Université de Liège Faculté des Sciences

The International Liquid Mirror Telescope: optical testing and alignment using a Nijboer-Zernike aberration retrieval approach

Thèse de doctorat

présentée pour l'obtention du diplôme de

Docteur en sciences

par

Arnaud Magette

Soutenue publiquement le 22 mars 2010 devant le Jury composé de:

Président :	Prof. Serge HABRAKEN
Superviseur :	Prof. Jean SURDEJ
Examinateurs :	Prof. Ermanno BORRA Prof. Claude JAMAR Dr. Jean MANFROID Dr. Pierre RIAUD Prof. Jean-Pierre SWINGS

Département d'Astrophysique, Géophysique et Océanographie

Abstract

In this thesis, we approach several aspects of the International Liquid Mirror Telescope (ILMT). In particular, we are interested in the optical disturbances that affect the quality of this type of telescope as well as the the means of measuring it.

First of all, the deformations of the mirror surface due to various phenomena are studied in detail.

Then, several aberration measurement methods are studied. We show that a new approach, based on the theory of Nijboer-Zernike (NZ), is very promising. This technique is thus developed and adapted in order to be appropriate for our particular needs (related to the ILMT). Using numerical simulations, we study this method and its limitations. The aberrations are measured very precisely with this method, even when the images are strongly disturbed by noise. As far as the number of measurable aberrations is concerned, it mainly depends on the number of PSF rings that are usable for measurement. It arises that in most of the cases, the low order aberrations (the 36 first Zernike coefficients) can be retrieved with an accuracy better than $\lambda/100$.

Finally, several applications related to the ILMT are considered. We develop an alignment method for the camera and its corrector with respect to the liquid mirror. This one is based on a map system (numerical model) of the aberration variations as a function of the camera misalignment. Simulations show that this method is accurate and reliable as long as at least two orthogonal aberrations can be accurately measured, which is made possible with the NZ method.

We also illustrate how to measure lens aberrations in laboratory and investigate the possibility to validate the corrector lenses in this way. We also show how to use the NZ method to test the quality of parabolic mirrors, without any auxiliary optics (null lenses), with a small variation of the method. This should allow to measure the liquid mirror surface.

Résumé

Dans cette thèse, nous abordons plusieurs aspects du Télescope à Miroir Liquide International (ILMT). En particulier, nous nous intéressons aux perturbations qui affectent la qualité optique de ce type de télescope ainsi qu'aux moyens de les mesurer.

Tout d'abord, les déformations de la surface du miroir dues à différents phénomènes sont étudiées en détail de manière théorique.

Ensuite, plusieurs méthodes de mesure des aberrations optiques sont étudiées. Nous montrons qu'une nouvelle approche, basée sur la théorie de Nijboer-Zernike (NZ), est très prometteuse. Cette technique est donc développée et adaptée de manière à convenir à nos besoins particuliers (liés à l'ILMT). A l'aide de simulations numériques, nous étudions cette méthode ainsi que ses limitations. Les aberrations sont mesurées très précisément grâce à cette méthode, même lorsque les images sont fortement bruitées. Le nombre d'aberrations mesurables, quant à lui, dépend principalement du nombre d'anneaux de la PSF qui sont utilisables pour la mesure. Il ressort que dans la majorité des cas, les aberrations de bas ordres (les 36 premiers coefficients de Zernike) peuvent être retrouvées avec une précision meilleure que $\lambda/100$.

Enfin, plusieurs applications en relation avec l'ILMT sont envisagées. Nous développons une méthode d'alignement pour la caméra et son correcteur par rapport au miroir liquide. Celle-ci est basée sur un système de cartographie (modèle numérique) de la variation des aberrations en fonction du désalignement de la caméra. Les simulations montrent que cette méthode est précise et fiable pour autant qu'au moins deux aberrations orthogonales puissent être mesurées avec précision, ce qui est réalisable grâce à la méthode NZ.

Nous illustrons également le moyen de mesurer les aberrations dans les lentilles et nous étudions la possibilité de valider les lentilles du correcteur de cette manière. Nous montrons aussi comment utiliser la méthode NZ pour tester la qualité de miroirs paraboliques, sans optiques auxiliaires (lentilles d'Offner), au moyen d'une légère adaptation. Ceci devrait permettre de mesurer la surface du miroir liquide.

Université de Liège **Faculté des Sciences**

The International Liquid Mirror Telescope: optical testing and alignment using a Nijboer-Zernike aberration retrieval approach

Thèse de doctorat

présenté pour l'obtention du diplôme de

Docteur en sciences

par

Arnaud Magette

Soutenue publiquement le 22 mars 2010 devant le Jury composé de:

Président :	Prof. Serge HABRAKEN
Directeur de thèse :	Prof. Jean SURDEJ
Examinateurs :	Prof. Ermanno BORRA Prof. Claude JAMAR Dr. Jean MANFROID Dr. Pierre RIAUD Prof. Jean-Pierre SWINGS

Département d'Astrophysique, Géophysique et Océanographie

Mis en page avec la classe thloria.

Acknowledgments

I would like to thank a few persons for their precious help without which this work would not have been possible.

First and foremost, I would like to express my deepest gratefulness to my supervisor, Prof. Jean Surdej for his guidance and his attention. He had a very important role in my scientific formation with his advice, but also with his exemplary motivation to lead the project and the confidence he gave me. I am also grateful to him for his careful reading of this document and the help he gave me to improve it.

I wish to express my special thanks to Dr. Pierre Riaud for all the time he spent with me, for his advice and the many discussions we had about my work. His attentive reading of the present document and the invaluable comments he made about it were very appreciated.

Let me also acknowledge Prof. Ermano Borra, Prof. Serge Habraken, Prof. Claude Jamar, Dr. Jean Manfroid, Dr. Pierre Riaud, Prof. Jean Surdej and Prof. Jean Pierre Swings, the members of my thesis committee, for having accepted to read and evaluate this PhD thesis.

I am also grateful to the people who helped me to achieve the practical parts of this work. For their assistance in repairing and operating the 2m Liquid Mirror Telescope of the "Centre Spatial de Liège", I whish to thank Joel Poels, Przemyslaw Bartczack, Vincent Vandeweert, Davide Ricci, Nathalie Ninane, Thiery Jaquemart and Francis Monfort. François Finet, Charles Hanot and Pierre Riaud are acknowledged for their collaboration in the optical experiments we made in laboratory.

Prof Jean-Pierre Swings, Dr. Natacha Linder and Denis Defrère are particularly thanked for their careful reading of the present document and their help to correct it, to complete it and to make it clearer.

For his general help on the International Liquid Mirror Telescope project and the discussion we had about many subjects, I wish to thank Prof. Paul Hickson from the University of British Columbia.

These four years of PhD thesis would not have been such a pleasant work without the friends that I made at the institue of astrophysics and geophysics of the University of Liège. Thanks to Den, Nat, Méla, Antho, Bertrand, Charles, Davide, Emilie, François, Gi, Laurent, Oliver, for the funny "whist" and other card games we play during the lunch time and for the general good atmosphere at work they create. I am particularly grateful to Denis, my office-mate, for so many things that it is impossible to remind them all here, but some of the most important being the many fruitful discussions we had, the so-called "small-fluid" and "super-banco" break times, the exchange of TV shows, the "Civ4" evenings, ... I hope we will keep in touch after our professional ways will separate.

Finally, I would like to thank my wife Virginie, who supported me during this work with her love and kindness. Thanks also go to my mother, Régine, for the time she spent listening to the problems I had with this work and for her wise advice. I am very grateful to both of them for their help in correcting this manuscript.

Thanks very much to all of you for what you have done for me.

I cannot finish these acknowledgments without having a deep thought for Oliver Garcet. I hope you can keep doing astronomy from wherever you are.

This research was supported by fellowships from the University of Liège ("Bourse de doctorat") during the first year, from the Belgian National Science Foundation ("FRIA") for the next three years and from the Communauté française de Belgique - Actions de recherche concertées - Académie universitaire Wallonie Europe during the last few months.

Contents

Acknowledgments **i**

Acronyms **1**

Introduction **3**

I Liquid Mirror Telescopes **7**

1 The International Liquid Mirror Telescope **9**

1.1 Introduction to the ILMT project 9

1.2 The liquid mirror 14

1.2.1 Principle 14

1.2.2 Reflecting liquids 16

1.3 The CCD Camera of the ILMT 20

1.3.1 Time Delay Integration 23

1.3.2 Filters 24

1.4 A specific corrector 25

1.5 Liquid Mirror Telescopes scientific roadmap 30

1.5.1 Specificities of observations with LMTs 30

1.5.2 The first liquid mirror telescopes 31

1.5.3 The CSL 2m liquid mirror telescope 32

1.5.4 NASA Orbital Debris Observatory - (NODO) 35

1.5.5 The Large Zenithal Telescope - (LZT) 37

1.5.6 The International Liquid Mirror Telescope (ILMT) 40

2 The Charge Coupled Device camera **43**

2.1 CCD chip selection 43

2.1.1 Overview of the possible chips 44

2.1.2 Quantum efficiency 45

2.1.3 Noise level 46

2.1.4 Full well capacity . . . 48
2.1.5 Defect specifications . . . 49
2.1.6 Conclusions: choice of the chip . . . 52
2.2 Technical specifications of the ILMT CCD camera . . . 53
2.2.1 Overview of the 1100 series general characteristics . . . 53
2.2.2 Linearity of the imaging system . . . 53
2.2.3 Chip cooling system . . . 54
2.2.4 Vacuum . . . 54
2.2.5 Readout mode . . . 55
2.2.6 Time tagging . . . 55
2.2.7 Electronic interface . . . 59
2.2.8 Camera sizes . . . 59
2.2.9 Environmental conditions . . . 60
2.2.10 Mechanical interface and constraints . . . 60
2.2.11 Package contents . . . 63
2.3 The 2m-LMT CCD camera . . . 64
2.3.1 Presentation of the camera . . . 64
2.3.2 Imaging capabilities . . . 67
2.3.3 Time Delay Integration (TDI) behavior testing . . . 71
2.3.4 Discussion . . . 79

II Optical considerations 81

3 Disturbances of the liquid mirror 83
3.1 Zero order equilibrium . . . 84
3.2 Gravitational field gradient and Earth curvature . . . 84
3.3 Tilt of the rotation axis of the liquid mirror . . . 89
3.4 Earth rotation: the Coriolis effect . . . 92
3.5 Wind induced spiral waves . . . 101
3.6 Vibration induced concentric waves . . . 105
3.6.1 Modeling . . . 106
3.7 Waves damping . . . 112
3.8 Testing the liquid mirror . . . 114

4 Point Spread Function - (PSF) 117
4.1 Introduction . . . 117
4.1.1 Diffraction . . . 117
4.1.2 Convolution . . . 119

4.1.3 Atmospheric turbulence . . . 120
4.1.4 PSF fitting . . . 122
4.1.5 Speckle interferometry . . . 124
4.1.6 Adaptive optics . . . 126
4.2 Optical aberrations . . . 130
4.2.1 Zernike polynomials . . . 131
4.2.2 The pupil function . . . 133
4.3 Calculation of the PSF from the pupil function . . . 133
4.3.1 The Fourier approach . . . 133
4.3.2 The Nijboer-Zernike approach . . . 135

5 Aberration retrieval **143**
5.1 Optical testing and classical phase retrieval method . . . 143
5.1.1 The shape of the tested surface . . . 144
5.1.2 The Foucault test . . . 144
5.1.3 The Ronchi test . . . 151
5.1.4 Computer Generated Holograms - (CGHs) . . . 153
5.1.5 The Hartmann test . . . 154
5.1.6 The Roddier test . . . 156
5.1.7 Summary . . . 158
5.2 The Nijboer-Zernike (NZ) approach . . . 159
5.2.1 Intuitive approach using classical Zernike expansion (α_n^m) . . . 160
5.2.2 NZ retrieval theory based on the general Zernike coefficients (β_n^m) . . . 162
5.2.3 Predictor-Corrector approach . . . 165

III Implementation **167**

6 Implementation of the Nijboer-Zernike theory **169**
6.1 Computation of the PSF based on the aberration coefficients . . . 169
6.1.1 General Zernike coefficients (β) . . . 169
6.1.2 Numerical calculation of the $V_n^m(r, f)$ functions . . . 171
6.1.3 Comparison between the FFT PSF and NZ PSF . . . 173
6.2 Retrieval of the aberration coefficients from the PSF analysis . . . 179
6.2.1 First step: the input images . . . 179
6.2.2 The inner product . . . 183
6.2.3 Resolution of the systems for $m \neq 0$ . . . 184
6.2.4 Resolution of the systems for $m = 0$ . . . 185
6.3 Study of the Nijboer-Zernike retrieval method . . . 186

6.3.1 Predictor-corrector convergence . . . 186
6.3.2 Imperfect cases of retrieval . . . 193
6.3.3 Impact of the image sampling on the retrieval process . . . 201
6.3.4 Conclusions . . . 203

7 Application 1: Alignment of the ILMT 205
7.1 The 2m Liquid Mirror Telescope alignment method . . . 206
7.1.1 Optical design . . . 206
7.1.2 Alignment method . . . 208
7.1.3 Simulations and results . . . 213
7.1.4 Conclusion . . . 215
7.2 The ILMT alignment: an optimized approach using the phase-retrieval technique 216
7.2.1 Principle . . . 216
7.2.2 Optical design . . . 217
7.2.3 Aberration maps . . . 218
7.2.4 Simulations and results . . . 219
7.2.5 Conclusion . . . 221

8 Application 2: Aberration measurements 223
8.1 NACO aberration measurement . . . 224
8.1.1 Description of the on-sky experiment . . . 224
8.1.2 Aberration retrieval . . . 225
8.1.3 Results . . . 229
8.1.4 Conclusion . . . 230
8.2 Lens testing . . . 233
8.2.1 Equipment . . . 233
8.2.2 Zemax models . . . 235
8.2.3 The images . . . 237
8.2.4 Aberration retrieval . . . 238
8.2.5 Results . . . 240
8.2.6 First additional application: Test of the ILMT corrector . . . 243
8.2.7 Second additional application: Test of parabolic mirrors . . . 245
8.2.8 Conclusions . . . 248

Conclusions 249

Bibliography 253

IV Appendices **259**

A Focus tolerancing **261**

B Computation of the intensity **265**

Calculation of the intensity (5.22) 265

Computation of Ψ^m_{meas} 5.27 and 5.28 268

C Numerical expression of the V^m_n function **273**

D Generalization of the NZ equations when the complex β^0_N coefficients is dominant **279**

Acronyms

ADU Analog to Digital Unit

AO adaptive optics

CCD Charge Coupled Device

CONICA COuder Near-Infrared CAmera

CSL Centre Spatial de Liège (Liège Space Center)

EASO Extragalactic Astrophysics and Space Observations (AEOS - Astrophysique Extragalactique et Observations Spatiales)

FFT Fast Fourier Transform

FPGA Field Programable Gate Array

FWHM Full Width at Half Maximum

ILMT International Liquid Mirror Telescope

JSC Johnson Space Center

GPI Gemini Planet Imager

GPS Global Positioning system

LED Light Emitting Diode

LIDAR LIght Detection and RAnging

LGS Laser Guide Star

LM Liquid Mirror

LMT Liquid Mirror Telescope

LZT Large Zenithal Telescope

MPP Multi-Pinned Phase

MTF Modulation Transfer Function

NAOS Nasmyth Adaptive Optics System

NODO NASA Orbital Debris Observatory

NZ Nijboer-Zernike

PSD Power Spectral Density

PSF Point Spread Function

QE Quantum Efficiency

RMS root mean square

SDSS Sloan Digital Sky Survey

SI Spectral Instruments

SINFONI Spectrograph for INtegral Field Observations in the Near Infrared

SPHERE Spectro-Polarimetric High-contrast Exoplanet REsearch

TDI Time Delay Integration

UT Unit Telescope

VLT Very Large Telescope

ZPL Zemax Programming Language

Introduction

The basic concept of liquid mirrors is well known since Newton discovered that the free surface of a rotating liquid is a paraboloid of revolution. It has then been developed by Ernesto Capocci (1856) and the first prototype of a liquid mirror was built in 1872 by Henry Skey. He encountered some technical difficulties related to the angular speed stability of the mirror. In 1909, Robert Wood built a 51cm liquid mirror but vibrations transmitted to the liquid by the ball-bearing made it unusable. Moreover, at that time, the impossibility to tilt the mirror was a very important limitation that prevented liquid mirror telescopes to be useful for astronomy.

It is only since the eighties that the technology has sufficiently evolved to circumvent those technical issues. Particularly, the use of CCD sensors allows to electronically track stars thanks to the Time Delay Integration (TDI) acquisition mode. In 1982, Ermanno Borra has constructed the first working 1.5m liquid mirror giving diffraction limited images. This was the beginning of the modern liquid mirror technology era. In the following decades, Borra and Hickson built several 3m class liquid mirror telescopes. The largest one, the 6m Large Zenithal Telescope (LZT), was achieved in 2005 by Paul Hickson, in Vancouver, Canada.

The main advantage of liquid mirror telescopes is their relatively low cost compared to conventional glass mirror telescopes. However the limitation to the zenith observation is a serious drawback for some astrophysical investigations. Nevertheless, it is perfectly appropriate for survey applications. For example, cosmological phenomena are supposed to be isotropically distributed in the sky. Studying them at the zenith is thus not a problem.

The International Liquid Mirror Telescope (ILMT) project consists of a 4m spinning pool of mercury used as the primary mirror associated with a 4096×4096 pixels CCD camera installed at its focus and a five lens corrector designed to compensate for the field aberrations and the TDI distortions. This telescope will survey a 22 arcmin wide strip of sky down to the deep magnitude of 22.5 in the near infrared every night. This observational strategy will allow the detection and study of many new variable objects. The ILMT scientific drivers are mainly the study of gravitational lenses and supernovae.

In this thesis, we address several aspects of the ILMT. Particularly, we study the disturbances that affect the optical quality of this kind of telescope as well as the means of measuring them. The deformations of the mirror due to the rotation or to the curvature of the Earth are considered and entirely characterized because they imperatively have to be taken into account, in particular during the precise calculation of the angular speed of the mirror. We also model the wavelets propagating through the mercury surface to study their effect on the image quality of the liquid mirror.

When the telescope will be finished (hopefully in 2011), it will be necessary to measure the liquid mirror surface in order to precisely determine the disturbances it undergoes. For this

purpose, several aberration measurement methods are studied (Foucault, Rodier, ...). We show that a new approach based on the Nijboer-Zernike (NZ) theory is very promising.

This aberration measurement technique is thus developed and adapted to be appropriate for our particular needs. Using numerical simulations, we have completely characterized this method and its limitations. The accuracy obtained on the aberrations retrieved with this method is very high, even in case of strongly disturbed images. The number of aberrations that can be accurately measured mainly depends on the number of PFS rings that present a sufficient signal to noise ratio. In most cases, the low order aberrations (the 36 first Zernike coefficients) can be retrieved with an accuracy better than $\lambda/100$. This accurate aberration retrieval method should be very convenient to use with liquid mirror telescopes because of its simplicity of implementation and its capability to work with a large number of aberrations.

Several applications of this new measurement method are considered. One of them is obviously the characterization of the liquid mirror. We also apply it for the alignment of the upper-end unit with respect to the primary mirror and for the validation of the corrector lenses.

As far as the alignment is concerned, we develop a method based on maps that describe the variation of the aberrations of the system as a function of the misalignments of the camera and its corrector with respect to the liquid mirror. These maps are obtained from a theoretical model of the telescope. Simulations show that this method is accurate (position computed within 0.5mm and 1 arcmin accuracy) and reliable (95 % of the cases are treated correctly) as long as at least two orthogonal aberrations (ex: tip and tilt) can be measured accurately. The use of the NZ method should allow to achieve this accuracy.

Regarding the classical aberration measurement, we first present the application of the NZ retrieval method on VLT-NACO images that will then be compared to previous calibration results. We will then illustrate the classical use of this technique to measure the aberrations present in lenses. This process could be used to validate the lenses of the corrector. We then show how a small modification of the NZ method can be used to test the quality of aspheric optical elements (i.e. parabolic mirrors) without using any auxiliary optics (null lenses).

The present thesis is divided in three parts. The first one (chapters 1 and 2) contains a general description of Liquid Mirror Telescopes (LMT), and of the ILMT project. The second part (chapters 3 to 5) is related to optical considerations. We study the disturbances related to liquid mirrors that would decrease the image quality of the telescope and we explore the optics theory to find a way of measuring this quality. The third part (chapters 6 to 8) is dedicated to the implementation of the method developed in the previous one. We first investigate the method itself in order to discover its capabilities and limitations. We then present two types of applications, the alignment of the camera and the testing of optical elements.

A general presentation of the ILMT is given in chapter 1. The related basic concepts are exposed in this chapter. The principle of the liquid mirror is approached and the characteristics of the reflecting liquid are reviewed. The CCD sensor principle is also briefly reminded and the particular Time Delay Integration (TDI) acquisition mode is described. Other liquid mirror telescopes are also presented in this chapter. The end of this chapter briefly describes other LMTs and their scientific contributions. The ILMT science drivers are also briefly reviewed.

The upper-end instrument is completely described in chapter 2. The detailed specifications constituting the call for tender we made for the CCD camera as well as the comparison of several sensors are presented in this chapter. Some tests we had performed in order to characterize a smaller camera (belonging to CSL) are also described at the end of the chapter. This camera

was expected to be useful for the ILMT testing purpose.

The difficulties inherent to the liquid mirrors are introduced in chapter 3. They are related to the Earth curvature, the Coriolis effect, the tilt of the mirror axis and wavelets over the liquid surface. We detail the calculation of the related aberrations and give the results related to the ILMT.

Chapter 4 presents the concept and properties of the Point Spread Function (PSF). We address the way to measure it and the causes that can disturb it. We also present the concept of aberration and the particular basis of the Zernike polynomials that will be used to describe those aberrations. This chapter ends with the development of a convenient way to compute the PSF from the aberrations it contains, based on the Nijbeor-Zernike theory.

Several optical aberration measurement methods (Foucault, Ronchi, Roddier, Hartmann,...) are investigated in chapter 5. The reverse theory of Nijboer-Zernike is then presented. It can be used to accurately compute aberrations contained in PSFs. We then develop an extension of this theory to asymmetrical systems containing high order aberrations.

Chapter 6 discusses the implementation of the NZ retrieval theory and its properties. We used numerical simulations to study many behaviours of the algorithm such as its convergence, the effects of completeness and noise, and the sampling of the images. The capabilities and limitations of this method are investigated.

In chapter 7, we develop a first application of the NZ aberration measurement method, the alignment of the upper-end unit (camera + corrector) with respect to the liquid mirror.

Chapter 8 presents more classical applications such as the determination of the aberrations contained in an optical system (lenses). We test our NZ method on images taken with the VLT-NACO instrument and compare the results with previous calibrations. We also develop a convenient way of testing parabolic mirrors from their center of curvature without the need of an auxiliary optics.

Part I

Liquid Mirror Telescopes

Chapter 1

The International Liquid Mirror Telescope

1.1 Introduction to the ILMT project

As indicated by the name, the International Liquid Mirror Telescope (ILMT) is an instrument that uses a reflecting liquid as the primary mirror. We describe here the specificities of this kind of telescopes. Among these, a primary mirror made of rotating liquid mercury, whose principle is introduced in section 1.2.

The International Liquid Mirror Telescope project began a bit more than a decade ago with the participation of many countries that were interested in its scientific potential. Scientists met in Marseille (France) in 1997 during a workshop where the scientific applications of such a telescope were discussed.

During the past recent years (1997-2009), the complete funding for the construction of the ILMT has been gathered by the Extragalactic Astrophysics and Space Observations (EASO) group of the University of Liège. The ILMT project consists in designing, building and operating a Liquid Mirror Telescope (LMT) using a mercury mirror with a diameter of four meters in order to perform a deep survey of a strip of sky, the goal being to study all the variable objects that are in that strip. The scientific objectives of the ILMT as well as the science studies already achieved with other LMTs will be presented at the end of this chapter. The project did really start with the construction at the Liège Space Center of a 2m LMT prototype completed in 2001.

In addition to the liquid mirror, the ILMT will be equipped with a 4096×4096 pixel CCD camera located eight meters above the liquid mirror at its prime focus. It provides a field of view of about 0.5 square degree. This camera is introduced in section 1.3 and extensively described in chapter 2.

An optical corrector is placed before the camera. It aims at compensating the off-axis coma of the parabolic mirror in order to get a "large" well-corrected field of view as well as getting rid of the LMT specific aberrations. This will be presented in section 1.4.

Unlike the classical telescopes that are movable, the liquid mirror must be kept horizontal, looking toward the zenith. The upper-end elements (camera + corrector), that are located at the focus of the primary mirror, are supported by a simple static structure shown in fig. 1.1 (background). This stell structure is almost 10m high.

The whole telescope is housed in a very simple enclosing with a sliding-opening roof. One of the preliminary drawings of this building is presented in fig. 1.2.

Figure 1.1: The ILMT structure that will support the CCD camera and the optical corrector. It is far more simple than the mobile mount of a classical telescope. As the liquid mirror cannot be tilted, the structure does not need to be movable. The 4m diameter mirror container is shown in front of the ILMT structure. It is a carbon-fiber structure that will be covered with polyurethane by spin casting. It will contain up to 500kg of mercury during the startup of the mirror. This picture has been taken during the construction of the ILMT at AMOS.

The liquid mirror itself is composed of several elements. The most used reflecting liquid is the elemental mercury (Hg). Its properties as well as the reasons for this choice are exposed in section 1.2.2. The mercury is poured in a dish (fig. 1.1 at the foreground) that is four meters wide. It is composed of a foam core surrounded by carbon-fiber clothes that ensure its rigidity. Fig. 1.3 (left) shows the gluing of extra carbon-layers that aim at increasing the stiffness of the central part of the bowl. It is very important to have a stiff structure that will not get deformed under the weight of mercury. Such deformations could lead to important flows of mercury that would cause the dish to collapse.

The carbon bowl, that has a rough spherical shape, will be covered with polyurethane by spin casting. This layer will have a shape that is very near to the final paraboloidal shape of the mirror, allowing to spread a very thin (1mm) layer of mercury. It is important to minimize the liquid thickness in order to reduce the mass of the mirror and to minimize the perturbations of its surface.

The container lays on an air bearing (fig. 1.3 right) that has to support the weight of the table and the mercury. Up to 500kg of mercury will be used during the start up of the mirror.

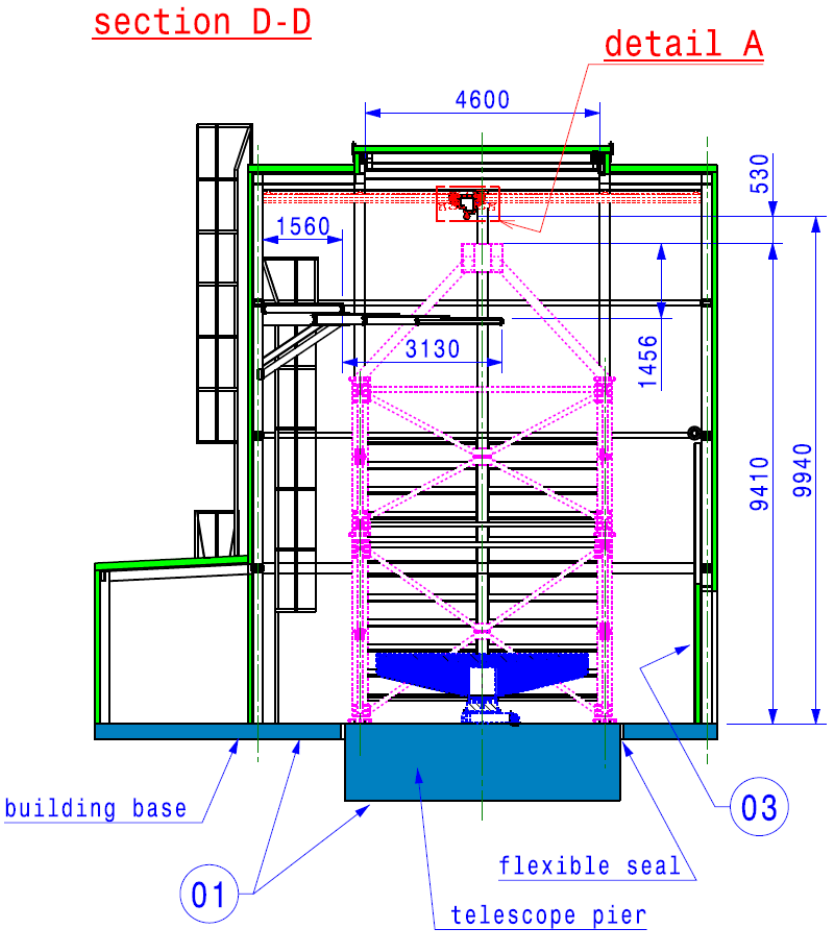


Figure 1.2: Drawing of the enclosure protecting the telescope structure and the mirror dish. Sizes are given in millimeters. The housing simply consists in a building with an opening roof. This drawing was provided by AMOS.

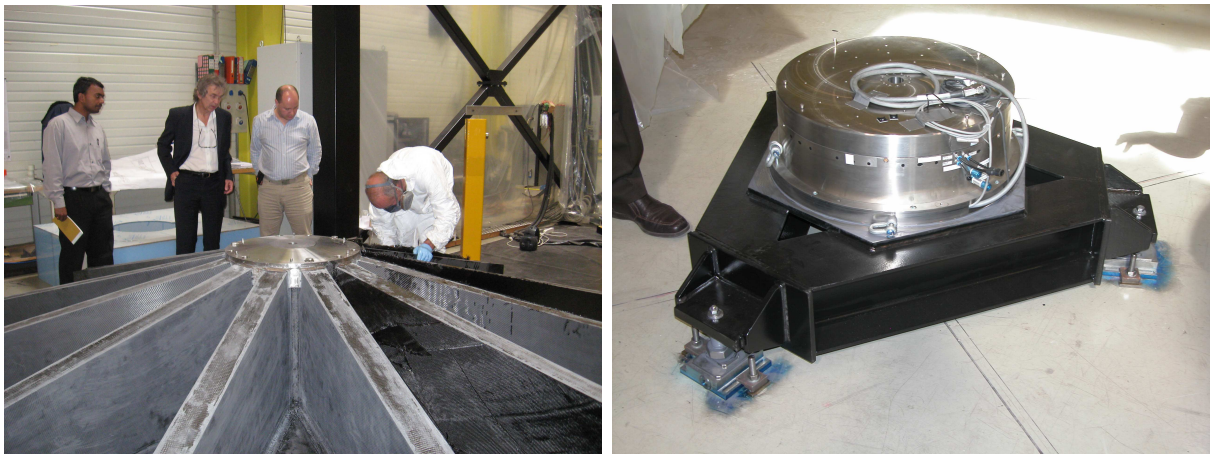


Figure 1.3: Left: The mirror container during the gluing of extra carbon-fiber layers. This process aims at increasing the stiffness of the dish, that must be capable of supporting an unbalanced weight without deformations that would cause instabilities. The people present in this picture are from left to right: B. Kumar, PhD student; Pr. J. Surdej, project scientist of the ILMT; S. Denis, ILMT project manager at AMOS; E. Plumacker, technician responsible for the gluing of the carbon fibers. Right: Picture of the air-bearing that supports the mirror dish. The electrical micro-stepping motor is inside the bearing.

Figure 1.4: Map of India. The future location of the ILMT observatory is indicate by the sign (in the North of the map). Image from Google Map.

This bearing contains the electrical micro-stepping motor that will rotate the dish.

The whole mirror assembly rests on a three point mount that is used to precisely align the mirror axis with the local gravity. This mount is also visible on fig. 1.3 (right). The effect of a misalignment of the mirror axis will be investigated in chapter 3

It was first planned to install the ILMT in La Silla (Chile) but a new Indian observatory was finally chosen. It is the Devasthal observatory located near the town of Nainital in the North of India, in the South-West of the Himalaya mountains. Information about this site can be found in

Telescope	
Resolution	0.6" FWHM
Field of view	$22' \times 22'$
Life expectancy	10 years
Effective focal length	< 10 m
Mirror	
Diameter	4 m
Focal length [1]	8 m (F/2)
Detector	
CCD	4096×4096 pixels
Pixel size	15 μm/pixel
Pixel angular size	0.4"/pixel
Filters	g' r' i' in a filter slide (i' permanent)
Read out noise	4 e^-/pixel
Dark current	$5 \cdot 10^{-4}$ e^-/pixel/second
Full-well capacity	$> 250ke^-$
Cooler	Cryo-tiger
Corrector	
Resolution	FWHM $\leq$ 0.2" (at focus)
Number of lenses	5
Glass	N-BK7
Conic constant	0 (spherical)
Location	Devasthal (India)
Latitude	29°22'46" North
Longitude	79°40'57" East
Median seeing	< 1.2"
Altitude	2400m
Atmospheric air pressure	750 ± 100 mbar
Relative humidity	$< 80\%$ (70% of the time)
Limiting magnitude	22.5 (in one TDI time: $\sim 102s$)

Table 1.2: Technical characteristics of the ILMT.

Sagar et al. (2000) and is summarized in Table 1.2 with the general characteristics of the ILMT. The location of the observatory is presented on the map of India shown in fig. 1.4

A 2m LMT prototype has been built by the Liège Space Center in order to demonstrate that it was realistic to use a liquid mirror as the primary mirror of an astronomical telescope. This project and some other liquid mirror telescopes around the world are presented at the end of this chapter.

1.2 The liquid mirror

1.2.1 Principle

A liquid mirror is an optical device that is based on the following principle (exposed for example in Borra 1982). When a liquid is in an equilibrium state, its surface follows an equipotential surface. It is thus perpendicular to net forces that apply on the liquid. If the liquid is rotating in the constant field of gravity, the shape of its surface is a paraboloid. Indeed, the liquid undergoes two different forces; the gravity that follows a constant vertical downward direction and the centrifugal pseudo-force that is horizontal and increases as the square of the radius. The vectorial combination of these two forces results in a vector that is vertical on the rotation axis of the fluid and gets more and more inclined when the radius is increasing. Hence, the surface of the liquid sets in a paraboloid shape. This principle is represented in fig. 1.5.

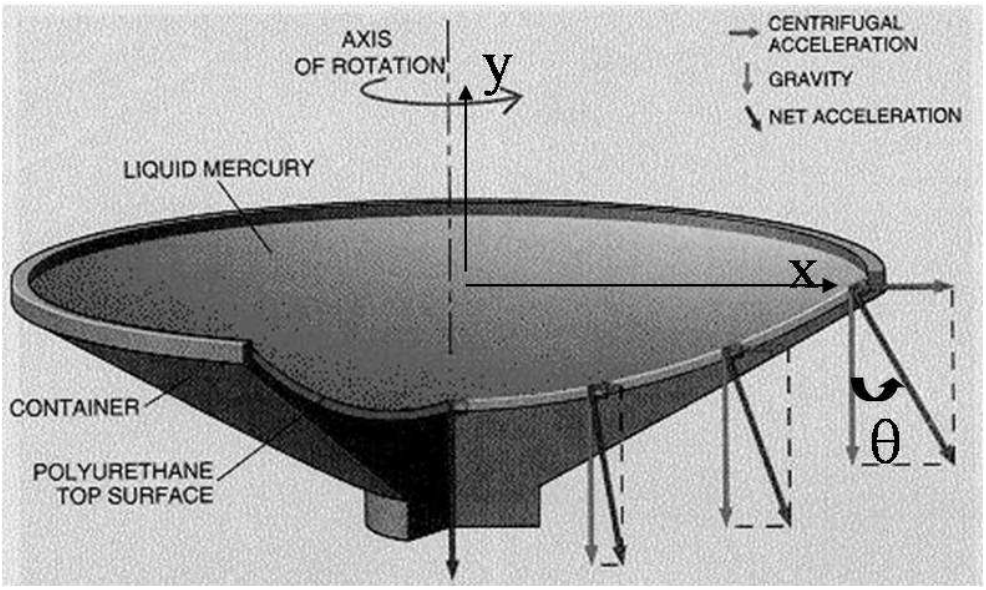


Figure 1.5: Principle of the liquid mirror. A reflecting liquid is spread out in a rotating dish. The liquid then undergoes two accelerations, the centrifugal (horizontal arrows) and the gravitational (vertical arrows) ones. The free surface of the liquid sets itself perpendicularly to the net acceleration it undergoes (the vectorial sum of the two others.). Combination of the rotation and the gravitational field of the Earth thus generates a parabolic surface. Image from the website of "futura-sciences" (http://www.futura-sciences.com/).

Let us now demonstrate that the shape is indeed a paraboloïd. The tangent of the angle between the vertical axis and the net force (θ) is given by

$$\tan(\theta) = \frac{dy}{dx} = \frac{\omega^2 x}{g} \tag{1.1}$$

where $\omega^2 x$ is the expression of the centrifugal acceleration and g is the acceleration of gravity. The vertical position y of the fluid element as a function of its radial position x is found by simple integration of the previous equation

$$y = \frac{\omega^2 x^2}{2g} \tag{1.2}$$

which is the equation of a parabola with a focal length given by

$$F = \frac{g}{2\omega^2} \tag{1.3}$$

Such a parabolic shape is very interesting for the primary mirror of an astronomical telescope. Indeed, the light coming from a source located infinitely far from the mirror (as a star) and parallelly to its symmetry axis converges to the focus of the mirror. The parabolic mirror is said to be aberration free for the rays of light that are parallel to its axis. In the same case, a spherical mirror would have several foci depending on the radius where the rays would be reflected. The marginal rays converge further away from the mirror than the central rays. This is called the spherical aberration. On the other hand, when the observed object is not located on the axis of the mirror, field aberrations, such as coma, appear in the image. A parabolic mirror thus requires a field corrector, i.e. an optical device designed to correct for the off axis coma aberration in order to get a larger "well-corrected" field of view.

Liquid mirrors thus offer a practical way to easily get cheap large primary mirrors for telescopes. However, telescopes based on this technology present several limitations, the main one being the impossibility to orient it. It is obvious that the mirror cannot be tilted for a physical reason, the rotation axis has to be aligned with the local gravity and a practical reason, the liquid has to stay in the bowl. As a consequence, such a liquid mirror telescope cannot track stars in the sky, it can just look at them while they are crossing its field of view. This is why the LMTs are often called zenithal telescopes.

Fortunately, the progress in imaging technology allows to use this type of telescope. Indeed, using a CCD camera with a particular readout mode called "Time Delay Integration" allows to follow the field displacement. This particular readout mode will be explained in section 1.3.

Moreover, the total field of view of the telescope is not limited to the $22' \times 22'$ announced in Table 1.2. Indeed, as the Earth is rotating around the polar axis, the field of view of the telescope makes a 360° turn giving access to about 156 square degrees (88 squared degrees at high galactic latitude - $|bII| > 30°$ - see fig. 1.6) of sky for the specific latitude of Devasthal (29°22'46"N). Each night, the same strip of sky crosses the field of view of the telescope. It fact, that would be true if the Earth was only rotating on its axis. Because of its revolution around the Sun and various phenomenon of nutation and precession, the strip of sky will be slightly different from one night to another.

Even if the theoretical shape of the mirror is a perfect paraboloid (interferometric tests presented in Borra et al. 1992, show an rms precision of the surface of $\lambda/20$), the surfaces of the mirrors obtained in practical cases are disturbed by many causes. These are presented in chapter 3.

The best way to minimize these disturbances consists in reducing the thickness of the liquid layer. This explains why the shape of the container should be as near from the final parabola as possible. This is achieved with the spin-casting of the polyurethane. This material is liquid when its components are mixed together and it hardens within half an hour. It is poured in the dish while it is still liquid. The dish is then rotated at the same speed as with the mercury to form the final mirror. The polyurethane thus takes the final paraboloidal shape and hardens keeping this shape. The mercury then only serves to make the surface reflective and still closer to the final paraboloid.

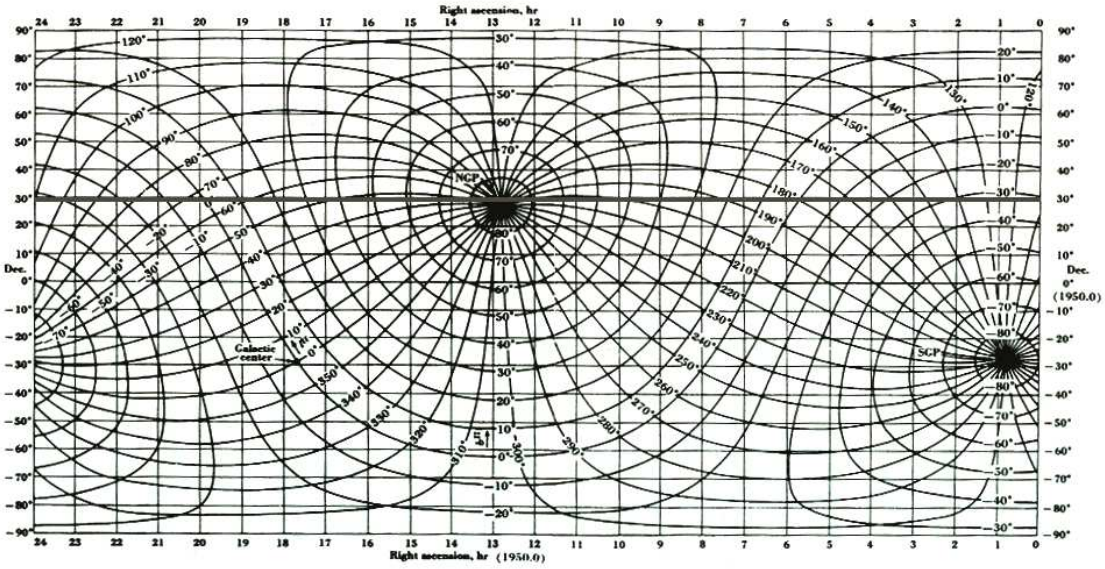


Figure 1.6: Graphical representation of the relation between celestial (RA, Dec) and galactic coordinates. The thick line represents the area of sky observed by the ILMT installed at Devasthal observatory. Image from the "New Jersey Science and Technology University" website (http://web.njit.edu/ gary/321/Lecture18.html).

1.2.2 Reflecting liquids

As shown in the previous section, the surface of a liquid rotating in a constant field of gravity takes a parabolic shape. When the liquid that is used is reflective, one gets a parabolic mirror.

Now, we have to find a suitable liquid. Only metals have a sufficient reflectivity. There are only a few metals that are liquid at room temperature. Among these, the mercury is the most common, but other liquids have been considered for LMTs. They will be reviewed at the end of this section.

Mercury (Hg)

The elemental mercury presents all the required characteristics to make a good liquid mirror. Its reflectivity is above 75% at optical and infrared wavelengths (fig. 1.7). This corresponds to about 90% of the reflectivity of an aluminum-coated glass mirror.

Moreover, it is liquid above -38.8°C, enabling its use at room temperature. Another advantage of mercury is its relatively inexpensive cost ($\sim$ 15\$/kg) compared to gallium for example ($\sim$ 4000\$/kg).

Mercury reacts with ambient air to form a transparent oxide layer. This oxidization happens during the first few hours of the surface stabilization. Once it is created, the oxide layer "reinforces" the mirror as it prevents the surface to break because of small perturbations such as insects falling on the mercury. It also decreases the evaporation rate of the mercury, which is a very interesting property since these vapors are highly toxic.

Fig. 1.9 shows the central part of the Large Zenithal Telescope (LZT - presented at the end of this chapter) while it is filled with mercury. The loading system consists in a peristaltic

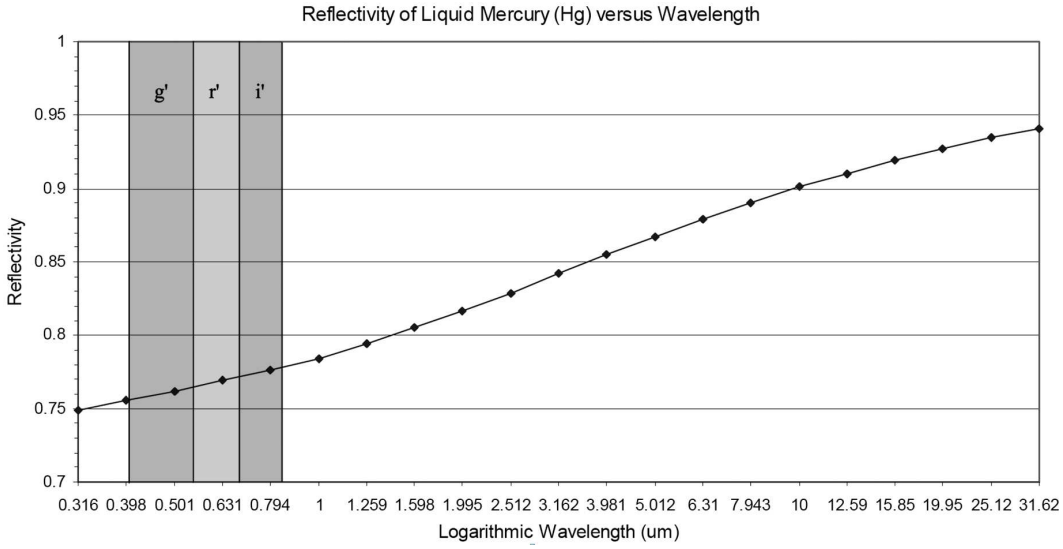


Figure 1.7: Reflectivity of oxidized mercury as a function of the wavelength. It is higher than 75% at optical and infrared wavelengths (The spectral bands we are interested in are represented). This is lower than classical coated glass. However, the coating of a glass mirror degrades and is expensive to refresh. A liquid mirror can be cleaned very easily. Image from Mulrooney's thesis. (Mulrooney 2000)

pump that brings the mercury directly on the surface of the mirror. Indeed, because of its high density (13.6 g/cm^3), handling of the mercury can reveal to be difficult especially because of the high amount of kinetic energy released on impact. Pouring the mercury from a too large height results in the creation of many droplets that can rebound everywhere, even higher than the original height.

A total of about 45 liters (500-600kg) of mercury will be loaded in the ILMT dish. Such a quantity will correspond to a layer with a thickness of about 3mm that is necessary at the start up of the mirror in order to spread it out more easily. Indeed, the surface tension of the mercury is very high and it prevents an easy spreading of the liquid. A large amount of mercury is thus necessary to close the mirror surface.

However, it has already been said that the disturbances in the mercury layer are better damped with a thinner layer of liquid. During the observations, the thickness of mercury should not exceed 1mm. This means that once the mirror will have been formed, some of the mercury will be pumped out of the bowl in order to reduce the layer thickness. The amount of mercury during operation should be around 170kg (12.5l).

To support the large quantity of mercury required for the start-up of the mirror, the air bearing must be very robust, both in compression (to support the weight) and in tilt (to resist to unbalanced weight due to non-uniform distribution of the mercury).

Mercury and safety

Mercury cannot be mentioned without talking about its almost legendary toxicity. Here, two forms of mercury should be distinguished. The elemental form, that is an inorganic salt, and the organic compounds. The latter is easily absorbed by the biological systems and it especially affects the central nervous system. These organic forms (mainly methylated mercury and mercury

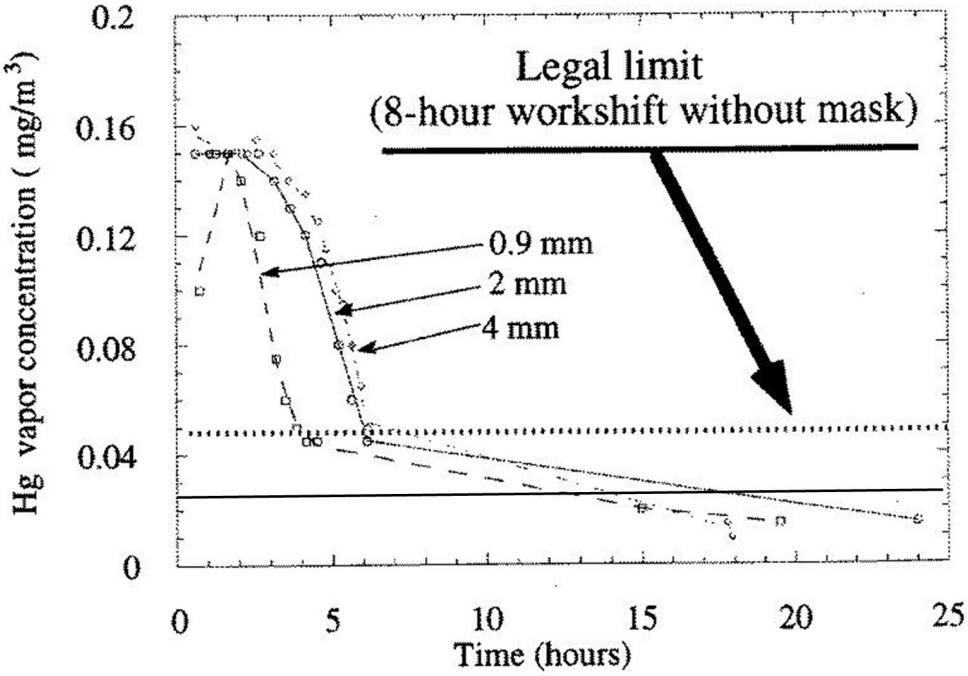


Figure 1.8: Mercury vapor concentration as a function of time for several thicknesses of mercury layers. The solid line represents the legal limit below which the level of vapor has to be in order to work eight hours without a gas-mask. The vapor concentration quickly drops below this limit thanks to the mercury oxide layer that rapidly forms at the surface and that prevents evaporation to occur. The current limit of $25\mu g/m^3$ is represented by the solid line, whereas the dotted line represents an older limit.

nitrate) are associated with mercury poisoning and death.

As far as the elemental inorganic form is concerned, it is mainly assimilated via the inhalation of its vapors that are absorbed in blood and transported through the organism. It is however not absorbed through the skin or digestive system, there is thus almost no risk of contact or ingestion contamination. The long term inhalation of mercury vapors is associated with chronicle health effects such as muscle tremors, loss of motor skills, memory loss, hallucinations or personality changes.

A half-mask respirator equipped with a sulfur impregnated activated carbon cartridge is completely sufficient to work safely during 8 hours with a mercury vapor concentration between 25 and $500\mu g/m^3$. Above this upper limit a supplied-air respirator is mandatory. However, such quantities of vapors are rarely reached during the operation of a liquid mirror. Indeed, as previously said, the elemental mercury oxidizes when it is in contact with the air. The oxide layer drastically decreases the evaporation of mercury, even below the $25\mu g/m^3$ limit. The only periods during which the vapor concentration gets high correspond to the pumping (in or out) of the mercury or when the surface breaks, i.e. when the oxide layer does not cover any longer the whole mercury surface. A complete study of the mercury vapor concentration is presented in Hickson et al. (1993), and the evolution of the vapor concentration as a function of time is presented in fig. 1.8.

Even if the contact with mercury should not be an issue, transporting it outside the confined room of the telescope should be avoided. In order to ensure a complete protection of the worker

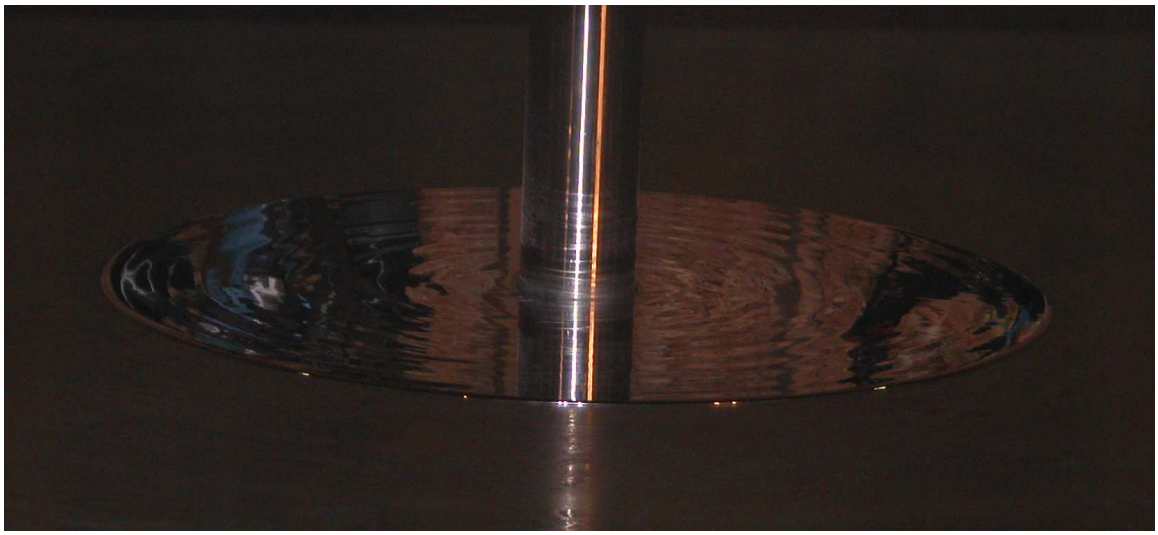


Figure 1.9: Mercury flowing in the LZT dish during the filling of the mirror.

and of the environment, complete protective suits and gloves, in addition to the gas mask, have to be worn when the mercury is manipulated. Fig. 1.10 shows two people from the EASO group working with the 2m LMT prototype with their full equipment.

Figure 1.10: Mercury in the 2m LMT dish. A gas mask is necessary when working with mercury until an oxide layer has been formed on the surface. Gloves and protective suits are mandatory to avoid contamination. Davide Ricci and the author are seen on this photograph.

Other Liquids

As it has been presented, mercury, that is usually used as the reflective liquid for liquid mirror telescopes, has two main disadvantages: its high density and its toxicity. Other liquid metals have thus been considered to replace mercury as the reflecting liquid.

Gallium is another liquid metal with a good reflectivity. Its properties have been studied in Borra et al. (1997). It presents the advantage of not being toxic but it is very expensive ($\sim 4000\$/$ kg) and it is not liquid at room temperature since its melting point is around $30°C$. The use of gallium thus requires a particular processing, like using an eutectic alloy or supercooling, to be usable as a reflecting liquid. Eutectic alloy has a melting temperature that is lower than those of its components (example: bronze is an eutectic alloy of copper and tin). Supercooling is a technique that allows keeping a material liquid below its melting point. These disadvantages would not be too important in regard of the great advantage of non toxicity, but it has been shown in Borra et al. (1997) that the gallium oxide is not transparent. That renders it completely useless for liquid mirrors.

Rubidium and Cesium are also liquid metals at room temperature but their low reflectivity and high chemical reactivity make them unsuitable for use as a liquid mirror.

Borra's team has also studied the possibility to make a deposition of a reflective film on a viscous fluid (see for example Borra et al. (1999) and Gagné et al. (2008)), but as we do not plan to use them for the ILMT, we will not detail them here.

1.3 The CCD Camera of the ILMT

A Charge Coupled Device (CCD) camera is a modern imaging system, composed of an electronic device that is able to measure the intensity of the light it receives. The detector is subdivided in pixels (picture elements): this gives information about the spatial distribution of the photons falling on the detector and enables the camera to generate an image.

The principle is quite simple, it is illustrated in fig. 1.11. Let us consider the light as rain. The pixels of the CCD would be buckets. As different droplets fall in different buckets, different photons hit different pixels. When the rain stops, it is possible to measure the volume of water in each bucket, that gives the spatial distribution of rain. In the same way, when the acquisition stops, the CCD is readout and the spatial distribution of the photons is determined, an image is produced.

Each pixel is made of a photo sensitive semi-conductor. Thanks to the photo-electric effect, when a photon hits the semi-conductor, an electron is released and captured by the potential well associated with this particular pixel. When the acquisition time is finished, the CCD is obscured so that no more light can reach it.

Then the readout begins. It happens sequentially, the first row of the CCD (the first series of buckets at the right of the detector) is transferred to the output register (the three extreme right buckets), the second row is transferred to the first one and the third one is transferred to the second one.

The output register is then readout. The first "bucket" is transferred to a measurement unit that "counts" the number of electrons that were contained in that bucket. The second bucket is

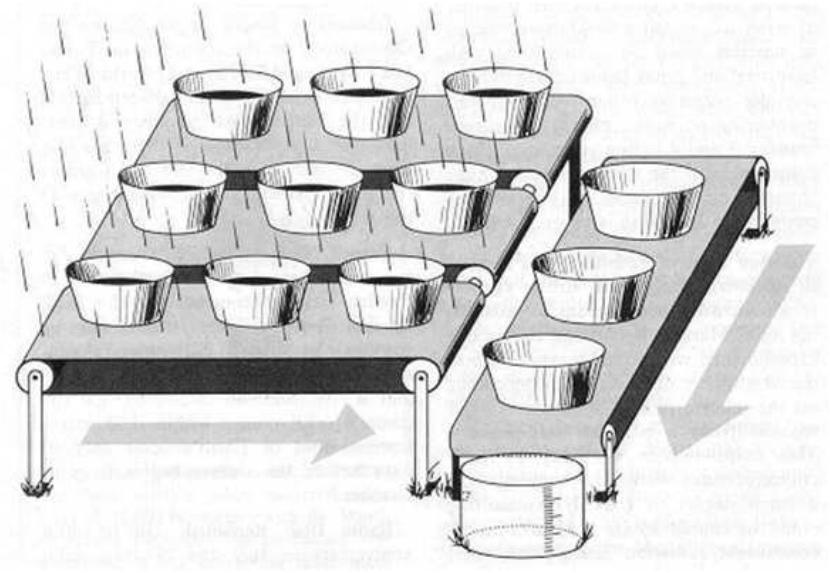


Figure 1.11: Principle of the CCD detector. If the photons are considered as drops of rain, each pixel behaves as a bucket, it collects the droplets that fall on it. It is then possible to determine how much rain fell in each bucket. The 3×3 buckets on the horizontal bands constitute the imaging area of the detector and the three right buckets represent the output register that is used during the readout. Image from Kristian and Blouke (1982).

transferred into the first one and the third one into the second one. This process is repeated until all the buckets of the output register have been measured. Then the next row of the detector is transferred to the output register and the same process is repeated until the whole CCD has been read.

In a classical CCD camera, the chip is readout as fast as possible (limited by the readout noise). However, it is also possible to adjust the readout speed, especially the transfers between the rows, in such a way that the charge transfer occurs at the same speed as the objects crossing the field of view of the camera. This is called the "Time Delay Integration" (TDI) mode. The principle of this readout mode and its interests are detailed in section 1.3.1.

In practical cases, the transfer of the charges from one pixel to another is slightly more complicated. Each pixel is composed of several (at least three) electrodes that are used to move the charges from one pixel to another (fig. 1.12). These pixels are separated by insulating material in order to avoid unintentional transmission of charges.

During the acquisition time (t1), electrodes P3 are brought to a positive potential while the two outer electrodes are negative. The released electrons are thus trapped in the potential well created by P3 (fig. 1.12). During the readout (t1 - t4), the potential of the electrodes on the right of the wells (P1) becomes positive and the one of the electrodes P3 becomes negative (t2). The electrons are thus shifted to this new position at the right of their original pixel. The same process occur between each pair of electrodes (P3-P1, P1-P2, P2-P3) until the electrons reached the next pixel. At that time, the right-most row of the CCD has been transferred to the output registry, which is then read in the same way. At the output if this registry, the electrons are "counted" by the output amplifier, and the signal measured at this amplifier is proportional to the number of photons that reached the CCD.

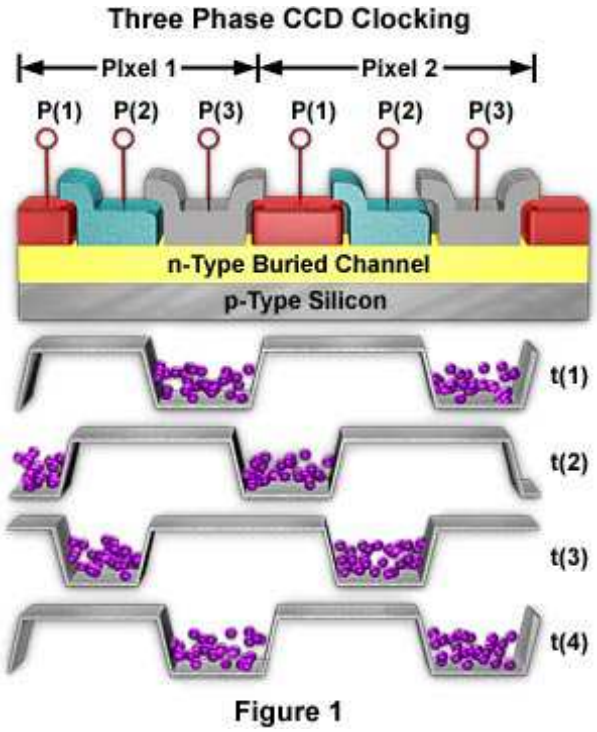


Figure 1.12: Illustration of the charge transfer between the pixels of a CCD. Each pixel is divided by three electrodes P1, P2, P3, that create a potential well. During the acquisition phase (t1), electrodes P3 are set to a positive potential, while the two others correspond to a negative potential. The electrons reaching the pixels are thus trapped in the potential wells of the electrodes P3. During the readout of the chip, the potential well is successively moved from one electrode to another (sequence t2, t3, t4), simply by modifying the potential of the electrodes. From t1 to t2, the electrodes P1 are set to a positive potential, which enlarge the potential well, and the electrodes P3 are then set to a negative potential. Hence the potential well has been displaced from P3 to P1. The same process then occurs between the electrodes P1 and P2. The image comes from http://micro.magnet.fsu.edu/primer/digitalimaging/concepts/threephase.html

The current CCD detectors are far more sensitive than the photographic plates that were previously used in astronomy. This type of detector also replaced the silver photographic films that were used in classical photography. The quantum efficiency, that represents the fraction of photons that are converted into electrons, can reach up to 95% with back illuminated devices instead of a few percent for the photographic plates. The other advantages are the very good linearity of the detector, a very low amount of noise, a high dynamic range of detection and a high photometric precision.

The drawbacks are related to the electronics and consist of bad pixels (pixels that do not react to photons as expected), dark current (electrons can be thermally excited) and readout noise (during the transfer of the charges, the variation of the potential can excite electrons). However, these noises are relatively low.

A CCD detector is generally characterized by the number of pixels it is composed of, the size of these pixels and the associated full well capacity, that is the maximum number of electrons that can be stored in a single pixel. The full well capacity defines the dynamic of the signals that can be recorded.

The number and size of pixels determine the size of the detector and the field of view (Fov) of the instrument together with the sampling of the image.

$$Fov = \frac{n_{\text{pix}} \cdot w_{\text{pix}}}{f} \tag{1.4}$$

where n_{pix} is the number of rows on which the signal is integrated, w_{pix} is the pixel size expressed in meters and f is the effective focal length of the telescope in meters. The field of view is then given in radians.

Chapter 2 details the characteristics of the CCD chip that has been chosen to build the camera that will equip the ILMT. All the specifications required for the camera are also presented in that chapter. It also presents some tests that have been performed on a 2048×2048 CCD camera that could have been used to perform optical tests of the ILMT. This camera was initially installed on the 2m CSL LMT prototype.

1.3.1 Time Delay Integration

As previously explained, a liquid mirror telescope cannot be tilted. A telescope based on such a mirror is thus unable to track celestial objects that seem to move in the sky as the Earth rotates. This is a major problem for a telescope. Looking at the zenith, an LMT only sees the stars crossing its field of view. This is not convenient at all and a method to artificially follow the motion of stars is mandatory.

The Time Delay Integration (TDI) technique, also called drift scanning, consists in adjusting the rate of charge transfer from a row to the next one to the crossing speed of the stars (the sidereal rate). The rows of the CCD are shifted in such a way that the image formed on the detector follows the object moving with the sky. Fig. 1.13 illustrates the principle of the TDI technique.

Let us note that the rows of pixels of the CCD chip correspond to the columns of the images. It is indeed the rows of pixels that are shifted during the readout and not the columns even if on the computer display the columns seem to be shifted during the acquisition.

This readout mode is presented in Gibson and Hickson (1992b) where the history and principle of the method are reviewed and the related image deformations are analyzed.

The integration lasts during the whole crossing time of the observed object. The integration time directly depends on the CCD field of view and inversely on the apparent motion speed of the star that only depends on the latitude of the observatory (since the telescope only looks at the zenith). It is given by:

$$T = 1.37 \cdot 10^{-2} \frac{n_{\text{pix}} \cdot w_{\text{pix}}}{f \cdot \cos(l_{at})} \tag{1.5}$$

where n_{pix} is the number of rows on which the signal is integrated, w_{pix} is the pixel size in microns, f is the effective focal length of the telescope (in meters) and l_{at} is the latitude of the observatory. T is then given in seconds. The integration time will thus be about 102.19 seconds (effective focal length: 9.4524m) for a liquid mirror telescope located at Devasthal and equipped with a 4096×4096 15μm-wide pixel CCD camera.

This allows to reach a limiting magnitude around 22.5 in the i' band. However, this limit holds for a single integration time but, as the same strip of sky is observed night after night,

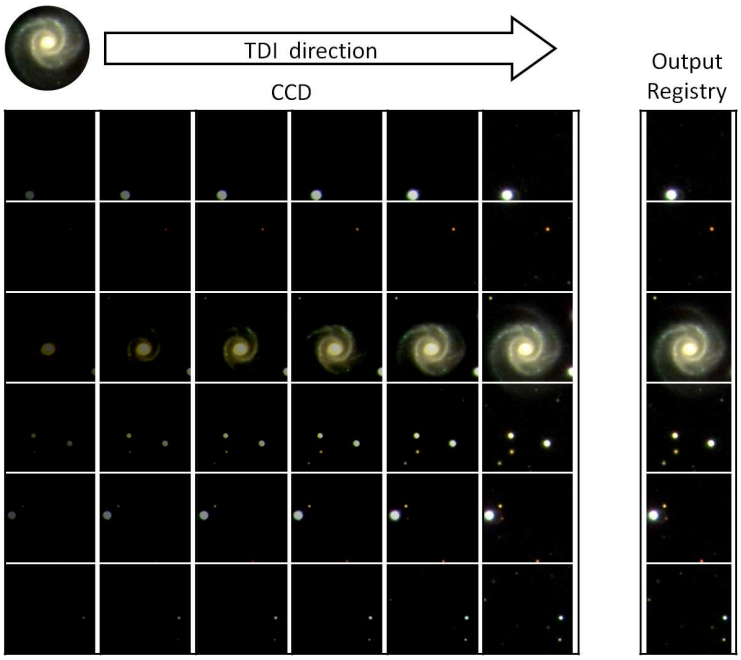


Figure 1.13: Principle of the Time Delay Integration. The rows of the CCD camera are shifted at a particular rate that corresponds to the sidereal rate. The integration of photons continues during the shift. The integration time thus depends on the time needed for the observed object to cross the CCD field of view. The final image is made of the sum of all intensities accumulated on each row. The original sky image shown above has been acquired with the LZT (http://www.astro.ubc.ca/LMT/lzt/)

these observations can be co-added to increase the limiting magnitude. The cumulation of 60 nights of observation increases the limiting magnitude by 2.2[2].

1.3.2 Filters

The CCD is a detector that is very well suited for photometric and astrometric measurements. However, this type of detector is not able to distinguish the wavelength of the photons that fall on it. The images captured with such a detector are thus monochromatic. Nevertheless, some applications, as supernovae study, require color information that can be acquired with a spectrometer, but it is not foreseen to equip the ILMT with such an instrument.

Another solution consists in inserting filters in the path of the light, typically just before the entrance window of the camera, to limit the acquisition to a particular range of wavelengths. Using such filters allows to obtain photometric information related to the corresponding spectral bands. Recording the fluxes for several colors corresponds to create a low resolution spectrum of the observed objects.

As far as the variability survey is concerned, we are mainly interested in observations in one spectral band, but secondary objectives would be better served with some color information. The scientific objectives of the ILMT are presented at the end of this chapter.

The observational strategy consists in observing mainly with one filter (i') and to change for

[2]The magnitude increases as $2.5\log(\sqrt{n})$, where n is the number of co-added frames.

example every few days to the r' or g' filters. The filter would not be changed during the night but only between two nights. The g', r', i' filters, equivalent to the Sloan ones (Fukugita et al. 1996) are defined in Table 1.3.

Color	Symbol	Central wavelength	Spectral width	Average transmission
Green	(g')	477.0nm	140nm	
Red	(r')	623.1nm	140nm	
Near-infrared	(i')	762.5nm	150nm	

Table 1.3: Spectral bands of interest as defined by the Sloan Digital Sky Survey filters (Fukugita et al. 1996).

1.4 A specific corrector

An LMT is a zenithal telescope that looks at the sky passing above it. The tracking is artificially achieved by electronically stepping the rows at the sidereal rate. This is however not sufficient to get a good image of the observed objects. Indeed, the tracks of the objects are not rectilinear. This curvature is extensively studied in Hickson and Richardson (1998). The curvature of the trails depends on the latitude of the observatory, the radius of curvature at the center of the CCD is given by

$$R = F \cot \delta \tag{1.6}$$

where δ is the declination of the star, that corresponds approximately to the latitude of the observatory and F is the focal length of the telescope. The displacement along the North-South axis (Y) as a function of the East-West position (X) is given by

$$Y = \frac{\sin \delta_0 \cos \delta_0}{\sin^2 \delta - \cos^2 \delta_0} \times \left(1 - \left\{ 1 - \frac{(\sin^2 \delta - \cos^2 \delta_0)[(1 + X^2) \sin^2 \delta - \sin^2 \delta_0]}{\sin^2 \delta_0 \cos^2 \delta_0} \right\}^{1/2} \right) \tag{1.7}$$

where δ_0 is the declination of the observatory and δ is the declination of the star.

Such a curved trajectory on the sky is visible on the CCD (see fig. 1.14). It introduces an additional aberration in the images. These deformations are presented and discussed in Vangeyte et al. (2002). Moreover, the transit speed of the stars depends on the North-South location of their track on the CCD. The stars at the North move more slowly than those at the South. However, the rows of the CCD are shifted at a constant rate over the whole North-South extension of the chip. The drift scan is thus too fast at the top of the chip and too slow at the bottom. This generates aberrations on the images. These distortions relative to the curved trajectories and to the differences between the star speeds and the row drift rate are called the TDI distortions. They are presented in fig. 1.15.

The star trail curvatures have thus to be compensated so that they are aligned with the CCD column and the variation of the star crossing speed should also be accounted for. Hickson and Richardson (1998) have proposed an optical corrector that "anti-distorts" the field of the

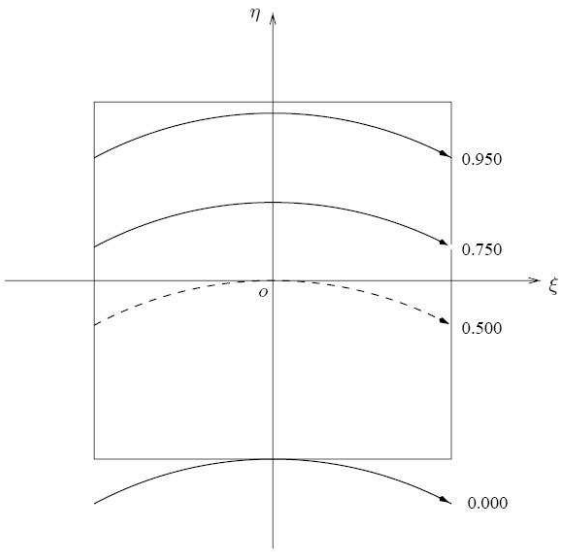
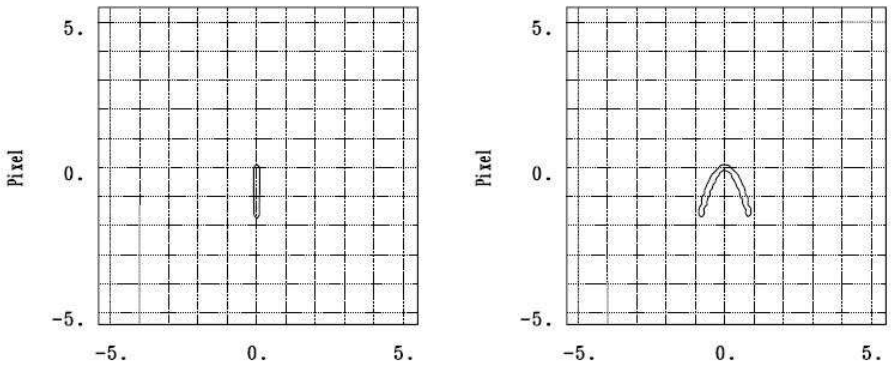
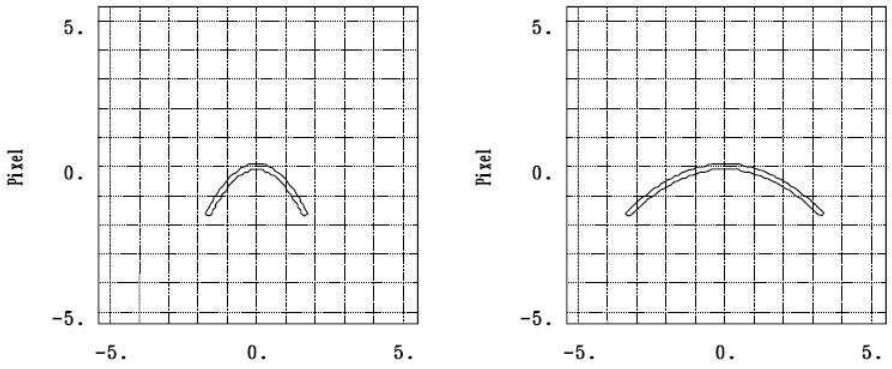


Figure 1.14: Star trajectories passing across the ILMT field (in the southern hemisphere) and aligned along the direction of the star apparent motion ($o\xi$ axis); the $o\eta$ axis points toward the North. The dashed curve illustrates the apparent trajectory of a star passing through the optical center. Image from Vangeyte et al. (2002).



(a) $4\times4C$ and 0.500-trail.

(b) $4\times4C$ and 0.625-trail.



(c) $4\times4C$ and 0.750-trail.

(d) $4\times4C$ and 1.000-trail.

Figure 1.15: Uncorrected star traces on the ILMT CCD for various positions of the star trail. These traces have been computed for a telescope located at El Toco (Chile). (a) case of a star passing at the center of the CCD. (d) case of a star passing at the top of the CCD. (b) and (c) correspond to intermediate positions. Image from Vangeyte et al. (2002).

	Diameter	R_1	R_2	t
L1	550	336.63	336.63	65
L2	250	2002.85	213.32	15
L3	250	1619.63	-601.72	25
L4	200	193.93	220.68	30
L5	125	-472.39	-251.54	15

Table 1.4: Characteristics of the lenses composing the ILMT corrector. All these lenses are spherical (conic number=0) and made of "N-BK7" glass. R_1 and R_2 are the radii of curvature of the first (in) and second (out) surfaces of the lenses respectively, t is the thickness of the lenses. All these values are expressed in millimeters.

telescope in such a way that the TDI distortions are compensated. This corrector also accounts for the classical field aberration of the parabolic mirror.

Two models of corrector are presented in figs. 1.16 and 1.17. The first one was designed to equip the 2m LMT prototype of the Liège Space Center (CSL), it is composed of four lenses, some of them being aspherical or/and made of exotic glass. All these lenses are co-axial. The second model, that will equip the ILMT, is composed of five spherical N-BK7 lenses. They are decentered and tilted to correct for the TDI distortion. This second approach is easier to implement, especially when large elements are needed, as it only requires spherical lenses. The first lens of the ILMT corrector has a diameter of 55cm, that is very large, even for a spherical lens. The characteristics of the ILMT corrector lenses are summarized in Table 1.4.

The corrector designs presented in figs. 1.16 and 1.17 have been performed with the optical simulation software, Zemax. This is a very complete tool that allows to perform tolerance studies, optimization and optical tests. It is used for the design and analysis of optical systems. Zemax can perform standard sequential ray tracing through optical elements, non-sequential ray tracing for analysis of stray light, and physical optics beam propagation.

It can model the propagation of rays through optical elements such as lenses (including aspheres and gradient index lenses), mirrors, and diffractive optical elements.The effect of optical coatings can also be accounted for.

Zemax can generate Point Spread Functions (PSFs), spot diagrams or other standard analyses. It includes a complete library of glasses and lenses from several manufacturers.

We used it intensively to model the ILMT optical system in order to analyze it as well as to simulate optical tests for several elements that compose this system. We also managed to have Matlab interacting with Zemax in order to automatize the testing procedure, especially as far as alignment methods are concerned. This particular point will be detailed in section 7.2 of chapter 7.

Fig. 1.18 presents spot diagrams of the ILMT optical system with the corrector installed and simulating the motion of several stars in the field of view. A spot diagram consists of a geometrical optics ray tracing simulation of the system. The sources are positioned and a large number of rays are "shot" from them. The propagation of these rays is computed from geometrical optics rules. The distribution, in the focal plane, of rays coming from the source is called the spot diagram. This type of analysis can be achieved for various angles off the axis, wavelengths and positions of the focal plane, to study the effect of defocusing. Such a study gives an idea of the shape of the images formed by the optical system, neglecting the diffraction phenomenon.

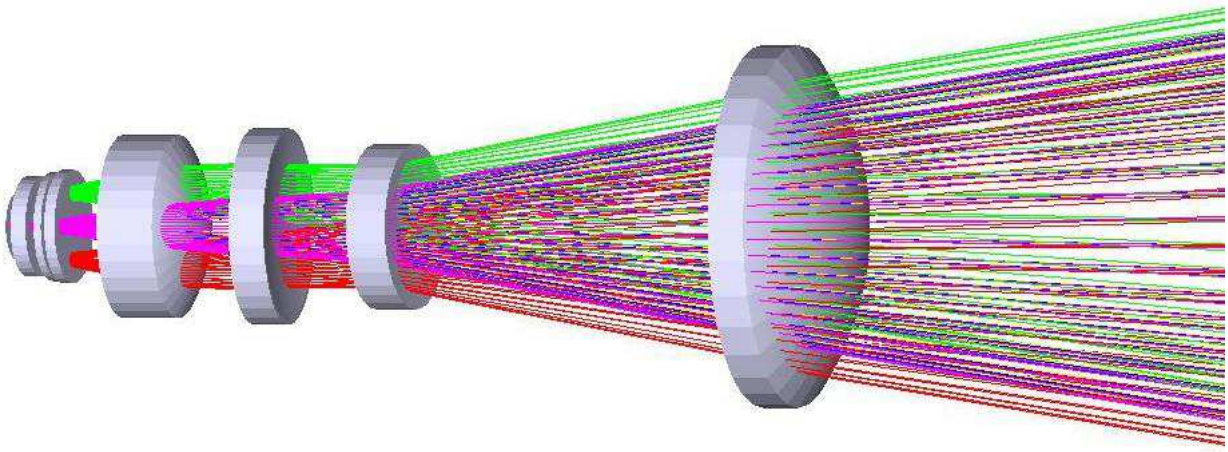


Figure 1.16: Three dimensional view of the corrector of the 2m LMT obtained from the Zemax model. The four lenses on the right constitute the corrector. They are aspheric but they have all the same axis. The 3 elements to the left correspond to the filter, the entrance window of the detector and the CCD chip. The diameter of the first lens on the right is 166mm and the entrance window of the camera is 46mm wide. The distance between the first lens and the focal plane is around 407mm.

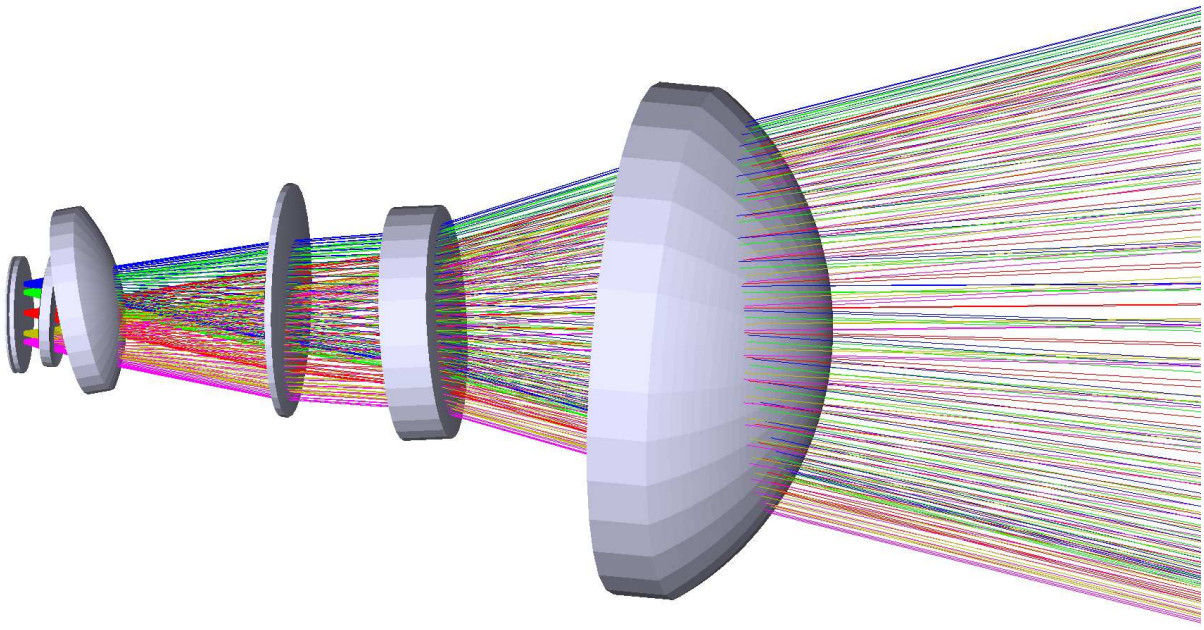


Figure 1.17: Three dimensional view of the TDI corrector for the ILMT obtained from the Zemax model. The five lenses are spherical but they are tilted and displaced from the axis of the corrector. The 3 elements (only 2 are visible) on the left correspond to the filter, and the entrance window of the detector. The diameter of the first lens on the right is 550mm and the entrance window of the camera is 125mm wide. The distance between the first lens and the focal plane is around 885mm.

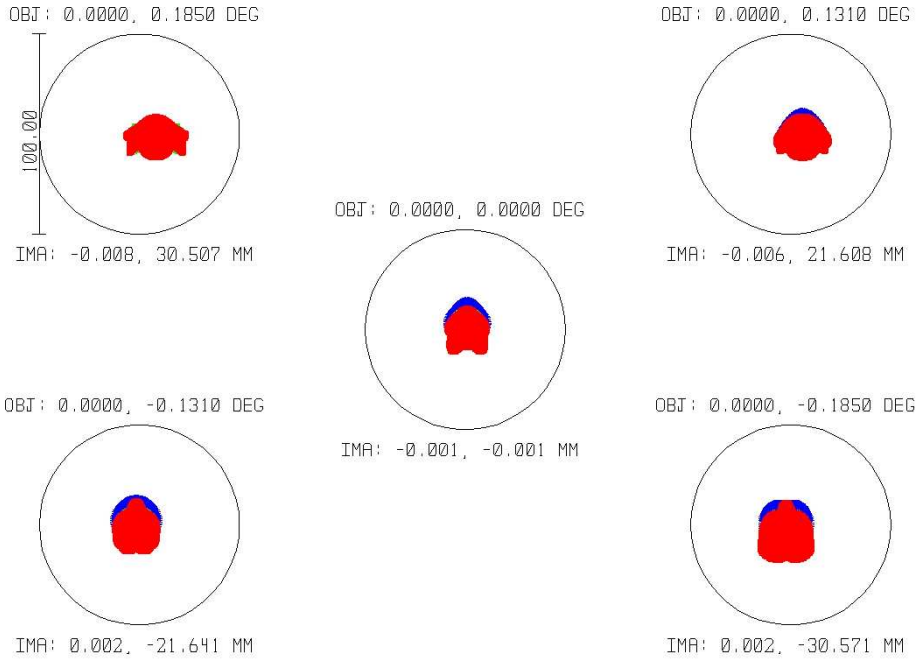


Figure 1.18: Spot diagrams, obtained from the Zemax model, of the ILMT system with the optical corrector installed. The circles that surround the spots have a diameter of $100\mu m$. The five spots represent five different star positions along the North-South axis. This aims at optimizing the corrector for the whole height of the field of view. These spots have a diameter of about three pixels and a roughly symmetrical shape that testifies a good correction.

The different spots are related to several positions of the source in the field of view. They correspond to geometrical images formed by the ILMT equipped with its corrector. These images are quite symmetrical and their diameters correspond to about three pixels. They present small chromatic aberrations.

Even if the simulations are promising, such an optical device is difficult to test, mainly because of the size of the lenses, but also because of the need of a moving source. Indeed, a converging beam that has a diameter larger than 50cm at the level of the first lens of the corrector can hardly be obtained without the primary mirror of the telescope. Moreover, it is difficult to get a moving source that mimics the motion of the stars. The easiest way of testing the corrector is thus to install it at the top of the ILMT structure and to look at the sky. However, the correction is optimized for India and it will not be good in Liège. Another solution is thus required to test the corrector. One, that involves the Nijboer-Zernike theory of wavefront phase retrieval, is proposed at the end of this thesis.

1.5 Liquid Mirror Telescopes scientific roadmap

The basic principle of liquid mirrors has been known for a long time. Newton was the first to realize that the free surface of a liquid always sets perpendicularly to the net force it undergoes. Following the same reasoning as the one presented in section 1.2.1, he deduced that a rotating liquid could be used as the primary mirror of a telescope. The concept has been developed by Ernesto Capocci in 1856 and a first liquid mirror was built in 1872 by Henry Skey. However he encountered technical difficulties due to angular speed stability and motor-mirror coupling. More than thirty years later, in 1909, Robert Wood built a 51cm diameter liquid mirror. He discovered that its surface was disturbed by vibration induced wavelets that made it completely useless. The ball-bearing was the cause of the transmission of these vibrations. In addition to these mechanical problems, the impossibility to tilt the mirror, and thus to track stars on photographic plates, was a show-stopper for this type of telescope.

It is only since the eighties that the technology has sufficiently evolved to circumvent those technical issues. Particularly the use of CCD sensors allows to electronically track stars thanks to the Time Delay Integration acquisition mode (TDI, see section 1.3.1). Given the advantages of new technologies, Ermanno Borra revived the liquid mirror telescope concept (Borra 1982). At that time began the modern era of Liquid Mirror Telescopes (LMTs).

The telescopes built since then are presented in the remainder of this section along with the scientific projects they were used for. We first present the specificities of the observations made with liquid mirror telescopes as well as their advantages for astronomical observations. The different LMTs that will be introduced in this section are chronologically presented on a time-line in fig. 1.19.

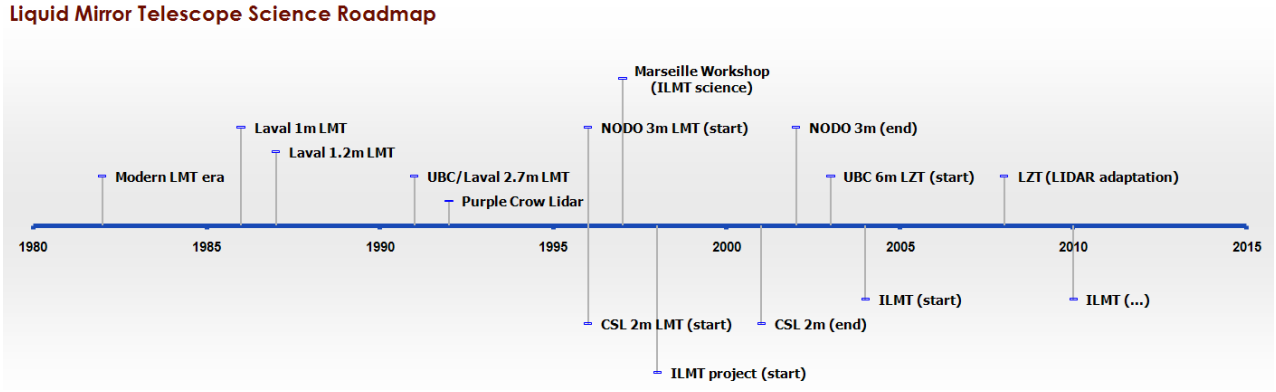


Figure 1.19: Time-line of the liquid mirror telescope projects around the world.

1.5.1 Specificities of observations with LMTs

Given their very low cost for large collecting apertures, LMTs can be dedicated to very specialized time-consuming projects such as variability surveys. They are optimal survey instruments able to provide valuable time-evolution information concerning celestial objects crossing their field of view. Such transit instruments are thus very well suited for a large number of projects involving the study of photometric and astrometric variability of the observed sources, like the study of galaxies, supernovae, quasars,... Several of these goals are considered in Borra (1982), especially

the study of cosmological phenomenon, supposed to be isotropically distributed and that would thus not suffer from exclusive zenithal observations.

The main limitation of an LMT comes from the relatively small region of the sky it can access. Indeed an LMT continuously monitors the same strip of sky, although it slightly varies from one night to another, due to the Earth motion. The zenithal observation implies several particularities. The integration time is determined by the time objects take to cross the CCD detector, but the data can be co-added from one night to the next ones to artificially increase this integration time as the same area is observed every night. However, in this case the variability information is lost.

The zenithal TDI observation mode also presents several advantages listed in Borra (1997). Flatfielding and defringing[3] are much more accurate because the images are actually formed by averaging over entire CCD columns. Atmospheric conditions are optimal at the zenith. Observing efficiency is high because there is no overhead from slewing the telescope, reading out the CCD, taking flatfielding frames, etc.

Tracking can only be performed in the TDI readout mode which limits the use of LMTs to imagers. However, using several filters can lead to low resolution spectroscopy. We will see that several LMTs have been used in that way.

1.5.2 The first liquid mirror telescopes

Since modern technology made the liquid mirror telescopes useful for astronomy, Borra's team is the first to have built and operated, mostly in the laboratory, such telescopes. Borra et al. (1985b) reports the construction of a 1m diameter f/4.7 prototype LMT that was used to obtain star trails on photographic plates. The image quality obtained with this telescope (2 arcsec) seemed to meet the optical requirements even if vibration induced waves diffracted about 30% of the light outside the central peak of the PSF. Those results lead them to build an improved 1.2m diameter f/4.58 LMT (Borra et al. 1988).

A few years later, a 2.7m LMT was jointly built by the Universities of British Columbia and Laval (Canada). Gibson and Hickson (1991) present the interest of a quasar survey carried out with this telescope. Low resolution spectroscopy has been carried out using 40 narrow-band filters. Hickson et al. (1994) report the successful construction and operation of this telescope. Its 2048 x 2048 CCD detector images a field of view of 0.5 degree in diameter. The 2 minute integration time provided images within a 20 arcmin wide strip of sky down to the limiting magnitude of 21 with a seeing limited resolution of 2 arcsec. The primary scientific program of this instrument was to obtain spectral energy distributions of all objects in the survey area, thanks to the 40 interference filters spanning the entire visible wavelength range $0.4 - 1.0\mu$m.

Science

The 1m class LMTs were used to obtain 300 hours of data on photographic film. The objectives were to search for rapid phenomena in the sky and to evaluate the potential of the LMT as a research instrument. The vibration induced wings of diffracted light were not detected in these data. Those experiments aimed at completing the optical shop tests performed on these mirrors, especially to detect perturbations on short time scales.

[3]Fringing is a phenomenon due to interferences occurring inside the CCD detector and resulting in a non-uniformity of the background of the images acquired with such a detector.

The first use of an LMT with a well defined scientific goal is presented in Content et al. (1989). They used some of the data previously obtained to search for optical flares and flashes. No events were observed and it was concluded that optical flashes and flares were rare events.

Apart from the few projects presented above, another 2.7m class LMT has been used to equip a LIDAR facility. A 2.65m liquid mirror based on the same design as the 2.7m mirror introduced before is used as a light collector in the Purple Crow Lidar facility in Delaware observatory (University of Western Ontario).

A LIDAR (LIght Detection and RAnging) is an atmosphere study facility that collects the light emitted by a powerful laser that is diffused by the atmospheric layers. The measurements allow air density, pressure, temperature and composition (for instance, water vapour) to be determined. Using liquid mirrors instead of classical ones allows to get very large inexpensive collectors which is an important performance parameter for a LIDAR. Details on the Purple Crow LIDAR and its science driven can be found in Sica et al. (1992).

1.5.3 The CSL 2m liquid mirror telescope

The Liège Space Center (CSL) used liquid mirrors for optical shop testing (Ninane and Jamar 1996): they can indeed be used as aspherical reference surfaces. Knowing the liquid mirror technology, it was decided that CSL would build a demonstrator telescope using such a mirror. The objective was to get images of the sky with a liquid mirror.

The telescope that was built is composed of the same type of elements as previously described for the ILMT (presented in fig. 1.20), except for the motor that is not coaxial with the bearing. The turntable and the motor are linked together by means of a belt. This system presents the advantages of isolating the dish from vibrations generated by the motor and smoothing the motor rotation. The mastering of the belt manufacturing and its use were thus required to operate the system. It is not so easy to build a belt that lasts long enough to properly operate the mirror.

A plastic tent has of course to be used to confine the mercury vapors in the vicinity of the mirror.

A CCD camera and an optical corrector were specifically designed and constructed. The detector was installed in the focal plane of the mirror, 7m above the latter one, and roughly aligned. The corrector has never been installed. The images taken look like those presented in fig. 1.21.

These images were considered as sufficient to demonstrate that building a telescope based on a liquid mirror was feasible. The demonstration objective was thus reached and the project stopped because of the lack of interest in improving a zenithal telescope in Liège, where the quality of the sky is quite poor and the cloud cover important. A final report (Ninane 2001) presents the details of the project and the technical characteristics of the Liège Space Center 2m Liquid Mirror Telescope are summarized in Table 1.5.

During the development phase of the ILMT project, we thought about using the 2m LMT as a testing tool to prepare the road for the ILMT. However, the abandoned LMT required refreshing and repairing. We were in charge of this work. Once the whole system had been fixed and the parts that were too old had replaced, the system was restarted.

We then worked at obtaining a stable mirror. This required the manufacturing of a good

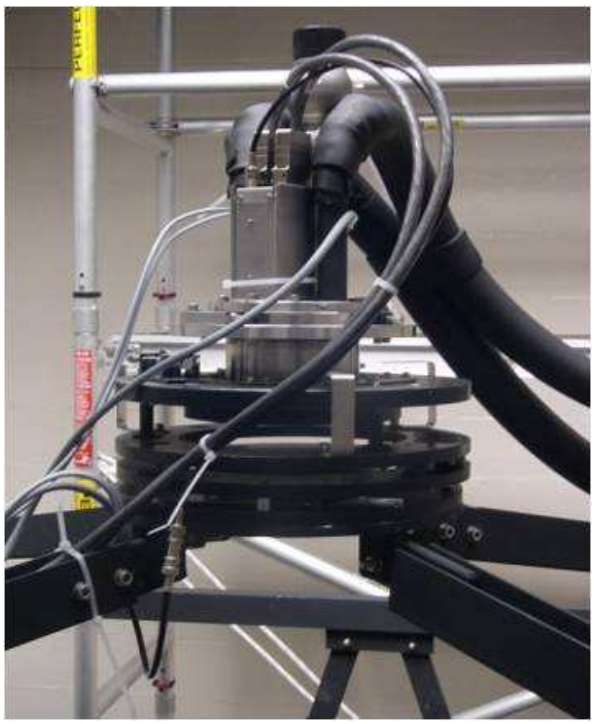


Figure 1.20: Top left: The bottom part of the 2m LMT structure that supports the focal instruments and the confinement tent. The turntable rests on the air-bearing that lays on a three point mount. The motor is not included in the bearing as it is the case for the ILMT. Instead, the angular momentum of the motor is transmitted by means of a belt. Bottom left: CCD camera of the 2m LMT with its alignment mechanisms, the corrector is not installed. Bottom right: The 2m LMT mirror during startup. The mercury is spread out in the dish by manually increasing and decreasing the rotation speed of the table.

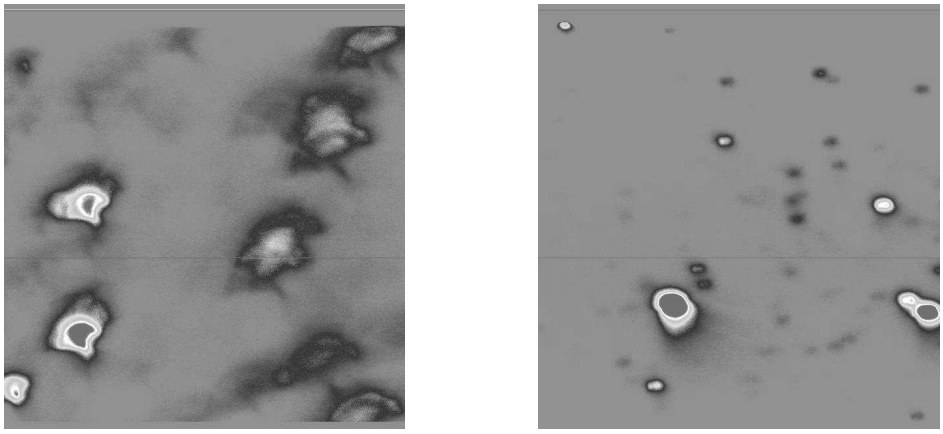


Figure 1.21: Images of stars taken by N. Ninane with the 2m LMT of CSL. The pictures correspond to the full field of view of the telescope, that is 14×14 arcmin. Two types of images are visible. Left: Large spots with an asymmetric shape. Right: Small and large spots with an approximately symmetric shape. Courtesy of N. Ninane.

Telescope	
Mirror Diameter	2m
Focal length	$\sim$ 7m
Rotation speed	$\sim$ 8rpm
Mercury volume	$\sim$ 10l (startup)
Thickness of Hg	1.6mm
CCD	
Number of pixels	2048×2048
Pixel size	15μm
Field of view	14 arcmin
Integration time	$\sim$ 71s
Location	Liège (Belgium)
Longitude	5°34'33.6" East
Latitude	50°35'56.4" North

Table 1.5: Characteristics of the CSL 2m LMT. This telescope aimed at demonstrating the feasibility of telescopes based on liquid mirrors. We had plans to use this telescope to test the methods that we developped (see text).

belt for the motor transmission and the monitoring of the rotation speed of the mirror. When the rotation aspect was acceptable, the mercury was loaded in the mirror recipient. However, once the mercury was on the turn-table, the belt suffered a lot and broke down more easily. The gluing of the belt had thus to be improved. We never managed to keep the mirror working for more than 24h (because of the belt problem). We thus never had the opportunity to obtain a thin mercury layer.

Unfortunately, several other troubles occurred with this LMT, especially with the compressed air system, that finally led to the end of the project. However, during these manipulations, the basics of liquid mirror operation have been reviewed and assimilated as well as the functioning of the camera and its cooling and vacuum system. Last but not least, using this liquid mirror introduced us to the safe manipulation of mercury.

The CCD camera has been brought to the Hololab laboratory in order to estimate its usability with the ILMT. These tests are presented in the second part of chapter 2.

Science

Apart from a few images of the sky of Liège, the 2m LMT of CSL did not fulfil scientific objectives. However, this prototype telescope aimed at demonstrating the possibility to use a liquid mirror as the primary mirror of a zenithal telescope.

1.5.4 NASA Orbital Debris Observatory - (NODO)

The NASA Orbital Debris Observatory (NODO), shown in fig. 1.22, is a three meter class liquid mirror telescope located near Cloudcroft in New Mexico. It is extensively described in Mulrooney's Ph.D. thesis (Mulrooney 2000). The observatory is located in a good astronomical site, far from cities and seismically stable. About 200 clear nights per year could be dedicated to observations.

Started in October 1996, the operation of this LMT ended in September 2002, but many of its components have been incorporated into the 6m Large Zenithal Telescope (LZT) that is described in the following section. During its operation, the NODO has demonstrated the interest of the Liquid Mirror Telescope as an astronomical tool. Its technical characteristics are presented in Table 1.6.

The NODO LMT was mainly dedicated to study the population distribution of orbiting space debris (Potter and Mulrooney 1997). Such surveys are performed using radio telescope arrays or small optical telescopes. However, these instruments are not able to observe debris smaller than 10cm and only in a Low or Medium Earth Orbit (LEO/MEO). There was a need to detect objects as small as 1cm even in Geostationary Earth Orbits (GEO), which required a larger dedicated optical telescope. However, the cost of a classical glass telescope is so high that it would be far too expensive to only use it for orbital debris observations. A convenient solution came from the fairly inexpensive liquid mirror telescope. The performances of the NODO allowed NASA astronomers to reach the 1cm objective detection.

Science

Although the main goal of the NODO was the survey of the space neighborhood of the Earth, it was also used to demonstrate the interest of using a 3m class liquid mirror telescope for scientific objectives. It has been used to perform several astronomical surveys such as the detection of Near-Earth Objects and studies of galaxies and QSOs at redshift smaller than 0.5 (Hickson and Mulrooney 1998). Several hundred thousands of objects have been measured with a set of 33 narrow-band filters spanning the optical spectrum.

The first preliminary work consisted in generating a catalog of the objects present in the 20 arcmin strip of sky observed with this telescope (Cabanac 1997). Using the data obtained, Cabanac et al. (1998) presented a search for objects showing peculiar spectral energy distributions.

Based on this preliminary catalog, a multi-narrowband survey of 10 000 galaxies and quasars was then performed using intermediate-bandwidth filters ranging from 455 to 948nm. This survey is described in Hickson and Mulrooney (1998). It was aimed at studying galaxy and

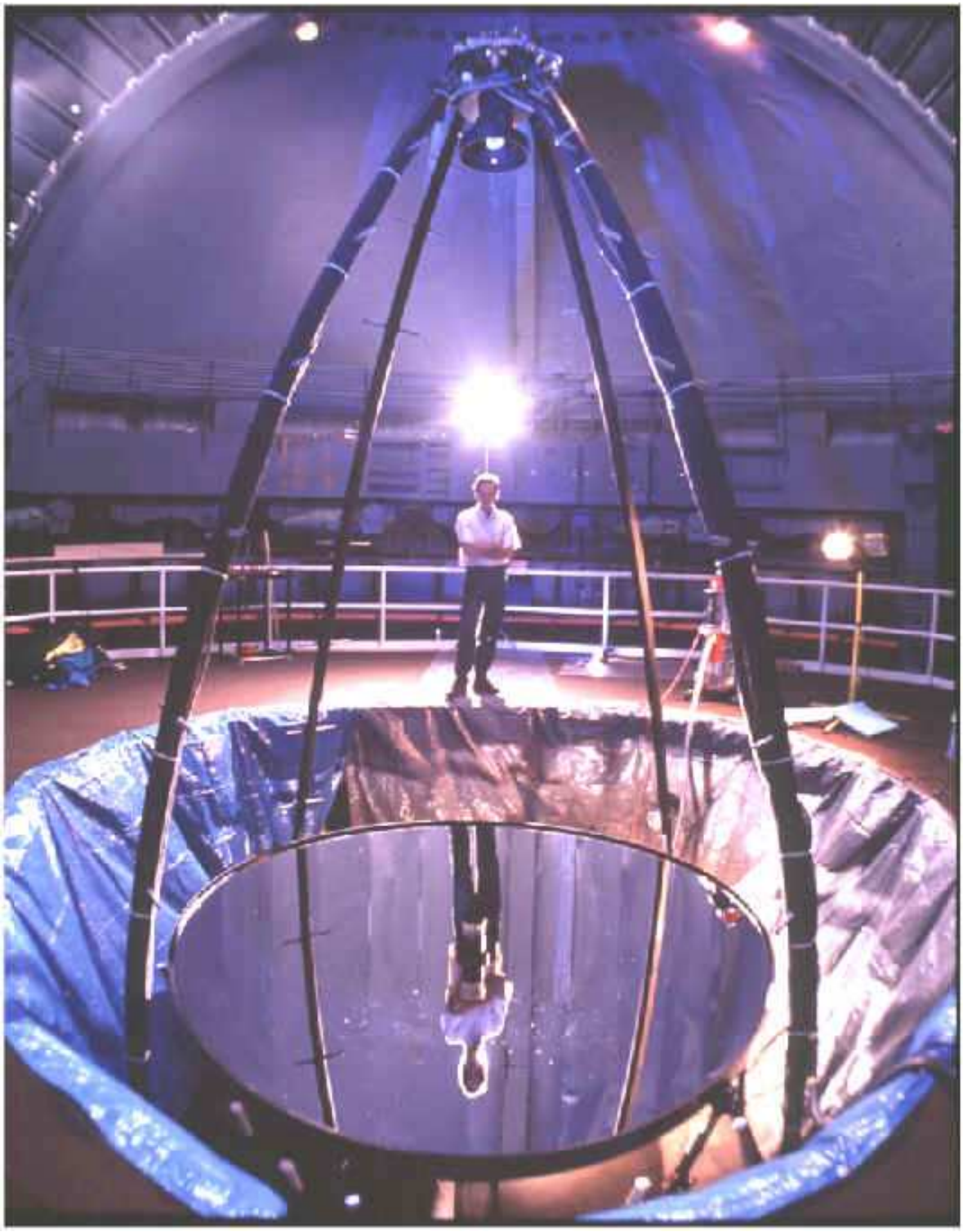


Figure 1.22: The NASA Orbital Debris Observatory telescope. Image from the NODO website.

