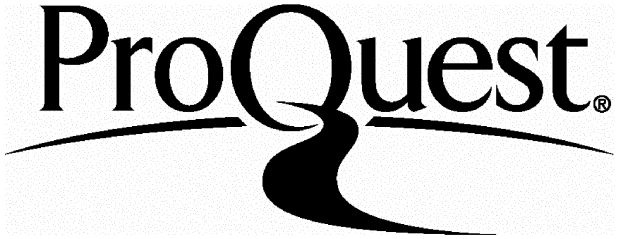
Control of Flexure in Large Astronomical Spectrographs

A Thesis Submitted for the Degree
of
Doctor of Philosophy of the University of London
by
Paolo D'Arrigo

Optical Science Laboratory
Department of Physics and Astronomy
University College
University of London
1996



Abstract

This thesis describes the design, construction and testing of an experimental system for improving the imaging stability on the detectors of the Intermediate-dispersion Spectroscopic and Imaging System (ISIS), a Cassegrain spectrograph at the 4.2 metre William Hershel Telescope. This system, called ISAAC (ISIS Spectrograph Automatic Active Collimator) is based on the new concept of active compensation, where spectrum shifts, due to the spectrograph flexing under the effect of gravity, are compensated by the movement of an active optical element. ISAAC is a fine steering tip-tilt collimator mirror.

The thesis provides an extensive introduction on astronomical spectrographs, active optics and actuator systems. The new concept of active compensation of flexure is also described. The problem of spectrograph flexure is analyzed, focusing in particular on the case of ISIS and on how an active compensation system can help to solve it. The development of ISAAC is explained, from the component specification and design, to the construction and laboratory testing. The performance and successful testing of the instrument at the William Herschel Telescope is then described in detail. The implications for the future of ISIS and of new spectrograph designs are then discussed, with particular stress on the new High Resolution Optical Spectrograph (HROS) for the 8-metre Gemini telescopes.

Contents

List of Tables

List of Figures

Glossary

ADU Analogue Digital Unit.

CCD Charge-Coupled Device (photosensitive array).

CFHT Canada France Hawaii Telescope.

DAC Digital to Analog Converter.

DEIMOS Dual-beam Echelle Imager and Multi Object Spectrograph.

ES Electro-Strictive (actuator type).

ESI Echellette Spectrograph and Imager.

ESO European Southern Observatory.

FSR Free Spectral Range.

FWHM Full Width at Half Maximum.

GMOS Gemini Multi-Object Spectrograph.

HIRES HIgh Resolution Echelle Spectrograph.

HROS High Resolution Optical Spectrograph.

HV High Voltage.

IRAF Image Reduction Analysis Facility.

ISAAC Isis Spectrograph Automatic Active Collimator.

ISIS Intermediate-dispersion Spectroscopic and Imaging System.

LDSS-2 Low Dispersion Survey Spectrograph.

LEMO Coaxial cable connector type and manufacturer.

LPO La Palma Observatory.

LRIS Low Resolution Imaging Spectrograph.

LUT Look-Up Table.

LVDT Linear Variable Differential Transformer.

MMT Multi Mirror Telescope.

MOS/SIS Multi-Object Spectrograph/Subarcsecond Imaging Spectrograph.

MS Magneto-Strictive (actuator type).

NTT New Technology Telescope.

OSL Optical Science Laboratory.

PA Position Angle (on Cassegrain turntable).

PC Personal Computer.

PCB Printed Circuit Board.

P.I. Physik Instrumente GmbH.

PZT Piezoelectric (from lead zirconate-titanate).

RGO Royal Greenwich Observatory.

S/N Signal to Noise.

Terfenol Terbium, Dysprosium, Iron alloy.

UCL University College London.

VLT Very Large Telescope.

WFS Wave-Front Sensor.

WHT William Herschel Telescope.

Zerodur Near-zero thermal expansion glass ceramic.

Chapter 1

Introduction: Telescopes and Spectrographs

1.1 Preface

When I started this thesis, the study of astronomy and astrophysics was undergoing major changes (and still is), mainly thanks to revolutionary technological developments in astronomical instrumentation. Of these technological advances, I want to mention just two, which I think are among the most important: one is the development of space telescopes, the other is the new generation of ground-based telescopes. In April 1990 the Hubble Space Telescope was launched, the first optical telescope to operate from space. After the spherical aberration of its primary mirror was successfully corrected during a Space Shuttle mission in December 1993 [1], the space telescope has delivered superb images of the universe. Taking full advantage of the absence of atmospheric turbulence, its imaging capability has been far superior to any ground based telescope. Since then space astronomy has expanded very quickly, and other telescopes covering the ultraviolet and infrared regions of the spectrum have been built. This has greatly extended the wavelength range available for astronomical observations, previously limited to smaller windows in which the Earth's atmosphere is transparent.

The development of space astronomy has not been in competition with ground-based observation for several reasons. Space telescopes are still limited in size: even the

2.4 metre Hubble Space Telescope is quite small when compared with the largest telescopes on the ground. Their lifetime is also limited and the cost is far higher than ground based systems. Therefore, at the same time space astronomy was rapidly developing, ground-based astronomy has also seen dramatic changes. Since January 1994, the 10-metre Keck telescope has been operational on Mt. Mauna Kea in the Hawaii island. This is the largest telescope in the world and the first fully operational telescope to surpass in diameter the 5-metre Hale telescope on mount Palomar, which was completed almost 50 years ago. The russian 6m telescope and the Multi-Mirror Telescope, earlier attempts to surpass the Hale telescope, ran into trouble in trying to produce good imaging quality, so until the late 1980s 3 to 4 metres was considered to be a safe technical limit for the size of a ground based telescope. Even more important is the fact that the Keck telescope is not one off, but just the first of a new generation of telescopes, in the 8 to 10 metre range, which are currently planned or being built. The reason for this giant leap in telescope size has been the consequence of a series of technological advances, of which the development of computers is the most important.

These fundamental changes in the construction of telescopes have been accompanied by new challenges in the relative instrumentation. Most of modern astronomy and astrophysics originated from spectroscopic studies of the objects in the sky. At present time, spectroscopy is still the major tool for analysing the light coming from these objects. It is therefore safe to say that the astronomical spectrograph is still the most important instrument for our understanding of the universe. The new telescopes, both ground-based and in space, demand innovative design and better performance from spectrographs. The work in this thesis was aimed at solving an important problem affecting ground-based instruments: the stability of the spectrum on the detectors of the spectrograph. This problem is particularly serious for instruments at the Cassegrain focus of the telescope, because the spectrograph follows the motion of the telescope when tracking a star, and bends under its own weight. This effect is commonly called flexure and, while always detrimental, its relevance has become more critical with the new generation of large telescopes.

My work was focussed on reducing flexure in the ISIS (Intermediate-dispersion Spectroscopic and Imaging System) spectrograph at the 4.2 metre William Herschel

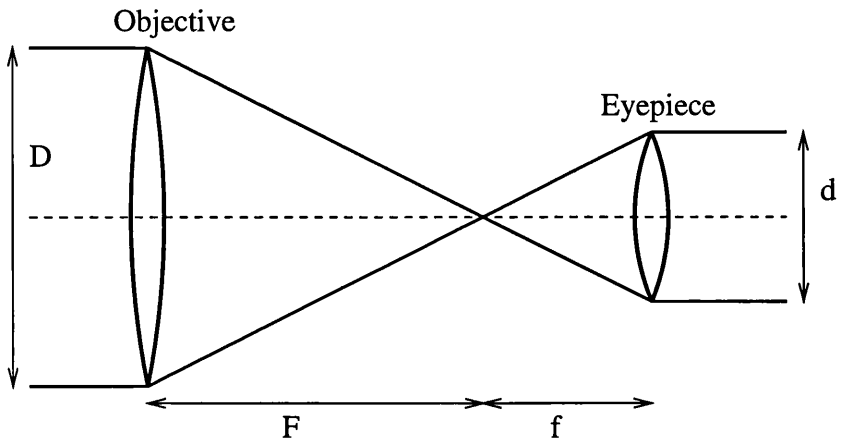


Figure 1.1: Schematic optical layout of a refracting telescope. The foci of the objective and of the eyepiece coincide, and an object observed through the two lenses appears magnified by the ratio F/f.

Telescope. However, the main drive behind this was to find a solution for the new generation of large Cassegrain instruments for 8 to 10 metre telescopes, and in particular for the High Resolution Optical Spectrograph (HROS) for the 8-metre Gemini telescope. In the case of HROS, it was concluded that unless some new way of solving the problem of flexure could be found, it probably could not be built as a Cassegrain instrument. The work in this thesis was therefore essential to the future design of HROS, and had the potential of tackling a long-standing issue in ISIS. In my last year of this thesis I was employed by the Gemini project as a part-time consultant for the issue of flexure compensation. As a consequence of this work I produced some Technical Notes, which are included in appendix C, and I presented a paper at a conference on the next generation of optical telescopes (appendix B).

