Master thesis and internship[BR]- Master's thesis : High-contrast imaging of protoplanetary disks with VLT/NaCo[BR]- Integration internship

Auteur : Savonet, Robin
Promoteur(s) : Christiaens, Valentin
Faculté : Faculté des Sciences appliquées
Diplôme : Master en ingénieur civil en aérospatiale, à finalité spécialisée en "aerospace engineering"
Année académique : 2024-2025
URI/URL : http://hdl.handle.net/2268.2/23301

University of Liège
School of Engineering and Computer Science

High-contrast imaging of protoplanetary disks with VLT/NaCo

SAVONET Robin

Master's thesis completed in order to obtain the degree of:
Master of Science in Aerospace Engineering

Thesis supervisor:
CHRISTIAENS Valentin

Reading committee:
ABSIL Olivier
CIOPPA Anthony
VAN DROOGENBROECK Marc

Academic year 2024 - 2025

Abstract

Despite the significant number of extra-solar planets that have already been discovered, our understanding of their formation and early evolution remains incomplete. Most detection methods are effective at identifying mature exoplanets, often orbiting relatively old stars. However, these techniques provide limited insight into the processes that govern the birth and development of planetary systems. Among the various available approaches, one stands out for its unique ability to observe exoplanets in their formative stages and within their birth environments: high-contrast imaging.

High-contrast imaging allows astronomers to spatially resolve circumstellar environments and detect faint planetary companions embedded in the disks of gas and dust that surround young stars. These protoplanetary disks are crucial to understanding the mechanisms that lead to planet formation, disk clearing, and the subsequent evolution of planetary systems.

The aim of this thesis is to apply state-of-the-art post-processing algorithms to archival data. More specifically, this work focuses on data obtained with the now-decommissioned NaCo instrument, formerly installed on the Very Large Telescope (VLT) in Chile. Although NaCo is no longer operational, its extensive data archive remains a valuable resource for exploring young star systems with modern computational tools.

This approach demonstrates that even archival observations can yield novel scientific insights when reprocessed with advanced methods. As a result, the combination of legacy data and modern analysis holds promise for expanding our understanding of planet formation without the need for new observations - an efficient and sustainable approach in the context of limited telescope time.

Acknowledgments

First, I would like to thank the entire team at PSILAB for warmly welcoming me during the semester. In particular, I am grateful to Olivier Absil for introducing me to the field of exoplanets and their observation through his insightful lectures. I am especially thankful to Valentin Christiaens for proposing this project and for his continuous guidance, support, and valuable advice throughout its realization. I would also like to thank Justin Latour for welcoming me into his office and for including me in the daily lunches with the amazing team at the Astrophysics Department. Thank you all for always being available to answer my questions and for offering helpful advice.

On a more personal note, I want to express my gratitude to some of the most important people in my life who are always there for me. Thank you, Robin, for being an incredible support - not only for this project but throughout the years and across so many other endeavors. Your unwavering friendship has always encouraged me to persevere. To Zoé, thank you for your constant support and motivation, especially during the final stages of writing. You are my locomotive - you push me forward every single day. Finally, I extend my deepest thanks to my parents, my two sisters, and my entire family for always believing in me, even through there were bumps along the way. You taught me to follow my dreams and to look up to the stars with admiration.

A special thank you goes to Markus Bonse and his team for providing the pre-processing pipeline used to reduce the ESO datasets that served as the foundation for this work.

Contents

Acronyms

ADI Angular Differential Imaging. v, 22–27, 29–32, 42, 47, 48, 61, 66, 67

AGPM Annular Groove Phase Mask (type of coronagraph). 20, 21, 34, 38–40, 42, 61

ALMA Atacama Large Millimeter Array. 15, 17

ANDROMEDA ANgular Differential OptiMal Exoplanet Detection. 26

AO Adaptive Optics. 18–21, 23

ARDI Angular and Reference-star Differential Imaging. vi, 24, 27, 31, 42–45, 49, 51, 53, 58, 61, 66, 67, 72

AU Astronomical Unit. 1, 51

DI Differential Imaging. 20

ESO European Southern Observartory. 7

FWHM Full-Width at Half Maximum. 26, 37, 39

I-PCA Iterative Principal Component Analysis. vi, 25, 31–33, 42–45, 48–51, 53, 54, 56–60, 63–65, 67, 68, 70, 71

IAU International Astronomical Union. 3, 4

JWST James Webb Space Telescope. 46

LLSG Local Low-rank plus Sparse plus Gaussian-noise decomposition. 26

LOCI Locally Optimized linear Combination of Images. v, 25

MSE Mean Squared Error. 37, 40, 42

MUSE Multi Unit Spectroscopic Explorer: VLT instrument. 48

NaCo Nasmyth Adaptive Optics System (NAOS) - Near-Infrared Imager and Spectrograph (CONICA): VLT instrument. vi, 2, 20, 33, 34, 45, 48, 49, 52, 53, 56–59, 62–64, 66, 69–74

PA Parallactic Angles. 25, 26, 31, 48, 49, 51, 52, 56, 61, 62

PACO Patch Covariances. 27

PC Principal Component. 27–31, 43, 44, 49, 51–53, 56–59, 61, 63, 64, 66, 68, 69, 72–74

PCA Principal Component Analysis. 25, 27–32, 34, 36, 37, 40, 43, 44, 56, 62, 68, 70, 73, 74

PCC Pearson Correlation Coefficient. 36, 37, 40, 42

PSF Point Spread Function. 18–21, 25, 27–32, 68, 69

RDI Reference-star Differential Imaging. 23, 24, 27, 29–31, 35, 40, 42, 56, 59, 61, 63–67, 70

SDI Spectral Differential Imaging. 27

SNR Signal-to-Noise Ratio. 38, 40–42, 70

SPHERE Spectro-Polarimetric High-Contrast Exoplanet Research: VLT instrument. 2, 14, 20, 48, 52, 57, 63

SSIM Structural Similarity Index Measure. 37, 40

STIM Standardized Trajectory Intensity Mean. 44, 68

SVD Singular Value Decomposition. 28

VLT Very Large Telescope. vii, viii, 2, 20, 33, 48, 52, 57

Chapter 1

Introduction

Although speculative claims about the existence of extrasolar planets date back to the 16^{th} century, the true challenges of detecting such worlds only became evident as astronomers came to understand the limitations of available technology. The prospect of directly imaging a distant object up to a billion times fainter than its host star once seemed more like science fiction than scientific reality. Yet, throughout history, humanity has never ceased to wonder what lies beyond. As our understanding of the universe expanded and its vastness became increasingly clear, a fundamental question began to emerge: could the sequence of events that gave rise to intelligent life on Earth truly be unique in the immensity of the cosmos? It is this very question that has fueled humanity's enduring pursuit of exoplanets and the search for worlds capable of supporting life.

To that end, the 20th century saw the emergence of increasingly accessible technology for detecting potential exoplanets. Larger and more sophisticated telescopes were being built, while alternative detection techniques were actively researched. Throughout the century, multiple claims of exoplanet discoveries were made, but it was not until 1992 that Aleksander Wolszczan achieved the first confirmed detection of an exoplanet (Wolszczan & Frail [1992]). This planet was detected due to its orbit around a pulsar, which enabled precise timing measurements. From that point onward, numerous other exoplanets were discovered and confirmed using a variety of detection techniques. Notably, the first exoplanet orbiting a main-sequence star was detected in 1995 (Mayor et al. [1995]), marking a significant milestone in the field. This first planet was a major discovery for the field, as it is roughly the size of Jupiter but orbits its host star at only $\sim$0.052 AU. For comparison, the semi-major axis of Mercury is 0.387 AU, meaning this Jupiter-sized planet is extremely close to its star. At the time of its discovery, this was completely inconsistent with the prevailing theories of planet formation. Since then, more of these so-called "hot Jupiters" have been discovered, prompting new theories about their formation.

Nowadays, using various detection techniques, nearly 6000 exoplanets have been discovered and confirmed (Dressing et al. [2025]). Each method has its own strengths and limitations, making it highly valuable to combine multiple techniques to validate observations - though some methods have mutually ex-

clusive requirements (e.g., the transit method favors short orbital periods, while direct imaging requires wide orbital separations). Among these, one particularly promising approach is direct imaging, as it is one of the few techniques capable of estimating key parameters such as an exoplanet's temperature, mass, and atmospheric composition. However, only about a hundred exoplanets have been discovered and confirmed using this method. This is primarily due to two major challenges: the limited angular resolution and the extreme contrast between a bright host star and its faint planetary companion. To overcome these challenges, several technologies are employed to enhance contrast and resolution. These include adaptive optics systems, which correct for atmospheric turbulence in ground-based observations, and coronagraphs, which block out most of the host star's light. In addition, observational strategies such as differential imaging and advanced post-processing algorithms are used to further improve performance. Despite these advancements, direct imaging has so far been limited to massive planets at large orbital separations. However, with rapid technological progress - including the development of extreme adaptive optics and upcoming facilities like the Extremely Large Telescope - the future of direct imaging holds great promise.

The main objective of this Master's Thesis is to image young stellar systems, as studying such systems can help us better understand the formation of our own Solar System. It is known that young stars are surrounded by a cloud of gas and dust called a circumstellar disk, and observing these disks - along with potential protoplanets forming within them - is a particularly challenging task (to date, only two protoplanets have been discovered and confirmed by the scientific community). Direct imaging is one of the few methods currently capable of efficiently resolving protoplanetary features such as disks and protoplanets.

Recently, new post-processing algorithms have been proposed to enhance the quality of direct imaging results for young stellar systems. These algorithms have been successfully tested on VLT/SPHERE datasets (Juillard et al. [2024]), but have not yet been applied to older archival data such as that from VLT/NaCo. The objective of this work is to assess how much these newly developed algorithms can improve the quality of images obtained from the now-decommissioned VLT/NaCo instrument.

In Chapter 2, the scientific background related to exoplanetology and young stellar systems is introduced. This is followed by a technical overview of high-contrast imaging methods in Chapter 3. The objective of the thesis is then outlined in Chapter 4, with a detailed description of the methodology provided in Chapter 5. Before presenting the results, the selected targets of interest are introduced in Chapter 6. The outcomes of the study are then presented in Chapter 7 and further discussed in Chapter 8. Finally, Chapter 9 concludes the thesis with a summary of the work and a perspective on future directions.

Chapter 2

Exoplanets and young stellar systems

Thanks to space missions like Gaia and Hipparcos (now both retired), new stars are discovered every day. It is not a stretch to assume that many of these stars host their own systems of orbiting planets. This assumption has proven to be accurate, as nearly 6000 exoplanets in over 4000 planetary systems have been detected and confirmed to date.

Unfortunately, due to the distance between us and any potential exoplanet, the task of detecting them can be daunting. However, thanks to clever technologies and detection techniques, the number of detected planets increases every year, as shown in Figure 2.1. It appears that the most effective method for detecting exoplanets is by far the transit method, followed by the radial velocity method. Microlensing comes next, and finally, direct imaging. This confirms that direct imaging is a challenging technique; however, it remains essential for studying young stellar systems as it is the only method capable of doing so effectively. All these methods are explained individually in Section 2.2, but first, we must better understand the science behind the different targets we aim to analyze.

2.1 Exoplanets

Another word for exoplanet is extrasolar planet, referring to planets located outside of our Solar System. To begin, we must first understand what constitutes a planet. Despite seeming straightforward at first glance, defining the term "planet" has historically been a significant scientific debate. The most well-known controversy surrounds the reclassification of Pluto from a planet to a dwarf planet in 2006, though few are familiar with the exact criteria that define a planet. According to the definition adopted by the IAU in 2006, a celestial body is considered a planet if it:

1. is in orbit around the Sun,

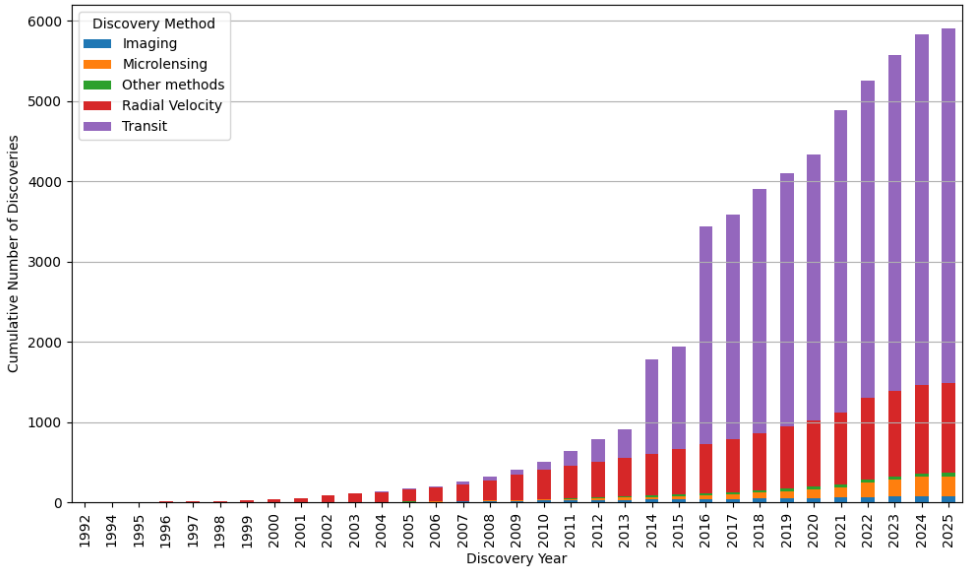


Figure 2.1: Cumulative exoplanet discoveries per year by method. (Data retrieved from Dressing et al. [2025] on May 21, 2025)

2. has sufficient mass for its self-gravity to overcome rigid body forces, so that it assumes a nearly round shape (hydrostatic equilibrium), and

3. has cleared the neighbourhood around its orbit.

If a celestial body satisfies only the first two criteria, it is classified as a dwarf planet - Pluto being the most famous example.

However, it is clear that this definition cannot be directly applied to extrasolar planets, primarily because exoplanets, by definition, do not orbit the Sun. Additionally, due to the vast distances separating us from these objects (with the closest, Proxima Centauri b, located 4.25 light years away), it is extremely difficult to determine whether they are spherical[1] or have cleared their orbits. As such, a different definition must be adopted for the classification of exoplanets. The currently accepted definition is the one most recently amended by the IAU in 2018 (Lecavelier des Etangs & Lissauer [2022]):

- Objects with true masses below the limiting mass for thermonuclear fusion of deuterium (currently calculated to be 13 Jupiter masses for objects of solar metallicity) that orbit stars, brown dwarfs or stellar remnants and that have a mass ratio with the central object[2] below the L_4/L_5 instability ($M/M_{central} < 2/(25 + \sqrt{621}) \approx 1/25$) are "planets" (no matter how they formed). The minimum mass/size required for an extrasolar object

[1] In practice, detected exoplanets must be sufficiently large to be observable, and are therefore very likely to be approximately spherical in shape.

[2] The term "central object" refers to the host stellar object whether it is a star, brown dwarf, or stellar remnant.

to be considered a planet should be the same as that used in our Solar System.

- Substellar objects with true masses above the limiting mass for thermonuclear fusion of deuterium are "brown dwarfs", no matter how they formed nor where they are located.
- Free-floating objects in young star clusters with masses below the limiting mass for thermonuclear fusion of deuterium are not "planets", but are "sub-brown dwarfs" (or another more appropriate term, depending on the context).

The reason why the upper mass limit is defined by the threshold for deuterium fusion rather than a fixed numerical value is because each celestial object has a different metallicity. This variation affects the conditions under which deuterium fusion occurs, causing the mass limit to vary from one object to another and making it unsuitable to impose a universal threshold.

It is also notable that any free-floating object - meaning one that does not orbit a star - is never considered a planet, regardless of its mass.

2.2 Indirect detection methods

As already introduced in Figure 2.1, there are several methods used to detect exoplanets in stellar systems. Most of these are classified as indirect detection methods, as they rely on observing the effects of a planet on its host star or environment rather than capturing the planet itself. In this section, we will briefly present a selection of these indirect detection techniques. The direct imaging method, which is central to this work, is discussed in detail in Chapter 3.

2.2.1 Transit

The transit photometry detection method relies on observing the periodic dimming of a star's light caused by a planet passing in front of it, partially blocking its emitted flux, as illustrated in Figure 2.2. This flux drop can reach a few percent for large planets orbiting small stars, but is generally much smaller - around 0.008% in the case of an Earth-sized planet transiting a Sun-like star. Consequently, like most detection techniques, the transit method is only effective under specific conditions. First, the planetary system must be oriented such that the orbital plane is aligned with our line of sight, allowing transits to occur from our vantage point. Second, the ratio between the planetary and stellar radii must be favorable - i.e., the planet must be large enough relative to the star for the flux drop to be detectable with available photometric precision. Finally, the planet's orbital period must be sufficiently short to ensure that a complete transit event can be observed within a single observing window. For all these reasons, hot Jupiters - gas giants in close-in orbits - are by far the most commonly detected exoplanets via this method.

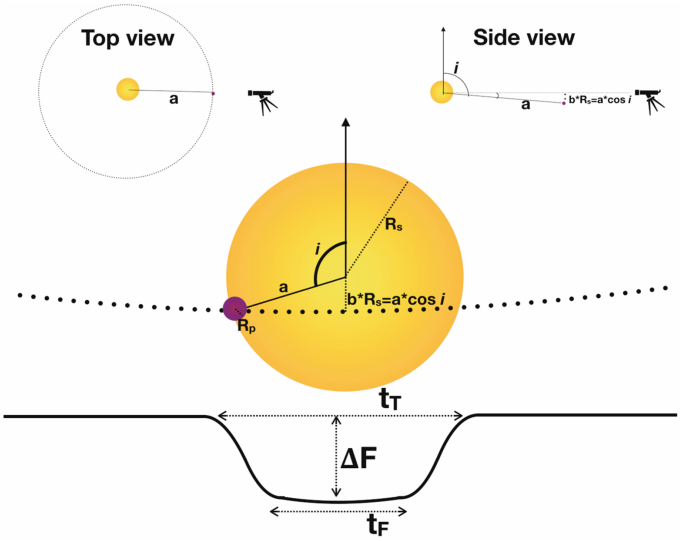


Figure 2.2: Schematic representation of a planetary transit across its host star, shown from both top and side views to illustrate the observer's perspective. The total duration of the transit is denoted by t_T, while t_F represents the interval during which the planet is fully in front of the stellar disk, corresponding to the period of maximum light dimming. (Credits: Deeg & Alonso [2018])

One of the primary advantages of the transit method over other indirect detection techniques is that it enables the determination of the planet's radius directly from the light curve. In some cases, it even allows for the characterization of a planet's atmosphere by analyzing the starlight that filters through its upper atmospheric layers during transit. However, the method also suffers from a notably high false positive rate. For instance, observations from the Kepler space telescope revealed that approximately 35% of close-in giant planet candidates were likely false positives (Santerne et al. [2012]).

2.2.2 Radial velocity

The second most productive method in terms of exoplanet detection is the radial velocity method. It was notably used in 1995 to discover the first exoplanet orbiting a Sun-like star (Mayor et al. [1995]). This technique relies on the fact that exoplanets are gravitationally bound to their host stars. As a planet orbits, it exerts a gravitational pull on the star, causing the star to move in a small, periodic motion. If this motion is large enough, it can be detected through shifts in the star's spectral lines, a phenomenon known as the Doppler effect. These Doppler shifts can be measured by comparing the star's spectrum to reference measurements of similar stars (Lindegren & Dravins [2003]). An illustration of this method is shown in Figure 2.3.

However, this method requires the system to be oriented such that the star's motion has a component along our line of sight - similar to the geometric con-

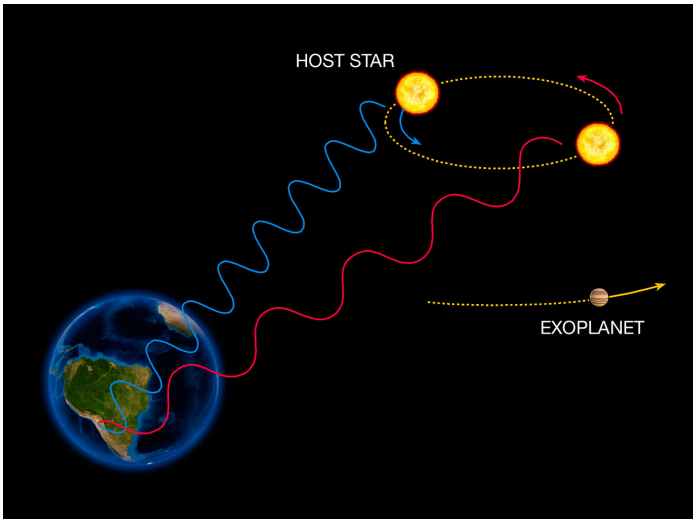


Figure 2.3: Illustration of the radial velocity method for exoplanet detection. A blueshift is observed in the star's spectrum when it moves toward the observer, and a redshift is observed when it moves away. (Credits: ESO)

straint of the transit method. Additionally, the exoplanet must be sufficiently massive relative to the star to induce a detectable gravitational effect. And if multiple massive companions orbit the star, it can become very challenging to disentangle the motion caused by each one. One notable advantage of the radial velocity method is that it allows astronomers to estimate the exoplanet's minimum mass and orbital eccentricity based on the observed Doppler shifts.

2.2.3 Gravitational microlensing

The gravitational microlensing detection method is rooted in the principles of general relativity, as illustrated in Figure 2.4. When a foreground star (the "lens") passes in front of a more distant background star, its gravitational field bends the light from the background source, causing a temporary increase in its observed brightness. If the lensing star hosts a planet, the planet's gravitational field can introduce a brief anomaly in the light curve of the event. These anomalies typically last from a few hours to several days, depending on the planet's mass, and are superimposed on microlensing events that themselves span weeks to months. This allows for the indirect detection of exoplanets orbiting the lensing star.

This method relies on a highly improbable event - the precise alignment of a background source star with a foreground lensing system. As a result, a very large number of background stars must be monitored continuously to capture potential microlensing events. However, unlike other detection techniques, microlensing allows for the discovery of exoplanets orbiting faint and distant stars, and can provide estimates of both the planet's and the host star's mass. Despite these advantages, the method has notable limitations. Each microlensing

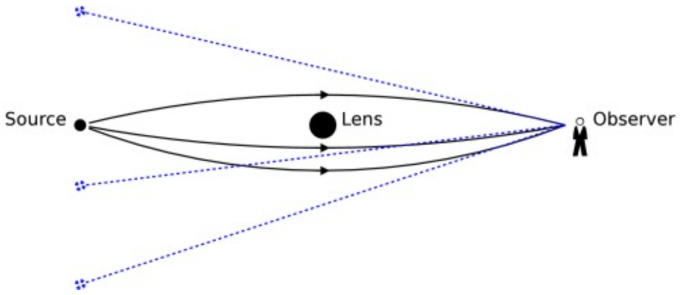


Figure 2.4: Illustration of the principle behind the gravitational microlensing method. Due to the gravitational field of a massive foreground object - referred to as the "lens" - the light from a distant background source is bent. To an observer, this bending causes the light from the source to appear as if it is coming from multiple positions around the lensing object, resulting in a temporary magnification of the source's brightness. (Credits: Kabdiyev [2023])

event is a one-time occurrence, offering no opportunity for repeated observations. Moreover, due to the faintness and distance of the targets, follow-up studies or confirmation using other detection techniques are typically not feasible.

2.2.4 Comparing methods

Now that the main detection methods are introduced, their respective advantages and limitations can be compared. A summary is presented in Table 2.1.

Finally, using the database of confirmed exoplanets provided by NASA, we can produce a scatter plot of exoplanet mass versus semi-major axis, sorted by detection method, to visualize the parameter space in which each method is most effective. This is shown in Figure 2.5. As previously discussed, the transit method yields the most detections for large planets - ranging from Earth-sized up to 10^4 M_{Earth} - and is primarily limited to planets in close orbits around their host stars. The microlensing technique extends to slightly larger semi-major axes but also tends to detect more massive planets, with relatively few Earth-sized detections. The radial velocity method is most effective for massive planets located between 1 and 10 AU from their stars, although some smaller planets have also been detected closer in. Finally, direct imaging is only capable of identifying the most massive planets - those exceeding 500 M_{Earth} (or 1.5 $M_{Jupiter}$) - and typically at large separations from the host star. This is because, for an exoplanet to be directly imaged, it must be both sufficiently large and located far enough from its host star for the planet-star contrast to be detectable with current instrumentation.

Detection method	Pros	Cons	Measurable quantities
Transit	Can detect Earth-sized planets	- Large exoplanets around small stars - Short periods - High false positive rate - Need particular orientation of the observed planetary system	- Radius - Atmospheric composition
Radial velocity	Can detect Earth-sized planets	- Large exoplanets - Need particular orientation of the observed planetary system	- Eccentricity - Mass
Microlensing	Can detect faint and distant source	- Unique detection - No confirmation possible - Based on an unlikely event; need constant monitoring to detect	- Mass - Host star mass

Table 2.1: Recapitulative table of the different indirect detection methods.

2.3 Protoplanetary disks

One might wonder about the relevance of studying young stellar systems. Beyond the pursuit of scientific knowledge about our universe, examining young planetary systems is key to uncovering how our own Solar System formed and evolved in its earliest stages.

2.3.1 Star formation

Nowadays, stars form from the gas present in dense molecular clouds, which are often remnants of supernovae - the powerful explosions of dying massive stars. Stars are classified into three populations based on their period of formation. The first stars, labeled "Population III," are believed to have formed at the dawn of the Universe, shortly after the Big Bang. They were composed only of primordial gas - hydrogen and helium - and lacked the heavier elements (metals) that are produced later in stellar evolution.

"Population II" stars formed from the debris of the first generation and thus contain more heavy elements than Population III stars, though still far fewer than Population I stars like our Sun. Unlike Population III stars, Population II stars have been directly detected through surveys targeting metal-poor stars in the Milky Way and neighboring galaxies (Klessen & Glover [2023]). However, since most stars currently under observation are Population I stars, the formation process of this population will be the focus of this section.

