the initial Darwin and TPF-I projects: an array of four to five telescopes arranged on a circle and providing a θ^4 central transmission was suggested for the Darwin mission (Léger et al. 1996), while a four-telescope linear array with a θ^6 central transmission was proposed for the TPF-I mission (Angel & Woolf 1997). Later on, a general condition was derived by Absil (2001) to reach a θ^{2p} transmission for an arbitrary telescope array.

Coronagraphs. Coronagraphs are also affected by the finite size of the star. This stellar leakage can be treated by considering the stellar disk as a sum of off-axis point sources (Riaud et al. 2003), therefore assimilating the spatial leak to pointing errors (i.e., *tip-tilt*). The sensitivity of coronagraphs to such *low-order aberrations* strongly depends on the coronagraph design and whether it acts on the amplitude or the phase. It can however be minimized in both approaches by increasing the power n of off-axis θ^n transmission (higher-order masks, Foo et al. 2005; Lloyd & Sivaramakrishnan 2005; Mawet et al. 2005b; Palacios 2005; Shaklan & Green 2005). In the case of pure phase-mask coronagraphs, the family of coronagraphs the most sensitive to the finite size of stars, it is to be noted that the stellar leakage can anyway always be partially removed by adding a Lyot dot in the center. It is worth noting that the physical extension of this opaque mask is by far smaller than in pure amplitude masks (Roddier & Roddier 1997; Riaud et al. 2003).

2.1.2 Spectral constraints: achromaticity

Expression 2.2 for the rejection rate is valid only for a perfect Bracewell interferometer. In practice, for nulling interferometers but also for coronagraphs, the rejection rate is degraded by atmospheric turbulence (for ground-based facilities) and various instrumental effects causing imperfect co-phasing of the light beams, wavefront errors, intensity mismatches and polarization errors (Ollivier 1999). This contribution, called *instrumental leakage* and independent of the stellar diameter to the first order, adds to the geometric leakage at the destructive output of the interferometer (resp. the final image of the coronagraphic instrument). Moreover, in the framework of high contrast imaging, one usually deals with large spectral bandwidths. In order to preserve the full efficiency of the destructive interference process, the phase shift must be *achromatic*. If not the case, the destructive interference is not complete at all wavelengths and some light from the central bright object is not suppressed. This instrumental leakage could prevent the detection of the light from the faint feature, because of the noise created by the residual light escaping the destructive process. Approximate expressions can be derived for the various chromatic contributors to instrumental leakage, although it is dominated by nonlinear, second-order error terms (Lay 2004). In practice, it is commonly admitted that "achromaticity" is reached as soon as the sensitivity to wavelength change is faint enough to be tolerable in the concerned application.

Let

$$U_j(\lambda) = A_j(\lambda) e^{i\frac{2\pi}{\lambda}h_j + i\phi_j(\lambda)} \tag{2.3}$$

be the complex amplitude of the wavefront number j. A_j is its amplitude ($A_j^2 = I_j$) whereas h_j and ϕ_j are the optical path delay (OPD) and phase shift experienced by the wavefront, respectively. In a two-way interferometer, with $A_1^2 = A_2^2 = A^2 = I_0$, the transmitted intensity at the recombination step is

$$I(\lambda) = |U_1(\lambda) + U_2(\lambda)|^2 = 2I_0 \left[1 + \cos\left(\frac{2\pi}{\lambda}(h_2 - h_1) + (\phi_2(\lambda) - \phi_1(\lambda))\right)\right] \tag{2.4}$$

Phase control

Let us assume that the zero-OPD ($h_2 - h_1 = 0$) is always satisfied by a proper set-up adjustment. The ideal phase shift for nulling in a two-way interferometer (or phase coronagraph) requires that $\Delta\phi(\lambda) = \phi_2(\lambda) - \phi_1(\lambda) = \pi$. In this perfect case $I(\lambda) = I_{min} = 0$ such that the on-axis light is totally suppressed, leading to an infinite rejection ratio. However, at least for some wavelengths, the phase shift is in practice not perfectly equal to π but rather to $\Delta\phi(\lambda) = \pi + \epsilon(\lambda)$, where $\epsilon(\lambda)$ is the wavelength dependent residual phase shift error. The corresponding total residual light over the bandwidth $\Delta\lambda$ is subsequently given by

$$I_{min} = \int_{\Delta\lambda} I(\lambda) d\lambda = 2I_0 \int_{\Delta\lambda} \left[1 + \cos\left(\pi + \epsilon(\lambda)\right)\right] d\lambda \tag{2.5}$$

If ϵ is small, the average null depth over the whole considered spectral bandwidth $\Delta\lambda$ resumes to

$$N = \frac{I_{min}}{I_{max}} = \frac{\int_{\Delta\lambda} \epsilon(\lambda)^2 d\lambda}{4\Delta\lambda} = \frac{\sigma^2}{4} \tag{2.6}$$

with σ^2, the variance of the phase shift over the bandpass $\Delta\lambda$.

For example, the target null depth of 10^{-5} requires a maximum phase-shift standard deviation $\sigma = 6.325 \times 10^{-3}$ radian rms over the whole bandwidth. A more severe figure of merit would be obtained by imposing a given null depth at each wavelength instead of an average one. In this case, Eq. 2.6 holds true provided that the standard deviation σ is replaced by the residual phase-shift error $\epsilon(\lambda)$ at each wavelength.

Intensity mismatch

The effect of unequal intensities in the interfering wavefronts is to induce a non-null transmission on the optical axis: the additional amount of light in one of the wavefronts will not interfere and thus contribute as a background emission. In case of intensity mismatch $A_1^2 \neq A_2^2$, Eq. 2.4 becomes

$$I_{min} = |U_1(\lambda) + U_2(\lambda)|^2 = |A_1(\lambda) + A_2(\lambda) e^{i(\pi + \epsilon(\lambda))}|^2 \tag{2.7}$$

Assuming an intensity ratio $q(\lambda) = \frac{A_2^2}{A_1^2}$, the null depth becomes

$$N(\lambda) = \frac{(1 - \sqrt{q(\lambda)})^2 + \epsilon^2(\lambda)\sqrt{q(\lambda)}}{(1 + \sqrt{q(\lambda)})^2} \tag{2.8}$$

For example, to obtain a single-wavelength target null depth of 10^{-5} assuming a phase error ϵ of 6.300×10^{-3} radian, the intensity ratio q must be kept above 0.999 (an intensity mismatch of 0.1%). Assuming $\epsilon = 0$, q has to remain larger than 0.99, i.e., a 1% intensity mismatch.

Polarization mismatch

There are various types of polarization errors, but they can always be converted into phase or intensity errors (Serabyn 2000). In case of alignment mismatches between linear polarization vector angles $\theta_1 \neq \theta_2$, Eq. 2.4 transforms into

$$I_{min} = |U_1(\lambda) + U_2(\lambda)|^2 = |A_1(\lambda)\cos\theta_1 + A_2(\lambda)\cos\theta_2 e^{i(\pi+\epsilon(\lambda))}|^2 \tag{2.9}$$

Assuming this time a misalignment $\Delta\theta = \theta_2 - \theta_1$, Eq. 2.8 still applies but with $q = \cos\Delta\theta^2$. Specifying $q > 0.999$ means that $\Delta\theta$ must be smaller than 0.032 radian, for example.

Wavefront errors

When high rejection ratios are needed, requirements directly translate into drastic optical constraints on the instrument optical components but also on incoming wavefront qualities. One of the most harmful contributors to instrumental leakage is the corrugation of wavefronts, especially for ground-based instruments where atmospheric turbulence produces large phase errors across the pupils.

Nulling interferometer. In interferometry, wavefront errors produce a mismatch between the shapes of the two beams to be combined. If $\sigma^2_{\phi_{1,2}}(\lambda)$ designates the wavelength-dependent phase variance over each pupil, the null depth is (Mennesson et al. 2002)

$$N(\lambda) = \frac{1}{4}\left(\sigma^2_{\phi_1}(\lambda) + \sigma^2_{\phi_2}(\lambda)\right) \ . \tag{2.10}$$

For instance, if a rejection ratio of 10^3 (or equivalently a null depth of 10^{-3}) is required, Eq. 2.10 shows that the standard deviation σ_ϕ of the phase across the pupil cannot be larger than 4.5×10^{-2} radian, which corresponds to a $\lambda/140$ rms wavefront quality. Such a low wavefront aberration is extremely challenging to reach from the ground, even in the mid-infrared. It corresponds to a Strehl ratio of 99.8% while state-of-the-art adaptive optics systems on large telescopes provide a Strehl ratio hardly above 50% in the near-infrared K band, or about 97% in the mid-infrared N band. *Spatial filtering* is therefore necessary for removing high-frequency wavefront aberrations.

Better than the classical spatial filtering by simple pinholes (Ollivier & Mariotti 1997), *single-mode waveguides* efficiently correct wavefront defects of both high- and low-order spatial frequencies, ensure a perfect matching of the amplitude profiles coming from the various beams, and can be used with almost optimum coupling efficiency over a broad optical bandpass. When a single-mode waveguide is used, the incoming wavefront excites the fundamental mode of the guide. The latter is the only one that is allowed to propagate. Its amplitude profile is fully determined by the physical properties of the guide, independently of the incoming wavefront. Corrugations will only affect the amount of energy coupled into the guide: phase errors convert into intensity errors.

Single-mode waveguides greatly enhance the feasibility of high dynamic range interferometry. Mennesson et al. (2002) demonstrate that wavefront qualities required for a 10^{-6} null depth with single-mode waveguide filtering are relaxed from $\lambda/4400$ to $\lambda/63$ (λ is the infrared observing wavelength), corresponding to optical commercial standards. It is to be noted that using a pinhole still relaxes the uncorrected case by a factor of 10, leading to a tolerance of $\lambda/400$.

Coronagraphs. As far as coronagraphs are concerned, defining a sensitivity to wavefront aberrations is not as straightforward as in pupil plane interferometry. Indeed, according to the spatial frequency of the aberration, the final coronagraphic image will be affected at a certain location in the detector focal plane. As the useful working zones, i.e., the so-called *discovery space*, are concentrated around the optical axis, high-spatial-frequency corrugations are not critical while low-order aberrations like tip-tilt strongly contaminates the central field of view. For this reason, wavefront corrugations as well as stellar leakage are considered as low-order aberrations that must be overcome by adapting the coronagraph design and improving the *power spectral density* (PSD) of wavefront generated by the optical train before the coronagraph, e.g., thanks to a dedicated fast-steering (or tip-tilt) mirror. In fact, whatever the coronagraphic system, an imperfect wavefront will feed the coronagraph with light partly incoherent that will be spread in a speckle halo. The final signal-to-noise ratio, and subsequently the integration time, will depend on the level of this PSF halo up to a certain point as we will discuss below. For this reason, the performance of the wavefront correction systems (adaptive optics, for example) is a critical point which must be addressed thoroughly.

2.1.3 Temporal constraints: stability

Nulling interferometer. In the above discussion, we have neglected another important source of noise: instrumental and stellar leakage contributes both as a bias (referred to as *null floor* leakage), by introducing a non-null average amount of additional stellar light, and as a noise, which relates to its temporal variability. Phase-chopping techniques allow for the subtraction of the null-floor leakage, which is the same in the two chopped states provided that there is no systematic difference between them. The bias will therefore contribute only as an additional source of shot noise. On the other hand, most of the instrumental noise is not suppressed by phase chopping, as proven by Lay (2004). This contribution, now referred to as *instability noise*, but as systematic noise in Lay (2004) or variability noise in Chazelas et al. (2006), is dominated by nonlinear, second-order terms related to the perturbations in the amplitudes, phases and polarisation angles of the electric fields from each telescope. The disturbance power spectra mix with each other so that perturbations at all frequencies, including DC, have an effect. Although a simple binary phase chop removes a number of these systematic errors, it has no effect on the dominant amplitude-phase cross terms and on the co-phasing errors. There is no phase chopping scheme that can remove the systematic errors without also removing the planet signal (Lay 2004).

The two independent studies cited here above (Lay 2004; Chazelas et al. 2006) have recently revisited the instrumental requirements for the Darwin and TPF-I missions in order to reduce instrumental stellar leakage down to a sufficiently low level for Earth-like planet detection. Both showed that the requirements on amplitude and phase control do not come from the null-floor leakage, but from instability noise. According to Lay (2004), the phase of the signal from each collector should be controlled to about 1 millirad ($\sim$ 1.5 nm at 10 μm) and the amplitude to about 0.1%. It must be noted that it is in fact impossible to obtain a stable null floor without also having a deep instrumental null, because instrumental perturbations are expected to have a significant dynamic component.

A major property of instability noise is that it does not depend on the actual architecture of the interferometer to the first order: θ^2 and θ^4 configurations produce similar amounts of instability noise if they are subject to the same amplitude and phase perturbations. Another major property is that instability noise significantly increases at short wavelengths. It is in fact expected that instability noise could exceed shot noise at short wavelengths and become the dominant source

of noise. In that case, the advantage of θ^4 configurations with respect to θ^2 configurations would almost disappear.

Coronagraphs. Coronagraphic techniques possess their own "instability noise", known as the *speckle noise*. Speckles arise from diffraction residuals due to uncorrected atmospheric aberrations left by the adaptive optics system as well as quasi-static aberrations mostly originating from the optical train. Several smart techniques of speckle calibrations were already proposed. Each of them takes advantage of the planet properties like its spectral signatures (see below), its polarization (Baba & Murakami 2003) or its coherence with respect to the star (Guyon 2004). The most developed technique so far is probably the spectral differential imaging (SDI, Marois et al. 2000) first implemented inside the TRIDENT camera at the Canada France Hawaii telescope (CFHT, Marois et al. 2005) and more recently at the VLT (Lenzen et al. 2004, 2005). The idea is to record simultaneously two images in two different but spectrally close filters on the same detector. If phase aberrations are small, the images are identical once rescaled in intensity and spatially matched. The mutual subtraction of the two images removes the residual speckles except those resulting from the so-called common and *non-common path errors*. The latter are induced by the optical train downstream of the coronagraph (dichroic, beam splitters, etc.). Cavarroc et al. (2006) showed that any study where common static aberrations are neglected leads to unrealistic results where the variance of the residual intensity converges to 0 for an infinitely long exposure. Moreover, this study also showed that the fundamental detection limit has a quadratic dependence in common path errors and a linear one in non-common ones. As expected, improvement of the detectability can be obtained if the PSD of the static phase aberrations is decreased, especially at low frequencies (low-order aberrations).

2.2 Achromatic phase shifters

The most simple way to induce a phase shift $\Delta\phi$ between two coherent wavefronts 1 and 2 is to add an optical path delay (OPD) either of air or of a given material of refractive index n

$$\Delta\phi = \frac{2\pi}{\lambda}\left(n_1(\lambda)e_1 - n_2(\lambda)e_2\right) \tag{2.11}$$

In this equation, $n_1(\lambda)e_1$ and $n_2(\lambda)e_2$ are the optical path experienced by the wavefront 1 and 2, respectively, i.e., the product of the physical distance e_i by the refractive index n_i of the propagating medium. However, it clearly appears that such a phase shift is highly chromatic because of the hyperbolic dependence in λ and because of the intrinsic refractive-index dispersions of the used materials. One needs achromatic phase shifters that would allow inducing a constant phase shift over a predefined wavelength range. Several techniques already exist and we will briefly describe them.

2.2.1 Dispersive plate APS

This approach, directly inspired by the techniques used by optical designers to minimize lens chromatic aberrations, uses a given number of glass or dielectric plates, whose materials and thicknesses are optimized together with the free-air OPD[9], mutually neutralizing the various dispersion gradients in order to introduce a given achromatic phase difference between the interfering

[9]The free-air OPD is regulated by the so-called delay lines.

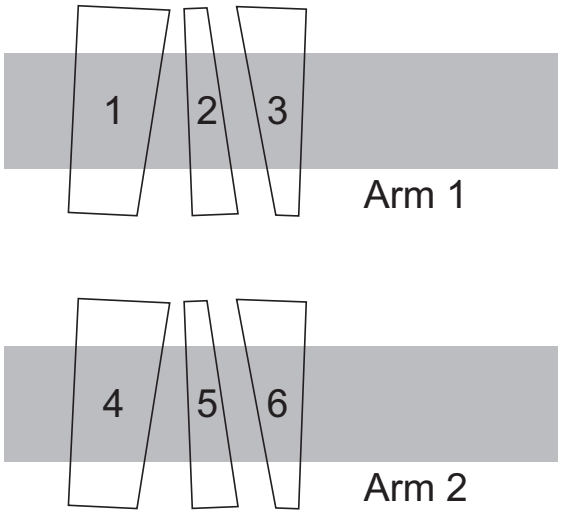

Figure 2.1: APS with dispersive wedged plates. Here is shown a 6-plate arrangement.

wavefronts. In other words, this gives an OPD varying as linearly as possible with wavelength over a given spectral range, resulting in an achromatic phase shift. The number of elements, their thickness and nature can used to minimize the residuals. Fine tuning towards minimum is obtained by slightly adjusting the thickness of some elements either by tilting them in case of parallel plates, or by translating them in case of wedged plates (Mieremet et al. 2000, Fig. 2.1).

Assuming N pairs of e_i-thick plates made of N materials of different refractive index n_i, the expression of the differential phase shift between the considered wavefronts is therefore

$$\Delta\phi(\lambda) = \frac{2\pi}{\lambda}\left(OPD + \sum_{i=1}^{N} e_i\left[n_i(\lambda) - 1\right]\right) \tag{2.12}$$

where OPD is the optical path delay when no plate is inserted. The design optimization rule consists in minimizing the residual phase shift error with respect to its nominal value ϕ, $\epsilon(\lambda) = \Delta\phi(\lambda) - \phi$ over the specified wavelength range. The free parameters are the different thicknesses e_i and the choice of materials with different $n_i(\lambda)$.

2.2.2 Focus-crossing APS

A beam crossing a focus is achromatically phase shifted by π radians, with respect to a parallel beam following an equal optical path. This property allows for an extinction by destructive interference when recombining two such beams, with the additional effect of a centro-symmetric rotation of pupilla on one beam. It is to be noted that the AIC is based on this property for a single aperture (Gay & Rabbia 1996).

The AIC has been validated in the lab and tested on the sky (Baudoz et al. 2000a,b). Nulling can be achieved with a 2-aperture interferometer using this phase-shifting property in one arm. The π-phase shift is obtained means of a cat's eye device (see Fig. 2.2). Since this set-up only uses mirrors, no dispersion occurs and the produced phase shift is intrinsically achromatic.

2.2.3 Field-reversal APS

This method provides a phase shift of π radians, based on the achromatic reversal of the electric-field vector on one of two interfering waves. This reversal is performed by a rotational shearing

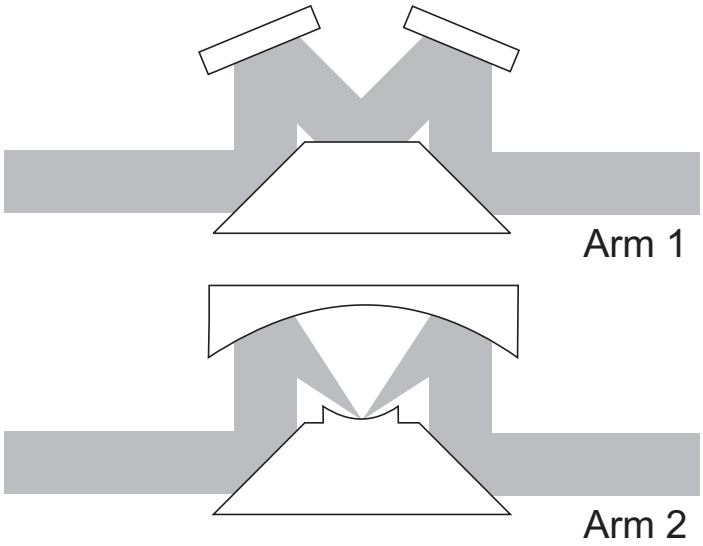

Figure 2.2: Basic set-up for an APS using focus crossing. An extra-focus is inserted in one of the two arms of the interferometer, by means of a cat's eye made of curved mirrors, thus yielding an achromatic phase shift of π. In the other arm, an optical train made of flat mirrors affords the optical path balance.

interferometer with a fixed shear of 180°, by means of rooftop mirrors whose summit lines are seen orthogonal by the wavefronts (hence, the incidence planes are perpendicular). Indeed, arranging two successive 45° reflections in such a way that their incidence planes are orthogonal with respect to each other results in a combination of the s component of the first reflection with the p component[10] of the second one and vice versa, resulting in an effective reversal of the electric-field vector (Serabyn & Colavita 2001, Fig. 2.3).

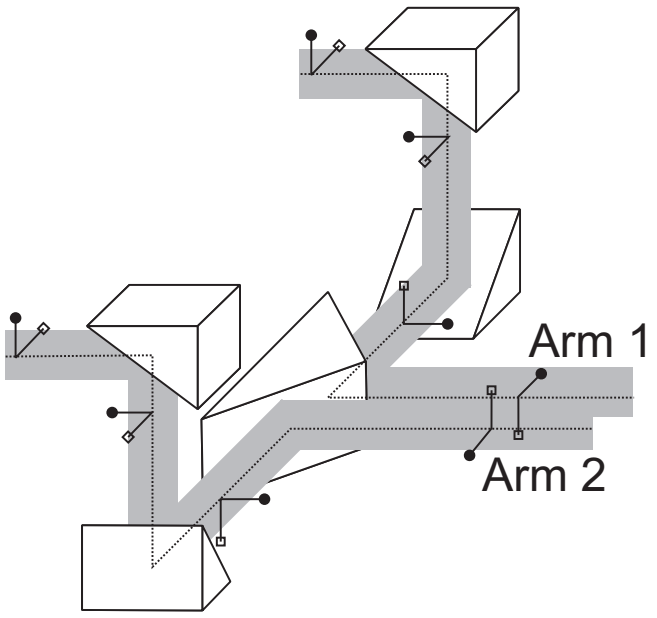

Figure 2.3: Field-reversal APS in the two-periscope configuration.

[10]The components of the electric field parallel and perpendicular to the incidence plane, which is defined by the wave vector and the surface normal, are termed "p-like" (parallel) and "s-like" (senkrecht, i.e., perpendicular in German).

As long as the reflection properties of both mirrors are identical and the angle of incidence is exactly 45° on both mirrors, the electric-field vector reversal is perfect over the full reflective spectral range of the mirror coating. It is to be noted that, like the focus-crossing APS, this setup induces a centro-symmetric rotation of the pupilla of one beam with respect to the other.

2.2.4 Quarterwave-mirror APS

Let us consider a simple boundary separating two dielectric media. Reflectance is identical on both sides except that on one side, the reflected beam undergoes a π-phase shift while it undergoes no phase shift on the other side. The design principle of thin-film APS is based on the extension of this boundary property to a whole multi-layer stack (Lemarquis & Riaud 2003).

2.2.5 Vectorial APS

Implementation of a vectorial phase shift, i.e., a phase retardance taking place between the orthogonal polarization states, is straightforward in a nulling interferometer (resp. phase-mask coronagraph). Considering two identical components belonging to the two distinct interferometer arms 1 and 2 (resp. adjacent quadrants of the FQPM), rotated by ninety degrees around the optical axis and from one another, then the parallel polarization states s_1 and p_2, s_2 and p_1 from the two interferometer arms (resp. ajdacent quadrants) are two by two in phase opposition. It must be noted that there is a strong constraint on the alignment of the components (see Sect. 8.4).

Birefringent plates

This approach uses a set of plane parallel retardation plates made of birefringent crystals, cascaded along the propagation axis and oriented perpendicular to it. The achromaticity is once again given by combining plates of different birefringent materials with properly-chosen thicknesses. Since the dispersion of the birefringence is different for the different materials, it is possible to make the subsequent phase shift achromatic within a given wavelength range (Hariharan 1996). The interest in using naturally birefringent phase shifters and a reason for their commercial success is that natural birefringence is usually two orders of magnitudes smaller than the indices. This constatation implies that thickness errors contribute two orders of magnitude less to the phase shift. On the other hand, very accurate knowledge of the birefringence is necessary.

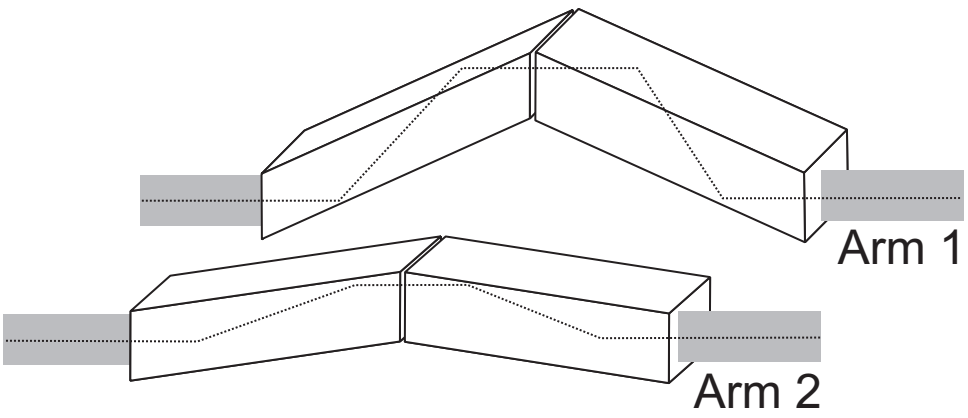

Figure 2.4: Fresnel-rhomb APS. Here is shown the double rhomb configuration implemented in the nulling mode (see Sect. 8.4).

Fresnel rhombs

Fresnel rhombs are conventional devices in optical engineering, used to transform a linearly-polarized incident field into a circularly-polarized one at the output of the device. This transformation results from a $\pi/2$-phase shift performed between the s and p components of the incident field, suitably oriented to have s and p components of equal amplitudes. Cascading two such rhombs usually provides a π-phase shift (Fig. 2.4). In fact, the use of Fresnel rhombs is based on properties of light at total reflection, where the s and p components undergo different phase behaviors. The total internal reflection (TIR) phenomenon comes indeed with a differential phase shift between the vectorial orthogonal polarization states. This vectorial phase shift takes the following form (Born & Wolf 1999a)

$$\Delta\phi_{s-p} = 2\arctan\left[\frac{\cos\theta\sqrt{\sin^2\theta - n_{ti}^2}}{\sin^2\theta}\right] \tag{2.13}$$

where θ is the angle of incidence, greater or equal to θ_c, the critical angle defined as $\sin\theta_c = n_{ti} = n_t/n_i$ and where n_i and n_t are the refractive indices of the incident and external media, respectively. Despite the natural dispersion of the material indices, this phase shift is considered as achromatic for standard applications (Rochford et al. 1997; Anderson 1988). It is to be noted that Fresnel rhombs are also commercially available. It is a well-proven technology.

2.3 Article: *The Four-Quadrant Phase-Mask Coronagraph: white-light laboratory results with an achromatic device*

As an illustration of the implementation of "classical" APS in a nulling setup, let us present the following paper (published in *Astronomy & Astrophysics*) where we report a laboratory experiment assessing the performance of a four-quadrant phase-mask coronagraph implemented with achromatic halfwave plates (see Fig. 2.5, and Mawet et al. 2006).

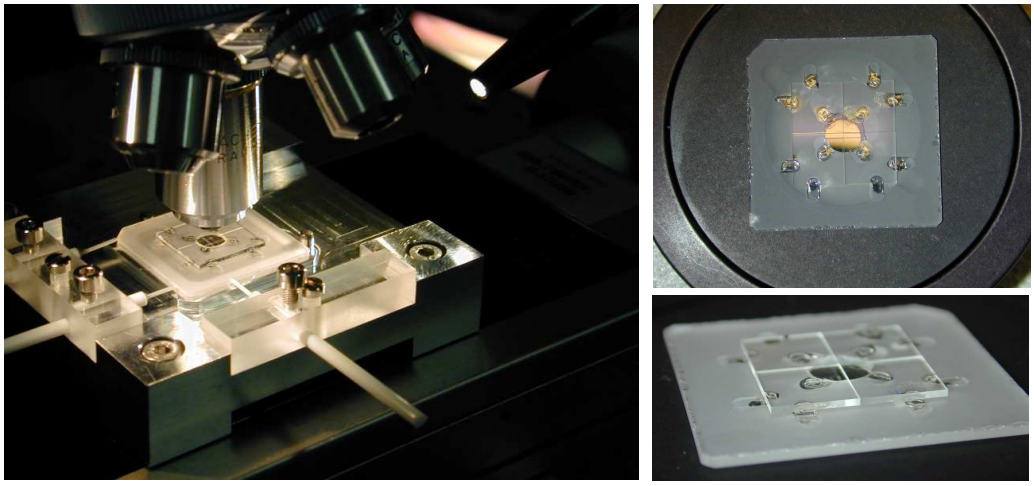

Figure 2.5: Four-quadrant phase-mask with achromatic halfwave plates. Left: mounting pad under microscope. Right: the assembled component, consisting of a silicate substrate with a hole at its center, and the four MgF_2 and four quartz plates glued on its both sides.

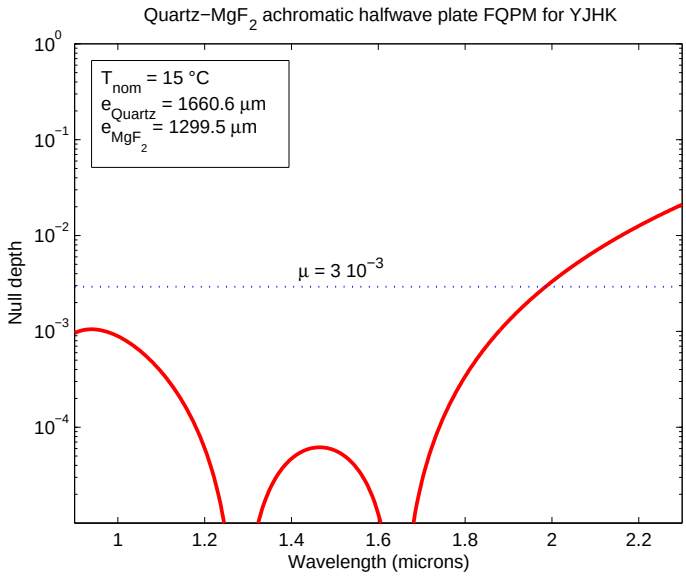

Figure 2.6: Four-quadrant phase mask with achromatic halfwave plates for VLT-PF/SPHERE, from Y to K band. Here is shown the theoretical performance of the component in terms of starlight residuals due to the imperfect destructive interference process, or in other words, the null depth. Note that the plate optimal thicknesses for a nominal temperature of 15°C are also shown.

In this work, we demonstrate the feasibility of the FQPM coronagraph achromatization by means of birefringent elements, paving the way towards more complex solution involving sub-wavelength gratings. The halfwave-plate technique (HWP) was proved more convenient both to manufacture and to implement than the dispersive achromatization such as the dispersive plate concept. As already mentioned, this conclusion comes directly from the fact that the fundamental constraint on the thickness control is relaxed by a factor comparable to the ratio between material indices and birefringences (often $< 10^{-2}$) leading to a control at the 100-nm level versus 1-nm with the dispersive plate scheme. On the other hand, birefringences must be known with a very good precision as well as their variations with temperature (Hale & Day 1988; Etzel et al. 2000) as illustrated in Fig. 2.7 (right).

In our experiment, we selected the Quartz-MgF_2 doublet for application up to the near-infrared domain on ground-based telescopes (up to $2.6\,\mu m$). Results in white light are promising: we obtained a residual level of 10^{-4} at $2.5\lambda/D$. The chromatic residue is sufficiently small to allow the use of this plate stack on a very large spectral range from the R band to the K band for moderate Strehl ratios. We also propose this type of achromatic mask for use on the ESO's VLT-planet finder instrument (SPHERE). For that, a specifically-optimized component has been designed for simultaneous use in the Y, J, H (and K) bands (see Fig. 2.6). Theoretical results are well within specifications and the experimentation in the visible has confirmed the feasibility and the interest of the technique. A new component under assembly for this wavelength regime will soon be tested (see here below).

The drawback of the technique is the very delicate cutting and assembly of the different plates into the four quadrants and on two stacked stages (see Fig. 2.5) but also the huge chromatic defocus induced by the mounting optical thickness (see Fig. 2.7, left). This defocus is responsible for very complicate Fresnel diffraction effects that were already proven to affect the measurements. Moreover, the defocus is different for the two orthogonal polarization states and is wavelength dependent. Another issue concerns the long-term stability of the mounting since the different plates are simply glued on both side of a thick substrate (see Fig. 2.5).

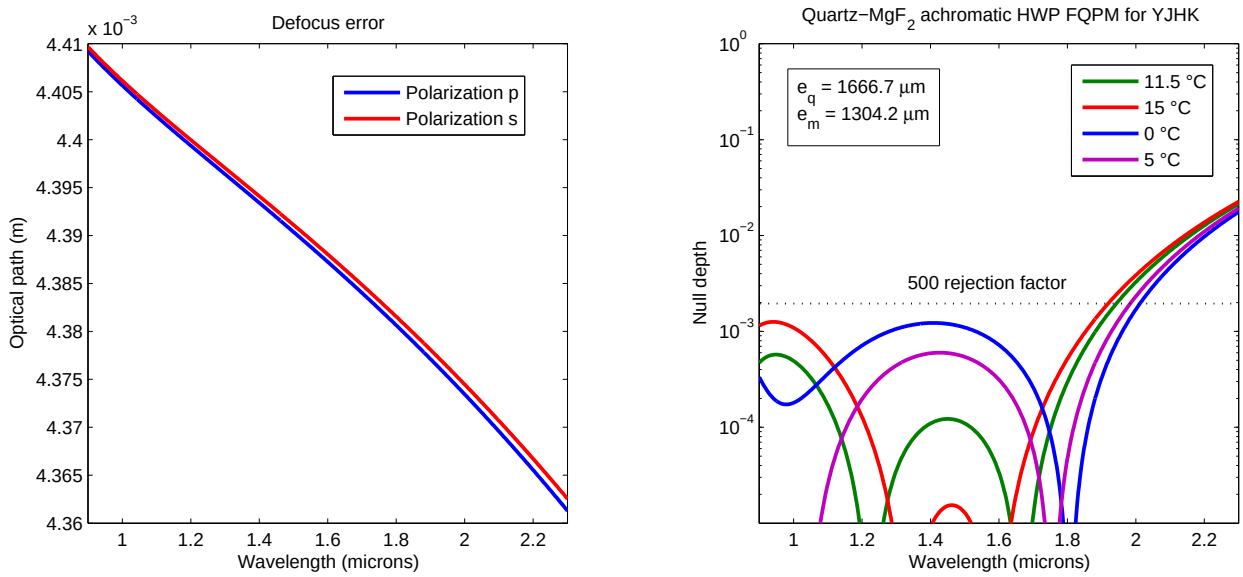

Figure 2.7: Left: defocus error induced by the Quartz-MgF_2 mounting optimized for VLT-PF for Y, J, H and K bands. Right: temperature sensitivity of the same component.

As far as the tests of these achromatic near-infrared components are concerned, a specific optical bench is under assembly at LESIA[11]. This testbed is similar to the one presented in the following paper except that it uses mirrors instead of lenses, i.e., it is a pure reflective bench (see the optical scheme in Fig. 2.8). It will be devoted to the qualification of the VLT-PF/SPHERE coronagraphs and will therefore operate in the H and K bands.

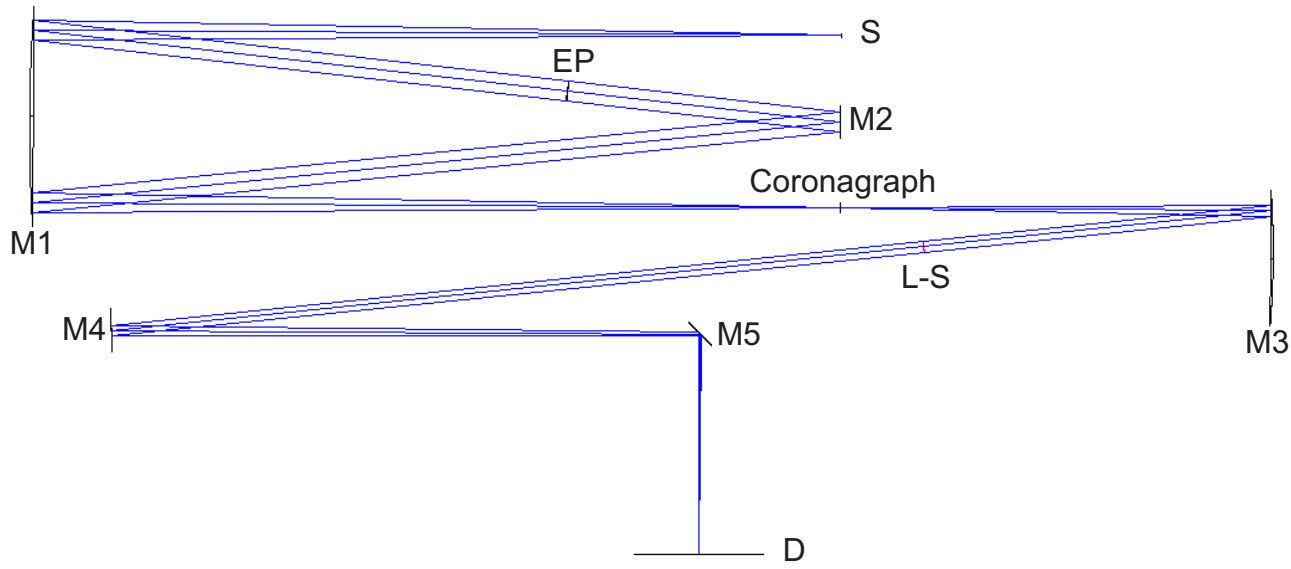

Figure 2.8: Ray-tracing scheme of the H/K-band test bench for the VLT-PF/SPHERE coronagraph qualifications. S is the spatially filtered wideband source (black body at 3400 K). M1, M3 and M4 are paraboloid mirrors with 750 mm, 400 mm and 750 mm focal lengths, respectively. M2 and M5 are plane mirrors. EP is the entrance pupil of diameter $D = 18.75$ mm, providing a $f/D = 40$ at the coronagraphic focal plane, which is the SPHERE specification. L-S is the Lyot stop. Courtesy of Jacques Baudrand and Pierre Riaud.

[11] Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique, Paris-Meudon Observatory.

A&A 448, 801–808 (2006)
DOI: 10.1051/0004-6361:20054158

Astronomy & Astrophysics

The four-quadrant phase-mask coronagraph: white light laboratory results with an achromatic device

D. Mawet[1], P. Riaud[1], J. Baudrand[2], P. Baudoz[2], A. Boccaletti[2], O. Dupuis[2], and D. Rouan[2]

[1] Université de Liège, 17 Allée du 6 Août, 4000 Sart-Tilman, Belgium
[2] LESIA, Observatoire de Paris-Meudon, 5 pl J. Janssen, 92195 Meudon, France
e-mail: mawet@astro.ulg.ac.be

Received 6 September 2005/ Accepted 8 November 2005

ABSTRACT

Achromatic coronagraphs are the subject of intensive research since they will be mandatory for many programs which aim at detecting and characterizing exoplanets. We report a laboratory experiment assessing the performance of the Four-Quadrant Phase-Mask coronagraph (FQPM) over a broadband wavelength range ($R \approx 2$). The achromatization of the FQPM is provided by achromatic halfwave plates (HWP). These phase shifters combine birefringent plates made of different materials with properly chosen thicknesses. The HWP thickness control is relaxed by two orders of magnitudes with respect to the classical (non-birefringent) dispersive plate approach. In our experiment we used a two stage stack of Quartz and MgF_2. This combination allows to cover a large spectral range in the visible (500−900 nm) with a small phase error residual around π ($\approx$0.12 rad rms). With this achromatization, we obtained an attenuation of 755 on the white light PSF peak. This solution is directly applicable to ground-based telescopes using high order adaptive optics such as the ESO's VLT-Planet Finder project and could easily be transposed in the mid-infrared domain for future space-based missions like DARWIN/TPF.

Key words. instrumentation: adaptive optics – techniques: high angular resolution – stars: planetary systems – methods: observational

1. Introduction

Direct detection of faint sources around bright astrophysical objects such as exoplanets orbiting their parent star is very difficult due to the large flux ratio. For example, extrasolar planets are typically 10^4–10^7 times fainter than their host star in the infrared. Therefore, the study of such objects necessitates coronagraphic instruments and nearly perfect wavefronts for optimal operations. Subsequently, this challenging science case requires:

– for Extrasolar Earth-like planets, dedicated space-based observatories like the DARWIN/TPF projects (see Léger et al. 1996, for instance);
– for Extrasolar Giant Planets (EGPs), space-based observatories or Extreme-Adaptive Optics ground-based imaging facilities, like the ESO's VLT-Planet Finder (Mouillet et al. 2003).

A few years ago, we proposed a new device, the Four-Quadrant Phase-Mask coronagraph (FQPM), to advantageously replace Lyot-type amplitude coronagraphs. The FQPM uses a four quadrant π phase shift distribution in the focal plane to provide a self-destructive interference for a centered point like monochromatic source. The FQPM provides significantly smaller working distances than Lyot-type coronagraphs as well as better rejection factors at high Strehl ratios. After presenting the principles and an extensive numerical study (Rouan et al. 2000; Riaud et al. 2001), a third paper (Riaud et al. 2003) reported the laboratory results with a real component in monochromatic light. Since then, the actual performance of the setup has been largely improved: we have obtained a very stable peak attenuation of 10^5 and a total rejection factor of 10^4. The current setup only works accurately for a single wavelength. Observing with a large bandwidth like the K band (2–2.4 μm) with a monochromatic mask would yield a theoretical rejection of only 150. However, for ground-based telescopes using today's state of the art adaptive optics, such as NACO (VLT's NAOS-CONICA, Rousset et al. 1998), a monochromatic phase mask is sufficient for the full K band, since the main limitation comes from the residual uncompensated wavefront errors (Boccaletti et al. 2004).

For higher order corrections such as planned for the ESO's VLT-Planet Finder second generation instrument on the VLT, the chromaticity issue of the FQPM is no longer negligible. Moreover, this instrument requires multi-wavelength operations (0.9–2.5 μm). An achromatic device is thus needed. We therefore propose to implement commercially available achromatic halfwave plates (HWP) in an original way to reproduce the particular FQPM focal plane phase shift distribution. Section 2 describes the principle of birefringent achromatic retarders and presents numerical simulations to assess theoretical

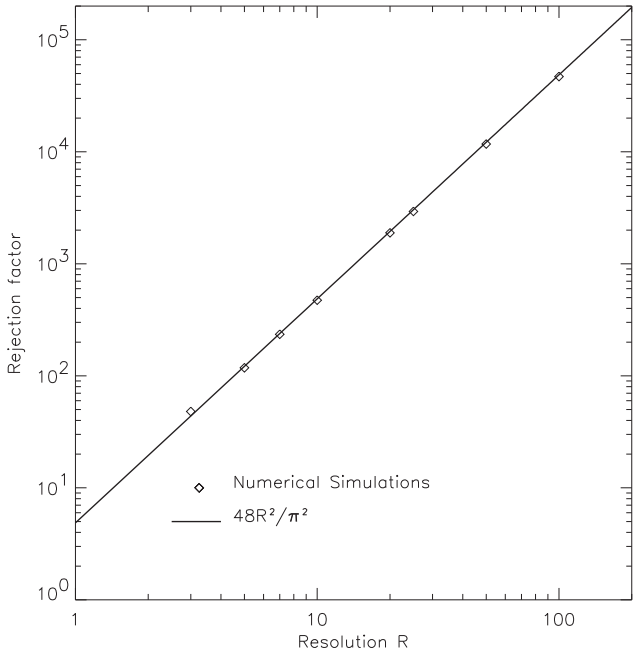

Fig. 1. Rejection τ of a monochromatic component as a function of the spectral resolution R. The continuous line is for the analytical formula whereas the diamonds show coronagraphic numerical simulation results.

limitations of this technique regarding the total rejection. In Sect. 3, after reviewing some manufacturing issues, we present the results of our white light coronagraphic experiment. We discuss in Sect. 4 the possibility of extending this technique to other wavelength ranges, like the mid-infrared domain. We also discuss a very promising technique using synthetic birefringence created on a unique substrate made of any transparent material.

2. Principle

Before presenting the principle of an achromatic FQPM, let us first define the metrics we shall use (Boccaletti et al. 2004). A coronagraph's ability to suppress the on-axis starlight can be quantified with two different parameters:

- *the total rejection*: ratio of total intensity of the direct image to that of the coronagraphic image;
- *the peak attenuation*: ratio of the maximum (on-axis) intensity of the direct image to that of the coronagraphic image.

It is important to note that these metrics are not always related in the same way according to the working conditions.

2.1. Monochromatic phase masks

In the past, our team only manufactured monochromatic phase masks using thin film deposition or etching techniques based on the "index step" principle: a step of height h in a material of refractive index n at the wavelength λ induces a phase shift

$$\Delta\phi = \frac{2\pi}{\lambda}(n-1)h \tag{1}$$

The hyperbolic phase shift dependence in λ makes the technique unapplicable for high performance use over a large spectral bandwidth. In fact, it was shown (Riaud et al. 2003) that the broadband total rejection for a monochromatic mask is

$$\tau \approx \frac{48}{\pi^2}R^2 \tag{2}$$

where $R = \lambda/\Delta\lambda$ is the spectral resolution. Figure 1 shows the total rejection expected for a monochromatic component with respect to the spectral resolution. We clearly notice its rapid deterioration in the broadband cases. For example, in the 500−900 nm wavelength range ($R = 1.75$), the highest total rejection expected with a monochromatic mask is only $\tau = 15$. In this case, the coronagraphic numerical simulation gives a peak attenuation of 35 (see Sect. 3.4).

In order to detect and characterize young EGPs, next generation AO instruments like the ESO's VLT-Planet Finder will require larger stellar rejections for several broadband filters (from J to K). Numerical simulation in H band assuming a Strehl ratio of 80%[1] gave a coronagraphic peak attenuation of about 450. This result was obtained assuming a phase shift error with respect to π of 0.01 rad. This value can be considered as a specification on the mask chromatic residuals since the same simulation with a 0.1 rad error showed an evident degradation of the peak attenuation of about 3.5 after speckle calibration. A specific broadband coronagraph is therefore necessary. Achromatic waveplates are the solution we consider in this paper.

2.2. Achromatic waveplates

Waveplates are optical elements which introduce a phase shift between the polarization components s and p of the incident light. Most of them use the birefringence phenomenon. Birefringence is a natural property of anisotropic crystals but can also be artificially created using one-dimensional subwavelength gratings (Mawet et al. 2005b). The birefringence is defined as follows

$$\Delta n = n_e - n_o \tag{3}$$

where n_o and n_e are the ordinary and extraordinary indices associated with the two privileged directions of vibration, s and p. If the birefringence Δn is negative, the medium is said "negative uniaxial". In this case, one defines the fast (resp. slow) axis as the axis associated with the extraordinary (resp. ordinary) index. Conversely, if the birefringence is positive, the medium is said "positive uniaxial". In this case, the fast (resp. slow) axis is defined as the axis associated with the ordinary (resp. extraordinary) index.

Achromatic waveplates (halfwave, quarterwave,...) are commonly produced by combining two plates of different birefringent materials with properly chosen thicknesses. Since the dispersion of the birefringence is different for the two materials, it is possible to make Optical Path Difference (OPD) values linear with λ within a given wavelength range. Hence,

[1] Expected performance for the 1340 actuators of the VLT-PF Adaptive Optics.

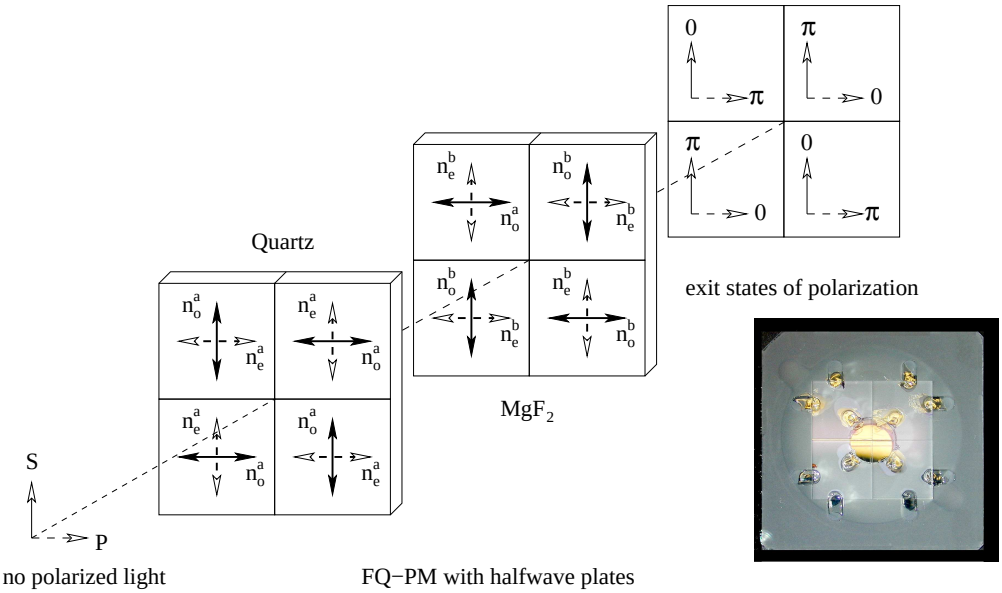

Fig. 2. Principle of a FQPM using halfwave plates. Each individual quadrant consists of a two material stack (Quartz, MgF_2). Two quadrants along one diagonal are rotated by 90° around their normals with respect to the two others. The subscript a and b stand for the Quartz and MgF_2 plates whereas the o and e ones, for the ordinary and extraordinary indices, respectively. The bottom right picture shows the final component assembled on a silicate substrate. Fixation points (optical glue) are clearly visible at the waveplate external edges. The 8 mm-diameter circular working area is at the center of the component.

the retardation of the resulting waveplate can be made little sensitive to the wavelength (Hariharan 1996). We consider a combination of two such birefringent plates of thicknesses d_a and d_b and of birefringences Δn_a and Δn_b. The usual condition for achromatism is that the retardation of the system $\Delta\phi = \frac{2\pi}{\lambda}(d_a\Delta n_a + d_b\Delta n_b)$ should be equal to half a wave (for a halfwave retarder) at two selected wavelengths λ_2 and λ_3 while minimizing the phase shift error with respect to the chosen retardation value over the $\lambda_1 - \lambda_4$ wavelength range. Therefore we have

$$d_a\Delta n_a(\lambda_2) + d_b\Delta n_b(\lambda_2) = \frac{\lambda_2}{2} \quad (4)$$

$$d_a\Delta n_a(\lambda_3) + d_b\Delta n_b(\lambda_3) = \frac{\lambda_3}{2} \quad (5)$$

$$\epsilon = \min \int_{\lambda_1}^{\lambda_4} |\Delta\phi(\lambda) - \pi|^2 \, \mathrm{d}\lambda \quad (6)$$

where $\Delta n_a(\lambda_2)$, $\Delta n_b(\lambda_2)$, $\Delta n_a(\lambda_3)$ et $\Delta n_b(\lambda_3)$ are, respectively, the values of the birefringence of the two materials at these two wavelengths. A solution for Eqs. (4)–(6) can be obtained by combining a material with a positive birefringence with one having a negative birefringence. Conversely, two materials whose birefringences have the same sign can be combined if their fast axis are set perpendicular. Let us mention the special case of the so-called "superachromatic" retarder (Frecker et al. 1976): it consists of three pairs of two materials. This configuration is very performing but is bulky and would therefore induce a huge defocus error and a ghost problem (12 interfaces).

The interest in using naturally birefringent phase shifters and a reason why they are commercially available comes from the fact that, usually, the birefringence is two orders of magnitudes smaller than the indices. So a thickness error contributes to the phase shift two orders of magnitude less. On the other hand, we need very accurate measurements of the birefringence for the used materials, over the range of wavelengths to be covered. This is generally the case for commercially available devices.

2.3. FQPM with waveplates

This section explains how the concept of polarization phase shifters is adapted to the FQPM. Let *s* and *p* be the global polarization components of the incoming light. In each of the four quadrants made of cut halfwave plates, the *s* and *p* global polarization states are projected according to the local fast and slow axis orientations of the two-level stack. We have already seen that two indices, the ordinary index n_o and the extraordinary index n_e, can be assigned to these directions. Let us now assume that the four cut quadrants are strictly identical (same two-level stack) and implemented in the following way: two of them in two opposite quadrants along one diagonal are rotated by 90° around their normals with respect to the two others. This antisymmetrical configuration mimics the FQPM particular focal plane π-phase distribution for each parallel and potentially interfering state of polarization (see Fig. 2). Such a design works with natural light. In practice, all plates have to be polished and cut in the same material blank to obtain homogenous components in terms of refractive index and surface quality.

Another way to make the FQPM achromatic was proposed and tested by Baba et al. (2002): the polarization interferometric coronagraph. It is in fact a FQPM whose phase shift is provided by a Liquid-Crystal (LC) device sandwiched between two crossed polarizers. This method allows a broadband use

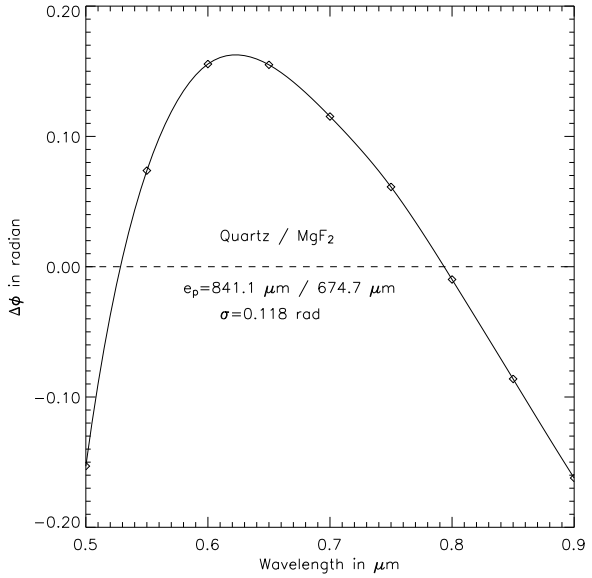

Fig. 3. Numerical simulation result for achromatic halfwave plates. The curve shows the phase residuals around π in radian versus wavelength (microns). The diamonds indicate the points where the birefringences have been precisely measured by the manufacturer. The phase error standard deviation over the full wavelength range is 0.118 rad rms.

of the coronagraph, but only with linearly polarized light in the visible.

2.4. Theoretical performance

The selected retarder in our experiment is a combination of Quartz (SiO_2) and MgF_2 crystal plates. Quartz and MgF_2 are often associated for the good matching of their respective birefringence dispersions. Both are transparent from the ultraviolet to the near infrared, have low indices (low Fresnel losses) and can be easily polished (low wavefront errors). The mean birefringences over the wavelength range (λ_1(=500 nm)−λ_4(=900 nm)) are respectively $\Delta n_a \approx 0.00900$ for Quartz (positive uniaxial) and $\Delta n_b \approx -0.01190$ for MgF_2 (negative uniaxial). Note that birefringence values at the 10^{-5} level of precision can be easily found in the literature for these materials. For the working spectral range, we found the following couple of optimized thicknesses: $d_a = 841.1\,\mu$m for Quartz and $d_b = 674.7\,\mu$m for MgF_2. In Fig. 3, we present the residual phase shift error obtained for this optimal solution, with a standard deviation $\sigma = 0.118$ rad rms over the full spectral band. The total rejection for a nulling phase mask coronagraph being related to the phase error standard deviation σ by $\tau = \frac{4}{\sigma^2}$ (Riaud et al. 2003), we therefore obtain $\tau \approx 290$ assuming phase errors only.

3. Laboratory results

3.1. Achromatic FQPM assembly

The achromatic phase mask assembly is mounted on a silica substrate 1 mm thick with a central hole of 8 mm in diameter. The four Quartz plates are assembled on one side and the four remaining MgF_2 plates on the other one. According to the implementation scheme presented here above, the π phase shift between adjacent quadrants is provided by rotating the fast axis of two quadrant stacks along one diagonal by 90° around their normals. Each of the 8 individual plates is polished ($\lambda/10$ PTV at 550 nm) and cut parallel to the optical axis with a micrometric precision. The edge parallelism cutting error with respect to each plate fast/slow axis is below 30 arcsecs. All plates are anti-reflection coated for the considered bandwidth (reflectivity <1%). One of us (O. D.) assembled the eight plates with a high precision (<10 μm) while respecting orientations, alignments and coplanarity (<10 arcmin) for the two stages. This task revealed to be practically difficult and long to achieve within the specifications. For example, the coplanarity issue was overcome only by performing the assembly in a clean environment and under in situ interferometric metrology. Indeed, a 4 μm dust particle, for instance, can induce a 14 arcmin out-of-plane deviation.

In order to hide the imperfections at the edges resulting from the cutting process, we added a thin opaque spider (a tungsten wire 18 μm in diameter) over the four quadrant transitions. Indeed, the rough cut edges at the transitions would otherwise diffract too much light inside the geometric pupil area.

3.2. Optical setup

The optical bench is the one used for the FQPM monochromatic tests (see Riaud et al. 2003). The white light source is a halogen lamp at a temperature of 3400 K (100 W) and the setup is feeded with a 4 μm core fiber (single mode for wavelengths longer than 620 nm). All lenses are in silica (infrasil 301) with low frequency surface errors lower than $\lambda/20$ PTV or $\lambda/80$ rms across 10 mm in diameter and AR coated for use between 400 and 1100 nm (reflectivity <1% per interface). The entrance pupil is a hole with a diameter of 1.78 mm. In this configuration the $f-$number is 253 and the size of the Airy pattern projected on the achromatic phase mask is $\lambda/d = 176\,\mu$m at 700 nm. This size is one order of magnitude above the precision of the plate adjustment (≈10 μm). Finally, to filter the coronagraphic diffraction at the edge of the pupil, we used a diaphragm of 350 μm in diameter, designed to undersize the pupil by about 23% in diameter. This size for the Lyot stop (i.e. nearly 80%) corresponds to a trade-off between the diffraction residuals filtering and global throughput (see Riaud et al. 2001). The Strehl ratio estimated for the 532 nm PSF image is 99.8%, which is equivalent to $\lambda/180$ rms for the total wavefront error.

3.3. Acquisition and data processing

All images were recorded with a ST8 XE Sbig camera equipped by a KAF 1600 E CCD chip cooled by one Peltier module. The sampling in our configuration (3 × 3 binning) is 43 pixels per λ_0/d, with $\lambda_0 = 700$ nm being the central wavelength of the 500−900 nm bandpass filter. To assess the broadband coronagraphic performance of the component, we first

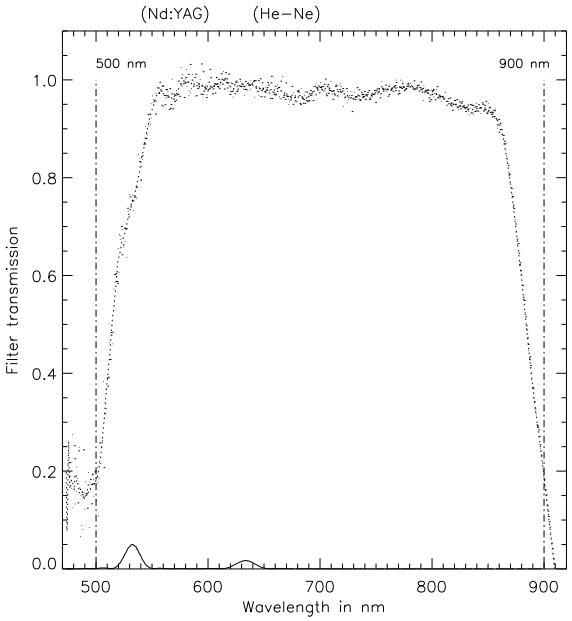

Fig. 4. This figure presents the spectral filter response as measured on our coronagraphic bench. The spectral sampling is 0.55 nm/pixel. The two peaks below, convolved with the PSF, correspond to calibration lines obtained with a YAG:Nd and a He − Ne laser, respectively. The maximum transmission of the filter is 96 ± 3%.

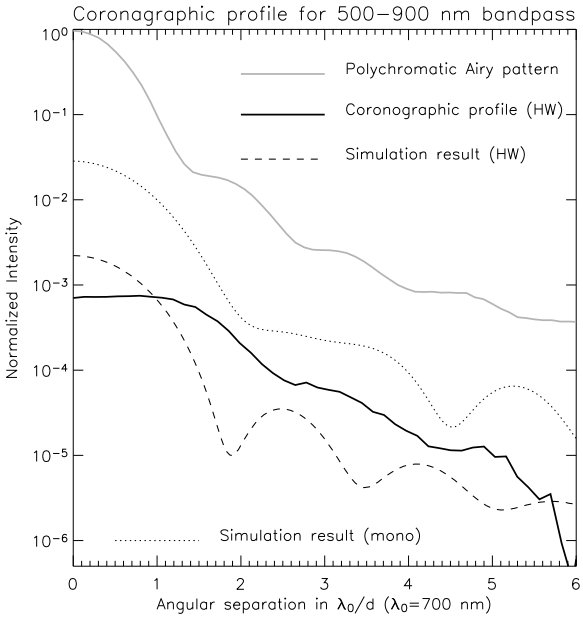

Fig. 5. Experimental and theoretical coronagraphic profiles: the grey solid curve is the experimental PSF for the 500−900 nm bandpass. The total exposure time for the PSF is 60 s. The continuous black line shows the coronagrapic profile obtained with our achromatic waveplate mask. The coronagraphic image was obtained adding forty exposures of 30 s each. The dashed line presents the coronagraphic simulation results taking into account the waveplate phase residuals, the defocus error, the spectral response of the halogen lamp and the camera. For comparison, the dotted line presents the simulation results for a monochromatic mask used in the same conditions. All curves are azimuthally averaged (for this reason, the on-axis attenuation seems to be greater than the measured peak attenuation value of 755).

acquired a coronagraphic image with the white halogen lamp and the bandpass filter characterized in Fig. 4. The exposure time was about 30 s. Forty exposures were co-added to improve the signal-to-noise ratio up to 1000. The next step was to record direct non-coronagraphic images in the same conditions but with a shorter exposure time to avoid saturation (1 s). All images were then subtracted with a median dark frame. Results in terms of coronagraphic rejection normalized profiles are displayed in Figs. 5 and 7.

The visible waveplates are optimized for the 500−900 nm wavelength range but they are also quite good for the 700−1000 nm range. Therefore, we have also tested the component with a Schott RG 645 highpass filter with a maximum transmission of 91%. The cutoff at long wavelengths is ensured by the CCD low sensitivity around 1.1 μm. In Fig. 7, the corresponding coronagraphic rejection profile is also shown.

3.4. Performance

The attenuation on the stellar peak is about 755 for the 500−900 nm bandpass (see Fig. 5) and 561 for the 700−1000 nm one (see Fig. 7). The total rejection is 294 for the first bandpass and 256 for the second one. The results we obtained indicate that the attenuation on the central peak is mostly limited by the mask residual chromatism rather than the surface roughness ($\approx\lambda_0/1000$ rms) or the size of the source. Indeed, before each polychromatic measurements, the optical bench was calibrated with a monochromatic mask and a laser diode source, routinely giving a peak attenuation of about 10^5 (see Riaud et al. 2003). Moreover, the measurements of the total rejection are in agreement with the theoretical performance (see Sect. 2.4). In Fig. 6, we notice, in agreement with the coronagraphic profile of Fig. 5, that the residual level of 10^{-4} is quickly reached at $2.5\lambda/d$.

The shape of the polychromatic coronagraphic image (Fig. 6) is similar to the classical FQPM case except that it is somewhat blurred compared to the monochromatic one (Riaud et al. 2003). To understand the blurring effect due to the large bandwidth ($R \approx 1.75$), we simulated a polychromatic PSF passing through the coronagraph in order to quantify the residuals. In the Fourier transform calculations, we took into account the size of the pupil for each wavelength as well as the diameter of the Lyot stop in the relayed pupil. The dashed line in Figs. 5 and 7 presents simulation results taking into account the cumulative effects of the halfwave plates phase residuals, defocus errors induced by the use of simple lenses (no achromats) and the presence of a small spider on the mask transitions. For comparison, the dotted line presents the simulation results for a monochromatic mask used in the same conditions.

Given the spectral response of our Halogen lamp and bandpass filter, the calculated total rejection of the achromatic coronagraphic device for the bandpass 500−900 nm is 340 and the expected attenuation on the stellar peak is roughly 450. The first value (total) is not far from measurements (294) whereas the second one (peak) is better in practice (755). For comparison, the theoretical total rejection for the monochromatic mask

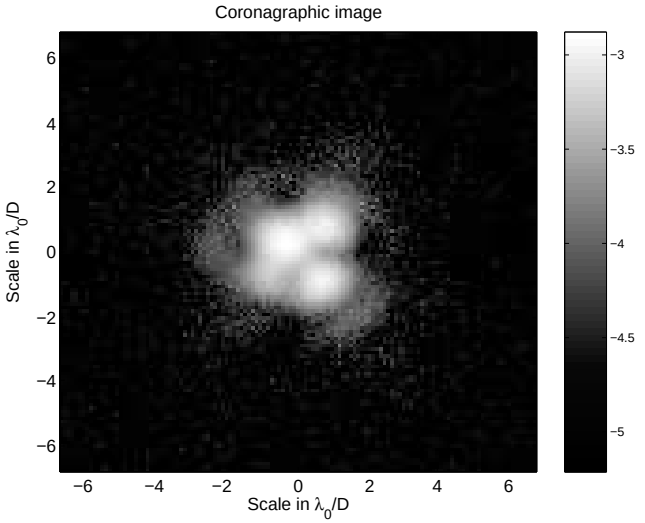

Fig. 6. Polychromatic coronagraphic residual image (logarithmic scale) for 500−900 nm bandpass. We use a 77% Lyot stop for spatial filtering. The angular resolution is given for $\lambda_0 = 700$ nm and for the entrance pupil (full pupil without diaphragm). The classical FQPM four-spot pattern is easily recognizable though somewhat blurred.

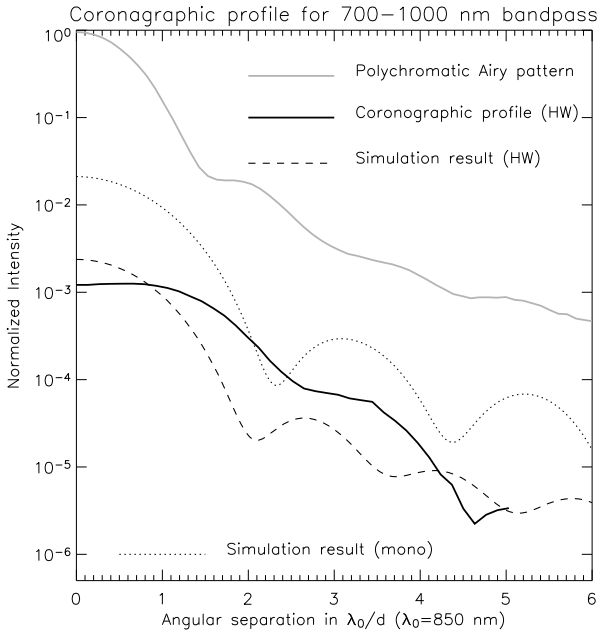

Fig. 7. Experimental and theoretical coronagraphic profiles: the grey solid curve is the experimental PSF for the 700−1000 nm bandpass. The total exposure time for the PSF is 60 s. The continuous black line shows the coronagrapic profile obtained with our achromatic waveplate mask. The coronagraphic image was obtained adding forty exposures of 30 s each. The dashed line presents the coronagraphic simulation results taking into account the waveplate phase residuals, the defocus error, the spectral response of both the halogen lamp and the camera. For comparison, the dotted line presents the simulation results for a monochromatic mask used in the same conditions. All curves are azimuthally averaged (for this reason, the on-axis attenuation seems to be greater than the measured peak attenuation value of 561).

in the 500−900 nm range is only 35. The discrepancy between the simulated and measured peak attenuation value could be explained by complex Fresnel diffraction effects induced by the spider intended to mask cutting imperfections. Indeed, classical Fourier propagation is not able to reproduce the four-spot shape of the actual broadband coronagraphic image (see Fig. 6). It is to be noted that the small width of the spider wires (18 μm in diameter corresponds to $\lambda/7d$ at $\lambda = 500$ nm) does not affect the optical throughput nor the inner working distance.

The ratio between the two values (peak and total) is low in the polychromatic case (≈1.32) compared to previous ratio obtained in the narrow band case with a laser diode (≈10, see Riaud et al. 2003). This observation is directly related to the polychromatic blurring effect.

For an actual stellar source, the rejection factor could be more important, depending on the maximum emissivity of the star. For example with the proposed 500−900 nm achromatization, a G2V and a M2V type stars would give better rejection factors. Indeed the maximum black body emissivity for these stars coincides with the two points of exact π phase shift (see Fig. 3).

It is interesting to compare the results of our achromatic FQPM to the polarization interferometric coronagraph proposed by Baba et al. (2002). The peak attenuation is 6.5 better in our case but this value has to be balanced because of our slightly smaller bandpass (500−900 nm compared to 370−830 nm). It must also be noted that the throughput of the experiment presented in Baba et al. (2002) is limited by the two polarizers to only <25%.

4. Perspectives

4.1. Near infrared

In the framework of the second generation instrumentation for the VLT (VLT-Planet Finder project) and following the results presented hereabove, we have proceeded with the manufacturing of near infrared components, using the same birefringent combination. Indeed, the Quartz transparency is well suited for wavelength ranges up to 2.6 μm and MgF_2 is transparent up to 8.5 μm. In Table 1, we present three different optimizations in the range 700−2500 nm. The first one which covers the whole spectral range was assembled and will be tested soon. The large spectral range is achieved at the cost of a lower rejection factor. However, this component is well adapted to an instrument like NACO for which Strehl ratios of 50% are routinely obtained.

The situation is quite different for VLT-PF, an instrument specifically optimized to image and characterize EGPs. For that, VLT-PF will include several coronagraphs each dedicated to a specific set of filters. The principle of detection relies on the Simultaneous Differential Imaging (SDI) method which consists in subtracting images taken simultaneously at different wavelengths (Marois et al. 2005). Such an instrument is already included in NACO although without coronagraphic capabilities (Hartung et al. 2004). This method provides self

Table 1. Halfwave plate achromatization for various wavelength ranges from the *V* band up to the *N* band. The table summarizes the optimized thicknesses for the chosen doublet materials. The last line gives the total rejection factor.

Filters	500–900 nm ($R = 1.75$)	700–2500 nm ($R = 0.9$)	H ($R = 4.7$)	K ($R = 5.5$)	N ($R = 4.8$)
Couple (a/b)		Quartz/MgF_2			CdS/CdSe
d_a (μm)	841.1	1488.7	1332.2	1219	1798.1
d_b (μm)	674.7	1168	1057.9	977.1	1618.8
τ (total rejection)	340	260	7900	1.47×10^5	2.16×10^5

reference subtraction while minimizing speckle residuals. It is expected to significantly improve the direct detection of EGPs and brown dwarfs. Basically, two or more wavelengths are observed simultaneously like for instance in the *H* band: 1.575, 1.600 and 1.625 μm. The 1.625 μm line corresponds to the methane feature which, if present, shall be revealed by subtraction (stars have no methane signature contrary to low mass objects). In the *K* band, the same method could also determine EGP CO_2/CH_4 proportions, which would constrain models of their atmospheres (Burrows et al. 2004; Chabrier et al. 2004; Sudarsky et al. 2003). However, the SDI method requires very low phase errors for proper subtraction of the adjacent wavelength narrow band images. For that reasons, we have searched for optimal solutions in the *H* and *K* bands and the results are presented in Table 1. Larger rejection rates are clearly feasible when the spectral range is narrower since the chromatic phase residual is also reduced.

4.2. Towards the thermal infrared

Earth-like planet detection is optimal at the planet's maximum emissivity, i.e. in the thermal infrared, around 12 μm. One of the most relevant biosignature is the strong Ozone spectral feature (DesMarais et al. 2003), around 9.65 μm. For this reason, the *N* band appears to be a well-suited choice.

Infrared nulling interferometry (Bracewell 1978), considered for space-based missions like DARWIN/TPF (Léger et al. 1996) or ground-based experiments like Keck-I (Serabyn et al. 2004), is an interesting technique to achieve both high angular resolution and high contrast detection/characterization of exoplanets. This technique consists in adjusting the phases of the beams coming from various telescopes (two in the most simple configuration) to produce a pupil plane destructive interference on the optical axis. The phase shift (π, for instance) may be provided in broad bands thanks to achromatic HWP.

An alternative to nulling interferometry is to recombine coherently all the telescopes in a single image plane. The HWP achromatic FQPM should then be regarded as an interesting solution if implemented at the Fizeau or densified focus of such an interferometer (Riaud et al. 2002). This approach would necessitate a minimum of three telescopes to be efficient (the main limitation of this method would be the cross-talk between sub-pupils). Preliminary numerical simulations show starlight attenuations of more than 10^6 for a dilution factor (ratio of the interferometer baseline to the sub-pupil diameter) greater than 15 (Riaud et al. in preparation).

However, in the mid-infrared domain, the lack of birefringent materials would be the major problem for both approaches. Moreover, the few ones that can be found have large refractive indices, thereby inducing more losses due to spurious Fresnel reflections. Nonetheless, the couple CdS/CdSe is a good candidate to perform achromatization optimizations in the thermal infrared. The phase shift residuals for this doublet with optimized thicknesses give a total rejection $\tau = 2.1 \times 10^5$ for a theoretical spectral resolution $R = 2.75$ in the *N* band (9−13 μm, see Table 1). If a third material was incorporated, the equivalent total rejection could reach 10^6. As said before, there are not many candidates, but a solution with $AgGaSe_2$, for instance, is feasible.

4.3. Artificially birefringent waveplates

Another solution to achromatize the phase shift of the FQPM relies in the subwavelength grating technology. The period of subwavelength gratings is smaller than the wavelength of the incident light. They do not diffract light as classical spectroscopic gratings in the sense that only the zeroth transmitted and reflected orders are allowed to propagate outside the grating region, leaving incident wavefronts free from any further aberrations. For this reason, they are often called Zeroth Order Gratings (ZOGs). ZOGs present very specific properties: 1-D modulated ZOGs artificially create unique anisotropic and dispersive characteristics which can be used to synthesize achromatic waveplates from isotropic materials. This phenomenon is referred to as form birefringence. Two extensive studies on the implementation and optimization of such structures in coronagraphy were recently presented (Mawet et al. 2005a; Mawet et al. 2005b). Results of these studies are very promising with total rejections larger than 10^5 for usual astrophysical bandfilters. The ZOG technology does not require any delicate assembly since the four gratings are engraved on a unique substrate. This solution is monolithic and therefore much more thermally stable. Furthermore, this technology is sufficiently flexible to accommodate a great variety of dielectric materials and is very well adapted to mid-infrared wavelength ranges.

5. Conclusions

We have demonstrated the feasibility of the FQPM coronagraph achromatization by means of birefringent elements. The halfwave plates technique was proved more convenient both to manufacture and to implement than the dispersive achromatization such as the dispersive plate concept (Mieremet et al. 2000). This conclusion comes directly from the fact that the fundamental constraint on the thickness control is relaxed by a factor comparable to the ratio between material indices and

birefringences (often $<10^{-2}$) leading to a control at the $\approx$100 nm level versus 1 nm with the dispersive plate scheme.

In our experiment, we selected the Quartz-MgF_2 doublet for application up to the near infrared domain on ground-based telescopes (until 2.6 μm). Results in white light are promising: we obtained a residual level of 10^{-4} at $2.5\lambda/d$. The chromatic residue is sufficiently small to allow the use of this plate stack on a very large spectral range from the *R* band to the *K* band for moderate Strehl ratios. We also propose this type of achromatic mask optimized for the ESO's VLT-Planet Finder project. Theoretical results are well within its specifications and the experimentation in the visible has confirmed the feasibility and the interest of the technique.

We also expect to manufacture a monolithic component made of artificially birefringent waveplates using the more flexible ZOG technology. This would open further possibilities in high contrast imaging for ground/space-based facilities like the VLT-PF and TPF/DARWIN projects.

Acknowledgements. D.M. acknowledges the financial support of the Belgian "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture". P.R. acknowledges the financial support of the "Pôle d'Attraction Inter-Universitaire". We also are very grateful to the financial support of the "BQR" contract from the Paris-Meudon Observatory. We warmly acknowledge Olivier Absil and the anonymous referee for useful comments.

References

Baba, N., Murakami, N., Ishigaki, T., & Hashimoto, N. 2002, Opt. Lett., 27, 1373

Boccaletti, A., Riaud, P., Baudoz, P., et al. 2004, PASP, 116, 1061

Bracewell, R. N. 1978, Nature, 274, 780

Burrows, A., Sudarsky, D., & Hubeny, I. 2004, ApJ, 609, 407

Chabrier, G., Allard, F., Baraffe, I., Barman, T., & Hauschildt, P. H. 2004, in Extrasolar Planets: Today and Tomorrow, ed. J.-P. Beaulieu, A. Lecavelier des Etangs, & C. Terquem, Proc. ASP Conf., 321, 131

DesMarais, D. J., Lin, D., Harwit, M., et al. 2001, Biosignatures and Planetary Properties to be Investigated by the TPF Mission, JPL Publication 01-008, California Inst. Of Technology, Pasadena, CA, 57

Frecker, J. E., & Serkowski, K. 1976, Appl. Opt., 15, 605

Hariharan, P. 1996, Opt. Engineering, 35, 3335

Hartung, M., Herbst, T. M., Close, L. M., et al. 2004, A&A, 421, L17

Léger, A., Mariotti, J. M., Mennesson, B., et al. 1996, Icarus, 123, 249

Marois, C., Doyon, R., Nadeau, D., et al. 2005, PASP, 117, 745

Mawet, D., Riaud, P., Absil, O., & Surdej, J. 2005a, ApJ, 633, 1191

Mawet, D., Riaud, P., Surdej, J., & Baudrand, J. 2005b, Appl. Opt., 44, 7313

Mieremet, A. L., Braat, J. J., Bokhove, H., & Ravel, K. 2000, in Interferometry in Optical Astronomy, ed. P. J. Léna, & A. Quirrenbach, Proc. SPIE, 4006, 1035

Mouillet, D., Fusco, T., Lagrange, A.-M., & Beuzit, J.-L. 2003, EAS Pub. Ser., 8, 193

Riaud, P., Boccaletti, A., Rouan, D., Lemarquis, F., & Labeyrie, A. 2001, PASP, 113, 1145

Riaud, P., Baudrand, J., Boccaletti, A., & Rouan, D. 2003, PASP, 115, 712

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

Rousset, G., Lacombe, F., Puget, P., et al. 1998, in Adaptive Optical System Technologies, ed. D. Bonaccini, & R. K. Tyson, Proc. SPIE, 3353, 508

Serabyn, E., Booth, A. J., Colavita, M. M., et al. 2004, in New Frontiers in Stellar Interferometry, ed. W. Traub, Proc. SPIE, 5491, 806

Sudarsky, D., Burrows, A., & Hubeny, I. 2003, ApJ, 588, 1121

Part II

Subwavelength gratings

3

Theory and manufacturing of subwavelength gratings

Contents

Abstract. *As the spatial period of a grating tapers off with respect to the wavelength of the incident electromagnetic wave, the interaction between the field and the structural modulation becomes more and more complex. From resonance effects to homogenization, the grating acquires very special properties that were historically first considered as defects but are now seen as keys to engineer new optical media. In the subwavelength domain, scalar theories of diffraction fail so that the vectorial nature of light must be taken into account. Several approaches to rigorous diffraction theory exist. The most popular one is the so-called rigorous coupled-wave analysis (RCWA), which is a resolution of Maxwell's equations in the frequency space. We chose to use RCWA for its efficiency to model the type of gratings under study in this work. This chapter is devoted to the mathematical formulation of RCWA, as well as the description of subwavelength gratings, together with the techniques to manufacture them.*

3.1 Diffraction by a grating

3.1.1 Scalar diffraction theory

For a large variety of diffractive elements such as classical gratings, it is possible to consider the wavefield as scalar. The hypothesis is to only consider one transverse component of the electric or magnetic field. The mathematical treatment starts with the time-independent Helmholtz equation derived from the Maxwell's equations

$$(\nabla^2 + n^2k^2)U(x,y,z) = 0 \tag{3.1}$$

In order to describe wave propagation, the scalar field $U(x,y,z)$ is decomposed into plane waves by means of a Fourier transformation

$$U(x,y,z) = \int\int \tilde{U}(\nu_x,\nu_y,\nu_z)e^{2\pi i(\nu_x x+\nu_y y+\nu_z z)}d\nu_x d\nu_y d\nu_z \tag{3.2}$$

ν_x, ν_y and ν_z are called spatial frequencies and denote the propagation directions of the components of the wavefield. Inserting Eq. 3.2 into Eq. 3.1, we derive the differential equation

$$\frac{\partial^2\tilde{U}(\nu_x,\nu_y,\nu_z)}{\partial z^2} + k^2\left[1-\lambda^2(\nu_x^2+\nu_y^2)\right]\tilde{U}(\nu_x,\nu_y,\nu_z) = 0 \tag{3.3}$$

By solving this equation, we obtain the general equation for light propagation in a homogeneous medium

$$U(x,y,z) = \int\int\int\int U(x_0,y_0,z_0)e^{2\pi i\left[\nu_x(x-x_0)+\nu_y(y-y_0)+\frac{z-z_0}{\lambda}\sqrt{1-\lambda^2(\nu_x^2+\nu_y^2)}\right]}d\nu_x d\nu_y dx_0 dy_0 \tag{3.4}$$

This equation describes the complex amplitude propagation from an object plane z_0 to a plane z. In the paraxial approximation, i.e., for small angles θ to the optical axis, this equation can be written as

$$U(x,y,z) \propto \frac{i}{\lambda}\int\int U(x_0,y_0,z_0)\frac{e^{ikr}}{r}\cos\theta dx_0 dy_0 \tag{3.5}$$

Equation 3.5 is known as Kirchhoff's diffraction integral. It mathematically describes Huygens principle, which states that the wavefield at a location z behind a diffraction screen is described by a superposition of spherical waves $\frac{e^{ikr}}{r}$ originating at the occulting screen.

3.1.2 Fresnel and Fraunhofer diffraction

Kirchhoff's diffraction integral can be approximated to describe the two most important regimes in scalar diffraction theory. In order to describe near-field diffraction or "Fresnel diffraction", we approximate the distance r between object and observation plane by a quadratic Taylor expansion

$$r = (z-z_0)\left[1+\left(\frac{x-x_0}{z-z_0}\right)^2+\left(\frac{y-y_0}{z-z_0}\right)^2\right]^{1/2} \tag{3.6}$$

$$\approx (z-z_0)\left[1+\frac{1}{2}\left(\frac{x-x_0}{z-z_0}\right)^2+\frac{1}{2}\left(\frac{y-y_0}{z-z_0}\right)^2\right] \tag{3.7}$$

In this case, corresponding to a paraboloidal approximation of spherical waves, the diffraction integral is approximated by a Fresnel transform

$$U(x,y,z) \approx \frac{i}{\lambda}\frac{e^{ik(z-z_0)}}{z-z_0}\int\int U(x_0,y_0,z_0)e^{\frac{i\pi}{\lambda(z-z_0)}\left[(x-x_0)^2+(y-y_0)^2\right]}dx_0dy_0 \tag{3.8}$$

If we approximate Kirchhoff's integral for even larger distances $(z-z_0)$ we reach the regime of "far-field" of "Fraunhofer" diffraction. This approximation of spherical waves by plane waves is valid if $z-z_0 \gg \frac{1}{2}k(x_0^2+y_0^2)$. In this case, the diffraction integral can be written as

$$U(x,y,z) \propto \frac{i}{\lambda}\frac{e^{ik(z-z_0)}}{z-z_0}e^{\frac{i\pi}{\lambda(z-z_0)}(x^2+y^2)}\int\int U(x_0,y_0,z_0)e^{-2\pi i\frac{xx_0+yy_0}{\lambda(z-z_0)}}dx_0dy_0 \tag{3.9}$$

This equation simply resumes to a Fourier transform of the input amplitude distribution $U(x_0,y_0,z_0)$. For this reason, the Fraunhofer approximation is widely used for the diffraction analysis of most of classical optical applications.

3.1.3 Grating equation

Although diffraction was originally considered as a limitation to the performance of optical systems, the advantageous use of the phenomenon has been known for two centuries in the form of diffraction gratings, elements that periodically modulate the incident wavefront. Their particular significance lies in the characteristics of the diffracted field: an ideal grating generates a set of waves called diffraction orders that propagate into discrete directions. The grating equation that gives the diffraction angles θ_m for the different diffraction orders m, naturally derives from the scalar diffraction theory applied to a grating of N slits but can also be easily retrieved by simple geometrical considerations. It can be written as

$$n_{I,III}\sin\theta_m \pm n_I\sin\theta_0 = \frac{m\lambda}{\Lambda} \tag{3.10}$$

where θ_0 is the incidence angle of the plane wave of wavelength λ impinging upon the grating of spatial period Λ. n_I and n_{III} are the refractive indices of the incident (superstrate) and transmitting (substrate) media, respectively (see Fig. 3.1). By convention, the plus sign with the index n_I corresponds to the reflection and the minus sign with the index n_{III} to the transmission.

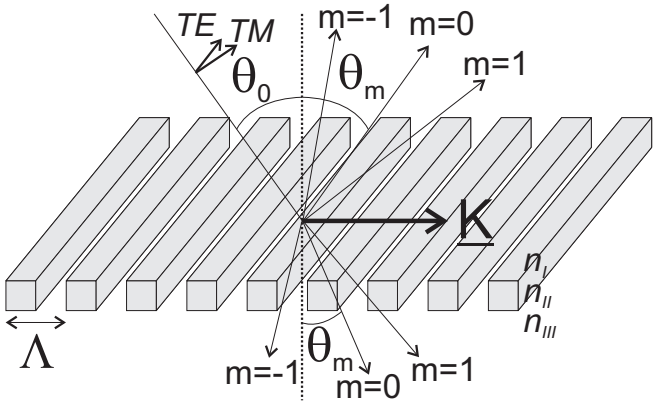

Figure 3.1: Illustration of the grating equation for the first order of diffraction in reflection and in transmission. The fundamental polarization modes TE and TM are shown together with the grating vector $\underline{K}$ which is perpendicular to the grating lines and whose modulus $|\underline{K}| = 2\pi/\Lambda$.

The amplitudes of the diffraction orders are then determined by the physical structure of the periodic modulation. In 90% of the practical cases, a simple Fourier transform is sufficient to give reliable results provided that the Fraunhofer diffraction hypothesis are met. The use of diffraction gratings in spectroscopic applications is motivated by their wavelength selectivity and by their capability to decompose the incident light into a spectrum.

3.2 The vectorial nature of light

Since the development of electromagnetic theory, it has been known that any form of light can be represented by two orthogonal linearly polarized waves, reflecting the boson nature of electromagnetic interactions. These two mutually independent fundamental polarization modes are commonly named TE, or "transverse electric'" polarization where the $\underline{E}$ vector is perpendicular to the plane of incidence[12] and TM, or "transverse magnetic" polarization whereas the same condition holds for the $\underline{H}$ vector (see Fig. 3.1). Once the transverse field component is known, the other field vector that lies in the plane of incidence is obtained from Maxwell's equations.

3.2.1 Wood anomalies

Wood (1902) discovered abrupt variations of the diffraction efficiency with wavelength for a grating illuminated by a continuous-spectrum light source, which could not be explained using scalar diffraction theories. This effect, the so-called Wood (or Rayleigh-Wood) anomaly (Rayleigh 1907), originates from the vanishing or appearance of a higher diffraction order propagating at the grazing angle, i.e., along the surface of the grating. Another type of anomaly are the so-called resonance anomalies, such as the plasmon anomalies of metallic gratings and the anomalies of dielectric coated gratings like guided-mode resonances, which are due to the excitation of leaky waves or surface waves (Nevière 1980; Lenaerts 2005). Common to these anomalous effects is that they are highly polarization dependent, and the full vectorial electromagnetic diffraction theory has to be applied to predict their qualitative and quantitative form (Andrewartha et al. 1979).

To explain the newly-observed resonance phenomena, Lord Rayleigh made the first attempt to solve the electromagnetic problem of gratings in 1907 (Rayleigh 1907). He suggested that the field may be expressed as simple expansion series both inside and outside the modulated region of the grating. The Rayleigh expansion is still used to describe the propagation of the electromagnetic field outside the grating, but the latter assumption was found numerically unsatisfactory in certain cases after over half a century.

In addition to the question of anomalies, the demand for electromagnetic analysis arose with the advance of the fabrication technologies to smaller and smaller groove spacings with respect to the wavelength. Scalar theories fail for such gratings that are operating in and below what is called the resonance domain. Vectorial nature of light must definitely be taken into account.

3.2.2 Vectorial theories of diffraction

In the vectorial theories of diffraction, the vector characteristics of the electromagnetic field are considered via the boundary conditions resulting from Maxwell's equations. They describe the coupling between the individual components of the electromagnetic wavefield. These theories are,

[12] The plane of incidence is defined by the wave vector and the surface normal. In case of normal incidence, it is defined by the wave vector and the grating vector $\underline{K}$ (see Fig. 3.1).

for this reason, often quite abusively qualified as "rigorous". They are rigorous only in the sense that they arrive at the solution without simplifying assumptions and without iterative techniques. But, in general, the accuracy of the solution still depends on the fidelity of the physical problem description. A large number of rigorous vectorial theories have been presented since the 1960's, the majority of which can be classified as integral or differential (Petit 1980; Maystre 1984).

Integral methods differ significantly from the differential methods that are based on the solution of differential equations. In the method developed by Maystre (1984), the field is represented as an integral of an unknown function $\phi(x)$ over a given boundary that separates two half spaces. In the case of a perfectly conducting grating, $\phi(x)$ corresponds to the surface current density, but it has no direct physical significance for dielectric or partially-conducting gratings. The field outside the surface-relief grating is obtained by solving $\phi(x)$ from the integral equations derived from the Helmholtz equation and the boundary conditions. The method can be generalized to gratings containing several layers of different materials. Integral methods are in general rather complicated theoretically and numerically, but they are very stable and perform well even for deep metallic gratings. One advantage of integral methods is that continuous boundaries can be handled accurately, whereas in the differential methods the boundaries must be discretized into slabs, thereby being approximated by staircase-like structures. On the other hand, integral methods are not suitable for graded-index gratings such as sinusoidal holographic volume gratings. Their applicability to discontinuous profiles such as binary gratings is questionable as well; the profile may be approximated by its truncated Fourier-series representation, but the convergence is not always satisfactory due to Gibbs phenomenon.

3.2.3 Rigorous Coupled-Wave Analysis

In this section, we present a stable and efficient three-dimensional RCWA algorithm for the diffraction of multi-layer stacks with two-dimensional gratings and homogeneous layers. The RCWA is part of the family of the differential methods, it is also called the Fourier modal method (FMM), because it is based on a decomposition of the field in a Fourier basis. The accuracy of the solution depends solely on the number of terms in the field space-harmonic expansion, with conservation of energy always being satisfied. The basic formulation of the algorithm (Moharam & Gaylord 1981) is to express the electromagnetic fields in the regions bounding the grating structure (the incident and transmitting regions) as solutions to Maxwell's equations and to express the fields within the grating region as space-harmonic expansions which are also solutions to Maxwell's equations. The tangential field components are then matched at each layer boundary within the structure to ensure continuity.

The RCWA algorithm was first exposed by Moharam & Gaylord (1981) for the investigation of holographic gratings. They then extended its application to surface relief gratings (Moharam & Gaylord 1982, 1983a) in conical incidence (Moharam & Gaylord 1983b; Moharam et al. 1995a), for metallic gratings (Moharam & Gaylord 1986), and later on to two-dimensional gratings (Moharam 1988). However, despite its versatility and simplicity, the coupled-wave method was used with caution because of very slow convergence in TM polarization (see Peng & Morris (1995) for instance). This numerical issue, also affecting the differential theory, was identified by Li & Haggans (1993) and Turunen (1996), and later on empirically solved for the RCWA by Lalanne & Morris (1996); Lalanne (1997a) and Granet & Guizal (1996). Li (1996b) gave both reformulations a firm mathematical foundation while Popov & Nevière (2000) applied it to the differential theory. The new formulation converges faster because it uniformly satisfies the boundary conditions in the grating region, whereas the old formulations do so only nonuniformly. In other words, it uniformly

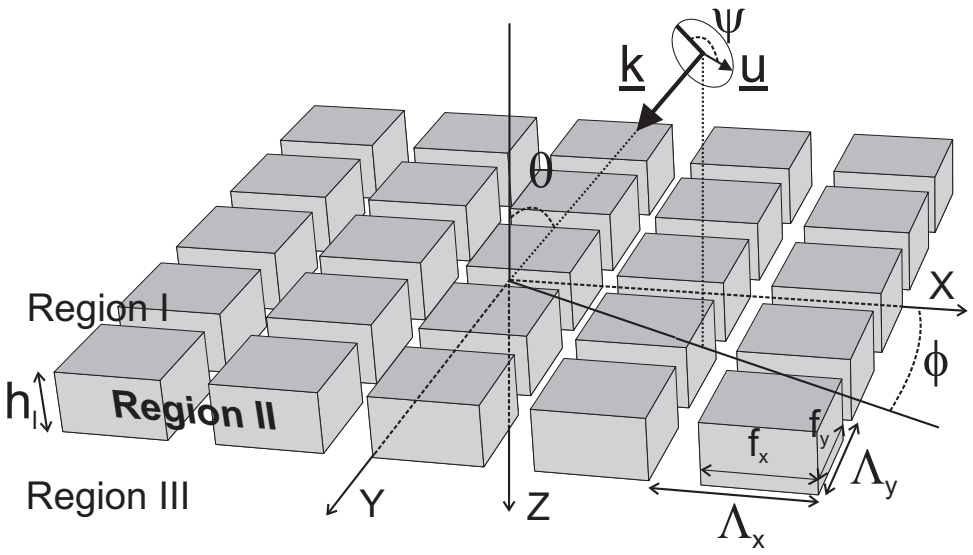

Figure 3.2: Three-dimensional geometry of the diffraction problem by a two-dimensional grating.

preserves the continuity of the electromagnetic field quantities that should be continuous across permittivity discontinuities.

Arbitrary-profile gratings must nevertheless still be discretized into layers of homogeneous permittivity modulation where the electromagnetic field is determined and then matched in sequence at the interfaces. Even if the staircase approximation has recently been questioned (Popov et al. 2002), it is valid for the majority of conventional problems. However, as for multilevel gratings, the analysis methods of discretized gratings face a common difficulty associated with the exponential functions of the spatial variable in the direction of wave propagation, i.e., perpendicular to the grating plane. In order to solve this issue, Moharam et al. (1995b) proposed the "enhanced transmittance matrix method" and Li (1996a, 1997), the "S-Matrix" and "R-Matrix" algorithms. The most important criterion for achieving unconditional numerical stability with these algorithms is to avoid the exponentially-growing functions in every step of the matrix recursion. From the point of view of numerical efficiency, the S-matrix algorithm is generally preferred even if exceptional cases are noted.

We will now expose the three-dimensional RCWA algorithm for two-dimensional gratings in its latest version implementing the latest numerical patches presented here above. Let us consider the geometry of Fig. 3.2 illustrating a grating structure consisting of a two-dimensional grating layer. For presentation purpose, we have only sketched one layer in the region II, but in general, the structure can be a stack of any number of grating layers embedded with homogeneous ones in any arbitrary order. To apply RCWA to the stack, all grating layers must have the same periodicity Λ_x along the X direction, and the same periodicity Λ_y along the Y direction. f_x and f_y are the so-called "filling factor" along the X and Y directions, respectively. They are defined as the ratio of the width of the grating ridges over the grating period. The thickness of the l_{th} layer is h_l, with l the layer index. The number of layers in the stack is L and the total thickness of the stack is $H_L = \sum_{l=1}^{L} h_l$ (see Fig. 3.3).

The whole stack can be divided into three regions: the incident region (Region I), the stack region (Region II) and the exit region (Region III). Region I and III correspond to the superstrate and substrate and are isotropic dielectric media characterized by optical refractive indices n_I and n_{III}, respectively. Instead of refractive indices, we shall use the permittivities $\epsilon_I = n_I^2$ and $\epsilon_{III} = n_{III}^2$. The permittivity distribution in each grating layer of the intermediate region II is

periodically modulated in the X and Y direction (periods Λ_x and Λ_y). Since the RCWA (or the FMM) requires the expansion of the field and permittivity in Fourier series, let us expand the permittivity and its inverse in Fourier series of the spatial harmonics as follows

$$\epsilon_l(x,y) = \sum_{g,h} \epsilon_{l,gh} e^{\left(j\frac{2\pi g x}{\Lambda_x} + j\frac{2\pi h y}{\Lambda_y}\right)} \tag{3.11}$$

$$\epsilon_l^{-1}(x,y) = \sum_{g,h} A_{l,gh} e^{\left(j\frac{2\pi g x}{\Lambda_x} + j\frac{2\pi h y}{\Lambda_y}\right)} \tag{3.12}$$

in which $\epsilon_{l,gh}$ and $A_{l,gh}$ are the Fourier coefficients for the l_{th} layer. They are therefore easily calculated by a two-dimensional Fourier transform.