1.2 Optical Telescopes

Since Galileo started using the telescope for astronomical observations at the beginning of the 17th century, the optical telescope has become the most important instrument in astronomy. Originally telescopes were *refractors*, focusing the light entirely by lenses. The refractor essentially consists of two lenses, an objective and an eyepiece (fig. 1.1). The objective lens has a larger diameter D and a longer

focal length F than the eyepiece (focal length f). The ratio of the focal lengths determines the magnification M of the refractor, that is:

$$M = F/f \tag{1.1}$$

The diameter of the objective lens, larger than the pupil of the eye (diameter d), increases the brightness of the observed object, due to its larger collecting area. If a star seen by the naked eye has a brightness B_e (total energy collected per unit time), then if viewed through a telescope it will appear to have brightness B_t given by [2]:

$$B_t = \frac{D^2}{d^2} B_e \tag{1.2}$$

The telescope therefore increases the (angular) size and the brightness of the observed object, with respect to the human eye. The use of refracting telescopes revolutionized observational astronomy and several large refractors are still in use today.

Most modern telescopes are, however, *reflectors*: they focus light trough reflections from mirrors with a conic surface. A conic curve (paraboloid or hyperboloid) has a pair of conjugated points along its axis, which represent the foci of the mirror, so that a point source placed at one of these points is re-imaged at the other by reflection in the mirror. In a paraboloid one of the two foci is at infinity (where stars are), so a parabolic mirror transforms an incoming plane-parallel wavefront into a spherical wavefront converging to a point. This explains why primary mirrors of reflecting telescopes are mostly paraboloids. Reflectors are usually preferred to refractors for several reasons. Mirrors are easier to build than lenses in large sizes and less expensive. Lenses suffer from light losses, due to the air-glass interfaces, and from chromatic aberrations, that is light of different wavelengths is focused in a slightly different position. Refractors have also mechanical drawbacks which limit their size. Because the objective can be supported only around its edge, large lenses tend to bend in the centre, distorting the image. Refractors also tend to be very long, which complicates the support system and increases the dome size.

1.2.1 Angular Resolution and Seeing

The plane-parallel wavefront from a distant point source is diffracted by the circular aperture of the telescope. As a result, the image at the telescope focus is not a point, but a circular diffraction pattern with a central bright spot containing 84% of the light (the Airy disc) and the rest in concentric rings. The first minimum of the diffraction pattern corresponds to an angle of incidence α_d given by [3]:

$$\alpha_d = \frac{1.22\lambda}{D} \tag{1.3}$$

In the equation λ is the wavelength of the incident radiation, while the factor 1.22 arises because the aperture is circular. The size of the Airy disc puts a limit on the theoretical angular resolution of the telescope: two point sources can only be resolved if their Airy discs are sufficiently separated to be seen as distinct. On the basis of Rayleigh's criterion this happens when the central maximum of one lies on the first minimum of the other. When the angular resolution of a telescope, at a given wavelength, is determined by the radius of the Airy disc, it is said "diffraction-limited".

For modern large telescopes, however, angular resolution is almost always limited by *seeing.* Seeing is the term used to describe the broadening of stellar images by atmospheric turbulence. As the light from an extraterrestrial source passes through different layers in the atmosphere, a small fraction of it is refracted, thus broadening the apparent angular size of the source. It is possible to define a seeing disc as the angular size of this broadened stellar image. For a point source, the angular distribution of light intensity is called seeing profile. To a first approximation, the seeing profile $\Sigma(\theta)$ is a Gaussian curve of the form:

$$\Sigma(\theta) = I_o \exp\left(-\frac{\theta^2}{\sigma^2}\right) \tag{1.4}$$

The diameter of the seeing disc α_s can then be defined as the full width at half maximum (FWHM) of the Gaussian distribution, obtaining:

$$\alpha_s = 2\sqrt{\ln 2}\sigma \tag{1.5}$$

For 4m and 8m class telescopes the seeing disc is much larger than the Airy disc at all but the longest wavelengths (several μm). As an example, for an 8-metre

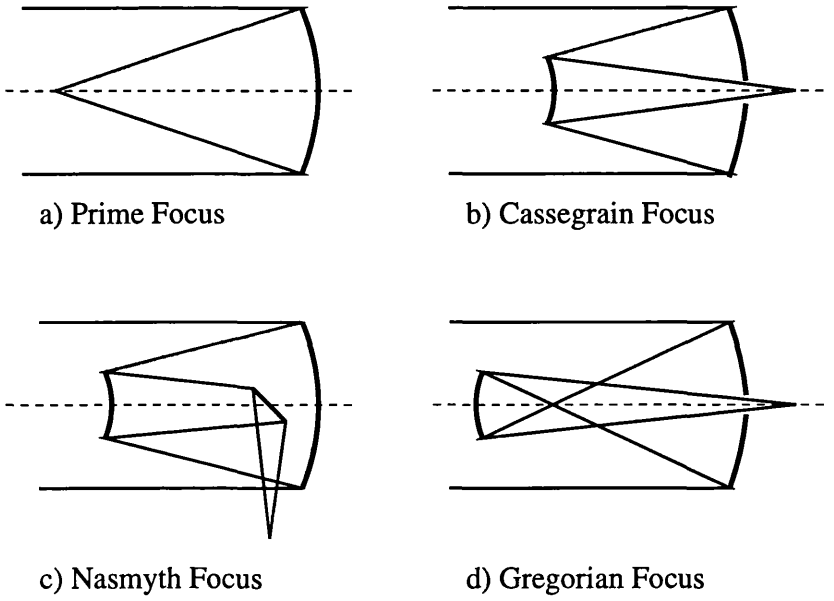


Figure 1.2: Different types of reflecting telescopes and their focal points: a) Prime focus - used mainly for imaging over wide fields. b) Cassegrain focus - the secondary mirror is an hyperboloid, the focus is below the primary mirror. c) Nasmyth focus - practically a "bent" Cassegrain. d) Gregorian focus - the secondary is an ellipsoid.

telescope and a seeing disc $\alpha_s = 1arcsec$, we have $\alpha_s = 2\alpha_d$ for $\lambda = 16\mu m$. The typical size of the seeing disc at a telescope site is very important, particularly for spectroscopic observations, because it is linked to the slit width of a spectrograph and ultimately to the instrument resolution. Typical values for a good seeing are $\alpha_s = 1arcsec$, although some of the best sites in the world experience seeing of $0.5arcsec$ or less. On Mauna Kea, Hawaii island, a seeing of $0.25arcsec$ is observed 10% of the time [4].

1.2.2 Reflecting Telescopes

I mentioned earlier that the main component for a modern reflecting telescope is a parabolic or hyperbolic primary mirror. Telescopes usually differ from each other in how the focussed light reflected by the primary is collected. The focal point of the primary mirror is called *prime* focus and it is the one with the least light losses (see fig 1.2). It is commonly used for imaging of very faint objects. In smaller telescopes the prime focus is sometimes bent at 90^o with a flat mirror and it is called

Newtonian focus. This arrangement is however impractical for large telescopes, due to mechanical unbalance in this asymmetric configuration. The commonest focus on large telescopes is the *Cassegrain*, because in this case the focal plane, right below the primary mirror, is in a convenient place for attaching instruments. In this configuration, the converging beam from the primary is reflected by a hyperbolic secondary mirror. The secondary folds the beam back and focuses it, through a hole in the primary, behind the mirror itself. The secondary is also used to change the focal length of the telescope: if h is the separation between primary and secondary and F_{pri} and F_{sec} the focal lengths, then the combined focal length is given by [5]:

$$\frac{1}{F_{cas}} = \frac{1}{F_{pri}} + \frac{1}{F_{sec}} - \frac{h}{F_{pri}F_{sec}} \tag{1.6}$$

In some telescopes the beam reflected by the secondary is bent at 90^o with a flat mirror to converge to a point on the side of the primary, called the *Nasmyth* focus. Usually two platforms, at each side of the primary mirror are used for telescope instrumentation at the Nasmyth foci. The Nasmyth focus has the disadvantage, with respect to Cassegrain, of one extra folding flat, which reduces throughput and disrupts polarization. Nonetheless, Nasmyth foci are frequently used, often together with a Cassegrain focus. Finally, the *Gregorian* telescope is a setup similar to the Cassegrain, where the secondary hyperboloid is substituted by an ellipsoid. This configuration implies a longer telescope length than in the Cassegrain, which is therefore usually preferred.