An interstellar cloud remains in equilibrium as long as the gas pressure bal-

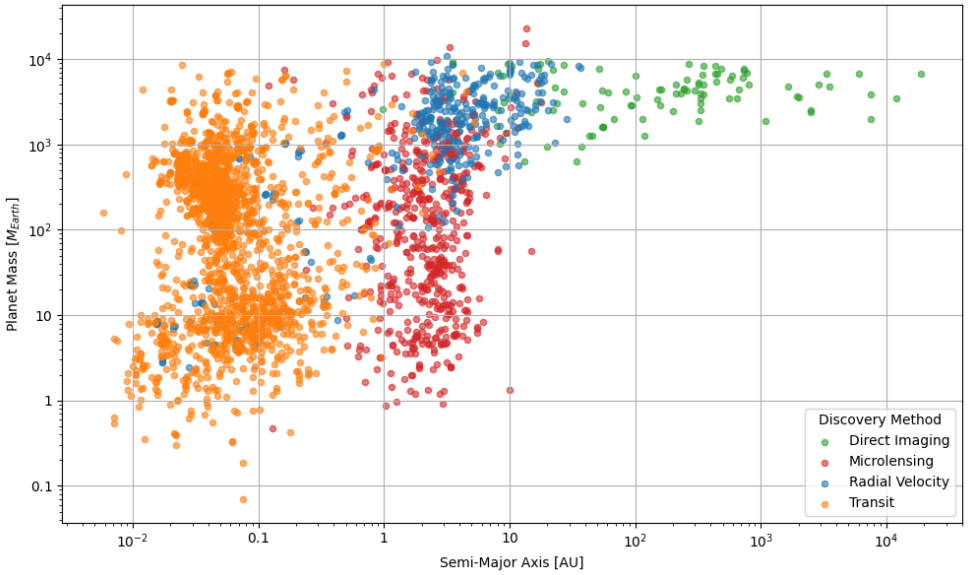


Figure 2.5: Exoplanet mass versus semi-major axis, color-coded by detection method. (Data retrieved from Dressing et al. [2025] on May 29, 2025)

ances its internal gravitational force. Once the cloud becomes sufficiently massive and pressure can no longer counteract gravity, it undergoes a process known as gravitational collapse. Such events can lead to the formation of thousands of stars at once, resulting in stellar clusters.

Each fragment of the collapsing cloud destined to become a star continues contracting under its own gravity until the temperature and pressure are high enough to ignite nuclear fusion reactions. These reactions release energy that raises temperature and pressure further, generating an outward expansion that balances gravity. When this hydrostatic equilibrium is achieved, the central region of the cloud is called a protostar.

During the collapse, the gas and dust orbiting the forming star - initially rotating slowly - begin to spin faster as it collapses due to the conservation of angular momentum. This rotation causes the material to flatten into a disk, forming a protoplanetary (or circumstellar) disk. This disk plays a crucial role in the stellar system's formation - not only by supplying the raw material for planet formation, but also by influencing the system's structure and long-term evolution.

2.3.2 Planet formation

Historically, models of planet formation have largely neglected dust evolution, assuming that planetesimals - the initial building blocks of planets - formed rapidly through gravitational effects (Safronov [1969]). However, recent models focusing on dust evolution have shown great promise (Drążkowska et al. [2023]). Figure 2.6 illustrates this evolution, from micrometer-sized dust grains to planets

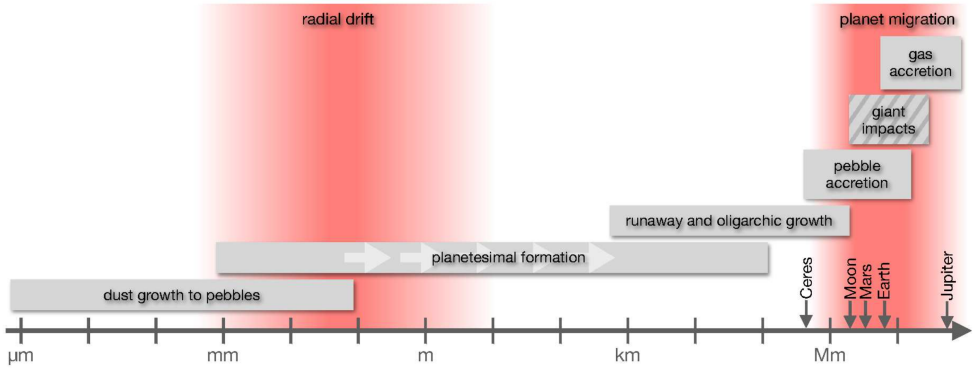


Figure 2.6: Overview of processes involved in planet formation and size scales at which they operate. The gray boxes represent growth processes and the red shaded regions represent where the radial mass redistribution is effective. The white arrows in the planetesimal formation box represent the lack of the intermediate sizes during the collapse of pebble clouds. The giant impacts may happen both before and after the protoplanetary gas disk disperses, while the other processes happen in the presence of gas. (Credits: Drążkowska et al. [2023])

measuring thousands of kilometers across. The process by which dust grows into millimeter-sized pebbles has been extensively studied for decades (Birnstiel et al. [2016]), but it remains poorly understood. Several scenarios have been proposed over the years. The first suggests that dust grows into larger particles through mass transfer during high-speed collisions. Although supported by laboratory experiments (Blum [2018]), this mechanism is too slow to prevent particles from spiraling into the star due to radial drift - a process in which solid material loses angular momentum through interaction with gas and migrates inward. The second scenario involves porous growth, where collisions between fluffy, porous ice aggregates allow growth beyond the bouncing and fragmentation barriers thanks to efficient sticking, even at high velocities (Okuzumi et al. [2012] and Kataoka et al. [2013]). However, this idea was challenged by the fact that the inner disk is too hot for ice to remain solid at relevant distances from the star (Homma & Nakamoto [2018]). However, the first scenario remains the most probable mechanism for pebble formation, since it is now commonly assumed that the drift process is halted at local pressure maxima (e.g., rings), as such substructures have been observed in the vast majority of resolved disks.

Following the formation of small pebbles, planetesimals are generally thought to form through the gravitational collapse of dense pebble clumps (Johansen et al. [2014]). Studies have shown that an enhanced solid-to-gas ratio in the disk could create favorable conditions for this to occur (Drążkowska et al. [2016]).

Once planetesimals - typically a few tens of kilometers in size and comparable to objects found in the asteroid belts of our own Solar System - have formed within the disk, they can grow further through collisions and accretion. Two main

regimes govern this stage (Armitage [2024]): *runaway growth* (Greenberg et al. [1978]), where the smallest protoplanets are not yet massive enough to influence the local planetesimal velocity dispersion, allowing any initial mass difference to amplify over time and favor the rapid growth of the larger bodies; and *oligarchic growth* (Kokubo & Ida [1998]), where more massive protoplanets stir the velocity dispersion of nearby planetesimals, which in turn slows their own accretion and allows neighboring, smaller protoplanets to catch up. As a result, planetary embryos grow at a more uniform rate.

Finally, large planetary embryos can collide with one another - either before or after the dispersion of the protoplanetary gas - to form terrestrial planets or the cores of giant planets. If these collisions occur before the gas has fully dispersed, the resulting giant planet cores can accrete surrounding gas, leading to the formation of giant planets (Bitsch et al. [2018]). Such collisions can also result in the formation of moons orbiting the planets, as is believed to have occurred in the case of Earth, which, according to this theory, collided with a Mars-sized body named Theia, producing a debris field that eventually could have coalesced into the Moon.

These theories about how planets form can sometimes be corroborated by observations of circumstellar disk structures. Indeed, if a spiral arm or a cavity devoid of material is observed in certain regions of the disk, it is likely that a protoplanet may have carved out these patterns. This topic will be explored further in the next section.

2.3.3 Protoplanetary disk structures

As previously mentioned, imaging and studying protoplanetary disks is a complex task. Their diffuse nature makes them more difficult to image than bright companions. However, one advantage of studying disks over point sources is that different regions of the disk emit or reflect light at different wavelengths of the electromagnetic spectrum, as illustrated in Figure 2.7. This allows observers to select instruments tailored to the specific spectral range corresponding to the part of the disk they wish to study. For example, dust suspended in the upper layers of the disk scatters light primarily in the infrared (IR) spectrum, while the thermal emission from solid particles within the disk is observed in the sub-millimeter to centimeter range. In contrast, gas within the disk emits light at specific emission lines, such as those from carbon monoxide (CO). A comparison of observations of the same disk around TW Hya at different wavelengths can be seen in Figure 2.8. It is clear that protoplanetary disks must be observed across multiple wavelengths to obtain a comprehensive view of their various components.

As for their structure, there are almost as many unique configurations as there are disks. Each planet-forming disk exhibits distinct morphological features, some of which are illustrated in Figure 2.9. However, despite this diversity, these structures can generally be grouped into common categories such as "rings" or "spirals". Disks observed around stars of different spectral types -

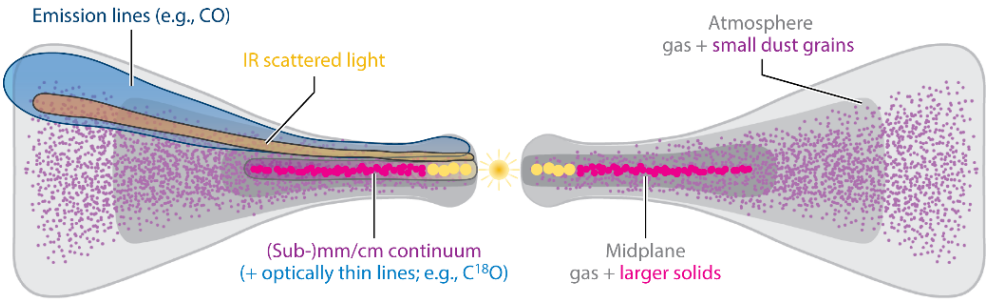


Figure 2.7: Diagram of a protoplanetary disk cross-section, with different colors indicating various wavelengths of emission. Gray areas represent gas, while the colored regions correspond to solid matter. (Credits: Andrews [2020])

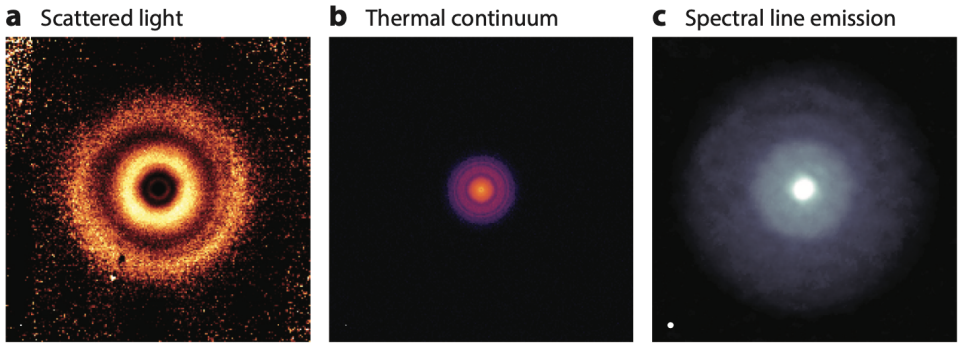


Figure 2.8: Observations of the same disk around TW Hya at different wavelengths. For scale, the outer disk is thought to extend to a radius of up to 200 AU (Vacca & Sandell [2011]). Scattered light from the suspended dust ($\lambda = 1.6\,\mu m$), thermal continuum from the disk pebble-sized solids ($\lambda = 0.9\,mm$) and spectral line emission from the gas (in this case, CO spectral line). (Credits: Andrews [2020])

from cool M-type stars to hotter A/B-type stars[3] - can all exhibit these characteristic structures, as shown in the figure. Observations of protoplanetary disks are often carried out in the H-band, and in some cases the J-band - atmospheric transmission windows in the infrared centered on 1.65 µm and 1.25 µm, respectively. But other wavebands like the R-band, the Y-band (e.g. Benisty et al. [2015]; Stolker et al. [2016]) or the K-band (Ren et al. [2023]) are also used as they are all part of the atmospheric transmission window. This choice of wavelength is consistent with the emission properties illustrated in Figure 2.8, as it corresponds to the scattered light produced by dust suspended in the upper layers of the disk.

[3]The Morgan-Keenan classification system categorizes stars using the letters O, B, A, F, G, K, and M, arranged from hottest (O) to coolest (M).

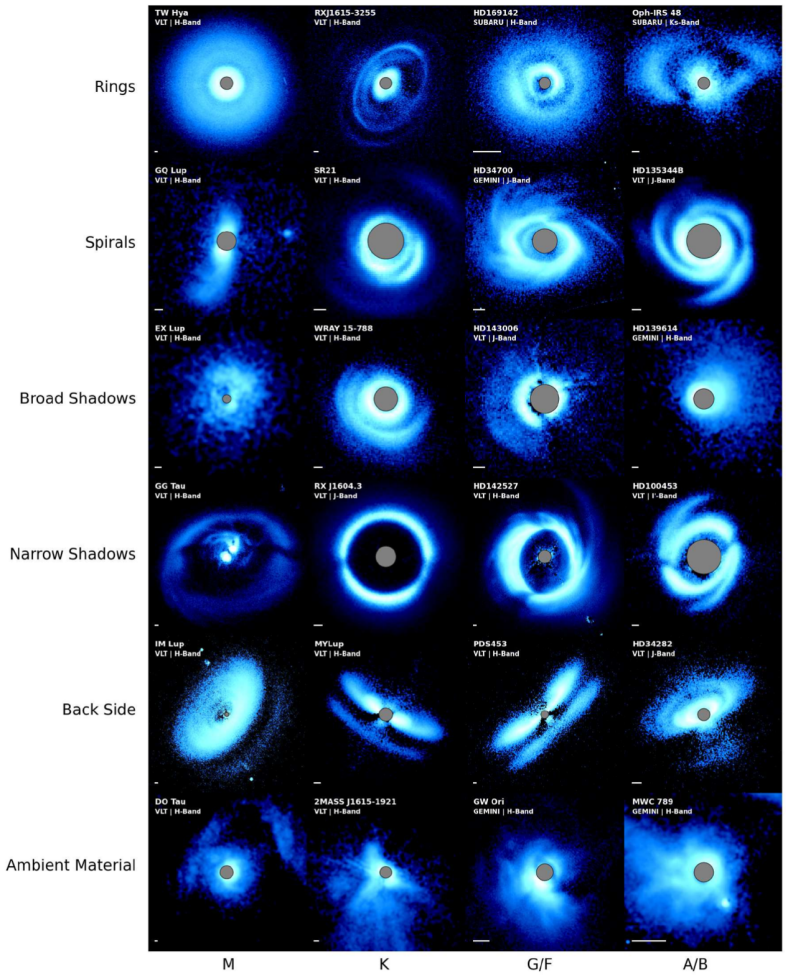


Figure 2.9: Gallery of disks observed with the SPHERE, GPI, and HiCIAO instruments, highlighting the wide range of morphological features present in planet-forming disks and their ubiquity across spectral types. All images are normalized within the depicted stamp and colorbars are arcsinh scaled. A 50 au scalebar appears in the lower left corner of each image, and grey central circles mark regions obscured by a coronagraph. North is up and East to the left in all images. The star name, instrument, and wavelength are indicated in the upper left corner of each image. (Credits: Benisty et al. [2023])

2.4 Protoplanet detection

As discussed in Section 2.2, some methods are inherently more effective than others for exoplanet detection. For instance, the transit method is known to have a relatively high false positive rate. While confirming or rejecting exoplanet candidates is often straightforward, this is not the case for protoplanet candidates. Most of these can only be detected through direct imaging, which limits

the ability to confirm them using complementary techniques. This context explains why, despite the numerous candidates proposed, the only system that is unanimously confirmed to host protoplanets remains PDS 70. Its two planetary companions, PDS 70 b and PDS 70 c, were discovered in 2018 (Keppler et al. [2018]; Müller et al. [2018]) and 2019 (Haffert et al. [2019]) respectively, and have since been confirmed by many other publications (Wagner et al. [2018]; Mesa et al. [2019]; Hashimoto et al. [2020]; Stolker et al. [2020]; Wang et al. [2020]; Zhou et al. [2021]).

2.4.1 Disk kinematics

Direct imaging of protoplanetary disks can sometimes reveal structural features that suggest the presence of a forming protoplanet. In disks without any embedded companions, the structure is typically expected to be asymmetrical in the orbital plane, producing what is commonly referred to as the "butterfly" pattern - an effect resulting from observing specific bands, showing the spatial emission morphology at specific frequencies[4] (Pinte et al. [2023]). However, when deviations from this pattern are observed, it may indicate the presence of a protoplanet actively forming through accretion of disk material but may also be due to other sources of deviation. Observations of such "butterfly" patterns and their potential perturbations have notably been conducted with the Atacama Large Millimeter Array (ALMA), as illustrated in Figure 2.10. This detection approach,

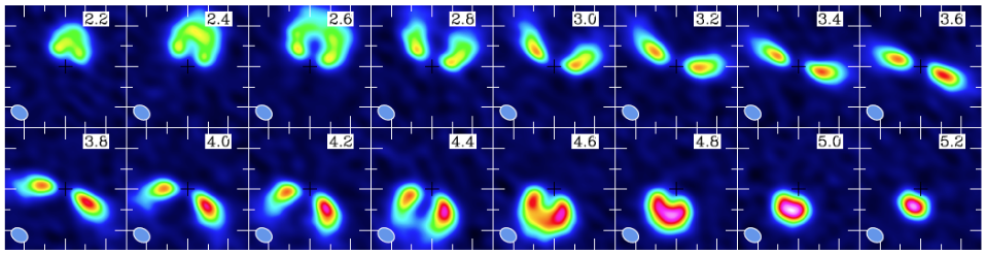


Figure 2.10: Observation of the circumstellar disk around HD 142527 using ALMA showing the "butterfly" effect that is usual in disk structure by observing ^{13}CO band transition. The LSR[5] velocity for each channel is presented in the top right corner of each panel. (Credits: Perez et al. [2015])

known as disk kinematics, is a promising method for identifying protoplanets in circumstellar disks. Some companions were already detected in the HD 97048 and HD 163296 (Pinte et al. [2018, 2019]) for example. Nonetheless, it remains necessary to conduct follow-up studies to confirm such detections.

[4]The reason behind this is the varied materials that constitute the different parts of the disk, gas of matter, emit in varied wavelengths.

2.4.2 Detection claims

The complex structures within the disks that host candidates often make it difficult to disentangle the planetary signal from that of the disk. Additionally, the reliability of post-processing algorithms is frequently questioned, and since each discovery typically relies on its own set of algorithms, it is common for subsequent studies to challenge the claims made in earlier publications (Benisty et al. [2023]). In this section, we present a few detections claims whose validity has either been disproven or remains under debate.

LkCa 15

LkCa 15 is a young solar analog star located in the Taurus-Auriga star-forming region, surrounded by a protoplanetary disk composed of three main belts with a total mass estimated at approximately 50 times that of Jupiter. The first detection of a potential protoplanetary companion, LkCa 15 "b", was reported by Kraus & Ireland [2012] using the NIRC2 instrument on the Keck II telescope. This detection was later supported by observations from Sallum et al. [2015] and Sallum et al. [2016], the latter of which also proposed the possible presence of two additional protoplanets in the system - LkCa 15 "c" and LkCa 15 "d". However, subsequent studies by Thalmann et al. [2016] and Currie et al. [2019] challenged these claims, arguing that the observed signals likely originated from scattered light produced by asymmetries in the disk rather than from orbiting protoplanets.

AB Aurigae

The AB Aurigae system, located in the Taurus-Auriga star-forming region, is a particularly promising candidate for hosting a forming planet. The central star is estimated to be approximately 2.4 times the mass of the Sun. A potential protoplanet has been identified in this system, situated within a gap in the circumstellar disk, as reported in Currie et al. [2022]. However, this candidate has yet to be either confirmed or definitively refuted.

Other Detections

Several other systems - such as HD 100546, MWC 758, and HD 169142 - have also been proposed as hosts of protoplanets. These claims, however, remain highly debated within the scientific community (Benisty et al. [2023]). In many cases, what appear to be planetary signals may instead be artifacts arising from complex disk structures or the limitations of post-processing algorithms used in direct imaging. Consequently, future observations and improved data analysis techniques will be essential to confirm or reject the existence of protoplanets in these systems. Some of these systems are analyzed in greater detail in the subsequent sections of this work.

Chapter 3

High contrast imaging

High contrast imaging of protoplanetary disks and potential planetary companions is a challenging task due to numerous factors that hinder the acquisition of clear images of such targets. However, it is the only techniques capable of capturing images of protoplanetary disks and potential protoplanets forming within their structures[1]. It is therefore essential to consider using high contrast imaging in order to study protoplanetary systems.

Given the challenges for high contrast imaging of protoplanetary systems, only a few protoplanets have been confirmed, with another dozen detected but yet to be verified (cf Section 2.4). Not only is detection challenging due to the circumstellar disk obstructing part of the protoplanet's light, but confirmation is also particularly difficult, as high contrast imaging remains the primary effective method for identifying protoplanets within circumstellar disks. In some cases, studying the orbit and movement of significant point sources within the circumstellar disk can lead to the detection or confirmation of a candidate companion through the analysis of its Keplerian motion (Pinte et al. [2018, 2019]).

Unfortunately, as with all companions orbiting distant stars, current state-of-the-art instruments can only detect those with large semi-major axes and high masses, significantly limiting the detectable population. Expanding the range of detectable companions will therefore require the next generation of telescopes currently under construction, particularly those equipped with extreme adaptive optics and larger mirror diameters, such as the Extremely Large Telescope (ELT) in Chile. To grasp the limitations of current technology, it is essential to understand the two key challenges posed by high contrast imaging: angular resolution and contrast.

First, the angular resolution θ can be calculated using the Rayleigh criterion, defined as $\theta = \frac{1.22\lambda}{D}$, where λ is the wavelength and D is the telescope diameter. Therefore, achieving the angular resolution necessary to observe companions close to a star requires larger mirrors.

Second, the contrast refers to the flux difference between the extremely bright star and its orbiting companion. For Earth-like planets, this contrast can

[1] Other forms of direct imaging including interferometry, using ALMA for exemple, are also capable of detecting protoplanets.

be as low as 10^{-10}, which is beyond the reach of current technology. This limitation explains why only large Jupiter-like planets can be directly imaged, as their size and sometimes self-luminescence increase the contrast up to 10^{-3}. Nevertheless, the planet's flux remains very small compared to the star's flux - an analogy often used is trying to spot a firefly next to a lighthouse - but with advanced techniques, this is now achievable.

In this section, several techniques will be described, each aiming to overcome these two challenges by improving both the angular resolution and contrast of the observations.

3.1 Adaptive Optics

The first technique used to enhance image quality is called Adaptive Optics (AO) . Its objective is to minimize distortion caused by atmospheric turbulence as much as possible. Atmospheric turbulence caused by temperature variations and wind distorts the wavefront reaching the telescope from the observed target. On the detector, this appears as a pattern of light speckles rather than a classic Point Spread Function (PSF), which, when averaged over time, produces a significantly blurred image. In conditions where atmospheric turbulence is not accounted for, the angular resolution is no longer constrained by the Rayleigh criterion over the telescope diameter but rather by the Fried parameter: $\theta = \frac{\lambda}{r_0}$ where the Fried parameter r_0 is strictly dependent on the atmospheric conditions at a given observation site. Consequently, the angular resolution of an uncorrected optical system is determined not by the telescope diameter but by the Fried parameter, which can only reach 10 cm at the best astronomical sites. This underscores the importance of using adaptive optics to correct the distorted wavefront, as relying solely on the Fried parameter would limit the measurements to the equivalent of those obtained with a 10 cm class telescope.

This is done by analyzing the wavefront coming through the optical instrument and compensating for the distortions using a deformable mirror. This mirror functions as a standard reflective surface equipped with actuators that adjust its shape to correct the waveform as needed. Figure 3.1 presents a schematic of a classical adaptive optics system. As previously described, the distorted wavefront enters the telescope and reaches the deformable mirror (here named adaptive mirror). The corrected wavefront from the deformable mirror is then directed to a beamsplitter. Part of the flux continues through the optical system to capture light from the telescope's target, while the remainder of the flux is sent to a wavefront sensor, which provides input to the control system that adjusts the mirror's actuators. Despite some of the light flux being lost at the beamsplitter to enable wavefront measurement, the image enhancement provided by using an AO system far outweighs this loss in light intensity.

To demonstrate the effectiveness of an adaptive optics system, we can examine the PSF of a binary star, as shown in Figure 3.2. The left side of the image presents a long-exposure PSF without any correction from an AO system,

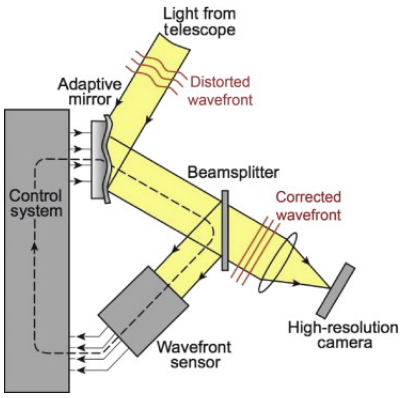


Figure 3.1: Simplified diagram of a classical adaptive optics system. (Credits: Tokunaga [2014])

while the right side shows the PSF with AO correction. Without an AO system to correct the distorted wavefront, it would be impossible to discern that this star is, in fact, a binary. The same principle applies to high contrast imaging of exoplanets - without AO, the angular resolution of large telescopes is significantly reduced, making the detection of companions nearly impossible. Thus, adaptive optics is indispensable for ground-based observations, particularly in the context of high contrast imaging of exoplanetary systems, where high angular resolution is crucial.

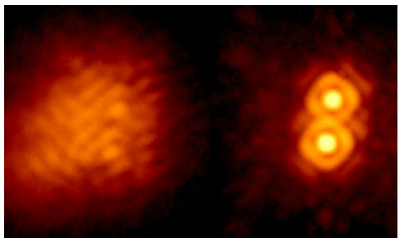


Figure 3.2: PSF of a binary star with (right) and without (left) an AO system. (Credits: Palomar Observatory / NASA-JPL)

3.2 Coronagraphs

The second technique used to enhance image quality focuses on increasing the achieved contrast between an exoplanet and its host star. Since the star is several orders of magnitude brighter than its companions, achieving high contrast imaging of exoplanets requires effectively enhancing this contrast. To address this issue, a coronagraph can be introduced into the optical system. A coronagraph is an optical device designed to block the majority of direct light from the central region of a field of view, allowing for the observation of faint surrounding objects. One might argue that using both an AO system and a coronagraph is redundant. However, AO systems are not without flaws. As mentioned earlier,

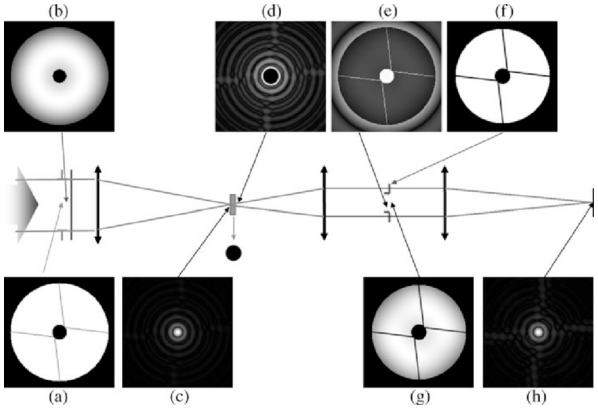


Figure 3.3: Principle of SPHERE Apodized Lyot Coronagraph: (a) entrance pupil, (b) apodizer, (c) PSF at the focus of the telescope, (d) PSF when the Lyot occulting coronagraphic mask is settled, (e) pupil image before the Lyot stop introduction, (f) Lyot stop, (g) Pupil image with the Lyot stop, (h) final coronagraphic PSF. (Credits: Guerri et al. [2011])

the PSF produced by such systems is not a perfect point but rather exhibits features like the rings visible in Figure 3.2, known as the Airy pattern, resulting from light diffraction through a circular aperture. These features can obscure potential companions if they happen to coincide with the rings. The principle of a coronagraph is illustrated in Figure 3.3. The specific coronagraph presented here is the one installed on the SPHERE instrument of the VLT and is called a Lyot Coronograph.