A unit-amplitude monochromatic plane wave with vacuum wavelength λ_0 is assumed to impinge from region I with arbitrary linear polarization. Suppressing the time dependence $e^{j\omega t}$, its electric field is given by

$$\underline{E}_{inc}(x,y,z) = \underline{u} e^{-j\underline{k}\cdot\underline{r}} \tag{3.13}$$

The wave vector in the incident region $\underline{k}$ is given by

$$\underline{k} = k_0 n_I \left(\sin\theta\cos\phi\,\underline{x} + \sin\theta\sin\phi\,\underline{y} + \cos\theta\,\underline{z}\right) \tag{3.14}$$

in which $k_0 = \frac{2\pi}{\lambda}$. The unit polarization vector $\underline{u}$ of the incident wave is given in terms of the angles θ, ϕ, ψ

$$\begin{aligned} \underline{u} &= u_x\underline{x} + u_y\underline{y} + u_z\underline{z} && (3.15)\\ &= (\cos\psi\cos\theta\cos\phi - \sin\psi\sin\phi)\,\underline{x} && (3.16)\\ &\quad + (\cos\psi\cos\theta\sin\phi + \sin\psi\cos\phi)\,\underline{y} && (3.17)\\ &\quad - (\cos\psi\sin\theta)\,\underline{z} && (3.18) \end{aligned}$$

in which θ is the polar angle, and ϕ the azimuth angle. The angle between the electric-field vector and the incident plane is ψ.

Field expressions in Regions I and III

In the regions I and III with constant refractive index, wave propagation is described by the Helmholtz equation 3.1. The simplest general solutions are plane waves, of which only a discrete set is allowed for a grating. The diffracted fields in regions I and III may be expressed in the form of the so-called Rayleigh expansions (Rayleigh 1907; Maystre 1984)

$$\underline{E}_I = \underline{E}_{inc} + \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \underline{R}_{mn} e^{-j\left(\underline{k}_{xm}x + \underline{k}_{yn}y - \underline{k}_{Iz,mn}z\right)} \tag{3.19}$$

$$\underline{E}_{III} = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \underline{T}_{mn} e^{-j\left(\underline{k}_{xm}x + \underline{k}_{yn}y - \underline{k}_{IIIz,mn}(z-Z_L)\right)} \tag{3.20}$$

in which $\underline{R}_{mn}$ and $\underline{T}_{mn}$ are the complex amplitudes of the electric fields of mn-th reflected and transmitted orders, respectively. The wave vector components k_{xm} and k_{yn} arise from phase matching and the Floquet conditions[13] and are given by

$$k_{xm} = k_0\left(n_I\sin\theta\cos\phi - m\frac{\lambda_0}{\Lambda_x}\right) \tag{3.21}$$

[13]For the mathematical demonstration of the Floquet theorem and the subsequent Rayleigh expansion, see for instance Moreau (2002).

$$k_{yn} = k_0 \left(n_I \sin\theta \cos\phi - n\frac{\lambda_0}{\Lambda_y} \right) \tag{3.22}$$

and

$$k_{lz,mn} = \begin{cases} k_0 \left[n_l^2 - \left(\frac{k_{xm}}{k_0}\right)^2 - \left(\frac{k_{yn}}{k_0}\right)^2 \right]^{1/2} & n_l > \left[\left(\frac{k_{xm}}{k_0}\right)^2 + \left(\frac{k_{yn}}{k_0}\right)^2 \right]^{1/2} \\ -jk_0 \left[\left(\frac{k_{xm}}{k_0}\right)^2 + \left(\frac{k_{yn}}{k_0}\right)^2 - n_l^2 \right]^{1/2} & n_l < \left[\left(\frac{k_{xm}}{k_0}\right)^2 + \left(\frac{k_{yn}}{k_0}\right)^2 \right]^{1/2} \end{cases} \tag{3.23}$$

with l =I, III. Diffraction orders with real-valued $k_{lz,mn}$ correspond to plane waves propagating in the regions I and III, respectively, or homogeneous waves.

Field expressions in Region II

In the grating region, the fields may be expressed as a Fourier expansions in terms of the spatial harmonics

$$\underline{E}_l = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \underline{S}_{l,mn}(z) e^{-j(k_{xm}x + k_{yn}y)} \tag{3.24}$$

$$\underline{H}_l = -j\sqrt{\frac{\epsilon_0}{\mu_0}} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \underline{U}_{l,mn}(z) e^{-j(k_{xm}x + k_{yn}y)} \tag{3.25}$$

in which ϵ_0 and μ_0 are the permittivity and permeability of free space, respectively. $\underline{S}_{l,mn}$ and $\underline{U}_{l,mn}$ are the amplitudes of the spatial harmonics of the fields such that $\underline{E}_l$ and $\underline{H}_l$ satisfy Maxwell's equations in the l_{th} grating layer

$$\underline{\nabla} \times \underline{E}_l = -j\omega\mu \underline{H}_l \tag{3.26}$$

$$\underline{\nabla} \times \underline{H}_l = -j\omega\epsilon_l(x,y)\epsilon_0 \underline{E}_l \tag{3.27}$$

Let us emphasize that the name "rigorous coupled-wave analysis" stems from the interpretation of $\underline{S}_{l,mn}$ and $\underline{U}_{l,mn}$ as amplitude functions of waves that are mutually coupled so that only the complete set rather than any single wave fulfills Maxwell's equations. Substituting Eqs. 3.24 and 3.25 into Eqs. 3.26 and 3.27 and eliminating the z components of the fields, an infinite set of first-order differential equations can be derived

$$\begin{aligned}
\frac{\partial S_{l,ymn}(z)}{\partial z'} &= U_{l,xmn}(z) + \frac{k_{yn}}{k_0^2} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} A_{l,m-p,n-q}(-k_{yq}U_{l,xpq} + k_{xp}U_{l,ypq}) \\
\frac{\partial S_{l,xmn}(z)}{\partial z'} &= -U_{l,ymn}(z) + \frac{k_{xm}}{k_0^2} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} A_{l,m-p,n-q}(-k_{yq}U_{l,xpq} + k_{xp}U_{l,ypq}) \\
\frac{\partial U_{l,ymn}(z)}{\partial z'} &= \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \epsilon_{l,m-p,n-q} S_{l,xpq} + \frac{k_{yn}}{k_0^2}(k_{xm}S_{l,ymn} + k_{yp}S_{l,xmn}) \\
\frac{\partial U_{l,xmn}(z)}{\partial z'} &= \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \epsilon_{l,m-p,n-q} S_{l,ypq} + \frac{k_{xm}}{k_0^2}(k_{xm}S_{l,ymn} + k_{yp}S_{l,xmn})
\end{aligned} \tag{3.28}$$

in which $z' = k_0 z$, $p = m-g$, and $q = n-h$. The indices m and n run over the different diffraction orders, while g and h run over the Fourier harmonics of the permittivity and its inverse. In order

to numerically solve this set of coupled differential equations, the set is truncated to finite size and expressed in matrix form

$$\frac{\partial}{\partial z'}\begin{bmatrix} \mathbf{S_{l,y}} \\ \mathbf{S_{l,x}} \\ \mathbf{U_{l,y}} \\ \mathbf{U_{l,x}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{P_{l,11}} & \mathbf{P_{l,12}} \\ \mathbf{0} & \mathbf{0} & \mathbf{P_{l,21}} & \mathbf{P_{l,22}} \\ \mathbf{Q_{l,11}} & \mathbf{Q_{l,12}} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q_{l,21}} & \mathbf{Q_{l,22}} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{S_{l,y}} \\ \mathbf{S_{l,x}} \\ \mathbf{U_{l,y}} \\ \mathbf{U_{l,x}} \end{bmatrix} \tag{3.29}$$

with

$$\mathbf{P_l} = \begin{bmatrix} \mathbf{P_{l,11}} & \mathbf{P_{l,12}} \\ \mathbf{P_{l,21}} & \mathbf{P_{l,22}} \end{bmatrix} = \begin{bmatrix} \mathbf{k_y}\epsilon_\mathbf{l}^{-\mathbf{1}}\mathbf{k_x} & \mathbf{I} - \mathbf{k_y}\epsilon_\mathbf{l}^{-\mathbf{1}}\mathbf{k_y} \\ \mathbf{k_x}\epsilon_\mathbf{l}^{-\mathbf{1}}\mathbf{k_x} - \mathbf{I} & -\mathbf{k_x}\epsilon_\mathbf{l}^{-\mathbf{1}}\mathbf{k_y} \end{bmatrix} \tag{3.30}$$

and

$$\mathbf{Q_l} = \begin{bmatrix} \mathbf{Q_{l,11}} & \mathbf{Q_{l,12}} \\ \mathbf{Q_{l,21}} & \mathbf{Q_{l,22}} \end{bmatrix} = \begin{bmatrix} \mathbf{k_x k_y} & \alpha\mathbf{A_l^{-1}} + (\mathbf{1}-\alpha)\epsilon_\mathbf{l} - \mathbf{k_y^2} \\ \mathbf{k_x^2} - \alpha\epsilon_\mathbf{l} - (\mathbf{1}-\alpha)\mathbf{A_l^{-1}} & -\mathbf{k_x k_y} \end{bmatrix} \tag{3.31}$$

To improve the computational efficiency, Eq. 3.29 can be further reduced to two second-order equations, shown in matrix form as

$$\begin{bmatrix} \frac{\partial^2 \mathbf{S_{l,y}}}{\partial \mathbf{z'^2}} \\ \frac{\partial^2 \mathbf{S_{l,x}}}{\partial \mathbf{z'^2}} \end{bmatrix} = \mathbf{\Omega} \begin{bmatrix} \mathbf{S_{l,y}} \\ \mathbf{S_{l,x}} \end{bmatrix} \tag{3.32}$$

with

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{k_x^2} + \mathbf{D}\left[\alpha\epsilon_\mathbf{l} + (\mathbf{1}-\alpha)\mathbf{A_l^{-1}}\right] & \mathbf{k_y}\left(\epsilon_\mathbf{l}^{-\mathbf{1}}\mathbf{k_x}\left[\alpha\mathbf{A_l^{-1}} + (\mathbf{1}-\alpha)\epsilon_\mathbf{l}\right] - \mathbf{k_x}\right) \\ \mathbf{k_x}\left(\epsilon_\mathbf{l}^{-\mathbf{1}}\mathbf{k_y}\left[\alpha\epsilon_\mathbf{l} + (\mathbf{1}-\alpha)\mathbf{A_l^{-1}}\right] - \mathbf{k_y}\right) & \mathbf{k_y^2} + \mathbf{B}\left[\alpha\mathbf{A_l^{-1}} + (\mathbf{1}-\alpha)\epsilon_\mathbf{l}\right] \end{bmatrix} \tag{3.33}$$

in which $\mathbf{B} = \mathbf{k_x}\epsilon_\mathbf{l}^{-\mathbf{1}}\mathbf{k_x} - \mathbf{I}$ and $\mathbf{D} = \mathbf{k_y}\epsilon_\mathbf{l}^{-\mathbf{1}}\mathbf{k_y} - \mathbf{I}$. $\epsilon_\mathbf{l}$ and $\mathbf{A_l}$ are the permittivity and inverse permittivity matrices which consist of the harmonic coefficients $\epsilon_{l,gh}$ and $A_{l,gh}$ defined in Eqs. 3.11 and 3.12. The diagonal matrices $\mathbf{k_x}$ and $\mathbf{k_y}$ are formed by the elements k_{xm}/k_0 and k_{yn}/k_0. If M and N are the numbers of spatial harmonics retained along the X and Y directions, then $\mathbf{I}$ is the identity matrix of dimension MN.

The improved eigenvalue formulation proposed by Lalanne (1997b) is here adopted. The real positive α parameter depends on the grating geometry and ranges in the interval $[0, 1]$. For example, in the case of symmetric gratings, i.e., with $f_x = f_y$, α has to be optimally at 0.5. In the asymptotic case of $f_y = 1$, i.e., when the two-dimensional grating becomes a one-dimensional grating with a periodicity along the X direction, α has to be unitary, leading to the case depicted in Lalanne & Morris (1996).

The coupled-wave equations 3.32 are then solved by finding the eigenvalues and eigenvectors of $\mathbf{\Omega}$, which is a matrix of rank 2MN. Compared to the eigenvalue problem of Eq. 3.28, which is of rank 4MN, the computational efficiency is improved by a factor 8. The spatial harmonics of the tangential electric field in the grating layers may be written as

$$S_{l,ymn}(z) = \sum_{i=1}^{2MN} w_{l1,mn}\left[d_{l,i}e^{-k_0\sigma_{l,i}(z-H_{l-1})} + u_{l,i}e^{k_0\sigma_{l,i}(z-H_l)}\right] \tag{3.34}$$

$$S_{l,xmn}(z) = \sum_{i=1}^{2MN} w_{l2,mn}\left[d_{l,i}e^{-k_0\sigma_{l,i}(z-H_{l-1})} + u_{l,i}e^{k_0\sigma_{l,i}(z-H_l)}\right] \tag{3.35}$$

in which the w's are elements of $\mathbf{W_l}$, the eigenvector matrix, and the σ's are elements of $\mathbf{\Sigma}$, which is the diagonal matrix of the positive square roots of the eigenvalues of $\mathbf{\Omega_l}$. $\mathbf{u^{(l)}}$ and $\mathbf{d^{(l)}}$ are column

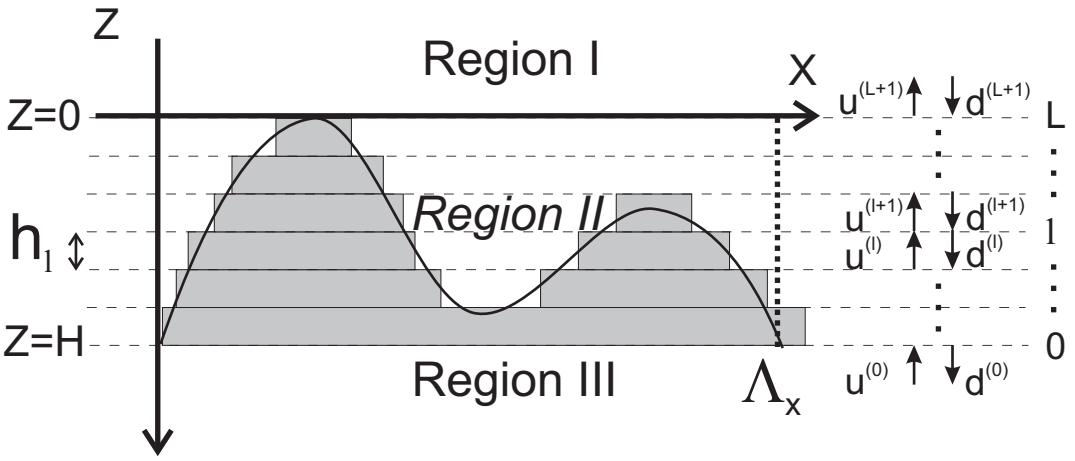

Figure 3.3: Staircase approximation of arbitrary-profile gratings.

vectors whose elements are the complex amplitudes of the modes propagating upward (in the -Z direction) and downward (in the Z direction), respectively. Their 4MN unknown coefficients will be determined from matching the boundary conditions at the appropriate interfaces. H_l represents the cumulative depth of the structure to the l_{th} layer and is given by $H_l = \sum_{l'=1}^{l} h_{l'}$, in which $h_{l'}$ is the thickness of the l_{th} layer. Substituting Eqs. 3.34 and 3.35 into Eq. 3.28 to deduce the magnetic fields in the grating layers, we can write in matrix form

$$\begin{bmatrix} \mathbf{S_{l,y}} \\ \mathbf{S_{l,x}} \\ \mathbf{U_{l,y}} \\ \mathbf{U_{l,x}} \end{bmatrix} = \begin{bmatrix} \mathbf{W_{l,1}X_{l,1}} & \mathbf{W_{l,1}X_{l,2}} \\ \mathbf{W_{l,2}X_{l,1}} & \mathbf{W_{l,2}X_{l,2}} \\ -\mathbf{V_{l,1}X_{l,1}} & \mathbf{V_{l,1}X_{l,2}} \\ -\mathbf{V_{l,2}X_{l,1}} & \mathbf{V_{l,2}X_{l,2}} \end{bmatrix} \begin{bmatrix} \mathbf{d^{(l)}} \\ \mathbf{u^{(l)}} \end{bmatrix} \tag{3.36}$$

where $\mathbf{X_{l,1}}$ and $\mathbf{X_{l,2}}$ are 2MNx2MN diagonal matrices with diagonal elements equal to $e^{-k_0\sigma_{l,i}(z-H_{l-1})}$ and $e^{k_0\sigma_{l,i}(z-H_l)}$, respectively, and in which

$$\mathbf{V_{l,1}} = (\mathbf{Q_{l,11}W_{l,1}} + \mathbf{Q_{l,12}W_{l,2}})\mathbf{\Sigma_l^{-1}} \tag{3.37}$$

$$\mathbf{V_{l,2}} = (\mathbf{Q_{l,21}W_{l,1}} + \mathbf{Q_{l,22}W_{l,2}})\mathbf{\Sigma_l^{-1}} \tag{3.38}$$

Once the eigenmodes and their propagating constants are known in every layer, we use the S-matrix algorithm to compute the up-wave $\mathbf{u}^{(l)}$ and down-wave $\mathbf{d}^{(l)}$ amplitude vectors (see Fig. 3.3). We define the S-matrix $\mathbf{S^{(l)}}$ that links the waves in the section $l+1$ and those of medium 0 in this way

$$\begin{bmatrix} \mathbf{u^{(l+1)}} \\ \mathbf{d^{(0)}} \end{bmatrix} = \mathbf{S^{(l)}} \begin{bmatrix} \mathbf{u^{(0)}} \\ \mathbf{d^{(l+1)}} \end{bmatrix} \tag{3.39}$$

with

$$\mathbf{S^{(l)}} = \begin{bmatrix} \mathbf{T_{uu}^{(l)}} & \mathbf{R_{ud}^{(l)}} \\ \mathbf{R_{du}^{(l)}} & \mathbf{T_{dd}^{(l)}} \end{bmatrix} \tag{3.40}$$

The transmission and reflection submatrices in Eq. 3.40 are computed in a recursive way. If, as usual, no light is impinging on the grating from the output region, i.e., $\mathbf{u^{(0)}} = \mathbf{0}$, only $\mathbf{R_{ud}^{(l)}}$ and $\mathbf{T_{dd}^{(l)}}$ are required for computing the reflected and transmitted waves, $\mathbf{u^{(L+1)}}$ and $\mathbf{d^{(0)}}$. The recursion formulae are

$$\mathbf{R_{ud}^{(l)}} = \mathbf{r_{ud}^{(l)}} + \mathbf{t_{ud}^{(l)}R_{ud}^{(l-1)}}\left[1 - \mathbf{r_{du}^{(l)}R_{ud}^{(l-1)}}\right]^{-1}\mathbf{t_{dd}^{(l)}} \tag{3.41}$$

$$\mathbf{T_{dd}^{(l)}} = \mathbf{T_{dd}^{(l-1)}}\left[1 - \mathbf{r_{du}^{(l)}R_{ud}^{(l-1)}}\right]^{-1}\mathbf{t_{dd}^{(l)}} \tag{3.42}$$

The matrices denoted by lowercase letters are the reflection and transmission submatrices that relate the waves in the two adjacent layers l and $l+1$. They are given by

$$\begin{aligned}
\mathbf{t_{uu}^{(l)}} &= 2\left[\mathbf{W^{(l)-1}W^{(l+1)}} + \mathbf{V^{(l)-1}V^{(l+1)}}\right]^{-1}\mathbf{X^{(l)}} \\
\mathbf{r_{ud}^{(l)}} &= \left[\mathbf{W^{(l)-1}W^{(l+1)}} + \mathbf{V^{(l)-1}V^{(l+1)}}\right]^{-1}\left[\mathbf{W^{(l)-1}W^{(l+1)}} - \mathbf{V^{(l)-1}V^{(l+1)}}\right] \\
\mathbf{r_{du}^{(l)}} &= \mathbf{X^{(l)}}\left[\mathbf{V^{(l+1)-1}V^{(l)}} + \mathbf{W^{(l+1)-1}W^{(l)}}\right]^{-1}\left[\mathbf{V^{(l+1)-1}V^{(l)}} - \mathbf{W^{(l+1)-1}W^{(l)}}\right]\mathbf{X^{(l)}} \\
\mathbf{t_{dd}^{(l)}} &= 2\mathbf{X^{(l)}}\left[\mathbf{V^{(l+1)-1}V^{(l)}} + \mathbf{W^{(l+1)-1}W^{(l)}}\right]^{-1}
\end{aligned}$$

in which $\mathbf{W^{(l)}} = [\mathbf{W_{l,1}}; \mathbf{W_{l,2}}]$, $\mathbf{V^{(l)}} = [\mathbf{V_{l,1}}; \mathbf{V_{l,2}}]$ and $\mathbf{X^{(l)}} = [\mathbf{X_{l,1}}]$. The recursion is initialized by setting $\mathbf{S^{(-1)}}$ equal to the identity matrix. Matrices $\mathbf{S^{(l)}}$ are computed recursively until matrix $\mathbf{S^{(L)}}$ is obtained. The amplitudes of the modes propagating backward in the input region I, $\mathbf{u^{(L+1)}}$ and forward in the output region III, $\mathbf{d^{(0)}}$, are obtained from Eq. 3.39, where $\mathbf{u^{(0)}} = \mathbf{0}$ and $\mathbf{d^{(L+1)}}$ represent the components of the incident wave.

The numerical stability of the S-matrix algorithm is rooted in the construction of the propagation matrix S itself. Indeed, the problem-causing, growing-exponential function, $\mathbf{X_{l,2}} = e^{k_0\sigma_{l,i}(z-H_l)}$ that could lead to truncation-error propagation, is absent from the S-matrix algorithm. If we define the diffraction efficiency in reflection (resp. transmission) $DE_{R,mn}$ (resp. $DE_{T,mn}$) by the amount of light that is back-reflected (resp. transmitted) into the diffraction order mn, $DE_{R,mn}$ (resp. $DE_{T,mn}$) is simply equal to the z component of the time-averaged Poynting vector and is related to the electric field components by the following relations:

$$DE_{R,mn} = Re\left(\frac{k_{Iz,mn}}{k_I\cos\theta}\right)|\underline{R}_{mn}|^2 \tag{3.43}$$

$$DE_{T,mn} = Re\left(\frac{k_{IIIz,mn}}{k_{III}\cos\theta}\right)|\underline{T}_{mn}|^2 \tag{3.44}$$

Thanks to numerous improvements, RCWA can be now considered as a well-established numerical technique for the study of grating-diffraction problems. It can efficiently solve almost all grating-stack configurations and has even recently been extended to waveguide and integrated optics computational problems (Lalanne & Silberstein 2000; Silberstein et al. 2001; Cao et al. 2002).

3.3 Subwavelength gratings

3.3.1 Definition

Subwavelength gratings are gratings which spatial period is smaller than the wavelength of the incident light. They do not diffract light in the sense that only the zeroth transmitted and reflected orders are allowed to propagate outside the grating region. All but the zeroth order are evanescent, leaving incident wavefronts free from any further aberrations (Richter et al. 1995). For this reason, subwavelength gratings are often called zero-order gratings (ZOGs). The condition under which a diffraction order propagates or not is determined by the grating equation 3.10 from which a "ZOG condition" on the grating period-to-wavelength ratio can be derived

$$\frac{\Lambda}{\lambda} \leq \frac{1}{n_I\sin\theta + \max(n_I, n_{III})} \tag{3.45}$$

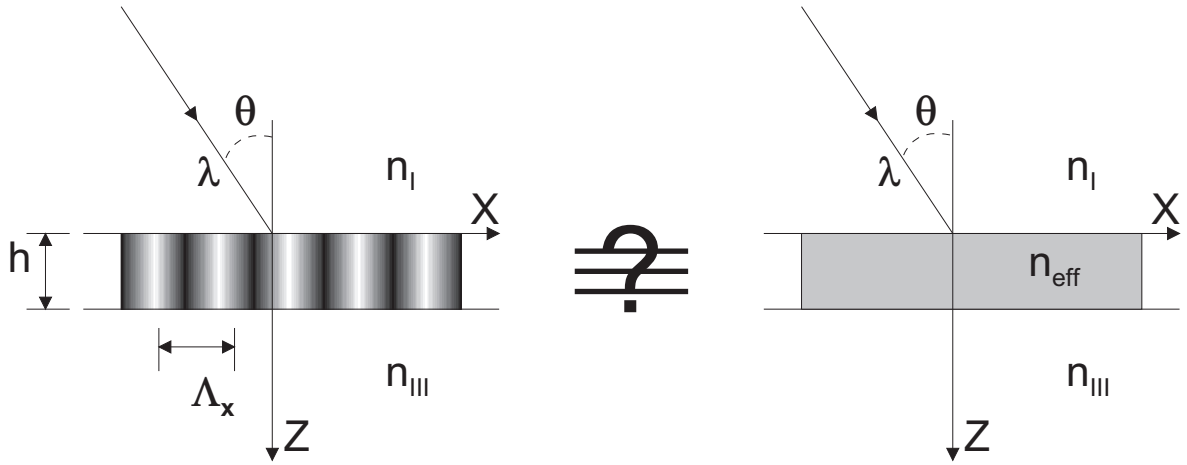

Figure 3.4: Homogenization of subwavelength gratings. Subwavelength gratings are considered as homogeneous media with effective properties that can be determined with more or less precisions according to the complexity of the chosen theoretical approach.

The zeroth order is still influenced by structure of the period and subwavelength gratings present very specific properties like, for example, unique anisotropic and dispersive characteristics which can be wisely used for numerous applications. One can really speak of refractive index engineering since the optical effective properties of the structures can be controlled by their geometry.

3.3.2 Effective medium theories

The foundations of the effective medium approximations to zero-order gratings rely on the fact that upon transmission through a subwavelength grating, the zeroth order experiences effective refractive indices resulting from the averaging of the dielectric constants of the grating media. This phenomenon is also known as the homogenization of subwavelength gratings (Fig. 3.4). To calculate these so-called effective refractive indices n_{eff}, scalar theories of diffraction are ineffective (see, e.g. Glytsis 2002). The vectorial nature of light must of course be taken into account. Before using a tool as complex as RCWA, several simpler approaches are possible depending on the grating period-to-wavelength ratio and the precision required. In fact, the objective of homogenization is to provide a simplified model of composite materials whose rigorous analysis is computationally difficult and sometimes even impossible. It is also often a practical way to qualitatively describe the complex behaviors of subwavelength gratings.

For a binary 1D surface-relief grating in the quasi-static limit, i.e., when the ratio $\frac{\Lambda}{\lambda} << 1$, a simple averaging treatment leads to the two following effective indices for the TE and TM states of polarization (Born & Wolf 1999b)

$$n_{eff,0}^{TE} = \left(F n_a^2 + (1-F) n_b^2\right)^{1/2} \tag{3.46}$$

$$n_{eff,0}^{TM} = \left(\frac{n_a^2 n_b^2}{F n_b^2 + (1-F) n_a^2}\right)^{1/2} \tag{3.47}$$

where n_a and n_b are the structure's real indices, and where F is the 1D filling factor. This straightforward approach that only considers the propagation of the zeroth order inside the modulated region neglecting higher-order modes is called the zeroth-order effective medium theory (EMT0).

But when the ratio $\frac{\Lambda}{\lambda}$ is no longer negligible, the latter closed-form expressions for the effective refractive indices are no longer correct. In such a case, higher-order effective medium theories

like the second-order effective medium theory (EMT2) which is deduced from the electromagnetic propagation in stratified media theory, allows to derive the following expressions for the effective indices (Rytov 1956; Yariv & Yhe 1984; Brundrett et al. 1994)

$$n_{eff,2}^{TE} = \left[(n_{eff,0}^{TE})^2 + \frac{1}{3} \left(\frac{\Lambda}{\lambda} \right)^2 \pi^2 F^2 (1-F)^2 (n_a^2 - n_b^2)^2 \right]^{1/2} \tag{3.48}$$

$$n_{eff,2}^{TM} = \left[(n_{eff,0}^{TM})^2 + \frac{1}{3} \left(\frac{\Lambda}{\lambda} \right)^2 \pi^2 F^2 (1-F)^2 \left(\frac{1}{n_a^2} - \frac{1}{n_b^2} \right)^2 (n_{eff,0}^{TM})^6 (n_{eff,0}^{TE})^2 \right]^{1/2} \tag{3.49}$$

In addition to the dependence on the wavelength, we also notice the dependence of the effective indices versus other parameters available in a possible design procedure: the grating period Λ, the filling factor F and the grating real indices n_a and n_b. It is important to note that in some cases, the grating effective properties are not intrinsic to the periodic structure but are rather related to the whole diffraction problem (Lalanne & Lemercier-Lalanne 1997).

Homogenization theories have extensively been studied in the 1990's. They have been applied to a large variety of grating problems: for lamellar gratings (Gu & Yeh 1996), in conical mounting (Haggans et al. 1993), for two-dimensional subwavelength binary gratings (Grann et al. 1994; Kikuta et al. 1998), for volume gratings (Campbell & Kostuk 1995; Gu & Yeh 1995; Lalanne & Hugonin 1998; Joubert et al. 2002). But nowadays, and thanks to increased computing efficiencies, these approximate and sometimes unfortunately misleading approaches are abandoned in favor of more complex and accurate theories, like RCWA.

3.4 Manufacturing techniques

Subwavelength gratings and micro-optics fabrication in general make use of lithographic technologies. Lithography is the name for a sequence of processing steps for structuring the surfaces of planar substrates (Herzig 1997; Sinzinger & Jahns 2003). Two types of lithographic fabrication procedures can be distinguished: mask lithography and scanning lithography. In mask lithography the pattern of the component is encoded as an amplitude distribution in a lithographic mask. Uniform illumination of the mask is used to expose a photosensitive coating on the substrate. In scanning lithography, no mask is used. Rather, local variation of the photoresist exposure is achieved in a so-called direct-writing process. To this end, a laser or electron beam is scanned over the substrate, while the beam intensity and exposure time of the beam are modulated. After the exposure of the photoresist layer, a development step converts the exposed photoresist into a surface profile. In a further processing step, the surface profile of the photoresist pattern can be transferred into the substrate.

3.4.1 Lithography of resists

Mask lithography

Widely used in micro-electronics or for high-volume fabrication of micro-chips, this technique forms the contrasted image through a mask previously manufactured according to the pattern to be projected onto a photosensitive material (Fig. 3.5). Different variants exist such as the contact and non-contact copy. This technique is especially useful for large-batch process and when alignment between different masks has to be performed in a multi-layer process (Sinzinger & Jahns 2003).

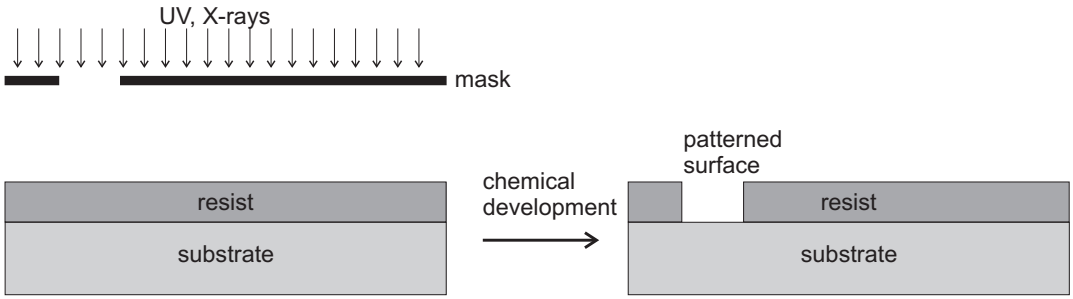

Figure 3.5: Principle of mask lithography. According to the position of the amplitude mask relative to the substrate and the illumination configuration, we can distinguish three types of mask lithography techniques: 1- contact printing where the mask is brought in contact with the resist, 2- proximity printing where a gap is maintained between the mask and the substrate, and 3- projection printing where the mask pattern is imaged onto the wafer optically. In this case very sophisticated optical (reduction) systems are necessary.

Scanning lithography

Direct Writing Laser (DWL). This is the most flexible technique based on a UV or blue laser beam scanned over the surface to be patterned (Fig. 3.6). Any kind of binary or multilevel pattern can be recorded thanks to the combination of a laser focussing system with an accurate moving table underneath. DWL process therefore uses the laser beam as a controlled writing tool for generating patterns onto a photosensitive layer (photoresist). A chemical processing of the exposed material is necessary in order to transform the previously contrasted illuminating pattern into a relief structure after the exposure step.

Electron/focused ion-beam lithography. In electron-beam (e-beam) lithography, a pattern on top of an electroresist-coated substrate is derived from imagery techniques based on an electron-beam scanning (e.g. Kajanto et al. 1989). Electrons are used for changing the structure of electroresist (Fig. 3.6). Illuminated areas become more soluble than other ones, making the subsequent chemical process more efficient onto exposed parts. Nanometric resolutions are achievable onto rather small substrates. Moreover, the substrate has to be prepared with an underneath conductive layer in order to avoid any charging effect. In focused ion-beam (FIB) writing, a focused ion beam instead of electron beam is used for patterning the surface at nanometric scale.

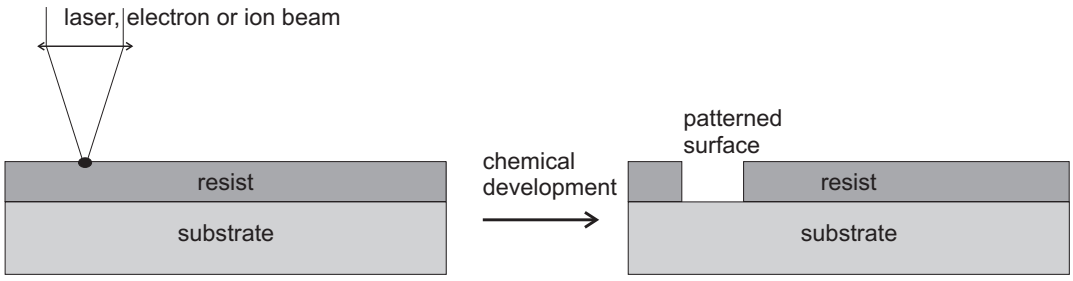

Figure 3.6: Principle of Scanning lithography.

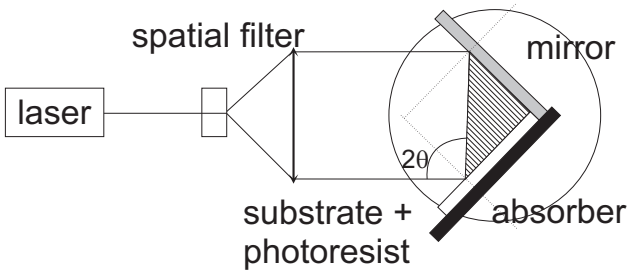

Figure 3.7: Principle of holographic lithography.

Holographic lithography

Holographic recording is a maskless photolithographic technique particularly suitable for recording diffractive structures into photoresists or gelatins. Periodic patterns ranging from a few hundreds up to several thousands lines per millimeter can be recorded in a single process. The holographic recording of diffraction gratings is a very efficient manufacturing technique, especially for large-scale gratings where direct ruling (one line at a time) becomes unpractical. It consists in recording in a photosensitive material the contrast pattern of two mutually-coherent interfering laser beams (Fig. 3.7). The recording material is usually dichromated gelatin (DCG) or a photoresist. After a wet chemical process, the resulting grating is a volume-phase holograms (VPH) or a surface-relief grating, respectively.

Nano-imprint lithography

In recent years, researchers have investigated a number of alternative and potentially low-cost methods: most notably, micro-contact printing, nano-imprint technology, AFM lithography and dip-pen lithography. Since the mid-1990s, nano-imprint lithography (NIL), initially proposed and developed by the Chou group (Chou et al. 1996), has emerged as one of the most promising technologies for high-throughput nanoscale patterning. In this method and its variants, such as step-and-flash imprint lithography (S-FIL), pattern replication is done nontraditionally by deforming mechanically the resist materials, which makes them completely free from the resolution-limiting factors such as light diffraction and beam scattering that are often inherent with the more traditional approaches (Guo 2004).

The principle of nano-imprint lithography is quite simple. As shown in Fig. 3.8, NIL uses a hard mould that contains nanoscale features defined on its surface to emboss into polymer material cast on the wafer substrate under controlled temperature and pressure conditions, thereby creating

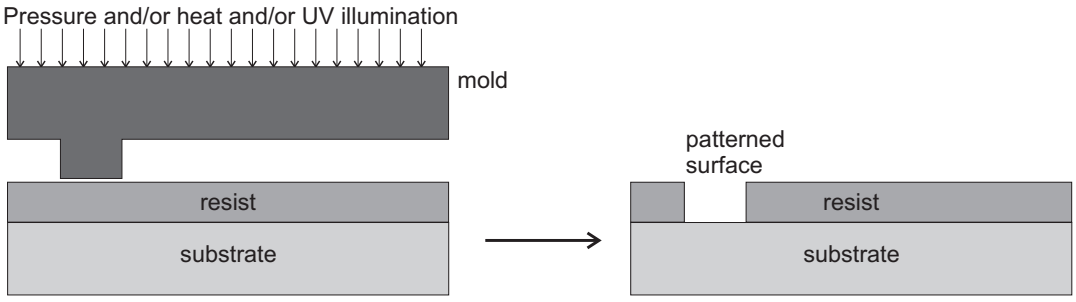

Figure 3.8: Principles of nano-imprint (NIL) and step-and-flash imprint lithography (S-FIL) lithography.

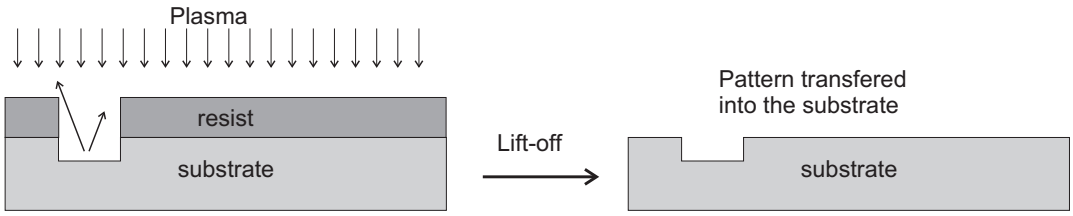

Figure 3.9: Etching principle.

a thickness contrast in the polymer material, which can be further transferred through the resist layer via an anisotropic etching based on an O_2-plasma process. Nano-imprint lithography has the capability of patterning sub-10 nm features, yet it only entails simple equipment and easy processing. This is the key reason why NIL has attracted wide attention within only a few years after its inception (Guo 2004). S-FIL is another great success as a mechanical printing method. In the S-FIL process, the substrate is first coated with an organic transfer layer; then a surface treated, transparent template with surface relief patterns is brought close and aligned to the coated substrate. Once in the proximity, a drop of low viscosity, photopolymerizable organosilicon solution is introduced into the gap between the template and the substrate. The organosilicon fluid spreads out and fills the gap under capillary action. Next the template is pressed against the substrate to close the gap, and the assembly is irradiated with UV light, which cures the photopolymer to make it a solidified and a silicon-rich replica of the template.

3.4.2 Pattern transfer into the substrate

In the previous section, we have presented the processing steps necessary to prepare the substrate for the actual structuring step. We have seen how the desired pattern is transferred into a mask, or directly written into the photosensitive layer on the substrate. In this section, we will focus on techniques applicable for the transfer of the mask pattern into the substrate.

Etching

Generally, we distinguish between isotropic and anisotropic etching processes (Sinzinger & Jahns 2003). For isotropic etching the substrate material is removed at the same speed (i.e., the etch rate) in all directions. In the case of anisotropic etching, the etch rates are different in different directions. Both anisotropic and isotropic etching processes are important for the fabrication of micro-optical components. But for the fabrication of diffractive optical elements (DOEs) such as subwavelength gratings, steep edges have to be etched. For this, anisotropic etching processes are used. There exists three main families of etching techniques: wet etching, dry etching and direct ablation sometimes called laser micro-machining.

Wet etching consists in immersing the masked sample into a liquid corrosive solution. Due to its high isotropy, the wet etching technique solution was not considered in this work since it appears to be incompatible with the kind of structures we will deal with.

Dry or plasma-enhanced etching, on the other hand, is a pure physical etching process, highly anisotropic (Fig. 3.9). Typical representatives of this group of technologies are sputter etching and ion-beam etching, which is sometimes referred as to ion-beam milling. A plasma is formed in the chamber containing the substrate or in a separate chamber for ion-beam processes. The chemically-inert plasma ions (e.g. Ar^+) are used to physically destroy the bonds of the substrate material and thus deplete the material. This etching is anisotropic because the ballistic effect

occurs mainly in the direction of the accelerated plasma ions (see Fig. 3.9). A chemical dimension can be added in order to increase the selectivity of the process between two different materials. Reactive ion etching (RIE) and reactive ion/plasma-beam etching (RIBE/RPBE) sometimes also referred to as chemically assisted ion-beam etching (CAIBE) makes use of both the ballistic effect and chemical reactivity of a beam of reactive ions to remove material or create structures into a substrate. The various parameters (gas melanges, beam energy, beam incidence, etc.) characterizing the etching process are optimized for the transfer of structures previously recorded in a masking layer into various materials. The interest of these techniques particularly comes from their directionality (vertical structures) and their high selectivity, which is the potential to efficiently etch one material and not another coexisting one.

For laser micro-machining applications excimer (excited dimmer) pulsed lasers of frequency-doubled Nd:YAG are generally used ($\lambda = 193$ nm). For both types of laser the fundamental principle of laser ablation is the same. It is initiated by the fast deposition of a large amount of energy directly into the substrate material, causing a dissociation of the substrate molecules. If this dissociation occurs fast enough, the material is removed from the substrate in a micro-explosion.

3.4.3 In situ monitoring

The different steps leading to diffractive components involve a design phase that must take technological limitations into account, a fabrication phase, and a final qualification phase. With current manufacturing techniques, feature lines and periods of the order of a few tens or hundreds of nanometers are manufacturable. For the determination of the grating profiles, for instance, scanning electron microscopy (SEM) is the reference metrology method. This approach is however time consuming, destructive (the wafer has in general to be cleaved and metallized) and limited to a small subset of the diffractive components of the etched wafer.

A well-known method for nondestructive testing based on the measurement and analysis of light scattered from a corrugated surface is optical scatterometry. In the case of periodic structures, light scattering takes the form of diffraction into orders. The information about the structure is routinely obtained from the distribution of intensities in the various diffracted orders. The inverse problem, which consists of reconstructing the grating geometry from the observed far-field data, is ill posed. In practice, a priori knowledge is introduced, and the grating profile is described by a finite number of parameters. Many independent measurements are performed by means of scanning over the angles of incidence and wavelength and by use of two orthogonal polarizations. Then the geometry that corresponds to the smallest deviation from the measurements is chosen.

In diffractive optics, this approach has been applied to a large variety of problems: metallic gratings, large-period gratings (Naqvi et al. 1994), subwavelength gratings (Marx & Psaltis 1997; Lalanne et al. 1999; Marciante et al. 2003; Yu et al. 2004). In semiconductor metrology, intensity measurements and ellipsometry are now widely used for in situ control and monitoring of growth and etching of periodic silicon surfaces (Blayo et al. 1995).

4

Use of subwavelength gratings

Contents

Abstract. *Developments in micro-lithography and associated techniques now make possible to practically use the special ability of subwavelength gratings to synthesize artificial media. From achromatic waveplates and anti-reflective structures to polarization-manipulation elements, the number of potential applications now attracts many research interests in various fields. In this chapter, we will review briefly the principal applications of subwavelength while lingering on those that are potentially interesting for astrophysical subjects like phase coronagraphy and nulling interferometry.*

4.1 Subwavelength gratings as phase retarders

One-dimensional subwavelength gratings, i.e., subwavelength gratings that are modulated along a single dimension and therefore supposed infinite along the other, can act as vectorial phase retarders, i.e., leading to a phase difference taking place between the orthogonal ordinary and extraordinary polarization states TE (or s) and TM (or p). For this reason, they can be engineered to provide artificial birefringent waveplates with very specific properties, in transmission as well as in reflection. As seen before, the physical reason for their artificial anisotropy or "form birefringence" is that the impinging light literally sees two different media as its wavefield vibrates in parallel or perpendicularly to the grating grooves, leading to two different effective indices for the orthogonal polarization states TE and TM.

4.1.1 Transmission mounting

The propagation through the artificial birefringent medium leads to the following phase shift

$$\Delta\phi_{TE-TM} = \frac{2\pi}{\lambda}\, h\, \Delta n_{TE-TM} \tag{4.1}$$

$\Delta n_{form} = \Delta n_{TE-TM}$ is called the "form birefringence". The term "form" has to be emphasized. Indeed, this property is essentially given by the geometry of the structure, no longer only by the intrinsic characteristics of the materials (Flanders 1983; Xu et al. 1995; Jin Kim et al. 1995; Schmitz et al. 1995; Brundrett et al. 1996).

The wavelength dependence of the effective indices n_{eff}^{TE} and n_{eff}^{TE} (see Eqs. 3.48 and 3.49) is also found in the form birefringence. This phenomenon, which is amplified just before the frontier of the resonant domain, is called the *dispersion of form birefringence*. Kikuta et al. (1997) were the first to propose achromatic quarterwave plates made of subwavelength grating structures using this property. They showed that the key point is to carefully control the geometry of the grating structure to tune the form birefringence in order to compensate for the hyperbolic dependence of the phase shift ($\propto 1/\lambda$, see Eq. 4.1) and thus make it achromatic. In practice, one has to search to make the form birefringence Δn_{form} be proportional to $\lambda/2h$ (see Fig. 4.2).

The subsequent phase shift can be made quasi-achromatic over reasonable spectral ranges. Typically, π-phase shifts (halfwave plates) with error standard deviations varying between $\sigma =$

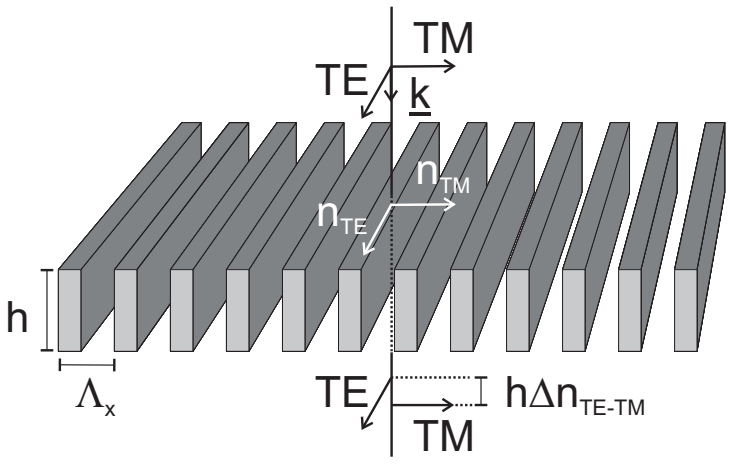

Figure 4.1: The transmission through a subwavelength grating leads to a differential phase shift between the polarization components TE and TM.

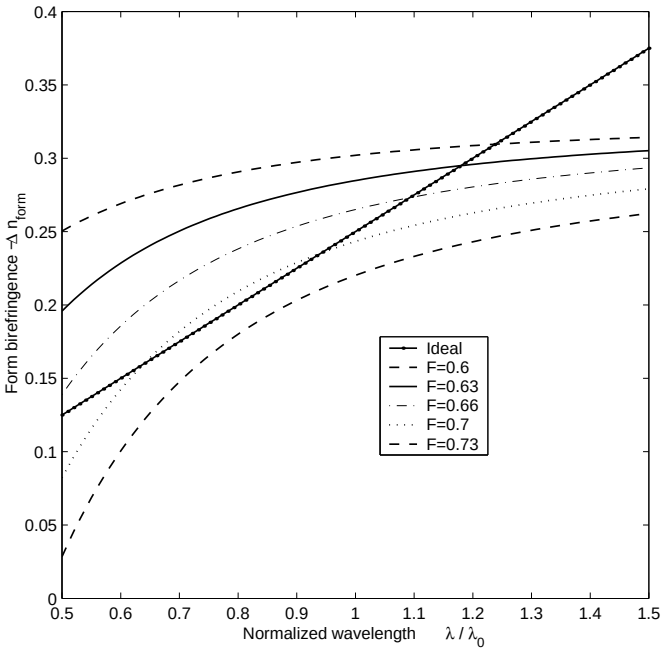

Figure 4.2: RCWA-calculated form birefringence $\Delta n_{form} = \Delta n_{TE-TM}$ of a silicon subwavelength grating, with respect to the normalized wavelength for various filling factors F between 0.6 and 0.73. The "ideal" quoted curve is the ideal birefringence that would lead to a perfect achromatization, i.e., proportional to $\lambda/2h$. Unfortunately, this idyllic behavior is only tangentially approached in practice.

5×10^{-3} and $\sigma = 10^{-2}$ radian rms can be obtained over bandpass with a spectral resolution $R_\lambda \approx \lambda/\Delta\lambda \approx 5$. A representative example in the K band ($2 - 2.4\,\mu$m) is shown in Fig. 4.3 (left) where standard deviation of the phase-shift error with respect to π is $\sigma = 7.5 \times 10^{-3}$ radian rms. The structure corresponding to this performance is a silicon surface-relief grating with a period $\Lambda_x = 475$ nm, a thickness of 2.956 μm and a silicon/silica filling factor $F = 85\%$, i.e., with an aspect ratio[14] of ~ 7. The transmittances shown in Fig. 4.3 (right), which remains substantially different for the orthogonal polarization states TE and TM, is above 90% for both of them despite the high refractive index of the silicon substrate ($n_{Si} \approx 3.4$). Such a high value is reached thanks to a 360 nm-thick ($\lambda/4n$) silicate anti-reflective layer on top of the grating. Optimization of the transmittance properties of subwavelength gratings in the transmission mounting probably constitutes their main weakness as we will discuss in the following chapter.

Bokor et al. (2001) theoretically investigated the possibility of making achromatic phase retarders with subwavelength gratings in slanted illumination. They numerically obtained the relatively modest performance of $90° \pm 5.9°$ over the $470 - 630$ nm wavelength range, and with a design somewhat uneasy to handle since the component is optimized at a high incidence angle ($\sim 60°$). Kikuta's design, by contrast, is optimized for normal incidence making it easier to insert in an existing optical train. Moreover, Kikuta's approach was also practically demonstrated in the mid-infrared by Nordin & Deguzman (1999) and Deguzman & Nordin (2001). They indeed manufactured a broadband form-birefringent quarterwave plate for the 3.5 to 5 μm wavelength range by engraving an optimized subwavelength grating in a silicon substrate. They measured a phase retardation varying from 89° to 102° over their selected bandwidth, in accordance to the RCWA predictions. Recently, the technology was implemented in the visible wavelength range by

[14]The aspect ratio of a grating is defined as the ratio of its thickness to the feature line: the higher the aspect ratio, the more difficult the fabrication.

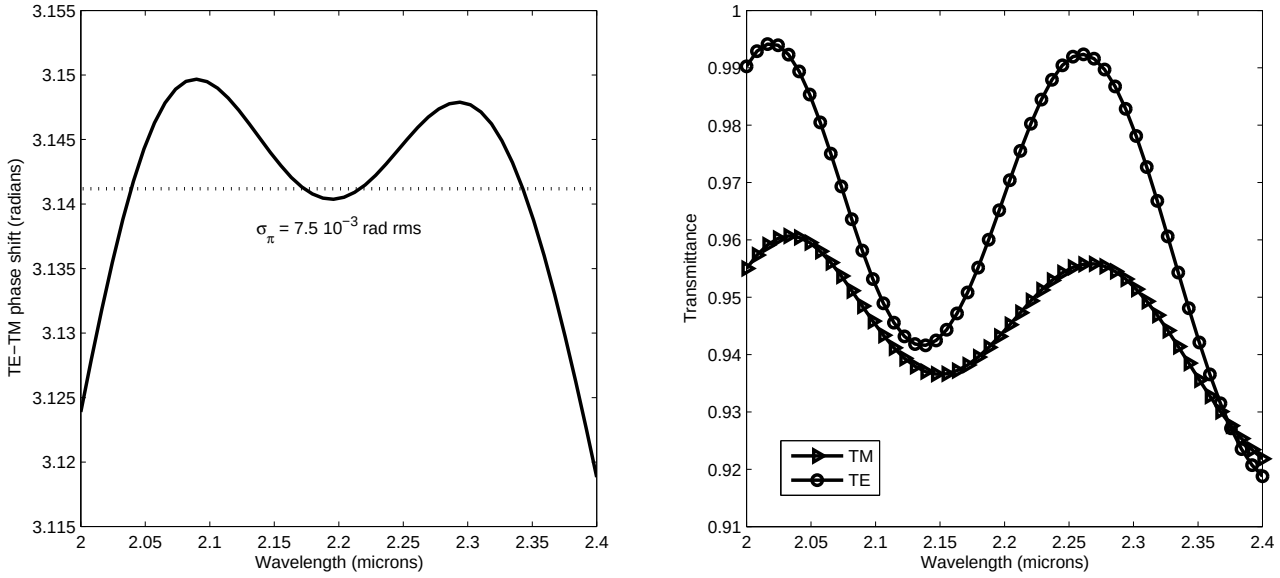

Figure 4.3: RCWA-calculation results (*TE-TM* phase shift on the left, and *TE-TM* transmittances on the right) in the transmission mounting for a silicon/silica subwavelength grating halfwave plate. The grating parameters were optimized to provide the best achromatic phase shift for the K band (2-2.4 μm) with the selected materials.

Yi et al. (2003) and Yu et al. (2006). It is even commercialized for optical pickup units (Deng et al. 2005). The performance of the optical pickup device is the current state-of-the-art reference for this technology with a measured retardance of 90°±3° and a mean transmittance greater than 95% over the 640-800 nm wavelength range, which is somewhat close to the theoretical limit derived here above.

It is worth noting that Yang & Yeh (1996, 1997) and Wang et al. (2005a) proposed to use, and successfully measured photoinduced form birefringence in photopolymers. Unfortunately, due to the faint index modulation, volume gratings are inappropriate to synthesize achromatic waveplates. Another idea to further enhance the form birefringence was suggested by Han & Kostuk (1996): engraving a subwavelength grating in a uniaxial birefringent material would allow reducing the thickness of the sample. But, they also observed nonlinear effects when the natural and artificial optical axes were crossed.

4.1.2 Reflection mounting

High-spatial-frequency lamellar gratings are also known to function as phase compensators, quarterwave and halfwave phase retarders, and polarization rotators that operate on zero-order specularly reflected beams (see Fig. 4.4). In other words, subwavelength gratings are also able to handle the reflection mode to provide vectorial phase shifts. Just as in the transmission mounting, controlling the geometrical parameters of these gratings allows engineering the phase retardation and polarization conversion introduced to the reflected beam. However, it is also known that variations of wavelength and polar angle of incidence affect the performance of these elements more strongly than variations of other geometrical and operational parameters (Haggans et al. 1993; Kleemann & Guther 1993).

In the particular total internal reflection (TIR) configuration, we have discovered that the

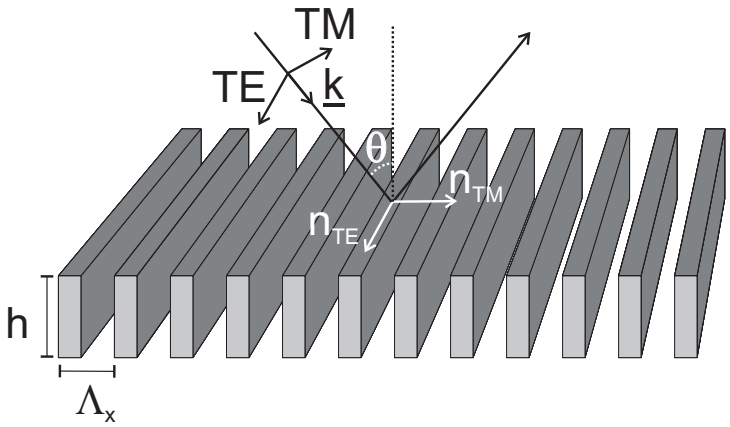

Figure 4.4: Reflection upon a subwavelength grating leading to a differential phase shift between the polarization components TE and TM.

behavior of subwavelength gratings was appropriate to generate super-achromatic phase shifts (Mawet et al. 2003, 2005a). In fact, the TIR phenomenon is known to come with a differential retardation between the vectorial TE and TM polarization components taking the following classical form (Born & Wolf 1999a)

$$\Delta\phi_{TE-TM} = 2\arctan\left[\frac{\sqrt{\sin^2\theta - n_{ti}^2}}{n_{ti}^2\cos\theta}\right] - 2\arctan\left[\frac{\sqrt{\sin^2\theta - n_{ti}^2}}{\cos\theta}\right] \tag{4.2}$$

where θ is the angle of incidence, greater or equal to θ_c, the critical angle defined as $\sin\theta_c = n_{ti} = n_t/n_i$, with n_i and n_t the refractive indices of the incident and external media, respectively. This property is exploited in the well-known Fresnel rhombs (Anderson 1974, 1988; Rochford et al. 1997; Rochford & Wang 1997; Wang et al. 1997). We found that engraving a subwavelength grating on the TIR interface leads to a significant improvement over the Fresnel-rhomb technology which is limited by the intrinsic index dispersion of the material used (Mawet et al. 2003, 2005a). Indeed, interaction between the subwavelength grating and the vectorial electromagnetic field leads to interesting effects on the phase of the external propagating fields, the TIR configuration ensuring the 100% diffraction efficiency in the reflected specular beam. The principle of what we named the total internal reflection grating achromatic phase shifter (TIRG APS) is to optimize a subwavelength grating in the TIR incidence condition to induce a super-achromatic phase retardance.

Subwavelength gratings in TIR can act as achromatic waveplates on their own but they can also be optimized to compensate for untreated surfaces in the optical train. Super-achromaticity can indeed be reached by using traditional Fresnel-rhomb phase retarders, i.e., several bare TIR interfaces (4 in the case of a double Fresnel rhomb), provided that at least one of them is engraved with an optimized subwavelength grating. In Fig. 4.5, we show the $\pi/2$-phase shifts induced by the cascade combination of two bare $ZnSe$/air interfaces in total internal reflection ($\theta = 65.06°$), the same with engineered subwavelength gratings engraved on the TIR interfaces, and the combination of one bare interface and the optimized grating. It clearly appears that the optimized subwavelength grating exactly compensates the natural drift of the bare $ZnSe$ TIR phase shift. Flexibility given by each design parameter defining the grating structure indeed allows a fine tuning of the phase-shift dispersion by the required amount for a substantial and benefic compensation.

Quantitatively, achromatizations of $\sigma \leq 10^{-3}$ radian rms can theoretically be obtained over wavelength ranges with a spectral resolution as low as $R_\lambda \approx \lambda/\Delta\lambda \approx 1$. An application of such a

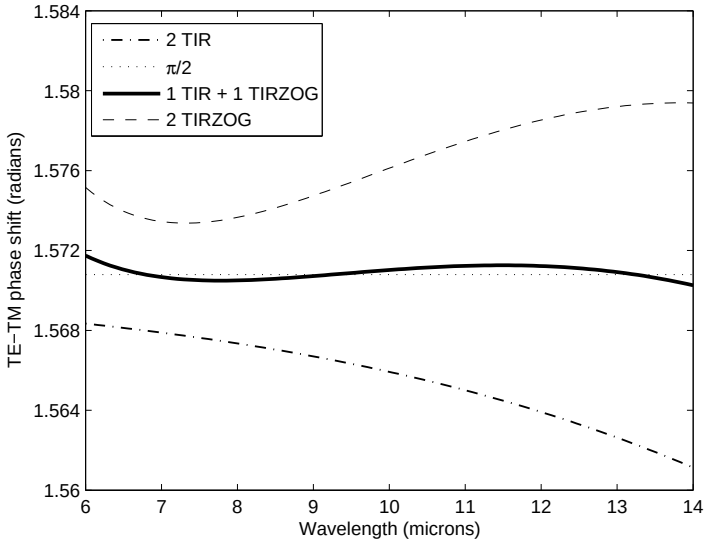

Figure 4.5: RCWA-calculation results in the reflection mounting. $ZnSe$ subwavelength grating quarterwave plate in TIR incidence in the 6-14 μm wavelength range. A single subwavelength grating is able to compensate for the intrinsic dispersion of the material.

structure will be presented in Chapter 8 where we will perform the design of an optimized TIRG APS for nulling interferometry.

Phenomenologically, the use of subwavelength gratings in TIR incidence implies that the field interacts with the structure only by means of its evanescent waves. The phase shift between the orthogonal polarization components arises due to the fact that their associated vanishing fields penetrate more or less deeply into the modulated region, inducing a pseudo-optical path difference responsible for the phase shift (Fig. 4.6). This differential skin effect can be explained by the anisotropy of the TE-TM zero-order effective indices. The achromaticity of the subsequent phase retardance can be understood from the particular grating-induced artificial dispersions of

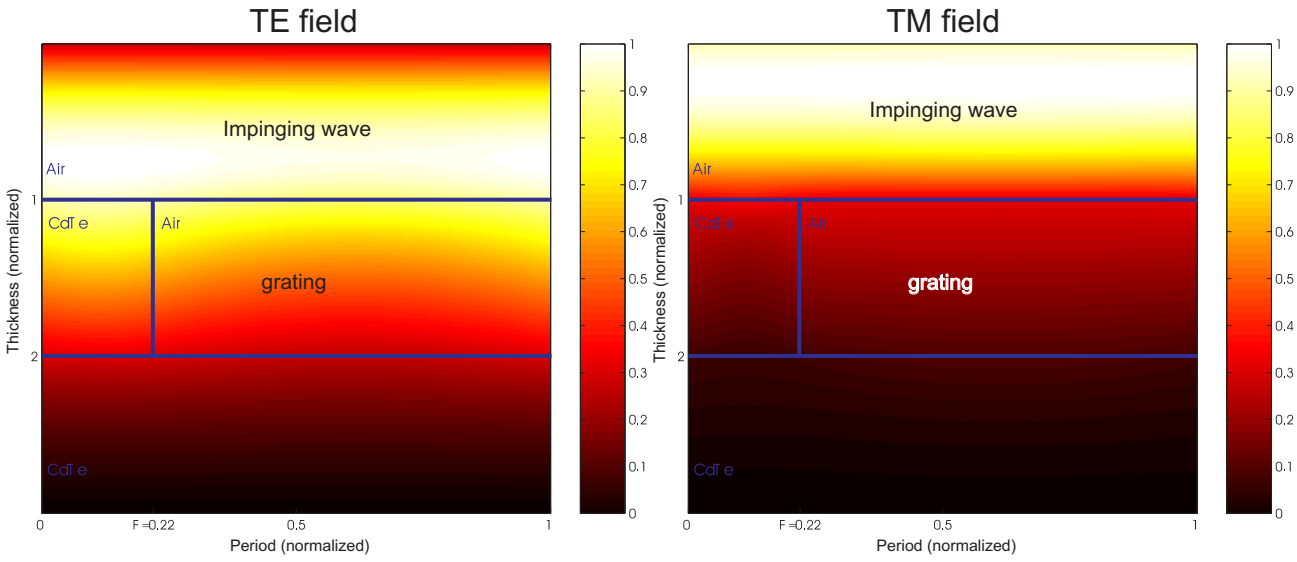

Figure 4.6: Electromagnetic field visualization for the TE and TM amplitudes illustrating the differential skin effect responsible for the phase shift. Indeed, it can easily be seen that the TE field penetrates more deeply than the TM one.

the form birefringence but, most of all in the reflection mounting case, by the complex interaction between the higher order vanishing modes.

It is to be mentioned that Liu & Azzam (1996) proposed one- and two-dimensional subwavelength surface-relief gratings on gold substrates to perform a quarterwave retardance upon reflection at 10.6 μm. However, their theoretical modelling is questionable because their use of effective medium theory is inappropriate. Kettunen & Wyrowski (1998) also suggested a hybrid approach where the grating is optimized as if it were to be used in transmission, but is in fact illuminated in a quasi-normal incidence reflection mode. This configuration is possible thanks to the grating implementation on top of a dielectric high-reflection stack, or a reflective metal layer. The advantage of this method over the purely transmission-mode phase retarder is the reduced aspect ratio of the structures, making them easier to manufacture. Indeed, the incoming beam passes twice in the structure as it goes down and up so that the effective thickness is twice the real one.

4.2 Subwavelength gratings as anti-reflective structures

Bernhard (1967) discovered that the corneas of night-flying moths was covered with a fine regular hexagonal array of protuberances which has a period of about 200 nm, similar depth, and with a cross section that is approximately sinusoidal (Fig. 4.7). This geometry induces a natural index gradient which was identified as a natural mean to reduce the reflection over a wide spectral and angular bandwidth, improving the moth's camouflage. This principle is the same as the one used for anechoic chambers.

4.2.1 Fresnel parasitic reflections

Fresnel reflections from surfaces in optical systems are cumbersome, particularly when there are many surfaces in the system: the total power loss can be considerable, and stray light that is due to the reflections tends to reduce the contrast in imaging systems. Optical coatings are the standard way to reduce reflections from optical surfaces, but they are rather expensive and unfortunately

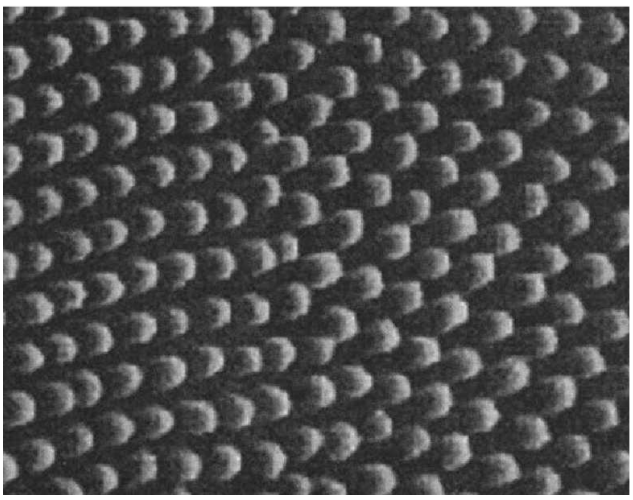

Figure 4.7: SEM picture of the cornea of a night-flying moth showing an hexagonal array of protuberances which has a period of about 200 nm, similar depth, and with a cross section that is approximately sinusoidal.

often a source of critical issues in numerous applications like broadband treatments for instance. Indeed, multi-layer thin films have problems associated with limitations in the coating materials, especially in the infrared (Shanbhogue et al. 1997). Classical coatings also exhibit various physical and chemical side effects affecting adhesion, thermal mismatch and the durability of the thin-film stack (Traylor Kruschwitz & Pawlewicz 1997). These problems are of course exacerbated in astrophysical applications, especially in space where stability constraints are emphasized.

In fact, unwanted Fresnel reflections come from the sudden transition from an optical medium of refractive index n_I to another one characterized by a different index n_{III}. The fraction R of light intensity which is reflected at normal incidence is given by the well-known Fresnel relation

$$R = \frac{(n_{III} - n_I)^2}{(n_{III} + n_I)^2} \tag{4.3}$$

If, like on the moth's eye, there is a gradual change of index, we can regard the net reflectance as the resultant of an infinite series of infinitesimal reflections at each incremental change of index. Since the reflections come from different depths from the surface, all of them will have different phases. If the transition takes place over at least a significant fraction of a wavelength, the resultant reflection can be significantly reduced because all phases will be present and destructively interfere with each other (Wilson & Hutley 1982).

This kind of "impedance matching" between the incident and substrate indices can in fact be implemented via two methods: the classical quarterwave transformer design and the graded- or tapered-index design (Collins 1966). The quarterwave transformer design requires a stack of discrete index layers on a substrate, where the depth of each layer is a quarterwave in the layer, and the index of refraction of each layer is designed with use of the optimal Tschebyscheff synthesis technique. Indeed, the so-called quarter-wavelength Tschebyscheff impedance-transformer method, widely used in antenna applications, allows the matching of the impedances (refractive indices in this context) of the external medium and the substrate, and can theoretically achieve any desired attenuation and bandwidth specification (Riblet 1957; Young 1961). The resultant interference coating produces a bandpass filter, where the bandwidth of the bandpass region and the

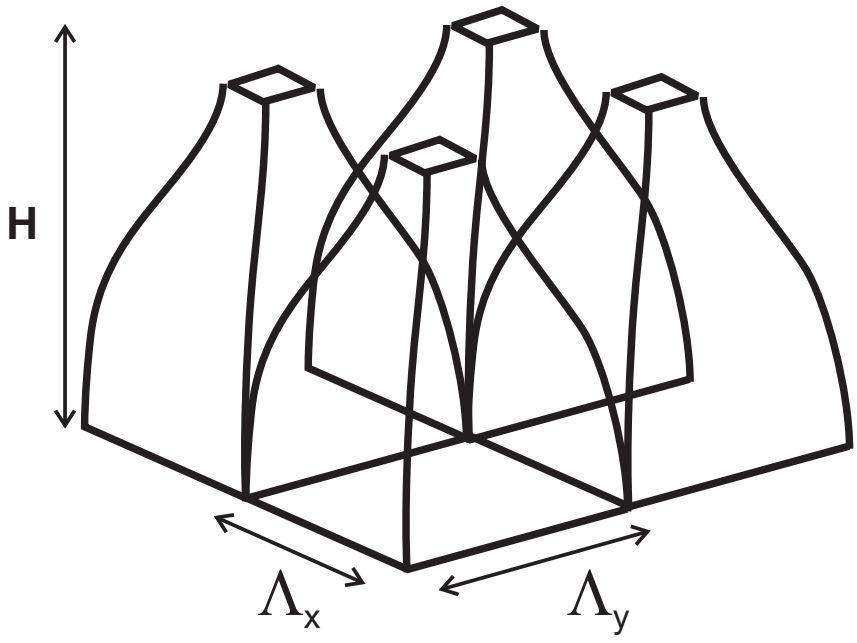

Figure 4.8: Klopfenstein tapered subwavelength grating structure as depicted in Grann & Moharam (1996).

maximum threshold level depend on the number of discrete quarterwave layers placed on the substrate. The graded-index matching design requires to continuously vary the refractive index from the incident region into the substrate region. The length of the taper and the index-distribution function are determined with use of the optimal Klopfenstein tapering technique (Klopfenstein 1956). This particular graded-index function (see Fig. 4.8 for its implementation with subwavelength gratings) will produce a high-pass or short-wave filter. The cutoff frequency is the lowest frequency (longest wavelength) at which the reflectivity is below the desired maximum threshold level.

Regardless of the design method chosen, it may not be possible to obtain actual materials with the properties that are needed, thus reducing the effectiveness of the AR surface. The subwavelength grating artificial properties can be envisaged to synthesize such materials, just like nature did for night-flying moths. It was in fact experimentally verified very soon[15] by Derrick et al. (1979) and Wilson & Hutley (1982) who proposed to use "crossed gratings" as anti-reflective structures. Later on, with the advent of powerful numerical tools like the RCWA, this very important application of subwavelength gratings was extensively studied in the 1D case. According to the adopted geometry, the reflection of the surface can indeed be reduced to a level as low as that achieved with very complex multilayer antireflection design (Gaylord et al. 1987; Brundrett et al. 1994; Brauer & Brygdahl 1994; Dos Santos & Bernardo 1997). The subwavelength properties of gratings involved here insure that the transmitted wavefront is NOT altered (Raguin & Morris 1993a).

Although 1D gratings are anisotropic and therefore can only be used as anti-reflecting structures for a single polarization at a time, the principle gave birth to numerous successful practical demonstrations (see, e.g. Smith et al. 1996; Kanamori et al. 2001; Brundrett et al. 1998) that have succeeded in convincing the industry of the potential applications of such components, e.g., for solar energy applications (Heine & Morf 1995). However, it quickly appeared that the dual-polarization use was mandatory in several fields where natural light is predominant like, for example, infrared optics for military and astrophysical instruments (Grann et al. 1995).

4.2.2 Structure of the anti-reflective subwavelength grating

Soon arose the following question: what is the surface-relief pattern for the optimum anti-reflection properties ? What surface-relief shape produces the least reflection over the broadest bandwidth ? Southwell (1991) tried to answer it with "pyramid-array" surface-relief structures. However, Southwell used approximate 2D-EMT (two-dimensional effective medium theory) closed-form expressions that were later on proven to significantly differ from the exact results obtained by the RCWA (Grann et al. 1995). This is a good example of the limitation of the effective medium techniques (Sect. 3.3.2) and a reason why they are no longer so used. Two years after Southwell, Raguin & Morris (1993b) determined the optimum design dimensions for multilevel two-dimensional subwavelength binary gratings with triangular or pyramidal profiles to minimize reflections. The optimum design dimensions were obtained by systematic adjustments of the depth and the filling factor of the profile.

Later on, Grann et al. (1995) established that, for the optimum design of tapered-grating structures, the best graded-index profile was expectedly produced by the Klopfenstein taper. They subsequently developed a technique to design such a taper with subwavelength gratings, using RCWA (Fig. 4.8). The theoretical performances of this optimal profile are extremely good with

[15] In fact, it is worth noting that Joseph Fraunhofer already noted that acid etching of certain glass surfaces substantially reduced the reflection.

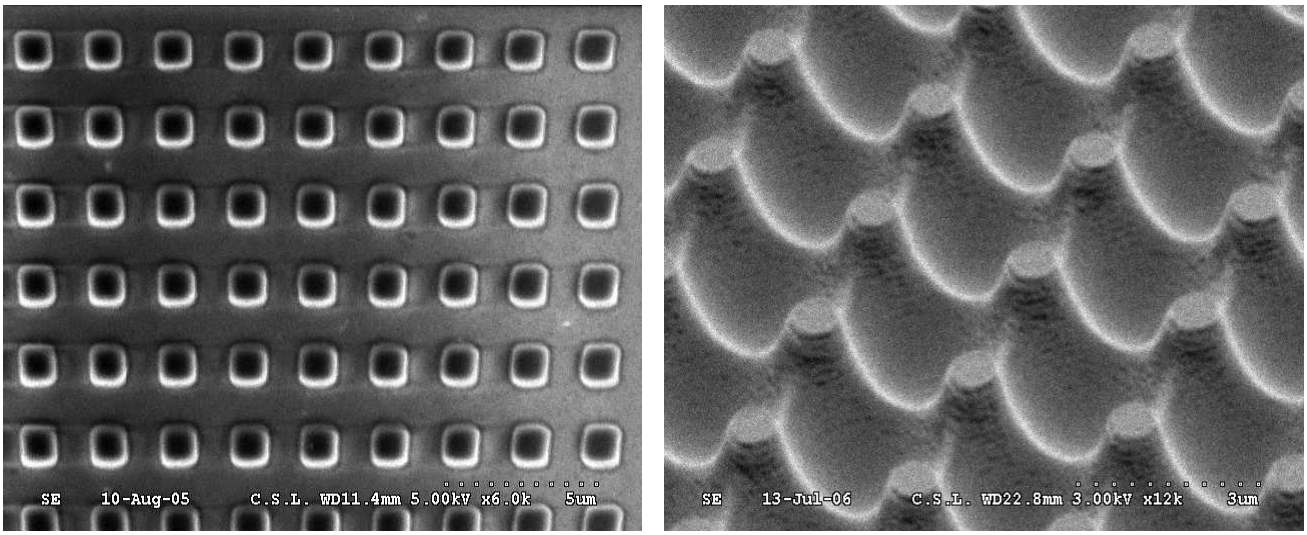

Figure 4.9: Left: photoresist pattern originated by DWL (SEM, view from the top). Right: crossed-sinusoidal profile obtained by holography on photoresist (SEM, oblique view). Courtesy of Denis Vandormael and Jérôme Loicq (CSL).

simulated results (Grann & Moharam 1996) characterized by reflectivities $\leq 0.3\%$ over bandwidths corresponding to a spectral resolution $R_\lambda \approx 0.8$. However, in practice, such a performance cannot be achieved, and the actual profile depends on the manufacturing process.

When we were asked to study anti-reflective subwavelength gratings in the framework of thermal infrared applications (i.e., the AR treatment of the dispersive plate APS for Darwin, see Sect. 2.2.1 and Nulltimate consortium 2002), we chose a more pragmatic approach, based on the actual available techniques for manufacturing subwavelength gratings (see Sect. 3.4): photolithography (Fig. 4.9, left), holography (Fig. 4.9, right) and lase micro-machining (Fig. 4.10). Even if photolithography can provide smooth profiles thanks to gray-level lithography (Sinzinger & Jahns 2003), usually, binary profiles are better mastered in terms of reproducibility and resolution. Photolithography has also the advantage that large areas can be treated (according to plasma-process capabilities), but large depths are in general difficult to achieve (Nikolajeff et al. 2000). On the contrary, holography naturally provides sinusoidal profiles. Two-dimensional symmetric profiles can be obtained with this technique by recording two 1D profiles, flipped by 90 degrees with respect to each other.

The use of excimer laser to directly ablate substrates naturally provides smooth profiles (see Fig. 4.10). It offers many other advantages (Dubreuil et al. 1998): new substrates can be tested very easily provided that the ablation curve (the ablated depth versus the laser fluence) is known, and high depths can be reached. However, the achievable resolution depends on the substrate, the excimer laser wavelength, and on the numerical aperture, NA, i.e., the optics used to image the mask on the sample. For example, working at $\lambda = 193$ nm with NA $= 0.15$, leads to a resolution W $= 1\,\mu$m. It is also to be noted that the technique can be investigated for curved surfaces. A disadvantage of excimer direct ablation is the small field available. This problem is balanced by the high repetition rate of excimer laser (100 Hz). Another issue concerns the inhomogeneity in the laser beam which has to be taken into account to achieve regular profiles since it is likely to create an overlapping between elementary grooved areas.

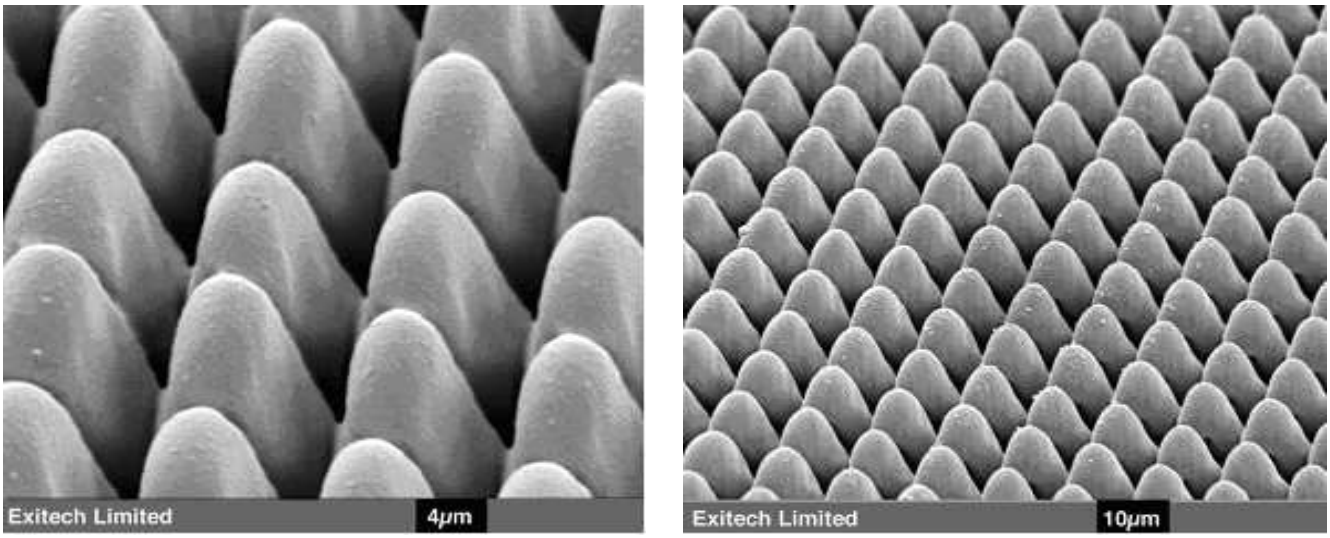

Figure 4.10: Egg-box, moth-eye, sinusoidal profile obtained by excimer direct laser ablation (SEM picture, © Excitech Limited).

4.2.3 Performance assessment

As already stated, our guideline for the performance optimization of subwavelength grating anti-reflective structures is pragmatic and in direct relation with standard manufacturing processes. We will indeed consider and simulate two types of gratings: according to photolithographic processes which naturally produce binary rectangular profiles, we will consider rectangular profiles and discretized pyramid profiles; according to the excimer laser direct ablation and holographic techniques which naturally produces sinusoidal profiles, we will consider sinusoidal profiles. Illustrative results are presented in Fig. 4.11 for the 6-11 micron wavelength range, corresponding to the Darwin first sub-band (see Sect. 8.3).

In order to theoretically demonstrate the capability of this technology, the calculations are performed for the two common infrared dielectric materials possessing the highest indices, Cadmium Telluride ($CdTe$: $n \approx 2.7$, see Annex B) and Germanium (Ge: $n \approx 4$, see Annex B). According to Eq. 4.3, without any treatment, a bare $CdTe$ substrate would lead to a parasitic reflection of 21% for one interface. It would even be worse for Germanium, leading to a single-face reflection of 36%.

The advantage of using such anti-reflection structures clearly appears in Fig. 4.11 with nevertheless strong differences according to the chosen profile. The simple rectangular profile in fact corresponds to a single quarterwave layer with an effective index depending on the filling factor. By appropriately tuning the filling factor, the layer effective index can be chosen to provide the exact impendance matching ($n_{eff} = \sqrt{n_I n_{III}}$) between the substrate and the incident medium, leading to a substantially better performance than with a traditional single-layer treatment where the actual choice of materials is always limited (Fig. 4.11, top left). Still, it is not comparable to the moth-eye profile performance, which, by realizing a perfectly smooth transition, decreases the parasitic reflection down to the percent level over very large bandwidths (Fig. 4.11, bottom left and right). An intermediate solution consists in approximating continuous profiles with stacks of binary gratings, leading to multi-level profiles (Fig. 4.11, top right).

Chalcogenide AR structures. In the framework of imaging optics for infrared applications, we have also performed RCWA calculations of anti-reflective structures for chalcogenide optical

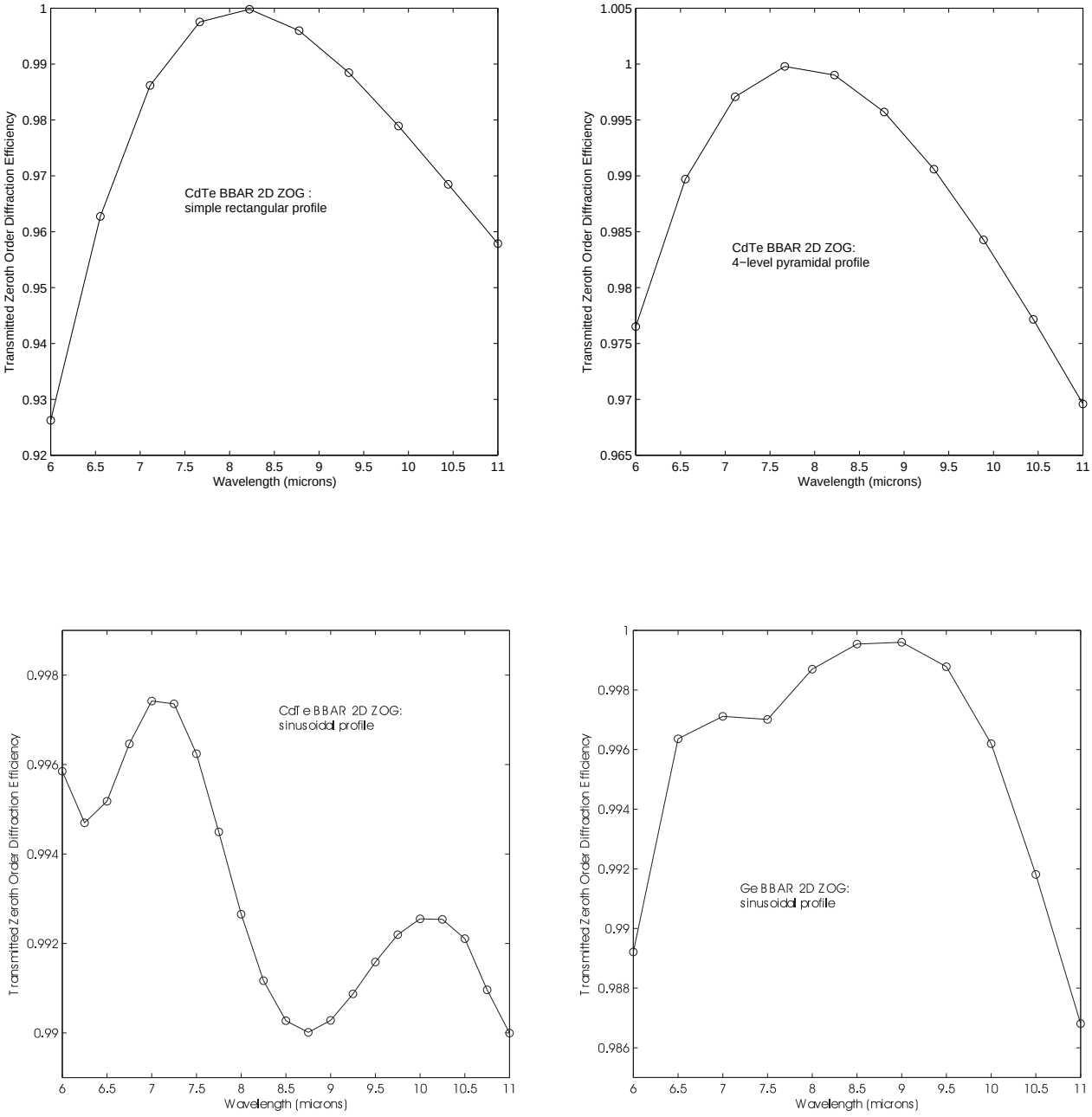

Figure 4.11: Anti-reflective performance of subwavelength gratings in terms of transmitted zeroth-order diffraction efficiencies, or equivalently single-interface transmittances (RCWA calculations). Top left: simple rectangular profile for a $CdTe$ substrate in the 6-11 μm wavelength range. The parameters of the structure geometry are: period of 2 μm, filling factor of 71%, thickness of 1.25 μm. Top right: idem with a 4-level pyramidal profile with a total thickness of 4 μm. Bottom left: idem with a crossed-sinusoidal, egg-box, or "moth-eye" profile with a total thickness of 7 μm. Bottom right: crossed-sinusoidal profile in the most constraining case of a Germanium substrate. The period is 1 μm and the thickness 9.5 μm, making it particularly difficult to manufacture.

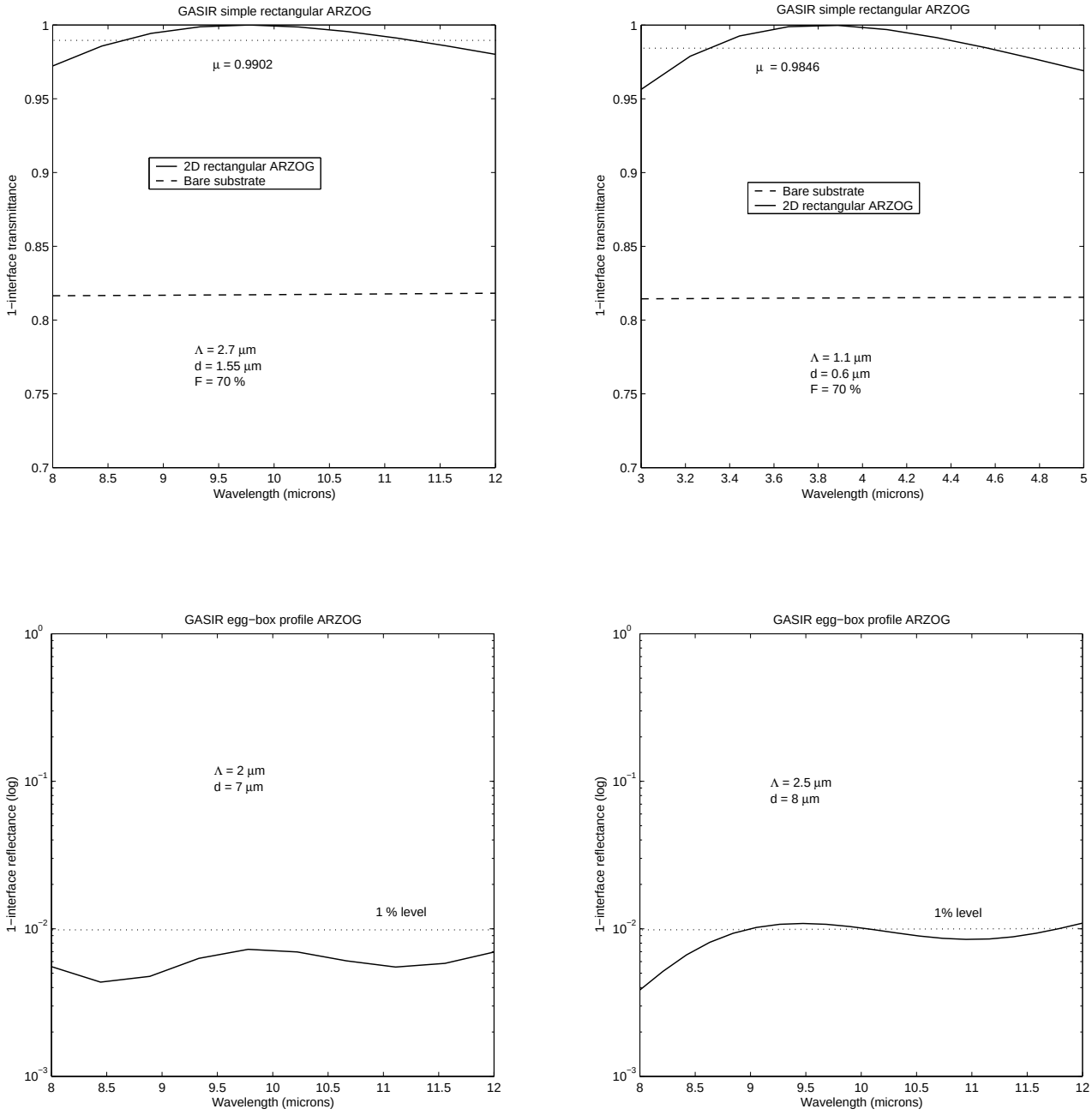

Figure 4.12: Anti-reflective performance of subwavelength gratings for GASIR (RCWA calculations). Top left: single-interface transmittance of a simple rectangular profile for a GASIR substrate in the 8-12 μm wavelength range. The parameters of the structure geometry are given in the different figures. Top right: idem for the 3-5 μm range. Bottom left: single-interface reflectance (logarithmic scale) for a crossed-sinusoidal, egg-box, or "moth-eye" profile. Bottom right: idem but another solution.

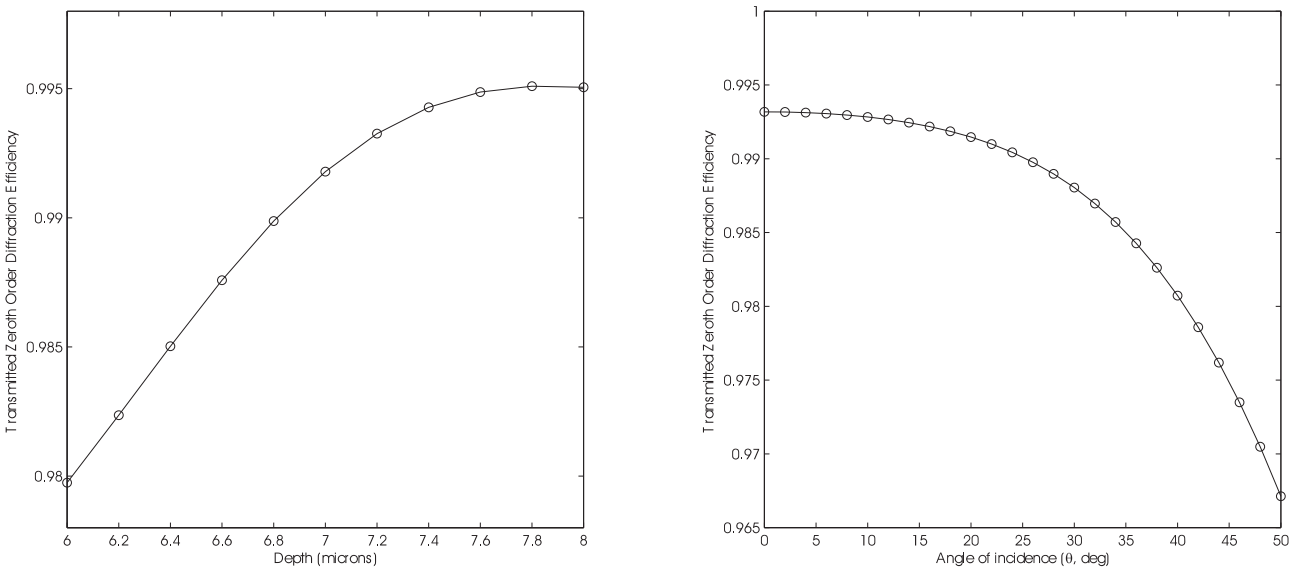

Figure 4.13: Anti-reflective subwavelength grating sensitivity to thickness/depth errors (left) and incidence variations (right) in the moth-eye case (RCWA calculations).

components (Fig. 4.12). The continuously increasing interest in the improvement of thermal imaging systems for spectral ranges extending up to the third window of atmospheric transparency (8-12 μm) has led to the development of suitable and low-cost optical materials (As_2Se_3, GASIR, IG6, etc.). Chalcogenide glasses are extensively studied for this purpose, and used both as bulk or fibred optical component.

4.2.4 Parameter tolerancing

We shall in this paragraph consider the AR performance sensitivity to parameter changes. The only restriction on the period comes, on one (upper) side, from the ZOG condition, and on the other (lower) side, from the manufacturing process capabilities. On the contrary, the response of the AR structure to a change in the total depth of the grating in the case of a moth-eye profile is sensitive at the one-micron level for thermal infrared applications (Fig. 4.13, left). Concerning the tolerance to incidence angle variations, the well-known wide acceptance of such structures is verified (Fig. 4.13, right). However, it is clear that the structure should be optimized for the chosen working incidence.

Concerning the lateral dimensions of the profiles it is known that, if the desired overall filling factor of the profile is preserved in the fabrication process, then nominal shape errors have a minimal effect on the desired diffraction characteristics (Pommet et al. 1995). The tolerance of simple rectangular treatments to parameter changes is more critical and, for thermal infrared applications, is at the 100-nm level for the thickness and feature line control. Moreover, the tolerance to incidence variations for simple rectangular profiles is of a few degrees only.

4.2.5 Diamond demonstrator

In collaboration with LESIA, we considered the manufacturing of a subwavelength grating anti-reflective demonstrator in diamond. The framework of this prototyping operation was the AR treatment of the Lyot coronagraph expected to equip the mid-infrared instrument (MIRI) of the

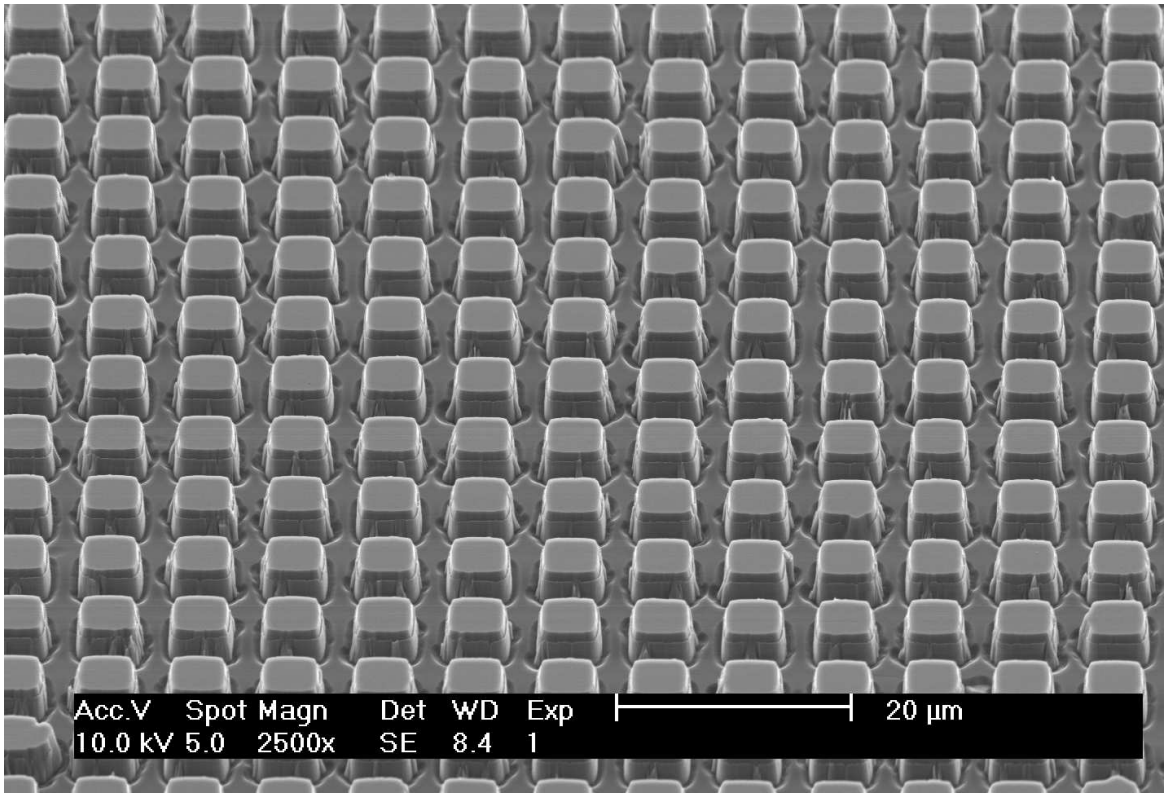

Figure 4.14: SEM picture of the anti-reflective subwavelength-grating demonstrator in diamond optimized for 24 μm. Courtesy of Mikael Karlsson (ADAMANTIS AB).

