Technologies for Astronomical Wide-Field Adaptive Optics

GUZMAN, CHRISTIAN,DANI

Technologies for Astronomical Wide-Field Adaptive Optics

Christian Dani Guzmán

A Thesis presented for the degree of

Doctor of Philosophy

Centre for Advanced Instrumentation

Department of Physics

Durham University

United Kingdom

December 2010

Technologies for Astronomical Wide-Field Adaptive Optics

Christian Dani Guzmán

Submitted for the degree of Doctor of Philosophy

December 2010

Abstract

Adaptive Optics (AO) can greatly enhance the resolution of astronomical images, achieving close to diffraction-limited performance in the near infrared; however there are a number of areas where significant improvements can be made, one of them being the very limited field of view that current AO systems can achieve. 'Wide-field AO' encompasses those techniques devised to widen the corrected field of view, from a few tens of arcseconds in 'classical AO' systems to several arcminutes in Multi-Object AO (MOAO).

This thesis researches some topics within 'wide-field AO' for astronomy, concentrating its experimental work in some of the key technologies required to implement MOAO: open-loop models to run deformable mirrors (DM) in a MOAO system and a 'Figure Sensor' to measure the shape of a DM with required accuracy and at high-speed, in order to incorporate it into the AO control system.

Declaration

The work in this thesis is based on research carried out at the Centre for Advanced Instrumentation, the Department of Physics, Durham University, United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text. Part of this work has been published in two *Optics Express* papers, in collaboration with other co-authors; however I was the lead investigator and author. These publications are:

1) Guzmán D., de Cos F.J., Sánchez F., Myers R. and Young L. (2010) "Deformable mirror model for open-loop adaptive optics using multivariate adaptive regression splines". *Optics Express*, 18, 7, 6492-6505.
2) Guzmán D., de Cos F.J., Myers R., Guesalaga A. and Sánchez Lasheras F. (2010) "Modelling a MEMS deformable mirror using non-parametric estimation techniques". *Optics Express*, 18, 20, 21356-21369.

Acknowledgments

I would like to start by thanking my supervisors, Professor Ray Sharples and Dr. Richard Myers, from the Centre for Advanced Instrumentation. To Ray, for giving me the opportunity to pursue my dream of doing a PhD in Astronomical Instrumentation; also for his dedication, in spite of his multiple commitments, in our bi-weekly meetings for the last three years; and for teaching me that unique perspective that only an experimented instrumentalist can have. To Richard, for his constant support and time spent explaining Adaptive Optics; for his friendship and close support throughout difficult moments in the last three years and for letting me enjoy the 'Virtual Instrumentation Centre' (VIC), where we discussed all kind of topics, not only AO related…

My sincere thanks go also to my mother and mother-in-law. My mom came from Chile to help us when we most needed and my mother-in-law has given us constant, endless support throughout all these years.

I would like to take this opportunity to remember two very close people for me: my father-in-law, Osvaldo, who sadly passed away earlier this year. He taught me to have a practical and effective point of view in life, which has helped me in completing this PhD. And my beloved father, Sergio, who also passed away this year. He taught me to put passion in all we do. I dedicate this degree to my Dad; I know he would have been proud.

I would like to also thank some of the good fellows in Durham: Nigel Dipper, Tim Morris, Ali Basden, Tim Butterley, Laura Young, Colin Dunlop, Nik Looker, Gary McCallum, Prof. Simon Morris and Paul Clark. You all made my time in Durham truly enjoyable!

Finally, I would like to mention the two most important persons in my life: my wife Claudia and my son, Benjamín. I truly thank you both for supporting me throughout all this time. Claudia gave me new strength every time I needed it and has been always there for me. Benjamin, the happiest boy on Earth, has taught me the difference between what is important and what it isn't. I dedicate this degree and thesis to you, my beloved family.

Contents

Chapter 1: Adaptive Optics Systems

1.1 Introduction

Astronomy has studied celestial objects since ancient times. The first astronomical observatories used for actual research can be dated back to BC times: for example in Rhodes, with the work of the Greek astronomer Hipparchus. The fields of theoretical and observational astronomy are deeply linked to other natural sciences, such as physics, chemistry and even biology. For example, the latest discoveries of planets around other stars are giving us the opportunity to think of life beyond the Earth, connecting astronomy and biology. We may be able to answer soon that old question “are we alone?” with a scientific basis for the first time in human history.

Observational astronomy is the part of astronomy that deals with research on objects in the sky through scientific observations. The telescope, developed 400 years ago, is the prime instrument for this purpose. Its steady growth in size and precision has enabled observations of extremely faint objects, allowing astronomers to make transcendental discoveries, such as the accelerated expansion of the Universe or the presence of a super massive black hole at the centre of our galaxy. New discoveries as exciting as these ones are waiting for the next generation of telescopes to be built in selected locations on the Earth, where the thin air above the telescope allows us observations of an unprecedented quality. The next generation of telescopes has been conveniently named the “Extremely Large Telescopes” or ELTs, to denote the leap forward with respect to their predecessors. From the telescope built by Galileo 400 years ago with a lens diameter of a few centimetres, we have progressed to building mirrors of 10 meters diameter for the current generation of telescopes. We have ambitions to use mirror diameters of 25 to 42 meters for the next generation. In order to build these giant machines, we need to develop new technologies as well as overcome monetary and political obstacles. The work reported in this thesis is part of the development of new technologies for the next generation of telescopes.

Ground-based telescopes suffer the negative effects of the atmosphere, even at the best astronomical sites on the Earth. Starlight is distorted by the turbulence present in the atmosphere,

causing a degrading effect on the astronomical images we can acquire from the sky. Early ideas to remove this effect can be traced back to the 1950's, from a seminal paper by H. Babcock in 1953, which contained an early proposal for adaptive optics. Adaptive optics is the technique that deals with the effect of atmospheric turbulence on telescope performance, with the goal of compensating the turbulence in real-time, while collecting the light from the sky for astronomical observations. A working adaptive optics system in a telescope increases the spatial resolution of the images obtained. The spatial resolution in a telescope image improves as the diameter of the telescope gets larger. This ultimately allows the resolution of individual point sources (stars for example) that are close together on the sky. High spatial resolution allows us to observe the constituent elements of a distant galaxy, enabling the study of its internal structure. This is science not possible if the resolution were limited such that we observed the galaxy as a diffuse nebula. As we shall see in this chapter, the spatial resolution is a function of the wavelength and size of the optics, therefore bigger telescopes are preferred not only because of their larger collecting area, but also because of the spatial resolution they can achieve.

Atmospheric turbulence acts by adding "small lenses" in the light path, limiting the resolution of a big telescope to not more than what an amateur telescope of 20 cm diameter would see. The turbulence is a dynamic and stochastic process, fundamentally caused by the sun's heat during the day. Adaptive optics systems rely on measuring the deformation of the wavefront caused by turbulence and correcting it using a device such as a deformable mirror. As this chapter will describe, this correction is different for each point in the field of a telescope, so a simple adaptive optics system is not capable of wide-field compensation. In fact, extending the corrected field-of-view is one of the main areas of research in adaptive optics, with this project being one of them.

This thesis is organized as follows: chapter 1 introduces the most important concepts behind turbulence and adaptive optics systems designed to compensate it, including the most complex systems for wide-field compensation. Chapter 2 describes a particular type of wide-field system, multi-object adaptive optics, which is the variant of adaptive optics behind this project. Chapter 3 includes the experimental work implemented to model deformable mirrors for multi-object adaptive optics, work published in two pier-reviewed papers this year, and included in the appendix of this

thesis. Chapter 4 presents a preliminary study of a figure sensor to be used to control a deformable mirror for multi-object adaptive optics. Chapter 5 presents the general conclusions of this thesis as well as future areas of research that can be conducted based on the results presented here.

1.2 Telescope Resolution

When a telescope focuses an image from the sky, it deals with the ray and wave nature of light. A correct optical figure in the telescope optics produces an image free of aberrations, however, the wave nature of light causes diffraction at the telescope, which gives rise to the upper limit in resolution the image can achieve (the *diffraction limit*). When the telescope focuses a point source, the image at the focal plane is the *point-spread function (PSF)*, which, due to diffraction is not a dimensionless point anymore, but a complex irradiance pattern that depends on the size and shape of the telescope aperture. Given that a point source accounts for all possible spatial frequencies[1], the PSF is the response of the optical system.

Fraunhofer modelled the irradiance distribution of light caused by diffraction for a circular aperture, which applies in the case of most telescopes. The irradiance pattern E at the focal plane can be described by a Bessel function of the first kind J_1 (Driggers 2003). The mathematical expression, also called the Airy pattern, is

$$E(r) = E_A \left(\frac{A}{\lambda f}\right)^2 \left[\frac{2J_1\left(\frac{\pi r D}{\lambda f}\right)}{\frac{\pi r D}{\lambda f}}\right]^2 \quad (1.1),$$

where:

E_A : is the irradiance at the aperture

$E(r)$: is the irradiance intensity at a distance r from the optical axis

λ : is the wavelength of light under study

D : diameter of the telescope

A : area of the aperture

[1] A point source in the spatial domain can be decomposed in components with individual spatial frequencies. This is equivalent to the Fourier transform of a Dirac delta, which has components at all frequencies and with the same magnitude. This is why the PSF represents the response of the system, since it shows how an optical system responds to all possible frequencies.

f : focal length

An irradiance pattern for the case of an 8-m telescope[2] is presented in Fig 1.1. The first ring occurs at $r \cong \pm 1.22\,\lambda f/D$. This is $r = \pm 9.15$ µm for the case in Fig 1.1.

The ability of an optical system to image two separate point sources at its focal plane is a function of its diameter. Lord Rayleigh defined a criterion for the resolution limit of a telescope, which is in relation to the first ring of the PSF. The expression for the angular resolution is then

$$R = \frac{1.22\lambda}{D}\,(\text{rad}) = \frac{251640\lambda}{D}\,(\text{arcsec}) \quad (1.2).$$

Two PSFs having the distance separation of Eq. 1.2 are presented in Fig. 1.2. This is an upper limit in terms of the resolving power a telescope can achieve. In the next section, we shall see that atmospheric turbulence limits the resolution of any ground-based optical telescope, to the equivalent of a few tens of centimetres at best.

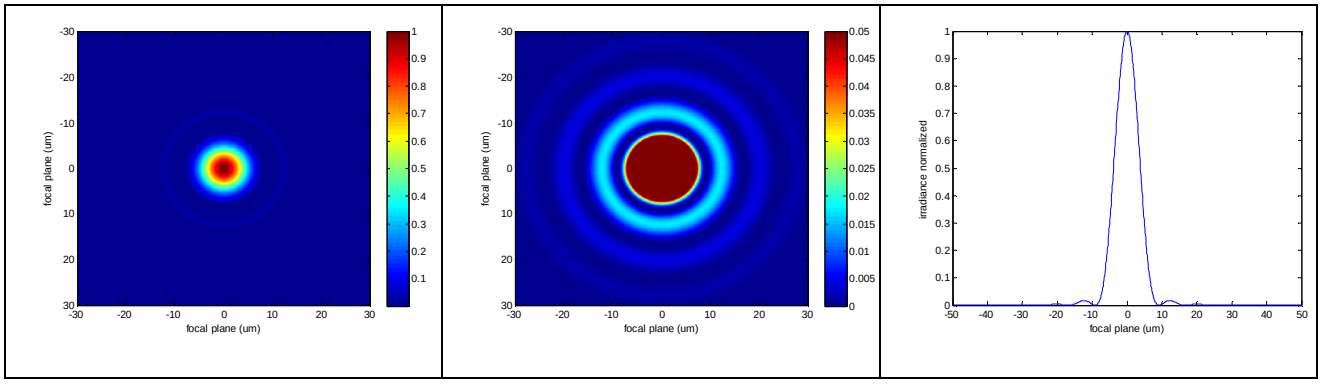


Fig. 1.1: Diffraction-limited irradiance pattern produced by an 8-m telescope at its Nasmyth focus (with F-ratio = 15), for wavelength = 500 nm. Left panel: Airy disk with linear colour scale. Central panel: scale adjusted to see the rings. Right panel: slice of the Airy disk through the origin.

[2] This is an example of a single mirror telescope working at f/# = 15 and without secondary mirror.

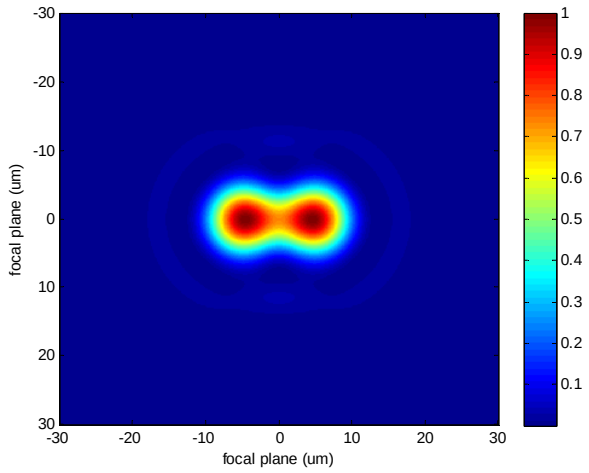


Fig. 1.2: Rayleigh resolution criterion, illustrated for two point sources on an 8-m telescope working at F-ratio = 15 and 500 nm wavelength.

1.3 Atmospheric Turbulence

The turbulence in the atmosphere is the result of mixing of air at different temperatures, which is caused by solar heating during the day. Having a mix of air molecules at different temperatures, the air density is constantly changing in a random fashion, which in turn produces a change in the index of refraction of the air. Light coming from stars is slightly refracted by the atmosphere, producing degradation in the resolution of the image. Atmospheric turbulence at

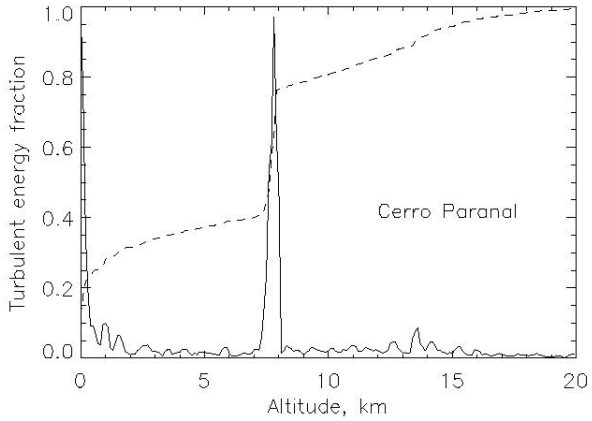


Fig. 1.3: Turbulence profile at Cerro Paranal in northern Chile. The solid line represents the turbulence energy (in normalized units) and the segmented line is the accumulated energy. Plot courtesy A. Tokovinin (CTIO).

observatories has been extensively studied (Wilson 2009) in the last decade, indicating that the turbulence is mostly limited to thin layers within the first 15 km of atmosphere above the observatory.

Fig. 1.3 shows this effect at Cerro Paranal in Chile, where the VLT is located. The figure is a kinetic energy plot with respect to altitude above the telescope. The two distinctive peaks are from the two strong turbulent layers, one immediately above the telescope (the "Ground Layer") and a second around 8 km of altitude. The two turbulent layers account for about two thirds of the total turbulence, but there is one third that is distributed along the whole column of turbulence.

The most commonly used model for atmospheric turbulence was proposed by Kolmogorov (1941) and developed by Tatarski (1961). This model is general for turbulence in a fluid medium, in which the turbulence is a consequence of adding energy to the medium in large spatial scales: solar heating in the case of atmospheric turbulence. The Kolmogorov model describes the turbulent behaviour between the "outer scale" L_0 and the "inner scale" l_0. The former corresponds to the scale where energy is added to the medium (usually tens of meters), while the latter occurs when the energy is finally dissipated as heat by viscous friction (within millimetres).

Atmospheric turbulence is a random process. As such, it can be modelled with the structure function $D_n(\Delta x)$, defined as the average difference between two values of a random variable for a large number of points, with the random variable being the index of refraction $n(x)$, i.e.

$$D_n(\Delta x) = \langle [n(x) - n(x')^2] \rangle \quad (1.3).$$

The structure function for the refractive index variation of turbulent air in the Kolmogorov model is

$$D_n(\Delta x) = C_n^2(h)\Delta x^{2/3} \quad l_0 < \Delta x < L_0 \ (1.4).$$

In Eq. 1.4, $C_n^2(h)$ is the vertical refractive index structure parameter, which is a function of the altitude h.

Developed from this model, the *Fried* parameter r_0 allows one to characterize the strength of the whole atmosphere (Fried 1965). The mathematical expression for r_0 is

$$r_0 = \left[0.423 \left(\frac{2\pi}{\lambda}\right)^2 (\sec \xi) \int C_n^2(h) dh\right]^{-3/5} \quad (1.5),$$

where ξ is the zenith angle of observation and λ is the observational wavelength.

The effect of turbulence on a telescope of diameter *D* is seen on the variance[3] of the wavefront σ_ϕ^2, which can be computed as

$$\sigma_\phi^2 = 1.0299 \left(\frac{D}{r_0}\right)^{5/3} \quad (1.6).$$

Alternatively, one can use the resolution *R* of the telescope in the presence of turbulence.

$$R = \frac{1.22\lambda}{r_0} \quad (1.7).$$

The main interpretation of Eq. 1.7 and its similarity with Eq. 1.2 is straight forward: in the presence of turbulence, the resolution is set by the parameter r_0, which represents the diameter of a telescope that produces the same resolution. Typical values for the *Fried* parameter are $5\ cm < r_0 < 20\ cm$. When a telescope of such a diameter is used, only the first order of the turbulence, i.e. the wavefront gradient or "tip and tilt" needs to be compensated for achieving the diffraction limit.

An alternative physical interpretation for the Fried parameter is that r_0 is the aperture size that produces a mean square wavefront error of around 1 rad^2.

1.4 Adaptive Optics Systems

A general scheme for an adaptive optics system is presented in Fig. 1.4. The telescope is pointed to a guide star, which is used for measuring the atmospheric turbulence. The light beam from the telescope is received by the "deformable mirror" (DM) or "wavefront corrector", which corrects the aberrations caused by the turbulence, shaping its surface with a figure that nulls the shape of the wavefront that passed through the atmosphere.

[3] This applies when the *piston* term (or linear offset level of the wavefront) has been removed.

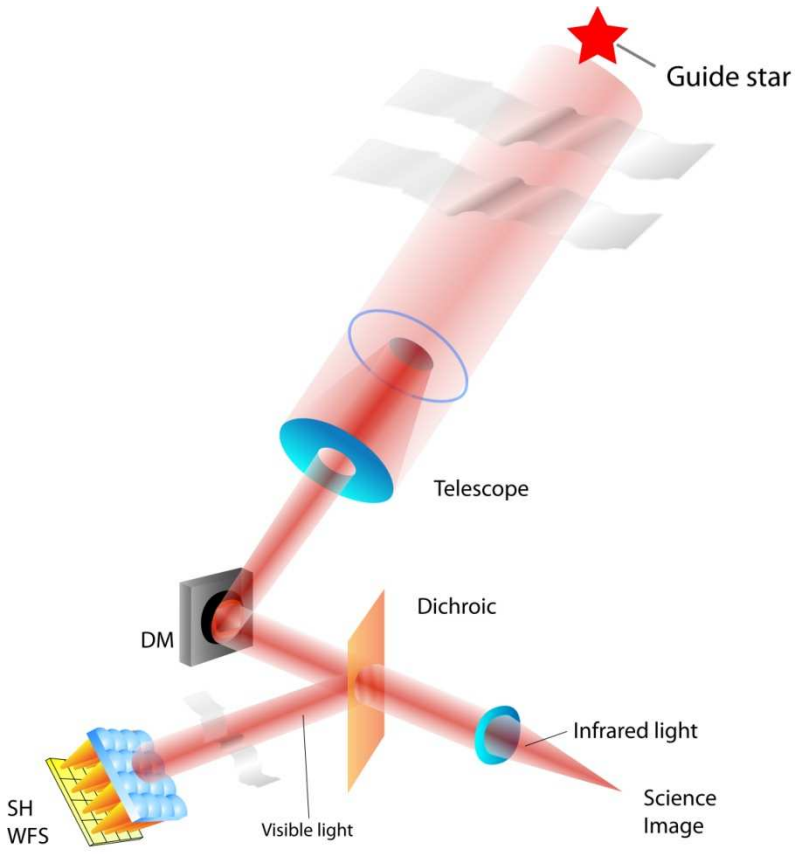


Fig. 1.4: Diagram of a typical AO system placed after a telescope. The AO system is composed of a deformable mirror and a phase or wavefront sensor.

After the deformable mirror, an AO system includes a phase or wavefront sensor (WFS) to measure the residual aberrations after the wavefront has been corrected by the DM. The Shack-Hartmann WFS (SH-WFS) is the most commonly used in AO systems for astronomy and it is the type sketched in Fig 1.4. The WFS uses optical wavelengths[4] for measuring the wavefront. This is usually implemented by diverting the light with a dichroic, letting the infrared light pass to a science instrument.

The AO system works in closed-loop by trying to keep a flat (or null) residual wavefront at the science image. This is orchestrated by a powerful computer, capable of processing the large amount of information coming from the wavefront sensor and calculating the commands to the deformable mirror. A “real-time” computer is used, meaning one which is capable of doing these

[4] Optical wavelengths are commonly called to those between 350 and 1000 nm.

computations in real-time, at rates of potentially thousands of updates per second. Current personal computers have become so powerful that they are starting to be used as AO computers.

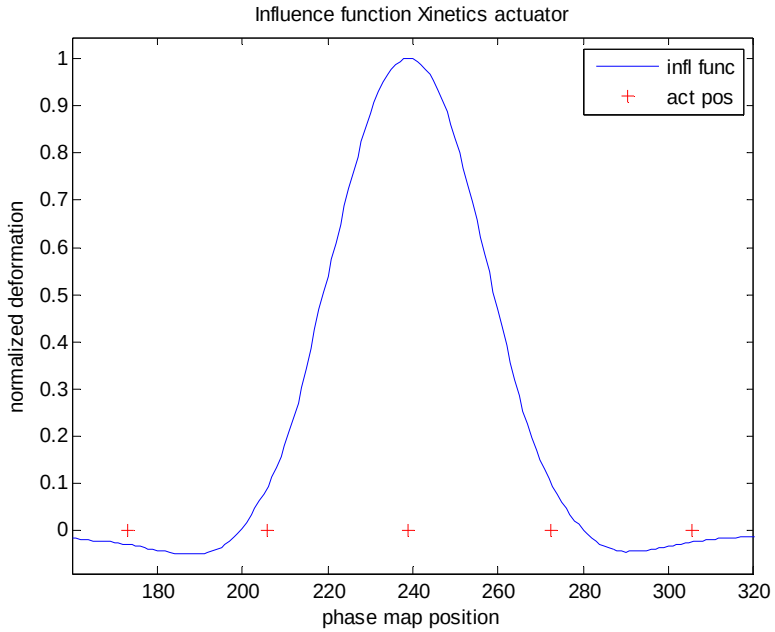


Fig. 1.5: A typical influence function, showing the position of the actuators. This plot is taken from interferometric data obtained with one of the DMs modelled in Chapter 3. The X axis is presented in interferometer phase map units, while the Y axis is normalized deformation. A 10% cross-coupling to neighbour actuators can be seen as the amount of deformation on the membrane at immediate adjacent actuators (the ones at positions 205 and 272). Adjacent actuators are 7 mm apart on this mirror.

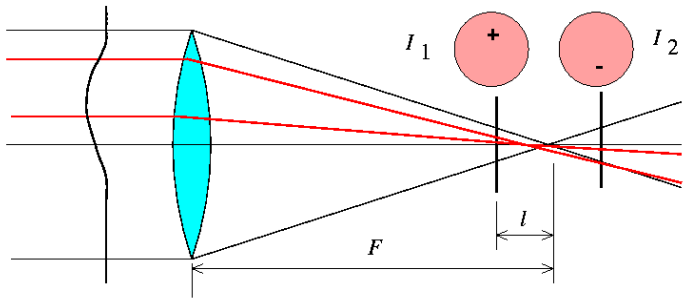
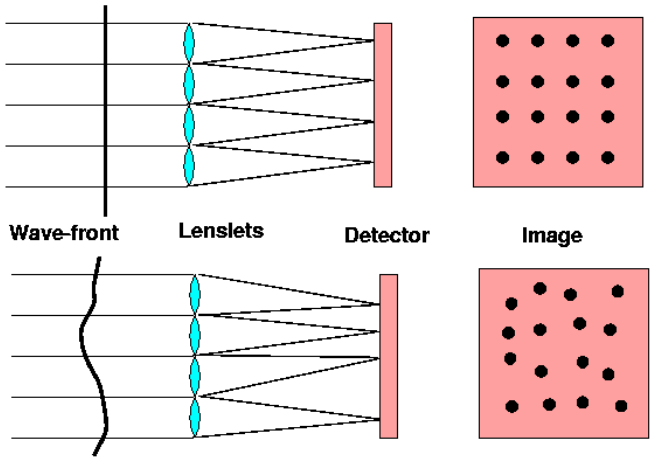
Different types of deformable mirrors have been developed over the last 30 years. They are usually classified depending on the technology used in the actuators. Most DMs have actuators in an equally spaced rectangular grid, where their acting direction is parallel amongst them and perpendicular to the mirror membrane. By varying their length by a few microns, the membrane becomes locally deformed at the actuator position. The local deformation of the membrane is usually called the "influence function" produced by the actuator. A typical influence function on the membrane spans neighbouring actuators, having a 10% to 20% effect on the immediately adjacent actuators. Fig 1.5 presents a typical influence function.

The main types are listed here and described in more detail in Chapter 3.

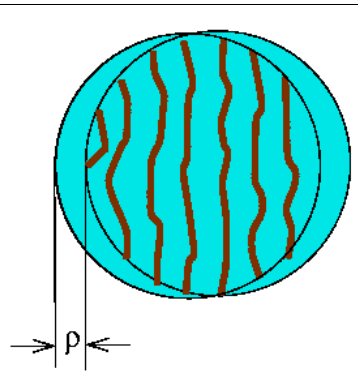
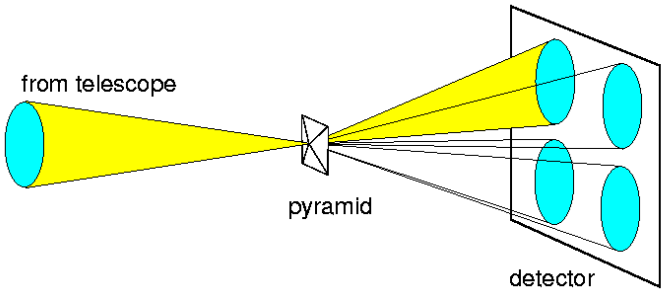
- Piezo stack: this is a DM made of discrete actuators, composed of piezo-electric ceramics

- Electrostrictive: this is a DM made of discrete actuators, composed of an electrosctrictive material
- Electrostatic: MEMS (Micro-Electro-Mechanical-System) type of devices, made on a single piece of silicon, where electrostatic actuators are built
- Magnetic: these are integrated devices using voice-coil actuators, composed of a magnet and a solenoid

The main types of wavefront sensors are briefly described here, with representative diagrams taken from Andrei Tokovinin's AO Tutorial[5].

Curvature: The wavefront aberration is measured at two planes around the focal plane, where the aberration causes excess illumination (at I_1) and lack of illumination (at I_2).	
Shack-Hartmann: The pupil is divided into "subapertures" by a lenslet array. The aberrations produce displacements of the spots produced by each subaperture. The position of the spot is proportional to the gradient of the wavefront at each subaperture. '*Centroiding*' is the process of finding the position of each spot, computing the Centre of Gravity of the subaperture in pixel units (or using a more advanced technique). We implemented a Shack-Hartmann WFS, which we describe in detail in Chapter 4.	

[5] This is an online tutorial, available at http://www.ctio.noao.edu/~atokovin/tutorial/index.html.

Lateral Shearing Interferometer: An interferogram can be formed by making the wavefront interfere with a shifted version of itself. The aberrations can be measured by analysing the fringe shapes.	
Pyramid: A glass pyramid is placed at the focus of a lens, dissecting the beam into 4 apertures. It is possible to measure the wavefront aberrations by comparing the intensity on equivalent positions on each subaperture.	

1.5 Angular Anisoplanatism

One of the main limitations of AO is angular anisoplanatism. This is the effect of not measuring the turbulence along the same line-of-sight as the science object which is being imaged, as can be seen on the left panel of Fig. 1.6, for a single layer of turbulence at height *h*.

The mean square anisoplanatic error at angle α in Fig 1.6 (left panel) for $D \gg r_0$ (i.e. when the telescope diameter is much larger than the equivalent size of turbulence) is

$$\langle \sigma_\alpha^2 \rangle = 2.914 \left(\frac{2\pi}{\lambda}\right)^2 (\sec\zeta)^{8/3} \alpha^{5/3} \int h^{5/3} C_n^2(h) dh \qquad (1.8).$$

Anisoplanatism causes a rapid deterioration of the correction as the angle between the science object and the guide star increases, a situation presented in the right panel of Fig. 1.6 for real observations at Keck Observatory in Mauna Kea (Hawaii), using the standard metric of *Strehl* ratio[6] for determining the level of correction.

[6] The Strehl ratio is a figure of merit to measure the quality of an optical system. It is defined as the ratio of central intensities of the PSF formed by the optical system under evaluation and the PSF formed by a diffraction-limited theoretical system. The Strehl ratio is a standard figure for AO system performance evaluation.

There is an angle within which the turbulence does not change significantly; this is called the *isoplanatic angle* θ_0 that depends on the location of the turbulence as well as the strength. Its mathematical expression is

$$\theta_0 = \left[2.914\left(\frac{2\pi}{\lambda}\right)^2 (\sec\zeta)^{8/3} \int h^{5/3} C_n^2(h) dh\right]^{-3/5} \quad (1.9).$$

For the simple case of one turbulent layer at an altitude h (Fig 1.6 left) the isoplanatic angle is

$$\theta_0 = 0.314(\cos\alpha)\frac{r_0}{h} \quad (1.10).$$

It can be easily confirmed that high altitude turbulence reduces the isoplanatic angle as does a small value of r_0. On the contrary, low altitude turbulence increases the isoplanatic angle, while large values of r_0 produce a similar effect.

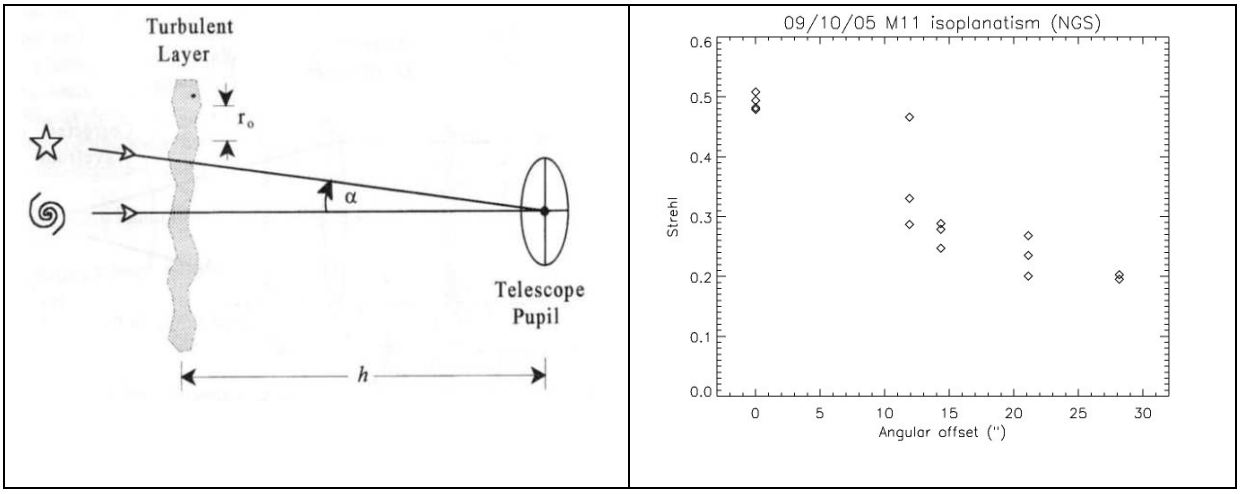


Fig. 1.6: Angular anisoplanatism in adaptive optics. Left panel: the concept of decorrelation between a turbulence measurement (using a guide star) and the turbulence 'seen' by a science object (a galaxy); diagram reproduced from Hardy (1998). Right panel: Strehl ratio for various angular offset measurements. Diagram reproduced from W.M. Keck Observatory's web page.

1.6 Laser Guide Stars

Using real stars for wavefront sensing suffers from the limitation that only bright stars can be used, depending on the sensitivity of the wavefront sensor. This is because wavefront sensing implies some type of aperture sub-division for measuring the local deformation of the wavefront. In a

Shack-Hartmann wavefront sensor[7] for example, the pupil is divided into sub-pupils, each of which measures the local gradient of the wavefront. Having sub-pupils implies dividing the limited number of incoming photons from the guide star amongst them and, depending on the noise floor of the sensor as well as the collecting power of the telescope, only relatively bright stars can therefore be used. The final implication of using real stars or "natural" guide stars (NGS) is a limited sky-coverage, which severely restricts the type of observation an adaptive optics system can preform. For example, observing distant galaxies at sky positions away from the Milky Way is generally not possible, since the guide star must normally come from the Milky Way.

To overcome this limitation, "artificial" stars have been created by means of launching high-powered lasers into the atmosphere. The laser is launched from the telescope structure and therefore it will point to where the telescope is observing, creating what is called a Laser Guide Star or LGS.

There are two distinctive types of LGS in AO: Sodium and Rayleigh. A brief description of each of them is given here:

- Sodium LGS: There is a layer of sodium atoms in the atmosphere, present at an altitude between 80 and 100 km. A laser tuned to an electronic transition within the sodium atom is used, typically the D2 line, centred at 589nm. When the sodium atoms are illuminated at this wavelength, they are stimulated and emit a sufficient number of photons back to the telescope to perform wavefront sensing.
- Rayleigh LGS: Photons can be produced by Rayleigh scattering from molecules within the atmosphere. This type of laser guide star is focused at lower altitudes than Sodium LGS, therefore they cannot sample the complete volume of turbulence.

Although having a LGS allows us to count on a bright enough guide stars to observe at any point in the sky, it comes with some limitations, described here:

- Tip-tilt Indetermination: The laser is launched to the atmosphere from a small telescope, mounted either at the side or at the centre of the main telescope. As it goes up through the atmosphere, the beam is affected by the same turbulence that it is trying to measure. As the

[7] The Shack-Hartmann wavefront sensor will be studied in detail in Chapter 4.

laser beacon is small (its diameter is comparable to a few r_0), its position on the sky gets displaced by turbulence (affected by the global tip-tilt or first order of turbulence). The same effect occurs for the returning photons from the LGS, therefore the tip and tilt information is lost. Because of this, a laser guide star cannot be used to measure the global tip and tilt in an AO system and a real star needs to be used for this task. A tip/tilt sensor uses fainter stars in the field compared to a NGS WFS, so this is not so strong a limitation in terms of sky coverage for a LGS system.

- Focal Anisoplanatism: The laser is focused at a finite altitude, so unlike real stars, they do not produce a truly collimated beam (as coming from infinity), even in the case of sodium LGS at 90 km. The finite altitude of the LGS causes the "cone effect" or focal anisoplanatism, where sections of the atmosphere are not sampled by a LGS, since the turbulence is not sampled with a cylinder, as a NGS does. The strength of the cone effect depends on the turbulence profile C_n^2. Fried and Belsher (1994) derived a single parameter d_0 that permits one to quantify the wavefront error produced by the focal anisoplanatism (FA). Its effect is given by $\sigma_{FA}^2 = (D/d_0)^{5/3}$, where d_0 is the LGS equivalent to the Fried parameter r_0. d_0 is proportional to $\lambda^{6/5}$.

Multiple LGS systems are being designed to alleviate focal anisoplanatism, since they can reduce the cone effect that one LGS has. These artificial guide stars can be placed anywhere in the field of the telescope, but they are usually located around the scientific Field of View (FoV). Having several LGS allows one to implement more advanced techniques to enhance the corrected field of an AO system. These techniques are referred to as "Wide Field Adaptive Optics" and are described in the next section.

1.7 Wide Field Adaptive Optics

Due to angular anisoplanatism, a simple AO system such as the one shown in Fig. 1.4, is only capable of achieving high resolution in a field-of-view of a few arcseconds. This system, now

called “classical AO”, is also known as “Single Conjugated AO” (SCAO), since there is only DM in the system, usually conjugated to the telescope pupil.

In order to increase the corrected field, it is necessary to measure the turbulence at more than one position in the field, probing the real volumetric character of the atmosphere. The first proposal to expand the corrected FoV was “Multi-Conjugated Adaptive Optics” (MCAO) (Tallon 1990), in which the volumetric aspect of the turbulence (with several layers of turbulence above the telescope) is taking into account, by using a number of deformable mirrors, conjugated at different altitudes, in order to compensate the turbulence by layers and thus extend the FoV.