For high contrast imaging of exoplanets and protoplanetary disks, specialized types of coronagraphs can be employed, tailored to specific observational requirements. In the context of this work, part of the data is obtained from the NaCo instrument on the VLT, utilizing an Annular Groove Phase Mask (AGPM) coronagraph (Mawet et al. [2005]), a specific type of vortex coronagraph. This technology makes use of optical interference rather than physically blocking the starlight with a mask, and is illustrated in Figure 3.4.

3.3 Observation strategies

Despite the use of the previously mentioned techniques, residual speckles from the AO system reduces the achieved contrast between the star and a potential companion often preventing the detection of exoplanets. To address this, these techniques must be combined with observational strategies known as Differential Imaging (DI). This approach involves subtracting representative images, that capture the stellar contribution, from the science image cube, with the goal of revealing a faint companion. Several types of DI can be implemented. In this section, we focus on those most relevant to this project.

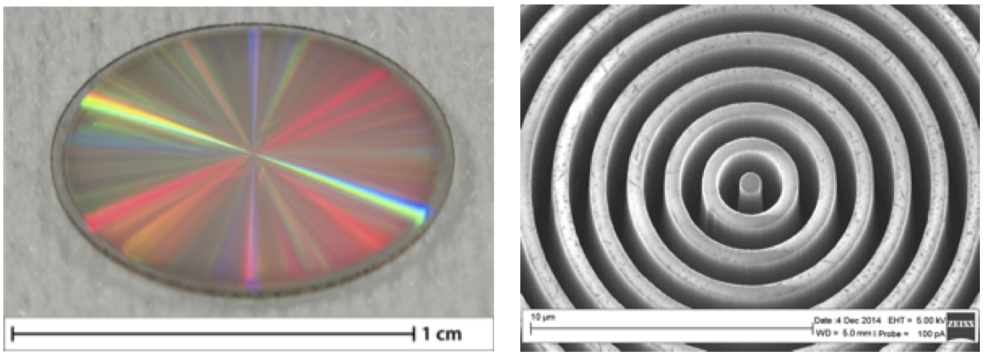


Figure 3.4: Photograph of an AGPM vortex coronagraph. *Left*: 1cm scale; *Right*: 10 µm scale using a SEM micrograph. (Credits: Absil et al. [2016])

3.3.1 Spectral Differential Imaging

Spectral Differential Imaging (SDI) leverages the fact that certain types of companions are spectrally distinguishable from the central star (Smith [1987]). Jovian planets, which are the primary exoplanets detectable via high contrast imaging, for example, exhibit strong CH_4 absorption bands that contrast with the spectral signature of the star. This is also true for other types of exoplanets, each exhibiting distinct spectral properties depending on the composition of their atmosphere.

Another approach leverages the fact that the diffraction pattern (visible as rings in the PSF) scales with wavelength: λ/D (Sparks & Ford [2002]). As the wavelength increases, the ring pattern expands outward from the central pixel, while the planet signal remains fixed in radial distance. The main advantage of this method is that it does not require long observations, as wavelength replaces time as the third dimension of the image cube. Consequently, PSF stability is only required over the short duration of the observation. However, a large wavelength coverage of the instrument is required and shorter wavelength see the quality of the AO system degraded.

3.3.2 Angular Differential Imaging

An alternative method, Angular Differential Imaging (ADI) (Marois et al. [2006]), avoids this limitation. By tracking the star's motion in the sky over a longer time while disabling the instrument's field derotator[2], a sequence of images is acquired with a rotating field of view. Indeed, by using the Earth's rotation, the sky appears to revolve around the North Celestial Pole, but if an instrument's field of view is fixed to a given star, the same phenomenon appears as the sky seems to rotate around that target star. While the starlight remains largely constant over time, the apparent motion caused by this rotation causes any potential exoplanets

[2]A field derotator is a component of the optical instrument that rotates the detector to compensate for sky rotation around the target during long-exposure observations.

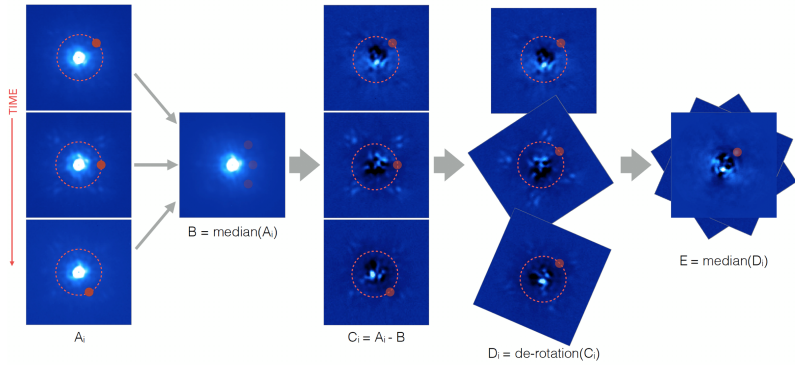


Figure 3.5: Median-ADI concept. A sequence of images A_i is obtained in pupil-tracking mode. Any putative planet would rotate throughout the images, along with the field of view, while the stellar halo and speckle pattern remains quasi-static. The median B of the sequence is subtracted to A_i to create a residual cube C_i. The residual cube is derotated (D_i) to align the field-of-view. The derotated residual cube is then stacked (typically with a median) to create a final image E. (Credits: Marois et al. [2006])

orbiting the target star to shift position in the image sequence. This relative movement enables the separation of the planetary signal from the quasi-static stellar light, halo and speckle pattern. The typical observation time required to achieve meaningful results is at least 30 minutes, depending on the planet's radial distance from the host star and the star's declination.

A diagram illustrating the ADI method is shown in Figure 3.5. The specific case depicted is known as *median-ADI*, in which the median of the image cube is subtracted from each frame in the sequence to isolate any potential planetary signal from the quasi-static stellar light.

This method, however, presents several disadvantages. In addition to requiring longer observation times - typically around 30 minutes to an hour, depending on the target's distance from the North celestial pole, and thus relying on stable sky conditions over extended periods - it also depends heavily on the radial separation between the planet and its host star. Specifically, the closer a planet is to its star, the greater the field rotation angle must be between successive images to avoid self-subtraction of the companion. If the companion is too close and the sky does not rotate sufficiently between frames, the planet may move less than a Full-Width at Half Maximum (FWHM). This increases the likelihood that the planet's signal will be included in the median (or any other reference quantity, depending on the variant of the method used), thereby diminishing its detectability.

One final comment can be made regarding ADI: within the scope of this thesis, which focuses on imaging protoplanetary disks and potential protoplanets embedded within their structures, ADI may produce unsatisfactory results (Milli

et al. [2012]). This is because the extended signal from a circumstellar disk is often captured and subsequently subtracted by standard algorithms during image processing. The degree of deformation in the final image of the disk depends on its morphology (Juillard et al. [2023]). Therefore, to accurately image disk features, alternative observational methods or combinations of techniques should be considered.

3.3.3 Reference-star Differential Imaging

The last observation method presented in this section is the Reference-star Differential Imaging (RDI) method (Xie et al. [2022]). Compared to ADI, this technique is better suited for directly imaging circumstellar disks around stars. It works by collecting stellar references through observations of calibration stars with no known circumstellar material. The reference library is typically planned in advance during the observation preparation process, with similar stars to the target observed on the same night to obtain the most correlated images possible. Additionally, a reference library can be constructed from the science images themselves, but this requires further post-processing to select the best reference images. This approach is employed in this thesis and is detailed in Section 5.1.

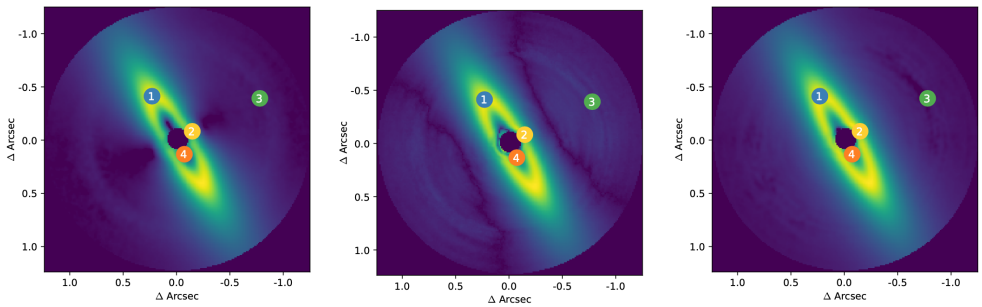
Compared to ADI, where the subtracted images are derived from the original science cube, in RDI the subtracted images come from reference image cubes. Since these reference stars are assumed to be devoid of circumstellar features, subtracting their images from the science cube should ideally remove all of the host star's starlight, leaving only the circumstellar features and potential companions of the target. This technique therefore avoids any self-subtraction of disk signals and potential companions.

However, RDI presents a disadvantage when applied to observations made with ground-based instruments. Due to the chaotic and unpredictable nature of atmospheric turbulence, it can be difficult to construct a library of reference images that are highly correlated with the science images. Nevertheless, as AO systems continue to improve over time, RDI is becoming increasingly reliable. Finally, this issue can be mitigated by building the reference library from archival data of the instrument, as will be done later in this work.

3.3.4 Combining strategies

As highlighted throughout this section on observation strategies, no single method is optimal on its own. Therefore, we will explore the combination of techniques, focusing particularly on ADI and RDI.

As already stated, ADI is most effective for imaging point-like sources with sufficient field rotation between images, but it struggles with extended structures like circumstellar disks. In contrast, RDI is well suited for imaging such extended sources without self-subtraction but requires observations of well-matched reference stars. RDI can therefore be less effective for challenging datasets, such as those involving faint disks or companions, or in cases with time-varying speckle



(a) ADI (b) RDI (c) ARDI

Figure 3.6: Best disk images using ADI and RDI on their own or together (ARDI). Four regions of importance have been represented: (1) the brightest part of the disk, (2) a region of the disk that is rotation invariant, (3) a noise region, and (4) the region near the coronagraphic region where the variation of the speckle field is higher. (Credits: Juillard et al. [2024])

patterns, especially when no appropriate reference star observations are available. In order to leverage the strengths of both methods, a novel approach presented in Juillard et al. [2024] is to consider a combination of RDI and ADI, referred to as ARDI.

Promising results have been obtained when applying ARDI to extended sources, with performance surpassing that of any individual strategy. This improvement is particularly significant when attempting to image planets embedded within a disk. In such cases, RDI often struggles due to the low contrast of the planetary signal, while ADI tends to distort or subtract the disk structure. By combining both approaches, ARDI effectively mitigates these limitations simultaneously.

Figure 3.6 presents a comparison of the post-processing results obtained using ADI, RDI, and ARDI. It is clear that the image produced with ARDI outperforms the other two: the disk structure appears less distorted, and there is significantly less noise propagating outside the disk region.

The quality of the resulting image is of course only partly due to the choice of the observing strategy. Indeed, another very important factor is the choice of post-processing algorithm to be used. The next section will be on this subject.

3.4 Post-processing algorithms

There exists a wide range of post-processing algorithms, each with their own strengths, weaknesses, computational efficiency, and resource requirements. The choice of post-processing algorithm is closely tied to the chosen observation method, as certain algorithms perform better with specific observation strategies. It is therefore crucial to understand both the underlying principles of the algorithms and the observation strategies in order to ensure an optimal match.

In this section, several algorithms will be introduced, with a particular focus

on PCA and I-PCA, which will be used in the context of this work.

3.4.1 Median-ADI

This method was briefly introduced in Section 3.3.2. Median-ADI is the simplest algorithm that can be used with the ADI technique. As illustrated in Figure 3.5, the median of the science cube is subtracted from each individual frame. Thanks to the ADI observation method, planetary signals can be revealed after derotation and stacking of the residual images.

This straightforward algorithm has the advantage of being relatively resource-efficient, requiring minimal computational power and processing time compared to more advanced techniques. However, as one might expect, the results are often unsatisfactory, especially for protoplanetary signals. The extended nature of disk emissions means that disks signals are frequently included in the median, leading to significant self-subtraction in the final image.

Some more advanced versions of this algorithm attempt to mitigate these issues by computing a different model PSF for each frame. One such variant, known as "smart median-ADI", selects the model PSF for a given image by computing the median only over the temporally closest frames that have sufficiently rotated from it, according to a set parallactic angle[3] (PA) threshold. This variant aims to improve the modeling of quasi-static speckles, which vary significantly over time and are often not well captured by a median computed over the entire cube.

3.4.2 Least-squares approximation using LOCI

This method builds on the concept of using PA thresholds but employs a more sophisticated model PSF. In the version of the algorithm introduced in Lafrenière et al. [2007], the PSF is constructed using a Locally Optimized Combination of Images (LOCI). This technique still incorporates a PA threshold by excluding frames that have not rotated sufficiently relative to the frame being processed.

The terms "linear combination of images" refer to the use of a library of reference frames to construct a model PSF through a linear combination of those frames, where the coefficients are chosen to minimize residual noise when subtracted from the target frame. The terms "locally optimized" mean that each science image is divided into multiple subsections and that the optimization is performed independently within each of these regions. This local optimization improves the quality of the PSF model.

[3] The parallactic angle refers to the angle of rotation between images caused by the apparent motion of the sky due to Earth's rotation. In the context of ADI datasets, each science cube must be accompanied by an array of parallactic angles corresponding to each frame. These angles are crucial for the derotation process, allowing all frames to be aligned to a common orientation before combining them to enhance planetary signals.

3.4.3 Local low-rank plus sparse plus gaussian-noise decomposition

The Local Low-rank plus Sparse plus Gaussian-noise Decomposition (LLSG) algorithm proposes a three-term decomposition of the ADI science cube into low-rank, sparse, and Gaussian noise components (Gomez Gonzalez et al. [2016]). The objective of this decomposition is to effectively isolate the planetary signal, which mostly remains in the sparse component, from the dominant starlight and quasi-static speckle noise.

By isolating the planetary signals in the sparse component, LLSG enhances the signal-to-noise ratio (SNR) of potential exoplanets, making faint companions easier to detect. The decomposition is achieved using low-rank approximation techniques inspired by methods from the machine learning literature.

3.4.4 Angular differential optimal exoplanet detection algorithm

The ANgular Differential OptiMal Exoplanet Detection Algorithm (ANDROMEDA) algorithm Cantalloube et al. [2015] is designed to detect point sources in an ADI science cube by subtracting pairs of images that are separated by a PA corresponding to approximately half the FWHM of the instrument. This targeted subtraction results in a characteristic residual pattern - a bright-positive lobe adjacent to a dark-negative lobe - representing a partial cancellation of the planetary signal. This effect is illustrated in Figure 3.7.

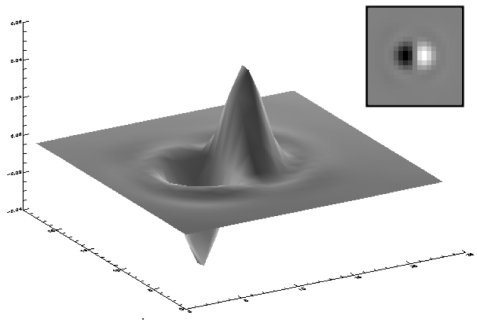


Figure 3.7: Subtraction of two images separated by a rotation of about half the FWHM as is done using the ANDROMEDA algorithm. (Credits: Cantalloube et al. [2015])

The resulting residual image is then processed using a maximum match-filter optimized for detecting this specific signature. By scanning the field of view, the algorithm evaluates the likelihood of a point source being present at each pixel location based on the strength of the match to the expected positive-negative pattern.

3.4.5 Patch covariances

The Patch Covariances (PACO) algorithm Flasseur et al. [2018] offers a statistically robust method for extracting faint planetary signals from background noise in high-contrast imaging data. It does so by constructing an empirical model of the residual starlight - comprising both speckles and background noise - through statistically grounded estimations of the noise. By leveraging these statistical models and the motion of the planet in the ADI frame, PACO can identify point sources that deviate significantly from the background noise distribution. Furthermore, the algorithm includes iterative fitting procedures to estimate the companion's flux, enabling both detection and characterization within a unified statistical framework.

3.4.6 Principal component analysis

The Principal Component Analysis (PCA) algorithm (Soummer et al. [2012]; Amara & Quanz [2012]) operates on a principle similar to the median-ADI algorithm presented in Section 3.4.1, namely through PSF subtraction. However, the model subtracted from the science cube is refined. Specifically, it is constructed by projecting each image onto the first PCs calculated from either a reference cube or the science cube itself. This algorithm, widely used in high-contrast imaging, is compatible with various observation strategies, including ADI, RDI, and SDI, as well as their combinations such as ARDI.

Concept

PCA (Pearson [1901]) is a linear dimensionality reduction technique widely used in data analysis. Its purpose is to reduce the number of dimensions in large datasets by transforming potentially correlated variables into a smaller set of uncorrelated variables called PCs. These PCs represent the directions along which the data varies the most. The first PC captures the largest variance, followed by the second, and so on. Each PC is orthogonal to the others, and for an n-dimensional dataset, up to n PCs can be computed.

Of course, constructing all n components does not reduce dimensionality; it merely results in a change of basis that retains all variance. In most applications, only the first few components are retained, as they typically explain most of the variance in the data. The first PC is a linear combination of the original variables that captures the greatest variance; the second explains the most variance remaining after accounting for the first, and so on. By selecting only the leading components, the dataset is effectively pruned of dimensions contributing less to its structure and variability.

Mathematical reasoning

Mathematically, PCs are derived from the covariance matrix of the data by computing its variance-weighted eigenvectors. Consider a dataset with K features,

$x_1, \ldots, x_K$, observed over N samples. A mean-subtracted data matrix $\mathbf{X}_m$ of size $K \times N$ is formed by subtracting the mean $\bar{x}_i$ from each sample: $\mathbf{x}_{i,m} = \mathbf{x}_i - \bar{x}_i$, with $\bar{x}_i = \frac{1}{n_{var}} \sum_i x_i$; n_{var} is the number of variables in $\mathbf{x}_i$. The covariance matrix is then given by:

$$\mathbf{C}_X = \frac{1}{N-1} \mathbf{X}_m \mathbf{X}_m^T$$

To compute the eigenvectors, Singular Value Decomposition (SVD) is used. SVD allows us to express any matrix $\mathbf{S}$ of size $N \times M$ as a product of three matrices:

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}}$$

where $\mathbf{U}$ is an $N \times M$ column-orthogonal matrix, $\mathbf{\Sigma}$ is an $M \times M$ diagonal matrix with non-negative singular values, and $\mathbf{V}^{\mathsf{T}}$ is the transpose of an $M \times M$ orthogonal matrix. The diagonal elements of $\mathbf{\Sigma}$ are the singular values of $\mathbf{S}$.

Since the covariance matrix $\mathbf{C}_X$ is symmetric and square, it can always be diagonalized, allowing the SVD to be rewritten as:

$$\mathbf{C}_X = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{-1}$$

The columns of $\mathbf{U}$ are the orthogonal eigenvectors of the covariance matrix, representing the directions of the PCs. The eigenvector corresponding to the largest eigenvalue is the first PC, as it accounts for the most variance in the dataset. The number of PCs to retain depends on the specific application and the desired balance between dimensionality reduction and information preservation.

PCA in high contrast imaging

In the context of high contrast imaging, the primary goal of PCA is to model and subtract the stellar PSF from the science cube in order to reveal faint circumstellar or companion signals. This is achieved by reconstructing the PSF using a limited number of the first PCs, which capture the dominant spatial structures in the PSF of the star while ideally excluding signals from companions or circumstellar disks.

However, using PCA for high contrast imaging presents several challenges. The post-processing procedure can introduce deformations that may lead to false claims of exoplanet detections. These deformations primarily arise from two major effects: self-subtraction and over-subtraction (Pueyo [2016]).

Self-subtraction occurs when part of the astrophysical signal is included in the first few PCs, due to the use of reference frames that contain the signal itself to construct the PSF model. Conversely, over-subtraction happens when the subtracted PSF is computed from reference images that do not contain any astrophysical signal. In this case, the resulting PSF may be too bright compared to that of the target observation, leading to excessive subtraction that removes part of the signal of interest. These two effects are illustrated in Figures 3.8 and 3.9 for self- and over-subtraction respectively.

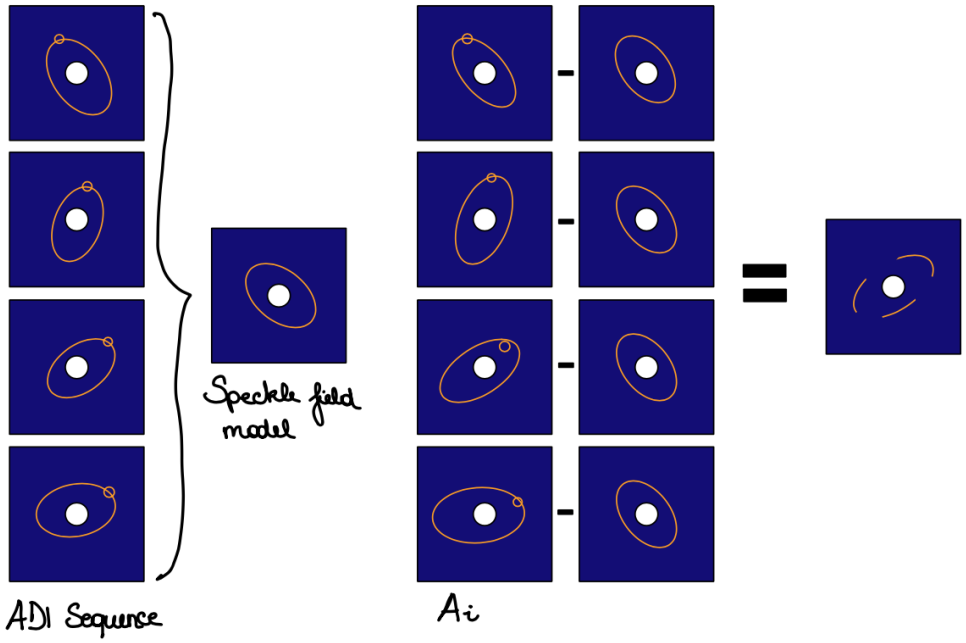


Figure 3.8: Illustration of the principle of self-subtraction.

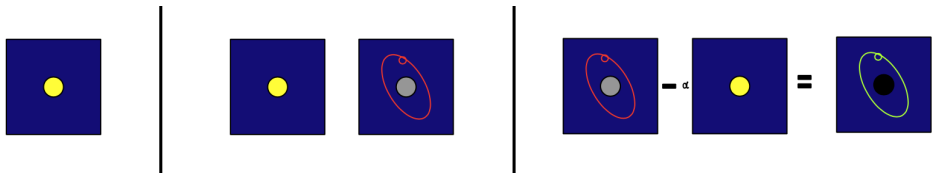


Figure 3.9: Illustration of the principle of self-subtraction.

The number of PCs retained is therefore a critical parameter, strongly dependent on both the dataset and the nature of the target signal. This sensitivity to the number of components makes PCA-based reduction a delicate balance: using too few components can leave significant residual starlight, reducing image contrast, while using too many can capture part of the significant signal in the subtracted PSF.

As already stated, the PCA algorithm is compatible with a variety of observing strategies. This thesis focuses on a combination of ADI and RDI, for which the behaviour in combination of PCA is detailed below:

- With an ADI observing strategy, the PCs used to construct the subtracted PSF are derived directly from the science image cube, as illustrated in Figure 3.10. This approach carries a significant risk of self-subtraction, particularly when the field rotation between frames is insufficient. Protoplanetary disk signals are especially susceptible to this issue - even when only a small number of PCs are used. Due to their often rotationally invariant morphology, these signals can appear stationary across frames and

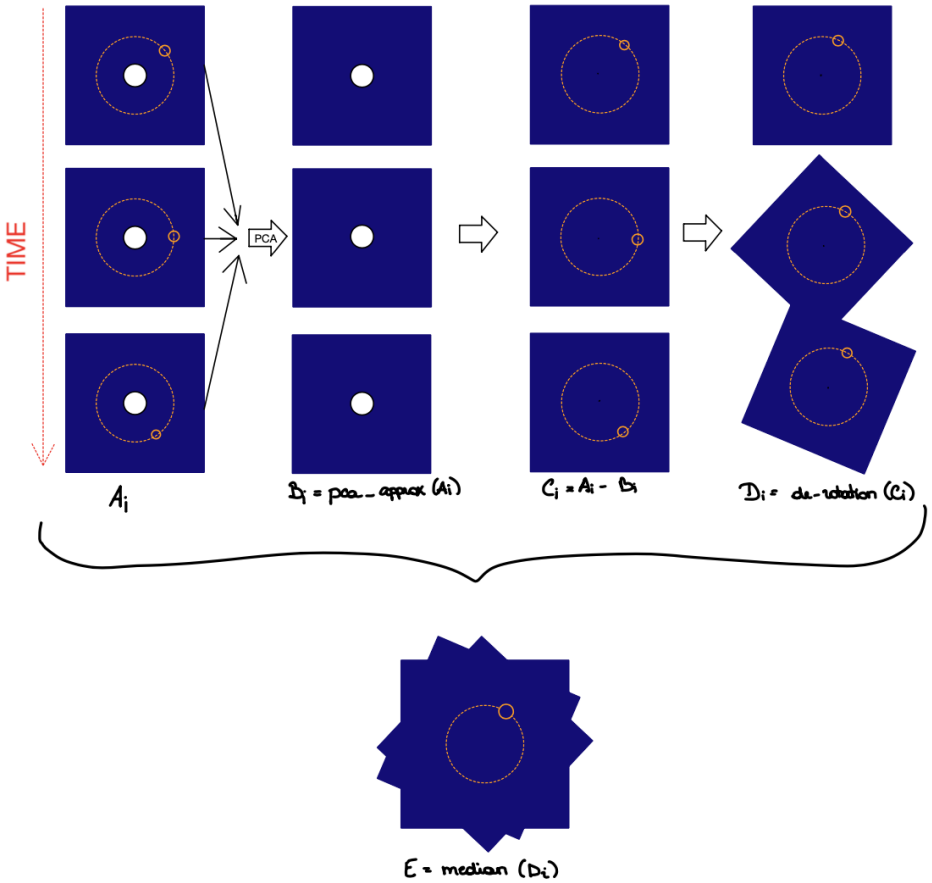


Figure 3.10: PSF subtraction using PCA-ADI. Similar to median-ADI except the PSF is computed individually for each science frame using the PCA algorithm. ϕ_j are the PCs, columns of the matrix **U**. (Credits: Soummer et al. [2012]; Amara & Quanz [2012])

may therefore be partially captured by the PCs, leading to their partial removal during PSF subtraction.

