

Near Infra-Red Focal Plane Wavefront Sensing

Rafael Machado Salgueiro

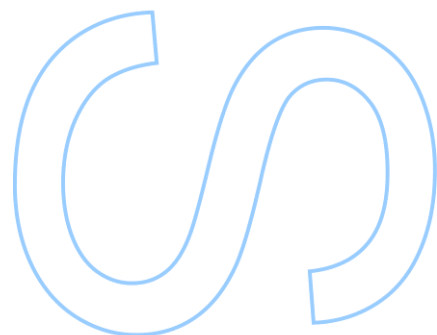
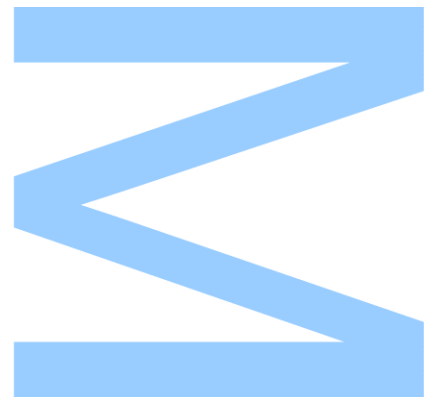
Master in Engineering Physics

Department of Physics and Astronomy

Faculty of Sciences of the University of Porto

Faculty of Engineering of the University of Porto

2024



Near Infra-Red Focal Plane Wavefront Sensing

Rafael Machado Salgueiro

Project report carried out as part of the Master
in Engineering Physics

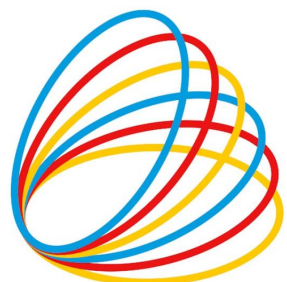
Department of Physics and Astronomy
2024

Supervisor

Dr. Carlos Correia, Faculty of Engineering

Co-supervisor

Prof. Paulo Garcia, Faculty of Engineering



centra

Dedicated to my family

Acknowledgments

Em primeiro lugar quero agradecer à minha família por todo o apoio durante os meus cinco anos de estudos de Engenharia Física. Obrigado por terem estado sempre presentes e pelo vosso apoio incondicional.

Quero também agradecer aos meus orientadores, Doutor Carlos Correia e Professor Paulo Garcia por todas as recomendações e conselhos dados ao longo deste projeto. Agradecer também por todos os conselhos "extra" ciência que me foram muito úteis durante este projeto e certamente o continuarão a ser no meu futuro profissional.

Agradecer também a todos os outros membros da equipa de investigação na qual estive inserido neste projeto. Obrigado por toda a vossa ajuda e pela atitude acolhedora desde o primeiro dia.

Por último, agradeço também a todos os amigos e professores que me acompanharam durante esta jornada.

Resumo

Contexto: Quando se realizam observações astronómicas em telescópios terrestres, a correção da turbulência requer uma estrela suficientemente brilhante próxima do alvo científico. No entanto, o número de estrelas-guia naturais (NGS) suficientemente brilhantes é bastante limitado, o que limitaria bastante a cobertura do céu. Para aumentar a cobertura do céu dos telescópios terrestres, estrelas-guia laser (LGS) são usadas como solução. Infelizmente, as LGS não fornecem uma solução completa, pois a detecção de frente de onda (WFS) de LGS é insensível a medições de tip/tilt (TT) e a medição de foco é comprometida pela variabilidade da altitude da camada de sódio (Na). Consequentemente, as medições de TT e seguimento do foco têm de ser feitos com uma NGS ténue próxima ao alvo científico onde a LGS está apon-tada. Os sensores de frente de onda no plano focal (FPWFS) são uma solução atraente para esse problema, pois são uma técnica de pupila completa que usa eficientemente todo o fluxo da NGS. Isso significa que estrelas muito mais ténues podem ser usadas para medições de TT e seguimento de foco, aumentando ainda mais a cobertura do céu atualmente alcançada com sistemas LGS.

Objetivos: Este estudo teve como objetivo explorar o desempenho de diferentes técnicas de FPWFS para medir modos de baixa ordem, como TT e foco, no infravermelho próximo (NIR). Foi dada atenção especial ao problema de seguimento de foco, pois este é o principal problema operacional em sistemas LGS. Estes algoritmos foram estudados em condições de trabalho desafiantes, porém realistas, ou seja, baixo fluxo (estrelas ténues) e resíduos de alta ordem, já que a NGS geralmente é encontrada a uma considerável separação angular da LGS e amostragem sub-Nyquist. Essencialmente, o principal objetivo desta pesquisa foi avaliar o potencial de diferentes técnicas de FPWFS para uma implementação real em um sistema LGS-AO para seguimento de foco lento.

Métodos: Três algoritmos de FPWFS foram explorados, nomeadamente, Gerchberg-Saxton (GS), a Linearized Focal Plane Technique (LiFT) e um algoritmo de ajuste, que chamei de Full Width at Half Maximum (FWHM) fit. Todos estes algoritmos pertencem à categoria de recuperação de fase com diversidade de fase, também conhecida como técnicas de diversidade de fase de imagem única. Estes algoritmos são capazes de extrair a fase aberrada, nomeadamente o modo de foco, de uma imagem de uma fonte pontual (point spread function) no plano focal. Além da fase aberrada, uma aberração adicional é intencionalmente adicionada, ou seja, a diversidade de fase, que é necessária para resolver a ambiguidade de sinal dos modos pares, como o foco. Os testes foram conduzidos por meio de simulações numéricas e também

com dados reais do telescópio Keck. O algoritmo de ajuste FWHM foi testado apenas com dados reais (de bancada e céu), pois foi considerado já numa fase avançada do trabalho. As simulações numéricas foram feitas para condições idealistas e realistas, incluindo condições de baixo fluxo (estrelas tênues), ruído, resíduos de alta ordem e amostragem sub-Nyquist.

Resultados: Simulações numéricas mostraram que GS e LiFT são capazes de realizar detecção de modos de baixa ordem (TT e foco) em condições realistas com erros de fase abaixo de 0.4 rad rms. O algoritmo GS atual utilizado não está totalmente preparado para lidar com imagens sub-amostradas. Como consequência, quando aplicamos GS com imagens sub-amostradas, é necessário aplicar um ajuste de inclinação considerável. No entanto, a resposta linear do GS para medições de foco é bastante impressionante e robusta. Estimou-se que para telescópios de classe de 10 metros de diâmetro, as correções de foco precisam ser feitas aproximadamente a cada 15 segundos para manter o erro de foco abaixo de 30 nm rms. Ainda assim, essa cadência de correção não é atualmente alcançável no Keck devido a restrições mecânicas, e as correções de foco terão que ser feitas aproximadamente a cada 30 segundos. Isso significa que teremos que lidar com erros de foco em torno de 50 nm rms. Todos os algoritmos são rápidos o suficiente nesses intervalos de tempo, o que significa que os custos computacionais não são um problema. Os resultados com dados reais também provaram a aplicabilidade de GS, LiFT e ajuste FWHM para o seguimento de foco.

Conclusões: Neste trabalho, demonstrei todos os diferentes procedimentos de pré processamento, calibração, operação em tempo real necessários para uma implementação no céu com GS, LiFT e o ajuste FWHM para seguimento de foco lento. Após este processo de maturação, estamos mais próximos de testar um destes algoritmos no céu. As técnicas de FPWFS estudadas mostraram potencial para realizar seguimento de foco em um contexto de sistema LGS-AO. No entanto, no geral, LiFT é o que apresenta melhor desempenho. O GS também mostra um comportamento linear impressionante, mas como requer um ajuste considerável de ganho óptico, é relegado em relação ao LiFT. O ajuste FWHM também apresentou bom desempenho com dados reais, mas com um regime de linearidade mais baixo.

Palavras-chave: Sensores de Frente de Onda no Plano Focal, Algoritmos de otimização não-linear, Óptica adaptativa

Abstract

Context: When performing astronomical observations in ground-based telescopes, turbulence correction requires a sufficiently bright star near the scientific target. However, the number of sufficiently bright natural guide stars (NGS) is quite limited, which would highly limit sky coverage. In order to increase the sky coverage of ground-based telescopes, laser guide stars (LGS) are deployed as a solution. Unfortunately, LGS does not provide a complete solution since LGS wavefront sensing (WFS) is insensitive to tip/tilt (TT), and focus measurements are compromised by the variability of the sodium (Na) layer altitude. Consequently, TT measurements and focus tracking must be done with a faint NGS near the scientific target where the LGS is pointed. Focal plane wavefront sensing (FPWFS) is an attractive solution to this problem since it is a full-pupil technique that efficiently uses the whole NGS flux. This means that much fainter stars can be used for TT measurements and focus tracking, further increasing the current sky coverage achieved with LGS systems.

Aims: This study aimed to explore the performance of different FPWFS techniques for measuring low-order modes such as TT and focus in the near-infrared (NIR). Special attention was given to the focus tracking problem since this is the main operational issue in LGS-AO systems. These algorithms were studied under challenging working conditions, yet realistic ones, namely low flux (faint stars) and high-order residuals, since the NGS is usually only found at considerable angular separation from the LGS and sub-Nyquist sampling, which is usually present for the sake of sky coverage. In essence, the main goal of this research was to assess the potential of different FPWFS techniques for an on-sky deployment in a real LGS-AO system for slow focus tracking.

Methods: Three different FPWFS algorithms were explored, namely Gerchberg-Saxton (GS), the Linearized Focal Plane Technique (LiFT), and a fitting algorithm, which I called the Full Width at Half Maximum (FWHM) fit. All these algorithms fall within the category of phase retrieval with phase diversity, also known as single image phase diversity techniques. These algorithms are capable of extracting the aberrated phase, namely the focus mode, from a single point source image (point spread function) at the focal plane. In addition to the aberrated phase, an additional aberration is intentionally added, i.e., the phase diversity, which is necessary to solve for the sign ambiguity of the even modes such as focus. Tests were conducted through numerical simulations and also with real data from the Keck telescope. The FWHM fit algorithm was only tested with real data (bench and on-sky) since it was only considered at a late stage of the work. The numerical simulations were done for idealistic and realistic conditions, including

low-flux conditions (faint stars), noise, high-order residuals, and sub-Nyquist sampling.

Results: Numerical simulations showed that GS and LiFT are capable of performing low-order mode (TT and focus) sensing under realistic conditions with rms phase errors below 0.4 rad rms. The current GS algorithm used is not fully prepared to deal with sub-Nyquist sampled images. As a consequence, when we apply GS with under-sampled images, a considerable slope adjustment needs to be applied. However, the linear response of GS for defocus measurements is quite impressive and robust. It was estimated that for 10-m diameter class telescopes, the focus corrections need to be approximately every 15 seconds to keep the defocus error below 30 nm rms. Still, this correction cadency is not currently achievable in Keck due to mechanical constraints, and focus corrections will have to be done approximately every 30 seconds. This means we'll have to deal with focus errors around 50 nm rms. All the algorithms perform fast enough in these kinds of time frames, meaning that the computational costs are not a problem. The results with real data also proved the applicability of GS, LiFT, and FWHM fit for focus tracking.

Conclusions: In this work, I showed and demonstrated all the different preprocessing, calibration, algorithmic, and real-time operation procedures necessary for an on-sky deployment with GS, LiFT, and the FWHM fit for slow focus tracking. After this maturation process, we are closer to testing one of these algorithms on sky. The FPWFS techniques studied showed the potential to perform focus tracking in an LGS-AO system context. However, overall, LiFT is the one that performs better. GS also shows an impressive linear performance, but since it requires a considerable optical gain adjustment, it is relegated in relation to LiFT. The FWHM fit also performed well with real data but with lower linearity ranges.

Keywords: Focal Plane Wavefront Sensing, Non-linear optimization algorithms, Adaptive optics

Contents

List of Figures	xi
List of Tables	xii
List of Abbreviations	xiii
1 Introduction	1
1.1 Context	1
1.2 Research Goals	2
1.3 Document Structure	2
2 State of the Art	4
2.1 Imaging through Turbulence	4
2.1.1 Optical Image formation	5
2.1.2 Angular Resolution	7
2.1.3 Zernike Polynomials and Wavefront representation	9
2.1.4 Image quality metrics	11
2.2 Adaptive Optics Systems	12
2.2.1 Main Components and Operation	12
2.2.2 Wavefront Sensing	13
2.3 Sky Coverage	15
2.3.1 Anisoplanatism	15
2.3.2 Laser Guide Stars	17
2.3.3 Tip-Tilt estimation and focus tracking	20
2.4 Focal Plane Wavefront Sensing	22
2.4.1 Motivation and associated Challenges	22
2.4.2 Sign ambiguity for even modes and phase diversity	25
2.4.3 Small-phase solution	27
2.4.4 Classification of Focal Plane Wavefront Sensing Techniques	29
3 Methods	32
3.1 Gerchberg-Saxton	32
3.2 Linearized Focal Plane Technique	33
3.3 Fitting methods	36

4	Results	39
4.1	Numerical simulations	39
4.1.1	Tests without noise	39
4.1.2	Tests with noise and no high-order residuals	47
4.1.3	Tests with noise and high-order residuals	49
4.1.4	Focus ramps	53
4.1.5	Close loop simulation: towards an on-sky implementation	60
4.2	Tests with real data	67
4.2.1	Results with bench data	67
4.2.2	Results with on-sky data	77
5	Conclusions and Outlook	83
5.1	Future steps and recommendations	84
6	Attachments	86
6.1	Proof of relationship between Zernike coefficients and wavefront error	86
6.2	Defocus error due to Sodium layer height variability	88
	References	94

List of Figures

2.1	Different atmospheric layers separated by isothermal boundaries.	4
2.2	Illustration of Fraunhofer diffraction for a circular aperture and the Airy pattern. . .	6
2.3	Illustration of the effects of AO on a point source image, i.e. PSF.	7
2.4	Separability of two point sources of equal intensity in diffraction limit.	8
2.5	First 15 Zernike polynomials according to Noll scheme.	10
2.6	Visual representation of Zernike polynomials.	10
2.7	Adaptive optics (AO) system scheme.	12
2.8	Illustration of Shack-Hartmann's wavefront sensor working principle.	14
2.9	Angular anisoplanatism illustration	16
2.10	Illustration of sky coverage as a function of the Strehl ratio for K-band with NGS and LGS.	18
2.11	Comparison of the cone effect between Rayleigh and Na LGS and example of LGS deployment at VLT.	19
2.12	Cone effect compensation by applying several LGS.	19
2.13	Laser Tomographic Adaptive Optics system.	20
2.14	Illustration of laser guide star insensitivity to Tip/Tilt and focus measurement problems due to Na layer mean height variability.	21
2.15	Illustration of local minima problem associated with non-linear optimization algo- rithms.	24
2.16	Illustration of the sign ambiguity for even modes.	26
2.17	Defocus sign ambiguity lifted with the usage of astigmatism as phase diversity. . .	27
2.18	Classification of inverse imaging problems.	30
2.19	Optical setup of a phase retrieval system.	30
2.20	Typical optical layout of a joint estimation system.	31
3.1	Illustration of Gerchberg-Saxton algorithm.	32
3.2	Illustration of how a Gaussian fit can be used to estimate defocus.	38
4.1	Random generation of Zernike coefficients using a Gaussian distribution with mean zero and standard deviation σ_n	40
4.2	GS and LiFT performance for a 1.0 Nyquist sampling without noise or high-order residuals.	42
4.3	Mean iterations ($\langle N_{it} \rangle$) cost for GS according to different amounts of phase diversity applied.	43

4.4	Comparison between the computational costs of GS and LiFT.	43
4.5	GS performance without noise or high-order residuals for 2.0 Nyquist sampling. . .	44
4.6	Performance of GS and LiFT for 0.5 Nyquist sampling without noise or high-order residuals.	45
4.7	LiFT performance for 0.25 Nyquist sampling without noise or high-order residuals.	46
4.8	GS performance for high-order estimation with 1.0 and 0.5 Nyquist sampling. . .	47
4.9	GS and LiFT performance according to the presence of noise with 0.5 Nyquist sampling.	48
4.10	Effects of noise in the performance of GS.	49
4.11	GS and LiFT performance on the estimation of TT and focus for 0.5 Nyquist sampling, noise, and high-order residuals.	50
4.12	GS and LiFT performance on the estimation of the first ten modes for 0.5 Nyquist sampling, noise, and high-order residuals.	51
4.13	GS performance according to the different number of iterations.	51
4.14	LiFT performance on the estimation of TT and focus by allowing a higher number of modes to be estimated.	52
4.15	LiFT performance on the estimation of TT and focus by allowing a higher number of modes to be estimated. The maximum a posteriori (MAP) estimator was used.	53
4.16	Illustration of full width at half maximum (FWHM) of the PSF for positive and negative values of defocus introduced.	54
4.17	Focus ramps for GS and LiFT for a sampling of 1 Nyquist without noise or high-order residuals. An amount of 200 nm rms of defocus was used as phase diversity.	55
4.18	Focus ramps for GS and LiFT for a sampling of 1 Nyquist without noise or high-order residuals. An amount of 200 nm rms of astigmatism was used as phase diversity.	56
4.19	Illustration of attained defocus linearity ranges with GS for different amounts of astigmatism introduced.	56
4.20	Crosstalk effects between defocus and astigmatism with GS.	57
4.21	Crosstalk effects in LiFT when estimating 5 modes (including the term corresponding to the diversity).	58
4.22	Focus ramps for GS and LiFT for 0.5 Nyquist sampling with photon noise and 100 nm rms of high-order residuals.	59
4.23	Focus ramps for GS and LiFT for 0.5 Nyquist sampling with photon noise and 200 nm rms of high-order residuals.	59

4.24 Focus ramps for LiFT for 0.5 Nyquist sampling with photon noise and 200 nm rms of high-order residuals. Performance comparison for when the diversity mode is estimated or not.	60
4.25 Random Sodium layer altitude variation time series according to theoretical PSD.	61
4.26 Simulation of predictable defocus error as a function of the correction cadency of the slow focus in Keck telescope.	62
4.27 Illustration of rolling average for focus tracking.	63
4.28 SNR study simulating a 50 mas sampled PSF as happens with the TRICK sensor.	64
4.29 Defocus error as a function of the correction cadency period (Δt_{CC}) for the closed-loop simulation.	65
4.30 Focus tracking example with LiFT and GS with $\Delta t_{CC} = 2.0$ s.	66
4.31 Results with rolling average for $\Delta t_{CC} = 2.0$ s and $\Delta t_{CC} = 5.0$ s.	67
4.32 Focus tracking example with LiFT and GS with $\Delta t_{CC} = 2.0$ s and $\Delta t_{LE} = 10.0$ s. Rolling average was performed.	67
4.33 Illustration of TRICK PSF for a specific data file (band K and $m = 12$) with defocus values of opposite signs.	69
4.34 Preprocessing steps used with bench data (applied to all images).	70
4.35 Example of mean in-focus image for a specific data file (band K and $m=12$). In-focus images are used to perform diversity calibration.	71
4.36 Illustration of the calibration process for GS. Results obtained with Keck bench data for band K with 11.2 of magnitude.	72
4.37 Results obtained with Keck bench data for band K with 11.2 of magnitude.	72
4.38 Results obtained with Keck bench data for band K with 12.3 of magnitude.	73
4.39 Results obtained with Keck bench data for band K with 13.8 of magnitude.	73
4.40 Example of raw image highly consumed by noise for the file corresponding to band K and $m=13.8$	74
4.41 Results obtained with Keck bench data for band H with 8.6 of magnitude.	74
4.42 Results obtained with Keck bench data for band H with 13.1 of magnitude.	75
4.43 Results obtained with Keck bench data for band H with 11.0 of magnitude.	75
4.44 Mean images corresponding to defocus offsets of -250, 0, and +250 nm rms for on-sky data.	77
4.45 Results with the FWHM fit with on-sky data.	78
4.46 Results of GS and LiFT for on-sky data using short-exposure PSFs.	79
4.47 Results for GS and LiFT using LE PSFs.	80

4.48 Mean images corresponding to defocus offsets of -250, 0, and +250 nm rms with on-sky data (4*4 pixels images).	81
4.49 Results for Trick on-sky data from 2020 (4*4 images) with GS and LE GS. No calibration was done.	82

List of Tables

2.1	Advantages and disadvantages associated with FPWS techniques.	25
4.1	Trick sensor characteristics.	68
4.2	Angular pixel sizes in mas corresponding to 1.0, 0.5, and 0.25 Nyquist sampling both for band H ($\lambda = 1.654 \mu\text{m}$) and band K ($\lambda = 2.179 \mu\text{m}$). The telescope's diameter used for these calculi was 11.732 m (the same as Keck).	68
4.3	Time consumption in seconds for on-sky data considering short and long exposures for a round pupil with 64*64 pixels.	80
4.4	Time consumption in seconds for on-sky data considering short and long exposures for a round pupil with 32*32 pixels.	80

List of Abbreviations

AO	Adaptive Optics
DM	Deformable Mirror
FPWFS	Focal Plane Wavefront Sensing
FWHM	Full Width Half Maximum
GS	Gerchberg-Saxton
LGS	Laser Guide Star
LiFT	Linearized Focal Plane Technique
NCPA	Non-Common Path Aberrations
NGS	Natural Guide Star
PSF	Point Spread Function
SNR	Signal-to-noise Ratio
S	Strehl ratio
TT	Tip/Tilt
WFS	Wavefront Sensing/Sensor

1 Introduction

1.1 Context

Ground-based astronomical telescopes must surpass several challenges in order to image the desired scientific targets efficiently. One of those challenges is optical turbulence provoked by the atmosphere (turbulent medium), which highly degrades the angular resolution of the obtained images, which is highly undesirable for scientific purposes. Optical turbulence is a consequence of the different atmospheric temperatures along the light's path, which results in variations in the index of refraction. To overcome this problem, adaptive optics (AO) systems are deployed to compensate for the wavefront distortions of the light provoked by turbulence. Essentially, an AO system measures the aberrated wavefront and then sends commands to a deformable mirror (DM) that is adjusted to compensate for the aberrated wavefront.

Another big challenge associated with ground-based telescopes is the availability of sufficiently bright natural guide stars (NGS) near the object of interest. These guide stars are necessary to perform wavefront corrections with the AO systems and they must be sufficiently bright; otherwise, low signal-to-noise ratios can severely affect the AO system operation. The problem is that the availability of sufficiently bright NGS is quite limited, meaning a low sky coverage would be achieved with ground-based telescopes.

In order to increase the sky coverage, laser guide stars (LGS) are deployed as a solution. These artificial beacons are created at approximately 90 km of altitude in the sodium (Na) layer, where Na atoms are excited by a laser tuned specifically to 589 nm. When the Na atoms return to the ground state, they emit the desired backscattered radiation. Unfortunately, LGS wavefront sensing (WFS) is insensitive to tip/tilt (TT), and the variability of the sodium layer altitude in the atmosphere compromises focus estimation.

Consequently, LGS-AO systems need to perform TT measurements and focus tracking using a faint NGS. Since we are dealing with very faint stars, focal plane wavefront sensing (FP-WFS) techniques are an extremely appealing solution because they are full-pupil techniques, meaning that the whole star flux is used for wavefront sensing. The idea behind FPWFS is then to extract TT and focus from a single image at the focal plane of a point source (star). The image of a point source is known as the point spread function (PSF). This type of FPWFS, where the phase is extracted from a single image at the focal plane, is usually known as phase retrieval, and this will be the main FPWS type discussed in this work.

The presence of TT provokes a change in the center of gravity of the PSF. For that rea-

son, TT can be efficiently estimated using center-of-gravity-like algorithms (also an FPWFS technique), which are known to work quite efficiently. From an operational point of view, focus tracking is a more challenging task, and for that reason, it will be particularly explored in this work. The first challenge associated with it is that the PSF manifests itself equally for defocus values with different signs but with the same absolute value. This is usually known as sign ambiguity, and it is not exclusive to defocus; it occurs for all even modes. However, this sign ambiguity can be lifted by using phase diversity, where a known amount of a certain aberration is applied allowing to break the degeneracy for even modes, such as defocus.

1.2 Research Goals

This research is specifically oriented towards LGS-AO systems for ground-based telescopes, where the measurement of TT and focus tracking need to be done with an NGS. In this work, different FPWFS techniques are explored through numerical simulations and tests with real data. Specific attention will be given to focus measurements since it is currently the main operational issue in LGS-AO systems.

The main goal of the research that will be presented is to assess the performance of different FPWFS algorithms for slow focus measurements and analyze what might be the best solution for on-sky deployments on real LGS-AO systems. The three FPWFS algorithms explored were Gerchberg-Saxton (GS), the linearized focal plane technique (LiFT), and also a fitting algorithm, which I called the Full width at Half Maximum (FWHM) fit. The FWHM fit was only used with real data tests as it was not initially considered for the study, but due to its simple algorithm and implementation, it was decided to explore its performance with real data.

1.3 Document Structure

This first section was dedicated to giving a short introduction about the research work that will be presented, its motivation, and its main goals. After this brief presentation, I will give a literature revision about some general topics of AO, describe the motivation and associated problems related to the usage of LGS for ground-based telescopes, and give a general review of FPWFS techniques with special attention to how they can be used to overcome the limitations of LGS related to TT measurements and focus tracking. Next, in the methods section, I will describe in detail the FPWFS algorithms that were studied in this work, namely GS, LiFT, and the FWHM fit.

In the results section, I will present all the results obtained in the study with the different FPWFS algorithms. There are two different sets of results: the ones obtained through numerical

simulations and the ones obtained with real data. In the numerical simulations, only results for GS and LiFT will be presented since the FWHM fit was only considered in an advanced stage of this work, where it was tested only with real data. In the numerical simulations, GS and LiFT were studied under ideal and also realistic conditions. The four main aspects that were considered for simulating real conditions were low-flux conditions (faint NGS), noise, high-order residuals, and sub-Nyquist sampling, which are the expected working conditions in a real scenario. After all the numerical simulations' results, tests with real will be presented. Data were provided by the Keck Observatory at Mauna Kea. These tests were done through a collaboration with Keck Observatory, for which a new slow focus tool is being studied. After presenting all the results, I will finally present all the main conclusions of this work and present some future recommendations.

2 State of the Art

2.1 Imaging through Turbulence

Earth's atmosphere poses one of the biggest challenges for ground-based telescopes since it distorts light's wavefront and, as a consequence, degrades the quality of astronomical images, which is highly undesirable for scientific purposes. Another important limitation related to the atmosphere for astronomical purposes is that it is opaque to radiation at wavelengths below $0.3 \mu m$ and above $30 \mu m$, although it becomes transmissive to radio waves longer than approximately 1 mm. On the other hand, its transparency is good between 0.3 and $1.0 \mu m$, covering the ultra-violet (UV), visible, and near-infrared (NIR). Between 1 and $10 \mu m$, the atmospheric transmissivity varies, but there are good NIR windows at $1.25 \mu m$ (J-band), $1.6 \mu m$ (H-band), and $2.2 \mu m$ (K-band), which are very important in AO [1].

In Fig. 2.1, the different atmospheric layers (based on temperature variations) are presented along with their isothermal boundaries. These temperature variations along the light's optical path result in refraction-index variations (optical turbulence), highly affecting the light's wavefront [2]. Refractive-index fluctuations' amplitude depends not only on the range of temperature variations but also on the air's density. Air density decays exponentially with height, which means that optical turbulence is usually greater near ground and falls exponentially with altitude [1]. This is actually one of the reasons for building ground-based telescopes on top of mountains at high altitudes, where most ground turbulence is avoided.

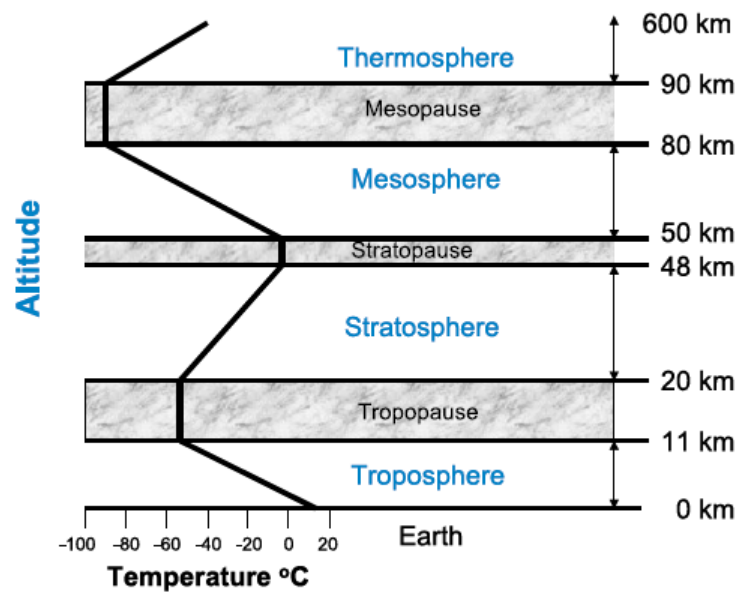


Figure 2.1: Different atmospheric layers separated by isothermal boundaries [2].

The Fried parameter (r_0) is commonly used as a quantitative measure of optical turbulence. It is defined as the diameter over which the optical phase distortion has a mean-square value of 1 rad^2 at a wavelength of $0.5 \mu\text{m}$, and it is given by [3]:

$$r_0 = \left[0.423 k^2 \sec(\zeta) \int_h C_n^2(h) dh \right]^{-3/5}, \quad (2.1)$$

where $k = \frac{2\pi}{\lambda}$ with λ being the wavelength, ζ is the zenith angle, and $C_n^2(h)$ is the turbulence profile or structure parameter (units of $\text{m}^{-2/3}$) with altitude (h). A high value of r_0 is a synonym for good seeing (weak turbulence), whereas low values of r_0 mean bad seeing (high turbulence). As a general rule, an r_0 value of 5 cm is considered poor seeing, while an r_0 of 20 cm is considered good seeing [4].

2.1.1 Optical Image formation

In Fourier Optics, an imaging system is linear and time-invariant (LTI). The system is linear in complex amplitude for spatial coherent imaging systems, whereas, for spatial incoherent systems, the system is linear in intensity [5]. Since the reference sources used in AO emit incoherent radiation [1], only the formalism for incoherent systems will be presented here.

For an LTI system, the output is given by the convolution between the input and the system's impulse response. For incoherent radiation, we then have [6]:

$$I_i = PSF * I_0, \quad (2.2)$$

where I_i and I_0 are the image and object irradiance respectively, $*$ represents the convolution operator, and PSF is the point spread function, the impulse response for incoherent illumination, which is given by [5]:

$$PSF = |\mathcal{F}\{U e^{i\Phi}\}|^2, \quad (2.3)$$

with $U e^{i\Phi}$ being the electric field at the pupil plane, and $\mathcal{F}$ denotes the Fourier transform operator.

According to Eq. 2.2, I_i would be the same as I_0 if the PSF were a Dirac function, meaning an absolutely perfect imaging system. However, this idealistic scenario is not physically possible since the imaging process experiences diffraction. The image of a point source is always of finite size (diffraction limit). Considering a uniform circular aperture and no phase aberrations, the PSF is given by the following expression [7]:

$$I(\theta) = I(0) \left[\frac{2J_1(ka \sin(\theta))}{ka \sin(\theta)} \right]^2, \quad (2.4)$$

where $k = \frac{2\pi}{\lambda}$ with λ being the radiation wavelength, a is the radius of the circular aperture, J_1 is a Bessel function, $I(0)$ is the irradiance at the center and θ is the angular distance from the center. This expression, commonly known as the Airy pattern, is obtained considering Fraunhofer diffraction, which is valid for most practical optical imaging systems [1].

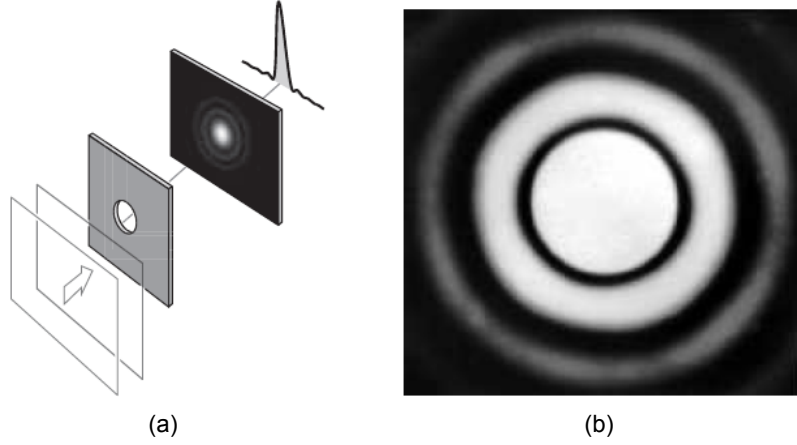


Figure 2.2: Illustration of Fraunhofer diffraction for a circular aperture (a) and the Airy pattern (b) [7].

In Fig. 2.2, we have a representation of the Fraunhofer diffraction through a circular aperture and the Airy pattern. By analyzing Eq. 2.4, we can explain the existence of several dark rings in the Airy pattern due to the zeros of the Bessel function. The first zero occurs when $\sin(\theta) = 1.22 \pi$, which means the angular distance from the center to the first dark ring (θ_1) satisfies the following relationship [7]:

$$\sin(\theta_1) = 1.22 \frac{\lambda}{D}, \quad (2.5)$$

with D being the diameter of the circular aperture. The encircled energy within this first ring contains 84 % of all energy [4]. This central spot, which is delimited by the first dark ring, is known as the Airy disk. According to Eq. 2.5, increasing the aperture's diameter reduces the size of the Airy disk. In essence, having a bigger circular aperture allows, at the same time, more energy to be collected since more light is gathered, and it also reduces the size of the Airy disk.

The diffraction-limited PSF (Airy pattern) can only occur when there are no wavefront aberrations, which is invalid for astronomical ground-based telescopes due to atmospheric turbulence. If no correction is applied, the intensity pattern with the presence of aberrations creates several speckles considering a short exposure image. For long exposure, we would obtain an

Airy-like pattern where the peak intensity is much lower and the FWHM of the central region is much bigger. More precisely, the FWHM would be equal to λ/r_0 with r_0 being the Fried's parameter, instead of λ/D , as in the diffraction limit [1].

In order to compensate for atmospheric turbulence, adaptive optics (AO) systems are deployed to improve image quality, which would otherwise be highly degraded. In Fig. 2.3, it is shown two short exposure PSFs with and without AO to illustrate the benefits of AO. Without AO, the final image (the convolution between the PSF and the object irradiance) would be extremely blurred. AO systems will be presented in more detail in section 2.2.

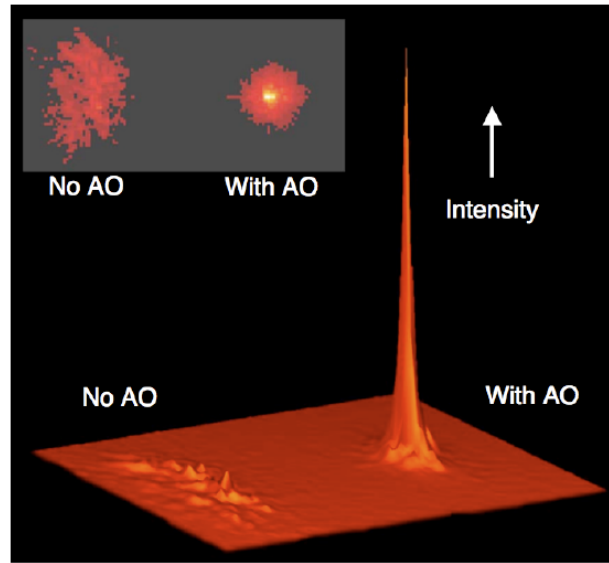


Figure 2.3: Illustration of the effects of AO on a point source image, i.e. PSF (Credit: Claire E. Max, UCSC).

