The 2017 conjunction of Beta Pictoris b : the life and death of Picsat, followed by a VLTI/GRAVITY observation of the re-emergence

Mathias Nowak

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Observatoire de Paris

The 2017 conjunction of beta Pictoris b

The life and death of PicSat,

followed by

A VLTI/GRAVITY observation of the re-emergence

Soutenue par
Mathias Nowak
Le 18 Septembre 2019

École doctorale n°127
Astronomie et Astrophysique d'Île-de-France

Spécialité
Astronomie & Astrophysique

Composition du jury :

Bruce Macintosh	Professor, Stanford University	*Rapporteur*
David Mouillet	Astronome, Institut de Planétologie et d'Astrophysique de Grenoble	*Rapporteur*
Carine Briand	Astronome, Observatoire de Paris	*Examinateur*
Julien Girard	Support Scientist, Space Telescope Science Institute	*Examinateur*
Matthew Kenworthy	Associate Professor, Leiden Observatory	*Examinateur*
Anne-Marie Lagrange	Directrice de recherche, Institut de Planétologie et d'Astrophysique de Grenoble	*Examinateur*
Samantha Thompson	Research Associate, University of Cambridge	*Examinateur*
Sylvestre Lacour	Chargé de recherche, Observatoire de Paris	*Directeur de thèse*

l'Observatoire de Paris | PSL

À ma maman,
qui n'y comprendra sûrement pas grand chose,
mais qui pourtant y est pour beaucoup.

Et à mes deux frères,
qui, poussés par je ne sais quelle intuition,
m'offrirent, il y à une quizaine d'années, une
carte du ciel et une boussole...

Contents

List of Figures

List of Tables

List of Acronyms

ADC	Analog-to-Digital Converter
ADCS	Attitude Determination and Control System
ADU	Arbitrary Digital Unit
AIT	Assembly, Integration, and Tests
ALMA	Atacama Large Millimeter/submillimeter Array
ANFR	Agence Nationale des Fréquences
AO	Adaptive Optics
APB	Advanced Peripheral Bus
ARIEL	Atmospheric Remote-sensing Infrared Exoplanet Large-survey
ASTERIA	Arcsecond Space Telescope Enabling Research in Astrophysics
AT	Auxiliary Telescope
AU	Astronomical Unit
BRITE	Bright(-star) Target Explorer
CAESAR	Conjunction Analysis and Evaluation Service, Alerts and Recommendations
CCSDS	Consultative Committee for Space Data Systems
CHARA	Center for High Angular Resolution Astronomy
CNES	Centre National d'Études Spatiales
CONICA	Coudé Near Infrared Camera
COTS	Commercial Off-The-Shelf
CRIRES	Cryogenic high-resolution Infrared Echelle Spectrograph
DAC	Digital-to-Analog Converter
DDT	Director's Discretionary Time
DEC	Declination
DIT	Detector Integration Time

DMA Direct Memory Access

DVD Digital Versatile Disk

EPS Electrical Power System

ESA European Space Agency

ESO European Southern Observatory

FIRST Fibered Instrument for Single Telescope

FM Flight Model

FTE Full-Time Equivalent

FWHM Full Width at Half-Maximum

GENIE Ground-based European Nulling Interferometery Experiment

GERICOS Generic On-Board Software

GPI Gemini Planet Imager

GPIO General Purpose Intput/Output

GRAVITY General Relativity Analysis via VLT Interferometry

HAL Hardware Abstraction Layer

HK House-Keeping

HOSTS Hunt for Observable Signatures of Terrestrial Systems

HSE High Speed External (clock signal)

HSI High Speed Internal (clock signal)

HST Hubble Space Telescope

HTTP Hypertext Transfer Protocol

HV High Voltage

HZ Habitable Zone

I2C Inter-Integrated Circuit

IARU International Amateur Radio Union

IC Integrated Circuit

IMU Inertial Measurement Unit

IR Infrared

IRAS Infrared Astronomical Satellite

IRS Infrared Spectrograph

ISRO Indian Space Research Organisation

ITU International Telecommunication Union

JSpOC Joint Space Operations Center

JWST James Webb Space Telescope

KISS Keep It Simple, Stupid

LBTI Large Binocular Telescope Interferometer

LEO Low Earth Orbit

LESIA Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique

LSE Low Speed External (clock signal)

MCMC Monte Carlo Markov Chain

NAOMI New Adaptive Optics Module for Interferometry

NAOS Nasmyth Adaptive Optics System

NASA National Aeronautics and Space Administration

NDIT Number of Detector Integrations

NORAD North American Aerospace Defense Command

NVIC Nested Vectored Interrupt Controller

OBC On-board Computer

OPD Optical Path Difference

PA Position Angle

PCA Principal Component Analysis

PDF Probability Density Function

PFI Planet Formation Imager

PFN Palomar Fiber Nuller

PID Process Identifier

PIONIER Precision Integrated-Optics Near-infrared Imaging Experiment

PLL Phase-Locked Loop

PPB Picsat Payload Bootloader

PR (drag) Poynting-Robertson (drag)

PRIMA Phase-Referenced Imaging and Micro-arcsecond Astrometry

PSF Point-Spread Function

PSLV Polar Satellite Launch Vehicle

QM Qualification Model

RA Right Ascension

RAM Random Access Memory

REF Réseau des Émetteurs Français

RS(422) Recommended Standard (422)

RTC Real Time Clock

RTOS Real Time Operating System

RW Reaction Wheel

RX Reception

SAGE Sagnac interferometer for Gravitational wave

SiDS Simple Downlink Share

SINFONI Spectrograph for Integral Field Observations in the Near-Infrared

SMF Single-Mode Fiber

SPAD Single Photon Avalanche Diode

SPHERE Spectro-Polarimetric High-contrast Exoplanet Research

SPI Serial Peripheral Interface

SSO Sun-Synchronous Orbit

STIS Space Telescope Imaging Spectrograph

TC Telecommand

TEC Thermo-Electric Cooler

TLE Two-Line Elements

TM Telemetry

TNC Terminal Node Controller

TPF Terrestrial Planet Finder

TX Transmission

UART Universal Asynchronous Receiver/Transmitter

UHF Ultra High Frequency

UT Unit Telescope

UTC Universal Time Coordinated

VHF Very High Frequency

VINCI VLT Interferometer Commissioning Instrument

VLT Very Large Telescope

VLTI Very Large Telescope Interferometer

WGS World Geodetic System

Foreword - About this thesis

My work at the Observatoire de Paris was mostly instrumental and observational. I tried to find new ways to observe the young β Pictoris stellar system, in order to better understand the origin of its giant planet β Pic b.

Since the discovery of its circumstellar disk, the β Pic system has been extensively studied. The literature on this astronomical object is abundant, with more than 100 dedicated papers published. In Part I of this thesis, I propose a review of the β Pictoris system, in which I go through what I think are the most important studies and discoveries made so far, and what are the current major open questions.

Initially, my work was supposed to be exclusively centered on PicSat, a three-unit CubeSat whose goal was to monitor the photometry of β Pictoris, and to observe the transit of β Pic b. I took the role of the *instrument scientist* for the mission during the first $\sim$ 2 yr of my Ph.D. The team working on PicSat was small ($\sim$ 5 people), and desigining a satellite, be it a small one, involves many different activities. So, I ended up working on several aspects of the mission, which spaned many different sub-fields of science and engineering (control theory, electronics, embedded programming, optical modeling, data analysis, database design and management, etc.). This heterogeneous work on PicSat is presented in Chapters 2 to 9 of this thesis, all grouped together under Part II.

Chapters 2 and 3 give a general introduction to the mission, and to its on-board instrument: a fibered photometer.

In Chapter 4, I present my solution to what is certainly one of the main problem of single mode photometry from space: how to maitain a stable injection in the fiber in the presence of residual satellite pointing errors?

There is an entire world between a conceptual algorithm to stabilize the injection in a single-mode fiber, as presented in Chapter 4, and an actual fibered instrument capable of tracking a target. This world is the world of *embedded programming*. When I started my Ph.D., *registers*, *peripherals*, *linker scripts*, *SPI* and *I2C protocols* were so many words foreign to me. Even a *microcontroller* was nothing more but a vague concept. I had to learn, and this took me quite some time. So much time, in fact, that I wanted to inflict to my reader some of the horrible details of progamming the STM32F303RT6 microcontroller of the PicSat instrument. So I wrote Chapter 5.

Chapter 6 focuses on the question of the data reduction, a topic probably more familiar to the astronomer. I spent some time studying how variations of the thermal environment of the satelilte could affect the optical quality of the instrument, how this optical quality was related to the photometric stability, and how to filter the data to guarantee its quality. The PicSat mission did not produce any science data, so this work remains a purely theoretical exercice. I decided to include it in this lengthy thesis anyway, as I think it can be of interest to some readers.

In Chapter 7, I give some details about the test campaigns which were performed on the PicSat instrument, and in which I was strongly involved. The goal of these test campaigns

was to verify that the instrument could withstand the launch, and the harsh conditions of the space environment. I also included my work on the optical alignment of the PicSat instrument in this chapter, simply because I did not know where else to include it.

I spent most of my time working on the science instrument, but I was also involved in the development of the ground segment for the mission. I was intially brought to it by necessity, simply because I needed a reliable format to send data in and out of the payload microcontroller. I started to dig into space communication protocols, and got a foothold in the ground segment. I then jumped into it when we later decided to involve the amateur radio community and the general public in the project, to retrieve data from the satellite. I got my amateur radio licence, which was required to operate the satellite, and I started to work on the mission website. This website rapidly became an important part of the colaboration with the amateur radio community, and since it was used to retrieve and display the satellite data, it was directly connected to the database. Before I realized what I was doing, it was already too late: I had become the official webmaster and database manager for the mission. I quickly present my activties on the ground segment in Chapter 8. This chapter severely deviates from the general astronomical topic of this thesis, and the reader can safely skip it if he or she wishes.

Chapter 9 is definitely shorter than I would have liked. In this chapter, I present the operation phase of the mission, which unfortunately lasted only 10 weeks. I give some details about the end of mission, and point to some of the possible causes. The reader will also find a conclusion about CubeSats, astronomy, and future perspectives at the end of this chapter.

The clear objective of this entire Part II was to answer key questions about the potential transit of β Pictoris b. I was also hoping that by producing a high-quality and consistent dataset over ~ 1 yr, PicSat would also help shed new light on the circumstellar disk and its current dynamics. But the harsh reality is that it did not work out that way.

A direct consequence for me was that I had to find another way to study β Pic b. I was fortunate enough to have the opportunity to observe the giant planet with the long-baseline interferometer GRAVITY at the Very Large Telescope. This observation, as well as the observation of HR 9799 e [Gravity Collaboration et al., 2019], were the first of their kind, and I had to develop a new method to reduce them. Part III, which contains Chapters 10 to 13 of this thesis, is dedicated to these long-baseline spectro-interferometric observations of β Pic b with VLTI/GRAVITY. This part is completely independant from Part II, and the reader does not need to read them in any particular order. In fact, she or he does not need to read both of them at all.

Chapter 10 is a general introduction to the role played so far by interferometry in exoplanet science, and Chapter 11 is a detailed introduction to interferometric observations in general, and to the GRAVITY instrument in particular. The reader already familiar with VLTI/GRAVITY can probably skip this chapter.

The observing strategy and the data reduction used to observe exoplanets with GRAVITY are detailed in Chapter 4. I developed the entire data reduction algorithm in a complex linear algebra formalism, which I believe is well suited to interferometry. This formalism makes use of complex random variables, which are presented in Annex B. The data reduction method itself is presented in Sections 12.4 and 12.5, in all its lengthy mathematical details.

The results obtained on β Pic b are presented in Chapter 13. This chapter arguably contains the main astrophysical contribution of this thesis. In particular, Section 13.5, which present a possible interpretation of the β Pic b observations in terms of formation history, is a direct attempt to contribute to some of the key questions outlined at the end of Part I.

Finally, a short general conclusion is given in Part IV.

Part I

BETA PICTORIS

Chapter 1

The beta Pictoris stellar system

Content of this chapter

1.1 A black swan in the sky?

It is estimated that our galaxy, the Milky-Way, contains about 250 billions stars. And although its exact shape is hard to determine precisely, we know that it is roughly a flat disc, with a diameter of about 200 000 light-years (ly), and a thickness ratio of 1/100th[1]. This represents a volume of about 6.3×10^{13} ly^3 (cubic light-years), leading to a stellar density of 3.9×10^{-3} stars/ly^3 in our galaxy.

Stellar density in the Milky-Way: $\rho_\star = 3.9 \times 10^{-3}$ stars/ly^3

The first planet outside of our Solar System was discovered in 1995, and since then, the number of confirmed exoplanets has steadily increased over the years. A few thousands are known today, suggesting that planet formation is not a rare phenomenon. In fact, a statistical analysis of microlensing data gathered during the period 2002-2007, and reported in Cassan et al. [2012], suggests that the majority of stars host at least one planet of any kind, with up to 20% hosting at least one giant planet.

Fraction of stars with a giant planet: $f_\mathrm{p} = 0.2$

As a general rule, small stars live calm and long lives (> 100 Gyr), whereas massive stars live spectacular but short lives ($\ll 1$ Gyr). Let's take 10 Gyr as the lifetime of a star. The typical duration of planet formation processes (from the formation of a planetary core, to the end of the cooling phase) is estimated to be a few tens of million years, say 50 Myr. If we consider that the total lifetime of a planet is the same as its host star, then the fraction of planets still in their cooling phase is ~ 0.005.

Fraction of systems in cooling phase: $f_\mathrm{c} = 0.005$

Considering a planetary system anywhere in the galaxy, what is the probability of being able to observe it in transit? The answer depends on the exact shape of the "planetary system". If we consider the entire protoplanetary or debris disk, studies suggest extensions of ~ 100 AU, with thickness increasing with radius, to reach ~ 10 AU at the external edge. As such, a debris disk would cast a shadow covering ~ 0.6 sr. The system is seen "in transit" if the Earth lies in these 0.6 sr, which should happen with a probability of $0.6/4\pi = 0.05$.

Fraction of transiting planetary systems: $f_\mathrm{t} = 0.05$

So, what is the probability of having a system seen edge-on (in transit), with a young giant planet orbiting around the star, close-enough to be easily detectable even with the naked eye (mag < 4.0)? To estimate this probability, we first need to take into account the fact that stars are not created equal, and that brighter stars can be detected from farther away. An estimate of the distribution of stars per magnitude in the solar neighborhood can be found in Ledrew [2001]. For each magnitude, we can calculate the maximum distance allowed for the star to have an apparent magnitude < 4:

Abs. mag.	0	1	2	3	4	5	6	7	≤ 8
Fraction f_m	0.001	0.003	0.005	0.01	0.015	0.03	0.03	0.03	0.876
Max dist. $D_{\mathrm{max},m}$ (ly)	190	120	77	49	30	19	12	7	< 5

The probability of having a star with a planet seen edge-on with a magnitude < 4 is then given by:

$$P = \sum_m \frac{4\pi}{3} (D_{\mathrm{max},m})^3 \times f_m \times \rho_\star \times f_\mathrm{p} \times f_\mathrm{c} \times f_\mathrm{t} \simeq \frac{1}{100}$$

And yet, such a system does exist in our Galaxy: β Pictoris. Are we really *that* lucky?

[1] Picture a DVD...

1.2 Beta Pictoris

With a visible band magnitude of $M_V = 3.86$, β Pictoris is the second brightest star of the Pictor constellation (Figure 1.1), and can easily be seen with the naked eye, under medium conditions. It is a star of the Southern sky, though, inaccessible to an observer in the Nothern hemisphere.

The basic parameters of the star are relatively well constrained (see Table 1.1). It is an A6V star [Gray et al., 2006], with a temperature of ~ 8000 K, a surface gravity of $\log(g) \simeq 4.1$ to 4.2, a near Solar metalicity of [Fe/H] $\simeq 0.05$ [Lanz et al., 1995, Gray et al., 2006]), and a mass of $\simeq 1.8\,M_\odot$ [Lanz et al., 1995, Crifo et al., 1997]. It was also noted by Crifo et al. [1997] that β Pictoris was a very young star, which was recently confirmed by a study of the β Pictoris Moving Group [Mamajek and Bell, 2014], setting the age of the star to 23 ± 3 Myr. The diameter of β Pic has been measured by long baseline interferometry: Di Folco et al. [2004], Kervella et al. [2004] found and angular diameter of 0.84 ± 0.04 mas with VLTI/VINCI, whereas a more recent observation with VLTI/GRAVITY is reported in Defrère et al. [2012], who obtained an angular diameter of 0.736 ± 0.019 mas. Given the distance of β Pic (19.3 pc, from Hipparcos measurements), this translates to a physical radius of ~ 1.5 to $1.7\,R_\odot$.

The star is also a δ-Scuti variable, with frequencies ranging from $\simeq 30$ d^{-1} to $\simeq 70$ d^{-1} [Koen, 2003, Galland et al., 2006, Mékarnia et al., 2017].

Interestingly, if the overall metalicity of β Pictoris seems to be well constrained, its chemical composition, and in particular its Carbon-to-Oxygen ratio, which is of prime interest for the study of the planetary system, remains to be determined. The only estimate of the chemical composition of β Pic currently available in the literature comes from Holweger et al. [1997], who determined via spectroscopic observations that the abundances of C, Ca, Ti, Cr, Fe, Sr, and Ba were all solar. The O abundance is still unknown.

Magnitude (V band)	Age (Myr)	Distance (pc)	Diameter ($R_\odot$)	Temperature (K)	Surface grav. $\log(g/g_0)$	Mass ($M_\odot$)	Metalicity [Fe/H]
3.86	23 ± 3	19.3	1.5 to 1.7	8000	4.1	1.8	0.05

Table 1.1: Main characterestics of β Pictoris

1.3 Circumstellar-disk around beta Pictoris

1.3.1 Initial discovery

Initial clues for the presence of circumstellar material around β Pictoris were found in 1984 when observations with the Infra-Red Astronomical Satellite (IRAS) led to the discovery of infrared excess emission around many nearby stars. In 1984, it was announced that circumstellar material had been detected around β Pictoris, as well as three other stars: Vega, Fomalhaut, and ϵ Eridani (see Aumann [1985]). In the case of Vega (α Lyrae), the same investigators also suggested that the detected circumstellar material could be associated with planet formation.

In 1984, no exoplanet had ever been detected. The first planet outside the solar system was only discovered in 1995 [Mayor and Queloz, 1995]. Thus, the idea of being able to connect the infrared emission detected with IRAS with the presence of potential planets sparked the interest of a few different teams, and in 1984, Smith and Terrile started to look for an optical counterpart to the infrared excess detected around β Pictoris. They reported that "using a special optical instrument known as a coronagraph", they were "successful in obtaining optical

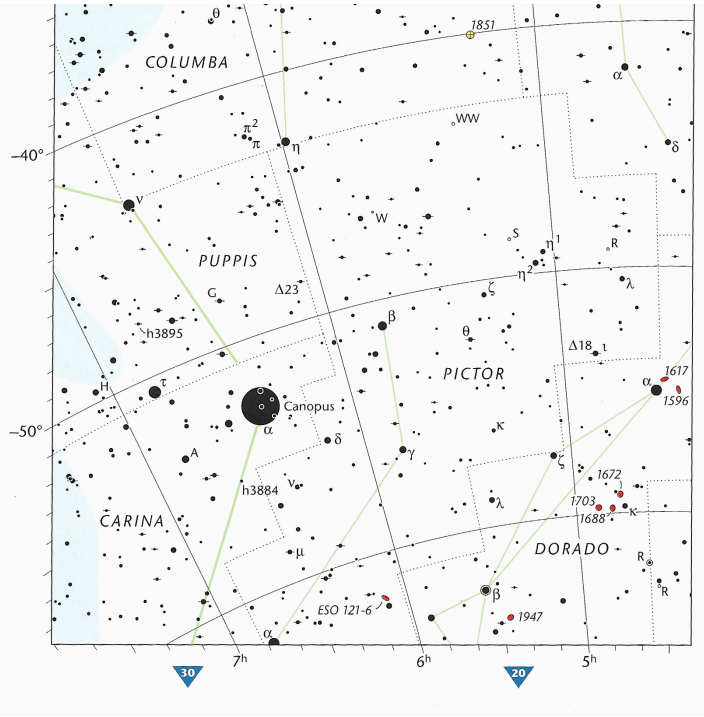


Figure 1.1: The small Pictor constellation (reproduced from the Sky & Telescope's Pocket Sky Atlas)

images of the disk [around β Pictoris]". This was the beginning of a quest to understand the origin of the disk which is still ongoing, more than three decades after this initial discovery.

To satisfy the undoubtedly curious reader, this first image of the β Pictoris disk, obtained by Smith and Terrile with their optical coronagraph mounted on the 2.5 meter Dupont Telescope is presented in Figure 1.2, along with a more recent picture, obtained with STIS, on-board the Hubble Space Telescope (HST)

1.3.2 General structure

The general structure of the circumstellar material around β Pictoris is that of a circumstellar disk oriented North-East (NE) to South-West (SW) (position angle of $\sim 29°$, Apai et al. [2015]), extending to about 400 AU [Smith and Terrile, 1984], and rotating toward us on the SW side [Olofsson et al., 2001].

The first series of observations of the circumstellar disk around β Pictoris were mainly aimed at determining the origin of the detected emission, as well as the disk radial profile (in terms of emission and density). It rapidly became clear that the detected emission was light from the star scattered by dust grains, which had to be relatively large ($> 1\mu$m) to explain why the scattering was achromatic [Paresce and Burrows, 1987]. Some peculiar characteristics were also found in the photometric radial profile, with abrupt changes in the profile detected by Golimowski et al. [1993]. Lagage and Pantin [1994] also reported a dust-depletion region within 40 AU of the system, which had already been suggested by Backman et al. [1992]. Lagage and Pantin suggested that this depletion zone could be linked to the presence of at least one planet orbiting in the system.

As observational capabilities progressed, images also started to unveil a set of asymmetries in the disk. Kalas and Jewitt [1995] observed β Pictoris in R band with a 2.2 m telescope

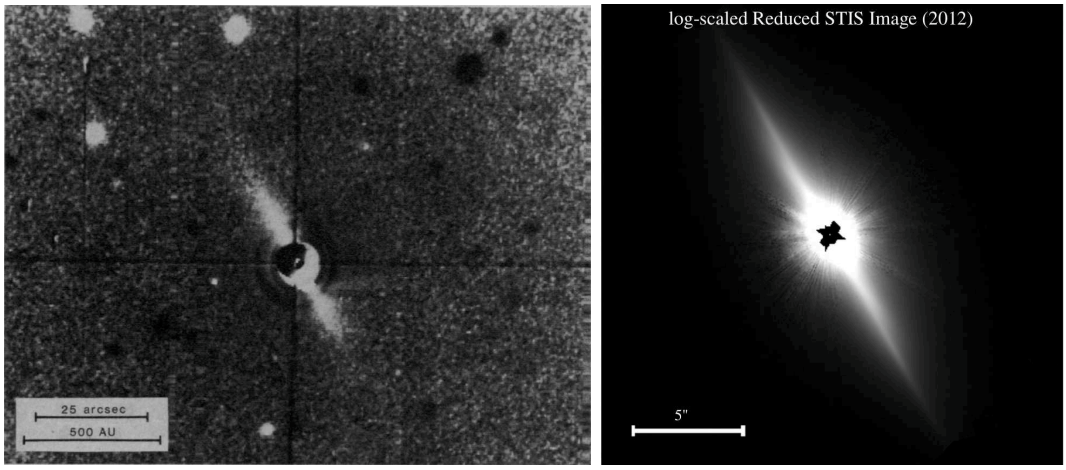


Figure 1.2: Left: the historical image of the β Pictoris circumstellar disk obtained by Smith and Terrile [1984] in optical light, with a 2.5 meter telescope and a coronagraph. Right panel: a more recent image, obtained in 2012 with STIS, on-board Hubble, and presented in Apai et al. [2015]. In both images, North is up, East is left.

and a coronagraph. They reported a total of 5 different asymmetries, consistently observed in their images: a size asymmetry, the NE size of the disk being more extended than the SW side; a brightness asymmetry, with the NE side brighter than the SW side; a width asymmetry, with the SW side becoming larger than the NE side beyond 7 as (arcsecond); a butterfly asymmetry, which is a more complex axial asymmetry around the mid-plane; and a wing-tilt asymmetry, with a 1.3 deg offset between the two sides of the disk. They suggested that the wing-tilt asymmetry was due to the fact that the disk was not seen exactly edge-on from the Earth, but with a slight tilt of $\sim$ 2 to 4 degrees. But the other asymmetries could not be explained by any effect other than intrinsic asymmetries of the disk structure. The butterfly asymmetry was explained ten years later, with the discovery of a secondary disk inclined by 5 degrees with respect to the primary disk [Golimowski et al., 2006].

A set of three planetesimal belts, located at approximately 6, 16, and 30 AU were found by Okamoto et al. [2004], and another larger one at 90 AU was later found by Wilner et al. [2011]. These belts are thought to act as reservoirs, continuously enriching the disk in gas and dust, and are thus arguably the main structuring elements. Some of these planetesimal belts also seem to be co-located, or at least very close to, the dust rings found by Wahhaj et al. [2003] (at 14, 28, 52, and 92 AU). Although the 52 AU ring does not seem to have a planetesimal belt counterpart, Telesco et al. [2005] reported observations of a dust clump on the SW side, at the same radial distance. They suggested that this large clump could either be due to the grinding of a group of planetesimals resonantly trapped, or the consequence of a cataclysmic breakup of planetesimals. In either case, the clump seems to indicate the presence of planetesimals at $\sim$ 50 AU, even though no belt was reported by Okamoto et al. The presence of this bright clump of dust particles on the SW side can also explain the brightness asymmetry initially reported in Kalas and Jewitt [1995].

The location of the giant planet β Pictoris b, determined via an extensive orbital survey, is somewhere in between the 6 and 16 AU belts, the best estimate for the semi major-axis being $\sim$ 10 AU [Lagrange et al., 2019]. The presence of a planet between the 6 and 16 AU belts was actually already suggested by Okamoto et al. in their 2004 paper (see Figure 1.3).

More recently, Matrà et al. [2019] found that the external Kuiper belt at $\sim$ 100 AU was

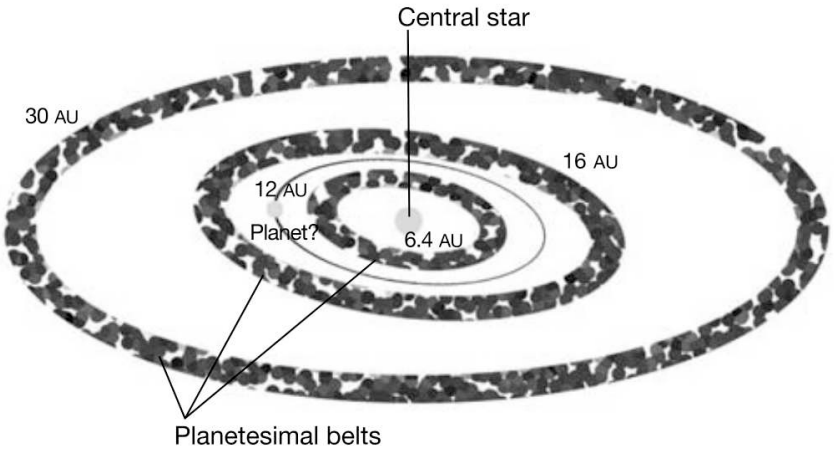


Figure 1.3: A depiction of the three inner planetary belts around β Pictoris, with a suggested planet orbiting between the first and second belts. The giant planet β Pic b was not known at the time, although the general structure of the disk was highly suggestive of the existence of such planets. The planet was discovered a few years later, and its semi major-axis was estimated to be ~ 10 AU, in perfect agreement with the general structure shown here. Figure taken from Okamoto et al. [2004].

actually made of two different populations of objects, with different orbital inclinations. They reported the discovery of a "dynamically hot" population with an inclination of $\sim 8.9°$, and a "dynamically cold" population, with an inclination of $\sim 1.1°$.

1.3.3 Dust in the disk

The disk orbiting around β Pictoris is known to be a debris-disk, in which planetesimal collisions is the main source of dust. In fact, according to Wilner et al. [2011], the outer belt, located at ~ 90 AU is the main reservoir of dust.

In a debris disk, the dynamics of the dust is constrained by a few mechanisms. Dust is produced through a collision cascade in planetesimal belts: collisions between large objects create smaller objects which collide to create even smaller debris, etc. The size distribution resulting from such a cascade has been shown to follow a power-law:

$$\rho(D) \propto D^{-\frac{7}{2}}$$

with $\rho(D)$ the density of particles of size D.

The evolution of a dust particle is mainly governed by two different forces: the gravitational force from the central star, and the radiation pressure. The radiation pressure can be further subdivided in two components with opposing effects: the radial component of radiation pressure tends to push the particles out of the system, whereas the tangential component tends to brake the particles, and thus to drive them inward into the system. This tangential part of radiation pressure is known as Poynting-Robertson drag (PR drag), and can be intuitively understood as the result of the dust particle moving into the photon flow: from the particle's point of view, light from the star is coming slightly from the forward direction, effectively creating a drag.

For a particle of typical size D orbiting at a distance R from the star, the gravitational force is given by:

$$F_{\rm grav} = \frac{\pi G M_\star \rho D^3}{6R^2}$$

with G the gravitational constant, $M_\star$ the mass of the central star, and ρ the density of the grain material.

Assuming that the grains perfectly absorb the light coming from the star, radiation pressure is given by:

$$F_{\rm rad} = \frac{L_\star D^2}{16R^2 c}$$

with $L_\star$ the bolometric luminosity of the central star, in Watts, and c the speed of light.

The ratio between those two forces is usually denoted β. It is independent of the orbital distance R, and given by:

$$\beta = \frac{F_{\rm rad}}{F_{\rm grav}} = \frac{6L_\star}{16\pi G c M_\star D\rho} = A \times \left(\frac{L_\star}{L_\odot}\right)\left(\frac{M_\star}{M_\odot}\right)^{-1} \times \rho^{-1} D^{-1}$$

with $A \simeq 1.1 \times 10^{-3}$ m^{2}s^{-2}kg^{-1}.

Depending on the value of its associated parameter β, a dust particle can have three different behaviors:

- if $\beta > 0.5$, the particle is blown out of the system[2].
- if $\beta < 0.5$, the particle remains on a bound orbit, although the effective gravitational potential felt by the particle is reduced (multiplied by a factor $1-\beta < 1$ compared to the normal case without radiation pressure). If the particle results from a collision in a planetesimal belt, it will move to an eccentric orbit, with a perihelion located somewhere in the belt. PR drag will then tend to circularize its orbit, while at the same time make it fall inward.
- for $\beta \ll 1$ (large grains), the radiation pressure is negligible compared to the gravitational force. These particles follow traditional Keplerian orbits. If the particle is created from collisions in a planetesimal belt, it will stay in the belt.

Assuming a typical dust density $\rho = 2500$ gm^{-3} (silicate grains), in the β Pic system where $L_\star \simeq 9\,L_\odot$ and $M_\star \simeq 1.7\,M_\odot$, the critical size below which a particle is blown away by radiation pressure is $\simeq 2$ μm. Depending on the exact density and optical properties of the grains, this critical size might vary significantly, but we can still expect all grains of size < 0.1 μm to be blown out of the system.

Despite this general theoretical understanding of how the dust behaves in a disk, it is very difficult to obtain observational confirmations. Grains of different sizes tend to radiate

[2]It may seem counter-intuitive that the limiting value is 0.5 and not 1. But this is because the initial conditions must be taken into account. A particle initially at rest would indeed be blown away for $\beta > 1$, and would fall on the star for $\beta < 1$. But if the particle is created by a collision between two planetesimals in a belt, its initial velocity is the Keplerian velocity of said planetesimals (the collisional velocity is negligible compared to orbital velocities). The belt being in a circular orbit, if we denote R its distance to the central star, then the orbital velocity is $v_{\rm orb} = \sqrt{GM_\star/R}$, whereas the escape velocity at the same distance R is $v_{\rm esc} = \sqrt{2GM/R}$. After the collision, the dust particle of given parameter β is moving at $v_{\rm orb}$, but because of the radiation pressure, the gravitational potential it feels is reduced. The escape velocity of this particle is now $v_{\rm esc} = \sqrt{2(1-\beta)GM_\star/R}$, hence the threshold at $\beta = 0.5$.

at different temperatures, and thus in different parts of the spectrum. This thermal emission adds to the scattered emission, and so far it proved very difficult to correctly predict both emissions [Ballering et al., 2016]. The presence of multiple planetesimal belts also tend to blur the situation. But some attempts were made to try to build a model of the dust distribution and the associated emission in the β Pic system (see Ahmic et al. [2009], Ballering et al. [2016]). The situation in the inner part of the system, were warm dust has been detected in mid-infrared and hot dust in the near-infrared [Chen et al., 2007, Li et al., 2012, Defrère et al., 2012] is even more complicated to study observationally, because it often remains unresolved by our instruments.

Different minerals have been identified in the β Pictoris disk. Observing the disk in the mid-infrared, Li et al. [2012] noted the presence of two components within the first 5 to 10 AU of the disk: glassy olivine and crystalline forsterite. Small grains of olivine (~ 0.1 μm) were observed on the two sides of the star, but not in the central part, while 2 μm grains were found in the central part. This was already noted by Okamoto et al. [2004], who suggested that this was indicative of the presence of a belt at 6 AU, which would be the origin of the glassy olivine. In this scenario, small particles are ejected out of the belt, while bigger particles spiral inward, creating the observed distribution. Li et al. [2012] confirmed that this was a plausible scenario, although not the only one. They also suggested alteration and destruction of the dust grains by sublimation as a possible explanation to some of the observed structures. They show that smaller grains of olivine tend to reach higher temperatures when exposed to stellar radiation, thus making them more susceptible to destruction in the inner part of the system. This can potentially create a deficit of small grains, leading to a similar distribution as the planetesimal belt scenario.

Trying to model the dust in the outer parts of the system, Ballering et al. [2016] found a quite different composition, with a large amount of silicates and astronomical organics, and small amounts of water ice.

1.3.4 Gas disk

Observations of CO in the disk around β Pic [Dent et al., 2014] have shown that gas is continuously being produced in the system, as the lifetime of CO (< 150 yr) is much smaller than the age of the system (~ 20 Myr). Most of the gas which is observed today probably comes from planetesimal collisions, photo-desorption, or other phenomena continuously creating it. In fact, in their observations performed with ALMA, Dent et al. [2014] found a rather peculiar distribution of CO, with a large clump located on the SW side of the disk, which could be a sign of a recent cataclysmic collision (see Figure 1.4).

The disk around β Pic is seen almost edge on, which makes possible to observe the gas in absorption spectroscopy. Many different species have been found using this technique, and we know today that the gas contains C, O, Na, Mg, Al, Si, S, Ca, Cr, Mn, Fe, Ni [Roberge et al., 2006]. But this is extremely puzzling, as some of those elements (like Na) are very sensitive to radiation pressure, and should be blown out of the system too quickly to reach observable concentrations. Yet Olofsson et al. [2001] found Na I in Keplerian orbit around β Pic.

The question of how metals could stay long enough to reach observable quantities in the β Pic system has been a long-standing question. The most widely accepted explanation is that these metals are subjected to another force which adds to gravity and radiation pressure: Coulomb forces between gaseous ions and neutral atoms [Roberge et al., 2006, Fernández et al., 2006]. Some elements, like C and O are much less sensitive to radiation pressure. An overabundance of carbon, ionized by the stellar radiation (observations of ionized carbon are reported in Cataldi et al. [2014]), can effectively create an ionized disk which brake metallic elements and keep them in the system.

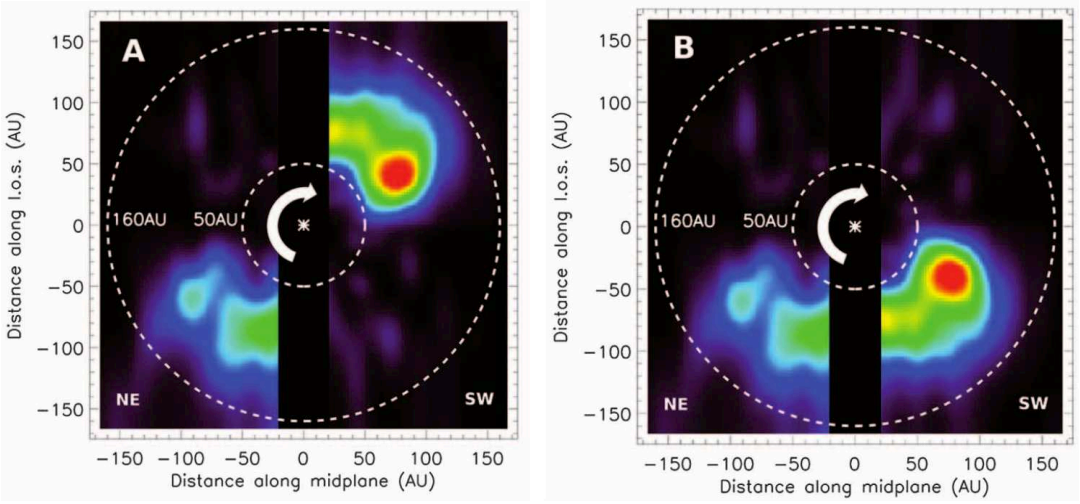


Figure 1.4: Two possible distributions of the CO gas detected with ALMA, and de-projected assuming Keplerian circular orbital motion. The left-hand panel shows a distribution with two clumps on opposite sides of the star, indicative of a resonance trapping mechanism with a giant planet other than β Pic b. The right-hand panel shows the second possible distribution, with only one clump, resulting from the cataclysmic collision of Mars-sized objects. Figure taken from Dent et al. [2014]

Contrary to what was suggested by Roberge et al. [2006], it has been shown that C is not the only element which seems to be overabundant in the disk. Using the Hershel space telescope, Brandeker et al. [2016] observed Oxygen around β Pic and although they could not provide a reliable estimate of the C/O ratio, they noted that it is present in large quantities. Oxygen is also only weakly affected by radiation pressure [Fernández et al., 2006], and its presence in ionized forms can also participate to the braking of metallic elements.

To explain this apparent overabundance of C and O, Xie et al. [2013] introduced two hypotheses: preferential production of these two elements (for example if most of the gas comes from the release of CO gas by solid bodies, see Kral et al. [2016]) or preferential depletion, because other metallic elements can still be blown away more quickly by the radiation pressure. An interesting argument in favor of the preferential production mechanism came from observations with the Hubble Space Telescope reported in Wilson et al. [2019]. Using the Cosmic Origin Spectrograph, they detected N gas, and estimated it to be in roughly Solar abundance. Since N is also very weakly affected by radiation pressure, if the overabundance of C and O is explained by a preferential depletion mechanism, N should also be overabundant, which it is not.

In the same study, Wilson et al. also noticed that the N gas does not seem to be falling onto the star, contrary to what they observed when looking at hydrogen gas (see Wilson et al. [2017]). To explain this difference, they suggested that the gas was produced in the inner system by the two different populations of comets reported in Kiefer et al. [2014]: in this scenario, N gas is mainly produced by comets S, and H gas by comets D (see after).

1.3.5 The inner system

The fact that the circumstellar disk is seen almost edge-on around β Pic makes it much easier to observe the inner part (< 1 AU) by absorption spectroscopy. Early-on, high-resolution spectroscopic observations revealed the presence of transient changes in the red wing of the

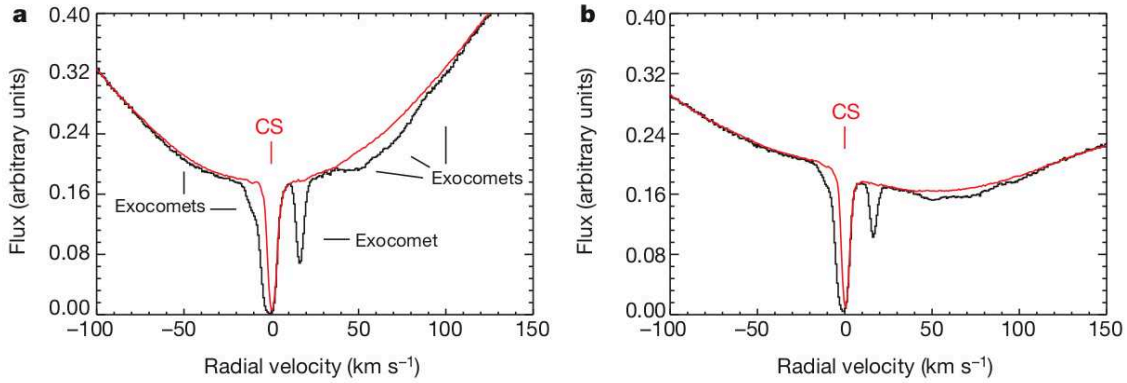


Figure 1.5: Spectrum of β Pictoris in the Ca II K-line (left), and Ca II H-line (right), showing transient events. The red line represent the spectrum obtained when transient event are subtracted (circumstellar disk only), and the black line represent the spectrum obtained with transient events. Figure taken from Kiefer et al. [2014]

Ca K line (see Figure 2.2 for an example of a spectrum), which were interpreted as the presence of kilometer sized bodies falling towards the star: the so-called "falling evaporating bodies" [Beust and Lissauer, 1994, Beust et al., 1996]. Transient events in the blue wing of the line were also discovered a few years later [Crawford et al., 1998], which were again interpreted as another population of comets falling onto the star.

It was shown by Kiefer et al. [2014] that two populations of transiting exocomets were coexisting in the inner β Pictoris system. Some comets share very similar orbits, with periastron at ~ 0.15 AU, and constitute the population "D", which most resemble the Kreutz family of comets in our own Solar System. The other population "S" is made of bodies with periastron < 0.1 AU, spread over a large interval of longitudes. Kiefer et al. suggested that the population "S" could be made of exocomets trapped in a mean motion resonance by β Pic b, and which could have slowly evolved to highly eccentric orbits. This scenario is supported by the lower evaporation rate observed for this population, a result of exhaustion by numerous passes at perihelion.

These falling evaporating bodies appear to be mainly composed of an icy core surrounded by a refractory crust [Karmann et al., 2001]. They clearly seem to be the main component driving the dynamics of the inner system, as they provide a source of constant gas enrichment (Vidal-Madjar et al. [2017], Wilson et al. [2017, 2019], Kiefer et al. [2019]).

One last component which has been detected in the inner system is hot dust, radiating at a temperature of 1500 K, seen via long-baseline interferometry in the H band [Defrère et al., 2012].

1.4 Discovery and long-term monitoring of beta Pic b

1.4.1 Multiple hints from the disk structure

After the initial discovery of the circumstellar disk around β Pic, it rapidly became clear that many of the observed structures and morphological asymmetries could be explained by the presence of a planetary system. Golimowski et al. [1993], Lagage and Pantin [1994], Kalas and Jewitt [1995] all reported anomalies in the shape of the disk, which could be linked to the presence of a planet, and Roques et al. [1994], Lazzaro et al. [1994], Mouillet et al. [1997]

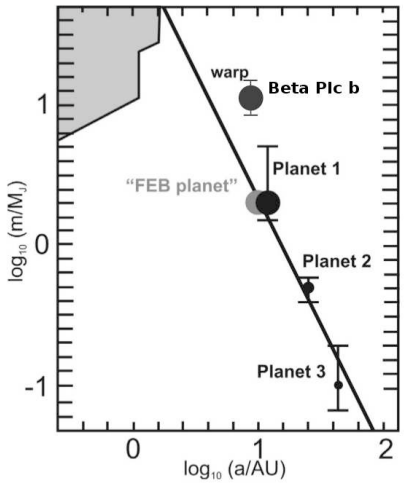


Figure 1.6: Potential planets proposed by Freistetter et al. [2007] to explain some of the asymmetries seen in the β Pic disk ("Planet" dots). The gray area is the locus excluded by radial velocity measurements available at the time [Galland et al., 2006]. The black line corresponds to the planet suggested in Mouillet et al. [1997], while the gray dot was suggested in Beust and Morbidelli [2000] to explain the falling evaporating bodies. For comparison, a "Beta Pic b" dot has been added here which corresponds to the best estimate for β Pic b (semi major-axis from Lagrange et al. [2019], and mass from Snellen and Brown [2018]). Original figure from Freistetter et al. [2007].

started to investigate the hypothesis from a modeling perspective. Mouillet et al. suggested the presence of a giant planet (up to $10\,M_{\rm Jup}$) located below 20 AU.

But asymmetries in the disk were not the only elements indicative of the presence of a potential planet. Lecavelier Des Etangs et al. [1995] reported observations of variations in the photometry of β Pictoris, and suggested that this could be explained either be a transiting dust cloud [Lamers et al., 1997], or a transiting planet [Lecavelier Des Etangs et al., 1997].

After more than two decades of study, no planet was found in the β Pic disk, but evidence for the presence of a planetary system were pilling up. Heap et al. [2000], Wahhaj et al. [2003], Telesco et al. [2005] uncovered new asymmetries and structures, and Freistetter et al. [2007] proposed a model explaining several of them, as well as the presence of falling evaporating bodies, under the assumption that at least one planet of a few Jupiter masses was orbiting β Pic at $\sim$ 12 AU. They also suggested that two other smaller planets were hiding at greater distances from the star (see Figure 1.6).

1.4.2 High-contrast observations and confirmation

It is often the case in Astronomy that new instrumental technologies bring new discoveries. Exoplanet science and debris disk study are certainly no exception. With the advent of Adaptive Optics (AO) in the 1990s, and its rapid progress during the 2000s, it became possible to observe faint targets at high resolution. Combined with coronagraphic instruments, this technology also made possible to reach extremely high-contrast, enabling the detection of faint companions to bright stars.

Lagrange et al. [2009a] reported on observations acquired in 2003 with NAOS-CONICA (NaCo) on the Very Large Telescope, in which a faint point-like source was detected, close to β Pictoris (Figure 1.7). In this initial publication, Lagrange et al. could not definitely

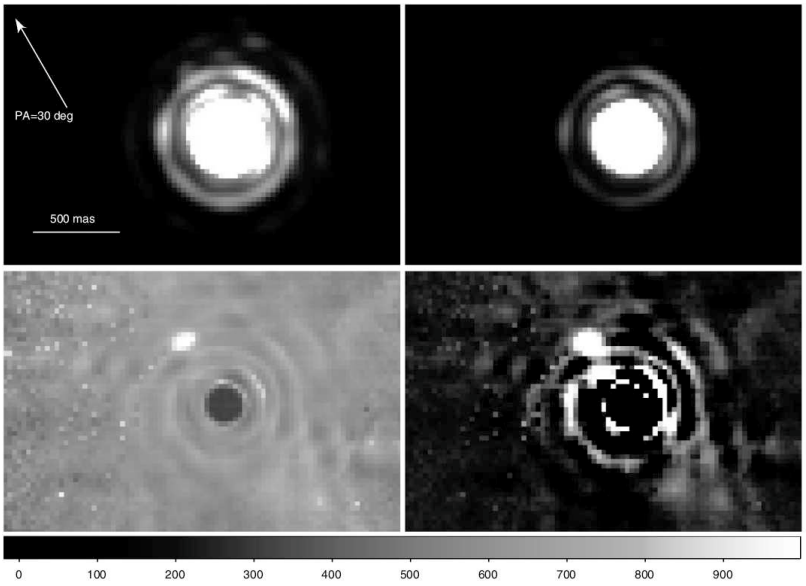


Figure 1.7: First image of the giant planet β Pictoris b obtained with NAOS-CONICA on the VLT, in 2003. The upper-left panel shows the image of β Pic obtained in L' band. The upper-right panel is a calibrator (HR 2435). The lower-left image is the divided image (β Pic divided by HR 2435), and the lower-right is a subtracted image (β Pic minus HR 2435, with a scaling factor applied to HR 2435). Figure taken from Lagrange et al. [2009a].

rule out a possible instrumental contamination or a background source, and new observations in 2009 did not show any point-source close to β Pic [Lagrange et al., 2009b]. But in 2010, the planet candidate reappeared on the other side of the star, in agreement with a Keplerian motion, and this confirmed that it was really a bound object [Lagrange et al., 2010].

1.4.3 Monitoring and orbit determination

After its discovery and confirmation, the giant planet β Pictoris b became the target of an extensive monitoring to characterize its orbital motion. Since it had been suggested by Lecavelier Des Etangs and Vidal-Madjar [2009] that β Pic b could be a transiting planet, a key question quickly became: will it transit again, and when?

A transit of β Pic b in front of its bright host star would be extremely valuable to determine its exact size, as well as to study its Hill sphere. Lecavelier des Etangs and Vidal-Madjar [2016] showed that two families of orbits were compatible with the measurements obtained on β Pic b and a transit in 1981. The planet could be on a 36 yr low-eccentricity, or on an 18 yr high-eccentricity (0.3) orbit, respectively leading to a new transit in early 2017 or 2018. But soon after, new measurements obtained with the Gemini Planet Imager (GPI, Macintosh et al. [2014]) strongly disfavored the hypothesis of a transit of β Pic b [Wang et al., 2016]. A transit of the planet's Hill sphere, though, is still possible. In fact, Wang et al. predicted a transit of the Hill sphere, starting in 2017, and ending in 2018 (see Figure 1.8).

This prediction sparked the interest of the exoplanet community, and triggered a large international campaign to monitor the star during the 2017-2018 period. The work presented in Part II of this manuscript was part of this campaign.

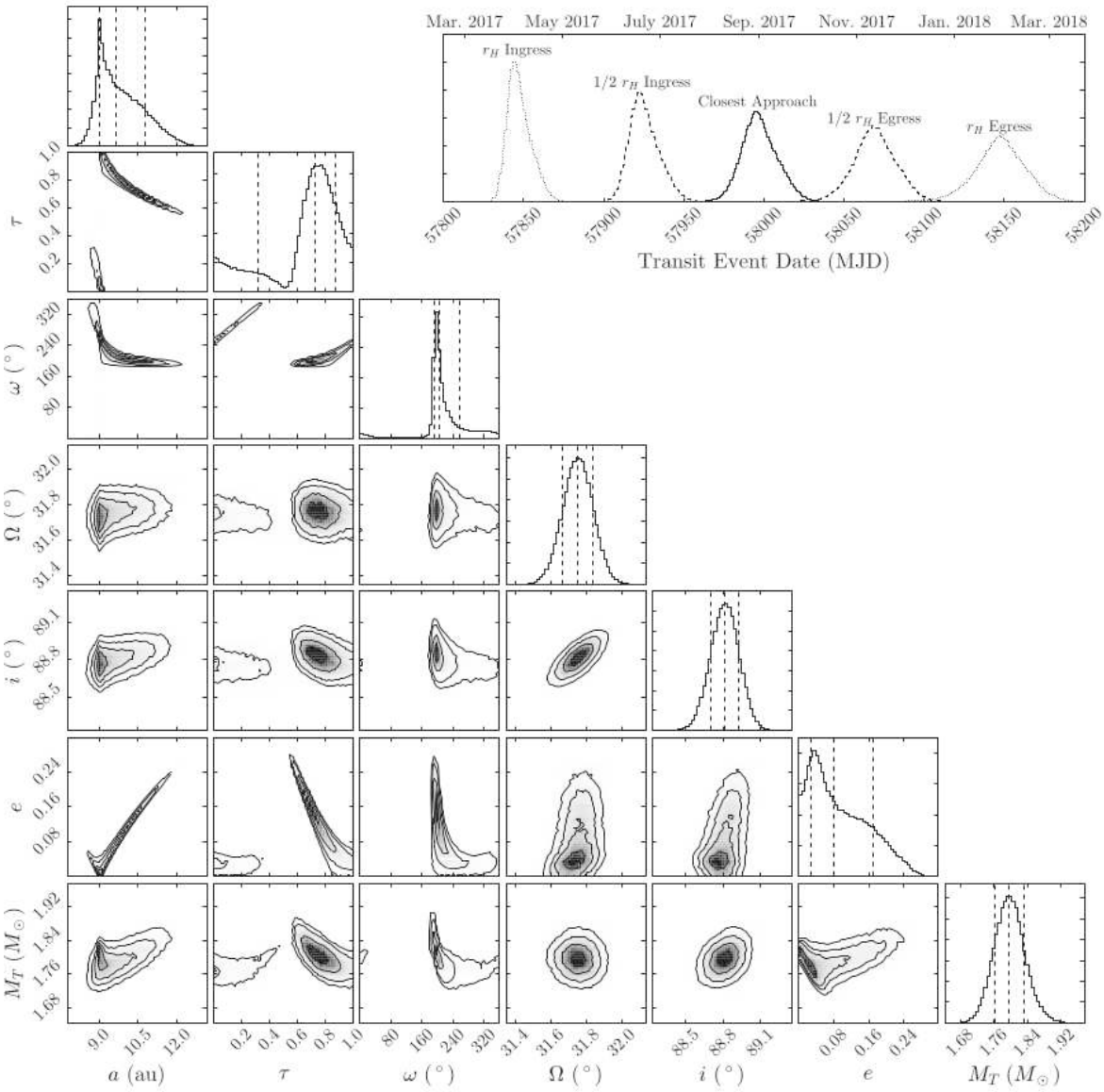


Figure 1.8: Results of a Monte-Carlo simulation showing the posterior distributions of the seven orbital parameters (including the mass of the central star), as well as the predicted dates for the transit events. Figure taken from Wang et al. [2016].

1.5 Atmosphere, physical characteristics, and formation of β Pictoris b

1.5.1 Model-dependent estimates of bulk parameters

Marley et al. [2007] and Fortney et al. [2008] showed that the two widely accepted possible formation mechanisms for giant planets ("hot start" and "cold start") should lead to different evolution scenarios during the first few $\sim$ 10 Myr of the life of said planets. This made apparent that the measurements of bulk parameters, like temperature, mass, and radius, could be used to determine by which mechanism a planet had formed. And although it was obvious that this was not an easy task, as there were still a lot of uncertainties affecting planetary evolution models, it brought considerable interest in the measurement of these parameters.

A number of studies have been performed to try to estimate some of the most important bulk parameters of the giant planet β Pic b. The general methodology behind all these studies is similar: photometric data in different wavebands and/or spectra of the planet are compared to atmospheric models, most commonly using model grids and a χ^2 based approach. Given a certain set of photometric and/or spectral data points $y_{\mathrm{data},k}, k \in \{1, \ldots, n\}$, with there associated error bars σ_k, and a model $y_{\mathrm{model},k}$ which usually depends on temperature T, surface gravity g, metallicity m, and possibly some other parameters, a χ^2 grid is calculated as following:

$$\chi^2(T_i, g_j, m_k) = \sum_{l=0}^{n} \frac{(y_{\mathrm{data},l} - y_{\mathrm{model},l})^2}{{\sigma_l}^2}$$

where the indices i, j, and k are used to represent the model grid sampling. From this, the minimum of the χ^2 can be extracted, as well as the area of 68% confidence, to give the best estimate of the parameters and their error bars.

Quanz et al. [2010] used a different approach, in which they directly compared a color magnitude (L' minus a narrow band at 4.05 μm) with evolutionary models. They reported the first estimate of the temperature of β Pic b: $T \simeq 1470\ K$. With another approach, in which they compared the color magnitude to empirical data from field brown dwarfs, they found $T \simeq 1700\ K$. Given their uncertainties, the two results were not necessarily incompatible.

Bonnefoy et al. [2011] also used a color magnitude ($\mathrm{K_S} - \mathrm{L'}$), but they compared it to atmospheric models rather than evolutionary models. They obtained $T = 1600 \pm 300$ K, but could not constrain the surface gravity. Two years later, however, improving on the number of datapoints in their analysis, Bonnefoy et al. [2013] provided the first estimate of the surface gravity on β Pic b: $\log(g/g_0) = 4.0 \pm 0.5$, with an improved temperature estimate $T = 1700 \pm 100$ K

When using atmospheric models, the surface gravity is an important parameter, as it is related to the pressure and temperature in the atmosphere. Changing the surface gravity can affect the resulting atmospheric spectrum, which also means that it can be well constrained by the model fitting. The surface gravity is given by:

$$g = \frac{M}{4\pi R^2}$$

which means that from the radius and surface gravity, it is also possible to extract an estimate of the mass of the planet.

Now, considering that an atmospheric model predicts the spectral emittance, not total luminosity, if the spectral shape can help to constrain the temperature and surface gravity,

Reference	Temperature K	$\log(g/g_0)$	Radius $R_{\rm Jup}$	Mass $M_{\rm Jup}$	Model used
Bonnefoy et al. [2011]	1600 ± 300	-	-	-	BT
	-	-	-	7 to 11	Evo. Track
Bonnefoy et al. [2013]	1700 ± 100	4.0 ± 0.5	1.4 ± 0.2	-	PHX
	-	-	-	10^{+3}_{-2}	Evo. Track
Currie et al. [2013]	1575 to 1600	3.8 ± 0.2	1.65 ± 0.06	7^{+4}_{-3}	DUSTY
Bonnefoy et al. [2014]	1650 ± 150	≤ 4.7	1.5 ± 0.2	-	PHX
	-	-	-	11.2 ± 0.3	Evo. Track
Morzinski et al. [2015]	1708 ± 23	4.2	1.45 ± 0.02	12.7 ± 0.3	Evo. Track
Baudino et al. [2015]	1550 ± 150	3.5 ± 1.0	1.76 ± 0.24	-	ExoREM
Chilcote et al. [2015]	1600 to 1700	3.5 to 4.5	-	-	BT
	-	-	-	10 to 12	Evo. Track
Chilcote et al. [2017]	1724 ± 15	4.18 ± 0.01	1.46 ± 0.01	12.9 ± 0.2	Evo. Track
	1700 to 1800	3.5 to 4.0	-	-	DUSTY, BT, PHX

BT: BT-SETTL; PHX: PHOENIX; DUSTY: AMES-DUSTY; Evo. Track: Evolutionary Track

Table 1.2: Estimates of the main parameters for β Pictoris b available in the literature.

the absolute photometry can constrain the surface of the planet, and hence its radius. Thus, it is possible to derive a mass estimate purely from atmospheric model fitting.

Following this approach, and using different datasets, Currie et al. [2013] obtained the following parameters: $T = 1575$ to 1600 K, $\log(g/g_0) = 3.8 \pm 0.2$, $R = 1.65 \pm 0.06\ R_{\rm Jup}$, and $M = 7^{+4}_{-3}\ M_{\rm Jup}$.

Bonnefoy et al. [2014], Baudino et al. [2015], Chilcote et al. [2015, 2017] all reported other estimates of the temperature, surface gravity, radius and mass of the planet. These values are compiled in Table 1.2.

1.5.2 A model-independent measurement of the mass?

All the different mass estimates reported in Table 1.2 depend either on a radius and surface gravity estimated using atmospheric modeling, or on a planetary evolution model. In either case, this makes these estimates *model-dependent*.

In the case of β Pictoris b, there is only one parameter which was estimated using a model-independent methodology: its mass. It was measured by Snellen and Brown [2018] using astrometric measurements made with the Hipparcos and Gaia missions. If we denote $M_{\rm P}$ the mass of the planet β Pic b, $M_\star$ the mass of the star, a the semi major-axis, and assuming a zero-eccentricity (circular orbit), and 90 degree inclination, the on-sky motion of the star and planet are given by:

$$\begin{cases} \alpha_\star(t) &= -\frac{1}{D}\sin({\rm PA}) \times \frac{M_\star}{M_\star+M_{\rm P}} \times a \times \cos\left(2\pi\frac{t}{P}+\phi\right) \\ \delta_\star(t) &= -\frac{1}{D}\cos({\rm PA}) \times \frac{M_\star}{M_\star+M_{\rm P}} \times a \times \cos\left(2\pi\frac{t}{P}+\phi\right) \end{cases}$$
$$\begin{cases} \alpha_{\rm P}(t) &= \frac{1}{D}\sin({\rm PA}) \times \frac{M_{\rm P}}{M_\star+M_{\rm P}} \times a \times \cos\left(2\pi\frac{t}{P}+\phi\right) \\ \delta_{\rm P}(t) &= \frac{1}{D}\cos({\rm PA}) \times \frac{M_{\rm P}}{M_\star+M_{\rm P}} \times a \times \cos\left(2\pi\frac{t}{P}+\phi\right) \end{cases}$$

Where D is the distance of the star, PA is the position angle of the planet (which does not vary since the system is seen edge-on), P the orbital period, and ϕ and unknown phase at the origin of time. To this motion, one should also add the system barycenter motion, and the parallax effect to get the true on-sky displacement as seen from the Earth. A non zero eccentricity or an inclination different than 90 degrees also need to be taken into account.

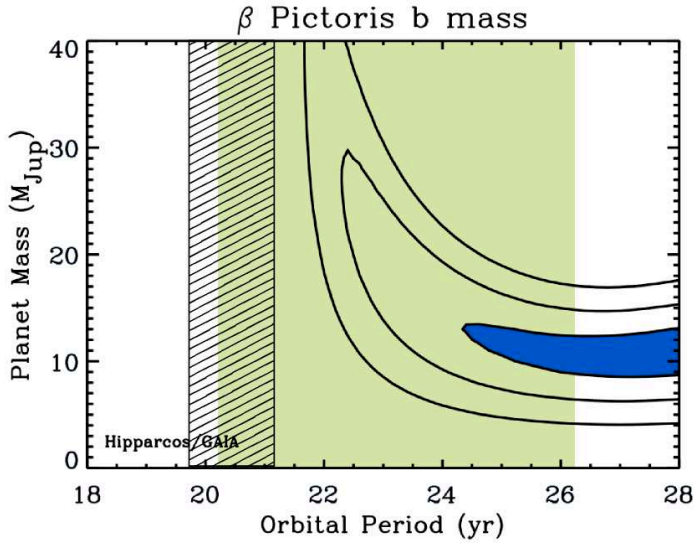


Figure 1.9: Constraints on the mass and orbital period of β Pictoris b obtained by Snellen and Brown [2018]. The lines show the 1σ, 2σ, and 3σ intervals (1σ highlighted in blue). The green area represents the constraint on orbital period from direct imaging available at the time (Wang et al. [2016]). The hatched area has been added to the original figure and represents the current best estimate from Lagrange et al. [2019], which takes into account the post-conjunction observation. Original figure from Snellen and Brown [2018].

The orbital period, stellar mass, and semi major-axis are related: $a = a(P, M_\star)$. The distance D of the star is known from the Hipparcos parallax, and the stellar mass $M_\star$ is known from other studies. Thus, the only two unknowns in the problem are the orbital period P, planet mass M_P, and phase ϕ. Snellen and Brown [2018] performed a χ^2 analysis on a mass-period grid, to obtain their best-estimate of the planet mass: $M_\mathrm{P} = 11 \pm 2\ M_\mathrm{Jup}$, in good agreement with model-based estimates.

There are two caveats to this study, though. First, looking at their Figure 3 (reproduced here in Figure 1.9), one can see that there mass estimate goes along an orbital period estimate of > 20 yr. Although this was compatible with the best estimate from direct imaging available at the time ($20.21 < P < 26.24$ at 68% confidence, from Wang et al. [2016]), it seems to be in disagreement with the current best estimate from Lagrange et al. [2019], which uses the last post-conjunction data point from SPHERE ($P = 20.29^{+0.86}_{-0.41}$). And secondly, their analysis assumes a stellar mass of $M_\star = 1.75\ M_\odot$, a value which was obtained using modelled stellar evolutionnary tracks by Crifo et al. [1997]. One can wonder how "model independant" their planet mass estimate really is, in the end...

1.5.3 Atmospheric composition

Determining the chemical composition of an exoplanet remains a challenging task, especially for directly imaged planets. High-contrast imaging instrument, such as SPHERE, on the Very Large Telescope [Beuzit et al., 2019], or the Gemini Planet Imager [Macintosh et al., 2014], only provides low-resolution spectroscopic capabilities ($R \sim 30$), in a few wavebands (J, H, K).

In an attempt to demonstrate the possibility of detecting planets with their molecular mapping technique, Hoeijmakers et al. [2018] obtained a medium resolution ($R \sim 5000$)

spectral map of the β Pic system with VLT/SINFONI. They found a strong correlation of the signal at the planet's location with the spectra of H_2O and CO, indicative of the presence of these two species in large quantities in the atmosphere of β Pic b. They also noticed that no significant correlation with CH_4 or NH_3 could be detected, suggesting that these two molecules are probably absent of the atmosphere.

Yes, the only thing we really know today about the chemistry of the atmosphere of β Pic b is that there is some water and some carbon monoxide, and probably no methane or ammonia. And β Pic b is an easy target...

Interestingly, although we know very little about the composition of the atmosphere, we do have an estimate of its rotation velocity. Measuring the rotational broadening of the absorption signal of CO with VLT/CRIRES high-resolution spectroscopy ($R \sim 10^5$), Snellen et al. [2014] estimated an equatorial spin rotation velocity of ~ 25 km/s, for a day length on β Pic b of ~ 8 hr.

1.5.4 The formation of beta Pictoris b

There are basically two competing scenarios to explain the formation of planets: core-accretion, and gravitational instability (or gravitational collapse).

In the core-accreation scenario [Lissauer and Stevenson, 2007], an initial solid core forms, and accretes gas and dust from the disk until it reaches a certain critical mass. At this point, a phase of runaway gas accretion occurs, during which the planet accretes most of the gas which will form its final atmosphere.

Gravitational collapse [Bodenheimer, 1974] is a more brutal mechanism, closer to what happens when stars form. In this scenario, an entire region of the circumstellar disk becomes unstable, and rapidly collapses to form a protoplanet, which then slowly contracts and cools down.

It was initially thought that the core-accretion mechanism, in which a large fraction of the entropy of the accreted material can be dissipated, would necessarily lead to colder planets, with lower entropy. Thus, core-accretion is sometimes refer to as the "cold-start" mechanism, as opposed to the gravitational instability "hot start" [Spiegel and Burrows, 2012].

Given this hot start/cold start distinction, it has been suggested by a few authors that β Pic b would have more likely formed through the disk instability planetary formation pathway [Bonnefoy et al., 2014, Chilcote et al., 2017]. However, the main arguments supporting this theory are the high bolometric luminosity of the planet and its high initial entropy, and recent work have shown that both could also be possible in a core-accretion scenario [Mordasini et al., 2017, Marleau et al., 2017]. Thus, the exact mechanism by which β Pic b has formed remains largely unknown.

1.6 Key questions

The β Pictoris system has quite a unique place in exoplanet science: the proximity of the star, combined with its high luminosity, its young age, and the fact that the circumstellar disk is seen edge-on, all contribute to make this system a perfect target for studying the formation of planets, the evolution of circumstellar disks, and disk-planet interactions.

Today, β Pictoris is arguably the best-known stellar system outside of our own. More than three decades of study have revealed a highly structured circumstellar disk, with rings, belts, and a giant planet. And yet, very little is known about how it came into being, and several important questions remain unanswered.

1 - How did the planet form? Where in the disk? Did it migrate? There are still conflicting theories to explain planet formation, especially in the case of massive planets. Core-accretion and disk instability are two possible scenarios which can explain the formation of β Pic b, and there is still no definitive argument in favor of one or the other. We also know that different mechanisms can change the orbit of a giant planet, but is is still unclear how frequent these migrations are, and what consequences they can have on the rest of the system. Today, we have no idea whether the giant planet β Pic b formed at its current location, at $\sim$ 10 AU from its host star, or if it formed elsewhere in the protoplanetary disk and later migrated to its position.

2 - What is the chemical composition of the planet? How can we link this composition to the formation history? A new body of research is currently emerging in exoplanet science, and several authors now aim at linking giant planet and brown dwarf formation processes to observable quantities like elemental ratios (e.g. Madhusudhan et al. [2014], Mordasini et al. [2016], Mollière and Snellen [2019]). For example, it seems that the carbon-to-oxygen number ratio (C/O) can hold crucial information about the fractional content of solid and gaseous material accreted by the planet/brown dwarf. Beta Pictoris is certainly an attractive target to test this idea, but some progress is required to first determine its chemical composition.

3 - How did the disk around β Pic evolved? What was the original disk like? If the current composition of the giant planet β Pic b is poorly known, the disk is a completely different story. Numerous chemical species have been found, and our understanding of the creation and destruction of these species is improving steadily. But it seems clear today that the disk around β Pic which we currently observe is quite different from the original protoplanetary disk. But if we want to understand the formation of planets, we also need to understand the evolution of the disk in which they form.

4 - Is β Pic b a transiting planet? If not, what was the 1981 event? Is there anything in the Hill sphere of the planet? The orbit of the giant planet β Pic b has an inclination very close to 90 degrees. Although the most recent orbit determination seems to exclude a transit of the planet, it also seems that the planet Hill sphere is transiting. If this is the case, the question remains to know whether or not the photometric variations observed in 1981 are related to such an event. And if there is something to be seen in the Hill sphere, what is it? And what can it tell us about the planet itself?

Part II

LIFE AND DEATH OF PICSAT

Chapter 2

Mission architecture

Content of this chapter

2.1 Half of an interferometer

Even for a small scale mission, assembling a team and getting the funds to design, develop, build, and launch a satellite is not easy, and a solid science case is paramount to the success of such an endeavor.

The science case for PicSat emerged during the workshop "Thirty years of β Pic and debris disks studies", held in Paris, from September, 8, 2014 to September 12, 2014. During this workshop, the astronomical community acknowledged the strong science interest of catching the upcoming β Pic b transit. But high-precision photometry from the ground is notoriously difficult to achieve, and it was very appealing to have a small telescope in space entirely dedicated to the monitoring of β Pic.

At that time, the idea of developing a small satellite to demonstrate the feasibility of space-based fiber-fed optical interferometry was also in the air at the Observatoire de Paris [Lacour et al., 2014]. The FIRST-S mission had been deemed too risky and too difficult for a first CubeSat mission, and it then appeared that by removing the interferometric combination stage, and leaving only a single telescope, the FIRST-S interferometer could be transformed in a fiber-fed photometer, which would make for a perfect first technological demonstrator, and a perfect β Pic monitoring mission.

Yes, PicSat was born on the premise that half of an interferometer can be a photometer.

2.2 Science objectives

2.2.1 The transit of beta Pictoris b

The main science objective of PicSat was the detection and characterization of the transit of β Pic b. When the development of the mission started, the exact date and time of transit was still largely unknown [Lecavelier des Etangs and Vidal-Madjar, 2016]. It was not even clear whether or not the planet itself would transit in front of its host star, and there was a possibility of a transit of the Hill sphere only [Lecavelier des Etangs and Vidal-Madjar, 2016, Wang et al., 2016]. The orbital inclination of β Pic b is now estimated to be $i = 89.04 \pm 0.03$ deg (see Chapter 13), and we know today that the planet itself did not transit.

A detection of a transit of the giant planet would have led to a model-independent measurement of its radius, which is of significant interest to constrain the formation mechanism [Baraffe et al., 2003]. But even if only the Hill sphere was actually transiting, this would have been a beautiful opportunity to study the close-by environment of β Pic b. Kenworthy and Mamajek [2015] had already shown with their observations of J1407B that it was possible to infer information about a potential ring system from a transit event (see Figure 2.1).

Model-dependent estimates of the radius of β Pic b had already been derived from photometric observations of the planet. The radius was thought to be $\sim 1.5\ R_{\rm Jup}$ (see Table 1.2 in Part I). The angular diameter was known from measurements made by Defrère et al. [2012]: $\theta_\star = 0.736 \times 10^{-3}$ arcsec, and the distance of β Pic was known from Hipparcos data: $d_\star \sim 20$ pc. From this, it was possible to calculate the expected photometric variation for the transit, given by the surface ratio between the planet and the star:

$$\frac{\Delta F}{F} = \left(\frac{R_{\rm planet}}{\frac{1}{2}\, d_\star \theta_\star} \right)^2 \simeq 1\% \tag{2.1}$$

Since β Pic b orbits at $a \sim 10$ pc from its host star, with an orbital period of ~ 20 yr, the transit duration could be calculated using:

$$\tau = \frac{P}{2\pi a} \times \theta_\star\, d_\star \sim 40 \text{ hr} \tag{2.2}$$

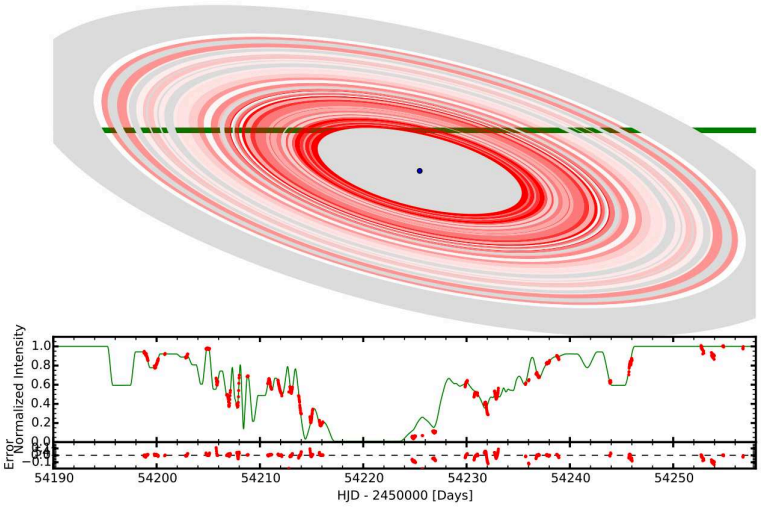


Figure 2.1: Nested ring model (top panel) used to interpret the photometric data on J1407B (bottom panel). On the top panel, the green line represents the path of the star behind the ring system. The intensity of the red color is proportional to the opacity of the ring. The gray area corresponds to unconstrained parts of the system. On the bottom panel, the red points are the measured data points, and the line is the model fit. The ring system is thought to occupy the Hill sphere of an unseen companion. Figure taken from Kenworthy and Mamajek [2015].

In order to ensure a detection at the 5σ confidence level, the instrumental photometric precision required is at least $0.01 \times 5^{-1} \times \sqrt{40} = 1\%/\text{hr}$. If we want to be able to measure the diameter of the planet at a precision of 1% (at $1\,\sigma$), then the required photometric precision needs to be at least $0.01 \times 1\% \times \sqrt{40} = 630$ ppm/hr (part per million per hour).

2.2.2 Exocomets in the beta Pictoris system

A second objective completed the astrophysical science case for the PicSat mission: exocomets and debris disk science. Kiefer et al. [2014] detected and studied exocomets using transit spectroscopy. They detected exocomets by looking for variations in the calcium lines, which means they detected the gas tails of the comets. An objective of PicSat was to detect exocomets in the broad visible band, in order to observe their dust tails. Lecavelier Des Etangs et al. [1999] modeled the photometric variations which could be expected from exocometary transits, and found typical transit depths of a few 10^{-4}, with durations of a few hours (see Figure 2.2). To be able to detect such events, a photometric precision of 100 ppm/hr was required for the mission.

The photometric precision requirements for exocomet detection was more stringent than the requirement for the giant planet transit detection. But since exocomets are transiting on a regular basis in the β Pic system, ensuring that PicSat could detect them was a way to guarantee a science return even in the case of a no-transit scenario.

2.2.3 Single-mode photometry from space

In addition to its science objectives, PicSat also had a technical objective: demonstrate the possibility of injecting starlight into a single-mode fiber, from a CubeSat platform. Single-mode injection is a key technology for many applications, from optical communication to

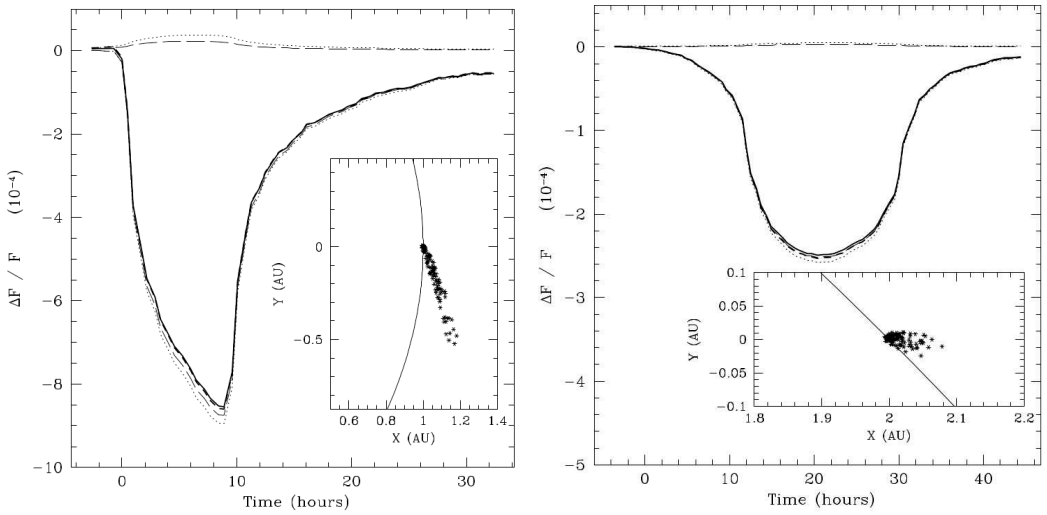


Figure 2.2: Models of two possible transits of an exocomet. The solid line gives the photometric variation in red light (800 nm), and the short dashed line gives the variations in blue light (400 nm). Figure taken from Lecavelier Des Etangs et al. [1999].

nulling interferometry, but it is notoriously difficult, as it puts serious constraints on the pointing stability of the instrument.

To enable further developments toward space-based interferometry, PicSat should demonstrate the possibility of achieving an on-target coupling ratio of $> 10\%$ with a stability of at least 1%/hr, and maintain it for the entire orbit duration.

2.3 Satellite and orbit

2.3.1 Science payload

The scientific payload of PicSat was designed to achieve stable single-mode injection and precision photometry on a bright star. The design, development, and tests of this payload are discussed in details in subsequent chapters. A general overview of the system is given in Figure 2.3.

2.3.2 CubeSat platform

PicSat was developed as a three-unit CubeSat (3U CubeSat). The total satellite footprint is 30 cm $\times$ 10 cm $\times$ 10 cm. In this reduced space, the platform hosts all the necessary subsystems: power generation and power distribution, on-board computer, radio-communication system, attitude control. A detailed view of the platform is given in Figure 2.4, along with a picture of the satellite flight-model.

Satellite bus The mechanical 3U structure has a standard form factor, and comes from Innovative Solutions in Space[1] (ISIS). Their standard product was slightly modified to accommodate the PicSat payload, and especially the bulky two-axis piezo stage. The communication system is the ISIS TRxVU transceiver. Space-to-ground communication (downlink) occurs at 435.25 MHz, and ground-to-space (uplink) at 145.910 MHz. A cooperation with the amateur

[1] www.isispace.nl

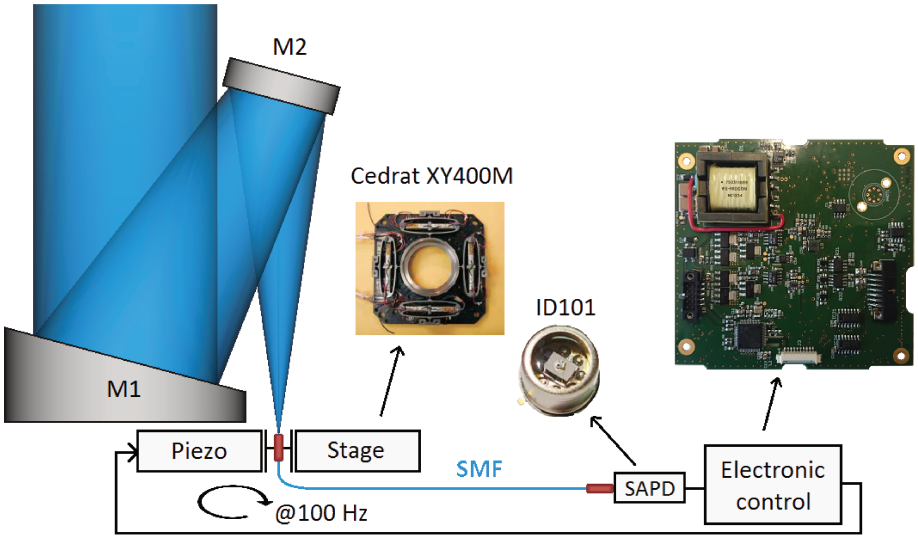


Figure 2.3: Overview of the principle of the PicSat payload. A small optical telescope collects the starlight, which is injected into a single-mode fiber, and brought to a Single Photon Avalanche Diode (SPAD). The head of the optical fiber is mounted on a two-axis piezo actuator, which is used to lock the fiber position on the star image in the focal plane. The entire payload is controlled by a small dedicated electronic board.

radio community allowed us to use these frequencies (see Chapter 8). The Electrical Power System (EPS) is the GomSpace[2] P31U, with a GomSpace BP4 battery pack. The total capacity is 37.4 Wh, and the energy is provided by 32 solar cells, spread over 4 fixed panels, and 2 deployable panels. The on-board computer is the ISIS OBC, based on an ARM9 processor. Two 32 GB SD cards are used to store science data.

Flight software The flight software for PicSat has been developed in-house at the Observatoire de Paris. The general architecture is based on the GERICOS framework [Plasson et al., 2016]. This framework implements an active objects design pattern, in which different high-level tasks correspond to different "active" objects. The processor time is shared between these objects by a Real Time Operating System (RTOS). The framework, initially developed for LEON architectures, was ported to the AT91SAM9G20 processor (ISIS on-board computer) and STM32F303 microcontroller (payload electronic board). In the case of PicSat, GERICOS uses FreeRTOS, but other RTOS are available. The flight software is further splitted into two levels, as allowed by the framework. Level-Zero (L0) contains the most critical parts of the software (management of the electrical power system, memory, communications, and detumbling). Level-One (L1) contains high-level non-critical tasks (target pointing, payload management, SD Card management, etc.). The L0 part of the software is written once and for all on the processor before the launch, while the L1 part can be updated remotely during the mission.

Attitude determination and control system (ADCS) Single-mode injection comes with serious constraints on pointing precision and stability. In order for the fine pointing system of the payload to work properly, the main ADCS needs to provide a pointing precision < 1 arcmin. The ADCS selected for the PicSat mission was the iADCS100 from Hyperion

[2]https://gomspace.com

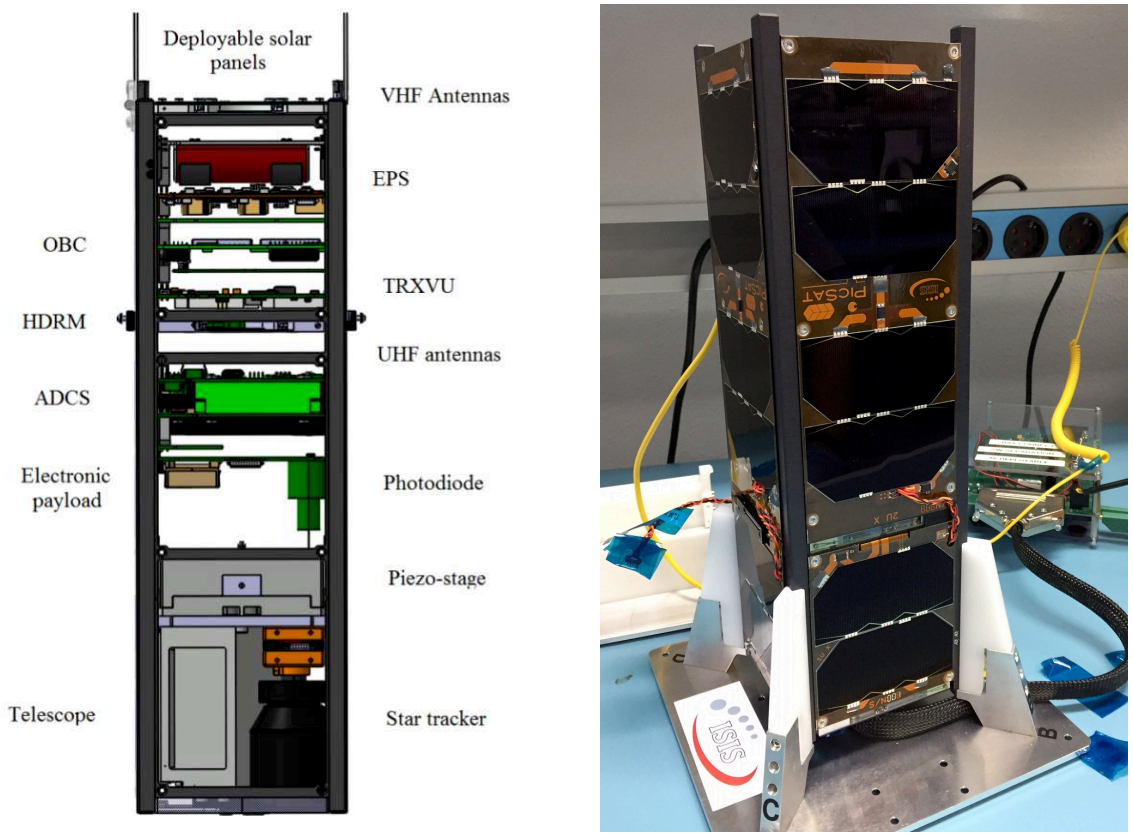


Figure 2.4: Left panel: a detailed overview of the layout of the 3U platform. EPS: Electrical Power System; OBC: On-Board Computer: TRxVU: communication system; HDRM: data-handling; ADCS: Attitude Control System. Right panel: A picture of the flight-model.

Technologies[3], featuring their ST200 star-tracker, and RW200 reaction wheels. A modification to the base system was made for PicSat, in which the star tracker was deported next to the optical telescope, and roughly aligned on the same optical axis. This system did not deliver the expected performance (see Chapter 9).

2.3.3 Orbit

CubeSats are not typically the main payload of a rocket. They are piggybacked to orbit. Consequently, CubeSat mission planners have limited freedom on the choice of the orbit. PicSat was designed to accommodate many different orbits, ranging form Sun-Synchronous Orbits (SSO), which are a kind of polar retrograde orbit widely used by Earth observing satellites, to equatorial orbits. It was finally launched on a SSO, at 505 km altitude. The two-line elements[4] for PicSat at the beginning of the mission were:

```
PICSAT
1 43131U 18004W   18021.60551509  .00001940  00000-0  87505-4 0  9991
2 43131  97.5546  83.7702 0008572 247.1439 112.8892 15.22227881  1414
```

[3] www.hyperiontechnologies.nl

[4] Two-line elements (or TLEs, for short), are a standard way of specifying orbital elements broadly used in the space industry. For a complete description of the format, please refer to: https://www.celestrak.com/NORAD/documentation/tle-fmt.php

The orbital period was $\sim$ 90 min, with a maximum Sun-eclipse duration of $\sim$ 20 min, and a similar target eclipse duration. Figure 2.5 gives an overview of the satellite orbit around the Earth.

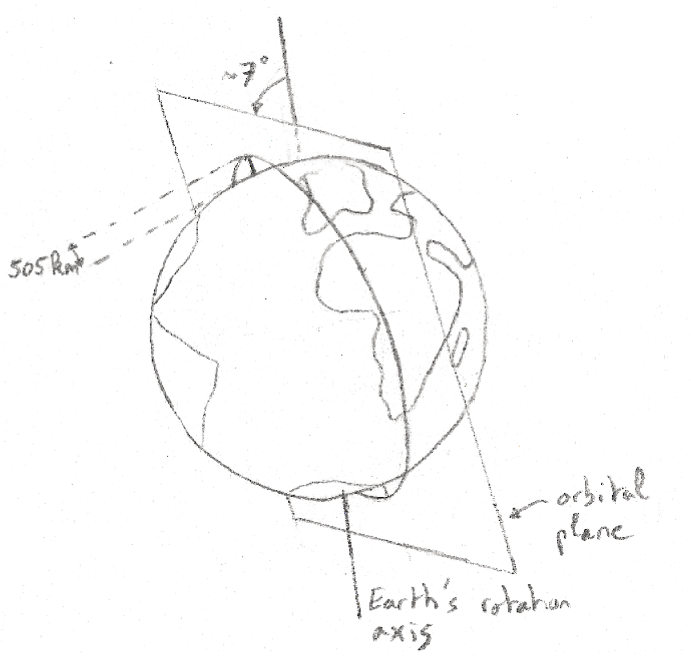


Figure 2.5: Overview of the orbit of PicSat around the Earth. The orbit is a Sun-Synchronous polar orbit, at 505 km altitude.

2.4 Budgets

2.4.1 Mass budget

A simplified mass budget of PicSat is given in Table 2.1. The total mass of the satellite is $\simeq$ 3.5 kg, largely dominated by the power subsystem ($\simeq$ 1.1 kg, in which 800 g accounts for the solar panels), and the payload ($\simeq$ 1.3 kg, in which 745 g accounts for the piezo actuator).

Subsystem	Mass	Uncertainty	Nominal	Min	Max
	g	%	g	g	g
Structure	226	2	226	221	231
Communication	223	2	223	218	228
Data handling	162	2	162	159	165
Power	1141	5	1141	1084	1198
ADCS	469	10	469	422	516
Payload	1333	5	1333	1266	1400
Total			3554	3370	3738

Table 2.1: Simplified mass budget for PicSat.

2.4.2 Power budget

The power budget is given in Table 2.2. The total power consumption for the satellite, including all margins, is 5730 mW. The 16 units of solar panel (4×3 on the sides of the satellite, and 2×2 deployables) are enough to guarantee a power generation of 5860 mW on the worst orbit, which leaves a positive balance of 130 mW.

Subsystem		Power mW	Margin %	Power+Margin mW	Duty cycle %	Consumption mW
Power System		210	2	214	100	214
On-board computer		380	2	388	100	388
Com.	Reception	560	5	588	100	588
	Transm.	3000	5	3150	5	158
Payload	Science	2040	5	2142	65	1392
	Standby	150	5	157	35	55
ADCS	Pointing	1800	10	1980	100	1980
	Standby	200	10	220	0	0
Consumption						4775
Margin						10%
EPS efficiency						0.9
Total consumption						5730

Table 2.2: Power budget for the PicSat mission.

2.4.3 Data budget

The data budget for PicSat is given in Table 2.3. Considering that the raw science data are generated at 1 kHz, the total amount of data generated on the satellite (450 MB/day) far exceed the capability of the ground segment (2.2 MB/day).

This problem is solved by reducing the science data on-board the satellite. Only the low-resolution science product is sent to the ground, which reduces the volume of the science data in normal operations to 127 kB/day. All raw science data is also copied on the two 32 GB SD cards, with full redundancy (if one card fails, no data is lost). The two SD cards allow the satellite to store 9 weeks of continuous raw science data. Some of this data can be transferred to the ground station to verify the performance of the tracking algorithm and/or the performance of the on-board data reduction software. In case of the detection of an anomaly in the light-curve of the target star, part of this data can also be downloaded in raw format, or reduced to a higher resolution, to improve the science data quality.

2.5 Schedule and costs

The idea of building a small CubeSat to monitor β Pictoris and study the transit of the giant planet β Pictoris b emerged in 2014, and activity on the project started in early 2015, less than 3 years before the expected time of transit. The overall mission development schedule was constrained by the need to have the satellite in orbit as soon as possible (mid 2017 best case, early 2018 worst case). The general timeline is given in Figure 2.6, the cumulative budget in Figure 2.7, and the cumulative number of FTEs (Full-Time Equivalents) in Figure 2.8.

Satellite transmission					
Type	Rate b/s	Margin %	Rate+Margin b/s	Duty cycle %	Volume kB/day
Science low-res	16	10	17.3	67	127
Payload monitoring	6	10	6.6	100	71
Satellite monitoring	32	10	35.2	100	380
Total					578
Ground station reception					
	Data rate		9600 b/s		
	Visibility per day		36 min		
	Error rate		5%		
	Link availability		90%		
	Volume		2216 kB/day		
On-board memory					
Type	Rate kb/s	Margin %	Rate+Margin kb/s	Duty cycle %	Volume MB/day
Science raw	56	10	61.6	67	446
Monitoring raw	0.40	10	0.44	100	4.7
Total					450.7
			9 weeks total		28.4 GB
			Redundancy		28.4 GB
			Available memory		64 GB
			Margin		7.2 GB

Table 2.3: Data budget for PicSat.

In summer 2015, a first prototype of the payload was built and tested, to validate the concept of doing photometry with a single-mode fiber. Early environmental tests on the payload were performed in December 2015, and revealed a problem with the piezo actuator, leading to a new and slightly bigger design of the payload. This new design was quickly assembled to be tested again in vibration and thermal vacuum, in order to validate it, and to move forward with the assembly of the complete engineering model. In October 2016, about a year after the project had started, we passed a Critical Design Review, and two months later, we performed a full engineering model vibration test. This validated the final design of PicSat, and we moved on to the assembly of the flight model. We also started to work on the ground segment (hardware and software) in early 2017. A Mission Readiness Review was passed in mid 2017. At the time, the mission was not ready to fly, but the review helped to clarify where to put the most effort in order to have a satellite ready to be delivered in December 2017. The satellite was successfully delivered on December, 8, 2017.

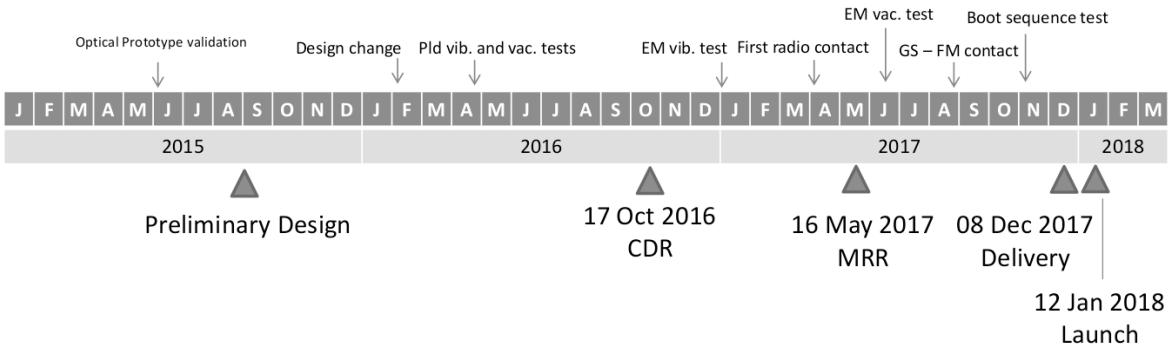


Figure 2.6: Timeline of the development of the PicSat project.

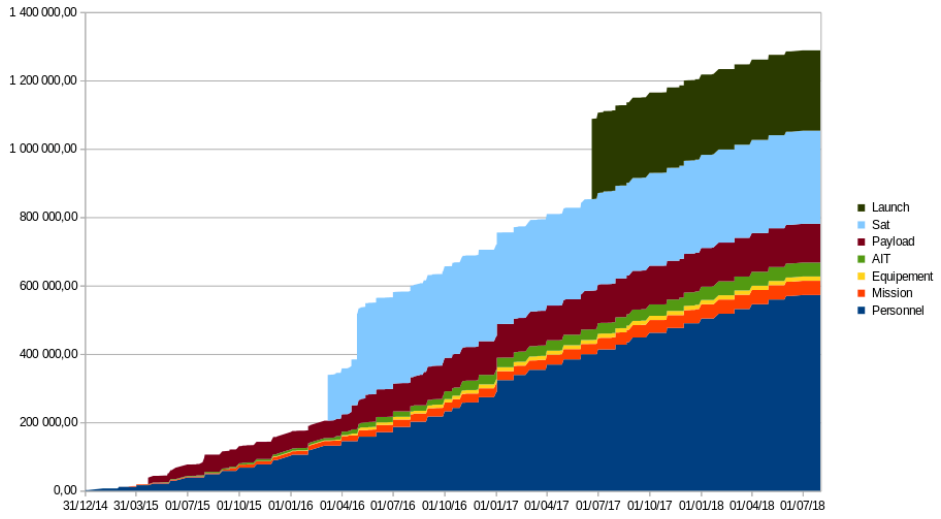


Figure 2.7: Cumulative budget as a function of time for the PicSat project, in euros. The budget is separated in salaries (blue), travels (orange), lab equipment (yellow), AIT activities (green), payload hardware (red), satellite hardware (light blue), launch (dark green).

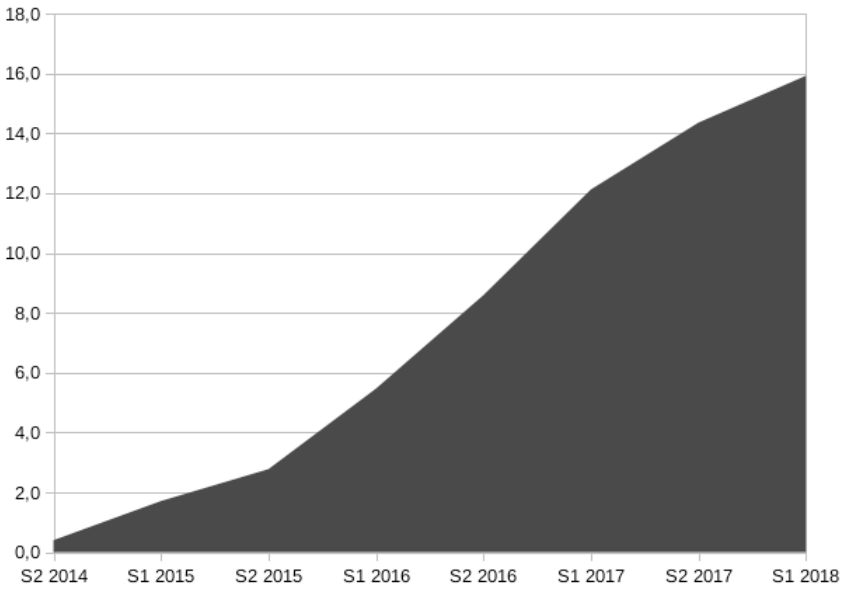


Figure 2.8: Cumulative number of FTEs (Full-Time Equivalent) as a function of time for the PicSat project.

Chapter 3

Payload overview

Content of this chapter

3.1 Concept

To answer its science and technical objectives, the PicSat mission required an instrument capable of simultaneously demonstrating the use of a single-mode fiber in space, and achieving high-precision photometry. This lead to the design of a "fibered photometer", in which, contrary to all usual photometers, the 2 dimensional detector array is replaced by a Single Photon Avalanche Diode (SPAD).

The light coming from the star is collected by a small optical telescope (the effective diameter is 3.72 cm), injected into a Single Mode Fiber (SMF) placed in the focal plane of the telescope, and brought to the SPAD for photon counting. The instrument operates between 400 and 700 nm, and the fiber has a core diameter of only ~ 3 μm, similar to the diameter of the star image on the focal plane of the instrument (Point-Spread Function, PSF). Thus, to properly inject light into this fiber, a specific tracking mechanism is required, to ensure that the small fiber stays centered on the star, even in the presence of platform pointing errors and/or jitter. This is done using a two-axis piezoelectric actuator on which the head of the fiber is mounted. This actuator can move the fiber in the focal plane, over a 450 μm $\times$ 450 μm area (about $10' \times 10'$ in terms of field of view). The ADCS performance level (30 as) ensures that the satellite is able to point β Pictoris with sufficient precision so that its image always falls into the accessible piezo range. The piezo actuator can then move the fiber to "scan" the entire field, find the star, and track it.

To reach a level of precision of 100 ppm/hr on the photometry of the star, an excellent tracking is not enough. Some variations of the PSF of the instrument are expected (mainly due to thermal stress of the optics), and will result in instrumental variations of the photometry. To correct these, it is necessary that the tracking algorithm also regularly samples the instrumental PSF at different positions, to be able to estimate some of its most critical parameters (mainly its size in two orthogonal directions). To do so, the piezo actuator will constantly modulate the position of the fiber around the central position of the star.

All payload activities are controlled and managed by a dedicated electronic board, which embeds a 72 MHz STM32F303 microchip, and the overall concept is illustrated in Figure 2.3 of Chapter 2

3.2 Opto-mechanics and reference frame

3.2.1 Optical setup

For proper injection into the fiber, the aperture ratio of the telescope must match the one of the fiber. Thus, the aperture ratio of the PicSat telescope is constrained to an F/D value of 4. The telescope must also fit into a single CubeSat unit. The payload is thus based on a compact 30 degree off-axis Newtonian design, with an effective diameter of 3.72 cm, and a focal length of 14.84 cm. The primary off-axis parabola used is oversized (50 mm diameter) to ensure optimum optical quality on the edges, and made of pure aluminum. The optical tube and the baseplate on which the instrument is mounted are also made in aluminum. This ensures homogeneous thermal dilation, making the position of the focal plane of the telescope independent of temperature. The secondary mirror is a plane mirror, of 22 mm diameter. A mechanical drawing of the primary mirror, containing the exact dimensions, and a rendering of the optical path in the telescope are given in Figure 3.1.

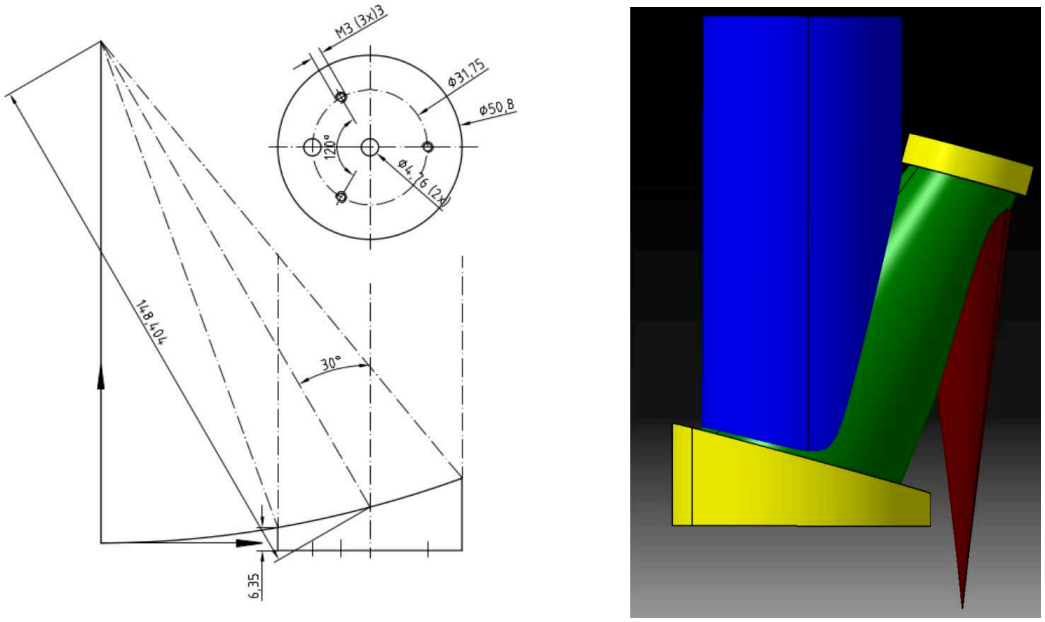


Figure 3.1: Left panel: mechanical drawing of the primary parabola. Right panel: optical path in the complete telescope. The telescope is an off-axis Newtonian design, with a 50 mm diameter primary parabola (bottom left), and a 22 mm diameter plane folding secondary (upper right).

3.2.2 Mechanical assembly and reference frame

The payload assembly is designed to fit into a single CubeSat unit. The primary parabola is directly mounted on a baseplate, which comes on top of the piezo stage. The optical tube encircles the parabola, and supports the secondary mirror assembly. This mechanical assembly, integrated into a standard CubeSat unit is shown in Figure 3.2. The payload frame of reference is superimposed to this image. The Z axis is parallel to the optical axis, and points out of the telescope. The whole structure is made out of aluminum, so that thermal variations are expected to produce homogeneous deformations, minimizing their impact on optical quality of the system[1].

One of the most critical part is the secondary mirror assembly, which is presented in Figure 3.3. The M2 mirror is glued on a support element, (Figure 3.3d), which is itself interfaced on the head block (Figure 3.3b) with three spacers. The exact height of these spacers can be adjusted to align the telescope (tip, tilt, and focus). The head block is fixed on top of the optical tube.

3.2.3 Piezoelectric stage

The fiber is mounted on the two-axis piezo actuator, and positioned in the focal plane of the telescope. The piezo actuator is based on space qualified components, and made by CEDRAT Technologies[2]. The two axes are controlled independently, with two different voltages gener-

[1] *A posteriori*, this design choice is... debatable, to say the least. It is true that an optical telescope fully made of aluminum should be less sensitive to thermal variations, but only in the case were the instrument is not subject to any thermal gradients, which is hardly the case in space. A better option may have been to select materials less sensitive to thermal deformations. See Chapter 6.

[2] https://www.cedrat-technologies.com

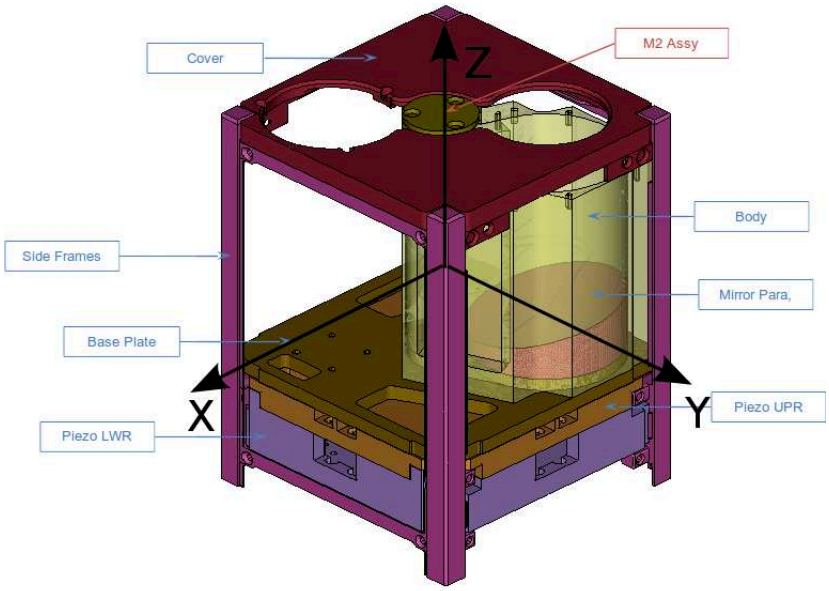
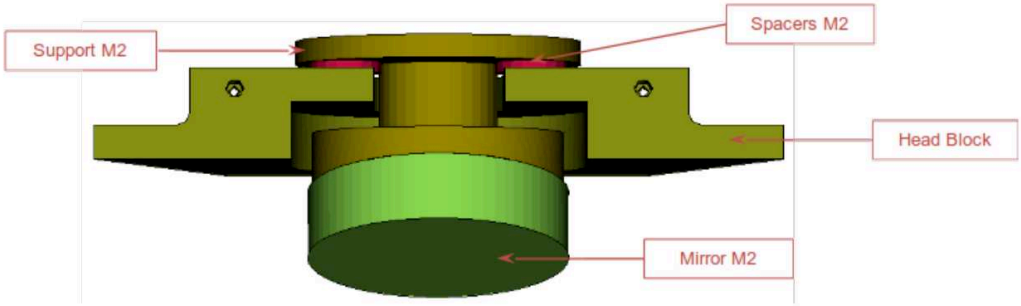
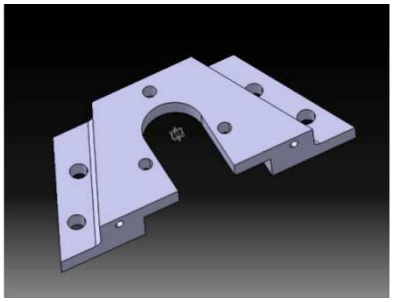


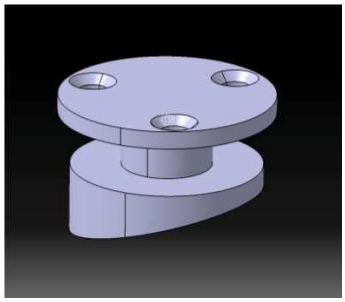
Figure 3.2: A 3D view of the payload integrated into its CubeSat unit, with the payload reference frame.



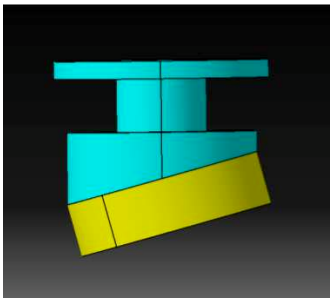
(a) Complete assembly



(b) Head Block



(c) Support



(d) Mirror glued on support

Figure 3.3: Mechanical assembly supporting the secondary mirror.

ated via two dedicated DACs on the payload board. The two axes of the actuator are also equipped with strain gauges, to measure the position of the actuator. The total range of the piezo is ~ 450 μm on both axes.

3.2.4 Fiber

The fiber selected for the PicSat mission is the S405-XP, from Nufern[3]. These fibers have pure silica cores, which make them resistant to the radiation environment of Low-Earth Orbit [Alam et al., 2017]. The overall characteristics of this fiber are compiled in Table 3.1.

Operatng wavelength	400-680 nm
Core NA	0.120
Mode-field diameter	3.3 ± 0.5 μm at 405 nm 4.6 ± 0.5 μm at 630 nm
Cutoff	380 ± 20 nm
Core attenuation	≤ 30 dB/km at 488 nm ≤ 30 dB/km at 630 nm

Table 3.1: Main characteristics of the Nufern S405-XP single-mode fiber used in the PicSat mission.

3.3 Electronics and detector

3.3.1 Architecture

The entire payload is controlled by a dedicated electronic board. The general architecture of this electronic board is presented in Figure 3.4. All the electronics are controlled by an STM32F3 microchip, running at 72 MHz. The board has two power-lines: a 3.3 V line feeding the microcontroller and several sensors, and a 5 V line feeding the photodiode, temperature control module, and the piezo stage (after up-conversion to 150 V). Several sensors are available on the board, to monitor voltages, amperages and temperatures. The board is connected to the main computer of the satellite via a 400 kB/s UART-RS422 line.

3.3.2 Avalanche diode

The detector used for the PicSat mission is the Single Photon Avalanche Diode (SPAD) ID101, made by IDQuantique[4]. The detector operates in the visible band (350 nm to 900 nm), with a wavelength response curve given in Figure 3.5.

The SPAD is powered by a regulated 5 V line. The bias voltage is created by a regulated -25 V line. The output signal is linked to a hardware 2 bits binary counter, whose output goes to one of the hardware counter of the microcontroller. The integration cycle is entirely managed at the level of this binary counter, which can be activated/deactivated using a GPIO of the microcontroller. A gate signal is created on the GPIO using a dedicated hardware timer. This regulates the integration time. A delay is then introduced before the next integration gate occurs, to regulate the total integration cycle. At the end of each integration, during the cycle delay time, the counter value is read, and the coutner is reset for the next integration gate. Figure 3.6 gives a diagram illustrating the acquisition sequence.

[3] https://www.nufern.com/
[4] https://www.idquantique.com

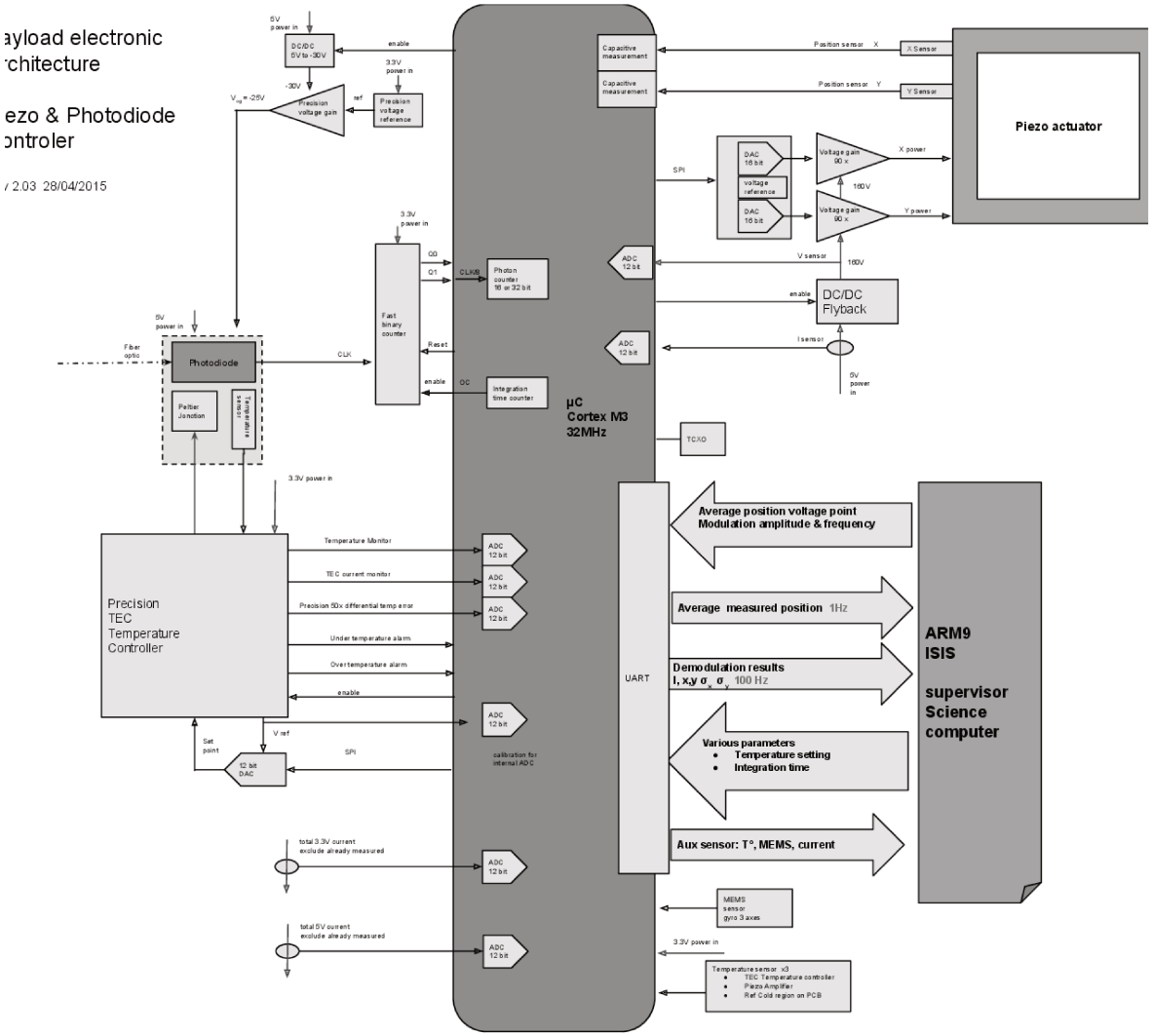


Figure 3.4: General architecture of the PicSat payload board.

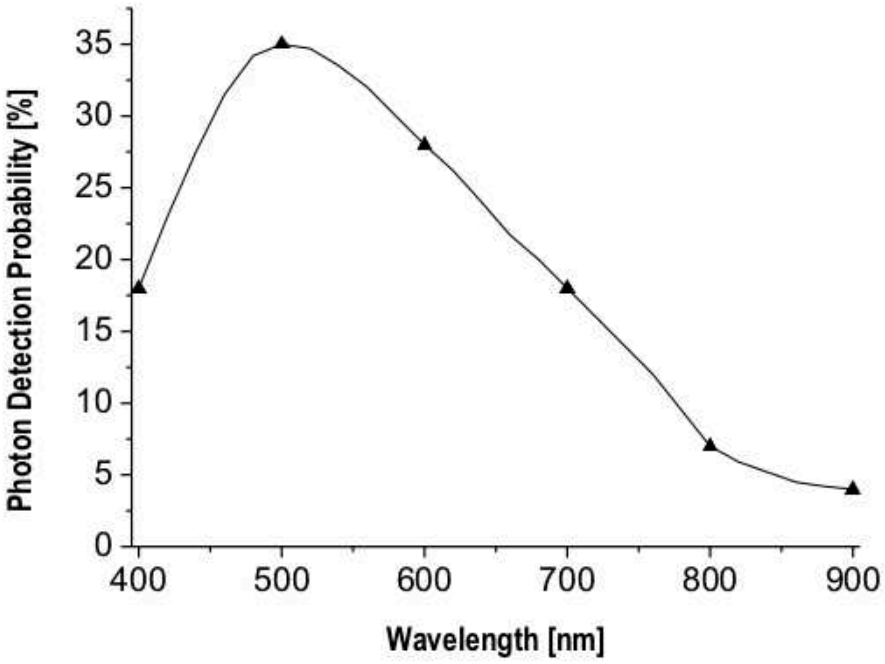


Figure 3.5: Photon detection probability as a function of wavelength for the ID101 Single Photon Avalanche Diode used in PicSat.

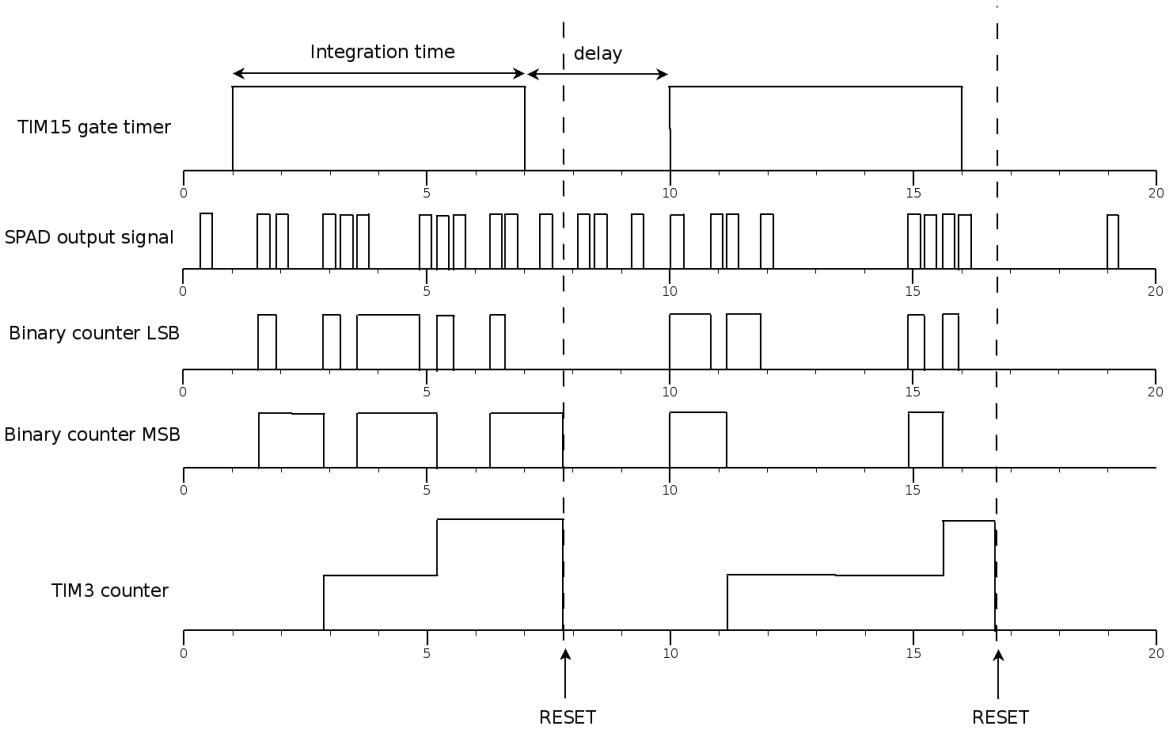


Figure 3.6: Diagram illustrating the acquisition sequence of the photodiode

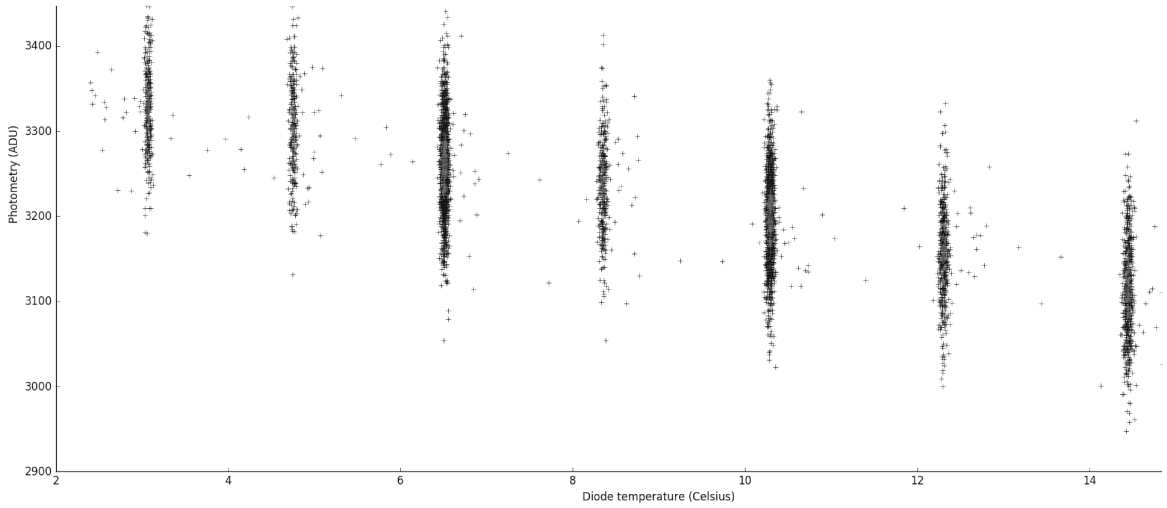


Figure 3.7: Evolution of the photometry measured on a regulated source with the temperature of the diode

3.3.3 TEC Controller

The SPAD used in the PicSat payload has a built-in thermo-electric device (TEC) which can be used to regulate the temperature of the single pixel. The regulation is managed by a dedicated TEC controller module, part of the payload board. The temperature setpoint can be selected using a single 15-bit DAC. The effect of the temperature of the pixel on the photometry measured by the diode has been calibrated during a thermal vacuum test campaign, and the result is given in Figure 3.7.

3.4 Functional description

3.4.1 Payload modes

The payload has different modes, in which its subsystems are activated/deactivated as necessary, to minimize power consumption. From a user perspective, only a subset of these modes are accessible, and all other modes are internal to the software. The user-accessible modes (also called "main modes") are: Standby (SBY), Idle (IDL), Slave (SLV), Science (SCI), and Busy (BSY). Within the payload software, the SCI mode is sub-divided into 3 different "acquisition modes": Imaging (IMG), Searching (SCH), and Tracking (TRK). The idea behind this subdivision is that the user only needs to trigger a transition to the SCI mode. The whole acquisition sequence (finding the star in the focal plane, acquiring it, tracking it) is taken care of by the payload itself.

SBY – Standby mode

A low power consumption mode, in which the only action of the payload is to gather and transmit house-keeping (HK) data to the OBC and beacons to the ground, if requested. The payload board is powered-up, and HK data are automatically acquired and sent to the OBC on a regular basis, if requested.

IDL – Idle

Higher consumption mode, in which the SPAD and the piezo are active but not running. The piezo is at a corner position, and the photodiode timer is never started. HK data are automatically acquired and sent to the OBC, if requested. The photodiode temperature is regulated in this mode.

IMG – Imaging

High power consumption, high data rate mode. In this mode, the fiber scans the focal plane to create a 2D image of the field of view and look for the star. The photodiode is used to get a measurement of the flux at different positions in the focal plane. The piezo sensors are used to monitor the position of the fiber. The photodiode is regulated in temperature. HK data are still acquired and transmitted to the OBC, if requested, along with photometric data and piezo positions.

SCH – Searching

High power consumption mode in which the payload is looking for the star around a particular position of the focal plane. This mode is used when the star has just been lost, to allow for a fast recovery, and also as a transition between IMG and TRK, during intial acquisition.

TRK – Tracking

High power consumption, high data rate mode, which constitutes the main science mode. The fiber is locked on the star, and science data are continuously acquired and transmitted to the OBC. Different algorithms can be used to track the star. HK data are acquired and transmitted to the OBC, if requested. The SPAD temperature is regulated.

SLV – Slave mode

Mainly intended as a test mode. The payload is passive and does not do anything on its own. The photodiode and the piezo are active but not running. The board waits for specific instructions from the OBC. By default, the photodiode temperature is regulated in this mode.

BSY – Busy

Mainly intended as a test mode, complimentary to the SLV mode. The payload will enter this mode from SLV when a long duration operation is requested by the user (for example: “get a 2D image of the focal plane”). The payload can still receive telecommands from the OBC, and reply to short duration commands (“get HK”, for example), but it will not accept any other long duration command until the current one if finished.

3.4.2 House-keeping data

A number of sensors are present on the electronic board, and can collect general data (voltages, amperages, temperatures, etc.), used to check the health of the system. These are called House-Keeping (HK) data (and HK sensors).

From a hardware point of view, most of these sensors use one of the numerous Analog-to-Digital Converters (ADCs) of the STM32 microchip. Notable exceptions are: the three temperature sensors (whose values are readily available on the I2C bus from the microcontroller), and the 3-axis gyroscope (also on the I2C bus). These ADCs are used in combination

Data	Sensor	Accessible on
Voltage on 5 V line	ADC2 Channel 12	Memory
Amperage on 5 V line	ADC2 Channel 8	Memory
Amperage on 3 V line	ADC2 Channel 7	Memory
Voltage on High-Volt. line	ADC2 Channel 4	Memory
Amperage on High-Volt. line	ADC2 Channel 3	Memory
Amperage on TEC line	ADC2 Channel 5	Memory
Used for temperature of diode	ADC2 Channel 11	Memory
Error between diode temp. and setpoint	ADC2 Channel 6	Memory
Reference voltage for TEC measurements	ADC2 Channel 9	Memory
Temperature 1	IC1	I2C bus
Temperature 2	IC2	I2C bus
Temperature 3	IC3	I2C bus
Microcontroller temperature	internal sensor	Memory
Rotational speed	Gyroscope	I2C bus

Table 3.2: List of all House-Keeping data of the payload, with their associated sensors.

with one of the two Direct Memory Access controllers (DMA), so that when the House-Keeping data acquisition starts, all ADCs related to HK data are continuously sampled and their values are copied to specific memory addresses. This way, the freshest data is always accessible with a simple memory read access. The three temperature sensors and the gyroscope work in a similar way, only they are independent of the main processor, and reading their measurements requires an I2C read access. To provide higher accuracy, all HK data can be binned by up to 16 points before being sent to the main computer.

A complete list of all the HK data and sensors can be found in Table 3.2. The position of the three temperature sensors and of the gyroscope are given in Figure 3.8.

3.4.3 Science acquisitions

In science mode, the payload autonomously switches between the three acquisition sub-modes (imaging, searching, and tracking).

3.4.3.1 Imaging phase

The imaging phase (IMG) is the initial phase of the science sequence. In this phase, the payload has no clue where the star is located in the focal plane of the telescope. Thus, the payload starts by scanning the entire field in search for some bright sources. When the scan is finished, if a source bright enough has been found, its position is memorized, and the payload switches to the searching phase.

1. The payload scans the entire field, starting at $x = 0$ μm, $y = 0$ μm, up to $x \simeq 430$ μm, $y \simeq 430$ μm, with a given step (typically $\simeq 3$ μm).

2. Denoting $c_{\max}$ the highest count value found, the payload determines whether or not $c_{\max} > c_{\text{star}}/2$, with c_{star} the typical count value expected for the target object, and the given integration time. If it is, the position (x_0, y_0) of this maximum is memorized, and the payload moves to the searching phase. If not, the payload starts over again at step 1.

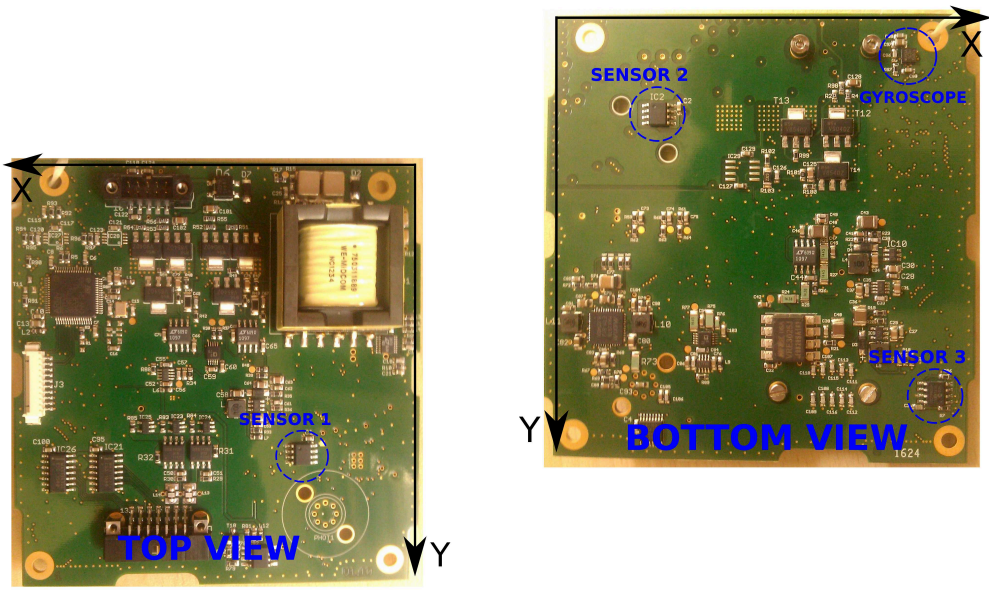


Figure 3.8: Top and bottom view of the electronic board, with the position of the three temperature sensors.

3.4.3.2 Searching phase

In the searching phase, the payload has a rough idea of the location of the star in the focal plane, but not precise enough to initiate the tracking algorithm. The payload will use a spiral pattern to help narrow down the exact location of the star.

The equation of the spiral used to search for the star is:

$$x(t) = x_0 + a \times bt \times \cos\left(\sqrt{bt}\right)$$
$$y(t) = y_0 + a \times bt \times \sin\left(\sqrt{bt}\right)$$

where t is the time, and a and b are parameters which can be modified to obtain a slow or fast spiral, of the desired compacity.

The spiral extends in the outward direction up to a given maximum radius. If no star is found at this point, the payload goes back to imaging. If a star is found (i.e. high count values are detected at multiple successive positions), the payload immediately switches to the tracking algorithm.

3.4.3.3 Tracking phase

In tracking mode, the payload knows were the star is located in the focal plane, and the tracking algorithm keeps the fiber centered on it. It will continue to do so until the star is lost (whichever comes first: end of target visibility, destruction of the satellite by advanced aliens, destruction of the target star by very advanced aliens, or a bug in the tracking algorithm). A detailed description of the algorithms used for tracking is given in Chapter 4.

Noise source	Noise level	Note
Photon noise	60 ppm/hr	-
Readout noise	0 ppm/hr	Photon counting detector
Dark current	< 1 ppm/hr	-
Gain stability	20 ppm/hr	100 μV stability of bias voltage
Scattered light	< 1 ppm/hr	from Moon or Earth, filtered by the fiber
Thermal noise	40 ppm/hr	Not including induced optical deformations
Pointing stability	80 ppm/hr	One of my thesis objectives
Total	108 ppm/hr	calculated using $\sqrt{\sum {N_k}^2}$

Table 3.3: Photometric noise budget for the PicSat instrument.

3.5 Photometric budget

For PicSat to be able to achieve its science objectives, the science instrument needs to be able to reach a photometric precision of ~ 100 ppm/hr. The complete noise budget can be broken in different parts, discussed below. The total budget is given in Table 3.3.

Photon noise The visible apparent magnitude of β Pic is $M_{\mathrm{V}} = 3.86$. Taking into account the fiber injection ratio η, the transmission of the instrument T, and the wavelength width of the detector sensitivity curve $\Delta\lambda$, the total flux of energy propagating into the fiber is given by:

$$F = F_{\mathrm{Vega}} \times 10^{-\frac{3.86}{2.5}} \times \frac{\pi D^2}{4} \times \eta \times T \times \Delta\lambda \tag{3.1}$$

Using a reference wavelength of $\lambda_{\mathrm{eff}} = 545$ nm for the visible photons, the photon flux is:

$$F_\gamma = \frac{F_{\mathrm{Vega}} \times 10^{-\frac{3.86}{2.5}} \times \frac{\pi D^2}{4} \times \eta \times T \times \Delta\lambda}{\frac{hc}{\lambda_{\mathrm{eff}}}} \tag{3.2}$$

Using $F_{\mathrm{Vega}} \simeq 360 \times 10^{-14}$ W cm^{-2} s^{-1} μm^{-1}, $\eta \simeq 0.8$, $\Delta\lambda \simeq 250$ nm and $T = 0.2$ (which includes optical transmission and quantum efficiency of the detector), we have:

$$F_\gamma \simeq 16 \times 10^4 \text{ photons/s} \tag{3.3}$$

In practice, though, the fiber will not be kept centered on the star image, but it will rather be kept at a given offset (see Chapter 4 for details). The effective injection ratio will be closer to ~ 0.4, and the photon flux will be $F_{\gamma,\mathrm{eff}} = 8 \times 10^4$ photons/s. The noise level is thus:

$$N_{\mathrm{photon}} = \sqrt{\frac{1}{F_{\gamma,\mathrm{eff}} \times 1 \text{ hr}}} \simeq 60 \text{ ppm/hr} \tag{3.4}$$

Detector noises The detector noise can be broken down into three different contributions: readout noise, dark current, and gain stability. The payload uses a photon counting detector, so there is no readout noise. The ID101 photodiode has a dark current of $F_{\mathrm{dark}} = 1000$ photons/s. The photon flux is given in Equation 3.3. The dark current noise level is given by:

$$N_{\mathrm{dark}} = \frac{\sqrt{F_{\mathrm{dark}} \times 1 \text{ hr}}}{F_\gamma \times 1 \text{ hr}} < 1 \text{ ppm/hr} \tag{3.5}$$

The gain of the SPAD depends on the bias voltage applied. The payload electronic board was designed with the requirement of having a bias voltage stable down to 100 μV, ensuring a gain stability of 20 ppm. This is an absolute stability, independent of time.