JWST (see Sect. 1.5.3). The working wavelength of this coronagraphic mode is around 24 μm with a bandwidth $R_\lambda \approx 5$. The anti-reflective structure should therefore be effective between 21.6 and 26.4 μm. Diamond was chosen because of its exceptional properties (see Annex B): hardness, transparency (from 200 nm to 1 mm), high thermal conductibility, low thermal expansion, etc. Progress in CVD (chemical vapor deposition) techniques make this material more and more available to continuously lower prices.

We designed a grating to satisfy the bandwidth requirements using the RCWA code of Sect. 3.2.3. Assuming a rectangular binary profile, optimizations led to the following parameters: a period of 7.8 μm with a thickness of 3.65 μm and a filling factor of 74%. The theoretical results are shown in Fig. 4.15 (left and right). The manufacturing was done by ADAMANTIS AB[16]. This small enterprize has indeed developed a very interesting know-how in the field of diamond micro-structuring (see e.g. Karlsson et al. 2001; Karlsson & Nikolajeff 2002, 2003).

The subwavelength grating was manufactured by standard lithography and etching techniques. A 125-nm thin aluminum film was first sputtered on top of a diamond substrate. Next, a 180-nm thin polymethylmethacrylate (PMMA) layer was spin-coated on top of the aluminum film, and a two-dimensional binary grating was structured in the PMMA layer by electron-beam lithography. PMMA was then used as an etch mask in an inductively coupled plasma (ICP) etching system to open up the aluminum. Etch parameters were: ICP power of 500 W, bias of 10 V, chamber pressure of 5 mTorr, flow rates of 45 sccm for BCl_3, and 5 sccm for Cl_2, with a total etch time of 5 min. This process was very stable and allowed a well-controlled pattern transfer from PMMA

[16]Adamantis AB is a small Swedish start-up company (Sept. 2003) issued from the technology department of engineering sciences of the Uppsala university. Adamantis is developing techniques for and manufactures diamond micro-structures based on customer demands. The company is currently working on diamond optical components for space applications.

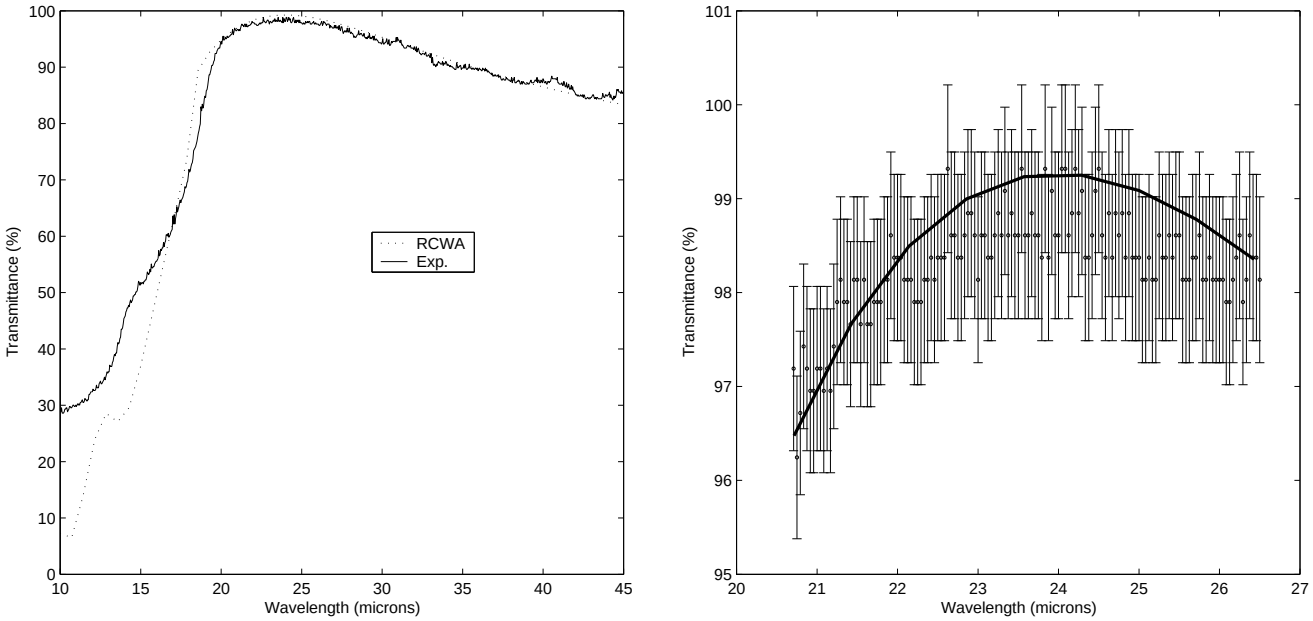

Figure 4.15: Anti-reflective subwavelength grating demonstrator in diamond optimized for 24 μm. Left: 10 to 45 μm spectrometer data (continuous line) and theoretical RCWA expected results (dotted line). Right: zoom on the 21-26 μm working wavelength range where the data points are shown with their error bars, and the RCWA calculations shown with a continuous line.

to Al.

The transfer into the underlying diamond substrates was done in an oxygen plasma in another ICP-etching system. The advantage of using ICP over other etch systems, such as reactive ion etching (RIE), is that ICP gives a high ion density and therefore short etch times. ICP systems also yield better anisotropy, due to low process pressure, and smoother etched surfaces than RIE. Carbon will easily form volatile compounds with oxygen radicals so that an oxygen plasma is therefore suitable for diamond etching. Etch parameters were: ICP power of 600 W, bias of -140 V, chamber pressure of 2.5 mTorr, flow rates of 7 sccm for O_2 and 8 sccm for Ar, with total etch times of 10-14 minutes. All samples were mounted with vacuum grease on the water cooled aluminum RF-chuck to enhance the thermal conductivity (to avoid burning of the resist). By first measuring the etch rate of partly covered diamond and knowing the desired grating depth, the etch time needed for fabricating the subwavelength grating could be calculated (the etch rate of diamond was measured to be 200 nm/min). Finally, the Al was stripped by wet etching. The final result is shown in Fig. 4.14. The excellent definition of the rectangular profile is clearly seen.

Optical characterization of the sample was done in Sweden before sending. Spectroscopy in reflection was used. The results are excellent and in very good agreement with theoretical predictions (Fig. 4.15). We asked the "Institut d'Astrophysique Spatiale" of Orsay (L. d'Hendecourt) to realize an independent measure with a different technique, the Fourier transform spectrometer (FTS). They obtained a slightly different result that was a few percent lower in transmittance. This discrepancy originates from the different illumination conditions in the two devices.

4.3 Other applications of subwavelength gratings

Apart from the phase retarders and anti-reflective structures, there are many other applications of subwavelength gratings. Their very special properties of artificial medium synthesis and their sensitivity to polarization have been known for quite a long time and seems at last to catch the interest of the scientific and engineer communities. Let us now briefly review the panel of possible applications of subwavelength gratings without going into details and skipping the applications of resonant phenomenon (for that, see e.g. Lenaerts 2005) like biosensors (see, e.g. Yih et al. 2006).

4.3.1 Polarization-selective diffractive optical elements

Phase-only diffractive optical elements (DOEs) are considered as an attractive technology for a variety of applications in photonics and optoelectronics where arbitrary wavefront generation is desired. The small size and ability of DOEs to generate complex wavefronts contributes to their popularity, particularly for applications involving monochromatic illumination. Most DOEs (Sinzinger & Jahns 2003) are designed by algorithms based on scalar diffraction theory and are fabricated with large features relative to the operational wavelength. Unfortunately, these characteristics render the DOE insensitive to polarization, which would offer an additional degree of freedom in optical and photonic system design. Exploiting the polarization indeed enables applications of such as polarimetric imaging and polarization-based switching.

Fabrication of form-birefringent micro-structure from a single substrate appeared very attractive. Indeed, as already proven, form birefringence permits a designer to engineer desired anisotropic properties onto a single substrate. Moreover, since the substrate does not require natural birefringence, conventional materials can be employed, which allows mature fabrication processes like photolithography to be used. Indeed, advances in modern lithography have removed the previous limitation by which only long wavelengths in the infrared were accessible to ZOG-based DOEs. Xu et al. (1995) and Schmitz et al. (1995) were the first groups to apply form birefringence to the fabrication of computer-generated hologram (CGH). Yu et al. (2000, 2002) extended form-birefringent design to two-dimensional polarization-selective CGHs. However, the individual phase tranformations were restricted to binary values (0 or π). Later on, Mirotznik et al. (2004) further extended the method to synthesize DOEs that nearly have an arbitrary number of phase levels, eliminating the artifacts produced by binary phase elements and allowing designers to manufacture more general polarization-selective DOEs.

4.3.2 Polarizers

High-spatial-frequency metal gratings have long been recognized as an effective polarizer option for the IR portion of the spectrum (Bird & Parrish 1960; Young et al. 1965). Wire-grid polarizers (WGP) are indeed extensively used as polarization-sensitive elements in various applications, ranging from remote sensing to displays for biomedical engineering, because of the excellent polarization performance and planar structure that allow them to be easily pixilated and integrated into other optoelectronic devices. The wire grids of a WGP are made of a good conductor; if the electric field is parallel to the wire grids (TE polarization), they absorb and do not sustain an electric vector at the surface. This makes the field strength of a TE polarization order propagating parallel to the surface much smaller than that of a TM order whose electric field is orthogonal to the wires of a WGP. In other words, an electric field oscillates orthogonally for the most part with respect to grating wires as the light is transmitted through a wire-grid grating polarizer.

For recent application examples, see, for instance, Nordin et al. (1999), and their micro-polarizers for infrared imaging polarimetry, or Deguzman & Nordin (2001) for their stacked subwavelength gratings as circular polarization filters. Let us also cite Wang et al. (2005b) who describe and demonstrate nanowire-grid polarizers for the near-infrared with excellent extinction ratios. It is to be noted that WGP are commercially available (Creech-Eakman et al. 2003). Polarizers are one of the most prolific applications of subwavelength gratings.

4.3.3 Polarizing beam splitters

Polarizing beam splitters (PBSs) have numerous applications, such as magneto-optic data storage in optical information processing and optical switching in optical communication. Conventional PBSs, such as Wollaston prisms and PBS cubes, are both bulky and heavy, or applicable in a narrow wavelength range. The compact size and light weight of subwavelength gratings are advantageous to the miniaturization and integration of optical systems. Lopez & Craighead (1998) calculated and manufactured a device that acts as a quarterwave plate at normal incidence and as a polarizing beam splitter at an angle of incidence of ~ 40 degrees. Their device is made of a multi-layer SiO_2/Si_3N_4 surface-relief subwavelength grating with a period of 0.3 μm, designed for an operating wavelength of 632.8 nm. They measured encouraging extinction ratio of about 25. Yi et al. (2004) recently proposed a novel broadband polarizing beam splitter. Their compact design is made of a sandwiched SiO_2/Si subwavelength grating with a period of 100 nm, theoretically providing a high polarization extinction ratio (≥ 1000) in a broad spectral range (from 1.3 to 2.3 μm).

4.3.4 Distributed index medium

Blazed binary subwavelength gratings

For beam deflection, the so-called blazed gratings can be designed to divert nearly 100% of the incident power into a single diffracted order. However, it is not easy to fabricate the continuously varying surface profile of the blazed grating with existing technology. To facilitate the manufacturing of such components, structures with multiple discrete surface levels were introduced. According to the theory, a 16-level structure can deflect 99% of input beam power to a designated direction. But the multi-step alignment, particularly when the required feature size is less than one wavelength, makes the manufacturing difficult once again. A binary-level subwavelength grating with a space-variant (see here below) filling factor can mimic the quasi-linear phase transmittance of a continuous blazed grating (Zhou & Drabik 1995). The advantage is of course at the manufacturing step, which is easier for binary structures than for continuous ones. In some cases, subwavelength gratings have also been demonstrated to provide efficiencies larger than those of conventional discretized blazed gratings (Lalanne et al. 1998; Astilean et al. 1998; Lee et al. 2002).

Artificial graded-index medium

Graded-index medium are used for guiding, imaging, optical signal processing, mode matching, coupling, and other applications. A graded-index profile can be achieved by gradually modifying the filling factor of a one-dimensional subwavelength grating engraved into a slab waveguide along one dimension, while the beam is confined in the other dimension by the waveguide. The use of

subwavelength gratings offers the intrinsic advantages of on-chip integration such as miniaturization, eliminating the need to align each component separately, and compatibility with standard micro-fabrication techniques for manufacturability (for further details, see Levy et al. 2005).

4.3.5 Space-variant implementation of subwavelength gratings

Subwavelength gratings are said to be space-variant when the local characteristics (period, orientation of the grating lines, etc.) of the structure spatially vary from point to point. Such components are now extensively studied as polarization-control elements. The idea of using subwavelength gratings with a space-variant implementation was first introduced by Davidson et al. (1992). They showed that, by controlling the local direction and geometry of the subwavelength grating, any polarization change and continuity can be obtained. It is of first importance to note that sometimes, elements that can provide nonuniform space-variant polarization are required and that such space-variant polarization manipulators are difficult to produce with natural birefringent elements. Here are some example of practical applications:

- elements for transforming the polarization of high-power CO_2 laser beams, e.g. converting the azimuthal polarization into a linear polarization (Oron et al. 2000; Bomzon et al. 2002a) or producing linearly polarized light with axial symmetry (Niv et al. 2003);
- computer-generated space-variant polarization elements with subwavelength metal stripes for polarization coding of data in optical communication, optical computers and neural networks, optical encryption (Dahan et al. 2005), tight focusing, beam shaping (Levy et al. 2004a), and particle trapping and acceleration (Bomzon et al. 2001b);
- elements for the generation of Pancharatnam-Berry geometrical phase (see the definition in Sect. 6.3) in space-variant polarization-state manipulations (Bomzon et al. 2001c) for the formation of helical beams (useful in atoms/particles-trapping applications, see Biener et al. 2002), polarization-dependent focusing lens (Hasman et al. 2003) or propagation-invariant vectorial Bessel beams (Niv et al. 2004), for instance;
- integrated components for real-time polarimetry (Bomzon et al. 2001a, 2002b,c; Biener et al. 2003b) which is very useful in military and civil remote-sensing applications;
- computer-generated depolarizers (Biener et al. 2003a) which are optical elements reducing the degree of polarization of beams, independently of their incident polarization state with application in optical measurement equipments, for instance;
- point-spread function shaper (Tsai et al. 2006);
- optical vortices (Niv et al. 2005b,a; Levy et al. 2004b) which will be considered in the following chapter for their application in coronagraphy.

Let us conclude this chapter on the very interesting subject of optical vortices. It will be extensively studied in Chapter 6. We will present and demonstrate the interest of using space-variant subwavelength gratings optimized for coronagraphy thanks to the generation of achromatic optical vortices.

Part III

Phase-mask coronagraphy

5

Four-quadrant zero-order grating phase-mask coronagraph

Contents

Abstract. *We present in this chapter an original concept of phase-mask coronagraph implemented with diffractive optical elements consisting of optimized subwavelength surface-relief gratings. This integrated component is an evolution of the four-quadrant phase-mask coronagraph (FQPM), which solves the known π-phase shift chromaticity issue of the latter. We call it the four-quadrant zero-order grating phase-mask coronagraph (4QZOG).*

5.1 FQPM with ZOGs: 4QZOG

We have already presented the FQPM coronagraph implemented by means of achromatic halfwave plates (see Sect. 2.3). The concept of using birefringent elements in a nulling coronagraph has therefore been successfully qualified. However, the lack of birefringent materials in certain wavelength ranges like the thermal infrared, and some annoying issues like the bulkiness of the mounting, or simply the difficulty of cutting and assembling the different plates together made us consider the use of artificial birefringent elements (Mawet et al. 2006).

5.1.1 Implementation of the FQPM by means of subwavelength gratings

As stressed out in the previous chapter, subwavelength gratings are known to synthesize artificial birefringent elements. Their implementation into an integrated FQPM can therefore be considered

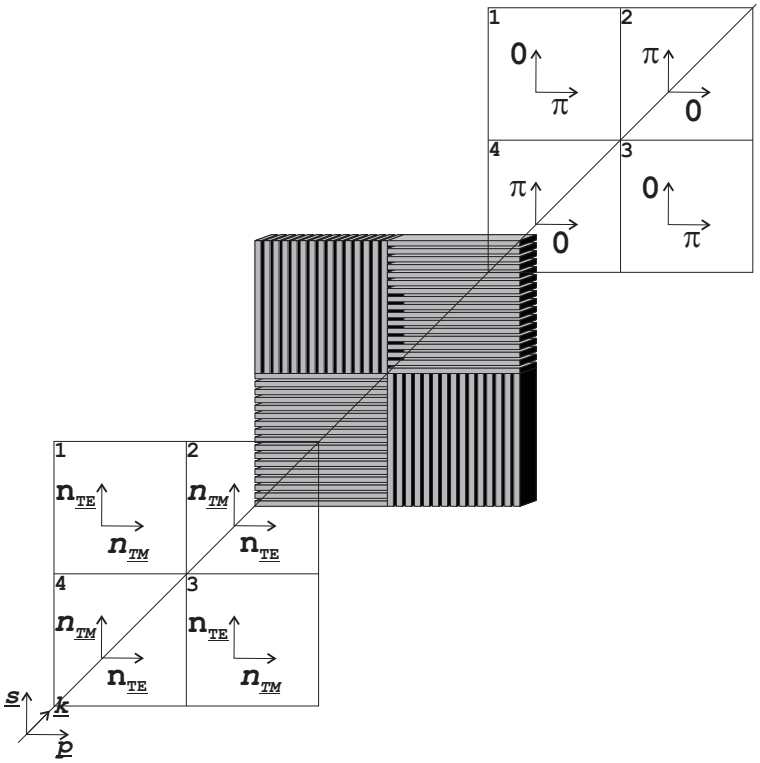

Figure 5.1: 4QZOG implementation. The four gratings engraved on a unique substrate are strictly identical and integrated in the following way: two of them in two quadrants along one diagonal are rotated by 90° around their normals with respect to the two others. This anti-symmetrical configuration achieves the FQPM particular focal plane π-phase shift distribution (see text for explanations).

and is illustrated in Fig. 5.1. Let $\underline{s}$ and $\underline{p}$ be the complex amplitude vectorial components of the incoming light. In each of the four quadrants, the $\underline{s}$ and $\underline{p}$ global polarization states can be decomposed in the $\underline{TE_i}$ and $\underline{TM_i}$ vectorial complex amplitudes according to the line orientations of the local grating in the i_{th} quadrant ($i = 1, ..., 4$). Indeed, the convention in normal incidence specifies that the electric field of the TE and TM components vibrates in parallel and perpendicularly to the grating lines, respectively. We have already shown in Sect. 3.3 that two effective indices n_{TE} and n_{TM} can be assigned to the corresponding TE-TM perpendicular polarization states. Let us now assume that:

1. the four gratings engraved on a unique substrate are strictly identical;
2. they are integrated with the following implementation: two of them in two quadrants along one diagonal are rotated by 90 degrees around their normals with respect to the two others.

The consequence of the first hypothesis is that the effective indices n_{TE_i} and n_{TM_i} are identical in each quadrant, i.e., $n_{TE_i} = n_{TE}$ and $n_{TM_i} = n_{TM}$, with $i = 1, ..., 4$. The second one gives the particular anti-symmetrical configuration of Fig. 5.1 with the following result.

If, by construction,

$$\begin{aligned} \Delta\phi_{TE_i-TM_i} &= \frac{2\pi}{\lambda} h\, \Delta n_{TE_i-TM_i} \\ &= \frac{2\pi}{\lambda} h\, \Delta n_{TE-TM} \approx \pi \end{aligned}$$

Then the potentially-interfering parallel polarization states along the $\underline{s}$-component direction for example are phase shifted in the following way (j being an adjacent quadrant of i)

$$\begin{aligned}\Delta\phi_{TE_i-TM_j} &= \frac{2\pi}{\lambda}\,h(n_{TE_i}-n_{TM_j})\\ &= \frac{2\pi}{\lambda}\,h(n_{TE_i}-n_{TM_i})\\ &= \Delta\phi_{TE_i-TM_i}\approx\pi\end{aligned}$$

The same goes for the $\underline{p}$-component direction. Consequently, for the $\underline{s}$- and $\underline{p}$-component directions respectively, the FQPM particular focal plane π-phase distribution is achieved, i.e., adjacent quadrants are out of phase. Therefore, it works in natural light.

We will call this new structure the 4QZOG (four-quadrant zero-order grating). It must be noted that the precision on the perpendicularity between the gratings is directly related to the coronagraphic performance. It has to be finely controlled (see Sect. 8.4).

5.1.2 ZOG specific optimizations

The ZOG optimization consists in optimizing the π-phase shift quality within a given spectral range for a well-chosen material. It also requires the equalization of the interfering fluxes, which is difficult to achieve. The merit function to be minimized is the null depth over the considered bandpass. The latter measures the darkness of the destructive interference taking place in the pupil plane following the phase-mask coronagraph focal plane and is directly related to the phase-shift error with respect to π, i.e., ϵ, and the flux ratio q between the orthogonal polarization states, according to Eq. 2.6.

A well-adapted structure for example consists in a rectangular profile grating covered with an AR layer which settles at the bottom of the grooves and on top of the ridges. As already mentioned, subwavelength gratings can accommodate a large variety of materials and wavelength ranges. The free parameters of the optimization concern the geometry of the grating: the period, filling factor and thickness of the grating itself, and finally the thickness of the AR layer. Several optimization algorithms can be used. Three were implemented and tested with our home-made RCWA tool.

The first one is a so-called "medium scale optimization". It uses the BFGS (Broyden 1970; Fletcher 1970; Goldfarb 1970; Shanno 1970) quasi-Newton method with a mixed quadratic and cubic line-search procedure. The BFGS formula is used for updating the approximation of the Hessian matrix. The DFP (Davidson-Fletcher-Powell) formula, which approximates the inverse Hessian matrix can also be used (Fletcher & Powell 1963; Fletcher 1980). The default line-search algorithm, is a safeguarded mixed quadratic and cubic polynomial interpolation and extrapolation method. This method generally requires fewer function evaluations but more gradient evaluations. It is worth noting that the function to be minimized must be continuous, and that this algorithm might only give local solutions.

The second one uses the "simplex-search method" as in Lagarias et al. (1998). This is a direct-search method that does not use numerical or analytic gradients as the medium scale optimization. If n is the length of x, a simplex in an n-dimensional space is characterized by the n+1 distinct vectors that are its vertices. In a 2D space, a simplex is a triangle; in a 3D space, it is a pyramid. At each step of the search, a new point in or near the current simplex is generated. The function value at the new point is compared with the function's values at the vertices of the simplex and, usually, one of the vertices is replaced by the new point, giving a new simplex. This step is

repeated until the diameter of the simplex is less than the specified tolerance. This algorithm is generally less efficient than the first one for problems of order greater than two. However, when the problem is highly discontinuous, it might be more robust, i.e., it can often handle discontinuity, particularly if it does not occur near the solution which is sometimes the case with our application. Also, it might only give local solutions.

"Genetic algorithms" have also been tried but with less success than the simplex-search method which practically revealed to be the most efficient, i.e., the fastest to converge. The *simplex* method was therefore used in most of the optimization procedures presented in this work.

5.2 Article: *Subwavelength surface-relief gratings for stellar coronagraphy*

In the following paper, published in *Applied Optics*, we present and discuss the implementation of a FQPM with subwavelength gratings. Optimized designs are presented and completely studied thanks to the RCWA code presented in Sect. 3.2.3. A complete tolerance analysis is also performed with the goal of assessing the feasibility of the component manufacturing by traditional lithographic techniques.

Subwavelength surface-relief gratings for stellar coronagraphy

Dimitri Mawet, Pierre Riaud, Jean Surdej, and Jacques Baudrand

We present a new design of a phase mask coronagraph implemented with subwavelength diffractive optical elements consisting of optimized surface-relief gratings. Phase mask coronagraphy is a recent technique that seeks to accommodate both high dynamic and high angular resolution imaging of faint sources around bright astrophysical objects such as exoplanets orbiting their host stars. The original design we propose is a new, integrated, and flexible solution to the π phase-shift chromaticity of the phase mask coronagraphs. It will allow broadband observations, i.e., shorter integration times and object characterizations, by means of spectroscopic analysis. The feasibility of the component manufacturing is also considered through a tolerance study.

OCIS codes: 220.4830, 050.5080.

1. Introduction

Direct detection of faint sources around bright astrophysical objects such as stars or active galactic nuclei is difficult because of the large flux ratio between them. For example, β-Pictoris debris disk is 1000 times fainter than its host star in the visible spectrum. The study of such objects requires dedicated instruments called coronagraphs. Coronagraphy has been used in high dynamic imaging since its invention by Lyot in 1939.[1] The Lyot coronagraph principle is simple: it consists of an occulting mask at the telescope focus, which is centered on the bright object, and a well-dimensioned diaphragm, the so-called Lyot stop, in the relayed pupil to remove diffraction residuals. Lyot's design proved to be efficient for solar and stellar coronagraphy[2] and prolific, giving birth to an entire new family of improved byproducts, the so-called amplitude coronagraphs. Unfortunately, amplitude coronagraphs possess a major inherent weakness: The physical extension of the opaque zone occults quite a significant region centered on the optical axis and thus all the sources behind it. For example, let us mention the case of the recent notch-filter mask.[3] This design presents a large central opaque zone extending in the focal plane up to $3\lambda/D$ (D is the telescope diameter, whereas λ is the wavelength), where the potential companion is still attenuated by at least 50%.

New types of coronagraph have been designed as alternative solutions to the classical amplitude coronagraphs. This family of components acts on the phase instead of acting on the amplitude; they are, therefore, called phase coronagraphs.[4] The phase coronagraphs are not affected by the inherent central "dead zone" of the amplitude coronagraphs. Rouan *et al.*[5] proposed in 2000 a promising design, the so-called four-quadrant phase mask coronagraph (FQPM). The principle is to divide the focal plane into four equal areas centered on the optical axis, with two of them providing a π phase shift. This causes destructive interference ("nulling") to occur inside the geometric pupil area (Fig. 1). The nulling phenomenon's quality determines the coronagraph's performance and is quantified by the value of the so-called "nulling (or null) depth" coefficient over the considered spectral band.

The FQPM principle has been validated on the bench in monochromatic light[6] and installed on the NAOS-CONICA adaptive optics at the VLT (Very Large Telescope) UT4. It has given promising preliminary scientific results[7] and perspectives for future instruments such as the James Webb Space Telescope (JWST) (the infrared successor of the Hubble Space Telescope) and VLT–PF[8] (Planet-Finder, a proposal for the second generation instruments of the

D. Mawet (mawet@astro.ulg.ac.be), P. Riaud, and J. Surdej are with the Institut d'Astrophysique et de Géophysique, Université de Liège, 17 Allée du 6 Août, B-4000 Sart Tilman, Belgium. J. Baudrand is with the Observatoire de Paris-Meudon, 5 place Jules Janssen, F-92195 Meudon, France.

Received 28 February 2005; revised manuscript received 23 June 2005; accepted 24 June 2005.

0003-6935/05/347313-09$15.00/0

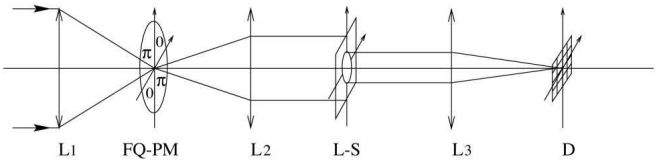

Fig. 1. Basic FQPM coronagraphic optical bench scheme: L1, L2, and L3 are three lenses in the optical system. L1 provides a large (to minimize spatial defects) F/D ratio on the FQPM; L2 images the pupil in the second plane. The Lyot stop (L-S) suppresses the diffracted starlight, and, finally, L3 forms the coronagraphic image on detector D.

Very Large Telescope). Manufacturing challenges of the FQPM come from the π phase shift achievement. So far only monochromatic masks have been realized using thin-film deposition or etching techniques on the "index step" principle: a step height h in a material of refractive index n at wavelength λ induces a phase shift:

$$\Delta\phi = \frac{2\pi}{\lambda}(n-1)h. \tag{1}$$

Unfortunately, this only works perfectly for a given wavelength. The hyperbolic phase-shift dependence in λ makes it inapplicable for high-performance use of such devices over a large spectral bandwidth . For instance, the nominal performance of the monochromatic mask used on NAOS-CONICA in the K band (2–2.4 μm) under well-corrected atmospheric turbulence is about 2000 times weaker than in monochromatic laboratory experiments.[6] In this paper we propose an original solution to achromatize and implement the π phase shift on the FQPM so it could be used over regular spectral bands corresponding to classical astrophysical filters allowing high nulling performances. Large band observations permit shorter integration times and object characterizations by spectroscopic analysis. Our solution is based on the dispersion of the so-called form birefringence of subwavelength surface-relief gratings.

Section 2 is devoted to the presentation of the zero-order gratings and to their very specific properties. In Section 3 we will describe the original implementation of the gratings into the final coronagraphic component. The theoretical design will be considered in detail in Section 4, followed by discussion of a tolerance study in Section 5 to demonstrate the manufacturing feasibility. Finally, Section 6 will be dedicated to the coronagraphic performance assessments.

2. Zero-Order Gratings

Zero-order gratings (ZOGs) consist of subwavelength gratings, i.e., the period of the structure is smaller than the wavelength of the incident light. They do not diffract light in the sense that only the zero transmitted and reflected orders are allowed to propagate outside the grating region, leaving incident wavefronts free from any further aberrations.

The condition under which a diffraction order propagates or not is determined by the well-known grating equation

$$\frac{\Lambda}{\lambda} = \frac{m}{n_I \sin\theta + n_{I,III}\sin\theta_m}, \tag{2}$$

where θ is the angle of incidence; θ_m is the exit angle of the mth order of diffraction; and n_I and n_{III} are the refractive indices of the incident (superstrate) and transmitting (substrate) media, respectively. By convention, the plus sign with the index n_I corresponds to the reflection, and the minus sign corresponds to the index n_{III} to the transmission. From this relation, we can immediately derive a ZOG condition on the grating period to the wavelength ratio

$$\frac{\Lambda}{\lambda} \le \frac{1}{n_I \sin\theta + \max(n_I, n_{III})}. \tag{3}$$

ZOGs present very specific properties: One-dimensionally modulated ZOGs artificially create unique anisotropic and dispersive characteristics that can be used to synthesize achromatic wave plates leading to a vectorial phase difference, i.e., between the vectorial polarization states TE (transverse electric: the electrical field is perpendicular to the plane of incidence, where the plane of incidence is defined by the grating normal and the direction of the incoming light) and TM (transverse magnetic: the electrical field lies in the plane of incidence),

$$\Delta\phi_{\mathrm{TE-TM}} = \frac{2\pi}{\lambda} h\Delta n_{\mathrm{TE-TM}}. \tag{4}$$

To simulate grating responses and calculate the so-called form birefringence $\Delta n_{\mathrm{form}} = \Delta n_{\mathrm{TE-TM}}$ in the subwavelength and resonant domains ($\Lambda \approx \lambda$), scalar theories of diffraction dramatically fail. The vectorial nature of light must be taken into account. Several approaches are possible depending on the grating period to wavelength ratio. In the quasi-static limit, i.e., when the ratio $\Lambda/\lambda \ll 1$, a simple homogenization treatment leads to the two following effective indices for the TE and TM states of polarization (see Born and Wolf[9]):

$$n_{\mathrm{eff},0}^{\mathrm{TE}} = \left[F n_a^2 + (1-F) n_b^2\right]^{1/2}, \tag{5}$$

$$n_{\mathrm{eff},0}^{\mathrm{TM}} = \left[\frac{n_a^2 n_b^2}{F n_b^2 + (1-F) n_a^2}\right]^{1/2}, \tag{6}$$

where n_a and n_b are the structure's real indices, and where F is the filling factor (ratio of the width of grating ridges to the period, a/Λ) (see Fig. 2). This straightforward approach is called the zero-order effective-medium theory (EMT0). When the ratio Λ/λ is no longer negligible, the latter closed-form expressions for the effective refractive indices are no longer correct. In such cases, the second-order effective-

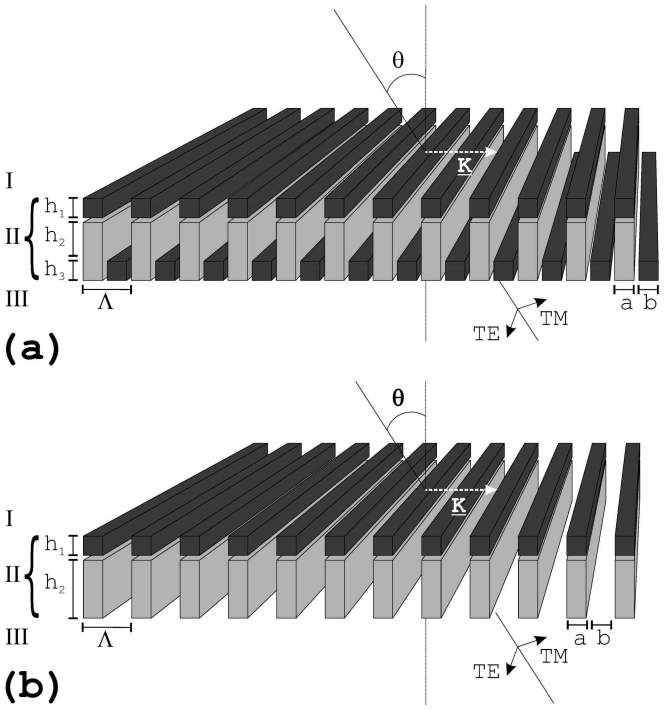

Fig. 2. ZOG schematic presenting the main parameters of the grating: the grating vector $|\underline{K}| = 2\pi/\Lambda$ perpendicular to the grating lines with Λ being the period, the grating depth h, and the filling factor $F = a/\Lambda$. (a) Two AR-layer design, where the AR-layer thickness is $h_1 = h_3$. and $h_2 + h_3 = h$. (b) One AR-layer design, where the AR-layer thickness is h_1 and $h_2 = h$. The grating (medium II) is surrounded by the media I (superstrate) and III (substrate). The incident light, making an angle θ with the grating normal, can be decomposed in its TE (transverse electric) and TM (transverse magnetic) states of polarization.

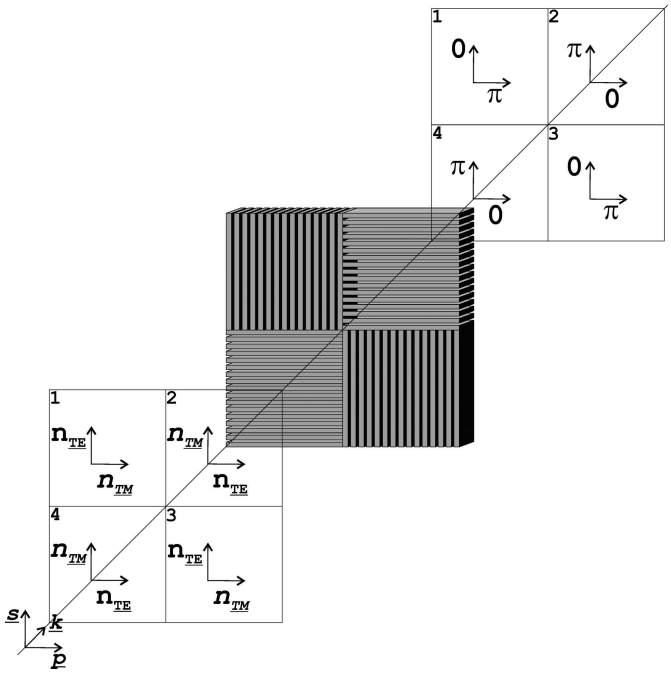

Fig. 3. 4QZOG implementation: $\underline{s}$ and $\underline{p}$ are the vectorial complex amplitude components of the incoming light of wave vector $\underline{k}$. In each of the four quadrants, the $\underline{s}$ and $\underline{p}$ global polarization states are decomposed in the corresponding $\underline{TE}_i$ and $\underline{TM}_i$ vectorial complex amplitudes according to the local grating line orientations (i is the quadrant number). Two effective indices n_{TE_i} and n_{TM_i} can be assigned to the corresponding perpendicular polarization states. The four gratings engraved on a unique substrate are strictly identical and implemented in the following way: two of them in two quadrants along one diagonal are rotated by 90° around their normals with respect to the two others. This antisymmetrical configuration achieves the FQPM particular focal plane π-phase distribution (see text for explanations).

medium theory (EMT2), which is deduced from the electromagnetic propagation in stratified media theory, allows us to derive these expressions for the effective indices[10]:

$$n_{\mathrm{eff},2}{}^{\mathrm{TE}} = \left[(n_{\mathrm{eff},0}{}^{\mathrm{TE}})^2 + \frac{1}{3}\left(\frac{\Lambda}{\lambda}\right)^2 \pi^2 F^2 (1-F)^2 \times (n_a{}^2 - n_b{}^2)^2 \right]^{1/2}, \quad (7)$$

$$n_{\mathrm{eff},2}{}^{\mathrm{TM}} = \left[(n_{\mathrm{eff},0}{}^{\mathrm{TM}})^2 + \frac{1}{3}\left(\frac{\Lambda}{\lambda}\right)^2 \pi^2 F^2 (1-F)^2 \times \left(\frac{1}{n_a{}^2} - \frac{1}{n_b{}^2}\right)^2 (n_{\mathrm{eff},0}{}^{\mathrm{TM}})^6 (n_{\mathrm{eff},0}{}^{\mathrm{TE}})^2 \right]^{1/2}. \quad (8)$$

In addition to the dependence on the wavelength, we also note the dependence of the effective indices versus other parameters available in a design procedure: the grating period Λ, the filling factor F, and the grating real indices n_a and n_b (see Fig. 2). The wavelength dependence of the effective indices is consequently also found in the form birefringence Δn_{form}. This phenomenon is called the dispersion of form birefringence. We must emphasize the term "form." Indeed, this property appearing just before the frontier of the resonant domain is essentially given by the geometry, no longer only by the intrinsic characteristics of the materials. The key point is that, by carefully controlling the geometry of the grating structure, we will be able to tune the form birefringence to compensate for the hyperbolic dependence of the phase shift [$\propto 1/\lambda$, see Eq. (4)] and thus make it achromatic. This is a good example of refractive-index engineering.

3. FQPM Implementation: the Four-Quadrant ZOG (4QZOG)

The implementation of the ZOG phase shifters into the FQPM is straightforward and represented in Fig. 3. Let $\underline{s}$ and $\underline{p}$ be the vectorial complex amplitude components of the incoming light. In each of the four quadrants, the $\underline{s}$ and $\underline{p}$ global polarization states can be decomposed in the $\underline{TE}_i$ and $\underline{TM}_i$ vectorial complex amplitudes according to the line orientations of the local grating in the ith quadrant, with $i = 1, \ldots, 4$. Indeed, the convention under normal incidence specifies that the electric field of the TE and TM components vibrates parallel and perpendicular to the grating lines, respectively. We have shown above that two effective indices, n_{TE_i} and n_{TM_i}, can be assigned to the corresponding perpendicular polarization states.

Let us assume now that (i) the four gratings engraved on a unique substrate are strictly identical and (ii) they are implemented in the following way: two of them in two quadrants along one diagonal are rotated by 90° around their normal with respect to the two others.

The consequence of the first hypothesis is that the effective indices n_{TE_i} and n_{TM_i} are identical in each quadrant, i.e., $n_{\mathrm{TE}_i} = n_{\mathrm{TE}}$ and $n_{\mathrm{TM}_i} = n_{\mathrm{TM}}$, with $i = 1, \ldots, 4$. The second one gives this particular antisymmetrical configuration (Fig. 3) with the following result. If, by construction,

$$\triangle\phi_{\mathrm{TE}_i\text{–}\mathrm{TM}_i} = \frac{2\pi}{\lambda} h\Delta n_{\mathrm{TE}_i\text{–}\mathrm{TM}_i}$$

$$= \frac{2\pi}{\lambda} h\Delta n_{\mathrm{TE}\text{–}\mathrm{TM}} \approx \pi,$$

then the potentially interfering parallel polarization states along the $\underline{s}$ component direction, for example, are phase shifted in the following way (j being an adjacent quadrant of i):

$$\Delta\phi_{\mathrm{TE}_i\text{–}\mathrm{TM}_j} = \frac{2\pi}{\lambda} h(n_{\mathrm{TE}_i} - n_{\mathrm{TM}_j})$$

$$= \frac{2\pi}{\lambda} h(n_{\mathrm{TE}_i} - n_{\mathrm{TM}_i})$$

$$= \Delta\phi_{\mathrm{TE}_i\text{–}\mathrm{TM}_i} \approx \pi.$$

The same goes for the $\underline{p}$ component direction. Consequently, for the $\underline{s}$ and $\underline{p}$ component directions, respectively, the FQPM particular focal plane π-phase distribution is achieved, i.e., adjacent quadrants are in phase opposition. Therefore, it works in natural light. We will call this new structure the four-quadrant ZOG (4QZOG). It must be noted that the precision on the perpendicularity between the gratings is directly related to the nulling performance. It has to be finely controlled.

4. Design of the 4QZOG

Effective-medium theories are qualitatively interesting to understand the underlying physical phenomena, but only deal with zero-order diffracted waves and disregard the evanescent higher-order ones. To obtain the exact response of this kind of structure, they have to be taken into account together with external media influences in terms of multiple-beam interferences, just as in the so-called rigorous coupled wave analyses (RCWA).[11] Moreover, this theory gives, in addition to the phase, the intensity of the diffracted waves (diffraction efficiencies).

A. Preliminary Considerations

The design procedure consists of the optimization of the π phase-shift quality within a given spectral range for a well-chosen material. It also requires the equalization of the interfering fluxes, which is difficult to achieve. Indeed, the difference between the effective indices that induces the form birefringence is also responsible for the resultant unequal TE/TM–Fresnel reflection coefficients at the interfaces between the modulated region and the external media. The merit function to be minimized is the null depth $N(\lambda)$. The latter is directly related to the phase-shift error with respect to π, $\varepsilon(\lambda) = \Delta\phi_{\mathrm{TE}\text{–}\mathrm{TM}}(\lambda) - \pi$, and the flux ratio $q(\lambda) = n_{\mathrm{TE}}(\lambda)/n_{\mathrm{TM}}(\lambda)$:

$$N(\lambda) = \frac{\left[1 - \sqrt{q(\lambda)}\,\right]^2 + \varepsilon(\lambda)^2\sqrt{q(\lambda)}}{\left[1 + \sqrt{q(\lambda)}\,\right]^2}. \qquad (9)$$

We have investigated several ways to ensure a proper flux balance, i.e., $q \approx 1$. The principle is always the same and consists of trying to minimize the index discontinuities (impedance matching) at the interfaces by the following:

(i) Varying the grating profile: The structure can have trapezoidal profiles instead of rectangular ones. It reduces the parasitic reflection but keeps the anisotropy and a sufficient form birefringence dispersion up to a certain point.

(ii) "Sandwiching" the grating: The superstrate material can be chosen to avoid the brutal air–ZOG interface.[12] It is conceptually interesting and gives good numerical results, but presents extreme manufacturing challenges, such as how to keep the grating openings clear.

(iii) Depositing a classical antireflective (AR) layer: The ZOG is AR treated with a single λ/4 layer. The latter settles in the bottom of the openings and on the top of the grating ridges [Fig. 2(a)]. This leads to an original configuration, a three-layer grating with a complex behavior governed by diffraction. We demonstrate that this particular structure can be optimized to give interesting results.

Currently, subwavelength gratings are manufactured more readily for the near-infrared and mid-infrared domains because there are still difficulties in imprinting submicrometer features. Hopefully, the near-term potential applications (VLT–PF, for example) of this technology concern this wavelength domain. We will focus on the astrophysical band filters H (1.475–1.825 μm), K (2–2.4 μm), and N (9.3–11.6 μm). We have to choose among the restricted list of infrared materials, but the ZOG technology is sufficiently flexible to accommodate the majority of them. We will present results for diamond, zinc selenide (ZnSe), cadmium telluride (CdTe), silicon (Si), and germanium (Ge). This choice is justified because these selected materials are common in infrared applications, they cover a large refractive-index spectrum (2–4), and their etching processes are well known.

B. 4QZOG Theoretical Analysis

The indices for diamond,[13] ZnSe,[14] CdTe,[15] Si,[15] and Ge (Ref. 15) will be taken from the following repre-

Table 1. Coefficients for Material Refractive-Index Representations

Coefficient	Diamond	ZnSe	CdTe	Si	Ge
A	1	1	$-2.373 \times 10^{-4}T + 3.8466$	$1.600 \times 10^{-4}T + 3.431$	$-6.040 \times 10^{-3}T + 11.051\,28$
B	0.3306	4.463 95	$8.057 \times 10^{-4}T + 3.2215$	-2.643×10^{-2}	$9.295 \times 10^{-3}T + 4.005\,36$
C	30625	0.0404	$-1.10 \times 10^{-4}T + 0.1866$	4.324×10^{-3}	$-5.392 \times 10^{-4}T + 0.599\,034$
D	4.3356	0.461 32	$-2.160 \times 10^{-2}T + 12.718$	-3.194×10^{-4}	$4.151 \times 10^{-4}T + 0.091\,45$
E	11 236	0.1538	$-3.160 \times 10^{1}T + 18753$	8.835×10^{-6}	$1.514\,08T + 3426.5$
F	—	2.882 89	—	—	—
G	—	2213.5	—	—	—

sentations:

$$n_{\text{diamond,ZnSe,CdTe,Ge}}(\lambda) = \left(A + \frac{B\lambda^2}{\lambda^2 - C} + \frac{D\lambda^2}{\lambda^2 - E} + \frac{F\lambda^2}{\lambda^2 - G}\right)^{1/2}, \quad (10)$$

$$n_{\text{Si}}(\lambda) = A + B\lambda + C\lambda^2 + D\lambda^3 + E\lambda^4. \quad (11)$$

The corresponding coefficients are given in Table 1. We assume an ambient temperature (T = 298 K), keeping in mind that corrections are needed if T is different. Note that λ is expressed in micrometers in all representations except for diamond, where it is expressed in nanometers.

We have performed simulations for the *H, K,* and *N* bands, assuming in each case a ZOG AR treated with a YF_3 layer.[16] Absorption and dispersion of YF_3 have been taken into account (see Fig. 13 in Lemarquis *et al.*[17]). We chose YF_3 as a coating material, but several other possibilities exist such as BaF_2 and LaF_2, which could be better choices according to the substrate index. The grating structure in all these simulations consists of a single rectangular profile discretized into three levels to take the deposited AR layer into account [Fig. 2(a)]. The results are always discussed in terms of the null depth [Eq. (9)]. Despite the flexibility in the design of the 4QZOG, there will be a compromise on the substrate refractive properties. Indeed, the dispersion of the form birefringence is proportional to the raw index n. Practically, n must be greater than approximately 2 to sufficiently compensate for the wavelength hyperbolic dependence of the phase shift [Eq. (4)] and thus achromatize it. Unfortunately, the intensity of the parasitic reflection phenomenon will also increase accordingly. We have already discussed the flux imbalance problem, but the Fresnel reflection will also be responsible for a so-called second-order ghost in the final coronagraphic image. This ghost comes from the double reflection on the back face of the substrate and on the internal interface of the AR-treated ZOG. Its intensity can be quantified assuming a reasonable AR performance for the back side of the component, i.e., about 0.5% of reflectivity.

We can see in Fig. 4 that the ghost intensity level significantly increases with the substrate refractive index. We also note that the rough null depth performance is inversely proportional to the material index. Indeed, our AR-layer solution for the differential flux problem shows here its limitations. For this reason, and for the sake of simplicity, we will restrain our analysis to ZnSe and diamond, both possessing the lowest-index values (near 2.4).

Results for both selected materials in terms of null depth are excellent either for the *H, K,* or *N* bands (Figs. 5, 6, and 13). Indeed, the mean null depth performances are at the 10^{-5} level over the whole spectral range of the three considered band filters. Without losing generality, we shall restrict further analysis to the particular case of the *K* band. We have explicated the performances in terms of phase-shift quality and differential transmittances between the TE and the TM states of polarization (Figs. 7 and 8). The standard deviation of the phase around π is only 7×10^{-3} rad rms, which is good and indicates a small residual chromaticity. As far as the transmittances through the ZOG are concerned (zero-order diffrac-

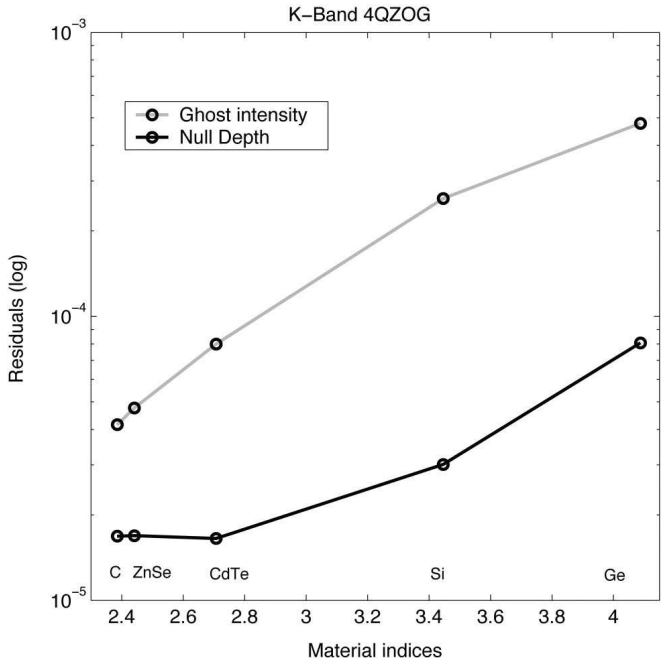

Fig. 4. Design level of performance according to the substrate refractive index n. The ghost intensity level increases significantly with n. We also note that the rough null depth performance is inversely proportional to n. The best compromise is therefore for the low-index values.

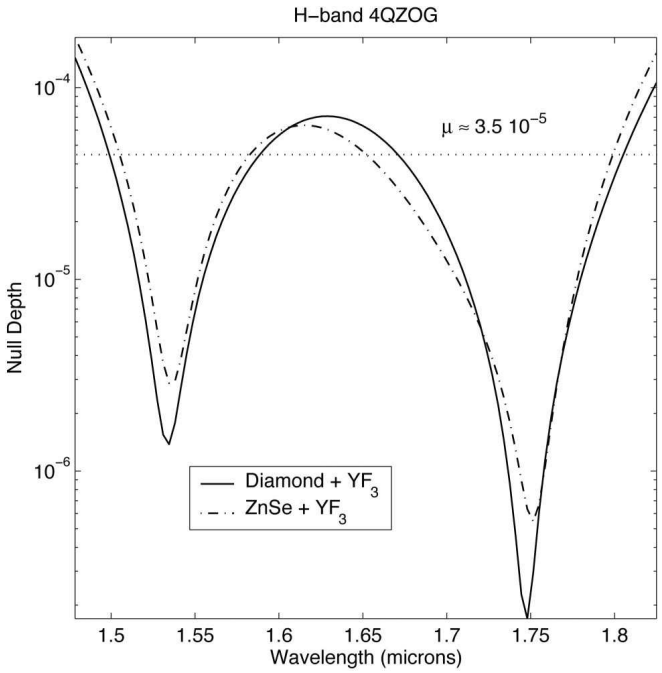

Fig. 5. *H*-band 4QZOG null depth (logarithmic scale) vs wavelength. The continuous curve is for the diamond YF_3 AR-coated 4QZOG, the dashed curve for the ZnSe YF_3 AR-coated one. $\mu \approx 3.5 \times 10^{-5}$ is the mean null depth over the whole *H* band.

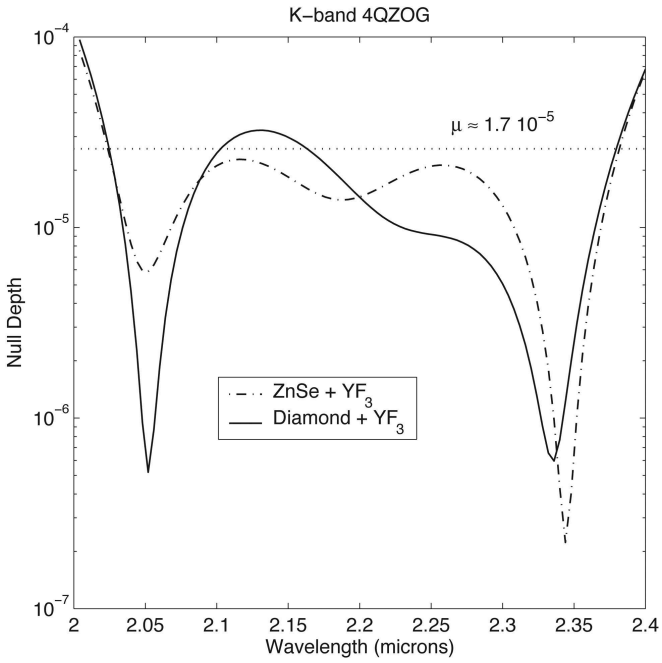

Fig. 6. *K*-band 4QZOG null depth (logarithmic scale) vs wavelength. The continuous curve is for the diamond YF_3 AR-coated 4QZOG, the dashed curve for the ZnSe YF_3 AR-coated one. $\mu \approx 1.7 \times 10^{-5}$ is the mean null depth over the whole *K* band.

tion efficiencies in transmission for the TE and TM components), we note an important absorption due to the YF_3 layer. This energy leakage affects the optical throughput of the system but favorably reduces the ghost level by absorbing a fraction of the internal higher-order parasitic reflections.

The grating parameters of the *K*-band optimized solutions for the ZnSe and diamond materials are given in Table 2. All the fabrication aspects will be considered in Section 5. These optimal parameter sets have been calculated by implementing the simplex search method[17] with our RCWA algorithm. One

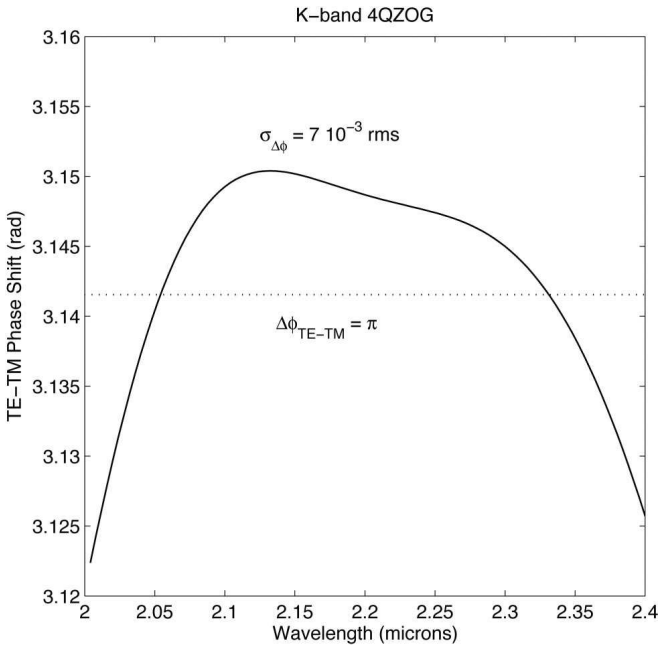

Fig. 7. *K*-band 4QZOG phase shift between the TE and TM polarization states vs wavelength: the phase-shift quality is quantified by the phase-shift standard deviation $\sigma_{\Delta\phi}$ (the less, the better), which is at the very good level of 7×10^{-3} rad rms.

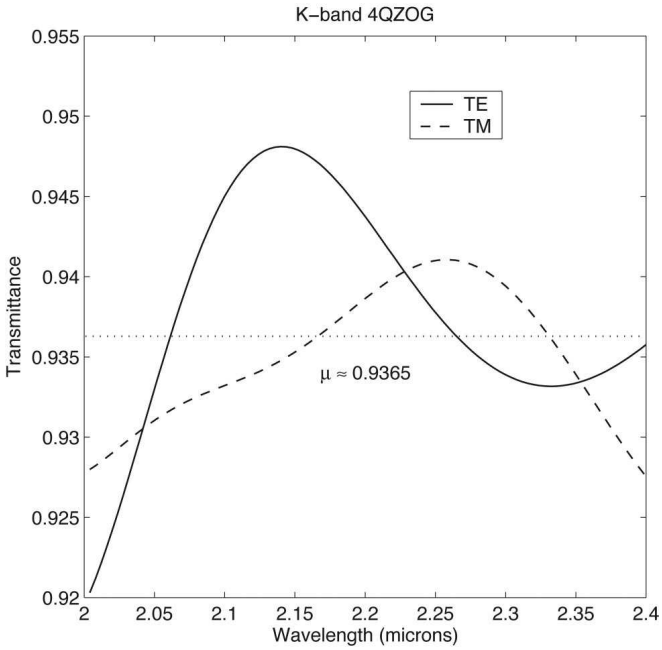

Fig. 8. *K*-band 4QZOG transmittances (zero-order diffraction efficiencies in transmission for the TE and TM components) vs wavelength, including the mean optical throughput of the component, which is ≈93.6%, taking absorption into account. The amplitude peak-to-valley variations for the two polarizations TE and TM over the whole band range from 1.5% to 3%.

Table 2. K-Band Diamond and ZnSe 4QZOG Parameters (Two AR-Layer Design)

Parameters	Diamond	ZnSe
Grating period Λ	0.740 μm	0.708 μm
Grating depth $h\ (= h_2 + h_3)$	3.238 μm	3.116 μm
Grating filling factor F $(= a/\Lambda)$	70%	72%
YF_3 AR-layer thickness $h_{AR}\ (= h_1 = h_3)$	423 nm	404 nm

can see that only minor corrections are needed between the ZnSe and the diamond designs; though their dispersions are quite different, their indices are very close. Parameters for other bands may be derived by simple scaling operations, with little corrections to take the index dispersion and/or absorption of the substrate and AR layer again into account.

In conclusion, the results we obtained, taking phase defects and amplitude inequalities into account, are very good with deep ($\approx 10^{-5}$) average nulls over the considered wavelength bands. This level of performance is sufficient in the case of phase coronagraphy, as we will see in Section 6.

5. Tolerancing and Manufacturing

The filling factor (F) is the determinant parameter of the grating structure and the most difficult to control during the fabrication process. Its influence can be analyzed by varying it and finding the best solutions in terms of the free parameters that remain; the grating period, depth, and the AR-layer thickness. Figure 9 shows the optimal configuration null depth as a function of the grating filling factor F. This calculation clearly points out that the best solutions occur for the largest F with some local minima. The corresponding parameter adjustments are given in Fig. 10, where, for each F, the optimal period and depth have been recomputed. We logically note that the updated depth increases with F. Indeed, the form birefringence that can be considered as an optical anisotropy is inversely proportional to F, whereas its wavelength dispersion increases accordingly. Consequently, a given phase shift will be obtained for a deeper grating. These considerations illustrate the design's strong sensitivity to the parameters.

To assess the actual feasibility of the component fabrication, we have performed several tolerance simulations. By varying each parameter independently around an optimal solution, the specifications would impose an unrealistic tight control of the profile and depth of the grating (at the nanometer level). We have adopted a more physical approach that requires the scanning of the parameter space searching affordable solutions in terms of a well-chosen merit figure. In our application, the latter is the null depth as defined in Eq. (9) and integrated over the whole considered wavelength band. These time-consuming RCWA calculations have led to interesting results (Fig. 11), demonstrating that the manufacturing is

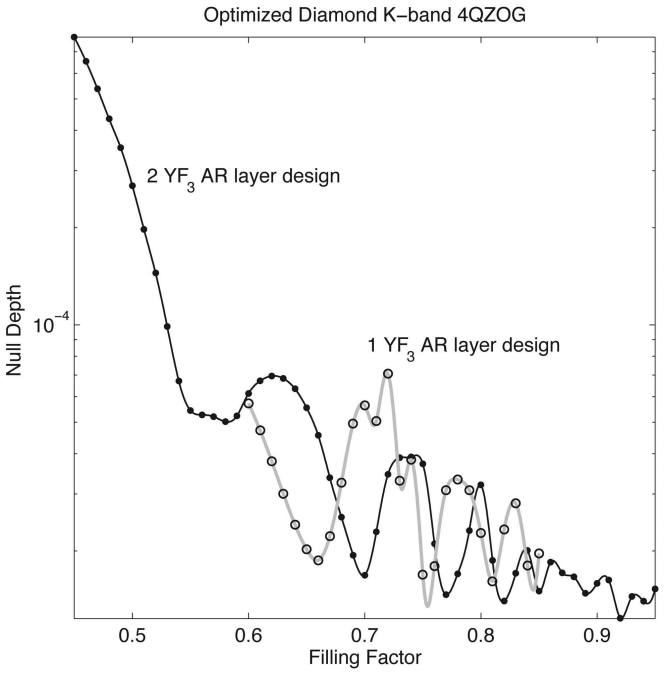

Fig. 9. Null depth (logarithmic scale) versus the grating filling factor (F). An optimal solution has been calculated at each F value. We note that best null depths occur for the largest F values with some local minima. We have overplotted in gray the one AR-layer case in which we notice a tiny average degradation of the null depth (negligible in our application) but a nonnegligible shift in optimal filling factors for the manufacturing.

within the reach of current microelectronic technology. The fabrication implies the traditional lithography step defining the profile mask either with direct laser writing or e-beam lithography and then the transfer of the pattern into the substrate by plasma

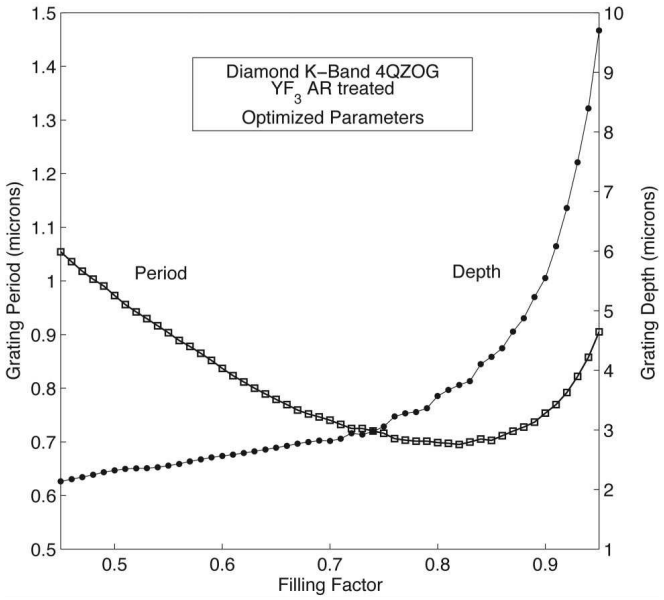

Fig. 10. Double plot of the grating depth and period vs the filling factor (F). The calculations show the optimal recomputed grating period (left) and depth (right) for the best null depth at each F. We note that the depth increases with F. The period shows a parabolic behavior.

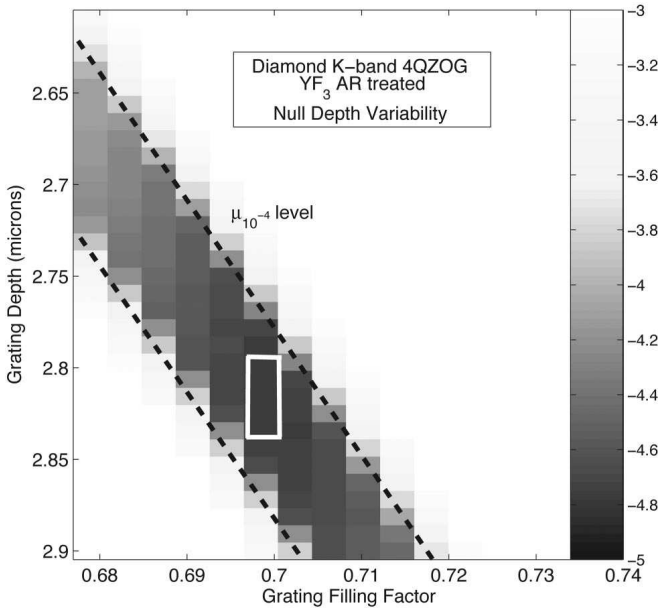

Fig. 11. 4QZOG null depth sensitivity (logarithmic scale) to the grating depth h and filling factor F parameter variations. Each point in this figure represents the average null depth over the whole K band for a given set of h and F but retaining the best solution among several AR-layer thicknesses (with a 2% tolerance) to accommodate this *a posteriori* possibility of correction.

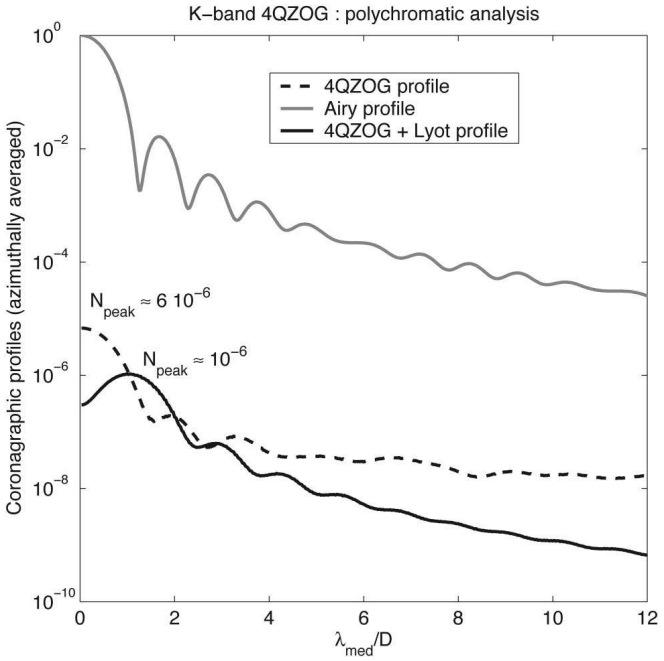

Fig. 12. K-band 4QZOG coronagraphic profile (logarithmic scale) vs the angular separation in λ_{med}/D unity, D being the telescope diameter. The gray curve corresponds to the polychromatic Airy disk. The dashed curve represents the wide band 4QZOG azimuthally averaged profile where the peak-to-peak null depth is about 6×10^{-6}. The continuous curve corresponds to the previous case with a tiny circular Lyot opaque mask ($0.55\ \lambda_{med}/D$ of diameter) to get rid of the chromatic residuals and ghost image. The peak-to-peak null depth is approximately 10^{-6}.

etching. These techniques are more adapted to small filling factors. This is the reason why we have always privileged the lowest ones in our optimal results, i.e., around $F \approx 70\%$ (see Fig. 11). It is to be noted that such structures with similar aspect ratios (the aspect ratio is the ratio between the grating depth and the width of the grating ridges) have already been successfully produced in silicon[18] using these methods.

Assuming a classical realistic resolution of 10 nm in thickness (2%) for an *a posteriori* correction with the AR layer, we can ensure a grating etching depth tolerance of about 100 nm at the null depth level of 10^{-4} (affordable solution). This result also assumes a perfect *a priori* metrology of the photolithographic mask using either the scanning electron microscope or the atomic force microscope. These conclusions have led us to seriously consider the fabrication of the component that is under study with the collaboration of "Adamantis AB" at the Angström Laboratory (University of Uppsala, Sweden). Indeed, they have developed a very robust and well-proven process for diamond microstructuring.[19,20] The preliminary manufacturing assessments have already led us to consider the possibility of depositing the AR layer on the bare substrate before etching. In such a case, there will only be one layer settling at the top of the grating ridges [see Fig. 2(b)]. Corresponding numerical simulations have only tackled a tiny average degradation of the final performances but negligible in our application (Fig. 9). The reason is that the greatest contribution to the parasitic Fresnel reflections comes from the first interface (air–ZOG) where the index discontinuities are the largest. It should be noted that in the one AR-layer design, there is a nonnegligible correction of the other grating parameters with respect to the two-layer one. In conclusion, both designs work and the choice belongs to the manufacturer.

6. Coronagraphic Simulations

We have performed polychromatic coronagraphic simulations for the K-band 4QZOG. These calculations try to simulate the broadband response of the component on a perfect optical bench. It thus only tackles chromatic effects. We have linked the output results of our RCWA code in terms of phase and amplitude transmittances to a traditional Fourier propagation coronagraphic simulator.[21] The results are excellent (Fig. 12) and are in good agreement with the corresponding null depth curve (Fig. 6).

The polychromatic null depth on the Airy peak is about 6×10^{-6}. The 10^{-7} speckle level is very quickly reached at $2\lambda_{med}/D$ ($\lambda_{med} = 2.2\ \mu m$), where D is the telescope diameter. Better performances (10^{-6} on the peak) can be achieved by placing at the center of the component, a tiny[4] circular Lyot opaque mask of $0.55\lambda_{med}/D$, which is the diameter encircling half of the Airy disk energy. By doing this, we remove a large part of the 4QZOG chromatic residuals and also prevent the annoying ghost problem to a certain point. These performances are comfortably 1 order of magnitude above the H-band/K-band VLT–PF specification. Indeed, the main limitation for ground-based observations comes from the atmospheric turbulence

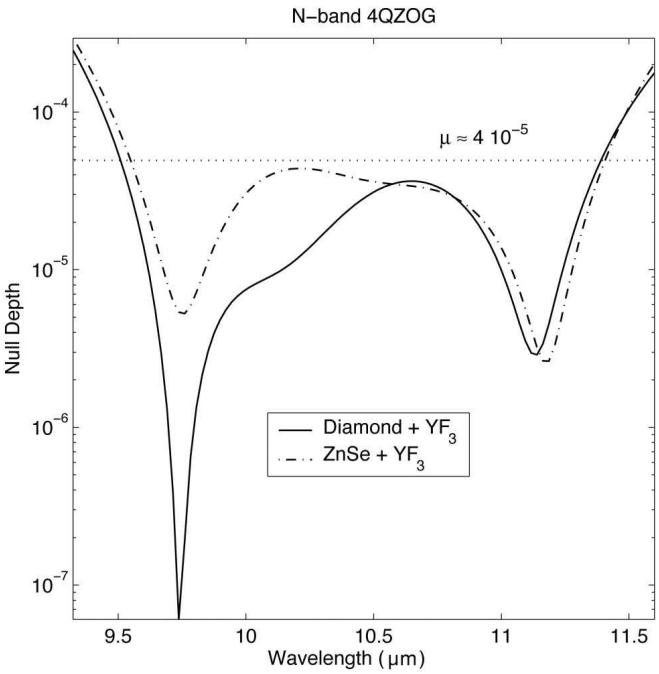

Fig. 13. *N*-band 4QZOG null depth (logarithmic scale) versus wavelength. The continuous curve is for the diamond YF_3 AR-coated 4QZOG, the dashed curve for the ZnSe YF_3 AR-coated one. $\mu \approx 4 \times 10^{-5}$ is the mean null depth over the whole *K* band.

residuals (the so-called speckles) after adaptive optics corrections of the incoming fluctuating wavefronts. For this independent reason, the baseline dynamic will be limited to a maximum null depth of 10^{-4}.

Another potential application of the 4QZOG could be an infrared interferometer in the *N* band for NASA's Terrestrial Planet Finder and ESA's Infra-Red Space Interferometer–DARWIN missions; the main limitation of the nulling interferometer will be the leakages from the partially resolved stellar diameter. Therefore, the performance obtained in our theoretical simulations (Fig. 13) would again be sufficient.

7. Conclusion

We have presented an original solution to the chromaticity problem of the π phase shift for the FQPM coronagraph by using subwavelength gratings (ZOGs). Indeed, the birefringence of ZOGs can be controlled by the geometry to compensate for the hyperbolic dependence of the phase shift and intrinsic material dispersions. We call this new design the "4QZOG." The results (null depth $\approx 10^{-5}$) are within ground-based instrument specifications for exoplanet detection (VLT–PF) and could also be envisaged for future ambitious space-based missions (TPF/DARWIN). We have also demonstrated the feasibility of the component fabrication with realistic tolerances. Its manufacturing is under assessment by "Adamantis AB" (Angström Laboratory, Uppsala University, Sweden). It will be the subject of a forthcoming paper.

D. Mawet acknowledges the financial support of the Belgian "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture." P. Riaud and J. Surdej acknowledge the financial support of the "Pôle d'Attraction Inter-Universitaire."

References

1. B. Lyot, "A study of the solar corona and prominences without eclipses," Mon. Not. R. Astron. Soc. **99,** 580–594 (1939).
2. B. A. Smith and R. Terrile, "A circumstellar disk around Beta Pictoris," Science **226,** 1421–1424 (1984).
3. M. J. Kuchner and D. M. Spergel, "Notch-filter masks: practical image masks for planet-finding coronagraphs," Astrophys. J. **594,** 617–626 (2003).
4. F. Roddier and Cl. Roddier, "Stellar coronagraph with phase mask," Publ. Astron. Soc. Pac. **109,** 815–820 (1997).
5. D. Rouan, P. Riaud, A. Boccaletti, Y. Clénet, and A. Labeyrie, "The four-quadrant phase-mask coronagraph. I. Principle," Publ. Astron. Soc. Pac. **112,** 1479–1486 (2000).
6. P. Riaud, J. Baudrand, A. Boccaletti, and D. Rouan, "The four-quadrant phase-mask coronagraph. III. Laboratory performance," Publ. Astron. Soc. Pac. **115,** 712–719 (2003).
7. D. Gratadour, D. Rouan, A. Boccaletti, P. Riaud, and Y. Clénet, "Four quadrant phase mask K-band coronagraphy of NGC 1068 with NAOS-CONICA at VLT," Astron. Astrophys. **429,** 433–437 (2005).
8. D. Mouillet, T. Fusco, A.-M. Lagrange, and J.-L. Beuzit, "Planet Finder on the VLT: context, goals, and critical specifications for adaptive optics," EAS Publ. Ser. **8,** 193–200 (2003).
9. M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, 1999), Chap. 15, pp. 837–840.
10. S. M. Rytov, "Electromagnetic properties of a finely stratified medium," Sov. Phys. JETP **2,** 466–475 (1956).
11. M. G. Moharam and T. K. Gaylord, "Rigourous coupled-wave analysis of planar grating diffraction," J. Opt. Soc. Am. **71,** 811–818 (1981).
12. H. Kikuta, Y. Ohira, and K. Iwata, "Achromatic quarter-wave plates using the dispersion of form birefringence," Appl. Opt. **36,** 1566–1572 (1997).
13. F. P. Bundy, "Melting point of graphite at high pressure: heat of fusion," Science **137,** 1055–1057 (1962).
14. W. J. Tropf, "Temperature-dependent refractive index models, for BaF_2, CaF_2, MgF_2, SrF_2, LiF, NaF, KCl, ZnS, and ZnSe", Opt. Eng. **34,** 1369–1373 (1995).
15. G. J. Hawkins, "Spectral Characterisation of Infrared Optical Materials and Filters," Ph.D. dissertation (University of Reading, United Kingdom, 1998).
16. F. Lemarquis, G. Marchand, and C. Amra, "Design and manufacture of low-absorption ZnS–YF_3 antireflection coatings in the 3.5–16 μm spectral range," Appl. Opt. **37,** 4239–4244 (1998).
17. J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," SIAM J Optim **9,** 112–147 (1998).
18. G. P. Nordin and P. C. Deguzman, "Broadband form birefringent quarter-wave plate for the mid-infrared wavelength region," Opt. Express **5,** 163–168 (1999).
19. M. Karlsson and Fr. Nikolajeff, "Diamond micro-optics: microlenses and antireflection structured surfaces for the infrared spectral region," Opt. Express **11,** 502–507 (2003).
20. M. Karlsson, K. Hjort, and Fr. Nikolajeff, "Transfer of continuous-relief diffractive structures into diamond by use of inductively coupled plasma dry etching," Opt. Lett. **26,** 1752–1754 (2001).
21. P. Riaud, A. Boccaletti, D Rouan, F. Lemarquis, and A. Labeyrie, "The four-quadrant phase-mask coronagraph. II. Simulations," Publ. Astron. Soc. Pac. **113,** 1145–1154 (2001).

5.3 Diamond FQPM and 4QZOG

It soon appeared that diamond was a privileged material for near-infrared coronagraphy because of its exceptional properties. Indeed, as already mentioned in Sect. 4.2.5, diamond is not only the hardest material on Earth, it also exhibits the highest thermal conductivity of all solids, shows a record wide optical transmission window (from 200 nm to ~ 1 mm, with some absorption features between 3 and 6 μm, see Annex B), is inert to most chemicals and extremely resistant. Whereas those unique properties have been known for a long time there have been mainly two hurdles limiting the usage of diamond in various applications: the lack of an industrially viable process for producing synthetic diamond, and the extreme difficulty in machining diamond to desired surface structures. In recent years there has however been considerable progress in the development of chemical vapour deposition (CVD) processes for producing synthetic diamond, resulting in high-quality diamond that can now be bought from several vendors.

5.3.1 Diamond FQPM

In the framework of our fruitful collaboration with ADAMANTIS AB (see Sect. 4.2.5), they were asked by the LESIA to manufacture prototypes of FQPM in diamond for the coronagraphs of MIRI (the mid-infrared camera and spectrometer of the JWST, see Sect. 1.5.3). The prototypes had to be monochromatic, i.e., based on the index-step principle, and designed to operate at 4.77 and 15.5 μm. The step specifications were therefore $h_{4.77} = 4.77/2(n_{diamond}(4.77) - 1) = 1.7271\,\mu$m for the smallest wavelength and $h_{15.5} = 15.5/2(n_{diamond}(15.5) - 1) = 5.6142\,\mu$m for the largest one. The manufacturing was done by standard lithography and etching techniques adapted to diamond micro-structuring (see Sect. 4.2.5). The final result was really convincing with very steep transitions between adjacent quadrants. The WYKO profilometer measurements gave a few tens of nanometers of precision for the heights given above, leading to slopes approaching the ninety degrees without any difficulty, which is one of the particularities of the used ICP-etching technique.

The two diamond FQPM were manufactured for two distinct wavelengths, i.e., with a given step, which had to be accurately controlled for qualification. For that purpose, a low-resolution spectroscopic device to disperse the coronagraphic image and hence derive the operating wavelength was implemented on the LESIA optical bench. The principle is to use the coherence length of the operating laser to measure the wavelengths of different orders k of the null, given that the phase shift $\phi = (2k + 1)\pi$, and according to the following relation

$$(2k + 1)\pi = \frac{2\pi}{\lambda_k}(n(\lambda, T) - 1)h \tag{5.1}$$

A transmission grating (110 lines/mm) with about 25% of transmission in the first order was used to obtain a pixel sampling of about 0.26 nm on the CCD between 500 and 800 nm. The wavelength calibration was performed on the $He - Ne$ laser line at 632.8 nm and the sodium yellow lines at 589 and 589.6 nm. A coronagraphic image was first acquired with a white halogen lamp at 3400 K with a long exposure time (30-120 s). Then a direct image was recorded without the coronagraphic mask (off-axis image) and a shorter exposure time (1-5 s). These two spectra were then divided to derive the wavelength dependence of the rejection factor. By applying this method for the control of the step height accuracy in the two diamond components, very precise values could be measured, and were only a few nanometers out of the specifications (see Fig. 5.2).

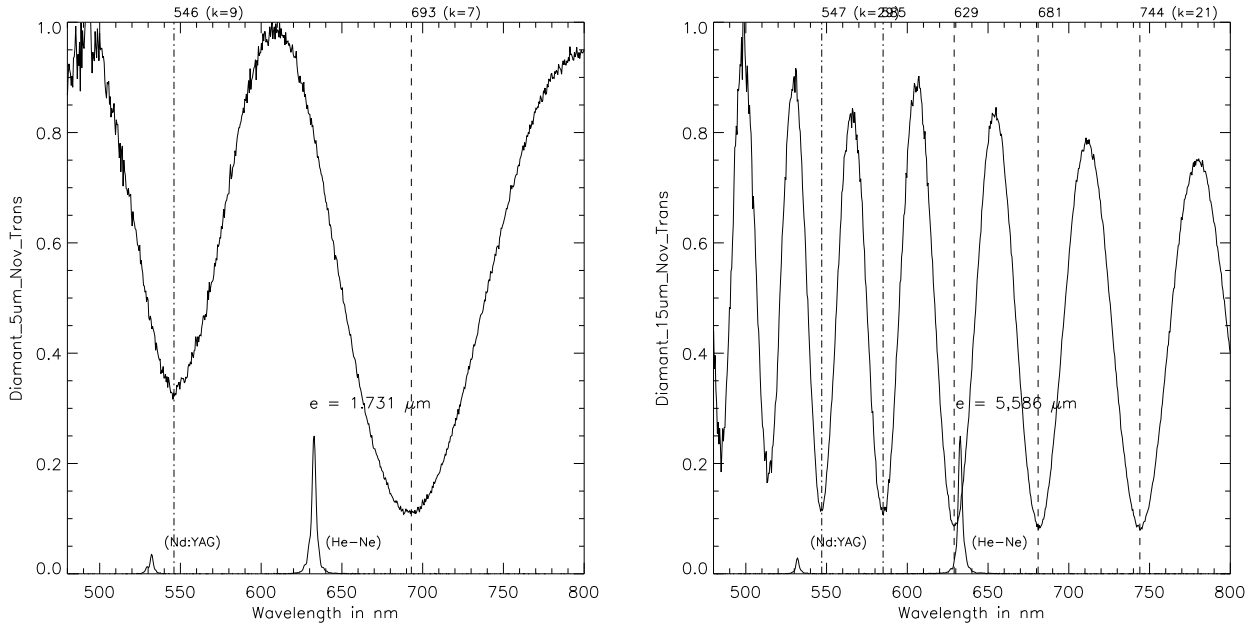

Figure 5.2: Low-resolution visible spectroscopy of diamond IR masks. Left: 4.77 μm mask with a step measured at 1.731 μm. Right: 15.5 μm mask with a step measured at 5.586 μm. Courtesy of Jacques Baudrand (LESIA).

The first component, optimized for 4.77 μm was then measured in the nulling mode at its operating wavelength on the LESIA cryogenic testbed (Baudoz et al. 2004, 2006). For that, a three-layer $YF_3/ZnSe/YF_3$ anti-reflection coating was first deposited on the masks by the Fresnel institute[17]. The measurements were again very convincing with stellar peak attenuations of ~ 400 while the measured contrast at $2\lambda/D$ was as good as 10^{-4}. Still, this performance was limited by the spectral resolution ($R_\lambda = 10$) and the angular size of the artificial source, not perfectly point-like.

5.3.2 Diamond 4QZOG

This successful demonstration of the micro-structured diamond reliability for astrophysical purposes has led us to envisage the manufacturing of a 4QZOG in diamond according to our calculated design for the K band, and using the same advanced lithographyic techniques available and apparently well mastered by ADAMANTIS AB. The specification were taken from Mawet et al. (2005c) and are summarized in Fig. 5.3.

The first trials of e-beam mask writing in PMMA were encouraging and after some iterations finally approached the feature line specifications (Fig. 5.4). Unfortunately, for some very sad reasons, this operation was not pursued further. We would like to warmly thank Mikael Karlsson of ADAMANTIS for his fantastic work on diamond components. We are looking forward to working again together.

[17] Laboratory "Unité mixte de recherche" based in Marseille.

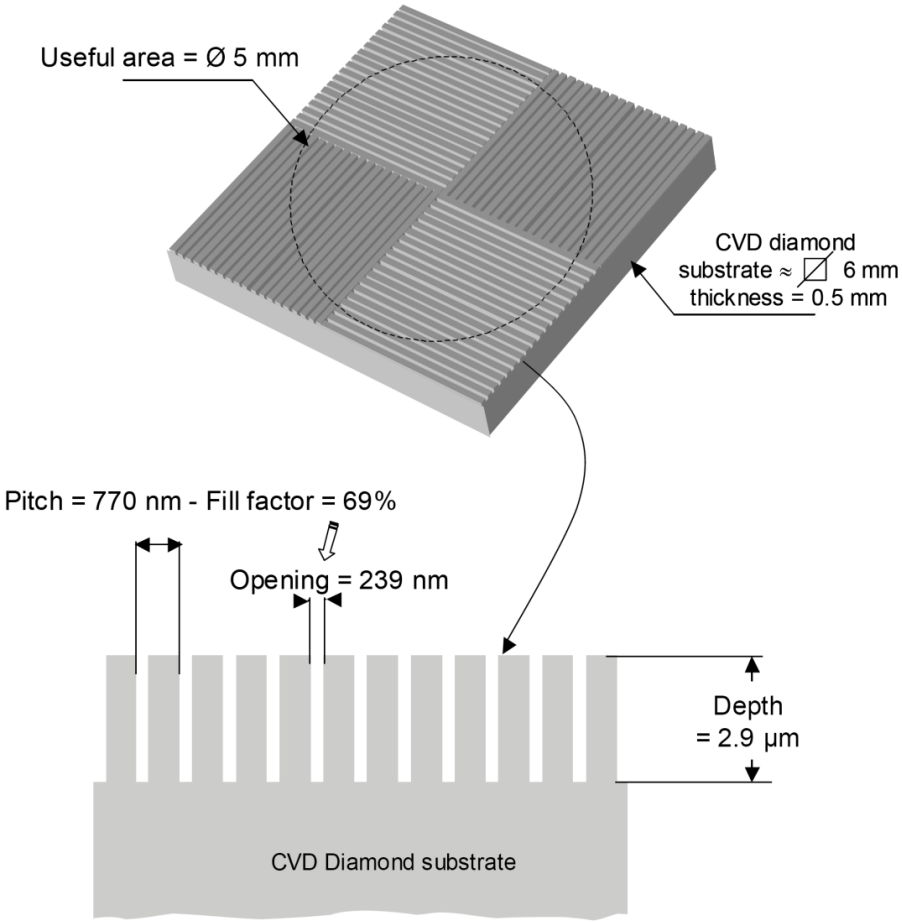

Figure 5.3: Schematic of the diamond 4QZOG specifications: period = 770 nm ± 50 nm; filling factor = 69%, i.e., openings = 239 nm ± 10 nm; depth = 2900 nm ± 20 nm. Orientation between quadrants: ±10 arcsec. Transitions at quadrant boundaries $< 1\mu$m.

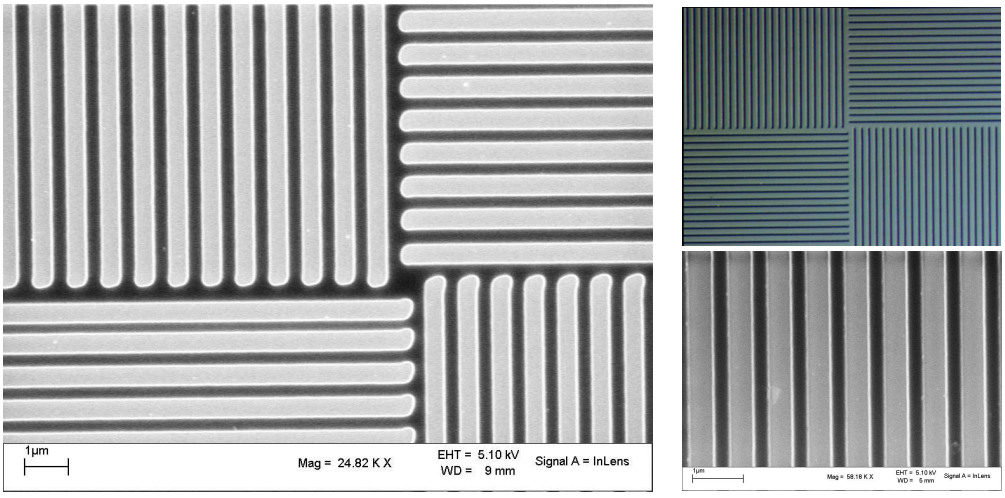

Figure 5.4: SEM pictures of a 4QZOG pattern imprinted in a PMMA resin mask using e-beam lithography. Courtesy of Mikael Karlsson (ADAMANTIS AB).

6

Annular groove phase-mask coronagraph

Contents

Abstract. *We present in this chapter a totally new concept of phase-mask coronagraph implemented with optimized subwavelength surface-relief gratings that provide an achromatic vectorial phase shift. This original design consists in a space-variant subwavelength grating inducing an optical vortex: the annular groove phase-mask (AGPM) coronagraph is fully symmetric and free from any "dead zones", i.e., focal-plane areas where potential companions or circumstellar features are attenuated, or hidden.*

6.1 Principle of the AGPM

The idea of the annular groove phase-mask (AGPM) coronagraph was to suppress the "dead zones" resulting from the quadrant transitions of the FQPM/4QZOG. Indeed, the latter induce a non-negligible attenuation of the superimposed circumstellar features or potential companions lying on them (Riaud et al. 2001). These dead zones represent quite a significant portion of the focal plane (about 10% at $6\lambda/D$). They induce a substantial loss of discovery space and information as well as obvious misleading artifacts in the case of extended objects like circumstellar disks (see, e.g. Sect. 1.3.3).

6.1.1 Space-variant subwavelength gratings leading to optical vortices

The concentric grooves of the AGPM coronagraph (Fig. 6.1) are in fact a space-variant subwavelength grating that synthesizes a spiral-phase plate leading to the subsequent formation of an optical vortex. Indeed, at the center of the component, the phase possesses a screw dislocation inducing a phase singularity, i.e., an optical vortex. The central singularity forces the intensity to vanish by a total destructive interference, creating a dark core. The optical-vortex dark core propagates and is conserved along the optical axis.

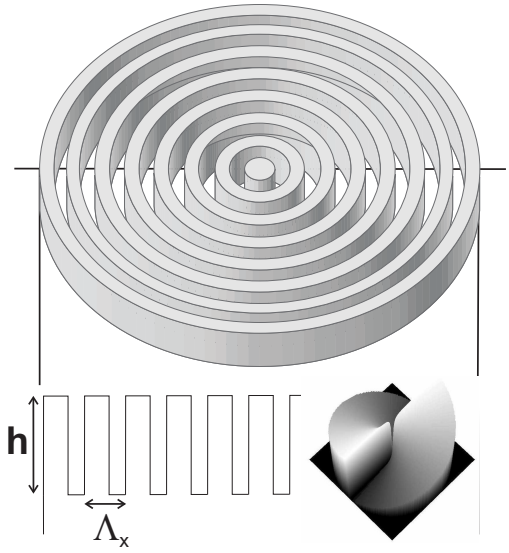

Figure 6.1: Schematic of the AGPM specific space-variant implementation of subwavelength gratings together, with the induced spiral phase ramp of 4π.

6.1.2 Optical vortices as coronagraphs

Vortices are fascinating features of waves and are commonly found in nature. In fluids, they appear as whirlpools, smoke rings, hurricanes, tornadoes. In cosmology the spirals of galaxies can be considered as vortices with their central singularity as well: supermassive black holes are suspected to be at the center of massive galaxies. In quantum mechanics, they are known to form in Bose-Einstein condensates, superfluids and superconductors. Optical vortices have been for twenty years at the heart of a resurgence leading to new potential applications, and among them, one of particular interest in modern astrophysics.

Whether a dark core is created in the pupil or focal plane of a telescope will determine the way it further evolves. In Swartzlander (2001), the author already proposed to create an optical vortex in the pupil plane to peer at the faint monochromatic signal in the relayed focal plane with appropriate filtering. In Mawet et al. (2005b), we proposed to do the inverse, i.e., creating an optical vortex in the focal plane, filtering in the relayed pupil plane and making the detection in a final image plane. This solution is theoretically much more attractive. Indeed, unlike the "optical vortex coronagraph" presented in Swartzlander (2001), it can be analytically demonstrated (Mawet et al. 2005b, see here below) that the theoretical attenuation of the AGPM is infinite in the ideal case[18]. Furthermore, the ZOG unique properties permit a broadband use as in the 4QZOG case whereas the component proposed in Swartzlander (2001) is monochromatic.

[18]This remarkable property has later on been independently confirmed in Foo et al. (2005).

Indeed, the ZOG parameters (period, depth, filling factor) are the same in the 4QZOG and AGPM configurations. The only difference between them concerns the geometry of the subwavelength grating implementation.

6.2 4QZOG vs AGPM

As can be seen in Fig. 6.2, the phase behavior of the AGPM is totally different from that of the 4QZOG (or FQPM) one. In the focal plane, the spatial variation of the phase is completely smooth for the AGPM while abrupt at the quadrant transitions for the 4QZOG (or FQPM). In the relayed pupil plane the rejection outside the pupil area is circularly symmetric for the AGPM while concentrated in four zones for the 4QZOG (FQPM). These simulated images stress out the advantage of the AGPM versus 4QZOG (FQPM) as far as discovery zone is concerned. In the case of extended objects, such an advantage of the AGPM over the 4QZOG (FQPM) is significant and not to be neglected.

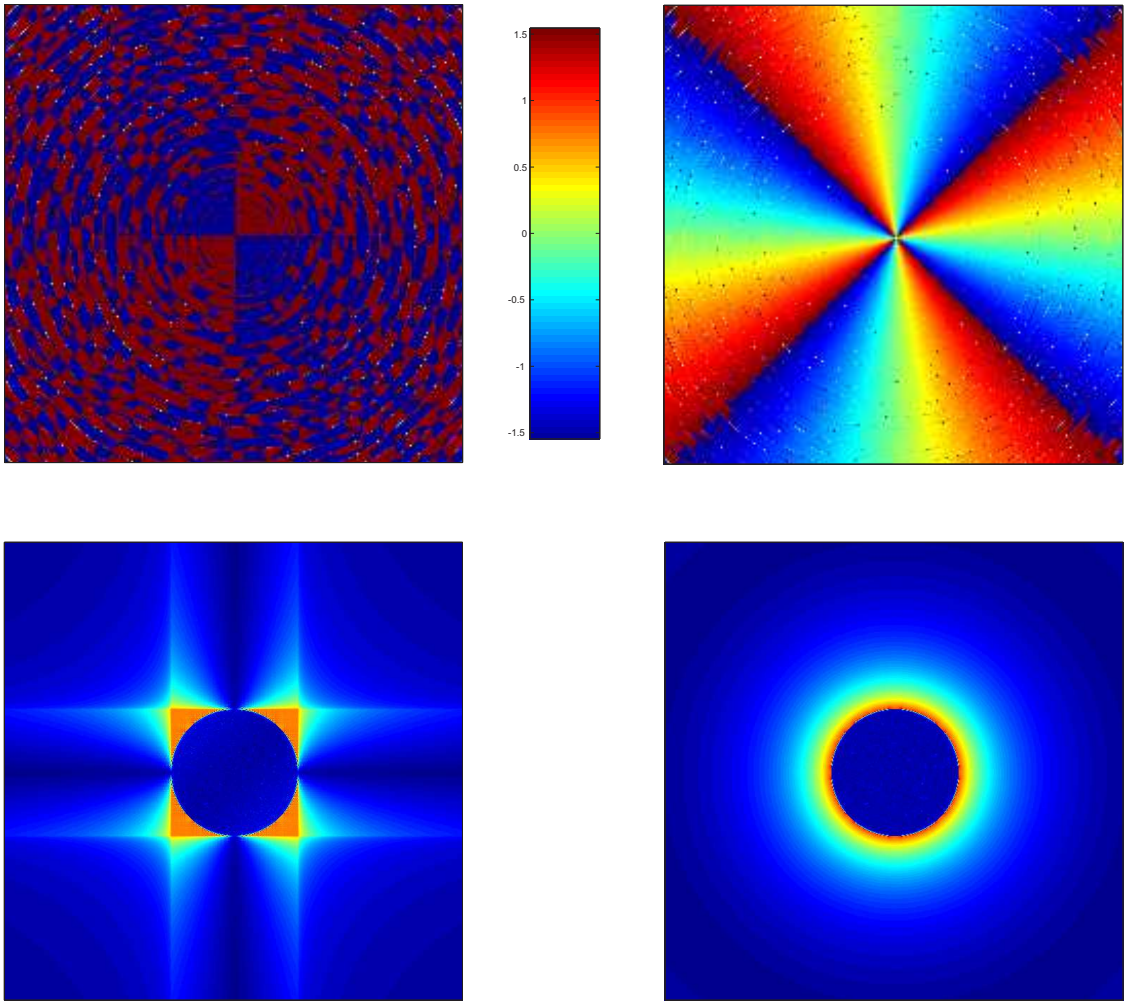

Figure 6.2: 4QZOG versus AGPM. Top: 4QZOG (left) vs AGPM (right) focal plane phase distribution of the product of the mask with a slightly aberrated Airy pattern (the phase scale in radian goes from $-\pi/2$ to $\pi/2$). We distinguish the brutal vs smooth phase transitions of the first with respect to the second. Bottom: pupil-plane intensity distribution. In both cases, the light is rejected outside the geometric pupil area but in the AGPM case (right), it is completely circularly symmetric.