2.1.2 Angular Resolution

The goal of any imaging system is to obtain images as sharp as possible, i.e., with the highest angular resolution possible. Angular resolution measures an imaging system's performance that defines its ability to distinguish between two closely spaced point sources [2]. As shown in the previous section, a point source will be imaged as a diffraction pattern, where the central core, known as the Airy disk, has a radial angular size of $\Delta\theta = 1.22 \lambda/D$ neglecting the existence of wavefront aberrations. Let's imagine the case where two point sources of equal brightness are separated by an angular distance of $\Delta\alpha$. When $\Delta\alpha \ll \Delta\theta$, the two point sources will not be resolved by the imaging system (Fig. 2.4 b)), whereas when $\Delta\alpha \gg \Delta\theta$ they will be clearly resolved (Fig. 2.4 a)).

Different criteria were proposed in order to quantify the angular resolution of an optical imaging system, i.e., the minimum angular distance between two equally bright point sources that the imaging system can resolve. Two widely known criteria are the Rayleigh and Sparrow

criteria (Fig. 2.4 c), d)). According to Rayleigh criteria, two point sources of equal brightness are resolved when each peak coincides with the first dark ring of the other. Considering a circular aperture, this happens when the angular distance between the sources is given by [8]:

$$\Delta\alpha = 1.22 \frac{\lambda}{D} \text{ (Rayleigh)} \quad (2.6)$$

According to Sparrow's criteria, the angular resolution of an imaging system is given by [9]:

$$\Delta\alpha = 1.0 \frac{\lambda}{D} \text{ (Sparrow)}. \quad (2.7)$$

In this configuration, the resulting irradiance profile has a broad flat top, in which the second derivative is zero [7].

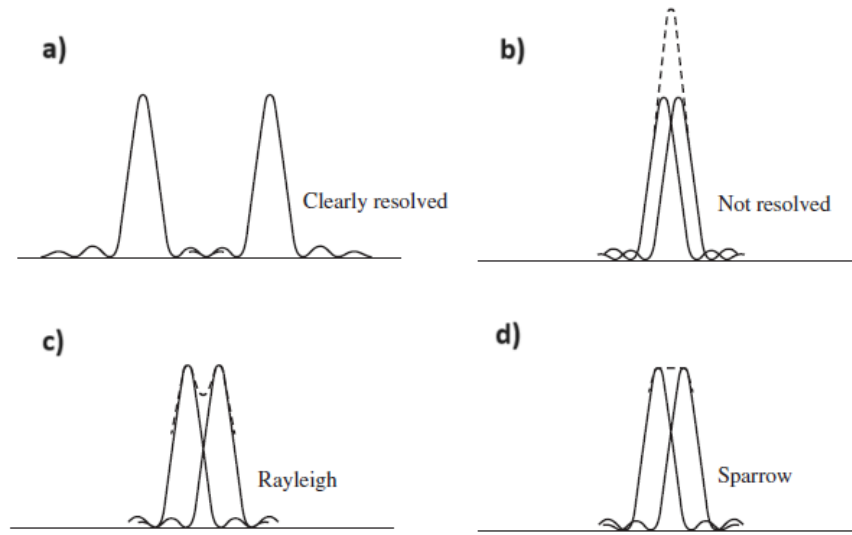


Figure 2.4: Separability of two point sources of equal intensity in diffraction limit [7].

From Eq. 2.6 and 2.7, we see that the angular resolution is higher (lower angular distance) as the diameter of the aperture increases or if we use radiation with lower wavelengths. Essentially, for some given wavelength, the angular resolution will increase with the aperture's diameter, which is one of the main reasons (the other is concerned with the larger amount of light collected) modern ground-based telescopes such as the Very Large Telescope (VLT) have considerable diameters (8-10 meters). Future telescopes will have even bigger apertures, such as the Extremely Large Telescope (ELT), which will have a primary mirror with 39.3 m of diameter. To give a more quantitative comparison, let's consider we are observing the visible at 500 nm. According to Sparrow criteria, the angular resolution for a telescope with $D = 10 \text{ m}$ is 10 mas, whereas for $D = 40 \text{ m}$, it is 2.6 mas, which is a considerably better resolution.

The previous values were obtained without considering the existence of wavefront aberrations.

tions, i.e., in the diffraction limit. As seen in the last section, the angular size for a long exposure image has an FWHM given by λ/r_0 rather than λ/D , which highly deteriorates the achievable angular resolution. Even at the best-seeing sites, the median seeing disk size is about 500 mas at visible wavelengths [10]. Without AO to compensate for wavefront aberrations, increasing the telescopes' diameters would be useless since atmospheric seeing would always limit the angular resolution.

2.1.3 Zernike Polynomials and Wavefront representation

Zernike polynomials, first introduced by Frits Zernike, have many applications in optics, namely for adaptive optics. In 1976, Noll introduced a specific normalization and ordering for these polynomials to facilitate statistical analysis of wavefront error due to atmospheric turbulence [11]. This indexing and normalization of the Zernike polynomials is widely used in AO since it facilitates the description of the wavefront error and gives a logical ordering of the modes.

Zernike polynomials are defined on a unit circle. They are most conveniently expressed in polar coordinates (ρ, θ) , where ρ is the normalized radial component ($0 \leq \rho \leq 1$) and θ is the polar angle ($0 \leq \theta \leq 2\pi$). According to the Noll scheme, these polynomials are defined as follows [11]:

$$Z_j(\rho, \theta) = Z_n^m(\rho, \theta) = \begin{cases} \sqrt{2(n+1)} R_n^m(\rho) \cos m\theta, m \neq 0, j \text{ is even}, \\ \sqrt{2(n+1)} R_n^m(\rho) \sin m\theta, m \neq 0, j \text{ is odd}, \\ \sqrt{(n+1)} R_n^m(\rho), m = 0, \end{cases} \quad (2.8)$$

where the radial polynomials $R_n^m(\rho)$ are defined as:

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s! \left(\frac{n+m}{2} - s\right)! \left(\frac{n-m}{2} - s\right)!} \rho^{n-2s} \quad (2.9)$$

We can see that these polynomials are defined as the product of normalization factors, radial polynomials, and azimuthal (angular) functions [12]. Index n is the degree of the radial polynomials, $R_n^m(\rho)$; the index m is the azimuthal frequency; n and m are non-negative integers and satisfy $n-m \geq 0$ and $n-m = \text{even}$ [12]. In Fig. 2.5, we have the formulas for the first 15 Zernike polynomials along with the aberration names used for them in optics, and in Fig. 2.6, we have a visual representation of these polynomials. Some of these aberrations will be referred to extensively during this work, namely tip/tilt, defocus, and primary astigmatisms.

j	n	m	$Z_j(\rho, \theta)$	$Z_j(x, y)$	Aberration
1	0	0	1	1	Piston
2	1	1	$2\rho \cos \theta$	$2x$	x -tilt
3	1	1	$2\rho \sin \theta$	$2y$	y -tilt
4	2	0	$\sqrt{3}(2\rho^2 - 1)$	$\sqrt{3}[2(x^2 + y^2) - 1]$	Defocus
5	2	2	$\sqrt{6}\rho^2 \sin 2\theta$	$2\sqrt{6}xy$	45° Primary astigmatism
6	2	2	$\sqrt{6}\rho^2 \cos 2\theta$	$\sqrt{6}(x^2 - y^2)$	0° Primary astigmatism
7	3	1	$\sqrt{8}(3\rho^3 - 2\rho) \sin \theta$	$\sqrt{8}y[3(x^2 + y^2) - 2]$	Primary y-coma
8	3	1	$\sqrt{8}(3\rho^3 - 2\rho) \cos \theta$	$\sqrt{8}x[3(x^2 + y^2) - 2]$	Primary x-coma
9	3	3	$\sqrt{8}\rho^3 \sin 3\theta$	$\sqrt{8}y(3x^2 - y^2)$	
10	3	3	$\sqrt{8}\rho^3 \cos 3\theta$	$\sqrt{8}x(x^2 - 3y^2)$	
11	4	0	$\sqrt{5}(6\rho^4 - 6\rho^2 + 1)$	$\sqrt{5}[6(x^2 + y^2)^2 - 6(x^2 + y^2) + 1]$	Primary spherical aberration
12	4	2	$\sqrt{10}(4\rho^4 - 3\rho^2) \cos 2\theta$	$\sqrt{10}(x^2 - y^2)[4(x^2 + y^2) - 3]$	0° Secondary astigmatism
13	4	2	$\sqrt{10}(4\rho^4 - 3\rho^2) \sin 2\theta$	$2\sqrt{10}xy[4(x^2 + y^2) - 3]$	45° Secondary astigmatism
14	4	4	$\sqrt{10}\rho^4 \cos 4\theta$	$\sqrt{10}[(x^2 + y^2)^2 - 8x^2y^2]$	
15	4	4	$\sqrt{10}\rho^4 \sin 4\theta$	$4\sqrt{10}xy(x^2 - y^2)$	

Figure 2.5: First 15 Zernike polynomials according to Noll scheme [12].

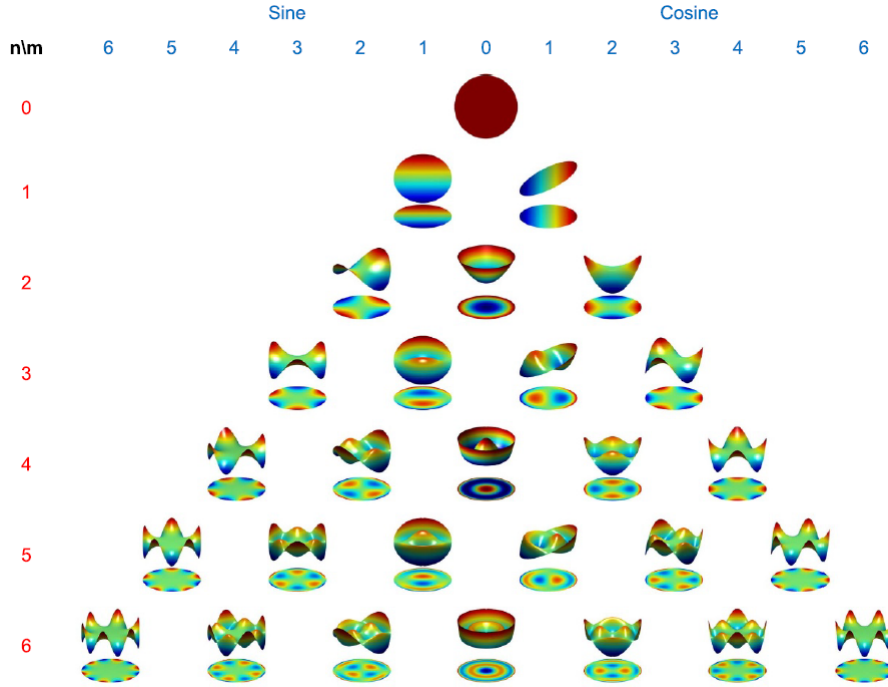


Figure 2.6: Visual representation of Zernike polynomials [12]

Noll indexing scheme allows writing Zernike polynomials' orthonormality as follows:

$$\frac{\int_A Z_j(\rho, \theta) Z_{j'}(\rho, \theta) dA}{\int_A dA} = \frac{\int_0^{2\pi} \int_0^1 Z_j(\rho, \theta) Z_{j'}(\rho, \theta) \rho d\rho d\theta}{\int_0^{2\pi} \int_0^1 \rho d\rho d\theta} = \delta_{jj'}, \quad (2.10)$$

where $\delta_{jj'}$ represents the Kronecker delta function and $\int_0^{2\pi} \int_0^1 \rho d\rho d\theta = \pi$.

One of the biggest interests of Zernike polynomials is to describe an aberrated phase (Φ) as a sum of such polynomials:

$$\Phi(\rho, \theta) = \sum_{j=1}^{\infty} a_j Z_j(\rho, \theta), \quad (2.11)$$

where the coefficients a_j are given by:

$$a_j = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 \Phi Z_j \rho \, d\rho \, d\theta \quad (2.12)$$

2.1.4 Image quality metrics

One advantage of using Zernike polynomials is that we can express the wavefront variance (σ_Φ^2) as the sum of squared Zernike coefficients, excluding the piston coefficient. Consequently, the rms phase error can be easily calculated using the following equation (see proof in the Appendices - 6.1):

$$\text{rms phase error} = \sqrt{\sigma_\Phi^2} = \sqrt{\sum_{j=2}^{\infty} a_j^2} \quad (2.13)$$

The rms phase error is a usual metric to evaluate an AO system's performance; the lower its value, the better the AO performance. Another very common AO performance metric, known as the Strehl ratio, was first introduced by K. Strehl. When optical aberrations are present in the wavefront, light will be diffracted away from the center of the PSF and consequently reduce its peak intensity [1]. The Strehl ratio (S), which is the ratio of the central irradiance of the PSF ($I(0)$) to that of a diffraction-limited PSF ($I_0(0)$) is a metric used to quantify this reduction in peak intensity [13]. The formula for the Strehl ratio is given by:

$$S = \frac{I(0)}{I_0(0)} \quad (2.14)$$

From Eq. 2.14, we can infer that S is always between 0 and 1, where a value of 1 means a perfect performance of the AO system, where all the optical aberrations were corrected. The closer S is to the value of one, the better the performance of an AO system. However, when significant aberrations occur, several off-axis peaks appear, and the central peak can itself be shifted. In such cases, S ends up being a poor quality metric [14]. However, in modern AO systems, where near-diffraction limit performance is achieved, this problem is not encountered, and the Strehl ratio can be safely used as a quality metric.

The Strehl ratio can be related to the phase variance (σ_Φ^2) through the following relation:

$$S \approx e^{-\sigma_\Phi^2}, \quad (2.15)$$

which is known as the Maréchal approximation [15]. This approximation is valid up to phase errors of 2 rad rms [1]. When we have a Strehl value ratio greater or equal to 0.8, that is usually considered as almost diffraction-limit operation [14], which in terms of rms phase error corresponds to approximately 0.45 rad rms.

2.2 Adaptive Optics Systems

2.2.1 Main Components and Operation

To achieve near-diffraction-limited resolution, AO systems must compensate for wavefront aberrations resulting from propagation through atmospheric turbulence. In Fig. 2.7, we have an illustration of an AO system. The main principle behind wavefront correction is phase conjugation [4]. Phase conjugation is achieved with the deformable mirror (DM), which is placed in the conjugated plane of the telescope's pupil and whose shape should be the conjugate of the wavefront error to null the phase error. The DM has several actuators that are driven according to the wavefront aberrations measured by a wavefront sensor (WFS). Currently, there are different types of DM, such as piezoelectric, micro-electro-mechanical systems (MEMS), bimorph, or segmented mirrors [13]. In addition to the DM, AO systems also use a separate tip/tilt mirror to correct for TT. This mirror is flat and as the name suggests, its goal is to correct for tip/tilt, i.e., compensate for the image motion. Usually, the slow tip/tilt is corrected by the tip/tilt mirror, while the fast tip/tilt is corrected via DM.

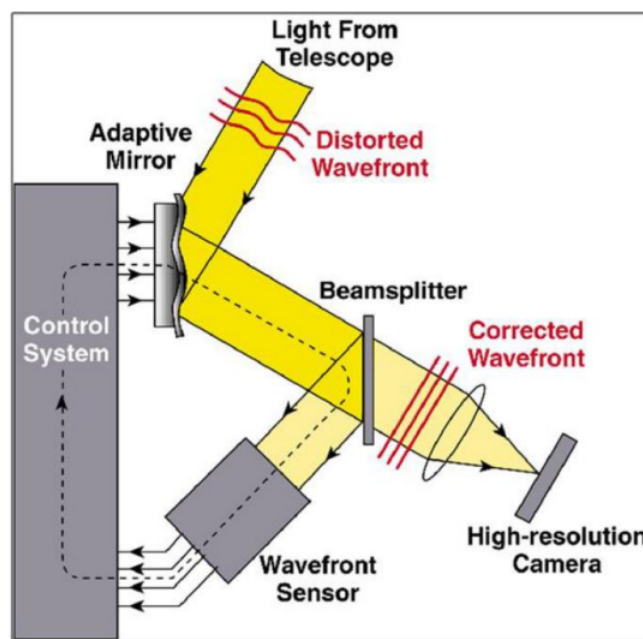


Figure 2.7: Adaptive optics (AO) system scheme [16].

The WFS and the DM communicate through a control system that sends the commands to the DM according to the WFS measurements. There are two main types of control systems: open-loop and closed-loop. In closed-loop, light is sent to the WFS after correction on the DM, meaning the WFS only sees the residual error according to the last correction [1]. The effect of the DM can be monitored, and further corrections can be applied to reduce the error, meaning feedback is applied [13]. In Fig. 2.7, the AO system represented is in closed-loop mode. On the other hand, in an open-loop configuration, light goes first through the WFS, where it is measured in its raw state, and then that information is sent to the DM. There is no feedback in this kind of configuration, and adequate compensation depends on accurate wavefront measurement and correction, meaning the WFS and DM need to be accurately calibrated over the whole dynamic range of the aberrated wavefront [1]. In general, closed-loop operation gives better performance [13].

These AO systems must keep up with atmospheric turbulence. Otherwise, there will be a considerable lag, and AO performance will be severely compromised. The rate at which turbulence changes is usually defined by the Greenwood frequency f_G , which is given by the following expression [17]:

$$f_G = \left[0.102 k^2 \sec(\zeta) \int_h C_n^2(h) v_w(h)^{\frac{5}{3}} dh \right]^{\frac{3}{5}}, \quad (2.16)$$

with $k = \frac{2\pi}{\lambda}$, ζ the zenith angle, C_n^2 the turbulence profile, and v_w the wind velocity profile. For stronger turbulence, f_G will be higher. The Greenwood frequency is used to determine how fast an AO system should work. As a general rule, an AO system should operate at a frequency 10 times higher than f_G [18], which usually means operating at a frequency of 1 kHz.

2.2.2 Wavefront Sensing

Wavefront sensors (WFS) play a crucial role in AO systems for astronomical telescopes, and they must fulfill several requirements. First, a WFS needs to sense the wavefront with high spatial resolution and operate with enough speed to keep up with atmospheric turbulence. Another requirement is that they must work with incoherent light (broadband and extended sources) since most astronomical sources emit this type of radiation [19]. As a consequence, absolute measures of optical phase are impossible to make [1]. The solution is then to measure intensities and then convert them to the desired phase.

A common type of WFS used in AO are the so-called pupil plane WFS, which measure specific properties of the wavefront at the pupil using zonal or modal decomposition. The most common approach is the zonal approach, where the wavefront is determined on a discrete set of

points. In each zone, the first-order spatial derivative (slope) is obtained, and from this gradient map, we can estimate the wavefront in a process known as wavefront reconstruction. Several types of pupil WFS exist, namely the Shack-Hartmann (SKH), the shearing interferometer, or the curvature sensor [1].

The most common WFS used in AO for ground-based telescopes is the Shack-Hartmann wavefront sensor (SH-WFS). It is a pupil plane WFS and its operational principle is depicted in Fig. 2.8. As we can see, the SH-WFS samples the incoming wavefront using a lenslet array placed at a pupil conjugate plane. Each lenslet will form an image at the focal plane, where a detector is placed to measure the obtained intensity pattern. This can be done either by using a separate detector for each sub-aperture (usually a four quadrant cell - quad cell) or by using a single charged couple device (CCD) detector. Assuming we have several detectors (one per sub-aperture), the obtained spots will be perfectly centered at each detector in the presence of a planar wavefront. When an aberrated wavefront is present, the spots won't be centered anymore; they will be shifted. Each lenslet will image a small fraction of the wavefront, which can be locally assumed as a tilted wavefront by a small angle α . Consequently, each spot will be displaced by a distance αf in relation to the reference positions, where f is the lenslet focal length [19].

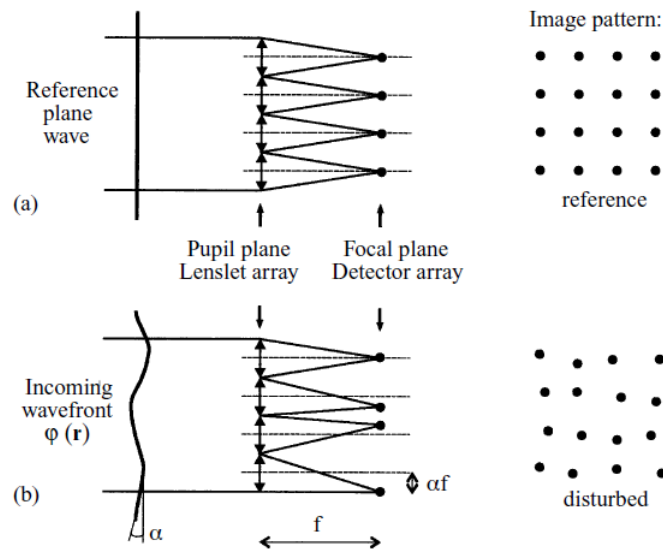


Figure 2.8: Illustration of Shack-Hartmann's wavefront sensor working principle: a) planar wavefront, b) aberrated wavefront [19].

These shifts (δ_x, δ_y) can be used to obtain the wavefront gradient, which can be used to obtain the wavefront ($\mathbf{W}$) by performing wavefront reconstruction. This wavefront reconstruction can be done by performing a linear integration of the wavefront slopes, which can be done by a simple matrix multiplication of this type:

$$\mathbf{W} = \mathbf{BS}, \quad (2.17)$$

where $\mathbf{B}$ is known as the reconstruction matrix and $\mathbf{S}$ is a vector of slopes [19].

Another possible way to perform wavefront sensing is to derive the phase from the whole-aperture intensity at the focal plane. These kinds of techniques are known as focal plane techniques [19]. However, the relationship between the pupil phase and the intensity pattern is highly non-linear, and non-linear iterative algorithms are required to estimate the phase. Due to the computational burden associated with these kinds of algorithms, other types of WFS are preferred for AO since the WFSs need to operate at very high rates. However, for low-order mode sensing, focal plane techniques can be a good solution with many points in their favor, as we'll see in this work. Focal plane techniques will be discussed in more detail in section 2.4 since they will be the center of all the work presented.

Although wavefront sensors are usually divided into two categories, focal and pupil plane techniques, I should emphasize that wavefront sensing is a much broader field of study, and there are other types of WFS that operate neither at the pupil nor at the focal plane. Wavefront sensing is a quite flexible field of study, and different techniques can be used for the purpose.

2.3 Sky Coverage

2.3.1 Anisoplanatism

When doing astronomical observations, a reference star is needed near the object of interest. This reference star is used to drive the AO systems for real-time compensation of atmospheric turbulence. However, if the science object is at a certain angular distance from the reference star, then the turbulence sampled by the wavefront won't be the same as the one from the imaging path, which means there will always exist inherent residual turbulence. This phenomenon is known as angular anisoplanatism and the wavefront error caused by it is known as anisoplanatic error [1]. Fig. 2.9 provides an illustration of angular anisoplanatism.

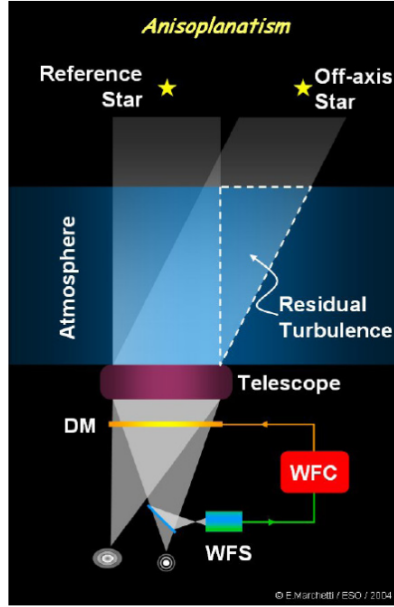


Figure 2.9: Angular anisoplanatism illustration [20]. The off-axis star is representative of the science object (it does not need to be a star).

The anisoplanatic error is conveniently expressed in terms of the isoplanatic angle θ_0 , which is the angular distance resulting in a mean-square wavefront error of 1.0 rad^2 [1]. When the telescope's diameter (D) is much bigger than the Fried coherence length (r_0), which is the case for most astronomical telescopes, the isoplanatic angle (θ_0) can be calculated through the following equation [21]:

$$\theta_0 = \left[2.91 k^2 \sec^{\frac{8}{3}}(\zeta) \int_h C_n^2(h) h^{\frac{5}{3}} dh \right]^{-\frac{3}{5}} \quad (2.18)$$

where $k = \frac{2\pi}{\lambda}$, ζ is the zenith angle, and h is the altitude. As we can see from Eq. 2.18, the isoplanatic angle highly depends on the turbulence profile $C_n^2(h)$, and it will be lower as the turbulence increases. Typical values for the isoplanatic angle θ_0 are $2''$ for the visible and $10''$ in the NIR. The isoplanatic angle can be calculated in a more simple way if we consider a single dominant turbulent layer located at height h from the ground, in which case θ_0 is given by [19]:

$$\theta_0 = 0.314 (\cos \zeta) \frac{r_0}{h} \quad (2.19)$$

The anisoplanatic error σ_θ^2 is related to the isoplanatic angle θ_0 through the following equation [19]:

$$\sigma_\theta^2 = \left(\frac{\theta}{\theta_0} \right)^{\frac{5}{3}}, \quad (2.20)$$

where θ is the angular distance between our object of interest and the reference star.

The angular anisoplanatism error is a big limitation to sky coverage as we can only observe near the reference star (within the isoplanatic angle). However, this problem wouldn't limit sky coverage if there were many sufficiently bright natural guide stars (NGS) over the whole sky, but unfortunately, that is not the case. Since the number of sufficiently bright NGS is limited, the solution to improve sky coverage is to create artificial stars generated by laser backscatter within the atmosphere. These artificial stars are commonly known as laser guide stars (LGS).

2.3.2 Laser Guide Stars

In order to increase sky coverage, laser guide stars (LGS) have been deployed in the last decades in almost all 8-10 m diameter ground-based telescopes such as the Very Large Telescope for instance (Fig. 2.11 b). LGS consist of the backscattered light from ground-based laser beams, and there are two main types, namely Rayleigh and Sodium (Na) LGS. Rayleigh LGS result from Rayleigh backscattering produced by air molecules. For this kind of LGS, the beacon's altitude is limited to about 20 km [1]. On the other hand, Sodium LGS are created at higher altitudes, more precisely in the sodium layer located at approximately 90 km of altitude in the mesosphere due to meteoritic ablation. These beacons are produced using lasers specifically tuned to 589.2 nm to excite the sodium atoms present in the Na layer. When these Na atoms return to the ground state, the desired backscattered radiation is emitted (sodium resonance fluorescence backscattering) [22].

In Fig. 2.10, we have a comparison between the sky coverage using NGS and using LGS from analytical computations for the K-band. From this graph, we can see that LGS allows for achieving considerably higher sky coverage. It should be noted that sky coverage never reaches 100 % since an LGS AO system always needs a complementary NGS for TT sensing and focus tracking. However, the star used for TT measurements and focus tracking can be considerably fainter, allowing higher sky coverage. This will be explored in more detail in section 2.3.3.

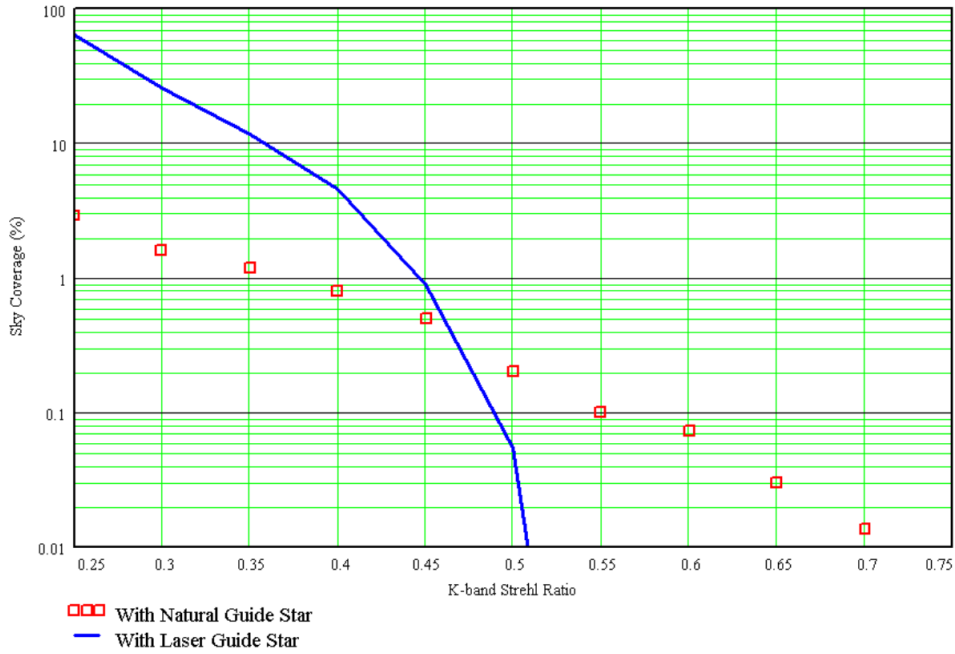


Figure 2.10: Illustration of sky coverage as a function of the Strehl ratio for K-band with NGS and LGS [23].

Another aspect that stands out from the results in Fig. 2.10 is that LGS AO systems cannot achieve such high Strehl ratios as NGS AO systems because LGS are located in the atmosphere at a finite altitude, generating the so-called cone effect also known as focal anisoplanism. Since LGS are located at finite altitudes within the atmosphere, only part of the turbulence will be sensed, which means light from the science target will go through turbulent regions that are not traversed by light from the beacon [24]. Since Rayleigh LGS are located closer to the telescope at lower altitudes, the cone effect will be much worse than what happens with Na LGS (Fig. 2.11 a). For this reason, Na LGS are preferable to Rayleigh LGS, and these are the types of LGS that are deployed in most modern ground-based telescopes. From now on, when referring to LGS, we subtend that they are Na LGS for notation simplicity.

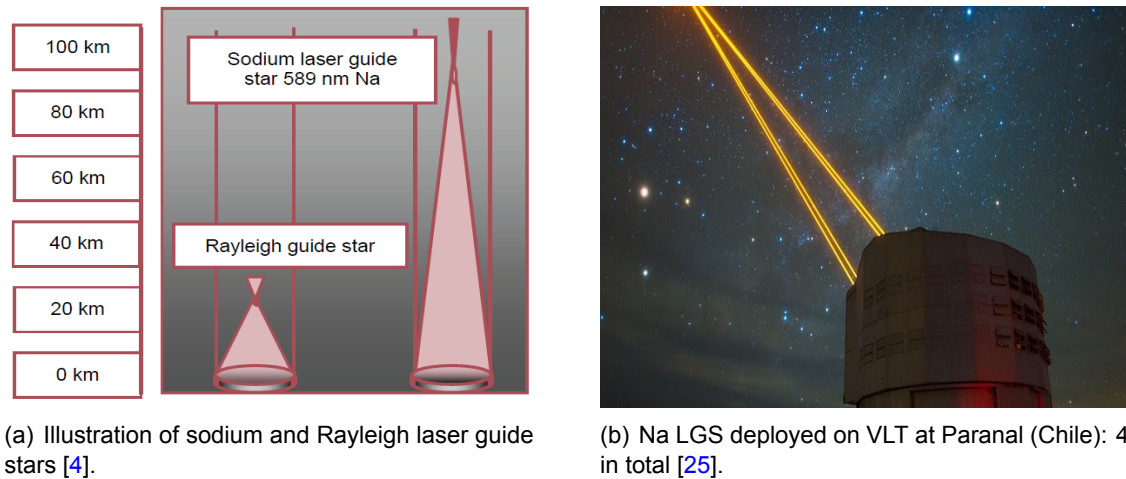


Figure 2.11: Comparison of the cone effect between Rayleigh and Na LGS (a) and example of LGS deployment at VLT (b).

To compensate for the cone effect characteristic of LGS, a suitable solution is to apply several LGS instead of just one. In this way, more turbulence will be sensed, mitigating the cone effect as illustrated in Fig. 2.12. However, even though the turbulence below the laser beacons is better sensed, the turbulence above them will never possibly be sensed using LGS. Still, since most of the turbulence is located in the lower atmosphere, applying several LGS can significantly improve the AO performance, yet never as good as with NGS.

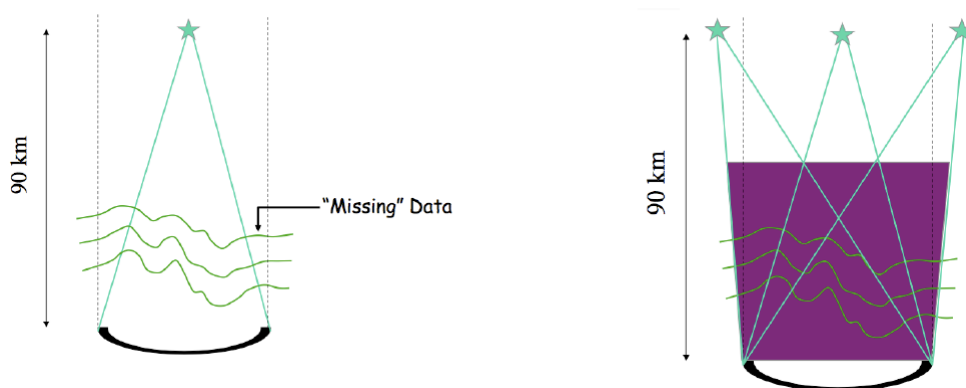


Figure 2.12: Cone effect compensation by applying several LGS [13]

Since LGS appeared in ground-based astronomy, different laser AO systems have been proposed and developed, and many are already in operation. One such system example is the Laser Tomographic Adaptive Optics (LTAO) illustrated in Fig. 2.13. In this LGS AO system, several off-axis wavefront sensors are deployed to get tomographic information on the turbulence above the telescope. In the case of LTAO, this tomographic information is projected into a single deformable mirror, optimizing the performance in a single direction across a small field of view (FoV) [26]. The corrected FoV is similar to what would be achieved with an NGS, but

LTAO systems have much higher sky coverage. Another common type of LGS AO system is the multi-conjugate adaptive optics (MCAO), in which several off-axis wavefront sensors are once more deployed, but each one controls a different deformable mirror. This way, MCAO systems allow achieving a much larger corrected FoV (~ 0.5 -1 arcmin in the NIR) while at the same time covering almost all sky [20]. On the other hand, the correction over the FoV is not as good as what happens with LTAO.

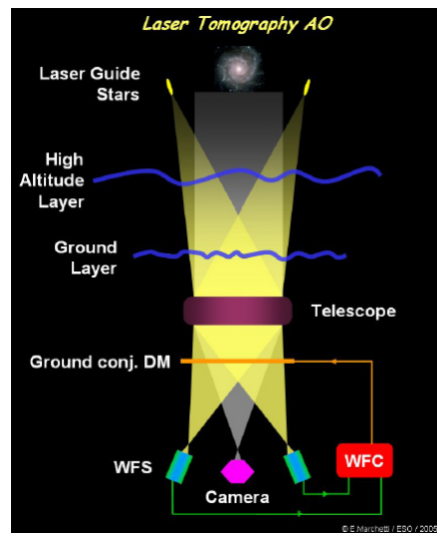


Figure 2.13: Laser Tomographic Adaptive Optics system [20].

LGS AO systems other than just LTAO and MCAO exist, and in the future, many others may be proposed and developed. The increase in sky coverage provided by LGS is a huge attraction to modern astronomy, and these systems are essential for all existing and future ground-based telescopes. However, LGS alone provide only a partial solution since there are some issues associated, one of them being the cone effect, as already discussed before, which, by applying several LGS, can be mitigated.