Scattered light Estimating the impact of scattered light on the photometric measurement of the PicSat instrument is difficult, mainly because it requires to calculate the coupling between the single-mode fiber and the light scattered by the mechanical surfaces within the instrument. An upper-limit order of magnitude can be proposed, though. Considering the case of Figure 3.9, in which a surface element $\mathrm{d}S$ of the primary mirror is illuminated by an incoming flux F, and scatters a fraction Q of this flux, the total amount of power which is scattered is simply given by:

$$P_{\text{scattered}} = QF\mathrm{d}S \tag{3.6}$$

In the case of an isotropic scattering limited to the half-space not occupied by the rest of the parabola, the flux received at the fiber level is:

$$F_{\text{at fiber}} = \frac{QF}{2\pi f^2}\mathrm{d}S \tag{3.7}$$

We can make the pessimistic assumption that all photons hitting the head of the fiber core are then guided into the fiber. The total amount of power that penetrates into the fiber is given by the incoherent sum of the power from each surface element:

$$P = \iint_{\text{parabola}} \frac{QF}{2}\left(\frac{R_{\text{fib}}}{f}\right)^2 \mathrm{d}S = \frac{Q}{2}\left(\frac{R_{\text{fib}}}{f}\right)^2 FS_{\text{parabola}} \tag{3.8}$$

where S_{parabola} is the total area of the parabola. We can see that the total flux which penetrates into the fiber is given by the flux received by the parabola FS_{parabola}, attenuated by a certain factor which depends on the focal-length of the parabola, and the core radius. The fiber used for PicSat has a core radius of $\sim 2\mu$m, and the focal length is $f = 15$ cm. Assuming a scattering coefficient of 1%, the attenuation factor is:

$$\frac{Q}{2}\left(\frac{R_{\text{fib}}}{f}\right)^2 \simeq \times 10^{-12} \quad \text{i.e.} \quad -120 \text{ dB} \tag{3.9}$$

With this, the ratio of the light scattered by the Moon (visual magnitude of -12.6) to the light from β Pic is $\sim 10^{-8}$. In a single-mode instrument, scattered light is negligible.

Thermal stability The SPAD is also sensitive to temperature, which is why it is mounted on a thermo-electric system to stabilize its temperature. The target temperature stability is 0.01°C, and the detector error is 0.4%/°C. Thermal noise thus accounts for $\sim$ 40 ppm/hr, assuming that temperature fluctuations occur on orbital timescales. This noise does not take into account the photometric variations induced by thermal deformations of the optics, which can be a serious source of concerns (see Chapter 6).

Pointing stability To limit the total photon noise to 100 ppm/hr, the noise from the pointing stability needs to be below 80 ppm/hr. Limiting this noise was one of the main goals of my thesis (see Chapter 4)

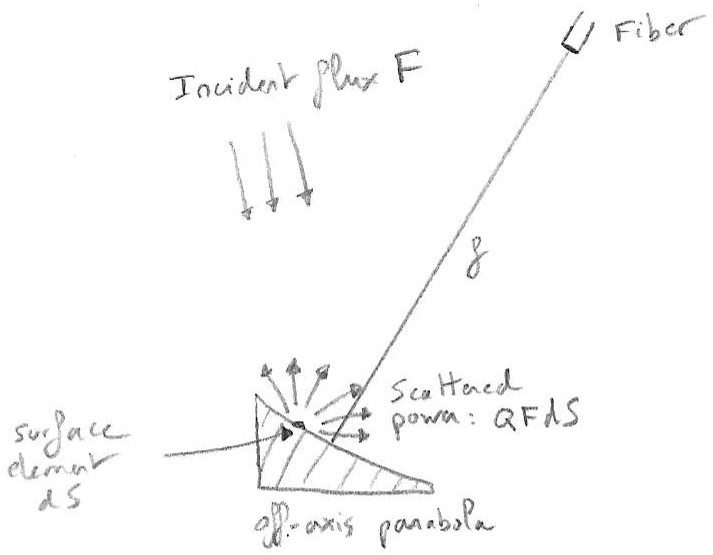


Figure 3.9: An incident flux F illuminates the primary mirror. The scattered light illuminates the fiber located at a distance f.

Chapter 4

Tracking algorithm

Content of this chapter

4.1 Fiber injection and photometric stability

4.1.1 Single-mode fiber injection in PicSat

4.1.1.1 Theoretical injection for PicSat

In the ideal case, when the instrument is perfectly pointed toward its target, the injection ratio as a function of the fiber position (X, Y) is nearly Gaussian, with a Full-Width at Half-Maximum (FWHM) which depends on the wavelength of the observation. In Figure 4.1, I give the injection ratio as a function of the fiber distance to the center of the star image, for a monochromatic observation at 500 nm, and a large band observation of β Pictoris (taking into account the detector's response and the stellar spectrum).

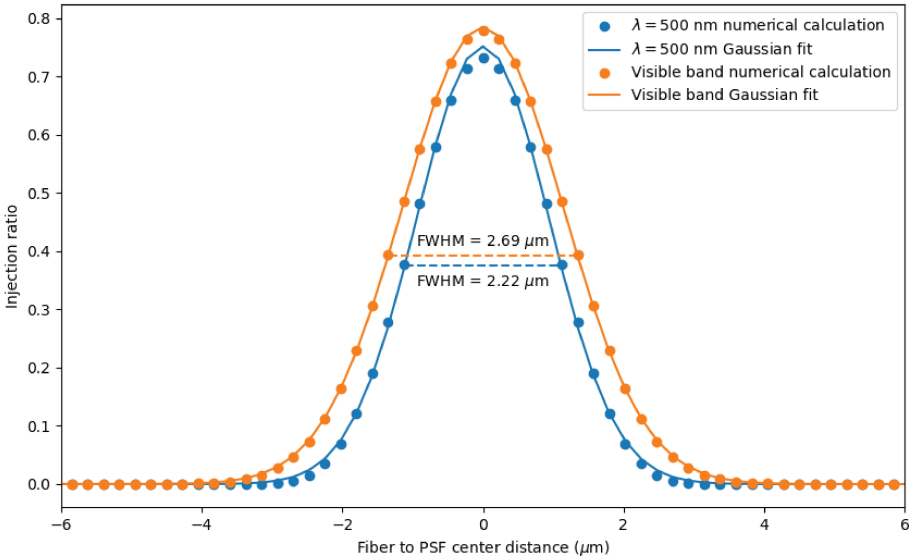


Figure 4.1: Injection ratio as a function of the distance between the fiber and the center of the Airy disk, calculated for the PicSat instrument. The orange dots correspond to an observation at $\lambda = 500$ nm, and the blue dots correspond to a large band visible observation, taking into accounts the quantum efficiency of the detector, and for a target spectrum similar to β Pictoris. The two lines corresponds to the best Gaussian fits to each series. These calculations are valid for an on-axis target. For an off-axis target, a term accounting for the astigmatism of the off-axis parabola needs to be added.

4.1.2 Tracking errors and photometric noise budget

The visible injection for the PicSat instrument is well approximated by a Gaussian function:

$$\eta_V(X, Y) = \eta_0 \times \exp\left(-\frac{X^2 + Y^2}{2\sigma^2}\right) \tag{4.1}$$

where X and Y represent the positioning error of the fiber with respect to the center of the star image, $\eta_0 \simeq 80\%$, and $\sigma = 1.14$ μm corresponds to a FWHM of 2.69 μm. This injection function can be converted to angular coordinates by dividing by the focal length $f = 14.84$ cm and then to arcseconds. In angular coordinates, we have: $\sigma = 1.58$ as.

I will explain later in this section that tracking the star image with the fiber requires to modulate the fiber position around the center of the image. Thus, the fiber is never centered on the star image, but is kept at a certain position X_m, Y_m, typically at a distance of $\sim 0.5 \times$ FWHM from the center. The injection is then $\sim \eta_0/2$, and the derivatives are given by:

$$\frac{\partial \eta_V}{\partial X}(X_m, Y_m) = -\frac{\eta_0 X_m}{2\sigma^2} \tag{4.2}$$

$$\frac{\partial \eta_V}{\partial Y}(X_m, Y_m) = -\frac{\eta_0 Y_m}{2\sigma^2} \tag{4.3}$$

Small pointing errors of standard deviations σ_X, σ_Y will turn into a relative injection error given by:

$$\frac{\sigma_\eta}{\eta_V} = \frac{\sqrt{(X_m \sigma_X)^2 + (Y_m \sigma_Y)^2}}{\sigma^2} \tag{4.4}$$

If we consider that the error is the same on the two axes ($\sigma_X = \sigma_Y = \sigma_{\text{pos}}$), and assuming $\sqrt{{X_m}^2 + {Y_m}^2} \simeq \text{FWHM}/2$, we have:

$$\frac{\sigma_\eta}{\eta_V} \sim \frac{\text{FWHM} \times \sigma_{\text{pos}}}{2\sigma^2} = \frac{\sqrt{2\ln(2)} \times \sigma_{\text{pos}}}{\sigma} \tag{4.5}$$

In the photometric budget presented in Chapter 3, the noise induced by tracking errors was assumed to be of the order of 80 ppm/hr, or 13% at 1 kHz (the frequency of the control loop). Using the above equation, this number can be directly translated to an objective for the tracking precision of about 0.13 μm or 0.19 as, at 1 kHz. For a CubeSat platform, this constitutes an ambitious objective.

4.2 A two stage pointing system for PicSat

4.2.1 Satellite attitude control

The orientation of the satellite in space, and the pointing of the science target, are done using a dedicated system on-board the satellite: the Attitude Determination and Control System (ADCS).

4.2.1.1 Sensors

Different sensors can be used to measure the orientation of a satellite in space. The most common sensors are gyroscopes, magnetometers, star trackers, and Sun/Earth sensors.

Gyroscopes are sensitive to the rotation speed of the satellite, and are typically part of an Inertial Measurement Unit (IMU) which integrates the measurements to provide an estimate of the orientation. These sensors are usually very accurate and can acquire data at high frequency (100 Hz or even more). The main drawback of IMUs is their sensitivity to drift errors.

Magnetometers and star trackers are two sensors which directly measure the orientation of the satellite. A magnetometer measures the orientation of the magnetic field in the satellite frame, which can be converted to an absolute orientation of the satellite by knowing the absolute orientation of the magnetic field itself. Thus, magnetometer are usually reserved for satellite orbiting in Low Earth Orbit (LEO), where the Earth's magnetic field is strong enough

to provide accurate measurements. A model of the Earth's magnetic field and a good estimate of the orbital position of the satellite are necessary to convert the magnetic measurement to an attitude measurement.

Star trackers are by far the most accurate sensors available for satellite attitude control systems. A star tracker is a camera which looks at the stars. An algorithm is used to recognize the stars in the camera field-of-view, and to deduce the orientation of its line-of-sight, and thus of the satellite. Star trackers have two important drawbacks: they usually have a low refresh rate (a few Hz), and can only operate if the satellite is stable enough. If the satellite is rotating too quickly, the stars are blurred on the images, and the algorithm cannot recognize any pattern. For CubeSat star trackers, the maximum tolerated slew rate is typically ~ 0.1 to 1 deg/s, and the accuracy of the attitude measurement is < 1 arcmin.

Finally, the last common type of sensor used in ADCS are Sun and/or Earth sensors. A Sun sensor can be as simple as a set of photodiodes placed on different parts of the satellite, and used to determine the direction of the Sun relative to the satellite. This measurement can be converted to an estimate of the orientation of the satellite if the position of the satellite relative to the Sun is known. Similarly, Earth sensors are used to detect the direction of the Earth, usually using infrared detectors.

4.2.1.2 Actuators

On the actuator side, different options are available depending on the mission requirements.

The simplest actuators which can be used are probably the magnetorquers. Magnetorquers are coils which are used as electro-magnets: a current is sent in the coil to generate a magnetic field, which will interact with the Earth's magnetic field to generate a torque on the satellite. Magnetorquers contain no moving parts, and are very reliable. However, they can only generate small torques, and only in the plane orthogonal to the Earth's magnetic field, which is why they are not used for fast maneuvering and precise tracking. Furthermore, these actuators are reserved to satellites in LEO, where the magnetic field is strong enough to generate useful torques.

For missions in which precision pointing is required, reaction wheels can be used. A reaction wheel is a wheel of high moment of inertia, which can be set in rotation by a motor. Accelerating or decelerating the rotation of the wheel creates a reaction torque on the satellite. The main drawback of such systems is that they are subject to *saturation*. If the system is used to compensate for a constant external torque (residual atmospheric friction, for example), the wheel will be constantly accelerated in one direction, and will saturate, i.e. reach its maximum allowed speed. Another similar system which is not subject to this restriction is the control moment gyroscope. This actuator is similar to a gyroscope, and uses a wheel rotating at a constant speed. The rotation axis of the wheel can be changed to create a gyroscopic torque to control the orientation of the satellite. Both reaction wheels and control moment gyroscopes are also subject to potential vibrations, if the wheels are not perfectly balanced (which they never are). Reaction wheels are much more common on CubeSat platforms than control moment gyroscopes.

Finally, thrusters can also be used to control the orientation of a satellite. But these systems are fairly uncommon on CubeSat missions.

4.2.1.3 iADCS100 on-board PicSat

The attitude determination and control system selected for the PicSat mission is the iADCS100 from Hyperion Technologies. This system uses a combination of magnetometers, gyroscopes,

SPECIFICATIONS

Performance				
Total momentum storage per axis		+/-1.5, +/- 3.0, +/- 6.0[1]		mN.m.s
Maximum torque		> 0.087		mN.m
Nominal magnetic moment		0.2 (X, Y), 0.1 (Z)		$A.m^2$
Attitude determination accuracy		30		arcseconds (3σ)
Pointing accuracy		<< 1		°
Slew rate		> 1.5[2]		°/s
Radiation tolerance		> 45[9]		krad (Si)
Operating temperature		- 45 / - 20 to + 40 / + 85[7]		°C
Dimensions				
Outer dimensions		95 x 90 x 32		mm
Mass[1]		400 / 435 / 470		g
Electrical specifications				
	Min.	Typ.	Max.	
Supply voltage	4.0	5.0[3]	15[4]	V
Bus logic level voltage		Referenced to Vsys[5]		V
Power consumption:				
Idle	-	1150	900[6]	mW
Nominal[7]	-	1400	-	mW
Peak[8]	-	4000[1]	4500[6]	mW

[1] Depending on the reaction wheel models
[2] For a 3U CubeSat with the RW210.15 reaction wheel complement, over all axes
[3] When using the 5V system power pins on the standard CubeSat header
[4] When using the VBAT pin on the standard CubeSat header
[5] Vsys can range from 3.3 to 5.1V for I²C applications.
[6] To be confirmed
[7] Depends on use case
[8] Peak values are given to size the power supply. Power consumption can be limited by the iADCS to match supply
[9] Not accounting for Star Trackers and Reaction Wheels used

Figure 4.2: Specifications of the iADCS100 used in the PicSat mission. Extracted from the flyer HT-IADCS100-V1.02F, publicly available on Hyperion Technologies website.

and the ST200 star tracker for attitude determination, and a combination of 3 RW210 reaction wheels, and 3 magnetometers for attitude control. The specifications of the system are given in Figure 4.2.

4.2.1.4 ADCS of the ASTERIA mission

The Arcsecond Space Telescope Enabling Research in Astrophysics (ASTERIA) mission [Smith et al., 2018] was deployed from the International Space Station in Q4 2017. The objective of the mission was to demonstrate the feasibility of high pointing precision with a nanosatellite. The ADCS selected for this 6U CubeSat mission was the Blue Canyon Technology (BCT) fleXible Attitude Control Technology (XACT), which uses the same types of sensors and actuators as the iADCS100 selected for PicSat. Pong [2018] published on-sky results obtained with ASTERIA, where they show that the XACT achieved a pointing precision of ~ 6 arcsec. In particular, they give the power spectral density of the pointing errors obtained with this attitude control system. The power spectral density is reproduced in Figure 4.3. Since no such data could be obtained with the iADCS100 on-board PicSat, this dataset from ASTERIA remains the best reference dataset to simulate the PicSat payload performances.

4.2.1.5 A Matlab/Simulink ADCS model

Since the data from the ASTERIA mission were not available at the time of the development of the PicSat mission, and since no other comparable set of data exists in the literature, I

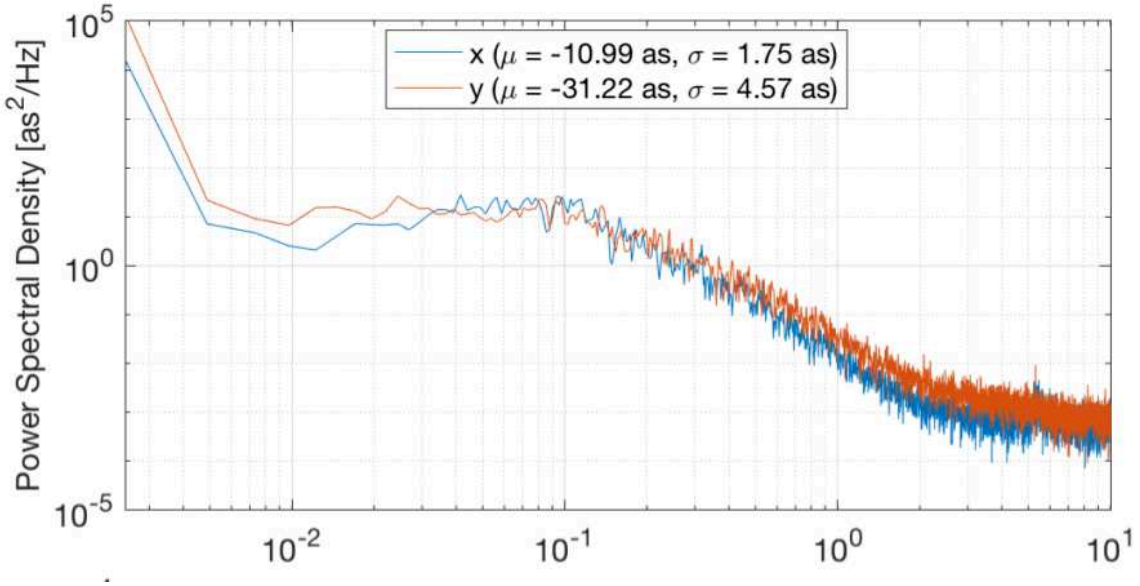


Figure 4.3: Power spectral density for the pointing errors obtained on-sky with the ASTERIA mission (Figure taken from Pong [2018]).

had to resort to making my own ADCS model, in order to generate plausible realistic errors to test the payload control loop. The simple model I developed is presented in Annex C. In this model, the ADCS uses an IMU and a star tracker as its two sensors, and a set of reaction wheels for which I neglected any vibrations due to a potential imbalance.

4.2.2 Conceptual payload control loop

The single-mode fiber in the PicSat instrument is mounted on a two-axis piezo actuator, which can be used to control the fiber position in the focal plane, in order to correct the pointing errors of the attitude control system. But since the detector used in PicSat has only a single pixel, determining the best command to apply on the piezo stage in order to keep the fiber centered on the star is tricky. Assuming that the center of the star image is located at a position $(x_{\text{star}}, y_{\text{star}})$ in the focal plane, taking a measurement with the fiber located at a position $(x_{\text{fib}}, y_{\text{fib}})$ gives a count value proportional to the injection ratio:

$$c = c_0 \times \eta_0 \exp\left(-\frac{(x_{\text{fib}} - x_{\text{star}})^2 + (y_{\text{fib}} - y_{\text{star}})^2}{2\sigma^2}\right) \tag{4.6}$$

Assuming that the coefficient of proportionality c_0, related to the stellar magnitude, is known, as are the instrumental coefficients η_0 and σ, this measurement c can be inverted to obtain an estimate of the quantity:

$$r^2 = (x_{\text{fib}} - x_{\text{star}})^2 + (y_{\text{fib}} - y_{\text{star}})^2 \tag{4.7}$$

which is the squared distance between the fiber and the star

Even ignoring any potential source of noise, a single measurement c is not enough to determine with certainty the position of the star in the focal plane. Multiple measurements obtained at different positions are required to infer the position of the star. In the PicSat

instrument, the position of the fiber is constantly modulated around a central position:

$$x_{\text{fib}}(t_k) = x_0(t_k) + x_{\text{mod}}(t_k) \tag{4.8}$$

$$y_{\text{fib}}(t_k) = y_0(t_k) + y_{\text{mod}}(t_k) \tag{4.9}$$

where t_k is a discrete time step, (x_0, y_0) the target central position, and $(x_{\text{mod}}, y_{\text{mod}})$ is the modulation applied on the fiber. The simplest modulation is a circular modulation of radius r_{mod} and frequency f_{mod}, for which:

$$x_{\text{mod}}(t_k) = r_{\text{mod}} \times \cos(2\pi f_{\text{mod}} t_k) \tag{4.10}$$

$$y_{\text{mod}}(t_k) = r_{\text{mod}} \times \sin(2\pi f_{\text{mod}} t_k) \tag{4.11}$$

The frequency of the modulation is typically ~ 100 Hz, and the radius is of the order of half the FWHM of the injection ratio: $r_{\text{mod}} \sim 1.35$ μm.

Combining measurements made at different times, and thus at different points along the modulation radius, it is possible to derive an estimate the position of the star in the focal plane: $(\hat{x}_{\text{star}}, \hat{y}_{\text{star}})$. This estimated star position can be used as the reference position for the fiber: $(x_0, y_0) = (\hat{x}_{\text{star}}, \hat{y}_{\text{star}})$.

In practice, the algorithm used to estimate the position of the star can take the form of an extended Kalman filter (see Section 4.5), but this algorithm also needs an estimate of the shape of the injection function η, which can change slightly due to environmental variations. Variations of the injection function are expected to be slow compared to pointing noise, and a dedicated soft real time algorithm can be used to estimate these parameters. In practice, this soft real time algorithm runs of the satellite's main computer, and not on the payload microcontroller.

Integrations with the photodiode are done at 1 kHz, which constitutes a reasonable upper limit. Higher frequencies for data acquisition are difficult to reach for three different reasons:

- The flux expected from β Pictoris taking into account optical transmission, quantum efficiency of the detector, peak injection ratio, etc. is ~ 160 photons/ms (see photometric budget in Section 3.5). Using a modulation radius of half the FWHM of the injection function reduces this value to ~ 80 photons per integration. Higher integration frequencies would reduce this value even further, making the tracking loop extremely sensitive to photon noise.

- Running the Kalman filter at higher frequencies requires more calculation on the STM32F3 microcontroller, which can only run at 72 MHz maximum. In practice, while it might be possible to run the filter at slightly higher frequencies (a few kHz), a speed test has shown that the filter can take up to $\simeq 235$ μs to perform an integration, with a mean time of $\simeq 200$ μs. Other tasks are also devoted to the microcontroller (gathering housekeeping data, sending data to the main computer, etc.), and limiting the Kalman filter frequency to 1 kHz seemed safe.

- The data link between the payload and the satellite main computer has a maximum data rate of $\simeq 400$ kb/s. A single photometric data point, which comes with a value for the number of photons measured, two values for the command applied on the piezo, two values for the position measured by the strain gauges, and a value to reference the position on the modulation pattern, represents 9 bytes of data. Offloading photometric data acquired at 1 kHz requires a data rate of at least ~ 100 kb/s, without counting the overhead of the communication protocol, and the other types of data that need to be sent to the computer (beacons, monitoring data, etc.). Again, 1 kHz for the data acquisition seemed to be a safe limit to guarantee the availability of the OBC-payload link for other types of data.

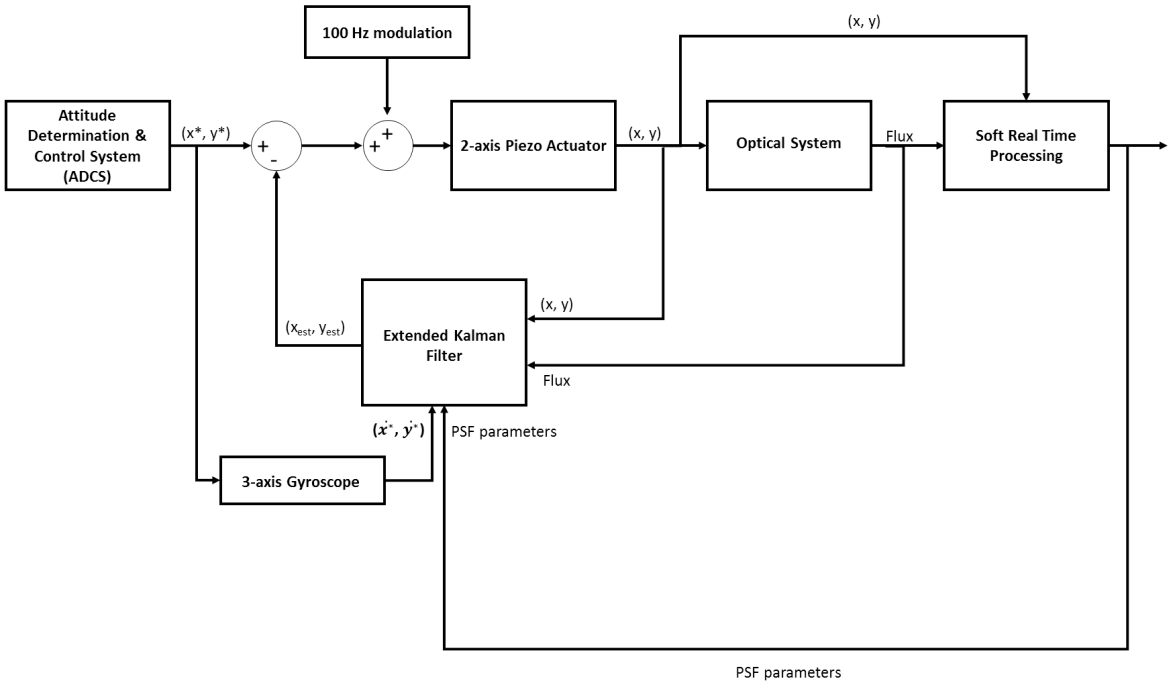


Figure 4.4: Block diagram of the conceptual control loop used as a second stage fine pointing system in the PicSat payload. The ADCS residual pointing errors are the input of this control loop, whose objective is to correct them. To do so, the position of the fiber is modulated in the focal plane, and multiple photometric measurements acquired at different positions along the modulation pattern are combined in an extended Kalman filter to derive a real time estimate of the star position. The loop is closed by using this best estimate of the star position as a reference position for the fiber.

Finally, the 3-axis gyroscope embedded on the payload electronic board can be used to estimate the rotation speed of the satellite, which provides a measurement of the drift speed of the star in the focal plane. This measurement can also be fed to the Kalman filter for improved accuracy.

A block diagram illustrating this conceptual control loop is given in Figure 4.4.

4.2.3 Testing strategy

ADCS test bench

Testing the two-stage control loop for PicSat in a realistic situation is difficult. The best way to do so requires a working attitude control system to generate the pointing errors. At the beginning of the PicSat project, a complete test bench was envisioned to test the control loop in a realistic situation. The concept included an air-bearing table to simulate the frictionless space environment, and a tablet to simulate the night sky.

In this setup, illustrated in Figure 4.5, the satellite is fixed on the air-bearing table, and faces the tablet screen. The star tracker, which is the most accurate sensor of the attitude control system, sees the tablet screen through a lens of appropriate focal lens to project that screen at infinity. By illuminating certain pixels on the screen, it is possible to simulate a night sky. The star tracker generates measurements which are fed to the attitude control

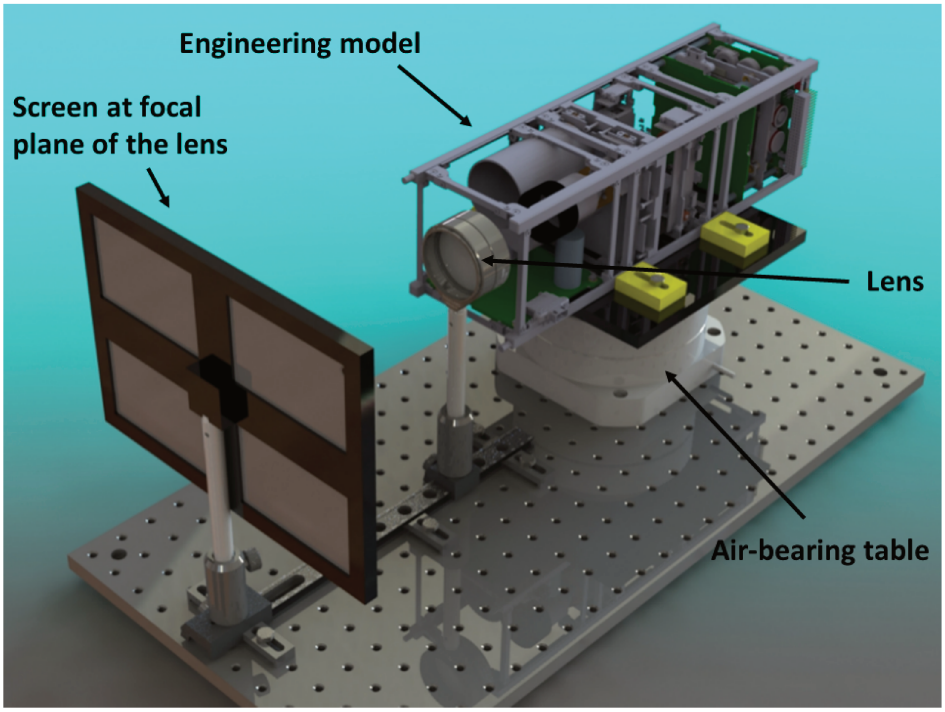


Figure 4.5: A 3D rendering of the test bench envisioned at the beginning of the project to test the attitude control system of the satellite, and the payload control loop.

system. The system reacts and control the free axis of the satellite to keep it stable. In this setup, the ADCS itself generates pointing noise on the free axis, and a collimated light source can be used to illuminate the payload telescope and test the interaction between the ADCS and the second stage system.

Unfortunately, balancing the air-bearing table, using the ADCS in a single axis mode, and tricking the star tracker to act like if it was under the night sky proved to be too complicated. This test bench never provided the expected performances, and although it proved to be useful to test the star tracker itself, it could not be used to test the full pointing system.

Payload tracking test bench

Since I could not use the ADCS bench to test the payload, I had to set up my own bench dedicated to the tracking loop. One of the main problem encountered in the development of this test bench was the problem of the gyroscopes.

The payload electronic embeds a small 3-axis gyroscope, whose measurements can be fed into the tracking algorithm to improve its performance. Simulating displacements of the star in the focal plane of the telescope can be easily done using a tip/tilt mirror between the collimated light source and the payload telescope. However, while doing so, the telescope itself is kept fixed, and the measurements returned by the gyroscope are unrelated with the star displacements. To properly simulate payload pointing errors, the telescope and the electronic board themselves should be put on a tip/tilt mount. Although I attempted to set up such a bench, it proved to be too time consuming, and could not be done within the imparted time.

Since it was not possible to set up a bench in which the telescope itself was moving to test the complete payload algorithm, and because the time was limited to develop the bench, I opted for a radically different strategy: I did not use the test bench to directly test the

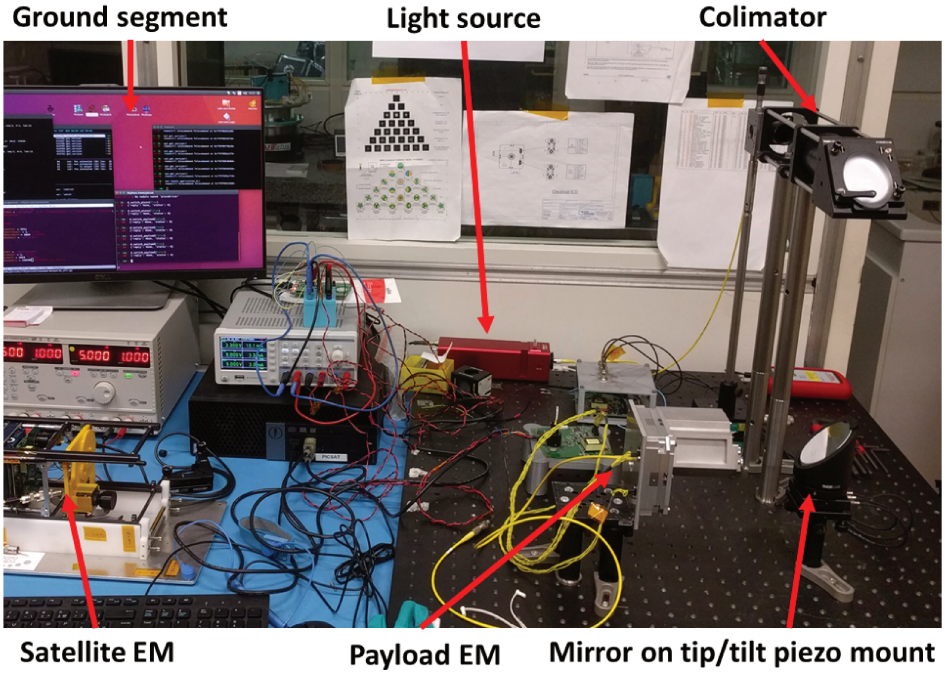


Figure 4.6: Payload test bench

algorithm, but rather to validate the MATLAB/SIMULINK model. This decision lightened the requirements on the test bench, and made possible to build it rapidly.

A picture of the test bench is given in Figure 4.6. The payload telescope, mounted on the two-axis piezo stage, is fixed on a square mount, and faces a 45 degree mirror on a tip/tilt mount controlled by a piezo. A collimated light-beam is sent on the 45 degree mirror, and illuminates the payload telescope. By controlling the tip/tilt piezo actuators, it is possible to move the image created by the collimated beam in the focal plane of the telescope. To simplify the development of the software used for the bench, a payload electronic board was used to control the tip/tilt actuators. Using a payload board to control the tip/tilt actuators came at a price, though, as it was not possible to inject arbitrary displacements. Since the payload control loop operates at 1 kHz, the tip/tilt itself needs to be controlled at least at this same frequency, and ideally at a higher rate. A total of 4 bytes are necessary to set the values of the two high-voltage lines used to control the tip/tilt mount. Thus, a sequence of 1 s of tip/tilt motion takes at least 4 kB of memory space, and ideally should take $\simeq$ 16 to 20 kB (oversampling by a factor 4 or 5 compared to the payload board). The STM32F3 payload microcontroller has only 80 kB of RAM, which can store 5 s of data at most.

To overcome this issue, the payload board was programmed to control the tip/tilt mount using simple sequences generated by the microcontroller itself (sine waves, ramps, etc.). These simple waveforms can be used to compare the performance of the tracking algorithm on the test bench to results obtained with similar waveforms on the MATLAB/SIMULINK model. If the results are similar for sinewaves of representative frequencies (0.1 Hz $< f <$ 10 Hz), the model can be considered to be reliable, and the results it generates can be trusted.

Finally, I also had to develop my own software to provide real time data visualization for the payload on the test bench. The software, programmed in Python, reads and unpacks the data packets from the payload. All quantities of interest for the user are then automatically displayed in different graphs, to provide a real-time feedback to the user.

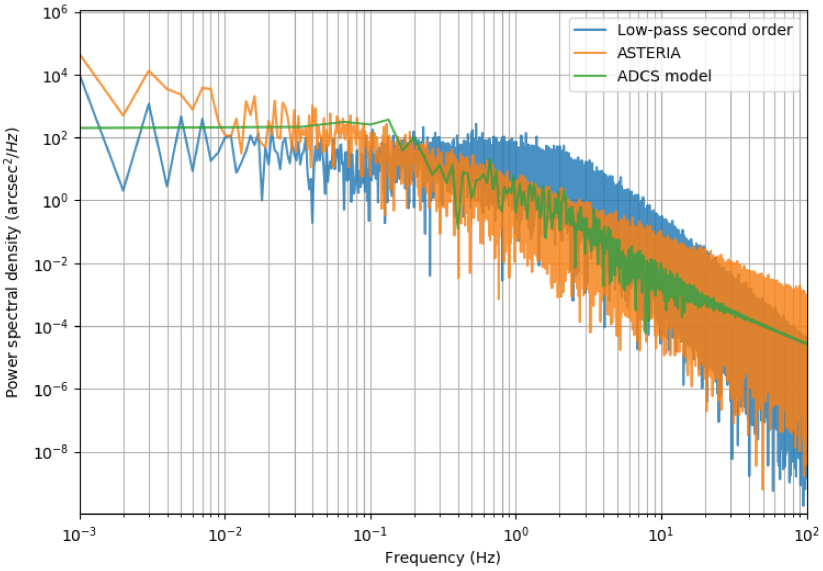


Figure 4.7: Power spectral densities for the pointing residuals in the three test scenarios. All curves have be normalized to the same standard deviation: 10 arcsec.

4.2.4 Testing scenarios

Since the exact performance of the Hyperion Technologies iADCS100 were unknown, I decided to use two different residual error profiles to test the tracking algorithm, to which I added a third scenario when the data from the ASTERIA mission were released.

The first scenario uses the pointing errors generated using the MATLAB/SIMULINK ADCS model I developed (see Annex C), normalized to a standard deviation of 10 arcsec, which was the expected $1\,\sigma$ error of the iADCS100.

The second scenario simulates a case in which the attitude control system generates more noise in the 1-10 Hz range. In this scenario, the power spectral density of the pointing residual errors is represented by a low-pass filter of second order, with a cutoff frequency of $f_0 = 1$ Hz:

$$H(f) = \frac{1}{1 + \left(\frac{f}{f_0}\right)^2} \tag{4.12}$$

In this scenario also, the sequence is normalized to a standard deviation of 10 arcsec.

For the third scenario, I used the power spectral density from the ASTERIA mission to generate the residual errors. The power spectral density published by the ASTERIA team is limited to frequencies < 10 Hz (higher frequencies could not be measured by ASTERIA), so instead of using the ASTERIA spectral density directly, I used a first-order model for which I adjusted the parameters to be as close as possible to the ASTERIA density. In this case also, the sequence of pointing errors was normalized to a standard deviation of 10 arcsec.

The three power spectral densities corresponding to these test scenarios are shown in Figure 4.7.

4.3 Kalman filtering for PicSat

Kalman filters are used in a variety of fields, from control theory to statistics, parameter estimation, data filtering, etc. The consequence is that there are probably as many ways to present and understand Kalman filters as there are field in which they are used. I personally tend to view Kalman filters as data fusion estimators, used to combine different sources of information on a system to provide the best estimate of their state vector.

In the case of the PicSat payload, the quantity of interest for the control loop is the position of the star in the focal plane. This position varies with time due to forces applied on the satellite by the environment and the attitude control system. The evolution of the star position could be described using a position-velocity-acceleration model. But since the payload has no information on the forces applied on the satellite, including the acceleration is of little interest, and increases the dimensions of the vectors and matrices manipulated by the algorithm. Since computing power was a concern on the small STM32F3 microcontroller, I adopted a position-velocity model.[1]

Integrations with the photodiode are done at a frequency f corresponding to a period dt, a discrete time: $t_k = k \times dt$ can be adopted. In a position-velocity model, the state vector can be defined as:

$$X_k = \begin{bmatrix} x_{\text{star}}(t_k) \\ y_{\text{star}}(t_k) \\ v_{x,\text{star}}(t_k)\, dt \\ v_{y,\text{star}}(t_k)\, dt \end{bmatrix} \tag{4.13}$$

where the velocity has been multiplied by dt to simplify the state evolution matrix, and to have all components with the same unit. The state vector at time t_{k+1} can then be modeled as a linear combination of the elements on the state vector at time t_k:

$$X_{k+1} = AX_k + \epsilon_k \tag{4.14}$$

Where A is the state evolution matrix, and where ϵ_k represents a source of "noise" applied on the system. In reality, ϵ_k corresponds to the effect on the state vector of the different forces applied on the satellite, and which are not taken into account by the evolution matrix A. From the perspective of the payload, these unknown forces can be considered as a random vector. This random vector ϵ_k will be assumed to follow a normal law of mean 0, and of constant variance-covariance matrix Q:

$$\epsilon_k \sim \mathcal{N}(0, Q) \tag{4.15}$$

The evolution matrix A have an important impact on the performance of the control loop, and must be carefully selected. The choice of A will be discussed later. But a possible example, corresponding to a constant velocity model, could be:

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.16}$$

[1] Going from a state vector with 4 dimensions to 6 is a big deal: in the Kalman filter, it is necessary to invert matrices of size $n \times n$, where n is the length of the state vector. Since it was not possible to use external linear algebra libraries, I had to program every matrix operation myself on the payload microcontroller (yes, from matrix addition to inversion...). And it turns out that for $n \leq 4$, the inversion can still be calculated using the Cayley-Hamilton formula. For $n = 6$, it is not the case anymore...

The role of the Kalman filter is to provide a real time estimate the state vector: $\hat{X}_k$. To do so, two sources of information are used: the model of the system, and a direct measurement of the position of the star. The basic concept behind the Kalman filter is to proceed in three steps: first an *a priori* estimate of the state vector is obtained using model propagation; secondly, a measurement is performed on the system, which depends linearly on the state vector, and the result obtained gives new information on the true state vector; finally, the estimate obtained with model propagation and the new information coming from the measurements are merged together to provide the final *a posteriori* estimate.

4.3.1 State vector propagation

We assume that an estimate of the state vector at time t_k is available. This estimate is not perfect, and comes with a variance-covariance matrix P_k describing the confidence interval. Since we have some understanding of how the system behaves (via the matrix A), we can propagate this previous knowledge of the state vector at time t_k to estimate the new state vector at time t_{k+1}:

$$\hat{X}_{k+1} = A\hat{X}_k \tag{4.17}$$

While propagating the state vector through the evolution matrix A, it is also necessary to propagate the errors. The previous state vector estimate had a variance-covariance matrix P_k. The variance-covariance matrix of the propagated state vector is thus:

$$P_{k+1} = AP_kA^T \tag{4.18}$$

Propagating the covariance matrix this way does not account for the fact that we know that the evolution model represented by the matrix A is not perfect, and neglects the impact of the process noise ϵ. To take into account the impact of this noise, a second term needs to be added to the new covariance matrix. The rigorous way of doing this is to add the covariance matrix of the process noise ϵ:

$$P_{k+1} = AP_kA^T + Q \tag{4.19}$$

The model propagation equations can be written:

$$\hat{X}_{k+1} = A\hat{X}_k \tag{4.20}$$

$$P_{k+1} = AP_kA^T + Q \tag{4.21}$$

Without any other source of information on the system, this propagation equations will give an estimate of the position which evolves exactly as prescribed by the propagation matrix A, with ever increasing uncertainties ($\lim_{k\to\infty} \|P_k\| = \infty$).

4.3.2 Measurement

The second step in the Kalman filter is to obtain a measurement on the system. In some cases, the state vector itself can be directly measured, but this may not always be the case. For example, in the case of the PicSat payload, the position of the star in the focal plane cannot be directly measured. In the Kalman filter formalism, the only assumption made is that the measurement Z obtained on the system depends linearly on the state vector X:

$$Z_{k+1} = HX_{k+1} + \zeta_{k+1} \tag{4.22}$$

Where H is a matrix representing the linear measurement, and ζ a random vector representing the measurement noise:

$$\zeta_{k+1} \sim \mathcal{N}(0, R) \tag{4.23}$$

4.3.3 Optimal data fusion

Given the *a priori* estimate of the state vector $\hat{X}^{\text{ap}}_{k+1}$, the expected measurement on the system is:

$$Y_{k+1} = H\hat{X}^{\text{ap}}_{k+1} \tag{4.24}$$

And the difference between the real measurement value and this expected measurement is related to the difference between the real state vector value and the *a priori* estimate:

$$Z_{k+1} - Y_{k+1} = H\left(X_{k+1} - \hat{X}^{\text{ap}}_{k+1}\right) + \zeta \tag{4.25}$$

The above quantity is sometimes refer to as the "innovation", in Kalman filter literature. If we multiply this quantity by an arbitrary matrix K and add it to the *a priori* estimate, we get:

$$\hat{X}_{k+1} = \hat{X}^{\text{ap}}_{k+1} + KH\left(X_{k+1} - \hat{X}^{\text{ap}}_{k+1}\right) + K\zeta \tag{4.26}$$

Since the expected value for $\hat{X}^{\text{ap}}_{k+1}$ is X_{k+1}, the expected value for $\hat{X}_{k+1}$ is also X_{k+1}, no matter what the value of K is. $\hat{X}_{k+1}$ is an estimate of the state vector, whose covariance can be minimized by carefully selecting the gain matrix K. Denoting P^{ap}_{k+1} the covariance matrix on the *a priori* estimate $\hat{X}^{\text{ap}}_{k+1}$, the covariance matrix on $\hat{X}_{k+1}$ is given by:

$$P_{k+1} = (I - KH)\,P^{\text{ap}}_{k+1}\left(I - H^T K^T\right) + KRK^T \tag{4.27}$$

The diagonal terms of this matrix give the variances on the individual state vector components. These variances are all minimum when the trace of the matrix is minimum. The trace of P_{k+1} is given by:

$$\text{Tr}\left(P_{k+1}\right) = \text{Tr}\left(P^{\text{ap}}_{k+1}\right) - \text{Tr}\left(KHP^{\text{ap}}_{k+1}\right) - \text{Tr}\left(P^{\text{ap}}_{k+1}H^T K^T\right) + \text{Tr}\left(K\left(HP^{\text{ap}}_{k+1}H^T + R\right)K^T\right)$$

Deriving with respect to the ij-the component of K leads to:

$$\frac{\partial \text{Tr}\left(P_{k+1}\right)}{\partial K_{ij}} = -\left[HP^{\text{ap}}_{k+1}\right]_{ji} - \left[{P^{\text{ap}}_{k+1}}^T H^T\right]_{ij} + \left[K\left(HP^{\text{ap}}_{k+1}H^T + R\right)\right]_{ij} + \left[K\left(HP^{\text{ap}}_{k+1}H^T + R\right)\right]_{ji}$$

And using the fact that covariance matrices are symmetric:

$$\frac{\partial \text{Tr}\left(P_{k+1}\right)}{\partial K_{ij}} = -2\left[P^{\text{ap}}_{k+1}H^T\right]_{ij} + 2\left[K\left(HP^{\text{ap}}_{k+1}H^T + R\right)\right]_{ij}$$

Setting all these derivatives to 0 to minimize the trace finally gives the optimal value for the gain K (sometimes also refer to as the "blending factor"):

$$K = P^{\text{ap}}_{k+1}H^T\left(HP^{\text{ap}}_{k+1}H^T + R\right)^{-1} \tag{4.28}$$

4.3.4 Summary: Kalman filter equations

To summarize, the Kalman filter works in several steps:

State vector propagation The estimate of the state vector at time t_k is propagated to the next timestep t_{k+1} using the following equations, which gives the *a priori* new state vector estimate:

$$\hat{X}^{\text{ap}}_{k+1} = A\hat{X}_k \tag{4.29}$$

$$P^{\text{ap}}_{k+1} = AP_kA^T + Q \tag{4.30}$$

Innovation, or measurement update A linear measurement Z is obtained on the system, and blended with the *a priori* estimate to give the final estimate of the state vector at time t_{k+1}:

$$\hat{X}_{k+1} = \hat{X}^{\text{ap}}_{k+1} + K\left(Z - H\hat{X}^{\text{ap}}_{k+1}\right) \tag{4.31}$$

With:

$$K = P^{\text{ap}}_{k+1} H^T \left(HP^{\text{ap}}_{k+1} H^T + R\right)^{-1} \tag{4.32}$$

The covariance matrix is then:

$$P_{k+1} = (I - KH)\, P^{\text{ap}}_{k+1} \left(I - H^T K^T\right) + KRK^T \tag{4.33}$$

$$\tag{4.34}$$

Which can be simplified into[2]:

$$P_{k+1} = (I - KH)\, P^{\text{ap}}_{k+1} \tag{4.35}$$

4.4 Fast and simple: barycenter-based Kalman algorithm

4.4.1 Photometric barycenter as a position measurement

As explained in the last part of Section 4.2, there is no easy way to obtain a measurement of the position of the star in the focal plane using the single pixel detector of the PicSat payload. One possible way to go around this issue is to use multiple measurements made on different points of the modulation cycle to "guess" the position of the star. The simplest way to do so is to proceed in a similar way as what is usually done to find the position of a star using a 2D detector: calculate the photometric barycenter.

At a given timestep t_{k+1}, the last n data points are collected from the system's memory, n corresponding to the number of points on the modulation pattern. Each data point contains the position of the fiber, set to the best star position estimate plus the modulation, and the photometric count value obtained:

$$(\hat{x}_j + x_{\text{mod},j}, \hat{y}_j + y_{\text{mod},j}, c_j) \tag{4.36}$$

From this set of n data points, the barycenter estimate of the star position is calculated using:

$$x_{\text{bary},k+1} = \frac{\sum_{j=(k-n+1)}^{k} c_j \left(\hat{x}_j + x_{\text{mod},j}\right)}{\sum_{j=(k-n+1)}^{k} c_j} \tag{4.37}$$

$$y_{\text{bary},k+1} = \frac{\sum_{j=(k-n+1)}^{k} c_j \left(\hat{y}_j + y_{\text{mod},j}\right)}{\sum_{j=(k-n+1)}^{k} c_j} \tag{4.38}$$

[2]This simplification uses the definition of the gain matrix K:

$$\begin{aligned} P_{k+1} &= P^{\text{ap}}_{k+1} - KHP^{\text{ap}}_{k+1} - PH^T K^T + KHP^{\text{ap}}_{k+1} H^T K^T + KRK^T \\ &= (I - KH)\, P^{\text{ap}}_{k+1} - P^{\text{ap}}_{k+1} H^T K^T + K(HP^{\text{ap}}_{k+1} H^T + R)K^T \\ &= (I - KH)\, P^{\text{ap}}_{k+1} - P^{\text{ap}}_{k+1} H^T K^T + P^{\text{ap}}_{k+1} H^T K^T \\ &= (I - KH)\, P^{\text{ap}}_{k+1} \end{aligned}$$

In practice, the values of $\hat{x}_j$ and $\hat{y}_j$ comes from the control loop itself, and are changing at each time step. At first approximation, they are evolving at constant velocity (Kalman propagation), and we can approximate $\hat{x}_j$ by $\hat{x}_k - (k-j)\hat{v}_{x,k}\,dt$, and write:

$$\begin{aligned} x_{\text{bary},k+1} &\simeq \frac{\sum_{j=(k-n+1)}^{k} c_j \left(\hat{x}_k - (k-j)\hat{v}_{x,k}\,dt + x_{\text{mod},j}\right)}{\sum_{j=(k-n+1)}^{k} c_j} \\ &= \frac{\sum_{j=(k-n+1)}^{k} c_j \left(\hat{x}_k + \hat{v}_{x,k}\,dt + x_{\text{mod},j}\right)}{\sum_{j=(k-n+1)}^{k} c_j} - \frac{\sum_{j=(k-n+1)}^{k}(k-j+1)c_j}{\sum_{j=(k-n+1)}^{k} c_j} \times \hat{v}_{x,k}dt \end{aligned}$$

with a similar equation for y.

Since $\hat{x}_k + \hat{v}_{x,k}\,dt$ corresponds to the a priori estimate of the position of the star at t_{k+1}, the first fraction in the above equation is exactly the photometric barycenter calculated around the estimated position of the star, and can be taken as a measurement of the position of the star. The second term is a velocity correction. This measurement can be directly integrated into the Kalman filter by using the following measurement equation:

$$\begin{aligned} Z_{k+1} &= H_{k+1} X_{k+1} \\ H_{k+1} &= \begin{pmatrix} 1 & 0 & -\frac{\sum_{j=(k-n+1)}^{k}(k-j+1)c_j}{\sum_{j=(k-n+1)}^{k} c_j} & 0 \\ 0 & 1 & 0 & -\frac{\sum_{j=(k-n+1)}^{k}(k-j+1)c_j}{\sum_{j=(k-n+1)}^{k} c_j} \end{pmatrix} \end{aligned}$$

The fact that the measurement matrix H is not a constant and varies at each iteration of the Kalman filter is not a problem. The reasoning and the proof of the optimality of the Kalman gain proposed at the beginning of this section does not require the assumption of a constant measurement matrix.

One last piece of information about the measurement which is required for the Kalman filter to operate is the measurement noise covariance matrix R. When the measurements are directly taken by sensors, the documented noise levels can be used in the covariance matrix. In the case of the barycenter estimate, the precision depends on the shape of the injection function, on the modulation pattern used, and on the photon noise level. Using integrations of 1 ms with the detector, the maximum number of counts expected on a star like β Pictoris is ~ 100. If the modulation pattern has 10 points spread over a circle of radius equals to half the FWHM of the injection function, the expected count level for each integration is $\simeq 50$. The typical noise on the barycenter measurement is then:

$$\sigma_{\text{bary}} \simeq \frac{\sqrt{\sum_{k=1}^{10} \left(\sqrt{50} \times \text{FWHM}/2\right)^2}}{\sum_{k=1}^{10} 50} \simeq \text{FWHM} \simeq 0.1\ \mu\text{m} \tag{4.39}$$

This noise value corresponds to the ideal case for of a modulation pattern perfectly centered on the star, with a fixed star, and a centro-symmetric injection function. If the star is moving, if the injection function is not symmetric, and if the modulation pattern is not centered on the star, the barycenter estimate can be biased, and higher noise values for the measurements can lead to better performance of the Kalman filter. The exact value of the matrix R needs to be adjusted empirically, but reasonable values could be:

$$R = \begin{pmatrix} 0.01 \text{ to } 1\ \mu\text{m}^2 & 0\ \mu\text{m}^2 \\ 0\ \mu\text{m}^2 & 0.01 \text{ to } 1\ \mu\text{m}^2 \end{pmatrix} \tag{4.40}$$

4.4.2 Propagation matrix and process noise

The propagation matrix A, which is used by the Kalman filter to predict the evolution of the system, needs to be carefully selected. Ideally, this matrix should reflect a physical understanding of the system. However, in the case of the PicSat instrument, variations of the state vector are induced by the ADCS residual pointing noise, which is largely unknown. The Hyperion iADCS100 has never flown before the PicSat mission, and we do not have access to on-sky data.

In this situation, the best way to proceed is to use an empirical model, which can be adjusted during the mission. Based on the simulated ADCS presented in Section 4.2, and especially from the power spectral density of the residual pointing errors presented in Figure C.2, a second order model with damping seems adequate.

Such a model has been proposed for the Kalman filter used to operate the GRAVITY Fringe Tracker [Menu et al., 2012]. It can be represented using an auto-regressive model of second order:

$$x_{k+1} = a_1 x_k - a_2 x_{k-1} + \epsilon \tag{4.41}$$

With ϵ_k a random noise, and a_1, a_2 two coefficients defined from the damping coefficient ξ and the cutoff frequency f_0 by:

$$a_1 = 2\exp\left(-2\pi\xi f_0 dt\right)\cos\left(2\pi f_0 dt\sqrt{1-\xi^2}\right) \tag{4.42}$$

$$a_2 = \exp\left(-4\pi\xi f_0 dt\right) \tag{4.43}$$

Equation 4.41 for the second order auto-regressive model can be adapted to the state vector representation used for the PicSat instrument by writing:

$$\begin{aligned} x_{k+1} &= a_1 x_k - a_2\left(x_k - v_{xk} dt\right) \\ &= (a_1 - a_2) x_k + a_2 v_{x,k} dt \end{aligned}$$

The equation for the velocity can be written:

$$\begin{aligned} v_{x,k+1} dt &\simeq x_{k+1} - x_k \\ &= (a_1 - a_2 - 1) x_{k+1} + a_2 v_{x,k} dt \end{aligned}$$

In the state vector representation selected for the PicSat Kalman filter, the second order model leads to a propagation matrix A given by:

$$A = \begin{pmatrix} a_1 - a_2 & 0 & a_2 & 0 \\ 0 & a1 - a2 & 0 & a_2 \\ a_1 - a_2 - 1 & 0 & a_2 & 0 \\ 0 & a1 - a2 - 1 & 0 & a_2 \end{pmatrix} \tag{4.44}$$

Together with the propagation matrix, it is also necessary to select a process noise covariance matrix, used to reflect the fact the the model is not perfect. This corresponds to the matrix Q introduced in the Kalman filter propagation equation (Equation 4.30). Combining the propagation matrix and the process noise covariance matrix, it is possible to simulate the power spectral density generated by the model, by calculating a series of state vectors using:

$$X_0 = 0 \tag{4.45}$$

$$X_{k+1} = A X_{k+1} + \mathcal{N}(0, Q) \tag{4.46}$$

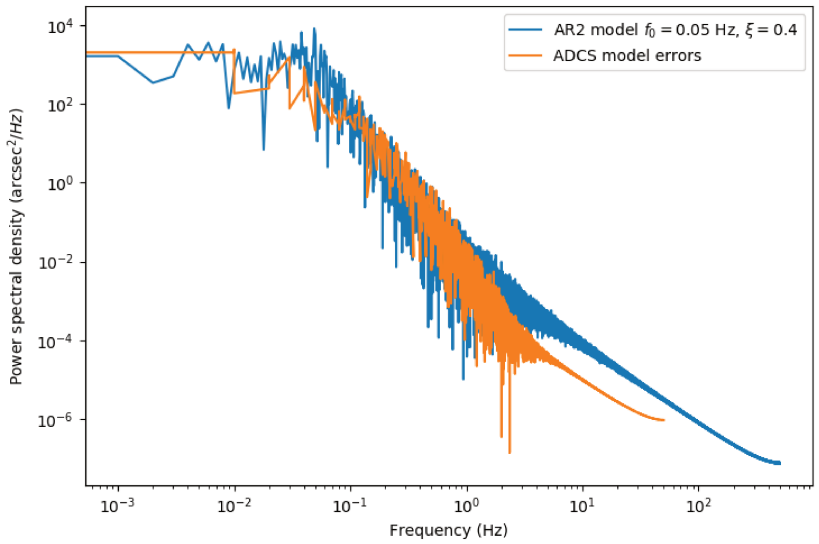


Figure 4.8: Comparison between the power spectral density of an autoregressive model of order 2 (AR2, $f_0 = 0.05$ Hz, $\xi = 0.4$), and the power spectral density of the residual pointing errors generated by the ADCS model, presented in Section 4.2 (x-axis only, normalized to a standard deviation of 10 arcsec).

A propagation matrix corresponding to a frequency $f_0 = 0.05$ Hz, a damping coefficient $\xi = 0.4$, and a process noise covariance matrix given by:

$$Q = \begin{pmatrix} 1\ \text{as}^2 & 0 & 0 & 0 \\ 0 & 1\ \text{as}^2 & 0 & 0 \\ 0 & 0 & 1\ \text{as}^2 & 0 \\ 0 & 0 & 0 & 1\ \text{as}^2 \end{pmatrix} \tag{4.47}$$

leads to a power spectral for the two first components of the state vector (star position) close to what the ADCS model generates (see Figure 4.8).

The two matrices A and Q can be adjusted empirically to get the best performance out of the Kalman filter. Generally speaking, the propagation matrix A controls the shape of the power spectral density, whereas the covariance matrix Q controls the level of noise (i.e. the overall position of the power spectral density on the y axis).

4.4.3 MATLAB/SIMULINK implementation

To test the performance of the photometric barycenter/Kalman filter combination for tracking the position of the star, I implemented a control loop based on this algorithm in MATLAB/SIMULINK. An overview of this model is given in Figure 4.9.

In this model, a central block is used to simulate the instrumental response. This block is fed by a set of parameters used to describe the instrumental PSF (maximum flux, FWHM in two directions to simulate a possible astigmatism, and orientation of the PSF in the focal plane). These parameters can be set to their theoretical values, and time variations can be added to simulate the effects of the environment (changes in the thermal environment can impact the shape of the PSF). From the current position of the fiber and the values of these

PSF parameters, the instrumental response block simulates integrations on the photodiode which include photon noise.

The photometric measurements generated by the instrument simulator are sent to the barycenter block which also uses the fiber position values to compute the barycenter position. This position goes into the Kalman filter, which estimates the position of the star. This position is used as the setpoint for the new fiber position.

The modulation frequency used is 100 Hz, which is well within the linear response range of the piezo, and the typical modulation radius is < 2 μm, too small for the hysteresis to have any impact. Thus, the response of the piezo is assumed to be perfect in this model, with no delay or hysteresis.

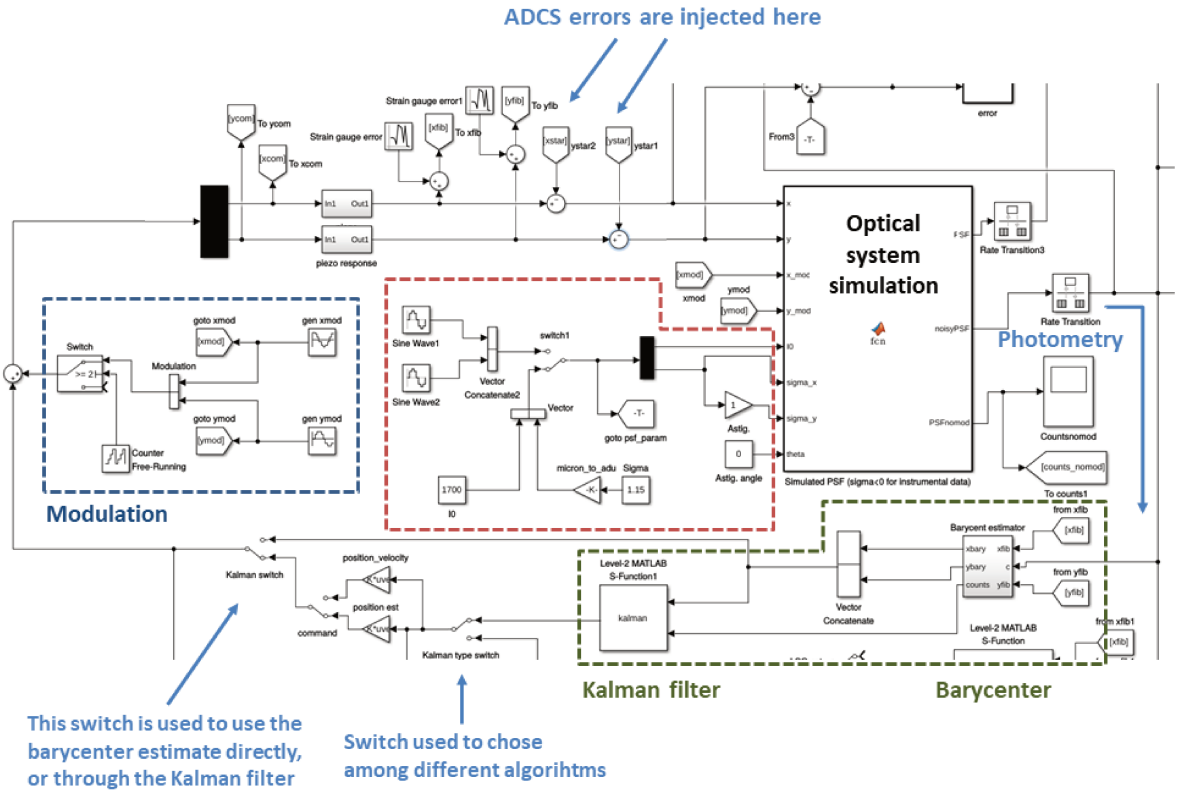


Figure 4.9: Annotated view of the MATLAB/SIMULINK model developed to test the PicSat control loop. This picture only shows the central part of the model, which includes the modulation, the instrument simulator, and the barycenter/Kalman filter. The complete model contains other blocks which will be described later in this Chapter.

4.4.4 Validation of the MATLAB/SIMULINK model

The MATLAB/SIMULINK model was validated on the testbench. Harmonic errors (pure sinewaves) were injected using the tip/tilt mirror, and the experimental results were compared to the prediction of the MATLAB/SIMULINK model. Figure 4.10 gives an example of the results obtained, in terms of photometric precision achieved when the system was submitted to sinewave errors of 30 μm amplitude (peak-to-peak), with frequencies ranging from

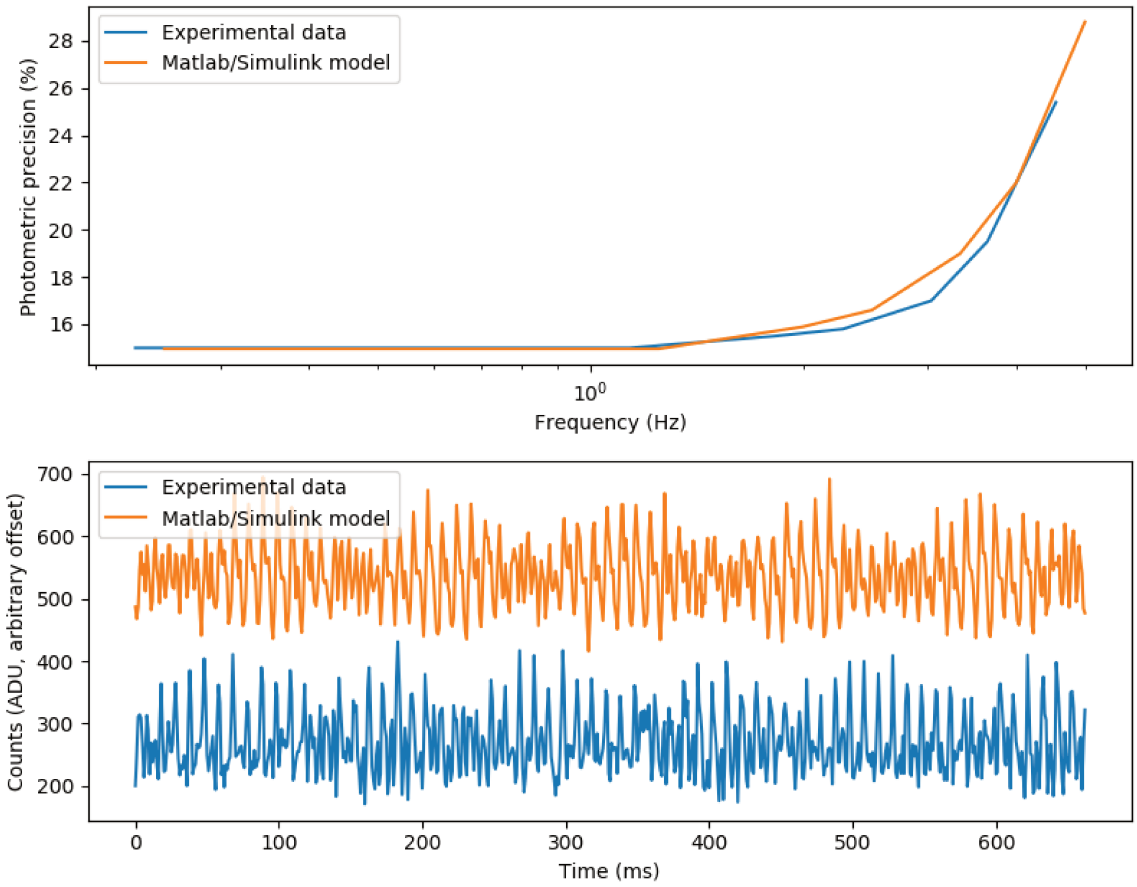


Figure 4.10: Validation of the MATLAB/SIMULINK model of the barycenter Kalman filter. The upper panel gives the evolution of the photometric precision achieved when submitting the system to a pure sinewave jitter (peak-to-peak amplitude of 30 μm), as predicted by the model, and as obtained experimentally using the testbench. The lower panel gives an example of a time series obtained using the model and the testbench, for a frequency of 5 Hz.

0.25 Hz to 5 Hz.

4.4.5 Simulation results

One of the main advantages of using simulations to study the performance of the tracking algorithm is that it is possible to separate the photon noise from the pointing noise. The instrument simulation block used in the model presented in Figure 4.9 has multiple outputs: one of them generates photometric measurements affected by photon noise, while another one generates the same photometric measurements without the photon noise. The tracking loop is fed by the measurements affected by photon noise, as it would be in the real world. But the other output, free of photon noise, can be used to determine the impact of the tracking errors only on the photometry.

The modulation injected is a circular modulation of frequency $f_{\mathrm{mod}} = 100$ Hz, and radius $r_{\mathrm{mod}} = 1.35$ μm (half the FWHM of the injection function). In practice, this means that the modulation added to the Kalman estimate of the star position oscillates between 10 different

values[3]. These values are given by:

$$\begin{cases} x_{\text{mod}} = r_{\text{mod}} \cos\left(\frac{2k\pi}{10}\right) \\ y_{\text{mod}} = r_{\text{mod}} \sin\left(\frac{2k\pi}{10}\right) \end{cases} \qquad k \in \{0 \dots 9\} \tag{4.48}$$

The results obtained with this modulation pattern and the barycenter/Kalman filter to control the position of the fiber are given in Figure 4.11, 4.12, and 4.13.

The Kalman filter control loop has no problem tracking the star in the two first scenarios, but the precision is not sufficient to reach the target 13% injection stability at 1 kHz. In the third scenario (ASTERIA errors, scaled up to 10 arcsec precision), the situation is even worse, with a very unstable photometry (mean value: 77 ADU, standard-deviation: 35 ADU). Although this algorithm can ensure that some light always reaches the detector, which may be useful for some applications, it is not good enough to allow for precision photometric observations.

4.5 Living the dangerous life: extended Kalman algorithm

4.5.1 Limits of the barycenter measurement

One of the biggest problems of the linear Kalman filter wich uses a photometric barycenter as a position estimate is its lack of reactivity, which makes it sensitive to high velocities of the star. Estimating the position of the star using a photometric barycenter of the points acquired on the modulation always gives a good estimate of the direction in which the star is moving, which makes the algorithm relatively robust. Even if the tracking performance is not great, at least the algorithm always goes in the "right direction". The downside of the barycenter measurement is that it is strongly biased toward the natural center of the geometric modulation pattern.

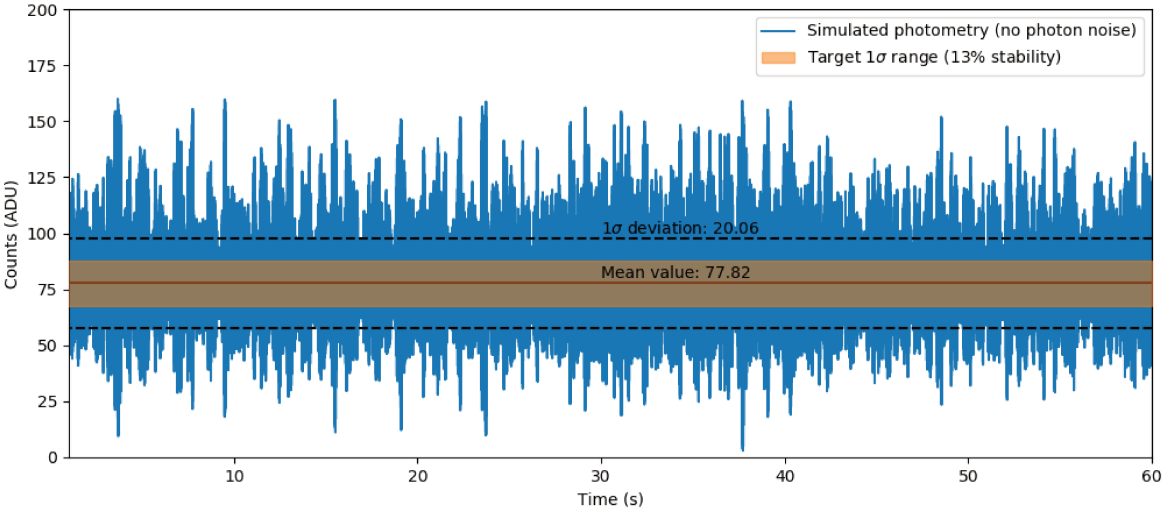
Figure 4.14 illustrates a situation in which the star is located at about 0.8 as away from the reference position estimated by the Kalman filter (origin of the reference frame). The Kalman filter takes 10 measurements along the modulation pattern, and combines them by computing the barycenter. The result is the new measurement of the star position, which is off by about 0.3 as. Even in this noise-free situation, the barycenter estimate of the star position is not good: the estimator is biased.

A possible solution for this problem is to change the way the photometric points are combined to give a measurement of the position of the star. The injection function is nearly Gaussian, and a photometric measurement gives a count value c described by:

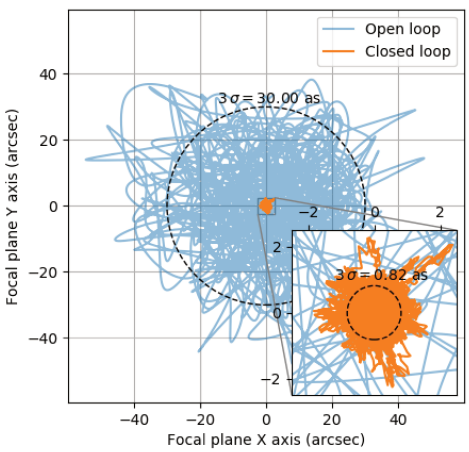
$$c(x_{\text{fib}}, y_{\text{fib}}, x_{\text{star}}, y_{\text{star}}) = c_{\max} \exp\left(-\frac{(x_{\text{fib}} - x_{\text{star}})^2 + (y_{\text{fib}} - y_{\text{star}})^2}{2\sigma^2}\right) \tag{4.49}$$

If the shape of the instrumental injection function is known (i.e. the parameters $c_{\max}$ and σ), this measurement can be inverted to give the value of the distance between the fiber positon and the star. A single measurement thus gives a circle on which the star is located, and multiple measurements along the modulation pattern can be used to trilaterate the position of the star (see Figure D.6). The main difficulty of this approach is that the calculations are difficult, especially if photon noise needs to be taken into account.

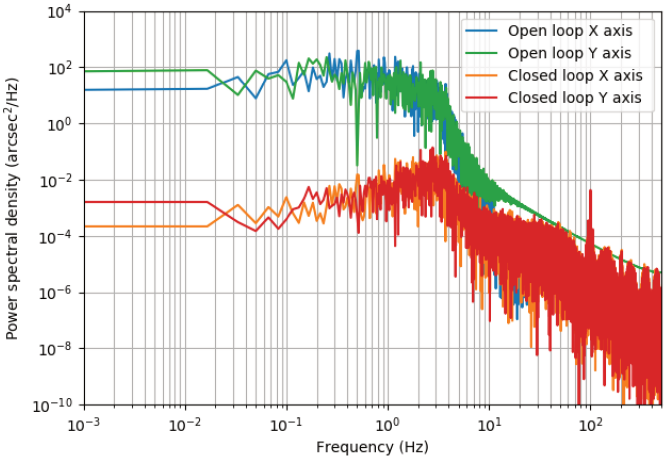
[3] Having an integration frequency which is an integer multiple of the modulation frequency guarantees that the modulation will cycle between a number of points given by the ratio of the two frequencies. This is particularly interesting, as it allows to pre-compute the modulation values before the algorithm starts, and reduces the calculation burden on the payload microcontroller.



(a) Photometry

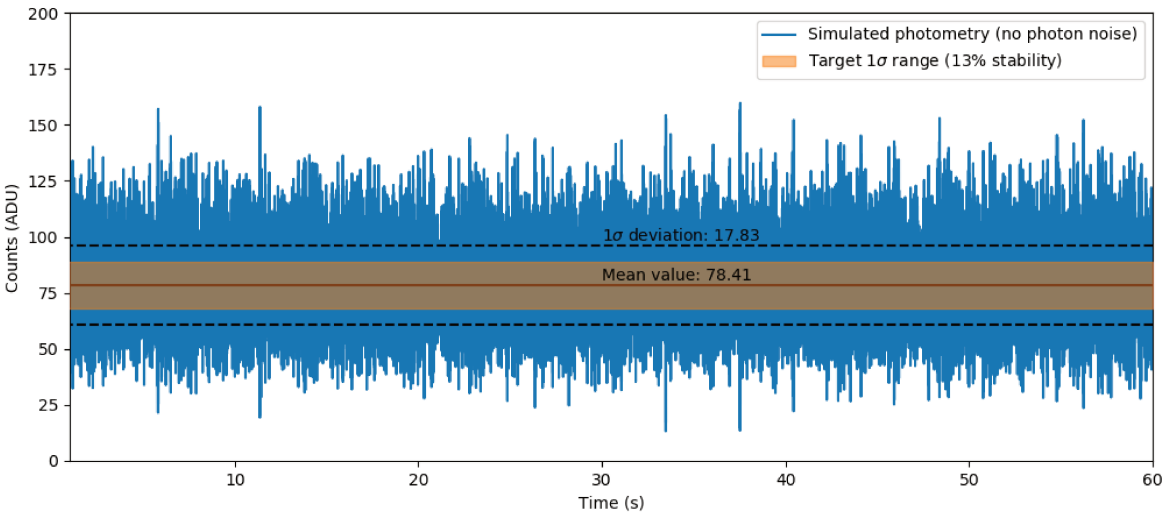


(b) Residual errors

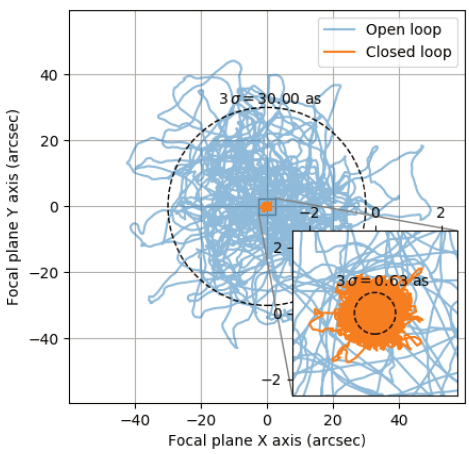


(c) Error spectral densities

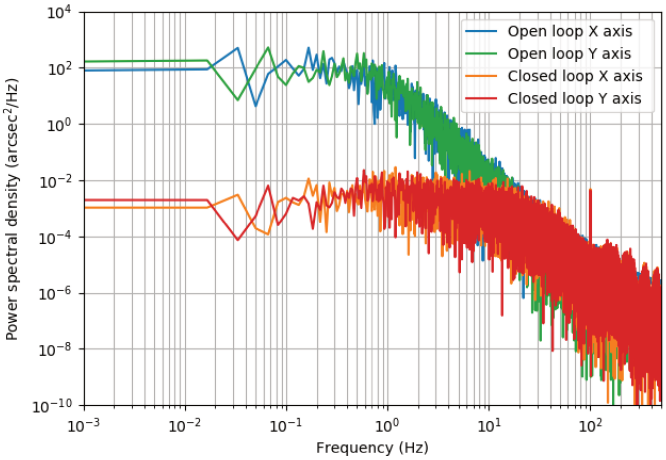
Figure 4.11: Results obtained with the MATLAB/SIMULINK model for the first test scenario (simple ADCS model). The input errors (open loop) have been normalized to a standard deviation of 10 as, and the maximum injection is 160 ADU (per ms). In this simulation, the modulation used is a circular modulation running at 100 Hz, with a radius of 1.35 μm corresponding to half the FWHM of the injection function. The expected number of ADU per integration is 80. Panel 4.11a gives the sequence of photometric points obtained during 1 min of simulation, ignoring the first 1 s (convergence phase of the Kalman filter). Panel 4.11b gives an overview of the star displacement in the focal plane (blue curve, open loop), compared to the residual errors downstream of the control loop (orange, closed loop). Panel 4.11c gives the power spectral densities of the pointing errors, in open and closed loops, for the two axes.



(a) Photometry

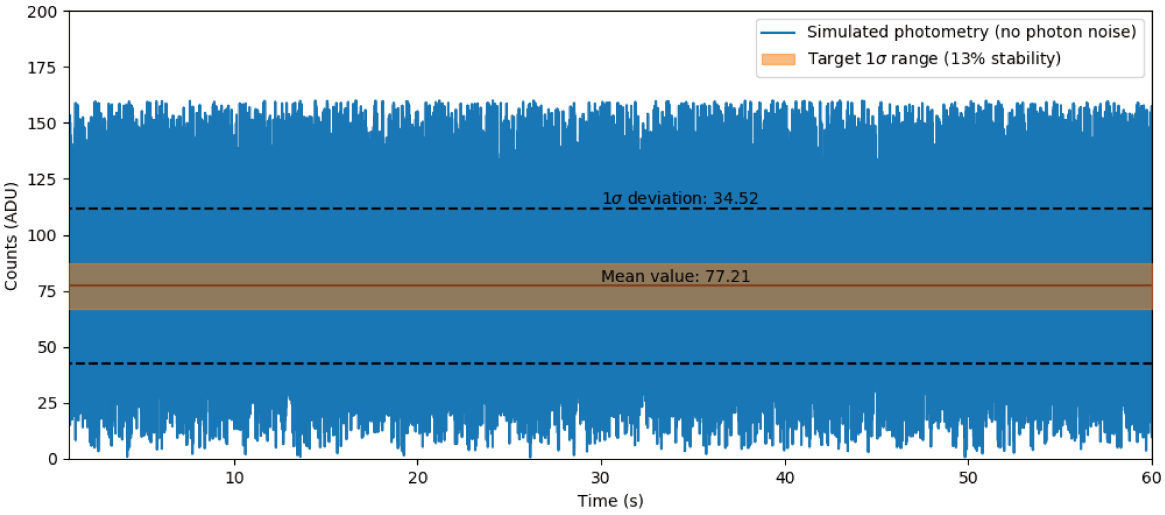


(b) Residual errors

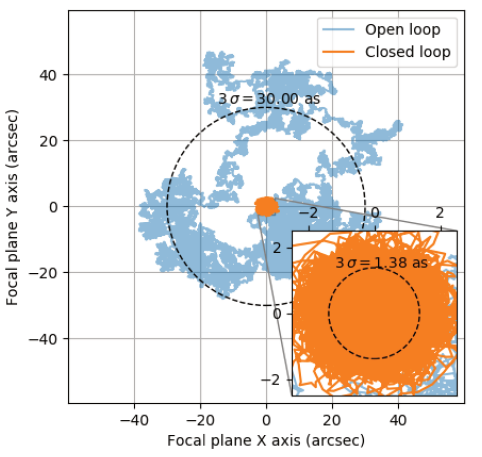


(c) Error spectral densities

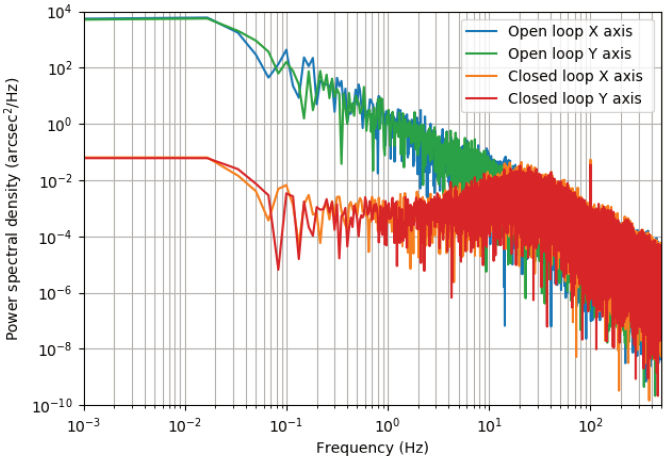
Figure 4.12: Same as Figure 4.11 for the second scenario, in which the injected ADCS errors are simulated using a second order low-pass filter with a cutoff frequency of 1 Hz. The injected pointing errors have been normalized to a standard deviation of 10 as.



(a) Photometry



(b) Residual errors



(c) Error spectral densities

Figure 4.13: Same as Figure 4.11 for the third scenario, in which the injected ADCS errors comes for the in-flight data of the ASTERIA mission, scaled up to a $1\,\sigma$ pointing error of 10 as.

A faster version of this same concept can be derived by *linearizing* the trilateration. If we consider that the location of the star is "not too far from the Kalman estimate" (whatever that means), Equation 4.49 can be linearized in $(x_{\text{star}}, y_{\text{star}})$. In the example considered here, the origin of the reference frame is taken at the Kalman estimate, and the linearized version of Equation 4.49 reads:

$$c(x_{\text{fib}}, y_{\text{fib}}, x_{\text{star}}, y_{\text{star}}) \simeq c(x_{\text{fib}}, y_{\text{fib}}, 0, 0) \left[1 + \frac{x_{\text{star}}\, x_{\text{fib}}}{\sigma^2} + \frac{y_{\text{star}}\, y_{\text{fib}}}{\sigma^2}\right] \tag{4.50}$$

If the fiber stays on the 0.5 FWHM circle, we have $c(x_{\text{fib}}, y_{\text{fib}}, 0, 0) = c_{\max}/2$, and thus:

$$\frac{x_{\text{star}}\, x_{\text{fib}}}{\sigma^2} + \frac{y_{\text{star}}\, y_{\text{fib}}}{\sigma^2} = \frac{c(x_{\text{fib}}, y_{\text{fib}}, x_{\text{star}}, y_{\text{star}})}{c_{\max}} - 1 \tag{4.51}$$

in which we recognized the equation of a line. In this linearized version of the star position trilateration, each measurement is converted to a line on which the star is supposed to be. These measurements can be combined to estimate the position of the star by finding the point which minimizes the sum of the squared distance to all lines (potentially weighting the different lines according to the photon noise). Since the problem is linear, the least-square solution has an explicit closed-form.

Figure 4.14b illustrates the concept of linearized trilateration, and shows that the resulting estimate of the star position is improved over the barycenter approach.

4.5.2 Linearized non-linear innovation

In the previous subsection, I proposed an intuitive explanation of how the control loop can be improved by switching to a different type of measurement. The formalism of Kalman filters actually provides a direct way to implement this approach, through what is known as an extended Kalman filter.

Extended Kalman filters were developped as a way to control non-linear systems. In the case of the PicSat instrument, the non-linearity comes from the measurements (Equation 4.49), which cannot be described as a linear operation on the state vector ($Z = HX$, as in Equation 4.23). But if we assume that the *a priori* estimate of the state vector is close to the real one, the measurement equation can be linearized around the *a priori* estimate. Denoting $Z_{k+1} = Z(X_{k+1})$ the non-linear measurement made at iteration $k+1$, when the state of the system is X_{k+1}, we can write:

$$Z(X_{k+1}) = Z(\hat{X}^{\text{ap}}_{k+1} + X_{k+1} - \hat{X}^{\text{ap}}_{k+1}) \tag{4.52}$$

The general linearized version of this equation reads[4]:

$$Z(X_{k+1}) = Z(\hat{X}^{\text{ap}}_{k+1}) + \left[\frac{\partial Z}{\partial X}\left(\hat{X}^{\text{ap}}_{k+1}\right)\right]\left[\hat{X}^{\text{ap}}_{k+1}\right] \tag{4.53}$$

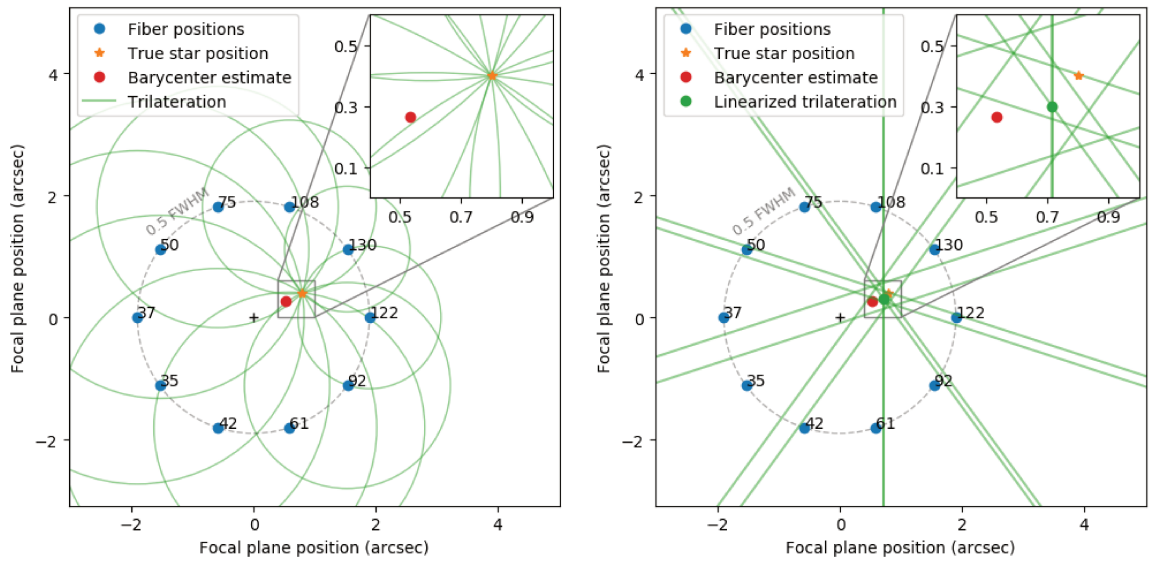
The expected measurement at iteration $k+1$ is simply given by:

$$Y_{k+1} = Z_{k+1}(\hat{X}^{\text{ap}}_{k+1}) \tag{4.54}$$

Which means that the difference between the real and the expected measurements is given by:

$$Z_{k+1} - Y_{k+1} = \left[\frac{\partial Z}{\partial X}\left(\hat{X}^{\text{ap}}_{k+1}\right)\right]\left[X_{k+1} - \hat{X}^{\text{ap}}_{k+1}\right] \tag{4.55}$$

[4] Again, this is the equation of a hyperplane, in the system's state space.



(a) Trilateration

(b) Linearized trilateration

Figure 4.14: Illustration of the trilateration and linearized trilateration approach to the estimation of the star position. In this illustration, the fiber is moved along a modulation circle around the current best estimate of the star position (taken at the origin of the reference frame), and ten photometric points are acquired. The true location of the star is set at $x = 0.8, y = 0.2$ (orange star), and the results of the measurements are given next to each blue point (no photon noise). In the true trilateration algorithm, each measurement can be inverted into a circle of presumed location for the star, and the intersection point of all circles is the true of location of the star. In the linearized version, each measurement is inverted into a line, not a circle, and the best estimate is chosen to minimize the sum of the squared distances to this line (green dot on the left panel). For each case, the location of the barycenter is also shown as a red dot.

Equation 4.55 has the exact same structure as Equation 4.25, replacing H by the Jacobian of Z. Consequently, the Kalman filter can be extended to directly take into account non-linear measurements by using the linearized innovation equation:

$$H_{k+1} = \frac{\partial Z}{\partial X}\left(\hat{X}^{\text{ap}}_{k+1}\right) \tag{4.56}$$

$$\hat{X}_{k+1} = \hat{X}^{\text{ap}}_{k+1} + K\left(Z_{k+1} - Y_{k+1}\right) \tag{4.57}$$

$$P_{k+1} = \left(I - KH_{k+1}\right) P^{\text{ap}}_{k+1} \tag{4.58}$$

In the case of the PicSat instrument, it should be noted that the function Z actually depends on the iteration number $k+1$, since it depends on the position on the modulation pattern. Denoting $(x_{\text{mod},k}, y_{\text{mod},k})$ the modulation at iteration k, we have:

$$H_k = \begin{pmatrix} \frac{\partial \eta}{\partial x}(x_{\text{mod},k}, y_{\text{mod},k}) & 0 & 0 & 0 \\ 0 & \frac{\partial \eta}{\partial y}(x_{\text{mod},k}, y_{\text{mod},k}) & 0 & 0 \end{pmatrix} \tag{4.59}$$

with $\eta(x, y)$ the injection function. The expected measurement is given by:

$$Y_k = \eta(x_{\text{mod},k}, y_{\text{mod},k}) \tag{4.60}$$

And the measurement noise matrix can be defined to take into account the photon noise, and the impact of pointing errors on the photometry:

$$R_k = \left(Z_k + (0.13 \times Y_k)^2\right) \tag{4.61}$$

The downside of the extended version of the Kalman fiter for the PicSat instrument is that it requires a model of the instrumental response η. Ideally, the gaussian injection function, normalized to the proper intentisty (which depends on the target star and on the integration time) should be used. However, the instrumental response can vary due to environmental effects, and a soft-real time feedback of the estimated shape of the instrumental response might be necessary. This can be provided by an algorithm running on the on-board computer.

4.5.3 MATLAB/SIMULINK implementation

The MATLAB/SIMULINK implementation of the extended Kalman filter is straightforward, and is presented in Figure 4.15. The entire algorithm is implemented in a single Level-2 S-Function block. The filter does not work in physical units (microns or arcseconds), but in ADU, corresponding to the piezo command space (0-65535 to represent the entire piezo range). The block is fed by the fiber position (in ADU), a measurement of the star velocity from the gyroscope (converted to ADU/iteration), the photometric measurements, and a set of instrumental parameters used to describe the instrumental response.

Due to the fact that the gyroscope is embedded on the payload board, separated from the optical system, the accuracy of its measurements in space is unknown. Thermal distorsion of the satellite can bias the gyroscopic measurements. Consequently, the MATLAB/SIMULINK implementation of the extended Kalman filter can be set to use or discard these measurements, to test the two scenarios. In practice, the two versions (with and without the gyro measurements) have been implemented in the payload flight software.

To simplify all calculations in the extended Kalman filter, the instrumental response is modeled using an astigmatic gaussian function:

$$\eta(x, y) = \eta_0 \exp\left\{ -\frac{\left(\cos(\theta)x - \sin(\theta)y\right)^2}{2{\sigma_1}^2} - \frac{\left(\cos(\theta)y + \sin(\theta)x\right)^2}{2{\sigma_2}^2} \right\} \tag{4.62}$$

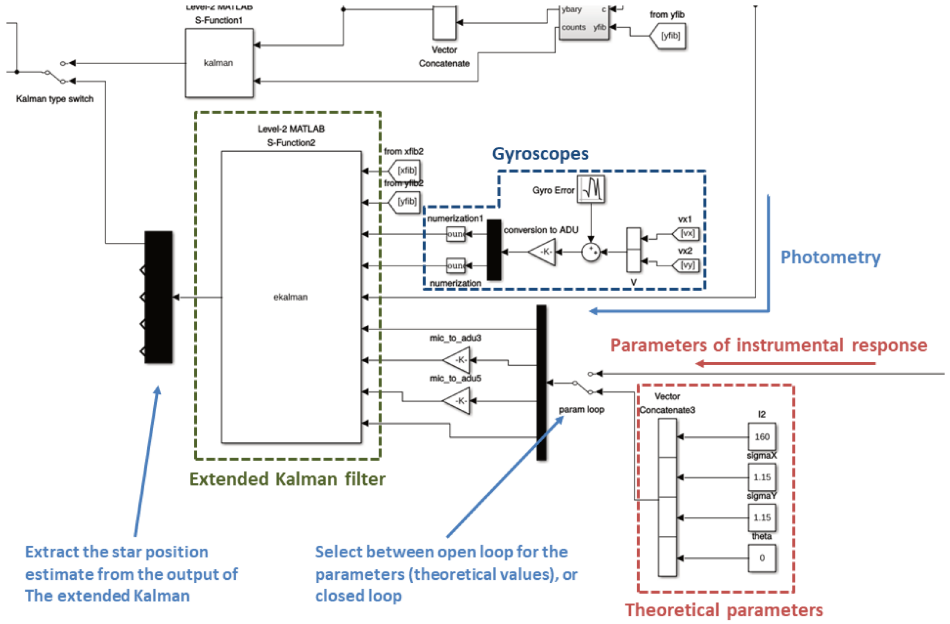


Figure 4.15: Annotated view of the MATLAB/SIMULINK implementation of the extended Kalman control loop. The algorithm is implemented in a Level-2 S-Function block, which uses the fiber position, the photometry, and possibly simulated measurements from the gyroscope. The algorithm also needs an estimate of the shape of the instrumental response, which can either be set to its theoretical value (open loop), or estimated by another block (closed loop, see Figure 4.16).

The three instrumental parameters σ_1, σ_2, and θ are used to describe the shape of the gaussian function, and the last parameter η_0 corresponds to the maximum number of counts.

These four parameters can be either set to their theoretical values (open loop), or estimated in soft-real time by a dedicated least-square algorithm, implemented in another Level-2 S-Function block (Figure 4.16). The instrumental response can be expected to vary over orbital timescales ($\sim$ 90 min), and, in practice, these parameters could be estimated once every $\sim$ 1 min by the satellite's main computer.

4.5.4 Validation of the MATLAB/SIMULINK implementation

In a similar way as for the barycenter based Kalman algorithm, the MATLAB/SIMULINK model of the extended Kalman filter was validated on the testbench, with sinewave errors injected with the tip/tilt mirror. The results of these tests are given in Figure 4.17

4.5.5 Simualtion results

The results of the extended version of the Kalman algorithm, tested in similar conditions as the barycenter Kalman (see Section 4.4.5), are given in Figure 4.18, 4.19, and 4.20. The performance is better than with the barycenter algorithm, and, for the scenarios 1 and 2, the photometric stability is close to the objective (17% versus 13%). However, for the third scenario, the photometry is still too unstable (33%) to guarantee the scientific performance of the mission.

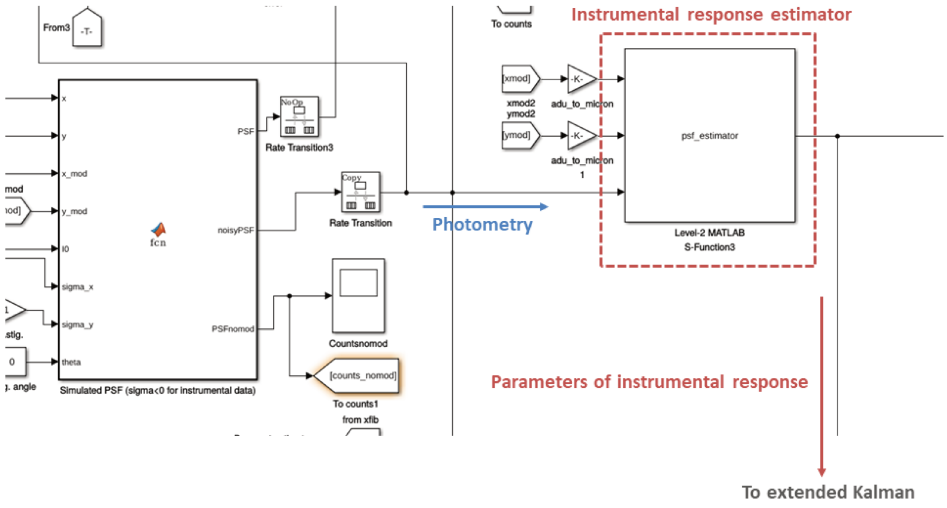


Figure 4.16: Annotated view of a MATLAB/SIMULINK implementation of the parameter estimator used to describe the instrumental response.

4.5.6 A better modulation pattern

One way to improve the results obtained with the control loop is to change the modulation strategy. The photometric variations resulting from a pointing error depends on the magnitude of the error, and on the derivative of the injection function around the modulation points (see Section 4.1.2). The derivative of the gaussian injection function is maxium at $\simeq 0.6\,\mathrm{FWHM}$ from the center. Consequently, using a circular modulation of diameter equals to the FWHM of the injection function makes all photometric measurements very sensitive to tracking errors. From the perspective of the control loop, this means that the measurements which are fed to the Kalman filter are more sensitive to the position of the star, which is a good point. But from a science perspective, this sensitivity has a negative impact on the quality of the data.

The two conflicting requirements can be brought together using a modulation pattern which alternates between parts of the injection function which are close to the center, where the photometry is less sensitive to tracking errors, and parts at 0.5 FWHM, which are more sensitive to the position of the star. The extreme case would be to alternate between positions on the circle of one FWHM diameter, and positions at the center of the injection function (see Figure 4.21a). However, in this case, every other measurements is useless for the tracking loop (the derivative is zero), which means that the Kalman filter will effectively run at 500 Hz instead of 1 kHz. A better alternative is to use a "star" pattern with positions alternating between an inner and an outer circles, as in Figure 4.21b.

Using this modulation pattern yields similar results in terms of pointing stability. But the photometric points obtained can be separated in two sets: the points obtained on the outer modulation circle, and the points obtained on the inner modulation circle. For the same pointing errors, the photometric stability for these two sets can be significantly different.

Of course, filtering half of the points (all points on the outer circle) effectively reduces the integration time by a factor 2, and thus increases the photon noise. But the loss is partly compensated by the better injection obtained on the inner circle, and the gain in terms of photometric stability is substantial.

The results obtained on each scenario are given in Figure 4.22, 4.23, and 4.24. Even taking into account the increased photon noise, the improvement in terms of photometric stability is

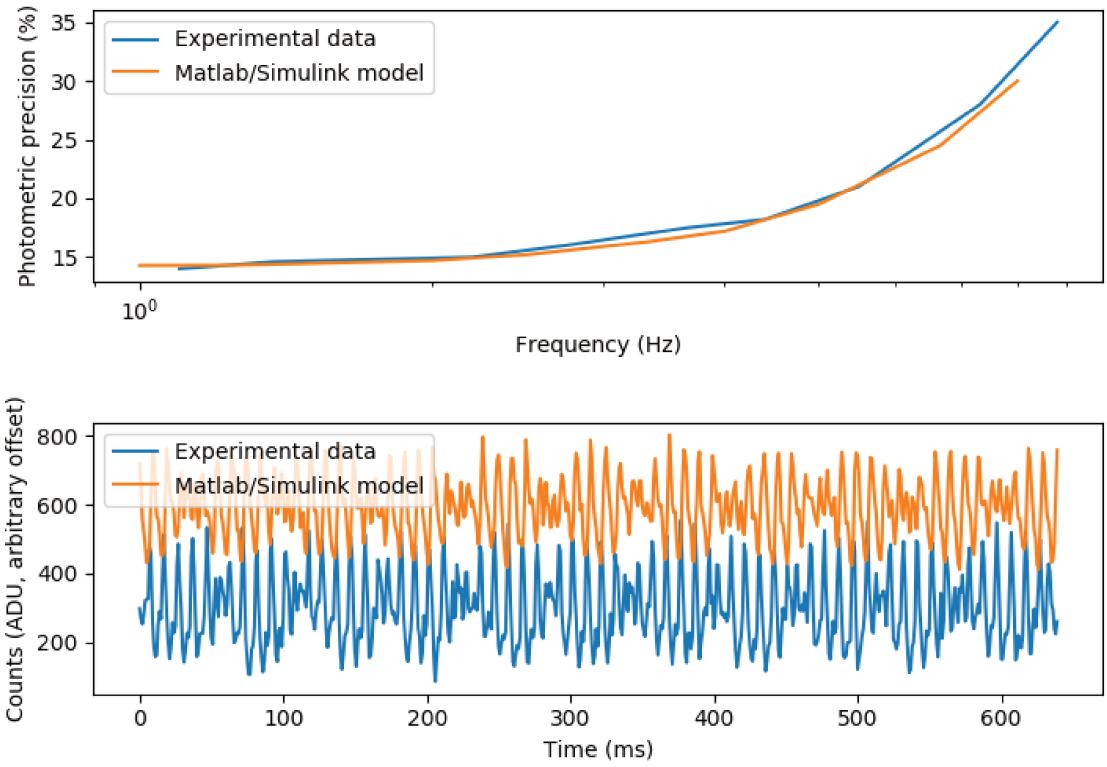
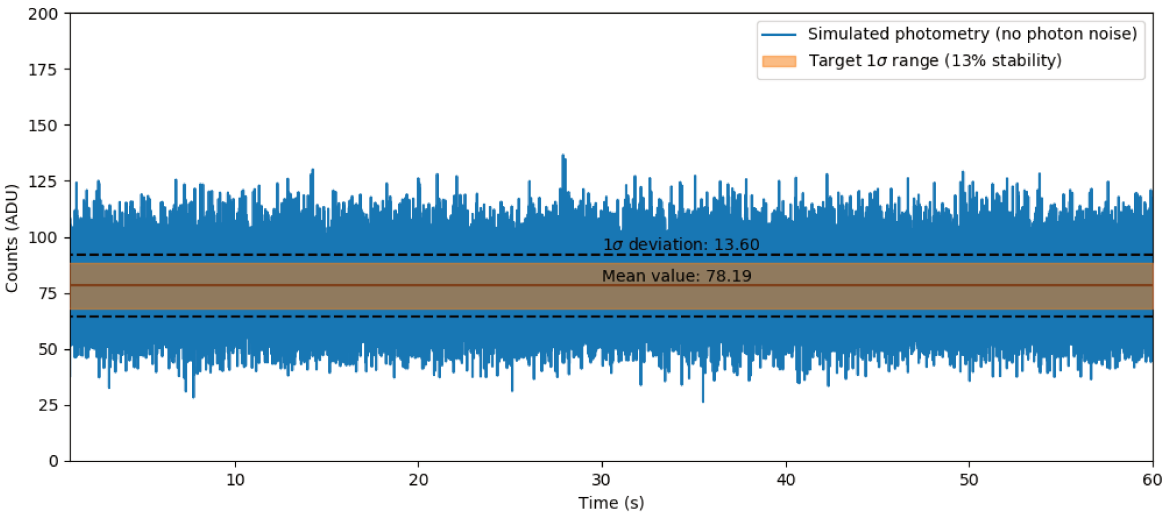
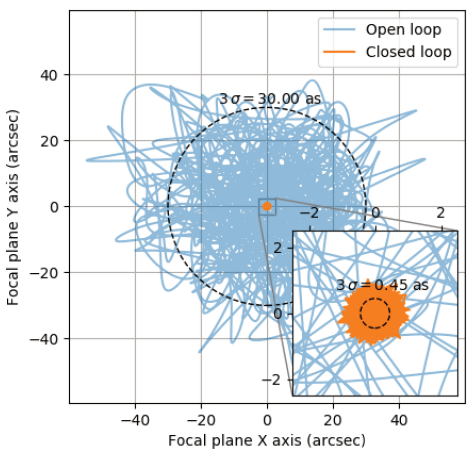


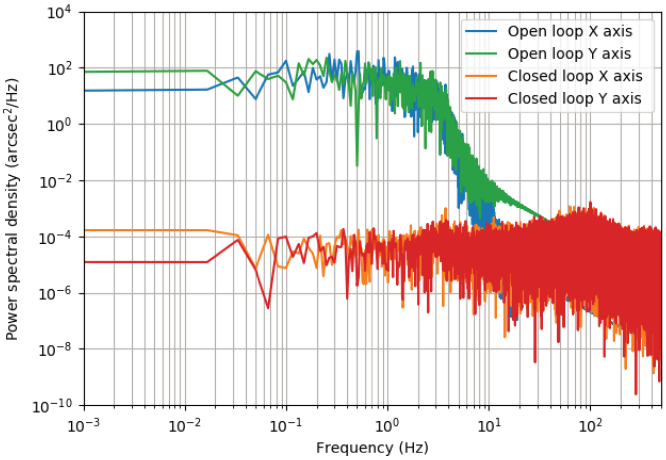
Figure 4.17: Validation of the MATLAB/SIMULINK model of the extended Kalman filter. The upper panel gives the evolution of the photometric precision achieved when submitting the system to a pure sinewave jitter (peak-to-peak amplitude of 30 μm), as predicted by the model, and as obtained experimentally using the testbench. The lower panel gives an example of a time series obtained using the model and the testbench, for a frequency of 8 Hz (the slight frequency offset between the experimental and the simulated curves is due a drift of the RTC of the electronics controlling the tip/tilt mirror, and has been taken into account when plotting the the upper panel figure).



(a) Photometry

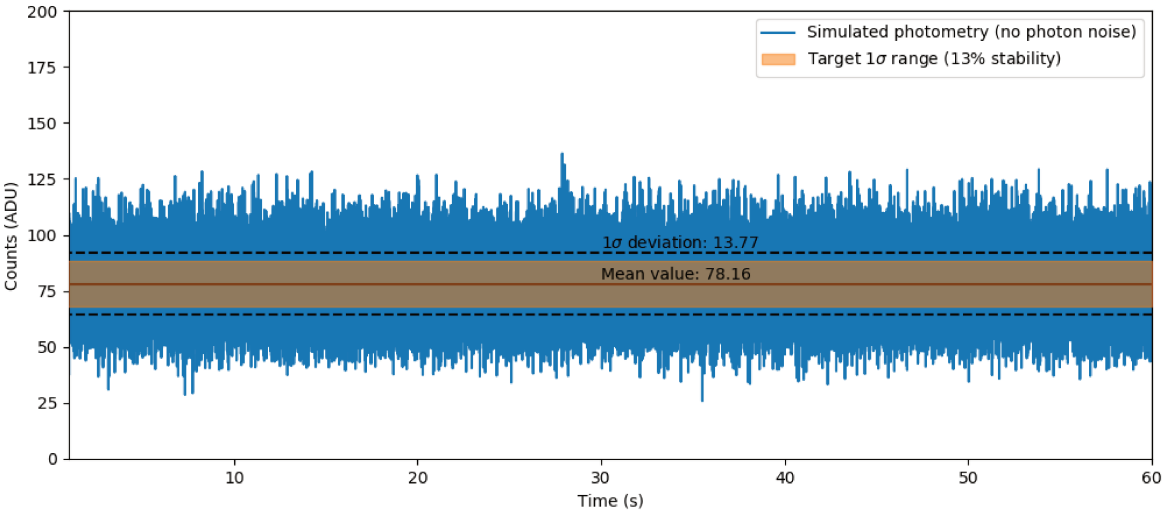


(b) Residual errors

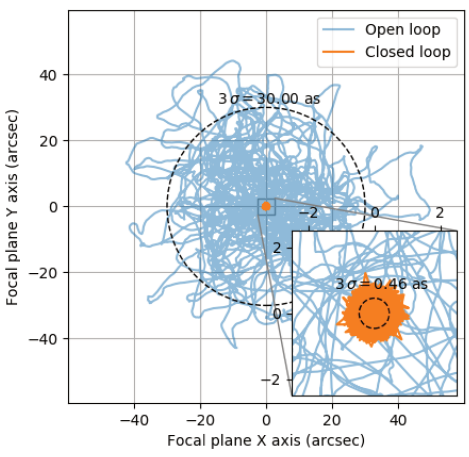


(c) Error spectral densities

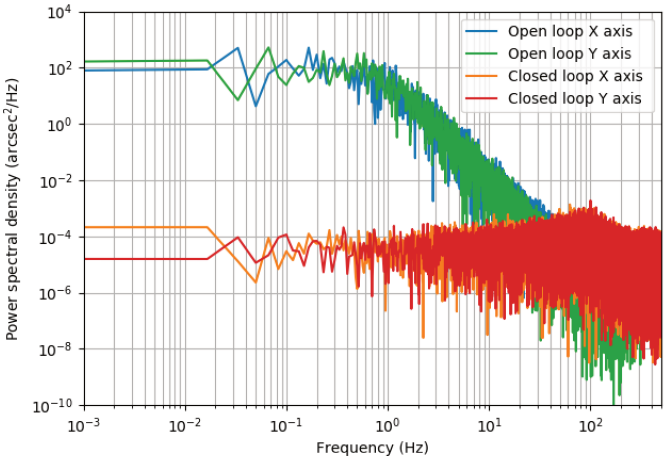
Figure 4.18: Same as Figure 4.11 for the first scenario, in which the injected ADCS errors are simulated using a simple MATLAB/SIMULINK model. The injected pointing errors have been normalized to a standard deviation of 10 as.



(a) Photometry

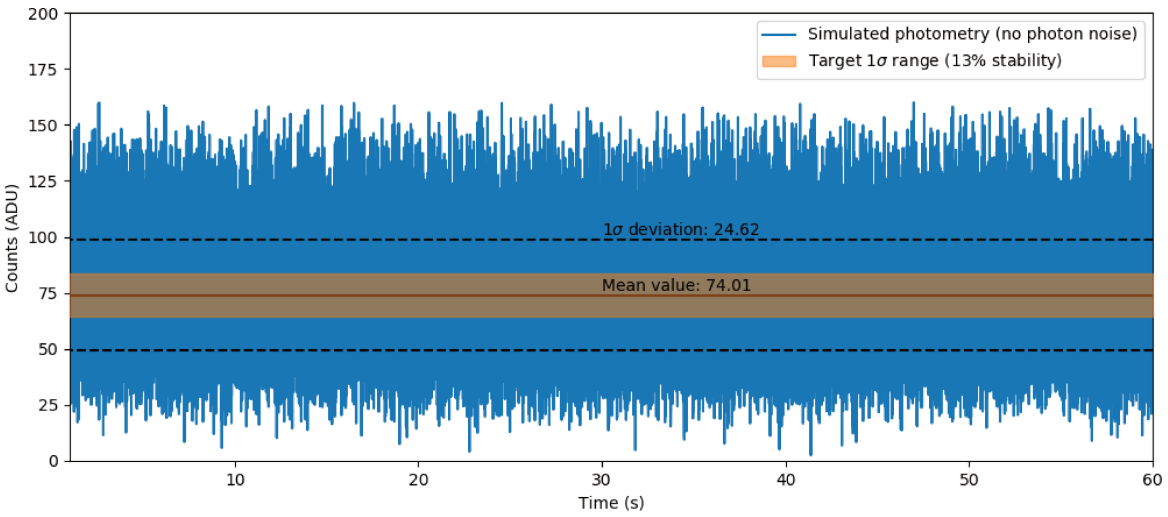


(b) Residual errors

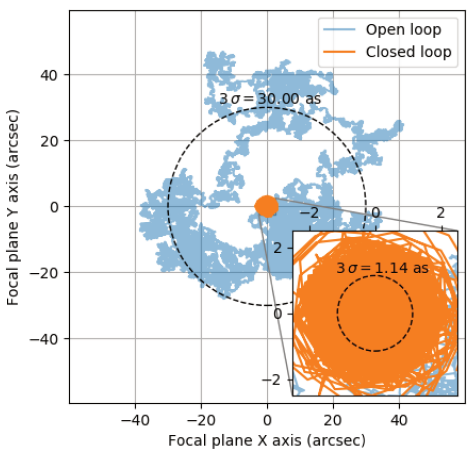


(c) Error spectral densities

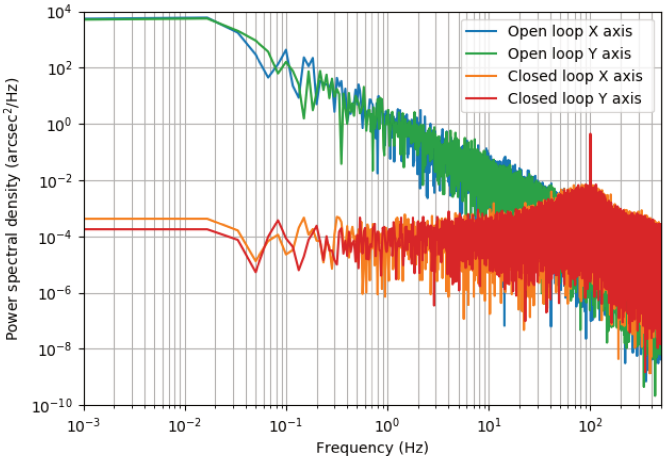
Figure 4.19: Same as Figure 4.11 for the second scenario, in which the injected ADCS errors are simulated using a second order low-pass filter with a cutoff frequency of 1 Hz. The injected pointing errors have been normalized to a standard deviation of 10 as.



(a) Photometry

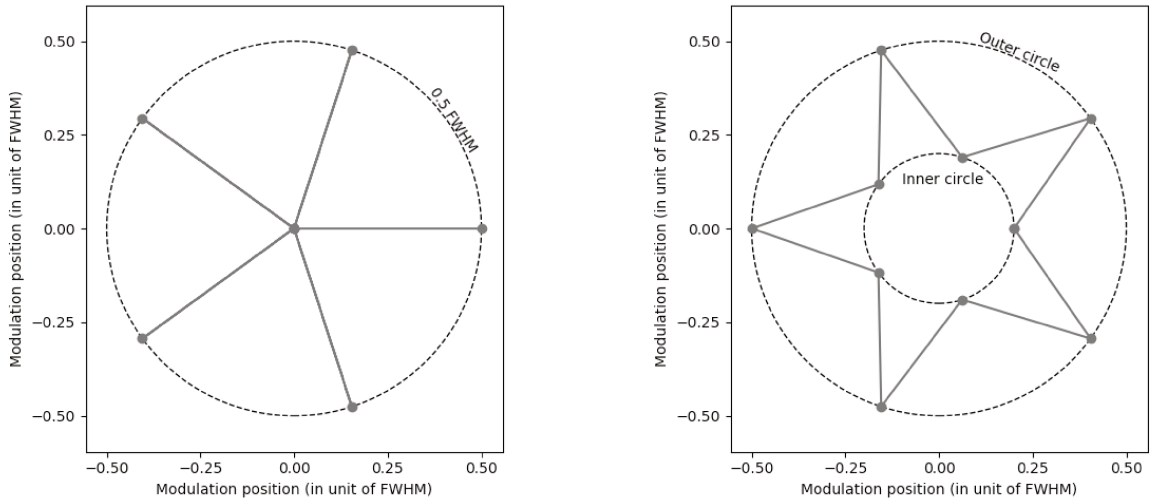


(b) Residual errors



(c) Error spectral densities

Figure 4.20: Same as Figure 4.11 for the third scenario, in which the injected ADCS errors comes for the in-flight data of the ASTERIA mission, scaled up to a $1\,\sigma$ pointing error of 10 as.



(a) Circle-center

(b) Star

Figure 4.21: Two alternative modulation patterns which can be used to improve the performance of the algorithm.

enough to reach the target objective in the case of Scenario 1 and 2. In Scenario 3, though, the photometric variations are still too strong. Further reducing the diameter of the inner circle could potentially help to reach the target stability, but simulations show that it also makes the Kalman filter unstable.

4.5.7 Adding in the gyroscopes

To further improve the performance of the Kalman algorithm, it is also possible to use the data from the gyroscopes in the control loop. The measurements from this sensor can be directly converted to measurements of the velocity of the star in the focal plane, i.e. measurements of the third and fourth components of the state vector.

Denoting $v_{x,\mathrm{gyro}}$ and $v_{y,\mathrm{gyro}}$ the two measurements of the velocity of the star (on the two axes), and $\sigma_{x,\mathrm{gyro}}$, $\sigma_{y,\mathrm{gyro}}$ the associated errors (standard deviations), the new measurement equations H_k are given by:

$$H_k = \begin{pmatrix} \frac{\partial \eta}{\partial x}(x_{\mathrm{mod},k}, y_{\mathrm{mod},k}) & 0 & 0 & 0 \\ 0 & \frac{\partial \eta}{\partial y}(x_{\mathrm{mod},k}, y_{\mathrm{mod},k}) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.63}$$

The effective measurement Z, expected measurement Y, and measurement noise matrices are given by:

$$Z_k = \begin{pmatrix} c_k \\ v_{x,\mathrm{gyro}} \\ v_{y,\mathrm{gyro}} \end{pmatrix} \quad Y_k = \begin{pmatrix} \eta(x_{\mathrm{mod},k}, y_{\mathrm{mod},k}) \\ \hat{v}x, k^{\mathrm{ap}} \\ \hat{v}y, k^{\mathrm{ap}} \end{pmatrix} \quad R_k = \begin{pmatrix} Z_k + (0.13 \times Y_k)^2 & 0 & 0 \\ 0 & {\sigma_{x,\mathrm{gyro}}}^2 & 0 \\ 0 & 0 & {\sigma_{y,\mathrm{gyro}}}^2 \end{pmatrix}$$

Although the measurements from the gyroscopes itself does not really improve the pointing accuracy, it helps to stabilize the control loop, and makes possible to reduce the inner diameter

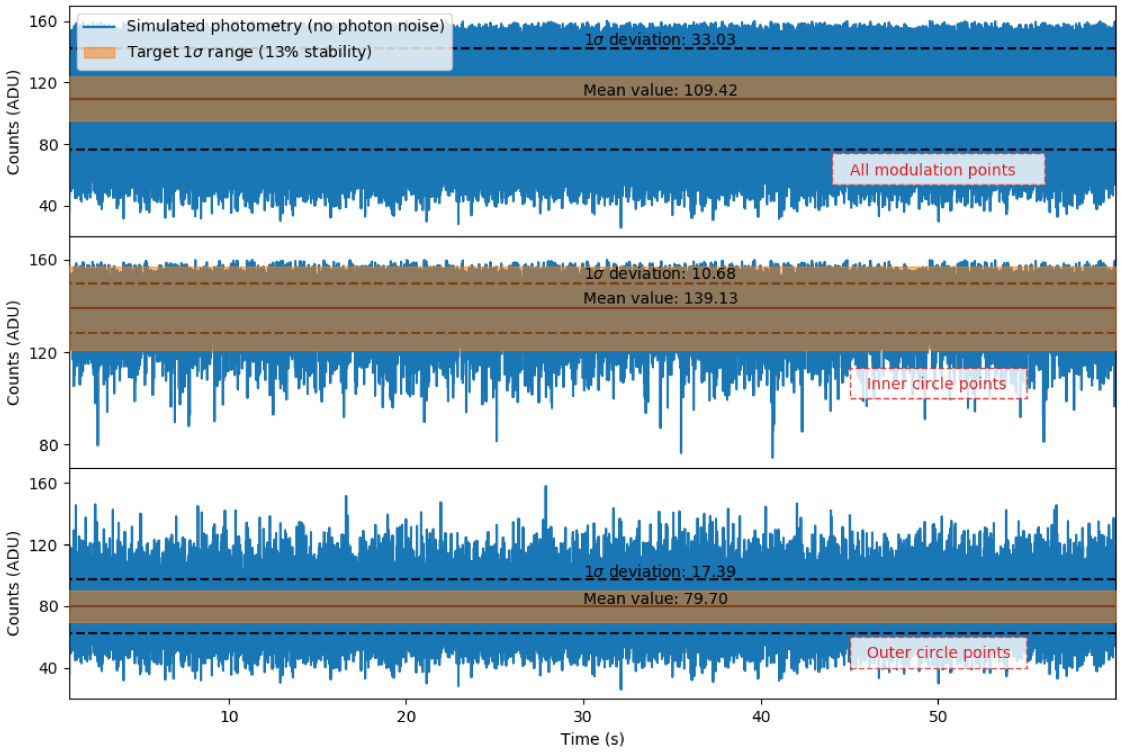


Figure 4.22: Photometry obtained on Scenario 1 (simple ADCS model), using the extended Kalman filter with a star pattern for modulation. The outer circle of the star has a 0.5 FWHM radius, and the inner circle has a 0.2 FWHM radius. The upper panel gives the total photometric sequence, its mean value and standard deviation, and the target 13% stability range, counting all points in the sequence. The middle panel is similar, but counting only the points in the inner circle of the modulation pattern. The bottom panel gives the result for the outer circle only.

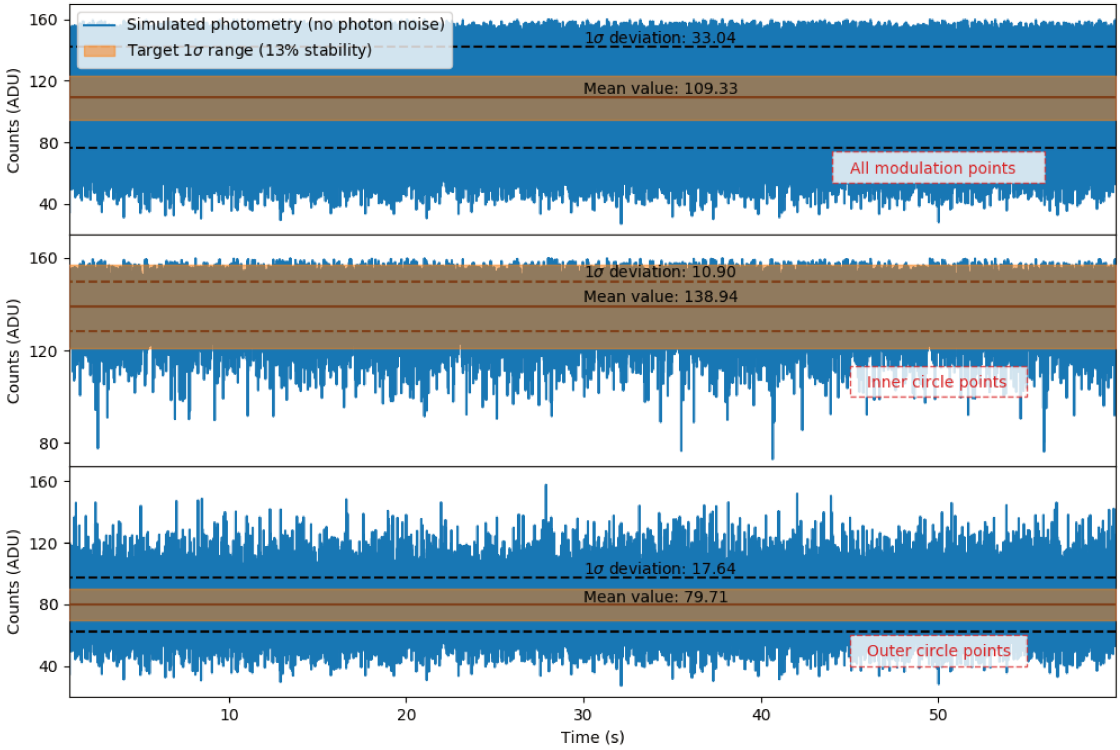


Figure 4.23: Same as Figure 4.22, for Scenario 2.

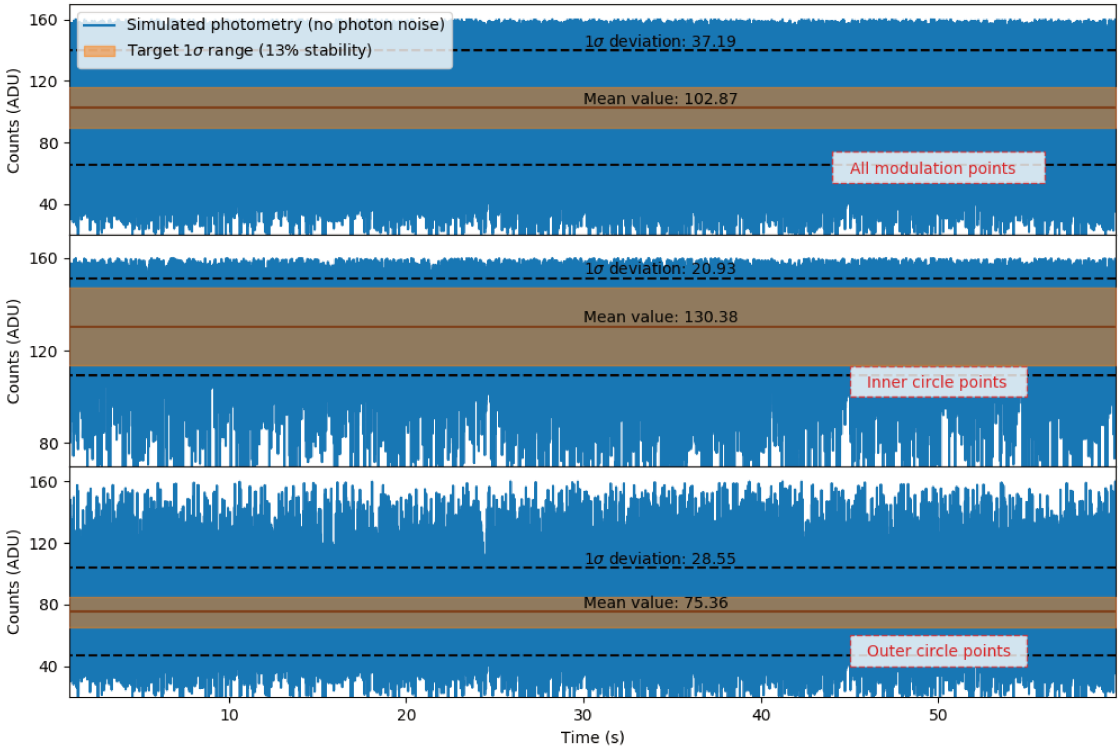


Figure 4.24: Same as Figure 4.22, for Scenario 3.

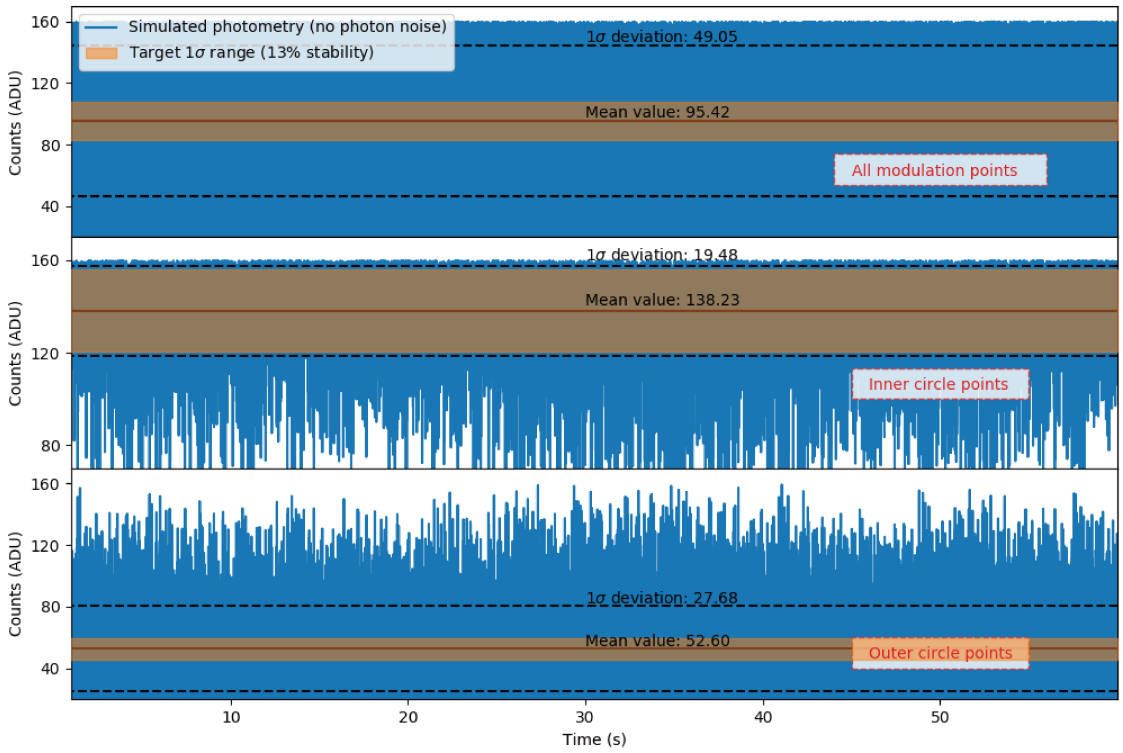


Figure 4.25: Same as Figure 4.22, for Scenario 3 (ASTERIA errors, scaled up to 10 as at $1\,\sigma$, but now using a star modulation pattern with an inner circle radius of 0.1 FWHM, an outer circle radius of 0.5 FWHM, and adding in the measurements from the gyroscope to help stabilize the whole control loop.

of the modulation pattern. The best results on Scenario 3 (ASTERIA) are obtained with an inner diameter to 0.1 FWHM, an outer diameter of 0.5 FWHM, and adjusting the gains of the filter (in particular adjusting the propagation matrix to an AR2 of frequency $f_0 = 3$ Hz and damping $\xi = 0.4$, to have a slightly more agressive filter). The resulting photometry is given in Figure 4.25. The photometric stability on the inner circle of the modulation pattern is 13%.

4.5.8 Instrumental parameter errors

The extended Kalman filter uses a model of the instrumental response to estimate the position of the star in real time. Any error on this model can potentially lead to increased tracking errors. In Figure 4.26, I show the evolution of the pointing precision as a function of the error on the two main parameters describing the instrumental response: the peak intensity, and the FWHM of the gaussian. In order to ensure good performance of the tracking loop, the error on the parameter estimates should be kept $< 10\%$. But it is interesting to note that even with large errors (up to 30% on the FWHM, 50% on the peak intensity), the control loop is still able to track the star.

The MATLAB/SIMULINK model of the control algorithm also implements a closed-loop to estimate in soft-real time the instrumental parameters. The instrumental parameters are estimated at regular intervals Δt, using the following method:

- For each position $x_{k_{\mathrm{mod}}}, y_{k_{\mathrm{mod}}}$ along the modulation pattern, the photometric points

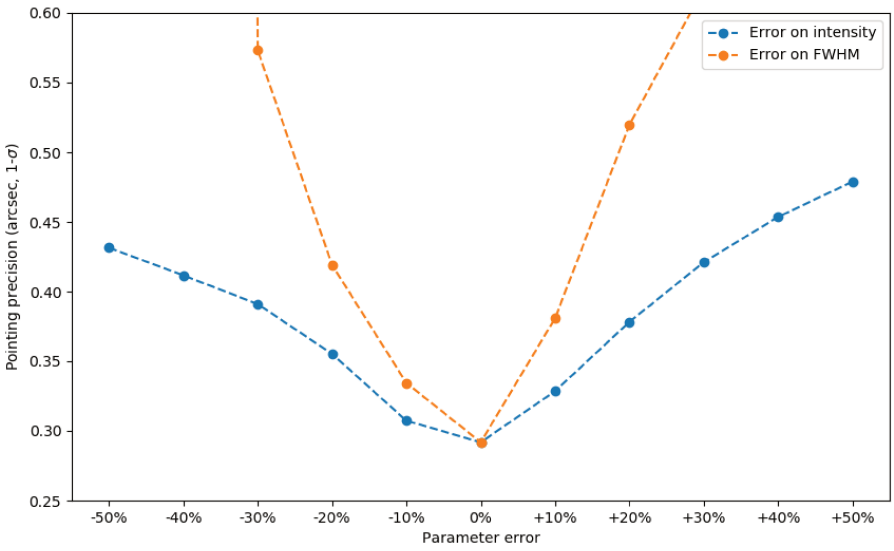


Figure 4.26: Pointing precision of the extended Kalman filter as a function of parameter error, plotted for Scenario 3 (ASTERIA errors), using a circular modulation of radius 0.5 FWHM.

acquired at this position over the period $[t - \Delta t, t]$ are combined to give a mean value $c_{k\text{mod}}$.

- A parameterized model of the instrumental response (given in Equation 4.62) is adjusted on the points $x_{k_{\text{mod}}}, y_{k_{\text{mod}}}, c_{k_{\text{mod}}}$, and the best parameters (in the least-square sense) are estimated.

In practice, the instrumental response is expected to vary over orbital timescales ($\sim$ 90 min), and the parameter estimator could run every few minutes. Simulating such long periods in the MATLAB/SIMULINK model is not possible, and much faster variations of the instrumental response were injected, with errors of $\pm 30\%$ on the intensity and the FWHM, with a period of 60 s. The parameter estimator was set to run at 5 Hz, and the results are given in Figure 4.27.