The telescope mounting is also an important element in the characteristics of refractors. The telescope mounting can be equatorial or altazimuthal; figure 1.3 shows the difference between the two. Until recently most large telescopes had an equatorial mounting. In this configuration one of the two axes about which the telescope is driven is set parallel to the axis of the Earth's rotation and is called the polar axis. The second axis, the declination axis, is set at 90^o to the first. This mounting is ideally suited to follow the motion of stars on the sky because, once the object's declination is chosen, a single rotation is used to track the star. This was the main reason for preferring equatorial mounting in the past. A single rotation at constant speed is enough to follow the star in the sky, greatly simplifying the telescope control system. The main drawback is that for a large telescope the mount structure becomes huge, with most of the telescope mass hanging from the base at an angle

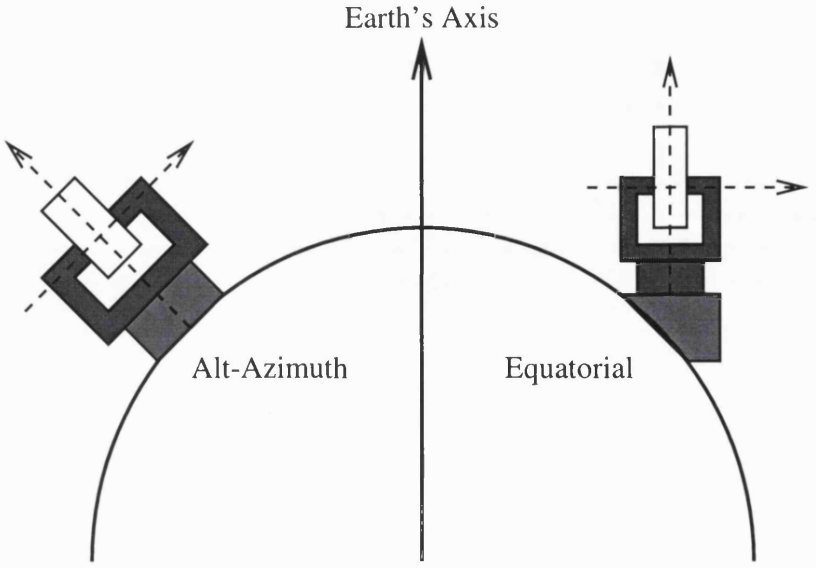


Figure 1.3: Equatorial and altazimuthal mounting for telescopes. In the equatorial mounting one axis is parallel to the rotation axis of the Earth and the other at a right angle with respect to it. In the altazimuth mounting the axes of rotation are vertical and horizontal (with respect to the ground).

(see fig 1.3); this in turn requires a large dome to contain it.

The altazimuthal mounting is much more compact and is recently becoming the preferred choice of mounting for large telescopes. In an altazimuthal telescope the two axes of rotation are simply vertical and horizontal. As a consequence, tracking a star requires a complicated combination of the two rotations, and the speed of rotation needs to vary, too. This discouraged astronomers in the past, but with the development of computers, combined with sophisticated guiding techniques this is not an issue anymore. The altazimuthal mounting has the advantage of compactness and robustness, greatly improving the mechanical stability, and it is practically the only option for 8-metre class telescopes. Indeed, the development of altazimuthal mounting and computer-controlled pointing and guiding of the telescope (pioneered at the Russian 6 m telescope) has been a necessary prelude to this new generation of optical telescopes.

1.2.3 Modern Telescopes

For many years, and until relatively recently, the 5-metre telescope of the Palomar Observatory has been the largest telescope in service in the world [6]. The 6-metre russian telescope has had great problems with the setting of the large primary mirror and its overall performance did not match that of smaller telescopes [7]. One of the reasons for this was that 5-6 metres diameter was close to the limit to which a mirror could be figured and polished with sufficient accuracy. Moreover, for the mirror to keep the correct shape, it had to be thick and extremely heavy. The primary mirror for the 6 m Russian telescope, for example, is $65cm$ thick and weights 42 tons. In addition to that, I have already mentioned that altazimuthal mounting was not a feasible option because of the problem in tracking a star. The combination of equatorial mounting with a thick solid mirror blank of large diameter made the engineering of large telescope structures extremely difficult. However, several technical and scientific improvements have allowed in recent years the diameter of the largest telescopes to increase by a factor two.

The first step was the adoption of altazimuthal mounting and computer-controlled telescope guiding which I mentioned earlier. Then there was the need to reduce the mass of the telescope primary mirror while maintaining high surface accuracy, both in terms of mechanical stiffness and stability in presence of temperature changes. Three independent strategies for building telescope mirrors in the 8-metre range have been developed:

- Hollow honeycomb mirrors
- Thin meniscus monolithic mirrors
- Segmented mirrors

Each one of these approaches has been a result of improvements both in the mechanical support and in the mirror manufacture and polishing. The different types of telescope primary mirrors are shown schematically in figure 1.4. Each of these strategies has advantages and shortcomings and in fact, all of them have been adopted for at least one major telescope project. I will briefly examine these three different strategies.

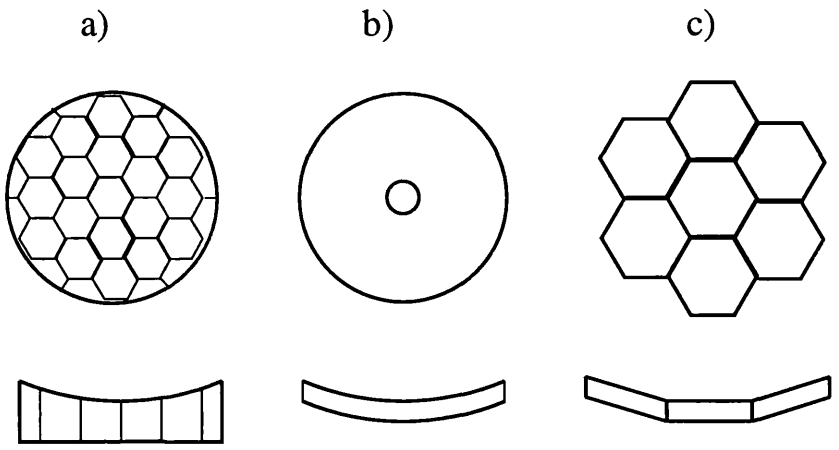


Figure 1.4: Different types of primary mirrors for modern telescopes: a) hollow honeycomb borosilicate mirror. b) thin meniscus. c) segmented mirror. The diagrams are only a schematic representation and they are not to scale. The segmented primary for the Keck telescope has 36 segments, for example.

Hollow honeycomb mirrors. The idea of producing a lightweight mirror by introducing holes in the mirror blank is not new. The challenge however, is how to put this idea into practice for an 8 metre primary mirror. Recently, techniques where a single solid blank is made out of many small hollow components, arranged in a honeycomb structure have allowed the casting of large lightweight honeycomb mirrors [8]. Such a structure has the advantage of low mass and good mechanical stiffness. However, because the borosilicate glass used in manufacturing has a relatively high thermal expansion coefficient, the mirror temperature needs to be controlled to an accuracy of around 0.1^{o} [7]. This is achieved through a sophisticated control system which detects the mirror temperature in several locations and blows air at a suitable temperature into the mirror cavities. A honeycomb borosilicate mirror was the choice of Multiple Mirror Telescope (MMT) conversion and of the Magellan project [8, 9].

Thin monolithic mirrors. An alternative way of reducing the weight of the primary mirror is to reduce its thickness. By reducing the mirror thickness, its mechanical stiffness is also decreased, so thin primary mirrors have to rely on a complex active support system, with actuators and position sensors that guarantee the accuracy of the mirror surface is maintained. An active support for a telescope

mirror normally consists in a large number (tens if not hundreds) of actuators on the mirror back surface, which are controlled by a computer so that the mirror adapts to the different telescope positions (with respect to gravity) to keep the surface in the correct shape. Until the development of computers which could handle a large amount of data in a relatively short time, this type of support system was not feasible. Moreover, sensor and actuator technology has improved considerably in recent years and at the same time several new glasses with high rigidity and low thermal expansion have been developed, which together with better polishing techniques, have produced larger, thinner mirrors of high optical quality [10]. The low thermal expansion of the mirror is very important, to avoid introducing additional distortions and stresses in the blank (apart from the active distortions that keep the correct mirror shape). Several 8-metre telescopes, like the two Gemini, the ESO Very Large Telescope (VLT) and the Subaru [11, 12, 13], have adopted the thin meniscus concept.