- An RDI strategy avoids including the signal of interest in the subtracted PSF, since the PCs are computed from a reference image cube that does not contain the target (e.g., exoplanet or disk). However, over-subtraction can still occur during the subtraction step. The subtracted PSF is built from reference images that do not contain any signal of interest, which means that, in the case of protoplanetary disks, the starlight is not partially obscured by the disk. As a result, subtracting this reference PSF can lead to negative areas in the final image due to over-subtraction of the starlight. In such cases, part of the astrophysical signal can still be unintentionally removed.

The choice of observing strategy depends on the characteristics of the target and the dataset. If the science image cube shows a sufficiently large variation in PA between successive frames, but the correlation with a reference cube is low, then ADI is preferred. Conversely, if the variation in PA is limited but the correlation between the science and reference image cubes is high, then RDI becomes more effective. Finally, when both a large PA variation and high correlation with the reference cube are present, a combined approach - ARDI - can be employed to maximize the benefits of both strategies.

3.4.7 Iterative principal component analysis

The Iterative Principal Component Analysis (IPCA) algorithm is a more recent adaptation of PCA for high contrast imaging and has proven significantly more effective than traditional PCA for detecting faint signals in young stellar systems - such as protoplanets and protoplanetary disks (Pairet et al. [2021]; Juillard et al. [2023]). It has been shown that protoplanetary disks are often partially captured in the low-order PCs due to the rotation-invariant nature of their signal, causing them to be incorporated into the subtracted PSF and thus partially removed from the final image (Milli et al. [2012]; Christiaens et al. [2019b]).

To mitigate this issue, the I-PCA algorithm iteratively refines the estimate of the stellar speckle field by progressively subtracting the disk signal from the subtracted PSF at each step. This prevents the signal of interest from being incorporated into the PCs. The core concept is illustrated in Figure 3.11.

The process begins by computing a first approximation of the PSF using a fixed number of principal components. This initial subtraction leaves behind residual astrophysical signal, which is then used to refine the PSF estimate in subsequent iterations. At each step, a portion of the residual signal remaining after subtracting the PSF built from the initial PCs is excluded from the residual cube. This prevents the algorithm from incorporating that signal into the subtracted PSF when higher-order PCs are added in subsequent iterations. A threshold is provided to the iterative algorithm to define the minimum intensity of the signal that is considered significant and excluded from the residual cube.

Naturally, the initial number of principal components chosen is crucial, as it determines which part of the signal will be preserved and refined. A poorly chosen number could either fail to remove enough stellar residuals or over-subtract the target signal. Therefore, it must be selected carefully depending on the target and dataset characteristics.

The I-PCA algorithm will be employed in this thesis to directly image protoplanetary disk structures around promising stellar systems.

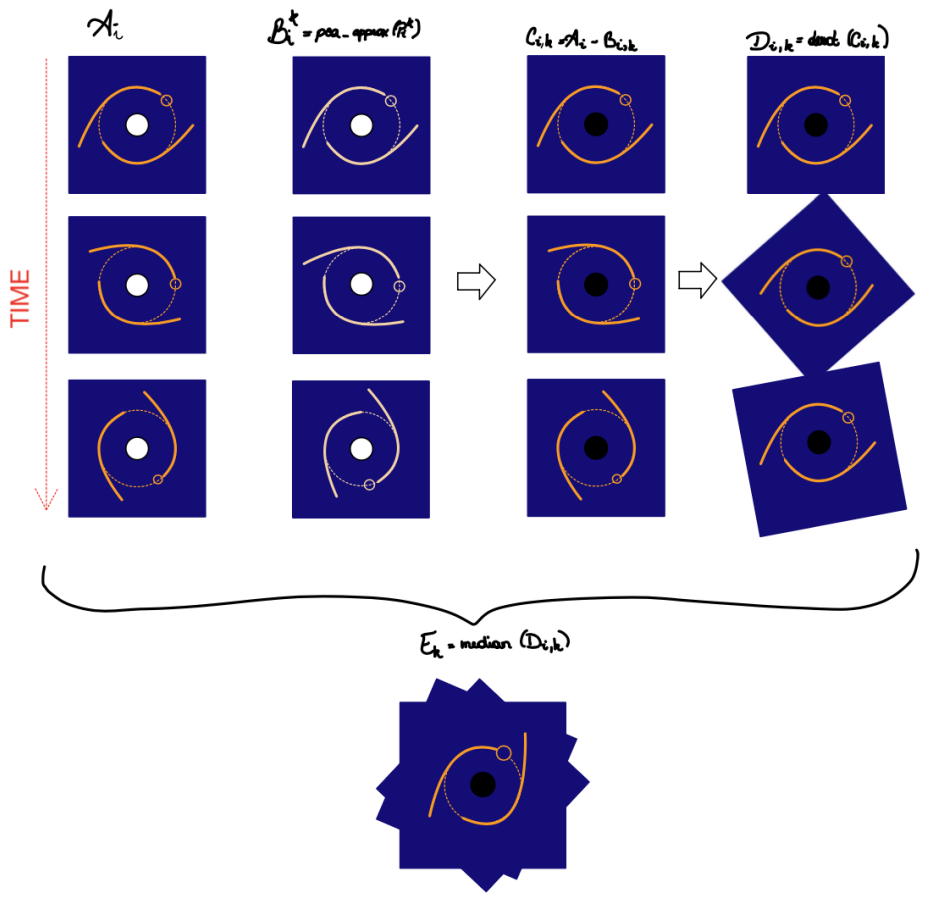


Figure 3.11: PSF subtraction using I-PCA-ADI. The principle is similar to standard PCA-ADI, with the addition of an iterative signal isolation process. A protoplanetary disk, shown in red, is progressively extracted from the subtracted PSF to reduce self-subtraction. (Credits: Pairet et al. [2018, 2021]; Juillard et al. [2023]; Christiaens et al. [2024])

Chapter 4

Research objectives

This chapter outlines the scope of the master's thesis by presenting its main objectives, the dataset employed, and the selected targets of interest. It provides the contextual background necessary to understand the motivation behind the study and the criteria guiding target selection. These elements together establish the foundation for the work presented in the subsequent sections.

4.1 Main Objectives

As discussed in Section 3, direct imaging of exoplanets and protoplanetary disks relies heavily on various techniques to enhance contrast and angular resolution. While many of these methods - such as adaptive optics, coronagraphs, and specific observing strategies - are applied during the observation phase, their benefits cannot be retrospectively applied to archival data. The only technique capable of improving archival observations is post-processing.

The primary objective of this master's thesis is to attempt the direct imaging of young exoplanets and their surrounding disk structures using archival data obtained from a now-decommissioned instrument of the Very Large Telescope (VLT) in Chile. This is achieved by leveraging the new I-PCA algorithms introduced in Section 3.4.7.

4.2 The POPCORN Dataset

The dataset used in this work, referred to as the POPCORN dataset, is an archival collection of observations obtained with the VLT instrument NaCo. The observations span from as early as November 2008 until the decommissioning of NaCo in September 2019. These archival data have been preprocessed by a team at ETH Zurich[1] using the `PynPoint` pipeline (Stolker et al. [2019]; Bonse et al. [2025]).

[1]M. J. Bonse et al. - Institute for Particle Physics and Astrophysics, ETH Zurich, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland

4.2.1 NaCo Observation Modes

The NaCo L'-band ($\lambda = 3.8\,\mu m$) observations used in this work were conducted using two distinct observing modes: an AGPM coronagraph and four-point dithering pattern.

Coronagraphic data

The coronagraph employs an AGPM optical vortex which, as described in Section 3.2, uses optical phase manipulation to destructively interfere with the on-axis starlight. This significantly attenuates the central stellar flux, thereby enhancing the detectability of nearby planetary companions (Mawet et al. [2013]).

Dither data

The "dither" data is obtained using a dithering strategy. This technique consists of observing the star at different positions on the detector, rather than keeping it centered. The main purpose of dithering is to help distinguish the stellar signal - which moves across the field - from the instrumental background noise, which remains fixed on the detector.

4.2.2 Preprocessing Pipeline

The preprocessing of the archival data is performed using the `PynPoint` pipeline (Stolker et al. [2019]). Each observation is processed independently following these steps:

1. The image cube undergoes standard calibrations, including dark and flat-field corrections to normalize pixel intensities. Dead pixels are identified using a bad pixel map and corrected through interpolation.

2. Image centering is applied: for coronagraphic (AGPM) data, centering is done with respect to the vortex center; for dither data, the image is centered around the brightest pixel.

3. A frame selection procedure is performed using a PCA-based algorithm. Each image is projected onto the set of principal components computed from the image cube, and only the frames most similar to the overall cube are retained. This is explained in more details in Section 5.1.1.

Chapter 5

Methodology

5.1 Reference-star selection

The first step in performing any kind of RDI is to identify the best references for each source of interest.

We begin by building a master cube of reference observations by selecting 10 representative images from each relevant reference. A star observation is considered relevant for a given target if it shares the same observation mode (cf. Section 4.2.1) and does not host a protoplanetary disk or bright companion.

To select the 10 representative images for each reference, the image cube is divided into 10 sections of equal frame count. In each section, the 3 central frames are selected and compared to the median of the full cube using a Pearson correlation coefficient (explained in detail below), with a threshold set at 0.95. If none of the three frames exceed this threshold relative to the median, the search is expanded to the 5 central frames of the section. This process is repeated recursively until one sufficiently correlated frame is found for each of the 10 sections of the image cube.

5.1.1 Selection models

Once the two master reference cubes are constructed - one for each observation mode - the next step is to identify the most suitable reference frames for each target of interest. To achieve this, four different selection methods were evaluated in the scope of this work. In each case, the k^{th} potential reference frame, denoted $X_i^{(k)}$, is compared to the temporal median of the science frames, M. The methods used for this comparison are described below:

- **Principal Component Analysis**: As described in Section 3.4.6, Principal Component Analysis (PCA) can be employed to identify suitable reference frames by projecting the median of the science cube onto its first principal components. In this case, the first five components were considered. The same projection is applied to each frame in the reference master cube.

To determine the most appropriate reference frames, the distance between the projection coefficients of each reference frame and those of the median science frame is computed for every principal component. The distance metric used is the sum of absolute differences (Manhattan distance), which provides a measure of how far each reference frame deviates from the median of the science data in the reduced PCA space.

The total distance for the k$^{\text{th}}$ reference frame is calculated as a weighted sum of the individual distances across the five components. The weighting decreases exponentially, assigning greater importance to the lower-order components, as shown below:

$$dist^{(k)}_{PCA,tot} = \sum_{p=1}^{5} 0.8^{p-1} \cdot dist^{(k)}_{PCA,p}$$

where $dist_{p,k}$ denotes the distance corresponding to the p$^{\text{th}}$ principal component for the k$^{\text{th}}$ reference frame. This formulation ensures that discrepancies in the first principal components contribute more significantly to the total distance than those in higher-order components.

The reference cubes that contain the 10 frames with the lowest PCA distance are the best references.

- **Mean Squared Error**: The Mean Squared Error (MSE) method calculates the distance between each frame in the reference master cube and the median of the science cube. This distance is computed by taking the sum over all pixels of the squared differences between each reference frame and the mean science frame, and then dividing the result by the total number of pixels N_{px} in the frame. The resulting distance for the k$^{\text{th}}$ reference frame, denoted as $dist_i$, is given by:

$$dist^{(k)}_{MSE} = \frac{1}{N_{px}} \sum_{i=1}^{N_{px}} \left(M_i - X_i^{(k)} \right)^2$$

- **Pearson Correlation Coefficient**: The Pearson Correlation Coefficient (PCC, Pearson [1895]) is a statistical measure of the linear correlation between two datasets. It represents a normalized form of covariance, with values ranging between -1 and 1. In the context of this work, the distance based on the PCC method is defined as (Ruane et al. [2019]):

$$dist^{(k)}_{PCC} = \frac{cov\left(X_i^{(k)}, M_i \right)}{std\left(X_i^{(k)} \right) \cdot std\left(M_i \right)}$$

where $cov(\cdot, \cdot)$ and $std(\cdot)$ denote the covariance and standard deviation,

respectively, defined as:

$$cov(P,Q) = \frac{1}{N_{px}-1}\sum_{i=1}^{N_{px}}\left(P_i-\bar{P}\right)\left(Q_i-\bar{Q}\right)$$

$$std(P) = \sqrt{\frac{1}{N_{px}-1}\sum_{i=1}^{N_{px}}\left(P_i-\bar{P}\right)^2}$$

- **Structural Similarity Index Measure**: The Structural Similarity Index Measure (SSIM, Wang et al. [2004]) is a perceptually motivated metric that assesses image quality by analyzing variations in structural content. It relies on the notion that pixels exhibit strong spatial dependencies when located near one another. SSIM incorporates three components: luminance, contrast, and structure. The luminance term functions similarly to the MSE, while the structural component is akin to the PCC. Consequently, SSIM can be interpreted as a combination of the MSE and PCC approaches. The distance according to this method is given by:

$$dist_{SSIM}^{(k)} = \frac{1}{N_{px}}\sum_{i=1}^{N_{px}} L_i^{(k)} C_i^{(k)} S_i^{(k)}$$

where L, C and S are the luminance, contrast, and structural terms respectively, computed over a FWHM×FWHM square centered on a pixel i and are given by:

$$L_i = \frac{2\bar{X}\bar{M}+c_1}{\bar{X}^2+\bar{M}^2+c_1}$$

$$C_i = \frac{2std(X)std(M)+c_2}{std(X)^2+std(M)^2+c_2}$$

$$S_i = \frac{cov(X,M)+c_3}{std(X)std(M)+c_3}$$

with c_1, c_2 and c_3 are small constants added to prevent instability when the denominator reaches values close to zero.

For each method, every reference cube in the master cube is assigned a score based on the best distance among its 10 representative frames. The 10 reference cubes with the most favorable scores are then selected as references for the provided science cube. It is important to note that for the PCA and MSE methods, lower distance values indicate better matches, whereas for the PCC and SSIM methods, higher values correspond to better similarity.

5.1.2 Models Comparison

With four different methods available for finding the best references and no a priori knowledge of which performs best, a comparative analysis is necessary to

identify the most effective approach. This comparison follows a methodology similar to that presented in Sanghi et al. [2024], in which fake companions are injected into observations and the SNR is evaluated at their locations. The method yielding the highest SNR is deemed the most effective.

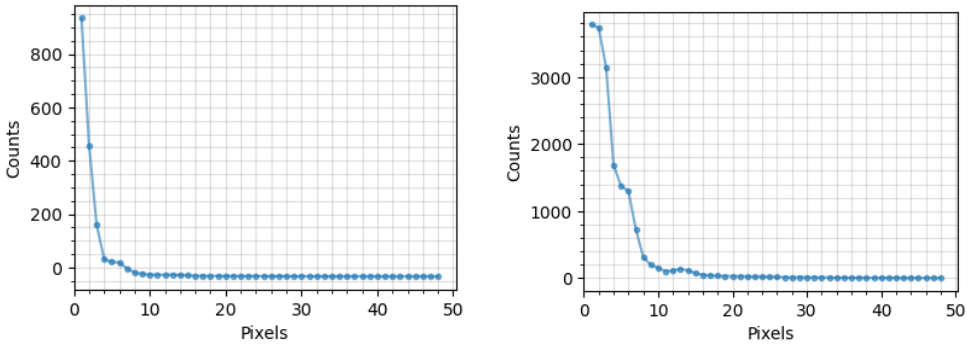
Identification of Suitable Stars

The first step is to identify suitable observations for fake companion injection. Observations are considered suitable - hereafter referred to as "blank stars" - if they do not exhibit known bright companions or detectable protoplanetary structures. This ensures that any detected signal can be confidently attributed to the injected fake companion.

Companion Flux Estimation

The flux of the injected fake companion must be realistically estimated from the observational data. To correctly scale the companion's flux, it is necessary to base it on the flux of the host star. The procedure for estimating the stellar flux depends on the observation mode - whether it was acquired in dither or coronagraphic (AGPM) mode.

For dither observations, the stellar flux can often be measured directly from the image cube, as no coronagraph is used to suppress the central signal. However, some observations are affected by saturation, which distorts the recorded flux and renders it unreliable. To identify such cases, a manual inspection was carried out for all 101 potentially affected dither observations. This involved analyzing the average radial profile of each image cube to detect anomalies indicative of saturation. A comparison between saturated and non-saturated observations is presented in Figure 5.1. Of the 101 dither observations that could potentially



(a) Non-saturated observation (b) Saturated observation

Figure 5.1: Comparison between a non-saturated (left) and a saturated (right) dither observation. The saturated observation shows similar flux at the center and adjacent pixels, while a non-saturated observation displays a steep flux drop-off near the center, as expected.

be saturated, 43 are identified as saturated and excluded from the blank star pool due to the inability to accurately compute their flux.

For the remaining dither observations, a corresponding shorter observation is available for flux estimation. If the detector is saturated during the long science exposure, the short observation - taken with an attenuation factor of 80 - avoids saturation. The true stellar flux can then be recovered by multiplying the measured flux in the short observation by the inverse of this attenuation factor.

For AGPM coronagraphic observations, direct flux estimation from the main science image is not feasible because of starlight suppression by the coronagraph. Instead, flux must, once again, be estimated from secondary shorter observations, in which the star is placed off-center to avoid the attenuation caused by the coronagraph and allow for an unobstructed measurement. In some cases, the exposure time of these secondary observations differs from that of the main science frames to prevent saturation, and this must be taken into account when computing the stellar flux.

Of the 736 available observations in the POPCORN dataset, only 9 AGPM observations are unusable due to the absence of a non-coronagraphic companion sequence. In addition to the 43 saturated dither observations, 48 AGPM observations that include multiple short sequences are excluded to simplify the analysis and avoid the need for further selection criteria. This still leaves a largely sufficient number of blank stars across both observing modes for robust model comparison.

Fake companion insertion

Once the suitable blank stars are identified, 50 are selected for each of the two observation modes to create a sufficiently large and statistically meaningful sample. Fake companions are then inserted into each of these blank star datasets. To simulate various observational scenarios, the parameters for the injected companions are varied across a range of values.

First, the angular position θ of the fake companion is varied from 0° to 270° in increments of 90°, resulting in four different angular positions around the star. For each of these positions, a set of radial distances is determined based on contrast curves, which are computed using the previously estimated stellar flux. A 5σ contrast threshold is used, as it is widely accepted as a reliable benchmark for exoplanet detection in direct imaging.

The contrast curves provide a set of radial distances starting from a FWHM from the center of the image, increasing by one pixel per contrast level beyond the first. Radial distances are chosen to closely match multiples of the FWHM. For instance, if the FWHM is 4.16 pixels, the closest practical distances (from the image center) used for fake companion insertion are:

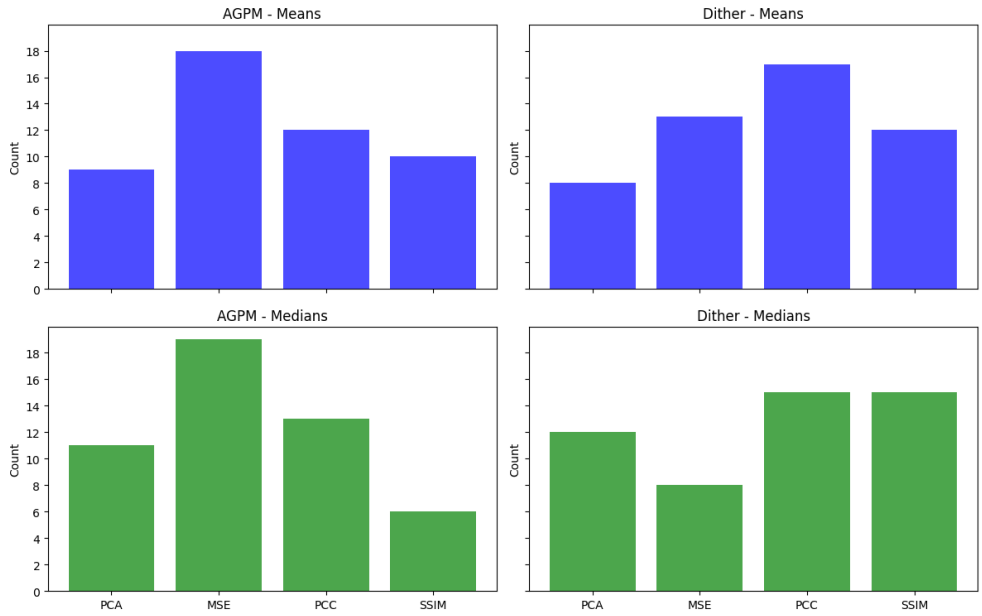


Figure 5.2: Comparison of the four reference models by counting the number of times each methods returned the best SNR detection value for a given target. Two statistical values are analyzed; the mean (top) and the median (bottom) of the SNR values over the 16 injected fake companions. The process is done for both observation modes; the coronagraphic AGPM observation mode (left) and the dither observation mode (left).

$$\begin{bmatrix} 2 \cdot \text{FWHM} = 8.32 \text{ px} \approx 8.16 \text{ px}, \\ 3 \cdot \text{FWHM} = 12.48 \text{ px} \approx 12.16 \text{ px}, \\ 4 \cdot \text{FWHM} = 16.64 \text{ px} \approx 16.16 \text{ px}, \\ 5 \cdot \text{FWHM} = 20.80 \text{ px} \approx 21.16 \text{ px} \end{bmatrix}$$

SNR computation

Finally, after inserting a fake companion in the image cube of the given blank star, a SNR of the fake companion is computed using a classic PCA-RDI algorithm. RDI is done using the reference stars provided by one of the fours methods cited above. The method that returns the best SNR for the most fake companion positions and most blank stars will be considered the best method for identifying reference stars. Figure 5.2 presents the results of that analysis. As we can see, for the AGPM observation mode, the MSE method seems to be more effective than the others at finding the best references. However, for the dither observations, it is less clear which method is best. The PCC method seems to be a bit more effective with respect to the mean SNR values returned but appears as effective as the SSIM method with respect to the median values.

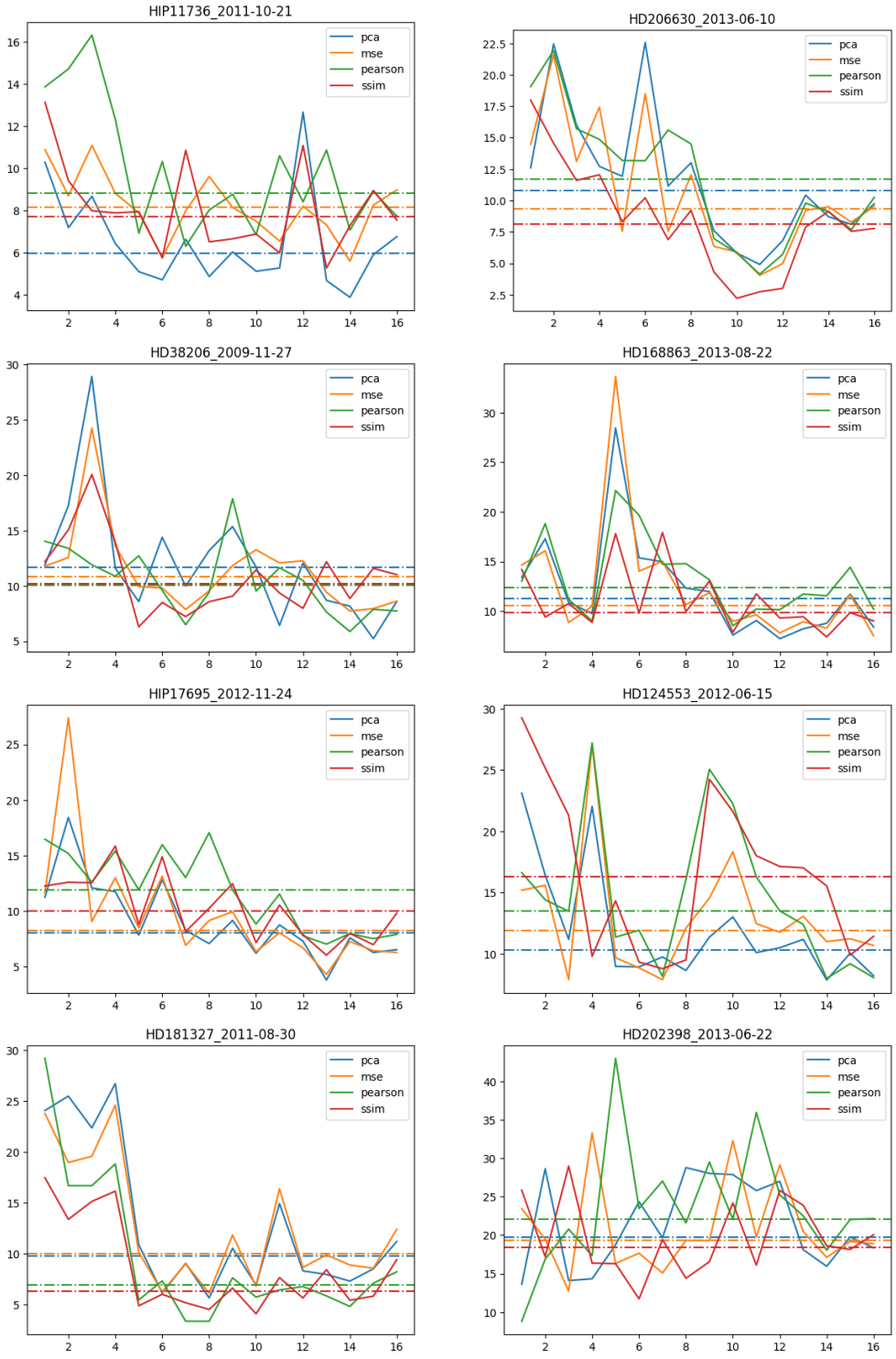


Figure 5.3: SNR variation for different fake companion insertions on dither data using the 4 different reference identification models. The fake companions are sorted by radial distance; e.g. the first 4 on the x-axis represent the closest inserted companions at the 4 different angles.