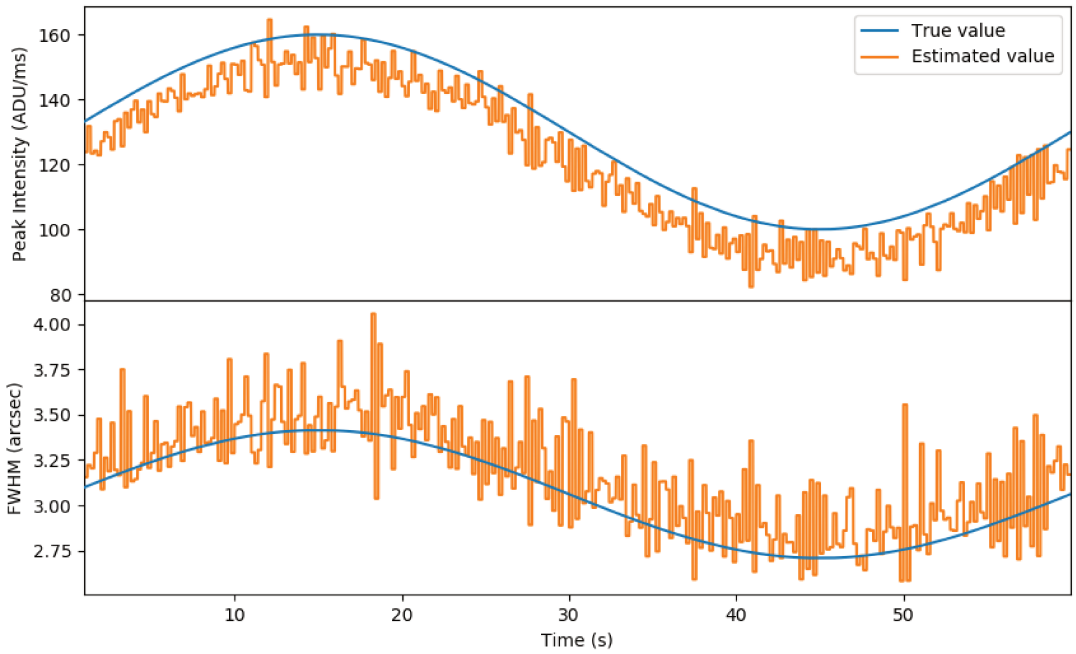
4.5.9 Pointing precision and target magnitude

The level of noise on each individual integration made with the photodiode has a direct impact on the quality of the tracking.

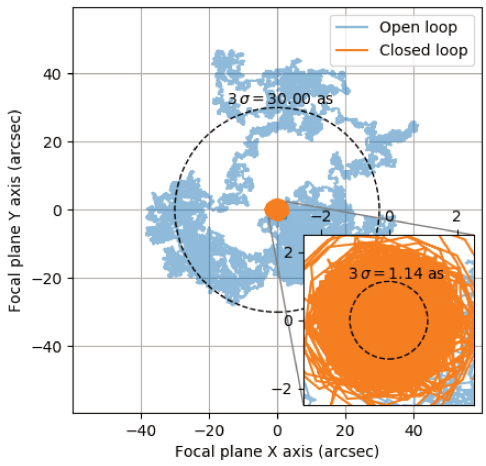
Figure 4.28 shows the results obtained using the MATLAB/SIMULINK model of the PicSat control loop, in the case of Scenario 3 (ASTERIA attitude errors, normalized to 10 arcsec at $1\,\sigma$), for different level of incoming flux. The noise on each measurement can be separated in two terms: the photon noise, which varies as the square-root of the total number of counts, and an error term related to the pointing error, which varies linearly with the total number of counts. The Signal-to-Noise Ratio (S/N) has the form:

$$\text{S/N} = \frac{S}{\sqrt{S + \alpha^2 S^2}} \tag{4.64}$$

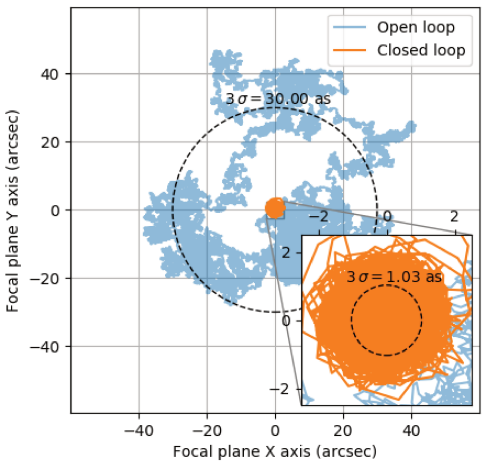
with α typically of the order of 0.05 to 0.1. This particular noise structure explains the presence of a threshold at high flux.



(a) Parameter variations



(b) Pointing residuals



(c) Reference case

Figure 4.27: Upper panel: variations of the two main instrumental parameters (peak intensity and FWHM of the gaussian injection function), and values estimated by the "parameter loop". Lower panel: residual pointing errors in the reference case (ASTERIA ADCS errors, star modulation pattern with an inner circle of radius 0.5 FWHM and an outer circle of radius 0.2 FWHM), and in the same case with parameter variations and parameter loop closed.

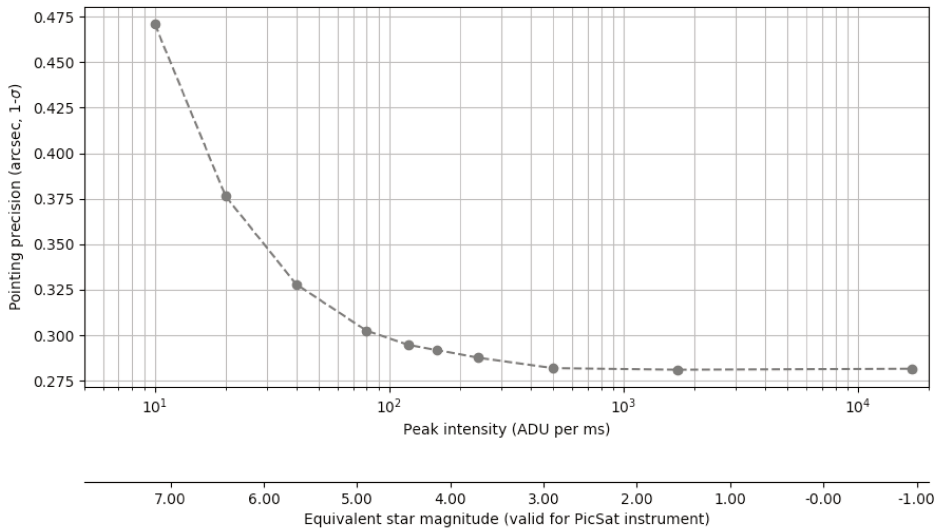


Figure 4.28: Evolution of the tracking precision as a function of the number of counts per integration (taken at the peak of the injection function, NOT on the modulation positions). This plot was obtained using 1 ms integrations, in the case of Scenario 3 (ASTERIA errors). The bottom axis gives the equivalent star magnitude, assuming an aperture size of 3.71 cm, an injection ratio $\eta = 0.8$, and an instrumental transmision $T = 0.2$ (which includes the quantum efficiency of the detector).

The presence of a threshold at high flux means that for very bright targets, or if using the algorithm with a larger aperture size (more light-collecting power), it may be interesting to reduce the integration time. If the instrument observes in a regime where the pointing precision is limited by this threshold, then reducing the integration time will not change the S/N of individual integrations, but the increased frequency of the filter will help to correct the higher-frequencies of the ADCS errors.

In the case of the PicSat instrument, the value of the optimum integration time is probably around 1 ms, but also depends on the exact behavior of the attitude control system. This parameter was also a parameter which could be changed in space.

4.6 Conclusion

The performances of the different algorithms presented in this Chapter are summarized in Table 4.1, which also gives the resulting photometric performance which can be achieved (taking into account the photon noise and the tracking noise). The initial objective (see Section 3.5) was 60 ppm/hr of photon noise, and 80 ppm/hr of tracking noise. Although the use of an extended Kalman filter with a star pattern for the modulation pattern adds more photon noise, it also reduces the tracking noise in case of Scenario 1 and 2. The total budget is reached for these two scenarios. In scenario 3, which corresponds to the real flight data obtained with the ASTERIA ADCS [Pong, 2018], the total noise is close to the objective, although slightly above. It should be noted, though, that the errors from the ASTERIA mission have been scaled up to a standard deviation of 10 arcsec, to take into account a

		Scenario 1		Scenario 2		Scenario 3	
Barycenter Kalman		Track.: 0.27 as Phot.: 148 ppm		Track.: 0.21 as Phot.: 133 ppm		Track.: 0.46 as Phot.: 243 ppm	
Extended Kalman		Track.: 0.15 as Phot.: 110 ppm		Track.: 0.15 as Phot.: 110 ppm		Track.: 0.38 as Phot.: 186 ppm	
Extended Kalman Star modulation	Outer	Track.: 0.18 as Phot.: 130 ppm	Outer	Track.: 0.18 as Phot.: 131 ppm	Outer	Track.: 0.37 as Phot.: 208 ppm	
	Inner	Track.: 0.18 as Phot.: 70 ppm	Inner	Track.: 0.18 as Phot.: 70 ppm	Inner	Track.: 0.37 as Phot.: 96 ppm	
	All	Track.: 0.18 as Phot.: 61 ppm	All	Track.: 0.18 as Phot.: 61 ppm	All	Track.: 0.37 as Phot.: 87 ppm	

Table 4.1: Summary of the performance of the PicSat control loop for each scenario. For the extended Kalman filter and the star modulation pattern, the performance is given for the outer circle, the inner circle, and a noise-weighted combination of the two ("all point"). The tracking performance is the $1\,\sigma$ uncertainty on the distance between the star and the estimated position (at 1 kHz). The photometric performance is the $1\,\sigma$ uncertainty on the photometry, which includes photon noise and tracking errors, converted to a value in ppm/hr by dividing by $\sqrt{1\text{ hr}/1\text{ ms}}$.

possible lower performance of the iADCS100 used in the PicSat mission. The real ASTERIA errors leads to a total noise which is below the budgeted noise.

Overall, with a little bit of tweaking not unfamiliar to control algorihtms, the use of a Kalman control loop running at 1 kHz can guarantee the photometric performance of the mission[5].

[5]At least from a tracking perspective. I will show in Chapter 6 that tracking errors may not be the dominating source of noise, in the end...

Chapter 5

Embedded Software

Content of this chapter

5.1 Introduction and general philosophy

5.1.1 Development

In Chapter 4, I presented the tracking algorithm used to keep the fiber centered on the target star, and stabilize the injection. But designing a control loop and having a payload effectively tracking β Pictoris are two *very* different things. Between the two lies the STM3F3 microcontroller, and a daunting task: programming it. I spent a significant fraction of my thesis time developing the flight-software for the payload.

One of the most important design decision which was taken during the development process was to divide the flight software in two parts: a bootloader, and an application software. The objective was to be able to update the flight software with new algorithms, or to reprogram part of the payload while in flight.

The general development philosophy I decided to adopt for the payload software (bootloader and application software) also follows a dual approach. On the one hand, I developed all low-level functions (drivers for the different hardware elements of the board, processor clocks and bus management, etc.) from scratch, by directly accessing the required microcontroller registers. This was difficult, and very labor intensive. But with this strategy, I had a complete control on the functionality offered by the drivers, and I avoided the unnecessary and/or uncontrolled side effects of external libraries (such as the HAL library for the STM32 processor). On the other hand, to avoid unmanageable complexity, and to improve the readability, testability, and maintainability of the high-level functionality, I based the general architecture and development of the payload software on the GERICOS framework [Plasson et al., 2016]. The use of this framework was particularly facilitated by the fact that I had no external library to interface with it.

5.1.2 Testing

Except for some of the mathematical functions used in the tracking algorithm, most of the low-level functions are highly hardware dependent, which makes them difficult to test. Because most of these drivers also have very limited functionality and/or adjustable parameters, the tests performed "on the flow" during the development were usually enough to ensure their reliability.

The small size of the payload as well as its limited scope of operations also made possible for me to perform a "full-scale" on-ground test. To do so, a dedicated test-bench was built at LESIA, and I tested the payload electronics, software, algorithm, and its interface with the other systems on this bench, using the real hardware.

5.2 Bootloader

5.2.1 Overview

The PicSat Payload Bootloader (PPB) has been developed for two different purposes: to mitigate the risk of having bugs in the on-board software (difficult to test thoroughly in the laboratory environment), and to keep the possibility of observing targets other than beta Pictoris during the late stages of the mission (which could potentially require a completely different tracking algorithm, for example if observing a double star[1]).

[1] α Cen? Yes, I had ambitious dreams about this payload...

The main point of having a bootloader for the payload microchip was to be able to re-write the Flash memory while in-flight. This is critical for any software update. As this piece of software is highly critical, I conducted its development with the following guidelines in mind:

- simplicity: I kept the number of functionalities as low as possible, to avoid complexity in the development and maximize reliability.
- compacity: I wrote the bootloader as small as possible, in order to leave enough space for the application software in the limited Flash memory.
- hardware minimalism: the bootloader uses only the absolutely necessary hardware of the board. This ensures that even in the case of hardware deficiency, the PPB will still operate nominally.
- safety: the PPB implements dedicated software checks to make sure that a stupid user (command) cannot damage it.

To ensure that even in the case of a failed update of the payload application code, the payload could still operate, the PPB also handles two applications (A and B), written in two different parts of the Flash memory. This can be used to always keep a functioning application on the board.

Finally, I made the PPB itself compatible with the CCSDS standard for communication. CCSDS packets can be directly routed from the ground to the payload bootloader by the OBC.

5.2.2 Usage

Startup

On power-up, the STM32 microchip of the payload board automatically starts executing the PPB code from the Flash memory. The RS422a/UART4 interface and its RX interrupt are configured, and the PPB automatically sends its version number in the form of a CCSDS telecommand answer packet (answering a fake `get_version` command). It is then ready to accept commands, in the form of CCSDS telecommand (TC) packets.

Starting an application

Starting one of the two applications (A or B) stored on the payload board is as easy as sending the appropriate telecommand to the payload bootloader. The aplpication code is loaded from the Flash memory, and the application starts.

Updating the Flash memory

To write a new application in the Flash memory, the user simply needs to alternate TC packets to erase the first sector to be re-programmed, write it (usually requires two packets because of the limitation in packet size), re-read to check the content if necessary, and move on to the next sector.

5.2.3 Memory map

The Flash memory embedded on the STM32F303 microchip has a total size of 512 kB, and is divided into 250 sectors of 2048 B each. The PicSat Payload Bootloader is written at the very beginning of the Flash memory. A total of 59 kB (the first 29 sectors, i.e. sectors 0 to 28

included) are allocated to this PPB. The PPB implements software Flash access protection, which makes impossible for the user to erase, write, or even read these first 29 sectors. The PPB itself has no access to this part of the Flash memory.

Sector 29, directly following the PPB, is the application determination sector, used in a previous version of the bootloader, but unused in the most recent one. This sector is not software protected. The user can access it in either write, read, or erase.

The rest of the memory (the last 220 sectors) is divided in two parts of 225 kB each (sectors 30 up to 139 included; sectors 140 to 249 included), in which the two applications can be stored. The PicSat Payload Application A starts at the beginning of sector 30 (address: 0x0800F000), and ends at the end of sector 139 (address: 0x08045FFF). Application B starts at the beginning of sector 140 (0x08046000), and ends at the last address of the memory (0x0807CFFF).

Section	Section Size	Sector	Begins at (HEX)	Ends at (HEX)	Sector Size	
Picsat Payload Bootloader	59392	0 1 2 ... 28	8000000 8000800 8001000 ... 800E000	80007FF 8000FFF 800E7FF ... 800E7FF	2048 2048 2048 ... 2048	PROTECTED
Application determ.	2048	29	800E800	800EFFF	2048	
Picsat Payload App. A	225280	30 24 25 ... 139	**800F000** 800F800 8010000 ... 8045800	800F7FF 800FFFF 80107FF ... 8045FFF	2048 2048 2048 ... 2048	
Picsat Payload App. B	225280	140 141 142 ... 249	**8046000** 8046800 8047000 ... 807C800	80467FF 8046FFF 807CFFF ... 807CFFF	2048 2048 2048 ... 2048	

Figure 5.1: Map of the Flash memory of the STM32 microchip embedded on the PicSat payload board.

5.2.4 Hardware involved

The PPB uses only the minimum required hardware to perform its activities. On the payload board, only the RS422a interface is used, and no other system is activated. In the STM microchip itself, the UART4 interface, as well as the NVIC (used to manage interrupts on UART4 RX) are required for communication. The timer TIM1 is also used. Finally, the RTC can also be configured to ensure proper timetag values of the telemetries sent by the PPB, but this is optional (the RTC is only configured on user request). The RTC runs on the LSE. Of course, proper operation of the embedded Flash memory access hardware is also required for updating the memory.

Task	Hardware involved
Gathering and monitoring HK data	ADCs, Temperature Sensors, I2C, Gyroscope
Regulating SPAD temperature	ADCs, SPAD, SPI, DAC
Acquiring photometry	TIM15, TIM3, SPAD, binary counter
Tracking star	HV transf., DACs, I2C, SPI, Gyro., piezo ADCs
Sending data and report to OBC	UART/RS-422 module, UART, DMA

Table 5.1: List of the tasks dedicated to the payload board during operations.

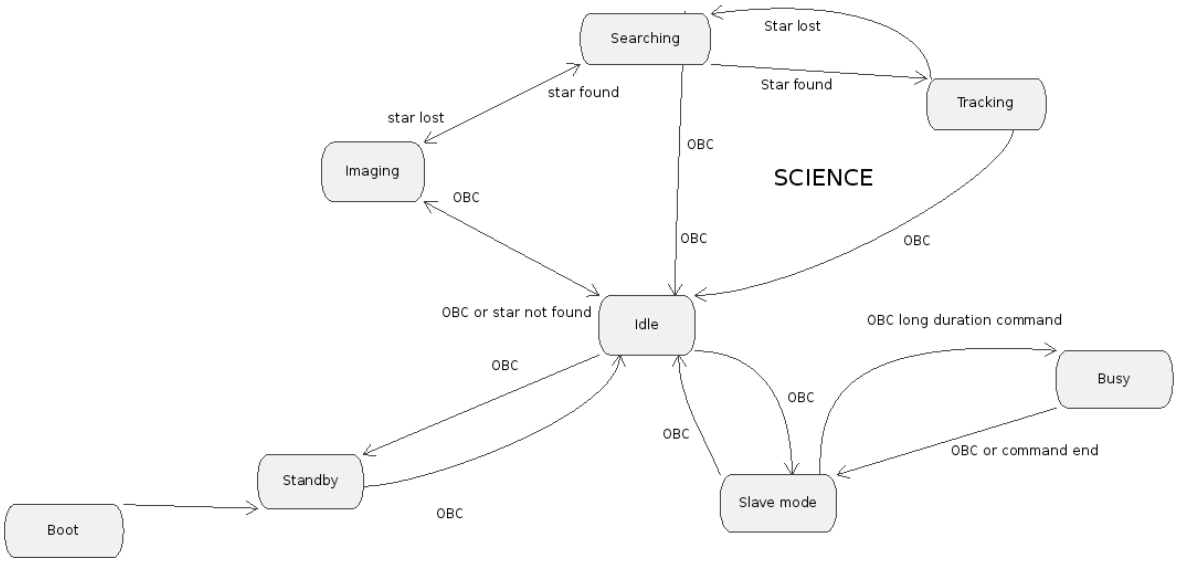


Figure 5.2: State diagram of the payload board. Most of the transitions must be explicitly requested by the On-Board Computer (OBC)

5.3 Application software: High-level design

5.3.1 Tasks devoted to the payload board

During operations, a certain number of tasks are devoted to the payload board. These tasks can be related to the science activities and to House-Keeping (HK) monitoring and management. A list of these tasks can be found in Table 5.1.

5.3.2 State machine

Based on the list of tasks described in Table 5.1, I decided to use a limited number of states for the state machine of the payload board. In Figure 5.2, I give the complete state diagram I adopted fir the payload software. The detailed description of the different modes have been given in Section 3.4.

The transitions from one state to another are mainly managed externally by the OBC. During science operations, transitions are automatically triggered by the payload between IMG, SCH and TRK modes when the star is lost/found. These transitions have minimal effects on the different subsystems of the board (they mainly switch from one algorithm to another), and are managed internally. All other transitions must be triggered explicitly by

the OBC. A transition from SBY to IDL is expected just before β Pic visibility (mainly to start the SPAD temperature regulation and have a stable temperature when β Pic is visible). Another transition from IDL to IMG is expected at the beginning of visibility. At the end of visibility, the payload will automatically cascade from any of the science modes back to IDL (TRK to SCH to IMG to IDL). The minimal behavior expected from the OBC is to trigger an IDL to SBY transition.

5.3.3 Overview of the design

To ensure that no single activity of the payload uses all the computing resources, the processor time is divided between multiple tasks, running in separate threads. The RTOS allocates computer time to these tasks based on their priority levels.

I divided the payload software in the following 7 tasks, given by decreasing priority level:

- ObcInterfaceManager: this task manages the communication interface between the payload and the OBC. It receives data on the RS422 interface, detects telecommand packets, verifies their format, and transmits them to the BoardDirector for execution. To avoid any situation where the user would be unable to send any command to the payload because another task is trapped in a function that never returns, this task has the highest priority level.

- BoardDirector: this task manages most of the payload board subsystems. This is the main task in which the telecommands are executed. This task is also responsible for managing the state machine.

- AcquisitionManager: responsible for all the science activities of the payload, and in particular for tracking the star and acquiring the photometry. This task has a relatively high priority level because any interruption in the data acquisition can lead to a loss of tracking, which should be avoided.

- PacketsManager: this task manages the telemetry packets and forward them to the ObcInterfaceManager when they are ready to be sent to the OBC.

- HkManager: a low priority task whose job is to manage the different sensors and gather the HK data.

- BeaconManager: a very low priority task whose job is to regularly send a beacon to the OBC.

- Idle: the default Gericos task which is called when no other task require processor time.

Two different mechanisms are used within the payload software to wake up these tasks: software timers and hardware interrupts. A total of four different timers are used to clock different activities. A hkManagerTimer wakes up the HkManager when HK data must be collected (which happens on a regular basis). Two timers (sciencePacketTimer and hkPacketTimer) are used to regularly empty the data buffers and send packets to the OBC. These two timers wakes up the packetsManager task. Finally, a beaconTimer is used to wake up the beaconManager task on a regular basis (if requested by the user).

To these timers, two important interrupts are added to manage inbound communication and science acquisitions. These two interrupts are the following, in decreasing priority order:

- obcRxInterrupt: this interrupt is triggered when a data byte is received on the payload-OBC communication line. The handling of this interrupt is totally independent of the

general GERICOS/RTOS task management system. An incoming byte will trigger this interrupt, which will ALWAYS be immediately serviced by the processor, no matter which task is currently running.

- diodeIntegrationInterrupt: this interrupt is triggered upon termination of an integration of the photodiode. It is serviced immediately by the processor, but its only effect is to wake up the AcquisitionManager task. This means that proper action upon this interrupt takes place within the GERICOS/RTOS flow, when time is allocated to the acqusitionManager task.

5.3.4 Telecommand management

As stated in the previous section, the telecommands from the OBC are received through the USART1/RS422a line. The bytes are received one by one using a dedicated hardware interrupt. When a new byte is received, the ObcInterfaceManager adds it to a dedicated buffer, and checks if an "end packet sequence" has been received. If so, the buffer is emptied into a telecommand packet, and the format of the packet is checked. Is the packet is found to be valid, it is transferred to the BoardDirector for execution. If not, the packet is dropped. In either case, if a format acknowledge has been requested, a request is sent to the PacketsManager task for creating and sending it.

A generic mechanism is provided in the payload source code to add new telecommands: the CommandHolder interface. In theory, any task inheriting form this CommandHolder class, can register its own telecommands which can then be executed. However, as no telecommand forwarding mechanism between the different task of the payload software is provided, the BoardDirector is in practice the only task which can effectively inherits from CommandHolder.

To give some details about this mechanism: a task inheriting from CommandHolder has access to a `void executeTcPacket(TcPacket*)` to execute a command, and also to a `void pushCommandToTable(uint8_t, void_func)` method to register a new function (of type `void_func`) in the table of available commands. The table is created at startup of the application in the RAM, and gives the correspondance between a telecommand id and the memory address at which the telecommand code is located (a "pointer" to the telecommand). When the `executeTcPacket` is called, the telecommand id is extracted from the packet, the corresponding registered function is found and called with the whole tcPacket as argument (to ensure that the necessary parameters are passed to the command). The CommandHolder interface also automatically manages the emission of format and execution acknowledgement packets, as per CCSDS specifications.

The class diagram corresponding to these activities is given in Figure 5.3.

5.3.5 Science activities

5.3.5.1 Architecture

At software level, all the science activities are managed by a single GERICOS active object: acquisitionManager. This task is linked to several different workers (classical objects which are not tasks[2]), each-one encapsulating all the necessary low-level functions to drive one particular hardware subsystem. Thus, all these hardware subsystems are fully separated and independent at software level[3].

The acquisitionManager is also linked to three different ProcessingWorker objects, which are responsible for performing all the mathematical operations for guiding the fiber. One

[2]The workers are... well, inactive, I guess? Only the "managers" have their own threads.

[3]This is standard division of labor applied in software...

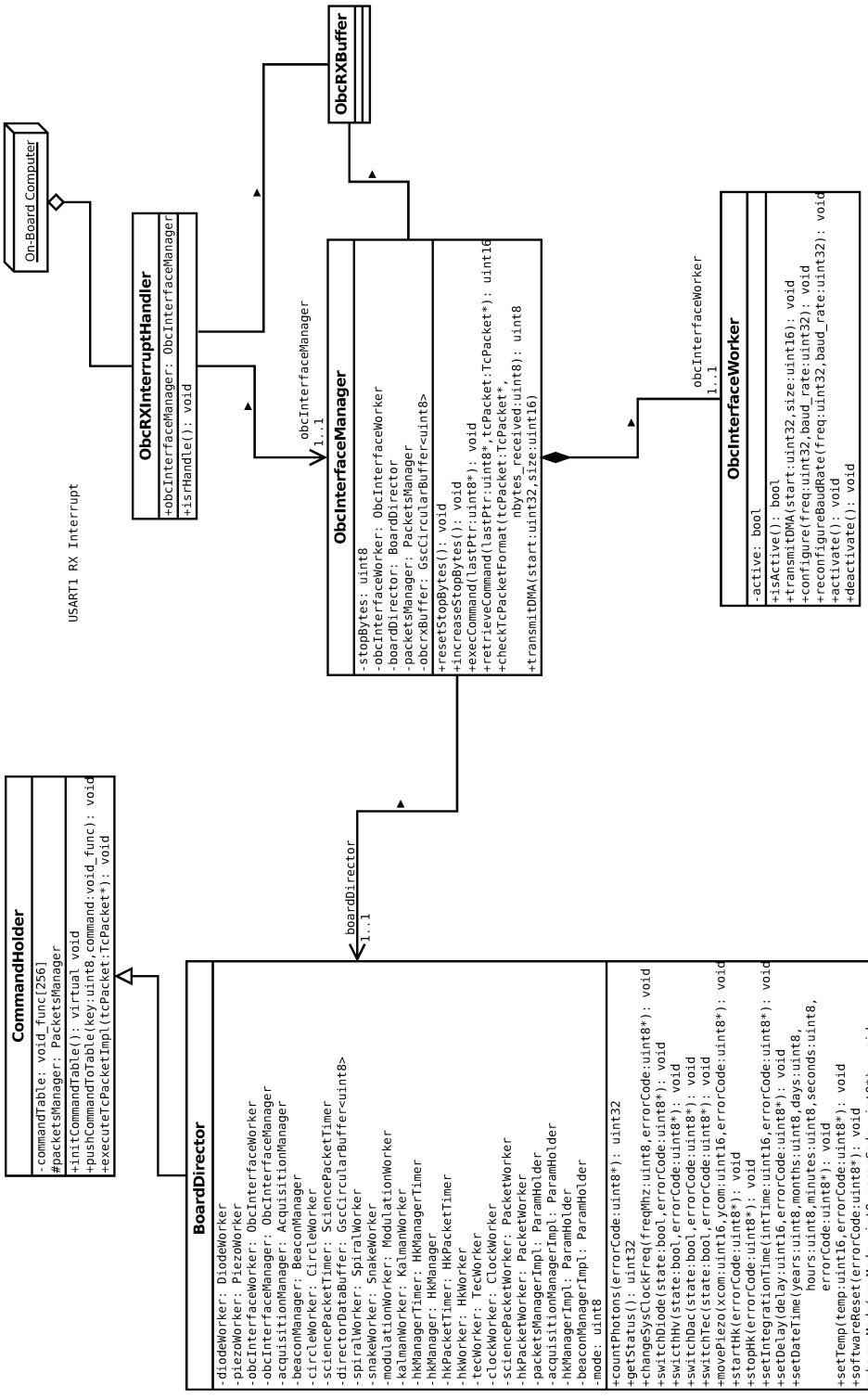


Figure 5.3: Class diagram showing the telecommand management within the payload software.

of these ProcessingWorker objects is in charge of the "imaging" phase, another manages the "searching" phase, and the final one is responsible for the "tracking" phase.

Finally, the acquisitionManager also has access to a dedicated PacketWorker, with its own circular data buffer to store all the photometric measurements waiting to be sent to the OBC.

All the activities of the acquisitionManager are clocked by a hardware interrupt coming from the photodiode integration timer TIM15. When the integration ends, the interrupt automatically loads the TIM3 counter register value (measurement from the SPAD) into the `uint32_t counts` attribute of the DiodeWorker, and then requests a call to a `void nextAcquisition(void)` method of the acquisitionManager. Thus, starting or stopping the acquisition sequence is as simple as activating or deactivating a hardware interrupt.

The class diagram corresponding to these activities is given in Figure 5.4.

5.3.5.2 The ProcessingWorker interface

In an effort to simplify the implementation of new tracking algorithms, I implemented a dedicated interface[4], called ProcessingWorker, to encapsulate all the algorithms. Each implementation of an algorithm takes the form of a class, inheriting from the ProcessingWorker interface, and thus needs to provide a `void computeNextCommand(void)` method to compute, using internal variables, the next command to apply on the piezo actuator. A way to retrieve the X and Y commands must also be provided, in the form of a `uint16_t getXCom(void)` and a `uint16_t getYCom(void)` methods. Thanks to this inheritance mechanism, each of the sub-classes will also has a direct access to several attributes: `uint16_t xpos`, `uint16_t ypos`, `uint32_t counts`, `uint16_t x0`, `uint16_t y0`, containing respectively the last known position of the piezo actuators, the last measurement from the SPAD, as well as the last known position of the star. This inheritance mechanism ensures that the detailed mechanics of a particular algorithm is completely irrelevant at software-level, and the only thing that matters is that each and every object implementing a tracking or searching algorithm must have the attributes of the ProcessingWorker class. This has the practical advantage of making any change of the algorithms at any point of the project relatively easy, by completely decoupling them from the rest of the code.

5.3.5.3 Sequence diagrams

The instruction flow occurring within the payload during a single acquisition is given in Figure 5.5. To maximize the total integration time on the SPAD, all the heavy tracking calculations are performed when the diode is integrating (integration on the photodiode is completely done at hardware level, and the processor if free during this time). Thus, the calculations is always done one step in advance: the diode integrates at position n when the processor computes the position n+1 to be applied at the next iteration.

The expected instructions required to start the science activities (i.e. switching from IDL mode to IMG) is given in Figure 5.6. Note that in this diagram, the BoardDirector is a particular GERICOS Task which has access to all the subsystems of the board, and which is in charge of organizing the operations. In some sense, it is the "Mode Manager" of the payload.

[4] An "interface", in object-oriented programming, is a class used to specify the general structure of an object, its input and outputs, and the methods it is required to have. An interface class is never meant to be instantiated, and the methods declared in its structure are devoided of all content. It only provides a common structure, sort of like a "contract", which guarantees that any class inheriting from it will have certain methods and/or attributes. The acquisitionManager does not need to know the details of the algorithm implemented by any given ProcessingWorker. As long it the ProcessingWorker has a "computeNextCommand" method, the acquisitionManager can use it, no matter what it really does.

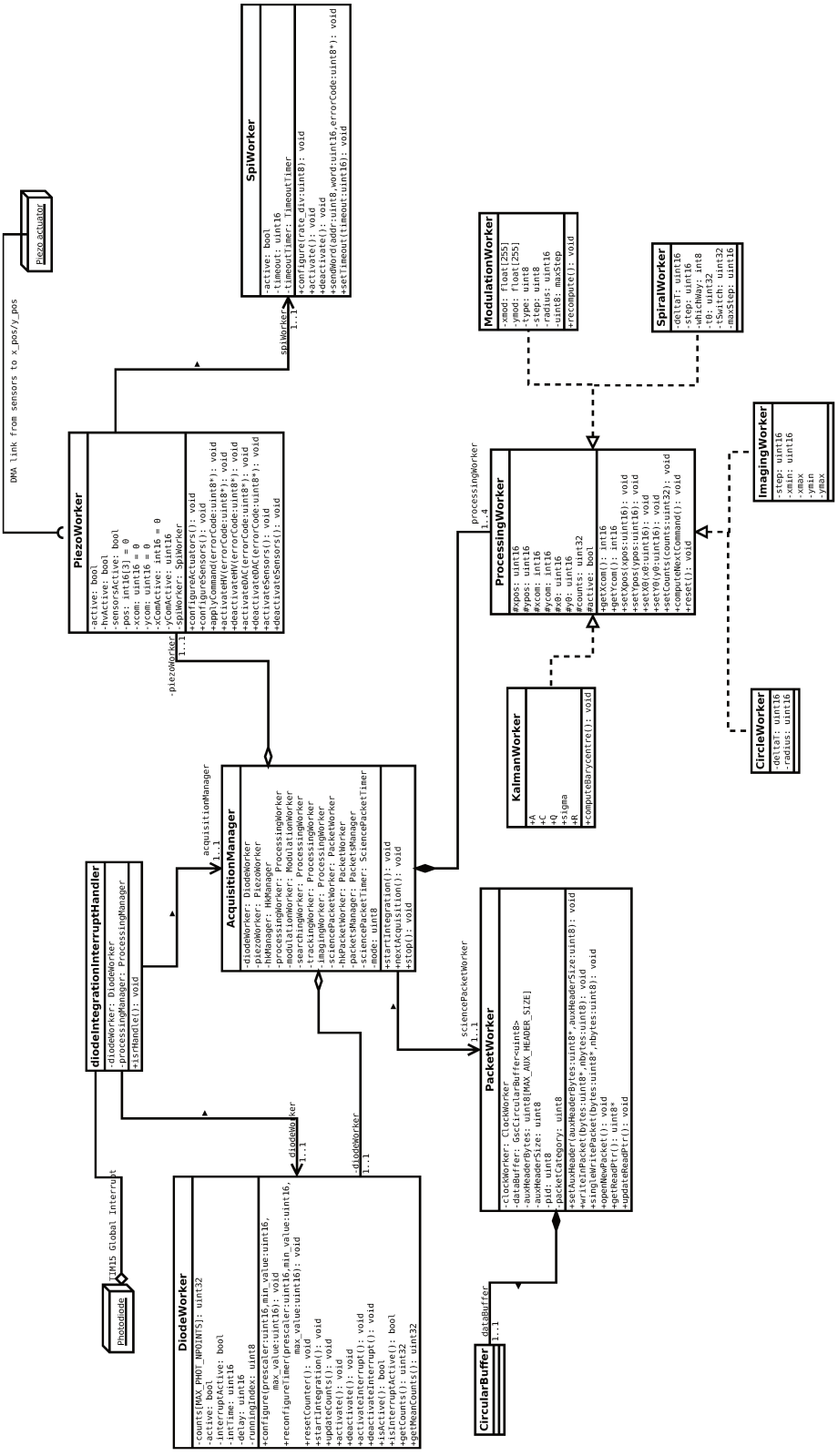


Figure 5.4: Class diagram showing the organization of the different software objects involved in science data acquisition

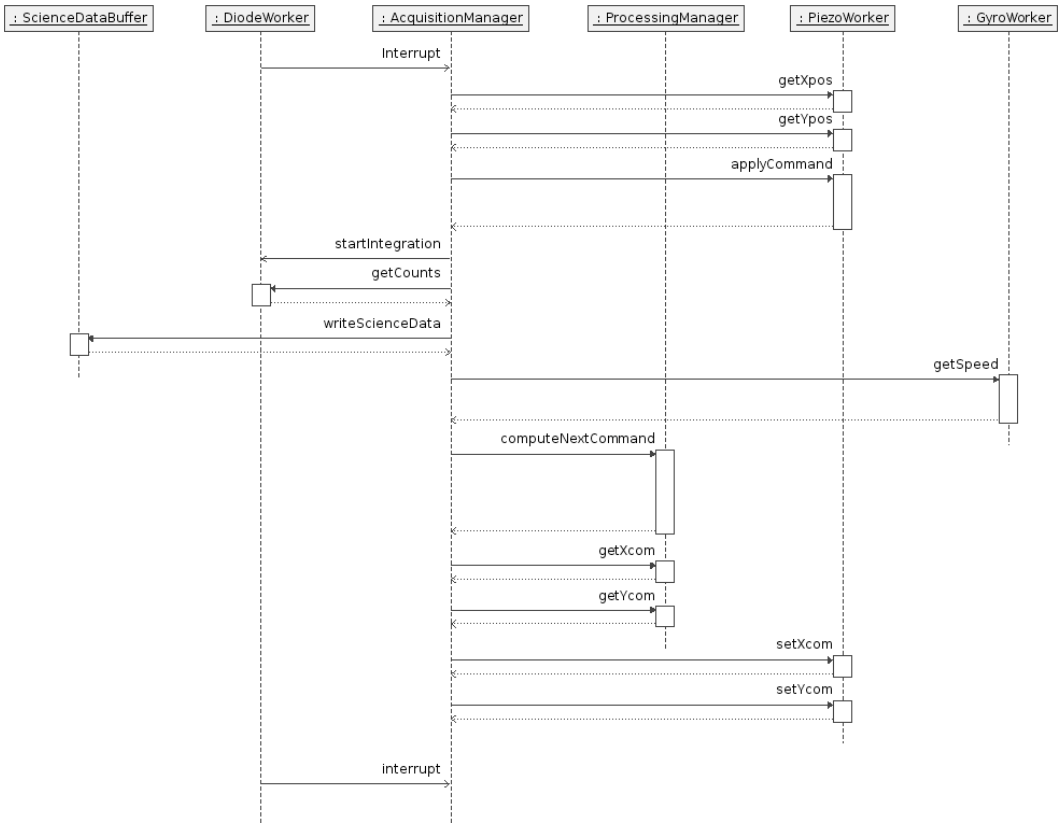


Figure 5.5: Sequence diagram for one acquisition

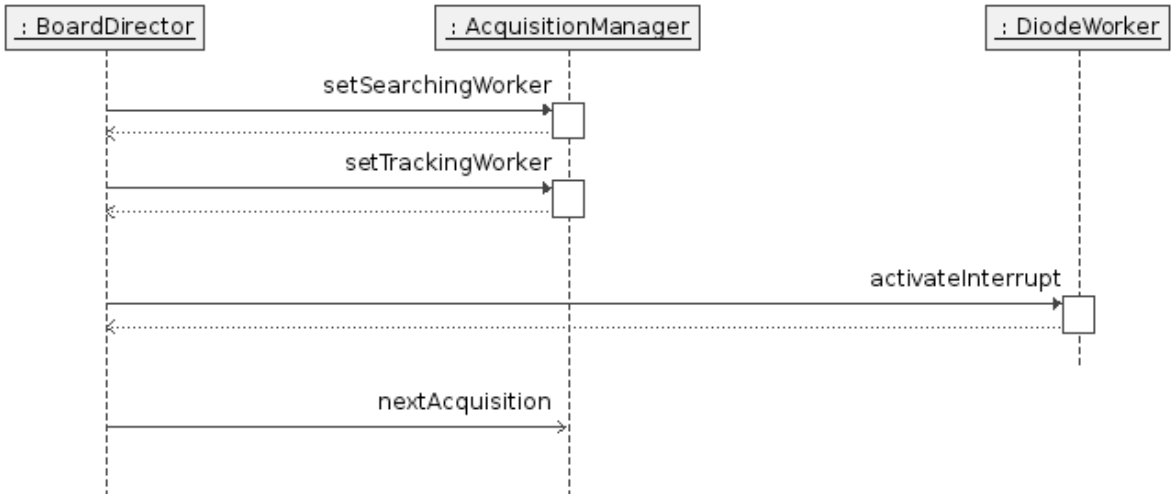


Figure 5.6: Expected sequence diagram for starting the science activities

5.3.6 House-Keeping management

5.3.6.1 Architecture

Due to the limited number of information available (14 sensors in total), the house keeping activities on the payload board are limited. The task in charge of managing those activities is the HkManager. This task is directly linked to a HkWorker which does all the actual work[5]. A HkPacketWorker, with its own data buffer, is used to write the collected data in correctly formatted CCSDS packets before sending them to the OBC. The class diagram is given in Figure 5.7. HK data acquisition are timed by a software timer.

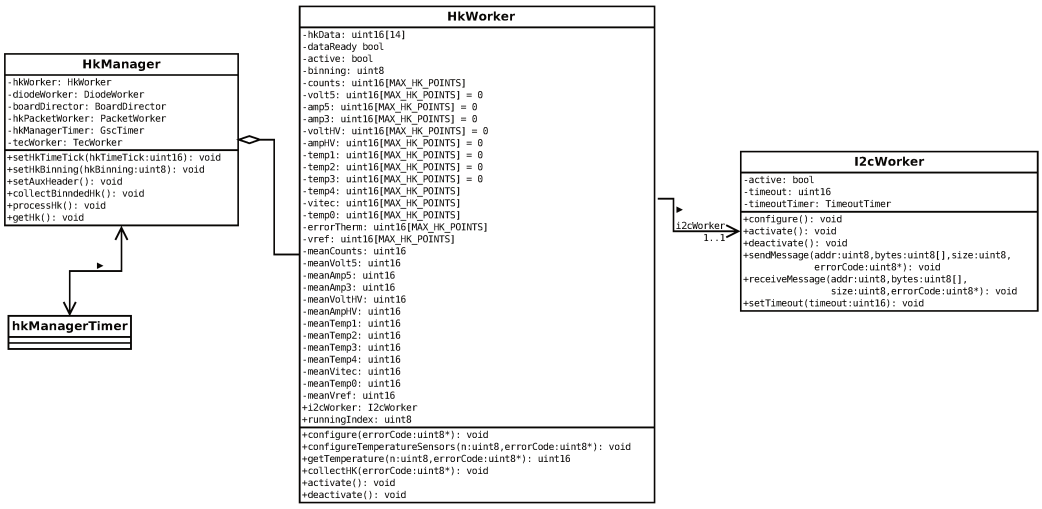


Figure 5.7: Class diagram showing the organization of the different software objects involved in house-keeping activities

5.3.6.2 Sequence diagrams

A typical sequence diagram for HK data acquisition is given in Figure 5.8.

5.3.7 Beacon

5.3.7.1 Architecture

I implemented a simple automatic beacon transmission in the payload software[6]. The BeaconManager task is responsible for creating the beacon and sending it via the PacketsManager class on a timely basis. The class diagram is provided in Figure 5.10.

[5] According to proper division of labor...

[6] The usefulness of a payload beacon was initially debated within the PicSat team, as there was already a beacon machanism for the satellite itself. It turned out to be pretty useful when, on Feb, 12, 2018, during one of the morning passes, why nobody in the operation team had ever requested the satellite to turn on the science instrument, the satellite rose from the North-East horizon emitting two beacons: the normal satellite beacon, and the payload beacon. A small bug in the L1 satellite state machine was triggering untimely power on of the payload.

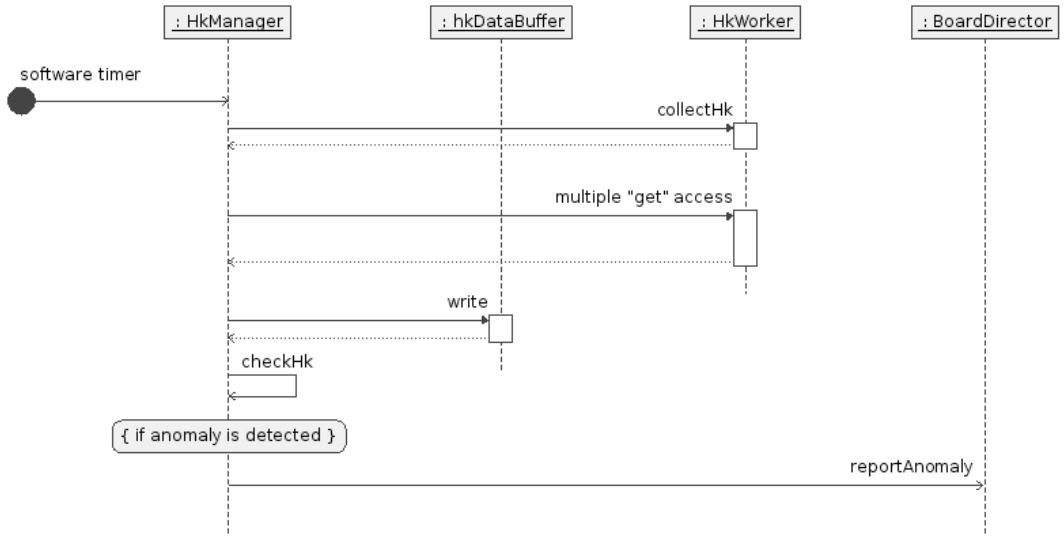


Figure 5.8: Sequence diagram for HK data collection

5.3.7.2 Sequence diagram

The typical sequence diagram for a beacon emission is given in Figure 5.9.

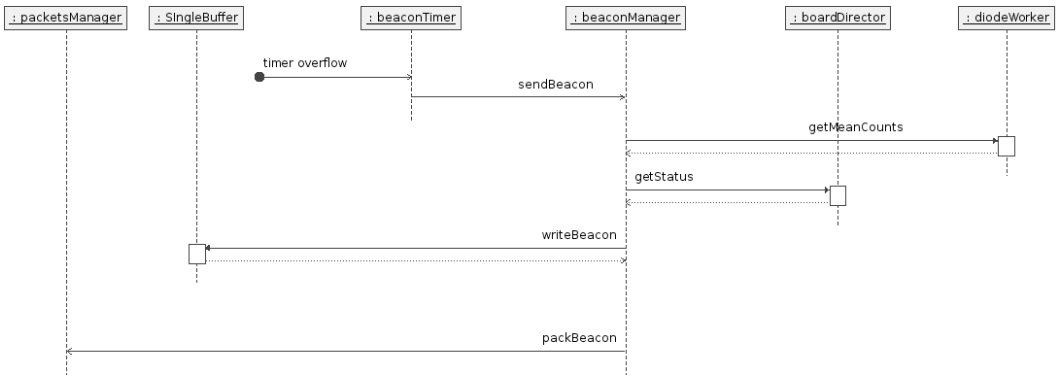


Figure 5.9: Sequence diagram for beacon emission

5.3.8 Communications

5.3.8.1 Architecture

During the science phase of the mission, the communication between the payload board and the OBC can reach up to > 100 kbits/s. This is not negligible for the payload processor, and does require some careful management of the data. The main objective of the payload-OBC communication architecture is to maximize the use of the hardware DMA[7] in communication management. To do so, the science data are first stored on the payload board, then organized

[7]DMA, or Direct Memory Access, is a very useful feature of microcontrollers, which can be used to automatically transfer data from one part of the internal memory to another, or between the internal memory and a peripheral. A DMA is very useful to transfer long streams data on a communication link, as the microcontroller only needs to execute a few instructions to set up the DMA. All the data handling is then performed by the DMA itself, while the microcontroller is free of doing something else.

into communication packets which can be automatically handled by the DMA at hardware level.

Communication between the payload and the OBC are handled by two different tasks: ObcInterfaceManager and PacketsManager. ObcInterfaceManager controls the communications and PacketsManager is in charge of organizing the creation of the packets that will be sent by DMA on request of the ObcInterfaceManager.

In this architecture, data transmission can either be timed by a software timer (which times the creation of the packets at PacketManager level, which, in turn, triggers a request to the `void transmitDMA` method of the ObcInterfaceManager), or by a buffer event (e.g. "circular science buffer half-full"). Data reception is timed by hardware interrupts (USART1 RX interrupts).

The class diagram representing the architecture of the OBC communication system is given in Figure 5.11. Note that to enhance readability this diagram only includes science data transmission. House-keeping data transmission, log transmission, etc. are handled the same way, using different packetsWorker, and different software timers.

5.3.8.2 Sequence diagrams

The exact instruction flow behind a DMA packet transfer is shown in Figure 5.12. For command reception, the sequence diagram is given in Figure 5.13.

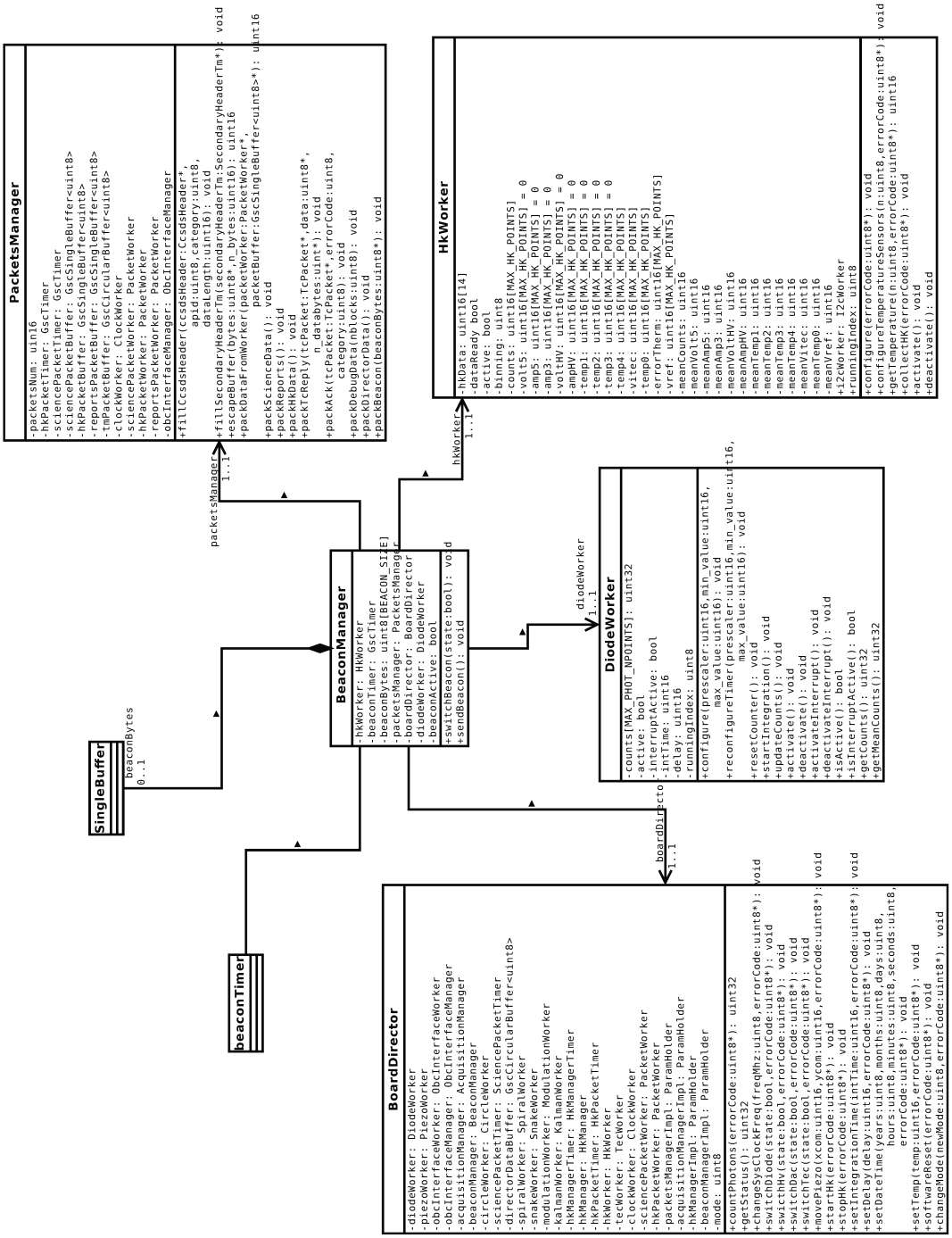


Figure 5.10: Class diagram for the payload beacon management

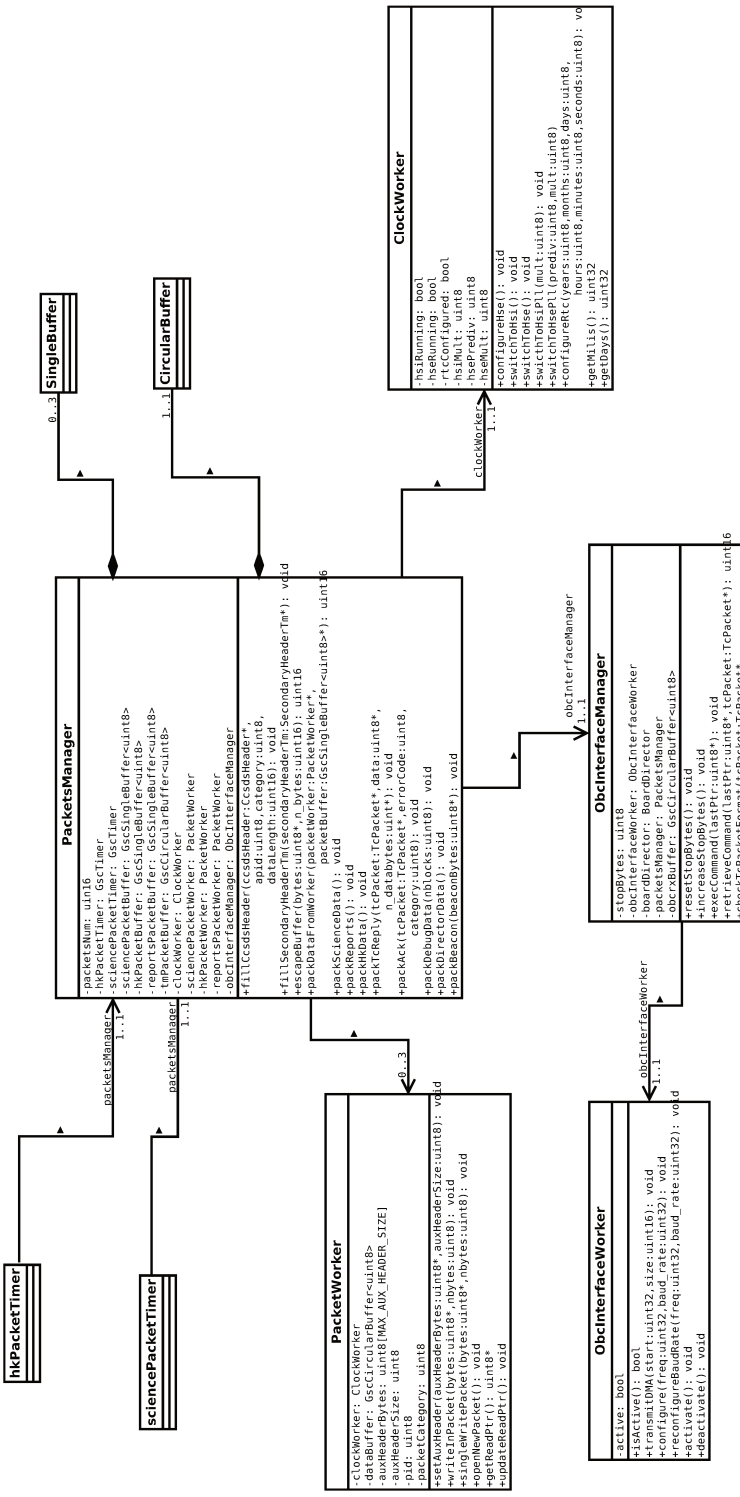


Figure 5.11: Class diagram for the payload-OBC communication management system

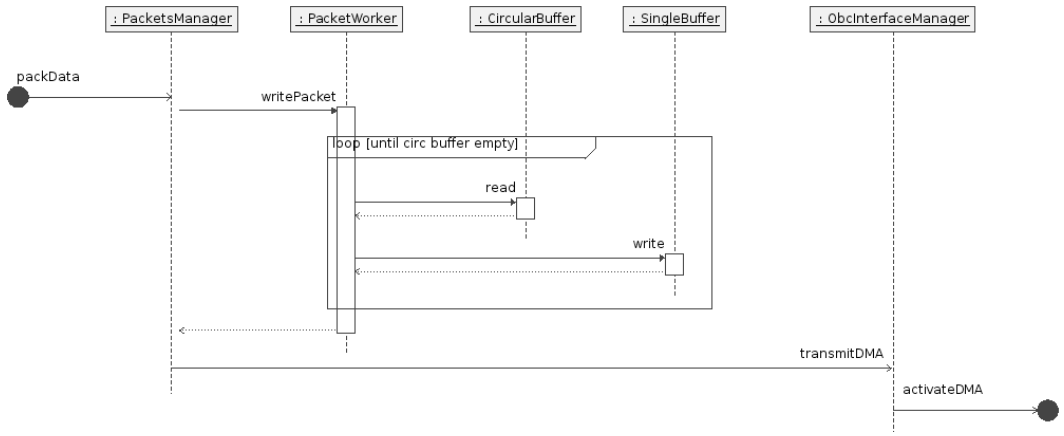


Figure 5.12: Sequence diagram for packet transmission

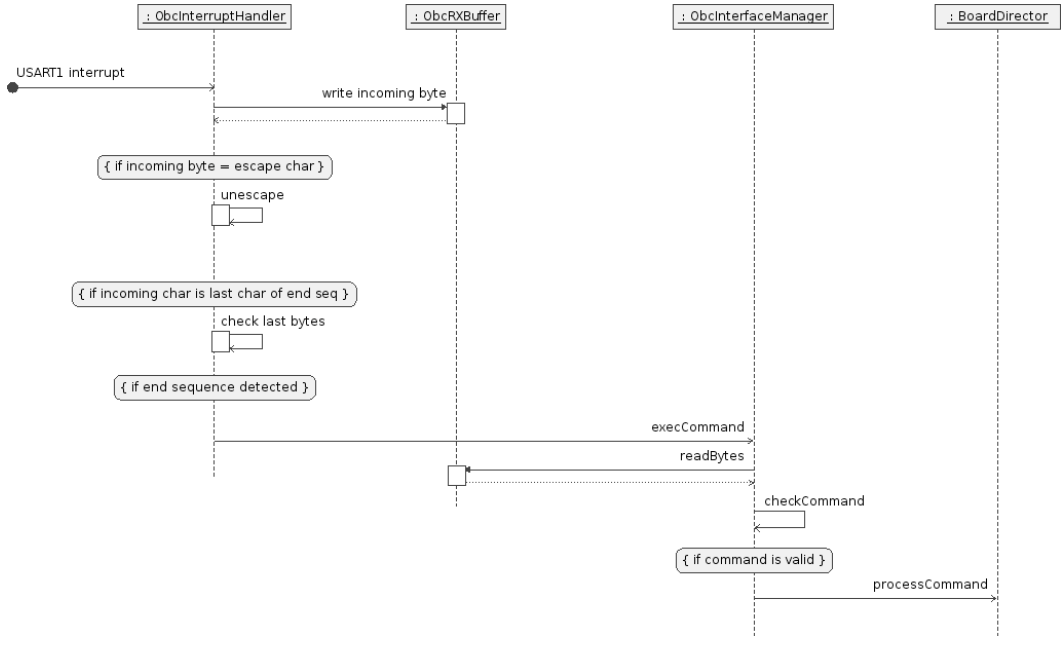


Figure 5.13: Sequence diagram for OBC command reception

5.4 Application software: Low-level implementation

As explained in Section 5.1, I developed all the low-level drivers for the hardware subsystems from scratch, and I directly implemented them in the GERICOS framework. Hereafter, I give a short description of these drivers, organized by classes in which they are implemented.

SpiWorker

1. The SpiWorker class provides a `void configure(uint8_t)` method to configure the SPI bus with a rate division given by the argument, and its associated GPIOs (PB13 and PB15). This method only configures the bus, and does not activate it. The bus is configured in write mode only.

2. The SpiWorker class provides a `void activate(void)` and a `void deactivate(void)` methods to turn the SPI bus on and off. When called, the `deactivate` method will wait until the end of the ongoing transmission, if any, before turning the bus off.

3. The SpiWorker class provides a `void sendWord(uint8_t, uint16_t, uint8_t*)` function which takes an 8-bit address and a 16-bit word as argument, and send them through the SPI bus. This method does not manage the routing of the message on the bus. Activating/deactivating the different subsystems connected to the bus must be done beforehand. If an error occurs on the bus, its value is returned in the third argument.

4. The SpiWorker class provided a `void setTimeout(uint16_t)` method to set the timeout value for the bus, if required. Timeout are managed at software level, using a timeoutTimer.

I2cWorker[8]

1. The I2cWorker class provides a `void configure(void)` method to configure the I2C bus and its associated GPIOs (PB6 and PB7). This method only configures the bus, and does not activate it. The bus is configured in read and write mode.

2. The I2cWorker class provides a `void activate(void)` and `void deactivate(void)` methods to turn the I2C bus on and off. When called, the `deactivate` method will wait until the end of the ongoing transmission if any before turning the bus off.

3. The I2cWorker class provides a:

   ```
   void sendMessage(uint8_t, uint8_t[], uint8_t, uint8_t*)|
   ```

   method which takes an 8-bit address, an array of bytes, and the number of bytes of the array to send, and send these bytes on the I2C bus, to the given address. If an error occur, the error code is returned in the last argument.

4. The I2cWorker class provides a

   ```
   void receiveMessage(uint8_t, uint8_t[], uint8_t, uint8_t*)
   ```

   method which takes an 8-bit address, an empty array of bytes, and a given number of bytes to read. The method requests a read access to the I2C peripheral of the given address, and wait until the correct number of bytes have been received. The bytes are stored in the given array. If an error occur, its value is returned in the last argument.

[8] If you've read the stuff on the SpiWorker, you can probably stop here, and jump to the next chapter...

5. The I2cWorker class provides a `void setTimeout(uint16_t)` method to set the timeout value for the bus, if required. Timeout are managed at software level, using a timeoutTimer.

DiodeWorker

1. The DiodeWorker class provides a `void configure(uint16_t, uint16_t, uint16_t)` method which configures all the pins used by the SPAD, the hardware binary counter, and the timers TIM15 and TIM3 (this includes: PC6, PB1, PB14, PD2). This method also configures TIM15 as a gate generator, with a prescaler value given by the first argument, a switch value given by the second argument, and an overflow value given by the third argument (for a length of the pulse given by arg2 $\times$ arg1/APB2CLK, and a cycle period of (arg3 + arg2) $\times$ arg1/APB2CLK). This method also configures and activates the interrupt on TIM15. The net end result of a call to this method is a powered-off SPAD, and a reset binary counter and TIM3 counter.

2. The DiodeWorker class provides a
`void reconfigureTimer(uint16_t, uint16_t, uint16_t)`
method which changes the values of the prescaler and of the switch and overflow limits of TIM15 respectively for its first, second, and third argument values. This method stops any ongoing integration, resets the TIM15 timer, and updates the registers at the end. This method only accesses TIM15 registers.

3. The DiodeWorker class provides two methods, `void activate(void)` and the opposite `void deactivate(void)` methods to power up and power down the SPAD and its TIM3 counter.

4. The DiodeWorker class provides two methods, `void activateInterrupt(void)` and `void deactivateInterrupt(void)` to activate or deactivate the end of integration interrupt on TIM15. These methods do not change the configuration of this interrupt, but only enable or disable the interrupt request.

5. The DiodeWorker class provides a `void resetCounter(void)` method to reset both the binary counter and the TIM3 counter.

6. The DiodeWorker class provides a `void startIntegration(void)` method which starts the integration by activating TIM15. The method does NOT reset the binary and/or the TIM3 counters beforehand.

7. The DiodeWorker class provides a `void updateCounts(void)` method which reads the value of the TIM3 counter, and load it into a `uint32_t counts` attribute.

8. The DiodeWorker class provides a `uint32_t getCounts(void)` method which returns the result of the last integration.

TecWorker

1. The TecWorker class provides a `void configure(void)` method to configure the temperature regulation system of the SPAD. This method configures all the GPIOs related to TEC regulation (PB8, PB9, PB12, PB5, PA4, PA5, PC6, PC3, PC4, PC5). At the end of the call, the TEC regulation is configured, but not active.

2. The TecWorker class provides a `void setTemp(uint16_t)` method to select the temperature setpoint for the SPAD. This method only changes the value of the `temp` attribute, and does not actually change the setpoint. To do so, a call to the `applyCommand` method is required.

3. The TecWorker class provides a `void applyCommand(uint8_t*)` method to send the value of the attribute `temp` to the DAC register of the temperature control system using the SPI bus. This method does not change the state (active/inactive) of the system. This method uses the SPI bus. If an error occurs on the bus, the error code is returned in the argument.

4. The TecWorker class provides a `void activate(uint8_t*)` as well as the opposite `void deactivate(uint8_t*)` methods to activate and deactivate the regulation system. These functions do not change the current configuration of the TEC system, and use the SPI bus. If an error occurs on the bus, the error code is returned in the argument.

PiezoWorker

1. The PiezoWorker provides a `void configureActuators(void)` method to initialize all the GPIOs related to the two piezo actuators (PA8, PA11, PA12 PC7, PC8). The GPIOs are configured, but the high voltage and DAC modules are NOT started.

2. The PiezoWorker class provides a `void configureSensors(void)` method to initialize all the GPIOs related to the two piezo sensors (PA2, and PA3).

3. The PiezoWorker class provides a `void activateHV(uint8_t*)` as well as the opposite `void deactivateHV(uint8_t*)` methods to safely activate and deactivate the high voltage module. These methods check at hardware level that at no time one of the two piezo DAC is OFF while the high voltage is ON. Possible errors are returned in the argument.

4. The PiezoWorker provides a `void activateDAC(uint8_t*)` as well as the opposite `void deactivateDAC(uint8_t*)` methods to safely activate and deactivate the two piezo DACs. The activation method uses the SPI bus to activate the internal reference voltage source of the DACs. The deactivation method does not need to use the SPI bus to turn off this internal reference source. If the SPI bus returns an error, the error value is returned in the argument.

5. The PiezoWorker provides a `void setXcom(uint16_t)` and a `void setYcom(uint16_t)` methods to change the target values of the command on X and Y axis for the piezo. These commands do not change the command applied to the piezo. A call to `applyCommand` is necessary to do so.

6. The PiezoWorker provides a `void applyCommand(uint8_t*)` method which uses the SPI bus to set the actual command of the two piezo DACs to the values contained in the attributes `xcom` and `ycom`. If the SPI bus returns an error, the error value is returned in the argument.

HkWorker

1. The HkWorker class provides a `void configure(void)` method to configure all the GPIOs and ADCs involved in the acquisition of HK data. This method also initializes

the 3 temperature sensors using the I2C bus. If an error occurs on this bus, its value is returned in the argument.

2. The HkWorker class provides a `void configureTemperatureSensor(uint8_t, uint8_t*)` method which can be used to configure any of the three sensors (given in the first argument) using the I2C bus. If the bus returns an error, its value is put in the second argument.

3. The HkWorker class provides a `void collectHk(uint8_t*)` method to retrieve all the different HKs of the payload board. This method uses the I2C bus (for temperature sensors). If an error occurs, its value is returned in the argument.

4. The HkWorker class provides a `void activate(void)` and a `void deactivate(void)` methods to activate or deactivate to collection of HK data.

ObcInterfaceWorker

1. The ObcInterfaceWorker class provides a `void configure(uint32_t, uint32_t)` method to configure the UART4-RS422a link. This method configures all the related GPIOs (PC9, PC10, PC11, PC13). The UART4 is configured to send 8-bit words, with 1 stop bit, with an oversampling of 16. The first argument of this method is the frequency of the SYSCLOCK (in Hz), and the second one is the requested baud rate. This method also configures and enables the RX interrupt. Finally, this method also configures DMA TX management. At the end of the call, the RS422a is ready to use in both RX and TX.

2. The ObcInterfaceWorker provides a `void reconfigureBaudRate(uint32_t, uint32_t)` method which takes the SYSCLOCK frequency as first argument, and the requested baudrate as second argument, and reconfigures the UART4/RS422a line to achieve this baudrate.

3. The ObcInterfaceWorker provides a `void transmitDMA(uint32_t, uint16_t)` method which takes the beginning address of a data buffer as first argument, and a number of bytes as second argument, and transmits the given number of bytes from the beginning of the buffer through the UART4 line using DMA.

4. The ObcInterfaceWorker class provides a `void activate(void)` as well as the opposite `void deactivate(void)` methods to turn on or off the UART4/RS422a line.

GyroWorker

1. The GyroWorker class provides a `void configure(uint8_t, uint8_t)` method to configure the gyroscope system. This method initializes all the related GPIOs (PA0 and PA5), and configures the gyroscope through the I2C bus. The first value given in argument is directly loaded into the CTRL2 register of the gyroscope (which controls the different hardware filters), and the second is loaded into the CTRL4 register (which controls the acquisition frequency).

2. The GyroWorker class provides a `void activate(void)` and a `void deactivate(void)` methods to activate and deactivate the gyroscope through the I2C bus. These methods do not change the current configuration of the system.

3. The GyroWorker class provides a `getSpeeds(void)` method to request and retrieve the last speed measurements from the gyroscope through the I2C bus, and load them in a `uint16_t speed[3]` attribute. The speed measurements are sorted in the following order: x, y, z.

ClockWorker

1. The ClockWorker class provides a `void configureHse(void)` method to configure the HSE external oscillator. The method configures GPIOs PF0 and PF1, starts the oscillator, wait for its stabilization, but does not switch the external 20 MHz clock as the SYSCLK.

2. The ClockWorker class provides a `void switchToHsi(void)` as well as the opposite `void switchToHse(void)` methods to set the HSE or HSI as the main SYSCLK.

3. The ClockWorker class provides a `void switchToHsiPll(uint8_t)` as well as a `void switchToHsePll(uint8_t, uint8_t)` methods to set the HSE or HSI as the input clock source to the PLL, and the PLL output as the main SYSCLK. Each of these two methods first checks that the HSE or HSI oscillator is running correctly before switching to it. Then, they set the PLL as the main source clock, with a predivider value given by the first argument (in case of HSE), and a multiplier value given by the first (HSI) and second argument (HSE).

4. The ClockWorker provides a method `void configureRtc(6 x uint8_t)` to configure the RTC of the STM32 microprocessor and initialize it with a given date and time (the format is the following: first argument is the number of years, then months, days, hours, minutes, and seconds). This method accesses back domain registers to activate the RTC, to set the HSE clock as its input, and to load the value given in argument into the date and time registers. It then relocks the back domain registers, and starts the RTC.

Chapter 6

Single mode photometry data reduction

Content of this chapter

6.1 Single-mode injection and optical aberrations

6.1.1 Photometric measurements in PicSat

Equation D.4 gives the fiber injection ratio as a function of the electric field in the fiber plane and of the fundamental mode of the fiber. Combined with Equation A.14 which gives the expression of the fiber fundamental mode, and the Kirchoff integral (Equation A.6), which gives a way to calculate the electric field in the fiber plane, this Equation D.4 can be used to calculate the injection function $\eta(x, y)$, i.e. the injection ratio as a function of the fiber position in the focal plane.

For a circular aperture, with no obstruction, the injection function is nearly Gaussian (see Figure 4.1):

$$\eta(x, y) = \eta_0 \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{6.1}$$

In the PicSat instrument, the fiber does not stay centered on the star image, but is forced to move along a modulation pattern by the tracking loop (see Chapter 4 for details). Consequently, the PicSat instrument does not measure the photometry at the center of the injection function $\eta(0, 0)$, but rather at different positions along the modulation pattern. A PicSat photometric measurement can be described as a vector of size n equals to the number of points on the modulation pattern (typically $n = 10$):

$$C = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = I_{\text{star}} \times T \times \begin{pmatrix} \eta(x_1, y_1) \\ \vdots \\ \eta(x_n, y_n) \end{pmatrix} \tag{6.2}$$

with $(x_1, y_1), \dots (x_n, y_n)$ the different points of the modulation pattern, I_{star} the intensity of the star, and T a transmission coefficient, which can be ignored as long as the instrument is not used for absolute photometry.

Of course, the quantity of interest is the stellar intensity I_{star}. If the injection function is known, each component of the photometry vector C can be divided by the proper value $\eta(x_k, y_k)$ to obtain an estimate of the intensity of the star:

$$\begin{aligned} \hat{I}_{\text{star},1} &= \tfrac{c_1}{\eta(x_1, y_1)} \\ &\vdots \\ \hat{I}_{\text{star},n} &= \tfrac{c_n}{\eta(x_n, y_n)} \end{aligned}$$

These estimates can then be combined using a photon noise-weighted average to obtain the final estimate of the stellar intensity:

$$\begin{aligned} \hat{I}_{\text{star}} &= \sum_{k=1}^{n} \alpha_k \hat{I}_{\text{star}} \\ \alpha_k &= \frac{\eta(x_k, y_k)^{-\frac{1}{2}}}{\sum_{i=1}^{n} \eta(x_i, y_i)^{-\frac{1}{2}}} \end{aligned}$$

This estimate of the stellar intensity is a linear combination of all the measurements $(c_1, \dots, c_n)$. Thus, it can be described in a single operation in the formalism of linear algebra

by introducing a weight vector V given by:

$$V = n^{-1} \begin{pmatrix} \frac{\eta(x_1,y_1)^{-\frac{3}{2}}}{\sum_{i=1}^{n} \eta(x_i,y_i)^{-\frac{1}{2}}} \\ \vdots \\ \frac{\eta(x_n,y_n)^{-\frac{3}{2}}}{\sum_{i=1}^{n} \eta(x_i,y_i)^{-\frac{1}{2}}} \end{pmatrix} \tag{6.3}$$

and writing the linear combination as a projection against this vector:

$$\hat{I}_{\text{star}} = V^T C \tag{6.4}$$

The photometric points in the PicSat instrument are acquired at 1 kHz, and the stellar intensity estimate can theoretically be calculated at a frequency of $1/n$ kHz. But the space to ground communication link limits the amount of data which can be downloaded from the satellite. Thus, the photometric points are binned on each modulation position, over ~ 1 min, and this sets the time resolution of the stellar intensity estimate calculated from the PicSat data.

6.1.2 Fiber injection and Zernike polynomials

The estimate of the stellar intensity obtained using the projection against the vector V defined in Equation 6.3 can be calculated from the theoretical injection function, and is only valid for this particular function. Any variation of the instrumental injection function will result in an error on the estimated intensity.

The main sources of variations of the instrumental injection function are thermal deformations of the telescope. Changes in the shape of the primary mirror will distort the wavefront of the incoming light, and change the expression of the electric field in the plane of the fiber. These aberrations can be described in the pupil plane by using the so-called Zernike polynomials $Z_n^m(\rho,\theta)$ and $Z_n^{-m}(\rho,\theta)$, with $0 \leq m \leq n$, and (ρ,θ) the polar coordinates. These polynomials form a basis of the functions defined over the unit circle, and any deformation of the surface of the primary parabola can be described as a sum of scaled Zernike polynomials:

$$\Delta z(r,\theta) = \sum_{n,m} a_n^m Z_n^m \left(\frac{r}{R},\theta\right) + b_n^m Z_n^{-m} \left(\frac{r}{R},\theta\right) \tag{6.5}$$

with R the radius of the primary mirror, a_n^m the coefficients of the "even" Zernike polynomials, and b_n^m the coefficients of the "odd" polynomials. The even and odd polynomials are similar, to within a rotation of $\pi/2m$.

To properly calculate the injection function, this potential deformation of the surface of the parabola needs to be taken into account in the Kirchoff integral (Equation A.6). The deformation changes the expression of the electric field in the image plane, and thus changes the injection. Figure 6.1 gives the shape of the pupil plane aberration, of the Point-Spread Function (squared modulus of the electric field in the image plane), and of the injection function, for the second and third order polynomials. The Zernike polynomial of order 0 (Z_0^0, piston) has no impact on the PSF or on the injection function, and the two polynomials of first order (Z_1^1, Z_1^{-1}, tip/tilt) are compensated by the tracking algorithm.

6.1.3 Impact of optical aberrations on the photometric stability

Since the vector V used to calculate the naive estimate of the stellar intensity is only valid for the theoretical non-aberrated injection function, any variation of the shape of the parabola

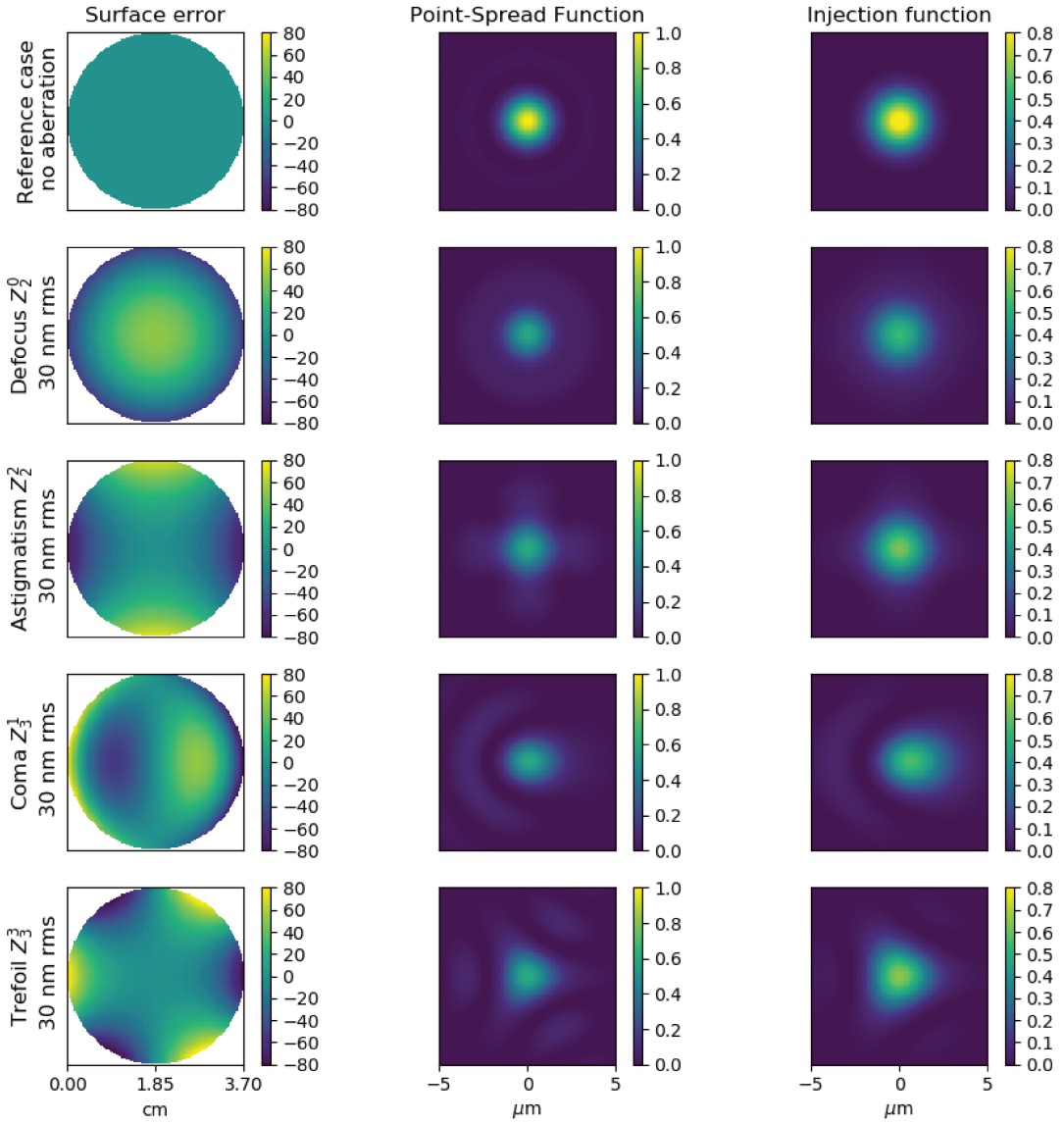


Figure 6.1: Shape of the aberration in the pupil plane (left column, scale in nm), of the Point-Spread Function (squared modulus of the electric field in the image plane, middle column), and of the injection function (right column), for the Zernike polynomials of order 2 and 3, calculated for the PicSat telescope, using the Kirchoff integral. This calculation was done for a monochromatic observation at 500 nm.

will bias this estimate. Figure 6.2 shows the bias on the estimated stellar intensity calculated using a weight vector V (Equation 6.3) corresponding to a non-aberrated injection function at 500 nm. This Figure shows that optical aberrations can have a strong impact on the photometry, with error of up to 10% for $\sim$ 30 nm rms of surface aberration. The effect of surface aberrations also depends on the modulation pattern selected for the tracking algorithm, with a circular modulation pattern apparenty less affected that the star modulation pattern.

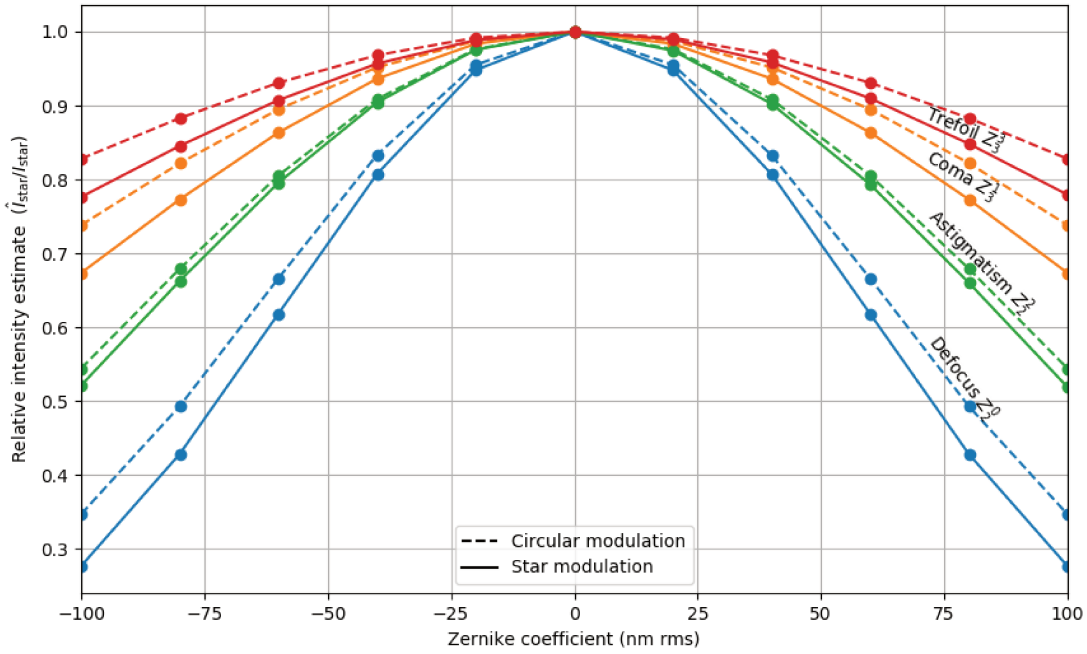


Figure 6.2: Evolution of the relative naive intensity estimate $\hat{I}_{\text{star}}/I_{\text{star}}$ with the level of optical aberration, for the first orders Zernike polynomials. These curves were calculated assuming an observation at 500 nm, and for two different modulation patterns: a 10 point circular modulation pattern of radius 0.5 FWHM (of the reference non-aberrated injection function), and a 10 point star modulation pattern with an inner circle of 0.2 FWHM and an outer circle of 0.5 FWHM.

6.2 Optical deformations of the PicSat telescope

6.2.1 Thermal study

While orbiting around the Earth, on its 505 km Sun-synchronous orbit, the PicSat nanosatellite is affected by significant thermal variations. Optical radiation from the Sun and the Moon, and infrared radiation from the Earth are absorbed by the satellite, which also radiates its own thermal flux. The different electronics inside the satellite dissipate energy, and also affect the overall thermal budget.

A complete thermal study was performed on PicSat, with the aim of estimating the temperature variations of the different subsystems of the satellite, over several orbital cycles. Different cases were considered, and the most realistic one ("cold case") is illustrated in Figure 6.3.

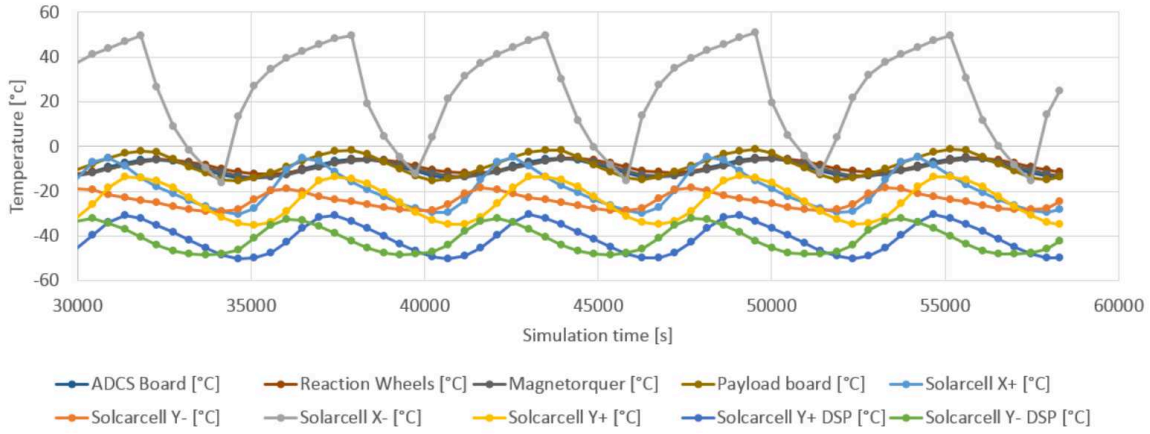


Figure 6.3: Temperature variations for different subsystems of the satellite, simulated over a few orbital cycles, in the cold-case scenario. Figure from Flecht [2016].

A detailed thermal study of the optical telescope was then subcontracted to ECM[1]. Using the general results obtained with the orbital simulation, this study focused on modeling the thermal deformations of the optical telescope, and in particular of the primary parabola. A few images illustrating the types of thermal deformations obtained are given in Figure 6.4.

6.2.2 Zernike fit of the primary parabola

In the detailed study, the surface of the primary parabola was modeled by a set of 13 points, for which the displacements induced by thermal variations were simulated over an entire orbital cycle.

I fitted the positions in 3D space of these 13 points using a model including the theoretical shape of the primary parabola, and additional Zernike polynomials. I performed a least-square optimization, minimizing the quantity:

$$\chi^2(a_n^m, b_n^m) = \sum_{k=1}^{13} \left(z_k - z_{\mathrm{M}_1}(x_k, y_k) - \sum_{n,m} \left[a_n^m Z_n^m \left(\frac{r_k}{R}, \theta_k \right) + a_n^{-m} Z_n^{-m} \left(\frac{r_k}{R}, \theta_k \right) \right] \right)^2$$

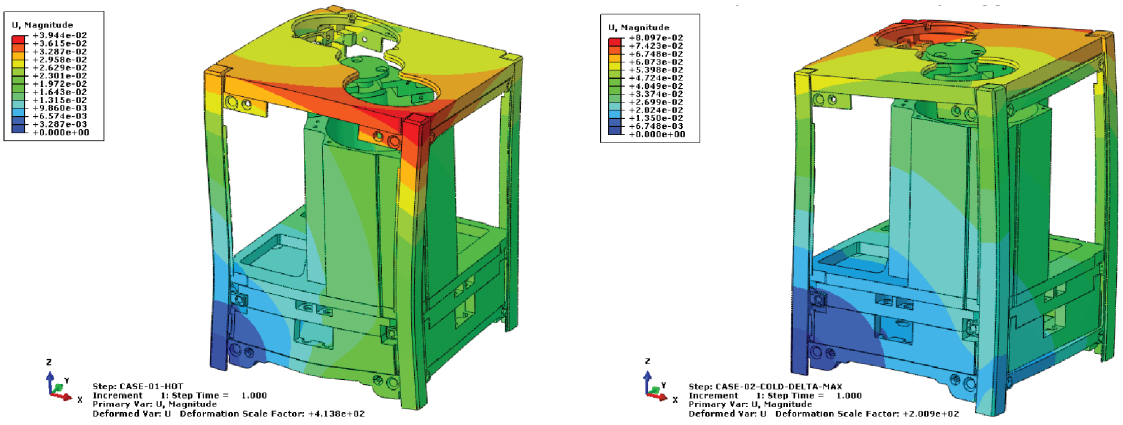
where (z_k, x_k, y_k) and $(z_k, r_k \cos(\theta_k), r_k \sin(\theta_k))$ are the Cartesian and cylindrical coordinates of the 13 points describing the surface of the parabola, and $z_{\mathrm{M}_1}(x_k, y_k)$ is the theoretical perfect shape of the primary mirror at these points.

The least-square fit was limited to the first few orders of optical aberrations ($n \leq 3$). and the results are presented in Figure 6.5. In all this work, the Zernike polynomials are normalized to a root-mean-square (rms) value of 1 nm over the 3.7 cm circular aperture of the telescope, and the coefficients a_n^m and b_n^m are always reported in units of nm rms.

6.2.3 Induced photometric variations

The direct consequence of the thermal variations of the surface of the primary mirror is that the naive stellar intensity estimate, which uses a weight vector set by the shape of the reference non-aberrated injection function is biased (see Figure 6.2). Variations of the thermal environment of the satellite lead to fluctuations of the photometry.

[1] http://www.ecm-be.com/



(a) Hot case (b) Cold case

Figure 6.4: Thermal deformations (exaggerated) of the PicSat optical telescope, in two different cases (hot case, no eclipse of the Sun over the orbit, and cold case, with eclipses). Figure by ECM.

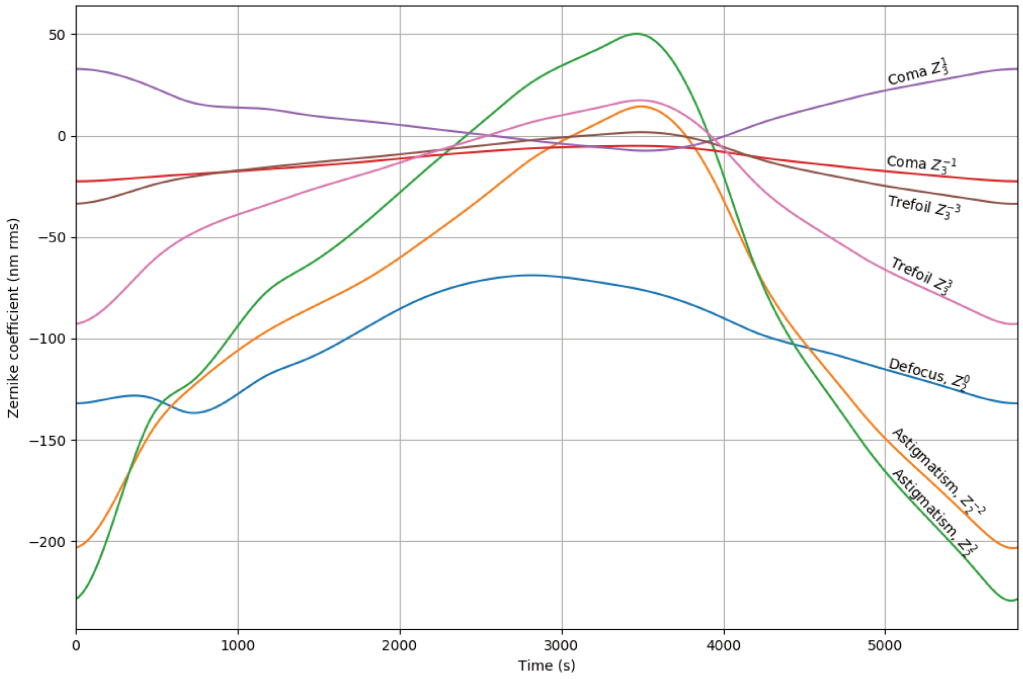


Figure 6.5: Variations of the Zernike coefficients a_n^m and b_n^m as a function of time for a complete orbit, obtained by fitting the results of the detailed thermal study performed by ECM ("cold case", corresponding to a case where the Sun is eclipsed by the Earth once every orbit).

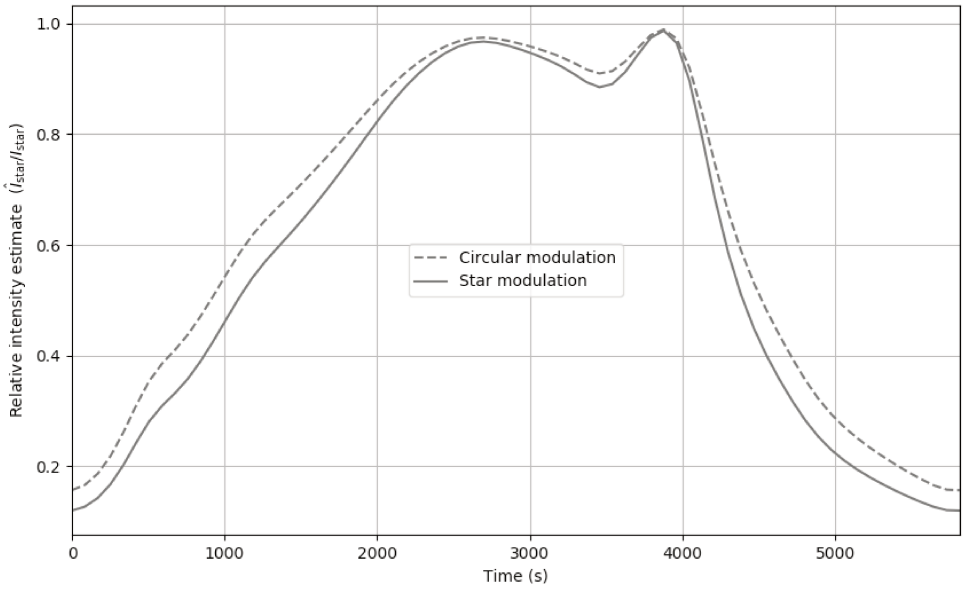


Figure 6.6: Typical photometric variations over a complete orbit (5817 s), calculated with the thermal deformations presented in Figure 6.5 (defocus term ignored), for two possible modulation patterns: a circle of radius 0.5 FWHM, and a star with an inner radius of 0.2 FWHM, and an outer radius of 0.5 FWHM.

The mechanical structure of the PicSat telescope, which supports the primary mirror, the secondary mirror, and which sets the distance between the pupil plane and the fiber, was entirely made in aluminum. This design was chosen in order to minimize the defocus: a global change in temperature of the optical telescope leads to homothetic deformations of the instrument, keeping the fiber in the focal plane of the primary parabola. Consequently, the variations of the Z_2^0 coefficient presented in Figure 6.5 should be compensated by the change in distance between the fiber, the secondary mirror, and the primary parabola.

Ignoring any tip/tilt variation (compensated by the tracking algorithm), and any defocus, the evolution of the naive estimate of the stellar intensity with time can be calculated over an entire orbit. The results are presented in Figure 7.3, for the two possible modulation patterns considered so far.

From Figure 6.6, it follows that without a proper data reduction scheme, the photometric data from the PicSat mission will be unusable for any science program. The impact of thermal fluctuations is simply too large, and the photometry is completely unstable.

6.3 Calibrating the photometry

6.3.1 Thermal fluctuations and thermal noise

Thermal fluctuations over the orbit of the satellite can introduce large variations of the measured photometry. But this may not be a problem, as long as these fluctuations are the same from one orbit to the other. The space environment is very stable, especially for a satellite kept in inertial pointing, and located on a Sun-synchronous orbit (the solar illumination angle is constant on the satellite). Furthermore, the orbital period T_{orbit} is only affected by residual atmospheric drag, and evolves on long timescales ($\sim$ 1 yr). Consequently, most of the pho-

tometric variations induced by the thermal fluctuations should be easy to remove by filtering out the orbital frequency and its harmonics from the signal.

But even though the space environment can be expected to be very stable, some random or pseudo-random factors can still affect the temperature of the satellite: power consumption of the different electronics, Earth's albedo, variations of the position of the Moon, etc. These pseudo-random phenomena affect the repeatability of the orbital thermal variations. To account for these effects, the time variations of the optical aberrations can be described using:

$$\begin{cases} a_n^m(t) = a_n^m(\phi(t)) + \alpha_n^m(t) \\ b_n^m(t) = b_n^m(\phi(t)) + \beta_n^m(t) \end{cases} \tag{6.6}$$

where $\phi(t)$ is the orbital phase at time t, and α and β are two random terms.

An order of magnitude for the repeatability of the temperature variations of the satellite from one orbit to another is $\sim 10\%$ (from other CubeSat missions, and confirmed by flight data from PicSat, see Figure 9.3). As an approximation, I model the random terms α_n^m and β_n^m in the above equation as random variables, with a standard deviation equals to 10% of the total amplitude of the variations of a_n^m and b_n^m:

$$\begin{cases} \alpha_n^m(t) \sim \mathcal{N}\left(0, \left[\max\left\{a_n^m(\phi(t))\right\} - \min\left\{a_n^m(\phi(t))\right\}\right]^2\right) \\ \beta_n^m(t) \sim \mathcal{N}\left(0, \left[\max\left\{b_n^m(\phi(t))\right\} - \min\left\{b_n^m(\phi(t))\right\}\right]^2\right) \end{cases} \tag{6.7}$$

In theory, these random noises α and β should show a strong short-term temporal correlation (i.e. $\alpha(t) \simeq \alpha(t + dt)$, for a small dt), and these noises should be characterized by their full frequency spectra. But this temporal correlation does not significantly affect the data reduction scheme proposed in the rest of this section, and will be ignored for the sake of simplicity.

From now on, I will refer to the $a_n^m(\phi(t))$ and $b_n^m(\phi(t))$ as "thermal fluctuations", and to α_n^m and β_n^m as "thermal noise".

With this definition, the thermal fluctuations are perfectly repeatible from one orbit to another, and can easily be removed by filtering the data. One simple way of doing so is to fold the data over a single orbital period, calculate a mean orbital photometry, and divide each point by this reference value.

6.3.2 Weight vector calibration

In Section 6.1, I introduced the weight vector V and the associated intensity estimate:

$$\hat{I}_{\text{star}} = V^T C \tag{6.8}$$

The weight vector V can be defined using the non-aberrated injection function, as in Equation 6.3. But in this case, the intensity estimate is biased in the presence of optical aberrations. If the optical system is aberrated, another weight vector should be used, defined from the aberrated injection function η_{ab}:

$$V_{\text{ab}} = n^{-1} \begin{pmatrix} \frac{\eta_{\text{ab}}(x_1,y_1)^{-\frac{3}{2}}}{\sum_{i=1}^n \eta_{\text{ab}}(x_i,y_i)^{-\frac{1}{2}}} \\ \vdots \\ \frac{\eta_{\text{ab}}(x_n,y_n)^{-\frac{3}{2}}}{\sum_{i=1}^n \eta_{\text{ab}}(x_i,y_i)^{-\frac{1}{2}}} \end{pmatrix} \tag{6.9}$$

This weight vector can theoretically be calculated by careful thermal and optical modeling of the system. In practice, though, this can prove extremelly difficult, and a better approach is to calibrate it.

Keeping the linear algebra formalism, we can introduce an "injection vector", which we denote H, defined as:

$$H = \begin{pmatrix} \eta(x_1, y_1) \\ \vdots \\ \eta(x_n, y_n) \end{pmatrix} \tag{6.10}$$

This injection vector depends on the time, through the injection function which itself depends on the evolution of the optical aberrations: $H = H(t)$. Taking into acount the separation of optical aberrations between a "fluctuation" term and a "noise" term (see Equation 6.6), we can neglect the "noise" part, and consider that the injection function depends on the orbital phase, through the thermal fluctuations (the $a_n^m(\phi)$ and $b_n^m(\phi)$ terms in Equation 6.6): $H = H(\phi)$. The difference between $H(t)$ and $H(\phi)$ comes from the thermal noise, and we can write the first order approximation:

$$H(t) = H(\phi) + \sum_{n,m} \alpha_n^m \frac{\partial H}{\partial a_n^m}(\phi) + \beta_n^m \frac{\partial H}{\partial b_n^m}(\phi) \tag{6.11}$$

If we consider a sequence of data, acquired at time $t_0, \ldots, t_N$, we can denote $\phi_0, \ldots, \phi_N$ the associated orbital phases, and $C_0, \ldots, C_N$ the measured photometric vectors. We have the relation:

$$C_k = I_{\text{star}}(t_k) \times \left[H(\phi_k) + \sum_{n,m} \alpha_n^m(t_k) \frac{\partial H}{\partial a_n^m}(\phi_k) + \beta_n^m(t_k) \frac{\partial H}{\partial b_n^m}(\phi_k) \right] \tag{6.12}$$

The photometric measurements can be grouped by orbital phases, i.e. we can create groups of indices k such that all of the associated orbital phases are equals[2]. Since ϕ is constant in such a group, the values of $H(\phi)$ and its derivative are also constant. If the sequence is long enough (a few 100 orbits), and if the stellar intensity is constant over this sequence, the mean of the C_k should be proportional to $H(\phi)$:

$$\text{mean}(\{C_k, \quad \forall k, \phi(t_k) = \phi\}) \propto H(\phi) \tag{6.13}$$

This mean value contains all the values $\eta[\phi](x_k, y_k)$ required to calculate the aberrated weight vector $V(\phi)$ as in Equation 6.9, to within a multiplicative normalization coefficient.

To summarize, it is possible to use a calibration scheme to calculate a set of unbiased weight vector, by doing the following:

- Observe a stable target during ~ 100 orbits.
- Group the photometric points per value of orbital phase ϕ, and calculate the mean of each group. Since the injection function is stable at constant orbital phase, this gives an estimate of the injection at each modulation position $\eta(x_k, y_k)$, valid for the aberrations at orbital phase ϕ.
- Using the injection values, calculate a weight vector $V(\phi)$ valid for the aberrations at orbital phase ϕ using Equation 6.9.

[2]The photometric data are downloaded from PicSat with a time resolution of 1 min, and the orbital is ~ 90 min, which means that we can create about 90 such groups.

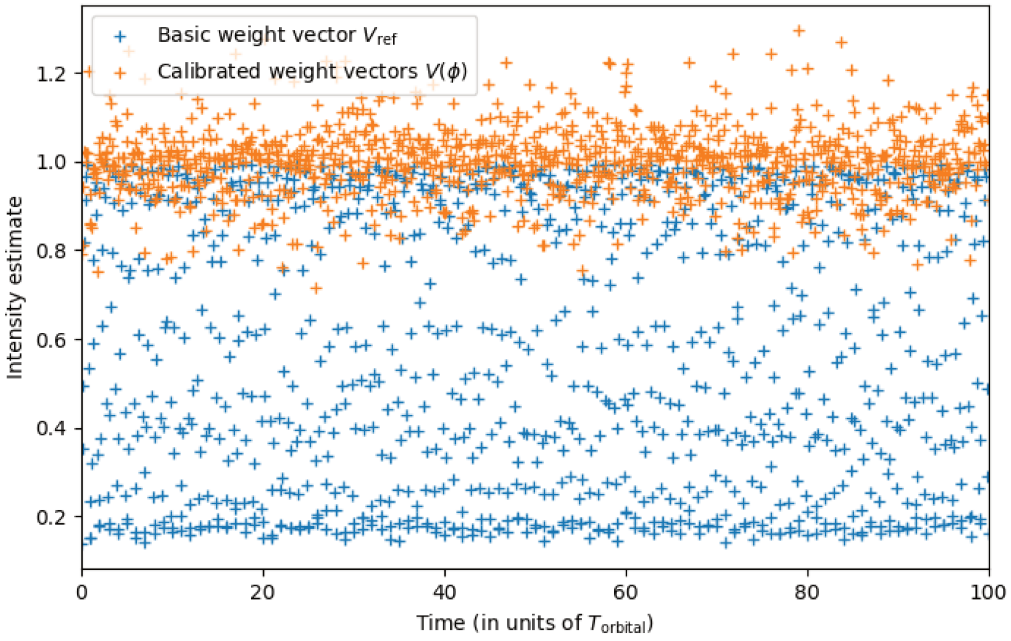


Figure 6.7: Comparison between the photometry reduced with the reference weight vector (not taking into account the fluctuations of the optical aberrations), and the calibrated noise vectors. Data were simulated using the Kirchoff integral, with optical aberrations defined as in Equations 6.6 and 6.7, at $\lambda = 500$ nm, for a circular modulation pattern of radius 0.5 FWHM. Each point represents 1 min of data, affected by photon noise. The simulations was done for 10 different orbital phases, and 100 orbits.

The results of this procedure over a simulated sequence obtained using a noise structure given by Equations 6.6 and 6.7 (plus photon noise) are given in Figure 6.7. This calibration scheme is useful to remove the photometric bias introduced by thermal fluctuations, but the noise level in the reduced data is still too high, due to thermal noise (1%/hr).

6.3.3 Optimum weight vector and PCA calibration

Although the calibration of the weight vectors used to calculate the photometry can already smooth out the huge fluctuations of the photometry induced by the thermal environment of the satellite, the precision reached is not enough for the ambitious science objectives of the PicSat mission.

The calibrated weight vectors $V(\phi)$ guarantees that the photometry obtained is not biased by the thermal fluctuations, but it does nothing toward optimizing the Signal-to-Noise Ratio (S/N). In fact, given their definition, these calibrated weight vectors do not factor in any information on the structure of the thermal noise. To optimize the S/N of the reduced photometry, while keeping the simple linear algebra formalism used so far, it is necessary to perform a Principal Component Analysis (PCA) on the data.

We consider a given orbital phase ϕ, for which a number of photometric points are available: $C_1, \ldots, C_p$ (p being the number of orbits worth of data which are reduced at once). The

goal is to find the optimum weight vector V_{opt}, which means:

$$\text{mean}\left(V_{\text{opt}}^T C_k\right) = 1 \qquad \text{(unbiased, to within a normalization factor)} \tag{6.14}$$

$$\text{cov}\left(V_{\text{opt}}^T C_k\right) \text{ is minimum} \tag{6.15}$$

Finding this optimum weight vector is straightforward using a principal component analysis. The measurement covariance matrix $W = \text{cov}(C_k)$ is symmetric and positive, and thus can be diagonalized in an orthonormal basis. We can introduce the positive eigenvalues $\lambda_1, \ldots, \lambda_n$, and the orthonormal eigenvectors $V_1, \ldots, V_n$ (n being the length of the individual vectors C_k, set by the number of modulation points):

$$WV_i = \lambda_i V_i$$

$$V_i^T V_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

The optimum weight vector can be written as an arbitrary sum of these vectors:

$$V_{\text{opt}} = \sum_{i=1}^{n} v_i V_i \tag{6.16}$$

The mean photometry obtained with this weight vector is given by:

$$\hat{I}_{\text{star}} = \sum_{i=1}^{n} v_i \gamma_i$$

With:

$$\gamma_i = \text{mean}\left(V_i^T C_k\right) \tag{6.17}$$

And the final measurement variance can be written:

$$\sigma_I{}^2 = \sum_{i=1}^{n} v_i{}^2 \lambda_i \tag{6.18}$$

With these notations, the unbias and optimum conditions of Equations 6.14 and 6.19 writes:

$$\begin{cases} \sum_{i=1}^{n} \gamma_i v_i & = 1 \\ \sum_{i=1}^{n} \lambda_i v_i{}^2 & \text{is minimum} \end{cases}$$

These two equations define a constrained minimization problem, easily solved by introducing and optimizing the Lagrange multiplier $\mathcal{L}(v_i, \mu) = \sum_i \lambda_i v_i{}^2 - \mu\left(1 - \sum_i \gamma_i v_i\right)$:

$$\frac{\partial \mathcal{L}}{\partial v_k} = 2\lambda_k v_k - \mu v_k = 0 \quad \Rightarrow \quad v_k = \frac{\mu \gamma_k}{2\lambda_k}$$

$$\frac{\partial \mathcal{L}}{\partial \mu} = 1 - \sum_{i=1}^{n} \gamma_i v_i = 0 \quad \Rightarrow \quad \frac{\mu}{2} = \left(\sum_{i=1}^{n} \frac{\gamma_i{}^2}{\lambda_i}\right)^{-1}$$

Which gives the solution:

$$v_k = \frac{1}{\sum_{i=1}^{n} \frac{\gamma_i{}^2}{\lambda_i}} \times \frac{\gamma_k}{\lambda_k} \tag{6.19}$$

To summarize, a set of optimum weight vectors can be obtained from a calibration observation sequence, by doing the following:

- Group the photometric points per orbital phase ϕ.
- For each group, perform a principal component analysis to obtain the vectors $V_k(\phi)$ and the eigenvalues $\lambda_k(\phi)$ of the covariance matrix.
- For each group, use the entire set of data points $C_k(\phi)$ to calculate the values $\gamma_i(\phi) = \text{mean}\left(V_i^T C_k(\phi)\right)$.
- For each orbital phase, form the optimum calibrated weight vector $V_{\text{opt}}(\phi)$ using Equation 6.19.
- The set of $V_{\text{opt}}(\phi)$ can then be used to reduce another dataset. If the noise structure of the second dataset is similar to the first one, the PCA guarantees that the reduced photometry is unbiased, with minimal variance.[3]

The result of this PCA calibration obtained on the same data sequence as presented in Figure 6.7 is given in Figure 6.8. The resulting photometric stability is $\simeq 0.1\%$/hr. But this result, calculated over the entire sequence, hides huge disparities between the different orbital phases. A box plot obtained by folding the data, and grouping the points per orbital phase, clearly reveals that all orbital phases are not equivalent in terms of data quality (see Figure 6.9). Of course, this is related to the fact that the aberrations of the optical telescope are significantly lower at certain orbital phases.

The overall photometric stability can be lowered to $\simeq 500$ ppm/hr simply by rejecting the less favorable orbital phases, and keeping only the data associated to an orbital phase $80 \text{ deg} \leq \phi \leq 240$ deg.

6.4 Conclusion

In a classical photometric instrument, which uses a two-dimensional detector in the focal plane of a telescope, the photometry is calculated by integrating all pixels which are receiving photons from the star. If the PSF is deformed by optical aberrations, the integration box can always be enlarged, to ensure that all photons coming from the star are properly accounted for. The detector noise is increased, but the photometry is not biased by the optical aberrations of the telescope. In the case of a single-mode instrument, the number of photons reaching the detector depends on the telescope to fiber coupling, which varies with optical aberrations. This makes fibered instruments much more sensitive to optical aberrations.

To a large extent, the results of this chapter remain highly speculative, as no real data was obtained to test the data reduction method. But despite this, the results obtained show, both from a theoretical and a simulation standpoints, that it is possible to obtain a stable photometry with a fibered instrument. Yet, strictly speaking, the objective in terms of photometric stability of the PicSat mission is not reached. Taking into account the large optical deformations induced by thermal variations over the orbit of the satellite, and using a PCA based data reduction, the performance achieved is $\simeq 500$ ppm/hr, which already guarantees two of the main science objectives of the mission: β Pic b transit detection, and the technical demonstration (see Section 2.1). The noise budget of the fibered photometer used in PicSat is not dominated by tracking errors or by photon noise, but rather by the optical aberrations coupled to the fiber injection.

[3] Note that this does not mean that this data reduction is the best data reduction possible. The weight vectors give the optimal linear combination of the different modulation points. But a more complex data reduction, involving non-linear operations could potentially yield better results.

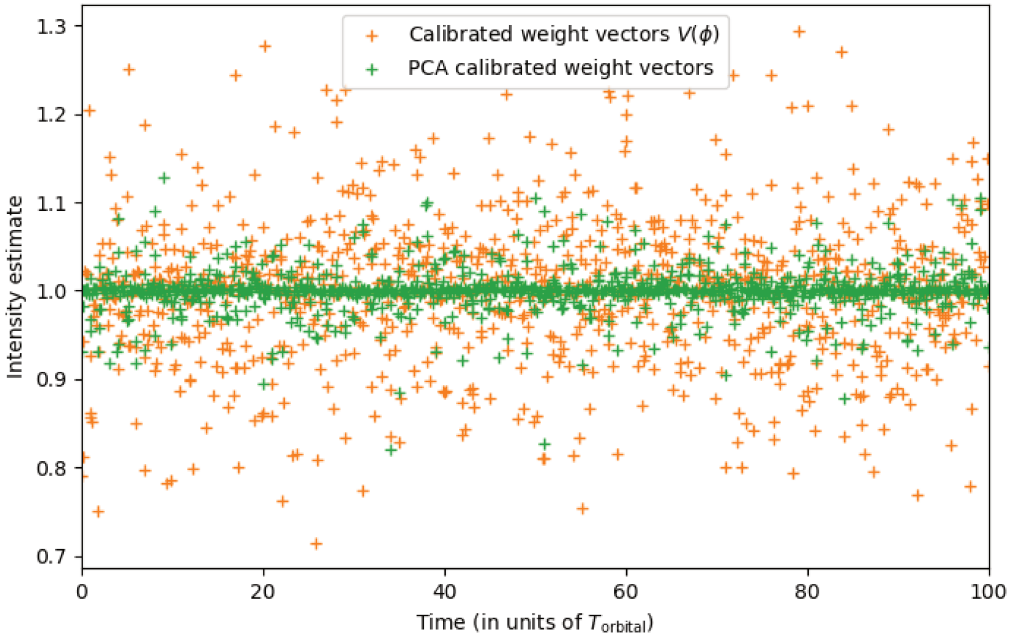


Figure 6.8: Same as Figure 6.7, but comparing the photometry obtained using the simple calibrated weight vectors of Section 6.3.2 and the PCA calibration.

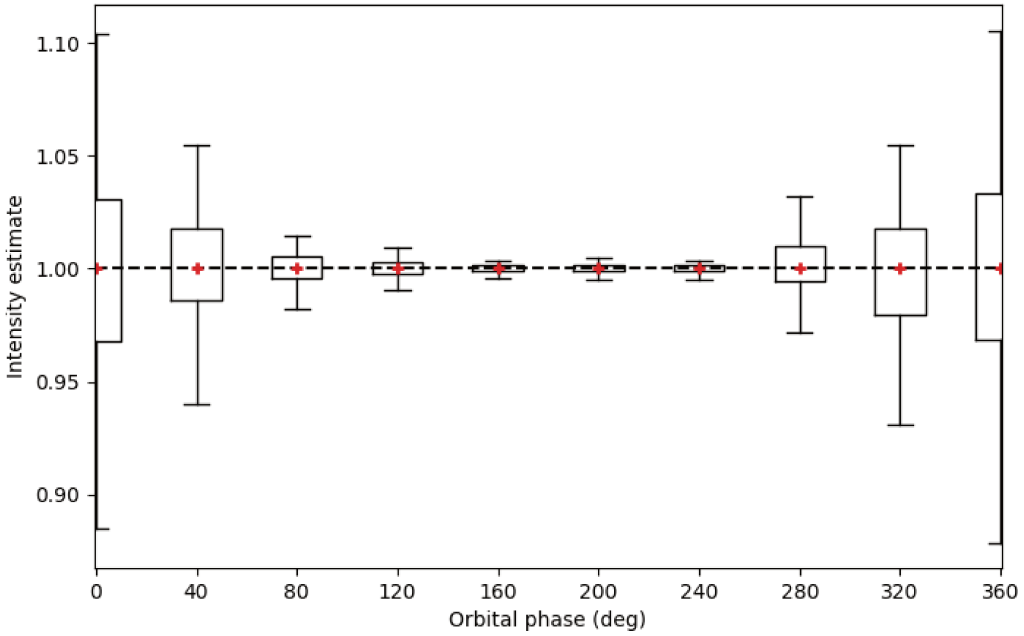


Figure 6.9: Box plot obtained by folding the PCA reduced photometric data (green crosses in Figure 6.8) over a single orbit, and grouping the points per orbital phase. The red crosses give the mean value on each orbital phase. The rectangular boxes give the lower and upper quartiles, and the vertical lines extends from the lowest to the highest values.

In order to reach the objective of 100 ppm/hr of the mission, a more careful design of the optical telescope seems necessary. The initial design choice of building the entire telescope in aluminum to minimize optical aberrations may have been an error. This only guarantees homothetic deformations for a constant temperature throughout the entire instrument. And the large thermal expansion coefficient of aluminum $\alpha \sim 20 \times 10^{-6}$ K^{-1} makes the whole setup very sensitive to thermal gradients. Simulations show that it is necessary to reduce the optical aberrations given in Figure 6.5 by a factor ~ 5 to reach the target objective of 100 ppm/hr in terms of photometric stability. Designing the instrument using materials with lower thermal expansion coefficients (Invar, $\alpha \sim 1$ to 5×10^{-6} K^{-1}, Zerodur, $\alpha < 10^{-7}$ K^{-1}), and/or ensuring that strong thermal gradients can be avoided is key to reach this level of photometric precision with a fibered instrument.

Chapter 7

Test campaigns

Content of this chapter

Date	Type	Hardware tested	Objective	Result
Jan 2016	Vibrations	Electronics, SPAD, piezo	Test non-qualified hardware	Dead SPAD (during transport?), broken piezo
Apr 2016	Vibrations	Electronics, SPAD	Test SPAD (not tested previously)	SPAD resisted vibrations
Apr 2016	Thermal vac.	Electronics, SPAD (not regulated)	Test electronics	No heat point on electronics
Jan 2017	Vibrations	Satellite QM	Check resistance of optical alignment	Alignment resisted to vibrations
Feb 2017	Thermal vac.	Satellite QM	Check thermal regulation of photodiode	Broken M2, strong variations of photometry
Jun 2017	Thermal vac.	Satellite QM, new M2 and elec.	Verify new M2 and elec.	M2 was OK, less photometric variations
Aug 2017	Thermal vac.	Satellite FM	Verify flight model	Success
Oct 2017	Shocks	Payload QM	Verify optical alignment	Optical alignment withstand shocks

Table 7.1: Description of the 8 test campaigns carried out on the PicSat payload. The objectives and results listed in this table relate to the payload only. This table does not include the acceptance tests performed on the FM before delivery.

7.1 Payload test plan

Space is a notoriously harsh environment. The combination of strong thermal gradients due to the Sun illumination, quick thermal variations from the orbital motion of the satellite, and vacuum, creates a thermally dangerous environment. Radiations from the Sun or from cosmic rays, unshielded by the atmosphere, are all dangerous threats for electronic systems. And during launch, vibrations from the rocket and strong shocks resulting from stage separations can easily damage mechanical elements inside the satellite.

The science objectives of the mission did not require the satellite to operate for several years, and long-term reliability was not a critical issue. So, in order to cut cost and development time, the PicSat payload was developed using conventional electronics, and some non space-qualified components. To compensate for some of the risks inherent to this approach, a total of 8 test campaigns were carried out on the payload, spread over all development stages. These test campaigns and their objectives are described in Table 7.1.

7.2 Vibrations

The electronic board, the photodiode, and the piezo actuators initially chosen for the payload were not space-qualified components. A first vibration test campaign was thus performed early on, a month after having received the different elements. The main objective of this campaign was to verify that the piezo stage and the photodiode could withstand vibrations at a level representative of the launch environment. Unfortunately, the photodiode was not working properly when it arrived at the test facility, so this component could not be tested. But this test campaign revealed a major problem with the piezo stage, which did not fully survive the test: the displacement range on one of the two axes had been significantly reduced by the vibrations.

The consequences of the failure of the piezo stage were heavy: the payload was completely redesigned to accommodate a new and bulkier two-axis piezo actuator, which was based on space-qualified components. To give a sense of the importance of the design change, I show in Figure 7.1 a 3D overview of the first design of the payload, to be compared to Figure 3.2. In

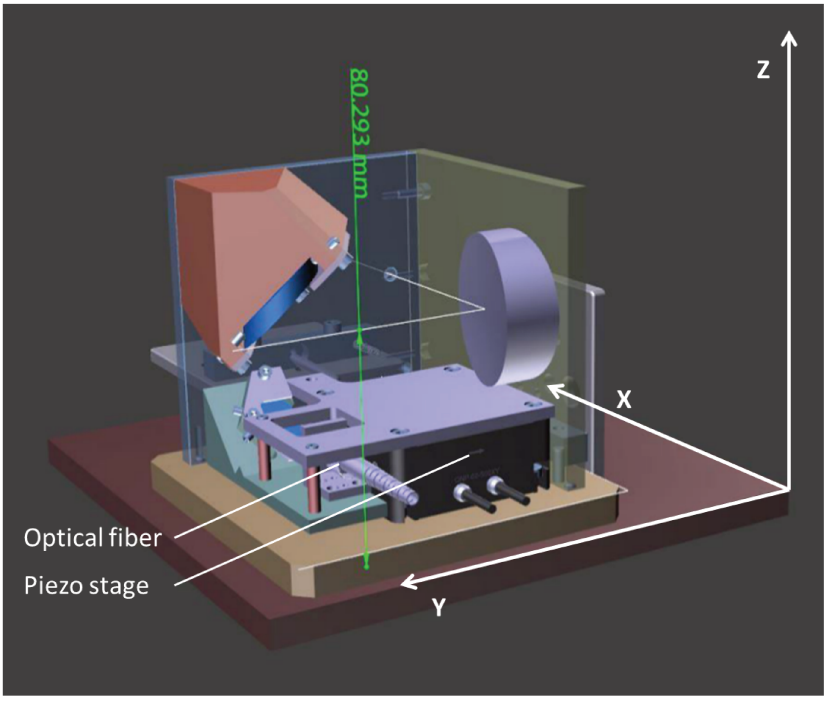


Figure 7.1: A 3D rendering of the initial design of the PicSat payload, which had to undergo drastic changes after the failure of the piezo stage in vibrations.

this initial setup, the optical axis of the telescope was oriented on the Y axis of the satellite (i.e. the side), not along the Z axis. The optical fiber was attached to the side of the piezo actuator, which was $\sim 30\%$ smaller than the new piezo stage (in each dimension).

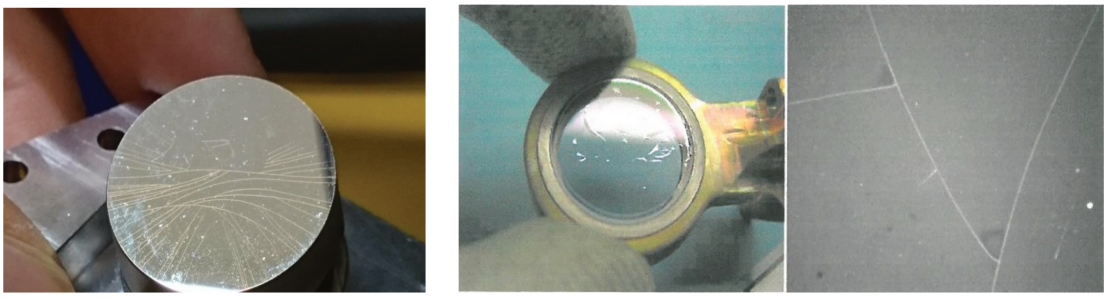
Subsequent vibration campaigns (with the new piezo, and then on the complete satellite qualification model) went without any major issue. One of the most important objectives of these tests was to verify that the optical alignment could resist the launch environment. All tests performed on the payload indicated that it would.

7.3 Thermal vacuum

Multiple tests were performed in a thermal vacuum chamber with the PicSat payload, in order to verify the photometric stability of the instrument. During these test campaigns, three major issues were encountered, each one requiring up to several months to be solved.

Secondary mirror coating The M2 mirror used in the PicSat optical telescope is a 22 mm diameter plane mirror. The initial version had a simple aluminum coating on a Zerodur mirror. During the second thermal test campaign, performed in February 2017 (6 months before the expected launch date at the time, and 1 yr before the actual launch date), the mirror failed (see Figure 7.2a). A review of the literature revealed that the most likely cause was a failure of the coating (see Figure 7.2b), and a new mirror was ordered, with a coating adapted to the space environment. This adds to the failure of the piezo stage in vibrations in showing the importance of careful testing when using non-space qualified components.

Temperature induced photometric variations Large variations of the photometry recorded by the photodiode on a stable source were noticed during the February 2017 thermal vac-



(a) Secondary mirror failure

(b) Coating failure from literature

Figure 7.2: A picture of the secondary mirror (M2) of the PicSat instrument after the thermal vacuum test campaign of February 2017, and a similar example of coating failure found in the literature (Wernham [2011])

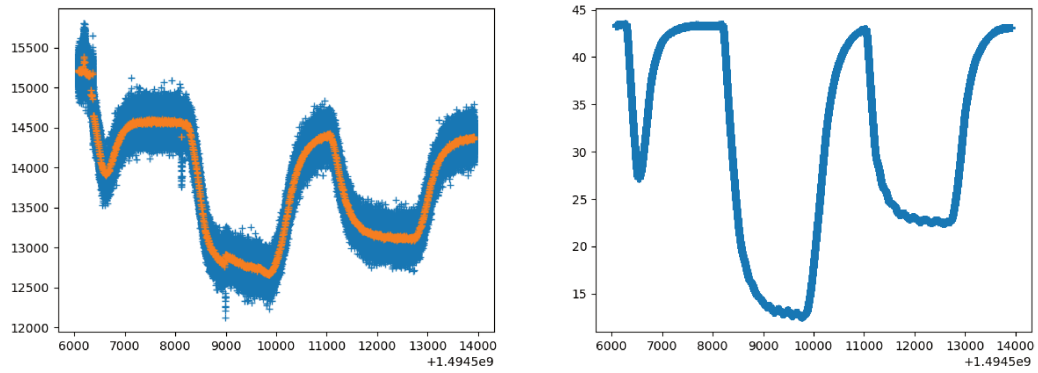


Figure 7.3: The obvious correlation between the photometry measured on a stable source and the temperature of the electronic board. Left panel: photometry (arbitrary unit) versus time (s). The blue points are the raw data, acquired at high-frequency. The orange points represents the same data, binned over 10 points. Right panel: temperature (°C) versus time (s). All data acquired during the February 2017 thermal vacuum test campaign.

uum campaign. These variations were strongly correlated with the temperature recorded by the sensors on the electronic board (see Figure 7.3). Smaller variations correlated with the high-voltage command applied to the piezo stage were also noted during this test campaign, indicative of an electrical problem. Further investigation confirmed this hypothesis, revealing a design error caused by a misunderstanding between our team and IDQuantique, the manufacturer of the photodiode. The photodiode is connected to two voltage lines: a $V_{\mathrm{DD}} = 5\ V$ line, and a polarization voltage. The gain of the detector is directly related to the polarization voltage. For PicSat, this voltage was regulated at $V_{\mathrm{OP}} = -25$ V, with a precision regulator (error of $< 20\ \mu$V). However, it turned out that the internal wiring of the photodiode was designed in such a way that the effective polarization voltage was not V_{OP} , but $V_{\mathrm{OP}} - V_{\mathrm{DD}}$. Thus, any variations caused in the main 5 V line of the payload board directly translated into photometric variations. Since both the piezo stage and the thermo-electric cooler were connected to the main 5 V, this caused photometric variations correlated with the piezo command, and with the temperature.

A new version of the electronic board corrected this problem. But further testing showed

that thermal variations of the photometry had not completely disappeared. A long and tedious investigation, involving several days of testing in a climatic chamber, as well as days of testing with a hair-dryer (heating different parts of the payload one at a time) finally led me to the conclusion that the most likely cause of these remaining variations was thermal deformations of the photodiode, and in particular of its internal optics used to couple the fiber with the single-pixel detector.

Having reached this conclusion, I built a small setup to test the 5 photodiodes still available from the initial batch delivered at the beginning of the project (excluding diodes extensively tested in thermal vacuum, in the climatic chamber, in vibrations, etc.). One photodiode was found to have small and linear temperature variations, with no hysteresis. This detector was mounted on the flight-model of the payload, and is now orbiting the Earth.

Hair-dryer bug The last problem uncovered during thermal testing of the payload was a software bug. This bug in the on-board software was particularly difficult to trace down, as it would only manifest itself in certain situation of high-processor load and increased temperature of the electronic board. Several problems (communication errors with the payload board, unexpected crashes of the software, etc) were encountered during thermal vacuum test campaigns, or even during vibration test campaigns (due to the heat generated by the vibrating platform) which were linked to other causes, but which most probably were all related to this bug.[1]

7.4 Radiations

Due to the complexity involved in such tests and the low requirement on the mission lifetime (< 1 yr was enough to follow the potentail transit), we did not perform any dedicated radiation test on the satellite. Except for the careful selection of the single-mode fiber (with a pure silica core, see Chapter 3), the material and electronics selected for the payload were all standard components, with no particular radiation tolerance.

7.5 Optical alignment

7.5.1 General considerations

The optical alignment campaign for the flight model of PicSat took place during summer 2017. The instrument is simple, with only two mirrors and 3 adjustment points. But several particularities made it extremely difficult to align.

First, the PicSat photometer was a fibered instrument. The combination of the off-axis parabola and the single-mode fiber made the instrument extremely sensitive to alignment errors. And the fact that fiber could not be replaced by a camera made difficult to have a general picture of the focal plane. The tracking algorithm, designed to compensate for the

[1] For the reader interested in all the juicy details of embedded programming: the STM32F3 microcontroller can be clocked at different frequencies using an internal or external oscillator, and a Phase-Locked Loop frequency multiplier. But when increasing the clock frequency, it is also necessary to instruct the microcontroller to add wait cycles when accessing the internal Flash memory, to account for any latency. Failure to do so can result in unstable read access to the memory. In the case of the PicSat payload, at lab temperature, the latency was small enough for the processor to be able to cope with it, which is why the problem only appeared in thermal vacuum tests. The higher temperature would then slightly change the physical characteristics of the memory, to a point where the microcontroller would simply crash without the proper wait cycles. I found this bug when heating different parts of the payload with a hair-dryer to look for the cause of the photometric variations, hence the name. It probably took me at least 10 days, full time, to find this bug...

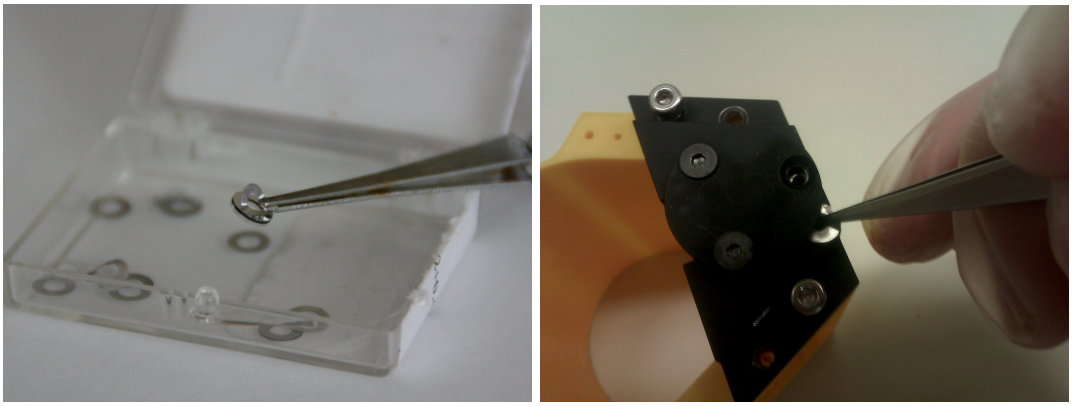


Figure 7.4: Left panel: picture of one of the peelable washers used for aligning the PicSat instrument. Right panel: The author of this thesis is trying to find the best way to add a few layers of washer back on one of the three spacers suporting the secondary mirror mount. Picture taken during a training exercise, on a 3D printed copy of the telescope.

small jitter of the satellite was unable to keep the fiber centered on the star while I was adjusting the tip/tilt or focus. I had to resort to using the "scanning mode" only. In this mode, the fiber constantly scans the focal plane, creating a virtual "image". But the process is slow, and depending on the resolution of the scan, it can take up to several minutes to take an image. The direct consequence is that the whole alignment process is extremely slow and tedious: move a screw half a turn one way, wait a few minutes for the scan to finish, move the screw back a quarter of a turn, wait again, etc.

Secondly, due to the design of the telescope, the alignment was made on the three mounting points of the secondary mirror. Three screws were used to fix the support of the secondary mirror on the head block, itself mounted at the end of the optical tube. On each of these three screws, a spacer was used to adjust the height of the secondary mirror. By adjusting the height of these spacers, it was possible to adjust the tip, tilt, and focus of the instrument. The spacers used for PicSat were "peelable". At first glance, these peelable spacers looked very much like normal washers, except that they had an aluminum side, and a transparent plastic side. In reality, though, these spacers were not made of a single piece of material, but of multiple thin layers stacked together (10 μm on the aluminum side, 50 μm on the plastic side). To adjust the height, one needed to remove the washer from the telescope, and with a scalpel, or any other sharp tool, carefully remove one layer (i.e. "peel" the spacer). Simple enough, except that it turned out surprisingly difficult to reliably remove a single layer. And not only that, but optical alignment is often an iterative process, which means that sometimes it was necessary to remove a layer, and sometimes it was necessary to *add* a few *back*. Then it was like walking a tightrope: carefully adding the 10 μm-thin layers with small pliers, hoping than none would fall down the optical tube onto the primary mirror[2]...

To make the alignment at least possible, I had to find a general procedure minimizing the number of manipulations required with the peelable washers.

[2]Yes, aligning the PicSat payload was an excruciatingly difficult and tedious exercice. By far the most difficult task I had to do during this thesis...

7.5.2 Alignment with springs

Replacing the peelable washers with springs, it was then much easier to change the height of the 3 secondary mirror mounting screws. With some training, it was even possible to manipulate the screws without loosing the fiber tracking. So I decided to use only springs during the first step of the alignment procedure.

The objective of this first step was to adjust the tip/tilt, in order to align the parabola on the center of the piezo-actuator. Off-axis parabolas generate astigmatism when observing off-axis targets, which reduce the fiber injection factor. Thus, it was necessary to align the parabola to have a central part of piezo field-of-view free of astigmatism.

The procedure I used was the following:

- Using a camera, and removing the secondary mirror mount completely, I first aligned a collimated source on the parabola by minimizing the astigmatism (Figure 7.5a). When the collimator is not aligned with the parabola, the two foci (sagittal and transverse) are clearly visible by changing the focus of the camera. I adjusted the tip/tilt of the collimated source until the two foci converged to a single one.

- Then I added back the secondary mirror, with springs in place of the peelable washers, and I adjusted the tip/tilt to have the image of the collimated source at the center of the piezo (see Figure 7.5b for the setup). At this point, the instrument is aligned in tip/tilt.

- Using the springs, it is possible to fine-tune the alignment of the instrument by iterating the following sequence: 1- Adjust the focus of the telescope by changing the height of all three screws while maintaining the image of the collimator still in the piezo field, until the injection is at a maximum; 2- Change the tip/tilt of the collimated source by a given quantity. This moves the image at another point in piezo field. 3- Adjust the tip/tilt of the secondary mirror to re-center the image. 4- Adjust the focus again, to reach maximum injection. 5- If the injection is better than at step 1, this means that the alignment of the collimator with the primary parabola has improved. Iterate the process again, starting from this new position. Otherwise, adjust the collimator back to its original position, and try adjusting in the opposite direction.