6.3 Article: *Annular Groove Phase-Mask Coronagraph*

In the following paper, published in the *Astrophysical Journal*, we present the AGPM through a complete vectorial analysis, mixing a Jones-matrix formalism to RCWA. From the principle and RCWA calculations to realistic coronagraphic simulations, this original component is proven to be very interesting.

The Astrophysical Journal, 633:1191–1200, 2005 November 10

ANNULAR GROOVE PHASE MASK CORONAGRAPH

D. Mawet,[1] P. Riaud,[2] O. Absil,[3] and J. Surdej
Institut d'Astrophysique et de Géophysique, Université de Liège, 17 Allée du 6 Août, Bât B5c, B-4000 Liège, Belgium;
mawet@astro.ulg.ac.be, riaud@astro.ulg.ac.be, absil@astro.ulg.ac.be, surdej@astro.ulg.ac.be
Received 2005 May 10; accepted 2005 July 17

ABSTRACT

We present a new phase mask coronagraph consisting in an optical vortex induced by a space-variant surface relief subwavelength grating. Phase mask coronagraphy is a recent technique aiming at accommodating both high dynamic and high angular resolution imaging of faint sources around bright objects such as exoplanets orbiting their parent stars or host galaxies of active galactic nuclei. Subwavelength gratings are known to be artificially birefringent. Their unique dispersive characteristics can be controlled through the grating geometry in order to synthesize achromatic phase shifters. We show that implementing them in a ring-shaped way produces a fully symmetric and achromatic coronagraph without any gap or "dead zone." The practical manufacturing of the device is also discussed.

Subject headings: circumstellar matter — planetary systems — techniques: high angular resolution

1. INTRODUCTION

Direct detection of faint sources around bright astrophysical objects such as stars or active galactic nuclei (AGN) is very difficult due to the large flux ratio between them. For example, an Earth-like exoplanet is typically 6×10^9 times fainter than its host star in the visible and 7×10^6 times fainter in the thermal infrared, while the contrast of already known debris disks around main-sequence stars is generally larger than 1000 in the visible (e.g., β Pictoris's disk; see Smith & Terrile 1984). The circumnuclear structures of AGNs (obscuring torus, jet-induced structures, etc.) are at least 100 times less luminous than the central engine at visible and near-IR wavelengths (e.g., NGC 1068; see Rouan et al. 2004, for instance). The study of such objects therefore requires dedicated instruments such as coronagraphs. Current coronagraph designs are either pure amplitude masks (Lyot 1939) or pure phase masks (Roddier & Roddier 1997; Rouan et al. 2000; Soummer et al. 2003). Let us mention the special case of the achromatic interferocoronagraph (AIC; Gay & Rabbia 1996), which consists of a single-pupil achromatic nulling interferometer and also the so-called vortex spatial filter, which is a monochromatic pupil plane mask (Swartzlander 2001). The phase mask coronagraphs have been designed as alternative solutions to the amplitude coronagraphs to correct their inherent weakness: the physical extension of the opaque zone occults quite a significant fraction of the central field and thus all sources located behind it, i.e., near the bright object.

The four-quadrant phase mask coronagraph (FQ-PM) proposed by Rouan et al. (2000) is a very well performing design. The principle is to divide the focal plane into four equal areas centered on the optical axis, with two of them on a diagonal providing a π phase shift. This causes destructive interference ("nulling") to occur inside the geometric pupil area. The FQ-PM coronagraph has been validated on a laboratory bench in monochromatic light (Riaud et al. 2003) and installed on the NAOS-CONICA adaptive optics instrument (Boccaletti et al. 2004) at the ESO's Very Large Telescope (VLT). It has given promising preliminary scientific results (Gratadour et al. 2005) and perspectives for future instruments such as the European Mid-IR Instrument for NASA's *James Webb Space Telescope* (Baudoz et al. 2005) or the VLT Planet Finder, a second-generation instrument for the VLT (Mouillet et al. 2003). Unfortunately, the FQ-PM still possesses two drawbacks. First, the π phase shift is difficult to achieve in practice without or with a low chromaticity. Several solutions have nevertheless been studied; a very promising one we have recently proposed uses the unique dispersive characteristics of subwavelength gratings (Mawet et al. 2005). Second, the four phase transitions between adjacent quadrants create four λ/D-large "dead zones," where the potential circumstellar signal or companion is attenuated by up to 4 mag (Riaud et al. 2001).

In this paper, we propose a new design of a phase mask coronagraph derived from the FQ-PM that inherently allows the reduction of the chromaticity issues down to an acceptable level and totally suppresses the annoying dead zones of the latter. This new coronagraph is referred to as the annular groove phase mask (AGPM) coronagraph, since it is made up of a concentric circular subwavelength grating (see Fig. 1). The paper is organized as follows. In § 2 we present the principles of the AGPM coronagraph by introducing the subwavelength gratings, describing their so-called space-variant implementation, and finally discussing the chosen design. Section 3 is devoted to the realistic numerical simulations of the AGPM performance based on a three-stage modeling. In § 4 we briefly provide some manufacturing hints. Finally, we conclude by giving some perspectives on future applications in § 5. Some results and mathematical developments are detailed in the appendices.

2. PRINCIPLES OF THE AGPM CORONAGRAPH

The AGPM coronagraph is a focal plane microcomponent consisting of a concentric circular surface-relief grating with rectangular grooves of depth h equally separated by the period Λ (see Fig. 1). This coronagraph, working in natural light, is a pure vectorial phase mask, i.e., it induces a differential phase shift between the local polarization components of the incident natural (or polarized) light. As for every other coronagraph, the AGPM coronagraph is complemented by a well-dimensioned diaphragm

[1] Ph.D. student, under "FRIA" contract.
[2] Postdoctoral position, under "PAI" contract.
[3] Ph.D. student, under "FNRS" contract.

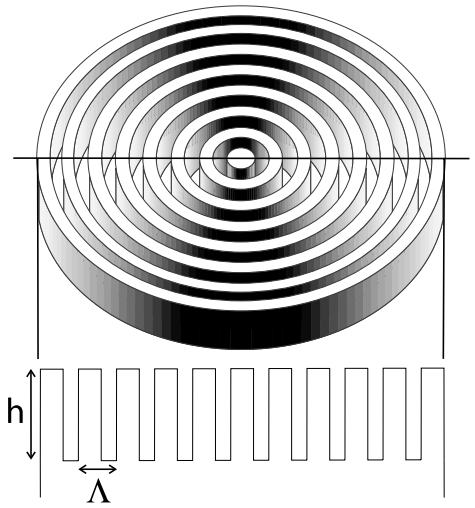

FIG. 1.—AGPM coronagraph scheme. The AGPM consists of a concentric circular surface-relief subwavelength grating with rectangular grooves of depth h and a periodicity Λ.

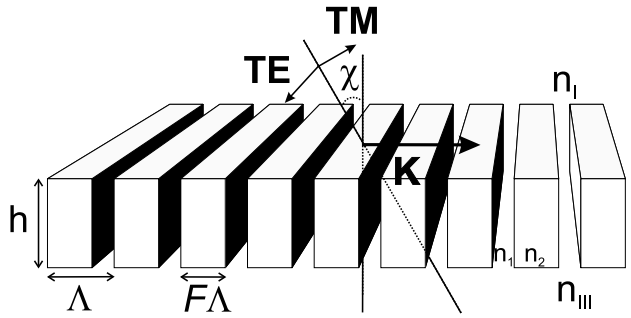

FIG. 3.—ZOG scheme presenting the main grating parameters: the grating vector $|\boldsymbol{K}| = 2\pi/\Lambda$, perpendicular to the grating lines, where Λ is the period; the grating depth h; and the so-called filling factor F, such that $F\Lambda$ is the width of the grating ridges. The vectors **TE** and **TM** are the vectorial orthogonal polarization components of the χ-incident light. Here $n_{\rm I}$ and $n_{\rm III}$ are the refractive indices of the incident and transmitting media, respectively. The parameters n_1 and n_2 are the refractive indices of the grating itself (in our case, $n_1 = n_{\rm I}$ and $n_2 = n_{\rm III}$). Finally, **TE** is the transverse electric vibration, where the electric field vector is perpendicular to the plane of incidence (the plane of incidence is defined by the grating normal and the direction of the incoming light, in our case by the grating normal and the grating vector), and **TM** is the transverse magnetic one (the electric field vector lies in the plane of incidence).

in the relayed pupil plane (Lyot stop) to suppress the diffracted starlight (for the optical implementation, see Fig. 2).

2.1. *Subwavelength Gratings*

When the period Λ of the grating is smaller than the wavelength of the incident light, it does not diffract as a classical spectroscopic grating. All the incident energy is forced to propagate only in the zeroth order, leaving incident wave fronts free from any further aberrations. The subwavelength gratings are therefore often called zeroth-order gratings (ZOGs). Whether a diffraction order propagates or not is determined by the well-known grating equation, from which a "ZOG condition" on the grating period to wavelength ratio can be derived,

$$\frac{\Lambda}{\lambda} \leq \frac{1}{n_{\rm I} \sin\chi + \max(n_{\rm I}, n_{\rm III})}, \tag{1}$$

where χ is the angle of incidence and $n_{\rm I}$ and $n_{\rm III}$ are the refractive indices of the incident (superstrate) and transmitting (substrate) media, respectively (see Fig. 3). This type of grating behaves like homogeneous media with unique characteristics, which can be used to synthesize artificial birefringent achromatic wave plates (Kikuta et al. 1997; Nordin & Deguzman 1999) or monolithic antireflective structures (see, e.g., Karlsson & Nikolajeff 2003). Quarter-wave or half-wave plates are extensively used in astrophysics for polarimetric studies. Subwavelength gratings constitute an elegant and flexible solution to produce these plates.

The key point is that by carefully controlling the geometry of the grating structure (via the grating parameters: the period Λ, the depth h, and the width of the grating ridges $F\Lambda$, where F is the filling factor), one can finely adjust the so-called form birefringence

$$\Delta n_{\rm TE-TM}(\lambda) = n_{\rm TE}(\lambda) - n_{\rm TM}(\lambda), \tag{2}$$

where $n_{\rm TE}$ and $n_{\rm TM}$ are the two effective indices associated with the subwavelength structure, one for each polarization state: **TE** (transverse electric, see Fig. 3) and **TM** (transverse magnetic). Intuitively, one can understand this artificial anisotropy and the existence of two distinct effective indices: the incident light sees two different media as its vectorial components vibrate parallel or orthogonal to the grating lines. The goal is to make the form birefringence proportional to the wavelength in order to compensate for the hyperbolic dependence of the subsequent differential phase shift between the two polarization components **TE** and **TM** and thus achromatize it at the required value of π,

$$\Delta\phi_{\rm TE-TM}(\lambda) = \frac{2\pi}{\lambda} h \Delta n_{\rm TE-TM}(\lambda) \approx \pi, \tag{3}$$

where h is the optical path through the birefringent medium.

2.2. *Space-variant ZOGs*

The concentric circular grooves of the AGPM coronagraph are in fact what is called a "space-variant" ZOG: when the local characteristics (period, orientation of the grating lines, etc.) of the structure vary from point to point, it is said to be space variant. Such components were recently extensively studied as polarization control elements (Niv et al. 2003; Biener et al. 2002;

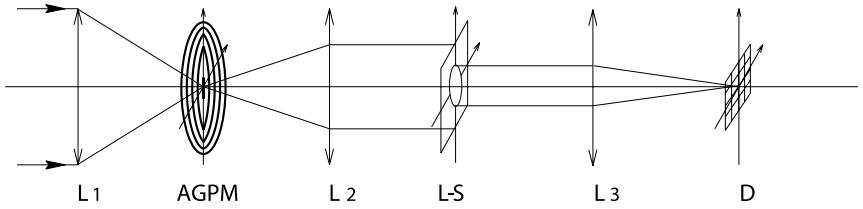

FIG. 2.—Basic AGPM coronagraphic optical bench scheme. L1, L2, and L3 are three lenses in the optical system. L1 provides a large (to minimize spatial defects) F/d ratio on the AGPM, L2 images the pupil in the second plane, the Lyot stop (L-S) suppresses the diffracted starlight, and finally L3 forms the coronagraphic image on the detector D.

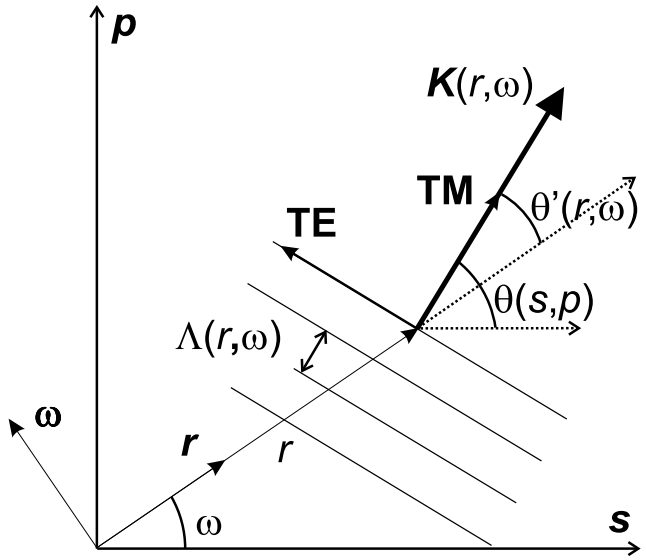

FIG. 4.—Space-variant ZOG vectorial analysis. Here $\boldsymbol{s}$ and $\boldsymbol{p}$ are the unit vectors of the chosen Cartesian basis, whereas $\boldsymbol{\omega}$ and $\boldsymbol{r}$ are the polar coordinate unit vectors. In addition, **TE** (transverse electric) and **TM** (transverse magnetic) are the polarization unit vectors according to the local grating line orientations. By definition, in normal incidence, the **TE** (**TM**) components are orthogonal (parallel) to the local grating vector $\boldsymbol{K}(s,p)$ [$= \boldsymbol{K}(\omega,r)$], spanning angles $\theta(s,p)$ and $\theta'(\omega,r)$ with respect to $\boldsymbol{s}$ and $\boldsymbol{r}$, respectively. Finally, $\Lambda(s,p)$ [$= \Lambda(\omega,r)$] is the grating period.

Bomzon et al. 2002; Levy et al. 2004). Applications are numerous: for example, polarimetry, laser-beam shaping, laser machining, tight focusing, particle acceleration, and atom trapping. Space-variant ZOGs are typically described by a function representing the grating vector spatial variation,

$$\boldsymbol{K}(s,p) = K_0(s,p)[\cos\theta(s,p)\boldsymbol{s} + \sin\theta(s,p)\boldsymbol{p}], \qquad (4)$$

where $\boldsymbol{s}$ and $\boldsymbol{p}$ are the Cartesian unit vectors and $K_0(s,p) = 2\pi/\Lambda(s,p)$ is the grating vector modulus corresponding to the local period $\Lambda(s,p)$. Here $\theta(s,p)$ is the local direction of the grating vector with respect to $\boldsymbol{s}$, the space-variant grating vector always being perpendicular to the local grating lines (see Fig. 4). In polar coordinates, we have

$$\boldsymbol{K}(r,\omega) = K_0(r,\omega)[\cos\theta'(r,\omega)\boldsymbol{r} + \sin\theta'(r,\omega)\boldsymbol{\omega}], \qquad (5)$$

where $\boldsymbol{r}$ and $\boldsymbol{\omega}$ are the polar coordinate unit vectors. Here $\theta'(\omega,r)$ is the local direction of the grating vector with respect to $\boldsymbol{r}$ (see Fig. 4). Let us now consider the general case of the spiral geometric phase space-variant ZOGs. The grating groove direction in this case is given by $\theta(s,p) = l_p\omega/2$ or $\theta'(r,\omega) = (l_p/2-1)\omega$, where l_p is the so-called topological Pancharatnam charge (a nonsigned integer; see Appendix B). The grating vector therefore becomes

$$\boldsymbol{K}(r,\omega) = K_0(r,\omega)\{\cos[(l_p/2-1)\omega]\boldsymbol{r} + \sin[(l_p/2-1)\omega]\boldsymbol{\omega}\}. \qquad (6)$$

The continuity of the grating grooves is ensured by imposing $\nabla \times \boldsymbol{K} = 0$, which also implies that the grating vector derives from a grating function Φ ($\boldsymbol{K} = \nabla\Phi$). Integration over an arbitrary path yields

$$\Phi(r,\omega) = \begin{cases} 2\pi\dfrac{r_0}{\Lambda_0}\left(\dfrac{r_0}{r}\right)^{l_p/2-1}\dfrac{\cos(l_p/2-1)\omega}{(l_p/2-1)}, & l_p \neq 2, \\ 2\pi\dfrac{r_0}{\Lambda_0}f(r), & l_p = 2. \end{cases} \qquad (7)$$

This function describes a family of binary gratings depending on the topological charge l_p (Fig. 5). Let us remark that the continuity criterion has been introduced for manufacturing convenience. In the $l_p = 2$ case, the circular symmetry allows the choice of any pure radial function. The AGPM corresponds to

$$\Phi_{\mathrm{AGPM}}(r,\omega) = 2\pi\frac{r}{\Lambda_0}, \quad l_p = 2. \qquad (8)$$

The family of spiral phase space-variant ZOGs creates an "optical vortex." Indeed, at the center of these components, the phase possesses a screw dislocation inducing a phase singularity, i.e., an optical vortex. The central singularity forces the intensity to vanish by a total destructive interference, creating a dark core. This dark core propagates and is conserved along the optical axis. Whether a dark core is created in the pupil or focal plane of a telescope will determine the way it further evolves. Swartzlander (2001) proposed to create an optical vortex in the pupil plane to peer at the faint monochromatic signal in the relayed focal plane with appropriate filtering. In this paper, we propose to do the inverse, i.e., to create an optical vortex in the focal plane, filter in the relayed pupil plane, and make the detection in a final image plane. This solution is theoretically much more attractive, as we will see. Furthermore, the ZOG's unique properties permit an efficient broadband use.

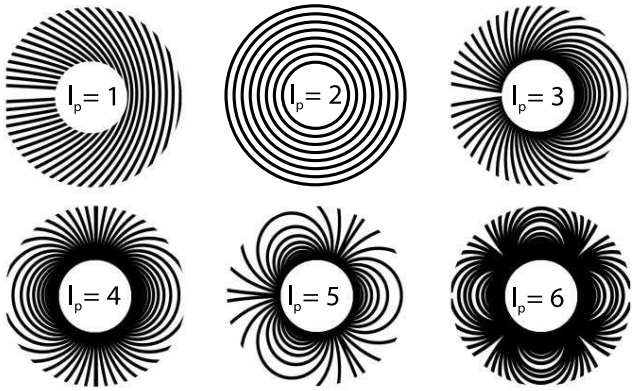

FIG. 5.—Binary grating geometry for topological charges l_p ranging from 1 to 6. Only the $l_p = 2$ geometry possesses the required circular symmetry for use with constant ZOG parameters, which permit achromatization. The component centers have been occulted for presentation purposes.

2.3. *AGPM Coronagraph*

The AGPM coronagraph corresponds to the spiral phase of topological charge $l_p = 2$, implying that the Pancharatnam phase (see Appendix B) undergoes two 2π phase jumps within one revolution around the optical axis (see Fig. 6). This phase modification results solely from the polarization manipulation and is purely geometrical in nature. In the $l_p = 2$ case, a given polarization state repeats itself $2l_p = 4$ times. This point is argued for the linear polarization case in § A1 (a full analytical treatment of the polarization using space-variant Jones matrices is presented in Appendix A). We also show in § A2 that the

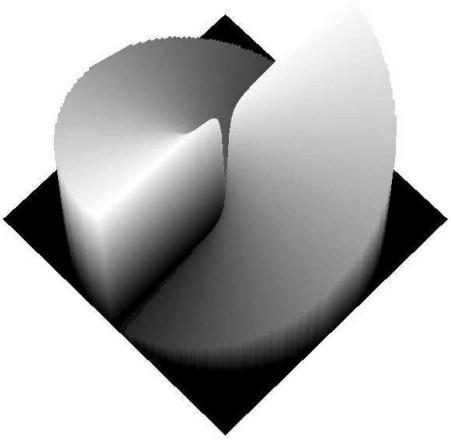

FIG. 6.—Pancharatnam phase ramp of the AGPM coronagraph: $\phi_p = 2\omega$. The associated topological charge is $l_p = 2$. Within one revolution around the optical axis, i.e., $\omega = 2\pi$, one easily confirms that $\phi_p = 2(2\pi)$.

Jones vector for the output components can also be described in a helical polarization basis, with right-handed ($\boldsymbol{R}^{\circlearrowright}$) and left-handed ($\boldsymbol{L}^{\circlearrowleft}$) circularly polarized input fields. In this particular case and under ideal conditions, we obtain at the output

$$\boldsymbol{R}^{\circlearrowright} = \begin{bmatrix} 0 \\ e^{i(2\omega-\pi/2)} \end{bmatrix}, \quad \boldsymbol{L}^{\circlearrowleft} = \begin{bmatrix} e^{-i(2\omega+\pi/2)} \\ 0 \end{bmatrix}. \tag{9}$$

The Pancharatnam phase clearly appears as the argument of the exponential, $\phi_p = 2\omega$. Therefore, within one revolution, i.e., $\omega = 2\pi$, one easily confirms that $\phi_p = 2(2\pi)$. In addition, the helical basis allows us to decouple the output polarization components. This facilitates the forthcoming Fourier analysis.

We demonstrate in Appendix C, thanks to Sonine's integral (Sneddon 1951, p. 55), that in the $l_p = 2$ configuration the vortex propagation up to the relayed pupil plane evolves into a perfect destructive interference, totally rejecting the starlight outside the geometric pupil area (we also analytically demonstrate that the perfect attenuation holds true for even values of l_p). Like the FQ-PM, the theoretical attenuation of the AGPM is therefore infinite in the perfect achromatic and circular filled pupil case (Riaud et al. 2001). We have also chosen the $l_p = 2$ case for the following reason: in order to be achromatic, the space-variant ZOG local characteristics (grating period, depth, and filling factor) are well defined and do not tolerate any departure from optimal values within the tolerances (see § 4). We note in Figure 5 that only the $l_p = 2$ case affords the required symmetry to fulfill this constraint. The other configurations ($l_p \neq 2$) all imply a variation of the grating period that would destroy the achromatic characteristics of the phase shift. Moreover, such a variation of the period could lead the grating to exit the subwavelength domain with dramatic consequences: higher diffraction orders would show up.

The AGPM implementation of the space-variant ZOG is thus totally circularly symmetric. The grating vector is constant in modulus and aligned with the radius. In other words, the AGPM coronagraph can be seen as a FQ-PM coronagraph in polarization. Indeed, if we consider the four cardinal points on the AGPM, the resulting phase shift distribution is analogous to the FQ-PM for each parallel potentially interfering polarization state (see Fig. 7). This argument holds true for each azimuth angle and for each radius, and thus for the whole focal plane.

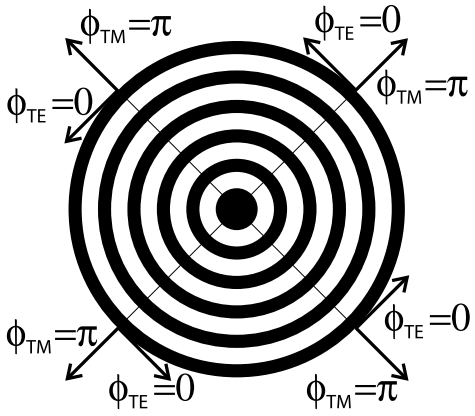

FIG. 7.—AGPM scheme and analogy with the FQ-PM coronagraph. The AGPM can be seen as a polarization FQ-PM. The parallel potentially interfering polarization states are out of phase according to the FQ-PM focal plane phase shift distribution. Here $\phi_{\rm TE}$ and $\phi_{\rm TM}$ are the output phases of the polarization components **TE** and **TM** such that $\Delta\phi_{\rm TE-TM} = |\phi_{\rm TE} - \phi_{\rm TM}| = \pi$.

3. NUMERICAL RESULTS IN A REALISTIC CASE

We have performed realistic numerical simulations that rely on a three-stage modeling:

1. A "rigorous coupled wave analysis" stage, where the form birefringence of the local grating is optimized, leading to the space-variant ZOG Jones matrix $J_{\rm ZOG}(s, p)$. At this stage, the final performance of the coronagraph can already be quantified by the null depth.
2. The analytical polarization treatment based on Jones calculus, which gives the spatial distribution of the linear/helical polarization components of the incident light. We use for this step the results obtained in Appendix A.
3. A scalar far-field Fourier propagation coronagraphic code for each polarization state.

To simulate the grating response and calculate the form birefringence $\Delta n_{\rm form} = \Delta n_{\rm TE-TM}$ in the subwavelength and resonant domain ($\Lambda \approx \lambda$), scalar theories of diffraction dramatically fail. The vectorial nature of light must be taken into account, implying a resolution of the Maxwell equations by the so-called rigorous coupled wave analysis (RCWA; Moharam & Gaylord 1981). RCWA gives the full diffractive characteristics of the simulated structure.

The ZOG form birefringence optimization has already been extensively presented in Mawet et al. (2005) in the context of the FQ-PM achromatization for the *H*, *K*, and *N* bands (4QZOG). We focus here on the mostly used *K* band, but the conclusions are applicable to other band filters. In Figure 8 we present the RCWA results for a subwavelength surface-relief grating engraved on the surface of a diamond (C) or ZnSe substrate and covered by a $\approx\lambda/4$ antireflective (AR) layer of YF_3. The latter settles at the bottom of the grooves and on top of the grating ridges. The null depth, which characterizes the darkness of the destructive interference taking place in the relayed pupil plane of the telescope, takes into account the phase errors with respect to π, $\epsilon(\lambda) = \Delta\phi_{\rm TE-TM}(\lambda) - \pi$, and amplitude mismatches $q(\lambda) = \eta_{\rm TE}(\lambda)/\eta_{\rm TM}(\lambda)$ in the following way:

$$N(\lambda) = \frac{\left[1 - \sqrt{q(\lambda)}\right]^2 + \epsilon(\lambda)^2\sqrt{q(\lambda)}}{\left[1 + \sqrt{q(\lambda)}\right]^2}. \tag{10}$$

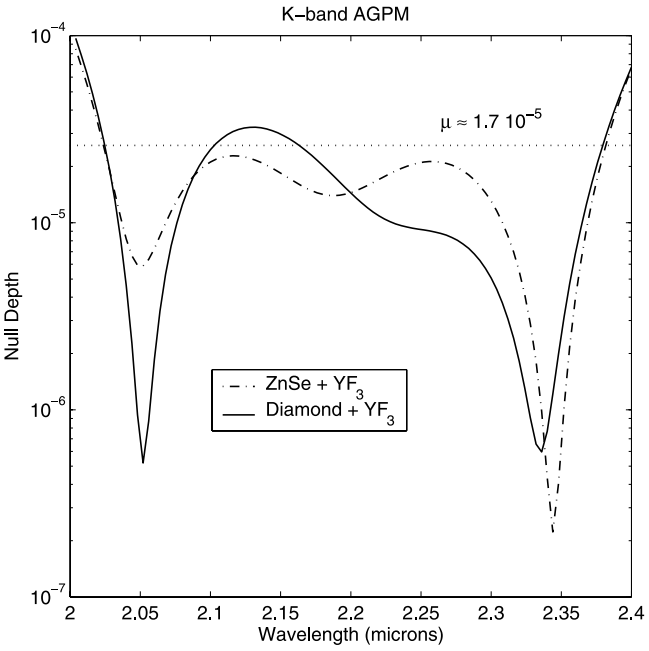

FIG. 8.—*K*-band AGPM coronagraph null depth vs. wavelength. The solid curve is for the diamond YF_3 AR coated ZOG. The dot-dashed curve is for the ZnSe YF_3 AR coated one. The mean null depth over the whole *K* band is $\mu \approx 1.7\times10^{-5}$.

We can notice in Figure 7 the efficiency of the ZOG solution in solving the chromaticity problem for the *K* band with a mean null depth $\mu \approx 1.7\times10^{-5}$. As said before, deep nulls can also be achieved for other usual band filters (see Table 1). It must be noted that the optical throughput efficiency of the optimized ZOG is >90%. For the sake of clarity, the third stage of the AGPM simulations (Fourier propagation) has been performed for the *K* band only. We have used an IDL code for Fraunhofer diffraction analysis (Riaud et al. 2001). To minimize the aliasing effect of the fast Fourier transform, we have used large arrays (up to 2048×2048) for the calculation. The intrinsic performance of the coronagraph will be limited by the phase residuals with respect to π and the transmittance mismatches between the two polarization states **TE** and **TM**, as well. We have also assumed wave front qualities of $\lambda_0/250$ rms, where λ_0 is the central wavelength of the considered filter. In our case (*K* band: $\lambda_0 = 2.2$ μm), this hypothesis leads therefore to wave front qualities of $\approx\lambda/70$ rms, with $\lambda = 632.8$ nm. This is quite a severe but somewhat realistic figure. Indeed, the Virgo team (Mackowski et al. 1999; Brillet et al. 2003) has demonstrated state-of-the-art mirror quality with an excellent polishing realization ($\lambda/226$ rms at 632.8 nm) within the framework of gravitational wave detection. This technology is directly applicable to our case.

The final results of our three-stage calculation are excellent (see Fig. 9). As demonstrated in Appendix C in an analytical way, the starlight is rejected outside the geometric pupil area in a fully symmetric annular shape. The smoothness of the phase shift occurring in the focal plane ensures the absence of the dead zones induced by the FQ-PM/4QZOG quadrant transitions. The final *K*-band polychromatic image reveals the simulated companion 15 mag fainter. The coronagraphic profile functions of the angular separation in λ/d (d is the telescope diameter) show a peak-to-peak attenuation of about 10^{-5} (Fig. 10). The speckle level of $\approx10^{-7}$ is quickly reached at a few λ/d. The AGPM coronagraphic behavior is very similar to the achromatic 4QZOG coronagraph (Mawet et al. 2005), but with a total symmetry.

Figure 11 presents the attenuation of the off-axis simulated companion, which is also quite similar to the FQ-PM/4QZOG in its best configuration, i.e., only along the two diagonals. Indeed, as mentioned in the introduction, the FQ-PM/4QZOG quadrant transitions induce a nonnegligible attenuation of the superimposed circumstellar features lying on them. These dead zones represent quite a significant portion of the focal plane (about 10% at $6\lambda/d$). Thanks to the perfect AGPM circular symmetry, this problem does not exist anymore. We also note that the inner working angle of the AGPM is very good, peering well under λ/d. As far as stellar leakage is concerned, numerical simulations show that it increases as $\theta^2_{\lambda/d}$ close to the optical axis, just as the FQ-PM/4QZOG (where $\theta_{\lambda/d}$ is the angle from the optical axis). In fact, calculations show that for a vortex of topological charge l_p, the stellar leakage grows as $\theta^{l_p}_{\lambda/d}$.

4. MANUFACTURING

In the *K*-band diamond AGPM case, for instance, the local ZOG optimal parameters are

1. period, $\Lambda = 740$ nm;
2. filling factor, $F = 70\%$;
3. total depth, $h = 3.240$ μm; and
4. AR layer thickness, 420 nm.

The fabrication of the AGPM coronagraph implies no developments other than those for the 4QZOG (Mawet et al. 2005), which is currently under assessment in diamond by Uppsala University's Angström Laboratory "Adamantis AB" (Karlsson

TABLE 1
AGPM ACHROMATIZATION (NULL DEPTH) FOR VARIOUS BAND FILTERS: VISIBLE, NEAR-IR AND MID-IR

PARAMETER	FILTER			
	V ($R_\lambda = 5.5$)	I ($R_\lambda = 3.75$)	J ($R_\lambda = 4.16$)	H ($R_\lambda = 4.7$)
Null depth (on-axis).....	2×10^{-5}	1.35×10^{-4}	6.5×10^{-5}	3.5×10^{-5}
Expected contrast (at $3\lambda/d$).....	1.66×10^{-7}	1.12×10^{-6}	5.42×10^{-7}	2.92×10^{-7}
Grating period (nm).....	280 (n-LAF32)	305 (C)	400 (C)	525 (C)
	K ($R_\lambda = 5.5$)	L' ($R_\lambda = 6.3$)	M ($R_\lambda = 16.6$)	N ($R_\lambda = 4.86$)
Null depth (on-axis).....	1.7×10^{-5}	8.4×10^{-6}	5×10^{-7}	4×10^{-5}
Expected contrast (at $3\lambda/d$).....	1.42×10^{-7}	7×10^{-8}	4.2×10^{-9}	3.3×10^{-7}
Grating period (μm).....	0.740 (C)	1.28 (C)	1.7194 (C)	3.29 (C)

NOTES.—Here $R_\lambda = \lambda/\Delta\lambda$ is the spectral resolution. C stands for CVD diamond, while n-LAF32 refers to a high-index Schott glass.

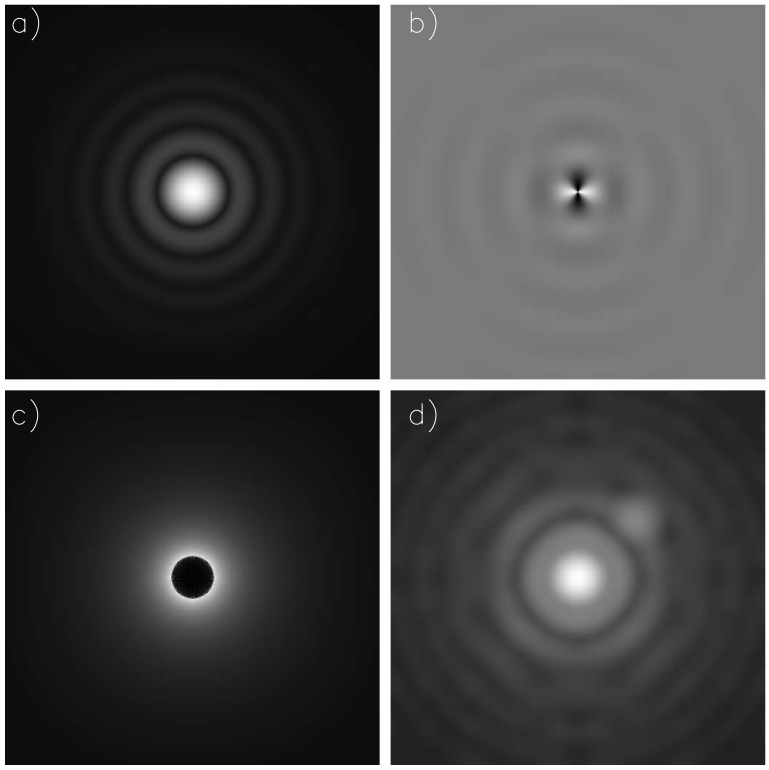

FIG. 9.—Numerical simulation illustrating the diffractive behavior of the AGPM coronagraph. (*a*) Airy pattern provided by a perfect telescope without central obscuration. We have also added a companion 15 mag fainter located $2\lambda/d$ ($\lambda = 2.2$ μm) away from the star. (*b*) Complex amplitude of the star phase shifted by the mask. We note that the phase mask effect is close to that obtained for the FQ-PM coronagraph, but without any discontinuity left. (*c*) Picture showing the starlight rejection in the relayed pupil plane. The diffraction pattern is annular and symmetric in this configuration. (*d*) Resulting coronagraphic image for the full *K* band, where the fainter companion is clearly visible. All images are displayed with a nonlinear scale.

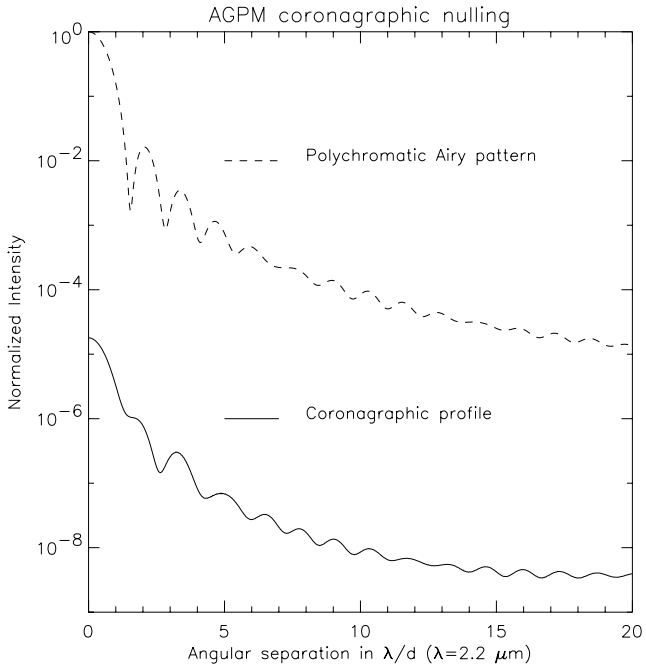

FIG. 10.—Theoretical radial profiles obtained with the *K*-band AGPM. The solid line shows the coronagraphic profile. The residual central peak is due to the effect of the phase errors (residual chromatism) around the π phase shift. In this case, a starlight speckle level of 10^{-7} is reached at $3\lambda/d$. The dashed line shows the polychromatic Airy pattern for the full *K* band. The diaphragm (Lyot stop) is open at 80%.

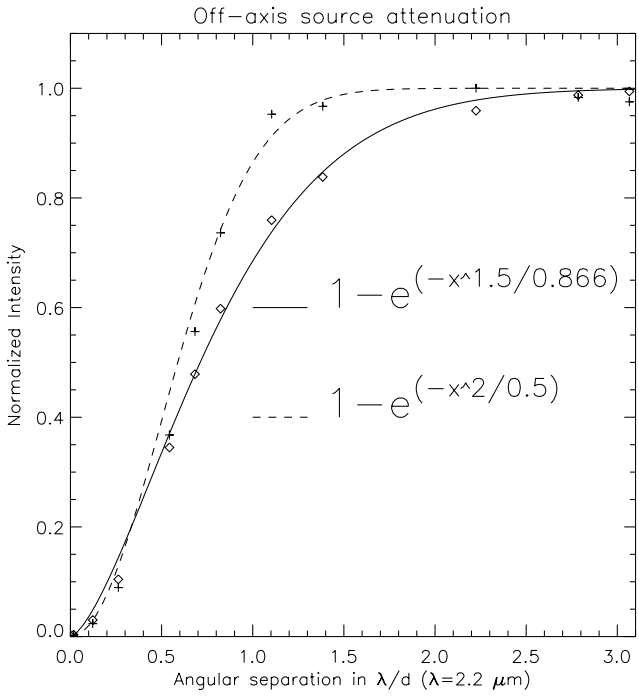

FIG. 11.—Degradation of the coronagraphic performance function of angular separation. This figure compares the companion attenuation for the AGPM vs. the classical 4QZOG in its best configuration, i.e., at least $1\lambda/d$ away from a quadrant transition. This degradation is measured on the total energy. The solid line shows the exponential fit on the simulated data (*open diamonds*) for the AGPM coronagraph. The dashed line shows the exponential fit on the simulated data (*plus signs*) for the FQ-PM/4QZOG coronagraph.

& Nikolajeff 2003; Karlsson et al. 2001). The manufacturing is based on microelectronic technologies. The first step consists in the definition of the lithographic mask: one has to imprint the concentric annular pattern in a resin coated on the chosen substrate material. It can be realized by laser direct writing or *e*-beam lithography. The precision of this step is critical, because it defines once and for all the lateral dimensions of the ZOG, i.e., the filling factor (F). The latter is the determinant parameter of the grating structure and the most difficult to control during the fabrication process. A complete study of the design sensitivity to the parameters has been presented in Mawet et al. (2005). The conclusion was that the tolerance on the filling factor was at the 1% level but also that if the manufacturing process was interactively conducted, then errors on the filling factor definition could be compensated a posteriori. The next fabrication steps consist in transferring the mask pattern into the substrate by an appropriate reactive plasma beam etching down to the desired depth, followed by the $\lambda/4$ AR layer deposition. Assuming a classical realistic resolution of 10 nm in thickness (2%) for the AR layer sputtering, we can ensure a grating etching depth tolerance of about 20 nm at the null depth level of 10^{-5}. This value is well within reach with current technologies, especially if in situ real-time monitoring of the grating parameters is implemented during the fabrication process (Lalanne et al. 1999).

5. DISCUSSION

In this paper, we have presented a new phase mask coronagraph that is free from any "dead zone," thanks to its perfect circular symmetry, and inherently quasi-achromatic. The AGPM coronagraph consists in an optical vortex induced by a space-variant subwavelength grating. The potential performance of the AGPM coronagraph is very good, ensuring, for instance, a theoretical contrast of 1.4×10^{-7} at $3\lambda/d$ over the whole K band with inherent perfect symmetry. The inner working angle of the mask is $\approx\lambda/d$, still with a good contrast of $\approx10^{-5}$. Thanks to the ZOG design flexibility, the AGPM coronagraph can accommodate a large variety of materials and wavelength bands (see Table 1), thus making it an attractive solution for future ambitious high-resolution/high-contrast space- or ground-based imaging facilities. For instance, the AGPM coronagraph could be used alone on a single-pupil telescope either in space or on the ground (with an adaptive optics system) to dramatically enhance the dynamics. It could also be used at the Fizeau or densified focus of an interferometer (Labeyrie 1996; Boccaletti et al. 2000; Riaud et al. 2002) to take advantage of the increased resolution. However, in the Fizeau configuration, phase coronagraphs are limited by the cross-talk between the different interferometer subpupils, whereas in the densified one, the limitation comes from diffraction effects induced by residual gaps between the joined subpupils (P. Riaud et al., in preparation). Nevertheless, the AGPM should be seriously regarded as an integrated high-contrast solution to be implemented in NASA's *Terrestrial Planet Finder* and/or ESA's infrared space interferometer, *DARWIN*, missions.

D. M. acknowledges the financial support of the Belgian "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture." P. R. and J. S. acknowledge the financial support of the "Pôle d'Attraction Inter-Universitaire."

APPENDIX A

POLARIZATION TREATMENTS

Let us perform a full space-variant polarization and phase analysis of the component, assuming first that it induces an optical vortex of l_pth order. Representing it by a space-variant Jones matrix, one can find the resulting wave front for any incident polarization,

$$J_{\mathrm{vortex}}(s,p) = M[\theta(s,p)]J_{\mathrm{ZOG}}(s,p)M[\theta(s,p)]^{-1}. \tag{A1}$$

Here $J_{\mathrm{ZOG}}(s, p)$ actually describes the effects of the local ZOG form birefringence that transforms the phase (**TE** − **TM** vectorial phase shift $\Delta\phi_{\mathrm{TE-TM}}$) and amplitude (**TE** − **TM** differential Fresnel parasitic reflection) of the outgoing beam,

$$J_{\mathrm{ZOG}}(s,p) = \begin{bmatrix} \eta_{\mathrm{TE}} & 0 \\ 0 & \eta_{\mathrm{TM}}e^{i\Delta\phi_{\mathrm{TE-TM}}} \end{bmatrix}, \tag{A2}$$

where η_{TE} and η_{TM} are the local grating transmittances along the **TE** and **TM** directions of polarization, respectively (the transmittances can be assimilated to diffraction efficiencies). These transmittances are different because of the different Fresnel reflection coefficients resulting from the existence of the two effective indices that give birth to the form birefringence (see eq. [2]). For the sake of simplicity, we write $\Delta\phi = \Delta\phi_{\mathrm{TE-TM}}$. The parameter $M[\theta(s, p)]$ is the θ-dependent rotation matrix

$$M[\theta(s,p)] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \tag{A3}$$

where $\theta(s, p)$ is the grating vector angle that defines the local grating line orientations. Thus, we have

$$J_{\mathrm{vortex}}(s,p) = \begin{bmatrix} \eta_{\mathrm{TE}}\cos^2\theta + \eta_{\mathrm{TM}}\sin^2\theta e^{i\Delta\phi} & \sin\theta\cos\theta\left(\eta_{\mathrm{TE}} - \eta_{\mathrm{TM}}e^{i\Delta\phi}\right) \\ \sin\theta\cos\theta\left(\eta_{\mathrm{TE}} - \eta_{\mathrm{TM}}e^{i\Delta\phi}\right) & \eta_{\mathrm{TM}}\sin^2\theta + \eta_{\mathrm{TE}}\cos^2\theta e^{i\Delta\phi} \end{bmatrix}. \tag{A4}$$

In order to carry out the analysis of the component with the scalar Fourier coronagraphic propagation code, we have to chose a basis to project the incident polarization (natural or not).

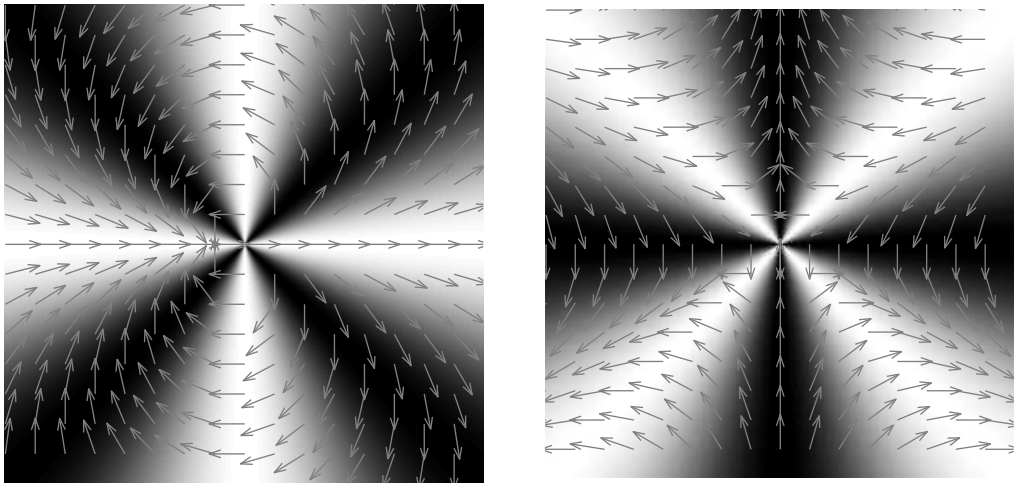

FIG. 12.—AGPM response to linear polarization. *Left:* Intensity map for an input linear horizontal polarization seen by a horizontal analyzer. *Right:* Same, but for an input linear vertical polarization. Arrows show the corresponding vectorial polarization field that has been submitted to the rotation given by eqs. (A9) and (A10). We clearly notice that a given polarization state repeats itself four times.

A1. LINEAR BASIS

We can decompose the problem by projecting the incident polarization on the orthogonal linear ($\boldsymbol{s}, \boldsymbol{p}$) basis. Therefore, we have the following Jones vectors as linear polarization inputs:

$$\boldsymbol{E}_p = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{A5}$$

$$\boldsymbol{E}_s = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{A6}$$

Multiplying them both by the vortex Jones matrix $J_{\rm vortex}(s, p)$, we obtain

$$\boldsymbol{E}_p = \begin{bmatrix} \eta_{\rm TE}\cos^2\theta + \eta_{\rm TM}\sin^2\theta e^{i\Delta\phi} \\ \sin\theta\cos\theta\left(\eta_{\rm TE} - \eta_{\rm TM}e^{i\Delta\phi}\right) \end{bmatrix}, \tag{A7}$$

$$\boldsymbol{E}_s = \begin{bmatrix} \sin\theta\cos\theta\left(\eta_{\rm TE} - \eta_{\rm TM}e^{i\Delta\phi}\right) \\ \eta_{\rm TM}\sin^2\theta + \eta_{\rm TE}\cos^2\theta e^{i\Delta\phi} \end{bmatrix}, \tag{A8}$$

respectively, corresponding to the output polarization states to be injected in a subsequent coronagraphic code. In the perfect case (exact π phase shift, i.e., $\Delta\phi_{\rm TE-TM} = \pi$ and unitary matched efficiencies $\eta_{\rm TE} = \eta_{\rm TM} = 1$) for the AGPM configuration ($l_p = 2$ and thus $\theta = l_p\omega/2 = \omega$, where ω is the azimuthal polar coordinate), we have

$$\boldsymbol{E}_p = \begin{bmatrix} \cos 2\omega \\ \sin 2\omega \end{bmatrix}, \tag{A9}$$

$$\boldsymbol{E}_s = \begin{bmatrix} \sin 2\omega \\ -\cos 2\omega \end{bmatrix}. \tag{A10}$$

This implies that in the AGPM case ($l_p = 2$) an input linear polarization, which is horizontal, for example, will locally rotate by twice the azimuthal angle ω, as shown in Figure 12, where the corresponding output vectorial field and intensity response to linear polarization are displayed.

A2. HELICAL BASIS

The analysis can be decoupled by projecting the incident vectorial field on a helical basis, i.e., with right- and left-handed circular polarization unit vectors

$$\boldsymbol{R}^{\circlearrowleft} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{A11}$$

$$\boldsymbol{L}^{\circlearrowleft} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{A12}$$

In such a case, the vortex component Jones matrix must be transformed by

$$J^{\circlearrowright\circlearrowleft}_{\text{vortex}}(s,p) = UJ_{\text{vortex}}(s,p)U^{-1}, \tag{A13}$$

with the helical-basis transformation matrix

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix}. \tag{A14}$$

Finally, we have

$$J^{\circlearrowright\circlearrowleft}_{\text{vortex}}(s,p) = \frac{1}{2}\left(\eta_{\text{TE}} + \eta_{\text{TM}} e^{i\Delta\phi}\right) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2}\left(\eta_{\text{TE}} - \eta_{\text{TM}} e^{i\Delta\phi}\right) \begin{bmatrix} 0 & e^{i2\theta} \\ e^{-i2\theta} & 1 \end{bmatrix}. \tag{A15}$$

In the perfect case where $\eta_{\text{TE}} = \eta_{\text{TM}} = 1$ and $\Delta\phi = \pi$, and in the AGPM configuration case where $l_p = 2$ and thus $\theta = l_p\omega/2 = \omega$, where ω is the azimuthal polar coordinate, we obtain as output

$$\boldsymbol{R}^{\circlearrowright} = \begin{bmatrix} 0 \\ e^{i(2\omega - \pi/2)} \end{bmatrix}, \tag{A16}$$

$$\boldsymbol{L}^{\circlearrowleft} = \begin{bmatrix} e^{-i(2\omega + \pi/2)} \\ 0 \end{bmatrix}. \tag{A17}$$

Therefore, the two output polarization beams are orthogonal and decoupled in the helical basis.

APPENDIX B

PANCHARATNAM TOPOLOGICAL CHARGE

The so-called Pancharatnam phase has been introduced to measure the comparison of the phases of two light beams in different states of polarization. It is defined as the argument of the inner product of the two Jones vectors describing the two light beams to be phase compared,

$$\phi_p = \arg\langle E(\omega, r), E(0, r)\rangle. \tag{B1}$$

We can also define the associated topological charge of the beam, which is a nonsigned integer giving the number of times that the azimuthal angle rotates about the phase disclination (topological defect),

$$l_p = \frac{1}{2\pi} \oint \nabla\phi_p \, ds. \tag{B2}$$

In the AGPM case, $\phi_p = 2\omega$ ($l_p = 2$), which implies that the polarization state repeats itself $2l_p = 4$ times (see Fig. 12).

APPENDIX C

PERFECT REJECTION PROOF

Let us now analytically compute the pupil plane intensity distribution. The latter can be expressed as the Fourier transform of the product of the Airy disk function (a filled circular pupil is assumed) and the mask phase ramp. We have seen in § A2 that the mask phase in the helical basis takes the simple decoupled form $e^{i(l_p\omega - \pi/2)}$. Therefore, in the Fourier-plane polar coordinates (ρ, ψ) we have

$$A_{\text{pup}}(\rho, \psi, l_p) = \text{FT}\left[\frac{2J_1(2\pi R_{\text{tel}} r)}{2\pi R_{\text{tel}} r} e^{i(l_p\omega - \pi/2)}\right](\rho, \psi). \tag{C1}$$

Explicitly,

$$A_{\text{pup}}(\rho, \psi, l_p) = -i \int_0^\infty \int_0^{2\pi} \frac{2J_1(2\pi R_{\text{tel}} r)}{2\pi R_{\text{tel}} r} e^{i(l_p\omega)} e^{-2i\rho r \cos(\omega - \psi)} r \, dr \, d\omega, \tag{C2}$$

where we recognize the nth order Bessel function J_n. Indeed, various integrals can be expressed in terms of Bessel functions,

$$J_n(z) = \frac{1}{2\pi i^{-n}} \int_0^{2\pi} e^{iz\cos\phi} e^{in\phi}\, d\phi, \tag{C3}$$

and thus we have

$$A_{\rm pup}(\rho,\psi,l_p) = -i^{l_p-1}\frac{2e^{il_p\psi}}{R_{\rm tel}} \int_0^{\infty} J_1(2\pi R_{\rm tel} r) J_{l_p}(2\pi\rho r)\, dr. \tag{C4}$$

C1. AGPM CORONAGRAPH: $l_p = 2$

The previous result in the $l_p = 2$ case is the so-called Sonine's integral (Sneddon 1951, p. 55),

$$S = \int_0^{\infty} y^{1+\mu-\lambda} J_\lambda(ay) J_\mu(by)\, dy = \begin{cases} 0, & 0 < a < b, \\ \dfrac{b^\mu (a^2-b^2)^{\lambda-\mu-1}}{2^{\lambda-\mu-1} a^\lambda \Gamma(\lambda-\mu)}, & 0 < b < a. \end{cases} \tag{C5}$$

Thus, taking $l_p = 2$, we have

$$A_{\rm pup}(\rho,\psi,l_p=2) = \begin{cases} 0, & 0 < \rho < R_{\rm tel}, \\ \dfrac{e^{i2\psi}}{\pi\rho^2}, & 0 < R_{\rm tel} < \rho. \end{cases} \tag{C6}$$

We have demonstrated that in the perfect case for $l_p = 2$ (AGPM), the light is entirely rejected outside the geometric pupil area.

C2. GENERALIZATION TO l_pTH-ORDER VORTICES

Equation (C4) corresponds to the so-called Hankel transform of $l_{\rm P}$th order of the Bessel J_1 function. This transform has an analytical solution (Abramowitz & Stegun 1972, p. 487),

$$A_{\rm pup}(\rho,\psi,l_p) = -i^{1-l_p}\frac{2e^{il_p\psi}}{R_{\rm tel}} \begin{cases} (2\pi\rho)^{l_p}(2\pi R_{\rm tel})^{-l_p-1}\dfrac{\Gamma(1+l_p/2)}{\Gamma(l_p+1)\Gamma(1-l_p/2)}\, {}_2F_1\left(\dfrac{l_p+1}{2}, \dfrac{l_p}{2}; l_p+1; \dfrac{\rho^2}{R_{\rm tel}^2}\right), & 0 < \rho < R_{\rm tel}, \\ (2\pi\rho)^{-2}(2\pi R_{\rm tel})\dfrac{\Gamma(1+l_p/2)}{\Gamma(2)\Gamma(l_p/2)}\, {}_2F_1\left(\dfrac{l_p+1}{2}, \dfrac{2-l_p}{2}; 2; \dfrac{\rho^2}{R_{\rm tel}^2}\right), & \rho > R_{\rm tel}, \end{cases} \tag{C7}$$

where we recognize the gamma Γ and hypergeometric ${}_2F_1$ functions. This function shows perfect attenuation for even l_p values only,

$$A_{\rm pup}(\rho,\psi,l_p) = 0, \qquad \rho < R_{\rm tel} \text{ and } l_p = 2, 4, 6, \ldots. \tag{C8}$$

REFERENCES

Abramowitz, M., & Stegun, I. A. 1972, Handbook of Mathematical Functions (New York: Dover)
Baudoz, P., Boccaletti, A., Riaud, P., Cavarroc, C., Baudrand, J., Reess, J. M., & Rouan, D. 2005, PASP, submitted
Biener, G., Niv, A., Kleiner, V., & Hasman, E. 2002, Opt. Lett., 27, 1875
Boccaletti, A., Riaud, P., Baudoz, P., Baudrand, J., Rouan, D., Gratadour, D., Lacombe, F., & Lagrange, A.-M. 2004, PASP, 116, 1061
Boccaletti, A., Riand, P., Moutou, C., & Labeyrie, A. 2000, Icarus, 145, 628
Bomzon, Z., Niv, A., Biener, G., Kleiner, V., & Hasman, E. 2002, Appl. Opt., 41, 5218
Brillet, A., Vinet, J.-Y., Loriette, V., Mackowski, J.-M., Pinard, L., & Remillieux, A. 2003, Phys. Rev. D, 67, 102006
Gay, J., & Rabbia, Y. 1996, CR Acad. Sci. Paris, 332, 265
Gratadour, D., Rouan, D., Boccaletti, A., Riaud, P., & Clénet, Y. 2005, A&A, 429, 433
Karlsson, M., Hjort, K., & Nikolajeff, F. 2001, Opt. Lett., 26, 1752
Karlsson, M., & Nikolajeff, F. 2003, Opt. Express, 11, 502
Kikuta, H., Ohira, Y., & Iwata, K. 1997, Appl. Opt., 36, 1566
Labeyrie, A. 1996, A&AS, 118, 517
Lalanne, P., Pichon, P., Chavel, P., Cambril, E., & Launois, H. 1999, Appl. Opt., 38, 4980
Levy, U., Tsai, C.-H., Pang, L., & Fainman, Y. 2004, Opt. Lett., 29, 1718
Lyot, B. 1939, MNRAS, 99, 580
Mackowski, J. M., Pinard, L., Dognin, L., Ganau, P., Lagrange, B., Michel, C., & Morgue M. 1999, Opt. Quantum Electron., 31, 507
Mawet, D., et al. 2005, Appl. Opt., in press
Moharam, M. G., & Gaylord, T. K. 1981, J. Opt. Soc. Am., 71, 811
Mouillet, D., Fusco, T., Lagrange, A.-M., & Beuzit, J.-L. 2003, in Astronomy with High Contrast Imaging, ed. C. Aime & R. Soummer (Les Ulis: EDP Sciences), 193
Niv, A., Biener, G., Kleiner, V., & Hasman, E. 2003, Opt. Lett., 28, 510
Nordin, G. P., & Deguzman, P. C. 1999, Opt. Express, 5, 163
Riaud, P., Boccaletti, A., Baudrand, J., & Rouan, D. 2003, PASP, 115, 712
Riaud, P., Boccaletti, A., Rouan, D., Lemarquis, F., & Labeyrie, A. 2001, PASP, 113, 1145
Riaud, P., et al. 2002, A&A, 396, 345
Roddier, F., & Roddier, C. 1997, PASP, 109, 815
Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., & Labeyrie, A. 2000, PASP, 112, 1479
Rouan, D., et al. 2004, A&A, 417, L1
Smith, B. A., & Terrile, R. 1984, Science, 226, 1421
Sneddon, I. N. 1951, Fourier Transforms (New York: McGraw-Hill)
Soummer, R., Dohlen, K., & Aime, C. 2003, A&A, 403, 369
Swartzlander, G. A., Jr. 2001, Opt. Lett., 26, 497

6.4 AGPM coronagraphs onboard SEE-COAST ?

This section is devoted to the presentation of the SEE-COAST project. It has to be taken as an informative preliminary presentation illustrating a potential application of the AGPM coronagraph. The scientific goals of the mission are only superficially addressed, and the design considerations must be taken as rough drafts.

6.4.1 Mission philosophy

SEE-COAST stands for Super-Earth Explorer – Coronagraphic Off-Axis Space Telescope. It is a joint proposal of the University of Liège (IAGL and CSL), the Paris-Meudon Observatory (LUTH and LESIA), OHP, UNSA (Nice), ETH Zurich, LAOG, Geneva, Amsterdam and Torun in response to an ESA call for proposals for a small-medium mission (budget of $\sim$ 350 Meuros) in the framework of the Cosmic Vision program. SEE-COAST is a medium-class (1.5-2 meter) ultraviolet-visible space telescope (Fig. 6.3) intended at comparative exoplanetology. SEE-COAST will be optimized for high-contrast imaging. Its primary mirror and the downstream optical train should provide a very low wavefront error of $\lambda/100$ rms ($\lambda = 632.8$ nm) together with an off-axis implementation. The goal is to detect and characterize extrasolar planets thanks to low/medium-resolution spectroscopy in the visible.

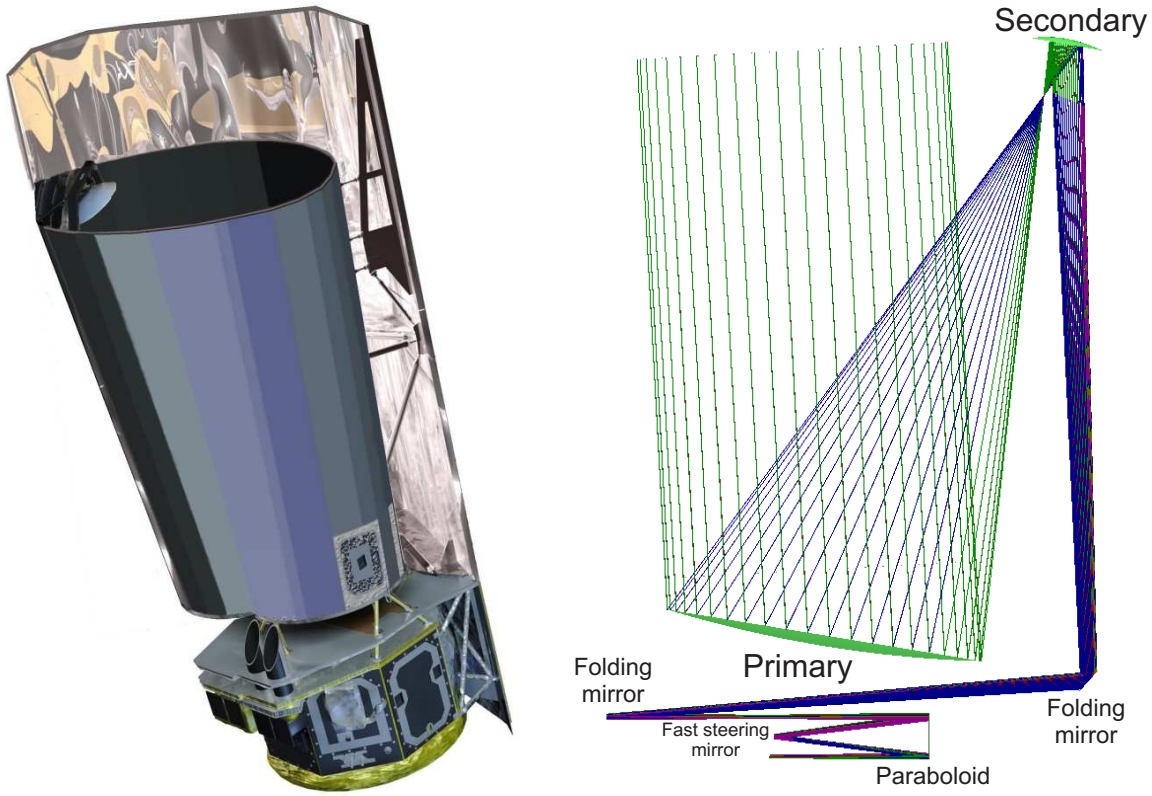

Figure 6.3: SEE-COAST satellite concept. Left: artistic view of the sun-shielded off-axis visible 1.5-meter telescope with a 0.2-meter secondary. Right: preliminary optical concept with fast-steering mirror for tip-tilt corrections. Courtesy of Pierre Riaud.

6.4.2 Science case overview

The science case of a very good space telescope working in the visible is extremely rich. The Hubble heritage can not suffer any comparison, for example. However, the main purpose of SEE-COAST is to do comparative exoplanetology. Lessons from the last ten years of exoplanet hunting are that this new field is certainly a great source of positive surprises. The kind of surprises that make science progresses. However, the science case of SEE-COAST is very precise since it will be conceived as a precursor mission to more ambitious and long-term projects like Darwin/TPF. As such, SEE-COAST must prepare the work of its successor for example by sampling the exozodiacal disks of Darwin's potential targets for better characterization down to a few zodis. Indeed, as already stressed out in Sect. 1.3.2, exozodiacal light mixes with the faint planetary signal and might prevent its detection. Its knowledge is therefore mandatory to avoid the potential stellar sources where it is too bright for the modulation techniques to be efficient (Absil et al. 2006).

The characteristics of SEE-COAST will also be unique for exoplanet hunting. High contrast and high angular resolution imaging will be provided by its exceptional optical quality and thanks to specific advanced coronagraphic instruments. SEE-COAST should be able to image dozens of Jupiter-like planets in the 1-5 AU range around nearby stars up to 15 pc and provide the first direct measurements of a broad range of fundamental characteristics. These characteristics include orbital inclination, mass, brightness, orbital and/or rotational variabilities, etc. Low-resolution spectroscopic ($R_\lambda \approx 15$) capabilities will permit a first characterization of their atmosphere in the visible, searching for H_2O, CH_4 and NH_3 features among others. In this respect, SEE-COAST will be complementary to JWST/MIRI (mid-infrared) and VLT-PF/SPHERE (near-infrared).

These observations will be of primarily importance for the comprehension of planetary formation mechanisms but also will answer many questions about extrasolar planets: are zodiacal dust disks rare or common, and how are they related to the presence of planets? What is the frequency of planets that are inaccessible to radial velocity survey? Do stars with known radial velocity detected planets also host Jupiter-like planets at large semimajor axes? Do ice worlds, ocean planets or ringed planets exist?

But the most exciting goal of SEE-COAST is the potential discovery of **one** super-Earth ($\sim 3R_\oplus$) around another star, preferably in the habitable zone. Such a discovery would definitely revolutionize our perception of the world and lead to unpredictable consequences in numerous fields of science but with one particular community that will be overwhelmed: the astrobiologists. Of course, the feasibility of such a putative goal is still to be technologically demonstrated because of the extremely stringent optical constraints that have *a priori* to be tackled.

Additional science can be conducted by extending the imaging capabilities of SEE-COAST towards the ultraviolet regime (100-300 nm). For example, the angular resolution of 20 mas in the $Ly\alpha$ emission line would allow imaging stellar forming regions in nearby galaxies, but also wind interactions in cataclysmic variables, OB associations and Wolf-Rayet stars, leading to new insights about magnetic field-plasma processes. In extragalactic astrophysics, such a telescope would allow imaging the torus and accretion zones of nearby AGN ($z < 2 \approx 750$ pc), constraining the three-dimensional geometry of the narrow line regions (NLRs) and the nuclear disks of the host galaxies in order to probe the connection between the accreting matter, the outflowing gas, and the ionizing radiation. Moreover, while broad emission line profiles are a characteristic signature of the energetic processes in the center of active galaxies, the spectra of some systems also show absorption lines. Such absorption features are generally found against the ultraviolet continuum. Whereas the properties of the emitting clouds are reasonably well understood, less is known about the absorbing gas, especially how it is accelerated to the high velocities observed. The main problem in understanding such systems is that the background continuum necessary for

absorption in active galaxies is often present only at the nucleus and at distributed, compact hot spots. A space telescope with coronagraphic capabilities would permit the study of active galaxies with much better spatial resolution and at UV wavelengths. Such a study of selected nearby systems could clarify both the background against which the absorbers are seen and the connection between the emitting and the absorbing gas.

6.4.3 Optical concept

The entire optical concept of SEE-COAST is dedicated and optimized to high contrast imaging. For that, extreme optical qualities are mandatory, with resultant wavefront errors of $\lambda/100$ rms and an excellent stability (10 hours - 3 days). Phase-mask coronagraphs indeed require non-obscured perfectly smooth pupils to provide their full potential of starlight rejection (Riaud et al. 2001). This is the main reason for requiring an off-axis telescope (see Fig. 6.3).

6.4.4 Coronagraphic instrument

The priority instrument is the high resolution coronagraphic camera (HRCC), which is a narrow-field ($\sim 27'' \times 27''$) high dynamic range instrument optimized for providing the envisaged detection capabilities for exoplanet imaging. This camera shall be equipped with modern coronagraphs like the FQPM or AGPM (Fig. 6.4). Let us derive some very preliminary results assuming the use of optimized AGPM coronagraphs and a detection scheme analog to the VLT-PF/SPHERE one, i.e., coronagraphy with spectral differential imaging (SDI). Let us assume that the HRCC comprises three main filters spanning the 650-760, 750-860 and 850-960 nm wavelength ranges. Each of them is divided into three $R_\lambda = 15$ bands. This makes 9 simultaneous images that encompass the 2048×2048 CCD detector, and should allow efficient SDI speckle subtractions. Preliminary dimensioning of the three AGPM is illustrated in Fig. 6.5 together with cross subtraction profiles. The raw nulling provided by the different coronagraph is at the $\sim 5 \times 10^{-6}$ level. Spectral

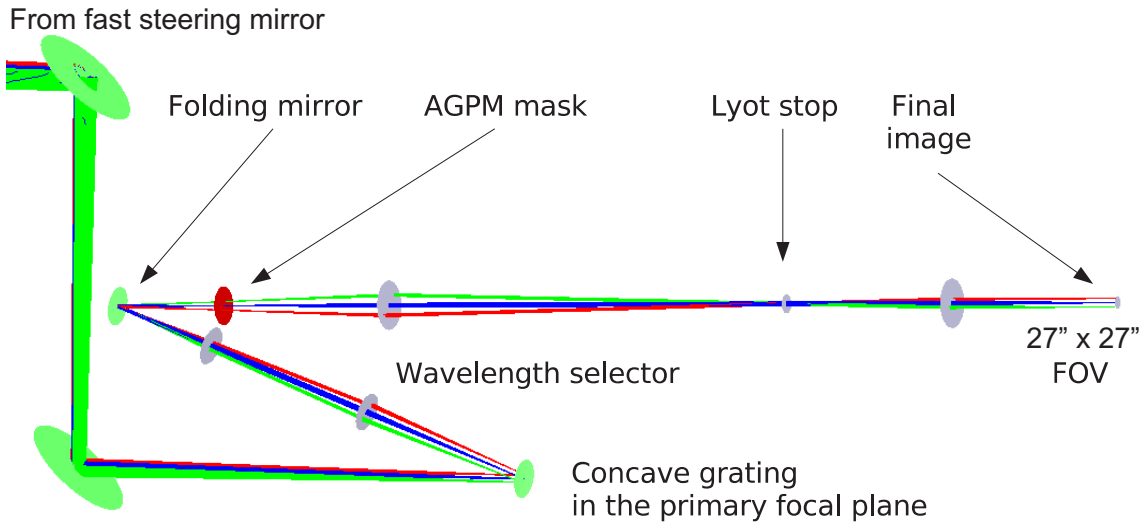

Figure 6.4: High resolution coronagraphic camera (HRCC) concept (rough draft). The wavelength selection is done thanks to the proper combination of filters and gratings. It is worth noting that a grating placed in a focal plane will disperse the light into several pupils of different colors which can then be selected by well-dimensioned and well-positioned diaphragms. Courtesy of Pierre Riaud.

subtraction assuming residual WFE of $\lambda/100$ rms ($\lambda = 632.8$ nm) allows gaining a factor from 10 to 100, leading to a residual level at $2\lambda/D$ from $\sim 2 \times 10^{-7}$ to $\sim 2 \times 10^{-8}$, respectively.

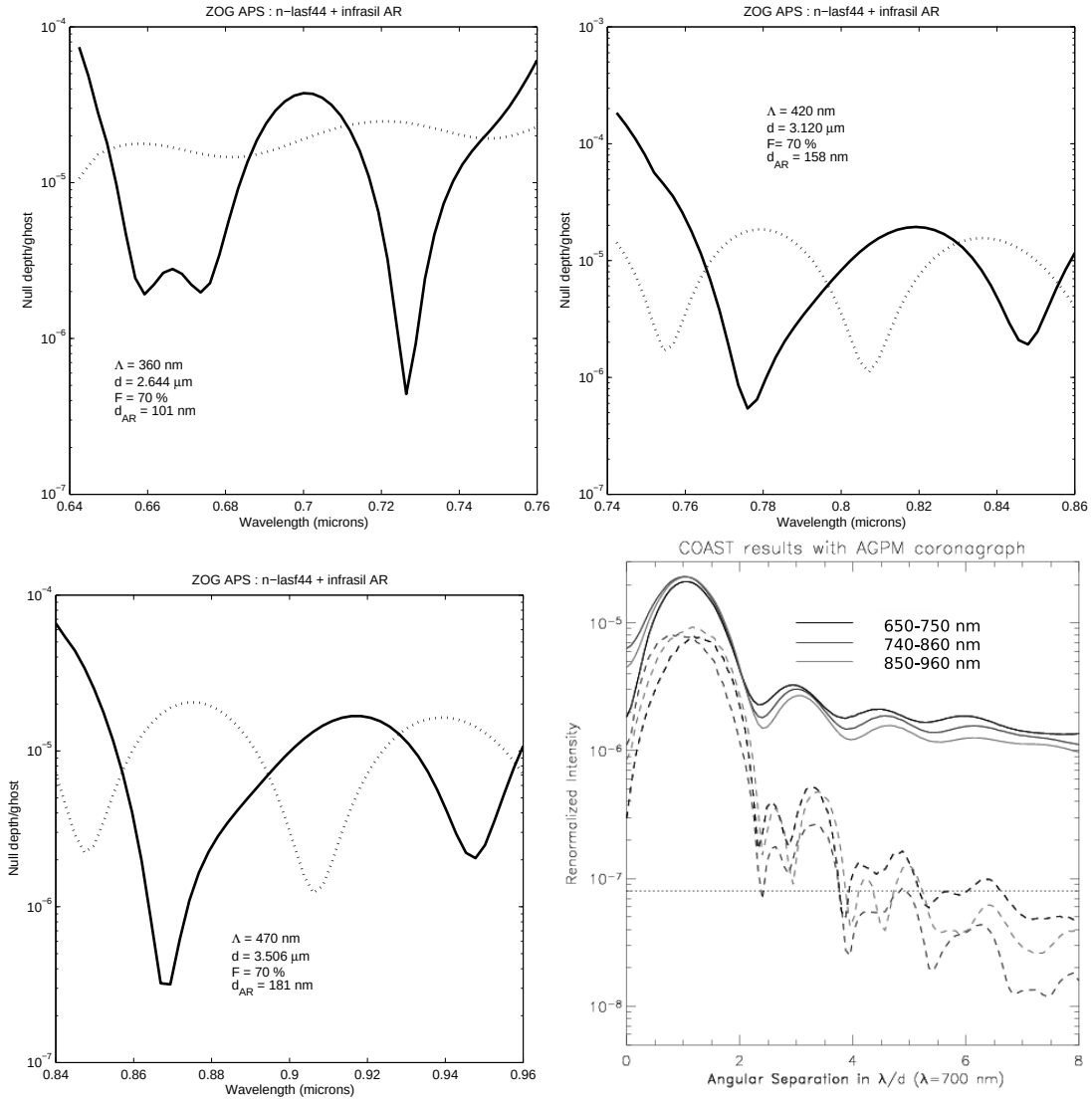

Figure 6.5: Optimized AGPM for SEE-COAST. Top: 640-760 nm (left) and 740-860 nm (right) wavelength range null depth. Bottom left: 840-960 nm wavelength range null depth. (Infrasil is the amorphous form of SiO_2). Bottom right: numerical simulation assuming residual WFE of $\lambda/100$ rms ($\lambda = 632.8$ nm) and phase-mask imperfections. The dashed lines show the residual levels of cross subtractions between filters.

Proper detection capability assessment necessitates a complex simulator (COASTSIM) under evaluation. The latter should assess the SDI subtraction real gain with respect to speckle residue removals. The simulator should provide the detection capability of a certain scientific target under well defined conditions (stellar type, proximity, contrast, etc.) and accordingly the optical train error budget (WFE, jitter, etc.).

6.4.5 Feasibility

To conclude this introductory section about the SEE-COAST project, let us discuss briefly some technological feasibility issues.

Feasibility of the telescope

Preliminary assessments concluded that the required $\lambda/200$ rms ($\lambda = 632.8$ nm) surfacing error should be feasible assuming that a specific measurement method is developed that would ensure an accuracy of $\lambda/400$ rms ($\lambda = 632.8$ nm). A secondary active mirror or dedicated fast-steering mirror should also be considered for active compensation of centering (tip-tilt) and defocus errors. Industrial partners (Alcatel and Astrium) are being consulted on these matters. The question of feasibility is of course more a question of budget.

Feasibility of the coronagraphic masks

Mask feasibility with subwavelength grating is questionable since the periods required range from 360 nm to 470 nm, assuming a n-LASf44 schott glass (high -index glass: $n \approx 1.8$). The subsequent thicknesses range from $2.5\,\mu$m to $3.5\,\mu$m, inducing therefore aspect ratios of ~ 10 (assuming a filling factor of 70 %). Such aspect ratios are within reach of the current technologies. The obstacle concerns more the substrate material which must be compatible with plasma-etching recipes. Unfortunately, visible transparent high-index glasses like n-LASf44 (Schott) are composed of a lot of impurities that can prevent a proper etching by contaminating the vacuum chambers. Therefore, a R&D program assessing these matters is absolutely necessary. It is worth noting that the experience with VLT-PF/SPHERE from the ground and the JWST from space should provide valuable information. First efforts in this direction will be presented in the following chapter.

7

Manufacturing of 4QZOG and AGPM coronagraphs

Contents

Abstract. *This chapter is devoted to the presentation of the manufacturing of 4QZOG and AGPM coronagraph prototypes that has been initiated at CEA-LETI with the collaboration of LESIA and LAOG in the framework of the R&D for VLT-PF/SPHERE. The fabrication philosophy is exposed as well as a tolerance study assessing the feasibility and constraints for each step of LETI's specific technological pileup.*

7.1 LETI operation

7.1.1 Context of the operation

We have decided, in collaboration with our colleagues of LESIA and LAOG[19] to initiate the realization of prototypes of 4QZOG and AGPM by CEA-LETI [20]. LETI is the ideal technological partner. Apart from their mastership of state-of-the-art technologies of micro-electronics, they indeed possess a good knowledge of the constraints of instrumentation for astrophysical applications since they have already worked on integrated optics for interferometry (Fig. 7.1). Let us quote

[19] Laboratoire d'Astrophysique de l'Observatoire de Grenoble.

[20] Based in Grenoble, LETI (laboratoire d'électronique de technologie de l'information) is a laboratory operated by the Technology Research Directorate (DRT) of the French Atomic Energy Commission (CEA). It is one of the largest applied research laboratories in electronics in Europe.

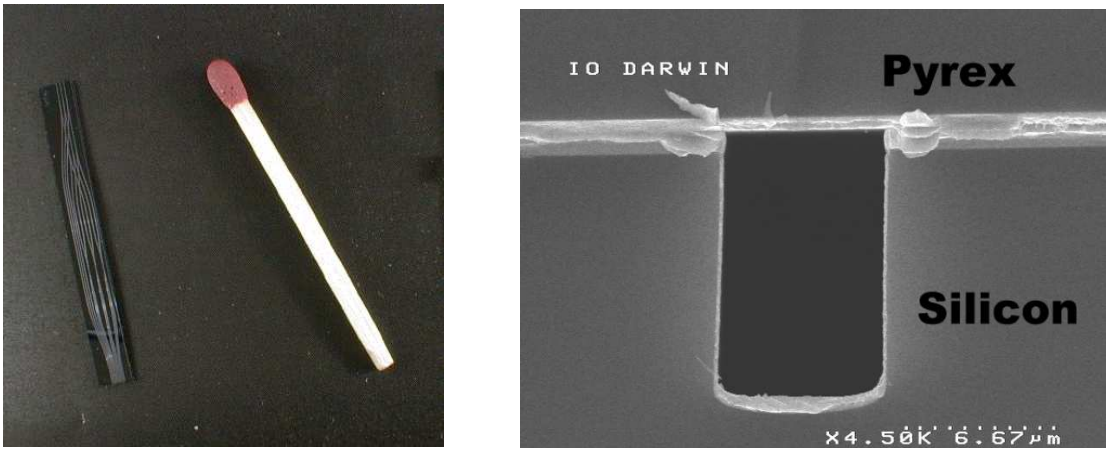

Figure 7.1: CEA-LETI products for astrophysical applications. Left picture: photograph of one prototype of integrated optics component together with a match to give the spatial scale (Malbet et al. 1999). The component features a 3-way beam combiner with photometric calibration channels manufactured with the silicon etching technology. Right picture: photograph of a hollow metallic waveguide input using a scanning electron microscope. The Pyrex cover is maintained by anodic bonding on the silicon substrate. The etching depth is 10 μm. The gold deposition is thicker at the bottom of the waveguide than along the lateral walls (Labadie et al. 2006).

the IONIC beam combiner (Monnier et al. 2004, integrated optics near-infrared interferometric camera), successfully implemented at the IOTA interferometer (Mount Hopkins, Arizona), the IONIC-2TK beam combiner (Lebouquin et al. 2006) which is the current beam combiner of the VINCI instrument (VLTI), the VITRUV/VSI (VLTI Spectro-Imager) second-generation instrument for the VLTI (Malbet et al. 2004) and the studies for modal filtering and beam combination in the thermal infrared for the Darwin mission (Labadie et al. 2005, 2006).

7.1.2 Goals

The purpose of this operation fits in the framework of the development of achromatic coronagraphs for the VLT-planet finder (SPHERE) second-generation instrument for the detection and characterization of warm and young extrasolar planets (see Sect. 1.5.3). The goal is to experimentally validate the concepts of subwavelength grating achromatization and implementation in original phase coronagraphs such as the 4QZOG and AGPM, while relying on state-of-the-art standard micro-electronics manufacturing technologies, such as silicon lithography. The already mentioned flexibility of subwavelength gratings to implement achromatic waveplates will be exploited by designing and optimizing the components according to a trade-off between the technological pileup proposed by LETI and the required performances for SPHERE, while minimizing failure risks.

7.2 Design of the subwavelength grating

The first design trade-off concerns the wavelength range. On one hand, VLT-PF coronagraphic cameras will encompass wavelengths from band I (0.9 μm) to K (2.2 μm). On the other hand, the silicon electronic absorption loweredge starts at 1.2 micron. For this reason, we have chosen to design the subwavelength grating for optimal operations simultaneously in H (1.5-1.8 microns) and K bands (2-2.3 microns). Doing this, we will demonstrate the feasibility and applicability

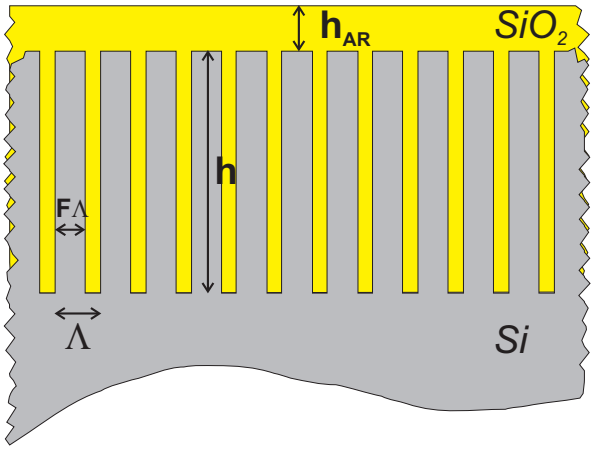

Figure 7.2: Scheme of the subwavelength grating structure resulting from a trade-off between LETI technological pileup constraints and performances.

of the subwavelength grating technology for phase coronagraphs while simplifying their potential implementation in the instrument by regrouping the two bands on a single component.

7.2.1 Technological pileup definition

The proprietary and confidential technological pileup definition, i.e., the definition of each elementary technological step[21] leading to the final component, proposed by the LETI and submitted to our approbation led to the structure presented in Fig. 7.2. It consists in a silicon substrate imprinted with the subwavelength grating made of silicon filled with silica. The whole structure is encapsulated with a silica thin layer acting both as a protection and an anti-reflective coating.

7.2.2 Grating optimization

Knowing the approximate structure definition, optimization procedures can be undertaken to find the best grating parameters (period, filling factor, depth, thickness of the AR layer) to minimize the null depth simultaneously in H and K bands. Optimization was performed using the simplex-search method (see Sect. 5.1.2) coupled to the RCWA code with the following calculation parameters: 21 orders of diffraction retained, 100 points in wavelength. The optimization led to the grating geometry parameters presented in Table 7.1, and the results displayed in terms of null depth and parasitic residual ghost level in Fig. 7.3.

The subwavelength grating technology once again demonstrates its flexibility by allowing simultaneous operation in two well-separated spectral bands. Indeed, the theoretical rejection factor in the H band is 1500 and 3500 in the K band, while being always above 500 from 1.5 μm to 2.35 μm.

As already stressed out in the paper presented in Sect. 5.2, Fresnel parasitic reflections, while diminishing the optical efficiency, can prevent the null depth from reaching very low levels because of the intensity mismatch due to the differential effects between the polarization components TE and TM (see Eq. 2.8). Indeed, the difference between the effective indices n_{eff}^{TE} and n_{eff}^{TM} at the origin of the form birefringence, also induces unequal TE/TM Fresnel reflection coefficients at

[21] Preliminary estimations by CEA-LETI has led to more than twenty process steps in the present case.

Table 7.1: Optimized parameters of the H/K band Si/SiO_2 subwavelength grating achromatic phase retarder.

Parameters	Value
Grating period Λ	0.402 μm
Grating depth/thickness h	2.1973 μm
Grating filling Factor F	80%
SiO_2 AR layer thickness h_{AR}	280 nm

the grating top and bottom interfaces. The SiO_2 AR layer equilibrates the differential effect, but only to a certain point.

Moreover, the Fresnel parasitic reflections are also responsible for a second-order so-called "ghost" in the final coronagraphic image. This ghost comes from the double reflection on the back face of the substrate and on the internal interface of the AR treated ZOG. Its intensity can be quantified assuming a reasonable AR performance for the back side of the component, i.e., about 0.5% of reflectivity. The intensity level of the ghost for the structure of Table 7.1 in Fig. 7.3 is maintained below 10^{-3} but better performances can be achieved by improving the back face anti-reflective efficiency. The substrate can also be tilted, or polished with a slight wedge in order to dump the stray light.

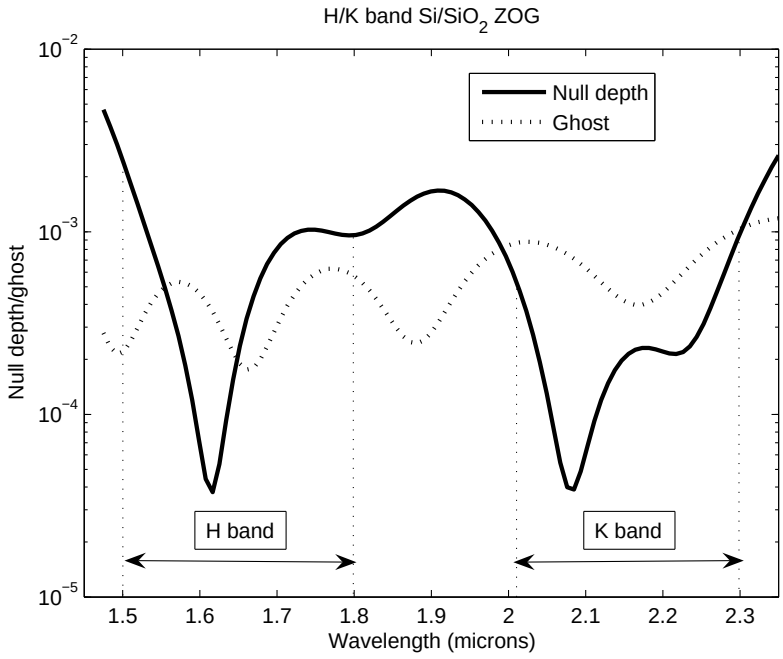

Figure 7.3: Performance (null depth/ghost) of the optimal subwavelength grating based on the LETI technological pileup. The rejection factor (inverse of the null depth) is $\sim$ 1500 in the H band and $\sim$ 3500 in the K band.

7.3 Tolerancing and manufacturing philosophy

Sensitivity to each geometrical parameter such as the period, filling factor, grating and AR layer thicknesses can be drawn. However, such a raw calculus would be useless and far from manufacturing reality. Let us take a different approach, based on the actual chosen process. Without going into details for intellectual property reasons, we shall nonetheless describe the philosophy we have selected according to the LETI know-how and suggestions.

Based on micro-electronic technologies, the components will be manufactured on silicon wafers. The size that has been chosen for the latter is 8 inches ($\sim$ 20 cm). The photolithography step where the primary mask shall be defined, will be carried out by a "micro-stepper" machine. This machine uses the projection-printing principle as presented in Sect. 3.4.1. The working wavelength of the CEA-LETI micro-stepper that will be used is $\lambda = 248$ nm and the reduction factor of the projection $\sim$ 4-5. The field of view (FOV) of the "stepper" is 22 mm $\times$ 22 mm. This means that on the entire wafer, 44 patterns can be imprinted. We have decided to spread the 4QZOG and AGPM over the wafer and to allocate the same number of patterns to both of them, i.e., 22 4QZOG and 22 AGPM (see Fig. 7.4). It is worth noting that the subwavelength grating micro-geometry is exactly the same for the 4QZOG and AGPM components. The only difference between them concerns the macroscopic implementation of the micro-structure: a four-quadrant anti-symmetrical implementation for the 4QZOG phase masks (see Chapter 5) and a circularly concentric one for the AGPM coronagraphs (see Chapter 6).

We have then decided to take advantage of the substantial size of the stepper field of view by imprinting 9 (3 by 3) patterns inside each of them. Each elementary pattern is therefore about 7 mm $\times$ 7 mm. The purpose of imprinting 9 different patterns in every FOV is to minimize the risk of failure by putting the solution into a robust frame. Indeed, having 9 patterns by stepper FOV gives us the possibility to scan the parameter space. It is to be noted that only parameters defining the lateral dimensions of the structure (period and filling factor) can be scanned that way since the stepper defines the lithographic mask before any etching process takes place. The

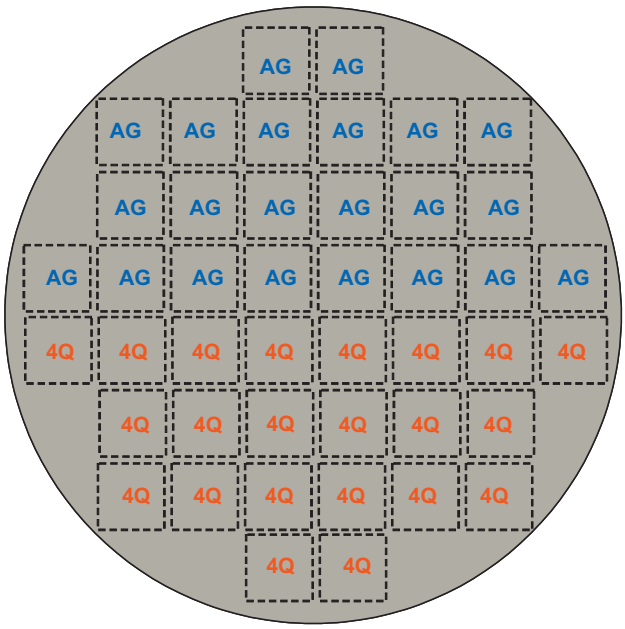

Figure 7.4: Allocation of the 22 4QZOG and 22 AGPM micro-stepper patterns on the 8-inch silicon wafer. Each pattern is 22 mm by 22 mm in size, corresponding to the micro-stepper field of view.

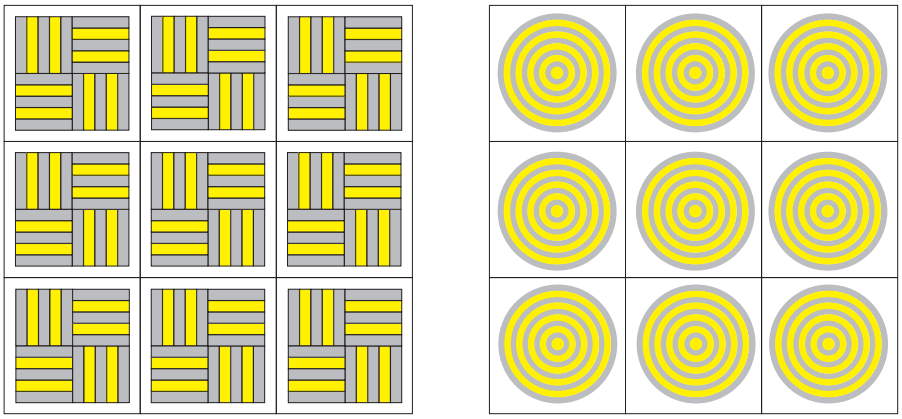

Figure 7.5: Micro-stepper 22 mm by 22 mm field of view for the 4QZOG (left) and AGPM (right) patterns.

reason for being so precautious is twofold:

- thickness precision and uniformity across the wafer cannot be guaranteed by the etching control better than a few tens of nanometers;
- the profile that is likely to emerge from the LETI confidential technology is a W-shaped profile as illustrated in Fig. 7.6. This departure from nominal configuration, quantified by the grating-wall slope angle α is very penalizing as we will discuss.

The nine patterns have now to be optimally determined. Referring to Fig. 7.6, let us define the silicon (resp. silica) so-called "feature line" d_2 (resp. d_1) as the product of the grating period Λ by the filling factor F (resp. $1-F$), i.e., $F\Lambda$ (resp. $(1-F)\Lambda$). The silica feature line d_1 has to be fixed once and for all since, for technological reasons, it will be common to all individual patterns in the stepper FOV and across the whole wafer as well.

Let us take the optimal silica feature line value indirectly given in Table 7.1, i.e., $d_1 = (1 - 0.8) * 0.402 = 0.08\,\mu$m, or 80 nm. Nine values for d_2 must now be chosen in order to optimally frame the solution assuming that the grating-wall slope α is likely to vary between 87.5 and 90 degrees. Indeed, numerical simulations show that below 87.5°, resonance phenomenons affect the results by a too strong amount, especially in the H band. This behavior is easily explainable since the slope associated with the W shape gets the lower part of the grating structure to exit the subwavelength domain, logically beginning by the shortest wavelengths (H band). Moreover, a double periodicity appears, enhancing overlapping effects. For these reasons, the specification on the grating slope is the "best effort towards 90 degrees".

The nine values for d_2 are then chosen to range from 250 nm to 330 nm with a 10-nm step. Doing this, a satisfactory solution exists for all assumed slopes, as can be verified in Fig. 7.7, Fig. 7.8, Fig. 7.9, Fig. 7.10, Fig. 7.11 and Fig. 7.12. Each of these five panels contains four graphs in the "landscape" disposition:

1. top left, there is shown the best rejection ratio in H and K bands versus the silicon feature line;
2. top right, this graph shows the average transmittance between the orthogonal polarization states in H and K bands;
3. at bottom left is displayed the optimal adjustment of the grating thickness versus the feature line;
4. at bottom right is represented the null depth of the best solution for the current grating-wall slope angle versus wavelength.

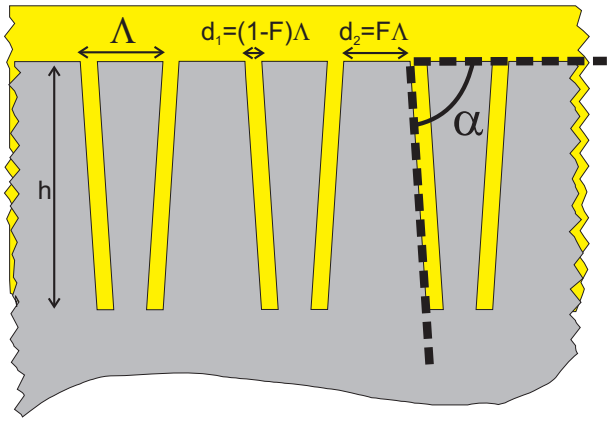

Figure 7.6: W-shaped profile likely to emerge from the LETI proprietary process pileup. The new parameter to be taken into account is the angle α.

Several conclusions can be drawn and/or confirmed from this analysis: in Fig. 7.7, i.e., for a 87.5-degree slope angle, the degradation of the performance in the H band is clearly visible while in Fig. 7.12, the 90-degree slope-angle solution appears to be the best one. In general, there is a strong tendency for the optimal feature line to increase with the slope angle. It is also the case for the optimal thickness adjustment. Indeed, the optimal thickness is proportional to the feature line. It is to be noted that the anti-reflective and protective layer of SiO_2 was assumed to be constant and fixed at an optimal $\lambda/4n$ value, i.e., 280 nm in all cases ($\lambda \approx 1.7\,\mu$m, $n \approx 1.5$).

To finally conclude this section, let us summary the design tolerances in Table 7.2.