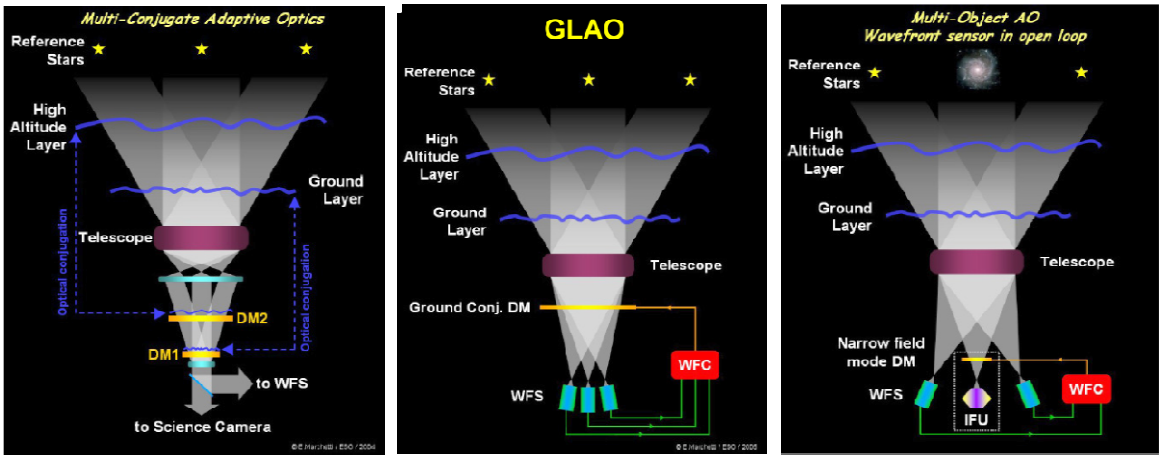


Fig. 1.7: Wide field adaptive optics systems. Left panel: multi-conjugated adaptive optics. Central panel: ground-layer adaptive optics. Right panel: multi-object adaptive optics. Images courtesy ESO.

In MCAO, several WFS are placed at different positions in the field to sample the turbulent volume, as the left panel of Fig. 1.7 depicts. These measurements allow the system to compute the corrections for each DM, taking into consideration that they will correct the turbulence in layers. The process behind this computation is usually called “tomography”, since it is similar to the concept of reconstructing a volume which is imaged at different angles.

A simplification of MCAO is Ground-Layer AO (GLAO) (Rigaut 2002), in which only one deformable mirror is used, to correct the ground layer of turbulence (see central panel of Fig. 1.7). The ground layer has been found to be the strongest in many observatories; see Fig 1.3. Multiple WFS are still needed to sample the turbulence common to all lines of sight. The tomography in GLAO is

implemented by computing the average turbulence from all WFS, which corresponds to the ground layer.

The last type of wide-field AO is the one relevant to this project, Multi-Object AO, or MOAO; see right panel in Fig. 1.7. This type of AO system departs from all previous systems described, in the sense that the turbulence is measured by several WFS in open-loop, i.e. they see raw incoming turbulence before it is corrected by a DM. The WFS data are used to compute a tomographic reconstruction of the turbulence applicable to certain positions in the FoV, where a DM is placed (conjugated to the telescope pupil) in order to correct all turbulence for that specific line-of-sight. These positions in the field correspond to science objects, while the WFS measure the turbulence at other positions in the field, where LGS (or NGS) are located. Besides being in open-loop, MOAO differs from the previous techniques in that it is not possible to obtain a continuous corrected patch of the sky, but only a few selected small regions. Therefore, this technique is interesting for spectroscopy mostly, where placing a slicing element such as an "Integral Field Unit" (*IFU* in Fig. 1.7) allows one to obtain 3D spectroscopy[8] of the corrected sub-field. This type of AO is described in detail in Chapter 2. One of the first on-sky demonstrators of this technique is CANARY, a one-channel MOAO system to be deployed at the 4.2 meter William Hershel Telescope in the Canary Islands. This and other MOAO instruments are described in Chapter 2.

1.8 Other Adaptive Optics Configurations

Other types of AO that have been developed in the last few years are Laser Tomographic AO and Extreme AO, which we briefly described here for completeness.

- Laser Tomographic AO: The goal of this technique is to achieve the diffraction limit of the telescope for a narrow FoV, usually not wider than the isoplanatic angle of a single-conjugated AO system. To achieve this level of correction, several LGS are placed in a narrow field, using tomographic reconstruction techniques to compute the whole volume of

[8] 3D spectroscopy is a technique developed in astronomical instrumentation to obtain spectra of each element of a two dimensional image, producing three-dimensional "data-cubes" where the wavelength is the third dimension.

turbulence and a DM with a large number of actuators: usually a thousand or more for an 8 m telescope.

- Extreme AO: This is a type of single-conjugated AO in which a very high level of correction is required, in order to achieve a very high contrast when imaging. The goal of this technique is to image extremely faint objects, such as planets around other stars (exoplanets) or faint companions to nearby stars. One such system is being built for the Gemini South 8 meter telescope, called the "Gemini Planet Imager". The high order DM of this system has 4000 actuators, the largest one in terms of number of actuators in an AO system for astronomy.

1.9 References

Driggers, R. "Encyclopedia of Optical Engineering", CRC, (2003)

Fried, D., "Statistics of a Geometric Representation of Wavefront Distortion", JOSA, 55 pp. 1427–1435, (1965)

Fried, D. and Belsher, J., "Analysis of fundamental limits to artificial-guide-star adaptive-optics-system performance for astronomical imaging", JOSA-A, 11, 1, pp. 277-287, (1994)

Hardy, J. "Adaptive Optics for Astronomical Telescopes", Oxford University Press, (1998)

Kolmogorov, A., "Dissipation of Energy in the Locally Isotropic Turbulence". Reprinted in Royal Society of London Proceedings Series A (1991), 434, pp. 15–17, (1941)

Rigaut, F., "Ground-Conjugate Wide Field Adaptive Optics for ELTs", Proceedings of the Topical Meeting held May 7-10, 2001, Venice, Italy. (2002)

Tallon, M. and Foy, R. "Adaptive Telescope with laser probe: anisoplanatism and cone effect" A&A 235, pp. 549-557 (1990)

Tatarski, V., "Wave propagation in a turbulent medium", McGraw-Hill, New York, (1961)

Wilson, R. W, Butterley, T, Sarazin, M, "The Durham/ESO SLODAR optical turbulence profiler", MNRAS, 399, 4, pp. 2129-2138, (2009)

Chapter 2: Multi-Object Adaptive Optics

2.1 Overview

This chapter discusses a particular type of adaptive optics: multi-object adaptive optics (MOAO), which is one of the techniques devised in recent years to overcome the limited FoV a standard adaptive optics system can achieve (Hammer 2001, Assemat 2007). The main characteristics of this technique are presented and discussed, followed by some descriptions of the most important aspects to take into account when dealing with a MOAO system. Finally, a brief survey of current systems in development is presented.

2.2 Introduction

Adaptive Optics (AO) was introduced in Chapter 1 as a technique to compensate for the loss in spatial resolution in astronomical images due to changes in the index of refraction of the atmosphere caused by turbulent mixing of air pockets at different temperatures. One of the main limitations of AO is angular anisoplanatism, which is a consequence of the decreasing correlation of turbulence effects between lines of sight through the atmosphere as their angular separation is increased.

Existing generations of AO systems have been very productive in obtaining images and spectra of astronomical objects close to the guide stars measuring turbulence, particularly at near-infrared wavelengths where the correlation length for the turbulence is larger. Laser guide stars are optimum in this respect, since they can be pointed close to the science target along with the telescope. However, these systems still require a natural guide star to measure tip-tilt, which limits the ability to point to any place in the sky.

For a number of science topics within observational astronomy, it is preferred to have a wide FoV available for observing. In particular, the multi-object spectrograph (MOS) is a type of wide field astronomical instrument, which improves the efficiency of observations by obtaining spectra for many

astronomical targets simultaneously. This is of particular interest for science such as extragalactic astronomy (galaxies beyond our own Milky Way) where there are often large numbers of galaxies available within a FoV of a few arcminutes diameter. 'Workhorse' MOS are common in observatories: some examples of such instruments are listed in Table 2.1.

As it can be seen in Table 2.1, each of these MOS has a FoV several arcminutes across, which is adequate for observing multiple galaxies. An example of an interesting extragalactic field is presented in Fig. 2.1. It is Abell 2218, known for its gravitational lensing[9] effect, with a number of gravitational lensed arcs seen in the image.

Table 2.1: Examples of multi-object spectrographs in the largest telescopes.

Observatory	Instrument	Wavelength	Field of View (armin^2)
Keck	DEIMOS	Optical	81.5
Keck	MOSFIRE	Near infrared	37.2
VLT	VIMOS	Optical	4 x 56
VLT	KMOS	Near infrared	40.7
Magellan	IMACS	Optical	737

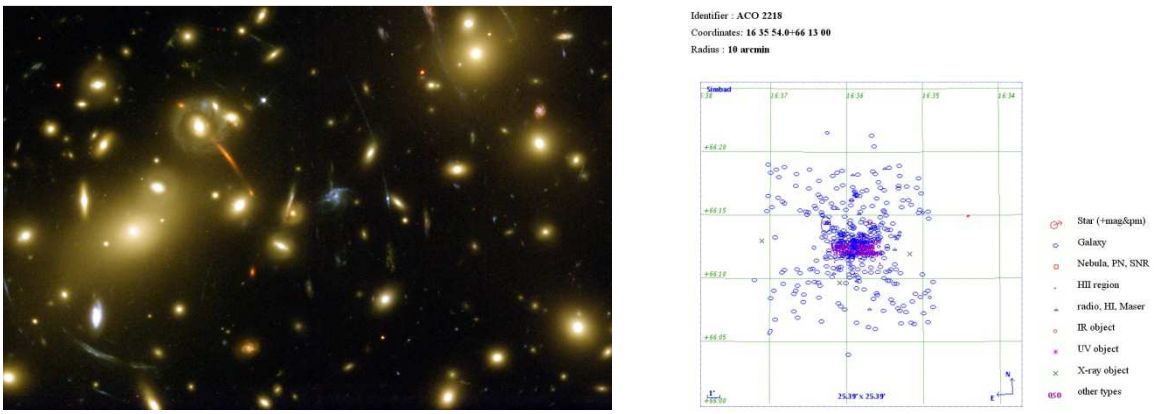


Fig. 2.1: Abell 2218 galaxy cluster. Left panel: Hubble Space Telescope (HST) image of the central region of the cluster (approx. 3 x 2 armin). Right Panel: SIMBAD10 map of the cluster, for the 20 arcmin diameter central region. Each blue dot in this map is a galaxy.

[9] Gravitational lensing is produced when the light coming from objects farther away gets bent by the gravitational effects of objects in front. A galaxy cluster with a huge mass is capable of producing such a vast gravitational field, acting as a giant lens for light of object behind it.
[10] SIMBAD is an astronomical database of objects, with cross-reference and visualization tools.

However, each individual galaxy in the cluster (a 20 arcmin diameter map is presented in the figure) can only be observed at the resolution limit given by the seeing at the time of the observations, which, for a ground-based observation, is typically not better than 0.7 arcsec. Whilst a spatial resolution of 0.1 arscec is achieved for the HST image, this is limited to a small field by the size of the available cameras and does not include spectroscopy. Better spatial resolution is required to allow resolving details of smaller galaxies over the full field, in imaging as well as spectroscopic modes.

To obtain a deep understanding of how galaxies are born, grow and evolve, it is necessary to observe their internal structures, usually requiring a spatial resolution better than the seeing limited condition. Integral field spectroscopy is one powerful technique used to obtain spatially resolved spectra of objects. If an object is extended (i.e. a galaxy), it allows spectra to be obtained for many parts of the object, enabling the spectroscopic observation of their internal structure, and the study of chemical abundances and velocity fields. A typical result of these observations is a radial velocities map, which measures the redshift of the different parts of the galaxy, with respect to the redshift of the whole body.

An example of a radial velocities map is shown in Fig. 2.2, showing that the nuclear star cluster in a nearby galaxy is rotating; this type of observation is now possible thanks to the progress in integral field spectroscopy.

The possibility of obtaining integral field spectroscopy at high spatial resolution of several targets (galaxies) in one exposure, selected from a region in the sky spanning several arcminutes, would open a vast window of unexplored science, certainly with deep implications for cosmology as well as astrophysics. Multi-object adaptive optics has been conceived to fulfil this desire.

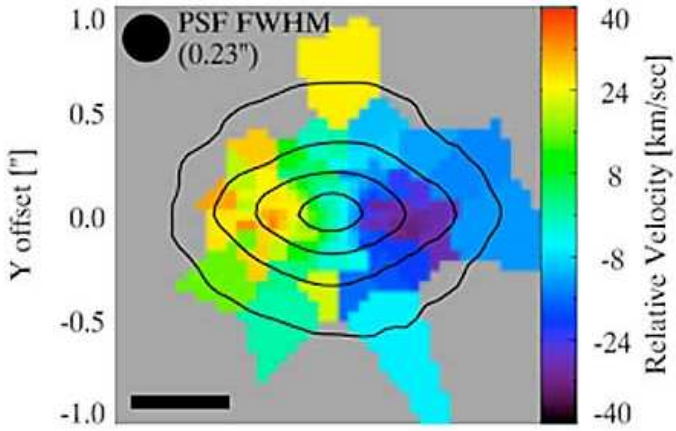


Fig. 2.2: Nuclear star cluster in NGC 4244. Blue approaches and green/orange recedes. Rotation of ~30 km/sec is clearly visible. These data were obtained at Gemini observatory, using the AO-fed 'Near Infrared Integral Spectrograph' (NIFS). Image courtesy Gemini Observatory.

2.3 The Wide-Field Problem

As seen in Chapter 1, a main limitation of adaptive optics is angular anisoplanatism. This is caused by the fact that the light from different objects in the field of a telescope go through different optical paths, encountering different turbulence and thus suffering different aberrations at any given moment.

We present a brief analysis of this problem from a purely geometrical standpoint (see Fig. 2.3). We assume a telescope with a certain radius R_0, which is observing an object on-axis (red dotted lines in the left panel of Fig. 2.3). There is a laser guide star (LGS) launched from the telescope at a certain off-axis angle FoV (Field-of-View) and at an altitude h_{LGS}. We are only considering one layer of turbulence at an altitude h_{Layer}. Having a LGS at a finite altitude implies we are in the presence of focal anisoplanatism as well. The LGS beam intersects the turbulent layer forming a meta-pupil of radius R_1.

If we make the assumption that all the turbulence is concentrated in one layer at some altitude, then it might be possible to fully correct it, if we can manage to measure it appropriately. This would ultimately depend on the interaction between metapupils at the layer. This problem is presented in the right panel of Fig. 2.3, seen at the turbulence layer plane. Due to focal anisoplanatism

(the sodium layer is not at infinity, but at about 80 km from sea level) and angular anisoplanatism (off-axis position of the LGS), the metapupils' overlap is limited and this has a direct consequence in the level of correction the AO system can achieve.

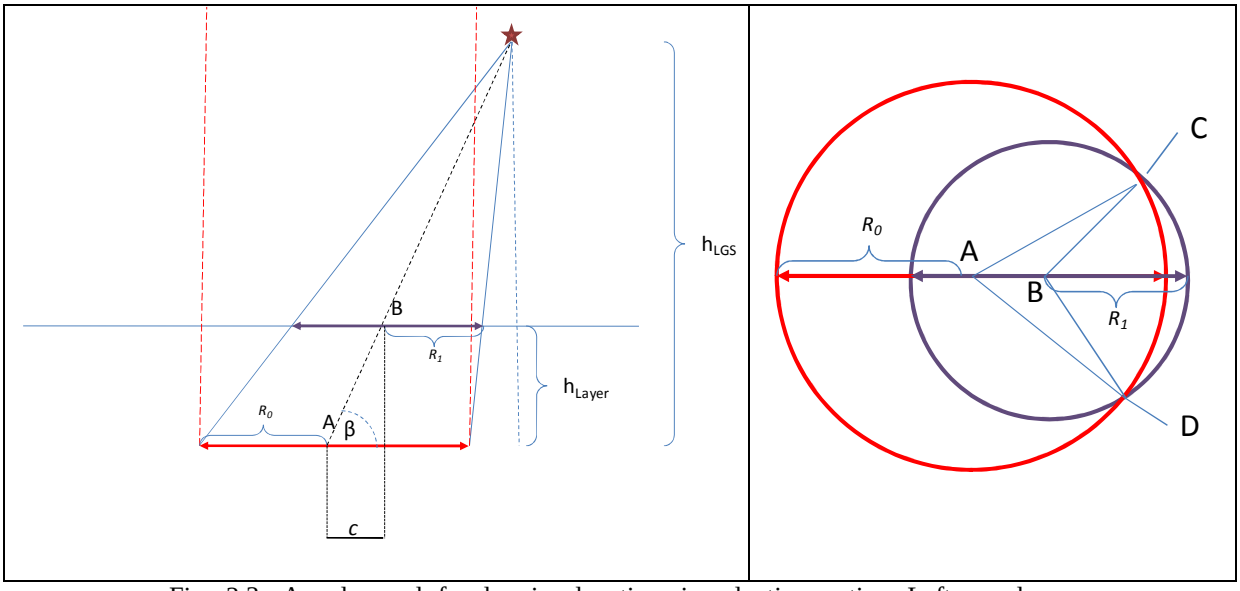


Fig. 2.3: Angular and focal anisoplanatism in adaptive optics. Left panel: science beams (in red) and a laser guide star at a certain angle (FoV) off-axis. Right panel: overlap of the meta-pupils at the altitude of the turbulence layer.

We derive in Appendix of this chapter some basic relationships between the distances in Fig. 2.3, which allows us to run some basic simulations to understand the net effect regarding metapupil overlap caused by the off-axis position of the LGS as well as the altitude of the turbulence. We performed simulations for the LGS (with the sodium layer at 80 km) and NGS (guide star at infinity) cases, see Fig. 2.4.

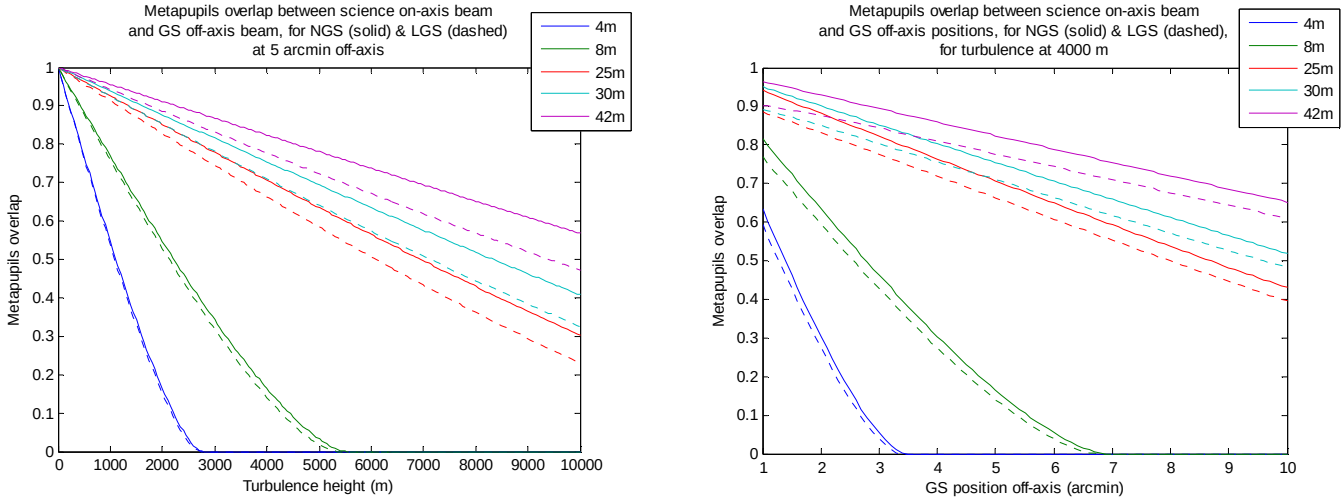


Fig. 2.4: Metapupil overlap for several situations. Left panel: fixing the guide star at 5 arcmin off-axis and varying the turbulence height. Right panel: fixing the turbulence height at 4000 meters and varying the position of the guide star. NGS results are solid lines and LGS results are dashed lines. Different color lines represent various telescope mirror diameters in meters.

The level of overlap between metapupils is an important factor in determining the final Strehl ratio an AO system can achieve[11]. Some specific implications from Fig. 2.4 are described here:

- The effect of focal anisoplanatism (the 'cone effect') is seen as the difference in position between solid and dashed lines, not is seen to be significant in current 4 to 8 meter telescopes. The cone effect also depends on wavelength[12] and may not be of significance for current telescopes in the near infrared, but it assumes real importance for extremely large telescopes (>25 m diameter).
- 4 meter telescopes have very poor metapupil overlap for any turbulence above the ground (see fast decay of blue curves in Fig. 2.4), which is consistent with these types of telescopes being good candidates to implement Ground-Layer AO (GLAO)[13].
- 8 meter telescopes have reasonably good metapupil overlap for medium height turbulence (see green curves in Fig. 2.4), which makes them candidates for a wide-field AO instrument

[11] Which depends on many other aspects, see section on MOAO Error Budget.
[12] As we have seen in Chapter 1, $\sigma_{FA}^2 = (D/d_0)^{5/3}$, with $d_0 \propto \lambda^{6/5}$.
[13] Ground-layer AO, introduced in Chapter 1, aims to correct only the turbulent layer immediately above the telescope. It is considered a 'seeing improver' only (with poor Strehl ratio) but it can provide correction over several arcminutes FoV.

such as MOAO, but there will be high altitude turbulence (i.e. at 10 km), which cannot be measured. This certainly depends on the position of the guide stars.

2.4 Multi-Object Adaptive Optics

In general, a MOAO instrument is a near infrared multi-object spectrograph with the following main sub-systems:

- A number[14] *N* of laser guide stars, launched from the telescope to a ring[15] surrounding the science field. This is strictly a telescope facility, but it is required by a MOAO instrument.
- A number *N* of wavefront sensors, measuring raw incoming turbulence from laser guide stars. Even though a system with only natural guide stars is feasible to build and use, there are only a few positions in the sky where a suitable star asterism can be found[16].
- A number M of wavefront sensors for natural guide stars, in order to determine tip and tilt.
- A pick-off opto-mechanism to observe each of the science objects. The light beam from each object is relayed to a deformable mirror. These DMs work in open-loop with respect to the WFS defined above.
- An integral field unit (IFU) to dissect the spatial components of each AO-corrected beam. An image slicer is typically selected for this application.
- A spectrograph to retrieve the spectra for each point in the AO-corrected objects. Depending on the number of objects, more than one spectrograph (working in parallel) may be required.
- A near infrared detector(s) to acquire the spectra.
- An adaptive optics computer, in charge of measuring the incoming turbulence and calculating the correction to each science target, to be applied on each DM.

A schematic diagram of a MOAO instrument is presented in Fig. 2.5.

[14] The number of LGS depends on many parameters, from required Strehl ratio to cost.
[15] Although other topologies can be used, a ring is a simple geometry to measure turbulence from a number of angles, producing metapupil overlap with the science objects at the centre of the field.
[16] CANARY phase A considers using NGS only and a few asterisms have been identified, but it is unlikely to have scientific productivity with such a configuration. See section 2.8 for a description of CANARY.

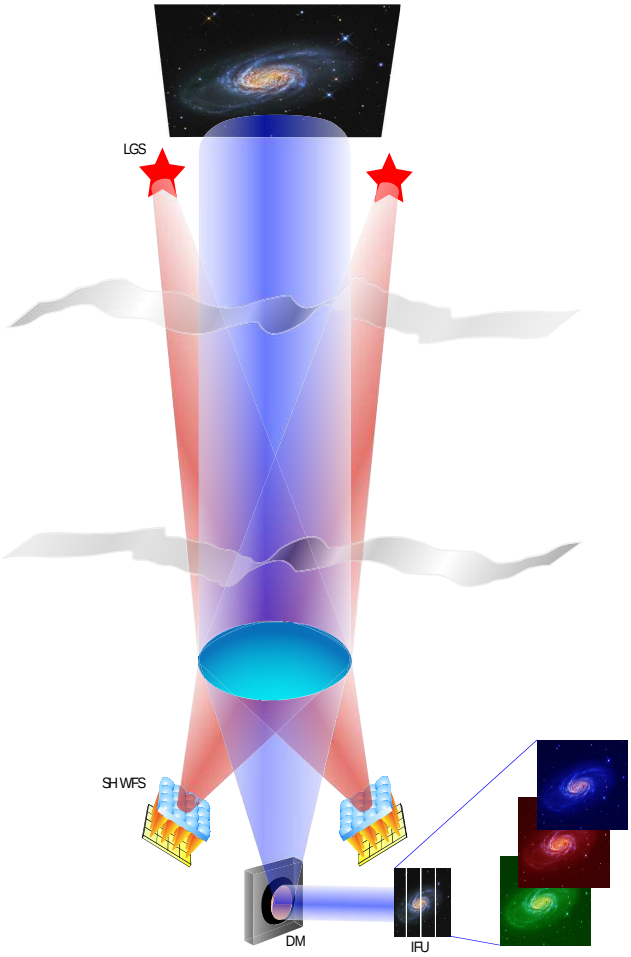


Fig. 2.5: Cartoon of a multi-object adaptive optics instrument. Several LGS are launched to the periphery of the field, to measure the incoming raw turbulence with Shack-Hartmann WFSs. The science beam (from a galaxy) is relayed to a deformable mirror, which corrects the aberrations caused by the turbulence, sending the corrected beam to an integral field unit (IFU), which slices the spatial image into its component wavelengths for scientific analysis.

2.5 Implications of Open-Loop Operation

Operation in open-loop is not new in adaptive optics: Hardy (1998) for example, includes some suggestions in running an AO system in open-loop; nevertheless, all AO systems found in 4 to 10 m class telescopes nowadays run in closed-loop. MOAO is the first instance of an astronomical AO system where open-loop operation is justified and studied in detail. We can identify the following areas of particular interest when designing and building an AO system for open-loop operation.

2.5.1 No optical feedback

In classical AO systems in closed-loop, a wavefront sensor measures the action of the deformable mirror, because it is placed "downstream" with respect to the deformable mirror in the optical path, so it measures the residual error and feeds it back to the system. See top panel of Fig. 2.6 for a comparison diagram.

Closed-loop operation is a basic feature in any control application, from a temperature controller to a nuclear power station. For instance, if the controller includes an integral[17] term in its mathematical formulation, then the steady-state residual error will be null[18].

In MOAO such an optical feedback does not exist, therefore, the system cannot self-recover when the system incurs errors after applying the wavefront correction. See bottom diagram in Fig. 2.6.

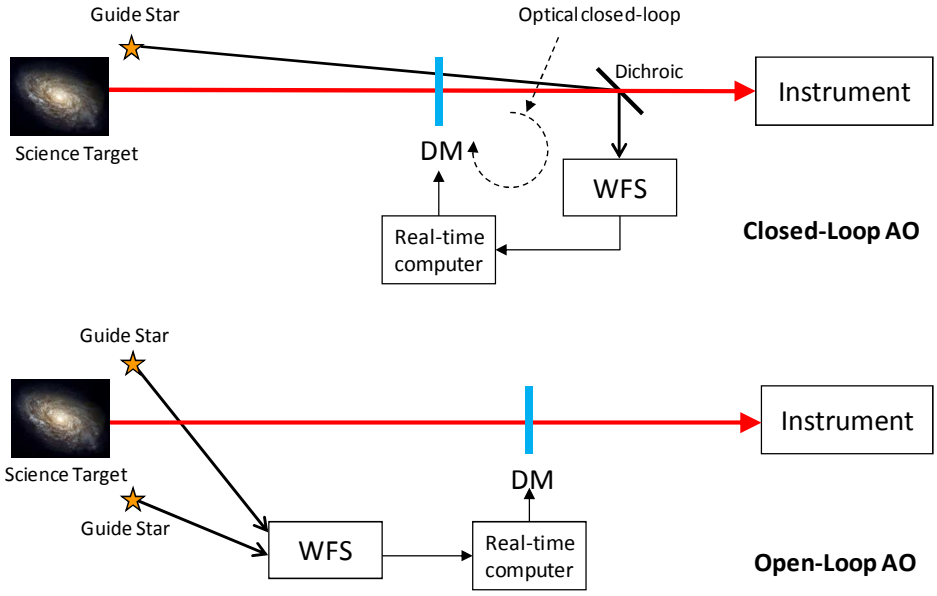


Fig. 2.6: Comparison diagrams for closed-loop and open-loop operation in AO. Top panel: closed-loop operation, showing how the wavefront sensor (WFS) closes the loop with respect to the deformable mirror (DM). Bottom panel: open-loop operation, where the WSF paths are different and unrelated with respect to the DM path.

[17] A typical controller is of the *PID* type, which stands for *Proportional* (a constant gain), *Integral* (an integrator) and *Derivative* (a differentiator). When the setpoint for the controller (its desired value) is constant and there is a non-zero integral term in the controller, the steady-state error of the system in closed-loop is zero.

[18] This is true when the 'set point' of the system is constant, which is not the case of an AO system, where the turbulence is permanently changing, but in principle, a very fast system in closed-loop could achieve null error while the turbulence remains static (in a time frame of milliseconds). This in practice is not possible, given that the number of photons to measure the turbulence is very limited.

2.5.2 Wavefront sensor dynamic range

In closed-loop systems, the WFS(s) measure the residual turbulence after the deformable mirror has corrected the wavefront. Therefore, the WFS(s) measure a null or flat wavefront on average. This is not demanding in terms of the dynamic range of the sensor, but it depends on the system having "locked-on", i.e. having achieved steady-state, closed-loop operation (Hardy 1998). WFSs for MOAO need more dynamic range, since they measure raw incoming turbulence in open-loop, as presented in the bottom panel of Fig. 2.6. The dynamic range required, expressed as the size of the pixel array to measure the position of each centroid in a SH-WFS, would ultimately be defined by the strength of the turbulence[19].

Wavefront sensors for AO systems working in closed-loop systems sometimes use the "Quad-Cell" or "Bicell" (Hardy 1998), a *2 x 2* pixels arrangement to determine the centroid position of each spot on a Shack-Hartmann[20] wavefront sensor. For greater accuracy there is the *4 x 4* pixels array for each SH subaperture, about which Hardy comments "… it is doubtful whether arrays larger than *4 x 4* can be justified." (Hardy's section 5.3.1). Facing the need of operation in open-loop, it becomes clear that larger arrays would be required. It is not in the scope of this section to establish the appropriate size for subapertures in MOAO; however, we do perform a related analysis in Chapter 4, for using a SH-WFS to measure the shape of a DM working in a MOAO system. Measuring the shape of the DM requires a similar dynamic range as which will be needed in order to measure the raw incoming turbulence that will request the deformation of the DM. From that analysis, it seems that 8 to 12 pixels per subaperture side are adequate for SH-WFS working in open-loop. Amongst other parameters, the accuracy of the sensor depends on the ability of the centroiding algorithm to determine the precise position of the centroid, working in the presence of photon and readout noise, and inherent noise sources associated with light and detectors at optical wavelengths.

Given the wider FoV of the WFS in open-loop, any non-linearities within that FoV would directly increase the wavefront error associated with the WFS. Furthermore, in the case of ELTs, there

[19] There can be other effects which affect the WFS dynamic range, for example wind shake.
[20] We presented the Shack-Hartmann wavefront sensor in Chapter 1, describing it in detail in Chapter 4.

is the 'spot elongation' in laser guide star wavefront sensors, caused by the sodium layer, which has a non-negligible thickness when seen by the off-axis subapertures. Elongated spots are more difficult to track when centroiding and even more so for MOAO systems, given the dynamic range is so stringent.

These effects are the subject of active research (Gilles 2006, Thomas 2006) with the goal of understanding them in depth and experimenting with techniques to reduce their impact in the MOAO error budget.

2.5.3 Deformable mirror model

Shaping the deformable mirror to flatten the wavefront is achieved automatically in closed-loop AO systems, through the inherent effect of the WFS running behind the DM (see top panel of Fig. 2.6) which when locked on produces a null wavefront. The 'Real-Time Computer (RTC)' (see Fig. 2.6) or 'AO Computer' will produce the appropriate values to command the DM actuators to achieve this effect. This is not the case for open-loop AO systems, where a tomographic reconstructor (see next section) will calculate the required shape for the DM in certain units, but it does not have the ability to compute the DM actuator voltages to achieve that shape. This is the purpose of a deformable mirror model, which acts as the interface between the tomographic reconstructor and the electronic drivers commanding the DM actuators. This is particularly relevant when the final shape of the DM facesheet is not found to be a linear combination of the individual DM actuator influence functions. We performed experiments with two deformable mirrors with such behaviour. Our results are presented in Chapter 3, as part of the experimental work of this project.

A basic issue with a DM model is the unit[21] defined to describe the shape of the DM surface. This unit must be consistent with the one used by the tomographic reconstructor. For example, the tomography can be computed in the equivalent of "centroiding pixels" from the centroiding process occurring in each wavefront sensor. Alternatively, it is possible to calibrate the reconstructor to request a DM shape described in nanometres, so a DM model can be calibrated using an optical

[21] For example, the displacement of the DM facesheet with respect to the body of the mirror, quantified in nanometer (*nm*) or Angstrom (*Å*).

interferometer. In our experiments, we used an optical interferometer to measure the DM surface, so our results are in nanometres. The step of finding common units between the reconstructor and DM model is related to the calibration issues of a MOAO system, described next.

In summary, it becomes clear that the accuracy of the DM model is a critical component of the MOAO error budget.

2.5.4 Calibration issues

The final goal of any AO system is to increase the spatial resolution of the science images, acquired by the instrument attached to the AO system. There must be a calibration step to make sure the AO system is correcting the turbulence adequately. This is particularly relevant for what is usually called "non-common path errors". In any AO system, there are different paths taken by the light, typically one beam is diverted towards the WFS and a second beam goes to the science instrument. The AO system is able to correct static aberrations in the optical path[22], as long as they can be measured and calibrated out. An example of a non-common path calibration is focusing the science instrument. If a reference source is imaged by the science instrument with the AO system not running, it is possible to command the DM to compensate for the focus term static aberration. This figure on the DM must be included when running the AO system in the presence of turbulence. This step is true for any AO system, irrespective of open or closed-loop operation, but in open-loop systems the non-common path is larger, since the WFS light is picked-off at the focal plane of the telescope and not after the DM (which is typically at the conjugated plane of the entrance pupil of the telescope) as in closed-loop systems. Furthermore, a MOAO system must have a mechanism to pick-off the science targets within the FoV, which may imply additional field rotations of the science images with respect to the DMs as well as WFS, thus a proper calibration procedure is essential to achieve a satisfactory level of correction.

A somewhat related area is the 'connection' between wavefront sensors and deformable mirrors. In classical closed-loop AO, there is the "interaction matrix", which is built experimentally

[22] An AO system can correct static aberrations as long as the required stroke is within the DM capabilities.

by poking each actuator of the DM and recording its effect in the wavefront seen by the WFS, in the presence of no turbulence. The interaction matrix takes into account any optical effects within the AO system (say pupil misregistrations) and it is an effective way to quantify the real effect of each actuator on the wavefront. In that sense, the interaction matrix is part of the system calibration. The interaction matrix must be pseudo-inverted to produce the 'control matrix', which serves for closing the loop in the most basic closed-loop implementation: the vector of wavefront corrections is calculated as the matrix product between the control matrix and the vector of wavefront measurements[23].

In MOAO, it is not straight forward to build an interaction matrix, since the WFS does not look at the DM. One can synthesise a theoretical interaction matrix based on knowing the influence functions of each actuator in the DM and the magnification of the system; however any instrumental defect (such as pupil misregistrations between WFS and DM) will directly degrade the correction. This problem is present in current MOAO experiments and it is challenging to solve it without adding a WFS in closed-loop. In fact, the MOAO demonstrator CANARY will implement a "Truth Sensor", which is another WFS running in closed-loop with respect to the DM. The Truth Sensor's first objective is to measure the level of correction achieved, but it can be used to calibrate the system, as done by Vidal (2010) in the "Learn & Apply" algorithm, which is a technique to implement tomographic reconstruction.

2.6 Tomographic Reconstruction for MOAO

Tomography is a mathematical formulation that allows the reconstruction of a volume from cross-sectional images taken at different angles. It was primarily developed for medical imaging, such as in X-ray computed tomography (Kak 1988). Its application to astronomical adaptive optics was first introduced when multi-conjugated AO systems started to be developed (Tallon 1990). In that case, an estimate of the turbulent volume was required, using several samples of the turbulence from guide stars at different positions in the FoV of the telescope. There are differences between the

[23] This basic implementation assumes static behaviour of the turbulence. A controller can be included to account for dynamic behaviour of the atmosphere and the DM.

medical tomography, where it is possible for the 'patient' to be static in time and imaged at a plethora of angles (for example, rotating the instrument around the patient), and atmospheric tomography, where the selection of angles is limited to tens of arcminutes at the most, the light sources are faint and the turbulent phenomena are random and dynamic. Nevertheless, the principle of reconstructing the volume from a number of penetrating measurements can still be applied.

In purely geometrical terms, the tomographic reconstruction consists of estimating the phase of a science object located somewhere in the FoV of the telescope, using the phase measurements from guide stars located somewhere else in the field, ideally around the science object. Because of the considerations described in section 2.3, due to angular anisoplanatism (and focal anisoplanatism if using laser guide stars), there is de-correlation between the turbulence seen by guide stars with respect to the science object. Finding the best estimate while projecting the measurements to turbulent layers at different heights in the atmosphere is the basis of what is called tomographic reconstruction in adaptive optics.