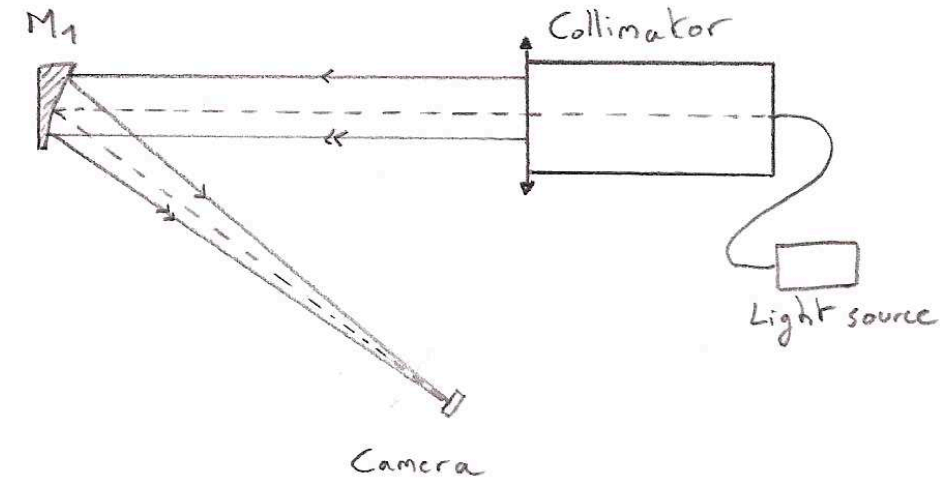
At the end of this procedure, the instrument is aligned. But wait! What about those springs? Of course they cannot go to space...

7.5.3 Introduction of peelable washers and tip/tilt alignment

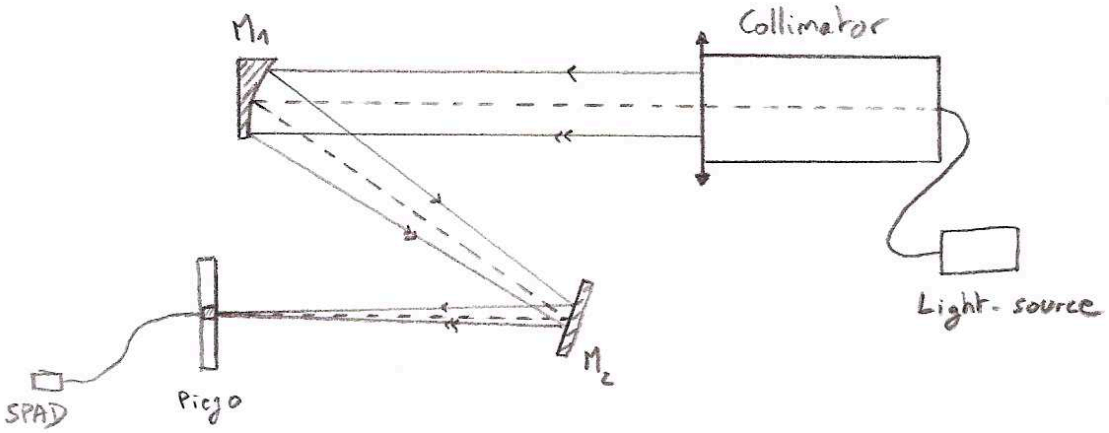
The instrument is perfectly aligned with the springs in place of the peelable washers. The objective of this step is to remove the springs and put the washers back, while minimizing the extent of the damage done to the alignment. To do so, I changed the optical setup, and I used a retro-injection setup (see Figure 7.5c). In this setup, the fiber is disconnected from the photodiode, and connected to a light source. Light propagates through the whole instrument, and exits the telescope as a collimated beam. Keeping everything firmly attached to the optical bench, it is possible to follow this procedure to add the peelable washers and adjust the tip/tilt again:

- Use the camera to take a picture of the light beam produced by the payload aligned with the springs.

- Dismount the secondary mirror, remove the springs, add the peelable washers, and mount the mirror back.

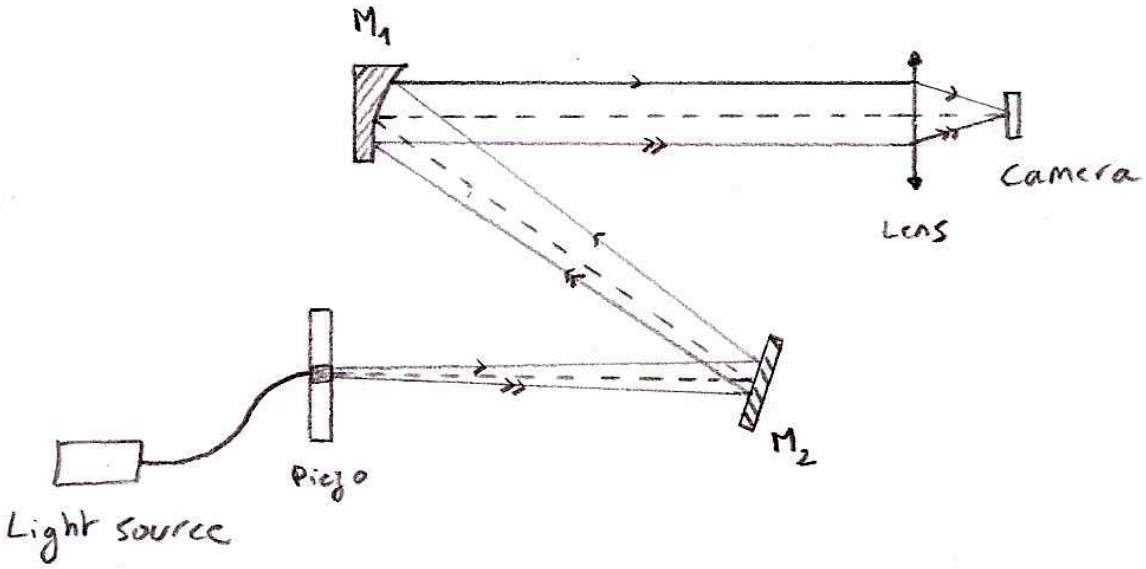


(a) Primary only

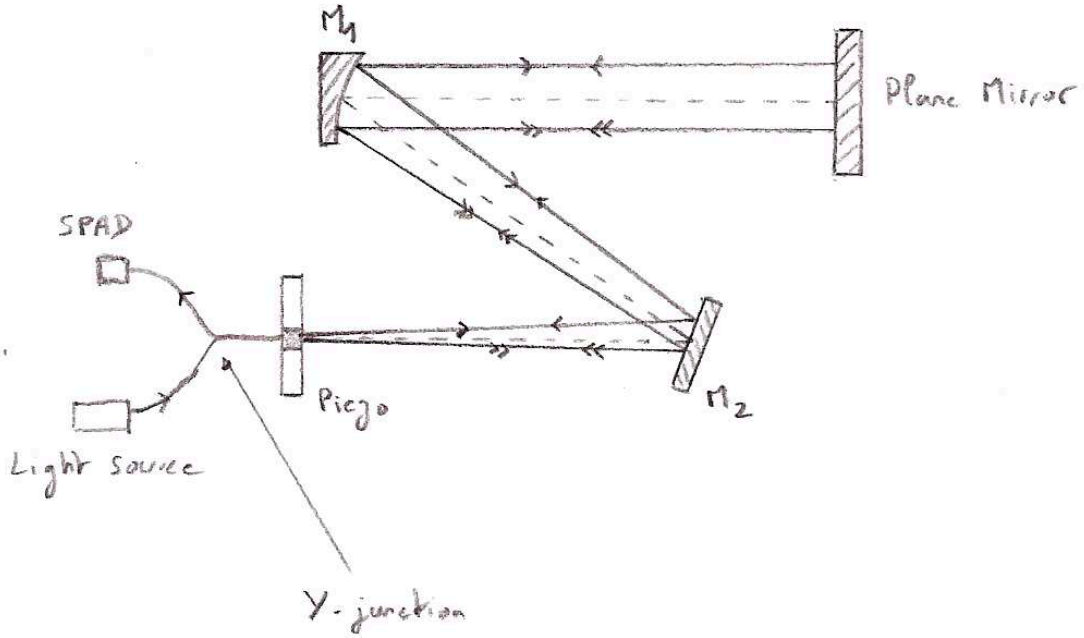


(b) Direct injection

Figure 7.5: The four different optical setups used to align the instrument of PicSat. Panel 7.5a: optical setup with only the primary mirror used to align the collimator on M1. Panel 7.5b: the direct injection setup, used to align the instrument in tip/tilt/focus with the springs, and in focus with the spacers.



(c) (continued) Retro-injection



(d) (continued) Self-collimation

Figure 7.4 (continued): Panel 7.5c: the retro-injection setup, used when replacing the springs with the spacers. Panel 7.5d: the self-collimation setup used to verify the final alignment.

- Take another image with the camera, and measure by how much the image of the fiber has moved in the focal plane of this camera.
- Dismount the secondary mirror again, peel one washer, mount the mirror back, take another image, and measure by how much the position of the image has changed. Repeat this step for each of the three washers.
- Using the data of the previous step, it is possible to calculate the displacement of the image induced by a change in the height of each washer. Knowing the initial position of the image when the instrument was correctly aligned in tip/tilt (from the image acquired at the first step), the amount of layers to be removed/added on each of the three washers can be calculated and removed/added.
- Small adjustments of the height of each washer may be necessary to ensure that the image of the fiber ends up exactly at its initial position.

After this sequence, the instrument is aligned in tip/tilt.

7.5.4 Focus

To align the instrument in focus, it is necessary to use the injection setup again. The collimated source is put back in place. Since the instrument is aligned in tip/tilt, the tip/tilt of the source can be adjusted to center the image in the piezo field. This ensures that the collimated source is aligned with the primary parabola. At this point, the only variable left is the focus, which can be changed by adjusting the height of all washers simultaneously by the same amount, to maximize the fiber injection. This step is slow and tedious, but I did not find any better option.

7.5.5 Fibered self-collimation

One of the main problem with the alignment sequence described above is that the final focus adjustment depends on the quality of the collimator. If the collimator is not perfectly tuned, the instrument may effectively end up slightly out of focus. A classic way of dealing with this problem is to use self-collimation, a process in which an optical instrument is aligned using the instrument itself as a collimator.

For verifying the alignment of the PicSat payload, I used a similar idea, but adapted to a "fibered version". Using a fibered Y junction, it is possible to simultaneously measure the light coming out of the instrument fiber, and retro-inject another beam in the telescope. By placing a mirror in front of the telescope, light can be retro-injected in the fiber and in the telescope (collimator), reflected on the mirror, focused again by the telescope, and measured at the other end of the Y junction. This setup is illustrated in Figure 7.5d.

In this setup, the optical aberrations measured at the output of the system are twice the aberrations of the instrument. This makes the alignment even more difficult, which is why it was only used as a last step, but it provides a way to unequivocally check the final alignment, and can potentially be used to fine-tune it if necessary.

7.5.6 Result of the alignment

With all the trials and errors on the engineering model, the changes in the optical setup, the need for several updates of my payload data visualization software, and the hassle of working in the clean room, it took me several weeks to come up with this alignment procedure. Aligning the flight-model itself took me about 3 weeks, full time. And the result of the alignment was

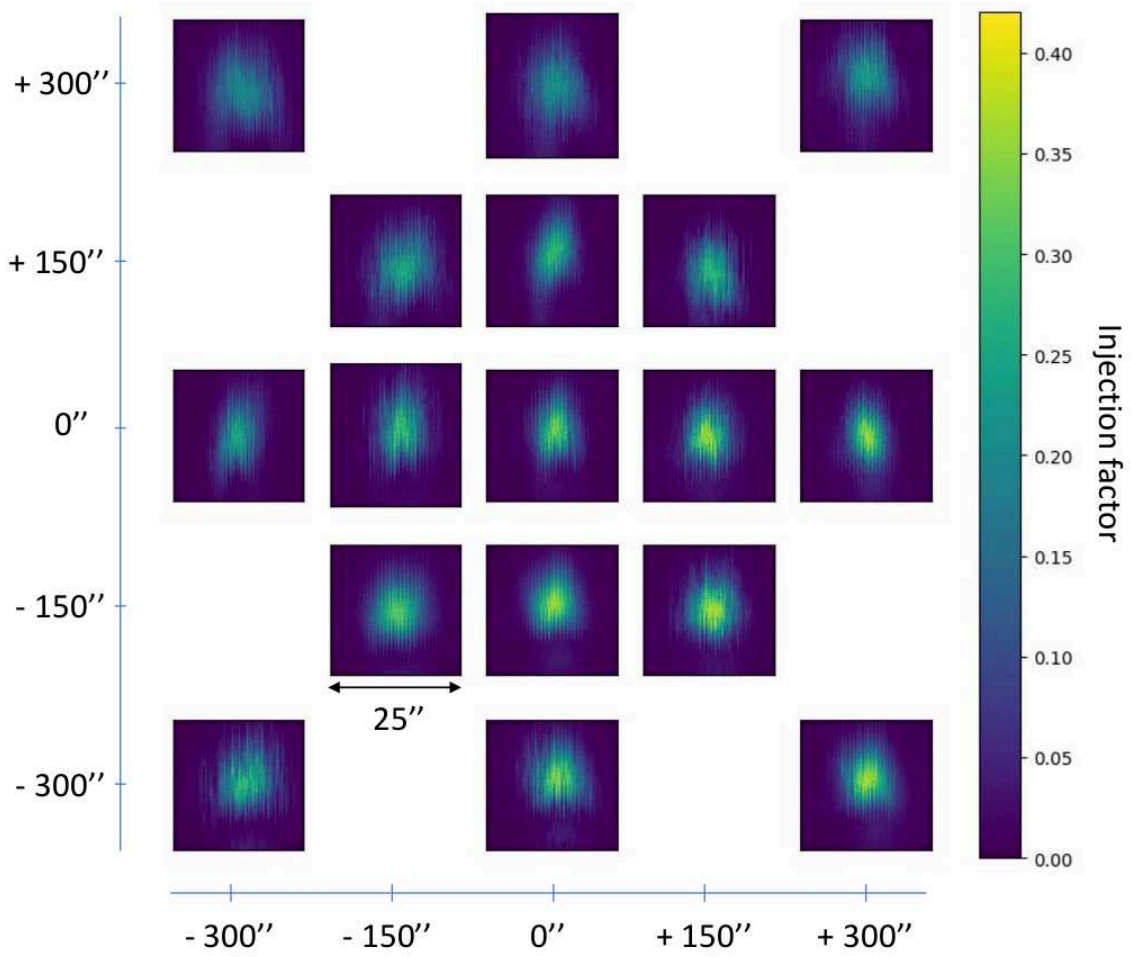


Figure 7.5: Overview of the shape of the equivalent Point-Spread Function of the instrument (which includes fiber-coupling) at different locations in the field-of-view. The upper-left part of the field-of-view show strong aberrations, which reduce the injection down to $\sim 10\%$.

not perfect. The final injection ratio obtained was $\sim 30\%$, far from the theoretical maximum which is close to $\sim 80\%$. Figure 7.5 gives an overview of the shape of the instrumental PSF at different positions in the focal plane, which clearly shows that the alignment is not perfect, with a PSF visibly better in the lower-right corner than in the upper-left.

But a moderate 30% injection was enough to guarantee the photometric performance of the instrument, and every attempt to improve the alignment came with a risk of ruining it[3] (since it was necessary to remove and mount back the secondary mirror, an operation not 100% repeatable). So it was decided to stop the alignment at this 30% injection level.

[3] In fact, the three weeks required to align the flight payload reported above do include a few cases of decent alignments ruined by an optimistic "if I just peel this washer one more time..."

Chapter 8

Ground segment and amateur collaboration

Content of this chapter

8.1 Ground segment

8.1.1 Packet protocol

The communication protocol used throughout the PicSat mission loosely follows the general CCSDS (Consultative Committee for Space Data Systems) format described in the Space Packet Protocol [CCSDS Secretariat, 2003]. The CCSDS protocol is complex, and we had to adapt it to the need of the PicSat mission. The protocol was first adapted as a way to send and receive data to the payload, and then became the general protocol used for the entire mission.

A communication packet can be broken into three parts (see Table 8.1): a primary header (same structure for all packets), a secondary header (one structure for telecommands, and one for telemetries), and a data section. The data section possibly starts with an auxiliary header (for telemetries only). These three parts taken together form the body of the packet, which due to limitations of the mission space-ground communication system, is limited to a length of 235 bytes maximum.

Field	CCSDS Header			Secondary Header TC/TM			Data		
Byte	0	...	5	6	...	23	24	...	N

Table 8.1: General structure of a TC packet.

CCSDS Header

The structure of the CCSDS Header is given in Table 8.2. The version number is always 0, the type is 0 for telemetries and 1 for telecommands. The secondary header flag is set to 1.

The PID (Process Identifier) field identifies the software task which is the origin or destination of the packet. Together with the level flag (0 for a packet related to Level-0 software, 1 for Level-1), and the payload flag (1 for a packet related to the payload, 0 otherwise), this PID is used to route packets in the satellite and within the software itself.

The category identifies the type of packet. This is an important number from the perspective of the ground segment software, as it indicates the structure of the data section, and thus how the packet should be unpacked, and in which table of the database the data should be stored. For ground-to-space packets, the category is always set to 0.

The sequence flag in the PicSat mission is not used, and is always set to `0b11`. The counter number is used to identify the packet, and is increasing for each packet emitted.

Finally, the data length is equal to the number of bytes contained in the body of the packet, CCSDS Header excluded (the number of bytes of the secondary header and data).

CCSDS Header															
Field	Version			Type	Sec. Hdr flag	PID			Lvl flag	Pld flag	Category			...	
Bits	0	1	2	3	4	5	...	8	9	10	11	...	15	...	

CCSDS Header									
Field	...	Seq. Flag		Counter			Data length		
Bits	...	16	17	18	...	31	32	...	47

Table 8.2: General structure of the CCSDS Header

Secondary Header TC

The general structure of the secondary header used for the telecommand packets is given in Table 8.3. The first three bits are used to request acknowledgment packets from the satellite, upon reception, format verification, and execution of the telecommand. A value of 1 is used when an acknowledgment is requested. A value of 0 is used when it is not requested. The telecommand identifier (tc id) is a number which unequivocally identifies the telecommand. The structure of the data section of the packet depends on this number (parameters of the telecommand). The emitter ID is actually ignored in the PicSat mission. The signature is a 16 byte cryptographic signature, generated with a key shared between the ground segment in Meudon and the satellite. It is used to authenticate the origin of the packet. Since the signature also depends on the content of the packet, it is also used as a way to verify the integrity of the telecommands received by the satellite.

Secondary Header TC												
Field	Ack reception	Ack format	Ack execution	Tc Id			Emitter id			Signature		
Bits	0	1	2	3	...	12	13	14	15	16	...	143

Table 8.3: General structure of the telecommand secondary header

Secondary Header TM

The structure of the secondary header used for the telemetry packets is given in Table 8.4. These fields are set by the on-board computer using an internal Real-Time Clock (RTC). As per CCSDS specifications, the time is given in full days elapsed since `January, 1, 2000`, and millisecond elapsed since the beginning of the current day ($0 \leq \text{ms} < 86\,400\,000$).

Secondary Header TM						
Field	Days since ref			Millisecond of day		
Bits	0	...	15	16	...	31

Table 8.4: General structure of the telemetry secondary header

Data section

The data section is an optional field. In case of telecommands, it contains the parameters of the command. In case of a telemetry packet, it contains the data, possibly preceded by an auxiliary header. The data section is specific to each packet category and to each telecommand id.

8.1.2 Nanoctrl (ground segment software)

A large part of the ground segment software for the PicSat mission was developed in-house at the Observatoire de Paris. Three different pieces of software were developed (see Figure 8.1 for a schematic overview). PicTalk[1] was developed by amateur radio F4GKR, and is used to decode the telemetries received. Nanoctrl[2] is the core of the ground segment software. It is used to send telecommands, to display simple information from the telemetries, to save all

[1] https://github.com/f4gkr/PicTalk
[2] https://github.com/ceyzeriat/nanosat-control

telecommands and telemetries in the database, etc. Finally, PicsatData is a Python package used to access and manipulate all information contained in the mission database. I was personally involved in maintaining and updating Nanoctrl after its initial development. I was also in charge of setting up and maintaining the mission database, and the associated PicsatData package.

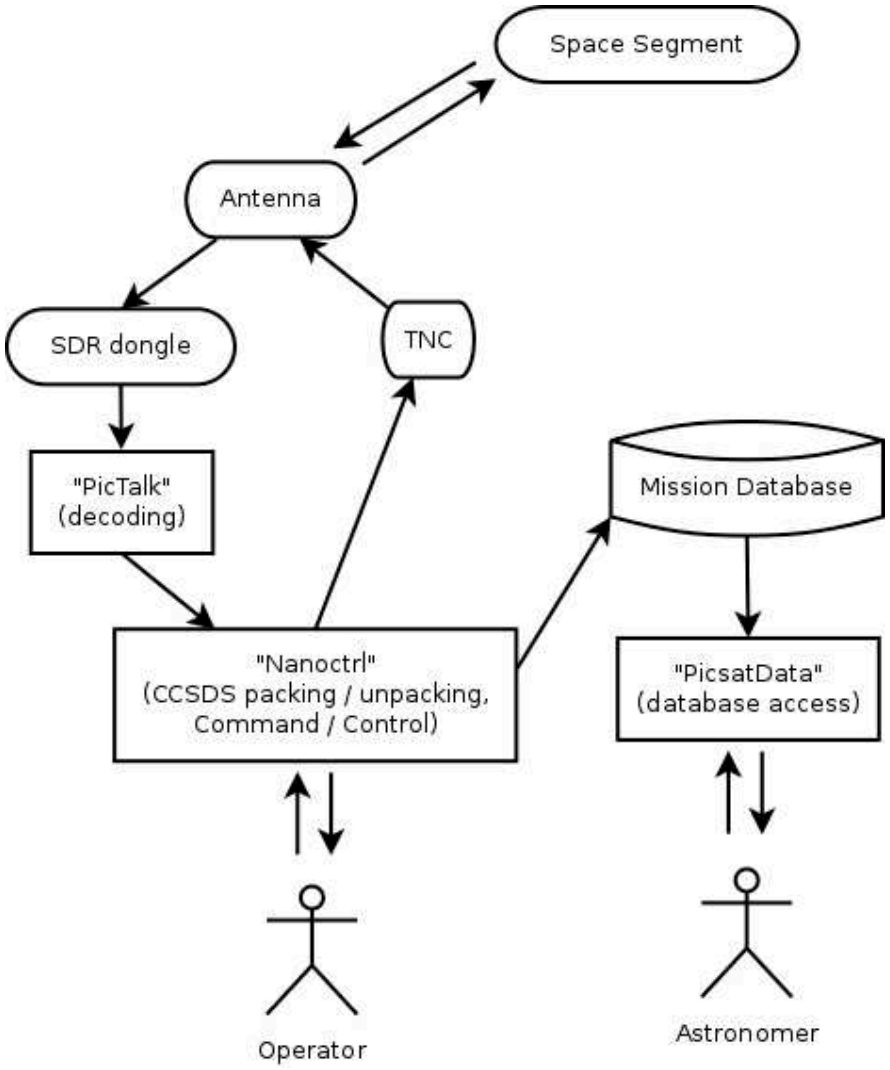


Figure 8.1: General architecture of the ground segment software.

8.1.3 Database architecture

The mission database of the PicSat mission reflects the architecture of the CCSDS packets.

Telemetries For the space-to-ground packets, a general `telemetries` table stores all the general information about the telemetry packets received. This information roughly corresponds to the primary and secondary CCSDS headers, plus the reception time, and a few quality flags. Then, for each packet category, a `tmcat` table is used to store all information contained in the auxiliary header, and a `data` table to store the actual data. The auxiliary

header typically corresponds to general information required to properly interpret the data. For example, in a high-frequency science data packet, each photometry point is given as a set of a count value measured by the diode, two numbers corresponding to the piezo XY position, and two numbers corresponding to the piezo XY command. To interpret this data, it is also necessary to have some information about the modulation pattern currently in use, the tracking mode, the integration time, etc. This information, which does not need to be repeated for each data point, is contained in the auxiliary header, and thus is stored in the `tmcat` table. This architecture is illustrated in Figure 8.2. Each line in the `tmcat` and `data` table contains a `telemetry_packet` key, linking it to the corresponding line of the `telemetry_packet` table.

Retrospectively, this architecture was not the most user-friendly, as it complicates the relationship between data points and their associated auxiliary header parameters. Given the amount of data received, repeating the auxiliary header data in each line of the `data` table would have been much simpler.

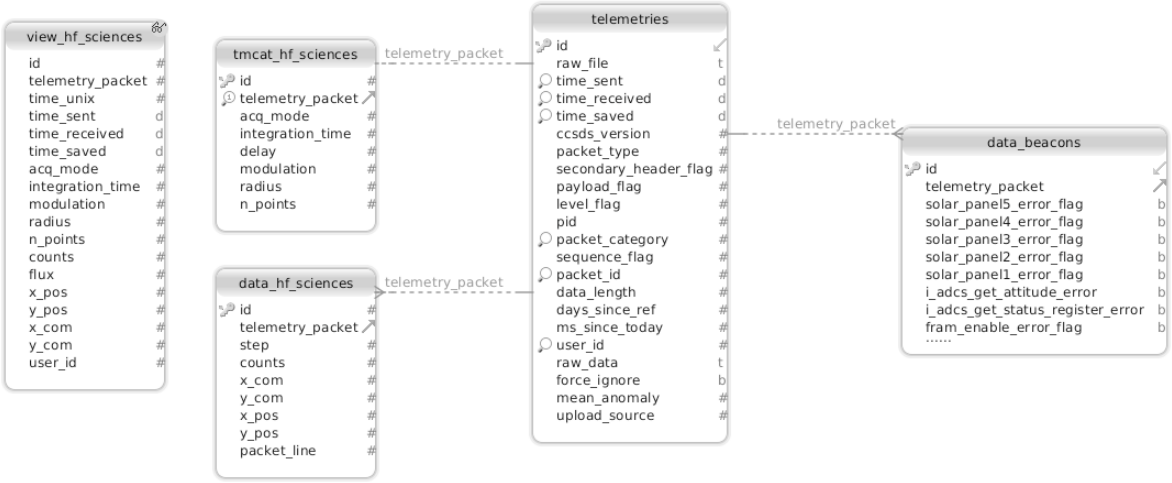


Figure 8.2: General architecture of the part of the mission database dedicated to storing the telemetries (see text).

Telecommands The database architecture for the telecommands is simpler than for the telemetries (see Figure 8.3). A general `telecommands` table is used to store all the general information about the telecommands, except its parameters and associated values, which are stored in a dedicated `telecommand_data` table. A small `command` table is used to convert the telecommand id numbers into human-readable names.

Users The last part of the database is dedicated to user management. A `users` table contains all the general information about the users registered through the PicSat mission website (including a hash of the password). The initial design of the mission database included an integer value `user_id` in the `telemetries` table, referencing the `users` table. The idea was to use this number to attribute each packet to its receiver. However, when the collaboration with the amateurs started, it rapidly became obvious that this was a bad design choice, as a single packet was often received by several people simultaneously. To properly attribute packets to each of its receivers, a dedicated `user_packets` table was added to the database.

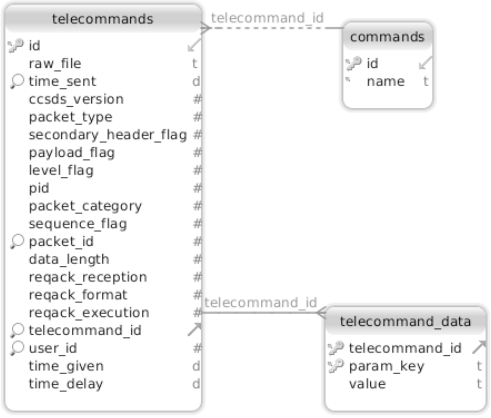


Figure 8.3: Same as Figure 8.2, for the telecommands.

Finally, a `requests` table was used to store requests of data download (see Section 8.3). This architecture is shown in Figure 8.4.

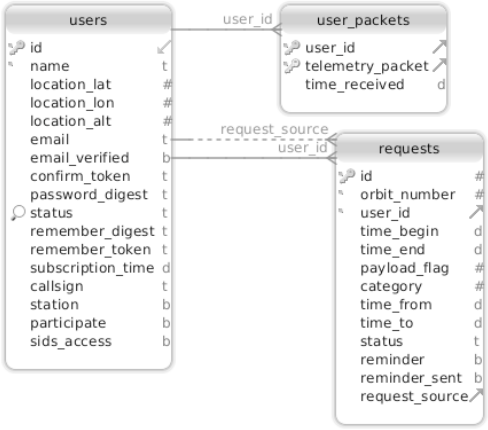


Figure 8.4: Same as Figure 8.2, for the user managements.

8.1.4 Ground station

Ground-to-space communication (telecommands) with PicSat was done at 145.910 MHz, and space-to-ground communication (telemetries) at 435.525 MHz. A dedicated ground station was built in Meudon to operate the satellite (see picture in Figure 8.5). It featured:

- An uplink Yagi antenna WX-220 Wimo (VHF)
- A downlink Yagi antenna WX-7036 Wimo (UHF), and a preamplifier MVV-432 VOX PREAMPLI
- A transceiver TS2000E TXHF/ 50/ 144/ 430MHZ, with a Terminal Node Controller (TNC) to convert the analog signal into a numerical signal
- A power supply for the TS2000 (SPS-250)

- An UHF/VHF power-meter CN103LN
- A set of rotors to control the orientation of the antennae (G5500 ROTOR SITE/ AZIMUT)
- A RS232 interface to the rotor (GS232B INTERFACE RS232C ROTOR YAESU)

Figure 8.5: The ground station in Meudon, by a cold winter day. The antenna on the right is the emitting antenna, operating at 145 MHz, and the antenna on the left is the receiving antenna, operating at 435 Mhz.

8.2 Collaboration with radioamateurs

8.2.1 Amateur radio service

A number of radio frequency bands in the world are reserved for non-commercial use by individuals interested in radio-communications, and wishing to study its different facets and/or to setup and use radio-installations. These frequencies are called *amateur frequencies*. Some frequencies are reserved for the *amateur radio service*, legally limited to operations below an altitude of 50 km. Other frequencies, part of the *amateur satellite service*, can be used for space communications. The International Telecommunication Union (ITU), an agency of the United Nations, coordinates the use of frequency bands worldwide. All amateur unions are collectively represented in the ITU by the International Amateur Radio Union (IARU).

To be allowed to emit on any amateur frequency, a license is required. In France, this license can be obtained from the *Agence Nationale des Fréquences* (ANFR), after passing both a technical and a law exam. Any new licensee is granted a *callsign*, which is used for

identification in all amateur radio-communications. I passed this exam during my PhD to be allowed to operate the satellite[3].

8.2.2 PicSat and the amateur radio community

Since the beginning of the mission, PicSat was heavily supported by the amateur radio community. The *Réseau des Émetteur Français* (REF) was involved in the development phases. They advised our team on the ground station setup, and on the process of obtaining a frequency.

PicSat was allowed to use amateur frequencies in exchange for embarking a radio transponder (repeater). Satellite time would be shared between the science instrument and the transponder. But we also decided to take advantage of the opportunity offered by the use of amateur frequencies to involve the amateur community in data-collection. Communication on social media (see our Twitter account IamPicSat[4]) and on our mission website, regular reports on the status of the mission and on the use of packets collected by amateurs, open format for the telemetries, and an honest effort to involve members of the team in the amateur community (see my talk at AMSAT-F[5], in french) were duly rewarded by an impressive participation of the amateur community worldwide (Figure 8.6 and 8.7).

8.2.3 Collecting packets

To retrieve telemetry packets from amateurs listening to the satellite, I developed two different services, which were used throughout the mission. The first one was a text-file based processing method, and the second one was a dedicated server.

8.2.3.1 Text-file packet processing

Amateurs listening at PicSat and retrieving packets could save them in a simple text file, in hexadecimal format. Different packets were separated by newlines, and the whole file could be sent to the web server based in Meudon via a dedicated tab in the mission website. Files were saved in folders named to identify their owner. Each night, at a given time, a CRON task would go through all files received during the day, and process packets one by one. The format of each packet was verified, and a few checks on the data contained were performed. All valid packets were then added to the mission database.

This method was quick and simple to setup at the beginning of the mission, but proved to be quite inconvenient for everyone. First, different people were using different formats for the packets (e.g. grouping hexadecimal symbols by two or four; separating them by whitespaces or not; capital letters or not; keeping the complete AX25 packet or removing the header; keeping the end sequence for the packet or not; etc.), and adapting the processing software to make sure that all formats could be accepted was tedious. Secondly, with such a system, it was impossible for amateurs to have an immediate feedback on how their station was performing. If, for any reason, the station was missing some bytes here and there, and/or was not decoding the packets correctly, they would only notice it by the next day, and a large part of the data could be lost. Similarly, it was not possible for the PicSat team to check the satellite beacons in real time when the satellite was passing over stations around the world.

[3] F4HZG
[4] https://twitter.com/IamPicSat
[5] https://www.youtube.com/watch?v=ODpF1plHfvo

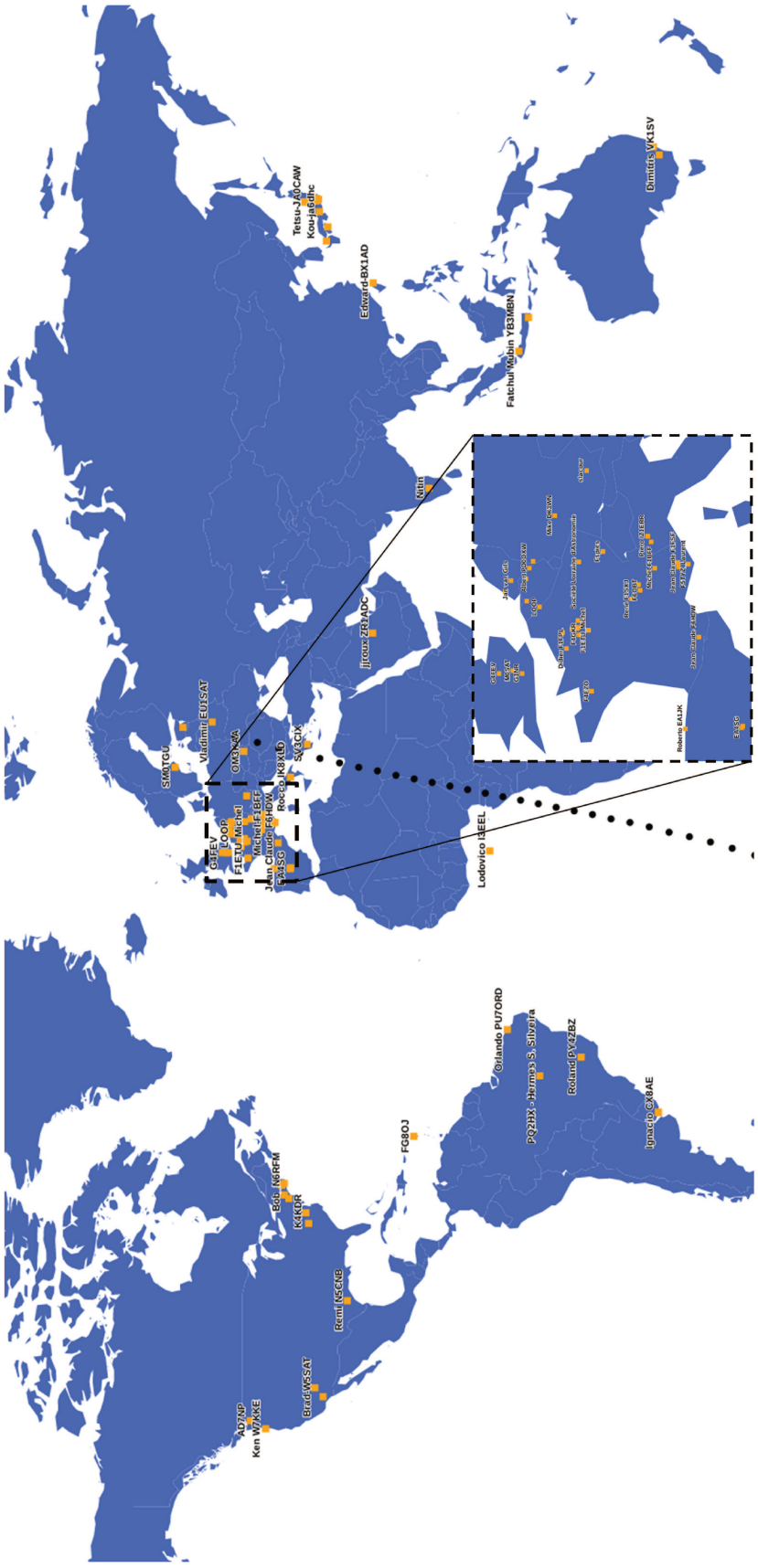


Figure 8.6: The worldwide amateur ground station network of the PicSat mission, comprising 79 stations spread over 5 continents.

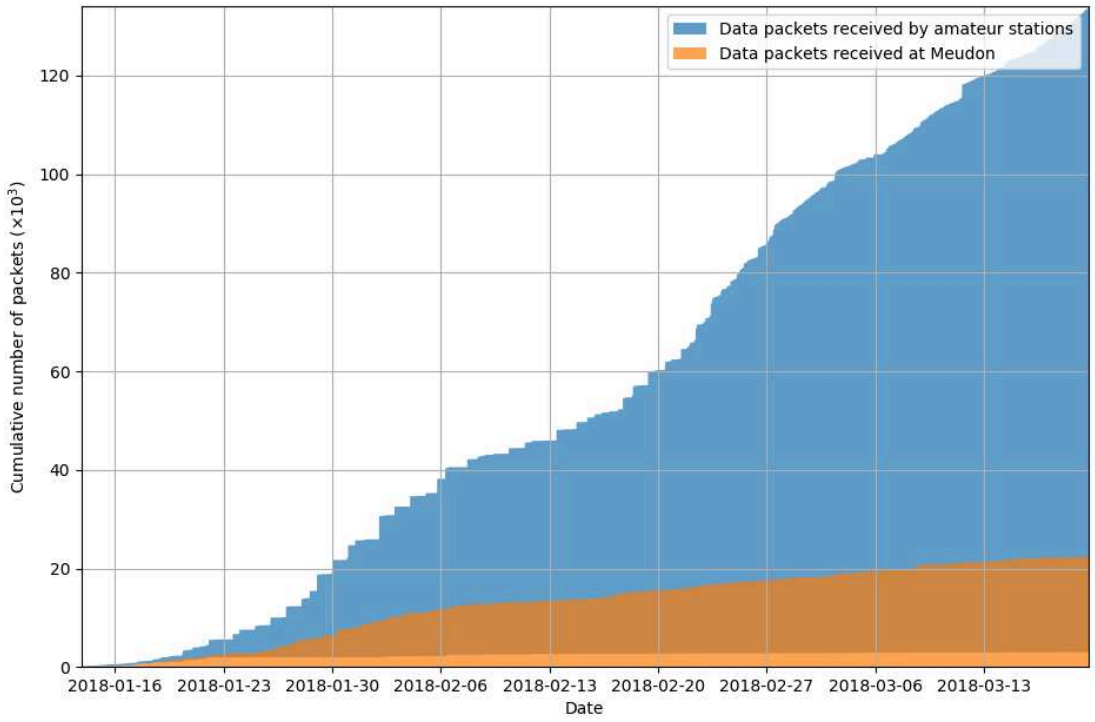


Figure 8.7: Cumulative number of packets received by our ground station station in Meudon and by amateurs, as a function of time. Packets received simultaneously by multiple amateurs are only counted once. The region were the two area are superimposed represents the packets received both in Meudon and by amateurs. The first inflection in the amateur curve (around 2018-01-23) corresponds to the website and SiDS server slowly getting up and running. The second inflection point (2018-02-20) corresponds to the beginning of dedicated "data download operations" over amateur stations, in which the satellite was requested to send a continuous stream of data to offload the memory cards. The decreasing rate at which packets are received in Meudon (orange curve) is due to increased telecommand activity, causing interference with the receiving antenna in our ground station setup.

Parameter	Optional	Type	Description	Example
noradID	No	int	Norad ID of the spacecraft	43131
source	No	str	Callsign or user name	F4HZG
timestamp	No	str	UTC timestamp (see ISO 8601)	2018-01-24T23:42:46Z
frame	No	str	The received data, in hexadecimal string (AX.25 packet with or without KISS 'C0 00 .. C0'. Whitespaces optional.	C0 00 A9 05 DE ...
locator	No	str	Type of the given receiver's location. Currently, only 'longLat' is supported.	longLat
longitude	No	str	Longitude of the receiver (WGS84)	8.95564E
latitude	No	str	Latitude of the receiver (WGS84)	49.73145N
tncPort	Yes	int	Optional as per SiDS standard specification, but not used in PicSat SiDS system	0
azimuth	Yes	float	azimuth angle (deg) of antenna (if avaiblable)	10.5
elevation	Yes	float	elevation angle (deg) of antenna (if avaiblable)	85.0
fDown	Yes	float	Frequency of the receiver's downlink channel during reception (with Doppler), in Hz	435525000

Table 8.5: Descriptions and formats of the parameters used by the PicSat SiDS server to collect telemetry packets from amateurs

8.2.3.2 SiDS server

To solve the issues inherent to the text-file based approach, and on the advice of amateurs themselves, I decided to set up an SiDS server for PicSat.

The SiDS (Simple Downlink Share) interface is an amateur standard used to forward data packets from any receiving station to a centralized mission control server. It was first introduced and used by the University of Würzburg, for the UWE-3 mission.

The concept behind the SiDS is to allow people to send data packets in hexadecimal format via HTTP POST requests, such as:

```
https://picsat.obspm.fr/sids/reportframe?noradID=100&source=F4HZG
&timestamp=2018-01-12T09:00:00Z&frame=FE DC BA 98 76 54 32 10
&locator=latlong&longitude=5E&latitude=45N
```

The list of required and optional parameters used in the SiDS standard (version 0.9, as used for PicSat) is given in Table 8.5. Upon reception of a frame, the server verifies its integrity, and add it to the mission database if correct. The user is then notified of the results of the integrity checks by an HTTP response. In case of success, the user receives an HTTP code 200 along with plain text message "OK". In case of failure of the request, the user receives an HTTP code 400 along with a human readable plain text message describing the error (example: "Bad AX.25 header: source should be XX and is YY").

The system seems simple to setup, but I faced a major issue while developing the PicSat SiDS: database collisions. When a user connects to the Python server, a new task is created, to handle his requests. Thus, when the satellite was passing over Europe, where several amateurs were listening to it and sending frames simultaneously to the server, many tasks were running in parallel, each trying to add data into the mission database. In this situation, write access collisions were frequent, and most requests were not properly handled by the system.

The problem, which had not been foreseen in the initial development of the server, was difficult to solve. The fix was to delegate all database access to a centralized task, and use a request queue. Each server task would limit itself to receiving the POST requests, verifying their format, sending answers to the users, and adding valid requests to a common queue. The centralized task would then take the requests in the queue one by one, and add them to

the mission database. A database locking mechanism was also implemented to avoid database write access collisions between the SiDS central task, and our own ground segment software. The overall system proved to be very robust.

8.3 Mission website

To facilitate interactions with the amateur community, and to share our scientific excitement with the general public, we decided to create a public mission website. One of the objectives of the website was also to display information from the database in graphs and other easy to read formats, in order to facilitate our day-to-day activities. Consequently, the website architecture was tightly connected to the mission database, and since I was maintaining the database, I was also in charge of creating and maintaining the website. This was not an easy task, which first required me to learn Ruby and Ruby On Rails.

The website contained a total of 11 different sections, each with its own objectives and target audience. As of 2019, the website is still accessible at:

`https://picsat.obspm.fr/`

Home This section contained the landing page of the website. The main feature of the landing page was a set of two eye-catching pie-graphs showing the types of telemetry packets received during the last 24 hours, and the contributors (i.e. amateurs who had received packets during these last 24 h). The page also displayed the general ranking of amateurs (by number of packets received since the beginning of the mission). The page contained a small text explaining what PicSat was, and a few links to the most recent news published under the "news" section of the website.

News The news section was linked to a non-public WordPress site. The goal of this section was to allow anyone in the team to easily publish small news about the project, without requiring an update of the main website code. The WordPress articles were automatically retrieved by the mission website, and displayed in this section. In practice, this section was mostly left unused, as the IamPicSat Twitter account was our main channel for publishing news.

Media resources This section was targeted to people interested in the project, and wishing to download and use images, or read articles or science papers written about it.

Science mission A section targeted toward the general public, and explaining what the science objectives of the mission were.

Satellite This section was subdivided in 3 parts: one part contained a publicly available general explanation about the satellite, its launch and its orbit. A hidden subsection was only accessible to registered user, and was made for contributors and team members. This section contained a dashboard used to visualize information about the satellite (monitoring data versus time in different graphs, LED like buttons displaying warnings or errors sent by the satellite, graphs showing the load factors of the memory buffers, etc). The PicSat team used this section on a regular basis to check the status of the satellite. All data were directly loaded from the mission database by the website.

Payload Similar to the "Satellite" section, but for the science instrument.

Operations This complex section of the website was used to facilitate operations of the satellite, and took a different aspect depending on the status of the user.

For a non-registered user, this section only contained a map showing the current position of the satellite over the Earth, and a web form to calculate the next visibility windows over a given position.

For a registered and contributing amateur, this section also contained the "Operation Map", a full screen map of the Earth displaying the ground station network. The user could click on any station and have a short description of the upcoming operations scheduled during the next passes over this station (e.g. data download operations, beacon reception, science acquisition, etc.). A set of click-based menus and web forms also allowed all PicSat team members to use this map to prepare data download operations on amateur stations. The website would automatically generate "data download requests", which would be loaded by the ground segment software during the next pass of the satellite over Meudon, and appropriate instructions would be uploaded to the satellite.

At the time of end-of-mission, a similar tool was being tested by a subset of amateurs, to allow them to directly request data download over their stations. A few graphs were used to inform amateurs of which types of data were missing in the database, which one could be requested during the next passes of the satellite over their stations, and which one were already scheduled for download over other stations. They could then request the data they wanted to receive, and on which pass. The website would automatically generate the appropriate commands, and suggest their upload to the satellite to the satellite operator at the next pass over Meudon. Development of this system was driven by two observations:

- A significant fraction of all data download operations over amateur stations were unsuccessful because amateurs were not constantly listening to the satellite.
- Looking at the database to select the right set of data to download on each station, without knowing exactly the characteristics of all stations was difficult and time-consuming for the small PicSat mission team.

A truly collaborative management of these operations with all amateurs involved would have been much more efficient.

Mission data This section contained two tables, showing the telecommands sent to the satellite, and the telemetry received. Any registered user could access a complete description of the data in any given packet from this table. A tab in this section was also dedicated to the β Pic light-curve. Any registered user could explore the light-curve displayed in a zoomable graph, and tools were in development to allow people to mark, comment, and ask questions about particular features they were seeing. Sadly, this tab was never used...

Communications The complete data protocol and packet formats were explained in this section, which allow amateur listening to the satellite to decode the data they were receiving.

Contribute This section contained all information about how to contribute to the PicSat mission by listening and uploading packets. Several people also requested information about how to set up a small ground station to listen to PicSat, and a dedicated page was in preparation at the time of end-of-mission.

User The user tab could be used to register to the website, log in, log out, upload packets, consult and download all data previously uploaded, etc.

Chapter 9

Operations, mission results and conclusion

Content of this chapter

9.1 Launch

The final integration of the flight model of PicSat was carried out in one of the clean room of the Observatoire de Paris, and the satellite was delivered to Innovative Space Logistics (ISL), in the Netherlands, on December, 11, 2017. The satellite was then sent to India, and integrated on the Polar Satellite Launch Vehicle PSLV-C40 rocket. The primary payload of the rocket was an Indian Earth observation satellite, Cartosat-2F.

The rocket lifted-off on January, 12, 2018, at 03:59 UTC (Figure 9.1), from the Satish Dawan Space Center, on the Indian East coast. The launch was successful, and PicSat was released on its 505 km Sun-synchronous orbit about 20 minutes after lift-off.

The first pass over Europe occurred at $\sim$ 07:45 UTC, and the satellite beacon[1] was detected from our ground station, in Meudon. Several amateurs also confirmed detection of the satellite beacon at 435.525 MHz.

Figure 9.1: Hoping for the best... (Credit: ISRO)

9.2 Operations

9.2.1 Initial operations

The first telecommands were sent to PicSat on the launch day, during the evening passes. Acknowledgments and replies were received as expected, confirming the proper functioning of the satellite. However, it rapidly became apparent that the tools and procedures developed to

[1] The on-board computer of PicSat was programmed to automatically emit a packet every 10 s, which contained several house-keeping data. The beacon was used to detect the beginning and the end of passes, and to check the general status of the satellite.

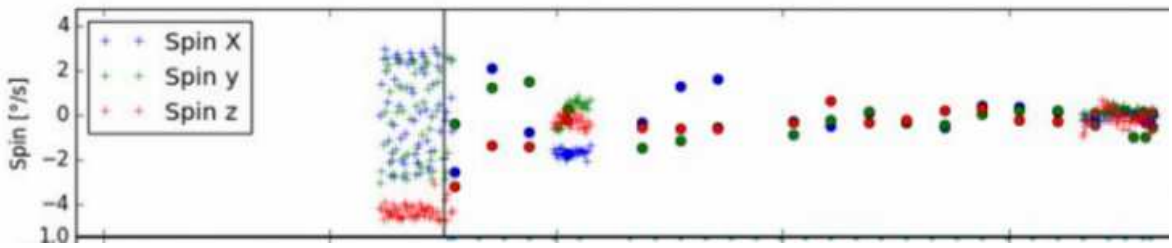


Figure 9.2: Evolution of the rotation speed of the satellite on its 3 axis when switching the ADCS to its "detumbling" mode. The rotation speed is initially of a few degrees per second, and rapidly drops when the detumbling is activated (black vertical line).

operate the satellite were not very well adapted to the real conditions. All our tools had been developed using the real ground segment and the engineering model of the satellite, located in a room at the Observatoire de Paris. The communication between the engineering model and the ground station was easy, and only a few packets were lost from time to time. When operating the satellite, the situation was completely different, and entire passes without a single data packet properly decoded was not uncommon (due to software mishaps, poor visibility of the satellite, hardware issues, errors on the orbital elements, etc.). Many of our scripts intended to simplify operations of the satellite by sending whole sequences of commands had been designed under the assumption that we could use the different acknowledgments from the satellite as indicators of the correct execution of commands. With a poor radio-communication link, all these scripts were completely ineffective.

The lack of automation in our ground segment software also made all our activities prone to human errors. Typing the command manually in the terminal under the pressure of running out of time often resulted in typos, bad argument formats, etc, further reducing the number of operations we could perform in a single pass.

But after a first week or two learning the hard way how to operate a satellite, we slowly got up to speed, and proper activities could start.

9.2.2 ADCS failure

One of the first critical operation to perform was the detumbling of the satellite. The initial rotational speed of PicSat was of the order of a few degrees per second, which was too large for the star tracker to get a fix. Thus, the Attitude Determination and Control System (ADCS) was used to reduce this rotational speed. The command to activate the detumbling sequence was sent 2 weeks after the launch, and resulted in a rotational speed of ~ 0.1 deg/s (see Figure 9.2). However, despite the fact that this rotational speed was below the maximum tolerated slew rate for the ST200, the system did not return a quaternion when switched to "target pointing".

During about 8 weeks, we worked with Hyperion Technologies to try to solve this issue. We confirmed that the ST200 was alive, as were all the independent subsystems of the iADCS100. But due to what appeared to be high-level software issues, we never obtained a reliable and stable quaternion. Despite 8 weeks of in-space debugging of the system (also using our engineering model on the ground to test possible fixes), PicSat never pointed to its science target.

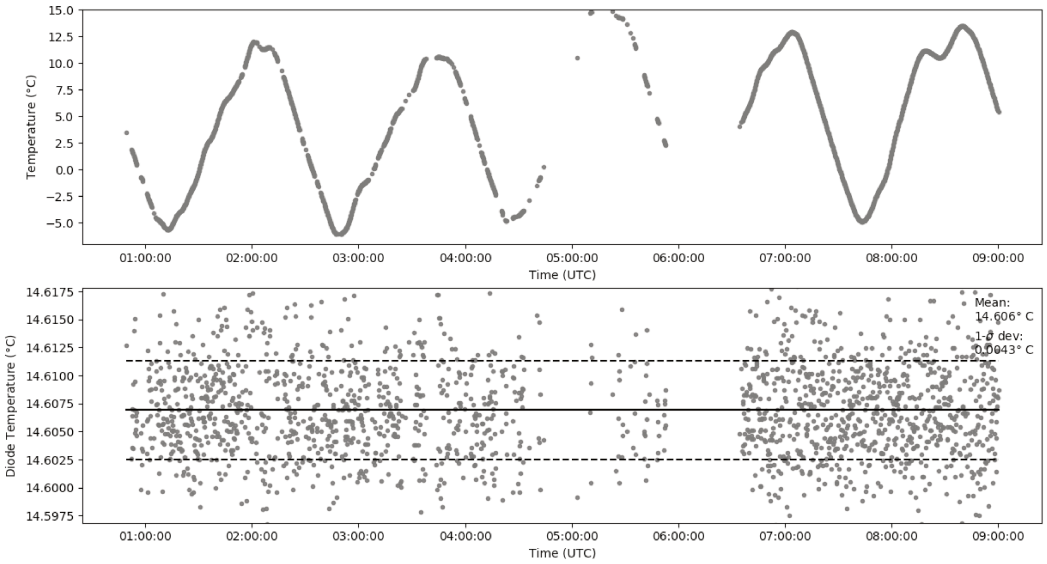


Figure 9.3: Evolution of the temperature of the payload electronic board (top panel), and of the temperature of the detector (bottom panel), on five consecutive orbital cycles.

9.2.3 Payload operations

Most of the satellite passes were devoted to trying to fix the ADCS problems. And since it was not possible to point to the target star, the possible payload activities remained limited. The payload was turned on several times, and the proper functioning of the electronics was confirmed. The TEC controller was also used to verify the temperature regulation of the detector, and we confirmed that the target stability of 0.01 °C was reached. During a test performed over 8 hr (several orbits), the temperature of the detector stayed stable at 14.606 ± 0.004°C (see Figure 9.3).

No major malfunction of the payload flight-software was identified over the duration of the mission. We also performed an upload of a new version of the flight-software, to fix minor issues related to the Real-Time Clock. This demonstrated that the bootloader was perfectly operational.

However, we decided to focus on fixing the ADCS to have a working target pointing mode before gathering photometric data. This turned out to be a very bad decision, as we did not obtain photometric data before the sudden platform failure. Consequently, we do not know if the avalanche detector was operational or not.

9.2.4 Platform failure

While we were still working on trying to get the ADCS to point the satellite towards β Pictoris, on March, 20, 2018 (after about 10 weeks of operations), our team was contacted by W2RTV[2], an amateur radio living on the East coast of the USA. W2RTV reported not hearing the satellite over his station, for the first time since the beginning of the mission.

In the evening pass over our own ground station, in Meudon, we confirmed that the satellite was indeed silent. We tried to trigger a reboot of the satellite by sending the appropriate telecommands, but this did not solve the problem. We were blinded, not even knowing if the

[2]Pronounced Whisky Two Romeo Tango Victor

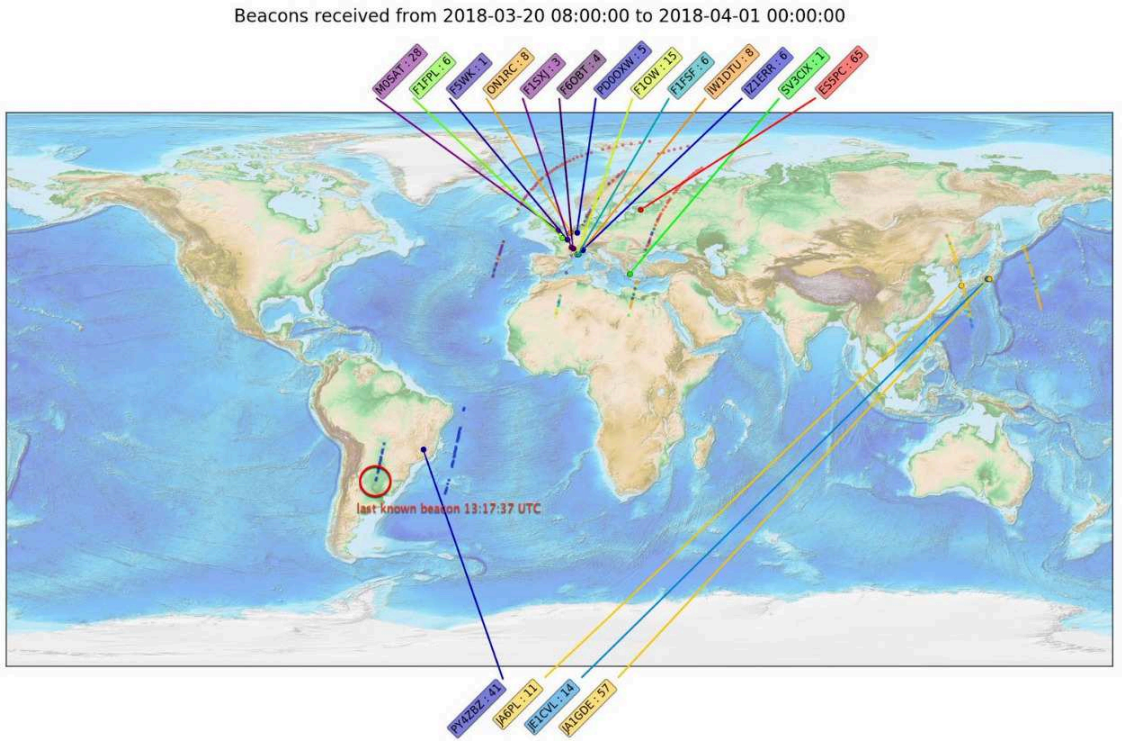


Figure 9.4: Map of the last beacons received by all amateurs listening to PicSat during the last few orbits before end-of-mission

satellite was receiving our telecommands or not, and we decided to stop all activity, in the hope that a reboot triggered by the EPS watchdog[3] would solve the problem.

It did not.

The satellite was emitting its radio beacon every 10 s on amateur frequencies, and we had a network of 80 radioamateurs following it every day. Thus, our ground coverage was extensive, and we managed to pinpoint the time and location at which PicSat stopped emitting. In Fig. 9.4, I show a map of the last few orbits of the satellite, with all the beacons received around the world. The last beacon was received above South America, at 13:17 UTC, by PY4ZBZ[4]. When we contacted him, PY4ZBZ reported a short (13 deg elevation) but nominal pass, with a clear signal. After that, the satellite was supposed to fly over the South pole, and then over south-east Asia. We contacted YB3MBN[5] and BX1AD[6], two radioamateurs respectively located in Indonesia and Taiwan, who reported not having heard the satellite when in visibility of their stations at 13:50 UTC and 13:58 UTC. We concluded that PicSat stopped working sometime between 13:20 UTC and 13:50 UTC on March 20, while flying over the South pole, between South America and Indonesia.

We performed an in-depth analysis of the possible causes of failure, but it was not possible to identify with certainty the origin of this early end-of-mission. Possible explanations are discussed below.

[3]The Electrical Power System on-board PicSat had a 3 day watchdog timer, that needed to be reset by sending a dedicated telecommand from the ground. If this command was not received in three days, the watchdog would trigger a complete electric reset (power cycle) of the entire satellite.

[4]Papa Yankee Four Zulu Bravo Zulu

[5]Yankee Bravo Three Mike Bravo November

[6]Bravo X-Ray One Alpha Delta

Figure 9.5: The 25 m antenna of the station PI9CAM, in Dwingeloo, Netherlands.

Power amplifier One of our first hypotheses to explain the loss of all communications was a malfunction of the satellite's transmission power amplifier. This power amplifier had been a source of concern during some tests on the ground, and although we solved the problem, it became an automatic suspect after the loss of contact. We reached out to the amateur radio community and asked for help to listen to the satellite with a high-gain antenna. An attempt was made with the 25 m dish of the PI9CAM station, in Dwingeloo, Netherlands (Figure 9.5). No signal was heard. We also contacted the NASA's Goddard Space Flight Center, and obtained a quick response. Two attempts to listen to PicSat were done with a 20 m class antenna, without success. The fact that no beacon could be detected with 20 to 25 m dishes ruled out a power-amplifier problem.

Aging Another possible cause of failure is aging of one of the numerous critical electronics (main computer, power system, communication system). We went through all the house-keeping data retrieved since the beginning of the mission, and found absolutely no warning sign of a premature failure. All data were nominal, even in the last beacon received. Although this is not enough to completely rule out this possibility, it makes it unlikely.

Collision We also considered the possibility of a collision with a space debris. We retrieved the history of the Two-Line-Elements (TLE) from the Space Situational Awarness service at JSpOC, to study the evolution of the orbital parameters of PicSat. In this data, we noted a strange and temporary change in inclination on March 20, as well as a peak in the mean motion derivative (see Figure 9.6). We inquired about these effects to the Conjunction Analysis and Evaluation Service, Alert and Recommendations (CAESAR) of CNES (French space agency), and we got a confirmation that this was not the sign of a collision. In fact, the change in inclination appeared to be a glitch, and no such change was found by the CAESAR team in their own data. The peak in mean motion derivative was due to a peak in geomagnetic

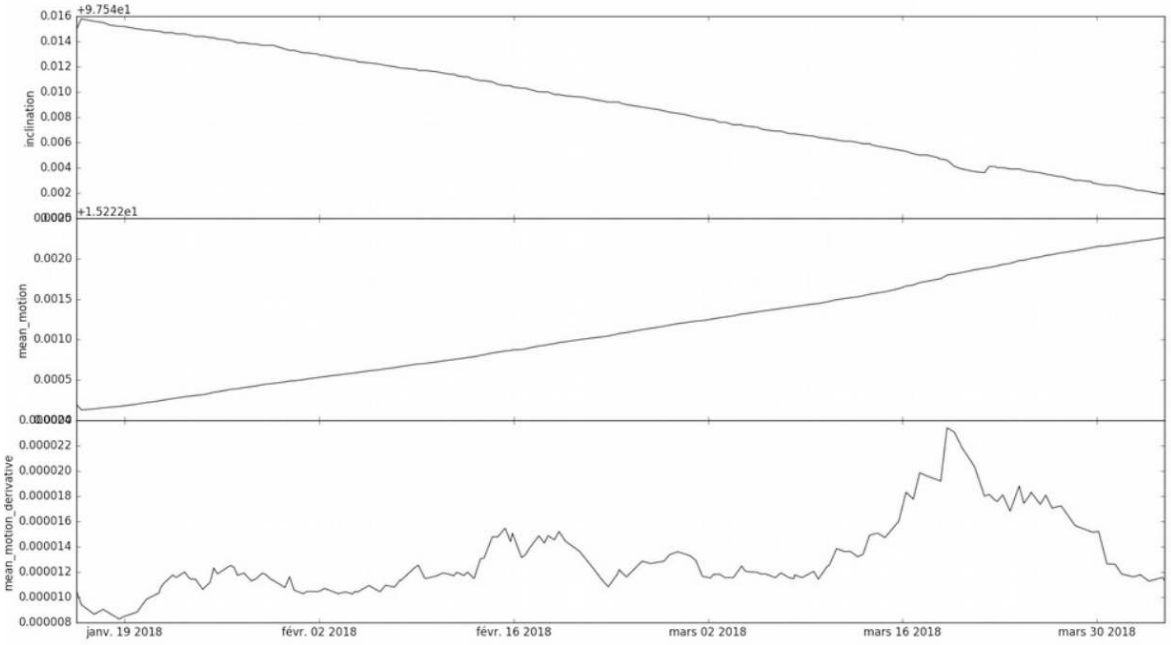


Figure 9.6: Evolution of some of the orbital parameters of PicSat during the mission. Upper panel: inclination (degrees). Middle panel: mean motion. Lower panel: mean motion derivative. Note the temporary change in inclination on March 20, which seems to coincide with a peak in the mean motion derivative.

activity, which had occurred on March, 19, 2018.

Radiation damage Geomagnetic activity was higher than normal for a few days before the end of mission, with a peak on March 19. The geomagnetic activity can raise the density of high-energy electrons in the outer Van Allen radiation belt. As can be seen in Fig. 9.7, the particle density increased from March 17 to March, 22, when it peaked, before going down to its normal level. The outer Van Allen belt is located at high altitude (between 4 to 6 Earth radii), much higher than Low-Earth Orbit (< 1000 km). However, the belt goes to lower altitudes at the poles (see Fig. 9.8). Thus, the satellite was subject to an increased flux of charged particles, which are known to be dangerous for electronics. This is a possible explanation for the early end of mission, coherent with the fact that the satellite stopped working over the South pole.

Software bug The last possible explanation for the failure of the satellite involves the software we developed at the Observatoire de Paris. During the last few passes of the satellite over our station in Meudon, we were performing a software update. As explained in Chapter 2, the on-board software of PicSat had a dual-layer architecture. A Level-0 software was in charge of all the critical tasks (power, communication, detumbling), whereas a second layer (Level-1) was in charge of mission specific non-critical tasks. The Level-0 software could not be changed on-orbit, but could be used to update the Level-1.

During the last pass of the satellite over Meudon, after having spent several days sending pieces of a new Level-1 software to the satellite, we requested the Level-0 to rewrite the Level-1. However, the new Level-1 software loaded on-board was found to be corrupted, and the command was rejected by the satellite, meaning that something went wrong during the

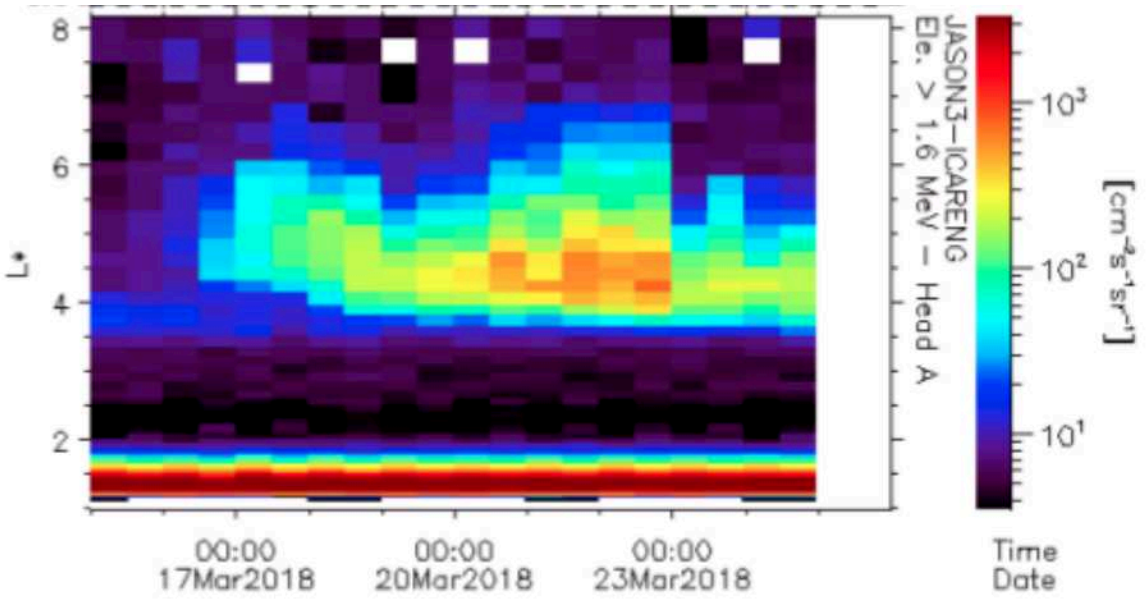


Figure 9.7: Evolution of the density of high-energy electrons in the outer Van Allen belt at the time of PicSat end-of-mission (data from JASON3 ICARE-NG)

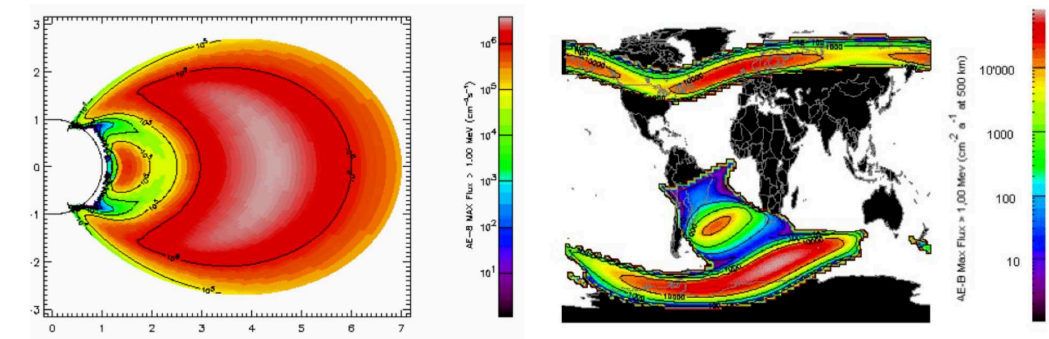


Figure 9.8: Map of the integral electron flux > 1 MeV in the outer belt (left panel), and at 500 km altitude above the Earth (right panel). This does not correspond to the situation at the time of end-of-mission, but only illustrates how the outer Van Allen belt connects to the poles. Figures from https://www.spenvis.oma.be/help/background/traprad/traprad.html, data from Vette [1991]

upload of the new software.

An extensive investigation of a possible Level-0 software bug led us to the discovery of a vulnerability in the code, which can indeed lead to a system bootloop (upon reboot, the Level-0 would try to execute a sequence of operations which would lead to a memory corruption, triggering a reboot, and thus the execution of the same sequence over and over again). In this case, the computer would reboot before even switching on the communication system, which would result in a complete loss of communications.

Despite all our greatest efforts, we never managed to trigger this bootloop on the engineering model when reproducing the sequence of commands used for the software update (which had already been tested on the engineering model before trying to update the flight model). The only way to trigger the bootloop was to send a sequence of command specifically engineered to exploit the L0 vulnerability.

9.3 Astrophysical results

This part intentionally left blank

9.4 Perspectives

Had the satellite platform and the attitude control system worked, it is still doubtful that PicSat would have achieved its 100 ppm accuracy in terms of photometric stability. The poor thermal design of the telescope, and the difficulty to align the telescope would have most certainly lead to a low injection. The photometric accuracy of PicSat would probably have been in the range of 500 ppm to 1000 pm. This would have been enough to cover the main science objective (detecting a transit of β Pic b, if it occurred), but not the secondary objective (exocomets).

The lab work and the simulations performed during the project all indicate that it is possible to achieve precision pointing and stable fiber coupling on a nanosatellite platform. The control algorithm developed for the PicSat mission, and presented in Chapter 4 is able to keep the fiber well centered on the star image, and it can guarantee an injection ratio $\eta \simeq 50 \pm 10\%$ (1-σ deviation at 1 kHz). This opens up numerous possibilities for exciting future missions.

Because they retain the coherency of the light propagating into them, single-mode fibers are of prime importance for interferometry. Having an algorithm able to control the fiber injection makes possible to embark a small interferometer on a CubeSat platform. This idea is at the core of SpaceFirst, a CubeSat-based nulling interferometer currently being studied at the Observatoire de Paris (see Lacour et al. [2014], and illustration in Figure 9.9a).

Another concept made possible by the work carried out during the development of PicSat is SAGE, a geostationary CubeSat constellation acting as a gravitational wave detector [Lacour et al., 2018]. This concept is illustrated in Figure 9.9b. Three satellites are placed in geostationary orbit, and form a roughly equilateral triangle. The satellites are connected by a bidirectional laser link, and time-delayed interferometry is used to monitor any change in the optical path created by a gravitational wave.

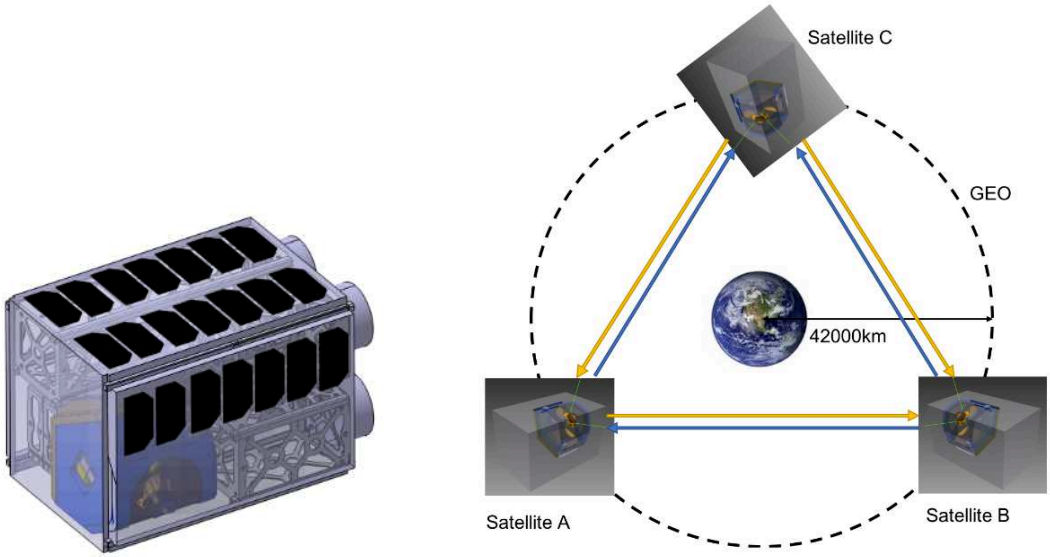
As a last idea, it can be noted that PicSat was designed with a small telescope (3.7 cm effective diameter) because its science target was the bright β Pictoris. Replacing this small optical telescope by a bigger one would open the possibility of doing exoplanet transit observations on fainter targets. A $\sim$ 9 cm aperture telescope, which could still fit in a 3U platform by using a compact modified Gregorian optical design (see Figure 9.9c), would make possible to look for transits as small as 10^{-4} around stars of magnitude < 9, or 10^{-3} around stars of magnitude < 12. Such a satellite could potentially support bigger missions (JWST, ARIEL), for example by refining the transit ephemeris to reduce the risk of missing them when observing with bigger instruments. In this context, the use of a single-mode fiber can have one major advantage over more classical approaches to photometry: it is very resilient to scattered light. During the entire development cycle of PicSat, no particular care was given to protecting the instrument from scattered light when testing it. The payload could easily operate in the lab, in ambient light, relying only on the excellent filtering capabilities of the fiber. This can alleviate the need for a bulky baffle, which can be difficult to fit on a small Cubesat type mission.

9.5 A final word on CubeSats and astronomy

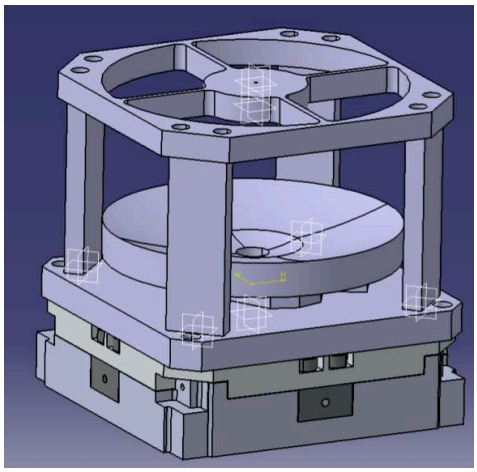
Astrophysically speaking, the PicSat mission was a failure. It did not deliver the data it was supposed to deliver on β Pictoris. A major malfunction of the ADCS (a commercial system) made impossible to point the target star during the 10 weeks of operations. The cause of the early end-of-life of the satellite remains unclear.

At a time when more and more CubeSat-based astronomy missions are being suggested,

(a) SpaceFirst: a 6U nulling interferometer



(b) SAGE: a constellation of three 12U CubeSats for gravitational wave detection



(c) A 9 cm modified Gregorian telescope for PicSatv2

Figure 9.9: Three concepts of future missions which would use the technology developed during the PicSat project.

this raises the question of the viability of this type of platforms, and more importantly, of the COTS components approach. Other projects have demonstrated that astronomical CubeSats are possible. For example, the BRITE constellation [Weiss et al., 2014], or the ASTERIA mission[7] were successful. But both of these teams either developed a large part of the hardware themselves, or used flight-proven components (the ASTERIA mission used the ADCS from Blue Canyon Technologies, already flight-proven). In the case of PicSat, we deliberately decided to turn to commercial hardware for most of the satellite subsystems, some of which (i.e. the ADCS) was not flight-proven.

Whether or not this was a good decision is hard to tell. This made possible to design, build, launch, and operate a mission in an extremely short time frame (3 yr), and with limited resources (16 FTEs, $1.4 M total). But the COTS components available today for CubeSats may not have quite the level of reliability suited for a space mission. And due to the fact that most of these systems are mission critical (power, communication, main computer), it is almost impossible to know for sure which one was faulty (if any; again, this could be a software failure). This makes difficult to decide how to improve the next mission based on this return of experience. The ADCS was clearly defective. But for the next mission, should we also change the On-Board Computer? The communication board? What about adding redundancies? To which sub-systems?

Looking back at this project, we believe that more testing may be the key to mission success. As long as the reliability of commercial systems remains unclear, they should probably be tested very thoroughly. Along all the development stages of the project, we did perform several environmental tests, but only in thermal vacuum, vibrations, and shocks. We did not perform any radiation test on the electronics, mainly because these are more costly, and much more difficult to set up. We also spent a large amount of time trying to develop a dedicated test bench for the ADCS, but this turned out to be too time-consuming, and we had to give up.

As astronomers and/or instrumentalists, we wish to focus only on the development of new and ambitious payloads, able to deliver high-quality science data. But to be able do so, we need to have reliable COTS hardware for the platform, that we can safely rely on. This may not be the case, yet. Then, early testing, even when the manufacturers claim certain levels of performance/reliability for their products, is probably the only way to know for sure in which systems we can be confident, and which systems should be avoided. This may also be the only way to push manufacturers to increase the reliability of their systems.

[7] https://www.jpl.nasa.gov/cubesat/missions/asteria.php

Part III

A GRAVITY OBSERVATION OF THE RE-EMERGENCE

Chapter 10

The role of interferometry in exoplanet science

Content of this chapter

10.1 An example: the diameter of HD 189733 b

Vulpecula (*le petit Renard*, for the French reader), is not a very famous constellation of the Northern sky. If it was not for the Dumbell Nebula, the constellation would probably be completely unknown. It is small, faint, and largely overshadowed by its much more impressive neighbors: Cygnus and Pegasus. Within this unremarkable constellation lies HD 189733 (or V452 Vulpeculae), an unimpressive binary system, invisible to the naked eye. Around HD 189733 A, the primary star of the system, orbits HD 189733 b, an exoplanet slightly bigger than Jupiter. $19 \pm 8\%$ bigger, to be precise.

It may not seem like much, but the fact that we know with confidence the diameter of HD 189733 b to within 10% is absolutely remarkable. The measurement made by Baines et al. [2007], who reported the value of $1.19 \pm 0.08\,R_{\rm Jup}$ for the radius of HD 189733 b, was indeed the first direct, model-independant, measurement of the diameter of an exoplanet ever made. It is also a great introduction to the complicated role played so far by optical interferometry in the quest for exoplanets, and a beautiful example of how current astronomical discoveries often rely on a diversity of instruments and methods.

How exactly could Baines et al. obtained such a measurement? Given that the HD 189733 system is located at about 20 parsecs (more than 60 light-years) from our Sun, the giant planet HD 189733 b only represents an angle of 30 μas on the sky. But the resolving power of a telescope of diameter D observing at a wavelength λ is set by the Rayleigh criterion, giving the angular resolution element:

$$\theta_{\rm Rayleigh} = 1.22 \times \frac{\lambda}{D} \tag{10.1}$$

which means that even at optical wavelengths, resolving HD 189733 b to measure its size would require a telescope aperture size of at least a few kilometers. We do not have that.

The answer to that puzzle lies in the fact that HD 189733 b is a transiting planet, meaning that once every orbit, the disk of the planet passes in front of its star relative to us, and blocks a fraction of the stellar light.

The initial discovery of HD 189733 b, reported in Bouchy et al. [2005], was not made by the transit method, but by radial velocity. The star was observed with the ELODIE spectrograph [Baranne et al., 1996], installed on the 193 cm telescope of the Observatoire de Haute-Provence, in France. But when Bouchy et al. found out that their radial velocity data showed the potential presence of a planet orbiting in 2.2 days around the star, they decided to attempt a transit detection. They used a 1.2 m telescope, also located at the Observatoire de Haute-Provence, and managed to observe three transits of the planet, confirming its existence, and giving a measurement of the relative transit depth.

The relative transit depth is defined as the fraction of the flux of the star blocked by the planet at the maximum of transit. Assuming that the two objects are circular in shape, the transit depth can be directly linked to the ratio of the planet area to the star area. For a planet of radius $R_{\rm planet}$ orbiting around a star of radius $R_\star$, $F_\star$ being the stellar flux, and $\Delta F_\star$ the drop in stellar flux at maximum of transit, we can write:

$$\frac{\Delta F_\star}{F_\star} = \left(\frac{R_{\rm planet}}{R_\star} \right)^2 \tag{10.2}$$

The transit depth measurement made by Bouchy et al. is a direct measurement of the size of the planet relative to its host star. It can then be converted to an absolute size by using a stellar model to link the observed photometry $F_\star$ and spectral type to the stellar diameter. But this only provides a model-dependent estimate of the planetary radius, with all the potential

biases implied by such an approach. A model-independent measurement of the planet radius requires a direct determination of the stellar diameter.

A star like HD 189733 A (an orange dwarf) is typically slightly smaller than our Sun, with a diameter of about 10^6 km. This is 10 times bigger than a typical giant planet, and thus only requires a telescope aperture size of a few 100 m to be resolved. Still, this is almost two orders of magnitude bigger than any telescope currently in existence, and one order of magnitude bigger than the size envisioned for the next generation telescopes. So, how do we measure a stellar diameter? Using long-baseline optical interferometry.

Take two telescopes of diameter D, observing at wavelength λ, separated by a distance B, and coherently combine the light coming from each one. The resolving power of such an instrument is not anymore set by the aperture size of each individual telescope, but rather by the separation distance B:

$$\theta_{\text{interf}} = \frac{\lambda}{2B} \tag{10.3}$$

Even relatively small 1 m telescopes, when spread a few 100 m apart and combined as a single interferometer, can be used to measure stellar diameters.

And this is exactly what Baines et al. did. They used the Georgia State University's Center for High Angular Resolution Astronomy (CHARA) array [McAlister et al., 2005], located atop Mount Wilson, California, to observe HD 189733 A and determine its angular size. The CHARA array is made of six 1 m class telescopes separated by up to 330 m, and operates in H (1.67 μm) or K' (2.13 μm) band. Using its H band mode, Baines et al. managed to determine the angular size of HD 189733 A. They found a value of $\Theta_\star = 0.377 \pm 0.024$ mas.

Of course, the CHARA interferometer, like virtually any astronomical instrument, is only sensitive to the *apparent* size of the star, not its *absolute* size. A star two times bigger than another (in radius), but located at twice the distance, will appear similar on the science detectors of CHARA. Converting the angular size of HD 189733 A to an absolute physical size requires to know its distance. Fortunately, HD 189733 A is on the Hipparcos catalog[1] [Perryman et al., 1997], which means that its annual parallax has been measured by the Hipparcos satellite: $\gamma_\star = 51.9 \pm 0.9$ mas. The annual parallax corresponds to the apparent angular displacement of a star seen from the Earth, due to the motion of our planet. It is a purely geometrical effect, whose amplitude is inversely proportional to the distance of the star (in the small-angle, or large-distance approximation). The annual parallax $\gamma_\star$ can be converted to a distance $d_\star$ from HD 189733 A to our Sun:

$$\gamma_\star = \frac{a_{\text{Earth}}}{d_\star} \tag{10.4}$$

where a_{Earth} is the average Sun-Earth distance (by definition equal to 1 astronomical unit, or 1 AU).

Putting together all of the above, we can write the direct determination of the radius of HD 189733 b in a single equation, which perfectly illustrates the role of the different techniques and instruments required to get it:

$$R_{\text{planet}} = \sqrt{\frac{\Delta F_\star}{F_\star}} \times \Theta \times \frac{1}{\gamma_\star} \times 1 \text{ AU} \tag{10.5}$$

- $\frac{\Delta F_\star}{F_\star}$ is the relative transit depth, measured using precision photometry on the star at the time of transit (this time of transit itself being determined with radial velocity measurements made with a high-resolution spectrograph)

[1] An now, also on the Gaia DR2 catalog: $\gamma_\star = 50.57$ mas

- $\Theta_\star$ is the angular size of the star, measured via long-baseline near-infrared interferometry
- $\gamma_\star$ is the annual parallax of the star, measured using precision astrometry from space

HD 189733 is not the only star which was observed with CHARA to determine its diameter. In fact, Baines et al. [2008] measured the diameters of 24 exoplanet host stars using the exact same technique, demonstrating the interest of long-baseline interferometry for exoplanet science.

10.2 Exozodiacal dust

Indirect determination of planetary parameters is not the only contribution made by interferometry in exoplanet science. The high-resolution capability of interferometric instruments has also been used to observe dust in debris disks.

Defrère et al. [2012] used VLTI/PIONIER to observe β Pictoris. They detected a small excess emission in the H band, indicative of the presence of hot dust (1500 K) in the close vicinity of the star (< 1 AU).

But this was not the first high-resolution detection of dust around a star, nor was it the most extensive study of this type of infrared emission. In fact, the infrared emission generated by exozodiacal dust, which can outshine small planets, has been recognized as a potential problem for future exoplanet missions. This has triggered an extensive program to observe and characterize exozodi, in which nulling interferometry proved extremely useful.

The concept of nulling interferometry, first introduced by Bracewell [1978], is simple: combine the light of two telescopes with a phase offset of π. This way, any on-axis target produces destructive interferences, and it becomes possible to look for faint objects slightly off-axis without having to deal with the light of a bright on-axis source.

Using the Keck Interferometer Nuller (KIN), Mennesson et al. [2014] found mid-IR excess emission around 5 stars known to exhibit far-IR excess. The NASA funded HOSTS survey [Danchi et al., 2014], which started in 2015 on the Large Binocular Telescope [Defrère et al., 2015], also looked for resolved mid-infrared emission around 30 carefully selected stars [Weinberger et al., 2015]. Early results of this study have shown that about 17% of solar type stars may host high level (≥ 100 zodis[2]) of exozodiacal dust [Ertel et al., 2018b]. In H-band, Ertel et al. [2014] used the VLTI/PIONIER instrument to look for evidence of debris disks around a total of 92 stars. At these wavelengths, the component probed is a hotter (≥ 1000 K) dust probably in the inner part of the disks, although the exact location of the emission is still debated. From their measurements, they estimated a near-IR excess rate of $10^{+4.3}_{-2.5}$%. This rate is significantly different than the 28^{+8}_{-6}% rate obtained with another sample of stars, observed in K-band with CHARA/FLUOR [Absil et al., 2013]. Ertel et al. [2014] suggested this could indicate that hot emission is dominated by scattering in the H band, while thermal emission gradually takes over in K band, making excesses more readily detectable in K-band. Perrin et al. [2016] and Mennesson et al. [2011] seem to agree with this hypothesis, but it still needs to be confirmed by spectrally resolved observations.

To this date, the exact connection between these different components (cold exo-Kuiper belts, warm HZ dust, hot component) remains largely unknown, and is still an active area of observational and theoretical research. One of the major question, already raised by Absil et al. [2013], is the question of the connection between the hot inner component of exozodiacal clouds, and the colder external component.

[2]The *zodi* is a measurement unit used in exozodi surveys. One zodi corresponds to the level of interplanetary dust emission in the Solar system.

Overall, our understanding of exozodiacal dust emission remains limited. But it makes no doubt that interferometry has made a significant contribution to this field (see Table 10.1 listing all surveys dedicated to exozodiacal dust).

Survey	Instrument	Type	Band	Targets	Detections	Reference
Spitzer	IRS	Slit spectrograph	30-34μm	152	$\simeq 16$	Lawler et al. 2009
Spitzer	IRS	Slit spectrograph	N	152+51	2	Lawler et al. 2009
HOSTS	LBTI	Long-baseline nuller	N	38	10	Ertel et al. 2018a
CHARA	FLUOR	Long-baseline	K	42	11	Absil et al. 2013
VLTI	PIONIER	Long-baseline	H	92	9	Ertel et al. 2014
Keck	KIN	Long-baseline	N	47	5	Mennesson et al. 2014
Palomar	PFN	Aperture masking nuller	K	10	1	Mennesson, priv. com. 2018

Table 10.1: Major surveys dedicated to exozodi. All instruments except Spitzer/IRS are interferometers.

10.3 Detecting planets?

More than 20 years ago, Angel and Woolf [1997] proposed to build a space-based nulling interferometer to detect exoplanets and characterize their atmospheres. The objective was to detect carbon dioxide, water, or ozone in the atmosphere of Earth-like planets. This apparently started a trend, and several projects based around a space interferometer were proposed to search for exoplanets. The NASA SIM-PlanetQuest mission, later renamed SIM Lite, was supposed to be a 6 mirror interferometer on a single satellite platform. It should have surveyed more than 60 stars, looking for Earth-sized planets using high-precision astrometry. The Terrestrial Planet Finder (TPF) was even more ambitious, with four satellites flying in formation, each carrying a 2 m class telescope. Array configurations with baselines up to 400 m have been suggested, to allow for an angular resolution of 2.4 mas at operating wavelengths (6 - 20 μm). The goal was again to detect exo-Earths in the habitable zone of their stars. TPF was also a NASA mission. But the European Space Agency, unwilling to stay behind, considered a similar concept: their Darwin mission. Here again, formation flying, 1 to 2 meter class telescopes, and baselines of up to 200 m were on the menu. None of these missions went off the drawing-board.

But grandiose has never been a privilege of space agencies. Back on Earth, the Planet Formation Imager (PFI) concept was proposed in 2013. The reference design is still preliminary, and consists in twelve 3 m class telescopes, in a "Y" array, with baselines of up to 1.2 km. The science objective differs from the space-based missions: PFI targets young planetary systems in L/M band, to study planet formation mechanisms.

On a smaller scale, Hi-5 [Defrère et al., 2018] is a concept for a new generation beam combiner for the VLTI, using Kernel nulling [Martinache and Ireland, 2018]. The concept is to combine long-baseline interferometry with nulling interferometry. It differs from another project proposed a few years before by Lacour et al. [2014], which consisted in combining nulling interferometry and closure phases, but the basic idea is the same: use nulling to get rid of photon noise, and find an observable robust to instrumental and atmospheric noise. In these proposals, the objective is again clearly stated: find new exoplanets.

Existing interferometers have also been proposed as exoplanet finders. For example, Sahlmann et al. [2012] proposed to use the astrometric capabilities of VLTI/PRIMA to look for planets around dwarf stars. The Very Large Telescope Interferometer, The Keck interferometer, the CHARA array, the Palomar Fiber nuller, the Large Binocular Telescope Interferometer, the Navy Prototype Interferometer... All of them have been suggested as

planet finders at one point or another, as illustrated by Beuzit et al. [2007] (see their Table 1, reproduced here in Table 10.2).

To this day, no planet has ever been detected by any of these projects.

10.4 GRAVITY, and the place of this work

If anything, Table 10.2 shows that detecting planets has been a long-standing goal for interferometrists. But such a detection was not possible with previous instruments, which could not reach the necessary contrast ratio.

The situation changed with GRAVITY, a second generation instrument on the Very Large Telescope Interferometer (VLTI), and I was fortunate enough to be involved in the two first interferometric detections of exoplanets. The first, HR 8799 e, is reported in Gravity Collaboration et al. [2019]. The second, β Pictoris b, is presented in Chapter 13 of this thesis. They were both made with GRAVITY, using a data reduction method I developed, and which is presented in Chapter 12.

But why is it now possible to detect exoplanets with long-baseline interferometry? What makes GRAVITY so special? The answer is short: *dual-field*. A much longer answer can be found in Chapter 11, in which I present the GRAVITY instrument, and this particular mode. In a few words, GRAVITY is the first interferometer capable of dual-field observations, a mode allowing to interferometrically observe a star-planet system for which the two components do not simultaneously enter the science fiber. The two components are observed one at a time, and the phase of the planet is referenced to the phase of the star. In this configuration, the combination of the single-mode fiber and the Adaptive Optics filters out a large part of the stellar light. The planet being slightly off-axis, its complex visibility is also modulated by the Optical Path Difference between the telescopes, which further helps to distinguish the planet coherent flux from the remaining stellar light. This overcomes the main barrier to exoplanet detection with interferometry: contrast.

Interferometer	Method	Goals
CHARA Interferometer	Differential Closure Phases	Measure low-res spectrum of hot Jupiters in H/K bands
	Precision Visibilities	Hot Jupiter detection using FLUOR combiner
Keck Interferometer	Differential Phase	Hot Jupiter detection using 1-5 μm wavelength range
	Nulling	Mid-infrared search for zodiacal dust might uncover close-in planet
	Narrow-Angle Astrometry	Planet detection using Keck-Keck and outrigger array
Navy Prototype Interferometer	Imaging	Detect hot Jupiter transit by imaging stellar photosphere in visible
Palomar Testbed Interferometer	Narrow-Angle Astrometry	Detect massive planets around sub-arcsecond binaries
Very Large Telescope Interferometer	Differential Phase	Low-res spectrum of hot/warm exoplanets (H/K bands)
	Differential Closure Phase	Hot Jupiters and perhaps more difficult planets
	Narrow-Angle Astrometry	PRIMA instrument will allow long-term astrometry
	Nulling	GENIE instrument meant to detect low-mass companions (K,L,M bands)
	Future Plans	
Antarctic Plateau Interferometer	Differential Phase/Closure Phase	Hot Jupiter planet characterization (3-5 μm)
Darwin (ESA)	Nulling Space Interferometer	Terrestrial Planet Finding (mid-infrared)
Large Binocular Telescope	Nulling	Detect (warm) massive exoplanets in thermal IR
Space Interferometry Mission Interferometer (NASA)	Astrometry	Detect low-mass exoplanets via induced planet wobble
Terrestrial Planet Finder Interferometer (NASA)	Nulling Space Interferometer	Terrestrial Planet Finding (mid-infrared)

Table 10.2: Proposed exoplanet searches using interferometry (Table from Beuzit et al. [2007]).

Chapter 11

Long-baseline interferometry with VLTI/GRAVITY

Content of this chapter

11.1 The concept of long-baseline interferometry

As stated in Chapter 10, the angular resolution of a telescope of diameter D observing at wavelength λ is given by the Rayleigh criterion, and corresponds to an angular resolution element $\theta_{\text{Rayleigh}} = 1.22 \times \lambda/D$. But the resolution element of two telescopes of diameter D separated by a distance B is given by $\theta_{\text{interf}} = 0.5 \times \lambda/B$. The goal of this section is to explain these two numbers.

11.1.1 Diffraction limit of a single-dish telescope

Image of a complex object by an astronomical telescope

Consider an astronomical object, whose angular intensity distribution on the sky is given by $F(\alpha, \beta)$, α and β being referred to the optical axis, the image created by the telescope is the incoherent sum of the images of all the angular elements of the source:

$$I(X,Y) = \iint_{\alpha,\beta} I_{\alpha,\beta}(X,Y) \tag{11.1}$$

where $I_{\alpha,\beta}(X,Y)$ is the image of the angular element located at α, β.

Considering only this angular element at (α, β) of the source, and with the hypothesis that the object is far enough so that the incoming wavefront is plane, we can determine the expression of the complex amplitude of the electric field in the pupil plane:

$$A_{\alpha,\beta}(x,y) = \sqrt{F(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta} \times e^{i\frac{2\pi}{\lambda}(\alpha x + \beta y)} \tag{11.2}$$

The same calculations as performed to obtain the result of Equation A.11 in Annex A would lead to the following expression for the intensity in the pupil plane:

$$I_{\alpha,\beta}(X,Y) = F(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta\,\mathrm{PSF}\,(X - f\alpha, Y - f\beta) \tag{11.3}$$

In which we recognize what could be expected: the instrumental Point-Spread Function (PSF), but centered on the geometrical image of the angular element, and with and intensity given by the source intensity.

The expression of the image of the full object is thus:

$$I(X,Y) = \iint_{\alpha,\beta} F(\alpha,\beta) \times \mathrm{PSF}(X - f\alpha, Y - f\beta)\mathrm{d}\alpha\mathrm{d}\beta \tag{11.4}$$

Or, with a simple change of variable:

$$I(X,Y) = f^2 \iint_{\alpha,\beta} F\left(\frac{x}{f}, \frac{y}{f}\right) \times \mathrm{PSF}(X - x, Y - y)\mathrm{d}x\mathrm{d}y \tag{11.5}$$

In which we recognize the convolution of the geometrical image of the source by the instrumental PSF:

$$I(X,Y) = \left(\left[f^2 F\left(\frac{x}{f}, \frac{y}{f}\right)\right] * \mathrm{PSF}(x,y)\right)(X,Y) \tag{11.6}$$

The resolving power of a single-dish telescope

If we consider a binary star, with two stars separated by an angle $\Delta\theta$ on the sky, we can use Equation 11.6 to determine the image through a simple telescope of diameter D and focal length f. If we consider that the primary star is centered on the optical axis of the instrument, with the X axis of the focal plane in the direction of the binary, then we have:

$$I(X,Y) \propto \mathrm{PSF}(X,Y) + \mathrm{PSF}(X - f\Delta\theta, Y) \tag{11.7}$$

Given that the instrumental PSF has a certain width (see Figure A.2 in Annex A), depending on the value of $\Delta\theta$ compared to the instrument diameter D, the observer may or may not be able to distinguish the two individual components of the binary system.

The first zero of the instrumental PSF occurs at $X = 1.22 \times \lambda f/D$. If the binary angular separation is $1.22 \times \lambda/D$, then the maximum of the secondary occurs exactly at the zero of the primary. The binary is *resolved*. Any binary with a smaller separation angle will remain unresolved by the instrument. This is the origin of the Rayleigh criterion.

In practice, with careful modeling of the instrumental response, it is possible to go slightly beyond the Rayleigh criterion. But not by much, and any telescope with a diameter D will only be able to resolve separations down to $\sim \lambda/D$. This is the *diffraction limit*.

11.1.2 Aperture masking

If a telescope of diameter D is not able to resolve a binary system with separation $\Delta\theta < \lambda/D$, this is because, at the scale of the mirror, the incident wavefront of the off-axis component of the binary appear similar to the wavefront of the on-axis component. The phase slope of the off-axis component (the exponential term of Equation 11.2) is not strong enough to have any impact on the image.

Intuitively, one way to deal with this issue could be to sample the incident wavefront on larger distances. A bigger telescope would of course do this. But even an instrument made of two small parts of a bigger telescope would, in principle, work. If we suppose we only have two small parts of diameter $D/\sqrt{2}$ from a larger parabola, separated by a distance B (see Figure 11.1), then the total collecting surface is equivalent to a telescope of diameter D, but the instrumental PSF is given by:

$$\mathrm{PSF}_2(X,Y) = \mathrm{TF}\left[P(x,y)\right](X,Y) \tag{11.8}$$

where P is the pupil function, which can be written:

$$P(x,y) = P' * (\delta(x - B/2, y) + \delta(x + B/2, y)) \tag{11.9}$$

with δ the δ-Dirac function, and P' the circular aperture function of diameter $D/\sqrt{2}$
Basic Fourier transform calculus then gives:

$$\mathrm{PSF}_2(X,Y) = \mathrm{PSF}(X,Y) \times \sin\left(\frac{\pi B x}{\lambda}\right) \tag{11.10}$$

The shape of this new instrument PSF is given in Figure 11.2.

The PSF of an aperture made of two sub-apertures of diameter $D/\sqrt{2}$ separated by a distance B has its first zero at $X = 0.5 \times \lambda/B$. The resolving power of such an instrument is thus of the order of λ/B, and not λ/D anymore, although the total surface area is $\sim D^2$

The situation described here is difficult to realize in practice without having an effective mirror of diameter B on which two sub-apertures are taken. This is the idea of aperture masking, which does not help to go beyond the diffraction limit *per se*, but which is extremely useful to actually reach this limit even in the presence of atmospheric perturbations (which we have completely neglected so far).

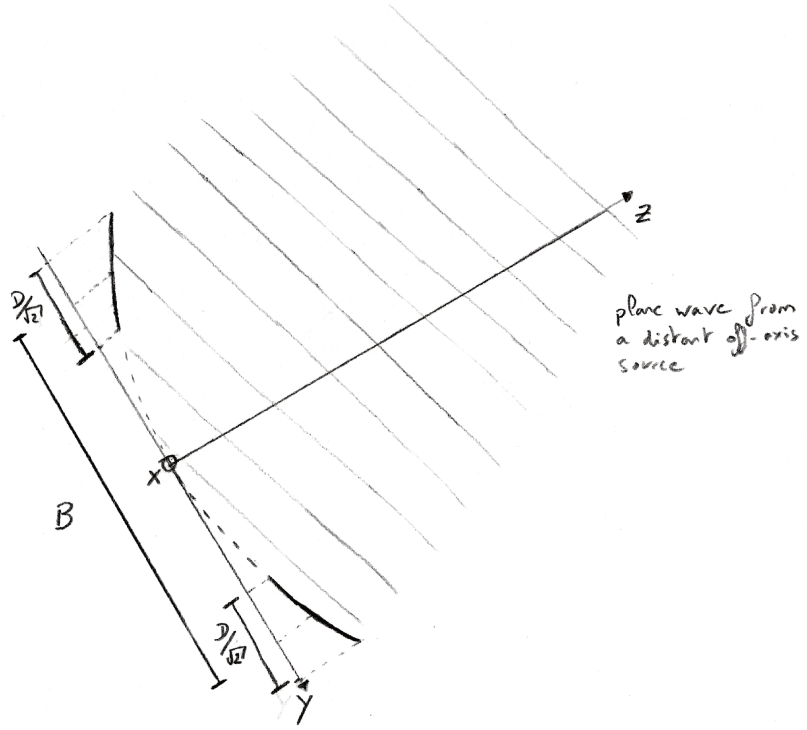


Figure 11.1: Two small telescopes of diameter $D/\sqrt{2}$ separated by a distance B. Each telescope is made of a subsection of a larger parabola (in dotted line). The whole setup is illuminated by a distant off-axis point source generating an incident plane wave.

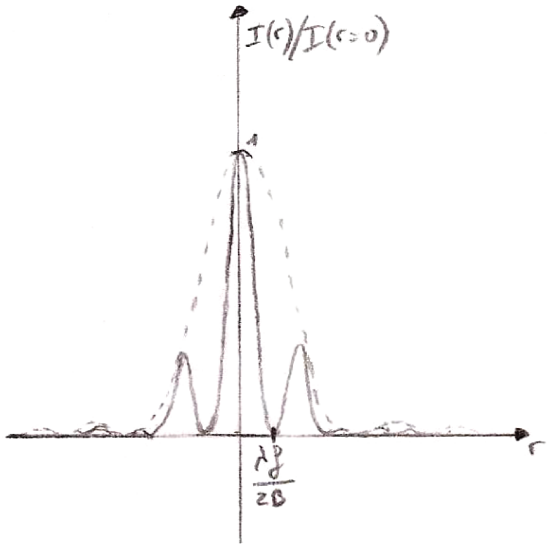


Figure 11.2: The shape of the instrumental Point-Spread-Function (PSF) for a telescope made of two sub-apertures of diameter $D/\sqrt{2}$ taken on a larger mirror and separated by a distance B.

11.1.3 Long-baseline interferometry

There is an alternate way to aperture masking, though, which consists in taking two independent telescopes, each one with its own mirror, and to recombine the light of the two in a *beam combiner*. In this case, the separation B of the two telescopes can be set independently of the size D of their mirrors. When $B \gg D$, the instrument is said to be in the "long-baseline" regime.

In the long-baseline regime, the telescope diameter D can be neglected compared to the interferometric baseline B. We can consider that the two telescope interferometer samples the wavefront at two points $(x - B/2, y)$ and $(x + B/2, y)$. If such an instrument observes a monochromatic point source in the direction (α, β), the electric field at the output of the beam combiner is the sum of the two fields collected by the telescopes T_1 and T_2:

$$E(t) = E_{T_1}(t) + E_{T_2}(t) \tag{11.11}$$

The complex amplitude $A(x, y, z = 0)$ of the incident field in the plane of the telescope is given by:

$$A(x, y) = E_0 \exp\left\{ i \frac{2\pi}{\lambda} (\alpha x + \beta y) \right\} \tag{11.12}$$

The total field after the beam combiner is the sum of two fields:

$$E(t) = e^{-i\omega t} \left[A(x - B/2, y) + A(x + B/2, y) \right] \tag{11.13}$$

$$= E_0 e^{-i\omega t} \left[\exp\left(-i \frac{2\pi\alpha B}{2\lambda} \right) + \exp\left(i \frac{2\pi\alpha B}{2\lambda} \right) \right] \tag{11.14}$$

Which simplifies in:

$$E(t) = 2 \times E_0 e^{-i\omega t} \cos\left(\frac{\pi\alpha B}{\lambda} \right) \tag{11.15}$$

We can get the intensity by squaring the modulus of the electric field:

$$I(t) = 2 \times I_0 \left[\cos\left(\frac{\pi\alpha B}{\lambda} \right) \right]^2 \tag{11.16}$$

And if the two-telescope interferometer observes a binary system, with one star of intensity I_A at zenith, and one star of intensity I_B at (α, β), then the intensity is the incoherent sum of the intensities produced by each star:

$$I(t) = 2 \left(I_A + I_B \left[\cos\left(\frac{\pi\alpha B}{\lambda} \right) \right]^2 \right) \tag{11.17}$$

Once again, we note that the system is fully resolved when the peak of component B corresponds to the first zero of component A, which corresponds to $\alpha = 0.5 \times \lambda/B$. A long-baseline two-telescope interferometer has an angular resolution element of $\sim \lambda/B$.

11.2 Basic formalism of long-baseline interferometry

11.2.1 The UV plane

Here we consider again the case of a long-baseline interferometer made of two telescopes T_1 and T_2, located at arbitrary positions $\mathbf{P_1} = (x_{T_1}, y_{T_1}, z_{T_1})$ and $\mathbf{P_2} = (x_{T_2}, y_{T_2}, z_{T_2})$ in space

(Figure 11.3). An object in the direction $\mathbf{k}$ in the sky will create an electric field at the two telescopes given by:

$$E(t, T_1) = A_{T_1} e^{-i\omega t} \tag{11.18}$$
$$E(t, T_2) = A_{T_2} e^{-i\omega t} \tag{11.19}$$

With A_{T_1} and A_{T_2} the complex ampltitude at the two telescopes:

$$A_{T_1} = \exp\left(i\frac{2\pi}{\lambda}\left(\mathbf{k}.\mathbf{P_1}\right)\right) E_0 \tag{11.20}$$
$$A_{T_2} = \exp\left(i\frac{2\pi}{\lambda}\left(\mathbf{k}.\mathbf{P_2}\right)\right) E_0 \tag{11.21}$$

Ignoring the problem of beam-combination, and assuming that the fields coming from the two telescopes can somehow be added, the intensity measured by the interferometer is:

$$I(T_1T_2, \lambda) = \|A(t, T_1) + A(t, T_2)\|^2 \tag{11.22}$$

Injecting the expression of the two fields:

$$I(T_1T_2, \lambda) = \|E_0\|^2 \left\|\exp\left(i\,\frac{2\pi}{\lambda}\left(\mathbf{k}.\mathbf{P_1}\right)\right)\left[1 + \exp\left(i\,\frac{2\pi}{\lambda}\left(\mathbf{k}.\left(\mathbf{P_2} - \mathbf{P_1}\right)\right)\right)\right]\right\|^2$$

This simplifies to:

$$\begin{aligned} I(T_1T_2, \lambda) &= I_0 \left\|1 + \exp\left(i\,\frac{2\pi}{\lambda}\left(\mathbf{k}.\left(\mathbf{P_2} - \mathbf{P_1}\right)\right)\right)\right\|^2 \\ &= 2I_0 + 2I_0 \cos\left(\frac{2\pi}{\lambda}\left(\mathbf{k}.\left(\mathbf{P_2} - \mathbf{P_1}\right)\right)\right) \end{aligned}$$

We can use two angular coordinates for $\mathbf{k}$, by using the Y axis as the polar axis of a spherical coordinate system. In this case, we have:

$$\mathbf{k} = \begin{pmatrix} \cos(\beta)\sin(\alpha) \\ \sin(\beta) \\ \cos(\alpha)\cos(\beta) \end{pmatrix} \tag{11.23}$$

Under the assumption that α and β are both small angles:

$$\mathbf{k} = \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} \tag{11.24}$$

At which point the expression for the measured intensity becomes:

$$I(T_1T_2, \lambda) = 2I_0 + 2I_0 \cos\left(\frac{2\pi}{\lambda}\left(\alpha(x_{T_2} - x_{T_1}) + \beta\left(y_{T_2} - y_{T_1}\right) + z_{T_2} - z_{T_1}\right)\right) \tag{11.25}$$

If the two telescopes are on the same mount (as it is the case for the Large Binocular Telescope Interferometer, in Arizona), the z-coordinates of the two telescopes in the pointing direction will always be the same. If the two telescopes are not on the same mount, there will always

be a dedicated system (delay line) to compensate for the $z_{T_2} - z_{T_1}$ term, which only depends on the pointing direction. This term can usually be ignored. We are left with:

$$I(T_1T_2, \lambda) = 2I_0 + 2I_0 \cos\left(\frac{2\pi}{\lambda}\left(\alpha(x_{T_2} - x_{T_1}) + \beta\left(y_{T_2} - y_{T_1}\right)\right)\right) \tag{11.26}$$

So far, no assumption was made on the exact orientation of the X and Y axes. We can chose the XY directions, orthogonal to the Z axis, to correspond to the image plane: X is oriented to the East, Y to the North. Then, the α and β introduced before actually corresponds to the astronomical ΔRA and ΔDEC (ignoring a cos(DEC) factor...).

We can now define the (u, v, w) coordinates of the baseline T_1T_2 by:

$$u(T_1T_2, \lambda) = \frac{x_{T_2} - x_{T_1}}{\lambda} \tag{11.27}$$

$$v(T_1T_2, \lambda) = \frac{y_{T_2} - y_{T_1}}{\lambda} \tag{11.28}$$

$$w(T_1T_2, \lambda) = \frac{z_{T_2} - z_{T_1}}{\lambda} \tag{11.29}$$

The (u, v, w) coordinates of a baseline are simply the baseline vector coordinates in the image plane & pointing axis reference frame, and expressed in unit of λ.

We can also introduce a dimensional but wavelength-independent version of these coordinates:

$$U(T_1T_2) = x_{T_2} - x_{T_1} \tag{11.30}$$

$$V(T_1T_2) = y_{T_2} - y_{T_1} \tag{11.31}$$

$$W(T_1T_2) = z_{T_2} - z_{T_1} \tag{11.32}$$

The expression for U, V, W may seem trivial, but it is not. For ground-based telescopes, the image-plane/pointing axis frame of reference is non-trivial, as it varies both with the pointing direction and with time, due to rotation of the Earth. And so do the UVW coordinates of the baseline.

With these coordinates, the expression of the interferometric intensity can now be written:

$$I(T_1T_2, \lambda) = 2I_0 + 2I_0 \cos\left(2\pi\left[\Delta\text{RA} \times u(T_1T_2, \lambda) + \Delta\text{DEC} \times v(T_1T_2, \lambda)\right]\right) \tag{11.33}$$

11.2.2 Introducing the complex visibility

Complex visibility and Fourier Transform formalism The above expression of the intensity is perfectly valid, and has the merit of a strong connection with what can actually be observed on the science detector of a long-baseline interferometer: the cosine term reflects the presence of oscillations when the intensity is dispersed in λ. These oscillations are not interferometric fringes *per se*, though. They occur because the phase difference between the two telescopes varies with λ. True interferometric fringes only appear when the phase difference between the two telescopes is effectively scanned at a given λ (using delay lines, for example).

When observing more complex objects, the expression of the intensity becomes more complicated. If we consider an object with angular intensity distribution $I_{\alpha,\beta}$, then the expression of the intensity on the science detector is the incoherent sum of each elementary source:

$$I(b_k, \lambda) = \iint_{\alpha,\beta} 2I_{\alpha,\beta} + 2I_{\alpha,\beta} \cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right)\right) \mathrm{d}\alpha\mathrm{d}\beta \tag{11.34}$$

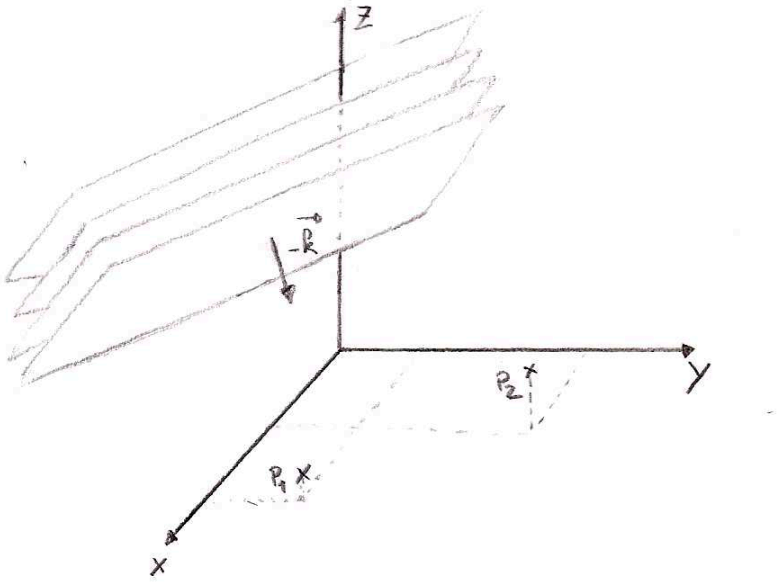


Figure 11.3: Two telescopes located at P_1 and P_2 are illuminated by a distant point source in direction $\mathbf{k}$.

In this expression we can see that when the source has some spatial complexity, the actual fringe pattern on the detector $I(b_k, \lambda)$ is difficult to read. In practice, the fringes are blurred. This expression can be separated in two terms:

$$I(b_k, \lambda) = \iint_{\alpha,\beta} 2I_{\alpha,\beta} \mathrm{d}\alpha\mathrm{d}\beta + \iint_{\alpha,\beta} 2I_{\alpha,\beta} \cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right)\right) \mathrm{d}\alpha\mathrm{d}\beta \tag{11.35}$$

The first term corresponds to what one would get if one was to sum the intensity acquired by the two telescopes separately. This term is the *incoherent flux*. The second term, called *coherent flux* is actually the real part of the Fourier transform of the source's angular intensity distribution:

$$\iint_{\alpha,\beta} 2I_{\alpha,\beta} \cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right)\right) \mathrm{d}\alpha\mathrm{d}\beta = \mathrm{Re}\left(\mathrm{TF}\left[I_{\alpha,\beta}\right]\left(\frac{U_k}{\lambda}, \frac{V_k}{\lambda}\right)\right) \tag{11.36}$$

This motivates the introduction of the *complex visibility* function:

$$V(b_k, \lambda) = \mathrm{TF}\left[I_{\alpha,\beta}\right]\left(\frac{U_k}{\lambda}, \frac{V_k}{\lambda}\right) \tag{11.37}$$

So we can write:

$$I(b_k, \lambda) = I_{\mathrm{incoherent}} + \mathrm{Re}(V(b_k, \lambda)) \tag{11.38}$$

Complex visibility and ABCD recombination Even though the introduction of the complex visibility is fully justified by the Fourier transform relationship between the observed fringe pattern and the angular intensity distribution of the source, introducing and working with this quantity can also be motivated purely from an instrumental/data reduction perspective, since GRAVITY uses an ABCD beam combiner.

In an ABCD beam combiner, the intensity is measured on 4 different outputs. For each output, a phase offset is added between the fields of the two telescopes. We have:

$$I_A(b_k,\lambda) = \left\| A(t,T_{k_1}) + A(t,T_{k_2})e^{i\phi_A} \right\|^2 \tag{11.39}$$

$$I_B(b_k,\lambda) = \left\| A(t,T_{k_1}) + A(t,T_{k_2})e^{i\phi_B} \right\|^2 \tag{11.40}$$

$$I_C(b_k,\lambda) = \left\| A(t,T_{k_1}) + A(t,T_{k_2})e^{i\phi_C} \right\|^2 \tag{11.41}$$

$$I_D(b_k,\lambda) = \left\| A(t,T_{k_1}) + A(t,T_{k_2})e^{i\phi_D} \right\|^2 \tag{11.42}$$

Which leads to:

$$I_A(b_k,\lambda) = 2I_0 + 2I_0\cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right) + \phi_A\right) \tag{11.43}$$

$$I_B(b_k,\lambda) = 2I_0 + 2I_0\cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right) + \phi_B\right) \tag{11.44}$$

$$I_C(b_k,\lambda) = 2I_0 + 2I_0\cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right) + \phi_C\right) \tag{11.45}$$

$$I_D(b_k,\lambda) = 2I_0 + 2I_0\cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right) + \phi_D\right) \tag{11.46}$$

By taking:

$$\phi_A = 0 \tag{11.47}$$

$$\phi_B = \frac{\pi}{2} \tag{11.48}$$

$$\phi_C = \pi \tag{11.49}$$

$$\phi_D = \frac{3\pi}{2} \tag{11.50}$$

We then have:

$$I_A(b_k,\lambda) = 2I_0 + 2I_0\cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right)\right) \tag{11.51}$$

$$I_B(b_k,\lambda) = 2I_0 - 2I_0\sin\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right)\right) \tag{11.52}$$

$$I_C(b_k,\lambda) = 2I_0 - 2I_0\cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right)\right) \tag{11.53}$$

$$I_D(b_k,\lambda) = 2I_0 + 2I_0\sin\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right)\right) \tag{11.54}$$

And we can form the quantity:

$$\frac{I_A - I_C}{2} + i \times \frac{I_D - I_B}{2} = 2I_0\, e^{-i\left(\frac{2\pi}{\lambda}(\alpha U_k + \beta V_k)\right)} \tag{11.55}$$

Generalizing to a complex object of spatial distribution $I_{\alpha,\beta}$, it is easy to see that this quantity is exactly the complex visibility introduced before:

$$\frac{I_A - I_C}{2} + i \times \frac{I_D - I_B}{2} = \iint_{\alpha,\beta} 2I_{\alpha,\beta}\, e^{-i\left(\frac{2\pi}{\lambda}(\alpha U_k + \beta V_k)\right)} \mathrm{d}\alpha \mathrm{d}\beta = V(b_k,\lambda) \tag{11.56}$$

Baseline	Length (meters)
UT 4 - UT 3	62.4
UT 4 - UT 2	89.3
UT 4 - UT 1	130.2
UT 3 - UT 2	46.6
UT 3 - UT 1	102.4
UT 2 - UT 1	56.5
AT x - AT y	from 8 to 202

Table 11.1: VLTI baseline lengths

The ABCD combiner samples the true interferometric fringes (i.e. the oscillation of the intensity as a function of the phase difference between the two fields) in 4 points separated by $\pi/2$ to directly recover the full complex visibility. At each λ, this complex visibility $V(b_k, \lambda)$ contains all the information encoded in the fringe pattern $\phi \rightarrow \left\| A_{T_{k_1}} + A_{T_{k_2}} e^{i\phi} \right\|^2$.

This is the beauty of ABCD-based modern interferometry: the interferometric fringes are never actually recorded by the instrument! They are sampled on 4 points, so as to reconstruct the full complex visibility, which is the only quantity accessible to the end-user.

11.3 Fiber-fed long-baseline interferometry with GRAVITY

11.3.1 GRAVITY and the VLTI

The Very Large Telescope Interferometer is a facility which can combine the light of the four 8.2 m Unit Telescope (UTs) or the four 1.8 m Auxiliary Telescopes (ATs) of the Very Large Telescope (VLT), located on top of Cerro Paranal, Chile. The VLTI includes a set of delay lines which can provide up to 60 m of optical path delay.

The 4 UTs are fixed, providing 6 different baselines (see Table 11.1). The 4 ATs can be moved at different positions prior to an observing run. A total of 30 different positions are accessible to the ATs, for baseline lengths ranging from 8 to 202 m (see Figure 11.4).

The VLTI is also equipped with MACAO, an adaptive optics system for the UTs which can deliver moderate Strehl ratio (~ 0.5, see Figure 11.5). Another AO system, CIAO, also operates on the UTs, but only in the infrared. As of 2019, a new adaptive optics system for the ATs is being commissioned: NAOMI.

GRAVITY is a 2nd generation instrument for the VLTI. It is a 4-beam K-band interferometer, which began commissioning on the ATs in 2015, and on the UTs in 2016. The science operations started in 2016 for the ATs, and in 2017 for the UTs. An overview of the working principle is given in Figure 11.6.

11.3.2 Fiber injection and integrated optics ABCD beam combination

Until now, I have purposedly ignored the problem of beam combination. I have just assumed that it was possible, somehow, to take the electric field at the position of one telescope, add it to the electric field at the position of another telescope, and measure the corresponding light intensity. I have even stated that, in the case of ABCD beam combination, one could do this while adding an additional phase offset to one of the two fields. In practice, this is not so easy to do, and the problem of beam combination is always at the heart of the development of an

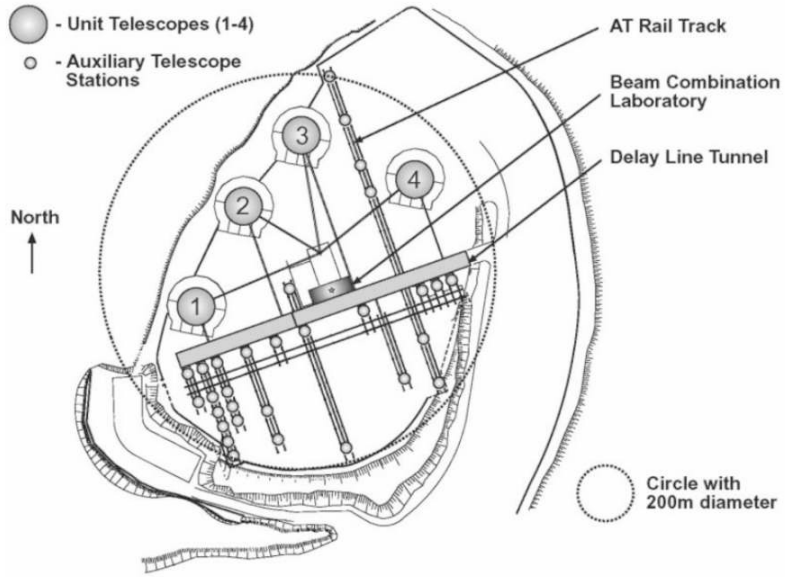


Figure 11.4: Map of the Very Large Telescope Interferometer, showing the 4 Unit Telescopes (numbered gray disks), and the 30 possible stations for the Auxiliary Telescopes (small red circles). Figure taken from Koehler [2001].

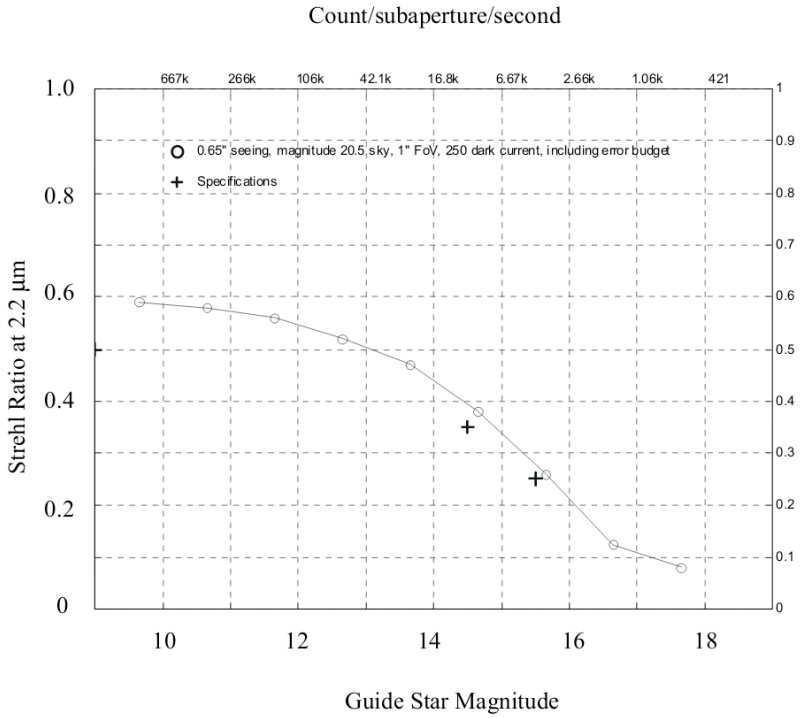


Figure 11.5: Performance of the MACAO adaptive optics installed on the Unit Telescopes and used with the Very Large Telescope Interferometer. Figure taken from Koehler [2001].

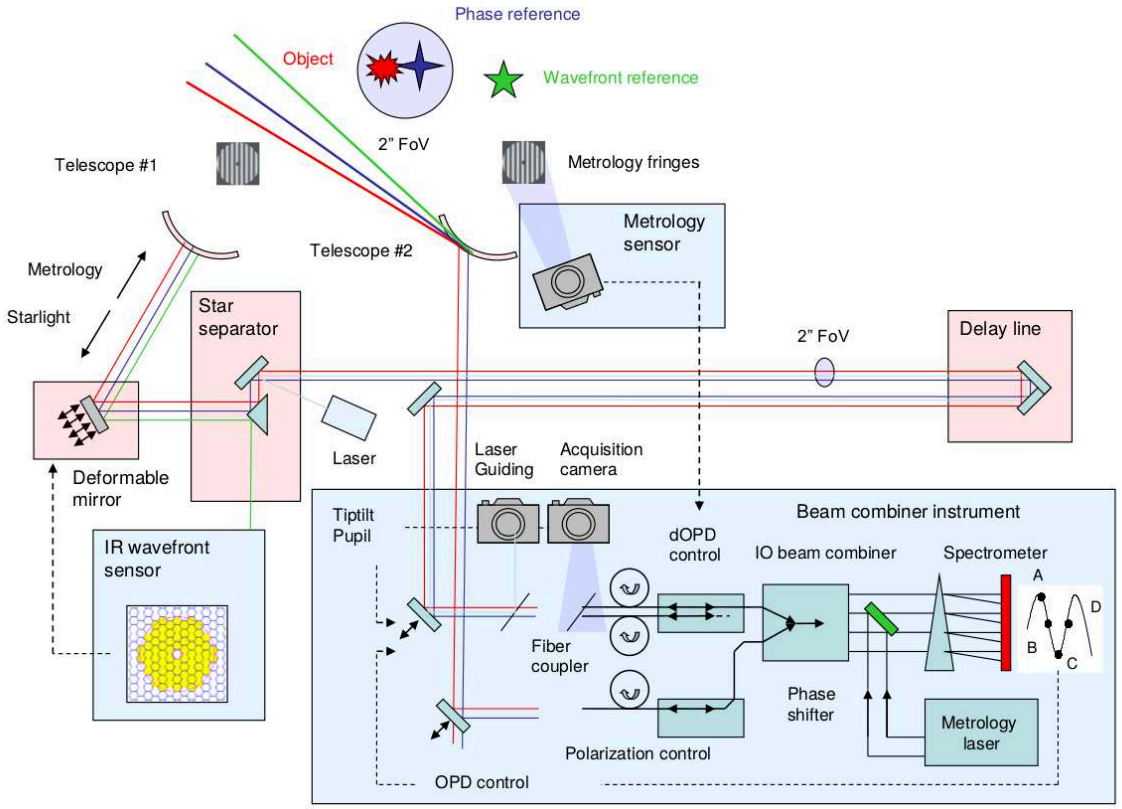


Figure 11.6: Working principle of VLTI/GRAVITY. This instrument can combine the light of 4 telescopes (4 UTs or 4 ATs), and extract the complex visibilities. An Adaptive Optics system equips each telescope, and provides moderate Strehl ratios. The light from each telescope is injected in a single-mode fiber, and the beam combiner is made with integrated optics. Interferometric fringes are spectrally dispersed on the detector to provide low, medium, or high spectral resolution capabilities in K band (R = 22, 500, or 4500). The VLTI delay lines are used to control large phase offsets between the beams of different telescopes, and the dOPD control system is used for fringe-tracking. The metrology laser counter-propagates in the optical train all the way to the telescopes. The laser light coming from the fringe tracking fiber (wavefront reference, in green), and science fiber (blue and red) produces interferometric fringes in the pupil plane of each telescope. These fringes are measured by dedicated diodes, and are used to control the OPD, and to correct for non-common path errors at data-reduction level. Figure from Gravity Collaboration et al. [2017].

interferometer. The approach taken by GRAVITY makes heavy use of fibers and integrated optics.

Consider a telescope of diameter D, focal length f, and put a single mode fiber in the focal plane of the telescope. The classical calculation of the electric field in the focal plane of the telescope (Fraunhoffer approximation) is given in Annex A, and leads to Equation A.9. At position (X, Y) in the image plane, the complex amplitude of the electric field is given by:

$$A_{\text{image}}(X,Y) = \text{TF}\left[P(x,y)A_{\text{pupil}}(x,y)\right]\left(\frac{X}{\lambda f}, \frac{Y}{\lambda f}\right)$$

where $P(x, y)$ is the pupil aperture function, and $A_{\text{pupil}}(x, y)$ is the complex amplitude of the field in the pupil plane.

The complex amplitude of the field injected into the fiber is then given by:

$$A_{\text{fiber}} = \iint A_{\text{image}}(X,Y)A_{\text{mode}}(X,Y)\text{d}X\text{d}Y \tag{11.57}$$

where A_{mode} is the complex amplitude of the fiber fundamental mode (see also Equation A.14 in Annex A). Using Fourier transform calculus[1], we can calculate the injected field directly in the pupil plane:

$$\begin{aligned} A_{\text{fiber}} &= \iint \text{TF}\left[A_{\text{image}}(X,Y)\right](x,y)\text{TF}\left[A_{\text{mode}}(X,Y)\right](x,y)\text{d}x\text{d}y \\ &= \iint P(x,y)A_{\text{pupil}}(x,y)\rho(x,y)\text{d}x\text{d}y \end{aligned}$$

Where we have denoted ρ the Fourier transform of the fiber fundamental mode field. This is called the *fiber pupil illumination*. If we can neglect the diameter of the telescope compared to the baseline of the interferometer (long-baseline), then we can write:

$$P(x,y) \simeq \delta(x,y) \tag{11.58}$$

And since G is normalized, $\rho(0,0) = 1$, and we have:

$$A_{\text{fiber}} = A_{\text{pupil}}(0,0) \tag{11.59}$$

The field injected into the fiber corresponds to the field at $(0, 0)$, the center of the telescope pupil[2].

11.3.2.1 Integrated optics beam combiner

The field coming from the astronomical source is injected separately in two single-mode fibers in two different telescopes T_{k_1} and T_{k_2}. One question remains: how to combine them?

One possible option is to re-image the field coming from the two fibers onto a third fiber, using a beamsplitter. With an optical setup similar to the one given in Figure 11.7, the total field in fiber 3 is the sum of the fields propagating in fiber 1 and 2.

This setup has one major disadvantage: it uses bulk optics (collimating lenses and beamsplitter), which makes it more difficult to control in temperature. There is another way to combine the field of two fibers, which does not require the use of bulk optics whatsoever: fibered beamsplitters. The basic concept of a fibered beam-combiner is to drive the two

[1] Namely the Parseval-Plancherel identity.

[2] This, of course, is a rough approximation. In particular, it neglects all possible phase errors at the fiber injection which may be due to an optically abberated pupil.

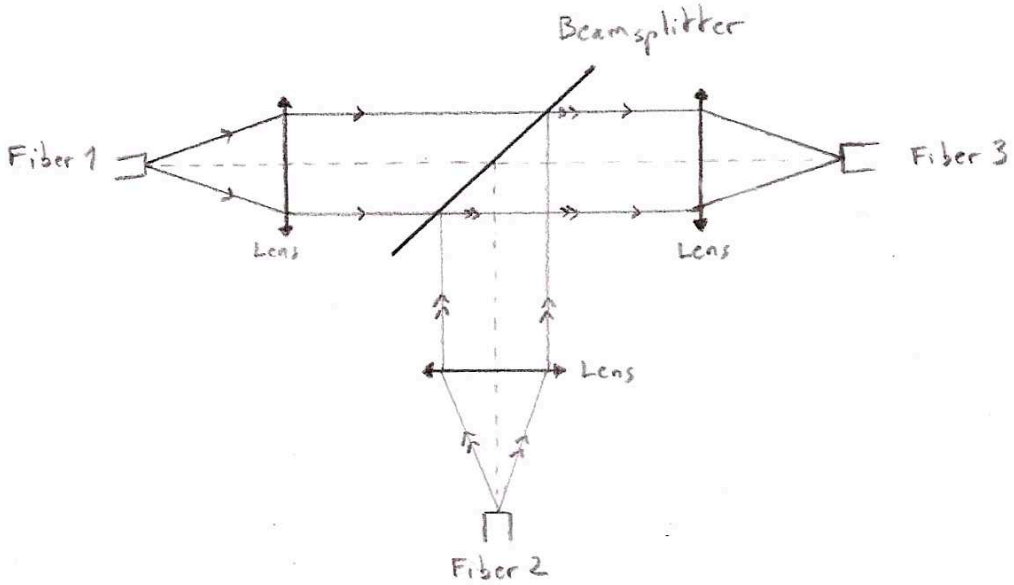


Figure 11.7: Principle of a bulk optics beam combiner for optical fibers. In this setup, the two input fibers 1 and 2 are located in the focal plane of two lenses, to create two collimated beams. The two beams are combined by the beamsplitter, are re-imaged onto the output fiber 3 by a third lens.

waveguides so close to one another that the evanescent field propagating around one fiber actually reaches the other fiber, and starts propagating into it. This couples the fields of the two waveguides together, and when the waveguides are separated again (see Figure 11.8), the fields at the two outputs are :

$$A_{\text{output 1}} = \frac{A_{\text{input 1}} + A_{\text{input 2}}}{\sqrt{2}} \tag{11.60}$$

$$A_{\text{output 2}} = \frac{A_{\text{input 1}} + A_{\text{input 2}} \times e^{i\pi}}{\sqrt{2}} \tag{11.61}$$

The presence of a phase offset term at one of the two outputs satisfies the conservation of energy:

$$\begin{aligned} I_{\text{output 1}} + I_{\text{output 2}} &= \frac{1}{2}\left\|A_{\text{input 1}} + A_{\text{input 2}}\right\|^2 + \frac{1}{2}\left\|A_{\text{input 1}} + A_{\text{input 2}} e^{i\pi}\right\|^2 \\ &= \frac{I_{\text{input 1}} + I_{\text{input 2}} + 2A_{\text{input 1}}A^*_{\text{input 2}}}{2} + \frac{I_{\text{input 1}} + I_{\text{input 2}} - 2A_{\text{input 1}}A^*_{\text{input 2}}}{2} \\ &= I_{\text{input 1}} + I_{\text{input 2}} \end{aligned}$$

Using fibered beamsplitters downstream of the beam-combiner, and adding a $\pi/2$ phase offset gives a complete integrated-optics ABCD beam-combiner (see Figure 11.9). This is exactly the approach taken in the GRAVITY instrument. The final beam-combiner is extremely small, and can easily be installed in a cryostat, for optimal performance.