Table 7.2: Subwavelength grating tolerancing conclusions and expected performances.

Slope angle (°)	Rejection in H	Rejection in K	Feature line	Thickness
87.5	100	1500	250 nm	2.1 μm
88	700	1700	250 nm	1.9 μm
88.5	500	600	250 nm	1.8 μm
89	500	700	270 nm	1.8 μm
89.5	600	2500	310 nm	2.1 μm
90	1500	3000	320 nm	2.2 μm

7.4 Selection and tests

As the process will deliver 44×9 coronagraphs per wafer, a huge number of components will have to be sorted in order to select the best ones. A coronagraphic test is time consuming (because of alignments, long integration times, etc.) so that we had to think about another procedure to efficiently perform the selection. It will consist in polarimetric measurements (see Sect. 9.3.2 for more details about these methods) of the vectorial *TE-TM* phase-shift quality provided by the subwavelength gratings, at certain laser wavelengths like the widespread telecom wavelengths 1.55 μm and 1.62 μm, for instance. The optical implementation of these tests is very simple since it only requires classical on-the-shelf near-IR optical components like diodes, polarizers, detectors, etc. (see Sect. 9.3.2). The measurements should be very fast, and lead to the selection of the best components for the coronagraphic tests on the VLT-PF/SPHERE H/K-band coronagraphic testbed (see Fig. 2.8 in Sect. 2.3).

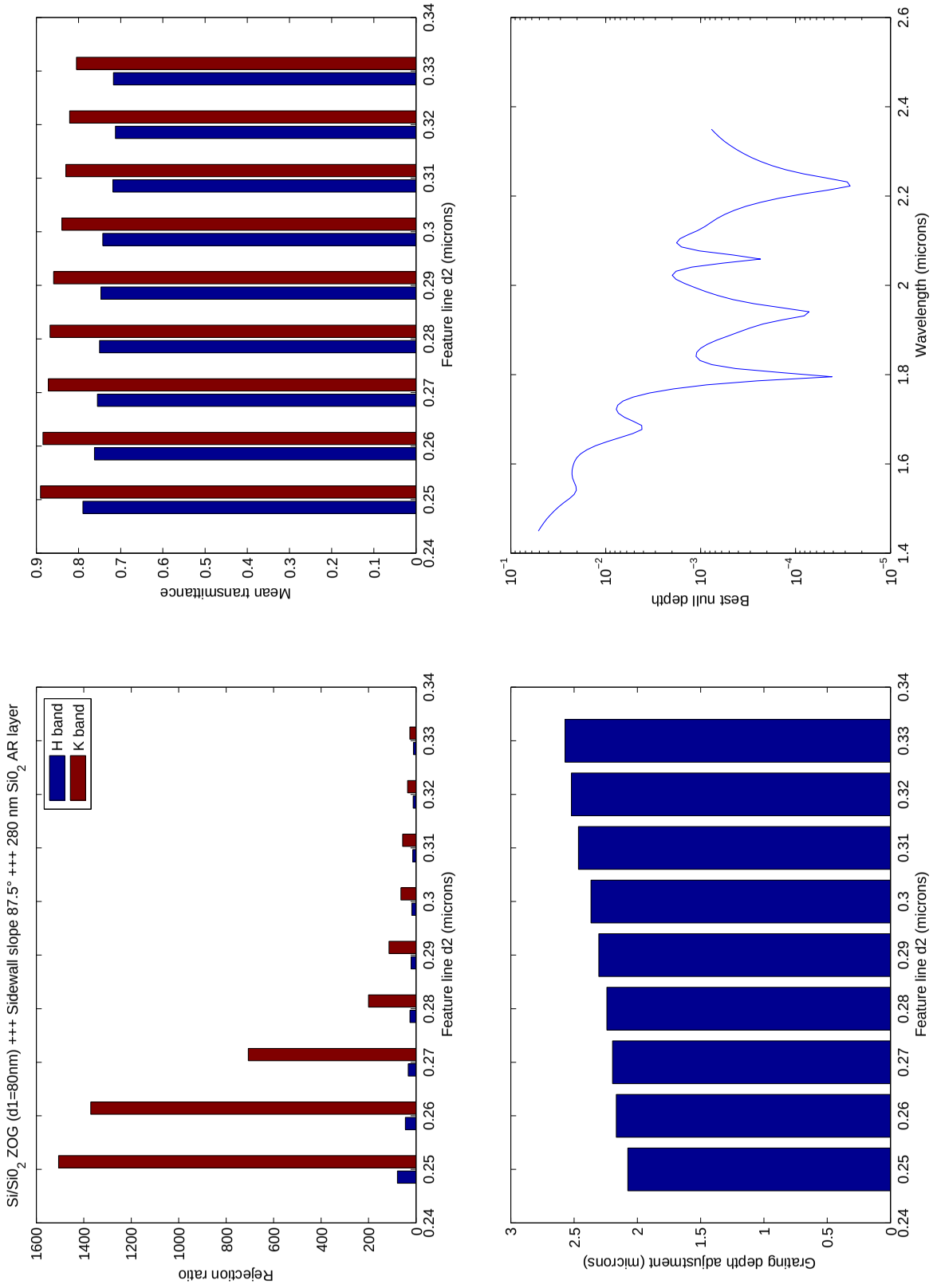

Figure 7.7: Panel for the 87.5° grating-wall slope angle.

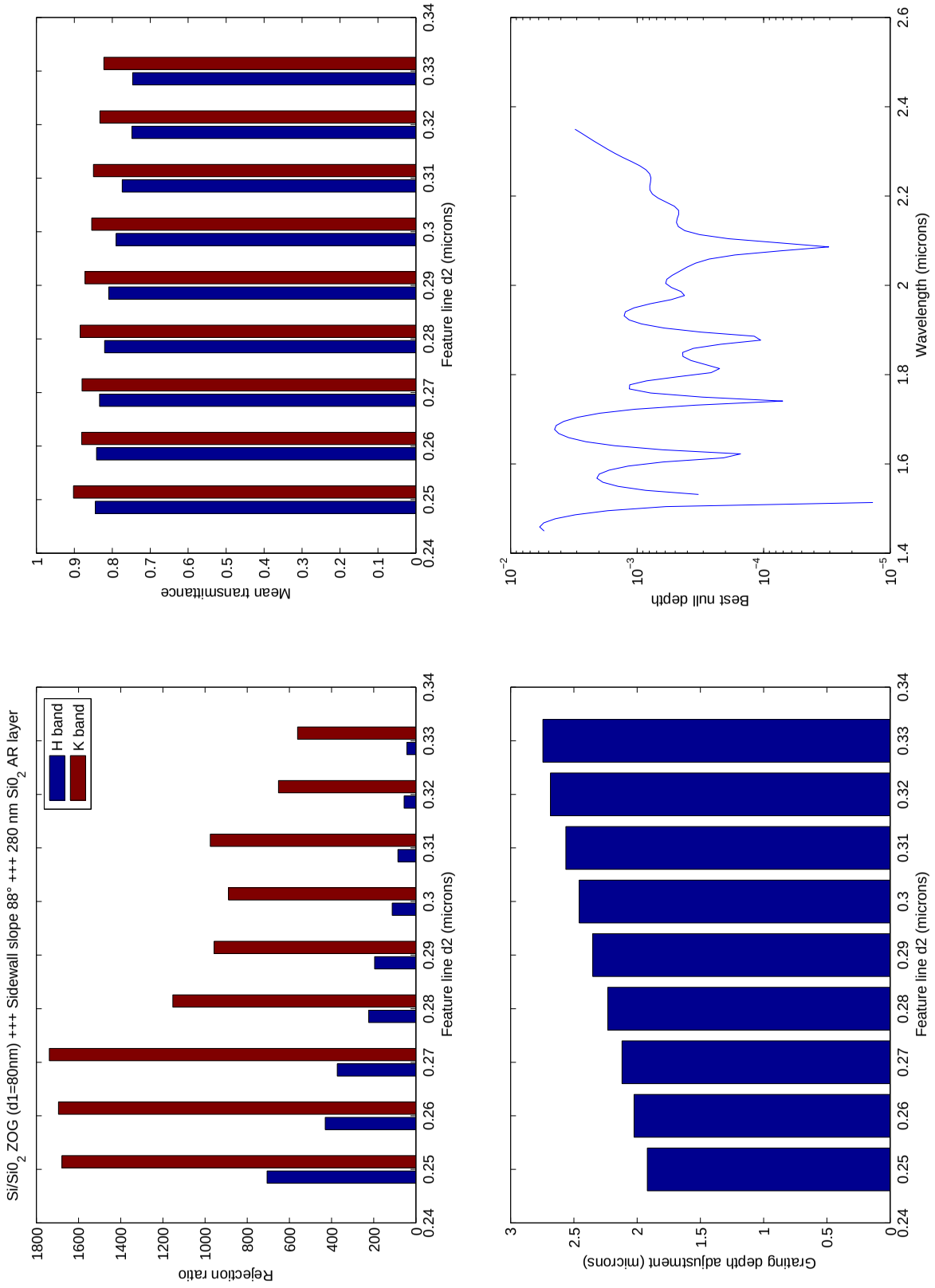

Figure 7.8: Panel for the 88° grating-wall slope angle.

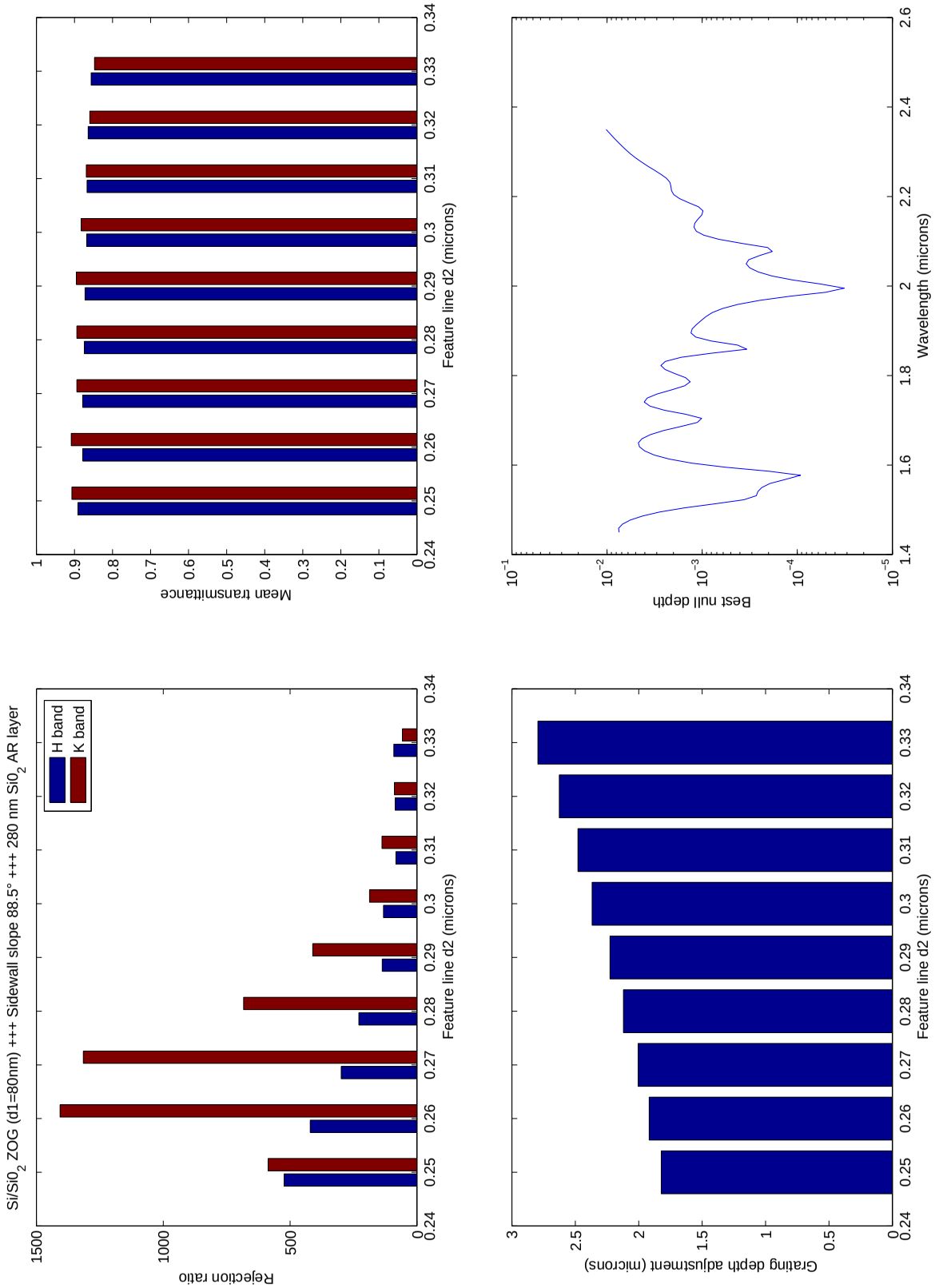

Figure 7.9: Panel for the 88.5° grating-wall slope angle.

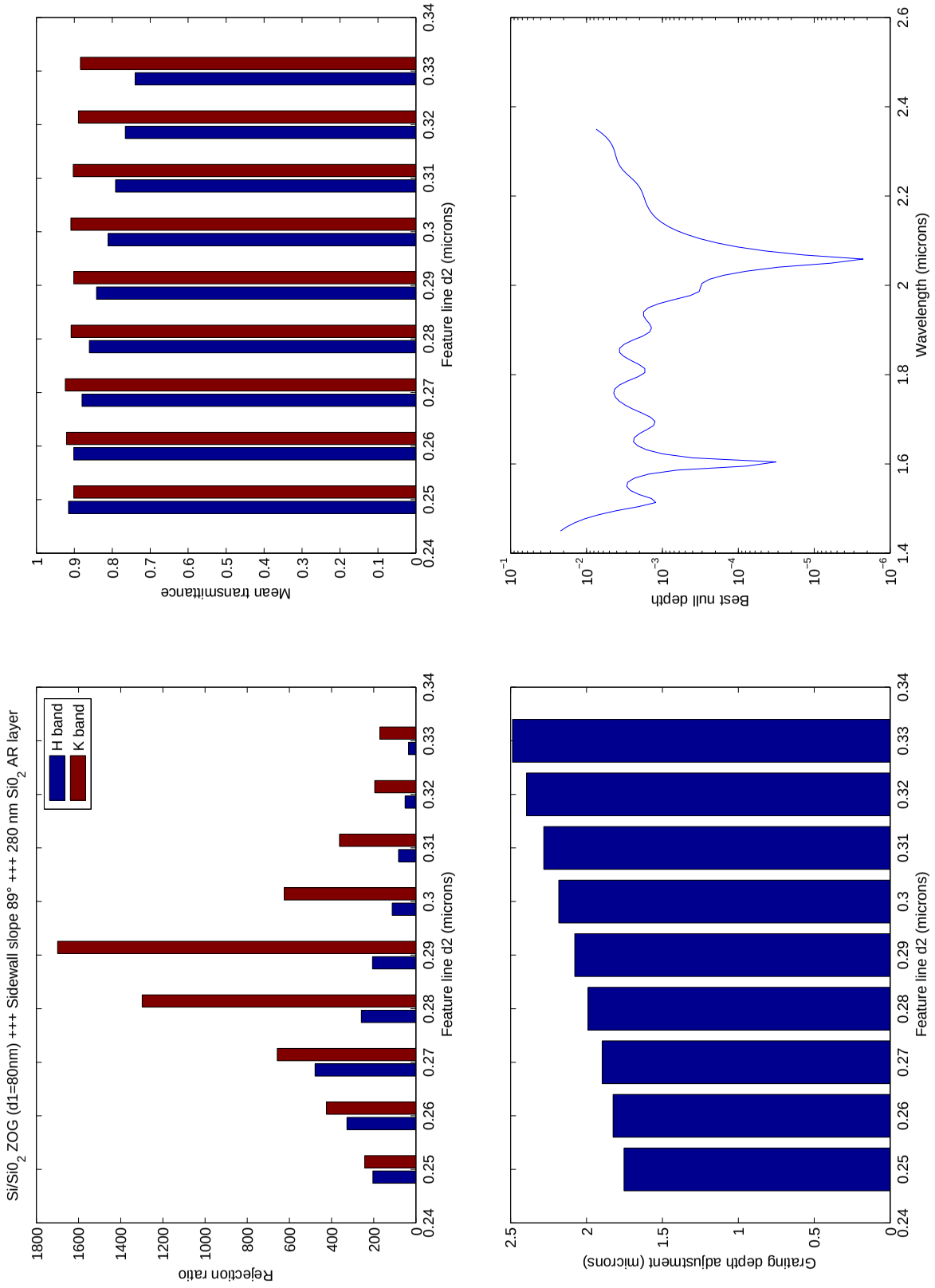

Figure 7.10: Panel for the 89° grating-wall slope angle.

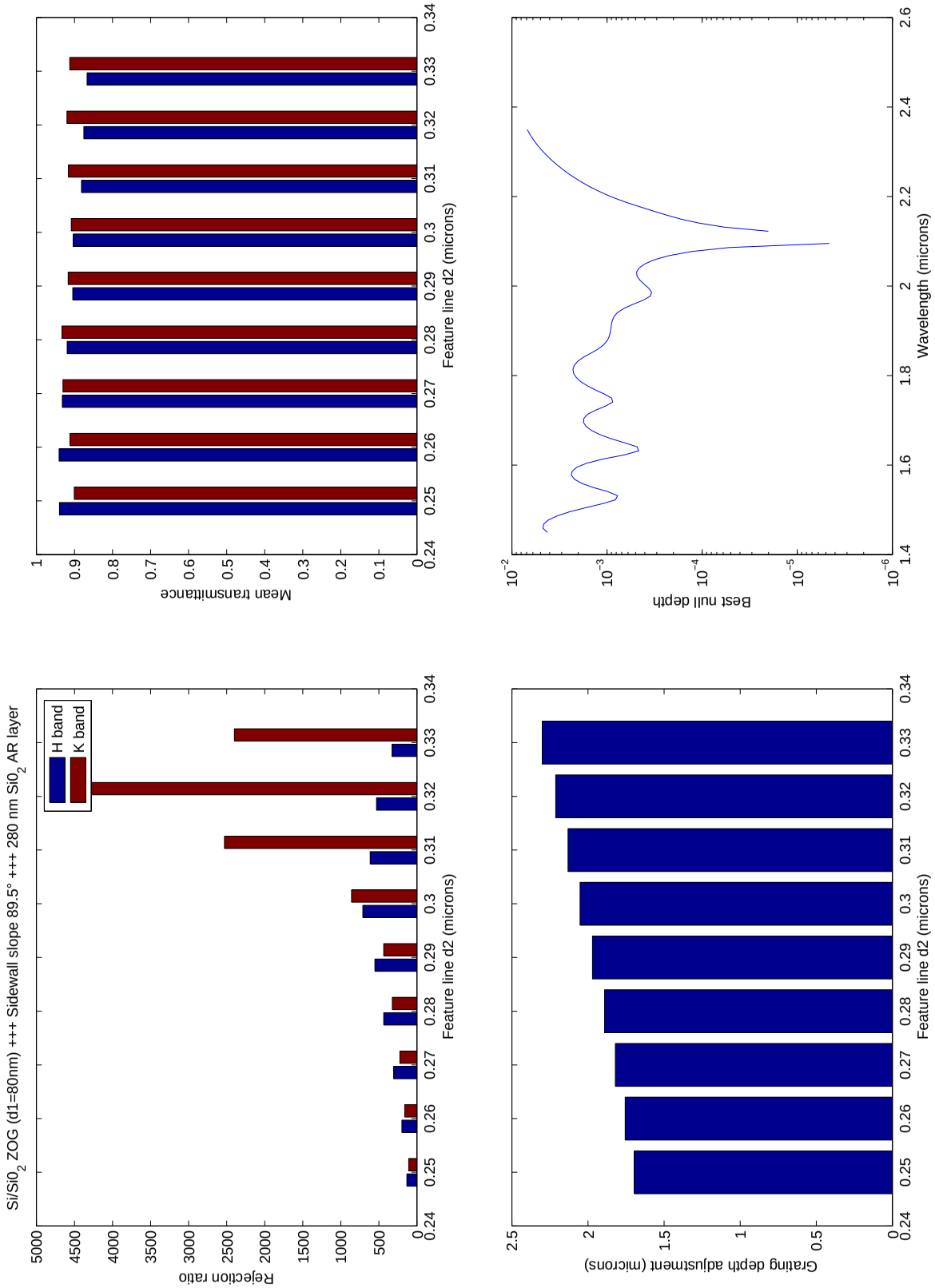

Figure 7.11: Panel for the 89.5° grating-wall slope angle.

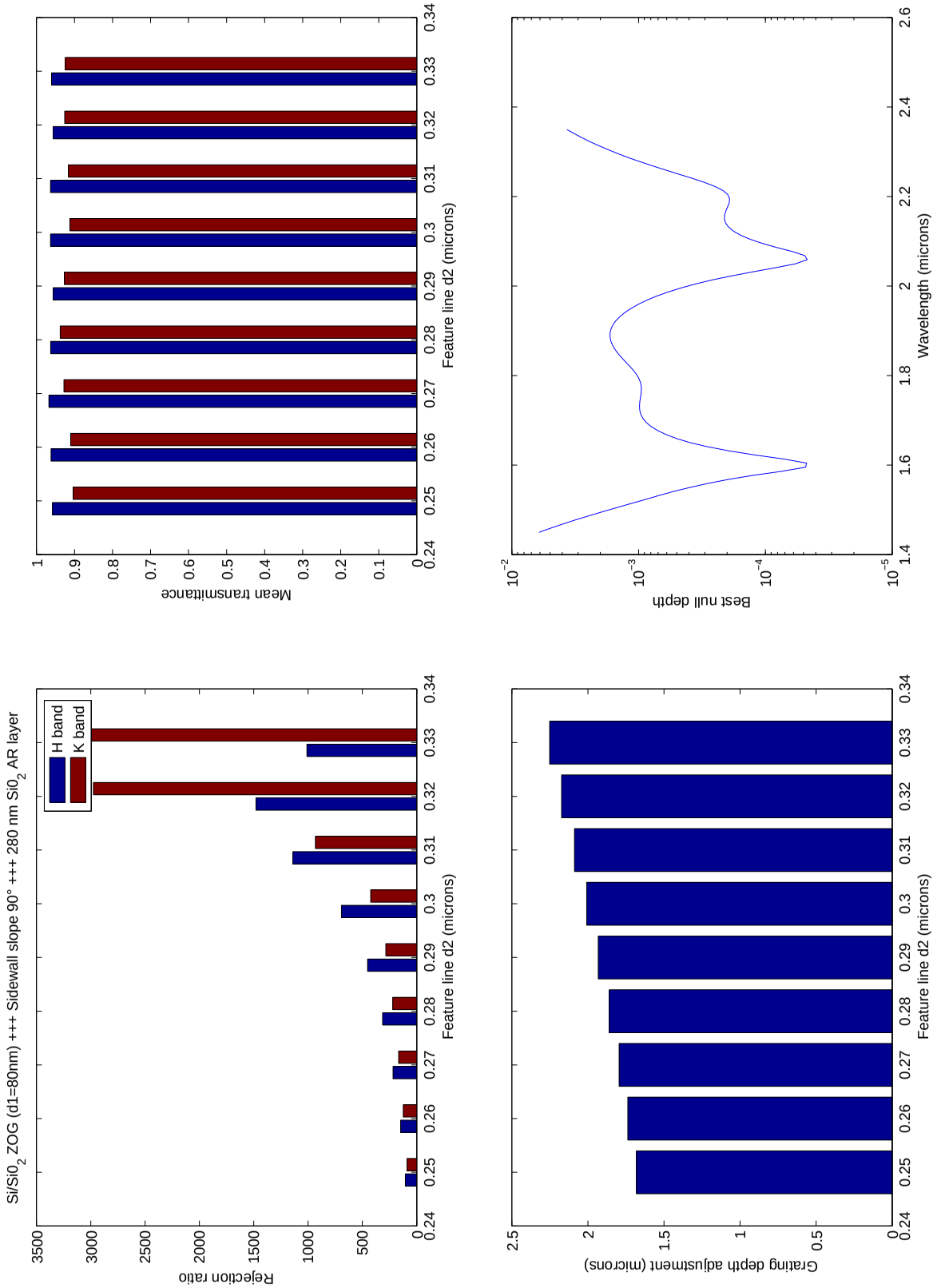

Figure 7.12: Panel for the 90° grating-wall slope angle.

Part IV

Nulling interferometry

8

Theoretical study of the total internal reflection grating APS

Contents

Abstract. *The so-called total internal reflection grating achromatic phase shifter (TIRG APS or APSZOG) consists of a back-illuminated subwavelength grating whose physical characteristics induce a super-achromatic phase shift between the vectorial components s (or TE) and p (or TM) of the incident natural or polarized light, while the total internal reflection insures a* 100%*-diffraction efficiency. As already stressed out, such high-performance phase shifters are required in the fields of nulling interferometry in order to achieve large dynamics (up to* 10^7 *for the detection of an exo-Earth in the thermal IR).*

8.1 Summary of the context

As already stated in the introductory part of this work, direct detection and characterization of faint sources around bright astrophysical objects is very difficult due to the large flux ratio between them. For example, an Earth-like exoplanet is typically $\sim 10^{10}$ times fainter than its host star in the visible spectrum, reducing to $\sim 10^{7}$ in the thermal infrared (Sect. 1.2.3). Infrared nulling interferometry proposed by Bracewell (1978) appears to be one of the most promising technique to achieve the high angular resolution and high dynamic range required to allow the ambitious detection of the first exobiological tracers on extrasolar Earth-like planets, if they exist. The nulling interferometry technique, as already introduced in Chapter 1, consists in adjusting the phases of the beams coming from various telescopes (two in the most simple configuration) to produce a fully destructive interference on the optical axis. The quality of the destructive interference, or the so-called null depth (N) relies on the optical component ability to induce a very precise phase shift (e.g. π for a two-telescope configuration) and a very low amplitude mismatch over the considered wavelength range.

Unfortunately, searching for biomarkers in exoplanet atmospheres requires spectroscopic characterization over large spectral bands. For example, the Darwin infrared space interferometer (Léger et al. 1996) considered by ESA will operate in a wavelength band between 6 and 18 μm, seeking biosignatures like the O_3-H_2O-CO_2 triplet (Sect. 1.2.3). The huge flux ratio between the parent star and the planet therefore requires unprecedented high-performance broadband achromatic phase shifters (APS). The performance requirement for the APS components directly comes from the 10^{-7} planet-star contrast but also from the system architecture and the associate amount of residual stellar leakage (Sect. 2.1.3). According to Chazelas et al. (2006), for an interferometer with stellar leakages similar to those of a Bracewell two-telescope interferometer, a sensible value for the mean instrumental leakage would be

$$N(\lambda) = 10^{-5} \left(\frac{\lambda}{7\,\mu m} \right)^{3.37} \tag{8.1}$$

In addition to the APS imperfections there are many systematic sources of reduction for the null depth (Lay 2004): e.g., telescope pointing errors, photometric unbalance, OPD errors, etc. Consequently, the requirement for the limitation on the null depth due to the APS alone, must be significantly lower than expressed in Eq. 8.1, i.e., $N = 10^{-6}$ or better. In practice, we will always apply a reasonable security coefficient for the performances of the APS by requiring a few 10^{-7}. Several concepts of APS have already been presented (Sect. 2.2). Each of them has its advantages and drawbacks. For instance, the mirror approaches (focus-crossing APS and field-reversal APS) are limited to π-phase shifts only (or $\pi/2$ for the focus crossing APS). Both induce a pupil rotation of 180° which could reveal very penalizing in terms of wavefront filtering, all the more if no monomode waveguide are available[22]. The dispersive plate approach is sensitive to the material choices and characterizations since three different ones are needed. Given the demanded qualities (transparency, knowledge of the dispersion, etc.), finding and selecting them is not obvious in the thermal infrared. Moreover, the working point of the dispersive-plate APS is quite tricky to find in practice. The solution we suggest can provide any phase-shift value and does not induce any pupil rotation while requiring a single material. Of course, it has also its own Achille's heels as we will discuss.

[22] As of today, no satisfactory monomode waveguide is available, but R&D is very active in this domain. Of course, the demand is very strong.

8.2 Modulating the total internal reflection

The principle of the TIRG APS is to make us of the ability of subwavelength gratings in the total internal reflection (TIR) configuration to induce achromatic phase shifts between the orthogonal polarization components TE (or s) and TM (or p) (see Sect. 4.1.2) and implement it into an APS to use in the framework of infrared nulling interferometry. For that, we propose to engrave a subwavelength grating on the TIR interfaces of rhombohedra in order to compensate for the natural index dispersion endured by classical Fresnel rhombs (King 1966; Bennett 1970), which is very penalizing in the present context of nulling interferometry, especially in thermal infrared where the choice of material is limited (Korte et al. 1988; Anderson 1988). We also demonstrate that depositing a single thin layer of a well-chosen material can also lead to a significant improvement over the classical Fresnel rhombs.

8.2.1 Total internal reflection grating

Interaction between the grating engraved on the TIR interface (Fig. 8.1) and the impinging vectorial electromagnetic field leads to interesting effects on the phase and amplitude of the totally reflected light. As already stated in Sect. 3.3, 1D subwavelength gratings, i.e., gratings only modulated along one dimension, turn out to be artificially birefringent. It means that the structure can be associated with two so-called effective indices, one for each polarization component TE and TM. These effective indices, n_{eff}^{TE} and n_{eff}^{TM} are totally dependent on the grating and incidence geometries. As the geometry can be controlled, one can really speak of refractive-index engineering. The induced vectorial phase shift can then be made very achromatic. Achromatic means in this case that the phase-shift value remains constant over the considered wavelength range. As the considered leading application of the TIRG APS component is nulling interferometry [23], we chose to optimize the grating design with the null depth as the figure of merit. The null depth somehow is the quantitative measure of the darkness of the destructive interference (see Sect. 2.1). The TIR configuration ensures a 100% efficiency for the back-reflected light whatever the polarization, and therefore no amplitude mismatch[24]. The figure of merit to be minimized consequently resumes

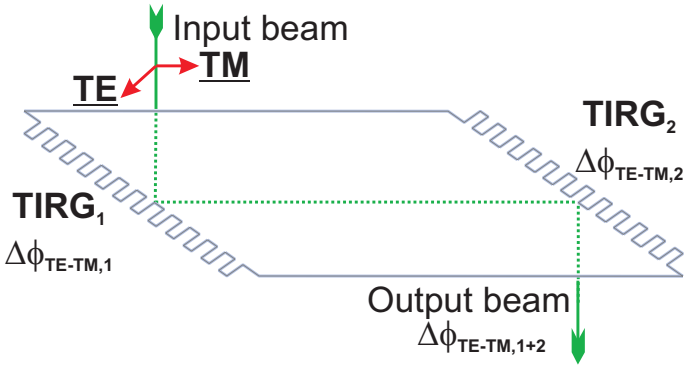

Figure 8.1: Schematic of the TIRG APS: it is analog to a Fresnel rhomb which TIR interfaces are engraved with an optimized subwavelength grating. A TIRG APS designed for a π-phase shift can possess for example two TIR interfaces, each providing a $\pi/2$-phase shift such that the resultant is $\Delta\Phi_{TE-TM,1+2} = \Delta\Phi_{TE-TM,1} + \Delta\Phi_{TE-TM,2} = \pi/2 + \pi/2 = \pi$. Such a component is to be inserted in each interferometer arm and orthogonally from one another (see Sect. 8.4).

[23] The TIRG APS high performances could as well be used in other applications like polarimetry or ellipsometry.

[24] We will see later in this chapter that, despite the TIR configuration, roughness can be a source of amplitude mismatch.

to $N(\lambda) = \sigma^2/4$, where σ is the phase-shift error standard deviation with respect to the nominal value of π (any other value is possible). This means that the null depth is directly proportional to the variance of the phase shift over the considered bandpass, i.e., the achromaticity.

8.2.2 Total internal reflection thin/thick film

Instead of engraving a subwavelength grating onto the TIR interface, one can deposit a layer of a well-chosen foreign material (Fig. 8.2).

Total internal reflection optically-thin film. The principle of the so-called total internal reflection thin film achromatic phase shifter (TIRTF APS) is to make use of such a thin film coated with an appropriate thickness on the TIR interface. This principle has been known in the visible regime for quite a long time. Indeed, MgF_2 thin films are commonly deposited on $BK7$- or silica-rhomb TIR interfaces to improve the angular and chromatic behavior of commercial Fresnel rhombs (Clapham et al. 1969; Filinski & Skettrup 1984; Spiller 1984; Nagib & Khodier 1995). The physical explanation of the compensation effect is the same as for subwavelength gratings (Sect. 4.1.2): the phase difference between the polarization components is modulated by the differential skin effect undergone by the evanescent waves and engendered by the presence of the layer of a foreign material. In other words, light does not see the rhomb/air interface, but rather the thin film/air interface plus the multiple interference effect between the evanescent waves inside the layer.

Total internal reflection optically-thick film. Azzam (2004) and Azzam & Spinu (2004) also proposed to use an optically-thick film forming a dielectric-dielectric interface with a well-chosen and known refractive-index ratio in order to minimize entrance beam configuration sensitivity, or more precisely, the phase shift $\Delta\phi$ dependence to incidence-angle variations. Indeed, it has been known for a long time (Bennett 1970) that the optimal index ratio for this purpose, i.e., nullifying $\partial\Delta\phi/\partial\theta = 0$ while imposing $\Delta\phi = \pi/4$, is $n = \sqrt{2} + 1 = 2.414214$. This index ratio roughly corresponds to a diamond/air interface or as suggested by Azzam & Spinu (2004) to a Si/SiO_2 interface working in the 1.2-4 μm range. This solution is unfortunately limited by the choice of existing materials.

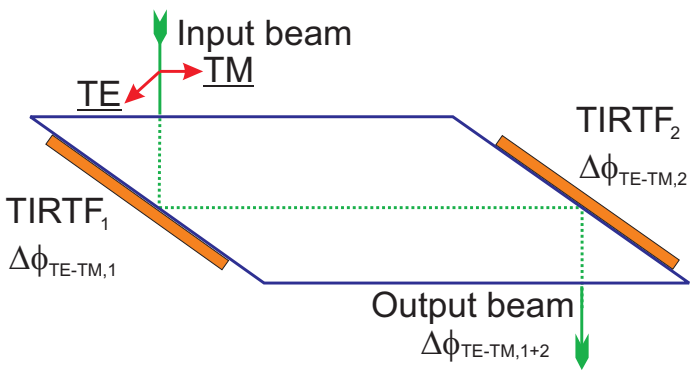

Figure 8.2: Schematic of the TIRTF APS component: it is analog to a Fresnel rhomb which TIR interfaces are coated with a thin or thick film of a well-chosen foreign material.

8.2.3 Double-rhomb configuration

Single Fresnel rhombs such as illustrated in Fig. 8.1 are known to be very sensitive to incidence variations (Nagib & El-Bahrawy 1994; Nagib 1997, 1998). This is unfortunately also the case for the TIRG APS. Fig. 8.3 (left) shows the standard deviation of the phase shift versus the incidence angle in the case of a $CdTe$ single-rhomb TIRG APS. Even if the absolute achromaticity σ_μ[25] is preserved, the achromaticity with respect to π, σ_π, is only maintained in the $N = 10^{-6} \Longleftrightarrow \sigma_\pi = 2 \times 10^{-3}$ specification within ~ 5 arcmin. Of course, this calculus is raw and the tolerance to beam divergence for instance has to be refined. To evaluate the real impact of the beam divergence on the null depth, we have finely modelled the incidence geometry for the incoming diverging beam. For that, it is worth noting that the tolerance with respect to conical angle variation[26] is of several degrees, therefore considered as negligible. Taking this into account, the incident beam can be discretized into rectangles (instead of rings) each associated with a specific corresponding polar angle θ (see Fig. 8.3, right). To each of these rectangles is also associated a null depth, function of the polar angle. The surface and therefore the energy incident on each of these rectangles is given by the following simple relation

$$S(i) = [R(i+1) - R(i)] \cdot 2R \sin \alpha(i) \tag{8.2}$$

where $R(i) = \theta(i) \cdot f$, $R = \theta(n) \cdot f$ and $\alpha(i) = \arccos \frac{R(i)}{R}$, with f, an arbitrary length. Intuitively, we can anticipate that the outer top and bottom portions of the beam will contribute less than the central ones where the optimal values are voluntarily centered. The actual resulting null depth is subsequently given by

$$N_{res} = \sum_i S(i) N(i) \tag{8.3}$$

where the profile of S (normalized to 1) is given by Eq. 8.2. The calculation shows that the tolerance of 5 arcmin mentioned here above is relaxed by a factor of 2, leading to a 10-arcmin sensitivity, whatever the beam dimension R.

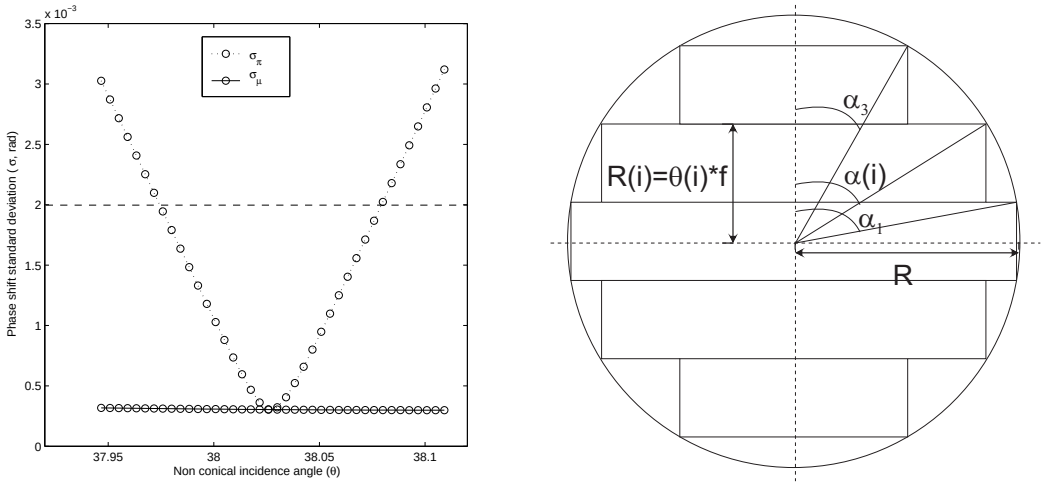

Figure 8.3: Left: phase-shift error standard deviation versus incidence-angle variations for a single-rhomb $CdTe$ TIRG APS. Right: beam specific discretization for beam-divergence analysis. R is the beam radius in this context.

[25] σ_μ is calculated with respect to the mean phase shift whereas σ_π is calculated with respect to π.

[26] The conical angle is the azimuth angle ϕ defined in Sect. 3.2.3, i.e., the angle defining the orientation of the plane of incidence when it does not contain the grating vector.

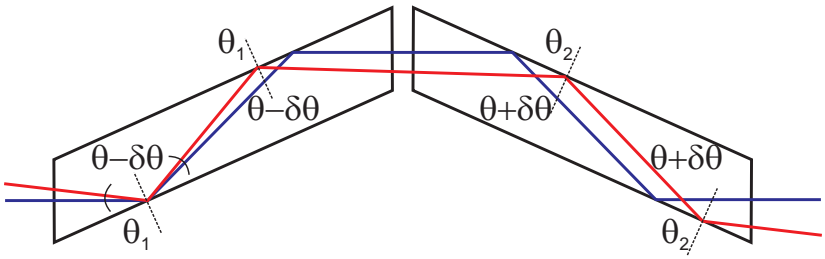

Figure 8.4: This scheme shows the double-rhomb configuration. θ_1 (resp. θ_2) is the angle of incidence upon the TIR interfaces of the first (resp. second) rhomb.

Nevertheless, such a substantial sensitivity would be penalizing in the present application since the thermal infrared interferometric beam is not only likely to diverge because of Fresnel diffraction (Chazelas 2003) but it is also likely to wander around because of vibrations (Brachet 2005). For this reason, we chose to consider other geometries that could alleviate the entrance beam configuration sensitivity. Note that this issue is surely not a new one since, to our knowledge, a first answer to the problem was proposed 50 years ago by Mooney (1952) and its famous so-called "Mooney rhomb" (pentaprism shape). Since then, there have been many proposed alternative geometries but we chose the so-called "double-rhomb configuration" for its simplicity (the beam remains on-axis, see Fig. 8.4) and very robust insensitivity to incidence-angle variations (Rochford et al. 1997). Indeed, the pairs of reflections in the two rhombs are complementary (see Fig. 8.4); an increase in the first two TIR angles due to departure from nominal incidence leads to a decrease in the last two angles. Thus if the phase shift varies linearly with the TIR angle, the retardance changes are cancelled (see Fig. 8.5). Therefore a wider range of incidence angle variations can be tolerated: up to several degrees.

It is to be noted that in the double-rhomb configuration, the number of subwavelength gratings to be imprinted (resp. thin film to be deposited) on the TIR interface may be limited to only two out of four. Indeed, the dispersion compensation artificially introduced by the modulated interface(s) is sufficient to reach the specifications (see Sect. 4.1.2).

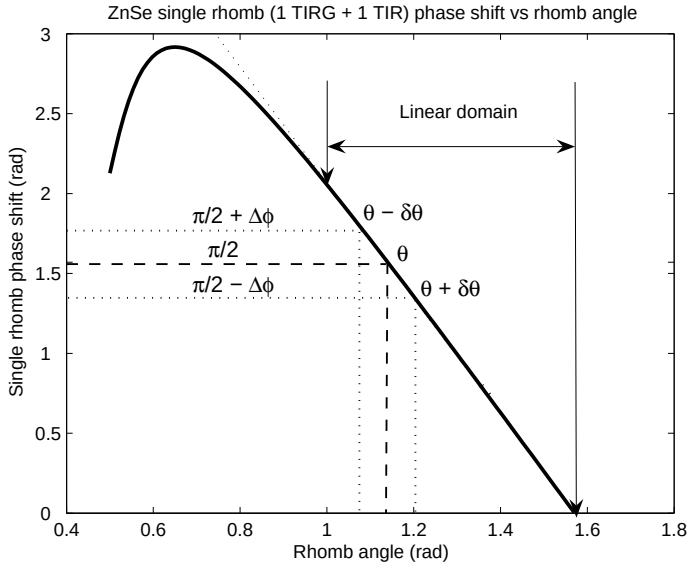

Figure 8.5: The double-rhomb geometry allows incidence-angle variations $\delta\theta$ to be compensated by the angle complementarity between the two rhombs and the linearity of the phase shift with respect to the rhomb angle.

8.3 Theoretical analysis

This section is devoted to the theoretical analysis of the TIRG and TIRTF APS. For that, we used a RCWA code based on the theory exposed in Sect. 3.2.3 and the simplex-optimization procedure when needed (see Sect. 5.1.2). The purpose of this analysis was to find the optimal parameters for the APS components "achromatizing" the Darwin working wavelength range, from 6 to 18 microns. However, due to practical constraints like wavefront filtering or dichroic specifications, the latter is expected to be divided into two or three sub-bands. In the two-band case, the first one ranges from 6 to 11 microns while the second one ranges from 11 to 18 microns.

Before going into the details of the theoretical analysis, one had to choose the materials according to the bandwidth specifications mentioned here above. Regarding this matter, we immediately had to discard common infrared materials like silicon (Si) and gallium arsenide ($GaAs$) for their strong multi-phonon absorption features beginning between 8 and 12 microns (see Annex B), i.e., the most interesting wavelengths for biosignature detection (remind that the ozone spectral signature, for instance, is centered at 9.65 μm). For the same reason, we also had to discard zinc sulfide (ZnS), another widespread infrared material (see Annex B). This is very unfortunate since these materials are very convenient to handle.

The choice for the bulk material constituting the rhomb in fact revealed to be very difficult and indeed severely limited since the material has to be perfectly transparent up to 18 microns and of course available in large ingots of very good optical quality (very good homogeneity, low impurity, etc.). For this reason, we focussed on the remaining infrared material worth considering for Darwin (see Annex B): zinc selenide ($ZnSe$), cadmium telluride ($CdTe$) and germanium (Ge). Although $ZnSe$ and Ge are in general used in the 11-18 μm range, $CdTe$ is the only one perfectly clear up to 18 microns. It is to be emphasized that KRS-5, common in infrared applications (see Annex B) despite its bad thermal conductivity (5.4×10^{-3} W/cm/K) and expansion coefficient (5.8×10^{-5}), was also considered since it is one of the rare materials transparent above 15 microns. However, it revealed to be less efficient than $CdTe$ and $ZnSe$. Moreover, it is very difficult to handle because of its toxicity and softness. Its polishing was never demonstrated at better qualities than $\lambda/4$ rms ($\lambda = 632.8$ nm). Finally, the lack of known micro-structuring processes for this material has therefore prevented its practical interest in the present application.

We therefore selected $ZnSe$, $CdTe$ and Ge for the rhomb bulk materials. It is worth insisting on the fact that these materials are common in infrared applications, that they cover a large refractive index spectrum ($n = 2 - 4$) and that their manufacturing processes are not totally unknown for most of them. Note that other materials like Si, $GaAs$, ZnS and even diamond can be considered as layer material candidates for the TIRTF solution, even in their phonon absorption range since the thicknesses needed are very thin and the subsequent absorptions therefore negligible. The refractive and absorption properties for the considered materials are summarized in Annex B. The working temperature is assumed to be either $T = 100$ K or $T = 298$ K keeping in mind that corrections are needed if T is different.

8.3.1 $ZnSe$ rhomb

$ZnSe$ is available in large quantities/volumes at a relatively low cost. $ZnSe$ is also easy to polish with very good surface qualities. Moreover, its thermal properties are very attractive, with a low thermal expansion coefficient (7.1×10^{-6}/K) and a good thermal conductivity (0.18 W/cm/K). Unfortunately, the chosen double-rhomb configuration inevitably lengthens the optical path in the material. For instance, in the present case, a working angle of ~ 1.139 radians would lead to a

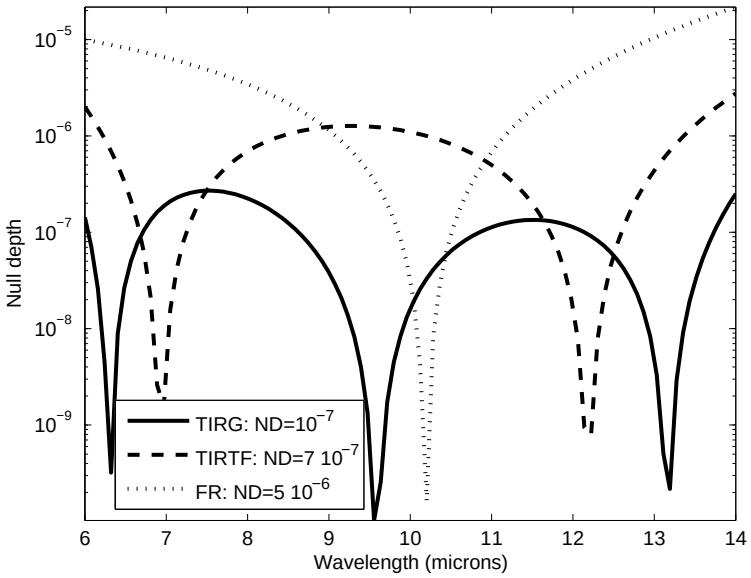

Figure 8.6: *ZnSe* double-rhomb APS: comparison between Fresnel rhomb (FR) with non-treated TIR interfaces, TIR thin film (TIRTF) and TIR grating (TIRG). More than substantially lowering the global (mean) null depth over the considered wavelength range, the TIRTF and the TIRG solutions significantly decrease the strong leakage at its edges, inevitable with the bare FR solution.

physical path of ~ 17.3 cm inside the material for an entrance beam diameter of 15-20 mm. In the second Darwin sub-band (11-18 μm), such a long path is penalizing since *ZnSe* begins its phonon absorption around 14 microns. Indeed, the absorption coefficient (Hawkins 1998, see Annex B) k is equal to 4.24×10^{-6} at 14 μm and 300 K, leading to an absorption of about 50%. This value reduces to 2.52×10^{-6} at 100 K, giving an absorption of 30%, which remains acceptable for demonstration purposes. For this reason, a *ZnSe* TIRG APS in the double-rhomb configuration will be limited to the 6 to 14 μm wavelength range.

Results of RCWA calculus and optimization are shown in Fig. 8.6 and summarized in Table 8.1. First of all, it is worth noting that the *ZnSe* bare Fresnel rhomb (without any ZOG nor coating on the TIR interface(s)) is not performing well enough according to the Darwin $N = 10^{-6}$ specification (Fig. 8.6, dotted line). However, depositing a foreign material at the TIR interface or engraving a subwavelength grating onto it allows us to overcome this limitation at least in the first 6-11 μm Darwin wavelength range. The performance is somewhat deteriorated in the larger 6-14 μm range, but it is still comfortable in the Darwin specification for the subwavelength grating solution (Fig. 8.6, continuous line) while only by a feeble margin for the thin film one (Fig. 8.6, dashed line). Note that in this latter case, CVD (Chemical Vapor Deposition) diamond would be particularly appropriate as the layer material (see Sect. 9.2.2).

8.3.2 *CdTe* rhomb

Availability of *CdTe* ingots is more limited than *ZnSe* ones and the price is higher. Polishing of *CdTe* is quite delicate but currently under evaluation for improvement by several manufacturers. *CdTe*, which also possesses good thermal characteristics (expansion coefficient of 5.9×10^{-6}/K and

Table 8.1: Null depths for the optimal Fresnel rhomb, TIRG APS, TIRTF APS components for the selected materials ($ZnSe$ and $CdTe$). CLD stands for Carbon like diamond (i.e., CVD diamond).

Material/band	**Fresnel rhomb**	***TIRG APS***	***TIRTF APS***
$ZnSe/6-11$ μm	1.6×10^{-6}	1×10^{-8}	CLD/ZnS layer: 1×10^{-7}
$ZnSe/6-14$ μm	6×10^{-6}	1×10^{-7}	CLD layer: 7×10^{-7}
$CdTe/6-11$ μm	2×10^{-7}	2×10^{-9}	CLD/$ZnSe$/ZnS layer: 1×10^{-8}
$CdTe/11-18$ μm	8×10^{-7}	1×10^{-8}	CLD/$ZnSe$/ZnS layer: 4×10^{-8}
$CdTe/6-18$ μm	1.6×10^{-6}	1×10^{-7}	CLD/$ZnSe$/ZnS layer: 4×10^{-7}

conductivity of 0.062 W/cm/K), nevertheless appears to be more than a viable solution[27]. Indeed, as already stated, $CdTe$ is the only selected rhomb material perfectly clear up to 18 microns (see Annex B). Moreover, theoretical RCWA results are excellent, showing better performance than with $ZnSe$ (see Table 8.1). First of all, the $CdTe$ bare Fresnel-rhomb solution is worth considering since it is theoretically performing well enough, at least for the Darwin first sub-band (6-11 μm) but unfortunately not for the second one (11-18 μm). Depositing a thin film of a foreign material (e.g., diamond, $ZnSe$ or ZnS) overcomes this limitation. As far as the subwavelength grating solution is concerned, results are comfortably in the specifications for both bands.

Let us emphasize that there exists a most interesting solution achromatizing the whole Darwin wavelength range from 6 to 18 microns with only one component: an optimized double rhomb in $CdTe$ modulated with a subwavelength grating on at least two of the four TIR interfaces (see Table 8.1).

8.3.3 Ge rhomb

Germanium is a very well suitable material for classical Fresnel rhombs since its natural dispersion is extremely low (Hawkins 1998, see Annex B). Theoretical results for the bare Fresnel rhomb are excellent: $N = 4\times10^{-9}$ over the 6-11 μm wavelength range. Adding a ZOG at the TIR interface further improves this already excellent behavior. Unfortunately, Ge is not perfectly transparent in the Darwin 11-18 μm band (its phonon absorption begins around 12 microns, see Annex B), making it a partial solution only. It is also worth noting that Ge thermal properties are, as every dielectric materials in general, quite satisfying with a rather low expansion coefficient (6×10^{-6}) and a good conductivity (0.0602 W/cm/K).

8.4 Interferometer implementation

Implementation of the vectorial phase shift in a nulling interferometer is straightforward. Considering two strictly identical components belonging to the two distinct interferometer arms 1 and 2, rotated by ninety degrees around the optical axis and from one another, then the potentially-interfering parallel polarization states (s_1 with p_2 and s_2 with p_1) are two by two in phase opposition. It must be noted that there is a strong constraint on the alignment of the components around the optical axis with respect to each other. Let $\Delta\chi$ be the misalignment angle between the rhombs along the optical axis (Fig. 8.7). $\Delta\chi$ is then directly related to the null depth N:

[27] It is to be noted that CSL possesses a $CdTe$ growing facility (Verstraeten 2002). Samples for the present application are being grown and will soon be qualified (dispersion and thermo-optic coefficients).

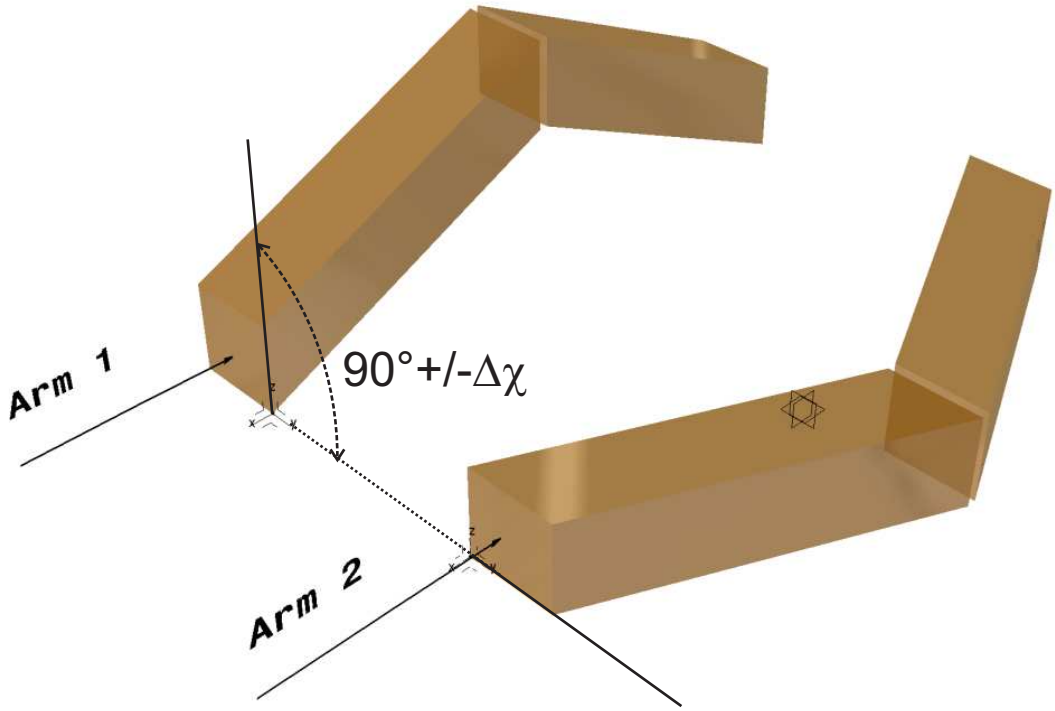

Figure 8.7: Interferometer implementation: two strictly identical components belonging to the two distinct interferometer arms, are rotated by ninety degrees around the optical axis and from one another. $\Delta\chi$ is the misalignment angle between the rhombs around the optical axis.

$N = (1 + \sin\Delta\chi)(1 - \cos\Delta\chi)/2$. Therefore, to fulfill the constraint $N < 10^{-7}$, we must impose that $\Delta\chi \leq 2$ *arcmin*.

Departure from the nominal zero optical path length between the two components belonging to the two interferometer arms due to manufacturing/surfacing errors and/or differential thermal dilatations can be compensated either by slightly tilting the rhombs or making use of beam splitters with adjustable thicknesses if in the same material. It is to be noted that a difference in the geometry of the paired rhomb segments (e.g. length) is not expected to be an issue provided that the segments are cut and polished together.

8.5 Tolerancing and design of a prototype

In this section, we will discuss the design of a prototype TIRG APS intended at being one of the four APS selected for study, fabrication and test in the framework of the Darwin R&D activities (Nulltimate consortium 2002; CSL & IAGL 2005). The first part of the present section will be devoted to the tolerancing of the subwavelength grating, the second one to the roughness and homogeneity issues. In the third and last sub-section, the rhomb design will be addressed and the final drawings presented. The choice of the material for the prototype has been made thanks to a trade-off matrix (CSL & IAGL 2006). Even if $CdTe$ shows the best theoretical characteristics in terms of transparency and performances, $ZnSe$ was the retained material for the following reasons:

- it is a quite cheap, easy to source, and convenient material to handle and polish (surface qualities of $\lambda/30$ rms with $\lambda = 632.8$ nm are routinely obtained);
- $ZnSe$ micro-structuring is well studied and referenced in the literature;
- it is compatible with the thickness-adjustable beam splitters of the NULLTIMATE test bench which are planned to be in $ZnSe$ (see Sect. 9.3.3);
- the gain provided by the TIRG solution is more obvious with $ZnSe$ than with $CdTe$, i.e., it should be easier to measure the gain of the TIRG solution in the $ZnSe$ case as compared to the $CdTe$ one[28].

8.5.1 Micro-structure tolerancing

The fabrication of the TIRG APS will be based on micro-electronic technologies (Sect. 3.4). The first classical step consists in imprinting a photomask of the grating in a resin coated on the chosen substrate material. The precision of this step is critical because it defines once and for all the lateral dimensions of the grating: its period Λ and the so-called feature line, i.e., the period multiplied by the filling factor F. This pattern will then uniformly be transferred into the substrate by an appropriate reactive plasma-beam etching down to the desired depth. The fabrication has to be interactive to properly compensate for process errors. In situ monitoring (Sect. 3.4.3) is a possible solution but not the only one as we will discuss.

Let us now consider the $ZnSe$ TIRG APS designed for the 6-14 μm wavelength range. A rough optimization for this range leads to a $\sim$ 900-nm period taking a fabrication constraint already into account. Indeed, the photomask is expected to be written by a direct writing laser (DWL, see Sect. 3.4.1) which has a 600-nm writing spot size (see Sect. 9.2). This specification of the writing apparatus in fact already constrains the period-filling factor relation which must satisfy $(1 - F)\Lambda > 0.6\,\mu$m. Then, assuming a fixed period of 900 nm, best solutions are searched using the *simplex*-optimization method coupled to the RCWA algorithm (see Sect. 5.1.2) with the free parameters left: the filling factor F, the grating thickness h and the incidence angle θ. Results of this optimization are displayed in Fig. 8.8 (left), where the optimal null depth is plotted versus the feature line, i.e., the product of the 900-nm fixed period and the varying filling factor. Continuous variations are imposed to the feature line while letting the optimization algorithm find the corresponding adjustment of the thickness that minimizes the null depth (see Fig. 8.8, right). The best null depth, in this case the minimum one, is $\sim 1.5 \times 10^{-7}$. This optimal value is

[28] The $CdTe$ bare Fresnel rhomb already provides a deep null without any ZOG or thin/thick film (Table 8.1) while for $ZnSe$, the null passes from $N_{FR}(6,\, 14\,\mu\text{m}) > 10^{-5}$ for the Fresnel rhomb to $N_{TIRG}(6,\, 14\,\mu\text{m}) \approx 10^{-7}$ for the TIRG APS (see Table 8.1 and Fig. 8.6).

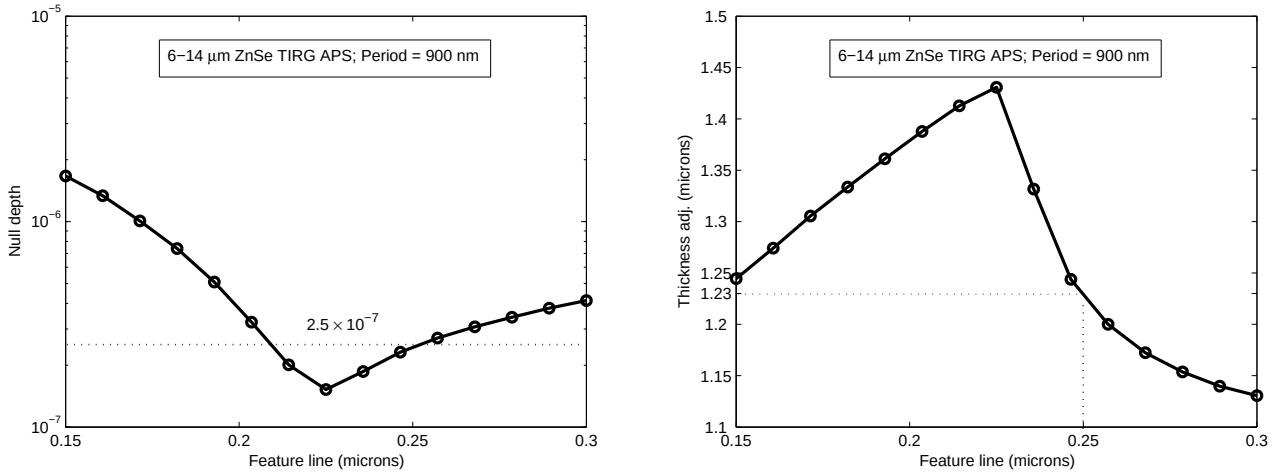

Figure 8.8: 6-14 μm *ZnSe* double-rhomb APS with 900 nm period. Left: optimized null depth versus feature line. Right: thickness adjustment (optimized) versus feature line.

obtained for the 220-nm feature line. The corresponding adjusted thickness is $\sim 1.43\,\mu$m. If we calculate the grating aspect ratio, we find 6.5 which is somewhat demanding (but not impossible) given the nature of the material to be subsequently etched. To relax this difficulty, we chose the conservative size of 250 nm for the feature line. Doing this, the adjusted thickness is $1.23\,\mu$m and the aspect ratio subsequently reduces to ~ 5, which is more comfortable. The 250-nm feature line choice can therefore be considered as the second trade-off between performance and feasibility. Let us now fix the feature line to 250 nm, keep the period at 900 nm and vary the thickness artificially, letting the optimization algorithm find the corresponding incidence polar angle θ adjustment that minimizes the null depth. Results of this analysis are shown in Fig. 8.9 left and right. Provided that the polar angle can be adjusted with a sub-arcmin precision, the $N \approx 2.5 \times 10^{-7}$ tolerance on the thickness definition is 1.23 ± 25 nm, i.e., $\sim 2\%$, which is feasible.

The conclusion of this tolerance analysis is that, provided that there are interactions between measurements and manufacturing at each key step, the $N \approx 2.5 \times 10^{-7}$ tolerance on the definition of the grating paramaters is comfortable, up to 50 nanometers for the feature line (if we accept

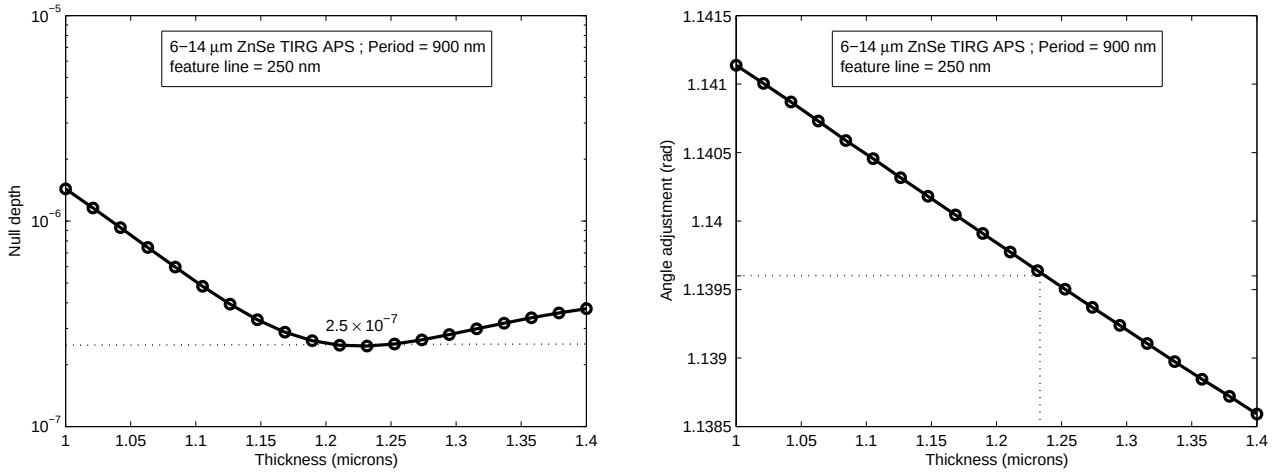

Figure 8.9: 6-14 μm *ZnSe* double-rhomb APS with a 900-nm period and 250-nm feature line. Left: angle-optimized null depth vs thickness. Right: corresponding angle adjustment vs thickness.

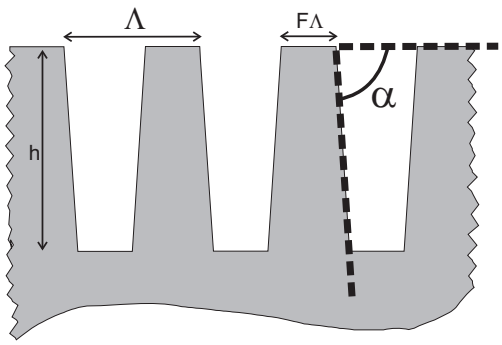

Figure 8.10: Trapezoidal profile likely to emerge from the plasma-etching process. The new parameter to be taken into account is the grating slope angle α.

to relax the aspect ratio manufacturing constraint) and about the same for the thickness. A posteriori correction is reported on the incidence polar angle with a sub-arcmin precision (i.e., $\sim 30''$) which is very convenient since the double-rhomb geometry ensures the insensitivity to the entrance beam incidence configuration (i.e., input angle and beam divergence) at the macroscopic level.

8.5.2 Grating slope angle

Departure from the nominal assumed grating profile, i.e., a perfectly rectangular one, is likely to naturally emerge from the plasma-beam etching process (Fig. 8.10). Generally, one can expect that if the desired overall filling factor of the profile is preserved in the fabrication process, then shape errors have a minimal effect on the desired diffraction characteristics (Pommet et al. 1995). However, as we will see, departure from rectangularity induces further constraints on the lithographic mask definition and the plasma-beam etching, as well as on the overall performances. Calculations assessing the sensitivity to the grating slope angle are summarized in Fig. 8.11 and Fig. 8.12. Null depth performance is significantly compromised for slope angles smaller than 85 degrees. Moreover, the grating-parameter adjustments (feature line, thickness and polar angle) are very penalizing. For example, a 88°-slope angle already leads to significantly lower the feature

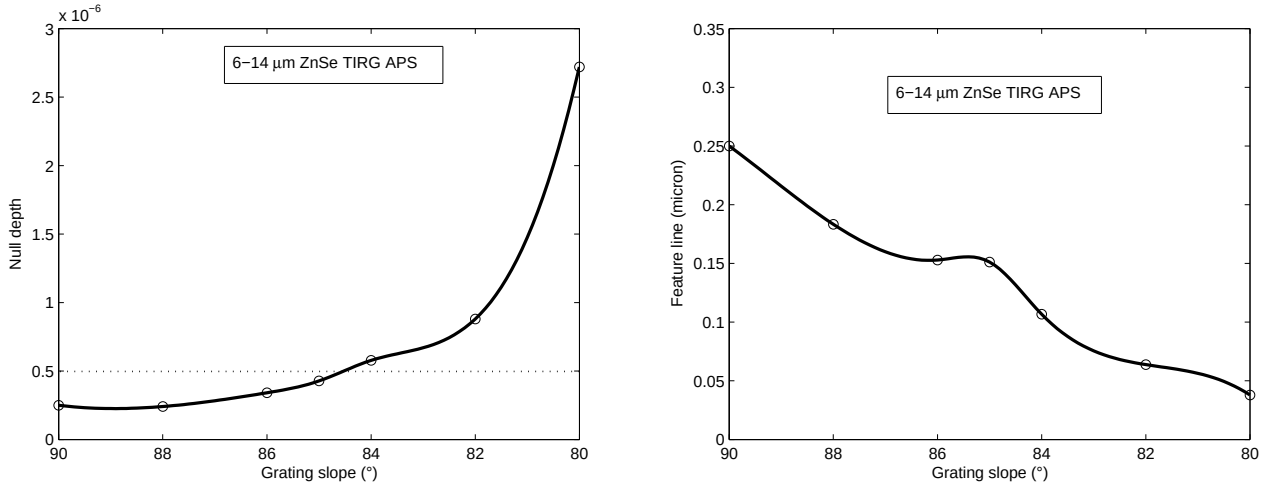

Figure 8.11: Grating slope-angle tolerancing. Left: null depth versus grating slope angle in degrees. Right: corresponding feature line adjustment versus grating slope angle.

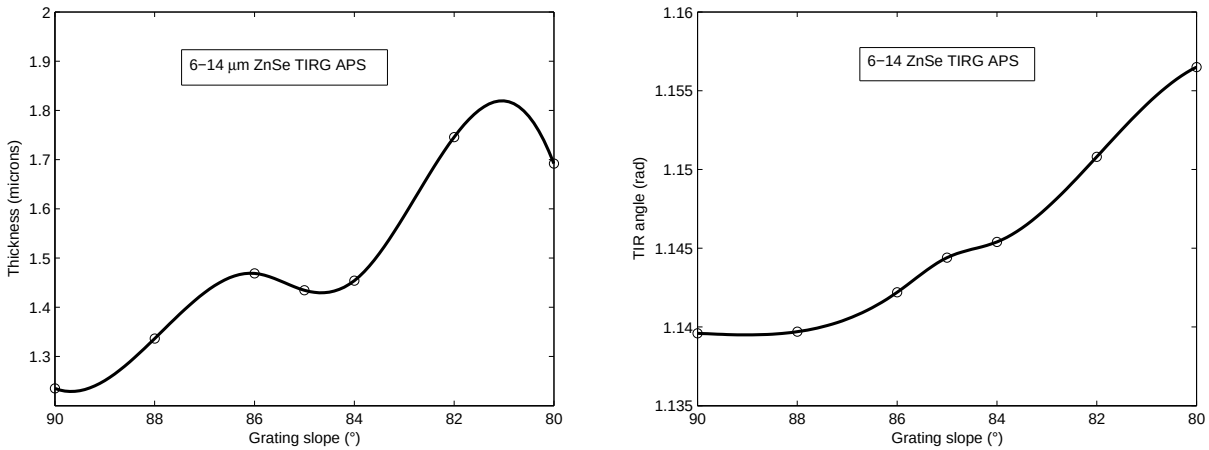

Figure 8.12: Grating slope-angle tolerancing. Left: corresponding thickness adjustment versus slope angle. Right: TIR angle adjustment versus slope angle.

line (down to ~ 200 nm) to keep the $N \approx 2.5 \times 10^{-7}$ performance. Moreover, the corresponding thickness adjustment is not in favor of keeping the aspect ratio in a reasonable range. At 80°, it is unrealistic, with a feature line smaller than 50 nm and a thickness of about 1.7 μm, all of them leading to a poor $N \approx 2.5 \times 10^{-6}$. For these reasons, the specification on the slope angle is once again: "best effort towards 90°". In any case, the **mandatory a priori knowledge of the global slope angle** is required.

8.5.3 Thin-film solution tolerancing

As far as the TIRTF APS component is concerned, tolerancing can be envisaged in another way. Since the parameter space is limited to two variables, i.e., the thickness of the layer and the angle of incidence, the working points can be drawn in two-dimensional maps. From Fig. 8.13, we can conclude that the tolerance on the thickness is several percent ($\sim 10\%$). It is to be noted that coatings are routinely deposited with a precision of $\sim$ 1-2% on the thicknesses (see Sect. 9.2).

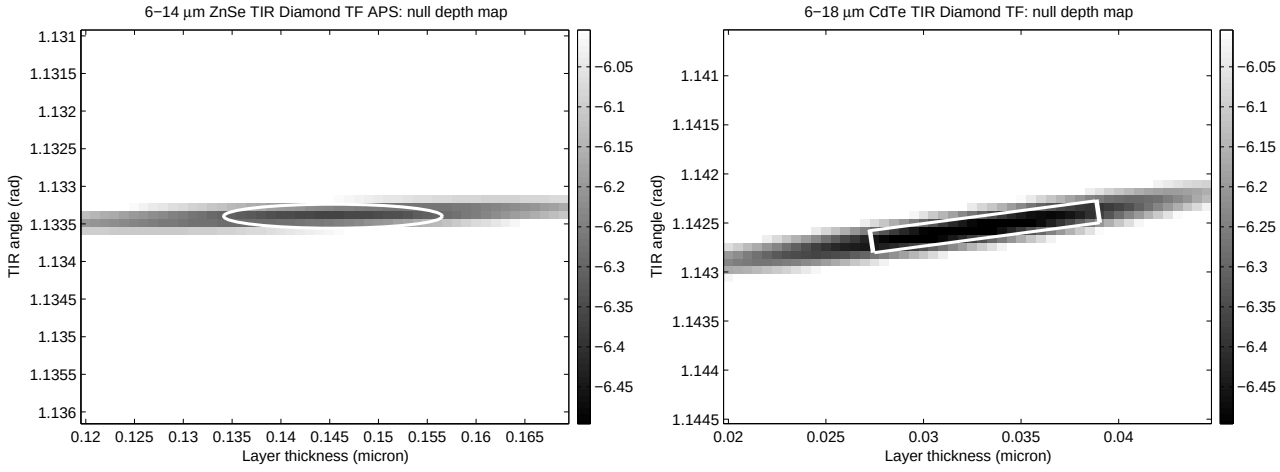

Figure 8.13: 2D maps of the null depth (log scale, $10^{-\alpha}$) according to the variables "thickness of the layer" and "incidence angle". Left: $ZnSe$ TIRTF APS coated with diamond for the 6 to 14 micron band. Right: $CdTe$ TIRTF APS coated with diamond for the 6 to 18 micron band.

8.5.4 Roughness and homogeneity

Roughness and material homogeneity both affect the wavefront quality and phase shift between the interferometer arms. Let us discuss them separately here below.

Roughness

Roughness is a general term for designing randomness in material surface topographies. Unfortunately, the spatial information is often "forgotten" and this can sometimes be misleading. In the present application, three spatial scales can be distinguished: the macroscopic scale with typical dimensions belonging to the 10 μm to centimeter range, 10-100 μm being the typical size of $ZnSe$ micro-crystal grains; the microscopic scale with dimensions ranging from the micrometer to 10-100 μm, and the nanoscopic scale with typical dimensions below the micrometer, down to a few nanometers.

The macroscopic-scale roughness is related to global wavefront errors (WFE) and large-scale micro-irregularities. Darwin specifications on WFE taking spatial filtering into account have led to the $\lambda_{632.8}/30$ rms surfacing constraint (see Sect. 2.1.2). As already mentioned here above, this tight specification was one of the reasons for the choice of $ZnSe$. Concerning large-scale micro-irregularities, the scalar scattering theory provides simple formulae yielding approximate quantities which are nonetheless sufficient in many applications (Hadaway et al. 2001). When a surface is illuminated, some light will be specularly reflected, some will be backscattered, and the rest will be transmitted or absorbed. The so-called total integrated scattering (TIS) is the fraction of backscattered light divided by the total light reflected from the sample (backscattered plus specularly reflected light). TIS can be directly related to a rms roughness δ if the scattering is caused by surface micro-irregularities over the entire surface, with heights small compared with the working wavelength, and with a correlation length (related to the spacing between roughness features) which is large compared to the working wavelength. The expression relating TIS and rms roughness is then

$$TIS = \frac{R_d}{R_0} = \frac{R_0 - R_s}{R_0} = \frac{\text{scatt. power}}{\text{scatt. power} + \text{spec. power}} = 1 - \mathrm{e}^{-(4\pi \mathrm{n}\delta\cos\theta/\lambda)^2} \tag{8.4}$$

where R_d is the diffuse reflectance (i.e., backscattered radiation), R_0 is the total reflectance (specular plus diffuse), R_s is the specular reflectance, n is the refraction index of the scattering material, θ the angle of incidence, and λ the wavelength of the impinging light. A numerical application with $\delta \approx 5$ nm (average between $ZnSe$ bare and treated surfaces), $n = 1$, $\theta = 65°$, and $\lambda = 6$ μm yields $TIS \approx 2 \times 10^{-5}$. This already small value must be considered as an upper bound and with extreme precautions since the TIS energy is spread out with an unknown angular dependence, i.e., the incoherent TIS fraction backscattered into the specular shall be much smaller than the calculated amount. As far as the grating dimensions are concerned on these macroscopic scales, they of course have to be homogeneous at the levels specified in the analysis presented here above.

The microscopic-scale roughness directly mixes with the grating dimension definition. In this respect, localized defects can be tolerated as far as the grating dimensions are globally respected. Indeed, spatial filtering precisely aims at filtering the high-spatial frequencies induced by small imperfections. Moreover, one interesting theoretical property of subwavelength gratings is that the global periodicity acts as an energy concentrator through the physical law depicted by the grating equation (see Eq. 3.45). Indeed, being subwavelength implies that energy cannot be diffracted into higher orders. Unfortunately, the subwavelength grating concept is a theoretical idealization.

Perturbations of any kind to the idealistic profile will always lead to diffusion and degradation of the nominal performances. Modelling micrometric defects is not an easy task and requires some tricks since, to our knowledge, there exists no simple method that allows mixing periodic structures, i.e., gratings, to randomness. In general, these problems are treated separately. Let us nevertheless attempt to quantify the effect of the present considered type of randomness. First of all, we will split the 1 μm to 10-100 μm range into the 1-10 μm and 10-100 μm ranges according to the minimum uniform grating size allowing the use of RCWA. Indeed, one of the working hypothesis of RCWA calculations is that structures are supposed to be infinitely periodic. In practice, the applicability of the algorithm can however be extended to finite-number-of-period gratings down to a minimum size of $\sim$ 10-20 periods, i.e., $\sim 15\,\mu$m in the present case (Wu & Glytsis 2002; Hirayama et al. 1997).

Roughness in the 10-100 μm scales can then be estimated by Monte-Carlo simulations dividing the beam into cells encompassing 10 to 20 periods of the subwavelength grating, i.e., with a typical size of $\sim 15\,\mu$m. As stressed out here above, this discretization allows the use of the RCWA algorithm within each elementary cell with a sufficient accuracy (Wu & Glytsis 2002; Hirayama et al. 1997). As far as roughness amplitudes are concerned, they will be chosen according to the nominal values routinely obtained by state-of-the-art $ZnSe$ surfacing techniques at these scales, i.e., 1-2 nm rms, but corrected by the expected degradation of the plasma-etching process that is likely to increase the roughness up to $\sim$ 5-10 nm rms (see Chapter 9). Taking the fact that

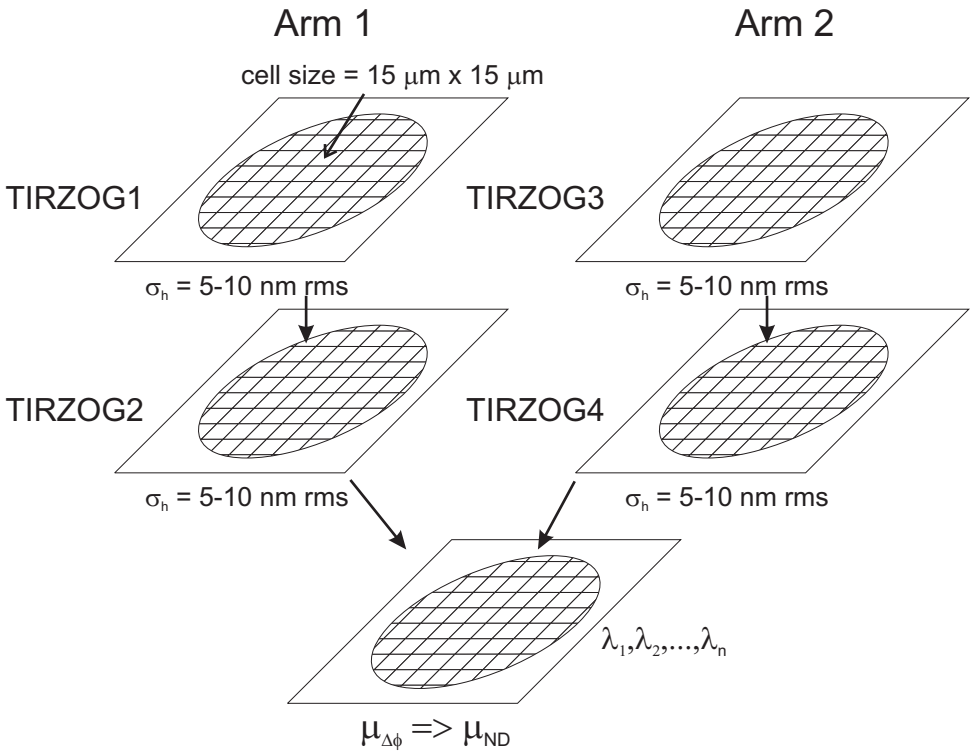

Figure 8.14: Principle of the Monte-Carlo micro-roughness analysis in the $10-100\,\mu$m scale. The 15 mm-pupil is discretized into cells of $15\,\mu\text{m} \times 15\,\mu$m in size, where the thickness h of the grating is assumed to be affected by a Gaussian random error of $\sigma \approx$ 5-10 nm rms. Since each arm of the interferometer undergoes two TIR on two subwavelength gratings, two independent phase screens per arm are generated and then recombined properly at one wavelength at a time, for a total of 20 wavelengths covering the 6-14 μm range. Then the mean phase-shift residual $\mu_{\Delta\phi}$ with respect to phase opposition is calculated within each cell and then averaged through the entire pupil to give the final mean null depth μ_N.

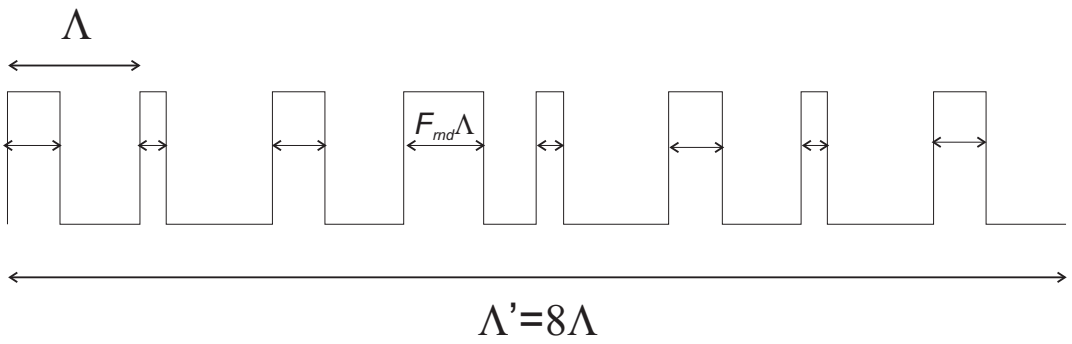

Figure 8.15: Illustration of a super-cell analog to the one used for evaluating the feature line rouhgness impact on performances. This example shows a $\Lambda' = 8\Lambda$ super-cell. $F_{rnd}\Lambda$ is a random variable defined by a gaussian distribution of variance σ^2. Here σ^2 is exaggerated for illustration purposes. In practice, cells of 32 or 64Λ have been used.

four surfaces will independently be affected by this type of roughness into account, Monte-Carlo simulations (see Fig. 8.14) led to a degradation of the mean null depth from the nominal value of 2.5×10^{-7} to 3.2×10^{-7} in the case of a 5-nm rms roughness and up to 5×10^{-7} in the 10-nm rms case.

Roughness in the smaller 1-10 μm scales concerns random errors on the lateral dimensions of the grating, i.e., the feature line. From the specifications of the direct writing laser apparatus (DWL) that is planned to be used in the fabrication of the present component (see Sect. 9.2), the expected line-width deviations from uniformity are of the order of $\sim$ 25 nm rms. The so-called super-cell method (Philippe Lalanne, private communication) was implemented to numerically estimate the influence of such errors. This tricky method is in fact derived from the application of RCWA to the study of rough inhomogeneous films (Giovannini & Amra 1997; Giovannini et al. 1998). Let us then define a super cell encompassing 32 grating periods, i.e., $32\Lambda \approx 30\,\mu$m. Next, each of the 32 feature lines is randomly affected by an error leading to the 25 nm rms uniformity specified by the DWL machine's vendor (Fig. 8.15). Each elementary period is then discretized into 64 parts so that the super-cell is discretized into $32 \times 64 = 2048$ data points. In order to ensure convergence in this particular case (see Annex A), we had to take $\sim$ 1000 orders into account, which is about two orders of magnitude more than in the nominal case. This has therefore involved very long computation times so that we could only sample a few points across the 6-14 μm wavelength range. The result of these heavy and time-consuming calculations was that the phase-shift quality seemed not to be affected by the considered type of roughness while the reflected amplitudes were. Indeed, micro-roughness induces diffusion in the reflected beam so that the TIR was not completely ensured for the TE and TM components, giving for instance, efficiencies of 99.91% for TE and 99.97% for TM at 6 microns in the $ZnSe$ double-rhomb case. Thanks to the super-cell method, we therefore managed to quantify the amplitude-mismatch influence on the performance (see Eq. 2.8) so that, in conclusion, the feature line roughness should only degrade the null depth by $\sim 2.25 \times 10^{-8}$.

Finally, the nanoscopic-scale roughness is not expected to induce significant perturbations of the performances in the thermal infrared since those tiny defects will be washed out by the large-scale periodicity (which is already sub-lambda) and the larger-scale defects treated hereabove.

Homogeneity

One of the drawbacks encountered when optical materials are used in transmission is that they should be homogeneous enough to keep, after transmission, the beam optical quality. It is evident

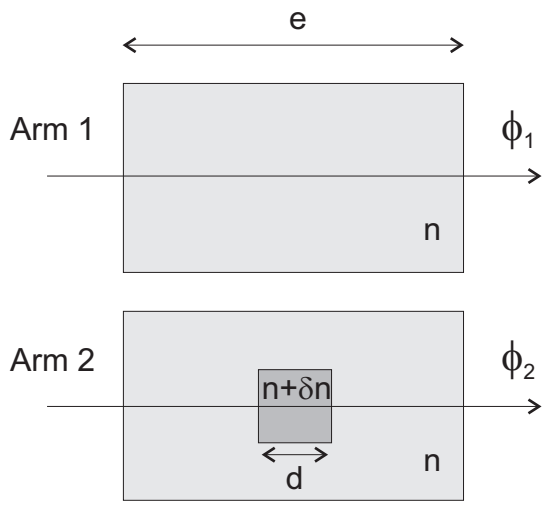

Figure 8.16: Phase variation induced by an index inhomogeneity δn of size d.

that the acceptable deviation from a perfect homogeneity depends strongly on the scale of the defects. Regarding the required null depth of 10^{-6} this may be evaluated in the following way (Mangin 2003): in a first approximation we consider that homogeneity defects alter the phase only, even if light scattering should also be considered. Both effects are reduced by performing optical filtering. Fig. 8.16 depicts the simple case where a beam crosses two plates, one of them exhibiting a defect of thickness d, index $n + \delta n$, for a plate of thickness e and average index n.

At a wavelength λ, the defect-induced phase difference can be written

$$\delta\phi = \phi_2 - \phi_1 = \frac{2\pi}{\lambda}[(n\ e + \delta n\ d) - n\ e] = \frac{2\pi}{\lambda}\delta n\ d \tag{8.5}$$

Over a complete optical path, Eq. 8.5 can be written

$$\delta\phi = \sum_i \frac{2\pi}{\lambda}\delta n_i\ d_i \tag{8.6}$$

Let us now compute the rejection rate that can be achieved with a two-beam nulling interferometer when this phase defect occurs. For the sake of simplicity, we will consider a cylindrical defect with radius r_{def} and uniform thickness $d = 2r_{def}$. From Eq. 2.4, we know that at the constructive output

$$I_{max} \approx 4I_0\pi r_{pupil}^2 \tag{8.7}$$

whereas at the destructive (or dark) one, we have

$$I_{min} \approx 2I_0\frac{\delta\phi^2}{2}\pi r_{def}^2 \tag{8.8}$$

The null depth therefore resumes to

$$N = \frac{I_{min}}{I_{max}} = \frac{2I_0\frac{\delta\phi^2}{2}\pi r_{def}^2}{4I_0\pi r_{pupil}^2} = \frac{\delta\phi^2 r_{def}^2}{4r_{pupil}^2} \tag{8.9}$$

substituting Eq. 8.5 into Eq. 8.9, we have

$$N = \frac{4\pi^2\delta n^2 r_{def}^4}{\lambda^2 r_{pupil}^2} \tag{8.10}$$

Given a null detph N, the maximum value for r_{def} is therefore

$$r_{def} = \left(\frac{\lambda^2 r_{pupil}^2 N}{4\pi^2 \delta n^2} \right)^{1/4} \tag{8.11}$$

Imposing $N = 10^{-7}$ and assuming the smallest wavelength $\lambda = 6\,\mu$m, a beam size of 15 mm so that $r_{pupil} = 7.5$ mm, and $\delta n \approx 3 \times 10^{-6}$ for the $ZnSe$ (homogeneity value routinely obtained), we have $r_{def,max} \approx 870\,\mu$m.

This value considers a single defect and is given for illustration. Of course several defects may affect a pupil. In that case, the budget has to be performed over all the defects. In the actual Darwin instrument, monomode optical filtering will further decrease very significantly the impact of such defects on the wavefront. During the present study, as no such modal filtering is available, we will therefore assume a classical pinhole filtering which diameter is a fraction $1/k$ of the Airy disk at 6 μm. Now, we estimate the improvement provided by this filtering and the resulting features of acceptable defects. In that case, the gain in terms of rejection rate can be computed, according to the size of the defect. In the focal plane, the PSF of a defect n times smaller (in diameter) than the pupil size will be n times larger (still in diameter) (this is a classical property of Fourier transforms, Ollivier & Mariotti 1997; Mennesson et al. 2002). As a consequence, the amount of defect energy kept after filtering is n^2 smaller than the pupil energy. In addition, the fact that the filter is about k times smaller than the Airy disk increases the nulling depth by an additional factor k. Thus one can write

$$N_{\text{after filtering}} \approx \frac{r_{def}^2}{k\ r_{pupil}^2} N_{\text{before filtering}} = \frac{4\pi^2 \delta n^2 r_{def}^6}{k \lambda^2 r_{pupil}^4} \tag{8.12}$$

The maximum value for r_{def} then becomes

$$r_{def} = \left(\frac{k \lambda^2 r_{pupil}^4 N}{4\pi^2 \delta n^2} \right)^{1/6} \tag{8.13}$$

Imposing again $N = 10^{-7}$ and assuming the smallest wavelength $\lambda = 6\,\mu$m, a beam size of 15 mm so that $r_{pupil} = 7.5$ mm, $\delta n \approx 3 \times 10^{-6}$ for the $ZnSe$ inhomogeneity and $k = 8$ (Mennesson et al. 2002), we have $r_{def,max} \approx 2.5$ mm, which is quite comfortable since the micro-crystals of CVD $ZnSe$ are known to be $\sim$ 40-60 μm in size.

Putting it the other way around is more significant since even without spatial filtering, the contribution to the null depth of a single $\delta n \approx 3 \times 10^{-6}$ defect of $\sim$ 40-60 μm is only of $\sim 10^{-12}$ (10^{-18} with spatial filtering).

8.5.5 Rhombohedra design

Some specific macroscopic tolerancing issues will now be considered in this section. The micro-structure sensitivity analysis conclusion was that the tolerance on the grating parameters mainly resumes to the sole fine control of the incidence polar angle which is in turn alleviated by the use of the double-rhomb configuration. However, the previous analysis neglected parasitic-reflection phenomenons due to the rhomb in/out interfaces. Thus, the macroscopic components have still to be optimized for stray-light dumping. Broadband anti-reflective treatments will be necessary as well as appropriate wedge angles at the different interfaces as we will discuss. For that, we decided to model the double-rhomb geometry (see Fig. 8.17) in order to perform a ray-tracing analysis.

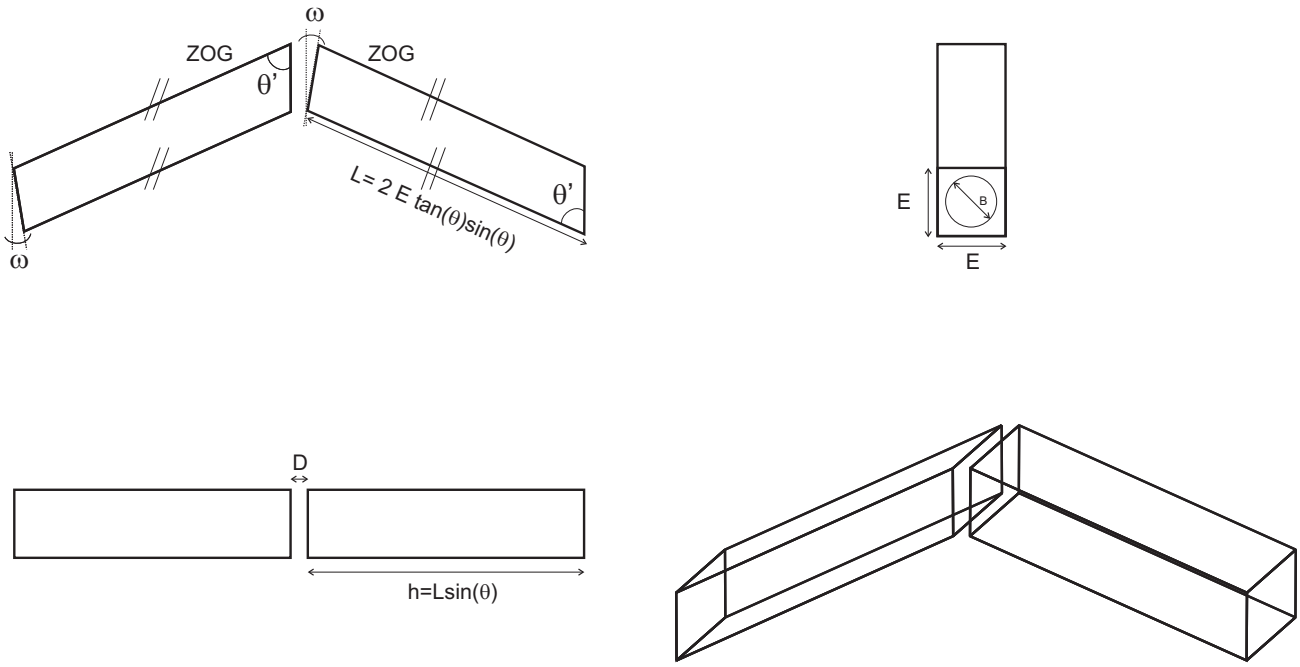

Figure 8.17: Double-rhomb geometrical scheme and definition of the wedge angle w.

Ray tracing

Ray-tracing analysis was performed with the ZEMAX optical analysis software[29]. The loaded geometry was that of the nominal $ZnSe$ double-rhomb configuration (Fig. 8.17). First results were expectedly affected by strong parasitic light issues originating from the parallelism of the in/out interfaces of the two rhombs and the relatively high refractive index of $ZnSe$ ($n \approx 2.4$, leading to a reflection coefficient of $\sim 17\%$ per interface). We therefore immediately chose to adopt the wedge solution with the main purpose of ejecting spurious reflections away from the main beam. Indeed, wedging the in/out interfaces with respect to each other by an angle w (see Fig. 8.17) allows deviating higher-order parasitic beams of at least

$$\theta \geq 2nw \tag{8.14}$$

with n, the refractive index of the material. The specification of the wedge angle w is not trivial and requires a fine analysis as we will discuss.

First, the wedge angle must be large enough so that the spurious reflections are rejected. Rejection occurs when the coupling efficiency of the output beam into the single-mode waveguide (or the pinhole) that acts as a wavefront filter becomes negligible. When a beam is tilted with respect to the maximum coupling into the single-mode waveguide, the coupling efficiency decreases with the distance of the Airy image to the center of the guide. In order to reject stray light, the overall transmission efficiency of the parasitic beam must be significantly below the nulling performance, i.e., $\sim 10^{-7}$. This transmission efficiency results from the throughput of the rhomb interfaces, and on the coupling efficiency into the single-mode waveguide. From numerical computations, we derive that the distance d between the center of the single-mode guide and the image of the parasitic beam must exceed k=25 times the mode size A in order to attenuate the parasitic deviated beam well below the 10^{-7} level. According to Ruilier (1999), the latter depends on the coupling optics focal length f, the wavelength λ, and the beam diameter B as follows

$$A = 0.71\lambda f/B \tag{8.15}$$

[29]Results presented is this section have independently been confirmed by Jérôme Loicq (CSL) on ASAP.

From $A = f\theta$, we finally derive the minimum value of θ

$$\theta > 0.71k\lambda/B \tag{8.16}$$

which in turn provides the minimum wedge angle w

$$w > 0.355k\frac{\lambda}{nB} \tag{8.17}$$

For example, if $B = 15$ mm and $\lambda = 14\,\mu$m, $w > 3.5$ mrad, i.e., ~ 10 arcmin.