Finally, we can take a look at a few SNR graphs presented in Figure 5.3 and how they vary with the different fake companion injected. We can see that for most blank stars, no single method is better for all fake companions injected hence why it is hard to distinguish a best method. Despite all this, the graphs presented in Figure 5.2 still indicate slightly better references from the MSE model for AGPM data and from the PCC model for dither data, so the references from these models will be the ones used for the I-PCA-ARDI imaging.

5.2 Pre-processing

The first step before proceeding to the post-processing involves pruning the science image cube of potentially corrupted or misaligned frames, referred to as "bad frames". To identify these, the median image of the cube is computed, and the Pearson correlation coefficient is calculated between each frame and this mean. Frames with a correlation below 90% are considered poor quality and are discarded. An example of this process is illustrated in Figure 5.4 for the target HD 169142. This figure clearly demonstrates that, although the science cube

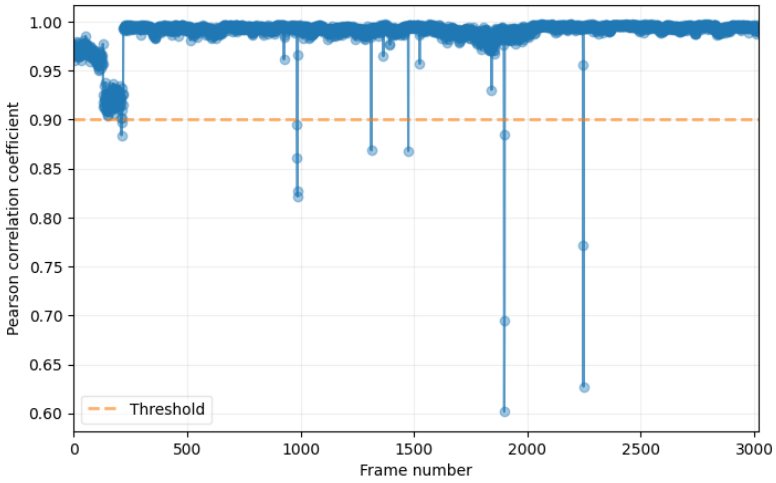


Figure 5.4: Bad frame binning for the target HD 169142 observed on May 19th, 2017. The horizontal line indicates the 90% correlation threshold, below which frames are excluded.

occasionally contains a few outlier frames that must be removed, the majority of the frames exhibit high quality and are retained for further analysis.

A similar procedure is applied to the reference image cube, which is built by stacking the cubes from the 10 selected reference observations. Since we are performing ARDI, it is crucial to match the number of reference frames to the number of science frames. This balance ensures the algorithm leverages both ADI and RDI components optimally. Using more reference frames than science frames biases the algorithm toward RDI, whereas fewer reference frames favor ADI.

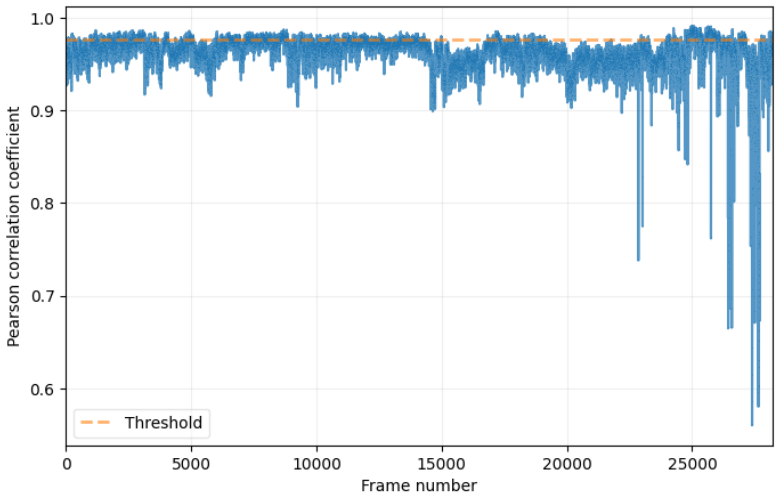


Figure 5.5: Bad frame trimming for the reference cube for target PDS 70, observed on July 9th, 2019. The horizontal line indicates the minimum correlation value among the top N most correlated frames, where N equals the number of science frames.

To select the most suitable reference frames, we compute the correlation between each reference image and the mean of the science cube. If the science cube contains N images, the N most correlated reference frames are selected for final processing. This ensures a well-matched reference set, and the correlation values can be visualized to confirm the selection quality. Figure 5.5 shows this for the same PDS 70 observation used in Figure 5.4. As shown in the figure, the selected reference frames exhibit a high correlation with the science cube, indicating that the reference selection process performed well.

5.3 I-PCA-ARDI

5.3.1 Optimizing the initial number of principal components

As previously mentioned, the choice of the initial PC used to start the iterative process is critical. If this value is set too high, a portion of the circumstellar signal may already have been captured by the PCs and subsequently subtracted from the image cube, making it unrecoverable in later iterations. Conversely, starting with too low a number of PCs can leave behind residual stellar light that was not fully captured, which may then be preserved across iterations and ultimately drown out the signal of interest.

To determine the optimal initial number of PCs, a standard PCA reduction is first performed across the first N PCs[1]. The resulting images are then manually

[1]In most cases, the optimal result is found before the 20th PC. Therefore, the initial PCA scan is typically performed from PC1 to PC25.

inspected to identify which one offers the most promising balance between stellar subtraction and signal preservation. This selected PC count serves as the starting point for the subsequent iterative process.

5.3.2 Significant signal threshold

To conserve significant signals at each iteration of the PCA process, we must determine an appropriate threshold above which the signal is identified as significant by the algorithm. This is achieved using a Standardized Trajectory Intensity Mean (STIM) map (Pairet et al. [2019]), a detection map derived from PCA residuals. Detection maps assign to each pixel a value representing the likelihood that a planetary signal is present at that location - effectively distinguishing between noise and potential astrophysical signals.

The STIM map, in particular, is computed directly from the residual cube output by the PCA. It provides a statistically grounded way of interpreting the residuals by highlighting pixels that deviate significantly from the noise distribution. This makes it especially suitable for defining a dynamic threshold for identifying meaningful signals during the iterative process. By scaling the detection threshold according to the STIM map, we ensure that only statistically significant features are retained for further iterations. For this work, a new two-dimensional STIM map was developed to better characterize the radial variation of the noise across the image.

5.3.3 Final considerations

With the methodology now fully outlined, we can turn our attention to the results obtained for selected targets of interest. As discussed throughout this section, numerous parameters can be optimized to produce the best possible final images using the I-PCA-ARDI approach. These include the choice of the initial and final number of PCs, the threshold applied to identify significant signals in the PCA residual cube, and the number of iterations performed per PC. The latter must be chosen carefully - large enough to capture most of the significant circumstellar signal, yet not so high as to reintroduce noise and speckle residuals into the final image.

Chapter 6

Targets of interest

Having set all the necessary components needed to perform an I-PCA-ARDI processing of observation datasets, we can now take a look at the result on specific targets. As already stated, the targets for this thesis are young stellar systems harboring circumstellar features. To that extent, the first dataset that is presented in this section is the famous PDS 70, the only system where protoplanets have been detected and confirmed at the moment. The objective is to try out the algorithm on a system that is well known to see how much of the detected features from the literature can be imaged. After that first system, other promising young stars are imaged using the NaCo data archive, most of which are targets of claims for potentials protoplanets orbiting the host star. For each target, the star and its circumstellar features will be described using the literature.

6.1 PDS 70

PDS 70 is a young (5.4 $\pm$ 1 Myr) K7 star located at a distance of 113.43 $\pm$ 0.52 pc. It is estimated to weight 0.76 $\pm$ 0.02 Solar mass (Müller et al. [2018]). The existance of a circumstellar disk around the star was first proposed in 2004 (Metchev et al. [2004]) and later resolved in 2006 (Riaud et al. [2006]). From there, plenty of other observations in a variety of wavelengths were performed to better characterize the disk and its structure. In total, it is estimated to have a radius of about 140 AU, and a large radial gap between 15 and 60 AU from the star.

The first protoplanet in the system, PDS 70 b, was detected in 2018 (Keppler et al. [2018]) within the disk cavity. This first detected companion is estimated to weight between 5 to 9 Jupiter masses. Later observations confirmed the planet and estimated its semi-major axis to be around 22 AU, its effective temperature between 1000 and 1600 K and its radius ranging from 1.4 to 3.5 the radius of Jupiter (Müller et al. [2018]). This unusually large radius might be due to the presence of circumplanetary material around the protoplanet which might explain its very red spectrum (Christiaens et al. [2019a,b]).

A second protoplanet, PDS 70 c, was detected in 2019 (Haffert et al. [2019]) closer to the outer edge of the gap in the disk. Due to its position, it is harder to estimate its characteristics as its signal mixes with the disk signal. It however estimated to weight between 4 to 12 Jupiter masses and to be situated at 34.5 $\pm$ 2 AU from the host star (Haffert et al. [2019]).

6.2 HD 36112

HD 36112, also known as MWC 758, is a young (3.5 $\pm$ 2 Myr) located at a distance of 151 $\pm$ 9 pc. It is estimated to weight 2 $\pm$ 0.2 Solar masses (Meeus et al. [2012]; Isella et al. [2010]). The star is surrounded by a protoplanetary disk with a large cavity from the star up to approximately 50 AU. The disk also has a spiral structure with a near-symmetric two-arms spiral between 30 and 75 AU from the star. Dong et al. [2015] have hypothesised that the presence of this spiral structure could be due to a companion outside the disk - at a radius close to 90 AU - weighing multiple times the mass of Jupiter. The hypothesis has been later reinforced by computations from Ren et al. [2018]. However follow-up studies failed to detect any companion to HD 36112.

6.3 HD 135344B

HD 135344B, also known as SAO 206462, is a young (9 $\pm$ 2 Myr) star located at a distance of 142 $\pm$ 27 pc. It is part of the binary system HD 135344 which include it and the primary star, HD 135344 (SAO 206463). The two stars are gravitationally bound and separated by 3000 $\pm$ 600 AU. HD 135344B is estimated to weight 1.7 $\pm$ 0.2 Solar masses (Müller et al. [2011]). The star is surrounded by a protoplanetary disk separated into two rings. The inner ring, located at 0.05 AU to 1.8 AU from the star, is highly inclined compared to the outer disk and the star might still be accreting some mass from it. The outer ring, located at 45 AU and spanning up to 200 AU from the star, is seen almost pole-on from Earth and is arranged in a two-arm spiral pattern. Between the two rings, a large gap spans from 1.8 AU to 45 AU (Müller et al. [2011]; Casassus et al. [2021]).

Once again, the presence of a gap in the disk is hypothesized to be caused by a sub-stellar companion having cleared that gap (Grady et al. [2009]). However, the presence of this companion has not yet been ruled out or confirmed. Recently, a second candidate companion (Companion Candidate 1 - CC1) has been detected using JWST observations (Cugno et al. [2024]), and is thought to orbit HD 135344B at a distance of $\sim$300 AU and, if it is confirmed, is probably the driver for the eastern spiral.

6.4 HD 169142

HD 169142 is a young (6^{+6}_{-3} Myr) star located at a approximate distance of 145 pc. It is estimated to weight approximately 1.65 Solar masses (Grady et al. [2007]; Quanz et al. [2013]). The circumstellar disk around this star is azimuthally symmetrical and is composed of two rings located at 25 and 60 AU from the star. The inner cavity up to 22 AU as well as the 38 AU annular gap between the rings are both depleted in gas and dust (Hammond et al. [2023]).

The existence of a protoplanetary companion, HD 169142 b, has been proposed in 2019 (Gratton et al. [2019]) and in 2023 (Hammond et al. [2023]) using direct imaging and disk dynamics. This companion has yet to be confirmed but is a very promising target to investigate. However, due to the symmetrical nature of the disk structure, the use of ADI is not adapted since most of the disk will be captured by the algorithm as it is not seen rotating in the image cube.

Chapter 7

Results

In this section, the different available observations are described, with an emphasis on the most important variables for determining whether the observation is suitable for direct imaging. These include the seeing - a measure of atmospheric turbulence during the observation - as well as the PA rotation of the image cube and the number of science frames remaining after binning the bad frames. After listing the available observations, the best images returned by the I-PCA algorithm for each target are presented and briefly commented on. A more in-depth discussion is provided in Section 8.

7.1 PSD 70

PDS 70 has been observed twice in the NaCo archive; the detailed observations are summarized in Table 7.1.

To compare the NaCo images with those in the literature, a figure illustrating the positions of both protoplanets and the structure of the disk in the PDS 70 system is presented in Figure 7.1.

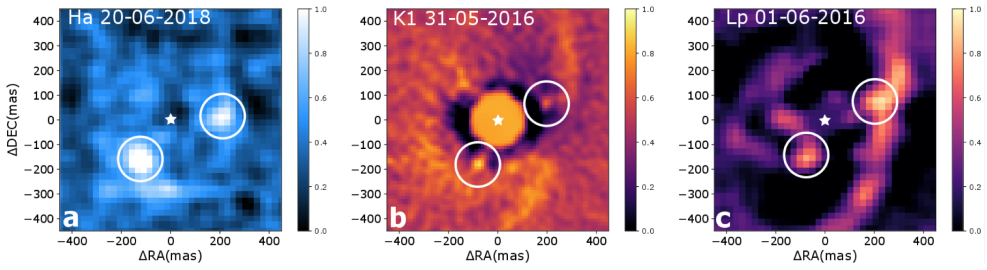


Figure 7.1: PDS 70 b (left white circle) and c (right white circle) multi-epoch and wavelength detections. The star is centered on the images and represented by a white star pattern. (a) Hα detection map using MUSE/VLT on June 20, 2018. (b) K1-band observation using SPHERE/VLT on May 14, 2016 using ADI post-processing. (c) L-band observations using NaCo/VLT on June 1, 2016. (Credits: Haffert et al. [2019])

Date	PA [°]	Seeing (median)	Seeing (std)	Number of frames	Observation mode
June 2, 2016	84.1	0.42	0.08	1615	Dither
July 9, 2019	103.7	0.58	0.11	4719	Dither

Table 7.1: List of NaCo observations for PDS 70.

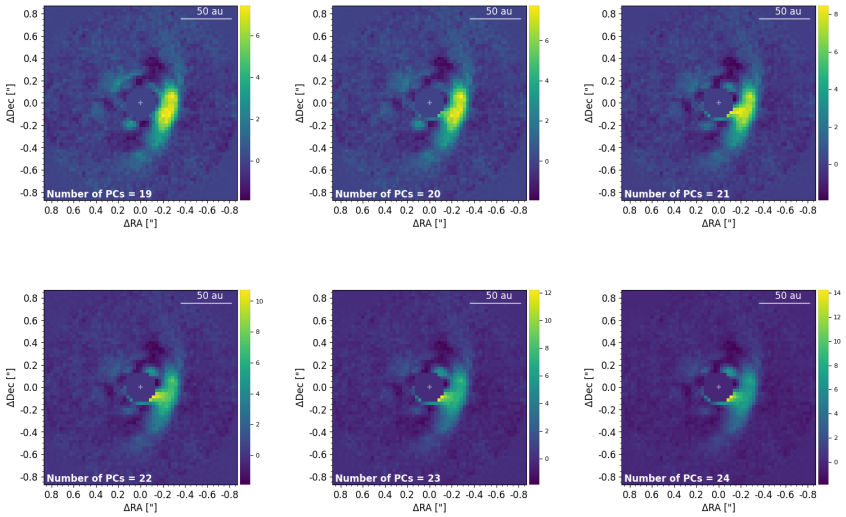


Figure 7.2: June 2, 2016 NaCo observation of PDS 70 processed using I-PCA-ARDI using thr = 1.5, 8 iterations per PC and starting at PC 18. Comparison of the different images when stopping at PC 19 to 24.

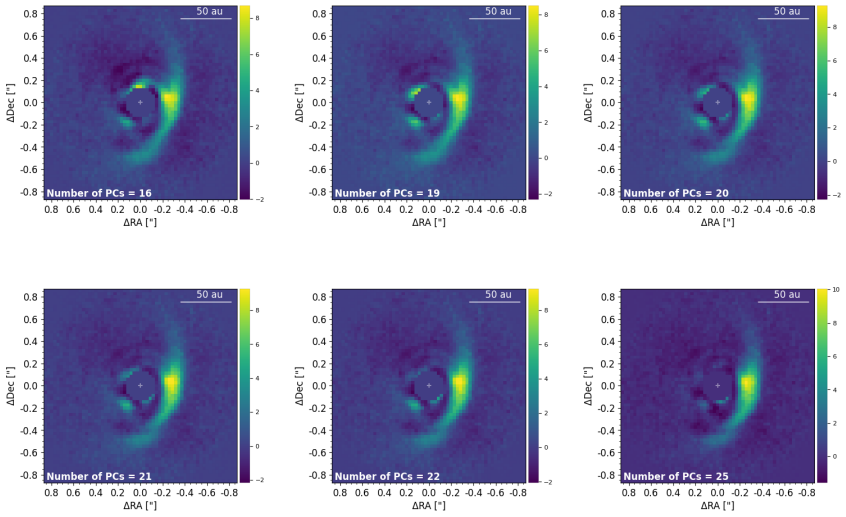


Figure 7.3: July 9, 2019 NaCo observation of PDS 70 processed using I-PCA-ARDI using thr = 1.5, 10 iterations per PC and starting at PC 9. Comparison of the different best images when stopping at PC 16 to 25.

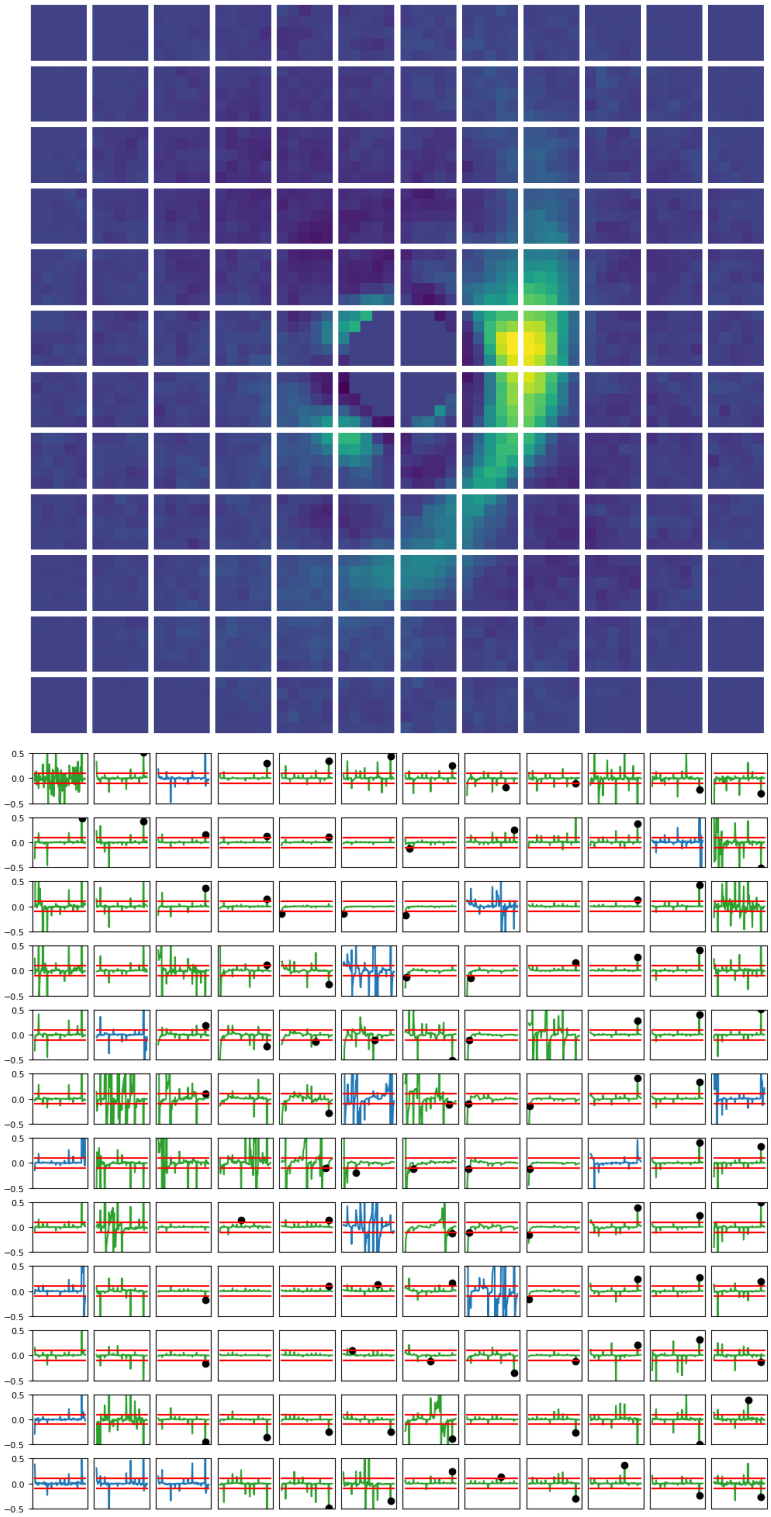
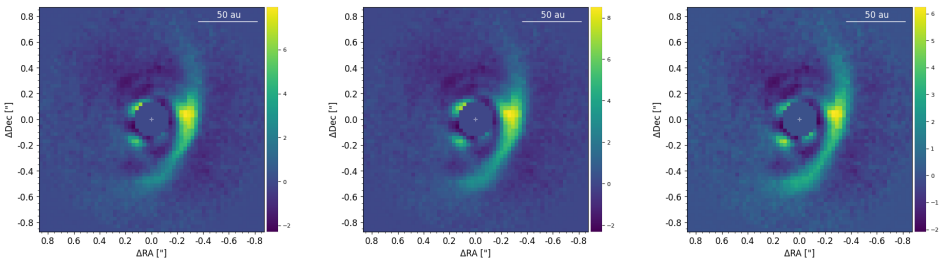


Figure 7.4: Convergence study of the I-PCA results for PDS 70 observed on July 9, 2019. Each cell corresponds to a square in the top image and the graphs plots the convergence of the I-PCA results until the top image is reached.



(a) Initial PC = 6 (b) Initial PC = 9 (c) Initial PC = 19

Figure 7.5: Study of the influence of the initial number of PC considered.

A high-quality image of the circumstellar disk around PDS 70 can be obtained from the observation of July 9, 2019. Despite favorable seeing conditions and a high PA difference, the images from the observation on June 2, 2016 are not of the same quality as those from the later observation. The I-PCA-ARDI images from both datasets are shown in Figures 7.2 and 7.3 for the 2016 and 2019 observations, respectively. In both images, the most prominent companion, PDS 70 b, is clearly visible at the bottom left of the central mask. The second companion, PDS 70 c, is visible only in the 2019 data; in the 2016 observation, it is buried within the disk signal, and its position can only be inferred. The disk itself displays the same structure as the one shown in Figure 7.1c.

Another interesting result is the apparent clockwise rotation of PDS 70 b by a few degrees between its 2016 and 2019 positions. Over a three-year interval, such a modest shift is expected, as the companion is located roughly 20 AU from the central star - about the same distance as Uranus from the Sun, which has an orbital period of 84 years.

The choice of parameters in the I-PCA algorithm plays a crucial role in determining the final image quality. As previously mentioned, the final PC considered is a tunable parameter, but so are the initial PC and the number of iterations. For this analysis, the initial number of principal components was set to 9 in order to retain as much signal as possible from both PDS 70 b and c. A comparison is provided in Figure 7.5. In this figure, it is challenging to identify a clearly superior result, though it appears that some of the signal from PDS 70 c is lost when starting at PC 19 or 6. The number of iterations is also a key factor: it must be large enough to ensure the recovery of all significant signal, yet not so large as to unnecessarily increase computation time without additional benefit.

Finally, a convergence study of the I-PCA process is shown in Figure 7.4. This graph tracks signal convergence within 5×5 pixel cells. The top image indicates the location of a given cell in the I-PCA image; the selected frame corresponds to the subjectively best result from the 2019 observation, obtained by stopping at PC 20. This graph helps assess whether the number of iterations used was sufficient for the signal to converge. In this case, ten iterations appear adequate for convergence.

7.2 HD 36112

HD 36112 has been observed only a single time in the NaCo archive, the detailed observation is summarized in Table 7.2.

To compare the NaCo images with those in the literature, a figure illustrating the structure of the disk in the HD 36112 system is presented in Figure 7.6.

Images of HD 36112 using the single NaCo dataset available are presented in Figure 7.7. These images show the same disk structure as reported in the literature. Each of the nine final images is computed by stopping at a given PC, from 3 to 11.

The downward spiral arm appears to be more affected by the increase in PC. This might be due to the fact that, being less visible than the upward spiral arm, the threshold used may more easily misclassify its signal as noise rather than significant.

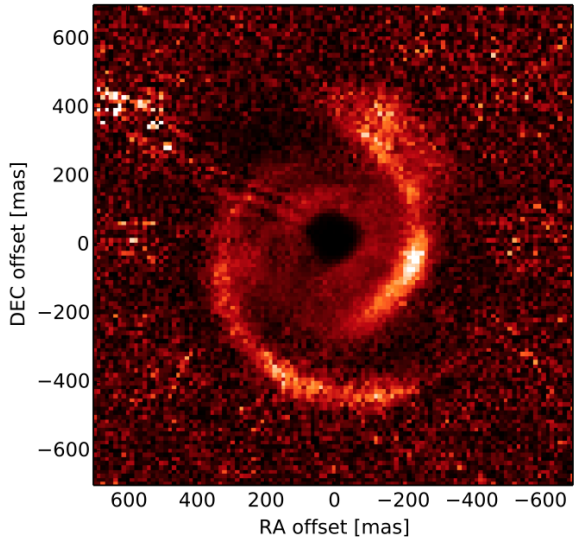


Figure 7.6: Direct image of HD 36112 using Y-band observations from SPHERE/VLT on December 5, 2014. (Credits: Benisty et al. [2015])

Date	PA [°]	Seeing (median)	Seeing (std)	Number of frames	Observation mode
January 17, 2019	62.8	0.58	0.11	3989	AGPM

Table 7.2: List of NaCo observations for HD 36112.