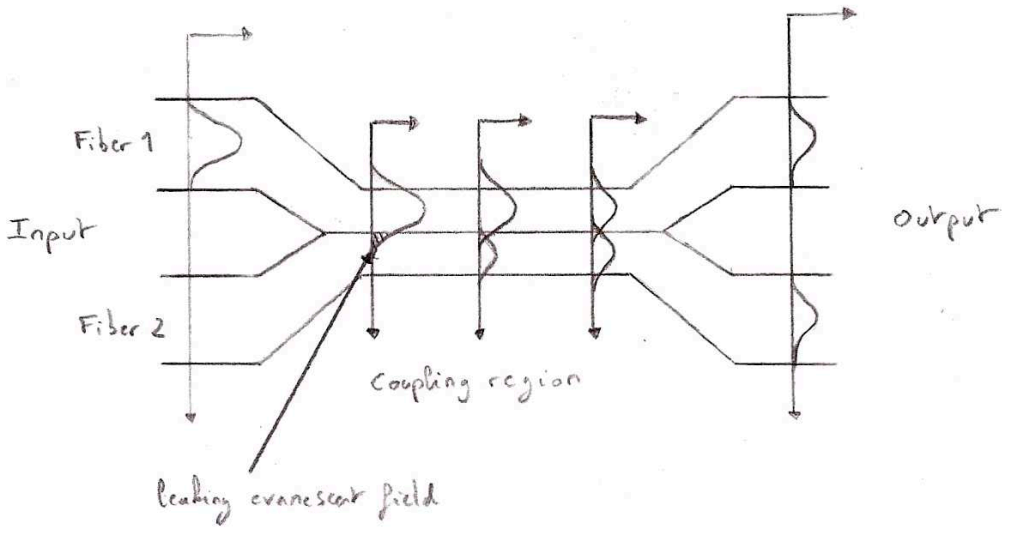


Figure 11.8: Principle of a fibered beam-combiner. Contrary to the bulk optics combiner shown in Figure 11.7, this setup does not use any lens or beamsplitter. The two input fibers are brought very close to each other, so that the evanescent field in the cladding of one fiber can "leak" into the other fiber. The small inset graphs show the evolution of the intensity distribution in the two fibers, when light is injected only in fiber 1. Over the coupling area, the leaking evanescent field slowly couples to the second fiber. Conservation of energy tells us that the field in the two output fibers are necessarily in phase opposition (phase difference of π).

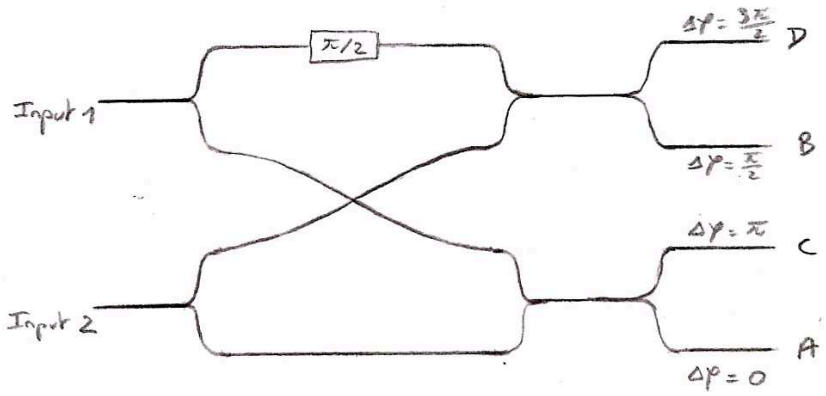


Figure 11.9: Integrated optics ABCD beam-combiner. This setup is made using two fibered beamsplitters to split each input in two parts, which are then recombined in a different order by two beamsplitters. A $\pi/2$ phase shift is added to one of the channel to obtain the correct 4 different phases at the output. In practice, controlling the exact phase of each output is difficult, and the beam combination must be calibrated.

11.3.3 Fringe tracking, metrology, and atmospheric phase noise rejection

Consider again the never-ending case of two telescopes observing a point-source in the sky, located at (α, β). The intensity retrieved at the output A of the ABCD beam-combiner is:

$$I_A(b_k, \lambda) = 2I_0 + 2I_0 \cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right)\right)$$

This equation assumes the ideal case of a perfectly flat incident wavefront. But the presence of the atmosphere makes this assumption invalid to a large extent. In practice, turbulence and variations of the air refraction index makes the incident wavefront unpredictable. The direct consequence is an additional atmospheric phase term at each telescope, which changes the complex amplitude of the electric field sampled by the two telescopes:

$$A_{T_{k_1}} = A_{T_{k_1}\,\text{ideal}} \times e^{i\phi_{\text{atm},T_{k_1}}} \tag{11.62}$$

$$A_{T_{k_2}} = A_{T_{k_2}\,\text{ideal}} \times e^{i\phi_{\text{atm},T_{k_2}}} \tag{11.63}$$

In this non-ideal case, the intensity I_A becomes:

$$I_A(b_k, \lambda) = I_{A,\text{ideal}}(b_k, \lambda) \times e^{i\left(\phi_{\text{atm},T_{k_2}} - \phi_{\text{atm},T_{k_1}}\right)} \tag{11.64}$$

Similarly, every output of the ABCD beam-combiner would be phase-shifted by the same quantity. Thus, the complex visibility measured by the ABCD combiner in the presence of atmospheric phase errors is the true astrophysical visibility of the observed target, affected by a phase error term due to the atmosphere:

$$V_{\text{mes}}(b_k, \lambda) = V(b_k, \lambda)_{\text{astro}} \times e^{i\left(\phi_{\text{atm},T_{k_2}} - \phi_{\text{atm},T_{k_1}}\right)} \tag{11.65}$$

In this equation, the subscript "astro" means that the complex visibility is related to the astrophysical source intensity distribution by the Fourier transform relation given in Equation 11.37.

The introduction of this atmospheric phase error has two major instrumental consequences. The first is obvious with the above equation: the measured visibility is not equal to the true astrophysical visibility anymore. The second effect is somewhat hidden by the fact that the interferometer only records the interferometric fringes at 4 different (phase) points through the ABCD combiner. Physically, the direct consequence of a phase error on the electric fields sampled by the two telescopes is to add a phase-offset to the fringes. And since the typical atmospheric coherence time t_0 is short (typically $t_0 \sim 1$ ms), if the instrument integrates over a certain duration $T_{\text{DIT}} \gg t_0$, the fringes will be blurred. In mathematical language, we can write:

$$I_A(b_k, t) = 2I_0 \left[1 + \int_t^{t+T_{\text{DIT}}} \cos\left(\frac{2\pi}{\lambda}\left(\alpha U_k + \beta V_k\right) + \Delta\phi_{atm}(t)\right)\mathrm{d}t\right] \tag{11.66}$$

with $\Delta\phi_{\text{atm}}$ a fast and randomly varying function of time t. The end result will be that the coherent part of the flux (the integral) will most probably vanish, and the final complex visibility measured by the ABCD beam combiner will be zero.

So we are left with two problems: limiting the coherence loss, and retrieving the astrophysical visibility from the corrupted measured visibilities.

11.3.3.1 Limiting the coherence loss: the GRAVITY fringe tracker

One way to limit the coherence loss due to the atmosphere is to limit the exposure time. If $T_{\rm dit} \leq t_0$, then there is no coherence loss. The measured visibility is still corrupted by an atmospheric phase error, but the fringes are not "blurred", and the coherent part of the flux can at least be measured. The obvious limit to this approach is that the detector noise will severely limit the capability of the instrument to observe faint targets. Even with a multi-meter class telescope, observing anything fainter that a magnitude 10 or 11 will become a challenge, even with excellent conditions.

The other option is to fight-back, and artificially introduce a phase offset between the two arms of the interferometer to compensate for the atmospheric phase noise. Such a system will monitor the interferometric fringes, and keep them still. The technique is referred to as *fringe tracking*. Of course, to be able to track the fringes, the fringe-tracker needs to measure the phase error $\Delta\phi_{\rm atm}$ at least every t_0. The fringe-tracker itself needs to observe a bright source. Fortunately, there is no need for the fringe-tracker to observe the same source as the science detector. The fringe-tracker of GRAVITY, described in Lacour et al. [2019], uses a second fiber, and its own ABCD combiner. The science and fringe-tracker fibers can be placed on two different targets in the focal plane of the instrument, making possible to use long exposures on a faint target for the science detector, and short exposures on a close-by bright target with the fringe-tracker. This may sound familiar to anyone who has ever worked with adaptive optics, as there are a lot of similarities between the two systems. One of the similarities is that the angular distance between the reference star used by the fringe-tracker and the science target is limited by the isoplanatic angle. Another similarity is that because the fringe-tracking and the science channels operate using different optical trains, the system becomes sensitive to non-common path errors.

11.3.3.2 Non-common path errors

To clarify the problem, let us look at the full optical path in each arm of the interferometer, and list all possible sources of phase errors. The light coming from each telescope first goes through the VLTI delay lines, which are used to add large optical delays into the system (up to 60 meters). These are in particular useful to offload the fringe-tracker actuators (see Lacour et al. [2019]). Then, the light goes into the proper GRAVITY instrument. It first encounters the fiber injection unit, then the OPD control unit (the main actuator used by the fringe-tracker), and finally enters the beam-combiner (see Figure 11.10). Due to thermal dilation or refraction index variations, each of these subsystems can be a source of phase error. For telescope T_{k_1}, the phase of the field at the entrance of the science beam-combiner can be written:

$$\phi_{\rm sci,T_{k_1}} = \phi_{{\rm atm},T_{k_1}} + \phi_{{\rm DL},T_{k_1}} + \phi_{{\rm prop,sci},T_{k_1}} + \phi_{{\rm OPD},T_{k_1}} \tag{11.67}$$

where $\phi_{\rm prop,sci}$ stands for all phase errors acquired when propagating in the science channel (including fiber injection), excluding the additional phase introduced by the OPD unit (which is common to the fringe-tracker channel, and thus has its own term, $\phi_{\rm OPD}$).
For the fringe-tracker (ft) channel:

$$\phi_{\rm ft,T_{k_1}} = \phi_{{\rm atm},T_{k_1}} + \phi_{{\rm DL},T_{k_1}} + \phi_{{\rm prop,ft},T_{k_1}} + \phi_{{\rm OPD},T_{k_1}} \tag{11.68}$$

And at telescope T_{k_2}:

$$\phi_{\rm sci,T_{k_2}} = \phi_{{\rm atm},T_{k_2}} + \phi_{{\rm DL},T_{k_2}} + \phi_{{\rm prop,sci},T_{k_2}} + \phi_{{\rm OPD},T_{k_2}} \tag{11.69}$$

$$\phi_{\rm ft,T_{k_2}} = \phi_{{\rm atm},T_{k_2}} + \phi_{{\rm DL},T_{k_2}} + \phi_{{\rm prop,ft},T_{k_2}} + \phi_{{\rm OPD},T_{k_2}} \tag{11.70}$$

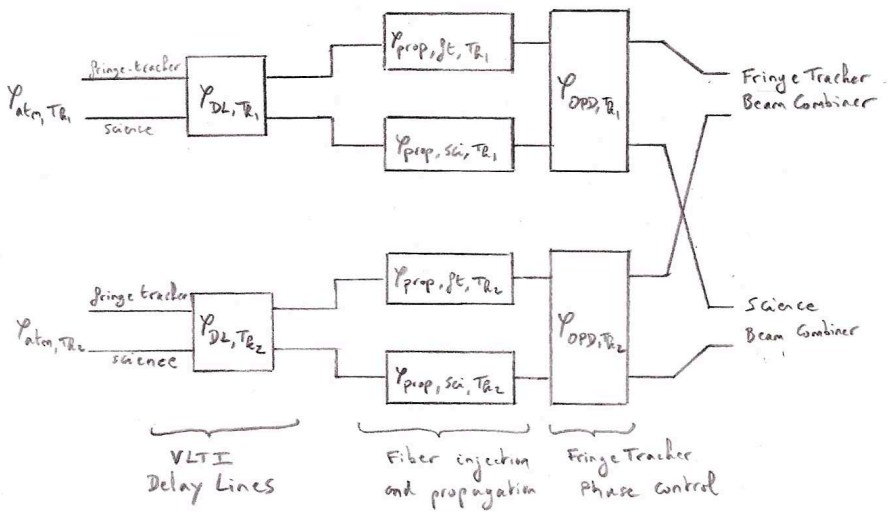


Figure 11.10: Simplified illustration of the optical path in the GRAVITY instrument, showing the different phase aberrations acquired by the beams coming from two telescopes.

In all these equations, the terms $\phi_{\rm DL}$ and $\phi_{\rm OPD}$ are actually similar, and serve the same purpose. We can group them in a $\phi_{\rm PC}$ (Phase Control) term. Then, the phase differences measured by each combiner write:

$$\Delta\Phi_{\rm ft} = \phi_{{\rm ft},T_{k_2}} - \phi_{{\rm ft},T_{k_1}} = \Delta\phi_{\rm atm} + \Delta\phi_{\rm PC} + \Delta\phi_{\rm prop,ft} \tag{11.71}$$

$$\Delta\Phi_{\rm sci} = \phi_{{\rm sci},T_{k_2}} - \phi_{{\rm sci},T_{k_1}} = \Delta\phi_{\rm atm} + \Delta\phi_{\rm PC} + \Delta\phi_{\rm prop,sci} \tag{11.72}$$

When the fringe-tracker is active, it keeps the phase difference at zero:

$$\Delta\phi_{\rm ft} = 0 \tag{11.73}$$

Thus, we have:

$$\Delta\phi_{\rm atm} + \Delta\phi_{\rm PC} = -\Delta\phi_{\rm prop,ft} \tag{11.74}$$

In which case the science channel phase difference can now be written:

$$\Delta\Phi_{\rm sci} = \Delta\phi_{\rm prop,sci} - \Delta\phi_{\rm prop,ft} \tag{11.75}$$

This remaining phase difference is a pure non-common path error.

Written in the form of Equation 11.75, this non-common path error may look extremely difficult to remove. In order to suppress it, one would need to measure the difference of $\Delta\phi$ between the two interferometric channels. This would require again another interferometric stage, with its own beam combiner, for each baseline. All of this within the cryogenic stage of the instrument. However, if we split the $\Delta\phi$, and reorganize the terms, we have the equivalent form:

$$\Delta\phi_{\rm sci} = \left(\phi_{{\rm prop,sci},T_{k_2}} - \phi_{{\rm prop,ft},T_{k_2}}\right) - \left(\phi_{{\rm prop,sci},T_{k_1}} - \phi_{{\rm prop,ft},T_{k_1}}\right) \tag{11.76}$$

In this form, the non-common path error appears as the difference of two telescope based errors. This means that the non-common path error can either be measured and corrected by a dedicated telescope-based metrology system, or removed using closure phases, to retrieve the astrophysical information.

11.3.3.3 Retrieving the astrophysical visibility

The remaining phase error for each baseline, as given in Equation 11.76, is the difference of two telescope based errors. For baseline b_k linking telescopes T_{k_1} to T_{k_2}, the complex visibility measured at the output of the ABCD beam-combiner is:

$$V(b_k,\lambda)_{\text{mes}} = V(b_k,\lambda)_{\text{astro}} \times e^{-i\left(\phi_{T_{k_2}}-\phi_{T_{k_1}}\right)} \tag{11.77}$$

with ϕ_{T_k} the non-common path phase error at telescope T_k: $\phi_{T_k} = \phi_{\text{prop,sci},T_k} - \phi_{\text{prop,ft},T_k}$.

This is actually very similar to the initial problem of atmospheric phase noise, which is also a telescope-based phase error. The big difference, though, is that the stability of this noise now depends on the stability of the instrumental setup, not of the atmosphere. This makes possible to take long exposures. But the problem of retrieving the true astrophysical visibility remains. Two routes can be followed: visibility amplitude/closure phase combination, or phase-referencing.

Visibility amplitude and closure phases The first possibility to retrieve astrophysical information from a measured visibility as given by Equation 11.77 is to find a quantity insensitive to the phase noise. An obvious example is the squared visibility modulus:

$$\|V(b_k,\lambda)_{\text{mes}}\|^2 = \left\|V(b_k,\lambda)_{\text{astro}} e^{-i\left(\phi_{T_{k_2}}-\phi_{T_{k_1}}\right)}\right\|^2 = \|V(b_k,\lambda)_{\text{astro}}\|^2 \tag{11.78}$$

A less obvious example is the closure phase. The concept can be traced back to the work of Jennison [1958] in radio-astronomy, and the "interferometric triangles", made of three telescopes. Consider three telescopes T_1, T_2, and T_3, and the measured visibilities on the three baselines linking the different telescopes together:

$$V(T_1T_2,\lambda)_{\text{mes}} = V(T_1T_2,\lambda)_{\text{astro}} \times e^{-i\left(\phi_{T_2}-\phi_{T_1}\right)} \tag{11.79}$$

$$V(T_1T_3,\lambda)_{\text{mes}} = V(T_1T_3,\lambda)_{\text{astro}} \times e^{-i\left(\phi_{T_3}-\phi_{T_1}\right)} \tag{11.80}$$

$$V(T_2T_3,\lambda)_{\text{mes}} = V(T_2T_3,\lambda)_{\text{astro}} \times e^{-i\left(\phi_{T_3}-\phi_{T_2}\right)} \tag{11.81}$$

Forming the quantity $B(T_1T_2T_3,\lambda) = V(T_1T_2,\lambda) \times V(T_2,T_3,\lambda) \times V(T_1,T_3,\lambda)^*$, where the * symbol denotes the complex conjugate, we have:

$$\begin{aligned} B(T_1T_2T_3,\lambda) =& V(T_1T_2,\lambda)_{\text{astro}} \times e^{-i\left(\phi_{T_2}-\phi_{T_1}\right)} \times V(T_2T_3,\lambda)_{\text{astro}} \times e^{-i\left(\phi_{T_3}-\phi_{T_2}\right)} \\ & \times V(T_1T_3,\lambda)^*_{\text{astro}} \times e^{+i\left(\phi_{T_3}-\phi_{T_1}\right)} \end{aligned} \tag{11.82}$$

which simplifies into:

$$B(T_1T_2T_3,\lambda) = V(T_1T_2,\lambda)_{\text{astro}} V(T_2T_3,\lambda)_{\text{astro}} V(T_1T_3,\lambda)^*_{\text{astro}} \times e^{-i\left(\phi_{T_2}-\phi_{T_1}+\phi_{T_3}-\phi_{T_2}-(\phi_{T_3}-\phi_{T_1})\right)} \tag{11.83}$$

$$= V(T_1T_2,\lambda)_{\text{astro}} V(T_2T_3,\lambda)_{\text{astro}} V(T_1T_3,\lambda)^*_{\text{astro}} \tag{11.84}$$

The quantity $B(T_1T_2T_3,\lambda)$ is called *closure visibility*, or *bispectrum*, and is insensitive to telescope-based phase errors. But as the visibility modulus can readily be extracted from the measured visibility, the closure visibility is mostly used as a way to extract the *closure phase* C_ϕ:

$$\begin{aligned} C_\phi(T_1T_2T_3,\lambda) &= \arg\left(B(T_1,T_2,T_3,\lambda)\right) \\ &= \arg\left(V(T_1T_2,\lambda)_{\text{astro}}\right) + \arg\left(V(T_2T_3,\lambda)_{\text{astro}}\right) - \arg\left(V(T_1T_3,\lambda)_{\text{astro}}\right) \end{aligned}$$

The closure phase holds information on the phase of the astrophysical complex visibility (but not *all* information), and is insensitive to telescope-based errors.

Using visibility amplitudes and/or closure phases, it is possible to extract useful information on the observed astrophysical object, even in the presence of telescope-based phase errors. When no fringe-tracker is available, this can be used to get rid of atmospheric phase noise (with short exposures to avoid coherence loss), and when a fringe tracker is present, this can get rid of non-common path errors.

Phase referencing with the metrology system Another approach to the problem of non-common path errors is to try to measure and correct them. In GRAVITY, the non-common path error terms of Equation 11.76 are measured by a dedicated metrology system. This metrology system uses a 1.908 μm laser, which is injected at the beam-combiner level (Figure 11.6), and counter-propagates in the entire optical train (beam-combiner, OPD control system, fiber injection, VLTI delay line, and up to the telescope). A system of diodes mounted on the telescope spider then measures the fringes formed in the pupil plane by the metrology laser light exiting the science and fringe-tracker fibers. This gives a direct measurement of the term $\phi_{\text{prop,sci}} - \phi_{\text{prop,ft}}$ at each telescope, which can be used to correct for this non-common path error term.

But of-course, injecting the laser light into the optical train creates another source of phase errors... This may seem endless, but the injection module is not perfect, and the laser light does not necessarily enter the science and fringe-tracking optical trains with the same phase. The phase difference actually measured by the metrology system at telescope T_k can be written:

$$\Delta\phi_{\text{met},T_k} = (\phi_{\text{prop,sci},T_k} - \phi_{\text{prop,ft},T_k}) + (\phi_{\text{met,sci},T_k} - \phi_{\text{met,ft},T_k}) \tag{11.85}$$

And so using the metrology measurements to correct the phase of the visibility, we get the "metrology corrected" visibility (for baseline b_k linking T_{k_1} to T_{k_2}):

$$V(b_k,\lambda)_{\text{metcorr}} = V(b_k,\lambda)_{\text{astro}} \times e^{i\left(\phi_{\text{met,sci},T_{k_2}} - \phi_{\text{met,ft},T_{k_2}} + \phi_{\text{met,ft},T_{k_1}} - \phi_{\text{met,sci},T_{k_1}}\right)} \tag{11.86}$$

Fortunately, this error does not dependent on anything external to the metrology injection unit. In particular, it does not depend on the fiber pointing, is insensitive to variations of optical fiber lengths in the main optical-train, etc. This error is expected to be very stable, and does not hamper observations in a *phase referencing* mode. The concept is to switch back and forth between two astrophysical targets (one of which could be an interferometric calibrator). The visibility obtained on each object, after the metrology correction, would be:

$$V_{\text{obj}_1,\text{metcorr}} = V_{\text{obj}_1,\text{astro}} \times e^{i\left(\phi_{\text{met,sci},T_{k_2}} - \phi_{\text{met,ft},T_{k_2}} + \phi_{\text{met,ft},T_{k_1}} - \phi_{\text{met,sci},T_{k_1}}\right)} \tag{11.87}$$

$$V_{\text{obj}_2,\text{metcorr}} = V_{\text{obj2},\text{astro}} \times e^{i\left(\phi_{\text{met,sci},T_{k_2}} - \phi_{\text{met,ft},T_{k_2}} + \phi_{\text{met,ft},T_{k_1}} - \phi_{\text{met,sci},T_{k_1}}\right)} \tag{11.88}$$

The idea behind "referencing the phase object 2 to object 1" is then simply to calculate the phase difference between the two, in which the metrology injection error term disappears:

$$\arg\left(V_{\text{obj}_2,\text{metcorr}}\right) - \arg\left(V_{\text{obj}_1,\text{metcorr}}\right) = \arg\left(V_{\text{obj}_2,\text{astro}}\right) - \arg\left(V_{\text{obj}_1,\text{astro}}\right) \tag{11.89}$$

This is especially useful, for example, to measure the differential astrometry (i.e. the separation) between two components of a resolved binary system (by moving the science fiber back and forth between the two components).

11.4 Single-field and dual-field observations

So far, I have described the general working principles of the GRAVITY instrument, but I have not explained its main difference with other interferometers, which is also what makes it capable of observing exoplanets: dual-field. In this section, I describe the difference between the single-field and dual-field modes. To do so, I will focus on the example of a simple binary system observed by the instrument.

11.4.1 Single-field ("classical interferometry")

To understand what dual-field observations are, it is necessary to understand how a classical (single-field) interferometric observation works. When observing a binary system in single-field, the science fiber is centered on the photocenter of the binary, and both components enter the fiber (see Figure 11.11a). This means that the separation of the target binary needs to be smaller than the fiber field-of-view, which corresponds to the instrument PSF (~ 60 mas when observing with the UTs, or ~ 300 mas when observing with the ATs). Denoting (α_A, β_A) the coordinates of the component A, and $I_A(\lambda)$ its intensity, the complex amplitude of the field sampled by the fiber at telescopes T_{k_1} and T_{k_2} separated by a baseline b_k is given by:

$$A_{A,T_{k_1}}(\lambda) = \sqrt{I_A(\lambda)} \times e^{-i\phi_0}$$
$$A_{A,T_{k_2}}(\lambda) = \sqrt{I_A(\lambda)} \times e^{-i\left[\phi_0 + \frac{2\pi}{\lambda}(\alpha_A U_k + \beta_A V_k)\right]}$$

where ϕ_0 is an arbitrary origin of phase. A similar formula holds for component B. The two sources are incoherent, and the complex visibility measured at the output beam combiner is the sum of the complex visibility of A and B:

$$V(b_k, \lambda) = I_A(\lambda) e^{-i\frac{2\pi}{\lambda}(\alpha_A U_k + \beta_A V_k)} + I_B(\lambda) e^{-i\frac{2\pi}{\lambda}(\alpha_B U_k + \beta_B V_k)}$$

Denoting $\Delta\alpha$ and $\Delta\beta$ the separation of the binary:

$$V(b_k, \lambda) = \left[I_A(\lambda) + I_B(\lambda) e^{-i\frac{2\pi}{\lambda}(\Delta\alpha U_k + \Delta\beta V_k)}\right] e^{-i\frac{2\pi}{\lambda}(\alpha_A U_k + \beta_A V_k)}$$

From this, it directly follows that the visibility modulus and the closure phase, which are the two interferometric observables, only depend on the separation vector. In both observables, the last exponential term vanishes. For example, the squared visibility modulus is given by:

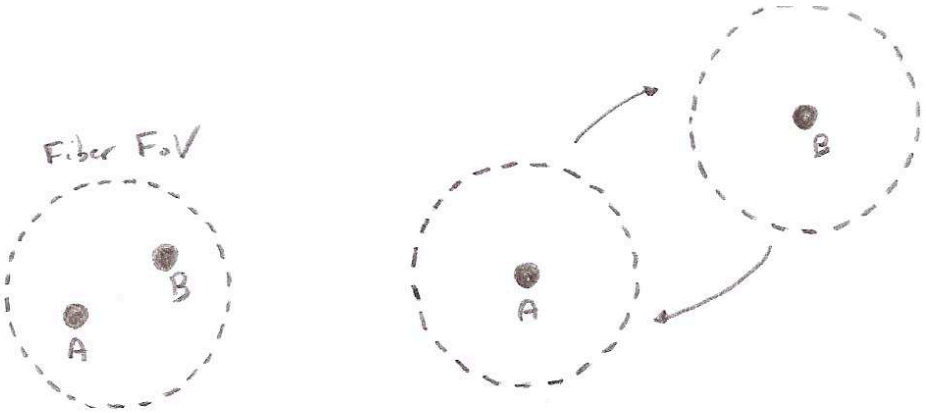
$$\|V(b_k, \lambda)\|^2 = \left\| I_A(\lambda) + I_B(\lambda) e^{-i\frac{2\pi}{\lambda}(\Delta\alpha U_k + \Delta\beta V_k)} \right\|^2$$

It is possible to extract the separation and the contrast ratio just from this visibility modulus.

11.4.2 Dual-field

But what if the two sources are too far away, and cannot enter the fiber simultaneously?

In this case it is necessary to observe the two components separately, moving the fiber back and forth (see Figure 11.11b). Without a constant phase origin between the two sets of observations, it would be impossible to reconstruct any useful information from these separate



(a) Single-field (b) Dual-field

Figure 11.11: Depiction of the difference between a single-field and a dual-field observations of a binary system (components A ad B). In single-field, the Field-of-View (FoV) of the science fiber is centered on the photometric center, and encircles both components. In dual-field, the science fiber can only see one component at a time, and it is necessary to alternate between observations of the two components.

observations. But if the fringe-tracker keeps tracking on component A, for example, then it sets a common phase origin, and the visibilities can be individually measured:

$$\begin{aligned} V_A(b_k,\lambda) &= I_A(\lambda)e^{-i\phi_0} \\ V_B(b_k,\lambda) &= I_B(\lambda)e^{-i\left[\phi_0+\frac{2\pi}{\lambda}(\Delta\alpha U_k+\Delta\beta V_k)\right]} \end{aligned}$$

From this, it is possible to numerically reconstruct the total visibility $V = V_A + V_B$. But in practice, it is often more convenient to use the *referenced visibilities*:

$$\begin{aligned} V_{\mathrm{refA}}(b_k,\lambda) &= V_A(b_k,\lambda)e^{-i\arg(V_A(b_k,\lambda))} \\ &= I_A(\lambda) \\ V_{\mathrm{refB}}(b_k,\lambda) &= V_B(b_k,\lambda)e^{-i\arg(V_A(b_k,\lambda))} \\ &= I_B(\lambda)e^{-i\frac{2\pi}{\lambda}(\Delta\alpha U_k+\Delta\beta V_k)} \end{aligned}$$

which easily give access to the contrast spectrum and to the separation.

Chapter 12

Dual-field direct detection of exoplanets with VLTI/GRAVITY

Content of this chapter

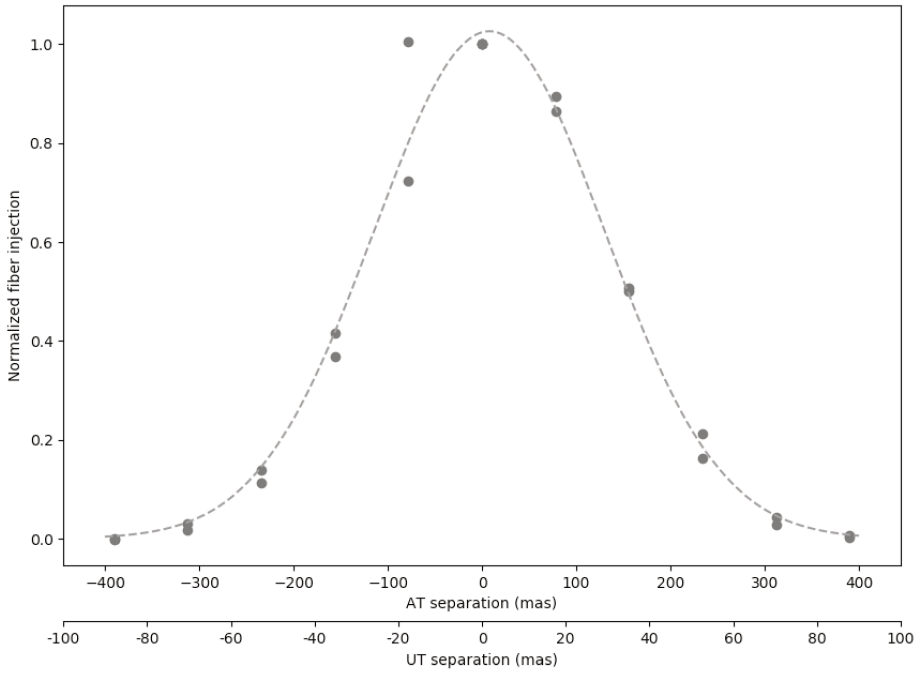


Figure 12.1: Fiber injection factor on the ATs (normalized to unity) as a function of the angular separation of the fiber from the center of the PSF. The dots represent real data acquired in February 2019, as part of a technical run on the ATs with the NAOMI adaptive optics. The line is the best gaussian fit. The second x-axis (UT separation) is obtained by applying a scaling factor of 4.44 to the AT separation values, neglecting any potential difference of performance between the UT adaptive optics and AT adaptive optics.

12.1 Observing strategy

12.1.1 High-contrast via dual-field observation

Long-baseline interferometry is not known for its high-dynamics. If it was not for its "dual-field" capability, and the use of single-mode fibers, the GRAVITY instrument would be no exception to this rule. In single-field observations, the typical precision on the measurement of the visibility modulus for a bright target (magnitude 6) is a few percents. This is far from the required 10^{-2} % or 10^{-3} % required for the direct detection of giant exoplanets.

But with GRAVITY, there is this possibility of observing in dual-field mode, in which the science fiber can be moved from one target to another. Combined to the use of an adaptive optics system, which delivers moderate Strehl ratio, the single mode-fiber effectively acts as a strong spatial filter. The injection factor rapidly drops when the fiber is offseted from the center of the telescope's PSF. In Figure 12.1, I give the evolution of the injection factor as a function of the fiber-PSF center separation.

The strategy to observe exoplanets with GRAVITY is to use the on-axis, dual-field mode. The fringe-tracker fiber is kept on the central star, while the science fiber alternates between the star and the planet positions, changing at each exposure.

When the science fiber is centered on the planet, the stellar flux is largely reduced by the combination of AO and single-mode filtering. From Figure 12.1, we can see that the typical reduction when observing at ~ 100 mas from the star is a factor of > 100. In itself, this is not enough to actually be able to see the planet. The total flux propagating into the fiber, and

reaching the science ABCD detectors, is still largely dominated by stellar light. However, this stellar light is only partly coherent, which means that only a fraction of this light will affect the complex visibility after the ABCD combination. Typically, for the β Pic b observations reported in Chapter 13, the coherent stellar light is a factor ~ 100 greater than the planet coherent light.

Fortunately, there is a final difference between the two sources: the on-sky position. The star, which is centered on the optical axis, produces a complex visibility with zero OPD. Its phase does not vary with wavelength λ. The planet, on the other hand, produces a complex visibility at a given OPD, which is set by its position: $\mathrm{OPD} = \Delta\mathrm{RA} \times U + \Delta\mathrm{DEC} \times V$ where (U, V) are the baseline coordinates (see Section 11.2). The phase is proportional to $1/\lambda$. At first-order, denoting λ_0 the central wavelength of the GRAVITY K-band observations, the star and planet visibilities have the form:

$$V_\star(\lambda) = G(\lambda) S_\star(\lambda) e^{i\phi} \tag{12.1}$$

$$V_{\mathrm{planet}}(\lambda) = G(\lambda) S_{\mathrm{planet}}(\lambda) e^{i\left(\phi + \frac{4\pi}{\lambda_0}\mathrm{OPD}\right)} e^{-i\frac{2\pi}{{\lambda_0}^2}\mathrm{OPD}\times\lambda} \tag{12.2}$$

Where S is the spectrum, G the instrumental spectral response, and ϕ an arbitrary and unknown phase zero-point
The measured visibility is the incoherent sum of the two:

$$V_{\mathrm{measured}}(\lambda) = V_\star(\lambda) + V_{\mathrm{planet}}(\lambda) \tag{12.3}$$

We are left with the problem of recovering $V_{\mathrm{planet}}(\lambda)$ from this sum, dominated (in power) by $V_\star$. But a close look at the above equations reveals that this is actually similar to a well-known problem in radio-communication: demodulating and amplitude-modulated signal. Considered as a function of λ, the planet visibility is actually an amplitude-modulated version of the spectrum S (and instrumental response G), with the carrier wave $\exp\left(-i\frac{2\pi}{{\lambda_0}^2}\mathrm{OPD} \times \lambda\right)$. The star visibility is directly given by its spectrum (which is equivalent to say that this is the spectrum amplitude-modulated with a carrier of frequency 0). What this all means is that in Fourier space, the two signals are not centered around the same frequency.

This gives an intuitive explanation of how long-baseline interferometry can leverage the small difference of position between the planet and its host star to recover the planet signal. In practice, though, a Fourier-based approach to the data reduction and analysis is not satisfactory. There is some overlap between the two Fourier components (the spectra are too broad compared to the carrier frequency), and the phase is not truly proportional to λ (it varies as $1/\lambda$). Other tools are needed, which are described in Section 12.5.

12.1.2 The central star as a calibrator

When observing a star-planet binary system in dual-field, the star itself provides a natural calibrator for the observations. The stellar visibility is used to reference the phase of the planet complex visibility (after the metrology correction, see Section 11.3.3.3 for details on phase referencing), but also as a spectral reference when extracting the planet spectrum. Extracting a *contrast* spectrum, rather than directly the planet spectrum, makes the process literally insensitive to any instrumental spectral response, much more resilient to atmospheric spectral distortions, and less affected by telluric lines (the contrast spectrum would also be insensitive to the atmosphere if the planet and the star were observed simultaneously, which is never the case).

12.1.3 Pipeline reduction and astroreduced file products

The data acquired with VLTI/GRAVITY can be reduced using the official ESO pipeline [Lapeyrere et al., 2014]. The pipeline takes care of all the necessary steps to retrieve the complex visibilities for each baseline, and the flux for each telescope, from the outputs of the different ABCD beam combiners. It takes care of background subtraction, flat-field correction, etc. It also manages the metrology correction of the phase. Finally, it combines all DITs within an exposure into a single "scivis" data file.

This last step (averaging the different DITs within an exposure file) is a problem for the exoplanet observations. As will be explained in Section 12.5, there is a degeneracy in the data reduction process when trying to extract the spectrum, due to the necessary star subtraction. This degeneracy is reduced by the sky rotation, which changes the shape of the planet visibility over time. Keeping all the points available separated helps to reduce the risk of bias in the extracted planet spectrum. The typical exposure time for observing an exoplanet with GRAVITY is $\sim$ 10 to 60 s. A typical exposure will contain about 10 to 30 DITs, thus representing about 5 to 10 min of observation. One hour of observing time will contain about 5 to 10 planet exposures (plus exposures obtained on the star, which are shorter due to the difference in brightness). What this means is that by keeping the individual DITs separated, the number of points in time available to remove the degeneracy can by increased from ~ 5 to ~ 50.

Thus, the entire data reduction procedure described in the next sections is based on the intermediate "astrored" file products, and not on the final "scivis" files produced by the GRAVITY pipeline. The "astrored" files can be obtained by setting a dedicated option when running the pipeline (`--astro-file`, see the GRAVITY pipeline User Manual for details).

When working with the "astrored" files, which were intended to be used for testing purposes by the GRAVITY collaboration, several additional steps not performed by the pipeline are necessary:

- The visibilities need to be corrected for the metrology.
- The visibilities need to be corrected for the fringe-tracker zero-point. To reduce the possibility of bias due to the detectors response, the fringe-tracker does not constantly keep the science fringes centered on the ABCD detectors (i.e. with OPD = 0 on the A detector), but alternates between different setpoints (so that the modulation seen in wavelength due to the λ-dependence of the OPD-induced phase moves on the detectors). This needs to be taken into account.
- The phase-referencing of the planet visibilities with the stellar visibilities needs to be done manually (this would also be the case with the fully reduced "scivis" data).

As a side note, I also want to point out that in the case of an exoplanet observations, there is no easy way to "flux-calibrate" the visibilities, i.e. to calculate the quantity:

$$V(b_k, \lambda)_{\rm cal} = \frac{V(b_k, \lambda)}{\sqrt{f_{T_{k_1}}(\lambda) \times f_{T_{k_2}}(\lambda)}} \tag{12.4}$$

where b_k is the baseline linking telescopes T_{k_1} to T_{k_2}, and $f_{T_{k_1}}, f_{T_{k_2}}$ are the flux observed by the two telescopes. The reason why this calibration is particularly difficult is that the flux at each telescope is dominated by the stellar flux, not by the planet light. The calibration would divide the planet visibility by (mostly) the stellar flux, which would inject a lot of photon noise. Plus, as the quantity actually extracted by the data reduction method described hereafter is the contrast spectrum of the planet with respect to the star, there is no actual need for such a calibration, if the mean injection is stable when switching between the planet and the star.

12.2 Some notations and important results

12.2.1 General notations

Before moving to the GRAVITY exoplanet data reduction, I first need to introduce some important notation rules that I will use in the rest of this Chapter.

- Complex numbers are underlined, real numbers are not. For example, the complex visibility is denoted $\underline{V}$, whereas the visibility modulus, which is real, is simply V.
- The real part of a complex number $\underline{X}$ is denoted $\mathrm{Re}(\underline{X})$, and its imaginary part $\mathrm{Im}(\underline{X})$.
- The complex conjugate of a quantity $\underline{X}$ is denoted $\underline{X}^*$.
- Physical quantities related to the planet are usually denoted with the subscript P, and quantities related to the star with the subscript $\star$ (e. g. $\underline{V}_\mathrm{P}$ and $\underline{V}_\star$).

Most of the GRAVITY data (i.e. spectra, complex visibilities or visibility modulus, etc.) are wavelength dependent quantities, for which the wavelength axis is discrete: $(\lambda_1, \dots, \lambda_{n_\lambda})$, n_λ being the number of wavelength points. To any arbitrary quantity $F(\lambda)$ that depends on λ, I associate a vector of length n_λ, which will be denoted in bold (i. e. $\mathbf{F}$), and defined as:

$$\mathbf{F} = \begin{pmatrix} F_{\lambda_1} \\ \vdots \\ F_{\lambda_{n_\lambda}} \end{pmatrix} = \begin{pmatrix} F(\lambda_1) \\ \vdots \\ F(\lambda_{n_\lambda}) \end{pmatrix}$$

For these matrices and vectors, I add the following notation rules:

- The transpose of a matrix $\mathbf{M}$ is denoted $\mathbf{M}^T$.
- The complex transpose of a matrix $\underline{\mathbf{M}}$ is denoted $\underline{\mathbf{M}}^\dagger$.
- The identity matrix is denoted $\mathbf{I}$.

From there, I also introduce the scalar product $\langle \, . \, , \, . \, \rangle$, defined as the natural scalar product over these vectors in λ:

$$\langle F(\lambda), G(\lambda) \rangle = \mathbf{F}^\dagger \mathbf{G} = \sum_{k=1}^{n_\lambda} F_k^* G_k$$

Now that I have introduced a matrix space and its scalar product, all mathematics related to euclidean space can be readily applied. In particular, I will often be interested in defining and using subspace projections. I will denote $\underline{\mathbf{P}}_S$ the projector onto the subspace S, and $\underline{\mathbf{P}}_S^\perp$ the projector orthogonal to S, defined as $\underline{\mathbf{P}}_S^\perp = \mathbf{I} - \underline{\mathbf{P}}_S$

12.2.2 The visibility zoo

In this Chapter, I will use several different visibilities, which all represent different mathematical or physical quantities. Hereafter, I give a short description of the different visibilities used, and what they represent. The letters b, t, and λ are used to represent the dependance of the different quantities to the instrumental baseline, the time, and the wavelength.

$\underline{V}_P(b,t,\lambda)$ — This represents the intrinsic planet visibility, which depends on the physical characteristics of the planet (spectrum and size), as well as on its position in the sky (for the phase). In practice, the planet is not resolved by the interferometer, and the dependance of this quantity to the planet size can be ignored (the planet is a point source).

$\underline{V}_\star(b,t,\lambda)$ This represents the intrinsic star visibility, which depends on the physical characteristics of the star (spectrum and size). Usually, this quantity does not depend on the position of the star, which is centered in the instrument (i.e. the star gives the phase reference).

$\underline{V}_{P,\text{mes}}(b,t,\lambda)$ This is used to mathematically represent the planet visibility $\underline{V}_P$, affected by the instrument response ("measured"). This quantity is only used conceptually in the calculations. $\underline{V}_P$ is never actually measured by the instrument, since when the fiber is centered on the planet, the planet visibility is always contaminated by the stellar visibility. Thus, this quantity should not be confused with $\underline{V}_{\text{refplanet}}$, which is the actual "on planet" measurement.

$\underline{V}_{\star,\text{theo}}$ This represents the theoretical visibility of star of same size as the target star, but with a flat spectrum. In other words, this represents the visibility of the target star $\underline{V}_\star$, divided by its spectrum. It is only used in Equation 12.29.

$\underline{V}_{\text{on planet}}(b,t,\lambda)$ This notation is used to mathematically represent the visibility observed by the instrument when the fiber is centered on the planet location (i.e. this is a model of $\underline{V}_{\text{refplanet}}$) . It is only a notation used to model the observation, and should not be confused with the actual data (given by $\underline{V}_{\text{refplanet}}$).

$\underline{V}_{\text{on star}}(b,t,\lambda)$ This notation is similar to the previous one, but for the star. It represents, in a mathematical sense, the visibility measured by the instrument when the science fiber is centered on the star (i.e. $\underline{V}_{\text{on star}}$ is a model of $\underline{V}_{\text{refstar}}$).

$\underline{V}_{\text{refplanet}}(b,t,\lambda)$ This represents the actual data extracted from the "on planet" observations, and is the only quantity (with its star couterpart $\underline{V}_{\text{refstar}}$) actually accessible to the user.

$\underline{V}_{\text{refstar}}(b,t,\lambda)$ This represents the actual data extracted from the "on star" observations.

$\widetilde{\underline{V}}_{\text{whatever}}$ The tilde notation is used to represent any visibility "shifted to the planet reference frame", i.e. multiplied by a quantity $\exp\{+i\phi_{\alpha,\delta}(b,t,\lambda)\}$, where $\phi_{\alpha,\delta}(b,t,n_\lambda) = 2\pi/\lambda\times(\alpha U+\delta V)$, with (α,δ) the coordinates of the planet with respect to the central star, U and V being the coordinates of baseline b in the UV plane. This is useful to remove the oscillations of the planet visibility due to its offset position.

V_{whatever} When the visibility is not underlined, this means that the quantity represented is the visibility modulus.

12.2.3 A summary of Annex B

In Annex B, I describe a way to extend the commonly used complex likelihood formalism to complex random variables. This is especially useful for interferometry, in which the quantity used are often complex, and affected by complex errors. Here is a short summary of the main results of this Annex.

Covariance and pseudo-covariance

When dealing with complex quantities and complex errors, some of the usual results from real variables cannot be used. In particular, the error propagation is different, and the usual propagation using the covariance matrix is not enough. In fact, when dealing with complex variables, the covariance matrix $\underline{\underline{\mathbf{W}}}$ needs to be completed by a pseudo-covariance matrix[1] $\underline{\underline{\mathbf{Z}}}$.

For a complex random vector $\underline{\mathbf{V}}$, the covariance and covariance matrices are defined by (Equation B.10 and B.11):

$$\underline{\underline{\mathbf{W}}} = \mathrm{cov}(\underline{\mathbf{V}}) = \mathrm{E}\Big(\left[\underline{\mathbf{V}} - \mathrm{mean}(\underline{\mathbf{V}})\right]\left[\underline{\mathbf{V}} - \mathrm{mean}(\underline{\mathbf{V}})\right]^{\dagger}\Big)$$
$$\underline{\underline{\mathbf{Z}}} = \mathrm{pcov}(\underline{\mathbf{V}}) = \mathrm{E}\Big(\left[\underline{\mathbf{V}} - \mathrm{mean}(\underline{\mathbf{V}})\right]\left[\underline{\mathbf{V}} - \mathrm{mean}(\underline{\mathbf{V}})\right]^{T}\Big)$$

With E denoting the expectancy. For the covariance matrix, this corresponds to the classical definition for real random vector.

Using these matrices, propagating the errors becomes very easy. Under a linear transformation $\underline{\mathbf{V}}' = \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{V}}$, the covariance matrices are changed into (Equations B.21 and B.22):

$$\mathrm{cov}\left(V'\right) = \underline{\underline{\mathbf{A}}}\,\mathrm{cov}(\underline{\mathbf{V}})\,\underline{\underline{\mathbf{A}}}^{\dagger}$$
$$\mathrm{pcov}\left(V'\right) = \underline{\underline{\mathbf{A}}}\,\mathrm{pcov}(\underline{\mathbf{V}})\,\underline{\underline{\mathbf{A}}}^{T}$$

Complex-normal law

It is also possible to define a complex normally distributed vector, which corresponds to a complex variable whose real and imaginary parts are gaussians, and potentially correlated. A random vector $\underline{\mathbf{V}}$ is distribued along the complex normal law of mean $\underline{\mathbf{V}}_0$, covariance $\underline{\underline{\mathbf{W}}}$, and pseudo-covariance $\underline{\underline{\mathbf{Z}}}$ (which I denote $\underline{\mathbf{V}} \sim \mathcal{CN}(\underline{\mathbf{V}}_0, \underline{\underline{\mathbf{W}}}, \underline{\underline{\mathbf{Z}}})$) if its Probability Density Function (PDF) obeys (Equation B.32):

$$\mathrm{PDF}\left[\underline{\mathbf{V}}\right](\underline{\mathbf{X}}) \propto \exp\left(-\frac{1}{2}\begin{bmatrix}(\underline{\mathbf{X}} - \underline{\mathbf{V}}_0)^{\dagger} & (\underline{\mathbf{X}}^* - \underline{\mathbf{V}}_0^*)^{\dagger}\end{bmatrix}\begin{pmatrix}\underline{\underline{\mathbf{W}}} & \underline{\underline{\mathbf{Z}}} \\ \underline{\underline{\mathbf{Z}}}^{\dagger} & \underline{\underline{\mathbf{W}}}^*\end{pmatrix}^{-1}\begin{bmatrix}\underline{\mathbf{X}} - \underline{\mathbf{V}}_0 \\ \underline{\mathbf{X}}^* - \underline{\mathbf{V}}_0^*\end{bmatrix}\right)$$

To condense this notation, I will usually introduce extended matrices (for example $\underline{\underline{\mathbf{A}}}_2$, the extended $\underline{\underline{\mathbf{A}}}$) and vectors ($\underline{\mathbf{X}}_2$, the extended $\underline{\mathbf{X}}$), defined by:

$$\underline{\underline{\mathbf{A}}}_2 = \begin{pmatrix}\underline{\underline{\mathbf{A}}} \\ \underline{\underline{\mathbf{A}}}^*\end{pmatrix}$$
$$\underline{\mathbf{X}}_2 = \begin{pmatrix}\underline{\mathbf{X}} \\ \underline{\mathbf{X}}^*\end{pmatrix}$$

The extended covariance matrix $\underline{\underline{\mathbf{W}}}_2$ is a special case of that, defined by:

$$\underline{\underline{\mathbf{W}}}_2 = \begin{pmatrix}\underline{\underline{\mathbf{W}}} & \underline{\underline{\mathbf{Z}}} \\ \underline{\underline{\mathbf{Z}}}^{\dagger} & \underline{\underline{\mathbf{W}}}^*\end{pmatrix}$$

This way, I can write:

$$\mathrm{PDF}\left[\underline{\mathbf{V}}\right](\underline{\mathbf{X}}) \propto \exp\left(-\frac{1}{2}\begin{bmatrix}(\underline{\mathbf{X}}_2 - \underline{\mathbf{V}}_{0,2})^{\dagger} & \left(\underline{\mathbf{X}}_2^* - \underline{\mathbf{V}}_{0,2}^*\right)^{\dagger}\end{bmatrix}\underline{\underline{\mathbf{W}}}_2^{-1}\begin{bmatrix}\underline{\mathbf{X}}_2 - \underline{\mathbf{V}}_{0,2} \\ \underline{\mathbf{X}}_2^* - \underline{\mathbf{V}}_{0,2}^*\end{bmatrix}\right)$$

[1] The need for another matrix can be intuitively understood. A complex random variable of size n can always be described as a real random variable of size $2n$, containing the real and imaginary parts. The real covariance matrix contains $2n \times 2n = 4n^2$ coefficients, and is symmetric. This leaves a total of $2n^2 + n$ real coefficients, which are all necessary to describe the individual covariances between the different elements of the random variable. In the complex case, the complex covariance contains $(n^2 + n)/2$ independant complex coefficients, which gives $n^2 + n$ independant real coefficients. A total of n^2 coefficients are missing...

Maximum likelihood estimator

The most important result of Annex B states that if a measured quantity $\underline{\mathbf{Y}}$ can be linearly modeled by:

$$\underline{\mathbf{Y}} = \underline{\mathbf{A}}\mathbf{X}$$

where $\mathbf{X}$ is a set of real parameters, then the best estimate of $\mathbf{X}$ given a set of measurement $\underline{\mathbf{Y}}$ affected by some gaussian noise of covariance $\underline{\mathbf{W}}$ and pseudo-covariance $\underline{\mathbf{Z}}$ is (Equation B.48):

$$\hat{\mathbf{X}} = \left(\mathrm{Re}\left(\underline{\mathbf{Y}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2 \right) \left[\mathrm{Re}\left(\underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2 \right) \right]^{-1} \right)^T$$

Using the notations from above.

12.3 GRAVITY measurement errors

The GRAVITY pipeline can be used to extract the complex visibility $\underline{\mathbf{V}}_{\mathrm{refplanet}}$ for each baseline b_i, and at each time t_k. The program also returns an estimate of the error on this visibility vector, in the form of a complex vector $\underline{\mathbf{E}} = (\underline{e}_1, \ldots, \underline{e}_{n_\lambda})^T$, where $\mathrm{Re}(\underline{e}_k)^2$ represents the variance of the real part of $\underline{\mathbf{V}}_{\mathrm{refplanet},k}$, and $\mathrm{Im}(\underline{e}_k)^2$ the variance of the imaginary part of $\underline{\mathbf{V}}_{\mathrm{refplanet},k}$. Thus, in the errors as returned by the GRAVITY pipeline, there is no correlation between the real part of the visibility and its imaginary part, nor between different wavelengths or baselines. The error values only take into account photon and detector noises. In our case, this is a critical problem, as any wavelength correlated noise is ignored. In particular, variations in the fiber coupling coefficient η, or atmospheric transmission variations, are not taken into account

To take into account these sources of noise, the errors from the GRAVITY pipeline can be replaced by empirical estimates of the covariance and pseudo-covariance matrices, obtained directly from the data. The following assumptions are made:

1. The values of $\underline{\mathbf{V}}_{\mathrm{refplanet},b,t}$ at different wavelengths can be correlated
2. The values of $\underline{\mathbf{V}}_{\mathrm{refplanet},b_i,t}$ and $\underline{\mathbf{V}}_{\mathrm{refplanet},b_j,t}$ can be correlated, for any two baselines b_i and b_j
3. There is no time correlation (i.e. $\underline{\mathbf{V}}_{\mathrm{refplanet},b_i,t_m}$ and $\underline{\mathbf{V}}_{\mathrm{refplanet},b_j,t_n}$ are always uncorrelated)
4. Within an exposure file (EXP), the covariance and pseudo-covariance matrices of the visibility stay constant, but they may vary from EXP to EXP.

With these assumptions, the covariance and pseudo-covariance matrices of a vector $\underline{\mathbb{V}}$ formed by the concatenation of all the measured planet visibilities:

$$\underline{\mathbb{V}} = \begin{pmatrix} \underline{\mathbf{V}}_{\mathrm{refplanet}_{b_1,t_1}} \\ \vdots \\ \underline{\mathbf{V}}_{\mathrm{refplanet}_{b_{n_b},t_1}} \\ \vdots \\ \underline{\mathbf{V}}_{\mathrm{refplanet}_{b_{n_b},t_{n_t}}} \end{pmatrix}$$

necessarily have the form:

$$\mathrm{cov}(\underline{\mathbb{V}}) = \mathrm{diag}\left(\underline{\mathbf{W}}_{\mathrm{EXP}_1}, \ldots, \underline{\mathbf{W}}_{\mathrm{EXP}_1}, \underline{\mathbf{W}}_{\mathrm{EXP}_2}, \ldots, \underline{\mathbf{W}}_{\mathrm{EXP}_2}, \ldots, \underline{\mathbf{W}}_{\mathrm{EXP}_n}, \ldots, \underline{\mathbf{W}}_{\mathrm{EXP}_n}\right) \quad (12.5)$$

Where each matrix $\underline{\mathbf{W}}_{\mathrm{EXP}_k}$ is repeated for each DIT within the EXP_k. Similarly:

$$\mathrm{pcov}(\underline{\mathbb{V}}) = \mathrm{diag}\left(\underline{\mathbf{Z}}_{\mathrm{EXP}_1}, \ldots, \underline{\mathbf{Z}}_{\mathrm{EXP}_1}, \underline{\mathbf{Z}}_{\mathrm{EXP}_2}, \ldots, \underline{\mathbf{Z}}_{\mathrm{EXP}_2}, \ldots, \underline{\mathbf{Z}}_{\mathrm{EXP}_n}, \ldots, \underline{\mathbf{Z}}_{\mathrm{EXP}_n}\right) \quad (12.6)$$

Using the visibilities acquired at all times $t_1, \ldots, t_{n_k}$ within EXP_k gives (the "refplanet" subscripts have been dropped):

$$\underline{\mathbf{W}}_{\mathrm{EXP}_k} = \frac{1}{n_k - 1} \sum_{j=1}^{n_k} \underline{\mathbf{V}}_{t_j} \underline{\mathbf{V}}_{t_j}^{\dagger} - \frac{1}{n_k(n_k-1)} \left(\sum_{j=1}^{n_k} \underline{\mathbf{V}}_{t_j}\right) \left(\sum_{j=1}^{n_k} \underline{\mathbf{V}}_{t_j}\right)^{\dagger} \quad (12.7)$$

$$\underline{\mathbf{Z}}_{\mathrm{EXP}_k} = \frac{1}{n_k - 1} \sum_{j=1}^{n_k} \underline{\mathbf{V}}_{t_j} \underline{\mathbf{V}}_{t_j}^{T} - \frac{1}{n_k(n_k-1)} \left(\sum_{j=1}^{n_k} \underline{\mathbf{V}}_{t_j}\right) \left(\sum_{j=1}^{n_k} \underline{\mathbf{V}}_{t_j}\right)^{T} \quad (12.8)$$

With:

$$\underline{\mathbf{V}}_{t_k} = \begin{pmatrix} \underline{\mathbf{V}}_{b_1,t_k} \\ \vdots \\ \underline{\mathbf{V}}_{b_{n_b},t_k} \end{pmatrix} \quad (12.9)$$

These general matrices $\mathrm{cov}(\underline{\mathbb{V}})$ and $\mathrm{pcov}(\underline{\mathbb{V}})$, as well as the individual $\mathbf{W}_{\mathrm{EXP}}$ and $\underline{\mathbf{Z}}_{\mathrm{EXP}}$ will be used in the extraction of the astrometry and spectrum.

12.4 Extracting the astrometry

12.4.1 Visibility model

The visibility $\underline{V}_{\text{on planet}}$ (referenced to the star) measured by GRAVITY when the fiber is centered at the planet location is a complex number, which can be seen as a function of the baseline b, the time t, and the wavelength λ: $\underline{V}_{\text{on planet}}(b, t, \lambda)$.

At each time t, and for each baseline b, the measured complex visibility is the sum of the planet visibility, and some remaining coherent star light. In theory, this visibility should be expressed as:

$$\underline{V}_{\text{on planet}}(b, t, \lambda) = \eta_\star(\lambda)\underline{V}_\star(b, \lambda) + \eta_{\mathrm{P}}(\lambda)\underline{V}_{\mathrm{P}} \quad (12.10)$$

With $\eta_\star$ and η_{P} the wavelength-dependent fiber coupling functions for the star and planet light. These two functions are different because the offsets between each object and the fiber are different.

The planet to fiber coupling function η_{P} can be calculated theoretically, but it depends on the offset between the fiber and the planet. It will be necessary to take this function into account when extracting the spectrum (see Section 12.5), but at this stage, the exact position of the planet is unknown. The best option is to use a constant injection:

$$\eta_{\mathrm{P}}(\lambda) = \gamma \quad (12.11)$$

For the star to fiber coupling function, the situation is different. This term does not represent the coupling between the stellar PSF and the fiber, as the offset between the two

is too large ($\simeq 140$ mas). This coupling function is dominated by speckle noise leaking into the fiber. Empirically, the exact shape of this coupling function appears highly variable and strongly chromatic. To properly represent the observed visibilities, I find that a low-order polynomial is required. The complete visibility model is:

$$\underline{V}_{\text{on planet}}(b,t,\lambda) = \underline{Q}(b,t)(\lambda)\underline{V}_{\star}(b,t,\lambda) + \gamma\underline{V}_{\text{P}}(b,t,\lambda) \tag{12.12}$$

where $\underline{V}_{\text{P}}$ is the planet complex visibility, $\underline{V}_{\star}$ the star complex visibility, $\lambda \to \underline{Q}(b,t)(\lambda)$ a polynomial function in λ, of order m (typically, $m \leq 4$).

If we consider both the planet and the star as point sources[2], we can write the planet complex visibility in the form:

$$\underline{V}_{\text{P}}(b,t,\lambda) = C_P(\lambda)V_{\star}(\lambda)e^{-i\phi_{(\alpha,\beta)}(b,t,\lambda)} \tag{12.13}$$

where C is the planet to star contrast spectrum, $V_{\star}$ the star visibility modulus, and $\phi_{(\alpha,\delta)}(b,t,\lambda)$ is the phase offset between the two telescopes of baseline b, at time t, and for wavelength λ. This phase function depends only on the coordinates of the planet relative to the star: (α,δ). When the fiber is directly placed on the star, we have:

$$\underline{V}_{\text{on star}}(b,t,\lambda) = \gamma' V_{\star}(b,t,\lambda) \tag{12.14}$$

And so we can write:

$$\underline{V}_{\text{on planet}}(b,t,\lambda) = \sum_{k=0}^{m} a_k\lambda^k\underline{V}_{\text{on star}}(b,t,\lambda) + \gamma C_{\text{P}}(\lambda)\underline{V}_{\text{on star}}(b,t,\lambda)e^{-i\phi_{\alpha,\delta}(b,t,\lambda)} \tag{12.15}$$

In which the coefficient γ' has been implicitly absorbed in γ and the different a_k.

Equation 12.15 is a multi-parameter model of the visibility data, which can possibly be used to extract the coordinates α,δ. There are still two major issues with this model:

- The on star data are not acquired rigorously at the same time as the data on the planet.
- The contrast spectrum of the planet C_λ is unknown.

For the "on star" visibility, the best way to proceed is to approximate the term $\underline{V}_{\text{on star}}$ at time t by taking the data at the closest available time. In practice, if the strategy described in Section 12.1 is followed, then, for each planet exposure file, there should be one star exposure file taken before, and one after. The "on star" visibility at the time of the planet exposure can simply be approximated by the mean of these two exposures.

For the contrast spectrum, two approaches can be used, depending on the situation. If the observed target has already been discovered, and if its main characteristics (temperature, surface gravity) have been estimated, then it is possible to calculate an approximate contrast spectrum using available models for the planet and star spectra (and a scale factor for the magnitude). If the planet or its main physical characteristics are not constrained by other observations, then an iterative process is required. The astrometry is extracted under an *a priori* assumption on the spectral shape (flat spectrum?). Using this astrometry, a contrast spectrum can be extracted following the recipe given in Section 12.5, which can then be injected back into the astrometry extraction code. At this point, it is just a matter of iterating between the two steps and hoping for convergence.

With these two approximations, the astrometry model has $m \times n_b n_t + n_b n_t + 2$ free parameters: $m \times n_b n_t$ parameters to describe the polynomial Q at each baseline and each exposure, $n_b n_t$ γ coefficients for the planet to fiber coupling (1 per baseline and exposure), and 2 parameters for the planet coordinates (α,δ).

[2]The star may actually be partly resolved by the interferometer. This has little impact on the astrometry, and I only take it into account for the spectrum extraction.

12.4.2 Parametric fitting

12.4.2.1 Global strategy

With $n_b \times n_t$ typically of the order of a few hundreds, if not taken cautiously, the model fitting with $(m+1) \times n_b n_t + 2$ parameters ($m = 3$ or 4) can rapidly become very complicated to perform. But the model given in Equation 12.15 has some interesting properties:

- The model is linear in all coefficients $\underline{a}_k$.
- The model is linear in all coefficients γ.
- The model only couples the different baselines and exposures through the parameters α and δ.

With these characteristics in mind, it is possible to design an efficient algorithm to calculate the astrometric solution. The idea is to use a grid in α, δ, and to calculate the minimum χ^2 (or log-likelihood) at each point of the grid by fitting a model restricted to all the linear parameters. The main advantage is that if α and δ are fixed, the model fitting is independent for each exposure and each baseline. This makes possible to do $n_t \times n_b$ separate linear fit with $(m+1)$ parameters each, which is much faster than a linear fit with all $(m+1)n_b n_t$ parameters. The astrometric solution is then retrieved from the (α, δ) χ^2 map. Mathematically, what this means is that there is a closed-form expression for the log-likelihood minimized over all parameters except α, δ.

12.4.2.2 A closed-form solution for the restricted likelihood

For this section, I will switch to matrix notations (see Section 12.2) and denote $\underline{\mathbf{V}}_{\text{refplanet}}(b_k, t_j)$ the complex visibility vector measured at the planet position. For baseline b_k and time t_j, I also denote $\underline{\theta}_{kj}$ the set of $m+1$ parameters $\underline{\theta}_{ij} = (\underline{a}_0, \ldots, \underline{a}_m, \gamma)$, and I introduce the complete parameter set $\underline{\theta} = (\underline{\theta}_{11}, \underline{\theta}_{21}, \ldots, \underline{\theta}_{n_b n_t}, \alpha, \delta)$. I also write $\underline{\mathbf{V}}_{\text{on planet}}(\underline{\theta}_{kj}, \alpha, \delta)$ for the model restricted to baseline b_k and time t_j (which only depends on $\underline{\theta}_{kj}$ and α, δ). I denote $\underline{\mathbf{W}}_{kj}$ and $\underline{\mathbf{Z}}_{kj}$ the covariance and pseudo-covariance matrices of $\underline{\mathbf{V}}_{\text{refplanet}}(b_k, t_j)$, which corresponds to the appropriate $\underline{\mathbf{W}}_{\text{EXP}}$ and $\underline{\mathbf{Z}}_{\text{EXP}}$ introduced in Section 12.3. The expression of the complete likelihood for parameters value $\underline{\theta}$ is the product of the likelihood restricted to each baseline and exposure:

$$\mathcal{L}(\underline{\theta}) = \prod_{k=1}^{n_b} \prod_{j=1}^{n_t} \mathcal{L}(\underline{\theta}_{kj}, \alpha, \delta) \tag{12.16}$$

With:

$$\mathcal{L}(\underline{\theta}_{kj}, \alpha, \delta) \propto \exp\left(- \left[\underline{\mathbf{V}}_{\text{refplanet}}(b_k, t_j) - \underline{\mathbf{V}}_{\text{on planet}}(\underline{\theta}_{kj}, \alpha, \delta) \right]^T \underline{\mathbf{W}}_{kj,2} \left[\underline{\mathbf{V}}_{\text{refplanet}}(b_k, t_j) - \underline{\mathbf{V}}_{\text{on planet}}(\underline{\theta}_{kj}, \alpha, \delta) \right] \right)$$

Taking the log-likelihood:

$$-\log \mathcal{L}(\underline{\theta}) = \sum_{k=1}^{n_b} \sum_{j=1}^{n_t} -\log \mathcal{L}(\underline{\theta}_{kj}, \alpha, \delta) \tag{12.17}$$

The quantity of interest is actually $-\log \mathcal{L}(\alpha, \delta)$, corresponding to the log-likelihood minimized over all parameters other than α and δ. Minimizing this quantity over all parameters is the same as minimizing independently all $\log \mathcal{L}(\underline{\theta}_{kj}, \alpha, \delta)$ over $\underline{\theta}_{kj}$.

Explicitly separating α, δ from the rest of the parameters, and dealing with each exposure one at a time, the model can be restricted to baseline b_k and time t_j and written:

$$\underline{\mathbf{V}}^{(\alpha,\delta)}_{\text{on planet},b_k,t_j}(\underline{a}_0, \underline{a}_1, \ldots, \underline{a}_m, \gamma) = \sum_{k=0}^{m} \underline{a}_k \Lambda^k \underline{\mathbf{V}}_{\text{refstar}}(b_k, t_j) + \gamma\, \mathbb{C}\underline{\Phi}_{(\alpha,\delta)}(b_k, t_j)\underline{\mathbf{V}}_{\text{refstar}}(b_k, t_j) \tag{12.18}$$

Where I have introduced the following matrices:

$$\underline{\Phi}_{(\alpha,\delta)}(b_k, t_j) = \begin{pmatrix} e^{-i\phi_{(\alpha,\delta)}(b_k,t,\lambda_1)} & 0 & \ldots & 0 \\ 0 & e^{-i\phi_{(\alpha,\delta)}(b_k,t,\lambda_2)} & \ldots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & e^{-i\phi_{(\alpha,\delta)}(b_k,t,\lambda_{n_\lambda})} \end{pmatrix} \tag{12.19}$$

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_{n_\lambda} \end{pmatrix} \tag{12.20}$$

$$\mathbb{C} = \begin{pmatrix} C(\lambda_1) & 0 & \ldots & 0 \\ 0 & C(\lambda_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & C(\lambda_{n_\lambda}) \end{pmatrix} \tag{12.21}$$

And where the reference visibility $\underline{\mathbf{V}}_{\text{refstar}}$ is a measured quantity which should corresponds to the model $\underline{\mathbf{V}}_{\text{on star}}$.

I am now almost ready to apply the result of Annex B to obtain a closed-form solution for the minimum restricted log-likelihood. But I still need to write the model in a pure matrix form, and to get rid of the complex nature of the parameters $\underline{\theta}_{kj}$. To do so, I introduce the alternate parameters:

$$\zeta_{kj} = \left(\mathrm{Re}(\underline{a}_0)\,, \mathrm{Im}(\underline{a}_0)\,, \ldots, \mathrm{Re}(\underline{a}_m)\,, \mathrm{Im}(\underline{a}_m)\,, \gamma\right) \tag{12.22}$$

And write:

$$\underline{\mathbf{V}}_{(\alpha,\delta),b_k,t_j}(\theta_{kj}) = \underline{\mathbf{A}}\zeta_{kj} \tag{12.23}$$

With:

$$\underline{\mathbf{A}} = \left[\Lambda^0\underline{\mathbf{V}}_{\text{refstar}}(b_k,t_j) \quad i\,\Lambda^0\underline{\mathbf{V}}_{\text{refstar}}(b_k,t_j) \quad \ldots \quad \Lambda^m\underline{\mathbf{V}}_{\text{refstar}}(b_k,t_j) \quad i\,\Lambda^m\underline{\mathbf{V}}_{\text{refstar}}(b_k,t_j) \quad \mathbb{C}\underline{\Phi}_{(\alpha,\delta)}(b_k,t_j)\underline{\mathbf{V}}_{\text{refstar}}\right]$$

Under these notations, the results of Annex B fully apply, and the best estimate for ζ_{kj} is:

$$\widehat{\zeta}_{kj} = \left(\mathrm{Re}\left(\left[\underline{\mathbf{V}}_{\text{refplanet}}\right]_2^\dagger \underline{\mathbf{W}}_2^{-1}\underline{\mathbf{A}}_2\right)\left[\mathrm{Re}\left(\underline{\mathbf{A}}_2^\dagger\underline{\mathbf{W}}_2^{-1}\underline{\mathbf{A}}_2\right)\right]^{-1}\right)^T \tag{12.24}$$

Where the dependence of $\underline{\mathbf{A}}$, $\underline{\mathbf{W}}_2$, and $\underline{\mathbf{V}}_{\text{refplanet}}$ to k, j is implicit. The associated minimum log-likelihood is:

$$\min_{\theta_{kj}}\left(-\log\mathcal{L}\left(\theta_{kj},\alpha,\delta\right)\right) = \left[\underline{\mathbf{V}}_{\text{refplanet}}(b_k,t_j) - \underline{\mathbf{A}}_{kj}\widehat{\zeta}_{kj}\right]_2^\dagger \underline{\mathbf{W}}_{kj,2}{}^{-1}\left[\underline{\mathbf{V}}_{\text{refplanet}}(b_k,t_j) - \underline{\mathbf{A}}_{kj}\widehat{\zeta}_{kj}\right]_2 \tag{12.25}$$

Which finally gives, for the log-likelihood restricted to α, δ:

$$-\log \mathcal{L}(\alpha,\delta) = \sum_{k=0}^{n_b}\sum_{j=1}^{n_t} \left[\underline{\mathbf{V}}_{\text{refplanet}}(b_k,t_j) - \underline{\mathbf{A}}_{kj}\widehat{\zeta_{kj}}\right]_2^{\dagger} \underline{\mathbf{W}}_{kj,2}{}^{-1} \left[\underline{\mathbf{V}}_{\text{refplanet}}(b_k,t_j) - \underline{\mathbf{A}}_{kj}\widehat{\zeta_{kj}}\right]_2 \tag{12.26}$$

This expression can be used together with a grid-based search algorithm to obtain numerically the maximum likelihood estimate of (α, δ).

12.4.3 Astrometry error bars

From the log-likelihood map $-\ln \mathcal{L}(\alpha,\delta)$, the best estimate of (α,δ) can be taken as:

$$\widehat{(\alpha,\delta)} = \text{argmin}\left\{-\ln \mathcal{L}(\alpha,\delta)\right\} \tag{12.27}$$

And the 95% interval confidence as the $\min(-\ln \mathcal{L}) + 5.99$ contour (see Annex B).

12.5 Extracting the spectrum

12.5.1 Visibility model

The visibility model already introduced in the previous section is:

$$\underline{V}_{\text{on planet}}(b,t,\lambda) = \underline{Q}(b,t)(\lambda)\underline{V}_{\text{on star}}(b,t,\lambda) + \eta_{\text{P}}(\lambda)\underline{V}_{\text{P}}(b,t,\lambda) \tag{12.28}$$

where $\underline{V}_{\text{P}}$ is the planet complex visibility, $\lambda \to \underline{Q}(b,t)(\lambda)$ a polynomial function in λ, of order m, and $\underline{V}_{\text{on star}}$ the complex visibility at the position of the star (and where the fiber injection coefficient from Equation 12.14 has been absorbed in the polynomial Q).

For estimating the astrometry, the planet to fiber coupling function $\eta_{\text{P}}(\lambda)$ was approximated as a constant: $\eta_{\text{P}}(\lambda) = \gamma$. To extract the spectrum, the correct shape of this wavelength-dependent function needs to be taken into account.

12.5.1.1 Planet visibility

The planet can be considered as a point source. In theory, its visibility sould be directly given by its spectrum S_{p}, modulated by the OPD between the different telescopes:

$$\underline{V}_{\text{P}}(b,t,\lambda) = S_P(\lambda)e^{-i\phi_{(\alpha,\delta)}(b,t,\lambda)}$$

In practice, though, the instrument is certainly not perfect, and the instrumental response R affects the measured planet visibility:

$$\underline{V}_{P,\text{mes}}(b,t,\lambda) = S_P(\lambda)R(b,t,\lambda)e^{-i\phi_{(\alpha,\delta)}(b,t,\lambda)}$$

The usual way to take into account this instrumental response is to observe a calibrator, of known visibility. In the case of exoplanet observation, the star provides a natural calibrator. However, there are a few differences between the planet and the star which needs to be taken into account when doing so:

- The spectrum of the star is neither flat, nor the same than the planet spectrum.
- Contrary to the planet, the star can be partly resolved by the interferometer.

- The star is detected by the acquisition camera, and the fiber can be correctly centered on its photocenter. This is not the case for the planet, and the fiber must then be placed at a given offset from the star. Since the position of the planet is not perfectly known, the fiber is not perfectly centered on the planet. Thus, the fiber injection and the instrumental response are different when observing the star and the planet.

These differences can all be taken account into by introducing the planet to star contrast spectrum $C(\lambda)$, and by taking into account the planet to fiber coupling ratio and the shape of the resolved star visibility in the definition of an "on star instrument response" G:

$$G(b,t,\lambda) = \frac{\eta_{\mathrm{P}}(b,t)V_{\mathrm{refstar}}(b,\lambda)}{V_{\star,\mathrm{theo}}(b,\lambda)} \tag{12.29}$$

where η_{P} is the injection function when observing the planet, which can be theoretically calculated (Annex A) from the effective position of the fiber and the planet astrometry (extracted using the method described in Section 12.4), $V_{\mathrm{refstar}}(b,\lambda)$ is the star visibility modulus measured by the instrument (i.e. the visibility at the position of the star), and $V_{\star,\mathrm{theo}}(b,\lambda)$ is the theoretical visibility of the star, due to its spatial extension (and ignoring its spectrum)[3].
In other word:

$$\begin{aligned} G(b,t,\lambda) &= \frac{\eta(\lambda)V_{\mathrm{refstar}}(b,t,\lambda)}{\frac{V_\star(b,t,\lambda)}{S_\star(\lambda)}} \\ &= \frac{\eta(\lambda)R(b,t,\lambda)V_\star(b,t,\lambda)}{\frac{V_\star(b,t,\lambda)}{S_\star(\lambda)}} \\ &= \eta(\lambda)R(b,t,\lambda)S_\star(\lambda) \end{aligned}$$

Which justifies that the measured planet visibility can be written in this form better suited to the spectrum extraction:

$$\underline{V}_{P,\mathrm{mes}}(b,t,\lambda) = C(\lambda)G(b,t,\lambda)e^{-i\phi_{(\alpha,\delta)}(b,t,\lambda)} \tag{12.30}$$

where C is the planet to star contrast spectrum, $V_\star$ the star visibility modulus, and $\phi_{(\alpha,\delta)}(b,t,\lambda)$ is the phase offset between the two telescopes of baseline b, at time t, and for wavelength λ.

The main interest of this Equation 12.30 is to explicitly introduce the contrast spectrum $C(\lambda)$, which is largely independant of atmospheric variations, and telluric lines. The two spectra S_{P} and $S_\star$ introduced in the previous equations are strongly affected by the atmosphere. But if the "on star" observations used to calculate $\underline{V}_{\mathrm{refstar}}$ are close in time to the "on planet" observations, then these two spectra are affected by the same atmospheric absorption, which disappears from the ratio C. The definition of G given by Equation 12.29 also alleviates the need to explicitly introduce the spectrum of the star (possibly affected by the atmosphere) at this stage of the reduction. This spectrum stays hidden in the V_{refstar} quantity.

The phase function in Equation 12.30 depends only on the coordinates of the planet relative to the star: (α,δ). From now on, I will denote:

$$\underline{\Phi}_{(\alpha,\delta)}(b,t,\lambda) = e^{-i\phi_{(\alpha,\delta)}(b,t,\lambda)} \tag{12.31}$$

I will also consider that the position of the planet is known (measured on the GRAVITY data using the method described in the previous section), and drop the (α,δ) subscript.

[3] This theoretical visibility can be calculated using the star diameter. In most cases, though, the star will remain mostly unresolved, and this term can be ignored. But if a measurement of its diameter exists in the literature, like it is the case for β Pictoris [Defrère et al., 2012], it can be injected in the definition of G.

Oscillations in the planet visibility can be removed by "shifting all visibilities to the planet reference frame", i. e. by multiplying them by $\underline{\Phi}^*$ (this removes the oscillations in the planet visibility, since $\underline{\Phi}\,\underline{\Phi}^* = 1$). I will denote $\underline{\widetilde{V}}$ these shifted visibilities: $\underline{\widetilde{V}} = \underline{\Phi}^* \underline{V}$. I can now write:

$$\underline{\widetilde{V}}_{\text{on planet}}(b,t,\lambda) = \underline{Q}(b,t)(\lambda)\underline{\widetilde{V}}_{\text{on star}}(b,t,\lambda) + C(\lambda)G(b,t,\lambda) \tag{12.32}$$

12.5.2 Orthogonal projection onto the stellar visibility space

Equation 12.32 shows that the total visibility is a combination of the stellar visibility multiplied by a polynomial in λ of order m, and the planet contrast spectrum C, which is the quantity I am trying to extract. The exact value of the polynomial $\underline{Q}$ depends on the fiber injection, which is strongly affected by the residual atmospheric turbulence (the equivalent of the the so-called "speckle noise" in direct imaging). Thus, I need a way to remove this term from the equation. To do so, I will project Equation 12.32 on a space orthogonal to the stellar visibility model.

Considering only a particular baseline b_k and time t_l, I can drop these indices, switch to a matrix notation, and write:

$$\underline{\widetilde{\mathbf{V}}}_{\text{on planet}} = \sum_{k=0}^{m} \underline{a}_k \Lambda^k \underline{\widetilde{\mathbf{V}}}_{\text{on star}} + \mathbf{GC} \tag{12.33}$$

where the $\underline{a}_k$ are the coefficients of the polynomial $Q(b,t)$, and where I have introduced:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_{n_\lambda} \end{pmatrix} \tag{12.34}$$

And the instrument response $\mathbf{G}$:

$$\mathbf{G} = \begin{pmatrix} G_{\lambda_1} & 0 & \dots & 0 \\ 0 & G_{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & G_{\lambda_{n_\lambda}} \end{pmatrix} \tag{12.35}$$

At this point, I can use the projector orthogonal to the subspace $\mathbb{C}^m[\Lambda]\,\underline{\widetilde{\mathbf{V}}}_{\text{on star}}$, generated by $(\Lambda^0 \underline{\widetilde{\mathbf{V}}}_{\text{on star}}, \dots, \Lambda^m \underline{\widetilde{\mathbf{V}}}_{\text{on star}})$, to get rid of the polynomial sum:

$$\underline{\mathbf{P}}^{\perp}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star}}} \underline{\widetilde{\mathbf{V}}} = \sum_{k=1}^{m} a_k \underline{\mathbf{P}}^{\perp}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star}}} \Lambda^k \underline{\widetilde{\mathbf{V}}}_{\text{on star}} + \underline{\mathbf{P}}^{\perp}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star}}} \mathbf{GC} \tag{12.36}$$

By definition:

$$\forall k \leq m, \qquad \underline{\mathbf{P}}^{\perp}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star}}} \Lambda^k \underline{\widetilde{\mathbf{V}}}_{\text{on star}} = 0 \tag{12.37}$$

And now, I have this very simple result, witten in the most abstruse notation:

$$\underline{\mathbf{P}}^{\perp}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star}}} \underline{\widetilde{\mathbf{V}}} = \underline{\mathbf{P}}^{\perp}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star}}} \mathbf{GC} \tag{12.38}$$

In common language, this Equation 12.38 is only taking advantage of the fact that the visibility model given in Equation 12.32, which is an equality between different vectors, is valid in any subspace of the complex space $\mathbb{C}^{n_\lambda}$ (in which all the vectors are defined). In particular, this visibility model can be written in the subspace orthogonal to all the $m+1 < n_\lambda$ vectors $\Lambda^0\underline{\widetilde{\mathbf{V}}}_{\text{on star}}, \ldots, \Lambda^m\underline{\widetilde{\mathbf{V}}}_{\text{on star}}$ in which the unknown speckle noise simply vanishes. Equation 12.38 is a projection of Equation 12.32 "orthogonally to the speckle noise subspace".

12.5.3 Sky rotation and problem inversion

Equation 12.38 is very elegant in the sense that the coefficients a_k, which were the residual of the unknown speckle noise and fiber injection variations, have completely disappeared. This came at a cost, though, as the orthogonal projector used is necessarily a matrix of rank $< n_\lambda$, ans thus cannot be inverted. The projection of the planet visibility has removed any part of the planet visibility colinear to the stellar visibility model. This means that part of the planet visibility is now impossible to recover.

Fortunately, the stellar visibility varies with time t and baseline b: $\underline{\widetilde{\mathbf{V}}}_{\text{on star}} = \underline{\Phi}^*(b,t)\underline{\mathbf{V}}_{\text{on star}}$, due to the sky rotation. Consequently, the "speckle noise subspace", defined by the set of vectors $\Lambda^0\underline{\mathbf{V}}_{\text{on star}}, \ldots, \Lambda^m\underline{\mathbf{V}}_{\text{on star}}$ is not constant, but varies with b and t (due to the sky rotation). Thus, the information on the planet visibility which is lost during the projection is different for each b and t, and by combining all the equations, it is possible to recover the complete information.

From now on, I will reintroduce the variables b and t that I previously dropped. To simplify the notation[4], I will also denote:

$$\underline{\mathbf{P}}^\perp_{b,t} = \underline{\mathbf{P}}^\perp_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star}}(b,t)} \tag{12.39}$$

$$\underline{\widetilde{\mathbf{V}}}_{b,t} = \underline{\widetilde{\mathbf{V}}}_{\text{on planet}}(b,t) \tag{12.40}$$

I have:

$$\forall (b,t) \qquad \underline{\mathbf{P}}^\perp_{b,t}\underline{\widetilde{\mathbf{V}}}_{b,t} = \underline{\mathbf{P}}^\perp_{b,t}\mathbf{G}_{b,t}\mathbf{C} \tag{12.41}$$

Denoting $(b_1, \ldots, b_{n_b})$ the different baselines, and $(t_1, \ldots, t_{n_t})$ the different times at which data have been acquired, I can concatenate all these equations by defining a few new matrices:

$$\underline{\mathbf{P}}^\perp = \begin{pmatrix} \underline{\mathbf{P}}^\perp_{b_1,t_1} & 0 & \ldots & 0 & 0 \\ 0 & \underline{\mathbf{P}}^\perp_{b_2,t_1} & \ddots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \underline{\mathbf{P}}^\perp_{b_{n_b},t_{n_t-1}} & 0 \\ 0 & 0 & \ldots & 0 & \underline{\mathbf{P}}^\perp_{b_{n_b},t_{n_t}} \end{pmatrix} \tag{12.42}$$

$$\underline{\widetilde{\mathbb{V}}} = \begin{pmatrix} \underline{\widetilde{\mathbf{V}}}_{b_1,t_1} \\ \vdots \\ \underline{\widetilde{\mathbf{V}}}_{b_{n_b},t_1} \\ \vdots \\ \underline{\widetilde{\mathbf{V}}}_{b_{n_b},t_{n_t}} \end{pmatrix} \tag{12.43}$$

[4] I'm trying, really...

$$\mathbb{G} = \begin{pmatrix} \mathbf{G}_{b_1,t_1} \\ \mathbf{G}_{b_2,t_1} \\ \vdots \\ \mathbf{G}_{b_{n_b},t_{n_t-1}} \\ \mathbf{G}_{b_{n_b},t_{n_t}} \end{pmatrix} \tag{12.44}$$

With these notations, the entire set of equations given by 12.41 is reduced to the very compact form:

$$\underline{\mathbf{P}}^{\perp}\underline{\widetilde{\mathbb{V}}} = \underline{\mathbf{P}}^{\perp}\mathbb{G}\mathbf{C} \tag{12.45}$$

Equation 12.45 is a huge improvement over equation 12.38, as solving for $\mathbf{C}$ (in the least square sense) now requires the matrix $\left(\mathbf{P}^{\perp}\mathbb{G}\right)^T\mathbf{P}^{\perp}\mathbb{G}$ to be invertible, which is a much less restrictive condition than all the individual $\underline{\mathbf{P}}^{\perp}_{b,t}\mathbf{G}$ being invertible.

To further simplify the notations, I introduce the projected visibilities $\underline{\widetilde{\mathbb{U}}} = \underline{\mathbf{P}}^{\perp}\underline{\widetilde{\mathbb{V}}}$ and the collapsed projection operator $\underline{\mathbb{P}} = \underline{\mathbf{P}}^{\perp}\mathbb{G}$. Equation 12.45 takes the final form:

$$\underline{\widetilde{\mathbb{U}}} = \underline{\mathbb{P}}\mathbf{C} \tag{12.46}$$

12.5.4 Independent system of equations

At this point, I may have to remind the reader that the objective of this section is to find the best estimate of the contrast spectrum $\mathbf{C}$ from the GRAVITY data...

To do so, I plan to use again the maximum likelihood formalism that I introduced in Annex B, and which I already used to extract the astrometry in Section 12.4. But Equation 12.46 has a serious problem, which needs to be taken care of. Denoting $\underline{\mathbf{W}}$ and $\underline{\mathbf{Z}}$ the covariance and pseudo-covariance matrices of $\underline{\widetilde{\mathbb{V}}}$, the covariance and pseudo-covariance matrices of $\underline{\widetilde{\mathbb{U}}}$ are respectively given by $\underline{\mathbf{P}}^{\perp}\underline{\mathbf{W}}\underline{\mathbf{P}}^{\perp\dagger}$ and $\underline{\mathbf{P}}^{\perp}\underline{\mathbf{Z}}\underline{\mathbf{P}}^{\perp T}$. $\underline{\mathbf{P}}^{\perp}$ being a projector matrix, it is not invertible, and thus the covariance and pseudo-covariance matrices of $\underline{\widetilde{\mathbb{V}}}$ are not invertible anymore.

As a consequence, to be able to use the maximum likelihood formalism, I first need to operate a change of basis to make sure that the system of equations is invertible. To do so, I need to go back to the individual Equations 12.41. For each (b,t), noting that $\underline{\mathbf{P}}^{\perp}_{b,t}$ is a projector, I can introduce an Hermitian matrix $\underline{\mathbf{H}}_{b,t}$ such that[5]:

$$\underline{\mathbf{H}}_{b,t}\underline{\mathbf{P}}^{\perp}_{b,t}\underline{\mathbf{H}}^{\dagger}_{b,t} = \mathbf{D}_{b,t} \tag{12.47}$$

$$\underline{\mathbf{H}}_{b,t}\underline{\mathbf{H}}^{\dagger}_{b,t} = \mathbf{I} \tag{12.48}$$

where the matrix $\mathbf{D}_{b,t}$ has the particular form $\mathbf{D}_{b,t} = \mathrm{diag}(1,\dots,1,0,\dots,0)$. This particular form of $\mathbf{D}_{b,t}$ comes from the fact that $\underline{\mathbf{P}}^{\perp}_{b,t}$ is an orthogonal projector. The number of 0s at the end of the diagonal of $\mathbf{D}_{b,t}$ depends on the rank of $\underline{\mathbf{P}}^{\perp}_{b,t}$, and thus on the rank of the set of vectors $\Lambda^k\underline{\widetilde{\mathbf{V}}}_{\star b,t}$, $k \in \{0,\dots,m\}$. It can range from 1 to $m+1$.

Multiplying each individual equation of 12.41 by $\underline{\mathbf{H}}_{b,t}$, noting that $\underline{\mathbf{H}}_{b,t}\underline{\mathbf{P}}^{\perp}_{b,t} = \mathbf{D}_{b,t}\underline{\mathbf{H}}_{b,t}$, and explicitly including the noise term $\underline{\mathbf{E}}_{b,t}$ on $\underline{\widetilde{\mathbf{V}}}_{b,t}$:

$$\mathbf{D}_{b,t}\underline{\mathbf{H}}_{b,t}\underline{\widetilde{\mathbf{V}}}_{b,t} = \mathbf{D}_{b,t}\underline{\mathbf{H}}_{b,t}\mathbf{G}_{b,t}\mathbf{C} + \mathbf{D}_{b,t}\underline{\mathbf{H}}_{b,t}\underline{\mathbf{E}}_{b,t} \tag{12.49}$$

[5]This corresponds to a diagonalization of the matrix $\underline{\mathbf{P}}^{\perp}_{b,t}$ in an orthonormal basis. Since $\underline{\mathbf{P}}^{\perp}_{b,t}$ is an orthogonal projector, the existence of such a basis is always guaranteed.

At this point, I can split the matrices $\mathbf{D}_{b,t}$ and $\underline{\mathbf{H}}_{b,t}$ in blocks:

$$\mathbf{D}_{b,t} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{12.50}$$

$$\underline{\mathbf{H}}_{b,t} = \begin{bmatrix} \underline{\mathbf{H}}_{b,t_{11}} & \underline{\mathbf{H}}_{b,t_{12}} \\ \underline{\mathbf{H}}_{b,t_{21}} & \underline{\mathbf{H}}_{b,t_{22}} \end{bmatrix} \tag{12.51}$$

A block calculation of Equation 12.49 then gives the following equation:

$$\begin{bmatrix} \underline{\mathbf{H}}_{b,t_{11}} & \underline{\mathbf{H}}_{b,t_{12}} \end{bmatrix} \widetilde{\underline{\mathbf{V}}}_{b,t} = \begin{bmatrix} \underline{\mathbf{H}}_{b,t_{11}} & \underline{\mathbf{H}}_{b,t_{12}} \end{bmatrix} \mathbf{G}_{b,t}\mathbf{C} + \begin{bmatrix} \underline{\mathbf{H}}_{b,t_{11}} & \underline{\mathbf{H}}_{b,t_{12}} \end{bmatrix} \underline{\mathbf{E}}_{b,t} \tag{12.52}$$

And a dummy equality $0 = 0 + 0$ for the second line.

The use of the change-of-basis matrix $\underline{\mathbf{H}}_{b,t}$ has led to the reduced Equation 12.52. The noise term $\begin{bmatrix} \underline{\mathbf{H}}_{b,t_{11}} & \underline{\mathbf{H}}_{b,t_{12}} \end{bmatrix} \underline{\mathbf{E}}_{b,t}$ which appears in this equation now has invertible covariance and pseudo-covariance matrices. This can be easily proven by writing:

$$\begin{aligned} \operatorname{cov}\left(\begin{bmatrix} \underline{\mathbf{H}}_{11} & \underline{\mathbf{H}}_{12} \end{bmatrix} \underline{\mathbf{E}}\right) &= \begin{bmatrix} \underline{\mathbf{H}}_{11} & \underline{\mathbf{H}}_{12} \end{bmatrix} \underline{\mathbf{W}} \begin{bmatrix} \underline{\mathbf{H}}_{11}^{\dagger} \\ \underline{\mathbf{H}}_{12}^{\dagger} \end{bmatrix} \\ &= \underline{\mathbf{H}}_{11}\underline{\mathbf{W}}\underline{\mathbf{H}}_{11}^{\dagger} + \underline{\mathbf{H}}_{11}\underline{\mathbf{W}}\underline{\mathbf{H}}_{12}^{\dagger} + \underline{\mathbf{H}}_{12}\underline{\mathbf{W}}\underline{\mathbf{H}}_{11}^{\dagger} + \underline{\mathbf{H}}_{12}\underline{\mathbf{W}}\underline{\mathbf{H}}_{12}^{\dagger} \\ &= \left(\underline{\mathbf{H}}\underline{\mathbf{W}}\underline{\mathbf{H}}^{\dagger}\right)_{11} \end{aligned}$$

in which the reader will have recognized the expression of the upper-left block of $\underline{\mathbf{H}}\underline{\mathbf{W}}\underline{\mathbf{H}}^{\dagger}$.

Since $\underline{\mathbf{H}}\underline{\mathbf{W}}\underline{\mathbf{H}}^{\dagger}$ is the covariance matrix of $\underline{\mathbf{H}}\underline{\mathbf{E}}$, its upper-left square matrix $\left(\underline{\mathbf{H}}\underline{\mathbf{W}}\underline{\mathbf{H}}^{\dagger}\right)_{11}$ is the covariance matrix of the first components of $\underline{\mathbf{H}}\underline{\mathbf{E}}$. Thus, it is also invertible. A similar reasoning holds true for $\operatorname{pcov}\left(\begin{bmatrix} \underline{\mathbf{H}}_{11} & \underline{\mathbf{H}}_{12} \end{bmatrix} \underline{\mathbf{E}}\right)$.

The entire system of Equations 12.41 can thus be reduced to the following system:

$$\forall (b,t), \quad \begin{bmatrix} \underline{\mathbf{H}}_{b,t_{11}} & \underline{\mathbf{H}}_{b,t_{12}} \end{bmatrix} \widetilde{\underline{\mathbf{V}}}_{b,t} = \begin{bmatrix} \underline{\mathbf{H}}_{b,t_{11}} & \underline{\mathbf{H}}_{b,t_{12}} \end{bmatrix} \mathbf{V}\mathbf{C} \tag{12.53}$$

which has the major advantage of being subject to a noise term which has invertible covariance and pseudo-covariance matrices.

12.5.5 Full problem inversion

Introducing:

$$\underline{\mathbf{H}} = \begin{bmatrix} (\underline{\mathbf{H}}_{b_1,t_1})_{11}(\underline{\mathbf{H}}_{b_1,t_1})_{12} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & (\underline{\mathbf{H}}_{b_2,t_1})_{11}(\underline{\mathbf{H}}_{b_2,t_1})_{12} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & (\underline{\mathbf{H}}_{b_{n_b},t_{n_t}})_{11}(\underline{\mathbf{H}}_{b_{n_b},t_{n_t}})_{12} \end{bmatrix} \tag{12.54}$$

I can redefine:

$$\widetilde{\underline{\mathbb{U}}} = \underline{\mathbf{H}}\widetilde{\underline{\mathbb{V}}} \tag{12.55}$$

Where:

$$\widetilde{\underline{\mathbb{V}}} = \begin{pmatrix} \widetilde{\underline{\mathbf{V}}}_{b_1,t_1} \\ \vdots \\ \widetilde{\underline{\mathbf{V}}}_{b_{n_b},t_1} \\ \vdots \\ \widetilde{\underline{\mathbf{V}}}_{b_{n_b},t_{n_t}} \end{pmatrix} \tag{12.56}$$

And also introduce:

$$\underline{\mathbb{H}} = \begin{pmatrix} \left[(\underline{\mathbf{H}}_{b_1,t_1})_{11} \quad (\underline{\mathbf{H}}_{b_1,t_1})_{12}\right] \mathbf{G}_{b_1,t_1} \\ \left[(\underline{\mathbf{H}}_{b_2,t_1})_{11} \quad (\underline{\mathbf{H}}_{b_2,t_1})_{12}\right] \mathbf{G}_{b_2,t_1} \\ \vdots \\ \left[(\underline{\mathbf{H}}_{b_{n_b},t_{n_t}})_{11} \quad (\underline{\mathbf{H}}_{b_{n_b},t_{n_t}})_{12}\right] \mathbf{G}_{b_{n_b},t_{n_t}} \end{pmatrix} \tag{12.57}$$

Then, the system of Equations 12.57 can be combined in:

$$\underline{\widetilde{\mathbb{U}}} = \underline{\mathbb{H}}\mathbf{C} \tag{12.58}$$

where all previous considerations guarantee that the noise term affecting $\underline{\widetilde{\mathbb{U}}}$ has invertible covariance and pseudo-covariance matrices.

And now the results of Annex B can finally be applied! The maximum likelihood estimate of the planet spectrum is given by:

$$\widehat{\mathbf{C}} = \left(\mathrm{Re}\left(\underline{\widetilde{\mathbb{U}}}_2^{\dagger} \underline{\mathbf{W}}_2^{-1} \underline{\mathbb{H}}_2\right) \left[\mathrm{Re}\left(\underline{\mathbb{H}}_2^{\dagger} \underline{\mathbf{W}}_2^{-1} \underline{\mathbb{H}}_2\right)\right]^{-1} \right)^T \tag{12.59}$$

Where a new matrix is implicitly defined:

$$\underline{\mathbf{W}}_2 = \begin{pmatrix} \underline{\mathbf{W}} & \underline{\mathbf{Z}} \\ \underline{\mathbf{Z}}^{\dagger} & \underline{\mathbf{W}}^{*} \end{pmatrix} \tag{12.60}$$

With $\underline{\mathbf{W}}$ and $\underline{\mathbf{Z}}$ the covariance matrices of $\underline{\widetilde{\mathbb{U}}}$, which can be calculated from the covariance matrices of $\underline{\widetilde{\mathbb{V}}}$ using Equation 12.55:

$$\underline{\mathbf{W}} = \mathrm{cov}\left(\underline{\widetilde{\mathbb{U}}}\right) = \underline{\mathbf{H}}\,\mathrm{cov}\left(\underline{\widetilde{\mathbb{V}}}\right) \underline{\mathbf{H}}^{\dagger} \tag{12.61}$$

$$\underline{\mathbf{Z}} = \mathrm{pcov}\left(\underline{\widetilde{\mathbb{U}}}\right) = \underline{\mathbf{H}}\,\mathrm{pcov}\left(\underline{\widetilde{\mathbb{V}}}\right) \underline{\mathbf{H}}^{T} \tag{12.62}$$

12.5.6 Spectrum error bars

Deriving the final error bars on the spectrum estimate $\hat{\mathbf{C}}$ from the covariance matrices of $\underline{\widetilde{\mathbb{U}}}$ is straightforward using Equation 12.59. I have:

$$\mathrm{cov}\left(\underline{\widetilde{\mathbb{U}}}_2\right) = \begin{pmatrix} \underline{\mathbf{W}} & \underline{\mathbf{Z}} \\ \underline{\mathbf{Z}}^{\dagger} & \underline{\mathbf{W}}^{*} \end{pmatrix} \tag{12.63}$$

$$\mathrm{pcov}\left(\underline{\widetilde{\mathbb{U}}}_2\right) = \begin{pmatrix} \underline{\mathbf{Z}} & \underline{\mathbf{W}} \\ \underline{\mathbf{W}}^{\dagger} & \underline{\mathbf{Z}}^{*} \end{pmatrix} \tag{12.64}$$

The covariance and pseudo-covariance matrices of $\underline{\widetilde{\mathbb{U}}_2}^{\dagger} \mathbf{W}_2 \underline{\mathbb{H}}_2$ are then given by:

$$\mathrm{cov}\left(\underline{\widetilde{\mathbb{U}}}_2^{\dagger} \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2\right) = \underline{\mathbb{H}}_2^{\dagger} \left(\underline{\mathbf{W}}_2{}^{-1}\right)^{\dagger} \mathrm{cov}\left(\underline{\widetilde{\mathbb{U}}}_2\right) \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2 \tag{12.65}$$

$$\mathrm{pcov}\left(\underline{\widetilde{\mathbb{U}}}_2^{\dagger} \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2\right) = \underline{\mathbb{H}}_2^{T} \left(\underline{\mathbf{W}}_2{}^{-1}\right)^{T} \mathrm{pcov}\left(\underline{\widetilde{\mathbb{U}}}_2\right) \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2 \tag{12.66}$$

And using Equations B.16, I can extract the covariance matrix of $\mathrm{Re}\left(\underline{\mathbb{U}}_2^{\dagger} \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2\right)$:

$$\mathrm{cov}\left(\mathrm{Re}\left(\underline{\mathbb{U}}_2^{\dagger} \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2\right)\right) = \frac{1}{2}\,\mathrm{Re}\left(\mathrm{cov}\left(\underline{\mathbb{U}}_2^{\dagger} \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2\right) + \mathrm{pcov}\left(\underline{\mathbb{U}}_2^{\dagger} \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2\right)\right) \tag{12.67}$$

Being back to real variables, the final covariance matrix on $\widehat{\mathbf{C}}$ is given by:

$$\mathrm{cov}\left(\widehat{\mathbf{C}}\right) = \left[\mathrm{Re}\left(\underline{\mathbb{H}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbb{H}}_2\right)\right]^{-1} \left[\mathrm{cov}\left(\mathrm{Re}\left(\underline{\mathbb{U}}_2^\dagger \underline{\mathbf{W}}_2{}^{-1} \underline{\mathbb{H}}_2\right)\right)\right] \left[\mathrm{Re}\left(\underline{\mathbb{H}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbb{H}}_2\right)^T\right]^{-1} \quad (12.68)$$

Equation 12.65 to 12.68 show how the final error bars on the planet spectrum can be derived from the covariance matrices of $\underline{\widetilde{\mathbb{U}}}$, and Equation 12.61 and 12.62 relate these matrices to the covariance matrices of $\underline{\widetilde{\mathbb{V}}}$. Noting that:

$$\underline{\widetilde{\mathbb{V}}} = \underline{\mathbf{\Omega}}^* \underline{\mathbb{V}} \quad (12.69)$$

Where:

$$\underline{\mathbf{\Omega}} = \begin{pmatrix} \Phi_{b_1,t_1} & 0 & \dots & 0 \\ 0 & \Phi_{b_2,t_1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \Phi_{b_{n_b},t_{n_t}} \end{pmatrix} \quad (12.70)$$

In terms of errors, Equation 12.69 translates to:

$$\mathrm{cov}\left(\underline{\widetilde{\mathbb{V}}}\right) = \underline{\mathbf{\Omega}}\, \mathrm{cov}\left(\underline{\mathbb{V}}\right) \underline{\mathbf{\Omega}}^\dagger \quad (12.71)$$

$$\mathrm{pcov}\left(\underline{\widetilde{\mathbb{V}}}\right) = \underline{\mathbf{\Omega}}\, \mathrm{pcov}\left(\underline{\mathbb{V}}\right) \underline{\mathbf{\Omega}}^T \quad (12.72)$$

Which is the last equation needed in order to estimate the final error bars on the spectrum $\mathbf{C}$ from the errors on the elementary visibilities $\underline{\mathbf{V}}_{b,t}$, since the covariance and pseudo-covariance matrices of $\underline{\mathbb{V}}$ have already been given in Section 12.3.

12.5.7 Wrapping up

12.5.7.1 How do I calculate $\underline{\mathbf{P}}^{\perp}_{\underline{\mathbb{C}}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}}$?

First of all, from the data, the best approximation of $\underline{\mathbf{V}}_{\text{on star}}$ available is $\underline{\mathbf{V}}_{\text{refstar}}$. Now, since the projector is supposed to correspond to a time t at which the planet is observed, there is no corresponding $\underline{\mathbf{V}}_{\text{refstar}}$. The star is not observed simultaneously to the planet. Thus, it is necessary to further approximate $\underline{\mathbf{V}}_{\text{on star},b,t}$ by using either the closest (in time) $\underline{\mathbf{V}}_{\text{refstar},b,t}$ available, or a mean of several of them, or another suitable combination (see Section 13.2.2 for a similar problem when referencing the visibilities).

Now, by definition, a projector orthogonal to a subspace S is obtained by subtracting a projector onto that said subspace from the identity matrix:

$$\underline{\mathbf{P}}^{\perp}_{S} = \mathbf{I} - \underline{\mathbf{P}}_{S}$$

And a projector onto the subspace generated by the family of $\Lambda^k \underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}$ can be obtained by simply projecting each of the individual vectors of the natural basis on the subspace generated by the $\Lambda^k \underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}$:

$$\underline{\mathbf{P}}_{\underline{\mathbb{C}}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}} = \begin{pmatrix} \underline{\mathbf{C}}_1 & \underline{\mathbf{C}}_2 & \dots & \underline{\mathbf{C}}_{n_\lambda} \end{pmatrix} \quad (12.73)$$

Where the $\underline{\mathbf{C}}_k$ are column vectors defined by:

$$\underline{\mathbf{C}}_k = \underline{\mathbf{P}}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \text{with a single 1 in the } k\text{th position} \tag{12.74}$$

And since the scalar product used here is the natural scalar product, the projection of this vector $\mathbf{X}_j$ (with only a single 1 at the j-th position) is the vector colinear to the $\Lambda^k\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}$, which also minimizes over $\underline{\mathbf{Y}}$ the quantity:

$$\|\mathbf{X}_j - \underline{\mathbf{Y}}\| \tag{12.75}$$

This means that the vector $\underline{\mathbf{C}}_k$ that we are looking for can be found by solving for $\underline{\mathbf{A}}$, in the least-square sense, the following linear problem:

$$\mathbf{X}_k = \left(\Lambda^0\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t} \quad \dots \quad \Lambda^m\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}\right)\underline{\mathbf{A}} \tag{12.76}$$

The solution is:

$$\hat{\underline{\mathbf{A}}} = \left(\underline{\mathbf{M}}^T\underline{\mathbf{M}}\right)^{-1}\underline{\mathbf{M}}^T\,\mathbf{X}_k \tag{12.77}$$

With $\underline{\mathbf{M}}$ defined as the concatenation of all the $\Lambda^k\underline{\widetilde{\mathbf{V}}}_{\text{on star}}$ in columns:

$$\underline{\mathbf{M}} = \left(\Lambda^0\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t} \quad \dots \quad \Lambda^m\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}\right) \tag{12.78}$$

And the vector $\underline{\mathbf{C}}_k$ is finally:

$$\underline{\mathbf{C}}_k = \underline{\mathbf{M}}^T\left(\underline{\mathbf{M}}^T\underline{\mathbf{M}}\right)^{-1}\underline{\mathbf{M}}^T\,\mathbf{X}_k \tag{12.79}$$

And since this is valid for all $0 \leq k \leq n_\lambda$, this means that:

$$\underline{\mathbf{P}}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}} = \underline{\mathbf{M}}^T\left(\underline{\mathbf{M}}^T\underline{\mathbf{M}}\right)^{-1}\underline{\mathbf{M}}^T \tag{12.80}$$

And finally:

$$\underline{\mathbf{P}}^{\perp}_{b,t} = \mathbf{I} - \underline{\mathbf{M}}^T\left(\underline{\mathbf{M}}^T\underline{\mathbf{M}}\right)^{-1}\underline{\mathbf{M}}^T \tag{12.81}$$

with $\underline{\mathbf{M}}$ defined in Equation 12.78.

As a final remark on the calculation of this matrix $\underline{\mathbf{P}}^{\perp}_{b,t}$, it can be noted that since the $\Lambda^k\underline{\widetilde{\mathbf{V}}}_{b,t}$ are related to the $\Lambda^k\underline{\mathbf{V}}_{b,t}$ by a change of basis:

$$\Lambda^k\underline{\widetilde{\mathbf{V}}}_{b,t} = \Phi^*\Lambda^k\underline{\mathbf{V}}_{b,t} \tag{12.82}$$

the projector $\underline{\mathbf{P}}^{\perp}_{b,t}$ can also be calculated in the space of the non-tilded vectors $\Lambda^k\underline{\mathbf{V}}_{\text{on star},b,t}$, and transformed by the appropriate change of basis afterwards:

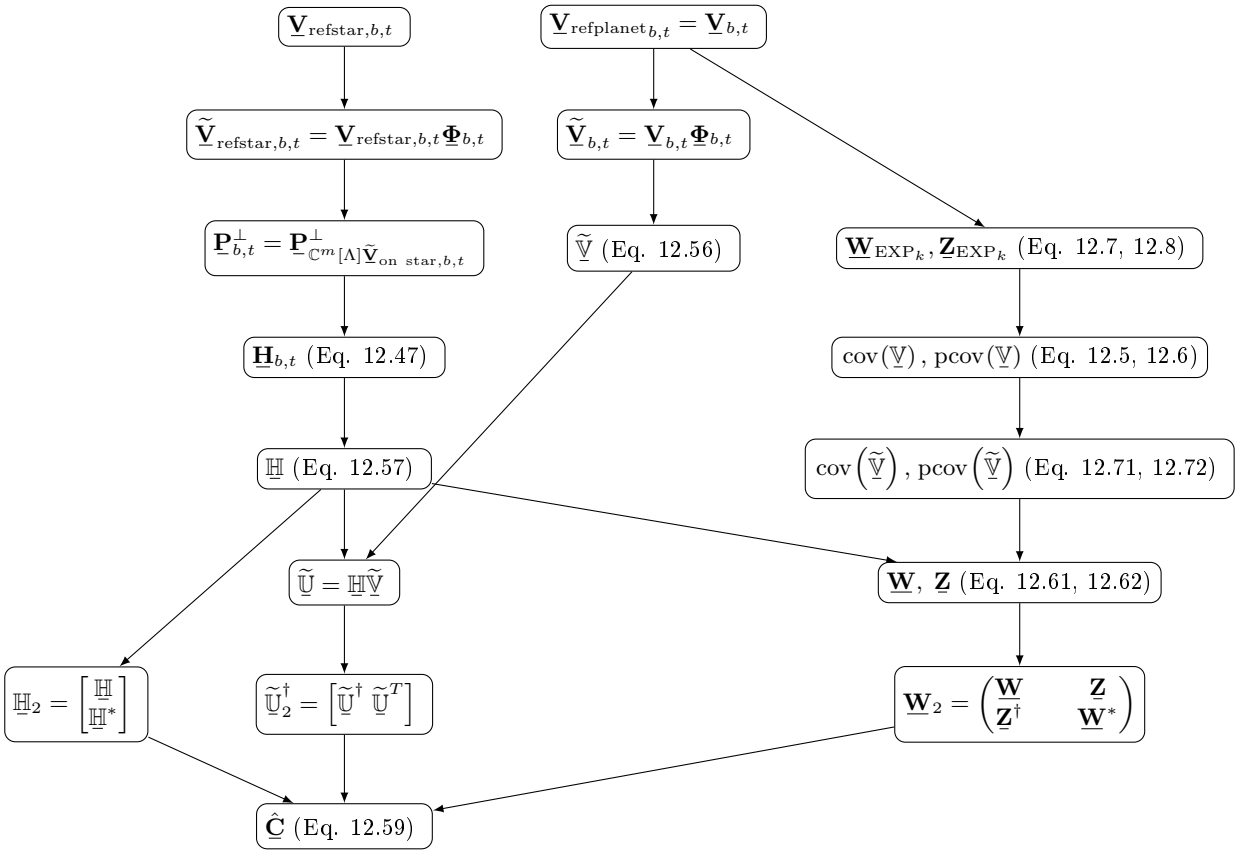
$$\underline{\mathbf{P}}^{\perp}_{\mathbb{C}^m[\Lambda]\underline{\widetilde{\mathbf{V}}}_{\text{on star},b,t}} = \Phi\left[\underline{\mathbf{P}}^{\perp}_{\mathbb{C}^m[\Lambda]\underline{\mathbf{V}}_{\text{on star},b,t}}\right]\Phi^* \tag{12.83}$$

12.5.7.2 How do I calculate $\underline{\mathbf{H}}_{b,t}$?

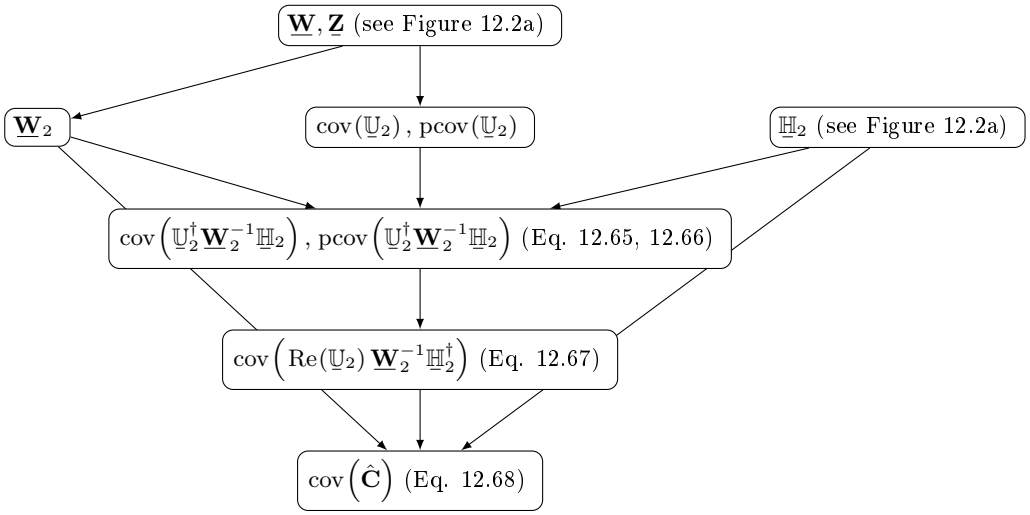
The matrix $\underline{\mathbf{H}}_{b,t}$, used to restrict the estimation problem to a subspace where the covariance matrices are invertible, is defined by Equation 12.47. This matrix corresponds to a diagonalization of $\underline{\mathbf{P}}^{\perp}_{b,t}$ in an orthonormal basis. It can be calculated, for example, by executing a singular value decomposition of $\underline{\mathbf{P}}^{\perp}_{b,t}$.

12.5.7.3 Flowcharts

The "recipe" to extract the spectrum can be summarized in two flowcharts, given in Figure 12.2.



(a) Flowchart for the calculation of $\widehat{\mathbf{C}}$



(b) Flowchart for the calculation of $\text{cov}\left(\widehat{\mathbf{C}}\right)$

Figure 12.2: Flowcharts showing the progressive calculation of the different matrices used to extract the maximum likelihood estimate of the planet spectrum $\widehat{\mathbf{C}}$ (upper panel), and its covariance matrix $\text{pcov}\left(\widehat{\mathbf{C}}\right)$ (lower panel).

Chapter 13

Beta Pictoris b seen with VLTI/GRAVITY

Content of this chapter

Beta Pictoris	
Magnitude K (science)	3.48
Magnitude V (AO)	3.86
Beta Pic b	
Magnitude K	~ 13
Separation (Sep. 2018)	140 mas

Table 13.1: Magnitude and separation of β Pictoris b when observed with VLTI/GRAVITY in September 2018.

13.1 Observations

The observations of β Pictoris b presented in this Chapter were made on Director's Discretionary Time, on September, 22, 2018, after the successful detection of HR 8799 e at the end of August 2018 (see Gravity Collaboration et al. [2019]). At the time, the planet was close to its host star (separation $\simeq 140$ mas), and this observing program had two main objectives:

- Obtain a precise post-conjunction astrometry.
- Obtain a high-quality medium resolution spectrum of the planet in K-band to investigate its atmospheric properties.

The observing strategy described in Chapter 12 of this thesis was followed. The planet β Pic b and its host star β Pic were observed alternatively. The main characteristics of the two targets are given in Table 13.1, and the complete observing log is given in Table 13.2.

13.2 Data reduction

The data reduction used for the β Pic b observations follows the method described in Chapter 12 of this thesis.

The initial reduction of the data was made using the official GRAVITY pipeline, in version 1.12. I explicitly requested the "astroreduced" intermediate data products, and discarded the final "scivis" files. All subsequent steps of the data reduction are performed on the astroreduced files.

13.2.1 Extracting and correcting the visibilities

In the astroreduced files, the raw complex visibilities can be extracted from the `OI_VIS`, in the field `VISDATA` (I will use the notation `OI_VIS->VISDATA` to indicate the origin in the astroreduced products of the different quantities I use). It is convenient to reshape `VISDATA` in an array of size $(n_{\rm DIT}, n_b, n_\lambda)$. These visibilities are not corrected for the fringe-tracker zero-point, dispersion, and metrology. This needs to be done manually. To do so, some data must be retrieved from the `OI_VIS` and `OI_FLUX`:

`PHASE_REF` = `OI_VIS->PHASE_REF`

`OPD_DISP` = `OI_VIS->OPD_DISP`

`TELFC_MCORR` = `OI_FLUX->OPD_MET_TELFC_MCORR`

`FC_CORR` = `OI_FLUX->OPD_MET_FC_CORR`

Target	UT Time (UTC)	DIT (s)	NDIT	Seeing (")	τ_0 (ms)	Airmass	Parallactic angle (deg)
PLANET	07:37:40	30.0	10	0.9	4.7	1.33	-66.4
STAR	07:43:55	0.3	50	0.8	4.7	1.31	-64.7
PLANET	07:44:55	30.0	10	0.8	4.7	1.31	-64.5
SKY	07:50:30	30.0	10	0.6	7.6	1.30	-62.9
STAR	07:56:25	0.3	50	0.5	9.7	1.28	-61.2
SKY	07:57:14	0.3	50	0.5	9.7	1.28	-61.0
PLANET	07:58:30	30.0	10	0.4	9.5	1.28	-60.6
STAR	08:04:15	0.3	50	0.5	10.4	1.26	-58.9
PLANET	08:05:12	30.0	10	0.5	8.1	1.26	-58.6
STAR	08:10:53	0.3	50	0.5	8.5	1.25	-56.9
PLANET	08:11:50	30.0	10	0.5	8.7	1.25	-56.6
STAR	08:17:34	0.3	50	1.0	5.4	1.24	-54.8
SKY	08:18:22	0.3	50	0.8	6.3	1.23	-54.5
PLANET	08:19:22	30.0	10	0.9	5.9	1.23	-54.2
SKY	08:24:56	30.0	10	0.9	5.3	1.22	-52.4
STAR	08:30:45	0.3	50	0.9	5.2	1.21	-50.4
PLANET	08:31:40	30.0	10	0.9	5.2	1.21	-50.1
STAR	08:37:21	0.3	50	0.7	7.4	1.20	-48.1
PLANET	08:38:31	10.0	30	0.8	5.9	1.20	-47.7
STAR	08:44:12	0.3	50	0.9	5.5	1.19	-45.7
PLANET	08:45:07	10.0	30	0.7	7.2	1.19	-45.3
SKY	08:50:41	10.0	30	0.7	7.6	1.18	-43.2
STAR	08:56:32	0.3	50	0.7	8.2	1.17	-41.0
SKY	08:57:21	0.3	50	0.7	8.2	1.17	-40.7
PLANET	08:58:22	10.0	30	0.7	7.4	1.17	-40.3
STAR	09:04:03	0.3	50	1.0	6.1	1.16	-38.0
PLANET	09:05:01	10.0	30	1.0	6.1	1.16	-37.6
STAR	09:10:47	0.3	50	0.9	6.6	1.15	-35.2
PLANET	09:11:46	10.0	30	1.2	6.6	1.15	-34.8
STAR	09:17:30	0.3	50	0.9	6.4	1.15	-32.4
SKY	09:18:21	0.3	50	0.9	6.4	1.15	-32.0
PLANET	09:19:26	10.0	30	0.9	5.9	1.15	-31.6
SKY	09:25:03	10.0	30	0.8	6.0	1.14	-29.1
STAR	09:30:53	0.3	50	0.8	6.0	1.14	-26.5
PLANET	09:31:48	10.0	30	0.8	6.7	1.14	-26.0
STAR	09:37:30	0.3	50	0.7	8.4	1.13	-23.4
PLANET	09:38:26	10.0	30	0.8	6.8	1.13	-23.0
STAR	09:44:10	0.3	50	0.9	6.0	1.13	-20.2
PLANET	09:45:05	10.0	30	0.6	6.0	1.13	-19.8
STAR	09:50:51	0.3	50	0.6	6.0	1.12	-17.0
PLANET	09:51:49	10.0	30	0.6	6.0	1.12	-16.6
STAR	09:57:33	0.3	50	0.7	6.0	1.12	-13.7
STAR	09:58:31	0.3	50	0.7	6.0	1.12	-13.2
SKY	09:59:20	0.3	50	0.8	6.0	1.12	-12.8

Table 13.2: Observing log for the DDT β Pic b program, carried out on September, 22, 2018.

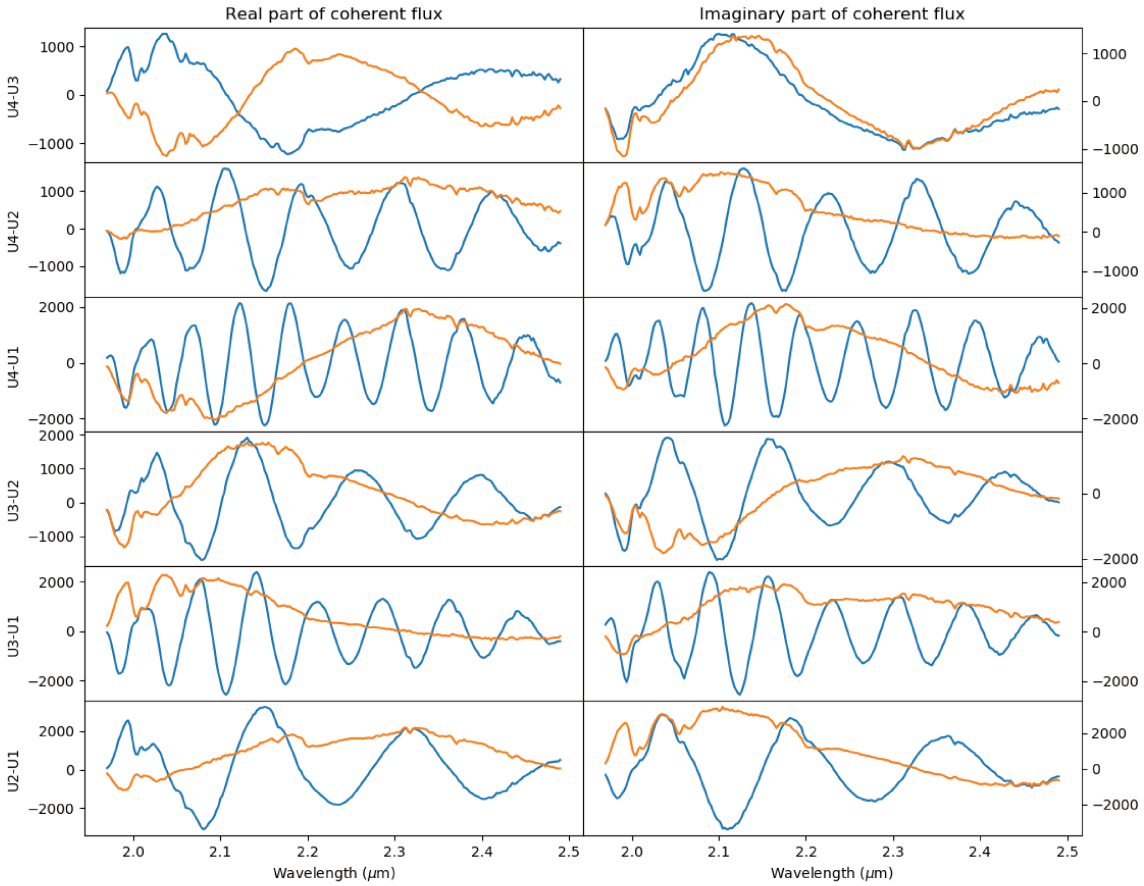


Figure 13.1: An example of complex visibilities acquired on β Pictoris b. The data shown here corresponds to the first observation on the planet, at 07:37:40 UT. The visibilities have been binned on all DITs, and are plotted in real part (first column) and imaginary part (second column). Each row represent a different baseline. The blue line shows the raw visibility before all corrections, and the orange line shows the result after dispersion, fringe-tracker zero point, and metrology correction. The visibilities are still dominated by the remaining stellar flux, which is why no planet fringes can be seen.

The OPD dispersion and fringe-tracker zero point correction can be done using the following:

$$\texttt{VISDATA} = \texttt{VISDATA} \times \exp\left(-i\frac{2\pi}{\lambda}\left[\texttt{OPD\_DISP} - \texttt{PHASE\_REF}\right]\right)$$

For the metrology correction, the two fields extracted from `OI_FLUX` represent a telescope based correction in OPD. For each baseline b_k linking telescopes T_{k_1} to T_{k_2}, the corresponding `VISDATA` must be corrected using:

$$\begin{aligned}\texttt{VISDATA} = \texttt{VISDATA}_{b_k} \\ \times \exp\left(-i\frac{2\pi}{\lambda}\left[\texttt{TELFC}_\texttt{M}\texttt{CORR}_{T_{k_2}} + \texttt{FC}_\texttt{M}\texttt{CORR}_{T_{k_2}} - \texttt{TELFC}_\texttt{M}\texttt{CORR}_{T_{k_1}} - \texttt{FC}_\texttt{M}\texttt{CORR}_{T_{k_1}}\right]\right)\end{aligned}$$

Figure 13.1 illustrates the impact of these corrections on the data.

13.2.2 Referencing the visibilities

When corrected visibilities are available for each exposure, it is possible to reference the visibilities of the planet to the star. This step aims at correcting for the unknown zero-point of the metrology (see Section 11.3.3.3). To do so, it is necessary to subtract the phase of the visibilities obtained on the star to the phase of the visibilities obtained on the planet:

$$\texttt{VISREF}_{\text{planet}} = \texttt{VISDATA}_{\text{planet}} \times \exp\left(-i \arg\left[\texttt{VISDATA}_{\text{star}}\right]\right)$$

The difficulty in this step is to decide which stellar exposure to use to calculate the $\texttt{VISDATA}_{\text{star}}$ term. Depending on the variability observed in the data, several choices are possible:

- Combine all stellar exposure files by calculating the complex mean, and use the argument of this "master stellar exposure" to reference each planet exposure. This method should only be used when there is very little variability in the data, to minimize detector and photon noise.
- If a strategy similar to the one suggested here is used, for each planet exposure there is a stellar exposure obtained just before, and a stellar exposure obtained just after. The mean of the two can be used to calculate the reference. This is the method I used for the observations presented here.
- If the variability is extremely high, it may be necessary to interpolate between the different stellar exposures. One simple possibility is to use different weights for the stellar exposures acquired before and after the planet exposure, depending on the exact time at which each DIT has been acquired.

An example of the impact of the phase referencing on actual data is given in Figure 13.2.

After this step, the resulting visibilities correspond to what was denoted $\underline{\mathbf{V}}_{\text{refplanet}}(b,t,\lambda)$ in Chapter 12:

$$\underline{\mathbf{V}}_{\text{refplanet}} = \texttt{VISREF}_{\text{planet}} \tag{13.1}$$

For the star, the situation is easier, as "referencing the star" simply means setting its phase to zero. In other words:

$$\underline{\mathbf{V}}_{\text{refstar}} = \left|\texttt{VISDATA}_{\text{star}}\right| \tag{13.2}$$

13.2.3 Extracting the astrometry

The method used to extract the astrometry in the β Pic b data follows the general method described in Section 12.4. For each exposure file, a model including a set of polynomials to take into account variability in the speckle noise, a set of parameters to take into account the variable injection factor, and the sky separation $\Delta\alpha, \Delta\delta$ is fitted to the data. The fit is linear in all parameters except the two sky coordinates. Creating a χ^2 (or rather, a "minus log-likelihood") map in $\Delta\alpha, \Delta\delta$ is straightforward, from Equation 12.26.

To illustrate the need for a polynomial to properly represent the leaking stellar flux, I show in Figure 13.3 an example of the best fits obtained with different orders of this polynomial.

In Figure 13.4, I show an example of an exposure in which I have subtracted the stellar part of the best fit to the data, to show the residual planet fringes and the way they are fitted by the model.

The log-likelihood, or equivalently the χ^2, is a quantity which can be summed. Log-likelihood maps can be created DIT by DIT, baseline per baseline, and added together to

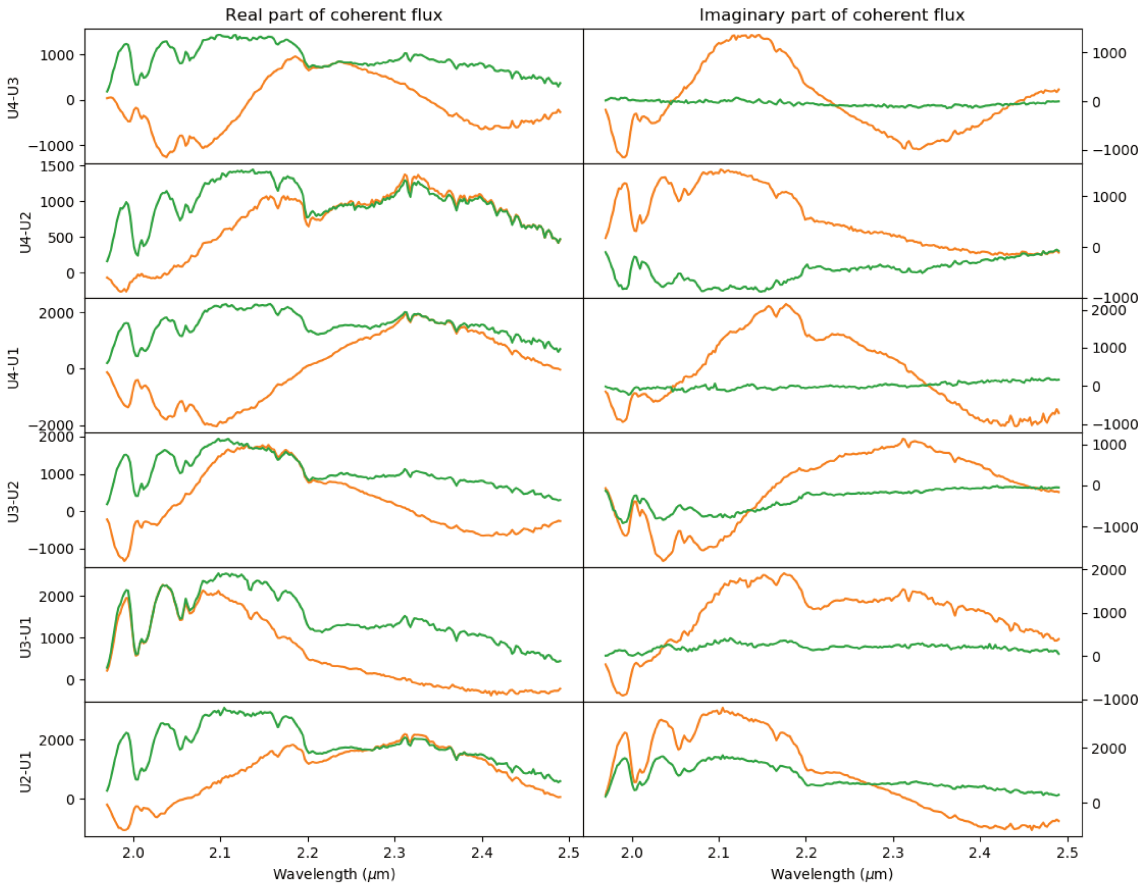


Figure 13.2: Same as Figure 13.1, but showing the visibilities before and after the phase referencing step. The orange line corresponds to the corrected visibilities (orange line in Figure 13.1), and the green line shows the complex visibilities referenced to the star. In the particular case of this figure, all stellar exposures have been combined to calculate the phase reference.