Segmented mirrors. A segmented primary mirror is a mosaic of several smaller mirrors which are elements of the same optical surface. The segments are supported by the same mechanical structure and by an active support system. Obviously the position of each segment needs to be accurately measured and controlled to keep the shape of the entire primary mirror to optical quality accuracy. This was the strategy adopted by the 10-metre Keck telescope [14]. The primary mirror of the Keck telescope is a mosaic of 36 hexagonal segments, which positions are controlled by three actuators each. An active control system, including several sensors which measure the relative displacement of adjacent segments, keeps the surface of the composite mirror in the correct shape. Again, the development of computers and of very accurate position sensors has been the foundation for large segmented primary mirrors. Here too, improvements in mirror manufacturing were needed: for a parabolic telescope, each segment is a relatively large non axially symmetric aspheric mirror, so new polishing techniques of large aspheric surfaces had to be developed [7].

Ultimately, each of the above strategies relies critically on innovative support systems and the advent of active optics. Active mirror supports and active and adaptive optics are described in more detail in the next chapter.

1.3 Spectroscopy and Spectral Analysers

Spectroscopy is the fundamental diagnostic tool in astrophysics, the basis for new discoveries and for confirmation of theoretical models. In this section and the next I will briefly overview the principles of spectroscopy an the relative instrumentation. The radiation from a source is characterized by its specific intensity $I(\nu)$ at a frequency ν; the *spectrum* of the source is given by the frequency dependance of $I(\nu)$. There is usually a distinction between a continuous spectrum, where $I(\nu)$ varies slowly with ν, and emission or absorption spectra, which are characterized by spectral lines (where $I(\nu)$ varies strongly in a narrow spectral interval). In practice, when measuring the spectrum $I_0(\nu)$ of any source, the instrument profile of the spectrograph $P(\nu)$ needs to be considered as well. The measured spectrum $I(\nu)$ will be a convolution of these two functions:

$$I(\nu) = \int_0^\infty P(\nu - \nu')I_0(\nu')d\nu' \tag{1.7}$$

The instrument can therefore contribute to broaden a spectral line. If the instrument profile is known, the original spectrum can be recovered by a deconvolution. In most cases the instrument profile is much narrower than the observed line and its contribution is negligible, so that the deconvolution is not necessary.

1.3.1 Spectral Resolution and Dispersion

The function of a spectral analyser is to measure the spectrum of a light source, or the quantity $I(\nu)$. In the optical fraction of the spectrum, this is achieved by *dispersing* the light through refraction or diffraction/interference, that is by using prisms, gratings or interferometers. I will concentrate on the first two dispersing elements, since they are relevant to astronomical optical spectrographs. The first parameter that characterizes a spectral analyzer is its angular dispersion. If rays of wavelengths λ and $\lambda + d\lambda$ emerge from the disperser at angles ϑ and $\vartheta + d\vartheta$ respectively, the angular dispersion is defined as:

$$A = \frac{d\vartheta}{d\lambda} \tag{1.8}$$

Together with the angular dispersion, one of the main parameters characterising a disperser, is its resolution. The resolution $\mathcal{R}$ is a measure of the ability of the

instrument to separate two distinct lines close in wavelength. If $d\lambda_m$ is the minimum wavelength separation, then the resolution is defined as:

$$\mathcal{R} = \frac{\lambda}{d\lambda_m} \tag{1.9}$$

Since dispersers are normally used with collimated (parallel) light, we can apply Fraunhofer's theory of diffraction to the light hitting the disperser. A monochromatic parallel beam passing through an aperture of linear dimension $2s$, is *diffracted* by the aperture. This means that, if a screen is placed after the aperture, a pattern of bright and dark areas is produced on the screen. The diffraction pattern, in this simple case of a linear (rectangular) aperture, has a characteristic shape. If we consider just one dimension, the intensity in the diffraction pattern is given by:

$$I(p) = \left(\frac{\sin kps}{kps}\right)^2 \tag{1.10}$$

Where $k = 2\pi/\lambda$ and ps is the optical path difference between a ray from the centre of the aperture and one emerging at the end. The function of equation 1.10 is shown in fig. 1.5. The diffraction pattern shows a maximum in the in the center and the first minimum for $kps = \pi$. If $B = 2s$ is the size of the disperser, the first minimum of the diffraction pattern would occur at an angle:

$$\vartheta_m = \lambda/B \tag{1.11}$$

Applying Rayleigh's criterion of resolution, two wavelengths separated by $d\lambda_m$ can be resolved when the maximum of the diffraction pattern of one wavelength coincides with the first minimum of the other. Given the angular dispersion A of the spectral analyzer, we have:

$$d\vartheta_m = \frac{d\vartheta}{d\lambda} d\lambda_m = \frac{\lambda}{B} \tag{1.12}$$

We can therefore derive the relation between resolution and angular dispersion in an ideal, diffraction limited disperser as:

$$\mathcal{R} = \frac{\lambda}{d\lambda_m} = B\frac{d\vartheta}{d\lambda} \tag{1.13}$$

Resolution gives also a measure of the accuracy with which it is possible to measure a line profile. If the quantity $d\lambda_m$ is much smaller than the line width $\Delta\lambda$, then the profile will be well resolved. Otherwise the instrument profile will contribute significantly to the line shape. At high resolution, with say $\mathcal{R} = 50000$, we have $d\lambda_m = 0.1\mathring{A}$ for $\lambda = 5000\mathring{A}$, so all but the narrowest lines are quite well resolved.

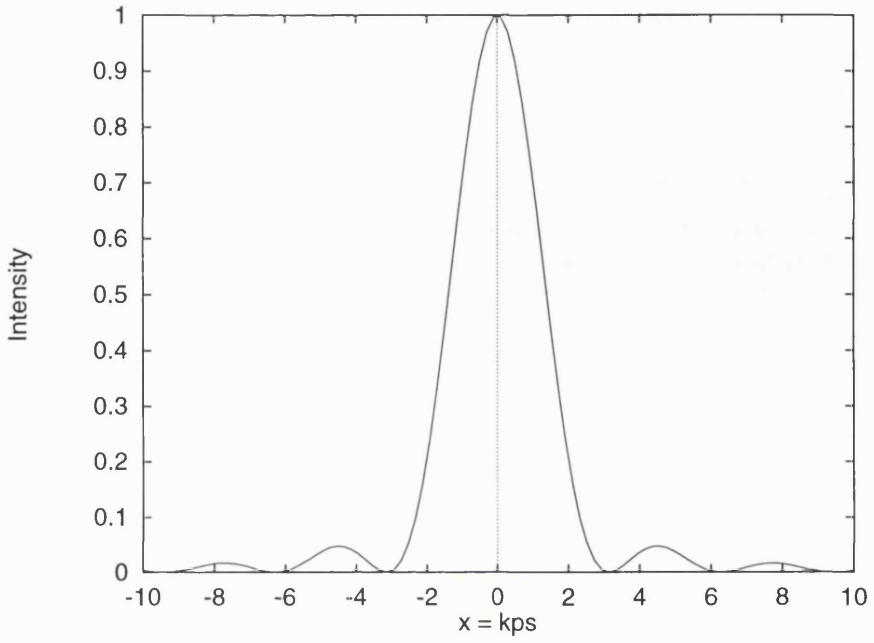


Figure 1.5: Fraunhofer diffraction from a rectangular aperture. The graph shows the intensity function in one dimension. The function has minima for $x = n\pi$ and the principal maximum for $x = 0$.

1.3.2 Prism Disperser

A prism of glass or other transmitting material has the property of dispersing the light because the refractive index n of the material is a function of the wavelength: $n = n(\lambda)$. When the prism is illuminated with light of different wavelengths, each one of them is refracted at a slightly different angle, according to the Snell law of refraction. If i and r are the incident and refracted angles at the interface between air and glass, then:

$$\frac{\sin i}{\sin r} = n(\lambda) \tag{1.14}$$

Therefore, the angle of deviation ϵ of a beam hitting the prism also depends on the wavelength. The optical setup for a dispersing prism is illustrated in fig. 1.6. If the beam has a diameter B and fills a prism with a base e, the angular dispersion of the prism A_p is given by [18]:

$$A_p = \frac{e}{B}\frac{dn}{d\lambda} \tag{1.15}$$

The resolution is immediately derived from equation 1.13 as:

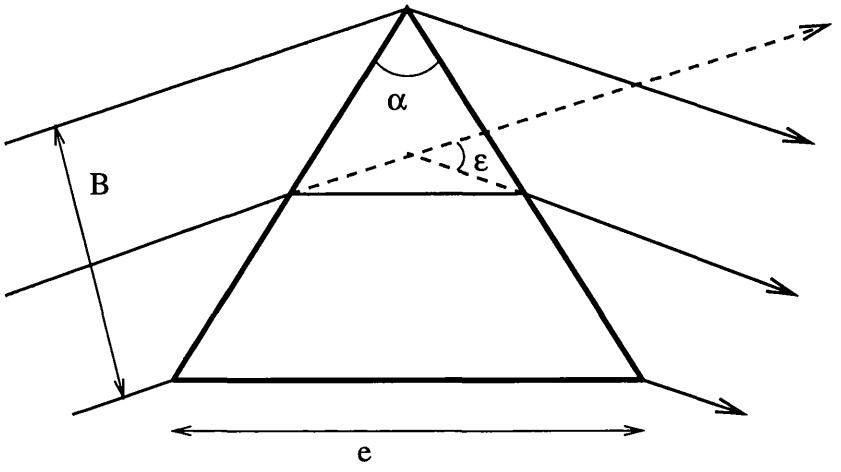


Figure 1.6: Optical diagram of a prism operating at minimum deviation angle. In this condition the passage of the rays through the prism is symmetrical. The deviation angle ϵ is defined with respect to the incoming direction. The prism base e is determined by the beam size B so that if the latter is reduced, e also decreases.

$$\mathcal{R} = e\frac{dn}{d\lambda} \tag{1.16}$$

Usually the index of refraction n depends on the wavelength in a way that can only be approximated from experimental measurements. For a prism made of fused silica at $\lambda = 4000\mathring{A}$, $dn/d\lambda = 1100cm^{-1}$ [19]. So, if the prism has a base $e = 10cm$, the (diffraction-limited) resolution would be $\mathcal{R} = 11000$.

1.3.3 Grating Disperser

A diffraction grating consists of a large number of very close, equally spaced grooves ruled on a plane or concave surface. The grating can be used in transmission or in reflection, at low (usually first) interference order or at high orders (in which case it is called echelle). The optical setup for a reflection grating is shown in fig. 1.7. The interference arises because there is a phase difference between the rays emerging from two successive grating grooves. The phase difference is determined by the difference in the optical path of the two rays, namely $BC - AD$ in the figure. The order of interference m, the wavelength λ, the groove spacing a, and the incident

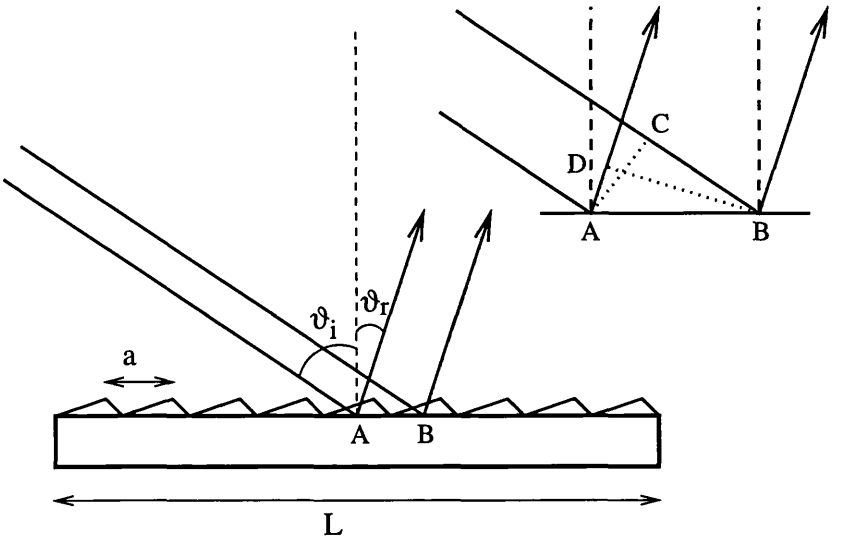


Figure 1.7: Schematic optical diagram of a reflection grating. The grating shown is of the blazed type, which is the most common. For a non-blazed grating, a represents the groove spacing. The angles of incidence and reflection are connected through the grating equation. The optical path difference for rays emerging from two successive grooves is given by $BC - AD$ and it is shown enlarged in the figure.

and reflected angles are then connected through the grating equation:

$$m\lambda = a(\sin\vartheta_i + \sin\vartheta_r) \tag{1.17}$$

Physically, the order of interference is determined by the number of wavelengths that can fit into the path difference of rays coming from successive grooves. This explains the distinction between gratings and echelles: the echelle has a much coarser ruling, so visible wavelengths are accommodated into high orders of interference ($m \sim 20$ to $m \sim 100$ typically). For a first order grating, only one wavelength can fit in the path difference and $m = 1$. The intensity profile of the light emerging from the grating is essentially the product of the combined interference of N apertures, which has a maximum (constructive interference) whenever eq. 1.17 is satisfied, and the diffraction from a single aperture (a single groove of the grating). The intensity function of the light dispersed by the grating, if s is the width of each groove and a the groove spacing, can then be written as:

$$I(\vartheta_r) = I_0 \left(\frac{\sin x}{x}\right)^2 \left(\frac{\sin Ny}{\sin y}\right)^2 \tag{1.18}$$

Where:

$$x = \frac{\pi s}{\lambda}(\sin\vartheta_i + \sin\vartheta_r) \tag{1.19}$$

$$y = \frac{\pi a}{\lambda}(\sin\vartheta_i + \sin\vartheta_r) \tag{1.20}$$

The angular dispersion of the grating can be derived by differentiating the grating equation, to obtain:

$$A = \frac{d\vartheta_r}{d\lambda} = \frac{m}{a\cos\vartheta_r} \tag{1.21}$$

The resolution of the grating can be calculated from eq. 1.13 and the expression for the angular dispersion, remembering that the beam (aperture) size is related to the grating length L by $B = L\cos\vartheta_r$, so that [20]:

$$\mathcal{R} = \frac{mB}{a\cos\vartheta_r} = \frac{mL}{a} \tag{1.22}$$

Since the ratio $L/a = N$ is the number of grooves in the grating, we have:

$$\mathcal{R} = mN \tag{1.23}$$

Using the grating equation, the above relation can also be rewritten as:

$$\mathcal{R} = \frac{L}{\lambda}(\sin\vartheta_i + \sin\vartheta_r) \tag{1.24}$$

The last expression is usually preferred, because it shows the linear dependance on the size of the grating L. A typical grating with $L = 100mm$ and having $1000lines/mm$ and working in the first order ($m = 1$) at a given wavelength, would therefore have a theoretical resolution of $\mathcal{R} = 100000$. An echelle working at the same wavelength with, say, a ruling of $50lines/mm$, would operate at $m \simeq 60$. If the ruled surface were the same, the echelle would have a resolution $\mathcal{R} = 300000$, much higher than the grating.

Another important grating parameter is the *free spectral range*, which is the range of wavelengths for which overlapping from adjacent orders does not occur. Free spectral range for the order m can be calculated directly from the definition, calculating the wavelength interval $\Delta\lambda$ for which the interference order is increased by one. This implies imposing that $m(\lambda + \Delta\lambda) = (m+1)\lambda$, obtaining:

$$\Delta\lambda_{FSR} = \lambda/m \tag{1.25}$$

The angular size of the free spectral range can also be determined from the angular dispersion as:

$$\Delta\vartheta_{FSR} = \frac{d\vartheta_r}{d\lambda}\Delta\lambda_{FSR} = \frac{\lambda}{a\cos\vartheta_r} \tag{1.26}$$

The angular FSR is important in the design of spectrographs, because it is related to the size of the optics and of the detector in the instrument.

1.3.4 Blazed Grating

The distribution of the light among the various orders depends on the shape of the individual grooves. Blazing is a technique that produces grooves with flat smooth faces, inclined to the surface by an angle ϑ_B, the blaze angle. Blazing is used to concentrate the diffracted light from the grating in some chosen angle. It is easy to demonstrate that, at the blaze peak, the blaze angle is related to the incident and reflected angles by [21]:

$$\vartheta_B = \frac{\vartheta_i + \vartheta_r}{2} \tag{1.27}$$

So if we define an angle ϑ as:

$$\vartheta = \frac{\vartheta_i - \vartheta_r}{2} \tag{1.28}$$

We can rewrite the grating equation:

$$m\lambda_B = 2a\sin\vartheta_B\cos\vartheta \tag{1.29}$$

A typical high resolution astronomical grating working in the blue (similar to the R1200B in the ISIS spectrograph), will have $a = 1/1200mm$ and, if working with $m = 1$, $\vartheta_B \simeq 15^o$ and $\vartheta \simeq 0$ to obtain a blaze wavelength $\lambda_B = 4300\AA$. An echelle instead, like the one proposed for High Resolution Optical Spectrograph for the Gemini telescope, will have $a = 1/100mm$ and $\vartheta_B = 63.4^o$. Again ϑ is very small, but the echelle works at higher orders, so in order $m = 42$ we have $\lambda_B = 4260\AA$. A blazed grating or echelle working with $\vartheta = 0$ is said to be in Littrow mode.