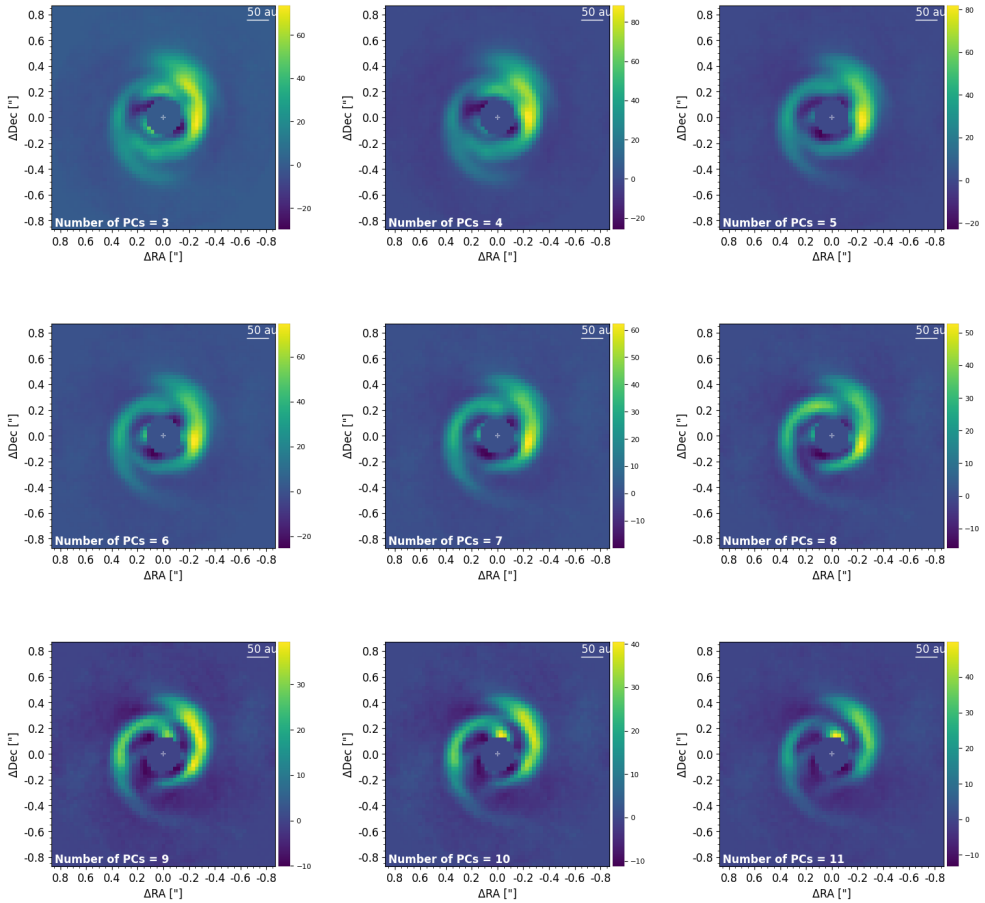
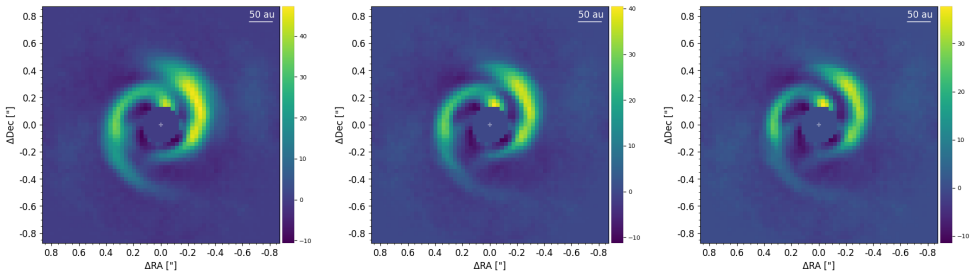


Figure 7.7: July 9, 2019 NaCo observation of HD 36112 processed using I-PCA-ARDI using thr = 7, 11 iterations per PC and starting at PC 2. Comparison of the different best images when stopping at PC 3 to 11.



(a) Threshold = 0 (b) Threshold = 7 (c) Threshold = 8

Figure 7.8: Study of the influence of the threshold for identifying significant signals. The number of PC considered is 10.

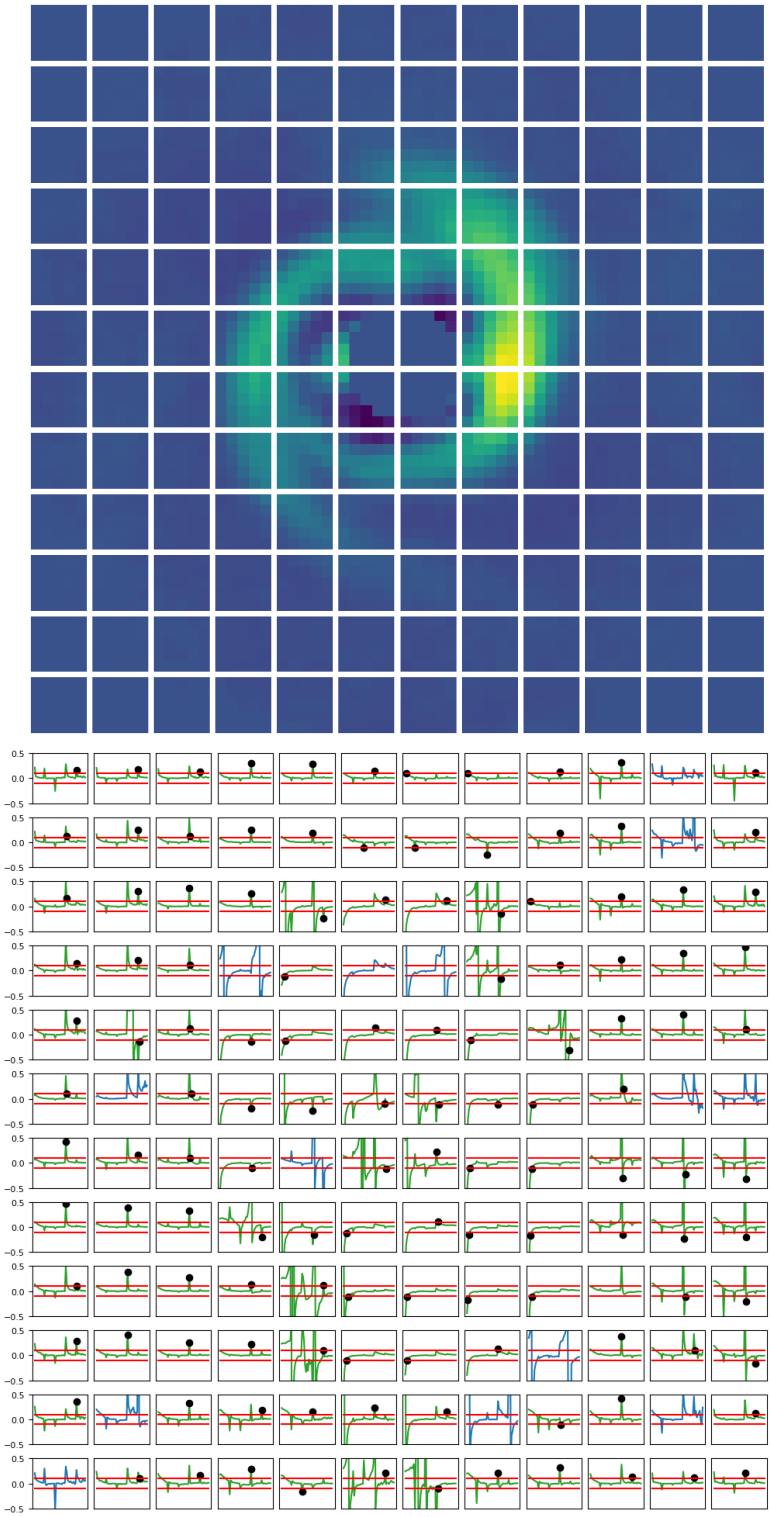


Figure 7.9: Convergence study of the I-PCA results for HD 36112 observed on January 17, 2019. Each cell corresponds to a square in the top image and the graphs plots the convergence of the I-PCA results until the top image is reached.

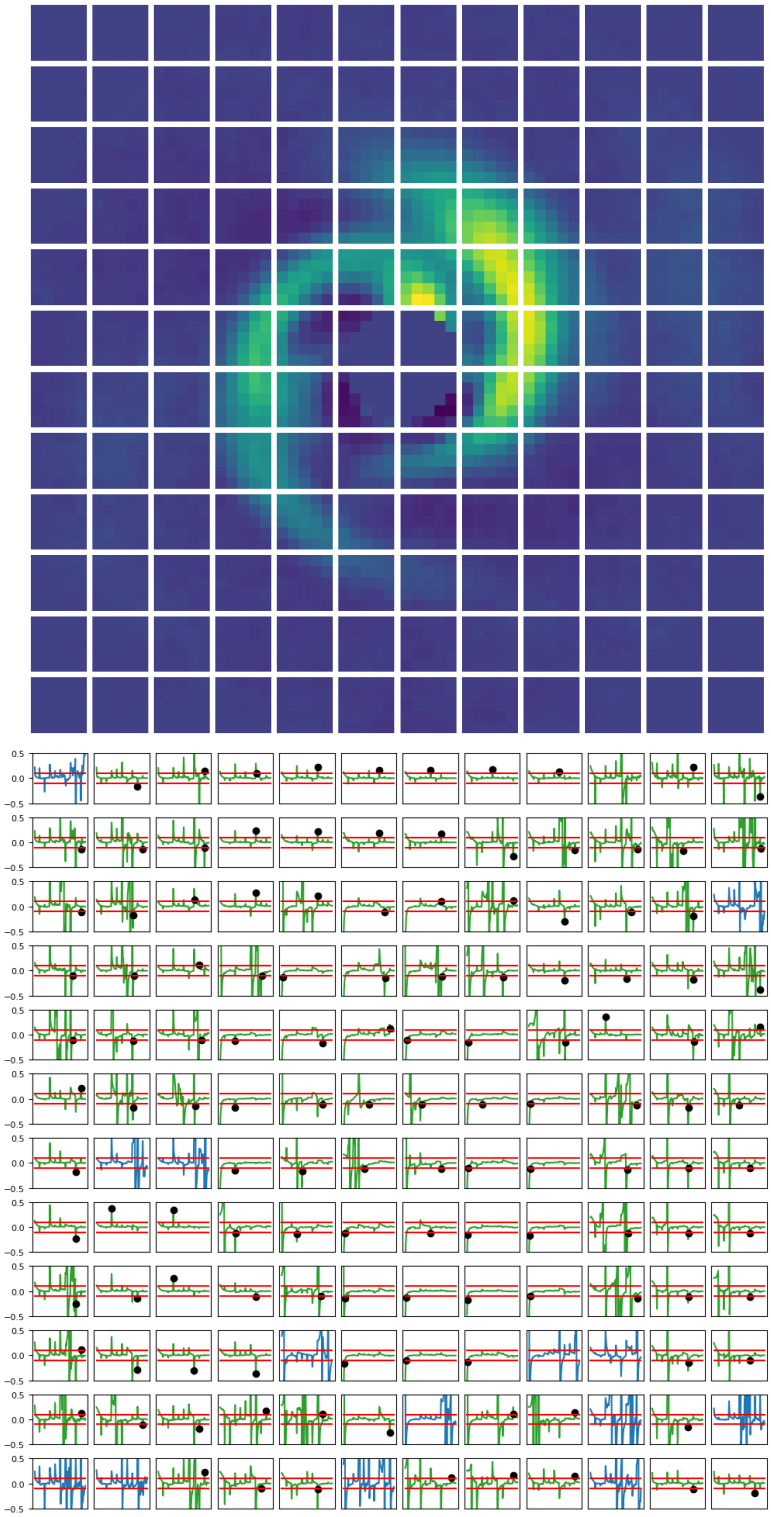


Figure 7.10: Convergence study done with the exact same parameters as Figure 7.9 but with threshold = 0.

Figure 7.8 presents a comparison of different thresholds for the image resulting from stopping at PC 10. Using a lower threshold (in this case, 0) reduced the amount of self-subtraction for the downward spiral arm but also increased the amount of noise propagated through the iterative process, resulting in a blurrier image, as can be seen in Figure 7.8a.

Finally, a convergence study of the I-PCA process can again be analyzed through the same graphs as for the previous target. The convergence is presented in Figure 7.9. This time, however, a second convergence graph is presented in Figure 7.10. This second graph is built from the I-PCA process using a threshold of 0. The inspiration behind using a null threshold comes from Juillard et al. [2024]. In that article, the threshold was set to 0 for all I-PCA reductions, and a consequence of this choice was the non-convergence of the signal in cells without significant signal, which were dominated by noise. This second convergence graph shows exactly that problem, as most of the outer cells have large peaks and sometimes do not even converge. On the other hand, the central cells mostly follow the same convergence pattern as the graphs in Figure 7.9, where most of the cells converge, including the outer ones.

7.3 HD 135344B

HD 135344B has been observed six times in the NaCo archive; the detailed observations are summarized in Table 7.3. Unfortunately, due to poor seeing and a low number of science frames, the two observations from 2017 and the one from June 30, 2018 are unusable. The PCA images resulting from processing these three observations do not show clear significant signals and are instead dominated by noise.[1]

To compare the NaCo images with those in the literature, a figure illustrating the structure of the disk in the HD 135344B system is presented in Figure 7.11.

Date	PA [°]	Seeing (median)	Seeing (std)	Number of frames	Observation mode
March 24, 2013	138.5	0.79	0.03	5356	Dither
May 8, 2017	31.9	1.26	0.25	586	AGPM
May 24, 2017	33.5	1.13	0.15	613	AGPM
June 2, 2018	86.8	0.40	0.06	2001	AGPM
June 3, 2018	139.1	0.37	0.05	2379	AGPM
June 30, 2018	105.8	2.48	0.66	1401	AGPM

Table 7.3: List of NaCo observations for HD 135344B. In gray, the unusable observations due to bad observation conditions.

[1] In the case of the 2017 observations, this could be due to a too low PA rotation, but even a reduction using only RDI does not yield better images.

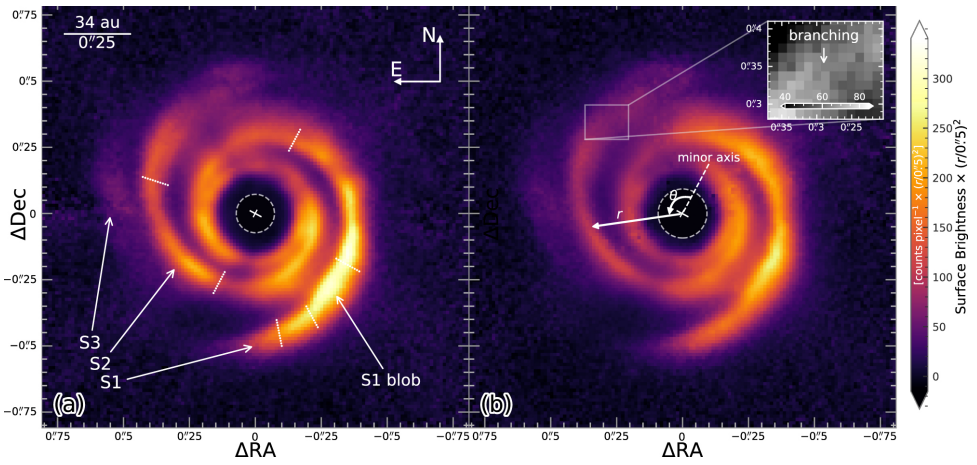


Figure 7.11: J-band observations from HD135344B using SPHERE/VLT on May 3, 2015 (left) and May 4, 2016 (right). (Credits: Xie et al. [2021])

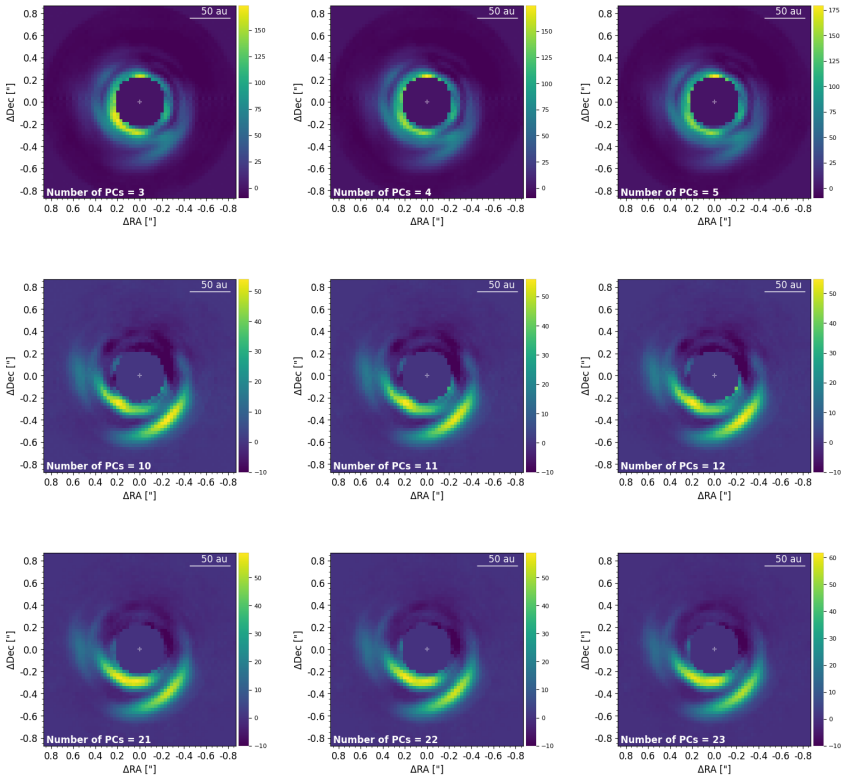


Figure 7.12: March 24, 2013 NaCo observation of HD 135344B processed using I-PCA-RADI using thr = 1.5, 10 iterations per PC and starting at PC 1. Comparison of the different best images when stopping at various PCs between 3 and 25.

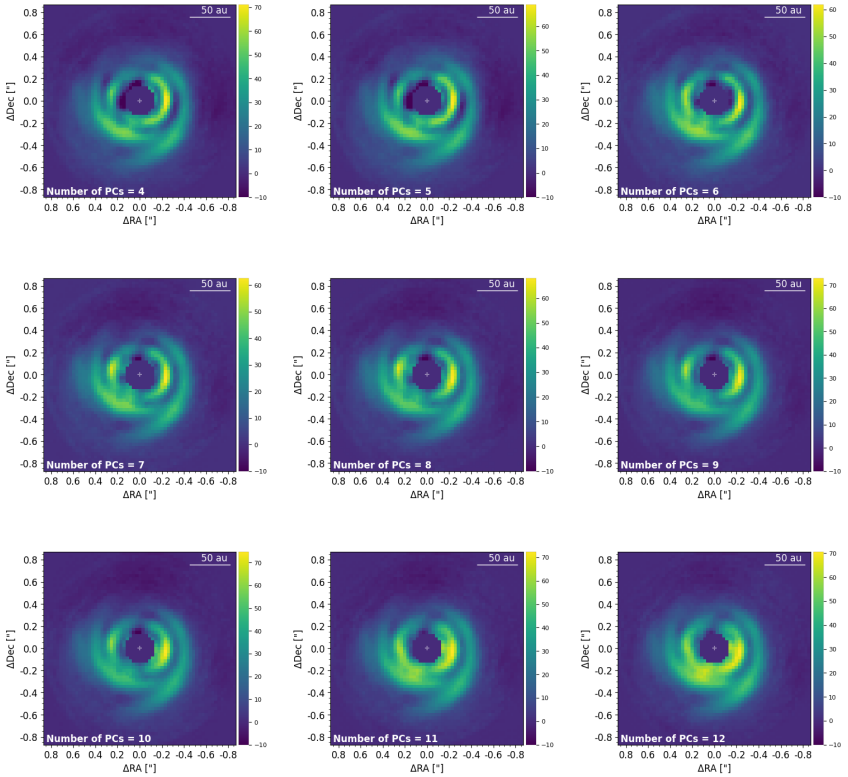


Figure 7.13: June 2, 2018 NaCo observation of HD 135344B processed using I-PCA-RADI using thr = 0.5, 11 iterations per PC and starting at PC 1. Comparison of the different best images when stopping at PC 4 to 12.

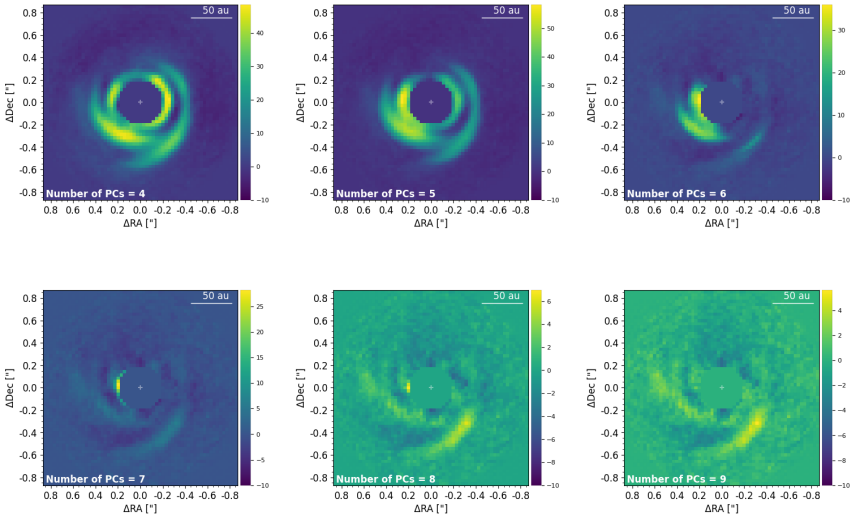


Figure 7.14: June 2, 2018 NaCo observation of HD 135344B processed using I-PCA-ARDI using thr = 10, 15 iterations per PC and starting at PC 1. Comparison of the different best images when stopping at PC 4 to 9.

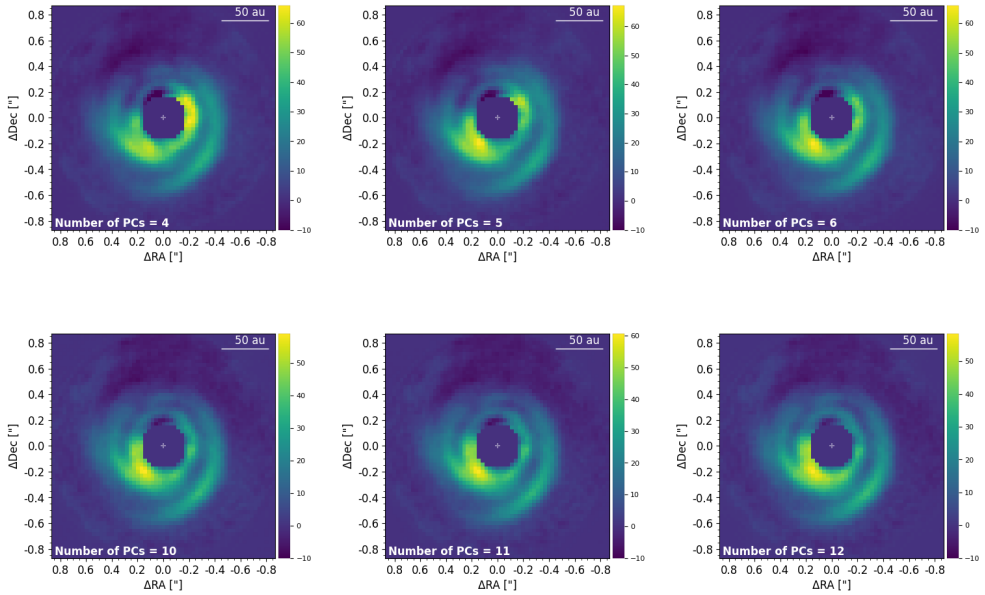


Figure 7.15: June 3, 2018 NaCo observation of HD 135344B processed using I-PCA-RADI using thr = 0, 10 iterations per PC and starting at PC 1. Comparison of the different best images when stopping at various PCs between 4 and 12.

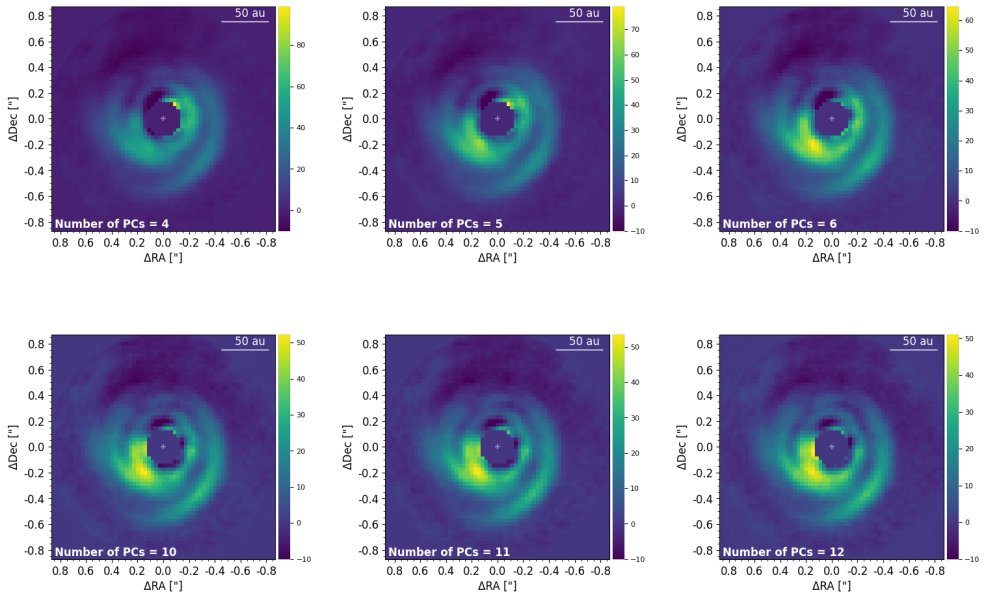


Figure 7.16: June 3, 2018 NaCo observation of HD 135344B processed using I-PCA-RDI using thr = 6, 10 iterations per PC and starting at PC 1. Comparison of the different best images when stopping at various PCs between 4 and 12.