QSO distributions, evolution and large-scale structure. It also provided photometry and spectral energy distribution for all objects in the strip.

Telescope	
Mirror Diameter	3m
Effective focal length	5.177m
Rotation speed	10rpm
Mercury volume	14l
Thickness of Hg	1.6mm
CCD 1	
Number of pixels	1024×1024
Pixel size	24μm
Field of view	0.27 deg (16 arcmin)
Integration time	78s
CCD 2	
Number of pixels	2048×2048
Pixel size	24μm
Field of view	0.54 deg (32.5 arcmin)
Integration time	156s
Location	Cloudcroft (New Mexico)
Longitude	105°42' 59" West
Latitude	32°58'43.5" North
Altitude	2772m

Table 1.6: Characteristics of the NODO. This telescope, dedicated to the observation of orbital debris, has been successively equipped with two CCD cameras.

1.5.5 The Large Zenithal Telescope - (LZT)

The Large Zenithal Telescope (LZT) project began in 1994 when the idea appeared that large liquid mirrors could be scientifically interesting. A new generation of 10m class LMTs was considered. However, such telescopes cannot be thought of as a simple scaling up of the existing 3m class telescopes. An intermediate step was required to pave the way.

A 6m class telescope, the LZT, has thus been built at the University of British Columbia (UBC) by Prof. P. Hickson. With its 6m, 3 ton mirror, the LZT is the largest liquid mirror ever built.

The main objective of this telescope was to improve the liquid mirror technology required for large mirrors. It had thus to be located in an easily accessible place. The site where the observatory was erected is located 50km away from UBC in the mountains to the North-East of Vancouver. This site has been monitored during about 20 years. From these observations 30 % of clear sky and a median seeing of 2.2" can be expected. Even if this is not a so-called "good" astronomical site, the accessibility and available infrastructure were very welcomed.

The LZT project is extensively described in Hickson et al. (1998) and Hickson and Racine (2007). An analysis of the image quality is presented in Hickson et al. (2007). Pictures of the telescope are presented in figs. 1.23 and 1.24. The technical characteristics are given in Table 1.7.

Telescope	
Mirror Diameter	6m
Effective focal length	10m
Rotation period	8.50812s
Thickness of Hg	1.2mm
Reflectivity ($0.6\mu m$)	77%
CCD	
Number of pixels	2048×2048
Pixel size	24μm
Field of view	23 arcmin
Integration time	100s
Location	Vancouver (Canada)
Longitude	49°17'17" West
Latitude	122°34'23" North
Altitude	395m
Median seeing	2.2 arcsec
Broadband limiting magnitude	25.4 (R)
Mediumband limiting magnitude	24.4 (750nm)

Table 1.7: Characteristics of the Large Zenithal Telescope.

Several differences exist between the designs of the LZT and the ILMT. The mirror especially required improvements. The dish is made of a shell of composite material, itself composed of seven hexagonal segments bold together. This shell is supported by a steel structure that provides stability and stiffness. The inner shape of the shell is a sphere of 18m curvature radius. The dish rests on a special air bearing specifically designed to support a 10 ton load.

The liquid mirror is protected from the wind (due to the rotation of the mirror) by a 12μm thick Mylar cover, which prevents wind induced waves to appear on the surface of the mercury. Indeed, these waves, that would diffract light out of the PSF, should be avoided, as it will be explained in chapter 3. However, the presence of the Mylar cover diffracts about 9% of the light in the halo surrounding the PSF. This amount can be reduced below 2% using thinner Mylar sheets (1.4μm) and an optimized supporting structure. Using this cover reduces the residual error on the surface of the mirror below 10nm ($\lambda/63$ rms).

Science

Completed in 2003, the LZT underwent engineering tests until 2005 when its regular operation began. Even if the site of the LZT is not well suited for astronomy, such a large telescope equipped with good instruments can have a very interesting scientific impact. Equipped with a four lens TDI corrector and a low noise 2048×2048 CCD camera, the LZT performs a deep multiband survey of a 17 arcmin wide strip of sky (Cabanac et al. 2002). The redshift and spectral energy distribution of more than one million galaxies and several thousands of quasars will be measured by photometry through 40 filters (one per night). Distant supernovae will also be detected.

Recently, the LZT has been transformed into a LIDAR in order to study the sodium layer of the upper atmosphere (Pfrommer et al. 2008). This layer is involved in the production of laser guide stars used for adaptive optics. These concepts are presented in chapter 4.

Figure 1.23: Left: The enclosure of the Large Zenithal Telescope located in the hills at the North of Vancouver. Right: Stored segments of mylar cover. These segments, holding 12μm mylar sheets, go on the liquid mirror to prevent spiral waves caused by the rotation wind. Mrs Suzanne Tremblay-Hickson is seen on the photograph.

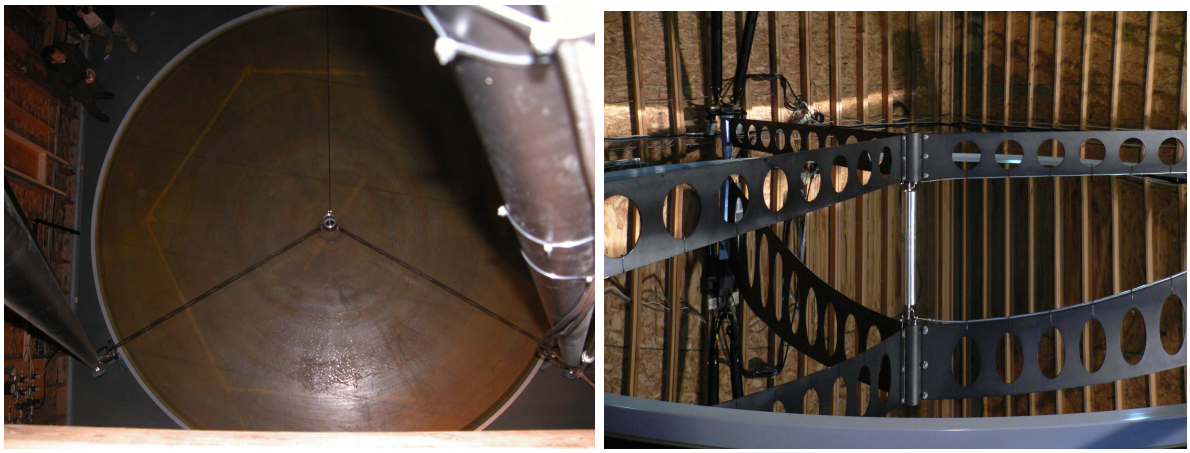


Figure 1.24: Left: The 6m mirror dish seem from the focal region. Side view of the LZT 6m liquid mirror while it is rotating.

1.5.6 The International Liquid Mirror Telescope (ILMT)

Borra (1997) reviewed the limitation and advantages of the zenithal observation and assessed the interest of building a 4m class liquid mirror telescope. Dedicating such an instrument to a full time project was an opportunity never encountered in astronomy. The idea of building an International Liquid Mirror Telescope was born. This project consists of a 4m spinning pool of mercury used as the primary mirror associated with a 4096×4096 pixel CCD camera installed at its focus and a five lens corrector designed to compensate for the field aberrations and the TDI distortions.

Scientific objectives of the ILMT

This telescope will be entirely dedicated to a photometric and astrometric variability survey of a 22 arcmin wide strip of sky reaching a magnitude of 22.5 in the i' band every night. This should allow to detect and study many new variable objects.

Several science objectives are described in Surdej et al. (2006). From the Nainital hills where the telescope will be installed, the ILMT will scan the Galaxy from the Northern galactic pole to the bulge and the survey will approximately cover 90 square degrees at high galactic latitude. This is particularly useful for extragalactic studies like gravitational lensing and supernovae.

Such a survey will also provide unique data for studies of the galactic objects, accurate measurements of stellar proper motions and trigonometric parallaxes useful for the detection of faint red, white and brown dwarfs and halo stars.

Gravitational lenses

The observational strategy for studies of gravitational lensing effects with the ILMT is described in Poels et al. (2001). It consists in surveying a sky area as deeply as possible to discover interesting targets and then to detect gravitational lens candidates among them.

Surdej and Claeskens (1997) and Jean et al. (1999) present estimations of the number of gravitational lensing effects that can be expected to be detected from a deep survey performed with a 4m LMT. Those simulations predict the possible detection of about 50 new multiply imaged quasars. Micro-lensing and weak-lensing effects are also investigated.

The daily monitoring of these lenses will permit to improve our knowledge of the parameters of those lenses (geometry, time delay, micro-lensing signature). An accurate modeling of those lenses will contribute to better understand the quasar source structure and the mass distribution in the deflecting galaxy, especially dark matter.

Supernovae

An estimate of the number of supernova events that could be detected in the strip of sky surveyed by the ILMT is addressed in Borra (2001). Such a monitoring will help discovering and observing several thousands of supernovae (z=1) per year. Obtaining the same results with classical telescopes would be far too expensive.

The interest of observing Type Ia supernovae[4] lies in the fact that they are bright standard candles that help measuring cosmological distances and that they are generally used for the determination of the Hubble constant (H_0).

[4]Supernovae are classified as a function of their spectra. Type "I" supernovae do not contain hydrogen lines in their spectrum whereas type "II" ones show such lines. Those types are also divided in sub-types. Subtype "a" contains silicon, sub-type "b" present an important quantity of helium whereas sub-type "c" is characterized by a small amount of helium.

Chapter 2

The Charge Coupled Device camera

An introduction to CCD technology has been given in section 1.3 of chapter 1. This chapter now details the specifications of the CCD camera that will equip the ILMT. We first present the selection of the chip based upon the characteristics we are interested in and then we introduce the specifications regarding the implementation of this chip into the camera.

These two steps are detailed here because we were in charge of collecting the information about the chips, that would help to make the final choice. We have also managed the contacts with the Liège Space Center and several CCD camera manufacturers like Apogee and Spectral Instruments. We first contacted Apogee but they did not make custom cameras any more, and they suggested to contact Spectral Instruments. After a series of mail exchanges aiming at determining whether Spectral Instruments would be able to provide the camera suited to the ILMT, it was decided to write a specification document that would describe all the requirements for the camera. We were in charge of writing this document, that has been used to define the work of Spectral Instruments.

The second part of this chapter presents the CCD camera that was installed on the CSL 2m LMT. This camera was planned to be used for testing the ILMT mirror before acceptance. We have thus performed several tests aimed at characterizing its imaging capability and its TDI behavior.

2.1 CCD chip selection

The main element of a camera is the CCD chip. It defines the field of view of the telescope, the sampling, the measurement dynamic, the noise level in the images, the spectral sensitivity and many other parameters of the instrument. Selecting the appropriate chip for our application is thus very important.

This section aims at presenting the characteristics of three CCD chips that were envisioned for the International Liquid Mirror Telescope. These chips had been preselected on the basis of their number and size of pixels. These values, combined with the focal length of the primary mirror, completely define the field of view of the telescope. All the chips presented hereafter have at least 4096×4096 pixels of 15μm. We will compare their advantages and drawbacks in order to decide which one is best suited.

2.1.1 Overview of the possible chips

Two models were first envisioned, an E2V and a Fairchild chip. Indicative quotes have been received directly from both manufacturers. We then added a third chip suggested by one of the camera manufacturers that we had contacted, Spectral Instruments (SI). It is another E2V chip. The preselected chips are:

- E2v CCD 231-84-1-E06 (85 k€)
- E2v CCD 230-84-1-E04 (no quote)
- Fairchild CCD 6161 ($\sim$ 75 k€)

These chips will be compared on the basis of several criteria, like their noise level or full well capacity, in order to make the best possible choice for the ILMT camera. The general characteristics of the chips are presented in Table 2.1.

	CCD 231-84	CCD 230-84	CCD 6161	Requirements
Number of pixels	4096×4112	4096×4112	4096×4096	4096×4096
Fill factor	100%	100%	100%	100%
Pixel size	15 μm	15 μm	15μm	12 μm
Flatness	$< 20\mu$m	$< 20\mu$m	20μm ($< 50\mu$m)[5]	$< \pm 15\mu$m
Illumination	BI	BI	BI	BI [6]
Pixel charge storage	350ke$^-$	150ke$^-$	90ke$^-$	100ke$^-$
Digitization	16 bit	16 bit	16 bit	16 bit

Table 2.1: Specifications of the chips that we are interested in. They are compared with general requirements suggested by the Liège Space Center that is in charge of the camera.

We see from Table 2.1 that several characteristics of the chips do not meet the suggested requirements. These are the pixel size, the flatness of the chips and their full well capacity. However, the initially decided pixel size was 15μm, and the corrector design is based on this value. This requirement is thus not a problem (confirmed by AMOS). Moreover, the impact of this size difference on the Full Width at Half Maximum (FWHM) of the Point Spread Function (PSF) is negligible. The contribution σ_p of the finite pixel size to the error budget can be computed from

$$\sigma_p = \frac{\theta_p}{\sqrt{12}} \tag{2.1}$$

where θ_p represents the angular size of the pixel. It is given by the ratio of the linear pixel size (15μm or 12μm) to the effective focal length of the telescope (9.4524m). The contribution to the FWHM in the case of a pixel size of 15μm is 3.7% whereas it is 2.4% in the 12μm case.

[6] FairChild specifies a typical value of 20μm and guarantees a value better than 50μm.

[6] BI stands for Back Illuminated. In the classical front illuminated CCD, the light that falls on the sensors has to go trough the electrode system that is used to generate the potential well and to transfer the charges. A fraction of the light is thus lost. In thinned back illuminated CCDs, the light falls on the other side of the chip, it thus does not cross the electronics. This ensures a higher quantum efficiency, especially for blue light.

As far as the flatness of the chips is concerned, a higher value corresponds to a less flat chip that results in a less good focusing. However, even if the E2V guarantees a better flatness (20 μm instead of 50 μm), both chips have about the same typical values (20 μm). It comes out that the flatness contribution to the total error is also negligible compared to the median seeing, it is given by

$$\sigma_f = \frac{D\Delta_f}{10\sqrt{2}F^2} \tag{2.2}$$

where D is the mirror diameter (4m), Δ_f is the focus variation related to the flatness of the chip (20μm P-V) and F is the effective focal length of the telescope (9.4524m). Using these values in equation 2.2 gives a contribution of about 0.1 arcsec to the FWHM, that may be neglected.

The full well capacity is the total number of electrons that can be stored in the potential well of each pixel. It determines the maximum number of photons that can be measured by a single pixel before it saturates. It thus also defines the dynamic range of the detector. This aspect of the chip will be considered in another section.

2.1.2 Quantum efficiency

The first criterion that we will use to compare the chips is their quantum efficiency (QE). It describes the response of the sensor at different wavelengths. The QE corresponds to the probability for a photon at a given wavelength to be absorbed by the pixel it falls on. A higher QE results in a more accurate detection and an increased sensitivity of the detector in the particular range of wavelength. It is important that the maximum sensitivity of the detector corresponds to the spectral bands that we are particularly interested in for the scientific objectives. These bands mainly correspond to the i' and r', g' bands as defined for the Sloan Digital Sky Survey (SDSS) in Fukugita et al. (1996). Other bands of the SDSS could also be interesting for us. These bands are defined hereafter.

Name of the band	Symbol	Central wavelength
Ultraviolet	u'	354.3nm
Green	g'	477.0nm
Red	r'	623.1nm
Near-infrared	i'	762.5nm
Infrared	z'	913.4nm

Table 2.2: Central wavelength of the spectral bands as defined by the Sloan Digital Sky Survey filters. The bands that most interest us are first i' and then g', r'. The bandwidth is about 150nm.

From these definitions, one can see that a high quantum efficiency between 600nm and 800nm is mandatory for the ILMT camera. The QE of the three chips as a function of the wavelength is plotted in figs. 2.1 to 2.3.

From the curves presented in fig. 2.2, it is clearly visible that the chip proposed by Spectral Instruments (E2V CCD 230-84) does not suit well our requirements. Indeed, its QE peak is located between the g' and r' bands. The QE at the center of the i' band is only about 85% and it quickly decreases with increasing wavelengths. As a high quantum efficiency is mandatory, this chip has been directly dismissed and will not be considered in the following comparisons.

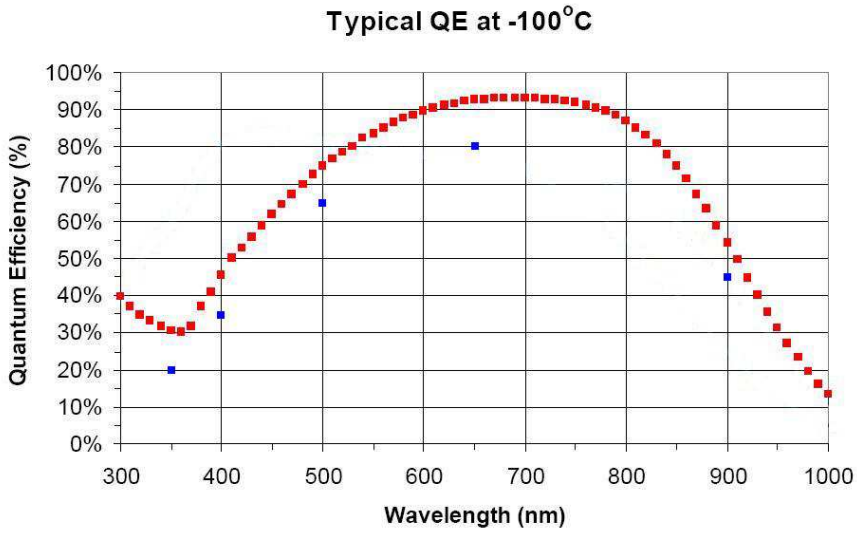


Figure 2.1: The red squares represent the typical quantum efficiency of the E2V ccd 231-84 for the coating that best fits our requirement. The blue squares (below the red curve) show the guaranteed minimum quantum efficiency. This graphic comes from the manufacturer datasheet.

The two other chips have a typical quantum efficiency larger than 90% between 600nm and 800nm. The Fairchild one presents a higher QE peak of about 98% near 750nm, which would be perfect for us.

On the other hand, the answer of the E2V chip is flatter over the r' and i' bands. This could be interesting but the E2V guaranteed minimum values are about 10% lower than the announced ones. This leads to a minimum guaranteed quantum efficiency in the spectral bands that we are interested in of about 80%. Fairchild does not specify their minimum values.

The Fairchild chip is thus more interesting as far as the quantum efficiency is concerned.

2.1.3 Noise level

The noise level caused by the CCD chip is another important criterion of comparison. The noise is a signal that is recorded by the sensor and that does not come from the source. The noise signal is thus not related to the light that falls on the CCD chip and disturbs the measurement of the real signal coming from the stars.

Two types of noise have to be considered as far as CCD chips are concerned. The readout noise corresponds to electrons created by the electronic device during the readout of the chip. It mainly depends on the readout frequency. The dark current corresponds to the thermal emission of electrons by the semiconducting material. It depends on the temperature of the chip and on the integration time. When the temperature is high, more electrons are thermally excited and captured in the pixel potential well. In order to reduce this type of noise, the cameras are generally cooled down. The noise characteristics of the CCD chips are presented in Table 2.3. It is directly seen from the datasheet values, that the E2V chip presents a smaller readout noise level than the Fairchild one.

As far as the dark current level is concerned, the specifications given in the manufacturer

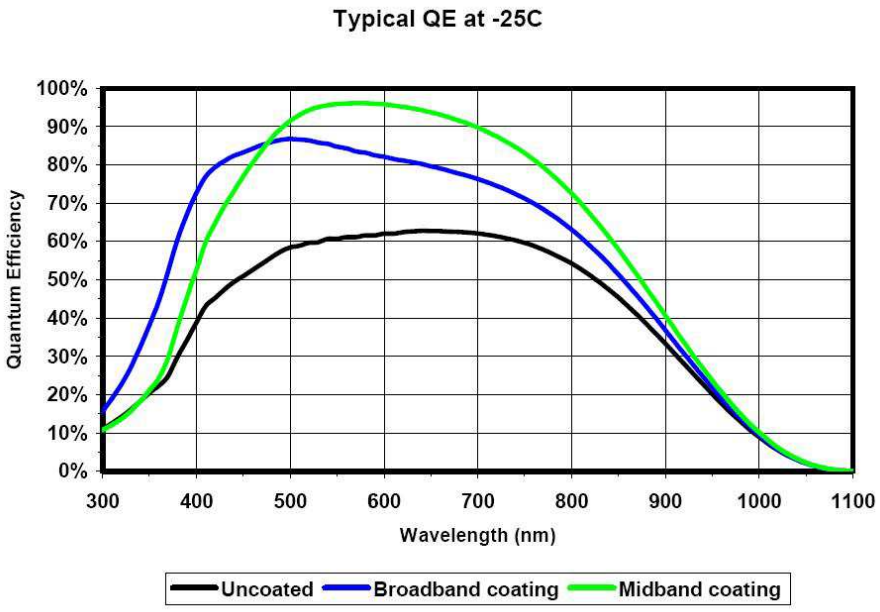


Figure 2.2: E2V CCD 230-84 quantum efficiency for different coatings. The QE in the range of wavelengths that we are interested in (600-800nm) is too low, the maximum being around 550nm. This graphic comes from the manufacturer datasheet.

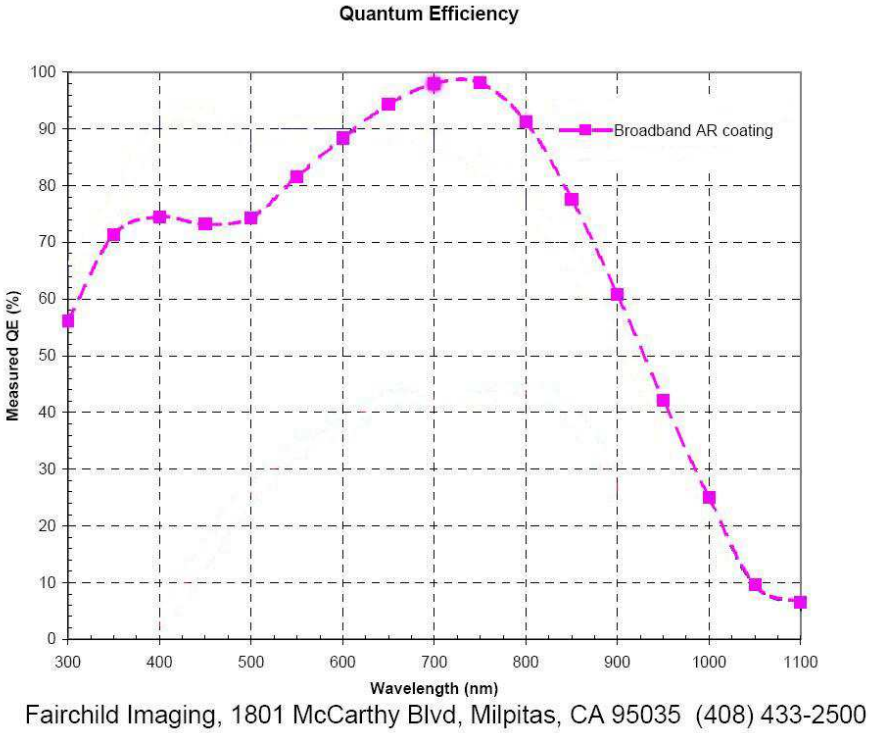


Figure 2.3: Fairchild CCD 6161 quantum efficiency (-60°C) for the coating of interest. The QE peak corresponds quite well to the center of the i' band. However, no minimum guaranteed QE is provided by Fairchild. This graphic comes from the manufacturer datasheet.

Noise type	CCD 231-84	CCD 6161	Requirements
Readout noise (e^- rms)	5 (1MHz)	12 (1MHz)	5
Readout noise (SI 100kHz)	$2e^-$	$3.5e^-$	$5e^-$ *rms*
Dark current ($e^-/pix/s$)	$0.83\,10^{-3}$	0.02	0.015
Dark Current (-100°C SI)	10^{-4}	10^{-4}	0.015

Table 2.3: Noise specifications of the envisioned CCD chips compared to CSL noise requirements. As the data concerning the CCD 231-84 are given at a temperature of -100°C whereas these regarding the CCD 6161 are given for -60°C, the noise values quoted by SI have been added to ease the comparison. The lines in normal format contain the information given in the manufacturer datasheets. The other lines, in italic format, present the information from the values quoted by Spectral Instrument.

datasheet are related to different temperatures for E2V and Fairchild. This makes it difficult to compare the noise level of both chips. A formula is given by E2V to estimate the dark noise level corresponding to a higher temperature (equation 2.3).

$$\frac{N_d}{N_{d0}} = 122T^3 \exp\left(-6400/T\right) \tag{2.3}$$

where N_d is the number of thermally generated electrons at temperature T and N_{d0} is the number of thermally generated electrons at the reference temperature (293K). However, the results of this formula are not consistent and they will thus not be considered. Indeed, the formula requires the knowledge of the noise level at a reference temperature, that is not given in the datasheets. However, two noise levels are given for two other different temperatures. Knowing them allows to compute N_{d0}, but using either one or the other values leads to very different results.

Instead, the dark currents quoted by Spectral Instruments are used to compare their chips, it is the same for both of them. The readout noise announced by SI for the E2V chip is still lower than the one of the Fairchild chip.

However, considering the small integration time of the ILMT (about 102s), the dark current levels are low enough for both sensors (E2V: 0.1 e-/pix and FC: 2.4 e-/pix) compared to the expected sky level and can thus be neglected.

The readout noises also seem negligible except in the U band where the sky level could be quite low and the corresponding shot noise would be of the same order of magnitude as the readout noise.

We conclude that, regarding the noise levels, the E2V chip thus seems preferable to the Fairchild one.

2.1.4 Full well capacity

As it has been previously exposed, the full well capacity of the chip defines the maximum number of electrons that can be stored in the potential well of each pixel. Two full well capacities are defined, one for the pixels and another one for the output register that is at least twice as large as the other one. The pixel full well capacity is the limitative parameter, that is considered hereafter. It is very important as it determines the dynamic of the sensor. Cameras using a chip

with a large full well capacity will allow to measure a larger range of magnitudes than a camera based on a chip with a smaller full well capacity.

The difference between the two chips, regarding the full well capacity is very important. Spectral Instruments may guarantee a full well capacity of $275ke^-$ for the E2V chip (goal 300ke-) but only $90ke^-$ for the FC (goal 120ke-).

It is also important to know that when the full well capacity of a pixel is exceeded, the pixel still generates photo-electrons, but these ones are not captured by the potential of the pixel and leak toward other pixels. This is the saturation phenomenon, which should be avoided as much as possible as it disturbs the measurements over a whole area of the sensor.

The number of stars which would saturate each chip should thus be estimated. Indeed, these stars will bleed along the rows in both directions, causing troubles in areas with a high stellar density.

A rough estimate of the saturation magnitude for the Fairchild CCD 6161 chip, as suggested by Paul Hickson (private communication), assumes that the central pixel contains about 10% of the flux, so the star image has about ten times as many electrons as the full well capacity allows. The number of photons per second, per square meter, per Angstrom is given by

$$n_{\text{phot}} = \frac{10 \times n_{fw}}{t \cdot BW \cdot A \cdot Q} \tag{2.4}$$

where n_{fw} is the full well capacity, t is the integration time, BW is is the filter Bandwidth in Angstroms (typically 1200 in the V band), A is the telescope area and Q is the product of the throughput of the telescope and of the detector quantum efficiency. This gives $n_{\text{phot}} \sim 1$ photon/s/A/m^2 which corresponds to a V magnitude of 17.5 for the ILMT case.

It is obviously highly preferable to use a chip that presents a high full well capacity. The E2V chip is then a better choice as far as this point is concerned.

2.1.5 Defect specifications

The CCD sensor quality is classified in grades, ranging from 0 (the best quality) to 5 (usable device). This quality concerns the number of bad pixels, which do not behave as expected. These pixels are either too luminous or too dark, meaning that they generate respectively too much or not enough electrons. A grade 0 detector presents very few defective pixels whereas a grade 5, even if it is functional, has plenty of bad pixels.

Grade 0 chips are very expensive and such a high quality is generally not useful. That is why we have chosen grade 1 chips for both manufacturers. However, important differences exist between their definitions of this grade. We present in Table 2.4 the specifications corresponding to the grade 1 of each manufacturer.

E2V and Fairchild do not use the same terminology to describe the defects of their chip, and, when the same terms are used for a particular defect, their definition is not the same. One should thus not only compare the numbers, but also the definitions of the defects. The latter are exposed hereafter.

	E2V	Fairchild
Column defect BW	10 (< 3)	5
White spots	800 (< 400)	
Total (BW) spots	1500 (< 750)	
Traps > $200e^-$	15 (< 10)	
Point defect		200
Cluster defect		25

Table 2.4: Grade 1 guaranteed maximum number of defects in the chips. Typical values are given in brackets. Both manufacturers use a different terminology to characterize their chip. It is thus difficult to compare them. BW stands for black and white.

E2V definitions

- *White spot:* a pixel whose dark generation rate is higher than $5e^-$/pix/sec at -100°C.
- *Black spot:* a pixel with photo response less than 50% of the local mean.
- *Column defect:* a column that contains at least 100 white or black single pixel defects.
- *Traps:* A trap causes charges to be temporarily held in a pixel and they are counted as defects if the quantity of trapped charges is larger than $200e^-$.

E2V grade 1 specification accepts 1500 pixels that generate more than 5 electrons per second at -100°C or that are 50% darker than their neighbors. These pixels will not be gathered in more than 10 columns containing 100 of them.

Fairchild definitions

- *Point Defect:* corresponds to two things.
  - *Dark pixel:* a pixel whose amplitude is below 50% of the mean signal
  - *Hot pixel:* a pixel generating more than 10 e-/pix/sec at -60°C
- *Cluster Defect:* a grouping of 2 to 9 adjacent point defects within a maximum area of 3 adjacent rows by 3 adjacent columns (excluding the pixels in defective columns).
- *Column Defect:* a grouping of more than 10 contiguous point defects in a single column, or a column which does not meet the minimum column transfer efficiency specification.

From these definitions, we deduce that the Fairchild chip will contain less than 200 pixels that generate more than 10 electrons per second at $-60°C$ or that are 50% darker than their neighbors. These pixels are mainly isolated, especially they cannot be gathered in more than 5 columns of 10 or more contiguous bad pixels and more than 25 groups of 9 or more adjacent (in the two directions) bad pixels.

Comparison

One directly sees that the black spot definition of E2V corresponds to the dark spot of Fairchild. However, even if the white spot definition of E2V is similar to the hot pixel definition of Fairchild, they are slightly different. Indeed, the allowed emission of electrons is twice as small (for the E2V) but at a lower temperature (-100°C instead of -60°C).

The E2V grade 1 thus accepts more white pixels but their definition of this defect is very strict. The total (BW) spot/ point defect, is thus hardly comparable.

Anyway, point defects are not really important in the TDI readout mode because the final image results from an averaging of all pixels in a column.

On the other hand, the definition of a bad column from Fairchild seems better than the one from E2V. Indeed they talk about 10 bad pixels in a bad column instead of 100 bad pixels for E2V. However the 10 pixels must be consecutive, which means that a column could have only one pixel over ten that is good and still be considered as "not bad".

The column defects can have important effects on the TDI readout mode and it is mandatory to minimize such defects. However, once again the definitions are hardly comparable and the values are of the same order of magnitude.

Because of the difficulty of comparing the defect definitions of both manufacturers, no conclusion can be drawn about the quality of each chip. They will be both considered as equivalent.

2.1.6 Conclusions: choice of the chip

To conclude, let us summarize the advantages and drawbacks of the CCD chips presented before. The Fairchild chip presents one main advantage; its higher quantum efficiency in the spectral band we are interested in. Its main disadvantage comes from its low full well capacity.

On the contrary, the main advantage of the E2V chip is its high full well capacity. Indeed, as previously written, the full well capacity of the detector is very important to minimize the risk of saturation. Another important advantage of this chip is its lower noise level, although, considering the small integration time (about 102s) of the ILMT, the noise level of both chips will be quite negligible.

Characteristic	E2V	FairChild
Flatness	✓	
Quantum efficiency		✓
Noise level	✓	
Full well capacity	✓	
Quality (defects)[7]	✓	✓
Total	4	2

Table 2.5: Summary of the comparison criteria. The E2V chips wins most of the comparisons, it is thus the best choice for the ILMT.

It comes out from this study that the E2V chip is probably a better choice for the ILMT. Indeed, even if its quantum efficiency is a bit lower that the one of the Fairchild chip, the much higher full well capacity is very interesting and important for us. The E2V manufacturer will be asked about the possibility to get a higher quantum efficiency and less column defects. The chip is illustrated in fig. 2.4

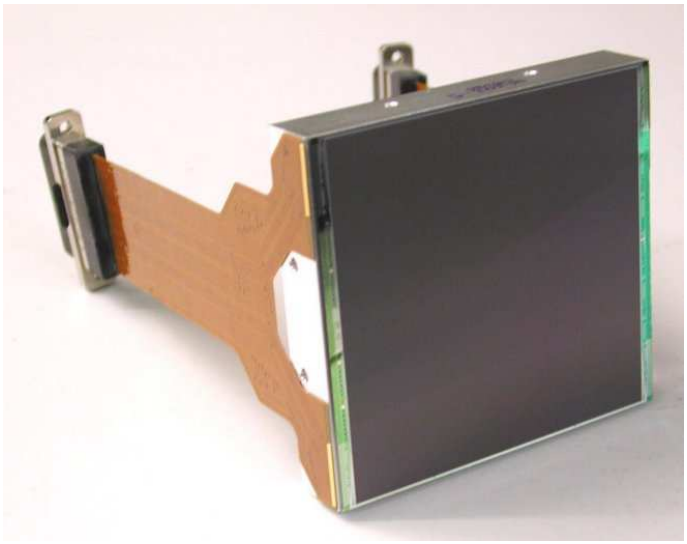


Figure 2.4: Illustration of the E2V231-84 chips

[7]It is not obvious which chip has the best quality, we thus check both chip in this table.

2.2 Technical specifications of the ILMT CCD camera

The ILMT will be equipped with a CCD camera based on the chip presented above. The camera and its equipment should be designed in order to provide the chip with its optimal conditions of functioning, especially a temperature that is sufficiently low to guarantee the requested level of dark current.

Moreover, the camera must include a time tag for each image row. It will be coupled with a GPS timer module that will generate these tags.

The camera will be installed in the focal plane of the liquid mirror telescope. The possibility of alignment of the camera with respect to the optical corrector, as well as along the east-west direction, is mandatory. The CCD camera will have to be aligned with respect to the optical corrector and the whole upper-end unit (detector + corrector) will have to be aligned with respect to the primary mirror. This last item is not part of the present specification. The alignment of the camera with respect to the corrector should be at least partially motorized (defined in a later section).

The Spectral Instruments' 1100S camera seems to be a good basis to fit our purposes. The supplier will be requested to quote for a custom version, including the above selected CCD chip, the following detailed specifications. The GPS module that will be best suited has still to be determined. The Liège Space Center (CSL) will take care of the mechanical parts of the global assembly and of the overall integration.

2.2.1 Overview of the 1100 series general characteristics

The 1100 Series CCD camera of Spectral Instruments (illustrated in fig. 2.5) can accommodate a wide variety of large scientific CCD chips. It is designed to minimize the readout and dark noise levels. Thanks to a cryocooling system, the CCD chip can be cooled down to $-100°C$ in order to ensure a dark current as low as $10^{-4}e^{-}$/pix/s.

The camera implements several readout modes with a frequency ranging from 50kHz to several megahertz depending on the chip. It can read the chip on up to four ports digitized on 16 bits to optimize the imaging speed with a readout noise as low as $3e^{-}$ rms. Specialized readout modes, such as TDI, are also possible. The electronics of the camera can be triggered either by an internal or by an external clock.

Connected to a computer through a fiber optic link, the camera is provided with a software interface which enables the drive of the camera and the reception of the acquired images.

2.2.2 Linearity of the imaging system

Each pixel of the CCD chip is made of a photoelectric semiconductor that emits a given number of electrons for each photon received. These electrons are then counted (by measuring the current they produce) to estimate the number of photons having reached the pixel. The relation between the number of photons and electrons is thus theoretically linear. However, it is not always true in practical systems, especially when the signal is very low (very few electrons) or near to the saturation limit (number of electrons near to the full well capacity).

The imaging system (CCD + electronics) designed by the camera provider should have a

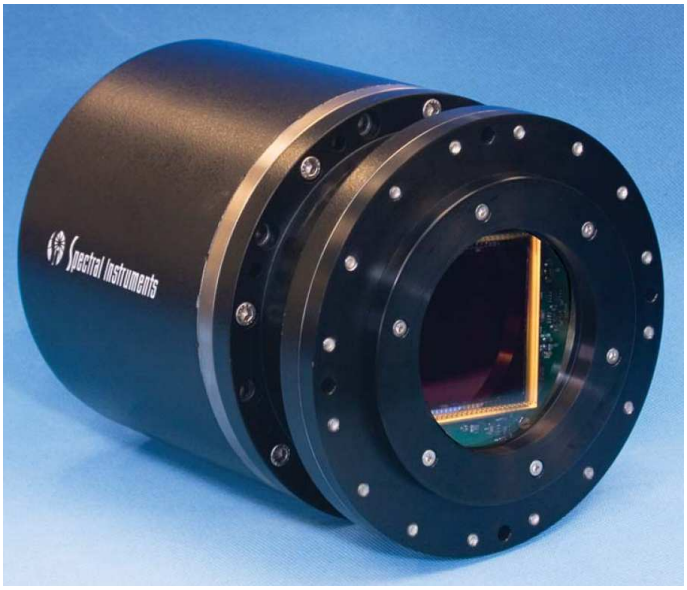


Figure 2.5: Picture of a 1100 Series Camera from Spectral Instruments.

linear response. Between $100e^-$ and 95% of the saturation value, the linearity should be better than 1% rms (cf. MegaCam specs in Borgeaud et al. 2000).

2.2.3 Chip cooling system

The cooling system is part of the camera that will be delivered by the camera provider. It should be sufficient to ensure that the dark current level respects the specifications of the CCD. A thermoelectric cooling system (Peltier) will be insufficient to reach this goal. For example, Spectral Instruments announced $2e^-$/pix/sec at -60°C with a Peltier system. This is far too much even considering the small integration time of the ILMT. A Cryo-tiger cooling system should thus be used (i.e. SI can guarantee $5 \cdot 10^{-4} e^-$/pix/sec at -100°C with this system).

Moreover, the cooling system (pump, compressor, and any moving element) should transmit NO vibration to the camera head and to the liquid mirror. Indeed, such vibrations would create waves on the mirror that are unacceptable. This condition seems realistic provided that the compressor is located far enough from the liquid mirror (in the camera supply unit). The pipes that transport the cooling fluid should thus be longer than 30m.

The total heat generation, from the camera and its components, around the focus should not exceed 100W. This budget does not include the heat dissipated by the alignment motors.

2.2.4 Vacuum

It is important that the CCD chip be housed in vacuum, at least during its operation. This aims at avoiding the deposition of condensates on the chip and to facilitate its cooling. However, having a vacuum pump running during the whole night should be avoided as it would be an undesirable source of heat. Hence, sealed vacuum is much more preferable.

The CCD camera manufacturer should thus provide a camera that houses the CCD chip in a

sealed vacuum ($< 10^{-2}$ mbar) chamber. This vacuum should not have to be refreshed more than once a year on site. Sending the camera back to the provider for more important maintenance could be tolerated, provided that it does not occur more often than once every 5 years.

No vacuum pump is expected to be provided by the CCD camera provider.

2.2.5 Readout mode

The CCD camera manufacturer should provide a camera that is able to read in both TDI and normal modes. The TDI readout mode will be used during normal observations whereas the regular readout mode will be used during the testing phases.

While operating in classical mode, the chip will be read on four ports with a 16-bit digitization. In this mode, the readout of the chip should be fast enough to ensure that no acquisition occurs during the readout. Indeed, no mechanical shutter will obscure the CCD during the readout, an "electronic" shutter is thus required. It simply consists in reading the chip fast enough so that no significant acquisition can occur during this period.

As far as the TDI readout mode is concerned, the chip will be read on one single port still with a 16-bit digitization. This mode is used to artificially track the stars crossing the field of view of the telescope. The drift rate of the rows have to be fine tuned to exactly match the apparent angular speed of the stars. In the ILMT case, they take about 102s to cross the entire chip. We should thus have a drift rate of 4096 rows/102s, corresponding to ~ 40 rows/s or to one row every 24.9487ms. An error on the drift period of 1μs would correspond to a shift of about 0.164 pixel at the end of the integration which is acceptable.

The drift rate should be adjustable between 1 and 100 rows per second, the TDI period should thus be adjustable from 10ms to 1s with a step smaller than $1\mu s$. Moreover, it has to be stable enough so that it does not vary during the acquisition. Its stability should be better than 10^{-5} relatively to the effective TDI speed.

The clock used to drive the camera (internal or external) should not drift over long periods. A drift smaller than $< 10^{-8}$s/s is allowed for every one hour. An internal 1 ppm thermally stabilized oscillator would be the best solution for us.

2.2.6 Time tagging

Importance of the time tagging

In order to achieve our scientific objectives, we need to identify each image row with a "time tag" that is the Universal Time (UT) corresponding to the time of the readout of that particular row. This time information will help us to precisely compare the data acquired during different nights of observation (i.e. subtraction and/or co-addition). These operations require that time information is associated to each row so that any subsequent image handling can be accurate. The accuracy of the time information should be better than 1 ms.

Time source

An external GPS will be coupled to the camera in order to generate the time information. To ensure the synchronization, the camera will send a pulse to the GPS at every TDI shift being applied to the CCD chip. The GPS will answer this pulse by generating and sending a time tag.

The typical time (based on the Spectracom EC31M GPS) needed by the GPS is about 55 μs to be ready to send the requested time information (once the pulse has been received) and about 16 ms to send the message through a serial port (RS232) at 19200 bauds. The shift time delay is about 25 ms; the time information should thus be available before the next image row is transferred.

Depending on the recombination solution chosen amongst those proposed below, a GPS module other than the EC31M could be used.

Time tag characteristics

The time tags will consist in a set of pixels coded on 16 bits, as the image pixels. The type of coding of these data is still to be determined in accordance with the GPS manufacturer, the camera provider and the customer. Many GPS manufacturers use proprietary formats that mostly consist in 31 ASCII (8-bits) characters.

Recombination

Hereafter, we propose (in order of preference) three different ways of using the time tags generated by the GPS. The first one consists of electronic recombination directly inside the camera. It is the most efficient and the most secure for us. The time and image information are combined before entering the computer, and no real-time issues may occur.

If such a solution is too difficult to be implemented (this decision should be taken in accordance with the customer), a second approach based on computer recombination can be envisioned. However, this one could cause some possible "real-time" processing issues because of the operating system I/O driver of the computer used to recombine the data. This solution still presents a high level of security as every row is tagged with time information.

The third, less desirable, solution should not be envisioned unless the two other ones are really not possible to be implemented and provided that the camera internal oscillator accuracy and stability can be guaranteed to the highest level. This solution consists in generating a time tag only for the first row and to rely on the camera master clock stability to compute the time information related to the following rows.

In case the solution 2 or 3 would be chosen, the Spectracom GPS module EC31M would not be optimal. A PCI GPS module would be better in order to ease the communication between the computer and the GPS.

Solution 1: Electronic recombination

Even if a computer recombination could be considered in the case it would be the only viable option, an electronic recombination would be more preferable as it can occur in real-time.

A solution should thus be found in order to combine the time tags with the corresponding image rows electronically inside the camera. This implies that the camera is equipped with

- a compatible input (to receive the time tag from the GPS)
- two buffers of the size of the time tag (32 bytes = 16 pixels)
- two buffers of the size of one image row (4096 pixels)

The time tag will be acquired by the camera through the compatible input. It will then be stored in the tag buffer until the acquisition of the image row is finished. Once the image row reaches the output buffer, it is stored in the image buffer until the time data is ready. When both waiting buffers are ready, the pieces of information they contain are combined by a dedicated FPGA[8] and sent to the computer as a single row of data that contains 4112 pixels = 4096 pixels of image + 16 pixels of time information.

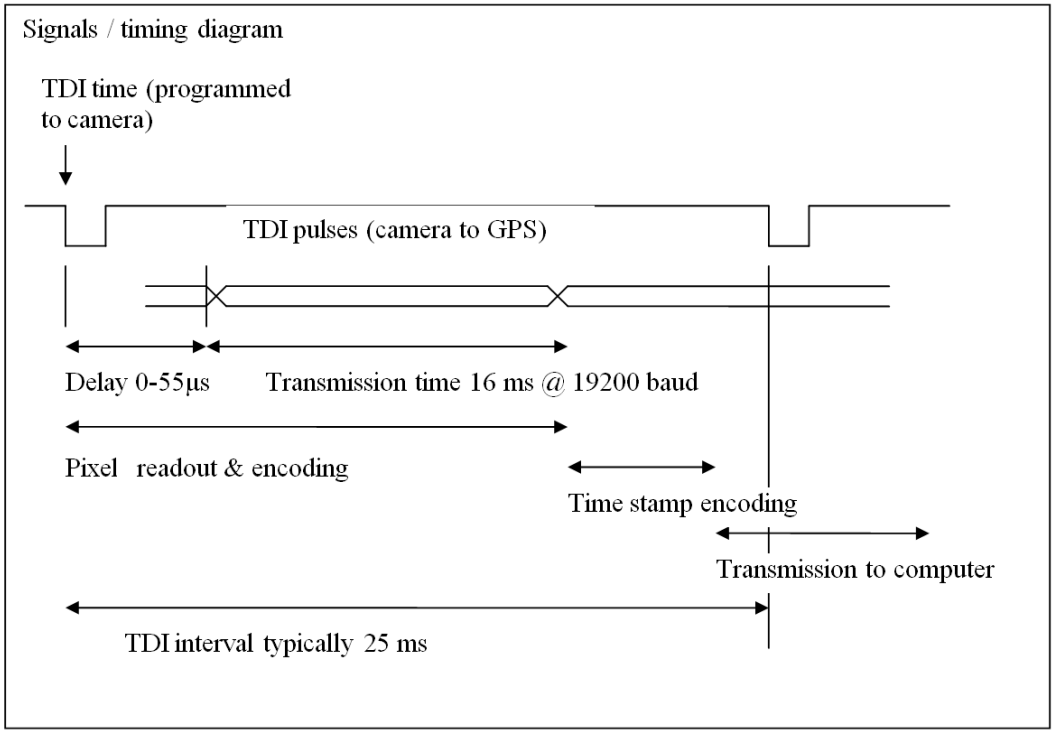


Figure 2.6: Timing diagram for the time tagging (based on the Spectracom GPS). It should be reviewed by the CCD camera provider in case the electronic recombination of the time tag is chosen.

The process starts again at each TDI shift. The image and time buffers ensure that whatever information arrives first, it can wait for the other one to be available so that the recombination occur in any case. Each buffer is double in order to ensure continuity in the acquisition. In case the transmission of the previous data is not finished, the following information can still be stored in the second buffer before being recombined and sent to the computer.

A timing diagram is presented in fig. 2.6. This timing should be agreed by the CCD camera provider in case the electronic recombination solution is chosen.

[8]FPGA: Field Programmable Gate Array: it is an electronic chip that can be programmed to perform a particular task.

Solution 2: Alternative computer recombination

The second possible solution consists in acquiring the time and image data separately and recombining them using a software application. Even if this solution is less attractive for us than the previous one, we present here an approach that could be satisfactory.

First, the camera receives, from its driving software, an order to send N image rows acquired in the TDI mode. The software then gathers these N rows in an image file at the transmission rate of the camera. Synchronized with each row (within the stability of the internal oscillator of the camera), a pulse is sent by the camera to the GPS.

The GPS then receives these pulses from the camera and sends the corresponding time messages to the computer (through the serial port in case of the EC31M GPS, another type of port would probably be better). The N messages corresponding to the N pulses related to the N TDI shifts are collected in another file.

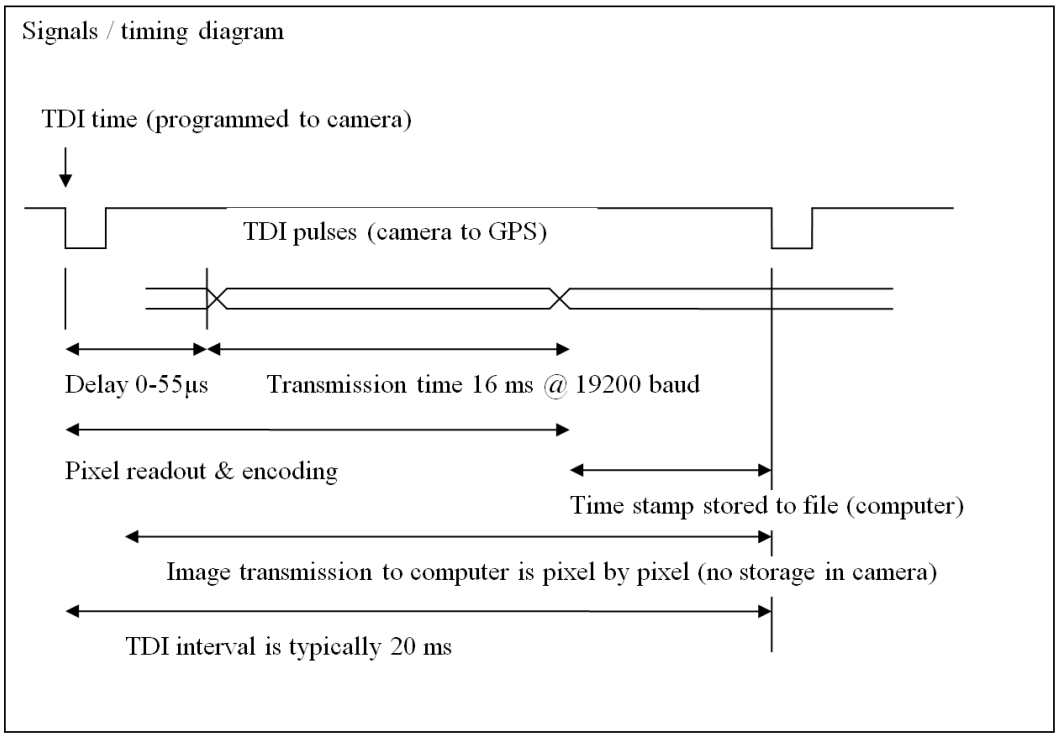


Figure 2.7: Timing diagram for the time tagging corresponding to the second solution.

Once the total image acquisition is finished, the software combines the N image rows and the corresponding N time tags contained in the two files. If the TDI shift is very accurate and a very small number of tags are lost, it is then possible to detect and compensate for these errors and to get a precise time tag for each image row.

The process begins again with the following groups of N rows. As in the previous case, two files of each type should be used in parallel to ensure the continuity of the acquisition. The new pieces of information are gathered in one single file while the previous information contained in the other file is combined with the time tag.

Solution 3: Alternative solution without recombination

The following alternative solution might only be used provided that the camera manufacturer can guarantee the accuracy and stability of the internal oscillator of the camera so that the drift after 24h is smaller than 100ms, which corresponds to 4 rows (see oscillator specifications).

In this case, only the first column would have to be tagged. The time information related to the following rows would be computed from the first one.

As in the previous cases, the camera sends a pulse to the GPS at every TDI shift. The GPS generates a time for each pulse it receives and sends it to the computer (the best solution would be to use a PCI-card GPS so that the link between the GPS and the computer would be easier). The first time tag generated by the GPS corresponds to the first image row. The time tags corresponding to the following rows are computed based on the known number of TDI shifts and on the time of the first row.

The GPS still generates the time related to each TDI pulse so that it is possible to crosscheck the computed time and GPS time in order to detect any drift. This verification does not have to be performed in real time and is thus easier than direct recombination.

This last alternative should only be used if the two first solutions are not possible.

2.2.7 Electronic interface

The camera shall provide the following interfaces:

- *Time tagging output:* negative pulse TTL 50 Ohm, falling edge coincident with the row shift operation at the TDI rate +/- 1 master clock period, duration TBD
- *Remote control:* manufacturer specifications
- *Pixel output:* manufacturer specifications
- *Others:* any other interface needed for the camera operation, monitoring and maintenance

2.2.8 Camera sizes

The allowed volume and weight (table 2.6) at the top of the structure is limited by the available space below the building roof and by the elements that are already designed, such as the corrector and its mechanical alignment system. These size requirements concern only the camera head. The cryo-tiger, power supply, and other "accessories" will be located on the floor and do not have to fit in these dimensions. The alignment mechanisms are not supposed to fit into these requirements, they only concern the camera.

Diameter	$200mm$
Length[9]	$400mm$
Weight	$< 6kg$

Table 2.6: Camera sizes

[9]This distance includes the connectors (cryo-lines, ...) located on the top of the camera head.

2.2.9 Environmental conditions

The camera will have to resist the environmental conditions of the site where the telescope will be installed. These are given in table 2.7.

Atmospheric air pressure	750 ± 100 mbar
Operating temperature	-10°C to +25°C
Survival temperature	-25°C to +30°C
Relative humidity	up to 100% (non condensable)

Table 2.7: Environmental conditions

2.2.10 Mechanical interface and constraints

The mechanical issues concern the interface between the camera and the upper-end unit, the alignment mechanism and the space available before the focal plane that should contain the filter slide system and the entrance window of the camera.

Definition

The mechanical interface is the part of the focal plane unit that supports the camera, allows alignment motions, provides a filter/shutter unit and attaches to the telescope upper end.

A drawing of the mechanical interface will be included in the specifications (by CSL). The front end of the camera should be compatible with this interface. A drawing of the allowed volume for the camera and its alignment mechanism will also be provided. It mainly concerns the volume of the mechanism.

Distance between the CCD and the entrance window

The distance available between the end of the optical corrector and the focal plane is 29mm. This spacing is very limitative as it has to contain many elements:

- the filter slide (to be designed),
- the mechanical shutter (combined with the filter slide),
- the entrance window of the camera,
- the front part of the vacuum chamber of the camera [10], which should be smaller than 7mm.

The total distance from the outside face of the entrance window to the CCD detector (including the window thickness) should be smaller than 12mm. This is compatible with the announced window thickness (4.7mm) and the minimum distance between the window and the CCD (7mm). This distance of 12mm is compatible with the addition of the filter slide. Fig. 2.8 presents a schematic of the distance available between the upper part of the corrector and the focal plane.

[10] This part extends from the entrance window of the camera to the CCD (the CCD has of course to be located in the focal plane).

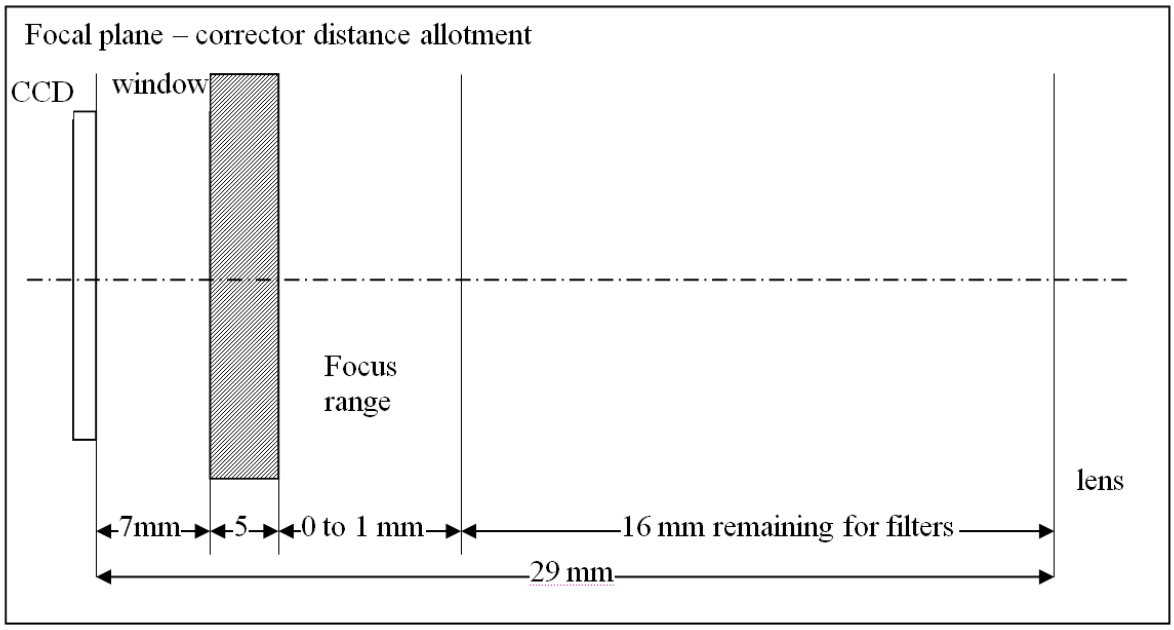


Figure 2.8: Available distance between the corrector and the focal plane.

Entrance window

The entrance window of the camera should be made of fused silica and be as thin as possible to maximize free space for the filter slide. Its optical characteristics and size requirements are given in table 2.8.

Polishing	$\lambda/4$ PTV
Scratch-Dig[11]	20-10
Diameter	4" ($\sim 102mm$)
Thickness	$< 5mm$
Clear aperture	(70x70mm)

Table 2.8: Entrance window specifications

Other mechanical constraints

The cables, pipes, and other connections delivered with the camera must be flexible, compatible with the alignment motions as specified in the following section. The camera shall be mounted with respect to the corrector such that the CCD columns are parallel with the apparent motion of the stars in the field of view, with a tolerance within the azimuth adjustment range. Marks shall be made on the camera body, mechanism unit and telescope upper end to ensure proper positioning of rotationally symmetrical parts.

[11]The scratch-dig represents the quality of the window polishing.

Alignment mechanisms

The alignment mechanisms aim at adjusting the position of the corrector and detector with respect to each other and with respect to the primary mirror. The relative positions of these elements have to be optimized in order to get the best possible image quality. The alignment mechanisms will be provided by the Liège Space Center.

Two stages of alignment are required:

- the alignment of the camera with respect to the corrector (part of the focal plane unit),
- the alignment of the upper end (camera + corrector) with respect to the liquid mirror (part of the telescope).

Tolerancing of the focus adjustment has been achieved (see appendix A) in order to estimate the impact of the focalization of the camera alone instead of the focal adjustment of the whole upper-end assembly (corrector + detector). It comes out of this analysis that it would be better to move the corrector and the detector together to adjust the focus. Two approaches can thus be considered. Either the camera is positioned once and for all with respect to the corrector, in the laboratory, and a high precision mechanical interface solidarizes them. Or an alignment mechanism is designed to precisely adjust the camera position with respect to the corrector. However the latter approach may turnout to be complicated as it would add some unnecessary degrees of freedom.

In case the second option is chosen, the azimuthal alignment and the focus control of the camera with respect to the corrector should be motorized and remotely controlled, as it will allow a more accurate positioning of the elements.

The camera alignment specifications are detailed in Table 2.9.

Motion	Range	Resolution/stability	Control	Notes
Focus	±1mm	50μm	motorized, remotely controlled	[12]
Azimuth	±2°	5 arcsec	motorized, remotely controlled	[13]
X,Y		100μm	manually	[14]
Tip, tilt		30 arcsec	manually	[14]

Table 2.9: Specifications for the alignment of the camera with respect to the corrector. Additional information is given as footnotes.

Filters and support, shutter (to be provided by CSL)

The science objectives require at least three filters (g', r', i') previously defined (reminded in Table 2.10). These filters do not have to be changed during the observation periods (night), but

[12]The focus adjustment does not move when it is not used or when it is turned on/off (non reversible drive or brake). No encoder is required (count steps). Mechanical end stops are required.

[13]The azimuthal adjustment is fixed when it is not used (locking screw or non reversible drive). No encoder is required (count steps). Mechanical end stops required.

[14]Intended for camera alignment with respect to the corrector, on an optical bench. It will not be used frequently and should be locked after adjustment.

it should be possible to change them between two successive nights. They should be installed in a filter slide that would allow to easily change the filter that is used for the observations.

Moreover, as the available space between the corrector and the detector (focal plane) is very small, the filter slide will also be used as a mechanical shutter that will protect the camera from very bright objects. The slide will thus contain at least four positions, one for each filter and a dark one, located between two other filters. The selection of the position should be motorized and remotely controlled.

Another slide will contain "technical" filters, which simply consist in glass plates with parallel faces. They will have different thicknesses so that they introduce different defocuses. They can be coated as the i' science filter, that corresponds to the optimal spectral range of the corrector and detector. The defocuses introduced by those "technical filters" are required to apply the Nijboer-Zernike phase retrieval method that will be used to measure the surface of the mirror as well as to perform the alignment of the upper-end unit. A fourth position with a dark filter (shutter) should also be provided in this slide.

Both slides should be interchangeable and do not have to be used at the same time. They both should be motorized in the same way. The slides and the scientific filters should be provided by CSL. The technical filters should be manufactured at the same time as the scientific ones in order to reduce the coating cost. Spare filters should also be provided.

The filter characteristics are gathered in the table hereafter, g', r', i' are the science filters and t1, t2, t3, are the technical filters.

Filters	g'	r'	i'	t1	t2	t3
Central wavelength (±2%) [nm]	477.0	623.1	762.5	762.5	762.5	762.5
Full Width at Half Maximum [nm]	140	140	150	150	150	150
Thickness [mm]	5	5	5	4.66	5	5.3394
Polishing (at 632nm)	$\lambda/4$					
Scratch-Dig	40-20 (goal 20-10)					
Diameter [mm]	100					
Substrate	BK7					

Table 2.10: Filter characteristics. The spectral band definitions are from Fukugita et al. (1996).

2.2.11 Package contents

All the items required for the operation of the camera have to be provided either by the camera provider (CP) or by CSL. Table 2.11 only contains the components to be delivered by the CCD camera provider.

Item	Provider
Camera	CP
Camera power supply	CP
Cables > 30m	CP
Optic fiber link to computer > 50m	CP
Cable for remote control > 50m	CP
Computer interface (PCI / PCIe card)	CP
Software compatible with Win XP/Vista	CP
Cryo-tiger	CP
Cooling line > 30m	CP
Cryo-tiger power supply/control unit	CP
Cable sensors > 30m	CP

Table 2.11: Package contents

2.3 The 2m-LMT CCD camera

The construction of the $4k \times 4k$ CCD camera presented above will require some time, and it is not expected to be available before summer 2010. The other parts of the ILMT, such as the liquid mirror could be ready before that time and it would be interesting to test them before their final acceptance.

Another camera is thus required to perform these validation tests, and the CSL 2m LMT camera might do the job. Reading in TDI mode, it has been designed to work with the 2m liquid mirrors. Moreover, the CSL LMT project having been given up, the $2k \times 2k$ camera can be borrowed to perform some tests of the ILMT elements.

Before using this camera as a measurement instrument, it was necessary to characterize it. After a short introduction of the characteristics of the CSL $2k \times 2k$ CCD camera, this section will present the tests that have been conducted on this device. They consisted in analyzing the quality of the images taken with the camera and in studying its TDI behavior. The images on which these analyses have been performed were obtained during laboratory sessions. The first part of the study, concerning the image quality, has been performed by François Finet (PhD student on the ILMT project in the EASO group). It is summarized hereafter.

As we will see at the end of this chapter, the $2k \times 2k$ CCD camera of CSL is not usable scientifically speaking.

2.3.1 Presentation of the camera

The camera has been built by the Liège Space Center to equip their 2m LMT presented in section 1.5.3 of chapter 1. It is based on an E2V chip, the *CCD42-40 Ceramic AIMO Back Illuminated* chip. The specifications of this chip are presented in Table 2.12, and the camera is presented in fig. 2.9.

In addition to the camera head that has to be located at the focus of the mirror, a vacuum pump, a cooling unit and a large electronic control unit compose the camera. The latter is connected to a computer that is supposed to control the data acquisition process.

[15]BI stands for Back Illuminated.

	CCD 42-40
Number of pixels	2048×2048
Field of view	$14' \times 14'$
Fill factor	100%
Pixel size	13.5μm
Flatness	(not specified)
Illumination	BI[15]
Pixel charge storage	$> 80ke^-$
Digitization	16 bits
Readout noise (e^- rms)	4.5
Dark current ($e^-/pix/s$)	250 (20°C)

Table 2.12: Specifications of the CCD used to build the CSL 2m LMT camera. These data come from the datasheet provided by E2V. The grade information is not known.

The CCD chamber is not in sealed vacuum, that is why the camera requires a vacuum pump to reach an internal pressure lower than $5 \cdot 10^{-3}$mbar. This pump has to run almost permanently while the camera is cooled down in order to avoid condensation on the CCD chip. In practical cases, we stopped the pump during the night as it became very hot when running during a few hours. Stopping the pump causes the pressure to increase from $5 \cdot 10^{-3}$mbar to ~ 0.5mbar, which is still less than one thousandth of the atmospheric pressure, which is still sufficient to ensure that no condensation occurs. The pump was started again before the beginning of the measurements.

The cooling system of the camera consists of a thermoelectric Peltier system, coupled with a chiller. The latter ensures the circulation and cooling of the coolant fluid that evacuates the heat from the hot plate of the Peltier. This system is able to cool the CCD ship down to -50°C. The cooling procedure is managed by a control unit and has to be sufficiently slow to avoid thermal shocks. First the coolant is cooled to -10°C, then the Peltier element is turned on. In case the initial temperature of the different elements is not suitable (fluid too cold, for example) a heating phase is added. The liquid tank is cooled or heated only by conduction and the temperature is measured at the bottom of the tank, which leads to very slow cooling/heating process. The heating is particularly slow, since the hot fluid is at the top of the tank and the temperature is measured at the bottom, it appears that when the probe measures an adequate temperature, most of the fluid is hotter than the measured temperature. When the circulation pump is activated, the fluid is mixed and its temperature becomes uniform. In order to avoid this over heating, it is appropriate to wait for the system to be in optimal conditions before turning on the cooling system. This means that when the cooling process is stopped, the system should rest for about half a day.

The complete cooling process takes about two hours when it is started in optimal conditions. This is why we decided to keep the cooling system running even when the camera was not used (between two test sessions). When more than one night was foreseen between two test sessions, both the cooling system and the vacuum pump were stopped.

Readout mode

The camera does not implement another readout mode than the TDI. Once started, the acquisition is done continuously. Even if it is possible to stop the drift of the rows to increase the acquisition time, the acquisition will continue during the readout. As the camera is not equipped

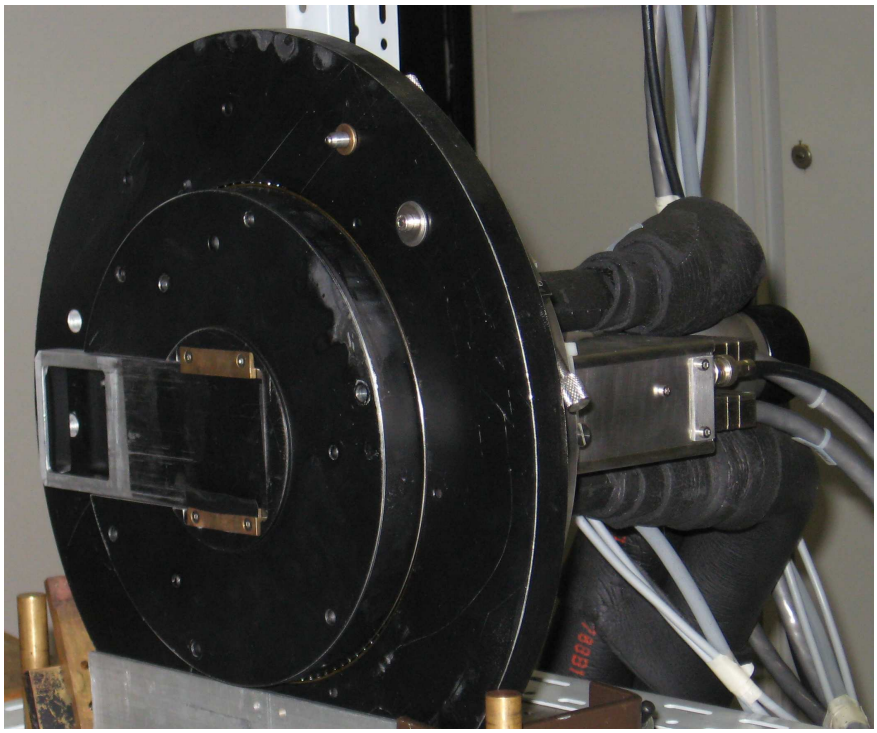


Figure 2.9: Picture of the CSL 2m LMT camera.

with a shutter, it is difficult to perform the tests that aim at characterizing the CCD chip.

In order to simulate a classical exposure, we let the readout running with the light source turned off in order to empty the charges that would have been stored in the pixels from the previous exposure. The drift is then stopped and the light source is turned on: the acquisition starts. After a given time, the source is turned off and the drift is started again to read the camera. The drift continues until the whole image has been read. However the beginning of the acquisition does not necessarily correspond to the first row of an image as it is randomly started (by the user). The 2048 exposed columns thus not necessarily belong to the same image: such a method is not practical as two images have generally to be combined in order to get the 2048×2048 exposed pixels.

Multi-Pinned Phase (MPP)

The CCD chip used in the camera has an inverted mode. This allows to use the camera in the Multi-Pinned Phase (MPP) mode. It is a CCD technology which significantly reduces the dark current generation rate, and that consists in doping the semiconductor that composes the CCD. This creates a potential barrier in each pixel, which allows charge integration (during the exposure) with a lower electrode potential.

The thermally excited electrons are not captured as the potential well is smaller, which presents the advantage of reducing the dark current. However, a lower potential well cannot store as many electrons as a higher one, the full well capacity of the chip is thus reduced.

2.3.2 Imaging capabilities

The images acquired with a CCD camera are generally used to make science. It is thus important that these images be reliable. This means that the images should contain information that mainly comes from the objects and not from other disturbing sources, as the camera.

Several tests have been conducted with the CSL camera in order to characterize it. These mainly consisted in flat-field and dark analysis. Hereafter, we present a summary of the results of these tests. A complete internal report has been written by François Finet.

Flat-field

A flat-field is an image acquired with a uniformly illuminated detector. It is used to determine the spatial variation of the response of the chip. The scientific images that are taken with the camera can be corrected for the non uniform response of the CCD simply by dividing it by the flat-field. This is a first characterization of the chip. The left part of fig. 2.10 shows a flat-field corresponding to an exposure of the CCD of four minutes under a uniform illumination. The right part of the same figure presents some horizontal sections of flat-fields corresponding to the different exposure times.

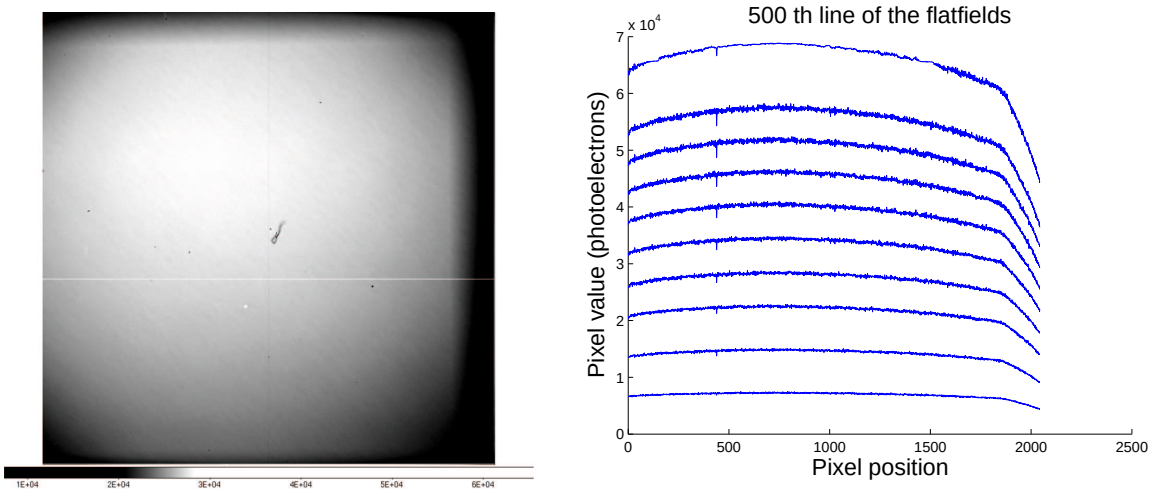


Figure 2.10: Left: 4 minute exposure flat-field. The upper edge points to the North direction and the TDI drift is done toward the left. The grey scale goes from 0 to 65535 ADU. Right: Horizontal section of several flat-fields corresponding to different exposure times. The spatial non uniformity of the chip is clearly visible. The sensitivity is better around the center of the CCD than near the edges.

The graphs show that the CCD response is not uniform at all. The signal varies by about 25% between the center of the image and its edges. Even, if the uniformity of the illumination could be partially faulty, it cannot explain such a high variation. Vignetting cannot be the cause of this non uniformity as the CCD is right behind the entrance window of the camera and nothing special obscures it.

Flat-fields can also be used to determine the response linearity the chip and its saturation limit (full well capacity). This is performed by acquiring a series of flat-fields with increasing integration time. The median intensity of each flat-field is then computed and the corresponding

values are plotted as a function of the integration time. Three curves, corresponding to different parts of the CCD are presented in fig. 2.11

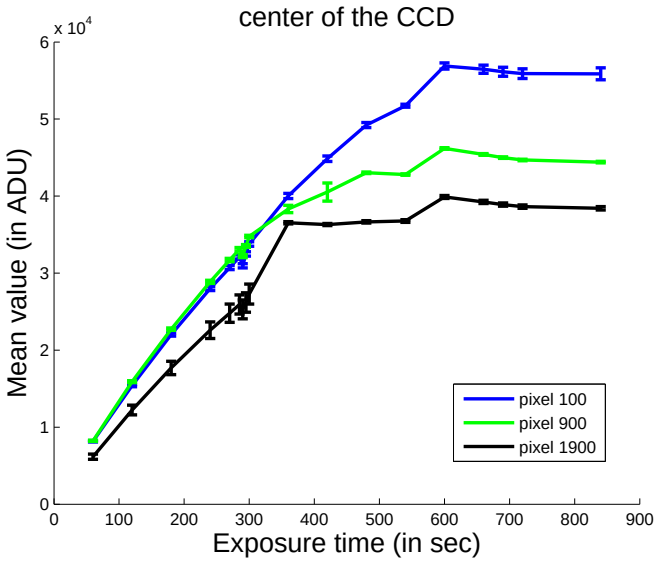


Figure 2.11: Curves showing the evolution of the mean flatfield values as a function of the exposure time in three different areas of the CCD. Those regions are located along the line 1024 of the CCD (see fig. 2.12 left).The linear behavior of the chip is clearly visible in the three regions. The saturation phenomenon also appears, but the limit is different in each case. The lower value is around 40k ADU and the higher one around 60k ADU.

These three curves clearly emphasize the linear behavior of the response of the chip (in each area) until the pixels reach their full-well capacity. However, the saturation limit is different in each region of the CCD. A mapping of the saturation has been achieved in order to estimate the variation of the full well capacity over the CCD. It is represented in fig. 2.12.

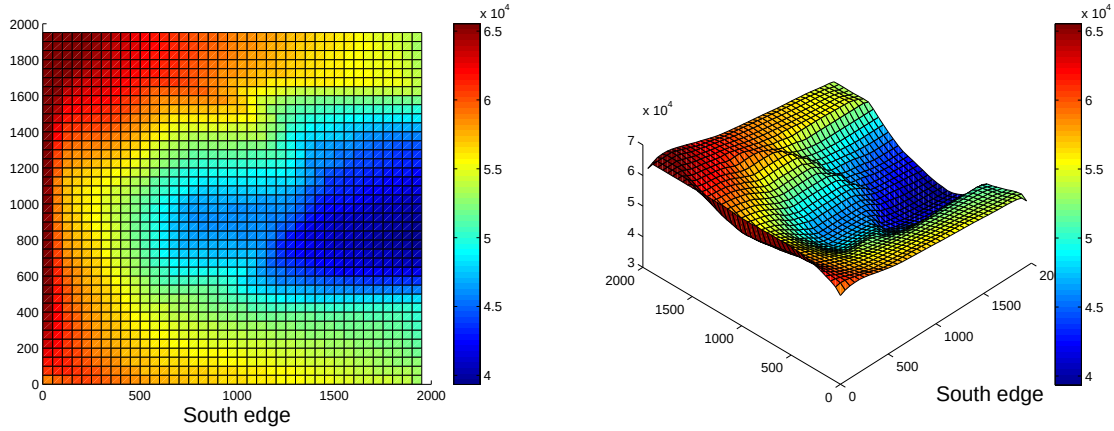


Figure 2.12: Mapping of the saturation values. The color bar is in ADU. The TDI drift is toward the left. The saturation value presents a local minimum around the center of the chip. The saturation values of the area located on the right are suspected to be affected by this local minimum because of the TDI readout mode.

The full well capacity presents a minimum around 40k ADU[16] near the center of the chip. Knowing that the gain of the output amplifier is 2, this minimum corresponds to $80ke^-$. This corresponds to the minimum full well capacity specified in the datasheet. The depletion of the full well capacity at the right of the CCD is suspected to be due to the readout. Indeed, the electrons corresponding to this part of the image have to cross the central region of the CCD during the readout of the chip. As the full-well capacity is smaller in that region, the exceeding electrons cannot be transferred. They thus stay in the right area of the CCD until the level of the signal gets sufficiently low to be under the saturation of the central area. Only at that moment they can cross this area and be read. They usually come with the next image that is thus disturbed. The leaking of electrons is clearly visible in the images where saturation has been reached as shown in fig. 2.13.

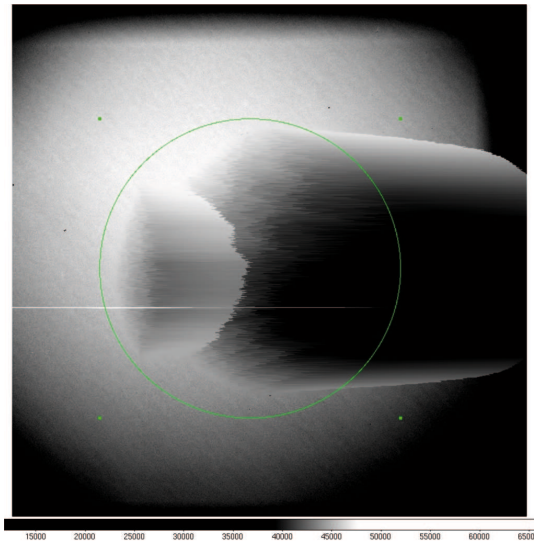


Figure 2.13: Saturated flat-field (exposure: 7 min). The leaking phenomenon related to the saturation is clearly visible. During the readout, the electrons related to the pixels at the left of the image cross a region of low saturation limit (the center of the image) that cannot accept them. They thus leak toward the nearby area of higher saturation limits.

Very few differences were noticed between the cases where the MPP mode was activated or not. Nevertheless, the saturation values corresponding to the MPP mode were, as expected, slightly smaller than those corresponding to the non-MPP mode.

The flat-field analysis also enables the study of the bad pixels (hot/cold spots). No particular result has been noticed, except that the very low sensitivity of the edges of the camera makes them appear as cold spots, especially at the right of the image.

Dark

A "dark" consists of an image taken with the CCD sensor occulted. As no light can reach it, the measured signal is only composed of the dark current and of the readout noise. Such dark images are acquired for several integration times in order to measure the mean thermal electron generation rate.

[16] ADU stands for analog to digital units. It is the basic unit of an analog to digital converter.

In order to suppress the contribution of the bulk of the readout noise, the signal corresponding to the readout alone (exposure time = 0.00s) is subtracted from the signal acquired with longer integration times. The median dark signal is then estimated and plotted as a function of the integration time (fig. 2.14).

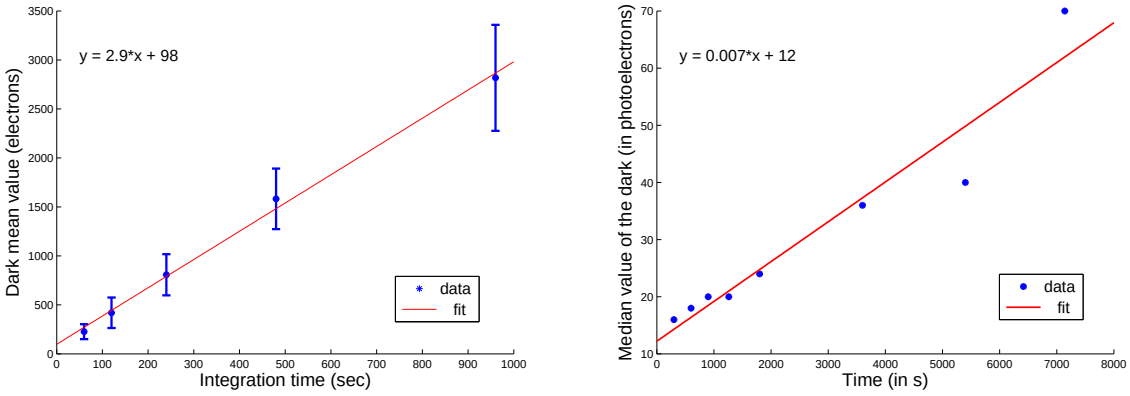


Figure 2.14: Mean dark signal as a function of the integration time. The slope in the linear fit corresponds to the number of electrons thermally generated in a pixel every second. The constant term is the addition of the bias and of the readout noise. Left: MMP mode disabled. Right: MPP mode activated. The values obtained in the MPP mode are in good agreement with the specifications. The linear fit gives a thermal electron generation rate of 0.007e^-/pix/s.

The thermal electron generation rate derived from the linear fit of the curve in the MPP mode (0.007e^-/pix/s) corresponds quite well to the specification of the chip given in the datasheet (typical dark current: 0.005e^-/pix/s; maximum 0.01e^-/pix/s). As far as the dark signal is concerned, an important difference is noticed when the MPP mode is activated or not. As expected, the dark current in the non-MPP mode is very high whereas it is quite low when the MPP mode is activated.

Summary

Even if the CSL camera respects most of its specifications, important non-uniformities in the response and in the saturation limits of the chip produce significant disturbances on the images. Furthermore although the non-uniformity of the response can theoretically be corrected for by dividing the images by the flat-field, its value ($\sim 25\%$) is found to be extremely large. In case of saturation, the perturbations can also affect the next observed image as electrons leak because of their travel trough pixels whose full well capacity varies significantly.

The dark signal study allows to emphasize the effect of the Multi-Pinned Phase mode. The important dark current measured in the non-MPP mode suggests to use the camera only in the MPP mode in which the dark signal is quite low.

2.3.3 Time Delay Integration (TDI) behavior testing

The Time Delay Integration readout mode, also called drift scan, allows to electronically track the stars crossing the CCD field of view. The photo-electrons produced are stepped along the column of the chip at the sidereal rate. In such a way, the electrons follow the image of the stars while they move over the CCD chip. This technique allows to integrate the flux of the stars while they are crossing the field of view of the camera.

It is obvious that the column drift rate must perfectly fit the sidereal rate in order to get the best possible image quality. It is thus very important to study the TDI behavior of the CCD camera. This implies to characterize the drift speed as well as its stability. We present hereafter a set of measurements that allowed us to determine these pieces of information.

Principle of the readout speed test

The drift speed measurement test is based on images of a flashing source. This flashing source has a known period that will be used as the time basis in the image drift analysis. This basis allows to measure the exposure time that corresponds to the reading of the entire chip. Indeed, as the flashes appear on the images, it is quite easy to estimate the total exposure time as it is related to the number of flash signatures contained in the images. This relation also depends on the flash frequency, this dependency will be expressed in the theoretical section hereafter.

Theory

The "flashing" signal is sampled in time by the readout of the CCD camera. The shifting of the rows translates the temporal flashing signal into a spatial signal. The sampling of this signal depends on the number of rows of the CCD as well as on the duration of the readout. The cutoff frequency of the acquisition system (the CCD camera) is given by:

$$f_{cutoff} = \frac{N_r}{2T_{read}} \tag{2.5}$$

where N_r is the number of rows in the CCD chip and T_{read} is the total readout time of the CCD. The 2048×2048 pixel CCD camera using a typical master clock frequency of 1MHz is read in about 71s which corresponds to a cutoff frequency of about 14.5 Hz.

According to the Nyquist-Shannon theorem, the minimum sampling of a function corresponds to two samples per period. The sample frequency is thus twice the cutoff frequency, so the column drift rate is given by:

$$f_{col} = 2f_{cutoff} = \frac{N_r}{T_{read}} \tag{2.6}$$

Our tests are aimed at verifying this.

The total readout time theoretically depends on the master clock frequency and on the number of pixels of the CCD. The CCD chip is supposed to have 2048×2048 pixels. However, the images we get from the camera contain 2168×2048 pixels. The rows have pixels that are not active, which means that they contain no photo-electron: they will be used to add information, time tag

for example. However these pixels have to be taken into account for the readout time computation as they also have to be read. The equation of the readout time is

$$T_{read} = \frac{16 \cdot N_p}{f} \tag{2.7}$$

where N_p is the number of pixels read (2168x2048) and f is the master clock frequency. The factor "16" comes from divisions of the clock frequency that are performed electronically. These divisions reduce the effective reading clock. The pixel readout frequency is only one sixteenth of the master clock frequency. As previously written, the total readout time of the CCD with a master clock frequency of 1MHz is about 71s. The adequation between this equation and the actual behavior of the camera will be tested hereafter.

Presentation of the test setup and procedure

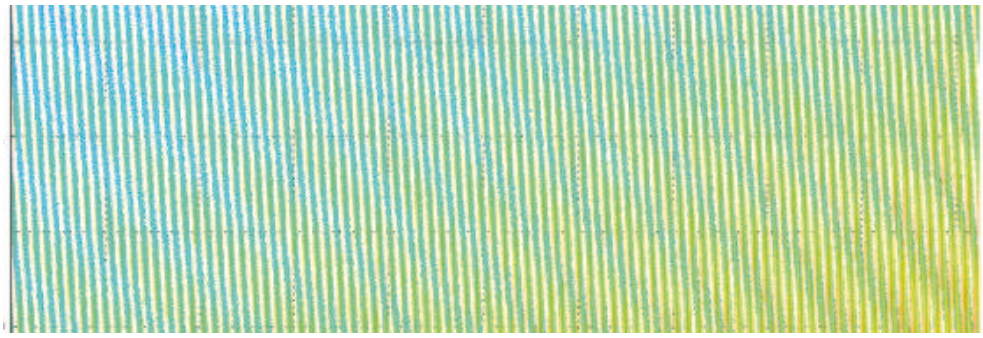


Figure 2.15: Image taken with the neon-ceiling lights of the room turned on. The vertical lines are interpreted as the signature of the 50Hz variation of the illuminating lights.

We first thought about this test by looking at images taken when the ceiling lights of the laboratory were turned on (fig. 2.15). The 50Hz variation of the light intensity is clearly detected on the images. However, it appears that for a typical readout time of about 70s, the cutoff frequency is much lower than 50Hz (see eq. 2.5). This bad sampling produces multiple aliasing which prevents us from getting interesting results from these images.

We then decided to use an adjustable light source consisting of the LEDs located on the back panel of a laptop. Their pulsation frequency and intensity are programmable with a software. These LEDs thus constitute a good light source for our tests. They were first programmed to emit a flash every second, but, a first analysis gave strange results. It appears that the programmed frequency of the flashes was not precise enough. Anyway, as the stability of the frequency seemed good, we decided to measure the effective frequency before doing the test again with the same source. The real frequency was in fact 0.89Hz instead of 1Hz. Entering this parameter in the analysis solved the problem and the results are presented hereafter.

We also used a diffusive screen, simply consisting of a white sheet of paper (fig. 2.16), intended to avoid direct lighting of the CCD camera. Indeed, the light emitted by the LEDs is too important and it would have saturated the sensor. Moreover, this screen aims at illuminating the camera as uniformly as possible, so that the impulse caused by the flashes can be distinguished.

It is important to clearly define the parameters of the runs in order to correctly interpret the results. The master clock frequencies applied are ranging from 0.5MHz to 1.5MHz with a step of

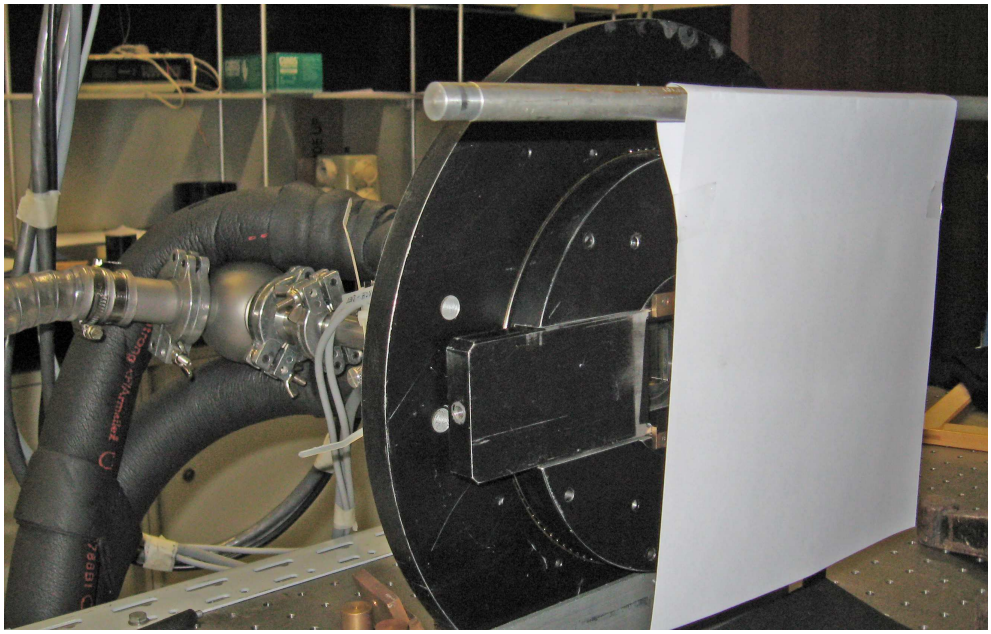


Figure 2.16: The CCD camera and the diffusing screen. The volume between the screen and the camera is covered with a black cloth to avoid parasitic light to enter the CCD.

0.1MHz. These frequencies correspond to cutoff frequencies quite larger than the source flashing frequency (see Table 2.13). The signal should then be correctly sampled and no aliasing problem should remain.

Master clock frequency (f) [MHz]	Cutoff frequency (f_{cutoff}) [Hz]	Readout time (T_{read}) [s]	Column drift rate (f_{col}) [col/s]
0.5	7.21	142.08	14.41
0.6	8.65	118.40	17.30
0.7	10.09	101.49	20.18
0.8	11.53	88.80	23.06
0.9	12.97	78.93	25.95
1.0	14.41	71.04	28.83
1.1	15.86	64.58	31.71
1.2	17.30	59.20	34.59
1.3	18.74	54.65	37.48
1.4	20.18	50.74	40.36
1.5	21.62	47.36	43.24

Table 2.13: This table shows the theoretical value of the cutoff frequency, the readout time and the drift rate for the $2k \times 2k$ CCD camera. The 0.89Hz frequency of the flashing LEDs is correctly sampled for each frequency of the master clock.

Several series of seven images were acquire for different frequencies of the camera master clock in order to study the effect of the readout frequency. The first two images are considered as transitions (readout speed not stabilized) that should not be used for the analysis. Indeed, the master clock frequency is changed during the acquisition of the first image of the series. After that change, a duration of about two images is required before the drift speed stabilizes again.

The influence of this perturbation will be presented hereafter.

Picture processing

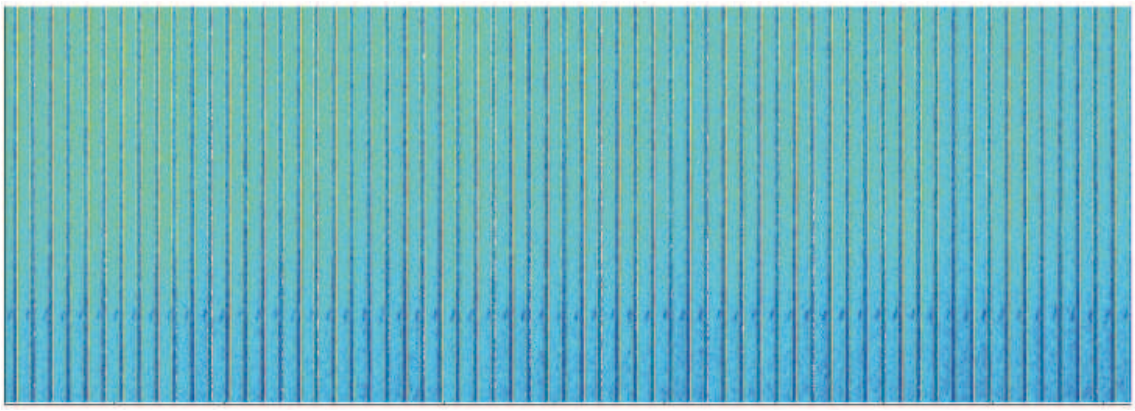


Figure 2.17: Slice of a typical image obtained with the source flashing at 0.89Hz and a master clock frequency of 1MHz. The vertical lines correspond to signatures of the flashes. The number of these lines informs us about the duration of the image. Note that only a central band of the image is presented for the sake of clarity.

Fig. 2.17 presents a horizontal slice of a typical image that we got from the testing setup. The flash signature is clearly visible in these images, and similitudes with the "neon" image are clearly noticeable.

The first step in our analysis consists in applying a two dimensional Fast Fourier Transform of the images so that a frequency analysis is possible. The power spectrum of the image should present some peaks that correspond to the flash frequency and to its harmonics. Our processing includes a subtraction of the average power spectrum so that only the peaks appear in the spectrum (fig. 2.18). The frequency positions of the peaks are then determined and used as the basis for the measurements.

It should be reminded that the x-axis frequency scale of the Fourier transform is determined by the cutoff frequency. In our case, the frequency axis has 1024 steps between 0 and the cutoff frequency. Then, one simply has to adjust the cutoff frequency so that the measured positions of the peaks correspond to the flash frequency and its harmonics. The corresponding cutoff limit is the real searched value. This process is repeated several times to improve the fitting of the comb and to get a precise value of the cutoff frequency. The latter is closely related to the readout time and to the column drift rate, as previously mentioned. In this way, we are able to verify that our CCD TDI-behavior fits the theoretical model.

Results

The total CCD chip readout time as a function of the master clock frequency is represented in fig. 2.19. The durations determined with the flash sequences are very well fitted with the theoretical curve.

Another approach consists in computing the readout time of each image from the date of creation of the image files. Although this process is not really accurate (the date accuracy is of

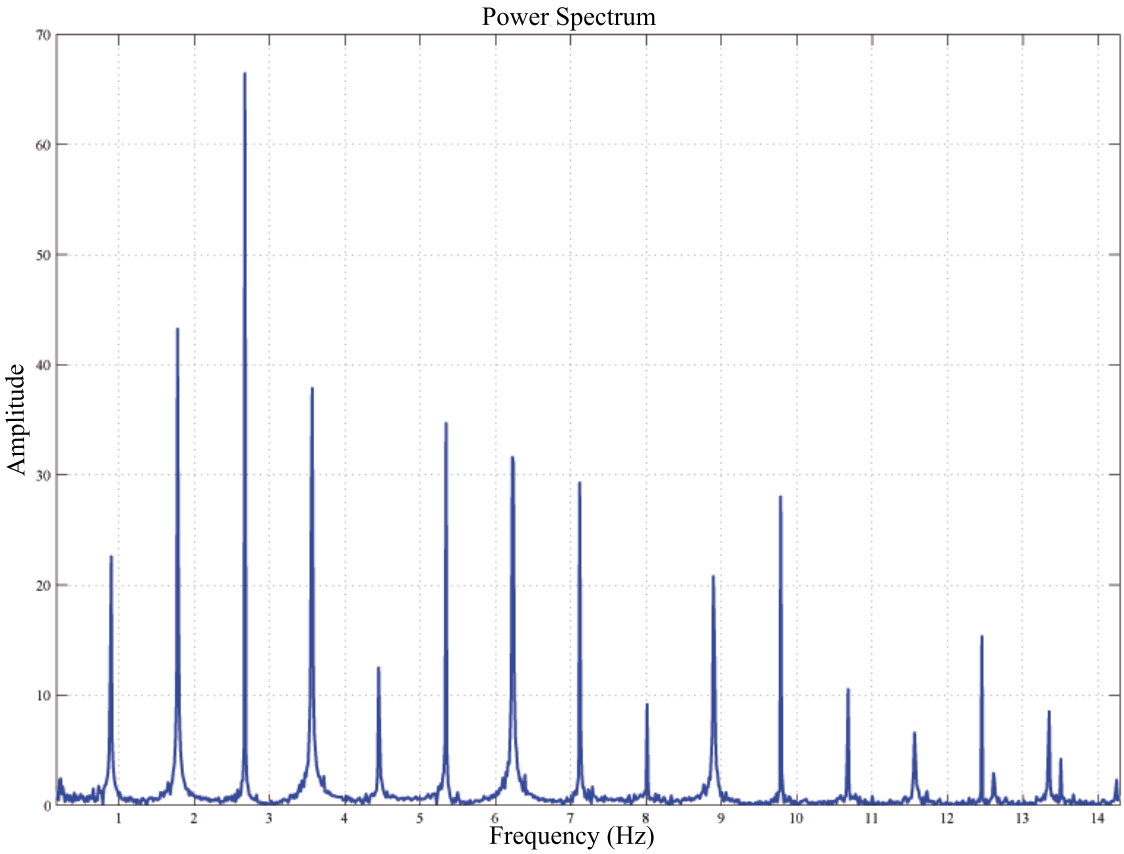


Figure 2.18: Power spectrum of the image presented in fig. 2.17. The mean values have been subtracted so that only the peaks appear on the graph. The x-axis scale is determined by the cutoff frequency of the CCD, depending itself on the CCD readout time.

the order of one second), it can give a qualitative idea. The values computed from the dates also correspond quite well to the theoretical values. Our $2k \times 2k$ CCD can then be said to fit the theoretical expectancy as far as the relation between the readout time and the master clock frequency is concerned.

The other interesting result of these tests is the verification of the relationship between the column drift rate and the master clock frequency. This relation can be obtained from equations 2.6 and 2.7

$$f_{col} = \frac{N_c}{N_p} \frac{f}{16} \tag{2.8}$$

Fig. 2.20 presents this relationship; the computed results fit the theoretical curve quite well. As expected, the drift rate of the CCD is a linear function of the master clock frequency. Equation 2.8 can be used to determine the master clock frequency that corresponds to the sidereal speed.

These results show that our CCD global TDI-behavior is satisfactory. However, it is now necessary to check the stability of the drift rate as well as the direction of the drift. This is achieved in the next set of tests.

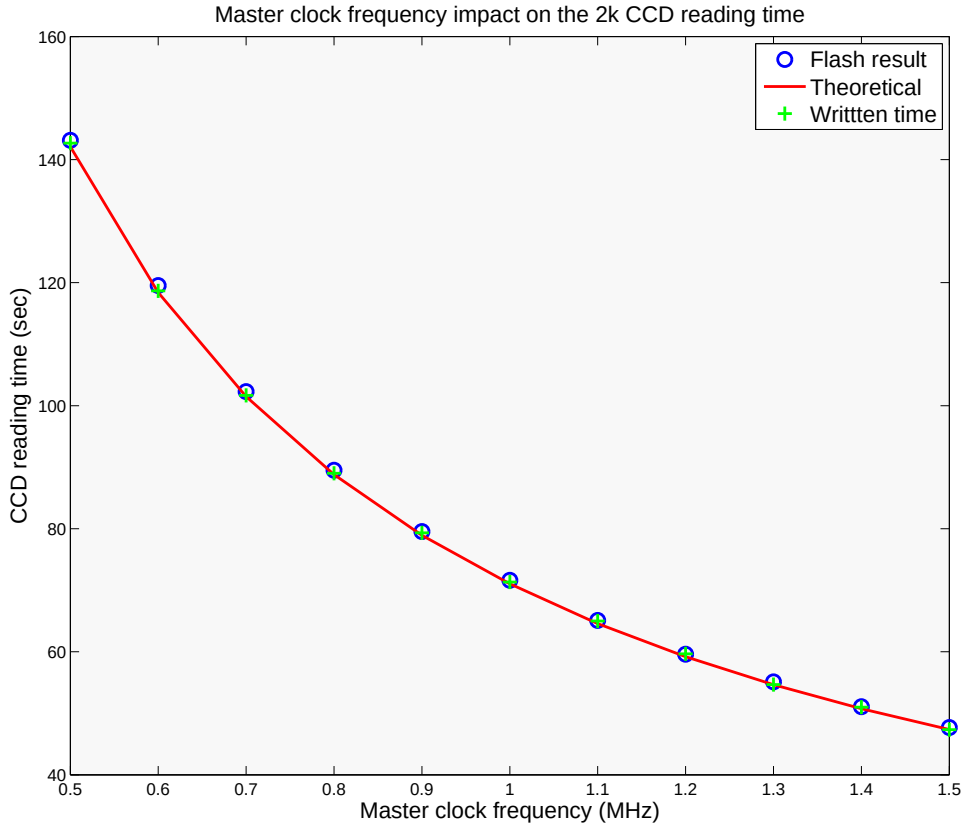


Figure 2.19: Relationship between the master clock frequency and the CCD readout time. The blue circles show the results computed from the flash analysis. The red solid curve represents the locus of the theoretical values computed from equation 2.7. The green crosses show the CCD readout time computed from the image dates of creation.

The drift stability test principle

Now that we have determined the cutoff frequency and the relation between the drift rate and the master clock frequency, we can perform a test that will inform us about the drift stability.

The principle of this test is quite simple. It first consists in expressing the flash period in terms of a number of column as the time sampling is converted in space sampling because of the TDI readout mode. We then determine this period for every flash present in each image. The average period corresponding to an image (T_{flash}) is computed as well as the standard deviation. This deviation is the measurement of interest as it will inform us about the variations of the period in a given image. A smaller standard deviation corresponds to a better drift stability.