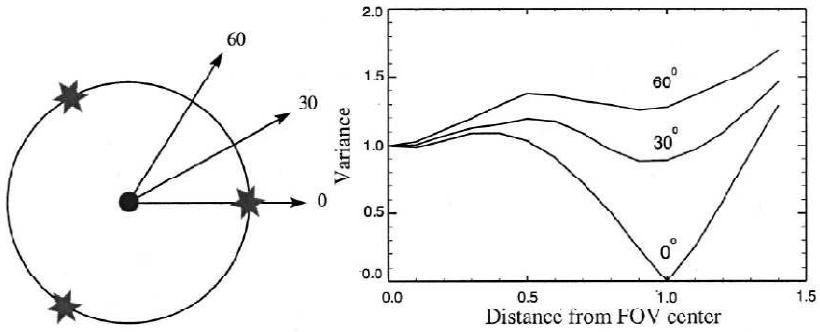


Fig. 2.8: Simulation results of the residual phase variance for a science object moving from on-axis to off-axis positions. Left panel: diagram of the guide stars and possible positions for the science object (in rays at different angles). Right panel: residual phase variance at different off-axis positions. Distance is normalised in units of guide-star angle (the guide star is at 1.0). Reproduced from Tokovinin (2001).

Let us consider a simple case of a ring of 3 guide stars and a science object somewhere inside the ring (see Fig. 2.8). If the object lay at the centre of this field, the 'contribution' of each guide star to the phase estimate for the science object would be similar, therefore a tomographic reconstructor should weigh their contributions equally. As soon as the science object position is

displaced towards one of the guide stars, this GS would be better placed for sampling the turbulence as seen by the science object and thus the tomographic reconstructor should weight its measurements more than the rest. This exercise is presented by A. Tokovinin in one of his papers on tomography (Tokovinin 2001), reproduced in Fig. 2.8.

It is interesting to note in Fig. 2.8 that the phase residual for an on-axis object (equal to *1.0* on the plot) increases when the science object position departs from the on-axis position until reaching about half the angle at which the guide stars are placed (the distance is normalized to the guide star angular position). It then drops when the object gets closer to the guide star. This is a consequence of partial metapupil overlap (see Fig. 2.3), even though the science object is approaching the guide stars.

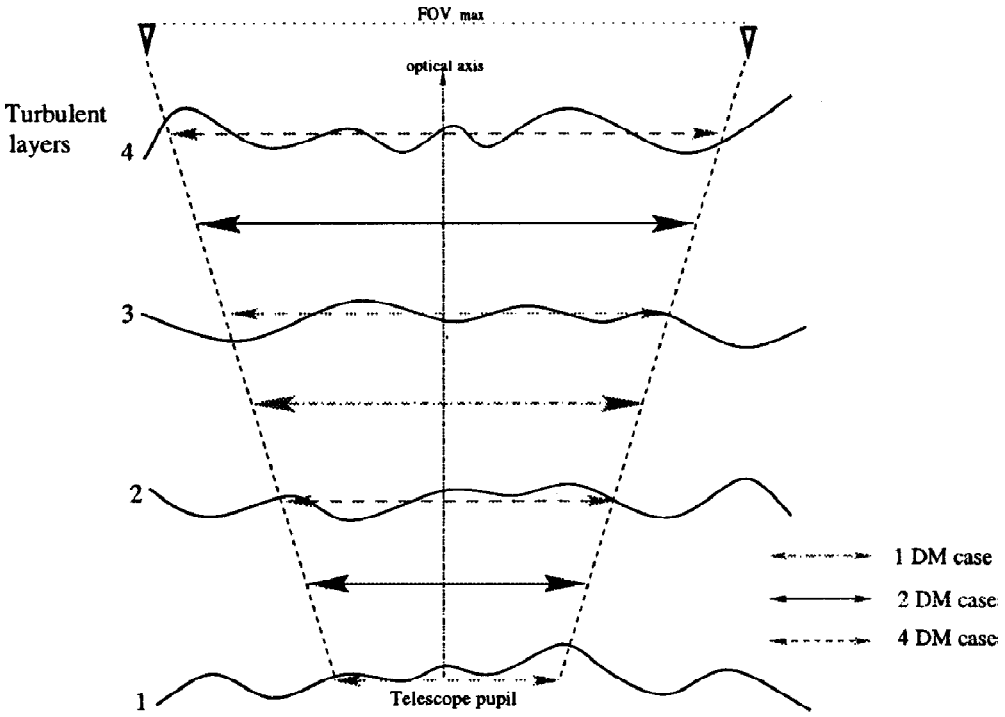


Fig. 2.9: Schematic diagram of the tomographic problem, as represented by Fusco et al (2001). Various turbulent layers, from Ground layer (#1) to high-altitude layer (#4) are present in the atmosphere, but a fewer number of DMs will be conjugated to certain altitudes. Possible conjugations when having from one to four DMs are presented. Reproduced from Fusco (2001).

The optimal tomographic reconstruction is presented by Fusco for MCAO (2001), where the general case of having a bigger number of turbulent layers than deformable mirrors is analysed (see Fig. 2.9). This work is very important in the development of tomography for AO, since the whole

concept of reconstructing the volumetric turbulence is turned into an optimization problem of finding the best combination of DM corrections that would minimize the resulting phase error. In that sense, tomography for AO differs from tomography applied to other fields, since one is not interested in finding the position and strengths of the turbulent layers, but finding the optimal solution to remove them.

Assemat *et al* use Fusco's approach for MOAO in his FALCON[24] paper (2007). Neichel (2008) and Gavel (2004) describe tomography for AO in two stages, which we use here to further explain the tomographic reconstruction procedure that shall be implemented in a MOAO system.

Under linear assumptions, one is interested in finding a vector φ_{DM} of DM(s) corrections, computed from a vector of wavefront measurements ϕ_{WFS}. Mathematically:

$$\varphi_{DM} = W\phi_{WFS} \quad (2.1).$$

Where $W_{2N_{gs}\times N_{DM}}$ is the phase volume reconstructor matrix that converts $2N_{gs}$ wavefront measurements[25] into N_{DM} correction phases. Fusco (2001) has shown that W can be split into two independent matrices:

$$W = P_{opt}W_{tomo} \quad (2.2).$$

The first stage is finding W_{tomo}, the tomographic reconstructor, which has dimensions = $2N_{gs} \times N_L$, with N_L the number of reconstructed turbulent layers. The second stage is to project the tomographic solution onto the DMs, computing the projection matrix P_{opt}. W_{tomo} only depends on the guide star configuration and the atmospheric conditions (recall Fig. 2.3). P_{opt} is a geometric operation that depends on the FoV to be corrected as well as the number and positions of the DMs with respect to the reconstructed layers.

The tomographic reconstructor W_{tomo} can be found with several methods, depending on the optimization criteria selected. The most used methods are *Minimum Mean Square Error* (MMSE), *Least Square Estimator* (LSE) and *Truncated LSE* (TLSE). MMSE finds the estimate of the turbulence in each reconstructed layer that minimizes the residual variance. LSE finds the best

[24] FALCON was the first MOAO instrument proposed; see section 2.8 for a survey of MOAO instruments.
[25] It is 2 times because of the *x* and *y* slope measurements in SH-WFS.

estimate of the turbulence that fits the measurements. TLSE is similar to LSE but some singular values are truncated to avoid non-invertible conditions, for example when there are unseen modes between the SH-WFS and DM. It is beyond the scope of this section to describe these techniques in detail; Neichel (2008) presents a good summary of the optimal solutions for the three methods. Similarly, Neichel presents the optimal projection matrix P_{opt}.

Vidal (2010) has proposed a new approach to tomography, the 'Learn and Apply' method (Vidal 2010). It is the first such methodology particularly devised for MOAO, which proposes a rather practical and elegant implementation, described qualitatively in the following paragraphs:

An additional 'Central' WFS is placed in the FoV after the DM (i.e. in closed-loop). This extra WFS is used for calibration purposes. In particular, it is possible to build an interaction matrix with the DM, as in standard closed-loop systems, to know the exact effect that the DM correction will produce on the science image.

The Central WFS, along with the off-axis WFSs running in open-loop, are used to observe synthetic aberrations produced at the telescope pupil. These aberrations can be generated by the deformable secondary mirror for example, as in the 42 metre European ELT (E-ELT) project[26]. This common information allows the system to take into account all possible misregistrations, rotations and scale factors among the WFSs. A linear relationship is built to describe the on-axis Central WFS measurements using the off-axis WFS measurements.

Real turbulence is recorded by all WFSs (including the Central one), which feed, along with the relationship found in (2), an algorithm that produces a linear tomographic reconstructor, based on the real measurements from the atmosphere and not on *apriori* information, such as historical turbulence profiles. This data are taken over a few seconds in on-sky 'calibration' operation. This is the 'Learn' step of the method.

The data recorded are used to fit a model, to be used later with real observations, when the Central WFS cannot be placed in the field. This is the 'Apply' step of the method.

[26] The E-ELT considers an optical design of the telescope with 5 mirrors, two of which will be adaptive. One of them will correct global tip and tilt, while the other will correct low-order aberrations, conjugated approximately to the Ground layer.

Laboratory experiments performed by the authors using this method showed the best results in terms of wavefront error, with respect to other more classical methods for tomography described here. The final on-sky tests will come with the deployment of CANARY later this year. See the section on MOAO instruments in this chapter for more details about this project.

2.7 MOAO Error Budget

The Error Budget in AO is a very useful tool for performance estimation when designing a system. It consists of an evaluation of all error sources that would degrade the final level of correction an AO system is able to achieve. These error sources are quantified using analytic expressions and/or the result of realistic simulations. Typical error budget components are, for example, the DM fitting error[27] and the wavefront measurement error[28], but there are many more. Assuming the error terms in the budget are uncorrelated, e.g. they are independent error sources and modifying one does not affect others, their contributions can be added up in quadrature; that is, the wavefront error variances can be summed up directly. The expected Strehl ratio of the AO system can be calculated using the Marechal approximation[29] from the total wavefront error in radians. It is not the purpose of this section to give a detailed overview of this tool, but to introduce the error budget analysis implemented for CANARY.

2.7.1 Performance Objective

This is a reference which the performance should be compared against. Some examples of performance objectives are:

- Reach diffraction limit at one line-of-sight
- Maintain a stable Strehl ratio over the field
- Maximize the FoV, given a lower limit for the Strehl ratio

In the case of CANARY, as it is a MOAO demonstrator with one channel in open-loop, the performance objective for the phase A of the project was to achieve the same performance as a single-

[27] This is produced by the limited number of actuators in a deformable mirror, when trying to correct an aberrated wavefront which cannot be fitted perfectly.
[28] For instance, a Shack-Hartmann WFS would have error sources in the detector noise.
[29] The Marechal approximation is $S = e^{-\sigma^2}$, with σ the total wavefront error in radians.

conjugated AO system running in closed-loop with a bright natural guide star. From theoretical analysis, it is easy to confirm that this baseline is not possible to reach, because the tomographic reconstructor and the DM model are extra error terms in an open-loop model, which inevitably contribute non-negligible error terms. Therefore the comparison is implemented removing these two terms.

2.7.2 Error budget terms in MOAO

Table 2.2 presents a list of components of the error budget developed for CANARY, used here as an example of an MOAO error budget.

We describe the main terms that drive the error budget and how they were obtained:

- WFS centroiding (item 1 with 156 nm): This error comes about when measuring the turbulence in open-loop, where the SH spots can be anywhere in the subaperture. This error was calculated from Monte Carlo simulations, similar to the ones implemented in Chapter 4 for the Figure Sensor development, where a subaperture collecting photons is modelled. The effects of photon and readout noise, besides non-linearities because of the large dynamic range required in the SH subapertures, are the main components of the wavefront error obtained for this item.
- Tomographic error (item 6 with 551 nm): This large error is calculated from Monte Carlo simulations, comparing a classical single-conjugated AO system and a MOAO system under the same conditions. This allows isolating the tomographic error from other common errors (such as DM fitting error for instance). The large number found is a worst case scenario, given the parameters used in the simulations: a wide separation for the guide stars (at 1.25 arcmin from the science target); the magnitude of the guide stars ($R = 14$), the turbulence conditions ($r_0 = 5\ cm$) and a non-optimal tomographic reconstructor

The rest of the terms in the budget have been estimated from theory or simulations, but none of them contributes to the total budget of 615 nm as the tomographic error and WFS open-loop

centroiding do. In fact, if we only consider an error budget with these two terms, the total wavefront error would be 573 nm, close to the 615 nm calculated.

This section does not pretend to investigate the details behind each of the terms of this error budget, but only presents the main considerations that are being studied for one of the first MOAO systems, as CANARY is.

Table 2.2: CANARY Error Budget.

Item	**Source of Error**	**Error Term**	**Wavefront error RMS (nm)**
1	WFS	Centroiding in open-loop	156
2	Tomographic reconstructor	Calibration of interaction matrix	30
3		Calibration of sensor reference slopes	10
4		Calibration of truth sensor reference slopes	10
5		Calibration of identification matrices	60
6		Tomographic error (including effect of asterism)	551
7	Adaptive component errors	DM fitting	150
8		Open-loop DM error	48
9		Open-loop Tip-tilt error	26
10	Real-time computer	Latency	17
11		Update rate	49
12		Bandwidth	100
13	Optical aberrations	Field aberrations	42
14		Static non-common path	30
15		Time-varying non-common path	0
16	Errors of conjugations, shifts or aberrations	DM misconjugation	30
17		WFS misconjugation	30
18		DM-WFS shift	30
19		Telescope –AO bench shift	50
20	TOTAL WAVEFRONT ERROR		615

2.8 Survey of MOAO projects as of 2010

2.8.1 FALCON

FALCON was a second-generation instrument proposed for an 8 meter telescope at the VLT. Its scientific objective was to observe galaxies at high redshift in a 10 x 10 arcmin FoV, therefore MOAO was proposed to achieve this goal. MOAO had been originally proposed by F. Hammer in 2001 and the FALCON proposal was co-authored by Hammer. In the most detailed paper about FALCON, Assemat (2007) proposed to develop miniature AO systems for the MOAO channels as well as WFSs, taking advantage of micro-DMs and small optics. See Fig. 2.10 for a concept for each MOAO channel, presented in Hammer (2004).

FALCON relies on having very sensitive WFSs in order to use natural guide stars only, which can be as faint as *R=17* magnitude. Numerical simulations presented by Assemat showed that it is possible to reach a resolution better than 0.25 arcsec in the near infrared *J* and *H* bands[30] with the MOAO system, for 95 % of the objects (taken from a real field on the sky). FALCON was never built for the VLT, but it pushed the interest in MOAO to further studies, which then became the other projects described in this section.

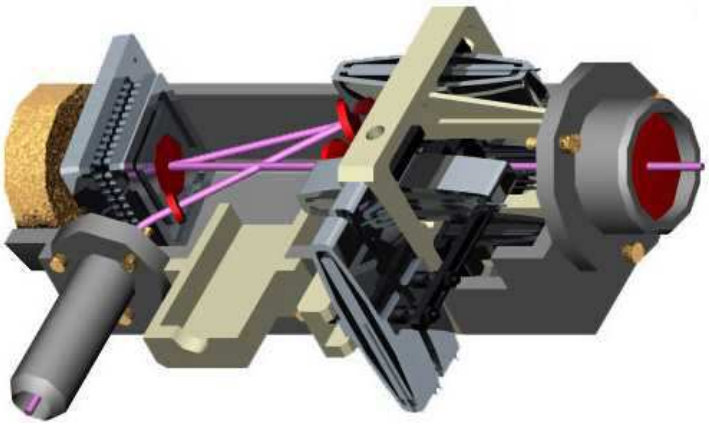


Fig. 2.10: Opto-mechanical concept for the MOAO channels, taken from Hammer et al (2004). The inner red spot in the model represents the DM, with the optical path shown in pink. Reproduced from Hammer (2004).

[30] *J* band is centred at 1.2 µm; *H* band is centred at 1.65 µm.

2.8.2 CANARY

CANARY is an on-sky MOAO demonstrator under construction by our group in *Durham* and our collaborators from *Paris Observatory* and the *UK Astronomy Technology Centre*, to be installed at the 4.2 meter *William Hershel Telescope* in the Canary Islands. It will test MOAO and tomographic reconstruction in open-loop on the sky for the first time in 2010. CANARY incorporates one MOAO channel. It has three development phases.

- *Phase A* is a purely natural guide stars system, with 3 WFS in open-loop and a low-order DM (7 x 7 subapertures, 8 x 8 actuators). See Fig. 2.11 for the optical design of this phase. Several NGS asterisms have been identified from star catalogues and will be used for on-sky tests in this purely technical phase.
- *Phase B* adds an asterism of 4 laser guide stars, produced by low altitude Rayleigh scattering from a laser, which is used in conjunction with a diffractive optical element[31] to form the asterism. The WFSs will be pointed to the LGSs.
- *Phase C* is like phase *B* in terms of LGS, but implements an E-ELT like configuration in terms of deformable mirrors. The low-order DM of phases *A* and *B* is now in closed-loop (conjugated to the ground) and a high-order DM (32 x 32 actuators) works in open-loop. This is one of the DMs that we characterize in Chapter 3.

[31] A diffractive optical element (DOE) is a passive element that operates by means of interference and diffraction to produce arbitrary distributions of light.

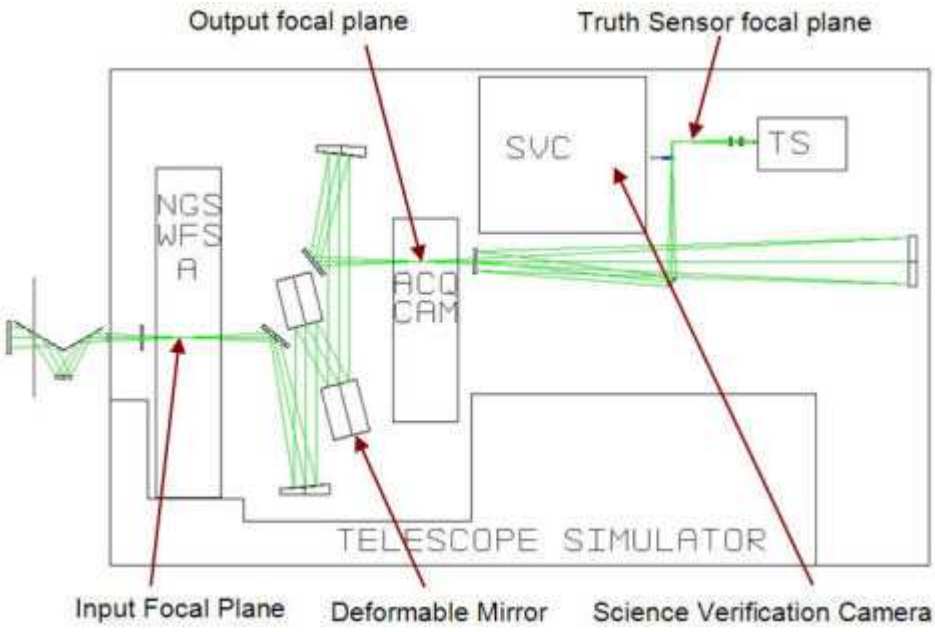


Fig. 2.11: Optical design for CANARY Phase A. The light from the telescopes enters the instrument at the left, after passing through the de-rotator. The block of NGS WFS operating in open-loop can be seen first, followed by the DM. Image courtesy T. Morris.

This project and in particular Phase C is a "pathfinder" to EAGLE, one of the instrument concepts under study for the E-ELT. The E-ELT provides a mirror for closed-loop correction (named 'M4' in the E-ELT optical design), while the instrument incorporates the MOAO channels with high-order DMs in open-loop in each one.

2.8.3 Victoria Open-Loop Testbed

The Victoria Open-Loop Testbed (VOLT) is a laboratory experiment implemented at a 1.2 meter telescope in Canada (Andersen 2008). It consisted of a single-channel, on axis, open-loop system. A photograph of the VOLT optical bench is presented in Fig. 2.12. It was tested on-sky in May 2008, achieving open-loop correction on *Arcturus* (*R=0.3*). The experiment had some difficulties in keeping the WFS noise low, and the observations were done in the presence of bad seeing of 2.5" FWHM (this is r_0 = 4 cm at 500 nm). In spite of all this, VOLT managed to improve the FWHM to 0.5" (see Fig. 2.13), with an estimated wavefront error of 370 nm rms.

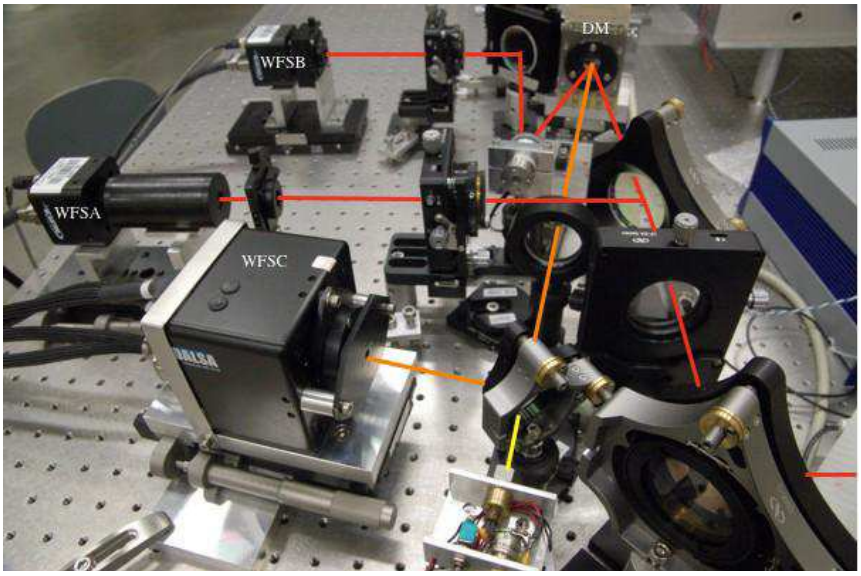


Fig. 2.12: VOLT optical bench, showing the different light paths. The light from the telescope arrives at the telescope from the right (red line), where a beam splitter diverts a part of the light to the WFS (WFSA in the image). The rest goes to the DM, after which it goes to the WFS in closed-loop (WFSB, a "truth" sensor to evaluate performance of the system). The paths in yellow and orange show the Figure Sensor WFS (WFSC), which includes a dedicated light source for illuminating the DM. Reproduced from Andersen (2008).

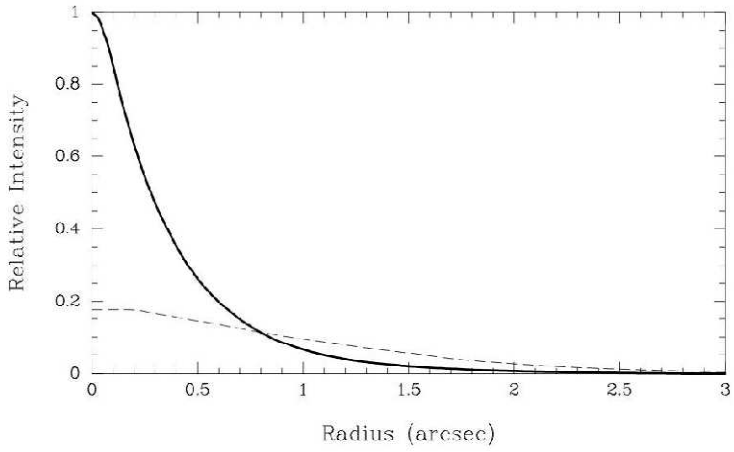


Fig. 2.13: Radial profiles of the two images of Arcturus obtained with VOLT, with open-loop AO system off (dashed) and on (solid). Reproduced from Andersen (2008).

2.8.4 RAVEN

RAVEN is a MOAO visitor[32] instrument for the Subaru 8 meter telescope, under construction by a consortium composed of Canadian and Japanese institutions. It will use only natural guide stars for wavefront sensing and it will have two MOAO channels. RAVEN is defined as a MOAO pathfinder (as CANARY) with two main goals: "…first, Raven has to demonstrate that MOAO technical challenges can be solved and implemented reliably for routine on–sky observations. Secondly, Raven must demonstrate that reliable science can be delivered with multiplexed AO systems" (Conan 2010). As an AO system, RAVEN will only have the MOAO channels, but the science spectra will be obtained with IRCS, an existing near infrared spectrograph at SUBARU. Table 2.3 presents the baseline requirements for RAVEN. See Fig 2.14 for the current optical design of RAVEN.

Table 2.3: RAVEN baseline requirements, from Conan *et al* (2010).

Parameter	**Requirement**
FoV	2 arcmin
Science FoV	4 arcsec per channel
# of science channels	2
# of WFS	3
Delivered Encircled Energy[33]	> 30 % in a 150 milliarcsec slit (H band, median r_0)
Throughput	> 0.32 in H band
Wavelength coverage	0.9 – 2.5 μm

[32] According to the project, its status is "in-carry instrument or PI instrument".
[33] Encircled Energy is a metric used in spectroscopy to quantify the amount of light that falls within the spectroscopic sampling element. It is analogous to the PSF size or Strehl ratio for imaging in AO.

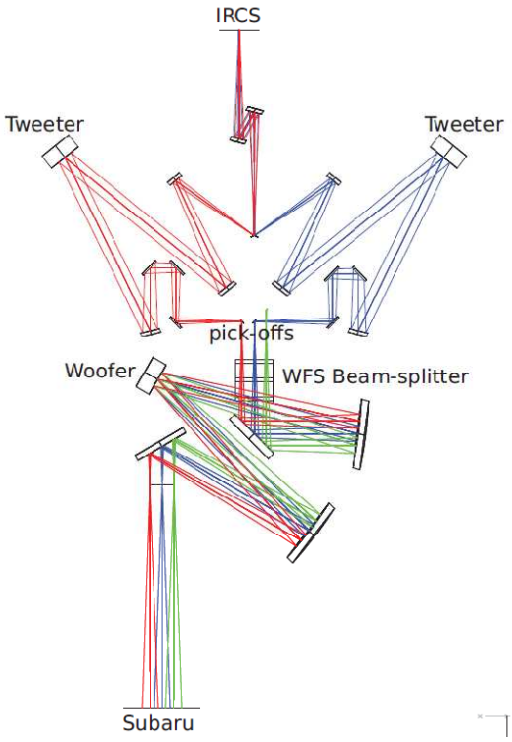


Fig. 2.14: RAVEN conceptual diagram. The 'Woofer' is a low-order DM working in closed-loop, while the two 'Tweeters' are high-order DMs working in open-loop. Reproduced from Conan (2010).

2.9 References

Andersen, D., Fischer, M., Conan, R., Fletcher, M., Veran, J., "VOLT: The Victoria Open-loop Testbed", SPIE Proceedings 7015, pp. 70150H, (2008)

Assémat, F., Gendron, E., Hammer, F., "The FALCON concept: multi-object adaptive optics and atmospheric tomography for integral field spectroscopy - principles and performance on an 8-m telescope", MNRAS, 376, 1, pp. 287-312, (2007)

Conan, R., Bradley, C., Lardière, O., "Raven, a harbinger of Multi–Object Adaptive Optics based instruments at the Subaru telescope", SPIE Proceedings 7736, pp. 77360T, (2010)

Fusco, T., Conan, J., Rousset, G., Mugnier, L. and Michau, V., "Optimal wave-front reconstruction strategies for multiconjugate adaptive optics", JOSA-A, 18, 10, pp. 2527-2538, (2001)

Gavel, D., "Tomography for multiconjugate adaptive optics systems using laser guide stars" SPIE Proceedings 5490, pp. 1356-1373, (2004)

Gilles, L. and Ellerbroek, B., “Shack – Hartmann wavefront sensing with elongated sodium laser beacons: centroiding versus matched filtering”, Applied Optics, 45, 25, pp. 6568-6576, (2006)

Hammer, F., Sayède, F., Gendron, E., *et al*, “The FALCON concept: multi-object spectroscopy combined with MCAO in near-IR”, ESO Workshop Proceedings: Scientific Drivers for ESO Future VLT/VLTI Instrumentation, arXiv: astro-ph/0109289v1, (2001)

Hammer, F., Puech, M., Assemat, F. *et al*, “FALCON: a concept to extend adaptive optics corrections to cosmological fields”, SPIE Proceedings 5382, pp. 727-736, (2004)

Hardy, J. “Adaptive Optics for Astronomical Telescopes” Oxford University Press, (1998)

Kak, A., and Slaney, M., “Principles of Computerized Tomographic Imaging”, IEEE Press (1988)

Neichel, B., Fusco, T., Conan, J., “Tomographic reconstruction for wide-field adaptive optics systems: Fourier domain analysis and fundamental limitations”, JOSA-A, 26, pp. 219-235, (2008)

Tallon, M. and Foy, R., “Adaptive telescope with laser probe: isoplanatism and cone effect”, A&A 235, 1-2, pp. 549-557, (1990)

Thomas, S., Fusco, T., Tokovinin, A., Nicolle, M., Michau, V and Rousset, G., “Comparison of centroid computation algorithms in a Shack – Hartmann sensor”, MNRAS 371, 1, pp. 323-336 (2006)

Tokovinin, A., and Viard, E., “Limiting precision of tomographic phase estimation”, JOSA-A 18, 4, pp- 873-882 (2001)

Vidal, F., Gendron, E., and Rousset, G., "Tomography approach for multi-object adaptive optics," JOSA-A, 27, 11, pp. A253-A264 (2010)

2.10 Appendix: Geometrical relationships in Fig. 2.3.

Circle $A = (0,0)$, radius R_0

Circle $B = (d, 0)$, radius R_1

$$R_1 = R_0 \frac{h_{LGS} - h_{Layer}}{h_{LGS}}$$

$$d = \frac{h_{Layer}}{\tan(\beta)}$$

The intersecting area between circles is[34]:

$$Area = R_1^2 \arccos\left(\frac{d^2 + R_1^2 - R_0^2}{2dR_1}\right) + R_0^2 \arccos\left(\frac{d^2 + R_0^2 - R_1^2}{2dR_0}\right)$$

[34] Obtained from Wolfram Research's website (www.wolfram.com).

Chapter 3: Modelling Deformable Mirrors for Open-Loop Operation

3.1 Overview

CANARY is a Multi-Object Adaptive Optics experiment to be installed at the 4 meter William Hershel telescope. As such, it relies on controlling its deformable mirrors in open-loop. This chapter describes the efforts made to develop models to control deformable mirrors in such a configuration. We present models for two types of DM, a 'Xinetics' mirror with electrostrictive actuators and a 'Boston Micromachines' MEMS DM. We implemented two different modelling techniques, achieving similar performance with respect to other types of models found in the literature. This work has been published in two peer-reviewed papers, included in the Appendix of this thesis.

3.2 Introduction

We introduced Multi-Object Adaptive Optics (MOAO) in the previous chapters. MOAO requires operation in an optical open-loop, i.e. the wavefront sensor does not measure residual turbulence but rather "raw" incoming aberrations caused by the atmosphere. The DM in charge of applying the inverse of the turbulent phase must therefore work with no feedback, unlike classical AO systems, where the wavefront sensor closes the optical loop for DM positioning.

MOAO is the subject of intensive study at present, with a number of research groups implementing laboratory experiments as well as on-sky demonstrators to confirm this technique works (Andersen 2008, Myers 2008). These groups have all acknowledged the need for a 'DM Model', which we propose should work as described in Fig. 3.1.

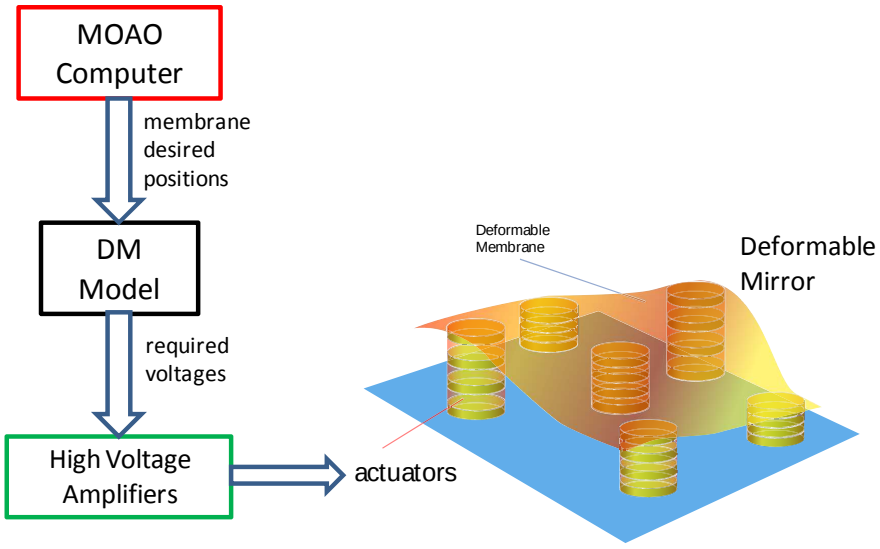


Fig. 3.1: Operation of a DM model within a MOAO system.

The MOAO computer receives the measurements of the distorted wavefronts from NGS or LGS and computes a correction to be applied by the DM. This is the input to the DM Model, expressed in some common units[35] between the MOAO computer and the DM Model. The DM model calculates the required voltages to apply to the DM actuators in order to achieve the desired shape on the DM membrane, sending a vector of voltages to the high voltage amplifiers that drive the device. It can be understood how important it is to have an accurate DM model that will compute the correct voltages for each deformation requested by the MOAO computer.

Other groups implementing MOAO systems have designed DM models that we call "*physical models*", since they take into account the physical properties and characteristics of the deformable mirror, for instance force and stress on the membrane, to produce a model that mimics the behaviour of the DM components (Stewart 2007). By researching the fairly limited literature on the subject, we concluded that at some point all models need to implement some simplifications in order to make the problem solvable[36] (Vogel 2006), which in the end limits the ability of the model to predict the DM shape with high accuracy. These physical models are described in a later section of this chapter.

[35] This can be some typical units used in optics, such as nanometres or angstroms.
[36] This is to solve the model's equations with standard techniques.

Looking for a novel and simpler approach to model the DM behaviour, and having worked with artificial neural networks (ANN) systems in other areas[37], it appeared sensible to try techniques such as ANN in the DM model problem. From existing literature and our own experiments, we knew that the behaviour of certain types of DM is not linear, i.e. the combination of actuator voltages produces a DM shape that is not the linear combination of the individual actuators. Therefore, it could be appropriate to use techniques that are efficient in modelling non-linear systems for this particular problem.

From past experience, we knew the main limitation of ANN is how the complexity of the network scales up with the number of inputs and outputs. As we had limited experience with ANN, we searched for an expert in this field, who could give us advice and support in handling large networks. We were fortunate to contact Dr. Javier de Cos, a professor at University of Oviedo (Spain), who joined us in an active and fruitful collaboration. Dr. de Cos has a wide experience in modelling with techniques such as ANN, among others, although he was not familiar with adaptive optics or deformable mirrors.

We tried two different techniques from within the 'non-parametric estimation' category. They are 'multivariate adaptive regression splines' (MARS) and 'feed forward multi-layer perceptron back-propagation' (MLP-BP) ANN. We have two DMs with non-linear behaviour: a Xinetics DM and a Boston Micromachines MEMS[38] DM. The MARS technique was used to model the Xinetics and the Boston MEMS DM was modelled using MARS and MLP-BP techniques. We obtained similar results with respect to other existing models, with the advantage of only having to take a large set of data with an interferometer in front of the DM to feed the model, as the only modelling stage.

The work described in this chapter was divided between Durham and Oviedo using the following scheme: Durham led the project, implementing the optical experiments and the associated metrology with an interferometer. We reduced all the data in Durham and analysed its validity. The reduced data were then transferred to a computer server in Durham where Dr. de Cos had access, so that he could run the training scripts to build the models. Once trained, the models were fed with new

[37] I have worked with ANNs for control applications, while doing my undergraduate engineering degree.
[38] MEMS stands for Micro-Electro-Mechanical System and it refers to those electro-mechanical systems built on a semiconductor substrate.

data, the results of which were taken back to the Durham optical table for obtaining final results. Dr. De Cos shared his expertise in tuning training parameters and implementing novel approaches for large models, as in section 3.7.3.

3.3 Description of Deformable Mirrors

For the purposes of this chapter, a DM is a device with a rectangular grid[39] of some type of electrically driven actuators and a deformable facesheet that lies on top of them. Fig. 3.2 is a representation of such a device.

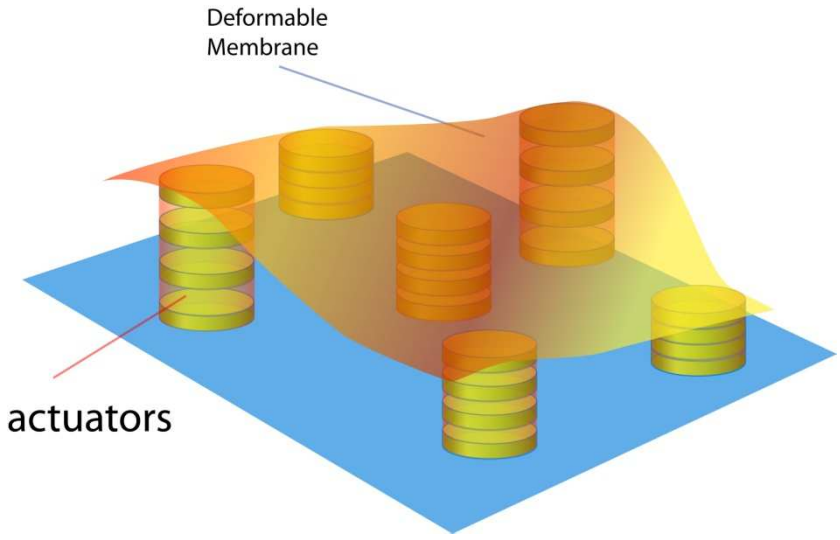


Fig. 3.2: A schematic view of a deformable mirror.

The actuators' sole function is to modify their length, with the consequence of changing the shape of the DM surface. The main types of DM are briefly listed in table 3.1.