In addition, LGS WFS is insensitive to Tip/Tilt (TT), meaning an NGS still needs to be used for that purpose. Currently, the sky coverage limiting factor concerns the availability of sufficiently bright NGS for TT measurements. Of course, these NGS can be much fainter than what would be required for high-order mode estimation. Another problem associated with the LGS is that the Na layer altitude varies with time, which, if no correction is applied, will affect focus measurements. These two last problems are further discussed in the following section.

2.3.3 Tip-Tilt estimation and focus tracking

As stated in the last section, one of the problems of LGS is their insensitivity to image motion, i.e., to tip/tilt (TT) [27]. This happens because the transmitted and received beams are bent

by the same amount (Fig. 2.14 a) [13], and, consequently, the apparent position of the LGS remains static.

There are, however, some favorable factors for determining TT. The first one is that the isoplanatic angle for low-order modes such as TT is an order of magnitude larger than for high-order modes [26], which means we can use NGS relatively far from the LGS. Another advantage is related to the possibility of using the whole NGS flux for TT determination, meaning that much fainter stars can be used in comparison with the usual NGS in classical AO. TT modes are usually measured using center-of-gravity-like algorithms applied directly to an image at the focal plane. One last favorable factor related to TT correction is that it can be safely made at a lower rate (about 1/4 of the high-order loop) [28], meaning that higher SNR can be achieved which is particularly important when dealing with very faint stars.

Another critical limitation of LGS relates to the Na layer mean altitude variability, which induces a focus error as depicted in Fig. 2.14 b). This means that an altitude drift may be seen by the LGS WFS as an atmospheric focus when, in reality, that is not the case, affecting the AO performance. The telescope optical system should then be refocused as the LGS altitude variations occur to avoid performance loss. The good news is that these altitude drifts occur very slowly compared to atmospheric focus, making it easier to distinguish between these two effects. However, a natural phenomenon can provoke much faster altitude changes in the Na layer, namely meteors, which can cause altitude variations of approximately 1 km in 1 second, an extremely fast variation compared to normal altitude drifts. However, after 10-20 seconds, the Na mean altitude is once more reestablished within about 1 second [29].

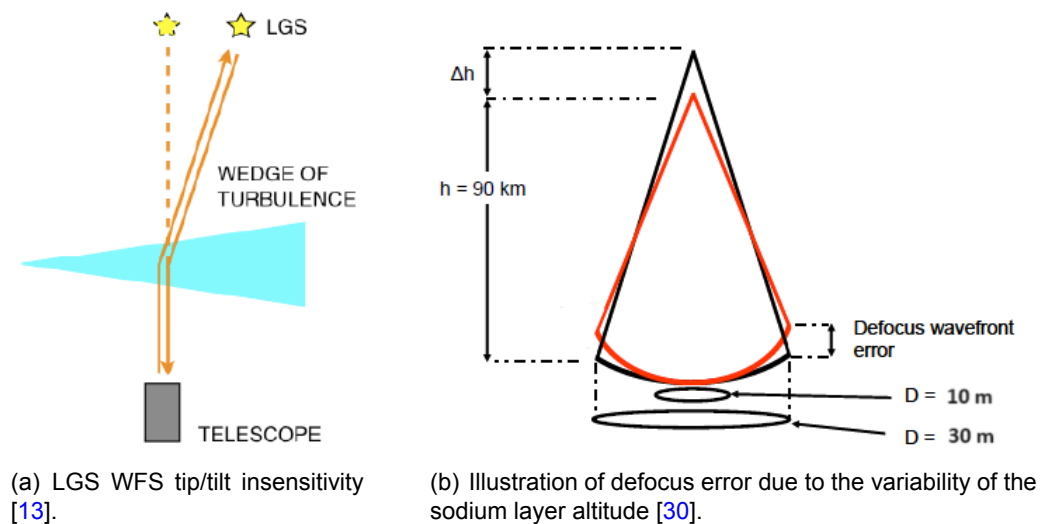


Figure 2.14: Illustration of laser guide star insensitivity to Tip/Tilt and focus measurement problems due to Na layer mean height variability.

Different studies were performed in an attempt to describe the mean altitude variations of

the Na layer, namely by Pfrommer & Hickson [31] and Neichel et al. [22]. It was found by both that the power spectral density (PSD) of the mean altitude variations is well-fitted by the following model:

$$PSD(f) = \alpha f^\beta, \quad (2.21)$$

with f representing the frequency, and α and β constants. An altitude variation Δh of the Na layer gives rise to an rms focus error σ_{Na} given by [31]:

$$\sigma_{Na} = \frac{1}{16\sqrt{3}} \left(\frac{D}{h - h_0} \right)^2 \Delta h, \quad (2.22)$$

where D is the telescope's diameter, h is the mean sodium layer altitude (approximately 90 km), and h_0 is the telescope's altitude. The proof for this formula is presented in section 6.2.

This error grows with D^2 , meaning that the refocusing operation needs to be done at higher frequencies for telescopes with large diameters. This is particularly important for future telescopes with 30-40 m diameter, such as the ELT. For example, let's imagine that we have two telescopes located at an altitude of 3000 m, with diameters of 10 and 30 m. Then, for $\Delta h = 100$ m, we have $\sigma_{Na} = 48$ nm rms and $\sigma_{Na} = 430$ nm rms for $D = 10$ and $D = 30$ respectively.

2.4 Focal Plane Wavefront Sensing

2.4.1 Motivation and associated Challenges

Focal Plane wavefront sensing (FPWFS) is an attractive choice for low-order mode sensing in an LGS AO system context. Perhaps the biggest attraction is that it is a full-pupil technique, meaning that the whole flux from the guide star can be utilized. This is an interesting scenario when dealing with very faint natural guide stars, as needed in LGS AO systems for estimating TT and focus tracking. Using FPWFS can thus help further increase the sky coverage.

Another advantage associated with FPWFS is that it does not require additional hardware, which is why FPWFS is sometimes referred to as a sensorless technique. Only the image at the focal plane provided by the science camera is needed, which makes the implementation of FPWFS in an AO system quite straightforward in terms of hardware. However, in some FPWFS techniques, an additional image other than the one from the science camera may be needed, as we'll see next in this section. Still, the extra hardware requirements are minimal.

In addition, FPWFS also provides the possibility of compensating for non-common path aberrations (NCPA), which cannot be sensed using conventional WFS techniques such as the Shack-Hartmann. NCPA are the WFS differences between the WFS and the science camera,

which are a consequence of the different light paths between the WFS and the science camera. These kinds of aberrations vary very slowly (in the order of minutes: quasi-static), meaning that the corrections can be done at low rates depending on the stability of the AO system [32].

Now, I will introduce the challenges associated with FPWFS. As already seen before, FPWFS is a specific type of WFS technique where the phase is estimated from the image at the focal plane [33]. The relationship between the phase and the intensity at the focal plane is highly non-linear. As seen in a previous section, for incoherent and monochromatic radiation, the image and the object intensities (I_i and I_o respectively) are related through the following relation:

$$I_i = p * I_o + n, \quad (2.23)$$

where p is the image of a point source (impulse response), known as the point spread function, and n is the noise. The point spread function is given by:

$$p = |\mathcal{F}\{U e^{i\Phi}\}|^2, \quad (2.24)$$

with $U e^{i\Phi}$ being the electric field at the pupil plane, where Φ is the phase we want to estimate.

Due to this non-linear relationship between the phase and the intensity pattern, FPWFS is a non-trivial inverse problem, which involves the usage of iterative non-linear optimization algorithms to estimate the phase Φ . One of the disadvantages of such algorithms is the computational burden associated, which can be a severe problem for real-time operation in an AO system. However, the recent algorithmic and computational hardware advances have greatly reduced this computational issue. These costs are expected to be even further reduced in the upcoming years, making it possible for FPWFS algorithms to operate at much higher rates [34].

Non-linear optimization algorithms search for a solution (set of parameters) that minimizes a particular metric error defined by the user. We start with an initial guess for the solution and then update it at each iteration, usually using gradient-based search algorithms as illustrated in Fig. 2.15. Higher-order search algorithms (Hessians, third-order, etc.) can also be used, but with significant computational costs associated.

In Fig. 2.15, we also illustrate one of the other challenges related to non-linear optimization algorithms, which is the existence of several local minima. These are undesirable since the real optimal solution is the global minima. The minimum at which the algorithm converges highly depends on the initial guess, as shown in Fig. 2.15. When there are many local minimums, our initial guess should be as close as possible to the final solution so we can converge to the

desired global minimum. However, an initial guess close to the solution is usually difficult to obtain.

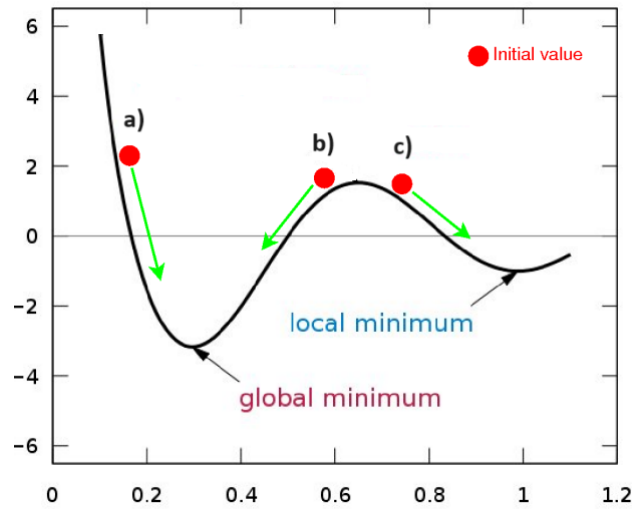


Figure 2.15: Illustration of local minima problem associated with non-linear optimization algorithms (Adapted from [35]). The initial guess value plays a crucial role in the minimum at which the algorithm converges.

Another important aspect affecting FPWFS performance is the sampling of the images. The detector's angular pixel size determines the sampling of the images. Ideally, the sampling should be of 1.0 Nyquist to avoid any loss of spatial information. For 1.0 Nyquist sampling, the detector's angular pixel size should be half the FWHM of the diffraction-limited PSF, i.e., $0.5 \frac{\lambda}{D}$. It is known that the FPWFS performance is better with well-sampled PSFs rather than undersampled ones.

The images' signal-to-noise ratio (SNR) will also affect FPWFS performance. If we have an image highly affected by noise (low SNR), then it is clear that the FPWFS performance will be severely compromised. There are different types of noise, namely photon noise, readout noise, and dark noise. The photon noise is independent of the detector and is "inherent in the light" due to its quantum nature. When light reaches the detector at a specific star flux, photons are converted into electrons. However, even at constant flux, the arrival of the photons follows Poisson statistics, meaning that the number of photons will have a certain standard deviation/error [36]. On the other hand, readout noise depends on the detector being used, and it is associated with non-idealities in the measurement process. As each pixel is being read out, an analog voltage needs to be quantified by A/D conversion. This conversion process is not perfect, and errors arise, which grow as the readout speed increases. Another noise term associated with detectors is dark noise, which is associated with thermally generated electrons. A simple way to reduce this noise is to cool the detector to very low temperatures [37].

As seen in Eq. 2.23, the image formation model accounts for a noise term. The noise

can be modeled by using statistical models, namely Gaussian or Poisson statistics. At high flux, when the readout or dark noises are minimal, assuming Gaussian statistics for the noise allows us to model the noise well. If no assumption is made about the noise or if the model is not the most correct, then FPWFS algorithms can be severely affected since they are quite model-dependent. This can be particularly important when we have low SNR.

Despite all the challenges associated with FPWFS, its advantages for LGS AO systems are quite exciting, which motivates all the research around these types of algorithms for WFS in AO. Tab. 2.1 has a resume of all the main advantages and challenges associated with FPWFS discussed in this section.

Advantages	Disadvantages
No hardware requirements	Computationally demanding
Possibility to measure NCPA	Local minima
Full-pupil technique (more sky coverage)	Noise, sampling, and model dependence

Table 2.1: Advantages and disadvantages associated with FPWS techniques.

2.4.2 Sign ambiguity for even modes and phase diversity

The relationship between the PSF and the pupil plane phase given by Eq. 2.24 results in a sign ambiguity for even Zernike modes. The even modes are the ones with even radial order, such as defocus or astigmatism. Assuming a symmetric pupil, and only the presence of even modes (even phase), we have that [38]:

$$\left| \mathcal{F} \left\{ U e^{i\Phi_{\text{even}}(+x)} \right\} \right|^2 = \left| \left(\mathcal{F} \left\{ U e^{i\Phi_{\text{even}}(+x)} \right\} \right)^* \right|^2 = \left| \mathcal{F} \left\{ \left(U e^{i\Phi_{\text{even}}(-x)} \right)^* \right\} \right|^2 \quad (2.25)$$

Knowing that $\Phi_{\text{even}}(-x) = \Phi_{\text{even}}(+x)$, we conclude that:

$$\left| \mathcal{F} \left\{ U e^{i\Phi_{\text{even}}(+x)} \right\} \right|^2 = \left| \mathcal{F} \left\{ U e^{-i\Phi_{\text{even}}(+x)} \right\} \right|^2, \quad (2.26)$$

which allows us to conclude that the intensity pattern at the focal plane is the same for $\pm\Phi_{\text{even}}(x)$, resulting in a sign ambiguity for even modes. In Fig. 2.16, I have an illustration of this sign ambiguity for the specific case of defocus, where it is possible to see that the PSFs are the same whether the sign is positive or negative.

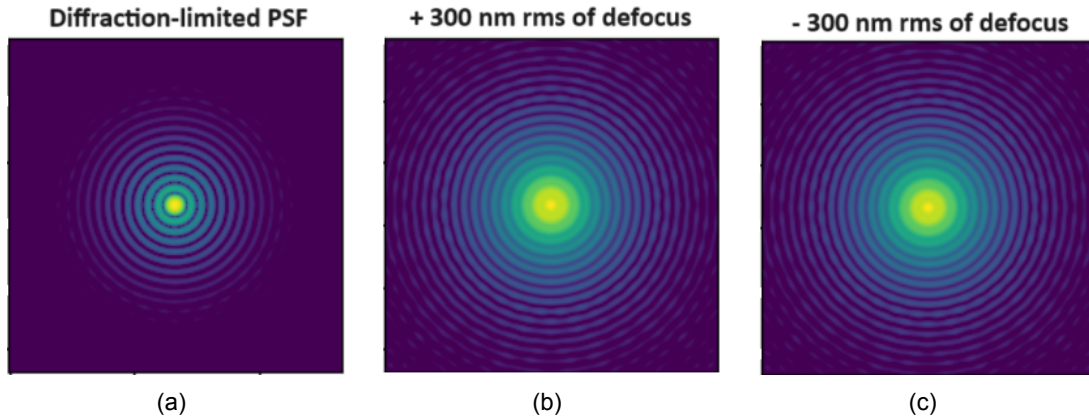


Figure 2.16: Illustration of the sign ambiguity for even modes, namely for defocus. The PSF is the same regardless of whether the defocus introduced is positive or negative. The PSFs are presented in logarithmic scale.

The solution to this sign ambiguity is to use diversity, which is typically applied in the amplitude or phase. Amplitude diversity would imply using a non-uniform pupil, allowing to break the degeneracy for even modes. However, this kind of diversity won't be used in this work, so it won't be further discussed.

Diversity in the phase implies applying a known amount of a certain aberration in the optical system, which will solve the sign ambiguity. The estimation of defocus will be particularly explored in this work, so the following discussion about phase diversity will be oriented explicitly toward defocus.

The most common type of aberration applied as phase diversity is defocus since it is straightforward to apply and control (we just need to move the detector's position back or forth by a certain amount) [39]. Applying phase diversity changes our working point. Let's imagine we have 300 nm rms of defocus as diversity. If we add +300 nm rms of defocus, we'll have a total of 600 nm rms of defocus, while if we add -300 nm rms of defocus, we'll have a total of 0 nm rms of defocus. The obtained PSFs will now differ, meaning the sign ambiguity was lifted. However, if we add -400 nm rms, we'll have a total of -100 nm rms. There will be a sign ambiguity again since the PSF for +100 nm rms is the same as the one for -100 nm rms. In this way, the range without sign ambiguity will always be one-sided bounded when using defocus as diversity to estimate defocus.

Other than just defocus, other aberrations can be used as phase diversity, namely astigmatism, to lift the sign ambiguity for defocus. In Fig. 2.17, I have two PSFs with 200 nm rms of astigmatism (45°) with defocus values of opposite signs. We can see that the PSF will stretch in different directions according to the sign of defocus introduced. Consequently, using astigmatism as diversity allows for achieving higher linearity ranges when compared with defocus.

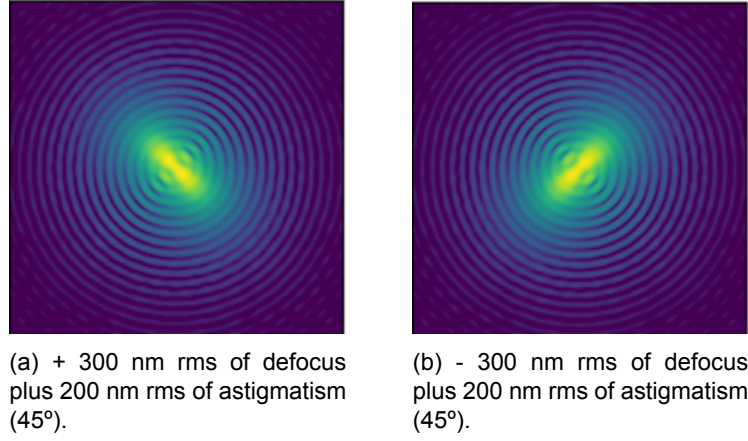


Figure 2.17: PSF stretches in perpendicular directions according to the sign of the defocus introduced in the presence of astigmatism. The amount of astigmatism (45°) used for these images was 200 nm rms. The PSFs are presented in logarithmic scale.

2.4.3 Small-phase solution

As stated before, FPWFS algorithms use iterative non-linear optimization algorithms to estimate the phase. However, Gonsalves proposed a linear and non-iterative solution for FPWFS, which he called the small-phase solution since it assumes that the unknown phase is small [39]. Next, I will explain and show the small-phase solution proposed by Gonsalves. In the following demonstrations, the convention used is that uppercase letters represent quantities at the pupil plane, whereas lowercase letters represent quantities at the focal plane.

The point spread function (p) for the small phase can be approximated by:

$$p = |\mathcal{F}\{U e^{i\Phi}\}|^2 \approx |\mathcal{F}\{U(1 + i\Phi)\}|^2 = |\mathcal{F}\{U\} + i\mathcal{F}\{U\Phi\}|^2, \quad (2.27)$$

where the phase (Φ) can be obtained as a sum of its even and odd components: $\Phi = \Phi_e + \Phi_o$, allowing to rewrite p as:

$$p = |\mathcal{F}\{U\} + i(\mathcal{F}\{U\Phi_e\} + \mathcal{F}\{U\Phi_o\})|^2 = |u + i(v + w)|^2 \quad (2.28)$$

Assuming that U is real and even, we know from Fourier symmetry properties that u and v will be real and even and w imaginary and odd. We now define $y = iw$ such that y is real and odd. We have then that p is given by:

$$p = |u + y + iv|^2 \quad (2.29)$$

Since u, y , and v are all real, we have $p = (u + y)^2 + v^2$. The even and odd parts of p are given by Eq. 2.30 and 2.31, respectively:

$$p_e = u^2 + y^2 + v^2 \quad (2.30)$$

$$p_o = 2uy \quad (2.31)$$

Eq. 2.31 allows us to determine y :

$$y = \frac{p_o}{2u} \quad (2.32)$$

However, to avoid divisions by excessively small values (or even zero) that would cause divergence problems, we should use the following formula instead:

$$y = 0.5 \frac{p_o u}{u^2 + \epsilon}, \quad (2.33)$$

where ϵ is a suitable small constant. Knowing that $y = i\mathcal{F}\{U\Phi_o\}$, we can then calculate Φ_o through the following relation:

$$U\Phi_o = \text{Im}\{Y\}, \quad (2.34)$$

with $Y = \mathcal{F}^{-1}\{y\}$. Having y , we can now solve Eq. 2.30 for v :

$$v = \sqrt{p_e - u^2 - y^2} \quad (2.35)$$

However, the sign is ambiguous in this equation, which makes the even component of the phase undetermined [39]. We need to use phase diversity to solve this sign ambiguity for the even part of the phase. In essence, a second image (diverse image) is required to solve this undetermination. Typically, the diverse image is obtained by applying a known amount of defocus. The phase of the diverse image will have the diversity Φ_d in addition to the aberrated phase Φ , and the image at the focal plane will be given by:

$$p_d = \left| \mathcal{F}\{U e^{i(\Phi + \Phi_d)}\} \right|^2, \quad (2.36)$$

Following a similar procedure to the one followed to obtain Eq. 2.29, we have that:

$$p_d = u^2 + 2u(y + y_d) + (y + y_d)^2 + (v + v_d)^2, \quad (2.37)$$

where $y_d = iw_d = i\mathcal{F}\{U\Phi_{o,d}\}$, and $v_d = \mathcal{F}\{U\Phi_{e,d}\}$ which are quantities easily calculated since we know the phase diversity Φ_d . From Fourier symmetry properties, we know that y_d is real

and odd, while v_d is real and even. The even part of p_d is then given by:

$$p_{e,d} = p_e + 2yy_d + y_d^2 + 2vv_d + v_d^2 \quad (2.38)$$

Solving for v , we have:

$$v = \frac{p_{e,d} - p_e - (y_d^2 + v_d^2 + 2yy_d)}{2v_d} \quad (2.39)$$

Eq. 2.39 is highly sensitive to noise due to the subtraction of two similar images ($p_{e,d}$ and p_e). Consequently, this equation is used solely to compute the signs of v . For the absolute values of v , we use Eq. 2.35 instead [40].

This algorithm provides a closed-form and non-iterative solution for the phase estimation from a focal plane image. It can be viewed as an attempt to linearize a non-linear problem. However, this speed increase comes with some associated costs. First, this algorithm assumes a small phase and can only work with symmetric pupils. Secondly, this technique also demands the usage of two images, meaning that the light needs to be split into two separate channels or alternatively record images with and without diversity [41].

In 2002, Gonsalves proposed the sequential phase diversity technique, which eliminated the need to use two separate images [42]. This technique relies on a closed-loop system, where the diverse image at iteration x is the previous image at iteration $x - 1$, which has a phase diversity in relation to the image at x , given by the change in DM commands [41]—an extension of this sequential phase diversity technique known as Fast&Furious was recently proposed with some on-sky implementations [41], [43].

2.4.4 Classification of Focal Plane Wavefront Sensing Techniques

Focal Plane Wavefront sensing (FPWFS) algorithms can be divided into two different categories: phase retrieval and joint estimation [44]. Joint estimation is also referred to as phase diversity in the literature. Still, this term can be confused with the phase diversity corresponding to a known optical aberration, as introduced before. To avoid confusion, phase diversity will always refer to an optical aberration throughout this work.

In Fig. 2.18, we have a classification scheme that distinguishes between different inverse problems regarding available object and system knowledge [34]. In a WFS scenario, the system, i.e., the optical aberrations, is always unknown. If that weren't the case, WFS wouldn't be necessary at all. We are then interested in the bottom part of the scheme, where we have phase retrieval and joint estimation. In the case of phase retrieval, the object needs to be

known, while for joint estimation, that is not the case.

		Object	
		known	unknown
System	known	omniscience	image restoration
	unknown	system identification (phase retrieval)	joint estimation (phase diversity)

Figure 2.18: Classification of inverse imaging problems [34].

Phase retrieval algorithms usually extract the phase from an image of a point source, i.e., the PSF. An illustration of the typical optical setup for a phase retrieval algorithm is presented in Fig. 2.19. In this figure, we can also see the application of phase diversity needed to lift the sign ambiguity for even modes, as already seen before. It should also be noted that in phase retrieval, the phase information is extracted from a single image. For these reasons, these phase retrieval algorithms can also be called single-image phase diversity algorithms. The FPWFS algorithms that will be studied in this work fall into this category and will be presented in more detail in the following section.

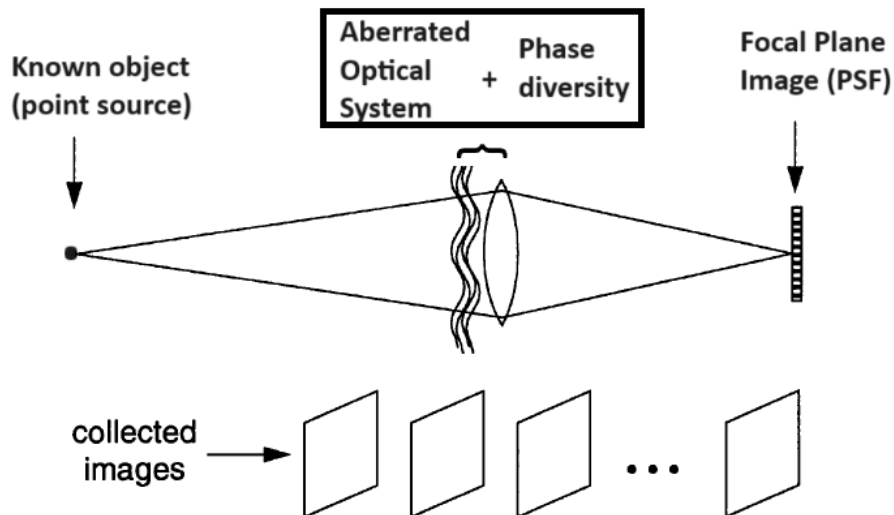


Figure 2.19: Optical setup of a phase retrieval system (Adapted from [19]).

Contrary to phase retrieval, joint estimation can also work when the object is unknown, which is particularly common when imaging extended objects [44]. Joint estimation was first

proposed by Gonsalves [45] and the standard optical setup is illustrated in Fig. 2.20. One important aspect concerning joint estimation is that it always requires two images: the conventional image at the focal plane and a diverse image. The diverse image is degraded by the optical aberrations plus a known aberration, i.e., phase diversity, which usually is defocus due to its implementation simplicity. Joint estimation can also work when the object is known. However, when that happens, phase retrieval is usually preferred. Only when there is uncertainty about the object being imaged should we use joint estimation [34].

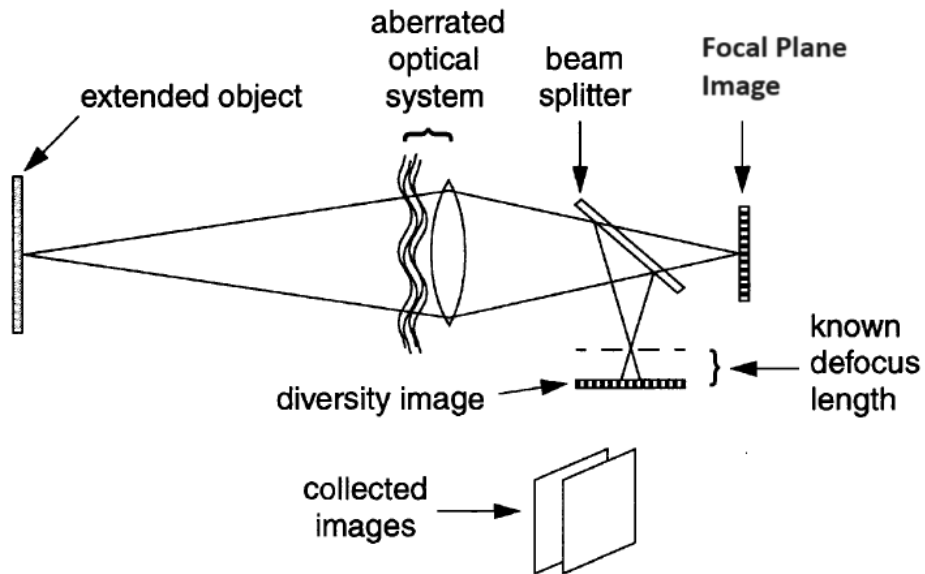


Figure 2.20: Typical optical layout of a joint estimation system [44]. The diversity image is degraded by the optical aberrations plus a known amount of defocus (phase diversity).

3 Methods

In this section, I will describe the phase retrieval algorithms that will be analyzed and studied in the next section, namely the Gerchberg-Saxton (GS), the Linearized Focal Plane Technique (LiFT), and the Full Width at Half Maximum (FWHM) fit.

3.1 Gerchberg-Saxton

Gerchberg-Saxton (GS) algorithm was first proposed by Gerchberg and Saxton in 1972 [46], and it is the most standard and foundational phase retrieval algorithm. GS algorithm uses the Fourier relationship between the electric field at the pupil and at the focal planes to perform phase retrieval. In addition to that relation, partial constraints known a priori are required at each plane. At the focal plane, that constraint comes from the intensity measurement (p), which determines that the amplitude of the electric field at the focal plane is given by $\sqrt{p}$. At the pupil plane, despite not knowing the exact amplitude value, we know that the electric field only exists within the pupil, so we apply that constraint by using a mask [47].

The GS algorithm is presented schematically in Fig. 3.1. In GS, we iterate back and forth between the pupil and focal plane using a Fourier transform, satisfying known constraints before returning each plane [47].

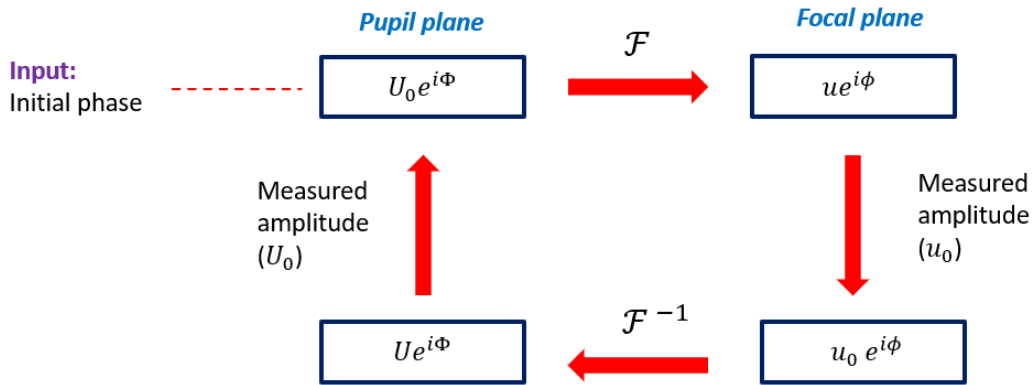


Figure 3.1: Illustration of Gerchberg-Saxton algorithm.

The initial phase guess given to GS can be random, although using a plane phase proved more effective according to the tests I performed. Of course, if we have more a priori knowledge of our phase, we should give that information to the algorithm, namely the phase diversity. If phase diversity is not provided as our initial guess, the sign of even modes will be undetermined/ambiguous.

The result of this GS algorithm is a phase that will be within $-\pi$ and $+\pi$ (wrapped phase). To obtain the Zernike coefficients corresponding to that phase, we need to unwrap the phase and

project our phase on the Zernike basis. I use a phase unwrapping function provided by scikit-image [48], which is based on a phase unwrapping algorithm proposed by Miguel Arevallilo Herraes et al. [49].

The algorithm's convergence is monitored by calculating the squared error at the Fourier plane, i.e., the focal plane. At iteration n , the error is given by:

$$E_n^2 = ||u_0^2 - u^2||_F, \quad (3.1)$$

which is the Frobenius norm of the difference between the current estimate of the focal plane intensity and the measured one. This error metric can be used to exit the GS cycle before reaching the maximum number of iterations. In my studies, I used a threshold of 10^{-9} for the error metric. When $E_n^2 < 10^{-9}$, then we exit the cycle. This is, of course, a very exigent threshold that, in real conditions with noise and high-order residuals, is never reached, but with simulated data, it can be achieved.

Fienup showed in 1982 that the squared error E_n^2 can only decrease or stay the same at each iteration [47]. For that reason, this algorithm is also known as the error-reduction algorithm. Despite that, the algorithm's convergence is sometimes slow. Usually, the algorithm converges very fast within the first iterations but then starts converging much more slowly. A particular problem is the occurrence of plateaus, where the error reduces extremely slowly, which is quite a common occurrence. Sometimes, if we continue iterating, we can surpass these plateaus and start converging fast again, but that sometimes requires hundreds or even thousands of iterations, which means a huge increase in computational costs.

GS algorithm does not explicitly apply any non-linear optimization algorithm as was expected from a phase retrieval technique. However, it was shown by Fienup that the GS algorithm is closely related to the steepest descent method [47], which is a specific type of non-linear optimization algorithm. This ends up proving why the GS algorithm works despite not directly applying any non-linear optimization algorithm.

3.2 Linearized Focal Plane Technique

The Linearized Focal Plane Technique (LiFT) is a much more recent phase retrieval technique proposed by Serge Meimon et al. in 2010 [50]. Despite having the word linearized in it, I should enforce that this is not a linear technique. As we'll see, it is a relatively standard non-linear optimization technique similar to what happens with the steepest descent algorithm. An example of a real linear phase retrieval technique is the small angle solution proposed by Gonsalves, already introduced in section 2.4.3.

As we know, the image at the focal (I) plane is given by the following formula:

$$I(\mathbf{r}, \Phi) = \left| \mathcal{F}\{U e^{i(\Phi + \Phi_d)}\} \right|^2 + n(\mathbf{r}) = \left| \mathcal{F}\{U_d e^{i\Phi}\} \right|^2 + n(\mathbf{r}), \quad (3.2)$$

where $U_d = U e^{i\Phi_d}$ is the electric field at the pupil plane, and $n(\mathbf{r})$ is assumed to be a zero-mean Gaussian noise of the detector [50]. The phase Φ can be decomposed in the Zernike basis: $\Phi = \sum_j a_j Z_j$. Let $\mathbf{A}$ represent the vector of Zernike coefficients, then we have:

$$I(\mathbf{A}) = \left| \mathcal{F}\{U_d e^{i(\sum_j a_j Z_j)}\} \right|^2 + n(\mathbf{r}) = \mathcal{F}\{U_d e^{i(\sum_j a_j Z_j)}\} \times \mathcal{F}\{U_d e^{i(\sum_j a_j Z_j)}\}^* + n(\mathbf{r}) \quad (3.3)$$

Doing a first-order Taylor approximation for $I(\mathbf{r})$ with regard to $\mathbf{A}$:

$$I(\mathbf{A}) = I(\mathbf{0}) + \mathbf{A} \left. \frac{\partial I}{\partial \mathbf{A}} \right|_{\mathbf{A}=\mathbf{0}} + n(\mathbf{r}), \quad (3.4)$$

where $I(\mathbf{0})$ is the image accounting just with the phase diversity and

$$\begin{aligned} \frac{\partial I}{\partial a_k} &= \mathcal{F}\{U_d e^{i(\sum_j a_j Z_j)} i Z_k\} \times \mathcal{F}\{U_d e^{i(\sum_j a_j Z_j)}\}^* \\ &\quad + \mathcal{F}\{U_d e^{i(\sum_j a_j Z_j)}\} \times \mathcal{F}\{U_d e^{i(\sum_j a_j Z_j)} i Z_k\}^* \end{aligned} \quad (3.5)$$

This way, we have that

$$\left. \frac{\partial I}{\partial \mathbf{A}} \right|_{\mathbf{A}=\mathbf{0}} = \mathcal{F}\{i U_d Z_k\} \times \mathcal{F}\{U_d\}^* + \mathcal{F}\{U_d\} \times \mathcal{F}\{i U_d Z_k\}^* \quad (3.6)$$

$$I'_k(\mathbf{0}) = \left. \frac{\partial I}{\partial \mathbf{A}} \right|_{\mathbf{A}=\mathbf{0}} = 2\Re\{\mathcal{F}\{U_d\}^* \times i \mathcal{F}\{U_d Z_k\}\} \quad (3.7)$$

From Eq. 3.4, we have then that:

$$I(\mathbf{A}) - I(\mathbf{0}) = \sum_k a_k I'_k(\mathbf{0}) + n \quad (3.8)$$

We assume that the images are formed on a matrix of pixels. For the p^{th} pixel, we define $\Delta I[p] = I(\mathbf{A})[p] - I(\mathbf{0})[p]$ and $\mathbf{H}[p, k] = I'_k(\mathbf{0})[p]$.

$$\Delta \mathbf{I} = \mathbf{H}(\mathbf{0}) \mathbf{A} + \mathbf{n} \quad (3.9)$$

The goal of LiFT, as we know, is to estimate the vector of coefficients $\mathbf{A}$, and for that purpose,

we need to take into account the noise $\mathbf{n}$, which is assumed to be Gaussian with mean zero, and its probability law can be written as:

$$p(\mathbf{X}) \propto \exp\left(-\frac{1}{2}\mathbf{X}^T \mathbf{R}_n^{-1} \mathbf{X}\right) \quad (3.10)$$

where $\mathbf{R}_n = \langle \mathbf{n}\mathbf{n}^T \rangle$ is the noise covariance matrix. According to Eq. 3.9, we have that the probability of having $\Delta \mathbf{I}$ knowing $\mathbf{A}$ is given by:

$$p(\Delta \mathbf{I} | \mathbf{A}) \propto \exp\left(-\frac{1}{2}(\Delta \mathbf{I} - \mathbf{H}\mathbf{A})^T \mathbf{R}_n^{-1} (\Delta \mathbf{I} - \mathbf{H}\mathbf{A})\right) \quad (3.11)$$

This probability is maximum when $(\Delta \mathbf{I} - \mathbf{H}\mathbf{A})^T \mathbf{R}_n^{-1} (\Delta \mathbf{I} - \mathbf{H}\mathbf{A})$ is minimum. This means that the derivative in relation to $\mathbf{A}$ equals zero. Doing some algebraic operations, we have that [51]:

$$2\mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H}\mathbf{A} - 2\mathbf{H}^T \mathbf{R}_n^{-1} \Delta \mathbf{I} = 0 \quad (3.12)$$

The maximum likelihood estimate is then given by:

$$\hat{\mathbf{A}}_{ML} = \mathbf{P}_{ML} \Delta \mathbf{I}; \quad \mathbf{P}_{ML} = (\mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}_n^{-1}, \quad (3.13)$$

The maximum likelihood estimator $\hat{\mathbf{A}}_{ML}$ is a ponderated estimate that is applied to the image difference. This ponderation is adaptable according to the flux conditions. At intense flux, the noise is low, and all pixels are considered. On the other hand, at low flux, the noise is high, and only the pixels around the center are considered.