The theoretical flash period expressed as a number of rows can be computed as a function of the master clock frequency. Indeed, the number of rows per flash can be computed from the number of rows read per second, that is the column drift rate, and from the flash frequency:

$$T_{\text{flash}} = \frac{f_{col}}{f_{in}} \tag{2.9}$$

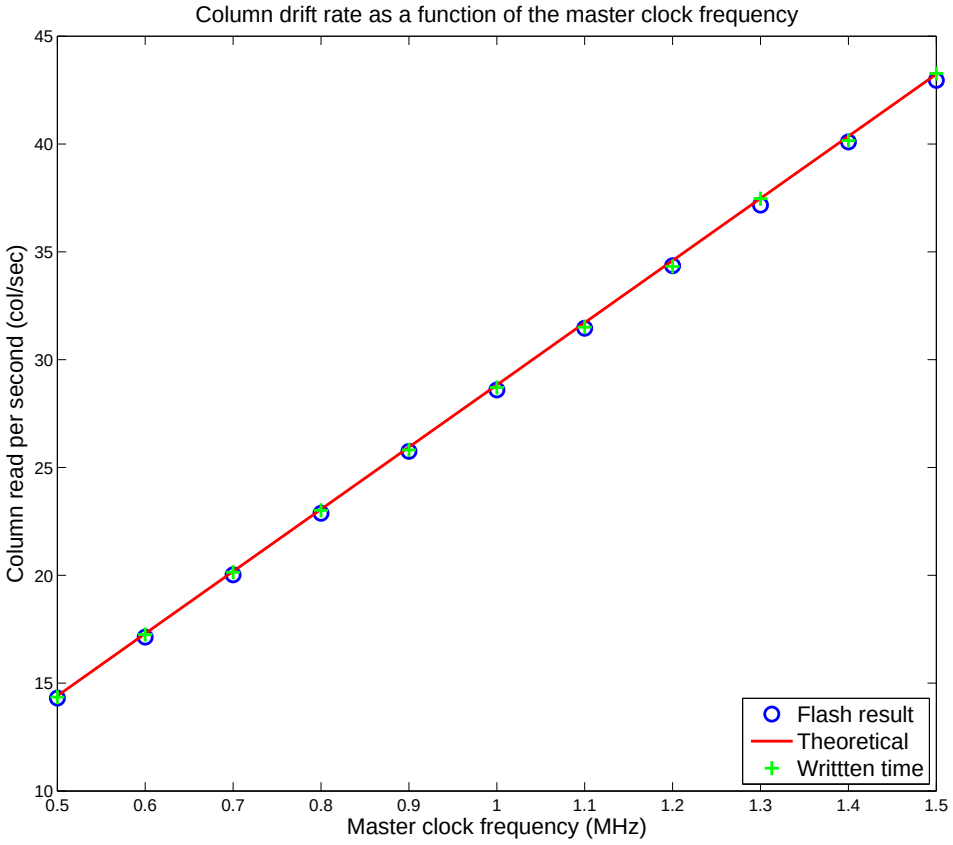


Figure 2.20: Column drift frequency as a function of the input master clock frequency. The color code is the same as in fig. 2.20.

where f_{in} is the frequency of the flashes. Moreover, we have shown earlier that the column drift rate (f_{col}) is a linear function of the master clock frequency. As we already know that the measured f_{col} follows the theoretical law, we expect the measurements of T_{flash} to be well fitted by the theoretical curve.

The top graph of fig. 2.21 presents the results of the measurements concerning the drift stability. As expected, the theoretical curve (in red) fits the measurements (blue circles) quite well. The vertical blue error bars represent the standard deviation related to the mean values. These deviations, that represent the drift rate instability, are plotted in the bottom graph as a function of the master clock frequency. The dotted red line represents the limiting measurement accuracy. Indeed, the flash period cannot be measured with an accuracy better than half a column. The results show that the drift rate is very stable since the average drift instability remains lower than half a column.

Fig. 2.22 shows the global effect of a perturbation on the master clock frequency on the drift stability in the images that follow this perturbation. The values presented are averages of the different sets of images corresponding to each frequency applied to the master clock. The value corresponding to the images during which the frequency perturbation is applied does not appear on the graph, only the subsequent images are presented. The perturbation simply consists in a one step increase of the frequency.

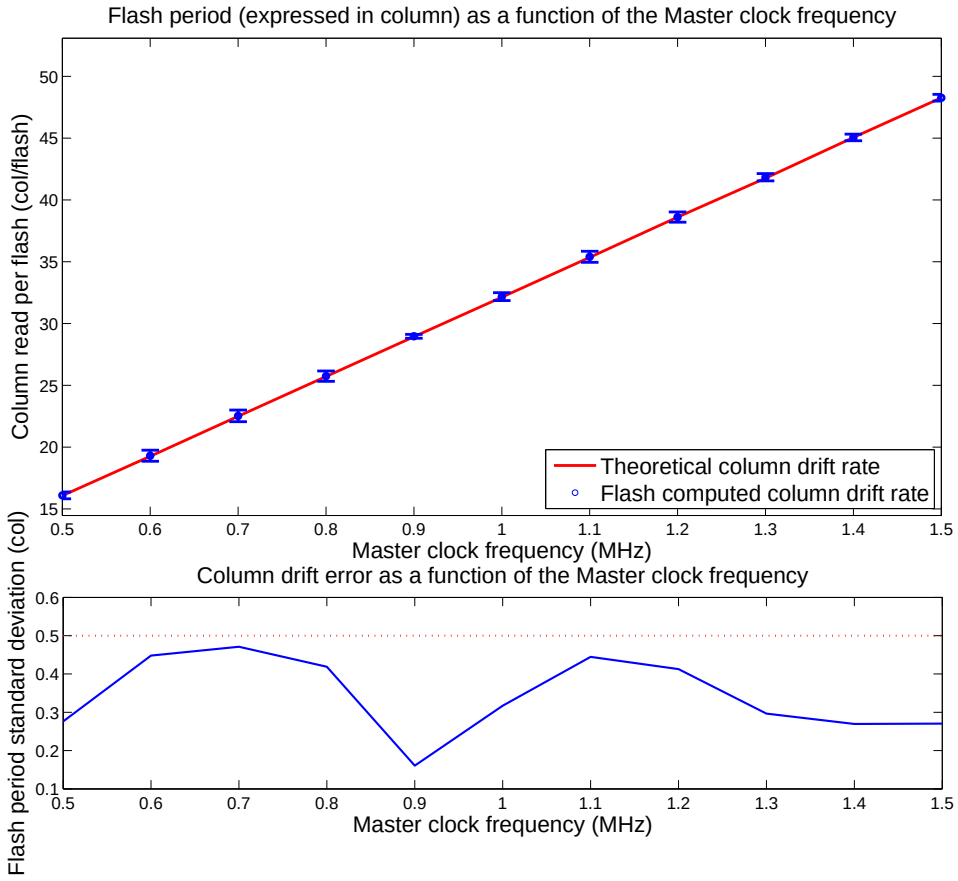


Figure 2.21: Top: The red solid curve represents the theoretical relation between the flash period measured as a number of rows and the master clock frequency. The blue points and error bars respectively correspond to the mean periods computed from the flash analysis and to the standard deviation of each set of measurements. Bottom: The blue solid. curve represents the mean standard deviation of the flash period computed from the flash analysis as a function of the master clock frequency. The values plotted here correspond to those of the error bars shown on the top graph. The red dotted line shows the limit of the measurement accuracy (half a column).

This graph aims at demonstrating the ability of this measurement to detect drift instability. Indeed, we expected that modifying the master clock frequency would perturbate the drift rate and that the effect of the perturbation would vanish after a few images. This is confirmed by our results. We also deduce from the graph in fig. 2.22 that the three images following the image during which the frequency has been modified are also disturbed.

Conclusion

As a conclusion, we have shown that the TDI mode of the $2k \times 2k$ CCD camera matches the expectations. It has first been demonstrated that the relation between the master clock frequency and the column drift rate fits the theory quite well. This will allow to precisely adjust the drift frequency to the sidereal rate.

After that, the stability of the column stepping has been studied for stable and disturbed images. Once again, the results are optimistic since we have shown that the stability is better

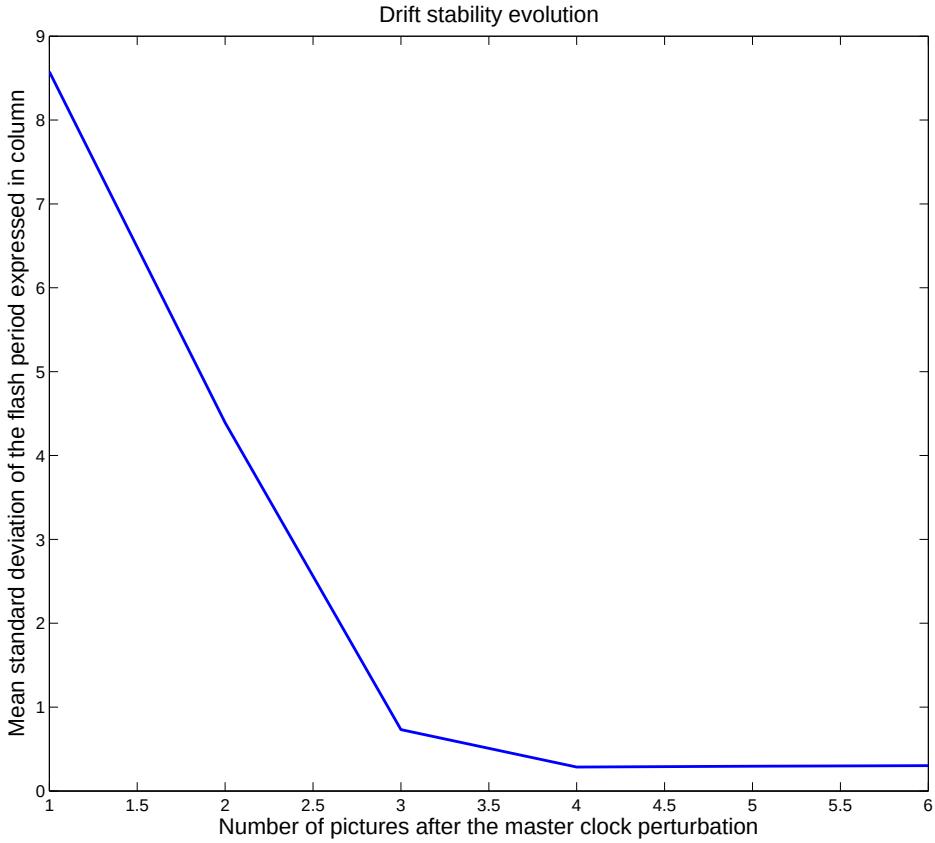


Figure 2.22: Evolution of the drift stability after a perturbation of the master clock frequency. The x-axis corresponds to the number of images read after the perturbation. The graph represents a mean of the standard deviations corresponding to each master clock frequency.

than the measurement accuracy. Moreover, it appeared from our tests that a perturbation of the master clock frequency can disturb the drift rate of up to three images after the image during which the perturbation occurred.

2.3.4 Discussion

The first aspect to consider is the general use of the CCD camera. Several drawbacks were found during our testing session. The first one is the long preparation time required before taking any image. The chip enclosure of the camera has to be pumped to create the vacuum. It has also to be cooled down to 223K (-50°C). The whole process takes about two and a half hours. Moreover, the vacuum pump has to run permanently to maintain the vacuum level. This results in the generation of vibrations and heating that may disturb the image quality. Commercial CCD cameras are generally sealed and pumped once a year only, they require only a few minutes before they can be used. It would be interesting to use such a solution for the ILMT ($4k \times 4k$) camera.

Another difficulty comes from the readout mode. As only the TDI readout mode is available, it is not easy to get long exposures. Indeed, even if the drift can be stopped to increase the

acquisition time, the CCD is nevertheless always read in TDI. Therefore, the acquisition continues during the readout of the chip.

Finally, it should be noted that the CCD camera is provided with all the material that is required for the readout, driving, cooling and pumping. These modules are separated from the camera and the electronic is gathered in a voluminous rack. The whole camera thus occupies an important volume because of all these devices. All this makes this camera not very practical to use.

Despite these difficulties, we have conducted several tests aiming at characterizing the imaging capability of the camera and the scientific interest of the acquired data. These tests can be classified in two categories: the image quality and the TDI behavior. The first part consists in the study of flat-fields, saturation limits, darks and bad pixels. The second one investigates the TDI drift speed and stability.

On the one hand, the imaging capability of the camera seems compromised because of several reasons. The response of the pixels varies a lot over the whole CCD. The difference can reach 25% between the center and the edge of the detector. Moreover the saturation limit of the chip is not uniform at all. This causes trouble when an image successively crosses a region of high saturation limit (edges of the chip) and then another one of low saturation limit (center of the chip). The excess of electrons acquired in the high limit area are released on the CCD when it reaches a low saturation area. This produces a charge drift across the CCD. The images that are saturated are thus not reliable. Because the electron drift continues on the next images, these images are also lost. As no mechanical shutter is provided, it is impossible to protect the CCD from luminous objets and the saturation effects cannot be prevented. It can also appear necessary to restart the camera in order to clear the saturation disturbances caused by the variable saturation level.

On the other hand, considering the TDI, both the speed of the column stepping and its stability behave as theoretically expected. The drift rate is proportional to the master clock frequency and the period errors that would exist are smaller than the accuracy of our tests.

Finally, even if the camera is not very practical to use, it complies with its specifications, except for the edge effects affecting the sensitivity of the chip. However, the presence of several difficulties such as the saturation makes it unreliable for scientific purposes as it is almost impossible to know whether the images were disturbed by saturation phenomena.

Part II

Optical considerations

Chapter 3

Disturbances of the liquid mirror

The existing literature about liquid mirrors presents those as instruments of very high optical quality that provide seeing-limited scientific images (e.g. Hickson et al. 2007; Hickson and Racine 2007). Indeed, interferometric tests conducted by Borra et al. (1992) show that the liquid surface is parabolic within an RMS error of the order of $\lambda/20$ (at 632nm); as far as the NODO is concerned, the RMS surface error was 37nm whereas the LZT even reaches the value of 9nm.

However, the surface of a liquid mirror is not formed once and for all. It can be disturbed by many phenomena. Some of them may cause small permanent deformations, like a change of the shape of the mirror that results in a small change of focal length. Other ones may induce time variable disturbances, like waves traveling through the mercury. All these have effects on the image quality of the liquid mirror telescope.

In the first chapter we have shown that the free surface of a liquid is in equilibrium when its normal is parallel to the resultant of the net acceleration felt by this liquid. Based on this property, we will show in this chapter the effects on the surface shape induced by several causes that were neglected in the first approach given in chapter 1. We first remind the zero order result and then increase the problem complexity. We consider the case where the gravitation acceleration varies in direction and intensity as a function of the position of the fluid element on the mirror (due to the Earth's curvature and altitude difference). The effect of the tilt of the rotation axis of the mirror is also investigated. We finally study the effect of the Coriolis force due to the Earth rotation. We will see that we can definitely not neglect the perturbations causing a variation in the focal length. In addition to this defocusing, the Coriolis force introduces a non negligible astigmatism. Fortunately, almost all the other perturbations are so small that they do not have to be considered.

A matrix approach of these studies is presented in Gibson and Hickson (1992a) and their results are applied to the 2.7m UBC/Laval LMT. Mulrooney (2000) presents the same type of results applied to the NODO LMT. The study we present in this section consists of an intuitive vectorial approach, based on the projection of the different accelerations on the axis of the mirror reference frame. Our results are consistent with those of Gibson and Hickson (1992a) and Mulrooney (2000).

We then decompose the difference between the actual shape of the mirror and the perfect parabola in terms of Zernike polynomials. Indeed, any wavefront can be decomposed in fundamental aberrations. The Zernike polynomials constitute a particular basis for wavefront series expansion, each of them representing a particular aberration. Expressing a wavefront in terms of

Zernike polynomials thus allows to know what are the errors affecting this wavefront. In the case of the disturbances on the liquid mirror, expressing them in series of Zernike polynomials allows to precisely determine what type of aberrations will be contained in the images obtained with this mirror. That is why it is so important to study these perturbations. The Zernike polynomial properties will be extensively presented in chapter 4 (section 4.2.1).

After that, we investigate the time variable disturbances. Two types of waves have been previously observed on liquid mirrors. Spiral waves are caused by the differential wind between the mercury and the air above (Bernoulli's equation), while concentric waves are generated by vibration transmitted to the mirror dish. Both types are modeled and their effects are studied hereafter.

Using typical parameters of other liquid mirrors, we study the effect of these waves on the quality of the images produced by a disturbed mirror. Once the ILMT will have been constructed, the parameters will be measured and the models will be completed. They will then be applied to precisely determine the aberrations of the mirror and the resulting images.

3.1 Zero order equilibrium

It has already been demonstrated in chapter 1 (section 1.2.1) that the equipotential surface of a rotating liquid in the uniform field of gravity of the Earth is given by equation 1.2, recalled here:

$$z = \frac{\omega^2 r^2}{2g_0} \qquad\qquad F_0 = \frac{g_0}{2\omega^2} \tag{3.1}$$

The first expression is the equation of a parabola height (z) as a function of the axial distance (r) of focal length F_0, whose value is given by the second formula, g_0 is the Earth gravitation (considered uniform and constant is this first approximation) and ω is the angular speed of the mirror.

The equipotential surface adopted by the fluid is computed after determining the total acceleration felt by the liquid. In this particular zero order case, the only accelerations to consider are related to the constant rotation of the mirror (the centrifugal acceleration) and the uniform acceleration of gravity. Even if this is a good, first approximation, real cases involve other parameters such as the curvature of the Earth and the Coriolis effect related to the Earth rotation. These effects are explored in the following sections in order to estimate the aberrations that they induce on the mirror.

3.2 Gravitational field gradient and Earth curvature

The first effect we consider is the non-uniformity of the gravitational field $\overrightarrow{g}$, related to the curvature of the Earth. There are two things to be considered here, the variation in intensity of the vector $\overrightarrow{g}$ with the altitude and its variation in inclination. Indeed, $\overrightarrow{g}$ is not uniformly vertical, it is a radial vector pointing toward the center of the Earth.

Let us first consider the case for which the rotation axis of the mirror is perfectly aligned with the local gravitational acceleration that is considered to be uniformly vertical. However, we consider that the intensity of $\overrightarrow{g}$ varies with the altitude.

It comes out of the equivalence principle between the inertial mass and the gravitational mass, that the intensity of the gravitational force F_g that applies on a fluid element of mass m is

$$F_g = mg = G\frac{M_E m}{R_E^2} \tag{3.2}$$

where G is Newton's constant of gravitation and M_E, R_E are the mass and radius of the Earth (between the center of the Earth and the surface of the center of the mirror). Now, as the mirror is not flat, the height (z) of the mirror surface should be taken into account. We then get,

$$\frac{1}{g_E} = \frac{(R_E + z)^2}{GM_E} = \frac{1}{g_0}\left(1 + \frac{z}{R_E}\right)^2 \tag{3.3}$$

where g_E is the general gravitational acceleration of the Earth corresponding to an altitude z above the surface of the planet ($z = 0$, see fig. 3.1), g_0 is the gravitational acceleration at the surface of the planet. Replacing g_0 with this expression of g_E in the zero order computation gives

$$\frac{dz}{dr} = \frac{\omega^2 r}{g_0}\left(1 + \frac{z}{R_E}\right)^2 \tag{3.4}$$

Integrating this expression with respect to dr gives the equation of the equipotential surface

$$z = R_E\left(\frac{1}{1 - \frac{\omega^2 r^2}{2g_0 R_E}} - 1\right) \tag{3.5}$$

Using the series development for $1/(1-\varepsilon)$ limited to the fourth order, we get

$$z = \frac{\omega^2 r^2}{2g_0} + \frac{\omega^4 r^4}{4g_0^2 R_E} + O(r^5) \tag{3.6}$$

The first term of this equation is the parabola of the zeroth order case, and the second term corresponds to a combination of Zernike aberrations (piston, defocus and spherical)

$$\frac{\omega^4 r^4}{4g_0^2 R_E} = \frac{\omega^4}{4g_0^2 R_E}\left(\frac{Z_1}{3} + \frac{Z_4}{2\sqrt{3}} + \frac{Z_{11}}{6\sqrt{5}}\right) \tag{3.7}$$

where Z_i is the i^{th} Zernike polynomial, Z_1 being piston term, Z_4 the defocus aberration and Z_{11} the spherical aberration. The coefficients of the Zernike polynomial Z_i represent the intensity of the aberration associated to this polynomial. In the case of the ILMT ($\omega = 0.783$rad/s, $g_0 = 9.81\text{m/s}^2$ and $R_E \approx 6.37 \cdot 10^6$m), the intensities of these aberrations are listed in Table 3.1. All three aberrations are found to be negligible.

Aberration	Intensity (nanometers)
Piston	0.051
Defocus	0.044
Spherical	0.011

Table 3.1: Intensity of the aberrations due to the non uniformity of the gravitational field intensity expressed in nanometers. All these aberrations are smaller than $\lambda/10000$. These aberrations are illustrated in fig. 4.10 (section 4.2.1).

Let us now consider the case for which the intensity of $\overrightarrow{g}$ is not dependent on the altitude but the vector $\overrightarrow{g}$ is oriented toward the center of the Earth instead of being uniformly vertical. Indeed, due to the Earth curvature, the orientation of $\overrightarrow{g}$ at the edge of the mirror is different from the one at the center. This means that $\overrightarrow{g}$ has an horizontal component in the mirror reference frame. This component points toward the center of the mirror. Hence, it has to be subtracted from the centrifugal acceleration. Fig. 3.1 shows a representation of the reference frame and the accelerations that have to be taken into account.

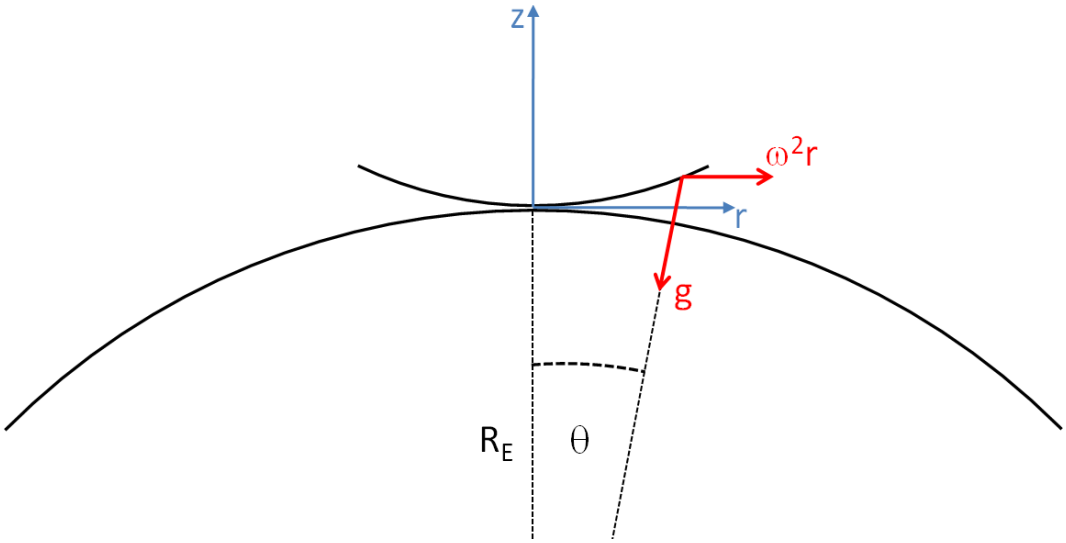


Figure 3.1: Representation of the variation of the orientation of the g vector in the mirror reference frame.

The tangent of the θ angle is given to a first approximation by the ratio of the position r on the axis and the radius of the Earth R_E. Knowing this, it is possible to calculate the vertical and horizontal projections of $\overrightarrow{g}$ (g_v, g_h)

$$g_v = g_0 \Big/ \sqrt{1 - \frac{r^2}{R_E^2}} \qquad\qquad g_h = g_0 \frac{r}{R_E} \Big/ \sqrt{1 - \frac{r^2}{R_E^2}} \tag{3.8}$$

We can then deduce

$$\frac{dz}{dr} = \frac{\omega^2 r}{g_0} \sqrt{1 - \frac{r^2}{R_E^2}} - \frac{r}{R_E} \tag{3.9}$$

that can be integrated with respect to dr

$$z = \frac{\omega^2}{g_0} \left(\frac{-R_E^2}{3} + \frac{1}{3} \sqrt{1 + \frac{r^2}{R_E^2}} (r^2 + R_E^2) \right) - \frac{r^2}{2R_E} \tag{3.10}$$

Using the series development of $\sqrt{1+\varepsilon}$ limited to the fourth order, we get

$$z = \frac{\omega^2 r^2}{2g_0} - \frac{r^2}{2R_E} + \frac{\omega^2 r^4}{8 g_0 R_E^2} + O(r^5) \tag{3.11}$$

Once again, the first term corresponds to the zeroth order parabola. The second one is also a function of r^2 and can be combined with the first term to correspond to another effective parabola

$$z = \frac{\omega^2}{2g_0}\left(1 - \frac{g_0}{\omega^2 R_E}\right) r^2 + \frac{\omega^2 r^4}{8 g_0 R_E^2} + O(r^5) \tag{3.12}$$

whose focal length is given by

$$F = F_0 \left(1 - \frac{g_0}{\omega^2 R_E}\right)^{-1} \tag{3.13}$$

where $F_0 = g_0/2\omega^2$ is the initial focal length of the zeroth order parabola. The new focal length is thus slightly larger than the initial one and the variation is proportional to F_0. In the case of the ILMT, the variation is about 2.5ppm of the focal length, which corresponds to about $20\mu m$.

The third term of equation 3.11 can be decomposed as a sum of Zernike aberrations (piston, defocus and spherical) as in the previous case.

$$\frac{\omega^2 r^4}{8 g_0 R_E^2} = \frac{\omega^2}{8 g_0 R_E^2}\left(\frac{Z_{11}}{6\sqrt{5}} + \frac{Z_4}{2\sqrt{3}} + \frac{Z_1}{3}\right) \tag{3.14}$$

The intensities of these aberrations are given in Table 3.2. Once again, they are totally negligible. The main effect of the Earth curvature is thus a small variation in the focal length of the parabola, that can be compensated for either by changing the angular speed (ω) of the mirror or by simply adjusting the detector focal position.

Aberration	Intensity (femtometers)
Piston	0.058
Defocus	0.050
Spherical	0.013

Table 3.2: Intensities of the aberrations due to the curvature of the gravitational field expressed in femtometers. They are all smaller than $\lambda/10^{10}$, and are thus negligible.

The two previous approaches were oversimplified. Indeed, it is obvious that if the curvature of the Earth is to be considered, the gradient of the gravitational acceleration should also be taken into account. We now explore the combination of these two effects. Looking again at fig. 3.1, we now have to add the distance z to the radius of the Earth. The equations of $\tan\theta$, g_v and g_h become

$$\begin{aligned} \tan\theta &= \frac{r}{R_E + z} \\ g_v &= g_0\left(1 - \frac{r^2}{(R_E+z)^2}\right)^{-1/2} \\ g_h &= g_0 \frac{r}{R_E+z}\left(1 - \frac{r^2}{(R_E+z)^2}\right)^{-1/2} \end{aligned} \tag{3.15}$$

Using these expressions, and denoting ΔR the variation of altitude with respect to the surface of the planet (the distance between the mirror and the surface of the Earth along $\overrightarrow{g}$), the derivative of z with respect to r can be written

$$\frac{dz}{dr} = \frac{\omega^2 r}{g_0}\left[1+\left(\frac{r}{R_E+z}\right)^2\right]^{1/2}\left(1+\frac{\Delta R}{R_E}\right)^2 - \frac{r}{R_E+z} \tag{3.16}$$

where ΔR is given by

$$\Delta R = \sqrt{(R_E+z)^2+r^2} - R_E \tag{3.17}$$

Integration of equation 3.16 (performed with the Mathematica software) leads to

$$\begin{aligned} z &= (-2g_0R_E+\omega^2r^2)^{-2}\big[-4g_0^2R_E^3+4\omega^2g_0R_E^2r^2+\big\{-\omega^2R_Er^4 & (3.18)\\ &+ \sqrt{\frac{[-2g_0R_E+\omega^2r^2]^2[2g_0(R_E-r)+\omega^2R_Er^3][-\omega^2r^3+2g_0R_E(R_E+r)]}{R_E}}\big\}\big] & (3.19)\end{aligned}$$

This equation is not really usable as it is. However, its series expansion limited to the fourth order (also computed with Mathematica) is much more interesting

$$z = \frac{\omega^2}{2g_0}\left(1-\frac{g_0}{\omega^2R}\right)+\left(\frac{\omega^2}{4gR_E^2}+\frac{\omega^4}{4g^2R_E}-\frac{1}{8R_E^3}\right)r^4+O(r^5) \tag{3.20}$$

The parabola obtained is the same as the one that we get when considering only the curvature of the Earth. As far as the fourth order term is concerned, it is much more complicated but it can again be expressed as a sum of Zernike aberration polynomials (piston, defocus and spherical)

$$\left(\frac{\omega^2}{4gR_E^2}+\frac{\omega^4}{4g^2R_E}-\frac{1}{8R_E^3}\right)r^4 = \left(\frac{\omega^2}{4gR_E^2}+\frac{\omega^4}{4g^2R_E}-\frac{1}{8R_E^3}\right)\left(\frac{Z_1}{3}+\frac{Z_4}{2\sqrt{3}}+\frac{Z_{11}}{6\sqrt{5}}\right) \tag{3.21}$$

The aberration coefficients are given in Table 3.3 for the ILMT.

Aberration	Intensity (nanometers)
Piston	0.048
Defocus	0.042
Spherical	0.011

Table 3.3: Intensity of the aberrations due to the curvature and non uniformity of the gravitational field expressed in nanometers. Again they are all smaller than $\lambda/10000$.

As expected from the previous results, the aberrations related to the fourth order term are negligible, the only noticeable effect of the curvature of the Earth is thus the very slight variation ($\sim 20\mu$m) in the focal length.

3.3 Tilt of the rotation axis of the liquid mirror

The calculations presented in the previous section assumed that the axis of rotation of the mirror was perfectly aligned with the local gravity vector (for $r = 0$). We investigate now the effect of a misalignment between the rotation axis and the gravity vector, the latter being considered constant and uniform.

A tilt of the rotation axis of the mirror in the gravitational reference frame corresponds to a global tilt of the gravitational acceleration in the mirror reference frame. The situation is represented in fig. 3.2.

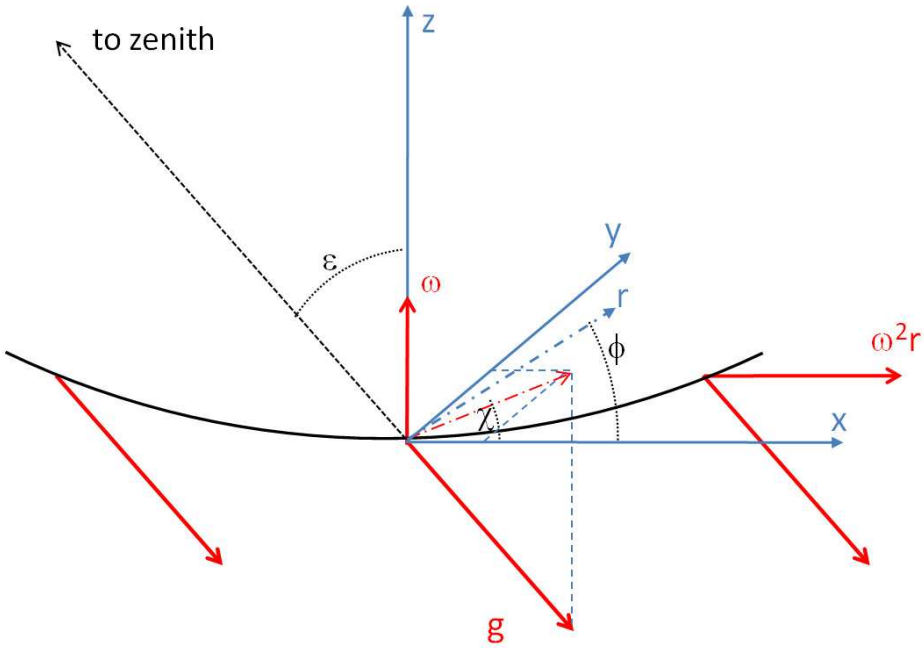


Figure 3.2: A tilt of the axis with respect to the local gravity corresponds to a tilt of the gravitational field vector in the reference frame of the mirror. This **g** vector has to be projected onto the (x, y, z) cartesian coordinates and then onto the (r, ϕ, z) polar coordinates.

The acceleration of the gravity vector $\overrightarrow{g}$ has to be projected on the mirror reference frame (x,y,z) and then on the cylindrical coordinates (r, ϕ, z) where $\overrightarrow{r}$ is the position of a particular infinitesimal element of fluid. The projections are

$$\begin{aligned} g_z &= g_0 \cos(\varepsilon) \\ g_r &= g_0 \sin(\varepsilon) \sin(\chi - \phi) \\ g_\phi &= g_0 r \sin(\varepsilon) \cos(\chi - \phi) \end{aligned} \tag{3.22}$$

(3.23)

where ε is the angle between the axis of the mirror and the zenith direction, ϕ is the angle between $\overrightarrow{x}$ and $\overrightarrow{r}$ and χ is the angle between $\overrightarrow{x}$ and the projection of $\overrightarrow{g}$ in the horizontal plane $(\overrightarrow{x}, \overrightarrow{y})$. The slope of the resulting surface with respect to r and ϕ are respectively

$$\begin{aligned} \frac{dz}{dr} &= \frac{\omega^2 r + g_0 \sin(\varepsilon)\sin(\chi-\phi)}{g_0\cos(\varepsilon)} \\ \frac{dz}{d\phi} &= r\tan(\varepsilon)\cos(\chi-\phi) \end{aligned} \tag{3.24}$$

Integration of these equations leads to

$$\begin{aligned} z &= \frac{\omega^2 r^2}{g_0\cos\varepsilon} + \tan(\varepsilon)\sin(\chi-\phi)r + f(\phi) \\ z &= r\sin(\varepsilon)\sin(\chi-\phi) + f(r) \end{aligned} \tag{3.25}$$

These equations are consistent with each other and can be combined to give

$$z = \frac{\omega^2 r^2}{g_0\cos\varepsilon} + \tan(\varepsilon)\sin(\chi-\phi)r \tag{3.26}$$

The first term represents a parabola with a focal length

$$F = \frac{g_0\cos(\varepsilon)}{2\omega^2} = F_0\cos(\varepsilon) \tag{3.27}$$

The variation in the focal length as a function of the tilt angle (ε) is shown in fig. 3.3. As for the previous cases, the second term in equation 3.26 can be expressed as a function of the Zernike polynomials

$$\tan(\varepsilon)\sin(\chi-\phi)r = \tan(\varepsilon)\left(\cos\chi\frac{Z_2}{2} + \sin\chi\frac{Z_3}{2}\right) \tag{3.28}$$

where Z_2, Z_3 are the Zernike polynomials relative to tip/tilt aberrations. This is illustrated in fig. 3.4.

The impact on the focal length can be important as a tilt of 0.9 degree results in a variation of the focal length of 1mm. However, using a precision level, it should be possible to align the rotation axis to better than 1 arcsec. Such a tilt would induce 0.1nm (10^{-4} wave) of defocus, that is negligible. The effect of the second term (between parentheses) is a global tilt of the wavefront that does not introduce any aberration. Mulrooney (2000) presents the effect of an axis misalignment for the case of the NODO telescope.

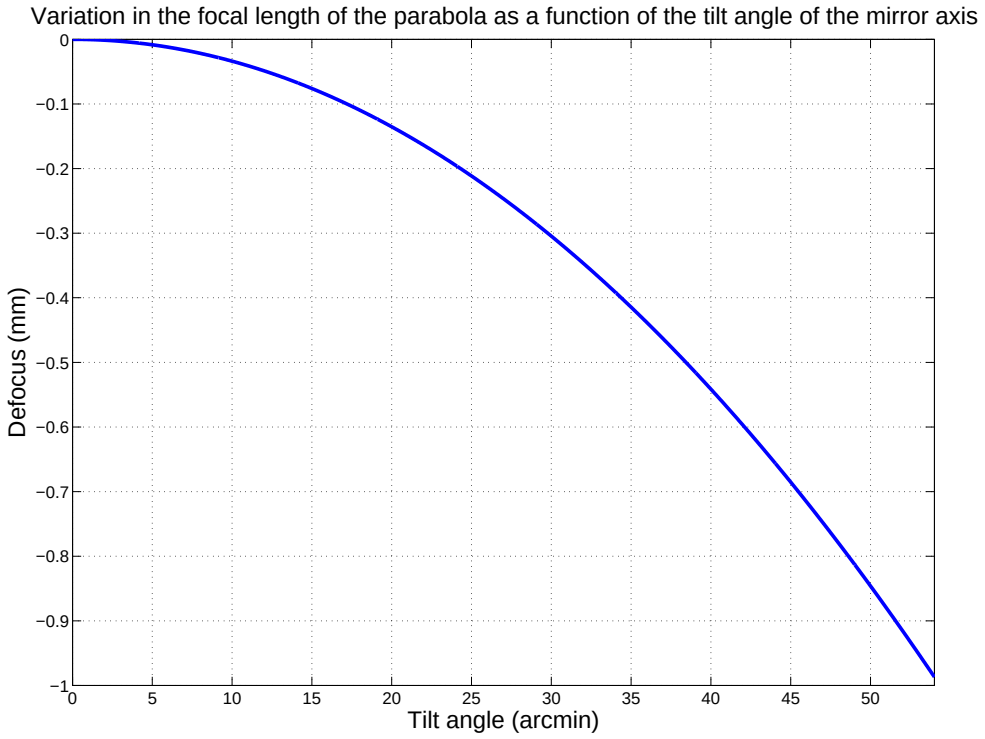


Figure 3.3: Evolution of the focal length of the parabola as a function of the inclination of the mirror rotation axis with respect to the vertical axis.

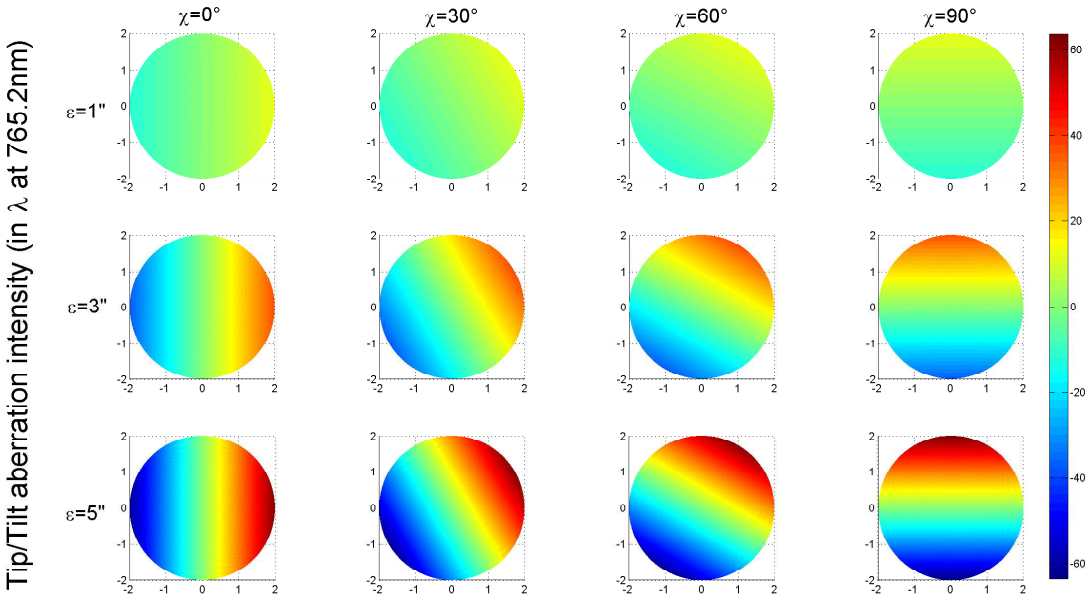


Figure 3.4: Illustration of the tip/tilt aberration for several values of χ (0°,30°,60°,90°) and ε (1,3,5 arcmin). The intensity is expressed in wavelength units (at 765.2nm) and the (x,y) coordinates represent the position on the mirror (in meters).

3.4 Earth rotation: the Coriolis effect

Until now, we did consider that the mirror was in a fixed reference frame, however, it is moving around the Earth. The fictive forces related to the Earth rotation can have an impact on the mirror surface. This section aims at investigating this effect.

This case is more difficult to consider intuitively as it implies many projections that can reveal difficult to handle. We thus summarize here the convenient method detailed in Gibson and Hickson (1992a). It consists in using transformation matrices (rotation and translation) in order to establish a relation between the mirror reference frame ($\mathbf{X}$) and the Earth reference frame ($\mathbf{U}$). Nevertheless, it seems that there is a small error in the rotation matrix $\mathbf{R_3}$ and its derivative used by Gibson and Hickson. We use here the same approach, but have corrected these errors. The situation is described in fig. 3.5.

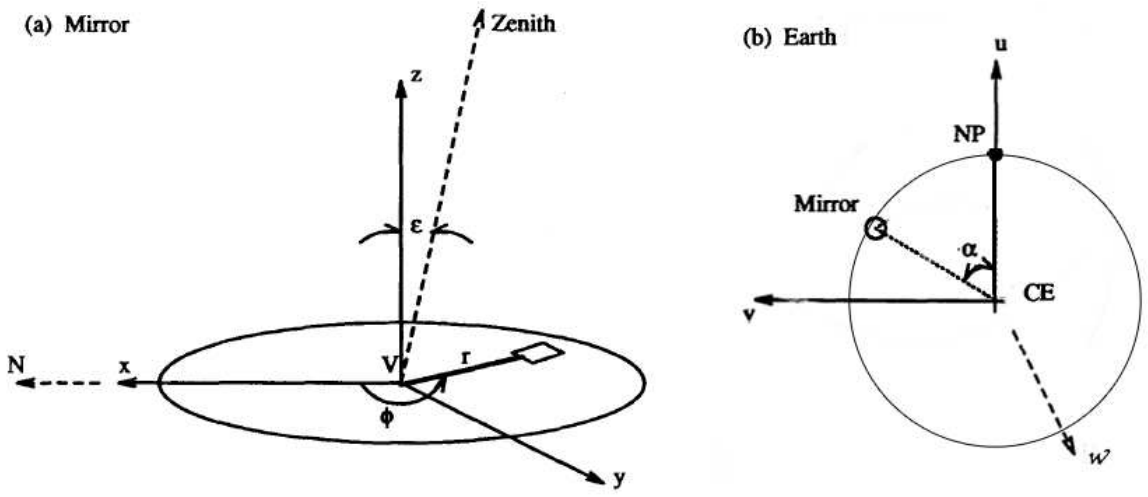


Figure 3.5: Illustration of the mirror reference frame (left) and the Earth reference frame (right). The original figure comes from Gibson and Hickson (1992a)

The relation between the two reference frames can be expressed as

$$\mathbf{U} = \mathbf{R_3}\mathbf{R_2}(\mathbf{X} + \mathbf{T}) \tag{3.29}$$

where $\mathbf{U}$ is the coordinate system related to the Earth reference frame (u, v, w) whereas $\mathbf{X}$ is the one corresponding to the mirror (x, y, z). $\mathbf{T}$ is the translation matrix that links the origin of both coordinate systems, and $\mathbf{R_2}, \mathbf{R_3}$ are rotation matrices that are described hereafter.

$$\mathbf{X} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad \mathbf{U} = \begin{pmatrix} u \\ v \\ w \end{pmatrix} \qquad \mathbf{T} = \begin{pmatrix} 0 \\ 0 \\ R_E \end{pmatrix} \tag{3.30}$$

The vector $\overrightarrow{u}$ is aligned with the North-South axis and points toward North. Both $\overrightarrow{x}$ and $\overrightarrow{u}$ are tangent to the terrestrial spheroid and are oriented respectively toward North and West and $\overrightarrow{z}$ points toward the zenith of the mirror. Moreover, the following relation exists between the cartesian coordinates (x,y,z) and the cylindrical coordinates (r,ϕ,$\zeta(x, y)$)

$$\begin{aligned} x &= r\cos\phi \\ y &= r\sin\phi \\ z &= \zeta(x,y) \end{aligned} \tag{3.31}$$

The transformation matrix between (x,y,z) and (r,ϕ,ζ) is

$$\begin{pmatrix} r \\ \phi \\ \zeta \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{3.32}$$

Let us now define the rotation matrices. $\mathbf{R_2}$ applies a rotation around the $\overrightarrow{y}$-axis to point the $\overrightarrow{z}$-axis toward the north pole and $\mathbf{R_3}$ applies a rotation around the $\overrightarrow{z}$-axis to make $\overrightarrow{x}$ lie in the $\overrightarrow{u}$-$\overrightarrow{w}$ plane. The $\mathbf{R_2}$ and $\mathbf{R_3}$ matrices are defined as

$$\mathbf{R_2} = \begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix} \qquad \mathbf{R_3} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.33}$$

where α is the co-latitude of the telescope and θ is the hour angle at the vertex. Only the matrix $\mathbf{R_3}$ varies with time since $\theta = \Omega t$, where Ω is the angular rotation speed of the Earth ($7.27 \cdot 10^{-5}$ rad/s). The first and second derivatives of this matrix are given by

$$\dot{\mathbf{R}}_\mathbf{3} = \Omega \begin{pmatrix} -\sin\theta & -\cos\theta & 0 \\ \cos\theta & -\sin\theta & 0 \\ 0 & 0 & 0 \end{pmatrix} \equiv \Omega\mathbf{R_5} \qquad \ddot{\mathbf{R}}_\mathbf{3} = -\Omega^2 \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 \end{pmatrix} \equiv -\Omega^2\mathbf{R_6} \tag{3.34}$$

As we have seen in the previous cases, the slope of the equipotential surface is proportional to the resultant of the gravitational acceleration $\mathbf{G}$ and the centrifugal force $\mathbf{F}$:

$$\begin{pmatrix} \partial\zeta/\partial r \\ \partial\zeta/\partial\phi \\ -1 \end{pmatrix} = k(\mathbf{G} + \mathbf{F}) \tag{3.35}$$

If we consider the acceleration of gravity as being constant, uniformly oriented along $\overrightarrow{z}$ and pointing downward, we have

$$\mathbf{G} = \begin{pmatrix} 0 \\ 0 \\ -g_0 \end{pmatrix} \tag{3.36}$$

The centrifugal acceleration is equal to the opposite of the second derivative of the position, that is given in the $\mathbf{U}$ frame by

$$\ddot{\mathbf{U}} = -\Omega^2\mathbf{R_6}\mathbf{R_2}(\mathbf{X} + \mathbf{T}) + 2\Omega\mathbf{R_5}\mathbf{R_2}\dot{\mathbf{X}} + \mathbf{R_3}\mathbf{R_2}\ddot{\mathbf{X}} \tag{3.37}$$

Projecting this acceleration into the $\mathbf{X}$ reference frame by multiplying $\ddot{\mathbf{U}}$ by $\mathbf{R_2^{-1}}\mathbf{R_3^{-1}}$, we get the centrifugal force $\mathbf{F}$

$$\mathbf{F} = \Omega^2(R_2^{-1}R_3^{-1}R_6R_2)(\mathbf{X}+\mathbf{T}) - 2\Omega\mathbf{R_2}\mathbf{R_3^{-1}}\mathbf{R_5}\mathbf{R_2}\dot{\mathbf{X}} + \ddot{\mathbf{X}} \tag{3.38}$$

where $\dot{\mathbf{X}}$, $\ddot{\mathbf{X}}$ are defined by

$$\dot{\mathbf{X}} = \omega \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix} \qquad \ddot{\mathbf{X}} = -\omega^2 \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \tag{3.39}$$

Replacing $\mathbf{G}$ and $\mathbf{F}$ in equation 3.35 and computing the matrix products $\mathbf{R_2^{-1}}\mathbf{R_3^{-1}}\mathbf{R_6}\mathbf{R_2}$ and $\mathbf{R_2^{-1}}\mathbf{R_3^{-1}}\mathbf{R_5}\mathbf{R_2}$, we get

$$\begin{pmatrix} \partial\zeta/\partial r \\ \partial\zeta/\partial\phi \\ -1 \end{pmatrix} = k \left\{ \begin{pmatrix} \omega^2 r \\ 0 \\ -g_0 \end{pmatrix} + 2\Omega\omega r \begin{pmatrix} \cos\alpha \\ 0 \\ \cos\phi\sin\alpha \end{pmatrix} \right.$$

$$\left. +\Omega^2 \begin{pmatrix} r\cos^2\phi\cos^2\alpha + r\sin^2\phi - R_E\cos\phi\cos\alpha\sin\alpha \\ -r^2\sin\phi\cos\phi\cos^2\alpha + r^2\sin\phi\cos\phi + rR_E\sin\phi\cos\alpha\sin\alpha \\ -r\cos\phi\cos\alpha\sin\alpha + R_E\sin^2\alpha \end{pmatrix} \right\} \tag{3.40}$$

We solve the z-equation to find k

$$k = \frac{1}{g'}(1 + hr\cos\phi)^{-1} \tag{3.41}$$

where $g' \equiv g_0 - \Omega^2 R_E \sin^2\alpha$ and $h \equiv \Omega\sin\alpha(\Omega\cos\alpha - 2\omega)/g'$. The r-equation becomes

$$\frac{\partial\zeta}{\partial r} = \frac{[\omega^2 + \Omega^2\sin^2\phi + \Omega^2\cos^2\phi\cos^2\alpha + 2\omega\Omega\cos\alpha]r - [\Omega^2 R_E\cos\phi\cos\alpha\sin\alpha]}{g'(1+hr\cos\phi)} \tag{3.42}$$

Integrating this equation along r gives

$$\begin{aligned} \zeta &= \frac{\omega'^2}{g'}\frac{1}{\cos\phi h}\left[r - \frac{\ln(1+r\cos\phi h)}{\cos\phi h}\right] \\ &- \frac{\Omega^2\sin^2\alpha}{g'}\frac{\cos\phi}{h}\left[r - \frac{\ln(1+r\cos\phi h)}{\cos\phi h}\right] \\ &- \frac{\Omega^2 R_E\cos\alpha\sin\alpha}{g'}\frac{\ln(1+r\cos\phi h)}{h} \end{aligned} \tag{3.43}$$

Using the series expansion of the Napierian logarithm limited to the third order and simple trigonometry, we finally get

$$\begin{aligned}
\zeta &= r^2\left(\frac{\omega'^2}{2g'} - \frac{\Omega^2\sin^2\alpha}{4g'} + \frac{\Omega^2 R_E\cos\alpha\sin\alpha}{4g'}h\right) \\
&- r\cos\phi\left(\frac{\Omega^2 R_E\cos\alpha\sin\alpha}{g'}\right) \\
&+ r^2\cos(2\phi)\left(-\frac{\Omega^2\sin^2\alpha}{4g'} + \frac{\Omega^2 R_E\cos\alpha\sin\alpha}{4g'}h\right) \\
&+ r^3\cos\phi\left(-\frac{\omega'^2}{3g'}h + \frac{\Omega^2\sin^2\alpha}{4g'}h - \frac{\Omega^2 R_E\cos\alpha\sin\alpha}{4g'}h^2\right) \\
&+ r^3\cos(3\phi)\left(\frac{\Omega^2\sin^2\alpha}{12g'}h - \frac{\Omega^2 R_E\cos\alpha\sin\alpha}{12g'}h^2\right) \qquad (3.44)
\end{aligned}$$

where $\omega'^2 = \omega^2 + \Omega^2 + 2\omega\Omega\cos\alpha$ is the square modulus of the vectorial sum of the angular speeds of the mirror $\overrightarrow{\omega}$ and of the Earth $\overrightarrow{\Omega}$. A simple analysis of the first term reveals that the main effects of the Earth rotation are changes in the rotation speed and in the rotation axis (the vectorial sum of both rotations involved) and a "variation" in the gravitational acceleration that is reduced by the centrifugal acceleration on the surface of the Earth. The first order shape (first term) is still a parabola with a slightly different focal length.

$$F = F_0\frac{1 - \Omega^2 R_E\sin^2\alpha/g}{1 + [\Omega^2 + 2\Omega\omega\cos\alpha - 0.5\Omega^2\sin^2\alpha + 0.5\Omega^2 R_E\cos\alpha\sin\alpha h/2]/\omega^2} \qquad (3.45)$$

where F_0 is the zero order focal length. In the case of the ILMT, located at the Devasthal observatory, the focal length is reduced by about 2.77cm. The variation of the focal length as a function of the colatitude of the observatory is presented in fig. 3.6. Even at the poles, the focal length is modified, this is due to the addition of the Earth rotation to the mirror rotation.

The other terms in equation 3.44 can be expressed as a function of the Zernike polynomials (tilt, astigmatism, coma and trefoil). The conversions are

$$\begin{aligned}
r\cos\phi &= \frac{Z_2}{2} \\
r^2\cos(2\phi) &= \frac{Z_6}{\sqrt{6}} \\
r^3\cos(\phi) &= \frac{Z_8}{3\sqrt{8}} + \frac{Z_2}{3} \\
r^3\cos(3\phi) &= \frac{Z_{10}}{\sqrt{8}} \qquad (3.46)
\end{aligned}$$

$$(3.47)$$

In the case of the ILMT (at Devasthal), the values of these aberrations are summarized in Table 3.4. Their variations as a function of the colatitude are shown in fig. 3.7 to 3.10. The aberration surface (without the tilt) is presented in fig. 3.11.

Aberration	Polynomial	Intensity (λ)
Tilt	Z_2	-54.72
Astigmatism	Z_6	$-8.338 \cdot 10^{-5}$
Coma	Z_8	$4.712 \cdot 10^{-3}$
Trefoil	Z_{10}	$-1.621 \cdot 10^{-10}$

Table 3.4: Intensities of the aberrations caused by the rotation of the Earth expressed in wavelength unit (at 762.5 nm). Apart from the tilt, these aberrations are very small. Their variations with the colatitude of the mirror are presented in fig. 3.7 to 3.10

The perturbations introduced by the Coriolis force could cause significant aberrations in the images produced by the ILMT. Hickson (2001) presented a convenient way of compensating for the Coriolis effect by simply tilting the axis of rotation of the liquid mirror. The required angle is a function of the latitude and of the mirror rotation rate. We have seen that the effective rotation vector $\overrightarrow{\omega'}$ of the mirror is the vectorial sum of the rotation vector of the mirror $\overrightarrow{\omega}$ and the one of the Earth $\overrightarrow{\Omega}$. The effective axis and rotation rate are thus slightly different. Moreover, we have seen that the effective acceleration of gravity $\overrightarrow{g'}$ is also composed of two parts, the actual acceleration of gravity $\overrightarrow{g}$ and a vector $\Omega^2 R_E \overrightarrow{e_r}$ that is the centrifugal acceleration due to the rotation of the Earth, at the level of the mirror ($\overrightarrow{e_r}$ is a unitary radial vector perpendicular to the rotation axis of the Earth). The effect of this acceleration is to reduce the effective acceleration of gravity $\overrightarrow{g'}$ felt by the liquid and to modify its orientation. In order to cancel the Coriolis force, the vector $\overrightarrow{\omega'}$ should be aligned with the vector $\overrightarrow{g'}$, as it is the case at the poles. We already saw that as these locations, the different aberrations vanish, precisely because of the alignment between the rotation axis of the mirror and the effective gravity. This alignment can be obtained anywhere, simply by tilting the actual axis of rotation of the mirror.

Let us call γ the angle between $\overrightarrow{g}$ and $\overrightarrow{g'}$ and β the angle between $\overrightarrow{\omega}$ and $\overrightarrow{\omega'}$, the required tilt angle is thus $\epsilon = \gamma - \beta$ with

$$\begin{aligned} \gamma &= \frac{\|\overrightarrow{g'} \times \overrightarrow{g}\|}{g'g} \approx \frac{\Omega^2 R_E \sin(2l)}{2g_0} \\ \beta &= \frac{\|\overrightarrow{\omega'} \times \overrightarrow{\omega}\|}{\omega'\omega} \approx \frac{\Omega \cos(l)}{\omega} \end{aligned} \tag{3.48}$$

where l is the latitude of the observatory. The value of the tilt angle as a function of the latitude is shown in fig. 3.12. The compensation tilt (ϵ) has to be applied toward the nearest pole (North in the northern hemisphere and South in the southern hemisphere).

In practical cases, the local vertical is measured with a bubble level that is sensitive to the effective gravity $\overrightarrow{g'}$. The required correction angle is then given by β instead of ϵ and the tilt has to be done toward the equator. The ILMT located at Devasthal will require a tilt of its rotation axis by ~ 16.7" measured from the effective vertical toward the South. The tilt angle from the zenith will be about $4'02$" toward the North. It is important to note that this tilt implies that the strip of sky that will be observed with the telescope is shifted by 4 arcmin Northward.

Let us now evaluate what would be the exact rotation period of the mirror to ensure that the mirror has a given focal length (8m in our case). We have computed the values corresponding to

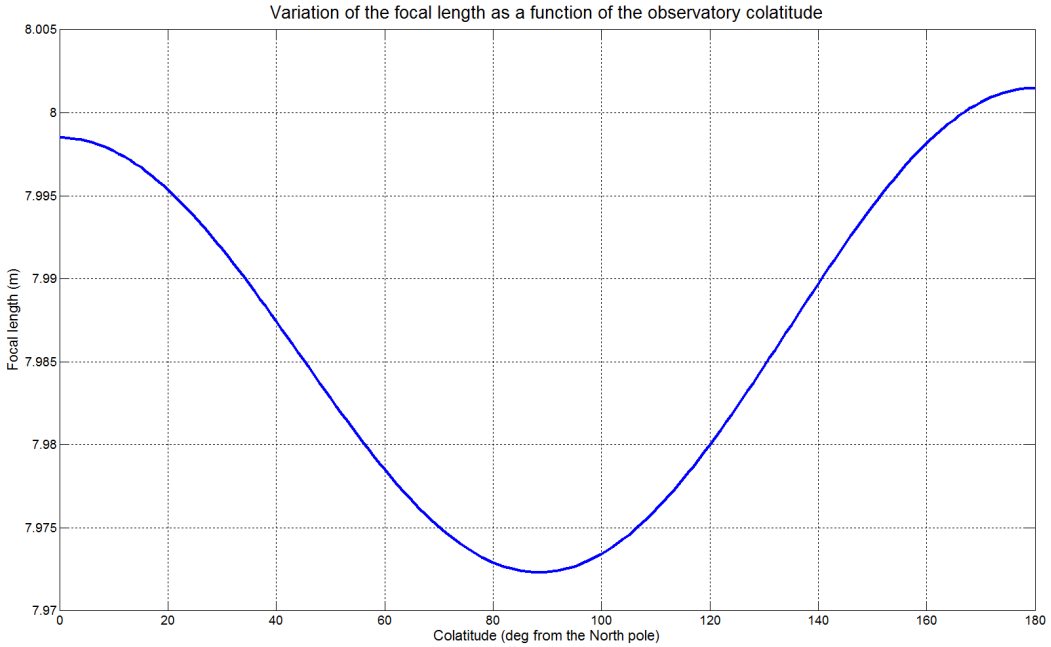


Figure 3.6: Variation of the focal length as a function of the observatory colatitude. A liquid mirror located at one of the pole ($\alpha = 0°or180°$) would still present a modified focal length, because of the addition of the rotation of the Earth to the one of the mirror.

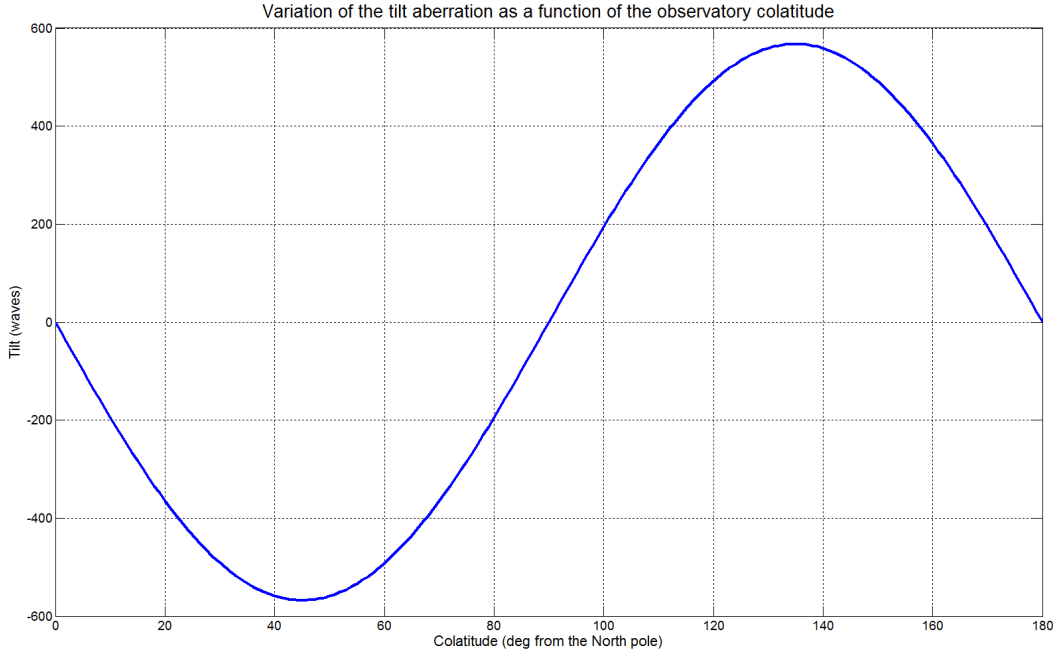


Figure 3.7: Variation of the tilt aberration due to the Coriolis effect as a function of the colatitude of the observatory. At Devasthal, the aberration would reach -54.72λ. As the centrifugal acceleration related to the rotation of the earth is null at the poles, the local gravity is not deviated and the two rotations occur around the same axis, the tilt thus vanishes at those locations. At the equator, the centrifugal acceleration is oriented in the opposite direction to that of gravity. The direction of effective gravity is thus the same as the actual gravity and the tilt aberration also vanishes.

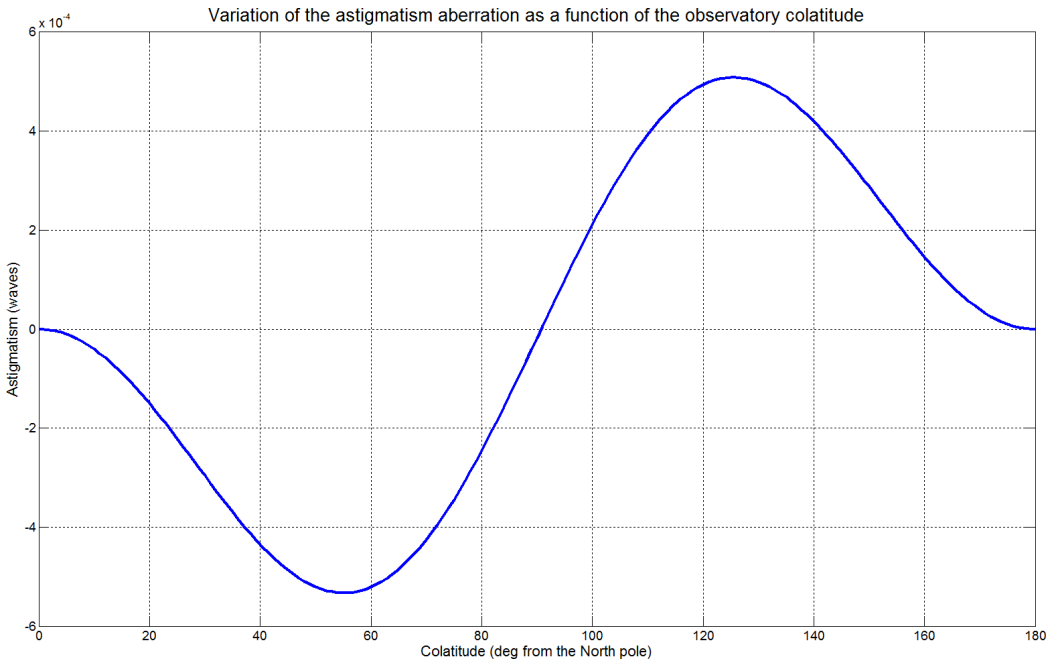


Figure 3.8: Variation of the astigmatism aberration due to the Coriolis effect as a function of the colatitude of the observatory. At Devasthal, the aberration will be $-8.338 \cdot 10^{-5}\lambda$. This aberration vanishes both at the poles and at the equator, this is due to the alignment of the rotation axis of the mirror with the effective local gravity at those particular locations, as in the tilt case.

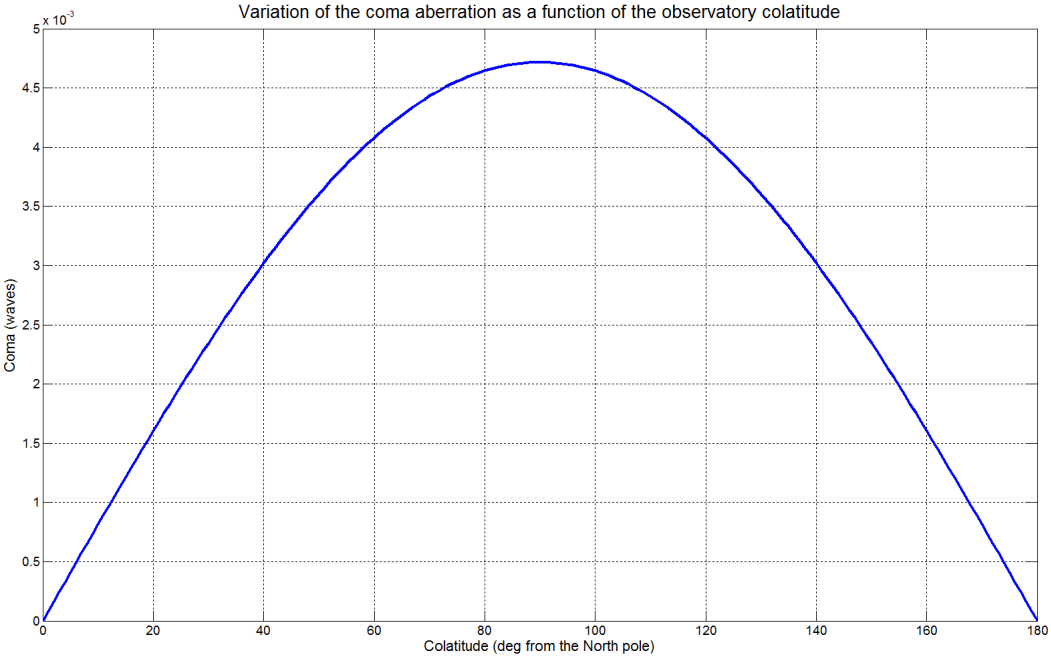


Figure 3.9: Variation of the coma aberration due to the Coriolis effect as a function of the colatitude of the observatory. At Devasthal, the aberration will be $4.712 \cdot 10^{-3}\lambda$, it is null at the poles and maximum at the equator. The variation of this aberration is related to the difference between the vectorial sum of the two rotations and the projection of this sum on the mirror rotation axis. This one is maximal at the equator and null at the poles.

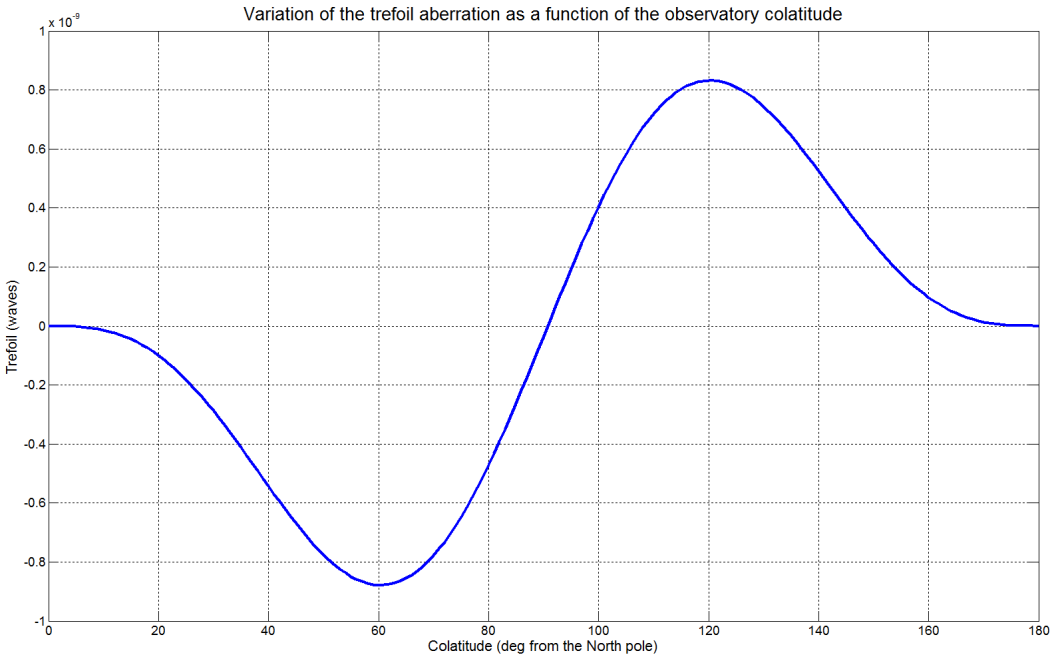


Figure 3.10: Variation of the trefoil aberration due to the Coriolis effect as a function of the colatitude of the observatory. At Devasthal, the aberration will be $-1.621 \cdot 10^{-10}\lambda$, As for the other even aberrations (cosine function), this one is null at the poles and at the equator.

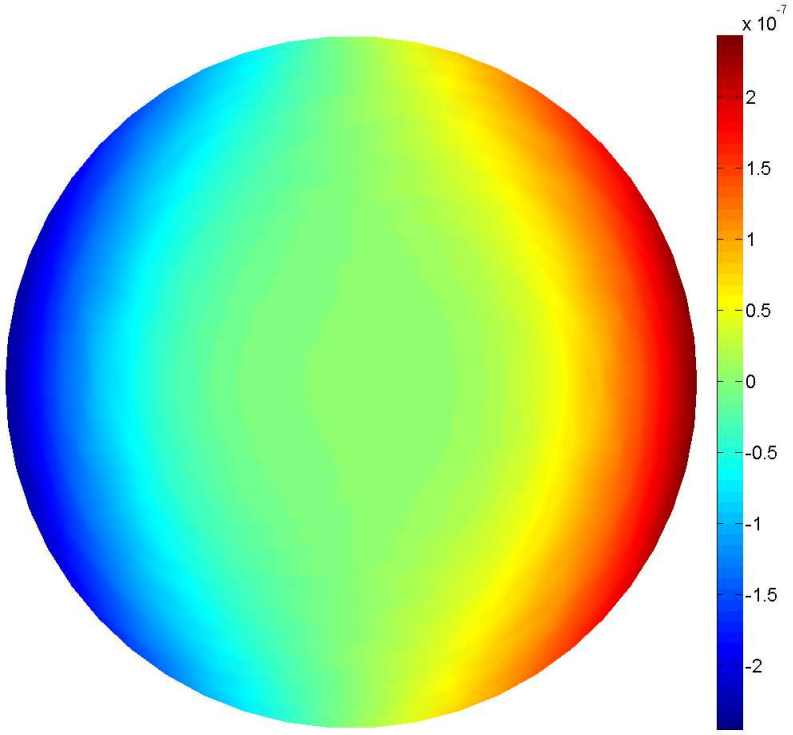


Figure 3.11: Difference between the surface disturbed by the Earth rotation and the perfect parabola. The tilt aberration has also been subtracted. This surface represents the sum of the astigmatism, coma and trefoil aberrations for a liquid mirror located at Devasthal. The intensity is represented in wavelength units.

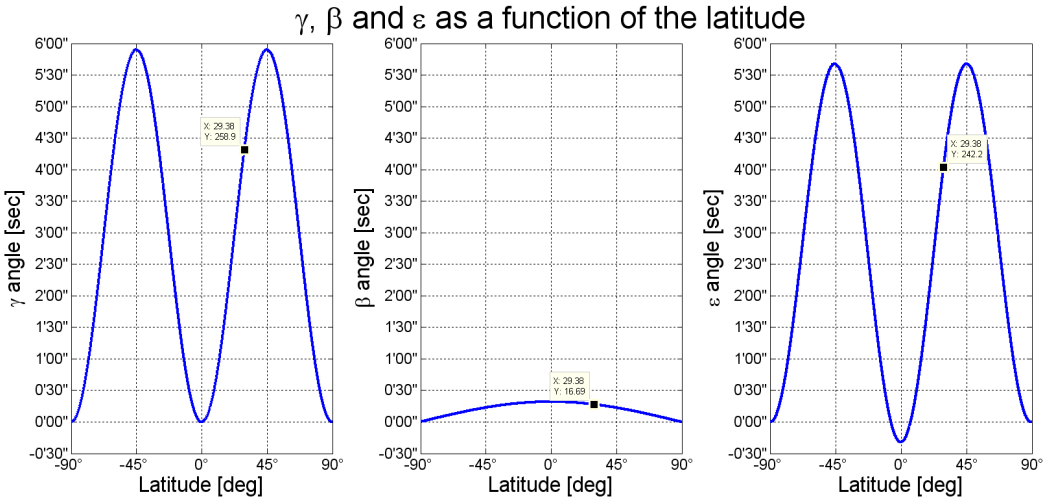


Figure 3.12: Variation of the γ, β, ϵ angles as a function of the latitude of the observatory. The $\epsilon = \gamma - \beta$ angle is the tilt angle, measured from the zenith, that is required to compensate for the Coriolis effect. The tilt angle required at Devasthal is about $4'02"$ toward the North. The compensation angle is toward the North in the Northern hemisphere and toward the South in the southern hemisphere (Chile observatories). In practical cases, the rotation axis is aligned with the local gravity that is measured with a bubble level. Such a device is sensitive to the effective gravity g'. In this case, the correction required is only the β angle toward the equator. The ILMT will require a β tilt of $\sim 16.7"$ toward the South at Devasthal.

an ILMT located in Liège (AMOS) and at Devasthal (Observatory). The two sets of results are presented in Table 3.5, with and without the tilt correction presented above. The angular speeds calculated here only account for the Coriolis effect, but we have seen that this is the dominant one.

Location	Liège		Devasthal	
Latitude	50°35'58.6"		29°23'46"	
Altitude [m]	215		2403	
$\overrightarrow{g_e ff}$ $[m/s^2]$	9.81057764		9.78540973	
Tilt correction	No	Yes	No	Yes
Angular speed $[rad/s]$	0.78296526	0.78296518	0.78198081	0.78198071
Rotation period [s]	8.02485832	8.02485915	8.03496104	8.03496206
Correction angle ϵ	5'37.5646"		4'49.0832"	
Correction angle β	12.1937"		16.7581"	

Table 3.5: Optimal angular speed in order to obtain a precise focal length (8m) as well as the parameters used to compute it. The results are given for Liège and Devasthal with and without the tilt correction presented before. The geoid model used to compute the effective gravitational acceleration can be found on the web page http://www.gfy.ku.dk/~iag/handbook/geodeti.htm

3.5 Wind induced spiral waves

The rotation of the mirror results in the creation of an apparent wind above the mercury layer. Indeed, the air layer above the mercury does not move while the mercury itself is rotating. In the reference frame of the mercury, the air thus seems to move. This apparent wind may disturb the mercury layers by exciting spiral waves. These waves have been observed on several liquid mirrors (the NODO mirror (Mulrooney 2000), the 3.7m mirror at Laval University mirror (Tremblay and Borra 2000), the LZT mirror (Hickson and Racine 2007), ...)

This type of waves only appears for radii larger than a limit value corresponding to a critical wind speed. Two approaches are considered in Mulrooney (2000) to determine the speed that corresponds to the appearance of the turbulence. The first one consists in comparing the expression of the Reynolds number (equation 3.49) as a function of the radius with a critical value (R_e^{crit}) that defines the limit between the laminar and turbulent regimes.

$$R_e = \frac{\omega r^2}{\nu} \tag{3.49}$$

where ω is the angular rotation speed of the system and ν is the kinematic viscosity of the mercury. For radii corresponding to $R_e > R_e^{\text{crit}}$ the regime is turbulent and spiral waves appear on the mercury layer. The radius computed this way for the NODO ($\sim 47cm$) is slightly underestimated when compared to the measured radius ($\sim 64.5cm$) where waves effectively appear.

The other approach consists in using the expression derived in Landau and Lifshitz (1959) for the stability of a tangential discontinuity between two fluids of different densities (ρ_1, ρ_2) and surface tensions (α_1, α_2), moving relatively to each other. If α_2 is negligible, the relation is

$$U_{\text{crit}} \leq \left(\frac{4\alpha_1 g(\rho_1 - \rho_2)(\rho_1 + \rho 2)^2}{\rho_1^2 \rho_2^2} \right)^{1/4} \tag{3.50}$$

The flow is laminar when the relative velocity (U) between the fluids is smaller than the critical value U_{crit} (equation 3.50). Considering the NODO case, this approach leads to a critical radius of 67.3 cm that fits ($\sim 64.5cm$) the measured value quite well. The same equation leads to a value around 92.5cm for the ILMT.

The other parameters like frequency or wavelength are not well described by the theoretical models considered by Mulrooney (2000). Nevertheless, he gives empirical equations that describe the shape of the spirals. Using these equations, François Finet proposed a basic three dimensional parametric model to describe the variation of the LM shape with respect to the parabola.

$$\begin{aligned} z &= A(r)\cos(N(\phi + \phi_m)) \\ \phi_m &= 2\left(1 - \frac{r}{R}\right) \end{aligned} \tag{3.51}$$

where $A(r)$ is the amplitude function, ϕ_m is the angular position of the maximum of the wave, N is the number of waves and R is the radius of the mirror. The orientation of the spirals depends on the sign of ϕ_m. If it is positive, the spirals are oriented counter clock-wise, when it is negative, the spirals are oriented clock-wise.

The final model has been improved to account for time variations. Following the measurements of Mulrooney, these waves are almost stationary. They rotate somewhat more slowly than

the mirror. We thus implement a parameter of rotation for the mirror (ωt) in the cosine term of the first equation 3.51. The other effect is the pulsation of the anti-nodes that is modeled by adding a cosine oscillation in the amplitude function, which then becomes a function of (r, t). The amplitude measurements achieved by Mulrooney on the NODO mirror (using the Ronchi technique) allow to think that the amplitude increases linearly with the radius. However, adding a point with null amplitude at the radius where the waves begin to appear (determined either by the Reynolds or the Landau approach) leads to a quadratic increase of the amplitude. Both are implemented in our model but the quadratic law is used for the simulations that follow.

$$\begin{aligned} A(r,t) &= (C_1 r^2 + C_2 r + C_3) \cdot \cos(2\pi t/T) \qquad r \geq r_{\text{crit}} \\ A(r,t) &= 0 \qquad r < r_{\text{crit}} \end{aligned} \tag{3.52}$$

where C_1, C_2, C_3 are constants that are computed to account for the critical radius (r_{crit}) where the waves appear and for the maximum amplitude of the waves (A_{max}) at the edge of the mirror. The derivative of the amplitude function with respect to r has to be continuous at $r = r_{\text{crit}}$, and is thus set to zero. T is the period of the amplitude variation.

The model we have developed allows to simulate the shape of the mirror (the variation with respect to a perfect parabola) for various cases. We have parameterized the number (N) and maximum amplitude (A_{max} - at the edge of the mirror) of the waves, the minimum radius (r_{crit}) where they appear and their amplitude variation period (T). We have studied the effect of these parameters on the quality of the image. The results are presented hereafter. A comparison between the model (right) and the spiral waves observed on a 1.5m f/2 liquid mirror in Borra's laboratory (left) is presented in fig. 3.13. The similitudes between the model and the observations are clearly visible.

In order to compute the PSF obtained during a given integration time (typically 100s for the ILMT) with a liquid mirror whose surface is disturbed by spiral waves, we sum up the PSFs corresponding to several particular times t (500 steps between 0 and 100s). We then compute the corresponding Strehl ratio.

Let us now use this model the investigate the impact of the maximal amplitude of the wave (at the edge of the mirror) on the image quality. We let A_{max} vary between 0 and 4 λ ($\lambda = 762.5nm$ - center of the i' band) and we measure the corresponding Strehl ratio. The result obtained with 37 waves beginning at 1m (approximated value estimated from equation 3.50 for $U_{lim} = 72.4cm/s$) of the center and having a pulsation period $T = 1/3s$ is presented in fig. 3.14. As expected, the Strehl ratio decreases when the amplitude of the waves increases. A realistic amplitude of 300nm $\sim \lambda/2$ (see Mulrooney 2000) causes about 5% of the light to be diffused in a halo surrounding the central peak. More and more light is diffused outside the peak when the amplitude of the waves increases.

The following step consists in studying the effect of the critical radius (r_{crit}). Letting it vary between 10cm and 1m, once again we compute the corresponding Strehl ratio. The result presented in fig. 3.14, shows that the Strehl ratio increases when the wave-free zone radius increases, as logically expected. We can thus deduce that a slowly rotating mirror would give better images (with higher Strehl ratio) than a rapidly rotating mirror as the critical radius is inversely proportional to the rotation speed of the mirror.

Finally, we studied the impact of the number of waves disturbing the mirror on the image quality. The Strehl ratio does not seem to vary with this parameter, in the tested range (between 1 and 190 waves). The upper limit we have chosen corresponds to spiral waves having a wavelength

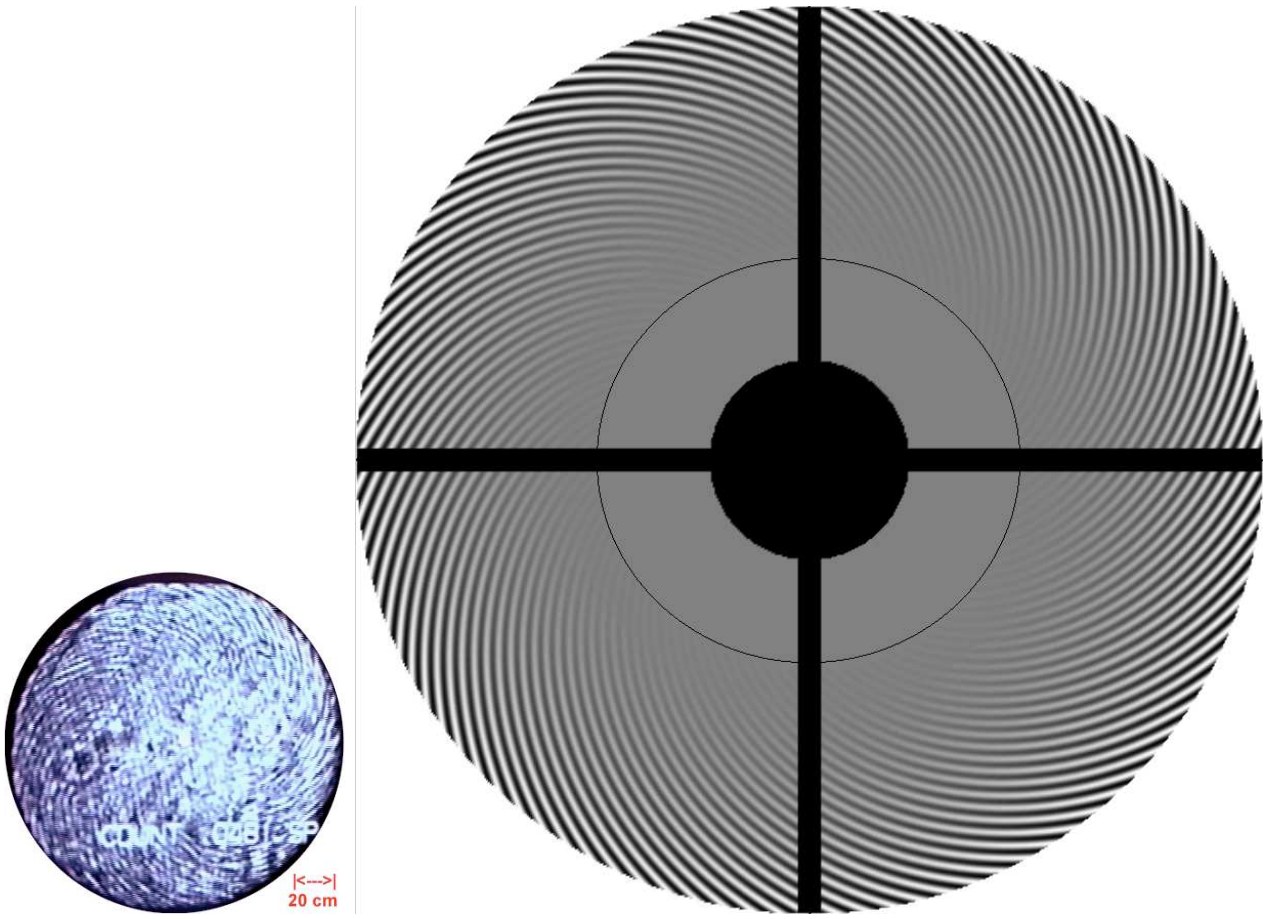


Figure 3.13: Left: Spiral waves observed on a 1.5m f/2 liquid mirror in Borra's laboratory. The mirror is rotating with an angular velocity $\omega = 1.28\,rad/s$ (figure from Mulrooney (2000)). Right: our model for the spiral waves on the surface of the mirror. This model is based on the observations presented in Mulrooney (2000). It contains 100 waves with a maximum amplitude of 300nm. The black circle represents the limit radius where the waves begin (90cm on this model). The similitudes between the model and the observations are clearly seen. The central obstruction and the spider of the ILMT telescope are also represented on the model.

of 5cm at the edge of a 3m mirror (the NODO) that fits with the observations of Mulrooney.

We then had a look at PSFs corresponding to seven different numbers of waves (1, 10, 35, 70, 110, 150, 190). As shown in fig. 3.15 the diffuse halo is farther and farther away from the central peak but the central peak does not vary. Indeed, increasing the number of waves is equivalent to increasing the space frequency of the disturbance but it does not really modify this disturbance. The location of the halo thus corresponds to higher frequencies in the Fourier plane (image), this means, further away from the center.

We conclude from these simulations that the spiral waves diffuse the light from the central peak of the PSF in a surrounding halo. The amount of light that is scattered depends on the amplitude ($A_{\max}$) of the waves and on the fraction of the radius (r_{crit}) that is disturbed. The mean radius of the halo depends on the number of waves disturbing the mirror. They are given in table 3.6.

Hickson and Racine (2007) showed that these wind-induced spiral-waves could be almost suppressed by covering the mirror with a mylar sheet. Even if this protective mylar and the structure that supports it also diffuses some light, the scattering related to a thin mylar sheet is

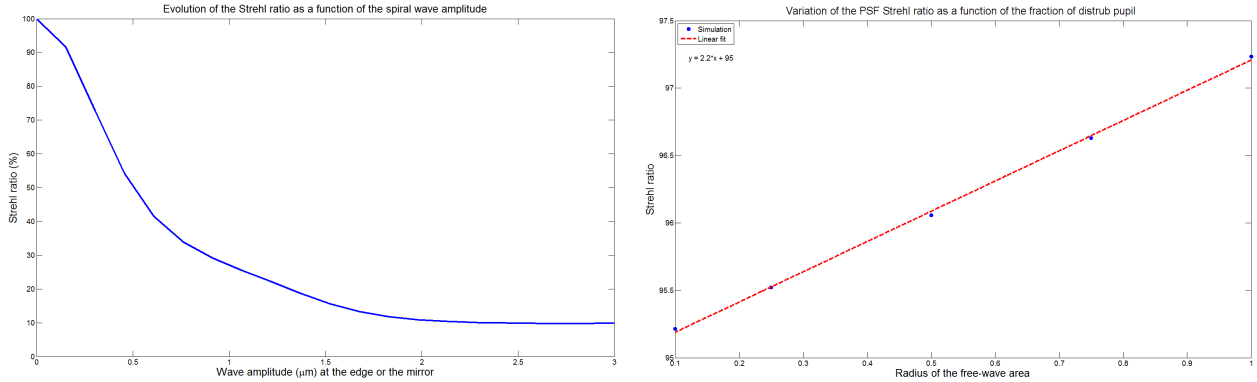


Figure 3.14: Left: Evolution of the Strehl ratio of the images obtained at the focus of a 4m liquid mirror disturbed with 37 spiral waves as a function of the edge amplitude of these waves. Spiral waves of realistic edge amplitude (300nm) cause about 5% of the light to be diffused outside the peak of the PSF. Right: Evolution of the Strehl ratio as a function of the critical radius ($A_{max} = 300nm$). A wave free area of about 1-m in radius would lead to a Strehl ratio of 97% (provided the other parameters correctly fit the real waves). The equation of the linear fit is $y = 2.2x + 95$.

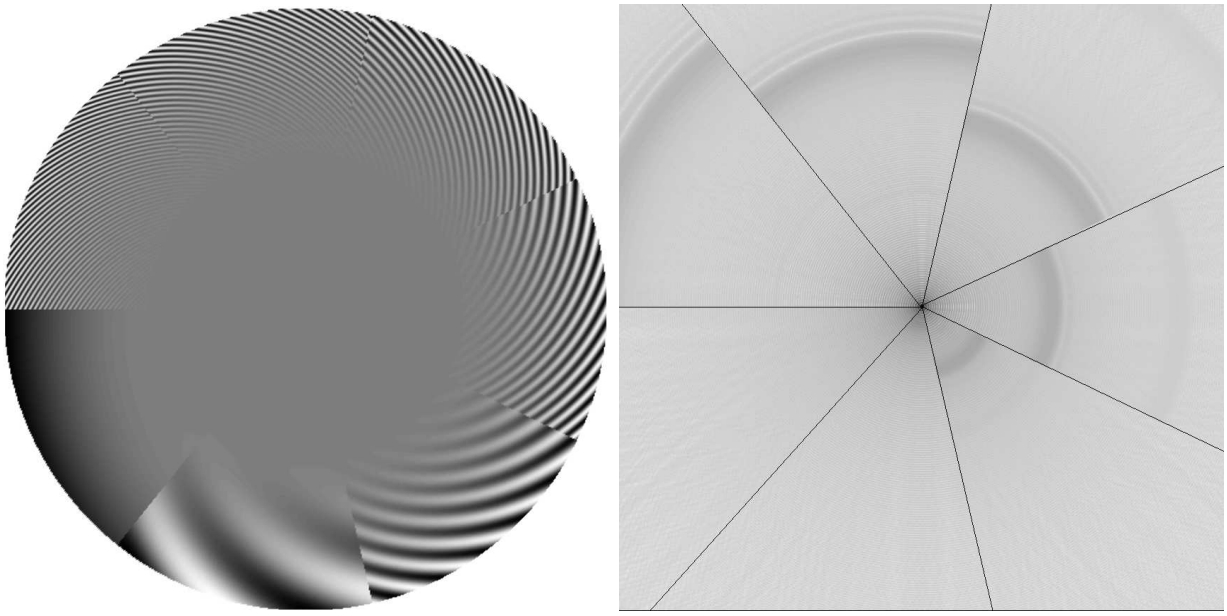


Figure 3.15: Left: Samples of the seven pupils used to compute the influence of the number of waves on the Strehl ratio. The white curves correspond to the peak of the waves. The number N of waves are, from the bottom left segment in counter-clockwise direction: 1, 10, 35, 70, 110, 150, 190, waves. The diameter of the pupil is 4m (as it corresponds to the ILMT primary mirror). Right: Segments of PSFs corresponding to the number of waves in the left figure. The presence of waves on the mirror diffuses light out of the central peak of the PSF. The diffused halo is farther and farther away from the center with more waves on the pupil (respectively 2.44, 20, 61.5, 126, 184, 250, and 314 λ/D). However, the Strehl ratio itself does not vary as a function of the number of waves ($\sim$ 95% in this particular case). Let us note that the colors have been inverted in order to emphasize the halo.

smaller than the one caused by the spiral waves present on the LZT. In the case of the ILMT, some measurements will be necessary to determine whether covering the mirror will be useful or not.

Number of waves	1	10	35	70	110	150	190
Distance (λ/D)	2.44	20	61.5	126	184	250	314

Table 3.6: Distance (λ/D) between the halo and the center of the PSF for the different cases considered in fig. 3.15.

3.6 Vibration induced concentric waves

Another type of waves has been observed on several liquid mirrors (Mulrooney 2000; Tremblay and Borra 2000). Vibrations transmitted to the dish produce concentric waves both at the center of the mirror (inner hub) and at its outer wall, propagating radially trough the mercury layer. Two types of concentric waves have been observed by Mulrooney (2000). Short wavelength waves emitted at the outer-edge of the mirror and propagating inward have been noticed on the NODO mirror when it was tested at the Johnson Space Center (JSC). These waves only exist at large radii and dissipate before reaching the center of the mirror. The characteristics of these waves, observed by Mulrooney, are given in Table 3.7. An illustration of such waves is presented in fig. 3.16 (left). These waves have not been observed during the operation of the NODO. Mulrooney concluded that they were probably forced by some environmental effects, specific to the JSC.

Long wavelength concentric waves have been observed at both locations of the mirror (JSC and NODO) with identical frequencies. These waves, also observed by Tremblay and Borra (2000), are probably related to the natural oscillation of the mirror container. Hickson et al. (1993) stated that the 2.7m UBC/laval LMT had a natural oscillation frequency of 18Hz when it was filled with a 2mm layer of mercury and 27Hz when it was empty. This is consistent with the 15Hz measured on the NODO mirror with 1.6mm of Hg. Obviously, the natural resonance frequency of the mirror depends on the thickness of the mercury layer.

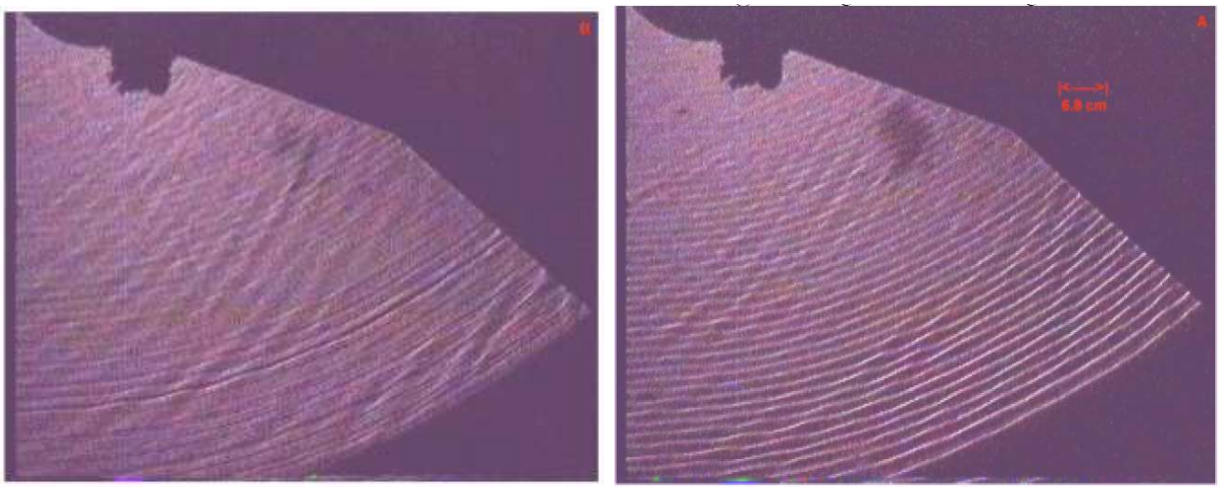


Figure 3.16: Illustration of the two types of concentric waves observed at larger radii ($r > 75cm$) on the NASA-LMT (figures from Mulrooney 2000). Left: Short wavelength concentric waves. Right: Long wavelength concentric waves. Both images have been acquired under point-source illumination at the radius of curvature. Spiral waves are also visible on these images.

The characteristics of the long wavelength concentric waves observed on the 3m NODO mirror are also given in Table 3.7, and they are illustrated in fig. 3.16 (right). They originate from both

the central hub (propagating outward) and the periphery (propagating inward).

Characteristics	Short wavelength	Long wavelength
Wavelength (λ)	0.68cm	1.15 - 2 cm
Frequency ($\omega/2\pi$)	45Hz	15Hz

Table 3.7: Characteristics of the short and long wavelength waves observed on the NODO liquid mirror.

The correspondence between the observational measurements presented in Mulrooney (2000) and the theoretical models they used is not obvious. Since the physical models do not seem to describe the observed behavior of the concentric waves, we decided to build a parametrical model that will be used to estimate the impact of concentric waves on the image quality. As for the spiral waves, the values obtained with the NODO observations will be used to calibrate our model. The values that will be measured with the ILMT will be used, in order to fine-tune the model and to simulate realistic perturbations.

3.6.1 Modeling

Our model is based on the one dimension differential equation that describes the evolution of damped waves.

$$\frac{\partial^2 z}{\partial t^2} + 2\zeta\omega\frac{\partial t}{\partial r} + \omega^2 z = F \tag{3.53}$$

where ζ is the damping parameter, ω is the natural pulsation [17] of the system and F is the forcing term that can be written as a sum of oscillations of given amplitude (A_i) and frequency (ω_i)

$$F = \sum_i A_i \cos(\omega_i t + \phi_i) \tag{3.54}$$

The solution of this type of differential equation is given by the sum of the general solution of the homogeneous equation (z_g 3.55) and the particular solution of the heterogenous equation (z_p 3.56)

$$z_g = \exp -\zeta\omega t[C_1\cos(\sqrt{1-\zeta}\omega t) + C_2 sin(\sqrt{1-\zeta}\omega t)] \tag{3.55}$$

$$z_p = \sum_i \frac{A_i}{\sqrt{(\omega^2-\omega_i^2)^2 + 4\zeta^2\omega^2\omega_i^2}}\cos(\omega_i t + \phi_i - \delta_i) \tag{3.56}$$

$$\delta_i = \arctan\left(\frac{2\zeta\omega\omega_i}{\omega^2-\omega_i^2}\right)$$

where C_1, C_2 are constants that can be determined using the boundary conditions. The general solution z_g is a transitory solution as it is damped with time. Since we want to study the effect of the concentric waves on a liquid mirror during the operation of the telescope, we can assume that the transitory phase is over. Indeed, the oxide layer should be formed and the exceeding

[17] The pulsation ω is related to the frequency f by the formula $\omega = 2\pi f$. In the following, we will mainly use the term "frequency" instead of pulsation for ω as the frequency is more intuitive than the pulsation. Anyway, the 2π term is taken into account in the simulations.

mercury should have been pumped out so that a sufficient amount of time has passed to consider that the transitory waves have disappeared.

The only waves to consider are thus given by the particular solution z_p (equation 3.56) that is the permanent response of the system. The square root in the denominator of the fraction represents the amplitude coupling between the forcing oscillation and the oscillation propagating in the fluid. This coupling is presented in fig. 3.17 (left) as a function of the natural frequency of the fluid and of the excitation frequency. δ_i is the phase coupling between the forcing and the fluid response. It is presented in fig. 3.17 (right).

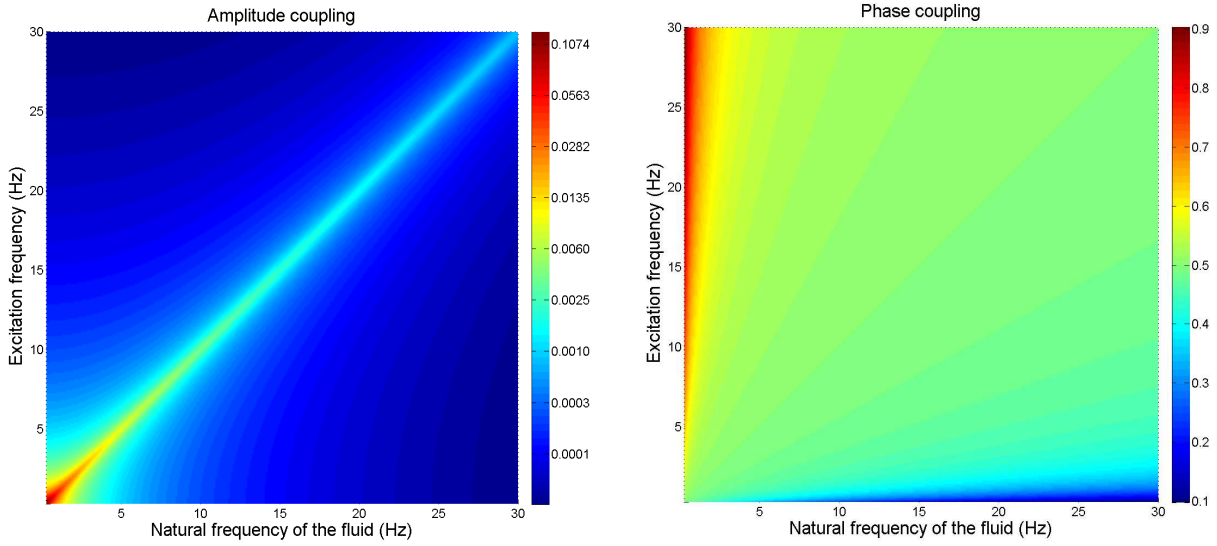


Figure 3.17: Left: Amplitude coupling (color scale $10/\sqrt{}$). Right: Phase coupling (colorbar expressed in π unit).

Moreover, the concentric waves we want to model propagate along the radius of the mirror. In one dimension, the equation of a propagating wave must satisfy the following differential equation

$$\frac{\partial^2 z}{\partial r^2} - \frac{1}{c^2}\frac{\partial^2 z}{\partial t^2} = 0 \tag{3.57}$$

where c is the velocity of the wave. The solution of this type of equation must be in the form

$$z(r,t) = f(r+ct) + g(r-ct) \tag{3.58}$$

where $f(r,t), g(r,t)$ are two functions that propagate in opposite directions. In our case, these functions are the particular solutions of the non homogeneous equation (z_p). For sake of simplicity, we consider only the g function that propagates from the center of the mirror to its outer edge (from small radius toward large radius). We simulate wave propagating inward ($f(r,t)$) by inverting our radius axis.

We also add a term of space damping ($\exp(-dr)$). Indeed, even if the forcing oscillation does not damp with time, the wave that propagates through mercury is damped during its travel. Each wave in our model is thus represented by

$$z_i(r,t) = \frac{A_i \exp(-dr)}{\sqrt{(\omega^2-\omega_i^2)^2 + 4\zeta^2\omega^2\omega_i^2}} \cos(\omega_i t - kr + \delta_i + \phi_i) \tag{3.59}$$

These waves are supposed to be reflected (if they are not damped before) when they reach either the outer edge or the central hub of the mirror. Each reflection is total, the reflected wave has thus the same amplitude as the incident wave and the phase is conserved.

In order to cope with the observations of Mulrooney (2000) and Tremblay and Borra (2000) our model allows to generate several waves at different frequencies and amplitudes both at the center and at the outer edge of the mirror. The parameters of our model are thus the inner and outer radius of the mirror (position of the sources and reflections - r_{min}, r_{max}), the natural frequency of the fluid (ω), the damping term (d) and the description of the forcing terms (A_i, ω_i, ϕ_i).

Some of these parameters are linked to physical parameters of the mirror, as the damping that is a function of the thickness of mercury. However, as these relations are not precisely known, they are optional in our model.

The wavelength and frequency are related together by a dispersion relation. Landau and Lifshitz (1959) give an equation for waves propagating on a liquid surface

$$\omega^2 = \left(gk + \frac{\alpha k^3}{\rho} \right) \tanh(kh) \tag{3.60}$$

where g is the local acceleration of gravity, $k = 2\pi/\lambda$ is the wave-number, α is the surface tension of the fluid ($\alpha_{Hg} = 0.485\text{N/m}$) and ρ is the density of the fluid ($\rho_{Hg} = 13.546 \cdot 10^3 \text{Kg/m}^3$). Mulrooney (2000) proposed another equation (derived by Hickson) that takes the rotation of the fluid (the rotation of the mirror, $\Omega = 0.781\, rad/sec$) into account

$$\omega^2 = 4\Omega^2 + ghk^2 + \frac{\alpha h k^4}{\rho} \tag{3.61}$$

The first term is related to the rotation of the mirror, the second term represents the gravity waves and the third one refers to capillarity waves.

The modeled surface of the mirror corresponds to the sum of the waves created by all the excitation sources and all the reflections of those waves. Fig. 3.18 shows the radial waves emitted from the center (top left) and propagating outward, those emitted from the outer edge (middle left) and propagating inward. Those waves are reflected on each boundary as many times as necessary so that the damping reaches 10^{-6} of the initial amplitude. The sum of both waves is presented in the bottom left part of the same figure, and the complete surface of the mirror is shown on the right. Values used to generate these images are typical values that do not represent real measurements. The amplitudes of the excitation are of the order of a few millimeters and the frequencies of the order of a dozen of Hertz. The natural frequency of the liquid is set to 2Hz and the damping coefficient is computed using the empirical law presented in Tremblay and Borra (2000), which based on their own measurements is

$$d = 15.8\ h^{-1.38} \tag{3.62}$$

where h is the thickness of mercury (mm). The simulation presented in fig. 3.18 assumes a thickness of 1mm. The values used in our model are consistent with the observations presented in Tremblay and Borra (2000).