The first classification comes from the physical principle used in the actuators. The first two technologies available were piezo-electric ceramics (made of lead, zirconate and titanate or PZT) as well as electrostrictive material (made of lead magnesium niobate or PMN). Both actuator types suffer from *hysteresis*, which is a type of 'mechanical memory' where the actuator's position depends on the previous position (whether it was above or below the future position) and therefore it becomes

[39] There are other configurations of actuators, most notably in bimorph mirrors, but they are not treated in this chapter. Please review Hardy (1998), chapter 6, for more details on that topology.

difficult to predict its final position. One of the fundamental differences between PZT and PMN materials is their Curie point[40]. While PZT actuators have a Curie point between 200 C and 350 C, PMN actuators have a Curie point at 0 C. This makes hysteresis behaviour very different: for PZT actuators, hysteresis is of the order of 10% and constant at typical operating temperatures, whilst PMN hysteresis is very low at room temperature, but it can become very high at colder temperatures, which are common in observatory environments.

Table 3.1: Types of Deformable Mirrors, classified according to actuator technology.

Actuator type	**Main manufacturer**	**Optical aperture (mm)**	**Number of actuators**	**Main characteristics**
Electrostrictive	Xinetics (USA)	(*)	(*)	Low hysteresis, but temperature dependent
Piezo stack	Cilas (France)	80 - 360	52 - 4200	High hysteresis, not-dependent on temperature
Electrostatic	Boston Micromachines (USA)	1.5 - 25	140 - 4096	MEMS, Hysteresis-free
Magnetic	Alpao (France)	9 - 40	52 - 241	Highly linear Hysteresis-free

(*) Note: Xinetics DM are highly customizable and are used in classified projects, so we do not have updated information on their current specifications

Later actuator types are electrostatic and magnetic, both of which are hysteresis-free. The former have been developed in MEMS, which have been very well received by the adaptive optics community, because they can have a very high density of hysteresis-free actuators. Magnetic actuators are a fairly novel technology, consisting of voice-coil actuators (magnet and solenoid). They have found good acceptance as well, given their lower cost and high linearity. Deformable mirrors made of both types of actuators (from Boston Micromachines and Alpao) have been incorporated into MOAO systems in the last few years (Andersen 2008).

[40] The Curie point in piezoelectric materials is the temperature at which the material loses its spontaneous polarization and piezoelectric characeristics.

Unlike magnetic actuators, which are easily driven by a control voltage of 0-10 volts or similar, PZT, PMN and electrostatic actuators require high voltages to produce a few microns of displacement, typically 100 volts for PMN, 200 volts for electrostatic and 400 volts for PZT[41]. This makes the DM driver a bulky component, which cannot provide more bandwidth[42] than a few KHz at the most.

Table 3.2: Deformable mirrors modelled, with their total number of actuators and number of actuators modelled.

Deformable mirror	**Pupil size**	**Number actuators**	**Sector modelled**	**Actuators modelled**
Xinetics	75 mm	97	11 x 5	55
Boston MEMS	10 mm	1020	14 x 9	126

Table 3.2 presents the characteristics of the two DMs modelled in this chapter. For reasons explained further below in the chapter, we did not model the complete pupil of the mirror, but a sector, so the table also includes the number of actuators modelled.

3.4 Metrology Equipment

Before progressing to the description of the DMs, it is pertinent to introduce the metrology equipment we used to measure the shape of the DM surface. This was a commercial Twyman-Green interferometer[43], a '*Fisba*', which uses a 633 nm He-Ne stabilized laser source and a 512 x 512 pixels CCD camera for interferogram acquisition. Fisba computer software calculates the phase map from a set of interferograms acquired with the instrument. Fig. 3.3 presents a diagram of how the DM surface is measured with the Fisba interferometer. All of the phase maps presented in figures throughout this chapter were obtained with this instrument. We tested the repeatability of the measurements by acquiring a large number of interferograms of a static DM and found it to be 6 nm (this value corresponds to the standard deviation of the phase maps).

[41] The voltage on PZT depends on the hardness of the material. Harder materials need higher voltage but they have less hysteresis.
[42] An adaptive optics bandwidth is fundamentally limited by the wavefront sensor frame time, which depends on the magnitude of the guide star. Typical bandwidths for current astronomical AO systems are a few hundred Hertz.
[43] http://www.fisba.ch.

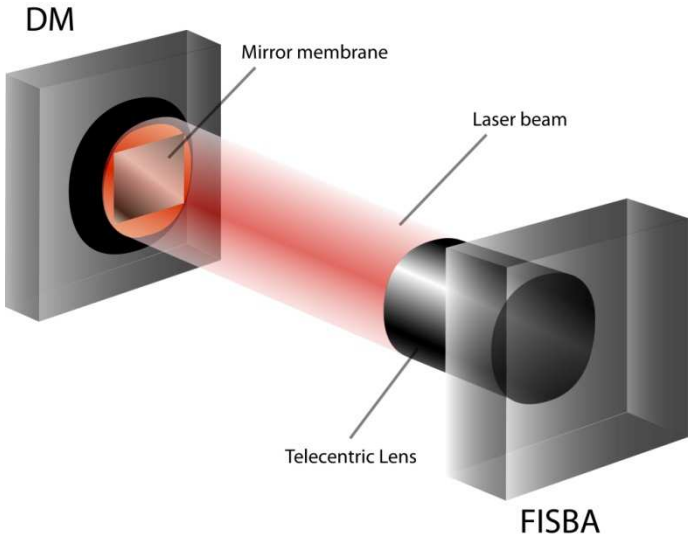


Fig. 3.3: Schematic representation of the Fisba – DM setup on an optical bench.

3.5 DM Characterization

The relationship between the position of the DM (for example, on top of an actuator) and the voltage applied to that actuator is in general non-linear. In particular, the Xinetics DM and the Boston DM have a quadratic relationship between voltage and actuator stroke. See Fig. 3.4, which shows this relationship for the case of the Xinetics DM.

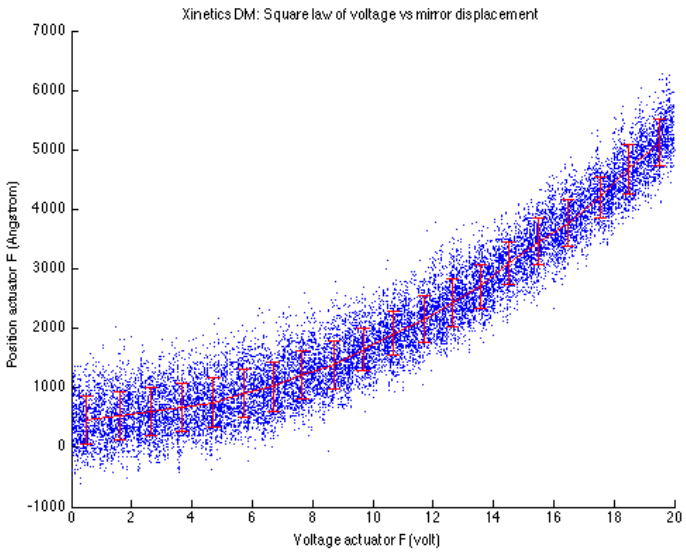


Fig. 3.4: Voltage versus stroke, Xinetics DM. This plot was obtained with one of the data sets used for the results in Fig. 3.7, further below in this section.

Besides the non-linear relationship between individual actuator displacements and their voltages, there are other non-linear effects. For instance, when trying to represent the final position of the facesheet as a linear combination of individual actuator deformations, one can incur large errors. An example of this non-linearity is presented in Fig. 3.5: We experimented with the sector of 9 x 5 actuators (described in table 3.2), which are centrally located on our Xinetics DM. We applied a half-range offset to all actuators to 'raise' the DM, placing the facesheet at a 'zero' phase condition, to then exercise each actuator individually, applying +12 volts with respect to this offset. We obtained the sum of all individual actuators (top-left panel in Fig. 3.5). The second experiment was to apply +12 volts to all actuators at once ("joint poke", top-right panel in Fig. 3.5). We evaluated the difference between results, the residual phase map (bottom-left panel in Fig. 3.5) and a slice of that residual (bottom-right panel in Fig. 3.5).

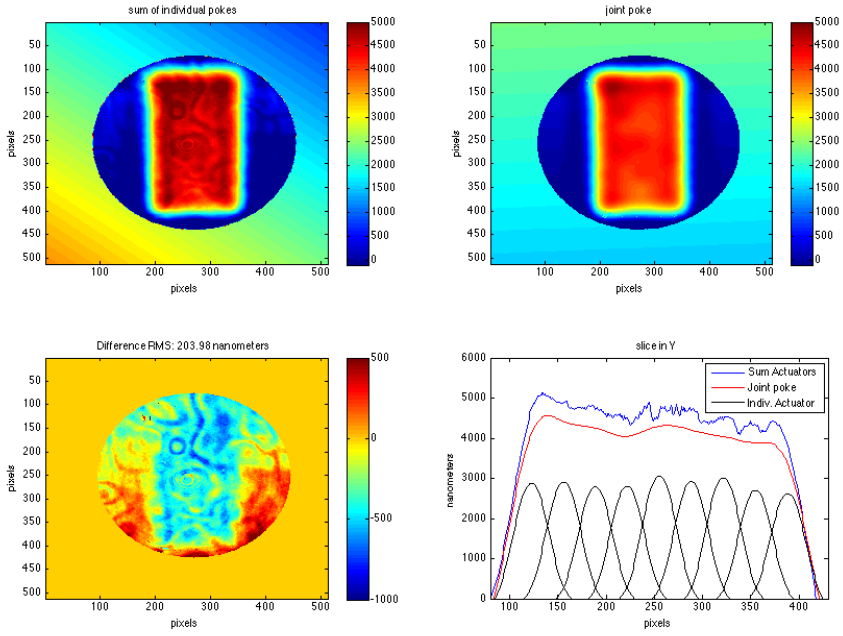


Fig. 3.5: Non-linear behaviour of DM actuators, in a 9 x 5 actuators example. Top-left panel: sum of individual actuators, poked to +12v. Top-right panel: combined effect when poking all actuators together to +12v. Bottom-left panel: difference between top panels, with residual RMS for the area being poked. Bottom-right panel: slice of the bottom-right panel along the central column, showing individual actuators, the sum of them and the joint poke. The Z coordinate unit (color bar) is nanometres.

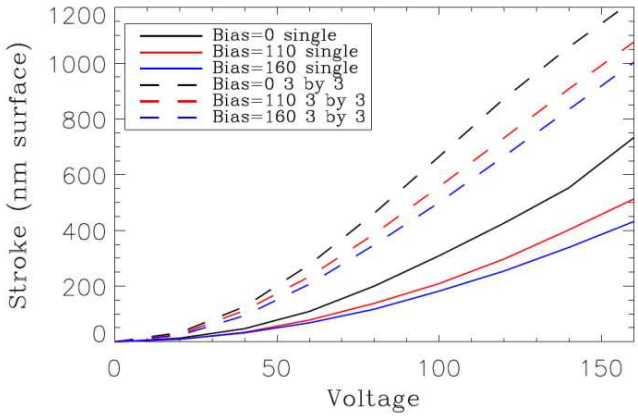


Fig. 3.6: Non-linear behaviour of a MEMS DM, where individual and joint poke are compared. Reproduced from Evans (2006). Solid lines correspond to poking one actuator; dashed lines correspond to poking a set of 3 x 3 actuators.

This simple exercise shows that the sum of actuators produces an excessive result of about 200 nm with respect to the joint poke. Therefore, modelling the final shape of a DM is not as simple as combining the individual contributions of the actuators.

Similarly, MEMS DMs have a non-linear effect when doing joint pokes. In Fig. 3.6 we reproduced a result from Evans (2006), which is an excellent example of this behaviour: individual pokes (solid lines in Fig. 3.6) produce about half the stroke than joint pokes (dashed lines in Fig. 3.6).

Traditionally, the 'influence function' is the concept used to describe the shape of the deformed facesheet around an actuator being poked. However, the final shape of any influence function would ultimately depend on the positions of the neighbour actuators. Depending on the physical characteristics of the facesheet, this effect may or may not extend beyond adjacent neighbour actuators. This cross-talk effect implies that it is not enough to characterize the individual influence functions, but to look for a more thorough way of understanding the 'effect' of neighbour actuators. From the modelling experience we will describe further below, we found a satisfactory way of characterizing the effect of neighbour actuators: we poke an area of the DM with random values and compute the correlation coefficients matrix R (or 'cross-correlation'). If C is the covariance matrix of the set of data, the cross correlation R is defined with individual elements as

$$R_{i,j} = \frac{C_{i,j}}{\sqrt{C_{i,i} \cdot C_{j,j}}} \quad (3.1).$$

Each row and column of *R* represents one actuator, and the values at positions *(i,j)* in *R* correspond to the correlation between actuators *i* and *j*. Using this definition, it follows naturally that the diagonal elements of R will be unity. We present *R* for a 5 x 5 actuator area of the Xinetics DM in Fig. 3.7. We also show in that figure a comparison between the cross-correlation and the traditional influence function (normalized) for the same actuator. The cross-correlation results are slightly higher for immediate neighbours than a standard influence function[44].

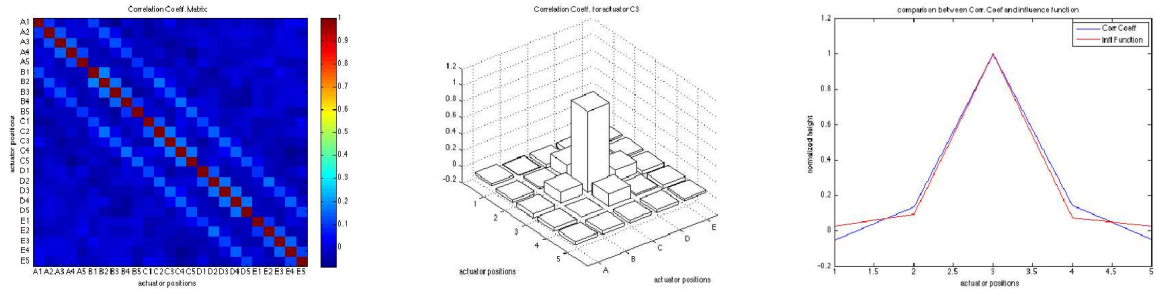


Fig. 3.7: Left panel: correlation coefficients matrix R; central panel: correlation coefficients for the central actuator, taken from matrix R; right panel: comparison between a traditional influence function and the correlation coefficients.

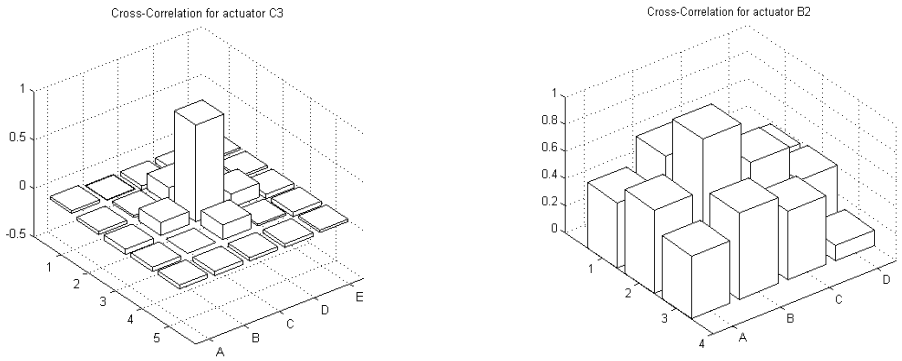


Fig. 3.8: Left panel: correlation coefficients for Xinetics actuator; right panel: correlation coefficients for Boston MEMS actuator.

In Fig. 3.8 we compare cross-correlations for the Xinetics and Boston MEMS. While for the Xinetics DM, the immediate neighbours actuators (the ones forming a 'cross' with the central actuator in Fig. 3.8) are the only relevant ones for the final shape of the mirror, in the Boston MEMS the

[44] In an influence function, the adjacent actuators are not being actuated, so they are passive. However, in this experiment all actuators are actuated, so they are active and with all the possible range of voltages they accept, so their contribution to the final position of the actuator under analysis spans all possible cases.

situation is more complex, because not only the immediate actuators are important (and they are relatively more important than in the Xinetics case), but the actuators on the corners as well as actuators two rows/columns apart can have an effect on the final position of the mirror. These two different situations are consequences of the very different physical characteristics of the mirrors: the Xinetics DM has a stiff membrane and the Boston MEMS has a very thin and floppy membrane, therefore the level of deformation of the membrane at neighbour actuator positions is very different. For a real scenario of multiple voltages on all actuators, the final shape of the mirror would not be easy to predict.

This section gives a fairly good idea of the complexity involved in modelling these deformable mirrors. As we see later in the chapter, if we think in terms of the number of variables that an algorithm such as MARS will incorporate into the model[45], it would probably be the central actuator and the 4 immediate neighbours for the case of the Xinetics, but in the case of the Boston MEMS, it may include the 8 surrounding neighbours and possibly some others from the next rows and columns. One of the main advantages in the non-parametric regression technique selected is that it does the job of selecting the relevant variables for us.

3.6 Physical Models

Before going on to describe the techniques we used to model our deformable mirror, we describe the most relevant physical models one can find in the literature.

The main driver behind the need for a DM model in adaptive optics systems before MOAO was the importance of handling hysteresis and non-linear behaviour adequately. A good example of these efforts is the paper by Hom (1999) in which the non-linear behaviour of electrostrictive actuators is modelled. This is relevant to our own study, since our Xinetics mirror has electrostrictive actuators. The authors employ Finite Elements Analysis (FEA) to model the DM facesheet and contact points with the actuators (see Fig. 3.9). They use an electrical equivalent for each actuator, as Fig. 3.10 presents.

[45] The MARS algorithm automatically discards variables that do not affect the output being modelled.

In figures 3.9 and 3.10, we reproduce some relevant figures from that paper.

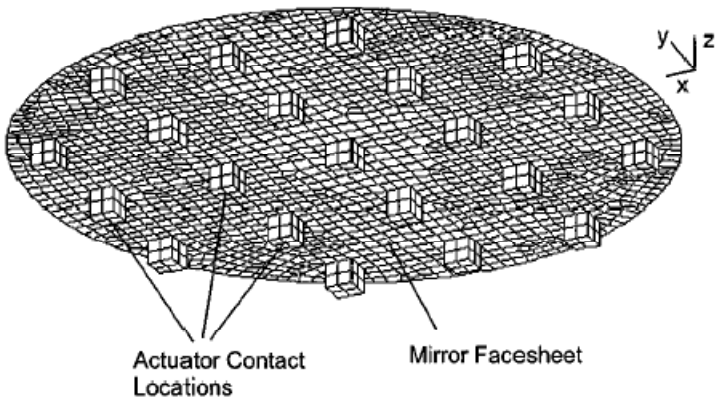


Fig. 3.9: The finite element mesh for computing the structural stiffness of the mirror. Reproduced from Hom (1999).

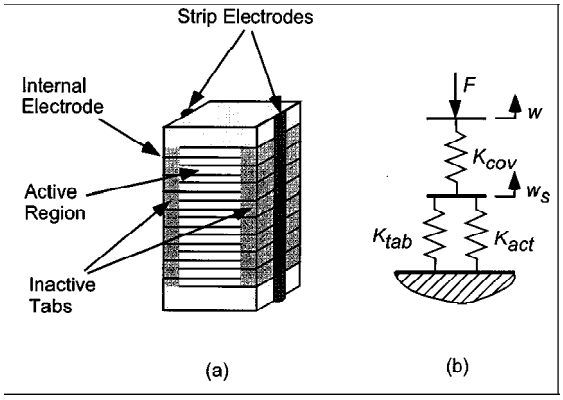


Fig. 3.10: (a) the multilayered, co-fired actuator and (b) the one-dimensional model of the actuator. Reproduced from Hom (1999).

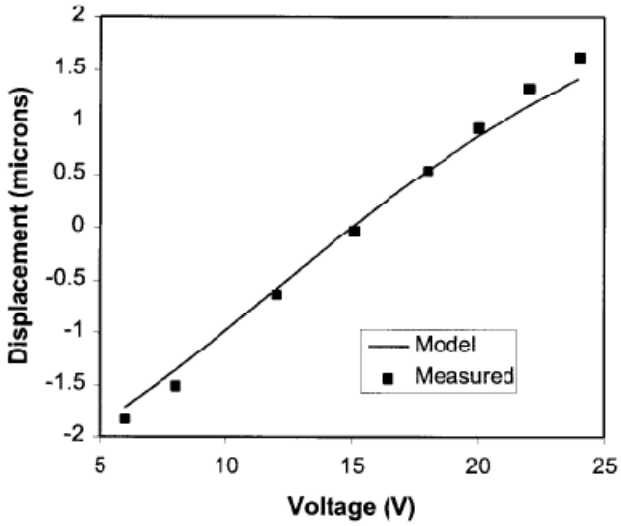


Fig. 3.11: Comparison between the static model and experimental results, when displacing the inner actuators of the mirror. Reproduced from Hom (1999).

Hom reports incurring a 13% underestimation error with respect to the real mirror position in the result reproduced in Fig. 3.11.

MEMS deformable mirrors have become popular in AO because of their high density of actuators and negligible hysteresis. They fit MOAO needs very well, given their very small size compared to other DMs. Various groups have produced physical models of this type of device for

open-loop operation. The model by Stewart (2007) is probably the best example of such a physical model.

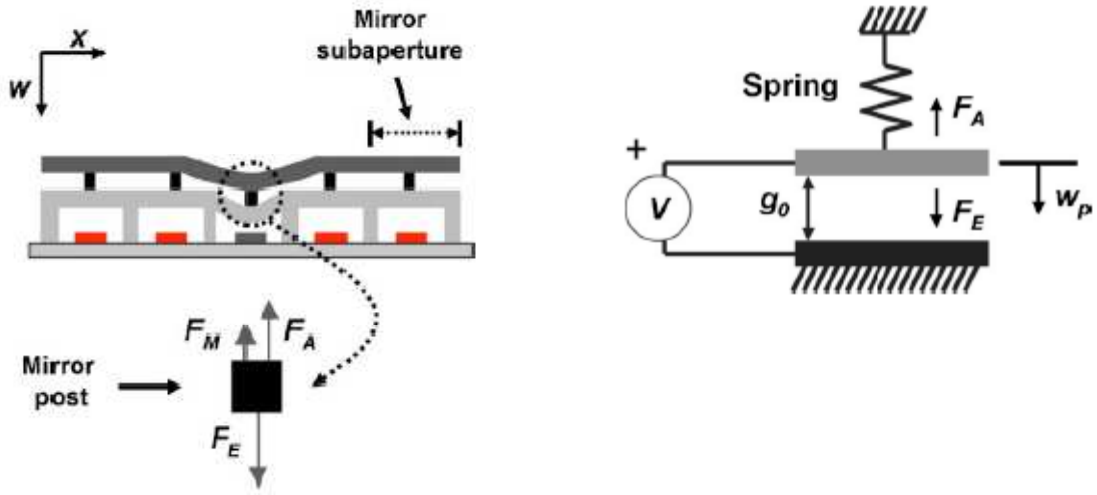


Fig. 3.12: MEMS DM model. Left panel: model of the forces interacting on each actuator. Right panel: electro-mechanical model for the electrostatic actuator. Reproduced from Stewart (2007).

In Fig. 3.12, we reproduce two figures from Stewart's paper, in which the mechanical forces and model are described. MEMS actuators work as parallel-plate electrical capacitors, where an electrostatic force is generated by the electrical field between the plates when there is a voltage difference between the plates. This force deforms the actuator (as seen in the left panel of Fig. 3.12), which, in turn deforms the mirror membrane. The whole system can be modelled as the above mentioned capacitor and a spring, representing the restoring force on the actuator. However, the assumption of parallel plates cannot be maintained for large strokes, because of the deformation suffered by the actuator. Effects like this one ultimately imply non-constant terms, such as the distance between the capacitor's plates. The model uses thin plate theory to model the facesheet, which may also result in non-linear terms when the membrane displacement goes beyond the thickness of the plate. In spite of these complexities, Stewart's model achieves errors of the order of 15 nm RMS, when commanding a MEMS DM to follow a Zernike focus term.

3.7 Non-Parametric Estimation Techniques

At the beginning of the CANARY project and this work (early 2007), it was thought that we would need to model a MEMS DM solely, so we studied the literature mentioned in the previous section, which attempts to model the MEMS DM behaviour following physical principles. However, the non-linearities incurred in the Boston MEMS seemed to be too large to be able to find such a model without employing linearization processes or trying to fit non-linear equations to the data. The number of actuators which define the final position of a MEMS DM at any given actuator seemed to be too large, making the problem quite complex[46]. On the other hand, we were aware that there are a number of techniques to model non-linear problems with multiple inputs and outputs, probably one of the most popular being Artificial Neural Networks (ANN). This is an example of 'Non-parametric estimation techniques', which are those algorithms that adapt their functionality to the data they are exposed to. The same topology of neurons can be used to model a deformable mirror or to detect breast cancer in adult women[47], depending on the data the net is being exposed to. After considering the best algorithms for our required modelling, we performed experiments and settled on using two different ones: Multivariate Adaptive Regression Splines (MARS) and a particular type of neural network, multi-layer perceptron back-propagation (MLP-BP). In collaboration with Dr. De Cos, we published two peer-reviewed papers: in the first paper we modelled our Xinetics DM using MARS; in the second paper we modelled the Boston MEMS DM using MARS and ANN. Both papers are included in the Appendix of this thesis.

To implement these models, we used the *R* Language[48], which is the open-source implementation of *S*, a language for statistical computing developed by Dr. John Chambers at Bell Labs. *R* comes with packages, of which MARS and several ANN ones are standard.

[46] See previous section on DM characterization.
[47] These two very different examples are now within Dr. De Cos' own expertise.
[48] http://www.r-project.org.

3.7.1 Multivariate Adaptive Regression Splines (MARS)

Multivariate Adaptive Regression Splines is a multivariate nonparametric regression technique introduced by Jerome H. Friedman (1991). A very clear description of MARS can be found in the abstract of a paper by Friedman and Roosen (1995), which we reproduce here:

> Multivariate Adaptive Regression Splines (MARS) is a method for flexible modelling of high dimensional data. The model takes the form of an expansion in product spline basis functions, where the number of basis functions as well as the parameters associated with each one (product degree and knot locations) are automatically determined by the data. This procedure is motivated by recursive partitioning (e.g. CART)[49] and shares its ability to capture high order interactions. However, it has more power and flexibility to model relationships that are nearly additive or involve interactions in at most a few variables, and produces continuous models with continuous derivatives. In addition, the model can be represented in a form that separately identifies the additive contributions and those associated with different multivariable interactions.

Mathematically, the purpose of MARS is to predict the values of a continuous dependent variable, $y_{n\times 1}$, from a set of independent explanatory variables, $X_{n\times p}$. The MARS model can be represented as

$$y = f(X) + e \qquad (3.2),$$

where $e_{n\times 1}$ is the error vector.

MARS does not require any *a priori* assumptions about the underlying functional relationship between dependent and independent variables. Instead, this relation is uncovered from a set of coefficients and piecewise polynomials (basis functions) of degree q that are entirely driven from the regression data *(y, X)*. The MARS regression model is constructed by fitting *basis functions* to distinct intervals of the independent variables. Generally, the basis functions are piecewise polynomials, also called splines. In MARS terminology, the joining points of the polynomials are called 'knots' and will be denoted by *t*. For a spline of degree q each segment is a polynomial function. MARS uses two-sided truncated power functions as spline basis functions, described by the following equations:

$$[-(x-t)]^q = \begin{cases} (t-x)^q, & x < t \\ 0, & x \geq t \end{cases} \qquad (3.3)$$

[49] CART stands for Classification And Regression Trees, a set of techniques using binary trees for tackling classification and regression problems.

$$[+(x-t)]^q = \begin{cases} 0, & x < t \\ (t-x)^q, & x \geq t \end{cases} \quad (3.4),$$

where $q \geq 0$ is the power to which the splines are raised and determines the smoothness of the resultant function estimate.

When $q = 1$, which is the case for the models implemented in this chapter, only simple linear splines are considered. As an example, a pair of splines for $q = 1$ and one knot at $t = 0.5$ are presented in Fig. 3.13.

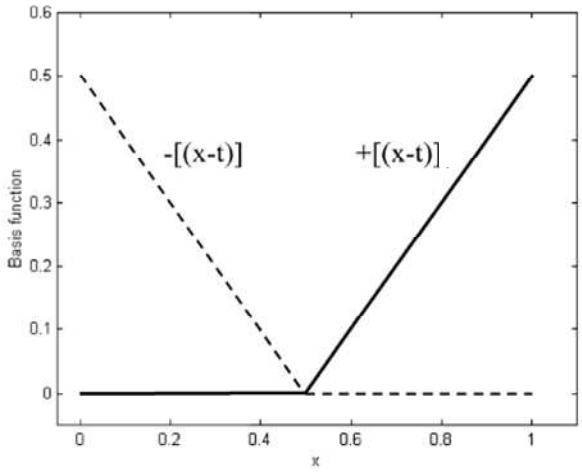


Fig. 3.13: A graphical representation of two spline basis functions. The left spline $x < t, -(x-t)$ is shown as a dashed line; the right spline $x \geq t, +(x-t)$ as a solid line.

The solid line represents the right-sided spline, $x \geq t, +(x-t)$, which is positive for all values located at the right side of the knot *t*. The dashed line represents the left-sided spline, $x < t, -(x-t)$, which is positive for all values located at the left side of the knot *t*.

The MARS model of a dependent variable y with *M* basis functions can be written as

$$\hat{y} = \hat{f}_M(x) = c_0 + \sum_{m=1}^{M} c_m B_m(x) \quad (3.5),$$

where $\hat{y}$ is the dependent variable predicted by the MARS model, c_0 is a constant, $B_m(x)$ is the *m-th* basis function, which may be a single spline function or a product (interaction) of two or more spline basis functions. c_m is the coefficient of the *m-th* basis function.

MARS will optimize the number of basis functions (*M*) introduced into the model and the number and positions of the knots introduced into the model. The MARS algorithm is recursive and was designed to be computationally feasible.

To determine which basis functions should be included in the model and to measure the 'lack-of-fit' of the model, MARS utilizes the 'generalized cross-validation' (*GCV*). The *GCV* is the mean squared residual error divided by a penalty dependent on the model complexity. The *GCV* criterion is defined as

$$GCV(M) = \frac{1}{n}\frac{\sum_{i=1}^{n}[y_i - \hat{f}_M(x_i)]^2}{[1 - C(M)/n]^2} \quad (3.6),$$

where $C(M)$ is defined as

$$C(M) = M + dM \quad (3.7),$$

where *M* is the number of basis functions in Eq. 3.5 and the parameter *d* is a penalty for each basis function included into the model. $C(M)$ can be seen as a complexity penalty that increases with the number of basis functions in the model. The parameter *d* can be regarded as a smoothing parameter. Large values of *d* lead to fewer basis functions and therefore smoother function estimates. *d = 2* in our models.

3.7.2 Artificial Neural Networks (ANN)

Historically, ANNs were meant to operate through simulating the activity of the human brain at a simplified level. The ANN accomplishes this through a large number of highly interconnected processing elements (neurons), working together to solve specific problems, such as forecasting and pattern recognition. Each neuron is connected to some of its neighbours with varying coefficients or weights that represent the relative influence of the different neuron inputs compared to other neurons.

The feed forward multi-layer perceptron back-propagation (MLP-BP) network is one of the most popular techniques in the field of ANN. The common topology of a MLP-BP neural network model is illustrated in Fig. 3.14. The source nodes in the input layer of the network supply respective elements of the activation pattern or input vector, which constitute the input signals applied to the

neurons in the hidden layer. The output signals of the hidden layer are used as inputs to the output layer. The output signals at the output layer constitute the overall response of the network to the activation patterns applied by the input layer neurons.

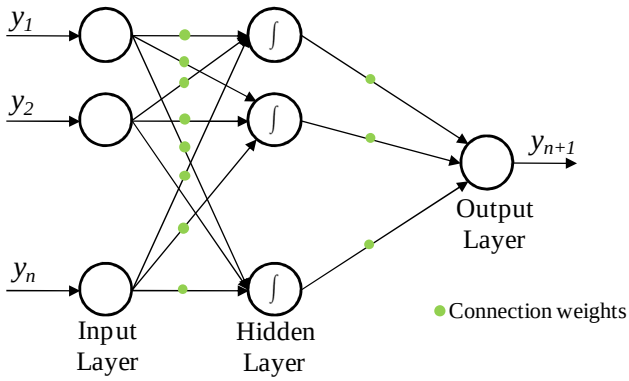


Fig. 3.14: Topology of a simple feed forward multi-layer perceptron back-propagation ANN.

With *n* input neurons, *m* hidden neurons, and one output neuron (similar to Fig. 3.14, with m neurons in the hidden layer), the training process of MLP-BP network can be described as follows:

1) Calculate the outputs of all hidden layer nodes, using:

$$net_j = \sum_{i=0}^{n} w_{i,j} y_i \quad (j = 1, \dots, m) \quad (3.8),$$

$$z_j = f_H(net_j) \quad (j = 1, \dots, m) \quad (3.9),$$

where net_j is the activation value of the j^{th} node, $w_{i,j}$ is the connection weight from input node i to hidden node j, y_i is the i^{th} input, z_j is the corresponding output of the j^{th} node in the hidden layer, and f_H is called the transfer function of a node, which is usually a sigmoid function. Mathematically

$$f_H(x) = \frac{1}{1+e^{-x}} \quad (3.10).$$

2) Calculate the output *O* of the network from the output layer neurons using

$$O = f_O\left(\sum_{j=0}^{m} w_{j,k} z_j\right) \quad (3.11).$$

where f_O is the activation function of the network (usually a linear function), $w_{j,k}$ is the connection weight from hidden node j to output node k ($k=1$ for this example) and z_j is the corresponding output of the j^{th} node in the hidden layer. All the connection weights and bias values are assigned with random values initially, and then modified according to the results of MLP-BP training process.

3) Minimize the 'cost function'. The instantaneous error for each pattern from the training set exposed to the network is as

$$\mathcal{E}(t) = \frac{1}{2}(O - y_t)^2 \quad (3.12),$$

where *t* is the t^{th} term of the training set. The *average squared error* is defined as

$$E_{av} = \frac{1}{N}\sum_{t=1}^{N} \mathcal{E}(t) \quad (3.13),$$

where *N* is the size of the training set. E_{av} represents the *cost function* and it measures the learning performance for a given training set.

3.7.3 ANN models implemented

"Network architecture" denotes the number of input/output variables, the number of hidden layers and the number of neurons in each hidden layer. It determines the number of connection weights and the way information flows through the network. One of the most important decisions when defining the ANN structure is to determine the appropriate number of hidden layers and the number of neurons in each layer. There is no systematic way to establish a suitable architecture, and the selection of the appropriate number of neurons is basically problem specific. Using a single hidden layer is a popular option, so we used it for our models. Regarding the number of neurons, we followed a trial and error procedure until we obtained satisfactory results. ANN inputs and outputs are straightforward to determine: the number of inputs is equal to the number of outputs and it is the number of actuators being modelled, provided we are measuring the DM membrane at actuator coordinates.

Before fitting an ANN model, the data should be pre-processed. There are two reasons for doing this: pre-processing can ensure that all variables receive equal attention during the training process. Otherwise, input variables measured on different scales will dominate training to a greater or lesser extent because initial weights within a network are randomized to the same finite range. Also, pre-processing is important for the efficiency of training algorithms. For example, the *gradient descent algorithm* used to train the MLP is particularly sensitive to the scale of data used. Due to the nature of this algorithm, large values slow training down because the gradient of the sigmoid function

at extreme values reaches zero. In general, there are fundamentally two types of pre-processing methods: The first one is to rescale the data to a small interval (referred to as *rescaling*), such as *[−1, 1]* or *[0, 1]*, depending on the transfer function used in the neurons. Some transfer functions are bounded (*logic* and *hyperbolic tangent* functions for example). The second method is to standardize the data by subtracting the mean and dividing by the standard deviation to make the data have a mean of *0* and variance of *1* (referred to as *standardization*).

MEMS DMs are very attractive for AO because of the large number of actuators they can offer. This presents the problem of scalability particular to ANN, which become very large when there are multiple inputs and outputs. Given that this work attempts to explore the applicability of these techniques, we devised a strategy to reduce the dimensionality of the problem. We designed a simple procedure to 'clone' the ANN models, which relies on the isotropic behaviour of MEMS DMs, i.e. a subsector of the facesheet behaves similarly irrespective of where this subsector is placed within the limits of the facesheet, as long as it is not at the edge of the pupil, where there are boundary effects. To model the Boston MEMS DM, we experimented with two different sizes of ANN:

- **ANN small (ANNs)** : MLP with one hidden layer and ANN structure: 12 x 16 x 12 neurons
- **ANN big (ANNb)** : MLP with one hidden layer and ANN structure: 30 x 40 x 30 neurons

As an example, the cloning strategy for ANNb is illustrated in Figures 3.15 and 3.16. We can see the DM modelled area of *14 x 9 = 126* actuators (see Table 3.2), which has been divided in seven 6 x 5 actuators sectors, illustrated in different colours in Fig. 3.15. Data from the central sector (white in the figure) were used for training the ANNb model. As built, the ANNb model can only model a sector of 6 x 5 actuators, so the *cloning procedure* consists of applying this small model to the coloured sectors in Fig. 3.15. The different sectors must overlap in order to give continuity to the model. The six different sectors within the 14 x 9 actuator region are identified in Fig. 3.16.