On the other hand, the disadvantage of having wedged plates is that they introduce deviation and dispersion. Deviation being, by definition, identical at all wavelengths, it can be compensated for by the adjustment of a mirror, for all wavelengths at the same time. The maximum wedge-angle requirement is therefore driven by dispersion. As written above, the deviation of a beam translates into a variation of the coupling efficiency. Dispersion therefore creates a differential coupling efficiency according to wavelength. In principles, if both effects are identical in the two interferometer arms, the consequence on the nulling is negligible. Moreover, the double-rhomb configuration, if perfectly anti-symmetrical as depicted in Fig. 8.17, allows cancelling the deviation and therefore dispersion since the wedges are assumed identical for both rhombs, in the same material and opposed.

In practice, wedging the plate at the minimum ~ 10 arcmin specification is not sufficient. This result was at first very surprising given the reasoning presented here above. The explanation is nevertheless very logical. Indeed, the analytical analysis does not take interactions between the two rhombs into account. These multiple interactions prevent certain parasitic rays from being deviated outside the main beam. Empirical ray-tracing analysis has led us to fix the wedge angle at 1° for efficiently reject the stray light as can be seen in Fig. 8.18. Such a substantial value is nevertheless a compromise, leaving some ghost residuals in the pupilla (at the 10^{-4} level). The reason for the trade-off is to prevent interfaces from being too tilted with respect to the main beam. Indeed, transmission through a tilted surface leads to differentiation between s and p perpendicular polarization components of the impinging light. Above one degree, the latter would lead to polarization disturbances above the required level of 10^{-7} (for 1°, the exact disturbance taking the 4 interfaces into account is 3.4×10^{-8}). In any case, stray-light residuals can be prevented thanks to the use of almost classical anti-reflection coatings with a reasonable specification on the reflectance which has to remain below 2% over the working wavelength range of 6-14 μm (see here below).

Let θ_1 (resp. θ_2) be the angle of incidence upon the TIR interfaces of the first (resp. second) rhomb (see Fig. 8.4). Another consequence of the wedge-angle presence is that upon refraction through the input face, the TIR angle on the two first TIR interfaces θ_1 is disturbed by the angle $\epsilon(\lambda) = \arcsin\frac{1}{n(\lambda)}\sin w$, leading to $\theta_1 = \theta + \epsilon$ in the first rhomb while in the second no perturbation occurs because of the deviation compensation provided by the anti-symmetrical wedge implementation, so that $\theta_2 = \theta$. This effect must be corrected by manufacturing the rhomb with an effective angle θ' different from the designed TIR angle θ. The required transformation restoring symmetry is $\theta' = \theta - \epsilon/2$. Doing this, in the nominal configuration, we have $\theta_1 = \theta + \epsilon/2$ and $\theta_2 = \theta - \epsilon/2$, and the linearity of the phase shift with respect to the incidence angle allows a quasi-perfect compensation, provided that θ is centered on the optimal nominal value. Once again, it is worth emphasizing the robustness of the double-rhomb geometry that allows relaxing incidence-related design obstacles.

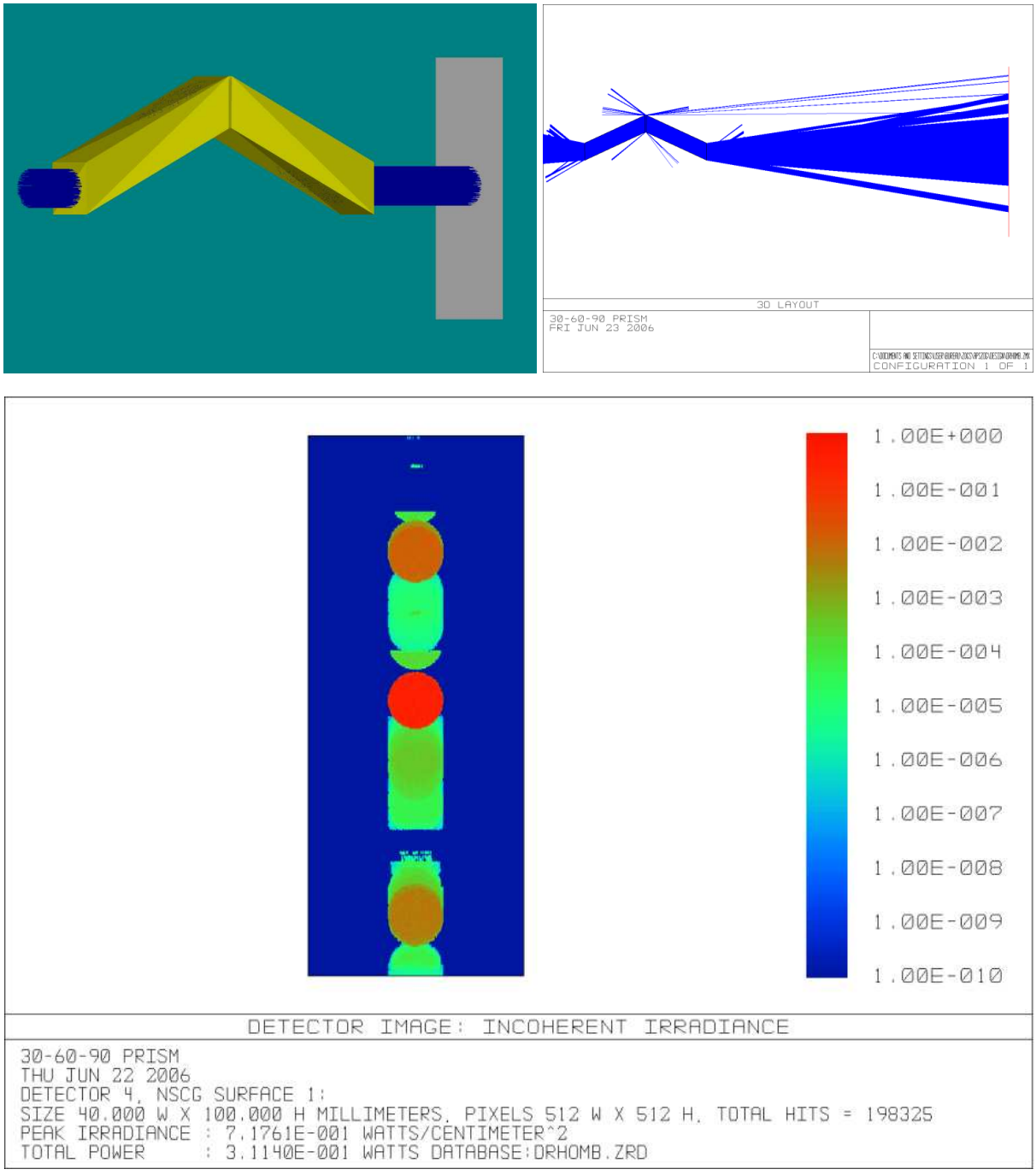

Figure 8.18: Ray-tracing analysis of the *ZnSe* double-rhomb with a wedge angle $w = 1°$. Main ghosts (orange) are ejected sufficiently far from the main beam (red). Residuals at the $10^{-3} - 10^{-4}$ level are spread around the optical axis. Some of them reenters the pupil near the edges. Isolation shows those penalizing rays undergoing at least four reflections. This means that a 2% anti-reflection coating is sufficient to level them down to the 10^{-7} specification.

Comments about anti-reflective treatments

One of the conclusions of the stray-light analysis presented here above is the necessity of treating the different input/output faces of the rhombs by appropriate anti-reflective structures. Although subwavelength gratings are mostly appropriate for this purpose (see Sect. 4.2), we have chosen classical multi-layer coatings for two main reasons:

- these traditional treatments are widespread and proposed by many manufacturers for the thermal infrared (8-12 μm range);
- they are currently cheaper than subwavelength gratings that still need further R&D for practical implementation and cost reduction.

We firstly propose to use the ZnS-YF_3 couple as coating materials for $ZnSe$ substrates (Fig. 8.19). Refractive indices used for calculations are given in Annex B for ZnS, and are taken from Lemarquis et al. (1998) for YF_3. It is to be noted that thin film refractive indices, which depend on deposition technologies and parameters, may be slightly different so that it becomes difficult to choose one dispersion law rather than another. Now, even if an accurate knowledge of thin film indices is mandatory for manufacturing, the optimum performance is not very sensitive to a slight index deviation, provided that the layer thicknesses are re-optimized.

The case of YF_3 is a bit different since fluoride materials in thin film form are known to be sensitive to moisture. According to the deposition parameters, fluoride films are more or less porous, and water adsorption can be observed. This induces absorption bands in the infrared spectral range that are characteristic of water, especially at 2.9 μm, 6 μm and beyond 11 μm. Obviously, this absorption will decrease coating transmittance. To take this phenomenon into account, we used dispersion laws for both the refractive index and the extinction coefficient (Lemarquis et al. 1998). It is worth mentioning that these calculations still demand optimization from thin film specialists. The results presented in this paragraph are preliminary and for illustration purposes only. Knowing this, let us present another solution using BaF_2 instead of YF_3 (Fig. 8.20), alleviating

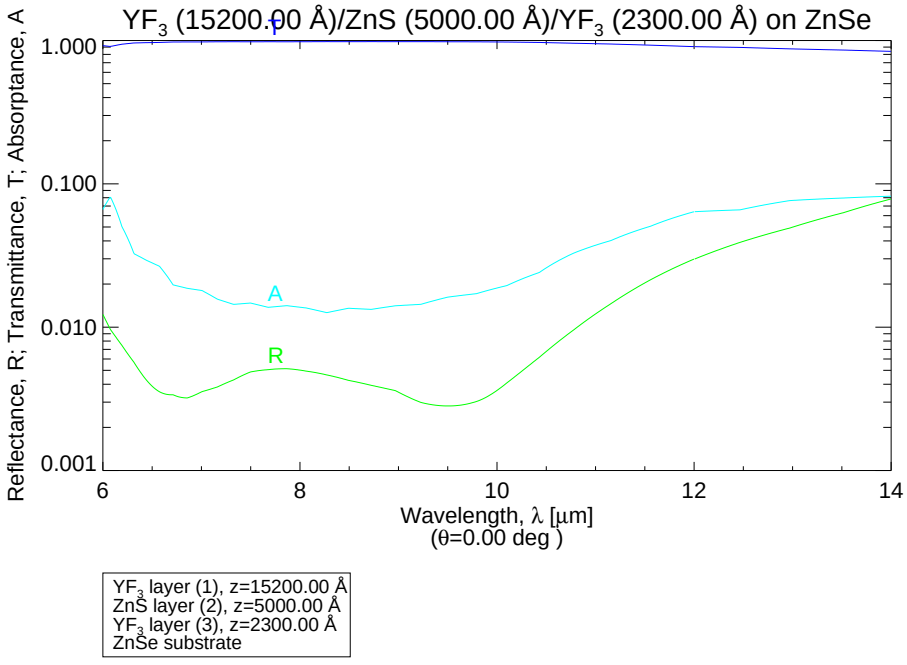

Figure 8.19: 3-layer ZnS-YF_3 anti-reflective coating for $ZnSe$ for the 6-14 μm wavelength range. Absorption due to YF_3 is quite significant. Courtesy of Pierre Riaud (IAGL).

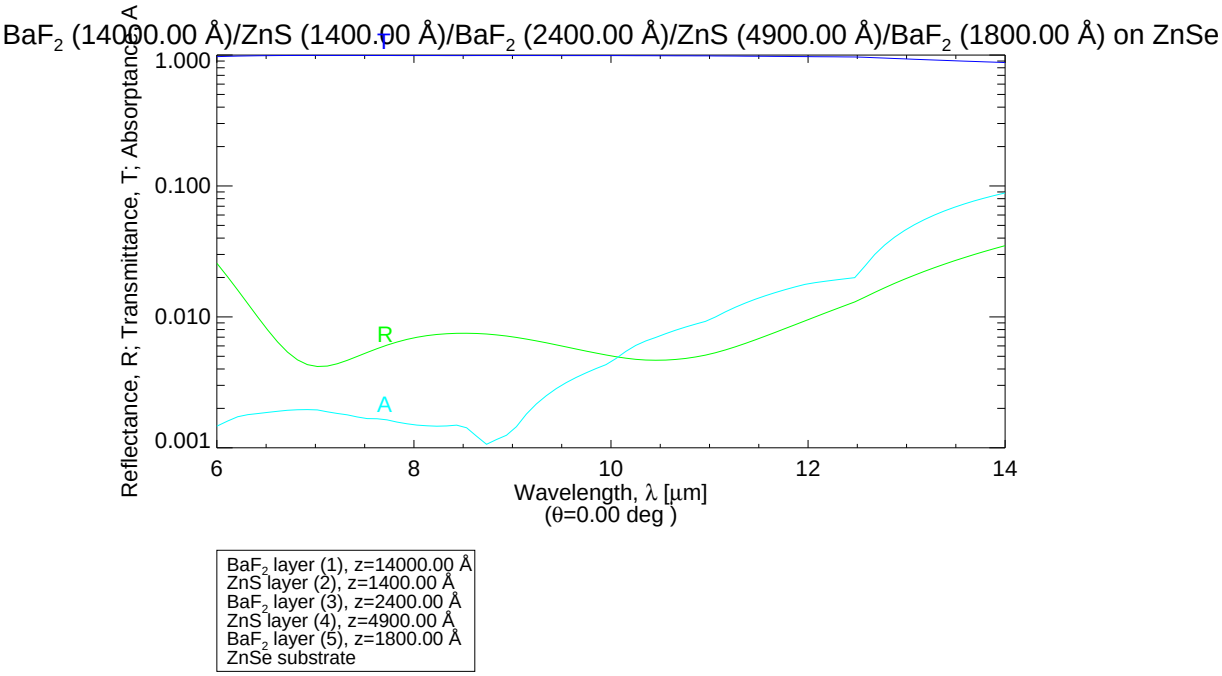

Figure 8.20: 5-layer ZnS-BaF_2 anti-reflective coating for $ZnSe$ for the 6-14 μm wavelength range. Courtesy of Pierre Riaud (IAGL).

short-wavelength absorption issues.

In nulling interferometry, identical components have to be placed on each arm of the interferometer in order to maintain a perfect balance for both phase and intensity properties. To obtain such a result for multi-layer coatings, similar coatings must be manufactured simultaneously during a single-deposition process. Similar coatings will have identical optical properties if the layer thicknesses are rigorously identical. However, since the samples cannot be located at the same place in the deposition chamber, a slight thickness mismatch can appear between the coatings. These considerations have specifically been studied for Darwin in Lemarquis (2003) with the conclusions that lateral uniformity of the coating does not seem to be a critical problem. On the contrary, radial uniformity should be regarded carefully. Taking the uniformity of standard deposition chambers into account, and considering a simple rotation of the substrates during deposition, coatings cannot be considered as uniform for the Darwin requirements. In that case, the orientation of the components inside the deposition chamber must be taken in account in order to have a correct compensation of phase shifts between the two arms of the interferometer. Another solution to suppress this constraint is to use planetary movements of the substrates during deposition. However, this solution is not of frequent use for manufacturers.

Final geometry of the prototype

The prototype was chosen to be a $ZnSe$ double rhomb engraved with two subwavelength gratings, one on each rhomb, leaving one bare TIR interface per rhomb. The prototype was optimized for the 6-14 μm wavelength range and not above because of the absorption of $ZnSe$. The beam size is assumed to be 15 mm in diameter, according to the NULLTIMATE test bench specification (see Sect. 9.3.3).

Assuming a rectangular profile for the baseline subwavelength grating (period $\Lambda = 900$ nm, feature line of 250 nm and thickness of 1.23 μm), the optimal TIR incidence polar angle was

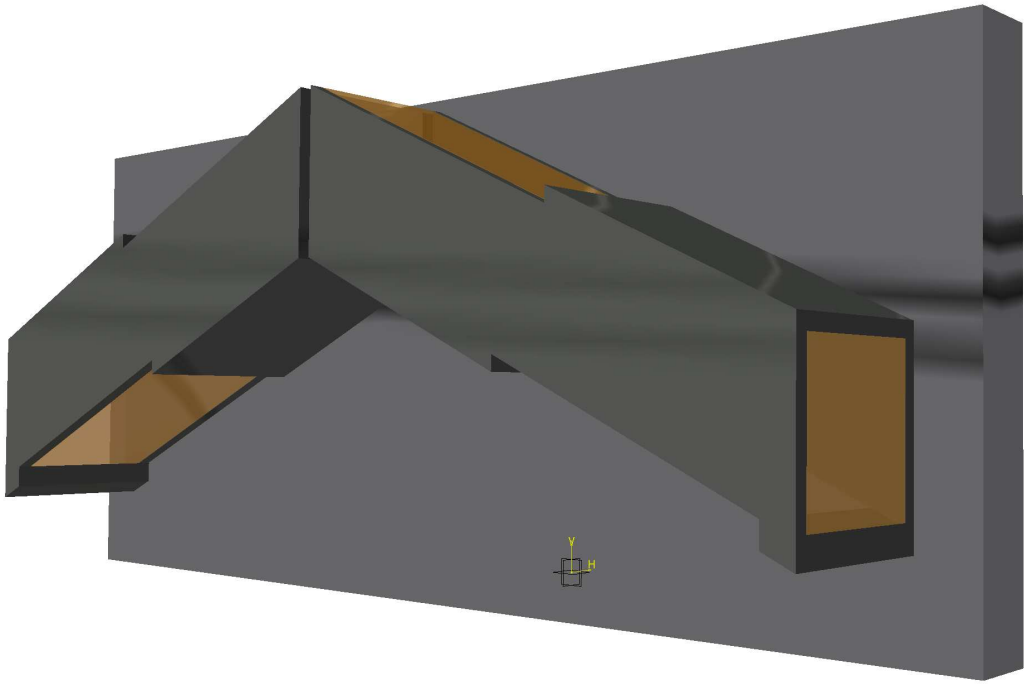

Figure 8.21: 3D view of the *ZnSe* double-rhomb TIRG APS prototype, and draft of the mechanical mount under optimization for alleviating thermal and mechanical stresses.

found by RCWA analysis to be 1.1396 radian (see Fig. 8.9, right). Since the stray-light analysis led to wedges of 1°, the latter must be corrected accordingly as explained here above, leading to a rhomb angle of 1.1352 radian. Taking a comfortable clearance margin, the input faces are chosen to oversize the beam diameter of the NULLTIMATE test bench (15 mm) by several millimeters: we will take 20 mm x 20 mm. The geometry is now completely constrained and final drawings defining the rhomb dimensions can be made (see Annex C). An artistic 3D view of the component mounted in a rough draft of mechanical mount is presented in Fig. 8.21.

Mechanical mount

In this paragraph, we shall draw some specific design-dependent constraints on thermal and mechanical aspects to be taken into account in the mechanical-mount conception (see the rough draft presented in Fig. 8.21).

Rhomb spacing. The wedge-induced deviation of the chief beam at the output of the first rhomb is a function of wavelength. As a direct consequence, the second rhomb thickness "seen" by each of the wavelengths is different. The amplitude of the thickness variation depends on the spacing D between the two rhombs (Fig. 8.17). Nevertheless, if this spacing is exactly the same in the two arms of the interferometer, whichever its actual value, no performance loss occurs since the interferometer remains symmetrical. When the spacing differs between the arms, a differential thickness difference appears, which cannot be compensated for nor by the delay lines nor by the thickness adjustable beam splitters.

Let us derive the constraint on the differential spacing δD. The thickness difference $e_{\lambda_1} - e_{\lambda_2}$ in the second rhomb between two beams of wavelengths λ_1 and λ_2 is related to the spacing D, the angular dispersion of the beam going out of the first rhomb ξ and the wedge angle of the second

one w

$$e_{\lambda_1} - e_{\lambda_2} = wD\xi \tag{8.18}$$

Differentiating the former equation with respect to D leads to

$$\frac{\partial e}{\partial D} = w\xi \tag{8.19}$$

Let us now assess the rhomb thickness variation δt which generates a nulling degradation of 10^{-8} at a given wavelength λ. This value directly comes from the sensitivity of the null depth to the OPD, and the relation between the OPD and the rhomb dimension, as follows

$$\delta t = \frac{\lambda\sqrt{N}}{\pi(n-1)} \tag{8.20}$$

For the overall null depth to remain insignificantly degraded, the differential thickness difference δe must remain lower than δt. If we suppose that the bench configuration (delay lines, beam splitters, etc.) is optimized for the chief ray, then ξ is taken to be equal to the PTV deviation divided by 2. A numerical application for the 6 μm wavelength and the $ZnSe$ 1°-wedged double rhomb yields $\delta t = 0.136$ nm, $\xi_{ptv} = 6 \times 10^{-5}$ rad, $w = 1$ degree, so that

$$\delta D < 0.1\,\text{mm} \tag{8.21}$$

Temperature effects. The prototype will first be tested at room temperature. It has been optimized accordingly. However, it is planned to be tested at cryogenic temperatures as well on the NULLTIMATE testbed, at 100 K precisely (see Sect. 9.3.3). Three effects of temperature will be considered:

- refractive-index variations with temperature;
- thermal expansion of the component;
- heat transfer during cooling.

Once again, the double-rhomb configuration robustness allows for a comfortable tolerance as far as refractive-index variations are concerned since cooling will impact on the performance through a mere global shift in the $ZnSe$ refractive index while conserving the global achromaticity. Indeed, the dispersion of the thermo-optic coefficients of $ZnSe$ is low (see Annex B). Therefore, a simple geometrical modelling of the main beam using the Snell-Descartes refraction law at each interface coupled to the RCWA code simulating the grating response has led to the conclusion that adjustments within a 10-arcmin range are sufficient for switching between the cold and hot configuration of the double-rhomb TIRG APS. This modelling also take chromatic refraction into account, and departure from nominal performance was found to be kept within a few 10^{-8} between the two configurations. The lesson of this analysis is that the mechanical mount of the double rhomb must allow varying the angle between the two rhombs with a subarcmin precision, or that two different mounts are needed. The two interferometer arm mounts are planned to be implemented on a 3-rotation axis deck so that the general orientation of the double rhomb as a whole can be controlled independently. This control is convenient for possible OPD adjustment though the NULLTIMATE test bench is expected to be equipped with $ZnSe$ beam splitters which are adjustable in thickness (see Sect. 9.3.3).

Thermal dilatations/contractions will only impact the geometry of the grating to a negligible level since the thermal expansion coefficient γ of $ZnSe$ at ambient temperature is 7.1×10^{-6}/K leading for the typical dimension of the grating, i.e., $\sim 1\,\mu$m, and a $\Delta T = 200$ K to a dilatation/contraction of only $10^{-6} \times 7.1 \times 10^{-6} \times 200 \approx 1$ nm. As far as the dilatation/contraction

of the whole rhomb is concerned, the mount has to be conceived for tolerating differences of $0.08 \times 7.1 \times 10^{-6} \times 200 \approx 0.12$ mm. In this respect, the choice of the mount material is of critical importance. Indeed, it must be chosen with a thermal expansion coefficient very similar to that of $ZnSe$ in order to avoid differential effects. Let us suggest the use of Titanium (for example, the $TA6V$ alloy) since its thermal expansion coefficient is 8×10^{-6}/K, i.e., very close to $ZnSe$, just like its thermal conductivity: 0.16 W/cm/K. It is to be noted that the thermal expansion coefficient is actually temperature-dependent, which was neglected here above. It indeed decreases with the temperature. Our calculations are therefore conservative.

The cooling time necessary to lower the temperature of a $ZnSe$ rhomb from 298 K to 100 K thanks to conduction can be estimated by simple heat-transfer considerations and knowing the $ZnSe$ thermal properties:

- heat capacity $\rho_c = 0.356$ J/gr/K;
- thermal conductivity $k = 0.18$ W/cm/K.

Newton's law of cooling indeed states the following

$$T(t) = T_{env} + (T(0) - T_{env})\, e^{-t/\tau} \tag{8.22}$$

where $T(t)$ is the rhomb temperature at time t, $T(0)$ is its initial temperature, T_{env} is the temperature of the environment of the rhomb (i.e., its thermal contact), and τ is a characteristic time defined as $\tau = \frac{\rho_c M}{Ak/d}$, with M the total mass of the rhomb, d its thickness, and A the area of the thermal contact surface. In the present case, since the density of $ZnSe$ is $\sim 5.2\,\mathrm{gr/cm^3}$, the mass of a single rhomb is 166 gr, its thickness is ~ 2 cm, and the contact surface is assumed to be a lateral one, i.e., with an area of $\sim 16\,\mathrm{cm^2}$. A numerical application with $T(0) = 298$ K and $T_{env} = 100$ K give $\sim$ 300 s for cooling the component down to 100 K $\pm$ 0.1 K. This is of course a raw calculation with the sole purpose of demonstrating that the characteristic cooling time of a $ZnSe$ rhomb is much smaller than the expected cooling time of the bench (several days) assuming that the contact is at least made on one lateral face of the rhomb. However, it is worth noting that cooling time is also driven by the admissible strain caused by internal thermal gradients. When the surface of a solid body is subjected to a temperature change ΔT, a compressive stress $\Delta T \gamma E/(1-\nu)$ indeed appears (Roark's formulas, Young 1989), with $E = 6.72 \times 10^{10}\,\mathrm{N/m^2}$ the $ZnSe$ Young modulus, and $\nu = 0.28$ the $ZnSe$ Poisson ratio. This stress must remain below the rupture modulus of $ZnSe$, which is $5.51 \times 10^7\,\mathrm{N/m^2}$. The maximum admissible internal temperature gradient is therefore $\Delta T_{max} \leq 83$ K, which is quite comfortable.

Now, we possess all the necessary tools and data for realistically calculating a minimum cooling time respecting the maximum admissible thermal gradient of 83 K. The heat to extract from the $ZnSe$ rhomb of 166 gr to cool it from 298 K down to 100 K ($\Delta T_{tot} \approx 200$ K) amounts to $Q = \rho_c M \Delta T_{tot} \approx 1.1 \times 10^4$ J. In a first step, let us impose the $\Delta T_{max} = 83$ K gradient to cool the component down to 183 K, or equivalently to evacuate $Q' \approx 6800$ J. The needed cooling time for that resumes to $t = dQ'/(kA\Delta T_{max}) \approx 57$ s. Then, using the Newton's law (Eq. 8.22), the remaining time to pass from 183 K to 100 K $\pm$ 0.1 K, is $\sim$ 276 s, leading to a total minimum cooling time of 333 s, which remains much smaller than the expected cooling time of the test bench.

A more detailed study of sensitivity to temperature gradient or variability should rather define the temperature regulation within the bench than the design of the APS by itself. For more details about these particular considerations, see Valette (2004). The conclusion of the latter work is that thermalization of the support is strongly improved by good thermal screening if thermal conductivity of the component is high and temperature low. Shape of optical components should not vary if good thermal shielding is achieved.

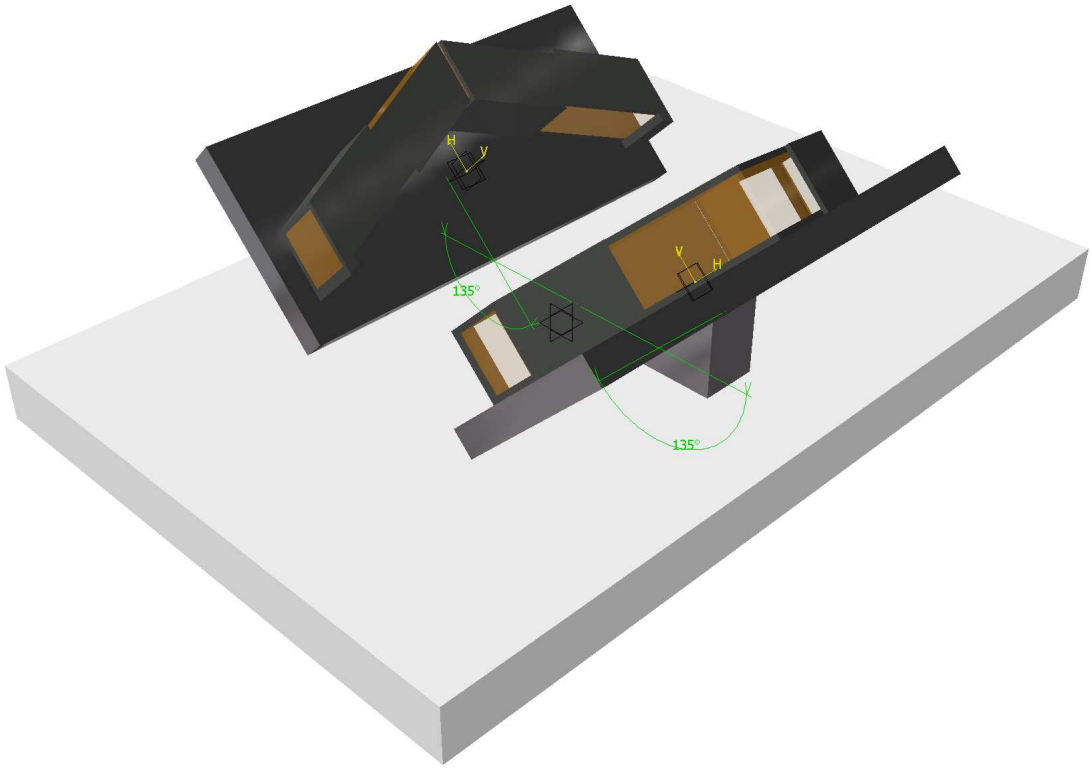

Figure 8.22: *ZnSe* double-rhomb TIRG APS prototype. Draft of mechanical implementation at 45° for alleviating stress-birefringence issues by letting the rhombs laying under their own weight. The mechanical mount seems to enclose them but there is in fact no contact. The apparent enclosure is just to prevent them from falling.

Stress-induced birefringence. Stress birefringence can be a strong obstacle to the final performance of the TIRG APS. Indeed, the macroscopic optical length inside the rhomb is as long as $\sim$ 17 cm. *ZnSe*, by its cubic and polycristalline nature, is not birefringent but as every existing material, when submitted to stresses, it becomes anisotropic. Stress-optic coefficient of *ZnSe* in the thermal infrared has been measured by Mariner & Vedam (1981): $C_{\lambda=10.6\,\mu\mathrm{m}} \approx -12$ brewsters[30]. The stress-optic coefficient is directly related to the stress-induced birefringence by

$$\Delta n = C_\lambda \sigma \tag{8.23}$$

where σ is the applied uniaxial stress. Imposing that the phase shift perturbation must remain below the 10^{-3} radian level at 6 μm[31], i.e., below the $\sim 10^{-7}$ null depth, implies that an uniaxial applied load must remain below $\sim 470\,\mathrm{N/m^2} = 47\,\mathrm{Pa} = 4.7\,\mathrm{gr/cm^2}$ which is very stringent and should be a sufficient motivation for a finite element analysis of the rhomb stresses in its working conditions. Indeed, applied loads will be localized and the propagation of their associated stresses in the bulk material up to the optical zones are *a priori* unknown. Moreover, the weight of a single rhomb is 166 gr. Assuming that the rhombs will lay on their whole lateral face of $\sim 16\,\mathrm{cm^2}$, this nevertheless makes it twice too strong. Manufacturing mounts such that both rhombs lay at a 45° angle from the horizontal in the bench should relax this constraint roughly by distributing the rhomb weight on the perpendicular face where one half can be sacrificed since no light impinges

[30] 1 brewster $= 10^{-12}\,\mathrm{m^2/N}$.

[31] We will assume that the stress-optic coefficient does not depend on λ.

on it (see Fig. 8.22). A preliminary finite-element analysis was therefore performed (see Annex D). The conclusion of this quick analysis is that the expected phase-perturbation upper bound resumes to $\sim 6 \times 10^{-4}$ radian at 6 μm, corresponding to a null-depth degradation of $\sim 1.6 \times 10^{-7}$.

Of course, these problems are not expected in space. In any case, they must be taken into account by the mechanical mount conception but also during manufacturing, i.e., cutting and polishing of the rhombs, in order to avoid residual stresses. For that, the elastic limit of *ZnSe* (55.1 MPa) must not be exceeded, which, actually, should not be a problem.

8.6 Summary

As the conclusion of the present chapter, we will tentatively draw an error budget for the *ZnSe* 6-14 μm double-rhomb TIRG APS summarizing all the points described in the different sections here above. The hypothesis of independency of every contribution to the resultant null depth is of course questionable but this is quite a conservative way of assessing the final performance of the component. The error budget is constructed on the basis of the nominal performance of 1.5×10^{-7} resulting from the sole optimization of the *ZnSe* subwavelength grating in the double-rhomb configuration (2 TIRG and 2 TIR interfaces) for the 6-14 μm wavelength range. Two situations are considered: an optimistic and a pessimistic one. The main difference between them concerns the degree of mastership in the control of the micro-structure design parameters.

Table 8.2: *ZnSe* 6-14 μm double-rhomb TIRG APS tentative error budget.

	Optimistic	**Pessimistic**
ZOG control	2.5×10^{-7} (2° slope)	5×10^{-7} (5° slope)
Thickness roughness	7×10^{-8} (5 nm rms)	2.5×10^{-7} (10 nm rms)
Lateral roughness	2.25×10^{-8} (25 nm rms)	2.25×10^{-8} (25 nm rms)
Homogeneity	10^{-18} (with spat. filt.)	10^{-12} (w/o spat. filt.)
Polarization disturbance (1° wedge)	3.4×10^{-8}	3.4×10^{-8}
Ghost	10^{-7}	10^{-7}
Rhomb spacing (<0.1 mm)	10^{-8}	10^{-8}
Temperature effects	10^{-8}	10^{-8}
Stress birefringence	10^{-7}	10^{-6}
Total	6×10^{-7}	2×10^{-6}

8.7 Article: *Use of subwavelength gratings in total internal reflection as achromatic phase shifters*

In this paper, published in the peer-reviewed journal *Optics Express*, the idea of using subwavelength gratings in total internal reflection was exposed for the first time. The principle of the TIRG APS is described and its theoretical performances are assessed. Emphasis is put on the phenomenological aspects of the modulated TIR. Some manufacturing hints are also given.

Use of subwavelength gratings in TIR incidence as achromatic phase shifters

Dimitri Mawet, Cédric Lenaerts, Pierre Riaud and Jean Surdej

University of Liège, 17 allée du 6 Août,
B-4000, Sart Tilman, Belgium

mawet@astro.ulg.ac.be

Serge Habraken and Denis Vandormael

Centre Spatial de Liège, Avenue du Pré-Aily,
B-4031, Liège-Angleur, Belgium

Abstract: Nulling interferometry constitutes a very promising technique in observational astrophysics. This method consists in attenuating the signal of a bright astrophysical object in order to detect much fainter nearby features, e.g. exoplanets around their host star. An on-axis destructive interference is created by adjusting the phases of the beams coming from various telescopes. The huge flux ratio between the parent star and the planet (10^6 in the thermal infrared) requires unprecedented high performance broadband phase shifters. We present a new design for these key components called Achromatic Phase Shifters (APS). We propose to use subwavelength diffractive optical elements under total internal reflection (TIR) incidence. Our component can be seen as an evolution of the Fresnel Rhomb technology.

OCIS codes: (120.3180) Interferometry; (050.5080) Phase Shift.

References and links

1. R.N. Bracewell, "Detecting Non Solar Planets by Spinning Infrared Interferometer," *Nature* **274,** 780–781 (1978).
2. A. Léger, J.M. Mariotti, B. Mennesson, M. Ollivier, J.L. Puget, D. Rouan, J. Schneider, "Could We Search for Primitive Life on Extrasolar Planets in the Near Future," *Icarus* **123,** 249–255 (1996).
3. R. Anderson, "Quarterwaveplate and Fresnel rhomb compared in the 10-Mum CO2 laser emission region," Appl. Opt. **27,** 13, 2746–2747 (1988).
4. K. B. Rochford, A. H. Rose, P. A. Williams, C. M. Wang, I. G. Clarke, P. D. Hale, G. W. Day, "Design and performance of a stable linear retarder," Appl. Opt. **36,** 25, 6458–6465 (1997).
5. M. Born & E. Wolf, "Reflection and refraction of a plane wave," in *Principles of Optics* eds. (Cambridge University press, seventh edition, 1997), pp. 49–53.
6. F. Peter, *Z Phys* **15**, 358-368 (1923).
7. W. J. Tropf, "Temperature-dependent refractive index models for BaF2, CaF2, MgF2, SrF2, LiF, NaF, KCl, ZnS and ZnSe," Opt. Eng. **34**, 1369-1373 (1995).
8. Gary J. Hawkins, *Spectral Characterisation of Infrared Optical Materials and Filters* (PhD Thesis - The University of Reading UK, 1998).
9. M.G. Moharam and T.K Gaylord, "Rigourous Coupled-Wave Analysis of planar grating diffraction," J. Opt. Soc. Am. **71**, 811-818 (1981).
10. P. Lalanne, P. Pichon, P. Chavel, E. Cambril, H. Launois, "Interferometric Characterization of Subwavelength Lamellar Gratings," Appl. Opt. **38,** 23, 4980–4984 (1999).

#7908 - $15.00 USD
Received 23 June 2005; revised 28 July 2005; accepted 25 August 2005

1. Introduction

Direct detection and characterization of faint sources around bright astrophysical objects like stars or Active Galactic Nuclei is very difficult due to the large flux ratio between them. For example, an Earth-like exoplanet is typically 10^9 times fainter than its host star in the visible spectrum, 10^6 in the thermal infrared. Infrared nulling interferometry proposed by R. Bracewell in 1978 [1], appears to be the most promising technique to achieve the high angular resolution and high dynamic range required to allow the ambitious detection of the first exobiological tracers. The nulling interferometry technique consists in adjusting the phases of the beams coming from various telescopes (two in the most simple configuration) to produce a fully destructive interference on the optical axis. The quality of the destructive interference, or the so-called Null Depth (ND) relies on the optical components ability to induce a very precise phase shift (e.g. π) and a very low amplitude mismatch over the considered wavelength range. Unfortunately, searching biomarkers in the exoplanet atmospheres requires spectroscopic characterization over large spectral bands. For example, the Darwin Infrared Space Interferometer [2] considered by the European Space Agency will operate in a wavelength band between 6 and 18 microns. High performance Achromatic Phase Shifters (APS) are therefore needed. For such broadband interferometers, we propose to use subwavelength gratings in a total internal reflection (TIR) configuration. These original components fit in the technological evolution of the well-known Fresnel Rhomb technology [3] [4]. Our theoretical calculations using the Rigorous Coupled Wave Analysis point to very promising results.

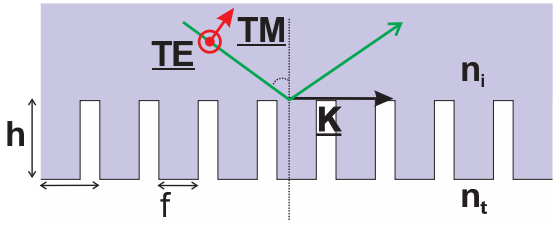

Fig. 1. Schematic representation of a subwavelength grating. The main parameters of the structure are: the grating vector $|\underline{K}| = 2\pi/\Lambda$, perpendicular to the grating lines, with Λ being the spatial period, the grating depth h and the filling factor f, such that $f\Lambda$ is the width of the grating ridges. $\underline{TE}$ and $\underline{TM}$ are the vectorial orthogonal polarization components of the θ-incident light. n_i and n_t are the refractive indices of the incident (substrate) and emergent (transmitting) media, respectively.

2. Principle of the TIRG APS

It is a well-known fact that the TIR phenomenon comes with a differential phase shift between the vectorial TE and TM polarization components. This vectorial phase shift takes the following form [5]

$$\Delta\phi_{TE-TM} = 2\arctan\left[\frac{\sqrt{\sin^2\theta - n_{ti}^2}}{n_{ti}^2\cos\theta}\right] - 2\arctan\left[\frac{\sqrt{\sin^2\theta - n_{ti}^2}}{\cos\theta}\right] \quad (1)$$

where θ is the angle of incidence, greater or equal to θ_c, the critical angle defined as $\sin\theta_c = n_{ti} = n_t/n_i$ and where n_i and n_t are the refractive indices of the incident and emergent media, respectively. This property is exploited in the well-known APS component called Fresnel Rhomb. We demonstrate that engraving a subwavelength grating on the TIR interface leads to a significant improvement over the Fresnel Rhomb technology which is limited by the intrinsic index

#7908 - $15.00 USD
Received 23 June 2005; revised 28 July 2005; accepted 25 August 2005

dispersion of the material used. When the period of a grating becomes smaller than the wavelength of the incident light, it does not diffract light as usual in the sense that only the zeroth transmitted and reflected orders are allowed to propagate outside the modulated regions, leaving wavefronts free from any further aberrations. Furthermore, interaction between the grating and the vectorial electromagnetic field leads to interesting effects on the phase and amplitude of the external propagating fields. Indeed, one dimensional subwavelength gratings, i.e., gratings only modulated along one dimension, turn out to be artificially birefringent. It means that the structure can be associated with two so-called effective indices, one for each polarization component TE and TM. These effective indices, n_{TE} and n_{TM} are totally dependent on the grating and incidence geometries (see Fig. 1). One can really speak of refractive index engineering.

The principle of the Total Internal Reflection Grating Achromatic Phase Shifter (TIRG APS) is to use a subwavelength grating in the TIR incidence condition (see Fig. 2). We will show in the next section that by carefully controlling the grating parameters, the induced vectorial phase shift can be made as nearly achromatic as possible. Achromatic means that the phase shift value has to remain constant over the considered wavelength range. As the leading application of the TIRG APS component concerns nulling interferometry (the TIRG APS high performances could as well be used in other applications like polarimetry), we have chosen to optimize the grating design with a figure of merit called the Null Depth. The Null Depth somehow is the darkness of the destructive interference. The TIR configuration ensures a one hundred percent efficiency for the back reflected light whatever the polarization, therefore no amplitude mismatches. The figure of merit to be minimized consequently resumes to $ND(\lambda) = \sigma^2/4$ where $\sigma(\lambda)$ is the phase shift error standard deviation with respect to the nominal value of π (other values are possible). This relation, derived from the interference between two plane waves theory, simply means that the Null Depth is directly proportional to the variance of the phase shift error, i.e., the achromaticity.

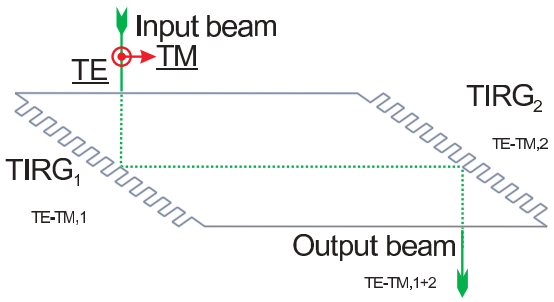

Fig. 2. Schematic of the TIRG APS component. The TIRG APS is analog to a Fresnel rhomb which TIR interfaces are engraved with an optimized subwavelength grating. A TIRG APS component calculated for a π phase shift possesses two TIR interfaces, each providing a $\pi/2$ phase shift such that the resultant is $\Delta\Phi_{TE-TM,1+2} = \Delta\Phi_{TE-TM,1} + \Delta\Phi_{TE-TM,2} = \pi/2 + \pi/2 = \pi$. Such a component is to be inserted in each interferometer arm and orthogonally from one another.

3. Theoretical design

We have carried out extensive numerical simulations focused on the two spectral bands of the Darwin mission: the first one ranges from 6 to 11 microns, the second one from 11 to 18 microns. We have now to choose among the restricted list of IR materials keeping in mind that the ZOG technology is flexible enough to accommodate the majority of them provided that their transparency is sufficient. We will present simulation results for Diamond, Zinc Selenide

#7908 - $15.00 USD
Received 23 June 2005; revised 28 July 2005; accepted 25 August 2005

Table 1. Temperature-dependant coefficients for material index representations.

Coeff.	Diamond	*ZnSe*	*CdTe*	*Ge*
A	1	1	$-2.373\times10^{-4}T+3.8466$	$-6.040\times10^{-3}T+11.05128$
B	0.3306	4.46395	$8.057\times10^{-4}T+3.2215$	$9.295\times10^{-3}T+4.00536$
C	0.030625	0.0404	$-1.10\times10^{-4}T+0.1866$	$-5.392\times10^{-4}T+0.599034$
D	4.3356	0.46132	$-2.160\times10^{-2}T+12.718$	$4.151\times10^{-4}T+0.09145$
E	0.011236	0.1538	$-3.160\times10^{1}T+18753$	$1.51408T+3426.5$
F	0	2.88289	0	0
G	0	2213.5	0	0

(*ZnSe*), Cadmium Telluride (*CdTe*) and Germanium (*Ge*). This choice is justified in three ways: these selected materials are common in IR applications, they cover a large refractive index spectrum (2-4) and their etching processes are well-known. The indices and corresponding dispersions for Diamond[6], *ZnSe*[7], *CdTe*[8], *Ge*[8] will be taken from the following Sellmeier-type relation

$$n_{Diamond,ZnSe,CdTe,Ge}(\lambda,T)=\left(A+\frac{B\lambda^2}{\lambda^2-C}+\frac{D\lambda^2}{\lambda^2-E}+\frac{F\lambda^2}{\lambda^2-G}\right)^{1/2} \tag{2}$$

The corresponding coefficients are given in Table 1. We will assume a refrigerant temperature $T=100K$, keeping in mind that corrections are needed if T is different. Note that λ is expressed in microns in all representations. To simulate grating responses in the subwavelength domain, scalar theories of diffraction dramatically fail. The vectorial nature of light must be taken into account implying a resolution of the Maxwell equations. We use an algorithm based on the Rigorous Coupled Wave Analysis [9] (RCWA) for the simulations of the grating responses. This algorithm also allows, for each polarization, the visualization of the electromagnetic field distribution (amplitudes and phases). We have performed simplex optimization of the component parameters for the selected materials and for the wavelength band ranging from 6 to 11 microns. The results are summarized in Table 2. We notice that the improvement of the Null Depth performance between the Fresnel Rhomb and TIRG APS technology depends on the selected material, ranging from 10^2 for the *CdTe* case up to 10^4 for the Diamond one. Let us note that in the particular case of the Diamond TIRG APS, the rhomb bulk material can be, for example, *ZnSe* on which a thin Diamond layer is deposited by Chemical Vapor Deposition (CVD). In Fig. 3, we explicitly show the *CdTe* TIRG APS results for the two Darwin spectral bands. These results are very good with deep nulls around 5×10^{-9} in average. We have also plotted, for comparison, the Null Depth of the corresponding Fresnel Rhomb.

Table 2. Null Depths for the optimal Fresnel Rhomb configurations and TIRG APS ones and corresponding grating periods for the selected materials.

	Diamond	*ZnSe*	*CdTe*	*Ge*
Fresnel Rhomb $6-11$ μm ND	4×10^{-4}	1.73×10^{-5}	6.4×10^{-7}	7×10^{-9}
TIRG APS $6-11$ μm ND	1.4×10^{-8}	3.4×10^{-7}	3.8×10^{-9}	3.5×10^{-10}
Period	980 nm	1100 nm	1400 nm	662 nm

#7908 - $15.00 USD
Received 23 June 2005; revised 28 July 2005; accepted 25 August 2005

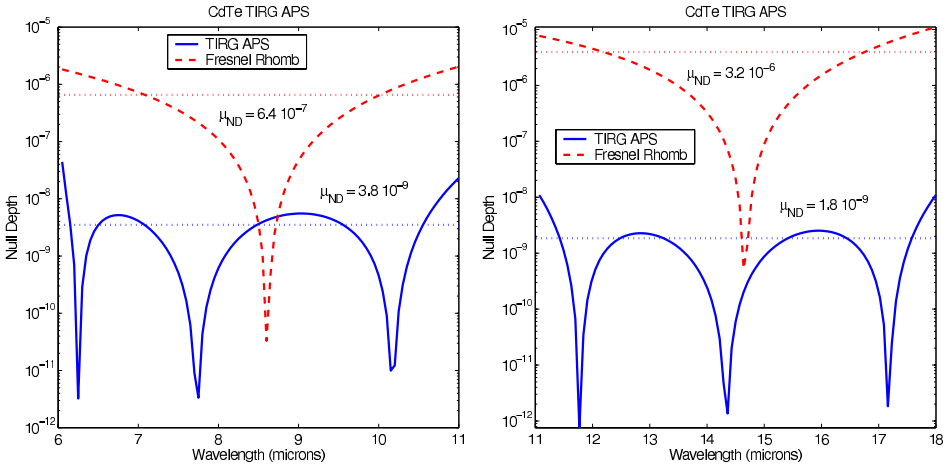

Fig. 3. Continuous lines: *CdTe* TIRG APS performances in terms of Null Depth (logarithmic scale). Dotted lines: *CdTe* Fresnel Rhomb Null Depths (logarithmic scale). Left: 6-11 microns band, the TIRG APS mean Null Depth is $\mu_{ND} = 3.8 \times 10^{-9}$. Right: 11-18 microns band, the TIRG APS mean Null Depth is $\mu_{ND} = 1.8 \times 10^{-9}$.

4. Phase shift and electromagnetic field visualization

To better understand the exact nature of the phase shift and its achromaticity, we propose to visualize the distribution of the electromagnetic field (amplitudes and phases). For this purpose, the structure is divided in three different regions: the first one is the incident medium, in this case the substrate, the second one is the modulated region, i.e., the grating and finally, the last one is the external medium, usually air or vacuum. Each of these three layers is coded with 384×128 ($x \times z$) pixels. The x-axis range corresponds to one period of the grating (Λ) whereas the z-axis range corresponds to 3 units of depth (h). The use of the ZOG in TIR incidence implies that the field interacts with the structure only by means of its evanescent waves. The phase shift between the TE and TM polarization components arises due to the fact that their associated vanishing fields penetrate more or less deeply into the modulated region. Figure 4 shows that the TM component penetrates less deeper than the TE one inducing a pseudo-Optical Path Delay responsible for the phase shift. This differential skin effect can be explained by the difference in the $TE-TM$ zeroth order effective indices. The achromaticity of the subsequent phase shift can be understood from the particular grating-induced artificial dispersions of the effective indices (see Fig. 4 bottom right) but, most of all, by the complex interaction between the higher order vanishing modes.

5. Discussion: implementation and manufacturing

Implementation of the vectorial phase shift in a nulling interferometer is straightforward. Considering two identical components belonging to the two distinct interferometer arms, rotated by ninety degrees around the optical axis and from one another, then the potentially-interfering parallel polarization states are two by two in phase opposition. It must be noted that there is a strong constraint on the alignment of the components. Let $\Delta\chi$ be the misalignement angle. $\Delta\chi$ is geometrically related to the Null Depth ND by vectorial additions of the phase shifted polarization components: $ND = (1+\sin\Delta\chi)(1-\cos\Delta\chi)/2$. Thus, in order to fulfill the constraint $ND = 10^{-6}$, we must reach $\Delta\chi \leq 6$ *arcmin*.

The fabrication of the TIRG APS is based on micro-electronic technologies. The first step

#7908 - $15.00 USD
Received 23 June 2005; revised 28 July 2005; accepted 25 August 2005

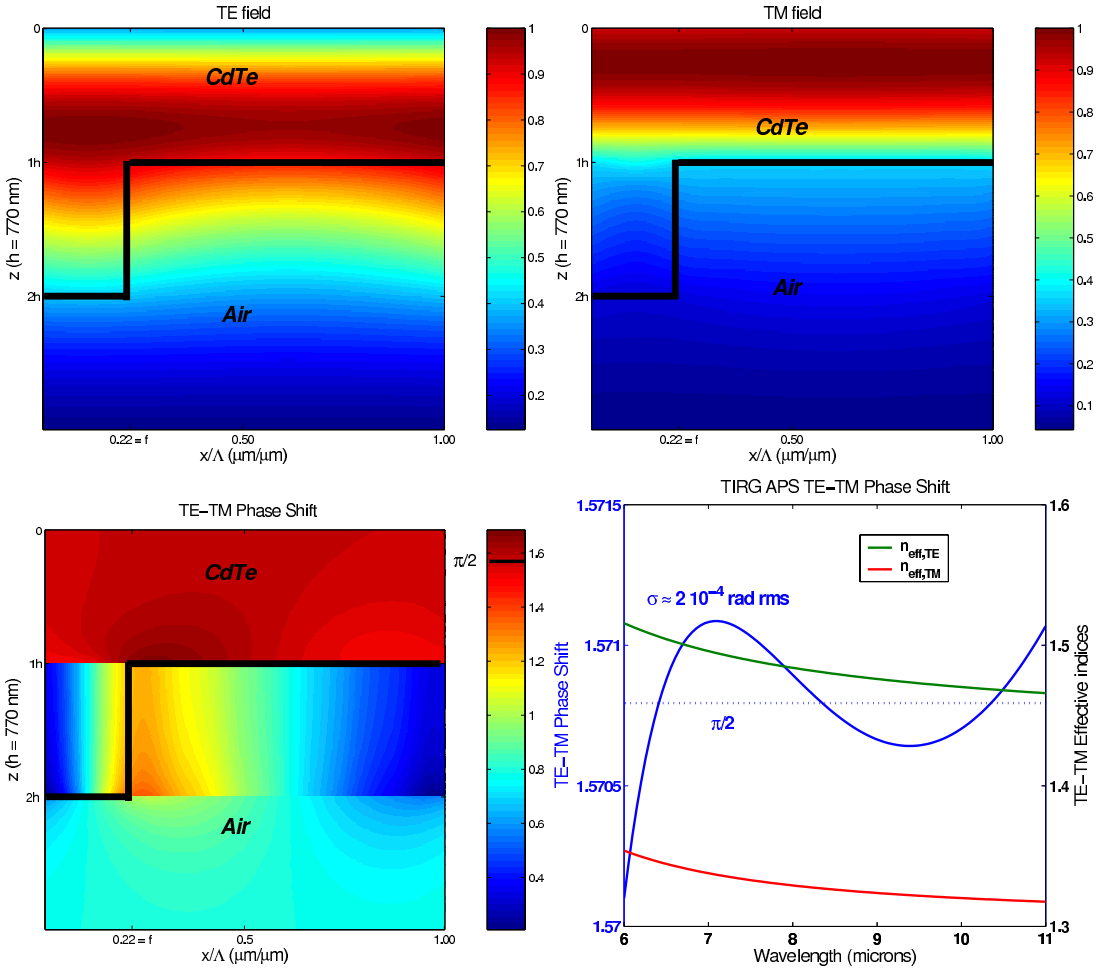

Fig. 4. Electromagnetic field RCWA visualization in the *CdTe* TIRG APS case at 6 microns. Top left: TE component field. Top right: TM component. Bottom left: $TE-TM$ phase shift field visualization. Bottom right: TIRG APS $TE-TM$ phase shift versus wavelength together with the effective indices.

consists in imprinting a grating mask in a resin coated on the chosen substrate material. It can be realized by laser direct writing or e-beam lithography. The precision of this step is critical because it defines once and for all the lateral dimensions of the ZOG, i.e., the filling factor f. This pattern is then uniformly transferred in the substrate by an appropriate reactive plasma beam etching down to the desired depth. The fabrication must be interactive to properly compensate for process errors, e.g. by using in situ monitoring [10]. To summarize, we have presented a new concept of APS relying on subwavelength grating technology. The TIRG APS consists in a grating optimized at TIR incidence. Theoretical results are excellent with deep nulls around (less than 5×10^{-9} for the CdTe) over wavelength ranges corresponding to a spectral resolution $R_\lambda = \lambda/\Delta\lambda \leq 1.5$. A prototype is currently under realization in collaboration with the "Centre Spatial de Liège".

#7908 - $15.00 USD
Received 23 June 2005; revised 28 July 2005; accepted 25 August 2005

9

Manufacturing and test considerations for the TIRG APS

Contents

Abstract. *The TIRG APS is a hybrid component, comprising micro-structures engraved into traditional bulk optics. Manufacturing of hybrid optics is very demanding since the techniques involved are not unique and since their combination or synergy makes it very complex. A long learning phase is therefore required. Compared to micro-electronics industry, the processes are somehow similar or derived. However, the materials and constraints are not the same, and in general, not limited to wafered substrates. A process optimization taking into account the knowledge of physical and chemical properties is also required. The University environment is helpful to develop those techniques, but an industrial support is absolutely required.*

9.1 Preliminary attempts

Prior to the manufacturing of the TIRG APS and before receiving the needed resources, we have conducted a preparatory work to assess the difficulties associated with the following main fabrication steps:

- subwavelength-grating mask making and fine control;
- etching of exotic materials.

The preliminary attempts we performed, revealed to be time consuming on one hand, but also very fruitful in terms of learning on the other hand.

9.1.1 Mask making using holography

As already emphasized in Sect. 3.4.1, the holographic recording is a very efficient manufacturing technique for producing masks of diffraction gratings since it allows treating very large areas at once. The grating period can also be easily defined by the holographic method since one only has to properly superpose two coherent plane waves of wavelength λ_0 while imposing between them an angle θ_0 defined by the well-known Bragg relation

$$\Lambda = \frac{\lambda_0}{2\sin\theta_0} \tag{9.1}$$

Unfortunately, the filling factor that characterizes the feature line consistent with the chosen period, cannot be directly defined since the cross interference pattern between the two plane waves induces a sinusoidal intensity distribution. According to the recording-medium choice, this sinusoidal distribution will induce a surface pattern which actual feature-line width is hardly predictable. This limitation can prevent the use of the holographic technique in the present application as we will discuss.

During Summer 2003, we performed many holographic recording trials at HOLOLAB (more than 150 samples were produced), trying to develop a reproducible process for controlling the feature line of holographic patterns. For that, we chose to use the very popular Microposit S-1805 photoresist from Shipley. This synthetic photosensitive polymer is part of the positive resin family. This means that illuminated zones are removed in a dedicated developer bath (e.g., Microposit MF-319 developer). The underlying process of chemically developable resins is very simple: through a proper illumination in the resin sensitive wavelength regime (for the S-1805: 350-480 nm, i.e., encompassing the i and g lines), polymeric chains are destroyed by the impinging light so that a solvent (the developer) can easily remove them.

The conditioning of photoresin is a liquid with a given viscosity. By adjunction of a solvent (e.g., Microposit EC Solvent), the latter can be modified. A layer of a given thickness is deposited on a cleaned substrate by spin-coating. According to the resin viscosity, the duration and rotation speed of the spin-coater, different thicknesses can be obtained. A post-bake processing is also necessary after the deposition in order to evaporate solvent residuals. We performed the exposure step by using the Lloyd-mirror mounting as illustrated in Fig. 3.7. The angle θ_0 was chosen to provide a period of 1 μm, i.e., according to Eq. 9.1, $\theta_0 = 12°$. The coherent field was provided by a spatially filtered Kr^+ laser emitting at 413 nm (available at HOLOLAB). An important parameter we had to control was the exposure time. It indeed directly determines the amount of energy received by the photoresist, and therefore the degree of destruction of the polymeric chains, which directly translates into reactivity to the development process (i.e., solubility of the exposed parts of the photoresist).

In fact, the goal of this work was to produce a chrome master for subsequent copy-contact replication (see Sect. 3.4.1). For that, we evaporated a thin chrome layer ($\sim$ 50 nm) by thermal deposition on top of the photoresist grating. The chrome layer settles indeed on the cleared part of the substrate and on top of the photoresist "domes" (see Fig. 9.1, left). In order to leave the substrate with the chrome strips only (see Fig. 9.1, right), we then removed the photoresist thanks to a lift-off process involving an intermediary layer which was *a priori* deposited between the substrate and the photoresist layer. Thanks to an accurate control of the process, the lift-off layer (LOL2000, from Shipley) is etched more readily than the photoresist layer during the chemical development. This results in the formation of small cavities underneath the photoresist layer, allowing the coated areas not to be attached to the residual parts which have to be removed with the mask at the end of the process (see Fig. 9.1, left). The best results we obtained to lead to the chrome mask depicted in Fig. 9.1 (right) were obtained for the following process parameters:

- total exposure energy of 50 mJ/cm^2;
- 25 s development in a 4 : 1 Microposit MF-319 : distilled water bath;
- rinsing of the samples with deionized water and drying.

Given the number of determining steps in the process, the obtention of a given feature line was prevented at that time by the extremely varying environmental conditions of the lab (Fig. 9.3, top and middle). Indeed, the chemical reactions leading to the photoresist-profile definition are very sensitive to the illumination stability. Moreover, the development reactions are kinetically very fast and therefore difficult to manage. The conclusion of this work was that a very stable environment in terms of temperature, humidity, cleanliness, laser stability, etc., is necessary for reproducibility.

In fact, we experimentally noticed what has already been predicted: commercial photoresists are engineered and optimized for binary illumination, e.g., with direct writing laser apparatus (DWL, see Sect. 3.4.1). Unfortunately, the intrinsic sinusoidal nature of the holographic process prevents the required binarity up to a certain point (Fig. 9.2). Any drift in the exposure or developing time will indeed induce an error on the filling factor definition achievable onto the photosensitive surface. Some recent results obtained at HOLOLAB (C. Lénaerts, private commu-

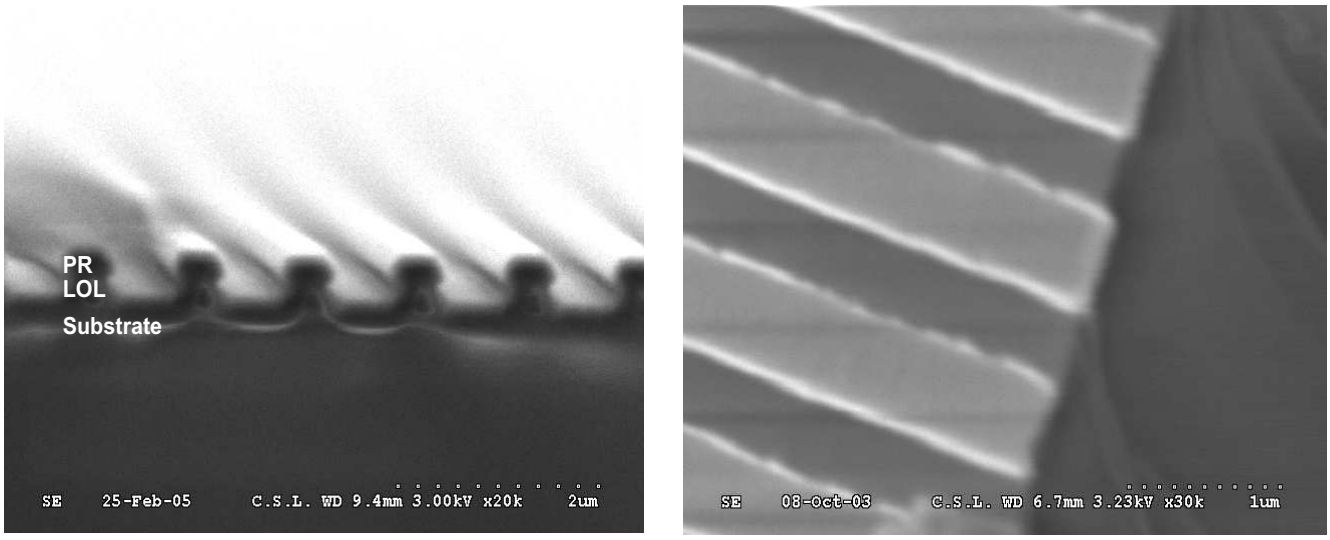

Figure 9.1: Left: developed photoresist grating where the lift-off layer cavities are clearly visible. Right: chrome mask originated from the previous holographically recorded photoresist grating. SEM measurements courtesy of Denis Vandormael (CSL).

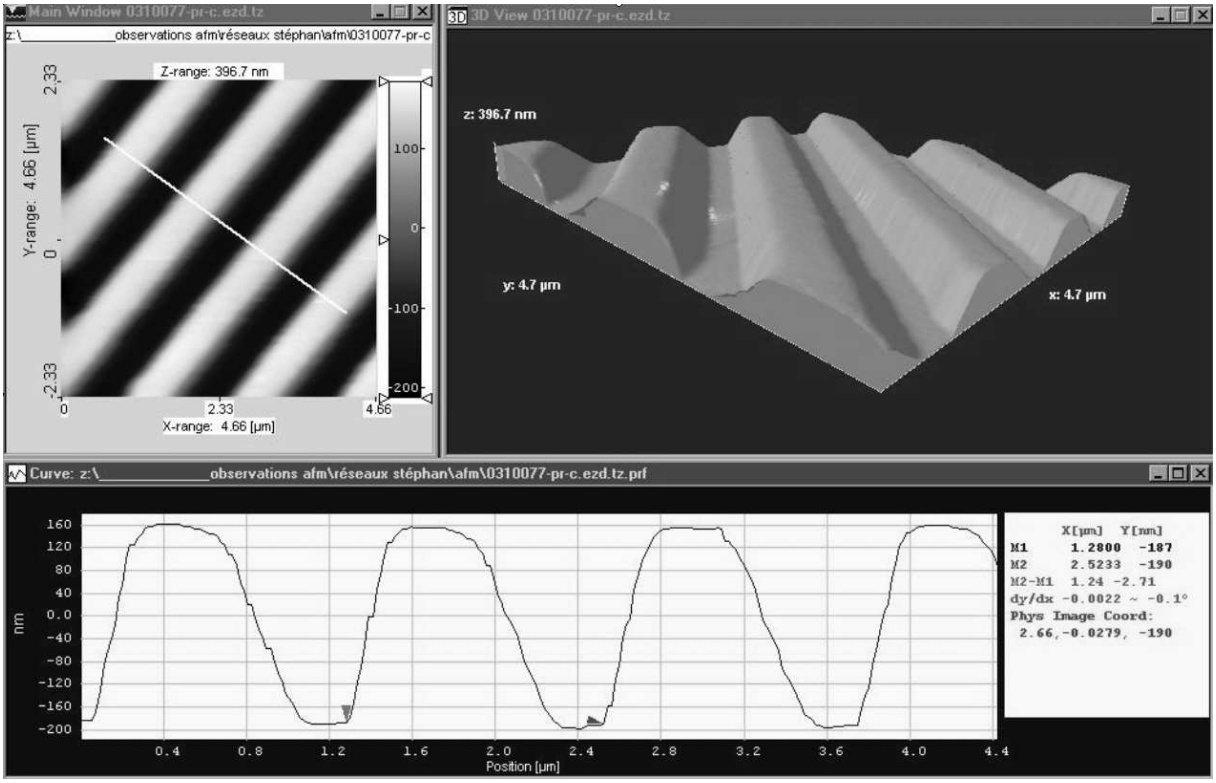

Figure 9.2: AFM picture of a holographically recorded photoresist grating showing the sinusoidal surface pattern reaching the substrate. The measurements were performed at CSL (Courtesy of Denis Vandormael).

nication) are in this respect nevertheless encouraging (Fig. 9.3, bottom). The difference between these attempts and the previous ones was the absence of the lift-off layer. The latter seems to induce a greater variability in the process control. The explanation directly comes from the reactivity of the LOL2000 compound in the developer bath. Here are the details of the new process:

- total exposure energy of 40 mJ/cm^2;

- 20 s development in pure Microposit MF-319 developer;

- rinsing of the samples with deionized water and drying.

The rather same geometrical illumination conditions as presented here above were used. These resulted in a grating which period is 900 nm, and a feature line of $\sim$ 250-300 nm. It is worth noting that there are still some uncertainties about the feature-line definition, and the roughness of the grating walls is clearly visible (see Fig. 9.3, bottom). It is also important to note that some resists can show a different behavior with respect to the illumination pattern. For example the SU-8 photoresist (MicroChem Corp.) has recently been reported to present vertical walls and reproducible feature lines in case of holographic lithography (Pang et al. 2005).

In conclusion, for applications where the definition of the filling factor is relaxed or not critical, this method is extremely efficient. For information, these masks were then used as lithographic mask to etch SiO_2 substrates (Mawet 2004). In the present application and despite the recent improvements obtained in the lab, the sensitivity of the process still prevents its practical utility.

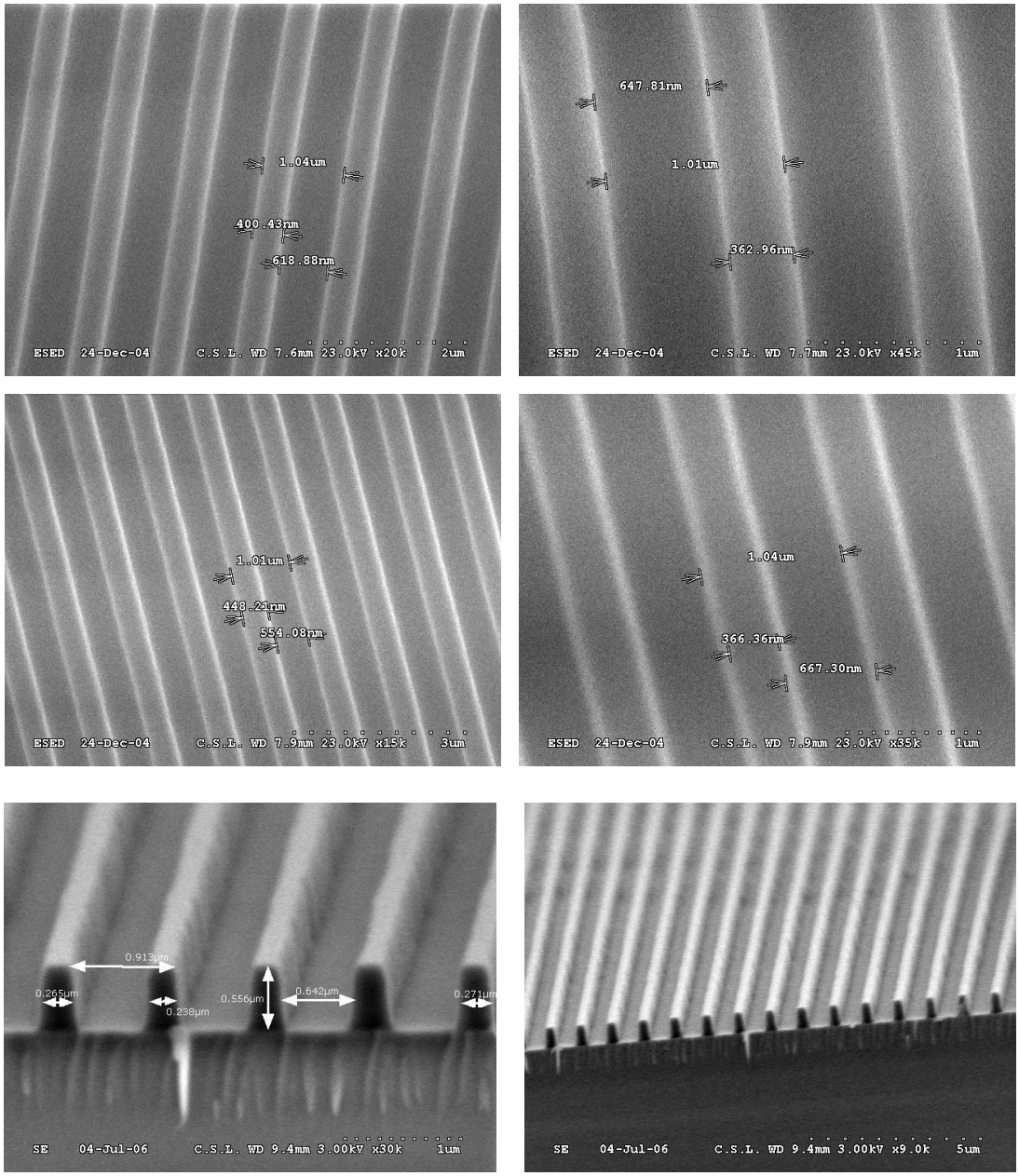

Figure 9.3: SEM pictures of photoresist (S-1805) holographic gratings showing the reproducibility issue of this process. The two SEM images above were taken from a unique grating at two different locations on the substrate. Idem for the two images in the middle for a second sample processed exactly the same way (see text for process details). Bottom: photoresist grating obtained at HOLOLAB with the latest optimized process (courtesy of Cédric Lenaerts, HOLOLAB). At last, a certain reproducibility has been reached. This is in contrast with the top picture case. The difference between them is the absence of the lift-off layer in the latest attempts.

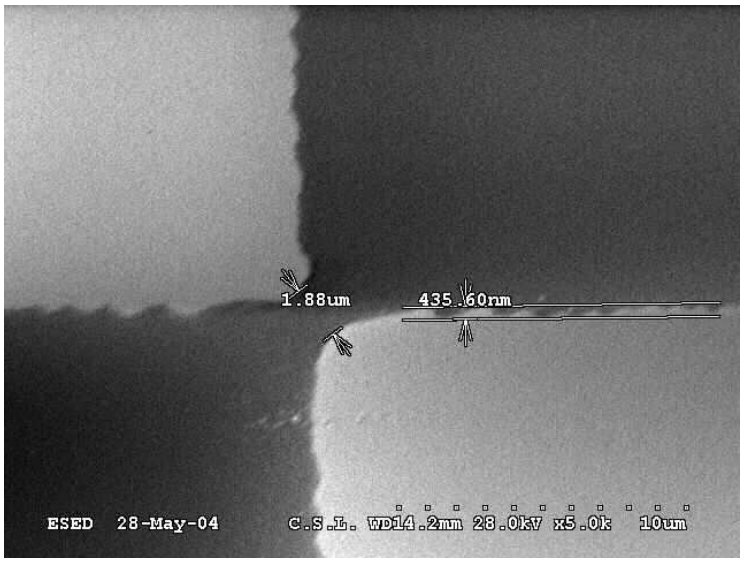

Figure 9.4: SEM pictures of the photoresist (S-1805) mask of the FQPM pattern manufactured by contact lithography showing a submicron resolution in the edge definition and and central gap size lower than the $3\,\mu$m specification. Courtesy of Denis Vandormael (CSL).

9.1.2 Reactive ion etching of $ZnSe$

Reactive ion etching of II-VI compounds has been the subject of extensive researches since the early 1990s (Clausen et al. 1988; Pearton & Ren 1993; Sparing et al. 1996; Chen et al. 2000; Legge et al. 2001). Most of these studies referred to a methane/hydrogen-based chemistry. In the framework of the technological developments of the FQPM coronagraphs for the mid-infrared imager (MIRI) of the JWST (see Sect. 1.5.3), CSL and IAGL proposed to realize the FQPM phase mask by reactive ion etching of $ZnSe$. This work was coordinated by LESIA who also initiated concurrent studies on other substrate compounds like Germanium (performed by CEA-Saclay) and diamond (performed by ADAMANTIS AB from Uppsala, see Sect. 5.3.1). The $ZnSe$ prototypes had to be monochromatic, i.e., based on the index-step principle, and designed to operate at 4.77 and 15.5 μm. The step specifications were therefore $h_{4.77} = 4.77/2(n_{ZnSe}(4.77) - 1) = 1.63\,\mu$m for the smallest wavelength and $h_{15.5} = 15.5/2(n_{ZnSe}(15.5) - 1) = 5.62\,\mu$m for the largest one.

We contributed to these developments by performing the photolithographic resist masks using the Kr^+ laser available at HOLOLAB. The optical mounting we used was quite simple and implemented the contact lithography technique (see Sect. 3.4.1) by employing a spatially filtered and collimated coherent beam. The chrome mask made by a subcontractor (Optimask) was brought into contact with the resist-coated substrate by means of an applied load. This transfer step revealed critical in the definition of the four-quadrant feature since the edges at the transitions between adjacent quadrants had to be as steep as possible, i.e., with a lateral width lower than $1\,\mu$m. Moreover, the center of the quadrant had also to be defined as precisely as possible because it is the spot of focalization of the Airy pattern corresponding to the star to be extinguished. Gaps or bumps, which are inevitable due to the limited resolution of contact lithography ($\sim 1\,\mu$m), are not allowed to be more than $3\,\mu$m-wide in principle. After some trials, we finally managed this critical step reproducibly while achieving a submicron resolution in the transfer process (see Fig. 9.4).

The etching attempts performed at CSL on $ZnSe$ (and $CdTe$) substrates using the produced photomask have shown that the crystalline state of the material to be etched is of primary importance. Fig. 9.5 illustrates the dramatic effect of the CH_4-H_2 process on the surface quality of

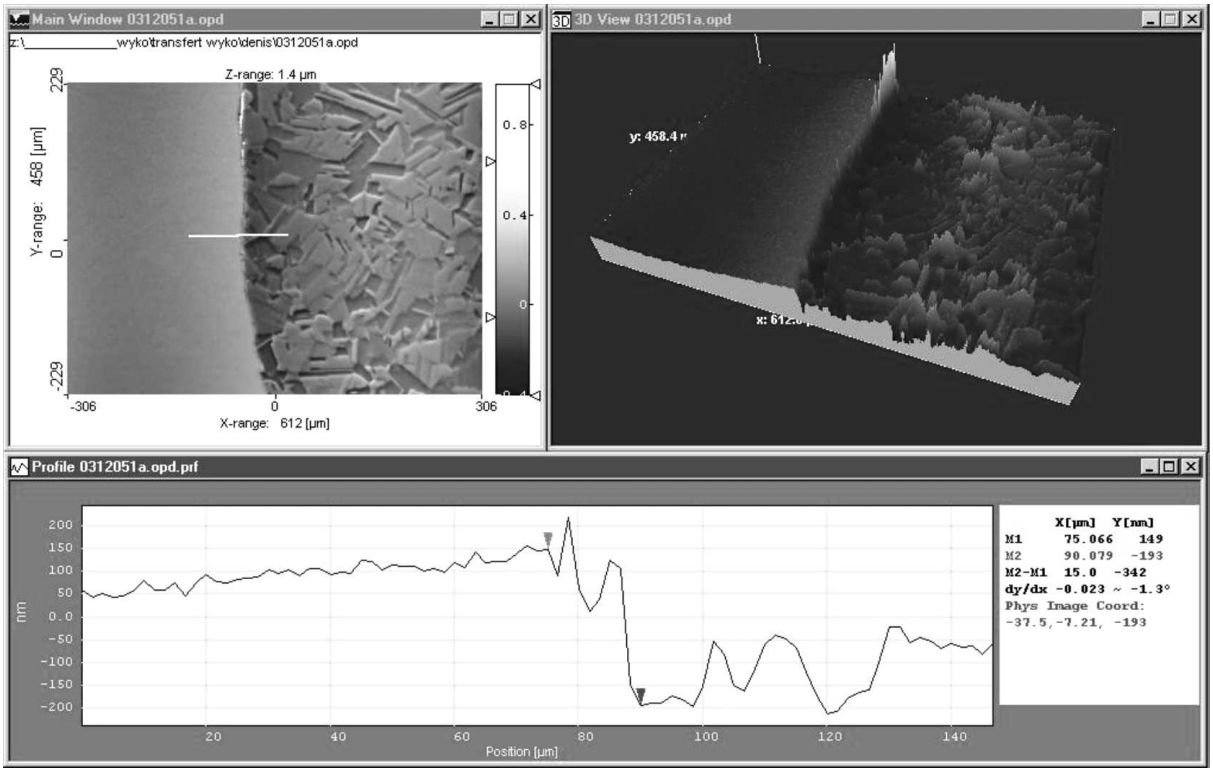

Figure 9.5: WYKO profilometer image of a polycristalline *ZnSe* sample etched with a methane/hydrogen chemistry. This image shows the differential etching selectivity induced by the crystallite orientations. Courtesy of Denis Vandormael (CSL).

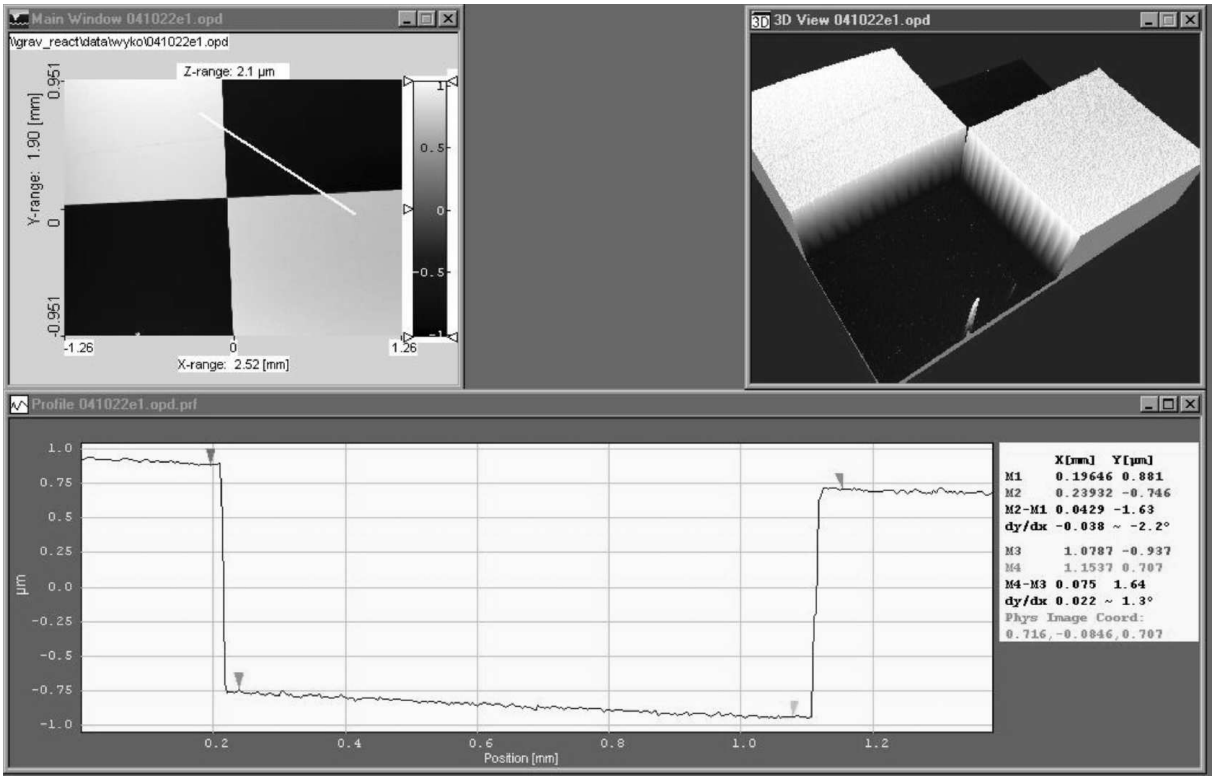

Figure 9.6: WYKO profilometer picture of the methane/hydrogen etched single- crystal *ZnSe* FQPM mask for 4.77 μm. Courtesy of Denis Vandormael (CSL).

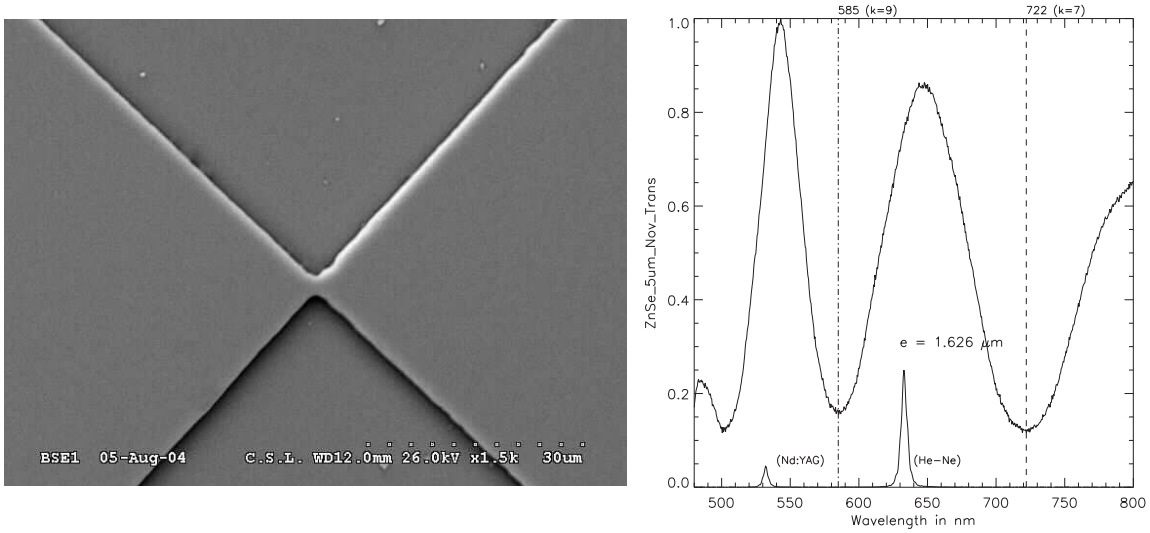

Figure 9.7: Left: SEM picture of the $ZnSe$ 4.77 μm FQPM (Courtesy of Denis Vandormael). Right: low-resolution spectroscopy (see Sect. 5.3.1 for details about this technique) of the $ZnSe$ 4.77 μm mask with a step measured at 1.626 μm. Courtesy of Jacques Baudrand (LESIA).

the sample: a CVD-grown sample of polycrystalline $ZnSe$ was indeed etched by reactive plasma-beam etching (RPBE, see Sect. 3.4.2) according to a CH_4-H_2-Ar chemistry. A polymer Kapton tape was used to mask half a sample in order to delimit a step onto the etched substrate, for metrology purposes. One can easily observe the roughness alteration, due to a differential etching rate between crystallites with different local orientations. The same behavior was observed with polycrystalline $CdTe$ samples. For this reason, we switched to single-crystal samples and managed finally to etch the FQPM pattern to the specifications. Results are displayed in Fig. 9.6 for the WYKO (see Annex E) measurement and in Fig. 9.7 for the SEM picture and low-resolution spectroscopy (see Sect. 5.3.1).

From these preliminary attempts, it appears that the methane/hydrogen-based chemistry is efficient for single-crystal compounds only. Unfortunately, this kind of samples is quite difficult to source. Indeed, our experience showed that manufacturers are not always able to ensure a 100% single- crystal/single- grain/twin-free quality. A residual twin or grain boundary can have dramatic effects after an etching process. Moreover, the raw price for a single- crystal material is rather prohibitive: $\sim 10,000.00$ dollars for a $18 \times 18 \times 18\,mm^3$ optically polished sample. And, as mentioned earlier, there is virtually no guarantee that the material quality is correct prior to the etching process.

Hopefully, a BCl_3-based chemistry appears to be a very promising alternative for etching polycrystalline II-VI compounds. Indeed, it has been recently shown in the literature that, for $ZnSe$, this chemistry is rather insensitive to the local crystallite orientations (Kurisu et al. 2002). Very recent attempts at CSL were encouraging in this respect. Efforts are now put to increase the mastering of this process as far as selectivity of the chosen photomask is concerned.

9.1.3 Mask high-resolution replication by contact printing

Contact printing is the simplest way to transfer a given feature upon a photoresit-coated substrate (see Sect. 3.4.1 and the FQPM example here above). We tried to perform a contact copy of a flat, binary grating mask onto another photoresist-coated substrate. The master used was the one originated from the process exposed here above (Fig. 9.1, left). This could have been a

way of structuring a mask coated onto a bulky substrate thanks to an intermediate flat mask. Transferring a 1 μm-period grating was not an issue despite the fringing effects occurring from the residual air trapped in between the photoresist layer and the mask. Indeed, Newton-ring fringes can alter significantly the copy into the photoresist. The dimension of the features to be copied (of the order of one wavelength) can prevent to use a coherent laser light because of these multiple interference effects. Moreover, reproducing the filling factor between the master and the replica was impossible even in non-coherent light illumination because of near-field Fresnel diffraction. Contact copy needs an intimate contact between the two parts (pseudo molecular adhesion), which is very difficult to achieve in our case, especially with materials as fragile as $ZnSe$ or $CdTe$. Actually, state-of-the-art contact photolithography achievable resolutions are at the micrometer level, which is not sufficient for the present application, unfortunately.

9.2 Manufacturing of the TIRG APS

Manufacturing the TIRG APS is a completely original work since it implies conjugating micro-optics to macro-optics. Indeed, one has to engrave an extremely fine and precise structure onto the surface of macroscopic rhombohedra. The rhombs also possess a precise geometry and must be conceived for facing tough thermal and mechanical constraints.

9.2.1 Fabrication plan

The most appropriate method to manufacture the micro-pattern into the rhomb material is based on nano-imprint and dry-etching processes (see Sect. 3.4). The first one is necessary for masking the parts of the substrates to be protected during the etching step. Direct writing laser (DWL, see Sect. 3.4.1) or mask-exposure techniques are not directly applicable onto the rhomb facets since those classical micro-lithography methods are designed for accommodating thin and flat substrates only, such as wafers for micro-electronics. Indeed, the maximum clearance between the substrate holder and the writing head (6 mm) of standard DWL tool is not compatible with the considered components ($\sim$ 20 mm-thick rhombs). To our knowledge, this is the case for all the commercially-available equipments. As far as the contact and holographic lithography are concerned, our preliminary investigations have shown that they are not adapted in this context. Hopefully, the nano-imprint alternative allows replicating onto a thick substrate a mask previously originated by conventional photolithography onto a thin substrate.

Fig. 9.8 illustrates the principle. A master is originated by illuminating (with a contrasted pattern) a photoresist layer coated on a thin substrate. For this step, we have chosen to use the direct writing laser (DWL, see Sect. 3.4.1) tool available at CSL. This technique appeared indeed to be particularly appropriate for recording the kind of pattern details we have to deal with. Electron-beam lithography and focused ion-beam writing are too expensive, and time-consuming techniques, generally dedicated to the recording of very high-resolution but small-area patterns. In the present case, there is no need for such high-resolution processes. In addition, these techniques are not easily available for the type of R&D study we have to perform. Although holographic recording is fully oriented towards the recording of linear diffractive structures, it does not allow an easy control of the filling factor, as noticed in the preliminary investigations presented here above.

State-of-the-art DWL machines, on the other hand, can achieve the filling-factor control and reproducibility that is needed in the present case. Moreover, the DWL minimum feature line

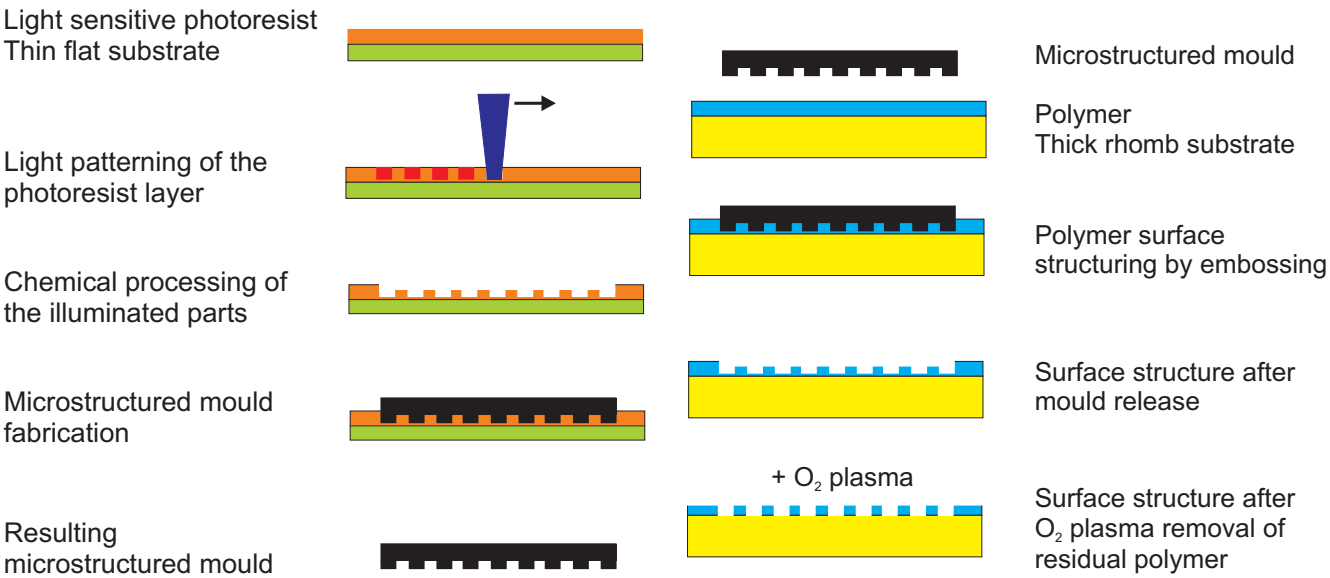

Figure 9.8: Nano-imprint lithography for the TIRG APS.

characteristic is within the specifications of the component to be fabricated (linewidth $\sim$ 0.6 μm). The DWL machine available at CSL (DWL66 from Heidelberg instrument, Fig. 9.9) uses a collimated 442-nm $HeCd$ laser beam as a controlled writing tool for imprinting features with sizes down to the 0.6 μm-linewidth spec onto photoresit-coated substrates as large as 140 $\times$ 140 mm. Writing pattern is performed by accurately moving the substrate underneath the focused laser[32] (position accuracy $\sim$ 40 nm). The photoresist relief structure is then classically obtained after a chemical process (development), as shown in Fig. 9.10.

The rest of the process pileup illustrated in Fig. 9.8 follows these steps: a mould is generated from the produced photoresist master and it is used to emboss a polymer layer deposited onto the rhomb facet, using either conventional NIL or S-FIL (see Sect. 3.4.1). After demoulding, the polymer residual layer is removed by an oxygen plasma in order to clear the substrate where it is supposed to be subsequently etched. The remaining polymer pattern serves as a lithographic mask for the subsequent reactive plasma-beam etching (RPBE, see Fig. 9.11) process. As already stated in Sect. 3.4.2, RPBE makes use of both the ballistic effect and chemical reactivity of a beam of reactive ions to remove or create structures into a substrate. The various parameters (gas melanges, beam energy, beam incidence, etc.) characterizing the etching process are optimized for the transfer into various materials. The interest of such a technique particularly comes from its high selectivity, i.e., the potential to efficiently etch one material and not another coexisting one, and its directionality (anisotropy). RPBE is widely used in the fabrication of micro-optical elements, micro-sensors and other micro-machines. The infrared material ($ZnSe$ or $CdTe$) is engraved using a selective chemistry which leaves intact the polymer mask but etches the infrared substrate. Our preliminary investigation have shown that a CH_4-H_2-Ar chemistry is convenient for single-crystal samples while BCl_3-based plasma are more adapted to polycristalline forms. It is worth mentioning again that the etched depth can be controlled in real time with an interferometric in situ device (Lalanne et al. 1999).

Although mask selectivity is of primary importance, we need to pay attention to the mask-removal step. This remark is particularly critical for the etching of materials such as $ZnSe$ and

[32]The laser beam is apodized with a Gaussian. The sensitized profile characteristic is subsequently defined by the intensity of exposition (writing speed and laser power). The laser must therefore be extremely stable and the installation conditioned in temperature (at the 0.1° C level).

Figure 9.9: Direct writing laser 66 (DWL66) from Heidelberg instrument available at CSL.

$CdTe$. Indeed, for these materials, classical corrosive mask removers are prohibited (only alcohol or acetone-based solutions, for instance, can be used). Therefore metal-based masks are not allowed and polymer-based materials such as photoresist appear to be the only suitable solutions since most of them are solvable in acetone. Preliminary attempts have also shown that photoresist is appropriate for methane/hydrogen processes. According to the literature, photoresist shall also be compatible with BCl_3-based dry-etching processes. If an additional lithographic process is

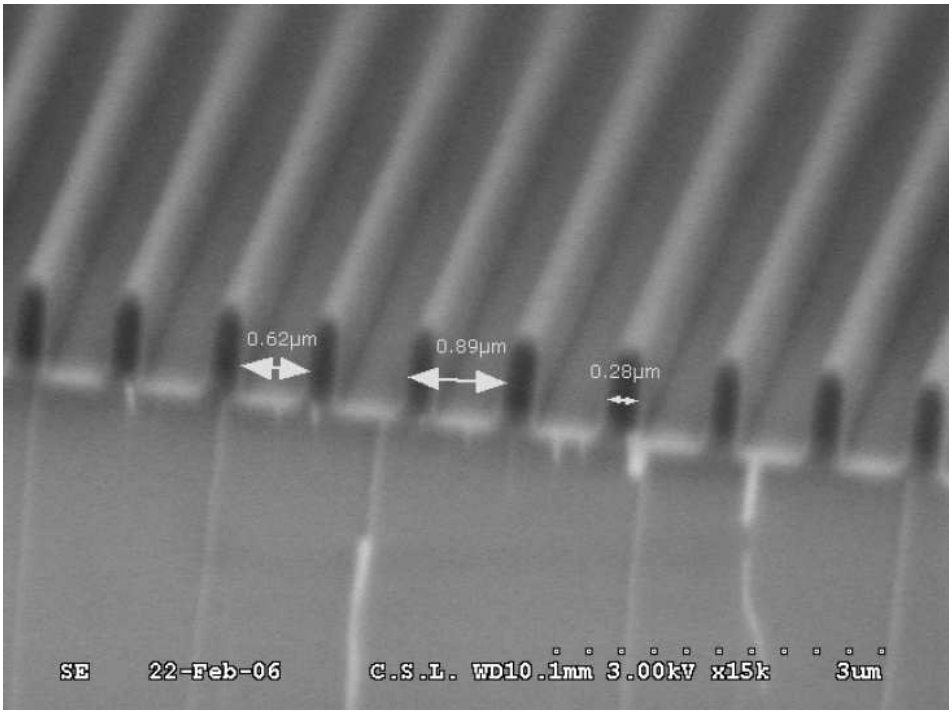

Figure 9.10: Master mould showing the micro-pattern on photoresist to be transferred by nano-imprint. The period and filling factors correspond to the design specifications, i.e., $\Lambda = 0.9\,\mu$m for the period and ~ 250 nm for the feature line. Courtesy of Jérôme Loicq (CSL).