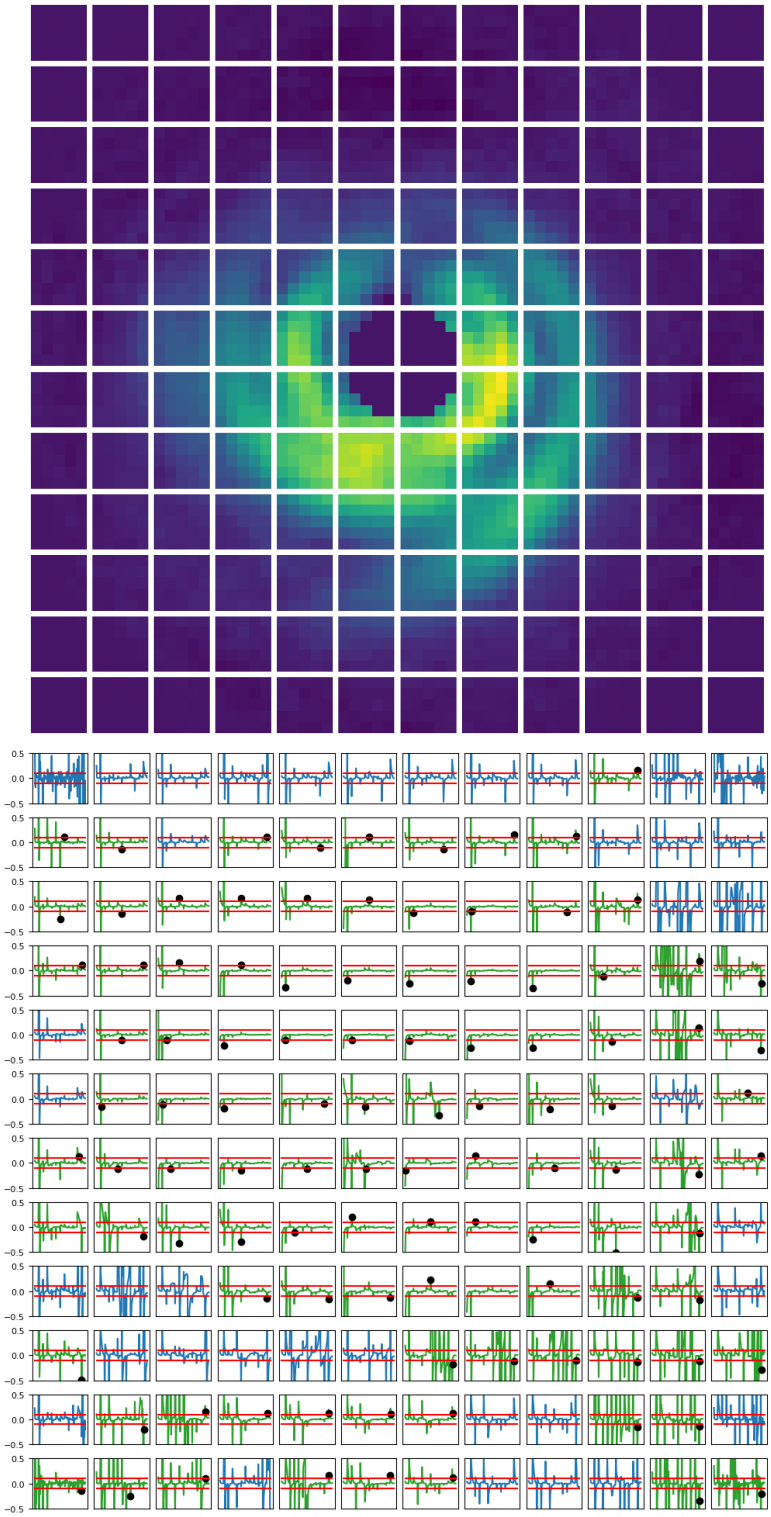


Figure 7.17: Convergence study of the I-PCA-RADI results for HD 135344B observed on June 2, 2018. Each cell corresponds to a square in the top image and the graphs plots the convergence of the I-PCA results until the top image is reached.

The first observation analyzed, the only dither observation, is from March 24, 2013. A mosaic with different final PCs is presented in Figure 7.12. This figure was produced using the RADI algorithm. The only difference with the ARDI algorithm is that RADI first iterates on RDI alone before starting to use ADI. This means that if the result is already very good after the RDI step, not much ADI will be applied, thereby reducing the amount of self-subtraction induced. This approach is particularly efficient for datasets with low PA difference or for disks that are highly rotationally invariant (disk structures with a significant amount of circular symmetry). From these images, only the lower part of the disk can be properly resolved. This observation has a lot of residual starlight that is not processed by the algorithm at the edge of the central cavity. This might be due to the fact that this is a dither observation, as the other AGPM datasets do not have this problem.

The second observation analyzed is from June 2, 2018. A mosaic with different final PCs is presented in Figure 7.13. These images are again obtained leveraging the RADI strategy to compensate for the high rotational invariance of the target's disk. Images obtained using a classic ARDI strategy are presented in Figure 7.14. It is apparent that using the RADI strategy is necessary for this dataset. Indeed, a large point-source appears on the center left of the image and grows with iteration, completely deleting the signal from the disk within a few PCs. This point-source is very likely an artifact from the algorithm rather than a companion, as it is very large and also changes shape, position, and brightness as the PC increases. It is also not detected in any other observation, which is one of the main criteria for a companion candidate.

For this June 2nd observation, a convergence graph is presented in Figure 7.17. Due to the low threshold of 0.5, the phenomenon discussed with the previous target - where noise is captured by the iterative procedure - is also observed here: most of the outer cells exhibit high peaks and non-convergence of the signal.

The last observation analyzed is from June 3, 2018, the night after the previous observation. A mosaic with different final PCs is presented in Figure 7.15. These images are again obtained leveraging the RADI strategy, but a comparison with images obtained using an RDI strategy can be made by referring to Figure 7.16. Both strategies return very similar images, indicating that the algorithm using the RADI strategy mostly relied on RDI to reach the final result.

7.4 HD 169142

HD 169142 has been observed six times in the NaCo archive. The detailed observations are summarized in Table 7.4. Once again, some observations are unusable due to excessively high seeing. No quality image of the protoplanetary disk could be retrieved from the observations on July 14, 2013, and May 31, 2016. Another observation is unfortunately unusable, as two large dark spots are visible in the cube throughout the entire duration. This could be due to a binary star being in the field of view, with the pre-processing pipeline attempting to mask it. Figure 7.18 presents an image from this unusable dataset.[2]

Date	PA [°]	Seeing (median)	Seeing (std)	Number of frames	Observation mode
June 29, 2013	158.9	0.80	0.08	4551	AGPM
July 14, 2013	135.4	1.36	0.22	2736	AGPM
May 31, 2016	138.9	1.13	0.15	1481	AGPM
May 5, 2017	140.4	0.47	0.10	2406	AGPM
May 18, 2017	94.1	0.90	0.05	1491	AGPM
May 19, 2017	109.5	0.55	0.12	3005	AGPM

Table 7.4: List of NaCo observations for HD 169142. In gray, the unusable observations due to bad observation conditions and in red, the unusable observation due to the presence of a pair of dark spots in the image cube.

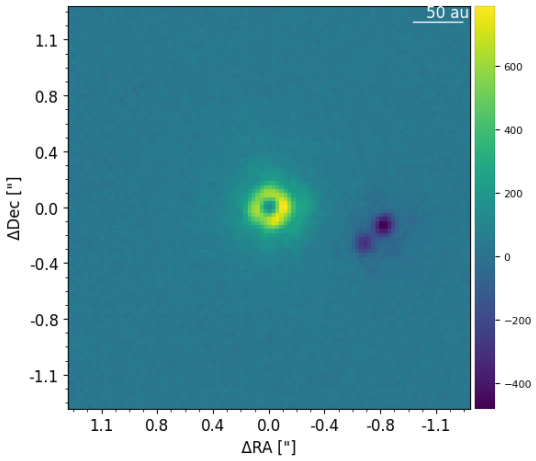


Figure 7.18: First frame of the raw observation image cube of HD 169142 from May 5, 2017. A pair of dark spots can be seen all throughout the cube, rendering the observation unusable.

[2] In an attempt to salvage the observation, the image cube was cropped to a small enough pixel size to exclude the dark spots. Unfortunately, this approach failed to yield any usable result using a PCA algorithm on the cropped cube.

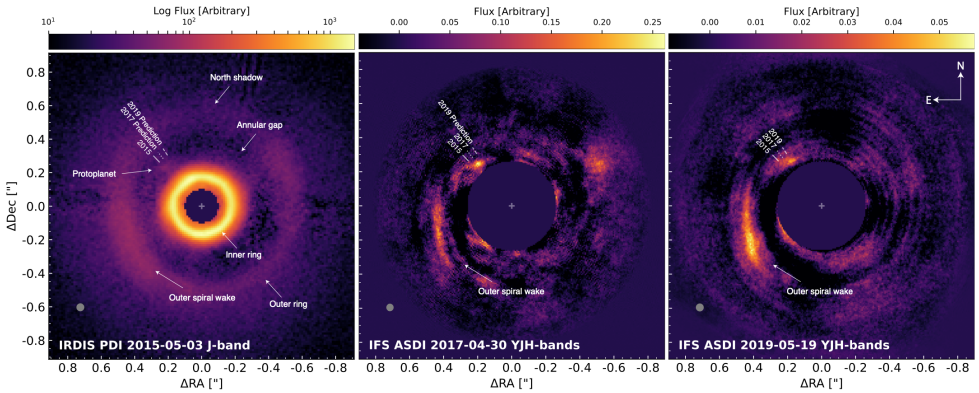


Figure 7.19: SPHERE/NaCo observations of the circumstellar feature of HD 169142 at three different epoch with detection of a companion candidate HD 169142 b. (Credits: (Hammond et al. [2023])

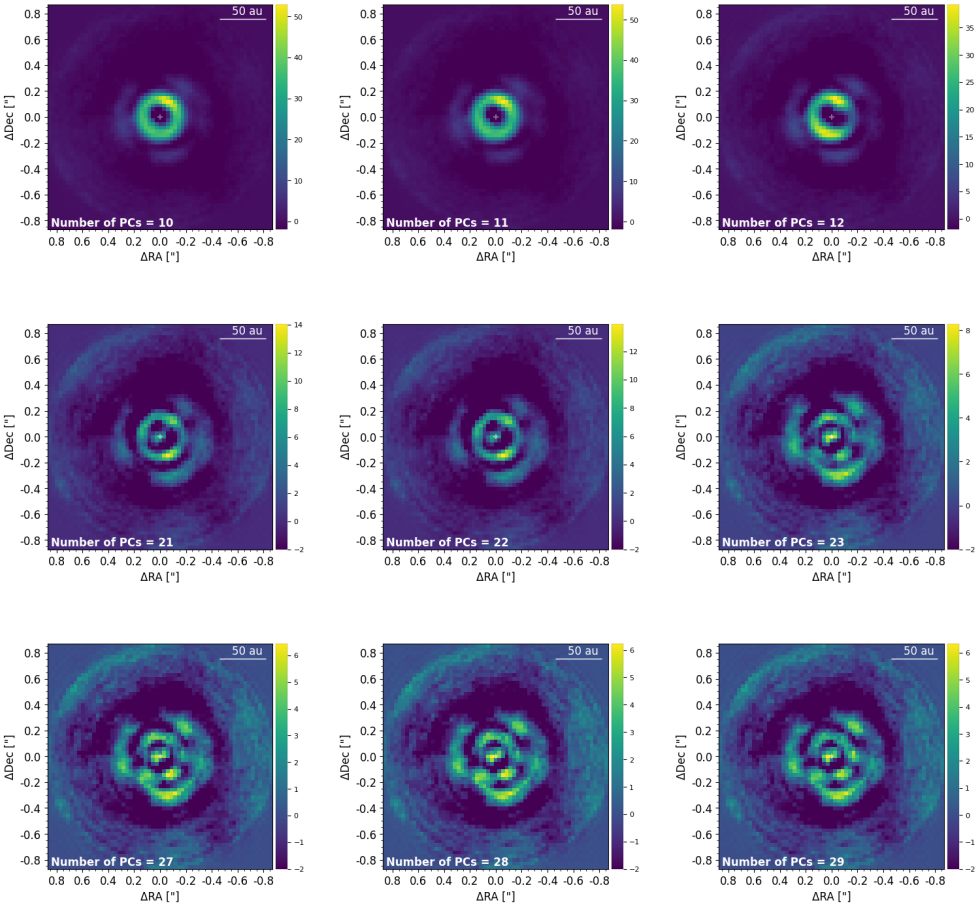


Figure 7.20: June 29, 2013 NaCo observation of HD 169142 processed using l-PCA-RDI using thr = 14, 11 iterations per PC and starting at PC 1. Comparison of the different best images when stopping at various PCs between 10 and 29.

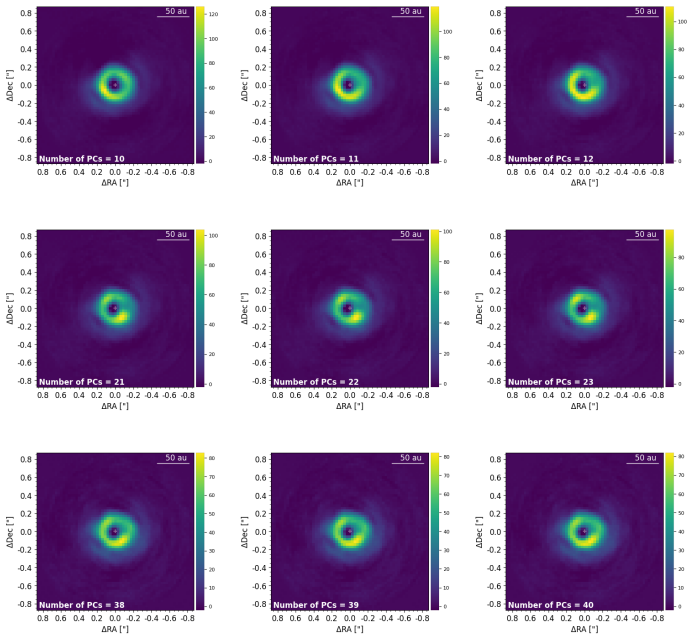


Figure 7.21: May 18, 2017 NaCo observation of HD 169142 processed using I-PCA-RDI using thr = 0, 10 iterations per PC and starting at PC 1. Comparison of the different best images when stopping at various PCs between 10 and 40.

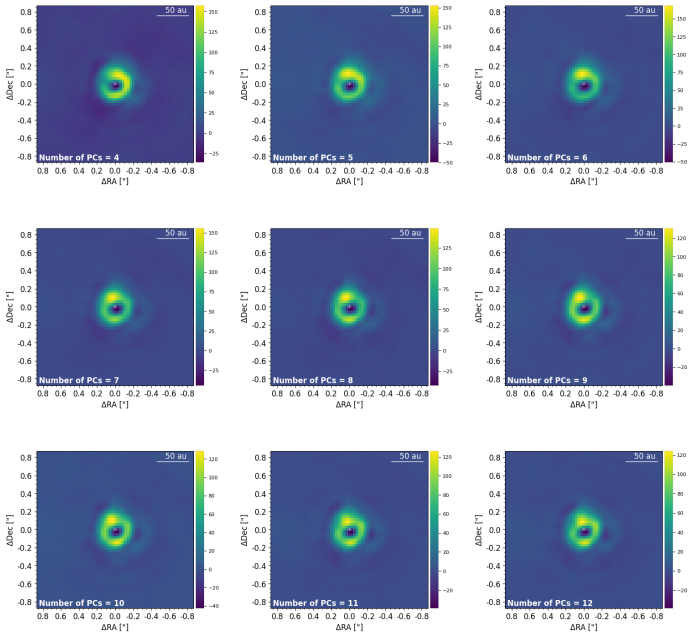


Figure 7.22: May 19, 2017 NaCo observation of HD 169142 processed using I-PCA-RDI using thr = 12, 10 iterations per PC and starting at PC 1. Comparison of the different best images when stopping at various PCs between 4 and 12.

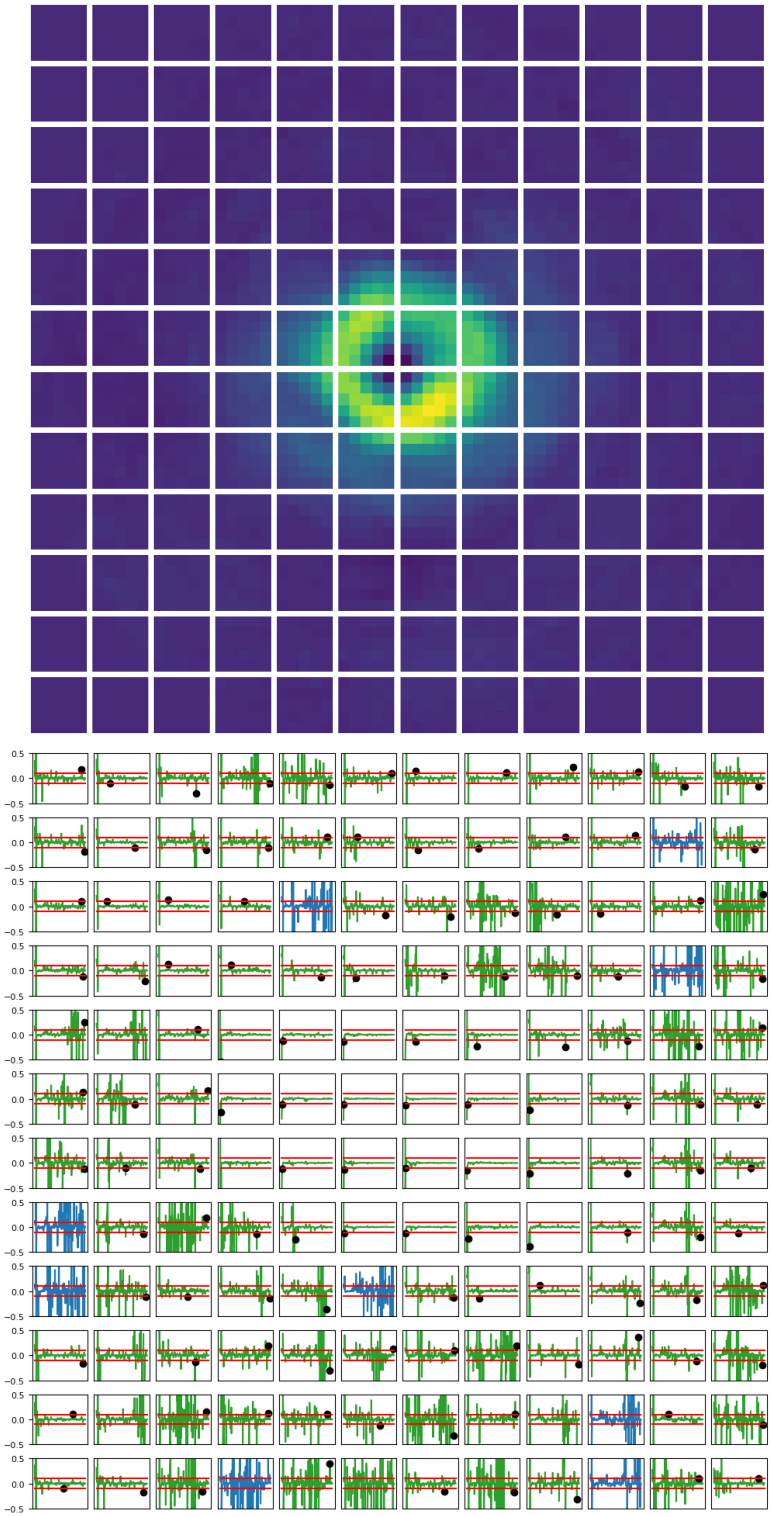


Figure 7.23: Convergence study of the I-PCA-RDI results for HD 169142 observed on May 18, 2017. Each cell corresponds to a square in the top image and the graphs plots the convergence of the I-PCA results until the top image is reached.

To compare the NaCo images with those in the literature, a figure illustrating the disk structure and the position of the candidate companion in the HD 169142 system is presented in Figure 7.19. Due to the highly rotationally invariant structure of the disk in this target, all images were obtained using the RDI strategy. Using any kind of ADI, results in a rapid capture of the disk within the PCs, effectively removing the signal. An example of this effect is shown in the Appendix in Figure 1.

The first observation analyzed is from June 29, 2013, and is presented in Figure 7.20. The inner ring disk structure is clearly visible in the first few PCs. However, since the threshold is not equal to zero, the disk begins to be excluded as a significant signal as it fades through iterations. At some point, some structures appear just outside the disk. These resemble a secondary ring but seem too close to the inner ring to correspond to the outer ring observed in the literature. This is likely residual starlight that was not properly subtracted; using only RDI is not as robust as an ARDI strategy.

The two final observations, from May 18 and 19, 2017 - presented in Figures 7.21 and 7.22, respectively - show minimal change throughout the iterative process. The first observation uses a threshold of 0, so the disk is always identified as a significant signal. The second uses a low enough threshold to keep the disk visible even after 12 PCs. No outer ring is visible in either observation, even at higher PCs, suggesting that the ring seen in the 2013 observation is likely an artifact of the post-processing.

Finally, a convergence study of the May 18, 2017 observation is presented in Figure 7.23. Convergence is clearly instantaneous for the central cells containing the ring, while in the rest of the image, the signal does not converge easily - if at all - once again due to noise propagation.

Chapter 8

Discussion

Now that the results have been presented, this chapter discusses their implications. It begins with a general analysis of how the I-PCA algorithm depends on its parameters, followed by an interpretation of the observed convergence behavior. Finally, the limitations of this study are discussed in relation to its objectives.

8.1 Results Interpretation

For all four targets of interest, the application of the I-PCA algorithm yields images comparable to those found in the literature. In the case of PDS 70, some processed images even reveal both confirmed protoplanets in the system. Unfortunately, no additional companions could be detected or confirmed for the other three targets. However, further work on optimizing the algorithm's parameters may enhance detectability in future analyses.

8.1.1 Algorithm parameter dependence

First, a few remarks can be made regarding the processing techniques used for the datasets, such as RDI, ARDI, and RADI. Most image cubes were processed using the ARDI technique, which combines both ADI and RDI to leverage the strengths of each method. However, the rotational-invariance of some protoplanetary disks can make ADI suboptimal due to self-subtraction effects, which can significantly reduce the disk signal.

To mitigate this, the RADI technique - where the RDI step is performed first, followed by the angular step - can offer a better compromise by reducing the impact of self-subtraction while still benefiting from angular diversity. For cases where even less reliance on ADI is desired, pure RDI processing can be considered. However, this comes with its own limitations, such as increased risk of over-subtraction of the disk signal or incomplete subtraction of stellar speckles.

Initial principal component

The initial PC at which the iterative process begins is one of the tunable parameters of the I-PCA algorithm. While an analysis of its influence is performed on one of the PDS 70 datasets in Section 7.1, similar comparisons were conducted on other datasets from different targets, although these are not presented in this thesis.

Overall, the choice of the initial PC does not appear to significantly affect the final result. Nonetheless, it remains important to select a low enough value to avoid excluding too much of the signal of interest, which may be captured in the lowest-rank components. Starting from the first PC is not always necessary - especially for protoplanetary disks, which often span on a wide range of components. Some computational time can be saved by skipping the first few PCs, provided the disk signal is still recovered in later components. Ultimately, the optimal choice will depend on the specific target, and an initial inspection using a grid of non-iterative PCA images remains a recommended step to determine the most effective strategy.

Significant signal threshold

Another critical parameter in the I-PCA algorithm is the threshold used to determine which signals are considered significant and should be excluded from the subtracted PSF. To retain as much of the astrophysical signal as possible, the threshold should be set to a low value. However, setting it too low increases the risk of including noise, which may then be propagated throughout the image.

Fortunately, the use of a 2-dimensional STIM map somewhat mitigates this issue, as the threshold is not scaled globally but rather based on annular maxima. This adaptive approach helps to prevent noise from being treated as significant signal, especially in regions dominated by it.

Number of iterations

The number of iterations has been briefly discussed in the results section for some targets. It must be set high enough to allow the pixels containing signals of interest to converge during each step of the iterative process. If the number of iterations is too low, the algorithm may not fully capture the significant components of the signal, potentially leading to the loss of parts of the disk that cannot be recovered once subtracted.

According to the convergence graphs presented for various datasets, 10 iterations generally appear to be sufficient. Beyond this point, the changes between iterations become minimal, indicating that the signal has stabilized. Increasing the number of iterations further results in only marginal improvements while significantly increasing computation time. Therefore, a balance must be struck between computational efficiency and sufficient convergence of the signal.

8.1.2 Convergence

All convergence figures presented in the previous section exhibit a similar pattern. Sharp spikes appear each time the algorithm moves on to the next PC; these occur because the iterative process requires a few steps to remove new significant signals revealed by adding the next component to the subtracted PSF.

The most informative aspect of these convergence graphs is whether the value within a given cell stabilizes over time. A converging value suggests that the signal is consistently present across iterations and likely corresponds to a real astrophysical feature. In contrast, a non-converging value - one that varies significantly from iteration to iteration - may indicate that the signal is dominated by noise, rather than being part of the signal of interest.

8.2 Limitations of this study

This study did not result in the detection of any new protoplanets candidates. Its primary objective was to image protoplanetary disks and, if possible, identify companions embedded within them. However, compared to modern instruments, NaCo provides significantly lower spatial resolution and contrast. These limitations are particularly restrictive in direct imaging, where detecting planetary-mass companions often requires the highest possible angular resolution and contrast performance.

Despite these limitations, as demonstrated in the case of PDS 70, additional observations remain valuable. Archival datasets can be instrumental in confirming the positions and characteristics of candidate companions. Furthermore, applying modern post-processing algorithms to data from decommissioned instruments - like NaCo - may uncover previously undetectable features or help confirm previous detections.

Chapter 9

Conclusion and Future Prospects

The initial stages of this work focused on building a solid understanding of protoplanetary disks and exoplanetary science more broadly. A particular emphasis was placed on direct imaging techniques, which are among the most powerful methods for detecting exoplanets and characterizing protoplanetary systems. In this context, the VIP Python library was used extensively. As it contains a wide range of post-processing algorithms, becoming familiar with its structure and functionality was a key early objective.

The first technical step involved working with NaCo datasets that had been pre-processed by a previous team. To facilitate later stages of analysis, a system was developed to automatically extract and compile relevant data into organized tables. Additional parameters required for the post-processing phase were also computed at this stage.

A critical component of the work was the selection of suitable reference stars for use in RDI techniques. A significant effort was invested in developing a robust pipeline to identify optimal references. Four different similarity metrics were evaluated, including one - based on PCA - developed specifically for this project. To validate these metrics, a comprehensive injection-recovery test was performed. Fake companions were inserted into 50 reference stars at 16 different positions, and the resulting SNR values were analyzed to determine which metric yielded the most accurate recoveries.

Once the best reference stars were identified, they were used to process the targets of interest using I-PCA algorithms. Each dataset underwent multiple iterations of reduction with varying parameters to optimize image quality. This fine-tuning ensured that the final images extracted the maximum amount of relevant signal. These results were then compared with existing images in the literature to assess the strengths and limitations of the applied techniques in the context of protoplanetary disk and companion imaging.