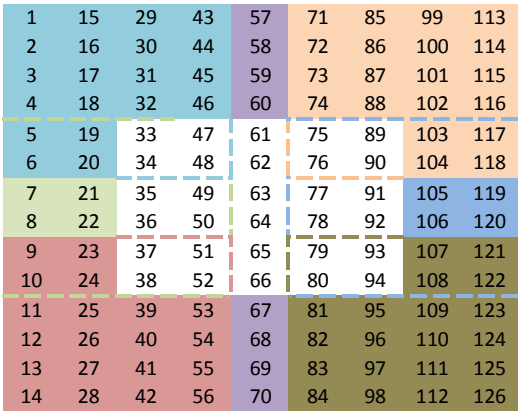


Fig. 3.15: Cloning strategy to model the 14 x 9 actuator region, using the same model of 6 x 5 actuators, replicated at different positions.

The cloning procedure works as follows: The actuators in white in Fig 3.15 are modelled with the ANNb placed at the central sector. For all other actuators, the ANNb is placed at one of six positions (depicted in Fig. 3.16) from where the following two cases can occur:

- The actuator belongs to only one of the 6 sectors, for example, any of the actuators on the bottom left sector in 'red', i.e. the ones forming this matrix: $\begin{bmatrix} 11 & \cdots & 53 \\ \vdots & \ddots & \vdots \\ 14 & \cdots & 56 \end{bmatrix}$. In this case, the ANNb model is used as it is, but placed at the sector and fed with the corresponding inputs (actuator positions). The outputs are calculated directly from the model.
- The actuator belongs to one of the overlapped areas, i.e. one of those that are covered by more than one sector. For example, actuators that formed this matrix: $\begin{bmatrix} 9 & 23 \\ 10 & 24 \end{bmatrix}$, which correspond to the overlap between the 'red' and 'light green' sectors. For these cases, the ANNb is run twice, one placed at each of the two sectors (using the corresponding inputs). The result for each of the actuators in the overlapped area is calculated as the average of the two results from each run.

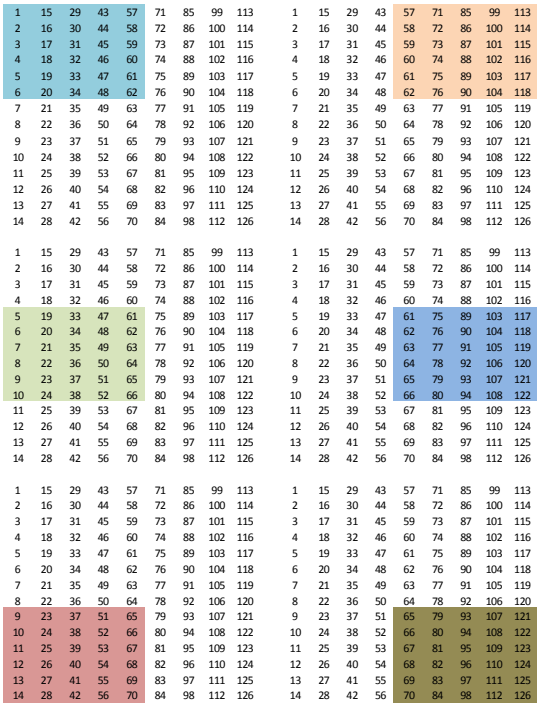


Fig. 3.16: Each of the 6 areas where the central area was cloned is shown individually here, with the same colours used in Fig. 3.15.

3.7.4 One-dimensional example

To finish the introduction to the non-parametric regression techniques used to model the DMs in this chapter, we present a one-dimensional data set, which we model with both techniques. The original equations to produce the data set are:

$$y = 0.5x + 3\ ,\ \ 0 < x < 2$$

$$y = -0.75x + 5.5\ ,\ \ 2 < x < 5$$

$$y = 0.15x + 1\ ,\ \ 5 < x < 10$$

Eq. (3.14) Set of linear equations for the one-dimensional example

We add a Gaussian noise with amplitude *0.075* to the linear system of equations in Eq. 3.14. Plots of the linear system and noisy "real" data are presented in Fig. 3.15.

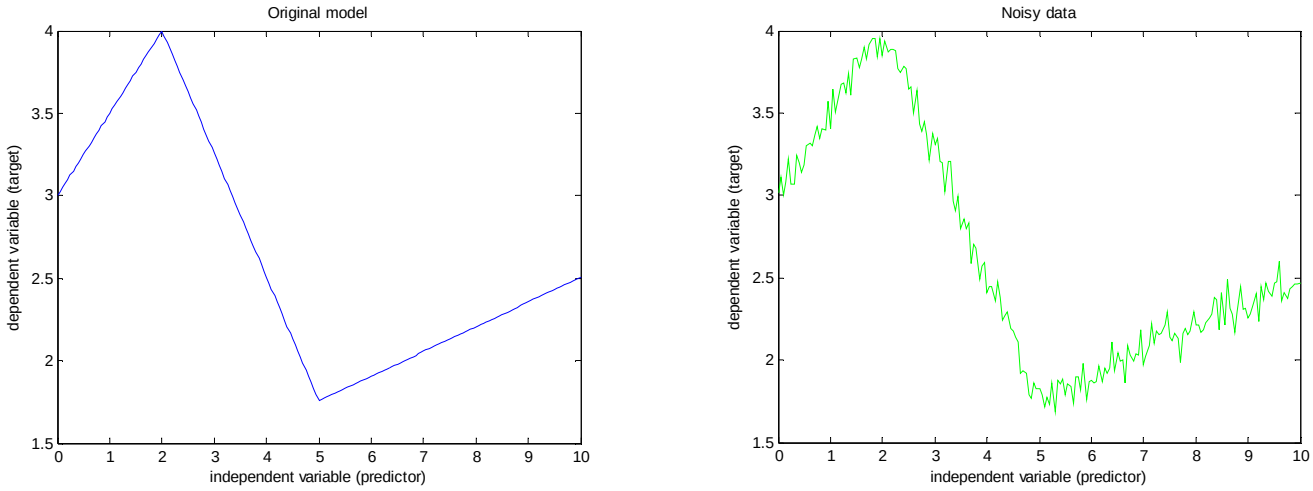


Fig. 3.15: Original noiseless model and noisy data used to train the models.

The noisy data set is used to feed a MARS algorithm. The algorithm produces a set of equations that are reproduced in Eq. 3.15, where 'BF' represents each of the Basis Functions used by the model.

BF1 = max(0, X - 5.9)

BF2 = max(0, 5.9 - X)

BF3 = max(0, X - 1.65)

BF7 = max(0, X - 2.15)

BF9 = max(0, X - 4.9)

BF11 = max(0, X - 7.15)

$$Y = 6.11177 + 0.685929 * BF1 - 0.529479 * BF2 - 0.386432 * BF3 - 0.917712 * BF7 + 0.825651 * BF9 - 0.0812522 * BF11 \quad (3.15)$$

The left panel in Fig. 3.16 presents the MARS model on top of the data set and the original model used to produce the data set. The residual RMS error is included in the title of each plot. The MARS model produces a smaller error with respect to the original model. This is because the MARS model is intended to reproduce the noisy data set.

Similarly, we implemented a standard 'feed-forward back-propagation' neural network and trained it with the noisy data. The trained network was fed with the noisy data and produced the output presented in the right panel of Fig. 3.16. The residual RMS for the ANN model is marginally worse than the MARS model, but both are able to model this data set accurately. Although the data are linear, they have two discontinuities, which complicate the modelling process.

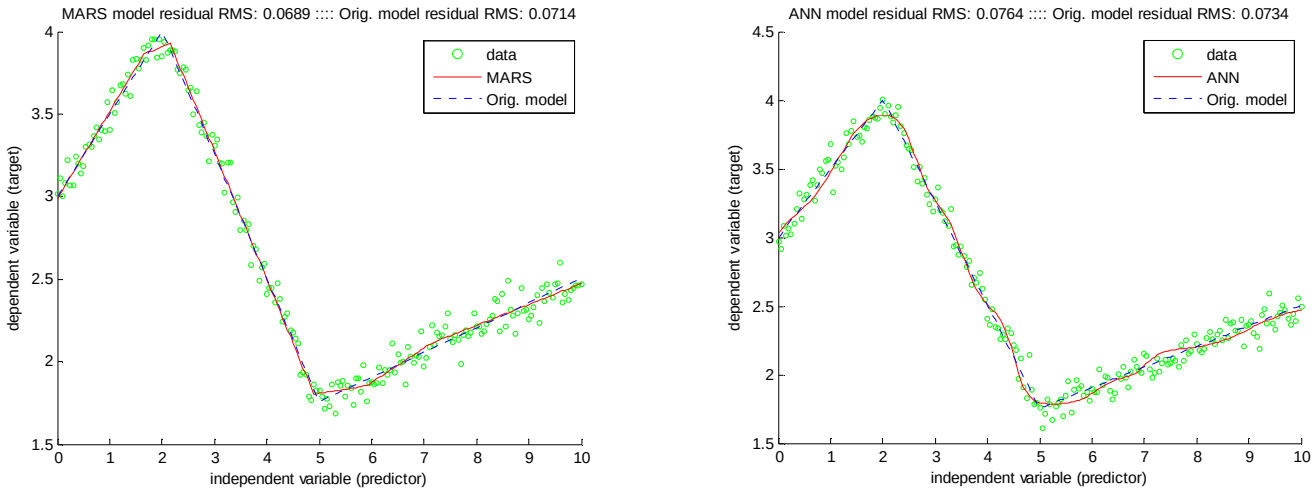


Fig. 3.16: MARS and ANN models plotted along with the original noisy data and the noiseless model.

3.8 Training and Estimation Processes

The non-parametric estimation techniques need to be exposed to real data from the process they will be modelling; this is called the 'training' process. Once they are trained, they can '*predict*' or '*estimate*' an output when an input is exposed to them. The estimation process represents the main operation of the model, when the non-parametric estimation technique is used to model the process. Both aspects of using these techniques are presented in Fig. 3.17.

In a real AO system, the DM surfaces are computed by the AO computer and passed on to the DM model (see Fig. 3.1). We did not have a complete AO system, so we had to produce DM surfaces somehow. We implemented two methods: for smooth surfaces, we used a Zernike polynomial generator, which we scaled to have interferometer units and amplitudes. For a case similar to a real AO system operation, we gathered new random shapes with the interferometer and fed our DM models with these data sets. The voltages generated by the models were used to feed the DM

once again, obtaining new sets of phase maps. The difference between both phase maps corresponds to the residual error attributable to the DM model. These concepts are relevant to the next section.

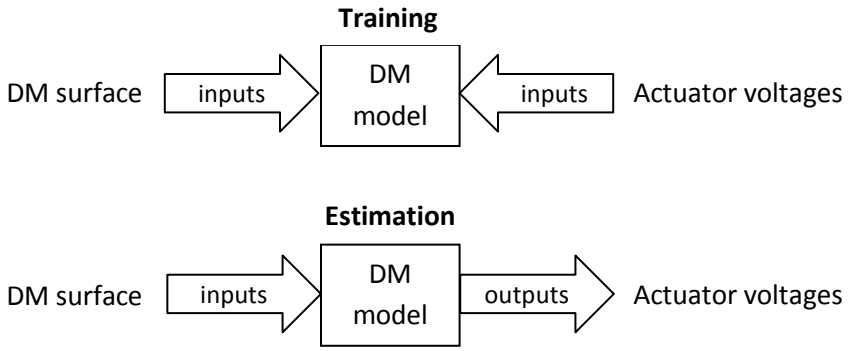


Fig. 3.17: Training and Estimation processes, for the case of Non-parametric techniques modelling a DM.

3.9 Experimental Results

3.9.1 Data Acquisition Methodology

Our data acquisition methodology for gathering the training sets was very simple: to collect a large number of random positions from the deformable mirrors with our interferometer. This is effective since we are interested in exposing the non-parametric regression algorithms to as much information as possible about the process we are modelling. In principle, we could have exposed the algorithms to ALL possible positions a DM is able to achieve, although this number is completely unrealistic. For example, for the case of a mirror with 37 actuators (a third of our Xinetics), the total number of combinations for actuators controlled with 12 bits DACs would be

$$4096^{37} = 4.54 \times 10^{133} \quad (3.16).$$

By using a random generator for the voltages to apply to the actuators, we obtained a statistical sample of the whole universe of combinations the set of actuators can have. The data acquisition consisted of generating a vector of random voltages, sending it to the DM high voltage driver for moving the DM and acquiring an interferogram of the DM. A combination of scripts in Matlab and Python did all the processing automatically, recording the voltage vector and the phase map for each random vector generated.

The existing Xinetics DM driver was part of an electronic crate with VME[50] boards running DSP computers. We wanted to run the system from a single computer, so we modified the existing system[51], designing and implementing a simple electronic interface for a personal computer's parallel port to send the digital numbers to the high voltage amplifiers that drive the DM.

In the case of the Boston MEMS DM, the high voltage driver came with a PCI board and control application so the electronic hardware was ready to be used from a Personal Computer. With support from a Software Engineer of our group, we modified the control application to receive external commands, sent from our scripts.

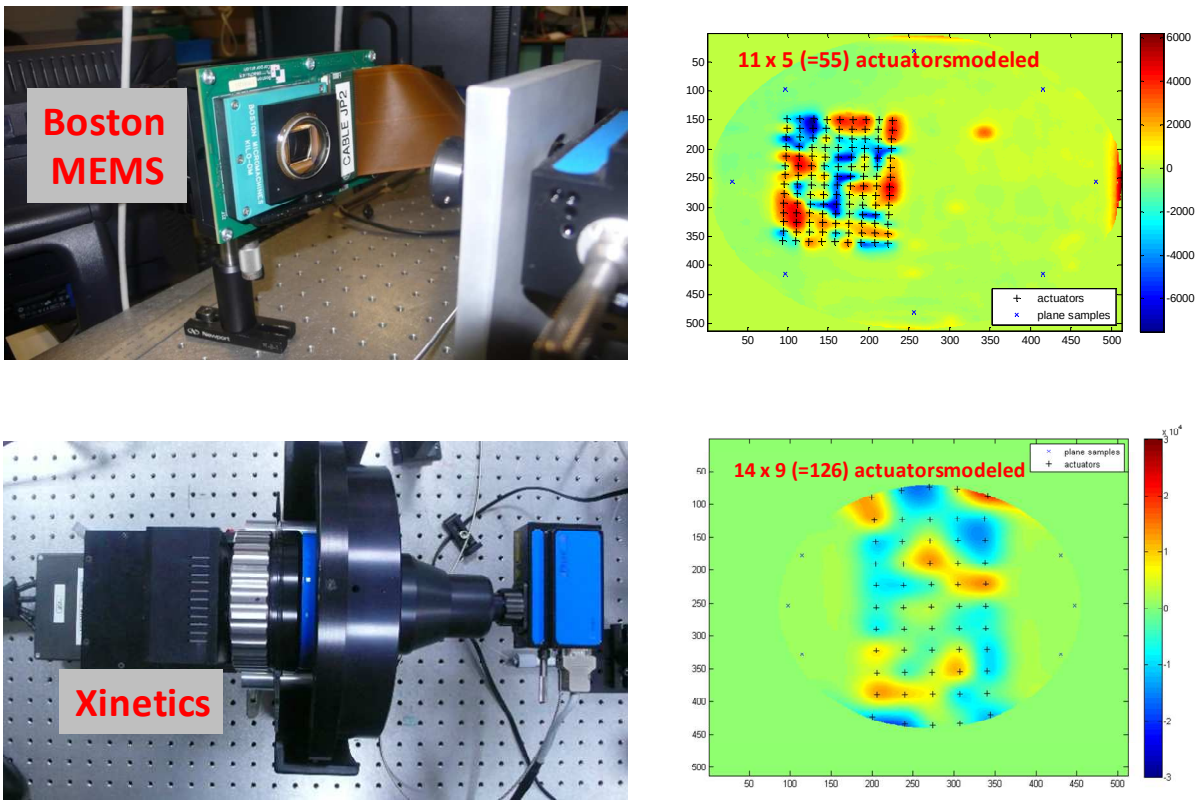


Fig 3.18: The experimental setup for each of the deformable mirrors modelled.
The right panels show the actuators being modelled in each case.

[50] VME is a high performance, asynchronous bus very popular in the 1990's.
[51] We kept backward compatibility with the existing system.

The data acquisition procedure comprises the following steps:

1) The DM is raised to a 'zero point' phase, applying half of the maximum voltage. This allows us to run the DM as a nominal wavefront corrector, able to produce positive and negative phase excursions (around the zero point)
2) A zero point phase map is acquired with the Fisba interferometer, and recorded
3) A random set of voltages are generated as a vector and sent to the DM high voltage drivers
4) A phase map is acquired, subtracting the zero point phase map obtained in (2)
5) The phase map and the vector of voltages is recorded on disk
6) The process is repeated from point (3) onwards

There was a limitation in the Fisba interferometer, which we encountered some time after we started taking measurements following the procedure described. The data acquisition script lasted typically 12 hours. During this time, there were small drifts on the phase maps, which were seen as deviations from the zero phase point, so the subtraction of this plateau did not produce the expected effect. We remedied the problem by leaving areas of the DM without being poked; therefore they were static and we took sample points in those sectors as a reference plane for the true plateau of the experiment. As can be seen in the right panels of Fig. 3.18, these points are marked as *X* in blue. We fitted a plane in three-dimensional space to these points and subtracted it from the phase maps, as part of the data reduction process. We then used a subset of the DM for modelling purposes, in order to allow areas of the DM to remain static and to limit the number of actuators to model, since we were experimenting with these new techniques and we were aware the training process would be long.

As part of the experiment, we found the position of each actuator in phase map space. For that, we poked each actuator individually and took a phase map of the DM, getting the influence function for each actuator. We then fit a 2D Gaussian to each influence function using the Nelder-Mead Simplex optimization algorithm (Nelder 1965). The main purpose of this exercise was to position the 2D Gaussian[52] at the best position in phase map space. This allowed us to find

[52] The 2D Gaussian was fixed in full-width at half-maximum, using a typical value for the influence function. It was not our purpose to fit the best Gaussian, but to find the position of the actuator in the phase map.

coordinates (x_i, y_i) for each actuator. The coordinates are given in terms of phase map pixels, e.g. $0 < x_i, y_i < 512$.

We were aware that the number of inputs to the model is one of the main factors defining its complexity and scalability. Given that the DM shape is fully defined by the actuator positions, we decided on measuring the DM at actuator positions solely. In spite of having many more measurement points (one per phase map pixel), we have to limit the number of inputs to limit the size of the models, so this was a good compromise between measuring a number of points sufficient to define the shape of the facesheet and handling the minimum number of inputs to our models. Therefore, all of our measurements have the number of points defined by the number of actuators being modelled. In that way, the models receive the same number of inputs as they produce outputs, since the model should calculate the voltages for the actuators.

Our non-parametric models were trained with large data sets. Table 3.3 gives some relevant numbers from the training process.

Table 3.3: Parameters for the training process.

Deformable mirror	**Number of actuators**	**Stroke excursion**	**training set acquired**
Xinetics	55	± 2700 nm	6000 random
Boston MEMS	126	± 700 nm	12000 random

3.9.2 Figures of Merit

When researching the existing literature, we realized the lack of a common figure to assess the quality of a DM model. For AO error budget purposes, we are interested in knowing the error in DM position in terms of nanometres, nevertheless, it is not comparable to incur a, say, 10 nm error when the wavefront to correct spans 1000 nm, rather than, for instance 100 nm. Therefore, we think it is relevant to define, in general, a figure of merit to assess the quality of our results as a ratio between error in nanometres and excursion requested. In a very recent article, Blain *et al* uses a figure of merit slightly different from our own (Blain 2010), so we have included both figures in order to do direct comparisons between our results and from other groups. See section 3.9.5 for a comparison summary.

3.9.2.a Ratio residual – desired peak-to-valley

This figure of merit relates the residual error with the peak-to-valley excursion of the desired wavefront correction. It is defined as

$$\frac{Residual_{RMS}}{Desired_{PV}} \quad (3.17).$$

3.9.2.b Ratio residual – desired RMS

This figure of merit relates the residual error with respect to the RMS value of the desired wavefront correction. It is defined as

$$\frac{Residual_{RMS}}{Desired_{RMS}} \quad (3.18).$$

Table 3.4: Residual errors when commanding a Zernike focus terms, for the three different models.

Model	**Focus term peak-to-valley (nm)**	**Residual error peak-to-valley (nm)**	**Residual error RMS (nm)**	$\boldsymbol{\frac{Residual_{RMS}}{Desired_{PV}}}$
MARS	613.5	108.0	19.0	3.1 %
ANNs	562.8	157.2	31.7	5.6 %
ANNb	524.5	103.3	23.7	4.5 %

3.9.3 Reproducing Zernike Polynomials

The motivation for this set of experiments was to assess the ability of the models to reproduce known shapes, even though the training has been based on random shapes on the DM. For the first model implemented, the MARS model of the Xinetics, we requested a series of combinations of Zernike polynomials, picked arbitrarily. The results are shown in Fig. 3.19. We did not try to obtain residual errors on the shapes of Fig. 3.19, because we were not modelling the whole mirror, so the error would be dominated by boundary effects, not caused by the model. We can see in Fig. 3.19 the model reproduces the Zernike polynomials effectively.

We did a similar experiment with the Boston MEMS, but given the density of actuators, it was possible to ‘draw’ a complete Zernike focus term on the DM sector being modelled and obtain some residual errors. This is also interesting, as we can compare the results with a similar experiment performed in Stewart (2007), see table 3.6. Fig. 3.20 presents the results, as slices along rows and

columns, for the three non-parametric models we implemented for the Boston MEMS, a MARS, ANN small and ANN big.

Table 3.5: Boston MEMS results for the three different models and using the two figures of merit.

Model	**Desired excursion PV (nm)**	**Desired excursion RMS (nm)**	**Residual RMS (nm)**	$\frac{Residual_{RMS}}{Desired_{PV}}$	$\frac{Residual_{RMS}}{Desired_{RMS}}$
MARS	1320.8	320.0	40.6	3.1 %	12.7 %
ANNs	1339.0	327.5	63.2	4.7 %	19.3 %
ANNb	1303.6	321.2	46.8	3.6 %	14.6 %

The colour plots in Fig 3.20 have different amplitudes; this is a feature of the models, since they were trained with the same data set. This effect should be possible to calibrate by pre-processing the training data appropriately, although we did not attempt to do it. We removed the effect by adjusting the theoretical focus term in each case to minimize the residual error and thus leave systematic effects out. The results are summarized in table 3.4.

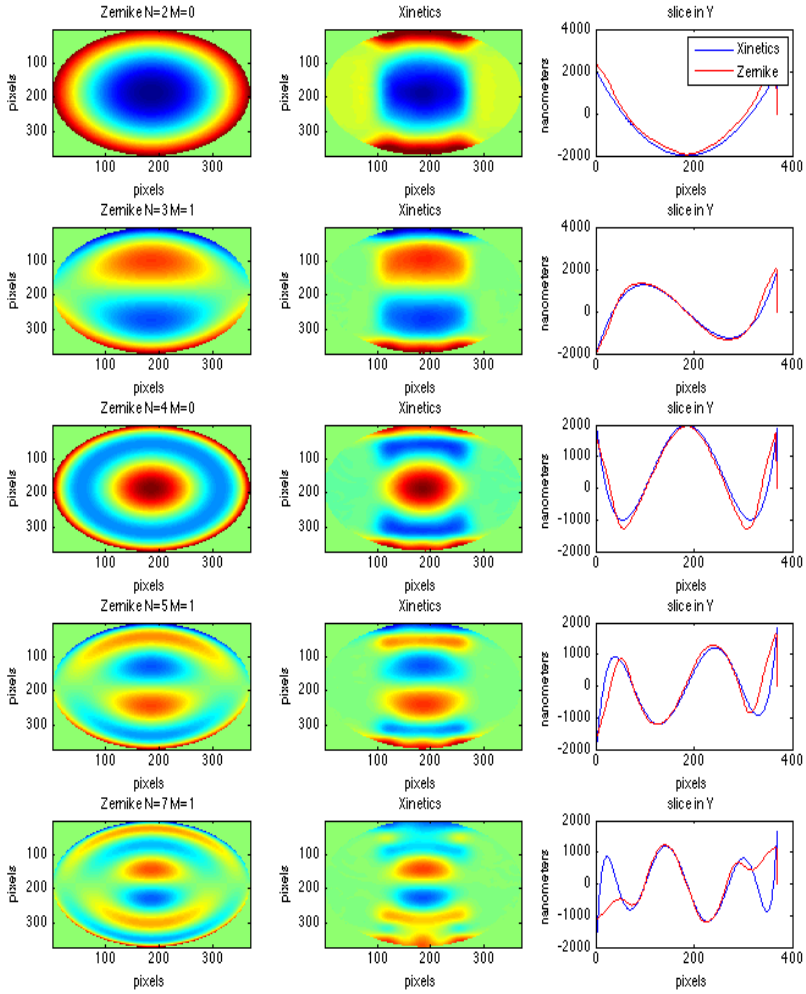


Fig. 3.19: Results using various Zernike polynomials. The left panels are theoretical Zernike polynomials; the central panels are the output of the DM for the 11 x 5 actuators being modelled. The right panels present a slice in Y from the central panels, to appreciate the differences between theoretical and experimental polynomials (red plot: Zernike polynomial; blue plot: DM surface). The colour map for the left and central panels is common, but it has not been included in order to simplify the figure.

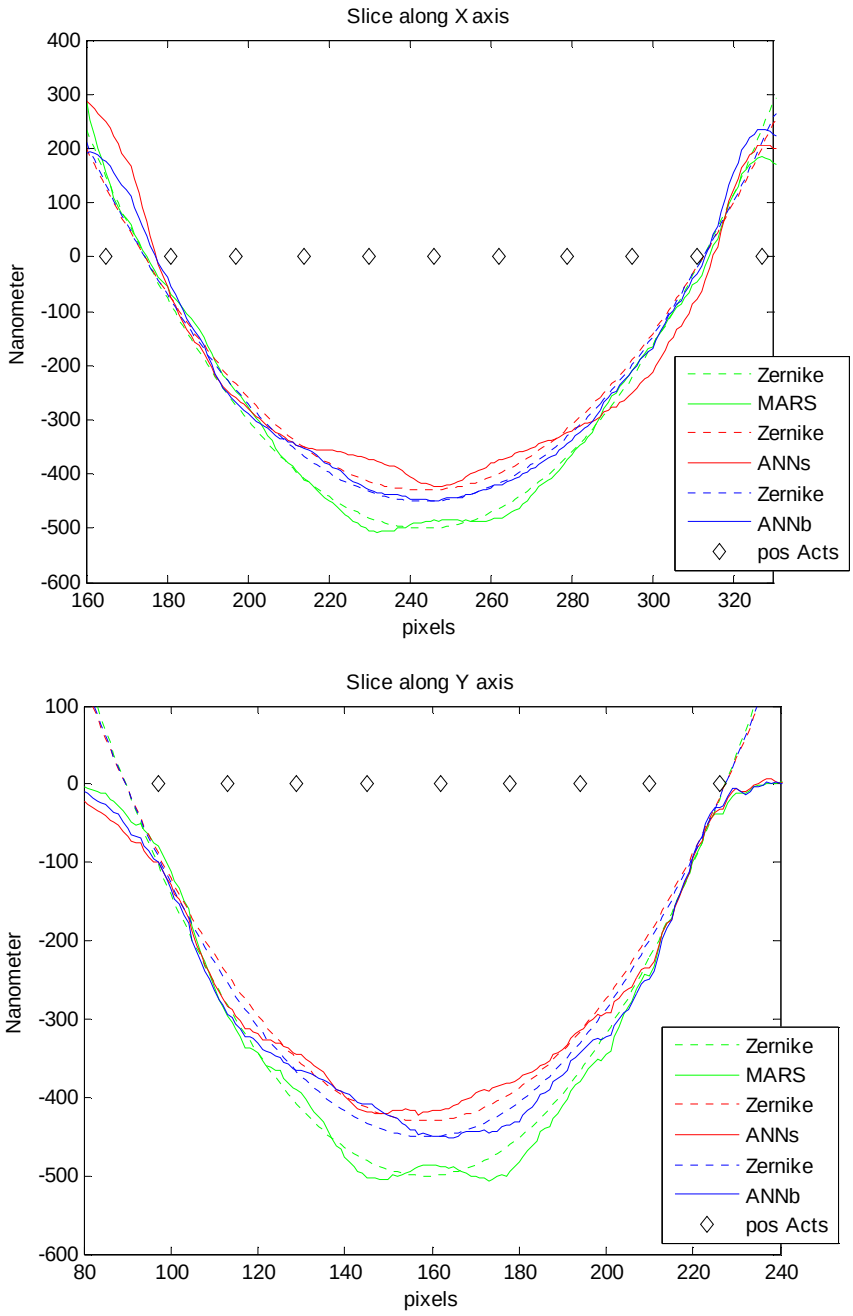


Fig. 3.20: Focus term ‘drawn’ by the MEMS DM. Slices in *X* and *Y*.

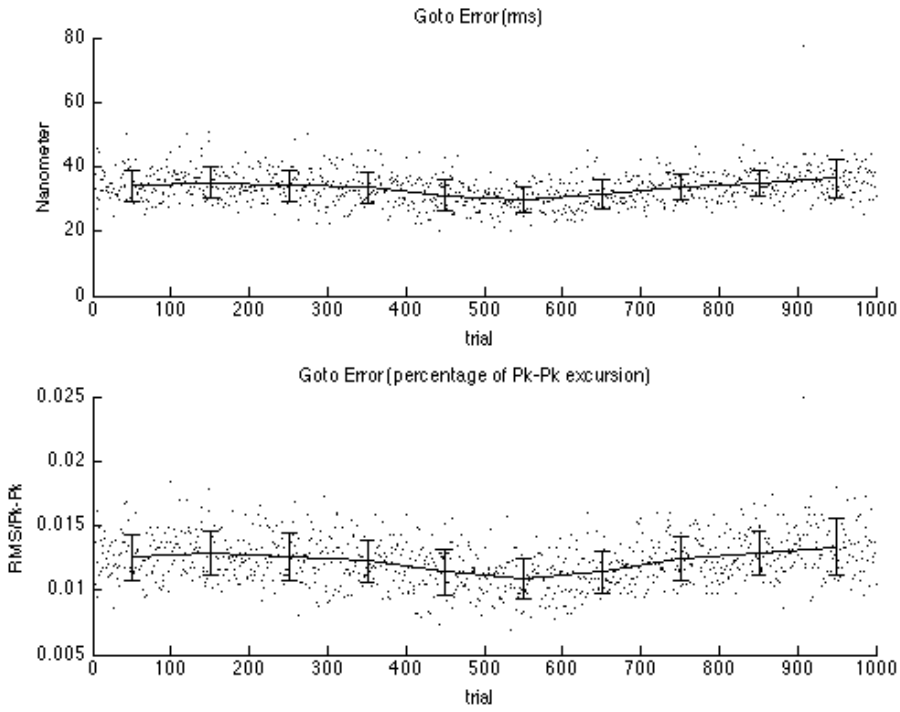


Fig. 3.21: Residual error for the random run with the Xinetics and MARS model: top panel, error expressed in nanometres; bottom panel, error expressed as a percentage, using the figure of merit in Eq. 3.17.

3.9.4 Reproducing Random Phases

The second phase of experiments consisted in requesting the model to shape the DM membrane following new random patterns generated by the same DM in a previous run (see section 3.8 for details of the implementation). Blain (2010) performed a similar experiment but reproducing turbulent phase screens. The difficulty of this approach is that the *DM fitting error*[53] term appears as an additional error term, not associated with the DM model under evaluation. By using individual random numbers on each actuator we prevented the DM from incurring a fitting error when trying to shape its surface to follow a pupil-wide pattern.

The results for the Xinetics' MARS model for 1000 random points are presented in Fig. 3.21. The average residual RMS is 1.2 %, using the figure of merit in Eq. 3.17.

[53] The DM fitting error takes into account the limited spatial resolution that a deformable mirror encounters when trying to shape its surface. This error term depends on the number of actuators principally.

Boston MEMS' MARS and ANN models results are presented in Fig. 3.22, for 3000 random points.

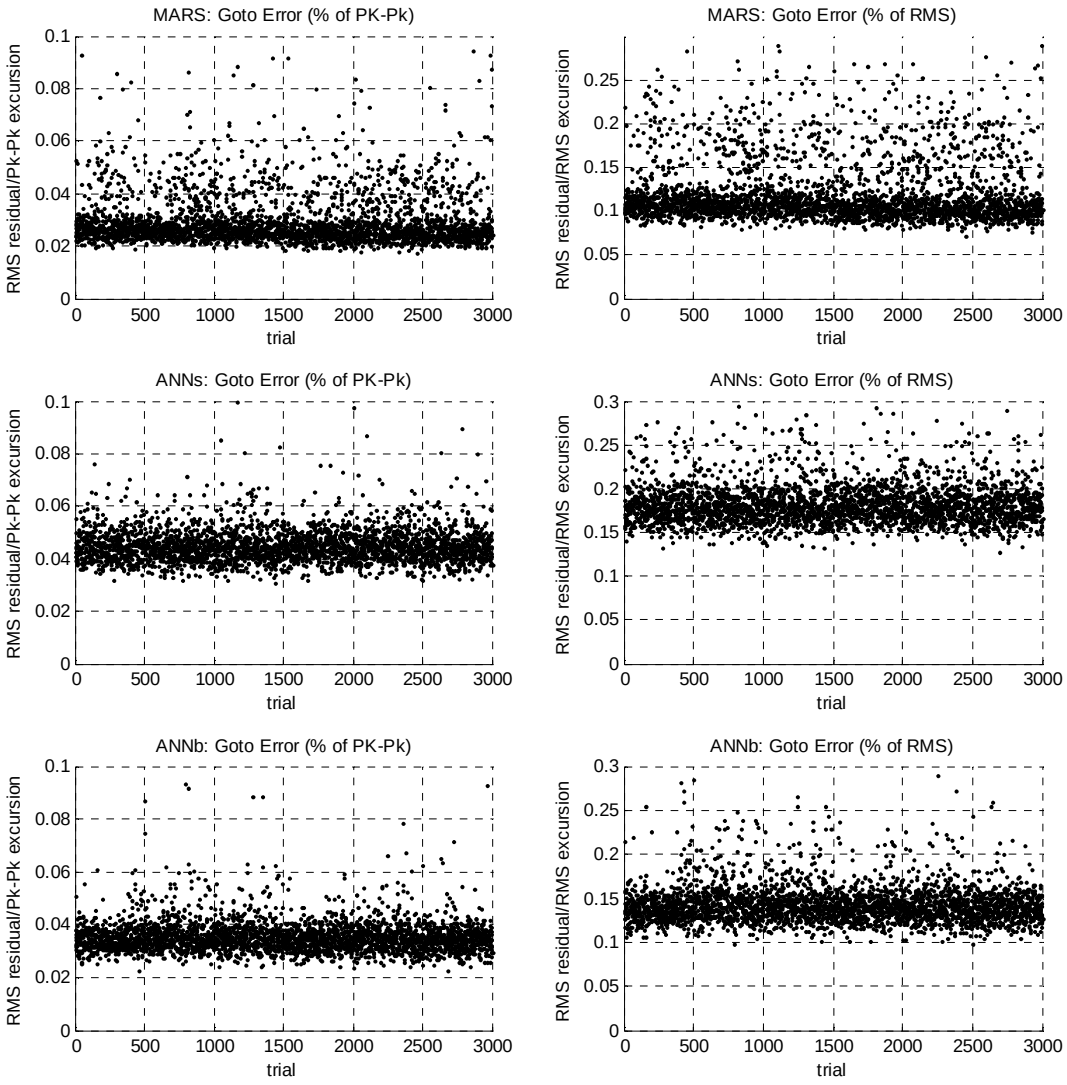


Fig. 3.22: Residual error for random run with Boston MEMS: MARS model in top panels; ANN small in middle panels and ANN big in bottom panels. Residual error using figure of merit in Eq. 3.17 in left panels and using Eq. 3.18 in right panels.

3.9.5 Comparison with Previous Works

As we described above, we performed two types of experiments, Zernike polynomial (focus term) and random shapes with our mirrors, to compare with similar experiments found in the literature. The other relevant works, Stewart (2007) and Blain (2010) experimented with Boston MEMS DM as well, so we compare our results and theirs in Tables 3.6 and 3.7:

Table 3.6: Comparison between our non-parametric models and Stewart's model, using figure of merit: $\mathrm{Residual}_{\mathrm{RMS}}/\mathrm{Desired}_{\mathrm{PV}}$.

Model	$\boldsymbol{\frac{Residual_{RMS}}{Desired_{PV}}}$
Stewart	2.7 %
MARS	3.1 %
ANN	4.5 %

Table 3.7: Comparison between our non-parametric models and Blain's model, using figure of merit: $\mathrm{Residual}_{\mathrm{RMS}}/\mathrm{Desired}_{\mathrm{RMS}}$.

Model	**Size of the random set**	$\boldsymbol{\frac{Residual_{RMS}}{Desired_{RMS}}}$
Blain	100	11.1 %
MARS	3000	12.7 %
ANN	3000	14.6 %

3.10 Conclusions

This chapter presents our efforts to devise an alternative modelling approach to control deformable mirrors for multi-object adaptive optics. Instead of using a model based on physical parameters, which would be only useful for a certain type of DM, we have utilized generic non-parametric estimation techniques and use them successfully with two different types of mirrors. The Xinetics and Boston MEMS only share in common a quadratic relationship between actuator voltage and stroke, but the rest of their characteristics are very different. In spite of this, our models based on

multivariate adaptive regression splines and artificial neural networks perform at similar level than previous physical models.

We can see that MARS performs extremely well for all models, although the Boston MEMS model has a significant number of 'outliers' (see Fig. 3.22). On the other hand, the ANN big model does perform marginally worse than MARS, but the number of outliers is limited. This is an encouraging result, which points us to consider refining the modelling technique using ANN for MEMS DMs, a topic to explore in future research.