According to Eq. 3.4, the linearity range of LiFT would be limited to around 0 since we only make a first-order Taylor expansion. To obtain higher linearity ranges, we iterate several times. The following algorithm is used [52]:

- Iteration 0: $\Delta \mathbf{I} = \mathbf{I}_{\text{data}} - \mathbf{I}(\mathbf{0}); \hat{\mathbf{A}}^0 = \mathbf{P}_{ML}(\mathbf{0}) \Delta \mathbf{I}$
- Iteration 1: $\Delta \mathbf{I} = \mathbf{I}_{\text{data}} - \mathbf{I}(\hat{\mathbf{A}}^0 + \mathbf{0}); \hat{\mathbf{A}}^1 = \mathbf{P}_{ML}(\hat{\mathbf{A}}^0 + \mathbf{0}) \Delta \mathbf{I} + \hat{\mathbf{A}}^0$
- Iteration 2: $\Delta \mathbf{I} = \mathbf{I}_{\text{data}} - \mathbf{I}(\hat{\mathbf{A}}^1 + \mathbf{0}); \hat{\mathbf{A}}^2 = \mathbf{P}_{ML}(\hat{\mathbf{A}}^1 + \mathbf{0}) \Delta \mathbf{I} + \hat{\mathbf{A}}^1$
- ...
- Iteration k: ...

where $\mathbf{I}_{\text{data}}$ is the image recorded on the detector and $\mathbf{I}(\mathbf{X})$ a simulated image from LiFT model. The initial working point of our algorithm $(\mathbf{0})$ contains only the phase diversity, but at each iteration, that point is updated [51]. With this iterative algorithm, we can achieve higher linearity ranges.

LiFT algorithm will stagnate in a certain solution once $\Delta \mathbf{I}$ equals zero, which means that the modeled image is equal to the one obtained from the detector. From this point on, the solution

won't be updated anymore, so we can stop iterating. More precisely, we stop iterating when $\|\hat{\mathbf{A}}^k - \hat{\mathbf{A}}^{k-1}\|$ is less than 10^{-3} nm rms.

The maximum likelihood estimator $\hat{\mathbf{A}}_{ML}$, without the linearization is given by [52]:

$$\hat{\mathbf{A}}_{ML} = \arg \min_{\mathbf{A}} \left([\mathbf{I}_{\text{data}} - \mathbf{I}(\mathbf{A} + \mathbf{0})]^T \mathbf{R}_n^{-1} [\mathbf{I}_{\text{data}} - \mathbf{I}(\mathbf{A} + \mathbf{0})] \right) \quad (3.14)$$

At each iteration k , we can evaluate the following criteria:

$$C_k = \left[\mathbf{I}_{\text{data}} - \mathbf{I}(\hat{\mathbf{A}}^{k-1} + \mathbf{0}) \right]^T \mathbf{R}_n^{-1} \left[\mathbf{I}_{\text{data}} - \mathbf{I}(\hat{\mathbf{A}}^{k-1} + \mathbf{0}) \right] \quad (3.15)$$

The algorithm iterates until $\|C_k - C_{k-1}\|$ is less than 10^{-6} .

LiFT is meant to sense low-order modes such as tip/tilt and focus. One way to limit aliasing effects due to the presence of high-order modes is to truncate the image (consider only a few central pixels) because high-order modes create speckles away from the PSF's center [51]. However, image truncation is insufficient to limit these aliasing effects [53]. Another way to limit aliasing effects is to estimate more modes than just TT and focus. However, if no prior assumption is made, LiFT is likely to diverge since there are many different modes' combinations that lead to the same image (or very similar ones). A more efficient solution is to perform a regularized estimation as proposed by Plantet et al. [53]. This implies using a maximum a posteriori (MAP) estimation instead of a maximum likelihood (ML) estimation. In the MAP estimation, We assume that the statistics of residual aberrations are known and the MAP solution is given by:

$$\hat{\mathbf{A}}_{\text{MAP}} = \left(\mathbf{H}(\mathbf{0})^T \mathbf{R}_n^{-1} \mathbf{H}(\mathbf{0}) + \mathbf{C}_\phi^{-1} \right)^{-1} \left(\mathbf{H}(\mathbf{0})^T \mathbf{R}_n^{-1} \Delta \mathbf{I} + \mathbf{C}_\phi^{-1} \bar{\mathbf{A}} \right) \quad (3.16)$$

with $\mathbf{C}_\phi$ the covariance matrix of Zernike coefficients and $\bar{\mathbf{A}}$ their mean. By default, the LiFT results presented will always correspond to the ML estimator. When using the MAP estimator, that will be referred to.

3.3 Fitting methods

Fitting algorithms can offer a more simple and intuitive solution in relation to the algorithms presented before namely GS and LiFT. These algorithms can be particularly effective for low-order mode sensing where the effect of those modes on the PSF is easily predictable. For example, TT provokes a change in the center of gravity of the PSF, whereas defocus contributes to an increase in the FWHM (broadening) of the PSF. In fact, TT is usually estimated using fitting algorithms, namely center-of-gravity-like algorithms, and they offer a simple and robust solution

that is used in almost all ground-based telescopes.

Concerning defocus, a simple way to estimate the amount of defocus would be to fit the PSF with a 2D Gaussian ($G(x, y)$) function given by the following expression:

$$G(x, y) = A \exp \left(-\frac{1}{2} \left(\frac{(x - x_0)^2}{\sigma_x^2} + \frac{(y - y_0)^2}{\sigma_y^2} \right) \right), \quad (3.17)$$

where A is the amplitude (maximum value/height), (x_0, y_0) is the center position, and σ_x, σ_y are the standard deviations in directions x and y respectively. This function is closely related to the Airy pattern function. Doing a gaussian fit could also be used to estimate TT since the center position is also estimated.

The values of σ_x and σ_y are related to the broadening of the Gaussian function. The higher their values, the more broadened will be the Gaussian function. Therefore, it is possible to define a relationship between the standard deviation of the Gaussian fit and the defocus introduced. However, as we've seen before, as a consequence of being an even mode, defocus needs to be estimated using phase diversity. The usage of astigmatism allows for higher linearity ranges. With astigmatism, the PSF will stretch in perpendicular directions according to the sign of the introduced defocus. In the case of astigmatism (0°), the stretching occurs in x and y directions. This kind of behavior can be easily simulated using a Gaussian function as depicted in Fig. 3.2. As we can see, when $\sigma_x > \sigma_y$, the Gaussian function stretches in the x direction, and when $\sigma_x < \sigma_y$, it stretches in the y direction.

As a consequence, the ratio between the values of σ_x and σ_y can be used to determine the defocus value (a_4). An expression of this type can be used:

$$a_4 = \text{sign}(ast) \left(\frac{\sigma_y}{\sigma_x} - 1 \right) \times \alpha. \quad (3.18)$$

The sign function is needed because, according to the sign of the astigmatism introduced, the stretching for positive or negative defocus occurs in different directions. For example, with a positive astigmatism and positive defocus, the PSF stretches in the y direction, but with a negative astigmatism, the stretching would be in the x direction.

Let's consider some cases in order to motivate this equation. According to this expression, the returned defocus is zero when $\sigma_x = \sigma_y$. When this occurs, it means there is no stretching, and zero defocus should indeed be returned. The factor $\frac{\sigma_y}{\sigma_x}$ allows us to determine the stretching direction, and from that, the defocus sign is adjusted accordingly. For example, when $\frac{\sigma_y}{\sigma_x} < 1$, that means the PSF is stretched in the x direction. Finally, the factor α allows us to draw a linear relationship between $\frac{\sigma_y}{\sigma_x}$ and the defocus introduced. This assumption of a linearity

relation between the amount of stretching ($\frac{\sigma_y}{\sigma_x}$) and the defocus introduced is somehow robust and may not be the best. Still, it was used in this work and with reasonably good results in the estimation of defocus.

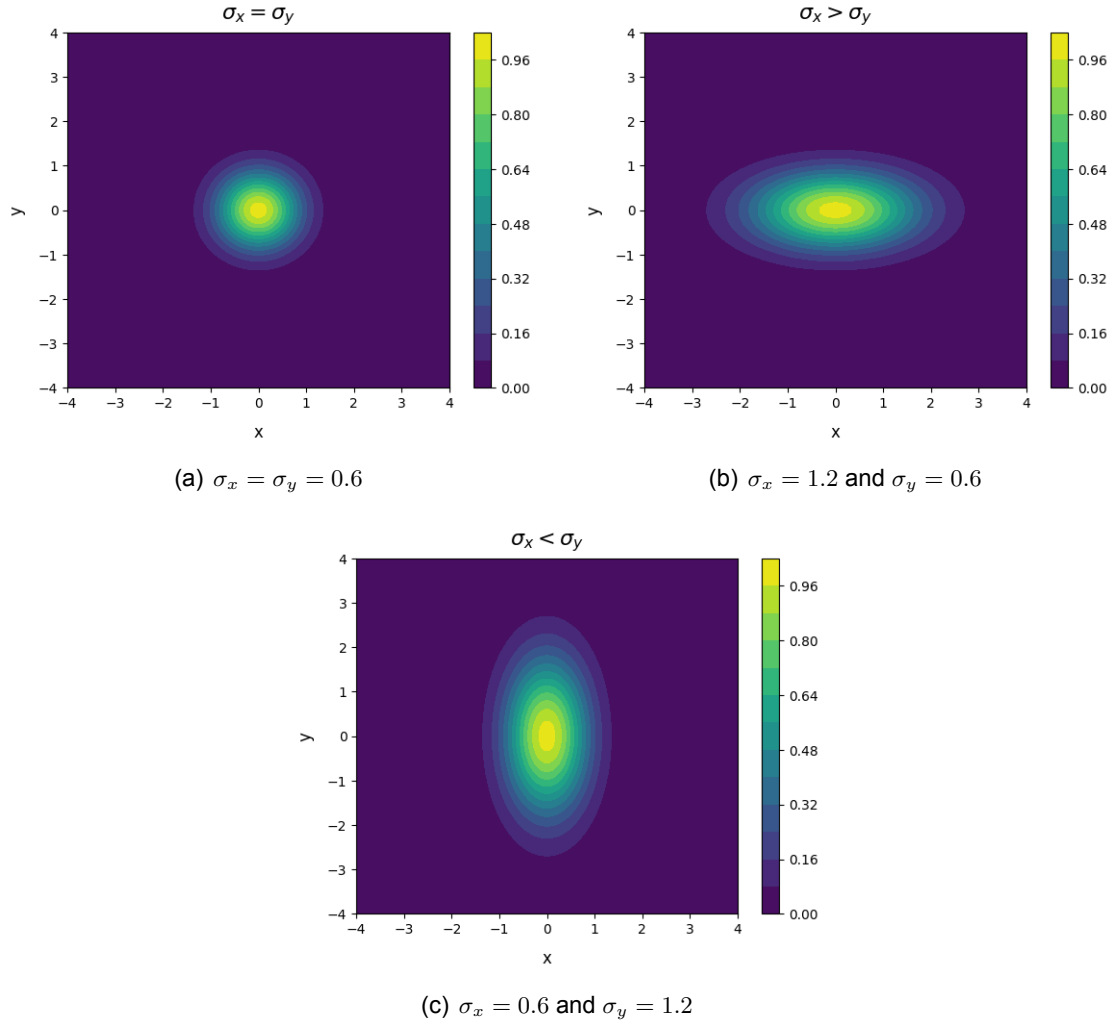


Figure 3.2: Gaussian function pattern for different values σ_x and σ_y . The value used for A was 1, and for the center position (0,0).

From now on, this algorithm will be called the FWHM fit algorithm. This algorithm was not tested in the numerical simulations because it was only considered at a later stage of this work. Still, it was tested with bench and on-sky data.

4 Results

4.1 Numerical simulations

In this section, the performance of two different focal plane wavefront sensing (FPWFS) techniques, namely Gerchberg-Saxton (GS) and the Linearized Focal Plane technique (LiFT), will be presented. These studies were performed through numerical simulations using OOPAO (Object Oriented Python Adaptive Optics), developed by C.T. Heritier [54]. The algorithm used for LiFT was based on a code provided by Arseniy Kuznetsov [55].

GS and LiFT were explicitly studied for estimating TT and focus using faint NGS, simulating an LGS-AO scenario where the measurement of TT and the focus tracking must be done with a faint NGS. Special attention was given to the defocus mode estimation since focus tracking is the main operational issue in LGS-AO systems for ground-based telescopes.

All the simulations presented next were done considering a 10 m diameter telescope with a round pupil, and for the source, an NGS in the band H ($\lambda = 1.654 \mu\text{m}$) was used.

4.1.1 Tests without noise

The main goal of LiFT and GS is to estimate low-order modes from the point spread function (PSF) with a known aberration (phase diversity) applied. In this work, astigmatism (45°) is the aberration used as our phase diversity to solve the classical sign ambiguity for even modes. LiFT and GS are capable of retrieving the modal coefficients of the aberrated phase (Φ), whose modal decomposition is given by

$$\Phi = \sum_j^{N_Z} a_j Z_j, \quad (4.1)$$

where Z_j are the different Zernike polynomials, a_j are the modal coefficients given in rad rms, and N_Z is the number of Zernike modes considered. In the following simulations, different aberrated phase screens corresponding to certain a_j modal coefficients (plus the phase diversity) are generated, and the corresponding PSFs are then computed.

The coefficients a_j are generated using a Gaussian distribution with mean zero and a certain standard deviation (σ_n). The standard deviation values follow a $(n + 1)^{-2}$ power law, with n being the radial order of the Zernike polynomial:

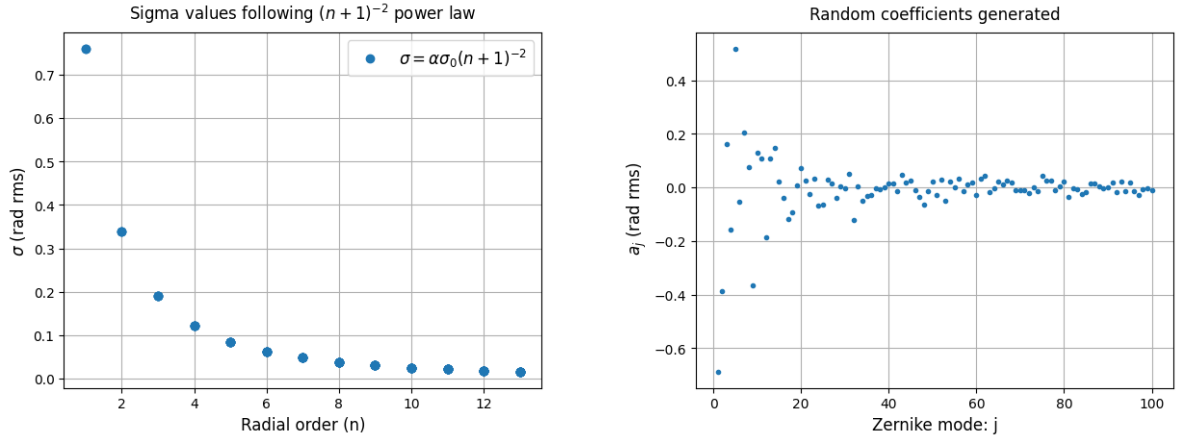
$$\sigma_n = \alpha \sigma_0 (n + 1)^{-2}, \quad (4.2)$$

where α is a constant chosen to be equal to four so that $\sigma_1 = \sigma_0$. This kind of power law for the different Zernike modes is a good approximation to a real AO system operation [10].

In Fig. 4.1 a), we can see the variation of the σ values according to the radial order n , and in Fig. 4.1 b), I present a random sample of Zernike coefficients up to $Z = 100$. We can see that the amplitudes of the modes with higher radial orders are lower. The radial order n is related to the Noll indices j according to the following equation:

$$n = \lfloor (\sqrt{2j - 1} + 0.5) - 1 \rfloor, \quad (4.3)$$

where $\lfloor x \rfloor$ represents the floor function that gives as the output the greatest integer less than or equal to x [12]. For instance, tip and tilt correspond to $n = 1$, defocus and astigmatism (0° and 45°) correspond to $n = 2$.



(a) Standard deviation variation according to $(n+1)^{-2}$ power law. The values of σ_0 and α are 0.76 rad rms and 4 respectively.

(b) Random sample of Zernike coefficients (up to $Z = 100$) following a Gaussian distribution: $N(\mu=0, \sigma=\sigma_n)$.

Figure 4.1: Random generation of Zernike coefficients using a Gaussian distribution with mean zero and standard deviation σ_n . The constant α was chosen to be 4 so that $\sigma_1 = \sigma_0$.

The random amplitudes for the different modes (a_j) are used to generate different aberrated phase screens. From these phase screens, the corresponding PSFs are then generated and used as input for GS and LiFT, which, from the PSFs, will retrieve the different sets of amplitudes for the different modes. Let $\hat{a}_j$ be the retrieved coefficients from LiFT or GS, then the phase error obtained in rad rms is given by:

$$\sigma_\epsilon = \sqrt{\sum_j^{N_Z} (a_j - \hat{a}_j)^2} \quad (4.4)$$

This will be the metric used to evaluate the performance of the algorithms. I should note that the piston term is not taken into account in these studies, and in fact, it would not have any

effect on the phase error. As a general rule, a phase error of 0.45 rad rms is an indicator of a very good performance [14].

The tests without noise were performed for the first 10 Zernike modes and different sampling conditions. For each sampling condition, I studied the algorithms' performances for increasing amounts of aberrations introduced for three different values of phase diversity: 100, 200, and 300 nm rms, which correspond to 0.38, 0.76, and 1.14 rad rms for band H ($\lambda = 1.654 \mu\text{m}$).

The Zernike coefficients were generated following a Gaussian distribution with mean zero and standard deviation given by Eq. 4.2. To simulate different phase screens with different amounts of aberrations I simply varied the value of σ_0 . Let f be a variable representing the ratio between σ_0 to the diversity introduced a_d . I varied $\log_{10}(f)$ from -2 to 0.5, which corresponds to f values between 0.01 and 3.16. For each f value, I simulated 25 random phase screen samples. A new variable, which I call r , should now be introduced since it will be used extensively throughout the plots in this report:

$$r = \frac{\langle rms \rangle}{a_d}, \quad (4.5)$$

where $\langle rms \rangle$ represents the mean rms value of the phase screens obtained with the 25 samples, and a_d represents the amount of phase diversity. I should note that $\langle rms \rangle$ is calculated taking into consideration all the aberrations introduced except the phase diversity. For example, an r value of 1 means that the mean rms phase value of the 25 samples is equal to the rms value of the phase diversity introduced, astigmatism (45°), in the case of this study.

In Fig. 4.2, I present the results for both GS and LiFT for a sampling of 1.0 Nyquist (2 pixels per FWHM at diffraction limit) for different amounts of phase diversity applied. We can see that for low r values, the performance of both algorithms is excellent (σ_ϵ of the order 10^{-8}), but when reaching values near 1, the phase error starts rapidly increasing. From these results, we can also conclude that for a certain amount of phase diversity applied, the algorithms can only perform well when the amount of aberrations is below the amount of phase diversity applied. This means that a higher amount of phase diversity applied allows for retrieving phase screens with more aberrations.

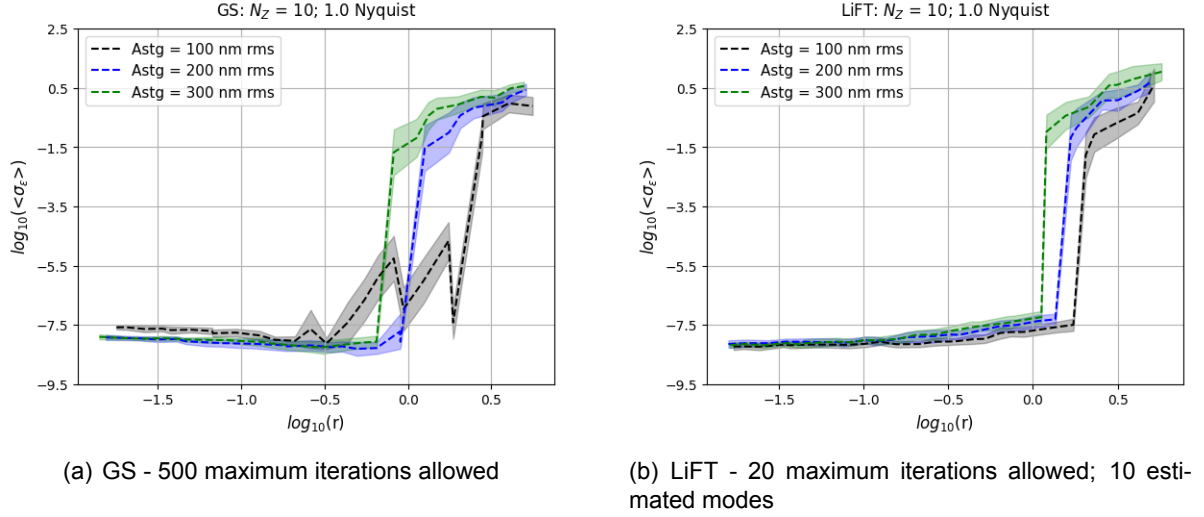


Figure 4.2: Performance of GS and LiFT according to phase error (see Eq. 4.4). For each value of r and astigmatism (phase diversity), 25 random samples were generated and the mean of the logarithmic phase error was calculated. The values used for the error bars are the standard deviations obtained.

One important aspect concerning GS and LiFT performance is the number of required iterations to achieve convergence. The convergence criteria for GS and LiFT were already presented in section 3. The number of iterations needed to achieve convergence with GS highly depends on the amount of phase diversity applied and on the value of r as shown in Fig. 4.3, which reflects the variation of the mean number of iterations for the algorithm to converge from the simulation regarding Fig. 4.2 a). We conclude that higher values of r require more iterations to converge for all amounts of phase diversity applied and that higher amounts of phase diversity mean fewer iterations to achieve convergence. On the other hand, LiFT doesn't have such a big dependence on the number of iterations. In fact, the results shown in Fig. 4.2 b) were obtained with a maximum number of iterations of just 20. Of course, as r increases, the number of iterations also slightly increases, but not in the same proportion as happens with GS.

Even though LiFT requires fewer iterations to converge, that does not necessarily mean that it is faster than GS since the computational costs per iteration are not the same for GS and LiFT. In order to quantify the computational costs of GS and LiFT, I measured the elapsed time for GS and LiFT according to the number of iterations. In the case of LiFT, the costs also depend on the number of estimated modes, so that was also taken into account for LiFT. The results are presented in Fig. 4.4. This study about computational costs was performed considering a Nyquist sampled PSF and pupil resolution of 128 pixels. I should note that using a higher sampled PSF or pupil resolution would increase the computational costs of both GS and LiFT. This happens because both algorithms use 2D Fourier transforms.

One interesting aspect is verifying that LiFT computational costs increase considerably

when we estimate more modes. On the other hand, GS does not have that dependence since it estimates a phase rather than specific modes. From that phase, we can extract as many modes as we want by projecting the phase on a specific modal basis, such as the Zernike basis, for instance, and this is a big advantage compared to LiFT. This study shows that LiFT performs faster when estimating a few modes, but when we increase the number of estimated modes, the computational costs of LiFT and GS tend to balance themselves.

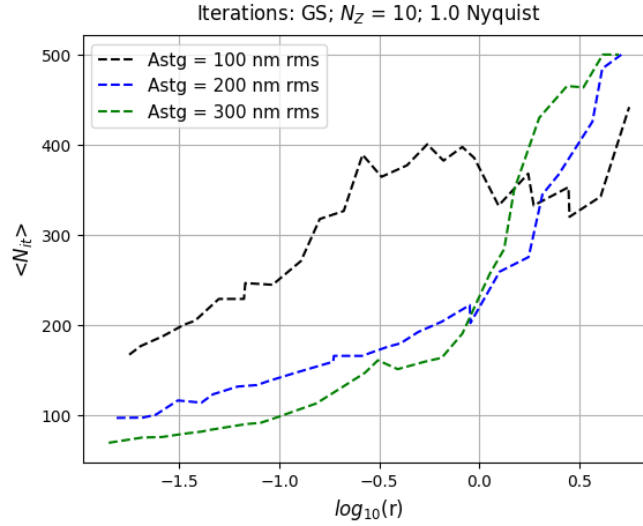
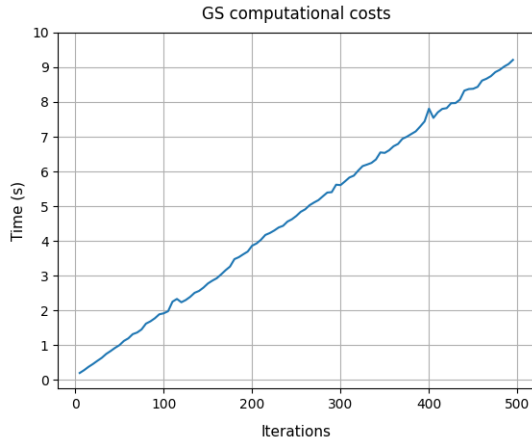
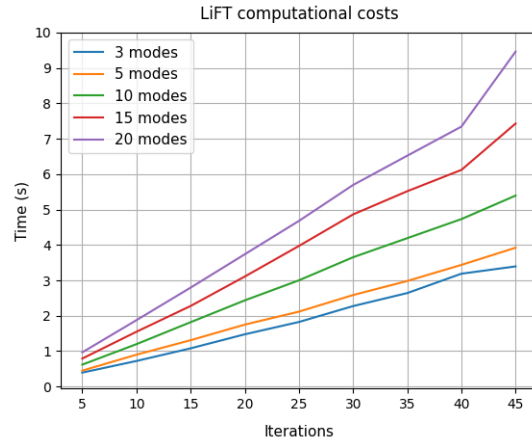


Figure 4.3: Mean iterations ($\langle N_{it} \rangle$) cost for GS with r for different amounts of phase diversity applied.



(a) Computational costs of GS according to number of iterations.



(b) Computational costs of LiFT according to the number of iterations for different number of estimated modes.

Figure 4.4: Computational costs in time for GS and LiFT considering 1.0 Nyquist PSF and a pupil with 128×128 pixels.

One last aspect concerning the data shown in Fig. 4.3 should be pointed out. For an astigmatism of 100 nm rms, the number of iterations stagnates and then decreases from $\log_{10}(r)$ values from -0.5 to 0.5. This behavior was not verified for 200 and 300 nm rms, as the mean

number of iterations always increases with increasing values of r . If we look at the results presented in Fig. 4.2 a), we can verify that within this r range, the results for 100 nm rms indeed follow a different pattern relative to the others. It seems that GS is interpreting the higher amount of aberrations as an "extra phase diversity", although this term is not scientifically correct. A plausible explanation for this behavior has not been found yet. However, to investigate this further, I performed the same tests for GS, but this time, I did so with a sampling of 2 Nyquist (4 pixels per FWHM at the diffraction limit). The results are shown in Fig 4.5.

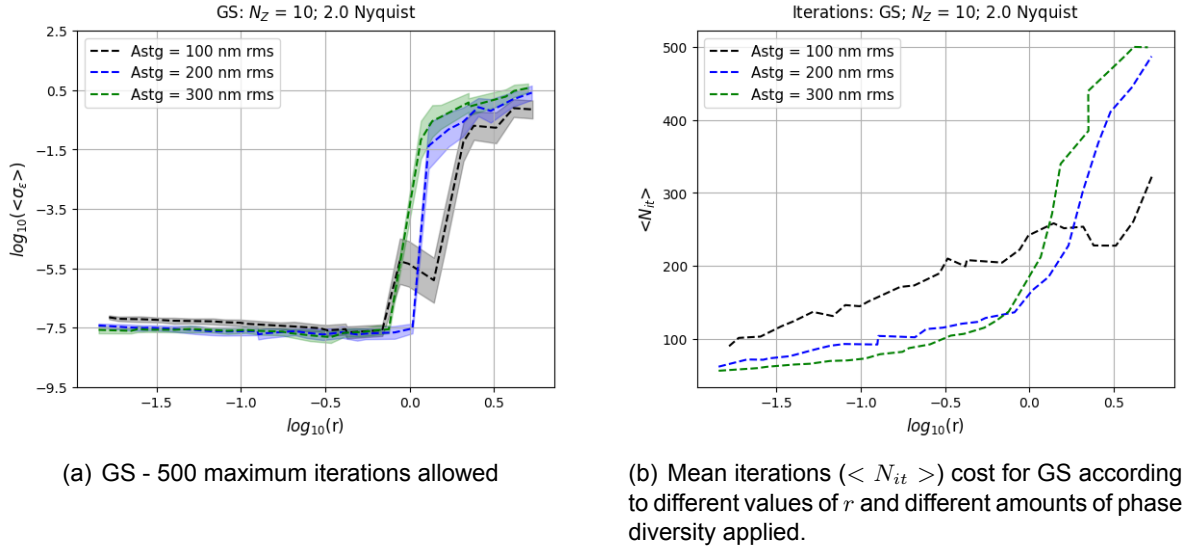


Figure 4.5: Results obtained in the same conditions as the ones presented in Fig. 4.2 a) except for the sampling, which is 2.0 Nyquist instead of 1.0 Nyquist.

From the results obtained, we can see that for 2 Nyquist, the number of iterations to achieve convergence is significantly lower mainly for an astigmatism of 100 nm rms. However, this doesn't mean that the algorithm performs faster with 2 Nyquist in comparison to 1 Nyquist. In fact, it is the opposite. The computational cost per iteration for 2 Nyquist is much higher which means that the algorithm performs slower even though it requires fewer iterations.

Regarding the results obtained for 100 nm rms, we see once again that there is a certain stagnation of the mean iterations required for convergence contrary to what happens with the higher amounts of diversity. The results for 2 Nyquist seem to be in line with the ones obtained for 1 Nyquist in what concerns the mean number of iterations, which doesn't provide much useful information for explaining the results for 100 nm rms of diversity.

Next, I will show the results concerning sub-Nyquist sampling, namely for 0.5 Nyquist (1 pixel per FWHM at diffraction limit) as depicted in Fig. 4.6. With this data, we see a clear performance difference between GS and LiFT. On one hand, LiFT performs just as well as in the case of 1.0 Nyquist, which shows that LiFT is not very affected by the loss of spatial

information in the PSF. On the other hand, GS has a clear performance decline when we go to 0.5 Nyquist. At this sampling, GS is never capable of achieving the performance of LiFT even if I allowed a much higher number of iterations. These results were somehow expected since GS is not fully prepared to work with sub-Nyquist sampled PSFs.

Another aspect concerning GS performance at 0.5 Nyquist should be pointed out. From Fig. 4.6 a), we can see that the presence of more aberrations (higher r values) slightly improves the algorithm's performance. Of course, when the r values are near one, that behavior disappears. It seems that GS interprets the presence of more aberrations as a kind of "extra phase diversity". Once again, it should be noted that the phase diversity term should not be used in this context since diversity assumes that we are dealing with a known amount of phase aberration, which is not the case here.

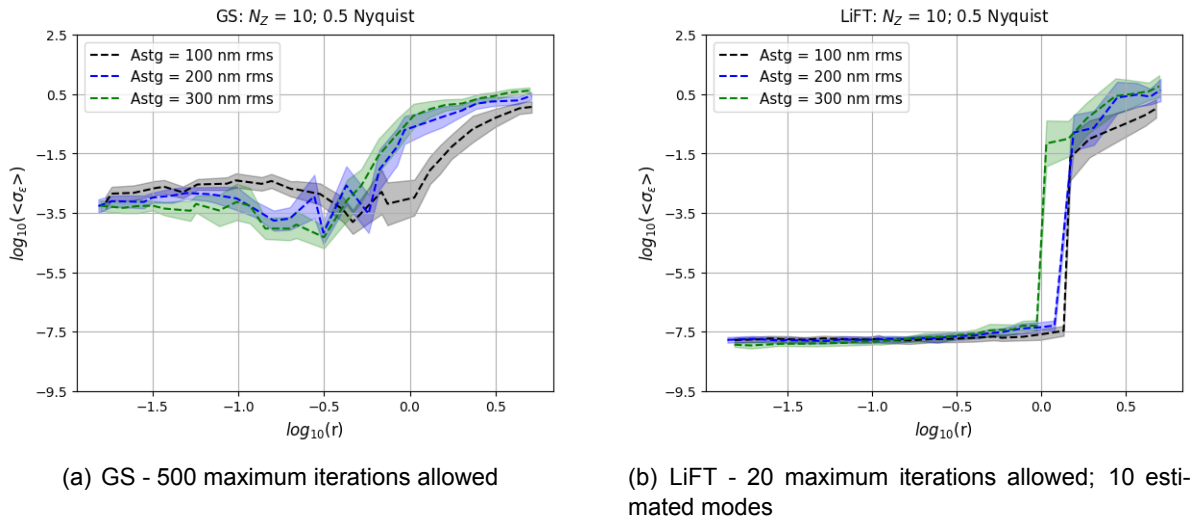


Figure 4.6: Performance of GS and LiFT according to phase error (see Eq. 4.4). For each value of r and astigmatism (phase diversity), 25 random samples were generated and the mean phase error was calculated. The values used for the error bars are the standard deviations obtained.

Below I also have the results obtained for 0.25 Nyquist sampling. Once more, we see that LiFT is still capable of excellent performance well below Nyquist sampling. I only have results for LiFT since my current GS algorithm cannot deal with PSFs sampled below 0.5 Nyquist. However, as I'll show in the following sections, we can still use GS in these cases by assuming a "wrong" sampling. Of course, optical gain errors will arise, and calibration is needed.