The point spread function corresponding to 100s of integration (typical ILMT integration time) with a primary mirror disturbed by concentric waves is presented in fig. 3.19. The top

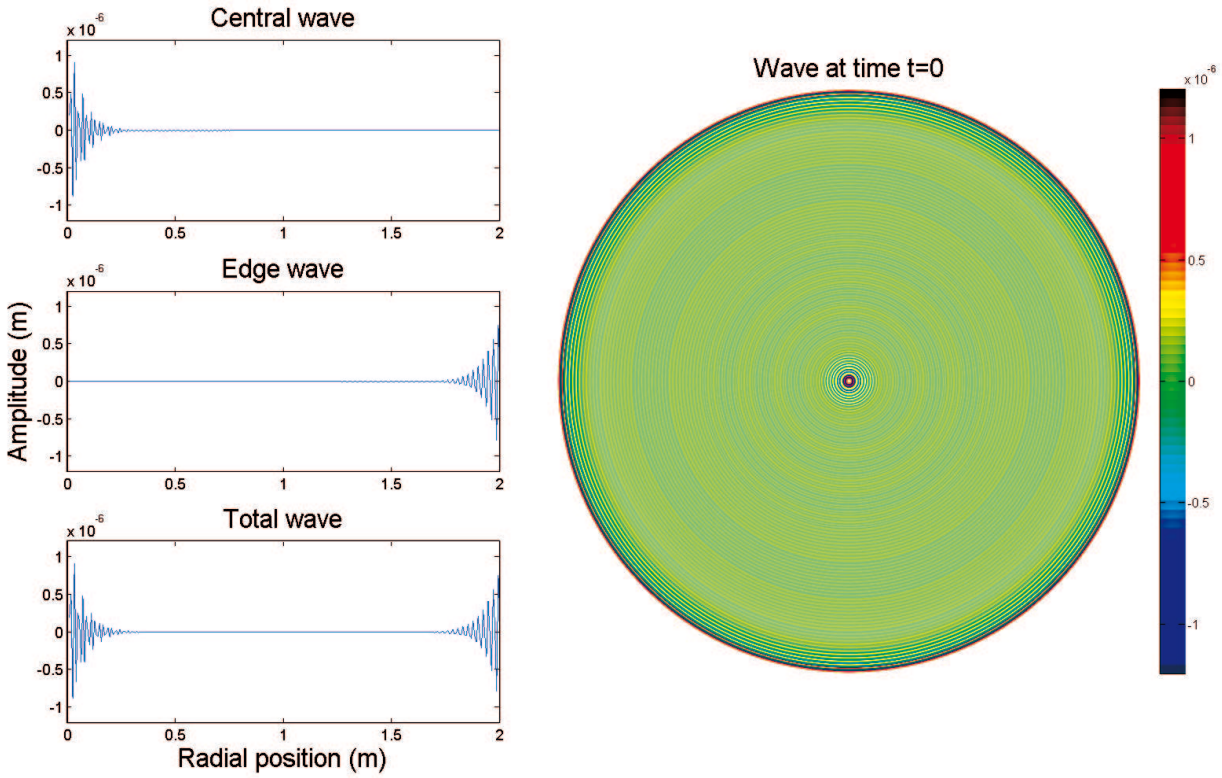


Figure 3.18: Model of concentric waves assuming two sources of excitation (at the center and at the edge of the mirror) with amplitudes of respectively 5mm and 6mm and frequencies of 10Hz and 12Hz. The mercury thickness is 1mm. The contrast of the waves on the disk has been artificially increased so that they become more visible.

figure presents the PSF that corresponds to a mirror disturbed by an excitation source whose amplitude is 1mm. The bottom figure corresponds to an excitation with 10mm amplitude.

As in the case of spiral waves, the concentric waves diffuse a given amount of light in a halo surrounding the central peak. The PSF corresponding to 10mm amplitude has a brighter halo (at $630\lambda/D$) and a second less intense halo (at $375\lambda/D$), while the 1mm amplitude PSF only present the bright halo at $630\lambda/D$. These distances are at least twice as far as in the spiral case, which is related to the wavelength of the waves disturbing the mercury. In the concentric case, they are of the order of 1cm, while in the spiral case is concerned, the smaller wavelengths are of the order of 6cm at the periphery. The frequencies of the waves are thus higher in the concentric case which explains why the halo in further in the corresponding PSFs (i.e. the Fourier transform of the pupil).

The 10mm amplitude PSF also presents, in its central part (see smaller pictures in the top left corner of the large ones), a double annular structure, probably related to the larger extension of the disturbed zones. Indeed, as the initial amplitude of the waves is larger, they travel over longer distances before disappearing. Fig. 3.20 (left) shows the evolution of the Strehl ratio as a function of the amplitude of excitation at the center of the mirror. In order to keep both amplitudes (center and edge) of the same order, we forced the second one to be 1.1 times the first one ($A_e = 1.1\,A_c$). The PSFs presented in fig. 3.19 correspond to the first and last cases (A_e=1mm and A_e=10mm) respectively.

As expected, the Strehl ratio decreases when the amplitude increases. Moreover the evolution

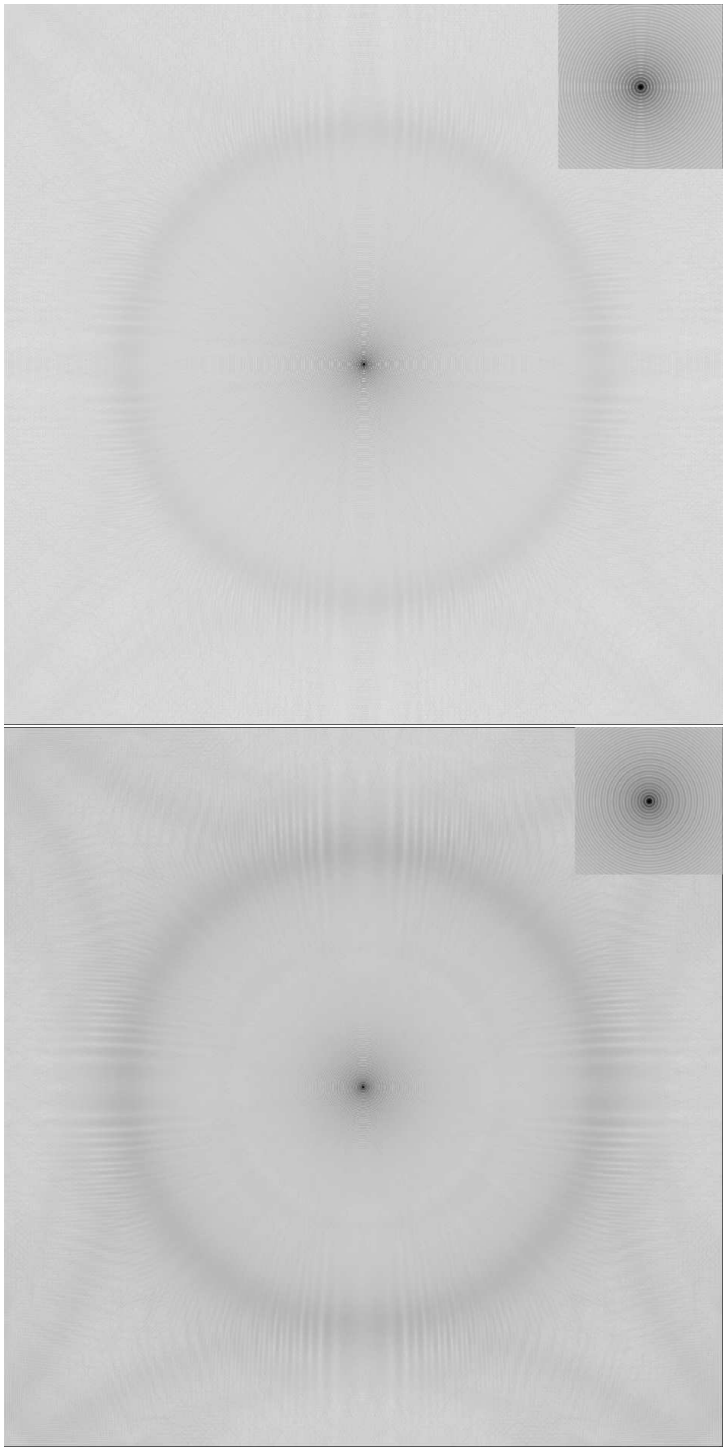


Figure 3.19: Point spread functions corresponding to the mirror modeled in fig. 3.18 (100s integration time). Top: PSF corresponding to an amplitude of excitation of 1mm. The Strehl ratio is 99.5%. Bottom: PSF corresponding to an amplitude of excitation of 10mm. In this case, the Strehl ratio is 80%. In both cases, a diffuse halo surrounds the central peak (at $630\lambda/D$). The halo contains more light in the 10mm case and a second smaller halo is also present at lower frequency (at $375\lambda/D$). In both cases, the top-right smaller picture presents a zoomed view of the central peak. The 10mm case shows a structure like a double ring that could be related to a larger extension of the disturbed area of the mirror.

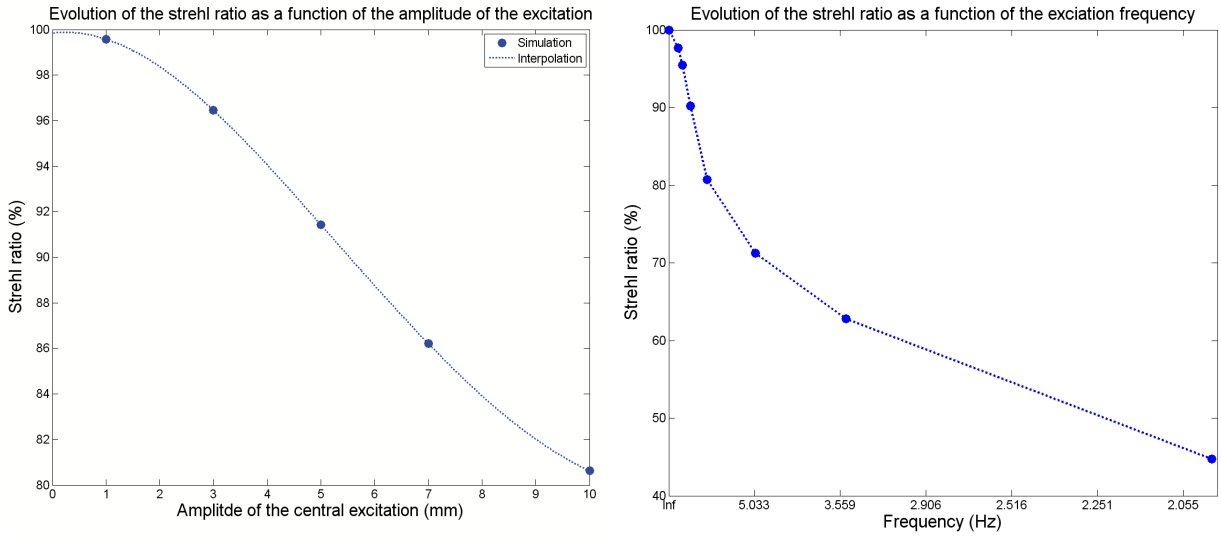


Figure 3.20: Evolution of the strehl ratio corresponding to a liquid mirror disturbed by concentric waves. Left: Strehl ratio as a function of the excitation amplitude. The center and edge excitation frequencies are respectively 10 and 12Hz. Right: Strehl ratio as a function of the excitation frequency. The scale of frequency is expressed in ω^{-2} unit, which corresponds to the amplitude coupling term variation with "ω". Using this scale allows to compare the curve related to the amplitude (left) with the one corresponding to the frequency (right). The behavior is similar in both cases. The amplitudes of the excitation are 5mm at the center and 6mm at the edge.

in the concentric case has the same "gaussian" shape as in the spiral case. The second graphic of fig. 3.20 presents the evolution of the Strehl ratio as a function of the excitation frequencies, again the edge frequency is computed from the central frequency ($\omega_e = 1.2\,\omega_c$). The scale of the x-axis is a function of ω^{-2} so that it follows the amplitude coupling function. Using this scale allows us to compare the amplitude and frequency graphs. It appears that the excitation frequency influences the amplitude in an important way. When the frequency is high, with respect to the natural frequency of the system, the amplitude coupling between the excitation source and the mercury is low and the Strehl ratio is very good. On the contrary, when the excitation frequency is near to the natural frequency, the amplitude coupling is important (resonance) and the amplitude of the waves gets important causing a quick decrease of the Strehl ratio.

3.7 Waves damping

As we saw in the previous section, the waves that propagate through the mercury are damped during their travel. It has been known for a long time that this damping depends on the mercury thickness (Borra et al. 1992; Girard and Borra 1997). Equations that establish the relation between the damping coefficient and the mercury thickness are presented hereafter.

Mulrooney (2000) proposed the derivation of a formula based on the Landau and Lifshitz (1959) computation for the damping of a fluid

$$\gamma = 2\nu k^2 \tag{3.63}$$

and on the dispersion relation presented above (equation 3.61). His relation between the damping, the thickness of the mercury layer, and the wave frequency is

$$\gamma = \frac{\nu\rho}{\alpha}\left[\sqrt{g^2 + \frac{1}{h\rho}(4\alpha\omega^2 - 16\alpha\Omega^2)}\right] \tag{3.64}$$

where ν is the kinematic viscosity of the mercury ($\nu_{Hg} = 114 \cdot 10^{-3}\mathrm{N/m}$), ρ is its density, α its surface tension, ω is the pulsation of the wave that propagates through the mercury and Ω is the angular velocity of the system. A mapping of the damping coefficient as a function of the wave frequency and of the mercury thickness is presented in fig. 3.21. The waves get more and more damped as their frequency increases and as the thickness of mercury decreases.

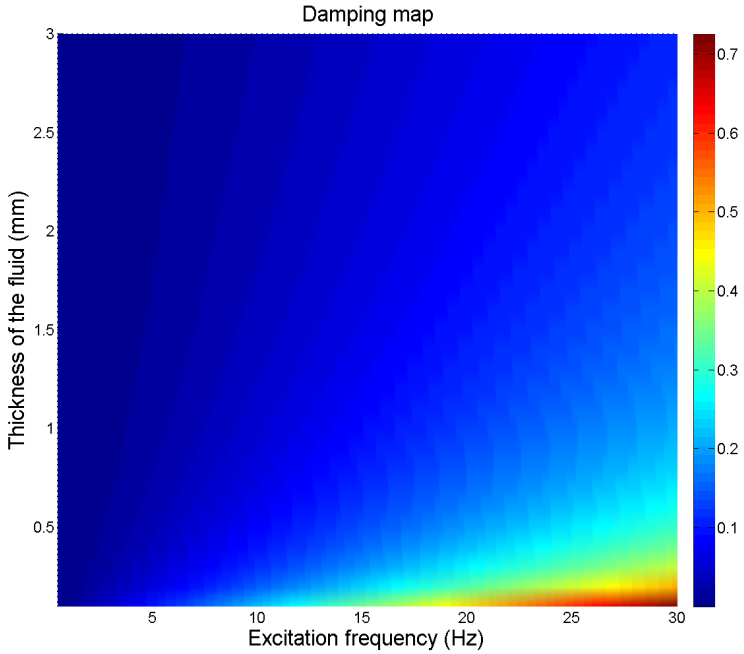


Figure 3.21: Damping as a function of the thickness of mercury and the frequency of the waves (equation 3.64). The damping coefficient if given in Hertz. The damping increases with the frequency and decreases with the thickness of the layer. The damping values presented here correspond to unoxidized mercury, the actual damping is thus expected to be more important.

Equation 3.64 does not take oxidization of the mercury into account. Tremblay and Borra (2000) propose an empirical law (3.65) for the dependence of the damping with respect to the thickness of the liquid layer.

$$d = 15.8\, h^{-1.38} \tag{3.65}$$

where h is the mercury thickness expressed in millimeter. As it has been established from measurements, this power law applies to an oxidized layer of mercury. However it does not take the dispersion into account. It is thus probably only valid for frequencies around 12Hz, the frequency at which the law has been determined. In our simulations of the concentric waves, we used this relation as the excitation frequency was of the same order of magnitude as those presented in Tremblay and Borra (2000). The ratio (%) of the remaining amplitude as a function of the traveled distance and of the mercury thickness is presented in fig. 3.22 (left).

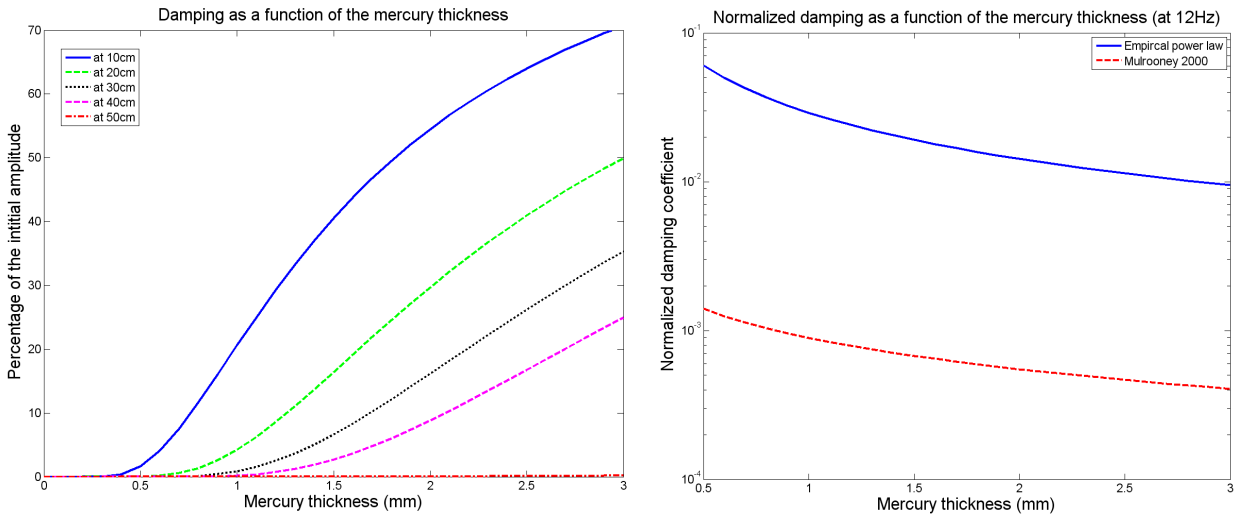


Figure 3.22: Left: Percentage of the initial excitation amplitude as a function of the mercury layer thickness for several propagation distances (10-50 cm). Right: Comparison of the normalized damping coefficients. The empirical results are normalized by the wave-number and the theoretical results are normalized by the wave pulsation. It is clearly visible that the two relations do not depend on the mercury thickness in the same way. In any case, the empirical results are much higher than the other ones. This difference can be partially explained by the extra damping due to the oxide layer.

It should be noted that the values for the damping coefficient given by equation 3.65 are expressed in m^{-1} and not in Hz as in equation 3.64. In order to be able to compare them, we have normalized both results respectively with the wave-number and the pulsation of the wave. The comparison is presented in fig. 3.22 (left). The values obtained with the empirical law are higher than those obtained with Mulrooney's equation. This difference is partially due to the extra damping caused by the mercury oxide layer. However, both expressions do not depend on the mercury layer thickness in the same way. Indeed, Mulrooney's equation is in $h^{-0.5}$ whereas Tremblay's empirical law is in $h^{-1.38}$. The latter being normalized by k, the dependence in h slightly changes because of the relation between k and h ($k = f(h^{-0.25})$). The normalized empirical relation is thus a function of $h^{-1.13}$.

Nevertheless, we can still conclude that the thickness of mercury should be reduced at maximum in order to improve the damping of the waves.

3.8 Testing the liquid mirror

We have shown in the previous sections that many causes may affect the surface of a liquid mirror. However, it has been computed that the deviation from a perfect parabola is mainly caused by wavelets propagating through the mercury layer. These waves are related to two different causes. The vibrations of the container are the first source of waves, they are mainly due to the instability of the rotation speed that are transmitted to the dish (at its resonance frequency). The waves created this way are concentric and have a wavelength around 1-2 cm. They can be reduced by improving the rotation stability. Waves are also generated by the relative wind between the air and the mercury that is created because of the rotation of the dish. They present a spiral shape and their wavelength is around 3-5 cm. Using a mylar cover on top of the bowl will isolate the rotating mercury from the ambient air so that the spiral waves will almost disappear. In both cases, the amplitude of the wave depends on the thickness of the mercury layer that is used to make the mirror: the latter should thus be as thin as possible. A thickness of about one millimeter is probably the best that can be practically obtained with a large mirror.

All these waves represent the main mirror disturbance and cause most of the image degradations. It is thus important to be able to detect and characterize them in order to identify their origin and to solve it.

Several classical test methods can reveal the presence of waves on the mirror, like Fourier's knife-edge test (Borra et al. 1985a) presented in fig. 3.23. However these tests require an easy access to the center of curvature of the mirror or to its focus, in the case that stars could be used as sources. Indeed, the material required for the tests must be located at one of these points. Even if it seems easy in the case of a small focal length (2-3 m), it becomes quite complicated for a focal length of eight meters and a center of curvature at sixteen meters above a mirror that has to stay horizontal.

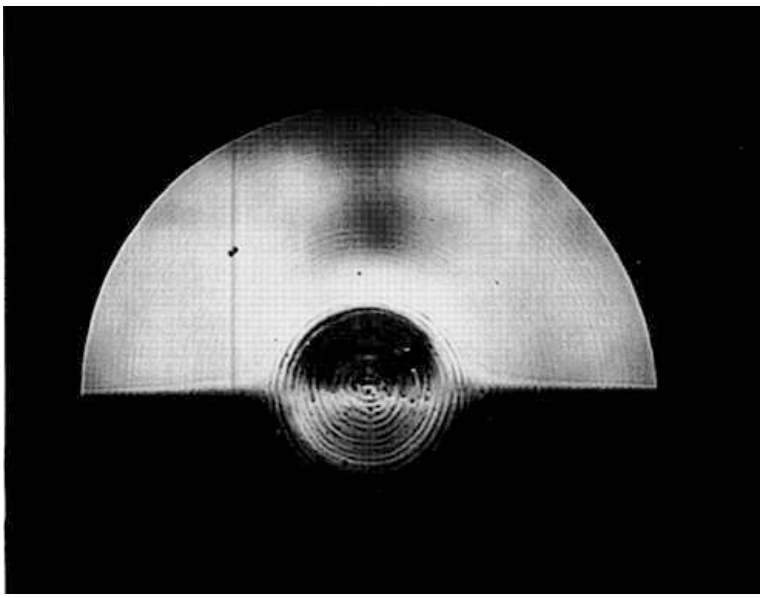


Figure 3.23: Fourier knife-edge test on the one meter liquid mirror at Laval university. Concentric waves are clearly visible in the center of the mirror. Image from Borra et al. (1985a).

Tests that could be implemented easily are preferable to discriminate between the different

types of waves. We are thus looking for tests that do not require the use of the center of curvature of the primary mirror or stars as sources. We review hereafter several tests that have been proposed. They are presented in order of increasing sensitivity.

The first qualitative test simply consists in looking at the outer edge of the liquid mirror in grazing incidence with naked eyes, or a camera with a zoom capacity. The waves, that present a higher amplitude at the edge of the mirror, could already be visible with this technique, and it may not be useful to undertake more sophisticated tests. However, even if this method can reveal the presence of waves, it can be quite difficult to distinguish between the two types.

In order to get more sensitivity, another qualitative test can be conducted. It consists in studying the motions of a laser spot that is reflected on the disturbed surface. It can be shown that the motion of the spot depends on the motion of the reflecting surface and it is thus possible to characterize the waves that propagate in the mercury. This method can be improved using a more complex pattern than a single spot. For example, using a "horizontal" line (perpendicular to the radius of the spinning mirror) introduces length information in the test and using a grating of several lines also introduces time information.

This test is extensively described in another document (complete report by François Finet). Its expected accuracy is of the order of one wavelength. The same type of technique has been used to measure the slope of the waves on the 3.7-m mirror of Laval University (Tremblay and Borra 2000).

Even if the qualitative approach introduced above can give interesting results, a more quantitative test is mandatory to complete the measurements of the disturbances on the mirror. In order to develop this test, we investigate the concepts and theory of optics (chapter 4), particularly the phase retrieval methods (chapter 5). After these reminders, we study a particular phase retrieval technique based on the Nijboer-Zernike theory (second part of chapter 5 and chapter 6): a third more accurate testing method could be based on this approach.

This method can be related to Roddier's method for optical testing. Indeed, the defocus range used by Roddier is sufficiently important so that the defocused images are in pupil planes. In the Nijboer-Zernike (NZ) approach, the defocus is small enough so that all images are still in the image planes. The NZ testing method and its implementation will be described later (chapters 7 and 8), after the corresponding theory will have been explained, but we can already say that using a laser and two lenses it would be possible to illuminate small areas of the mirror (10 cm), and to image them on a CCD camera. The pictures, taken in a given range of focus, will then be used to compute the aberrations of these parts of the mirror.

Chapter 4

Point Spread Function - (PSF)

4.1 Introduction

The system theory teaches us that any system is fully described by its impulse response. The latter allows to compute the output of the system for any given input. Indeed, the impulse response corresponds to the signal given by the system when a pulse has been given as input.

When an optical system is considered, the impulse response is called the "Point Spread Function" (PSF). It corresponds to the image (the output) of a point source (the input pulse) generated by the optical system and establishes a direct link between the input object and the output image. As an impulse response, the PSF completely characterizes the optical system, it thus constitutes an important piece of information.

This chapter presents several aspects of the PSF. First of all we will begin with some theoretical concepts such as diffraction and convolution. The first one is related to the propagation of light, whereas the second one is the mathematical link between the impulse response of a system and the output that corresponds to an arbitrary input.

PSF measurement techniques will be briefly introduced. We will then talk about atmospheric disturbance effects and about some methods that can be used to compensate for them. Finally two ways of mathematically computing the point spread function of a system will be presented. These are based on Fourier transform and on the Nijboer-Zernike theory.

4.1.1 Diffraction

Diffraction is an undulatory phenomenon happening when a wavefront reaches an obstacle (an optical system). It results in the deviation of the rays passing near this object. Obviously, every optical instrument is affected by the effects of diffraction.

The first theoretical explanation of this phenomenon was given by Huygens in 1678. He said that "every point of the space reached by a wavefront acts as a source of secondary waves" (fig. 4.1). The basic equation of the mathematical theory of diffraction is the Fresnel-Kirchhoff integral

$$U_p(r') = \frac{-ikU_0}{4\pi} \int_A \frac{e^{ik(r+r')}}{rr'} (cos(n,r) - cos(n,r'))dS \tag{4.1}$$

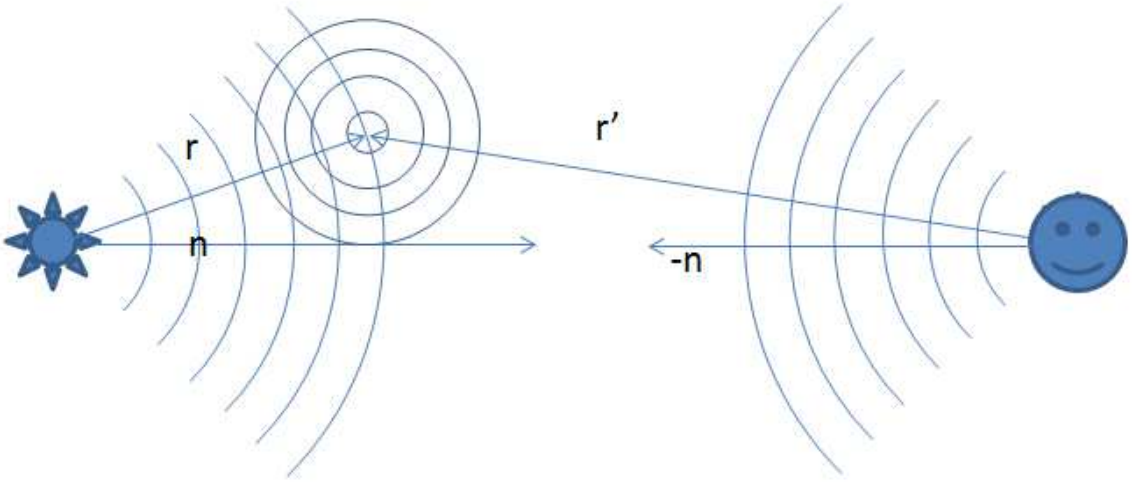


Figure 4.1: Illustration of the diffraction principle. The source at left emits a spherical wavefront. Each point of this wavefront act as a source. The signal that reaches the observer at right is composed of the sum of all the wavefronts coming from the source.

The wave received at the observer U_p is the sum of all the spherical waves, coming from each point of the source, reaching the observer. $\mathbf{n}$ is the vector that links the source and the observer and $\mathbf{r}, \mathbf{r}'$ are the position vectors of each "sub-source" with respect to the source and the observer (fig. 4.1). Even if this equation can be solved numerically, it is more interesting to consider the two following approximations, that can be treated analytically.

The first one is the Fresnel approximation, it is obtained by expanding the cosine terms of the Fresnel-Kirchhoff integral and limiting the series to the quadratic terms. The Fresnel approach thus consists in considering the wavefront as being parabolic instead of spherical.

$$U_p = \frac{-ikU_0}{2\pi zz'} e^{ik|PS|} \int_A e^{\frac{ik}{2z}\left[(x_0-x_m)^2+(y_0-y_m)^2\right]} dS \tag{4.2}$$

where x_0, y_0, z are the coordinates of the observer and $x_m, y_m z'$ are the coordinates of the points of the source. This approximation is valid for an observer located near to the diffracting aperture, that is in the Fresnel region also called the near-field[18].

The second, more stringent, case is given by the Fraunhofer approximation.

$$U_p = \frac{-ik}{2\pi z} e^{ik|PS|} e^{\left(\frac{iz_a}{2k}[u^2+v^2]\right)} \int_A U_0(x_0, y_0) e^{(-i[ux_0+vy_0])} dx_0 dy_0 \tag{4.3}$$

where U_0 is the amplitude distribution at the pupil, x_0, y_0 are the coordinates in the pupil plane and u, v are the coordinates in the image plane (observer). The Fraunhofer approximation consists in working with plane waves. In this case the cosine Taylor series are truncated at the first order. This corresponds to a source located at an infinite distance from the observer, as it is the case in observational astronomy, for example. In practical cases, this approximation applies when the observer is sufficiently far from the source. It can be seen from equation 4.3 that, under

[18]The near-field is the immediate neighborhood of the diffracting object (from the object to a few tens of centimeters) where the curvature of the wavefront has to be accounted for.

these conditions, the image corresponds to the Fourier transform of the wavefront coming out of the diffracting aperture (the pupil).

Because of the diffraction phenomenon, the image of a point source through an optical instrument is not a perfect point. Indeed, knowing that the image is given by the Fourier transform of the wavefront at the pupil (in the Fraunhofer approximation), and assuming that the object is a point source (a star) and that the optical instrument has a circular aperture, the image is thus the Fourier transform of a uniformly illuminated circular pupil. The result is the well-known Airy diffraction pattern (fig. 4.2). Looking at this picture, one easily understand why it is called a "point spread function" as it shows how a point object is spread when seen through an optical system.

The angular resolution of an optical instrument is defined by the angular radius of the Airy disc, i.e. the central spot of the pattern presented in fig. 4.2. In a perfect case, the resolution only depends on the telescope aperture (the primary mirror diameter), it is given by

$$\theta_t = 1.22\frac{\lambda}{D}, \tag{4.4}$$

where θ_t is the theoretical angular size resolved by a perfect telescope, λ is the wavelength of the observed light and D is the diameter of the primary mirror.

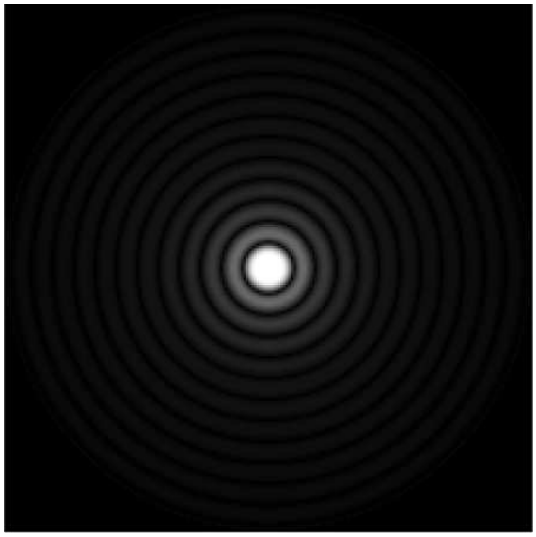


Figure 4.2: Airy pattern for a circular aperture, simulated with the Nijboer-Zernike equation (section 4.3.2). The intensity is expressed to the power 1/2.

4.1.2 Convolution

As previously stated, knowledge of the point spread function allows to compute the image produced by the optical system for any kind of source. This operation requires a mathematical tool called the convolution. The source image (that is different from a point or not) is obtained simply by convolving the object with the PSF.

The convolution is defined by

$$PSF \otimes Source(t) = \int_{-\infty}^{\infty} PSF(\tau)Source(t-\tau)d\tau. \tag{4.5}$$

Among the properties of this product, the following one is very interesting. The Fourier transform of the convolution product of two terms is equivalent to the classical product of the Fourier transform of these terms.

$$FT(PSF \otimes Source) = FT(PSF) \cdot FT(Source) \tag{4.6}$$

This property drastically simplifies the calculation of the convolution product.

Let us note that the PSF being the Fourier transform of the pupil, the term $FT(PSF)$ is the autocorrelation of the pupil. The Fourier transform of an impulse response is called the transfer function of the system.

Fig. 4.3 shows a simulated star field (left). The convolution of this field by the PSF of the instrument introduces a blurring of the images as all sources are spread out by the instrument (right). The diffraction phenomenon thus introduces a loss of information.

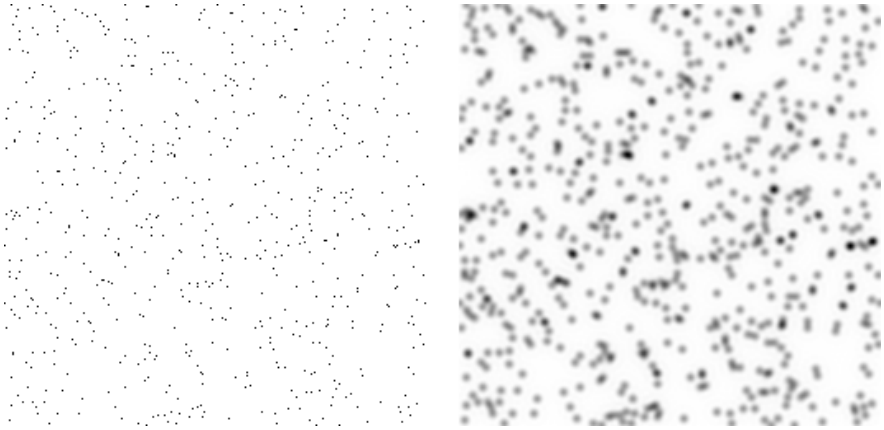


Figure 4.3: Left: Simulated field of stars as it is actually. Right: The image of the same field of stars after convolution by the PSF of the instrument. The only effect is thus due to the diffraction related to the diameter of the entrance pupil. The instrument introduces blurring into the images it produces. The diffraction thus causes a loss of information.

4.1.3 Atmospheric turbulence

We have presented in the previous section that any instrument degrades the images (they are spread out). However, the symmetry of the image is kept, and this degradation only consists in blurring. Another source of perturbation, that disturbs ground-based observations, is the atmospheric turbulence.

The atmospheric turbulence has been extensively studied and described, in particular in Roddier (1981). It is well known that ground-based observations are disturbed by the atmospheric turbulence. This one scatters the light which results in blurring (loss of resolution) and in lowering the mean surface brightness. Going through the atmospheric layers, the wavefront encounters turbulent regions where the air temperature, density and index of refraction vary. This results in changes of the propagation speed of the light. Different parts of the wavefront travel through different regions of the atmosphere. As their propagation speed is different, the wavefront get distorted, it is said to be "aberrated".

During its propagation through the turbulent atmosphere, a wave accumulates phase errors of the order of a few microns. The total error on the wavefront that reaches the ground is given by

$$\varphi = \frac{2\pi}{\lambda} \int n ds, \tag{4.7}$$

where n represents the variation of the index of refraction. This error is different at every point on the wavefront. The mean square difference between two points varies as the 5/3 power of the distance between these points.

$$\langle[\varphi(0) - \varphi(r)]^2\rangle \simeq 6.88 \left(\frac{r}{r_0}\right)^{5/3}. \tag{4.8}$$

This is called the phase structure function, where r is the distance between two points of the wavefront and r_0 is the characteristic length of the turbulence cell, called the Fried parameter. It corresponds to the typical size of the region where turbulence can be considered as homogeneous. The phase error is smaller when r_0 is large, since the region of homogeneous turbulence is larger. The Fried parameter depends on the wavelength as

$$r_0 \propto \lambda^{6/5}. \tag{4.9}$$

On the ground, r_0 is typically of the order of 10 cm in the visible and 80 cm in the K band.

Kolmogorov was the first to propose an expression for the Power Spectral Density (PSD) of the atmospheric turbulence.

$$\Phi_K(k) = 0.033\, C_N^2 k^{-11/3}. \tag{4.10}$$

This formula is valid for medium size turbulence cells of the order of r_0, but Von Karman improved it to take the small and large scale structures into account. Defining l_0 as the inner scale, i.e. the smaller turbulence structure size (0.1 - 1 cm) and L_0 as the outer scale, i.e. the largest turbulence structure (25 - 100m), the PSD expression becomes

$$\Phi_{VK}(k) = 0.033\, C_N^2\, (k^2 + k_0^2)^{-11/6} e^{-k^2/k_m^2}, \tag{4.11}$$

where C_N^2 is the refractive index structure function, $k_0 = 2\pi/L_0$ characterizes the PSD for large scale structure and $k_m = 5,92/l_0$ defines its behavior for the small scale structure.

The random phase distortion of the wavefront results in a blurring of the images. Two effects are responsible for this blurring, the distortion and the tilt of the wavefront. The first one induces imperfect focusing while the second one creates random motion of the image. This image blurring is referred to as the atmospheric seeing. A good astronomical site has a typical seeing better than one arcsecond.

For small telescopes ($D \lesssim r_0$) the dominant effect is the image motion and the telescope is almost diffraction limited (equation 4.4), as the phase error on scales smaller than r_0 is small. In the case of larger telescopes ($D >> r_0$) the turbulence cells are smaller than the telescope aperture and the image quality is limited by the seeing

$$\theta \approx \lambda/r_0 \tag{4.12}$$

Another (small) effect of the atmosphere is the scintillation. It is due to intensity changes during the propagation of the wavefront coming from the star through the turbulent atmosphere. Indeed, the energy flow is perpendicular to the wavefront. If the latter is distorted, there are regions where the energy converges and others where the energy diverges, this causes variations

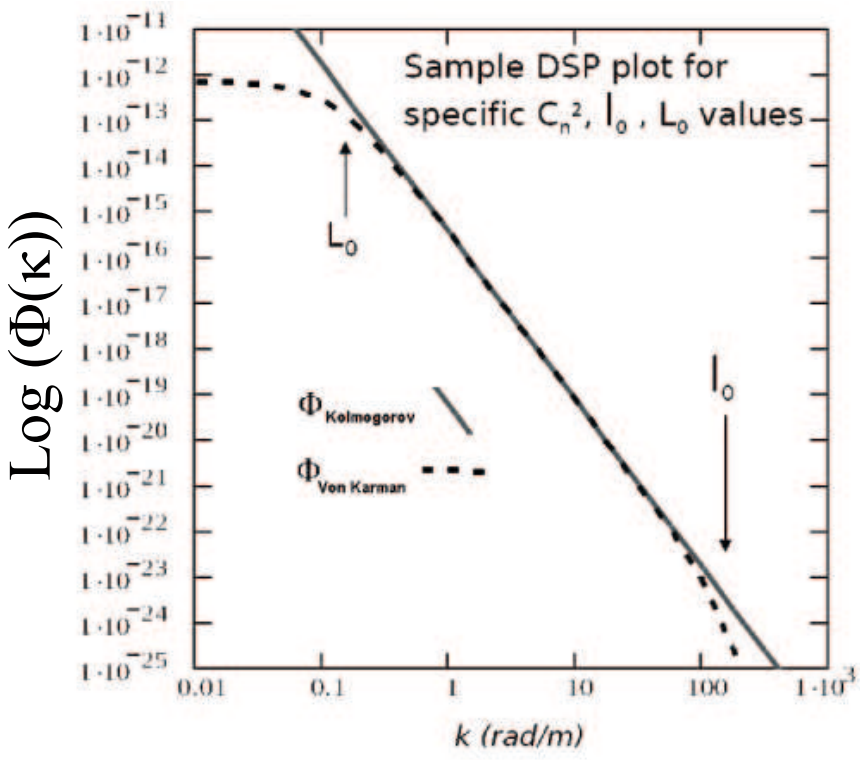


Figure 4.4: Power Spectral Density. The solid line represents the Kolmogorov spectrum. The dashed line corresponds to the Von Karman spectrum. (P. Riaud 2008, optics lesson: adaptive optics introduction.)

of the intensity. As the turbulence zones are moving with the wind, this phenomenon varies with a characteristic time scale given by

$$\tau_0 = 0,314 \frac{r_0}{V_{wind}} \tag{4.13}$$

where V_{wind} is the speed of the wind (m/s) in the turbulent layers of the atmosphere ($\sim 10m/s$). In the visible spectrum, τ_0 is of the order of 1-5 $\cdot 10^{-3}$s, while it is around 0.1s in the K band.

4.1.4 PSF fitting

In section 4.1.2, we showed a picture of a simulated field of stars as seen by an optical device. This field was not very crowded and there was almost no overlap between the different objects.

However, in real images, this is not often the case. Stars can be very close to each other and because of the spreading due to the instrument, their images may merge. It may also happen that a point-like object would be superimposed over an extended one.

These cases may turn out to be very difficult to treat as far as photometry and astrometry are concerned. It is indeed a big challenge to determine whether the measured flux belongs to the star or to the background and if the measurement of the centroid position is affected by the presence of a nearby star.

Knowing the PSF of the instrument can help to solve these problems. Using the technique called "PSF fitting", it becomes possible to model the objects and see how the adjusted models

combine to correspond to the observed image. This allows to determine the flux of all the objects and their positions by adjusting modeled PSFs.

Two ways exist to create the model; the analytical method and the empirical one. We introduce hereafter the most common analytical functions used to model PSF; Gaussian, Moffat, Lorentzian. These models are normalized, and the central intensity of the fitted objects is used as an additional parameter.

The gaussian PSF model is defined by the equation

$$PSF_{\text{Gaussian}} = \frac{1}{2\pi\sigma^2} e^{\left[-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right]} \tag{4.14}$$

where σ is the dispersion parameter and r is the radial position. For a two dimensional image, the gaussian profile is characterized by four parameters, two of them (x,y) describe the position of the central peak of the image and the two others (σ_x, σ_y) define the extensions of the gaussian along the x and y directions. This type of model gives a good fitting for the central part of the PSF. Nevertheless, a real PSF also possesses wings that are much more pronounced than in a simple gaussian (Trujillo et al. 2001a).

A lorentzian profile is indeed much better to model the wings of the PSF. It is defined as

$$PSF_{\text{Lorentzian}} = \frac{1}{\pi} \frac{\frac{1}{2}\Gamma}{(r)^2 + \left(\frac{1}{2}\Gamma\right)^2} \tag{4.15}$$

where Γ is the parameter specifying the width of the profile. In this case the wings are more important, but the peak is less well fitted.

The Moffat distribution combines the advantages of both the gaussian and the lorentzian profiles as the wings are more customizable. It is defined by

$$PSF_{\text{Moffat}} = \frac{\beta - 1}{\pi\alpha^2} \left[1 + \left(\frac{r}{\alpha}\right)^2\right]^{-\beta} \tag{4.16}$$

where α and β are two free parameters depending on the seeing and the magnitude of the star. This last profile is often used to model PSFs. Both the wings and the central peak are well fitted. Moffat functions are thus very good to model narrow PSFs. Moreover, it is interesting to note that this profile includes the Gaussian as a limit case, when $\beta \to \infty$ (Trujillo et al. 2001b). It also presents the advantage of being very stable numerically speaking. Common astrophysical softwares such as IRAF or SEXTRACTOR implement this type of profile.

All these profiles can be used for unidirectional or bidirectional PSF fitting. This type of fitting is generally implemented with non-linear methods, such as the Levenberg-Marquardt algorithm, that allows to compute the minimum of a non-linear function of several variables. It is based on interpolation of the Gauss-Newton algorithm and uses the gradient method. It is more stable than the classical Gauss-Newton algorithm even if it is slightly slower in particular cases. An example of implementation of this algorithm can be found in Markwardt (2008) (downloadable on http://www.physics.wisc.edu/∼craigm/idl/fitting.html).

In practical cases, it may be difficult to use a purely analytical model. Indeed, in addition to the spread out of the image due to diffraction, other deformations of the images are due to

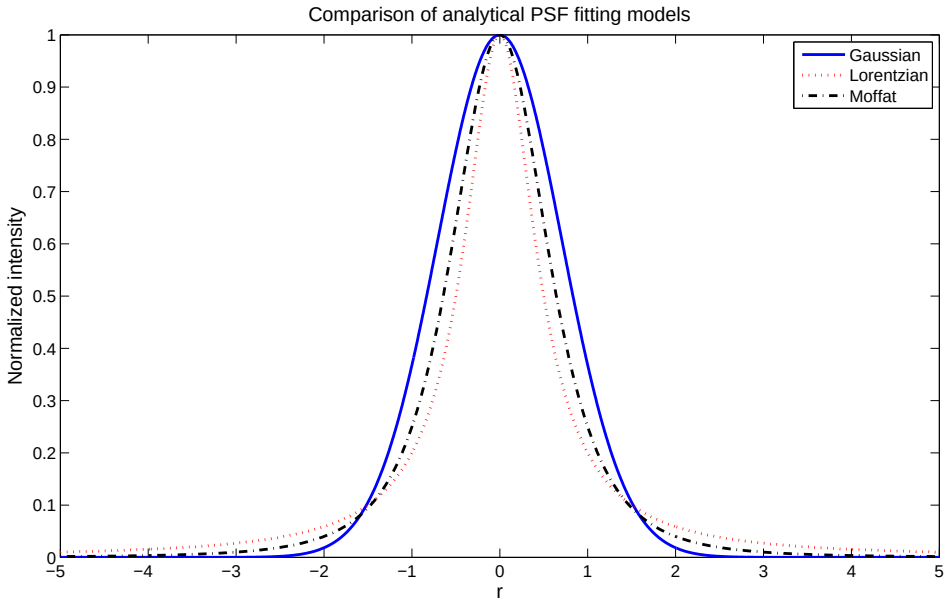


Figure 4.5: Comparison between a Gaussian, a Lorentzian and a Moffat profile. The Moffat is often the most interesting solution for PSF fitting as it is between the two other ones and it is more parameterizable.

imperfections of the instrument or to the atmospheric turbulence. In this case, the PSFs are difficult to estimate and an empirical model should be constructed.

4.1.5 Speckle interferometry

Speckles are little luminous spots created in the focal plane of a telescope by the interference of the rays coming from different parts of a wavefront distorted by the atmosphere. During a long exposure, these speckles are mixed together because of the varying tilt of the wavefront. This mixing results in a large image spot. However, the speckles can be seen in a short exposure image (20ms in V). Their characteristic angular size corresponds to the limit of diffraction of the telescope $\theta \approx \lambda/D$.

In 1970, Labeyrie suggested that the speckle-affected images contain more information than the long exposure blurred images (Labeyrie 1970). The image itself is difficult to exploit. However, it is possible to study the interference fringes of the diffraction pattern of the speckle images, obtained by shooting a laser beam through them. Nowadays, this pattern is calculated from the Fourier transform (FT) of the image (see fig. 4.6).

The characteristics of the fringes contain a lot of information. In the case where the object is a double star, for example, the visibility depends on the difference of magnitude between the two stars. The inter-fringe spacing is related to the angular distance between the stars and the orientation of the fringes gives their relative positions (fig. 4.6).

The signal to noise ratio of a speckle image is relatively low (see fig. 4.6). However it is possible to sum up a large number of images, in order to improve the signal, the analysis is thus

performed on the diffraction pattern of the sum of all the images.

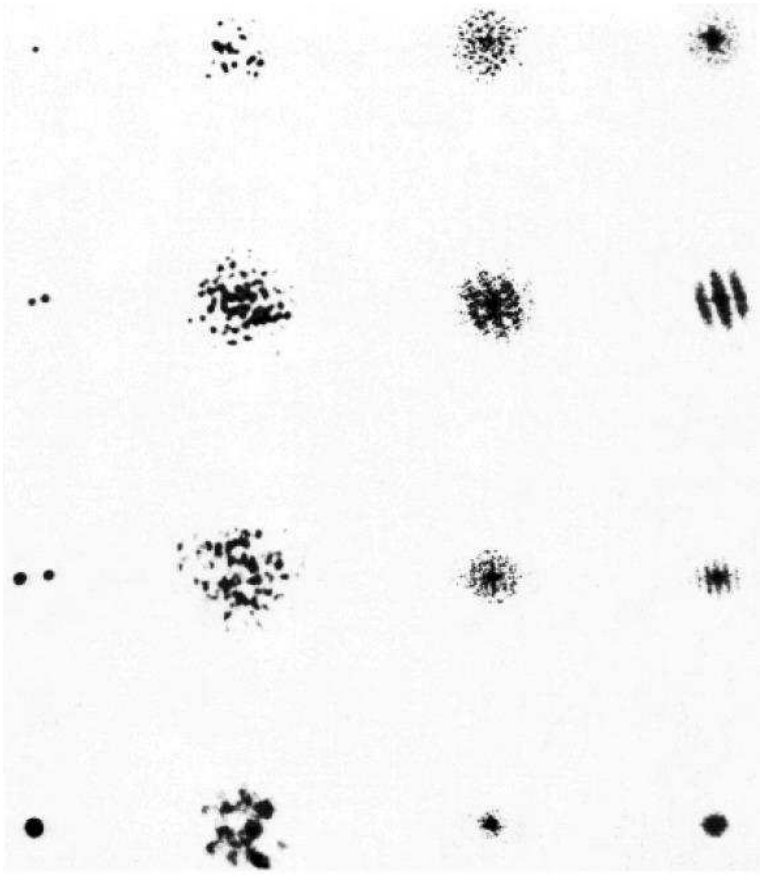


Figure 4.6: The first column presents different types of object. The second one shows one speckle image of the corresponding object. The third column is the Fourier transform of the speckle image presented in the previous one. The last column shows the Fourier transform of the sum of 20 speckle images of the object. On the first line, the object is an unresolved star. The second and third lines present double stars. Interference fringes appear in the last column. The orientation of the fringes is related to the relative position of the stars, the visibility of the fringes depends on the difference of magnitude between these stars and the inter-fringe-space is related to the angular distance between the latter. The last line shows the case of a resolved star. (P. Riaud Private communication)

Let us consider a general case, defining $O(\alpha, \beta)$ and $I(\alpha, \beta)$, the intensity distributions respectively in the object and the image, the observed image is given by the convolution of the object function with the speckle image of a point object. Indeed, the latter is the Fourier transform $p(\alpha, \beta)$ of the perturbed telescope pupil $P(x, y)$ (i.e. the perturbed PSF of the telescope),

$$I(\alpha, \beta) = O(\alpha, \beta) \otimes |p(\alpha, \beta)|^2. \tag{4.17}$$

The Fourier transform of the image is given by

$$i(x, y) = o(x, y) \cdot \mathcal{A}[P(x, y)], \tag{4.18}$$

and its intensity,

$$|i(x,y)|^2 = |o(x,y)|^2 \cdot |\mathcal{A}[P(x,y)]|^2, \tag{4.19}$$

where $\mathcal{A}[P(x,y)]$ is the autocorrelation function of $P(x,y)$. Its square modulus is the modulation transfer function (MTF) of the perturbed instrument. Only the time-average of this function can be determined. It is thus useful to add many instantaneous images before calculating the Fourier transform,

$$\sum |i(x,y)|^2 = |o(x,y)|^2 \cdot \sum |\mathcal{A}[P(x,y)]|^2. \tag{4.20}$$

The Fourier transform of the object is given by

$$|o(x,y)|^2 = \frac{\sum |i(x,y)|^2}{\sum |\mathcal{A}[P(x,y)]|^2}, \tag{4.21}$$

the time-averaged intensity of the pupils's autocorrelation function being known, this gives the intensity of the object's Fourier transform. This is the Van Cittert-Zernike theorem.

This method was used on large resolved stars to measure their angular diameters and mainly on double stars to measure their angular distance. To be usable, speckle interferometry requires that the target stars are bright enough ($m \sim 9$ for an 8-meter telescope). Moreover, in case the target is a double star, the difference of magnitudes must be small enough so that the speckles of both stars can be seen ($\delta m \sim 5$). Speckle interferometry was used on various targets such as Eta Carinae (Weigelt and Ebersberger (1986) and Hofmann and Weigelt (1988)), NGC 1068 (Wittkowski et al. 1998), IRC+10216 (Weigelt et al. 2002).

4.1.6 Adaptive optics

It has been presented in section 4.1.3 that the atmospheric turbulence degrades the image quality of any ground based telescope. Section 4.1.5 presented a solution to retrieve part of the information lost because of the atmosphere, but as long as the phase information was lost, it is impossible to reconstruct an image with this method.

Adaptive optics is a technology developed to improve the quality of images obtained with ground-based telescopes by removing, at least partially, the atmospheric effects. It consists in measuring and correcting the distortion of the incoming wavefront in real time, by using mobile and deformable mirrors.

The first stage of corrections corresponds to the position of the images. It consists in compensating the global tip-tilt errors of the wavefront. This is performed by a mobile tip-tilt mirror that can be rotated along two axes. It allows to correctly position the images on the optical axis of the system. Moreover, this mirror is generally mobile along the optical axis so that it is possible to correct for the defocusing. These corrections already remove a large amount of distortions corresponding to the blurring of the image. The tip-tilt corresponds to large wavefront errors, it thus has to be corrected separately in order to avoid saturating the deformable mirror. Indeed, trying to compensate for large tip-tilt errors would require the whole dynamic of this mirror which would not be able to correct other errors.

The next step consists in correcting the phase errors. A deformable mirror is shaped in such a way that it compensates for the wavefront deformations. The reflected wave is thus

almost perfectly planar, as it would have been in the absence of turbulence effects. In order to compensate for the turbulence in real time, the correction process has to be at least twice as fast as the characteristic time of the turbulence (for sampling requirement). As stated in the section concerning the atmospheric turbulence 4.1.3, the characteristic time of the turbulence depends on the wavelength, the required frequency of correction thus also depends on that parameter ($\sim$ 2000Hz in visible light; $\sim$ 200Hz in H band; $\sim$ 100Hz in K band). This also explains why it is easier to use adaptive optics in the infra-red than in visible light. The limiting frequency comes from the sensor frequency more than from the deformable mirror. Indeed, it is necessary that the analyzed wavefront corresponds to the incoming one, so that the calculated corrections can be applied on the right wavefront. On the one hand, if the sensor were too slow, the corrections sent to the deformable mirror would correspond to an older wavefront and would not fit the new one at all. On the other hand, if the mirror frequency was too low, the correction would not be optimal, but it still would correspond to the right wavefront.

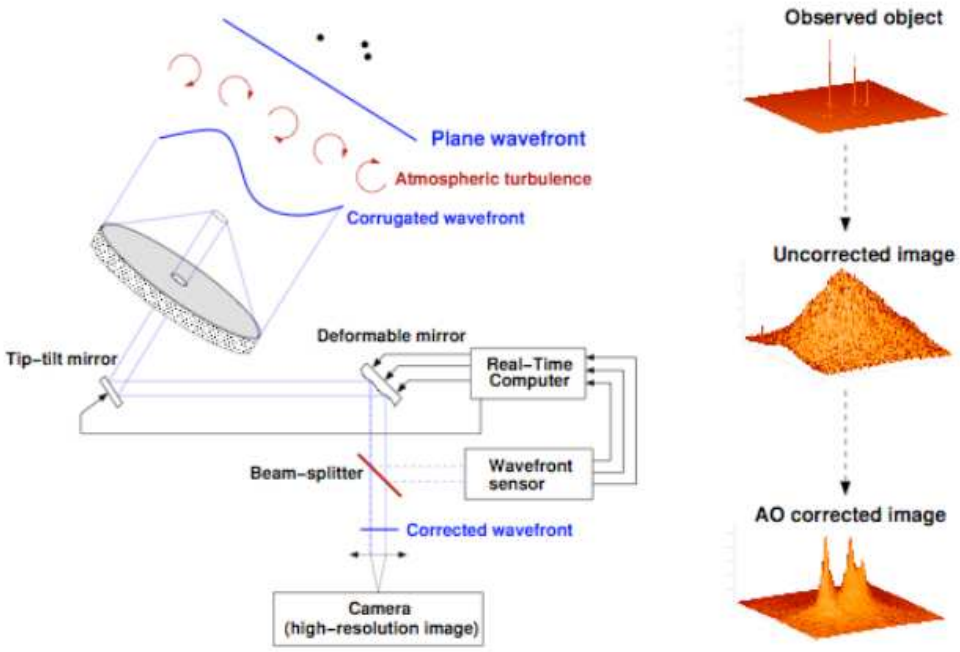


Figure 4.7: Principle of adaptive optics. The light coming from the star is first reflected on the primary mirror. It then reaches a tip-tilt mirror which corrects the tip-tilt and focus errors. After that, the light enters the adaptive optics system, where it is reflected on the deformable mirror. A fraction of the light is then sent to a wavefront sensor which measures the residual wavefront phase aberrations. A powerful computer calculates the deformations of the mirror needed to correct the wavefront and sends them to the tip-tilt and deformable mirrors, which mirrors then compensate for the deformations. Finally, the output of the system is an almost planar wavefront which is imaged in the focal plane. The shape of the mirrors is adapted in real time. (Source: NACO user's manual)

The adaptive optics, as illustrated in fig. 4.7, is based on wavefront measurements so that it is possible to compute the optimal shape that a deformable mirror should have in order to correct the incoming wavefront errors. To achieve this measurement, the light coming from the star, after having been reflected on the primary mirror, is sent to a tip-tilt mirror and then to a deformable mirror. After this last mirror, a fraction of the light is sent to a wavefront sensor. Once the shape of the wavefront is known, a control system calculates the position of the tip-tilt mirror as well as the optimal shape of the deformable mirror in order to compensate for the

phase errors. The tip-tilt mirror corrects the position of the images and the deformable mirror corrects the phase of the incoming wavefront so that the images appear sharp. These operations are performed as fast as possible in order to compensate for the turbulence in real time.

This technique shows impressive results as the resolution can be improved from 1 arcsec without AO to a few milli-arcsec with AO (see fig. 4.8). The adaptive optics image is sharpened and the intensity is concentrated in the central peak.

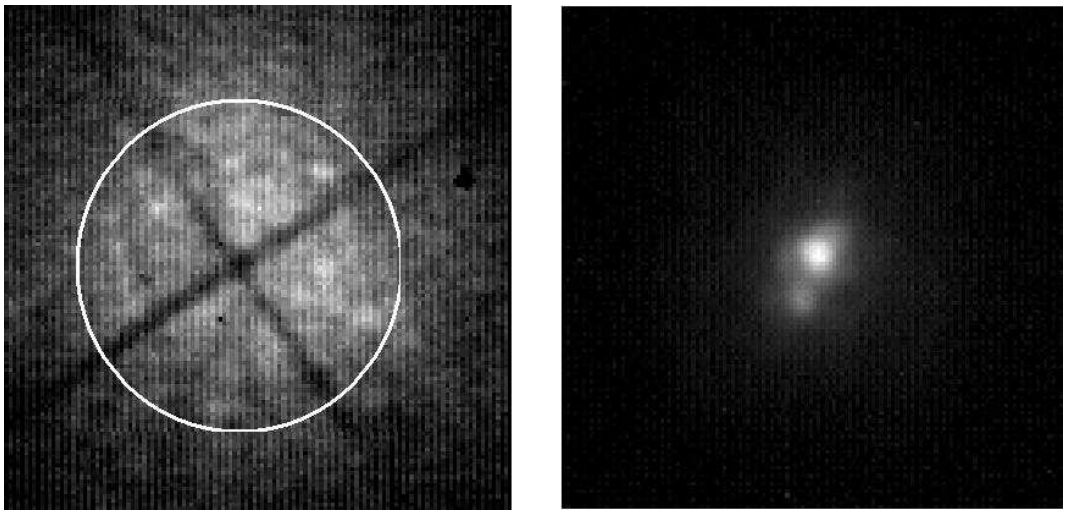


Figure 4.8: Left: Uncorrected NACO-VLT image in K band limited by the turbulence. Each speckle is diffraction limited. The circle represent the size of the seeing (about 0.8 arcsec) and the dark cross comes from the spider supporting the secondary mirror. Right: Corrected NACO image, it is almost diffraction limited, with a Strehl ratio of about 50%. (Image from P.Riaud 2008, Optics lesson: adaptive optics introduction.)

Wavefont sensor

In order to perform adaptive optics, the incoming wavefront has to be measured by a wavefront sensor. Several types of sensors exist: the Shack-Hartmann sensor, the curvature sensor, the pyramid sensor... The first two are being mainly used, the third one will not be presented in this document.

The Shack-Hartmann sensor consists of an array of small lenses which act as sub-apertures (fig. 4.9). Each of these lenses produces an image of the reference source, which position with respect to the optical axis of the corresponding lens is related to the local slope of the wavefront. Measuring these positions allows to determine the shape of the wavefront. This type of sensor is used at the VLT (Very Large Telescope) in the instrument NAOS/CONICA for example. A Shack-Hartmann sensor has to be used with a deformable mirror shaped by individual piezoelectric actuators. More details on the phase retrieval using the Shack-Hartmann method will be found in section 5.1.5

The advantages of this type of wavefront are its achromatism, its low sensitivity to scintillation, its large isoplanatic angle (30" in K band, see definition hereafter) and its good sensitivity. The CCD matrix must have at least four pixels per lens in order to respect the sampling theorem. Currently, the Shack-Hartmann is the most used sensor.

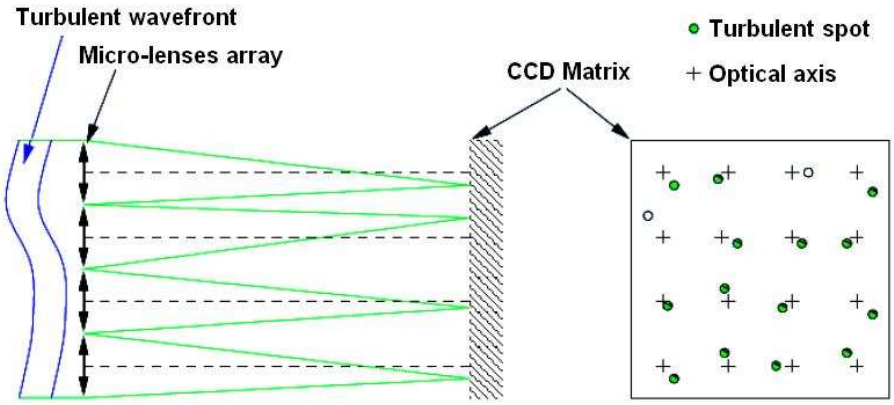


Figure 4.9: The Shack-Hartmann wavefront sensor is composed of an array of micro-lenses which sample the wavefront in sub-pupils. Each lens images a small part of the wavefront on a CCD. The position of the image relative to the optical axis of the lens is related to the slope of the corresponding part of the wavefront. Measuring each image position allows to determine the wavefront shape. (Image from P. Riaud 2008, Optics lesson: adaptive optics introduction.)

A method for reconstructing the wavefront using the Shack-Hartmann sensor may be found in Lane and Tallon (1992).

The curvature sensor (Roddier (1988)) is based on measurements of the intensity of defocused images (figs. 5.14 and 5.15 in section 5.1.6). Combination of the intra-focal and extra-focal intensities is related to the curvature of the incoming wavefront, (see the section 5.1.6, relative to the Roddier phase retrieval method for details).

The defocusing is performed thanks to a membrane mirror which oscillates between two positions, creating images alternatively located before and after the focal plane. The correction system used with the Roddier sensor is a bimorph mirror, which is made of two bonded piezoelectric plates. At the beginning Roddier thought that his method would need no intermediary calculation as the result of the sensing could be directly applied to the deformable mirror. Indeed, the curvature sensor measures the second derivative of the incoming wavefront, which is exactly what is required by the bimorph mirror. However, because of error propagation, it appears to be more difficult. Nevertheless, this method is still usable and is implemented, for example, in the SINFONI instrument installed on one of the VLT telescope (UT4).

The main physical limitation of adaptive optics is due to the dilution of light involved in the sensor. Even if this "loss" of light is necessary, the wavefront sensor has to sample the turbulence, it implies that the adaptive optics system can only be used with bright sources.

Let us notice that the sub-aperture size should correspond to the Fried parameter of the turbulence in the wavelength band of interest. Moreover the sampling requires that at least four pixels (2x2) of the CCD are dedicated to each sub-aperture.

Reference star

As previously written, the measurement of the shape of the wavefront has to be done on a bright star because of the large amount of photons required by the sensor. However, science targets are

often too faint to be used as a reference star, and another bright star, located in the neighborhood of the target of interest, has to be used as a reference. Indeed, the light from both the target and the reference follow approximately the same path through atmospheric turbulence. Hence, the image of the target is almost as well corrected as the reference image.

However, this presents two limitations; the corrected field of view is quite small and the adaptive optics cannot be used everywhere on the sky.

Concerning the first point, the size of the corrected field of view is characterized by the isoplanatic angle. It corresponds to the angle where the rms residual wavefront error is smaller than one radian. The field of view can be increased by using multi-conjugate adaptive optics, but this technique will not be described here.

A convenient solution allows to overcome the second problem. It consists in using a laser beam to generate a bright reference star (Laser Guide Star - LGS) in the vicinity of the target. A sodium laser produces a highly energetic beam at $589nm$, which excites the sodium atoms present in the upper layer of the atmosphere (90-120km), making them glow. The LGS can be used as a reference star in the same way as the natural guide star except that another natural reference star, however much fainter, is still needed for tip-tilt correction. Indeed, as the transmitted and received laser beams are both affected in the same way by the atmosphere. Such a system thus cannot sense the tip-tilt errors.

Information about adaptive optics can be found among other things in Roddier (2004) or Merkle (1993), see also Hickson (2008).

4.2 Optical aberrations

At the beginning of this chapter, we have presented the propagation equation of a perfect wavefront. This one is perfectly planar because its source is located very far away from the observer. When this wavefront comes into a perfect telescope, it is focused into a single point, which means that the planar wavefront is transformed in a spherical wavefront.

We have also seen that the atmosphere could disturb the incoming wavefront so that it is not planar any longer. Moreover, defects may also affect the optical instruments such that a planar incoming wavefront is not perfectly transformed into a spherical one. These defects are visible in images obtained with the instruments, and they are called optical aberrations.

This means that in the general case, not only the incoming wavefront is not planar but also the instrument introduces aberrations. If the atmosphere is considered as part of the instrument, the errors of the incoming wavefront can be added to the instrument aberrations.

Moreover, we know that the PSF characterizes the instrument. The aberration should thus be visible in this impulse response. We will show later in this section how to mathematically include these aberrations into the point spread function.

The basic concept that measures the importance of aberrations is called the Strehl ratio, defined as the ratio of the central intensity of the aberrated PSF to that of the diffraction limited PSF.

$$Sr = \frac{I(0)}{I_0(0)} \tag{4.22}$$

The Maréchal approximation (Terrien and Maréchal 1964) gives a relationship between the Strehl ratio and the RMS phase variance of the wavefront

$$Sr = \exp(-\sigma_\phi^2) \tag{4.23}$$

This equation is valid only for the low aberrations. Table 4.1 presents the Strehl ratio for several optical cases.

	Strehl ratio	WFE (rms)	λ	Width of the Airy disk
Classical optics	95%	$\lambda/28$	$0.55 - 5\mu$m	0.009"
Extreme adaptive optics	90%	$\lambda/20$	2.2μm	0.035"
Adaptive optics	60%	$\lambda/9$	2.2μm	0.05 "
Speckle interferometry	30%	$\lambda/6$	$0.55 - 12\mu$m	0.035"
Seeing	$1 - 3$%	$\lambda/3$	550nm	1.39 "

Table 4.1: Typical Strehl ratio corresponding to several optical cases (8m telescope). It should be noticed that the Strehl concept is well defined only for small aberrations (Strehl > 30%). In the case of speckle interferometry, the given Strehl ratio value is related to one particular speckle, but the total energy is spread out between all the speckles. An 8m aperture has been considered in all cases, except for the seeing where it is not applicable.

Beyond this basic concept, the aberrations are defined by polynomials. The wavefront consists of their sum. One of the most popular series of polynomials was invented by Zernike (Zernike 1934). They will be presented in section 4.2.1. Using these polynomials as a basis to express the aberrations in polar coordinates we will show a mathematical relation between the aberration and the PSF in section 4.2.2.

4.2.1 Zernike polynomials

The wave aberrations can be expanded in polynomial series. The Zernike polynomials correspond to a particular set of orthogonal polynomials, defined by:

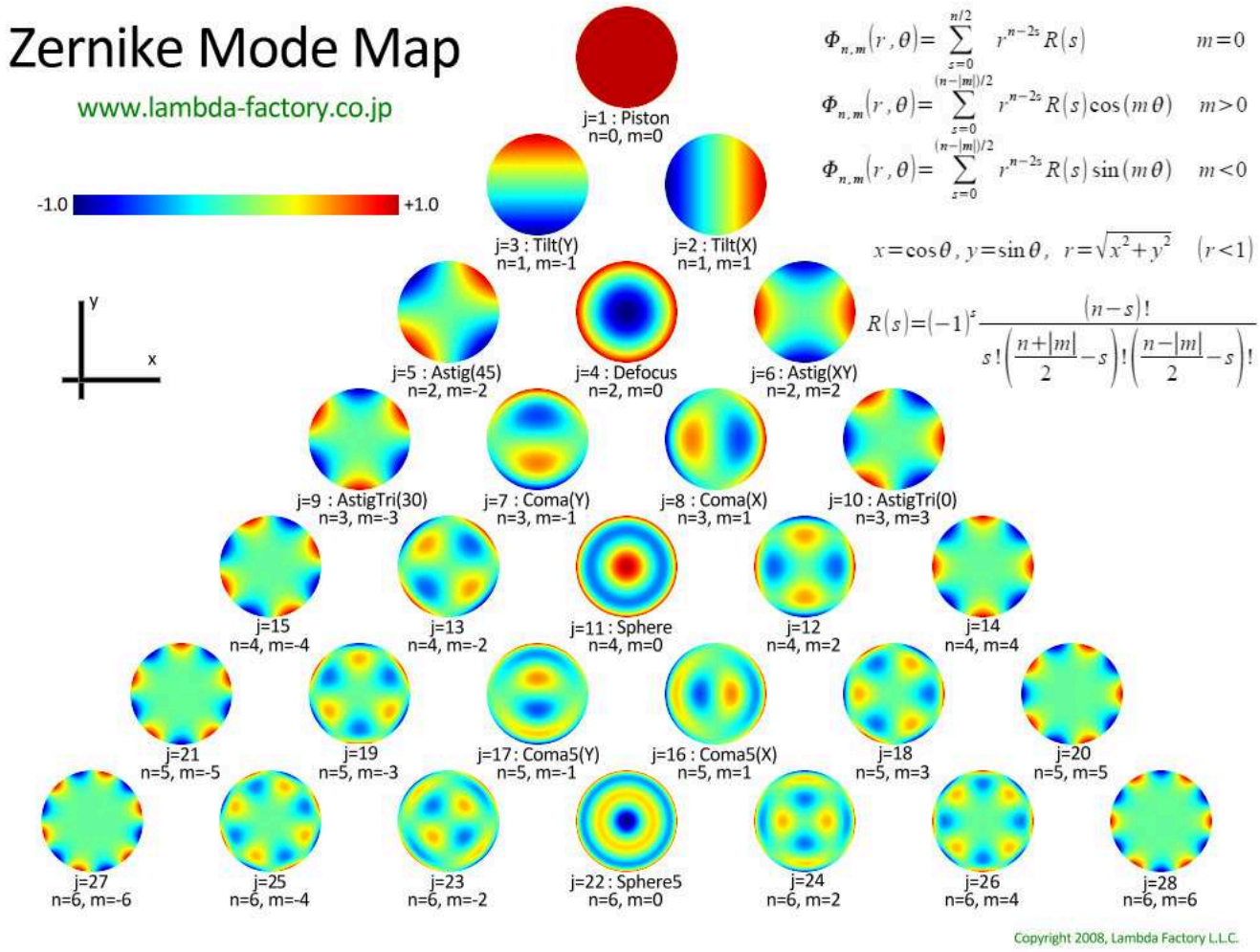
$$\begin{aligned} &R_n^m(\rho)\cos(m\theta) \\ &R_n^m(\rho)\sin(m\theta) \end{aligned} \tag{4.24}$$

for even and odd polynomials respectively. R_n^m is a radial polynomial of degree n in ρ containing terms ρ^n, ρ^{n-2}, ..., and ρ^m. It is defined as:

$$R_n^m(\rho) = \sum_{s=0}^{\frac{n-m}{2}} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}\rho^{n-2s} \tag{4.25}$$

The radial polynomials $R_n^m(\rho)$ are even or odd in ρ depending on the n, m values.

The particularity of these Zernike polynomials is their relation with the optical aberrations. Each of them corresponds to a particular aberration and the coefficients related to that polynomial represent the weight of the particular aberration in the wavefront. The aberrations are expressed in polar coordinates. Fig . 4.10 shows the first few polynomials and their corresponding aberrations.



$$\Phi_{n,m}(r,\theta)=\sum_{s=0}^{n/2} r^{n-2s}R(s) \qquad m=0$$

$$\Phi_{n,m}(r,\theta)=\sum_{s=0}^{(n-|m|)/2} r^{n-2s}R(s)\cos(m\theta) \qquad m>0$$

$$\Phi_{n,m}(r,\theta)=\sum_{s=0}^{(n-|m|)/2} r^{n-2s}R(s)\sin(m\theta) \qquad m<0$$

$$x=\cos\theta,\ y=\sin\theta,\ r=\sqrt{x^2+y^2} \qquad (r<1)$$

$$R(s)=(-1)^s\frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}$$

Figure 4.10: Phase aberrations corresponding the first Zernike polynomials. (www.lambda-factory.co.jp)

A recurrence relation exists between the radial polynomials

$$\begin{aligned}
\rho R_n^m(\rho) &= \frac{1}{2(n+1)}\left[(n+m+2)R_{n+1}^{m+1}(\rho)+(n-m)R_{n-1}^{m+1}(\rho)\right] \\
\rho R_{n+2}^m(\rho) &= \frac{n+2}{(n+2)^2-m^2}\left\{\left[4(n+1)\rho^2-\frac{(n+m)^2}{n}-\frac{(n-m+2)^2}{n+2}\right]R_n^m(\rho)\right. \\
&- \left.\frac{n^2-m^2}{n}R_{n-2}^m(\rho)\right\} \qquad (4.26)
\end{aligned}$$

More information about the Zernike polynomials can be found in Zernike (1934) (in German), Mahajan (1998) or Wyant and Creath (1992).

The polynomials presented above are well suited for a full circular pupil. However, most telescopes have a central obscuration. Mahajan (1981) presents a particular form of annular Zernike polynomials that can be more convenient to use is such cases.

4.2.2 The pupil function

The pupil function describes the aberrated optical system. It corresponds to the shape of the wavefront at the output pupil. The general expression of this function in polar coordinates is:

$$P(\rho,\theta) = A(\rho,\theta)\exp(i\Phi(\rho,\theta)) \tag{4.27}$$

where $A(\rho,\theta)$ is the transmission function and $\Phi(\rho,\theta)$ is the actual aberrated phase.

The classical approach consists in expanding the phase term alone in a sum of Zernike polynomials:

$$\Phi(\rho,\theta) = \sum_{n,m} \alpha_n^m Z_n^m(\rho,\theta) \tag{4.28}$$

where $Z_n^m(\rho,\theta)$ is the Zernike polynomial which corresponds to the particular degree n and order m, and α_n^m is the Zernike coefficient that measures the weight of the corresponding aberration. The important limitation in this approach is its unability to represent amplitude errors due for example to transmission trouble or apodization.

We thus consider a more general way of using the Zernike polynomials. Instead of only expanding the phase, the whole pupil function is decomposed as

$$P(\rho,\theta) = A(\rho,\theta)exp(i\Phi(\rho,\theta)) = \sum_{n,m} \beta_n^m Z_n^m(\rho,\theta) \tag{4.29}$$

where the $Z_n^m(\rho,\theta)$ are the Zernike polynomials and the β_n^m are the complex general Zernike coefficients. In this case, the interpretation of the β coefficients is far less obvious than for the α terms.

4.3 Calculation of the PSF from the pupil function

The point spread function of a system can be computed using the Fraunhofer approximated equation 4.3 for the diffractive propagation of light. This calculation may take several forms. It has been shown in section 4.1.1 that the propagation of light can be computed as the Fourier transform of the pupil function. This is the Fourier optics approach that will be exposed hereafter (section 4.3.1). We also showed that the pupil function can be expressed as a series of Zernike polynomials that constitute a polar basis. This particularity will be used in the Nijboer-Zernike (hearafter NZ) approach that will be presented in section 4.3.2. The NZ computation of the PSF is based on a polar Fourier transform called Hankel transform of the pupil function.

4.3.1 The Fourier approach

As previously discussed, the propagation of waves is ruled by the diffraction theory. In the Fraunhofer approximation this propagation is described by the Fourier optics. The relation between the exit pupil and the image of an optical system is the Fourier transform of the pupil (see section 4.1.1).

$$U(u,v) = \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)\ \exp(-2i\pi\ (ux+vy))\,dx\,dy \tag{4.30}$$

where (u,v) and (x,y) are the coordinates respectively in the image plane and in the pupil plane and $P(x,y)$ is the pupil function. The $1/\pi$ factor comes from the diffraction equation 4.3. We showed that the pupil function can be expressed as a function of the Zernike polynomials

$$U(u,v) = \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x,y)\exp(i\Phi(x,y))\ \exp(-2i\pi\ (ux+vy))\,dx\,dy \tag{4.31}$$

In practice the Fourier transform is computed using the "Fast Fourier Transform" (FFT) algorithm. Using this algorithm requires the adequate sampling in the image plane (the Nyquist-Shannon theorem requires $\lambda/D > 2$pix). Under these conditions, the size of the table (matrix) containing the pupil pixels[19] is given by

$$S_{\text{pup}} = \frac{S_{\text{matrix}}}{2Samp_{\text{image}}} \tag{4.32}$$

where S_{pup} is the diameter of the pupil in pixels, S_{matrix} is the size of the table that should be used for the Fourier transform and $Samp_{\text{image}}$ is the sampling of the final image. This relation means that, in order to get a sampling of 2 pixels, the pupil should be contained in a matrix that is four times larger than the pupil diameter. Table 4.2 shows the typical matrix size for several optical applications.

Sampling (pixels)	1024x1024	2048x2048	4096x4096
Application	classical optics	coronagraphy	Extremely Large Telescope
Reason		high sampling required	large pupil

Table 4.2: Typical sizes of the matrix that should be used as input to apply the FFT. The size of the input matrix determines the sampling of both the image plane and the pupil plane.

This sampling consideration can be an issue when expressing the pupil as a function of the Zernike polynomials. Indeed the latter require a good sampling of the pupil, especially for high order polynomials that have high azimuthal frequencies. This means that appling an FFT to this type of pupil would require a very large table and could take much computation time.

[19]Numerical images consist of matrices of pixels containing intensity information. The size of these matrices influences the sampling of the image.

4.3.2 The Nijboer-Zernike approach

Most optical systems are composed of circular optical elements. It is thus obviously interesting to use polar coordinates for the analytical computation of the point spread function produced by such systems. The Nijboer-Zernike approach, developed in Janssen (2002), consists in expressing the equation of the propagation of light in terms of Bessel functions in a polar coordinate basis, which is convenient to manipulate the pupil function in terms of Zernike polynomials. Another good point of this method is its invertibility, which will be presented in section 5.2.

Our approach is slightly different from that of Janssen (2002) since he starts with an equation of diffraction whereas we use the Fourier transform as a basis for the propagation of waves. In particular, our approach involves an inversion of the image coordinate axes which makes the Nijboer-Zernike theory consistent with the Fourier approach. Indeed, the development used in Janssen (2002) leads to the inversion of the sign of aberrations related to odd radial orders. Moreover, he considers symmetrical cases for which a part of the aberration is neglected. In our case, we consider a general development involving all possible aberrations.

In this section, we derive a general expression to compute the complex amplitude of a system, defined by the aberrations it introduces in a wavefront, for several positions around its focus. The expression we are going to establish is

$$U(r,\phi,f) = \sum_{n,m} \beta_{cn}^m \, 2i^m \, \cos(m\phi) \, V_n^m(r,f) + \sum_{n,m} \beta_{sn}^m \, 2i^m \, \sin(m\phi) \, V_n^m(r,f) \tag{4.33}$$

where r, ϕ are the polar coordinates in the image plane, β_{cn}^m and β_{sn}^m are the general Zernike aberration coefficients and f is the distance between the focus and the position of the image (defocus). The $V_n^m(r,f)$ function is defined by

$$V_n^m(r,f) = (-1)^m \int_0^1 \rho \, \exp(if\rho^2) R_n^m(\rho) J_m(2\pi r\rho) \, d\rho. \tag{4.34}$$

We base the further reasoning on the Fourier transform, presented in the previous section (equation 4.30), that we convert in polar coordinates. Expressing the pupil as a function of its aberrations, we can extract the defocus term, which will allow to compute the PSF of the system for any position around its focus.

After several mathematical manipulations, the Fourier transform can be expressed in the basis of the Bessel functions of the first kind ($J_m(r)$). The result is given in equation 4.33. The detailed mathematical development, that leads to this result, is presented here.

To derive the Nijboer-Zernike equations, we begin with the conversion of the Fourier transform (equation 4.30) in polar coordinates by using the following variable change

$$\begin{aligned} x + iy = \rho \exp((i\theta)) \Rightarrow x = \rho\cos(\theta), \quad y = \rho\sin(\theta), \quad \rho = \sqrt{x^2+y^2} \\ u + iv = r \exp((i\phi)) \Rightarrow u = r\cos(\phi), \quad v = r\sin(\phi), \quad r = \sqrt{u^2+v^2} \end{aligned} \tag{4.35}$$

we get

$$U(r,\phi) = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^{2\pi} P(\rho,\theta) \, \exp\{-2i\pi \, [r\cos(\phi)\rho\cos(\theta) + r\sin(\phi)\rho\sin(\theta)]\} \, \rho \, d\rho \, d\theta \tag{4.36}$$

Rearranging the terms and using the cosine addition formula, we obtain

$$U(r,\phi) = \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{0}^{2\pi} P(\rho,\theta)\ \exp(-2i\pi\ r\ \rho\ \cos(\theta-\phi))\ \rho\, d\rho\, d\theta \tag{4.37}$$

The integration limits are defined by the pupil function, their ranges cover $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$. Replacing these limits in the equation, we get

$$U(r,\phi) = \frac{1}{\pi}\int_{0}^{1}\int_{0}^{2\pi} P(\rho,\theta)\ \exp(-2i\pi\ r\ \rho\ \cos(\theta-\phi))\ \rho\, d\rho\, d\theta \tag{4.38}$$

The defocus term can be extracted from the phase of the pupil. In this a way, we get access to the focal position of the images, which is very interesting, especially for the inverse problem that will be treated in section 5.2.

$$U(r,\phi,f) = \frac{1}{\pi}\int_{0}^{1}\int_{0}^{2\pi} \exp(if\rho^2)P(\rho,\theta)\ \exp(-2i\pi\ r\ \rho\ \cos(\theta-\phi))\ \rho\, d\rho\, d\theta \tag{4.39}$$

were $\exp(if\rho^2)$ represents the phase difference in the exit pupil due to a focal shift (f) in the image plane. Indeed, the Zernike polynomial corresponding to the defocus aberration is $\rho^2 - 1$. As the pupil is described by the exponential of a sum of Zernike polynomials, it is possible to extract $\exp(if(\rho^2-1))$ from the pupil term. The second part of this term ($\exp(-if)$) corresponds to a piston term that can be included in the pupil function $P(\rho,\theta)$. The f parameter appearing in the exponential is the amount of defocus introduced for the computation of the PSF in the focal region. This defocus is expressed in wavelength units.

Equation 4.39 expresses the propagation of waves in polar coordinates according to the diffraction theory and to the Fraunhofer conditions. In order to compute the complex amplitude in the focal region, the explicit pupil function has to be introduced. We will consider both the general and the classical Zernike expansions.

General Zernike expansion

Replacing the pupil function by its general Zernike expansion (4.29) described in section 4.2.2, and reminded here

$$P(\rho,\theta) = \sum_{n,m} \beta_n^m Z_n^m(\rho,\theta) \tag{4.40}$$

we find

$$U(r,\phi,f) = \frac{1}{\pi}\int_{0}^{1}\int_{0}^{2\pi} \exp(if\rho^2)\sum_{n,m}\beta_n^m Z_n^m(\rho,\theta)\ \exp(-2i\pi\ r\ \rho\ \cos(\theta-\phi))\ \rho\, d\rho\, d\theta \tag{4.41}$$

Denoting $U_n^m(r,\phi,f)$ the complex amplitude corresponding to the Zernike polynomial (n,m), its expression is given by

$$U_n^m(r,\phi,f) = \frac{1}{\pi}\int_{0}^{1}\int_{0}^{2\pi} \exp(if\rho^2)Z_n^m(\rho,\theta)\ \exp(-2i\pi\ r\ \rho\ \cos(\theta-\phi))\ \rho\, d\rho\, d\theta \tag{4.42}$$

Relation 4.41 can then be written as

$$U(r,\phi,f) = \sum_{n,m} \beta_n^m U_n^m(r,\phi,f) \tag{4.43}$$

Now let us consider equation 4.42. Applying a phase shift of $+\phi$ does not modify the complex amplitude $U_n^m(r,\phi,f)$, we can thus write

$$U_n^m(r,\phi,f) = \frac{1}{\pi}\int_0^1\int_0^{2\pi} \exp(if\rho^2) Z_n^m(\rho,(\theta+\phi))\ \exp(-2i\pi\ r\ \rho\ \cos(\theta))\ \rho\, d\rho\, d\theta \tag{4.44}$$

Two cases should be considered because of the definition of the Zernike polynomials $Z_n^m(\rho,\theta)$

$$Z_n^m(\rho,\theta) = R_n^m(\rho) \begin{cases} \cos(m\theta) \text{ even polynomials} \\ \sin(m\theta) \text{ odd polynomials} \\ 1 \text{ when } m=0 \end{cases} \tag{4.45}$$

The polynomials corresponding to $m = 0$ can be treated as even polynomials with $cos(0) = 1$. We thus will have two cases to develop, corresponding to the even Zernike polynomials and to the odd ones.

$$\begin{aligned} U_n^m(r,\phi,f) &= \frac{1}{\pi}\int_0^1\int_0^{2\pi} \exp(if\rho^2) R_n^m(\rho)\cos(m(\theta+\phi))\exp(-2i\pi r\rho\cos(\theta))\rho d\rho d\theta \ \textbf{even} \\ U_n^m(r,\phi,f) &= \frac{1}{\pi}\int_0^1\int_0^{2\pi} \exp(if\rho^2) R_n^m(\rho)\sin(m(\theta+\phi))\exp(-2i\pi r\rho\cos(\theta))\rho d\rho d\theta \ \textbf{odd} \end{aligned} \tag{4.46}$$

Using simple trigonometry formula, we find,

$$\begin{aligned} U_n^m(r,\phi,f) &= \frac{1}{\pi}\int_0^1\int_0^{2\pi} \exp(if\rho^2) R_n^m(\rho)(\cos(m\theta)\cos(m\phi) - \sin(m\theta)\sin(m\phi)) \\ &\quad \exp(-2i\pi r\rho\cos(\theta))\rho d\rho d\theta \ \textbf{even} \\ U_n^m(r,\phi,f) &= \frac{1}{\pi}\int_0^1\int_0^{2\pi} \exp(if\rho^2) R_n^m(\rho)(\sin(m\theta)\cos(m\phi) + \cos(m\theta)\sin(m\phi)) \\ &\quad \exp(-2i\pi r\rho\cos(\theta))\rho d\rho d\theta \ \textbf{odd} \end{aligned} \tag{4.47}$$

In both cases the integral between 0 and 2π of the term $\sin(m\theta)$ vanishes. Moreover using the definition of the Bessel functions,

$$J_m(-2\pi r\rho) = (-1)^m\, J_m(2\pi r\rho) = \frac{(-1)^m i^{-m}}{\pi}\int_0^{\pi} \exp(2i\pi r\rho\cos(\theta))\cos(m\theta)d\theta \tag{4.48}$$

and replacing in the previous equation, we find

$$\begin{aligned} U_n^m(r,\phi,f) &= 2i^m \cos(m\phi)(-1)^m \int_0^1 \rho \, \exp(if\rho^2) R_n^m(\rho) J_m(2\pi r\rho)\, d\rho \;\; \textbf{even} \\ U_n^m(r,\phi,f) &= 2i^m \sin(m\phi)(-1)^m \int_0^1 \rho \, \exp(if\rho^2) R_n^m(\rho) J_m(2\pi r\rho)\, d\rho \;\; \textbf{odd} \end{aligned} \tag{4.49}$$

The integral term in this equation is called $V_n^m(r,f)$ so that we have,

$$\begin{aligned} U_n^m(r,\phi,f) &= 2i^m \, \cos(m\phi)\, V_n^m(r,f) \;\; \textbf{even} \\ U_n^m(r,\phi,f) &= 2i^m \, \sin(m\phi)\, V_n^m(r,f) \;\; \textbf{odd} \end{aligned} \tag{4.50}$$

where,

$$V_n^m(r,f) = (-1)^m \int_0^1 \rho \, \exp(if\rho^2) R_n^m(\rho) J_m(2\pi r\rho)\, d\rho \tag{4.51}$$

corresponds to equation 4.34 that has been presented at the beginning of this section. The complex amplitude is thus given by the sum of all considered Zernike coefficients multiplied by the corresponding $V_n^m(r,f)$ and by a sine or cosine function depending on the parity of the aberration term. The transformation presented above (the Fourier transform expressed as the integral of the Bessel functions) is called the Hankel transform. The $V_n^m(r,f)$ functions represent a particular expression of the radial Zernike polynomials in the image plane. $V_n^m(r,f)$ is the Hankel transform of order m of the radial Zernike polynomial $R_n^m(r)$.

Based on the parity of their associated polynomials, the β_n^m coefficients can be separated in two groups. Let us define β_{cn}^m the set of coefficients associated with the even polynomials (in $\cos(m\theta)$ or with $m = 0$) and β_{sn}^m the coefficients related to the odd polynomials (in $\sin(m\theta)$, with $m \neq 0$). The complex amplitude can then be written as in equation 4.33:

$$U(r,\phi,f) = \sum_{n,m} \beta_{cn}^m \, 2i^m \, \cos(m\phi)\, V_n^m(r,f) + \sum_{n,m} \beta_{sn}^m \, 2i^m \, \sin(m\phi)\, V_n^m(r,f). \tag{4.52}$$

The intensity PSF, which is the only measurable quantity, is given by:

$$I_{PSF}(r,\phi,f) = |U(r,\phi,f)|^2 \tag{4.53}$$

Constant amplitude A

The particular case where the amplitude is uniform over the whole pupil can be treated using the classical Zernike expansion for the pupil function (equations 4.27 and 4.28). Considering $A(r,\theta) = 1$ for the sake of simplicity, we get the following pupil function:

$$P(\rho,\theta) = \exp(i\Phi(\rho,\theta)) \tag{4.54}$$

that can be inserted, in equation 4.39,which becomes

$$U(r,\phi,f) = \frac{1}{\pi} \int_0^1 \int_0^{2\pi} \rho \exp(if\rho^2) \, \exp(i\Phi(\rho,\theta)) \exp(-2\pi i\rho r \cos(\theta-\phi)) d\rho d\theta \tag{4.55}$$

Applying the same phase shift as in the previous case, we get

$$U(r,\phi,f) = \frac{1}{\pi}\int_0^1\int_0^{2\pi} \rho \exp(if\rho^2)\ \exp(i\Phi(\rho,\theta+\phi))\exp(-2i\pi\rho r\cos(\theta))d\rho d\theta \tag{4.56}$$

where the exponential can be expanded in Taylor series. Provided that the phase aberration is small, ($\Phi << 1$), this series can be limited to the first order

$$\exp(i\phi) = \sum_{k=0}^{\infty}\frac{i^k}{k!}\Phi^k \approx 1 + i\Phi \tag{4.57}$$

Equation 4.56 becomes

$$\begin{aligned} U(r,\phi,f) = \int_0^1 \rho\exp(if\rho^2)\ \Big[\quad & \frac{1}{\pi}\int_0^{2\pi}\exp(-2i\pi\rho r\cos(\theta))d\theta \\ & +\frac{i}{\pi}\int_0^{2\pi}\Phi(\rho,\theta+\phi)\exp(-2i\pi\rho r\cos(\theta))d\theta\Big]d\rho \end{aligned} \tag{4.58}$$

Let us call I_1 the first integral and I_2 the second one. For I_1 we find,

$$I_1 = \frac{1}{\pi}\int_0^{2\pi}\exp(-2i\pi\rho r\cos(\theta))d\theta \tag{4.59}$$

Using the expression of the Bessel function (equation 4.48) with $m = 0$

$$J_0(-2\pi\rho r) = J_0(2\pi\rho r) = \frac{1}{\pi}\int_0^{\pi}\exp(-2i\pi\rho r\cos(\theta))d\theta \tag{4.60}$$

We can thus express I_1 as a function of J_0

$$I_1 = 2J_0(2\pi\rho r) \tag{4.61}$$

As far as I_2 is concerned, replacing Φ by its classical Zernike expansion gives,

$$I_2 = \frac{i}{\pi}\sum_{n,m}\alpha_n^m\int_0^{2\pi} Z_n^m(\rho,\theta+\phi)\exp(-2i\pi\rho r\cos(\theta))d\theta \tag{4.62}$$

As in the general case, we now replace Z_n^m by its expression for both the even and the odd cases:

$$\begin{aligned} I_2 &= \frac{i}{\pi}\sum_{n,m}\alpha_n^m R_n^m(\rho)\int_0^{2\pi}\cos(m(\theta+\phi))\exp(-2i\pi\rho r\cos(\theta))d\theta \ \ \textbf{even} \\ I_2 &= \frac{i}{\pi}\sum_{n,m}\alpha_n^m R_n^m(\rho)\int_0^{2\pi}\sin(m(\theta+\phi))\exp(-2i\pi\rho r\cos(\theta))d\theta \ \ \textbf{odd} \end{aligned} \tag{4.63}$$

once again, simple trigonometry gives, for respectively the even and odd polynomials,

$$
\begin{aligned}
I_2 &= \frac{i}{\pi}\sum_{n,m}\alpha_n^m R_n^m(\rho)\,\Big\{\int_0^{2\pi}\cos(m\theta)\exp(-2i\pi\rho r\cos(\theta))d\theta\cos(m\phi)\\
&- \int_0^{2\pi}\sin(m\theta)\exp(-2i\pi\rho r\cos(\theta))d\theta\sin(m\phi)\Big\}\ \textbf{even}\\
I_2 &= \frac{i}{\pi}\sum_{n,m}\alpha_n^m R_n^m(\rho)\,\Big\{\int_0^{2\pi}\sin(m\theta)\exp(-2i\pi\rho r\cos(\theta))d\theta\cos(m\phi)\\
&+ \int_0^{2\pi}\cos(m\theta)\exp(-2i\pi\rho r\cos(\theta))d\theta\sin(m\phi)\Big\}\ \textbf{odd}
\end{aligned}
\tag{4.64}
$$

The integrals over θ for the term $\sin(m\theta)$ vanish.

$$
\begin{aligned}
I_2 &= \frac{i}{\pi}\sum_{n,m}\alpha_n^m R_n^m(\rho)\int_0^{2\pi}\cos(m\theta)\exp(-2i\pi\rho r\cos(\theta))d\theta\cos(m\phi)\ \textbf{even}\\
I_2 &= \frac{i}{\pi}\sum_{n,m}\alpha_n^m R_n^m(\rho)\int_0^{2\pi}\cos(m\theta)\exp(-2i\pi\rho r\cos(\theta))d\theta\sin(m\phi)\ \textbf{odd}
\end{aligned}
\tag{4.65}
$$

where the integrals can be transformed using the Bessel functions defined by equation 4.48

$$
\begin{aligned}
I_2 &= 2i^{m+1}(-1)^m\sum_{n,m}\alpha_n^m R_n^m(\rho)\cos(m\phi)J_m(2\pi\rho r)\ \textbf{even}\\
I_2 &= 2i^{m+1}(-1)^m\sum_{n,m}\alpha_n^m R_n^m(\rho)\sin(m\phi)J_m(2\pi\rho r)\ \textbf{odd}
\end{aligned}
\tag{4.66}
$$

Replacing the expression of the two integrals I_1, I_2 in equation 4.58, we get the expression of the complex amplitude,

$$
\begin{aligned}
U(r,\phi,f) &= 2\int_0^1\rho\exp(if\rho^2)J_0(2\pi\rho r)dr\\
&+ 2\pi i\sum_{n,m}\alpha_n^m i^m\cos(m\phi)(-1)^m\int_0^1\rho\exp(if\rho^2)R_n^m(\rho)\cos(m\phi)J_m(2\pi\rho r)d\rho\ \textbf{even}\\
U(r,\phi,f) &= 2\int_0^1\rho\exp(if\rho^2)J_0(2\pi\rho r)dr\\
&+ 2\pi i\sum_{n,m}\alpha_n^m i^m\cos(m\phi)(-1)^m\int_0^1\rho\exp(if\rho^2)R_n^m(\rho)\sin(m\phi)J_m(2\pi\rho r)d\rho\ \textbf{odd}
\end{aligned}
\tag{4.67}
$$

Defining the $V_n^m(r,f)$ function as in the general case (4.51), we get:

$$
\begin{aligned}
U(r,\phi,f) &= 2V_{00}(r,f) + 2i\sum_{n,m}\alpha_n^m i^m\cos(m\phi)V_n^m(r,f)\ \textbf{even}\\
U(r,\phi,f) &= 2V_{00}(r,f) + 2i\sum_{n,m}\alpha_n^m i^m\sin(m\phi)V_n^m(r,f)\ \textbf{odd}
\end{aligned}
\tag{4.68}
$$

As in the general case, we can separate the α_n^m coefficients in two sets: the α_{cn}^m coefficients associated with the even Zernike polynomials (and $m = 0$ term) and the α_{sn}^m associated with the odd ones (without the $m = 0$ term). The final expression of the complex amplitude when the transmission ($A(r, \theta)$) is uniform over the whole pupil and constant through the optical system is thus

$$U(r, \phi, f) = 2V_{00}(r, f) + 2i \sum_{n,m} \alpha_{cn}^m i^m \cos(m\phi) V_n^m(r, f) + 2i \sum_{n,m} \alpha_{sn}^m i^m \sin(m\phi) V_n^m(r, f) \quad (4.69)$$

where the α_{cn}^m and α_{sn}^m are the classical Zernike coefficients related to the cosine and sine polynomials.

Extension to the case of a finite source size

The previous development was made for the case of a point source as the origin of the wavefront. However, this cannot be used in practical cases. A convenient way of extending this development to a larger source was presented in Dirksen et al. (2003). They assume that the source size is small compared to the coherence radius of the source and translate the effect of this enlargement by multiplying the pupil function with the Fourier transform of a disk. This corresponds to the Zernike-Van cittert theorem that is usually applied in interferometry, for stellar disk measurements,

$$\frac{J_1(2\pi a\rho)}{\pi a\rho} \quad \text{with} 0 \leq \rho \leq 1 \quad (4.70)$$

where a is the normalized diameter of the source. This expression can be approximated by

$$\frac{J_1(2\pi a\rho)}{\pi a\rho} \approx \exp(c - d\rho^2) \quad (4.71)$$

where optimal c, d can be computed as a function of $b = 2\pi a$,

$$c = \frac{b^4}{2304} + \frac{b^6}{46080},\ d = \frac{b^2}{8} + \frac{b^4}{384} + \frac{b^6}{10240} \quad (4.72)$$

The extension of the method then consists in replacing the $V_n^m(r, f)$ function with

$$\exp(c)\, V_n^m(r, f + id) \quad (4.73)$$

The focus parameter is replaced by a complex focus for the computation of the V_n^m functions.

Chapter 5

Aberration retrieval

Aberration retrieval consists in measuring the phase errors in a wavefront that are caused by the system it goes through. Many techniques exist to measure these wavefront distortions with different accuracies, they will be presented in the following sections. Some of them use the PSF of the system to compute the aberrations. This is particularly convenient to determine the aberration of large optical systems, such as telescopes.

The aim of measuring optical system aberrations is to minimize them in order to obtain the best possible image quality, with this instrument.

5.1 Optical testing and classical phase retrieval method

This section presents several optical testing methods commonly used to characterize classical solid-mirror telescopes. It is important, when building a telescope, to be able to characterize the quality of its optics, not only of the primary mirror but also all the other optical systems aligned along the optical path of the incoming light (corrector,...). The tests used to characterize the optics are very important and it is mandatory to know exactly which test is best suited for each particular case and what is its expected precision.

Being able to measure the phase errors of the wavefront coming from the tested optics is also very useful for other applications such as adaptive optics. In this case, the optical system is assumed to introduce no aberration or very little and the measured wavefront error is supposed to correspond to the aberrations introduced by the atmosphere. It is thus possible to correct for the atmospheric perturbations (see section 4.1.6). This application requires real time measurements and corrections of the wavefront and must not consume too much incoming light.

Combination of the two previous cases can also be studied. Considering an imperfect instrument looking at stars through the atmosphere, both the imperfection and the atmospheric turbulence will distort the wavefront. Even if a single measurement does not allow to discriminate between the origin of the aberrations, it may be possible to isolate the aberrations due to the instrument, by using several observations. Indeed, the aberrations related to the atmosphere are rapidly variable (corrected by adaptive optics) while those due to the instrument do not vary with time (or very slowly, they are compensated by active optics[20]). Making several measurements thus allows to discriminate between the static and the variable aberrations and then to

[20] Active optics allows to compensate for small deformations of the mirror related to is position.

determine those related to the instrument and to correct them. This can be used with extreme adaptive optics such as GPI (Gemini Planet Imager) or SPHERE (Spectro-Polarimetric High-contrast Exoplanet REsearch). Measurement of the quasi-static aberrations (those that slowly evolve with time) allows to calibrate the instruments, and these aberrations can be corrected for by active optics. We will illustrate the measurement of the static aberrations with the NACO instrument in chapter 8.

Other astronomical instruments, such as coronagraphs also require highly accurate optics. An accuracy of one thousandth of a wavelength can be required. To reach such an accuracy it is necessary to have a very good and reliable testing method, and this one has to be executed in a thermally controlled environment to avoid external perturbations.

As far as the ILMT is concerned, the images will be acquired in the TDI mode. It would be very interesting to be able to measure the TDI deformation on a star background in order to calibrate the point spread function subtraction method.

5.1.1 The shape of the tested surface

It is important to adapt the tests performed to the specific tested surface. The common point between all wavefront measurement methods, is the need of a source and an observer. These two elements have to be correctly located in order to ensure the good quality of the results.

Depending on the surface to test, the position of the source and detector has thus to be carefully selected. Indeed, these elements must be located at the conjugated stigmatic points of the surface. These points have a particularity: the light coming from one of them converges to the other one. Using these points ensures that the observer is located near the point where all rays would perfectly converge if the surface had a perfect shape. For a parabolic surface, these points correspond to the focus of the parabola and at the infinity, whereas for a spherical surface they are both located at the center of curvature of the sphere. In the last case, the source and the observer are slightly shifted in opposite directions (perpendicularly to the optical axis) as they cannot be both exactly located at the same place.