1.4 Astronomical Grating Spectrographs

The great majority of astronomical spectrographs are grating or echelle spectrographs (depending on the required resolution). Although the dispersing element is

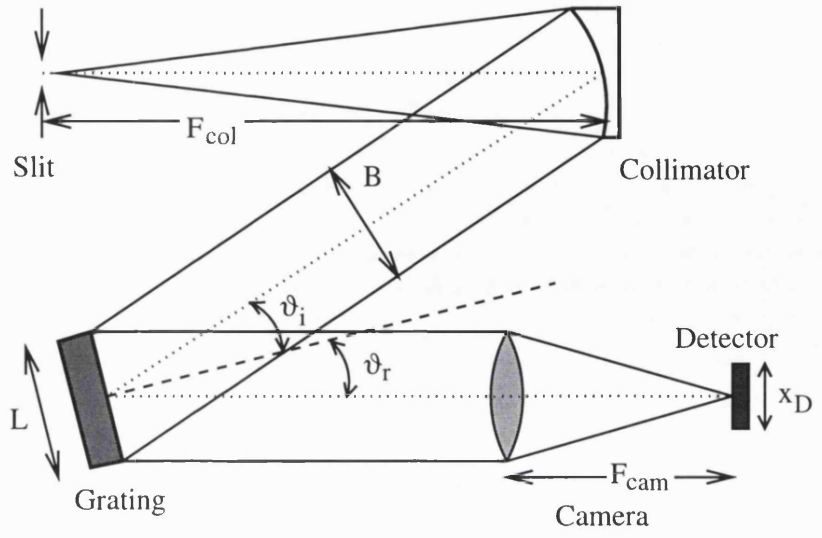


Figure 1.8: Schematic diagram of a grating spectrograph. The collimator has to match the focal ratio of the telescope (f). The slit is in the focal plane of the collimator, while the detector is in the focal plane of the camera. In a real spectrograph the camera has more optical components than the single lens shown in the diagram.

the most important component in the spectrograph, its properties depend on the overall design and even on the characteristics of the telescope that feeds the instrument. In particular, the effective resolution of the spectrograph can be significantly less than the theoretical resolution of the disperser because spectrographs operate far away from the diffraction limit. A schematic diagram of a typical grating spectrograph can be seen in fig. 1.8. The spectrograph has an input slit, which lies in the focal plane of the collimator. The collimator directs a parallel beam onto the disperser and the spectrum thus produced is collected by a camera and focused on a detector.

1.4.1 Spectrograph Characteristics

The choice of most of the parameters in a spectrograph is not free, but depends on few constants, which are usually the telescope diameter and focal ratio, the effective spectral resolution, the simultaneous spectral coverage on the detector and the detector characteristics (size and resolution element). The telescope provides the focal ratio f of the beam entering the spectrograph, and the focal ratio of the

collimator must match this. Moreover, the collimated beam, which hits the grating at an angle ϑ_i, must fill the disperser (of length L), so the collimator (and beam) size B, is determined by:

$$B = L \cos \vartheta_i \tag{1.30}$$

Therefore, the collimator focal length is also set:

$$F_{col} = fL \cos \vartheta_i \tag{1.31}$$

In the equations above, the free parameter is the grating length, which has to be chosen on the basis of the required spectral resolution. For an echelle spectrograph the camera focal length can be determined by choosing at which wavelength λ_D the free spectral length of the dispersed spectrum should fill the detector (of linear size x_D). The ratio x_D/F_{cam} of the detector size and camera focal length needs to equal the angular size of the free spectral range $\Delta\vartheta_{FSR}(\lambda_D)$, so that ($a$ is the ruling of the echelle):

$$F_{cam} = \frac{x_D a \cos \vartheta_r}{\lambda_D} \tag{1.32}$$

For a grating spectrograph other issues, like the need to account for interchangeable gratings may determine the camera focal length. When the collimator and camera focal lengths are chosen, the linear scale between the slit plane (w along slit width, l along slit length) and the detector plane (x spectral direction, y spatial direction) is determined by:

$$w = x \cdot \frac{F_{col}}{F_{cam}} \frac{\cos \vartheta_r}{\cos \vartheta_i} \tag{1.33}$$

$$l = y \cdot \frac{F_{col}}{F_{cam}} \tag{1.34}$$

In the dispersion direction (x), the additional factor with the cosines of incident and reflected rays is called *anamorphism* of the grating and it is a consequence of the dispersing properties of the grating. Equation 1.33 can be used to determine the minimum slit width. In fact, the detector will have a small, but finite detecting element size. For a charge-coupled device (CCD) or a photomultiplier, this is the pixel size, for a photographic emulsion is the grain size. For good sampling of the slit image on the detector, at least two of such elements need to be illuminated. If we indicate with p the size of the detecting element (pixel), the minimum slit width w is given by:

$$w = 2p \frac{F_{col}}{F_{cam}} \frac{\cos \vartheta_r}{\cos \vartheta_i} \tag{1.35}$$

The detector pixel size also determines the effective spectral resolution achievable with the spectrograph. Normally, two wavelengths will not be considered to be resolved if they project onto less than 2 pixels on the detector. This happens because two is considered to be minimum numbers of pixels necessary to sample a spectral line properly. The corresponding wavelength interval can be determined by imposing that the angular dispersion of this interval equals the angular size of the detector resolution element, that is:

$$\frac{d\vartheta_r}{d\lambda}\Delta\lambda_m = \frac{2p}{F_{cam}} \tag{1.36}$$

Using equation 1.21 we obtain:

$$\Delta\lambda_m = \frac{2p}{F_{cam}}\frac{a\cos\vartheta_r}{m} \tag{1.37}$$

And by definition of spectral resolution:

$$\mathcal{R}_{eff} = \frac{F_{cam}}{2p}\frac{m\lambda}{a\cos\vartheta_r} \tag{1.38}$$

Substituting the above equation into eq. 1.35 we obtain:

$$w = \frac{m\lambda}{a}\cdot\frac{fL}{\mathcal{R}_{eff}} \tag{1.39}$$

Or, using the grating equation:

$$w = \frac{fL}{\mathcal{R}_{eff}}(\sin\vartheta_i + \sin\vartheta_r) \tag{1.40}$$

We can see that, since f and $\mathcal{R}_{eff}$ are normally fixed, the minimum slit width depends essentially on the grating length L. To improve the spectrograph throughput this should be as large as possible, which implies large grating (or echelle) size. The above equation also shows that, at fixed L, widening the slit reduces the effective resolution of the instrument.

1.4.2 Spectrographs and Telescopes

The combined effect of the spectrograph and the telescope can be expressed by projecting the slit on the sky $\tilde{w}$, that is by calculating the angle under which the slit is seen through the telescope optics:

$$\tilde{w} = \frac{w}{F_{tel}} = \frac{w}{fD} \tag{1.41}$$

Where D is the telescope diameter. The projected slit width needs to be significantly smaller than the seeing disc α_s, so that the spectral lines on the detector are true images of the slit. When seeing and projected slit width are comparable in size, the image of the slit on the spectrograph detector is a product of the instrument and seeing profiles. From the astronomical point of view, the projected slit width is the one with more physical meaning. We can therefore rewrite eq. 1.40 as:

$$\tilde{w} = \frac{L}{D\mathcal{R}_{eff}}(\sin\vartheta_i + \sin\vartheta_r) \tag{1.42}$$

The above equation gives us a scale factor for telescopes and spectrographs. Since the projected slit width is limited by seeing on one side (there is no advantage in having a slit wider than the seeing disc) and by throughput considerations on the other (narrower slits have less throughput), it cannot vary much. Therefore, the above equation shows that, for a given spectral resolution, the grating size and often the whole spectrograph with it scales linearly with the telescope diameter. This is one of the major challenges that the new generation of 8-metre telescopes poses to instrumentalists: to be able to construct larger and larger spectrographs, while retaining the same imaging and stability performance of the smaller ones. The equation for the projected slit width on the sky can also be written in terms of the angular dispersion, by using the grating equation:

$$\tilde{w} = \frac{\lambda AL\cos\vartheta_r}{D\mathcal{R}_{eff}} \tag{1.43}$$

Since the projected slit on the sky is usually a parameter determined by the science requirements, the relation above gives us an alternative expression for the instrument resolution:

$$\mathcal{R}_{eff} = \frac{\lambda AL\cos\vartheta_r}{D\tilde{w}} \tag{1.44}$$

In astronomical spectrographs it is common to project on the sky not just the slit, but linear dimensions on the detector. The relation can be derived by combining equation 1.41 with equation 1.33 to obtain:

$$\tilde{x} = \frac{1}{fD}\cdot x\frac{F_{col}}{F_{cam}}\frac{\cos\vartheta_r}{\cos\vartheta_i} \tag{1.45}$$

By using the expression for the collimator focal length in eq. 1.31, the above expression simplifies to:

$$\tilde{x} = x\cdot\frac{L\cos\vartheta_r}{F_{cam}D} \tag{1.46}$$

Since the aperture of the camera must be able to cover the beam leaving the grating, we can consider that $D_{cam} = L\cos\vartheta_r$, so that we can introduce the camera focal ratio $f_{cam} = F_{cam}/D_{cam}$ to obtain simply [23]:

$$\tilde{x} = \frac{x}{f_{cam}D} \tag{1.47}$$

The above equation shows that if a given angle on the sky is to project on the same length on the detector, the product $f_{cam}D$ is to remain constant. This implies that larger telescopes (larger D), will need faster cameras (smaller focal ratios). This is also a challenge in designing spectrographs for large telescopes because faster cameras are more difficult to build and design.