3.11 References

Andersen, D., Fischer, M., Conan, R., Fletcher, M., Veran, J., "VOLT: The Victoria Open-loop Testbed", SPIE Proceedings 7015, pp 70150H-70150H-11, (2008)

Blain, C., Conan, R., Bradley, C., and Guyon, O., "Open-loop control demonstration of micro-electro-mechanical-system MEMS deformable mirror", Optics Express 18, 6, pp. 5433-5448 (2010)

Evans, J., Macintosh, B., Poyneer, L., Morzinski, K., Severson, S., Dillon, D., Gavel, D. and Reza, L., "Demonstrating sub-nm closed loop MEMS flattening", Optics Express, 14, 12, pp. 5558-5570, (2006)

Friedman, J., "Multivariate adaptive regression splines (with discussion)", Ann. Stat. 19, pp. 1-141, (1991)

Friedman, J. and Roosen C., "An Introduction to Multivariate adaptive regression splines", Statistical Methods in Medical Research, 4, pp 197-217, (1995)

Hom, C. , Dean., P. and Winzer, S., "Simulating electrostrictive deformable mirrors: I. Nonlinear static analysis", Smart Material Structures, 8, pp. 691-699, (1999)

Myers, R., Hubert, Z., Morris, T. *et al*, "CANARY: the on-sky NGS/LGS MOAO demonstrator for EAGLE", SPIE Proceedings 7015, pp. 70150E-70150E-9, (2008)

Nelder, J., and Mead, J., "A simplex method for function minimization", Computer Journal, 7, pp 308-313 (1965)

Stewart, J., Diouf, A., Zhou, Y. and Bifano, T., “Open-loop control of a MEMS deformable mirror for large-amplitude wavefront control”, JOSA-A, 24, 12, pp. 3827-3833, (2007)

Vogel, C. and Yang, Q., "Modelling, simulation, and open-loop control of a continuous facesheet MEMS deformable mirror," JOSA-A 23, 5, pp. 1074-1081 (2006)

Chapter 4: Figure Sensor

4.1 Overview

The Figure Sensor (FS) project consisted in developing a high-speed figure sensor to measure the shape of the CANARY deformable mirror, in order to confirm the position it reaches after being commanded by the CANARY AO computer. This chapter presents some aspects to be taken into consideration for the design of this sensor. An analysis of the expected performance is presented, along with simulation results implemented to understand the limitations in sensitivity that the FS can achieve. A lab experiment was implemented to confirm the sensitivity estimates.

4.2 Introduction

Open-loop adaptive optics was introduced in Chapter 1 and described in detail for the multi-object adaptive optics (MOAO) case in Chapter 2. In MOAO there is no optical feedback for the deformable mirror position, therefore the mirror must be controlled at minimum error but in open-loop. This can be implemented through a deformable mirror (DM) model, such as the ones presented in Chapter 3, or by adding a "Figure Sensor" in charge of measuring the shape of the DM surface continuously. This chapter is devoted to a preliminary study of such a sensor.

The figure sensor must measure the shape of the DM surface at a speed such that it is at least as fast as the AO computer commanding the DM. It should also obtain a map of the DM with at least the same spatial resolution used by the AO system, which corresponds to the resolution of the wavefront sensors measuring the atmospheric turbulence. The figure sensor should also measure the DM with a small error, so that when its measurements are incorporated in the AO control strategies[54], it does not impact the final error budget of the system. High speed, high spatial resolution and small error are therefore the main requirements for a figure sensor for a MOAO system.

[54] The figure sensor can be used for a local control loop in charge of positioning the DM, with set points defined by the AO computer.

As described in section 2.8, there are a number of groups implementing MOAO systems and they are considering the use of some type of figure sensor. For example, the "Victoria Open-loop Testbed" (VOLT) is an on-sky experiment in open-loop, measuring and correcting turbulence on-axis (Andersen 2008). The VOLT bench incorporates a "truth" wavefront sensor, which "provides an absolute reference of the DM shape" (Andersen 2008). They use a LED source to illuminate the DM face-on. The same group is now developing "RAVEN", a MOAO demonstrator for the 8 meter Subaru telescope in Mauna Kea, which would provide two MOAO channels (Conan 2010). RAVEN incorporates figure sensors for each DM.

To implement the figure sensor, some type of optical metrology instrument is required. A common instrument found in optical workshops for this purpose is an optical interferometer. Generally speaking, an interferometer is a very delicate instrument that needs careful alignment and vibration-free operation. Interferometers are usually slow in obtaining the resulting map of the surface, so, given the requirements and these limitations, we opted for using a Shack-Hartmann wavefront sensor (SH-WFS) as the figure sensor, since it comes with appealing characteristics for the project. In particular, a SH-WFS can be designed to be very compact and its principle of operation does not rely on wavefront interference so it is a robust instrument. If we have a fast CCD camera for image acquisition and we are not in a photon starvation regime[55], the exposure time is very short and therefore it is fast to compute the shape of the DM facesheet. Finally, for the purpose of this thesis, it was appealing to work with a SH-WFS, since it is the most common WFS in AO.

There are several high resolution SH-WFS commercially available, such as the ones offered by companies like Optocraft[56] (Germany) and Imagine Optics[57] (France), which have spatial resolution comparable to interferometers, but they all lack high-speed frame rates, imaging at no more than 30 frames per second.

Given these considerations, it appeared sensible to design and build a custom figure sensor for CANARY. This chapter is devoted to some important aspects of the sensor, from an AO error budget standpoint. The preliminary design concept of the FS is presented, followed by a description of

[55] Photon starvation is the case of SH-WFS in AO, which use starlight to measure the aberrations of the incoming wavefront.
[56] http://www.optocraft.de
[57] http:// www.imagine-optic.com

the SH-WFS. The relevant noise sources of the sensor are then introduced and discussed, followed by a brief introduction to charge-coupled devices (CCD). A basic wavefront reconstructor is introduced and some estimates of noise propagation are presented. A section with Monte Carlo simulations helps in understanding the trade-offs required when working at high-speed. The chapter ends with some laboratory experiments to confirm the sensitivity of the sensor, using a commercially available high-speed camera.

4.3 Figure Sensor Design

A design concept for the FS is presented in Fig. 4.1. A monochromatic light source (LED or laser) shines a collimated, off-axis beam onto the DM, which is reflected by the DM to illuminate the figure sensor optics, which consists of a lenslet array and a high-speed CCD. A science beam can be aligned to the DM as long as it is in a different angle with respect to the light source used by the FS.

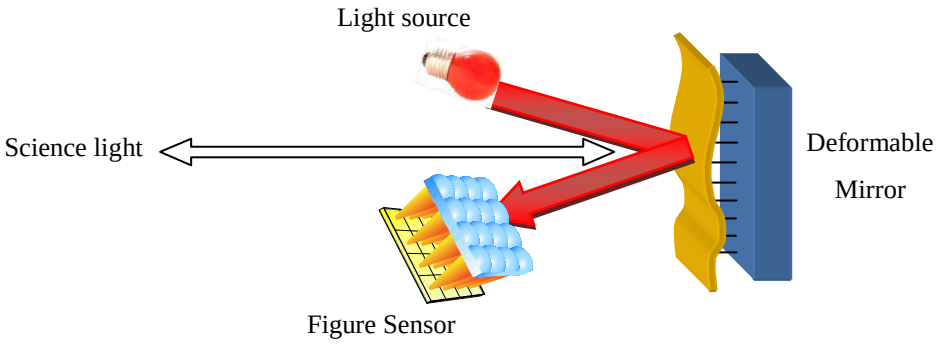


Fig. 4.1: Schematic view of the Figure Sensor to measure the shape of the deformable mirror surface.

4.4 Photon Noise

Light detection is a complex phenomenon. For the scope of this study, we will concentrate in considering light as composed of individual photons falling into a 'detector' device, which can produce a quantifiable measurement of the number of photons it is receiving in a given 'exposure' time. As many other processes in nature[58], the number of photons impacting a surface per unit of time

[58] The number of people arriving to a bank per hour is a typical example of a stochastic process, modelled with the same probability distribution as light.

is a stochastic process. For familiar wavelengths (optical and near infrared), they can be well modelled using a Poisson distribution (Janesick 2007), which has the expression

$$p(k,m) = \frac{m^k e^{-m}}{k!} \quad (4.1),$$

where p is the probability there are k occurrences in an interval (spatial or temporal) with an expected value (or mean) m.

The stochastic characteristic of the number of photons means that this process has to be accounted as a noise source, usually referred to as 'photon noise' or 'shot noise', the latter referring to the electrons generated on a detector by incident photons.

The images in Fig. 4.2 were produced with the simulation tool developed for this chapter (described in section 4.7). The figure shows two simulated CCD images with different expected levels, produced with a random number generator using a Poisson distribution.

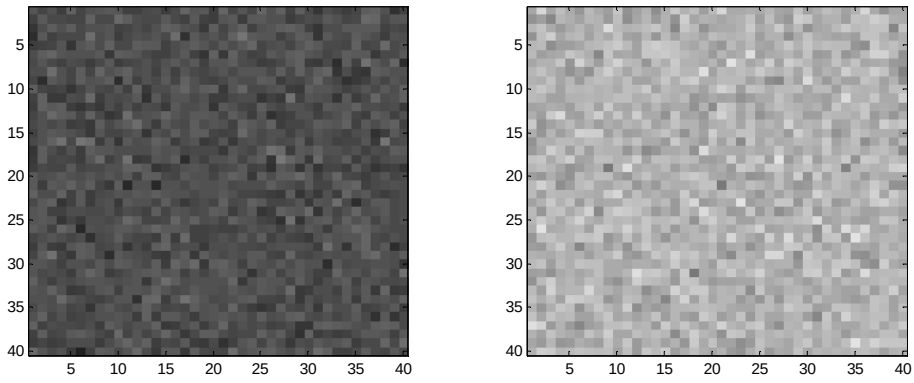


Fig. 4.2: Simulation of photon noise for two mean levels in a 40 x 40 pixels detector.

4.5 Shack-Hartmann Wavefront Sensor

4.5.1 Description

The original Hartmann screen mask was developed in 1904 by J. Hartmann to measure the optical aberrations in the 80 cm Potsdam refractor telescope (Wilson 1999). It consisted of a mask with holes installed in front of the objective lens, to produce a 'spot pattern' at the telescope focal plane. The Hartmann mask was later used in front of reflecting telescopes as well. This 'classical' Hartmann mask was further developed by Roland Shack at the University of Arizona in the late 60's,

who replaced the holes by an array of small lenses, allowing more efficiency in collecting photons and higher spatial sampling of the aperture (Platt 2001). Nowadays, it is the most commonly used wavefront sensor in adaptive optics, because of its easy calibration and consistent good results.

A SH-WFS consists of a lenslet array and a focal plane sensor, normally a CCD. Fig. 4.3 is an illustration of a SH-WFS, showing these two main components. The sensor images a point source in each subaperture, producing an array of spots. A distorted wavefront will produce light rays arriving at different angles, represented by the β angle in Fig. 4.3, causing a departure of each spot position from the location it would have had for an undistorted wavefront. This departure is shown as the subaperture position offset (Δx, Δy) in the figure. Wavefront aberrations are measured by accurately locating the spot positions across the CCD. This process is usually called 'centroiding' and there are various algorithms to do it. The wavefront can be retrieved by using a reconstruction algorithm, fed with the spot positions.

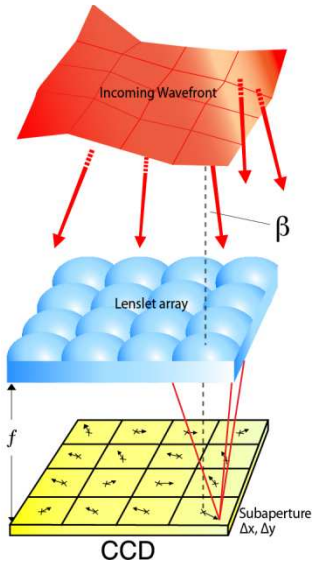


Fig. 4.3: A schematic diagram of a Shack-Hartmann wavefront sensor. A single subaperture is made up of an individual lens and the CCD pixels behind it.

4.5.2 Centroiding Algorithms

The simplest and most direct way to estimate the position $(\hat{x}, \hat{y})$ of the spot in each subaperture is to apply the Centre of Gravity (CoG) algorithm, expressed mathematicallyis

$$\hat{x} = \frac{\sum x I_{x,y}}{\sum I_{x,y}} \qquad \hat{y} = \frac{\sum y I_{x,y}}{\sum I_{x,y}} \quad (4.2),$$

where:

x is the x-coordinate of the (x,y) pixel

y is the y-coordinate of the (x,y) pixel

$I_{x,y}$ is the intensity of the (x,y) pixel

There are other centroiding algorithms, briefly described here:

- Thresholding Centre of Gravity (TCoG): it is similar to CoG, but a threshold is applied before computing Eq. 4.2, with the goal of including only pixels with a high flux. This helps in reducing the effect of background noise in the subapertures. We briefly explore its effect in the simulation section of this chapter.
- Weighted Centre of Gravity: it is similar to TCoG, but the threshold is a function of the flux level on each pixel – a kind of 'soft' thresholding (Thomas 2006).
- Correlation: this method uses the principle of correlation between the spot and a fixed template. It is also known as a 'matched-filter' in other areas of science and engineering. This method yields the best results for extended spots (Gilles 2006), for example in the case of laser guide stars in ELTs, where the cone effect produces elongation of the spots in peripheral subapertures.

4.5.3 Measurement Error in CoG

For the theoretical results in this section we assume the irradiance pattern of the spot in each subaperture to be approximately Gaussian. The error or noise in estimating the spot position comes from two main components:

- Photon noise, associated with the probabilistic arrival of photons in each subaperture
- Electronic noise, associated with detector effects, most importantly readout noise (RON). These are described in detail in the next section.

The contributions from each of these effects are mathematically (Thomas 2006)

$$\sigma^2_{\phi,N_{ph}} = \frac{\pi^2}{2ln2}\frac{1}{N_{ph}}\frac{{N_T}^2}{N^2_{samp}} \quad (rad) \quad (4.3),$$

$$\sigma^2_{\phi,N_r} = \frac{\pi^2}{3}\frac{N_r^2}{N_{ph}^2}\frac{N_s^4}{N^2_{samp}} \quad (rad) \quad (4.4),$$

where:

$\sigma_{\phi,N_{ph}}$: photon noise

σ_{ϕ,N_r} : electronic noise

N_{ph} : average number of photons per subaperture and per frame

N_r : readout noise

N_T : full-width at half-max of the spot

N_s : size of the subaperture in pixels

$N_{samp} = \frac{\lambda}{dp}$: relates the angular pixel size (p) to the half-width of the diffraction-limited spot, ($\frac{\lambda}{d}$) , where λ is the mean wavelength and d is the subaperture diameter

For the purposes of this study, we are not interested in the error in radians, but the error in pixel units at the SH focal plane. The relationship between the two is (Thomas 2006)

$$\sigma_x = \frac{N_{samp}}{2\pi}\,\sigma_\phi \quad (4.5).$$

Substituting Eq. 4.5 in Eqs. 4.3 and 4.4, we obtain the total measurement error

$$\sigma_x = \sqrt{\frac{1}{8ln2}\frac{N_T^2}{N_{ph}} + \frac{1}{12}\frac{N_r^2 N_s^4}{N_{ph}^2}} \quad (4.6).$$

We use Eq. 4.6 when experimenting with different values for the design parameters of the SH sensor, such as spot size and dynamic range. We will confirm the theoretical result in Eq. 4.6 when performing Monte Carlo simulation of a subaperture.

4.5.4 Dynamic Range, Spot Size and Sensitivity

SH-WFS in AO systems working in closed-loop measure approximately flat wavefronts. In such a case, spot positions on the SH focal plane are centred in their subapertures, therefore the dynamic range required by the sensor is limited. In the FS case, measuring the DM facesheet implies the spots can be anywhere in the subaperture, therefore the dynamic range comes to play an important role when evaluating the sensitivity of the FS.

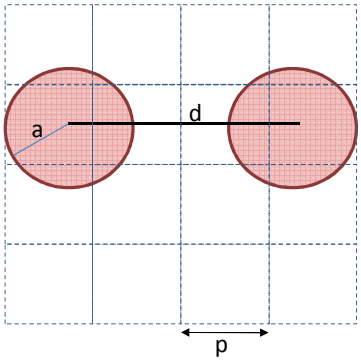


Fig. 4.4: Dynamic range (*d*) as a function of spot size (*a*) and pixel size (*p*).

The dynamic range problem is represented in Fig. 4.4. The spot[59] positions represented in the figure are the two limit[60] positions for the spot when displaced in the *X* coordinate. The dynamic range *d* depends on the number of pixels of size *p* in the subaperture and the spot radius *a*. Using a wider spot implies a reduction in dynamic range. The available dynamic range from the figure is

$$d = np - 2a \quad (4.7),$$

where *n* is the number of pixels per subaperture.

Sensitivity in a SH-WFS can be defined as the sensor's minimum measurable displacement in spot position. This would be ultimately limited by the quantum nature of light and readout noise, and is expressed by the *measurement error* of Eq. 4.6. When a SH works as a Figure Sensor, it must be able to measure the whole range of possible positions for the DM facesheet; therefore the FS should be designed to span the complete dynamic range of the DM actuators stroke. The FS sensitivity is the minimum displacement of the DM facesheet the sensor is able to measure. The minimum displacement of the DM facesheet is defined by the resolution at which the actuators are driven. The FS sensitivity (FS_s) can be defined then as

$$FS_s = \frac{d}{DM_b} \quad (4.8),$$

[59] The diagram in Fig. 4.4 is a simplified version of a spot, since in reality the spot follows an irradiance profile.
[60] The position is a limit in the sense the spot can only be sampled by the subaperture pixels.

where DM_b is the resolution at which the DM facesheet displaces, when commanded by the DM actuators. DM_b is ultimately the resolution of the digital-to-analog converters driving the amplifiers that move the actuators[61].

As an example, table 4.1 presents some preliminary design parameters of the FS for a Xinetics DM, based on an optical design done for CANARY:

Table 4.1: Design parameters of the FS for the Xinetics DM.

Parameter	**Value**	**Units**
Spot radius (a)	1.4	pixels
Subaperture size (n)	8	pixels
Dynamic range (d)	5.2	pixels
DM resolution (DM_b)	4096	levels
Sensitivity (FS_s)	0.0013	pixels
Measurement error (σ_x)	0.0015	pixels

The measurement error in Eq. 4.6 was obtained for a Gaussian spot. However, diffraction patterns in a SH-WFS follow Fraunhofer's diffraction patterns for a circular or a square aperture, which respectively corresponds to the well-known formulae (Driggers 2003):

$$E(x,y,z) = E_A\left(\frac{A}{\lambda z}\right)^2 \frac{sin^2\left(\frac{\pi a x}{\lambda z}\right) sin^2\left(\frac{\pi b y}{\lambda z}\right)}{\left(\frac{\pi^2 abxy}{\lambda^2 z^2}\right)^2} \quad (4.9),$$

$$E(r,z) = E_A\left(\frac{A}{\lambda z}\right)^2 \left[\frac{2J_1\left(\frac{\pi r D}{\lambda z}\right)}{\frac{\pi r D}{\lambda z}}\right]^2 \quad (4.10),$$

where:

E_A : Irradiance at the aperture

A : area of the aperture

D : diameter of the aperture (in Eq. 4.10)

a, b : width and height of the aperture (in Eq. 4.9)

J_1 : first-order Bessel function of the first kind

λ : working frequency

[61] Most DM need high voltages to displace their actuators (see table 3.1), however it is not the case of magnetic actuators.

x, y, r : spatial coordinates at the observation plane (perpendicular to the optical axis)

z : distance between observation plane and aperture (generally the focal length of the lens)

To use Eq. 4.6 for computing the measurement error in Table 4.1, we found the equivalent full-width at half-maximum (FWHM) for a Gaussian spot that would produce the closest irradiance pattern with respect to the spot produced by a square aperture (the most common one in a SH-WFS). To do this, we performed a quick simulation (in Fig. 4.5) varying the FWHM of a 2D Gaussian pattern and comparing it with a Fraunhofer irradiance pattern with unitary radius[62] (Eq. 4.9), and normalized to the total irradiance for the Airy disk region. The FWHM obtained (shown in the title of the plot) was used to obtain the measurement error[63] in Table 4.1. No readout noise was added.

As it can be seen in Table 4.1, the required sensitivity (FS_s) is marginally lower than the measurement error; therefore the sensitivity of the FS would be limited by the measurement error rather than the DM resolution. In such a case, we should pursue increasing the required sensitivity to get above the photon noise floor.

For a given spot size and measurement error, the design parameter to be modified to increase the required sensitivity is the number of pixels per subaperture. Fig. 4.6 shows comparative results for sensitivity at different subaperture sizes. From this figure, we can see that it is advisable to use subapertures of 9 pixels in size or more. However, this implies that the SH-WFS image becomes larger and therefore it may take longer to readout. A value of 9 pixels per subaperture is considered a good compromise between sensitivity and subaperture size.

[62] Having $a = 1$ in the Fraunhofer pattern allows one to determine a general relationship between FWHM of a Gaussian and a.

[63] We assumed photon noise only, with 10^5 photons in the subaperture.

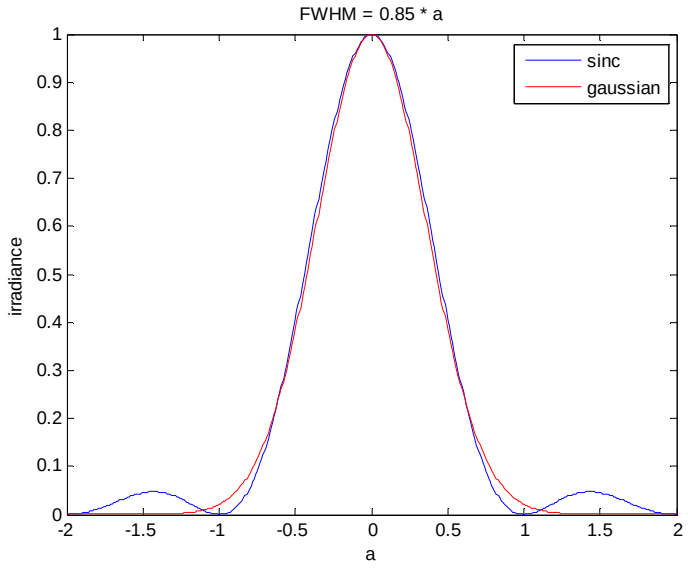


Fig. 4.5: Square aperture irradiance pattern compared with a Gaussian pattern which produces the same irradiance for the Airy disk.

If readout noise is included in the simulation of Fig. 4.6, the sensitivity defined by Eqs. 4.7 and 4.8 is not able to intersect the readout noise floor (see Fig. 4.7 as an example for a N_r = *10* electrons). Nevertheless, in such a case, thresholding centre of gravity (TCoG) can be applied to remove the effect of readout noise, in which case the effect is similar to Fig. 4.6.

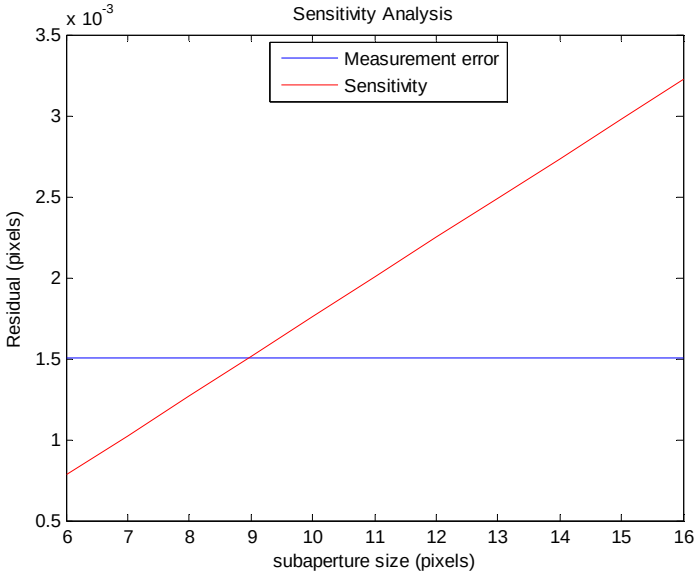


Fig. 4.6: Measurement error and FS sensitivity for different subaperture sizes, with zero readout noise.

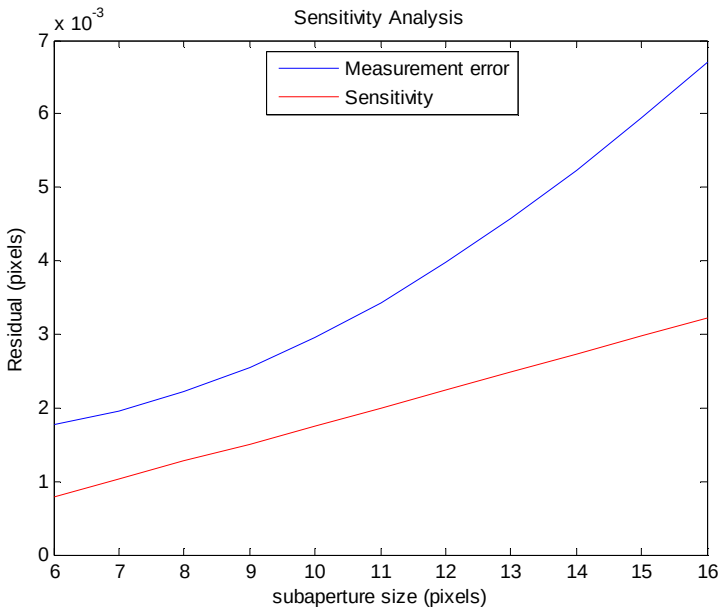


Fig. 4.7: Measurement error and FS sensitivity for different subaperture sizes, with readout noise of 10 electrons.

4.6 High-Speed CCDs

The CCD has been the preferred detector technology in astronomy since the 1990's, because of its high efficiency in collecting photons and excellent linearity (Janesick 2001). It is often the chosen detector technology for Shack-Hartmann wavefront sensors as well, where a high efficiency and high speed are the relevant requirements.

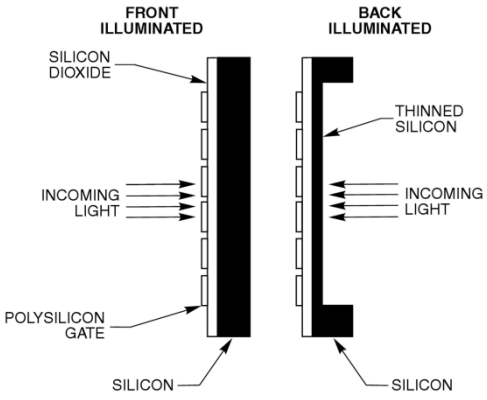


Fig.4.8: Front and back-illuminated CCDs. Reproduced from the Kodak web site

It is beyond the scope of this chapter to fully describe CCD characteristics, so we will only focus attention on those relevant characteristics for high speed operation. A full discussion of CCD imaging devices can be found in Janesick (2001) and Janesick (2007).

The relevant characteristics for high speed operation are:

Front/Back illumination and Fill-Factor: Diagrams of front-illuminated and back illuminated CCDs are presented in Fig. 4.8. The fundamental difference that results in a higher efficiency for back-illuminated CCDs, is the side of the device that receives incoming photons: in front-illuminated CCDs there are light-blocking layers (polysilicon and silicon dioxide) for reading out the charge, therefore photons arriving at those sectors are lost. On a back-illuminated CCD, virtually all photons are collected and generate electrons on each pixel. In order to work, a back-illuminated CCD's silicon is thinned, an unnecessary process in case of front-illuminated CCDs.

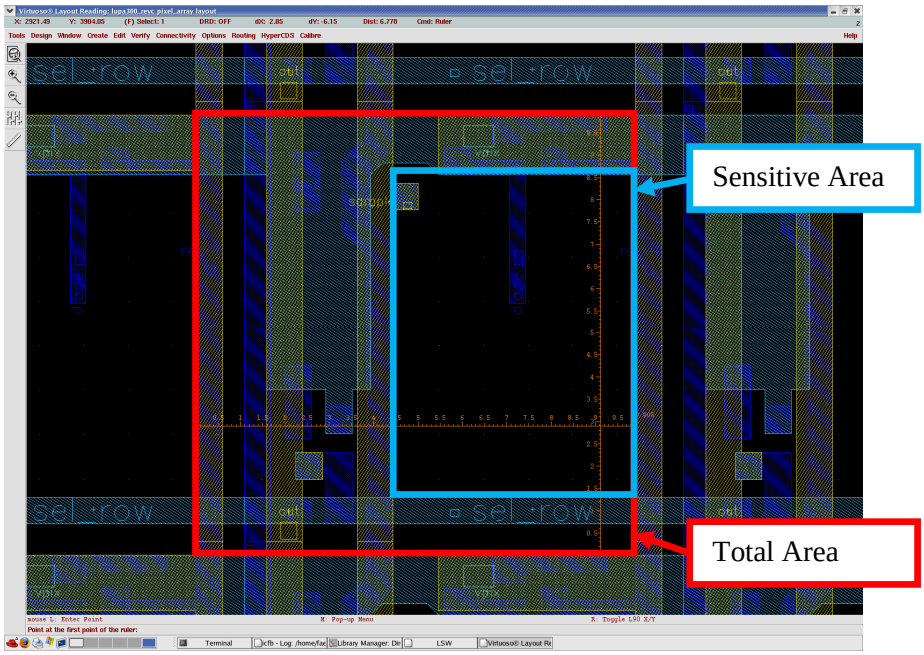


Fig. 4.9: Image from a software design application for CCDs, showing the sensitive area of a front-illuminated pixel (in blue) with respect to the total pixel surface (in red). The ratio of surfaces for this device is around 40 %. Image courtesy Cypress Semiconductors.

The poor efficiency of front-illuminated CCDs (only 40% of the photons can be collected in the CCD design depicted in Fig. 4.9) can be increased by adding micro-lens technology. This is a

layer of tiny convex-plane lenses deposited in front of each pixel, enlarging the collecting area of each pixel, as presented in Fig. 4.10.

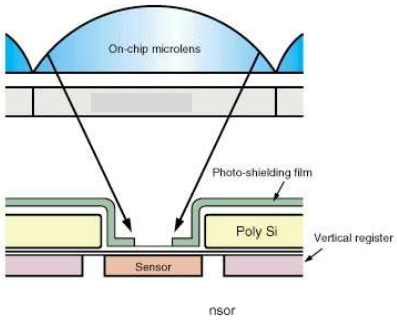


Fig. 4.10: Microlens on top of a front-illuminated pixel. Reproduced from the Kodak web site.

The ratio between the sensitive area of each pixel and its total area is called ‘fill-factor’ (FF) and plays a fundamental role when acquiring tiny irradiance patterns as the ones found in a SH WFS. Microlens technology allows increasing the FF up to around 90 %.

Readout Noise: CCD readout is the process of acquiring the image formed by electron accumulation in each pixel, from the CCD to a digital computer. In standard CCDs, readout is implemented by “moving” the electrons from row to row, and hence to a “serial register” at the edge of the device, from where each pixel can be read using an output amplifier, located at the corner of the device (see Fig. 4.11). In the output amplifier the electrons are converted to a voltage, which is sampled and digitized by external circuitry.

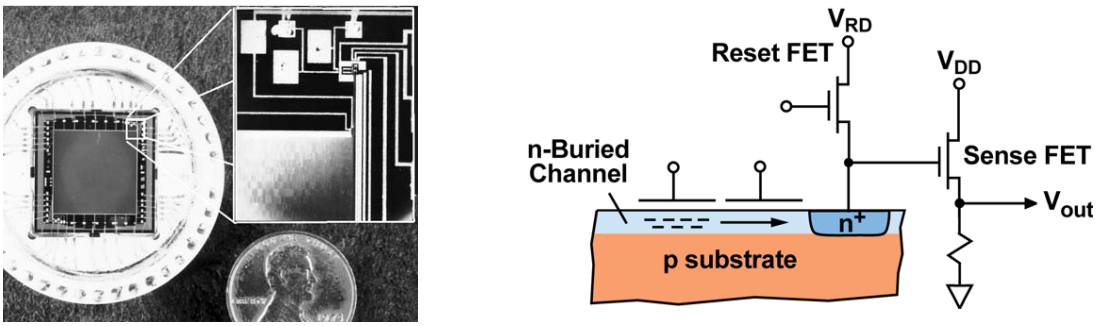


Fig. 4.11: Output amplifier imprinted at the corner of a CCD (left image) and equivalent electronic circuit (right image), composed of FET transistors. Images courtesy Texas Instruments and Lincoln Laboratories.

The output amplifier is the main source of noise in CCDs. This noise is known as '*Readout Noise*' (RON) and it is purely electronic. Sources of electronic noise are typically thermal noise as well as shot noise. As in other areas of electronics, the readout speed impacts RON, as it can be seen in Fig. 4.12, reproduced from a CCD detector datasheet used in astronomy.

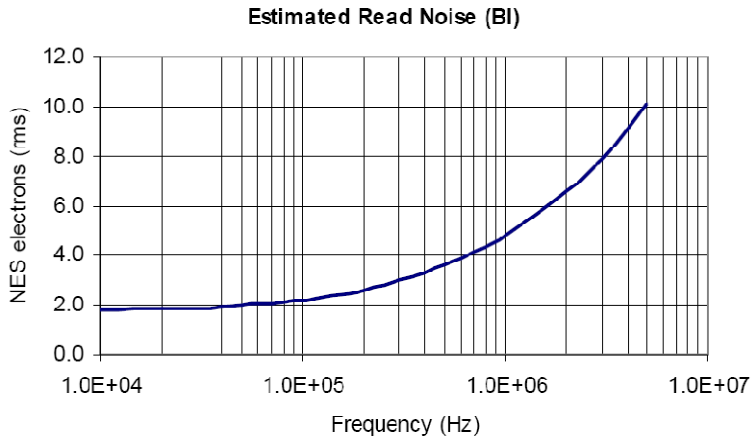


Fig. 4.12: Readout Noise for different pixel frequencies. Plot reproduced from a typical E2V CCD datasheet.

From Fig. 4.12, it is advisable to read a CCD at a low speed, i.e. spending a 'long' time per pixel. However, this makes the read time of the whole CCD too long for high speed applications; increasing the pixel readout frequency implies climbing up the line of Fig. 4.12 to higher RONs.

CCD sampling resolution: CCDs produce a voltage at the output amplifier, which needs to be sampled and digitized by external electronics. The contribution of this electronics is fundamental to maintain low electronic noise. The selection of sampling resolution is also important to have the correct measurement of the true number of electrons on each pixel. In astronomy, 16 bits analog-to-digital (A/D) converters are standard, but these devices run at slower sampling frequencies with respect to lower resolution converters. The FS requires high speed sampling, so a lower than 16 bits resolution converter may need to be selected. As we will show later in this chapter, we selected a commercial CCD camera that delivers digitized pixels and comes with 8 and 10 bits converters. The resolution of the converter is another noise source, called '*quantization noise*', since the true measurement gets distorted to a degree dependent upon the resolution of the converter. We explore the effect of converter resolution in the simulations section below.

4.7 Simulations

We have implemented Monte Carlo simulations of a single SH subaperture in order to understand the implications of using a high-speed (i.e. with high RON) and front-illuminated (i.e. with poor FF) camera.

4.7.1 Description of the Simulation Tool

The simulation tool implemented generates images of one Shack-Hartmann subaperture of 9 x 9 pixels, such as the one in Fig. 4.13. The tool generates images using the following procedure:

1) Poisson generator + Gaussian irradiance: each pixel in the subaperture is divided into 'high-resolution' (HR) pixels and a Poisson generator defines the total number of photons per HR pixel, which are then weighted by the amplitude of a Gaussian irradiance pattern. The tool allows defining the Full-Width at Half-Maximum (FWHM) of the Gaussian pattern.
2) Monte Carlo generation of photons: the value found in (1) for each HR pixel is used as the total number of photons that are spread by a random generator within each HR pixel.
3) The total number of photons per pixel in the subaperture is counted, masking out all those pixels outside the sensitive area of each pixel. The simple geometry shown in Fig. 4.14 is used and a fill-factor (FF) can be defined as a percentage for the ratio of the sensitive area to the total pixel area.
4) The accounted photons per pixel are converted to electrons and then sampled by an Analog-to-Digital Converter (ADC), whose resolution can be set as a parameter of the tool. The ADC output is in ADU (Analog-Digital Units).
5) Readout noise (RON) is added to the ADC output per pixel, using a random generator with a Gaussian distribution. A bias value is added to the ADC+RON value obtained.

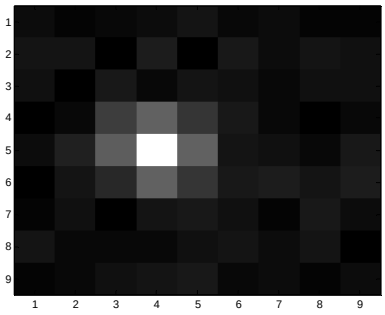


Fig. 4.13: Simulated Shack-Hartmann spot using a Gaussian irradiance pattern. The simulation includes added readout noise.

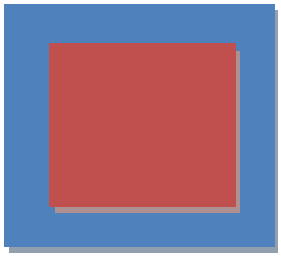


Fig. 4.14: Representation of each pixel. The red area is sensitive to photons, while the blue area is not. The ratio between these areas is the fill-factor of the pixel.