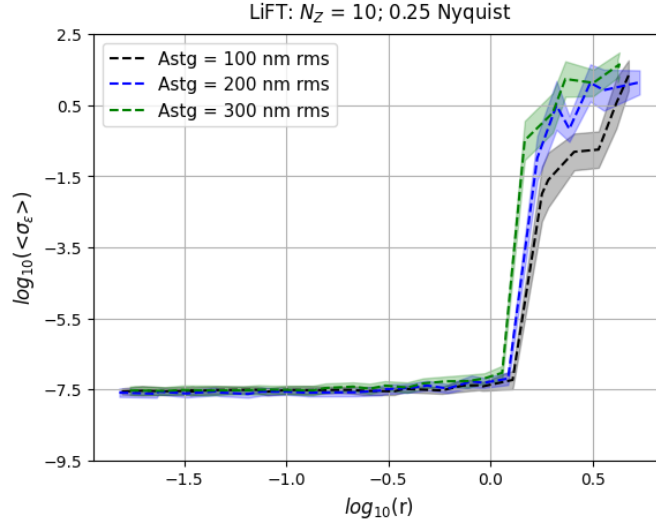


Figure 4.7: LiFT - 20 maximum iterations allowed; 10 estimated modes.

These noiseless tests allowed the identification of some differentiation aspects between LiFT and GS. The main difference between the two lies in their performance under sub-Nyquist sampling, where LiFT surpasses GS since it is prepared to deal well with those cases. Another disadvantage concerning GS is the high iteration cost when working with high r values and/or low amounts of phase diversity. Concerning LiFT, the sole disadvantage identified was the increase in computational costs as we increase the number of estimated modes. However, premature conclusions should not be taken at this point since noiseless tests are very far away from the real conditions where these algorithms are expected to be deployed. In the next section, more realistic tests will be presented and discussed.

Before going into the realistic tests, I will still present some results obtained with GS for high-order mode estimation. I will showcase GS performance for 1 and 0.5 Nyquist for the estimation of up to 200 Zernike modes. The computational costs of doing such simulations with LiFT would be tremendous, and this is one of the aspects that GS sets apart from LiFT.

The estimation of up to 200 modes requires diversity to be applied through more modes than just astigmatism as previously used. Phase diversity was applied in all 200 modes, whose modal coefficients follow a Gaussian distribution with mean zero and standard deviation (σ_d) given by Eq. 4.2. The value chosen for $\sigma_{0,d}$ was 0.76 rad rms, which corresponds to a mean diversity ($\langle rms \rangle_d$) of approximately 360 nm rms. It should be emphasized that this diversity was always different for each sample studied but always with the same value for $\sigma_{0,d}$. The results of GS performance for high-order estimation are shown in Fig. 4.8.

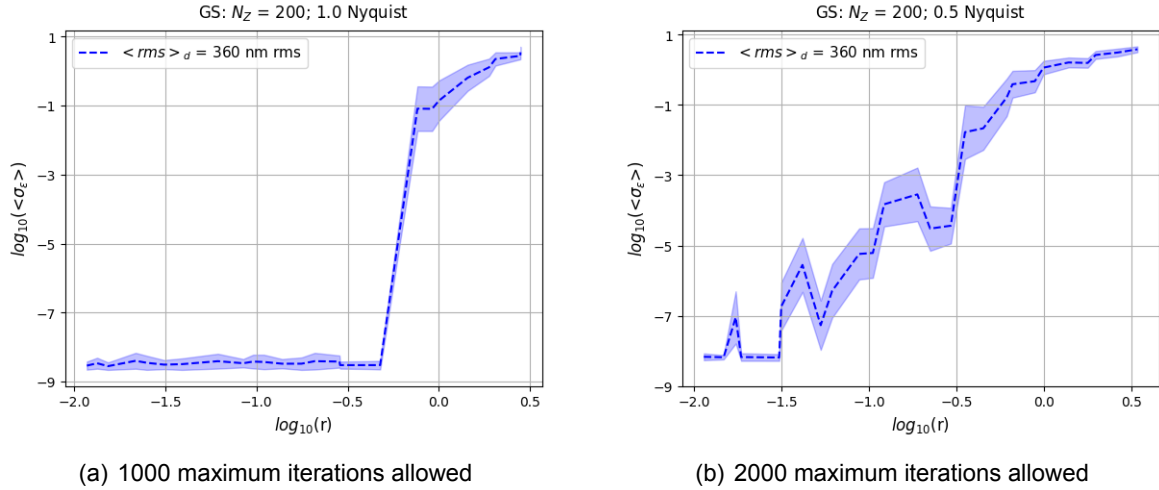


Figure 4.8: For each value of r , 25 random samples were generated and the mean of the logarithmic phase error was calculated. The values used for the error bars are the standard deviations obtained.

I should refer to the fact that diversity doesn't necessarily need to be applied in all 200 modes. Before running the simulation, I performed some preliminary studies where I verified that applying diversity only to a few modes (for example, from 10 to 10 modes) can be sufficient. From the studies I did, I would say that the presence of diversity on one mode per radial order should allow results similar to the ones presented in Fig. 4.8. Of course, the more diversity in more modes, the better.

The results shown in Fig. 4.8 show once again that GS performance is highly affected by sampling. As we can see, results for 1 Nyquist are much better, and the maximum number of iterations allowed for 1 Nyquist was just half of the ones allowed for 0.5 Nyquist. However, this doesn't mean that the performance at 0.5 Nyquist is bad since we have σ_ϵ values lower than 0.1 for r values almost up to 1.0, which is indeed very good. Of course, this study is done under ideal conditions with no noise, but it shows that GS is an algorithm that works not only for low-order estimation but also for high-order estimation as long as the necessary diversity is applied through those modes. This is done with no algorithmic changes but with an increase in computational cost due to a higher number of iterations required.

4.1.2 Tests with noise and no high-order residuals

In this section, GS and LiFT are studied under realistic conditions of low flux (faint stars), sub-Nyquist sampling, noise, and also with the presence of high-order residuals. The tests that will be presented were performed using a sampling time of 0.1 s and an exposure time of 1 s. The PSF with noise is simulated using the medium of 10 independent noise realizations. In this part, the amount of phase diversity (astigmatism 45°) applied was fixed at 200 nm rms for all

simulations.

In this part, we only apply tip/tilt (TT) and focus modes and completely ignore the presence of high-order modes (except, of course, the astigmatism that is our phase diversity). By high-order modes, we consider all modes after focus.

Here, the performances of GS and LiFT with 0.5 Nyquist sampling and the presence of photon noise are studied. In the next part, the presence of high-order residuals will also be investigated. The results obtained are shown in Fig. 4.9. Each algorithm was studied for different conditions of signal-to-noise ratio (SNR) by varying the magnitude (m) of the natural guide star (NGS). The variables r and σ_ϵ have already been introduced previously, and once again, each point in the graphs is obtained by calculating the medium values for the error using 25 random samples.

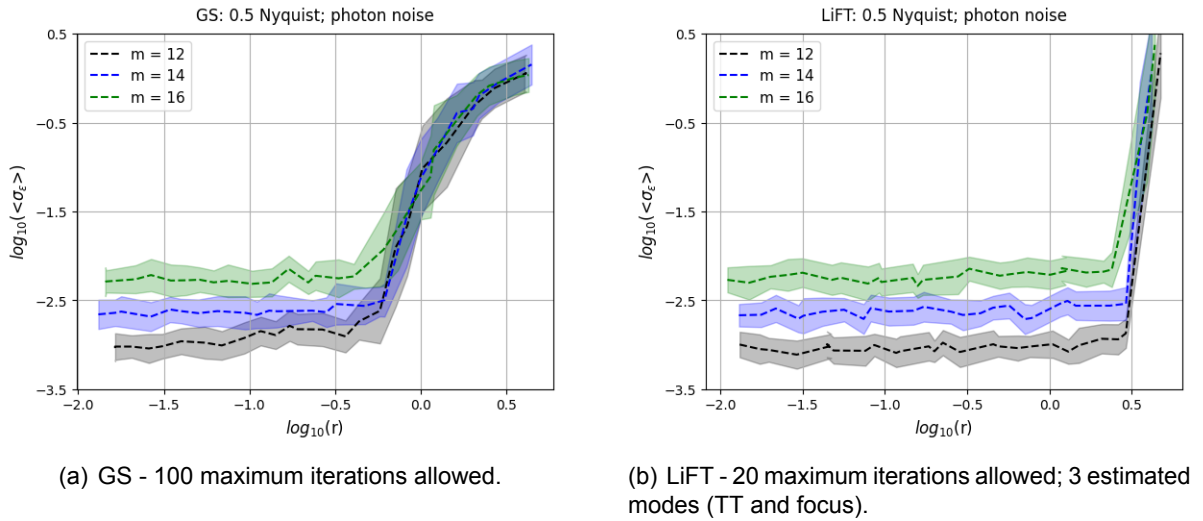


Figure 4.9: Results for both GS and LiFT for 0.5 Nyquist sampling with photon noise, but without high-order residuals. Performance was investigated for different NGS magnitudes.

One surprising result is the good performance of GS when applying photon noise. From the results obtained without noise for 0.5 Nyquist (see Fig. 4.6), the difference in performance between GS and LiFT was much bigger than now when considering photon noise. To further explore this I studied GS performance with and without noise for 1.0 Nyquist sampling. The results are presented in Fig. 4.10. These results show a huge performance decline when considering noise similar to what was verified for LiFT. With 0.5 Nyquist samplings, GS performance is already highly affected even without noise, and for that reason the performance decline when considering noise is not as evident, although it exists. These results seem to show that GS and LiFT are, in fact, equally affected by the presence of noise.

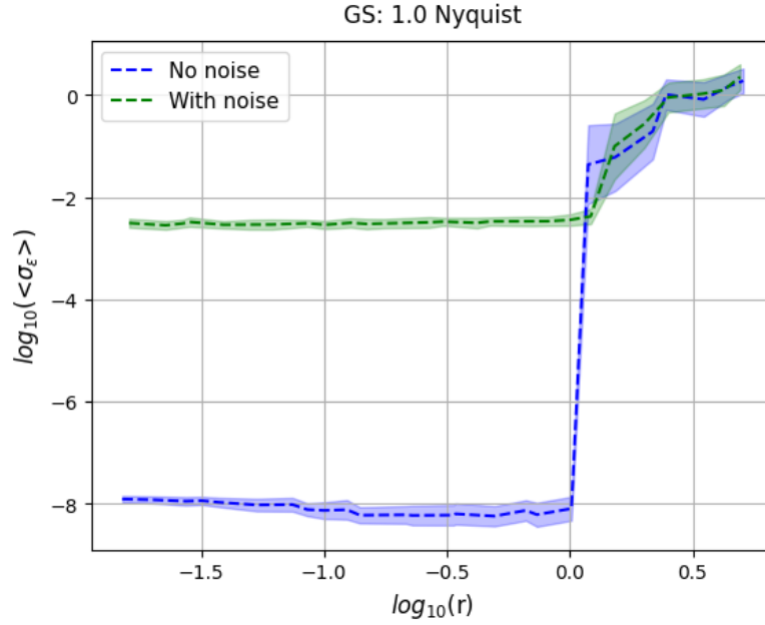


Figure 4.10: GS performance decline when considering the presence of noise. The values for σ_ϵ were calculated considering the first ten modes. A star with $m=14$ was used.

From the data of Fig. 4.9, I conclude that as expected, the higher the magnitude (m), the worse the algorithms's performances. Higher magnitude means lower flux, which also means lower SNR values, so these results were expected. Both algorithms can perform quite well (σ_ϵ of the order of -2.0) even for a magnitude 16 star. One main difference between GS and LiFT is the r range at which the algorithm starts failing. GS starts failing for lower r values contrary to LiFT, which holds its performance almost over the whole range.

It should be emphasized that the maximum number of iterations for GS was fixed at 100, which is five times lower compared to what we used in the noiseless tests. Allowing more iterations would not improve much the results since the algorithm only converges for approximately the first 100 iterations (tests quantifying GS performance for different numbers of iterations will be presented in the next section). Even though, the fact that it only converges approximately for the first 100 iterations does not limit its performance at all when we compare the results with LiFT. In the next section, the presence of high-order modes will also be taken into account to simulate even more challenging conditions, yet realistic ones.

4.1.3 Tests with noise and high-order residuals

In the previous section, we considered that the only modes present were TT and focus, which is unrealistic since that assumes a perfect correction of all higher-order modes from the LGS wavefront sensor being used. In addition, sometimes the astronomical objects that we wish to observe (at which LGS will be pointed) may be at a large angular separation from

the nearest NGS, which will result in even larger wavefront residuals. Performance under the presence of high-order residuals (Δ_r) is then an important aspect to study.

In the study that will be presented, all the values used for sampling time, exposure time, and amount of phase diversity are the same as in the previous section. I considered high-order modes up to $Z = 100$, whose coefficients were randomly generated by using a Gaussian distribution, whose standard deviation values follow a $(n + 1)^{-2}$ power law (see Fig. 4.1).

In Fig. 4.11 we have the results obtained for both GS and LiFT for a 0.5 Nyquist sampling using a magnitude 14 NGS. It is important to note that the values for σ_ϵ are calculated only taking into account TT and focus, which are the modes we are interested in. The r values, already introduced before, describe how much the amplitudes of TT and focus vary to the amount of phase diversity applied. According to the results obtained, the larger the amount of wavefront residuals, the worse the algorithms' performance as expected. Once again, GS starts decreasing in performance for lower values of r in comparison to LiFT. Even when no wavefront residuals are present, we verify that in GS results. From an attentive look at both graphs of Fig. 4.11, we can see that GS has a better performance (although for a lower r range) in comparison to LiFT when considering high-order residuals.

In Fig. 4.12, I present the results obtained for the estimation of the first ten modes, other than just TT and focus. Once more GS has a better performance under the presence of high-order residuals. Both algorithms can efficiently estimate more modes than just TT and focus, but in the case of LiFT, that comes associated with a high increase in computational costs, as already seen before.

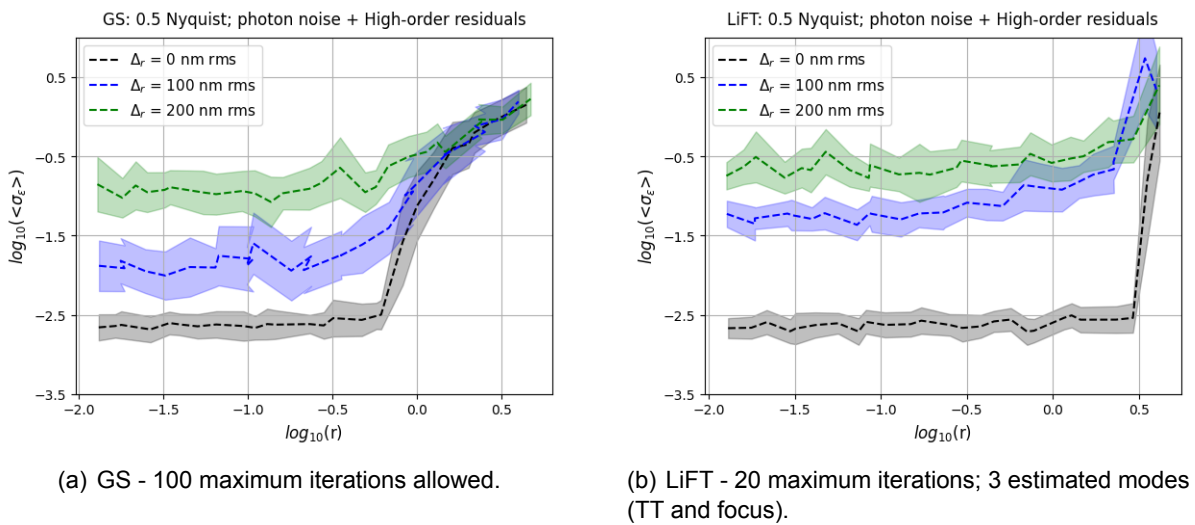


Figure 4.11: Results for both GS and LiFT for 0.5 Nyquist sampling with photon noise and also high-order residuals. The star's magnitude considered here was 14. Only values of TT and focus were used for the calculation of σ_ϵ .

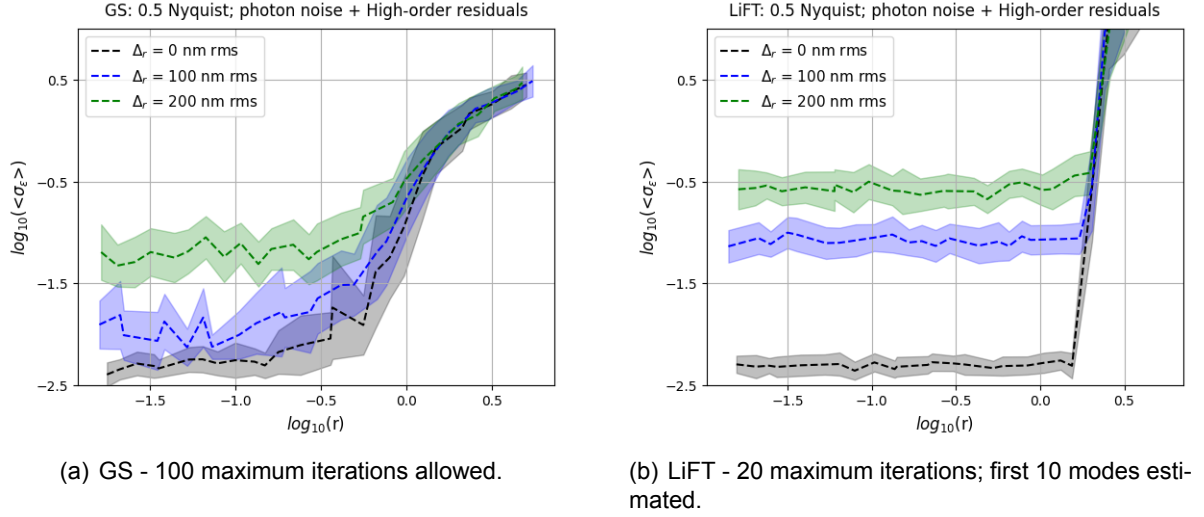


Figure 4.12: Results for both GS and LiFT for 0.5 Nyquist sampling with photon noise and also high-order residuals. The star's magnitude considered here was 14. The values of σ_ϵ were calculated taking into consideration all the first ten modes.

Now, I will present some tests conducted to evaluate the possibility of improving the results of LiFT and GS. In the case of GS, the only way to improve performance is to increase the number of iterations. In Fig. 4.13, I present the results obtained for two different numbers of allowed iterations: 100 and 1000. From the results obtained, we can see that 100 iterations are a good trade-off between computational cost and the algorithm's performance.

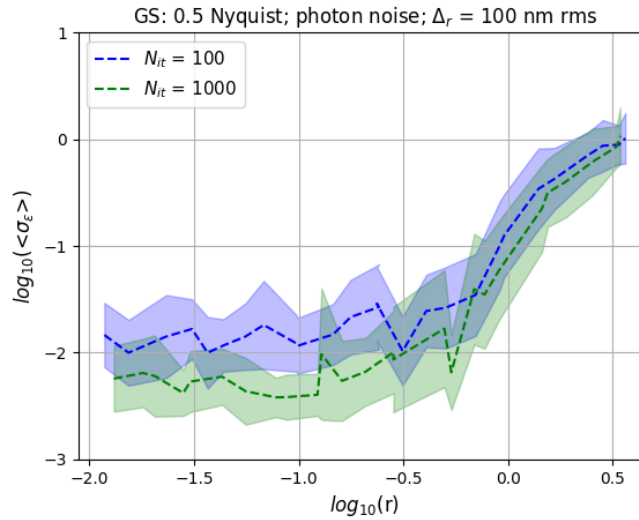


Figure 4.13: GS performance according to the different number of iterations. A natural star of magnitude 14 was used.

Perhaps, a more efficient way to improve GS performance would be by having more diversity through high-order modes. This could be done by using the deformable mirror to introduce that diversity. As already seen in the previous section, GS can be used to retrieve many more modes than just TT and focus without any algorithmic modifications or significant increases

in computational costs. By having more diversity through high-order modes, it is expected a more significant performance improvement of GS with more allowed iterations contrary to what happens without that extra diversity, where performance improvements are negligible.

Lastly, I now present some tests performed with LiFT in an attempt to study possible improvements, namely by allowing more modes to be estimated. The results obtained show a slightly better estimation of TT and focus by allowing the estimation of the first ten modes (Fig. 4.14). The estimation of more modes other than just TT and focus is expected to mainly help in the estimation of defocus whose estimation is affected by the presence of high-order modes (crosstalk effects). Even though the improvements are not very significant, it is clear that the estimation of up to 10 modes doesn't affect the estimation of TT and focus, which means that it can be safely used to correct NCPAs for instance. However, the estimation of more modes means a big increase in computational costs, which is one of the biggest disadvantages of LiFT when compared to GS.

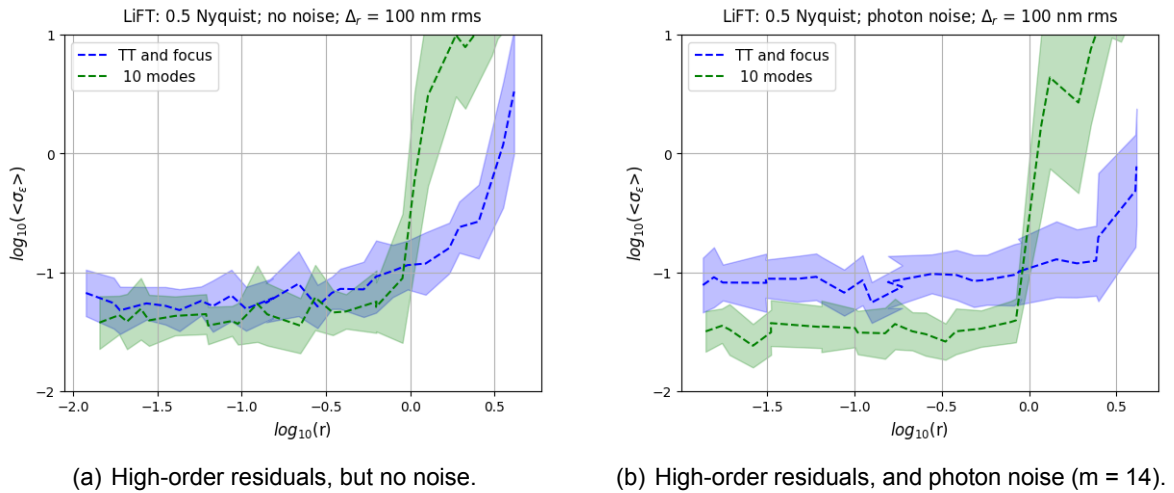


Figure 4.14: LiFT performance on the estimation of TT and focus by allowing a higher number of modes to be estimated. The values of σ_ϵ were calculated taking into consideration only TT and focus.

The results presented in Fig. 4.14 were obtained by using LiFT's maximum likelihood (ML) estimator. To investigate if, by using the maximum a posteriori (MAP) estimator, we could further improve LiFT's performance, the same simulation as the one used to obtain results from Fig 4.14 was performed with the MAP estimator. The results are shown in Fig. 4.15. Once more, we see that estimating more modes allows achieving a slightly better performance. However, the results obtained with the MAP estimator are very similar to the ones obtained with the ML estimator, meaning that the performance gains with the MAP estimator did not prove significant.

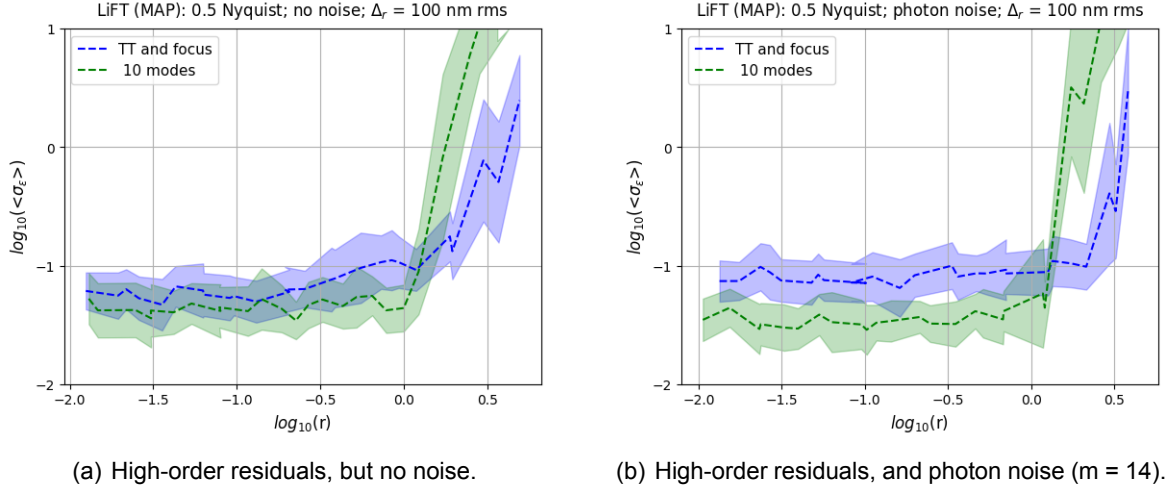


Figure 4.15: LiFT performance on the estimation of TT and focus by allowing a higher number of modes to be estimated. The maximum a posteriori (MAP) estimator was used. The values of σ_ϵ were calculated taking into consideration only TT and focus.

4.1.4 Focus ramps

As already introduced, LGS wavefront sensing is insensitive to TT, and focus estimates are compromised by the variability of the sodium layer's altitude. As a consequence, an NGS needs to be used for low-order sensing. Tip/Tilt modes are usually corrected by using algorithms based on center of gravity concepts, and they have already proven to work efficiently. On the other hand, current algorithms used for focus tracking aren't yet completely satisfactory. From a strictly operational point of view, focus estimation is indeed the main issue at present. For this reason, focus ramps for both GS and LiFT will be presented in this section.

As we know, focus is an even mode, which means that the PSF will manifest itself equally both for positive or negative values, provided that they have equal absolute values. The focus manifests itself in the PSF by provoking its enlargement, which means that the full width at half maximum (FWHM) increases. In Fig. 4.16, I present an illustration of the expected quadratic behavior of the FWHM of the PSF for increasing values of the introduced focus.

This quadratic behavior means that the PSF will be the same whether we apply +100 nm rms or -100 nm rms for instance. This is the known sign ambiguity for even modes. This ambiguity is usually solved by introducing phase diversity. Defocus itself can be used as phase diversity to solve its own sign ambiguity. However, as I will show next, using astigmatism as diversity can yield higher linearity ranges.

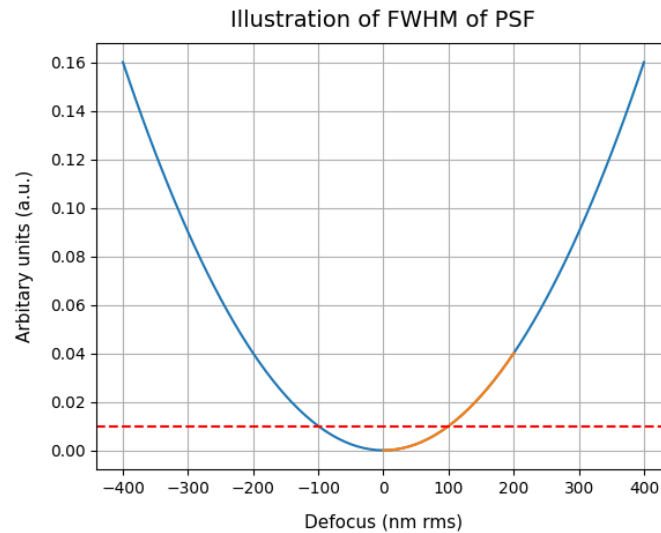


Figure 4.16: Illustration of full width at half maximum (FWHM) of the PSF for positive and negative values of defocus introduced.

In Fig. 4.17, I have focus ramps for both GS and LiFT under ideal conditions (no noise and no high-order residuals), with 200 nm rms of defocus as phase diversity. From the results obtained, we can clearly see that both algorithms start failing for values of defocus lower than -200 nm rms. This result was already expected, and to explain this further, I will give an example. Let's say, for instance, we introduce -300 nm rms; this means that we'll have an overall of -100 nm rms of defocus. On the other hand, if we had introduced -100 nm rms, we would have got an overall defocus of +100 nm rms. The PSF for ± 100 nm rms is the same (Fig. 4.16), which means that the algorithms will converge to the solution that is closer to the initial guess, which typically is phase diversity alone. In this way, the algorithms will retrieve -100 nm rms since it is the closest solution. I should note that this linearity limitation is not exclusive to defocus; it applies to all even modes. For instance, if we wanted to estimate astigmatism by using astigmatism as diversity, the same linearity problems would arise.

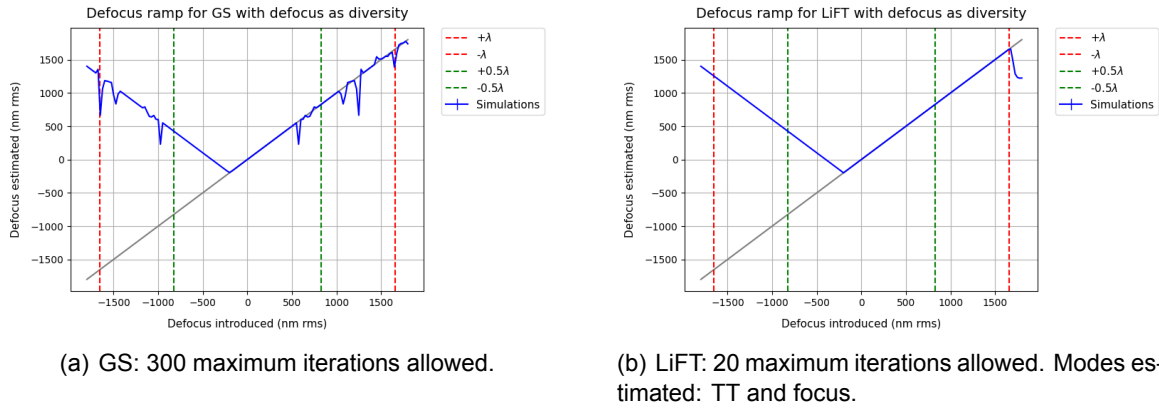


Figure 4.17: Focus ramps for GS and LiFT for a sampling of 1 Nyquist. An amount of 200 nm rms of defocus was used as phase diversity. Results were obtained without taking into consideration noise or high-order residuals. The value of λ is $1.654 \mu\text{m}$ (band H).

To obtain higher linearity ranges in defocus ramps, astigmatism can be used as phase diversity. When we have astigmatism as phase diversity, a positive defocus will stretch the PSF in a certain direction, and a negative defocus will stretch it in the perpendicular direction, thus completely solving the sign ambiguity. Let's imagine a similar example as the one presented before, where we have 200 nm rms of diversity, but this time astigmatism instead of defocus. Now, if we introduce -300 nm rms of defocus, the algorithms will converge to the correct solution since there is only one. With astigmatism, -100 and -300 nm rms of defocus yield different PSFs. For -300 nm rms, the PSF will be much more stretched.

The results obtained with astigmatism as phase diversity are shown in Fig. 4.18. For LiFT, the linearity range is indeed much higher as was expected, but the results of GS do not follow a similar pattern. GS results show linearity from -250 nm rms to +250 nm rms, which is much lower when compared with LiFT. For positive values of defocus, the results are worse when using astigmatism as phase diversity. In an attempt to seek the reason for such behavior, some more tests were performed and will be presented next.

First I tested the behavior of the GS's linearity range with two different amounts of phase diversity applied: 100 and 200 nm rms. The results are shown in Fig. 4.19 and we see that 100 nm rms of astigmatism yields a slightly higher linearity range. From my perspective, this may be associated with crosstalk between modes especially between defocus and astigmatism. If we introduce a very large amount of astigmatism as diversity, the PSF will get enlarged and the stretching due to the addition of defocus will not be evident at all. This means that GS will most certainly attribute the "fault" of PSF enlargement to astigmatism instead of defocus. For GS, higher amounts of astigmatism as diversity seem to increase crosstalk problems.

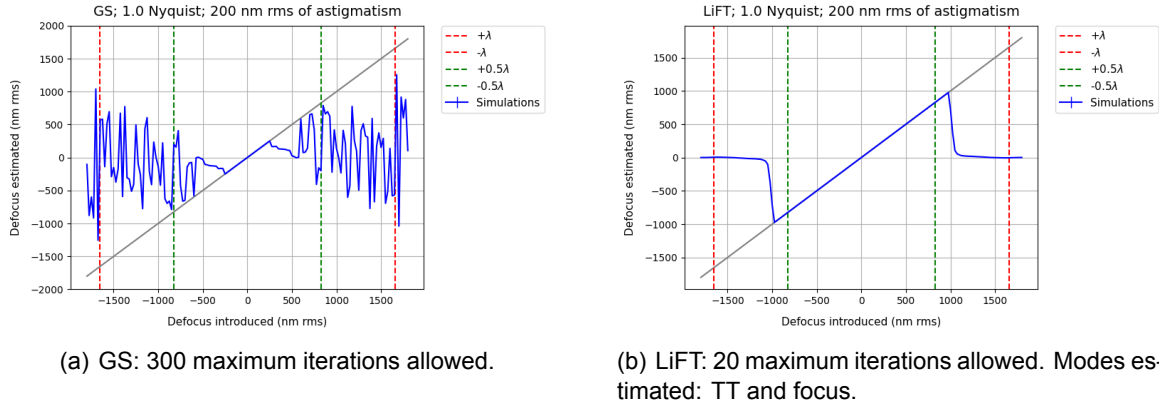


Figure 4.18: Focus ramps for GS and LiFT for a sampling of 1 Nyquist. An amount of 200 nm rms of astigmatism was used as phase diversity. Results were obtained without taking into consideration noise or high-order residuals. The value of λ is $1.654 \mu\text{m}$ (band H).

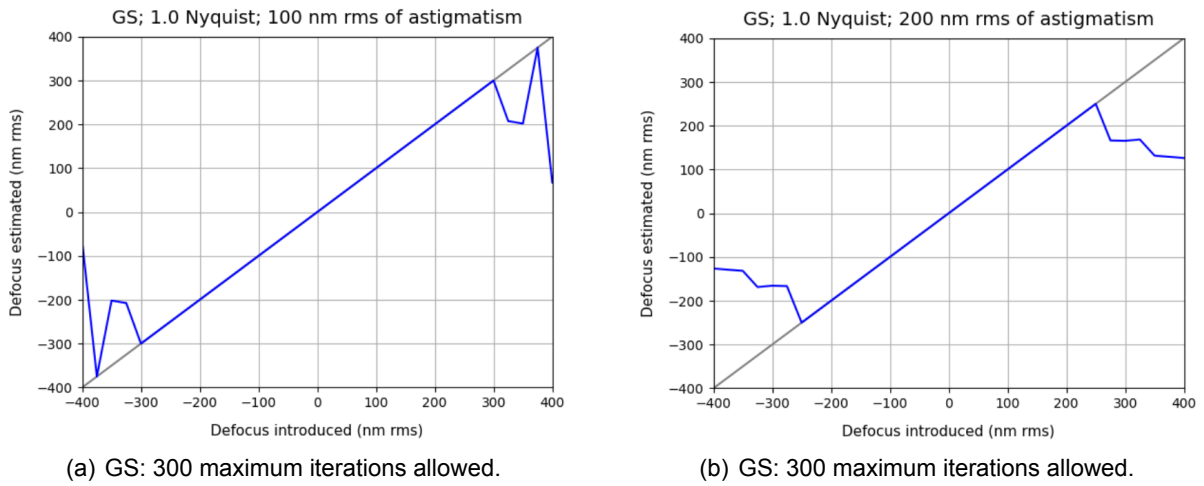


Figure 4.19: Focus ramps for GS for a sampling of 1 Nyquist for 100 and 200 nm rms of astigmatism as diversity. Results were obtained without taking into consideration noise or high-order residuals. The value of λ is $1.654 \mu\text{m}$ (band H).