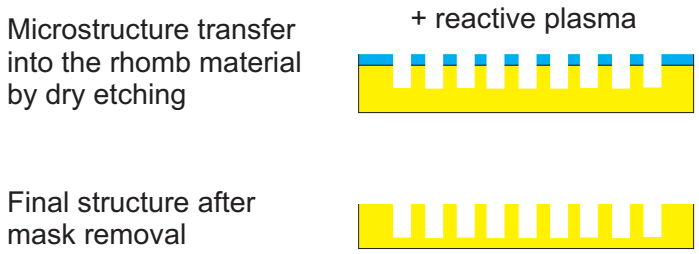

Figure 9.11: Pattern transfer from nano-imprinted mask into a substrate by dry etching.

considered (nano-imprint), we also have to perform the same kind of qualification for the masking material with respect to the chosen etching process.

9.2.2 Thin-film deposition

According to the material to be deposited and the required precision, different techniques can be considered. Indeed, classical evaporation is still used for simple coatings (like $ZnSe$), but advanced techniques allow preparing more compact-layer coatings such as ZnS, for instance, which necessitates specific ion assisted deposition (IAD) techniques (Lemarquis et al. 1998). It is to be noted that the well-known precision of the thickness control of these techniques can be an issue when one deals with very thin layers (typically less than 50 nm). However, the tolerance analysis presented in Sect. 8.5.3 shows that it is still within reach. As far as diamond deposition is concerned, commercial treatments by chemical vapor deposition (CVD) are already available for Ge or Si substrates. To our knownledge, it has never been demonstrated on $ZnSe$ and $CdTe$ substrates. This may be due to the very high temperature needed for the diamond growth, i.e., over 800°C (Dore et al. 1998). However, new promising approaches for diamond deposition are currently studied to lower the growing temperature. Good results have been reported with temperature around 400°C (Xiao et al. 2004). In the case of CVD diamond, thickness control is ensured by the relatively slow growing rate, i.e., $\sim$ 200 nm per hour (the final roughness is $\sim$ 5 nm rms).

9.3 Tests

The development of innovative technologies must obligatorily be accompanied by appropriate tests for qualification of the component expected performances, always keeping in mind the theoretical results against with the measurements have to be confronted.

9.3.1 Structural metrology

In situ monitoring (see Sect. 3.4.3) is often referred to as the optimal way to control each step of the manufacturing process. However, implementation of real-time measurements during fabrication is sometimes tricky. For this reason, we decided to independently control each fabrication step *ex situ* by nano-scale classical metrology based on the available techniques at CSL (see Fig. 9.12, left, and Annex E for the detailed characteristics of the different devices).

After each fabrication step and the related accurate nano-scale metrology, the theoretical model based on RCWA will be used to update the design parameters according to the measurement. The corrected parameters will be re-injected to the next manufacturing step, which will generate a new

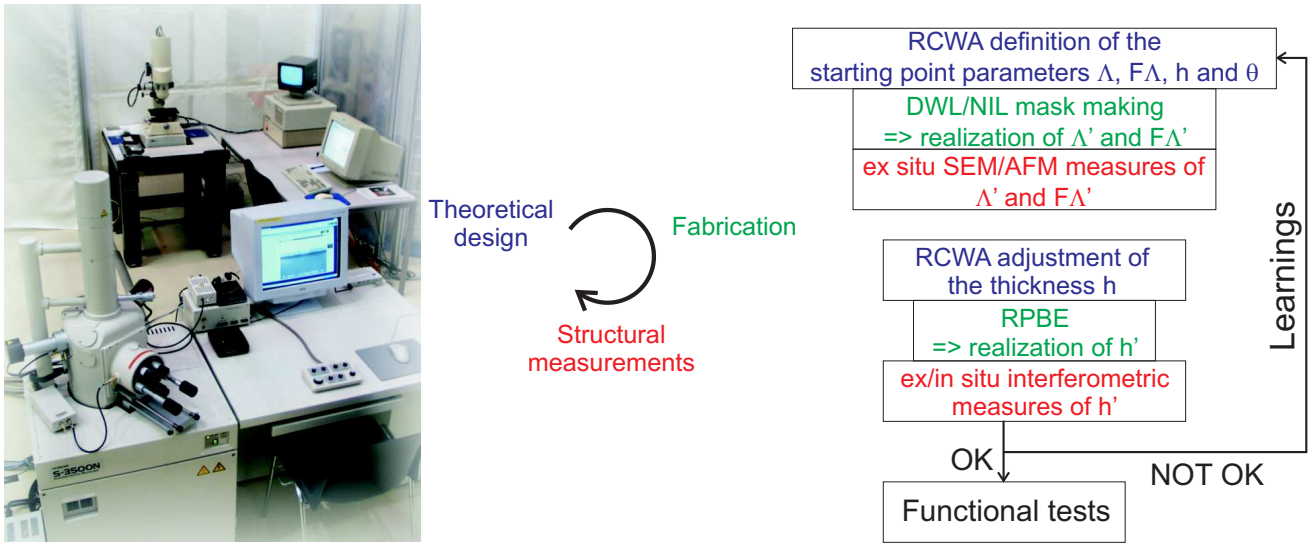

Figure 9.12: Left: scanning-electron microscopy and optical profilometry (WYKO) equipments at CSL (see Annex E for the specifications). Right: interaction between theory, fabrication and measurements.

measured error, and so on (see Fig. 9.12, right). As mentioned and illustrated in the previous section, this procedure, in conjunction with the design and tolerancing study, will finally demonstrate the TIRG APS manufacturing feasibility using the selected micro-fabrication technique.

In particular, initial and final roughness will be measured by the optical profilometry (WYKO). Scanning-electron and atomic force microscopy will check the mask quality and its relevance with respect to theoretical values. These techniques will allow us to validate the accuracy in the transfer of the micro-structure into the material.

9.3.2 Functional metrology

Several functional tests will be carried out to assess the functional performance of the component, i.e., its phase-shifting ability and its achromatic nature. Provided that the sample size and geometry is compatible with the apparatus, the direct and most appropriate tool for performing those measurements is an infrared ellipsometer. This kind of device (available at CSL, see Fig. 9.13, and Annex E for its characteristics) can indeed directly measure the phase shift between the perpendicular polarization components $\Delta\phi_{TE-TM}$ in the considered 6-14 μm range.

We can also use three different methods to measure the phase retardance between the polarization components TE and TM so that accuracy can be demonstrated and studied. Two methods make use of modified versions of standard polarimetric measurements and require rotating polarizers. These can be complemented by an interferometric method that is less sensitive to polarizer quality but is more sensitive to coherence effects in the retarders. Before exposing these methods, it is worth emphasizing that these measurements can advantageously be performed at other wavelengths than the working ones (6-14 μm in the present application), for the sake of simplicity and for cost reasons (thermal infrared optical components and cryogenic cameras are very costly). Indeed, reverse engineering can be used at the measurement wavelengths (e.g., 1.55 μm, the widespread telecom wavelength) to extrapolate the component performance in its working bandpass thanks to the RCWA code.

The first method makes use of a standard polarimetric technique modified to reduce error sources (see Fig. 9.14a). Linearly polarized light is incident upon the retarder, and the light

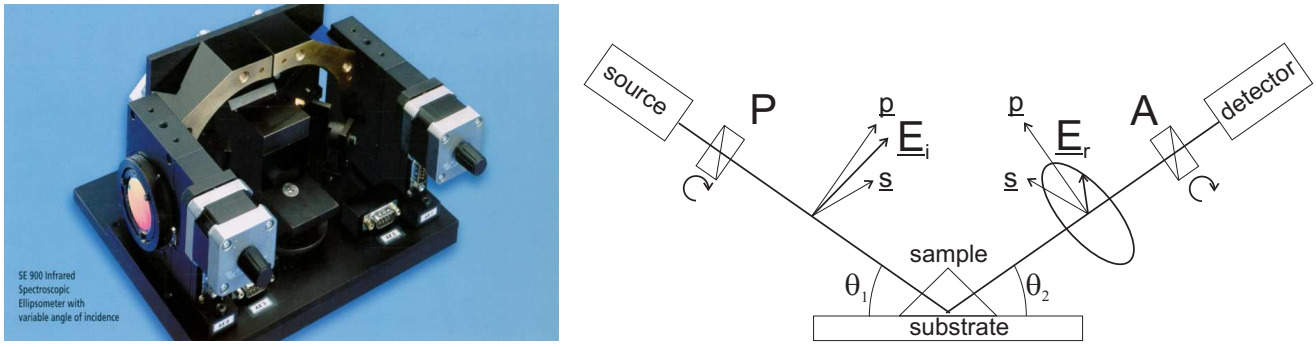

Figure 9.13: Left: infrared spectroscopic ellipsometer SE900 from SENTECH available at CSL. Right: ellipsometer scheme showing the source providing an incident linear polarization field $\underline{E_i}$ through polarizer P, which transforms into a elliptical polarization $\underline{E_r}$ upon reflection on the sample. The analyser A provides direct measurements of ψ and $\Delta\phi_{s-p}$ (or $\Delta\phi_{TE-TM}$) thanks to the so-called ellipsometric equation $\rho = R_p/R_s = \tan\psi e^{i\Delta\phi_{s-p}}$ where R_p and R_s are the complex reflection coefficient for s and p.

emerges with an elliptical polarization. The intensities of the two orthogonal polarization states are measured simultaneously with a Wollaston polarizer and photodiodes. A governing equation relates the ratio of these two intensities to the orientations of the input polarizer, the Wollaston polarizer, and the retardance. Use of the ratio of the transmitted intensities normalizes fluctuations in laser power as well as differences in detector gains. These intensities are measured as the input polarizer is rotated through a known range of angles, and a least-squares fit to the governing equation is performed to determine the phase retardance. Phase-retardance measurements in the visible with an expanded uncertainty between 0.047° and 0.11° have been obtained with this system (Williams et al. 1997).

The second method is a null technique adapted from ellipsometry. Before the retarder is inserted, the polarizer and the analyzer are crossed, and the biasing waveplate axes are aligned parallel to the polarizers (see Fig. 9.14b). Then the retarder is inserted and oriented with its retardance axes at 45° to those of the quarterwave plate. The polarizer is rotated until extinction occurs and the phase retardance is twice the angle of rotation. Retardance measurements with an expanded uncertainty less than 0.1° have been demonstrated in the visible. Advantages of this technique include a weak dependence on the stability of the laser, on the linearity of the photodiode, and on the accuracy of the quarterwave plates. The instrument is very sensitive to the accuracy with which the azimuthal positions of the polarizer and the analyzer can be read. An error of 1° in the orientation of the polarizer converts to an error of 2° in the measured phase retardance.

Finally, an accurate interferometer that directly measures the phase shift between s and p polarized fields can also be used (Rochford & Wang 1997). The retarder is placed in one arm of a Michelson interferometer (see Fig. 9.14c), and the input beam is linearly polarized at 45° so that the fast and the slow axes of the retarder are equally illuminated. At the output of the interferometer, a polarizing beam splitter with axes aligned to coincide with the retarder axes separates the light into two orthogonally polarized beams that are separately detected. The detected outputs follow the common sinusoidal transfer function typical of a Michelson interferometer, but the sinusoids differ by a phase bias equal to twice the actual retardance. Translating one of the mirrors at constant velocity with a piezoelectric driver scans the transfer function with time and produces the biased sinusoidal waveforms. The phase difference between the two waveforms is accurately determined by the acquisition of the data with a computer and the application of discrete Fourier analysis. Because the two beams traverse a common path, small path changes

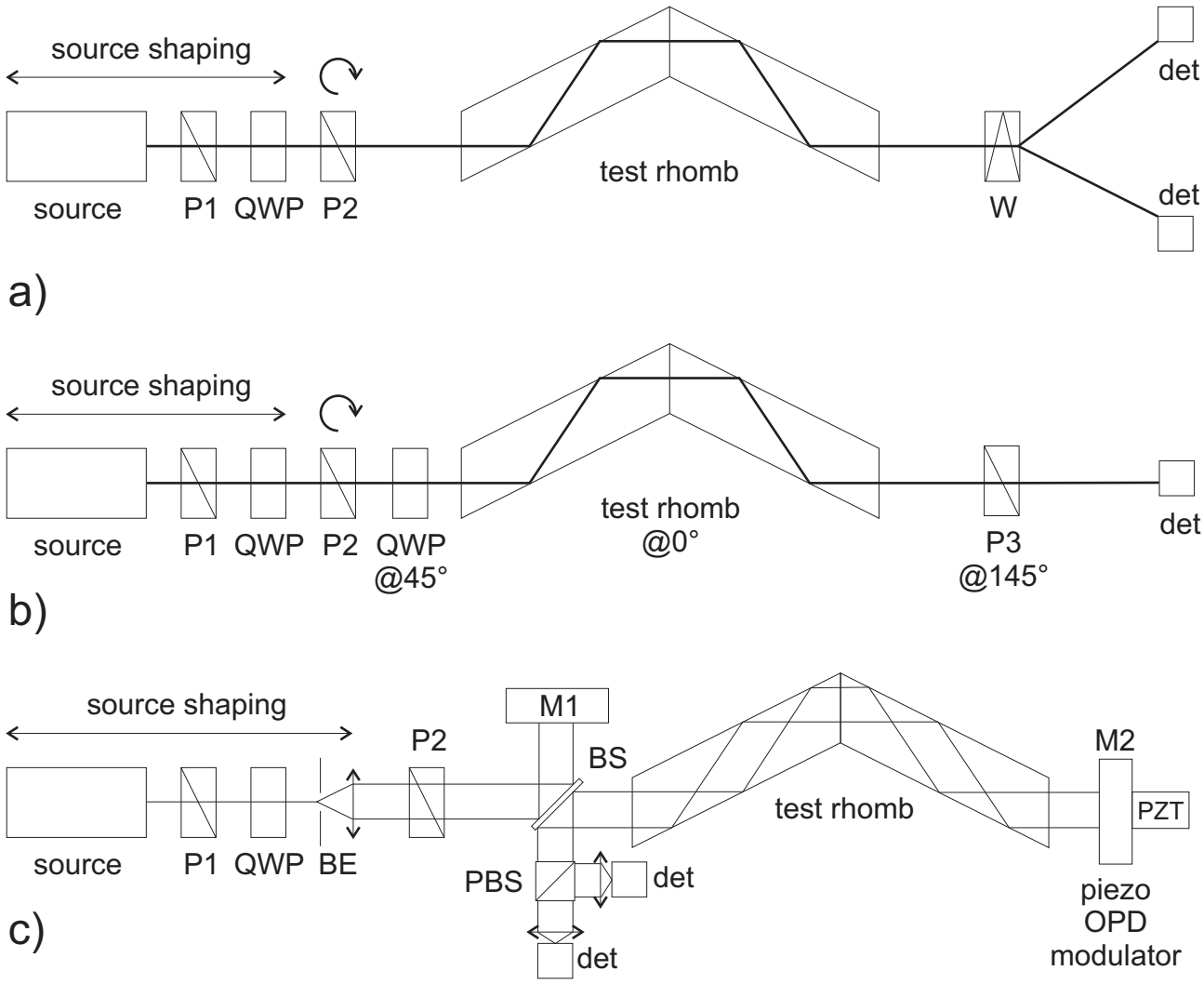

Figure 9.14: Layouts for polarization phase-shift measurements. a) Design of a polarimeter for retardance measurements. b) Null polarimeter for retardance measurements. c) Michelson interferometer of polarization retardance measurements. Components include polarizers P, mirrors M, quarterwave plates QWP, beam expander BE, beam splitters BS and polarizing beam splitters PBS.

caused by acoustics or temperature changes, for example affect both beams equally so the phase difference is unchanged. Retardance measurements with expanded uncertainties between 0.08° and 0.1° have been demonstrated in the visible using this method.

9.3.3 Final tests on the NULLTIMATE bench

Under the supervision and financing of ESA, IAS[33] is at the head of a large consortium involved in the development and characterization of different concepts of achromatic phase shifters for the Darwin mission. This project, called NULLTIMATE (NULling Low Temperature Interferometer for Multi-Achromatic phase shifters TEsting), aims at determining and characterizing the four

[33]The Institut d'Astrophysique Spatiale (IAS) is a laboratory of the National Center of Scientific Research (CNRS) and of the University of Paris-Sud 11. In addition to having the status of Observatory, the IAS comprises 140 scientists, engineers, technicians, administrators and graduate students. It is worth mentioning that IAS possesses a great experience in nulling interferometry breadboarding (Ollivier 1999; Brachet 2005).

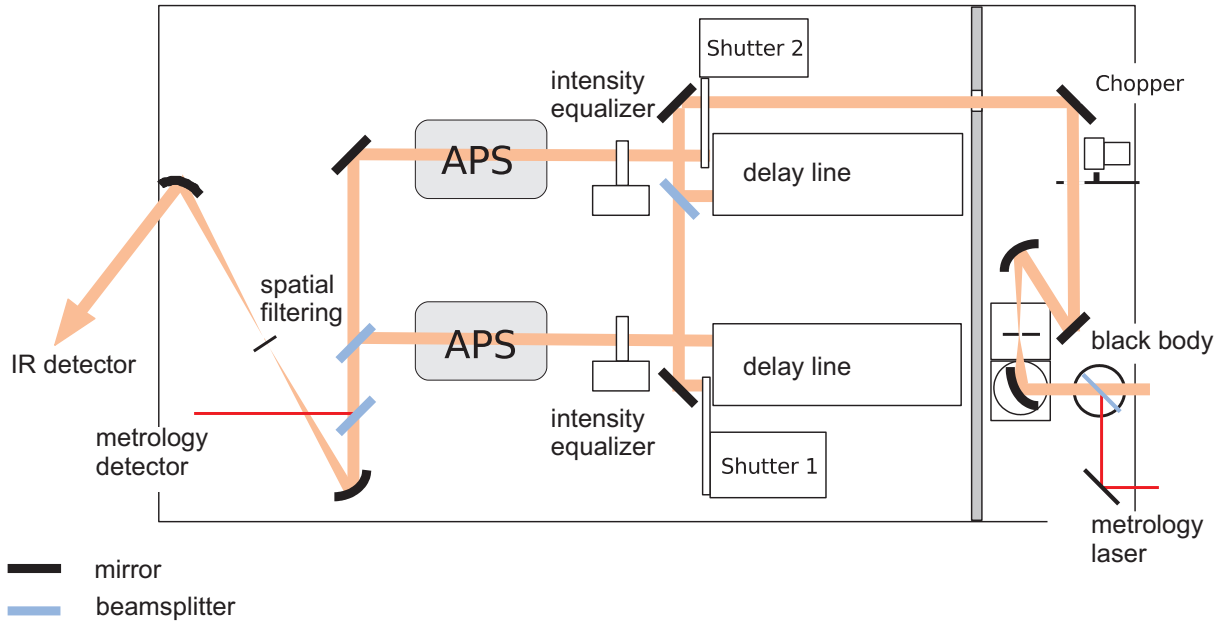

Figure 9.15: Scheme of the NULLTIMATE test bench. Courtesy of Bruno Chazelas (IAS).

APS solutions the most adapted to the current Darwin specifications. The different APS should allow rejection rates of 10^6 in the 6-18 μm wavelength range while ensuring a global 95% optical throughput. The four selected APS after a preliminary study are: the dispersive-plate APS, the focus-crossing APS, the field-reversal APS and the TIRG APS (see Sect. 2.2). The test bench is under definition and is expected to perform the first measurements in 2007. The measurements and tests are planned to be conducted firstly at ambient temperature and then at cryogenic temperature, i.e., 100 K (not before 2009). A preliminary scheme of the bench is shown in Fig. 9.15. It is worth noting that the beam splitters are planned to be constructed in *ZnSe* as well as adjustable in thickness for chromatic OPD compensation (on top of the classical delay lines).

Conclusion

Objectives and results

In this dissertation, we have provided a detailed discussion of the scientific case, theory, design, tolerancing and manufacturing of optical micro-components based on optimized subwavelength gratings. We have demonstrated the utility of these nano-engineered meta-materials in the framework of high dynamic range astrophysics, and in particular in the most appealing field of extrasolar planetary system imaging and characterization. Subwavelength gratings provide new and original solutions for the very demanding detection techniques involved in the fulfilment of this ambitious scientific goal, i.e., phase-mask coronagraphy and nulling interferometry.

The first chapter of this thesis, beyond the introduction of the scientific context, was an opportunity to present our contribution to the demonstration of the practical utility of new coronagraphic techniques for actual observations of young forming extrasolar systems. Using the four-quadrant phase-mask coronagraph installed at the focus of one of the most performing adaptive optics systems in the world (NACO at the VLT) as a high dynamic range imaging tool, we have contributed to the detection of an optically thick circumstellar disk around the young star PDS70, within which evidence of planetary formation was found. In Chapter 2, we went into the theoretical details of the high dynamic range imaging techniques while emphasizing the necessity of a family of components, the achromatic phase shifters, used both in phase coronagraphy and nulling interferometry. As an illustration of the practical implementation of existing techniques, we have demonstrated the concept, feasibility and actual laboratory performance of vectorial phase shifters (viz. commercial waveplates) in the fabrication of an achromatic four-quadrant phase-mask coronagraph, paving the way towards more complex solutions like the use of meta-material synthesized by optimized subwavelength gratings.

In the second part of this work, Chapter 3 was devoted to the presentation of the theoretical aspects and existing manufacturing techniques for subwavelength gratings. The theoretical tool for analyzing and designing subwavelength gratings, the so-called rigorous coupled-wave analysis (RCWA), is no less than a numerical resolution of the Maxwell's equations. Indeed, at the subwavelength scale, one has to take into account the vectorial nature of light since traditional scalar diffraction theories like the Fraunhofer or even the Fresnel approximations dramatically fail at describing the complex behavior of the micro-structured matter. The use of RCWA at University of Liège dates back to the early 1990s. It was firstly implemented and developed at HOLOLAB, benefiting from improvements by successive generations of researchers and taking advantage from the contributions of a very active scientific community. Motivated by new specific needs, our contribution to this endeavor in Liège was to entirely reprogram the algorithm in the modern, versatile and performing language of MATLAB. We have also implemented new functionalities like conical diffraction and three-dimensional analysis for multilayered two-dimensional gratings, making it a complete tool for treating almost every grating-related diffraction problem. In Chapter 4, we described as exhaustively as possible the different applications of subwavelength gratings, while

underlining their potential in high dynamic range imaging techniques. Our original contribution in this chapter consists in the study of anti-reflective structures made of specific two-dimensional subwavelength gratings for infrared astrophysical applications. In this regard, we have presented a theoretical analysis using RCWA as well as the result of a practical demonstration, involving the fabrication of a prototype in diamond with an industrial partner and the subsequent fruitful laboratory tests.

The third part of this dissertation contains the heart of our contribution to the field of new technologies for high dynamic range imaging, viz. the presentation and complete analysis of the new and original 4QZOG (Chapter 5) and AGPM components (Chapter 6). Both of them make use of optimized subwavelength gratings with the purpose of enlarging their useable bandwidth thanks to an integrated solution. While the 4QZOG is an achromatic evolution of the FQPM coronagraph, the AGPM is a totally new design of coronagraph providing a new way to reject starlight while being completely free from the dead zones inherent to the quadrant transitions of the FQPM. The AGPM creates a so-called optical vortex, which we have analytically demonstrated to entirely reject a perfectly centered coherent source. Unfortunately, the fabrication of such components requires state-of-the-art manufacturing technologies and related industrial support which is very costly. For these reasons, as of today, no prototype has been delivered but an ambitious programme, which has received financial support from public funds in Belgium and France, is underway at CEA-LETI in collaboration with our partners of LESIA and LAOG (Chapter 7).

In the last part of this dissertation, we present a new high performance achromatic phase shifter (APS) concept making also use of optimized subwavelength gratings. This concept aims at improving the performance of Fresnel rhomb components in the framework of nulling interferometry. Chapter 8 is dedicated to the presentation of the principle of the so-called total internal reflection grating APS (TIRG APS). It also describes the design and tolerancing of a prototype to be tested in the framework of the R&D activities for the Darwin mission (ESA). As far as the practical realization of this prototype is concerned, industrial and financial supports are also mandatory. Therefore, in collaboration with our partners at CSL, we asked —and finally obtained— a grant from ESA for the manufacturing of TIRG APS prototypes at the CSL, which is the subject of Chapter 9.

Perspectives

The main perspective of the present thesis is the use of new generation tools to conduct advanced observations in the fields of high dynamic range imaging, as exposed in Chapter 1. More pragmatically, the short-term prospects concern the implementation of one of the achromatic phase-mask coronagraphs developed here inside the VLT-PF/SPHERE instrument. It will be either an achromatic FQPM constructed from halfwave plates as presented in Chapter 2, or a 4QZOG/AGPM (Chapters 5 and 6) made of subwavelength gratings if the operation conducted at CEA-LETI is successful (Chapter 7), or even both of them. In fact, the two techniques appear to be complementary since the halfwave-plate FQPM concept offers extremely wideband capabilities ($R_\lambda \approx 2$) but at the cost of stability (delicate bulky mounting, temperature sensitivity, etc.) and discovery space (FQPM dead zones at the quadrant transitions), while the subwavelength grating solution is an integrated one (very compact, lightweight and stable) offering equal performances but on slightly smaller bandpass ($R_\lambda \approx 5$), and provides the possibility of manufacturing the AGPM, alleviating one of the most annoying Achilles'heel of the FQPM coronagraph.

Of course, VLT-PF/SPHERE is not the unique project of high dynamic range imaging from the ground. Many other extreme (or not) adaptive optics systems equipped with coronagraphs

are foreseen in almost every large observing facility in the world (Subaru, Keck, Palomar, Gemini, AEOS, etc.). The micro-components we are about to manufacture would advantageously replace classical Lyot coronagraphs that are still often envisaged for these instruments. As already mentioned in the first chapter of this dissertation, numerous projects of coronagraphic space telescopes are also being currently considered. The SEE-COAST project was introduced in Chapter 6. It will very soon be proposed to ESA as an answer to the call for small missions in the framework of the Cosmic Vision program. On the NASA side, no less than four similar projects have also been suggested to the Discovery program.

Unfortunately all these projects only partially make up for the recent announcement of the *sine die* deferment of the NASA's Terrestrial Planet Finder program (TPF-Coronagraph and TPF-Interferometer) and the cutback of a great part of the associated research budgets. Nevertheless, the direct detection of Earth-like extrasolar planets orbiting nearby stars and the search for signatures of biological activity in their atmospheres are still high-priority objectives in the community. Fortunately, ESA is still pursuing its road map program towards the ambitious Darwin space interferometer by funding many R&D activities for demonstrating the high dynamic range capabilities of infrared nulling interferometers. The development of the TIRG APS (Chapters 8 and 9) and the participation to the NULLTIMATE consortium is part of our contribution to this road map full of perspectives.

Part V

Appendices

A

RCWA convergence

Throughout this work, we have used the RCWA algorithm exposed in Chapter 3 for calculating the quantitative behavior of subwavelength gratings. Though the implementation of the RCWA code was performed using the latest refinements of the theory (Sect. 3.2.3), it must be tested against convergence. Since the main figure of merit used in this work was the null depth (Eq. 2.6), it appeared naturally as the test figure for convergence. In Fig. A.1, we show the residual between the null depth calculated for a given number of retained orders and the 513 retained order case. The test case was the $ZnSe$ double rhomb as depicted in Sect. 8.5 with the following parameters: period of 900 nm, feature line of 250 nm, thickness of 1.23 μm, incidence angle of $\theta = 1.1396$ rad, wavelength range from 6 to 14 μm (20 interpolation points). Assuming a convergence criterium of 10^{-8} on the null depth, we concluded that ~ 21 orders or more was a reasonable value for ensuring a valuable accuracy.

It is also to be noted that energy conservation is always checked in the algorithm with a threshold at 10^{-10}.

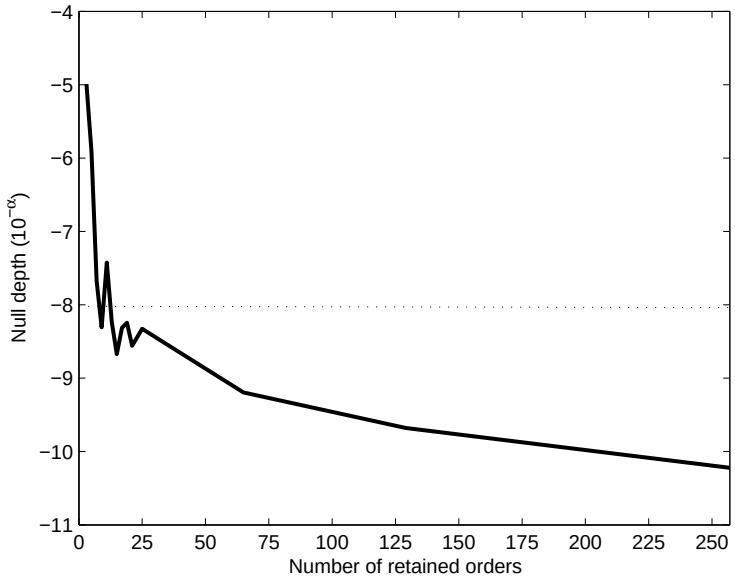

Figure A.1: Residual null depth error with respect to the 513 retained order case (logarithmic scale) versus the number of retained orders $(2N + 1)$.

B

Infrared materials for the TIRG APS

We have performed an exhaustive survey on transparent materials in the thermal infrared. Four main families of materials were investigated: chalcogenides, halides, fluorides and crystalline semiconductors. Since the TIRG APS solution that have been investigating requires the beam to travel quite a long path through the bulk material (around 20 cm), the most important criterion is their very good transparency in the two Darwin bands (6-11 and 11-18 μm). The second one concerns the homogeneity of available ingots since the considered APS requires large volumes. Optical characteristics like refractive index dispersions are not restraining since the TIRG APS has the ability and flexibility to compensate for natural dispersions in order to reach the proper effect on the phase. For the same reason, thermo-optic coefficients will not be considered here. Birefringent materials are avoided in order not to complicate the design (e.g., the alignment of optical axis with the rhomb and the subwavelength grating would be tricky). Thermo-mechanical stability together with hygroscopy may be limitative for some materials and should also be considered.

B.0.4 Chalcogenides

The continuously increasing interest in the improvement of thermal imaging systems for spectral ranges extending up to the third window of atmospheric transparency (8-12 μm) has lead to the development of suitable and low-cost optical materials (As_2Se_3, GASIR, IG6, etc.). Chalcogenide glasses are extensively studied for this purpose, and used both as bulk or fibred optical component. Chalcogenide cut-off wavelengths are in the second Darwin band (around 14 μm). Therefore, they can only be considered for the first band. These materials are soft and their behaviour at low temperature is unknown.

B.0.5 Halides

Halides (Ia, Ib/IIIb-VIIb) like $NaCl$ and KBr are known for their good transparency in the thermal infrared and also for their relatively low refractive indices. However, their hygroscopy, softness and cleavability makes them not suitable for our application except maybe KRS-5 which should be considered apart. KRS-5 is indeed transparent in both Darwin bands, is not sensitive to moisture and has no cleavage planes. On the other hand, microstructuring of KRS-5 is not common and therefore not well referenced in the literature. Silver halides ($AgCl$, $AgBr$, etc.) are in general hygroscopic. They also possess clivage planes and are soft. Their use in infrared optical fibers is less critical since they are covered by the surrounding protective layer. Moreover, $AgCl$

and $AgBr$ are photosensitive to visible light. This would lead to major difficulties for a proper implementation (manufacturing and laser alignments, etc.).

B.0.6 TeX glasses

As far as tellurium glasses are concerned, only the Rennes University in France actually manufactures these materials (TeX glasses). Again, these materials are widely used in fiber optics to replace chalcogenide glasses. Unfortunately, the losses between 7 and 9.5 μm are already at an unacceptable level (0.5-1 dB/m).

B.0.7 Fluorides

These materials (IIa $-$ $(\text{VIIb})_2$) are grown in large dimensions with good optical homogeneity (BaF_2, CaF_2, etc.) but possess a cleavage plane (111). Moreover their cut-off wavelengths (ranging from 9 to 12 μm) are too low for our application.

B.0.8 Crystalline semiconductors

Crystalline semiconductors are either pure elements IVb (Si, Ge) or IIb-VIb and IIIb-Vb compounds. Most of their relevant physical properties are already quite well known. For optical purposes Si, Ge, $GaAs$ and $CdTe$ are used as bulk single crystals while best optical properties for ZnS and $ZnSe$ are obtained with polycrystalline CVD compounds. $ZnSe$ can also be available in single crystals but at a higher cost and with homogeneity problems. $CdTe$ polycrystalline substrates can also be obtained. These materials are known to be thermally, mechanically and chemically stable. Moreover, processes to micro-structure semiconductors are well referenced though etching chemistries are delicate for some of them ($ZnSe$ and $CdTe$ for instance). The main characteristic of crystalline semiconductors is the quite high values of their refractive index. They depict also a rather complicated absorption structure in the region of the long wavelength cut-off that affects the true useful spectral range compared to the usually admitted transmission window. Let us detail here below the characteristics of some of the principal semiconductors.

Diamond

Diamond is a very interesting material for optical purposes. Indeed, its index wavelength dispersion is very low (see Fig. B.1, left). It is transparent from the UV to the mm waves, with some phonon absorption features between 2.5 and 6 μm (see Fig. B.1, right). It possesses the highest hardness and thermal conductivity of all materials. Quite apart from the cost, the only drawback is that large volumes are not available since its fabrication is based on Chemical Vapor Deposition (CVD) processes, unfortunately limited with diamond to thin wafer manufacturing (not more than a few mm thick).

We have used the following Sellmeier representation for the refractive index dispersion of CVD diamond (Bundy 1962)

$$n_{diamond}(\lambda) = \left(A + \frac{B\lambda^2}{\lambda^2 - C} + \frac{D\lambda^2}{\lambda^2 - E} \right)^{1/2} \tag{B.1}$$

with the corresponding coefficients given in Table B.1.

Table B.1: Sellmeier coefficients for diamond index representations at room temperature. CLD stands for Carbon like diamond.

Coeff.	CLD
A	1
B	0.3306
C	0.030625
D	4.3356
E	0.011236

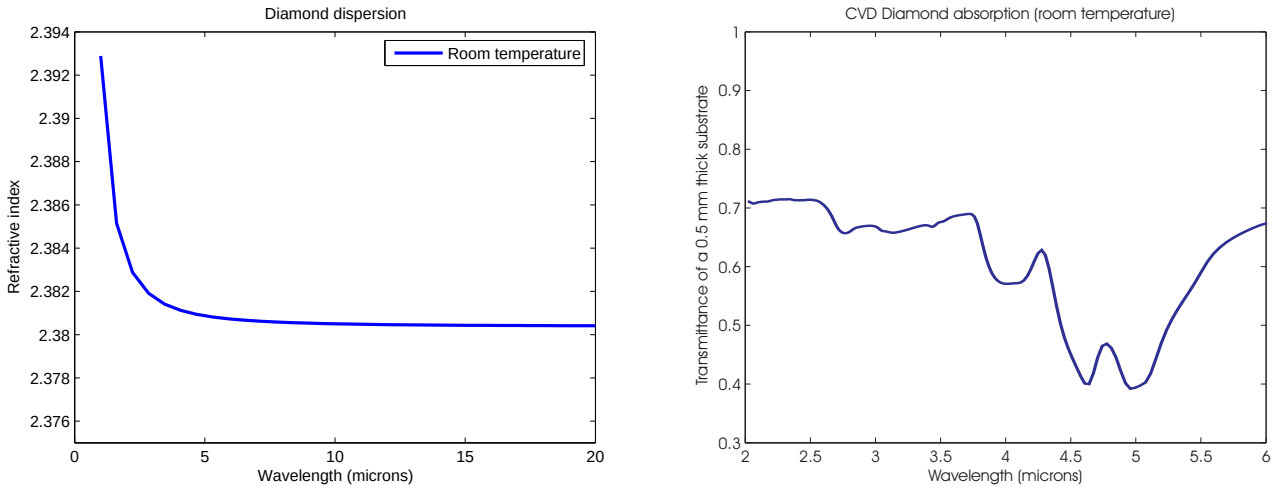

Figure B.1: Diamond refractive index dispersion at ambient temperature (left) and measured (courtesy Jacques Baudrand, LESIA) phonon absorption (right).

Silicon/Gallium Arsenide. Both *Si* and *GaAs* are very well known materials for their extensive use in micro-electronic technologies. However, more than their cut-off wavelengths (8 and 12 μm, respectively), their infrared multi-phonon absorption features makes them unsuitable for the Darwin application (see Fig. B.2, right). However, it is worth noting that *Si* is one of the most interesting materials for application up to ~ 9 μm since its processing is very well mastered. *Si* has been chosen for the fabrication of the prototypes 4QZOG and AGPM (see Chapter 7).

We have used the following Sellmeier representation for the refractive index dispersion of *Si* (Hawkins 1998)

$$n_{Si}(\lambda, T) = \left(A + B\lambda + C\lambda^2 + D\lambda^3 + E\lambda^4\right)^{1/2} \tag{B.2}$$

with the corresponding coefficients given in Table B.2.

Table B.2: Sellmeier coefficients for silicon temperature-dependent index representation.

Coeff.	Si
A	$1.600 \times 10^{-4}T + 3.431$
B	-2.643×10^{-2}
C	4.324×10^{-3}
D	-3.194×10^{-4}
E	8.835×10^{-6}

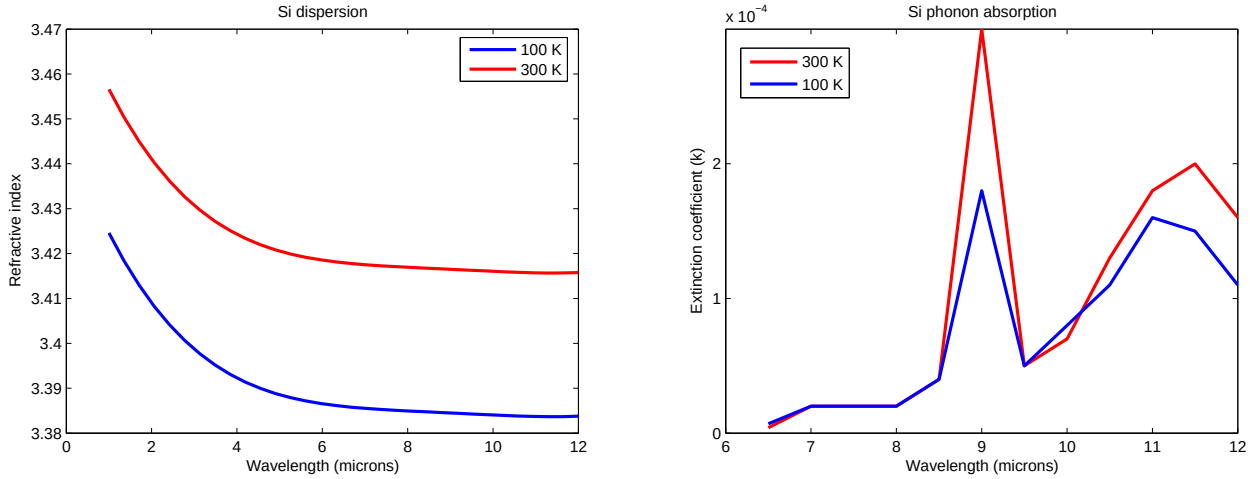

Figure B.2: Silicon refractive index dispersion (left) and predicted phonon absorption (right). Data from Hawkins (1998).

Germanium. *Ge* is a very interesting material for infrared applications. Apart from its very high index (around 4), its wavelength dispersion in the thermal infrared is extremely low (Fig. B.3, left). However, some absorption features in the second Darwin band makes it partly unsuitable for the Darwin specifications. Indeed, let us remind that the TIRG APS requires the beam to travel a long path into the material, making even a small absorption very penalizing.

We have used the following Sellmeier temperature-dependent representation for the refractive index dispersion of *Ge* (Hawkins 1998)

$$n_{Ge}(\lambda, T) = \left(A + \frac{B\lambda^2}{\lambda^2 - C} + \frac{D\lambda^2}{\lambda^2 - E} \right)^{1/2} \tag{B.3}$$

with the corresponding coefficients given in Table B.3.

Table B.3: Temperature-dependent coefficients for *Ge* index representation.

Coeff.	*Ge*
A	$-6.040 \times 10^{-3} T + 11.05128$
B	$9.295 \times 10^{-3} T + 4.00536$
C	$-5.392 \times 10^{-4} T + 0.599034$
D	$4.151 \times 10^{-4} T + 0.09145$
E	$1.51408 T + 3426.5$

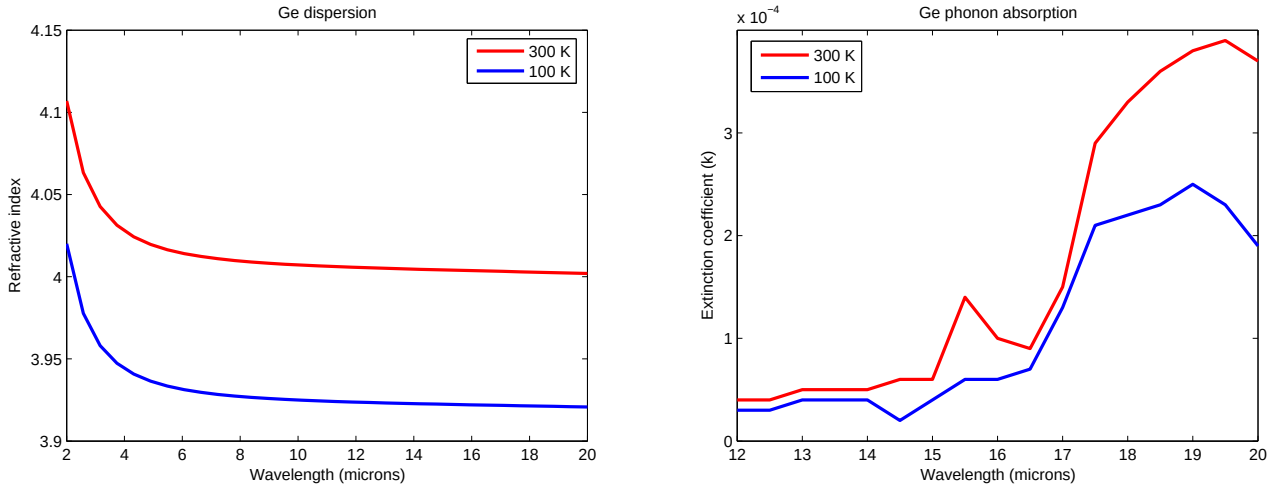

Figure B.3: Germanium refractive index dispersion (left) and predicted phonon absorption (right). Data from Hawkins (1998).

Zinc Selenide. *ZnSe* is a clear yellow polycrystalline material with a grain size of approximately 10 to 70 μm, transmitting in the range 0.5-15 μm providing extremely low bulk losses from scatter. Physical Vapor Transport (PVT) grown samples are monocrystalline but difficult to find with a good homogeneity. Having a very low absorption of energy makes it useful for optical components in high power laser window and multispectral applications, providing good imaging characteristics. *ZnSe* is also useful in high resolution thermal imaging systems. However, some absorption features starting at ~ 14 μm makes it unsuitable for the second Darwin band. *ZnSe* wavelength dispersion is quite high and temperature sensitive.

We have used the following Sellmeier temperature-dependent representation for the refractive index dispersion of *ZnSe* (Hawkins 1998)

$$n_{ZnSe}(\lambda, T) = \left(A + \frac{B\lambda^2}{\lambda^2 - C} + \frac{D\lambda^2}{\lambda^2 - E} \right)^{1/2} \tag{B.4}$$

with the corresponding coefficients given in Table B.4.

Table B.4: Temperature-dependent coefficients for *ZnSe* index representations.

Coeff.	*ZnSe*
A	$1.509 \times 10^{-4} T + 2.407$
B	$-1.801 \times 10^{-5} T - 2.564 \times 10^{-4}$
C	$1.300 \times 10^{-6} T - 1.308 \times 10^{-5}$
D	$-3.878 \times 10^{-8} T - 1.480 \times 10^{-5}$

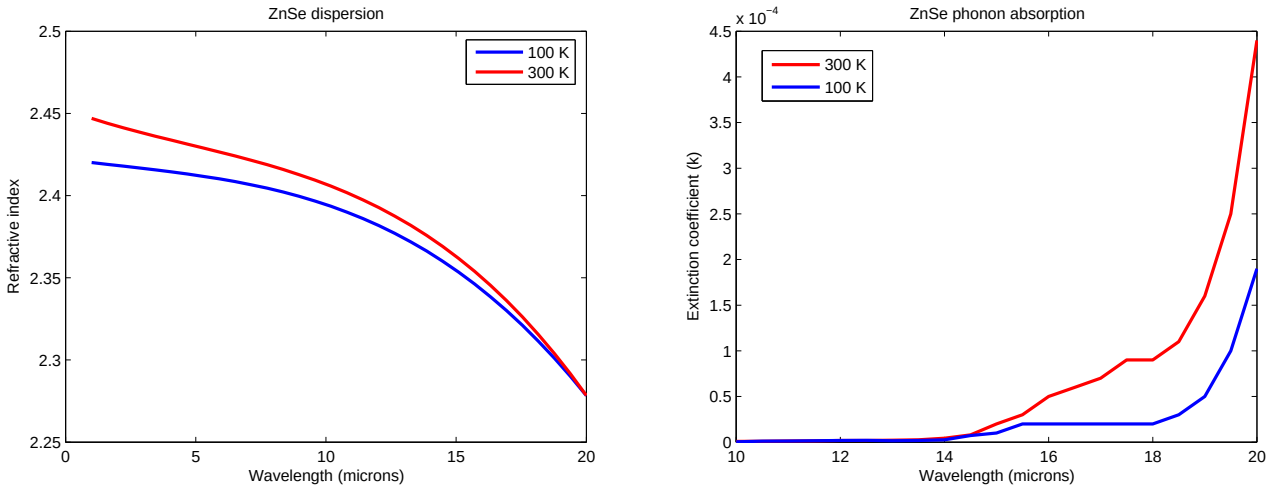

Figure B.4: *ZnSe* refractive index dispersion (left) and predicted phonon absorption (right). Data from Hawkins (1998).

Zinc Sulfide. ZnS exists both in a natural and synthetic crystalline form possessing cubic (Zinc blende) or hexagonal (Wurtzite) lattice structures. A variety of techniques have been developed to obtain synthetic ZnS in large, high purity crystals, including evaporation, sublimation, high pressure growth from molten ZnS, and sintered hot-pressed polycrystalline ZnS (IRTRAN 2). The most popular and easily available ZnS is the CVD ZnS (CleartranR) material. ZnS has its transparency window up to approximately 10 μm where multi-phonon absorption dominates. For this reason, it is not suitable for the Darwin application.

We have used the following Sellmeier temperature dependent representation for the refractive index dispersion of ZnS (Hawkins 1998)

$$n_{ZnS}(\lambda, T) = \left(A + B\lambda + C\lambda^2 + D\lambda^3 + E\lambda^4\right)^{1/2} \tag{B.5}$$

with the corresponding coefficients given in Table B.5.

Table B.5: Temperature-dependent coefficients for ZnS index representation.

Coeff.	***ZnS***
A	$5.608 \times 10^{-5}T + 2.282$
B	$-8.671 \times 10^{-6}T - 1.563 \times 10^{-2}$
C	$5.549 \times 10^{-7}T + 2.067 \times 10^{-3}$
D	$2.597 \times 10^{-8}T - 1.714 \times 10^{-4}$
E	$-9.798 \times 10^{-10}T + 2.884 \times 10^{-6}$

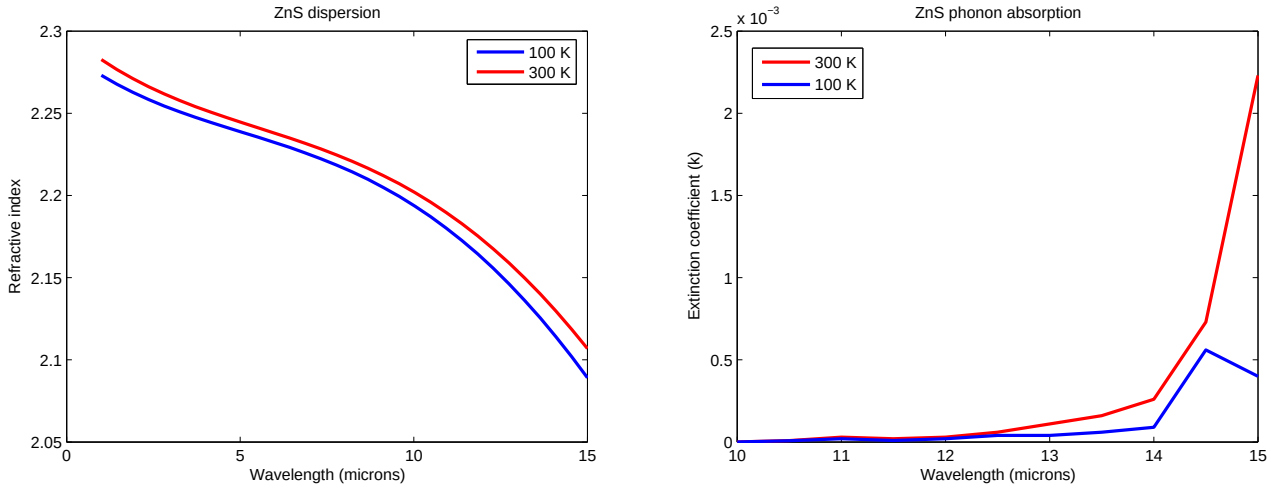

Figure B.5: ZnS refractive index dispersion (left) and predicted phonon absorption (right). Data from Hawkins (1998).

Cadmium Telluride. Among the long wavelength (above 18 μm) transparent II-VI materials available, cadmium telluride has proven to provide good optical performance across a wide range of temperatures and has adequate mechanical robustness to be used as a substrate material. Compared to the limited selection of alternative materials capable of transmitting at these long wavelengths (e.g., KRS-5, KRS-6, CsI, $CsBr$, diamond), $CdTe$ has a high resistance to moisture, is available at a reasonable price and can operate at elevated deposition temperatures without disassociating. It is however the softest of the II-VI materials and is most easily scratched or prone to cleaving (in its monocrystalline form). The external transmittance spectrum has a far-infrared multi-phonon absorption edge starting at approximately 26 μm. It is to be noted that its wavelength dispersion is quite low (DeBell et al. 1979; Hawkins 1998).

We have used the following Sellmeier temperature-dependent representation for the refractive-index dispersion of $CdTe$ (Hawkins 1998)

$$n_{CdTe}(\lambda, T) = \left(A + \frac{B\lambda^2}{\lambda^2 - C} + \frac{D\lambda^2}{\lambda^2 - E} + \frac{F\lambda^2}{\lambda^2 - G} \right)^{1/2} \tag{B.6}$$

with the corresponding coefficients given in Table B.6.

Table B.6: Temperature-dependent coefficients for $CdTe$ index representations.

Coeff.	$CdTe$
	$-2.973 \times 10^{-4}T + 3.8466$
B	$8.057 \times 10^{-4}T + 3.2215$
C	$-1.10 \times 10^{-4}T + 0.1866$
D	$-2.160 \times 10^{-2}T + 12.718$
E	$-3.160 \times 10^{1}T + 18753$

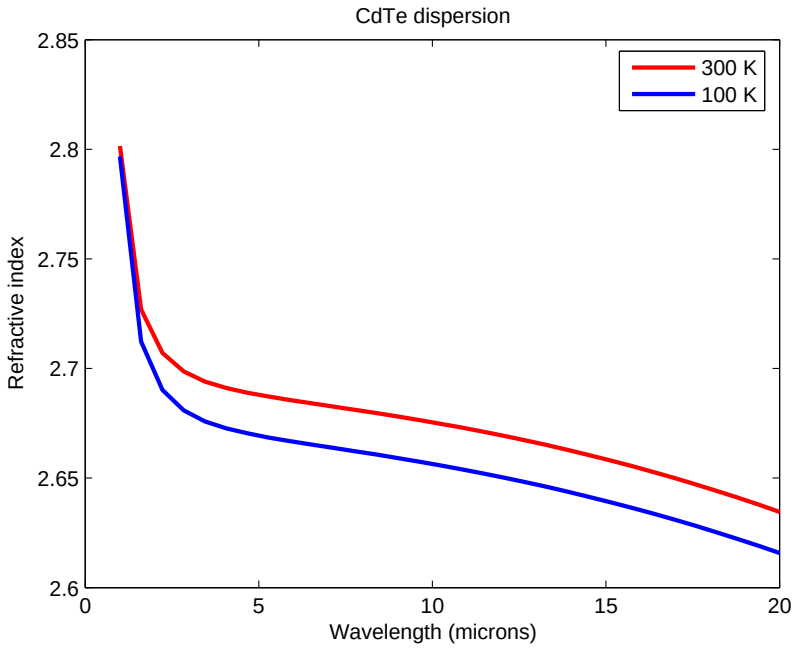

Figure B.6: $CdTe$ refractive index dispersion (left). Note that absorption is not shown because it does not occur in the bandpass of interest. Data from Hawkins (1998).

C

Drawings of the $ZnSe$ 6-14 μm TIRG APS prototype

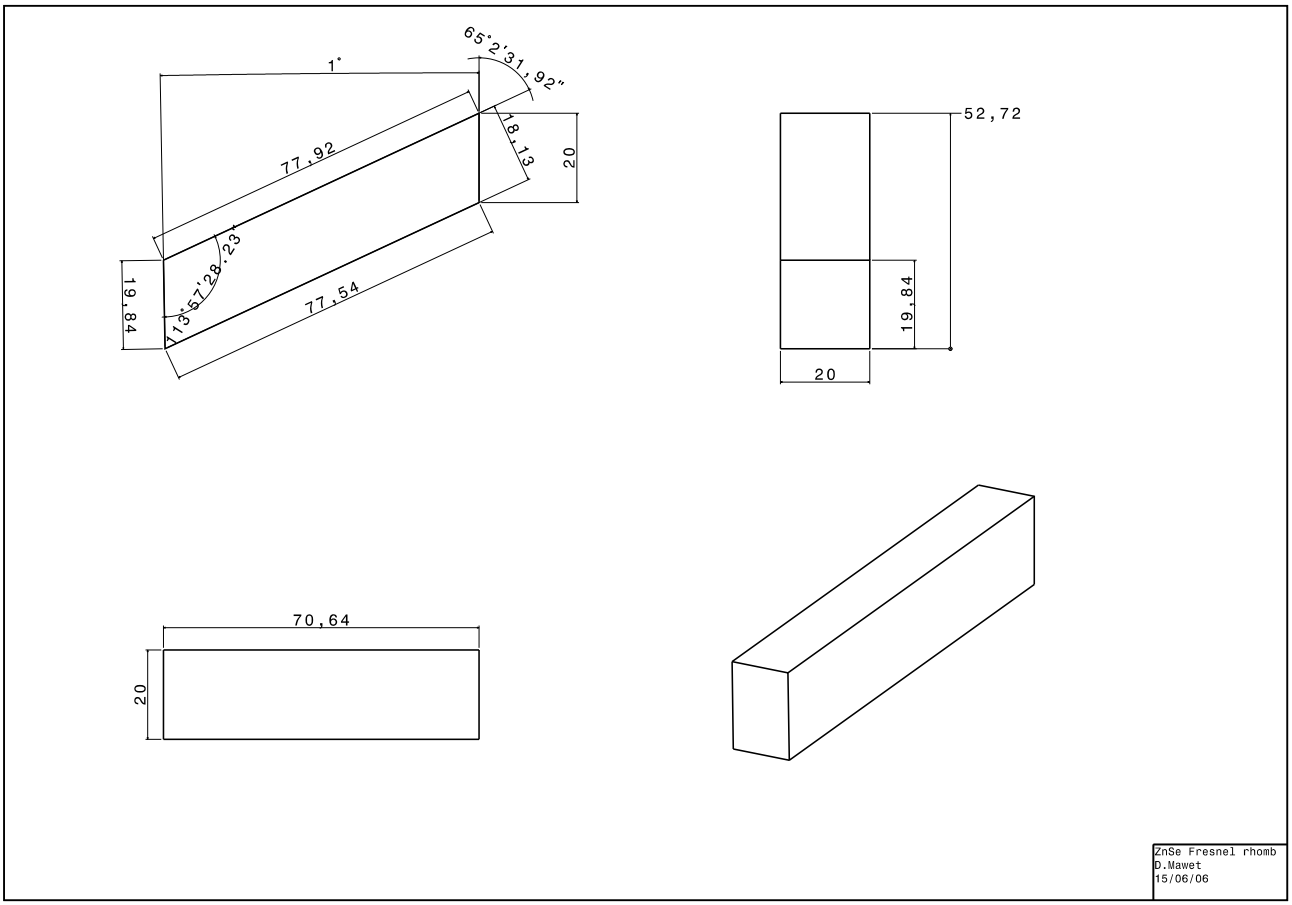

Figure C.1: $ZnSe$ rhomb TIRG APS prototype drawing.

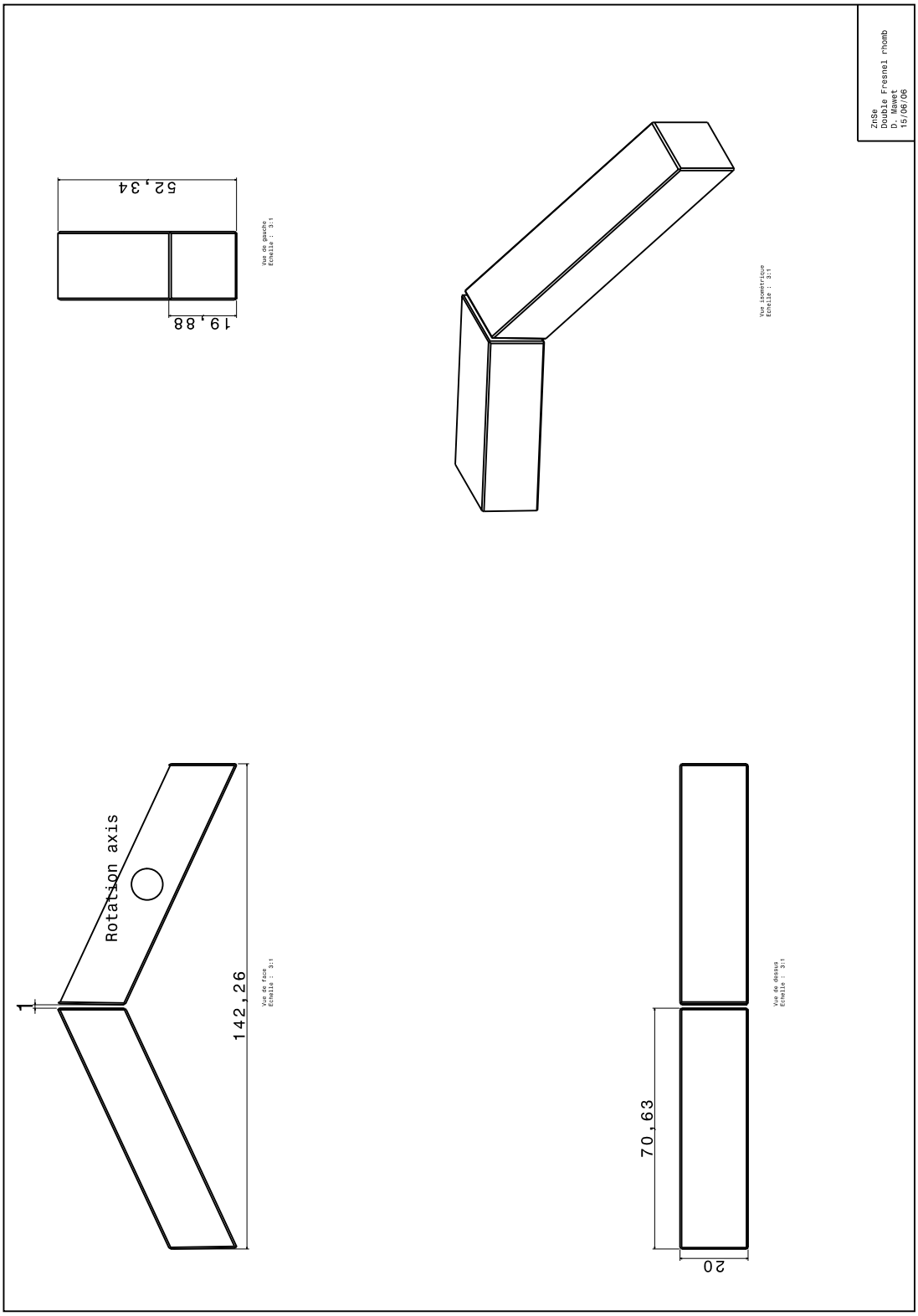

Figure C.2: *ZnSe* double rhomb TIRG APS prototype drawing.

D

Finite element analysis of a $ZnSe$ rhomb

Let us recall what has already been stressed out in Sect. 8.5.5. Given the stress-optic coefficient of $ZnSe$ ($C_{\lambda=10.6\,\mu\mathrm{m}} \approx -12$ brewsters), if we impose that the phase-shift perturbation must remain below the 10^{-3} radian level at 6 μm, i.e., below the $\sim 10^{-7}$ null depth, it implies that an uniaxial applied load must remain below $\sim 470\,\mathrm{N/m^2} = 47\,\mathrm{Pa} = 4.7\,\mathrm{gr/cm^2}$ which is very stringent.

To assess the propagation of the stresses up to the optical zones, a preliminary finite element analysis was performed with CATIA. The input configuration simulates the mount illustrated in Fig. 8.21 in the 45° configuration shown in Fig. 8.22. No applied loads other than the gravitational acceleration (9.81 $\mathrm{m/s^2}$) were considered. The mount of Fig. 8.21 was taken into account as a sliding support so that the resultant loads are normal to the contact surfaces. The rhomb therefore just leans in its mechanical mount. The contact surfaces are shown in Fig. D.2 together with the results of the finite element analysis. The following mechanical properties of $ZnSe$ were loaded in the program:

- density $5270\,\mathrm{kg/m^3}$;
- Young coefficient $E = 6.72 \times 10^{10}\,\mathrm{N/m^2}$;
- Poisson coefficient $\nu = 0.28$;
- elastic limit (rupture modulus) of $5.51 \times 10^7\,\mathrm{N/m^2}$.

The mesh for the static analysis was performed by 87177 "OCTREE" tetrahedral elements and 18758 nodes. Results in terms of Von Mises stresses[34] (mathematically defined here below) are displayed in Fig. D.2 and Fig. D.3 (cut along the major diagonal of the rhomb). The colors are chosen such that blue parts are under the 470 $\mathrm{N/m^2}$ criterium whereas the red ones are above. The calculus shows that the bottom of the rhomb is under stress beyond the specification calculated here above whereas the upper part is comfortably below.

This analysis allows us to quantify the effect of stress birefringence more precisely. Indeed, following the finite element analysis results, we can consider that one third of the rhomb (bottom) is at the mean stress value of $\sim 750\,\mathrm{N/m^2}$ while the remaining two thirds (top) are at the mean value of $\sim 150\,\mathrm{N/m^2}$. Then, the phase perturbation resumes to 6×10^{-4} radian at 6 μm, corresponding to a null depth of 1.6×10^{-7}. It is to be noted that this value is an upper bound since the Von Mises stress is a scalar defined as

$$\sigma_{VM} = \sqrt{\frac{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2}{2}} \tag{D.1}$$

[34] The von Mises stress is derived from the distortion energy theory and is a simple way to combine stresses in three dimensions to calculate failure criteria of ductile materials. In this way, the strength of material in a 3D state of stress can be compared to a test sample that was loaded in one dimension.

where σ_1, σ_2 and σ_3 are the principal uniaxial stresses[35]. Since only uniaxial stresses σ_s and σ_p perpendicular to the interferometric beam induce birefringence that can affect the propagation (birefringence along the beam direction does not enter into account), the Von Mises stress can only be an upper bound. A complete analysis linking the principal stresses (or the stress tensor) to the beam vectorial basis could directly deduce the resultant birefringence affecting the beam (see Fig. D.1). Unfortunately, such an analysis would demand modifying the software. Nevertheless, as already suggested, one can already state that stress birefringence should logically be minimized by the 45° configuration since in such a case, σ_s and σ_p should compensate for each other (to be confirmed). This symmetrization is another argument alleviating the stress birefringence issue since both components in the two interferometer arms are submitted to symmetrical constraints.

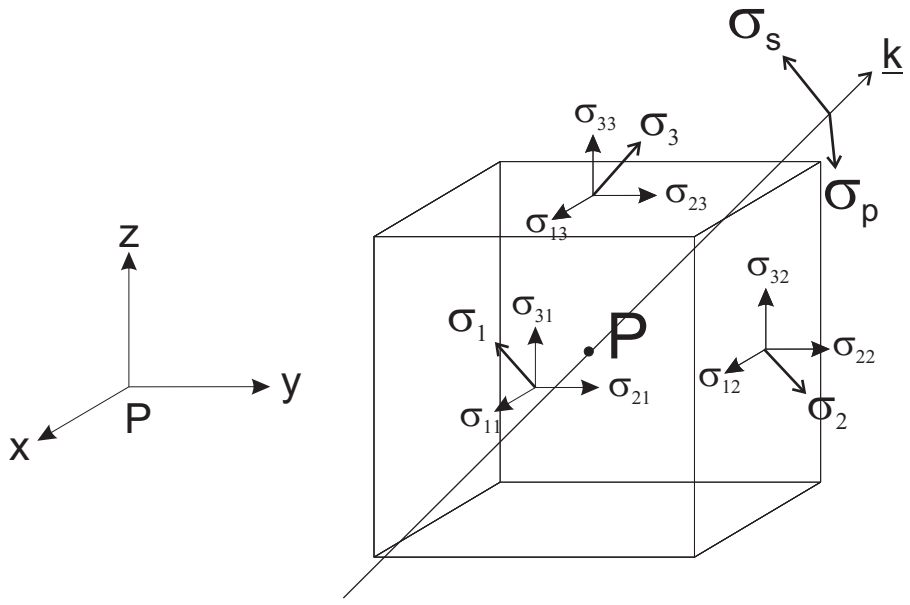

Figure D.1: Generalized stress tensor notation where the principal stresses σ_1, σ_2 and σ_3 are also shown. The beam of wave vector $\underline{k}$ passes by the point P and possesses its own reference frame defined by its direction, the s and p polarization components. The optical effect of mechanical constraints on stress birefringence is actually described by $\Delta n = C_\lambda(\sigma_s - \sigma_p)$. Symmetrization of the loads, like in the suggested 45° configuration, should equalize σ_s and σ_p, and subsequently minimize the stress-induced birefringence Δn.

[35]Principal stresses are the maximum and minimum values of the normal stresses. Eigenvalues of a stress tensor show the principal stresses, and the eigenvectors show the direction of the principal stresses (see Fig. D.1).

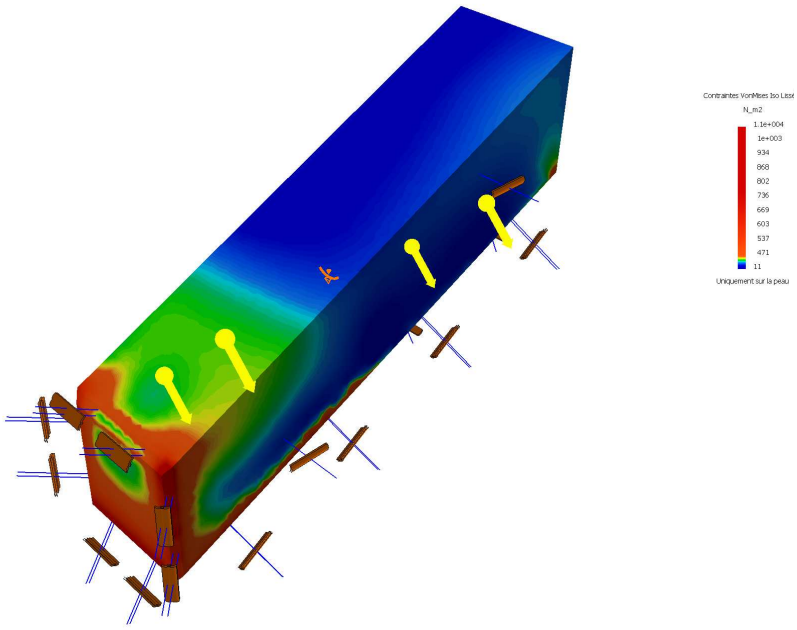

Figure D.2: Finite element analysis of the mounted $ZnSe$ rhomb showing the Von Mises stresses. The contact surfaces are pointed out by the blue double lines with a brown double shell symbolizing the sliding support. The contact are made on the whole lateral face, its bottom chamfer, on the upper half part of the bottom face and its chamfers, and on the 4 chamfers of the input face. The colors are chosen such that blue parts are under the 470 N/m^2 criterium whereas the red one are above.

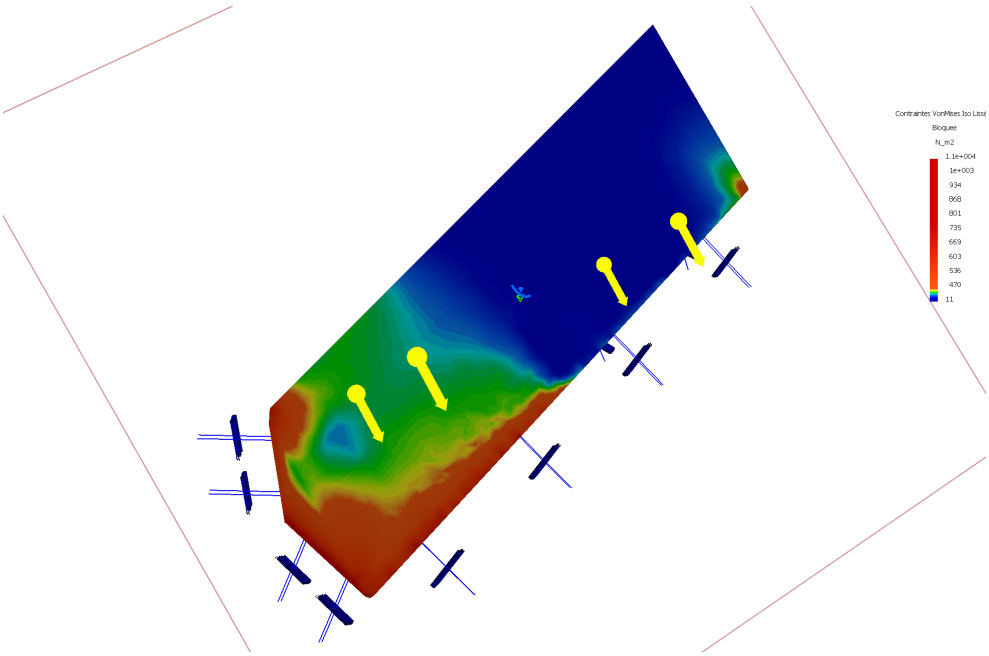

Figure D.3: Finite element analysis of the mounted $ZnSe$ rhomb: same as here above but cut along the long diagonal of the rhomb.

E

TIRG APS metrology: equipment characteristics

The surface morphology analysis is performed by Optical Interferometry thanks to a Zygo Mark IV and Minifiz:

- Fizeau interferometers;
- beam dimension of 4 inches ($\approx$ 101 mm);
- reference lenses (flat, $f/1.5$, $f/3.3$, $f/7.2$);
- accuracy of $\lambda/20$ ($\approx 0.03\,\mu$m) PV (flat reference), $\lambda/10$ ($\approx 0.06\,\mu$m) PV (spherical reference).

For the surface structure analysis, Optical Profilometry is performed with a Wyko optical profiler with the following characteristics:

- vertical resolution of 0.3 nm;
- lateral resolution of 0.1 – 13 μm;
- range of 60 μm up to 5 mm.

For surface topography and roughness analysis, a Scanning Electron Microscopy (SEM) Hitachi S3500-N Scanning Electron Microscope:

- classical "High Vacuum" mode;
- "Low Vacuum" mode for insulating samples (without metallization);
- secondary and backscattered electron detectors;
- Sample diameter of maximum 12 cm;
- resolution of 3 nm (high vac), a few tens nm (low vac);
- magnification: x30 up to x200,000.

For better vertical resolution and precision, surface topography and roughness analysis can also be performed by Atomic Force Microscopy (AFM) thanks to a Nanosurf EasyScan Dynamic Force Microscope, having the following characteristics:

- high resolution mode;
- Contact and non-contact mode;
- XY range $\leq 10\,\mu$m, Z range $\leq 2.6\,\mu$m;
- resolution: XY $\approx$ 0.3 nm , Z $\approx$ 0.05 nm.

For functional measurement, an infrared ellipsometer can be used. This kind of apparatus indeed directly measures the phase shift between the perpendicular polarization components $\Delta\phi_{TE-TM}$. The infrared spectroscopic ellipsometer SE900 from SENTECH has the following characteristics:

- spectral range from 2 to 25 μm;
- angle of incidence from 10° to 80°;
- repeatability on $\Delta\phi_{TE-TM} < 0.1° \sim 2 \times 10^{-3}$ rad;
- sample size: 30 mm x 30 mm.

F

Article: *A new family of planets? "Ocean-Planets"*

In the following paper, published in *Icarus*, "Ocean Planets" possible formation, composition, internal structure and atmosphere are considered. The existence of ocean is discussed as well as their possible exobiology interest. Our contribution to this work concerned the study of the influence of greenhouse-effect compounds on the existence of an ocean. Indeed, ice content in carbon dioxide for example would be a major obstacle to the formation of an ocean if there were not any potential physical mechanism to prevent it from being in the atmosphere. Let us summarize the content and our contribution to this paper here below.

Discussions of extrasolar planets often quietly assume that any object with mass $\sim 1 M_{\oplus}$ orbiting in a star's habitable zone will be terrestrial, i.e., composed mostly of silicates and iron-peak elements like the Earth. However, Léger et al. (2004) suggest that the habitable zones of nearby stars could harbor other similar-looking beasts. Planets composed substantially of volatiles can conceivably form with masses up to $\sim 10 M_{\oplus}$ and migrate inward to the habitable zone. Kuchner (2003) estimates that Earth-sized volatile-rich planets can retain their volatiles for billions of years at ~ 1 AU, protected by thick atmospheres that slowly undergo hydrodynamic escape. Depending on their distance to the star and properties of their atmospheres, some volatile-rich planets can develop a surface water ocean (with a depth of ≈ 100 km) and can be called Ocean Planets.

Let us consider that initially, 50% of the building blocks of the planet are made of ices that have a composition similar to that of comets, i.e., 90% H_2O, 5% NH_3, and 5% CO_2 by mass. Compared to the situation depicted in a pure ice ball case, this could cause major changes because the total amount of CO_2 and NH_3 is huge, several hundred times that present in Venus or the Earth (carbonates). If a significant part of these gases were in the atmosphere, it would have severe consequences: it would maintain it into a hot state and prevent the formation of an ocean of liquid water. This is what happens in Uranus and Neptune: because they contain 1 to 4 Earth masses of hydrogen and helium (Podolak et al. 2000), even the small intrinsic heat flux maintains the atmosphere in a state such that the temperature at its bottom is larger than water's critical temperature (647 K). Consequently, the interiors of Uranus and Neptune are fluid (Cavazzoni et al. 1999). However, four processes may limit the initial amount of CO_2 and NH_3 in the atmosphere and thus permit the existence of an ocean

- they are partially soluble in the liquid and solid. For example, at low pressure and $T \approx 300$ K, NH_4HCO_3 is soluble in water (12 g per 100 g water), as is $(NH_4)_2CO_3$ (100 g per 100 g water);

- NH_3 and CO_2 are known to easily form hydrates/clathrates when the temperature is not too high (less than 280 K for CO_2) (Leliwa-Kopystyński et al. 2002);
- phase separation in the interior may sequester elements at deeper levels, in which case they would be unavailable to form a massive atmosphere. The driving force for this sequestration is gravity. For instance, solid CO_2 is denser than solid H_2O, at least in the domain where experimental data are available (P <60 GPa; see, e.g. Yoo et al. 1999), e.g., by a factor 1.3 at 10 GPa and 1.2 at 50 GPa. If at high pressure H_2O and CO_2 separate into two phases the latter will sink down into the thick ice layer and most of the CO_2 will be locked into the solid ice mantle. How much carbon dioxide remains in the upper parts of the planet, including in the atmosphere, is an open question and conservatively, will be treated as a free parameter. However, it is pointed out that only a minute fraction of the CO_2 reservoir is in the atmosphere;
- evaporation processes may erode some of the more volatile gases, especially in the early period of high X and EUV activity of the central star (Lammer et al. 2003).

Apart from the detectability point of view, the interest of such planets is twofold, covering both planetology and exobiology. They would be a type of objects we do not have in the Solar System and would significantly extend the field of planetology. The search for a form of life similar to that which has developed on Earth would open a new field in exobiology because the conditions of the environment would be quite different from the terrestrial ones. The elements necessary to living bodies (P, S, Fe, Mg, Na, K, etc.) could be brought to the surface by micro-meteorites or found in the ocean as dissolved species.

It could even tell us something about the emergence of life on Earth. For instance, if some form of life is discovered on an Ocean Planet, it would indicate that it has occurred in the absence of black smokers because on these planets these structures are not expected, liquid water and silicates being separated by thousands of kilometers of ice. Indeed, around stars with not too high a C/O ratio, planetesimals built in the cold regions of the protoplanetary disk contain a significant fraction of water ice. In our Solar System, this is the case for all the moons of the giant planets except Io. Uranus and Neptune can themselves be considered as "ice giants": their interior density is indeed very similar to that of compressed water ice (Podolak et al. 2000). However, Uranus and Neptune also contain about 1 to 4 Earth masses of hydrogen and helium in the form of an outer envelope. The case of these new planets would be different because they contain much less hydrogen.

Mid-future space missions searching for planetary transits in the habitable zone (Kepler), possibly coupled with radial velocity follow-up, should provide us with valuable information about their existence and properties. Indeed, by coupling the radial velocity information of mass to the transit information about radius, density can be inferred and thus their composition. If they are as resistant with respect to the evaporation and photolysis of their atmospheres as some models predict (Kuchner & Spergel 2003; Selsis et al. 2006) CoRoT (launch scheduled in 2007) will detect the hottest ones.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Icarus 169 (2004) 499–504
ICARUS
www.elsevier.com/locate/icarus

Note

A new family of planets? "Ocean-Planets"

A. Léger,[a,*] F. Selsis,[b] C. Sotin,[c] T. Guillot,[d] D. Despois,[e] D. Mawet,[a] M. Ollivier,[a] A. Labèque,[a] C. Valette,[a] F. Brachet,[a] B. Chazelas,[a] and H. Lammer[f]

[a] *Institut d'Astrophysique Spatiale, bat 121, CNRS, University Paris-Sud, F-91405 Orsay, France*
[b] *Centro de Astrobiología, 28850 Torrejón de Ardoz, Madrid, Spain*
[c] *Géophysique, Université de Nantes, F-44321 Nantes cedex 3, France*
[d] *Observatoire de la Côte d'Azur, BP 4229, F-06304 Nice cedex 04, France*
[e] *Observatoire de Bordeaux (INSU/CNRS), B.P. 89, F-33270 Floirac, France*
[f] *Space Research Institute, Austrian Academy of Sciences, Schmiedlstrasse 6, 8042 Graz, Austria*

Received 10 July 2003; revised 22 December 2003

Available online 19 March 2004

Abstract

A new family of planets is considered which is in between the rocky terrestrial planets and the gaseous giants, "Ocean-Planets." We present the possible formation, composition and internal structure of these putative planets. We consider their oceans, as well as their possible Exobiology interest. These exoplanets should be detectable by Space missions such as Eddington, Kepler, and possibly COROT (launch scheduled in 2006). They have a density lower than that of rocky planets. Their rather large radius would make them attractive targets for exoplanet spectroscopic missions such as Darwin/TPF, all the more because a robust biosignature appears to exist.

Keywords: Extrasolar planets; Ices; Exobiology

1. Introduction

The extrasolar planetary systems discovered thus far show a surprising diversity of orbital parameters. Most of these systems do not resemble our own Solar System. There is presently no consensus as to why these planetary systems are so diverse, but it seems that migration due to interactions between planets and the protoplanetary disk is an important ingredient (Lin et al., 1996; Ward, 1997; Trilling et al., 1998). It seems reasonable to assume that planets resembling our Uranus and Neptune, or slightly less massive ones, may have formed in cold regions of a protoplanetary disk and migrated inward, possibly into the so-called "Habitable-Zone" where liquid water can be present at their surface. These planets would be extremely interesting as their large radius makes them rather easily detectable by transit missions (COROT, Eddington, Kepler) and analysable by Darwin/TPF. A planet with twice the Earth radius requires an integration time 16 times shorter than an Earth analogue, for the same distance and S/N conditions.

The interest of such planets is twofold, covering both Planetology and Exobiology. They would be a type of objects we do not have in the Solar System and would significantly extend the field of Planetology. The search for a form of life similar to that which has developed on Earth would open a new field in Exobiology because the conditions of the environment would be quite different from the terrestrial ones. The elements necessary to living bodies (P, S, Fe, Mg, Na, K, ...) could be brought to the surface by micrometeorites or found in the ocean as dissolved species. It could even tell us something about the emergence of life on Earth. For instance, if some form of life is discovered on an Ocean-Planet, it would indicate that it has occurred in the absence of Black Smokers because on these planets these structures are not expected, liquid water and silicates being separated by thousands of kilometres of ice.

Around stars with not too high a C/O ratio, planetesimals built in the cold regions of the protoplanetary disk contain a significant fraction of water ice. In our Solar System, this is the case for all the moons of the giant planets except Io. Uranus and Neptune can themselves be considered as "ice giants": their interior density is indeed very similar to that of compressed water ice (Podolak et al., 2000). However, Uranus and Neptune also contain about 1 to 4 Earth masses of hydrogen and helium in the form of an outer envelope. The case of the planets we consider in this Note is different because they contain much less hydrogen.

Of course, many parameters drive the structure and composition of planets. As planet formation is not well understood, and to limit the scope of this study, we choose to make the following assumptions:

(1) Only planets with masses in the range $1 < M/M_{\oplus} < 8$ are considered. The lower boundary is for the selection of planets that are easier to detect, and the upper one for objects that have not accreted a large amount of H_2 (Wuchterl et al., 2000). It is pointed out that the biggest objects of that type ($M \sim 6$–$8M_{\oplus}$) *are of special interest* because they are accessible to Transit detection and possibly to Radial Velocity measurements so that *their radius and mass could be determined simultaneously*.

* Corresponding author.
E-mail address: alain.leger@ias.u-psud.fr (A. Léger).

0019-1035/$ – see front matter
doi:10.1016/j.icarus.2004.01.001

(2) These planets are supposed to form relatively far from the star (further than the snowline, e.g., $\sim$ 5–10 AU) and migrate inward (to $\sim$ 1 AU) on a timescale which is of the order of $\leqslant$ 1 Myr (e.g., Lin et al., 1996).
(3) We assume that the planets are mostly made of refractory material (metals and silicates) and ices. From what we know of the composition of the protosolar nebula, we envision respective fractions of $\sim$ 50% of metals and silicates, and $\sim$ 50% ices, by mass. Because of the relatively small masses and closeness to the star after the initial migration, the planets should contain only trace amounts of hydrogen and helium (Lammer et al., 2003). Most of the ices should be in the form of H_2O, but their exact composition is crucial and will be discussed further.

The questions that arise are:

(i) What should be the internal structure of these planets?
(ii) Under which conditions will a water ocean be present, and what will be its depth?
(iii) What are their atmospheric properties?

In this Note, we attempt to provide the first steps towards answering these questions.

2. Internal structure

2.1. An outline of the cooling history of a planetary water ball

The knowledge of the temperature of the planetary interior is essential for determining its structure. Depending on the T and P profiles, the interior may be solid, liquid or gaseous. For simplicity, we first examine the case of a giant ball made of pure water.

The energy acquired by the planet (mass M and radius R) is of the order of, e.g., $\sim (1/2)\varpi GM^2/R$, G being the gravitational constant, ϖ a factor accounting for the density distribution inside the planet ($\varpi = 3/5$ for a uniform density) and assuming that half of the energy was radiated directly to space (perfect gas approximation). This would correspond to a temperature $T_i \sim \mu\varpi GM/k_B R \sim$ 25,000 to 75,000 K for $M = 1$ to $6M_\oplus$ if all this energy was converted into heat and stored in the gaseous planet ($\mu \sim 18$ is the mass of a water molecule, k_B the Boltzmann constant, we used $R \sim$ 10,000 to 20,000 km). This estimate is very rough (for example, and among many simplifications, it neglects the fact that accretion did not entirely take place at the same time and that water is partially decomposed at these temperatures) but it shows that these planets could have had an early stage in which they were at least partially gaseous. We now show that this stage was necessarily short.

Considering first an isolated planet and assuming a gaseous atmosphere of pure water (with a grey opacity $\kappa \sim 0.1$ cm^2 g^{-1}, e.g., Auman, 1966). The level at which the atmospheric temperature T is equal to the effective temperature $T_{\rm eff}$ is for an optical depth ~ 1 (Eddington approximation) or pressure $P_0 \sim 2g/3\kappa$, g being the surface gravity, i.e., a few mbars. The assumption of an atmosphere of pure water vapour implies, in order for water to remain in vapour form at P_0, that $T_{\rm eff} \geqslant 270$ K. The planet can spend at most $\tau \sim E_g/(4\pi R^2 \sigma T_{\rm eff}^4) \sim 5$ to 20 Myr in this partially gaseous state (this timescale could be shorter because the effective temperature is likely higher at the beginning).

When the cooling of the atmosphere has led to a temperature lower than 270 K water begins to condense and falls down towards the planetary interior where it vaporises, absorbing the corresponding latent heat. This process provides an efficient cooling of the interior. Progressively, it cools down until condensed water cannot vaporise anymore and sinks towards the planetary centre as high-pressure ice. A thick ice shell forms. If the accretion rate is not high enough, the planet may never reach this gaseous state for water.

The case of a planet at a given distance from its star is discussed in Section 2.4. The basic cooling mechanism remains the same but the planet effective temperature is now governed by the radiative exchange with its surroundings. The migration time ($\leqslant$ 1 Myr) is shorter than the upper limit for the cooling time that we derived. Then, it is possible that an Ocean-Planet cools mainly when at $\sim$ 1 AU.

The presence of silicates/iron does not qualitatively alter this analysis: "rocks" are expected to rapidly differentiate from ices, settle towards central regions, and adjust thermally to the surrounding envelope by conduction.

Note that Io, Europa, and Ganymede are differentiated, as are all other major solid planets of the Solar System.

2.2. Composition of ices

We expect that initially, ices had a composition similar to that of comets, i.e., 90% H_2O, 5% NH_3, and 5% CO_2 by mass. Compared to the situation depicted in the pure ice ball case, this could cause major changes because the total amount of CO_2 and NH_3 is huge, several hundred times that present in Venus or the Earth (carbonates). If a significant part of these gases were in the atmosphere, it would have severe consequences: it would maintain the atmosphere into a hot state and prevent the formation of an ocean of liquid water. This is what happens in Uranus and Neptune: because they contain 1 to 4 Earth masses of hydrogen and helium (Podolak et al., 2000), even the small intrinsic heat flux maintains the atmosphere in a state such that the temperature at its bottom is larger than water's critical temperature (647 K). Consequently, the interiors of Uranus and Neptune are fluid (e.g., Cavazzoni et al., 1999).

However, four processes may limit the initial amount of CO_2 and NH_3 in the atmosphere:

(i) they are partially soluble in the liquid and solid. For example, at low pressure and $T \sim 300$ K, NH_4HCO_3 is soluble in water (12 g per 100 g water), as is $(NH_4)_2CO_3$ (100 g per 100 g water);
(ii) NH_3 and CO_2 are known to easily form hydrates/clathrates when the temperature is not too high (less than 280 K for CO_2) (Leliwa-Kopystynski et al., 2002; Sloan, 1998);
(iii) phase separation in the interior may sequester elements at deeper levels, in which case they would be unavailable to form a massive atmosphere.

The driving force for this sequestration is gravity. For instance, solid CO_2 is denser than solid H_2O, at least in the domain where experimental data are available ($P < 60$ GPa; Yoo et al., 1999; Hemley et al., 1987), e.g., by a factor 1.3 at 10 GPa and 1.2 at 50 GPa. If at high pressure H_2O and CO_2 separate into two phases the latter will sink down into the thick ice layer and *most of the* CO_2 *will be locked into the solid ice mantle*. How much carbon dioxide remains in the upper parts of the planet, including in the atmosphere, is an open question and conservatively, will be treated as a free parameter. However, it is pointed out that only a minute fraction of the CO_2 reservoir is in the atmosphere;

(iv) the evaporation processes discussed in Section 4.1 may erode some of the more volatile gases, especially in the early period of high X and EUV activity of the central star (Lammer et al., 2003).

2.3. Internal structure modelling

A model for planet interior is used that has been developed for the Earth interior and expanded to extrasolar terrestrial planets (Dubois, 2002). Quantities depend upon the radius, r (1D model). They are the planetary mass located between 0 and r, local density, gravity and pressure, $m(r)$, $\rho(r)$, $g(r)$, and $P(r)$, respectively.

Density is a function of material, pressure and temperature. The latter dependence is not major and will be neglected. High-pressure laboratory experiments have been performed that provide equations of state $\rho(P)$ for the different materials. The adopted relations are as follows:

– main metals in the centre of a telluric planet are Fe and Ni. The equation of state of iron by Anderson and Ahrens (1994), is used;
– main kinds of silicates in the Solar System are the Fe and Mg rich silicates. Their phase at high pressure is Perovskite ($P > 23$ MPa; Anderson, 1997) with the $\rho(P)$ relation as measured by Duffy and Arhens (1995);
– the phase diagram of H_2O is complex, with 10 ice phases (Durham and Stern, 2001). The ice $\rho(P)$ relation has been measured up to 128 GPa by Hemley et al. (1987) and we extrapolate it for higher pressures.

Inputs are: the total mass of the planet, the fractions of metals, silicates and ices. Outputs are the outer planetary radius, the outer radius of the different material shells, variation with radius of the different physical quantities.

For a given mass of the planetary components, a self-consistent solution of the density, gravity and pressure functions, $\rho(r)$, $g(r)$, and $P(r)$, is required that fulfils the mass conditions

$$M_i = 4\pi \int_{R_i-1}^{R_i} \rho_i(r) r^2 dr, \tag{1}$$

where the index, i, corresponds to the metal, silicates and ice component, respectively, and R_i the outer radius of component (i); the gravity and pressure relations are

$$g(r) = Gm(r)/r^2, \tag{2}$$

$$P(r) = \int_{r}^{R_{\rm pl}} \rho(x) g(x)\, dx, \tag{3}$$

where $R_{\rm pl}$ is the planetary radius. *For a $6M_\oplus$ planet*, with the relative amounts of material as described in Section 1 and deduced from the Earth's values (Javoy, 1999), i.e., m 17% metals, 33% silicates, and 50% ices, the internal structure calculated is shown in Fig. 1. *The planetary radius is* $R_{\rm pl} = 2.0R_\oplus$, central pressure 1600 GPa and surface gravity $1.54g_\oplus$. The internal structure of a $6M_\oplus$ rocky planet and that of Earth are also shown for comparison.

2.4. Thermal structure

The interior structures may fall into three categories, depending on the stellar distance, orbital history and atmospheric composition:

(1) a hot gaseous planet where the temperature at $P = 22.1$ MPa (221 bars) is higher than the critical temperature of H_2O, 647 K, with a continuous transition from the outer atmospheric layers to the inner supercritical fluid (no surface);
(2) a planet with a liquid water surface;
(3) a planet with an icy surface. See also Stevenson (1999) for the case of isolated planets.

In the present paper we study case (2), that of an "Ocean-Planet," where an ocean is present with a surface temperature, T_S, in between the triple and critical temperature of water, 273 and 647 K. The Habitable Zone can be defined as distances to the star leading to case (2) and the fraction of case (3) where the surface ice can melt during the local summer.

Thereafter it is shown that the temperature profile $T(r)$, in the ice shell, follows the liquid/solid transition curve with pressure $T[P(r)]$ (Fig. 2 for low pressures).

The H_2O/silicate interface is derived to be at the temperature of the melting point of ice at the local pressure. For a $6M_\oplus$ Ocean-Planet, this pressure is 250 GPa (Fig. 1 caption) and the corresponding melting point of ice VII is 1150 K, as estimated from the Simon fusion equation in Mishima and Endo (1978). It must be noted that this equation predicts very well the later measurements by Fei et al. (1993).

It can be argued that the temperature at this interface is higher than the ice melting temperature and most of the water shell liquid. If so, a process analogous to that described in Section 2.1 would lead to the sinking of (high-pressure/high-density) ice and the fast build-up of an ice shell.

It can also be argued that the temperature at the H_2O/silicate interface is much lower than the melting temperature of ice VII and that internal heat is transferred by subsolidus convection in the ice layer. However, simple comparison with models built for the internal dynamics of the icy satellites of the giant planets in the Solar System shows that subsolidus convection in ice is not even sufficient to remove the heat produced by the decay of the radiogenic elements contained in the silicate shell (e.g., Grasset et al., 2000).

The only possibility is that heat is transferred by the melting of ice and migration of the melt towards the upper ocean. The amount of melt produced by the internal heating in the kind of planet described in this Note is on the order of 1 km thick layer per Myr at the interface. This is quite small compared to the size of the ice layer. Consequently, the temperature profile in the ice mantle follows the solid/liquid transition curve, i.e., from 1150 K at the silicate interface to the temperature at the bottom of the ocean.

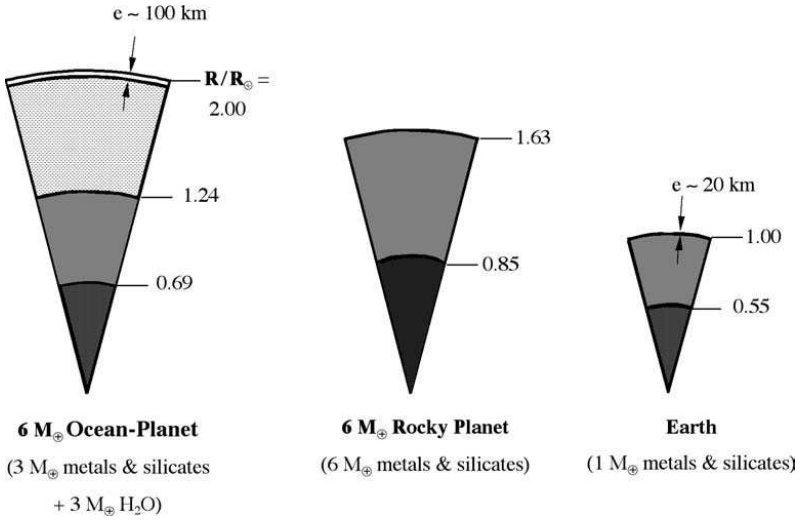

Fig. 1. From left to right: (1) calculated internal structure of a $6M_\oplus$ Ocean-Planet. Constituents are, from the centre to the outside, $1M_\oplus$ metals, $2M_\oplus$ silicates, and $3M_\oplus$ ice. The density (g cm^{-3}) at the centre, different interfaces and top is: 19.5, 15.6–8.2, 6.2–3.9, and 1.54; gravity ($g_\oplus$): 0, 2.1, 1.96, and 1.54; pressure (GPa): 1580, 735, 250, and ~ 1. The upper layer is a ~ 100 km thick ocean. The mean planetary density is 4.34 g cm^{-3}; (2) idem for a rocky planet with the same total mass ($2M_\oplus$ metals, $4M_\oplus$ silicates). Density is: 21.0, 15.5–8.1, and 4.1; gravity: 0, 2.72, and 2.24; pressure: 2200, 745, and 0. The mean planetary density is 7.74 g cm^{-3}; (3) for comparison, the structure of the Earth calculated with the same model is shown. It agrees fairly well with the actual one. Density is: 13, 9.5–5.2, 3.3; gravity: 0, 1.04, 1.00; pressure: 340, 130, 0. The mean planetary density is 5.57 g cm^{-3}.