4.7.2 Photon Transfer Curve and Detector Gain

The standard method to determine many parameters in a detector is the 'photon transfer curve' (Janesick 2007). It is beyond the scope of this chapter to describe this method in detail, but it is pertinent to give a very brief description for the purpose of this section.

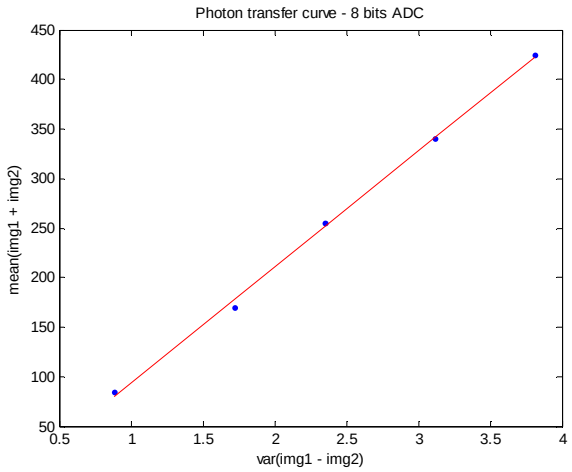


Fig. 4.15: Photon transfer curve for the 8 bits ADC. Gain = 117.1 e- / ADU.

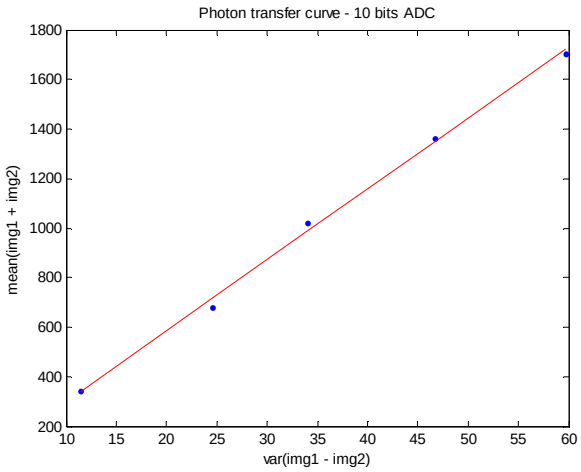


Fig. 4.16: Photon transfer curve for the 10 bits ADC. Gain = 28.6 e- / ADU.

A useful characteristic of a Poisson distribution is that its *mean* is equal to its *variance*. So, when taking images with a CCD detector at increasing levels of light, the variance of the pixel values will increase at the same rate as the mean value. Using this fact, it is possible to obtain the conversion factor between incoming photons and number of ADUs measured by the device. The details of this

method can be reviewed in Janesick (2001, 2007). Figures 4.15 and 4.16 show photon transfer curves for the two ADCs simulated and the corresponding value for the conversion factor, usually called 'gain'. These plots were obtained with a simplified version of the simulation tool, which only produces the total number of photons per pixel using a Poisson distribution. The subaperture size was increased from 9 x 9 to 30 x 30 in order to have a good statistical basis to compute mean and variance. This result is relevant since it confirms the simulation is producing realistic Poisson statistics.

4.7.3 Simulation Parameters and Figure of Merit

The simulation tool receives a number of inputs, which will be varied when exploring the influence of the design parameters. Some of the parameters were defined following the available settings in the high-speed camera used in the experiments (see Section 4.9) and thus we did not attempt to explore a vast parameter space since it would not have a real application. A good example of this is the analog-to-digital converter (ADC), which can be selected to be only 8 bits or 10 bits.

The parameters space for the simulations is detailed in table 4.2:

Table 4.2: Parameter space for the Monte Carlo simulation of a FS subaperture.

Parameter	**Units**	**Type**	**Range**
Mean number of photons[64]	unit	Optical	0 – 30,000
Full-Width at Half-Max (FWHM) of irradiance distribution	pixels	Optical	1.0 – 3.0
Fill Factor (FF) of each pixel	ratio	Optical	0.4 – 1.0
Analog-to-Digital Converter (ADC) resolution	bits	Electronic	8 – 10
Readout Noise (RON)	ADU	Electronic	0 – 3

As seen in table 4.2, we are defining two types of parameters:

- Optical: These parameters can be modified by means of the optical design and CCD selection[65]

[64] This is limited by the 'full-well' of each pixel, a parameter of CCDs.
[65] In practice, there are a very limited number of high-speed CCD detectors to choose from.

- Electronic: These parameters can be defined by design when selecting the readout electronics[66]

We explored the parameters space for the two types of parameters independently, in order to obtain results that are simpler to analyse.

The figure of merit used was the root-mean-square of the residual error, defined as

$$Residual_{err} = \sqrt{\frac{\sum(x\prime - \bar{x})^2}{n}} \qquad (4.11),$$

where:

x' : true position of the spot, after *gain factor* is computed (*)

$\bar{x}$: estimated position of the spot, using some centroiding algorithm

n : number of images in the run

(*) : The centroiding output may have a linear scale factor with respect to the *true* position of the centroid, so there is a *gain factor* that needs to be taken into account. This is described in Fig. 4.17. The gain factor is such that the slope of the lines is matched.

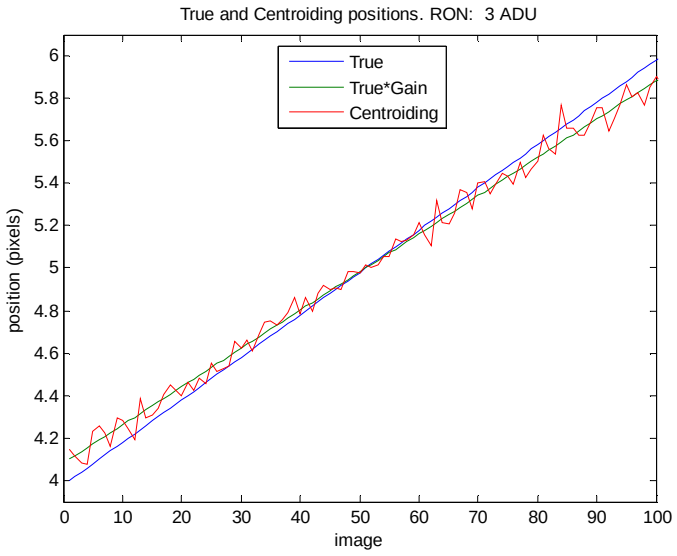


Figure 4.17: The true position of the spot (in blue) and the measured position from centroiding (in red) for a run of 100 images, varying the spot position between pixels #4 and #6 (in X). The green line is a linear regression from the centroiding data. A gain factor can be computed from the green and blue lines, so the slope of the green line matches the slope of the blue line and thus the measured and true position are similar.

[66] Similarly, high-speed cameras come with built-in electronics, so some of these parameters are fixed.

4.7.4 Parameters Space Exploration

For all simulations, the spot was displaced between pixels #4 and #6 (in *X*), of the 9 pixels in the subaperture, with a displacement resolution of 200 steps. The *Y* coordinate was kept constant at pixel 5 (centre of the subaperture). These dynamic parameters are summarized in table 4.3:

Table 4.3: Dynamic parameters for the simulation.

Parameter	**Value**	**Units**
Subaperture size	9 x 9	Pixels
Spot position (X)	$4 \rightarrow 6$	Pixel
Spot position (Y)	Fixed at 5	Pixel

We did not experiment with varying the number of photons in the subaperture, since the FS application will have a dedicated light source, so there will be plenty of photons to maintain a bright spot on each subaperture.

The following sections present the results when varying the optical and electronic parameters, which are relevant when selecting and using a high speed CCD camera.

4.7.4.a Optical Parameters

We explored the effect of varying the size of the spot as well as the fill-factor. The residual error is calculated as in Eq. 4.11. The plots in Fig 4.18 present the residual errors (on the title of each plot) for different combinations of FWHM and FF. Fig. 4.19 presents a surface, as the residual error from combinations of FF and FWHM. The electronic parameters were kept constant, ADC at 8 bits and RON at 0 ADU.

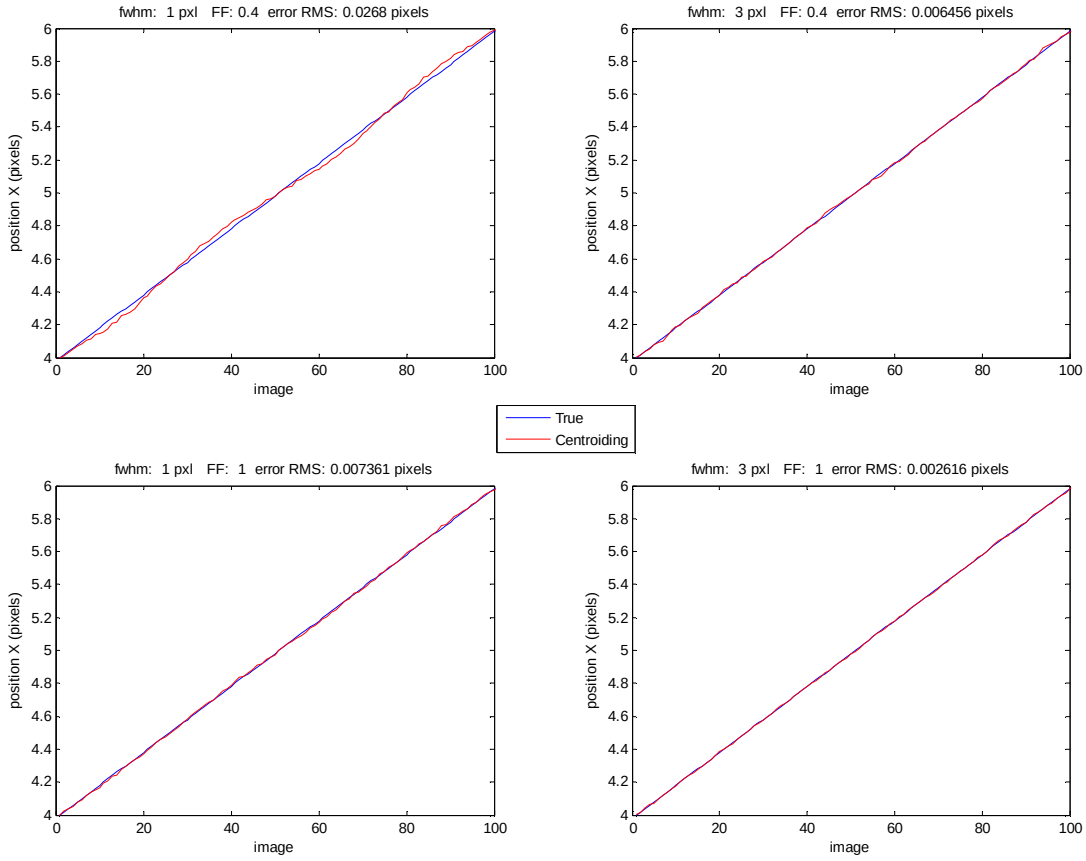


Fig. 4.18: The four limit conditions for the optical parameters FWHM and FF. RMS of the residual error is presented on the title of each plot.

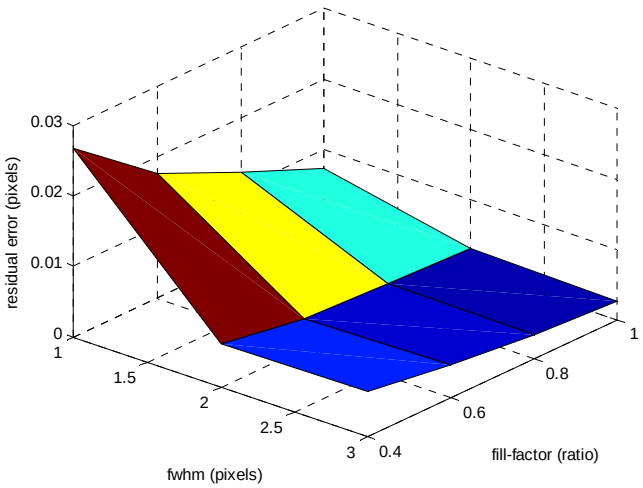


Figure 4.19: Residual error presented as a surface, when varying FWHM and FF simultaneously.

4.7.4.b Electronic Parameters

We also varied the two electronic parameters, ADC and RON, leaving FF at 100 % and FWHM at 2 pixels. The results are presented in figures 4.20 through 4.22. The residual error appears on the title of each plot.

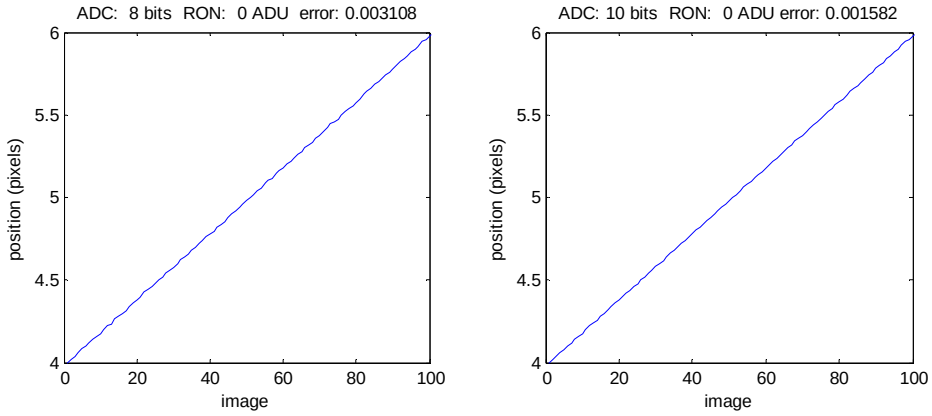


Fig. 4.20: Residual error for 8 bits and 10 bits ADC (RON = 0 ADU).

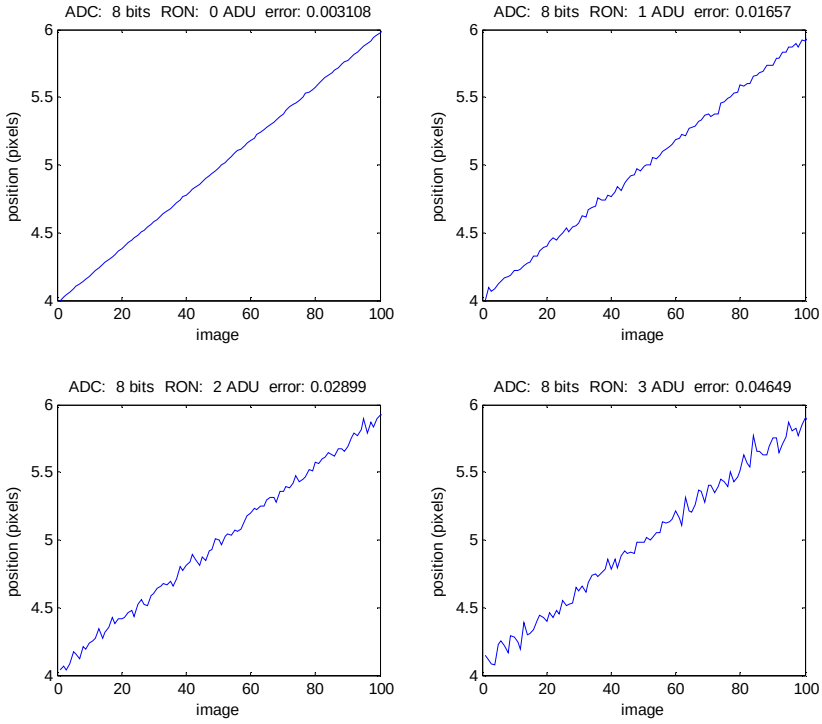


Fig. 4.21: Residual error for 8 bits ADC and RON = 0, 1, 2 & 3 ADU.

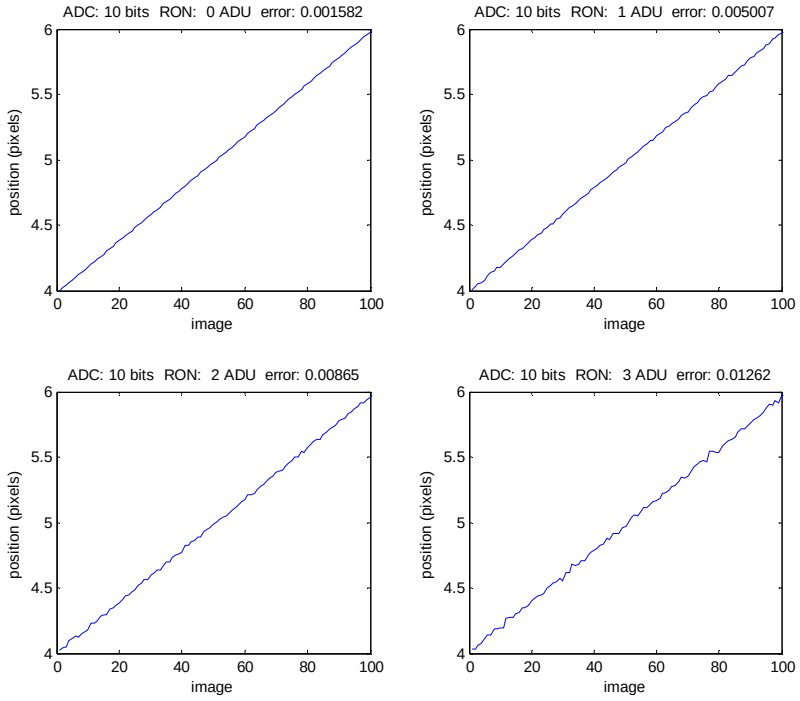


Fig. 4.22: Residual error for 10 bits ADC and RON = 0, 1, 2 & 3 ADU.

In the experimental section, we present the results from a high-speed CCD camera we selected for the figure sensor application. We run the simulation tool with the parameters listed in the next table as being representative of that camera. The residual error is plotted in the Fig. 4.23 for the case of RON = 0 ADU. We also experimented with two *threshold* values[67] (which can be seen on the plot titles), applied before centroiding.

Table 4.4: Parameters for the simulation with realistic values.

Parameter	**Value**
Fill-factor[68]	80 %
FWHM[69]	2 pixels
ADC	8 & 10 bits
RON	0, 1, 2 & 3 ADU

[67] Centre of Gravity (CoG) becomes Thresholding Centre of Gravity (TCoG), as mentioned in section 4.5.2.
[68] This is the most likely value for a front-illuminated CCD with microlens technology, as the one used in the experimental section
[69] We increased from the theoretical value of section 4.5.4, to be in agreement with the experimental results of this chapter

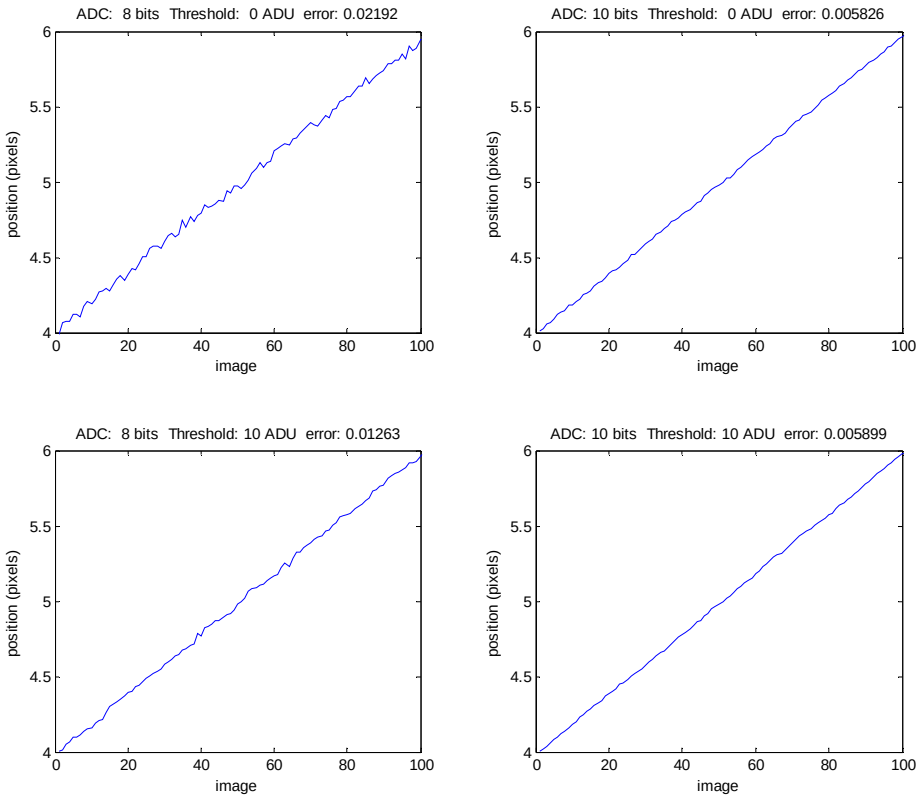


Fig. 4.23: Realistic parameters of the high-speed CCD camera used in the experimental section of this chapter.

Finally, Fig. 4.24 presents a comparison of the RON effect for different combinations of ADC and threshold, compared with the theoretical error using Eq. 4.6.

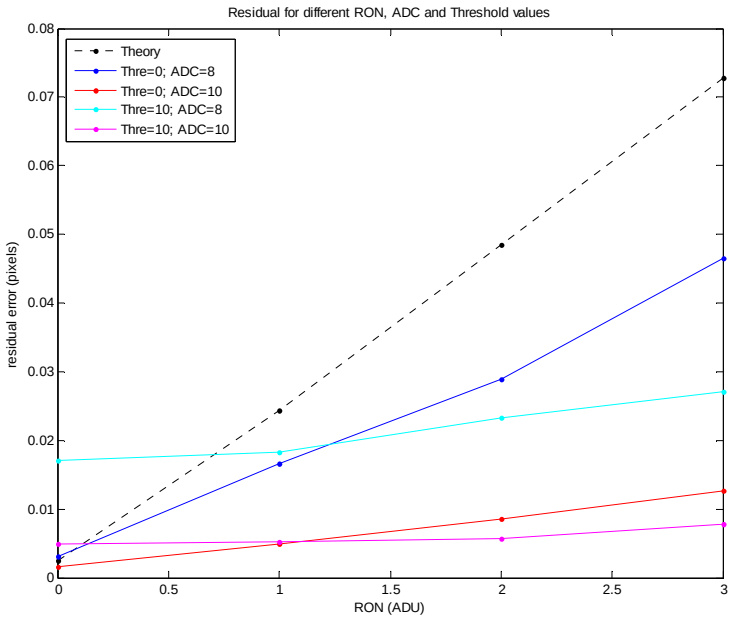


Fig. 4.24: Comparison of theoretical residual error and Monte Carlo simulations for ADC values, using two threshold values for the centroiding (CoG), for different RON values.

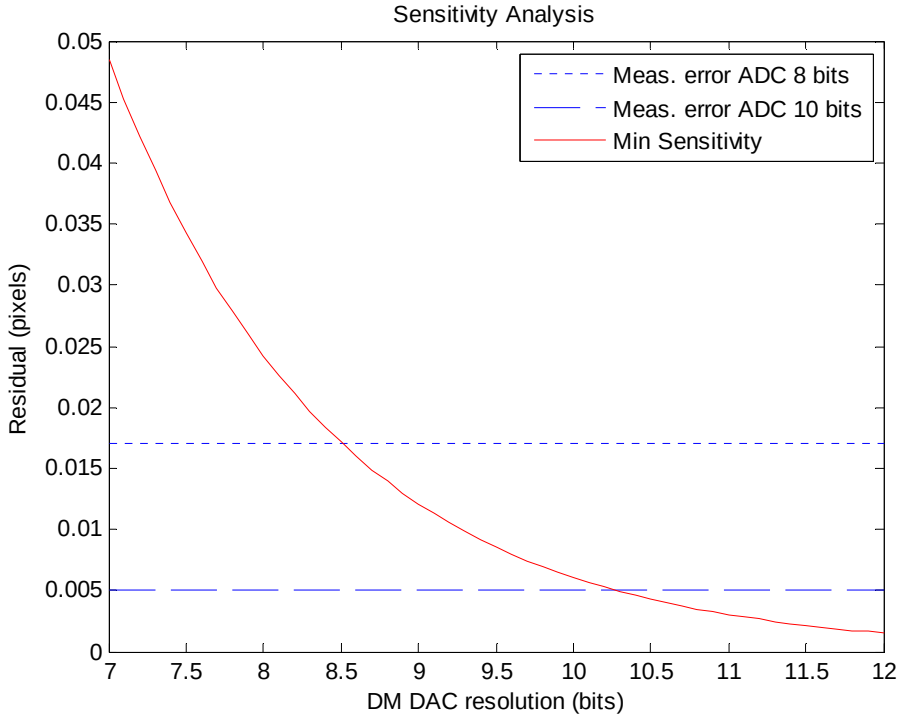


Fig. 4.25: Effective DM DAC resolution sensitivity, for RON = 1 ADU (from previous figure).

We returned to our sensitivity analysis (section 4.5.4) to understand the implications in sensitivity for the residual values of Fig. 4.24. For a given FS sensitivity, we obtained the equivalent

resolution for the DACs driving the DM actuators. This is, because the residual error for RON > 0 is higher than the one presented in Table 4.1, we computed the sensitivity for lower values of DAC resolution and compare it with the sensitivity obtained with the simulation tool. The result is presented in Fig. 4.25 for the case of RON = 1 ADU. The intersection between the sensitivity at each FS ADC resolution and the FS sensitivity represents the effective FS sensitivity required to measure displacements of the DM actuators. We use this analysis to find the final sensitivity of the FS with the experimental results at the end of this chapter.

4.7.5 Analysis of Simulation Results

The results of the simulations are largely as expected. The main conclusions derived from the simulations are the following:

- A poor fill-factor produces the expected 'wobbling' in the centroid position, as seen on Fig. 4.18. This result drives the selection of high-speed CCDs with microlens technology.
- The effect of ADC resolution when digitizing the CCD pixels is somewhat more important than previously expected. Going from 8 bits to 10 bits appears to be an appealing characteristic in defining the CCD camera to select. High-speed operation should not be limited by the resolution at these levels, since there are very fast ADC converters at these resolutions.
- From the results in Fig. 4.22, the most important effect for sensitivity is RON. This is an unavoidable noise source in CCDs, so care must be applied in selecting a device with respect to this parameter.
- RON = 1 ADU or thereabouts seems to be a realistic target, when looking at the results of the experimental section. This produces a sensitivity of around 0.015 pixel, e.g. 1.5 % of a pixel. This is in agreement with commercial Shack-Hartmann wavefront sensors from companies such as Optocraft and Imagine Optics, which claim a "hundredth of a pixel" spatial resolution limit.

4.8 Wavefront Reconstructor

It is necessary to go through a reconstruction step to recover the shape of the measured wavefront. The wavefront reconstructor receives the centroid positions from the SH-WFS and produces a surface output. It is therefore possible to have an estimate of the real mirror shape or wavefront, if the reconstructed wavefront is calibrated with an interferometer or a similar metrology device.

When the number of actuators in the deformable mirror is the same number of SH subapertures, the reconstruction becomes ill-conditioned, a typical situation in AO systems. In the FS case, the number of wavefront measurements (SH subapertures) is different from the number of wavefront estimates (wavefront 'nodes'), so we can produce a reconstruction matrix which is computed as the inverse of the geometric relationship between subapertures and nodes. This matrix is inverted using standard techniques, such as singular value decomposition (SVD).

The reconstruction process is well explained in Southwell (1980). The Eq. 4.12 allows one associating wavefront slopes and wavefront nodes, when using the geometric representation in Fig. 4.26.

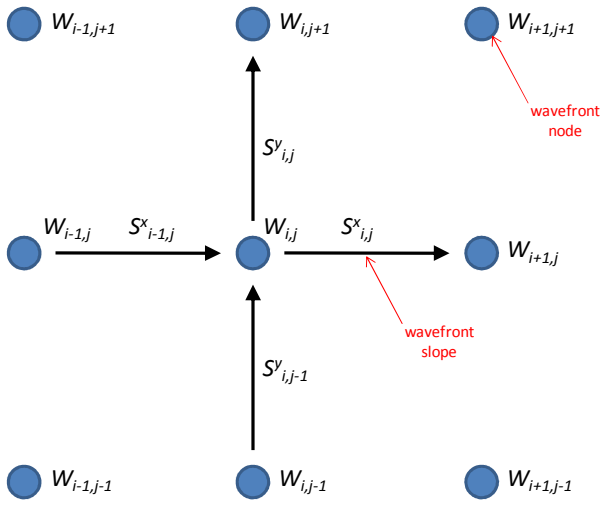


Fig. 4.26: Geometric representation of wavefront slopes (as arrows) and wavefront nodes (as blue dots). The slopes $S_{i,j}^{x,y}$ are the SH centroids. The reconstructor will find the wavefront phases $W_{i,j}$.

$$S_{i,j}^{x} = \frac{W_{i,j}+W_{i+1,j}}{2} \quad (4.12)$$

Mathematically, the reconstruction process can be represented in matrix form

$$s = Aw + n \quad (4.13),$$

where:

A : matrix defined by the geometry of nodes and subapertures

s : vector of slope measurements

w : vector of wavefront estimates

n : measurement noise

Solving for w yields to

$$w = A^{-1}s - A^{-1}n \quad (4.14).$$

As noted above, the applicability of this technique depends on the ability to invert A. The term $A^{-1}n$ represents the noise propagation through the reconstructor and it will ultimately limit the sensitivity of the figure sensor, which is the main aspect under study in this chapter, therefore we have implemented some simulations to quantify this term.

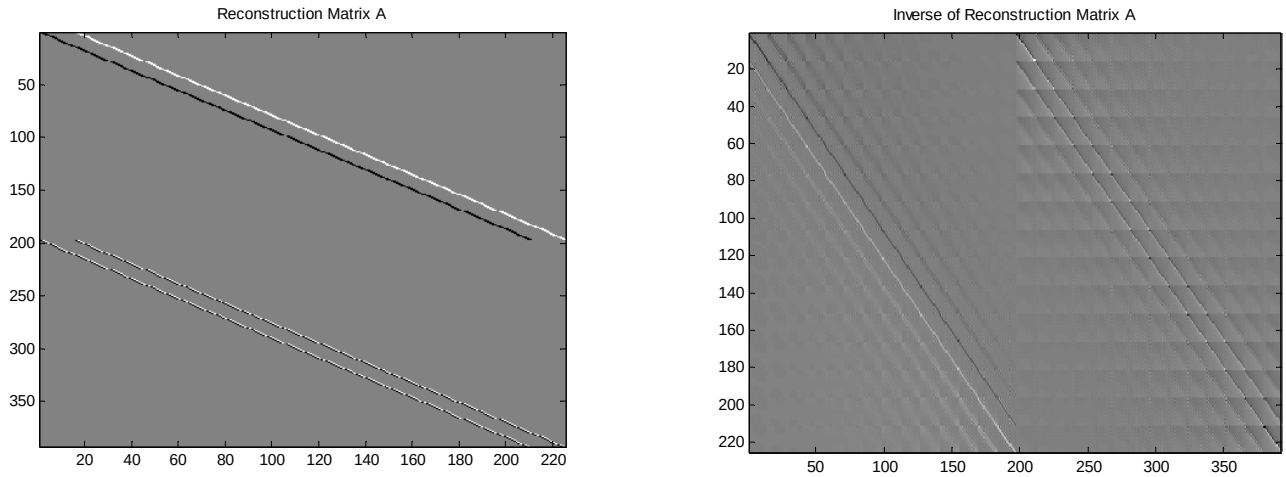


Fig 4.27: Reconstruction matrix A (left panel) and its pseudo-inverse (right panel), computer using SVD, for 14 x 14 subapertures figure sensor.

Our reconstructor accepts 14 x 14 subapertures[70], which should be an adequate number of sampling elements for measuring the CANARY DM, since the latter has 8 x 8 actuators. The dimensions of matrix *A* are:

- Rows: 392, which corresponds to the 14 x 14 subapertures and *X* and *Y* coordinates ($14 \times 14 \times 2 = 392$)
- Columns: 225, which corresponds to the 15 x 15 wavefront nodes[71] to be estimated ($15 \times 15 = 225$)

Matrix *A* is presented in Fig. 4.27 along with its inverse. The diagonal features in *A* come from the geometric factors (*+0.5* and *-0.5*) which relate the nodes and spots from Fig. 4.26.

To measure the noise propagation, we exercised the reconstructor with vectors of zero mean and Gaussian noise of increasing amplitude, measuring the RMS value of the output vector. The plot in Fig. 4.28 shows the result of this simulation. A linear regression in the plot (red line) produced a slope of *1.24*, which relates input and output noise. This is the net amplification effect of the reconstructor, usually called *noise propagation*.

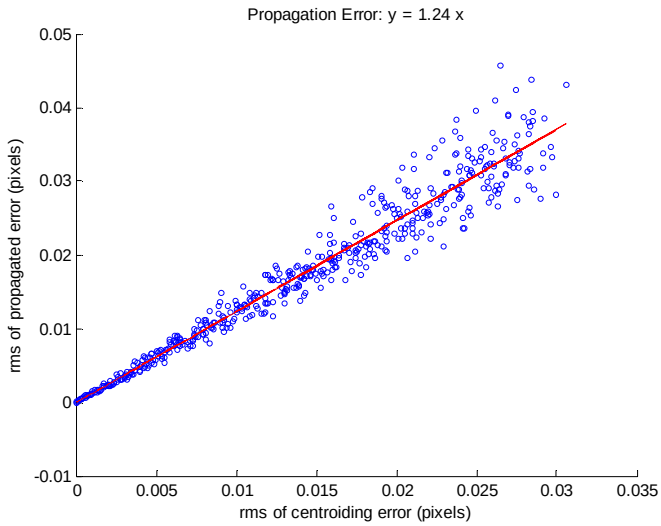


Fig. 4.28: Simulation of the propagation error for the standard reconstructor of Eq. 4.12 (blue dots) and linear regression of the data (red solid line). The increase in error due to the reconstructor is 24 %.

[70] The number of subapertures (14 x 14) was defined to be above the number of actuators in the DM but not excessive, so the total size of the image in the high-speed CCD camera chosen (see section 4.9) is adequate for high-speed imaging.
[71] The number of nodes is one more than the number of subapertures in the WFS, from Southwell's reconstructor (see Fig. 4.26).

4.9 Experiments

4.9.1 Description of the experiment

We undertook several experiments in the laboratory to understand the limitations a high-speed camera would impose on the FS. We needed to illuminate the CCD camera with SH spots, which we could move with respect to the CCD by known displacements. We tried several methods, and settled on using a piezo actuator to displace the CCD camera a small amount. Our piezo actuator was not calibrated so we devised an extension to the experiment to overcome this limitation. The optical experiment is shown in Fig. 4.29.

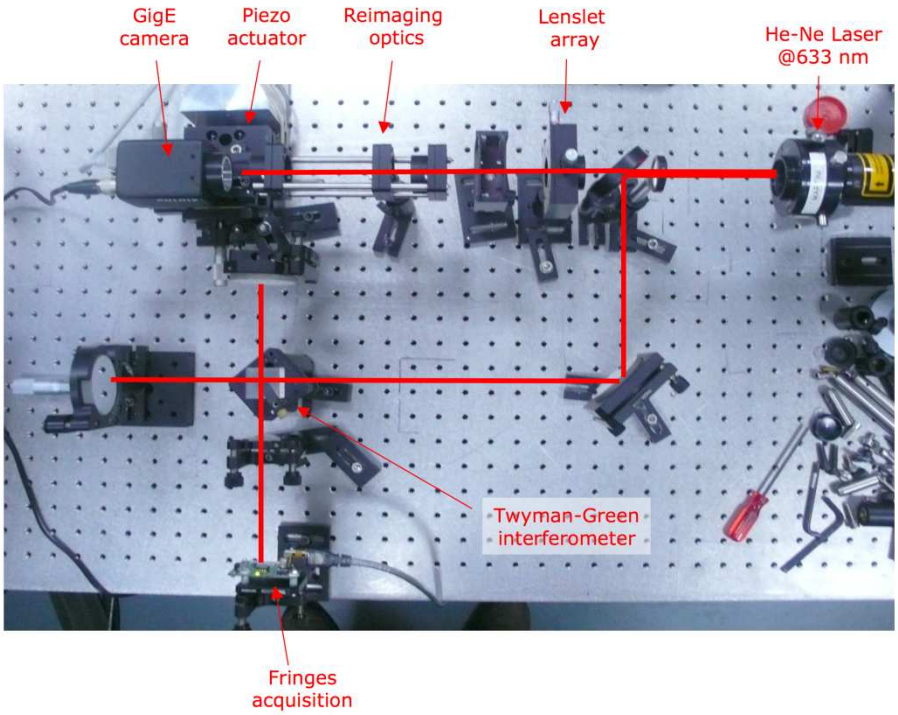


Fig. 4.29: Experiment to measure the sensitivity of the CCD camera when used with the SH-WFS.

The experiment consists of two branches: the laser beam is split into a straight through beam for the SH-WFS (going through the beamsplitter in the figure) and a diverting branch is used to implement a Twymann-Green interferometer with two flat mirrors and a second beam splitter. This interferometer allows us to measure the displacement of the FS camera (*GigE* camera in the figure) in 'fringe' units. A second camera in charge of '*Fringes acquisition*' (bottom of the figure) is installed

on the secondary branch; one of the flat mirrors is tilted in order to produce fringes on this camera. The other mirror is mounted on the linear stage where the FS camera is; both are displaced at the same time by the piezo actuator. We recorded this displacement with two computers simultaneously, while we applied a slow voltage ramp to the piezo actuator that takes around 15 seconds to complete. The raw data were saved, in order to do off-line processing.

Although we still lack an absolute calibration for the experiment, that is, to know the exact displacement in microns versus applied voltage to the piezo actuator, we can use the fringe positions as a relative calibration and use it to obtain a residual error for the centroiding.