In order to verify and prove the existence of crosstalk between defocus and astigmatism, I performed the same simulation used to obtain the results from Fig. 4.18 a), but plotted the retrieved values of astigmatism instead of defocus. The results are presented in Fig. 4.20. We can see that for values of defocus introduced between -250 nm rms and +250 nm rms, the retrieved value of astigmatism corresponds exactly to the amount of diversity introduced (200 nm rms), which means that there is no crosstalk between defocus and astigmatism within this range. This range represents exactly the linearity range obtained for the defocus ramps with GS, which means that crosstalk limits the linearity ranges of GS. For defocus values introduced with absolute values higher than 250 nm rms, we see that GS is retrieving more astigmatism than supposed since diversity is always a constant. When we apply high values of defocus, we are essentially contributing to the PSF enlargement (an increase of FWHM), and the stretching

effect verified in the PSF tends to be less evident. This is expected to induce possible crosstalk effects that are very evident in the case of GS. I should note that this crosstalk effect seems to be only happening between defocus and astigmatism since I plotted some other modes, and the retrieved values were always zero.

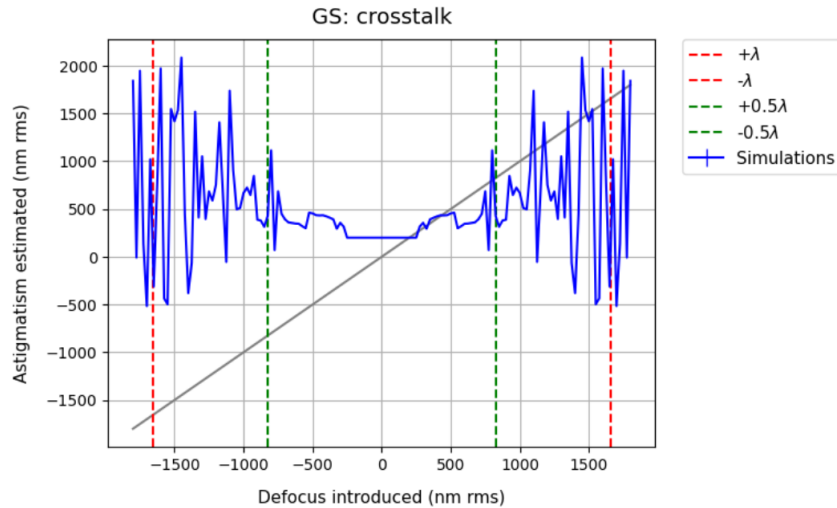


Figure 4.20: Estimated astigmatism values according to different defocus input values. Results obtained in the same conditions used for the results in Fig. 4.18 a). These results prove the existence of crosstalk effects between defocus and astigmatism.

It should be noted, however, that crosstalk effects are not exclusive to GS. In Fig. 4.21, I show the presence of crosstalk effects when estimating the first 5 modes instead of just TT and focus. We can see that the defocus linearity range is much more reduced now, approximately from -250 nm rms to +250 nm rms, similar to what was verified with GS. Once again, the crosstalk was only verified between defocus and the diversity mode (astigmatism 45°). A solution to avoid this kind of problem is to avoid the estimation of the mode corresponding to the diversity term. By estimating the first five modes, except for astigmatism (45°), we would obtain a defocus ramp equal to the one obtained in Fig. 4.18 b), where only TT and focus were estimated.

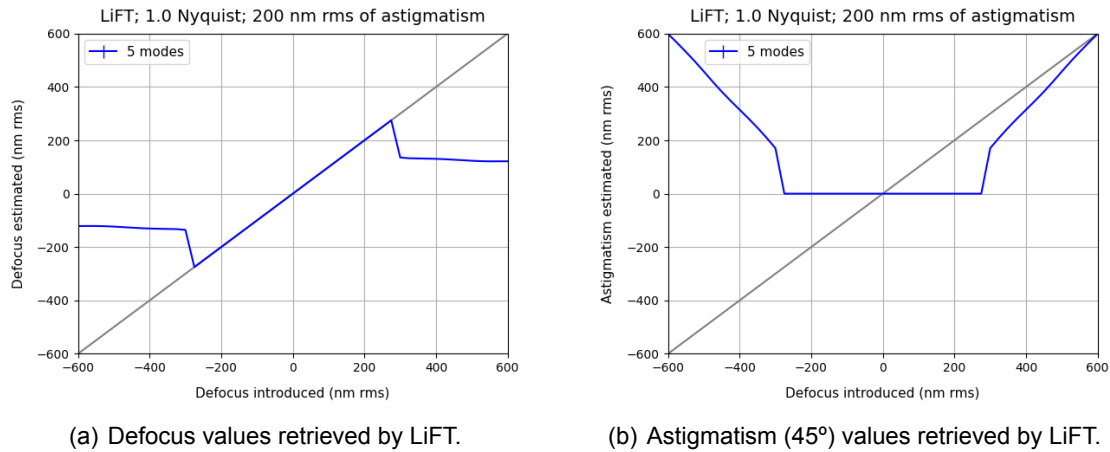
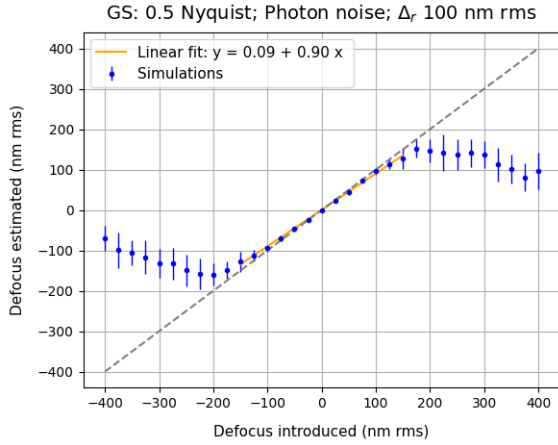


Figure 4.21: Crosstalk effects in LiFT when estimating 5 modes (including the term corresponding to the diversity).

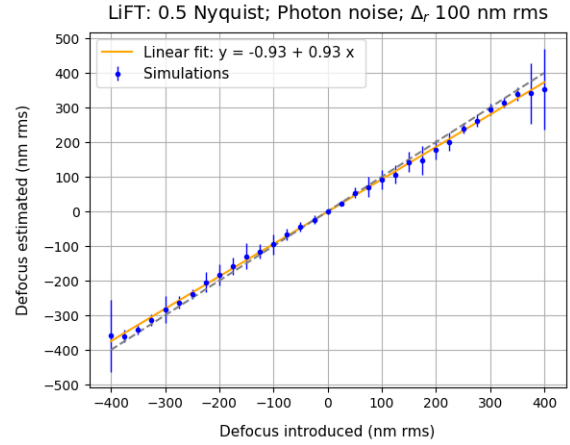
Now, I will showcase focus ramps obtained under realistic conditions, i.e., under sub-Nyquist sampling, photon noise, and the presence of high-order residuals. I made focus ramps from -400 nm rms to +400 nm rms, with step values of 25 nm rms. I used 25 random samples for each defocus value introduced and calculated the mean value obtained for each and the standard deviation. Results obtained for an amount of high-order residuals of 100 nm rms are presented in Fig. 4.22, while results for 200 nm rms are depicted in Fig. 4.23.

One of the results that stands out more prominently is the higher linearity ranges attained by LiFT. This was already verified in the previous tests without noise, so this was a predictable result. The GS results are limited once again by the effect of crosstalk. We see that in the edges of the linearity range, the defocus values estimated by GS tend to stagnate and decrease. This behavior is associated with crosstalk between astigmatism and defocus. In addition, adding even more astigmatism in the high-order residuals will highlight those effects even more. On the other hand, LiFT seems much more resistant to crosstalk effects, so higher linearity ranges can be achieved.

These results also show that high-order residuals tend to introduce a slope error in the focus ramps, which will be higher as more residuals are present. Consequently, slope calibrations must be applied to calibrate these focal plane wavefront sensors. In a real scenario, this is something that needs to be performed with all types of WFS, and FPWFS is not an exception.

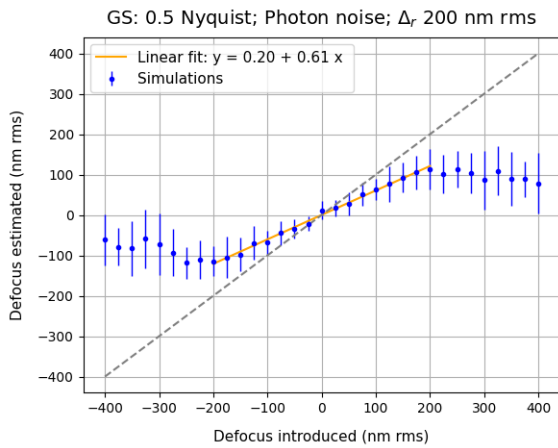


(a) GS: 200 maximum iterations allowed.

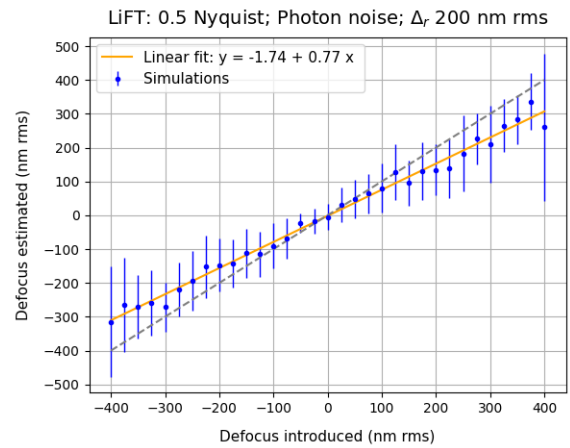


(b) LiFT: 20 maximum iterations allowed. Estimation of TT and focus.

Figure 4.22: Focus ramps for GS and LiFT for 0.5 Nyquist sampling with photon noise and high-order residuals. A natural guide star of magnitude 14 was used in band H. The values used for the error bars are the standard deviation values.



(a) GS: 200 maximum iterations allowed.

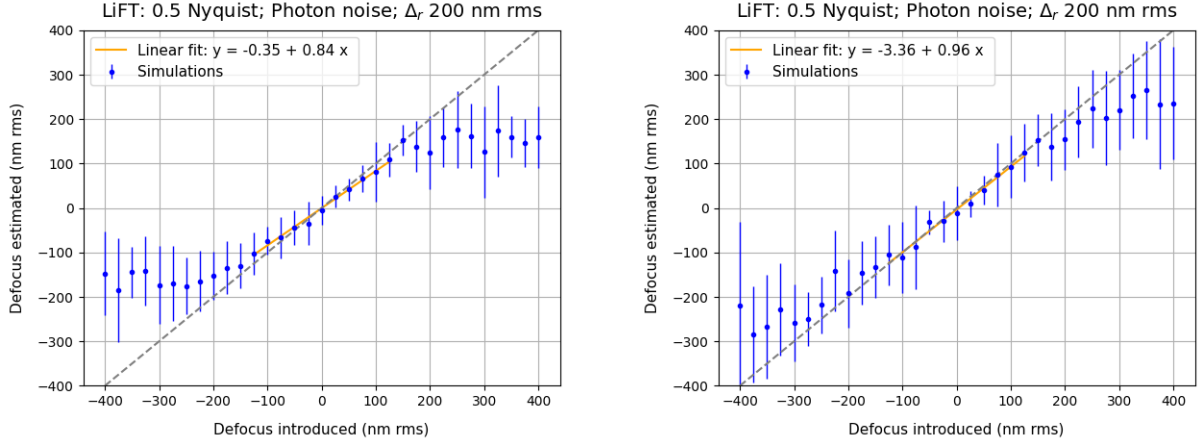


(b) LiFT: 20 maximum iterations allowed. Estimation of TT and focus.

Figure 4.23: Focus ramps for GS and LiFT for 0.5 Nyquist sampling with photon noise and high-order residuals. A natural guide star of magnitude 14 was used in band H. The values used for the error bars are the standard deviation values.

In the previous section, results have shown that estimating a few more modes other than just TT and focus helped the estimation of TT and focus with LiFT by compensating the presence of high-order modes. In Fig. 4.24, I have the focus ramps obtained when estimating five modes for the same conditions as the ones for Fig. 4.23. Results show improvements, especially between the range -100 to +100 nm rms, where the slope is closer to one. In Fig. 4.24 a), defocus values start to stagnate for absolute values over 200 nm rms, since the mode corresponding to the diversity (astigmatism 45°) is being estimated and crosstalk effects arise as already seen in this section. In Fig. 4.24 b), the mode corresponding to the diversity is not estimated, and higher

linearity ranges can be achieved. However, the standard deviations obtained are slightly higher.



(a) First five modes estimated (including astigmatism 45°).

(b) First five modes estimated except astigmatism 45°.

Figure 4.24: Focus ramps for LiFT for 0.5 Nyquist sampling with photon noise and 200 nm rms of high-order residuals. A natural guide star of magnitude 14 was used in band H. The values used for the error bars are the standard deviation values.

4.1.5 Close loop simulation: towards an on-sky implementation

In this section, numerical simulations will be performed for focus-tracking in an LGS-AO system scenario. The simulation results that will be presented refer to the specific case of the W. M. Keck Observatory at Mauna Kea, with whom I am collaborating toward a possible on-sky implementation of a new focus tracking tool for the Keck I LGS-AO system. It should be noted that this type of simulation can also be easily extrapolated to any other kind of LGS AO system on other telescopes, provided we change some input variables related to the telescope's characteristics.

The first step to simulate a real focus tracking scenario is to generate Na layer altitude variations as closely as possible to reality. As already introduced in the state of the art, the PSD for these mean altitude variations (Δh) is well-fitted by the following linear model [31]:

$$PSD(f) = \alpha f^\beta. \quad (4.6)$$

The values of $\alpha = 35 \text{ m}^2 \text{Hz}^{-1}$ and $\beta = -1.9$ were found from experimental data by Benoit et al. [22], and these are the values used in the simulations that follow. In Fig. 4.25, we have a random generation of such mean altitude variations for a 5-hour period.

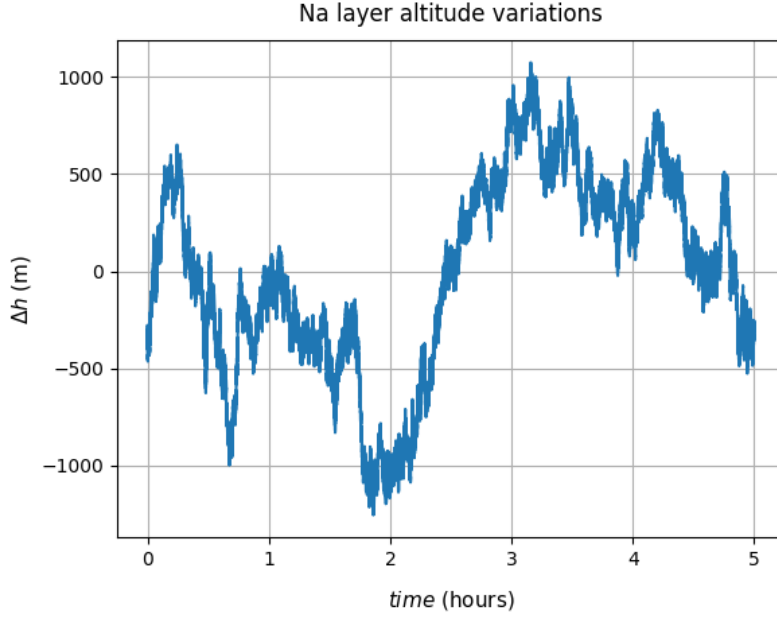


Figure 4.25: Random Sodium layer altitude variation taking into consideration the PSD given by Eq. 4.6 with $\alpha = 35 \text{ m}^2 \text{ Hz}^{-1}$ and $\beta = -1.9$.

These altitude variations (Δh) will result in a defocus error (σ_{Na}) given by [31]:

$$\sigma_{Na} = \frac{1}{16\sqrt{3}} \left(\frac{D}{h - h_0} \right)^2 \Delta h, \quad (4.7)$$

where D is the telescope's diameter, h is the mean height of the Na layer (90 km), and h_0 is the telescope's altitude. For the Keck case, we have $D = 11.732 \text{ m}$ and $h_0 = 4145 \text{ m}$. These defocus errors usually vary very slowly, and the correction is made at very low rates (every $\sim 20\text{-}30 \text{ s}$). However, since this focus error increases with D^2 , the focus tracking needs to be done faster for telescopes with bigger diameters. This issue is particularly important for the next generation of giant telescopes with diameters as big as 30-40 m.

One of the important operational aspects for focus tracking in LGS-AO systems is the correction cadency necessary to keep the system in focus as best as possible. In an attempt to quantify the correction cadency that may be needed for focus tracking in Keck, I studied the defocus error as a function of the correction cadency used. I generated a 5-hour time series of the Na layer altitude variations according to the PSD given by Eq. 4.6 and then generated the corresponding slow focus series according to Eq. 4.7. From this slow focus series, I studied the rms defocus error (σ_ϵ) as a function of the correction cadency period (Δt_{CC}). For example, for $\Delta t_{CC} = 1.0 \text{ s}$, that means the slow focus is averaged every 1 s, and a lag/delay of 1 s is applied. This is exactly what happens in a real scenario: we integrate for $\Delta t_{CC} \text{ s}$ and then apply a focus correction during the next $\Delta t_{CC} \text{ s}$. In this simulation, we assume that the focus averaging is perfect.

The results obtained for the rms defocus error (σ_ϵ) as a function of Δt_{CC} are presented in Fig. 4.26. According to the results, σ_ϵ grows with $\sqrt{\Delta t_{CC}}$. As expected, the error would be lower for lower correction cadence periods since we can keep up the slow focus variations better. However, in a real scenario, these focus corrections are expected to be done with very faint stars, and we may have to integrate longer to achieve good corrections. The ideal correction cadency needs to be chosen from a trade-off between SNR and delay.

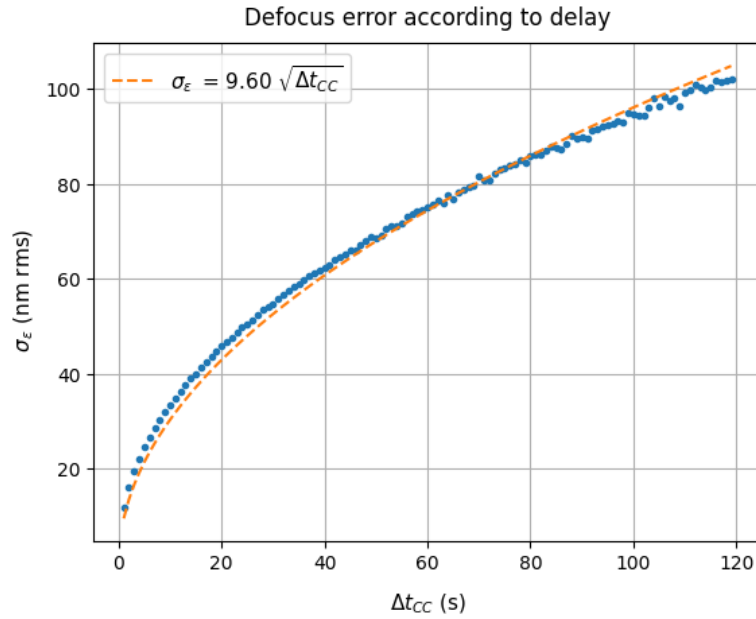


Figure 4.26: Simulation of expected defocus error (σ_ϵ) according to correction cadency used for focus tracking in the Keck telescope. The curve fit was performed using the function `optimize.curve_fit` from `scipy` [56]

There is, however, one way to perform fast focus corrections from PSFs with high SNR, even when dealing with faint stars. The solution is to perform a rolling average as depicted in Fig. 4.27. The idea is to integrate the PSF for a longer period than Δt_{CC} to provide more SNR and, at the same time, do corrections at high rates. In a rolling average scheme, Δt_{LE} needs to be equal to or higher than Δt_{CC} . In the case they are equal, no rolling average is, in fact, being done since the focus correction period is equal to the exposure time. Of course, this solution has its own costs as well since integrating for a longer period means that the delay effect will introduce an error.

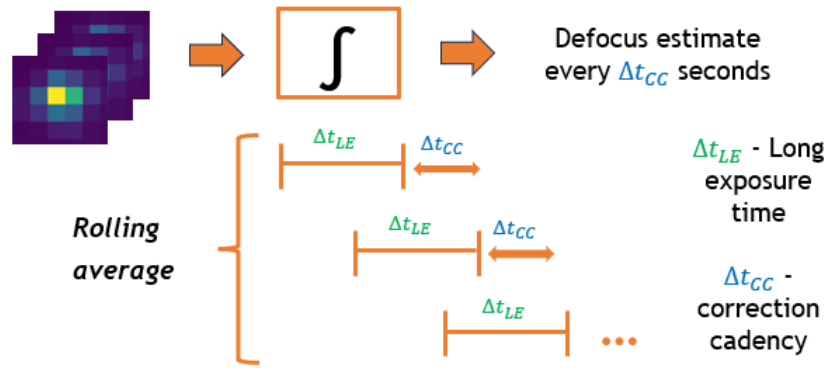


Figure 4.27: Illustration of rolling average for focus tracking. Corrections are made every Δt_{CC} seconds, where the defocus value is extracted from a PSF with an exposure time of Δt_{LE} s.

Now, I will present some results concerning the SNR achieved for different exposure times with faint stars in order to show what kind of exposure times are expected to allow good SNR. The PSF simulations were done considering a detector with an angular pixel size of 50 mas, which is the same as the one for the TRICK sensor (Near infra-red tip/tilt sensor) at the Keck I LGS-AO system. The idea at Keck is precisely to use TRICK to perform focus tracking other than just for measuring TT. Despite offering quite challenging sampling conditions, TRICK has a "natural" astigmatic PSF due to an optical error in the system, which can be efficiently used to solve for the sign ambiguity of the defocus mode, as already seen before. The astigmatism present is of type 0° and it has an rms value of approximately 200 nm. In Fig. 4.28, I show three different simulated TRICK PSFs that were obtained considering the presence of 200 nm rms of defocus and astigmatism 0° for an NGS of magnitude 16. Two types of noise were considered, namely photon noise and readout noise ($4e^-/pixel/s$). We can see that for an exposure of 1 ms, the PSF is highly consumed by noise, but for a one-second exposure, the SNR is already considerably high, and the difference with the noiseless PSF is almost non-existent. We can conclude that a 1-second exposure should be sufficient to provide sufficient SNR with faint stars up to $m = 16$, at least in the simulation environment.

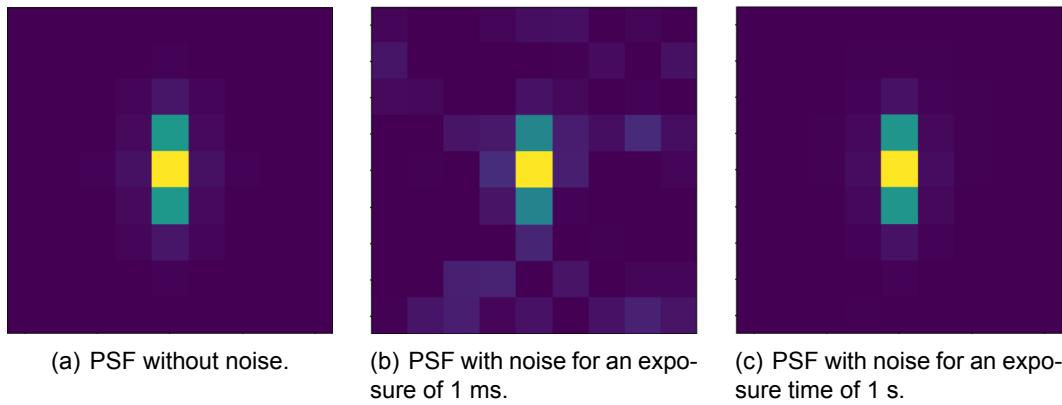


Figure 4.28: Simulation of PSF in OOPAO in a detector with 50 mas of angular pixel size for an NGS in H-band with magnitude 16 for different exposure times. The presence of photon noise and readout noise ($4 e^-/pixel/s$) are considered. The PSFs presented were obtained considering the presence of 200 nm rms of astigmatism (0°) and 200 nm rms of defocus.

Before showing the results for the focus tracking both for GS and LiFT, I should briefly describe the conditions of the close-loop study, which was done in OOPAO. All the different parameters used were chosen so as to simulate as closely as possible a real on-sky operation for focus tracking in keck. In this simulation, all components of the AO system were simulated, namely the different WFS, the DM, and even the atmosphere. The atmosphere was simulated considering five distinct turbulent layers and a Fried parameter (r_0) of 15 cm. This atmosphere was then used to simulate different aberrated screens. The light with the aberrated phase screens was then propagated through a telescope with a diameter of 11.732 m, assuming a round pupil. A Shack-Hartmann WFS with 20×20 sub-apertures was used to simulate the AO loop, where each spot was Nyquist sampled. This loop operated at a frequency of 1 kHz in a closed loop, which was the high-order loop.

A low-order loop was then created to simulate a focus tracking operation, where the type of WFS used was the FPWFS, namely LiFT/GS. In this loop, images were simulated in a detector with an angular pixel size of 50 mas, just as TRICK, and from those images, the focus values were extracted by GS/LiFT. These TRICK images were generated considering the presence of high-order residuals from the high-order loop (~ 150 nm rms), the presence of static NCPA (~ 80 nm rms simulated using a $(n + 1)^{-2}$ power law), the presence of 200 nm rms of astigmatism (0°), and of course the slow focus values. In addition, the presence of photon noise and readout noise ($4 e^-/pixel/s$) was also considered. In this simulation, we consider that the slow focus is not seen by the Shack-Hartmann and only appears in the low-order loop. In an areal scenario, the slow focus is usually filtered out since it varies much slower than the atmospheric focus, so this assumption is not entirely unrealistic.

In Fig. 4.29, I present the results for the closed-loop simulation considering an NGS in H-

band with magnitude 14 for the low-order loop. As expected, the defocus error (σ_ϵ) increases with Δt_{CC} since the lag effect becomes more and more evident. For illustration purposes, in Fig. 4.30, I show an example where the corrections are being made every 2 seconds and where the lag effect can be easily perceived. Since the SNR is already high for a 1-second exposure, increasing Δt_{CC} won't have any benefits as expected. The only exception was for GS, in which the error for a 1-second correction period was higher than for 2 and 5-second correction periods. This result was due to the existence of some divergent values, whose existence is somehow inexplicable for now. These results also show that LiFT has a better performance than GS, especially for larger Δt_{CC} values. In this figure, the theoretical limit of focus error for Keck is also plotted. As expected, GS and LiFT errors increase faster with Δt_{CC} since the theoretical limit is obtained assuming a perfect defocus averaging.

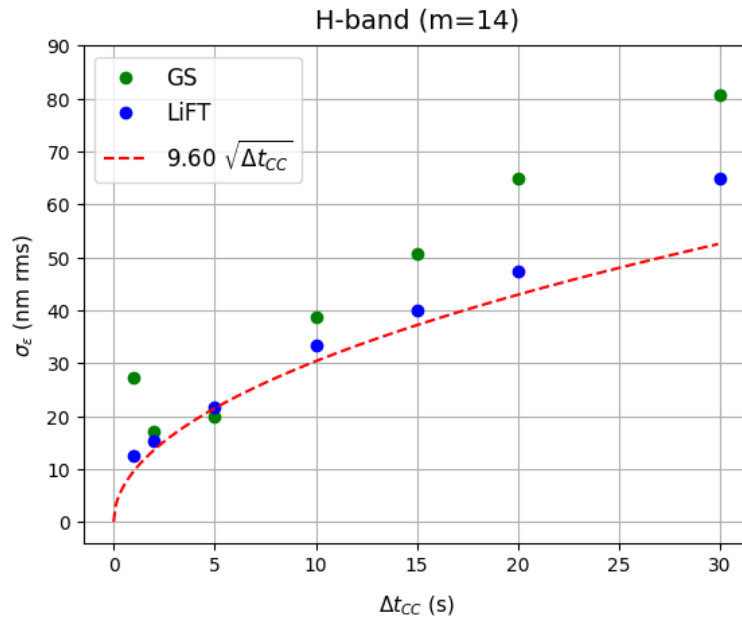


Figure 4.29: Defocus error as a function of the correction cadency period (Δt_{CC}) for the closed-loop simulation. Results for GS and LiFT are plotted along with the theoretical limit previously estimated for Keck.

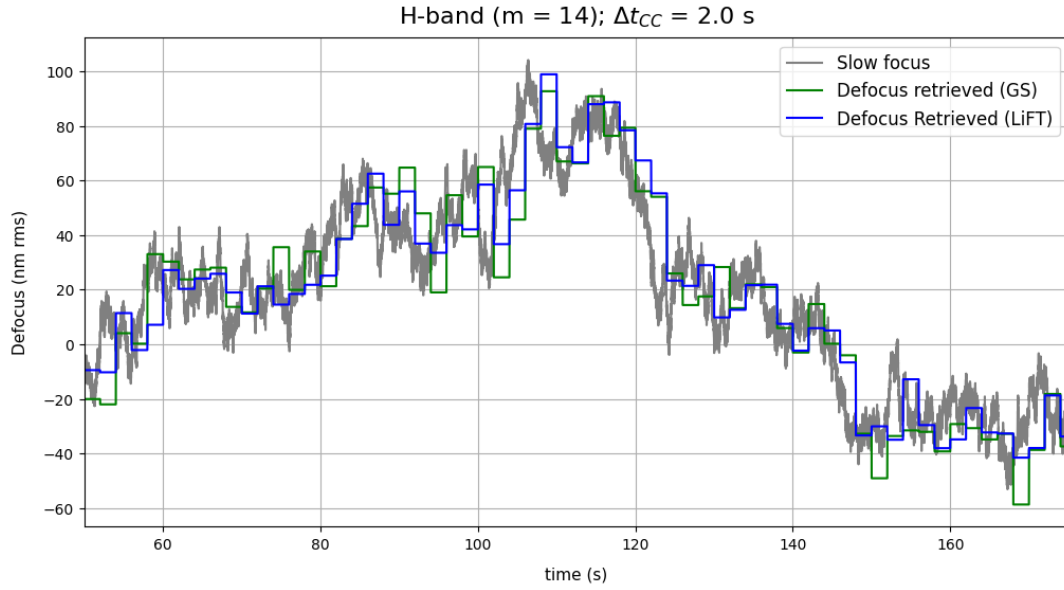


Figure 4.30: Focus tracking example with LiFT and GS with $\Delta t_{CC} = 2.0$ s.

Now, I will also show some results where the rolling average is performed. The results for $\Delta t_{CC} = 2.0$ s and $\Delta t_{CC} = 5.0$ s are presented in Fig. 4.31. For both cases, increasing the exposure time does not prove efficient since the error always increases. This could already be anticipated since for higher values of Δt_{LE} , the lag effect will be more considerable. As an illustrative example, in Fig. 4.32, I have an example with $\Delta t_{CC} = 2.0$ s and $\Delta t_{LE} = 10.0$ s. By comparing with Fig. 4.30, where no rolling average was performed, we can see that the lag effect is indeed more evident with the rolling average. Performing rolling average is only beneficial when the SNR is low for a certain Δt_{CC} .

However, some extra results should be pointed out. For example, doing a rolling average with $\Delta t_{CC} = 2.0$ s and $\Delta t_{LE} = 10$ s allows achieving almost the same performance as the one achieved with $\Delta t_{CC} = 5.0$ s without rolling average. This is an interesting result as we can have images with higher SNR and, at the same time, perform focus corrections faster. From this perspective, performing a rolling average can prove quite beneficial.

From an operational point of view, these correction cadencies (every 2-5 s) are difficult to achieve since focus corrections are made by acting directly in the lasers. Consequently, mechanical restrictions limit the rate at which focus corrections are made. For example, at Keck, focus corrections are currently made every 30 s. With this correction rate, performing a rolling average would not be beneficial. Still, in the near future, with the next giant telescope generation, focus corrections will need to be performed much faster, and faster ways to track focus will need to be encountered. Assuming those techniques will be found, the rolling average concept may prove an effective way to improve the focus tracking performance.

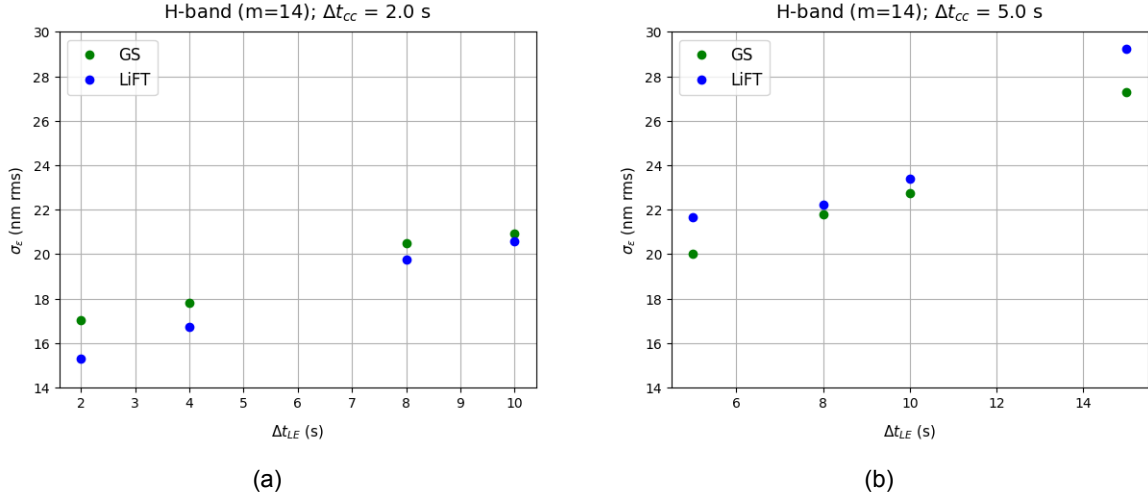


Figure 4.31: Results with rolling average for $\Delta t_{CC} = 2.0$ s and $\Delta t_{CC} = 5.0$ s.

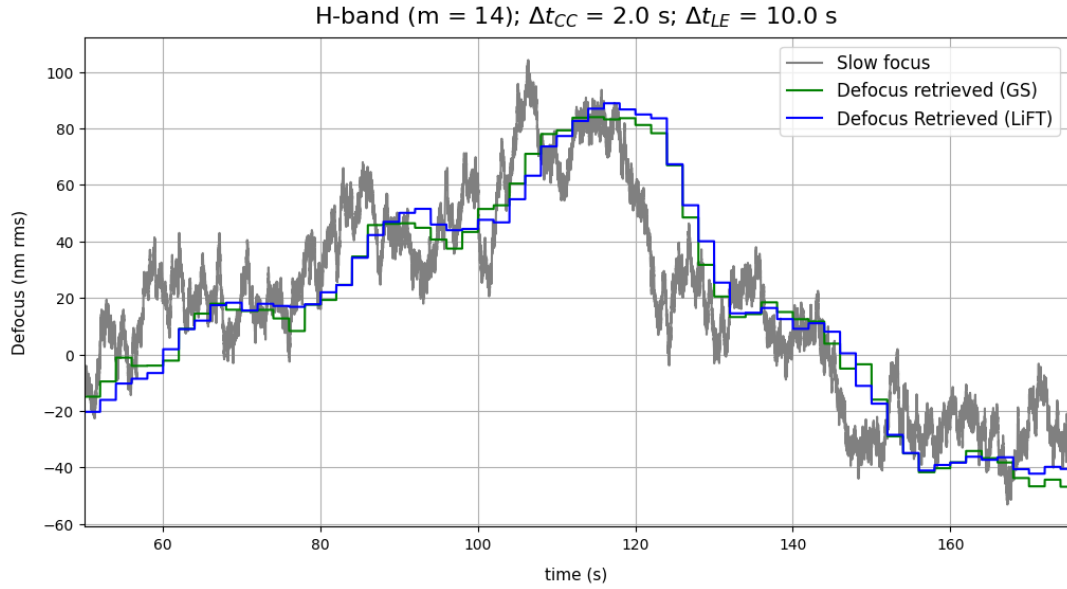


Figure 4.32: Focus tracking example with LiFT and GS with $\Delta t_{CC} = 2.0$ s and $\Delta t_{LE} = 10.0$ s. Rolling average was performed.