5.1.2 The Foucault test

The Foucault method consists in comparing the tested surface to reference osculating spheres. The measurement of the curvature of the mirror is performed by comparing it with these references, simply by occulting the reflected beam at particular positions that correspond to the center of curvature of these osculating spheres. This process is called the "cut"[21].

The principle of the cut

Let us consider a perfect spherical mirror, and a source located at its center of curvature. An observer located in the focusing region (near the center of curvature) will see a fully illuminated mirror (fig. 5.1(a)). Inserting a "knife-edge" in the reflected rays results in different cases that are illustrated in fig. 5.1. When the knife-edge cuts the reflected rays but does not completely

[21]This corresponds to the obscuration of a part of the light by a knife-edge.

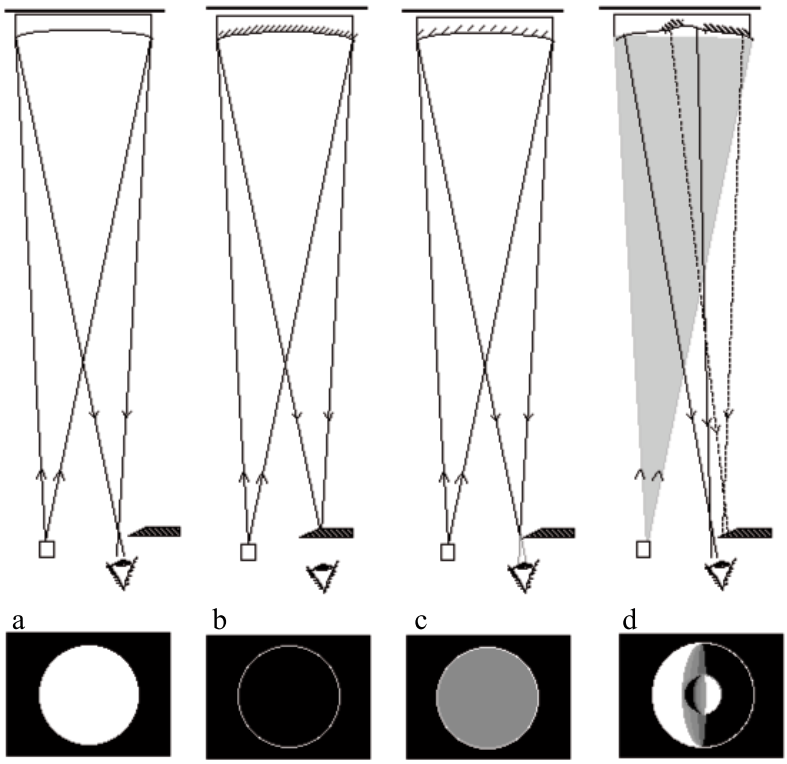


Figure 5.1: A spherical mirror illuminated by a source located slightly to the left of the center of curvature is observed slightly at the right of the center of curvature. The figure illustrates the effect of the introduction of a knife-edge into the reflected beam. From left to right: a) when the knife is outside the beam, the observer sees a fully illuminated mirror. b) In the case where the knife totally obstructs the beam, the mirror gets dark. c) If the knife edge is exactly on the axis of the beam where the rays converge, the latter are all partially attenuated in the same way over the whole pupil. In this case the observer will see the whole mirror "gray", because the darkening is uniform over the pupil, but the obstruction is not complete. d) If the mirror is not perfectly spherical, the introduction of the knife-edge reveals its defects. (figure from Koechlin 1990)

obscure them (fig. 5.1(d)), the observer sees zones of the mirror that are illuminated and others that are shadowed, this mixing of shadow and light looks like bumps and holes. In the particular case where the knife-edge is exactly where the rays converge, the illumination decreases uniformly over the whole pupil at the same time. This results in a global "graying out" (fig. 5.1(c)) and the mirror is said to appear flat, the bumps and holes have disappeared.

The distance between the knife-edge and the mirror defines the center of curvature of a particular osculating sphere to which the mirror is compared. In the situation presented in fig. 5.1(c) the mirror perfectly corresponds to the reference sphere (the rays converge to the center of curvature of the mirror and the knife-edge is precisely located on the focusing point). If the mirror is not perfect, it does not correspond to a unique sphere. Each area of the mirror has its own center of curvature that corresponds to a particular sphere, this is illustrated in fig. 5.1(d). The knife-edge stops some of the rays but not all of them as it was the case for a perfect mirror, which reveals the defects of the mirror to the observer.

Aberration measurement

As previously explained, using the cut technique allows to compare the mirror (or another surface to test) with reference spheres, defined by the position of the knife-edge. This causes the defects to appear and it allows to measure them. Indeed, moving the knife-edge along the mirror axis corresponds to a change of reference spheres. it is thus possible to compare the mirror with these references and then to measure the curvature radius of the defects present on the mirror.

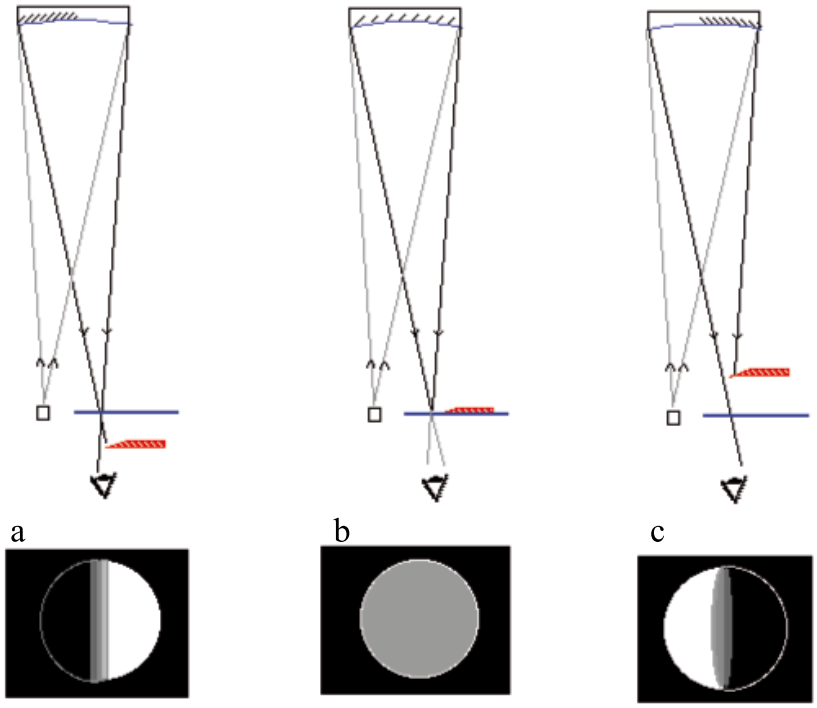


Figure 5.2: The "cut" allows to compare the mirror with osculating spheres. Moving the knife-edge along the axis of the mirror allows to measure the curvature of this mirror. Indeed, the position of the knife defines the center of curvature of a particular reference sphere. Moving this knife thus results in changing the reference. In case an area of the mirror has the same curvature as the particular reference sphere determined by the knife-edge position along the axis, this area appears grey to the observer. This, means that the difference between this zone and the osculating sphere is flat (b). If the sphere does not correspond to any part of the mirror, the latter seems only partially illuminated (a,c). This gives a visual effect of curvature of the surface. (figure from Koechlin 1990)

Let us move the knife-edge along the optical axis of a perfectly spherical mirror. When the knife is located before or after the focus of the reflected beam (fig. 5.2(a,c)), the observer sees a partially illuminated mirror. About half of the mirror disappears as the knife cuts a part of the rays corresponding to the left or the right part of the mirror, depending if the knife is introduced after (a) or before (c) the focus. Once again, if the cut occurs at the focus of the perfect sphere (fig. 5.2(b)), the whole mirror seems to be uniformly less illuminated. The position of the knife-edge along the optical axis determines the radius of curvature of the reference sphere, if it corresponds to a part of the mirror, this part will appear "flat". The Foucault test is thus based on the comparison of the surface to measure with a series of reference spheres to determine the curvature of the different parts of the mirror. It then becomes possible to measure the aberrations which affect the mirror.

Let us now consider an imperfect mirror, presenting a central hole. Moving the knife-edge along the optical axis allows to measure the defects in the surface as presented in fig. 5.3. The different parts of the mirror are compared to a series of osculating spheres.

The test of the surface then takes place as follows. The knife-edge is introduced in the beam well before the focus of the mirror (fig. 5.3(a)), it is then moved toward the focus until a part of the mirror appears flat (fig. 5.3(b)). The corresponding part of the mirror can be assimilated to a sphere with a radius of curvature that is given by the position of the knife-edge. While moving the knife toward him, the observer will see a succession of mirror illuminations, some of them show flat zones and others do not. The positions illustrated in fig. 5.3(c,d,e) respectively correspond to intermediate positions between the hole radius and the mirror radius. Each time a zone of the mirror appears flat, it corresponds to the particular reference sphere related to the position of the knife and its curvature can be measured. This process can be repeated until each zone of the mirror has been measured.

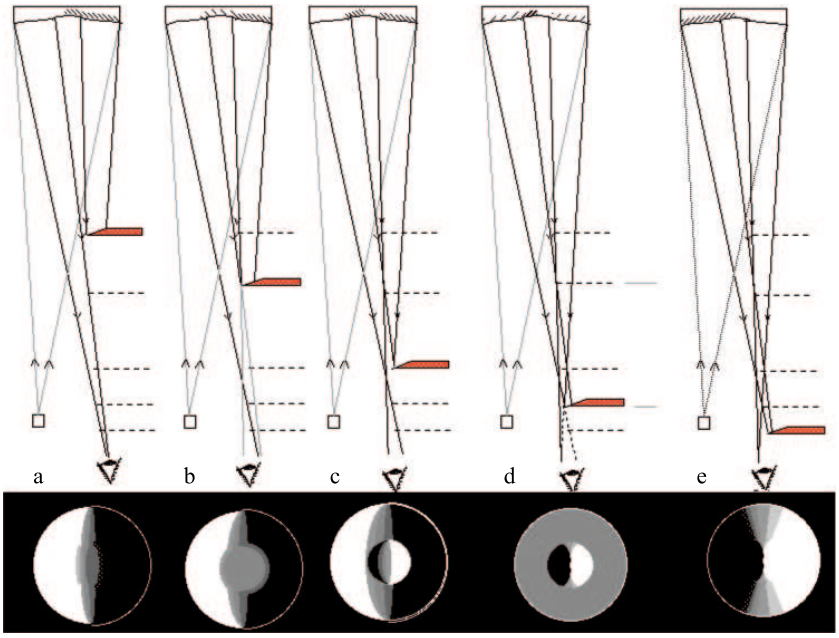


Figure 5.3: The curvature of a spherical mirror with a central hole is measured using the Foucault test. When the knife-edge position defines an osculating reference sphere that corresponds to a particular area of the mirror (the central hole in [b] and the outer ring in [d]) this zone appears flat to the observer. (figure from Koechlin 1990)

More information about the Foucault testing method can be found in Foucault (1859), Koechlin (1990) or Texereau (1961).

Paraboloid testing, the Couder screen

It is now obvious that the Foucault test is dedicated at analyzing spherical surfaces, at least in the form that has been presented before. Indeed, the test of paraboloids presents important difficulties since it would require a source located at infinity. However it is possible to use the Foucault approach to test the quality of a parabolic mirror, even by operating at its center of curvature. Indeed, this type of surface can be considered as a series of rings. Each of them corresponding to a particular sphere, that can be measured with the Foucault test. The test of a paraboloid then simply consists in verifying that these rings present the expected curvature.

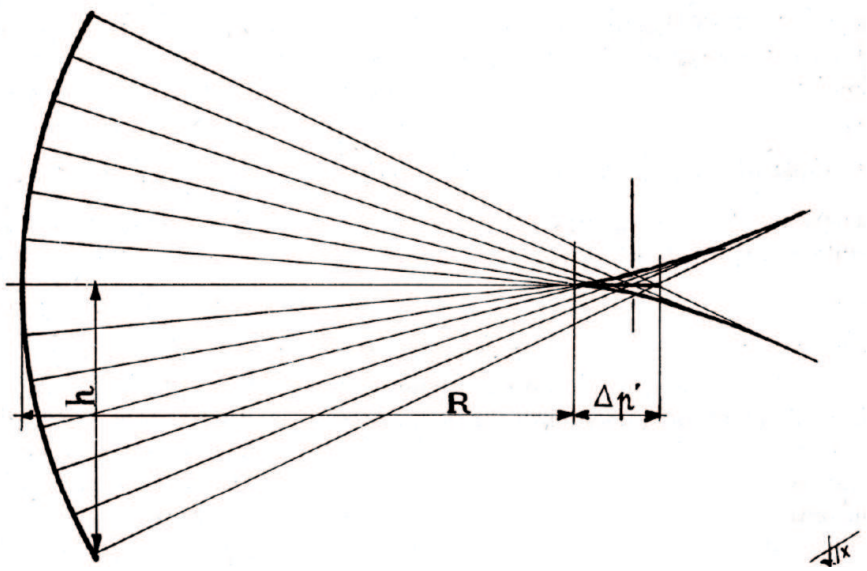


Figure 5.4: Illustration of the spherical aberration. When a light source is located at the center of curvature of the parabolic mirror, the paraxial rays converge before the marginal rays along the optical axis. The segment between the paraxial and the marginal focus is called the longitudinal aberration. This segment is seen as the central peak of the point spread function. After the convergence region, the rays almost merge resulting in a caustic whose effect is a reinforcement of the PSF ring. The vertical line at two thirds of the longitudinal aberration segment designates the circle of least confusion (figure from Texereau 1961).

Applying the classical Foucault test to a parabolic mirror will reveal an important spherical aberration, which corresponds to the difference between the sphere and the paraboloid. However, this aberration can be calculated theoretically for the particular tested paraboloid. It can then be subtracted from the measured spherical aberration to reveal the true defects of the surface.

Figure 5.4 illustrates the aberration of a parabolic mirror around its center of curvature. The segment between the two points where the central and the marginal rays are respectively crossing each other is called the longitudinal aberration. It completely determines the spherical aberration and can be calculated for a given mirror, with the formula

$$\Delta p' = \frac{h^2}{R} \tag{5.1}$$

where h is the radius of the considered zone and R is the radius of curvature of the mirror. Comparison of the computed and measured values of this longitudinal aberration allows to determine if the tested surface corresponds to the expected one.

We now explain how to measure this longitudinal aberration. Figure 5.5 shows the test of a paraboloid with a knife-edge located at several positions corresponding to the beginning, the half and the end of the longitudinal segment. Measuring these positions allows to determine the longitudinal aberration. However, the accurate measurement of positions A and C is not as easy as it could seem. Indeed, the central rays are almost parallel and their intersection is not always clearly visible. Moreover, measurement of position C is disturbed by the luminous ring due to the caustic of light shown in fig. 5.4.

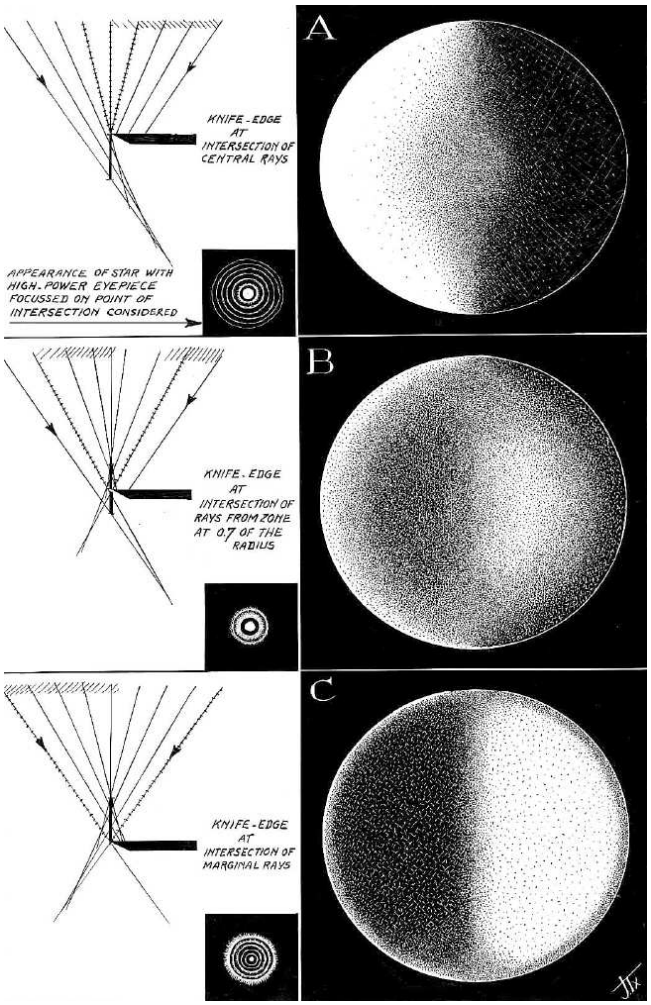


Figure 5.5: Characterization of a paraboloid with the Foucault method. [A] The knife edge is positioned at the paraxial focus, the central zone of the mirror appears flat to the observer. This position, corresponding to the beginning of the longitudinal aberration, is chosen as a reference. The knife is then moved away from the mirror. [C] Once it reaches the marginal focus, the outer ring of the mirror appears flat. This position, the end of the longitudinal aberration segment, is compared to the reference in order to measure the length of the longitudinal aberration. [B] Around the middle of the longitudinal aberration segment, the zone of the mirror located around $\sqrt{2}/2$ should appear flat. (figure from Texereau 1961)

These difficulties can be solved using a "Couder screen" (fig. 5.6) that defines zones on the mirror. The measurement of the longitudinal aberration is carried out by placing the screen in front of the surface, and then by determining the positions of the knife-edge so that each zone appears to be flat. As the mean radii of each zone are known, it is then possible to calculate the corresponding theoretical longitudinal aberration and to compare it with the real one. However, the use of the Couder screen limits the Foucault test to verifying that the shape of the tested optical surface departs from the sphere as a paraboloid would do. This means that the local defects of the mirror will probably not be seen. Moreover, this Couder screen may be difficult to use with large mirrors. More details about the Foucault testing method and Couder screen can be found in Texereau (1961).

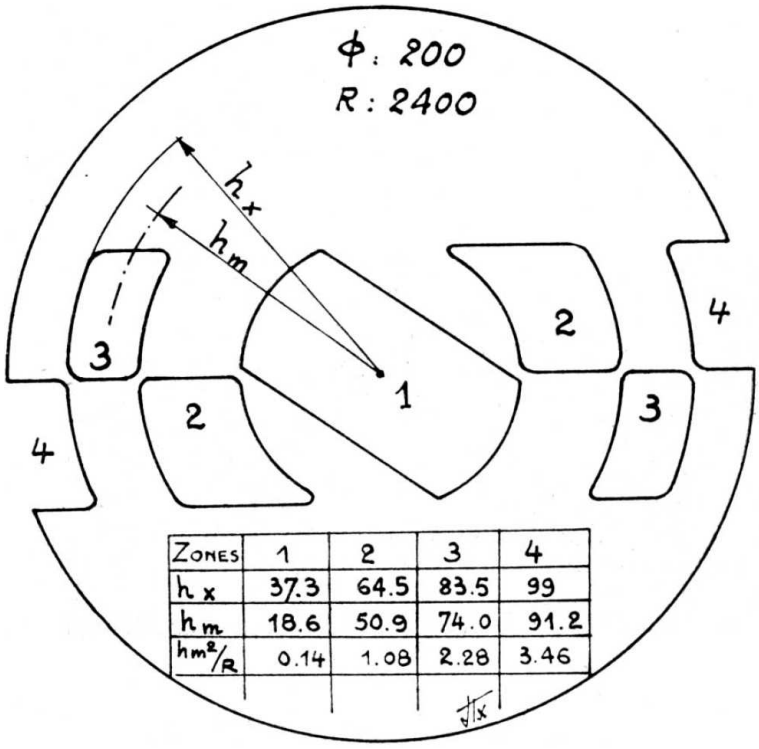


Figure 5.6: Example of a Couder screen used to test a paraboloid of 20 cm in diameter. The longitudinal aberration corresponding to each zone can be calculated analytically and compared to the measurement. This comparison allows to determine the difference between the expected paraboloid and the real mirror (figure from Texereau 1961).

Null test

A null test consists in comparing a reference wavefront with the particular surface to test. It is called "null" when the surface has the expected shape, and the difference between the wavefront created by the tested surface and the reference wavefront is null. Comparison between the wavefront and the surface can be performed with the Foucault method presented before, but other methods, such as interferometry, can also be used. The Foucault test is a null test for a sphere, when performed from its center of curvature, but null test can also be used to test aspheric surface.

In order to perform such a test, it is necessary to create a wavefront that corresponds to the expected shape of the tested surface. Optical systems can be designed to compensate for the spherical aberration corresponding to the difference between a sphere and a paraboloid, they are called null corrector and mainly consist of two lenses. The use of such lenses allows to test a paraboloid as if it were a sphere (this means at the center of curvature). The null lenses can then be used to transform the Foucault test in a null test for a paraboloid. In this case, the Couder screen is not necessary any longer. Null lenses can also be replaced by computer generated holograms, this will be discussed in section 5.1.4. Information about null correctors for paraboloids can be found in Offner (1963).

5.1.3 The Ronchi test

The Ronchi testing method is simple, fast and intuitive. It was first introduced in 1923 by Vasco Ronchi (Ronchi 1927, 1964) and consists in studying the deformation of a projected grating.

Principle

The Ronchi test is based on the following observation: in the case of a perfect optical instrument, the rays in the vicinity of the focus are symmetric and the images are thus also symmetric. However, once the instrument is not perfect anymore, this symmetry is broken and the images can be analyzed.

A Ronchi grating is placed in the beam near the image in intra-focal position (fig. 5.7), it simply consists of a series of opaque and transparent lines of the same width. A typical Ronchi screen has about five lines per millimeter. The observer will then see images of the grating deformed by the aberration affecting the mirror. With some training, it is possible to recognize the aberrations which cause the observed deformations.

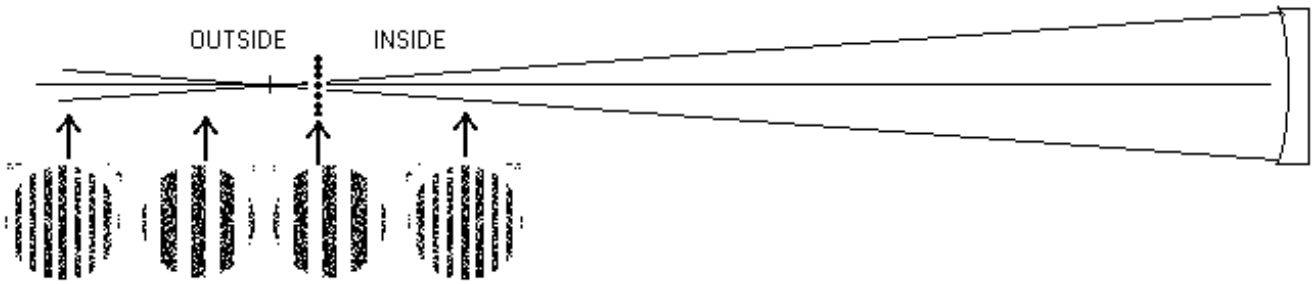


Figure 5.7: Ronchi testing setup, the source and observer are located around the center of curvature of the mirror located on the right. A Ronchi grating is inserted near the focus. The exact position of the grating influences the width of the observed grid. When the mirror is perfect, this grid is perfectly linear. Fig. 5.8 shows patterns corresponding to several types of aberrations.

The testing method consists in observing images of a source, typically a star, centered on the optical axis of the telescope, with a grating inserted near its focus. Comparison between the intra-focal and extra-focal images gives some qualitative information about the main defects of the mirror. It is important to note that the use of a Ronchi grating implies a special direction. A better characterization of the mirror is obtained by rotating the screen during the observation.

The accuracy of the method increases with the number of lines per millimeter. However, an upper limit exists, since the method becomes difficult to use because of diffraction effects occurring when too many lines are used (i.e. the size of the lines becomes to small). Another limitation comes from the fact that the Ronchi test is not performed at the focus of the instrument. The accuracy can be increased by approaching the focus. Finally, the sensitivity increases with the focal length over diameter ratio (F/D). Even, if the accuracy is quite low for an open instrument ($F/D \sim 1$), it can reach $\lambda/5$ for $F/D = 10$.

More sophisticated gratings can be used to generate particular wavefronts. They are called computer generated holograms and are described in the next section.

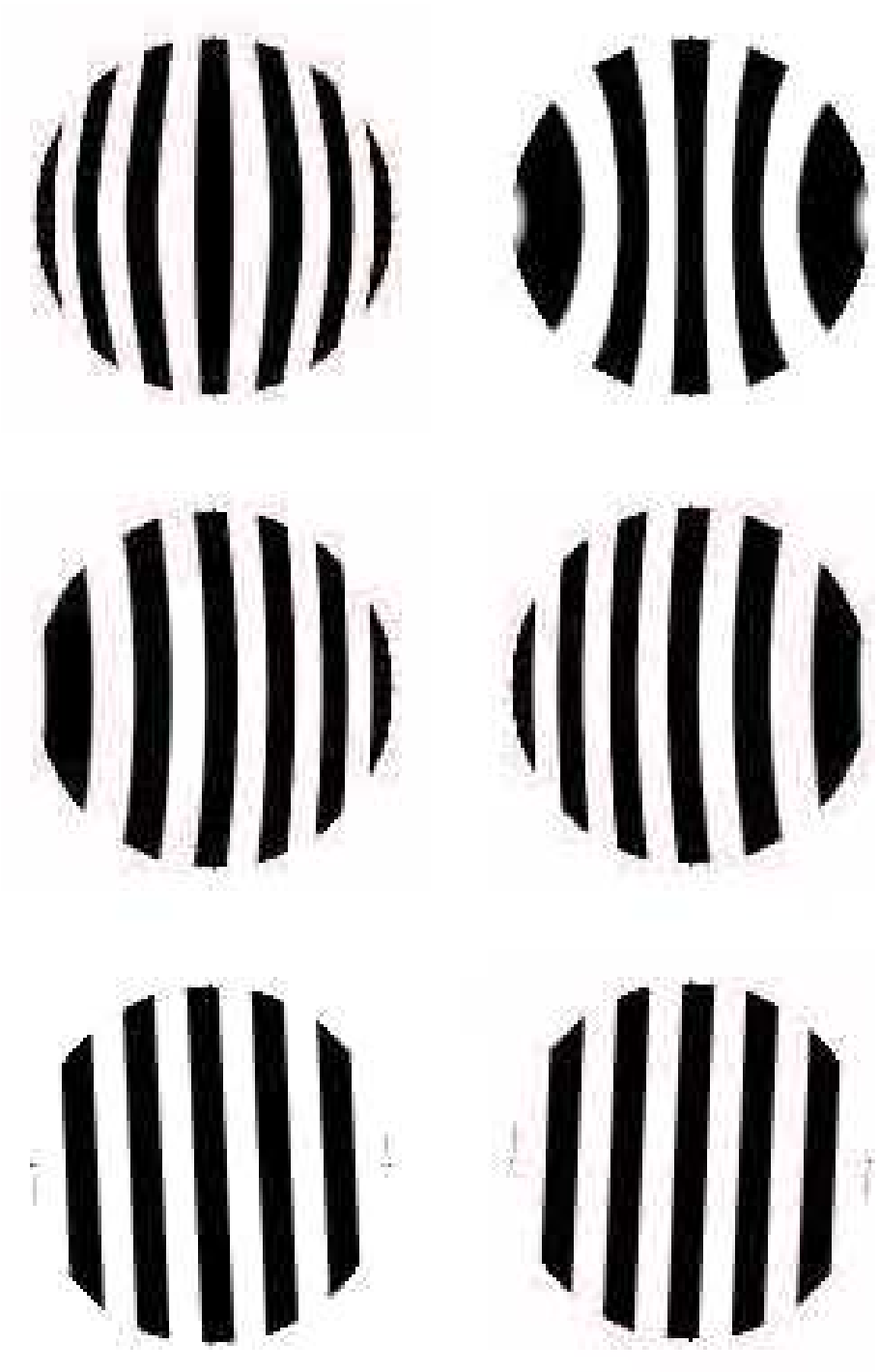


Figure 5.8: Intra and extra-focal Ronchi patterns corresponding to spherical aberration (top row), coma (middle row) and astigmatism (bottom row) (figure from http://www.astrosurf.com/tests/ronchi/ronchi.htm#haut).

5.1.4 Computer Generated Holograms - (CGHs)

Computer generated holograms (CGH) are distorted diffraction gratings made on transparent substrate (mainly glass), designed to produce the desired wavefront when illuminated with a suitable reference beam. The CGH patterns are computed using numerical simulations taking diffraction and propagation of light into account. Both the phase and the amplitude of the incoming wavefront can be modified using CGH, which leads to many interesting applications for this type of holograms. Figure 5.9 shows an example of computer generated hologram.

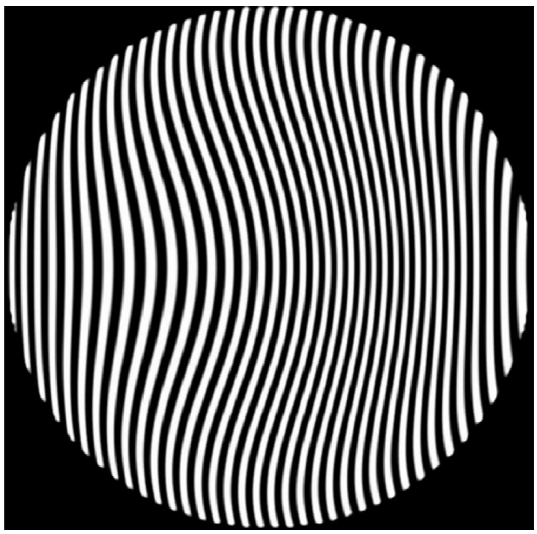


Figure 5.9: An example of Computer Generated Hologram (CGH) that creates a wavefront with spherical aberration (figure from Burge et al. 2007).

This ability of CGH to transform one wavefront into another with the desired shape is very useful for optical testing, especially as far as highly aspheric optical surfaces are concerned. These surfaces are very difficult to analyze because they cannot be tested with classical optical tests. However, the CGH ability to adjust a wavefront so that it corresponds to the surface to test allows to use this wavefront as a reference (fig. 5.10) for a null test.

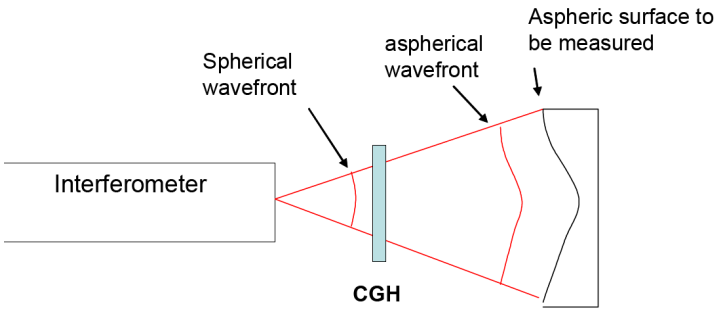


Figure 5.10: Null test of an aspheric surface using a CGH. The grating is computed so that the spherical incoming wavefront is transformed to correspond to the expected aspheric surface. The aspheric surface can then be compared to that wavefront using an interferometer for example (figure from Burge et al. 2007).

CGHs can also be used for alignment. Indeed, it is possible to design CGHs that generate sharp crosses, that can be used as precise reference for positioning optical elements (needed to test an unknown surface for example). Moreover, it is possible to multiplex several CGH patterns on the same substrate to create multiple wavefronts simultaneously out of one single reference beam. This allows, for example, to use a given CGH for both testing the surface and aligning it with the CGH at the same time. Figure 5.11 shows an example of multiplexed CGH which was used to test the New Solar Telescope (NST) at the Big Bear Solar Observatory. The central zone is used to test an aspheric surface whereas the outer patterns are used for alignment purpose. The four patterns in the small circles project reference spots to align subsequent optics, the ring pattern is used to align the CGH substrate and the four rectangular patterns project, a reference for the position of the tested surface.

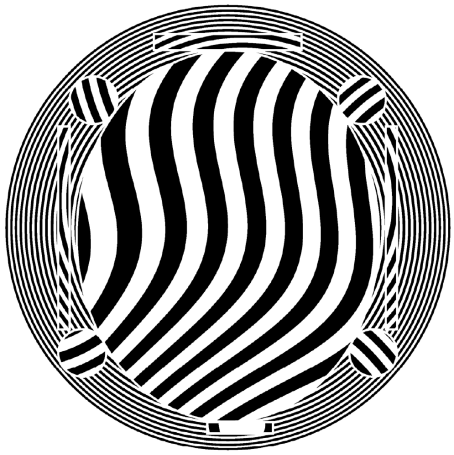


Figure 5.11: Example of multiplexed CGH. The central pattern creates the aspheric wavefront for surface testing. The outer regions are used to align the elements for the optical test (figure from Burge et al. 2007).

Current electron beam techniques can create 100mm wide CGH patterns with 0.1μm accuracy. Such holograms, can encode a lot of information. More information about CGHs and their application as alignment tool can be found in Burge et al. (2007) and as testing tool in Zhao et al. (2005).

5.1.5 The Hartmann test

At the beginning of the 20th century, Johannes Hartmann, working on the "Great Refractor" in Potsdam, developed his now famous screen test (Hartmann 1900). Information about this method can be found in Platt and Shack (2001).

Principle

The principle (fig. 5.12) of this test is to isolate rays of light in order to perform a sort of "ray tracing" analysis. A screen perforated with holes, called a Hartmann plate, is placed in front of the mirror to divide the reflected beam into rays. These isolated rays are then collected on two photographic plates successively located at two intra-focal and extra-focal positions. The points corresponding to the same hole are gathered in pairs. A simple linear interpolation between each pair of points enable the calculation of the coordinates of the piercing points in the focal planes of the corresponding rays. The result is equivalent to a spot diagram analysis of the focal region of the tested surface.

The Hartmann constant is defined as the average distance between the piercing points of each ray in the focal plane (intersection between the rays and the focal plane) and the optical axis. When the optical system is perfect, all rays cross the optical axis in the focal plane (the piercing points are all located on the axis) and the Hartmann constant is zero. In the other cases, the piercing points are not on the axis and the constant is larger than zero. The geometrical-optics aberrations are related to the coordinates of these piercing points, and the Nijboer relation

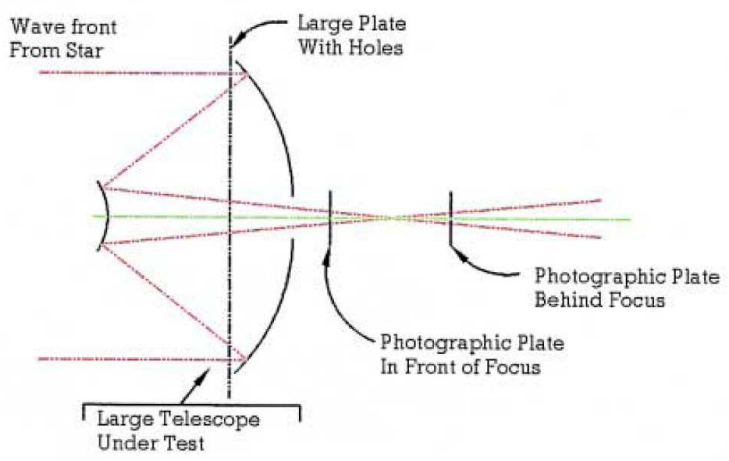


Figure 5.12: Hartmann test setup. A screen with regularly perforated holes is placed in front of the mirror, allowing only a few rays of light to reach the observer. Two photographic plates are successively disposed before and after the focus in order to locate the rays in these planes (figure from Platt and Shack 2001).

establishes a link between the geometrical and wavefront aberrations

$$\delta x = R\frac{\partial W}{\partial x}, \qquad \delta y = R\frac{\partial W}{\partial y} \tag{5.2}$$

where $\delta x, \delta y$ are the coordinates of a ray in the focal plane, R is the length between the exit pupil and the focal plane and W is the wavefront equation. Knowing this equation, it is possible to decompose the wavefront surface into Zernike polynomials, simply by projecting the surface on each aberration.

The Shack-Hartmann sensor

With the need to perform wavefront testing with very low illumination, Shack improved the Hartmann method by replacing the screen with an array of small lenses and the photographic plates with a CCD camera. In fact, the pupil is re-imaged on the lenslet array. Each of these lenses becomes a sub-aperture of the system and creates its proper spot on the CCD. If the wavefront entering the array is planar, the image of each lens is a spot located on its optical axis. On the other hand, when the wavefront is aberrated, the slope of the local wavefront corresponding to each particular lenslet causes the image to be out of the axis of its lens. Measuring the position of the spot thus indicates the slope of the wavefront which can be reconstructed. There must be at least 4 pixels dedicated to each spot in order to have a sufficient sampling.

A common mathematical theory for the Hartmann and Ronchi tests is developed in Cordero-Davila et al. (1992) and the way of retrieving the wavefront from the Hartmann testing is explained in Salas-Peimbert et al. (2005). A null test for aspherical surfaces based on the Hartmann method is presented in Malacara (1972).

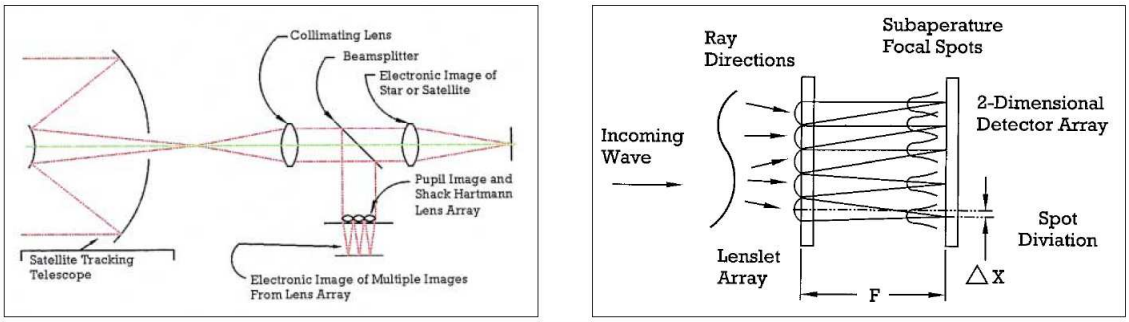


Figure 5.13: Shack-Hartmann wavefront sensor principle. An array of lenslets divides the incoming wavefront in small local wavefronts. If the incoming front is planar, each local wavefront will be focused on its lens axis. On the other hand, if the incoming front is distorted, the lenses will focus the local front somewhere else. The position of the spots is related to the slopes of the corresponding local wavefront (figure from Platt and Shack 2001).

5.1.6 The Roddier test

The Roddier test appeared around the early nineties with the search for a new type of wavefront sensor for adaptive optics (Roddier and Roddier 1989) based on curvature sensing instead of slope sensing. This approach revealed to be usable to test telescope optics (Roddier et al. 1990).

Principle

As in the case of the Ronchi testing method, the Roddier test is based on the symmetric measurement of the intensity before (intra-focal) and after (extra-focal) the focus of the optical system.

Let us first consider the case of a perfect instrument (fig. 5.14). If the input pupil is uniformly illuminated (intensity I_0), the light passing through a small area of the pupil illuminates the intra and the extra-focal image on the same area with the same intensity because of the symmetry around the focal point. The intensity is thus concentrated in the same way on both images and we get:

$$I_1 = I_2. \tag{5.3}$$

If the instrument is not perfect (see fig. 5.15), the wavefront in the small part of the pupil is curved and the focusing is not perfect. Indeed, for this part of the pupil the focus is slightly in front of the pupil focal point (in our example). The intensity in the intra-focal image (I_1) is thus different from the one in the extra-focal image (I_2).

Considering the full intra and extra focal images, the difference between their intensity is related to the curvature of the wavefront by the following equation.

$$\frac{I_1 - I_2}{I_1 + I_2} = \frac{\lambda L}{2\pi}\left(\partial\frac{\phi(r)}{\partial r}\cdot\delta_c - \nabla^2\phi(r)\right), \tag{5.4}$$

where $\phi(r)$ is the wavefront expression, L is the distance between the focal plane and the intra/extra focal images and δ_c is the edge function (it defines the edge of the aperture). Comparison

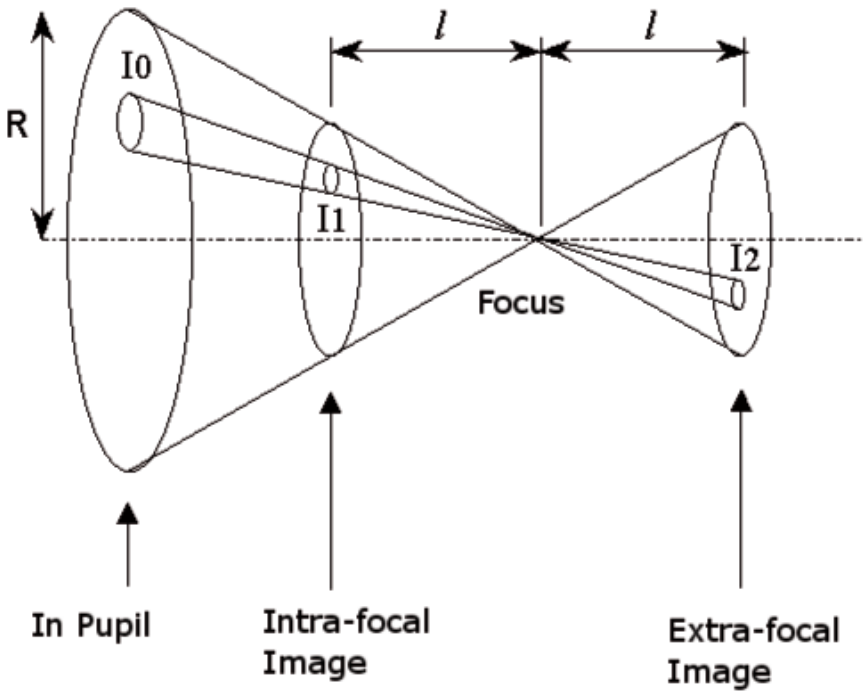


Figure 5.14: Principle of the Roddier testing method. The surface to characterize is uniformly illuminated and the intensity of light is measured symmetrically before and after the focus. If the incoming wavefront is aberration free, the intensity corresponding to a particular region of the input pupil I_0 collected on both positions will be the same $I_1 = I_2$. (Figure from www.astrosurf.com)

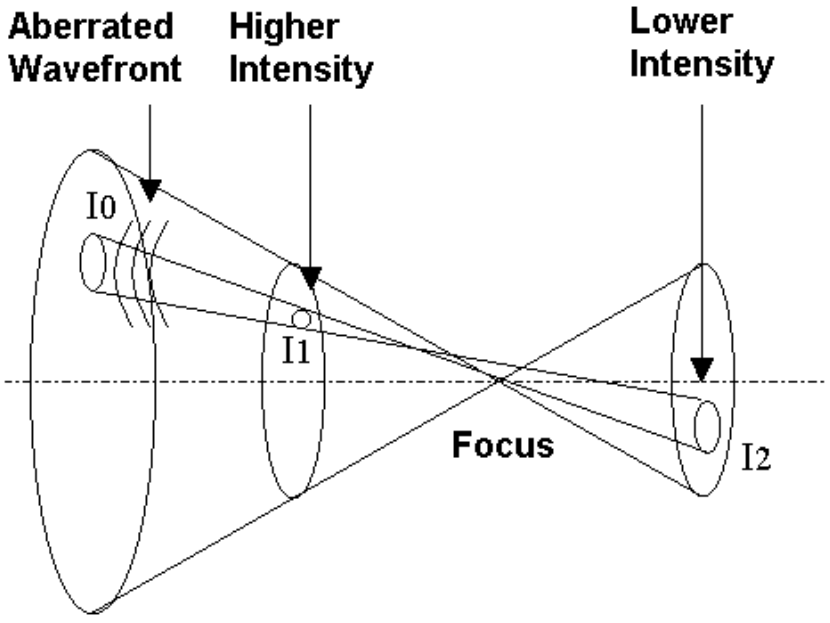


Figure 5.15: In the case where the incoming wavefront is aberrated, the convergence of the rays does not happen right at the focal point. This introduces an asymmetry in the intensity measured before and after the focus $I_1 \neq I_2$. (image from www.astrosurf.com)

of the intra and extra-focal intensities thus gives access to the wavefront aberrations, which can be retrieved from its Laplacian.

The Roddier testing method was used to detect the defects of the Hubble Space Telescope optics (Roddier and Roddier (1991), Roddier and Roddier (1993)).

Example: the spherical aberration

The principle of the Roddier test is illustrated for an instrument presenting spherical aberration (fig. 5.16). The calculations involved in this example were performed with a software called "winroddier" (available on the website: http://www.astrosurf.com/tests/roddier/projet.html).

In this case, the light coming from the regions around the optical axis converges before the focus whereas the light coming from the outside of the pupil converges after the focal point. The intra-focal image thus shows an excess of intensity in its central region whereas the extra-focal image presents an excess of intensity in the outer region, as illustrated in the two first diagrams of fig. 5.16.

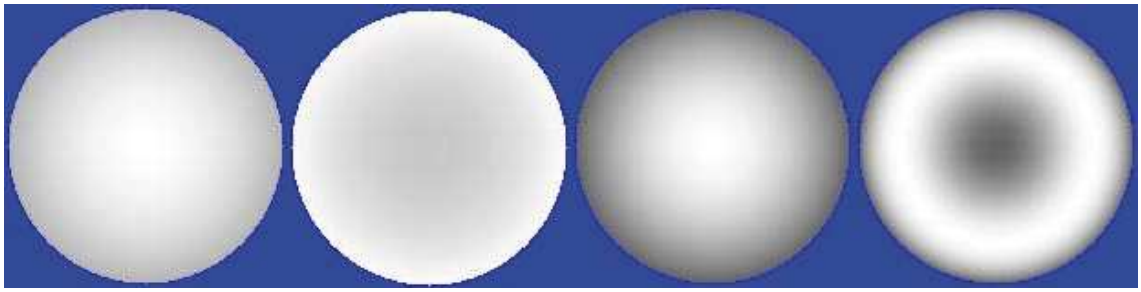


Figure 5.16: Example of computation of the wavefront using Roddier's method. From left to right: intra-focal and extra-focal intensities, laplacian of the wavefront and wavefront.

Relation between the Roddier and the Nijboer-Zernike theories

The defocus range used in the Roddier method is quite large, the intra and extra focal planes where the measurements are performed are thus pupil planes. Another approach consists in applying a very small defocusing so that the observations are made in the image plane allowing to study smaller aberrations. This is the Nijboer Zernike (NZ) approach that will be detailed in section 5.2.

5.1.7 Summary

Several testing methods have been presented in the previous sections, they are summarized in the table below. The next section will present another aberration measurement approach based on the Nijboer-Zernike theory. The expected accuracy of this method is included in the table as well as the one of a typical Zygo interferometer for comparison.

Method	Accuracy	Resolution
Foucault	$\lambda/60$ rms	5cm over 8m
Ronchi	$\lambda/5$ rms	10cm over 8m
Shack-Hartmann	$\lambda/100$ rms (UV $\lambda/500$ rms)	2cm over 8m
Roddier	$\lambda/100$ rms	2cm over 8m
Nijboer-Zernike	$\lambda/1000$ rms (expected)	2cm over 8m
Zygo	$\lambda/1000$ rms	1mm over 20 cm

Table 5.1: Summary of the testing methods presented in the previous sections with their accuracy and resolution. The Zygo is shown for comparison. The value given for the NZ method corresponds to 231 Zernike coefficients.

5.2 The Nijboer-Zernike (NZ) approach

The Nijboer-Zernike theory began with the work of Nijboer (1942) on the diffraction theory of aberration. He first introduced the relation between the diffraction equation and the optical aberrations expressed in the form of Zernike polynomials. However, the complexity of the equations forced him to limit his theory to small aberrations (well below one wavelength). Indeed, numerical computation was not possible at that time and the analytical developments required some simplifications.

At the beginning of the twenty first century, Janssen (2002) extended this theory. The problems encountered by Nijboer were circumvented by using an explicit Bessel series representation for the diffraction integral. He also proposed a convenient way to numerically compute the expressions involving these Bessel series (the V_n^m functions). This extended Nijboer-Zernike theory allows to quickly compute the intensity PSF of any system provided that its aberrations are known. This has been presented in chapter 4, section 4.3.2.

One year later, Dirksen et al. (2003) introduced the reverse approach that allows to perform aberration retrieval based on the classical Zernike expansion. They also proposed to apply this approach to the general complex Zernike coefficients. An interesting overview of the whole theory can be found in van der Avoort et al. (2005).

Janssen and his team used the NZ theory mainly for photolithography purposes. They made extensive developments for highly opened systems with a few, small, low order aberrations. For symmetry reasons, in their cases of interest, i.e. testing the lenses used for photolithography, they simplified the equation to only the cosines terms. Moreover, they used many defocused (up to fifteen) images to perform their retrieval.

For astronomical applications, some of their considerations cannot be applied anymore. Indeed, we can neither count on symmetry nor on aberrations limited to low orders when measuring complex optical instruments. We have thus generalized their developments to non symmetrical systems (using sine and cosine polynomials) with high orders (up to 231 Zernike polynomials) and larger (up to one wavelength with a Strehl ratio as low as 40%) aberrations. The aberrations are computed from only three images with different foci; one at the focal point, another before and a last one after the focus. Using less defocused images is much more convenient for astronomical applications where it is not always easy to get many images in the same conditions.

In the following section, we will first present an intuitive approach of the retrieval theory.

This one is based on the classical Zernike expansion with small aberrations in a symmetrical case. A general development, based on the general complex Zernike coefficients will follow in section 5.2.2 where very few hypotheses are introduced. The results of both approaches allow to compute the aberrations of an optical system that is characterized by a few PSFs (corresponding to different focal positions).

The approach we present hereafter consists in using the expression of the complex amplitude obtained in the previous chapter (equations 4.69 or 4.52) to derive an expression of the intensity PSF. Then a radial modal analysis of this equation is performed in order to decompose the PSF into its radial modes. These modes are compared to decoupled template radial modes. This step is equivalent to projecting the PSFs on the template radial modes basis, which leads to the formations of decoupled systems of linear equations. The resolution of these systems gives the aberration coefficients that we are looking for. The details of the demonstrations presented in these sections can be found in Appendix B, the principle of the retrieval is illustrated in fig. 5.17.

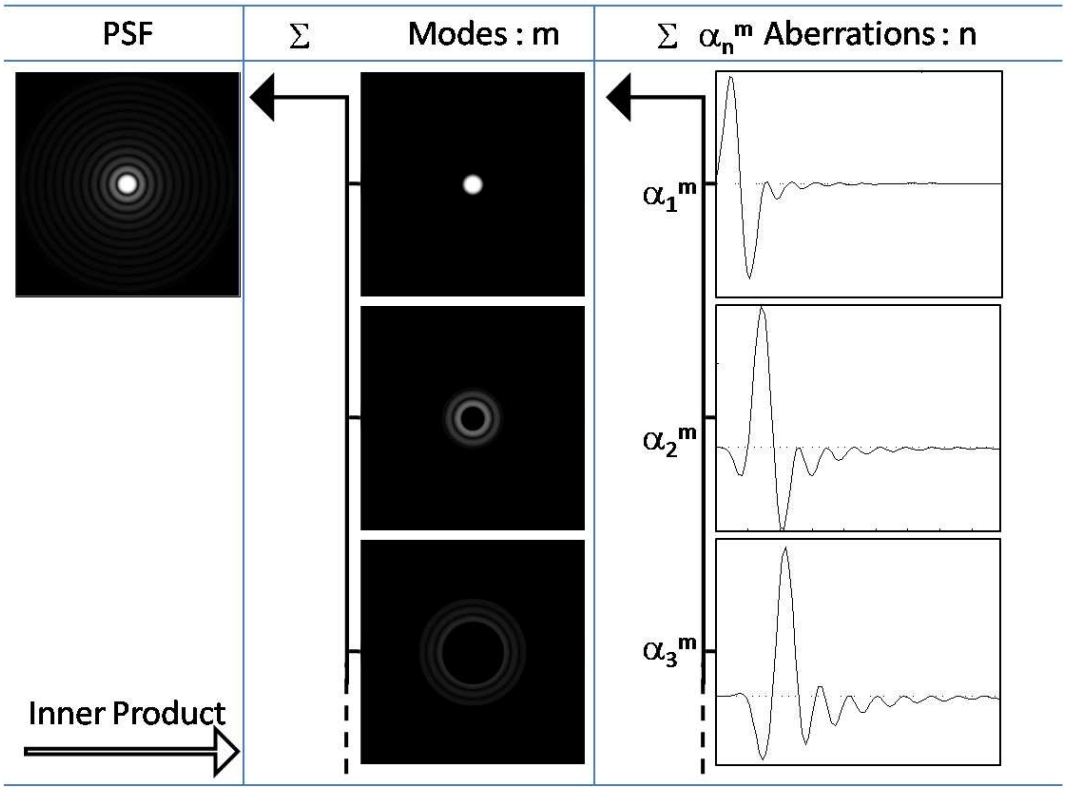


Figure 5.17: Illustration of the Nijboer-Zernike principle. The point spread function is composed of several modes illustrated by the different rings. Those modes result in the linear combination of aberration templates, where the combination coefficients correspond to the Zernike aberration coefficients. It is thus possible to compute the PSF from the Zernike aberration coefficients. On the other hand, the PSF can be projected onto modes using the inner product. These projections are related to the aberration templates by a simple linear system of equations. The diagonalization of this system thus allows to compute the aberration coefficients from the intensity PSF.

5.2.1 Intuitive approach using classical Zernike expansion (α_n^m)

As a first intuitive approach, let us consider the case where the aberrations are expressed in terms of the classical Zernike coefficients (α). We also drop the sine terms that can be treated in the

same way as the cosine terms. They will be included in the general development of section 5.2.2.

Let us first recall that in the case of small aberrations and constant amplitude, the PSF expression is given by (equation 4.69)

$$U(r,\phi,f) = 2V_{00}(r,f) + 2i\sum_{n,m}\alpha_n^m i^m \cos(m\phi)V_n^m(r,f) \tag{5.5}$$

where the sine terms have been neglected for the sake of simplicity. The intensity is thus given in first order approximation by,

$$I = U(r,\phi,f)\cdot U^*(r,\phi,f) \approx 4\left|V_0^0\right|^2 + 8\sum_{n,m}\alpha_n^m Re\left\{i^{m+1}V_0^0 * V_n^m\right\}\cos(m\phi) \tag{5.6}$$

where the quadratic term in $(\alpha_{n1}^{m1}\alpha_{n2}^{m2})$ has been omitted. Let us define the cosine modulated azimuthal mean of the measured intensity, $\Psi_{\text{meas}}^m(r,f)$, as

$$\Psi_{\text{meas}}^m(r,f) = \frac{1}{2\pi}\int_0^{2\pi} I_{\text{meas}}(r,\phi,f)\cos(m\phi)d\phi \tag{5.7}$$

and the normalized template aberrations

$$\Psi_n^m(r,f) = -8\varepsilon_m^{-1}\Im[i^m V_n^m(r,f)V_0^{0*}(r,f)] \tag{5.8}$$

where $\varepsilon_0 = 1, \varepsilon_1 = \varepsilon_2 = \varepsilon_3 = ... = 2$. Ψ_{meas}^m can be obtained directly from the real part of the Fourier transform of the PSF. Using these definitions, we can rewrite equation 5.6 in the form

$$\Psi_{\text{meas}}^m(r,f) \approx 4\delta_{m0}\left|V_0^0(r,f)\right|^2 + \sum_n \alpha_n^m\Psi_n^m(r,f) \tag{5.9}$$

where δ_{m0} is a Dirac delta ($\delta_{m0} = 1$ when $m = 0$ and $\delta_{m0} = 0$ when $m \neq 0$). We introduce the definition of an inner product for the functions $\Psi(r,f)$ and $\chi(r,f)$ as

$$(\Psi,\chi) = \int_0^\infty\int_{-\infty}^\infty \Psi(r,f)\chi^*(r,f)rdrdf. \tag{5.10}$$

This product acts as a projection of the Ψ function on the χ basis. Before going any further it is important to note some interesting properties of this product. It comes out of the definition of the V_n^m functions (equation 4.51) that

$$V_n^m(r,-f) = V_n^{m*}(r,f) \tag{5.11}$$

for all values of n, m. This implies that $|V_0^0|^2$ is even in f while $\Psi_{n'}^0$ is odd in f for all n'. It can then be deduced that the inner product of these two functions is null

$$\left(|V_0^0|^2, \Psi_{n'}^0\right) = 0,\ \forall n' \tag{5.12}$$

Taking the inner product of expression 5.9 with $\Psi_{n'}^m$, we get,

$$\sum_n \alpha_n^m(\Psi_n^m, \Psi_{n'}^m) \approx (\Psi_{\text{meas}}^m, \Psi_{n'}^m) \tag{5.13}$$

where the left and right hand sides have been switched. The α_n^m coefficients can be computed as follows, we first define a Gram matrix, G^m

$$G^m = ((\Psi_n^m, \Psi_{n'}^m))_{n',n=m,m+2,...} \tag{5.14}$$

and a right-hand side vector $\mathbf{r}^m$

$$\mathbf{r}^m = ((\Psi_{\text{meas}}^m, \Psi_{n'}^m))_{n',n=m,m+2,...}. \tag{5.15}$$

Replacing these expressions in equation 5.13, we get

$$\hat{\alpha}^m = (G^m)^{-1} r^m, \tag{5.16}$$

where $\hat{\alpha}^m$ is the vector of the estimated[22] α_n^m coefficients defined as $\hat{\alpha}^m = (\alpha_n^m)_{n=m,m+2,...}$ and where $(G^m)^{-1}$ is the inverse of the Gram matrix G^m. It is now obvious that the aberration retrieval based on the Nijboer Zernike approach simply reduces itself to the diagonalization of a linear system of equations.

5.2.2 NZ retrieval theory based on the general Zernike coefficients (β_n^m)

In the general case, the pupil function is given by

$$P(\rho,\theta) = A(\rho,\theta)e^{i\Phi(\rho,\theta)} = \sum_{n,m} \beta_n^m Z_n^m(\rho,\theta) = \sum_{n,m} \beta_{cn}^m Z_{cn}^m(\rho,\theta) + \sum_{n,m} \beta_{sn}^m Z_{sn}^m(\rho,\theta). \tag{5.17}$$

where we have defined β_{cn}^m and β_{sn}^m as the β_n^m coefficients corresponding respectively to the "cos" and "sin" Zernike polynomials, themselves written as Z_{cn}^m and Z_{sn}^m. Let us remind that the polynomials related to $m = 0$ are included in the cosine case as their behavior is well represented with the cosine function ($\cos 0 = 1$).

The equation of the PSF is (equation 4.52)

$$U(r,\phi,f) = \sum_{n,m} \beta_n^m U_n^m(r,\phi,f) = \sum_{n,m} \beta_{cn}^m U_{cn}^m(r,\phi,f) + \sum_{n,m} \beta_{sn}^m U_{sn}^m(r,\phi,f) \tag{5.18}$$

with

$$\begin{aligned} U_{cn}^m(r,\phi,f) &= 2i^m\, V_n^m(r,f)\, \cos(m\phi) \\ U_{sn}^m(r,\phi,f) &= 2i^m\, V_n^m(r,f)\, \sin(m\phi) \end{aligned} \tag{5.19}$$

and,

$$V_n^m(r,f) = (-1)^m \int_0^1 \rho\, e^{if\rho^2} R_n^m(\rho) J_m(2\pi r\rho)\, d\rho \tag{5.20}$$

[22]The equations used for the retrieval are approximated and the results obtained are estimations of the true values.

In the following, we will consider β_0^0 as real and positive. Assuming that β_0^0 is real corresponds to neglecting the piston phase aberration. In all practical cases, the real part of β_0^0 is positive as it is an amplitude term representing the transmission of the optical system ($0 < \beta_0^0 < 1$). In an optical system, the amplitude aberration is very weak and the β_0^0 term is close to 1.

Moreover, we can assume that β_0^0 is larger than the other coefficients β_n^m. In this case it can be extracted from the sum and the PSF equation becomes

$$\begin{aligned} U(r,\phi,f) = 2\beta_0^0 V_0^0(r,f) \qquad & +2{\sum_{n,m}}' \beta_{cn}^m i^m V_n^m(r,f)\ \cos(m\phi) \\ & +2{\sum_{n,m}}' \beta_{sn}^m i^m V_n^m(r,f)\ \sin(m\phi) \end{aligned} \tag{5.21}$$

where the "prime" sign on the sums indicates that the term with $n = m = 0$ has been removed from these sums. The intensity of the PSF is given by

$$\begin{aligned} I \quad &= |U|^2 = 4(\beta_0^0)^2 |V_0^0|^2 \\ &+8{\sum_{n,m}}' \beta_0^0 \Re(\beta_{cn}^m)\Re[i^m V_n^m V_0^{0*}]\cos(m\phi) - 8{\sum_{n,m}}' \beta_0^0 \Im(\beta_{cn}^m)\Im[i^m V_n^m V_0^{0*}]\cos(m\phi) \\ &+8{\sum_{n,m}}' \beta_0^0 \Re(\beta_{sn}^m)\Re[i^m V_n^m V_0^{0*}]\sin(m\phi) - 8{\sum_{n,m}}' \beta_0^0 \Im(\beta_{sn}^m)\Im[i^m V_n^m V_0^{0*}]\sin(m\phi) \\ &+O((\beta_n^m)^2) \end{aligned} \tag{5.22}$$

where the quadratic term in β_{cn}^m, β_{sn}^m, is defined as,

$$\begin{aligned} O((\beta_n^m)^2) \quad &= 4\left|{\sum_{n,m}}' \beta_{cn}^m i^m V_n^m \cos(m\phi)\right|^2 + 4\left|{\sum_{n,m}}' \beta_{sn}^m i^m V_n^m \sin(m\phi)\right|^2 \\ &+8\, Re\left({\sum_{n,m}}' \beta_{cn}^m i^m V_n^m \cos(m\phi)\right) \cdot Re\left({\sum_{n,m}}' \beta_{sn}^m i^m V_n^m \sin(m\phi)\right) \\ &+8\, Im\left({\sum_{n,m}}' \beta_{cn}^m i^m V_n^m \cos(m\phi)\right) \cdot Im\left({\sum_{n,m}}' \beta_{sn}^m i^m V_n^m \sin(m\phi)\right) \end{aligned} \tag{5.23}$$

As β_0^0 is considered to be larger than the other β_n^m, the quadratic terms in $\beta_{n1}^{m1} \cdot \beta_{n2}^{m2}$, with $n1, m1, n2, m2 \neq 0$ will be neglected in the following demonstration. These terms will be discussed later. The complete demonstration of equation 5.22 can be found in Appendix B. As in the intuitive approach, we introduce the modulated azimuthal means, $\Psi_{\text{meas}}^m(r,f)$, the phase aberration templates Ψ_n^m and the amplitude template aberrations, χ_n^m

$$\Psi_{\text{meas}}^m(r,f) = \frac{1}{2\pi}\int_0^{2\pi} I_{\text{meas}}(r,\phi,f) e^{im\phi} d\phi \tag{5.24}$$

$$\Psi_n^m(r,f) = -8\,\varepsilon_m^{-1}\,\Im[i^m V_n^m(r,f)V_0^{0*}(r,f)] \tag{5.25}$$

$$\chi_n^m(r,f) = 8\,\varepsilon_m^{-1}\,\Re[i^m V_n^m(r,f)V_0^{0*}(r,f)] \tag{5.26}$$

Replacing these expressions in the intensity (equation 5.22) where the quadratic term has been omitted (details will be found in Appendix B), we find

$$\Psi^0_{\text{meas}} \approx \frac{1}{2}(\beta_0^0)^2\chi_0^0 + \sum_n{}' \beta_0^0\,\Re(\beta_{cn}^0)\,\chi_n^0 + \sum_n{}' \beta_0^0\,\Im(\beta_{cn}^0)\,\Psi_n^0 \tag{5.27}$$

$$\Psi^m_{\text{meas}} \approx \left\{\sum_n \beta_0^0\,\Re(\beta_{cn}^m)\,\chi_n^m + \sum_n \beta_0^0\,\Im(\beta_{cn}^m)\,\Psi_n^m\right\} + i\left\{\sum_n \beta_0^0\,\Re(\beta_{sn}^m)\,\chi_n^m + \sum_n \beta_0^0\,\Im(\beta_{sn}^m)\,\Psi_n^m\right\} \tag{5.28}$$

The last equation can be split into $\Psi^m_{_c\text{meas}}$ and $\Psi^m_{_s\text{meas}}$ corresponding respectively to the real and imaginary part of Ψ^m_{meas}, which corresponds to the cosine and sine part of the imaginary exponential.

$$\Psi^m_{\text{meas}} = \Re(\Psi^n_{\text{meas}}) + i\Im(\Psi^n_{\text{meas}}) = \Psi^n_{c\text{meas}} + i\Psi^n_{s\text{meas}} \tag{5.29}$$

$$\Psi^m_{_c\text{meas}} \approx \sum_n \beta_0^0\,\Re(\beta_{cn}^m)\,\chi_n^m + \sum_n \beta_0^0\,\Im(\beta_{cn}^m)\,\Psi_n^m \tag{5.30}$$

$$\Psi^m_{_s\text{meas}} \approx \sum_n \beta_0^0\,\Re(\beta_{sn}^m)\,\chi_n^m + \sum_n \beta_0^0\,\Im(\beta_{sn}^m)\,\Psi_n^m \tag{5.31}$$

Now we should note that the inner product between χ_n^m and Ψ_n^m is null

$$(\chi_n^m, \Psi_{n'}^m) = 0, \qquad \forall n, n' = m, m+2, ... \tag{5.32}$$

Indeed, using equation 5.11, it can be demonstrated that the real part of the V_n^m functions is even in f whereas their imaginary part is odd. It can be deduced that, depending on the value of m, χ_n^m is even when Ψ_n^m is odd and inversely. The use of this base of functions (χ_n^m, Ψ_n^m) thus leads to linear algebra. This also explains the choice to use the general Zernike expansion $\sum \beta_n^m Z_n^m$ instead of the classical expansion $e^{i\sum \alpha_n^m Z_n^m}$, where an approximation was required to linearize the exponential.

Using the inner product, it is now possible to build linear systems for $m = 0, 1, ...$, as in the intuitive approach presented earlier. Multiplying equations 5.27, 5.30, 5.31, with $\chi_{n'}^m$ and $\Psi_{n'}^m$ gives the following decoupled systems. Here, we consider the case where $m = 0$ separately because the corresponding Zernike polynomials are not associated with neither sine nor cosine functions (piston, defocus, spherical aberration, ...). We get,

$$\begin{cases} \dfrac{1}{2}(\beta_0^0)^2(\chi_0^0, \chi_{n'}^0) + \displaystyle\sum_n{}' \beta_0^0\,\Re(\beta_{cn}^0)\,(\chi_n^0, \chi_{n'}^0) \approx (\Psi^0_{\text{meas}}, \chi_{n'}^0) \\ \displaystyle\sum_n{}' \beta_0^0\,\Im(\beta_{cn}^0)\,(\Psi_n^0, \Psi_{n'}^0) \approx (\Psi^0_{\text{meas}}, \Psi_{n'}^0) \end{cases} \tag{5.33}$$

for $m = 0$ where $n, n' = 0, 2, ...$ For the real part of $Re(\Psi^m_{\text{meas}}) = \Psi^m_{c\text{meas}}$ and $m \neq 0$, we get

$$\begin{cases} \sum_n \beta_0^0 \, \Re(\beta^m_{cn}) \, (\chi^m_n, \chi^m_{n'}) \approx (\Psi^m_{c\text{meas}}, \chi^m_{n'}) \\ \sum_n \beta_0^0 \, \Im(\beta^m_{cn}) \, (\Psi^m_n, \Psi^m_{n'}) \approx (\Psi^m_{c\text{meas}}, \Psi^m_{n'}) \end{cases} \tag{5.34}$$

where $n, n' = m, m+2, ...$ For the imaginary part of $Im(\Psi^m_{\text{meas}}) = \Psi^m_{s\text{meas}}$ and $m \neq 0$ we get,

$$\begin{cases} \sum_n \beta_0^0 \, \Re(\beta^m_{sn}) \, (\chi^m_n, \chi^m_{n'}) \approx (\Psi^m_{s\text{meas}}, \chi^m_{n'}) \\ \sum_n \beta_0^0 \, \Im(\beta^m_{sn}) \, (\Psi^m_n, \Psi^m_{n'}) \approx (\Psi^m_{s\text{meas}}, \Psi^m_{n'}) \end{cases} \tag{5.35}$$

where $n, n' = m, m+2, ...,$.

The terms in $m = 0$, β^m_{cn} and β^m_{sn} get decoupled. In order to compute the β^m_n, we proceed as follows. We first solve the first equation of the first system 5.33 involving $(\beta_0^0)^2$ and $\beta_0^0 Re(\beta_n^0)$. This gives β_0^0 and then $Re(\beta_n^0)$. Once β_0^0 is known, it is possible to solve the second equation of this first system and get $Im(\beta_n^0)$. It is also possible to solve the two equations of the second and third systems in the same way as for the intuitive approach, by inverting the Gram matrix. In the case of purely imaginary β^m_n, the retrieval corresponds to the pure-phase case (α^m_n).

5.2.3 Predictor-Corrector approach

The calculation presented in the previous section allows to retrieve the β^m_n coefficient with a precision of a few percent or worst depending on the case. This lack of accuracy is due to the omission of the quadratic terms. Hereafter, we introduce an improvement which allows to account for these terms. The corrections brought by this approach are small, as the errors are.

A coupling between phase and amplitude appears due to the crossed terms in "$\sin \cdot \cos$". The error introduced by these coupled terms are smaller than the error related to the quadratic phase terms as far as classical imagery is concerned. Regarding coronagraphy, where first order terms are attenuated, the quadratic phase terms and amplitude-phase coupled terms are of the same order of magnitude.

The phase effect due to the omission of the quadratic terms in equation 5.23 can be almost corrected using a predictor-corrector scheme, and the effect of the coupled terms can be neglected. These crossed terms lead to errors in the real parts of the β coefficients, but this effect is pretty well corrected with the predictor-corrector process. Its principle is based on the following reasoning. Based on equation 5.22, the intensity can be expressed as

$$I = f(\beta) + E(\beta), \tag{5.36}$$

where $f(\beta)$ is the function presented in eq. 5.22, where the quadratic term has been omitted. This function includes all the terms in β_0^0 ($(\beta_0^0)^2$, $(\beta_0^0) \cdot (\beta_n^m)$). The $E(\beta)$ function contains the quadratic terms in β^m_n ($O((\beta^m_n)^2)$). This term being neglected, it introduces an error depending on the β^m_n. The retrieval approach previously presented is based on the following simplification

$$I \approx f(\beta^{'}) \tag{5.37}$$

Solving this equation leads to a first approximation $\beta^{'}$ of β. These $\beta^{'}$s do not exactly correspond to the input image, since the equation is not exact because of the neglected term. However, they almost correspond to an image $I^{'}$ such that

$$I^{'} - E(\beta^{'}) = f(\beta^{'}). \tag{5.38}$$

Knowing $\beta^{'}$ allows to exactly compute $I^{'}$. Indeed, only the retrieval is approximated, the computation of the intensity from the β's is exact. Then, knowing I and $I^{'}$, it is possible to compute the error term $E(\beta^{'})$ corresponding to the approximated coefficients as

$$I = I^{'} - E(\beta^{'}). \tag{5.39}$$

We get,

$$E(\beta^{'}) = I^{'} - I \tag{5.40}$$

that can be considered as an approximation of the exact $E(\beta)$. This becomes exact when $\beta^{'}$ tends to β. It is then possible to compute $\beta^{''}$ using,

$$I - E(\beta^{'}) = f(\beta^{''}). \tag{5.41}$$

Each iteration gives a better approximation of the error related to the omission of the quadratic term. The predictor-corrector approach thus leads to a better estimation of the $\beta^{'}$.

It should theoretically be possible to reach any precision for the retrieved β as long as enough iterations are allowed for. Practically this is not the case because of the noise present in the images. In the following chapter, we will present some simulations where the iteration process is pursued until the relative error between the consecutive computed coefficients is lower than 10^{-4}.

Part III

Implementation

Chapter 6

Implementation of the Nijboer-Zernike theory

It is of high importance to understand how the "Nijboer-Zernike" (NZ) theory presented in the previous chapters can be implemented to compute point spread functions as well as to perform aberration retrieval and the related applications.

This chapter presents the way we have implemented the Nijboer-Zernike approach for the retrieval of aberrations as well as the tests that we have conducted to characterize the method and to determine its limitations and the influence of the parameters. The theory has been implemented in the Matlab language, in a code containing about 2000 lines distributed among some 15 functions.

6.1 Computation of the PSF based on the aberration coefficients

In this section, we present how the general Zernike coefficients (β) are used to compute the point spread function. A comparison between several PSF computing methods will also be addressed.

Section 4.3.2 described the Nijboer-Zernike theory and the way to use it to calculate the PSFs either from the classical Zernike coefficients (α) or from the generalized ones (β). The former can be used to express small phase aberrations (Strehl ratio $> 90\%$) whereas the latter could represent (theoretically) any type of aberrations. Practically, the β coefficients will be limited to large aberrations that correspond to a Strehl ratio higher than 30 %. The following sections describe the implementation of the way of computing PSFs from these general Zernike coefficients.

6.1.1 General Zernike coefficients (β)

Computation of the PSF using the β coefficients is based on equation 4.52 obtained in section 4.3.2. The use of this set of coefficients is far more interesting for our foreseen applications as they can be used with highly aberrated wavefronts, which can correspond either to larger individual aberrations or to more aberrations. The wavefront may contain up to 231 Zernike

polynomials (degrees (n) 0 to 20) in the β case whereas the α case should not be used with more than approximately 36 polynomials (degrees 0 to 7).

The use of the general Zernike β coefficients implies the following definition of the pupil function

$$P(\rho,\theta) = \sum_{n,m} \beta_n^m Z_n^m(\rho,\theta) \tag{6.1}$$

The basic equation, which must be used for the implementation of the PSF computation, is

$$U(r,\phi,f) = 2\pi i^m \sum_{n,m} \left[\beta_{cn}^m \cos(m\phi) + \beta_{sn}^m \sin(m\phi)\right]\ V_n^m(r,f) \tag{6.2}$$

where the $V_n^m(r,f)$ can be computed numerically as presented in subsection 6.1.2 hereafter.

Although the implementation of this equation could seem easy, it does introduce some difficulties, such as the need to use the same number of sine and cosine coefficients. Moreover, the case where $m = 0$ requires an adaptation of the equation and must be treated separately. Indeed, even if the sine part of the equation vanishes, odd coefficients with $m = 0$ may exist (piston-β_0^0, spherical aberration-β_4^0, ...). These coefficients thus have to be considered among the cosine coefficients.

The software that we have developed implements a slightly different approach. Analyzing equation 6.2, it is obvious that it represents a sum of all the aberrations balanced by the β_{cn}^m and the β_{sn}^m coefficients and expressed in terms of the V_n^m functions. The latter are nothing else than particular representations of the Zernike radial polynomials. The terms of the sum expressed in equation 6.2 can be rearranged using the Noll's classification of the Zernike polynomials (Noll 1976), and the equation of the PSF becomes a simple sum over all the Zernike coefficients.

$$U(r,\phi,f) = 2\pi i^m \sum_{nz} \beta_{nz}\, V_n^m(r,f) \begin{cases} \cos(m\phi) & \text{even coefficients} \\ \sin(m\phi) & \text{odd coefficients} \\ 1 & m = 0 \end{cases} \tag{6.3}$$

where the cosine is used with even coefficients and the sine is used with odd coefficients. In the special case where $m = 0$, the sine or cosine function is replaced by 1. This equation can be easily used for numerical applications, and the only inputs required are the Zernike coefficients sorted as in Noll's classification.

The intensity of the PSF can be computed as

$$I = |U(r,\phi,f)|^2 \tag{6.4}$$

Normalization of the intensity

The intensity computed this way has to be normalized, since the general approach using complex β coefficients allows to model amplitude variations; the coefficients influence the intensity. For example, the central peak intensity of the focused image is determined by the modulus of β_0^0. It is

thus convenient to normalize the intensity with the Strehl ratio corresponding to the aberration used to generate the image, which can be computed as a function of the β_n^m

$$Sr = \frac{|\beta_0^0|^2}{\sum_{n,m} |\beta_n^m|^2} \tag{6.5}$$

The central intensity of the PSF generated with the NZ approach is given by $|\beta_0^0|^2$. The PSF intensity is thus normalized by dividing it by the denominator of equation 6.5.

The normalization of the other parameters will be investigated in section 6.1.3 where the PSF computed from the Nijboer-Zernike approach will be compared to those computed from the Fourier transform. In that section, the Strehl ratios of the PSFs obtained from different methods will be compared in order to normalize the aberration coefficients involved in the Nijboer-Zernike computation, so that the NZ approach is consistent with the FFT method.

6.1.2 Numerical calculation of the $V_n^m(r, f)$ functions

The $V_n^m(r, f)$ functions form the basis of the Nijboer-Zernike theory. Indeed, they represent particular expressions of the radial Zernike polynomials in the basis of the Bessel polynomials. Unfortunately, the continuous analytical expression 4.51 derived in section 4.3.2 is difficult to use for numerical computations. However, Janssen (2002) has demonstrated the possibility to derive a numerically-usable discretized expression (see appendix C)

$$V_n^m(r, f) = (-1)^m e^{if} \sum_{l=1}^{\infty} (-2if)^{l-1} \sum_{j=0}^{\frac{n-m}{2}} v_{lj} \frac{J_{m+l+2j}(2\pi r)}{l(2\pi r)^l}, \tag{6.6}$$

where v_{lj} is given by

$$v_{lj} = (-1)^p (m + l + 2j) \frac{\binom{m+j+l-1}{l-1}\binom{j+l-1}{l-1}\binom{l-1}{p-j}}{\binom{q+l+j}{j}} \tag{6.7}$$

Fig. 6.1 shows several $V_n^m(r, f)$ functions for different n, m values that have been computed using equation 6.6.

Equation 6.6, involving two sums instead of an integral, can be easily used in numerical applications. However, the sum over l is infinite, which is not convenient for numerical computation. It is thus necessary to define an upper limit for this sum. Braat et al. (2002) showed that the number of terms ($L_{\max}$) to include into the sum should be of the order of $3f$ in order to reach an accuracy of 10^{-4}. They have established that most of the practical cases could be solved using $L_{\max} = 35$.

Another limitation of this discrete expression comes from the binomial terms. They require the computation of factorials of numbers that can become very large when many Zernike polynomials have to be used. The maximum number of polynomials that can be computed with this equation is 231, which corresponds to polynomials with a degree up to n=20.

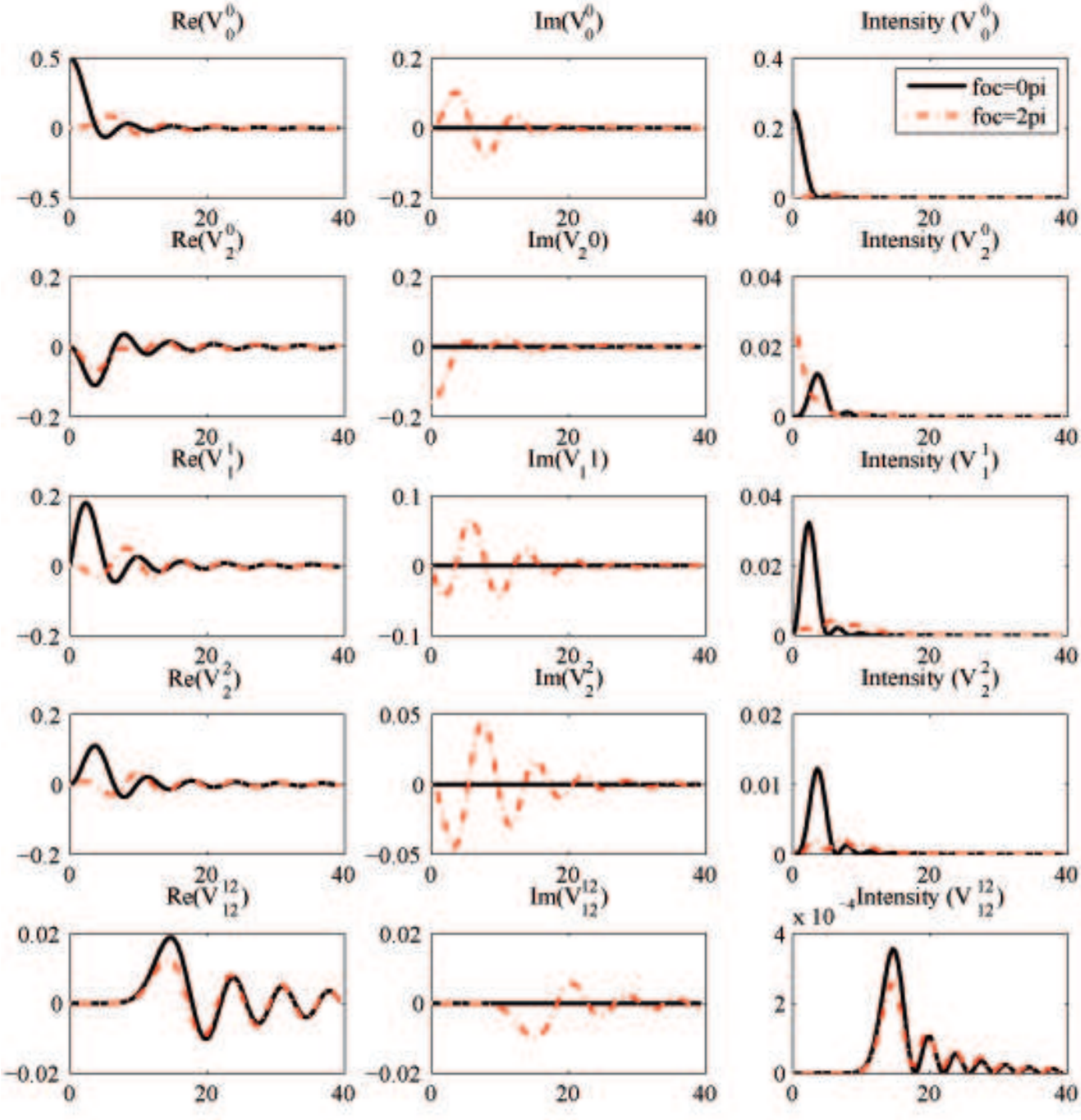


Figure 6.1: Illustration of the V_n^m functions for different values of n,m. The left column shows the real part of the functions, the second one shows their imaginary part and the last one represents the corresponding intensities. The black solid curves represent the function for zero defocus while the red dotted curves correspond to a defocus of 2π.

6.1.3 Comparison between the FFT PSF and NZ PSF

Now that we are able to compute the point spread function using the Nijboer-Zernike theory, it is interesting to compare it to the classical Fourier transform PSF computation.

This comparison does not aim at deciding which method is the best, but it will only help us to determine the normalization that should be applied to the aberration coefficients of the Nijboer-Zernike approach (β) to make it consistent with the Fourier transform method. It will be performed in two steps.

The first one consists in establishing a link between the classical Zernike coefficients (α) and the general coefficients (β). Then, using this relation, we will compare the Strehl ratios of the PSFs computed with both methods using corresponding (α - β) sets of aberrations.

Conversion between α and β coefficients

In order to compare the two PSF computation methods presented in chapter 4 (Fourier and Nijboer-Zernike), we have first to determine a relation between the classical and general Zernike coefficients. Indeed, the Fourier approach involves the α coefficients whereas the Nijboer-Zernike one uses the general β coefficients. The PSFs generated from those methods will be comparable only if the sets of coefficients used to compute them correspond to each other. In order to make such a comparison, let us start from the pupil function that can be written as (see section 4.2.2)

$$P(\rho,\theta) = A(\rho,\theta)\exp(i\sum_{n,m}\alpha_n^m Z_n^m(\rho,\theta)) = \sum_{n,m}\beta_n^m Z_n^m(\rho,\theta) \tag{6.8}$$

In the case where the aberrations are small and the amplitude is uniform ($A(\rho,\theta)=1$) over the whole pupil, the exponential term can be expanded in a Taylor series that can be limited to the first order.

$$\exp(i\Phi(\rho,\theta)) \approx 1 + i\Phi(\rho,\theta) \tag{6.9}$$

Replacing the phase ($\Phi(\rho,\theta)$) by its expression as a Zernike polynomial series, we get

$$1 + i\sum_{n,m}\alpha_n^m Z_n^m(\rho,\theta) = \sum_{n,m}\beta_n^m Z_n^m(\rho,\theta) \tag{6.10}$$

As the Zernike polynomials are orthogonal and $Z_0^0 = 1$, we find a simple relationship between the α and β coefficients that can be expressed as

$$1 + i\alpha_0^0 = \beta_0^0 \qquad n,m = 0 \tag{6.11}$$

$$i\alpha_m^n = \beta_n^m \qquad n,m \neq 0 \tag{6.12}$$

where all α coefficients are real and the β coefficients are complex. In the study hereafter, these relations cannot be used since we want to explore a wide range of aberrations, for which the "small aberrations" assumption is not true anymore.

Starting back from the exact equation 6.8, we should be able to determine a more accurate relation. However, an analytical relation would be very difficult to obtain as it would imply products of Zernike polynomials. We have thus decided to test two different numerical approaches.

The first one, based on the equation of the pupil itself, consists in computing the complex pupil corresponding to one type coefficient (α or β) and then in decomposing it in the other basis. The second way we tested is based on the retrieval process, a set of PSF is computed with the Fourier transform method based on α coefficient (for different focal positions) and the Nijboer-Zernike retrieval method is then used to determine the corresponding β coefficients. Both methods are detailed hereafter.

The pupil method is based on equation 6.8, which can be used to generate a complex pupil from both types of coefficients. It has thus the advantages of enabling the conversion of aberrations from the α to the β basis and inversely. In the following, we will assume that the amplitude is uniform over the whole pupil as it is the case for most optical instruments.

Let us first consider the conversion from classical α to general β coefficients. The pupil is computed from the equation

$$P(\rho,\theta) = \exp(i\sum_{n,m}\alpha_n^m Z_n^m(\rho,\theta)) \tag{6.13}$$

that can be written as

$$P(\rho,\theta) = \cos\sum_{n,m}\alpha_n^m Z_n^m(\rho,\theta) + i\sin\sum_{n,m}\alpha_n^m Z_n^m(\rho,\theta) \tag{6.14}$$

One thus get a real pupil and an imaginary pupil. Using simple identification of the real and imaginary terms, we get the relation between the α and β coefficients

$$\begin{aligned}\cos\sum_{n,m}\alpha_n^m Z_n^m(\rho,\theta) &= \sum_{n,m}\Re(\beta_n^m)Z_n^m(\rho,\theta)\\ \sin\sum_{n,m}\alpha_n^m Z_n^m(\rho,\theta) &= \sum_{n,m}\Im(\beta_n^m)Z_n^m(\rho,\theta)\end{aligned} \tag{6.15}$$

By decomposing the real part of the pupil in Zernike polynomials, it is thus possible to get the real part of the β coefficients whereas the decomposition of the imaginary part of the pupil gives access to the imaginary part of the β coefficients.

The decomposition is performed coefficient by coefficient by computing the integration of the product between the surface (real or imaginary part of the pupil in this case) and successively for each Zernike polynomial. Let us call $F(\rho,\theta)$ the surface to decompose, it is thus assumed to be writable in the form

$$F(\rho,\theta) = \sum_{n,m}\gamma_n^m Z_n^m(\rho,\theta) \tag{6.16}$$

where the γ_n^m coefficients represent the weights of the aberrations composing the surface. The integral described before is

$$\begin{aligned}\gamma_{n'}^{m'} &= \frac{1}{2\pi}\int_0^1\int_0^{2\pi}\sum_{n,m}\gamma_n^m Z_n^m(\rho,\theta)\cdot Z_{n'}^{m'}(\rho,\theta)\cdot\rho d\rho d\theta\\ &= \frac{1}{2\pi}\sum_{n,m}\int_0^1\int_0^{2\pi}\gamma_n^m Z_n^m(\rho,\theta)\cdot Z_{n'}^{m'}(\rho,\theta)\cdot\rho d\rho d\theta\end{aligned} \tag{6.17}$$

The integral of the sum is equivalent to the sum of the integrals, which is expressed in the second line. Using this property and because of the orthogonality of the Zernike polynomials, all the integrals corresponding to $m, n \neq m', n'$ vanish. Their normalization ensures that the integral corresponding to $m, n = m', n'$ is $2\pi\gamma_{n'}^{m'}$. Applying this formula successively with every Zernike polynomial allows to determine the corresponding γ_n^m. In our case, these γ_n^m coefficients correspond respectively to the real and imaginary parts of the β_n^m coefficients. This conversion from α to β terms may reveal difficult because of the sine and cosine in the phase that may vary rapidly when the phase errors become large. These high frequency variations can be difficult to represent with a limited sum of Zernike polynomials, and this method thus requires to use a larger number of β coefficients than the number of α coefficients used to compute the pupil.

The inverse conversion, from β to α coefficients, is based on the same principle. The complex pupil is computed using the equation

$$P(\rho, \theta) = \sum_{n,m} \beta_n^m Z_n^m(\rho, \theta) \tag{6.18}$$

and the phase function ($\Phi(\rho, \theta)$) of the pupil can be determined by means of

$$\Phi(\rho, \theta) = \arctan\left(\frac{\sum_{n,m} \Re(\beta_n^m) Z_n^m(\rho, \theta)}{\sum_{n,m} \Im(\beta_n^m) Z_n^m(\rho, \theta)}\right) \tag{6.19}$$

This phase can be decomposed in α coefficients by the same method as previously exposed. Even if this inverse decomposition does not present the sine and cosine difficulties previously encountered, another problem appears. The arctangent being defined between $-\pi$ and $+\pi$, the phase function may needs to be unwrapped before being decomposed. This method will be used to convert the aberrations from the β to the α basis.

The PSF method is based on the Nijboer-Zernike retrieval method presented in chapter 5. Point spread functions are computed from the α coefficients for several positions around the focus using the Fourier approach. They are then used in the retrieval process that determines the aberrations contained in the PSFs. As the retrieval is performed on the β basis, this approach allows to convert α to β coefficients. This conversion is possible only because we have used the same weighting of the polynomials in both methods.

This second approach presents the disadvantages of only being usable to convert α to β. As in the previous case, using a larger number of general coefficients for the retrieval than of classical ones (for the PSF computation) improves the correspondence between the sets. However, the process does not converge when too many β coefficients are being used. When small α aberrations are used to generate the PSFs, the results correspond to the simplified relation previously exposed. For larger aberrations, small crossed terms appear, both in the real and imaginary parts. In the following, we will use this method to convert the aberrations from the α to the β basis and the "pupil method" for the inverse conversion.

Comparison of the PSF computation methods

We now compare the different way of computing PSFs. The phase aberrations that will be used to perform the comparison are presented in fig. 6.2. They decrease as $1/n^2$ and the three first

aberrations, piston and tip/tilt, are set to zero, since they only displace the central peak but does not modify its shape.

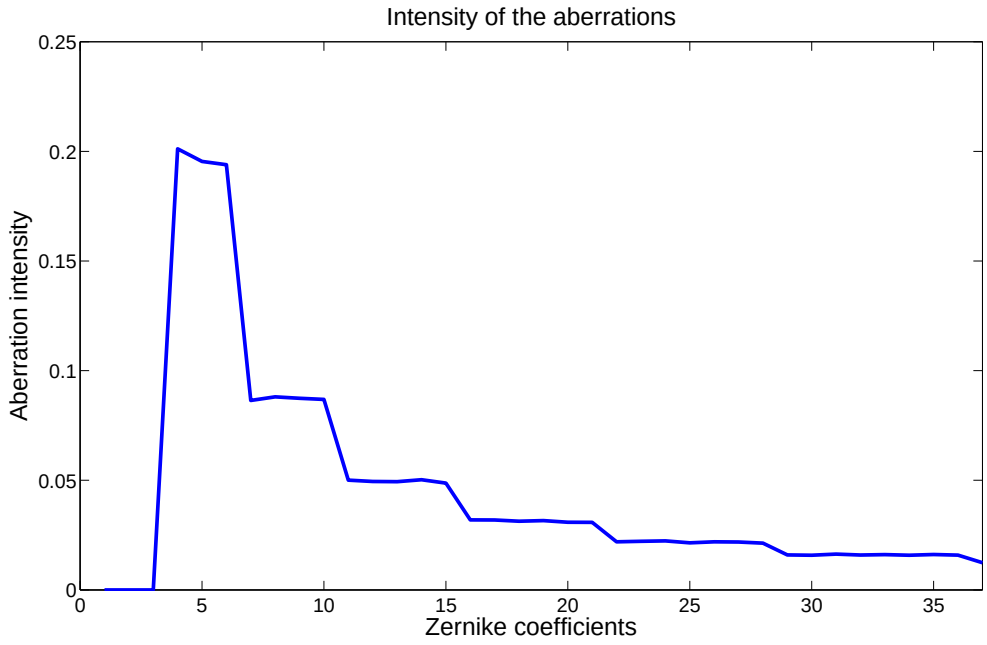


Figure 6.2: Imaginary part of the input coefficients (β_n^m) used to compare the different methods. This corresponds to the intensity of the different phase aberrations applied for the test. The values decrease as n^{-2}, the piston and tip/tilt aberrations have been set to zero.

These aberrations correspond to the α coefficients. We used the second approach described before to convert them into the β basis.

The first step of comparison is based on the PSF itself. Fig. 6.3 shows the PSFs computed with several methods. The first one (on the left) has been computed with the NZ approach using the β coefficients, the two others have been obtained by FFT. The difference between the latter two comes from the computation of the pupil function; the first one is based on the classical series expansion (α) whereas the second one uses the general expansion (β).

All three PSFs are very similar. This is due to the way of computing the relation between the α and β coefficients, since the latter come from the retrieval of the α PSF. This retrieval is supposed to fit the input PSF, it is thus logical that the PSFs are very similar. The PSFs that have been computed from the general Zernike coefficients (β) either using the NZ or the FFT method are very similar. The one computed with the classical coefficients (α) presents small differences coming from the imperfect conversion between the two types of coefficients.

It is well known that the aberrations defining the pupil in case of the Fourier approach are expressed in waves. It seems from this first observation that the β aberrations are also expressed in waves since the simplified relation is respected for small aberrations provided that the weighting of the polynomials is the same. No extra normalization is thus required.

The second comparison concerns the behavior of the Strehl ratio as a function of the amplitude of the highest aberrations. There are many ways of estimating a PSF Strehl ratio. In case of the Fourier computation, the phase is directly involved, and the Maréchal approximation (equation 4.22) can be used. The Strehl value computed this way will be called the "pupil" Strehl ratio. As

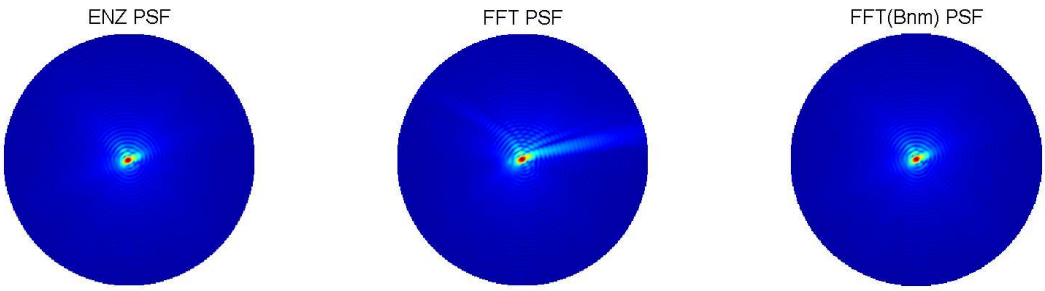


Figure 6.3: Comparison between Point Spread Functions computed in different ways. Left: ENZ approach using the general Zernike coefficients (β). Center: Fourier transform of the pupil built from the classical Zernike coefficients (α). Right: Fourier transform of the pupil built from the β coefficients. The intensity of the three PSFs are to the power 1/4 in order to improve the visibility of the rings.

far as the NZ PSF is concerned, the Strehl is given by formula 6.5. It will be called "Bnm" Strehl ratio. Another way of determining the Strehl ratio, for any normalized PSF, is to measure the value of the central peak of the PSF, these ones will be called the "Peak" Strehl ratios. However, in case of PSFs involving β coefficients (NZ and FFT(β)), both measurements will give the same results as the peak intensity is normalized by the same formula.

The results of the Strehl measurements are presented in figs. 6.4 and 6.5. The curves that correspond to the different methods of computation are almost the same for a Strehl ratio larger than 90%. Below that limit, differences due to the PSF computation method appear. The two so-called "pupil" curves are close to each other and the so-called "Bnm" and "peak" Strehl values are close to each other. As previously mentioned, the Strehl computed using equation 6.5 is used to normalize the intensity of the PSFs computed with β coefficients. It is thus logical to find the corresponding curves close to each other. However they are also close to the peak Strehl curve corresponding to the α PSF computed by Fourier transform, which means that the β related Strehl equation corresponds quite well to a measurement of the Strehl ratio on the peak. We have performed the same measurement using more β coefficients for the conversion. In this case, the curves of the same type (pupil or peak) get closer to each other.

A break appears, around 75% of Strehl, in the "peak" Strehl curve corresponding to the Fourier α PSF. The same break also occurs in all the β curves, because of the conversion method, since the β coefficients come from a fit of the α PSF. The break in the α curve thus also influences the β values and the corresponding curves.

It comes out of this comparison that the results related to the NZ approach correspond quite well to those of the Fourier method and that no extra normalization is required to ensure the consistence between the two methods. Both theories give similar PSFs and Strehl ratios, and we may conclude that the β_n^m coefficients are normalized in the same way as the α_n^m, i.e. in waves.

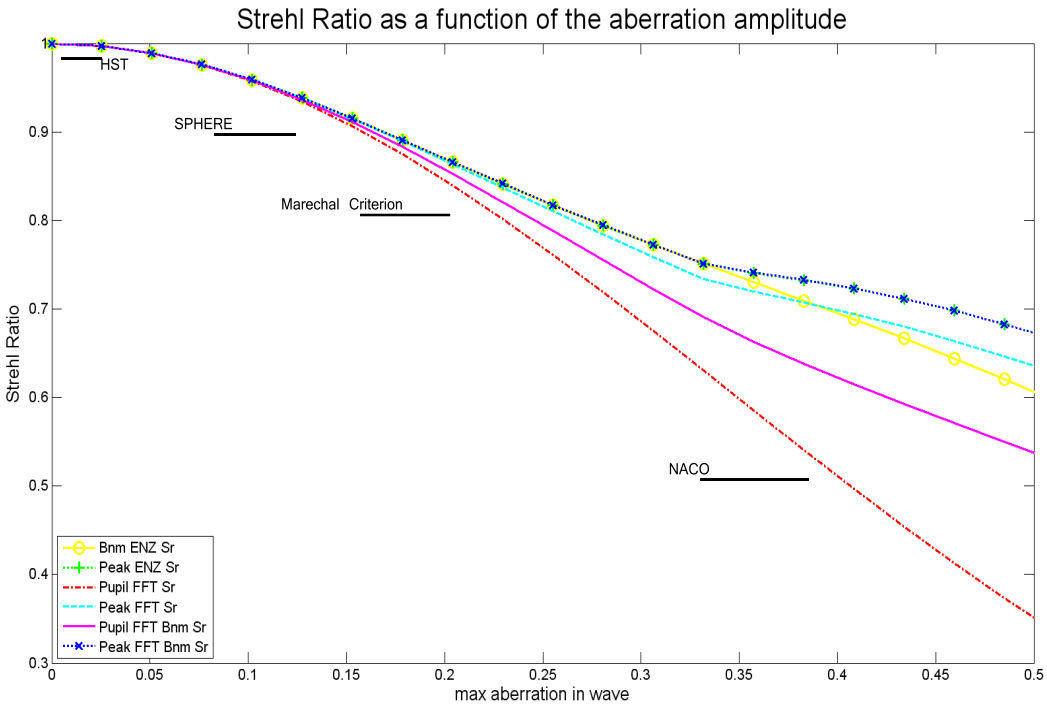


Figure 6.4: Comparison of the PSF Strehl ratios obtained with the different methods as a function of the amplitude of the aberrations. The curves corresponding to "Bnm" and "Pupil" Strehl ratios present results that have been computed internally. The "peak" curves present values of Strehl ratio that have been measured from the peak of the PSF. Typical Strehl ratio experiment values are indicated.

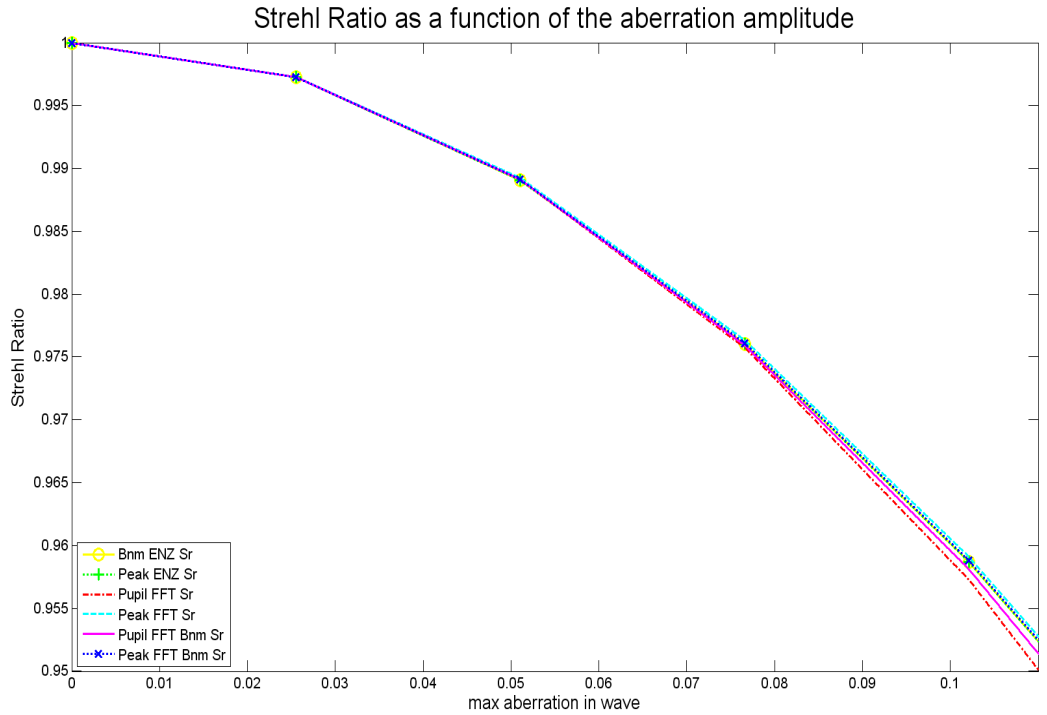


Figure 6.5: Detailed view of fig. 6.4 for the Strehl ratios between 95% and 100%.

6.2 Retrieval of the aberration coefficients from the PSF analysis

The Nijboer-Zernike retrieval method consists in analyzing through-focus images in order to compute the aberrations that affect these images. As demonstrated in chapter 5, the retrieval simply consists in the resolution of linear systems (see equations 5.33, 5.34, 5.35). The equations of these systems are based on several functions ($\Psi^m_{\text{meas}}(r,f)$, $\chi^m_n(r,f)$ and $\Psi^m_n(r,f)$), which can be easily computed from the V^m_n functions presented earlier and from the defocused intensity PSF. The implementation of these calculations is explained hereafter and the retrieval steps are illustrated with a practical case.