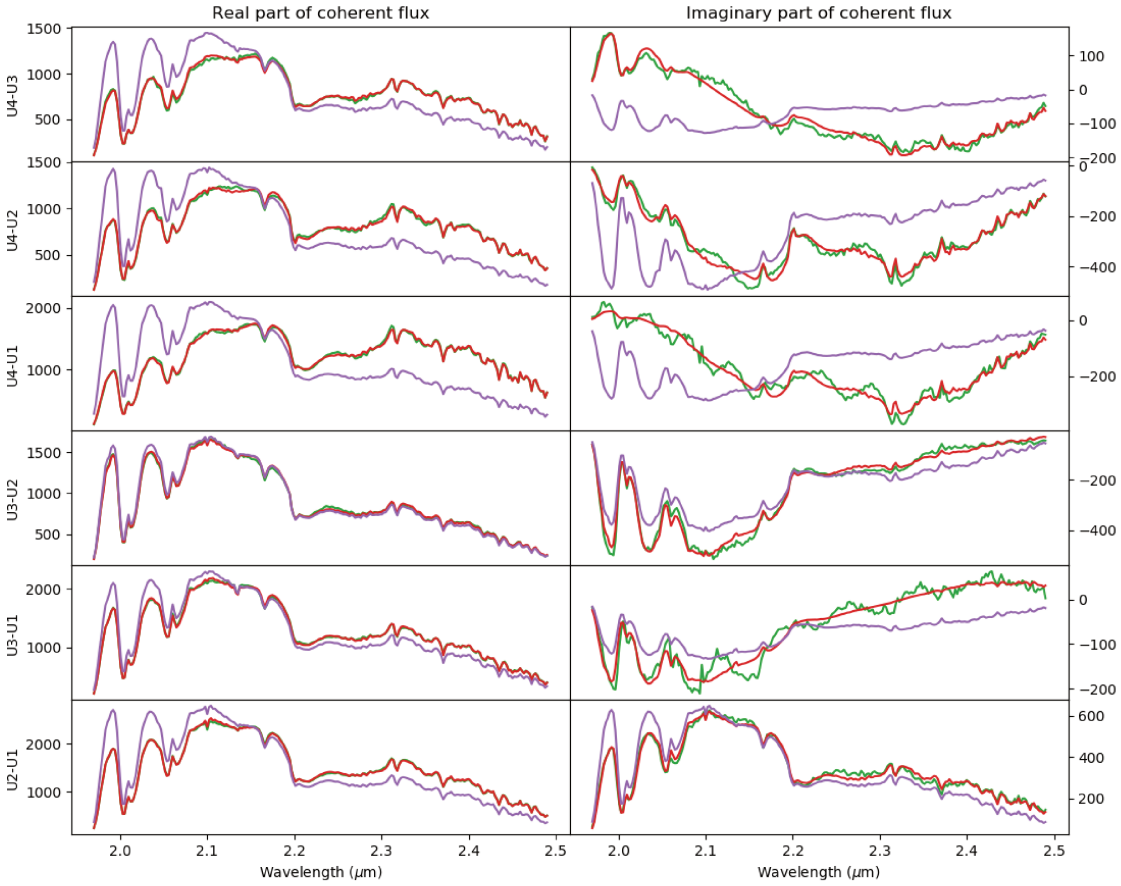


Figure 13.3: Complex visibilities (corrected and referenced to the star) for the first DIT of the 07:37:40 UT OB, and best fits obtained with different orders for the stellar flux polynomial (denoted m in Section 12.4). The green line corresponds to the data. The purple line corresponds to a fit using a first order polynomial. The red line corresponds to a 4th order polynomial.

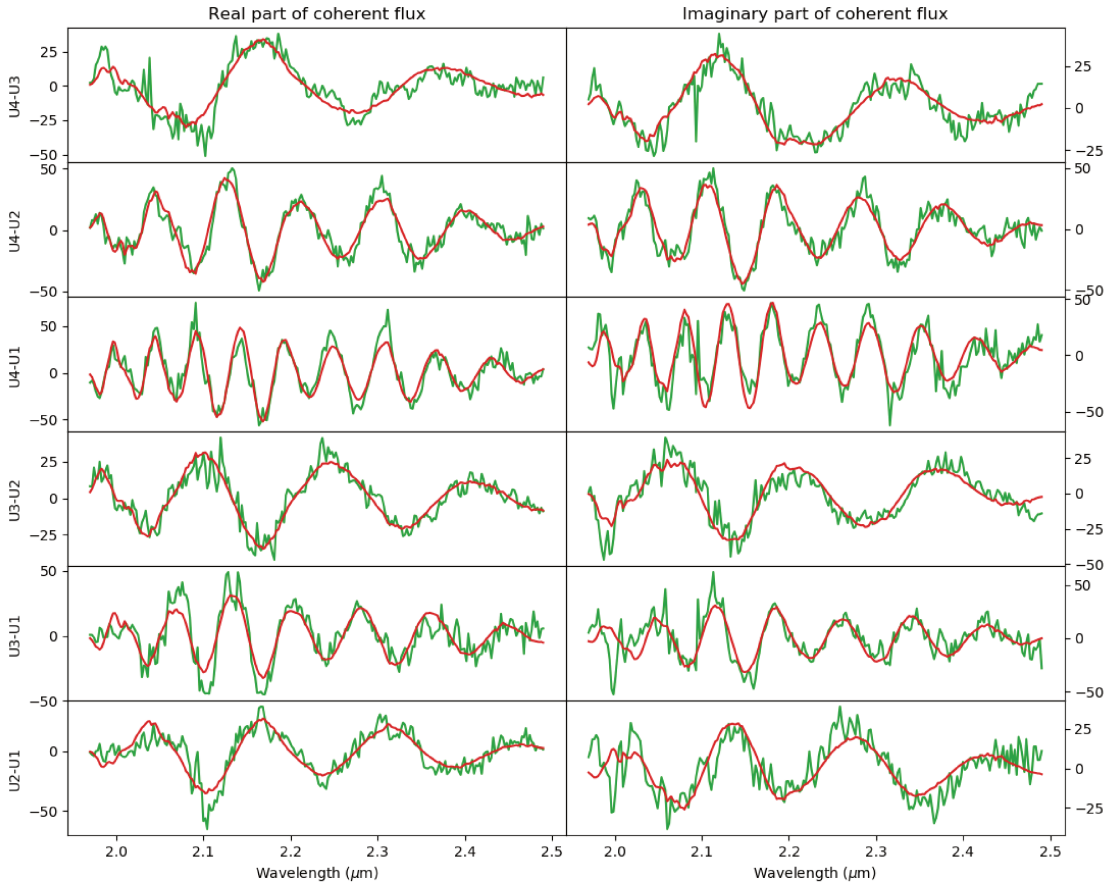


Figure 13.4: Complex visibilities (corrected and referenced to the star) for the 07:37:40 UT planet OB, binned over all DITs, and for which the stellar component of the best fit model (of order 4) has been subtracted. The green line corresponds to the data, and the red line is the best fit obtained when extracting the astrometry. In this plot, and using the notation of Chapter 12, the green line corresponds to the projected "on planet" visibility, i.e. to a vector $\underline{\underline{\mathbf{P}}}^{\perp}_{b,t}\underline{\mathbf{V}}_{\text{on planet},b,t}$.

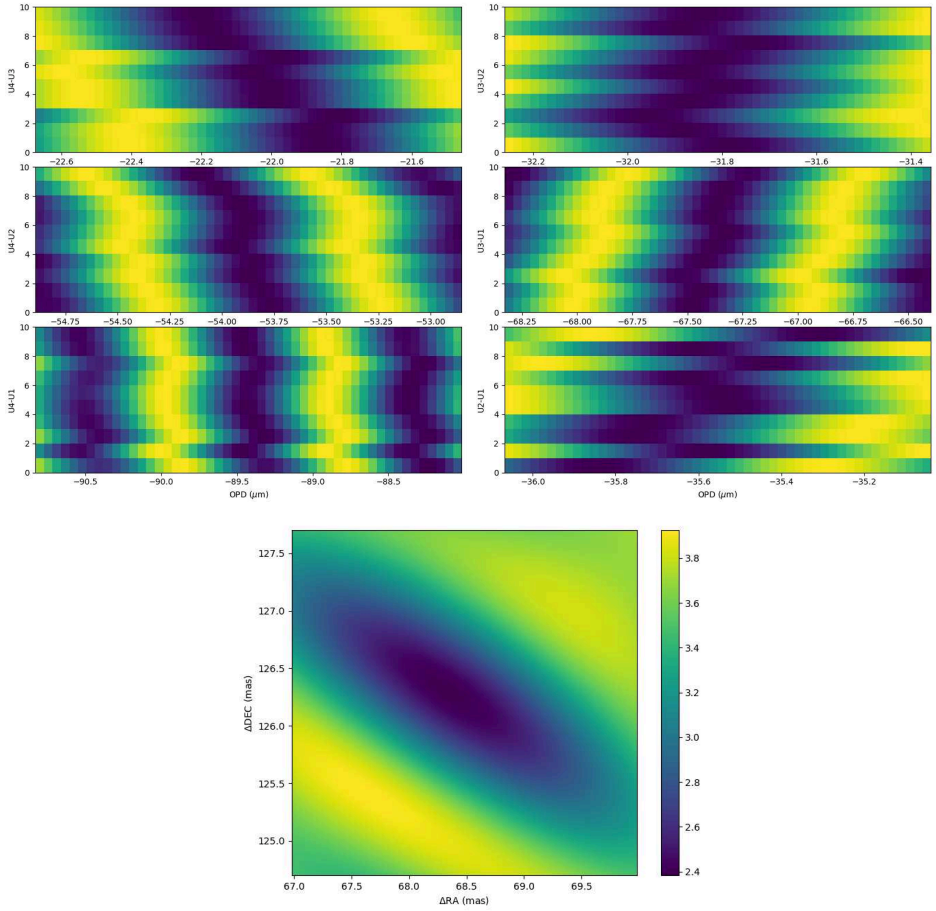


Figure 13.5: Top two panels: χ^2 maps obtained in OPD space for the 07:37:40 UT planet exposure. For each baseline, the image shows the OPD maps (OPD on x-axis) for each DIT (DIT value on y-axis). The color scale is arbitrary, and each row of each image has been normalized to the same value. The drift of the best OPD due to the sky rotation is obvious on these plots.
Bottom panel: Reduced χ^2 map in right-ascension and declination calculated from the OPD maps shown in the top panel.

create a master $-\log\mathcal{L}$ map for the complete observation. For a given baseline and a given DIT, the two sky coordinates only intervene in the form of an OPD (something like $\Delta\alpha\times U+\Delta\delta\times V$). Thus, if done DIT by DIT and baseline per baseline, the model-fitting can be done in OPD space. With only a single non-linear dimension, this reduces computing time. Then, when creating the final $-\log\mathcal{L}$ map, for each point in α,δ space, the corresponding OPD can be calculated for each baseline and each DIT, and the log-likelihood values retrieved from the OPD log-likelihood maps. This is a variant on the general method described in Section 12.4, which improves computing time.

In Figure 13.5, I give an example of the OPD χ^2 maps obtained for one exposure file, and of the corresponding (α,δ) χ^2-map calculated *a posteriori*.

13.2.4 Extracting the spectrum

For extracting the spectrum, the method used is an exact implementation of the mathematics described in Section 12.5, following the flowcharts given in Figures 12.2a and 12.2b.

The end-product extracted from GRAVITY observations is a *contrast* spectrum $C(\lambda)$. To be converted to an absolute planet spectrum $S_{\mathrm{P}}(\lambda)$, it is necessary to multiply it by a spectrum of the star. I used a synthetic BT-NextGen spectrum [Hauschildt et al., 1999], with a temperature of 8000 K, a $\log(g/g_0)$ of 4, and a solar metalicity. I scaled this synthetic spectrum to an ESO K-band magnitude of 3.495, taking into account the correct filter [van der Bliek et al., 1996]. The contrast spectrum and the calibrated absolute spectrum are given in Figure 13.6. The error bars in this figure only reflect the diagonal terms of the full covariance matrix.

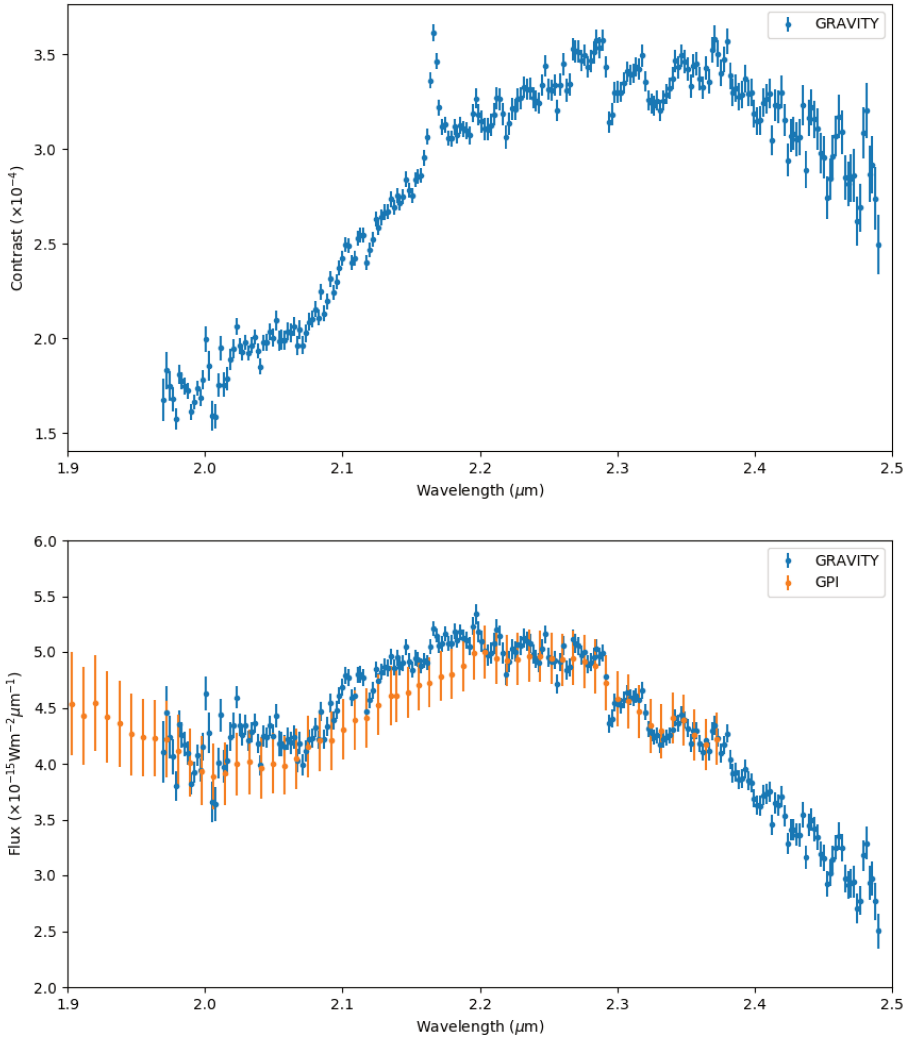


Figure 13.6: Top panel: K-band contrast spectrum of β Pictoris b relative to its host star, as extracted from the GRAVITY observations presented in this thesis. Bottom panel: Calibrated K-band spectrum of β Pictoris b, compared to GPI K-band spectrum from Chilcote et al. [2017]. Error bars on the GRAVITY spectra only reflect the diagonal terms of the complete covariance matrices.

Observation time	ΔRA	ΔDEC	$\sqrt{\text{Var}(\Delta\text{RA})}$	$\sqrt{\text{Var}(\Delta\text{DEC})}$	Covar (ΔRA, ΔDEC)
	mas	mas	mas	mas	mas^2
07:37:40	68.475	126.160	0.014	0.011	-0.00013
07:44:55	68.414	126.323	0.013	0.010	-0.00009
07:58:30	68.354	126.344	0.013	0.009	-0.00009
08:05:12	68.513	126.168	0.010	0.007	-0.00006
08:11:50	68.426	126.383	0.014	0.012	-0.00013
08:19:22	68.762	126.141	0.015	0.014	-0.00016
08:31:40	68.607	126.237	0.011	0.008	-0.00007
08:38:31	68.356	126.261	0.013	0.011	-0.00012
08:45:07	68.322	126.557	0.010	0.010	-0.00008
08:58:22	68.607	126.262	0.011	0.009	-0.00008
09:05:01	68.565	126.320	0.014	0.016	-0.00020
09:11:46	68.601	126.318	0.016	0.018	-0.00024
09:19:26	68.458	126.164	0.011	0.012	-0.00010
09:31:48	68.597	126.347	0.013	0.013	-0.00013
09:38:26	68.589	126.194	0.011	0.014	-0.00013
09:45:05	68.386	126.591	0.012	0.014	-0.00014
09:51:49	68.051	126.926	0.011	0.013	-0.00011
Global	68.48	126.34	0.04	0.05	-0.0014

Table 13.3: Astrometric position obtained for each exposure. Because the dispersion of all the exposures is much greater than the individual error bars, the global error bars are obtained by taking the variance and covariance of the series, rather than by combining the individual errors.

13.3 High-precision post-conjunction astrometry

13.3.1 Is GRAVITY broken?

The astrometric accuracy of GRAVITY fully benefits from the long baselines of the VLTI, and the typical astrometric precision is expected to be < 10 μas.

Yet, reducing the β Pictoris data exposure per exposure revealed a problem: the χ^2 valleys are indeed very narrow, leading to a high precision on the astrometry, but the position of the minimum moves from one exposure to the other. The complete results for the β Pic b astrometry data reduction are compiled in Table 13.3. The uncertainties of the right-ascension and declination depends on the orientation of the target with respect to the baselines of the instruments, and the two coordinates are not independent. Three numbers are thus reported for the error bars: the variance on RA and DEC, and the covariance between the two.

Table 13.3 clearly shows that something is wrong in the GRAVITY astrometry: the dispersion of the measurements obtained for each exposure is too large compared to the uncertainty. This is a textbook example of a *precise* but *inaccurate* measurement.

The root cause of this problem is still unclear, but a credible hypothesis is that the astrometry is affected by field-dependent phase errors, caused by a combination of fiber injection, optical aberrations, and sky rotation.

In any case, given the inaccuracy of the astrometric measurements, the best way to proceed is to calculate the error bars directly from the distribution of best positions found in each exposure. This reduces the final precision, which could be well below 10 μas without these phase errors, but this is the best solution available, for now.

13.3.2 Orbital parameters

The GRAVITY astrometry obtained on β Pic b was added to previous measurements, and a Monte-Carlo Markov-Chain algorithm was used to estimate the orbital parameters. This work was performed by Jason Wang (Caltech, USA), using a similar approach as in Wang et al. [2016]. The results are given in Figure 13.7, which clearly demonstrates the interest of high precision GRAVITY astrometry for constraining the orbit of β Pic b: with the new data point, all the orbital parameters are now well-constrained, with near-gaussian posterior distributions.

Contrary to previous study (e.g. Wang et al. [2016], Lagrange et al. [2019]) the results obtained when adding the GRAVITY astrometry point to an eccentric orbit, with $e \simeq 0.15 \pm 0.05$, and a slighly larger semi-major axis than expected: $a = 10.6 \pm 0.6$ AU. The epoch at perielion is now extremely well-constrained, as is the inclination of the orbit ($i = 89.04 \pm 0.03$ deg).

The β Pic b orbit fit can be further improved by adding Hipparcos/Gaia astrometry obtained on the star, as shown in Figure 13.8. The orbit fit with the Hipparcos/Gaia astrometric points is also sensitive to the mass of the planet. Figure 13.9 gives the distribution obtained for the mass of the planet. The value obtained is $M_{\rm planet} = 12.5 \pm 2\,M_{\rm Jup}$, in agreement with the work of Snellen and Brown [2018].

13.4 Atmospheric properties

Table 1.2 of Chapter 1 gives a list of temperature, surface gravity, radius, and mass estimates for the giant planet β Pictoris b obtained using atmospheric or planetary evolution models. Similar estimates can be extracted from the GRAVITY spectrum.

13.4.1 Fitting the spectrum with the Exo-REM model grid

The first approach I followed to extract estimates of bulk parameters is similar to what Chilcote et al. [2017] and others have done so far. I first used an Exo-REM atmospheric model grid from Charnay et al. [2018], at Solar metallicity, and parameterized in temperature and surface gravity, to calculate a χ^2 map, using:

$$\chi^2(g,T) = \left(S_{\lambda,\rm data} - S_{\lambda,\rm ExoREM}\right)^T W^{-1} \left(S_{\lambda,\rm data} - S_{\lambda,\rm ExoREM}\right) \tag{13.3}$$

Where S_λ is the spectrum (from the data, or from the model), and W is the covariance matrix on $S_{\lambda,\rm data}$.

The χ^2 map is given in Figure 13.10, and can be used to extract an estimate of the temperature and surface gravity.

The spectra from the model grid are given in W/cm/μm, and the GRAVITY K-band spectrum is calibrated in flux. Since the distance of β Pictoris is known ($D = 19.44$ pc), this makes possible to also estimate the radius of the planet. From the radius $R_{\rm P}$ and surface gravity $g_{\rm P}$, a model-dependent mass can be derived:

$$M_{\rm P} = \frac{G \times {R_{\rm P}}^2}{g_{\rm P}} \tag{13.4}$$

This mass can be compared to the estimate from Snellen and Brown [2018], who found a mass of $11 \pm 2\,M_{\rm Jup}$, or to the value reported in Figure 13.9 ($M = 12.5 \pm 2\,M_{\rm Jup}$). The $1\,\sigma$ and $3\,\sigma$ ranges are overplotted on top of the χ^2 map given in Figure 13.10. The best fit obtained without taking into account the mass of the planet corresponds to a temperature of

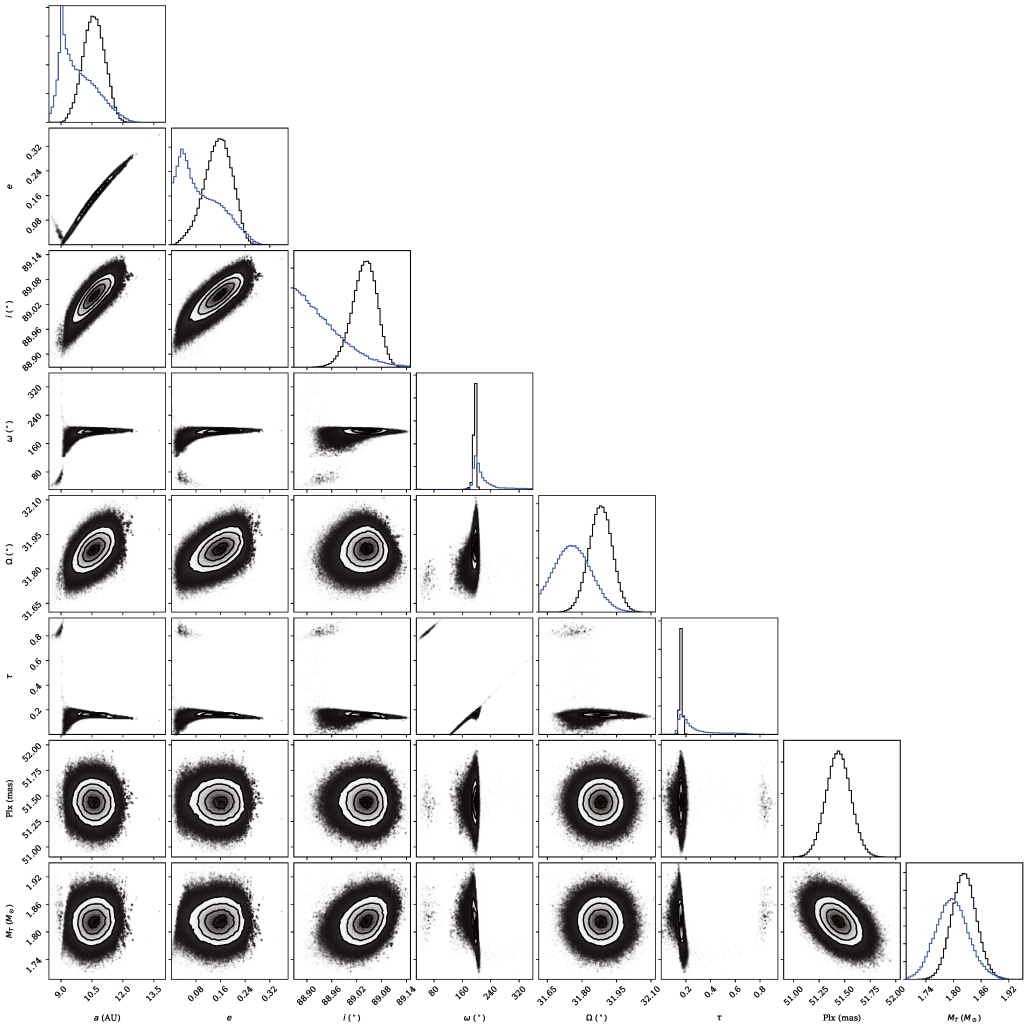


Figure 13.7: Posterior distributions for the Keplerian elements of β Pictoris b obtained with an MCMC approach. The upper row of each column gives the posterior distributions for each parameter obtained by Wang et al. [2016] in blue, and the new distributions obtained when adding the GRAVITY point in black. Analysis and Figure by J. Wang.

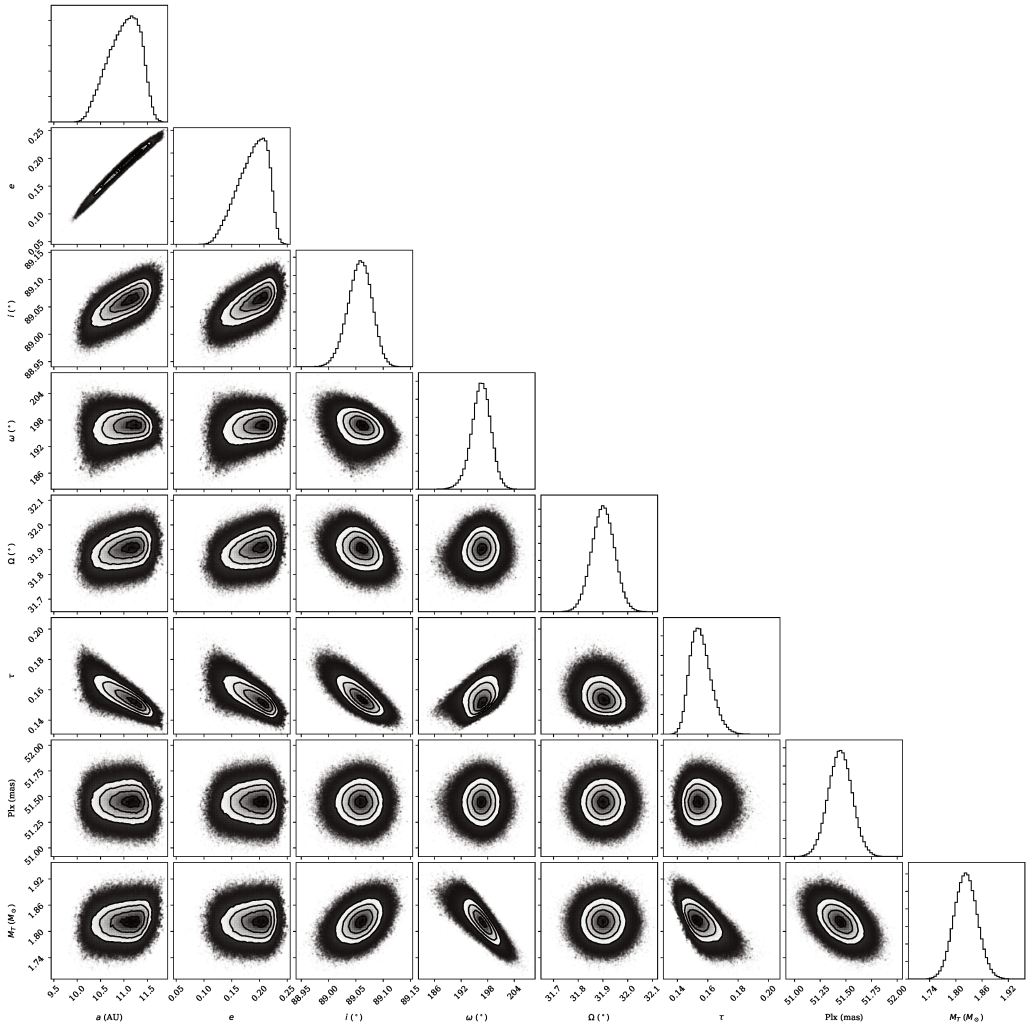


Figure 13.8: Posterior distributions for the Keplerian elements of β Pictoris b obtained with an MCMC approach, when including data from Wang et al. [2016], the new GRAVITY planet astrometry, and Hipparcos/Gaia star astrometry. Analysis and Figure by J. Wang.

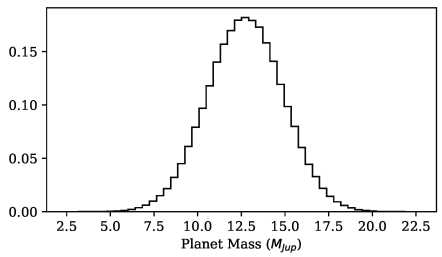


Figure 13.9: Posterior distributions for the mass of β Pic b. Figure by J. Wang.

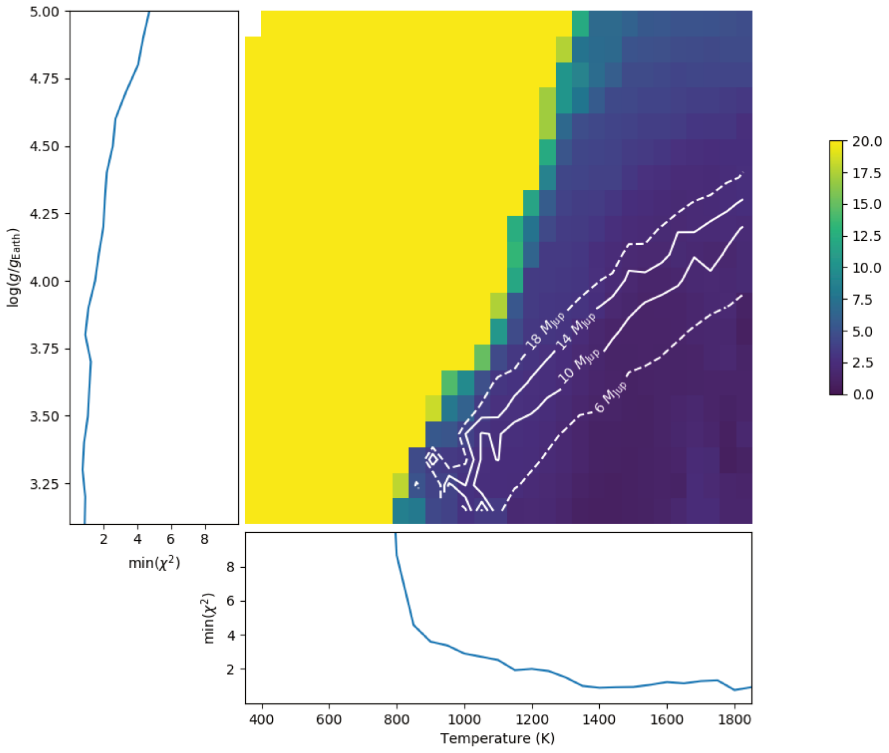


Figure 13.10: χ^2 map in surface gravity and temperature calculated using the Exo-REM grid of atmospheric models [Charnay et al., 2018]. The left and bottom panels give the χ^2 minimized over the surface gravity and temperature. The white plain and dashed lines delimit the $1\,\sigma$ and $3\,\sigma$ confidence intervals on the mass as measured when including Hipparcos/Gaia astrometry in the orbit fit (see Figure 13.9).

$T = 1750$ K, and a surface gravity of $\log(g/g_{\rm Earth}) = 3.3$. But this value corresponds to a mass of $1\,M_{\rm Jup}$, which is more than $5\,\sigma$ away from the value found with the orbit fit.

The mass measurement can be taken into account when determining which grid point corresponds to the best fit by adding a mass prior in the χ^2 map. In practice, this can be done by calculating a $\chi^2_{\rm mass}$ map using the following method:

- For each grid point of the ExoREM model, the best scaling factor to fit the GRAVITY data is extracted.
- This scaling factor is converted in a radius, and then in a corresponding planet mass $M(T, \log(g))$, using Equation 13.4.
- The $\chi^2_{\rm mass}$ value for this grid point is then set to $\left(M(T, \log(g)) - M_{\beta\ \rm Pic\ b}\right)^2 / {\sigma_M}^2$, where $M_{\beta\ \rm Pic\ b} = 12.5\,M_{\rm Jup}$ and $\sigma_M = 2\,M_{\rm Jup}$.

The $\chi^2_{\rm mass}$ map can then be combined to the initial χ^2 map. In theory, the combination of the two maps could be a simple addition: $\chi^2_{\rm tot} = \chi^2 + \chi^2_{\rm mass}$. But using such a combination effectively weights the mass measurement as much as a single GRAVITY data point. A

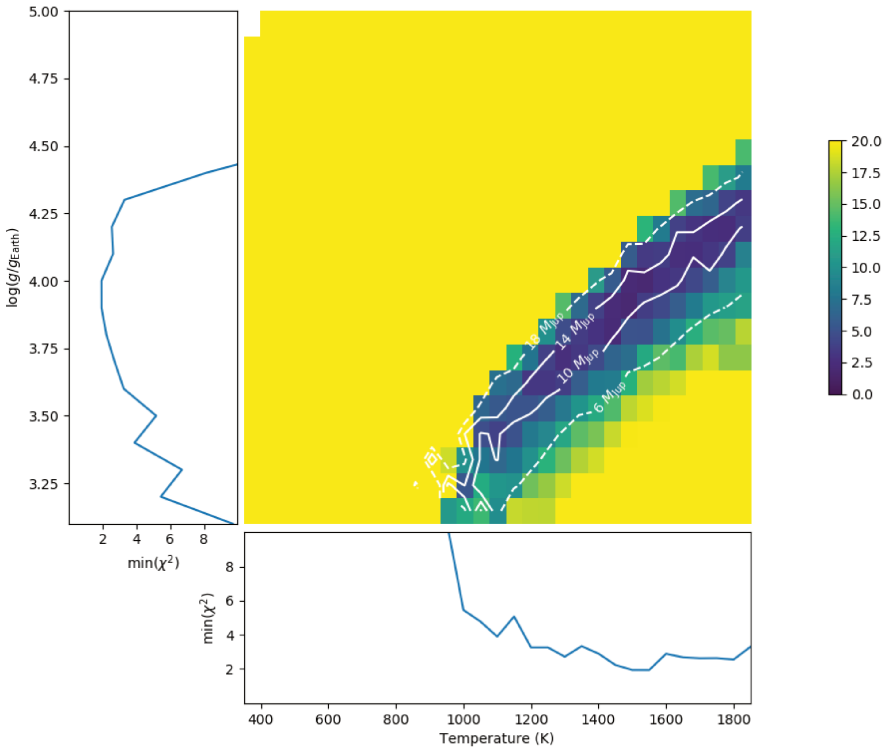


Figure 13.11: Same as Figure 13.10 with a prior on the planet mass to constrain the best fit to a mass compatible with the measurement by Snellen and Brown [2018] and with the value found in Section 13.3 of this work (see text for details)

.

better alternative is to consider that the mass measurement from Snellen and Brown [2018], or the one reported previously in this work, and the GRAVITY spectrum are two independant experiments, which can be weighted equally. In this case, the $\chi^2_{\rm mass}$ map needs to be scaled up by a factor n_λ, where n_λ is the number of data points in the GRAVITY spectrum.

Using this approach, the best fit is constrained to the region in which the mass is not significantly different from 11 to 12 $M_{\rm Jup}$. The total $\chi^2_{\rm tot}$ map obtained is given in Figure 13.11, and the best fit corresponds to a temperature of $T = 1550$ K, a surface gravity of $\log(g/g_{\rm Earth}) = 4.0$, and a planet mass of $M_{\rm Planet} = 13\,M_{\rm Jup}$.

Finally, the Exo-REM grid can also be parameterized in carbon to oxygen (C/O) ratio. A similar χ^2-based approach in temperature and C/O ratio, for a fixed surface gravity of $\log(g/g_{\rm Earth}) = 4.0$ (which corresponds to the best solution obtained with a mass prior), gives a best fit for a temperature of $T = 1500$ K and a C/O ratio of 0.16. The corresponding best fit is illustrated in Figure 13.12, in comparison with the Exo-REM spectrum obtained for the same temperature, metallicity, surface gravity, but at Solar C/O ratio.

The results of the different Exo-REM grid fits are summarized in Table 13.4.

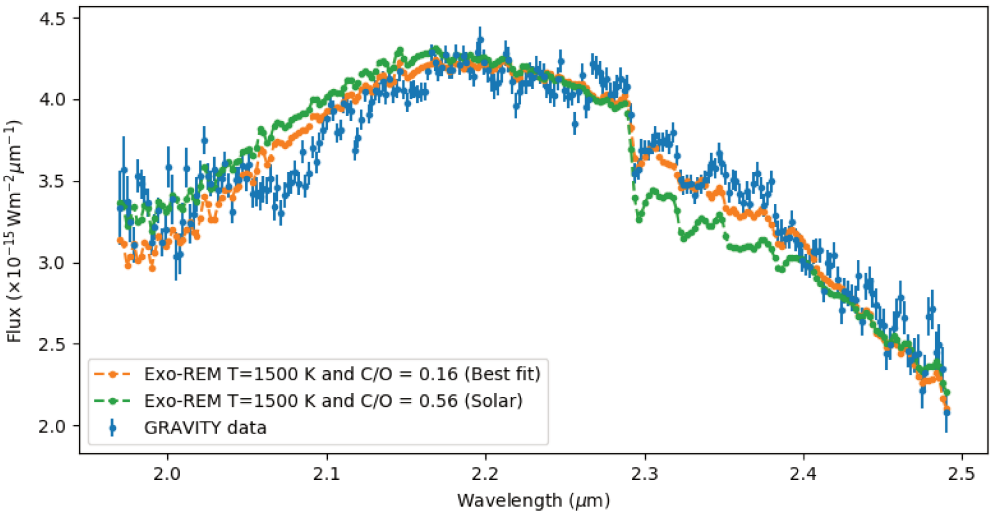


Figure 13.12: Best fit obtained with the Exo-REM model grid, at a fixed Solar metallicity and a fixed surface gravity of $\log(g/g_{\rm earth}) = 4.0$, corresponding to a temperature of 1500 K, and a C/O ratio of 0.16. For comparison, the Exo-REM spectrum obtained at the same temperature, but with a Solar C/O ratio is also overplotted.

Parameter	Base Exo-REM	With mass prior	Exo-REM with C/O	Note
Temperature (K)	1750	1550	1500	-
Log surface gravity	3.3	4.0	4.0 (fixed)	-
Mass ($M_{\rm Jup}$)	1	13	13	Derived with Eq 13.4
C/O	-	-	0.16	-

Table 13.4: List of parameters estimated using a χ^2 fit of the GRAVITY K-band spectrum with the Exo-REM grid of atmospheric model.

13.4.2 Retrieval with petitRADTRANS

A second approach can be followed to extract information about the atmosphere of β Pic b: atmospheric retrieval. Contrary to forward-modeling (e.g. with the Exo-REM grid model), in a retrieval code, the physicochemistry of the atmosphere is not calculated from the physical characteristics with a self-consistent model. Rather, these codes use a parametric model to calculate synthetic spectra, and then resort to statistical inference methods to extract these parameters from the data. A detailed description of what atmospheric retrieval consists in can be found in Madhusudhan [2018].

The analysis of β Pictoris b data was done using petitRADTRANS[1]. This part of the work was led by Paul Mollière (Leiden Observatory, Netherlands), and the results obtained are summarized in Figure 13.13.

[1]https://home.strw.leidenuniv.nl/ molliere/petitRADTRANS/index.html

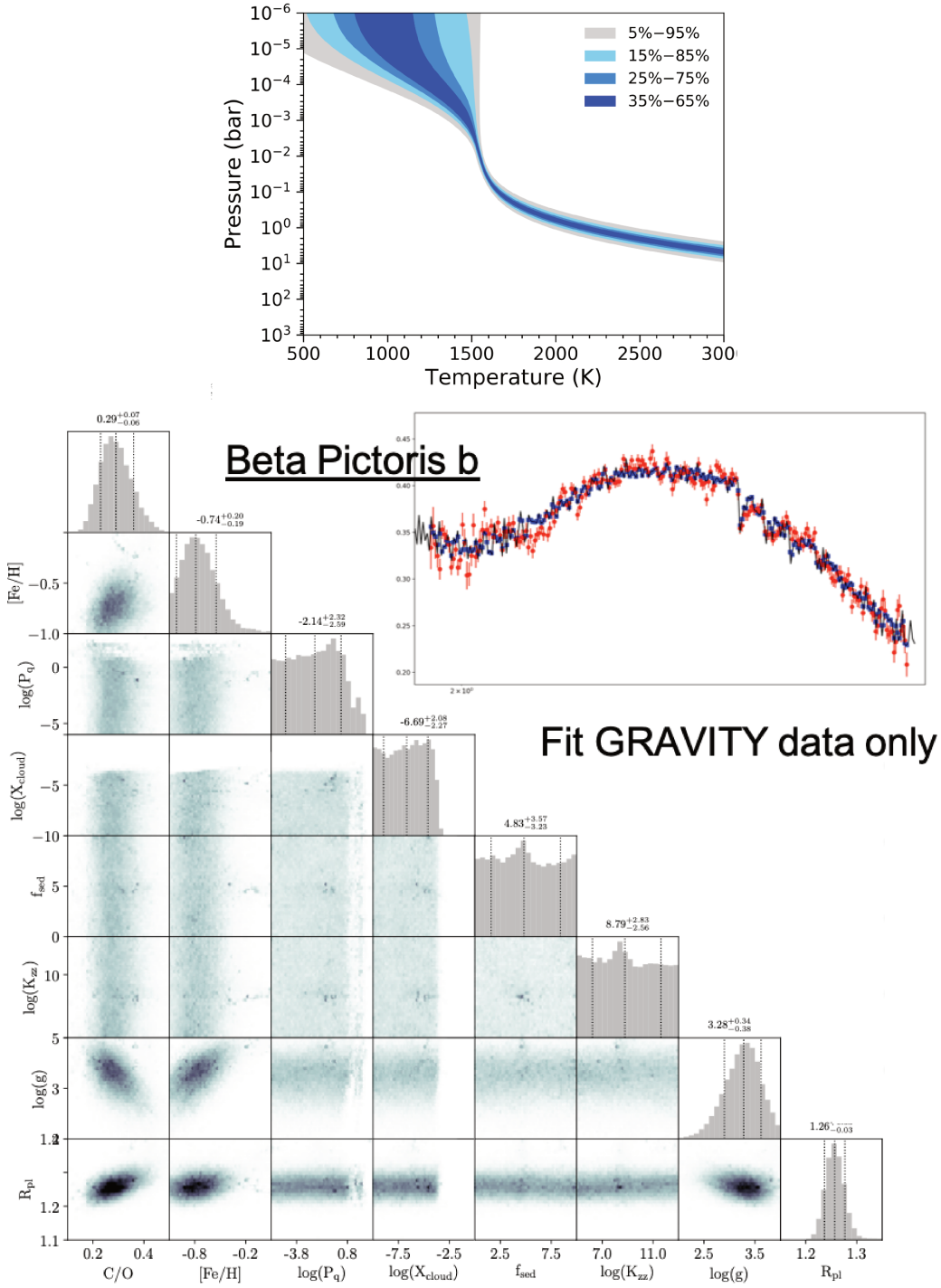


Figure 13.13: Results obtained by P. Mollière with the atmospheric retrieval code petitRADTRANS on the β Pic b GRAVITY data. The upper panel gives the pressure-temperature profile in the atmosphere, as retrieved with petitRADTRANS. On the bottom panel, the corner plot gives the posterior distributions of several parameters: the carbon to oxygen ratio (C/O), the matalicity (Fe/H), the log surface gravity ($\log g$), and the planet radius $R_{\rm pl}$. The inset next to the corner plot gives the GRAVITY spectrum (in red) and the fit obtained with petitRADTRANS (black line with blue dots).

13.5 C/O ratio and the formation history

The GRAVITY K-band spectrum obtained on β Pictoris b is the most resolved and the cleanest (highest S/N) obtained to date. I have shown in Section 13.4 that it could be used to extract the temperature of the atmosphere, the surface gravity, and the radius of the planet. These results are similar to what has been done in other studies (see Chilcote et al. [2017] for a recent example). But the GRAVITY K-band spectrum also gives an estimate of the atmospheric carbon to oxygen (C/O) ratio: $(\mathrm{C/O})_{\beta\ \mathrm{Pic\ b}} = 0.29^{+0.08}_{-0.07}$. In this section, I show how this measurement can bring new information related to the formation of the planet.

13.5.1 The link between element abundances and planet formation

Over the past few years, a whole new body of research has emerged in exoplanet science, with the explicit goal of linking planetary element ratios to formation processes. The question of the supersolar abundance of heavy elements in the atmosphere of Jupiter motivated the first attempts in that sense. Several studies looked at how planetesimal accretion in the gravitational instability and core-accretion scenarios could lead to heavy element enrichment [Helled et al., 2006, Helled and Schubert, 2009, Owen et al., 1999, Alibert et al., 2005]. On the exoplanet front, the work of Öberg et al. [2011] was the first general attempt to look at how element abundances and formation history are related. It forms the basis of the reasoning proposed in this section, and needs to be detailed.

At the core of the research carried out by Öberg et al. lies a simple idea, with far-reaching consequences: in a protoplanetary disk, material can be exchanged between the gas and solid phases, modifying the relative abundances of the different atomic elements. Their study is explicitly focused on the C/O ratio, which is the most easily accessible ratio, and the first one to have been measured in exoplanets [Madhusudhan et al., 2011, Konopacky et al., 2013]. Five molecular compounds are expected to contain most of the carbon and oxygen mass in protoplanetary disks: CO, CO_2, H_2O, carbon grains, and silicates (Si-O). Carbon grains and silicates are mostly solid, whereas the three other molecules are volatile, and can either be in a gaseous or condensed form. Far from the star, these three volatile molecules are in their condensed form. The solid phase is rich in both C and O. But when moving closer to the star, the temperature rises, and some molecules start to vaporize. At $\sim$ 20 K, the CO vaporizes, which depletes the solid phase in both C and O, and enriches the gas phase. At $\sim$ 50 K, CO_2 vaporizes, which strongly enriches the gas phase in oxygen (see illustration in Figure 13.14). The same happens when crossing the water snowline. A direct consequence is that even if the overall composition is homogeneous throughout the entire disk, the distribution of species between gas and solids is different. With some assumptions on the composition of the protoplanetary disk (molecular distribution, dust to gas ratio, etc.), it is possible to calculate the relative abundances of different elements as a function of the distance to the star. In the framework used by Öberg et al., where only the snowlines have an impact on the distribution of molecules, this gives characteristics "stepfunctions", each step corresponding to a given snowline (see example in Figure 13.15).

From this, it follows that the ratio of solid to gas accreted by the protoplanet to form its atmosphere will dictate its final atmospheric C/O ratio. If a planet forms between the CO_2 and CO snowlines, and accretes a considerable amount of solid material, then its final C/O ratio will be close to 0.3 (the C/O of the solid phase), whereas if it only accretes gaseous material, its C/O will be close to 1.0 (the C/O of the gas).

Of course, the step function presented by Öberg et al. [2011] is a gross simplification of the actual situation: it only considers three volatile species, and it does not take into account any possible deviation of the composition of the disk from the interstellar medium. Eistrup

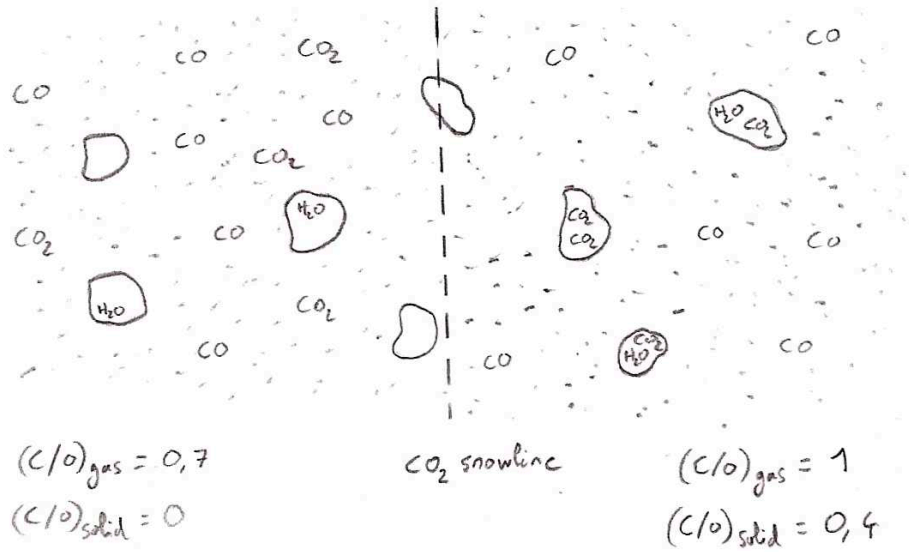


Figure 13.14: When crossing the CO_2 snowline, CO_2 molecules in the gas phase of the disk condense and move to the solid phase. This changes the C/O ratios of the gas and of the solid phases of the disk. The C/O number ratios given here are only for the sake of example.

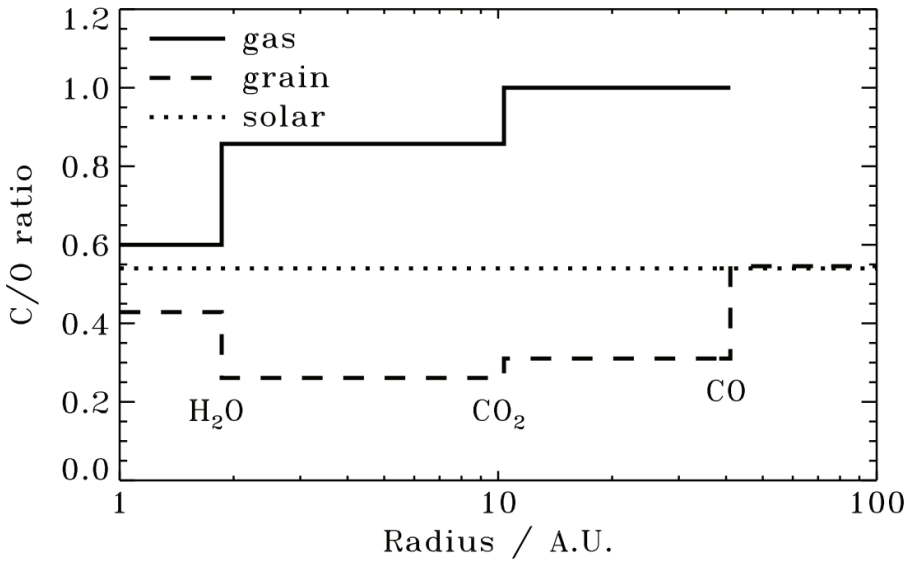


Figure 13.15: Carbon to Oxygen ratio as a function of distance to the star calculated for a typical protoplanetary disk (Sun like star, molecular abundances corresponding to what is observed in the Interstellar Medium). Figure taken from Öberg et al. [2011].

et al. [2016, 2018] have improved this by including a temporal dimension, taking into account the chemical evolution of the disk.

But even with a realistic model of the element distribution in the disk, calculating the element abundances in a hypothetical planetary atmosphere remains difficult. It strongly depends on the formation mechanism considered, in which many aspects remain poorly understood. It is worth mentioning three notable attempts to couple formation mechanisms to relative element abundances: Madhusudhan et al. [2014, 2017] studied the impact of planet migration and pebble accretion on the C/O ratio, and Mordasini et al. [2016] proposed an extensive model going from planet formation all the way to the generation of synthetic observations.

13.5.2 Beta Pic b C/O ratio in the gravitational collapse paradigm

Disk instability has been suggested as a possible pathway towards the formation of giant planets [Bodenheimer, 1974]. In this model, under specific conditions, the protoplanetary disk becomes unstable, leading to the creation of a self-gravitating cloud of gas and dust. Such a cloud contracts to form a protoplanet, which evolves into a giant planet. Contrary to the concurrent scenario of core-accretion, this model tends to lead to planets which are relatively hot at the beginning of their life ("hot-start").

I now reason in this gravitational collapse paradigm, and adopt a model similar to the one of Öberg et al. [2011] in which only the different snowlines have an effect on the composition of the solid and gas phases of the disk. My objective in this subsection is to calculate the C/O ratio in the atmosphere of a giant planet. To do so, I adopt the following notations:

- $n_{X,\mathrm{s}}$ (resp. $n_{X,\mathrm{g}}$) is the abundance of element X in the solid phase (resp. gas phase) of the disk, given relative to the total mass of the disk (i.e. in number of atoms per unit of disk mass).
- M_{solid} (resp. M_{gas}) is the total mass of solid (resp. gas) entering the composition of the atmosphere
- $f_{\mathrm{s/g}}$ is the dust-to-gas ratio in the disk

The total number of element X in the atmosphere is then given by the following equation, which is a different form of Equation 2 from Öberg et al. [2011]:

$$N_X = \frac{n_{X,\mathrm{s}}}{f_{\mathrm{s/g}}} \times M_{\mathrm{solid}} + \frac{n_{X,\mathrm{g}}}{1 - f_{\mathrm{s/g}}} \times M_{\mathrm{gas}}$$

The two factors $1/f_{\mathrm{s/g}}$ and $1/(1 - f_{\mathrm{s/g}})$ comes from the fact the the abundances are given per unit of disk mass, not per unit of solid or gas mass. The abundance of X per unit of solid mass is $n_{X,\mathrm{s}} \times M_{\mathrm{disk}}/M_{\mathrm{solid}} = n_{X,\mathrm{s}}/f_{\mathrm{s/g}}$, with a similar equation for the gas.
The C/O ratio in the planetary atmosphere is then:

$$\mathrm{C/O} = \frac{n_{\mathrm{C,s}}\, {f_{\mathrm{s/g}}}^{-1} M_{\mathrm{solid}} + n_{\mathrm{C,g}} \left(1 - f_{\mathrm{s/g}}\right)^{-1} M_{\mathrm{gas}}}{n_{\mathrm{O,s}}\, {f_{\mathrm{s/g}}}^{-1} M_{\mathrm{solid}} + n_{\mathrm{O,g}} \left(1 - f_{\mathrm{s/g}}\right)^{-1} M_{\mathrm{gas}}} \tag{13.5}$$

The disk composition I use is given in Table 13.5. Like Öberg et al. [2011], I assume that the main sources of C and O are H_2O, CO, CO_2, carbon grains, and silicate grains. I use a H_2O to CO ratio of 1:1 [Pontoppidan et al., 2005], and assume a CO rich disk (CO to CO_2 ratio 5:1, Marboeuf et al. [2014]). The C/O ratio is assumed to be 0.8 (in the range of values suggested in Öberg et al. [2011]). Holweger et al. [1997] have shown that the abundance of

Species	$n_{\mathrm{specie}}/n_{\mathrm{H_2O}}$	Note
H_2O	1	-
CO	1	Pontoppidan et al. [2005]
CO_2	0.2	5:1 ratio with CO from Marboeuf et al. [2014]
C (grains)	0.4	to ensure solar total C/O
O (silicates)	0.5	C/O = 0.8 in grains (compatible with values in Öberg et al. [2011])

Table 13.5: Relative abundances of the different species used for the young β Pic protoplanetary disk, given relative to H_2O

several elements (C, Ca, Ti, Cr, Fe, Sr, Ba) on the surface of β Pictoris were solar. Although the exact abundance of oxygen in β Pic remains unknown, it is reasonable to assume that it also is solar. To ensure a solar C/O ratio in the overall disk material, I take a relative abundance of C grains to H_2O of 0.4. The dust-to-gas ratio in the disk is fixed at $f_{\mathrm{s/g}} = 0.01$.

With all these assumpions, the overall disk composition is simillar to the one from Öberg et al. [2011], but with an increased proporption of water, which helps to lower the C/O ratio of the solid material exterior to the H_2O snowline.

The mass of β Pictoris b has been directly measured by Snellen and Brown [2018]: $M_{\mathrm{planet}} = 11 \pm 2\ M_{\mathrm{Jup}}$. In Section 13.3, I presented some results obtained using Hipparcos/Gaia astrometric measurements, and the new VLTI/GRAVITY position derived in this work. These results point to a planet mass of $M_{\mathrm{planet}} = 12.5 \pm 2\ M_{\mathrm{Jup}}$, in good agreement with the mass from Snellen and Brown. In the rest of this section, I use $M_{\mathrm{planet}} = 12\ M_{\mathrm{Jup}}$, together with the consistent measurement $M_{\mathrm{star}} = 1.82\ M_{\odot}$ (also from Section 13.3).

Studying the sedimentation of grains in a forming protoplanet, Helled and Schubert [2008] demonstrated that giant planets of mass $\geq 5 M_{\mathrm{Jup}}$ formed in the gravitational collapse scenario were not likely to have a core. The total mass of the atmosphere and the mass of the planet coincides, and I can write:

$$M_{\mathrm{planet}} = M_{\mathrm{gas}} + M_{\mathrm{solid}} \tag{13.6}$$

Using this total mass, the C/O ratio can be written as a function of the solid mass only:

$$\mathrm{C/O} = \frac{n_{\mathrm{C,s}}\, {f_{\mathrm{s/g}}}^{-1} M_{\mathrm{solid}} + n_{\mathrm{C,g}} \left(1 - f_{\mathrm{s/g}}\right)^{-1} \left[M_{\mathrm{planet}} - M_{\mathrm{solid}}\right]}{n_{\mathrm{O,s}}\, {f_{\mathrm{s/g}}}^{-1} M_{\mathrm{solid}} + n_{\mathrm{O,g}} \left(1 - f_{\mathrm{s/g}}\right)^{-1} \left[M_{\mathrm{planet}} - M_{\mathrm{solid}}\right]} \tag{13.7}$$

The remaining question is: what values to use for n_{O} and n_{C}? In the framework of the Öberg et al. model, this is equivalent to the question: where in the disk did β Pic b form? Adding a post-conjunction value to the extensive series of astrometric positions available for β Pic b, Lagrange et al. [2019] determined a value of $a = 8.90^{+0.25}_{-0.41}$ AU for the semi-major axis of the planet, and an eccentricity of $e = 0.01^{+0.029}_{-0.01}$. This gives a perihelion distance of ~ 8.8 AU, and and aphelion distance of ~ 9.0 AU. The results presented in Section 13.3 point to a more eccentric orbit, with a semi-major axis close to 11 AU, and an eccentricy closer to 0.15 or even 0.20. In this case, the perihelion and aphelion distances are respectively 9.5 AU and 13 AU.

To estimate the location of the different snowlines in the young β Pic b disk, I use the values for the only two disks observed around A-type stars in the list given by Andrews and Williams [2005]: V892 Tau, and AB Aur. Andrews and Williams provide an estimate of the

	H_2O snowline	CO_2 snowline	CO snowline
$T_{\text{evaporation}}$	135 K	47 K	20 K
AU Aur	9.2 AU	96.3 AU	642.8 AU
V892 Tau	8.3 AU	51.2 AU	223.6 AU

Table 13.6: Distance to the star for the snowlines of carbon monoxide (CO), carbon dioxide (CO_2), and water (H_2O) calculated for AU Aur and V892 Tau. The temperature laws used in these calculations are from Andrews and Williams [2005]. The evaporation temperatures are the same as the one used by Öberg et al. [2011].

temperature laws for these two systems:

$$T_{\text{AB Aur}}(r) = 367 \text{ K} \times \left(\frac{r}{1 \text{ AU}}\right)^{-0.45} \tag{13.8}$$

$$T_{\text{V892 Tau}}(r) = 461 \text{ K} \times \left(\frac{r}{1 \text{ AU}}\right)^{-0.58} \tag{13.9}$$

Using these temperature laws together with the evaporation temperatures of CO, CO_2, and H_2O reported by Öberg et al. [2011], is is possible to calculate the limit of the different snowlines in these two A systems. The results are given in Table 13.6. In the V892 Tau system, the current orbit of β Pic b would be very close to the water snowline. In AU Aur, the planet would be in between the water and carbon dioxide snowlines. Given that little is known about the relationship between the current orbit of β Pic b and its exact formation location (planets can migrate), it is hard to make a definitive claim as to where the planet formed in comparison to the water snowline at this point, and two options must be considered: a formation within the water snowline, and a formation between the water and the CO_2 snowlines.

The total mass of solid entering in the composition of the planet can be separated in two terms: the mass of solid initially contained in the disk fragment which collapsed to create the protoplanet, and the mass of planetesimals later accreted onto this protoplanet. The solid mass contained in the initial clump is directly related to the dust-to-gas ratio of the disk. Under the assumption that the mass of solid accreted by the planet is small compared to its total mass, I can write:

$$M_{\text{solid}} = f_{\text{s/g}} M_{\text{planet}} + M_{\text{accreted}} \tag{13.10}$$

This equation relies on the assumption that no core has formed in the young β Pic b protoplanet. Otherwise, some of the initial solid mass (the $f_{\text{s/g}} M_{\text{planet}}$ term) would have sedimented to form the core, and could no longer be counted as entering in the composition of the atmosphere.

In Figure 13.16, I give the evolution of the C/O ratio as a function of this accreted solid mass M_{accreted}, for two possible formation locations: within the water snowline (blue curve), and between the water and carbon dioxide snowlines (orange curve). This Figure indicates that in this model, β Pic b is unlikely to have formed within the water snowline.

Planetesimal accretion by a protoplanet in the gravitational collapse paradigm has been studied by Helled and Schubert [2009]. Under the assumption that the surface density in the disk in distributed in a power-law fashion, they show that the total mass of solid material available for accretion by the forming protoplanet is given by (combination of their Equations 2 and 4):

$$M_{\text{av}} = \frac{8}{70} M_{\text{disk}} (2 - \alpha) \left(30^{2-\alpha} - 0.1^{2-\alpha}\right)^{-1} \left(\frac{a}{\text{AU}}\right)^{2-\alpha} \left(\frac{M_{\text{planet}}}{3 M_{\text{star}}}\right)^{\frac{1}{3}} \tag{13.11}$$

in which α is the power-law coefficient in the surface density law. Neglecting the term in $0.1^{(2-\alpha)}$, and taking the log-derivative of this equation with respect to the coefficient α yields:

$$\frac{\partial \log(M_{\rm av}/M_{\rm disk})}{\partial \alpha} = -(2-\alpha)^{-1} - \ln\left(\frac{a}{\rm AU}\right) + \ln(30) \tag{13.12}$$

Solving $\frac{\partial \ln M_{\rm av}/M_{\rm disk}}{\partial \alpha} = 0$ for $a \simeq 9$ AU gives α for which the available mass is maximum at the orbit of β Pic b:

$$\alpha_{\rm mass\ max} = 2 - \frac{1}{\ln\left(\frac{10}{3}\right)} \simeq 1.17 \tag{13.13}$$

Using this value of α and $a = 9$ AU in Equation 13.11 gives an absolute maximum mass of planetesimals available for enriching the β Pic b protoplanet (which only relies on the power-law model for the surface density of the disk):

$$M_{\rm av,max} \simeq 4.4 \times 10^{-3}\ M_{\rm disk} \tag{13.14}$$

The two young disks found around A-type stars by Andrews and Williams [2005] have a mass of 0.004 $M_\odot$ and 0.009 $M_\odot$. Considering all stellar types, the median disk-to-star mass ratio they found is 0.5%. In the case of the young β Pic disk, this leads to a mass of $M_{\rm disk} \simeq 0.01 M_\odot$. The giant planet β Pic b being particularly massive, the initial disk around β Pic may have been in the massive tail of the distribution found by Andrews and Williams. A value of $M_{\rm disk} \simeq 0.1 M_\odot$ would put the disk in the few most massives among all their 175 targets. These two possible assumptions for the initial disk mass would lead to a total mass of solid available for accretion by the young β Pic b:

$$M_{\rm av,max} \simeq \begin{cases} 13\ M_{\rm Earth} & \text{for a common initial disk } (M_{\rm disk} = 0.01\ M_\odot) \\ 130\ M_{\rm Earth} & \text{for a massive initial disk } (M_{\rm disk} = 0.1\ M_\odot) \end{cases}$$

From Figure 13.16, it is clear that β Pic b cannot have formed in a light disk. The quantity of solid material available would have been far too small to reach the current C/O ratio. The limit corresponding to a massive disk is overplotted to the C/O curve. It shows that even in the case of a massive disk, it is not obvious that the forming β Pic b had enough solid material to lower its C/O ratio. And yet this limit does not even account for the fact that it is highly unlikely that β Pic b had enough *time* to accrete all this material at its disposal.

The formation of a planet by gravitational instability can be separated in a few different steps [Bodenheimer, 1974]: formation of the initial clump in the disk, quasi-equilibrium contraction, hydrodynamic collapse, and a new hydrostatic quasi-equilibrium phase. Accretion of planetesimals is thought to be efficient only during the pre-collapse phase, when the planet has not yet opened a gap in the disk (see [Helled and Schubert, 2009]). The duration of this phase decreases with increasing planet mass, and typical values ranges from a few 10^5 years for a Jupiter mass planet, down to a few 10^3 years for more massive planets [Decampli and Cameron, 1979, Bodenheimer et al., 1980]. Using the model proposed by Helled and Schubert [2009], the mass of planetesimal accreted during the pre-collapse phase of β Pic b can be estimated using:

$$M_{\rm accreted} = \int_0^{t_{\rm collapse}} \pi R^2_{\rm capture}(t)\sigma(a,t)\Omega(a){\rm d}t \tag{13.15}$$

Where $t_{\rm collapse}$ is the time of collapse, $R_{\rm capture}$ the protoplanet's capture radius, σ the surface density of solids in the disk at the location of the protoplanet, and Ω the orbital frequency.

Accepting the hypothesis of a massive disk ($M_{\rm disk} = 0.1\ M_\odot$), and using a power-law for the surface density ($\sigma = \sigma_0\,(a/5\ {\rm AU})^{-\alpha}$, with $\alpha = 1.17$), the solid density at $a = 9$ AU is:

$$\sigma(9\ {\rm AU}) \simeq 10\ {\rm g/cm^2} \tag{13.16}$$

The orbital period of the planet is $\sim$ 20 yr [Wang et al., 2016, Lagrange et al., 2018, Section 13.3 of this work]. The capture radius decreases with the contraction of the planet, but an optimistic value would be 2 to 3×10^{12} cm for a 1 $M_{\rm Jup}$ planet [Helled et al., 2006]. For a planet 10 times more massive, the effective radius would be $\simeq 5 \times 10^{12}$ cm. This yields:

$$M_{\rm accreted} \simeq 6.5 \times M_{\rm Earth} \times \frac{t_{\rm collapse}}{1000 \text{ yr}} \tag{13.17}$$

The corresponding accretion limits (for $t_{\rm collapse} = 10^3$ yr, a reasonable assumption, and $t_{\rm collapse} = 10^4$ yr, an extremely high value) have been added to the C/O ratio curve in Figure 13.16.

Taking into account the effective time available for efficient planetesimal accretion during the pre-collapse stage, the low C/O ratio measured with GRAVITY cannot be properly explained, even in the case of a planet forming oustide the H_2O snowline, in a massive disk. Significantly lowering the C/O ratio in a planet as massive as β Pic b requires the capture of a few 100 $M_{\rm Earth}$ of solid material from the disk, which is unlikely.

This can be seen as the general takeaway message of this study: contrary to the case of Jupiter-like planets, the formation of extremely massive planets (10 $M_{\rm Jup}$) through disk instability is so fast (1000 yr) that enrichment by planetesimals cannot play a significant role. The final C/O ratio is likely to reflect the overall C/O ratio of the disk, i.e. to be close to stellar. A large deviation from the stellar C/O ratio is evidence for a different formation mechanism.

13.5.3 Beta Pic b C/O ratio in the core-accretion paradigm

In the gravitational collapse paradigm, I have neglected any possible chemical evolution of the disk. Its molecular content was assumed to be close to the interstellar content [Öberg et al., 2011]. This is justified by the fact that disk instabilities typically occur in $\sim$ 100 to 1000 yr [Boss, 1997, Mayer et al., 2002], whereas chemical evolution can take a few Myr to change the composition [Eistrup et al., 2018]. The situation is different in the core-accretion paradigm.

Core-accretion is an alternate formation mechanism, in which an initial solid core forms, and slowly accretes gas from the disk. When the mass of gas is roughly the same as the mass of the core, the protoplanet enters a phase of "runaway gas accretion", during which it gains significant amount a gas over a short time [Lissauer and Stevenson, 2007]. In this scenario, the formation of a planet is a much longer process than it is with gravitational instability. This provides a plausible way to explain the final C/O ratio in β Pic b, as it allows for the disk to chemically evolve during the formation of the planet.

Eistrup et al. [2018] studied the chemical changes in gas and ices of a protoplanetary disk, under different conditions, taking into account the physical evolution of the disk mid-plane. Their study show that the C/O step function of Öberg et al. [2011] can be significantly affected by chemistry over a few Myr.

Mordasini et al. [2016] explored the effect of planetesimal enrichment coupled with disk composition, in a core-accretion scenario. They focused on the case of Jupiter mass planets migrating to short period orbits ("hot Jupiters"), which is a different archetype than β Pic b. But the general sequence of events they use to form their planets in the core-accretion paradigm can still be applied to β Pic b, only leaving out the inward migration part:

- The core of the planet forms from the accretion of solid material.
- When the core has formed, the protoplanet starts accreting a gaseous envelope.

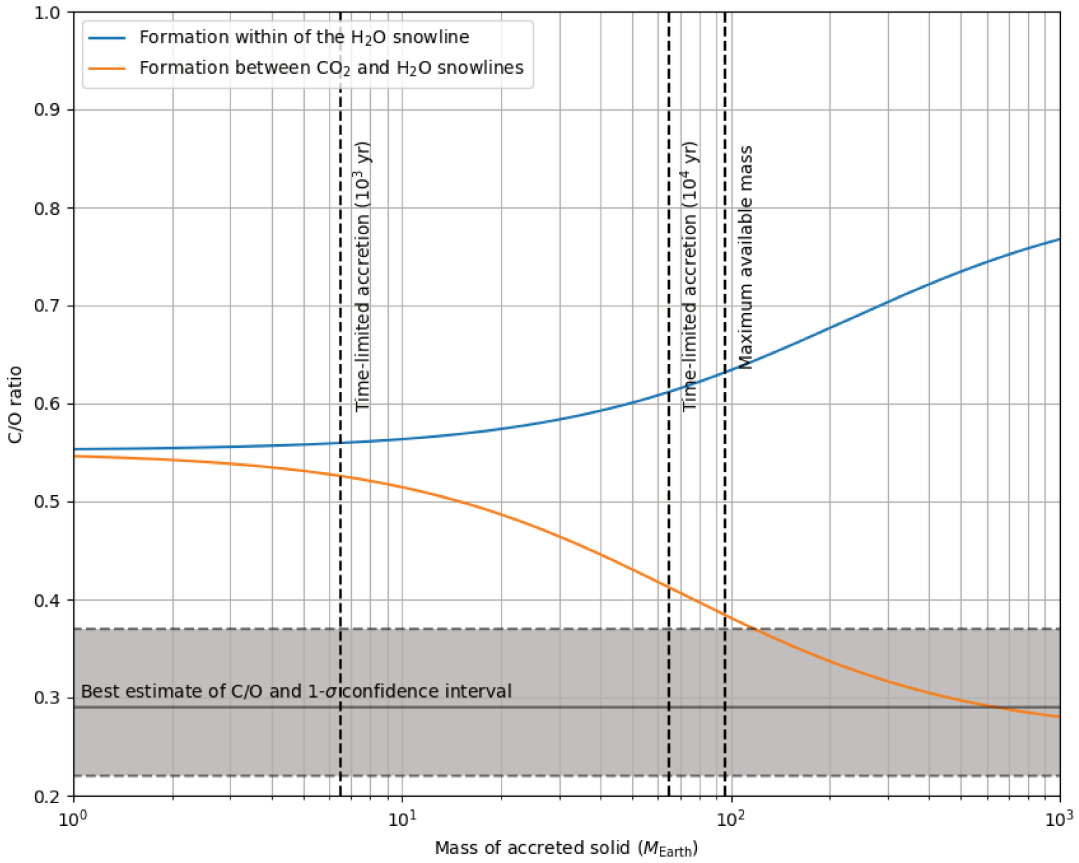


Figure 13.16: Evolution of the C/O ratio as a function of the total mass of solid accreted after the initial formation of the protoplanet. The blue curve corresponds to a formation within the H_2O snowline, and the orange curve corresponds to a formation between the H_2O and CO_2 snowlines. The shaded gray area is the 1-σ interval of our GRAVITY C/O measurement. Dashed vertical lines corresponds to different solid accretion limits discussed in the text. In this plot, the origin of the x-axis (no accretion of solid after the initial formation of the protoplanet) corresponds to a stellar C/O ratio because we assumed than no core was formed inside the protoplanet [Helled and Schubert, 2008]. Gas and dust in the initial protoplanet are disk gas-to-dust proportions, and the C/O ratio of the initial protoplanet is the same as the C/O ratio of the disk (i.e. stellar).

- During its formation, the proto-atmosphere is enriched by the accretion of disintegrating planetesimals.

- When the planet reaches a critical mass, runaway accretion occurs. At this step, the mass of the planet significantly increases.

- Runaway gas accretion clears a gap in the disk, ending the formation of the planet.

In the gravitational instability scenario, because the formation of the planet happens so quickly compared to typical timescales of disk evolution, the gas and solid making the atmosphere necessarily have a stellar combined C/O. If the solid and gas in the atmosphere are in the same proportion as they are in the disk ($M_{\rm solid} = f_{\rm s/g} M_{\rm gas}$), the C/O of the atmosphere is stellar. A deviation of the solid to gas proportion in the atmosphere is required to alter the C/O ratio.

In the case of core-accretion, the situation is different. Planetesimal enrichment of the proto-atmosphere occurs at an earlier time than runaway gas accretion, which provides the bulk of the gas mass. Hubickyj et al. [2005] have shown that, depending of grain opacity in the disk and mass of the core, runaway gas accretion can occur at any time between 1 Myr and 6 to 7 Myr. But 6 to 7 Myr is also the typical time required for the molecular composition of the disk to be significantly affected by chemistry. Consequently, if runaway accretion is a late event, then the solid and gaseous material forming the atmosphere of the planet comes from two virtually different disks: the initial protoplanetary disk for the solid, with molecular abundances close to the interstellar medium; and en evolved version of this disk for the gas, in which molecular abundances can be significantly altered.

Eistrup et al. [2018] have shown that a large fraction of water molecules can be transformed into dioxygen (O_2) over a few Myr, along a chemical pathway detailed in Walsh et al. [2015]. The binding energy of the dioxygen is smaller than the binding energy of CO_2 and thus, at the location considered for the forming β Pic b, O_2 is in gaseous form. As a result, the $H_2O \rightarrow O_2$ reaction can deplete the solid phase from its oxygen to enrich the gas phase. If most of the water turns to dioxygen in $\sim$ 7 Myr, then the exact time of runaway accretion can have a significant impact on the final C/O. In Figure 13.17, I give the evolution of the C/O ratio in the atmosphere as a function of the total mass of planetesimals accreted, for the case of an early runaway accretion (1 Myr, blue curve), and a late runaway accretion (7 Myr, red curve).

Using the temporal evolution curve for H_2O and O_2 given in Figure 7-a of Eistrup et al. [2018], and using the relative abundances of Table 13.5, it is possible to plot the evolution of the C/O in the gas and in the solid for the hypothetical young β Pic protoplanetary disk. The result, only valid between the H_2O and CO_2 snowlines, is given in Figure 13.18. In this Figure, I have also overplotted two different scenarios for the formation of a planet:

- The two blue curves corresponds to a scenario in which runaway accretion occurs at an early time. In this scenario, the planet accretes planetesimals before 1 Myr, when the material is still water-rich. The planet also accretes some gas, with a mixing gas to solid ratio of 1:1 (faded curve), or 1:3 (solid curve). At $t = 1$ Myr, the gas envelope reaches an arbitrary critical mass of 40 $M_{\rm Earth}$ ($\sim$ 1% of the final planet mass), and runaway accretion occurs. This stops the formation of the planet. The final C/O ratio depends on the mass of planetesimals accreted in the initial step (see blue curve on Figure 13.17).

- The two red curves corresponds to a scenario in which runaway accretion happens at a later time. In this scenario, the planet also accretes planetesimals before 1 Myr, but accretion of the gas is slowed down by a factor 7. It takes 7 Myr to reach $\sim$ 1% of the mass of the planet, which delays runaway accretion. I also assume that solid accretion

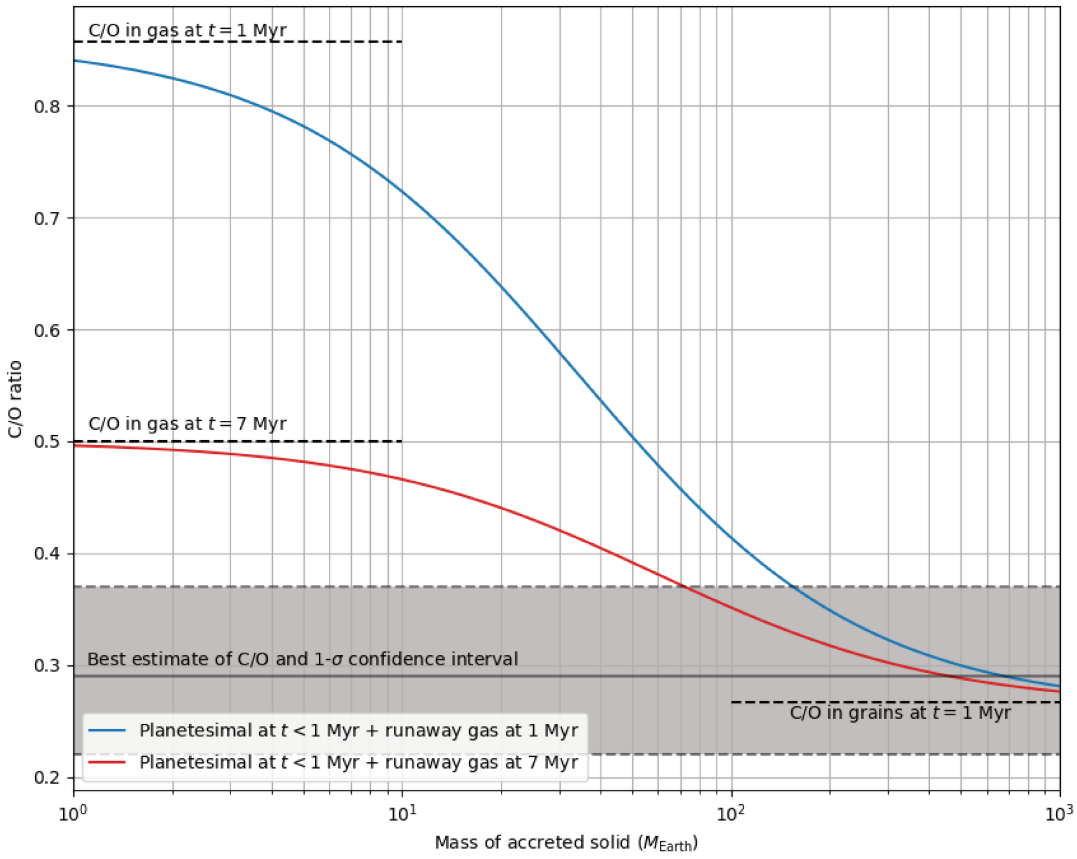


Figure 13.17: Evolution of the C/O ratio in the atmosphere of β Pic b as a function of the total mass of solid accreted by the protoplanet, for a formation between the H_2O and the CO_2 snowlines. The blue curve corresponds to an early runaway accretion (1 Myr), and the red curve corresponds to a late runaway accretion (7 Myr). In both cases, accretion of planetesimals happens during the early stage of the formation, when the disk is still water-rich (see Table 13.5 for the list of relative abundances used in this model). The gray area gives the 1-σ confidence interval of our GRAVITY C/O measurement.

stops at 1 Myr, for an unknown reason (depletion of the disk, competition with another planet?). In this case, the solid material accreted by the planet is water-rich, while most of the gas is O_2 rich. The final atmospheric C/O ratio is lower than for the previous scenario (see also Figrue 13.17).

This second scenario, in which runaway accretion occurs late in the formation of the planet, can explain the observed low C/O ratio with accretion of $< 100 M_{\mathrm{Earth}}$.

But this model comes with several important limitations. First, the exact reason why all planetesimal accretion would occur at the beginning of the formation (within the first Myr) in unclear. In this model, I found that an accretion equally spread over 7 Myr is not compatible with a final C/O of 0.3. In that case, a significant fraction of the accreted planetesimals are oxygen-poor, and the mean C/O of the accreted solid is too high.

Another major issue is the efficiency of the planetesimal enrichment. I have assumed that all solid material accreted after the formation of the core was converted into "atmospheric material". This is unlikely to be the case. A fraction of this material can be deposited into

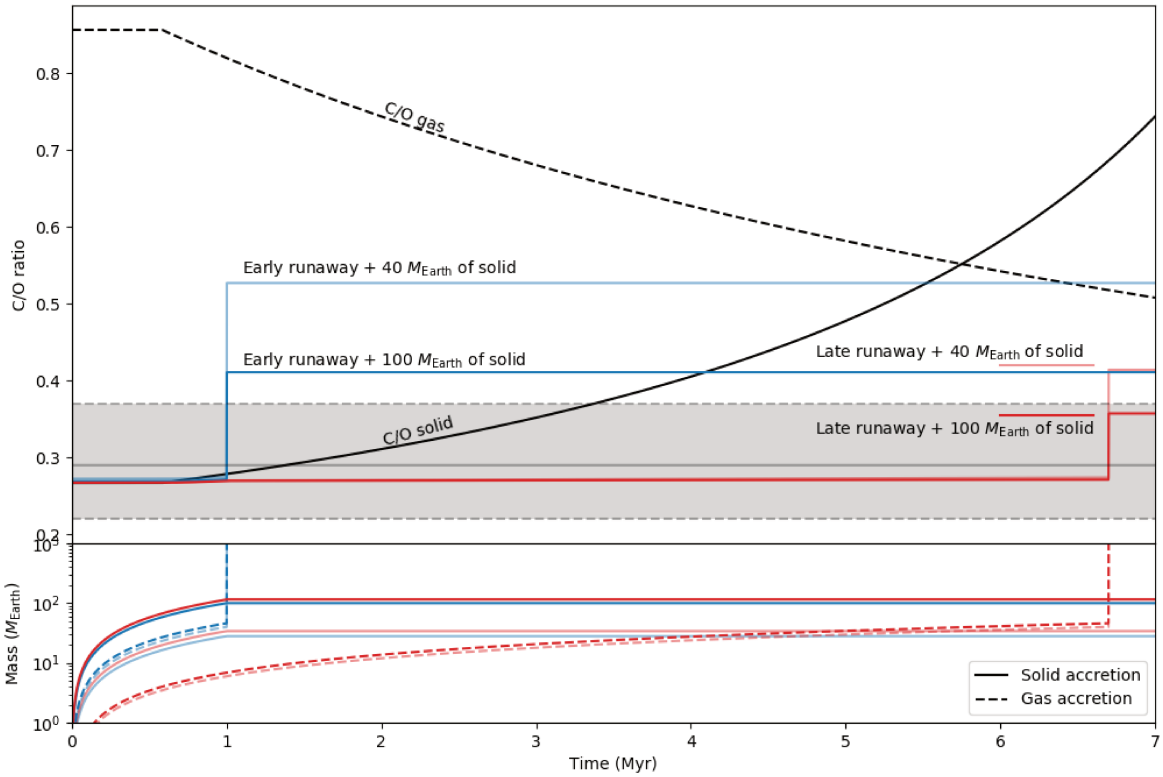


Figure 13.18: Evolution of the C/O of the proto-atmosphere of β Pic b for two possible scenarios. The upper panel gives the evolution of the C/O ratio, and the lower panel gives the evolution of the accreted mass of solid material (solid lines), and gas (dashed lines). In the upper panel, we also give the general evolution of the C/O in the solid and gas phases of the disk, taking into account the chemical transformation of H_2O into O_2 (kinetics of the reaction taken from Eistrup et al. [2018]). The blue scenario corresponds to an early runaway accretion (1 Myr), and the red scenario to a late runaway accretion (7 Myr). Faded curves correspond to an accretion of 40 $M_{\rm Earth}$ of solid, and solid lines correspond to 100 $M_{\rm Earth}$. In all cases, accretion of gas and solid are assumed to proceed at constant rate, with a sharp cutoff when the desired mass is reached. Runaway gas accretion occurs when the total mass of gas in the proto-atmosphere reaches 40 $M_{\rm Earth}$, which corresponds to $\sim$ 1% of the final planet mass. The gray area gives the 1-σ interval of our GRAVITY C/O measurement.

the planetary core, or can stay at the bottom of the atmosphere, especially given that in this scenario, solid material is accreted *before* most of the gas [Mordasini et al., 2016]. Strong vertical mixing can potentially mitigate this problem.

13.6 Conclusion

The K-band spectrum obtained on β Pic b with GRAVITY is extremely rich in information. This spectrum alone can provide an estimate of the temperature of the planet, its surface gravity, its radius, and mass, which are in good agreement with what previous studies have found. The temperature estimated with Exo-REM seems to be slightly lower than what Chilcote et al. [2017] found, but not totally inconsistent given that the error reported on these estimates do not include modeling errors.

The K-band is particularly interesting, because it contains an absorption band of water in its blue part (around 1.9 μm), and a CO band in its red part (2.3 μm). This makes the spectrum sensitive to the atmospheric C/O ratio, and allow for an estimate of this abundance ratio. I have shown in this Chapter that the C/O ratio could be used to put new constraints on the formation mechanism. Two different methods (forward modeling and atmospheric retrieval) lead to a consistent result: the C/O ratio in β Pic b is significantly inferior to the stellar C/O ratio. On the one hand, it is difficult to explain this when considering a formation through gravitational instability, mostly because this mechanism is extremely fast at creating planets. Lowering the C/O ratio requires to accrete a lot of solid ice-rich material from the initial protoplanetary disk, which takes time. On the other hand, a core-accretion scenario for the formation of β Pic b can potentially explain the low C/O, if coupled with the chemical evolution of the disk. This scenario, however, relies on many important assumptions, from the stellar abundances in the initial protoplanetary disk to the perfect mixing of planetesimals in the atmosphere. It needs to be taken cautiously, and the general picture may evolve significantly in the near future. Our understanding of interactions between forming planets and protoplanetary disks is still limited, as only recently did astronomical instruments become capable of measuring imprints and leftovers of these interactions.

Yet, the study proposed in this Chapter indicates that core-accretion, coupled with a chemical evolution of the protoplanetray disk, is a possible scenario to explain the formation of the giant planet β Pictoris b.

Part IV

GENERAL CONCLUSION & BIBLIOGRAPHY

The 2017 β Pic b transit campaign

PhD projects rarely go as planned, and this one was no exception. When I started my work on PicSat, I was expecting the central piece of this manuscript to be the light-curve of β Pic, possibly showing some interesting transit events. But PicSat failed, and there is no such light-curve.

Fortunately, other projects succeeded where PicSat failed, and monitored the photometry of β Pictoris during the year 2017. A large international campaign was formed to observe the Hill sphere transit, which involved many different instruments, on the ground and in space. Satellites from the BRITE constellation, made of five 8U CubeSats and monitoring stars across the celestial sphere [Weiss et al., 2014], were used to observe β Pic in late 2016 and Q1-Q2 2017. Also from space, an observation program used the Hubble Space Telescope to obtain precision photometry in January and February 2017 (baseline observation), and at two epochs around the time of ingress of the half Hill sphere. On the ground, ASTEP [Mékarnia et al., 2017], a 40 cm telescope located at the Concordia Research Station, at Dome C, Antarctica, and bRing [Stuik et al., 2017], a set of two cameras located in South Africa and Australia, continuously monitored the photometry of β Pic in 2017 and in the first half of 2018.

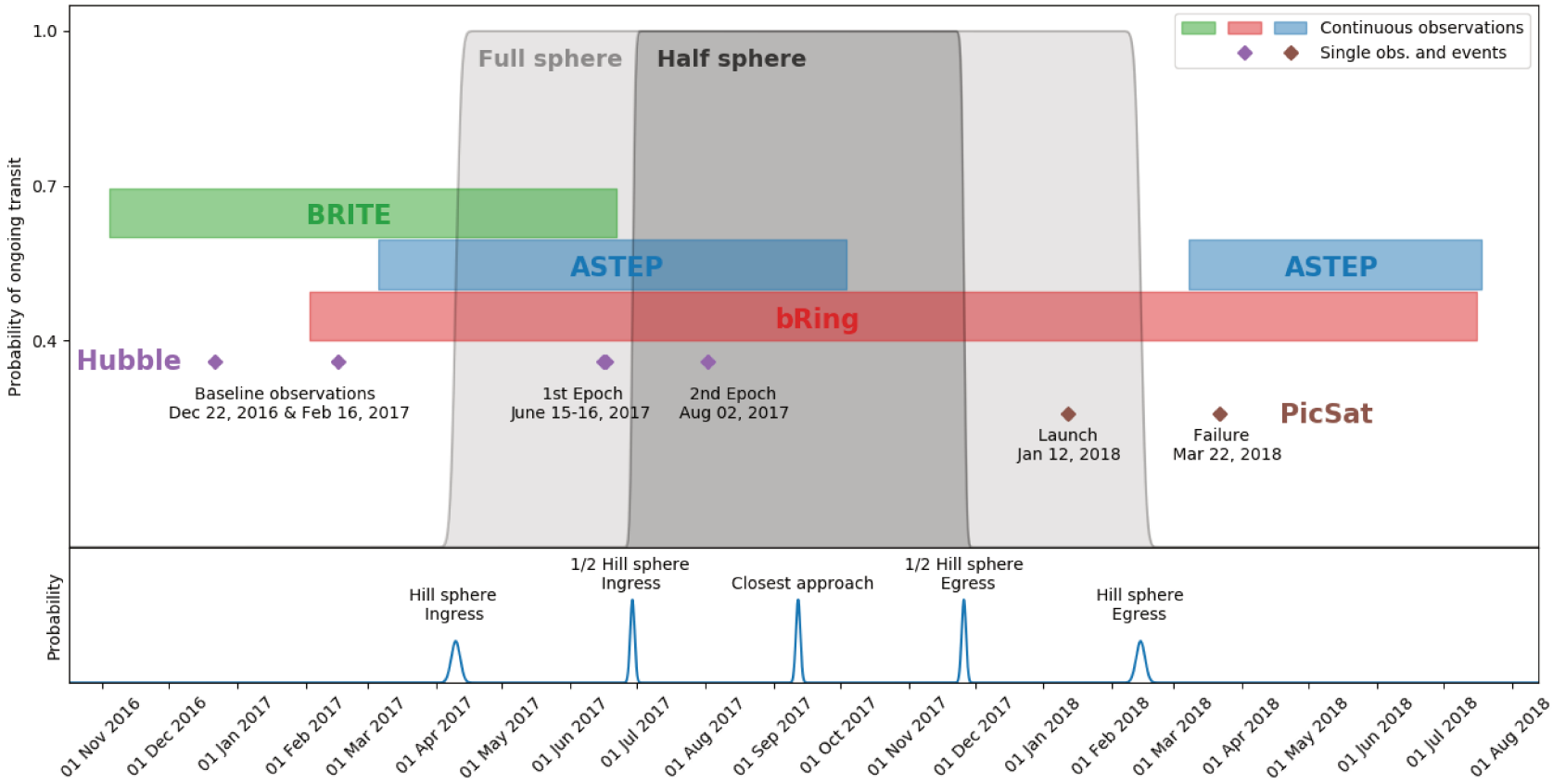
The precise GRAVITY astrometry obtained soon after the re-emergence of the planet, and presented in this work, strongly constrains the timing of the transit events. In fact, the uncertainty on the ingress and egress times of the Hill sphere transit are now limited by the uncertainty on the size of the Hill sphere, and not anymore by the uncertainty on the position of the planet. This gives an excellent *a posteriori* view of how the events unfolded during the two years of the international β Pic b transit campaign, which I show in Figure 13.19.

The Hill sphere transit was not detected, by any of the instrument part of this international campaign.

Exoplanets with GRAVITY

As PicSat did not return any scientifically useful data, the GRAVITY post-conjunction observation of β Pictoris b is arguably the main science result of my PhD work. This second direct detection of an exoplanet with long-baseline interferometry (the first one being reported in Gravity Collaboration et al. [2019]) provided the cleanest and most resolved spectrum of a directly image planet ever obtained. Combined with the most recent developments in exoplanet atmosphere modeling and retrieval, this GRAVITY observation made possible to estimate the Carbon-to-Oxygen ratio in the atmosphere of β Pic b, allowing us to peer for the very first time in its formation history. Given how recent these atmosphere retrieval techniques are, and how difficult it is to estimate exoplanetary element ratios, these results remain speculative. On the one hand, the discussion about the formation history of β Pic b proposed in Chapter 13 of this manuscript may very well be superseded by better results in the near future. But on the other hand, it also seems that the GRAVITY observation technique presented in Chapters 11 and 12 will make an important contribution to exoplanet science in the near future.

The exact role of GRAVITY for exoplanet research remains to be determined. The use of single-mode fibers and dual-field interferometry, which make such observations of high-contrast objects possible, also severely limits the field-of-view of the instrument, thereby wiping out the prospect of a large exploratory survey aimed at detecting new planets. Even with optimistic assumptions about the quality of observations, like considering that a single 5 minutes observation can be enough to claim a detection, a blind search with GRAVITY would be extremely time-consuming. The field-of-view of the fiber is ~ 60 mas in diameter,



A review of the international β Pic b transit campaign, with the timing of the transit events strongly constrained by the precise VLTI/GRAVITY post-conjunction astrometry. The bottom panel gives the probability distribution of the 5 main transit events (ingress and egress of the full and half Hill sphere, and closest approach. The top panel gives the probability of ongoing transit of the full and half Hill sphere (cumulative distribution of ingress minus cumulative distribution of egress), with the different observation windows and events of the transit campaign overplotted.

so exploring the entire region from 60 to 150 mas around a star would take approximately 10 hr on the 4 Unit Telescopes of the VLT.

A more clever approach is necessary to look for planets with GRAVITY. One possibility is to reduce the dimension of the search space by looking for planets in edge-on debris disks. This would restrain the search to a single line, and reduce the required observing time to ~ 1 hr. Such an observing strategy would also build on the low sensitivity of interferometry for the extended disk emission, which may make such instruments better suited for the task of finding planets in debris disks than direct imagers.

Given the very small number of planets directly detected so far ($\sim$ a dozen), looking for new ones is tempting. But there is certainly a case to be made for limiting GRAVITY to the role of a follow-up instrument, "only" used to obtain clean medium resolution spectroscopy on planet already detected. The large field-of-view of the direct imagers, such as VLT/SPHERE or Gemini/GPI, simply makes them better suited than GRAVITY for finding planets.

Finally, as exciting and promising as GRAVITY can be for exoplanets, it cannot observe planets at very small separation, which has been a long standing goal of interferometry. At the core of the dual-field technique is a trade of angular resolution for contrast, and despite its $\sim$ 100 m baselines, GRAVITY/dual-field is still limited to separations $\gtrapprox 60$ mas by the 8 m aperture size of the UTs.

Bibliography

O. Absil, D. Defrère, V. Coudé du Foresto, E. Di Folco, A. Mérand, J. C. Augereau, S. Ertel, C. Hanot, P. Kervella, B. Mollier, N. Scott, X. Che, J. D. Monnier, N. Thureau, P. G. Tuthill, T. A. ten Brummelaar, H. A. McAlister, J. Sturmann, L. Sturmann, and N. Turner. A near-infrared interferometric survey of debris-disc stars. III. First statistics based on 42 stars observed with CHARA/FLUOR. *Astronomy and Astrophysics*, 555:A104, July 2013. doi: 10.1051/0004-6361/201321673.

M. Ahmic, B. Croll, and P. Artymowicz. Dust Distribution in the β Pictoris Circumstellar Disks. *the Astrophysical Journal*, 705:529–542, November 2009. doi: 10.1088/0004-637X/705/1/529.

M. Alam, J. Abramczyk, U. Manyam, J. Farroni, and D. Guertin. Performance of optical fibers in space radiation environment. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 10567 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 105672M, Nov 2017. doi: 10.1117/12.2308184.

Y. Alibert, O. Mousis, C. Mordasini, and W. Benz. New Jupiter and Saturn Formation Models Meet Observations. *Astrophysical Journal Letters*, 626:L57–L60, June 2005. doi: 10.1086/431325.

S. M. Andrews and J. P. Williams. Submillimeter Array Observations of Disks in the SR 24 Multiple Star System. *Astrophysical Journal Letters*, 619:L175–L178, February 2005. doi: 10.1086/427325.

J. R. P. Angel and N. J. Woolf. An Imaging Nulling Interferometer to Study Extrasolar Planets. *the Astrophysical Journal*, 475:373–379, January 1997. doi: 10.1086/303529.

D. Apai, G. Schneider, C. A. Grady, M. C. Wyatt, A.-M. Lagrange, M. J. Kuchner, C. J. Stark, and S. H. Lubow. The Inner Disk Structure, Disk-Planet Interactions, and Temporal Evolution in the β Pictoris System: A Two-epoch HST/STIS Coronagraphic Study. *the Astrophysical Journal*, 800:136, February 2015. doi: 10.1088/0004-637X/800/2/136.

H. H. Aumann. Protoplanetary material around nearby stars. In M. D. Papagiannis, editor, *The Search for Extraterrestrial Life: Recent Developments*, volume 112 of *IAU Symposium*, pages 43–50, 1985.

D. E. Backman, F. C. Gillett, and F. C. Witteborn. Infrared observations and thermal models of the Beta Pictoris disk. *the Astrophysical Journal*, 385:670–679, February 1992. doi: 10.1086/170973.

E. K. Baines, G. T. van Belle, T. A. ten Brummelaar, H. A. McAlister, M. Swain, N. H. Turner, L. Sturmann, and J. Sturmann. Direct Measurement of the Radius and Density of the Transiting Exoplanet HD 189733b with the CHARA Array. *Astrophysical Journal Letters*, 661:L195–L198, June 2007. doi: 10.1086/519002.

E. K. Baines, H. A. McAlister, T. A. ten Brummelaar, N. H. Turner, J. Sturmann, L. Sturmann, P. J. Goldfinger, and S. T. Ridgway. CHARA Array Measurements of the Angular Diameters of Exoplanet Host Stars. *the Astrophysical Journal*, 680:728–733, June 2008. doi: 10.1086/588009.

N. P. Ballering, K. Y. L. Su, G. H. Rieke, and A. Gáspár. A Comprehensive Dust Model Applied to the Resolved Beta Pictoris Debris Disk from Optical to Radio Wavelengths. *the Astrophysical Journal*, 823:108, June 2016. doi: 10.3847/0004-637X/823/2/108.

I. Baraffe, G. Chabrier, T. S. Barman, F. Allard, and P. H. Hauschildt. Evolutionary models for cool brown dwarfs and extrasolar giant planets. The case of HD 209458. *Astronomy and Astrophysics*, 402:701–712, May 2003. doi: 10.1051/0004-6361:20030252.

A. Baranne, D. Queloz, M. Mayor, G. Adrianzyk, G. Knispel, D. Kohler, D. Lacroix, J.-P. Meunier, G. Rimbaud, and A. Vin. ELODIE: A spectrograph for accurate radial velocity measurements. *Astronomy and Astrophysics Supplement*, 119:373–390, October 1996.

J.-L. Baudino, B. Bézard, A. Boccaletti, M. Bonnefoy, A.-M. Lagrange, and R. Galicher. Interpreting the photometry and spectroscopy of directly imaged planets: a new atmospheric model applied to β Pictoris b and SPHERE observations. *Astronomy and Astrophysics*, 582:A83, October 2015. doi: 10.1051/0004-6361/201526332.

H. Beust and J. J. Lissauer. The effects of stellar rotation on the absorption spectra of comets orbiting Beta Pictoris. *Astronomy and Astrophysics*, 282:804–810, February 1994.

H. Beust and A. Morbidelli. Falling Evaporating Bodies as a Clue to Outline the Structure of the β Pictoris Young Planetary System. *Icarus*, 143:170–188, January 2000. doi: 10.1006/icar.1999.6238.

H. Beust, A.-M. Lagrange, F. Plazy, and D. Mouillet. The β Pictoris circumstellar disk. XXII. Investigating the model of multiple cometary infalls. *Astronomy and Astrophysics*, 310:181–198, June 1996.

J.-L. Beuzit, D. Mouillet, B. R. Oppenheimer, and J. D. Monnier. Direct Detection of Exoplanets. *Protostars and Planets V*, pages 717–732, 2007.

J.-L. Beuzit, A. Vigan, D. Mouillet, K. Dohlen, R. Gratton, A. Boccaletti, J.-F. Sauvage, H. M. Schmid, M. Langlois, C. Petit, A. Baruffolo, M. Feldt, J. Milli, Z. Wahhaj, L. Abe, U. Anselmi, J. Antichi, R. Barette, J. Baudrand, P. Baudoz, A. Bazzon, P. Bernardi, P. Blanchard, R. Brast, P. Bruno, T. Buey, M. Carbillet, M. Carle, E. Cascone, F. Chapron, G. Chauvin, J. Charton, R. Claudi, A. Costille, V. De Caprio, A. Delboulbé, S. Desidera, C. Dominik, M. Downing, O. Dupuis, C. Fabron, D. Fantinel, G. Farisato, P. Feautrier, E. Fedrigo, T. Fusco, P. Gigan, C. Ginski, J. Girard, E. Giro, D. Gisler, L. Gluck, C. Gry, T. Henning, N. Hubin, E. Hugot, S. Incorvaia, M. Jaquet, M. Kasper, E. Lagadec, A.-M. Lagrange, H. Le Coroller, D. Le Mignant, B. Le Ruyet, G. Lessio, J.-L. Lizon, M. Llored, L. Lundin, F. Madec, Y. Magnard, M. Marteaud, P. Martinez, D. Maurel, F. Ménard, D. Mesa, O. Möller-Nilsson, T. Moulin, C. Moutou, A. Origné, J. Parisot, A. Pavlov, D. Perret, J. Pragt, P. Puget, P. Rabou, J. Ramos, J.-M. Reess, F. Rigal, S. Rochat, R. Roelfsema, G. Rousset, A. Roux, M. Saisse, B. Salasnich, E. Santambrogio, S. Scuderi, D. Segransan, A. Sevin, R. Siebenmorgen, C. Soenke, E. Stadler, M. Suarez, D. Tiphène, M. Turatto, S. Udry, F. Vakili, L. B. F. M. Waters, L. Weber, F. Wildi, G. Zins, and A. Zurlo. SPHERE: the exoplanet imager for the Very Large Telescope. *arXiv e-prints*, February 2019.

P. Bodenheimer. Calculations of the early evolution of Jupiter. *Icarus*, 23:319–325, November 1974. doi: 10.1016/0019-1035(74)90050-5.

P. Bodenheimer, A. S. Grossman, W. M. Decampli, G. Marcy, and J. B. Pollack. Calculations of the evolution of the giant planets. *Icarus*, 41:293–308, February 1980. doi: 10.1016/0019-1035(80)90012-3.

M. Bonnefoy, A.-M. Lagrange, A. Boccaletti, G. Chauvin, D. Apai, F. Allard, D. Ehrenreich, J. H. V. Girard, D. Mouillet, D. Rouan, D. Gratadour, and M. Kasper. High angular resolution detection of β Pictoris b at 2.18 μm. *Astronomy and Astrophysics*, 528:L15, April 2011. doi: 10.1051/0004-6361/201016224.

M. Bonnefoy, A. Boccaletti, A.-M. Lagrange, F. Allard, C. Mordasini, H. Beust, G. Chauvin, J. H. V. Girard, D. Homeier, D. Apai, S. Lacour, and D. Rouan. The near-infrared spectral energy distribution of β Pictoris b. *Astronomy and Astrophysics*, 555:A107, July 2013. doi: 10.1051/0004-6361/201220838.

M. Bonnefoy, G.-D. Marleau, R. Galicher, H. Beust, A.-M. Lagrange, J.-L. Baudino, G. Chauvin, S. Borgniet, N. Meunier, J. Rameau, A. Boccaletti, A. Cumming, C. Helling, D. Homeier, F. Allard, and P. Delorme. Physical and orbital properties of β Pictoris b. *Astronomy and Astrophysics*, 567:L9, July 2014. doi: 10.1051/0004-6361/201424041.

A. P. Boss. Giant planet formation by gravitational instability. *Science*, 276:1836–1839, 1997. doi: 10.1126/science.276.5320.1836.

F. Bouchy, S. Udry, M. Mayor, C. Moutou, F. Pont, N. Iribarne, R. da Silva, S. Ilovaisky, D. Queloz, N. C. Santos, D. Ségransan, and S. Zucker. ELODIE metallicity-biased search for transiting Hot Jupiters. II. A very hot Jupiter transiting the bright K star HD 189733. *Astronomy and Astrophysics*, 444:L15–L19, December 2005. doi: 10.1051/0004-6361:200500201.

R. N. Bracewell. Detecting nonsolar planets by spinning infrared interferometer. *Nature*, 274:780, August 1978. doi: 10.1038/274780a0.

A. Brandeker, G. Cataldi, G. Olofsson, B. Vandenbussche, B. Acke, M. J. Barlow, J. A. D. L. Blommaert, M. Cohen, W. R. F. Dent, C. Dominik, J. Di Francesco, M. Fridlund, W. K. Gear, A. M. Glauser, J. S. Greaves, P. M. Harvey, A. M. Heras, M. R. Hogerheijde, W. S. Holland, R. Huygen, R. J. Ivison, S. J. Leeks, T. L. Lim, R. Liseau, B. C. Matthews, E. Pantin, G. L. Pilbratt, P. Royer, B. Sibthorpe, C. Waelkens, and H. J. Walker. Herschel detects oxygen in the β Pictoris debris disk. *Astronomy and Astrophysics*, 591:A27, June 2016. doi: 10.1051/0004-6361/201628395.

A. Cassan, D. Kubas, J.-P. Beaulieu, M. Dominik, K. Horne, J. Greenhill, J. Wambsganss, J. Menzies, A. Williams, U. G. Jørgensen, A. Udalski, D. P. Bennett, M. D. Albrow, V. Batista, S. Brillant, J. A. R. Caldwell, A. Cole, C. Coutures, K. H. Cook, S. Dieters, D. Dominis Prester, J. Donatowicz, P. Fouqué, K. Hill, N. Kains, S. Kane, J.-B. Marquette, R. Martin, K. R. Pollard, K. C. Sahu, C. Vinter, D. Warren, B. Watson, M. Zub, T. Sumi, M. K. Szymański, M. Kubiak, R. Poleski, I. Soszynski, K. Ulaczyk, G. Pietrzyński, and Ł. Wyrzykowski. One or more bound planets per Milky Way star from microlensing observations. *Nature*, 481:167–169, January 2012. doi: 10.1038/nature10684.

G. Cataldi, A. Brandeker, G. Olofsson, B. Larsson, R. Liseau, J. Blommaert, M. Fridlund, R. Ivison, E. Pantin, B. Sibthorpe, B. Vandenbussche, and Y. Wu. Herschel/HIFI observations of ionised carbon in the β Pictoris debris disk. *Astronomy and Astrophysics*, 563: A66, March 2014. doi: 10.1051/0004-6361/201323126.

CCSDS Secretariat. *Space Packet Protocol.* CCSDS Secretariat, 2003.

B. Charnay, B. Bézard, J.-L. Baudino, M. Bonnefoy, A. Boccaletti, and R. Galicher. A Self-consistent Cloud Model for Brown Dwarfs and Young Giant Exoplanets: Comparison with Photometric and Spectroscopic Observations. *the Astrophysical Journal*, 854:172, February 2018. doi: 10.3847/1538-4357/aaac7d.

C. H. Chen, A. Li, C. Bohac, K. H. Kim, D. M. Watson, J. van Cleve, J. Houck, K. Stapelfeldt, M. W. Werner, G. Rieke, K. Su, M. Marengo, D. Backman, C. Beichman, and G. Fazio. The Dust and Gas Around β Pictoris. *the Astrophysical Journal*, 666:466–474, September 2007. doi: 10.1086/519989.

J. Chilcote, T. Barman, M. P. Fitzgerald, J. R. Graham, J. E. Larkin, B. Macintosh, B. Bauman, A. S. Burrows, A. Cardwell, R. J. De Rosa, D. Dillon, R. Doyon, J. Dunn, D. Erikson, D. Gavel, S. J. Goodsell, M. Hartung, P. Hibon, P. Ingraham, P. Kalas, Q. Konopacky, J. Maire, F. Marchis, M. S. Marley, C. Marois, M. Millar-Blanchaer, K. Morzinski, A. Norton, R. Oppenheimer, D. Palmer, J. Patience, M. Perrin, L. Poyneer, L. Pueyo, F. T. Rantakyrö, N. Sadakuni, L. Saddlemyer, D. Savransky, A. Serio, A. Sivaramakrishnan, I. Song, R. Soummer, S. Thomas, J. K. Wallace, S. Wiktorowicz, and S. Wolff. The First H-band Spectrum of the Giant Planet β Pictoris b. *Astrophysical Journal Letters*, 798:L3, January 2015. doi: 10.1088/2041-8205/798/1/L3.

J. Chilcote, L. Pueyo, R. J. De Rosa, J. Vargas, B. Macintosh, V. P. Bailey, T. Barman, B. Bauman, S. Bruzzone, J. Bulger, A. S. Burrows, A. Cardwell, C. H. Chen, T. Cotten, D. Dillon, R. Doyon, Z. H. Draper, G. Duchêne, J. Dunn, D. Erikson, M. P. Fitzgerald, K. B. Follette, D. Gavel, S. J. Goodsell, J. R. Graham, A. Z. Greenbaum, M. Hartung, P. Hibon, L.-W. Hung, P. Ingraham, P. Kalas, Q. Konopacky, J. E. Larkin, J. Maire, F. Marchis, M. S. Marley, C. Marois, S. Metchev, M. A. Millar-Blanchaer, K. M. Morzinski, E. L. Nielsen, A. Norton, R. Oppenheimer, D. Palmer, J. Patience, M. Perrin, L. Poyneer, A. Rajan, J. Rameau, F. T. Rantakyrö, N. Sadakuni, L. Saddlemyer, D. Savransky, A. C. Schneider, A. Serio, A. Sivaramakrishnan, I. Song, R. Soummer, S. Thomas, J. K. Wallace, J. J. Wang, K. Ward-Duong, S. Wiktorowicz, and S. Wolff. 1-2.4 μm Near-IR Spectrum of the Giant Planet β Pictoris b Obtained with the Gemini Planet Imager. *Astronomical Journal*, 153: 182, April 2017. doi: 10.3847/1538-3881/aa63e9.

I. A. Crawford, H. Beust, and A.-M. Lagrange. Detection of a strong transient blueshifted absorption component in the Beta Pictoris disc. *Monthly Notices of the Royal Astronomical Society*, 294:L31–L34, February 1998. doi: 10.1046/j.1365-8711.1998.01373.x.

F. Crifo, A. Vidal-Madjar, R. Lallement, R. Ferlet, and M. Gerbaldi. β Pictoris revisited by Hipparcos. Star properties. *Astronomy and Astrophysics*, 320:L29–L32, April 1997.

T. Currie, A. Burrows, N. Madhusudhan, M. Fukagawa, J. H. Girard, R. Dawson, R. Murray-Clay, S. Kenyon, M. Kuchner, S. Matsumura, R. Jayawardhana, J. Chambers, and B. Bromley. A Combined Very Large Telescope and Gemini Study of the Atmosphere of the Directly Imaged Planet, β Pictoris b. *the Astrophysical Journal*, 776:15, October 2013. doi: 10.1088/0004-637X/776/1/15.

W. Danchi, V. Bailey, G. Bryden, D. Defrère, C. Haniff, P. Hinz, G. Kennedy, B. Mennesson, R. Millan-Gabet, G. Rieke, A. Roberge, E. Serabyn, A. Skemer, K. Stapelfeldt, A. Weinberger, and M. Wyatt. The LBTI hunt for observable signatures of terrestrial systems (HOSTS) survey: a key NASA science program on the road to exoplanet imaging missions. In *OII IV*, volume 9146, page 914607, July 2014. doi: 10.1117/12.2056681.

W. M. Decampli and A. G. W. Cameron. Structure and evolution of isolated giant gaseous protoplanets. *Icarus*, 38:367–391, June 1979. doi: 10.1016/0019-1035(79)90193-3.

D. Defrère, J. Lebreton, J.-B. Le Bouquin, A.-M. Lagrange, O. Absil, J.-C. Augereau, J.-P. Berger, E. di Folco, S. Ertel, J. Kluska, G. Montagnier, R. Millan-Gabet, W. Traub, and G. Zins. Hot circumstellar material resolved around <ASTROBJ>β Pic</ASTROBJ> with VLTI/PIONIER. *Astronomy and Astrophysics*, 546:L9, October 2012. doi: 10.1051/0004-6361/201220287.

D. Defrère, P. M. Hinz, A. J. Skemer, G. M. Kennedy, V. P. Bailey, W. F. Hoffmann, B. Mennesson, R. Millan-Gabet, W. C. Danchi, O. Absil, P. Arbo, C. Beichman, G. Brusa, G. Bryden, E. C. Downey, O. Durney, S. Esposito, A. Gaspar, P. Grenz, C. Haniff, J. M. Hill, J. Lebreton, J. M. Leisenring, J. R. Males, L. Marion, T. J. McMahon, M. Montoya, K. M. Morzinski, E. Pinna, A. Puglisi, G. Rieke, A. Roberge, E. Serabyn, R. Sosa, K. Stapelfeldt, K. Su, V. Vaitheeswaran, A. Vaz, A. J. Weinberger, and M. C. Wyatt. First-light LBT Nulling Interferometric Observations: Warm Exozodiacal Dust Resolved within a Few AU of η Crv. *the Astrophysical Journal*, 799:42, January 2015. doi: 10.1088/0004-637X/799/1/42.

D. Defrère, M. Ireland, O. Absil, J.-P. Berger, W. C. Danchi, S. Ertel, A. Gallenne, F. Hénault, P. Hinz, E. Huby, S. Kraus, L. Labadie, J.-B. Le Bouquin, G. Martin, A. Matter, B. Mennesson, A. Mérand, S. Minardi, J. D. Monnier, B. Norris, G. Orban de Xivry, E. Pedretti, J.-U. Pott, M. Reggiani, E. Serabyn, J. Surdej, K. R. W. Tristram, and J. Woillez. Hi-5: a potential high-contrast thermal near-infrared imager for the VLTI. In *Optical and Infrared Interferometry and Imaging VI*, volume 10701 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 107010U, July 2018. doi: 10.1117/12.2313700.

W. R. F. Dent, M. C. Wyatt, A. Roberge, J.-C. Augereau, S. Casassus, S. Corder, J. S. Greaves, I. de Gregorio-Monsalvo, A. Hales, A. P. Jackson, A. M. Hughes, A.-M. Lagrange, B. Matthews, and D. Wilner. Molecular Gas Clumps from the Destruction of Icy Bodies in the β Pictoris Debris Disk. *Science*, 343:1490–1492, March 2014. doi: 10.1126/science.1248726.

E. Di Folco, F. Thévenin, P. Kervella, A. Domiciano de Souza, V. Coudé du Foresto, D. Ségransan, and P. Morel. VLTI near-IR interferometric observations of Vega-like stars. Radius and age of α PsA, β Leo, β Pic, ϵ Eri and τ Cet. *Astronomy and Astrophysics*, 426:601–617, November 2004. doi: 10.1051/0004-6361:20047189.

C. Eistrup, C. Walsh, and E. F. van Dishoeck. Setting the volatile composition of (exo)planet-building material. Does chemical evolution in disk midplanes matter? *Astronomy and Astrophysics*, 595:A83, November 2016. doi: 10.1051/0004-6361/201628509.

C. Eistrup, C. Walsh, and E. F. van Dishoeck. Molecular abundances and C/O ratios in chemically evolving planet-forming disk midplanes. *Astronomy and Astrophysics*, 613:A14, May 2018. doi: 10.1051/0004-6361/201731302.

S. Ertel, O. Absil, D. Defrère, J. B. Le Bouquin, J. C. Augereau, L. Marion, N. Blind, A. Bonsor, G. Bryden, J. Lebreton, and J. Milli. A near-infrared interferometric survey

of debris-disk stars. IV. An unbiased sample of 92 southern stars observed in H band with VLTI/PIONIER. *Astronomy and Astrophysics*, 570:A128, October 2014. doi: 10.1051/0004-6361/201424438.

S. Ertel, D. Defrère, P. Hinz, B. Mennesson, G. M. Kennedy, W. C. Danchi, C. Gelino, J. M. Hill, W. F. Hoffmann, G. Rieke, A. Shannon, E. Spalding, J. M. Stone, A. Vaz, A. J. Weinberger, P. Willems, O. Absil, P. Arbo, V. P. Bailey, C. Beichman, G. Bryden, E. C. Downey, O. Durney, S. Esposito, A. Gaspar, P. Grenz, C. A. Haniff, J. M. Leisenring, L. Marion, T. J. McMahon, R. Millan-Gabet, M. Montoya, K. M. Morzinski, E. Pinna, J. Power, A. Puglisi, A. Roberge, E. Serabyn, A. J. Skemer, K. Stapelfeldt, K. Y. L. Su, V. Vaitheeswaran, and M. C. Wyatt. The HOSTS Survey—Exozodiacal Dust Measurements for 30 Stars. *Astronomical Journal*, 155:194, May 2018a. doi: 10.3847/1538-3881/aab717.

S. Ertel, G. M. Kennedy, D. Defrère, P. Hinz, A. B. Shannon, B. Mennesson, W. C. Danchi, C. Gelino, J. M. Hill, W. F. Hoffman, G. Rieke, E. Spalding, J. M. Stone, A. Vaz, A. J. Weinberger, P. Willems, O. Absil, P. Arbo, V. P. Bailey, C. Beichman, G. Bryden, E. C. Downey, O. Durney, S. Esposito, A. Gaspar, P. Grenz, C. A. Haniff, J. M. Leisenring, L. Marion, T. J. McMahon, R. Millan-Gabet, M. Montoya, K. M. Morzinski, E. Pinna, J. Power, A. Puglisi, A. Roberge, E. Serabyn, A. J. Skemer, K. Stapelfeldt, K. Y. L. Su, V. Vaitheeswaran, and M. C. Wyatt. The HOSTS survey for exo-zodiacal dust: preliminary results and future prospects. In *SPIE Conf. Series*, volume 10698, page 106981V, July 2018b. doi: 10.1117/12.2313685.

R. Fernández, A. Brandeker, and Y. Wu. Braking the Gas in the β Pictoris Disk. *the Astrophysical Journal*, 643:509–522, May 2006. doi: 10.1086/500788.

Tobias Flecht. Thermal modelling of the picsat nanosatellite platform and synergetic prestudies of the circus nanosatellite. Master's thesis, Lulea University of Technology, 2016.

J. J. Fortney, M. S. Marley, D. Saumon, and K. Lodders. Synthetic Spectra and Colors of Young Giant Planet Atmospheres: Effects of Initial Conditions and Atmospheric Metallicity. *the Astrophysical Journal*, 683:1104–1116, August 2008. doi: 10.1086/589942.

F. Freistetter, A. V. Krivov, and T. Löhne. Planets of β Pictoris revisited. *Astronomy and Astrophysics*, 466:389–393, April 2007. doi: 10.1051/0004-6361:20066746.

F. Galland, A.-M. Lagrange, S. Udry, A. Chelli, F. Pepe, J.-L. Beuzit, and M. Mayor. Extrasolar planets and brown dwarfs around A-F type stars. III. β Pictoris: looking for planets, finding pulsations. *Astronomy and Astrophysics*, 447:355–359, February 2006. doi: 10.1051/0004-6361:20054080.

D. A. Golimowski, S. T. Durrance, and M. Clampin. Coronagraphic imaging of the Beta Pictoris circumstellar disk - Evidence of changing disk structure within 100 AU. *Astrophysical Journal Letters*, 411:L41–L44, July 1993. doi: 10.1086/186907.

D. A. Golimowski, D. R. Ardila, J. E. Krist, M. Clampin, H. C. Ford, G. D. Illingworth, F. Bartko, N. Benítez, J. P. Blakeslee, R. J. Bouwens, L. D. Bradley, T. J. Broadhurst, R. A. Brown, C. J. Burrows, E. S. Cheng, N. J. G. Cross, R. Demarco, P. D. Feldman, M. Franx, T. Goto, C. Gronwall, G. F. Hartig, B. P. Holden, N. L. Homeier, L. Infante, M. J. Jee, R. A. Kimble, M. P. Lesser, A. R. Martel, S. Mei, F. Menanteau, G. R. Meurer, G. K. Miley, V. Motta, M. Postman, P. Rosati, M. Sirianni, W. B. Sparks, H. D. Tran, Z. I. Tsvetanov, R. L. White, W. Zheng, and A. W. Zirm. Hubble Space Telescope ACS

Multiband Coronagraphic Imaging of the Debris Disk around β Pictoris. *Astronomical Journal*, 131:3109–3130, June 2006. doi: 10.1086/503801.

Gravity Collaboration, R. Abuter, M. Accardo, A. Amorim, N. Anugu, G. Ávila, N. Azouaoui, M. Benisty, J. P. Berger, N. Blind, H. Bonnet, P. Bourget, W. Brandner, R. Brast, A. Buron, L. Burtscher, F. Cassaing, F. Chapron, É. Choquet, Y. Clénet, C. Collin, V. Coudé Du Foresto, W. de Wit, P. T. de Zeeuw, C. Deen, F. Delplancke-Ströbele, R. Dembet, F. Derie, J. Dexter, G. Duvert, M. Ebert, A. Eckart, F. Eisenhauer, M. Esselborn, P. Fédou, G. Finger, P. Garcia, C. E. Garcia Dabo, R. Garcia Lopez, E. Gendron, R. Genzel, S. Gillessen, F. Gonte, P. Gordo, M. Grould, U. Grözinger, S. Guieu, P. Haguenauer, O. Hans, X. Haubois, M. Haug, F. Haussmann, T. Henning, S. Hippler, M. Horrobin, A. Huber, Z. Hubert, N. Hubin, C. A. Hummel, G. Jakob, A. Janssen, L. Jochum, L. Jocou, A. Kaufer, S. Kellner, S. Kendrew, L. Kern, P. Kervella, M. Kiekebusch, R. Klein, Y. Kok, J. Kolb, M. Kulas, S. Lacour, V. Lapeyrère, B. Lazareff, J.-B. Le Bouquin, P. Lèna, R. Lenzen, S. Lévêque, M. Lippa, Y. Magnard, L. Mehrgan, M. Mellein, A. Mérand, J. Moreno-Ventas, T. Moulin, E. Müller, F. Müller, U. Neumann, S. Oberti, T. Ott, L. Pallanca, J. Panduro, L. Pasquini, T. Paumard, I. Percheron, K. Perraut, G. Perrin, A. Pflüger, O. Pfuhl, T. Phan Duc, P. M. Plewa, D. Popovic, S. Rabien, A. Ramírez, J. Ramos, C. Rau, M. Riquelme, R.-R. Rohloff, G. Rousset, J. Sanchez-Bermudez, S. Scheithauer, M. Schöller, N. Schuhler, J. Spyromilio, C. Straubmeier, E. Sturm, M. Suarez, K. R. W. Tristram, N. Ventura, F. Vincent, I. Waisberg, I. Wank, J. Weber, E. Wieprecht, M. Wiest, E. Wiezorrek, M. Wittkowski, J. Woillez, B. Wolff, S. Yazici, D. Ziegler, and G. Zins. First light for GRAVITY: Phase referencing optical interferometry for the Very Large Telescope Interferometer. *Astronomy and Astrophysics*, 602:A94, June 2017. doi: 10.1051/0004-6361/201730838.

Gravity Collaboration, S. Lacour, M. Nowak, J. Wang, O. Pfuhl, F. Eisenhauer, R. Abuter, A. Amorim, N. Anugu, M. Benisty, J. P. Berger, H. Beust, N. Blind, M. Bonnefoy, H. Bonnet, P. Bourget, W. Brandner, A. Buron, C. Collin, B. Charnay, F. Chapron, Y. Clénet, V. Coudé Du Foresto, P. T. de Zeeuw, C. Deen, R. Dembet, J. Dexter, G. Duvert, A. Eckart, N. M. Förster Schreiber, P. Fédou, P. Garcia, R. Garcia Lopez, F. Gao, E. Gendron, R. Genzel, S. Gillessen, P. Gordo, A. Greenbaum, M. Habibi, X. Haubois, F. Haußmann, T. Henning, S. Hippler, M. Horrobin, Z. Hubert, A. Jimenez Rosales, L. Jocou, S. Kendrew, P. Kervella, J. Kolb, A.-M. Lagrange, V. Lapeyrère, J.-B. Le Bouquin, P. Léna, M. Lippa, R. Lenzen, A.-L. Maire, P. Mollière, T. Ott, T. Paumard, K. Perraut, G. Perrin, L. Pueyo, S. Rabien, A. Ramírez, C. Rau, G. Rodríguez-Coira, G. Rousset, J. Sanchez-Bermudez, S. Scheithauer, N. Schuhler, O. Straub, C. Straubmeier, E. Sturm, L. J. Tacconi, F. Vincent, E. F. van Dishoeck, S. von Fellenberg, I. Wank, I. Waisberg, F. Widmann, E. Wieprecht, M. Wiest, E. Wiezorrek, J. Woillez, S. Yazici, D. Ziegler, and G. Zins. First direct detection of an exoplanet by optical interferometry. Astrometry and K-band spectroscopy of HR 8799 e. *Astronomy and Astrophysics*, 623:L11, March 2019. doi: 10.1051/0004-6361/201935253.

R. O. Gray, C. J. Corbally, R. F. Garrison, M. T. McFadden, E. J. Bubar, C. E. McGahee, A. A. O'Donoghue, and E. R. Knox. Contributions to the Nearby Stars (NStars) Project: Spectroscopy of Stars Earlier than M0 within 40 pc-The Southern Sample. *Astronomical Journal*, 132:161–170, July 2006. doi: 10.1086/504637.

Peter H. Hauschildt, France Allard, and E. Baron. The NextGen Model Atmosphere Grid for 3000<=T_{eff}<=10,000 K. *the Astrophysical Journal*, 512(1):377–385, Feb 1999. doi: 10.1086/306745.

S. R. Heap, D. J. Lindler, T. M. Lanz, R. H. Cornett, I. Hubeny, S. P. Maran, and B. Woodgate. Space Telescope Imaging Spectrograph Coronagraphic Observations of β Pictoris. *the Astrophysical Journal*, 539:435–444, August 2000. doi: 10.1086/309188.

R. Helled and G. Schubert. Core formation in giant gaseous protoplanets. *Icarus*, 198:156–162, November 2008. doi: 10.1016/j.icarus.2008.08.002.

R. Helled and G. Schubert. Heavy-element Enrichment of a Jupiter-mass Protoplanet as a Function of Orbital Location. *the Astrophysical Journal*, 697:1256–1262, June 2009. doi: 10.1088/0004-637X/697/2/1256.

R. Helled, M. Podolak, and A. Kovetz. Planetesimal capture in the disk instability model. *Icarus*, 185:64–71, November 2006. doi: 10.1016/j.icarus.2006.06.011.

H. J. Hoeijmakers, H. Schwarz, I. A. G. Snellen, R. J. de Kok, M. Bonnefoy, G. Chauvin, A. M. Lagrange, and J. H. Girard. Medium-resolution integral-field spectroscopy for high-contrast exoplanet imaging. Molecule maps of the β Pictoris system with SINFONI. *Astronomy and Astrophysics*, 617:A144, October 2018. doi: 10.1051/0004-6361/201832902.

H. Holweger, M. Hempel, T. van Thiel, and A. Kaufer. The surface composition of Beta Pictoris. *Astronomy and Astrophysics*, 320:L49–L52, April 1997.

O. Hubickyj, P. Bodenheimer, and J. J. Lissauer. Accretion of the gaseous envelope of Jupiter around a 5 10 Earth-mass core. *Icarus*, 179:415–431, December 2005. doi: 10.1016/j.icarus.2005.06.021.

R. C. Jennison. A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent. *Monthly Notices of the Royal Astronomical Society*, 118:276, 1958. doi: 10.1093/mnras/118.3.276.

P. Kalas and D. Jewitt. Asymmetries in the Beta Pictoris Dust Disk. *Astronomical Journal*, 110:794, August 1995. doi: 10.1086/117565.

C. Karmann, H. Beust, and J. Klinger. The physico-chemical history of Falling Evaporating Bodies around beta Pictoris: investigating the presence of volatiles. *Astronomy and Astrophysics*, 372:616–626, June 2001. doi: 10.1051/0004-6361:20010528.

M. A. Kenworthy and E. E. Mamajek. Modeling Giant Extrasolar Ring Systems in Eclipse and the Case of J1407b: Sculpting by Exomoons? *the Astrophysical Journal*, 800:126, February 2015. doi: 10.1088/0004-637X/800/2/126.

P. Kervella, F. Thévenin, P. Morel, J. Provost, G. Berthomieu, D. Sé Gransan, D. Queloz, P. Bordé, E. di Folco, and T. Forveille. VINCI/VLTI Observations of Main Sequence Stars. In A. K. Dupree and A. O. Benz, editors, *Stars as Suns : Activity, Evolution and Planets*, volume 219 of *IAU Symposium*, page 80, January 2004.

F. Kiefer, A. Lecavelier des Etangs, J. Boissier, A. Vidal-Madjar, H. Beust, A.-M. Lagrange, G. Hébrard, and R. Ferlet. Two families of exocomets in the β Pictoris system. *Nature*, 514:462–464, October 2014. doi: 10.1038/nature13849.

F. Kiefer, A. Vidal-Madjar, A. Lecavelier des Etangs, V. Bourrier, D. Ehrenreich, R. Ferlet, G. Hébrard, and P. A. Wilson. Fe I in the β Pictoris circumstellar gas disk. II. Time variations in the circumstellar iron gas. *Astronomy and Astrophysics*, 621:A58, January 2019. doi: 10.1051/0004-6361/201834274.

B Koehler. Vlti tutorial: Vlti concept and technical aspects. Technical report, ESO Garching, 2001.

C. Koen. δ Scuti pulsations in β Pictoris. *Monthly Notices of the Royal Astronomical Society*, 341:1385–1387, June 2003. doi: 10.1046/j.1365-8711.2003.06509.x.

Q. M. Konopacky, T. S. Barman, B. A. Macintosh, and C. Marois. Detection of Carbon Monoxide and Water Absorption Lines in an Exoplanet Atmosphere. *Science*, 339:1398–1401, March 2013. doi: 10.1126/science.1232003.

Q. Kral, M. Wyatt, R. F. Carswell, J. E. Pringle, L. Matrà, and A. Juhász. A self-consistent model for the evolution of the gas produced in the debris disc of β Pictoris. *Monthly Notices of the Royal Astronomical Society*, 461:845–858, September 2016. doi: 10.1093/mnras/stw1361.

S. Lacour, V. Lapeyrère, L. Gauchet, S. Arroud, R. Gourgues, G. Martin, S. Heidmann, X. Haubois, and G. Perrin. CubeSats as pathfinders for planetary detection: the FIRST-S satellite. In *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, volume 9143 of *Proceedings of the SPIE Conference*, page 91432N, August 2014. doi: 10.1117/12.2057381.

S. Lacour, M. Nowak, P. Bourget, F. Vincent, A. Kellerer, V. Lapeyrère, L. David, A. Le Tiec, O. Straub, and J. Woillez. SAGE: using CubeSats for gravitational wave detection. In *Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray*, volume 10699 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 106992R, July 2018. doi: 10.1117/12.2313639.

S. Lacour, R. Dembet, R. Abuter, P. Fédou, G. Perrin, É. Choquet, O. Pfuhl, F. Eisenhauer, J. Woillez, F. Cassaing, E. Wieprecht, T. Ott, E. Wiezorrek, K. R. W. Tristram, B. Wolff, A. Ramírez, X. Haubois, K. Perraut, C. Straubmeier, W. Brand ner, and A. Amorim. The GRAVITY fringe tracker. *Astronomy and Astrophysics*, 624:A99, Apr 2019. doi: 10.1051/0004-6361/201834981.

P. O. Lagage and E. Pantin. Dust depletion in the inner disk of β Pictoris as a possible indicator of planets. *Nature*, 369:628–630, June 1994. doi: 10.1038/369628a0.

A.-M. Lagrange, D. Gratadour, G. Chauvin, T. Fusco, D. Ehrenreich, D. Mouillet, G. Rousset, D. Rouan, F. Allard, É. Gendron, J. Charton, L. Mugnier, P. Rabou, J. Montri, and F. Lacombe. A probable giant planet imaged in the β Pictoris disk. VLT/NaCo deep L'-band imaging. *Astronomy and Astrophysics*, 493:L21–L25, January 2009a. doi: 10.1051/0004-6361:200811325.

A.-M. Lagrange, M. Kasper, A. Boccaletti, G. Chauvin, D. Gratadour, T. Fusco, D. Ehrenreich, D. Apai, D. Mouillet, and D. Rouan. Constraining the orbit of the possible companion to β Pictoris. New deep imaging observations. *Astronomy and Astrophysics*, 506:927–934, November 2009b. doi: 10.1051/0004-6361/200912098.

A.-M. Lagrange, M. Bonnefoy, G. Chauvin, D. Apai, D. Ehrenreich, A. Boccaletti, D. Gratadour, D. Rouan, D. Mouillet, S. Lacour, and M. Kasper. A Giant Planet Imaged in the Disk of the Young Star β Pictoris. *Science*, 329:57, July 2010. doi: 10.1126/science.1187187.

A.-M. Lagrange, M. Keppler, N. Meunier, J. Lannier, H. Beust, J. Milli, M. Bonnavita, M. Bonnefoy, S. Borgniet, G. Chauvin, P. Delorme, F. Galland, D. Iglesias, F. Kiefer,

S. Messina, A. Vidal-Madjar, and P. A. Wilson. Full exploration of the giant planet population around β Pictoris. *Astronomy and Astrophysics*, 612:A108, May 2018. doi: 10.1051/0004-6361/201730436.

A.-M. Lagrange, A. Boccaletti, M. Langlois, G. Chauvin, R. Gratton, H. Beust, S. Desidera, J. Milli, M. Bonnefoy, A. Cheetham, M. Feldt, M. Meyer, A. Vigan, B. Biller, M. Bonavita, J.-L. Baudino, F. Cantalloube, M. Cudel, S. Daemgen, P. Delorme, V. D'Orazi, J. Girard, C. Fontanive, J. Hagelberg, M. Janson, M. Keppler, T. Koypitova, R. Galicher, J. Lannier, H. Le Coroller, R. Ligi, A.-L. Maire, D. Mesa, S. Messina, A. Müeller, S. Peretti, C. Perrot, D. Rouan, G. Salter, M. Samland, T. Schmidt, E. Sissa, A. Zurlo, J.-L. Beuzit, D. Mouillet, C. Dominik, T. Henning, E. Lagadec, F. Ménard, H.-M. Schmid, M. Turatto, S. Udry, A. J. Bohn, B. Charnay, C. A. Gomez Gonzales, C. Gry, M. Kenworthy, Q. Kral, C. Mordasini, C. Moutou, G. van der Plas, J. E. Schlieder, L. Abe, J. Antichi, A. Baruffolo, P. Baudoz, J. Baudrand, P. Blanchard, A. Bazzon, T. Buey, M. Carbillet, M. Carle, J. Charton, E. Cascone, R. Claudi, A. Costille, A. Deboulbe, V. De Caprio, K. Dohlen, D. Fantinel, P. Feautrier, T. Fusco, P. Gigan, E. Giro, D. Gisler, L. Gluck, N. Hubin, E. Hugot, M. Jaquet, M. Kasper, F. Madec, Y. Magnard, P. Martinez, D. Maurel, D. Le Mignant, O. Möller-Nilsson, M. Llored, T. Moulin, A. Origné, A. Pavlov, D. Perret, C. Petit, J. Pragt, J. Szulagyi, and F. Wildi. Post-conjunction detection of β Pictoris b with VLT/SPHERE. *Astronomy and Astrophysics*, 621:L8, January 2019. doi: 10.1051/0004-6361/201834302.