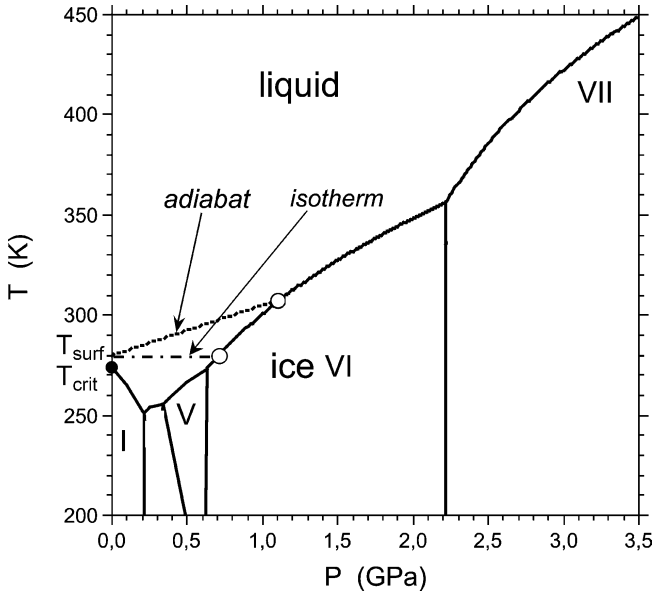

Fig. 2. Phase diagram for water (full lines). The dotted (dashed-dotted) line describes the temperature adiabatic (isothermal) profile in the ocean that results from a given surface value (here, 280 K = 7 °C). As explained in the text, the actual dependence is expected to be in between the adiabatic and the isothermal cases. In the adiabatic case, the ocean depth is 72 km, the ocean bottom temperature and pressure are 35 °C and 1.1 GPa. For a higher (lower) surface temperature, the ocean depth would be larger (smaller), i.e., $T_{\rm surf} = 30\,^{\circ}$C (0 °C) leads to a 133 km (60 km) ocean depth. A flatter T profile would lead to a shallower ocean. In the case of a isothermal profile, the depth is 40, 45, and 65 km for $T_{\rm surf} = 0$, 7, and 30 °C, respectively. The gas/solid or liquid transition curve is, at the scale of the figure, indistinguishable from the $P = 0$ vertical line. The filled circle is the triple point. The critical point ends that curve at ($T = 647$ K, $P = 22.1$ MPa). In the solid phase, the temperature follows that of the liquid/solid transition (Section 2.4).

The pressure in the silicate shell is larger than what is found in the Earth's interior. The behaviour of silicates in this domain is not known. By analogy with our planet, we assume that the silicates are composed of perovskite and magnesowustite and that the fossil heat plus that produced by the decay of long-lived radiogenic elements in the silicate shell, are transferred by subsolidus convection (e.g., Schubert et al., 2000). Without entering into the details of this convection one can say that the temperature profile within the silicate shell is made of a conductive temperature gradient in the cold thermal boundary at the top of the shell, an adiabatic temperature gradient in the well-mixed interior and a conductive temperature gradient at the hot thermal boundary layer at its bottom.

A picture of the temperature profile results. In the metal/silicate core, the temperature decreases from a high value in the centre (e.g., ~ 10,000 K for a $6M_{\oplus}$ Ocean-Planet by comparison between with the temperature at the centre of the Earth, 5000 K, and that for Jupiter, 20,000 K) to the ice fusion temperature at the silicate/ice boundary (e.g., 1150 K for that planet). In the ice shell, it follows the water liquid/solid transition to the ocean bottom temperature. The latter is discussed in Section 3.

3. Ocean

When such a planet is in the Habitable Zone of its star, its surface is liquid. A question arises, "how deep is the ocean?." As a first step, for the sake of simplicity, an ocean of pure water is considered.

The temperature is a function of depth, z, with $z = R_{\rm pl} - r$, and therefore pressure. The internal heat of the planet is transferred through the ocean. The *maximum* temperature gradient is adiabatic, otherwise convection in water would be triggered which is extremely efficient to transfer energy. We assume an adiabatic dependence for the temperature but it could be weaker, down to an isothermal one. One can note that for the Earth's ocean, the temperature at the sea floor is even lower than the temperature at the surface as a result of streams.

The adiabatic relation is:

$$dT/dP = \alpha T/(\rho_{\rm water} C_{\rm p}), \tag{4}$$

where α is the thermal expansion coefficient and $C_{\rm p}$ the heat capacity. For $P <$ 1–2 GPa, values are $\alpha \approx 3 \times 10^{-4}$ K^{-1}, $\rho_{\rm water} \approx 1$ g cm^{-3}, $C_{\rm p} \approx 4$ J K^{-1} g^{-1}, and the temperature is given by:

$$T_{\rm adiabat}(z) \approx T_{\rm surface} + 20P \ ({\rm GPa}). \tag{5}$$

The T [$P(z)$] position in the phase diagram indicates whether water is liquid or solid (Fig. 2). The $P(z)$ relation depends upon the surface gravity g, the liquid water density at low pressure ρ_0 and its compressibility K. It reads $P(z) = -K\,{\rm Ln}(1 - g\rho_0 z/K)$. The bottom of the ocean is obtained when the liquid water temperature crosses the solidification line. The ocean depth, bottom temperature and pressure result (Fig. 2). For an adiabatic dependence of the temperature, a surface temperature $T_{\rm surf} = 7\,^{\circ}$C leads to an ocean depth of 72 km. For a higher (lower) $T_{\rm surf}$, the ocean depth would be larger (smaller).

If the T dependence is less steep than adiabatic, e.g., nearly isothermal, the ocean would be shallower.

4. Atmosphere

4.1. Composition

On Ocean-Planets, the expected primary volatiles are H_2O, NH_3, and CO_2. As mentioned, we assume, as an example, an initial composition of ice m 90% H_2O, 5% NH_3, and 5% CO_2. For carbonaceous compounds, this is an important difference with the atmosphere of bodies further away from their stars, as Titan in the Solar System, where carbon is mainly in species such as CH_4 because of the lower temperatures (Reynolds et al., 1987).

NH_3 is very sensitive to UV, especially for $200 < \lambda < 300$ nm where the opacity due to other species is negligible. The fraction of its initial reservoir that happens to be in the atmosphere will be photodissociated and converted into N_2 and H_2 in less than 2 Myr for a planet located at 1 AU from a G2V star. The produced H_2 is subject to hydrodynamical escape and can sweep away a fraction of N_2 because the atmosphere develops an exosphere with a high temperature $T_{\rm exo}$ governed by heating by extreme UV and stellar particles (Lammer et al., 2003). $T_{\rm exo}$ obtained with the XUV heating alone can be over 10,000 K which is much higher than the planet effective temperature $T_{\rm eff}$ (255 K for Earth).

The presence of these radiations has been observed in solar proxies by numerous satellites (ASCA, ROSAT, EUVE, FUSE, IUE). It indicates that a young GV star has continuous flare events producing a radiation environment several hundred times more intense than the solar environment today (Guinan and Ribas, 2002).

After the majority of the hydrogen is lost and the heavier atmospheric compounds begin to dominate in the upper atmosphere, a part of the nitrogen should be dragged away by a diffusion limited hydrogen flow (Zahnle et al., 1990). Because the amount of nitrogen left is not sure, the partial pressure of N_2 in the atmosphere is treated as a free parameter in the present paper. So is the CO_2 pressure.

4.2. Planet IR emission

The planet IR emission must be calculated in a self-consistent way as a function of the atmospheric composition and the stellar irradiation. A key ingredient is the temperature profile. It is to be remembered that an isothermal atmosphere would have no spectral feature, whatever its composition. As a preliminary calculation, for a given ground temperature, an atmosphere

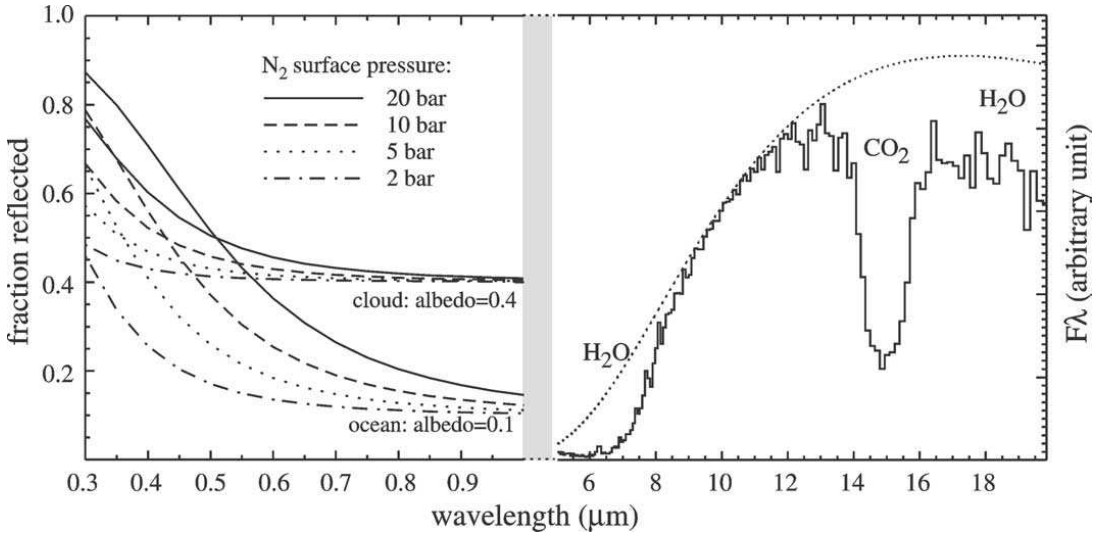

Fig. 3. Reflected and emitted light by a $6M_\oplus$ Ocean-Planet, versus wavelength. In the UV-visible, the fraction of the stellar flux reflected by the planet is shown for different values of the ground pressure of N_2, this gas being the main atmospheric component, and for situations with and without clouds. The light from the planet is the product of the stellar emission by this fraction. The former decreases with wavelength ($\propto \lambda^{-4}$) and becomes negligible at $\lambda > 3$ μm. In the IR, the emission is from the same planet with $T_{\text{surface}} = 320$ K $= 47\,^\circ$C, $P_{N_2} = 5$ bars and an atmosphere saturated with water vapour. Its cold tropospheric trap ($T \sim -65\,^\circ$C) is at 15 km. It contains CO_2 with a mixing ratio of 10^{-5}. As it is wet, and pretty warm at its surface, the planet is assumed to be cloudy with a full coverage of clouds at $h = 5$ km. Note the CO_2 absorption band at 15 μm and the strong H_2O ones at $\lambda < 8$ μm and $\lambda > 16$ μm.

saturated with water vapour is considered and an adiabatic decrease of T is calculated down to an arbitrary value of 180 K. This limitation is to prevent T from reaching unrealistic low values.

Figure 3 shows the interesting case of an Ocean-Planet with a 320 K (47 °C) surface temperature. This case of a rather hot planet with a super-humid atmosphere has no analogue in the Solar System. Both stellar reflected light and IR emission are shown.

4.3. Biosignatures?

Usually, O_2 or O_3, in association with H_2O and CO_2 are candidates biosignatures in a planetary atmosphere. To qualify them for an Ocean-Planet, one has to show that there are no abiotic processes that can produce a similar gas mixture.

Two cases of planetary atmospheres may result in a photochemical build-up of O_2 (Selsis et al., 2002):

(i) CO_2-rich atmospheres: $P_{CO_2} > 50$ mbar for dry atmospheres, $P_{CO_2} > 0.5$ bar for wet atmospheres, and
(ii) water-rich atmospheres loosing hydrogen to space after H_2O photolysis (for instance during a runaway greenhouse effect as described by Kasting, 1988).

Both cases require weak oxygen sinks (oxidation of rocks by soil weathering, and volcanic gases) to allow O_2 to reach significant levels.

Ocean-Planets may gather all these characteristics: a possible high CO_2 abundance, a water-rich upper atmosphere when the orbital distance is closer than the runaway greenhouse limit (Kasting et al., 1993) and low oxygen sinks in the absence of rocky surface and volcanism. Especially, due to their inexhaustible water reservoir, runaway greenhouse phases may last for most of their lifetime while, on Venus for instance, this phase is thought to be shorter than a few hundred million years (Kasting, 1988). Consequently, *the presence of* O_2 *is not a reliable biosignature on Ocean-Planets* because it can be the result of abiotic processes.

Can O_3 provide a better one? If O_2 is produced thanks to H_2O photolysis at high altitude, resulting hydrogenous radicals as $H^\bullet$, $OH^\bullet$, and $HO_2^\bullet$ react efficiently with O_3 and prevent the formation of a dense and detectable ozone layer. Producing abiotic O_2 from CO_2 photolysis implies a high CO_2 partial pressure that masks the O_3 band in the thermal infrared spectrum. Thus, as on terrestrial planets (Selsis et al., 2002), *the simultaneous detection of* O_3, H_2O, *and* CO_2 in the planetary mid-IR spectrum appears to be most reliable signature of an oxygen build-up of biological origin. This points out *the superiority of* O_3, *with respect to* O_2, *as a biosignature* (in presence of H_2O and CO_2).

5. Conclusion

We have shown that massive ice-rich planets possibly form in external regions of protoplanetary disks and migrate inward. Depending on their distance to the star and properties of their atmospheres, some of them can develop a surface water ocean and can be called "Ocean-Planets." Such oceans have a thickness of ~ 100 km. For a given mass, they have a density significantly lower than rocky planets and therefore a larger radius, a favourable feature for their detection and study.

Mid-future space missions searching for planetary transits in the Habitable Zone (Eddington, Kepler), possibly coupled with Radial Velocity follow-up, should provide us with valuable information about their existence and properties. If they are as resistant, with respect to the evaporation and photolysis of their atmospheres as some models predict (Kuchner, 2003) COROT (launch scheduled in 2006) will detect the hottest ones. They have a density lower than that of rocky planets. Their rather large radius makes them attractive targets.

Acknowledgments

We are grateful to Cécile Engrand, Olivier Grasset, Anthony Jones, Tobias Owen, and Daniele Pinti for their valuable help during this work. We thank N. Woolf and M. Marley for their relevant comments on the manuscript.

References

Anderson, O.L., 1997. Finding the isentropic density of Perovskite. J. Geophys. Res. Lett. 24, 213–216.

Anderson, W.W., Ahrens, T.J., 1994. An equation of state for liquid iron and implications for the Earth's core. J. Geophys. Res. 99, 4273–4284.

Auman Jr., J., 1966. The infrared opacity of hot water vapour. Astrophys. J. Suppl. 14, 171–206.

Cavazzoni, C., Chiarotti, G.L., Scandolo, S., Tosatti, E., Bernasconi, M., Parrinello, M., 1999. Superionic and metallic states of water and ammonia at giant planet conditions. Science 283, 44–46.

Dubois, V., 2002. Dynamique des satellites co-orbitaux et relations (M–R-composition) pour des exoplanètes de type terrestre. Thèse de doctorat. Université de Nantes, France.

Duffy, T.S., Arhens, T.J., 1995. Compressional sound velocity. J. Geophys. Res. 100, 529–542.

Durham, W.B., Stern, L.A., 2001. Rheological properties of water ice—applications to satellites of the outer planets. Annu. Rev. Earth Planet. Sci. 29, 295–330.

Fei, Y., Mao, H., Hemley, R., 1993. Thermal expansivity, bulk modulus, and melting curve of H_2O ice VII to 20 GPa. J. Chem. Phys. 99 (7), 5369–5373.

Grasset, O., Sotin, C., Deschamps, F., 2000. On the internal structure and dynamics of Titan. Planet. Space Sci. 48, 7–8, 617–636.

Guinan, E.F., Ribas, I., 2002. Our changing Sun the role of solar nuclear evolution and magnetic activity on Earth's atmosphere and climate. In: Montesinos, B., Gimenez, A., Guine, E.F. (Eds.), The Evolving Sun and its Influence on Planetary Environments. In: ASP Conf. Ser., vol. 269, pp. 85–107.

Hemley, R.J., Jephcoat, A.P., Mao, H.K., Zha, C.S., Finger, L.W., Cox, D.E., 1987. Static compression of H_2O–ice to 128 GPa (1.28 Mbar). Nature 330, 737–739.

Javoy, M., 1999. Chemical Earth models. C. R. Acad. Sci. Paris 329, 537–555.

Kasting, J.F., 1988. Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus. Icarus 74, 472–494.

Kasting, J.F., Whitmire, D.P., Reynolds, R.T., 1993. Habitable zones around main sequence stars. Icarus 101, 108–128.

Kuchner, M.J., 2003. Volatile-rich Earth-mass planets in the habitable zone. Astrophys. J. 596, L105–L108.

Lammer, H., Selsis, F., Ribas, I., Guinan, E.F., Bauer, S.J., Weiss, W.W., 2003. Atmospheric loss of exoplanets resulting from stellar X-ray and extreme-ultraviolet heating. Astrophys. J. 598, L121–L124.

Leliwa-Kopystynski, J., Maruyama, M., Nakajima, T., 2002. The water–amonia phase diagram up to 300 MPa: application to icy satellites. Icarus 159, 518–528.

Lin, D.N.C., Bodenheimer, P., Richardson, D.C., 1996. Orbital migration of the planetary companion of 51 Pegasis to its present location. Nature 380, 606–607.

Mishima, O., Endo, S., 1978. Melting curve of ice VII. J. Chem. Phys. 68 (10), 4417–4418.

Podolak, M., Podolak, J.I., Marley, M.S., 2000. Further investigations of random models of Uranus and Neptune. Planet. Space Sci. 48, 143–151.

Reynolds, R.T., McKay, C.P., Kasting, J.F., 1987. Europa, tidally heated oceans, and habitable zones around giant planets. Adv. Space Res. 7 (5), 125–132.

Schubert, G., Moore, W.B., Anderson, J.D., Jacobson, R.A., Lau, E.L., 2000. Io's gravity field and interior structure. AAS 32, 1046.

Selsis, F., Despois, D., Parisot, J.-P., 2002. Signature of life on exoplanets: can Darwin produce false positive detections? Astron. Astrophys. 388, 985–1003.

Sloan Jr., E.D., 1998. Clathrate Hydrates of Natural Gases. Marcel Decker, New York.

Stevenson, D.J., 1999. Possibility of life-sustaining interstellar planets. Nature 400, 32.

Trilling, D.E., Benz, W., Guillot, T., Lunine, J.I., Hubbard, W.B., Burrows, A., 1998. Orbital evolution and migration of giant planets. Astrophys. J. 500, 428–439.

Ward, W.R., 1997. Protoplanet migration by nebula tides. Icarus 126, 261–281.

Wuchterl, G., Guillot, T., Lissauer, J.J., 2000. Giant planet formation. In: Mannings, V., Boss, A.P., Russell, S.S. (Eds.), Protostars and Planets, vol. IV. Univ. of Arizona Press, Tucson, pp. 1081–1109.

Yoo, C.S., Cynn, H., Gygi, F., Galli, G., Iota, V., Nicol, M., Carlson, S., Häusermann, D., Mailhiot, C., 1999. Crystal structure of carbon dioxide at high pressure: "superhard" polymeric carbon dioxide. Phys. Rev. Lett. 83 (26), 5527–5530.

Zahnle, K., Kasting, J.F., Pollack, J.B., 1990. Mass fractionation of noble gases in diffusion limited hydrodynamic escape. Icarus 84, 502–527.

Bibliography

Abe, L., Domiciano de Souza, A., Vakili, F. & Gay, J., 2003, *Phase Knife Coronagraph. II - Laboratory results*, A&A, **400**, 385–392.

Abe, L., Vakili, F. & Boccaletti, A., 2001, *The achromatic phase knife coronagraph*, A&A, **374**, 1161–1168.

Absil, O., 2001, *Nulling Interferometry with IRSI-*Darwin*: Detection and Characterization of Earth-like Exoplanets*, Diploma thesis, University of Liège, http://vela.astro.ulg.ac.be/themes/telins/harigs/.

Absil, O., 2006, *Astrophysical studies of extrasolar planetary systems using infrared interferometric techniques*, Ph.D. thesis, University of Liège.

Absil, O., den Hartog, R., Gondoin, P. et al., 2006, *Performance study of ground-based infrared Bracewell interferometers. Application to the detection of exozodiacal dust disks with GENIE*, A&A, **448**, 787–800.

Agabi, A., Aristidi, E., Azouit, M. et al., 2006, *First Whole Atmosphere Nighttime Seeing Measurements at Dome C, Antarctica*, PASP, **118**, 344–348.

Aime, C., 2005a, *Principle of an Achromatic Prolate Apodized Lyot Coronagraph*, PASP, **117**, 1112–1119.

Aime, C., 2005b, *Radon approach to shaped and apodized apertures for imaging exoplanets*, A&A, **434**, 785–794.

Aime, C., Soummer, R. & Ferrari, A., 2002, *Total coronagraphic extinction of rectangular apertures using linear prolate apodizations*, A&A, **389**, 334–344.

Alibert, Y., Mordasini, C., Benz, W. & Winisdoerffer, C., 2005, *Models of giant planet formation with migration and disc evolution*, A&A, **434**, 343–353.

Allard, F., Hauschildt, P. H., Alexander, D. R. et al., 2001, *The Limiting Effects of Dust in Brown Dwarf Model Atmospheres*, ApJ, **556**, 357–372.

Anderson, R., 1974, *Polarization conserving light bending prisms and optimized Fresnel rhombs*, Appl. Opt., **13**, 1110–1114.

Anderson, R., 1988, *Quarterwaveplate and Fresnel rhomb compared in the 10-Mum CO2 laser emission region*, Appl. Opt., **27**, 2746–2747.

Andrewartha, J., Fox, J. & Wilson, I., 1979, *Resonance anomalies in the lamellar grating*, Opt. Acta, **26**, 68–89.

Angel, J., 1990, *Use of a 16-m Telescope to Detect Earthlike Planets*, in P.-Y. Bely, C. Burrows & G. Illingworth (eds.) *The Next Generation Space Telescope*, pages 81–88 (Space Telescope Science Institute, Baltimore).

Angel, J. R. P., Cheng, A. Y. S. & Woolf, N. J., 1986, *A space telescope for infrared spectroscopy of earth-like planets*, Nature, **322**, 341–343.

Angel, J. R. P. & Woolf, N. J., 1997, *An Imaging Nulling Interferometer to Study Extrasolar Planets*, ApJ, **475**, 373–379.

Arnold, L., Gillet, S., Lardière, O. et al., 2002, *A test for the search for life on extrasolar planets. Looking for the terrestrial vegetation signature in the Earthshine spectrum*, A&A, **392**, 231–237.

Astilean, S., Lalanne, P., Chavel, P. et al., 1998, *High-efficiencysubwavelength diffractive element patternedin a high-refractive-index material for 633nm*, Opt. Lett., **23**, 552–554.

Aumann, H., Beichman, C., Gillett, F. et al., 1984, *Discovery of a shell around Alpha Lyrae*, ApJ, **278**, L23–L27.

Aumann, H. H., 1988, *Spectral class distribution of circumstellar material in main-sequence stars*, AJ, **96**, 1415–1419.

Azzam, R., 2004, *Phase shifts that accompany total internal reflection at a dielectric-dielectric interface*, J. Opt. Soc. Am. A, **21**, 1559–1563.

Azzam, R. & Spinu, C., 2004, *Achromatic angle-insensitive infrared quarter-wave retarder based on total internal reflection at the Si-SiO2 interface*, J. Opt. Soc. Am. A, **21**, 2019–2022.

Baba, N. & Murakami, N., 2003, *A Method to Image Extrasolar Planets with Polarized Light*, PASP, **115**, 1363–1366.

Backman, D. E. & Paresce, F., 1993, *Main-sequence stars with circumstellar solid material - The VEGA phenomenon*, in *Protostars and Planets III*, pages 1253–1304.

Baraffe, I., Chabrier, G., Barman, T. S. et al., 2003, *Evolutionary models for cool brown dwarfs and extrasolar giant planets. The case of HD 209458*, A&A, **402**, 701–712.

Barman, T. S., Hauschildt, P. H. & Allard, F., 2001, *Irradiated Planets*, ApJ, **556**, 885–895.

Baudoz, P., Boccaletti, A., Riaud, P. et al., 2004, *Infrared laboratory tests of the 4-quadrant coronograph MIRI/JWST and VLT-PF*, in F. Combes, D. Barret, T. Contini, F. Meynadier & L. Pagani (eds.) *SF2A-2004: Semaine de l'Astrophysique Francaise*, page 175.

Baudoz, P., Boccaletti, A., Riaud, P. et al., 2006, *Feasibility of the Four-Quadrant Phase Mask in the Mid-Infrared on the James Webb Space Telescope*, PASP, **118**, 765–773.

Baudoz, P., Rabbia, Y. & Gay, J., 2000a, *Achromatic interfero coronagraphy I. Theoretical capabilities for ground-based observations*, A&AS, **141**, 319–329.

Baudoz, P., Rabbia, Y. & Gay, J., 2000b, *Study and test of a hybrid coronagraph*, in *Proc. SPIE Vol. 4007, Adaptive Optical Systems Technology, Peter L. Wizinowich; Ed.*, pages 971–979.

Beaulieu, J.-P., Bennett, D. P., Fouqué, P. et al., 2006, *Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing*, Nature, **439**, 437–440.

Beichman, C. A., Woolf, N. J. & Lindensmith, C. A. (eds.) 1999, *The Terrestrial Planet Finder (TPF): a NASA Origins Program to search for habitable planets* (JPL Publication 99-3, Pasadena).

Bennett, J., 1970, *A Critical Evaluation of Rhomb-Type Quarterwave Retarders*, Appl. Opt., **9**, 2123–2129.

Bernhard, G., 1967, *Structural and functional adaptation in a visual system*, Endeavour, **26**, 79.

Beuzit, J.-L., Mouillet, D., Dohlen, K. & Puget, P., 2005, *The VLT Planet Finder I: Status of the project*, in *SF2A-2005: Semaine de l'Astrophysique Francaise*, page 215.

Biener, G., Niv, A., Kleiner, V. & Hasman, E., 2002, *Formation of helical beams by use of Pancharatnam Berry phase optical elements*, Opt. Lett., **27**, 1875–1877.

Biener, G., Niv, A., Kleiner, V. & Hasman, E., 2003a, *Computer-generated infrared depolarizer using space-variant subwavelength dielectric gratings*, Opt. Lett., **28**, 1400–1402.

Biener, G., Niv, A., Kleiner, V. & Hasman, E., 2003b, *Near-field Fourier transform polarimetry by use of a discrete space-variant subwavelength grating*, J. Opt. Soc. Am. A, **20**, 1940–1948.

Biller, B. A., Kasper, M., Close, L. M. et al., 2006, *Discovery of a Brown Dwarf Very Close to the Sun: A Methane-rich Brown Dwarf Companion to the Low-Mass Star SCR 1845-6357*, ApJ, **641**, L141–L144.

Bird, G. R. & Parrish, M. J., 1960, *The wire grid as a near-infrared polarizer*, J. Opt. Soc. Am., **50**, 886–891.

Blayo, N., Cirelli, R., Klemens, F. & Lee, J., 1995, *Ultraviolet-visible ellipsometry for process control during the etching of submicrometer features*, J. Opt. Soc. Am. A, **12**, 591–599.

Boccaletti, A., Baudoz, P., Baudrand, J. et al., 2005, *Imaging exoplanets with the coronagraph of JWST/MIRI*, Advances in Space Research, **36**, 1099–1106.

Boccaletti, A. & Mouillet, D., 2005, *VLT Planet Finder III.: expected performance*, in *SF2A-2005: Semaine de l'Astrophysique Francaise*, page 223.

Boccaletti, A., Riaud, P., Baudoz, P. et al., 2004, *The Four-Quadrant Phase Mask Coronagraph. IV. First Light at the Very Large Telescope*, PASP, **116**, 1061–1071.

Bokor, N., Shechter, R., Davidson, N. et al., 2001, *Achromatic Phase Retarder by Slanted Illumination of a Dielectric Grating with Period Comparable with the Wavelength*, Appl. Opt., **40**, 2076–2080.

Bomzon, Z., Biener, G., Kleiner, V. & Hasman, E., 2001a, *Spatial Fourier-transform polarimetry using space-variant subwavelength metal-stripe polarizers*, Opt. Lett., **26**, 1711–1713.

Bomzon, Z., Biener, G., Kleiner, V. & Hasman, E., 2002a, *Radially and azimuthally polarized beams generated by space-variant dielectric subwavelength gratings*, Opt. Lett., **27**, 285–287.

Bomzon, Z., Biener, G., Kleiner, V. & Hasman, E., 2002b, *Real-time analysis of partially polarized light with a space-variant subwavelength dielectric grating*, Opt. Lett., **27**, 188–190.

Bomzon, Z., Kleiner, V. & Hasman, E., 2001b, *Computer-generated space-variant polarization elements with subwavelength metal stripes*, Opt. Lett., **26**, 33–35.

Bomzon, Z., Kleiner, V. & Hasman, E., 2001c, *Pancharatnam-Berry phase in space-variant polarization-state manipulations with subwavelength gratings*, Opt. Lett., **26**, 1424–1426.

Bomzon, Z., Niv, A., Biener, G. et al., 2002c, *Nondiffracting periodically space-variant polarization beams with subwavelength gratings*, Applied Physics Letters, **80**, 3685–3687.

Bordé, P., Rouan, D. & Léger, A., 2003, *Exoplanet detection capability of the COROT space mission*, A&A, **405**, 1137–1144.

Bordé, P. J. & Traub, W. A., 2006, *High-Contrast Imaging from Space: Speckle Nulling in a Low-Aberration Regime*, ApJ, **638**, 488–498.

Born, M. & Wolf, E., 1999a, *Principles of Optics* (Cambridge University Press), 7th edn.

Born, M. & Wolf, E., 1999b, *Principles of Optics*, chap. 15.5.2 Form birefringence (Cambridge University Press), 7 edn.

Boss, A. P., 1998, *Evolution of the Solar Nebula. IV. Giant Gaseous Protoplanet Formation*, ApJ, **503**, 923–937.

Boyle, B. J., 2001, *The Evolution of QSOs*, in I. Aretxaga, D. Kunth & R. Mújica (eds.) *Advanced Lectures on the Starburst-AGN*, page 325.

Bracewell, R. N., 1978, *Detecting nonsolar planets by spinning infrared interferometer*, Nature, **274**, 780–781.

Brachet, F., 2005, *Etude et développement d'un déphaseur achromatique pour l'interférométrie en frange noire*, Ph.D. thesis, University of Paris-Sud (XI).

Brauer, R. & Brygdahl, O., 1994, *Design of antireflection gratings with approximate and rigorous methods*, Appl. Opt., **33**, 7875–7882.

Brown, R. J. N., Forbes, D. A., Silva, D. et al., 2003, *Near-infrared imaging of ellipticals: surface brightness profiles and photometry*, MNRAS, **341**, 747–759.

Broyden, C., 1970, *The Convergence of a Class of Double-Rank Minimization Algorithms*, Journal Inst. Math. Applic., **6**, 76–90.

Brundrett, D., Gaylord, T. & Glytsis, E., 1998, *Polarizing Mirror/Absorber for Visible Wavelengths Based on a Silicon Subwavelength Grating: Design and Fabrication*, Appl. Opt., **37**, 2534–2541.

Brundrett, D., Glytsis, E. & Gaylord, T., 1994, *Homogeneous-layer models for high-spatial-frequency dielectric surface-relief gratings: conical diffraction and antireflection designs*, Appl. Opt., **33**, 2695–2706.

Brundrett, D., Glytsis, E. & Gaylord, T., 1996, *Subwavelength transmission grating retarders for use at 10.6 m*, Appl. Opt., **35**, 6195–6202.

Bundy, F. P., 1962, *Melting Point of Graphite at High Pressure: Heat of Fusion*, Science, **137**, 1055–1057.

Burrows, A., Marley, M., Hubbard, W. B. et al., 1997, *A Nongray Theory of Extrasolar Giant Planets and Brown Dwarfs*, ApJ, **491**, 856–875.

Burrows, A., Sudarsky, D. & Hubeny, I., 2004, *Spectra and Diagnostics for the Direct Detection of Wide-Separation Extrasolar Giant Planets*, ApJ, **609**, 407–416.

Campbell, G. & Kostuk, R., 1995, *Effective-medium theory of sinusoidally modulated volume holograms*, J. Opt. Soc. Am. A, **12**, 1113–1117.

Cao, Q., Lalanne, P. & Hugonin, J., 2002, *Stable and efficient Bloch-mode computational method for one-dimensional grating waveguides* , J. Opt. Soc. Am. A, **19**, 335–338.

Cavarroc, C., Boccaletti, A., Baudoz, P. et al., 2006, *Fundamental limitations on Earth-like planet detection with extremely large telescopes*, A&A, **447**, 397–403.

Cavazzoni, C., Chiarotti, G. L., Scandolo, S. et al., 1999, *Superionic and Metallic States of Water and Ammonia at Giant Planet Conditions*, Science, **283**, 44–56.

Chabrier, G., Barman, T., Baraffe, I. et al., 2004, *The Evolution of Irradiated Planets: Application to Transits*, ApJ, **603**, L53–L56.

Chauvin, G., Lagrange, A.-M., Dumas, C. et al., 2005a, *Giant planet companion to 2MASSW J1207334-393254*, A&A, **438**, L25–L28.

Chauvin, G., Lagrange, A.-M., Zuckerman, B. et al., 2005b, *A companion to AB Pic at the planet/brown dwarf boundary*, A&A, **438**, L29–L32.

Chazelas, B., 2003, *Préparation á la mission Darwin : Banc-test de déphaseurs achromatiques*, DEA thesis, University of Paris-Sud (XI),.

Chazelas, B., Brachet, F., Bordé, P. et al., 2006, *Instrumental stability requirements for exoplanet detection with a nulling interferometer: variability noise as a central issue*, Appl. Opt., **45**, 984–992.

Chen, W. R., Chang, S. J., Su, Y. K. et al., 2000, *Reactive Ion Etching of ZnSe, ZnSSe, ZnCdSe and ZnMgSSe by H2/Ar and CH4/H2/Ar*, Jpn. J. Appl. Phys., **39**, 3308–3313.

Chou, S., Krauss, P. & Renstrom, P., 1996, *Nanoimprint lithography*, J. Vac. Sci. Technol. B, **14**, 4129–4133.

Clampin, M., Melnick, G. J., Lyon, R. G. et al., 2004, *Extrasolar Planetary Imaging Coronagraph (EPIC)*, in J. C. Mather (ed.) *Microwave and Terahertz Photonics. Edited by Stohr, Andreas; Jager, Dieter; Iezekiel, Stavros. Proceedings of the SPIE, Volume 5487 (2004).*, pages 1538–1544.

Clapham, P., Downs, M. & King, R., 1969, *Some applications of thin films to polarization devices*, Appl. Opt., **8**, 1965–1974.

Clausen, E. M., Craighead, H. G., Schiavone, L. M. et al., 1988, *Reactive ion etching of epitaxial ZnSe thin films*, Journal of Vacuum Science Technology B: Microelectronics and Nanometer Structures, Volume 6, Issue 6, November 1988, **6**, 1889–1891.

Collins, R. E., 1966, *Foundations for Microwave Engineering*, chap. 5 (Mc-Graw Hill, Inc.).

Creech-Eakman, M. J., Leonard, T. A., Geis, J. & Serabyn, E., 2003, *Comparison of commercially available polarizing gratings for mid-infrared studies*, in *Polarimetry in Astronomy. Edited by Silvano Fineschi . Proceedings of the SPIE, Volume 4843 (2003)*, pages 510–516.

Crepp, J., Ge, J., Debes, J. et al., 2005, *Development of Binary Notch Filter Image Masks for the Coronagraphic Terrestrial Planet Finder*, in *SPIE, Vol. 5487*, pages 1383–1390.

CSL & IAGL, 2005, *Technical proposal in response to RFQ/3-11225/04/NL/PA: Lithographic Manufacturing of Zeroth Order Gratings for Innovative Achromatic IR Phase Shifters*, Tech. rep., between CSL/IAGL and ESA-ESTEC.

CSL & IAGL, 2006, *Lithographic Manufacturing of Zeroth Order Gratings for Innovative Achromatic IR Phase Shifters (contract 19.233): Technical Note 1, Investigation and selection of suitable IR materials*, Tech. rep., between CSL/IAGL and ESA-ESTEC.

Dahan, N., Niv, A., Biener, G. et al., 2005, *Thermal image encryption obtained with a SiO2 space-variant subwavelength grating supporting surface phonon-polaritons*, Opt. Lett., **30**, 3195–3197.

Davidson, N., Friesem, A. & Hasman, E., 1992, *Computer-generated relief gratings as space-variant polarization elements*, Opt. Lett., **17**, 1541–1543.

DeBell, A., Dereniak, E., Harvey, J. et al., 1979, *Cryogenic refractive indices and temperature coefficients of cadmium telluride from 6 Mum to 22 Mum*, Appl. Opt., **18**, 3114–3115.

Debes, J. H., Ge, J., Kuchner, M. J. & Rogosky, M., 2004, *Using Notch-Filter Masks for High-Contrast Imaging of Extrasolar Planets*, ApJ, **608**, 1095–1099.

Deguzman, P. & Nordin, G., 2001, *Stacked Subwavelength Gratings as Circular Polarization Filters*, Appl. Opt., **40**, 5731–5737.

Deng, X., Liu, F., Wang, J. et al., 2005, *Achromatic wave plates for optical pickup units fabricated by use of imprint lithography*, Opt. Lett., **30**, 2614–2616.

Derrick, G., McPhedran, R., Maystre, D. & Nevière, M., 1979, *Crossed Gratings: A theory and its Applications*, Appl. Phys., **18**, 39–52.

Disney, M. J., Boyce, P. J., Blades, J. C. et al., 1995, *Interacting Elliptical Galaxies as Hosts of Intermediate Redshift Quasars*, Nature, **376**, 150–153.

Dominik, C. & Decin, G., 2003, *Age Dependence of the Vega Phenomenon: Theory*, ApJ, **598**, 626–635.

Dore, P., Nucara, A., Cannav, D. et al., 1998, *Infrared Properties of Chemical-Vapor Deposition Polycrystalline Diamond Windows*, Appl. Opt., **37**, 5731–5736.

Dos Santos, J. & Bernardo, L., 1997, *Antireflection structures with use of multilevel subwavelength zero-order gratings*, Appl. Opt., **36**, 8935–8938.

Dubreuil, D., Harvey, E. C., Pigot, C. & Rizvi, N. H., 1998, *Antireflection-structured surfaces for mid-infrared entrance windows*, in *Proc. SPIE Vol. 3354, Infrared Astronomical Instrumentation, Albert M. Fowler; Ed.*, pages 269–280.

Dunlop, J. S., McLure, R. J., Kukula, M. J. et al., 2003, *Quasars, their host galaxies and their central black holes*, MNRAS, **340**, 1095–1135.

Ehrenreich, D., Tinetti, G., Lecavelier Des Etangs, A. et al., 2006, *The transmission spectrum of Earth-size transiting planets*, A&A, **448**, 379–393.

Etzel, S., Rose, A. & Wang, C., 2000, *Dispersion of the temperature dependence of the retardance in SiO 2 and MgF 2*, Appl. Opt., **39**, 5796–5800.

Fajardo-Acosta, S. B., Stencel, R. E., Backman, D. E. & Thakur, N., 1999, *Infrared Space Observatory Photometric Search of Main-Sequence Stars for Vega-Type Systems*, ApJ, **520**, 215–222.

Falomo, R., Kotilainen, J. & Treves, A., 2001, *Near-Infrared Imaging of the Host Galaxies of Three Radio-loud Quasars at z∼1.5*, ApJ, **547**, 124–128.

Falomo, R., Kotilainen, J. K., Pagani, C. et al., 2004, *The Cosmic Evolution of Quasar Host Galaxies*, ApJ, **604**, 495–507.

Falomo, R., Kotilainen, J. K., Scarpa, R. & Treves, A., 2005, *VLT adaptive optics imaging of QSO host galaxies and their close environment at z ∼ 2.5: Results from a pilot program*, A&A, **434**, 469–473.

Ferrarese, L., 2002, *Black Hole Demographics*, in C.-H. Lee & H.-Y. Chang (eds.) *Current high-energy emission around black holes*, pages 3–8.

Filinski, I. & Skettrup, T., 1984, *Achromatic phase retarders constructed from right-angle prisms: design*, Appl. Opt., **23**, 2747–2751.

Flanders, D. C., 1983, *Submicrometer periodicity gratings as artificial anisotropic dielectrics*, Appl. Phys. Lett., **42**, 492–494.

Fletcher, R., 1970, *A New Approach to Variable Metric Algorithms*, Computer Journal, **13**, 317–322.

Fletcher, R., 1980, *Unconstrained optimization*, vol. 1 of *Practical Methods of Optimization* (John Wiley and Sons).

Fletcher, R. & Powell, M., 1963, *A Rapidly Convergent Descent Method for Minimization*, Computer Journal, **6**, 163–168.

Foo, G., Palacios, D. M. & Swartzlander, G. A., 2005, *Optical vortex coronagraph*, Opt. Lett., **30**, 3308–3310.

Ford, V. G., Lisman, P. D., Shaklan, S. B. et al., 2004, *The Terrestrial Planet Finder coronagraph: technology and mission design studies*, in J. C. Mather (ed.) *Microwave and Terahertz Photonics. Edited by Stohr, Andreas; Jager, Dieter; Iezekiel, Stavros. Proceedings of the SPIE, Volume 5487 (2004)*, pages 1274–1283.

Franceschini, A., Hasinger, G., Miyaji, T. & Malquori, D., 1999, *On the relationship between galaxy formation and quasar evolution*, MNRAS, **310**, L5–L9.

Fridlund, C. V. M., 2004a, *Darwin and TPF: technology and prospects*, in W. Traub (ed.) *New Frontiers in Stellar Interferometry*, vol. 5491 of *Proc. SPIE*, pages 227–235.

Fridlund, C. V. M., 2004b, *The Darwin mission*, Advances in Space Research, **34**, 613–617.

Galicher, R., Guyon, O., Otsubo, M. et al., 2005, *Laboratory Demonstration and Numerical Simulations of the Phase-Induced Amplitude Apodization*, PASP, **117**, 411–420.

Gay, J. & Rabbia, Y., 1996, *Principe d'un coronographe interférentiel*, C. R. Acad. Sci. Paris, **322**, 265–271.

Gay, J., Rabbia, Y. & Manghini, C., 1997, *Interfero-Coronagraphy Using Pupil π-ROTATION*, in *ASSL Vol. 215: Infrared Space Interferometry : Astrophysics and the Study of Earth-Like Planets*, page 187.

Gaylord, T., Glytsis, E. & Moharam, M., 1987, *Zero-reflectivity homogeneous layers and high spatial-frequency surface-relief gratings on lossy materials*, Appl. Opt., **26**, 3123–3135.

Giovannini, H. & Amra, C., 1997, *Scattering-reduction effect with overcoated rough surfaces: Theory and experiment*, Appl. Opt., **36**, 5573–5579.

Giovannini, H., Saillard, M. & Sentenac, A., 1998, *Numerical study of scattering from rough inhomogeneous films*, Optical Society of America Journal A, **15**, 1182–1191.

Glytsis, E., 2002, *Two-dimensionally-periodic diffractive optical elements: limitations of scalar analysis*, J. Opt. Soc. Am. A, **19**, 702–715.

Goldfarb, D., 1970, *A Family of Variable Metric Updates Derived by Variational Means*, Mathematics of Computing, **24**, 23–26.

Goldreich, P. & Tremaine, S., 1979, *The excitation of density waves at the Lindblad and corotation resonances by an external potential*, ApJ, **233**, 857–871.

Granet, G. & Guizal, B., 1996, *Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization*, J. Opt. Soc. Am. A, **13**, 1019–1023.

Grann, E. & Moharam, M., 1996, *Comparison between continuous and discrete subwavelength grating structures for antireflection surfaces*, J. Opt. Soc. Am. A, **13**, 988–991.

Grann, E., Moharam, M. & Pommet, D., 1994, *Artificial uniaxial and biaxial dielectrics with use of two- dimensional subwavelength binary gratings*, J. Opt. Soc. Am. A, **11**, 2695–2703.

Grann, E., Moharam, M. & Pommet, D., 1995, *Optimal design for antireflective tapered two-dimensional subwavelength grating structures*, J. Opt. Soc. Am. A, **12**, 333–339.

Gratadour, D., Rouan, D., Boccaletti, A. et al., 2005, *Four quadrant phase mask K-band coronagraphy of NGC 1068 with NAOS-CONICA at VLT*, A&A, **429**, 433–437.

Greaves, J. S., Holland, W. S., Moriarty-Schieven, G. et al., 1998, *A Dust Ring around epsilon Eridani: Analog to the Young Solar System*, ApJ, **506**, L133–L137.

Gu, C. & Yeh, P., 1995, *Form birefringence of layered media and volume gratings*, J. Opt. Soc. Am. B, **12**, 1094–1099.

Gu, C. & Yeh, P., 1996, *Form birefringence dispersion in periodic layered media*, Opt. Lett., **21**, 504–506.

Guo, L. J., 2004, *TOPICAL REVIEW: Recent progress in nanoimprint technology and its applications*, Journal of Physics D Applied Physics, **37**, 123.

Guyon, O., 2003, *Phase-induced amplitude apodization of telescope pupils for extrasolar terrestrial planet imaging*, A&A, **404**, 379–387.

Guyon, O., 2004, *Imaging Faint Sources within a Speckle Halo with Synchronous Interferometric Speckle Subtraction*, ApJ, **615**, 562–572.

Guyon, O., Pluzhnik, E. A., Galicher, R. et al., 2005, *Exoplanet Imaging with a Phase-induced Amplitude Apodization Coronagraph. I. Principle*, ApJ, **622**, 744–758.

Guyon, O., Roddier, C., Graves, J. E. et al., 1999, *The Nulling Stellar Coronagraph: Laboratory Tests and Performance Evaluation*, PASP, **111**, 1321–1330.

Hadaway, J. B., Ahmad, A., Pezzaniti, J. L. et al., 2001, *Real-time total integrated scattering measurements on the Mir spacecraft to evaluate sample degradation in space*, Appl. Opt., **40**, 2755–2768.

Haggans, C., Li, L. & Kostuk, R., 1993, *Effective-medium theory of zeroth-order lamellar gratings in conical mountings*, J. Opt. Soc. Am. A, **10**, 2217–2225.

Halbwachs, J. L., Mayor, M. & Udry, S., 2005, *Statistical properties of exoplanets. IV. The period-eccentricity relations of exoplanets and of binary stars*, A&A, **431**, 1129–1137.

Hale, P. & Day, G., 1988, *Stability of birefringent linear retarders (waveplates)*, Appl. Opt., **27**, 5146–5153.

Han, C. & Kostuk, R., 1996, *Enhanced phase shift in a zeroth-order beam from subwavelength grating structures formed in uniaxial birefringent materials*, J. Opt. Soc. Am. A, **13**, 1728–1736.

Hariharan, P., 1996, *Achromatic and apochromatic halfwave and quarterwave retarders*, Optical Engineering, **35**, 3335–3337.

Hasman, E., Kleiner, V., Biener, G. & Niv, A., 2003, *Polarization dependent focusing lens by use of quantized Pancharatnam-Berry phase diffractive optics*, Applied Physics Letters, **82**, 328–330.

Hawkins, G., 1998, *Spectral Characterisation of Infrared Optical Materials and Filters*, Ph.D. thesis.

Heine, C. & Morf, R., 1995, *Submicrometer gratings for solar energy applications*, Appl. Opt., **34**, 2476–2482.

Herbst, T. M. & Hinz, P. M., 2004, *Interferometry on the Large Binocular Telescope*, in W. Traub (ed.) *New Frontiers in Stellar Interferometry*, vol. 5491 of *Proc. SPIE*, pages 383–390.

Herzig, H. (ed.), 1997, *MICRO-OPTICS*, chap. 4 Direct Writing of Continuous-relief Micro-optics (Taylor & Francis).

Hinz, P. M., Angel, J. R. P., Hoffmann, W. F. et al., 1998, *Imaging circumstellar environments with a nulling interferometer.*, Nature, **395**, 251–253.

Hinz, P. M., Angel, J. R. P., Woolf, N. J. et al., 2000, *BLINC: a testbed for nulling interferometry in the thermal infrared*, in P. Léna & A. Quirrenbach (eds.) *Interferometry in Optical Astronomy*, vol. 4006 of *Proc. SPIE*, pages 349–353.

Hirayama, K., Glytsis, E. N. & Gaylord, T. K., 1997, *Rigorous electromagnetic analysis of diffraction by finite-number-of-periods gratings*, Optical Society of America Journal A, **14**, 907–917.

Holland, W. S., Greaves, J. S., Zuckerman, B. et al., 1998, *Submillimetre images of dusty debris around nearby stars.*, Nature, **392**, 788–790.

Jaffe, W., Meisenheimer, K., Röttgering, H. J. A. et al., 2004, *The central dusty torus in the active nucleus of NGC 1068*, Nature, **429**, 47–49.

Jin Kim, T., Campbell, G. & Kostuk, R., 1995, *Volume holographic phase-retardation elements*, Opt. Lett., **20**, 2030–2032.

Joubert, C., Laude, V., Lee, M. & Plouin, J., 2002, *Volume Index Gratings in the Intermediate and Form-Birefringence Regimes*, Appl. Opt., **41**, 6751–6762.

Kajanto, M., Byckling, E., Fagerholm, J. et al., 1989, *Photolithographic fabrication method of computer-generated holographic interferograms*, Appl. Opt., **28**, 778–784.

Kalas, P. & Jewitt, D., 1995, *Asymmetries in the Beta Pictoris Dust Disk*, AJ, **110**, 794.

Kalas, P., Liu, M. C. & Matthews, B. C., 2004, *Discovery of a Large Dust Disk Around the Nearby Star AU Microscopii*, Science, **303**, 1990–1992.

Kanamori, Y., Hane, K., Sai, H. & Yugami, H., 2001, *100 nm period silicon antireflection structures fabricated using a porous alumina membrane mask*, Applied Physics Letters, **78**, 142–143.

Karlsson, M., Hjort, K. & Nikolajeff, F., 2001, *Transfer of continuous-relief diffractive structures into diamond by use of inductively coupled plasma dry etching*, Opt. Lett., **26**, 1752–1754.

Karlsson, M. & Nikolajeff, F., 2002, *Transfer of Micro-Optical Structures into GaAs by Use of Inductively Coupled Plasma Dry Etching*, Appl. Opt., **41**, 902–908.

Karlsson, M. & Nikolajeff, F., 2003, *Diamond micro-optics: microlenses and antireflection structured surfaces for the infrared spectral region*, Opt. Express, **11**, 502–507.

Kasdin, N. J., Vanderbei, R. J., Littman, M. G. & Spergel, D. N., 2005, *Optimal one-dimensional apodizations and shaped pupils for planet finding coronagraphy*, Appl. Opt., **44**, 1117–1128.

Kasdin, N. J., Vanderbei, R. J., Spergel, D. N. & Littman, M. G., 2003, *Extrasolar Planet Finding via Optimal Apodized-Pupil and Shaped-Pupil Coronagraphs*, ApJ, **582**, 1147–1161.

Kasting, J. F. & Catling, D., 2003, *Evolution of a Habitable Planet*, ARA&A, **41**, 429–463.

Kettunen, V. & Wyrowski, F., 1998, *Reflection-mode phase retardation by dielectric gratings*, Optics Communications, **158**, 41–44.

Kikuta, H., Ohira, Y. & Iwata, K., 1997, *Achromatic quarter-wave plates using the dispersion of form birefringence*, Appl. Opt., **36**, 1566–1572.

Kikuta, H., Ohira, Y., Kubo, H. & Iwata, K., 1998, *Effective medium theory of two-dimensional subwavelength gratings in the non-quasi-static limit*, J. Opt. Soc. Am. A, **15**, 1577–1585.

King, R. J., 1966, *Quarter-wave retardation systems based on the Fresnel rhomb principle* , Journal of Scientific Instruments, **43**, 617–622.

Kleemann, B. & Guther, R., 1993, *Phases of diffracted light for reflection and transmission gratings*, Journal of Modern Optics, **40**, 575 – 589.

Klopfenstein, R. W., 1956, *A transmission line taper of improved design*, in *Proc. IRE 44*.

Koerner, D. W., Sargent, A. I. & Ostroff, N. A., 2001, *Millimeter-Wave Aperture Synthesis Imaging of Vega: Evidence for a Ring Arc at 95 AU*, ApJ, **560**, L181–L184.

Korte, E. H., Jordanov, B., Kolev, D. & Tsankov, D., 1988, *Total reflection prisms as achromatic IR retarders*, Applied Spectroscopy, **42**, 1394–1400.

Kuchner, M. J., 2003, *Volatile-rich Earth-Mass Planets in the Habitable Zone*, ApJ, **596**, L105–L108.

Kuchner, M. J., Crepp, J. & Ge, J., 2005, *Eighth-Order Image Masks for Terrestrial Planet Finding*, ApJ, **628**, 466–473.

Kuchner, M. J. & Seager, S., 2005, *Extrasolar Carbon Planets*, submitted to ApJ.

Kuchner, M. J. & Spergel, D. N., 2003, *Notch-Filter Masks: Practical Image Masks for Planet-finding Coronagraphs*, ApJ, **594**, 617–626.

Kuchner, M. J. & Traub, W. A., 2002, *A Coronagraph with a Band-limited Mask for Finding Terrestrial Planets*, ApJ, **570**, 900–908.

Kuhlbrodt, B., Örndahl, E., Wisotzki, L. & Jahnke, K., 2005, *High-redshift quasar host galaxies with adaptive optics*, A&A, **439**, 497–512.

Kukula, M. J., Dunlop, J. S., McLure, R. J. et al., 2001, *A NICMOS imaging study of high-z quasar host galaxies*, MNRAS, **326**, 1533–1546.

Kurisu, K., Hirai, T., Fuse, K. et al., 2002, *Development of a Diffractive Optical Element for Laser Processing*, SEI Technical Review, **53**, 86–91.

Labadie, L., Broquin, J.-E., Kern, P. et al., 2005, *Results on the development of mid-infrared integrated optics for the Darwin Mission*, in D. R. Coulter (ed.) *Techniques and Instrumentation for Detection of Exoplanets II. Edited by Coulter, Daniel R. Proceedings of the SPIE, Volume 5905, (2005)*, pages 80–90.

Labadie, L., Labeye, P., Kern, P. et al., 2006, *Modal filtering for nulling interferometry. First single-mode conductive waveguides in the mid-infrared*, A&A, **450**, 1265–1266.

Lagarias, J., Reeds, J., Wright, M. H. & Wright, P. E., 1998, *Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions*, SIAM Journal of Optimization, **9**, 112–147.

Lalanne, P., 1997a, *Improved formulation of the coupled-wave method for two-dimensional gratings*, J. Opt. Soc. Am. A, **14**, 1592–1598.

Lalanne, P., 1997b, *Improved formulation of the coupled-wave method for two-dimensional gratings*, J. Opt. Soc. Am. A, **14**, 1592–1598.

Lalanne, P., Astilean, S., Chavel, P. et al., 1998, *Blazed binary subwavelengthgratings with efficiencies larger than thoseof conventional chelette gratings*, Opt. Lett., **23**, 1081–1083.

Lalanne, P. & Hugonin, J., 1998, *High-order effective-medium theory of subwavelength gratings in classical mounting: application to volume holograms*, J. Opt. Soc. Am. A, **15**, 1843–1851.

Lalanne, P. & Lemercier-Lalanne, D., 1997, *Depth dependence of the effective properties of subwavelength gratings*, J. Opt. Soc. Am. A, **14**, 450–458.

Lalanne, P. & Morris, G., 1996, *Highly improved convergence of the coupled-wave method for TM Polarization*, J. Opt. Soc. Am. A, **13**, 779–782.

Lalanne, P., Pichon, P., Chavel, P. et al., 1999, *Interferometric Characterization of Subwavelength Lamellar Gratings*, Appl. Opt., **38**, 4980–4984.

Lalanne, P. & Silberstein, E., 2000, *Fourier-modal methods applied to waveguide computational problems*, Opt. Lett., **25**, 1092–1094.

Lammer, H., Selsis, F., Ribas, I. et al., 2003, *Atmospheric Loss of Exoplanets Resulting from Stellar X-Ray and Extreme-Ultraviolet Heating*, ApJ, **598**, L121–L124.

Laureijs, R. J., Jourdain de Muizon, M., Leech, K. et al., 2002, *A 25 micron search for Vega-like disks around main-sequence stars with ISO*, A&A, **387**, 285–293.

Lay, O., 2004, *Systematic Errors in Nulling Interferometers*, Appl. Opt., **43**, 6100–6123.

Lebouquin, J.-B., Labeye, P., Malbet, F. et al., 2006, *Integrated optics for astronomical interferometry. VI. Coupling the light of the VLTI in K band*, A&A, **450**, 1259–1264.

Lee, M., Lalanne, P., Rodier, J. et al., 2002, *Imaging with blazed-binary diffractive elements*, J. Opt. A: Pure Appl. Opt., **4**, 119–124.

Léger, A., Mariotti, J., Mennesson, B. et al., 1996, *Could We Search for Primitive Life on Extrasolar Planets in the Near Future? – The DARWIN Project*, Icarus, **123**, 249–255.

Léger, A., Ollivier, M., Altwegg, K. & Woolf, N. J., 1999, *Is the presence of H_2O and O_3 in an exoplanet a reliable signature of a biological activity?*, A&A, **341**, 304–311.

Léger, A., Selsis, F., Sotin, C. et al., 2004, *A new family of planets? "Ocean-Planets"*, Icarus, **169**, 499–504.

Legge, M., Bacher, G., Bader, S. et al., 2001, *Selective ultrahigh vacuum dry etching process for ZnSe-based II-VI semiconductors*, Journal of Vacuum Science Technology B: Microelectronics and Nanometer Structures, Volume 19, Issue 3, May 2001, pp.692-694, **19**, 692–694.

Leliwa-Kopystyński, J., Maruyama, M. & Nakajima, T., 2002, *The Water-Ammonia Phase Diagram up to 300 MPa: Application to Icy Satellites*, Icarus, **159**, 518–528.

Lemarquis, F., 2003, *ESA AO/1-3946/02/NL/JA WP2112: Low reflexion coatings study*, Tech. rep., between ESA-ESTEC and Institut Fresnel.

Lemarquis, F., Marchand, G. & Amra, C., 1998, *Design and Manufacture of Low-Absorption ZnS YF 3 Antireflection Coatings in the 3.5 16-micron Spectral Range*, Appl. Opt., **37**, 4239–4244.

Lemarquis, F. & Riaud, P., 2003, *Thin-Film Achromatic Phase Shifters for Nulling Interferometry: Design Approach*, Appl. Opt., **42**, 6919–6928.

Lenaerts, C., 2005, *Etude et réalisation de filtres spectraux: application à la résonance de mode guidé et à la résonance de plasmons de surface*, Ph.D. thesis, Université de Liège.

Lenzen, R., Close, L., Brandner, W. et al., 2004, *A novel simultaneous differential imager for the direct imaging of giant planets*, in A. F. M. Moorwood & M. Iye (eds.) *Ground-based Instrumentation for Astronomy. Edited by Alan F. M. Moorwood and Iye Masanori. Proceedings of the SPIE, Volume 5492, (2004)*, pages 970–977.

Lenzen, R., Close, L., Brandner, W. et al., 2005, *NACO-SDI: A Novel Simultaneous Differential Imager for the Direct Imaging of Giant Extra-Solar Planets*, in W. Brandner & M. E. Kasper (eds.) *Science with Adaptive Optics*, page 46.

Levine, B. M., Aguayo, F., Bifano, T. et al., 2006, *The visible nulling coronagraph: architecture definition and technology development status*, in *High-Power Laser Ablation VI. Edited by Phipps, Claude R.. Proceedings of the SPIE, Volume 6265, pp. 1A (2006).*

Levy, U., Nezhad, M., Kim, H. et al., 2005, *Implementation of a graded-index medium by use of subwavelength structures with graded fill factor*, J. Opt. Soc. Am. A, **22**, 724–733.

Levy, U., Tsai, C., Kim, H. & Fainman, Y., 2004a, *Design, fabrication and characterization of subwavelength computer-generated holograms for spot array generation*, Opt. Express, **12**, 5345–5355.

Levy, U., Tsai, C., Pang, L. & Fainman, Y., 2004b, *Engineering space-variant inhomogeneous media for polarization control*, Opt. Lett., **29**, 1718–1720.

Li, L., 1996a, *Formulation and comparison of two recursive matrix algorithms for modeling layered diffrcation gratings*, J. Opt. Soc. Am. A, **13**, 1024–1035.

Li, L., 1996b, *Use of Fourier series in the analysis of discontinuous periodic structures*, J. Opt. Soc. Am. A, **13**, 1870–1876.

Li, L., 1997, *New formulation of the Fourier modal method for crossed surface-relief gratings*, J. Opt. Soc. Am. A, **14**, 2758–2767.

Li, L. & Haggans, C., 1993, *Convergence of the coupled-wave method for metallic lamellar diffraction gratings*, J. Opt. Soc. Am. A, **10**, 1184–1189.

Lin, D. N. C., Bodenheimer, P. & Richardson, D. C., 1996, *Orbital migration of the planetary companion of 51 Pegasi to its present location.*, Nature, **380**, 606–607.

Lineweaver, C. H. & Grether, D., 2003, *What Fraction of Sun-like Stars Have Planets?*, ApJ, **598**, 1350–1360.

Lissauer, J. J., 1993, *Planet formation*, ARA&A, **31**, 129–174.

Liu, J. & Azzam, R., 1996, *Infrared quarter-wave reflection retarders designed with high-spatial-frequency dielectric surface-relief gratings on a gold substrate at oblique incidence*, Appl. Opt., **35**, 5557–5562.

Lloyd, J. P. & Sivaramakrishnan, A., 2005, *Tip-Tilt Error in Lyot Coronagraphs*, ApJ, **621**, 1153–1158.

Lopez, A. & Craighead, H., 1998, *Wave-platepolarizing beam splitter based on a form-birefringentmultilayer grating*, Opt. Lett., **23**, 1627–1629.

Lopez, B., Schneider, J. & Danchi, W. C., 2005, *Can Life Develop in the Expanded Habitable Zones around Red Giant Stars?*, ApJ, **627**, 974–985.

Lovis, C., Mayor, M., Pepe, F. et al., 2006, *An extrasolar planetary system with three Neptune-mass planets*, Nature, **441**, 305–309.

Lyot, B., 1939, *The study of the solar corona and prominences without eclipses (George Darwin Lecture, 1939)*, MNRAS, **99**, 580–594.

Macintosh, B., Graham, J., Oppenheimer, B. et al., 2006, *MEMS-based extreme adaptive optics for planet detection*, in S. S. Olivier, S. A. Tadigadapa & A. K. Henning (eds.) *Microfluidics, BioMEMS, and Medical Microsystems IV. Edited by Papautsky, Ian; Wang, Wanjun. Proceedings of the SPIE, Volume 6113, (2006)*, pages 48–57.

Malbet, F., Berger, J.-P., Kern, P. Y. et al., 2004, *VITRUV, a second-generation VLTI instrument for aperture synthesis imaging with eight telescopes*, in W. A. Traub (ed.) *New Frontiers in Stellar Interferometry, Proceedings of SPIE Volume 5491. Edited by Wesley A. Traub. Bellingham, WA: The International Society for Optical Engineering, 2004., p.439*, pages 439–446.

Malbet, F., Kern, P., Schanen-Duport, I. et al., 1999, *Integrated optics for astronomical interferometry. I. Concept and astronomical applications*, A&AS, **138**, 135–145.

Mamajek, E. E., Meyer, M. R., Hinz, P. M. et al., 2004, *Constraining the Lifetime of Circumstellar Disks in the Terrestrial Planet Zone: A Mid-Infrared Survey of the 30 Myr old Tucana-Horologium Association*, ApJ, **612**, 496–510.

Mangin, J., 2003, *ESA AO/1-3946/02/NL/JA WP2220 : POTENTIAL AND SELECTED INFRARED MATERIALS*, Tech. rep., between ESA-ESTEC and Université of Bourgogne.

Mannings, V. & Barlow, M. J., 1998, *Candidate Main-Sequence Stars with Debris Disks: A New Sample of Vega-like Sources*, ApJ, **497**, 330–341.

Marciante, J., Farmiga, N., Hirsh, J. et al., 2003, *Optical Measurement of Depth and Duty Cycle for Binary Diffraction Gratings with Subwavelength Features*, Appl. Opt., **42**, 3234–3240.

Marcy, G., Butler, R. P., Fischer, D. et al., 2005, *Observed Properties of Exoplanets: Masses, Orbits, and Metallicities*, Progress of Theoretical Physics Supplement, **158**, 24–42.

Mariner, G. R. & Vedam, K., 1981, *Stress-optic coefficient of ZnSe at 10.6 microns*, Appl. Opt., **20**, 2878–2879.

Marois, C., Doyon, R., Nadeau, D. et al., 2005, *TRIDENT: An Infrared Differential Imaging Camera Optimized for the Detection of Methanated Substellar Companions*, PASP, **117**, 745–756.

Marois, C., Doyon, R., Racine, R. & Nadeau, D., 2000, *Efficient Speckle Noise Attenuation in Faint Companion Imaging*, PASP, **112**, 91–96.

Martel, A. R., Ford, H. C., Tran, H. D. et al., 2003, *Coronagraphic Imaging of 3C 273 with the Advanced Camera for Surveys*, AJ, **125**, 2964–2974.

Martinache, F., Guyon, O., Pluzhnik, E. A. et al., 2006, *Exoplanet Imaging with a Phase-induced Amplitude Apodization Coronograph. II. Performance*, ApJ, **639**, 1129–1137.

Marx, D. & Psaltis, D., 1997, *Polarization quadrature measurement of subwavelength diffracting structures*, Appl. Opt., **36**, 6434–6440.

Matsumoto, T., 2005, *Future Space Infrared Telescope Mission, SPICA*, Journal of Korean Astronomical Society, **38**, 89–91.

Mawet, D., 2002, *Etude d'un coronographe à masque de phase à 4 quadrants au moyen de l'optique diffractive* , Diploma thesis, University of Liège.

Mawet, D., 2004, *Application des réseaux sub-lambdas en interférométrie et en coronographie*, DEA thesis, University of Liège.

Mawet, D., Baudrand, J., Lenaerts, C. et al., 2003, *Birefringent achromatic phase shifters for nulling interferometry and phase coronagraphy*, in *ESA SP-539: Earths: DARWIN/TPF and the Search for Extrasolar Terrestrial Planets*, pages 519–524.

Mawet, D., Lenaerts, C., Riaud, P. et al., 2005a, *Use of subwavelength gratings in TIR incidence as achromatic phase shifters*, Opt. Express, **13**, 8686–8692.

Mawet, D., Riaud, P., Absil, O. & Surdej, J., 2005b, *Annular Groove Phase Mask Coronagraph*, ApJ, **633**, 1191–1200.

Mawet, D., Riaud, P., Baudrand, J. et al., 2006, *The four-quadrant phase-mask coronagraph: white light laboratory results with an achromatic device*, A&A, **448**, 801–808.

Mawet, D., Riaud, P., Surdej, J. & Baudrand, J., 2005c, *Subwavelength surface-relief gratings for stellar coronagraphy*, Appl. Opt., **44**, 7313–7321.

Mayor, M. & Queloz, D., 1995, *A Jupiter-Mass Companion to a Solar-Type Star*, Nature, **378**, 355–359.

Maystre, D., 1984, *Rigorous vector theories of diffraction gratings* (North-Holland, Amsterdam).

McLeod, K. K. & Rieke, G. H., 1994, *Near-infrared imaging of low-redshift quasar host galaxies. 2: High-luminosity quasars*, ApJ, **431**, 137–146.

Mennesson, B., M., O. & C., R., 2002, *Use of single-mode waveguides to correct the optical defects of a nulling interferometer*, J. Opt. Soc. Am. A, **19**, 596–602.

Michelson, A. A., 1920, *On the Application of Interference Methods to Astronomical Measurements*, ApJ, **51**, 257–262.

Mieremet, A. L., Braat, J. J., Bokhove, H. & Ravel, K., 2000, *Achromatic phase shifting using adjustable dispersive elements*, in P. J. Lena & A. Quirrenbach (eds.) *Proc. SPIE Vol. 4006, p. 1035-1041, Interferometry in Optical Astronomy, Pierre J. Lena; Andreas Quirrenbach; Eds.*, pages 1035–1041.

Mirotznik, M., Pustai, D., Prather, D. & Mait, J., 2004, *Design of Two-Dimensional Polarization-Selective Diffractive Optical Elements with Form-Birefringent Microstructures*, Appl. Opt., **43**, 5947–5954.

Moharam, M. & Gaylord, T., 1981, *Rigorous coupled-wave analysis of planar-grating diffraction*, J. Opt. Soc. Am., **71**, 811–818.

Moharam, M. & Gaylord, T., 1982, *Diffraction analysis of dielectric surface-relief gratings*, J. Opt. Soc. Am., **72**, 1385–1392.

Moharam, M. & Gaylord, T., 1983a, *Rigorous couple-wave analysis of grating diffraction-E-mode polarization and losses*, J. Opt. Soc. Am., **73**, 451–455.

Moharam, M. & Gaylord, T., 1983b, *Three dimensional vector coupled-wave analysis of planar-grating diffraction*, J. Opt. Soc. Am., **73**, 1105–1112.

Moharam, M. & Gaylord, T., 1986, *Rigorous coupled-wave analysis of metallic surface-relief gratings*, J. Opt. Soc. Am. A, **3**, 1780–1796.

Moharam, M., Grann, E., Pommet, D. & Gaylord, T., 1995a, *Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings*, J. Opt. Soc. Am. A, **12**, 1068–1076.

Moharam, M., Pommet, D., Grann, E. & Gaylord, T., 1995b, *Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmission matrix approach*, J. Opt. Soc. Am. A, **12**, 1077–1086.

Moharam, M. G., 1988, *Coupled-wave analysis of two-dimensional gratings*, in I. Cindrich (ed.) *Holographic Optics: Design and Applications*, vol. 883 of *Proc. SPIE*, pages 8–11.

Monnier, J. D., Traub, W. A., Schloerb, F. P. et al., 2004, *First Results with the IOTA3 Imaging Interferometer: The Spectroscopic Binaries λ Virginis and WR 140*, ApJ, **602**, L57–L60.

Mooney, F., 1952, *Modification of the Fresnel Rhomb*, J. Opt. Soc. Am., **42**, 181–182.

Moreau, V., 2002, *Intégration de systèmes d'interconnactions optiques reconfigurables*, Ph.D. thesis, Université de Liège.

Mouillet, D., Fusco, T., Lagrange, A.-M. & Beuzit, J.-L., 2003, *"Planet Finder" on the VLT: context, goals and critical specifications for adaptive optics*, in *EAS Publications Series*, pages 193–199.

Mouillet, D., Larwood, J. D., Papaloizou, J. C. B. & Lagrange, A. M., 1997, *A planet on an inclined orbit as an explanation of the warp in the Beta Pictoris disc*, MNRAS, **292**, 896–904.

Murakawa, K., Suto, H., Tamura, M. et al., 2004, *CIAO: Coronagraphic Imager with Adaptive Optics on the Subaru Telescope*, PASJ, **56**, 509–519.

Nagib, N., 1997, *Theory of oblique-incidence phase retarders*, Appl. Opt., **36**, 1547–1552.

Nagib, N., 1998, *Phase Retarders Highly Insensitive to the Input Angle*, Appl. Opt., **37**, 1231–1235.

Nagib, N. & El-Bahrawy, M., 1994, *Phase retarders with variable angles of total internal reflection*, Appl. Opt., **33**, 1218–1222.

Nagib, N. & Khodier, S., 1995, *Optimization of a rhomb-type quarter-wave phase retarder*, Appl. Opt., **34**, 2927–2929.

Naqvi, S., Krukar, R., McNeil, J. et al., 1994, *Etch depth estimation of large-period silicon gratings with multivariate calibration of rigorously simulated diffraction profiles*, J. Opt. Soc. Am. A, **11**, 2485–2493.

Neuhäuser, R., Guenther, E. W., Wuchterl, G. et al., 2005, *Evidence for a co-moving sub-stellar companion of GQ Lup*, A&A, **435**, L13–L16.

Nevière, M., 1980, *The homogeneous problem in Electromagnetic Theory of Gratings* (Springer-Verlag, Berlin).

Nikolajeff, F., Löfving, B., Johansson, J. et al., 2000, *Fabrication and Simulation of Diffractive Optical Elements with Superimposed Antireflection Subwavelength Gratings*, Appl. Opt., **39**, 4842–4846.

Nisenson, P. & Papaliolios, C., 2001, *Detection of Earth-like Planets Using Apodized Telescopes*, ApJ, **548**, L201–L205.

Niv, A., Biener, G., Kleiner, V. & Hasman, E., 2003, *Formation of linearly polarized light with axial symmetry by use of space-variant subwavelength gratings*, Opt. Lett., **28**, 510–512.

Niv, A., Biener, G., Kleiner, V. & Hasman, E., 2004, *Propagation-invariant vectorial Bessel beams obtained by use of quantized Pancharatnam-Berry phase optical elements*, Opt. Lett., **29**, 238–240.

Niv, A., Biener, G., Kleiner, V. & Hasman, E., 2005a, *Rotating vectorial vortices produced by space-variant subwavelength gratings*, Opt. Lett., **30**, 2933–2935.

Niv, A., Biener, G., Kleiner, V. & Hasman, E., 2005b, *Spiral phase elements obtained by use of discrete space-variant subwavelength gratings*, Optics Communications, **251**, 306–314.

Nordin, G. & Deguzman, P., 1999, *Broadband form birefringent quarter-wave plate for the mid-infrared wavelength region*, Opt. Express, **5**, 163–168.

Nordin, G., Meier, J., Deguzman, P. & Jones, M., 1999, *Micropolarizer array for infrared imaging polarimetry*, J. Opt. Soc. Am. A, **16**, 1168–1174.

Nulltimate consortium, 2002, *Technical proposal ESA AO/1-3946/02/NL/JA: APS for Darwin*, Tech. rep., between IAS and ESA-ESTEC.

Ollivier, M., 1999, *Contribution à la recherche d'exoplanètes - Coronographie interférentielle pour la mission DARWIN*, Ph.D. thesis, Université de Paris XI.

Ollivier, M., Le Duigou, J.-M., Mourard, D. et al., 2005, *PEGASE... towards DARWIN*, in F. Casoli, T. Contini, J. M. Hameury & L. Pagani (eds.) *SF2A-2005: Semaine de l'Astrophysique Francaise*, page 197.

Ollivier, M. & Mariotti, J.-M., 1997, *Improvement in the rejection rate of a nulling interferometer by spatial filtering*, Appl. Opt., **36**(22), 5340–5346.

Oron, R., Blit, S., Davidson, N. et al., 2000, *The formation of laser beams with pure azimuthal or radial polarization*, Applied Physics Letters, **77**, 3322–3324.

Owen, T., 1980, *The Search for Early Forms of Life in Other Planetary Systems - Future Possibilities Afforded by Spectroscopic Techniques*, in *ASSL Vol. 83: Strategies for the Search for Life in the Universe*, page 177.

Pagani, C., Falomo, R. & Treves, A., 2003, *Host Galaxies of Low-z Radio-loud Quasars: A Search of Hubble Space Telescope Archives*, ApJ, **596**, 830–839.

Palacios, D. M., 2005, *An optical vortex coronagraph*, in D. R. Coulter (ed.) *Techniques and Instrumentation for Detection of Exoplanets II. Edited by Coulter, Daniel R. Proceedings of the SPIE, Volume 5905, pp. 196-205 (2005).*, pages 196–205.

Pang, L., Nezhad, M., Levy, U. et al., 2005, *Form-birefringence structure fabrication in GaAs by use of SU-8 as a dry-etching mask*, Appl. Opt., **44**, 2377–2381.

Pearton, S. J. & Ren, F., 1993, *Plasma etching of ZnS, ZnSe, CdS, and CdTe in electron cyclotron resonance $CH_4/H_2/Ar$ and H_2/Ar discharges*, Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures, Volume 11, Issue 1, January 1993, pp.15-19, **11**, 15–19.

Peng, S. & Morris, G., 1995, *Efficient implementation of rigorous coupled-wave analysis for surface-relief gratings*, J. Opt. Soc. Am. A, **12**, 1087–1096.

Percival, W. J., Miller, L., McLure, R. J. & Dunlop, J. S., 2001, *The host galaxies of luminous radio-quiet quasars*, MNRAS, **322**, 843–858.

Perryman, M. A. C., 2000, *Extra-solar planets.*, Reports of Progress in Physics, **63**, 1209–1272.

Petit, R., 1980, *Electromagnetic Theory of Gratings* (Springer-Verlag, Berlin).

Pluzhnik, E. A., Guyon, O., Ridgway, S. T. et al., 2006, *Exoplanet Imaging with a Phase-induced Amplitude Apodization Coronagraph. III. Diffraction Effects and Coronagraph Design*, ApJ, **644**, 1246–1257.

Podolak, M., Podolak, J. I. & Marley, M. S., 2000, *Further investigations of random models of Uranus and Neptune*, Planet. Space Sci., **48**, 143–151.

Pollack, J. B., Hubickyj, O., Bodenheimer, P. et al., 1996, *Formation of the Giant Planets by Concurrent Accretion of Solids and Gas*, Icarus, **124**, 62–85.

Pommet, D., Grann, E. & Moharam, M., 1995, *Effects of process errors on the diffraction characteristics of binary dielectric gratings*, Appl. Opt., **34**, 2430–2435.

Popov, E. & Nevière, M., 2000, *Grating theory: new equations in Fourier space leading to fast converging results for TM polarization*, J. Opt. Soc. Am. A, **17**, 1773–1784.

Popov, E., Nevière, M., Gralak, B. & Tayeb, G., 2002, *Staircase approximation validity for arbitrary-shaped gratings*, J. Opt. Soc. Am. A, **19**, 33–42.

Quirrenbach, A., Aime, C., Barillot, M. et al., 2004, *Coronographic Methods for the Detection of Terrestrial Planets*, Tech. rep., Conclusions from a workshop held February 02-06, 2004 at Leiden University, Edited by A. Quirrenbach.

Racine, R., Walker, G. A. H., Nadeau, D. et al., 1999, *Speckle Noise and the Detection of Faint Companions*, PASP, **111**, 587–594.

Raguin, D. & Morris, G., 1993a, *Analysis of antireflection-structured surfaces with continuous one-dimensional surface profiles*, Appl. Opt., **32**, 2582–2598.

Raguin, D. & Morris, G., 1993b, *Antireflection structured surfaces for the infrared spectral region*, Appl. Opt., **32**, 1154–1167.

Rayleigh, 1907, *Note on the remarkable case of diffraction spectra described by Prof. Wood*, Philos. Mag., **14**, 60–65.

Ren, D. & Serabyn, E., 2005, *Symmetric nulling coronagraph based on a rotational shearing interferometer*, Appl. Opt., **44**, 7070–7073.

Riaud, P., Boccaletti, A., Baudrand, J. & Rouan, D., 2003, *The Four-Quadrant Phase Mask Coronagraph. III. Laboratory Performance*, PASP, **115**, 712–719.

Riaud, P., Boccaletti, A., Rouan, D. et al., 2001, *The Four-Quadrant Phase-Mask Coronagraph. II. Simulations*, PASP, **113**, 1145–1154.

Riaud, P., Mawet, D., Absil, O. et al., 2006, *Coronagraphic imaging of three Weak-line T-Tauri stars: evidence of planetary formation around PDS70*, accepted in A&A.

Riblet, H. J., 1957, *General synthesis of quarter-wave impedance transformers*, in *IRE Trans. on Microwave Theory Techniques MTT-5*, pages 36–43.

Richter, I., Sun, P., Xu, F. & Fainman, Y., 1995, *Design considerations of form birefringent microstructures*, Appl. Opt., **34**, 2421–2429.

Rivera, E. J., Lissauer, J. J., Butler, R. P. et al., 2005, *A 7.5 Earth mass Planet Orbiting the Nearby Star, GJ 876*, ApJ, **634**, 625–640.

Rochford, K., Rose, A., Williams, P. et al., 1997, *Design and performance of a stable linear retarder*, Appl. Opt., **36**, 6458–6465.

Rochford, K. & Wang, C., 1997, *Accurate interferometric retardance measurements*, Appl. Opt., **36**, 6473–6479.

Roddier, F. & Roddier, C., 1997, *Stellar Coronograph with Phase Mask*, PASP, **109**, 815–820.

Rouan, D., Riaud, P., Boccaletti, A. et al., 2000, *The Four-Quadrant Phase-Mask Coronagraph. I. Principle*, PASP, **112**, 1479–1486.

Ruilier, C., 1999, *Filtrage Modal et Recombinaison de Grands Télescopes. Contributions à l'Instrument FLUOR.*, Ph.D. thesis, University Denis Diderot - Paris VII.

Rytov, S. M., 1956, *Electromagnetic properties of a finely stratified medium*, JETP Soviet Physics, **2**, 466–475.

Safronov, V. S., 1969, *Evoliutsiia doplanetnogo oblaka.* (1969.).

Santos, N. C., Benz, W. & Mayor, M., 2005, *Extrasolar Planets: Constraints for Planet Formation Models*, Science, **310**, 251–255.

Santos, N. C., Bouchy, F., Mayor, M. et al., 2004, *The HARPS survey for southern extra-solar planets. II. A 14 Earth-masses exoplanet around μ Arae*, A&A, **426**, L19–L23.

Sasselov, D. D. & Lecar, M., 2000, *On the Snow Line in Dusty Protoplanetary Disks*, ApJ, **528**, 995–998.

Schade, D. J., Boyle, B. J. & Letawsky, M., 2000, *Hubble Space Telescope observations of X-ray-selected active galactic nuclei*, MNRAS, **315**, 498–516.

Schmitz, M., Brauer, R. & Bryngdahl, O., 1995, *Phase gratings with subwavelength structures*, J. Opt. Soc. Am. A, **12**, 2458–2462.

Schneider, J., 1994, *On the search for O2 in extrasolar planets*, Ap&SS, **212**, 321–325.

Schneider, J., 2006, *The Extrasolar Planet Encyclopedia*, http://www.obspm.fr/planets.

Selsis, F., Chazelas, B., Bordé, P. et al., 2006, *Could we characterize hot Ocean-Planets with COROT, Kepler and Doppler velocimetry ?*, submitted to Icarus.

Selsis, F., Despois, D. & Parisot, J.-P., 2002, *Signature of life on exoplanets: Can Darwin produce false positive detections?*, A&A, **388**, 985–1003.

Serabyn, E., 2000, *Nulling interferometry: symmetry requirements and experimental results*, in P. J. Lena & A. Quirrenbach (eds.) *Proc. SPIE Vol. 4006, Interferometry in Optical Astronomy, Pierre J. Lena; Andreas Quirrenbach; Eds.*, pages 328–339.

Serabyn, E., Booth, A. J., Colavita, M. M. et al., 2004, *The Keck interferometer nuller: system architecture and laboratory performance*, in W. Traub (ed.) *New Frontiers in Stellar Interferometry*, vol. 5491 of *Proc. SPIE*, pages 806–815.

Serabyn, E. & Colavita, M. M., 2001, *Fully Symmetric Nulling Beam Combiners*, Appl. Opt., **40**, 1668–1671.

Shaklan, S. B. & Green, J. J., 2005, *Low-Order Aberration Sensitivity of Eighth-Order Coronagraph Masks*, ApJ, **628**, 474–477.

Shanbhogue, H., Nagendra, C., Annapurna, M. et al., 1997, *Multilayer antireflection coatings for the visible and near-infrared regions*, Appl. Opt., **36**, 6339–6351.

Shanno, D., 1970, *Conditioning of Quasi-Newton Methods for Function Minimization*, Mathematics of Computing, **24**, 647–656.

Silberstein, E., Lalanne, P., Hugonin, J. & Cao, Q., 2001, *Use of grating theories in integrated optics*, J. Opt. Soc. Am. A, **18**, 2865–2875.

Sinzinger, S. & Jahns, J., 2003, *Microoptics* (Wiley-VCH, Weinheim).

Smith, B. A. & Terrile, R. J., 1984, *A circumstellar disk around Beta Pictoris*, Science, **226**, 1421–1424.

Smith, R., Warren, M., Wendt, J. & Vawter, G., 1996, *Polarization-sensitive subwavelength antireflection surfaces on a semiconductor for 975 nm*, Opt. Lett., **21**, 1201–1203.

Soummer, R., 2005, *Apodized Pupil Lyot Coronagraphs for Arbitrary Telescope Apertures*, ApJ, **618**, L161–L164.

Soummer, R., Aime, C. & Falloon, P. E., 2003a, *Stellar coronagraphy with prolate apodized circular apertures*, A&A, **397**, 1161–1172.

Soummer, R., Dohlen, K. & Aime, C., 2003b, *Achromatic dual-zone phase mask stellar coronagraph*, A&A, **403**, 369–381.

Southwell, W., 1991, *Pyramid-array surface-relief structures producing antireflection index matching on optical surfaces*, J. Opt. Soc. Am. A, **8**, 549–553.

Sparing, L. M., Wang, P. D., Xin, S. H. et al., 1996, *Photoluminescence blueshift induced by reactive ion etching of strained CdZnSe/ZnSe quantum well structures*, Journal of Vacuum Science Technology B: Microelectronics and Nanometer Structures, Volume 14, Issue 6, November 1996, **14**, 3654–3657.

Spiller, E., 1984, *Totally reflecting thin-film phase retarders*, Appl. Opt., **23**, 3544–3549.

Stam, D. M., Hovenier, J. W. & Waters, L. B. F. M., 2004, *Using polarimetry to detect and characterize Jupiter-like extrasolar planets*, A&A, **428**, 663–672.

Sudarsky, D., Burrows, A. & Hubeny, I., 2003, *Theoretical Spectra and Atmospheres of Extrasolar Giant Planets*, ApJ, **588**, 1121–1148.

Sudarsky, D., Burrows, A. & Pinto, P., 2000, *Albedo and Reflection Spectra of Extrasolar Giant Planets*, ApJ, **538**, 885–903.

Swain, M., Vasisht, G., Akeson, R. et al., 2003, *Interferometer Observations of Subparsec-Scale Infrared Emission in the Nucleus of NGC 4151*, ApJ, **596**, L163–L166.

Swain, M. R., Walker, C. K., Traub, W. A. et al., 2004, *The Antarctic planet interferometer*, in W. A. Traub (ed.) *New Frontiers in Stellar Interferometry, Proceedings of SPIE Volume 5491. Edited by Wesley A. Traub. Bellingham, WA: The International Society for Optical Engineering, 2004*, pages 176–185.

Swartzlander, G., 2001, *Peering into darkness with a vortex spatial filter*, Opt. Lett., **26**, 497–499.

Taylor, G. L., Dunlop, J. S., Hughes, D. H. & Robson, E. I., 1996, *A near-IR study of the host galaxies of radio-quiet quasars, radio-loud quasars and radio galaxies*, MNRAS, **283**, 930–968.

Trauger, J. T., Hull, T., Stapelfeldt, K. et al., 2003, *The Eclipse mission: a direct imaging survey of nearby planetary systems*, in J. C. Blades & O. H. W. Siegmund (eds.) *Future EUV/UV and Visible Space Astrophysics Missions and Instrumentation. Edited by J. Chris Blades, Oswald H. W. Siegmund. Proceedings of the SPIE, Volume 4854 (2003)*, pages 116–128.

Traylor Kruschwitz, J. & Pawlewicz, W., 1997, *Optical and durability properties of infrared transmitting thin films*, Appl. Opt., **36**, 2157–2159.

Tsai, C.-H., Levy, U., Pang, L. & Fainman, Y., 2006, *Form-birefringent space-variant inhomogeneous medium element for shaping point-spread functions*, Appl. Opt., **45**, 1777–1784.

Turunen, J., 1996, *Form-birefringence limits of Fourier-expansion methods in grating theory*, J. Opt. Soc. Am. A, **13**, 1013–1018.

Vakili, F., Aristidi, E., Abe, L. & Lopez, B., 2004, *Interferometric Remapped Array Nulling*, A&A, **421**, 147–156.

Valette, C., 2004, *Appendix A of progress meeting 1 of ESA AO/1-3946/02/NL/JA: vibration and temperature effects*, Tech. rep., between ESA-ESTEC and IAS.

Vanderbei, R. J., Kasdin, N. J. & Spergel, D. N., 2004, *Checkerboard-Mask Coronagraphs for High-Contrast Imaging*, ApJ, **615**, 555–561.

Vanderbei, R. J., Spergel, D. N. & Kasdin, N. J., 2003a, *Circularly Symmetric Apodization via Star-shaped Masks*, ApJ, **599**, 686–694.

Vanderbei, R. J., Spergel, D. N. & Kasdin, N. J., 2003b, *Spiderweb Masks for High-Contrast Imaging*, ApJ, **590**, 593–603.

Verstraeten, D., 2002, *Les II-VI photoréfractifs dans la bande 0,6 - 1,5 micron pour l'enregistrement holographique dynamique*, Ph.D. thesis, Université Sciences et Technologies - Bordeaux I.

Wang, J., Deng, X., Varghese, R. et al., 2005a, *High-performance optical retarders based on all-dielectric immersion nanogratings*, Opt. Lett., **30**, 1864–1866.

Wang, J. J., Zhang, W., Deng, X. et al., 2005b, *High-performance nanowire-grid polarizers*, Optics Letters, **30**, 195–197.

Wang, Z., Sun, W., Ruan, S. et al., 1997, *Polarization-preserving totally reflecting prisms with a single medium layer*, Appl. Opt., **36**, 2802–2806.

Ward, W. R., 1997, *Protoplanet Migration by Nebula Tides*, Icarus, **126**, 261–281.

Watson, S., Mills, J., Gaiser, S. & Diner, D., 1991, *Direct imaging of nonsolar planets with infrared telescopes using apodized coronographs*, Appl. Opt., **30**, 3253–3262.

Williams, P., Rose, A. & Wang, C., 1997, *Rotating-polarizer polarimeter for accurate retardance measurement*, Appl. Opt., **36**, 6466–6472.

Wilson, S. & Hutley, M., 1982, *The optical properties of "moth eye" antireflection surfaces*, OPTICA ACTA, **29**, 993–1009.

Wittkowski, M., Kervella, P., Arsenault, R. et al., 2004, *VLTI/VINCI observations of the nucleus of NGC 1068 using the adaptive optics system MACAO*, A&A, **418**, L39–L42.

Wolszczan, A. & Frail, D. A., 1992, *A planetary system around the millisecond pulsar PSR1257 + 12*, Nature, **355**, 145–147.

Wood, R., 1902, *On a remarkable case of uneven distribution of light in a diffraction grating spectrum*, Philos. Mag., **4**, 396–402.

Wu, S. & Glytsis, E., 2002, *Finite-number-of-periods holographic gratings with finite-width incident beams: analysis using the finite-difference frequency-domain method*, J. Opt. Soc. Am. A, **19**, 2018–2029.

Xiao, X., Birrell, J., Gerbi, J. E. et al., 2004, *Low temperature growth of ultrananocrystalline diamond*, Journal of Applied Physics, **96**(4), 2232–2239.

Xu, F., Tyan, R., Sun, P. et al., 1995, *Fabrication, modeling, and characterization of form-birefringent nanostructures*, Opt. Lett., **20**, 2457–2459.

Yang, C. & Yeh, P., 1996, *Form birefringence of volume gratings in photopolymers*, Applied Physics Letters, **69**, 3468–3470.

Yang, C. & Yeh, P., 1997, *Artificial uniaxial and biaxial dielectrics with the use of photoinduced gratings*, Journal of Applied Physics, **81**, 23–29.

Yariv, A. & Yhe, P., 1984, *Optical Waves in Crystals* (Wiley, New-York).

Yi, D., Yan, Y., Liu, H. et al., 2003, *Broadband achromatic phase retarder by subwavelength grating*, Optics Communications, **227**, 49–55.

Yi, D., Yan, Y., Liu, H. et al., 2004, *Broadband polarizing beam splitter based on the form birefringence of a subwavelength grating in the quasi-static domain*, Opt. Lett., **29**, 754–756.

Yih, J.-N., Chu, Y.-M., Mao, Y.-C. et al., 2006, *Optical waveguide biosensors constructed with subwavelength gratings*, Appl. Opt., **45**, 1938–1942.

Yoo, C. S., Cynn, H., Gygi, F. et al., 1999, *Crystal Structure of Carbon Dioxide at High Pressure: "Superhard" Polymeric Carbon Dioxide*, Physical Review Letters, **83**, 5527–5530.

Young, J. B., Graham, H. A. & Peterson, E. W., 1965, *Wire grid infrared polarizer*, Appl. Opt., **4**, 1023–1026.

Young, L., 1961, *Synthesis of Multiple Antireflection Films over a Prescribed Frequency Band*, J. Opt. Soc. Am., **51**, 967–974.

Young, W., 1989, *Roark's Formulas for Stress and Strain*, chap. Dynamic and Temperature Stresses (McGraw-Hill, Inc.), 6 edn.

Yu, W., Konishi, T., Hamamoto, T. et al., 2002, *Polarization-Multiplexed Diffractive Optical Elements Fabricated by Subwavelength Structures*, Appl. Opt., **41**, 96–100.

Yu, W., Mizutani, A., Kikuta, H. & Konishi, T., 2006, *Reduced wavelength-dependent quarter-wave plate fabricated by a multilayered subwavelength structure*, Appl. Opt., **45**, 2601–2606.

Yu, W., Takahara, K., Konishi, T. et al., 2000, *Fabrication of Multilevel Phase Computer-Generated Hologram Elements Based on Effective Medium Theory*, Appl. Opt., **39**, 3531–3536.

Yu, Z., Gao, H. & Chou, S. Y., 2004, *In situ real time process characterization in nanoimprint lithography using time-resolved diffractive scatterometry*, Applied Physics Letters, **85**, 4166–4168.

Zhou, Z. & Drabik, T., 1995, *Optimized binary, phase-only, diffractive optical element with subwavelength features for 1.55 μm*, J. Opt. Soc. Am. A, **12**, 1104–1112.

Résumé

A l'occasion du onzième anniversaire de la découverte de la première planète extrasolaire autour d'une étoile de type solaire, au moment où environ 200 planètes ont été découvertes hors de notre propre système, de passionnantes questions à propos de leur formation, leur évolution et pour certaines d'entre elles, leur aptitude à abriter la vie, sont plus que jamais posées. Ces interrogations sur nos origines ont déclenché l'émergence de nouveaux concepts technologiques et une très forte volonté pour pousser les technologies existantes à leur limite, tout cela pour répondre au fantastique défi observationnel posé. ELTs, interféromètres kilométriques au sol ou spatiaux, instruments de nouvelle génération: l'imagerie directe de systèmes extrasolaires et leur caractérisation est sans conteste l'un des thèmes observationnels les plus exigeants, tout cela à cause de l'énorme contraste et de la minuscule séparation angulaire entre les étoiles et leurs environements.

Cette thèse est dédiée à l'étude d'une classe de micro-composants basés sur la technologie des réseaux sub-lambda. Nous démontrons l'utilité des ces méta-matériaux intégrés et nano-structurés dans le domaine de l'imagerie à très grande dynamique. Les réseaux sub-lambda offrent en effet des solutions nouvelles et originales aux exigeantes contraintes induites par les objectifs scientifiques ambitieux de l'astrophysique à haut contraste. Après avoir montré l'utilité pratique des outils coronagraphiques modernes dans l'observation de systèmes planétaires en formation, nous présentons diverses solutions pour améliorer la capacité de détection de systèmes coronographiques existants, ainsi que de nouvelles totalement intégrées et susceptibles de surclasser les systèmes traditionnels au sein des instruments de nouvelle génération. Ensuite, toujours en profitant de la flexibilité optique des réseaux sub-lambda, nous proposons un nouveau concept de déphaseur achromatique pour l'interférométrie en frange noire, qui devra être construit et testé dans le cadre des activités de R&D censées ouvrir la voie à d'ambitieuses missions d'interféromètres spatiaux dédiés à la détection et la caractérisation de planètes semblables à la Terre.

Mots-clés: systèmes planétaires extrasolaires – imagerie à haut contraste – coronographie à masque de phase – interférométrie en frange noire – déphaseurs achromatiques – réseaux sub-lambda

Abstract

At the eleventh anniversary of the discovery of the first extrasolar planet around a solar-type star, at a time where about 200 planets have been discovered outside our own system, daunting questions about their formation, their evolution, and for a few of them, their ability of sustaining life, are more then ever posed. These interrogations about our origins have triggered the emergence of new technological concepts and a strong will for pushing existing technologies to their limit, with the purpose of tackling the fantastic implied observational challenges. ELTs, kilometric space-borne or ground-based interferometers, next-generation instrumentation: direct imaging of extrasolar planetary systems and their characterization is definitely one of the most demanding observational topic because of the huge contrast together with the small angular separation between stars and their environments.

This dissertation is devoted to the study of a class of micro-components based on the subwavelength grating technology. We demonstrate the utility of these integrated nano-engineered meta-materials in the framework of high dynamic range imaging. Subwavelength gratings indeed provide new and original solutions to the very demanding constraints induced by the ambitious scientific goals of high contrast astrophysics. After showing the practical utility of modern coronagraphic techniques in actual observations of young forming extrasolar planetary systems, we present diverse solutions for improving the existing coronagraph detection capabilities as well as totally new integrated ones that shall outperform the traditional systems within next-generation instruments. Then, still using the optical flexibility of subwavelength gratings, we propose a new concept of achromatic phase shifter for nulling interferometry to be manufactured and tested in the context of the R&D activities paving the way towards ambitious space-borne interferometers dedicated to Earth-like planet detection and characterization.

Keywords: extrasolar planetary systems – high dynamic range imaging – phase-mask coronagraphy – nulling interferometry – achromatic phase shifters – subwavelength gratings