6.2.1 First step: the input images

The first point consists in extracting the modes of the input PSFs, which are illustrated in fig 6.6. These example images have been simulated using the forward NZ calculation described at the beginning of this chapter. The aberrations which have been injected in these PSFs are presented in fig. 6.7, they are the same as those previously used for the comparison between the PSF-computing methods except that we added tip/tilt coefficients.

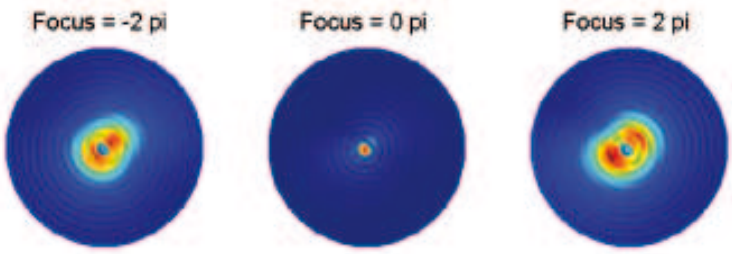


Figure 6.6: Point Spread Functions with known aberrations (presented before) that will be used to illustrate the retrieval method. The three images correspond to three positions around the focus. Left: one wave (2π) before the focus; Center: at the focus; Right: one wave after.

The aberrated images are composed of several modes in sine and cosine functions. The first step of the image study is the decoupling of these modes, which can be performed with a polar Fourier transform of the images (equation 5.24). It is given by:

$$\Psi^m_{\text{meas}}(r,f) = \frac{1}{2\pi}\int_0^{2\pi} I_{\text{meas}}(r,\phi,f)e^{im\phi}d\phi \tag{6.20}$$

where $I_{\text{meas}}(r,\phi,f)$ is the measured intensity expressed in polar coordinates. Separating the real and imaginary parts of the exponential term results in the decoupling of the sine and cosine terms.

$$\begin{aligned}\Psi^m_{\text{meas}}(r,f) \quad &= \frac{1}{2\pi}\int_0^{2\pi} I_{\text{meas}}(r,\phi,f)\cos(m\phi)d\phi \\ &+i\,\frac{1}{2\pi}\int_0^{2\pi} I_{\text{meas}}(r,\phi,f)\sin(m\phi)d\phi.\end{aligned} \tag{6.21}$$

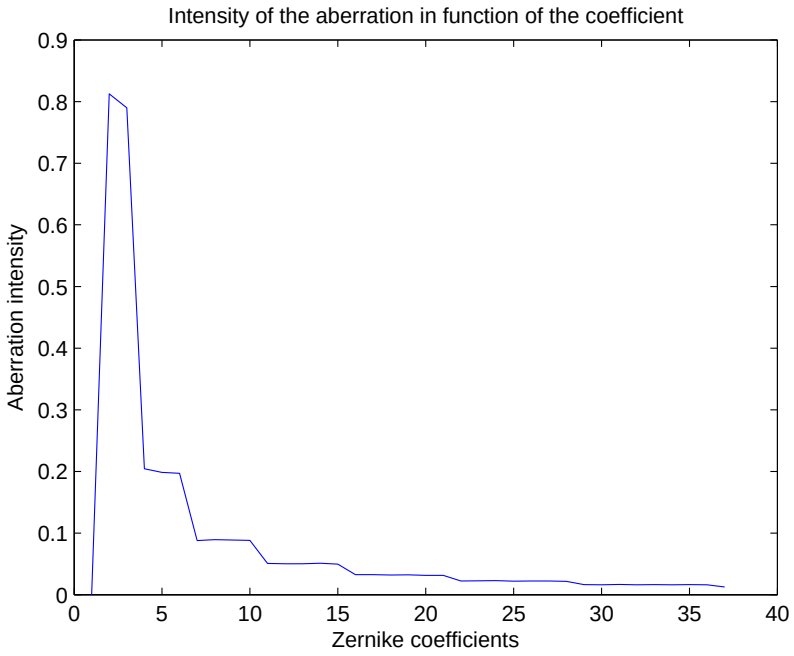


Figure 6.7: Intensity of the phase aberrations introduced to generate the PSFs presented in fig. 6.6. The same profile as in fig. 6.2 has been used, but in this case tip/tilt aberrations are present.

The real part is called $\Psi^m_{\text{cmeas}}(r,f)$ and the imaginary one $\Psi^m_{\text{smeas}}(r,f)$

$$\begin{aligned}\Psi^m_{\text{meas}}(r,f) &= Re(\Psi^m_{\text{meas}}(r,f)) + i\, Im(\Psi^m_{\text{meas}}(r,f)) \\ &= \Psi^m_{\text{cmeas}}(r,f) + i\,\Psi^m_{\text{smeas}}(r,f).\end{aligned} \tag{6.22}$$

The real and imaginary parts of $\Psi^m_{\text{meas}}(r,f)$ are thus decoupled in modes m, and in sine and cosine respectively. It is important to note that the intensity is measured with a CCD camera, and is thus acquired in cartesian (x,y) coordinates which must be converted. The whole integration is adapted in our software. It consists in an angular average of each image modulated with a sine or a cosine term. However, it is irrelevant to calculate this average for each particular value of the radius r. Indeed, the conversion from cartesian (x,y) to polar (r,ϕ) creates many values of r which are not distributed uniformly on circles. We present here the two different approaches that we have tested.

First approach

At first, the angular average mentioned above was computed for ranges of the radius centered on each value of the radius vector and whose width was equal to one radius step. As an example, let us assume that the radius vector is defined as $R = [0,1,2,3,4]$ and that we are computing $\Psi^m_{\text{meas}}(2,f)$, the r range defined previously corresponds to the interval $[1.5, 2.5[$, from which 2.5 is excluded.

The image pixels corresponding to that particular range of radius are selected and the corresponding angles are computed from the cartesian coordinates. These points are then sorted in order of ascending angles in order to create a variable vector containing the angles and a function

vector containing the data. The function vector is then integrated along the variable vector using the trapeze method.

This approach presents some particular cases when the considered values of the radius are extremal. When the radius is maximum, the range cannot be defined symmetrically around its value. Instead, r is chosen as the upper limit of the range. Taking the same example as previously with $r = 4$ the range of radii would be [3.5,4]. Apart from this change of definition, the computation of the integral is done in the same way.

The other particular case, corresponds to the minimum radius. Assuming that the image is centered on a particular pixel, this case corresponds to only one data point, which does not depend on ϕ anymore. The integral must then be evaluated analytically:

$$\Psi^m_{\text{meas}}(0,f) = \frac{I_{\text{meas}}(0,f)}{2\pi}\int_0^{2\pi} e^{im\phi}d\phi = \begin{cases} I_{\text{meas}}(0,f), & m=0 \\ 0, & m\neq 0 \end{cases}, \tag{6.23}$$

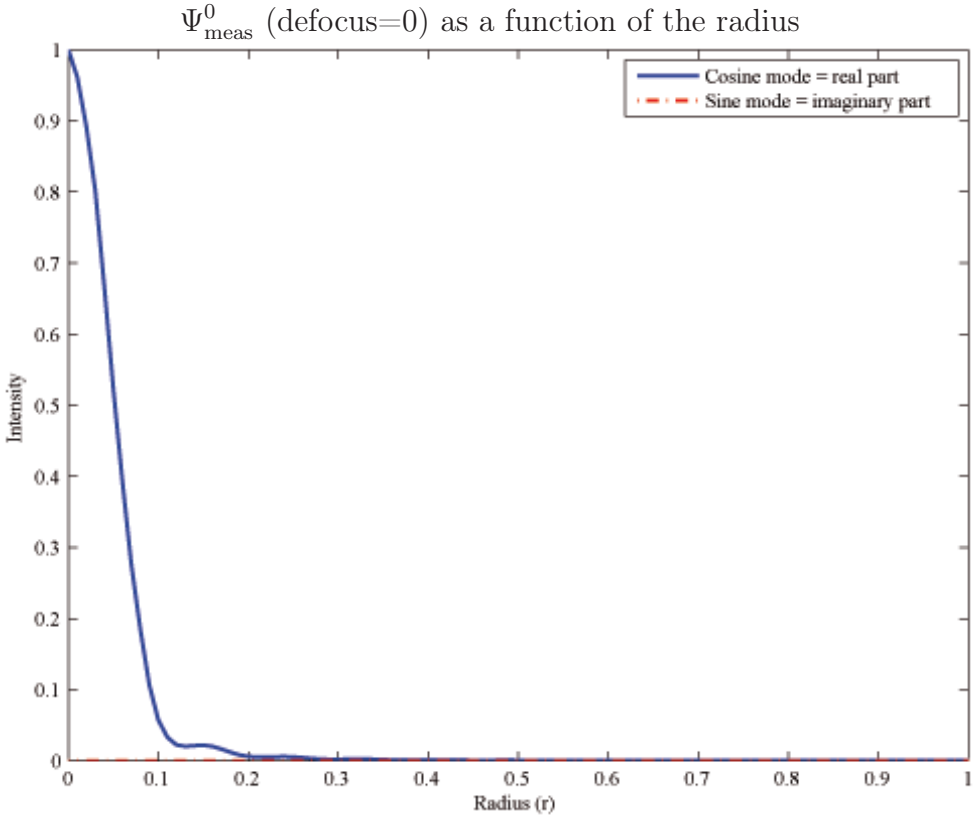


Figure 6.8: The Ψ^m_{meas} function intensity as a function of the radius for zero defocus and m=0. The blue solid curve represents the cosine part and the red dash-doted line (superimposed over the x-axis) corresponds to the sine part of the function which is null in this particular case.

This approach presents the disadvantage of computing averages over small ranges of the radius, which is not a problem for points far away from the center, where the intensity does not vary rapidly with the radius. However, as far as the center is concerned, this is not anymore the case and the averaging along the radius creates small differences that result in errors during the retrieval.

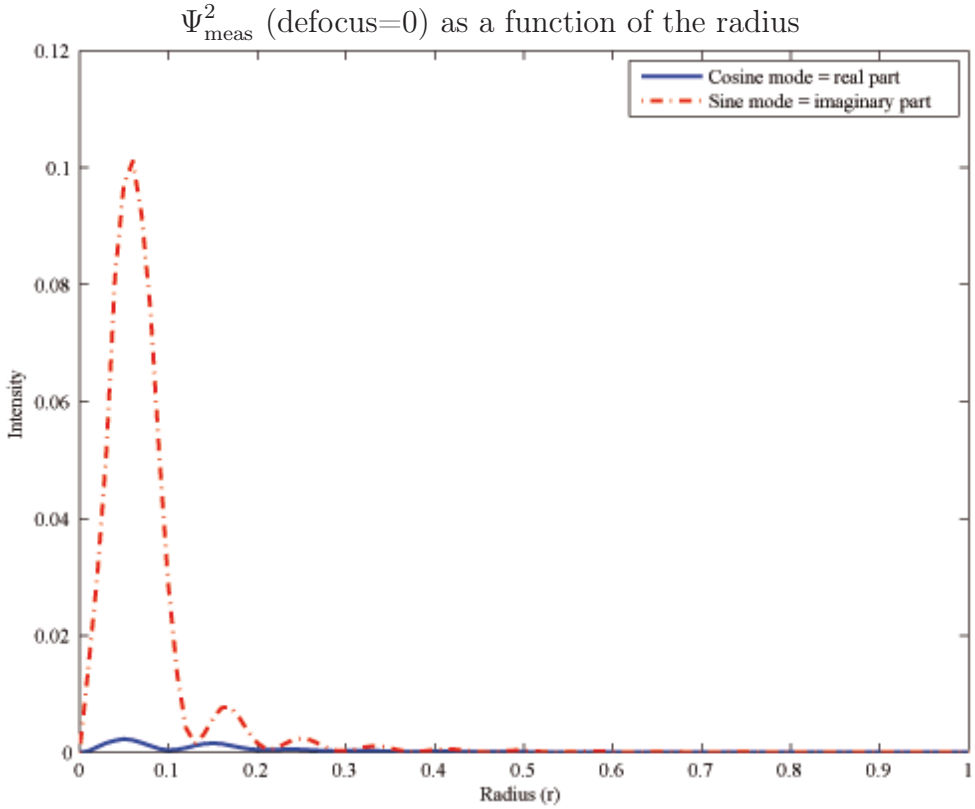


Figure 6.9: Same description as in fig. 6.8 with m=2. In this case, the imaginary part is more important than the real one.

Second approach

The other approach that has been considered simply consists in defining a range for the radius and another one for the azimuthal angle, which ranges constitute a polar grid. We convert the cartesian image by filing this polar grid: each position is associated with the intensity value from the cartesian image that best corresponds to these particular coordinates.

This gives good results for radial positions away from the center of the image, but around the center, the same positions of the cartesian image are used many times for several coordinates of the polar grid. In order to circumvent this issue, we interpolate the center of the image in such a way that the sampling is large enough to complete the polar grid. The number of interpolations is defined to optimize the result accuracy and to minimize the extra-time consumed. It should be noted that a Spline interpolation gives better results than a linear interpolation and that it does not take much time. It is also important to notice that the interpolation does not change the polar grid as only the points that best correspond to the polar coordinates are used for the conversion. It is thus obvious that this process does not increase the computation time too much as the number of elements used to convert the image does not depend on the interpolation. The only increase in time comes from the interpolation itself, and, as it is performed only at the center of the image (one tenth of the radius), this process is very fast (less than 0.5s for an 511×511 pixel image).

6.2.2 The inner product

Once the images have been decomposed in their basic modes, it is necessary to match these components with the corresponding aberrations (i.e. the V_n^m representation of the Zernike polynomials). This comparison is performed by the following inner product, defined by

$$(A, B) = \int_0^\infty \int_{-\infty}^\infty A(r, f)B^*(r, f)rdrdf. \tag{6.24}$$

This inner product corresponds to a particular correlation of the two functions A and B. It consists in projecting the function A in the B function basis. In practical cases, these functions are replaced by $\Psi^m_{\text{meas}}(r, f)$ and $\Psi^m_n(r, f)$ or $\chi^m_n(r, f)$. The two last functions, defined in equations 5.25 and 5.26, are normalization of the V_n^m functions that correspond respectively to the phase and to the amplitude aberration templates.

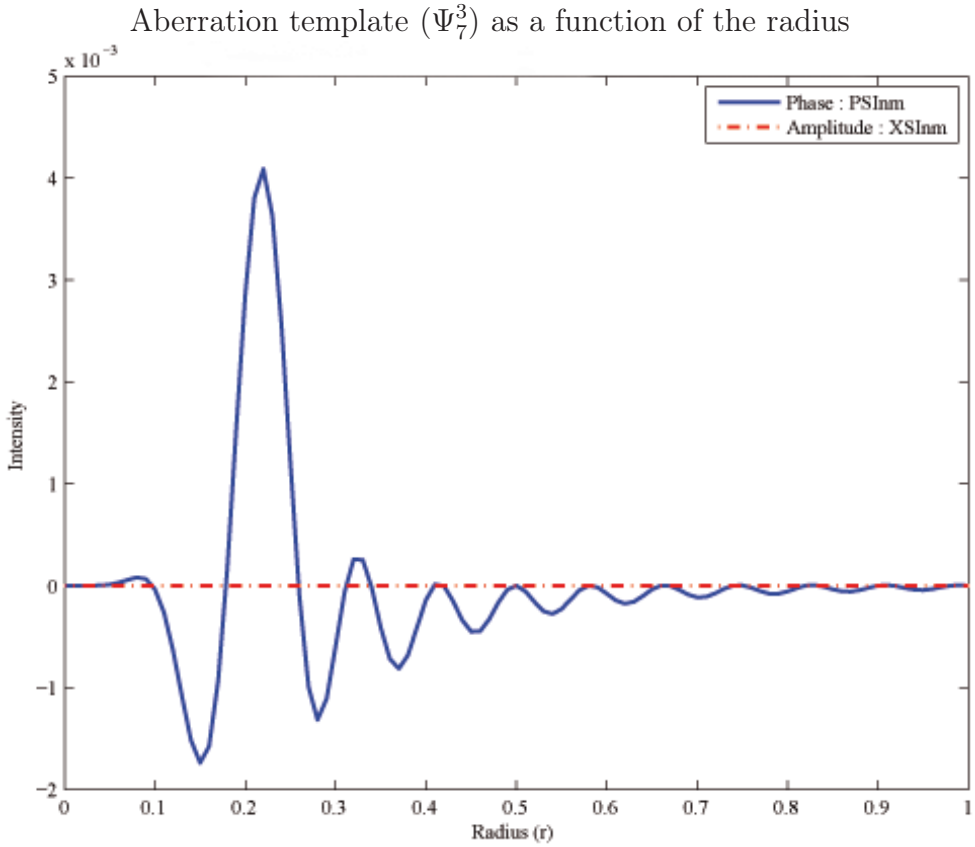


Figure 6.10: Ψ_n^m (solid blue line) and χ_n^m (dash-dotted red line) function as a function of the radius for zero defocus and n=7, m=3.

The projection of the image modes on the models using the inner product leads to the computation of the β aberration coefficients. Numerically, this inner product is computed using two encapsulated trapezium integration methods, with the radius and focus vectors as variables.

The introduction of the inner product allows to linearize the equations of the intensity, which can be expressed in matrix form as

$$\mathbf{G} \cdot \mathbf{u} = \mathbf{r}, \tag{6.25}$$

where the $\cdot$ sign involves a matrix product, $\mathbf{u}$ is the vector with unknown coefficients, $\mathbf{G}$ and $\mathbf{r}$ are defined using the inner product. The elements of the G matrix are defined as

$$G^m_{n,n'}(\chi) = (\chi^m_n, \chi^m_{n'}) \tag{6.26}$$
$$G^m_{n,n'}(\Psi) = (\Psi^m_n, \Psi^m_{n'}) \tag{6.27}$$

and the elements of the $\mathbf{r}$ vector are defined as

$$(r^m_{n'})_{\chi c} = (\Psi^m_{_c\text{meas}}, \chi^m_{n'}) \tag{6.28}$$
$$(r^m_{n'})_{\Psi c} = (\Psi^m_{_c\text{meas}}, \Psi^m_{n'}) \tag{6.29}$$
$$(r^m_{n'})_{\chi s} = (\Psi^m_{_s\text{meas}}, \chi^m_{n'}) \tag{6.30}$$
$$(r^m_{n'})_{\Psi s} = (\Psi^m_{_s\text{meas}}, \Psi^m_{n'}) \tag{6.31}$$

where the subscripts c and s correspond to an inner product involving the Ψ^m_{meas} related to the cosine and sine modes, respectively. The subscripts (χ) and (Ψ) denote a right hand side defined by an inner product involving $\chi^m_n(r,f)$ and $\Psi^m_n(r,f)$, respectively.

The systems defined in this way represent the relationship between the image modes projected on the aberration templates ($\Psi^m_n(r,f)$, $\chi^m_n(r,f)$) and the linear combinations of the aberration models. The coefficients involved in the combination are related to the unknown β coefficients. The resolution of these linear systems determines the aberration coefficients, it is described in the following sections.

6.2.3 Resolution of the systems for $m \neq 0$

The case where $m \neq 0$ is treated using the second and third groups of systems (equations 5.33 and 5.34), respectively related to the cosine and sine coefficients. Writing these systems using the matrix defined in the previous section, we get

$$\sum_n \beta^0_0 \, \Re(\beta^m_{cn}) \, G\chi^m_{n,n'} \approx (r^m_{n'}(\chi))_c \tag{6.32a}$$
$$\sum_n \beta^0_0 \, \Im(\beta^m_{cn}) \, G\Psi^m_{n,n'} \approx (r^m_{n'}(\Psi))_c \tag{6.32b}$$
$$\sum_n \beta^0_0 \, \Re(\beta^m_{sn}) \, G\chi^m_{n,n'} \approx (r^m_{n'}(\chi))_s \tag{6.32c}$$
$$\sum_n \beta^0_0 \, \Im(\beta^m_{sn}) \, G\Psi^m_{n,n'} \approx (r^m_{n'}(\Psi))_s \tag{6.32d}$$

for $m \neq 0$ and $n, n' = m, m+2, ...,$.

Let us call $\mathbf{u}_a, \mathbf{u}_b$ the unknowns related to the cosine terms and $\mathbf{u}_c, \mathbf{u}_d$ those related to the sine terms, their elements are defined as

$$u^m_{an} = \beta^0_0 \, \Re(\beta^m_{cn}) \tag{6.33a}$$
$$u^m_{bn} = \beta^0_0 \, \Im(\beta^m_{cn}) \tag{6.33b}$$
$$u^m_{cn} = \beta^0_0 \, \Re(\beta^m_{sn}) \tag{6.33c}$$
$$u^m_{dn} = \beta^0_0 \, \Im(\beta^m_{sn}) \tag{6.33d}$$

The unknown vectors are found by inverting the **G** matrix. Once the systems have been solved, the β coefficients can be found using the following relations

$$\beta^m_{cn} = \frac{u^m_{an} + i u^m_{bn}}{\beta^0_0} \tag{6.34a}$$
$$\beta^m_{sn} = \frac{u^m_{cn} + i u^m_{dn}}{\beta^0_0} \tag{6.34b}$$

Determination of the β coefficients require the knowledge of β^0_0, which is determined using the first group of systems corresponding to $m = 0$. This case is treated in the next section.

6.2.4 Resolution of the systems for $m = 0$

In the case where $m = 0$, the systems can also be written as in equation 6.32. Apart from their first elements that are respectively $\frac{(\beta^0_0)^2}{2}$ and 0, the unknown vectors are defined as in equations 6.33(a-d).

$$u^0_{en} = [\frac{1}{2}(\beta^0_0)^2, \beta^0_0 \, Re(\beta^0_{c2}), ..., \beta^0_0 \, Re(\beta^0_{cn})] \tag{6.35a}$$
$$u^0_{fn} = [0, \beta^0_0 \, Im(\beta^0_{c2}), ..., \beta^0_0 \, Im(\beta^0_{cn})] \tag{6.35b}$$

Computation of the unknown vector of both systems is performed as in the previous case. The β^0_0 term is assumed to be real and positive, which corresponds to a purely amplitude phenomenon as it is the case for the reflection of light on a mirror[23]. It is computed from the first element of u^0_{en} with the following equation

$$\beta^0_0 = \sqrt{2u^0_{en}}. \tag{6.36}$$

Once β^0_0 is known, it is possible to completely determine all the other values of β^m_{cn} and β^m_{sn}. It is then interesting to put the sine and cosine β coefficients together and to classify them according to Noll's approach (Noll 1976).

[23]This is the case in scalar light propagation theory. In the vectorial case, where polarization enters into account, a phase may exist between the two polarization bases. This phenomenon gets more important for off-axis mirror reflection.

6.3 Study of the Nijboer-Zernike retrieval method

This section aims at presenting the study of the impact of the parameters that appear in the retrieval process. We will begin with the limitation of the aberration retrieval method, such as the highest aberration coefficients that can be retrieved, and the accuracy that can be reached with this method. We then study the impact of the noise and of the completeness on the retrieval. Finally, we will investigate the improvement that can be obtained by optimizing the image sampling.

6.3.1 Predictor-corrector convergence

Use of the predictor-corrector iterative method aims at correcting the β coefficients from the error due to the omission of the quadratic term in the intensity equation of the retrieval. It is important to determine what parameters influence the convergence of this process.

The convergence of the iterative process depends mainly on two parameters; the number of unknown coefficients and the amplitude of the aberrations that can be expressed in terms of the Strehl ratio.

It appears from the tests that we have conducted, that when only the very low order terms (first ten) are to be retrieved, the limiting Strehl value is about 35%, which means that any better image can be decomposed into its aberrations. If higher order terms are present (first one hundred), only images with a Strehl ratio higher than 45% can lead to a converging retrieval process. However, this is a lower estimation as the tests we conducted used uniform aberration coefficients (all the same), which is not realistic since the amplitude of the coefficients decreases with the inverse of the space frequency ($1/n^2$). As high order coefficients are related to higher frequencies, they should be appreciably smaller. As an example, retrieving one hundred coefficients is possible as long as the coefficient values are under $0.3 + i\,0.3$ (Strehl 47%), but in practical cases, the first coefficients would be larger and the high order coefficients would be smaller. In case only six coefficients have to to be found, it is possible to retrieve them as long as they are smaller than $1.13 + i\,1.13$ (Strehl 32%) and no defocus is applied ($\beta_2^0 = 0$).

It is also important to note that a few specific coefficients are critical. It is the case mainly for the defocus, tip and tilt terms that can disturb the convergence, because of the asymmetry they cause in the images. Once this asymmetry is too large, we have noticed that the process hardly converges. However this is not very important since for practical cases, the best focus will be selected as a basis for the retrieval. The images used for the retrieval will thus be roughly symmetrically located around the focus.

When the aberration values are nearly critical, the convergence of the process can become very slow. Indeed, this case corresponds to considerably aberrated images for which the quadratic term in the expression of the intensity is of the same order of magnitude as the other terms. The error related to the omission of this term is not negligible anymore and the corrections applied are often too large. The aberration coefficients are thus over-corrected, which results in an oscillating behavior that may converge or not depending on the true values of the aberrations. If the process converges, it is very slow. The example presented by the blue solid curve (fig. 6.11) shows an oscillating case that has been stopped after one hundred iterations because it did not finish converging. The whole process took about eleven minutes of computing on a 2GHz dual

core processor.

However, the predictor-corrector iterations can be slightly modified to damp the oscillations and then increase the convergence speed. The red dotted curve (fig. 6.11) represents the same case treated with our oscillation damped process, which converged after 39 iterations for a total computation time of about four minutes and thirty seconds.

The principle of this accelerating process is quite simple. The presence of oscillations means that the retrieved values vary around the exact values. The convergence implies that the oscillation amplitude should become smaller with the increasing number of iterations. We thus assume that the mean value of the two last iterations is nearer to the exact result than the values computed during these iterations. The averaged coefficients are used as new values for the following two iterations and the process is repeated.

This accelerating process, that is very interesting in case of an oscillating convergence, may slow down in case of a normal (non-oscillating) convergence. In this case the classical approach gives better performances (fig. 6.12, blue solid line and red dash-dotted line).

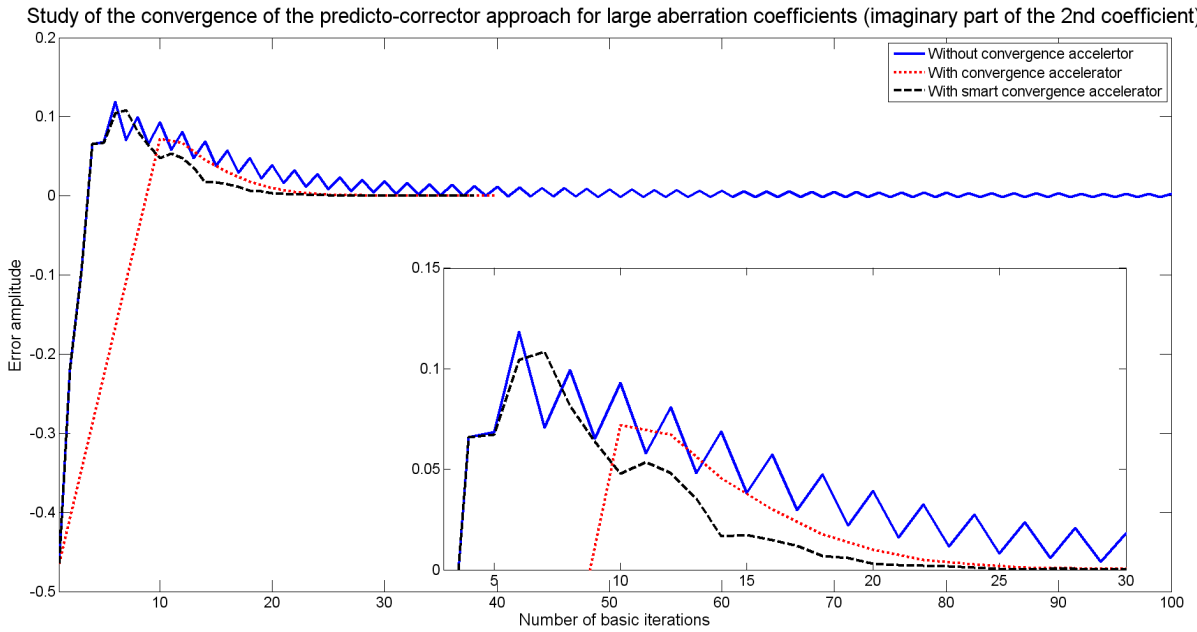


Figure 6.11: Convergence of the error on the imaginary part of the second coefficient in a critical case of very low convergence. The blue curve represents the classical convergence of the process, while the red dotted curve presents an accelerated process for the convergence. The process represented by the dotted curve is based on extra-computation from the "basic" iterations that compose the classical approach. The black dashed curve shows the result of the smart process that consists in a mix between the two other methods. This third method is even faster than the second one as the accelerating procedure can be triggered sooner.

Both approaches are blended in order to get a "smart" iterative process that damps the oscillation in case they would appear. When the aberrations are small enough so that the iterative process is not oscillating, the classical approach is used. On the other hand, as soon as oscillations appear in three successive iterations, the damping process is engaged (fig. 6.11 and fig. 6.12, black dashed curves). As far as the computation time is concerned, this approach is even faster than the previous attempt. Taking the same example as previously and applying the smart accelerator, the process converges in about four minutes.

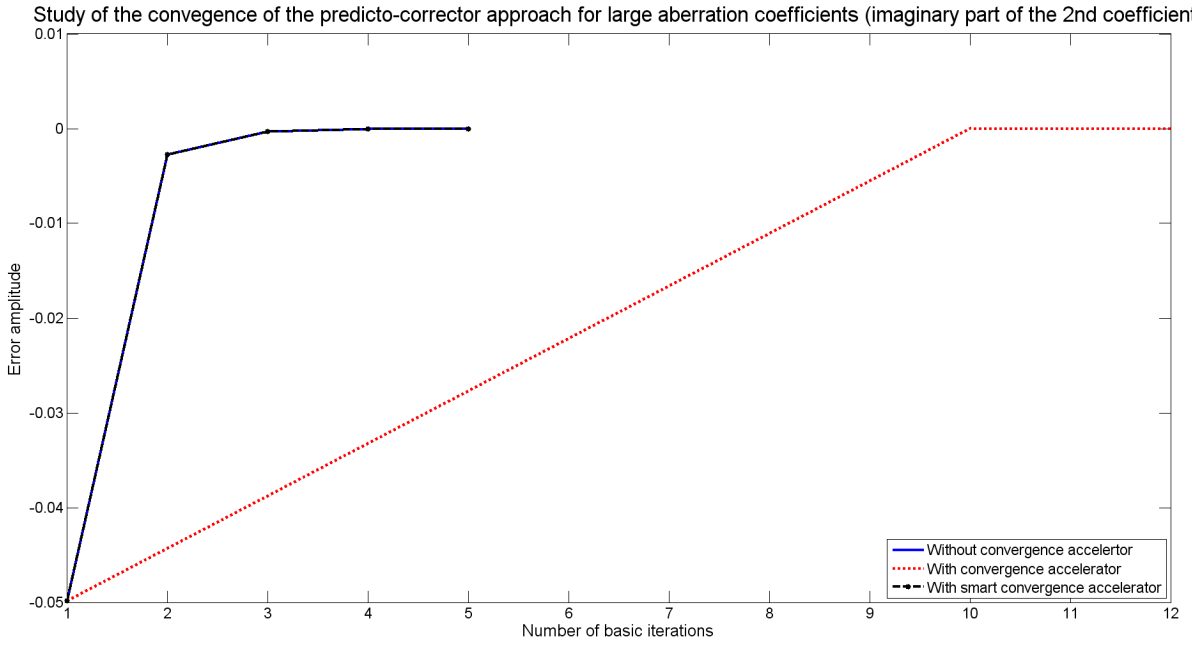


Figure 6.12: Comparison of the same methods as in fig. 6.11. However, this case is not critical as the coefficients corresponding to the input images are quite low (~ 0.1). The classical iterative process is faster than the accelerated one. The acceleration is useful only when the convergence is oscillating. The smart process is thus very interesting as it can discriminate between the two cases. It is illustrated by the black dashed curve, superimposed over the blue solid curve.

The use of this smart process allows to slightly increase the limiting aberrations that can be retrieved, which means that cases that are critical and diverging could converge when the oscillations are damped as the errors do not propagate from one iteration to another.

Now that we know the limitation of the predictor-corrector convergence, it is interesting to study the effect of the number of Zernike coefficients to be retrieved on the convergence. We present hereafter the evolution of the number of iterations required to reach the convergence criterion and the duration of these iterations, as a function of the number of aberrations used.

To conduct this study, we used a realistic power spectral density (PSD) for the aberration coefficients (Roddier 1981). The distribution is based on the Kolomorov atmospheric turbulence model

$$\sigma^2 = (n+1)^{(-11/3)} \cdot \left(\frac{D}{r_0}\right)^{(5/3)} \tag{6.37}$$

where n is the order of the Zernike coefficient, D is the diameter of the pupil and r_0 is the Fried parameter of the turbulence cell given by

$$\frac{\lambda}{r_0} = \theta_{\text{seeing}} \tag{6.38}$$

where λ is the wavelength and θ_{seeing} is the average seeing. In the case of the ILMT, we estimate a seeing of one arc-second, a wavelength of 633nm and a diameter of 4m. The corresponding PSD is shown in fig. 6.13 where it is compared to a $1/n^2$ PSD.

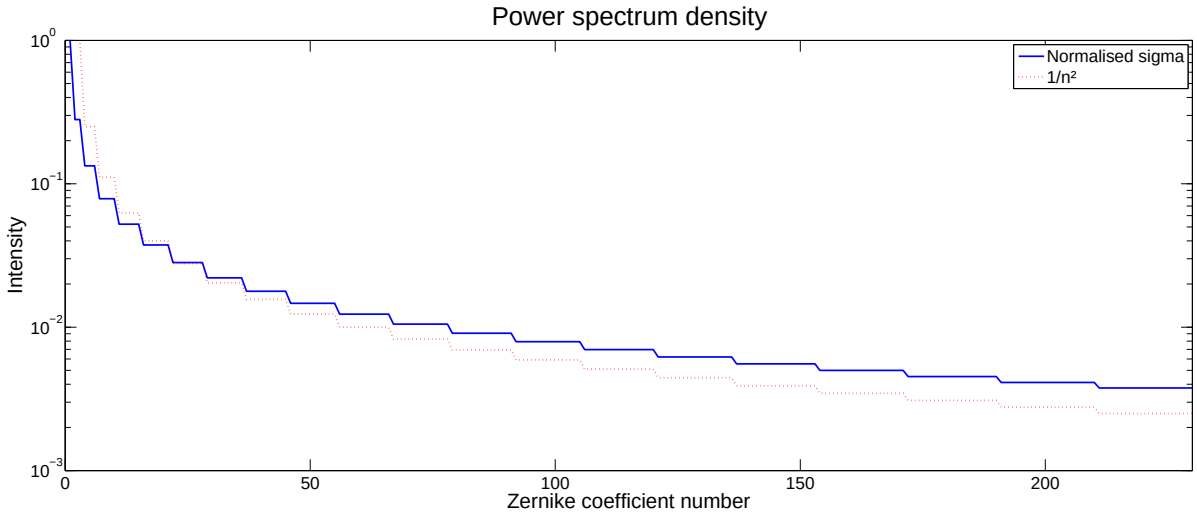


Figure 6.13: Power spectral density used to define realistic Zernike coefficients that will be used to test the convergence of the predictor-corrector method. The blue curve corresponds to $\sigma(n)$ as computed with equation 6.37, the red dotted curve presents a $1/n^2$ distribution.

The imaginary parts of the coefficients, that will be used to test the iterative process, will follow this PSD. Their real part will be set to zero as no amplitude phenomenon should occur. Only the real part of the first coefficient (β_0^0) is set to one while its imaginary part is set to zero as the piston term is not useful for a single aperture. The following tests will be conducted using the same number of aberration coefficients to generate the images and to be retrieved in order to ensure that no parasitic effects will appear.

The Strehl ratio of the point spread function computed from a given number of Zernike coefficients is illustrated in fig. 6.14. The use of all coefficients leads to a PSF with a Strehl ratio about 90%. The same tests have been conducted with coefficients that follow the same PSD but that are 1.5 and 2 times larger and which respectively correspond to a Strehl ratio of about 80% and 70%. The results will be presented hereafter.

The first test consists in determining the number of iterations that are required to reach the convergence criterion as a function of the numbers of aberrations it contains. The criterion that we used, corresponds to a maximum difference between the results of two successive iterations that is lower than 10^{-4}, which means that the largest difference between the β_n^m of two successive iterations is lower than 10^{-4} for every β_n^m. The results are presented in fig. 6.15.

As expected, the number of iterations that are required to converge increases with the number of coefficients that have to be retrieved. However, for a small number of coefficients (less than 50) the number of iterations is almost constant. When more than 60 aberrations have to be retrieved, the increase of the number of iterations is faster than the increase of the number of Zernike coefficients.

Moreover, the number of iterations increases as the Strehl ratio decreases, which is easily explained as the lower order coefficients are more important in case of a lower Strehl. This means that the higher order aberrations are masked by the low order ones. It is also interesting to notice that the retrieval process converges for any number of coefficients (less than 231) in the case of a Strehl ratio of about 90%. In the 70% and 80% cases, the process does not converge

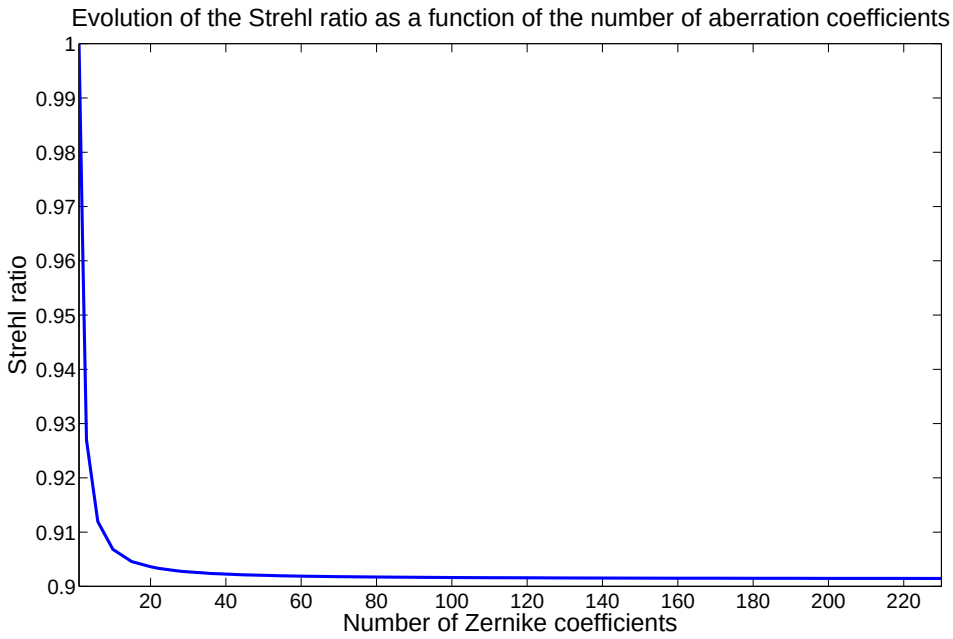


Figure 6.14: Strehl ratio of the point spread function as a function of the number of aberration coefficients used to create it. The set of aberrations is determined by the PSD defined in equation 6.37. The PSF computed using all the coefficients presents a Strehl ratio larger than 90%.

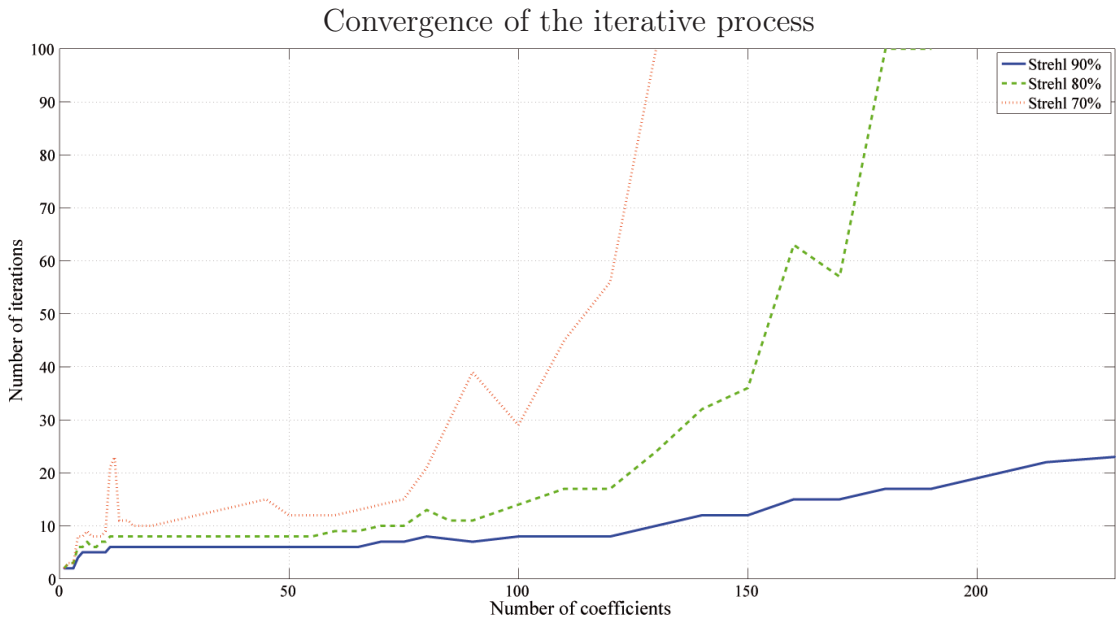


Figure 6.15: Number of iterations required to reach the convergence criterion as a function of the number of aberration coefficients that have to be retrieved. The three curves correspond to three sets of input coefficients, that respect the PSD presented earlier such that the Strehl ratio is 90% for the solid blue curve, 80% for the dash-dotted green curve and 70% for the dotted red curve. Two behaviors can be distinguished. If less than about 50 coefficients are considered, the number of iterations required is mostly constant. In case more than about 60 coefficients have to be retrieved, the number of iterations increases more than linearly.

beyond a given number of unknown coefficients. This is due to the same reason as the increase of the number of iterations required, as the quality of the PSF decreases, the higher order Zernike coefficients get lost, and the process cannot converge.

Another phenomenon that appears is the presence of bumps in the curves of fig. 6.15. This is due to the fact that the number of coefficients selected does not correspond to complete Zernike modes (n parameter). Indeed, each degree n is composed of several m orders. A degree is said to be complete when all orders that correspond to this degree are used in the retrieval (all possible values of m). The relation between the complete degree and the number of aberrations is given in table 6.1.

Degree	0	1	2	3	4	5	6
Number of aberrations in the degree	1	2	3	4	5	6	7
Cumulative number of aberrations	1	3	6	10	15	21	28
Degree	7	8	9	10	11	12	13
Number of aberrations in the degree	8	9	10	11	12	13	14
Cumulative number of aberrations	36	45	55	66	78	91	105
Degree	14	15	16	17	18	19	20
Number of aberrations in the degree	15	16	17	18	19	20	21
Cumulative number of aberrations	120	136	153	171	190	210	231

Table 6.1: Relation between the degree of the Zernike polynomial and the number of aberrations.

Once the degrees are not complete, the convergence is slower because some terms are missing to optimize the computation. Both complete and incomplete degrees are compared in fig. 6.16 for the case of a Strehl of 80%, the curve corresponding to the complete modes does not present any bump anymore. As far as the 70% Strehl curve is concerned, the bumps are still present even if they are smaller, they are due to the coupling between modes that are present and other that cannot be retrieved. This will be investigated later.

Fig. 6.17 shows the number of iterations required to converge as in fig. 6.15 but the numbers of Zernike coefficients used correspond to complete degrees. Once again, one can see that the bumps in the curves that correspond to Strehl ratios higher than 70% have almost disappeared. The curve at 77% of Strehl ratio that has been added corresponds to the same coefficients as those of the curve at 70% except for the tip-tilt coefficients that are those of the 80% curve.

The bumps that are still present in the 70% curve (red-dotted curve) reveal a coupling between the Zernike degrees (modes 4, 8 and 12 in this case). The use of Karhunen-Loeve polynomials could probably solve this type of coupling problem as they consist in collecting together the Zernike polynomials that are coupled. The Karhunen-Loeve basis is completely decoupled and can be found from the singular value decomposition (SVD) of the Zernike matrix. However, in the Zernike basis, the coupling depends on the PSD of the aberrations, which makes it difficult to convert from the Zernike basis to the Karhunen-Loeve basis in practical cases. More information about the Karhunen-Loeve approach can be found in Roddier et al. (1990) in the case of an atmospheric PSD.

Finally, it also comes out of our tests that the iteration duration is directly proportional to the number of coefficients to retrieve.

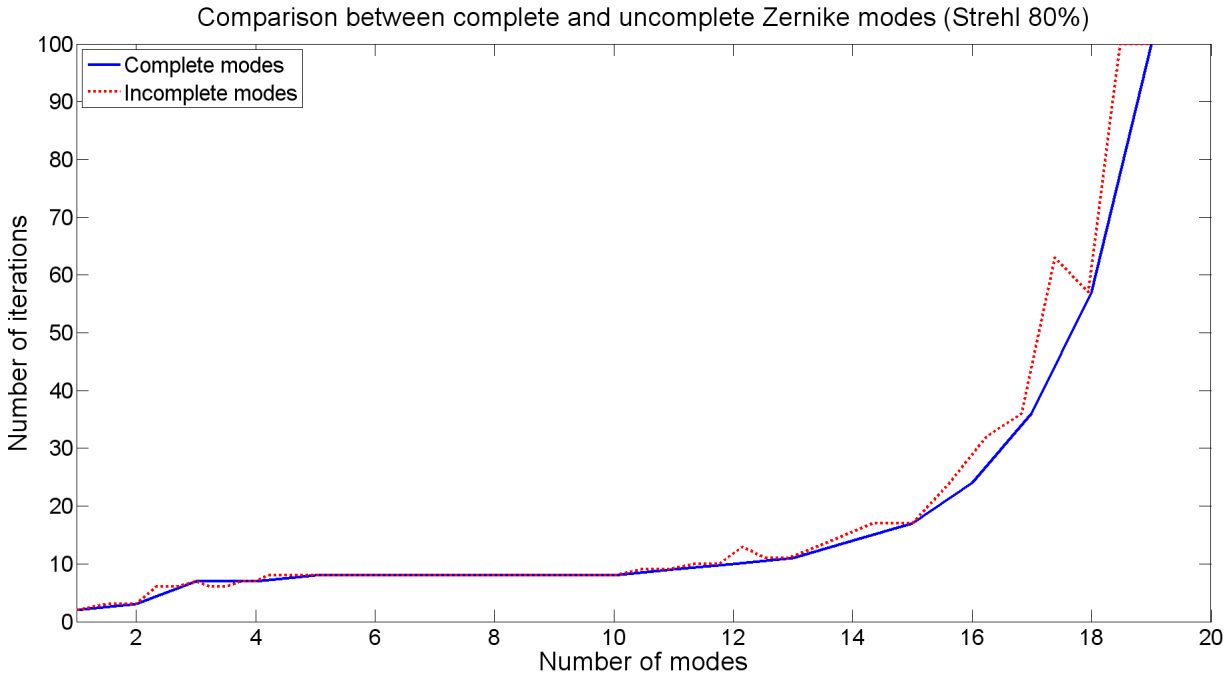


Figure 6.16: Comparison of the number of iterations required to reach the convergence criterion for a number of Zernike coefficients corresponding to complete (solid blue curve) and incomplete (dotted red curve) degrees. The resulting curve is smoother when the number of coefficients is such that the degrees are complete.

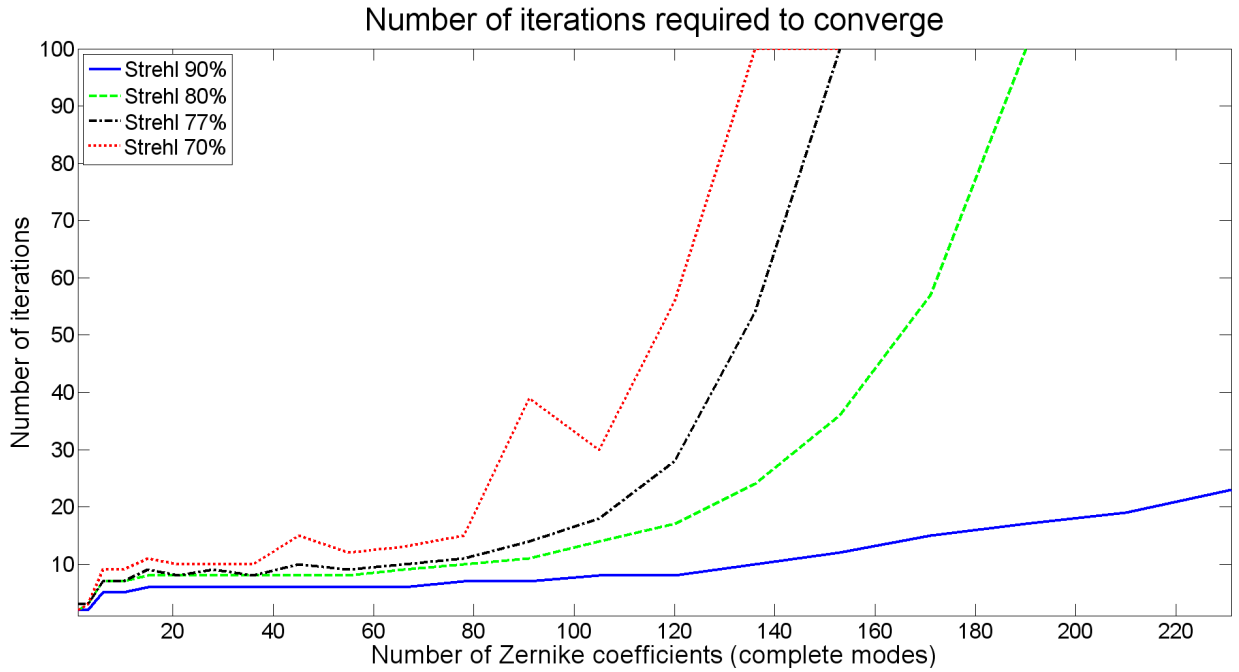


Figure 6.17: Study of the convergence as a function of the Strehl ratio and the number of Zernike degrees. The number of coefficients used to generate the PSF which have to be retrieved exactly corresponds to complete modes. Results corresponding to a PSF with 77% of Strehl have been added (dash-dotted black curve). The coefficients used to generate this PSF are the same as in the 70% case except for the tip-tilt aberrations that are the same as in the 80% case.

6.3.2 Imperfect cases of retrieval

So far we have studied the retrieval in perfect cases with PSFs that had been simulated with a known number of aberration coefficients and without any noise. The retrieval thus uses perfect images and exactly computes the same number of coefficients as the one that had been used to generate the PSF. In such perfect cases the accuracy on the retrieved coefficients could theoretically be very high as long as there is a sufficient number of iterations.

However, practical cases imply images that contain an infinite number of Zernike coefficients and that are disturbed with noise. It is thus very important to quantify these effects. Does neglecting the very small aberrations (represented by high order Zernike coefficients) influence the retrieval capability ? How does the noise in the image degrade the accuracy of the retrieved aberrations ? These issues are explored hereafter with two types of tests. The first one consists in retrieving a number of aberrations that is different from the one used to generate the PSF, which will show the effect of dropping the end of the series expansion in the retrieval. The second type of tests consists in adding some noise with variable intensity in the images. In this way, we should be able to characterize the response of the retrieval method in presence of perturbations in the images.

Study of the effect of an incomplete retrieval

The present analysis aims at studying the effect of an incomplete retrieval. As the input images are generated numerically, they are composed of a limited, but known, number of Zernike coefficients. On the other hand, real images, contain an infinity of Zernike polynomials, and it is thus impossible to retrieve all of them. That is what we call an incomplete retrieval, the number of aberration coefficients that is computed from the PSF is smaller than the number of coefficients in the images.

The retrieval accuracy is expected to be lower in such a case, since the aberrations that are characterized by the dropped coefficients will disturb the coefficients present in the retrieval computation.

Fig. 6.18 presents the results of the analysis of a PSF that is composed of 36 Zernike aberration coefficients, corresponding to polynomials of degree 0 to 7. This is a general low order case that is commonly used in surface testing (Zygo). The imaginary part of the β coefficients follows the PSD presented in fig. 6.13 and they bring the Strehl ratio of the PSF to 90%. The real part is set to zero except for $\beta_0^0 = 1$. Only the imaginary part of the output is considered.

As expected, the error on the coefficients increases when the retrieval is not complete. However, the error is not uniformly distributed over all the coefficients. In fig. 6.18, the degrees zero and one seem quite accurate, the second degree is quite accurate only when the fourth degree is not present or when it is accurate enough. The same coupling between degrees three and five appears even if the effect is less pronounced. It seems in fact that adding a new degree in the retrieval disturbs the accuracy of the degree of the same parity that is just below the new one.

The same type of analysis has been conducted with a set of PSFs generated from 231 Zernike coefficients, following the same PSD and giving about the same Strehl ratio. The errors, presented in fig. 6.18, represent an average of the relative error corresponding to triplets of degrees between 0 and 20.

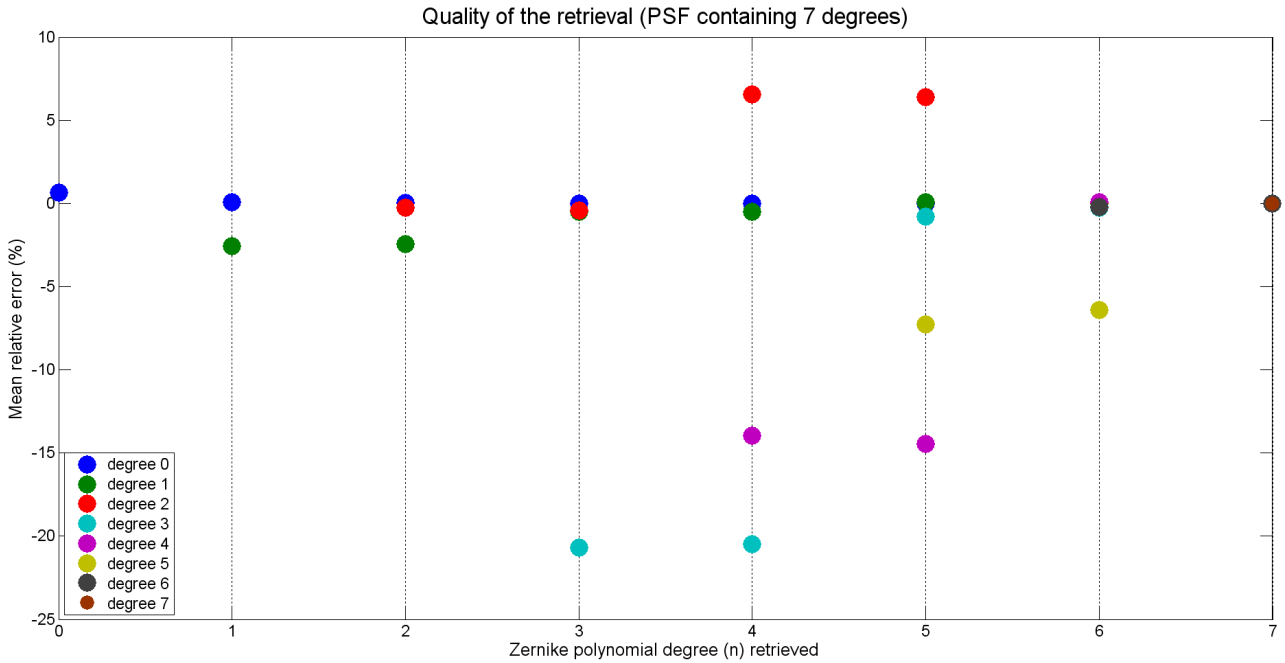


Figure 6.18: Mean relative error for each degree of the Zernike polynomials as a function of the number of complete degrees retrieved. The input PSFs were generated using 36 aberrations (8 Zernike degrees) that follow the PSD presented in fig. 6.13 ; they are characterized by a Strehl ratio of 90%. Coupling between Zernike degrees are clearly visible (degrees 2-4 and 3-5).

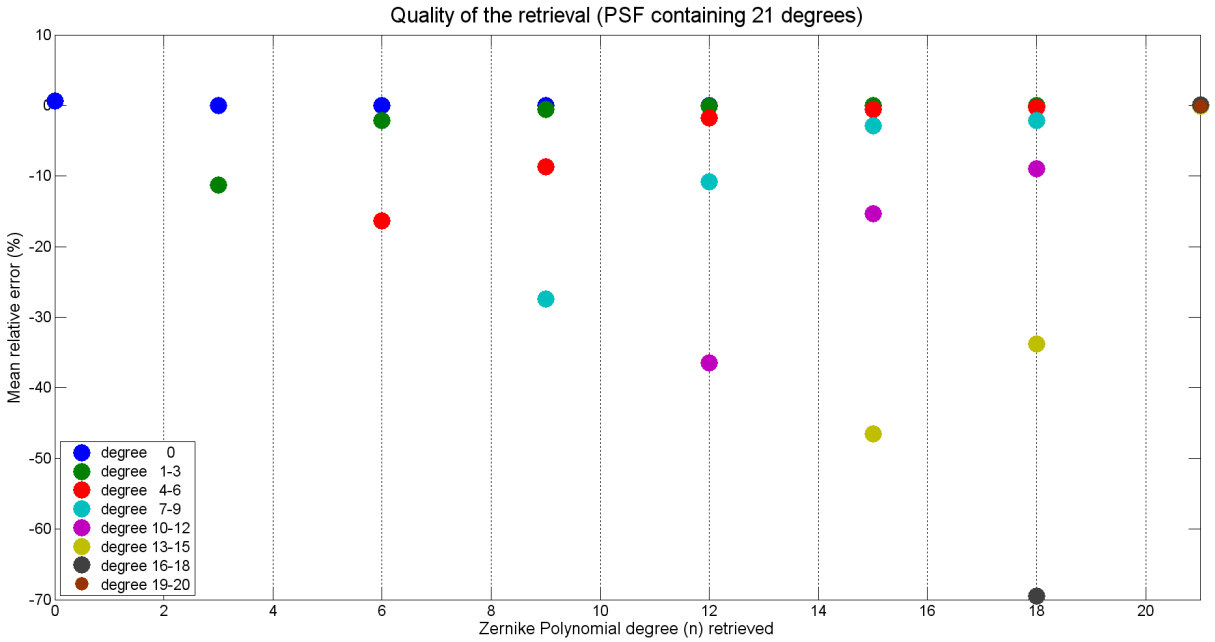


Figure 6.19: Mean relative error for every three degrees of the Zernike polynomials as a function of the number of complete degrees retrieved. The input PSFs were generated using 231 aberrations (21 Zernike degrees) that follow the same PSD. The Strehl ratio is still 90%. The relative error corresponding to a triplet of degrees decreases as the number of degrees to retrieve increases. When all the degrees have to be retrieved, the relative error falls to zero.

It comes out of fig. 6.19 that the accuracy on the retrieved coefficients corresponding to a particular triplet of degrees increases as the number of degrees used for retrieval increases. It is then obvious that using higher and higher degrees increases the accuracy on the retrieved lower degrees. Moreover, the number of high degrees, required to get a given accuracy on a particular lower degree, increases when this one gets higher. As an example, let us consider the first degree. The corresponding error falls below 2% when the second degree is present in the computation. On the other hand, to get the same accuracy on the second degree, the third and fourth degrees are required in the retrieval computation.

As a conclusion, when a retrieval cannot be complete, the accuracy of the measured aberrations increases when the number of aberrations accounted for in the process increases. Using as many aberrations as possible seems thus to be the best approach.

Study of the effect of the noise on the retrieval process

After having investigated the impact of the use of incomplete sets of aberrations on the retrieval process, we will now explore the photometric noise issue. This will be performed by artificially adding some noise to the images used in the previous completion tests.

This noise is composed of a 6 electron gaussian readout noise (this value is given for a single exposure; if the image is composed of several acquisitions, the 6 electron noise is applied to each of them) and a classical Poisson photon noise. The camera is supposed to have a full well capacity of 100k photo-electrons and a linear regime ranging from zero to 80% of this value. The CCD chip presents 1024 bad pixels randomly distributed over its surface, an offset of 500 photo-electrons has been considered, the flatness of the images has been corrected to 1% rms. These operations on the images have been carried out by Pierre Riaud (Riaud and Hanot 2010).

The PSF that was used for the tests has been considered for several different fluxes, corresponding to different noise levels, since the signal to noise ratio increases when the flux of the image increases. The flux is given in number of photo-electrons over the whole image. Fluxes ranging from 10^6 to 10^9 photo-electrons have been used, a 10^{10} case corresponding to a perfect image, that does not contain any noise, has been added.

An approximated relation between these fluxes and their corresponding signal to noise ratio is given in table 6.2. The cases with fluxes larger than 10^7 correspond to very high signal to noise ratios obtained in laboratory manipulations with the accumulation of many data integrations. Fluxes included between 10^6 and 10^7 corresponds to classical images with reasonable acquisition time and fluxes smaller than 10^6 are related to low signal to noise ratios. In chapter 8, we will apply the retrieval method on images acquired with NACO for which the total flux in those images is $3.37 \cdot 10^6$ electrons.

Fig. 6.20 presents radial profiles of the PSF for several flux values. A lower flux corresponds to a higher noise. When the latter becomes important, an offset appears between the noise-free images and the noisy ones. It is important to correct this offset before processing the images to retrieve their aberrations. This can easily be done by subtracting the median value of the image, and the offset corresponding to the noise-free images ($\sim 10^{-4}$) should be added in order to get a better accuracy on the retrieval. Moreover, because of the subtraction of the offset, negative values can appear in the PSF, which should be set to zero, as the intensity cannot be negative anywhere in the image. The results of this treatment are shown in fig. 6.21, for the same profiles as in fig. 6.20.

In case the noise is low, applying this correction makes the noisy profile fit the noise-free one

Flux	10^{10}	10^9	10^8	10^7	10^6
S/N (peak)	∞	NA	526	137	15,3
S/N (ring 1)	∞	NA	154	41	5,3
S/N (ring2)	∞	NA	40	11	1,75

Table 6.2: Relation between the fluxes and the signal to noise ratios at different locations of the PSF. "Peak" corresponds to the central points of the PSF, it is the same S/N concept as in astronomy. "Ring 1" corresponds to the S/N of the first ring of the PSF and "ring 2" to the S/N of the second ring. These concepts are more useful as far as aberration retrieval is concerned as they better represent the quality of the images. The case 10^{10} corresponds to pure signal without noise and the S/N values corresponding to 10^9 is so high that it is not relevant.

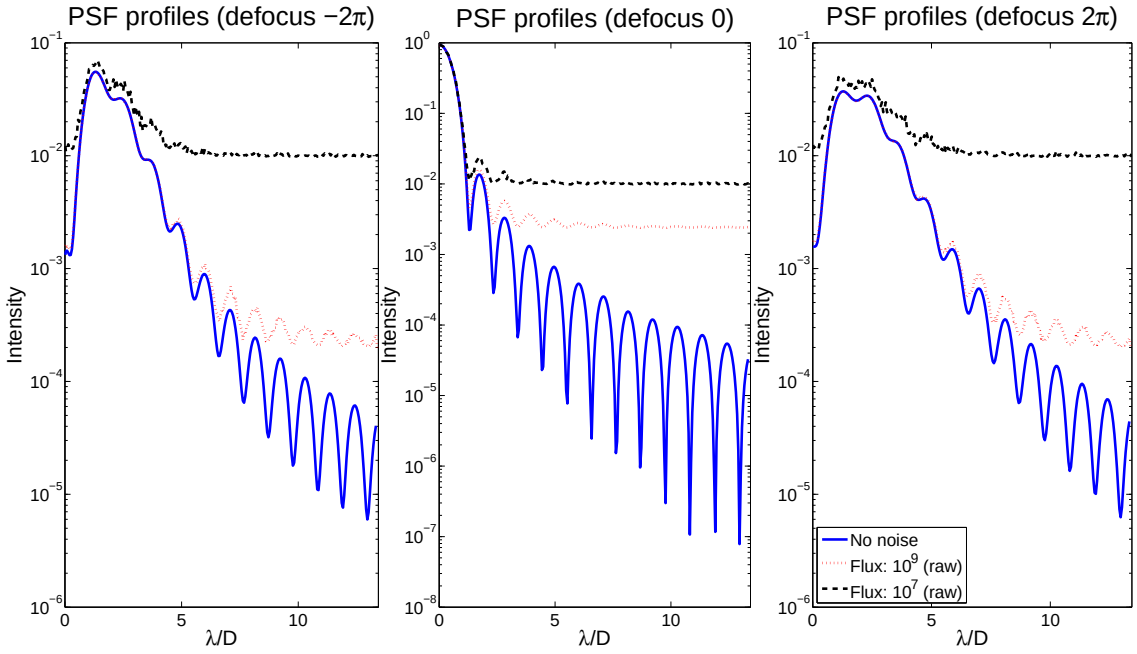


Figure 6.20: PSF profiles for three different fluxes and for three positions around the focal point. These profiles have been computed for monochromatic light (632nm). Flux 10^{10} (solid blue) represents a radial section of the noise-free aberrated images. Flux 10^9 (dotted red) represents the same section of a slightly noisy image and flux 10^7 (dash-dotted black) is that of a highly noisy image. It is clearly visible that the images that contain noise present a global offset that must be taken into account before the retrieval processing.

quite well. When more noise is present, only the first few rings still correspond. The others are lost in the noise perturbations, which means that only the first aberrations can be retrieved accurately. We can thus already expect that fewer aberrations will be retrieved from images that contain more noise and that the lower order aberrations will be more easily retrieved than the high order ones in case some noise is present in the images.

The tests we have performed consist in retrieving a given number of aberrations out of images that are composed of 36 and 231 general Zernike coefficients and that contain a given amount of noise. As for the previous tests, the number of aberrations that we try to compute corresponds to complete degrees. This is relevant since the aberrations that correspond to polynomials of the

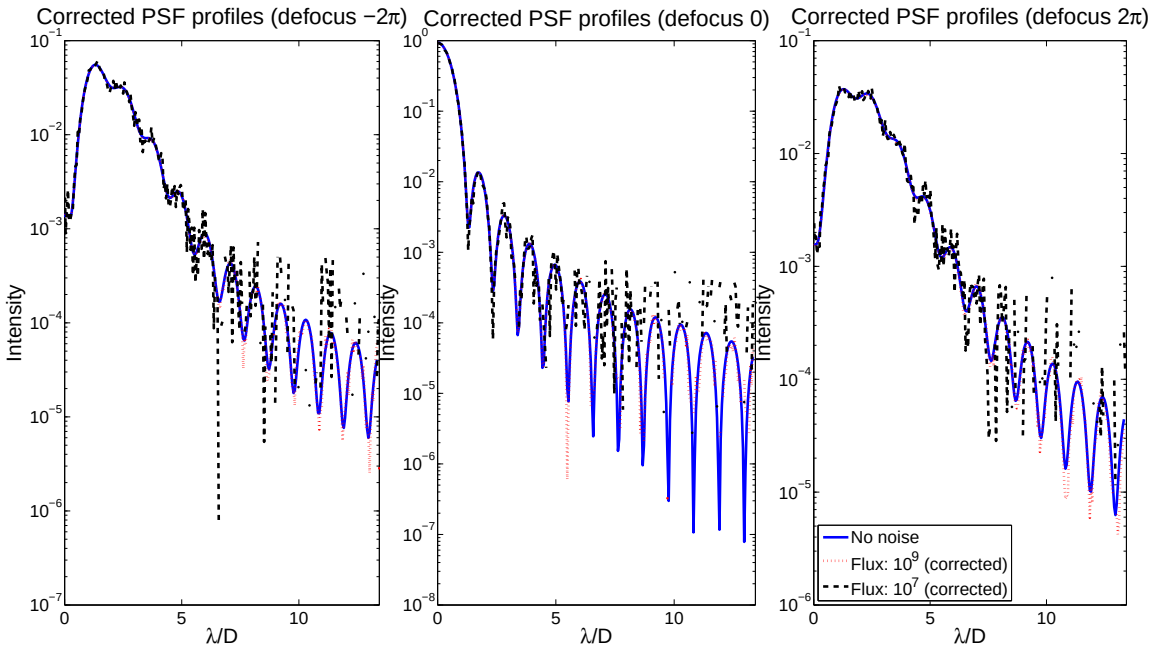


Figure 6.21: Same PSF profiles as in fig. 6.20 where the offset correction has been applied. The low noise curve (Flux 10^9 in dotted red) is almost superimposed over the noise-free curve (Flux 10^{10} in solid blue). For a higher noise level (flux 10^7 in dashed black), the offset correction still improves the profile, but the correction is less accurate. This figure also shows that the outer rings are more affected by the noise than the central peak and the rings. The low order aberrations that correspond to that central part of the PSF will thus be retrieved more easily than the higher orders, corresponding to the outer part of the PSF, which is lost in the noise.

same degree are of the same order of magnitude and should thus be disturbed in the same way by the noise. We have applied the retrieval process to the disturbed images in order to estimate the impact of the noise level for particular numbers of complete degrees. The maximum number of correctly retrieved coefficients is then measured.

A coefficient is said to be retrieved when the absolute difference between the computed value and the real value is lower than $\lambda/100$ rms. The first coefficient beyond that limit defines the end of the range of correctly retrieved aberrations that starts with β_0^0. What we call "the maximum number of retrieved coefficients", in the following, thus corresponds to the number of coefficients in that range. It should be noted that the first coefficient β_0^0 is included in the range even if the corresponding error is larger than the limit. Indeed the retrieved value of this coefficient only depends on the normalization of the input PSF, which can become very difficult when a lot of noise is present in the images. Moreover, an error on the retrieved value for this amplitude aberration does not affect the following values too much, and it is thus interesting to verify whether they can be retrieved even if β_0^0 is not accurate. When the noise is low (flux 10^9 to 10^7) no significative difference was noticed, since the error on β_0^0 is lower than $\lambda/100$ rms. The behavior changes when the noise level is higher (flux 10^6). In this case, the normalization of the peak is not accurate, but the low order aberrations can still be retrieved.

It is also interesting to note that some aberrations outside the so-called "correctly" retrieved range are also accurately computed (error below the $\lambda/100$ rms limit). The results presented here

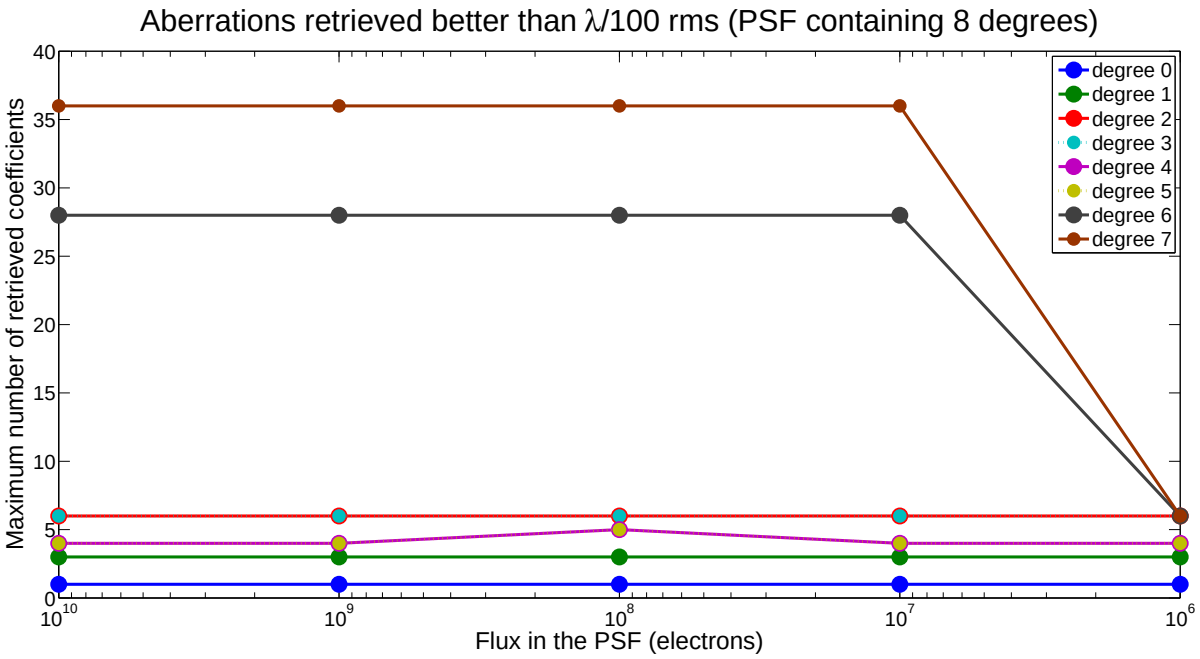


Figure 6.22: Results of the study of the retrieval capabilities using a realistic PSF that is disturbed with some noise. The original PSF is generated from 36 Zernike coefficients that follow the PSD previously presented. Some noise has been added to the PSF for several fluxes. The values given for a flux of 10^{10} photo-electrons correspond to noise-free images. The graphic represents the maximum number of Zernike coefficients that can be retrieved with an accuracy better than $\lambda/100$ rms as a function of the flux of the image. As the flux decreases, the signal to noise ratio decreases. Each curve corresponds to the use of more and more complete degrees for the retrieval. The cases where few coefficients are used in the retrieval process (degrees 0-2) are almost independent of the noise. A retrieval involving higher order aberrations (degrees 6-7) is also quite accurate up to 10^7 photo-electrons of flux. The intermediate cases (3-5) behave differently. They do not seem to be disturbed by the noise, but all the aberrations cannot be retrieved.

are thus pessimistic. However, it is interesting to consider only a full range of retrieved aberrations so that we know which coefficients are correctly computed. It would not be of practical interest to only know the number of aberrations that can be retrieved without knowing which are the correct ones. Considering only the defined range allows to be sure that the first aberrations are correctly computed. The case of the "out of range" coefficients will be discussed later.

Fig. 6.22 shows the results of the tests concerning the 36 aberration PSF. As it was expected from the completion tests, the intermediate degrees cause some trouble. It can be seen from fig. 6.18 that the accuracy on the low order aberrations (degrees 0-2) is quite good when they are used alone, which is also the case in fig. 6.22. Adding the third degree does not improve the results (fig. 6.22) as the error on this degree is important (fig. 6.18). The same remark applies to degrees four and five as their addition involves errors on degrees three to five. Using them thus does not improve the retrieval. In fig. 6.22, they are below degree two.

The retrieval process applied to PSFs containing only 36 aberrations seems to depend very little on the noise, since the number of correctly retrieved coefficients is mostly constant as a function of the noise level excepted for the most disturbed image (Flux 10^6).

However, this 36 aberration PSF case is not very realistic because real images are composed

of an infinite series of Zernike polynomials. Nevertheless, these aberrations are those usually retrieved with the classical surface testing methods (Zygo). Moreover, using fewer polynomials allows faster computation, which is quite useful to quickly test the NZ method.

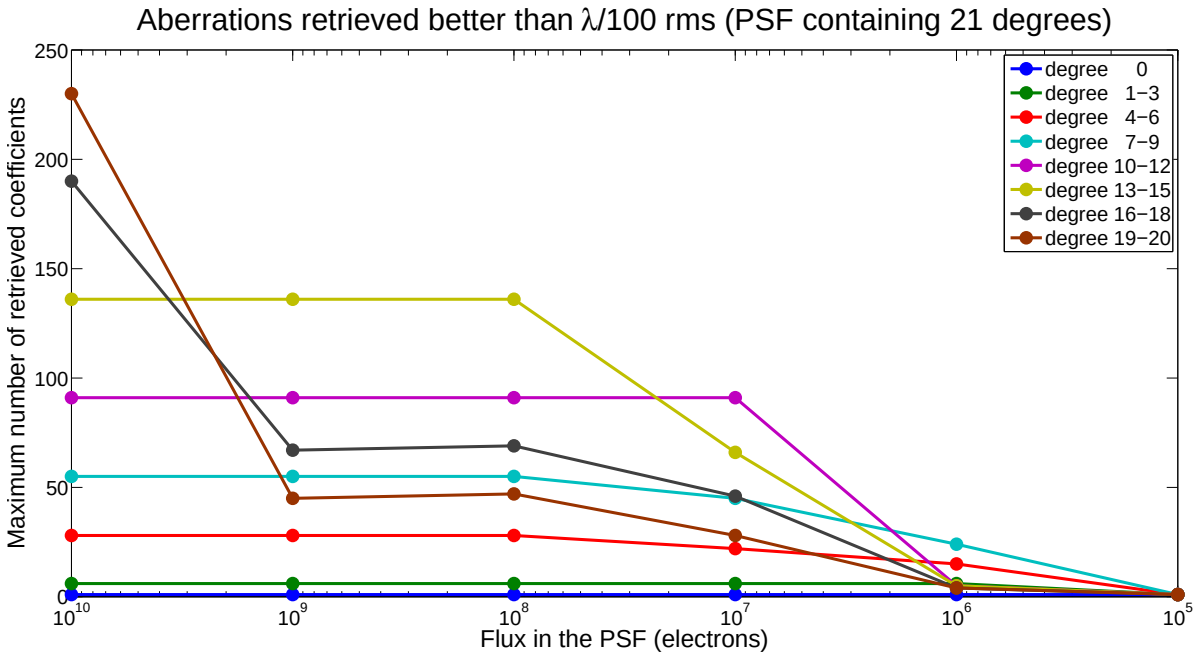


Figure 6.23: The same analysis as in fig. 6.22 has been conducted on a PSF generated from 231 Zernike coefficients. Because of their very low level, the high order coefficients are very sensitive to perturbations since even a very small amount of noise disturbs their retrieval. Working with only the lower order allows to accurately compute more aberrations, even with more noise in the input image.

Fig. 6.23 presents the same kind of results as in fig. 6.22, for the case where the PSFs contain 231 aberrations. Here, an important difference is visible between the case with and without noise. Indeed, without any noise (10^{10}), each coefficient computed can be accurately retrieved, whereas even with very low noise (flux 10^9) the higher degrees are lost. This is related to the location of these aberrations on the PSF, since they correspond to the outer rings, as soon as some noise is present in the images, these rings are highly disturbed and the aberrations cannot be retrieved anymore. The errors on the values of these aberrations induce other errors on the lower order aberrations, which explains why trying to compute too many aberrations is not recommended.

As far as the intermediate order aberrations are concerned, they tumble down one after another. When the noise level increases, their corresponding rings gradually become disturbed and they cannot be accurately computed.

These observations tend to show that an optimal number of coefficients should be used for the retrieval process in order to maximize the number of aberrations that can be correctly computed. This optimal value depends on the noise level, which can be seen as the highest number of aberrations that corresponds to the last visible ring.

Obviously, lowering the accuracy limit to $\lambda/10$ rms allows to retrieve more aberrations even with much more noise, which can be useful in case of high intensity aberrations. In this case, up to 55 aberration coefficients can be computed even with a flux as low as 10^5 photo-electrons.

As a conclusion, the noisy image tests show that when few low order aberrations are present in the images, the retrieval process does not really depend on the noise as the aberrations are related to the central part of the PSF, that is not much affected. However, this is not very realistic as practical PSFs will be composed of an infinity of aberrations. When higher order aberrations are present in the images, their corresponding rings quickly disappear in the noise and these aberrations cannot be accurately retrieved when the noise is too important.

Beyond a given noise level (flux 10^6, 10^5), the perturbations also affect the central peak and the retrieved first aberration (amplitude) can be inaccurate. This is due to a bad normalization of the PSF but does not affect the few following aberrations (2-6) as soon as the error on β_0^0 is not too important. Moreover, the defocused images get disturbed more quickly than the focused one as the energy is spread out over more pixels.

In order to avoid the loss of too much information in the outer rings, it is important to increase the signal to noise ratio, which can be achieved by modifying the sampling of the PSFs. Indeed, those used during our tests were very much over sampled, which is good for the retrieval process, but as the intensity is spread out over many pixels, the signal to noise ratio is quite low. The effect of the PSF sampling will be investigated in section 6.3.3.

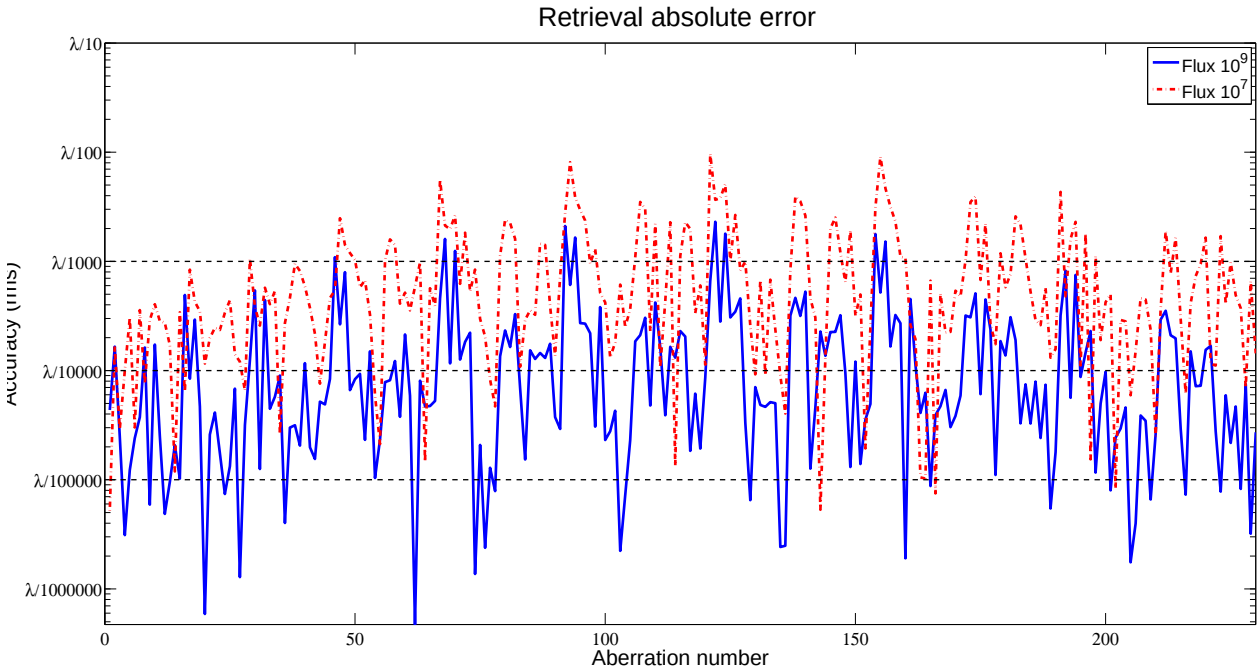


Figure 6.24: Absolute error on the retrieved aberrations using 231 coefficients for several noise levels. Structures are clearly visible. The limits represented on the graph correspond to good optics ($\lambda/100$ rms), coronagrpahy ($\lambda/1000$ rms) and to the VIRGO interferometer ($\lambda/10000$ rms), respectively. The noise free curve (not presented in this figure) shows a theoretical upper limit around $\lambda/10000$ rms.

As it has been previously mentioned, some aberrations outside the range of "correctly retrieved aberration" can also be retrieved with a good accuracy. Fig. 6.24 presents the absolute errors between the computed coefficients and the real injected values. Many "out of range" aberrations have errors below the $\lambda/100$ rms limit, which tends to show that our results are quite pessimistic.

These plots clearly show that the Nijboer-Zernike aberration measurement approach can be very accurate. The $\lambda/100$ rms limit corresponds to good optics, the $\lambda/1000$ rms one is the order of magnitude of what is encountered for coronagraphy (Riaud et al. 2005) and the $\lambda/10000$ rms limit is reached with the VIRGO gravitational wave interferometer (Mackowski et al. 1999).

Moreover, structures appear in the curves, that tend to indicate that some aberrations can be retrieved more accurately than others. The latter could be investigated in order to improve the retrieval beyond the current level.

From the noise and completeness presented here, one can deduce that an optimal number of aberrations could be determined in every cases. This optimum would balance the effects of noise (disturbs the high level aberrations) and the effect of the completeness (disturbs the low level aberrations because of missing high orders.) We suggest here a convenient way of finding this optimum. When the retrieval is successively performed with more and more aberrations, their value should stabilize since the retrieval is more and more complete. However, when the noise limit is reached (one tries to compute lost aberrations), the values get disturbed. Studying the evolution of the aberration values as a function of the number of aberrations computed, allows to determine this limit, which corresponds to the optimal searched value.

6.3.3 Impact of the image sampling on the retrieval process

The previous tests showed that, as expected, the number of aberrations than can be retrieved from a set of images is limited by the signal to noise ratio of these images. We then suggested that, all other conditions being the same, images with a smaller sampling (less pixels per λ/D) should have a higher signal to noise ratio. Indeed, if the detector receives a given amount of flux, the signal on each pixel will be lower if it is divided over many pixels than if it is concentrated on only a few of them. Contrarily, the noise is due to several sources that are not necessarily related to the signal (dark current, readout noise), which implies that when the energy is spread over a large number of pixels, the signal decreases faster than the noise and the signal to noise ratio decreases. The signal to noise is thus higher when the images are recorded on fewer pixels and we think that using a small sampling should allow to retrieve more aberrations.

However, the sampling of the images should respect the Nyquist-Shannon sampling theorem which states that any periodic function is completely determined if at least two points per period are known. As far as PSFs are concerned, the minimum sampling value corresponds to two pixels per λ/D. Moreover, let us remind that the NZ retrieval method presented here consists in fitting the modes of the images (sine and cosine projections) with aberration templates, one could thus think that a high sampling (more pixels per λ/D) would be helpful for this fitting.

The test presented in this section aims at determining the effect of the sampling on the retrieval process. We have generated two sets of images respectively with two and four pixels per λ/D, all other parameters being the same. In both cases, the Nyquist-Shannon theorem is respected and the information is thus correctly sampled. The 231 coefficients used to compute those sets of images follow the same PSD as those used for the tests presented earlier, and result in a Strehl ratio of 78%.

So far, the simulation (the PSF generator and the retrieval process) does not take energy considerations into account, which means that not matter what the sampling is (provided that it is Nyquist), there is no spreading of the energy over the pixels. The total energy is just increased with the sampling. We have thus decided to normalize both sets with the total flux contained in the two pixel images.

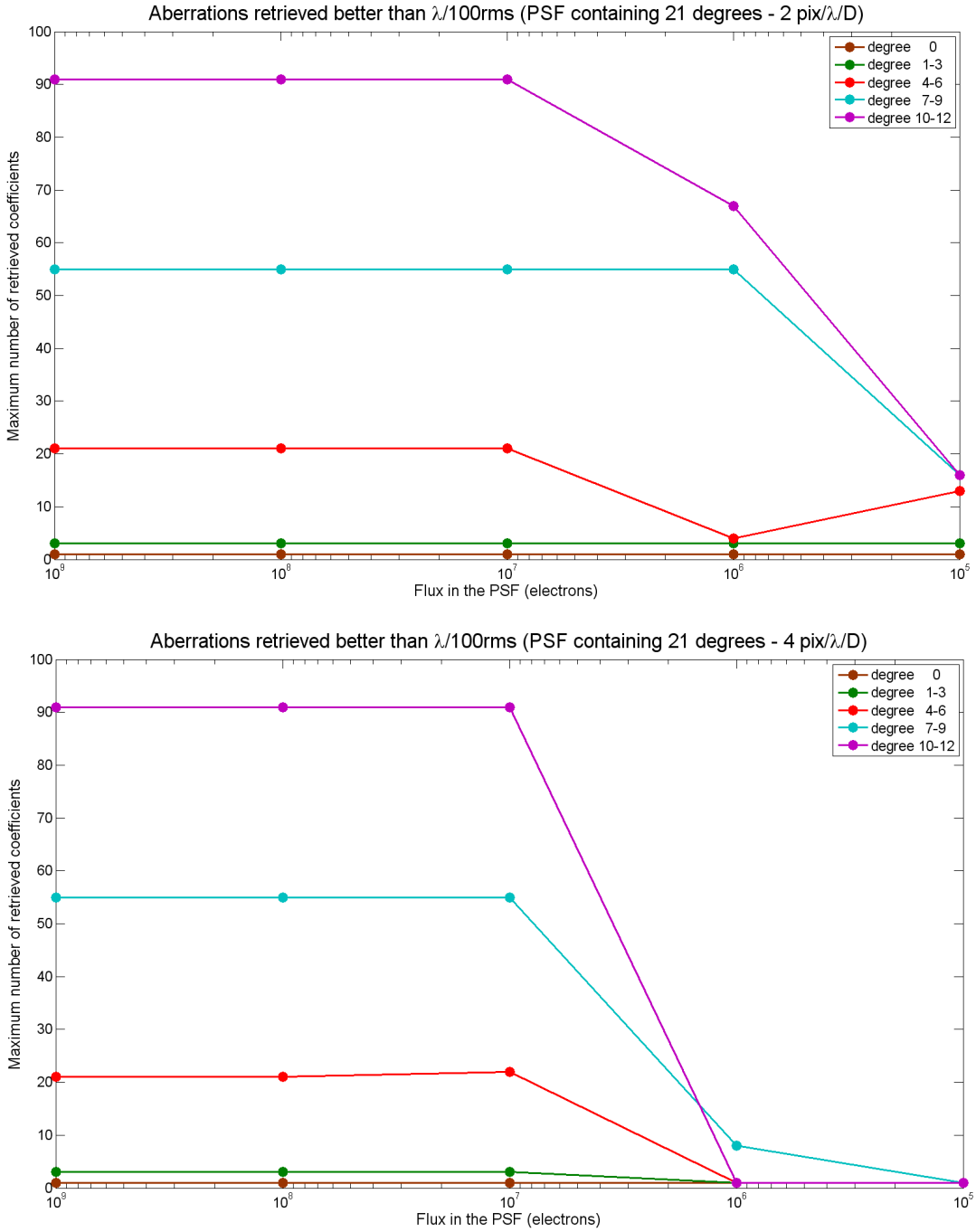


Figure 6.25: The same analysis as in figs. 6.22 and 6.23 has been conducted on two sets of PSFs generated from 231 Zernike coefficients, but using different sampling values (top: 2 pixels per λ/D, bottom: 4 pixels per λ/D) for the same total fux received. In case of small noise level (fluxes $10^9, 10^8, 10^7$), both image sets give the same results. For higher noise level (fluxes $10^6, 10^5$) the lower sampling images (2 pixels) allow a better retrieval than the higher sampling images, as it was expected. This is due to the larger spreading of the energy in the 4 pixels case leading to a smaller signal to noise ratio. On the contrary, the better sampling does not seem to improve the retrieval process. It should be noticed in the top figure, the point that is too low on the "degree 4-6" (flux 10^6) is limited by the fifth aberration at $\lambda/80$, the following aberrations fit the $\lambda/100$ criterion as in the 10^7 case.

Noise has been added to the normalized images in the same way as for the "noise" tests presented in the previous section. We should precise here that the aberrations used to generate the images are not exactly the same as for the previous test so the results of both tests should not be compared. As a consequence of the lower Strehl ratio, the retrieval process does not converge (or very slowly) with 136 aberrations or more. This phenomenon had already been observed when the convergence was tested in section 6.3.1 (fig. 6.17). Hence we applied the retrieval process only for 1,10,28,55, and 91 coefficients and determined the number of consecutive aberrations retrieved with a precision better than $\lambda/100$, as in the case of the noise test.

Fig. 6.25 shows the results of these tests. There is almost no difference between them for low noise levels (fluxes: $10^9, 10^8, 10^7$). This is not true anymore in the high noise regime (fluxes: $10^6, 10^5$) where the low sampling images allow to retrieve more aberrations than the high sampling images. This behavior was expected from signal to noise considerations previously mentioned.

On the contrary, a better sampling does not seem to be helpful for the retrieval process. Indeed, the same tests had been performed without normalizing the images, which means that the signal to noise ratios were the same in both cases. There was no difference at all between the results of the two sets of images, as if they were exactly the same.

As a conclusion, we have observed that decreasing the sampling is equivalent to concentrate the total flux received in fewer pixels and thus to increase the signal to noise ratio. More aberrations can thus be retrieved with a lower sampling. On the other hand, as long as the sampling is sufficient to respect the Nyquist-Shannon theorem, increasing it above this limit does not influence the retrieval process capabilities. Hence, the optimal number of retrieved aberrations is expected for a sampling very near to and higher than two pixels per λ/D.

6.3.4 Conclusions

In this section we have studied the capabilities of the aberration retrieval process and we have tested the impact of several parameters.

First of all we considered perfect cases of retrieval and just studied the speed of convergence. Obviously, we determined that the process required more iterations to converge when the number of computed aberrations is larger. The number of iterations also increases when the Strehl ratio of the image analyzed decreases. With low Strehl ratios, it is also possible that the process does not converge with any number of coefficients. We can hence conclude that when the quality of the PSF decreases, the higher order aberrations are lost and the process converge more slowly.

We have then tested the effect of the non-completeness of the retrieval. The completeness is the fact of computing the same number of aberrations during the retrieval as those included in the PSFs. It comes out of this study that using higher and higher degrees increases the accuracy on the retrieved lower degrees. Hence one should use the same number of coefficients for the retrieval as the one contained in the PSFs. This is not possible with actual PSFs which are composed of an infinity of aberrations. In this case, as many aberrations as possible should be computed.

In case noise would disturb the images, the previous affirmation is not true anymore. The tests we performed show that when high order aberrations are present in the images, their corresponding rings of PSF quickly disappear in the noise and they cannot be accurately retrieved. The errors on these aberrations are artificially compensated for by the software by adjusting errors on the low order aberrations, which also become inaccurate. It is thus important to

avoid computing the aberrations that are too much affected by noise. Hence, there is a trade off between the noise and completeness requirements which leads to an optimal number of coefficients to compute. This will be tested in chapter 8 in the case of the VLT NACO instrument.

Finally, we showed that the sampling of the images affects the signal to noise ratio and thus the retrieval capability. Decreasing the sampling is equivalent to concentrate the total flux received in fewer pixels. Hence the signal to noise ratio increases and more aberrations can be retrieved. Moreover, provided that the sampling is sufficient to respect the Nyquist-Shannon theorem, it is useless to increase it since this does not improve the retrieval. The optimal sampling is then near to two pixels per λ/D.

Chapter 7

Application 1: Alignment of the ILMT

One of the interesting applications of the Nijboer-Zernike phase retrieval theory consists in the measurement of the aberrations caused by a misalignment of the optical elements that compose a telescope. This is particularly interesting in the case of a Liquid Mirror Telescope that cannot be easily aligned with the usual methods, as they generally require to move the primary mirror.

The alignment of a telescope is very important, since a bad alignment results in bad image quality. A classical telescope is usually aligned by pointing it toward a very bright star (Vega in the Northern hemisphere). In this way, the aberrations can be directly observed in defocused images. Several images are thus taken while moving the different optical elements in order to find their optimal position, that minimizes the aberrations. This position corresponds to the best alignment of the system. However, depending on the system complexity (the number of optical surfaces, their off-axis angle and their size, the image acquisition mode, ...), reaching a good alignment may turn out to be difficult. Indeed, when a system containing many off-axis optical surfaces is misaligned, the image degradation quickly gets important and it becomes difficult to interpret the observed aberrations in order to align the system. As far as liquid mirror telescopes are concerned, the presence of a TDI optical corrector (5 large off-axis lenses) and the TDI acquisition mode, that involves two more degrees of freedom (East-West alignment and TDI drift speed), make it an unusual, complex instrument to align. Moreover, unlike glass mirrors, a liquid mirror can hardly be tested. Indeed, classical null tests (section 5.1) would require an access to the center of curvature of the mirror that is 16m above the mercury pool. The mirror is thus difficult to characterize. Finally, an LMT cannot track celestial objects, it can only look at stars moving in the sky. One thus has to wait for bright stars to cross the field of view of the telescope. In Liège, where the telescope should first be tested, less than one star of magnitude brigther than 15 crosses the field of view every hour. It is thus very difficult to use stellar objects as alignment source targets. Moreover, the observed fields of stars are always different and they are thus difficult to compare.

All these specificities make LMTs more difficult to align than classical telescopes. Nevertheless, LMTs currently in operation have been aligned on the sky, even if it was a time consuming operation. We present hereafter two alternative approaches, using an artificial source, which are more systematic, easily reproducible, highly accurate and fast to use. The first method, based on the 2m LMT model, was developed at the beginning of the present thesis. Even if some parts of this approach were interesting, it presented important issues that made it unusable for practical cases. After having studied the Nijboer-Zernike theory, we have conceived another approach to align the ILMT. It is based on the same principle as the previous one, but some important improvements, involving Nijboer-Zernike phase retrieval, have been included to make it more

simple to implement and to correct for the previous problems.

The same structure is used to describe both approaches. We first present the optical design used to simulate the artificial source that illuminates the liquid mirror. Then, using this scheme, we detail our first misalignment computation method. The latter is based on the comparison between characteristics (images in the first case and aberrations in the second one) measured in the actual case and the expected values of these characteristics computed from a theoretical model. The results of the first approach are very promising, however it presents some important limitations that have to be circumvented before it can be really useful. The Nijboer-Zernike aberration measurement technique is a convenient way of solving these problems, and it is used in the second method. Simulations of this improved approach show that it can be very accurate and that the improvements really solve the issues encountered with the first method.

7.1 The 2m Liquid Mirror Telescope alignment method

The procedure summarized in this section has been developed to align the 2m CSL LMT introduced in chapter 1, and has been extensively described in my master thesis (Magette 2007). It includes an optimization of an optical design aiming at illuminating the primary mirror of the telescope in the best possible way, with an artificial source. We then present a way of efficiently comparing actual and theoretical images, which allows to compute the position of the upper-end unit corresponding to these actual images. The results that can be expected from this technique are obtained from simulations and presented at the end of this section. Even if the results are promising, several drawbacks make this approach hardly usable in practical cases. However, it involves some techniques, as the use of maps, that are very interesting and will be re-used in the improved second method presented in section 7.2.

7.1.1 Optical design

Assuming that the CCD camera and the corrector have been previously aligned on an optical bench, we have determined the alignment requirements of the whole upper end unit (camera and corrector). This step has been realized by tolerancing the positions of these elements and by determining the induced image degradations. It results that an upper-end horizontality aligned better than 1mm in the focal plane and vertically aligned better than 6 arcmin with respect to the liquid mirror optical axis would produce images containing insignificant defects. We thus had to find a way of determining the position of the upper-end better than this required accuracy.

First of all, the tracking unability of the telescope has to be circumvented by using an artificial source. However, the latter cannot be located at infinity, as a star would be. Several optical elements are thus required to simulate a source. A series of optical designs have been tested in a previous work (Magette 2007), and we only remind, hereafter, the main results of this study.

Our first tests were based on a 10cm wide decentered collimated beam that was supposed to simulate a point source at infinity. However, the resulting images where diffraction limited and their degradation due to (even large) misalignments were not measurable. We have thus abandoned this type of design that was indeed not well suited for our purpose. The problem comes from the very limited area of the mirror that is illuminated by the collimated beam. The best solution would consist in using a very large collimated beam (as large as the mirror diameter) but this would require a large collimating lens (several meters), which is not realistic.

Our solution to cope with this problem is to use a diverging beam originating from the center of curvature of the parabola. In this case, the image is also located at the center of curvature. Even if this is only true for a spherical mirror, this property can also be used in the case of a parabolic mirror, but the images will contain spherical aberrations (as described in chapter 5). However, this is not a problem regarding the alignment, since this aberration is centrally-symmetric and a misalignment causes aberrations that break this symmetry.

The spot diagrams of this design provide us with some valuable information. The spot size (and thus the image size) is fairly larger than in the case of a parallel beam, and the changes in the spot shape and size should be measurable. However, this design is useless as the upper-end unit (corrector + detector) is located at the center of curvature of the mirror. Indeed, a telescope requires a detector located at the focal point of its primary mirror, and not at its center of curvature. In this case, the optical path of the light has thus to be increased while the upper-end unit is still located at the focus of the primary mirror. This type of "2f-2f" [24] design, as presented above, is usually used with null lenses in order to test the mirror (chapter 5).

The solution we propose consists in folding the beam with small flat mirrors in such a way that the upper end unit is located at the focus of the primary mirror while the light travels twice the focal length between the source and the mirror (and between the mirror and the detector). Several designs using different numbers of mirrors and different positions have been tested, and the final solution is presented in fig. 7.1. The source is composed of a fiber and two converging lenses, used to control the numerical aperture of the beam. The source is slightly shifted off-axis and oriented perpendicularly to the axis. Using converging lenses allows to locate the virtual origin of the diverging beam around the center of curvature where it is best suited.

The source is oriented perpendicularly to the primary mirror axis, and its vergence is adjusted in such a way that the focus of the system is near to the primary mirror axis. The beam then goes through a small beam splitter cube that will fold the reflected rays. This cube is preceded by a polarizer and followed by a quarter wave plate which rotates the polarization by 90 degrees (twice 45°). The overall configuration is such that the beam coming from the source is transmitted through the cube and the beam coming back from the primary mirror is reflected toward the corrector. Since the beams converge near the cube, the latter can be very small.

The spot diagrams corresponding to this final scheme are presented in the sub-figure of fig. 7.1. The nine spots correspond to point objects at different positions in the field of view of the CCD camera. Let us remind that the CCD has 2048×2048 pixels covering a field of view of 14×14 arcmin. The nine fields presented in the figure cover the full field of view of the camera. They are useful in order to break the symmetries of the system. This will be detailed in the next section.

The results are interesting because the spot size is larger than 10 pixels in radius, which means that they are measurable, and that their changes in shape are also distinguishable. This design should thus help us in aligning the upper-end unit since the images it produces have the appropriate properties. However, using it could be quite complicated as it requires several folding mirrors, that could be difficult to place correctly, even if we tried to simplify at maximum.

[24] We called this family of designs 2f-2f because both the source and the detector are located at a two focal length distance from the primary mirror on its optical axis.

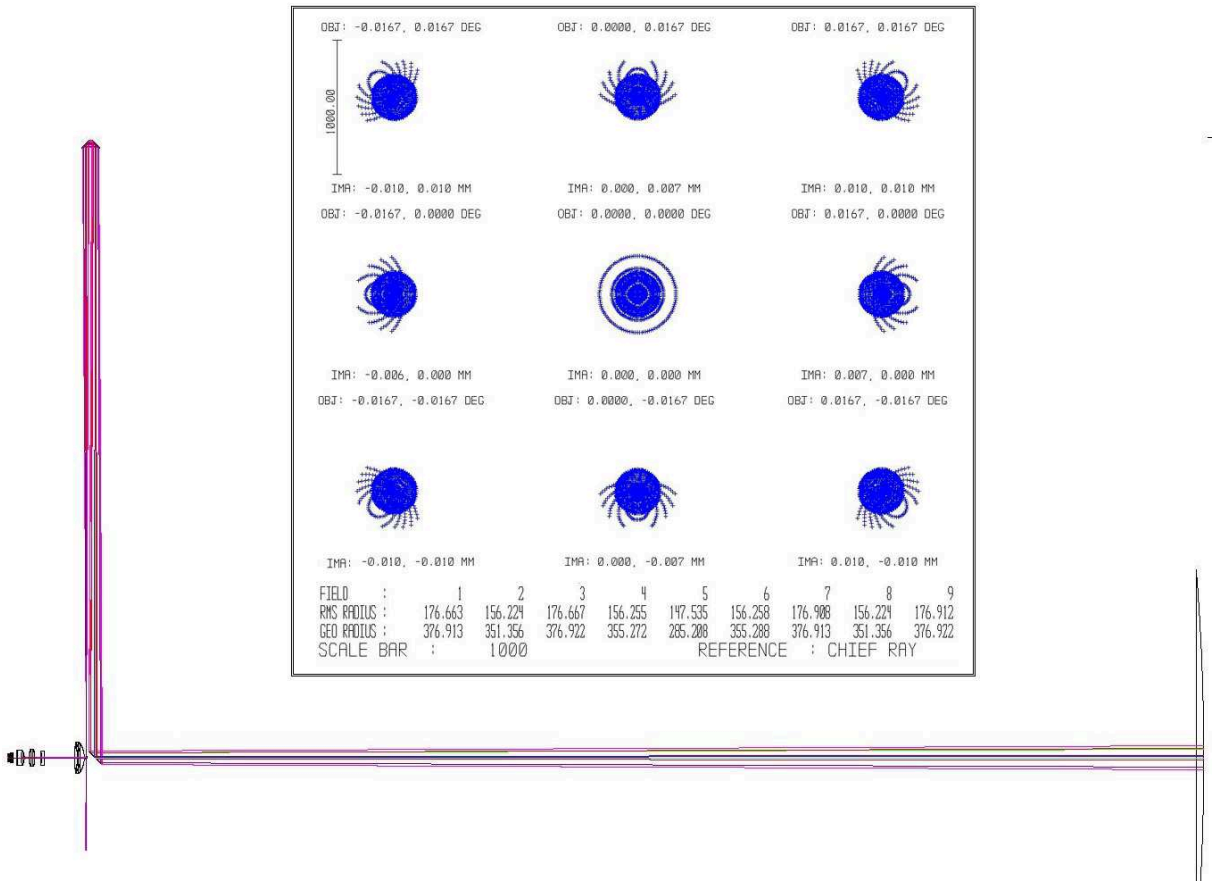


Figure 7.1: Main diagram: 2f-2f design using 4 folding mirrors. The optical device on the left is the corrector and the 2m mirror is on the right. The numerical aperture of the source is controlled in order to optimize the beam width at the level of the last folding mirror. Sub-figure: Spot diagram test of the controlled numerical aperture design. The sizes of the spots are large enough (more than 10 pixels in radius) and the changes in shape are detectable.