4.2 Tests with real data

In this section, tests with bench and on-sky data from the Keck I telescope will be presented and discussed. In addition to GS and LiFT, the FWHM fit was also studied in this part.

4.2.1 Results with bench data

In the previous section, several simulations were performed to assess the capabilities of GS and LiFT to measure slow focus since this is the main current operational issue in LGS-AO systems, such as the one in the Keck I telescope in Hawaii. The Keck I LGS-AO system uses a

low bandwidth wavefront sensor (LBWFS) to measure slow focus due to Na layer altitude variations. However, its performance is unsatisfactory, and FPWFS is being studied as a solution. In addition to GS and LiFT, the FWHM fit will also be studied here.

Part of my work was oriented towards this specific telescope, so tests with bench and on-sky data from Keck were performed to thoroughly test the capabilities of GS, LiFT, and the FWHM fit in measuring defocus. This section will be dedicated to presenting all the results with bench data, while in the next section, I will present results with on-sky data. All data was collected in the NIR (band K and H).

The NGS-AO (low-order) loop in Keck I uses a specific detector known as TRICK (Tip-tilt Removal with Ir Compensation at Keck) to correct TT and defocus. The idea is then to use the images provided by TRICK to estimate defocus errors ideally faster and more accurately than LBWFS using FPWFS techniques. In Tab. 4.1, I present the main features of TRICK. I should refer to the fact that the sampling time of TRICK is not always constant, and it is constantly updated according to the NGS magnitude. However, the bench data was collected with this specific sampling time. A round pupil with a diameter of 11.732 m was considered in the tests performed with bench data.

TRICK	
Pixel size (μm)	18
Angular pixel size (mas)	50
Sampling time (ms)	5

Table 4.1: Trick sensor characteristics.

As shown in Tab. 4.1, the pixel angular size of Trick is 50 mas, which is a pretty challenging scenario in terms of sampling both for band K and H, but most challenging for band H. In Table 4.2, I show the angular pixel sizes that correspond to 1.0, 0.5, and 0.25 Nyquist sampling both for H and K bands and we can see that to attain the same kind of sampling with band H, we would always need pixels with lower angular size. Since the pixel angular size is always the same, the PSFs from band H will always be worse sampled when compared to the ones from band K.

	Band H	Band K
1.0 Nyquist	14.5 mas	19.2 mas
0.5 Nyquist	29.1 mas	38.3 mas
0.25 Nyquist	58.2 mas	76.6 mas

Table 4.2: Angular pixel sizes in mas corresponding to 1.0, 0.5, and 0.25 Nyquist sampling both for band H ($\lambda = 1.654 \mu\text{m}$) and band K ($\lambda = 2.179 \mu\text{m}$). The telescope's diameter used for these calculi was 11.732 m (the same as Keck).

Before showing and discussing the results obtained with bench data from Keck, I will describe the type of data provided and the preprocessing steps that were used. The bench data provided consists of 6 files of data, where three concern band H data and the other three band K data. We have data files collected for different star magnitudes within each set corresponding to a specific band (H or K). Each file has images corresponding to inserted focus values from -300 nm rms to +300 nm rms with steps of 20 nm rms, where for each defocus value, we have 30 different images associated.

The phase diversity necessary to determine the sign of defocus is automatically provided by Keck, which has an astigmatism aberration in the NIR TT WFS path, i.e., on TRICK. The phase diversity consists of 200 nm rms of astigmatism (0°). This "natural" astigmatism can be efficiently used by FPWFS techniques to measure defocus. In Fig. 4.33, we can see the presence of this astigmatism aberration since the PSF stretches in perpendicular directions for defocus values of opposite signs.

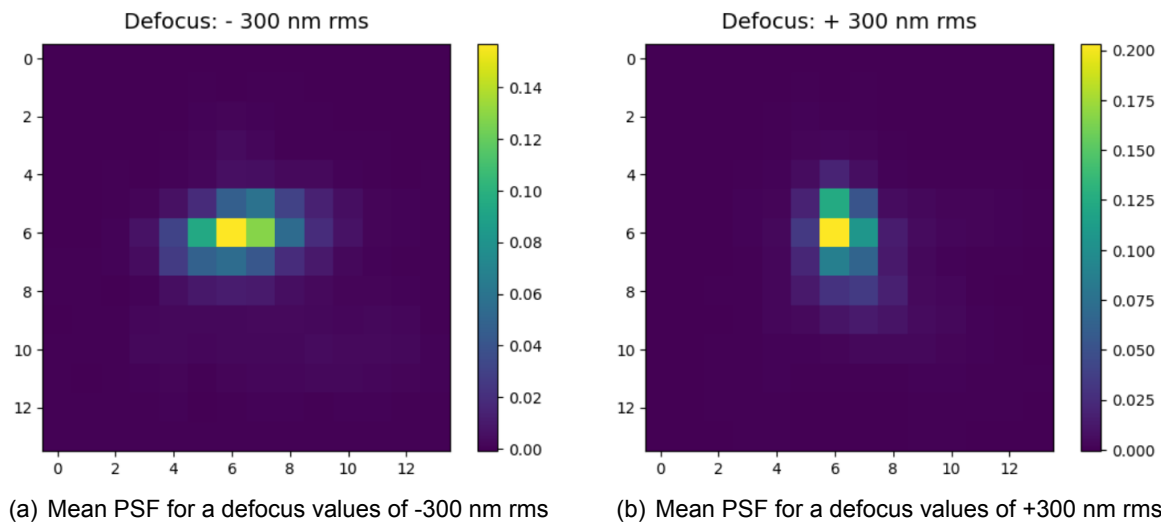


Figure 4.33: Illustration of TRICK PSF for a specific data file (band K and $m = 12$) with defocus values of opposite signs. The PSF stretches in perpendicular directions due to the 200 nm rms of astigmatism in the TRICK path.

Following the previous discussion, I should now introduce an algorithm developed by Arseniy Kuznetsov [55] that was also used to determine defocus values from the provided bench data, which I call FWHM fit. This method has already been introduced in the Methods section. As we have seen before, the PSF has a predictable behavior according to the value of defocus introduced, and fitting algorithms can be used to determine its values. This specific algorithm makes a Gaussian profile fit to the PSF to determine σ_x and σ_y . Then, these values are used to determine the value of defocus introduced employing the following empirical expressions:

$$defocus(K) = \left(\frac{\sigma_y}{\sigma_x} - 0.9 \right) \times 600 \quad (4.8)$$

$$defocus(H) = \left(\frac{\sigma_y}{\sigma_x} - 0.9 \right) \times 825 \quad (4.9)$$

Different coefficients are used according to the band. The coefficients for each equation were also empirically determined according to the bench data.

I will now showcase the preprocessing steps used with the bench data, described in Fig. 4.34. The preprocessing steps include background removal, image normalization, image cropping, and diversity calibration. The amount of cropping required depends on the SNR of the images. For these data files, cropping images to obtain a central region of 14*14 pixels worked for all expected for one file where the SNR was very low and more cropping was necessary. We need to consider fewer central pixels when we have low SNR to achieve good results. On the other hand, when the SNR is high, we should consider more pixels. Another crucial preprocessing step is diversity calibration. Ideally, the images in focus should only contain the expected 200 nm rms of astigmatism with no other aberrations. Still, more aberrations will come from non-common path aberrations (NCPA) or high-order residuals (for an on-sky scenario). High-order residuals cannot be calibrated since they will continuously change with time due to atmospheric turbulence. On the other hand, NCPA vary very slowly, so they can be more easily compensated.

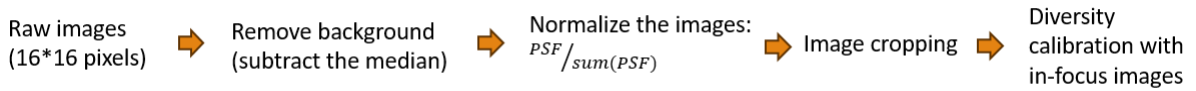


Figure 4.34: Preprocessing steps used with bench data (applied to all images).

In Fig. 4.35, I have an example of a mean image corresponding to the in-focus images of a particular data file. These images are used to calculate a new diversity that accounts for the presence of NCPA and helps improve defocus estimation. Basically, diversity calibration allows us to redefine our initial working point. This diversity calibration is particularly important to solve possible bias problems in the focus ramps. If a certain amount of defocus is detected with the in-focus images, where the defocus should be zero, then that defocus value should be subtracted from all results.

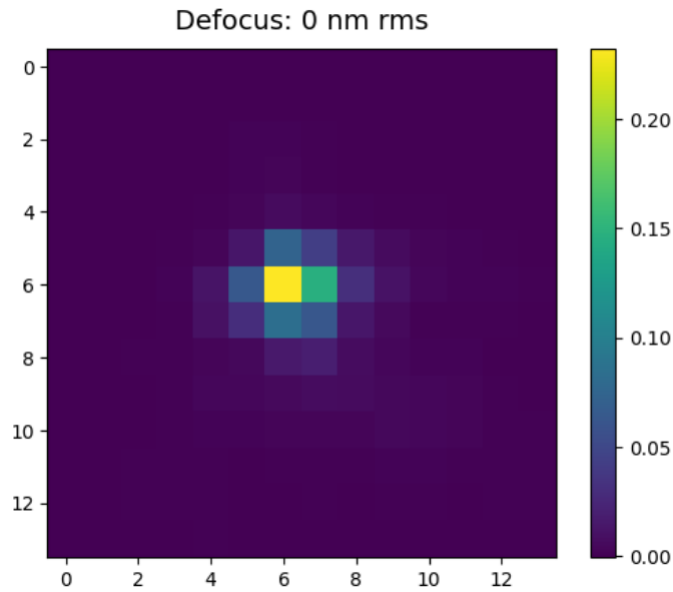


Figure 4.35: Example of mean in-focus image for a specific data file (band K and $m=12$). In-focus images are used to perform diversity calibration.

I will now showcase all the results obtained for all data files provided. In Fig. 4.37, 4.38, and 4.39, I have all the results relating to band K and in Fig. 4.41, 4.42, and 4.43 all the results referring to band H. For every defocus value introduced, all 30 images were used to estimate 30 defocus values. The median defocus values obtained are plotted with the corresponding 95-percentile interval. I should refer to the fact that LiFT results were obtained by estimating the first three modes for band H and the first four modes for band K. In addition, an initial guess for the defocus value coming from the FWHM fit was used for LiFT, whereas no initial defocus guess was made for GS. However, I should note that LiFT results would be similar even if no initial assumption was made.

In Fig. 4.36, I show a comparative example of GS results without calibration and with calibration. These slope adjustments were already expected since my current GS algorithm is unprepared to work under sub-Nyquist sampling. GS was applied as the PSFs were Nyquist sampled. However, an excellent linear behavior was attained with the bench data, which is a promising result. In an attempt to try solving this slope issue, I tested using better-sampled PSFs (1.0 Nyquist) by performing upsampling of the TRICK PSFs through the usage of interpolation algorithms. However, the obtained results were far from satisfactory, and this didn't prove an efficient solution. From my point of view, upsampling techniques end by introducing new aberrations to the PSFs that cannot be calibrated, and as a defocus estimation is highly affected.

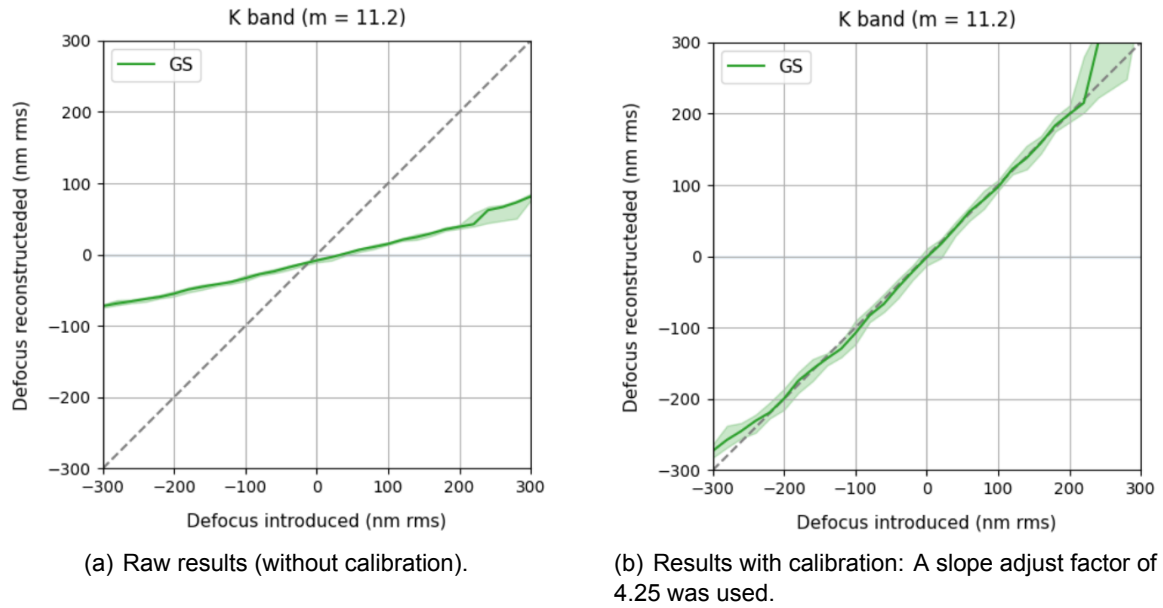


Figure 4.36: Illustration of the calibration process for GS. Results obtained with Keck bench data for band K with 11.2 of magnitude. The shaded area represents the 95th percentile interval, with the solid line indicating the median value.

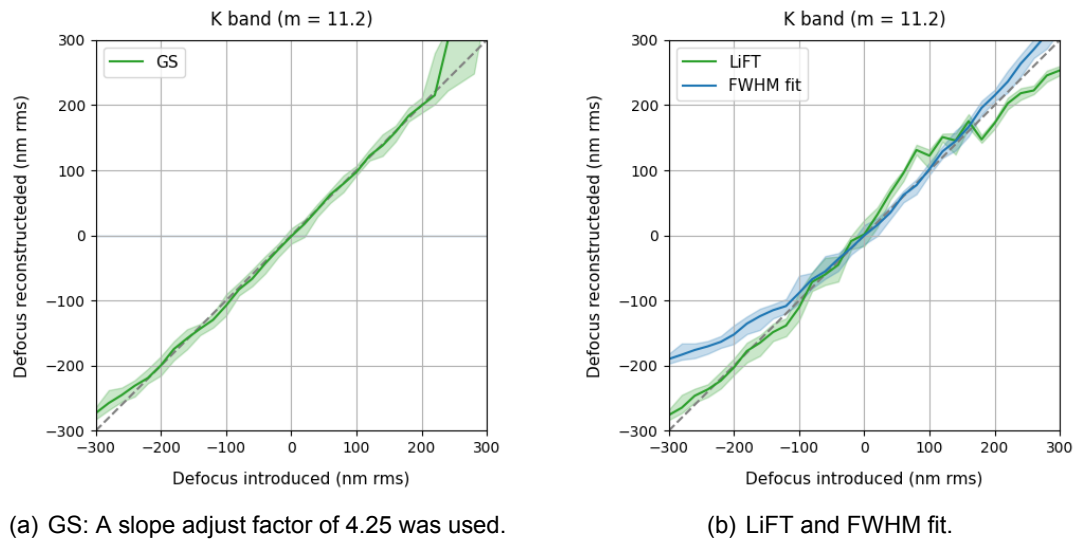


Figure 4.37: Results obtained with Keck bench data for band K with 11.2 of magnitude. The shaded area represents the 95th percentile interval, with the solid line indicating the median value.

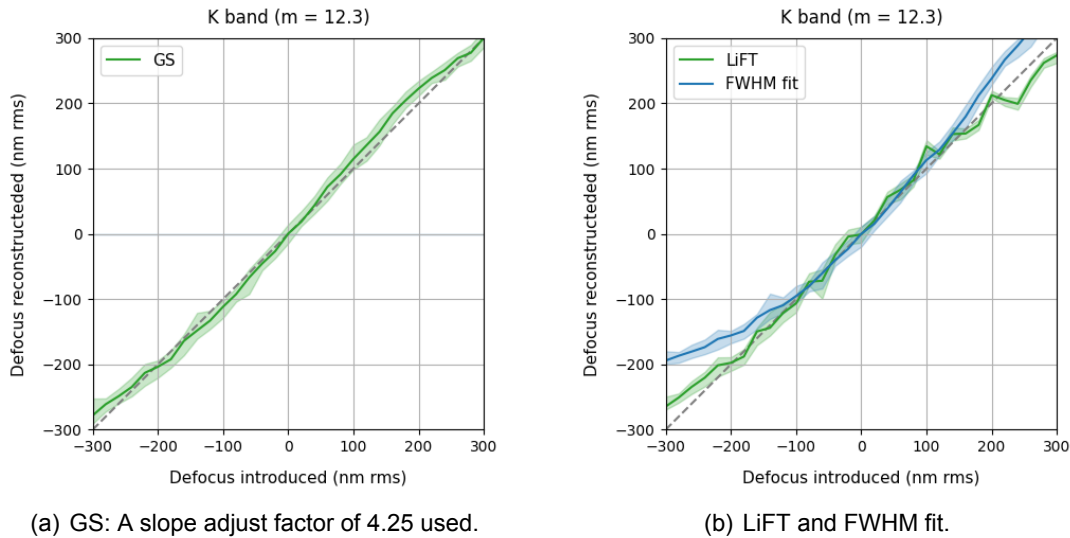


Figure 4.38: Results obtained with Keck bench data for band K with 12.3 of magnitude. The shaded area represents the 95th percentile interval, with the solid line indicating the median value.

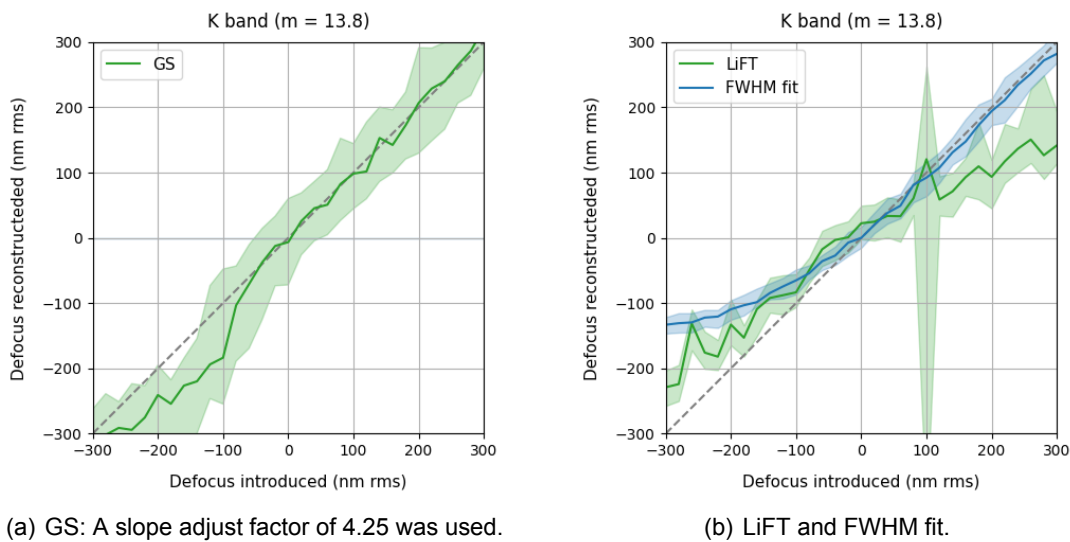


Figure 4.39: Results obtained with Keck bench data for band K with 13.8 of magnitude. The shaded area represents the 95th percentile interval, with the solid line indicating the median value.

By analyzing the results for band K, it is clear that the ones presented in Fig. 4.39 are worse than the others. The main reason for this result is that this specific file contained images highly consumed by noise, as we can see in Fig. 4.40, where an image example of this file is shown. Under such low SNR, LiFT, and GS are not capable of achieving as good results as the ones obtained with high SNR. One way to counter this SNR problem is to perform more cropping of the images, i.e., considering only fewer central pixels. The results presented in Fig. 4.39 were obtained by considering only a central region of 8*8 pixels of the TRICK images because otherwise, the results would have been even worse. These results show that low SNR affects

all algorithms' performance, and image cropping is insufficient to mitigate its harmful effects completely.

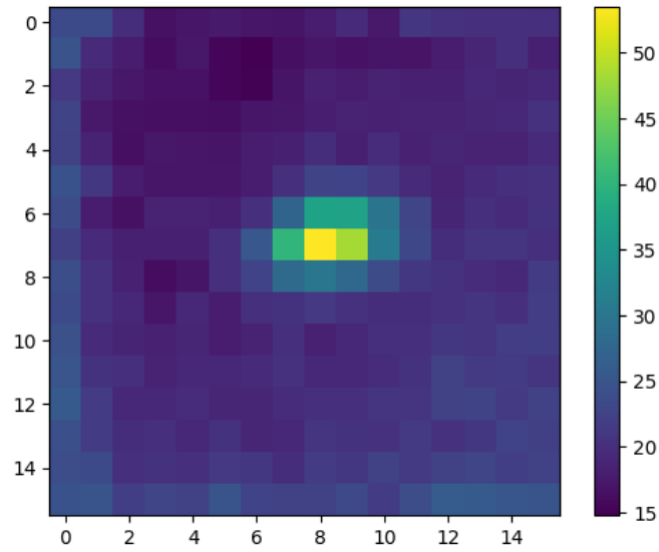


Figure 4.40: Example of raw image highly consumed by noise for the file corresponding to band K and $m=13.8$.

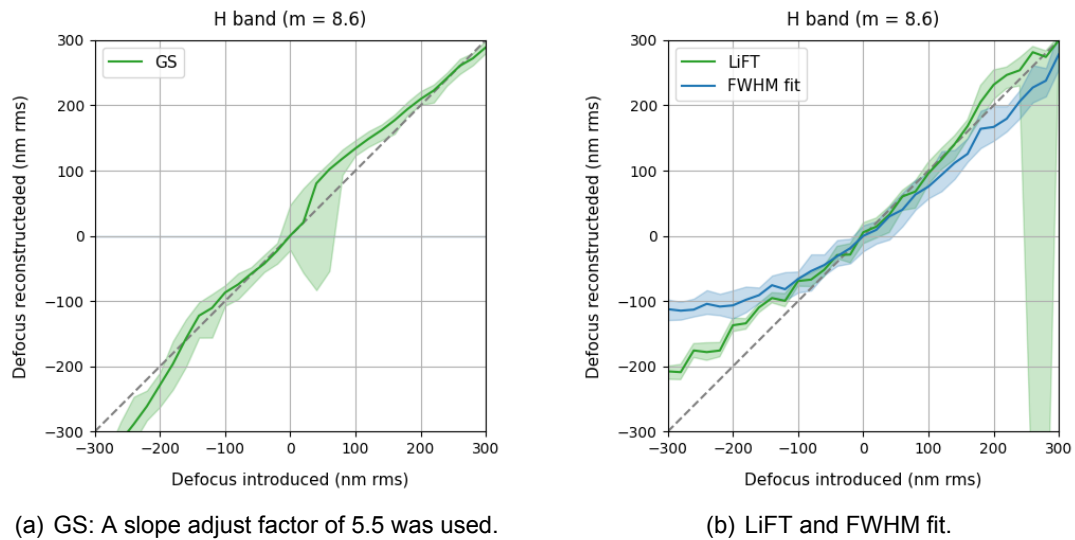


Figure 4.41: Results obtained with Keck bench data for band H with 8.6 of magnitude. The shaded area represents the 95th percentile interval, with the solid line indicating the median value.

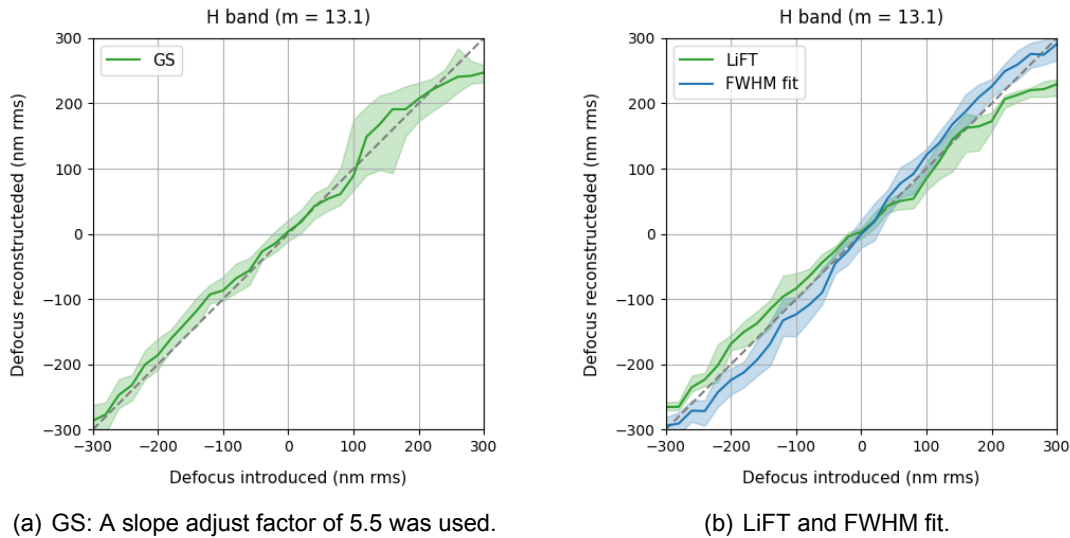


Figure 4.42: Results obtained with Keck bench data for band H with 13.1 of magnitude. The shaded area represents the 95th percentile interval, with the solid line indicating the median value.

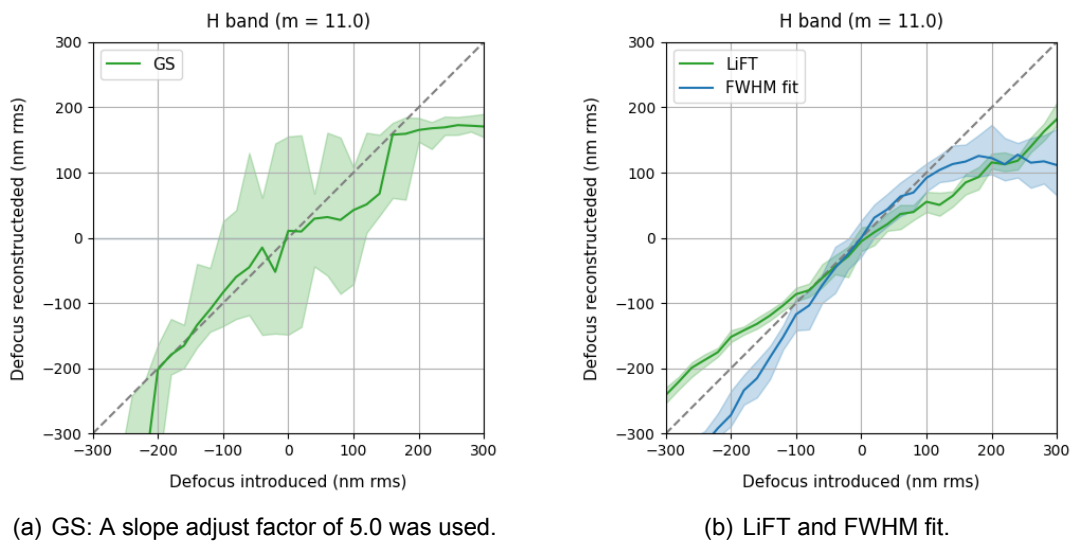


Figure 4.43: Results obtained with Keck bench data for band H with 11.0 of magnitude. The shaded area represents the 95th percentile interval, with the solid line indicating the median value.

In Fig. 4.41, 4.42, and 4.43, we have all the results obtained for the band H files. The results shown in Fig. 4.43 are much worse than the others. From an analysis of the images of the corresponding file, the SNR was high, so this is not the reason for such results. In this way, problems with data collection of this data file might be the reason for such results, but this was not verified in the images. The reason for these results is somehow inexplicable at the moment. Once again, GS results needed a slope calibration, but this time, the slope adjust factor was higher than that of band K. For band K, a slope adjust factor of 4.25 proved to efficiently calibrate all the defocus curves, whereas for band H the slope adjust factor was 5.5 (except for one file).

This was already expected since the PSF sampling in band H is lower than for band K.

All these results obtained with bench data allow us to outline some preliminary conclusions. GS showed excellent linear behavior in the defocus estimation even though the current algorithm is not fully prepared to work with undersampled images such as the ones from TRICK. However, GS requires slope calibration. The good news is that the optimal slope adjust factor proved to be approximately constant (4.25 for band K and 5.5 for band H). Of course, more tests would need to be performed to guarantee that these slope adjustments are always constant.

On the other side, LiFT showed good linear behavior without needing slope calibration, although a slight slope adjustment could have been applied in some cases. This is the main advantage of LiFT regarding GS. Another important aspect is the computational costs, where LiFT has a slight advantage in relation to GS. When using highly undersampled images such as the ones from TRICK, GS can only converge for the first few iterations (10-20), so the convergence to the final solution is fast. I should note that the computational costs of GS and LiFT could be even further reduced if we reduce the pupil sampling, for example, by considering 32*32 pixels instead of 64*64 pixels. I performed some tests, and similar results were obtained, whether using a resolution of 64 or 32 pixels for the pupil. Of course, pupil sampling cannot be reduced indefinitely since the results will start declining. This can be extremely useful from an operational point of view since we can decrease the algorithms' computational burden without additional hardware costs.

Regarding the FWHM fit algorithm, the results for the bench data were quite impressive. It is a simple solution and in terms of computational costs, it is the fastest. However, I should emphasize that this algorithm uses empirical expressions whose coefficient values were chosen according to this specific data set [55]. For an on-sky scenario, where the atmospheric residuals on TRICK vary constantly at a very high pace, these coefficients would need to be updated at that rate, which is unrealistic. This is a significant problem related to fitting algorithms for measuring defocus and why applying such an algorithm on-sky may not be easy.

Another important conclusion from these results is that images cannot be highly consumed by noise, i.e., we need images collected with good SNR; otherwise, performance will be highly affected. In case low SNR images are given, one way to mitigate noise effects is to crop the images and consider fewer central pixels, although this is not a perfect solution to the problem.

The following section will showcase results with Keck's on-sky data from the TRICK sensor. The on-sky conditions are much more challenging when compared to the bench, where all devices are more easily controlled, and there are no atmospheric residuals, which is a crucial aspect to be taken into account from an operational point of view.

4.2.2 Results with on-sky data

The next natural step in accessing LiFT, GS, and FWHM fit capabilities for measuring slow focus using TRICK is to test these algorithms with actual on-sky data from TRICK. This way, all real operational conditions, such as the detector's noise, NCPA, and high-order residuals, will be considered.

This section will present the results obtained with on-sky data from the TRICK sensor collected in 2020. All data refers to band K, composed of two data files; for the first one, we have 8*8 pixels images, and the second one has images with 4*4 pixels. The first file contains images with a larger FoV, whereas the second has lower FoV images. For each file, we have images corresponding to inserted defocus values from -250 nm rms to +250 nm rms with steps of 25 nm rms, where for each defocus value, we have 30 different images associated.

The preprocessing steps followed to analyze on-sky data are the same ones used with the bench data (Fig. 4.34), except for the image cropping, which was not necessary as the cropping of the images provided is already very significant.

First, I will present the results with the data files corresponding to the larger FoV images (8*8 pixels). In Fig. 4.44, I showcase the mean images corresponding to different values of inserted defocus, where the stretching of the PSF in perpendicular directions according to the sign of the defocus introduced is once more verified as expected. From the analysis of these images, it seems that a defocus offset is present since the PSF with -250 nm rms is much more stretched concerning the PSF corresponding to +250 nm rms. Bias offsets for the defocus values are expected to appear, not due to problems related to the algorithms but because a real and physical defocus offset seems to be present. This "natural" defocus offset is probably due to calibration errors when taking the data, assuming this bias offset was not intended.

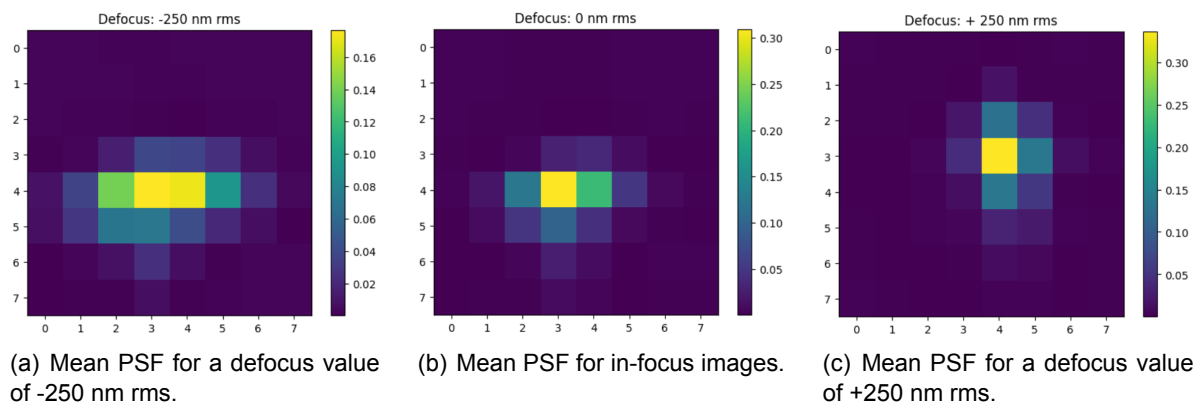


Figure 4.44: Mean images corresponding to defocus offsets of -250, 0, and +250 nm rms. These images correspond to the data file with 8*8 pixels images.

In Fig. 4.45, I present the on-sky results using the FWHM fit algorithm already explained in the previous section. In Fig. 4.45 a), I have the results obtained by using the same empirical formula used for the bench data, i.e., Eq. 4.8. We can see that the results are not as good as those obtained with bench data, which I already anticipated in the previous section. This fitting algorithm uses empirical formulas whose best coefficients vary according to the conditions and require constant calibration. In this case, the results show a high non-linear behavior which means that the choice of new coefficients is not enough to fully improve the results as shown in Fig. 4.45 b) where an empirical optimal coefficient was used to try improving the results. These results show that the FWHM fit algorithm does not seem prospective to be deployed on-sky due to its empirical nature and the consequent need for calibration for varying conditions, such as different NCPA or high-order residuals.