H. J. G. L. M. Lamers, A. Lecavelier Des Etangs, and A. Vidal-Madjar. Beta Pictoris light variations. II. Scattering by a dust cloud. *Astronomy and Astrophysics*, 328:321–330, December 1997.

T. Lanz, S. R. Heap, and I. Hubeny. HST/GHRS Observations of the beta Pictoris System: Basic Parameters of the Age of the System. *Astrophysical Journal Letters*, 447:L41, July 1995. doi: 10.1086/309561.

V. Lapeyrere, P. Kervella, S. Lacour, N. Azouaoui, C. E. Garcia-Dabo, G. Perrin, F. Eisenhauer, K. Perraut, C. Straubmeier, A. Amorim, and W. Brandner. GRAVITY data reduction software. In *Optical and Infrared Interferometry IV*, volume 9146 of *Proceedings of the SPIE Conference*, page 91462D, July 2014. doi: 10.1117/12.2056850.

S. M. Lawler, C. A. Beichman, G. Bryden, D. R. Ciardi, A. M. Tanner, K. Y. L. Su, K. R. Stapelfeldt, C. M. Lisse, and D. E. Harker. Explorations Beyond the Snow Line: Spitzer/IRS Spectra of Debris Disks Around Solar-type Stars. *the Astrophysical Journal*, 705:89–111, November 2009. doi: 10.1088/0004-637X/705/1/89.

D. Lazzaro, B. Sicardy, F. Roques, and R. Greenberg. Is there a planet around beta Pictoris? Perturbations of a planet circumstellar dust disk. 2: The analytical model. *Icarus*, 108: 59–80, March 1994. doi: 10.1006/icar.1994.1041.

A. Lecavelier Des Etangs and A. Vidal-Madjar. Is β Pictoris b the transiting planet of November 1981? *Astronomy and Astrophysics*, 497:557–562, April 2009. doi: 10.1051/0004-6361/200811528.

A. Lecavelier des Etangs and A. Vidal-Madjar. The orbit of beta Pictoris b as a transiting planet. *Astronomy and Astrophysics*, 588:A60, April 2016. doi: 10.1051/0004-6361/201527631.

A. Lecavelier Des Etangs, M. Deleuil, A. Vidal-Madjar, R. Ferlet, C. Nitschelm, B. Nicolet, and A. M. Lagrange-Henri. β Pictoris: evidence of light variations. *Astronomy and Astrophysics*, 299:557, July 1995.

A. Lecavelier Des Etangs, A. Vidal-Madjar, G. Burki, H. J. G. L. M. Lamers, R. Ferlet, C. Nitschelm, and F. Sevre. Beta Pictoris light variations. I. The planetary hypothesis. *Astronomy and Astrophysics*, 328:311–320, December 1997.

A. Lecavelier Des Etangs, A. Vidal-Madjar, and R. Ferlet. Photometric stellar variation due to extra-solar comets. *Astronomy and Astrophysics*, 343:916–922, March 1999.

G. Ledrew. The Real Starry Sky. *Journal of the Royal Astronomical Society of Canada*, 95: 32, February 2001.

D. Li, C. M. Telesco, and C. M. Wright. The Mineralogy and Structure of the Inner Debris Disk of β Pictoris. *the Astrophysical Journal*, 759:81, November 2012. doi: 10.1088/0004-637X/759/2/81.

J. J Lissauer and D. J. Stevenson. Formation of Giant Planets. In B. Reipurth, D. Jewitt, and K. Keil, editors, *Protostars and Planets V*, pages 591–606, April 2007.

B. Macintosh, J. R. Graham, P. Ingraham, Q. Konopacky, C. Marois, M. Perrin, L. Poyneer, B. Bauman, T. Barman, A. S. Burrows, A. Cardwell, J. Chilcote, R. J. De Rosa, D. Dillon, R. Doyon, J. Dunn, D. Erikson, M. P. Fitzgerald, D. Gavel, S. Goodsell, M. Hartung, P. Hibon, P. Kalas, J. Larkin, J. Maire, F. Marchis, M. S. Marley, J. McBride, M. Millar-Blanchaer, K. Morzinski, A. Norton, B. R. Oppenheimer, D. Palmer, J. Patience, L. Pueyo, F. Rantakyro, N. Sadakuni, L. Saddlemyer, D. Savransky, A. Serio, R. Soummer, A. Sivaramakrishnan, I. Song, S. Thomas, J. K. Wallace, S. Wiktorowicz, and S. Wolff. First light of the Gemini Planet Imager. *Proceedings of the National Academy of Science*, 111:12661–12666, September 2014. doi: 10.1073/pnas.1304215111.

N. Madhusudhan. *Atmospheric Retrieval of Exoplanets*, page 104. 2018. doi: 10.1007/978-3-319-55333-7.

N. Madhusudhan, J. Harrington, K. B. Stevenson, S. Nymeyer, C. J. Campo, P. J. Wheatley, D. Deming, J. Blecic, R. A. Hardy, N. B. Lust, D. R. Anderson, A. Collier-Cameron, C. B. T. Britt, W. C. Bowman, L. Hebb, C. Hellier, P. F. L. Maxted, D. Pollacco, and R. G. West. A high C/O ratio and weak thermal inversion in the atmosphere of exoplanet WASP-12b. *Nature*, 469:64–67, January 2011. doi: 10.1038/nature09602.

N. Madhusudhan, M. A. Amin, and G. M. Kennedy. Toward Chemical Constraints on Hot Jupiter Migration. *Astrophysical Journal Letters*, 794:L12, October 2014. doi: 10.1088/2041-8205/794/1/L12.

N. Madhusudhan, B. Bitsch, A. Johansen, and L. Eriksson. Atmospheric signatures of giant exoplanet formation by pebble accretion. *Monthly Notices of the Royal Astronomical Society*, 469:4102–4115, August 2017. doi: 10.1093/mnras/stx1139.

E. E. Mamajek and C. P. M. Bell. On the age of the β Pictoris moving group. *Monthly Notices of the Royal Astronomical Society*, 445:2169–2180, December 2014. doi: 10.1093/mnras/stu1894.

U. Marboeuf, A. Thiabaud, Y. Alibert, N. Cabral, and W. Benz. From stellar nebula to planetesimals. *Astronomy and Astrophysics*, 570:A35, October 2014. doi: 10.1051/0004-6361/201322207.

G.-D. Marleau, H. Klahr, R. Kuiper, and C. Mordasini. The Planetary Accretion Shock. I. Framework for Radiation-hydrodynamical Simulations and First Results. *the Astrophysical Journal*, 836:221, February 2017. doi: 10.3847/1538-4357/836/2/221.

M. S. Marley, J. J. Fortney, O. Hubickyj, P. Bodenheimer, and J. J. Lissauer. On the Luminosity of Young Jupiters. *the Astrophysical Journal*, 655:541–549, January 2007. doi: 10.1086/509759.

F. Martinache and M. J. Ireland. Kernel-nulling for a robust direct interferometric detection of extrasolar planets. *Astronomy and Astrophysics*, 619:A87, November 2018. doi: 10.1051/0004-6361/201832847.

L. Matrà, M. C. Wyatt, D. J. Wilner, W. R. F. Dent, S. Marino, G. M. Kennedy, and J. Milli. Kuiper Belt-Like Hot and Cold Populations of Planetesimal Inclinations in the β Pictoris Belt Revealed by ALMA. *arXiv e-prints*, February 2019.

L. Mayer, T. Quinn, J. Wadsley, and J. Stadel. Formation of Giant Planets by Fragmentation of Protoplanetary Disks. *Science*, 298:1756–1759, November 2002. doi: 10.1126/science.1077635.

M. Mayor and D. Queloz. A Jupiter-mass companion to a solar-type star. *Nature*, 378:355–359, November 1995. doi: 10.1038/378355a0.

H. A. McAlister, T. A. ten Brummelaar, D. R. Gies, W. Huang, W. G. Bagnuolo, Jr., M. A. Shure, J. Sturmann, L. Sturmann, N. H. Turner, S. F. Taylor, D. H. Berger, E. K. Baines, E. Grundstrom, C. Ogden, S. T. Ridgway, and G. van Belle. First Results from the CHARA Array. I. An Interferometric and Spectroscopic Study of the Fast Rotator α Leonis (Regulus). *the Astrophysical Journal*, 628:439–452, July 2005. doi: 10.1086/430730.

D. Mékarnia, E. Chapellier, T. Guillot, L. Abe, A. Agabi, Y. De Pra, F.-X. Schmider, K. Zwintz, K. B. Stevenson, J. J. Wang, A.-M. Lagrange, L. Bigot, N. Crouzet, Y. Fanteï-Caujolle, J.-M. Christille, and P. Kalas. The δ Scuti pulsations of β Pictoris as observed by ASTEP from Antarctica. *Astronomy and Astrophysics*, 608:L6, December 2017. doi: 10.1051/0004-6361/201732121.

B. Mennesson, E. Serabyn, C. Hanot, S. R. Martin, K. Liewer, and D. Mawet. New Constraints on Companions and Dust within a Few AU of Vega. *the Astrophysical Journal*, 736:14, July 2011. doi: 10.1088/0004-637X/736/1/14.

B. Mennesson, R. Millan-Gabet, E. Serabyn, M. M. Colavita, O. Absil, G. Bryden, M. Wyatt, W. Danchi, D. Defrère, O. Doré, P. Hinz, M. Kuchner, S. Ragland, N. Scott, K. Stapelfeldt, W. Traub, and J. Woillez. Constraining the Exozodiacal Luminosity Function of Main-sequence Stars: Complete Results from the Keck Nuller Mid-infrared Surveys. *the Astrophysical Journal*, 797:119, December 2014. doi: 10.1088/0004-637X/797/2/119.

J. Menu, G. Perrin, E. Choquet, and S. Lacour. Kalman-filter control schemes for fringe tracking. Development and application to VLTI/GRAVITY. *Astronomy and Astrophysics*, 541:A81, May 2012. doi: 10.1051/0004-6361/201218932.

P. Mollière and I. A. G. Snellen. Detecting isotopologues in exoplanet atmospheres using ground-based high-dispersion spectroscopy. *Astronomy and Astrophysics*, 622:A139, February 2019. doi: 10.1051/0004-6361/201834169.

C. Mordasini, R. van Boekel, P. Mollière, T. Henning, and B. Benneke. The Imprint of Exoplanet Formation History on Observable Present-day Spectra of Hot Jupiters. *the Astrophysical Journal*, 832:41, November 2016. doi: 10.3847/0004-637X/832/1/41.

C. Mordasini, G.-D. Marleau, and P. Mollière. Characterization of exoplanets from their formation. III. The statistics of planetary luminosities. *Astronomy and Astrophysics*, 608: A72, December 2017. doi: 10.1051/0004-6361/201630077.

K. M. Morzinski, J. R. Males, A. J. Skemer, L. M. Close, P. M. Hinz, T. J. Rodigas, A. Puglisi, S. Esposito, A. Riccardi, E. Pinna, M. Xompero, R. Briguglio, V. P. Bailey, K. B. Follette, D. Kopon, A. J. Weinberger, and Y.-L. Wu. Magellan Adaptive Optics First-light Observations of the Exoplanet β Pic b. II. 3-5 μm Direct Imaging with MagAO+Clio, and the Empirical Bolometric Luminosity of a Self-luminous Giant Planet. *the Astrophysical Journal*, 815:108, December 2015. doi: 10.1088/0004-637X/815/2/108.

D. Mouillet, J. D. Larwood, J. C. B. Papaloizou, and A. M. Lagrange. A planet on an inclined orbit as an explanation of the warp in the Beta Pictoris disc. *Monthly Notices of the Royal Astronomical Society*, 292:896, December 1997. doi: 10.1093/mnras/292.4.896.

E.-G. Neumann. *Single Mode Fibers Fundamentals*. Springer, 1988.

K. I. Öberg, R. Murray-Clay, and E. A. Bergin. The Effects of Snowlines on C/O in Planetary Atmospheres. *Astrophysical Journal Letters*, 743:L16, December 2011. doi: 10.1088/2041-8205/743/1/L16.

Y. K. Okamoto, H. Kataza, M. Honda, T. Yamashita, T. Onaka, J.-i. Watanabe, T. Miyata, S. Sako, T. Fujiyoshi, and I. Sakon. An early extrasolar planetary system revealed by planetesimal belts in β Pictoris. *Nature*, 431:660–663, October 2004. doi: 10.1038/nature02948.

G. Olofsson, R. Liseau, and A. Brandeker. Widespread Atomic Gas Emission Reveals the Rotation of the β Pictoris Disk. *Astrophysical Journal Letters*, 563:L77–L80, December 2001. doi: 10.1086/338354.

T. Owen, P. Mahaffy, H. B. Niemann, S. Atreya, T. Donahue, A. Bar-Nun, and I. de Pater. A low-temperature origin for the planetesimals that formed Jupiter. *Nature*, 402:269–270, November 1999. doi: 10.1038/46232.

F. Paresce and C. Burrows. Broad-band imaging of the Beta Pictoris circumstellar disk. *Astrophysical Journal Letters*, 319:L23–L25, August 1987. doi: 10.1086/184948.

M. D. Perrin, E. Choquet, A. Greenbaum, B. Ren, J. H. Debes, J. Mazoyer, M. Ygouf, L. Pueyo, J. Aguilar, C. Chen, D. A. Golimowski, D. C. Hines, M. N'Diaye, G. Schneider, R. Soummer, C. Stark, and S. Wolff. HST STIS NICMOS Coronagraphy of Four Debris Disks around Young Solar Analogs. In *AAS Meetings Abst. 227*, volume 227 of *American Astronomical Society Meeting Abstracts*, page 309.03, January 2016.

M. A. C. Perryman, L. Lindegren, J. Kovalevsky, E. Hoeg, U. Bastian, P. L. Bernacca, M. Crézé, F. Donati, M. Grenon, M. Grewing, F. van Leeuwen, H. van der Marel, F. Mignard, C. A. Murray, R. S. Le Poole, H. Schrijver, C. Turon, F. Arenou, M. Froeschlé, and C. S. Petersen. The HIPPARCOS Catalogue. *Astronomy and Astrophysics*, 323:L49–L52, July 1997.

P. Plasson, C. Cuomo, G. Gabriel, N. Gauthier, L. Gueguen, and L. Malac-Allain. GERICOS: A Generic Framework for the Development of On-Board Software. In *DASIA 2016 - Data Systems In Aerospace*, volume 736 of *ESA Special Publication*, page 39, August 2016.

C. M. Pong. On-Orbit Performance & Operation of the Attitude & Pointing Control Subsystems on ASTERIA. In *Poster Session I*, Delivering Mission Success, Small Satellite Conference, pages SSC18–PI–34, August 2018.

K. M. Pontoppidan, C. P. Dullemond, E. F. van Dishoeck, G. A. Blake, A. C. A. Boogert, N. J. Evans, II, J. E. Kessler-Silacci, and F. Lahuis. Ices in the Edge-on Disk CRBR 2422.8-3423: Spitzer Spectroscopy and Monte Carlo Radiative Transfer Modeling. *the Astrophysical Journal*, 622:463–481, March 2005. doi: 10.1086/427688.

S. P. Quanz, M. R. Meyer, M. A. Kenworthy, J. H. V. Girard, M. Kasper, A.-M. Lagrange, D. Apai, A. Boccaletti, M. Bonnefoy, G. Chauvin, P. M. Hinz, and R. Lenzen. First Results from Very Large Telescope NACO Apodizing Phase Plate: 4 μm Images of The Exoplanet β Pictoris b. *Astrophysical Journal Letters*, 722:L49–L53, October 2010. doi: 10.1088/2041-8205/722/1/L49.

A. Roberge, P. D. Feldman, A. J. Weinberger, M. Deleuil, and J.-C. Bouret. Stabilization of the disk around βPictoris by extremely carbon-rich gas. *Nature*, 441:724–726, June 2006. doi: 10.1038/nature04832.

F. Roques, H. Scholl, B. Sicardy, and B. A. Smith. Is there a planet around beta Pictoris? Perturbations of a planet on a circumstellar dust disk. 1: The numerical model. *Icarus*, 108:37–58, March 1994. doi: 10.1006/icar.1994.1040.

J. Sahlmann, D. Ségransan, A. Mérand, N. Zimmerman, R. Abuter, B. Chazelas, F. Delplancke, T. Henning, A. Kaminski, R. Köhler, R. Launhardt, M. Mohler, F. Pepe, D. Queloz, A. Quirrenbach, S. Reffert, C. Schmid, N. Schuhler, and T. Schulze-Hartung. Narrow-angle astrometry with PRIMA. In *Optical and Infrared Interferometry III*, volume 8445 of *Proceedings of the SPIE Conference*, page 84450S, July 2012. doi: 10.1117/12.926214.

B. A. Smith and R. J. Terrile. A circumstellar disk around Beta Pictoris. *Science*, 226:1421–1424, December 1984. doi: 10.1126/science.226.4681.1421.

M. W. Smith, A. Donner, M. Knapp, C. M. Pong, C. Smith, J. Luu, P. Di Pasquale, R. L. Bocchino, B. Campuzano, J. Loveland, C. Colley, A. Babuscia, M. White, J. Krajewski, and S. Seager. On-Orbit Performance & Operation of the Attitude & Pointing Control Subsystems on ASTERIA. In *The Year in Review*, Delivering Mission Success, Small Satellite Conference, pages SSC18–I–08, August 2018.

I. A. G. Snellen and A. G. A. Brown. The mass of the young planet Beta Pictoris b through the astrometric motion of its host star. *Nature Astronomy*, 2:883–886, August 2018. doi: 10.1038/s41550-018-0561-6.

I. A. G. Snellen, B. R. Brandl, R. J. de Kok, M. Brogi, J. Birkby, and H. Schwarz. Fast spin of the young extrasolar planet β Pictoris b. *Nature*, 509:63–65, May 2014. doi: 10.1038/nature13253.

D. S. Spiegel and A. Burrows. Spectral and Photometric Diagnostics of Giant Planet Formation Scenarios. *the Astrophysical Journal*, 745:174, February 2012. doi: 10.1088/0004-637X/745/2/174.

R. Stuik, J. I. Bailey, P. Dorval, G. J. J. Talens, I. Laginja, S. N. Mellon, B. B. D. Lomberg, S. M. Crawford, M. J. Ireland, E. E. Mamajek, and M. A. Kenworthy. bRing: An observatory dedicated to monitoring the β Pictoris b Hill sphere transit. *Astronomy and Astrophysics*, 607:A45, November 2017. doi: 10.1051/0004-6361/201731679.

C. M. Telesco, R. S. Fisher, M. C. Wyatt, S. F. Dermott, T. J. J. Kehoe, S. Novotny, N. Mariñas, J. T. Radomski, C. Packham, J. De Buizer, and T. L. Hayward. Mid-infrared images of β Pictoris and the possible role of planetesimal collisions in the central disk. *Nature*, 433:133–136, January 2005. doi: 10.1038/nature03255.

N. S. van der Bliek, J. Manfroid, and P. Bouchet. Infrared aperture photometry at ESO (1983-1994) and its future use. *Astronomy and Astrophysics Supplement*, 119:547–557, November 1996.

J. I. Vette. The AE-8 trapped electron model environment. *NASA STI/Recon Technical Report N*, 92, November 1991.

A. Vidal-Madjar, F. Kiefer, A. Lecavelier des Etangs, V. Bourrier, D. Ehrenreich, R. Ferlet, G. Hébrard, and P. A. Wilson. Fe I in the β Pictoris circumstellar gas disk. I. Physical properties of the neutral iron gas. *Astronomy and Astrophysics*, 607:A25, October 2017. doi: 10.1051/0004-6361/201630040.

Z. Wahhaj, D. W. Koerner, M. E. Ressler, M. W. Werner, D. E. Backman, and A. I. Sargent. The Inner Rings of β Pictoris. *Astrophysical Journal Letters*, 584:L27–L31, February 2003. doi: 10.1086/346123.

C. Walsh, H. Nomura, and E. van Dishoeck. The molecular composition of the planet-forming regions of protoplanetary disks across the luminosity regime. *Astronomy and Astrophysics*, 582:A88, October 2015. doi: 10.1051/0004-6361/201526751.

J. J. Wang, J. R. Graham, L. Pueyo, P. Kalas, M. A. Millar-Blanchaer, J.-B. Ruffio, R. J. De Rosa, S. M. Ammons, P. Arriaga, V. P. Bailey, T. S. Barman, J. Bulger, A. S. Burrows, A. Cardwell, C. H. Chen, J. K. Chilcote, T. Cotten, M. P. Fitzgerald, K. B. Follette, R. Doyon, G. Duchêne, A. Z. Greenbaum, P. Hibon, L.-W. Hung, P. Ingraham, Q. M. Konopacky, J. E. Larkin, B. Macintosh, J. Maire, F. Marchis, M. S. Marley, C. Marois, S. Metchev, E. L. Nielsen, R. Oppenheimer, D. W. Palmer, R. Patel, J. Patience, M. D. Perrin, L. A. Poyneer, A. Rajan, J. Rameau, F. T. Rantakyrö, D. Savransky, A. Sivaramakrishnan, I. Song, R. Soummer, S. Thomas, G. Vasisht, D. Vega, J. K. Wallace, K. Ward-Duong, S. J. Wiktorowicz, and S. G. Wolff. The Orbit and Transit Prospects for β Pictoris b Constrained with One Milliarcsecond Astrometry. *Astronomical Journal*, 152:97, October 2016. doi: 10.3847/0004-6256/152/4/97.

Alycia J. Weinberger, Geoff Bryden, Grant M. Kennedy, Aki Roberge, Denis Defrère, Philip M. Hinz, Rafael Millan-Gabet, George Rieke, Vanessa P. Bailey, William C. Danchi, Chris Haniff, Bertrand Mennesson, Eugene Serabyn, Andrew J. Skemer, Karl R. Stapelfeldt, and Mark C. Wyatt. Target Selection for the LBTI Exozodi Key Science Program. *The Astrophysical Journal Supplement Series*, 216:24, February 2015. doi: 10.1088/0067-0049/216/2/24.

W. W. Weiss, S. M. Rucinski, A. F. J. Moffat, A. Schwarzenberg-Czerny, O. F. Koudelka, C. C. Grant, R. E. Zee, R. Kuschnig, St. Mochnacki, J. M. Matthews, P. Orleanski, A. Pamyatnykh, A. Pigulski, J. Alves, M. Guedel, G. Handler, G. A. Wade, and K. Zwintz. BRITE-Constellation: Nanosatellites for Precision Photometry of Bright Stars. *Publications of the Astronomical Society of the Pacific*, 126:573, June 2014. doi: 10.1086/677236.

D. Wernham. Optical coatings in space. In *Advances in Optical Thin Films IV*, volume 8168 of *Proceedings of the SPIE Conference*, page 81680F, October 2011. doi: 10.1117/12.902318. URL `https://doi.org/10.1117/12.902318`.

D. J. Wilner, S. M. Andrews, and A. M. Hughes. Millimeter Imaging of the β Pictoris Debris Disk: Evidence for a Planetesimal Belt. *Astrophysical Journal Letters*, 727:L42, February 2011. doi: 10.1088/2041-8205/727/2/L42.

P. A. Wilson, A. Lecavelier des Etangs, A. Vidal-Madjar, V. Bourrier, G. Hébrard, F. Kiefer, H. Beust, R. Ferlet, and A.-M. Lagrange. First detection of hydrogen in the β Pictoris gas disk. *Astronomy and Astrophysics*, 599:A75, March 2017. doi: 10.1051/0004-6361/201629293.

P. A. Wilson, R. Kerr, A. Lecavelier des Etangs, V. Bourrier, A. Vidal-Madjar, F. Kiefer, and I. A. G. Snellen. Detection of nitrogen gas in the β Pictoris circumstellar disc. *Astronomy and Astrophysics*, 621:A121, January 2019. doi: 10.1051/0004-6361/201834346.

J.-W. Xie, A. Brandeker, and Y. Wu. On the Unusual Gas Composition in the β Pictoris Debris Disk. *the Astrophysical Journal*, 762:114, January 2013. doi: 10.1088/0004-637X/762/2/114.

Part V

APPENDICES

Appendix A

Fiber injection in PicSat and GRAVITY

A.1 Single-mode fiber injection

A.1.1 Injection equation

The injection ratio η in an optical fiber is defined as the fraction of the energy entering the telescope which actually ends up propagating into the fiber. It can be calculated by projecting the electric field illuminating the fiber plane onto the fiber's fundamental mode [Neumann, 1988]:

$$\eta = \frac{\left\| \iint E_{\text{fiber}}(x,y) \times E^*_{\text{mode}}(x,y) \mathrm{d}x\mathrm{d}y \right\|^2}{\left\| \iint E_{\text{fiber}} \mathrm{d}x\mathrm{d}y \right\|^2} \tag{A.1}$$

where x, y denotes the focal plane coordinates, E_{fiber} is the electric field in the fiber plane, E_{mode} the fundamental mode of the fiber, and * denotes the complex conjugate.

A.1.2 Electric field in the fiber plane

Depending on the situation, there are two main ways to calculate the electric field in the plane of the fiber: the Kirchoff integral, valid in all possible situations, or a simple Fourier transform, valid in the Fraunhoffer regime.

A.1.2.1 The Kirchoff integral

Here I consider the general scenario of a monochromatic observation at a given wavelength λ, with a telescope of arbitrary geometry (Figure A.1). The telescope is pointed at a star located very far away (i.e. the star can be considered as a point-source). The incident wavefront on the telescope is plane, and the incoming electric field as a function of time t and space coordinates (x, y, z) can be written:

$$E(t,x,y,z) = A(x,y,z)e^{-i\omega t} A(x,y,z) = A_0 e^{-i\frac{2\pi}{\lambda}z} \tag{A.2}$$

The Huyghens-Fresnel principle states that any given point P at the surface of the reflecting surface can be considered as a secondary source, emitting a spherical wavelet, whose complexe amplitude $\mathrm{d}A_P$ is a function of the reflected incoming wave:

$$\mathrm{d}A_P = A_0 e^{i\frac{2\pi}{\lambda}z_P} \mathrm{d}S_P \tag{A.3}$$

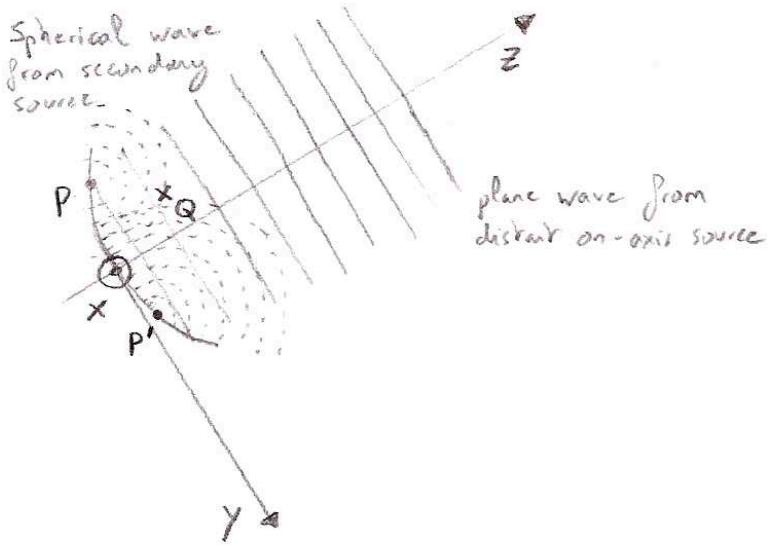


Figure A.1: A telescope of arbitrary geometry observes a monochromatic point-like object.

where $\mathrm{d}S_P$ is the surface element at position P.

At a point $Q = (x_Q, y_Q, z_Q)$ downstream of the reflection, the field produced by the spherical propagation of this wavelet is:

$$\mathrm{d}A_P(t, x_Q, y_Q, z_Q) = \frac{\mathrm{d}A_P}{PQ} e^{-i\frac{2\pi}{\lambda}PQ} \tag{A.4}$$

PQ being the distance from point P to point Q:

$$PQ = \sqrt{(x_Q - x_P)^2 + (y_Q - y_P)^2 + (z_Q - z_P)^2} \tag{A.5}$$

And the total complex amplitude of the field received in Q from the entire surface of the mirror is the coherent sum of the Huyghens-Fresnel wavelets:

$$A_Q = \iint_{\text{mirror}} \frac{A_0 e^{i\frac{2\pi}{\lambda}(z_P - PQ)}}{PQ} \mathrm{d}S_P \tag{A.6}$$

The above integral, which gives the complex field at point Q, is the so-called "Kirchoff integral". It can be used to numerically calculate the electric field in the plane of the fiber.

A.1.3 Fraunhoffer regime and Fourier transform formalism

Electric field

A particular expression of the Kirchoff which is very useful for astronomy can be obtained in the Fraunhoffer regime, which is a combination of two approximations: the paraxial approximation, and the focal plane approximation.

In the paraxial approximation ($z_Q \gg D$), we can write:

$$PQ \simeq z_Q \left(1 + \frac{1}{2}\left(\frac{x_P^2 - 2x_Q x_P}{z_Q^2} + \frac{y_P^2 - 2y_Q y_P}{z_Q^2} - \frac{2z_Q z_P}{z_Q^2}\right)\right) \tag{A.7}$$

The expression for the amplitude received in Q can be developed to first order:

$$A_Q \simeq \iint_{\text{mirror}} \frac{A_0}{z_q} \exp\left\{-i\,\frac{2\pi z_P}{\lambda}\right\} \exp\left\{i\frac{2\pi}{\lambda}\left(\frac{x_P^2+y_P^2}{2z_Q} - \frac{z_Q z_P}{z_Q} - \frac{x_P x_Q}{z_Q} - \frac{y_P y_Q}{z_Q}\right)\right\} \mathrm{d}S_P$$

Reordering some of the terms:

$$A_Q \simeq \frac{A_0}{z_Q} \iint_{\text{mirror}} \exp\left\{i\frac{2\pi}{\lambda}\left(\frac{x_P^2+y_P^2}{2z_Q} - \frac{2z_Q z_P}{z_Q}\right)\right\} \exp\left\{-i\,2\pi\left(\frac{x_P x_Q}{\lambda z_Q} + \frac{y_P y_Q}{\lambda z_Q}\right)\right\} \mathrm{d}S_P$$

At this point it becomes clear why the particular shape of a parabola is so important to astronomical telescopes. If the surface of the mirror is a parabola of focal length f, then we have the relation:

$$z_P = \frac{1}{4f}\left(x_P^2 + y_P^2\right) \tag{A.8}$$

And if we consider the field at a point Q in the focal plane ($z_Q = f$), the whole term in the first exponential vanishes, and we are left with an expression like:

$$A_Q \simeq \frac{A_0}{f} \iint_{\text{mirror}} \exp\left\{-i\,2\pi\left(\frac{x_P x_Q}{\lambda f} + \frac{y_P y_Q}{\lambda f}\right)\right\} \mathrm{d}S_P$$

We can note here that the particular shape of the incoming complex amplitude given in Equation A.3 plays only a minor role in the calculation. The exact same reasoning with a general incoming wavefront $A_P(x_P, y_P, z_P)e^{-i\omega t}$ yields a similar expression:

$$A_Q \simeq \frac{1}{f} \iint_{\text{mirror}} A(x_P, y_P, z_P) \exp\left\{-i\,2\pi\left(\frac{x_P x_Q}{\lambda f} + \frac{y_P y_Q}{\lambda f}\right)\right\} \mathrm{d}S_P$$

If we note that the coordinate z_P is set by x_P and y_P, we can reason in the "pupil plane", and consider that A_P only depends on the two coordinates x_P and y_P, to get the commonly known expression:

$$A_Q \simeq \frac{1}{f} \iint_{\text{pupil}} A(x_P, y_P) \exp\left\{-i\,2\pi\left(\frac{x_P x_Q}{\lambda f} + \frac{y_P y_Q}{\lambda f}\right)\right\} \mathrm{d}S_P$$

In which we recognize the expression of the Fourier Transform (FT) of the pupil plane field:

$$A(x_Q, y_Q) = \mathrm{TF}\left[P(x,y) \times A(x,y)\right]\left(\frac{x_Q}{\lambda f}, \frac{x_Q}{\lambda f}\right) \tag{A.9}$$

Where $P(x,y)$ is the pupil function, defined as (for a circular aperture of diameter D):

$$P(x,y) = \begin{cases} 1 & \text{if} \quad x^2 + y^2 < \frac{D^2}{4} \\ 0 & \text{otherwise} \end{cases} \tag{A.10}$$

Intensity

From this expression of the field at point Q, we can derive an expression for the intensity at (X, Y) in the focal plane of the telescope. To a normalization factor, we have:

$$I(X,Y) \propto \left\| \mathrm{TF}\left[P(x,y) \times A(x,y)\right]\left(\frac{x_Q}{\lambda f}, \frac{x_Q}{\lambda f}\right) \right\|^2 \tag{A.11}$$

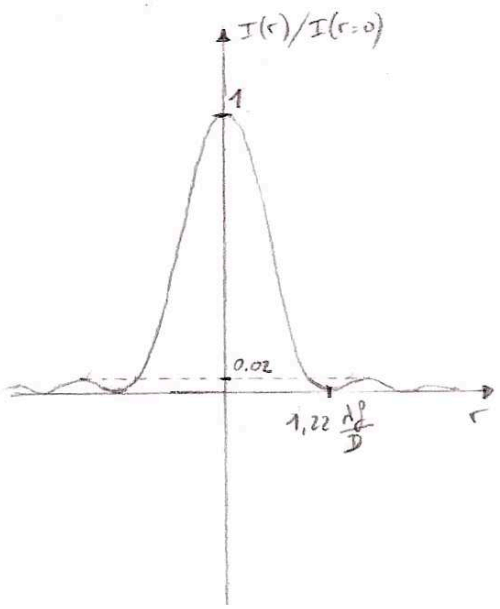


Figure A.2: Illustration of the Point-Spread Function (PSF), or image of a point-source by an astronomical instrument of diameter D.

For a point-source and a non-aberrated wavefront, the distribution of intensity in the image plane is directly given by the square modulus of the Fourier transform of the pupil function:

$$I \propto \|\mathrm{TF}\left[P(x,y)\right]\|^2 \tag{A.12}$$

For a circular aperture, this corresponds the the Airy Pattern:

$$I(X,Y) \propto \left(\frac{J_1\left(\pi\frac{D}{\lambda}\frac{r}{f}\right)}{\pi\frac{D}{\lambda}\frac{r}{f}}\right)^2 \tag{A.13}$$

D being the diameter of the aperture, f the focal length of the telescope, $r = \sqrt{X^2+Y^2}$, and J_1 the first order Bessel function of the first kind. A representation of the function $r \to I(r)$ is given in Figure A.2.

From all this, we see that the image of a point-source by an astronomical telescope is not just a point, but rather a complicated function of the aperture shape. This is usually referred to as the Point-Spread Function (PSF), to emphasize the fact that the diffraction of the light-waves "spreads" the image in the focal plane.

A.1.4 Fiber mode

A.1.4.1 Direct expression

The fundamental mode of a single-mode fiber depends on the exact size and nature of the core and of the cladding, but is usually very well approximated by a real Gaussian mode:

$$A_{\mathrm{mode}}(x,y) \simeq \frac{1}{\pi w_g}\exp\left(-\frac{x^2+y^2}{{w_\mathrm{g}}^2}\right) \tag{A.14}$$

with w_g the fiber mode field diameter, a wavelength dependent quantity which can be written:

$$w_g(\lambda) = w_g(\lambda_{\text{ref}}) \times \sqrt{\frac{\lambda}{\lambda_{\text{ref}}}} \tag{A.15}$$

This gaussian expression of the fiber fundamental mode is valid in the plane of the fiber itself, and can be used in combination with the Kirchoff integral to calculate the injection ratio.

A.1.4.2 Fiber illumination of the pupil plane

In the Fraunhoffer regime, the direct expression of the gaussian mode of the fiber can also be used in combination with the Fourier transform formalism to calculate the injection ratio. But in this particular regime, there is an alternate expression which is often very useful: the pupil plane overlap integral.

Consider a telescope of diameter D, focal length f, and put a single mode fiber in the focal plane of the telescope. The electric field in the focal plane of the telescope has already been calculated, and is given in Equation A.9. At position (X, Y) in the image plane, the complex amplitude of the field is given by:

$$A_{\text{focal}}(X, Y) = \text{TF}\left[P(x, y) A_{\text{pupil}}(x, y)\right] \left(\frac{X}{\lambda f}, \frac{Y}{\lambda f}\right)$$

where $P(x, y)$ is the pupil aperture function, and A_{pupil} is the complex amplitude of the field in the pupil plane.

If the fiber is located in the focal plane of the telescope, then the expression of the injection ratio can be calculated from Equation D.4:

$$\eta = \frac{\left\| \iint A_{\text{image}}(x, y) \times E^*_{\text{mode}}(x, y) \mathrm{d}x\mathrm{d}y \right\|^2}{\left\| \iint A_{\text{image}} \mathrm{d}x\mathrm{d}y \right\|^2}$$

But using Fourier transform calculus, this ratio can also we can also be calculated directly in the pupil plane:

$$\eta = \frac{\left\| \iint \text{TF}\left[A_{\text{image}}\right](X, Y) \times \rho\left(X, Y\right) \mathrm{d}X\mathrm{d}Y \right\|^2}{\left\| \iint \text{TF}\left[A_{\text{image}}\right](X, Y) \, \mathrm{d}X\mathrm{d}Y \right\|^2}$$

Where we have denoted ρ the inverse Fourier transform of the fiber fundamental mode field. This is called the *fiber pupil illumination*.

Since the complex amplitude in the image plane is the Fourier transform of the amplitude in the pupil plane, the above integral can actually be calculated entirely in the pupil plane:

$$\eta = \frac{\left\| \iint P(x, y) \times A_{\text{pupil}}(x, y) \times \rho(x, y) \mathrm{d}x\mathrm{d}y \right\|^2}{\left\| \iint P(x, y) \times A_{\text{pupil}}\left(x, y\right) \mathrm{d}x\mathrm{d}y \right\|^2} \tag{A.16}$$

And because the fiber mode is Gaussian, the fiber pupil illumination is also Gaussian, which makes the calculation of the overlap integral relatively easy.

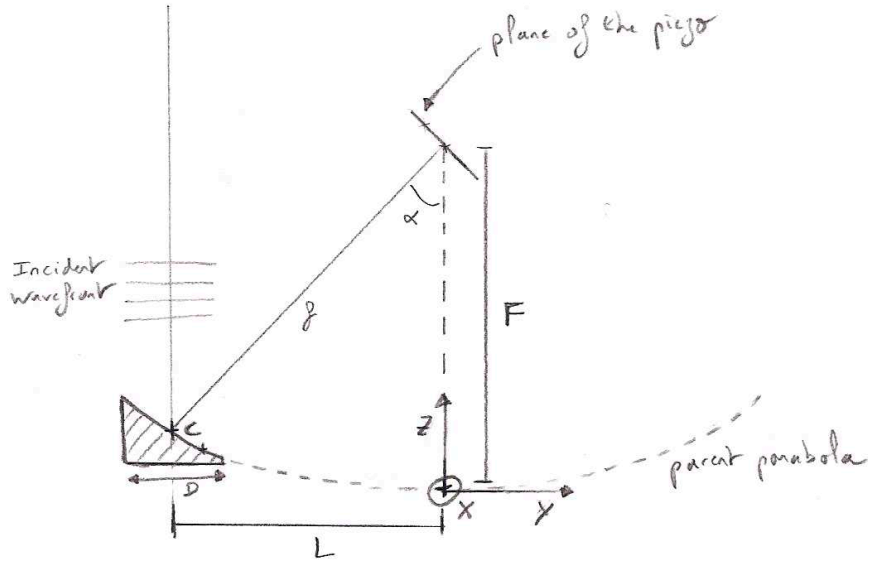


Figure A.3: Geometry of the unfolded PicSat telescope. The primary mirror is an off-axis parabola, which is part of a larger parent parabola. Observations are made in the plane of the piezo actuator, which is tilted with respect to the focal plane of the parent parabola.

A.2 Fiber injection for PicSat

The Fraunhoffer approximation does not account for the astigmatism introduced by off-axis parabolic mirrors when observing off-axis, which needs to be explicitly injected as a pupil-plane aberration. Plus, in the case of off-axis mirrors, the approximation is only valid if the diameter D of the mirror is much smaller than the focal length f, AND the distance from the center of the off-axis parabola to the center of the parent parabola is also much smaller than f. This condition is too restrictive for the PicSat instrument, and numerical calculations show a difference between the Fourier transform approach and the Kirchoff integral. Consequently, I opted for a Kirchoff-based calculation in all my simulations made on the PicSat instrument.

The telescope used in PicSat has an effective diameter of $D = 37.1$ mm, set by a diaphragm located on top of the optical tube, and a focal length of $f = 148.4$ mm. The aperture ratio is $f/D = 4.0$, and the main parabola has an off-axis angle of 30 deg. The fiber is mounted on the piezo actuator, which is inclined with respect to the focal plane of the parent parabola. The geometry is shown in Figure A.3. In this geometry, the field received in the plane of the fiber is:

$$A_{\text{fiber}}(X,Y) = \iint_{\text{mirror}} \frac{A_0 e^{i\frac{2\pi}{\lambda}\left(a(x^2+(y+y_c)^2)-\sqrt{[x-X]^2+[y-Y\cos(\alpha)]^2+[a(x^2+(y+y_c)^2)-Z\sin(\alpha)]^2}\right)}}{\sqrt{[x-X]^2+[y-Y\cos(\alpha)]^2+[a(x^2+(y+y_c)^2)-Z\sin(\alpha)]^2}}\mathrm{d}x\mathrm{d}y$$

If the source is slightly off-axis, coming at an angle $(d\alpha, d\beta)$, a phase slope needs to be added to the field A_0:

$$A_0 \leftrightarrow A_0 e^{-i\frac{2\pi}{\lambda}(xd\alpha+yd\beta)} \tag{A.17}$$

The injection for PicSat can the be calculated directly from Equation D.4.

A.3 Fiber injection in GRAVITY

The fiber coupling in GRAVITY function can be theoretically calculated. In the case of a non-aberrated wavefront, for a source located at $\Delta\text{RA}, \Delta\text{DEC}$ in the sky, the complex amplitude describing the wavefront at a given wavelength λ in the pupil plane of the telescope is:

$$A_\lambda(x,y) = P(x,y)\exp\left(i\,\frac{2\pi}{\lambda}\left(\Delta\text{RA}\times x + \Delta\text{DEC}\times y\right)\right) \tag{A.18}$$

For a fiber located at position (X,Y) in the image plane (i.e. at $(X/f, Y/f)$ on sky), the illumination in the pupil plane is:

$$\rho_\lambda(x,y) = G_\lambda(x,y)\exp\left(i\,\frac{2\pi}{\lambda}\left(\frac{xX}{f}+\frac{yY}{f}\right)\right) \tag{A.19}$$

with G_λ the normalized gaussian illumination pattern at wavelength λ. To within a normalization factor, the injection factor $\eta(\lambda)$ is then:

$$\begin{aligned}\eta(\lambda) &= \iint_{-\infty}^{+\infty} \rho_\lambda(x,y)A_\lambda(x,y)\mathrm{d}x\mathrm{d}y \\ &= \iint_{-\infty}^{+\infty} G_\lambda(x,y)P(x,y)\exp\left(i\,\frac{2\pi}{\lambda}\left[\left(\Delta\text{RA}-\frac{X}{f}\right)\times x + \left(\Delta\text{DEC}-\frac{Y}{f}\right)\times y\right]\right)\mathrm{d}x\mathrm{d}y\end{aligned}$$

Or, in terms of Fourier transform:

$$\eta(\lambda) = \text{TF}\left[G_\lambda(x,y)P(x,y)\right]\left(\frac{\Delta\text{RA}-X/f}{\lambda}, \frac{\Delta\text{DEC}-Y/f}{\lambda}\right) \tag{A.20}$$

The shape of this injection function is illustrated in Figure A.4, for different relative separations between the star image and the fiber.

To first order, over the wavelength range of GRAVITY, the fiber illumination can be assumed to be independent of λ. Neglecting the obstruction, we have:

$$P(x,y) = \begin{cases} 1 & \text{if } \sqrt{x^2+y^2} < D/2 \\ 0 & \text{otherwise}\end{cases}$$

$$G_\lambda(x,y) \propto \exp\left(-\frac{x^2+y^2}{2\left(0.3154\times D\right)^2}\right)$$

With D the aperture diameter (8.115 m for the UTs).

From this, it is possible to numerically calculate the shape of $\eta(\lambda)$ for different fiber offsets. The results are given in Figure A.4. The chromaticity of the injection becomes stronger with increasing fiber offset. For a given fiber offset, a second order polynomial is an excellent estimate of the chromaticity of the injection. Fitting a second-order polynomial for different fiber offsets, it is possible to see how the 3 coefficients vary with this offset. The result of this study is plotted in Figure A.5. It is possible to fit a 5-th order polynomial to each of these 3 curves to obtain a closed-form expression of the chromaticity of the injection for any fiber offset:

$$\eta(\lambda) = a_0(\theta) + a_1(\theta)(\lambda-\lambda_0) + a_2(\theta)(\lambda-\lambda_0)^2 \tag{A.21}$$

With:

$$a_0(\theta) = b_{0,0} + b_{0,1}\theta + b_{0,2}\theta^2 + b_{0,3}\theta^3 + b_{0,4}\theta^4 \tag{A.22}$$

$$a_1(\theta) = b_{0,1} + b_{1,1}\theta + b_{1,2}\theta^2 + b_{1,3}\theta^3 + b_{1,4}\theta^4 \tag{A.23}$$

$$a_2(\theta) = b_{0,2} + b_{2,1}\theta + b_{2,2}\theta^2 + b_{2,3}\theta^3 + b_{2,4}\theta^4 \tag{A.24}$$

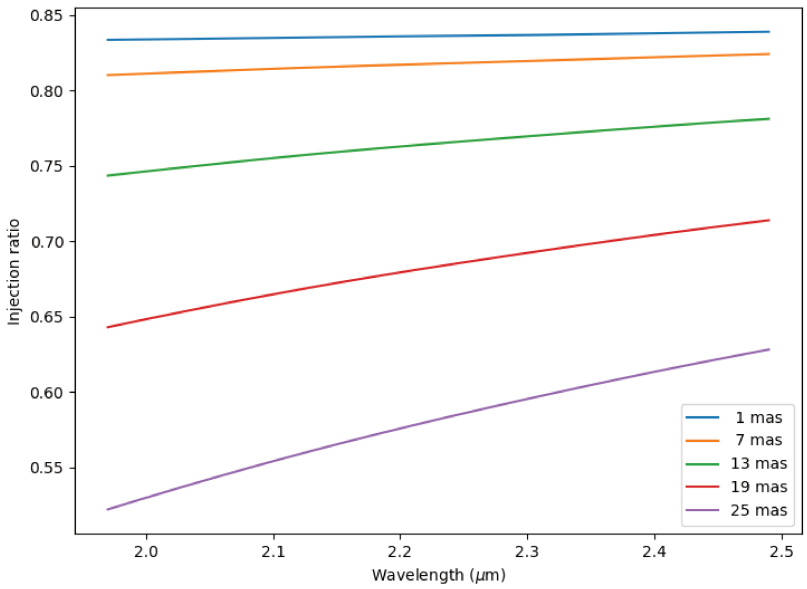


Figure A.4: Fiber injection $\eta(\lambda)$ versus wavelength for different offset between the fiber and the star image (converted to milli-arcseconds)

		j					
		0	1	2	3	4	5
i	0	8.32E-01	2.85E-03	-7.27E-04	7.75E-06	2.74E-08	-4.58E-10
	1	2.16E-02	-6.52E-03	1.14E-03	-2.83E-05	2.27E-07	-4.73E-10
	2	-1.26E-02	7.70E-03	-1.19E-03	4.03E-05	-4.85E-07	1.92E-09

Table A.1: Values of the $b_{i,j}$ of Equation A.22 to A.24 (valid in the approximate range 0 mas $< \theta <$ 80 mas)

θ being the offset between the fiber position and the star image, expressed in milli-arcsecond, and $\lambda_0 = 2.2$ μm. The values of the $b_{i,j}$ are given in Table A.1.

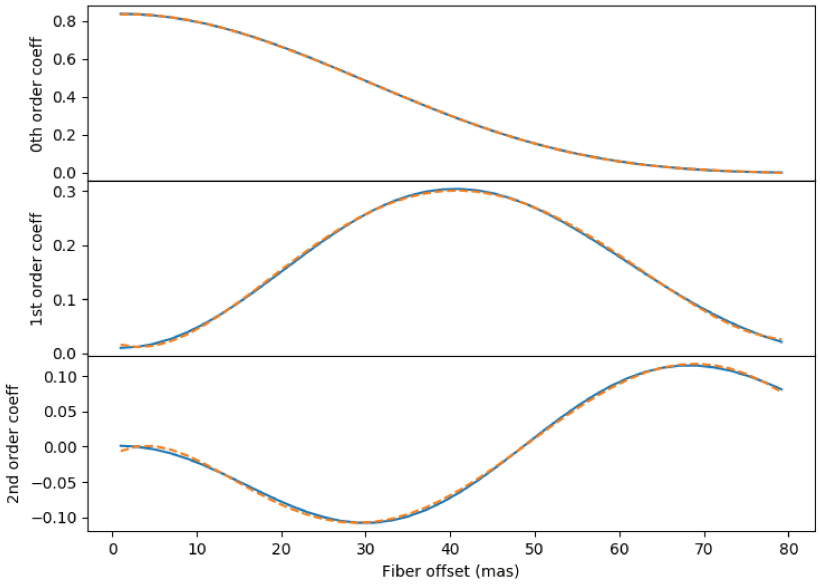


Figure A.5: Coefficients a_0 (top), a_1 (middle), and a_2 (bottom) versus fiber offset θ in milli-arcsecond. The blue line represent the numerical calculation (assuming a circular aperture), and the orange dotted line is a 5-th order polynomial fit. The coefficients of the different fits are given in Table A.1

Appendix B

An extension of maximum likelihood to complex errors

B.1 Notations

In this annex, I adopt similar notations as introduced in Section 12.2:

- Complex numbers, vectors, and matrices are underlined: $\underline{V}$.
- The real part of a complex number $\underline{V}$ is denoted $\mathrm{Re}(\underline{V})$, and its imaginary part $\mathrm{Im}(\underline{V})$.
- The complex conjugate of a quantity $\underline{C}$ is denoted $\underline{C}^*$.
- The conjugate transpose of a matrix or vector $\underline{H}$ is denoted $\underline{H}^\dagger$.
- Vectors are denoted in bold font: $\mathbf{X}$.

B.2 Complex errors and complex random variables

B.2.1 Covariance and pseudo-covariance matrices

The complex visibilities $\underline{\mathbf{V}}$ measured by GRAVITY are affected by random measurement noise. This noise affects both the real and the complex parts of the visibilities. We can write something like:

$$\underline{\mathbf{V}} = \underline{\mathbf{V}}_{\mathrm{true}} + \underline{\mathbf{E}} \tag{B.1}$$

where $\underline{\mathbf{V}}_{\mathrm{true}}$ is the true visibility, and $\underline{\mathbf{E}}$ is a complex random vector, combination of two real random vectors $\mathbf{E}_x$ and $\mathbf{E}_y$ representing respectively the measurement noise on the real and on the imaginary parts of the visibility:

$$\underline{\mathbf{E}} = \mathbf{E}_x + i\,\mathbf{E}_y \tag{B.2}$$

We can suppose both $\mathbf{E}_x$ and $\mathbf{E}_y$ normally distributed. If these two random vectors are independent, we can introduce their covariance matrices $\mathrm{cov}(\mathbf{E}_x)$ and $\mathrm{cov}(\mathbf{E}_y)$, and write:

$$\mathbf{E}_x \sim \mathcal{N}\left(0,\ \mathrm{cov}(\mathbf{E}_x)\right) \tag{B.3}$$
$$\mathbf{E}_y \sim \mathcal{N}\left(0,\ \mathrm{cov}(\mathbf{E}_y)\right) \tag{B.4}$$

Separating the real and the imaginary parts of the noise in two different vectors treated independently can be fine when there is no need to perform any complex operation on it. For

example, if we perform a real linear operation on $\underline{\mathbf{V}}$, by multiplying it by a real matrix $\mathbf{A}$, we have:

$$\mathrm{Re}(\mathbf{A}\underline{\mathbf{V}}) = \mathbf{A}\,\mathrm{Re}(\underline{\mathbf{V}}) \tag{B.5}$$
$$\mathrm{Im}(\mathbf{A}\underline{\mathbf{V}}) = \mathbf{A}\,\mathrm{Im}(\underline{\mathbf{V}}) \tag{B.6}$$

And thus the noise transforms as follows:

$$\mathbf{A}\underline{\mathbf{E}} \sim \mathcal{N}\left(0,\, \mathbf{A}\,\mathrm{cov}(\mathbf{E_x})\,\mathbf{A}^T\right) + i\,\mathcal{N}\left(0,\, \mathbf{A}\,\mathrm{cov}(\mathbf{E_y})\,\mathbf{A}^T\right) \tag{B.7}$$

However, as soon as we perform a complex transformation, for example multiplying $\underline{\mathbf{V}}$ by a complex matrix $\underline{\mathbf{A}}$, we introduce potential correlations between the real and imaginary parts of the transformed visibilities, and the above formalism becomes inefficient.

To be able to take into account any correlation between the real and the imaginary parts of the errors, we need to consider the real and the imaginary parts of the visibilities as part of a unique random vector. One way to deal with this problem is to switch to larger dimensions by defining:

$$\mathbf{V} = \begin{pmatrix}\mathrm{Re}(\underline{\mathbf{V}})\\ \mathrm{Im}(\underline{\mathbf{V}})\end{pmatrix};\quad \mathbf{V}_{\mathrm{true}} = \begin{pmatrix}\mathrm{Re}(\underline{\mathbf{V}}_{\mathrm{true}})\\ \mathrm{Im}(\underline{\mathbf{V}}_{\mathrm{true}})\end{pmatrix};\quad \mathbf{E} = \begin{pmatrix}\mathbf{E}_x\\ \mathbf{E}_y\end{pmatrix};\quad \mathbf{A} = \begin{pmatrix}\mathrm{Re}(\underline{\mathbf{A}}) & -\mathrm{Im}(\underline{\mathbf{A}})\\ \mathrm{Im}(\underline{\mathbf{A}}) & \mathrm{Re}(\underline{\mathbf{A}})\end{pmatrix} \tag{B.8}$$

The complex operation $\underline{\mathbf{A}}\,\underline{\mathbf{V}}$ is then equivalent to $\mathbf{AV}$. If the initial covariance matrix for $\mathbf{V}$ is:

$$\mathrm{cov}(\mathbf{V}) = \begin{pmatrix}\mathrm{cov}(\mathbf{E}_x) & 0\\ 0 & \mathrm{cov}(\mathbf{E}_y)\end{pmatrix} \tag{B.9}$$

Then, after the linear operation, we have:

$$\begin{aligned}\mathrm{cov}(\mathbf{AV}) &= \mathbf{A}\,\mathrm{cov}(\mathbf{V})\,\mathbf{A}^T\\ &= \begin{pmatrix}\mathrm{Re}(\underline{\mathbf{A}})\,\mathrm{cov}(\mathbf{E}_x)\,\mathrm{Re}(\underline{\mathbf{A}})^T + \mathrm{Im}(\underline{\mathbf{A}})\,\mathrm{cov}(\mathbf{E}_y)\,\mathrm{Im}(\underline{\mathbf{A}})^T & \mathrm{Re}(\underline{\mathbf{A}})\,\mathrm{cov}(\mathbf{E}_x)\,\mathrm{Im}(\underline{\mathbf{A}})^T - \mathrm{Im}(\underline{\mathbf{A}})\,\mathrm{cov}(\mathbf{E}_y)\,\mathrm{Re}(\underline{\mathbf{A}})^T\\ \mathrm{Im}(\underline{\mathbf{A}})\,\mathrm{cov}(\mathbf{E}_x)\,\mathrm{Re}(\underline{\mathbf{A}})^T - \mathrm{Re}(\underline{\mathbf{A}})\,\mathrm{cov}(\mathbf{E}_y)\,\mathrm{Im}(\underline{\mathbf{A}})^T & \mathrm{Im}(\underline{\mathbf{A}})\,\mathrm{cov}(\mathbf{E}_x)\,\mathrm{Im}(\underline{\mathbf{A}})^T + \mathrm{Re}(\underline{\mathbf{A}})\,\mathrm{cov}(\mathbf{E}_y)\,\mathrm{Re}(\underline{\mathbf{A}})^T\end{pmatrix}\end{aligned}$$

in which we clearly see the new non-diagonal correlation terms.

There is another way, though, which does not require to split the real and imaginary parts of the matrices. We can consider $\underline{\mathbf{E}}$ as a complex random vector, and introduce its covariance and pseudo-covariance matrices, defined as:

$$\mathrm{cov}(\underline{\mathbf{E}}) = \mathrm{E}\Big([\underline{\mathbf{E}} - \mathrm{mean}(\underline{\mathbf{E}})]\,[\underline{\mathbf{E}} - \mathrm{mean}(\underline{\mathbf{E}})]^{\dagger}\Big) \tag{B.10}$$
$$\mathrm{pcov}(\underline{\mathbf{E}}) = \mathrm{E}\Big([\underline{\mathbf{E}} - \mathrm{mean}(\underline{\mathbf{E}})]\,[\underline{\mathbf{E}} - \mathrm{mean}(\underline{\mathbf{E}})]^{T}\Big) \tag{B.11}$$

where $E(-)$ represents the expected value of a random vector.
These two matrices can be related to $\mathrm{cov}(\mathbf{E})$ by noting that, in the most general case:

$$\mathrm{cov}(\mathbf{E}) = \begin{pmatrix}\mathrm{cov}(\mathbf{E}_x) & \mathrm{cov}(\mathbf{E}_x, \mathbf{E}_y)\\ \mathrm{cov}(\mathbf{E}_y, \mathbf{E}_x) & \mathrm{cov}(\mathbf{E}_y)\end{pmatrix} \tag{B.12}$$

And (supposing that $\underline{\mathbf{E}}$ is centered around 0 to simplify the notations, but it is easy to show that this holds true even when $\underline{\mathbf{E}}$ is not centered):

$$\mathrm{cov}(\underline{\mathbf{E}}) = \mathrm{E}\big(\mathbf{E_x}\mathbf{E_x}^T\big) + \mathrm{E}\big(\mathbf{E_y}\mathbf{E_y}^T\big) + i\,\left[\mathrm{E}\big(\mathbf{E_y}\mathbf{E_x}^T\big) - \mathrm{E}\big(\mathbf{E_x}\mathbf{E_y}^T\big)\right] \tag{B.13}$$
$$\mathrm{pcov}(\underline{\mathbf{E}}) = \mathrm{E}\big(\mathbf{E_x}\mathbf{E_x}^T\big) - \mathrm{E}\big(\mathbf{E_y}\mathbf{E_y}^T\big) + i\,\left[\mathrm{E}\big(\mathbf{E_y}\mathbf{E_x}^T\big) + \mathrm{E}\big(\mathbf{E_x}\mathbf{E_y}^T\big)\right] \tag{B.14}$$

From which we deduce the following equalities:

$$\mathrm{cov}(\underline{\mathbf{E}}) = \mathrm{cov}(\mathbf{E_x}, \mathbf{E_x}) + \mathrm{cov}(\mathbf{E}_y, \mathbf{E}_y) + i\ [\mathrm{cov}(\mathbf{E_y}, \mathbf{E_x}) - \mathrm{cov}(\mathbf{E_x}, \mathbf{E_y})] \tag{B.15}$$

$$\mathrm{pcov}(\underline{\mathbf{E}}) = \mathrm{cov}(\mathbf{E_x}, \mathbf{E_x}) - \mathrm{cov}(\mathbf{E}_y, \mathbf{E}_y) + i\ [\mathrm{cov}(\mathbf{E_y}, \mathbf{E_x}) + \mathrm{cov}(\mathbf{E_x}, \mathbf{E_y})] \tag{B.16}$$

And:

$$\mathrm{cov}(\mathbf{E}_x, \mathbf{E}_x) = \frac{1}{2}\,\mathrm{Re}(\,\mathrm{cov}(\underline{\mathbf{E}}) + \mathrm{pcov}(\underline{\mathbf{E}})) \tag{B.17}$$

$$\mathrm{cov}(\mathbf{E}_y, \mathbf{E}_y) = \frac{1}{2}\,\mathrm{Re}(\,\mathrm{cov}(\underline{\mathbf{E}}) - \mathrm{pcov}(\underline{\mathbf{E}})) \tag{B.18}$$

$$\mathrm{cov}(\mathbf{E}_x, \mathbf{E}_y) = \frac{1}{2}\,\mathrm{Im}(-\mathrm{cov}(\underline{\mathbf{E}}) + \mathrm{pcov}(\underline{\mathbf{E}})) \tag{B.19}$$

$$\mathrm{cov}(\mathbf{E}_y, \mathbf{E}_x) = \frac{1}{2}\,\mathrm{Im}(\,\mathrm{cov}(\underline{\mathbf{E}}) + \mathrm{pcov}(\underline{\mathbf{E}})) \tag{B.20}$$

Also, from Equations B.10 and B.11, it is easy to show that for any complex matrix $\underline{\mathbf{A}}$:

$$\mathrm{cov}(\underline{\mathbf{A}}\underline{\mathbf{E}}) = \underline{\mathbf{A}}\,\mathrm{cov}(\underline{\mathbf{E}})\,\underline{\mathbf{A}}^{\dagger} \tag{B.21}$$

$$\mathrm{pcov}(\underline{\mathbf{A}}\underline{\mathbf{E}}) = \underline{\mathbf{A}}\,\mathrm{pcov}(\underline{\mathbf{E}})\,\mathbf{A}^{T} \tag{B.22}$$

The interest of using the covariance and pseudo-covariance formalism resides in these two last equalities: propagating the errors when performing a complex linear operation with $\underline{\mathbf{A}}$ is a matter of multiplying the covariance and pseudo-covariance matrix by $\underline{\mathbf{A}}$ and its transpose (or complex transpose), much like when dealing with classical real random vectors. At any point in time, these covariance and pseudo-covariance matrices can be converted to the actual variance-covariance matrix on the real and imaginary parts of the complex random vector, using Equation B.20.

B.2.2 The complex normal law

Let $\mathbf{E}$ be a real random vector following a centered normal law of covariance $\mathbf{W}$, which we note $\mathbf{E} \sim \mathcal{N}(0, \mathbf{W})$. Its Probability Density Function (PDF) can be expressed as:

$$\mathrm{PDF}[\mathbf{E}](\mathbf{X}) \propto \exp\left(-\frac{1}{2}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}\right) \tag{B.23}$$

Similarly, a random complex vector $\underline{\mathbf{E}}$ is said to follow a centered complex-normal distribution of covariance $\underline{\mathbf{W}}$ and pseudo-covariance $\underline{\mathbf{Z}}$ if the real random vector $(\mathbf{E}_x^T\ \mathbf{E}_y^T)^T$ is itself normally distributed, with mean 0 and covariance $\mathbf{W}$ defined by:

$$\mathbf{W} = \frac{1}{2}\begin{pmatrix} \mathrm{Re}(\underline{\mathbf{W}} + \underline{\mathbf{Z}}) & \mathrm{Im}(-\underline{\mathbf{W}} + \underline{\mathbf{Z}}) \\ \mathrm{Im}(\underline{\mathbf{W}} + \underline{\mathbf{Z}}) & \mathrm{Re}(\underline{\mathbf{W}} - \underline{\mathbf{Z}}) \end{pmatrix} \tag{B.24}$$

We note $\underline{\mathbf{E}} \sim \mathcal{CN}(0, \underline{\mathbf{W}}, \underline{\mathbf{Z}})$. The probability density function can be derived from the real normal law:

$$\mathrm{PDF}[\underline{\mathbf{E}}](\underline{\mathbf{X}}) \propto \exp\left(-\frac{1}{2}\begin{bmatrix}\mathrm{Re}(\underline{\mathbf{X}})^T & \mathrm{Im}(\underline{\mathbf{X}})^T\end{bmatrix}\mathbf{W}^{-1}\begin{bmatrix}\mathrm{Re}(\underline{\mathbf{X}}) \\ \mathrm{Im}(\underline{\mathbf{X}})\end{bmatrix}\right) \tag{B.25}$$

This PDF can also be expressed using the covariance and pseudo-covariance matrices. To do so, we first need to introduce the matrix L defined ad:

$$\underline{\mathbf{L}} = \frac{1}{2}\begin{pmatrix} \mathbf{I} & \mathbf{I} \\ -i\,\mathbf{I} & i\,\mathbf{I} \end{pmatrix} \tag{B.26}$$

We can note that:

$$\underline{\underline{\mathbf{L}}}^{-1} = \begin{pmatrix} \mathbf{I} & i\,\mathbf{I} \\ \mathbf{I} & -i\,\mathbf{I} \end{pmatrix} = \underline{\underline{\mathbf{L}}}^{\dagger} \tag{B.27}$$

Introducing the vector $\left(\underline{\mathbf{X}}^T\ \underline{\mathbf{X}}^{*T}\right)^T$, we have the following equalities:

$$\begin{pmatrix} \mathrm{Re}(\underline{\mathbf{X}}) \\ \mathrm{Im}(\underline{\mathbf{X}}) \end{pmatrix} = \underline{\underline{\mathbf{L}}} \begin{pmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{pmatrix} \tag{B.28}$$

$$\mathbf{W} = \underline{\underline{\mathbf{L}}} \begin{pmatrix} \underline{\underline{\mathbf{W}}} & \underline{\underline{\mathbf{Z}}} \\ \underline{\underline{\mathbf{Z}}}^* & \underline{\underline{\mathbf{W}}}^* \end{pmatrix} \underline{\underline{\mathbf{L}}}^{\dagger} \tag{B.29}$$

We also have (since the vector is real):

$$\left(\mathrm{Re}(\underline{\mathbf{X}})^T\ \mathrm{Im}(\underline{\mathbf{X}})^T\right)^T = \left(\mathrm{Re}(\underline{\mathbf{X}})^T\ \mathrm{Im}(\underline{\mathbf{X}})^T\right)^{\dagger} \tag{B.30}$$

Thus we can write:

$$\begin{aligned} \begin{bmatrix} \mathrm{Re}(\underline{\mathbf{X}})^T & \mathrm{Im}(\underline{\mathbf{X}})^T \end{bmatrix} \mathbf{W}^{-1} \begin{bmatrix} \mathrm{Re}(\underline{\mathbf{X}}) \\ \mathrm{Im}(\underline{\mathbf{X}}) \end{bmatrix} &= \left(\underline{\underline{\mathbf{L}}} \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix}\right)^{\dagger} \mathbf{W}^{-1} \left(\underline{\underline{\mathbf{L}}} \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix}\right) \\ &= \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix}^{\dagger} \underline{\underline{\mathbf{L}}}^{\dagger} \mathbf{W}^{-1} \underline{\underline{\mathbf{L}}} \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix} \\ &= \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix}^{\dagger} \left(\underline{\underline{\mathbf{L}}} \mathbf{W} \underline{\underline{\mathbf{L}}}^{\dagger}\right)^{-1} \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix} \\ &= \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix}^{\dagger} \begin{pmatrix} \underline{\underline{\mathbf{W}}} & \underline{\underline{\mathbf{Z}}} \\ \underline{\underline{\mathbf{Z}}}^* & \underline{\underline{\mathbf{W}}}^* \end{pmatrix}^{-1} \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix} \end{aligned}$$

From the definition of the pseudo-covariance matrix as given in Equation B.11, it is clear that $\underline{\underline{\mathbf{Z}}}$ is symmetric, and thus $\underline{\underline{\mathbf{Z}}}^* = \underline{\underline{\mathbf{Z}}}^{\dagger}$. This leads to the final expression for the PDF of a centered complex-normal random vector:

$$\mathrm{PDF}[\underline{\mathbf{E}}](\underline{\mathbf{X}}) \propto \exp\left(-\frac{1}{2} \begin{bmatrix} \underline{\mathbf{X}}^{\dagger} & \underline{\mathbf{X}}^{*\dagger} \end{bmatrix} \begin{pmatrix} \underline{\underline{\mathbf{W}}} & \underline{\underline{\mathbf{Z}}} \\ \underline{\underline{\mathbf{Z}}}^{\dagger} & \underline{\underline{\mathbf{W}}}^* \end{pmatrix}^{-1} \begin{bmatrix} \underline{\mathbf{X}} \\ \underline{\mathbf{X}}^* \end{bmatrix}\right) \tag{B.31}$$

This is easily extended to the non-centered case of mean $\underline{\mathbf{M}}$. If $\underline{\mathbf{E}} \sim \mathcal{CN}(\underline{\mathbf{M}}, \underline{\underline{\mathbf{W}}}, \underline{\underline{\mathbf{Z}}})$, then we have:

$$\mathrm{PDF}[\underline{\mathbf{E}}](\underline{\mathbf{X}}) \propto \exp\left(-\frac{1}{2} \begin{bmatrix} (\underline{\mathbf{X}} - \underline{\mathbf{M}})^{\dagger} & (\underline{\mathbf{X}}^* - \underline{\mathbf{M}}^*)^{\dagger} \end{bmatrix} \begin{pmatrix} \underline{\underline{\mathbf{W}}} & \underline{\underline{\mathbf{Z}}} \\ \underline{\underline{\mathbf{Z}}}^{\dagger} & \underline{\underline{\mathbf{W}}}^* \end{pmatrix}^{-1} \begin{bmatrix} \underline{\mathbf{X}} - \underline{\mathbf{M}} \\ \underline{\mathbf{X}}^* - \underline{\mathbf{M}}^* \end{bmatrix}\right) \tag{B.32}$$

Finally, using results known on real normal vectors, it is easy to demonstrate that complex linear operations on a random variable following a complex-normal law always yields a new complex-normally distributed normal variable, with a covariance and pseudo-covariance matrices propagated like in Equations B.21 and B.22:

$$\underline{\mathbf{E}} \sim \mathcal{CN}(\underline{\mathbf{M}}, \underline{\underline{\mathbf{W}}}, \underline{\underline{\mathbf{Z}}}) \quad \Rightarrow \quad \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{E}} \sim \mathcal{CN}\left(\underline{\underline{\mathbf{A}}}\,\underline{\mathbf{M}}, \underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{W}}}\,\underline{\underline{\mathbf{A}}}^{\dagger}, \underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{Z}}}\,\underline{\underline{\mathbf{A}}}^T\right) \tag{B.33}$$

B.2.3 Extending the maximum likelihood estimator to complex errors

In the previous section, I introduced the covariance and pseudo-covariance matrices of a complex random vector, useful to represent and propagate error measurements. In this section, I now consider the following problem: given a set of complex measurements, represented by a vector $\underline{\mathbf{Y}}$, and a linear model $\underline{\mathbf{Y}} = \underline{\mathbf{A}}\mathbf{X}$, what is the best estimate of the real parameter vector $\mathbf{X}$, in the presence of measurement noise on $\underline{\mathbf{Y}}$?

Since the measurement is affected by random noise, it is better to explicitly introduce this noise as a random variable:

$$\underline{\mathbf{Y}} = \underline{\mathbf{A}}\mathbf{X} + \underline{\mathbf{E}} \tag{B.34}$$

In the following, I assume that $\underline{\mathbf{E}}$ follows a complex-normal distribution, which is equivalent to assuming that the errors on the real and imaginary parts of the visibilities are gaussian, without any other assumption. In particular, I do not assume that the real and imaginary parts are independent.

$$\underline{\mathbf{E}} \sim \mathcal{CN}\left(0,\, \underline{\mathbf{W}},\, \underline{\mathbf{Z}}\right) \tag{B.35}$$

We can introduce the likelihood of the parameters values $\mathbf{X}$, defined as the probability of $\mathbf{X}$ given that we measured the data $\underline{\mathbf{Y}}$:

$$\mathcal{L}\left(\mathbf{X}\right) = \Pr(\mathbf{X}|\underline{\mathbf{Y}}) \tag{B.36}$$

A Bayesian reasoning yields:

$$\mathcal{L}\left(\mathbf{X}\right) = \frac{\Pr(\underline{\mathbf{Y}}|\mathbf{X})\,\Pr(\mathbf{X})}{\Pr(\underline{\mathbf{Y}})} \tag{B.37}$$

The probability of measuring data $\underline{\mathbf{Y}}$ given that the parameters value are $\mathbf{X}$ is exactly equals to $\Pr(\underline{\mathbf{E}} = \underline{\mathbf{Y}} - \underline{\mathbf{A}}\mathbf{X})$. Any *a priori* knownledge about $\mathbf{X}$ can be introduced at this point. Without any prior knowledge of $\mathbf{X}$, we can always take a uniform probability for $\mathbf{X}$ (which corresponds to the least favorable case), and write:

$$\mathcal{L}\left(\mathbf{X}\right) \propto \mathrm{PDF}[\underline{\mathbf{E}}]\left(\underline{\mathbf{Y}} - \underline{\mathbf{A}}\mathbf{X}\right) \tag{B.38}$$

The maximum likelihood estimate $\hat{\mathbf{X}}$ of $\mathbf{X}$ is the value that maximizes the likelihood $\mathcal{L}$, or, equivalently, that minimizes $-\log\mathcal{L}$:

$$\hat{\mathbf{X}} = \mathrm{argmin}\left(-\log\mathcal{L}\right) \tag{B.39}$$

Combining Equation B.38 with the fact that $\underline{\mathbf{E}}$ follows a complex-normal distribution, and with Equation B.32, we get:

$$\hat{\mathbf{X}} = \mathrm{argmin}\left\{\begin{bmatrix}(\underline{\mathbf{Y}} - \underline{\mathbf{A}}\mathbf{X})^\dagger & (\underline{\mathbf{Y}}^* - \underline{\mathbf{A}}^*\mathbf{X})^\dagger\end{bmatrix}\begin{pmatrix}\underline{\mathbf{W}} & \underline{\mathbf{Z}} \\ \underline{\mathbf{Z}}^\dagger & \underline{\mathbf{W}}^*\end{pmatrix}^{-1}\begin{bmatrix}\underline{\mathbf{Y}} - \underline{\mathbf{A}}\mathbf{X} \\ \underline{\mathbf{Y}}^* - \underline{\mathbf{A}}^*\mathbf{X}\end{bmatrix}\right\} \tag{B.40}$$

If we simplify the notations a little, and write:

$$\underline{\mathbf{Y}}_2 = \begin{pmatrix}\underline{\mathbf{Y}} \\ \underline{\mathbf{Y}}^*\end{pmatrix} \tag{B.41}$$

$$\underline{\mathbf{A}}_2 = \begin{pmatrix}\underline{\mathbf{A}} \\ \underline{\mathbf{A}}^*\end{pmatrix} \tag{B.42}$$

$$\underline{\mathbf{W}}_2 = \begin{pmatrix}\underline{\mathbf{W}} & \underline{\mathbf{Z}} \\ \underline{\mathbf{Z}}^\dagger & \underline{\mathbf{W}}^*\end{pmatrix} \tag{B.43}$$

Then the expression of $\hat{\mathbf{X}}$ becomes:

$$\hat{\mathbf{X}} = \operatorname{argmin}\left\{(\underline{\mathbf{U}}_2 - \underline{\mathbf{A}}_2\mathbf{X})^\dagger \underline{\mathbf{W}}_2{}^{-1} (\underline{\mathbf{Y}}_2 - \underline{\mathbf{A}}_2\mathbf{X})\right\} \tag{B.44}$$

And we recognize something very similar to the weighted χ^2 estimator, commonly used to deal with independent errors. The only difference here is that the matrix $\underline{\mathbf{W}}_2$ is complex, and not necessarily diagonal. But the problem can still be solved using similar calculations. We have:

$$\begin{aligned} -\log \mathcal{L}(\mathbf{X}) &= (\underline{\mathbf{Y}}_2 - \underline{\mathbf{A}}_2\mathbf{X})^\dagger \underline{\mathbf{W}}_2{}^{-1} (\underline{\mathbf{Y}}_2 - \underline{\mathbf{A}}_2\mathbf{X}) && \text{(B.45)} \\ &= \underline{\mathbf{Y}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{Y}}_2 - \mathbf{X}^\dagger \underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{Y}}_2 - \underline{\mathbf{Y}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2 \mathbf{X} + \mathbf{X}^\dagger \underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2 \mathbf{X} && \text{(B.46)} \end{aligned}$$

Which we can differentiate with respect to $\mathbf{X}$ to get:

$$\begin{aligned} \frac{\partial\left[-\log \mathcal{L}(X)\right]}{\partial \mathbf{X}} &= -\left(\underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{Y}}_2\right)^T - \left(\underline{\mathbf{Y}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right) + \mathbf{X}^T \left(\underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right)^T + \mathbf{X}^T \left(\underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right) \\ &= -\left(\underline{\mathbf{Y}}_2^\dagger \underline{\mathbf{W}}_2^{-1^\dagger} \underline{\mathbf{A}}_2\right)^* - \left(\underline{\mathbf{Y}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right) + \mathbf{X}^T \left(\underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1^\dagger} \underline{\mathbf{A}}_2\right)^* + \mathbf{X}^T \left(\underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right) \end{aligned}$$

And because $\underline{\mathbf{W}}_2$ is Hermitian (and thus its invert also is), this reduces to:

$$\frac{\partial\left[-\log \mathcal{L}(X)\right]}{\partial \mathbf{X}} = -2\,\mathrm{Re}\left(\underline{\mathbf{Y}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right) + 2\,\mathbf{X}^T\,\mathrm{Re}\left(\underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right) \tag{B.47}$$

Hence the expression for the maximum likelihood estimate of $\mathbf{X}$, corresponding to a zero of the derivative of $-\log \mathcal{L}$:

$$\hat{\mathbf{X}} = \left(\mathrm{Re}\left(\underline{\mathbf{Y}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right)\left[\mathrm{Re}\left(\underline{\mathbf{A}}_2^\dagger \underline{\mathbf{W}}_2^{-1} \underline{\mathbf{A}}_2\right)\right]^{-1}\right)^T \tag{B.48}$$

B.3 Error on the estimated parameters

Given a complex linear model $\underline{\mathbf{Y}} = \underline{\mathbf{A}}\mathbf{X}$, and a measurement $\underline{\mathbf{Y}}$ affected by random noise, Equation B.48 gives a way to retrieve an estimate of the parameters $\mathbf{X}$. But what about the error bars?

B.3.1 Unrestricted case

Let us first consider the unrestricted parameter estimation problem, in which all parameters $\mathbf{X}$ must be estimated (and not just some of them, as it is often the case). We know already that the best estimate $\hat{\mathbf{X}}$ for all these parameters can be obtained by maximizing the likelihood, or minimizing $-\ln \mathcal{L}(\mathbf{X})$.

The idea of defining error bars from $-\ln \mathcal{L}$ level contours finds its mathematical basis in Wilk's theorem, which states that if there is a "true value" $\mathbf{X}_0$ for the parameters (i.e. if the model correctly represents the data), then, in the limit of an infinite number of data points, the vector $(\mathbf{X}_0 - \hat{\mathbf{X}})$ behaves like a multi-variate random vector, with a covariance matrix given by:

$$\mathrm{var}\left(\mathbf{X}_0 - \hat{\mathbf{X}}\right) = -\left[\frac{\partial^2 \ln(\mathcal{L})}{\partial \mathbf{X}^2}(\hat{\mathbf{X}})\right]^{-1} \tag{B.49}$$

Since we know that $\hat{\mathbf{X}}$ is the location of the minimum of $-\ln\mathcal{L}$, if we suppose $\mathbf{X}_0$ close to $\hat{\mathbf{X}}$, we can write the second-order development:

$$\ln\mathcal{L}\left(\mathbf{X}_0\right) - \ln\mathcal{L}\left(\hat{\mathbf{X}}\right) = \left[\mathbf{X}_0 - \hat{\mathbf{X}}\right]^T \frac{1}{2}\frac{\partial^2 \ln(\mathcal{L})}{\partial \mathbf{X}^2}(\hat{\mathbf{X}}) \left[\mathbf{X}_0 - \hat{\mathbf{X}}\right] \tag{B.50}$$

Then we have:

$$-2\ln\left(\frac{\mathcal{L}\left(\mathbf{X}_0\right)}{\mathcal{L}\left(\hat{\mathbf{X}}\right)}\right) = \left[\mathbf{X}_0 - \hat{\mathbf{X}}\right]^T \mathrm{var}\left(\mathbf{X}_0 - \hat{\mathbf{X}}\right)^{-1} \left[\mathbf{X}_0 - \hat{\mathbf{X}}\right] \tag{B.51}$$

Which proves that the quantity:

$$T(\mathbf{X}_0) = -2\ln\left(\frac{\mathcal{L}\left(\mathbf{X}_0\right)}{\mathcal{L}\left(\hat{\mathbf{X}}\right)}\right) \tag{B.52}$$

follows a χ^2-distribution with p degrees of freedoms, where p is the number of independent parameters in $\mathbf{X}$. In other words, if one was to repeat the experiment a large number of times, each time with a different realization of the random errors affecting the measurement, and if one could also compute the value of $T(\mathbf{X}_0)$ (which would suppose that the true values of the parameters are known in advance), then one would obtain a distribution following a χ^2_p law.

From this, we know that the probability of having $T(\mathbf{X}_0)$ smaller than a given value β is given by the χ^2_p cumulative density function at p degrees of freedom:

$$\mathrm{Prob}\left(T(\mathbf{X}_0) < \beta\right) = \mathrm{CDF}\left[\chi^2_p\right](\beta) \tag{B.53}$$

But this probability is the same as the probability for $\mathbf{X}_0$ to be within the $\beta/2$ level contour of $-\ln\mathcal{L}$. Indeed, saying that $\mathbf{X}_0$ is such that:

$$-\ln\mathcal{L}\left(\mathbf{X}_0\right) \leq \min\left\{-\ln\mathcal{L}\right\} + \frac{\beta}{2} \tag{B.54}$$

is equivalent to saying that $\mathbf{X}_0$ is such that:

$$-2\ln\mathcal{L}\left(\mathbf{X}_0\right) \leq -2\ln\mathcal{L}\left(\hat{\mathbf{X}}\right) + \beta \tag{B.55}$$

And so the probability for $\mathbf{X}_0$ to be in the $\beta/2$ contour of the $-\ln\mathcal{L}$ map is the same as the one given in Equation B.53.

Finding the 95% confidence interval for the parameter $\mathbf{X}$ is now just a matter of finding the correct value of β. If we take $\beta = \mathrm{CDF}[\chi^2_p](0.95)$, then, by definition, the probability of having $\mathbf{X}_0$ within the contour region (i.e. of having the real value of the parameter in the confidence interval) is 0.95. This an acceptable definition for the 95% confidence interval[1].

[1] There are other possible definitions for the 95% confidence interval. One of which is based on the statistical result that, under some conditions of "not too much non-linearity of the model", the log-likelihood itself follows a χ^2_n distribution, with a dimensionnality n given by the number of datapoints. This can be used to define another confidence interval based on a pure log-likelihood threshold (i.e. taking the set of values giving $-2\ln\mathcal{L} < \beta$, with $\beta = \mathrm{CDF}[\chi^2_n][0.95]$), and not a contour around the minimum. This is a mathematically perfectly valid, but with a major disadvantage for astronomers: for a very unlucky realization of the noise, this confidence interval can possibly be... empty. I encourage the reader to medidate on this.

B.3.2 Restricted confidence interval

It is often the case in astronomy that some of the parameters contained in the model are of little scientific interest. This is typically the case when some parameters are introduced to take into accounts instrumental effects. When this is the case, $\mathbf{X}$ can be separated in two parts: $\mathbf{X} = (\mathbf{X}_{\text{useless}}, \mathbf{X}_{\text{useful}})$. In such a case, the best estimate of the useful parameters can of course still be estimated using Equation B.48, which gives a value for $\hat{\mathbf{X}} = \left(\hat{\mathbf{X}}_{\text{useless}}, \hat{\mathbf{X}}_{\text{useful}}\right)$. Throw away the useless part, and you have the best estimate of the useful parameters. But what about the error bars?

The 95% percent confidence interval defined above is valid for the entire set of parameters $\mathbf{X}$. One possibility is to project this interval onto the subspace defined by $\mathbf{X}_{\text{useful}}$, i.e. to take the extremum values of the different useful parameters over this confidence interval of $\hat{\mathbf{X}}$. But this projection grossly overestimates the confidence-interval.

There is a better way to proceed, based on the fact that the value of the restricted $T(\mathbf{X}_{0,\text{usefull}})$ also follows a χ^q_2 distribution (q degrees of freedom, with q the number of useful parameters). This restricted T is simply defined as:

$$T(\mathbf{X}_{\text{useful}}) = T_{\text{minimized over } \mathbf{X}_{\text{useless}}} \tag{B.56}$$

We already have:

$$T(\mathbf{X}_0) = \left[\mathbf{X}_0 - \hat{\mathbf{X}}\right]^T \text{var}\left(\mathbf{X}_0 - \hat{\mathbf{X}}\right)^{-1} \left[\mathbf{X}_0 - \hat{\mathbf{X}}\right] \tag{B.57}$$

So:

$$T(\mathbf{X}_{\text{useful}}) = \left\{ \left[\mathbf{X}_0 - \hat{\mathbf{X}}\right]^T \text{var}\left(\mathbf{X}_0 - \hat{\mathbf{X}}\right)^{-1} \left[\mathbf{X}_0 - \hat{\mathbf{X}}\right] \right\}_{\text{minimized over } \mathbf{X}_{\text{useless}}} \tag{B.58}$$

The proof relies on showing that:

$$\left[\mathbf{X} - \hat{\mathbf{X}}\right]^T \text{var}\left(\mathbf{X} - \hat{\mathbf{X}}\right)^{-1} \left[\mathbf{X} - \hat{\mathbf{X}}\right]_{\text{min over } \mathbf{X}_{\text{useless}}} = \left[\mathbf{X}_{\text{useful}} - \hat{\mathbf{X}}_{\text{useful}}\right]^T V^{-1} \left[\mathbf{X}_{\text{useful}} - \hat{\mathbf{X}}_{\text{useful}}\right] \tag{B.59}$$

with:

$$V = \text{var}\left(\mathbf{X}_{\text{useful},0} - \hat{\mathbf{X}}_{\text{useful}}\right) \tag{B.60}$$

To do so, we can write:

$$\text{var}\left(\mathbf{X}_0 - \hat{\mathbf{X}}\right)^{-1} = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \quad \text{and} \quad \text{var}\left(\mathbf{X}_0 - \hat{\mathbf{X}}\right) = \begin{pmatrix} F & G \\ G^T & H \end{pmatrix}$$

We have all the following relationships (matrix and its inverse):

$$AF + BG^T = I \tag{B.61}$$
$$AG + BH = 0 \tag{B.62}$$
$$B^T F + CG^T = 0 \tag{B.63}$$
$$B^T G + CH = I \tag{B.64}$$

If we expand the expression of $T(\mathbf{X})$ and derive it with respect to $\mathbf{X}_k$, we have:

$$\begin{aligned} \frac{\partial T}{\partial \mathbf{X}_k}(\mathbf{X}_{\text{useless}}, \mathbf{X}_{\text{useful}}) = &\left(\mathbf{X}^T_{\text{useless}} - \hat{\mathbf{X}}^T_{\text{useless}}\right) B \frac{\partial \mathbf{X}}{\partial \mathbf{X}_k} + \frac{\partial \mathbf{X}^T}{\partial \mathbf{X}_k} B^T \left(\mathbf{X}_{\text{useless}} - \hat{\mathbf{X}}_{\text{useless}}\right) \\ &+ \frac{\partial \mathbf{X}}{\partial \mathbf{X}_k} D \left(\mathbf{X}_{\text{useful}} - \hat{\mathbf{X}}_{\text{useful}}\right) + \left(\mathbf{X}^T_{\text{useful}} - \hat{\mathbf{X}}^T_{\text{useful}}\right) D \frac{\partial \mathbf{X}}{\partial \mathbf{X}_k} \end{aligned}$$

The minimum of $T(\mathbf{X})$ over all parameters in $\mathbf{X}_{\text{useless}}$ is obtained when all the derivatives are zero. This gives a set of $p-q$ equations. To shorten the notation, we can write $\Delta\mathbf{X}_{\text{useless}} = \mathbf{X}_{\text{useless}} - \hat{\mathbf{X}}_{\text{useless}}$, and $\Delta\mathbf{X}_{\text{useful}} = \mathbf{X}_{\text{useful}} - \hat{\mathbf{X}}_{\text{useful}}$. Systematically multiplying the equation for the $k-th$ derivative by $(C\Delta\mathbf{X}_{\text{useless}} - B\Delta\mathbf{X}_{\text{useless}})_k$ and adding everything gives:

$$\begin{aligned}0 =&\Delta\mathbf{X}_{\text{useless}}^T B\left(C\Delta\mathbf{X}_{\text{useful}} + B^T\mathbf{X}_{\text{useful}}\right) + \left(\Delta\mathbf{X}_{\text{useless}}^T + \Delta\mathbf{X}_{\text{useful}}^T C^T\right) B^T\Delta\mathbf{X}_{\text{useless}}\\ &+ \Delta\mathbf{X}_{\text{useful}}^T C^T\left(C\Delta\mathbf{X}_{\text{useful}} + B^T\Delta\mathbf{X}_{\text{useless}}\right) + \left(\Delta\mathbf{X}_{\text{useful}}^T C^T + \Delta\mathbf{X}_{\text{useless}}B^T\right) C\Delta\mathbf{X}_{\text{useful}}\end{aligned}$$

Covariance matrices are symmetric ($C^T = C$), so expanding the products and dividing by two gives:

$$\begin{aligned}0 =& \Delta\mathbf{X}_{\text{useless}}^T BC\Delta\mathbf{X}_{\text{useful}} + \Delta\mathbf{X}_{\text{useless}}^T BB^T\Delta\mathbf{X}_{\text{useless}}\\ &+ \Delta\mathbf{X}_{\text{useful}}^T CB^T\Delta\mathbf{X}_{\text{useless}} + \Delta\mathbf{X}_{\text{useful}}^T C^T C\Delta\mathbf{X}_{\text{useful}}\end{aligned}$$

Factorizing:

$$0 = \left(C\Delta\mathbf{X}_{\text{useful}} + B^T\Delta\mathbf{X}_{\text{useless}}\right)^T\left(C\Delta\mathbf{X}_{\text{useful}} + B^T\Delta\mathbf{X}_{\text{useless}}\right) \tag{B.65}$$

Thus:

$$0 = \left(C\Delta\mathbf{X}_{\text{useful}} + B^T\Delta\mathbf{X}_{\text{useless}}\right)$$

Solving it for $\Delta\mathbf{X}_{\text{useless}}$:

$$\Delta\mathbf{X}_{\text{useless}} = -C^{-1}B^T\Delta\mathbf{X}_{\text{useful}}$$

Which can be injected into the definition of T to get:

$$T(\mathbf{X}_{\text{useless}}, \mathbf{X}_{\text{useful}})_{\text{minim. over }\mathbf{X}_{\text{useless}}} = \Delta\mathbf{X}_{\text{useful}}^T\left(A - BC^{-1}B^T\right)\Delta\mathbf{X}_{\text{useful}} \tag{B.66}$$

From equations B.61 to B.64, we have:

$$\begin{aligned}&A = \left(I - BG^T\right)F^{-1}\\ &G^T = -C^{-1}B^T F \quad\Rightarrow\quad BG^T = -BC^{-1}B^T F\end{aligned}$$

And so:

$$A - BC^{-1}B^T = F^{-1}$$

Since with our previous definitions, we actually have $V = G$, this proves Equation B.59, which in turn demonstrates that:

$$T(\mathbf{X}_{0,\text{useful}}) = \left(\mathbf{X}_{0,\text{useful}}^T - \hat{\mathbf{X}}_{\text{useful}}^T\right)\text{var}\left(\mathbf{X}_{0,\text{useful}} - \hat{\mathbf{X}}_{\text{useful}}\right)^{-1}\left(\mathbf{X}_{0,\text{useful}} - \hat{\mathbf{X}}_{\text{useful}}\right)$$

Combining this to Wilk's theorem proves that $T(\mathbf{X}_{0,\text{useful}})$ follows a χ_q^2-distribution.

At this point, we can define a 95% confidence interval by using the region delimited by the contour-level of value CDF $\left[\chi_q^2\right](0.95)$ around $\hat{\mathbf{X}}_{\text{useful}}$, or a 68% interval by using CDF $\left[\chi_q^2\right](0.68)$ instead.

Appendix C

A simple ADCS model

C.1 Motivation

To develop the tracking algorithm for the PicSat instrument, I needed a way to model the pointing errors downstream of the attitude control system. Precision pointing on a CubeSat platform is something relatively new, and I could not find any real data in the literature. I had to resort to making my own ADCS model to generate pointing errors. The model was developed in MATLAB/SIMULINK, and is shown in Figure C.1. This model is intended to simulate the target pointing mode of the ADCS, in which the two main sensors used are the star tracker and the gyroscopes, and in which the satellite attitude is controlled with the reaction wheels.

C.2 Model

A torque τ is applied on the satellite, and is converted into a rotational acceleration $\dot{\boldsymbol{\omega}}$ using the matrix of inertia J_{PicSat}. This rotational acceleration is integrated two times to give the satellite rotational speed $\boldsymbol{\omega}$ and the angular position $\boldsymbol{\theta}$:

$$\dot{\boldsymbol{\omega}} = J_{\text{PicSat}}\,\boldsymbol{\tau} \tag{C.1}$$

$$\boldsymbol{\omega} = \dot{\boldsymbol{\theta}} \tag{C.2}$$

A random noise is added to the rotational speed and to the angular position to simulate measurements from a gyroscope and a star tracker:

$$\boldsymbol{\theta}_{\text{ST}} = \boldsymbol{\theta} + \mathcal{N}(0, \sigma_{\text{ST}}) \tag{C.3}$$

$$\boldsymbol{\omega}_{\text{gyro}} = \boldsymbol{\omega} + \mathcal{N}(0, \sigma_{\text{gyro}}) \tag{C.4}$$

with $\mathcal{N}(0, \sigma)$ representing a gaussian white noise of mean value 0 and standard deviation σ.

The measurement from the gyroscopes and from the star tracker are combined in the simplest possible way, to simulate a pseudo-IMU: the high-frequency gyroscope measurements are integrated to give an angular position estimate, which is reset to the low frequency star tracker measurement each time a new measurement is available. This gives an estimate of the satellite angular position $\hat{\boldsymbol{\theta}}_{\text{IMU}}$:

$$\hat{\boldsymbol{\theta}}_{\text{IMU}}(t) = \boldsymbol{\theta}_{\text{ST}}(t_{\text{ref}}) + \int_{t_{\text{ref}}}^{t} \boldsymbol{\omega}_{\text{gyro}}(t^*)\mathrm{d}t^* \tag{C.5}$$

where t_{ref} is the time at which the last measurement has been obtained with the Star Tracker.

Parameter	Value
Gyroscope	
Frequency	1 kHz
Accuracy	1 deg/s
Star Tracker	
Frequency	10 Hz
Accuracy	10 arcsec
External torque	
Torque on X axis	10 nN
Torque on Y axis	10 nN
Torque on Z axis	10 nN
Satellite moment of inertia	
1st principal axis	44×10^{-3} kg m^{-2}
2nd principal axis	44×10^{-3} kg m^{-2}
3rd principal axis	7×10^{-3} kg m^{-2}

Table C.1: Main parameters used in the MATLAB/SIMULINK ADCS model presented in this section.

The position estimate coming from the pseudo-IMU is then fed to a Kalman Filter to determine the best command to apply to the reaction wheels in order to keep the angular position at zero. The model assumes that the reaction wheels can be used to directly apply a torque to the satellite, with no delay and with an infinite precision. The torque $\boldsymbol{\tau}$ applied on the satellite is the sum of an external constant torque and of the torque from the reaction wheels:

$$\boldsymbol{\tau} = \boldsymbol{\tau}_{\text{external}} - \boldsymbol{\tau}_{\text{RW}} \tag{C.6}$$

with a minus sign since $\boldsymbol{\tau}_{\text{RW}}$ represents the torque applied on the reaction wheels, which is the opposite of the reaction torque applied onto the satellite.

Finally, the rotation of the satellite is then converted to a star position in the focal plane using appropriate geometrical transformations.

The main characteristics of this ADCS model are compiled in Table C.1.

C.3 Results

An example of a sequence of star position in the focal plane is given in Figure C.2, with a comparison to the real on-sky errors of the ASTERIA mission. The gain of the Kalman filters were empirically determined to obtain good performance of the control loop.

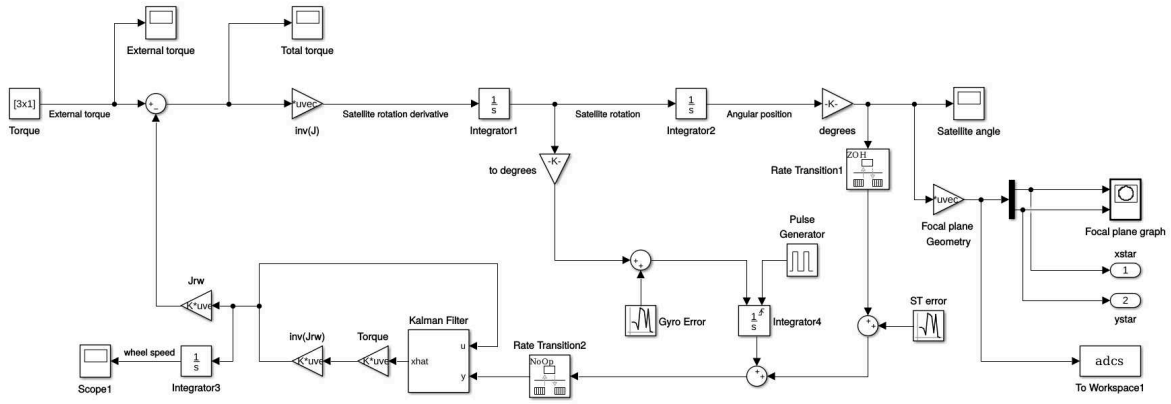
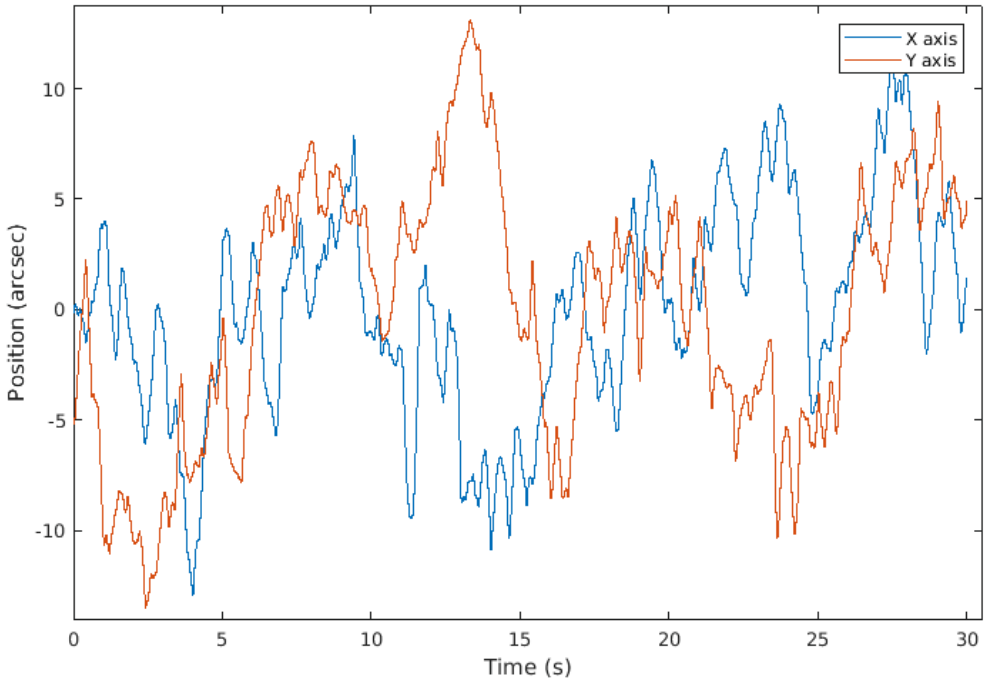
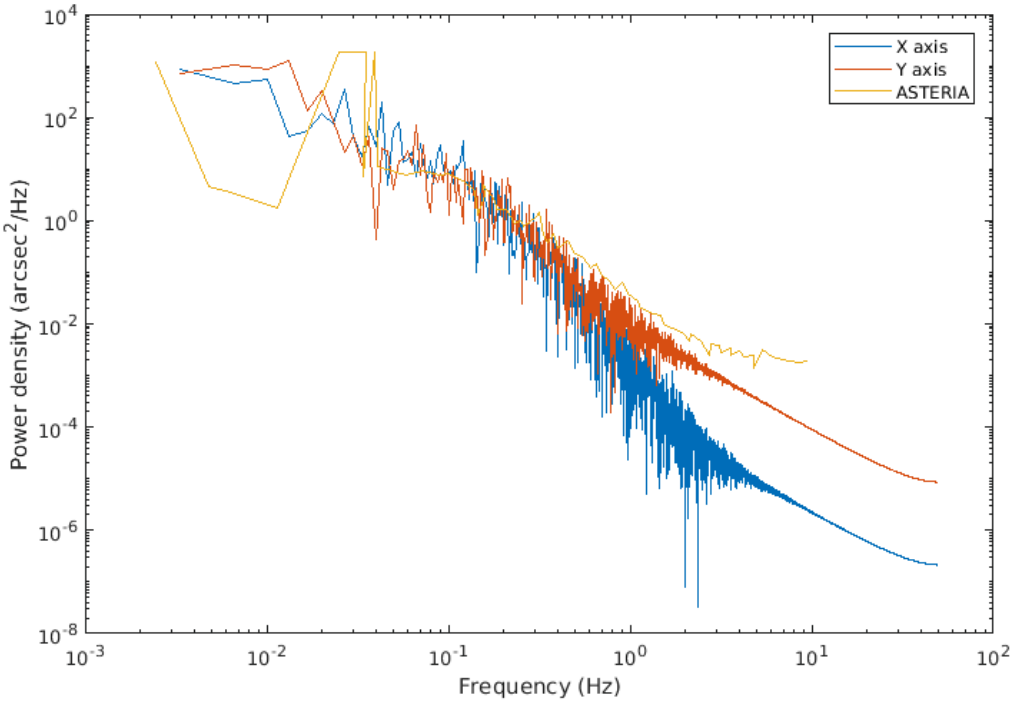


Figure C.1: A simple MATLAB/SIMULINK model of an ADCS used to simulate pointing errors. In this model, the ADCS uses a Star Tracker and a 3-axis gyroscope which are combined in a simple IMU to provide an estimate of the satellite angular position. This estimate is fed to a Kalman filter to determine the best torque to be applied to the satellite to counter-balance a constant external torque.



(a) Time series



(b) Power spectral density

Figure C.2: Time series and associated power spectral density of the star position in the focal plane of the PicSat telescope, generated by the simple ADCS model presented in this section. For comparison, the power spectral density obtained on-sky by the ASTERIA team is over-plotted (yellow curve, taken from Pong [2018]).

Annexe D

Résumé en français

D.1 Le système de beta Pictoris

D.1.1 L'étoile beta Pictoris

C'est depuis l'hémisphère sud, et seulement depuis cet hémisphère, que l'étoile beta[1] Pictoris est visible. Il s'agit d'une étoile de la constellation du Peintre (Pictor), située à 19.3 parsecs de la Terre (soit environ 60 années-lumière), et dont la magnitude apparente de 3.86 en fait une étoile de « $4^{\text{ème}}$ grandeur », facilement visible à l'oeil nu.

C'est une étoile de type A6, autrement dit une étoile relativement chaude, avec une température de surface d'environ 8000 K. Il s'agit d'une étoile relativement bien connue, puisque l'on possède actuellement des mesures de son rayon[2] (1.5 à 1.7 $R_{\odot}$), de sa gravité de surface ($\log(g) = 4.1$ dex), de sa masse ($M \simeq 1.8\ M_{\odot}$), et même de son âge (environ 23 millions d'années). Sa métallicité[3] est elle aussi connue [Fe/H] = 0.05. Le rapport d'abondance Carbone-Oxygène (rapport C/O) reste toutefois indéterminé, notamment parce que l'abondance d'oxygène n'a jusqu'à présent jamais été mesurée.

D.1.2 Le disque circumstellaire

L'étoile β Pictoris est somme toute une étoile relativement quelconque, qui ne doit finalement sa notoriété au sein de la communauté astronomique que par son environnement direct. C'est en 1984 que Smith et Terrile, guidés par des observations dans l'infrarouge faites quelques années auparavant, obtiennent la première image en bande visible du disque circumstellaire entourant β Pictoris (voir Figure D.1), qui est vu par la tranche depuis la Terre.

Étant donné le contexte de l'époque (aucune exoplanète n'a encore été découverte en 1984, puisque la première ne le sera qu'en 1995), cette découverte d'un disque autour d'une étoile attise la curiosité de plusieurs astronomes, et les observations se multiplient.

En 1987, Paresce et Burrows montrent que la lumière détectée n'est pas émise directement par le disque lui-même (rayonnement thermique), mais trouve plutôt son origine dans la réflexion de la lumière de l'étoile sur des grains d'une taille typique d'environ 1 micron contenus

[1] β, dans l'alphabet grec.

[2] En astronomie, les grandeurs sont souvent mesurées dans des unités rapportées au système Solaire, beaucoup plus parlantes que les unités du système métrique. On mesure les rayons stellaires en « rayons solaires » (symbole $R_{\odot}$), les masses stellaires en « masses solaires » ($M_{\odot}$), les rayons des planètes géantes en « rayons de Jupiter » (R_{Jup}), etc.

[3] Les astronomes nomment *métallicité* le rapport entre l'abondance des « métaux » (au sens astronomique, c'est-à-dire tout ce qui n'est pas de l'hydrogène ou de l'hélium), et l'abondance d'hydrogène. On donne souvent la métallicité en échelle $\log_{10}$, relativement à la valeur du Soleil ([Fe/H] = 0.05 correspond donc à une valeur de $10^{0.05} = 1.12$ fois celle du Soleil).

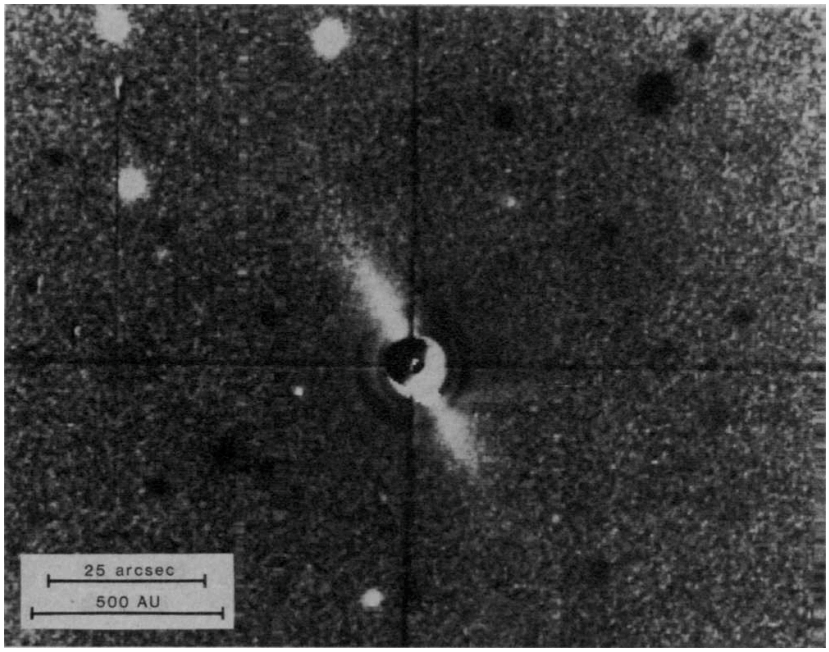


FIGURE D.1: Image historique du disque circumstellaire environnant l'étoile β Pictoris, obtenue par Smith et Terrile en 1984. L'étoile, située au centre du champ (intersection des lignes noires), est masquée par le coronographe, et le disque, vu par la tranche, apparaît comme une ligne. La structure apparente en demi-lune de l'étoile n'est qu'un effet instrumental, dû probablement à un défaut de centrage du masque coronographique.

dans le disque (lumière diffusée). Lagage et Pantin montrent, en 1994, que la zone proche de l'étoile ($\leq$ 40 UA, Unités Astronomiques) est appauvrie en poussière. Pour expliquer cette observation, ils suggèrent la présence d'une planète orbitant à l'intérieur de cette zone.

Des ceintures d'astéroïdes sont aussi identifiées dans le disque par Okamoto et al. en 2004, puis par Wilner et al. en 2011. Dans le système beaucoup plus interne, à moins d'une Unité Astronomique de distance de l'étoile, deux familles de comètes sont détectées, qui semblent être les principaux éléments structurants du système interne, notamment du fait que ces comètes enrichissent continuellement la région en gaz.

Le gaz lui-même, dans l'ensemble du disque de β Pictoris, est l'objet de nombreuses questions. Le système étant vu par la tranche, il est possible d'étudier la composition de ce gaz par spectroscopie d'absorption. Roberge et al. ont ainsi montré la présence de nombreux éléments chimiques différents, tous des *métaux*, au sens astronomique du terme : C, O, Na, Mg, Al, Si, S, Ca, Cr, Mn, Fe, Ni. Or, il se trouve que certains de ces éléments chimiques, à l'instar du sodium (Na), sont particulièrement sensibles à la pression de radiation[4]. Eût égard à la luminosité de l'étoile centrale, ces éléments devraient en théorie être rapidement soufflés en dehors du système, empêchant leur concentration en quantités suffisantes pour être détectables par nos instruments. L'explication la plus largement acceptée aujourd'hui est que ces différents éléments sensibles à la pression de radiation sont en fait retenus dans le disque par d'autres éléments, notamment le carbone et l'oxygène, peu sensibles à cette pression de radiation, et créant un effet de freinage piégeant les autres atomes.

Quoi qu'il en soit, il est clair que l'ensemble des éléments observés aujourd'hui dans le disque de β Pictoris (aussi bien pour ce qui concerne le gaz que la poussière) ne sont pas

[4]La pression de radiation est la pression exercée sur le disque par le flux lumineux provenant de l'étoile.

des résidus du disque originel, mais sont plutôt continuellement produits et détruits dans l'environnement actuel du disque. Le disque observé aujourd'hui est un *disque de débris*, résultat de l'évolution du disque protoplanétaire initial au cours des derniers 20 millions d'années environ.

D.1.3 La planète géante beta Pictoris b

La découverte, en 1995, de la première planète autour d'une étoile autre que le Soleil, puis les découvertes de nombreuses autres exoplanètes dans les années suivantes, ont probablement donné du poids aux hypothèses impliquant la présence d'une planète autour de β Pictoris pour expliquer certaines des structures observées. Au cours de la première moitié des années 2000, si aucune planète n'avait encore été découverte dans ce système, les indices de son existence s'accumulaient. Parmi ces indices, on notera l'observation de variations dans la photométrie (i.e. la luminosité) de l'étoile, rapportées par **?**, et interprétées comme le signe du passage de la planète devant l'étoile (voir aussi Section D.2).

Il faudra toutefois attendre la deuxième moitié des années 2000 pour que la planète géante β Pictoris b soit finalement détectée. Comme souvent en ce qui concerne l'astronomie « observationnelle », c'est une évolution *technique*, qui a, en grande partie, permis cette nouvelle observation. Le progrès rapide des systèmes d'optique adaptative, permettant de corriger en temps réel les effets de la turbulence atmosphérique sur les images astronomiques par l'utilisation d'un miroir déformable, a rendu possible l'obtention d'images extrêmement stables. Combinés à l'utilisation d'un coronographe, sorte de petit masque opaque placé au foyer du télescope pour bloquer la lumière de l'étoile, ces systèmes d'optique adaptative ont permis de former une nouvelle classe d'instruments, dits « imageurs à haut-contraste ». Ces instruments sont capables de masquer la lumière d'une source brillante (une étoile), pour faire apparaître la lumière d'un objet beaucoup plus ténu (une planète), placé dans son environnement immédiat.

En 2003, Lagrange et al. obtiennent la première "image" de β Pic b avec l'instrument NaCo, installé sur l'un des télescopes du Très Grand Télescope Européen (VLT), au Chili (voir Figure D.2). Toutefois, comme il est impossible sur le moment de certifier qu'il s'agit bien là d'une planète, et non pas, par exemple, d'une étoile d'arrière plan, la découverte ne sera publiée que quelques années plus tard. C'est en 2009 que la même équipe réussi à obtenir une deuxième image du même objet, permettant de constater qu'il se déplace autour de l'étoile, et donc d'en déduire qu'il s'agit bien d'une planète.

Après sa découverte, la planète β Pictoris b va rapidement faire l'objet d'une importante campagne d'observation, dans le but de déterminer son orbite, et certains de ses paramètres physiques. Les observations permettent à plusieurs équipes de mesurer[5] la température de la planète (entre 1550 et 1800 K, donc "très chaud"), la gravité de surface ($\log(g) \simeq 4\,\mathrm{dex}$), le

[5]Le terme de « mesure » tel qu'utilisé ici est tout à fait impropre. Les grandeurs physiques associées aux exoplanètes ne sont usuellement pas des observables directes. Autrement dit, il n'existe aucun instrument capable de *mesurer* par exemple la température à la surface d'une exoplanète. Les techniques utilisées reposent alors sur la recherche (théorique, ou empirique) de liens entre la grandeur d'intérêt (la température), et les grandeurs observables (la luminosité de la planète à différentes couleurs par exemple, ou mieux, son spectre lumineux). Notre compréhension de ces liens prend la forme de *modèles*, permettant d'imaginer, à partir d'une température, d'un rayon, d'une masse, à quoi doit ressembler la planète observée, et donc de calculer sa luminosité, son spectre, etc. En comparant les résultats des modèles obtenus avec différentes valeurs de température, rayon, masse, aux observations, on détermine alors quel est le modèle "le plus ressemblant", et donc quelles sont les valeurs des paramètres d'intérêt "les plus probables". Il existe des outils mathématiques permettant de quantifier rigoureusement cette "ressemblance" et cette "probabilité", mais il n'empêche que, *in fine*, les résultats que je présente ici comme des *mesures*, n'en sont pas. Ce sont plutôt des *estimations*, résultantes d'une interprétation des observations s'appuyant sur des modèles exoplanétaires, avec leur lot d'hypothèses, de simplifications, d'approximations, etc.

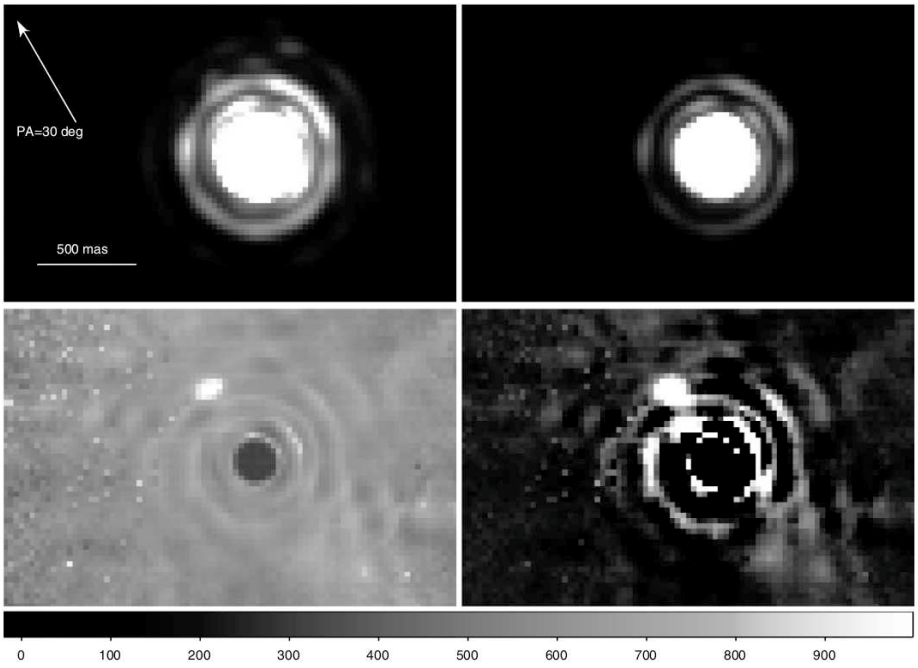


FIGURE D.2: La première observation de la planète géante β Pictoris b, obtenue en 2003 avec le VLT européen. L'image en haut à gauche est une image de l'étoile β Pictoris, obtenue en bande L' (infrarouge). L'étoile, très brillante, créé une « tâche de diffraction », dans laquelle rien n'est visible. Ce n'est qu'en observant une deuxième étoile de référence, dite « de calibration » (ici, l'étoile HR 2435, image en haut à droite), et en calculant le rapport (image en bas à gauche) ou la différence (en bas à droite) des deux, qu'il devient évident qu'un deuxième objet se cache dans la tâche de diffraction de β Pictoris. Figure tirée de Lagrange et al. [2009a].

rayon (environ 1.5 fois le rayon de Jupiter, soit 105 000 km), et la masse (10 à 12 fois la masse de Jupiter, ce qui fait de β Pic b une planète extrêmement massive).

En ce qui concerne l'orbite de β Pic b, une question importante est de savoir si oui ou non, vu depuis la Terre, la planète passe devant le disque de son étoile. Un tel phénomène, dit « transit photométrique », s'il se produit, est extrêmement intéressant pour la communauté astronomique, en ce qu'il permet de faire de nombreuses mesures importantes sur la planète en observant l'étoile au moment du transit. Un relevé de l'évolution de la position de β Pic b au cours du temps est effectué au cours de la première moitié des années 2010, et semble montrer que la planète va en effet passer devant son étoile entre l'été 2017 et l'été 2018. C'est le début de la campagne internationale d'observation de la conjonction de β Pic b, dans lequel s'inscrit pleinement le travail de thèse ici présenté.

Cette prédiction de transit est toutefois contredite fin 2016 par Wang et al., sur base de nouvelles observations obtenues avec l'instrument GPI, installé sur le télescope Gemini Sud. Wang et al. prédisent toutefois le passage de la sphère de Hill[6] de la planète devant l'étoile, en 2017-2018.

D.1.4 Les grandes questions en suspend

Le système de β Pictoris tient une place toute particulière dans le petit monde de la recherche sur les exoplanètes. Le fait que l'étoile soit proche, brillante, jeune, et que le système soit vu par la tranche font de β Pic une cible de prédilection pour l'étude de la formation des planètes géantes, de l'évolution des disques circumstellaires, et des interactions entre disques et planètes.

Aujourd'hui, β Pictoris semble être un système bien connu. Mais si les trois dernières décennies ont permis de se faire une idée de l'architecture générale du système, au moins dans ses grandes lignes, son origine reste bien mystérieuse, et de grandes questions restent aujourd'hui en suspend. Comment la planète géante β Pic b s'est-elle formée ? S'est-elle formée sur l'orbite où nous la voyons aujourd'hui, ou s'est elle formée ailleurs dans le système, pour ensuite migrer sur son orbite actuelle ? Quelle est sa composition chimique ? Peut-on lier cette composition actuelle à son origine, et à la nature du disque initial ? A quoi ressemblait le disque originel autour de β Pic, avant qu'il n'évolue pour donner le disque de débris qu'on connaît aujourd'hui ? Et β Pictoris b est-elle ou non une planète transitant devant son étoile ? Si non, qu'est-ce qui est à l'origine de l'évènement photométrique de 1981 ? Y a-t-il des satellites ou des anneaux autour de β Pic b ?

Il est clair que le travail présenté dans ce manuscrit ne saurait répondre à une telle avalanche de questions, pour la plupart extrêmement difficiles à aborder. Toutefois, ce sont bien ces questions là qui ont en grande partie motivé le travail ci-après présenté.

D.2 Vie et mort du nanosatellite PicSat

D.2.1 Objectifs scientifiques

Les objectifs scientifiques de la mission PicSat s'inscrivait directement dans le cadre des questions soulevées en Section D.1.4. L'objectif principal de la mission était la détection du transit de β Pictoris b, et la mesure[7] de son rayon. Le principe d'une telle mesure repose sur le fait

[6] La sphère de Hill d'une planète représente grossièrement la zone d'influence gravitationnelle de la planète. C'est une zone importante, puisqu'elle peut être l'endroit de la présence d'anneaux, lunes, ou autres petits corps environnant la planète. Bien sûr, la sphère de Hill peut tout aussi bien être vide...

[7] Et je parle bien ici de *mesure*, pas d'estimation comme précédemment. Les transits planétaires sont une des rares occasions de mesurer directement le rayon des exoplanètes, sans passer par la case "modèles".

que lorsqu'une planète de rayon R_P passe devant une étoile de rayon $R_\star$, elle en masque une partie. Précisément, elle masque une surface $\pi {R_\mathrm{P}}^2$ (surface projetée de la planète), sur un total de $\pi {R_\star}^2$ (surface projetée de l'étoile), soit une fraction $\alpha = (R_\mathrm{P}/R_\star)^2$. La conséquence directe est une diminution ΔF du flux lumineux de l'étoile donnée par :

$$\frac{\Delta F}{F} = \left(\frac{R_\mathrm{P}}{R_\star}\right)^2 \tag{D.1}$$

En mesurant la baisse de luminosité ΔF de l'étoile au moment du transit, et connaissant par ailleurs la taille de l'étoile (mesurée directement par des méthodes interférométriques), il est possible de déduire le rayon de la planète par une simple inversion de l'Équation D.1 :

$$R_\mathrm{P} = R_\star \times \sqrt{\frac{\Delta F}{F}} \tag{D.2}$$

Il est aussi possible de déterminer *a priori* quelle est la précision nécessaire sur la mesure de la luminosité de l'étoile ("précision photométrique") pour obtenir une certaine précision sur le rayon de la planète. Pour obtenir une mesure du rayon à 1% près, une précision photométrique de 630 ppm/hr (partie par million/heure) est nécessaire.

Un objectif secondaire de la mission PicSat était aussi la détection de comètes dans le système de β Pictoris, toujours par transit photométrique. De tels objets ont déjà été observés autour de β Pic, notamment par Kiefer et al., en 2014. Les variations photométriques relatives attendues sont de l'ordre de quelques 10^{-4}, pour des évènements durant typiquement quelques heures. En conséquence, leur détection requiert une précision photométrique de l'ordre de 100 ppm/hr.

Enfin, quoi que la mission PicSat ait souvent été présentée comme une mission astrophysique de mesure du transit photométrique, la vérité est qu'à l'origine, la mission avait surtout été pensée comme un démonstrateur technologique, précurseur nécessaire à de futures missions d'interférométrie spatiale. L'objectif technique de démontrer dans l'espace la stabilisation de l'injection de lumière stellaire dans une fibre optique monomode a largement contraint la conception de l'instrument. Quantitativement, pour espérer pouvoir poursuivre avec des missions d'interférométrie, une stabilité de l'injection à au moins 1%/hr était nécessaire.

D.2.2 Architecture de la mission

Pour parvenir à observer le transit de β Pictoris b initialement prévu pour fin 2017 - début 2018, il était nécessaire de développer et de lancer un satellite dédié en un très court laps de temps (moins de trois ans). Pour ce faire, il a été nécessaire de s'appuyer sur le concept de *CubeSat*, un standard de petit satellite né aux États-Unis.

PicSat a donc été pensé comme un CubeSat trois unités (3U), dont la taille totale est de $30 \times 10 \times 10\,\mathrm{cm}$. Au sein de ce volume très réduit, le satellite embarque des versions miniatures de l'ensemble des systèmes que l'on peut retrouver sur les plus gros satellites : système de contrôle d'attitude[8], système de communication radio, ordinateur de bord, et génération/distribution de puissance électrique (via les panneaux solaires). Une vue générale du satellite, ainsi qu'une photographie du modèle de vol de PicSat sont données en Figure D.3.

L'intérêt principal de respecter le standard CubeSat est de pouvoir bénéficier de lancements dits en « piggyback ». Les CubeSats lancés de la sorte sont simplement embarqués sur

[8]Le système de contrôle d'attitude (ou Attitude Determination and Control System, ADCS en anglais) est le système chargé de déterminer et de contrôler l'orientation du satellite dans l'espace. C'est lui qui a pour mission de pointer le satellite vers l'étoile.

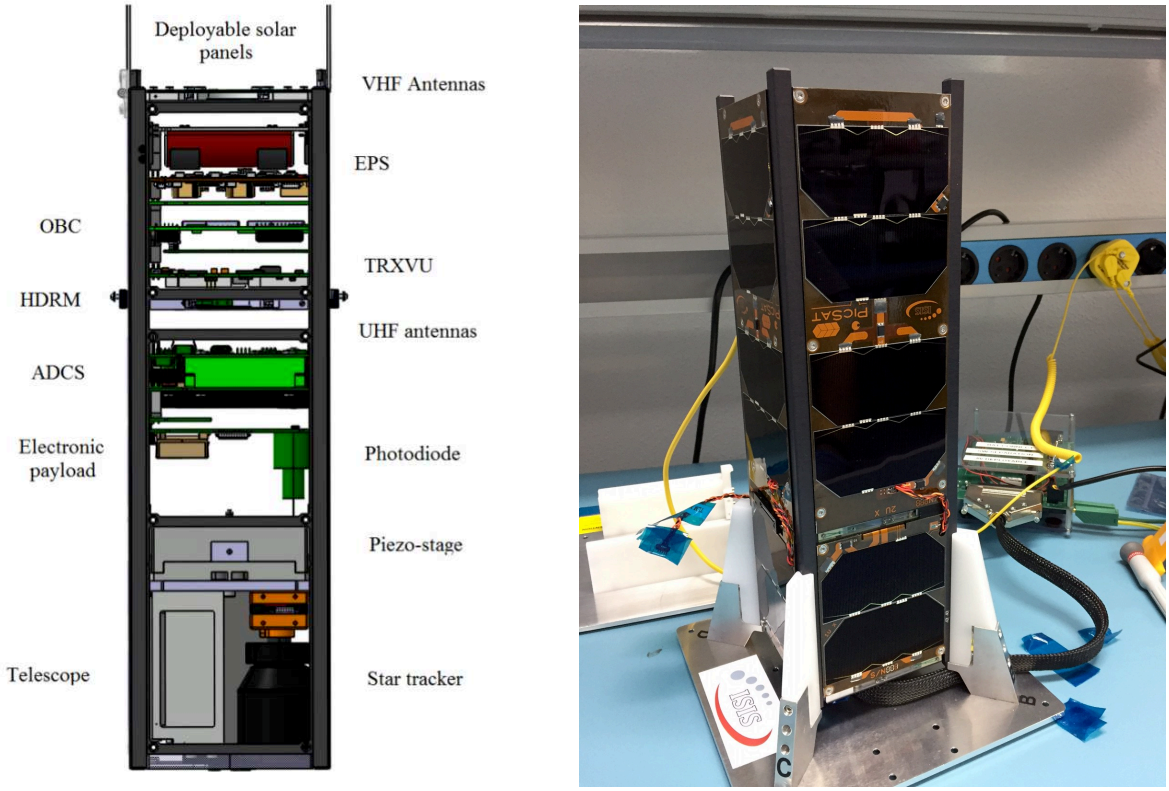


FIGURE D.3: À gauche : vue schématique du satellite PicSat. À droite : photographie du modèle de vol du satellite. Le satellite mesure 30 cm de hauteur.

un lanceur qui emporte un satellite de taille plus importante dans l'espace (la charge principale), et sur lequel il reste de la place. Une telle stratégie évite de devoir payer le coup complet du lancement, qui serait largement prohibitif. L'inconvénient est que le choix de l'orbite est limité, puisque largement fixé par la charge principale. PicSat a donc été développé en prenant soin de faire en sorte qu'il puisse fonctionner sur différents types d'orbites. Cela étant, la plupart des satellites lancés en orbite terrestre basse (de 500 a 800 km d'altitude environ) sont lancés sur des orbites héliosynchrones[9]. Ce fut aussi le cas de PicSat, lancé sur une orbite héliosynchrone à 505 km d'altitude.

Au total, PicSat est un satellite qui pèse environ 3.5 kg, et consomme environ 5 à 6 W de puissance électrique (l'équivalent d'une bonne lampe de poche). Son développement a pris environ 2 ans et demi, pour une équipe de 5 personnes. Le coût total du projet (incluant le satellite, la station de contrôle au sol, les tests, le lancement, les salaires, etc.) est d'environ 1.2 millions d'euros.

D.2.3 La charge-utile scientifique : un photomètre fibré

D.2.3.1 Concept

La *charge-utile* d'un satellite, est, par opposition à la *plateforme*, la partie "utile", du point de vue de l'utilisateur final (en l'occurrence, nous-mêmes). Si l'objectif de la plateforme est de permettre de contrôler le satellite, de communiquer avec lui, d'alimenter les sous-sytèmes en puissance électrique, cette plateforme elle-même n'est d'aucun intérêt vis-à-vis des objectifs directs de la mission. Elle n'existe que pour permettre de faire fonctionner la charge-utile en orbite.

Dans le cas de PicSat, la charge-utile est constituée d'un petit télescope optique, de 5 cm d'ouverture (diamètre du miroir primaire), et d'un détecteur à comptage de photons. Le schéma de principe est donnée en Figure D.4, et se résume de la manière suivante : lorsque le satellite est pointé vers son étoile cible, le miroir primaire (une parabole hors-axe) collecte la lumière de l'étoile, et la focalise sur une fibre optique monomode. Le miroir secondaire est un simple miroir plan, qui n'a d'autre but que de rendre l'instrument plus compact, en repliant le faisceau lumineux. Une fois injectée dans le fibre optique, la lumière est conduite vers un détecteur à avalanche, qui permet de détecter et de compter un par un les photons arrivant de l'étoile. Le nombre de photons collectés par unité de temps est une mesure directe de la luminosité de l'étoile. La difficulté majeure dans instrument comme celui-ci réside dans l'injection de la lumière au sein de la fibre optique.

Sans rentrer dans les nombreux détails techniques, une fibre optique peut être décrite comme une cavité optique résonnante, au sein de laquelle plusieurs modes peuvent se propager. Une fibre est dite « monomode » lorsque seul le mode fondamental de la cavité peut se propager. En pratique, une fibre monomode est une fibre dont le cœur est extrêmement petit.

La conséquence directe du caractère monomode d'un fibre optique, telle qu'utilisée par la charge-utile de PicSat, est qu'une onde lumineuse ne peut se propager à l'intérieur que si elle correspond parfaitement au mode fondamental. Si ce n'est pas le cas, seule une partie de l'énergie lumineuse se propage dans la fibre, le restant étant réfléchi au niveau de la tête de fibre. Le rapport de la puissance injectée dans la fibre sur la puissance du faisceau incident

[9]Une orbite héliosynchrone est, littéralement, une orbite synchronisée avec le Soleil. Il s'agit toujours d'orbites quasi-polaires (très inclinée, le satellite passant tour à tour du pôle Nord au pôle Sud), calculées de sorte que les effets induit par le léger aplatissement de la Terre fasse tourner le plan orbital au rythme d'un tour tous les ans. De ce fait, l'orbite conserver toujours la même orientation par rapport au Soleil. En conséquence, le satellite ne perçoit le défilement des saisons sur son orbite, ce qui permet de conserver les mêmes conditions d'éclairement tout au long de l'année.

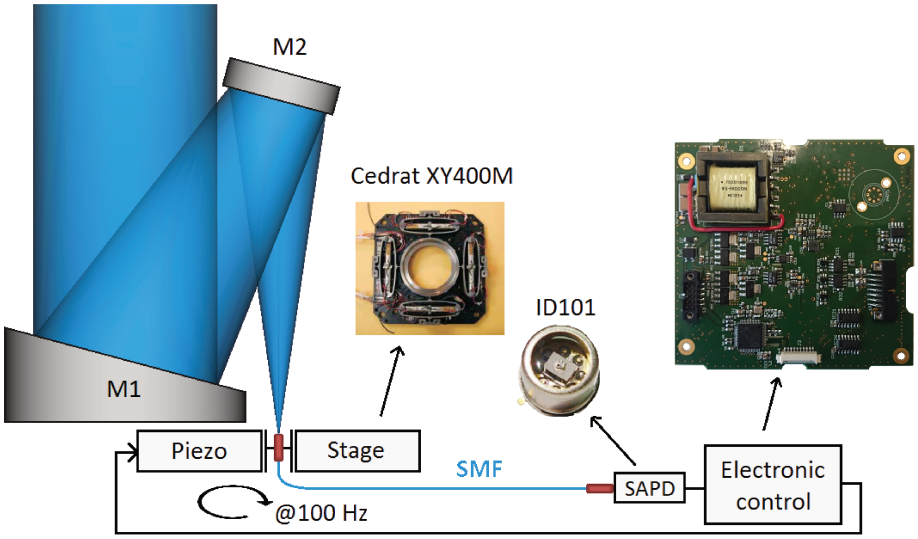


FIGURE D.4: Schéma de principe illustrant le fonctionnement de la charge-utile de PicSat. De gauche à droite : la lumière arrivant de l'étoile est collectée par un télescope constitué de deux miroir M1 et M2, puis injectée dans un fibre monomode (Single Mode Fiber, SMF en anglais). La fibre est montée sur une platine piezo-électrique deux axes (piezo stage), permettant de la déplacer pour la maintenir correctement centrée sur l'étoile. La lumière se propage dans la fibre jusqu'à au détecteur ID101, à comptage de photon. L'électronique de contrôle récupère les mesures de flux lumineux pour optimiser la position de la fibre.

est dit *rapport d'injection*, et est usuellement dénoté η[10]. En pratique, on démontre qu'en utilisant un télescope circulaire sans obstruction (i.e. un télescope hors-axe pour lequel le miroir secondaire ne fait pas d'ombre sur le primaire), il est possible d'atteindre un rapport d'injection $\eta = 0.8$, i.e. d'injecter 80% de la lumière dans le fibre. Toutefois, ce chiffre n'est valable que sous la condition d'avoir un télescope idéal, sans aberration optique d'aucune sorte, et d'avoir un alignement parfait. En présence d'une erreur d'alignement, le rapport d'injection chute très rapidement. La Figure D.5 donne l'évolution de ce rapport d'injection en fonction de l'erreur de pointage de la fibre, convertie en microns dans le plan focal du télescope de PicSat (ce qui correspond donc à la distance du centre de la fibre à l'image de l'étoile). Il suffit d'une erreur de moins de 1.5 microns pour faire passer l'injection de 80% à seulement 40%.