Chapter 2

Spectrograph Flexure and Active Optics

2.1 Stability of Cassegrain Spectrographs

An important problem that afflicts large astronomical instruments mounted at Cassegrain foci of modern telescopes is structural flexure. This happens because a Cassegrain instrument is placed right behind the primary mirror of the telescope (figure 2.1). Therefore while the telescope is tracking a star, the instrument and its supporting structure flex under their own weight, the optical alignment is disrupted and the image shifts on the detector. In the case of a spectrograph the problem is potentially serious because these spectrum shifts on the detector can significantly degrade the quality of the data. The issue is becoming more critical as telescopes, and spectrographs, grow in size and mass. As seen in the previous section, larger telescopes require larger spectrographs, but the gravitational bending of a structural beam of length L is proportional to L^3 [24] and flexure in a spectrograph can be expected to follow a similar scaling law. Therefore if the telescope diameter doubles and the spectrograph is simply rescaled, flexure could be expected to be 8 times worse. Moreover, the telescopes currently being built are in sites with exceptionally good seeing conditions which usually translates into even tighter specification for image stability. Cassegrain instrumentation for the new generation of telescopes will have to find innovative ways of solving this issue.

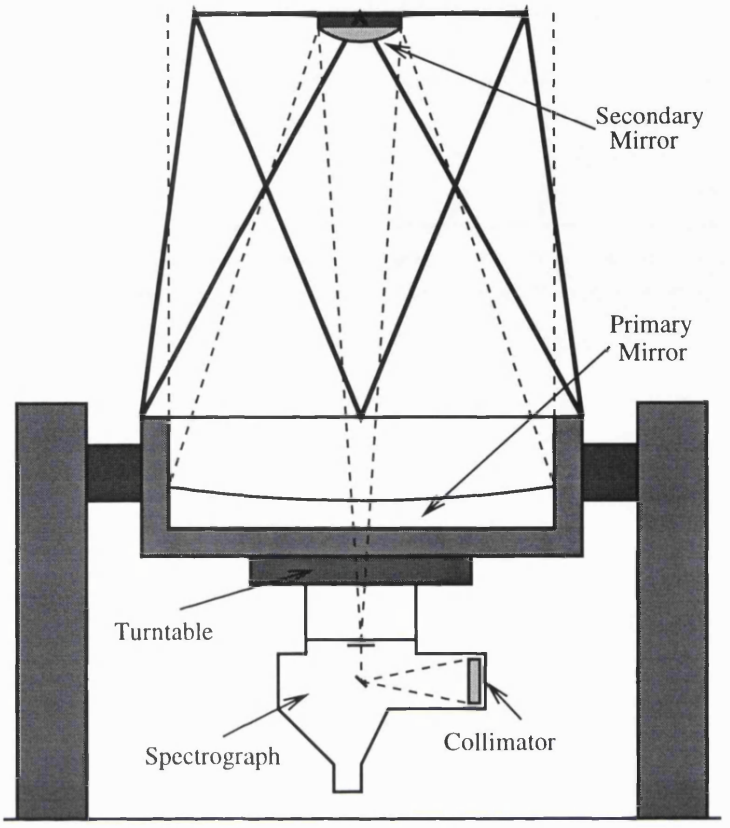


Figure 2.1: Spectrograph at the Cassegrain focus of a telescope. The instrument is placed behind the primary mirror and it is connected to the telescope structure through a turntable. This diagram is based on the ISIS spectrograph at the William Herschel Telescope, but it is typical of all Cassegrain instrumentation.

As a telescope tracks, a Cassegrain spectrograph is subject (in its local coordinate frame) to a varying gravity vector. In the case of an altazimuthal mounting, this results from the combination of altitude motion and Cassegrain turntable rotation. The spectrograph will also be subject to second-order effects due to the varying accelerations of all three telescope motions, which add small vector components to pure gravity. The spectrograph structure will therefore flex and deform, displacing the spectrum in the camera focal plane. This structural bending is mostly elastic and so it is essentially dependent on the orientation in space of the spectrograph. However, in the instrument there will be also non-elastic deformations which generate hysteresis in spectrum drifts. Hysteresis in the structure can result from a variety of causes, from microscopic motions and rubbing within bolted joints, through defects

in the crystal structure of the structural material, to the build up and release of stresses within it. Also, some internal components may suddenly move as the sign of the gravity vector in the component frame of reference reverses. For example, this may occur as the plane of a mirror surface passes through the zenith and the mirror "flips" in its mount, or a bearing-mounted mechanism may rock from one side to the other due to bearing play. Clamping moveable components can stress parts of a mechanism or structure, and the stresses can suddenly and unpredictably be relieved as the local gravity vector changes.

The combination of all the nonlinear processes has the effect that the current image position on the detector, depends not only on the instrument orientation in space, but on the recent history as well. 'History' includes the direction, speed of motion and acceleration to which the instrument was subject in reaching a certain position and also includes temperature changes. In general, these linear and nonlinear effects contrive to produce some or all of the following: a secular (smooth) image motion on the detector, modulated by higher frequency "wobbles" and occasional sudden "jumps". The combination of all these effects is commonly identified under the general name of flexure (rather than limiting it to the pure gravitational bending).

2.1.1 Flexure and Spectrograph Instrument Profile

There are several factors that may contribute to change the shape and width of the instrument profile of a spectrograph. I will not consider the effect of atmospheric seeing at this stage, assuming a uniformly illuminated slit. I will also assume to observe an unresolved spectral line, whose intrinsic width (when calculated in microns on the detector) is much smaller than the projected slit width. In this case the theoretical slit profile on the detector $w(x)$ is a rectangular step, with a width determined by the spectrograph resolution. The effective slit profile on the detector $\omega(x)$ however, is not exactly a rectangular step and will in general depend on:

- Theoretical slit profile $w(x)$
- Optical aberration function $a(x)$
- Flexure shifts Δx_f

- Diffraction
- Internal seeing

In practice the last two items are usually negligible in a spectrograph, but flexure and optical aberrations contribute significantly. In fact, the diffraction by the spectrograph optics will depend on the size B of the parallel beam illuminating the grating, producing a diffraction spread of the order of $\vartheta_d = \lambda/B$. The angular resolution of the spectrograph camera is instead determined by $\vartheta_c = 2p/F_{cam}$ ($2p$ is the projected slit width on the detector), which is usually much larger than the diffraction angle. Internal seeing is produced by air turbulence within the spectrograph and it is usually relevant only for relatively large temperature changes.