7.1.2 Alignment method

The alignment method we propose is based on the comparison between measurements and predictions from theoretical models. In the case of the 2m LMT, we compared the nine PSFs corresponding to the nine positions of the source in the field of view (cf. previous section).

Based on these comparisons, it is possible to precisely determine the actual position of the upper-end unit with respect to the optical axis of the liquid mirror. This position is characterized by an horizontal displacement (x,y) parallel to the focal plane and by a tip/tilt angle (θ_x,θ_y), both measured from the optical axis of the mirror. The camera and corrector can then be moved to their optimal positions. Let us note that the x and y axes are determined with respect to the corrector symmetry (depending on the TDI direction).

Although the alignment method presented here has been developed using the optical design introduced in the previous section for the 2m LMT, it is important to notice that it is completely independent of this design. Only the simulations presented hereafter are design-dependent, but the procedure is easily transposable to any other optical design. It will indeed be used as a basis for the alignment of the 4m ILMT presented in section 7.2 where it will be associated with

another, simpler illumination design and a completely different optical corrector.

Correlation maps

The optical model presented before allows us to compute the PSF for any position of the upper-end unit. However, Zemax is a ray tracing software that does not perform the optical propagation very well, and the computation of PSFs (by Zemax) depends on the source and image sampling and on the type of source. Nevertheless, the ray tracing allows to precisely compute the Zernike aberration coefficients that can be used to compute these PSFs. We have implemented the Fourier method presented in chapter 4 in the Matlab software. We can then compute the PSF that corresponds to the optimal alignment[25] (x, y, θ_x, θ_y=0) and compare it with the measured PSF.

The comparison is performed with a two dimensional correlation coefficient defined by

$$C = \frac{\sum_{m,n}(A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\left(\sqrt{\sum_{mn,}(A_{mn} - \bar{A})^2}\right)\left(\sqrt{\sum_{m,n}(B_{mn} - \bar{B})^2}\right)} \tag{7.1}$$

where A and B represent the matrices containing the intensities of the pixels of the two images and $\bar{A}$ is the two dimensional mean of the A matrix (the mean intensity). This correlation coefficient gives some information about the similarity between the two images. When they are exactly identical, its value is unity and it decreases when the differences between the two images increase. This coefficient thus provides an indication about the position of the upper-end unit with respect to the mirror axis. When its value is near to 1, the system is almost correctly aligned. On the contrary, if the value is small, the upper-end unit is far from its optimal position. However, "near" and "far" are not sufficient, it is mandatory to compute the exact location of the upper-end unit in order to bring the necessary correction.

We have thus introduced the concept of correlation maps, which are made in such a way that they establish a link between the position of the upper-end unit and the corresponding value of the correlation coefficient.

First of all, a grid of upper-end positions is defined, and the PSFs corresponding to these positions are then computed and compared with the aligned image. The value of the corresponding correlation coefficient, computed with equation 7.1, is then associated with the associated position. Each position defined by the grid is associated in the same way with a value of the correlation coefficient, which creates a map linking the positions and the coefficient values.

This concept is theoretically very interesting, since it allows to reach any accuracy, as the grid of position can be defined with any step (i.e. the difference between two consecutive positions). In practical cases, the only limitation comes from the possibility to distinguish between two images. However, when the differences between two images are so small that they cannot be distinguished, one can consider that an acceptable alignment is reached.

Two types of maps have been created; the first one consists in a horizontal displacement map, that contains the correlation values corresponding to an upper-end unit moved in the focal plane (along x and y axis), and the other one is a tilt map, that links the correlation values to rotations

[25] In the following, we will call "aligned image" or "aligned PSF" the PSF that corresponds to this perfect alignment.

of the upper-end unit about the x and y axes (tip/tilt). Both types of maps have to be combined in order to get all the information about the upper-end unit position.

Provided that the actual position of the upper-end unit is inside the map limits, the correlation coefficient values associated with the measured actual PSF will correspond to a position in the map, and the actual position (x, y, θ_x, θ_y) of the upper-end unit can be determined.

The meshing definition is an important step in the creation of the maps. The boundaries of the map should be such that any misalignment of the upper-end unit is included into the map limits. The upper and lower limits should thus be as large as possible.

Moreover, the accuracy of the measurement method depends on the smoothness of the position grid defined to create the map. It is thus preferable to compute maps with very small steps in order to get the highest possible accuracy. However, two limitations appear; as previously said, it is worthless to define positions that are so close to each other that it is not possible to distinguish the difference between their respective PSFs. Moreover, using small steps and large limits drastically increases the number of PSFs that has to be calculated, and, thus, the computing time.

In order to circumvent this second issue, we have created encapsulated maps. A large, rough map is first constructed with large upper and lower limits (i.e. 10mm for the horizontal displacement map) and a rough meshing (1mm) in order to approximately locate the upper-end unit. Then a second map with small limits corresponding to the previous step (1mm) and a more precise meshing (0.1mm) is used to compute the accurate position. Any number of encapsulated maps can be created to cope with any situation.

In the following simulations, we have used three maps: two encapsulated maps for the horizontal displacement and a single one for the tip/tilt. Their characteristics are summarized in table 7.1. In order to avoid error propagation, the useful limits of the inner horizontal map (presented in fig. 7.2) are twice as large as the step of the outer map. Moreover, the actual boundaries of the maps are slightly larger than the useful limit in order to avoid border effects. Obviously, other maps could be created if necessary.

Map name	Type	Boundary	Useful limit	Step
Mini	horizontal	1.2mm	1mm	0.1mm
Normal	horizontal	6mm	5mm	0.5mm
Tilt	tilt	30"	28"	2"

Table 7.1: Definitions and characteristics of the different maps used. The boundary limits are slightly larger than the useful limit in order to avoid border effects.

These limits have been set in order to cope with typical values of misalignment obtained with a manual assembly. We assumed that an horizontal displacement of ±5mm and tip/tilt angles of ±0.5 deg. should be easily achieved. The steps of the maps have been determined in order to get an accuracy well bellow the limit given by the tolerancing previously described. The horizontal displacement will be measured with an accuracy of 0.1mm in both x and y directions and the tip/tilt angles with an accuracy of 2 arcmin.

It should be noted that both types of displacement (horizontal and tip/tilt) are treated separately. The dominant effect (horizontal or tip/tilt) is first determined and then computed with the corresponding maps. Once it has been corrected for, the other type of map is used to

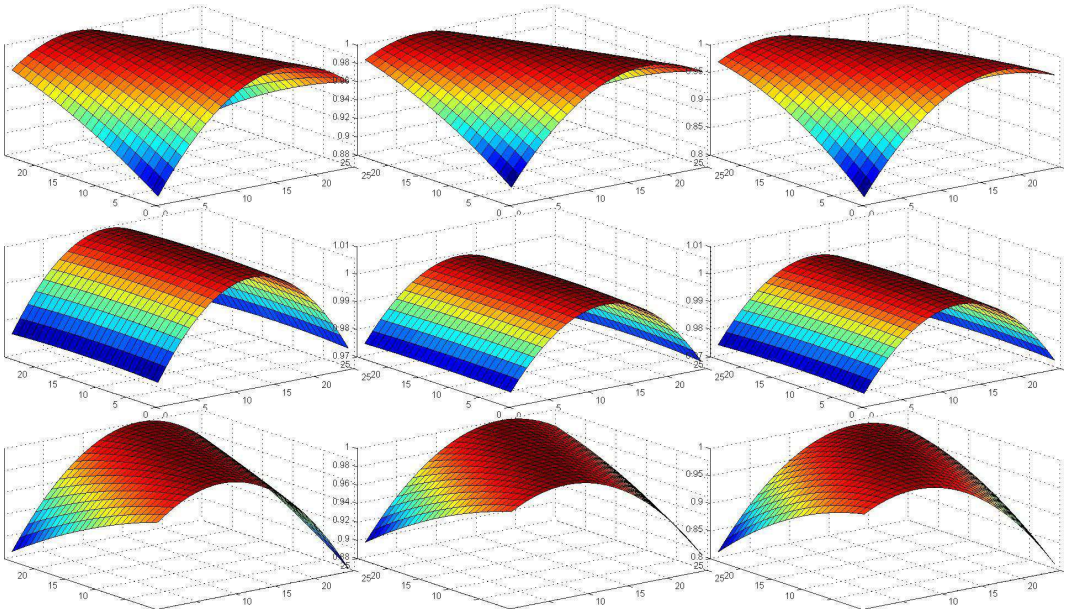


Figure 7.2: Three dimensional view of the set of mini maps. They are presented in normalized intensity. They give the relation between the value of the correlation coefficient and the horizontal displacement of the upper-end unit for positions between ±1.2mm around the optical axis with a step of 0.1mm. The other maps have the same kind of aspect.

compute and correct the other type of displacement. This process can be repeated until no more effect is measurable.

The creation of the correlation maps involves two main processes: the extraction of the information related to the position grid of the upper-end unit, and the computation of the PSFs and of their corresponding correlation coefficient.

The first part is performed with a Zemax macro that has been specifically programmed in Zemax Programming Language (ZPL). Using the model of the optical design presented before, it scans all the positions defined by the grid in order to extract the related Zernike coefficients. Each position is related to nine sets of coefficients corresponding to the nine fields. These aberration coefficients are stored in a file.

The computation part is performed by our Matlab routine, which reads the Zernike coefficients inside the files created by Zemax and computes the PSF corresponding to each position with the Fourier transform approach presented in chapter 4. The pupil function is given by the sum of the Zernike polynomials balanced by their respective coefficients and the PSF is obtained from the Fourier transform of this pupil function. Once the PSFs have been computed, they are compared with the aligned PSF (the PSF that corresponds to $x = y = \theta_x = \theta_y = 0$) thanks to the correlation coefficient. The Matlab software then associates the computed correlation values to their corresponding positions on the grid. Finally, we get some matrices containing values of correlation for each position defined by the grid. Each map is composed of nine sub-maps, for each of the nine fields considered previously.

Our maps thus associate measurements with a position of the upper-end unit. In the case presented here, those measurements are correlation coefficients but they could be any other measurable parameter. In section 7.2, we will apply this technique directly with aberration coefficients instead of PSF correlation coefficients in order to align the ILMT camera and corrector.

Computing the misalignment

The determination of the misalignment of the upper-end unit requires a measurement, which simply consists in acquiring an image (or a few images) with the telescope equipped with the optical design presented previously. The actual value of the correlation coefficient is then obtained by comparing this PSF with the theoretically computed aligned PSF. Using the maps presented before, it is possible to associate the actual value of this coefficient with the position of the upper-end unit. This step is performed by computing the intersection of the maps with a horizontal plane whose height corresponds to the actual correlation value (fig. 7.3).

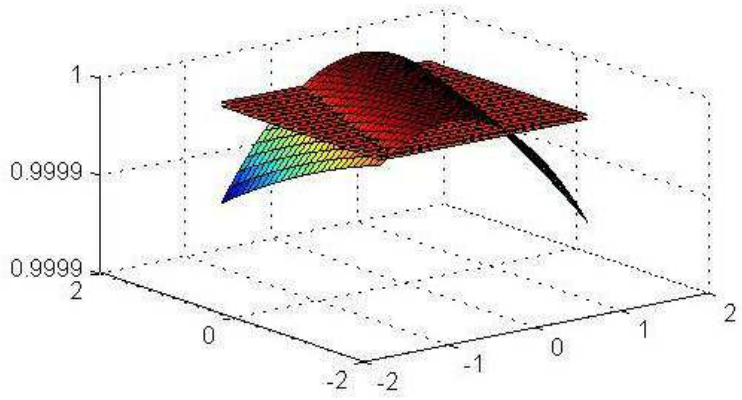


Figure 7.3: The intersection between the map and the correlation horizontal plane corresponds to the points belonging to the map which are located between the two planes.

Nine intersections are obtained in this way (fig. 7.4 left) for a particular position. As shown in the figure, these intersections are segments of curve, which means that several positions of the upper-end unit have the same correlation coefficient value. However, combining the nine intersections allows to break the symmetry of the maps and to reduce the number of effectively possible positions (fig. 7.4 right).

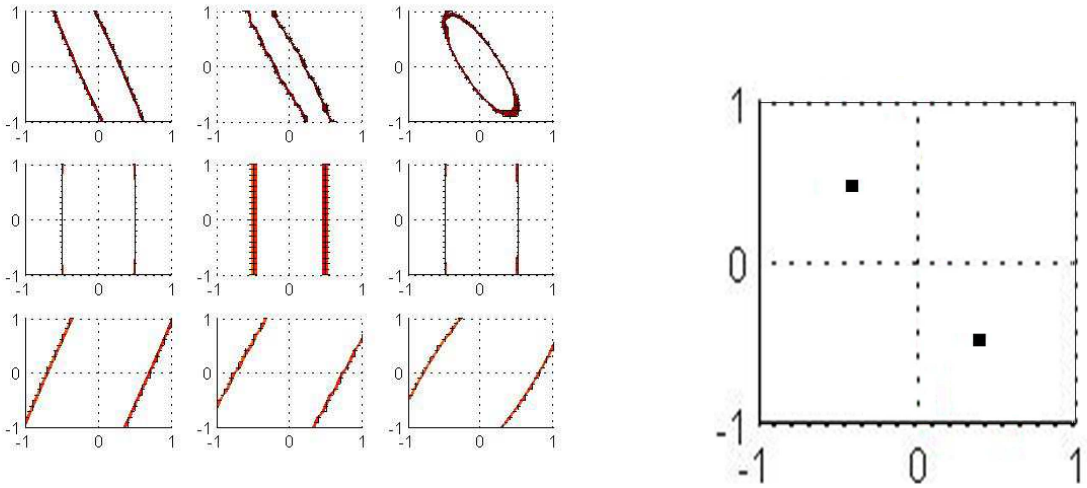


Figure 7.4: Left: The 9 field intersections. Right: Points in common between the 9 intersections.

However, one still generally gets two symmetrical positions (fig. 7.4 right), and it often

happens that the actual position is not precisely located on one node of the mesh. In this case, several points around the true location are found, and the actual position is given by the mean of their positions. The absolute value of the position is computed in this way but the signs of the displacements are still unknown because of the central-symmetry.

In order to break this symmetry, we apply a known displacement to the upper-end unit. The absolute value of this new position is then computed with the same map-intersection method, and, knowing the two positions and the distance between them, it is possible to determine the sign of both measurements. Even if this second measurement can be used to improve the precision on the measured actual position, it also renders the method more difficult. The improved approach presented in section 7.2 does not have this problem of symmetry since the aberrations used in the maps do not vary symmetrically.

This measurement technique has been extensively tested by simulating PSFs corresponding to random positions of the upper end. Hereafter, we present a summary of the results. All simulations and their results are presented in Magette (2007).

7.1.3 Simulations and results

The simulations consist in applying the measurement method presented above for many random positions of the upper-end unit (Monte-Carlo analysis). The reliability and accuracy of the alignment method are tested using two estimators; the position error and the shift error.

The position error is the absolute difference between the computed position and the true one, expressed in step of the map. The error magnitude is divided in three categories. The computed position is considered perfect when the position error is smaller than 0.5 step, which is the most accurate position that can be obtained with this map. A position error between 0.5 and 1 step corresponds to a correct computation of the position. In this case, the error is smaller than the resolution of the map. Beyond 1 step, the error is too large and the result is considered to be erroneous.

The shift error is the difference between the actual displacement applied to the upper-end unit (between the two images) and the measured one, expressed in map step. When this error is larger than 0.5 step, at least one of the measured positions is wrong. This can be corrected by using a third image related to a third position of the upper-end unit. This process has been implemented in a function called "autocheck". It consists in verifying the reliability of the measurements by using only those measurements. Using three images allows to know if the position has been correctly determined, but this third image increases the complexity of the method. We will see in section 7.2 that the new aberration based method requires only one image, since the symmetry is intrinsically broken in this method.

Hereafter, two types of simulations are presented. The first one does not take advantage of the "autocheck" function while the second ones does. The latter should thus present better results, but it is more complicated.

Table 7.2 summarizes the results related to the three types of map for the first estimator defined before. The results are expressed in percent of the total number of tests performed and the errors are measured in map step unit. The reliability of the method seems very good (more than 95% of the measurements are exact within the map resolution) as well as its accuracy (a large part of the correct measurements have an error below the measurement instrument half graduation). The interest of taking the shift error into account is obvious, since it improves both

the reliability (more acceptable cases) and the accuracy (more errors smaller than 0.5) of the method.

		Autocheck off			Autocheck on		
	error	< 0.5	< 1	> 1	< 0.5	< 1	> 1
Mini	X-position	97	3	0	99	1	0
	Y-position	89	9	2	92	7	1
Normal	X-position	95	2	3	96	3	1
	Y-position	85	10	5	92	8	0
Tilt	X-position				69	24	7
	Y-position				98	2	0

Table 7.2: Number of position errors as a function of their categroy ($< 0.5, < 1, > 1$) for the three types of maps. When the error is smaller than half a map step (< 0.5), the computed position is correct with the best possible accuracy (half of the graduation of the measure instrument). In the case where the error is smaller than the step of the map (< 1), the measured position can still be considered correct as the error is below the resolution of the map. Finally, if the error is larger than one step of the map (>1), the measured position is bad. The horizontal displacement map gives results that are acceptable in 99.5% of the cases, the tilt maps are reliable at 95%. The "Autocheck off" column gives the results of the simulations when the shift error is not taken into account. The "Autocheck on" column corresponds to simulations that took the shift error into account. The table shows that this "autocheck" function should always be used as it really improves the performance. That is why the tip/tilt results are presented only for this case.

Based on these results, it should be possible to measure the horizontal displacement of the upper-end unit in a range of ±5mm with an accuracy of 0.1mm. As far as the tip/tilt is concerned, it can be determined in a range of $\pm 0.5^\circ$ with a resolution of 2 arcmin. However, these are only simulation results and they only demonstrate the possibility of the method, not its feasibility.

Influence of the Zernike coefficients on the upper-end unit position measurement

The simulations presented previously were performed using the same number of Zernike coefficients to compute every PSF, this is the aligned PSF, the PSF used for the map creation (map-PSF) and the simulated PSF corresponding to the unknown position (unknown-PSF). However, this hypothesis is not necessarily true. We thus tested the "mini" maps using different numbers of Zernike coefficients in order to simulate the unknown-PSF and a constant number to compute the model PSFs (aligned-PSF and map-PSF). The first three rows of table 7.3 present the corresponding results. As previously demonstrated, the misalignment determination method is very reliable when all PSFs are computed from the same number of Zernike coefficients. The results are totally different when these numbers are different and the method seems completely useless in that case. This problem is quite detrimental for our method since the number of significant Zernike coefficients that composes the recorded PSF is nearly infinite (limited by the camera field of view).

Another test of this type has been conducted, which consists in using the same number of coefficients (40) for the aligned-PSF and for the unknown-PSF and a different number to create the maps (50). The last row of table 7.3 shows the errors related to this simulation. In this case, the results are better, which shows that provided the aligned-PSF and unknown-PSF are computed from the same number of Zernike coefficients, the results are good ($\sim$ 85% of

reliability), even if the map used was created with a different number of coefficients. We have made some other tests showing that each map has a range of usability, which means that it can be used with an aligned-PSF and unkown-PSF based on a given number of coefficients (the same for both PSFs), provided the difference in the number of coefficients is included in this range.

Number of coefficients			X-position		Y-position	
Map	Aligned	Unknown	< 1	> 1	< 1	> 1
50	50	30	0	100	0	100
50	50	50	99	1	98	2
50	50	75	35	65	20	80
50	40	40	82	18	90	10

Table 7.3: Results of the simulations using different numbers of Zernike coefficients to simulate unknown-PSF and to compute the model PSF (aligned-PSF and map-PSF). The X-position and Y-position columns present the percentage of error of the location process. Even if the misalignment is correctly estimated when all the PSFs are computed with the same number of coefficients, it is not the case anymore when the simulation and model PSFs use different numbers. The last row presents another test using the same number of Zernike coefficients for the aligned-PSF and unkown-PSF and a different one for the map-PSF. In this case, the results are better.

7.1.4 Conclusion

We have developed a numerical alignment method and demonstrated the high quality of the map concept. However, the approach presented in this section is quite complicated as it requires three images of nine fields. Moreover, the illumination design could be tricky to implement as it involves several optical elements located at positions difficult to reach. Finally, the issue related to the number of Zernike coefficients presented above is an important limitation to the method.

This work was achieved at the beginning of my PhD thesis. Because of its limitations, we have investigated phase retrieval techniques and particularly the Nijboer-Zernike theory presented in the previous chapters. The latter was not extensively used in the past although it presents several advantages. Among them is the interesting possibility of using it in the image plane, and its compatibility with the presence of large aberrations.

The Nijboer-Zernike phase retrieval method is used in combination with the concept of maps in the next section. This new alignment approach circumvents all the difficulties encountered with the method presented here.

7.2 The ILMT alignment: an optimized approach using the phase-retrieval technique

Taking advantage of the Nijboer-Zernike phase retrieval method, we present here an improvement of the alignment method introduced in the previous section. This new method has been developed for the 4m ILMT. We keep the concept of the map, but we replace the comparison of PSF with the direct comparison of aberrations (section 7.2.3). This new method requires the adaptation of the optical design, presented hereafter, in order to account for the specificities of the Nijboer-Zernike retrieval method and those of the ILMT. Finally, simulations of this upgraded alignment method and their results are shown. They demonstrate that this new method, while easier to use, is also more accurate and reliable than the previous one. Indeed, the problem related to the Zernike coefficients has been almost completely circumvented.

7.2.1 Principle

The previous method presented several problems, such as its complexity (3 images and 9 fields) or the difficulty of implementation of the optical design. However, its main limitation came from the number of aberration coefficients used to compute the PSF. In our upgraded method, we do not compare the PSF anymore but we directly use the aberrations in order to determine the misalignment. Images are obtained from the telescope with the upper-end unit in its actual (misaligned and unknown) configuration and then analyzed with the Nijboer-Zernike phase retrieval technique. Once the aberrations contained in the PSF have been determined (actual aberrations) they can be compared with the theoretical ones, obtained with the optical model.

This technique has many interesting advantages, the first one being its speed of execution. Indeed, once the aberrations have been extracted from the PSF, there is no need to compute any Fourier transform. The setting up of the maps is very fast, and additional ones can be constructed very easily, if required.

The other important advantage, that makes this method usable in practice, contrary to the method presented in the previous section, is the independence of the process with respect to the number of aberrations contained in the PSF. Indeed, as long as enough aberrations are accessible for the measurement process, the misalignment can be computed without the knowledge of all the aberrations. In principle the tip and tilt aberrations are highly sufficient for the process to take place, but the knowledge of more aberrations increases the reliability and accuracy of the method. Moreover, the retrieval capability of the Nijboer-Zernike retrieval approach has been extensively demonstrated in chapter 6, at least for low order aberrations, even with very noisy images.

Finally, this method requires only one image and only one field to provide accurate measurements, and, furthermore, the optical design could hardly be simpler.

However, there is still one drawback. The variation of the value of each aberration due to the misalignment of the upper-end unit has to be large enough to be measurable. This limitation did already exist in the previous version of the alignment method, but was more difficult to estimate. It should be noted that this limitation is not caused by the Nijboer-Zernike aberration measurement technique, since, as we have shown, the method can easily reach an accuracy of $\lambda/1000$ even with noisy images. However, many perturbations can cause aberrations which are small enough, that they can not be associated with a misalignment under a sufficient degree of

certainty. It is thus necessary that the variation of the aberrations related to a misalignment gets larger than those small uncontrolled variations. In our simulations, presented hereafter, the process only takes into account the aberrations larger than $\lambda/100$ and with variations larger than $3\lambda/100$ over the range of the map (P. Riaud, private communication).

The new optical scheme presented below takes these practical considerations into account, and has been designed in such a way that the low order aberrations present variations of a few hundredths of a wavelength.

7.2.2 Optical design

Since the design presented in section 7.1 for the 2m LMT does not meet the condition related to the aberration variation stated before, we had to modify it. Moreover, its complexity due to the important number of required optical elements (folding mirrors) makes it difficult to implement. Any misplacement of those elements would result in variations of the aberrations and would disturb the measurements.

We have thus gone back to the idea of a parallel decentered beam that is far more simpler to implement. Using defocused images should allow to get aberrations sufficiently large to be measurable. However, simulations show that their variation with the position of the upper-end unit are not large enough and cannot be observed.

Using a slightly diverging beam instead of a perfectly collimated one generates a larger defocus that generates measurable aberrations (larger than $\lambda/100$) with variations large enough to be detected (larger than $3\lambda/100$). Moreover, this optical design is the simplest that can be imagined as it only consists of a small diverging source (fiber laser) located a few meters above the liquid mirror. It is presented in fig. 7.5.

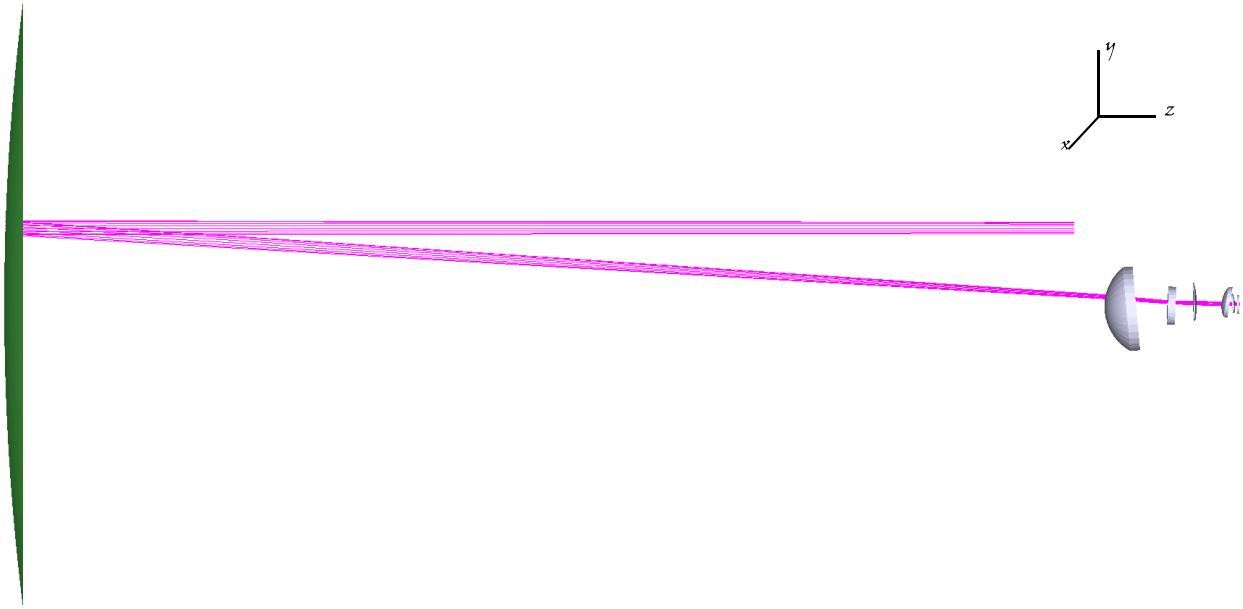


Figure 7.5: Optical design used to simulate the alignment of the upper-end unit of the ILMT. The source is decentered by 50cm along the y-axis and has a diameter of 7.5cm. It is sufficiently diverging ($\sim F/400$), to generate aberrations with measurable variations as a function of the upper-end unit displacements.

It is interesting to note here that the tilt aberration is much more important along one axis than the other because of the asymmetry of the corrector. The position of the source (the axis along which it is decentered) could thus be optimized in order to increase the tilt aberration along the other axis. However, the difference is not very important since only the median magnitude of the aberration is increased and not the amplitude of its variation.

7.2.3 Aberration maps

As in the previous case, we created some maps in order to establish a link between the value of the aberrations and the position of the upper-end unit. Each of these maps, called "aberration maps" (fig. 7.6), corresponds to a single aberration, and translates the position of the upper-end unit in terms of magnitude of this particular aberration. Contrary to the correlation maps defined for the previous version of the method, these aberration maps concern only a single central field. Indeed, the nine sub-maps are completely useless in this case since the symmetry they were supposed to break does not appear in the aberration maps. One single field is thus amply sufficient, provided that at least two orthogonal aberrations (i.e. tip and tilt, x-coma and y-coma,...) are available for the measurement process.

The computation time has also been decreased for these maps. The separate use of two independent softwares, for the previous method, implied time consuming read and write access on the hard drive disk. We have thus changed the architecture of the map calculator and of the simulator software in order to reduce the calculation time. The whole management is performed by a Matlab software that interacts with Zemax in order to modify the optical model (position of the upper-end unit) and to retrieve the theoretical information (aberrations) required by the process. The whole code is thus written in Matlab language.

Maps are computed for an arbitrary number of aberrations. The fifteen first are presented in fig. 7.6. However, some of these maps do not fit the magnitude or the variation amplitude requirement described earlier, and they should not be used by our algorithm in order to keep the simulations realistic. We have thus defined a few conditions that the maps should fulfill before being used for the measurements:

- the map should not be null everywhere (map significant),
- The variation of the aberration corresponding to the map should be larger than $3\lambda/100$ between the maximum and the minimum of the map (variation detectable),
- the measured aberration corresponding to the map should be larger than $\lambda/100$ (aberration significant),
- the value of the measured aberration should be in the range of its corresponding map (map adapted to the aberration).

These conditions ensure that the map can be used without any problem for the measurement of the upper-end unit position and that using this map is realistic in practical cases. Indeed, simulations can compute the precise location of the upper-end unit with aberrations as low as 10^{-15} but it would be neither true nor realistic for practical cases. We thus restrict the maps and aberrations used to those really usable for practical applications.

The ranges and steps of the new maps are the same as those of the old maps defined previously, they are reminded in table 7.4.

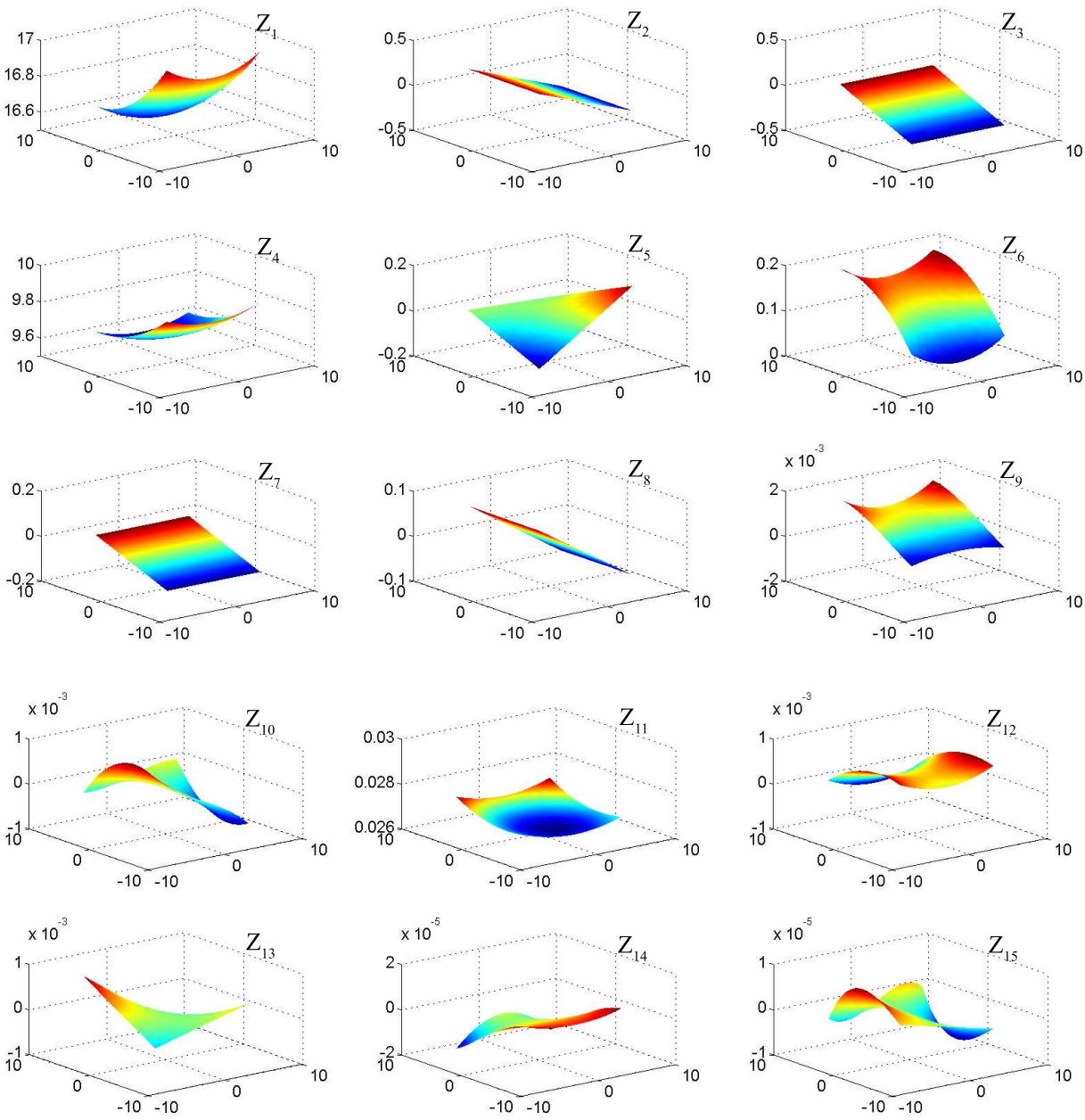


Figure 7.6: Maps of variation of the 15 first aberrations as a function of the horizontal displacement of the upper end unit. The range of displacement is between ±6mm with an accuracy of 0.5mm, which corresponds to what we have defined as the maps of "normal" size in the previous section. The variations of most of those maps are sufficiently important to be measurable.

7.2.4 Simulations and results

This new method has been extensively tested by simulating many random positions of the upper end unit. Fig. 7.7 shows the results of the measurements. The point forest represents the absolute errors on the position expressed in map step for one thousand unknown locations of the upper end unit measured with the normal map. More than 99% of the simulated unknown positions are measured with an accuracy better than one step of the map (the map precision). The error on the horizontal position is thus smaller than 0.5mm, which corresponds to one thousandth of the

Maps name	Type	Boundary	Useful limit	Step
Mini	horizontal	1.2mm	1mm	0.1mm
Normal	horizontal	6mm	5mm	0.5mm
Tilt	tilt	30"	28"	2"

Table 7.4: Summary of the characteristics of the different maps used. The boundary limits are slightly larger than the useful limits in order to avoid border effects.

corrector first lens diameter. Moreover, 94% of the positions are determined with an accuracy better than half of the step of the map. As in the previous case, it is the best precision that can be reached, and the method is thus optimal in 94% of the cases.

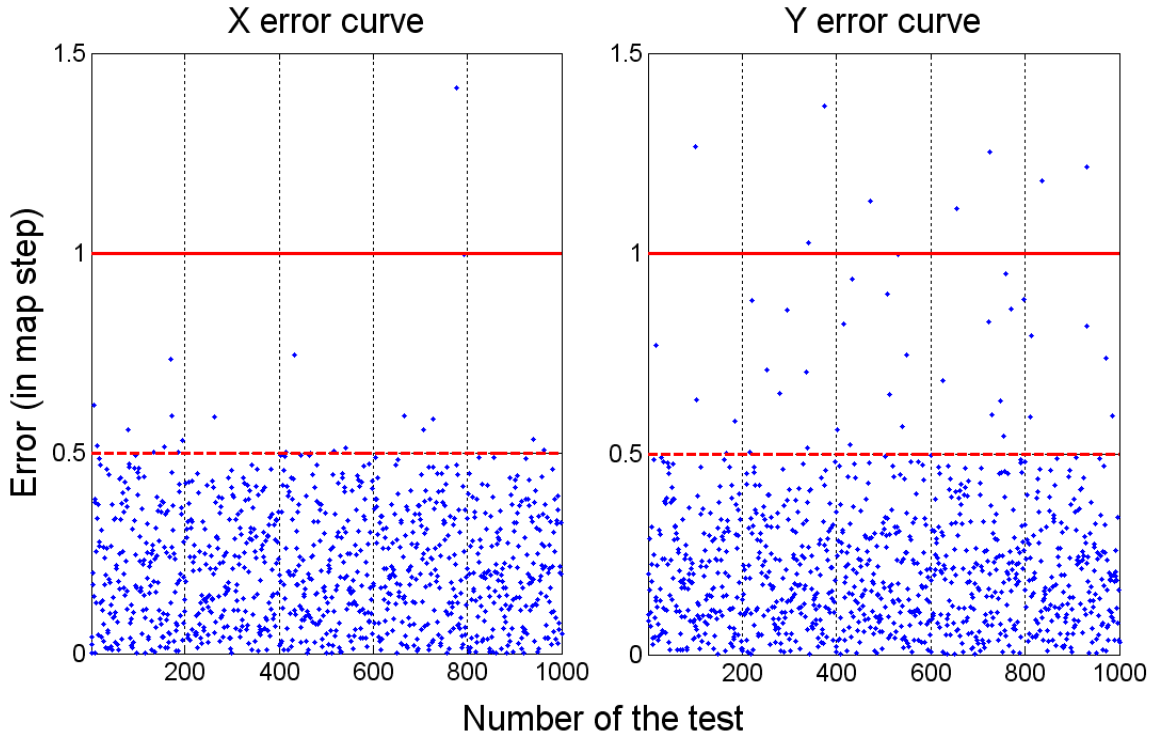


Figure 7.7: Results of Monte-Carlo simulations. They represent the error on the calculation of the location of the upper end unit using the maps presented in fig. 7.6. The red solid line corresponds to an error of one step of the map (0.5mm) and the dashed line corresponds to an error of half a step. The error is smaller than one step in 99% of the cases and smaller than half a step in 94%.

The results presented above are already quite accurate, but we pushed the tests further with the "mini map", which should allow to reach an accuracy of 100μm. The measurement process has been simulated for one hundred unknown positions. 80% of them were computed, with an accuracy better than one step and 76% better than half a step (50μm). This poorer result is due to the very small tilt aberration in one direction, already discussed before, which is so small that it cannot be not taken into account (third condition) by the location process. Hence, one of two orthogonal aberrations required are not available, and the process is not reliable; the positions cannot be measured accurately. Modifying the position of the source could help to increase this aberration and then the global reliability of the mini map. For example, a rotation of the position of the source by 90 degrees around the axis of the mirror should be sufficient.

The tilt map has also been tested and 99% of the positions have been measured with an accuracy better than one step (2 arcmin). 95% present errors smaller than half a step. The results were so accurate and reliable that it should be possible to increase the precision by using a map step of 1 arcmin.

7.2.5 Conclusion

We have presented an improved numerical alignment method, based on the map system developed in section 7.1.2. The use of the Nijboer-Zernike retrieval theory brings a convenient way of circumventing the trouble we had with the previous method concerning the number of coefficients used for the computation of the PSF.

Moreover, the new method is much more practical and easy to use, since the illumination design is far more simpler, and since the process only requires one image of one field.

Our simulations have demonstrated that the improved approach should allow to accurately compute the position of the upper-end within an accuracy of 0.5mm and 1 arcmin. The reliability of this method seems also to be extremely good since 99% of the results have the best possible accuracy. We thus have now an easy, usable in practice, reliable and accurate method to measure the misalignment of the ILMT upper-end unit.

Chapter 8

Application 2: Aberration measurements

In this chapter, we present two applications of the Nijboer-Zernike phase retrieval method. First of all, we will measure, a posteriori, the static aberrations of an adaptive optics instrument installed on the VLT-UT4: NACO. Three defocused images acquired with this instrument will be used to perform the retrieval. This first practical trial of the method on real images reveals technical difficulties that had not been encountered during the simulations. Like for many other telescopes, the pupil of the VLT contains a central obscuration (14% in diameter) and a spider (see fig. 8.2). To account for this, we have introduced a convenient way to consider pupils different from a full-disk in the retrieval process. Using this improvement, we are able to compute the phase error on the wavefront reaching the detector of NACO. The results of the retrieval can then be used in combination with coronagraphic data in order to improve them by calibrating the speckles related to the instrument. Indeed, static residuals that could get mixed up with star companions (planets, disks, ...) will be removed. The NACO coronagraph simulator has been used to estimate the amount of speckles due to these phase errors, it appears that they generate about 25% of the speckles.

We will then present a classical use of the NZ method: the measurement of the aberrations of a simple lens on an optical bench. We first determine the main aberrations with a numerical model of the system and then compare them with the actual results. This allows us to remove the aberrations related to the testing setup and to determine the actual aberrations of the lens. The same process can be used to characterize the lenses of the ILMT optical corrector, these tests will be introduced hereafter.

Finally, we will present an evolution of the method that allows to increase the measurement dynamic (in order to avoid saturation), provided that the dominating aberration is known. It is indeed possible to slightly adapt the NZ method to take this dominant aberration into account in order to prevent the measurements from saturation that would mask the fainter aberrations. We will present this adapted technique and the way to use it in order to measure the aberrations in parabolic mirrors without using auxiliary optics (Offner's lenses). This improvement is possible because, contrarily to other methods, the NZ approach does not require a reference surface. The accuracy on the measured aberrations is thus not limited by this reference and no saturation phenomenon should mask the small aberrations. It is thus possible to measure very small aberrations mixed together with large ones. This is particularly interesting to test aspherical optics from their center of curvature for which it may be difficult to generate the adequate reference wavefront. In other words, the NZ measurement method does not require the use of null lenses to test aspheric mirrors. The same approach could obviously be used to study larger parabolic mirrors, like the ILMT primary mirror, from their center of curvature with a method that is very

simple to implement.

8.1 NACO aberration measurement

NACO is the first adaptive optics (AO) system implemented on the Very Large Telescope (VLT) and it saw its first light in November 2001. It is composed of the Nasmyth Adaptive Optics System (NAOS) and the COudé Near Infrared ($1 - 5\mu m$) CAmera (CONICA).

NAOS receives an f/15 beam from the Nasmyth focus of the telescope, corrects it for atmospheric turbulence and transmits a 30 arcsec width image with a Strehl ratio of about 35-60% in the K-band to CONICA. Once in NAOS, the beam is collimated by an off-axis parabola. It is then successively reflected onto the tip-tilt and the deformable mirrors. A fraction of the corrected beam is sent to the Shack-Hartmann wavefront sensor while the main part of the beam is refocused onto the entrance focal plane of CONICA with the same numerical aperture. See section 4.1.6 for a detailed overview of adaptive optics.

The wavefront sensor directly sees the wavefront errors due to any degradation preceding it. The AO system can thus correct for these aberrations automatically. This is not the case for a degradation of the image quality induced after the sensor. The optical train that follows the sensor is composed of several optical elements that should be precisely aligned (especially the off-axis focusing parabola at the output of NAOS). As it cannot be perfectly done, some aberrations are generated in this part of the optical path. In order to reach optimal performances, those aberrations should also be compensated, which can be done by the deformable mirror provided that they are known. It is thus mandatory to calibrate the static aberrations of NACO (those that are not sensed by the Shack-Hartmann and corrected by the AO).

Previous calibrations have been performed by phase diversity. The method is presented in Blanc et al. (2003) and the experimental results are given in Hartung et al. (2003). Phase diversity is also based on the estimation of aberration from defocused images. However, this calibration was not performed directly on astrophysical objects but only on a calibration fiber source. We show hereafter an alternative way of measuring the static aberrations. Using a few stellar images corresponding to different positions of the detector around the focus, we determine the phase errors with the Nijboer-Zernike phase retrieval technique.

8.1.1 Description of the on-sky experiment

Through focus images (i.e. images before, on, and after the focus, from intra to extra focal) of HD25026 have been taken by P. Riaud during an observation run on September 26, 2009 (fig. 8.1). They consist of a co-addition of 160 frames of 0.35s exposure. Their characteristics are summarized in Table 8.1. These images will be used hereafter to measure the fixed residual aberrations of NACO.

Images pre-treatment

As any astronomical image, those used for the retrieval process have to be corrected for "dark" and "flat-field". The images provided by P. Riaud have already undergone these classical reductions.

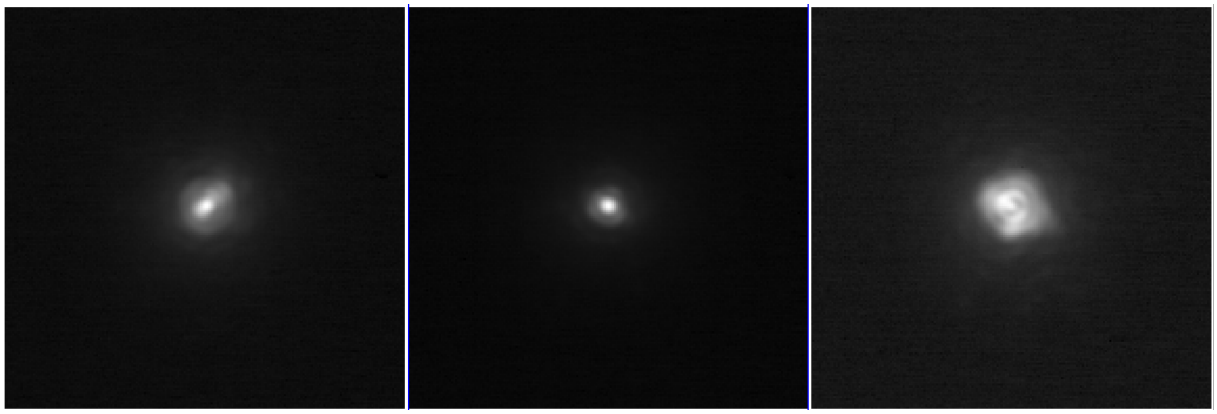


Figure 8.1: Images obtained with the NACO instrument on the VLT for several positions around the focus (-3, 0, 3mm respectively). Their characteristics are summarized in Table 8.1. Those images have been taken by P. Riaud during an observing run on September 26, 2009.

They thus only require to be centered[26] to be usable by the retrieval algorithm.

Size [pix]	192×192
Resolution [mas/pix]	13.27
Sampling [pix/(λ/D)][27]	4.65
Defocus [mm]	± 3
Number of exposures	160
Exposure time [s]	0.35
Wavelength [μm]	2.166
Spectral width ($\lambda/\Delta\lambda$)	70

Table 8.1: Characteristics of the images that will be used to perform the phase retrieval.

8.1.2 Aberration retrieval

Contrary to the simulated image case, applying the aberration retrieval process on science images requires more preparation, since the retrieval parameters are not known for these images. For example, the VLT pupil is not a full disk. We will explain how we took the actual pupil obscuration into account in the retrieval process. The determination of the retrieval parameters and of the optimal number of aberrations that should be computed will also be detailed. Finally, we will present the results of the treatment.

[26] The central peak of the PSF is not necessarily centered on the image, which is not convenient for the retrieval process. We thus place the original image in a larger "black" image (filled with zeros) in such a way that the central peak of the PSF is in the center of the new image. We prefer to add zeros around the original image in order to center the peak rather than removing a part of the original image, which would correspond to erase some information.

[27] The sampling can be computed from the formula: $205.265\lambda/D/\mathrm{R}$, where λ is the wavelength in μm, D is the effective diameter of the pupil in meters and R is the resolution of the instrument in mas/pix.

The VLT pupil

Our aberration measurement method presented in the previous chapters only works with a full circular pupil, which is not the case for the VLT. Like many other telescopes, its center is obscured by a disk, which radius is 14% that of the mirror ($R_{obs} = 0.14 \cdot 8 = 1.12$m), and a spider supporting the secondary mirror is also responsible for some obscuration (fig. 8.2 left). Moreover, the NACO coronagraph uses a "stop" which masks the outside 10% (in radius) of the entrance pupil (fig. 8.2 right).

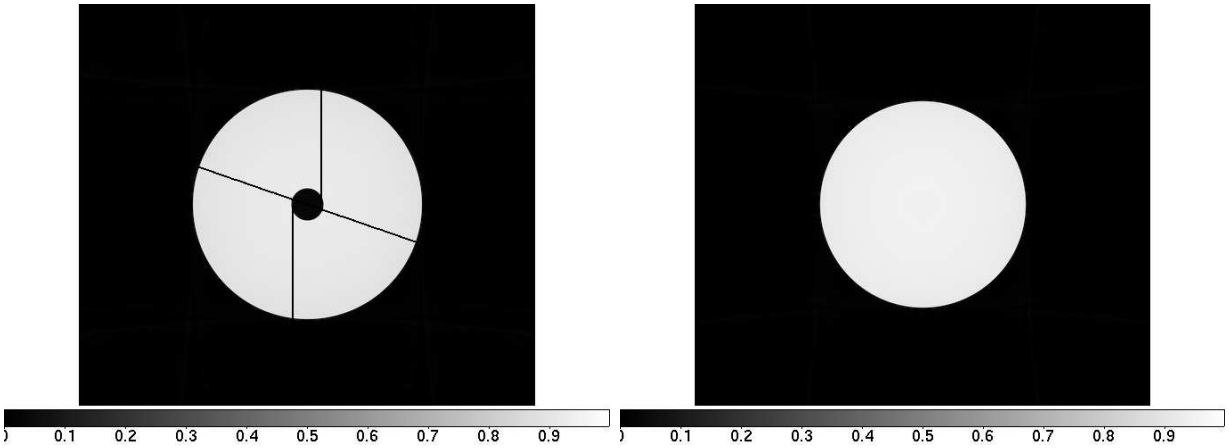


Figure 8.2: Left: Entrance pupil of the VLT presenting a central obscuration with a radius of 14% of that of the primary mirror, and a spider. Right: Coronagraph pupil stop of the NACO instrument, 10% of the radius the pupil is masked.

The particular pupil of the VLT must be taken into account during the retrieval as it modifies the weighing of the aberration coefficients. One solution would consist in changing the aberration decomposition basis. The annular Zernike polynomials (Mahajan 1981) are designed to work on annular pupils. However, they do not allow to account for the spider and require some additional mathematical development. Instead, we have used a slightly different approach.

Let us first remind the global principle of our aberration determination method. The accurate retrieval is based on an iterative process that consists in minimizing the differences between the input images and those computed from the measured aberrations. We thus first calculate the aberrations from the input images and then recompute new images from these aberrations. These reconstructed images are then subtracted from the initial ones and the error is determined. This error, corresponding to the truncated quadratic terms, is taken into account in the next iteration and the process restarts.

It is very difficult to account for a particular pupil during the aberration computation step. Indeed, the analytical calculation of the light propagation should take the particular shape into account and each different pupil would require a particular mathematical development. However, the pupil shape can easily be included in the PSF computation step. Once the images have been computed (for a complete disk pupil), a simple Fourier transform leads to the corresponding pupil. The latter can be masked with the actual VLT (or any other) pupil and a second Fourier transform of the masked pupil produces the corresponding PSF. This way of accounting for a particular pupil slows down the iterative process because the mask is accounted for only during the forward computation (from aberration to PSFs) and not during the backward computation (from PSFs to aberrations). Moreover, it requires two additional Fourier transforms for each

image. However, any kind of pupil can be accounted for in this way. This approach also has the advantage of using the classical Zernike polynomial basis (the most commonly used method), contrary to other one, using the annular Zernike polynomials. The latter consist in redefining the polynomial basis as a function of the pupil shape (annular), and would probably be faster as it would modify both the forward and the backward calculation, but it would only allow to account for annular pupils.

The creation of the mask-pupil requires two steps. The first one is the computation of the optimal inside and outside radii of the mask to ensure the correct sampling of the artificial pupil. An image of the mask (i.e. fig. 8.2) is then loaded and scaled to fit the radius requirement previously defined. The mask generated this way can then be used in the retrieval process.

In order to validate our method to account for a particular pupil (different from a full disk), we have generated PSFs using a simulated aberrated pupil based on the VLT model and applied the retrieval process to compute the aberrations. Those simulations clearly show that including the pupil in the retrieval greatly improves the accuracy on the retrieved aberrations. The residual errors are of the order of 10^{-4} when the pupil is used and larger than 10^{-2} when it is not. These errors mainly concern the real part of the coefficients. The simulations also showed that the pupil influence is larger than the one related to the other parameters (defocus, sampling).

Retrieval parameters

Two sets of parameters appear to be really important for the retrieval process; those related to the spatial sampling (the pixel size, the numerical aperture, the number of pixels per λ/D in the input image, ...) and those corresponding to the focal positions of the images.

The parameters of the first category can generally be determined from the technical characteristics of the instrument and/or from the images themselves. Moreover, the sampling used for the retrieval can easily be compared to the one of the input images by superimposing sections of the two types of PSFs. Fig. 8.3 shows such profiles for the three NACO images (columns 1-3) and their rebuilt counterparts, the two rows correspond to two different (orthogonal) sections.

Sampling parameters	
Pupil diameter[28]	$8 \cdot 0.9 = 7.2$m
Resolution	13.27 mas/pix
Wavelength	2.166μm
Sampling[29]	5.3 pix/(λ/D)
Normalized radius step	0.227
Normalized radius range	$0 - 28.83$
Focal parameters	Best values
Intra focal position	-3320μm
Best focus position	14μm
Extra focal position	2920μm

Table 8.2: Parameters used for the retrieval process.

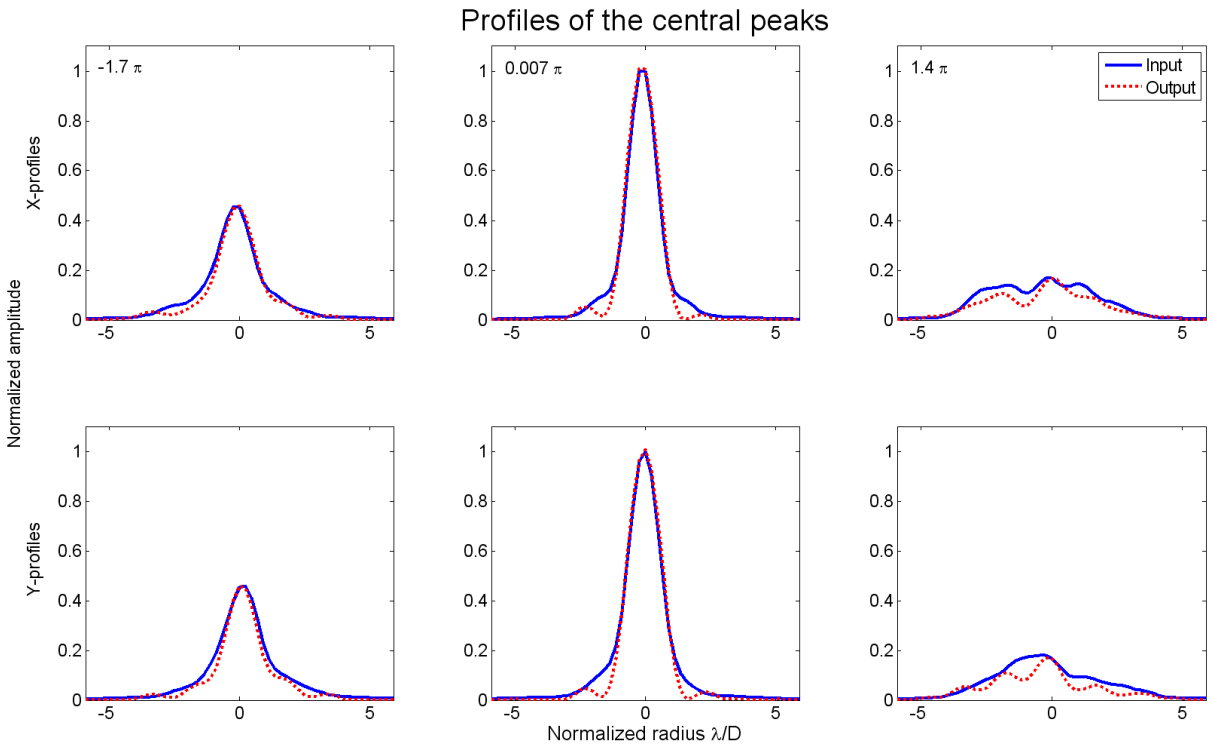


Figure 8.3: Superimposed normalized X and Y sections of the PSFs, corresponding to the central row and the central column of the images. The blue solid curves correspond to the input images and the red dotted ones to the re-computed images. The columns respectively correspond to the intra-focal, the focal and the extra-focal images from left to right. The residual differences between those curves are due to the co-addition related smoothing in the input images.

Wrong sampling parameters can lead to two different effects. The first one is a slowing down of the process convergence. The other effect is much more important, since it consists in perturbations of the retrieved pupil amplitude. Indeed, remember that the β coefficients (see section 4.2.2) allow the description of both the phase and amplitude aberrations, and a bad sampling results in a wrong retrieval of the amplitude part.

A correct determination of the sampling is also particularly important during the computation of the pupil (described in the previous section) in order to ensure its compatibility with the data.

We finally had to use a sampling slightly larger than theoretically expected, which is probably due to the enlargement of the central peak related to the co-addition of the 160 frames (due to the differential tip/tilt between the frame). The parameters actually used are summarized in Table 8.2.

As far as the focal positions of the images are concerned, they are theoretically known since the displacement of the camera is controlled by the observer. However, the accuracy of the mechanism is not infinite ($\sim 10\%$) and the actual displacements may not be the expected ones. In our case, the positions around the focus should be -3mm, 0mm and 3mm, but using these

[28]The VLT mirror has a diameter of 8m but the NACO coronagraph stop masks 10% of this diameter which influences the sampling on the NICMOS InSb Aladdin 3 detector. The effective diameter of $8 \cdot 0.9 = 7.2m$ should thus be taken into account, as far as the sampling is concerned.

[29]The sampling we used does not correspond to the result of the theoretical formula. Indeed, we also tuned this parameter in order to improve the convergence.

values results in a poor fitting between the input images and the computed ones. Hence, we had to tune those parameters to improve the results.

A simple but tiresome procedure can be used to estimate the optimal positions. The first step consists in applying the retrieval procedure using the theoretically applied defocus and then to compare the profiles as in the case of the sampling (fig. 8.3). The relative height of the peaks and of the rings helps in determining the correctness of the focal parameters. Now, using the retrieval software with slightly different focus parameters allows to estimate the way each parameter influences the height of the peaks. It is then possible to tune the focal positions, in a reasonable range (mechanical errors should be smaller than $\pm 10\%$), to reach the best fit between the profiles (Fig. 8.3). The corresponding focal values are summarized with the sampling parameters in table 8.2.

Simulations, performed to validate the way of accounting for the pupil shape, showed that both the sampling and focus parameters should be adjusted as well as possible to guarantee a good accuracy of the results.

Number of retrieved aberrations

We have shown in chapter 6 that it exists an optimal number of aberrations that should be computed, in order to ensure the good accuracy of the retrieval. This optimum is due to a trade-off between the completeness and the noise (photon noise and readout noise) effects. The first one requires as many coefficients as possible while the other one mostly disturbs the higher order coefficients. The optimal number of aberrations minimizes the disturbance due to the noise and the completeness effects, which means that all the important aberrations contained in the images are used in the model and those related to rings that are lost because of the noise are not used.

We apply here an easy way to determine this optimal number, that had previously been suggested. We know that the noise limits the number of PSF rings that can be used during the retrieval. The aberrations corresponding to those outer rings are thus imprecisely retrieved and the errors they contain propagate to the lower order aberrations. On the other hand, using too few aberrations does not allow to represent all the effects present in the images. Using more and more aberrations allows to fit more and more effects until the noise becomes too important with respect to the signal. The accuracy on the values of the aberrations thus progressively increases until the noise limit is reached. This is easily visible in the evolution of the aberration as a function of the number of computed aberrations. Hence, the optimal number of aberrations can be determined by applying the retrieval process several times with an increasing number of aberrations. Their values will first stabilize around the true values and then get disturbed, once the noise limit is exceeded. Fig. 8.4 shows the mean variation of the retrieved aberrations as a function of the number of aberrations used during the computation. The optimal number of coefficients appears to be around 45. This corresponds to eight complete degrees of Zernike polynomials.

8.1.3 Results

Using the parameters detailed earlier, we have applied the retrieval process on the images presented earlier. The aberrations we found, expressed in terms of the general Zernike coefficients β are not easily interpretable. Instead, we present the amplitude and phase of the wavefront as well as the real and imaginary part of the complex pupil that have been computed from the β coeffi-

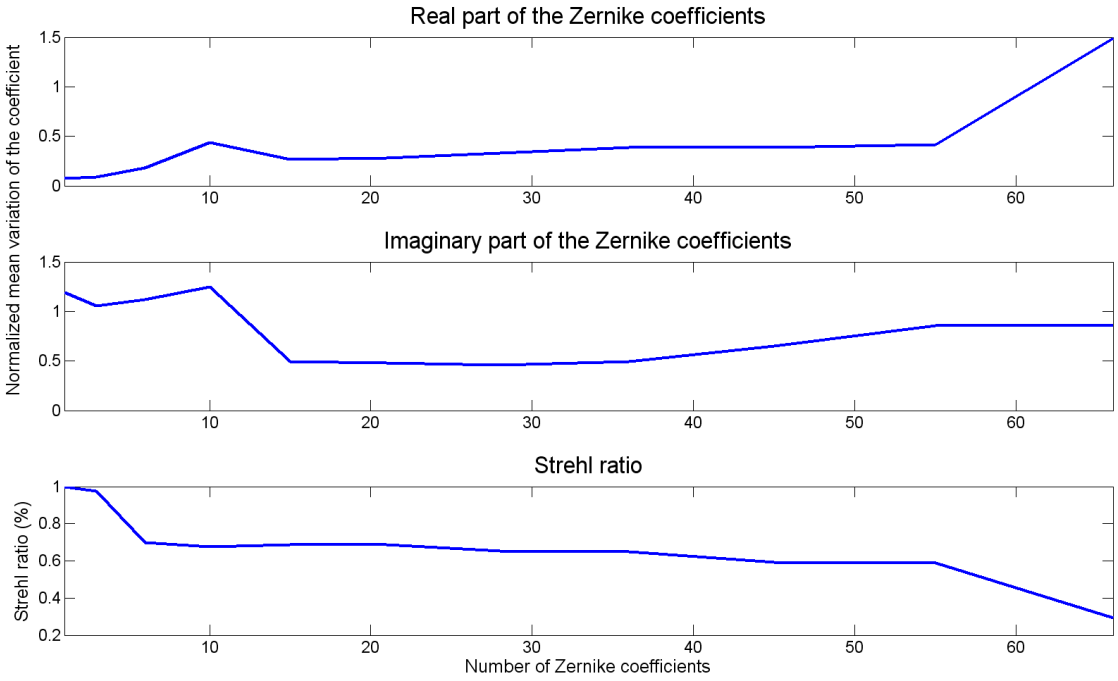


Figure 8.4: Evolution of the mean variation of the aberration values as a function of the number of coefficients used for the retrieval. The values are stable between 11 and 45 coefficients. The optimum is thus to compute the first 45 aberrations. The Strehl ratio given here only corresponds to the static aberrations. The Strehl corresponding to real images smoothed by the atmospheric turbulence will be smaller.

cients (fig. 8.5). The corresponding reconstructed images (fig. 8.7) are compared to the original images (fig. 8.6), whose X and Y profiles were presented in fig. 8.3. The reconstructed images are very similar to the original ones. The peaks and the rings have about the same amplitude, but the contrast between them is larger in the reconstructed images than in the original ones. This is due to the blurring of the original image due to the variation of the tip-tilt between the frames during the exposure. This blurring does not exist in the models. Hence, the aberrations we have measured correspond to the mean static aberrations of the VLT-NACO instrument.

The wavefront phase error presented in fig. 8.5 (top left) can be used to calibrate the speckles that appear on the images obtained with the coronagraph. It should then be possible to remove some of the speckles caused by the static aberrations we have determined, and thus improve the possibility to detect companions to the stellar images, such as exo-planets.

8.1.4 Conclusion

Using the Nijboer-Zernike aberration measurement technique with NACO images allowed to calibrate the static aberrations of the instrument with stellar images. In particular, the residual phase errors that were not corrected for by the adaptive optics have been determined.

Moreover, we have presented a convenient way of taking into account pupils different from a simple disk in the retrieval process without having to modify the aberration decomposition basis. Problems related to the image sampling, that had not been encountered in closed simulations, have also been addressed. Applying the Nijboer-Zernike retrieval theory to a practical case

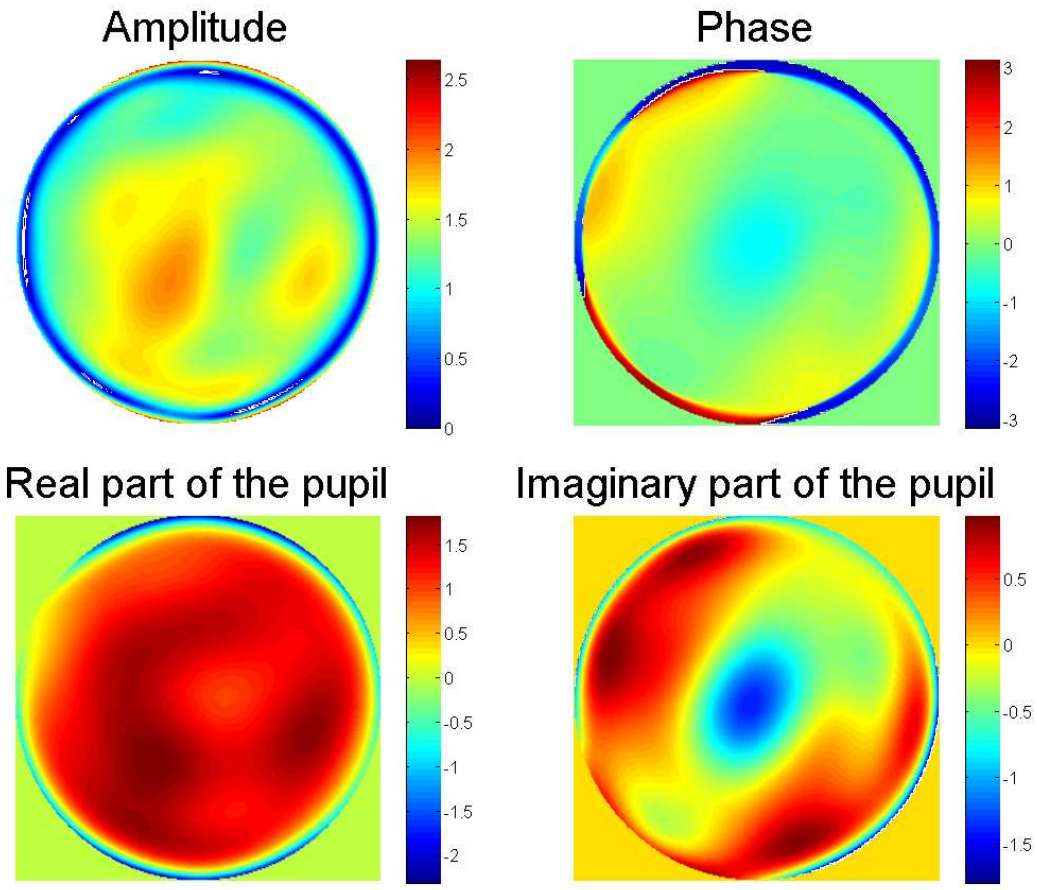


Figure 8.5: Top: Amplitude and phase computed from the retrieved aberrations. Bottom: Real and imaginary parts of NACO's complex pupil. All images present border effects that are probably related to the diaphragm of the NACO coronagraph. The effect of the central obscuration is also clearly visible, especially in the imaginary part of the pupil (bottom right). Finally, one should also notice that the amplitude (top left) is larger than 1 but relatively homogeneous. The first observation may be related to a global normalization of the energy in the PSF whereas the second one indicates that very few "amplitude aberrations" (reflection or transmission problems) are present in the system.

confronted us with the actual difficulties and allowed us to improve the method.

Using the retrieved aberrations, we have reconstructed the phase and amplitude of the pupil. The phase is particularly interesting as it represents the distortion of the wavefront propagating through the optical system (from the star to the detector) and that has not been corrected for by the adaptive optics. This information can be used to improve the quality of the results obtained with the instrument. For example, the NACO four quadrant phase mask coronagraph is supposed to suppress the light of the on-axis star. However, because of the imperfection of the wavefront reaching the mask, the light is not completely destroyed by interferences and residual speckles appear on the images. The phase computed from the measured aberration has been injected in a complete simulator of the NACO coronagraph. It showed that the retrieved phase corresponds to 25% of the total residual in the images, which means that 1/4 of the speckles are generated by the static aberrations of the instrument. This result, based on our approach, should help for the design of the next generation coronagraphic instrument that will be installed on the VLT-UT3: SPHERE.

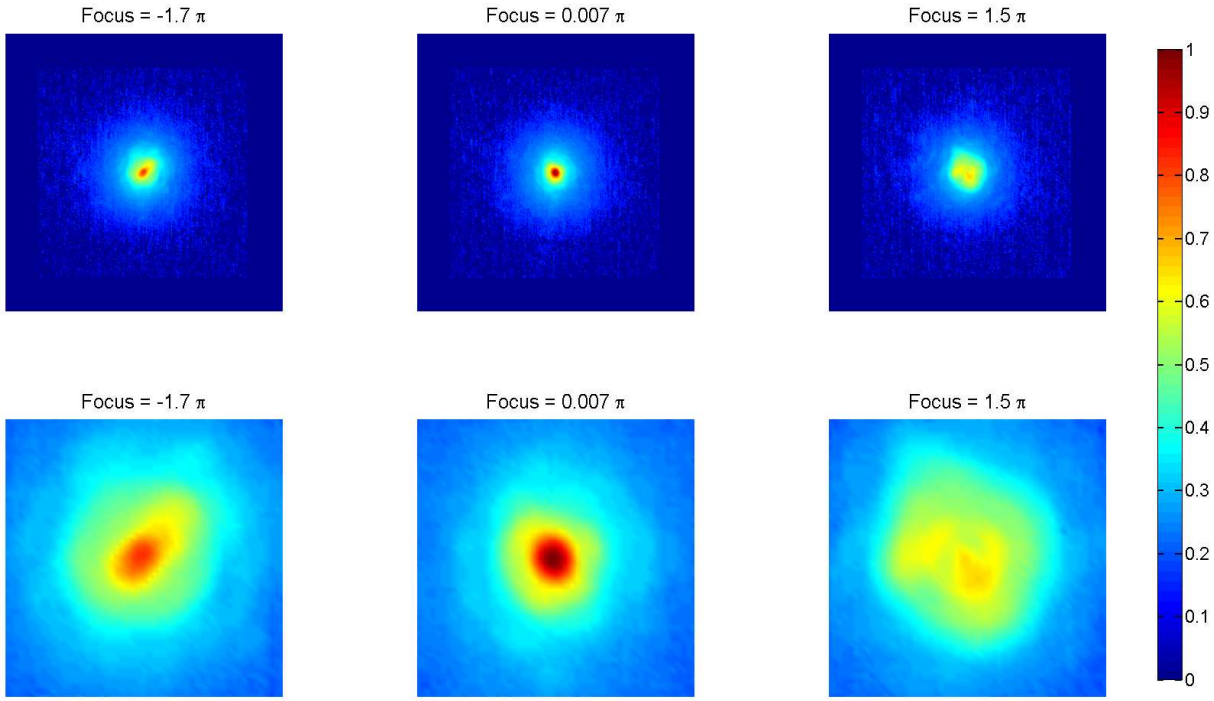


Figure 8.6: Input PSFs used for the retrieval. They were centered in 255×255 pixel tables. The second row shows zoomed views of the central peak. The intensities are represented to the power $1/4$ in order to improve the contrast of the images.

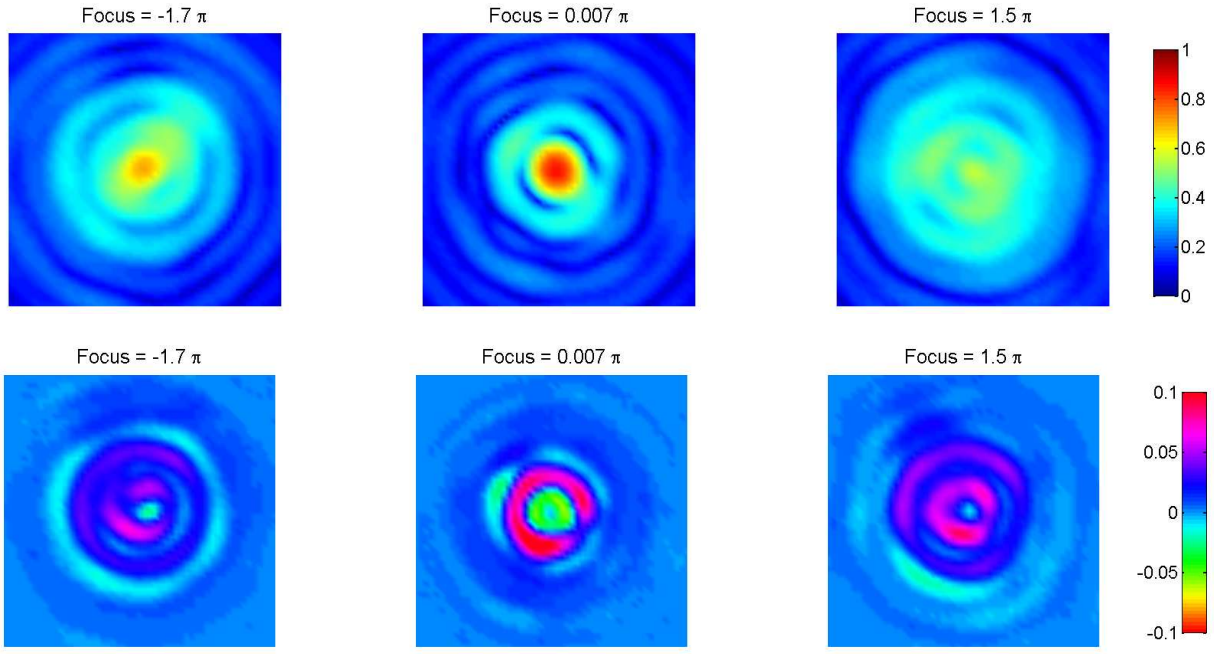


Figure 8.7: Top: Output PSFs computed from the measured aberrations. They are very similar to the input images. The intensities are represented to the power $1/4$ in order to improve the contrast of the images. Bottom: Differences between the input and output PSFs (actual image minus retrieved one). The low residuals ($<10\%$) are due to a blurring in the input images related to differential tip/tilt in the co-added frames.

8.2 Lens testing

The first development of the extended Nijboer-Zernike theory aimed at characterizing the lenses used in UV (193nm) photolithography in order to reach a high engraving resolution. In this section we use our retrieval software in order to determine the aberrations of a simple laboratory lens. The objectives of this manipulation are to test and qualify the method in real conditions as well as to determine whether it would be usable to test the ILMT optical corrector.

Indeed, this optical corrector is very difficult to test for several reasons, the first one being its large size: the entrance diameter is 550mm (first lens) and testing such a device would require a very large collimated or converging beam. We investigate here another solution consisting in testing the lenses, first separately, and then assembled (if possible), using a diverging source in a "2f-2f" configuration.

We first show how to test a simple lens in this configuration and then we review the possibility to extend the method to the corrector. The laboratory setup and the required equipment are detailed before presenting the images and the determination of the retrieval parameters. Finally, the measured phase is compared to the model in order to distinguish between the aberrations related to the testing setup and those inherent to the lens. After that, simulations and details of the measurements for the corrector are presented.

A brief presentation of a convenient way of testing parabolic mirrors with the same type of configuration is also addressed at the end of this section. Small adaptations of the theory allow to increase the dynamic of the method and thus to test aspherical surfaces without any auxiliary optics.

8.2.1 Equipment

The material required to test a simple lens with a "2f-2f[30]" setup is extremely simple. One needs a diverging source to illuminate the "unknown lens" and a detector to record the images. Of course, the lens that should be tested is also needed. These elements are detailed here.

The source

The source we used is a fiber laser at 650nm with adjustable intensity. Such a source is obviously diverging so that, at twice the lens focal length ($\sim$ 40cm), the whole lens is illuminated. The output-end of the fiber is shown on fig. 8.8 (right). The intensity of the laser was set to the minimum in order to avoid saturation of the CCD detector.

The camera

The detector, presented in the background of fig. 8.8 (left) is a SBIG camera equipped with a 2048×2048 pixel CCD sensor. It was installed on a micro-controller that allowed to precisely adjust its position along the optical axis of the system (to adjust the focal position). This controller allows an accuracy of 10μm on the displacement. The characteristics of the camera are summarized in Table 8.3.

[30] As in the previous chapter, a "2f-2f" setup is characterized by a source and a detector located at twice the focal length of the optical device (i.e. around its center of curvature).

Figure 8.8: Left: the tested lens in front of the CCD camera. The latter is installed on a micro-control to fine tune its focal displacements. Right: the fiber source installed on its support. In the background, one can see the laser source itself with its cooling system and power supply.

Pixel size	7.4μm
CCD chip size	2048 × 2048 pix
Full-well capacity	40k e^-
Dark current	0.07 e^-/pix/s
Readout noise	7.9 e^-
Cooling	Peltier

Table 8.3: Characteristics of the SBIG camera used as detector for this experiment.

The lens

The optics we have tested is a simple plano-convex F/2.6 lens made of BK7 glass. It presents an important curvature and is thus expected to have significant aberrations. Its characteristics are summarized in Table 8.4

Diameter	75
Radius 1	103
Radius 2	∞
Thickness	10.1
Focal length	200

Table 8.4: Characteristics of the lens we have tested. It is a spherical (conic number=0) lens made of BK7 glass (n=1.515). The radii of curvature respectively correspond to the first (in) and second (out) surfaces of the lenses. All values are given in mm.

The lens is shown on fig. 8.8 (left). We will see in a further section that the complete lens would create under-sampled images on a detector located at about twice its focal length. We

thus artificially reduced the diameter of the pupil by using a stop. The lens has been masked by a metallic piece with a sharp edge hole. This ensures a total covering of the lens (no ghost light) and a clean hole of 16mm. The complete sampling calculation will be presented in the "model" section hereafter. This mask was placed against the plane side of the lens, and it has not been perfectly centered on the lens (1.5mm along the vertical axis and 0.5mm along the horizontal one). This is not a problem since we aim here at illustrating the retrieval method and not at precisely measuring the lens aberrations. It is also the reason why we did not give too much importance to the alignment of the optical design. These misalignments will only generate additional aberrations and hence will lead to a better illustration of the method retrieval capabilities. We will use an improved numerical model where these misalignments are taken into account to remove their contribution in the final retrieved phase.

8.2.2 Zemax models

Manipulations performed in the laboratory have to be prepared. Hereafter we present the models and simulations that will help for the preparation of the measurements. First of all, we introduce a numerical model of the optimal test setup. It allows to determine the distances between the source, the lens and the camera in an ideal case. A second setup that takes into account the practical limitations encountered in the laboratory is then presented.

Optimal testing setup

Generally, a lens can easily be tested on a bench, a collimated beam is used to illuminate it and the lens then creates an image on a detector. This image can then be analyzed to unveil the defects of the lens. However, some lenses are too big to be tested with a parallel beam (they would require a high quality collimator as large as the lens). We investigate here another setup, using a diverging source and a CCD detector located symmetrically on both sides of the lens at a distance of around twice its focal length. This setup has the advantages of theoretically allowing to test lenses of any size, and it also requires the minimal size to test a lens with a non-collimated beam. It is presented in fig. 8.9. This type of system can be used to test any optical lens on a bench provided that is focal length is short enough (to fit in the bench length). The same type of scheme could thus theoretically be used to test the lenses of the ILMT corrector. This possibility is reviewed in section 8.2.6.

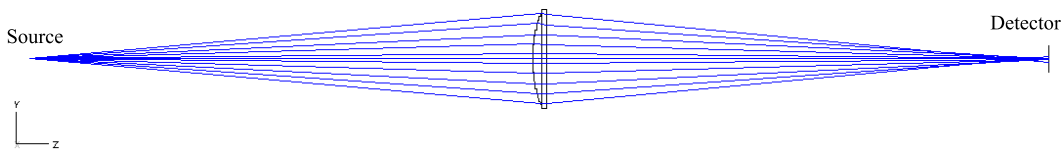


Figure 8.9: Zemax model of the optimal testing setup for the lens presented before. The fiber source is on the left and the camera is on the right.

The positions of the source and of the camera have been determined with the numerical model, in such a way that they minimize the defocus aberration. They are summarized in Table 8.5 along with the experimental parameters actually used.

	Optimal	Experimental
Source [mm]	383	425
Camera [mm]	383.4	368
Δ_f [mm]	± 0.56	± 2
Diameter [mm]	75	16

Table 8.5: Optimal and experimental parameters for the test of the lens. "Source" is the distance between the source and the first surface of the lens, "Camera" is the distance between the second surface, and the camera and Δ_f is the displacement that should be applied to the camera to obtain one wave of defocus. "Diameter" is the diameter of the unobscured part of the lens.

Practical testing setup

The actual optical scheme is slightly different from the optimal one (fig. 8.10). Indeed, the theoretical positions are not easily usable on the bench and we have opted for rounded values. Moreover, using the theoretical optimal position of the camera leads to a bad sampling of the PSFs on the detector. The sampling of the images is given by

$$S_a = \frac{F}{D}\frac{\lambda}{w_{\text{pix}}} \tag{8.1}$$

where S_a is the sampling of the PSF (i.e. the number of pixels per λ/D), F is the "pseudo-focal length" (i.e. the distance between the lens and the detector: 368mm), D is the diameter of the lens (75mm), λ is the wavelength of the source (650nm) and w_{pix} is the pixel size (7.4 μm). These values give a sampling of 0.42 pixel per λ/D which is very bad. We have thus used a diaphragm, as described before, to reduce the pupil size to 16mm. This corresponds to a sampling of 2.02 pixels. In our case, stopping down the lens is not a problem since we just want to illustrate the measurement principle. However, in case the whole lens should be tested, it is possible to increase the sampling by moving the camera away from the lens. In this case, the "2f-2f" optimal size for the optical system would not be respected anymore, but it may not be a problem provided the bench is sufficiently long.

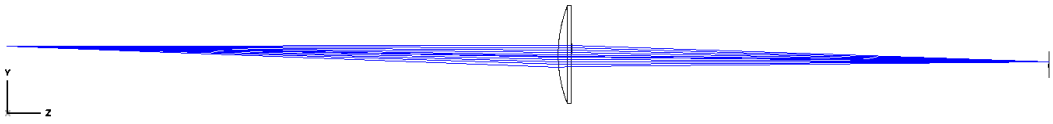


Figure 8.10: Zemax model of the actual testing setup. The lens has been stopped down in order to respect the sampling theorem (2 pixels per λ/D). Moreover the alignment of the diaphragm was done roughly, which induces additional aberrations. This is accounted for as much as possible in this model.

The main aberrations related to this optical system (perfectly aligned) are given in Table 8.6. Knowing these theoretical values it will be possible to estimate the actual efficiency of the retrieval method.

Aberrations	Values
Z_1	-0.5536
Z_4	0.2725
Z_{11}	0.0366

Table 8.6: Main aberrations (among the first 36) of the system (perfectly aligned) presented in fig. 8.10. They are expressed in wave units at 650nm. All other aberrations are smaller than 10^{-5}.

8.2.3 The images

The simulations presented before helped us in determining the best way to illuminate the lens and to position the detector in good conditions to perform the test. However, there is a difference between simulations and actual laboratory tests. We tried to fit our model as much as possible, but it was not possible, and we thus made some new simulations aiming at representing the real situation. The new model has also been introduced above, the actual testing setup is shown in fig. 8.11.

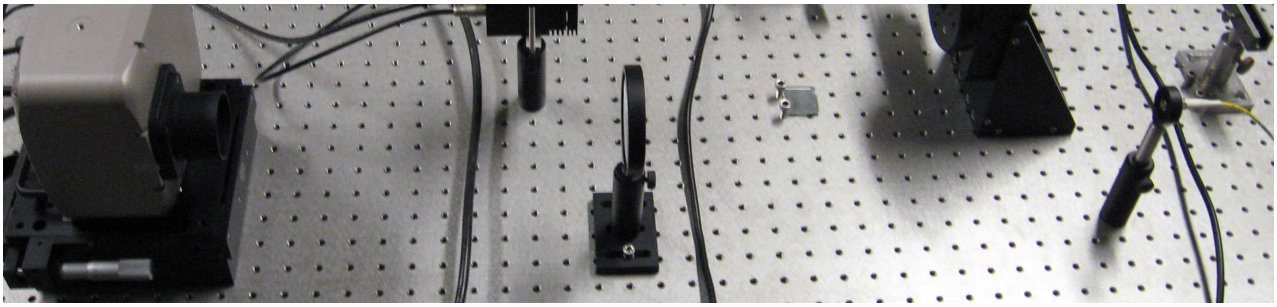


Figure 8.11: The complete testing setup. The distance between two holes on the table is 2.5cm.

The optical setup presented above aimed at having the lens imaging the source on the camera. In order to increase the signal to noise ratio, we have taken 50 pictures of 0.4s of exposure for each position of the camera, those frames were then co-added. "Skies"[31] have also been acquired and removed from the data. The resulting images have been centered and cropped to 255×255 pixel images. We then normalized the intensities of all the images by the maximum intensity of the focused image. The pictures finally available for the retrieval are shown in fig. 8.14 (top line).

[31] A "sky" is an image taken in the same conditions as the measurement pictures by with the laser source turn off. Such data allow to remove the light background from the target images.

8.2.4 Aberration retrieval

The images obtained in the laboratory have been used to measure the aberrations of the lens. In this section, we detail the retrieval process and the determination of the parameters, as we did in the NACO case. However, many aspects are very similar and they will just be reminded here.

Retrieval parameters

Contrary to the case of NACO where the parameters where not exactly known, in the case presented here, we are certain of the focal shift values within an accuracy of a few tens of micrometers. Errors of this order are not significant for the retrieval process. On the other hand, the sampling parameters were more difficult to determine accurately.

Indeed, the position of the CCD inside the camera is not perfectly known, but has been determined experimentally. We first measured the focal length of the lens and then adjusted the position of the camera to reach the best focus (smallest image). Once the camera was correctly located, we compared the focal length and the distance between the camera and the lens to estimate the location of the CCD detector inside the camera (with an accuracy of a few millimeters). The total distance between the lens and the detector is thus known within a few millimeters accuracy. Moreover, the diameter of the mask was measured to a fraction of a millimeter. Combination of those possible errors may lead to a theoretical sampling slightly different from the actual one.

For example, an error of 5mm (half of the lens thickness) on the focal length and 0.5mm on the pupil diameter would lead to a sampling of 2.11 pix/λ/D. The optimal parameter we found (2.09 pix/λ/D) is thus consistent with the accuracy of the measurements.

This value has been determined in the same way as for the NACO retrieval. We superimposed the profiles of the input PSFs with those of the reconstructed ones (fig. 8.12) and checked that the Full Width at Half Maximum (FWHM) of the model fits the one of the input. Moreover, the retrieval process converges faster when the sampling parameter becomes close to its optimal value. All the parameters used for the retrieval are summarized in Table 8.7.

Number of retrieved aberrations

As in the NACO case, we determined the optimal number of aberrations by applying the retrieval process several times with an increasing number of aberrations. The values of those aberrations first stabilized around their true value and then perturbations arose on the low order aberrations. Study of the evolution of the aberrations allows to determine the optimal number of aberrations that should be computed. Fig. 8.13 presents the mean evolution of the real and imaginary parts of the low order aberrations and the Strehl ratio computed from the β coefficients. From these graphics, we deduced that 36 aberrations is the optimal number that can be retrieved with a good accuracy. The corresponding Strehl ratio is about 80%.

It is interesting to note that the retrieval process converges as long as less than 78 aberrations are computed as in the case of the NACO analysis. This is probably related to the readout noise of the cameras used in the two cases. In the NACO cases, the readout noise is $15e^-$ and the images are composed of 160 frames. In the lens case, the readout noise is $7.9e^-$ and the images are composed of 50 frames. The signal to noise ratio varies as the square root of the number

Sampling parameters	
Focal length [mm]	368
Pupil diameter [mm]	16
Wavelength [μm]	0.650
Sampling [pix/(λ/D)]	2.09
Normalized radius step	1.03
Normalized radius range	$0 - 262.87$
Focal parameters	Best values
Intra focal position [μm]	-2000
Best focus position [μm]	0
Extra focal position [μm]	2000

Table 8.7: Parameters used for the lens aberration measurements. The focal length corresponds to the system, not to the lens alone, it represents the distance between the lens and the detector (in the camera). The sampling we used does not correspond to the result of the theoretical formula. The value given here corresponds to a slightly different focal length and pupil diameter within measurement accuracy and results in a better convergence of the process.

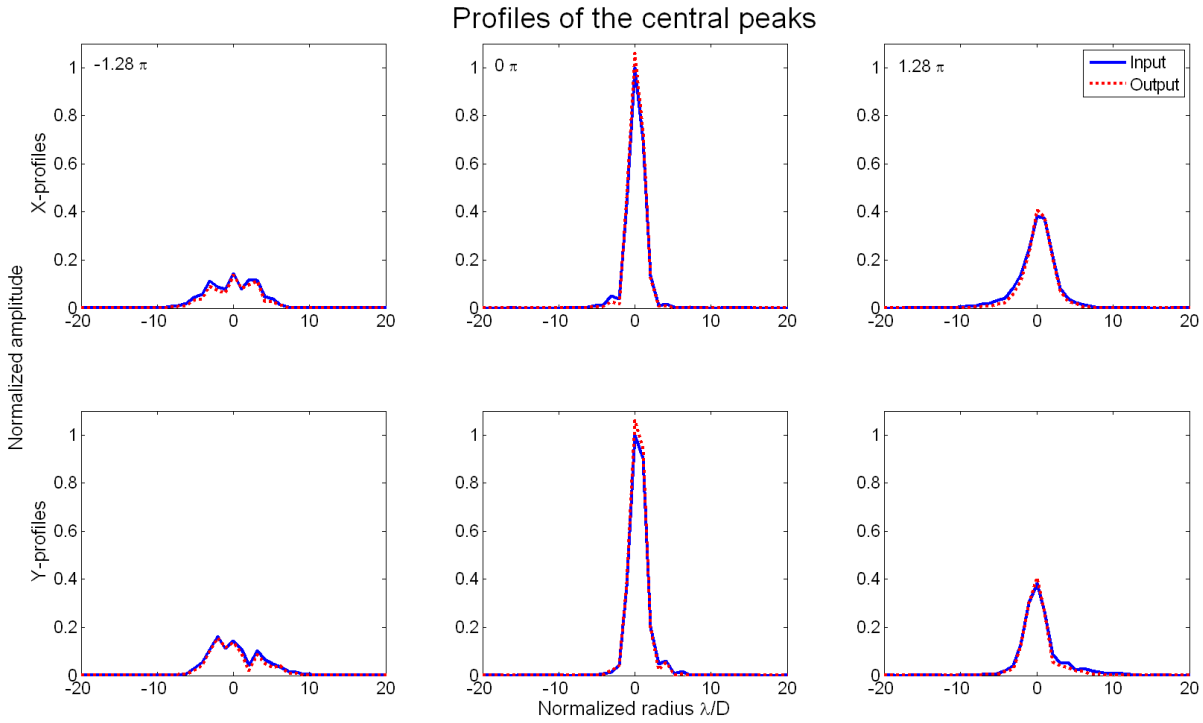


Figure 8.12: Superimposed normalized X and Y sections of the input PSFs and the model PSFs (reconstructed after retrieval), corresponding to the central row and the central column of those images. The blue solid curves correspond to the input images and the red dotted one to the recomputed images. The columns respectively correspond to the intra-focal, the focal and the extra-focal images (from left to right). The profiles are perfectly similar, permitting a high confidence in the results obtained. Let us note that no smoothing related to tip/tilt between the 50 frames appears, as it was the case for NACO, since the frames were all acquired in the same conditions.

of frames, and the noise ratio thus decreases as the inverse of the square root of the number of

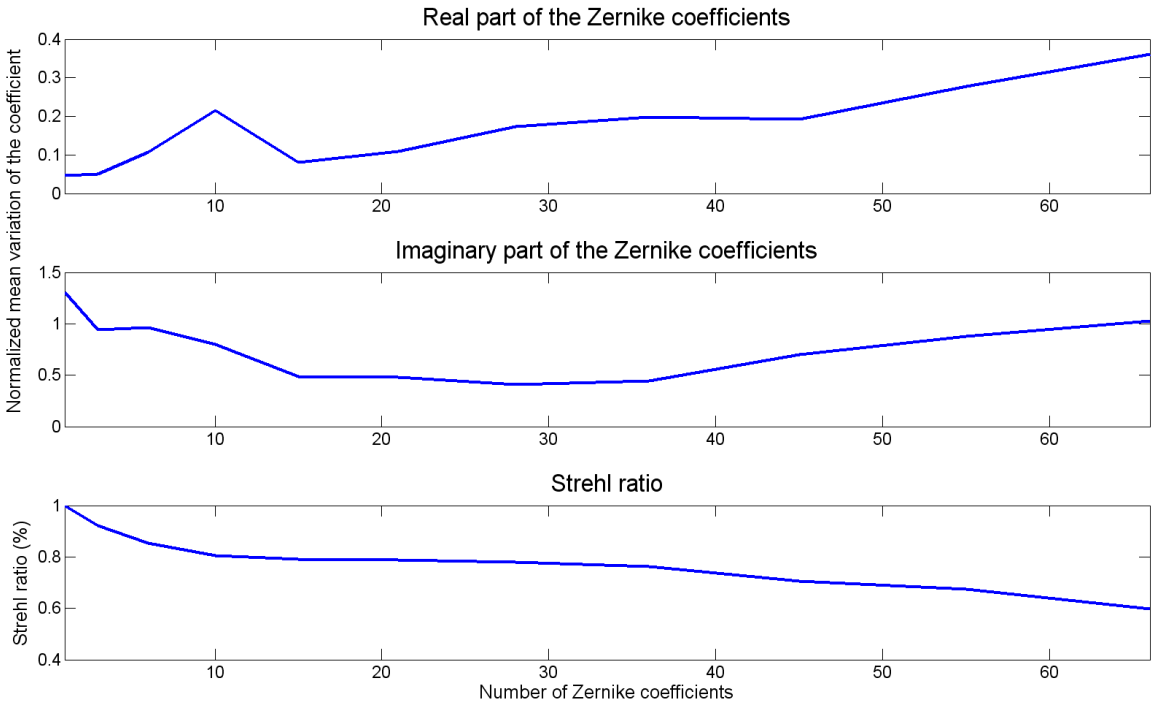


Figure 8.13: Evolution of the mean variation of the aberrations as a function of the number of coefficients used in the retrieval. The values are stable between 11 and 36 coefficients. The optimum is thus to compute the first 36 aberrations.

frames, we have

$$\frac{15}{\sqrt{160}} \sim \frac{7.9}{\sqrt{50}} \tag{8.2}$$

The noise ratio is equivalent in both cases, which explains why the convergence limits are the same in both cases. However, in the case of NACO, 45 aberrations were finally retrieved instead of 36 in the case presented here. This means that more rings are visible on NACO related images.

8.2.5 Results

Using the parameters presented before, the Nijboer-Zernike aberration measurement process has been used to compute the 36 first β coefficients describing the lens wavefront error. The input PSFs (fig. 8.14) are compared with those reconstructed from the aberrations computed (fig. 8.15). The differences between those PSFs (fig. 8.16) is very small (lest than 7%), which makes us confident in our results.

The pupil decomposition in the basis of the β coefficients is different from the classical one and can seem difficult to interpret. However the wavefront is the same as in the classical α coefficient basis, we thus present the results under the form of the amplitude and phase computed from the complex pupil function (fig. 8.17). The latter is computed through the pupil equation as a function of the β_m^n. Apart from border effects, the amplitude is almost uniform, which means that the tested lens contains almost no amplitude aberrations. The phase is also affected by border effects but it also presents identifiable aberrations.

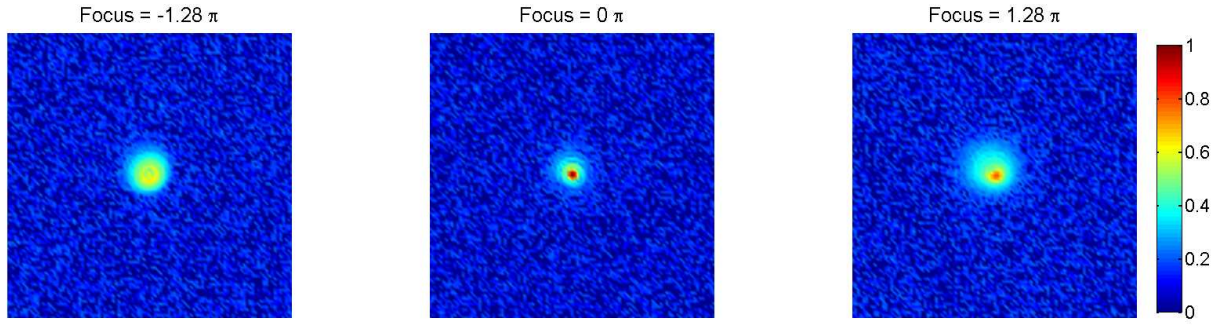


Figure 8.14: Top: Input PSFs used for the retrieval. They have been centered in 255×255 tables. The intensities are represented to the power $1/4$ to improve the contrast of the images.

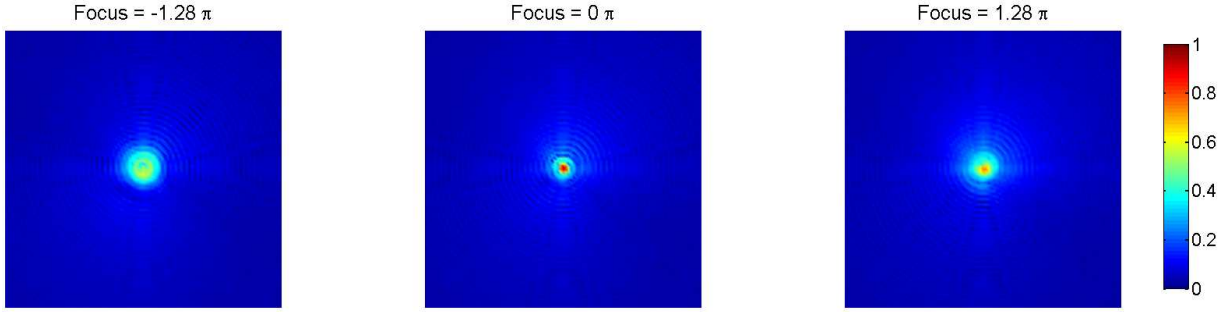


Figure 8.15: Output PSFs computed from the measured aberrations. They are very similar to the input images. The intensities are also represented to the power $1/4$ to improve the contrast of the images.

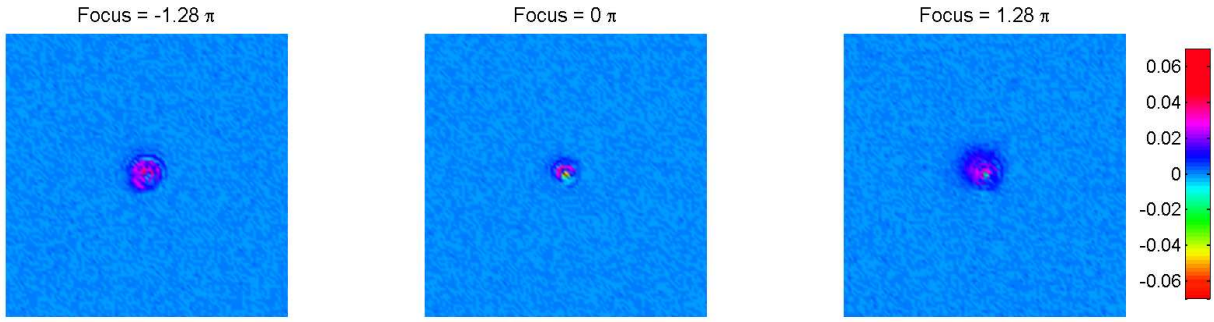


Figure 8.16: Differences between the input and output PSFs. The errors between the model and the images are smaller than 7%.

Let us remind that the alignment of the system was not perfect. We have modeled the phase corresponding to the measured alignment defects presented earlier (fig. 8.18 left). This phase is subtracted from the reconstructed one in order to isolate the actual aberrations of the lens from those due to the optical setup (fig. 8.18 right).

The remaining phase error corresponds to an RMS surface deviation of $\lambda/2.66$. Thorlabs (the lens manufacturer) does not specify the aberration of its lens but it is probably a common $\lambda/4$ rms lens. Moreover, it is quite thick and with a small f-number, the curvature of its surface is thus important and the lens could have more aberrations than expected. This is the reason why this lens was chosen for the test, and the result we found is thus coherent. The phase error is mainly composed of tip/tilt, coma and spherical aberrations.

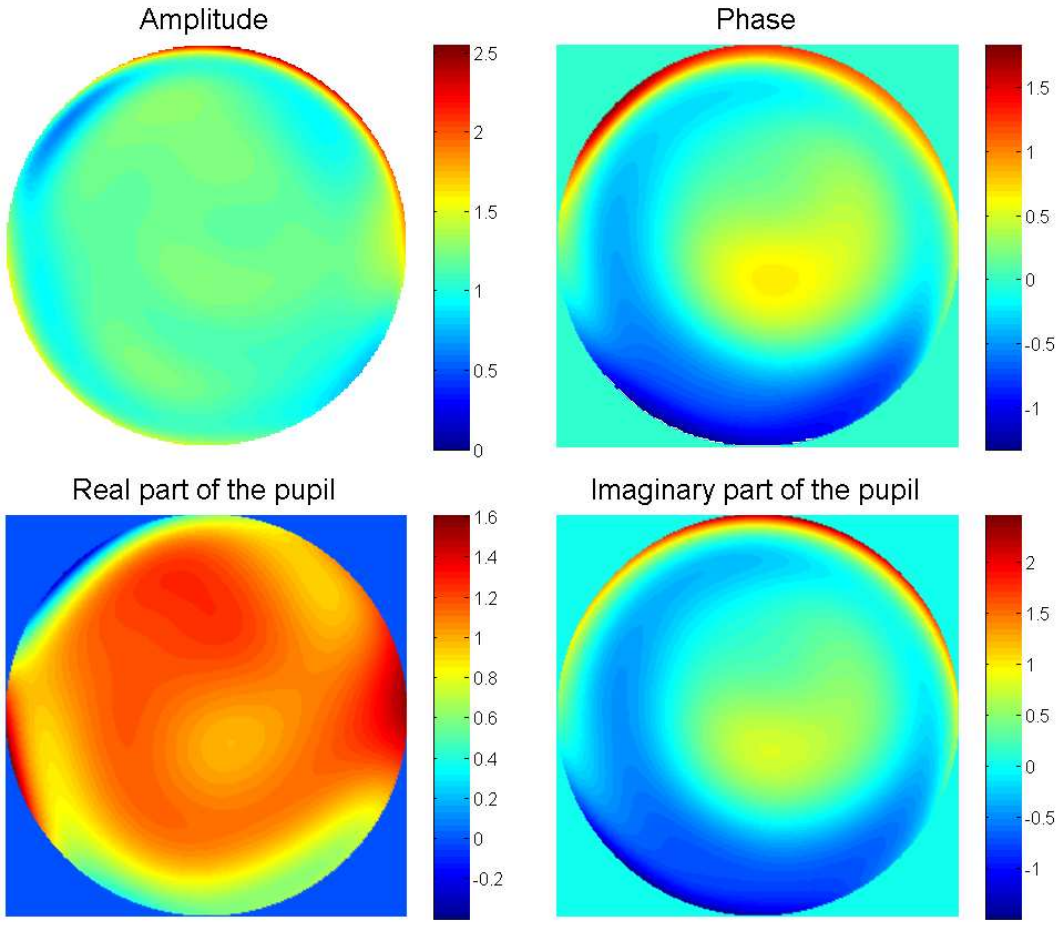


Figure 8.17: Top: Amplitude and phase computed from the retrieved aberrations. Bottom left: Real and imaginary parts of the complex pupil. Small border effects are visible. The amplitude is uniform and its value around 1.