Les spécifications données par *Hyperion Technologies*, constructeur du système de contrôle d'attitude de PicSat, conduisent à un défaut de pointage du satellite d'environ 30 secondes d'angle. Avec un instrument d'environ 15 cm de longueur focale, une telle erreur correspond à un peu plus de 20 microns dans le plan focal. Autrement dit, le système de contrôle d'attitude de PicSat n'est pas capable de garantir l'injection de la lumière de l'étoile dans la fibre optique. C'est pourquoi la fibre optique est montée sur une platine piezo-électrique deux axes, permettant de la déplacer afin de garantir son alignement sur l'étoile.

L'ensemble de l'instrument est contrôlé par une carte électronique dédiée, elle-même connectée à l'ordinateur de bord du satellite. Le développement de cet instrument, et en particulier de l'algorithmique de pointage permettant de conserver la fibre parfaitement alignée sur l'étoile en dépit des vibrations du satellite, ainsi que la programmation de la carte électronique embarquée, ont constitué une part majeure de mon travail.

[10]Lettre grecque « eta »

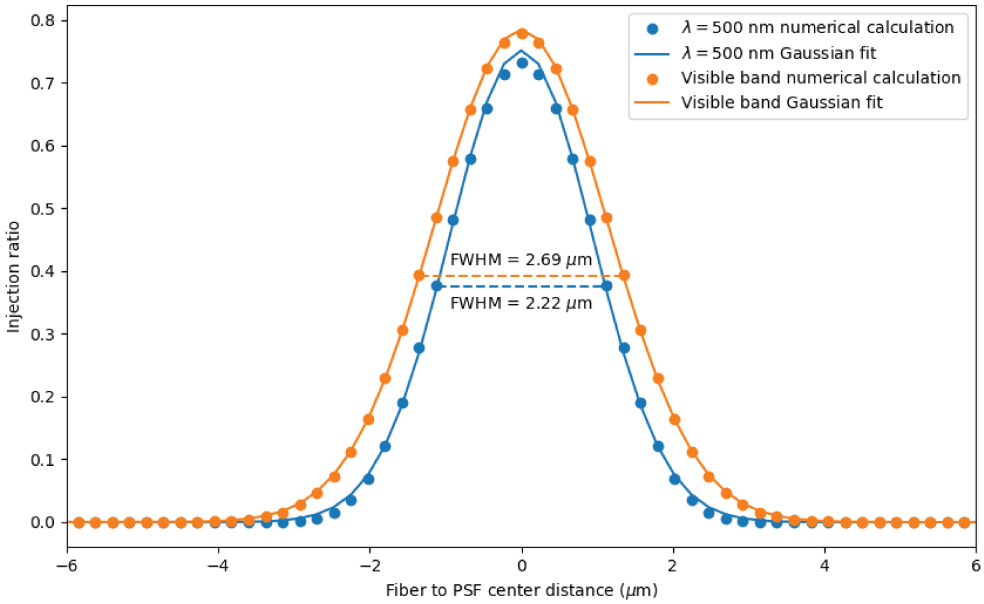


FIGURE D.5: Évolution du rapport d'injection η (fraction de la puissance lumineuse se propageant dans la fibre par rapport à la puissance lumineuse totale incidente sur le télescope) en fonction de la distance entre la fibre et le centre de l'image de l'étoile. La courbe bleue donne les valeurs à 500 nm de longueur d'onde (lumière bleue), et la courbe orange donne les valeurs intégrée sur l'ensemble du spectre visible, en tenant compte de la réponse instrumentale et du spectre typique de β Pictoris. Les points correspondent à des données calculées, et les lignes correspondent à un modèle Gaussien ajustant ces quelques points. La valeur de FWHM donne la largeur à mi-hauteur de chacun des deux courbes Gaussiennes. Deux choses sont visibles sur ces courbes : l'injection maximum atteignable est de 80%, et cette injection diminue rapidement avec l'erreur de positionnement de la fibre (il suffit d'un peu plus d'un micron pour tomber à 40%).

Source de bruit	Niveau	Commentaire
Lumière diffusée	1 ppm/hr	Largement limité par la fibre monomode
Courant d'obscurité	1 pmm/hr	-
Stabilité de gain	20 ppm/hr	Pour 100 μV de stabilité en tension
Bruit thermique	40 ppm/hr	-
Bruit de photon	60 ppm/hr	Bruit fondamental, indépassable
Stabilité de pointage	80 ppm/hr	Un objectif de ma thèse
Total	108 ppm/hr	Ce n'est pas une simple addition

TABLE D.1: Bilan photométrique de la mission PicSat

D.2.3.2 Bilan photométrique

La question fondamentale qui se pose lors de la conception d'un instrument scientifique est toujours : comment garantir les performances scientifiques de la mission ? Dans le cas d'une mission comme PicSat, garantir un retour scientifique revient à garantir un certain niveau de performance photométrique pour l'instrument (voir Section D.2.1). Pour ce faire, il est important d'identifier les principales sources possibles de contamination et/ou de bruit instrumental, via un « bilan photométrique ».

L'instrument à bord de PicSat a pour but de mesurer le flux lumineux en provenance de l'étoile (photométrie). Ce flux peut être contaminé par d'autres sources lumineuses, comme la lumière diffusée par la Lune, ou par le clair de Terre vu depuis l'orbite. Si cette lumière diffusée pénètre dans le télescope, puis dans la fibre optique, alors elle augmente artificiellement la photométrie mesurée par l'instrument, créant un biais de mesure. Dans le cas de PicSat, l'utilisation d'une fibre monomode, de très faible diamètre, minimise cette source de bruit. On peut montrer, par un calcul d'ordre de grandeur, que cette source de bruit reste limitée à moins de 1 ppm/hr.

Une deuxième source importante est formée par l'ensemble des bruits liés au détecteur lui-même. Une compréhension fine du fonctionnement des photodiodes à avalanche est nécessaire pour évaluer ces niveaux de bruits. Heureusement, le fournisseur de cet équipement fourni des données permettant de réaliser des estimations correctes. Une photodiode à avalanche nécessite une tension dite « de polarisation » permettant d'accélérer les électrons émis à l'arrivée d'un photon, pour créer l'avalanche à proprement parler, qui est détectée par l'électronique de lecture. Toute variation de tension résulte en une variation du gain du détecteur. L'électronique de PicSat doit garantir une stabilité en tension de 100 μV, conduisant à une stabilité de 20 ppm/hr au niveau de la photométrie. En parallèle de la tension, la température du détecteur a aussi un impact sur le gain de mesure. À nouveau, pour la mission PicSat, le détecteur est régulé en température à 0.01 degré Celsius de précision, afin de limiter le bruit à 40 ppm/hr. Enfin, des fluctuation électroniques sont aussi responsables d'un bruit dit de « courant d'obscurité », limité à 1 ppm/hr.

En plus du détecteur et des contamination extérieures, la précision de mesure est aussi limitée par un bruit beaucoup plus fondamental, inhérent à la nature corpusculaire de la lumière : le bruit de photon. Celui-ci dépend du flux observé, et, dans le cas d'une observation de β Pic avec un instrument de 3.5 cm de diamètre, il s'établit à environ 60 ppm/hr.

Finalement, la dernière source de bruit provient de la stabilité de pointage de la fibre optique, largement discutée dans la Section suivante. Pour limiter le bruit total à $\simeq$ 100 ppm/hr, ce bruit de pointage doit rester inférieur à 80 ppm/hr.

Ce bilan photométrique initial de la charge-utile est récapitulé en Table D.1.

D.2.4 Algorithmique de pointage

Comme indiqué précédemment, il est nécessaire, pour garantir une injection convenable de la lumière de l'étoile dans la fibre optique de PicSat, de faire en sorte que la fibre reste constamment alignée avec l'étoile, y compris en présence d'erreurs de pointage du satellite lui-même. C'est le rôle de l'algorithme de pointage de la charge-utile que de garantir cet alignement.

D.2.4.1 Quelle précision nécessaire ?

Le première question à laquelle il est nécessaire de répondre avant d'envisager les diverses possibilités pour réaliser un algorithme de pointage est : jusqu'où faut-il aller ? Quelle est la précision nécessaire pour la mission ?

PicSat étant une mission de photométrie, c'est-à-dire une mission visant à mesurer avec précision la luminosité de β Pictoris, il est non seulement nécessaire de garantir une bonne injection dans la fibre optique, mais aussi de faire ne sorte que cette injection soit stable. En effet, le flux lumineux mesuré par le détecteur à avalanche placé en aval de la fibre optique peut s'écrire comme le produit du flux de β Pictoris et du rapport d'injection :

$$F_{\text{mesuré}} = \eta \times F_{\beta \text{ Pic}} \tag{D.3}$$

En conséquence, toute fluctuation du rapport d'injection conduit à une fluctuation de flux mesuré, et donc à une augmentation de l'incertitude sur la mesure du flux de β Pictoris. Pour garantir une précision de mesure de ce flux à 100 ppm/hr, comme requis par les objectifs scientifiques de la mission, il est nécessaire de garantir une stabilité du rapport d'injection à 80 ppm/hr (voir bilan photométrique de la mission, en Table D.1), ce qui correspond à une stabilité de 13% à la fréquence de fonctionnement de la charge-utile (1 kHz).

La variation du rapport d'injection η en fonction de la position (X, Y) de la fibre par rapport au centre de l'étoile peut se décrire comme une fonction Gaussienne, d'amplitude $\eta_0 = 0.8$, et d'écart-type $\sigma = 1.14\,\mu$m :

$$\eta(X, Y) = \eta_0 \times \exp\left(-\frac{X^2 + Y^2}{2\sigma^2}\right) \tag{D.4}$$

Supposons maintenant que l'on cherche à maintenir la fibre autour d'une position de référence (X_m, Y_m) typiquement choisie à mi-hauteur de la fonction Gaussienne (i.e. $\eta(X_m, Y_m) = 0.5 \times \eta_0$, j'expliquerai ultérieurement pourquoi un tel choix). En présence d'une petite erreur de positionnement $(\delta X, \delta Y)$, l'injection peut s'écrire sous une forme linéarisée utilisant les dérivées de la fonction η :

$$\eta \simeq \frac{\eta_0}{2} + \frac{\partial \eta}{\partial X}(X_m, Y_m) \times \delta X + \frac{\partial \eta}{\partial Y}(Y_m, Y_m) \times \delta Y \tag{D.5}$$

À partir de là, il s'ensuit que des fluctuations du pointage en X, Y d'écart-types σ_X, σ_Y conduisent à des fluctuations d'injection d'écart-type relatif $\sigma_\eta/(\eta_0/2)$ donné par :

$$\frac{\sigma_\eta}{\eta_0/2} = \sqrt{\left(\frac{\partial \eta}{\partial X}\right)^2 {\sigma_X}^2 + \left(\frac{\partial \eta}{\partial Y}\right)^2 {\sigma_Y}^2} \tag{D.6}$$

Tous calculs faits, pour une erreur identique en X et Y : $\sigma_{\text{pos}} = \sigma_X = \sigma_Y$, on montre alors que :

$$\frac{\sigma_\eta}{\eta_0/2} = \frac{\sqrt{2\ln(2)} \times \sigma_{\text{pos}}}{\sigma} \tag{D.7}$$

Ce qui indique que pour garantir une stabilité à ±13% de l'injection, il est nécessaire de garantir que :

$$\sigma_{\text{pos}} < 0.13 \times \sqrt{2\ln(2)} \times \sigma = 0.13\ \mu\text{m} \tag{D.8}$$

Pour garantir les performances scientifiques de la mission, il faut donc que la boucle de régulation soit capable de contrôler la position de la fibre par rapport à l'étoile à une précision de 0.13 microns, ce qui n'est pas une mince affaire ! En pratique, cela signifie que des algorithmes relativement sophistiqués sont nécessaires.

D.2.4.2 Où est la difficulté ?

La précision de pointage requise pour les objectifs scientifiques est importante. Mais ce n'est pas dans ce chiffre de 0.13 microns de précision nécessaire que réside la principale difficulté de la boucle de pointage de PicSat. En réalité, le problème fondamental est que l'instrument de PicSat n'a pas de *champ de vue*. Autrement dit, l'instrument ne produit pas d'image du champ du ciel observé, à partir de laquelle il serait aisé de déterminer la position de l'étoile cible, et de positionner la fibre. Non, dans le cas de la charge-utile de PicSat, la seule mesure utilisable est la mesure de la photométrie, directement dépendante de l'injection η. Or, il n'est pas possible de déterminer de manière non-ambigüe la position de la fibre par rapport à l'étoile à partir d'une telle mesure photométrique. Dans un langage plus mathématique, il n'est pas possible d'inverser la fonction d'injection, afin de remonter à la position (X, Y) à partir d'une mesure de $\eta(X, Y)$. Au mieux, pour une fonction η gaussienne, telle que donnée en Équation D.4, une telle inversion ne peut donner que la *distance* à laquelle se situe la fibre par rapport à l'étoile. Ansi, si l'on constate par exemple que luminosité en sortie de fibre diminue, on peut en déduire que l'étoile s'éloigne de la fibre, mais sans savoir dans quelle direction.

D.2.4.3 Comment faire ?

La solution retenue pour la charge-utile de PicSat se base sur une *modulation* de la position de la fibre optique. Plutôt que d'essayer de garder la fibre parfaitement centrée sur l'étoile, la fibre est volontairement décentrée par rapport à l'étoile, et forcée à se déplacer autour. Si la fibre se déplace par exemple en cercle autour de l'étoile, alors il est possible, à partir de plusieurs mesures photométriques, de déterminer la position de l'étoile par « trilatération[11] », comme illustré en Figure D.6.

Bien sûr, la situation n'est pas si simple, et l'application d'une telle méthode de trilatération pose, en pratique, de nouvelles difficultés. Le bon fonctionnement de la trilatération suppose que les mesures de distance soient effectuées *simultanément*, ou bien, si ce n'est pas, que l'étoile reste fixe durant le temps nécessaire pour effectuer un tour de modulation avec la fibre. Dans les faits, la boucle de pointage tourne à 1 kHz, et un tour de modulation contient typiquement 10 points de mesure. La fréquence de la modulation est donc de 100 Hz environ. Les fréquences de vibration induites par le système de contrôle d'attitude du satellite peuvent aller jusqu'à 10 Hz. Dans les faits, ces vibrations sont trop importantes, et la fréquence de modulation n'est pas assez élevée pour garantir le bon fonctionnement de la trilatération.

La méthode développée pour contrôler la position de la fibre optique dans la charge-utile de PicSat s'inspire de cette idée de trilatération, mais l'intègre dans un algorithme

[11] La trilatération est une méthode mathématique de positionnement, cousine de la triangulation. Dans le cas de la triangulation, on détermine la position d'un objet à partir de mesures angulaires dans des triangles formés entre cet objet et un certain nombre des points de références. Dans le cas de la trilatération, procède de même, mais en remplaçant les mesures angulaires par des mesures de distance. Les systèmes de positionnement par satellite (comme par exemple le GPS) fonctionnent sur ce principe de trilatération.

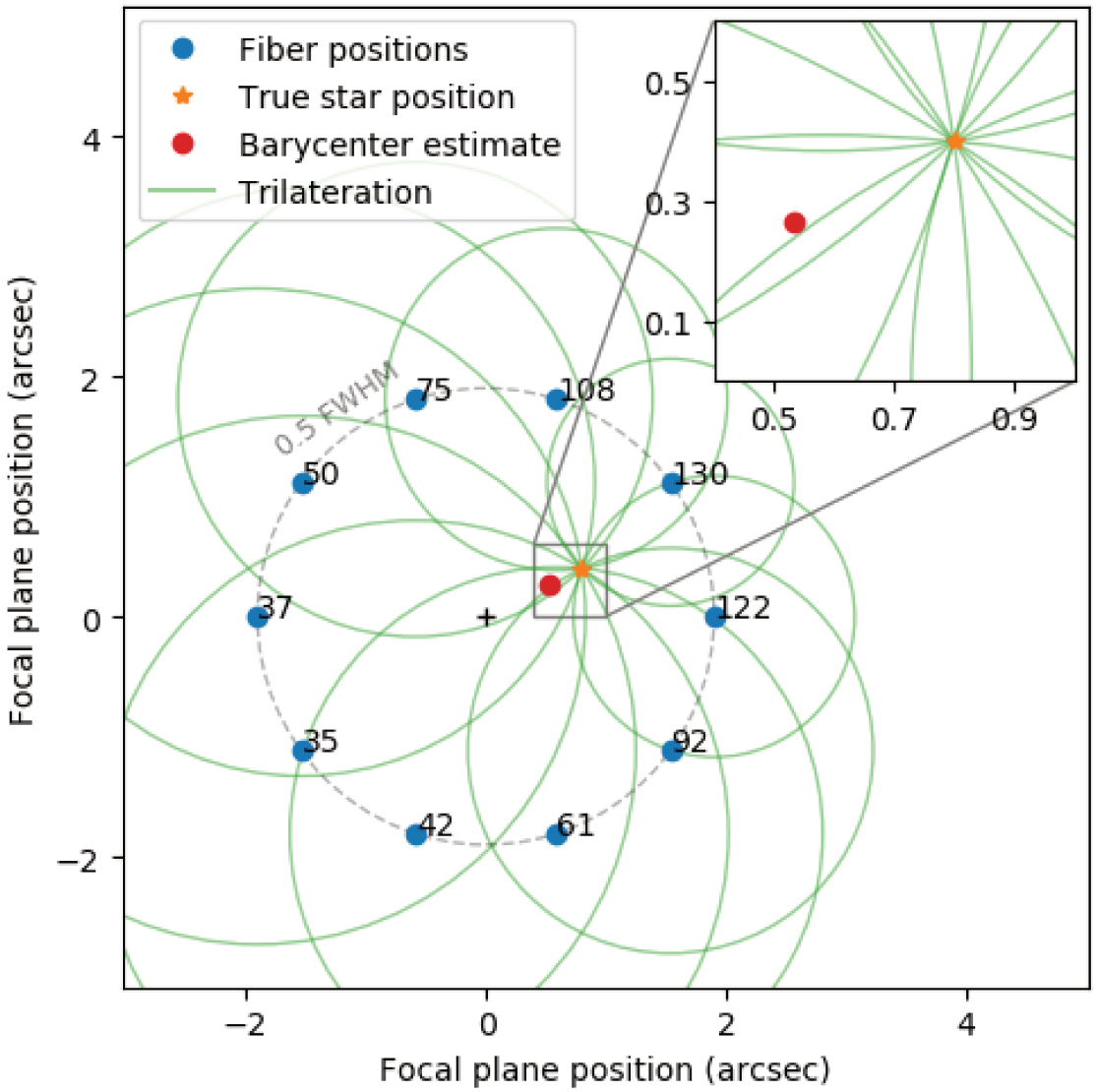


FIGURE D.6: Illustration du principe de la trilatération appliqué au cas de PicSat. L'étoile se situe à une position inconnue, matérialisée par le point orange. La fibre optique est déplacée sur un cercle centré sur une position arbitraire (cercle gris), et un certains nombre de mesures de photométrie sont prises (points bleus, avec des valeurs arbitraires données pour l'exemple). Plus la valeur de photométrie est élevée, plus l'étoile doit être proche de la position où la mesure a été prise. Chaque point correspond donc à une mesure de distance, qui donne donc un cercle sur lequel l'étoile se situe (cercles verts). En calculant l'intersection de tous ces cercle, on retrouve la position (inconnue) de l'étoile. Sur cette figure, le point rouge représente une autre estimation possible de la position de l'étoile, plus simple mais moins précise, qui consiste à calculer le centre de gravité des points de mesures.

ou chaque mesure est utilsée pour ce qu'elle est, indépendemment des mesures précédentes. Schématiquement, l'algorithme prend la forme suivante :

1. A partir de la connaissance *a priori* de la position et de la vitesse de l'étoile dans le champ de l'instrument, à l'instant t, on détermine la position de l'étoile à l'instant suivant $t + dt$.

2. On place la fibre optique autour de la position supposée de l'étoile à l'instant $t + dt$, en respectant le motif de modulation.

3. On mesure la photométrie en sortie de fibre, ce qui permet d'estimer l'injection η. Par inversion, on calcule la distance entre la fibre et l'étoile.

4. La seule connaissance de la distance entre la fibre et l'étoile ne donne normalement pas mieux qu'un cercle centré sur la fibre sur lequel peut se trouver l'étoile (voir Figure D.6 détaillant la trilatération). Toutefois, si l'on suppose que la position a priori de l'étoile (au point 2) n'est « pas trop mauvaise », on peut déterminer le point du cercle le plus proche de cette position supposée, et utiliser ce point comme mesure de position de l'étoile.

5. À partir de la connaissance initiale de la position de l'étoile et de sa vitesse, et en intégrant la nouvelle mesure de position obtenue au point précédent, on obtient une nouvelle estimation de la position et de la vitesse de déplacement de l'étoile dans le champ de l'instrument, valable à $t + dt$.

6. On reprend à partir de 1

L'intérêt d'un tel algorithme est de se débarrasser de l'hypothèse de fixité de l'étoile, nécessaire à la trilatération. Cette hypothèse est remplacée par une estimation de la vitesse de déplacement de l'étoile, faite conjointement à l'estimation de sa position.

D.2.4.4 Est-ce que ça marche ?

Déterminer la précision de pointage qu'il est possible d'atteindre à l'aide de la boucle de contrôle proposée s'est avéré délicat. Initialement, le plan de test incluait l'utilisation d'un banc dédié, sur lequel le satellite complet serait placé sur une table à coussin d'air, permettant une rotation libre sur un axe. Un écran placé au plan focal d'une lentille devait permettre de simuler un ciel étoilé à partir duquel le système de contrôle d'attitude se repère, et une source lumineuse collimatée et stable devait simuler la cible scientifique. L'objectif était de pouvoir faire fonctionner le système de contrôle d'attitude directement sur le banc, afin de déterminer les performances de la boucle de contrôle de la charge-utile en présence des vibrations réelles du satellite. Le développement de ce banc, difficile et particulièrement chronophage, a été abandonné.

Sans possibilité de tester la boucle de régulation en conditions réelles, le stratégie adoptée a fait la part belle aux simulations numériques. La boucle de contrôle et l'instrument scientifique ont été modélisés sous MATLAB/SIMULINK. Ce modèle, validé expérimentalement sur quelques cas simples à l'aide d'un banc de test approprié (Figure D.7), a ensuite été utilisé pour déterminer les performances de la charge-utile en présence de divers types de vibrations. Cette étude a permis de conclure que, modulo de nombreux détails passés sous silence dans ce résumé (choix du motif de modulation, modélisation de la réponse instrumentale, fréquence de fonctionnement de la boucle, etc.), les performances de la boucle de pointage seraient suffisantes pour garantir les objectifs scientifiques.

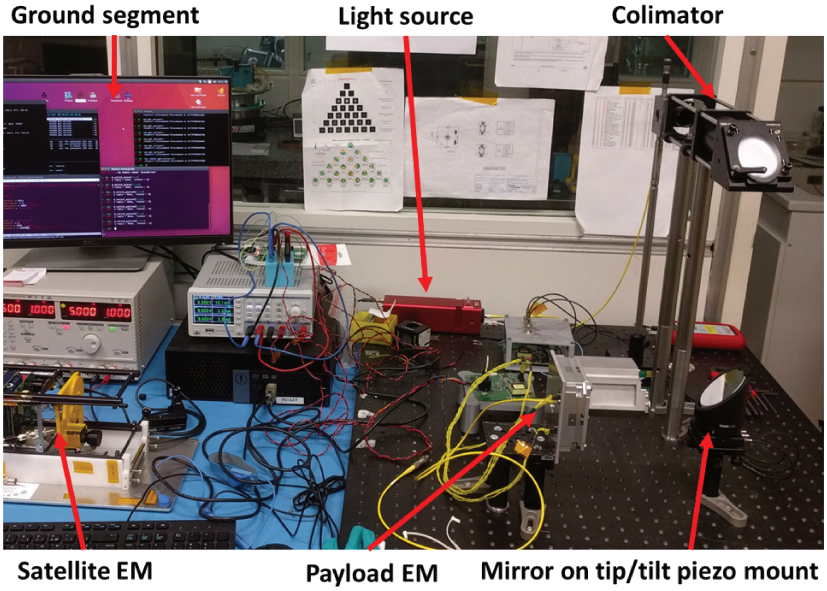


FIGURE D.7: Banc de test utilisé pour tester le système de pointage de la charge-utile. Une source blanche (boîtier rouge) est utilisée pour injecter de la lumière dans un système de collimation, simulant une source « à l'infini ». Le faisceau collimaté est dirigé vers un miroir contrôlé en tip/tilt, puis renvoyé vers le télescope de la charge-utile. En faisant vibrer le miroir tip/tilt, on simule des vibrations du satellites, et l'on peut vérifier que l'algorithme de pointage réagit en conséquence.

Mais cette conclusion, basée en grande partie sur des simulations, n'aura bien sûr jamais la valeur d'une démonstration en vol. Démonstration qui n'a malheureusement pas été possible avec la mission PicSat (voir Section D.2.9)

D.2.5 Logiciel embarqué

Entre le concept de boucle de contrôle introduit en Section D.2.4 et un système opérationnel sur la carte électronique de la charge-utile de PicSat, il y a une tâche importante, qui a aussi occupé une large part de ma thèse : la programmation embarquée.

Le développement du logiciel de la charge-utile s'est fait en deux niveaux : les *pilotes*[12] ont été écrits à partir de rien, directement en C/C++. Les fonctionnalités plus haut-niveau, incluant l'implémentation de l'algorithme lui-même, ont été écrites en s'appuyant sur les outils GERICOS, développés au LESIA.

En plus du logiciel de vol lui-même, il a aussi été décidé de développer un *chargeur d'amorçage* afin de permettre la reprogrammation en vol de l'électronique.

Les tâches principales du logiciel de vol de la charge-utile sont listées en Table D.2. Elles

[12]Dans le vocabulaire logiciel, un « pilote » est un petit logiciel, ou un morceau de code, faisant office d'interface bas-niveau avec les composants physiques. Par exemple, la position de la fibre optique dans la charge-utile est contrôlée en faisant varier la tension sur les actuateurs piezo-électriques, tension elle-même contrôlée par deux convertisseurs numérique-analogique connectés au microcontrôleur sur un bus SPI. L'envoi d'une série de commandes spécifiques sur le bus SPI est nécessaire pour modifier la tension, et déplacer la fibre. Le *pilote* de l'actuateur piezo-électrique est donc un petit logiciel prenant en charge ce processus très bas-niveau. L'utilisateur (ou le reste du logiciel) peut déplacer la fibre sans connaître les commandes spécifiques à envoyer au bus SPI, en déléguant cette tâche au pilote de l'actuateur piezo-électrique.

Tâche	Composants physiques impliqués
Récupérer les données de surveillance	Convertisseurs numériques, thermomètres, gyroscope
Régulation thermique du détecteur	Convertisseurs numériques et analogiques, photodiode
Acquisition de la photométrie	Horloges, compteurs, compteur binaire, photodiode
Pointage de l'étoile	Haute-tension, convertisseurs analogiques, gyroscope, piezo
Envoyer les données à l'ordinateur de bord	UART/RS422, DMA

TABLE D.2: Tâches principales du logiciel de vol de la charge-utile de PicSat.

ont trait essentiellement à l'acquisition des données scientifiques et des données de surveillance (tensions, températures, etc.). Le logiciel s'organise autour d'une machine d'états. Un mode principal, dit « science », est divisé en un ensemble de sous-modes permettant de séquencer la phase d'acquisition de l'étoile. Lorsque ce mode science est enclenché, sur réception d'une commande en provenance de l'utilisateur, le logiciel démarre une séquence d'imagerie, ou la fibre est déplacée progressivement sur tout le plan focal du télescope, afin de trouver la région du plan focal ou se situe l'étoile cible. Lorsque celle-ci est repérée, une bref transition par un mode de recherche approximative permet au logiciel de déterminer une estimation initiale plus précise de la position de l'étoile, nécessaire à l'initialisation et au verrouillage de la boucle de contrôle (mode *tracking*). Les transitions entre ces sous-modes sont entièrement automatiques, et la charge-utile réagit par elle-même lorsque l'étoile est perdue, et qu'il est nécessaire de repasser par une phase de recherche.

D.2.6 Réduction de données et performance scientifique

D.2.6.1 Le concept d'aberration optique

La Section D.2.3 expose le problème inhérent à l'injection de la lumière dans une fibre monomode : seule la partie de l'onde lumineuse qui correspond au mode fondamental de la fibre peut s'y propager, le reste étant réfléchi par la tête de fibre. L'une des difficultés engendrées par ce phénomène est celle du pointage, discutée en Section D.2.4. Toutefois, et contrairement à ce que pourrait laisser croire le bilan photométrique initial de la charge utile de PicSat, ce problème de pointage n'est qu'une facette du problème de l'injection dans la fibre.

Pour comprendre le problème dans son ensemble, il est nécessaire de passer par une description des *aberrations optiques* d'un télescope. Le principe de base d'un télescope optique est de focaliser la lumière en provenance d'une étoile en un point donné. En réalité, focaliser la lumière en un point unique est impossible. Au mieux, on peut focaliser la lumière dans une tâche, dont la taille typique est donnée par la limite de diffraction :

$$d \simeq \frac{\lambda}{D} \times F \tag{D.9}$$

Avec d la taille de l'image, D le diamètre du télescope, F la longueur focale, et λ la longueur d'onde de la lumière observée. Pour $D = 3.5\,\mathrm{cm}$, $F = 14\,\mathrm{cm}$, et $\lambda = 500\,\mathrm{nm}$, ce qui correspond au télescope de PicSat observant en lumière visible, on obtient la valeur de 2 microns, ce qui – et ce n'est pas un hasard ! – correspond grossièrement au diamètre de la fibre optique.

Pour focaliser de la sorte la lumière, le miroir primaire du télescope doit avoir une forme particulière. Pour un télescope comme celui de PicSat, dont le secondaire est plan, le miroir primaire est parabolique. Toute déviation par rapport à cette forme optimale résulte en une déformation de la tâche image du télescope. On parle d'*aberration optique*.

Mathématiquement, on peut décrire toute déformation d'un miroir comme la somme d'un certain nombre de déformations « élémentaires », décrites par les *polynômes de Zernike*. En

théorie, le nombre de ces polynômes est infini, mais en pratique, les déformations d'un miroir correspondent souvent aux polynômes de bas ordres. La Figure D.8 donne une représentation graphique des quelques premiers polynômes, et de l'impact d'une déformation de ce type sur la tâche image. On note sur cette Figure que le cas de référence (cas idéal, sans aberration), donne bien une tâche d'environ 2 microns, ronde, avec un anneau faiblement visible qui l'entoure.

Dès lors qu'il y a des aberrations dans le système optique, la tâche image se déforme. Il est toutefois important de noter que cette déformation n'entraîne pas de perte d'énergie. L'ensemble de l'énergie présente dans le cas de référence l'est toujours dans les cas suivants. La seule différence est que cette énergie est distribuée différemment dans l'image. En conséquence, si l'on utilse un instrument « classique », avec un détecteur bi-dimensionnel permettant de faire réellement une image, il est toujours possible d'intégrer l'ensemble de l'énergie reçue, simplement en additionnant les pixels contenus dans la tâche image. On peut donc dans ce cas toujours avoir une mesure correcte de la luminosité de l'étoile, même en présence de déformations du miroir.

Mais la situation change complètement avec l'utilisation d'une fibre optique. Dans ce cas, en effet, la seule mesure accessible est celle de l'énergie qui se propage dans la fibre. Or, si la tâche image se déforme, cela signifie aussi que le coefficient d'injection dans la fibre change, puisque la forme de l'image s'éloigne du mode fondamental. Bien sur, cette injection dépend toujours aussi de la position de la fibre par rapport à l'image de l'étoile. On peut représenter la situation en utilisant la fonction d'injection, qui donne le coefficient d'injection en fonction de la position de la fibre dans le plan image (colonne la plus à droite dans la Figure D.8). S'il était possible de mesurer l'ensemble de cette fonction d'injection, alors il serait potentiellement envisageable de reconstruire la forme de l'image, et par là d'obtenir une mesure correcte de la luminosité de l'étoile. Mais en réalité, la fibre ne permet de réaliser une mesure qu'à une seule position. Même en se plaçant au centre, là où est situé le maximum d'injection, la Figure D.8 montre bien que la présence d'aberrations optiques a un impact direct sur la mesure.

D.2.6.2 Variations thermiques et fluctuations de la photométrie

Les aberrations optiques conduisent à des variations de l'injection dans la fibre, voilà un fait établi. Mais quelles conséquences pour PicSat ? Pour répondre à cette question, il est nécessaire de passer par une étape de simulation thermique de l'instrument. Ce sont en effet les variations thermiques du télescope qui sont susceptibles d'engendrer des déformations du miroir. Une telle étude thermique a été réalisée pour PicSat, et s'est déroulée en deux temps. Une première étude a été faite pour déterminer quelles variations de température le satellite était susceptible de subir une fois sur son orbite. À partir des résultats de cette première phase (voir la Figure D.9), une deuxième étude a été effectuée pour déterminer l'étendue des déformations mécaniques du télescope.

En se basant sur les résultats de cette étude thermique, il est alors possible de modéliser l'évolution des différentes aberrations optiques de l'instrument au cours d'une orbite, en d'en déduire les variations de l'injection dans la fibre optique. Le résultat de ce travail, présenté en Figure D.10, n'est pas très encourageant : l'injection varie de 80% sur la durée d'une orbite ! Avec de telles fluctuations, il est impossible d'espérer pourvoir faire une mesure correcte de la luminosité de l'étoile cible.

D.2.6.3 Réduction des données et calibration photométrique

Pour améliorer la qualité des données PicSat, et les rendre utilisables d'un point de vue scientifique, il est nécessaire de corriger ce problème de fluctuations thermiques *a posteriori*, lors de la phase de réduction des données.

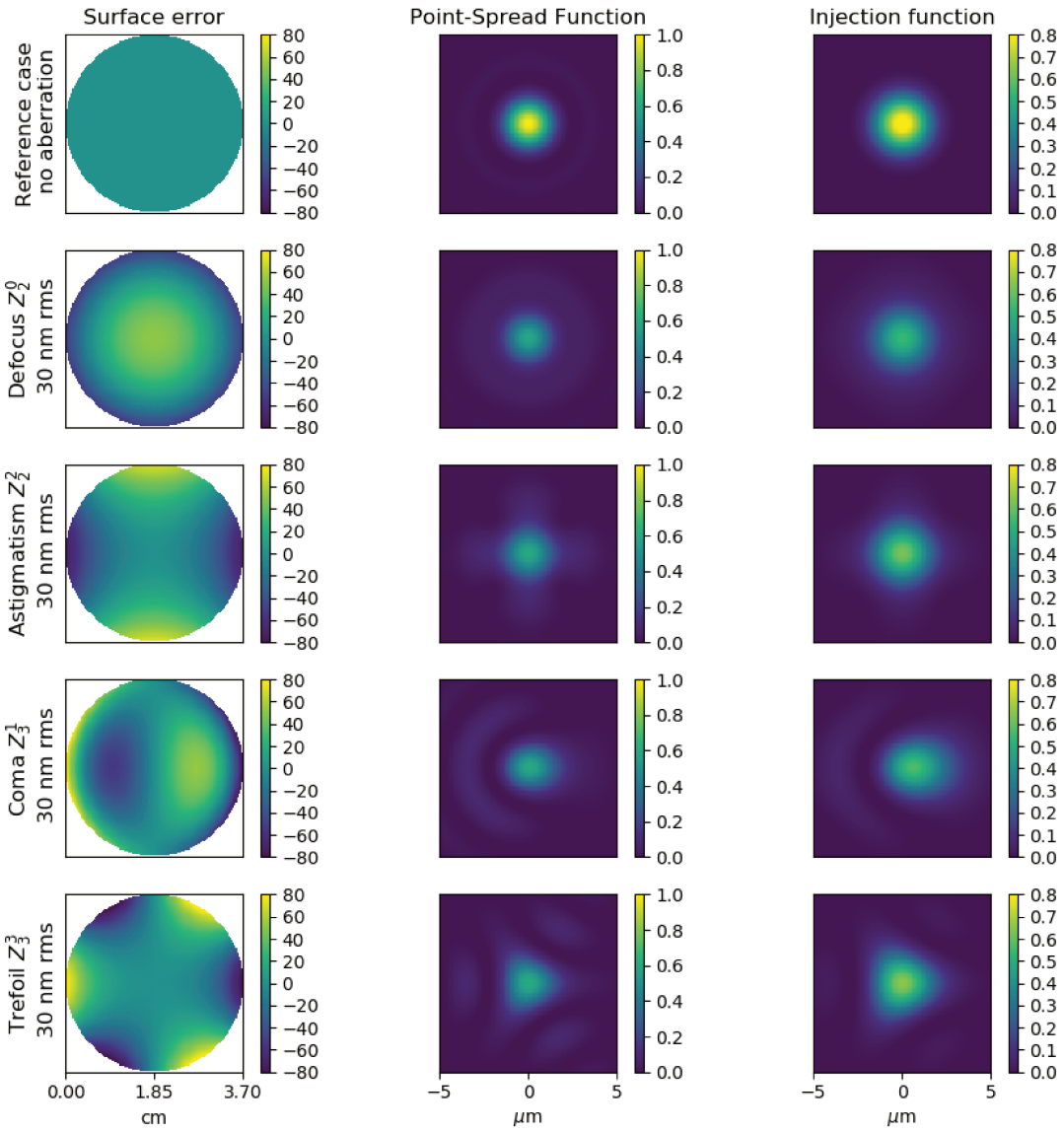


FIGURE D.8: Les polynômes de Zernike, utilisés pour représenter les aberrations optiques. La première colonne représente l'erreur de surface sur le miroir (différence entre la surface réelle et la surface parabolique de référence). La deuxième colonne donne la forme de la tâche image. La troisième colonne donne la forme de la fonction d'injection dans une fibre monomode (c'est-à-dire l'évolution du rapport d'injection en fonction de la position de la fibre par rapport au centre de la tâche image).

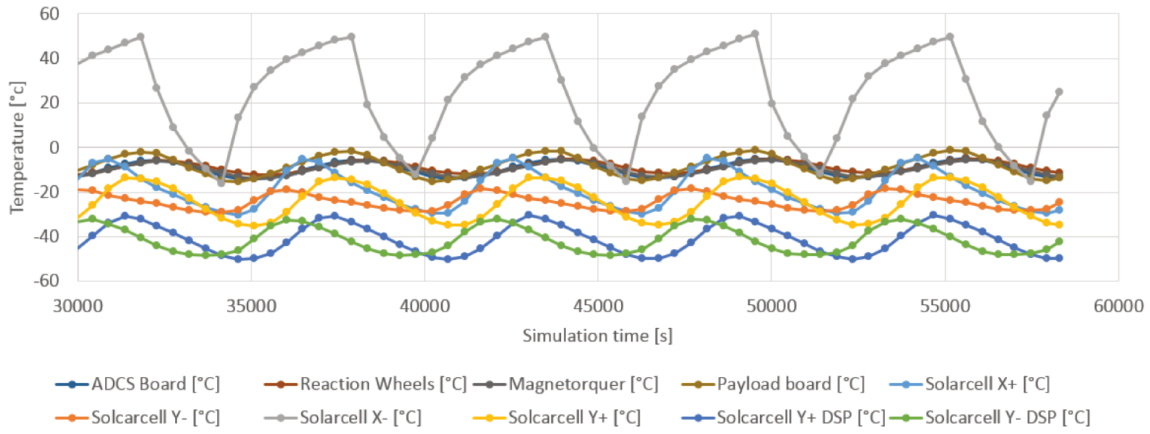


FIGURE D.9: Variations de la température au cours de l'orbite, pour différents éléments du satellite.

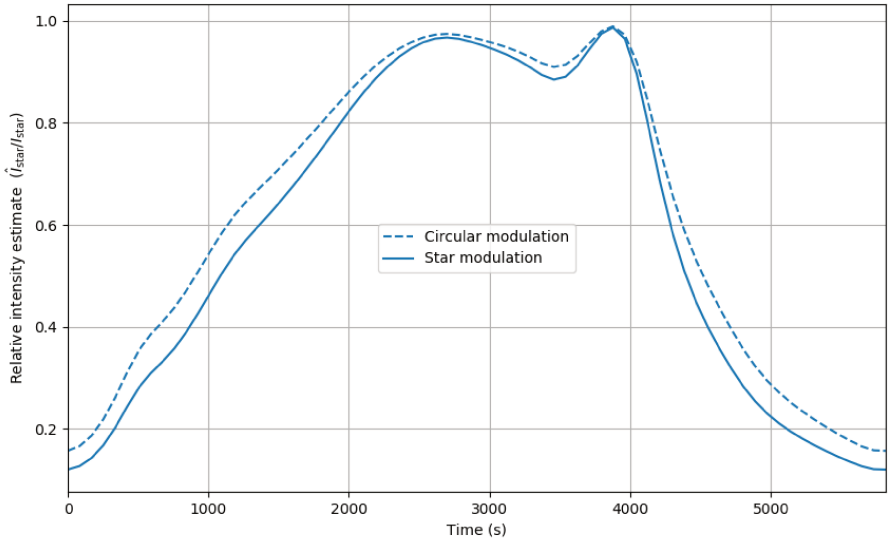


FIGURE D.10: Variations photométriques au cours d'une orbite, résultant des fluctuations thermiques du satellite (voir Figure D.9. Les deux courbes illustrent les résultats obtenus avec deux motifs de modulation possibles.

La Section D.2.4, qui décrit le développement de l'algorithme de pointage fin de la fibre optique, introduit le concept de « modulation » de la position de la fibre. Du fait de cette modulation, les mesures de photométrie effectuées par PicSat ne sont pas toutes obtenues au même point sur la « fonction d'injection » (voir Figure D.8). Si le motif de modulation utilisé contient une dizaine de points répartis en cercle, alors la photométrie sera mesurée sur une dizaine de points répartis en cercle autour de maximum de la fonction d'injection. En conséquence, il est envisageable de « deviner » la forme sous-jacente de la fonction d'injection à partir de ces mesures obtenues en différents points, et d'utiliser cette information pour corriger les mesures.

En réalité, deviner la forme complète de la fonction d'injection à partir de seulement quelques points est trop difficile. Toutefois, cette idée illustre bien le fait que les mesures effectuées contiennent une information quant à la forme réelle de la fonction d'injection, qu'il doit être possible d'utiliser pour améliorer les mesures. La méthode naïve de réduction des données, qui consiste à calculer la moyenne les points obtenus en chaque position, sans chercher à utiliser cette information encodée par les différences entre les points, est sous-optimale.

Une méthode plus avancée, mais restant toutefois suffisamment simple pour pouvoir être exécutée par l'ordinateur de bord du satellite, se fonde sur les principes de l'analyse en composantes principales. Sans rentrer dans la théorie mathématique sous-jacente, le principe d'une telle méthode repose sur l'hypothèse que les fluctuations de mesure, quoi qu'en partie aléatoires, ont tout de même une certaine régularité. En reprenant la Figure D.8, on peut noter que si les fluctuations sont, par exemple, dominées par les effets de tréfoil (aberration sur la ligne du bas), alors la fonction d'injection a une forme triangulaire, donc une certaine symétrie. Dans un tel cas, si les points de modulations étaient répartis sur un hexagone correctement orienté, la somme de deux points consécutifs serait probablement peu sensible aux fluctuations (si l'injection diminue sur un point, elle augmente sur le point suivant). D'un manière plus générale, selon le ou les type(s) d'aberration optique présent(s), il doit exister une combinaison optimale des points de modulation, pas forcément égale à la simple moyenne, qui doit donner l'estimation la plus robuste de la photométrie de l'étoile. Cette combinaison optimale peut être trouvée en analysant les structures des fluctuations de photométrie sur une série de données existantes (via la technique mathématique de décomposition en composantes principales). Toutefois, pour pouvoir être utilisée, cette méthode nécessite de réaliser au préalable une observation dite « de calibration », sur une étoile dont on sait que la luminosité est stable au cours du temps.

Cette technique, utilisée sur des données simulées en tenant compte des variations thermiques de l'instrument et des fluctuations thermique attendues sur l'orbite, montre tout de même que, contrairement au budget photométrique initialement établi, une précision supérieure à 500 ppm/hr aurait été difficile à obtenir avec un instrument comme celui de PicSat. Pour améliorer la situation, il serait nécessaire de revoir en partie la conception mécanique de l'instrument, en privilégiant l'utilisation de matériaux peu sensibles aux déformations thermomécaniques (alliages dits *invar*, Zerodur, etc.).

D.2.7 Les campagnes de tests et l'alignement optique

D.2.7.1 Campagnes de tests

L'environnement spatial est assez différent de l'environnement d'un laboratoire sur Terre. Du fait du vide spatial, la convection thermique est inexistante. Les systèmes placés en orbite ne peuvent donc se refroidir que par rayonnement. En ajoutant à cela les cycles orbitaux, au cours desquels un satellite passe successivement d'une phase éclairée par le Soleil à une phase d'éclipse, on obtient des cycles thermiques importants, qui mettent à rude épreuve les

Date	Type	Matériel testé	Objectif	Résultat
Jan 2016	Vibrations	Électronique, diode, piezo	Tester le matériel non-qualifié	Diode cassée (durant le transport), piezo cassé
Apr 2016	Vibrations	Électronique, diode	Tester la diode non-testée en Janvier	Diode en bon-état
Apr 2016	Vide thermique	Électronique, diode	Voir si ça crame pas	Pas de point chaud sur l'électronique
Jan 2017	Vibrations	Satellite (modèle de test)	Vérifier la robustesse de l'alignement	L'alignement tient le coup
Fév 2017	Vide thermique	Satellite (modèle de test)	Vérifier la régulation thermique de la diode	Forte variations, secondaire cassé
Jui 2017	Vide thermique	Satellite (modèle de test)	Vérifier le nouveau secondaire	Il tient le coup
Aou 2017	Vide thermique	Satellite (modèle de vol)	Vérifier le modèle de vol	Succès
Oct 2017	Chocs	Charge-utile (modèle de test)	Vérifier la robustesse de l'alignement	Ca tient le coup

TABLE D.3: Liste des essais (qualification et validation) effectués au cours du développement de la mission PicSat.

matériaux et les systèmes internes. Sans atmosphère pour bloquer les rayonnements cosmiques, les électroniques sont aussi soumises à d'importantes radiations. Enfin, pour rejoindre l'orbite terrestre, il est nécessaire d'en passer par la phase du lancement, au cours duquel l'ensemble du satellite est soumis à de très fortes vibrations, ainsi qu'à des chocs violents, lors de la séparation des étages de la fusée.

Non seulement l'environnement spatial est dangereux, mais en plus, sauf cas très particuliers, les coûts prohibitifs des lancements rendent toute réparation en orbite d'un satellite défectueux impossible. Il faut donc s'assurer que le satellite résistera au lancement et au vide spatial avant de le placer dans la fusée. Pour ce faire, des campagnes de tests sont réalisées au sol.

On distingue deux grands types de campagnes. Les campagnes « de qualification » sont des campagnes de tests effectuées dans des conditions similaires aux conditions rencontrées dans l'espace et dans le lanceur. Afin d'éviter de fatiguer inutilement le modèle de vol du satellite, au risque d'en réduire la durée de vie en orbite, ces tests de qualification sont réalisés sur un modèle dit « d'ingénierie », en tout point identique au modèle de vol, mais qui ne sera pas envoyer en orbite. Le deuxième type de campagne est constitué des campagnes dites « de validation », au cours desquelles le modèle de vol lui-même est testé, à des niveaux inférieurs aux niveaux des campagnes de qualification. Les tests de validation n'ont pas vocation à démontrer que le satellite résistera au lancement et aux conditions en orbite (c'est le but des essais de qualification sur le modèle d'ingénierie), mais sont plutôt réalisés pour vérifier l'absence d'anomalie majeure sur le modèle de vol.

La Table D.3 regroupe la liste des campagnes de tests effectuées tout au long du développement de la mission PicSat.

Plusieurs anomalies importantes ont été relevées au cours de ces campagnes de tests. La première anomalie a été détectée lors de tests en vibration effectués sur certains composants critiques de la charge-utile du satellite. Au cours de ces tests, la platine piezo-électrique qui avait initialement été sélectionnée pour la mission a été endommagée. Il a donc été nécessaire de reprendre toute la conception mécanique de la charge-utile, en utilisant une platine piezo-életcrique différente, dont le constructeur garantissait la résistance aux vibrations.

FIGURE D.11: État du miroir secondaire du télescope de PicSat après un test de cyclage thermique réalisé sous cloche à vide. L'aluminure du miroir est fortement endommagée, comme en témoignent les rayures bien visibles.

Les tests thermiques ont eux aussi révélé d'importants problèmes, notamment au niveau du miroir secondaire du télescope optique. La Figure D.11 donne un aperçu de l'état du miroir après une série de cycles thermiques effectués dans la cuve SiEnOm, au LESIA. Les « rayures » bien visibles traduisent un problème au niveau de l'aluminure, inadaptée à l'environnement spatial. À nouveau, il a été nécessaire de passer par un fournisseur de composants adaptés à l'environnement spatial.

Du fait des ressources limitées du projet, et de la difficulté inhérente à de tels tests, aucun essai en radiation n'a été effectué.

D.2.8 Segment sol et collaboration avec les radioamateurs

Un satellite, aussi performant et robuste soit-il, n'est d'aucune utilité si l'on ne peut pas communiquer avec pour lui envoyer des commandes, et recevoir des données. Cette problématique d'envoi et de réception de données est au cœur de la partie dite « segment sol » d'une mission spatiale.

Quoique ce ne soit usuellement pas la première chose qui vienne à l'esprit lorsque l'on pense « mission spatiale », le segment sol tient pourtant une place majeure tant dans la phase de développement que dans la phase d'opération. La mission PicSat a été la première mission spatiale opérée depuis le site de l'Observatoire de Meudon. En conséquence, il a fallu développer l'ensemble des outils nécessaires : standard de communication (format des paquets de données), base de données mission, logiciel de réception et de décodage, logiciel de commande et contrôle, logiciels de visualisations des données, et bien sûr, antennes au sol.

La Figure D.12 donne un aperçu de la station sol de la mission PicSat, sur le site de Meudon, par une froide journée de Janvier 2018. Les antennes utilisées sont des antennes de type Yagi[13], opérant à 145 MHz pour le lien montant (sol-espace), et à 435 Mhz pour le lien descendant (espace-sol). Ces deux fréquences, comme un certain nombre d'autres fréquences du spectre radio, sont des fréquences dites « amateures », réservées à une utilisation non-commerciale, par les personnes intéressées par l'étude des communications radio. Pour opérer des installations émettant sur les fréquences du service d'amateur, il est nécessaire d'obtenir un certificat de radio-amateur. En France, ce certificat est délivré par l'Agence Nationale des Fréquences (ANFR), après passage d'un examen de théorie radio, et d'un examen de

[13] Des râteaux, quoi...

FIGURE D.12: Les antennes de la station sol construite à Meudon, par une froide journée de Janvier. La plus petite des deux antennes est l'antenne UHF fonctionnant à 435 MHz pour la réception des données du satellite. L'antenne la plus grande fonctionne en VHF, à 145 MHz, et est utilisée pour l'envoi des commandes vers le satellite.

réglementation.

Le choix de l'utilisation de fréquences amateures pour la mission PicSat était motivé par deux considérations. D'abord, disons-le nettement, l'utilisation de ces fréquences est gratuite. Pour une mission aux moyens limités comme PicSat, c'est bien sûr un avantage important. Mais, outre cela, l'utilisation des fréquences amateures a aussi permis d'impliquer la communauté des radio-amateurs dans la réception des données. Le développement d'un site internet détaillant la mission, une communication honnête via Twitter sur la mission et l'utilisation des données récupérées, et un format de communication ouvert (i.e. n'importe qui écoutant les fréquences de PicSat pouvait non seulement recevoir les données, mais aussi les décoder), ont probablement été à l'origine de l'engouement important de la communauté amateure pour la mission PicSat. La Figure D.13 donne un aperçu des stations amateures ayant récupéré des données du satellite au cours de la mission.

Au total, ce sont plus de 120000 paquets de données qui ont été récupérés par les amateurs de par le monde, au cours des 10 semaines de fonctionnement de la mission.

D.2.9 Lancement, opérations, et résultats

Le 12 Janvier 2018, à 03 :59 UTC, la fusée PSLV-C40 décolle du centre spatial Satish Dawan, sur la côte Est de l'Inde (Figure D.14). À son bord se trouvent la mission d'observation de la Terre Cartosat-2F, et un certain nombres de petit satellites, dont PicSat. Après 20 minutes de vol environ, PicSat est déployé en orbite, à 505 km d'altitude.

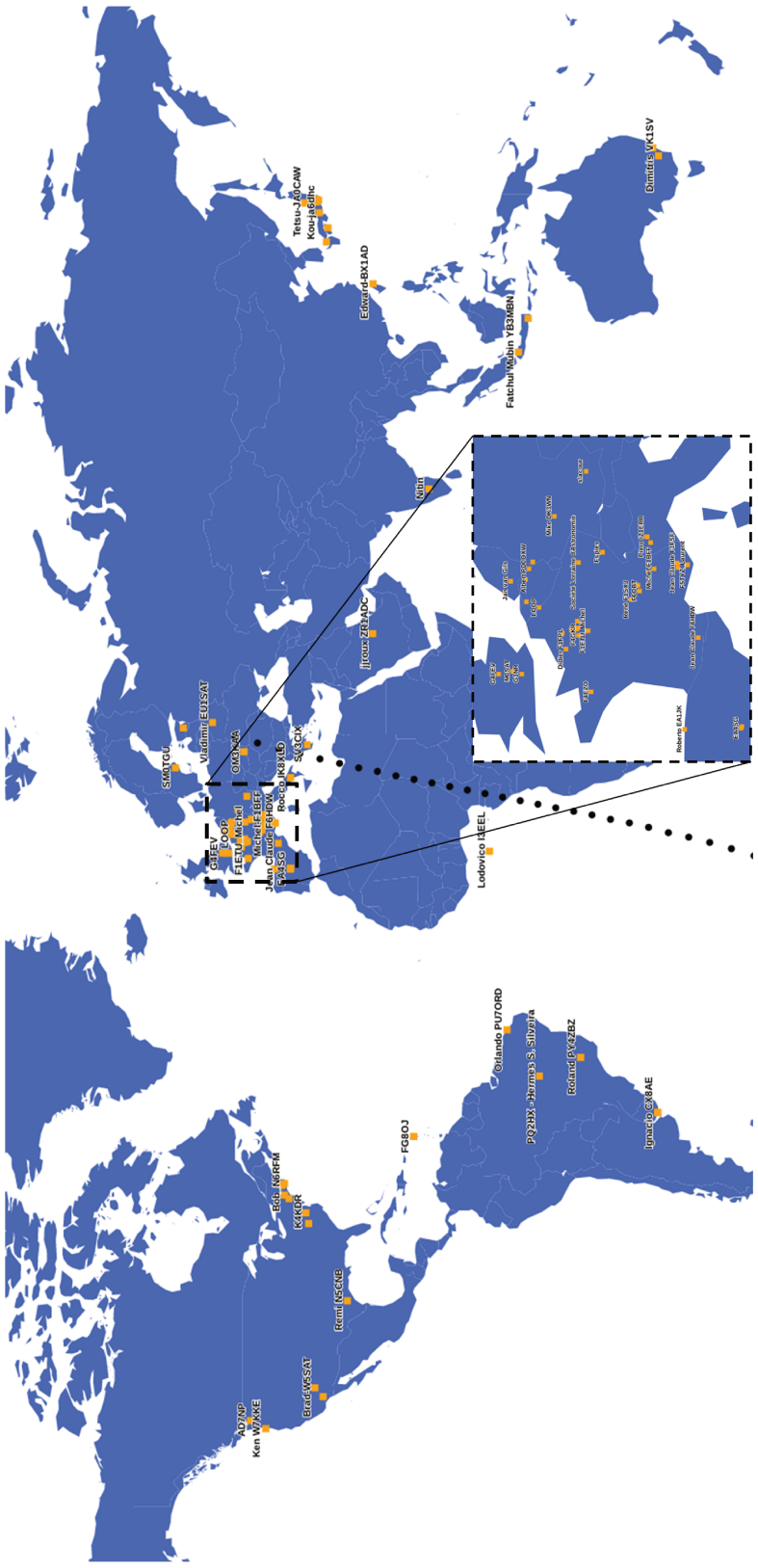


FIGURE D.13: Aperçu du réseau de stations amateures ayant rejoint le projet PicSat. Chaque point jaune représente une station amateur ayant écouté le satellite et transféré des données vers le centre de contrôle de Meudon au cours de la mission.

FIGURE D.14: Décollage de la fusée PSLV-C40 depuis le port spatial du Satish Dawan, sur la côté Est de l'Inde, le 12 Janvier 2018. Crédits : ISRO.

Le premier passage du satellite au dessus de la station sol de Meudon a lieu ce même jour, à 07 :45 UTC environ, et la balise radio[14] est détectée, confirmant l'allumage correct du satellite, et le déploiement des antennes. De nombreux amateurs confirment aussi très rapidement la bonne réception des balises PicSat, marquant le début de la phase d'opération en vol de la mission.

L'orbite polaire de PicSat le faisait passer au dessus de la station sol de Meudon deux fois le matin (entre 9h et midi), et deux fois le soir (entre 9h et minuit). Ces passages du satellite au-dessus de la station duraient au mieux 10 ou 15 minutes, et constituaient les seules fenêtres d'envoi de télécommandes au satellite. Force est de constater que les débuts de la phase d'opération ont été laborieux. Les outils logiciels du segment sol avaient été développés en utilisant le lien radio entre l'antenne du segment sol et le modèle d'ingénierie du satellite, situé au sol, dans une salle du LESIA. Le lien radio était alors de très bonne qualité, et l'ensemble des outils fonctionnaient correctement. En opération, toutefois, le lien radio s'est trouvé largement détérioré par des erreurs sur le calcul d'orbite, des erreurs de positionnement de l'antenne, ou plus simplement par la présence d'arbres autour de la station sol. Il aura fallu une dizaine de jours environ pour remettre en ordre les outils logiciels, et permettre leur fonctionnement correct en présence de ces différents problèmes.

Mais les problèmes de communication radio et de logiciel sol n'ont certainement pas été le souci principal de la phase d'opération en vol de PicSat. L'une des tâches critiques qu'il était nécessaire d'effectuer après le déploiement du satellite en orbite était de stopper la rotation du satellite, et de l'orienter vers l'étoile. Or, le système de contrôle d'attitude fourni par Hyperion

[14] L'ordinateur de bord de PicSat était programmé pour émettre automatiquement un paquet contenant diverses informations sur l'état du satellite, toutes les 10 secondes. Ce paquet, dit « balise » était utilisé pour détecter le début et la fin des passages de PicSat dans le ciel de Meudon, et pour vérifier le statut général du satellite.

FIGURE D.15: L'impressionnante antenne de 25 m de la station amateure de Dwingeloo (indicatif PI9CAM), utilisée pour tenter de rétablir la communication avec PicSat.

Technologies s'est avérer défectueux, et il n'a pas été possible d'orienter le satellite vers sa cible scientifique.

Après 8 semaines passées à essayer de comprendre l'origine du problème pour pouvoir le corriger, le satellite a soudain cessé d'émettre. C'est un radio-amateur situé aux États-Unis, W2RTV, qui le premier a indiqué, le 20 Mars 2018, ne plus recevoir les balises PicSat. Le soir de ce 20 Mars 2018, lors des passages du satellite au dessus de Meudon, aucun paquet n'a pu être reçu, confirmant l'anomalie. Toutes les tentatives faites pour rétablir le contact sont restées infructueuses.

Les problèmes pouvant potentiellement expliquer une perte complète du lien radio sont nombreux : problème d'alimentation en énergie, problème avec la carte radio, bug logiciel qui empêche l'ordinateur de bord de redémarrer correctement, etc. L'un des principaux suspects était initialement l'un des amplificateurs de puissance utilisé sur la carte de transmission radio du satellite, qui avait posé quelques problèmes au sol. Un amplificateur défectueux pouvait potentiellement expliquer une perte de puissance radio, et donc l'impossibilité de recevoir des données du satellite. Un appel à la communauté radio-amateure à permis de mobiliser l'antenne de 25 m de Dwingeloo, aux Pays-Bas (voir Figure D.15) pour écouter PicSat, malheureusement sans plus de succès. Même en l'absence d'amplification, une telle antenne aurait dû permettre de recevoir des données, ce qui de fait éliminait cette cause potentielle de panne.

Après une étude approfondie, il s'avère que le satellite a cessé d'émettre le 20 Mars 2018, entre 13 :20 et 13 :50 UTC, alors qu'il volait au dessus de pôle Nord. La cause exact de cette panne reste inconnue, quoi que les soupçons s'orientent plutôt vers une panne logiciel ou un dommage causé par les radiations.

Sans possibilité d'orienter le satellite vers sa cible scientifique durant les 10 semaines de fonctionnement avant la panne radio, il n'a pas non plus été possible d'injecter de la

lumière dans la fibre optique de l'instrument. La photodiode de la mission PicSat n'a donc malheureusement jamais vu le moindre photon en provenance de β Pic.

D.3 Observation de la ré-émergence de beta Pictoris b avec VLTI/GRAVITY

D.3.1 L'interférométrie et les exoplanètes

L'interférométrie est une technique d'observation consistant à recombiner la lumière obtenue par plusieurs télescopes. Il est malheureusement difficile d'appréhender pleinement cette méthode sans avoir recourt à un peu de mathématiques. Une explication détaillée, ou au moins une tentative d'explication détaillée, est proposée en Section D.3.2. Pour l'ores, l'objectif est de situer cette méthode dans le contexte plus général de l'observation des exoplanètes.

Commençons par un exemple. En 2007, Baines et ses collègues rapportent une mesure qui peut paraître intrigante : la mesure du rayon de la planète HD 189733 b, orbitant autour de l'étoile HD 189733. D'après leur étude, le rayon de cette planète est de 1.19 fois le rayon de Jupiter, à 0.08 près. Sachant que l'étoile en question est située à environ 20 parsecs de notre système Solaire (soit environ 65 années-lumière), le diamètre angulaire apparent de la planète se calcule aisément :

$$\theta_{\text{HD 189733 b}} = \frac{1.19\,R_{\text{Jup}}}{65\,\text{a-l}} \simeq \frac{8\times10^{4}\,\text{km}}{6\times10^{14}\,\text{km}} \simeq 1.3\times10^{-10}\,\text{rad} \tag{D.10}$$

Ce qui correspond à environ 3 micro-secondes d'arc.

En comparaison, un télescope de diamètre D, observant à une longueur d'onde λ, est limité à une résolution angulaire θ donnée par le critère de Rayleigh :

$$\theta_{\text{Rayleigh}} = \frac{1.22\times\lambda}{D} \tag{D.11}$$

Même en supposant une observation dans le visible, à 500 nm par exemple, il faudrait un télescope de plusieurs kilomètres de diamètre pour parvenir à résoudre[15] un objet aussi petit ! Comme aucun télescope de cette taille n'existe, la question se pose de savoir comment une telle mesure a pu être obtenue.

En réalité, l'étude en question procède en temps. Pour commencer, la taille de la planète par rapport à son étoile est obtenue via l'observation d'un transit photométrique : en mesurant la baisse de luminosité de l'étoile au moment du passage de la planète devant elle, on peut en déduire le rapport du rayon de la planète par rapport au rayon de l'étoile[16]. Une fois ce ratio obtenu, il est alors possible de déterminer le rayon de la planète en mesurant celui de l'étoile. Comme une étoile est beaucoup plus grosse qu'une planète, c'est évidemment beaucoup plus facile. Toutefois, à nouveau, en tenant compte du fait qu'une étoile est typiquement d'un rayon 10 fois supérieur à celui d'une planète comme Jupiter, le critère de Rayleigh indique qu'il est toujours nécessaire d'avoir à disposition un télescope d'au moins une centaine de mètres de

[15] Les objets astronomiques observés par un télescope apparaissent comme des petites tâches sur les détecteurs. À priori, plus un objet est petit, plus son image elle-même est petite. En pratique, toutefois, du fait de la diffraction de la lumière, pour un télescope de diamètre donné, il existe une taille minimale de la tâche image, qui correspond au critère de Rayleigh. Tous les objets plus petits que ce critère de Rayleigh apparaissent donc d'une taille identique sur la caméra, indépendamment de leur taille physique réelle. En conséquence, on ne peut pas mesurer la taille de ces objets trop petits à partir des images obtenues. On parle d'objets « non-résolus » par le télescope. À contrario, si l'objet est assez grand, il est « résolu », et on peut mesurer sa taille à partir de son image.

[16] Oui, c'est tout à fait ce que nous avons essayé de faire avec PicSat, sans succès.

diamètre pour pouvoir réaliser une telle mesure. Un tel télescope n'existe pas, puisque les plus grands à l'heure actuelle sont de l'ordre de 10 m de diamètre.

La solution est alors de *recombiner* la lumière provenant de plusieurs télescopes plus petits, mais espacés les uns des autres, au sein d'un interféromètre. Pour deux télescopes de diamètre D espacés l'un de l'autre d'une distance B, le critère de Rayleigh, donnant la résolution maximale des télescopes pris individuellement est toujours $\theta_{\text{Rayleigh}} = 1.22\lambda/D$, mais la résolution « interférométrique », elle, est de l'ordre de $\theta_{\text{interf}} = \lambda/(2B)$. Il « suffit » donc d'avoir à disposition un réseau de télescopes, potentiellement tous relativement petits, mais espacés de quelques centaines de mètres les uns des autres, et il devient possible de mesurer directement la taille d'un étoile comme HD 189733. C'est ce que l'équipe menée par Baines a fait en 2007, en utilisant le réseau de télescopes CHARA, situé en Californie, avec ses six télescopes de 1 m répandus sur plus de 300 m.

En plus de ce rôle de l'interférométrie comme moyen de mesurer les rayons d'étoiles abritant des planètes, on pourrait aussi citer les différents programmes visant à mieux comprendre les structures les plus internes des systèmes exoplanétaires, et notamment le rôle et l'origine de la poussière zodiacale. Ce sont là deux contributions importantes faites par l'interférométrie à la recherche sur les exoplanètes. Toutefois, force est d'admettre qu'au regard du nombres de projets qui ont été proposés pour tenter de détecter directement des planètes à partir de techniques interférométriques, la récolte reste bien maigre. La raison en est que la détection directe de planètes orbitant autour de leurs étoiles pose deux problèmes différents : un problème de *résolution angulaire*, puisqu'il s'agit de détecter une planète très proche de son étoile ; et un problème de *contraste*, puisqu'il s'agit aussi de détecter un objet très faiblement lumineux autour d'un objet extrêmement lumineux. En pratique, si les interféromètres excellent en termes de résolution angulaire, ils pêchent usuellement par leur manque de contraste.

Un certain nombre de méthodes ont été proposées pour pallier à ce problème, comme par exemple l'observation dite « en frange sombre[17] », mais jusqu'à présent, les performances n'ont pas été suffisantes pour observer des exoplanètes.

La situation a récemment changé, avec l'arrivée sur le ciel de l'instrument GRAVITY, installé sur le réseau de l'interféromètre du Très Grand Télescope (Very Large Télescope, ou VLT, en anglais) au Chili. J'ai eu la chance d'être impliqué, au cours de ma thèse, dans les deux premières détections interférométriques d'exoplanètes. La première, HR 8799 e, a été publiée en Mars 2019 (Gravity Collaboration, 2019). La deuxième, β Pictoris b, fait l'objet des sections ci-après.

D.3.2 L'instrument GRAVITY sur le VLTI

D.3.2.1 Lumière et interférences

Ondes lumineuses Il n'est pas possible de résumer et d'expliquer les observations réalisées durant cette deuxième partie de ma thèse sans s'arrêter d'abord sur le concept d'interférométrie optique. Or, pour comprendre l'interférométrie, il faut avoir en tête que ce que nous appelons « lumière » n'est rien d'autre qu'une onde électromagnétique, autrement dit des oscillations du champ électrique (ou, c'est équivalent, du champ magnétique), qui se propagent dans l'espace. Dans une assez large mesure, la lumière est finalement très semblable au son, à la différence près que le son consiste en des oscillations de pression, qui nécessitent donc un milieu matériel pour se propager. La lumière, elle, se propage aussi dans le vide.

[17] Encore connue sous le nom « d'interférométrie annulante », qui consiste à faire en sorte que la lumière de l'étoile reçue par les différents télescopes du réseau soit toujours recombinée avec un déphasage de π, afin de s'assurer que les interférences soient destructives

Puisque la lumière est une onde, elle peut être décrite par quelques propriétés importantes : l'« amplitude » des oscillations, leur « fréquence », ou, de manière équivalente, la « longueur d'onde », et la « vitesse de propagation » de l'onde. Comme dans le cas des ondes sonores, nous ne sommes pas capables de percevoir les oscillations elles-mêmes. Nous percevons simplement une « intensité », liée à l'amplitude des oscillations, ainsi qu'une « couleur », liée à la longueur d'onde. Cette couleur est à la lumière ce que la « hauteur » (aigu ou grave) est au son.

Mathématiquement, une onde lumineuse peut donc se représenter par une sinusoïde se propageant dans l'espace. Le champ électrique E peut s'écrire :

$$E(x,t) = A\sin\left(\frac{2\pi}{\lambda}(x-ct)\right) \tag{D.12}$$

où A est l'amplitude de l'onde, λ sa longueur d'onde, x la distance entre la source de l'onde et le point de l'espace auquel on s'intéresse, t le temps, et c la vitesse de la lumière.

Le terme $(x-ct)$ à l'intérieur de la sinusoïde montre que l'on peut voir l'onde lumineuse comme des oscillations dans l'espace, à temps t fixé, ou comme des oscillations dans le temps, à position x fixée.

Interférométrie optique Considérons maintenant deux télescopes T_1 et T_2, observant une étoile donnée. Les deux télescopes sont à des positions différentes, donc à des distances différentes de l'étoile. En conséquence, l'onde lumineuse reçue par le télescope T_2 est légèrement « décalée » par rapport à l'onde reçue par le télescope T_1. On parle de « déphasage ». On peut écrire :

$$E_{T_1}(t) = A\sin\left(\frac{2\pi}{\lambda}(x_1-ct)\right) \tag{D.13}$$

$$E_{T_2}(t) = A\sin\left(\frac{2\pi}{\lambda}(x_2-ct)\right) \tag{D.14}$$

avec x_1 et x_2 la distance des télescopes T_1 et T_2 à l'étoile. Le déphase $\Delta\phi$ entre les deux télescopes est alors donné par la différence entre les phases des deux sinusoïdes :

$$\begin{aligned}\Delta\phi &= \frac{2\pi}{\lambda}(x_2-ct) - \frac{2\pi}{\lambda}(x_1-ct)\\ &= \frac{2\pi}{\lambda}(x_2-x_1)\end{aligned}$$

La situation décrite par ces équations est représentée en Figure D.16. Notons, c'est important, que le déphasage $\Delta\phi$ dépend non seulement de la différence de distance entre les télescopes et l'étoile source, mais aussi de la longueur d'onde λ.

Supposons maintenant qu'il soit possible de « recombiner » les deux faisceaux lumineux, c'est-à-dire de combiner les deux champs électriques, pour obtenir leur somme. On obtient alors le champ total :

$$E_{\text{total}}(t) = E_{T_1}(t) + E_{T_2}(t) \tag{D.15}$$

$$= A\sin\left(\frac{2\pi}{\lambda}(x_1-ct)\right) + A\sin\left(\frac{2\pi}{\lambda}(x_2-ct)\right) \tag{D.16}$$

On peut montrer, avec un peu de calcul trigonométrique, que la somme de deux sinusoïdes de même fréquence est une sinusoïde, dont l'amplitude dépend du déphasage (voir aussi la représentation graphique de cette somme, en Figure D.16). On obtient :

$$E_{\text{total}}(t) = 2A\sin(\Delta\phi/2)\cos\left(\frac{2\pi}{\lambda}(x_1+x_2-2ct)\right) \tag{D.17}$$

Ce que dit cette Équation D.17, c'est que le champ électrique recombiné E_{total} est lui-même une sinusoïde, dont l'amplitude est donnée par $|2A\sin(\Delta\phi/2)|$. Cela signifie qu'en sommant les faisceaux en provenance des deux télescopes, les deux ondes interfèrent, et on obtient un signal lumineux dont l'amplitude dépend du déphasage. Si les ondes reçues par les télescopes sont en phase ($\Delta\phi = 0$), les interférences sont « constructives », et l'amplitude du signal recombiné est doublée (puisque $\sin(\Delta\phi/2) = 1$, dans ce cas). Et, aussi surprenant que cela puisse paraître, si les ondes reçues sont en opposition de phase ($\Delta\phi = \pi$), alors les interférences sont destructives, et l'amplitude du signal recombiné est nulle ($\sin(\Delta\phi/2) = 0$). Il n'y a plus de lumière ! Bien sûr, tous les cas intermédiaires sont possibles. Et en pratique, puisque le déphasage dépend de la longueur d'onde (donc de le couleur, on parle de déphasage « chromatique »), les interférences sont constructives dans certaines couleurs, et destructives dans d'autres (voir, encore une fois, la Figure D.16).

Exemple d'application astronomique Les deux paragraphes précédents décrivent le phénomène d'interférence entre la lumière en provenance de deux télescopes. Toutefois, il reste une question d'importance qu'il s'agit maintenant d'éclairer : à quoi ça sert ?

Pour comprendre l'intérêt de l'interférométrie pour l'astronomie, on peut repartir de l'Équation D.17, qui indique que l'amplitude du signal recombiné est donnée par :

$$A_{\text{total}} = 2A\sin(\Delta\phi/2) \tag{D.18}$$

$$= 2A\sin\left(\frac{\pi}{\lambda}(x_2 - x_1)\right) \tag{D.19}$$

où la deuxième ligne est obtenue en explicitant la valeur du déphasage (voir Équation D.15).

Mesurer l'intensité d'un signal lumineux en fonction de la longueur d'onde est une chose relativement facile en astronomie. On parle de spectroscopie, et il s'agit d'une technique vieille de plus de deux siècles. On peut donc facilement mesurer cette amplitude A_{total} en fonction de λ, et obtenir une courbe comme celle donnée en Figure D.16. Or, les oscillations sur cette courbe dépendent directement de la valeur du terme $(x_2 - x_1)$, dans l'équation ci-dessus. On peut donc, à partir du signal interférométrique, mesurer $x_2 - x_1$.

Le lecteur ayant une excellente vision géométrique aura peut-être déjà perçu la voie dans laquelle tout ça nous mène. Prenons deux télescopes T_1 et T_2, situés a deux positions différentes, et observant une étoile située dans une direction angulaire α, par rapport au zénith. Les termes x_1 et x_2 représentent la distance entre les télescopes T_1, T_2 et l'étoile. Nous allons les calculer. Pour ce faire, notons B la distance entre les télescopes, et prenons l'origine géométrique au milieu de ces deux télescopes. Les deux télescopes sont à une distance $B/2$, de part et d'autre de O. L'étoile S est à une distance D de l'origine O, dans une direction angulaire α. La situation est illustrée en Figure D.17.

Pour calculer la distance $x_1 = ST_1$, on peut se placer dans le triangle T_1OS. Dans ce triangle, deux distances sont connues ($SO = D$, et $T_1O = B/2$), ainsi qu'un angle ($T_1OS = \pi/2 - \alpha$, il s'agit d'un angle complémentaire). Pour qui se rappelle de ses cours de géométrie, deux distances et un angle suffisent à fixer un triangle, dont on peut calculer toutes les autres caractéristiques par les formules d'Al-Kashi. Pour la quantité qui nous intéresse, on obtient :

$$x_1 = \sqrt{D^2 + \frac{B^2}{4} - BD\cos\left(\frac{\pi}{2} - \alpha\right)} \tag{D.20}$$

$$= D \times \sqrt{1 + \frac{B^2}{4D^2} - \frac{B}{D}\sin(\alpha)} \tag{D.21}$$

La distance D étant largement supérieure à la distance entre les deux télescopes B, on peut

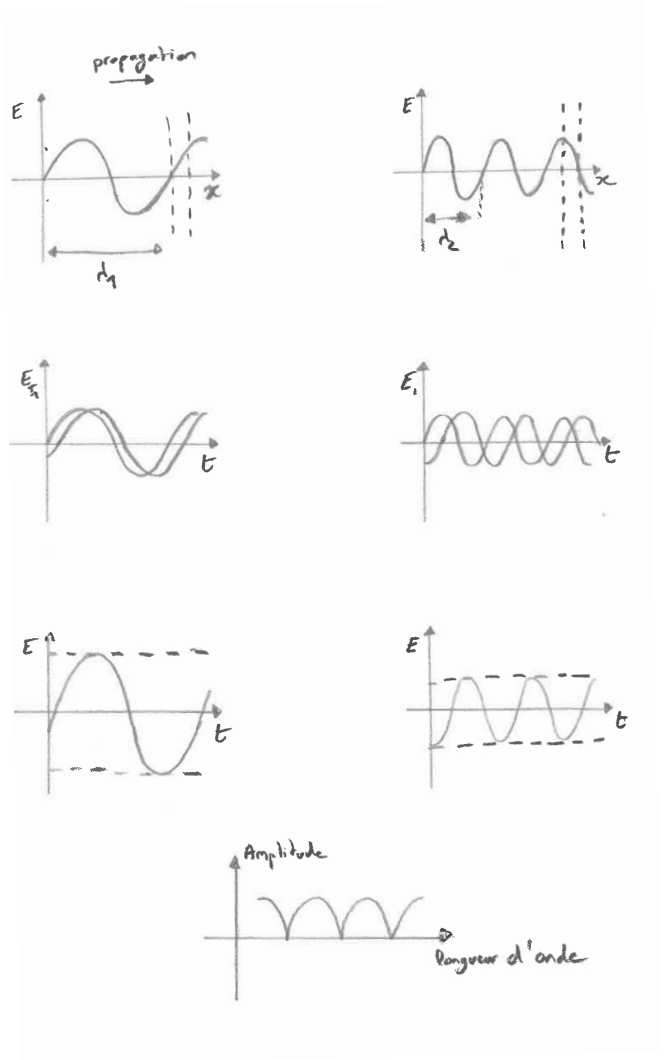


FIGURE D.16: Représentation schématique du phénomène d'interférences. La première ligne donne le champ électrique en fonction de la position, à un instant t fixé, dans le cas d'une onde lumineuse de longueur d'onde λ_1 (à gauche), ou $\lambda_2 < \lambda_1$ (à droite). Cette première ligne est une sorte de « photographie instantannée » du champ électrique. Si l'on pouvait représenter l'évolution temporelle sur une feuille de papier, on verrait l'onde « avancer » vers la droite.
La deuxième ligne montre l'évolution au cours du temps du champ électrique à deux positions fixées, symbolisées par les deux lignes pointillées verticales les plus à droite sur les figures de la première ligne. On constate que, selon la longueur d'onde, pour deux positions identiques, le déphasage entre les deux ondes est différentes (les deux ondes sur le graphique de gauche sont presque alignées, alors qu'elles sont nettement décalées à droite).
La troisième ligne donne la somme des deux ondes de la deuxième ligne, ce qui correspond au champ recombiné tel que mesuré par un interféromètre. On constate que plus les deux ondes sont « correctement alignées », plus l'amplitude du signal interférométrique est importante. Et comme cet « alignement » dépend de la longueur d'onde, l'amplitude du signal interférométrique dépend aussi de la longueur d'onde. C'est ce que représente la dernière figure, qui donne l'allure de l'évolution de l'amplitude interférométrique en fonction de la longueur d'onde. C'est ce signal que mesure un interféromètre.

simplifier cette équation en utilisation une approximation au premier ordre :

$$x_1 \simeq D \times \left(1 - \frac{B}{2D}\sin(\alpha)\right) \tag{D.22}$$

Pour la distance x_2, on peut raisonner dans le triangle SOT_2, pour lequel on connaît aussi deux longueurs ($SO = D$, et $OT_2 = B/2$), ainsi qu'un angle ($SOT_2 = \pi/2 + \alpha$). Les mêmes calculs donnent cette fois :

$$x_2 \simeq D \times \left(1 + \frac{B}{2D}\sin(\alpha)\right) \tag{D.23}$$

On combinant ces deux valeurs, on obtient alors une formule particulièrement élégante :

$$x_2 - x_1 = B\sin(\alpha) \tag{D.24}$$

Cette Équation D.24 montre que la quantité $x_2 - x_1$, que l'on peut mesurer par spectroscopie sur le signal interférométrique, est directement liée à la distance entre les deux télescopes (que l'on connaît, puisqu'on peut la mesurer), et à la position angulaire de l'étoile. On peut donc déterminer la position de l'étoile à partir de la recombinaison interférométrique des signaux issus des deux télescopes.

L'interférométrie peut donc notamment servir à réaliser des mesures astrométriques, c'est-à-dire des mesures de position des étoiles. L'avantage de cette méthode est sa très grande sensibilité : on peut en pratique mesurer des différence $x_2 - x_1$ bien inférieure à la longueur d'onde λ, soit typiquement quelques nanomètres. Comme on peut en plus utiliser des télescopes séparés d'une centaine de mètre, cela signifie qu'il est possible d'obtenir des précision astrométriques de l'ordre d'une dizaine de micro-arcsecondes. Nous sommes ici dans le régime de la « très haute résolution angulaire », ce qui explique pourquoi les interféromètres sont par exemple capables de mesurer des diamètres d'étoiles (voir Section D.3.1)

D.3.2.2 L'interféromètre du VLT, et l'instrument GRAVITY

Le Très Grand Télescope (Very Large Telescope, VLT) est l'une des installations de l'Observatoire austral européen (European Southern Observatory, ESO). Situé au Chili, sur le plateau désertique de l'Atacama, il est composé de quatre télescopes unitaires (Unit Telescope, UT) de 8.2 m de diamètre, et de quatre télescopes auxilaires (Auxiliary Telescopes, AT) de 1.8 m de diamètre. Les télescopes sont chacun équipés de plusieurs instruments, permettant toutes sortes d'observations. Mais le VLT dispose aussi d'un mode interférométrique (le VLTI), permettant la recombinaison des faisceaux issus des quatre UTs, ou des quatre ATs. Les UTs sont fixes, mais les ATs peuvent être déplacés, afin d'optimiser leurs positions pour les observations interférométriques. La Figure D.18 donne un plan du VLT.

Le VLTI lui-même dispose de plusieurs instruments permettant différents types d'observations interférométriques. L'instrument GRAVITY, installé au VLTI depuis 2015, est un instrument de deuxième génération, opérant en bande K (dans les longueurs d'ondes de l'infrarouge proche, entre 1.9 et 2.4 microns). Les observations scientifiques avec cet instrument ont commencé en 2016 sur les ATs, et en 2017 sur les UTs.

D.3.3 Comment observer des planètes avec GRAVITY ?

Via la recombinaison des faisceaux lumineux issus de plusieurs télescopes, l'interférométrie permet de faire des observations à haute résolution angulaire, sensibles à la position dans le ciel des astres observés. C'est cette particularité qui rend aussi l'interférométrie utile pour l'observation des planètes.

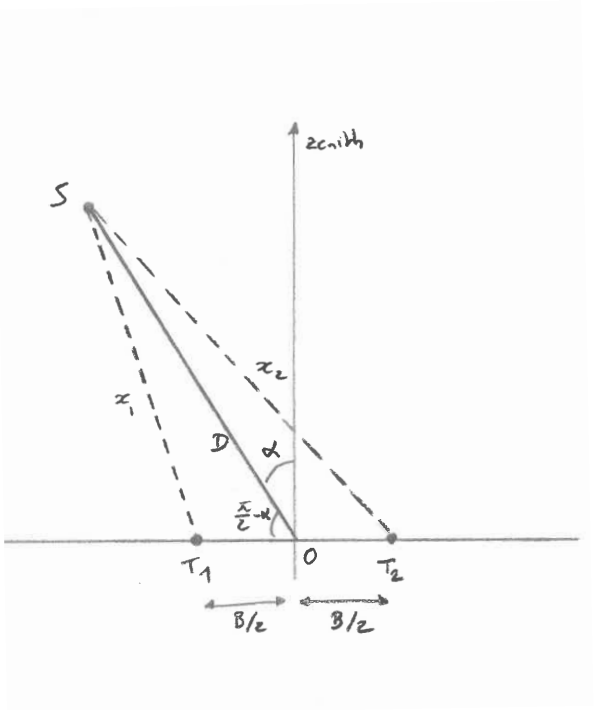


FIGURE D.17: Schéma représentant le situation géométrique d'un interféromère à deux télescopes (T_1 et T_2 observant une étoile située à une position S, dans une direction α.

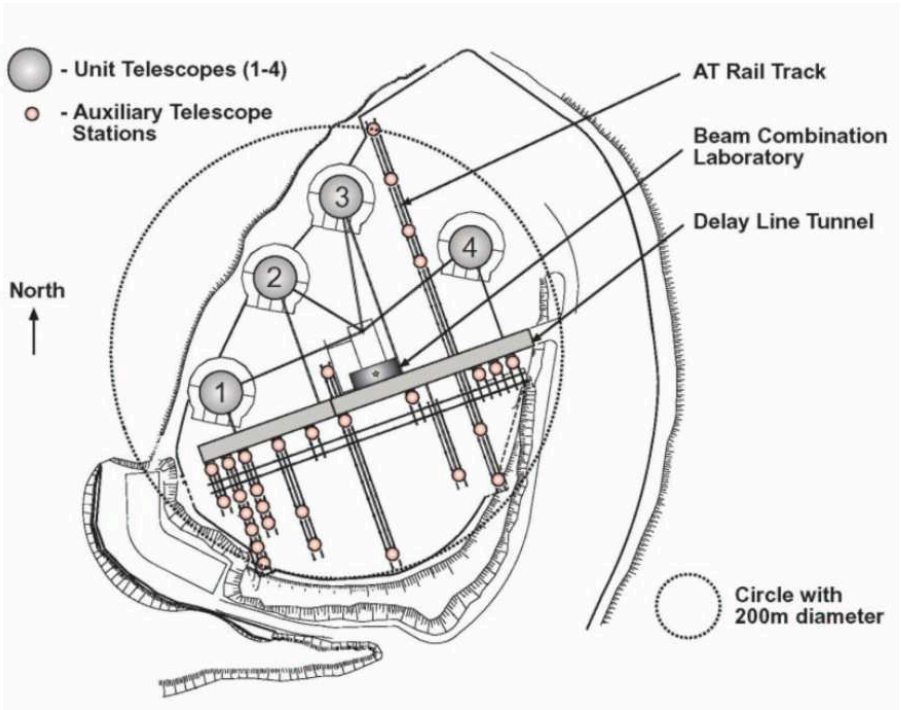


FIGURE D.18: Plan des infrastructures du VLT, au Chili. Les ronds numérotés représentent les télescopes unitaires de 8.2 m de diamètre. Les petits ronds représentent les positions possibles des télescopes auxiliaires de 1.8 m (qui peuvent être déplacés).

Pour comprendre l'intérêt d'un instrument comme GRAVITY pour l'observation directe des planètes, il est important d'avoir en tête que la difficulté principale de telles observations est de parvenir à distinguer la lumière de la planète de celle de l'étoile. Même dans les cas les plus favorables, une jeune planète géante est typiquement 100 000 fois moins brillante que son étoile. En conséquence, lorsque l'on pointe un télescope vers un tel système, la lumière de l'étoile domine très largement les images obtenues, et il est difficile de parvenir à distinguer une planète – faible point lumineux – dans ce halo brillant.

Toutefois, dans le cas d'un instrument interférométrique, si l'on reprend l'Équation D.19, on peut se rendre compte que le signal mesuré sur une planète est fondamentalement différent de celui mesuré sur une étoile. Les amplitudes interférométriques de la lumière de l'étoile et de celle de la planète, dispersées en longueur d'onde, sont données par :

$$A_{\text{total, étoile}} = 2A_{\text{étoile}} \sin\left(\frac{\pi}{\lambda} B\alpha_{\text{étoile}}\right) \tag{D.25}$$

$$A_{\text{total, planète}} = 2A_{\text{planète}} \sin\left(\frac{\pi}{\lambda} B\alpha_{\text{planète}}\right) \tag{D.26}$$

où l'Équation D.24 a aussi été utilisée pour expliciter la quantité « $x_2 - x_1$ » apparaissant dans l'Équation D.19.

Dans cette équation, les termes $A_{\text{étoile}}$ et $A_{\text{planète}}$ représentent l'intensité lumineuse de l'étoile et la planète[18]. En réalité, ces deux quantités dépendent de la longueur d'onde λ. C'est d'ailleurs cette dépendance à la longueur d'onde qui représente la mesure la plus intéressante, puisqu'elle trahit la température et la composition de ces astres (voir Section D.3.4).

L'instrument GRAVITY utilise une méthode de recombinaison un peu spécifique, qui permet de faire des observations « par rapport à l'étoile ». Autrement dit, les observations sont faites comme si l'étoile était la position de référence, à $\alpha = 0$. En réécrivant l'équation précédente avec la dépendance explicite en λ, et en adoptant l'étoile comme référence, on obtient :

$$A_{\text{total, étoile}}(\lambda) = 2A_{\text{étoile}}(\lambda) \tag{D.27}$$

$$A_{\text{total, planète}}(\lambda) = 2A_{\text{planète}}(\lambda) \sin\left(\frac{\pi}{\lambda} B\Delta\alpha\right) \tag{D.28}$$

où $\Delta\alpha = \alpha_{\text{planète}} - \alpha_{\text{étoile}}$ est la position de la planète par rapport à l'étoile.

Lorsque l'on réalise une observation, la planète étant très proche de son étoile, on obtient toujours un mélange des deux signaux :

$$A_{\text{observée}}(\lambda) = 2A_{\text{étoile}}(\lambda) + 2A_{\text{planète}}(\lambda) \sin\left(\frac{\pi}{\lambda} B\Delta\alpha\right) \tag{D.29}$$

C'est cette dernière équation qui fait apparaître clairement tout l'intérêt de l'interférométrie pour l'observation des exoplanètes : contrairement au cas de l'imagerie « classique », avec laquelle il est difficile de séparer la lumière de la planète de celle de l'étoile, l'interférométrie apporte une *modulation* naturelle du signal de la planète (le terme en « sinus »), qui permet de le séparer. On observe dans les données un signal « relativement plat », correspondant au

[18] En fait, à strictement parler, ces termes représentent l'amplitude de l'onde électromagnétique, qui est une quantité différente de l'intensité. Le problème est que pour établir le rapport exact entre l'amplitude et l'intensité, il est nécessaire d'introduire aussi la composante « magnétique » de l'onde lumineuse, dont je n'ai pas du tout parlé. L'intensité est alors la somme du carré de l'amplitude du champ électrique et du carré de l'amplitude du champ magnétique. En pratique, dans les équations, cela prend la forme de nombres complexes, dont on utilise la partie réelle pour réprésenter la composante électrique, et la partie imaginaire pour représenter la partie magnétique. Par soucis de simplification, tout ceci est ici brutalement ignoré, et je fais comme si « amplitude » et « intensité » étaient identiques. Cela ne change finalement pas grand chose à la discussion.

spectre de l'étoile, auquel se superpose un signal oscillant, qui correspond à la planète. Il est alors possible de séparer les deux, et d'extraire le spectre de la planète (le terme $A_{\text{planète}}$).

Évidemment, les données réelles contiennent leur lot de problèmes variés : l'atmosphère de la Terre déforme le spectre obtenu, la rotation du ciel fait varier les positions angulaires α au cours de la nuit, le système de métrologie instrumentale assurant la référence de phase n'est pas parfaitement stable, etc. Il n'est donc pas si simple d'extraire le signal de la planète des données obtenues. Tant mieux, cela donne du travail aux étudiants en thèse !

D.3.4 La planète géante beta Pictoris b vue par GRAVITY

D.3.4.1 Observations

Les observations de la planète géante β Pictoris b présentées dans la suite de cette Section ont été obtenues le 22 Septembre 2018, avec l'instrument GRAVITY sur l'interféromètre du VLTI (voir Sections précédentes). Au total, environ 90 min d'observation avec les quatre télescopes de 8 m du VLT auront été nécessaires. L'objectif de ce programme d'observation était d'obtenir une mesure précise de la position de la planète par rapport à son étoile, ainsi qu'un spectre de la lumière de la planète.

D.3.4.2 Spectre de beta Pictoris b et modèles atmosphériques

Il est donc possible, grâce à le modulation du signal interférométrique (voir Équation D.29), d'extraire un spectre de β Pictoris b à partir des observations GRAVITY. En réalité, les observations effectuée donne plutôt un spectre de *contraste* de la planète par rapport à l'étoile, autrement dit le rapport $A_{\text{planète}}(\lambda)/A_{\text{étoile}}(\lambda)$, qu'il est nécessaire de multiplier par un modèle de l'étoile afin de récupérer le spectre de la planète elle-même. Mais peu importe, le processus général de réduction de ces données interférométrique est long, laborieux, et je ne l'aborderai pas dans ce résumé. Le résultat, en revanche, est bien plus intéressant, et est présenté en Figure D.19.

Comparé aux résultats précédemment obtenus par d'autres équipes, avec un instrument d'imagerie directe, le spectre GRAVITY est mieux résolu, et globalement de meilleure qualité. À l'oeil, on distingue déjà quelques éléments intéressants : la bande d'absorption de l'eau autour de 2 microns est bien visible, tout comme les bandes du monoxyde de carbone, autour de 2.3 microns. Autour de 2.25 microns, la bande d'absorption du méthane reste peu, voire pas du tout, marquée. La seule analyse visuelle de ce spectre indique donc déjà la présence d'eau et de monoxyde de carbone en grande quantité dans l'atmosphère de β Pictoris b.

Pour aller plus loin, il est nécessaire d'avoir recourt à d'autres méthodes que la simple analyse visuelle. En astronomie, comme dans bien d'autres disciplines, pour tenter de comprendre les observations effectuées, il est courant de procéder par « ajustement de modèles ». L'exercice consiste à utiliser une table de modèles, qui, à partir de simulations des processus physico-chimiques ayant lieu à l'intérieur des atmosphères planétaires, donne une série de spectres possibles, en fonction de divers paramètres. Par exemple, la grille Exo-REM, développée à l'Observatoire de Paris, et utilisée dans ce travail sur β Pic b, donne une série de spectres valables pour diverses températures et gravités de surface. En comparant les données réelles à tous ces spectres simulés, on peut déterminer quel est le modèle « le plus ressemblant », et donc quelles sont les valeurs de température et de gravité à la surface de la planète « les plus probables ». Dans le cas des données β Pictoris b, toutefois, on constate que même le modèle le plus ressemblant diffère finalement assez significativement des observations (voir la courbe verte en Figure D.20).

Cette situation où l'on constate que les modèles ne correspondent pas aux données est une situation particulièrement intéressante, quoique souvent délicate. Elle signifie qu'il y a

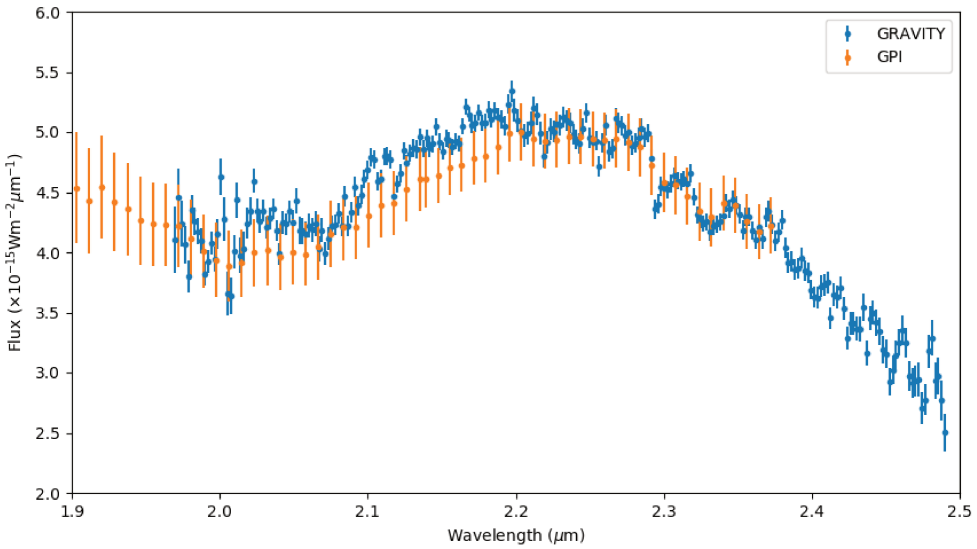


FIGURE D.19: Spectre de β Pictoris b en bande K (de 1.95 microns à 2.5 microns), obtenu à partir des observations VLTI/GRAVITY. Les points oranges (et les barres d'erreurs associées) représentent des données obtenues précédemment, par une autre équipe, avec l'instrument GPI.

quelque chose « qui cloche », ou bien dans les modèles eux-mêmes, ou bien dans la réduction des données. Sans plus de suspens, dans le cas de ces observations de β Pictoris b, la solution tient au fait que la grille Exo-REM utilisée initialement supposait un rapport C/O égal à celui du Soleil (soit environ 0.56). C'est usuellement une valeur raisonnable, mais, dans le cas précis de β Pic b, elle n'est pas correcte.

En adaptant Exo-REM pour pouvoir générer aussi des modèles de spectres avec des rapports C/O différents, on constate que l'ajustement des données est bien meilleur (voir courbe orange en Figure D.20). Dit autrement, la forme du spectre de β Pic b telle qu'observée donne une indication sur le rapport entre le nombre d'atomes de carbone et le nombre d'atomes d'oxygène dans l'atmosphère. Intuitivement, cela peut se comprendre, puisque la forme du spectre est dominée par la vapeur d'eau (H_2O), et le monoxyde de carbone (CO), deux des principaux porteurs d'atomes de carbone et d'oxygène. Pour un C/O de 0.56, comme supposé initialement dans la grille Exo-REM, il y a trop de carbone dans l'atmosphère, donc trop de CO, ce qui a tendance à créer une bande d'absorption trop profonde à 2.35 microns (courbe verte en Figure D.20). Si l'on réduit ce rapport à 0.16, on obtient un spectre beaucoup plus ressemblant (la courbe orange). Dans le jargon astronomique, on dit de façon inélégante que « les observations contraignent le rapport C/O ».

Une analyse mathématiques plus poussée permet d'extraire une estimation propre du rapport C/O, avec son incertitue associée. On obtient : $(C/O)_{\beta \text{ Pic b}} = 0.29^{+0.08}_{-0.07}$, ce qui constitue la première estimation du rapport C/O de β Pic b jamais obtenue.

D.3.4.3 La formation de beta Pictoris b

Il est remarquable que des observations comme celles présentées ici permettent de dire quelque chose du rapport entre la quantité de carbone et la quantité d'oxygène dans l'atmosphère d'une planète distante de plusieurs dizaines d'années-lumière. Mais il est sans doute encore

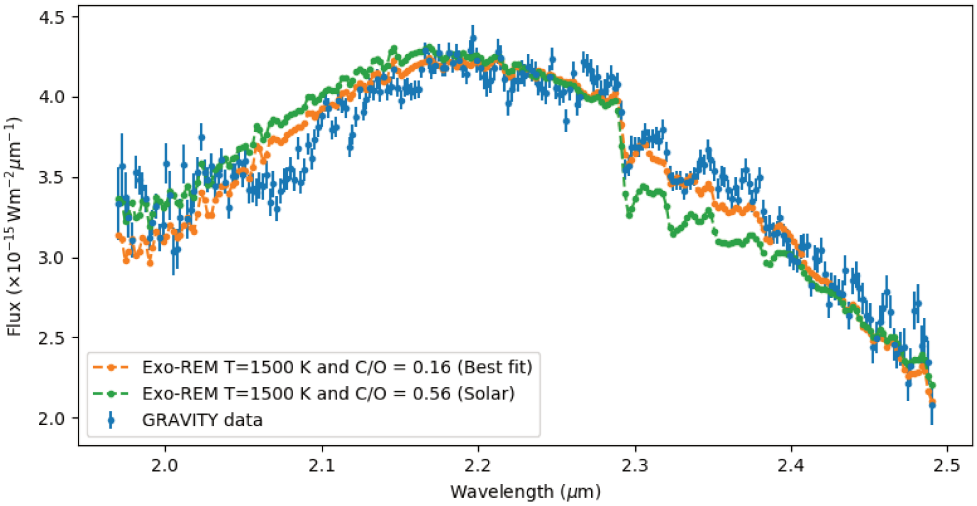


FIGURE D.20: Exemple de comparaison des données avec des modèles atmosphériques. La courbe bleue donne le spectre GRAVITY et ses barres d'erreurs. La courbe verte est un modèle Exo-REM, à une température de 1500 K et un rapport C/O de 0.56. La courbe orange est aussi un modèle Exo-REM à 1500 K, mais pour lequel le rapport C/O a été abaissé à 0.16. On remarque nettement que la région d'absorption du monoxyde de carbone (CO), autour de 2.3 microns, ressemble beaucoup plus au modèle à C/O = 0.16 qu'au modèle à C/O = 0.56.

plus étonnant de constater que cette mesure nous dit potentiellement quelque chose de ce qu'il s'est passé durant la formation de la planète elle-même, il y a plusieurs millions d'années.

Le C/O et la formation des planètes Récemment, plusieurs travaux ont tenté d'établir un lien entre les rapports atomiques des atmosphères planétaires et les processus de formation de ces mêmes planètes. On peut notamment citer les travaux menés Öberg et ses collègues, et publiés pour une part dans un premier article en 2011. Ces travaux sont essentiels à la suite de la discussion sur β Pictoris b, et j'en propose une reformulation dans ce paragraphe.

Même si les processus de formation des planètes restent encore assez mal compris, on sait toutefois que les planètes se forment au sein des disques circumstellaires, résidus de la formation des étoiles elles-mêmes. Ces disques sont composés de gaz d'une part, et de poussières microscopiques d'autre part. Leur chimie est très peu connue, mais il est raisonnable de supposer que ces disques sont grosso-modo homogènes, et de compositions similaires à la composition des étoiles elles-mêmes, puisque les étoiles se forment à partir de ce matériel. On peut donc supposer que le rapport C/O[19] dans les disques est homogène, proche de 0.56. Mais les disques sont composés de deux phases : un phase gazeuse, et une phase solide (la poussière). Tout le raisonnement de Öberg se fonde sur la distinction entre ces deux phases. En effet, quoique le C/O global soit homogène dans le disque, cela ne signifie pas que le C/O du gaz et le C/O du solide soient tous deux homogènes. Au contraire, ce n'est très certainement pas le cas.

La température dans un disque orbitant autour d'une étoile décroît à mesure que l'on s'éloigne du centre, puisque le flux lumineux se fait plus faible. En conséquence, à partir

[19] Je prends l'exemple du rapport C/O puisque c'est celui qui m'intéresse ici, mais on pourrait tenir le même raisonnement avec n'importe quel autre rapport atomique.

d'une certaine distance, la température est suffisamment faible pour que le dioxyde de carbone (CO_2), par exemple, devienne solide. On appelle cette distance la *ligne des neiges* du CO_2. Or, le CO_2 est une molécule qui contient une part importante de l'oxygène du disque, et, selon qu'il se trouve sous forme gazeuse ou solide, ce n'est pas la même phase du disque qui va bénéficier de cet enrichissement en oxygène.

Prenons par exemple un disque composé d'un certain nombre de molécules d'eau (H_2O), de monoxyde de carbone (CO), et de dioxyde de carbone (CO_2), comme illustré en Figure D.21, dans une proportion telle que $C/O = 0.62$ (dans cet exemple). On considère une région du disque autour de la ligne des neiges du CO_2. En deçà de cette limite, le CO_2 est gazeux. On peut alors compter le nombre d'atomes de carbone et d'oxygène dans le gaz, et obtenir $C/O = 0.7$. Si l'on procède de même dans la poussière, on trouve $C/O = 0$, puisque seule l'eau est présente en phase solide. D'ores et déjà, on constate que ni le C/O du gaz, ni celui du solide ne sont identiques au C/O global. Mais poursuivons : si l'on franchit la lignes des neiges du CO_2, celui-ci devient solide. Dans ce cas, le C/O de la phase solide passe à 0.4, tendis que celui de la phase gazeuse passe à 1. En conclusion, non seulement le C/O du gaz et celui de la poussière ne sont pas égaux, et diffèrent tous deux du C/O global, mais en plus ils sont variables selon la position dans le disque.

Dans ce raisonnement, le CO_2 est simplement pris pour exemple, mais ne constitue pas une molécule différente des autres. On peut définir des lignes des neiges pour chacune des molécules du disque, et calculer, pour n'importe quelle position, quelles sont les molécules présentent dans la poussière, et quelles sont celles présentent dans le gaz. À chaque franchissement d'une ligne des neiges, les C/O du gaz et de la poussière changent tout deux, du fait du changement de phase de la modélcule associée à cette ligne des neiges.

Si maintenant une planète se forme dans cet environnement, la quantité de carbone et d'oxygène qu'elle va pouvoir récupérer dépendra directement de sa localisation, ainsi que des fractions de matériau qu'elle va récupérer dans la phase gazeuse et dans la phase solide. Si par exemple on observait, dans le disque schématisé en Figure D.21, une planète dépourvue de carbone ($C/O = 0$), on pourrait en déduire qu'elle s'est nécessairement formée à gauche, en deçà de la ligne des neiges du CO_2, et uniquement à partir de poussières solides. Si on observait une planète à $C/O = 0.7$, on en déduirait que la planète s'est formée ou bien à gauche, essentiellement avec du gaz, ou bien à droite, à partir d'un mélange de gaz et de solide. Il devient possible, à partir d'une mesure du rapport C/O, de déduire quelque chose de la façon dont la planète s'est formée.

Un modèle de disque pour beta Pic Le disque de débris que l'on peut observer aujourd'hui autour de l'étoile β Pic est une version évoluée du disque circumstellaire dans lequel s'est formée la planète géante β Pic b. On ne peut donc pas s'appuyer sur les caractéristiques physico-chimiques du disque tel qu'observé aujourd'hui pour tâcher d'expliquer la formation de la planète. À la place, il est nécessaire de s'appuyer sur des observations d'autres disques, plus jeunes, ainsi que sur un certain nombre d'hypothèses.

On suppose donc que le rapport C/O global du disque est proche de celui de l'étoile, disons autour de 0.56. On suppose aussi que, comme cela semble être le cas dans de nombreux disques, les principaux porteurs d'oxygène et de carbone sont : l'eau, le monoxyde de carbone, le dioxyde de carbone, les silicates (SiO), et le carbone solide. En s'appuyant sur d'autres études de disques, ainsi que sur quelques hypothèses, on peut proposer les abondances relatives données en Table D.4.

À cela, on peut ajouter la Table D.5, qui donne la position des principales lignes des neiges dans deux jeunes disques connus situés autour d'étoiles de type A (comme β Pic), afin d'avoir une idée de la position possible de ces lignes dans le jeune disque de β Pic. Les silicates et le

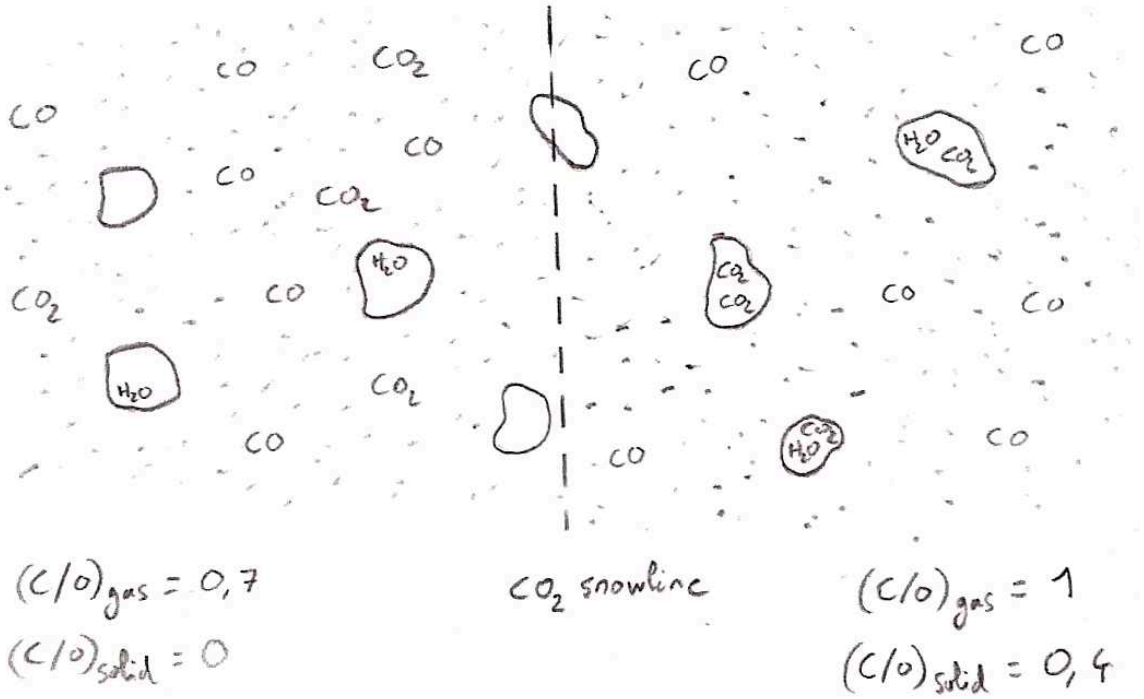


FIGURE D.21: Schéma illustration la situation en termes de distribution des atomes de carbone (C) et d'oxygène (O), autour de ligne des neiges (snowline) du CO_2, dans un jeune disque protoplanétaire. Les grains de poussière sont représentés comme des morceaux solides flottant dans la phase gazeuse (les petits points) du disque. Se référer au texte pour la discussion complète de de ce schéma.

Espèce chimique	Abondance	Commentaire
H_2O	1	prise comme référence arbitraire
CO	1	Pontoppidan et al. 2005
CO_2	0.2	Rapport 5 :1 avec le CO (Marboeuf et al. 2014)
C (grains)	0.4	Valeur choisie pour garantir un C/O global de 0.56
O (silicates)	0.5	Dans la plage proposée par Öberg et al. 2011

TABLE D.4: Table donnant les abondances (relatives à l'eau) des principaux porteurs de carbone et d'oxygène, dans un modèle simplifié du disque protoplanétaire de β Pic.

	Ligne des neiges		
	H_2O	CO_2	CO
$T_{\text{évap}}$	135 K	47 K	20 K
AU Aur	9.2 UA	96.3 UA	642.8 UA
V892 Tau	8.3 UA	51.2 UA	223.6 UA

TABLE D.5: Table donnant les températures d'évaporation des trois principales molécules porteuse de carbone et d'oxygène, ainsi que la position des lignes des neiges associées (en Unités Astronomiques), pour deux disques observés autour d'étoiles de type A (comme β Pic) : AU Auriga, et V892 Taurus. Valeurs tirées de Andrews et Williams [2005].

carbone solide sont des matériaux réfractaires, qui n'existent que sous forme solide.

Finalement, le rapport entre la quantité de gaz et la quantité de poussière, ou rapport $f_{s/g}$, peut être supposé égal à 0.01, une valeur courante dans les disques circumstellaires.

Á partir de là, les principales quantités sont posées. Place aux calculs !

Théorie de l'effondrement gravitationnel L'une des théories possibles pour expliquer la formation de planètes extrêmement massives comme β Pic b (environ 12 masses de Jupiter, $12\,M_{\text{Jup}}$), est la théorie de l'effondrement gravitationnel. Dans cette théorie, une protoplanète, contenant l'essentiel de la masse finale de la planète, se forme initialement à partir de l'effondrement d'une portion du disque. Suite à ça, la protoplanète, qui est initialement très peu dense, et donc de très grand diamètre, peut potentiellement accréter des grains de poussière, capturés par sa très massive atmosphère. Cette phase dure typiquement quelques milliers d'années, la protoplanète se contractant peu à peu par refroidissement. Au bout d'un certain temps, la planète atteint les conditions de la dissociation de l'hydrogène, évènement cataclysmique au cours duquel la protoplanète se contracte très brutalement, pour atteindre sa taille quasiment définitive. Après cette contraction, la planète est trop petite pour continuer d'accréter efficacement de la poussière, et la formation est terminée.

Dans ce scénario, la masse de la planète peut donc se décomposer en deux parties : la masse issue de la phase gazeuse du disque, et la masse issue de la phase solide. On écrit simplement :

$$M_{\text{planète}} = M_{\text{solide}} + M_{\text{gaz}} \tag{D.30}$$

La masse issue du solide peut elle-même se décomposer en deux parties : la masse initialement contenue dans le fragment du disque qui s'est effondré, et la masse accrétée suite à l'effondrement, avant la dissociation de l'hydrogène :

$$M_{\text{solide}} = M_{\text{solide,initiale}} + M_{\text{accrétée}} \tag{D.31}$$

Comme il n'y a pas d'effondrement préférentiel du gaz ou de la poussière, la protoplanète initialement formée contient du gaz et de la poussière dans les mêmes proportions que le disque lui-même. On a donc $M_{\text{solide,initiale}} = f_{s/g} M_{\text{gaz}}$, d'où :

$$M_{\text{solide}} = f_{s/g} M_{\text{gaz}} + M_{\text{accrétée}} \tag{D.32}$$

Dans ces conditions, quel est le rapport C/O possible de la planète ? Pour le savoir, il suffit de compter les atomes...

Introduisons encore quelques notations. On note $n_{C,s}$ (resp. $n_{O,s}$) le nombre d'atomes de carbone (resp. d'oxygène) par unité de masse solide du disque, et $n_{C,g}$ (resp. $n_{O,g}$) le nombre d'atomes de carbone (resp. d'oxygène) par unité de masse gazeuse. Le nombre d'atomes de carbone contenus dans la planète est donc :

$$\begin{aligned} N_{\text{carbone}} &= n_{C,s} \times M_{\text{solide}} + n_{C,g} M_{\text{gaz}} && \text{(D.33)} \\ &= \left(n_{C,g} + f_{s/g} n_{C,s}\right) \times M_{\text{gaz}} + n_{C,s} \times M_{\text{accrétée}} && \text{(D.34)} \end{aligned}$$

Et, pour l'oxygène :

$$N_{\text{oxygène}} = \left(n_{O,g} + f_{s/g} n_{O,s}\right) \times M_{\text{gaz}} + n_{O,s} \times M_{\text{accrétée}} \tag{D.35}$$

Ce qui donne un rapport C/O :

$$\text{C/O} = \frac{\left(n_{C,g} + f_{s/g} n_{C,s}\right) \times M_{\text{gaz}} + n_{C,s} \times M_{\text{accrétée}}}{\left(n_{O,g} + f_{s/g} n_{O,s}\right) \times M_{\text{gaz}} + n_{O,s} \times M_{\text{accrétée}}} \tag{D.36}$$

Tous les termes de cette équation, hormis $M_{\text{accrétée}}$, sont connus : les termes $n_{X,s/g}$ peuvent se calculer à partir des abondances données en Table D.4, en tenant compte de l'état (gazeux ou solide) de chaque composant ; $f_{s/g}$ est fixé à 0.01 ; et la masse de la planète étant largement dominée par le gaz, on a $M_{\text{gaz}} \simeq M_{\text{planète}} = 12\, M_{\text{Jup}}$. On peut donc tracer l'évolution du rapport C/O en fonction de la masse de solide accrétée $M_{\text{accrétée}}$, comme illustré en Figure D.22.

Les valeurs des termes $n_{X,s/g}$ étant dépendantes de la position dans le disque, par rapport aux différentes lignes des neiges, deux courbes sont tracées en Figure D.22. Ces courbes correspondent respectivement à une formation par effondrement gravitationnel en deçà de ligne des neiges de l'eau, et entre la ligne des neiges de l'eau et celle du CO_2 (qui sont les deux hypothèses raisonnables, étant données les valeurs de la Table D.5, et le fait que la planète β Pictoris b se situe à environ 10 UA de son étoile).

Sur cette même Figure D.22, plusieurs autres éléments ont été ajoutés. La mesure de C/O issue de la comparaison des modèles atmosphériques avec les observations est indiquée comme une bande grise (intervalle de confiance à 68%). Trois limites à la quantité de matière solide que la planète peut accréter au cours de sa formation sont données comme des lignes verticales : une limite absolue issue d'un calcul de la masse totale disponible dans le disque à la localisation de la protoplanète (la plus à droite), et deux limites tenant compte de la vitesse d'accrétion et du temps disponible (pour 10000 ans, ce qui est extrêmement optimiste, et 1000 ans, ce qui semble une durée plus raisonnable de la phase pré-dissociation).

Ce que montre cette Figure, c'est que même dans le cas le plus optimiste (formation entre les lignes de neige de l'eau et du CO_2, temps extrêmement long de la phase pré-dissociation permettant à la jeune protoplanète de ramasser toute la matière solide disponible), il est difficile d'expliquer un C/O aussi faible que celui mesuré à partir des observations GRAVITY. Ce scénario de formation tient difficilement la route.

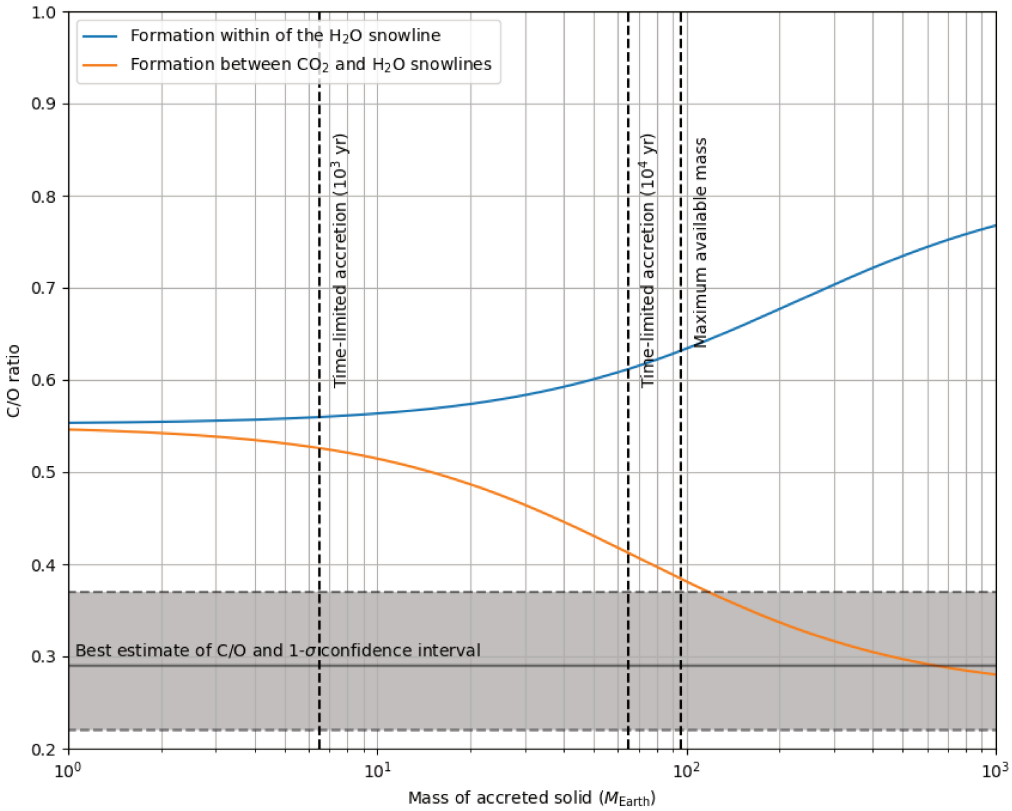


FIGURE D.22: Évolution du rapport C/O de l'atmosphère d'une planète formée par effondrement gravitationnel, en fonction de la masse de matière solide accrétée par la protoplanète avant la phase de dissociation de l'hydrogène, pour deux positions de formation possibles : avant la ligne des neiges de l'eau (courbe bleue), et entre la ligne des neiges de l'eau et celle du CO_2 (courbe orange). La zone grisée représente la mesure faite à partir du spectre GRAVITY, et son incertitude associée. Les différentes limites verticales sont discutées dans le texte. L'axe des abscisses est ici en échelle logarithmique, et en unité de masses terrestres.

Accrétion de cœur Il existe une théorie concurrente à la formation par effondrement gravitationnel : la théorie d'accrétion de cœur. Dans cette théorie, le processus de formation planétaire est beaucoup plus long, et se décompose en trois phases différentes : dans une première phase, un cœur planétaire de forme à partir de la coagulation de grains solides ; une fois le cœur formé, la protoplanète accrète lentement et régulièrement du gaz et des poussières ; à partir d'une certaine masse, une situation d'accrétion galopante se déclenche, et la planète accrète une énorme quantité de gaz en très peu de temps ; cette accrétion vide de son gaz la région du disque ou orbite la planète, stoppant toute accrétion ultérieure, et mettant un terme à la formation de la planète.

Ce scénario est assez différent du premier, en ce sens qu'une partie importante de la matière solide entrant dans la composition de la planète est piégée dans le cœur, inaccessible à l'observation. Le C/O observé est alors donné par le C/O de l'atmosphère de la planète seulement, composée du gaz et du solide accrété *après* la formation initiale du cœur. Autrement dit, par rapport au scénario précédent, la planète « démarre » avec une quantité de solide $M_{\text{solide,initiale}}$, mais son atmosphère ne contient que du solide accrété au fur et à mesure de la formation. Le C/O dans ce cas s'écrit :

$$\text{C/O} = \frac{n_{C,g} \times M_{\text{gaz}} + n_{C,s} \times M_{\text{accrétée}}}{n_{O,g} \times M_{\text{gaz}} + n_{O,s} \times M_{\text{accrétée}}} \tag{D.37}$$

La deuxième différence importante réside dans la vitesse du processus. Si la formation par effondrement est un processus extrêmement rapide (quelques milliers d'années tout au plus), la formation par accrétion de cœur est lente, et il faut typiquement quelques millions d'années pour former une planète par ce mécanisme.

La fait que l'accrétion de cœur soit lente implique un traitement un peu différent de l'Équation D.37 par rapport au traitement de l'Équation D.36 effectué précédemment. En effet, sur quelques millions d'années, le disque a le temps de se transformer chimiquement. La chimie des disques reste mal comprise, mais des travaux récents montrent que certaines réaction chimiques sont importantes. On peut notamment citer la transformation d'une fraction importante de l'eau en dihydrogène et dioxygène $2\,H_2O \rightarrow O_2 + 2\,H_2$. Vis-à-vis de l'étude menée ici, la conséquence directe ce cette réaction est de faire passer une part importante de l'oxygène de la phase solide (la glace d'eau), à la phase gazeuse (le dioxygène).

Si l'on se place entre les lignes des neiges de l'eau et du dioxyde de carbone, on peut donc imaginer deux scénarios. Dans une premier scénario, la formation de la planète est relativement rapide, avec une durée totale inférieure à 1 million d'années. Dans ce cas, le disque n'a pas suffisamment de temps pour évoluer chimiquement, et l'on peut utiliser l'Équation D.37 telle quelle, en calculant les termes $n_{X,s/g}$ comme précédemment, à partir de la Table D.4. On obtient à nouveau une courbe donnant le C/O en fonction de la masse de solide accrétée, donnée en Figure D.23 (courbe bleue). À nouveau, en plaçant la valeur du C/O mesurée avec GRAVITY, on obtient qu'il faut une accrétion extrêmement importante (au moins 200 masses terrestres de solide accrété) pour expliquer le C/O de 0.29.

Toutefois, dans ce paradigme de formation planétaire, on peut aussi proposer un scénario un peu différent, dans lequel la formation est plus lente. Dans ce cas, l'évolution chimique joue un rôle, et il faut tenir compte de la transformation de l'eau en dioxygène. Si la phase d'accrétion lente et régulière s'étale sur environ 7 millions d'années ou plus, alors la planète accrète une quantité importante de glace d'eau durant cette phase, ce qui lui apporte de nombreux atomes d'oxygène, mais, lorsque la phase d'accrétion galopante se déclenche, le disque a suffisamment évolué pour que la planète gagne aussi une grande quantité de dioxygène. Ce scénario, illustré par la courbe rouge dans la Figure D.23, est à même d'expliquer un C/O tel que mesuré par GRAVITY avec moins d'une centaine de masses terrestres de solide accrété.

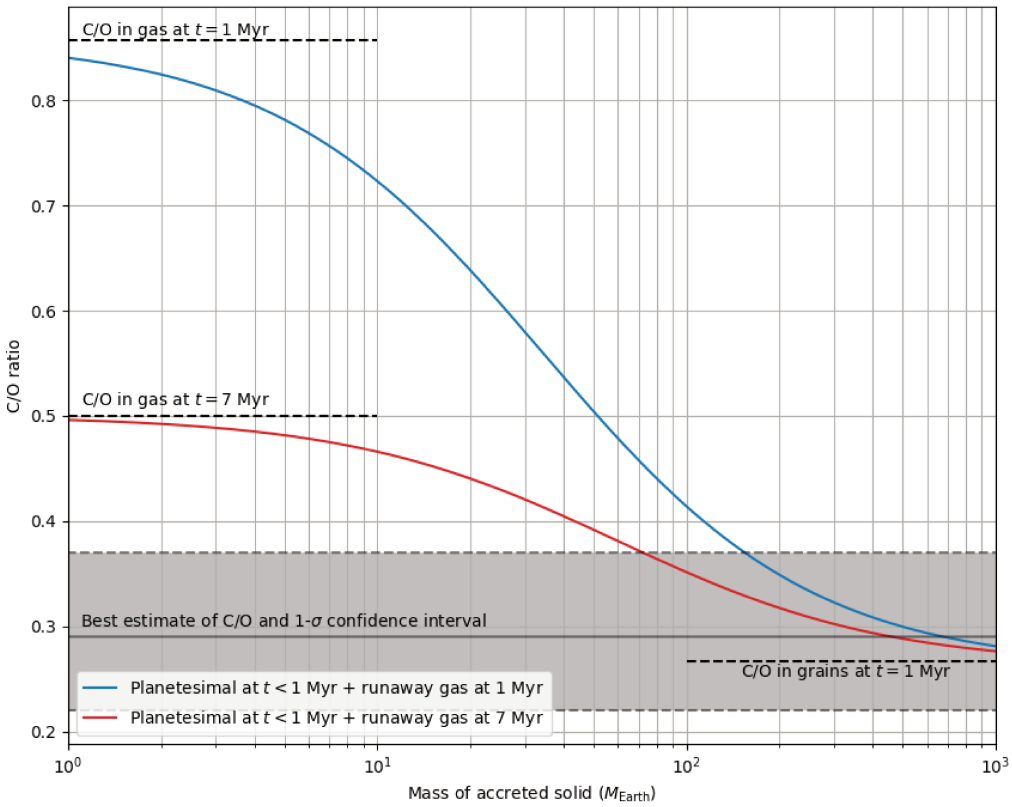


FIGURE D.23: Évolution du rapport C/O de l'atmosphère d'une planète formée par accrétion de cœur en fonction de la masse de solide accrétée avant la phase galopante. Ces courbes sont valables pour une formation dans la zone située entre la ligne des neiges de l'eau et celle du CO_2. Deux scénarios possibles sont représentés : une formation « classique », relativement rapide (1 million d'années, courbe bleue), et un scénario comme discuté dans le texte, pour lequel la phase galopante est retardée de 7 millions d'années (courbe rouge). À nouveau, la mesure GRAVITY et son incertitude associée sont figurées par la zone grise.

La raison pour laquelle la phase d'accrétion galopante pourrait être retardée de 6 ou 7 millions d'années dans un tel scénario n'est pas très claire, mais certains auteurs ont déjà suggéré par le passé que les processus de formation planétaire pouvaient être retardés par exemple en cas de compétition avec une autre planète en formation.

D.4 Conclusion

La mission PicSat, quoiqu'ayant permis de montrer qu'il était possible de concevoir et de lancer un nanosatellite avec une équipe restreinte, en moins de trois ans, n'a malheureusement pas donné les résultats scientifiques espérés sur la planète géante β Pictoris b.

En revanche, les observations de cette même planète avec l'interféromètre du VLT et l'instrument GRAVITY ont permis de mesurer pour la première fois le rapport C/O de son atmosphère. Ces données GRAVITY révèlent un C/O particulièrement bas de 0.29 environ, difficilement explicable dans un modèle de formation par effondrement gravitationnel. Cette mesure faite par GRAVITY pointe donc plutôt vers un scénario de formation par accrétion de cœur dans lequel, pour des raisons qui restent à éclaircir (compétition avec une autre planète

en formation ?), la phase d'accrétion galopante s'est trouvée retardée de plusieurs millions d'années. Dans un tel scénario, ce retard aurait permis à la planète de récupérer une quantité importante d'oxygène issu de la transformation de l'eau.

Bien sûr, un tel modèle est loin d'être parfait. La transformation de l'eau en dioxygène nécessite notamment que cette eau soit exposée au rayonnement de l'étoile, ce qui, en pratique, ne peut se faire qu'à la surface des grains de poussières. Il n'est pas évident, dans ces conditions, que la majorité de l'eau du disque puisse se transformer, comme supposé dans le scénario ici présenté. Il n'est pas non plus évident que la mesure du C/O telle qu'effectuée avec GRAVITY soit bien une mesure intégrée sur toute l'atmosphère. En pratique, la lumière qui nous parvient de la planète est issue des couches supérieures de l'atmosphère, et ne peut donc pas permettre de sonder les parties les plus en profondeur. Dans le cas de β Pic b, qui est une planète relativement chaude (1700 K), il est fort possible que ce problème soit largement atténué par le fait que de violents vents verticaux mélangent constamment l'atmosphère. Mais l'effet exact de ces vents reste incertain.

Quoi qu'il en ce soit, ces résultats obtenus avec GRAVITY forment une première tentative pour essayer de reconstruire l'histoire de la formation de β Pic b.

RÉSUMÉ

Beta Pictoris est probablement le système stellaire le plus étudié en dehors de notre propre système solaire. Ce système est une perle rare : l'étoile centrale est brillante, entourée d'un disque de débris vu par la tranche, et relativement jeune. Tout cela participe à faire de beta Pictoris une cible particulièrement intéressante pour l'étude de l'évolution des sytèmes planétaires. Trente ans d'observations par imagerie directe et par spectroscopie de transit ont déjà produit une moisson de résultats, qui dépeignent aujourd'hui un système extrêmement structuré, avec la présence d'anneaux, de ceintures, et d'une planète géante : beta Pictoris b. En revanche, pour ce qui est de la compréhension de l'origine du système et la façon dont il s'est formé, la récolte reste jusqu'à présent bien maigre, et l'on ne sait finalement que peu de choses sur la formation de beta Pic b, et la façon dont la planète intéragit avec le disque.
Cette thèse porte sur deux nouvelles tentatives faites pour observer la planète géante beta Pictoris b durant la période de sa conjonction inférieure, entre 2017 et 2018.
Dans la cadre de la campagne internationale d'observation du transit de la sphère de Hill de beta Pic b en 2017, un CubeSat trois unités, PicSat, a été lancé en orbite basse, avec à son bord un photomètre fibré. L'objectif de cette mission était de surveiller la luminosité de l'étoile beta Pictoris, afin de détecter et de caractériser le transit de la sphère de Hill de beta Pictoris b. La première motié de cette thèse détaille la conception, le développement, et les tests de l'instrument scientifique de la mission. Le satellite a été correctement déployé en orbite, mais un mauvais fonctionnement de son système commercial de contrôle d'attitude a rendu impossible toute obtention de données photométriques.
La seconde moitié de cette thèse présente une observation de la planète géante beta Pictoris b faite avec l'instrument GRAVITY, peu après la conjonction de la planète. Les quatre téléscopes de 8,2 m de l'Interféromètre du Très Grand Télescope (VLTI, en anglais) ont été combinés pour obtenir un spectre en bande K, à moyenne résolution, de la planète. Ce spectre donne notamment une première estimation du rapport planétaire carbone/oxygène, qui est fortement sous-stellaire. Combiné à un modèle d'evolution du rapport C/O au cours de la formation de la planète, cette observation indique que beta Pictoris b se serait formée par accrétion de cœur, plutôt que par effondrement gravitationnel du disque proto-planétaire.

MOTS CLÉS

beta Pictoris, Exoplanète, Cubesat, Interférométrie, Formation des planètes

ABSTRACT

Beta Pictoris is arguably one of the most studies stellar system outside of our own. The rare combination of a bright star, an edge-on debris disk, and the young age of the system make beta Pic a remarkable target to study the evolution of planetary systems. Thirty years of direct imaging and transit spectroscopy observations have revealed a highly structured disk, with rings, belts, and a giant planet: beta Pic b. But very little is known about how the system came into being. In particular, the formation history of beta Pic b and the way it interacts with the disk remain elusive.
This thesis focuses on two new attempts made to observe the giant planet, at the time of its inferior conjunction, in 2017/2018.
As part of the international observation campaign dedicated to the 2017 beta Pic b Hill sphere transit, PicSat, a small three-unit CubeSat, was launched in Low Earth Orbit. The satellite carried a fibered photometer, whose objective was to monitor the brightness of the star beta Pictoris, in order to detect and characterize the transit of the Hill sphere of beta Pictoris b. The design, development, and tests of this instrument are detailed in the first part of this thesis. The satellite was successfully deployed in orbit, but a malfunction of its commercial attitude control system made it unable to produce any photometric data.
The second part of this thesis introduces VLTI/GRAVITY spectro-interferometric observations of the giant planet beta Pictoris b, made soon after the conjunction. The combined four 8.2 m telescopes of the VLTI were used to obtain a high signal to noise, medium resolution, K-band spectrum of the planet, from which the first estimate of the planetary carbon-to-oxygen ratio was derived. The resulting substellar C/O ratio suggests that the giant planet is unlikely to have formed through gravitational instability, and that core-accretion is a more plausible scenario.

KEYWORDS

beta Pictoris, Exoplanet, Cubesat, Interferometry, Formation of planets