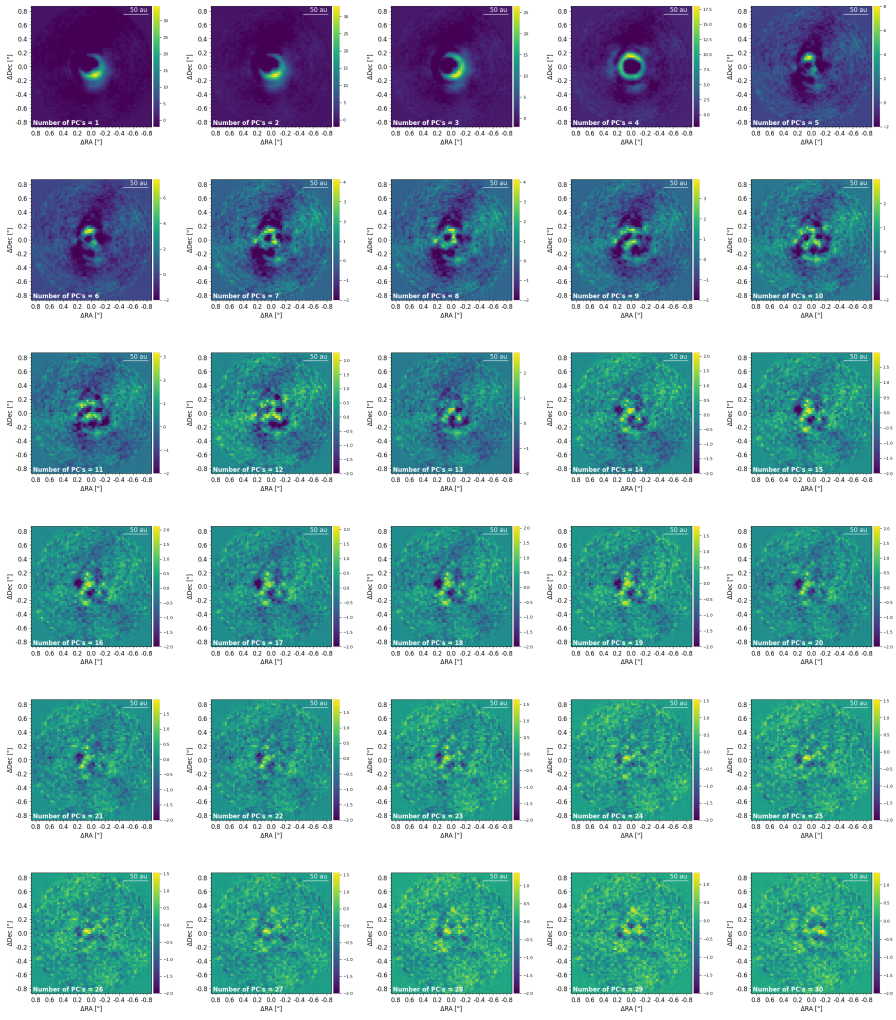
As in all areas of astronomy, the field of direct imaging continues to advance rapidly. The advent of next-generation observatories, such as the Extremely Large Telescope (ELT) currently under construction in Chile, will significantly

expand our capabilities. In particular, METIS - the ELT's mid-infrared instrument designed for high-contrast imaging and spectroscopy - will be a major leap forward. It could be seen as the conceptual successor to NaCo, but with dramatically improved resolution, sensitivity, and added spectroscopic capabilities.

Such instruments will allow us to study protoplanetary disks and embedded protoplanets in unprecedented detail, potentially leading to new discoveries and a deeper understanding of planet formation. The techniques developed and validated in this work can serve as a foundation for analyzing future data from METIS and similar instruments, ensuring that archival research continues to complement and enhance the scientific return of new observations.

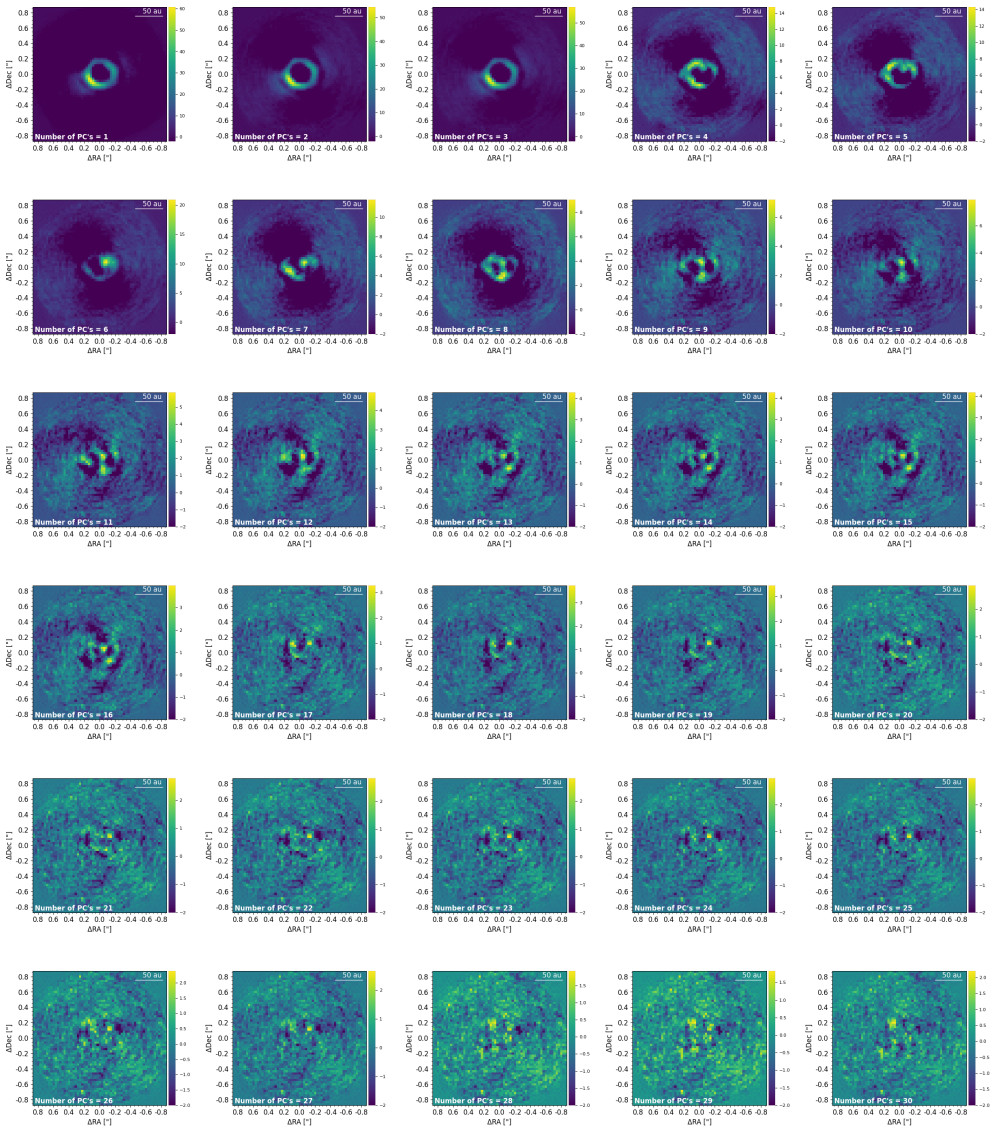
In conclusion, the application of I-PCA algorithms to the direct imaging of protoplanetary disks and embedded companions proves to be a promising approach. Moreover, applying these modern algorithms to archival data - such as the NaCo datasets - opens new opportunities for detection and confirmation without requiring new telescope time. This methodology can be particularly valuable for revisiting known systems or investigating new candidates within the existing archive.

Appendices



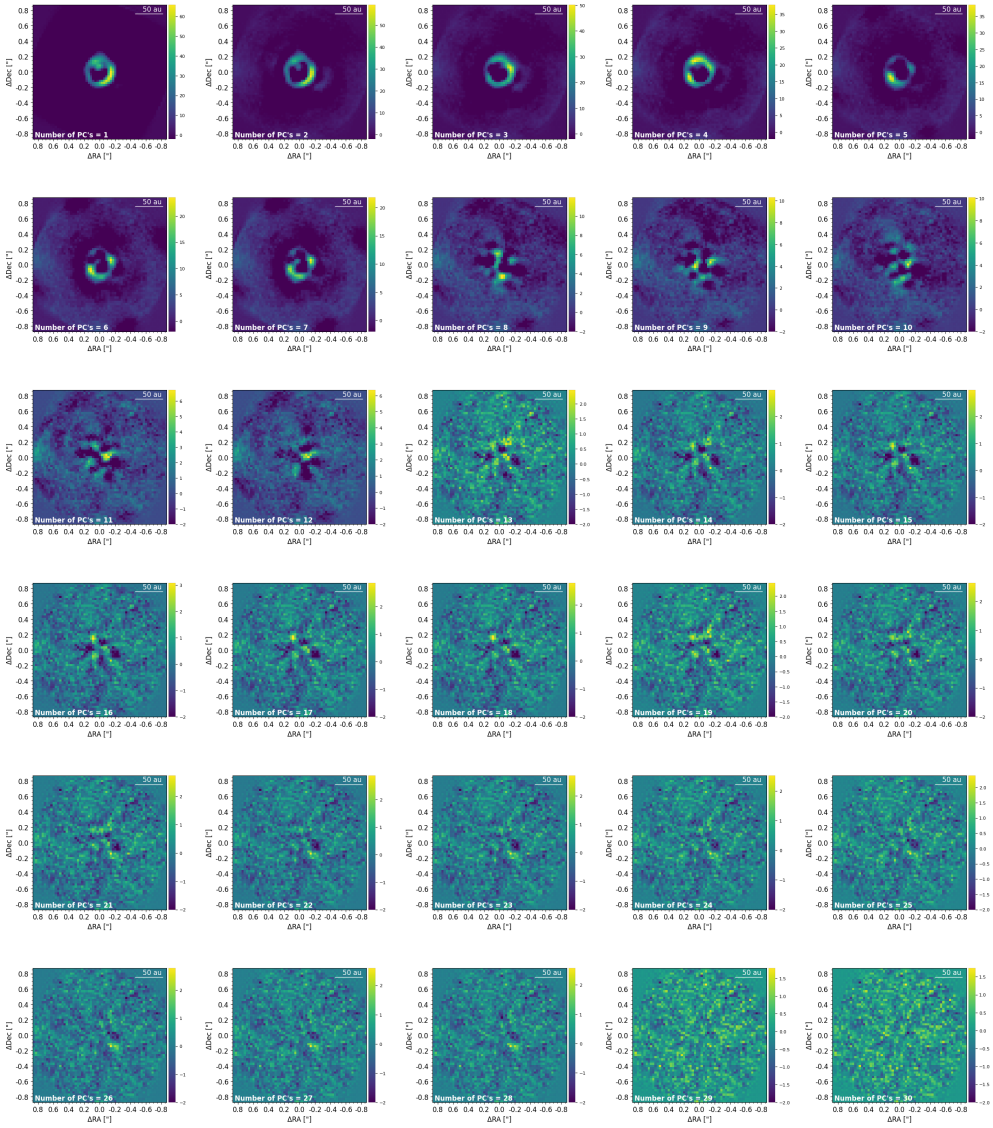
(a) June 29, 2013

Figure 1: ARDI images returned from the algorithm between PCs 1 and 30 for the 3 valid NaCo observations of HD 169142. No signal of interest apart from the disk can be seen, even at high PCs.



(b) May 18, 2017

Figure 1: PCA-RADI images returned from the algorithm between PCs 1 and 30 for the 3 valid NaCo observations of HD 169142. No signal of interest apart from the disk can be seen, even at high PCs.



(c) May 19, 2017

Figure 1: PCA-RADI images returned from the algorithm between PCs 1 and 30 for the 3 valid NaCo observations of HD 169142. No signal of interest apart from the disk can be seen, even at high PCs.

Bibliography

Absil, O., et al., *Three years of harvest with the vector vortex coronagraph in the thermal infrared*, 2016, Proc. SPIE, 9908, 99080Q

Amara, A., & Quanz, S. P., *PYNPOINT: an image processing package for finding exoplanets*, 2012, MNRAS, 427, 948

Andrews, S. M., *Observations of Protoplanetary Disk Structures*, 2020, ARA&A, 58, 483

Armitage, P. J., *Planet formation theory: an overview*, 2024, arXiv, arXiv:2412.11064

Benisty, M., et al., *Asymmetric features in the protoplanetary disk MWC 758*, 2015, A&A, 578, L6

Benisty, M., et al., *Optical and Near-infrared View of Planet-forming Disks and Protoplanets*, 2023, Astronomical Society of the Pacific Conference Series, 534, 605

Birnstiel, T., Fang, M., & Johansen, A., *Dust Evolution and the Formation of Planetesimals*, 2016, Space Sci. Rev., 205, 41

Bitsch, B., et al., *Pebble-isolation mass: Scaling law and implications for the formation of super-Earths and gas giants*, 2018, A&A, 612, A30

Blum, J., *Dust Evolution in Protoplanetary Discs and the Formation of Planetesimals. What Have We Learned from Laboratory Experiments?*, 2018, Space Sci. Rev., 214, 52

Bonse, M. J., et al., *Use the 4S (Signal-Safe Speckle Subtraction): Explainable Machine Learning Reveals the Giant Exoplanet AF Lep b in High-contrast Imaging Data from 2011*, 2025, AJ, 169, 194

Cantalloube, F., et al., *Direct exoplanet detection and characterization using the ANDROMEDA method: Performance on VLT/NaCo data*, 2015, A&A, 582, A89

Casassus, S., et al., *A dusty filament and turbulent CO spirals in HD 135344B - SAO 206462*, 2021, MNRAS, 507, 3789

Christiaens, V., et al., *Evidence for a Circumplanetary Disk around Protoplanet PDS 70 b*, 2019a, ApJ, 877, L33

—, *Separating extended disc features from the protoplanet in PDS 70 using VLT/SINFONI*, 2019b, MNRAS, 486, 5819

—, *MINDS: JWST/NIRCam imaging of the protoplanetary disk PDS 70. A spiral accretion stream and a potential third protoplanet*, 2024, A&A, 685, L1

Cugno, G., et al., *JWST/NIRCam Imaging of Young Stellar Objects. II. Deep Constraints on Giant Planets and a Planet Candidate Outside of the Spiral Disk Around SAO 206462*, 2024, AJ, 167, 182

Currie, T., et al., *No Clear, Direct Evidence for Multiple Protoplanets Orbiting LkCa 15: LkCa 15 bcd are Likely Inner Disk Signals*, 2019, ApJ, 877, L3

—, *Images of embedded Jovian planet formation at a wide separation around AB Aurigae*, 2022, NatAs, 6, 751

Deeg, H. J., & Alonso, R. 2018, in Handbook of Exoplanets, ed. H. J. Deeg & J. A. Belmonte, 117

Dong, R., Zhu, Z., Rafikov, R. R., & Stone, J. M., *Observational Signatures of Planets in Protoplanetary Disks: Spiral Arms Observed in Scattered Light Imaging Can be Induced by Planets*, 2015, ApJ, 809, L5

Dressing, C., et al., *NASA Exoplanet Archive*, 2025, NASA Exoplanet Archive

Drążkowska, J., Alibert, Y., & Moore, B., *Close-in planetesimal formation by pile-up of drifting pebbles*, 2016, A&A, 594, A105

Drążkowska, J., et al., *Planet Formation Theory in the Era of ALMA and Kepler: from Pebbles to Exoplanets*, 2023, Astronomical Society of the Pacific Conference Series, 534, 717

Flasseur, O., Denis, L., Thiébaut, É., & Langlois, M., *Exoplanet detection in angular differential imaging by statistical learning of the nonstationary patch covariances. The PACO algorithm*, 2018, A&A, 618, A138

Gomez Gonzalez, C. A., et al., *Low-rank plus sparse decomposition for exoplanet detection in direct-imaging ADI sequences. The LLSG algorithm*, 2016, A&A, 589, A54

Grady, C. A., et al., *The Disk and Environment of a Young Vega Analog: HD 169142*, 2007, ApJ, 665, 1391

—, *Revealing the Structure of a Pre-Transitional Disk: The Case of the Herbig F Star SAO 206462 (HD 135344B)*, 2009, ApJ, 699, 1822

Gratton, R., et al., *Blobs, spiral arms, and a possible planet around HD 169142*, 2019, A&A, 623, A140

Greenberg, R., Hartmann, W. K., Chapman, C. R., & Wacker, J. F., *The Accretion of Planets from Planetesimals*, 1978, 599

Guerri, G., et al., *Apodized Lyot coronagraph for SPHERE/VLT: II. Laboratory tests and performance*, 2011, ExA, 30, 59

Haffert, S. Y., et al., *Two accreting protoplanets around the young star PDS 70*, 2019, NatAs, 3, 749

Hammond, I., et al., *Confirmation and Keplerian motion of the gap-carving protoplanet HD 169142 b*, 2023, MNRAS, 522, L51

Hashimoto, J., et al., *Accretion Properties of PDS 70b with MUSE*, 2020, AJ, 159, 222

Homma, K., & Nakamoto, T., *Collisional Growth of Icy Dust Aggregates in the Disk Formation Stage: Difficulties for Planetesimal Formation via Direct Collisional Growth outside the Snowline*, 2018, ApJ, 868, 118

Isella, A., et al., *Millimeter Imaging of MWC 758: Probing the Disk Structure and Kinematics*, 2010, ApJ, 725, 1735

Johansen, A., et al., *The Multifaceted Planetesimal Formation Process*, 2014, 547

Juillard, S., Christiaens, V., & Absil, O., *Inverse-problem versus principal component analysis methods for angular differential imaging of circumstellar disks. The mustard algorithm*, 2023, A&A, 679, A52

Juillard, S., et al., *Combining reference-star and angular differential imaging for high-contrast imaging of extended sources*, 2024, A&A, 688, A185

Kabdiyev, A., *Performance comparison of neural networks in gravitational lensing detection*, 2023, Scientific Journal of Astana IT University, 5

Kataoka, A., Tanaka, H., Okuzumi, S., & Wada, K., *Static compression of porous dust aggregates*, 2013

Keppler, M., et al., *Discovery of a planetary-mass companion within the gap of the transition disk around PDS 70*, 2018, A&A, 617, A44

Klessen, R. S., & Glover, S. C. O., *The First Stars: Formation, Properties, and Impact*, 2023, ARA&A, 61, 65

Kokubo, E., & Ida, S., *Oligarchic Growth of Protoplanets*, 1998, Icarus, 131, 171

Kraus, A. L., & Ireland, M. J., *LkCa 15: A Young Exoplanet Caught at Formation?*, 2012, ApJ, 745, 5

Lafrenière, D., et al., *A New Algorithm for Point-Spread Function Subtraction in High-Contrast Imaging: A Demonstration with Angular Differential Imaging*, 2007, ApJ, 660, 770

Lecavelier des Etangs, A., & Lissauer, J. J., *The IAU working definition of an exoplanet*, 2022, New A Rev., 94, 101641

Lindegren, L., & Dravins, D., *The fundamental definition of "radial velocity"*, 2003, A&A, 401, 1185

Marois et al., C., *Angular Differential Imaging: A Powerful High-Contrast Imaging Technique*, 2006, ApJ, 641, 556

Mawet, D., Riaud, P., Absil, O., & Surdej, J., *Annular Groove Phase Mask Coronagraph*, 2005, ApJ, 633, 1191

Mawet, D., et al., *L'-band AGPM vector vortex coronagraph's first light on VLT/NACO. Discovery of a late-type companion at two beamwidths from an F0V star*, 2013, A&A, 552, L13

Mayor, M., et al., *51 Pegasi*, 1995, IAU Circ., 6251, 1

Meeus, G., et al., *Observations of Herbig Ae/Be stars with Herschel/PACS. The atomic and molecular contents of their protoplanetary discs*, 2012, A&A, 544, A78

Mesa, D., et al., *VLT/SPHERE exploration of the young multiplanetary system PDS70*, 2019, A&A, 632, A25

Metchev, S. A., Hillenbrand, L. A., & Meyer, M. R., *Ten Micron Observations of Nearby Young Stars*, 2004, ApJ, 600, 435

Milli, J., et al., *Impact of angular differential imaging on circumstellar disk images*, 2012, A&A, 545, A111

Müller, A., et al., *HD 135344B: a young star has reached its rotational limit*, 2011, A&A, 530, A85

—, *Orbital and atmospheric characterization of the planet within the gap of the PDS 70 transition disk*, 2018, A&A, 617, L2

Okuzumi, S., Tanaka, H., Kobayashi, H., & Wada, K., *Rapid Coagulation of Porous Dust Aggregates outside the Snow Line: A Pathway to Successful Icy Planetesimal Formation*, 2012, ApJ, 752, 106

Pairet, B., et al., *STIM map: detection map for exoplanets imaging beyond asymptotic Gaussian residual speckle noise*, 2019, MNRAS, 487, 2262

Pairet, B., Cantalloube, F., & Jacques, L., *Reference-less algorithm for circumstellar disks imaging*, 2018, arXiv, arXiv:1812.01333

—, *MAYONNAISE: a morphological components analysis pipeline for circumstellar discs and exoplanets imaging in the near-infrared*, 2021, MNRAS, 503, 3724

Pearson, K., *Note on Regression and Inheritance in the Case of Two Parents*, 1895, Proceedings of the Royal Society of London Series I, 58, 240

Pearson, K., *LIII. On lines and planes of closest fit to systems of points in space*, 1901, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2, 559

Perez, S., et al., *CO Gas Inside the Protoplanetary Disk Cavity in HD 142527: Disk Structure from ALMA*, 2015, ApJ, 798, 85

Pinte, C., et al., *Kinematic Structures in Planet-Forming Disks*, 2023, Astronomical Society of the Pacific Conference Series, 534, 645

Pinte, C., et al., *Kinematic Evidence for an Embedded Protoplanet in a Circumstellar Disk*, 2018, ApJ, 860, L13

—, *Kinematic detection of a planet carving a gap in a protoplanetary disk*, 2019, NatAs, 3, 1109

Pueyo, L., *Detection and Characterization of Exoplanets using Projections on Karhunen Loeve Eigenimages: Forward Modeling*, 2016, ApJ, 824, 117

Quanz, S. P., et al., *Gaps in the HD 169142 Protoplanetary Disk Revealed by Polarimetric Imaging: Signs of Ongoing Planet Formation?*, 2013, ApJ, 766, L2

Ren, B., et al., *A Decade of MWC 758 Disk Images: Where Are the Spiral-arm-driving Planets?*, 2018, ApJ, 857, L9

Ren, B. B., et al., *Protoplanetary disks in K_s-band total intensity and polarized light*, 2023, A&A, 680, A114

Riaud, P., et al., *Coronagraphic imaging of three weak-line T Tauri stars: evidence of planetary formation around PDS 70*, 2006, A&A, 458, 317

Ruane, G., et al., *Reference Star Differential Imaging of Close-in Companions and Circumstellar Disks with the NIRC2 Vortex Coronagraph at the W. M. Keck Observatory*, 2019, AJ, 157, 118

Safronov, V. S. 1969, Evoliutsiia doplanetnogo oblaka i obrazovanie Zemli i planet[1]

[1]Accessed through the translation from the Israel Program for Scientific Translations published in 1972 (Safronov [1972])

— 1972, Evolution of the protoplanetary cloud and formation of the earth and planets (Keter Publishing House)

Sallum, S., et al., *Accreting protoplanets in the LkCa 15 transition disk*, 2015, Nature, 527, 342

Sallum, S., et al., *Imaging protoplanets: observing transition disks with non-redundant masking*, 2016, Proc. SPIE, 9907, 99070D

Sanghi, A., et al., *Efficiently Searching for Close-in Companions Around Young M Dwarfs Using a Multiyear PSF Library*, 2024, AJ, 168, 215

Santerne, A., et al., *SOPHIE velocimetry of Kepler transit candidates. VII. A false-positive rate of 35% for Kepler close-in giant candidates*, 2012, A&A, 545, A76

Smith, W. H., *Spectral differential imaging detection of planets about nearby stars.*, 1987, PASP, 99, 1344

Soummer, R., Pueyo, L., & Larkin, J., *Detection and Characterization of Exoplanets and Disks Using Projections on Karhunen-Loève Eigenimages*, 2012, ApJ, 755, L28

Sparks, W. B., & Ford, H. C., *Imaging Spectroscopy for Extrasolar Planet Detection*, 2002, ApJ, 578, 543

Stolker, T., et al., *PynPoint: a modular pipeline architecture for processing and analysis of high-contrast imaging data*, 2019, A&A, 621, A59

—, *MIRACLES: atmospheric characterization of directly imaged planets and substellar companions at 4-5 μm. II. Constraints on the mass and radius of the enshrouded planet PDS 70 b*, 2020, A&A, 644, A13

—, *Shadows cast on the transition disk of HD 135344B. Multiwavelength VLT/SPHERE polarimetric differential imaging*, 2016, A&A, 595, A113

Thalmann, C., et al., *Resolving the Planet-hosting Inner Regions of the LkCa 15 Disk*, 2016, ApJ, 828, L17

Tokunaga, A. T. 2014, in Encyclopedia of the Solar System (Third Edition), third edition edn., ed. T. Spohn, D. Breuer, & T. V. Johnson (Boston: Elsevier), 1089

Vacca, W. D., & Sandell, G., *Near-infrared Spectroscopy of TW Hya: A Revised Spectral Type and Comparison with Magnetospheric Accretion Models*, 2011, ApJ, 732, 8

Wagner, K., et al., *Magellan Adaptive Optics Imaging of PDS 70: Measuring the Mass Accretion Rate of a Young Giant Planet within a Gapped Disk*, 2018, ApJ, 863, L8

Wang, J. J., et al., *Keck/NIRC2 L'-band Imaging of Jovian-mass Accreting Protoplanets around PDS 70*, 2020, AJ, 159, 263

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P., *Image Quality Assessment: From Error Visibility to Structural Similarity*, 2004, IEEE Transactions on Image Processing, 13, 600

Wolszczan, A., & Frail, D. A., *A planetary system around the millisecond pulsar PSR1257 + 12*, 1992, Nature, 355, 145

Xie, C., et al., *Spiral Arm Pattern Motion in the SAO 206462 Protoplanetary Disk*, 2021, ApJ, 906, L9

—, *Reference-star differential imaging on SPHERE/IRDIS*, 2022, A&A, 666, A32

Zhou, Y., et al., *Hubble Space Telescope UV and Hα Measurements of the Accretion Excess Emission from the Young Giant Planet PDS 70 b*, 2021, AJ, 161, 244