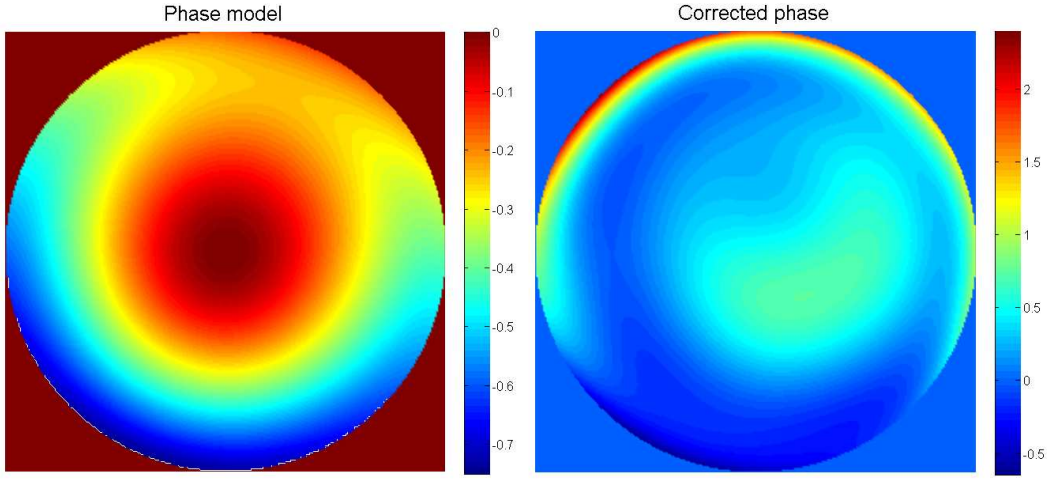


Figure 8.18: Left: Phase surface obtained from the numerical model of the system. It is used to remove aberrations related to the setup from the total measured aberrations. Right: Difference between the retrieved phase and the model on the left. This residual phase corresponds to the real aberrations of the lens. An almost complete ring of spherical aberration as well as signs of coma appear on this phase.

8.2.6 First additional application: Test of the ILMT corrector

The results obtained in the previous section are encouraging. They demonstrate that testing the ILMT corrector lenses should be feasible. In this section we determine the parameters related to the characterization of those lenses and those for the complete corrector.

Testing the lenses separately

We have presented a reliable and (fairly) straightforward way of testing lenses. The possibility to test the ILMT corrector lens in the same way is reviewed here. The main difficulty encountered in the previous lens test was to obtain an image sampling that respects the Nyquist-Shannon theorem. We saw that the sampling on the CCD is given by

$$S_a = \frac{F}{D}\frac{\lambda}{w_{\text{pix}}} \tag{8.3}$$

It is possible to determine a minimum F/D value required to reach a sampling of 2 pix/λ/D given a fixed wavelength (0.650μm) and pixel size (7.4μm)

$$\frac{F}{D} \geq 2\frac{\text{pix}}{\lambda} = 22.77 \tag{8.4}$$

The F/D ratios corresponding to each lens composing the corrector are summarized in Table 8.8, along with their other parameters.

	Diameter	R_1	R_2	t	F	P_{s}	P_{c}	Δ_f	F/D
L1[32]	550	336.63	336.63	65	~ 5000				
L2[33]	250	2002.85	213.32	15					
L3	250	1619.63	-601.72	25	850	1700	1689.4	$\pm$0.972	6.76
L4	200	193.93	220.68	30	1950	4000	3997.5	-9.201/+9.244	19.99
L5	125	-472.39	-251.54	15	1020	2000	2010.5	-5.361/+5.389	16.08

Table 8.8: Characteristics of the lenses composing the ILMT corrector. All these lenses are spherical (conic number=0) and made of "N-BK7" glass. R_1, R_2 are the radius of curvature of the first (in) and second (out) surfaces of the lenses respectively, t is the thickness of the lenses, P_{s} is the distance between the source and the first surface of the lens, P_{c} is the distance between the second surface and the camera and Δ_f is the displacement to apply to the camera to obtain one wave of defocus. All these values are only indicative. The F/D values are too small to reach the required sampling. All distances in this table are given in millimeters.

The first lens is not really converging, its focal length is about 5m which is far too long for an optical bench and it can thus not be tested with our method. The second one is even worse since it is diverging and cannot create images on a detector. The three other lenses have shorter focal lengths but they still require an optical bench whose length is between 3 and 8m which is not realistic. Moreover their F/D ratios are too small to ensure a good sampling. As in the case

[32] L1 has such a large focal length that it is not realistic to use our method to test it in the laboratory. That is why we did not calculate the positions of the source and camera.

[33] L2 is a diverging lens that cannot generate images on the detector. Hence it cannot be tested in the same way as the other ones.

of the lens presented earlier, a correct sampling could be obtained by obscuring the outer part of the lenses. In any case, the distances involved makes it unrealistic to envision testing those lenses with our "2f-2f" technique presented earlier. The main aberrations corresponding to each lens for such optical systems are given for information in Table 8.9.

Aberrations	L3	L4	L5
Z1	-23.20	-60.12	-8.3
Z4	560 10^{-6}	35 10^{-6}	-69 10^{-6}
Z11	10.38	28.09	3.75
Z22	750 10^{-6}	1.08	14 10^{-3}

Table 8.9: Main aberrations (among the first 36) corresponding to the corrector lenses. They are expressed in wave units at 650nm. The other aberrations are smaller than 10^{-5}.

Testing the complete corrector

Instead of testing the lenses separately, we investigate here the possibility to test the complete corrector. Obviously, the lenses have the same size as in the separated case, but they are tilted and/or decentered with respect to each other, and the combination of all the lenses results in a smaller focal length of about 93cm, which makes it easier to test on a bench.

The problem with this corrector is that it has been designed to receive a converging beam, and not a diverging one as it is the case with our "2f-2f" scheme. Illuminating the whole first lens with a diverging beam would result in a loss of light in the corrector since the lenses do not compensate enough the initial divergence of the beam. The light would thus not propagate through all lenses and would not come out of the corrector. In order to avoid any ghost light, the first lens should be stopped down to 50mm. Even if the lenses are not completely used (illuminated) in this way, such a test would allow to detect bad relative positions between the lenses.

The optical setup model is presented in fig. 8.19 and the corresponding parameters are summarized in Table 8.10.

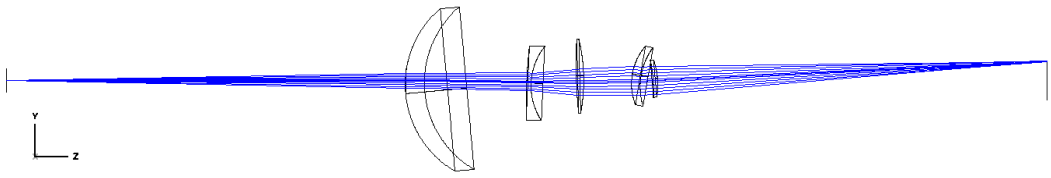


Figure 8.19: Testing setup for the complete corrector. The fiber source is on the left and the camera on the right. The corrector is stopped down to a diameter of 50mm to avoid ghost light due to the low convergence of the first lenses.

The total required distance is about 3.5m, which is still very large but could be realizable, and the corresponding F/D is large enough to ensure a good sampling. It is thus possible to test the complete corrector quite rapidly. These are good news since its characterization is mandatory to be certain that its specifications are respected.

Diameter [mm] - L1	550
Diameter [mm] - diaphragm	50
Focal length [mm]	930.6
Source [mm]	1350
Camera [mm]	1261.6
Δ_f [mm]	± 3
F/D	25.23

Table 8.10: Optical characteristics of the ILMT corrector testing setup. The focal length is the distance between the last lens and the detector.

8.2.7 Second additional application: Test of parabolic mirrors

A spherical mirror can easily be tested with a source and a detector located at its center of curvature. This is due to the ease of creating a spherical reference wavefront (a simple point source being sufficient). However, this is not the case anymore for parabolic mirrors. When such a mirror reflects a spherical wavefront, the focalization is not perfect and a so-called "spherical" aberration is generated. If the asphericity of the mirror increases, this spherical aberration can become dominant and mask the smaller aberrations. Offner (1963) presented an optical device, composed of two lenses (called Offner lenses or null lenses) designed to compensate for the spherical aberration due to the mirror. The Offner's lenses enable the measurement of parabolic mirrors from their center of curvature, just as if they were spherical. Nevertheless, the design of Offner lenses can be tricky and the compensator could perturbate the measurement if it is not correctly shaped.

In case such null lenses would not be used, the spherical aberration would be so large that it would saturate the measurement system (Roddier, Foucault,...), and the other, interesting aberrations of the mirror would be completely masked. The Nijboer-Zernike approach does not require a reference surface and hence does not risk any saturation. Adaptations can easily be done so that the NZ method provides a large dynamic of measurements. In this section we present the modifications we brought to the NZ theory presented before and how the "2f-2f" setup can be used to measure parabolic mirrors from their center of curvature without any auxiliary optics.

Theory of the high dynamic Nijboer-Zernike retrieval method

The development of the Nijboer-Zernike theory presented in chapter 5 (section 5.2) was based on the assumption that β_0^0 was the dominant term. This term was also supposed to be real and positive. Hence, all other aberration coefficients were much smaller than β_0^0. The complex amplitude of the PSF can be seen as a series of aberrations, which, in the classical theory, scale relatively to the β_0^0 coefficient.

Remember that the PSF is a sum of quadratic and crossed aberration terms. The scaling with respect to β_0^0 is mathematically expressed by neglecting, in the analytical development, the terms that do not involve β_0^0. The iterative process then consists in minimizing the error caused by the omission of these terms.

It may happen that another aberration (than β_0^0) would be larger than the others. This is particularly the case when a "2f-2f" setup is used with a parabolic mirror. The spherical aberration can then become dominant, and, the PSF being mostly shaped by this aberration,

the corresponding terms of the sum should not be neglected.

We introduce here our generalization of the classical retrieval theory to any dominant aberration. The detailed equations are presented in appendix D. We also temporarily relax the "real and positive" hypothesis stated earlier. Indeed, assuming that β_0^0 is real and positive was consistent, in the previous case, since β_0^0 represents the uniform term both in amplitude and phase. Since this global piston phase term does not influence the PSF shape, it is completely impossible to retrieve. The imaginary part of β_0^0 can thus be set arbitrarily to zero. Moreover, its real part corresponds to the global transmission of the optical system, hence it can be neither negative nor larger than 1. These particularities are not true anymore in the case where another aberration is dominant and we thus have relaxed these hypotheses.

The basic principle of our generalization is quite simple. The dominant aberration, that should not be omitted, is chosen to undergo the same special treatment as β_0^0 does in the classical theory. In all retrieval equations, the β_0^0 aberration term is replaced by another one that is expected to be very large, typically the spherical aberration. In this way, we develop the Zernike aberration basis around this largest aberration that is used as a new normalization. This is equivalent to rearranging the aberration terms so that the dominant one is in the first position. The following of the treatment is exactly the same provided that each occurrence of β_0^0 is replaced by the dominant β_n^m.

This change of basis is very simple when the dominant term is radial ($n = N, m = 0$, $\beta_n^m = \beta_N^0$), which was the case for β_0^0. The equations generalized for a dominant aberration β_N^0 are given in appendix D. The case where $m \neq 0$ is more complex since the corresponding β is also associated with an azimuthal modulation, which leads to crossed terms in sine or cosine functions ($\cos m \cdot \cos m'$, $\cos m \cdot \sin m'$, $\sin m \cdot \cos m'$, $\sin m \cdot \sin m'$). This case will not be detailed here.

Because of the generalization of the theory to a complex β_N^0, the variables of the classical method ($\beta_0^0 \Re(\beta_n^m)$ and $\beta_0^0 \Im(\beta_n^m)$) are replaced by

$$\begin{aligned} A_n^m &= \Re(\beta_N^0)\Re(\beta_n^m) + \Im(\beta_N^0)\Im(\beta_n^m) \\ B_n^m &= \Re(\beta_N^0)\Im(\beta_n^m) - \Im(\beta_N^0)\Re(\beta_n^m) \end{aligned} \tag{8.5}$$

that can be written in matrix form

$$\begin{pmatrix} A_n^m \\ B_n^m \end{pmatrix} = \begin{pmatrix} \Re(\beta_N^0) & \Im(\beta_N^0) \\ -\Im(\beta_N^0) & \Re(\beta_N^0) \end{pmatrix} \begin{pmatrix} \Re(\beta_n^m) \\ \Im(\beta_n^m) \end{pmatrix} \tag{8.6}$$

which lead to the classical variables when $\Im(\beta_N^0) = 0$ and $N = 0$. Let us note that the same equations link the sine and cosine related coefficients and what is explained here is valid for both cases. Solving the linear systems as in the classical theory gives the values of A_n^m and B_n^m.

As in the classical case, solving the first linear system for $m = 0$ allows to determine $|\beta_N^0|$ and then β_N^0 if an assumption is made on the relation between the real and imaginary parts. We have shown earlier that the classical α coefficients mostly correspond to the imaginary part of the β coefficients. We will thus assume that the real part of the dominant aberration is negligible and that β_N^0 is purely imaginary. With this assumption and inverting the matrix equation one can find each β_n^m.

This modified theory has been tested with the same type of simulations as those presented in chapter 6. Using the same set of aberrations but including tip/tilt and a larger spherical

aberration, retrieval tests have been performed for 36 and 231 aberrations coefficients. The results were similar to those obtained for the classical theory (without noise).

In pure simulated cases we thus manage to retrieve aberrations as large as 10 waves along with other smaller than a thousandth of a wavelength with a very high accuracy. Such a method would thus be very interesting to test aspherical surfaces where large spherical aberration are expected. We present hereafter a testing setup that could be used for such a task. Laboratory testing of parabolic mirrors is one of our projects for the near future.

Testing setup for small parabolic mirrors

Using a variation of the "2f-2f" setup presented earlier, it should be possible to test the quality of a parabolic mirror. A schematic of the principle is illustrated in fig. 8.20. The source and detector are located near to the center of curvature of the mirror. In that case, the dominant aberration should be the spherical one (β_4^0) provided that the source and camera are sufficiently near to the optical axis of the mirror. Let us note here that, in the model presented below, the mirror is aberration-free. The detector is used to acquire images around the best focus position and the improved NZ retrieval method is applied in the dynamic mode using β_4^0 as the dominant term. The typical aberrations, expected from such a setup are summarized in Table 8.11 and the typical wavefront phase is represented in fig. 8.21. The Z_1 aberration should not be taken into account since it only represents a global piston term that does not influence the images. Instead, the other aberrations should be compared to 1 ($\Re\beta_0^0 \sim 1$) in order to decide whether they are dominant or not. In this case, the term β_0^0 can still be considered as dominant since the aberrations Z_6, Z_{11} are smaller than 1. However, in case the parabola would have a smaller F/D (the ILMT mirror for example), these aberrations would become larger than 1 and the generalized method should be used.

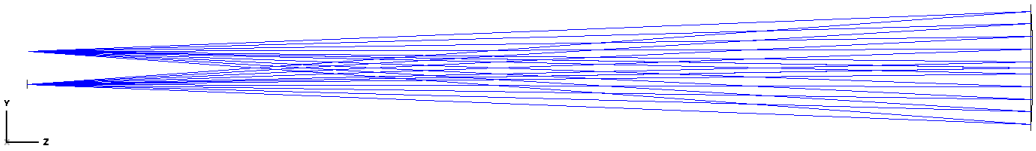


Figure 8.20: Typical testing setup for parabolic mirrors. The source and camera are located near the center of curvature of the mirror but slightly off-axis. In this scheme, the source is above the camera. The distance between the two is 4cm. The mirror has a diameter of 150mm and a focal length of 600mm.

Aberrations	Values
Z_1	0.795
Z_6	0.333
Z_{11}	**0.355**

Table 8.11: Typical aberrations expressed in wave rms (633nm) expected from the numerical model for the setup presented in fig. 8.20.

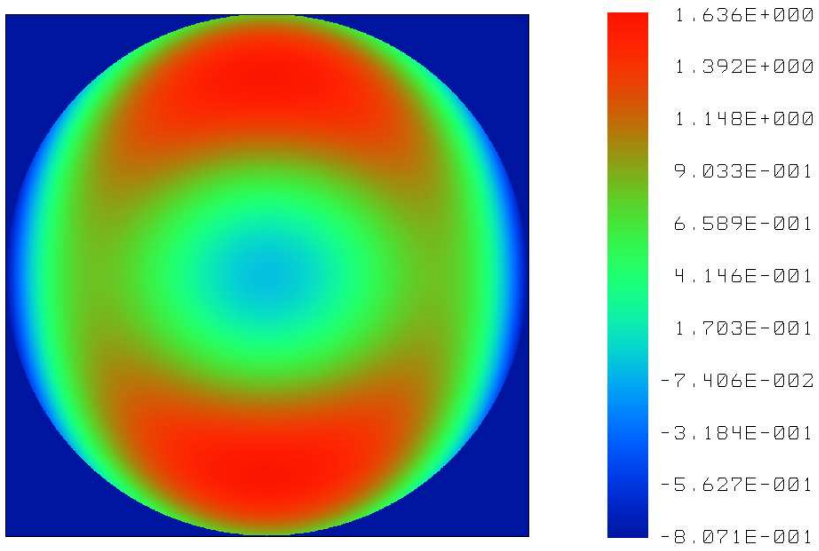


Figure 8.21: Typical phase of the wavefront corresponding to the setup presented in fig. 8.20. The corresponding P-V error is 2.44 waves whereas the rms error is about 0.5 wave.

8.2.8 Conclusions

We have presented a practical and easy way of testing small lenses on an optical bench using the Nijboer-Zernike retrieval method with a "2f-2f" optical setup. The main difficulty being related to the sampling of the images, the size of the tested lens has to be limited in order to ensure that the system respects the Nyquist-Shannon sampling theorem. The possibility to use this method to characterize the ILMT corrector has also been reviewed. Even if testing the lenses separately with this method does not seem realistic, interesting results could be obtained by testing the complete corrector.

We have also briefly introduced the possibility to use a "2f-2f" setup to characterize parabolic mirrors without any auxiliary optics (Offner's lenses). It should also be possible to test large aspheric mirrors thanks to the generalization of the Nijboer-Zernike retrieval theory we have developed here.

Conclusions

Objectives and results

The International Liquid Mirror Telescope (ILMT) constitutes a unique instrument project that will perform a deep survey of a 22 arcmin wide strip of sky in order to discover and to study many unknown variable objects such as gravitational lenses and supernovae. After a general presentation of the project and an introduction to the basic concepts it involves, this thesis mainly addressed the optical issues related to the ILMT.

We have first identified the factors that may affect the mirror surface. We have then developed mathematical expressions for the corresponding aberrations and used them to quantify the ILMT aberrations. We have also brought some corrections to the expression of the perturbations related to the rotation of the Earth that had been previously calculated. Apart from easily compensable defocus or tilt effects, most of those perturbations do not create aberrations sufficiently large to prevent the mirror to be usable.

However, the deviation from a perfect parabola caused by wavelets propagating through the mercury can be significant. We have proposed numerical models for the two types of waves that have been observed on liquid mirrors. They helped us to understand the effect of these waves on the image quality. Their main effect is to diffract the light outside the central peak of the PSF. The amount of light that is scattered in the halo depends on the amplitude of the waves and on the fraction of the radius that is disturbed (in the case of spiral wave), and the distance between the halo and the peak depends on the number of waves disturbing the mirror. Using a thin layer of mercury allows to damp these waves, and, thus, to decrease the amount of diffracted light.

After this theoretical approach of the problem, we have searched for a practical way of measuring these perturbations. We reviewed several aberration measurement techniques and particularly the Nijboer-Zernike theory. The latter can be used either to compute the PSF of an optical system from its known aberrations or inversely to determine the aberrations of an optical system from the PSFs it produces. This method had previously been developed for low order aberrations and symmetrical systems. We have extended it to asymmetrical systems involving a large number of aberrations (possibly large) and we wrote a software that implements this evolution.

Numerical simulations allowed us to characterize the capabilities of this method. We have shown that the iterative process converges faster when the strehl ratio is higher and we have imagined a way of accelerating the convergence when the strehl value is near to the divergence limit. This upgrade allows the process to converge with aberrated images presenting a strehl ratio as low as 35%.

We have also studied the effect of the completeness of the retrieval. We saw that the retrieval

process gives better results (more aberrations accurately measured) when more aberrations are searched, even if they are not present in the images. In this case, it is thus better to use as many aberrations as possible.

The study of the effects of the noise in the image leads to a very different conclusion. Indeed, the noise masks the outer rings of the PSF and the aberrations related to those rings cannot be retrieved accurately. When the level of noise increases, less aberrations can thus be measured. Trying to measure them anyway will result in a decrease of the accuracy on the low order aberrations because of error propagation. There is thus a trade off between the completeness and the noise perturbations, and we have shown a convenient way to determine the optimal number of aberrations.

The image sampling has also an effect on the signal to noise ratio as the noise level is mainly fixed for a given pixel but the amount of signal received depends on the number of pixel on which the total energy is spread out. Using an adequate sampling thus improves the signal to noise ratio, and the measurement accuracy.

Finally, we have shown that the accuracy of the measurement is quite high. The average error value for 231 measured aberrations is lower than $\lambda/1000$ even with poor image quality (only two PSF rings are visible). In case of a very high signal to noise ratio it can even falls below $\lambda/10000$ and noise free images lead to errors smaller than that, which demonstrates the theoretical potential of this method. Even with a lot of noise, the accuracy is always better than $\lambda/100$.

This accurate aberration retrieval method would be very convenient to use with liquid mirror telescopes because of its simplicity of implementation and its capability to work with many aberrations. Indeed, it only requires a source and a detector, contrarily to other methods that generally involve the introduction of specific devices in the light beam, which can reveal difficult to implement 16m above a mercury pool. Moreover, this technique does not present "aberration-saturation" issues that would prevent to measure the other aberrations. It is thus perfectly well suited to measure large complex optical systems.

Several applications of this method have been considered in the present dissertation.

First of all, we have presented an improved numerical alignment method, based on aberration maps. Monté-Carlo analyzes have demonstrated the accuracy (position computed within 0.5mm and 1 arcmin accuracy) and reliability (95 % of the cases are treated correctly) of this method, provided that at least two orthogonal aberrations (ex: tip and tilt) can be measured accurately, which should be possible with the NZ method. Moreover, this approach can easily be implemented as it only requires a diverging source (i.e. a fiber). We have thus developed an easy, practically usable, reliable and accurate method to measure the misalignment of the ILMT upper-end unit with respect to its primary mirror.

We have then applied the Nijboer-Zernike method to calibrate the NACO instrument. This type of adaptive optics instrument requires a good calibration in order to give the best possible performances. Even if this type of calibration has already been performed by other methods, the simplicity to use the Nijboer-Zernike technique provides an "on-the-fly" calibration tool. We have thus used this technique to measure the residual static aberrations of NACO. In order to achieve this goal, we have successfullu introduced an easy way to account for an obscured pupil in the retrieval process. Using a NACO coronagraph simulator, we saw that the measured aberrations correspond to about 25% of the speckles of the coronagraphic images obtained with the instrument. This means that 1/4 of the speckles, generated by the static aberrations, could

be removed with a suitable calibration of the deformable mirror to improve the images.

After that we used the Nijboer-Zernike theory in a more classical way, to determine the aberrations of a simple lens. This is a first step for the validation of the lenses of the ILMT corrector. We have determined the parameters of those tests, and showed that it would be preferable to test de corrector in its whole, because of the very long focal length of the separate lenses.

Finally, we have shown how to modify the Nijboer-Zernike approach in order to increase its dynamic range of measurement. Indeed, this phase retrieval method has the exceptional property of not making use of reference surfaces, which allows to measure aberrations in a wide range without risking that the strong aberrations mask the faint ones. This property is particularly convenient to test aspherical surfaces in a simple way, without any auxiliary optical element (i.e. null lenses).

Perspectives

The development of the Nijboer-Zernike retrieval method should not stop here. We have only began to explore its possibilities and the applications that could take advantage of this method are numerous. Theoretically as well as practically, there are several interesting things that should be developed.

First of all, it would be very interesting to include the annular Zernike polynomial theory (Mahajan 1981) in the phase retrieval process. Indeed, since the pupil of most telescopes has a central obscuration, the use of annular polynomials seems well suited. Even if the way we took the VLT pupil into account for the NACO calibration was interesting, using annular polynomials would probably increase the convergence speed of the retrieval process.

Another theoretical point that should be investigated is the determination of the precise analytical relation that exists between the classical Zernike coefficients (α) and general ones (β). In this thesis, we have used several conversion approaches that were good, but there remain still some problems. A mathematical relation would thus be a plus for the Nijboer-Zernike theory.

The applications we have simulated at the end of the present dissertation, should be practically implemented. Particularly the alignment of the upper-end unit could be realized using our method. We also plan to use the high dynamic theory in order to characterize two 15cm parabolic mirrors that will be used on an optical bench for interferometry. The same method could be applied to measure the liquid mirror of the ILMT from its center of curvature, once it will be ready. The classical way of measuring lenses could also be used to validate the ILMT corrector. This has been prepared in this thesis, and could easily be implemented.

Finally, among the applications that have not been approached here, we imagined one (but it is not the only one) that is probably worth some investigations. It consists in post-treating the images acquired with a telescope in order to try to improve their quality. Using stars in the field of view of the instrument as references, it should be possible to measure the aberrations present in this field, like it is the case with adaptive optics systems. It should then be possible to take them into account in order to improve the quality of the images. This will probably not be an easy task but it could be a convenient way to improve images acquired with telescopes that are not equipped with adaptive optics.

Bibliography

Abramovitch, M. and Stegun, I. A. 1972. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Dover Publications, New York.

Blanc, A., Fusco, T., Hartung, M., Mugnier, L. M., and Rousset, G. 2003. Calibration of NAOS and CONICA static aberrations. Application of the phase diversity technique. *A&A*, 399:373–383.

Borgeaud, P., Gallais, P., Boulade, O., Carton, P.-H., Gros, M., de Kat, J., Lee, P., and Nemee, L. 2000. 40 CCDs of the MegaCam wide-field camera: procurement, testing, and first laboratory results. In Iye, M. and Moorwood, A. F., editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 4008, pages 356–367.

Borra, E. 2001. Strip Searching for Supernovae. *astro-ph/0106264*.

Borra, E., Ritcey, A., and Artigau, E. 1999. Floating mirrors. *The Astrophysical Journal*, 516.

Borra, E. F. 1982. The liquid-mirror telescope as a viable astronomical tool. *Royal Astronomical Society of Canada, Journal*, 76:245–256.

Borra, E. F. 1997. Astronomical Research with liquid mirror telescopes. *ArXiv Astrophysics e-prints*.

Borra, E. F., Beauchemin, M., Arsenault, R., and Lalande, R. 1985a. Optical-shop testing of liquid mirrors. *PASP*, 97:454–464.

Borra, E. F., Beauchemin, M., and Lalande, R. 1985b. Liquid mirror telescopes - Observations with a 1 meter diameter prototype and scaling-up considerations. *ApJ*, 297:846–851.

Borra, E. F., Content, R., Girard, L., Szapiel, S., Tremblay, L. M., and Boily, E. 1992. Liquid mirrors - Optical-shop tests and contributions to the technology. *ApJ*, 393:829–847.

Borra, E. F., Content, R., Poirier, S., and Tremblay, L. 1988. Observing with liquid mirror telescopes - Evaluation of two observing seasons. *PASP*, 100:1015–1023.

Borra, E. F., Tremblay, G., Huot, Y., and Gauvin, J. 1997. Gallium Liquid Mirrors: Basic Technology, Optical-Shop Tests, and Observations. *PASP*, 109:319–325.

Braat, J., Dirksen, P., and Janssen, A. J. E. M. 2002. Assessment of an extended Nijboer-Zernike approach for the computation of optical point-spread functions. *Journal of the Optical Society of America A*, 19:858–870.

Burge, J. H., Zehnder, R., and Zhao, C. 2007. Optical alignment with computer-generated holograms. volume 6676, page 66760C. SPIE.

Cabanac, R. A. 1997. Science with Liquid Mirror Telescopes: Season 1996 of NASA Orbital Debris Observatory. *ArXiv Astrophysics e-prints*.

Cabanac, R. A., Borra, E. F., and Beauchemin, M. 1998. A Search for Peculiar Objects with the NASA Orbital Debris Observatory 3 Meter Liquid Mirror Telescope. *ApJ*, 509:309–323.

Cabanac, R. A., Hickson, P., and de Lapparent, V. 2002. The Large Zenith Telescope Survey: A Deep Survey Using a 6-m Liquid Mirror Telescope. In N. Metcalfe & T. Shanks, editor, *A New Era in Cosmology*, volume 283 of *Astronomical Society of the Pacific Conference Series*, pages 129–131.

Content, R., Borra, E. F., Drinkwater, M. J., Poirier, S., Poisson, E., Beauchemin, M., Boily, E., Gauthier, A., and Tremblay, L. M. 1989. A search for optical flares and flashes with a liquid-mirror telescope. *AJ*, 97:917–922.

Cordero-Davila, A., Cornejo-Rodriguez, A., and nez, O. C.-N. 1992. Ronchi and Hartmann tests with the same mathematical theory. *Appl. Opt.*, 31(13):2370–2376.

Dirksen, P., Braat, J., Janssen, A., and Juffermans, C. 2003. Aberration retrieval using the extended Nijboer-Zernike approach. *Journal of Microlithography, Microfabrication, and Microsystems 2*, pages 61–68.

Foucault, L. 1859. Mémoire sur la construction des telescopes en verre argenté. *Annales de l'Observatoire de Paris*, 5:197–237.

Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., and Schneider, D. P. 1996. The Sloan Digital Sky Survey Photometric System. *AJ*, 111:1748.

Gagné, G., Borra, E. F., and Ritcey, A. M. 2008. Tiltable rotating liquid mirrors: a progress report. *A&A*, 479:597–602.

Gibson, B. K. and Hickson, P. 1991. Outline of a liquid mirror telescope QSO survey. In Crampton, D., editor, *The Space Distribution of Quasars*, volume 21 of *Astronomical Society of the Pacific Conference Series*, pages 80–83.

Gibson, B. K. and Hickson, P. 1992a. Liquid mirror surface aberrations. I - Wavefront analysis. *ApJ*, 391:409–417.

Gibson, B. K. and Hickson, P. 1992b. Time-delay integration CCD read-out technique - Image deformation. *MNRAS*, 258:543–551.

Girard, L. and Borra, E. F. 1997. Optical tests of a 2.5-m-diameter liquid mirror: behavior under external perturbations and scattered-light measurements. *Appl. Opt.*, 36:6278–6288.

Hartmann, J. 1900. Bemerkungen über den Bau und die Justirung von Spektrographen. *Z Instrumentenkd*, 20:47–58.

Hartung, M., Blanc, A., Fusco, T., Lacombe, F., Mugnier, L. M., Rousset, G., and Lenzen, R. 2003. Calibration of NAOS and CONICA static aberrations. Experimental results. *A&A*, 399:385–394.

Hickson, P. 2001. Eliminating the Coriolis Effects in Liquid Mirrors. *Publications of the Astronomical Society of the Pacific*, 113:1511–1514.

Hickson, P. 2008. Fundamentals of atmospheric and adaptive optics.

Hickson, P., Borra, E. F., Cabanac, R., Chapman, S. C., de Lapparent, V., Mulrooney, M., and Walker, G. A. 1998. Large Zenith Telescope project: a 6-m mercury-mirror telescope. 3352:226–232.

Hickson, P., Borra, E. F., Cabanac, R., Content, R., Gibson, B. K., and Walker, G. A. H. 1994. UBC/Laval 2.7 meter liquid mirror telescope. *ApJ*, 436:L201–L204.

Hickson, P., Cabanac, R., and Watson, S. 1993. A study of mercury vapour concentrations at the UBC/Laval 2.7 meters liquid mirror observatory. Technical report, University of British Columbia.

Hickson, P., Gibson, B. K., and Hogg, D. W. 1993. Large astronomical liquid mirrors. *PASP*, 105:501–508.

Hickson, P. and Mulrooney, M. 1998. UBC - NASA Multi-Narrowband Survey. I - Description and Photometric Properties of the Survey. *The Astrophysical Journal Sypplement*, 115:35–42.

Hickson, P., Pfrommer, T., Cabanac, R., Crotts, A., Johnson, B., Lapparent, V. D., Lanzetta, K., Gromoll, S., Mulrooney, M., Sivanandam, S., and Truax, B. 2007. The Large Zenith Telescope: A 6 m Liquid-Mirror Telescope. *Publications of the Astronomical Society of the Pacific*, 119:444–455.

Hickson, P. and Racine, R. 2007. Image Quality of Liquid-Mirror Telescopes. *Publications of the Astronomical Society of the Pacific*, 119:456–465.

Hickson, P. and Richardson, E. 1998. A Curvature-Ccompensated Corrector for Drift-Scan Observations. *arXiv:astro-ph/9806303 v1.*

Hofmann, K.-H. and Weigelt, G. 1988. Speckle masking observation of Eta Carinae. *A&A*, 203:L21–L22.

Janssen, A. J. E. M. 2002. Extended Nijboer-Zernike approach for the computation of optical point-spread functions. *Journal of the Optical Society of America A*, 19:849–857.

Jean, C., Claeskens, J. ., and Surdej, J. 1999. Gravitational lensing studies with the 4-m International Liquid Mirror Telescope (ILMT). *ArXiv Astrophysics e-prints.*

Koechlin, L. 1990. *Taillez vous-mêmes votre miroir de télescope.* http://www.ast.obs-mip.fr/users/lkoechli/w3/miroir/Mi1.html.

Kristian, J. and Blouke, M. 1982. Charge-coupled devices in astronomy. *Scientific American*, 247:66–74.

Labeyrie, A. 1970. Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images. *A&A*, 6:85–87.

Landau, L. D. and Lifshitz, E. M. 1959. *Fluid mechanics.*

Lane, R. G. and Tallon, M. 1992. Wave-front reconstruction using a Shack-Hartmann sensor. *Appl. Opt.*, 31(32):6902–6908.

Mackowski, J., Pinard, L., Dognin, L., Ganau, P., Lagrange, B., Michel, C., and Morgue, M. 1999. Virgo mirrors: wavefront control. *Optical and Quantum Electronics*, 31:507–514(8).

Magette, A. 2007. Liquid Mirror Telescope Alignment. Master's thesis, University of Liège.

Mahajan, V. 1998. *Optical Imaging and Aberration, chapter 3*. SPIE Optical Engineering Press.

Mahajan, V. N. 1981. Zernike annular polynomials for imaging systems with annular pupils. *Journal of the Optical Society of America (1917-1983)*, 71:75–85.

Malacara, D. 1972. Hartmann test of aspherical mirrors. *Appl. Opt.*, 11(1):99–101.

Markwardt, C. B. 2008. Non-linear least squares fitting in idl with mpfit. In *Astronomical Data Analysis Software and Systems XVIII, Quebec, Canada, ASP Conference Series, Vol. TBD, eds. D. Bohlender, P. Dowler & D. Durand (Astronomical Society of the Pacific: San Francisco)*.

Merkle, F. 1993. Principles of adaptive optics. In Zuegge, H., editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 1780 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 142–151.

Mulrooney, M. 2000. *A 3.0 meter liquid mirror telescope*. PhD thesis, RICE UNIVERSITY.

Nijboer, B. 1942. *The Diffraction Theory of Aberrations*. PhD thesis.

Ninane, N. 2001. Télescope à miroir liquide; rapport final. Technical report, Centre Spatial de Liège.

Ninane, N. M. and Jamar, C. A. 1996. Parabolic liquid mirrors in optical shop testing. *Appl. Opt.*, 35:6131–6139.

Noll, R. J. 1976. Zernike polynomials and atmospheric turbulence. *Journal of the Optical Society of America (1917-1983)*, 66:207–211.

Offner, A. 1963. A null corrector for paraboloidal mirrors. *Appl. Opt.*, 2(2):153–155.

Pfrommer, T., Hickson, P., She, C., and Vance, J. D. 2008. High-resolution lidar experiment for the Thirty Meter Telescope. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7015 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*.

Platt, B. and Shack, R. 2001. History and principles of Shack-Hartmann wavefront sensing. *Journal of Refractive Surgery*, 17:573–577.

Poels, J., Moreau, O., Manfroid, J., Surdej, J., Borra, E., Claeskens, J. F., Jean, C., Montfort, F., Swings, J. P., Vangeyte, B., van Dessel, E., and Nakos, T. 2001. Surveys with the 4-m International Liquid Mirror Telescope. In A. J. Banday, S. Zaroubi, & M. Bartelmann, editor, *Mining the Sky*, pages 598–600.

Potter, A. E. and Mulrooney, M. 1997. Liquid metal mirror for optical measurements of orbital debris. *Advances in Space Research*, 19:213–219.

Riaud, P. and Hanot, C. 2010. Coronagraphy with interferometry as a tool for measuring stellar diameters. *Astrophysical Journal*. to be published.

Riaud, P., Mawet, D., and Absil, O. 2005. Limitation of the pupil replication technique in the presence of instrumental defects. *The Astrophysical Journal Letters*, 628(1):L81–L84.

Roddier, C. and Roddier, F. 1991. Reconstruction of the Hubble Space Telescope mirror figure from out-of-focus stellar images. In Bely, P. Y. and Breckinridge, J. B., editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 1494, pages 78–84.

Roddier, C. and Roddier, F. 1993. Combined approach to the Hubble Space Telescope wave-front distortion analysis. *Appl. Opt.*, 32:2992–3008.

Roddier, C., Roddier, F., Stockton, A., Pickles, A., and Roddier, N. 1990. Testing of telescope optics - A new approach. In Barr, L. D., editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 1236, pages 756–766.

Roddier, F. 1981. The Effects of Atmospheric Turbulence in Optical Astronomy. *Prog. Optics, Volume 19, p. 281-376*, 19:281–376.

Roddier, F. 1988. Curvature sensing and compensation: a new concept in adaptive optics. *Appl. Opt.*, 27:1223–1225.

Roddier, F. 2004. *Adaptive Optics in Astronomy*. Edited by François Roddier, pp. 419. ISBN 0521612144. Cambridge, UK: Cambridge University Press.

Roddier, N. and Roddier, F. 1989. Curvature sensing and compensation - A computer simulation. In Roddier, F. J., editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 1114, pages 92–96.

Ronchi, V. 1927. Due nuovi metodi per lo studio delle superficie e dei sistemi ottici. *Ann. Scuola Normale Superiore di Pisa*, 15.

Ronchi, V. 1964. Forty years of history of a grating interferometer. *Appl. Opt.*, 3(4):437–451.

Sagar, R., Stalin, C. S., Pandey, A. K., Uddin, W., Mohan, V., Sanwal, B. B., Gupta, S. K., Yadav, R. K. S., Durgapal, A. K., Joshi, S., Kumar, B., Gupta, A. C., Joshi, Y. C., Srivastava, J. B., Chaubey, U. S., Singh, M., Pant, P., and Gupta, K. G. 2000. Evaluation of Devasthal site for optical astronomical observations. *A&AS*, 144:349–362.

Salas-Peimbert, D. P., Malacara-Doblado, D., Durán-Ramírez, V. M., Trujillo-Schiaffino, G., and Malacara-Hernández, D. 2005. Wave-front retrieval from Hartmann test data. *Appl. Opt.*, 44(20):4228–4238.

Sica, R. J., Sargoytchev, S., Flatt, S., Borra, E., and Girard, L. 1992. Lidar measurements using large liquid mirror telescopes. In *NASA. Langley Research Center, 16th International Laser Radar Conference, Part 2 p 655-658 (SEE N92-31013 21-35)*, pages 655–658.

Surdej, J., Absil, O., Bartczak, P., Borra, E., Chisogne, J.-P., Claeskens, J.-F., Collin, B., De Becker, M., Defrère, D., Denis, S., Flebus, C., Garcet, O., Gloesener, P., Jean, C., Lampens, P., Libbrecht, C., Magette, A., Manfroid, J., Mawet, D., Nakos, T., Ninane, N., Poels, J., Pospieszalska, A., Riaud, P., Sprimont, P.-G., and Swings, J.-P. 2006. The 4m international liquid mirror telescope (ILMT). In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6267.

Surdej, J. and Claeskens, J.-F. 1997. Gravitational lens studies with a LMT. *Proceedings of the Workshop held at Marseille Observatory.*

Terrien, J. and Maréchal, A. 1964. *L'optique théorique*, volume 615 of *Que sais-je ?* PUF.

Texereau, J. 1961. *La Construction Du Télescope D'amateur*.

Tremblay, G. and Borra, E. F. 2000. Optical tests of a 3.7-m-diameter liquid mirror: behavior under external perturbations. *Appl. Opt.*, 39:5651–5663.

Trujillo, I., Aguerri, J. A. L., Cepa, J., and Gutiérrez, C. M. 2001a. The effects of seeing on Sérsic profiles - II. The Moffat PSF. *MNRAS*, 328:977–985.

Trujillo, I., Aguerri, J. A. L., Cepa, J., and Gutiérrez, C. M. 2001b. The effects of seeing on Sèrsic profiles I. *MNRAS*, 321:269–276.

van der Avoort, C., Braat, J. J. M., Dirksen, P., and Janssen, A. J. E. M. 2005. Aberration retrieval from the intensity point-spread function in the focal region using the extended Nijboer Zernike approach. *Journal of Modern Optics*, 52:1695–1728.

Vangeyte, B., J.Manfroid, and Surdej, J. 2002. Study of CCD mosaic configurations for the ILMT: Astrometry and photometry of point sources in the absence of a TDI corrector. *Astronomy and Astrophysics*, 388:712 – 731.

Weigelt, G., Balega, Y. Y., Blöcker, T., Hofmann, K.-H., Men'shchikov, A. B., and Winters, J. M. 2002. Bispectrum speckle interferometry of IRC +10216: The dynamic evolution of the innermost circumstellar environment from 1995 to 2001. *A&A*, 392,:131–141.

Weigelt, G. and Ebersberger, J. 1986. Eta Carinae resolved by speckle interferometry. *A&A*, 163:L5–L6.

Wittkowski, M., Balega, Y., Beckert, T., Duschl, W. J., Hofmann, K.-H., and Weigelt, G. 1998. Diffraction-limited 76mas speckle masking observations of the core of NGC 1068 with the SAO 6m telescope. *A&A*, 329:L45–L48.

Wyant, J. and Creath, K. 1992. Basic Wavefront Aberration Theory for Optical Metrology. *Applied Optics and Optical Engineering*, XI.

Zernike, F. 1934. Beugungstheorie des Schneidenverfahrens und seiner Verbesserten Form, der Phasenkontrastmethode. *Physica 1*, pages 689–704.

Zhao, C., Zehnder, R., Burge, J. H., and Martin, H. M. 2005. Testing an off-axis parabola with a CGH and a spherical mirror as null lens. In Stahl, H. P., editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5869, pages 261–272.

Part IV

Appendices

Appendix A

Focus tolerancing

The question has appeared to know whether it would be better to compensate for thermal induced defocus by moving the corrector and the detector together or by moving only the detector. Indeed, even if it can seem better to move the detector with the corrector, the latter is quite large ($\sim 55cm$ in diameter and $\sim 85cm$ long) and heavy. Hence, it would be simpler to move the camera alone if the image quality allows it.

In order to estimate the impact of a correction of a global defocusing, due to thermal expansion of the structure, by moving the camera alone, we used the Zemax model of the ILMT (shown in fig. A.1) to compute the correction required on the camera focal position for several defocuses introduced before the corrector (displacement of the whole upper-end unit: camera+corrector).

We thus moved the corrector with respect to the primary mirror (defocus) in order to simulate a change in size of the structure, and we then optimized the axial position of the camera to improve the focus (refocus correction). The displacement of the camera required to compensate for the global defocus is presented in fig. A.2.

Let us note that several fields have been considered along the North-South direction in order to study the image quality along this axis. This allows to analyze the quality of the correction of the TDI distortion related to a variation of the transit speed of the stars along the N-S axis. The re-optimization of the focal position of the camera is such that all fields are optimized together.

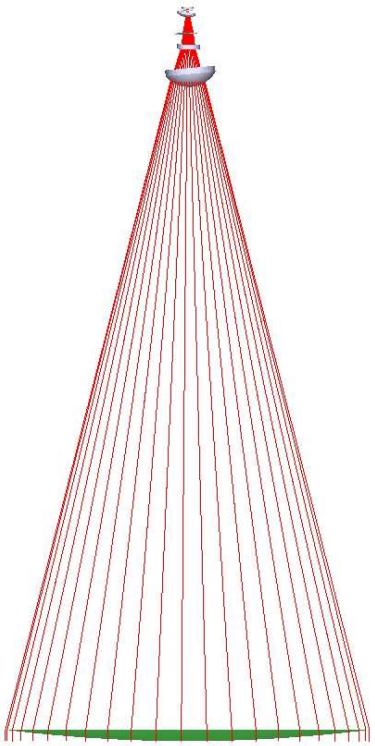


Figure A.1: Zemax model of the ILMT optics.

The relation between the correction (displacement of the camera) and the defocus (displacement of the whole upper-end unit) is quasi-linear with a very small quadratic term. The correction required is larger (~ 1.35 times) than the defocus applied, which turns out to be annoying since the spacing between the corrector and the camera is very small.

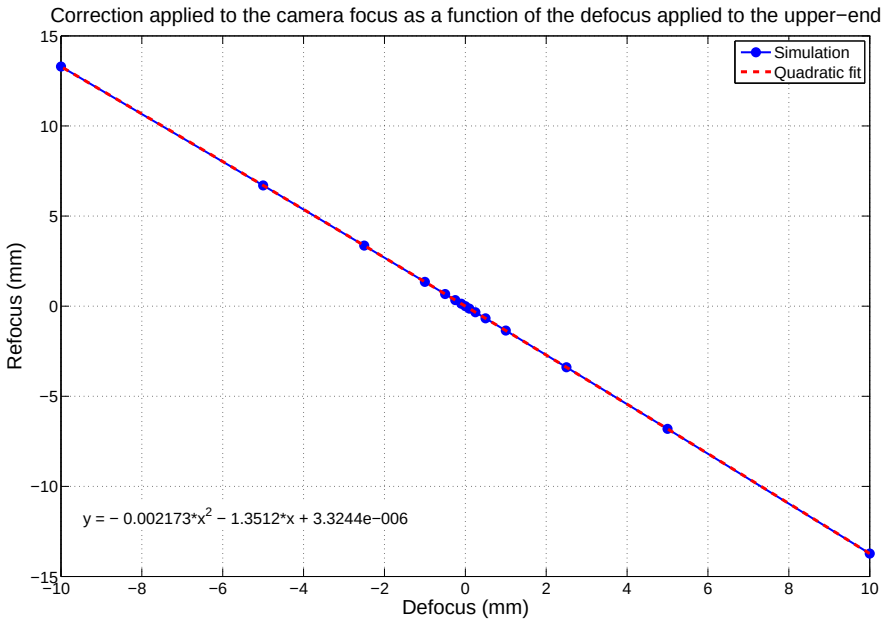


Figure A.2: Focal displacement of the camera required to compensate for the global upper-end defocusing. The correction needed on the camera position is larger than the focal perturbation applied before the corrector by a factor around 1.35.

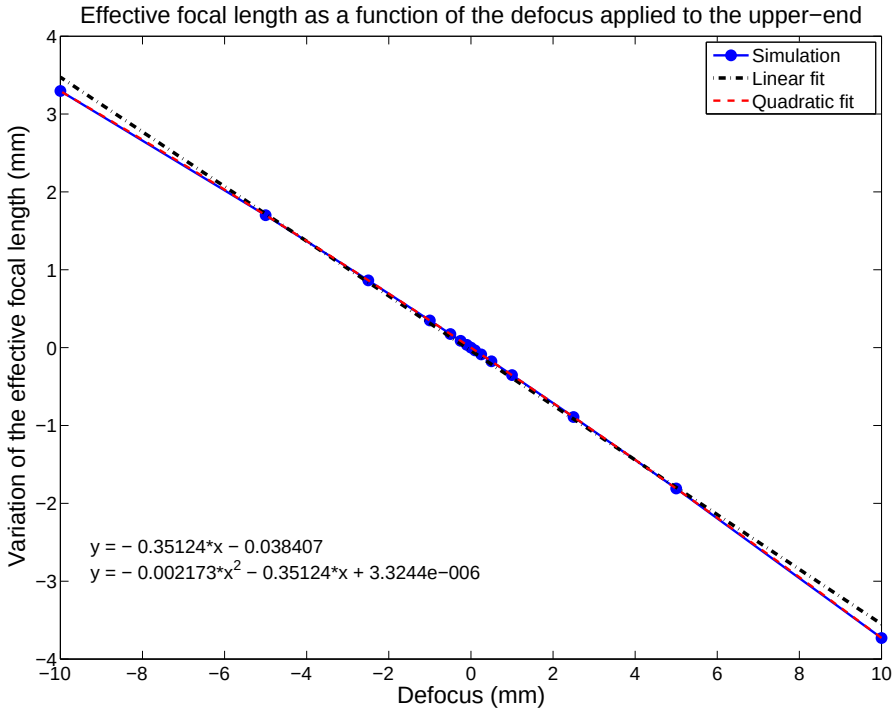


Figure A.3: Variation of the effective focal length (distance between the primary mirror and the detector) due to the difference between the defocus applied before the corrector and the focus correction applied to the camera. As the correction is larger than the perturbation (fig. A.2), the effective focal length changes.

Since the correction required on the camera focal position is larger than the defocus applied to the corrector, the effective focal length of the telescope (the distance between the primary mirror and the focal plane of the telescope) also varies with the defocus. This variation is presented in fig. A.3.

We have already seen that it will be difficult to compensate for a global defocusing by simply moving the camera, since the required correction is larger than the perturbation and the available space for the correction is quite limited. Let us now estimate the impact on the image quality of this manipulation (defocus + refocus). With that objective in mind, we performed a spot diagram analysis for each case of defocus/refocus. The quality of the image is represented by the spot radius. Fig. A.4 presents the evolution of the mean of the five geometrical spot radii corresponding to the five fields (along the N-S axis) as a function of the defocus applied to the corrector.

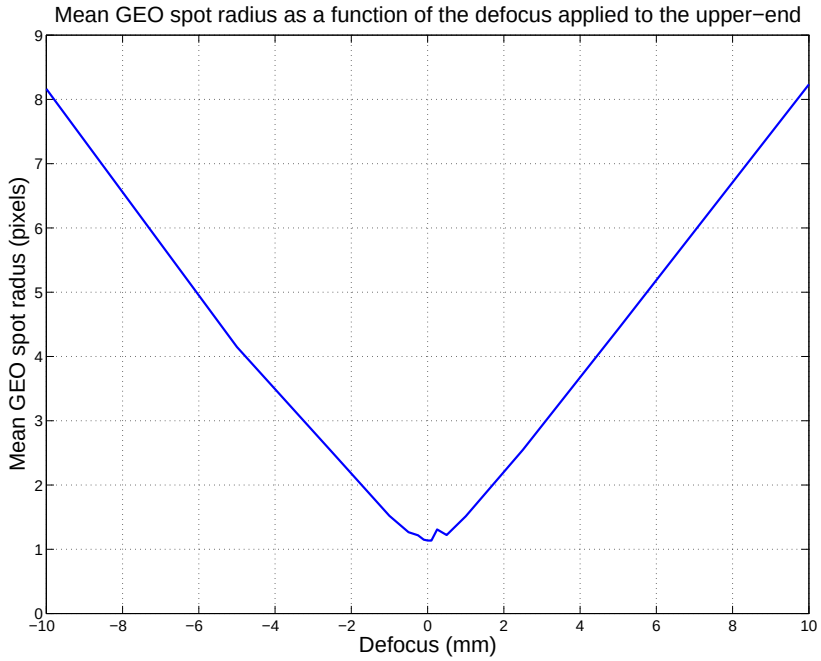


Figure A.4: Geometrical (GEO) spot radius (given by a spot diagram study) expressed in pixels ($15\mu m$) as a function of the defocus applied before the corrector. For a small defocus ($\pm$1mm) the size of the spot does not vary much. However, as soon as the defocus increases, the spot radius significantly increases, denoting a degradation of the image quality.

For a small defocus ($\pm 1mm$), the correction seems to be sufficient and the spot radius does not vary too much. However, as soon as the defocus gets larger, the image quality quickly degrades to a level that is not acceptable.

Moreover, we have noticed that the TDI drift had to be adapted in order that the correction applied to the camera position correctly balances the defocus introduced on the corrector. This is due to the change of the effective focal length related to the motion of the camera with respect to the corrector. It corresponds to a modification of the field of view that has to be compensated at the level of the TDI drift rate.

The Zemax model of the ILMT simulates the TDI drift by moving the CCD along the East-West direction. This displacement is determined by the two extremal positions of the chip (symmetrical with respect to the axis) and intermediary positions that are fractions of the extremal ones (i.e. position one corresponds to an extremity, position two to 90% of the extremity, position 3 to 80%,...). Defining a single extremity thus determines the entire movement of the CCD. Hereafter, we will call "TDI parameter" the extremal position that is used to determine this movement.

Hence, this TDI parameter has to be optimized for each case of defocus/refocus, which was performed manually, simply by looking at the superimposition of the spots corresponding to the different positions. The values are thus not perfectly accurate. The variation of the TDI parameter is presented in fig. A.5.

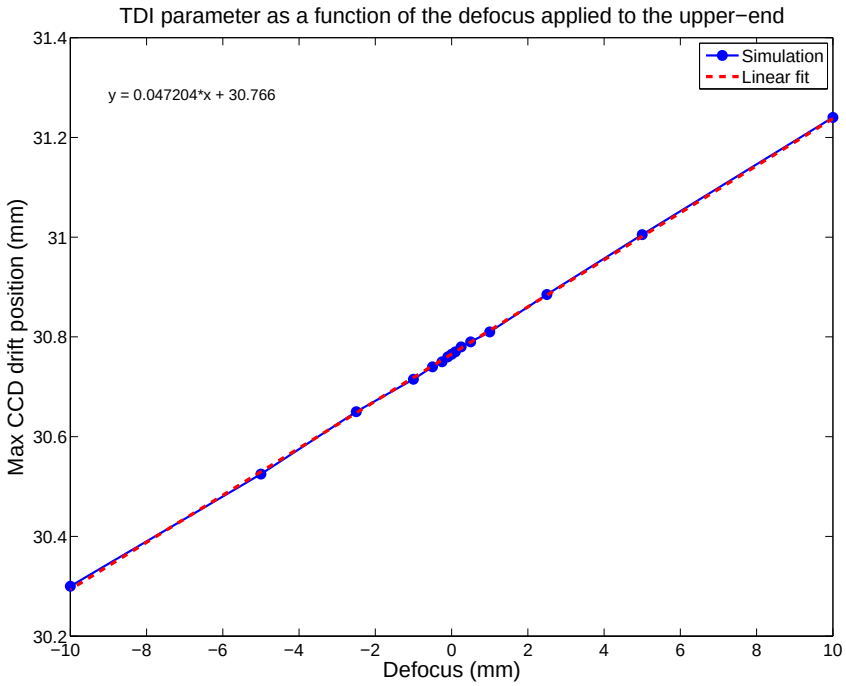


Figure A.5: Evolution of the extremal position of the camera (simulation of the TDI drift) as a function of the defocus applied before the corrector. The values presented here have been manually estimated and they are only first order approximations. However, the variation of the parameter seems to be a linear function of the defocus.

The relation between the defocus and the TDI parameter seems clearly linear and could thus be theoretically easily compensated. However, we showed in the 2k CCD camera TDI analysis (2.3.3) that modifying the TDI drift rate during the acquisition disturbs several images, and that such an adjustment during the readout is thus not desirable.

In conclusion, because of the problems that appear when the focal degree of freedom of the camera is used to compensate for a defocus of the whole upper-end assembly (corrector + detector), this solution should probably not be envisioned. Indeed, not only the movement of the camera could become large, as the correction required is larger than the introduced defocus, within a space that is small, but also the TDI drift rate has to be adapted. Even if these two corrections can be performed, the image quality becomes very bad for a defocus as large as 2mm.

Appendix B

Computation of the intensity

Calculation of the intensity

The expression of the complex amplitude $U(r, \phi, f)$ as a function of the general Zernike coefficients is given by equation 5.21, reminded here

$$U = 2\beta_0^0 V_0^0 + 2 \sum_{n,m}{}' \beta_{cn}^m i^m V_n^m \cos(m\phi) + 2 \sum_{n,m}{}' \beta_{sn}^m i^m V_n^m \sin(m\phi) \tag{B.1}$$

where the "prime" sign means that the term $n = m = 0$ has been removed from the sums. The variables (r, ϕ, f) of the functions have been omitted for the sake of simplicity. Let us define the following shortcuts for the sum over the sine and cosine terms

$$\begin{aligned} \sum \cos &= \sum_{n,m}{}' \beta_{cn}^m i^m V_n^m \cos(m\phi) \\ \sum \sin &= \sum_{n,m}{}' \beta_{sn}^m i^m V_n^m \sin(m\phi) \end{aligned} \tag{B.2}$$

The intensity of the PSF is given by the square modulus of the complex amplitude $|U|^2$. Using the shortcuts defined above, we get

$$|U|^2 = \left[\Re(2\beta_0^0 V_0^0 + 2\sum \cos + 2\sum \sin)\right]^2 + \left[\Im(2\beta_0^0 V_0^0 + 2\sum \cos + 2\sum \sin)\right]^2 \tag{B.3}$$

β_0^0 being assumed to be real and positive, this equation becomes

$$\begin{aligned} |U|^2 &= \left[2\beta_0^0 \Re(V_0^0) + \Re(2\sum \cos) + \Re(2\sum \sin)\right]^2 \\ &+ \left[2\beta_0^0 \Im(V_0^0) + \Im(2\sum \cos) + \Im(2\sum \sin)\right]^2 \end{aligned} \tag{B.4}$$

Developing the quadratic terms gives,

$$
\begin{aligned}
|U|^2 \quad &= 4(\beta_0^0)^2(\Re(V_0^0))^2 + (2\Re(\sum \cos))^2 + (2\Re(\sum \sin))^2 + 2(2\Re(\sum \cos))(2\Re(\sum \sin)) \\
&+8 \sum_{n,m}{}' \beta_0^0 \, \Re(V_0^0) \, \Re(\beta_{cn}^m i^m V_n^m) \cdot \cos(m\phi) \\
&+8 \sum_{n,m}{}' \beta_0^0 \, \Re(V_0^0) \, \Re(\beta_{sn}^m i^m V_n^m) \cdot \sin(m\phi) \\
&+4(\beta_0^0)^2(\Im(V_0^0))^2 + (2\Im(\sum \cos))^2 + (2\Im(\sum \sin))^2 + 2(2\Im(\sum \cos))(2\Im(\sum \sin)) \\
&+8 \sum_{n,m}{}' \beta_0^0 \, \Im(V_0^0) \, \Im(\beta_{cn}^m i^m V_n^m) \cdot \cos(m\phi) \\
&+8 \sum_{n,m}{}' \beta_0^0 \, \Im(V_0^0) \, \Im(\beta_{sn}^m i^m V_n^m) \cdot \sin(m\phi)
\end{aligned} \tag{B.5}
$$

Let us now gather the corresponding terms in real and imaginary parts. The intensity can be written as a sum of six terms,

$$I = |U|^2 = a + b + c + d + e + f \tag{B.6}$$

where a, b, c, d, e, f are developed hereafter.

$$a = 4(\beta_0^0)^2 \, (\Re(V_0^0))^2 + 4(\beta_0^0)^2 \, (\Im(V_0^0))^2 = 4(\beta_0^0)^2 \left|V_0^0\right|^2 \tag{B.7}$$

$$b = (2\Re(\sum \cos))^2 + (2\Im(\sum \cos))^2 = 4 \left| \sum_{n,m}{}' \beta_{cn}^m i^m V_n^m \cos(m\phi) \right|^2 \tag{B.8}$$

$$c = (2\Re(\sum \sin))^2 + (2\Im(\sum \sin))^2 = 4 \left| \sum_{n,m}{}' \beta_{cn}^m i^m V_n^m \sin(m\phi) \right|^2 \tag{B.9}$$

$$
\begin{aligned}
d \quad &= \quad 2(2\Re(\sum \cos))(2\Re(\sum \sin)) + 2(2\Im(\sum \cos))(2\Im(\sum \sin)) \\
&= \quad 8\Re(\sum \cos)\Re(\sum \sin) + 8\Im(\sum \cos)\Im(\sum \sin)
\end{aligned} \tag{B.10}
$$

$$e = 8 \sum_{n,m}{}' \beta_0^0 \, \Re(V_0^0) \, \Re(\beta_{cn}^m i^m V_n^m) \cdot \cos(m\phi) + 8 \sum_{n,m}{}' \beta_0^0 \, \Im(V_0^0) \, \Im(\beta_{cn}^m i^m V_n^m) \cdot \cos(m\phi) \tag{B.11}$$

$$f = 8 \sum_{n,m}{}' \beta_0^0 \, \Re(V_0^0) \, \Re(\beta_{sn}^m i^m V_n^m) \cdot \sin(m\phi) + 8 \sum_{n,m}{}' \beta_0^0 \, \Im(V_0^0) \, \Im(\beta_{sn}^m i^m V_n^m) \cdot \sin(m\phi) \tag{B.12}$$

The b, c, d terms constitute the quadratic term (equation 5.23) that is neglected in the retrieval process. Let us now develop the term called e. Combining the sums, we get

$$e = 8\beta_0^0 \sum_{n,m}{}' \left[\Re(V_0^0)\,\Re(\beta_{cn}^m i^m V_n^m) + \Im(V_0^0)\,\Im(\beta_{cn}^m i^m V_n^m)\right] \cdot \cos(m\phi) \tag{B.13}$$

We introduce the following shortcut notations,

$$\Re(V_0^0) = V_r \qquad \Im(V_0^0) = V_i \qquad \beta_{cn}^m = \beta_{rc} + i\beta_{ic} \tag{B.14}$$

Replacing these expression in equation B.13, we get

$$e = 8\beta_0^0 \sum_{n,m}{}' \left[V_r\,\Re((\beta_{rc} + i\beta_{ic})i^m V_n^m) + Vi\,\Im((\beta_{rc} + i\beta_{ic})i^m V_n^m)\right] \cdot \cos(m\phi) \tag{B.15}$$

Moreover, remember that, if A is a complex number, we have

$$\Re(i \cdot A) = -\Im(A) \qquad \Im(i \cdot A) = \Re(A) \tag{B.16}$$

Hence, equation B.15 becomes,

$$\begin{aligned} e = 8\beta_0^0 \sum_{n,m}{}' \cos(m\phi)\cdot \quad & \big[V_r\,\Re(\beta_{rc} i^m V_n^m) + V_r\,\Re(\beta_{ic} i i^m V_n^m) \\ & + V_i\,\Im(\beta_{rc} i^m V_n^m) + V_i\,\Im(\beta_{ic} i i^m V_n^m)\big] \end{aligned} \tag{B.17}$$

Extracting β_{rc}, β_{ic} and inserting V_r, V_i in the real and imaginary operators, and rearranging the terms, we get

$$\begin{aligned} e = 8\beta_0^0 \sum_{n,m}{}' \cos(m\phi)\cdot \quad & \big[\beta_{rc}\,(\Re(i^m V_n^m V_r) + \Im(i^m V_n^m V_i)) \\ & + \beta_{ic}\,(\Re(i i^m V_n^m V_r) + \Im(i^m V_n^m i V_i))\big] \end{aligned} \tag{B.18}$$

Then, using the result B.16, this equation can be written

$$\begin{aligned} e = 8\beta_0^0 \sum_{n,m}{}' \cos(m\phi)\cdot \quad & \big[\beta_{rc}\,(\Re(i^m V_n^m V_r) - \Re(i^m V_n^m i V_i)) \\ & - \beta_{ic}\,(\Im(i^m V_n^m V_r) - \Im(i^m V_n^m i V_i))\big] \end{aligned} \tag{B.19}$$

and gathering the real and imaginary terms, we get

$$e = 8\beta_0^0 \sum_{n,m}{}' \cos(m\phi) \cdot [\beta_{rc}\,\Re(i^m V_n^m (V_r - iV_i)) - \beta_{ic}\,\Im(i^m V_n^m (V_r - iV_i))] \tag{B.20}$$

From equation B.14, we have

$$V_0^0 = V_r + iVi \qquad ; \qquad V_0^{0*} = V_r - iV_i \tag{B.21}$$

Equation B.20 becomes,

$$\begin{aligned} e &= 8\beta_0^0 \sum_{n,m}{}' \cos(m\phi) \cdot \left[\beta_{rc}\, \Re(i^m V_n^m V_0^{0*}) - \beta_{ic}\, \Im(i^m V_n^m V_0^{0*})\right] \\ &= 8 \sum_{n,m}{}' \beta_0^0\, \Re(\beta_{cn}^m)\, \Re(i^m V_n^m V_0^{0*}) \cdot \cos(m\phi) \\ &- 8 \sum_{n,m}{}' \beta_0^0\, \Im(\beta_{cn}^m)\, \Im(i^m V_n^m V_0^{0*}) \cdot \cos(m\phi) \end{aligned} \tag{B.22}$$

The term called "f" can be treated in the same way replacing "$\cos(m\phi)$" with "$\sin(m\phi)$" and "β_{cn}^m" with "β_{sn}^m", we get

$$f = 8 \sum_{n,m}{}' \beta_0^0\, \Re(\beta_{sn}^m)\, \Re(i^m V_n^m V_0^{0*}) \cdot \sin(m\phi) - 8 \sum_{n,m}{}' \beta_0^0\, \Im(\beta_{sn}^m)\, \Im(i^m V_n^m V_0^{0*}) \cdot \sin(m\phi) \tag{B.23}$$

We have thus derived the expression of the intensity.

Derivation of Ψ^m_{meas} (equations 5.27 and 5.28)

Ψ^m_{meas} is defined by equation 5.24, replacing the intensity by its expression 5.22 where the quadratic term has been neglected,

$$\Psi^m_{\text{meas}} \approx \frac{1}{2\pi} \int_0^{2\pi} 4(\beta_0^0)^2 |V_0^0|^2 \cdot e^{im\phi} d\phi \tag{B.24a}$$

$$+ \frac{1}{2\pi} \int_0^{2\pi} 8 \sum_{n,m}{}' \beta_0^0\, \Re(\beta_{cn}^m)\, \Re(i^m V_n^m V_0^{0*}) \cdot \cos(m\phi) \cdot e^{im\phi} d\phi \tag{B.24b}$$

$$- \frac{1}{2\pi} \int_0^{2\pi} 8 \sum_{n,m}{}' \beta_0^0\, \Im(\beta_{cn}^m)\, \Im(i^m V_n^m V_0^{0*}) \cdot \cos(m\phi) \cdot e^{im\phi} d\phi \tag{B.24c}$$

$$+ \frac{1}{2\pi} \int_0^{2\pi} 8 \sum_{n,m}{}' \beta_0^0\, \Re(\beta_{cn}^m)\, \Re(i^m V_n^m V_0^{0*}) \cdot \sin(m\phi) \cdot e^{im\phi} d\phi \tag{B.24d}$$

$$- \frac{1}{2\pi} \int_0^{2\pi} 8 \sum_{n,m}{}' \beta_0^0\, \Im(\beta_{cn}^m)\, \Im(i^m V_n^m V_0^{0*}) \cdot \sin(m\phi) \cdot e^{im\phi} d\phi \tag{B.24e}$$

Let us consider each term separately, equation (B.24a) becomes

$$B.24a \Rightarrow \frac{4(\beta_0^0)^2 |V_0^0|^2}{2\pi} \left[\int_0^{2\pi} \cos(m\phi) d\phi + i \cdot \int_0^{2\pi} \sin(m\phi) d\phi \right] \tag{B.25}$$

In case $m = 0$, and noticing that,

$$\chi_0^0 = 8|V_0^0|^2 \tag{B.26}$$

we get,

$$\frac{4(\beta_0^0)^2|V_0^0|^2}{2\pi}\left[\int_0^{2\pi} 1d\phi + i\cdot\int_0^{2\pi} 0d\phi\right] = 4(\beta_0^0)^2|V_0^0|^2 = \frac{1}{2}(\beta_0^0)^2\chi_0^0 \tag{B.27}$$

In case $m \neq 0$, the results is

$$\frac{4(\beta_0^0)^2|V_0^0|^2}{2\pi}\left[\int_0^{2\pi} \cos(m\phi)d\phi + i\cdot\int_0^{2\pi} \sin(m\phi)d\phi\right] = 0 \tag{B.28}$$

Equation (B.24b) gives,

$$\begin{aligned} B.24b \;\Rightarrow\; & \frac{8}{2\pi}{\sum_{n,m}}' \beta_0^0\,\Re(\beta_{cn}^m)\,\Re(i^m V_n^m V_0^{0*}) \\ & \left[\int_0^{2\pi}\cos^2(m\phi)d\phi + i\cdot\int_0^{2\pi}\cos(m\phi)\,\sin(m\phi)d\phi\right] \end{aligned} \tag{B.29}$$

In case $m = 0$, the result is

$$\frac{1}{2\pi}\sum_n \beta_0^0\,\Re(\beta_{cn}^0)\,\chi_n^0\,\left[\int_0^{2\pi} d\phi + i\cdot\int_0^{2\pi} 0d\phi\right] = \sum_n \beta_0^0\,\Re(\beta_{cn}^0)\,\chi_n^0 \tag{B.30}$$

If $m \neq 0$, we get

$$\begin{aligned} & \frac{1}{\pi}\sum_n \beta_0^0\,\Re(\beta_{cn}^m)\,\chi_n^m\,\left[\int_0^{2\pi}\cos^2(m\phi)d\phi + i\cdot\int_0^{2\pi}\frac{\sin(2m\phi)}{2}d\phi\right] \\ & = \frac{1}{\pi}\sum_n \beta_0^0\,\Re(\beta_{cn}^m)\,\chi_n^m\,\left[\int_0^{2\pi}\frac{1}{2}d\phi + \int_0^{2\pi}\frac{\cos(2m\phi)}{2}d\phi + i\cdot\int_0^{2\pi}\frac{\sin(2m\phi)}{2}d\phi\right] \\ & = \sum_n \beta_0^0\,\Re(\beta_{cn}^m)\,\chi_n^m \end{aligned} \tag{B.31}$$

Equation (B.24c) becomes,

$$\begin{aligned} B.24c \;\Rightarrow\; & -\frac{8}{2\pi}\sum_n \beta_0^0\,\Im(\beta_{cn}^m)\,\Im(i^m V_n^m V_0^{0*}) \\ & \left[\int_0^{2\pi}\cos^2(m\phi)d\phi + i\cdot\int_0^{2\pi}\cos(m\phi)\,\sin(m\phi)d\phi\right] \end{aligned} \tag{B.32}$$

In case $m = 0$, the result is

$$\frac{1}{2\pi}\sum_n \beta_0^0\,\Im(\beta_{cn}^0)\,\Psi_n^0\,\left[\int_0^{2\pi} d\phi + i\cdot\int_0^{2\pi} 0d\phi\right] = \sum_n \beta_0^0\,\Im(\beta_{cn}^0)\,\Psi_n^0 \tag{B.33}$$

If $m \neq 0$, we get

$$\begin{aligned}
&\frac{1}{\pi}\sum_n \beta_0^0\, \Im(\beta_{cn}^m)\, \Psi_n^m \left[\int_0^{2\pi} \cos^2(m\phi)d\phi + i\cdot\int_0^{2\pi} \frac{\sin(2m\phi)}{2}d\phi\right] \\
&= \frac{1}{\pi}\sum_n \beta_0^0\, \Im(\beta_{cn}^m)\, \Psi_n^m \left[\int_0^{2\pi} \frac{1}{2}d\phi + \int_0^{2\pi} \frac{\cos(2m\phi)}{2}d\phi + i\cdot\int_0^{2\pi} \frac{\sin(2m\phi)}{2}d\phi\right] \\
&= \sum_n \beta_0^0\, \Im(\beta_{cn}^m)\, \Psi_n^m
\end{aligned} \tag{B.34}$$

Equation (B.24d) gives,

$$\begin{aligned}
B.24d \;\Rightarrow\; &\frac{8}{2\pi}\sum_{n,m}{}' \beta_0^0\, \Re(\beta_{sn}^m)\, \Re(i^m V_n^m V_0^{0*}) \\
&\left[\int_0^{2\pi} \cos 2(m\phi)\, \sin(m\phi)d\phi + i\cdot\int_0^{2\pi} \sin^2(m\phi)d\phi\right]
\end{aligned} \tag{B.35}$$

In case $m = 0$, the result is

$$\frac{1}{2\pi}\sum_n \beta_0^0\, \Re(\beta_{sn}^0\, \chi_n^0 \left[\int_0^{2\pi} 0d\phi + i\cdot\int_0^{2\pi} 0d\phi\right] = 0 \tag{B.36}$$

If $m \neq 0$, we get

$$\begin{aligned}
&\frac{1}{\pi}\sum_n \beta_0^0\, \Re(\beta_{sn}^m\, \chi_n^m \left[\int_0^{2\pi} \cos(m\phi)\, \sin(m\phi)d\phi + i\cdot\int_0^{2\pi} \sin^2(m\phi)d\phi\right] \\
&= i\sum_n \beta_0^0\, \Re(\beta_{sn}^m)\, \chi_n^m
\end{aligned} \tag{B.37}$$

As in the case of equation (B.24d), equation (B.24e) vanishes when $m = 0$, otherwise it becomes

$$\begin{aligned}
&\frac{1}{\pi}\sum_n \beta_0^0\, \Im(\beta_{sn}^m)\, \Psi_n^m \left[\int_0^{2\pi} \cos(m\phi)\, \sin(m\phi)d\phi + i\cdot\int_0^{2\pi} \sin^2(m\phi)d\phi\right] \\
&= i\sum_n \beta_0^0\, \Im(\beta_{sn}^m)\Psi_n^m
\end{aligned} \tag{B.38}$$

Gathering all these results, we get

$$\Psi_{\text{meas}}^0 = \frac{1}{2}(\beta_0^0)^2\chi_0^0 + \sum_n \beta_0^0\, \Re(\beta_{cn}^0\, \chi_0^0 + \sum_n \beta_0^0\, \Im(\beta_{cn}^0\, \Psi_0^0 \tag{B.39}$$

and

$$\begin{aligned}\Psi^m_{\text{meas}} \quad &= \sum_n \beta^0_0 \, \Re(\beta^m_{cn}) \, \chi^m_n + \sum_n \beta^0_0 \, \Im(\beta^m_{cn}) \, \Psi^m_n \\ &+ i \sum_n \beta^0_0 \, \Re(\beta^m_{sn}) \, \chi^m_n + i \sum_n \beta^0_0 \, \Im(\beta^m_{sn}) \, \Psi^m_n \end{aligned} \tag{B.40}$$

These equations correspond to equations 5.27 and 5.28, which had to be demonstrated.

Appendix C

Numerical expression of the V_n^m function

Preliminary computation: the T_n^m function

In this section we demonstrate that the T_n^m function, defined as

$$T_n^m = \int_0^1 \rho^{n+1} e^{if\rho^2} J_m(2\pi\rho r) d\rho \tag{C.1}$$

can be expressed as

$$T_n^m = e^{if} \sum_{l=1}^{\infty} (-2if)^{l-1} \sum_{j=0}^{p} t_{lj} \frac{J_{m+l+2j}(v)}{v^l} \tag{C.2}$$

with the new variables v, p and q defined as

$$v = 2\pi r \qquad p = \frac{n-m}{2} \qquad q = \frac{n+m}{2} \tag{C.3}$$

and t_{lj} is given by

$$t_{lj} = (-1)^j \frac{m+l+2j}{q+1} \binom{p}{j} \times \frac{\binom{m+j+l-1}{l-1}}{\binom{q+l+j}{q+1}} \tag{C.4}$$

where $\binom{p}{k}$ is the binomial expression, commonly defined by

$$\binom{p}{k} = \frac{p!}{k!(p-k)!} \tag{C.5}$$

Using the following variable change in equation C.1,

$$t = 2\pi\rho r \tag{C.6}$$

the expression of the function T_n^m can be written

$$T_n^m = \left(\frac{1}{v}\right)^{n+2} \int_0^v t^{n+1} e^{i\beta t^2} J_m(t) dt \tag{C.7}$$

with

$$\beta = \frac{f}{(2\pi r)^2} \tag{C.8}$$

The integral in equation C.7 is taken as the definition for a new function $H_{nm}(v)$.

We now define a general set of functions F_{nm}^l with

$$\begin{aligned} F_{nm}^{(0)}(2\pi r) &= (2\pi r)^n J_m(2\pi r) \\ F_{nm}^{(l)}(2\pi r) &= \int_0^{2\pi r} t F_{nm}^{(l-1)}(t)dt \end{aligned} \tag{C.9}$$

where n, m are real numbers such that $n + m + 1 > 0$. From these definitions, $H_{nm}(v)$ can be computed using partial integration, and we get

$$H_{nm}(v) = e^{i\beta v^2} \sum_{l=1}^{\infty} (-2i\beta)^{l-1} F_{nm}^{(l)}(v) \tag{C.10}$$

We will now compute the expression for $F_{nm}^{(l)}(v)$, and show that the equation

$$F_{nm}^{(l)}(v) = v^{n+1} \sum_{j=0}^{\infty} \frac{m+2j+l}{q+1} \times (-1)^j \frac{\binom{p}{j}\binom{m+j+l-1}{l-1}}{\binom{q+j+l}{q+1}} J_{m+2j+l}(v) \tag{C.11}$$

is true for l=1,2,... In order to show this we should first note that

$$\begin{aligned} \frac{m+2j+l}{q+1}(-1)^j \frac{\binom{p}{j}\binom{m+j+l-1}{l-1}}{\binom{q+j+l}{q+1}} = \; & \frac{\Gamma(-p+j)}{\Gamma(-p)} \frac{\Gamma(j+l)}{\Gamma(j+1)} \frac{\Gamma(q+1)}{\Gamma(q+j+l+1)} \\ & \binom{m+j+l-1}{l-1} (m+2j+l) \end{aligned} \tag{C.12}$$

This can be easily demonstrated by expanding the binomial terms and by noting that,

$$(-1)^j \binom{p}{j} = \frac{\Gamma(-p+j)}{\Gamma(-p)j!} \tag{C.13}$$

Let us imagine that equation C.11 is true for a particular l. From equations C.9 and C.12 we can compute the expression for $F_{nm}^{(l+1)}$

$$\begin{aligned} F_{nm}^{(l+1)}(v) &= \int_0^v t F_{nm}^{(l)}(t)dt \\ &= \sum_{j=0}^{\infty} \frac{\Gamma(-p+j)}{\Gamma(-p)} \frac{\Gamma(j+l)}{\Gamma(j+1)} \frac{\Gamma(q+1)}{\Gamma(q+j+l+1)} \\ &\times \binom{m+j+l-1}{l-1} (m+2j+l) \int_0^v t^{n+l+1} J_{m+2j+l}(t)dt \end{aligned} \tag{C.14}$$

Using equation 11.1.11 in Abramovitch and Stegun (1972) p.480 and replacing $\mu = n+1$, $\nu = m$ and $z = v$, we get,

$$\int_0^v t^{n+1} J_m(t)dt = v^{n+1} \sum_{k=0}^{\infty} \frac{(m+2k+1)\Gamma(-p+k)\Gamma(q+1)}{\Gamma(-p)\Gamma(q+k+2)} \cdot J_{m+2k+1}(v) \tag{C.15}$$

Replacing $n' = n+l$ and $m' = m+2j+l$, which gives $p' = p-j$ and $q' = q+j+l$

$$\begin{aligned} \int_0^v t^{n+l+1} J_{m+2j+l}(t)dt = \quad & v^{n+l+1} \sum_{k=0}^{\infty} (m+2(k+j)+l+1) \frac{\Gamma(-p+k+j)}{\Gamma(-p+j)} \\ & \times \frac{\Gamma(q+j+l+1)}{\Gamma(q+j+l+k+2)} \cdot J_{m+2(k+j)+l+1}(v) \end{aligned} \tag{C.16}$$

That can be inserted in equation C.14 and grouping the $j+k=s$ terms, we get

$$\begin{aligned} F_{nm}^{(l+1)} &= v^{n+l+1} \sum_{s=0}^{\infty} \frac{\Gamma(-p+s)}{\Gamma(-p)} \frac{\Gamma(q+1)}{\Gamma(q+l+s+2)} \\ &\times (m+l+2s+1) \times J_{m+l+2s+1}(v) \\ &\times \sum_{j=0}^{\infty} \frac{\Gamma(j+l)}{\Gamma(j+1)} \binom{m+j+l-1}{l-1} (m+2j+l) \end{aligned} \tag{C.17}$$

Considering the series on j of the last line in equation C.17, it can be shown by induction that

$$\sum_{j=0}^{\infty} \frac{\Gamma(j+l)}{\Gamma(j+1)} \binom{m+j+l-1}{l-1} (m+2j+l) = \frac{\Gamma(s+l+1)}{\Gamma(s+1)} \binom{m+s+l}{l} \tag{C.18}$$

Replacing in equation C.17, with $s = j$ and $l = l-1$ and using the expression C.12, we get

$$F_{nm}^{(l)}(v) = v^{n+l} \sum_{j=0}^{\infty} \frac{m+2j+l}{q+1} (-1)^j \binom{p}{j} \frac{\binom{m+j+l-1}{l-1}}{\binom{q+j+l}{q+1}} J_{m+l+j}(v) \tag{C.19}$$

which is equation C.11 that we wanted to demonstrate. Using this relation in equation C.10, we finally get the relation C.2 for T_{nm}.

Demonstration of the numerical expression of the V_n^m function

Let us now demonstrate that for $n, m \geq 0$ and such that $n-m \geq 0$ is even, the following equation is true

$$\begin{aligned} V_n^m(r,f) &= (-1)^m \int_0^1 \rho\, e^{if\rho^2} R_n^m(\rho) J_m(2\pi r\rho)\, d\rho. \\ &= (-1)^m e^{if} \sum_{l=1}^{\infty} (-2if)^{l-1} \sum_{j=0}^{p} v_{lj} \frac{J_{m+l+2j}(v)}{l(v)^l} \end{aligned} \tag{C.20}$$

where

$$v = 2\pi r \qquad p = \frac{n-m}{2} \qquad q = \frac{n+m}{2} \tag{C.21}$$

and v_{lj} is given by

$$v_{lj} = (-1)^p (m+l+2j) \frac{\binom{m+j+l-1}{l-1}\binom{j+l-1}{l-1}\binom{l-1}{p-j}}{\binom{q+l+j}{l}} \tag{C.22}$$

with $j = 0, 1, 2, ...$ and $l = 1, 2, ...$ Replacing the Zernike polynomials $R_n^m(\rho)$ in equation C.20 with their definition

$$R_n^m(\rho) = \sum_{s=0}^{p} \frac{(-1)^s (n-s)!}{s!(q-s)!(p-s)!} \rho^{n-2s} \tag{C.23}$$

and accounting for the definition of the T_n^m functions (equation C.1), we get

$$V_n^m(r,f) = (-1)^m \sum_{s=0}^{p} \frac{(-1)^s (n-s)!}{s!(q-s)!(p-s)!} T_{n-2s}^m \tag{C.24}$$

Using the expression of T_n^m previously demonstrated (equation C.2) and replacing $n = n - 2s$, that corresponds to $p = p - s$, $q = q - s$, we get,

$$V_n^m(r,f) = (-1)^m e^{if} \sum_{l=1}^{\infty} (-2if)^{l-1} \left[\sum_{s=0}^{p} \sum_{j=0}^{p-s} \frac{(-1)^s (n-s)!}{s!(q-s)!(p-j)!} t'_{lj} \frac{J_{m+l+2j}(v)}{v^l} \right] \tag{C.25}$$

where

$$t'_{lj} = (-1)^j \frac{m+l+2j}{q-s+1} \frac{\binom{p-s}{j}\binom{m+j+l-1}{l-1}}{\binom{q-s+l+j}{q+1-s}} \tag{C.26}$$

as defined in equation C.4. Manipulating the binomial terms, we find

$$t'_{lj} = (-1)^j (m+l+2j) \binom{m+j+l-1}{l-1} \frac{(p-s)!(q-s)!}{(p-j-s)!(j+q-s+l)!j!} \tag{C.27}$$

Let us now consider the double summation term in square brackets of equation C.25. It can be reorganized and simplified into:

$$\begin{aligned} [...] &= \sum_{j=0}^{p} (-1)^j (m+l+2j) \binom{m+j+l-1}{l-1} \frac{(j+l-1)!}{j!} \frac{J_{m+l+2j}(v)}{v^l} \\ &\times \sum_{s=0}^{p-j} (-1)^s \frac{(n-s)!}{s!(p-s-j)!(j+q-s+l)!} \end{aligned} \tag{C.28}$$

We call $S(p-j, n, p-j-l)$ the series over s. It comes form the definition of the binomial that

$$S(p-j,n,p-j-l)=\frac{1}{(p-j)!}\sum_{s=0}^{p-j}(-1)^j\binom{p-j}{s}\frac{(n-s)!}{(j+q-s+l)!} \tag{C.29}$$

It can be demonstrated, using a relation of recurrence (see Janssen 2002), that

$$S(p-j,n,p-j-l)=\begin{cases}0 & p-j\geq j\\ (-1)^k\binom{l-1}{p-j}\frac{(q+j)!}{(q+l+j)!} & p-j<l\end{cases} \tag{C.30}$$

We therefore get

$$\begin{aligned}V_n^m(r,f) &= (-1)^m e^{if}\sum_{l=1}^{\infty}(-2if)^{l-1}\sum_{j=0}^{p}(-1)^j(m+l+2j)\\ &\times \binom{m+j+l-1}{l-1}\frac{(j+l-1)!}{j!}\frac{J_{m+l+2j}(v)}{v^l}\\ &\times (-1)^(p-j)\binom{l-1}{p-j}\frac{(q+j)!}{(q+l+j)!}\end{aligned} \tag{C.31}$$

Rearranging the binomials and simplifying the terms that can be combined, this equation can be rewritten

$$V_n^m(r,f)=(-1)^m e^{if}\sum_{l=1}^{\infty}(-2if)^{l-1}\sum_{j=0}^{p}v_{lj}\frac{J_{m+l+2j}(v)}{lv^l} \tag{C.32}$$

with

$$v_{lj}=(-1)^p(m+l+2j)\binom{m+j+l-1}{l-1}\binom{j+l-1}{l-1}\binom{l-1}{p-j}\Big/\binom{q+l+j}{l} \tag{C.33}$$

which are the equations we wanted to demonstrate. We now have a convenient way of computing the $V_n^m(r,f)$ functions.

Appendix D

Generalization of the NZ equations when the complex β_N^0 coefficients is dominant

In this appendix, we generalize the development of the intensity function presented in appendix B. In this version we do not consider β_0^0 as the dominant aberration. Instead, any coefficient with $m = 0$ could be this dominant term. Moreover it does not have to be real anymore, the hypothesis on this aberration will be decided at the far end of the retrieval.

Let us start back from the general equation of the complex amplitude (equation 5.18)

$$U(r,\phi,f) = \sum_{n,m} \beta_n^m U_n^m(r,\phi,f) = \sum_{n,m} \beta_{cn}^m U_{cn}^m(r,\phi,f) + \sum_{n,m} \beta_{sn}^m U_{sn}^m(r,\phi,f) \tag{D.1}$$

with

$$\begin{aligned} U_{cn}^m(r,\phi,f) &= 2i^m\, V_n^m(r,f)\, \cos(m\phi) \\ U_{sn}^m(r,\phi,f) &= 2i^m\, V_n^m(r,f)\, \sin(m\phi) \end{aligned} \tag{D.2}$$

and,

$$V_n^m(r,f) = (-1)^m \int_0^1 \rho\, e^{if\rho^2} R_n^m(\rho) J_m(2\pi r\rho)\, d\rho \tag{D.3}$$

In this new "dynamic" case, β_N^0 is considered to be the dominant term, and can thus be extracted from the sum. The expression of the complex amplitude $U(r,\phi,f)$ as a function of the general Zernike coefficients is now given by

$$U = 2\beta_N^0 V_N^0 + 2{\sum_{n,m}}' \beta_{cn}^m i^m V_n^m \cos(m\phi) + 2{\sum_{n,m}}' \beta_{sn}^m i^m V_n^m \sin(m\phi) \tag{D.4}$$

where the "prime" sign means that the term $n = N, m = 0$ has been removed from the sums. The variables of the functions have been omitted for the sake of simplicity. We define the same shortcuts for the sum over the sine and cosine terms as in appendix B.

$$\begin{aligned}\sum \cos &= \sum_{n,m}{}' \beta_{cn}^m i^m V_n^m \cos(m\phi) \\ \sum \sin &= \sum_{n,m}{}' \beta_{sn}^m i^m V_n^m \sin(m\phi)\end{aligned} \tag{D.5}$$

The intensity is given by the square modulus of the complex amplitude $I = |U|^2$. Using the shortcuts defined above, we get

$$I = |U|^2 = \left[\Re(2\beta_N^0 V_N^0 + 2\sum \cos + 2\sum \sin)\right]^2 + \left[\Im(2\beta_N^0 V_N^0 + 2\sum \cos + 2\sum \sin)\right]^2 \tag{D.6}$$

Developing the squared terms gives,

$$\begin{aligned}|U|^2 \quad &= [2\Re(\beta_N^0 V_N^0)]^2 + [2\Re(\sum \cos)]^2 + [2\Re(\sum \sin)]^2 + 2[2\Re(\sum \cos)][2\Re(\sum \sin)] \\ &+8\sum_{n,m}{}' \Re(\beta_N^0 V_N^0)\, \Re(\beta_{cn}^m i^m V_n^m) \cdot \cos(m\phi) \\ &+8\sum_{n,m}{}' \Re(\beta_N^0 V_N^0)\, \Re(\beta_{sn}^m i^m V_n^m) \cdot \sin(m\phi) \\ &+[2\Im(\beta_N^0 V_N^0)]^2 + [2\Im(\sum \cos)]^2 + [2\Im(\sum \sin)]^2 + 2[2\Im(\sum \cos)][2\Im(\sum \sin)] \\ &+8\sum_{n,m}{}' \Im(\beta_N^0 V_N^0)\, \Im(\beta_{cn}^m i^m V_n^m) \cdot \cos(m\phi) \\ &+8\sum_{n,m}{}' \Im(\beta_N^0 V_N^0)\, \Im(\beta_{sn}^m i^m V_n^m) \cdot \sin(m\phi)\end{aligned} \tag{D.7}$$

We can gather the corresponding terms in real and imaginary parts and we express the intensity as a sum of six terms

$$I = |U|^2 = a + b + c + d + e + f \tag{D.8}$$

where a, b, c, d, e, f are developed hereafter.

$$a = 4[\Re(\beta_N^0 V_N^0)]^2 + 4[\Im(\beta_N^0 V_N^0)]^2 = 4\left|\beta_N^0 V_N^0\right|^2 \tag{D.9}$$

$$b = (2\Re(\sum \cos))^2 + (2\Im(\sum \cos))^2 = 4\left|\sum_{n,m}{}' \beta_{cn}^m i^m V_n^m \cos(m\phi)\right|^2 \tag{D.10}$$

$$c = (2\Re(\sum \sin))^2 + (2\Im(\sum \sin))^2 = 4\left|\sum_{n,m}{}' \beta_{cn}^m i^m V_n^m \sin(m\phi)\right|^2 \tag{D.11}$$

$$\begin{aligned} d &= 2(2\Re(\sum \cos))(2\Re(\sum \sin)) + 2(2\Im(\sum \cos))(2\Im(\sum \sin)) \\ &= 8\Re(\sum \cos)\Re(\sum \sin) + 8\Im(\sum \cos)\Im(\sum \sin) \end{aligned} \tag{D.12}$$

$$e = 8 \sum_{n,m}{}' \Re(\beta_N^0 V_N^0)\, \Re(\beta_{cn}^m i^m V_n^m) \cdot \cos(m\phi) + 8 \sum_{n,m}{}' \Im(\beta_N^0 V_N^0)\, \Im(\beta_{cn}^m i^m V_n^m) \cdot \cos(m\phi) \tag{D.13}$$

$$f = 8 \sum_{n,m}{}' \Re(\beta_N^0 V_N^0)\, \Re(\beta_{sn}^m i^m V_n^m) \cdot \sin(m\phi) + 8 \sum_{n,m}{}' \Im(\beta_N^0 V_N^0)\, \Im(\beta_{sn}^m i^m V_n^m) \cdot \sin(m\phi) \tag{D.14}$$

The terms b, c, d are quadratic terms that do not include the dominant aberration, we will not develop them since they are neglected during the retrieval process. Let us now develop the term called e. Gathering the sums, we get

$$e = 8\cos(m\phi) \sum_{n,m}{}' \left[\Re(\beta_N^0 V_N^0)\, \Re(\beta_{cn}^m i^m V_n^m) + \Im(\beta_N^0 V_N^0)\, \Im(\beta_{cn}^m i^m V_n^m)\right] \tag{D.15}$$

We introduce the following shortcut notations for the real and imaginary parts of V_N^0, β_N^0 and β_{cn}^m

$$V_N^0 = V_r + iV_i \qquad \beta_N^0 = \beta_{rN} + i\beta_{iN} \qquad \beta_{cn}^m = \beta_{rc} + i\beta_{ic} \tag{D.16}$$

where $V_r, V_i, \beta_{rN}, \beta_{iN}, \beta_{rc}$ and β_{ic} are real numbers. Let us now consider the product $\beta_N^0 V_N^0$, using the notation defined earlier, we get

$$\begin{aligned} \beta_0^0 V_N^0 &= (\beta_{rN} + i\beta_{iN})(V_r + iV_i) \\ &= \beta_{rN}V_r + i\beta_{iN}V_r + i\beta_{rN}V_i - \beta_{iN}V_i \\ &= (\beta_{rN}V_r - \beta_{iN}V_i) + i(\beta_{iN}V_r + \beta_{rN}V_i) \\ &= R + iI \end{aligned} \tag{D.17}$$

where R, I are real numbers representing respectively the real and imaginary part of the product. Replacing this expression of $\beta_N^0 V_N^0$ in equation D.15, we get

$$e = 8\cos(m\phi) \sum_{n,m}{}' \left[R\,\Re(\beta_{rc} i^m V_n^m) + R\,\Re(i\beta_{ic} i^m V_n^m) + I\,\Im(\beta_{rc} i^m V_n^m) + I\,\Im(i\beta_{ic} i^m V_n^m)\right] \tag{D.18}$$

Knowing that $R, I, \beta_{rc}, \beta_{ic}$ are real coefficients, the first ones (R, I) can enter and the second ones (β_{rc}, β_{ic}) can leave the $\Re$ and $\Im$ operators, which gives

$$e = 8\cos(m\phi) \sum_{n,m}{}' \left[\beta_{rc}\,\Re(i^m V_n^m R) + \beta_{rc}\,\Im(i^m V_n^m I) + \beta_{ic}\,\Re(i i^m V_n^m R) + \beta_{ic}\,\Im(i i^m V_n^m I)\right] \tag{D.19}$$

Let us now consider a complex number A, in this case, we have the following relations

$$\Re(i \cdot A) = -\Im(A) \qquad \Im(i \cdot A) = \Re(A) \tag{D.20}$$

and equation D.19 then becomes,

$$e = 8\cos(m\phi) \sum_{n,m}{}' \left[\beta_{rc}\, \Re(i^m V_n^m \{R - iI\}) - \beta_{ic}\, \Im(i^m V_n^m \{R - iI\})\right] \tag{D.21}$$

From the above definitions of R and I, we can deduce that

$$R - iI = (\beta_N^0 V_N^0)^* = \beta_N^{0*} V_N^{0*} \tag{D.22}$$

and equation D.21 becomes

$$e = 8\cos(m\phi) \sum_{n,m}{}' \left[\beta_{rc}\, \Re(\beta_N^{0*} i^m V_n^m V_N^{0*}) - \beta_{ic}\, \Im(\beta_N^{0*} i^m V_n^m V_N^{0*})\right] \tag{D.23}$$

we can also deduce from these definitions that

$$\beta_N^{0*} = \beta_{rN} - i\beta_{iN} \tag{D.24}$$

Replacing the expression of β_N^{0*} in equation D.23, taking β_{rN}, β_{iN} out of the $\Re, \Im$ operators and rearranging the terms leads to

$$e = 8\cos(m\phi) \sum_{n,m}{}' \left[(\beta_{rc}\beta_{rN} + \beta_{ic}\beta_{iN})\Re(i^m V_n^m V_N^{0*}) + (\beta_{rc}\beta_{iN} - \beta_{ic}\beta_{rN})\Im(i^m V_n^m V_N^{0*})\right] \tag{D.25}$$

Replacing all the shortcut notation we have defined and rearranging the terms, we get

$$\begin{aligned} e = 8\cos(m\phi) \sum_{n,m}{}' \big[\quad & [\quad \Re(\beta_N^0)\Re(\beta_{cn}^m) + \Im(\beta_N^0)\Im(\beta_{cn}^m)]\Re(i^m V_n^m V_N^{0*}) \\ - \quad & [\quad \Re(\beta_N^0)\Im(\beta_{cn}^m) - \Im(\beta_N^0)\Re(\beta_{cn}^m)]\Im(i^m V_n^m V_N^{0*})\big] \end{aligned} \tag{D.26}$$

Let us note that in the case where the dominant aberration would be β_0^0 and that $\Im(\beta_0^0) = 0$, the result we obtained here corresponds to the equation derived in appendix B. The parallelism between the two cases is then obvious, and the new equations can be obtained from the old ones by a simple variable change

$$\begin{aligned} \beta_0^0\, \Re(\beta_{cn}^m) \quad &\Longrightarrow \quad [\Re(\beta_N^0)\Re(\beta_{cn}^m) + \Im(\beta_N^0)\Im(\beta_{cn}^m)] \\ \beta_0^0\, \Im(\beta_{cn}^m) \quad &\Longrightarrow \quad [\Re(\beta_N^0)\Im(\beta_{cn}^m) - \Im(\beta_N^0)\Re(\beta_{cn}^m)] \end{aligned} \tag{D.27}$$

and where V_0^{0*} is replaced by V_N^{0*}. The term called f can be treated in the same way in order to obtain

$$f = 8\sin(m\phi)\sum_{n,m}{}' \Big[\quad [\quad \Re(\beta_N^0)\Re(\beta_{cn}^m) + \Im(\beta_N^0)\Im(\beta_{cn}^m)]\Re(i^m V_n^m V_N^{0*}) \\ - \quad [\quad \Re(\beta_N^0)\Im(\beta_{cn}^m) - \Im(\beta_N^0)\Re(\beta_{cn}^m)]\Im(i^m V_n^m V_N^{0*})\Big] \tag{D.28}$$

Based on the variable change expressed in equation D.27, the new Ψ^m_{meas} function can easily be obtained from the old one. It is given by

$$\Psi^0_{\text{meas}} \approx \frac{1}{2}(\beta_N^0)^2\chi_0^0 \quad + \quad \sum_n{}' [\Re(\beta_N^0)\Re(\beta_{cn}^m) + \Im(\beta_N^0)\Im(\beta_{cn}^m)]\ \chi_n^0 \\ + \quad \sum_n{}' [\Re(\beta_N^0)\Im(\beta_{cn}^m) - \Im(\beta_N^0)\Re(\beta_{cn}^m)]\ \Psi_n^0 \tag{D.29}$$

$$\begin{aligned}\Psi^m_{\text{meas}} \quad &\approx \quad \sum_n [\Re(\beta_N^0)\Re(\beta_{cn}^m) + \Im(\beta_N^0)\Im(\beta_{cn}^m)]\ \chi_n^m \\ &+ \quad \sum_n [\Re(\beta_N^0)\Im(\beta_{cn}^m) - \Im(\beta_N^0)\Re(\beta_{cn}^m)]\ \Psi_n^m \\ &+ \quad i\sum_n [\Re(\beta_N^0)\Re(\beta_{sn}^m) + \Im(\beta_N^0)\Im(\beta_{sn}^m)]\ \chi_n^m \\ &+ \quad i\sum_n [\Re(\beta_N^0)\Im(\beta_{sn}^m) - \Im(\beta_N^0)\Re(\beta_{sn}^m)]\ \Psi_n^m \end{aligned} \tag{D.30}$$

where the new Ψ_n^m and χ_n^m functions have the following new definitions, in which V_0^{0*} has been replaced by V_N^{0*},

$$\Psi_n^m(r,f) = -8\,\varepsilon_m^{-1}\,\Im[i^m V_n^m(r,f)V_N^{0*}(r,f)] \tag{D.31}$$

$$\chi_n^m(r,f) = 8\,\varepsilon_m^{-1}\,\Re[i^m V_n^m(r,f)V_N^{0*}(r,f)] \tag{D.32}$$

with $\varepsilon_m = 1$ for $m = 0$ and 2 in all other cases.

Calling the terms in square brackets respectively A_c, B_c, A_s, B_s, in equations D.29 and D.30, and applying the inner product as in the classical case leads to the same type of linear systems

$$\begin{cases} \dfrac{1}{2}(\beta_N^0)^2(\chi_N^0)^2 + \displaystyle\sum_n{}' A_c^0(\chi_n^0, \chi_{n'}^0) \approx (\Psi^0_{\text{meas}}, \chi_{n'}^0) \\ \displaystyle\sum_n{}' B_c^0(\Psi_n^0, \Psi_{n'}^0) \approx (\Psi^0_{\text{meas}}, \Psi_{n'}^0) \end{cases} \tag{D.33}$$

for $m = 0$ where $n, n' = 0, 2, ...$ For the real part of $Re(\Psi^m_{\text{meas}}) = \Psi^m_{c\text{meas}}$ and $m \neq 0$, we get

$$\begin{cases} \displaystyle\sum_n A_c(\chi_n^m, \chi_{n'}^m) \approx (\Psi^m_{c\text{meas}}, \chi_{n'}^m) \\ \displaystyle\sum_n B_c(\Psi_n^m, \Psi_{n'}^m) \approx (\Psi^m_{c\text{meas}}, \Psi_{n'}^m) \end{cases} \tag{D.34}$$

where $n, n' = m, m+2, ...$ For the imaginary part of $Im(\Psi^m_{\text{meas}}) = \Psi^m_{s\text{meas}}$ and $m \neq 0$ we get,

$$\begin{cases} \sum\limits_n A_s(\chi^m_n, \chi^m_{n'}) \approx (\Psi^m_{s\text{meas}}, \chi^m_{n'}) \\ \sum\limits_n B_s(\Psi^m_n, \Psi^m_{n'}) \approx (\Psi^m_{s\text{meas}}, \Psi^m_{n'}) \end{cases} \tag{D.35}$$

where $n, n' = m, m+2, ...,$. Solving these systems in the same way as for the classical theory (section 5.2.2) allows to determine $A^m_{cn}, B^m_{cn}, A^m_{sn}, B^m_{sn}$ and a simple transformation gives the β^m_n coefficients.