4.9.2 Fringe tracker

To know the position of the fringes within the images, we implemented a simple fringe tracker, using the following procedure:

1) For each fringe image, we collapsed multiple rows into a single vector, by summing all rows column by column. This increases the signal-to-noise of the interferogram, taking advantage of the fact that the fringes are almost vertical in the camera
2) We picked a fringe at the beginning of the run (see Fig. 4.30) and computed the centre of gravity of the fringe.
3) We tracked the same fringe by recalculating its centre of gravity for each new image
4) This simple algorithm gave us satisfactory results, given the high signal-to-noise of the interferograms.

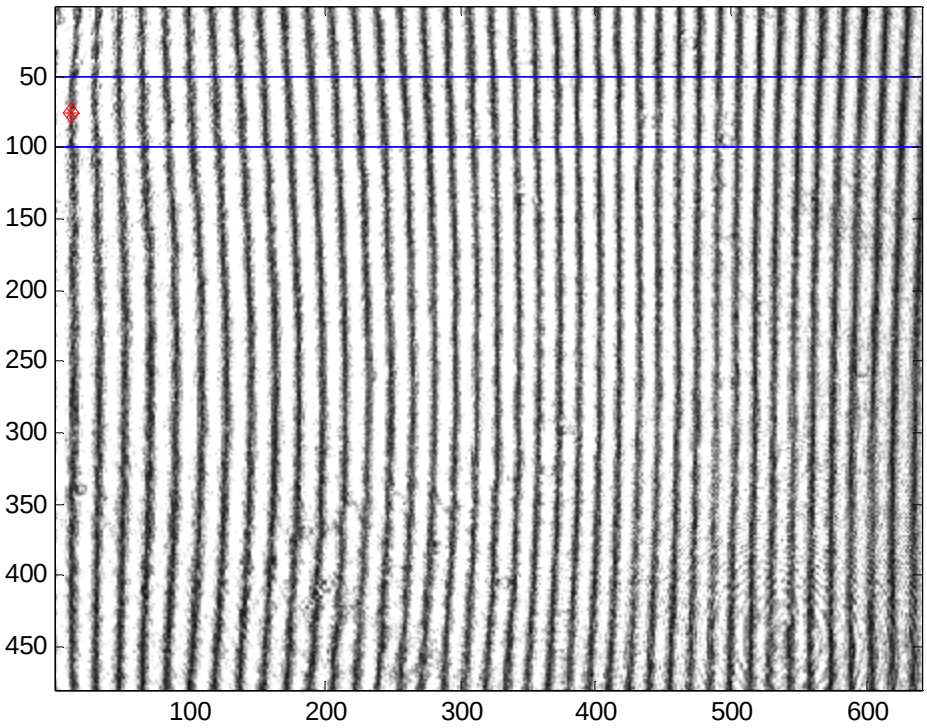


Fig. 4.30: First interferogram of a run, with the range of fringes used for the tracker (blue lines) and the fringe being tracked as a red diamond.

4.9.3 CCD Camera

The high-speed CCD camera selected for the FS is the JAI TM6740-GE, shown in Fig. 4.31:

Fig. 4.31: High-speed CCD camera JAI TM6740-GE.

This camera presents some appealing characteristics, described briefly here:

- Imaging rate of 200 frames-per-second for a full size image (640 x 480 pixels) and up to 1250 frames-per-second for a sub-windowed[72] image (224 x 160 pixels). Given the number of actuators in CANARY DM, it is possible to use the smaller image and hence obtain very high frame rates
- Data transfer through a Gigabit Ethernet[73] port, which is a standard, widely available protocol
- The CCD sensor of the camera is a front-illuminated Kodak CCD with micro-lens technology, i.e. high fill-factor

4.9.4 Centroiding

For this experiment, we are interested in understanding the error in position of the spot we are imaging with the high-speed camera, therefore we only need to perform centroiding in one spot of the SH pattern. For that purpose, we picked a central spot for the first image of the run (see Fig. 4.32) and defined a region of analysis around it. For the rest of the images in the run, we used this region to calculate the centroid position using CoG.

4.9.5 Calibration Methodology

Calibration for this experiment consists of establishing common units for both cameras recording the experiment. Given both cameras recorded relative positions in pixels, the calibration implies converting these units to a common system. For simplicity, we translate the fringe tracker to centroiding pixels. As figures 4.33 and 4.34 show, each camera produced its own measurement of the displacement. The calibration method we implement is a 'linear transformation' that allows converting fringe tracker pixels into centroiding pixels. In order to compute this transformation, we identify common features in the plots and define them as known points for the transformation, similar to passing a straight line through two points in the plane.

[72] This is not a binned image (i.e. where adjacent pixels get joined) but a section or window of the complete image.
[73] Gigabit Ethernet is a standard protocol used by computer and other components requiring high bandwidth.

As Fig. 4.34 shows, we used the beginning and the end of the plots as references for calculating these parameters. The piezo actuator was not displacing at these two moments during the experiment, therefore we can average the noisier centroid position during that time to improve the accuracy of the calibration. The two plots once calibrated are presented in Fig. 4.35. From this plot, it is now possible to obtain residual error figures (as the difference between them) in FS camera pixel units. See figures 4.36 and 4.38.

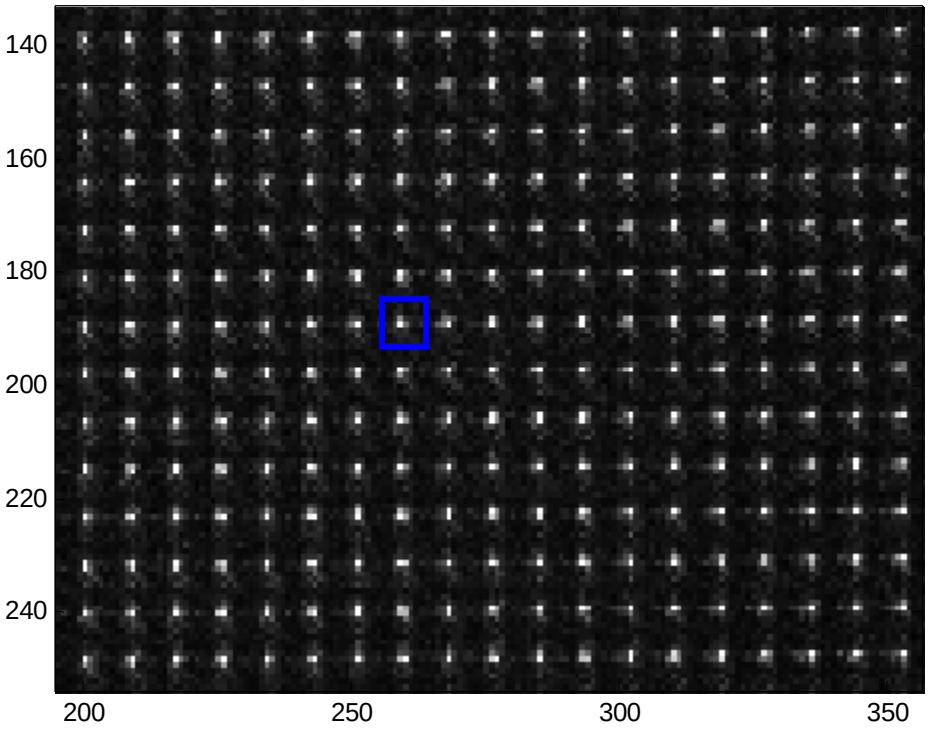


Fig. 4.32: A zoomed region of the SH image from the high-speed camera, showing the spot selected for centroiding and the analysis area (square in blue).

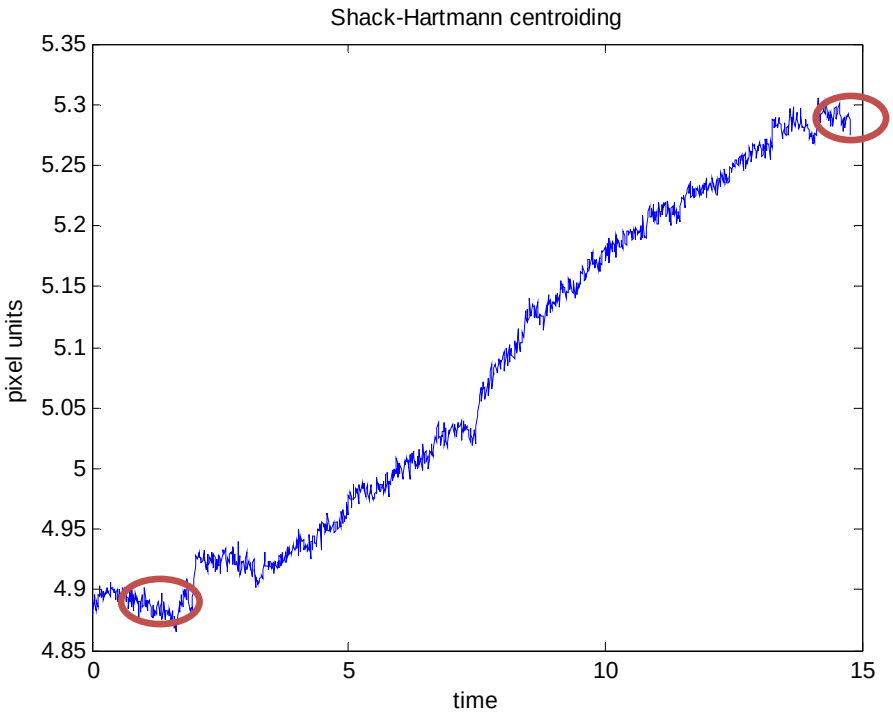


Fig. 4.33: Centroiding versus time for the FS camera (for ADC = 8 bits). The two red circles show the zones used for calibration purposes.

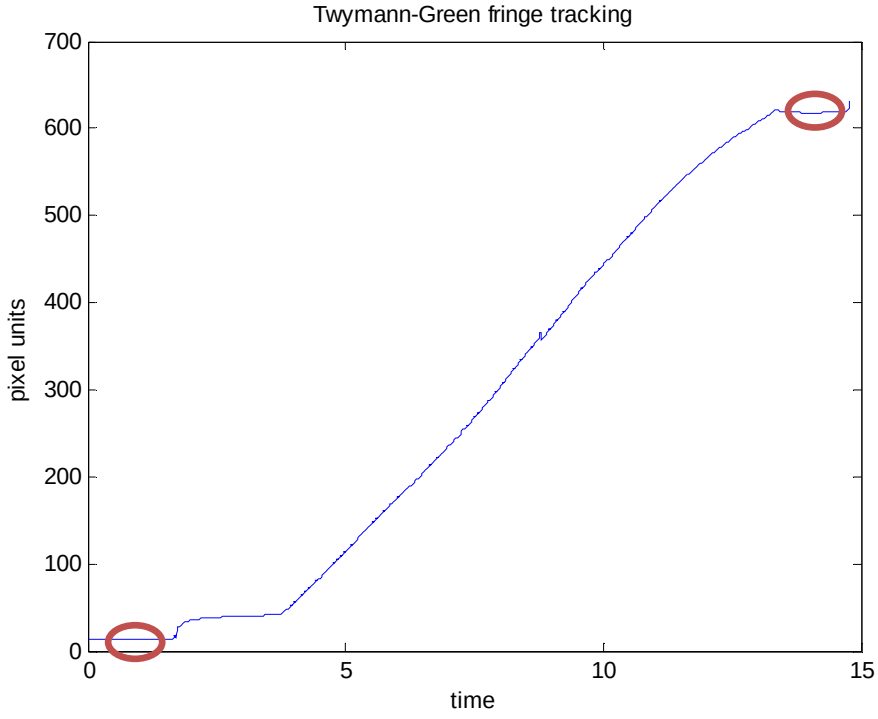


Fig. 4.34: Fringe tracking versus time for the fringe tracker camera. The two red circles show the zones used for calibration purposes.

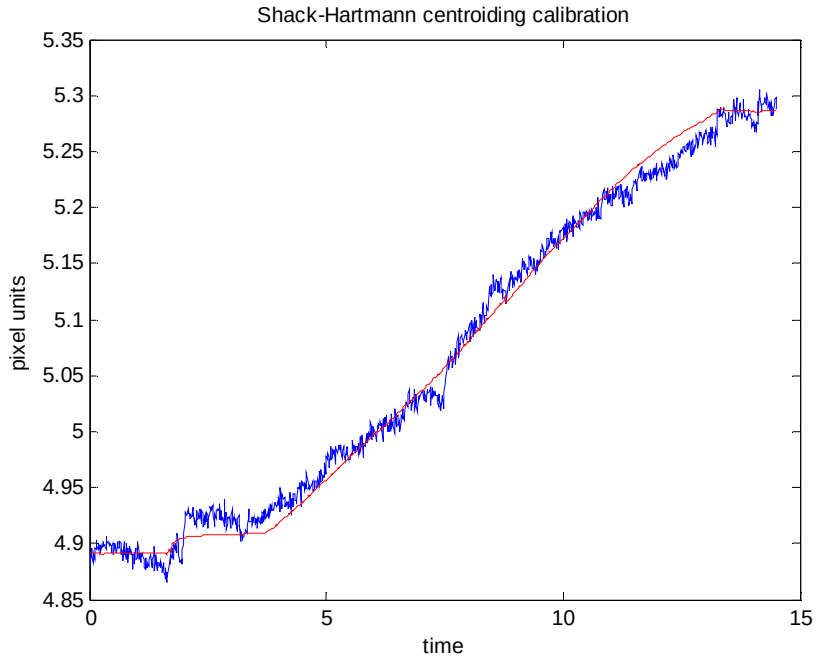


Fig. 4.35: Centroiding calibrated with respect to fringe tracking (for ADC = 8 bits), after applying the linear transformation.

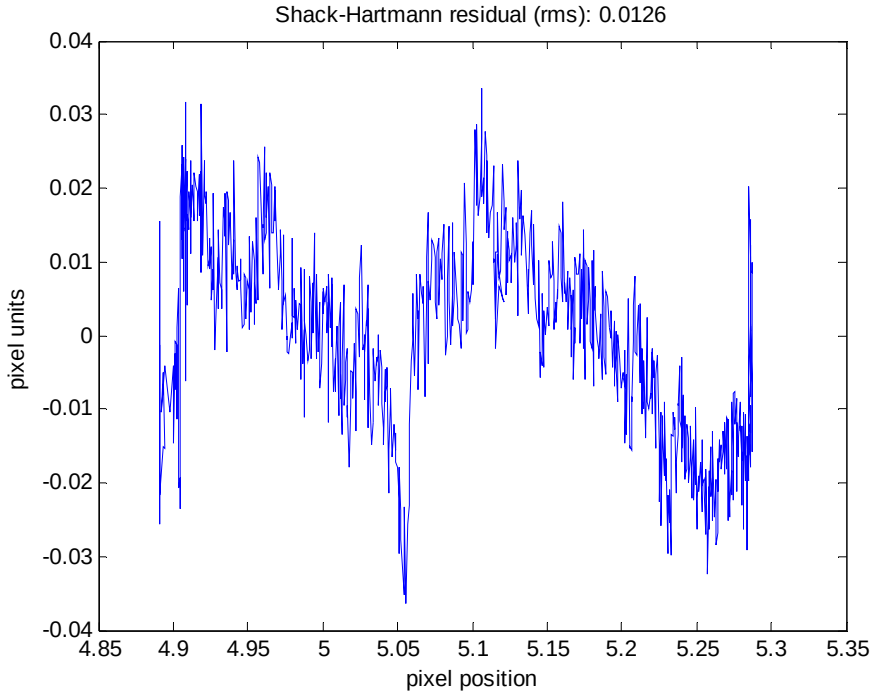


Fig. 4.36: SH residual error (value in the title of the plot) after calibration for ADC = 8 bits.

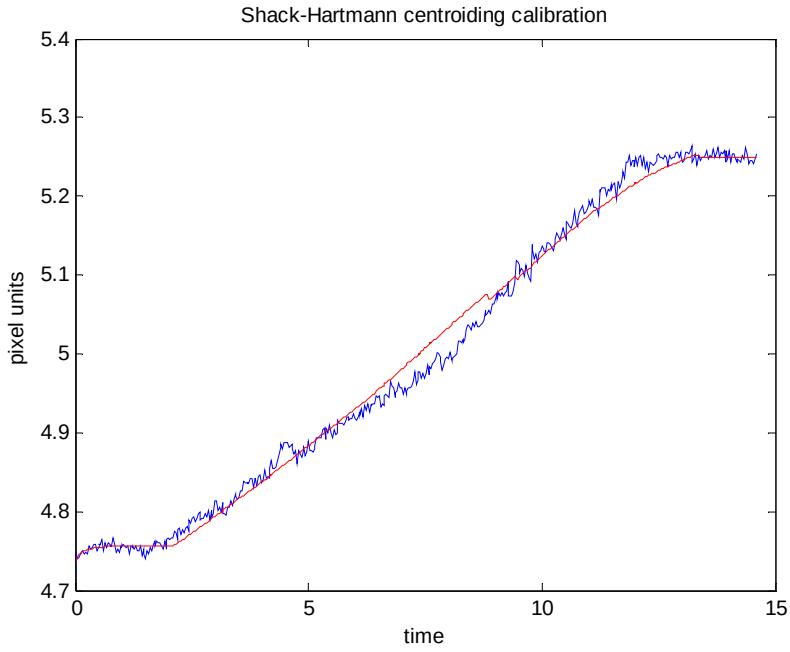


Fig. 4.37: Centroiding calibration for ADC = 10 bits.

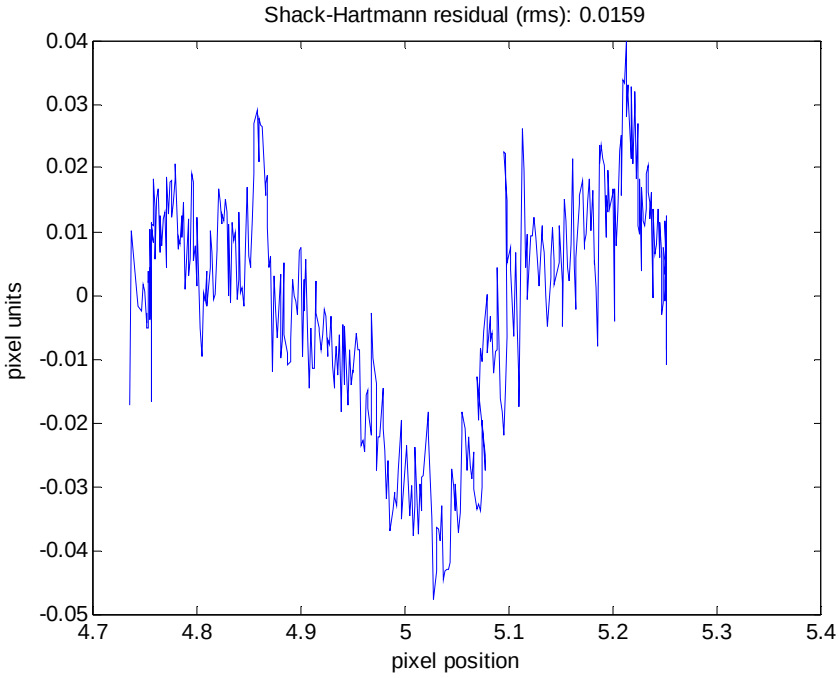


Fig. 4.38: SH residual error (value in the title of the plot) after calibration for ADC = 10 bits.

Although we could have expected to obtain a lower residual for the 10 bits, the experiment proves that the spot-to-spot uncertainty within one run dominates the residual over the difference

between 10 bits or 8 bits ADC. In fact, it is necessary to calibrate the systematic uncertainty in the rms values (i.e. by repeating the measurement for a different spot) to account for variations from spot to spot, due to different S-H lenslet aberrations, different CCD intra-pixel[74] responses, etc.

4.10 Final Sensitivity Estimate for the Figure Sensor

From our Monte Carlo simulations as well as the experiments presented in the previous section, we measure a residual error of about 0.015 pixel RMS for the centroiding process, in the presence of photon and readout noise.

A standard reconstructor will produce an error propagation of around 24 %. The final residual error of the FS would then be

$$0.015 \times 1.24 = 0.019 \text{ pixels} \quad (4.15).$$

In Fig. 4.39, we re-run the simulation in Fig. 4.25 to understand the equivalent number of bits in the DACs driving the DM actuators that the sensitivity found in Eq. 4.15 corresponds to.

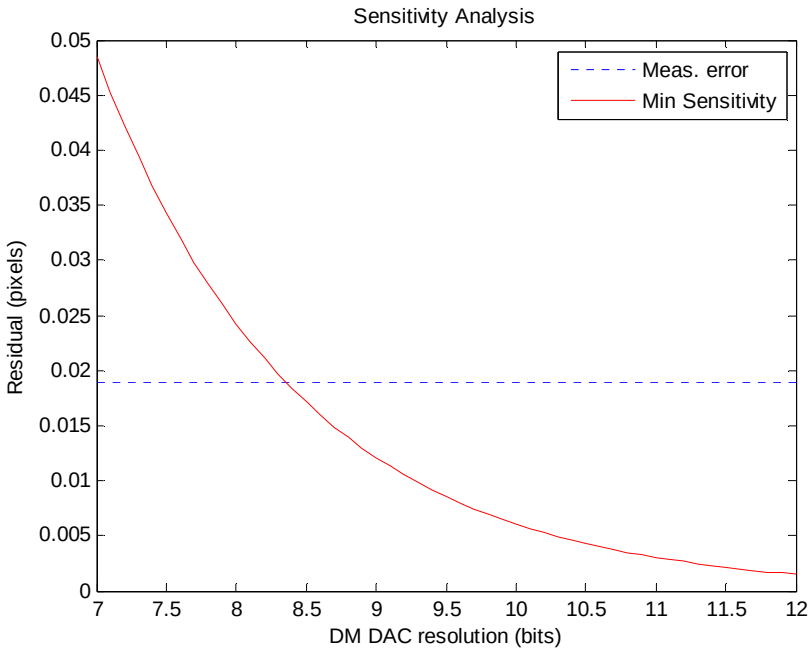


Fig. 4.39: Sensitivity analysis using the experimental residual error.

[74] CCD *intra-pixel response* corresponds to sensitivity differences between pixels in a CCD when collecting photons

We ultimately need to understand the sensitivity of the FS expressed in terms of DM actuator positions (nm). For the case of the Xinetics DM and assuming a linear relationship between DAC value to command the DM and its final position in nanometres, it is possible to find the sensitivity of the FS in terms of the DM actuator positions. This is calculated here:

For the Xinetics actuators the mechanical range p is

$$p = 5700\ nm.$$

And the DAC equivalent from the sensitivity analysis (from Fig. 4.39) is

$$DAC_e = 8.4\ bits.$$

The sensitivity of the figure sensor is then

$$FS_s = \frac{p}{2^{DAC_e}} = \frac{5700}{2^{8.4}} = 16.9\ nm \text{ (4.16).}$$

4.11 Conclusions

This chapter has presented a design for the CANARY figure sensor based on a Shack-Hartmann wavefront sensor. We have evaluated the sensitivity such a sensor can achieve when measuring the shape of the DM facesheet. In order to estimate this sensitivity, we have developed a 'first order' study, based on some theoretical aspects, Monte Carlo simulations and experimental results. We believe the results are accurate enough to be used as a baseline when implementing a figure sensor to measure a deformable mirror.

The final sensitivity number in Eq. 4.16 is comparable to the error the open-loop models implemented in Chapter 3 achieve; therefore the FS can be incorporated into the AO real-time computer for performance evaluation, keeping the level of the DM error term in the error budget of the system. Unlike some types of deformable mirrors such as the Xinetics[75], a Shack-Hartmann wavefront sensor is tolerant to changes in its operating temperature[76]; therefore it can be used to calibrate the DM model in operation.

[75] Xinetics DMs use electrostrictive actuators, which have a hysteresis that is very dependent on operating temperature, changing from < 1 % at 20 C, to 20 % at 0 C.

[76] The CCD's ability to collect photons persists under a very wide temperature range. Lenslets are made out of glass, which has a coefficient of linear expansion < 10 µm / C. For an operating range of 20 C and a lenslet of 10 mm in size, there would be an expansion of only 2 µm.

As briefly mentioned in the introduction of this chapter, one can also envision using the FS to control the DM in a local loop. The '*setpoint*' for this loop would come from the AO computer in charge of implementing MOAO. A controller in charge of this loop would force the DM to follow the setpoint positions from the AO computer, incurring an error that would be ultimately defined by the final sensitivity found in Eq. 4.16. If this closed-loop system is implemented in what can be called the 'Deformable Mirror Controller' (DMC), then the DMC error contribution to the error budget of the system would be *16.9 nm* RMS.

4.12 References

Andersen, D., Fischer, M., Conan, R., Fletcher, M. "VOLT: The Victoria Open Loop Testbed", SPIE Proceedings 7015, 70150H, (2008)

Conan, R., Bradley, C., Lardière, O., "Raven, a harbinger of Multi–Object Adaptive Optics based instruments at the Subaru telescope", SPIE Proceedings 7736, 77360T, (2010)

Driggers, R. "Encyclopedia of Optical Engineering", CRC, 2295-2297, (2003)

Gilles, L. and Ellerbroek, B. "Shack-Hartmann wavefront sensing with elongated sodium laser beacons: centroiding versus matched filtering", Applied Optics, 45, 25, pp.6568-6576, (2006)

Janesick, J. "Scientific Charged-Coupled Devices" SPIE Press, (2001)

Janesick, J. "Photon Transfer" SPIE Press, (2007)

Platt, B., Shack, R., "History and Principles of Shack-Hartmann Wavefront Sensing" Journal of Refractive Surgery 17, September/October 2001

Southwell, W., "Wave-front estimation from wave-front slope measurements", JOSA 70, 8, pp. 998-1006, (1980)

Thomas, S., Fusco, T., Tokovinin, A., Nicolle, M., Michau, V. and Rousset, G. "Comparison of centroid computation algorithms in a Shack-Hartmann sensor", MNRAS 371, pp. 323-336 (2006)

Wilson, R. "Reflecting Telescope Optics II", Springer, (1999)

Chapter 5: Conclusions and Future Work

5.1 MOAO

At present, there are adaptive optics systems running routinely on medium to large size telescopes, producing a steady growth in the number of publications[77] per year. All of these systems can be described as "first generation" systems, since they use one NGS / LGS and one DM for single-conjugated correction over a limited FoV. Some of the first wide-field AO systems have been the ground-layer AO systems at the 4.2 m WHT (Benn 2008) and at the 6.5 m MMT (Baranec 2009), or the Multi-Conjugate Adaptive Demonstrator "MAD" at the VLT (Marchetti 2008). These systems will be followed soon by Gemini's MCAO system (Neichel 2010), the first facility system delivering a wide-field AO correction to other astronomical instruments[78]. In fact in a few years, at least 5 large telescopes will have ground-layer AO systems built-in. These are: the AOF at the VLT, the twin LBT telescopes, the MMT and one of the Magellan telescopes.

Furthermore, there is agreement within the ELT projects[79] that, given the theoretical resolution a telescope of that diameter can achieve[80], an ELT should consider AO as an integral part of the telescope design. Indeed, the 25 m Great Magellan Telescope (GMT) and the 42 m European ELT have incorporated ground-layer AO as a central element of the telescope optics.

MOAO was first proposed by F. Hammer *et al* (2001), motivated by the observational objective of having multi-object spectroscopy with high spatial resolution. The concept was presented to the *Call for Proposals* for second generation instruments for the VLT as FALCON, but it was not considered for construction. Nevertheless, the FALCON concept was further developed and published by F. Assemat *et al* (2007). From then on, more interest in MOAO has flourished in the astronomical

[77] Private communication with Dr. Claire Max, Director of the *Center for Adaptive Optics*, University of California
[78] Gemini's MCAO system will produce a 70" x 70" AO-corrected FoV that will feed two instruments: a near-infrared imager, called GSAOI and a multi-slit spectrograph, called FLAMINGOS-II
[79] The three ELT projects under development now are: the E-ELT (European ELT, www.eso.org/public/teles-instr/e-elt.html); the GMT (Great Magellan Telescope, www.gmto.org) and the TMT (Thirty Meter Telescope, www.tmt.org)
[80] The Rayleigh limit for light at 500 nm is 16 milliarcsec for an 8 m telescope, while it is 3 milliarcsec for a 42 m telescope

and AO communities, in particular driven by the goal of achieving the full potential that an ELT can have.

Early concepts for the IRMOS instrument for the TMT[81] considered MOAO (Andersen 2006). This proposal considered a *woofer-tweeter*[82] configuration for the DMs. The woofer works in closed-loop and removes low-order aberrations, leaving to a tweeter the high-order correction afterwards. If there is more than one tweeter and they work in open-loop, the system becomes MOAO. One can consider the DM in ground-layer systems proposed or under construction as the woofer of a MOAO system. The EAGLE instrument[83] for the E-ELT is one such example. In this context, the CANARY demonstrator will be a stepping stone when it confirms that MOAO works on-sky[84].

It appears clear to us that the ground-layer AO systems that will come online in existing telescopes, and before any of the ELTs finishes construction, should consider a MOAO instrument in its instrumentation suite. Such a capability would allow high resolution multi-object spectroscopy in the near-infrared to become a reality for interested observers sooner than later. This means several years of observations before having access to a new and more powerful instrument such as EAGLE[85]. For example, if a MOAO instrument is built for the AOF at the VLT and can be ready by ~2014, such instrument would have at least 5 years of science (2015-2019) before it is outclassed by a better instrument on an ELT. This window of opportunity should definitely be explored.

5.2 Experimental Work

The experimental work presented in this thesis addresses the issue of deformable mirror control for Multi-Object Adaptive Optics using novel methods. MOAO represents a departure from all previous AO systems, since it works in an optical open-loop, as described in detail in Chapter 2.

[81] This concept was so-called "TiPi" (private communication with Dr. Keith Taylor)
[82] The "woofer-tweeter" concepts refers to splitting the AO system in low-order and high-order, using two different DMs, in analogy to an audio speaker
[83] http://eagle.oamp.fr/spip/
[84] Recall from section 2.8 that CANARY *Phase C* has a closed-loop woofer / open-loop tweeter configuration equivalent to EAGLE
[85] At the time this work was completed, there is no news on whether or not EAGLE will be built

Throughout this work, we have tackled the open-loop operation of the DM by proposing two different solutions, developed in Chapter 3 and 4.

In Chapter 3, we implemented purely open-loop models for DM control using so-called "non-parametric estimation techniques". The main motivations behind this implementation were of three kinds:

- To avoid a model that would depend on the physics of the DM components and materials (actuators, facesheet)
- To incorporate mathematical models that can handle the non-linear behaviour of the DM realistically
- To be able to apply the same techniques to any kind of DM

The results in Chapter 3 are quite satisfactory in all these regards. The following areas should be studied in order to improve and extend this work:

1) To use a faster and more accurate interferometer: we have gathered all of our experimental data using a "Fisba" Twymann-Green interferometer, which is a very good tool for optical metrology, but data acquisition is a very slow process, spanning days for the data used in this work. There are other interferometers, such as the ones from 4D Technologies[86], which can take interferograms at tens of Hertz rate. Incorporating an instrument like this would allow acquiring data for training an open-loop model in a matter of minutes.
2) The main "show stopper" for implementing models for DMs with a large number of actuators comes from the computational power required to train the model. For the case of the Boston MEMS DM, having a model with 126 inputs and 126 outputs was quite challenging for the computer resources available to us. The alleviation scheme implemented for the Artificial Neural Network, subdividing the area in smaller areas, worked rather well, although it did impact the accuracy of the model, when one

[86] http://www.4dtechnology.com

compares the two networks implemented. Most probably, a full-size network will obtain better results than we obtained with the MARS model. Nevertheless, there must be some type of subdivision scheme if we want to model big mirrors such as a full MEMS. It shall be confirmed whether these techniques are feasible to implement for a full DM.

In Chapter 4, we are proposing an alternative to a purely open-loop model for DM control in MOAO. This is a dedicated sensor to measure the shape of the DM and therefore being able to confirm that it achieved the correct figure. This is 'like' an interferometer (familiar from Chapter 3), but it must work at the speed at which the AO computer runs, usually not less than 500 Hz. Running at this speed rules out standard interferometers, even the high-speed ones from 4D Technologies. It also rules out typical detectors used in astronomy, say scientific CCDs, which are not sufficiently fast (tens of Hz at the most). The only remaining imaging technology is high-speed, interline CCDs. They are introduced in Chapter 4, with Monte Carlo simulations to understand the limitations in using this technology. We believe we have proposed a feasible implementation for tackling the task of measuring the DM shape, using a Shack-Hartmann Wavefront Sensor and the high-speed CCD camera. The main areas in which we see more work are listed here:

1) Complete SH-WFS simulations: we only implemented Monte Carlo simulations for a single sub-apertures of the Shack-Hartmann, but the final measurements depends on accurate centroiding in all subapertures, so this basic simulation tool should be extended to cover a realistic number of subapertures, say, 12 x 12 for a 8 x 8 actuators DM.
2) An absolute calibration procedure for centroiding: The experiments presented in Chapter 4 were adequate for the level of the study presented, but they lack absolute calibration. We alleviated this problem by implementing the Twymann-Green interferometer described in the Chapter; however, a final experiment should have a means of knowing the absolute displacement of the spots in microns. We experimented with a digital micrometer, but its hysteresis prevented us from gathering useful data. A laser-based micrometer is the appropriate tool for this job.

3) More experimental work measuring a real DM whole pupil: The next step in experiments with the Figure Sensor corresponds to obtaining experimental phase maps of a real DM and comparing them with data from a standard interferometer. Depending on the results and indeed the differences between the experimental results, the calibration procedure might be refined, in order to obtain “interferometer-like” outputs from the Figure Sensor.

5.3 Future Research

There is an interesting area which we plan to investigate as a follow-up to this work. It consists of applying the techniques used in Chapter 3 to the problem of open-loop tomography, necessary for MOAO and presented in the section 2.6 of Chapter 2.

The standard treatment of tomography is a linear system with variables and unknowns, the former being the guide stars WFS data and the latter being the turbulent layers. Nevertheless, although the turbulence presents mostly in layers, it is also continuously distributed in altitude, and this can constitute a third of the total turbulence (see Fig. 1.3). Therefore, it is a simplification to place layers at fixed altitudes and project the whole turbulence onto them. Instead, why not “learn” the atmosphere each night and apply a correction from that learning? This is in line with our experience from Chapter 3, where we “learnt” about the DM characteristics by exposing a statistical sample of its behaviour. This is similar, but more complex, because of the random character of the turbulence, and the nature of on-sky operation rather that the controlled environment of a laboratory. Nevertheless, we think that the idea is feasible and it should be explored around the following recipe:

1) Each turbulent measurement from a number of guide stars will become an element of the training data set for the model
2) For training, point the telescope to a crowded field, where a number of NGS can be chosen (assuming one uses one or more NGS WFS pick-off mechanisms). Use the constellation of LGS WFS around the field and acquire data from the NGS and LGS WFS simultaneously.

3) Train a neural network-like model with the following inputs and outputs:
    a. Inputs: LGS WFS data and NGS positions in the field.
    b. Outputs: NGS WFS data.
4) Exposed to WFS data from the Laser Guide Stars as inputs and the WFS data from NGS as outputs, the model will learn about the turbulence for the NGS positions in the field – this is the reason to incorporate the NGS positions in the field as inputs for training. The model will learn how to represent the turbulence seen by the NGS from what can be seen from the off-axis LGS WFS data, effectively computing a tomographic reconstructor from the real atmospheric data.
5) When observing, the model will be in 'Estimation' mode (recall Fig. 3.17), so when receiving new LGS WFS data and the position in the field where the correction needs to be applied, it will compute the best estimate for that particular position, effectively applying the tomographic reconstructor.

Vidal's "Learn and Apply" method is similar in concept, but we devised this method independently of theirs. We believe this approach would be able to cope well with the non-linear behaviour of turbulence, since it uses tools developed for such problems. A dedicated hardware platform, such as one based on programmable logic may be appropriate to train non-parametric models in time for nighttime operations. Another area of research will constitute the application of this type of computer systems to training requirements such as the ones described here.

5.3 References

Andersen, D., Eikenberry., S., Fletcher, M., *et al* "The MOAO system of the IRMOS near-infrared multi-object spectrograph for TMT", SPIE Proceedings, 6269, pp. 62694K1-62694K12, (2006)

Assémat, F., Gendron, E., Hammer, F., "The FALCON concept: multi-object adaptive optics and atmospheric tomography for integral field spectroscopy - principles and performance on an 8-m telescope", MNRAS, 376, 1, pp. 287-312, (2007)

Baranec, C., Hart, M., Milton, M., *et al* "On-Sky Wide-Field Adaptive Optics Correction Using Multiple Laser Guide Stars at the MMT", ApJ, 693, 2, pp. 1814-1820, (2009)

Benn, C., Don, A., Agocs, T., *et al* "GLAS/NAOMI: ground-layer AO at the William Herschel Telescope", SPIE Proceedings, 7015, pp. 701523-701523-10, (2008)

Hammer, F., Sayède, F., Gendron, E., *et al*, "The FALCON concept: multi-object spectroscopy combined with MCAO in near-IR", ESO Workshop Proceedings: Scientific Drivers for ESO Future VLT/VLTI Instrumentation, arXiv: astro-ph/0109289v1, (2001)

Marchetti, E., Brast, R., Delabre, R., "MAD on sky results in star oriented mode", SPIE Proceedings, 7015, pp. 70150F-70150F-12, (2008)

Neichel, B., Rigaut, F., Bec. M., *et al* "The Gemini MCAO System GeMS: nearing the end of a lab-story", SPIE Proceedings, 7736, pp. 773606-773606-14, (2010)

Appendix: Peer-reviewed Papers Published