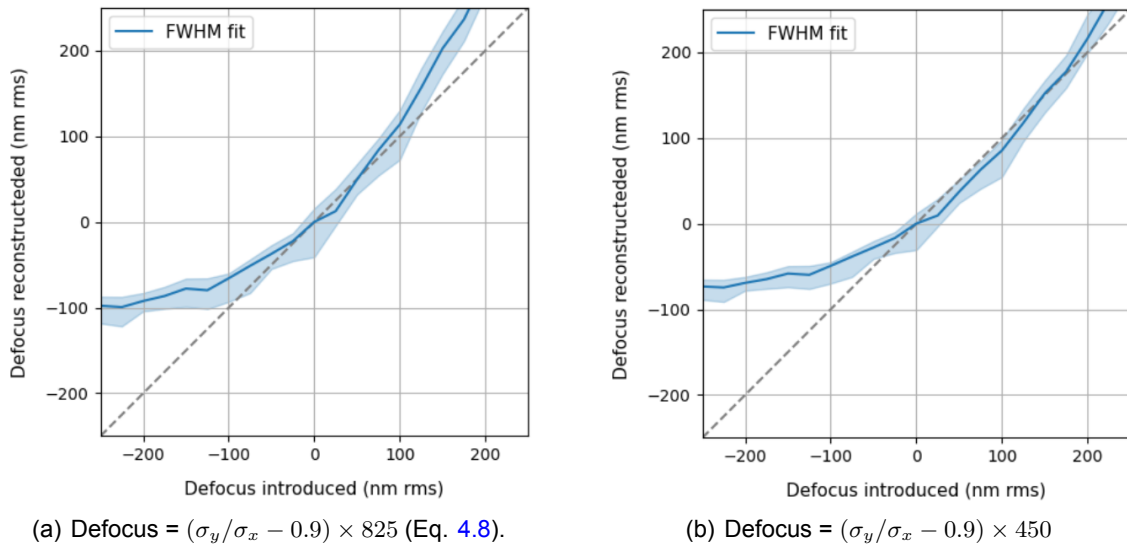


Figure 4.45: Defocus ramp obtained for on-sky data (8*8 pixels images) using the FWHM fit algorithm. The results presented on the left are obtained using the same expression as for the bench data, whereas the results on the right were obtained with a different coefficient value. The shaded area represents the 95th percentile interval, with the solid line indicating the median value.

The results obtained from GS and LiFT are depicted in Fig. 4.46. As expected from the bench results, GS showed again a slope error that needed to be corrected. The slope adjust factor that best calibrated the defocus curve was 4.75, a bit higher than the one verified with the bench data, which was 4.25. More tests with on-sky data from TRICK should be performed to confirm that the optimal slope adjustment is within a short range of values, which should be between 4.25 and 4.75 according to the obtained results with bench and on-sky data. Regarding LiFT, the results are similar to the ones obtained with bench data. No bias offset was verified (it is compensated in the diversity calibration with the in-focus images), and the slope obtained is almost near one.

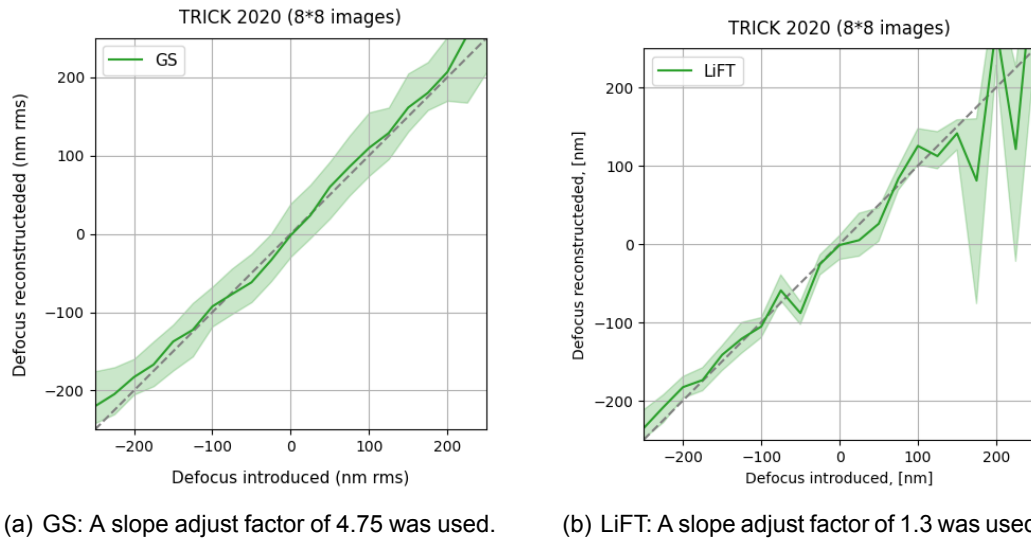


Figure 4.46: Results for Trick on-sky data from 2020 (8*8 images) with LiFT and GS. LiFT results were obtained by estimating the first three modes (TT and focus). The shaded area represents the 95th percentile, with the solid line indicating the median value.

The previous results were obtained by calculating 30 different defocus values from the 30 different images corresponding to each defocus offset. The final defocus can then be obtained by taking the median value of all measurements. Another possible way to calculate defocus is to do a single estimative using the average PSF corresponding to each defocus offset. This average PSF can be interpreted as a long-exposure (LE) PSF. Using an LE PSF may help mitigate the effects of noise or high-order residuals and, as a consequence, improve the defocus estimation. The results of such an approach are shown in Fig. 4.47. We can see that the defocus values retrieved using an LE PSF follow very closely the results of the median defocus when estimating 30 different values of defocus (Fig. 4.46). These results show that using LE PSFs is a viable procedure for defocus estimation.

From an operational point of view, this would mean that we would perform focus measurements at a lower rate. Still, on the other hand, we could significantly diminish the computational burden since we would need to perform much fewer defocus estimations and at much lower rates. Since the main goal is to measure slow focus, defocus estimates can be safely done at lower rates, and LE PSFs may be a viable solution. Of course, a trade-off must be done because we must be sufficiently fast to keep up with the slow focus variations. This type of study was made in simulations in a previous section.

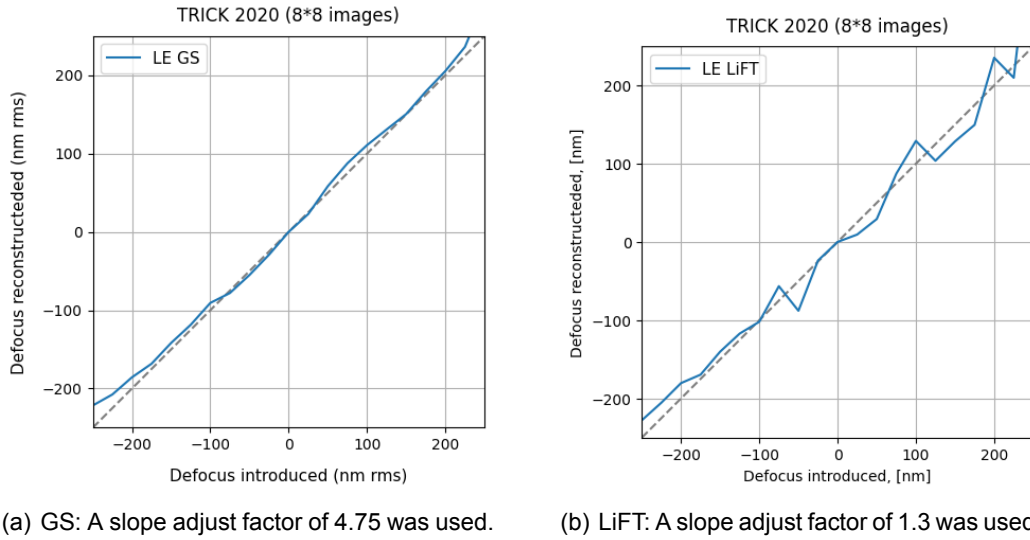


Figure 4.47: Results for Trick on-sky data from 2020 (8*8 images) with LE LiFT and LE GS. For each defocus offset, only one defocus estimate is made using an LE (mean) PSF.

In order to quantify the computational costs of the three algorithms, I measured the elapsed time for the whole simulation with on-sky data for GS, LiFT, and FWHM fit. In Tab. 4.3, I have the results when considering a round pupil with 64*64 pixels, while in Tab. 4.4, for a round pupil with 32*32 pixels. By analyzing these results, we see that the FWHM fit is clearly the fastest and its performance is not affected by the pupil sampling since it is not used in its algorithm. The elapsed time for the SE is approximately 30 times higher than the one for LE since we make 30 more defocus estimates. In the case of LE, we only do a defocus estimate per defocus offset. In addition, the computational costs can be reduced almost to half when we consider a pupil with 32*32 pixels instead of 64*64 pixels, and the obtained results are indistinguishable. Since the performance is the same, it is clearly preferable to use just 32*32 pixels for the pupil. In terms of optimization, this is an interesting and effective solution.

	Short Exposure (SE)	Long Exposure (LE)
LiFT	63 s	2.0 s
GS	101 s	3.5 s
FWHM fit	3.1 s	0.11 s

Table 4.3: Time consumption in seconds for on-sky data considering short and long exposures for a round pupil with 64*64 pixels.

	Short Exposure (SE)	Long Exposure (LE)
LiFT	24 s	0.8 s
GS	51 s	1.7 s
FWHM fit	3.1 s	0.11 s

Table 4.4: Time consumption in seconds for on-sky data considering short and long exposures for a round pupil with 32*32 pixels.

In the last part of this section, results concerning the TRICK data with 4*4 pixels images will be shown and analyzed. In Fig. 4.48, average images corresponding to defocus offsets of -250, 0, and +250 nm rms are presented. These images show that GS and LiFT will likely have difficulties estimating defocus values since there isn't enough spatial information with such a low FoV. The obtained results that will be shown next prove this prediction.

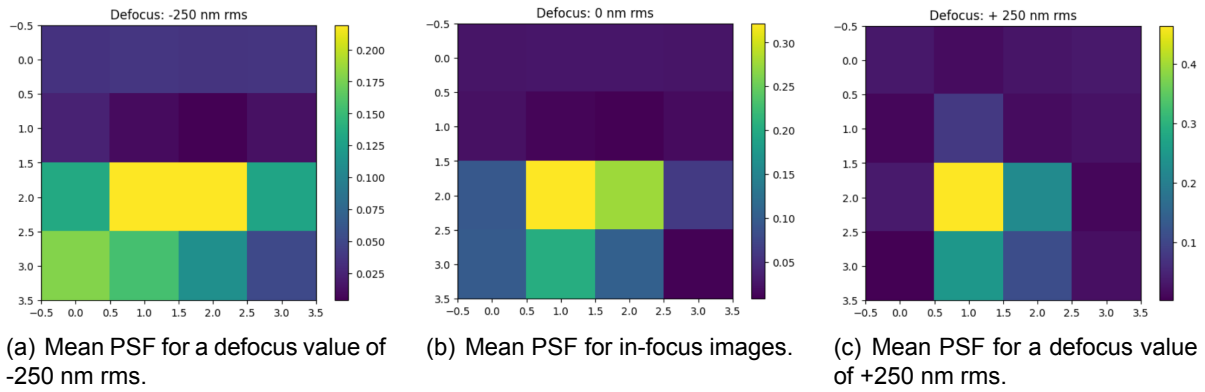


Figure 4.48: Mean images corresponding to defocus offsets of -250, 0, and +250 nm rms. These images correspond to the data file with 4*4 pixels images.

In Fig. 4.49, results referring to GS are depicted, and it is clear that linear behavior is far from satisfactory. This implies that calibration would be much more difficult with images with such low FoV, and as a result, we can conclude that a certain minimum FoV will be required to achieve calibration conditions. The results obtained with LiFT are not presented here since the LiFT algorithm completely diverged to unrealistic values with these images.

From the results obtained with bench and on-sky data, TRICK images should have a minimum of 8*8 pixels to achieve linear and calibratable behavior. However, collecting 16*16 pixels images could be preferable if possible, although that would require more time to read all the pixels. By having higher FoV, we can obtain more spatial information, and we can always crop those images if necessary in preprocessing steps in case, for example, the SNR of the images is low.

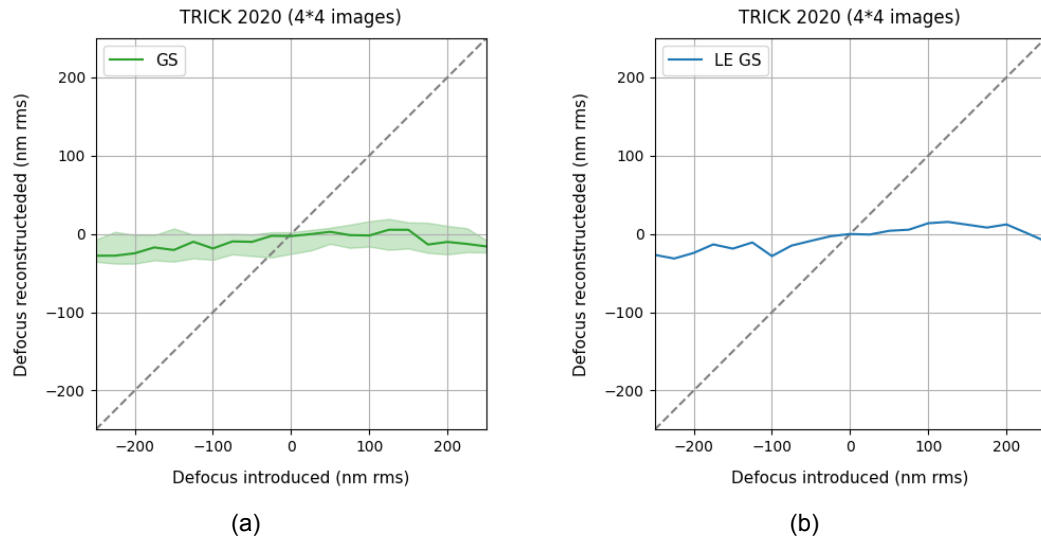


Figure 4.49: Results for Trick on-sky data from 2020 (4*4 images) with GS and LE GS. No calibration was done.

5 Conclusions and Outlook

In this work, different FPWFS techniques were studied for low-order mode sensing. This research is especially oriented towards LGS-AO systems, where TT measurements and focus tracking must be done with a faint NGS. FPWFS is an appealing solution to these problems since it is a full pupil technique, meaning that much fainter stars can be used, and no additional hardware is required.

Three different FPWFS algorithms were studied: Gerchberg-Saxton (GS), the Linearized Focal Plane Technique (LiFT), and also a fitting algorithm, which I called the Full Width at Half Maximum (FWHM) fit. These algorithms were tested through numerical simulations and with real data provided by W. M. Keck Observatory. Special attention was given to the focus measurements since the focus tracking problem poses more challenges in LGS-AO systems in relation to TT measurements.

One important differentiation aspect between the algorithms is their performance under sub-Nyquist sampling. LiFT is the one that is better prepared to deal with sub-Nyquist sampling, and optical gain errors in defocus ramps with LiFT are negligible. On the other hand, GS and the FWHM fit require considerable optical gain calibrations. The FWHM fit gives focus estimates based on a linear relationship between $\frac{\sigma_y}{\sigma_x}$, where σ_y and σ_x are obtained through the Gaussian fit. Consequently, the proportionality factor needs to be adjusted according to the sampling. In the case of GS, we apply it to sub-Nyquist sampled PSFs by assuming a "wrong sampling" of 1.0 Nyquist, i.e., we perform the algorithm as if the PSF was Nyquist sampled and then estimate the defocus. As expected, an optical gain error arises, and a slope adjustment needs to be applied. However, both GS and the FWHM fit showed that they can be calibrated since they show very good linear behavior. The linearity of the FWHM fit was not as good as the one for GS, but that may be related to the wrong assumption of proportionality between $\frac{\sigma_y}{\sigma_x}$ and defocus. This proportional behavior seems to be valid within a certain range but not for all defocus values.

Another important conclusion of this work is related to the preprocessing steps required when applying the algorithms with real data. One extremely important and mandatory step is the diversity calibration, which consists of updating our working point. Ideally, with zero defocus, the only aberrations present should be the diversity. However, the presence of NCPAs means the presence of extra aberrations that need to be compensated. By applying the algorithms with the in-focus images, we can efficiently compensate for these NCPA and solve any possible bias problems that would arise otherwise in the focus ramps.

In a real scenario the focus estimates can be done in two different ways: by doing several defocus estimates and then averaging them, or by doing a single focus estimate from a single long-exposure (LE) PSF. The tests performed with real data showed that performing a single defocus estimate from a LE PSF gives the same results as averaging several defocus measurements from short-exposure (SE) PSFs. As a consequence, it is preferable to use a LE PSF and do a single defocus estimate since the computational costs will be much more reduced, and the SNR will be higher. It was estimated that to operate below 30 nm rms of focus error in a 10 m diameter telescope, the focus tracking should be approximately done every 15 seconds. This provides more than sufficient time to achieve high SNR, which means that performing a rolling average in these conditions won't be beneficial. For a certain correction cadency, the rolling average is only beneficial if the SNR is low.

In terms of computational costs, the FWHM fit is the one with less costs, and GS is the one with more costs. However, it should be noted that LiFT only performs faster than GS when the number of estimated modes in LiFT is low (below ~ 10 modes), which is the case for low-order sensing. LiFT costs increase fast as we increase the number of estimated modes. For GS, dependence between computational costs and the number of estimated modes does not exist: GS estimates a phase rather than specific modes. The same happens with the FWHM fit, which performs a Gaussian fit for the PSF.

However, for the specific case of focus tracking, the computational costs are not a decisive factor since corrections can be made at very low rates. In fact, it was shown in this work that to achieve defocus errors in the order of 30 nm rms, the focus tracking can be done approximately every 10-15 seconds in 10-meter diameter class telescopes. For these correction cadencies, all algorithms can be safely applied since they perform well below these times. However, some tricks can be used to reduce GS and LiFT computational costs if that is needed. One simple and effective way is to reduce pupil sampling. It was shown with the keck data that using a 32*32 pix. pupil instead of a 64*64 pix. pupil, the costs can be approximately reduced to half, and the performance is exactly the same. For the case of the FWHM fit, the pupil sampling has no effect on the computational costs since the pupil is not an input.

5.1 Future steps and recommendations

From all the tests performed, a clear and unambiguous candidate for an on-sky deployment for focus tracking was not found. Fortunately, this difficult decision is due to the good performance of the algorithms rather than the contrary.

Still, overall, LiFT is the one with better performance, especially because it proved to be

more easily calibrated (in terms of optical gain). GS also proved excellent linear behavior even with real data but with a considerable optical gain adjustment needed. For that reason, it is relegated to second place. However, if this slope adjustment fully proves approximately constant (from the tests made it seems so), GS can be a good candidate for an on-sky deployment.

The FWHM fit showed to be more prone to optical gain (slope) errors than GS and LiFT. However, it seems that these errors might be associated with the wrong assumption of proportionality between $\frac{\sigma_y}{\sigma_x}$ with the amount of defocus introduced. The results show that within a certain defocus range (-100 to +100 nm rms), this assumption is valid, but it does not hold for the entire range. For that reason, the FWHM fit cannot be completely relegated for now. As a next step, the optimal relationship between $\frac{\sigma_y}{\sigma_x}$ with the introduced defocus will be studied and analyzed. In case a better relationship is found, the linearity range of this algorithm can be further increased. However, since the expected operational range in closed-loop operation is within -100 to +100 nm rms of focus, this increase in linearity may not even be necessary.

After further exploring the FWHM fit, a final decision for the best algorithm overall will have to be made based on the necessary requirements. After that, on-sky tests may be performed at Keck Observatory, where the final tests (under the most realistic conditions possible) are going to be made. If successful tests are performed, this technique may be extended to other types of LGS-AO systems around the globe.

FPWFS is a vast area of study, and many other types of algorithms can be applied for the specific task of focus tracking in other LGS-AO systems. As future recommendations, I would propose studying other kinds of FPWFS techniques for further comparisons.

6 Attachments

6.1 Proof of relationship between Zernike coefficients and wavefront error

The rms phase error of an aberrated phase is defined by [12]:

$$\sigma_{\Phi} = \sqrt{\langle \Phi^2 \rangle - \langle \Phi \rangle^2} \quad (6.1)$$

When the phase is described as an expansion of Zernike polynomials (Eq. 2.11), the mean phase value is given by:

$$\langle \Phi \rangle = \frac{1}{A} \int_A \Phi dA = \frac{1}{\pi} \sum_{j=1}^{\infty} a_j \int_A Z_j dA \quad (6.2)$$

Using the orthonormality condition (2.10), we have that:

$$\int_A Z_j dA = \int_A 1 \times Z_j dA = \int_A Z_1 \times Z_j dA = \pi \delta_{j1} \quad (6.3)$$

Using the previous two equations, we can conclude that the mean value of the phase equals the piston term (a_1):

$$\langle \Phi \rangle = a_1 \quad (6.4)$$

The phase mean square value is given by:

$$\langle \Phi^2 \rangle = \frac{1}{A} \int_A \Phi^2 dA = \frac{1}{\pi} \int_A \left(\sum_{j=1}^{\infty} a_j Z_j \right)^2 dA \quad (6.5)$$

$$\left(\sum_{j=1}^{\infty} a_j Z_j \right)^2 = \left(\sum_{j=1}^{\infty} a_j^2 Z_j^2 \right) + \left(\sum_{j=1}^{\infty} 2(a_j Z_j)(a_{j+1} Z_{j+1}) \right) \quad (6.6)$$

According to the orthonormality condition (Eq. 2.10), we have:

$$\langle \Phi^2 \rangle = \frac{1}{\pi} \left(\sum_{j=1}^{\infty} a_j^2 \int_A Z_j \times Z_j dA + \sum_{j=1}^{\infty} 2a_j a_{j+1} \int_A Z_j \times Z_{j+1} dA \right) = \sum_{j=1}^{\infty} a_j^2 \quad (6.7)$$

The phase variance is then given by:

$$\sigma_{\Phi}^2 = \langle \Phi^2 \rangle - \langle \Phi \rangle^2 = \left(\sum_{j=1}^{\infty} a_j^2 \right) - a_1^2 = \sum_{j=2}^{\infty} a_j^2, \quad (6.8)$$

which is the sum of all the squared Zernike coefficients except for the piston term. Finally, we have the formula for the rms phase error, which is equal to the square root the phase variance:

$$\text{rms phase error} = \sqrt{\sigma_{\Phi}^2} = \sqrt{\sum_{j=2}^{\infty} a_j^2} \quad (6.9)$$

6.2 Defocus error due to Sodium layer height variability

Let's assume that the telescope's pupil is at a distance of h from the sodium layer, which, neglecting the telescope's altitude, would be approximately 90 km. The artificial point source created in the Na layer emits a spherical wave, which arrives at the telescope's pupil with a radius equal to h . The longitudinal position (Z) of a point a transverse distance r from the pupil's center is given by:

$$Z = \frac{r^2}{2h}, \quad (6.10)$$

which is a parabolic approximation near the pupil's axis valid when $D \ll h$, which is true for all visible and near-infrared astronomical telescopes.

Suppose the artificial point source moves upwards a distance equal to Δh in the longitudinal direction. In that case, the spherical wavefront emitted will now have $h + \Delta h$ of radius when it arrives at the telescope's pupil. The difference between the original and the new wavefront is equal to:

$$\Delta Z = \frac{r^2}{2(h + \Delta h)} - \frac{r^2}{2h} = \frac{r^2}{2h} \left(\frac{1}{1 + \frac{\Delta h}{h}} - 1 \right) \quad (6.11)$$

Assuming that $\Delta h \ll h$, which is a realistic assumption, then ΔZ can be rewritten as:

$$\Delta Z \approx \frac{r^2}{2h} \left(1 - \frac{\Delta h}{h} - 1 \right) = -\frac{r^2}{2h^2} \Delta h \quad (6.12)$$

This term ΔZ represents the optical path difference (OPD). To calculate the rms wavefront error due to the Na layer altitude variability (σ_{Na}), we need to use the following formula:

$$\sigma_{Na} = \sqrt{\langle \Delta Z^2 \rangle - \langle \Delta Z \rangle^2} \quad (6.13)$$

Doing some further calculations, we have that:

$$\langle \Delta Z^2 \rangle = \frac{1}{A} \int_A (\Delta Z)^2 dA = \frac{4}{\pi D^2} \int_0^{\frac{D}{2}} (\Delta Z)^2 2\pi r dr = \frac{1}{2^6 \times 3} \left(\frac{D}{h} \right)^4 (\Delta h)^2 \quad (6.14)$$

$$\langle \Delta Z \rangle = \frac{1}{A} \int_A \Delta Z dA = \frac{4}{\pi D^2} \int_0^{\frac{D}{2}} (\Delta Z) 2\pi r dr = -\frac{1}{2^4} \left(\frac{D}{h} \right)^2 \Delta h \quad (6.15)$$

According to Eq. 6.13, 6.14, and 6.15, we conclude that an altitude variation (Δh) generates an rms wavefront error given by [31]:

$$\sigma_{Na} = \frac{1}{16\sqrt{3}} \left(\frac{D}{h - h_0} \right)^2 \Delta h, \quad (6.16)$$

where D is the telescope's diameter, h is the mean sodium layer altitude (approximately 90 km), and h_0 is the telescope's altitude.

References

- [1] John W Hardy. *Adaptive optics for astronomical telescopes*, volume 16. Oxford University Press, USA, 1998.
- [2] Larry C Andrews. *Field guide to atmospheric optics*. SPIE Filed Guides, 2019.
- [3] David L Fried. Statistics of a geometric representation of wavefront distortion. *JoSA*, 55(11):1427–1435, 1965.
- [4] Robert K Tyson and Benjamin West Frazier. *Field guide to adaptive optics*, volume FG03. SPIE Filed Guides, 2004.
- [5] Stephane Perrin and Paul Montgomery. Fourier optics: basic concepts. *arXiv preprint arXiv:1802.07161*, 2018.
- [6] Jason D Schmidt. *Numerical simulation of optical wave propagation with examples in MATLAB*. Spie, 2010.
- [7] Eugene Hecht. *Optics*, 5e. Pearson Education, 2002.
- [8] Lord Rayleigh. Investigations in optics, with special reference to the spectroscope. *Philosophical Magazine and Journal of science*, 1879.
- [9] Carroll Mason Sparrow. On spectroscopic resolving power. *Astrophysical Journal*, vol. 44, p. 76, 44:76, 1916.
- [10] Rene Racine and Brent L Ellerbroek. Profiles of nighttime turbulence above mauna kea and isoplanatism extension in adaptive optics. In *Adaptive Optical Systems and Applications*, volume 2534, pages 248–257. SPIE, 1995.
- [11] Robert J Noll. Zernike polynomials and atmospheric turbulence. *JOsA*, 66(3):207–211, 1976.
- [12] Kuo Niu and Chao Tian. Zernike polynomials and their applications. *Journal of Optics*, 24(12):123001, 2022.
- [13] Paul Hickson. *Fundamentals of atmospheric and adaptive optics*. The University of British Columbia, Department of Physics and Astronomy, 2008.
- [14] Jim Schwiegerling. Chapter 3.5. In *Optical specification, fabrication, and testing*. SPIE, 2014.

-
- [15] A Maréchal. Mechanical integrator for studying the distribution of light in the optical image. *JOSA*, 37(5):403_1–404, 1947.
- [16] Sergio Luis Suárez-Gómez, Carlos González-Gutiérrez, Fernando Sánchez-Lasheras, AG Basden, Icíar Montilla, FJ de Cos Juez, and M Collados-Vera. An approach using deep learning for tomographic reconstruction in solar observation. *Proceedings of the Adaptive Optics for Extremely Large Telescopes*, 5, 2017.
- [17] Darryl P Greenwood. Bandwidth specification for adaptive optics systems. *JOSA*, 67(3):390–393, 1977.
- [18] Darryl P Greenwood and David L Fried. Power spectra requirements for wave-front-compensative systems. *JOSA*, 66(3):193–206, 1976.
- [19] François Roddier. *Adaptive optics in astronomy*. Cambridge university press, 1999.
- [20] ESO. Ao modes. https://www.eso.org/sci/facilities/develop/ao/ao_modes.html, May 2024.
- [21] David L Fried. Anisoplanatism in adaptive optics. *JOSA*, 72(1):52–61, 1982.
- [22] Benoit Neichel, Celine d’Orgeville, Joseph Callingham, Francois Rigaut, Claudia Winge, and Gelys Trancho. Characterization of the sodium layer at cerro pachón, and impact on laser guide star performance. *Monthly Notices of the Royal Astronomical Society*, 429(4):3522–3532, 2013.
- [23] ESO. Laser guide star for adaptive optics. <https://www.eso.org/sci/facilities/develop/lgsf/overview.html>, May 2024.
- [24] David L Fried and John F Belsher. Analysis of fundamental limits to artificial-guide-star adaptive-optics-system performance for astronomical imaging. *JOSA A*, 11(1):277–287, 1994.
- [25] ESO. Vlt laser guide star facility. <https://www.eso.org/public/portugal/videos/eso1613b/?lang>, May 2024.
- [26] Thierry Fusco, Guido Agapito, Benoit Neichel, Sylvain Oberti, Carlos Correia, Pierre Haguenauer, Cédric Plantet, Felipe Pedreros, Zibo Ke, Anne Costille, et al. Key wave-front sensors features for laser-assisted tomographic adaptive optics systems on the extremely large telescope. *Journal of Astronomical Telescopes, Instruments, and Systems*, 8(2):021514–021514, 2022.

-
- [27] Francois Rigaut and Eric Gendron. Laser guide star in adaptive optics-the tilt determination problem. *Astronomy and Astrophysics (ISSN 0004-6361)*, vol. 261, no. 2, p. 677-684., 261:677–684, 1992.
 - [28] Jean-Marc Conan, Gérard Rousset, and Pierre-Yves Madec. Wave-front temporal spectra in high-resolution imaging through turbulence. *JOSA A*, 12(7):1559–1570, 1995.
 - [29] Glen Herriot and Craig Irvin. Mitigation of transient meteor events in sodium layer by tmt nfraos. In *Adaptive Optics Systems III*, volume 8447, pages 1102–1109. SPIE, 2012.
 - [30] G Herriot, P Hickson, B Ellerbroek, J Veran, C She, R Clare, and D Looze. Focus errors from tracking sodium layer altitude variations with laser guide star adaptive optics for the thirty meter telescope [6272-52]. In *PROCEEDINGS-SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING*, volume 6272, page 627211. International Society for Optical Engineering; 1999, 2006.
 - [31] Thomas Pfrommer, Paul Hickson, and C-Y She. High-resolution lidar experiment for the thirty meter telescope. In *1st AO4ELT conference-Adaptive Optics for Extremely Large Telescopes*, page 04001. EDP Sciences, 2010.
 - [32] Masen Lamb, David R Andersen, Jean-Pierre Véran, Carlos Correia, Glen Herriot, Matthias Rosensteiner, and Jason Fiege. Non-common path aberration corrections for current and future ao systems. In *Adaptive Optics Systems IV*, volume 9148, pages 1695–1707. SPIE, 2014.
 - [33] Robert A Gonsalves. Phase retrieval and diversity in adaptive optics. *Optical Engineering*, 21(5):829–832, 1982.
 - [34] Richard G Paxman, Brian J Thelen, Ryan J Murphy, Kurt W Gleichman, and James A Georges III. Phase-diverse adaptive optics for future telescopes. In *Advanced Wavefront Control: Methods, Devices, and Applications V*, volume 6711, pages 8–22. SPIE, 2007.
 - [35] Eko Minarto. *Optimization of Common Reflection Surface (CRS) attributes based on conjugate direction approach using Powell search method*. PhD thesis, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky, 2013.
 - [36] Arnold Daniels et al. *Field guide to infrared systems, detectors, and FPAs*, volume 161. SPIE press Bellingham, 2010.
 - [37] Pierre Léna, François Lebrun, and François Mignard. *Observational astrophysics*. Springer Science & Business Media, 2013.

-
- [38] Maxime Quesnel, G Orban de Xivry, Gilles Louppe, and Olivier Absil. A deep learning approach for focal-plane wavefront sensing using vortex phase diversity. *Astronomy & Astrophysics*, 668:A36, 2022.
 - [39] Robert A Gonsalves. Small-phase solution to the phase-retrieval problem. *Optics letters*, 26(10):684–685, 2001.
 - [40] Visa Korkiakoski, Christoph U Keller, Niek Doelman, Matthew Kenworthy, Gilles Otten, and Michel Verhaegen. Fast & furious focal-plane wavefront sensing. *Applied optics*, 53(20):4565–4579, 2014.
 - [41] Steven P Bos, Sébastien Vievard, Michaël J Wilby, Franck Snik, Julien Lozi, Olivier Guyon, Barnaby RM Norris, Nemanja Jovanovic, Frantz Martinache, J-F Sauvage, et al. On-sky verification of fast and furious focal-plane wavefront sensing: Moving forward toward controlling the island effect at subaru/scexao. *Astronomy & Astrophysics*, 639:A52, 2020.
 - [42] Robert A Gonsalves. Adaptive optics by sequential diversity imaging. In *Beyond conventional adaptive optics: a conference devoted to the development of adaptive optics for extremely large telescopes. Proceedings of the Topical Meeting held May 7-10, 2001, Venice, Italy. Edited by E. Vernet, R. Ragazzoni, S. Esposito, and N. Hubin. Garching, Germany: European Southern Observatory, 2002 ESO Conference and Workshop Proceedings, Vol. 58, ISBN 3923524617, p. 121, volume 58, page 121*, 2002.
 - [43] Steven P Bos, Michael Bottom, Sam Ragland, Jacques-Robert Delorme, Sylvain Cetre, and Laurent Pueyo. Fast and furious focal-plane wavefront sensing at wm keck observatory. In *Techniques and Instrumentation for Detection of Exoplanets X*, volume 11823, pages 434–445. SPIE, 2021.
 - [44] Richard G Paxman, Timothy J Schulz, and James R Fienup. Joint estimation of object and aberrations by using phase diversity. *JOSA A*, 9(7):1072–1085, 1992.
 - [45] Robert A Gonsalves and Robert Chidlaw. Wavefront sensing by phase retrieval. In *Applications of Digital Image Processing III*, volume 207, pages 32–39. SPIE, 1979.
 - [46] Ralph W Gerchberg and W. O Saxton. A practical algorithm for the determination of plane from image and diffraction pictures. *Optik*, 35(2):237–246, 1972.
 - [47] James R Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.

-
- [48] scikit image. Phase unwrapping. https://scikit-image.org/docs/stable/auto_examples/filters/plot_phase_unwrap.html, May 2024.
- [49] Miguel Arevallilo Herráez, David R Burton, Michael J Lalor, and Munther A Gdeisat. Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Applied optics*, 41(35):7437–7444, 2002.
- [50] Serge Meimon, Thierry Fusco, and Laurent M Mugnier. Lift: a focal-plane wavefront sensor for real-time low-order sensing on faint sources. *Optics letters*, 35(18):3036–3038, 2010.
- [51] Cédric Plantet. *Etude d'un nouveau concept d'analyse de front d'onde en plan focal appliqué à l'optique adaptative en astronomie*. PhD thesis, OBSERVATOIRE DE PARIS, 2014.
- [52] C Plantet, S Meimon, J-M Conan, and T Fusco. Experimental validation of lift for estimation of low-order modes in low-flux wavefront sensing. *Optics Express*, 21(14):16337–16352, 2013.
- [53] Cedric Plantet, Serge Meimon, Jean-Marc Conan, Benoît Neichel, and Thierry Fusco. Lift: analysis of closed loop performance in a laser assisted adaptive optics. In *Adaptive Optics Systems IV*, volume 9148, pages 824–836. SPIE, 2014.
- [54] Cedric Taissir Heritier, Christophe Verinaud, and Carlos M Correia. Oopao: Object oriented python adaptive optics. *Adaptive Optics for Extremely Large Telescopes (AO4ELT7)*, page 50, 2023.
- [55] Arseniy Kuznetsov. Lift. <https://github.com/EjjeSynho/LIFT>, May 2024.
- [56] Scipy. scipy.optimize.curve_fit. https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html, June 2024.