To calculate the effect of flexure on the theoretical slit profile $w(x)$, we will assume that flexure will shift the spectrum always in the same direction and at a constant rate χ. The unperturbed slit profile will be a rectangular step of width $2s$, as shown in fig. 2.2. The mathematical expression for that profile is (in the equation *sgn* is the sign function, which has a value $+1$ when its argument is positive and -1 when it is negative):

$$w(x) = \frac{I}{2}\left[\mathrm{sgn}(x+s) - \mathrm{sgn}(x-s)\right] \tag{2.1}$$

Due to flexure, the projected slit position on the detector will be shifted while the detector accumulates photons during the exposure. Thus at a time t, the photons will hit the area of the detector with ordinate x, such that:

$$-s + \chi t \leq x \leq s + \chi t \tag{2.2}$$

A the same time, the intensity also increases: $I = I_o t$. We can therefore write the time-dependent (instantaneous) slit profile as:

$$w(x,t) = \frac{I_o}{2}\left[\mathrm{sgn}(x+s-\chi t) - \mathrm{sgn}(x-s-\chi t)\right] \tag{2.3}$$

The resulting slit profile, as recorded by the detector is obtained by integrating the above expression over t:

$$\omega(x) = \frac{1}{T}\int_0^T w(x,t)dt \tag{2.4}$$

The time of integration T is arbitrary and we can take $T = 1$. The total flexure shift during the exposure will be given by $\Phi = \chi T$. We can also normalize the intensity

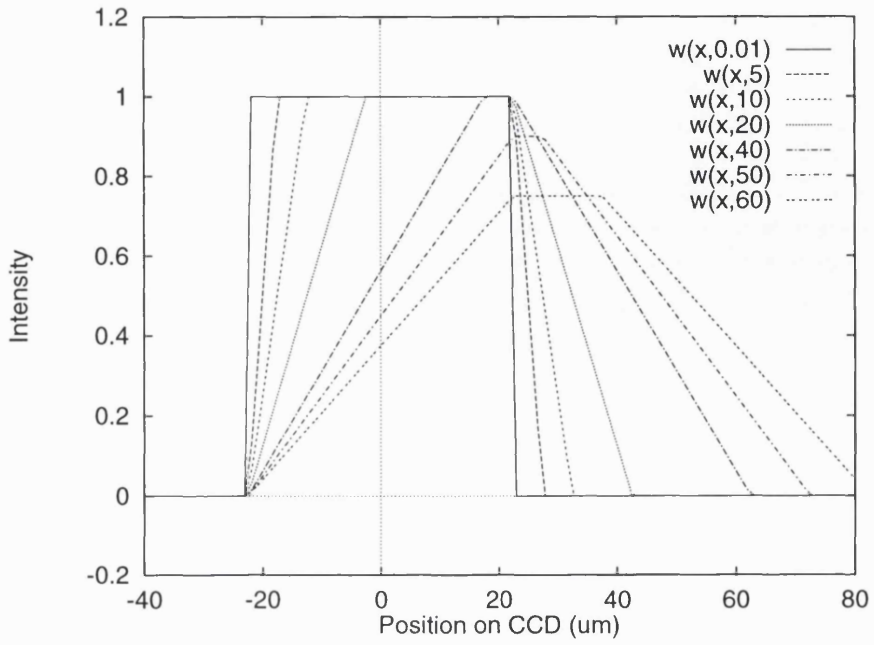


Figure 2.2: Convolution of the theoretical projected slit profile with several constant-rate flexure shifts. The projected slit width is $45\mu m$, the convoluted function has still the same FWHM as long as $\Phi \leq 45\mu m$.

to 1, so that the above integral becomes:

$$\omega(x) = \frac{1}{2\Phi}\left(\mid x - s - \Phi \mid - \mid x + s - \Phi \mid + \mid x + s \mid - \mid x - s \mid\right) \tag{2.5}$$

The above function is shown for different values of Φ in fig 2.2. The figure shows that as long as $\Phi \leq 2s$, both the peak and the FWHM of the profile is still unchanged. The profile as a whole is however changed in shape and its centre is shifted. The centre of the projected slit can be calculated to be shifted by:

$$\Delta x_f = \frac{\Phi}{2} \tag{2.6}$$

We can therefore see that, except for shifts larger than $\Delta x_f \sim 20\mu m$ flexure has a small impact on resolution, but it changes the absolute wavelength calibration of the spectrum. When optical aberrations are taken into account, we obtain a similar result, although the critical level of flexure is somewhat smaller. I will examine the implications of this result in section 2.5

2.1.2 Atmospheric Seeing, Flexure and Resolution

I mentioned earlier that when the size of the seeing disc becomes comparable with the projected slit $\tilde{w}$, the instrument profile is affected. When this happens, the slit is not uniformly illuminated anymore and the theoretical slit profile (without optical aberrations) is not a simple rectangular box. Algebraically, the slit profile $w(x)$ can be written as the product of the theoretical one (equation 2.1) and the seeing profile, that is:

$$w(x) = I_o \exp\left(-4\ln 2\frac{x^2}{\alpha_s^2}\right)[\mathrm{sgn}(x+s) - \mathrm{sgn}(x-s)] \tag{2.7}$$

In the case of ISIS, the William Herschel Telescope has $D = 4.2m$ and a $f/11$ focal ratio. Therefore, according to equation 1.41, a slit $w = 150\mu m$ wide would project to $\tilde{w} = 0.67arcsec$ on the sky. This value needs to be compared with typical seeing at the site: sometimes the seeing disc in La Palma can be smaller than $\alpha_s = 0.7arcsec$ even if the mean value is around $1.0arcsec$ [25]. The above values are comparable with the projected slit width, so we cannot ignore the effect of seeing on the instrument profile.

In principle a small seeing disc has a beneficial effect on the astronomical observation. Equation 2.7 shows that the main effect of a small seeing disc is to reduce the width of the instrument profile without reducing the energy throughput. For object of same brightness, the energy throughput is actually increased, as figure 2.3 shows. Most of the energy from the star is concentrated in an area smaller than the one covered by the slit, so throughput is maximized. This in turn increases the spectrograph resolution, which is defined in terms of the width of the instrument profile. In practice however, this small gain is nullified by a reduced sampling accuracy on the detector.

A very good seeing can, however, have also an indirect detrimental effect on the spectral stability. Since in this case the slit and the seeing profile are closely matched in size, it is extremely important that their relative position does not change during the exposure. When the seeing disc is much larger than the slit, any small shift of this disc would have very little impact on the slit illumination, which can still considered uniform. In good seeing conditions instead, a small shift of the seeing disc would change the illumination of the slit, thereby shifting the spectrum on the detector. To a first approximation, the (angular) shift required to produce a shift

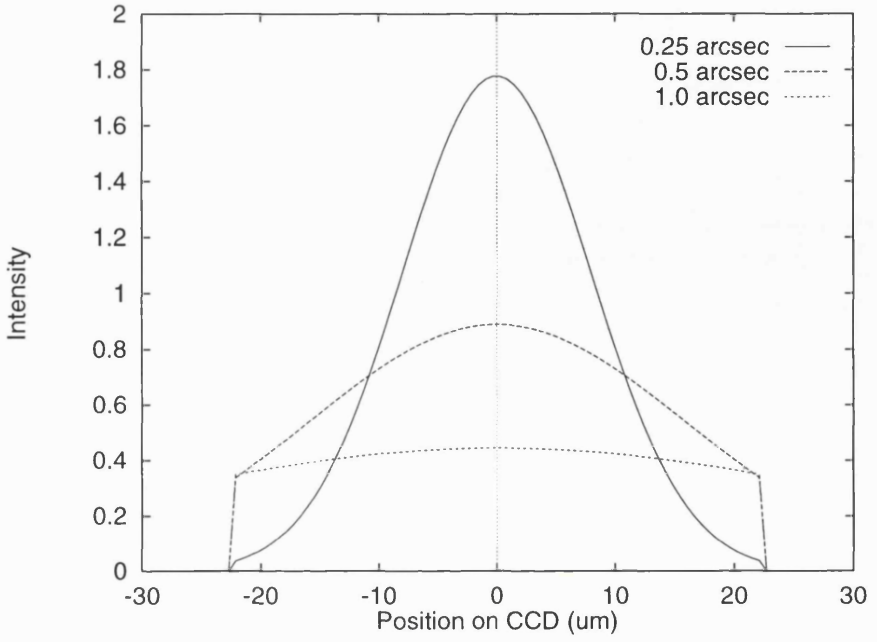


Figure 2.3: Effect of different seeing conditions on the instrument profile. The observed object is assumed to have the same brightness in the three curves. The profile corresponding to $0.25 arcsec$ seeing has a higher intensity because most of the light goes through the slit. A seeing disc $1.0 arcsec$ wide produces a profile close to the one expected for uniform illumination of the slit. The slit is assumed to project onto $0.6 arcsec$ on the sky and $45 \mu m$ on the detector.

x on the detector is given by equation 1.47. In the case of ISIS, with a camera focal ratio $f \sim 3$, a shift $x = 1 \mu m$ on the detector is produced by a shift of only $0.02 arcsec$ of the seeing disc.

The stability in the position of the seeing disc is determined by the guiding accuracy of the telescope and it is very important for the stability of the spectrum on the detector. Guiding accuracy, being a telescope pointing issue, is independent from the spectrograph stability and should be separated from the internal flexure in the spectrograph. It is necessary to be aware, however, that high guiding accuracy is of paramount importance for the image stability, and it is a basic assumption of the work in this thesis. High guiding accuracy in modern telescopes is normally achieved by using a star near to observed object as a reference for the telescope pointing. With this technique, the stability of the example above is not difficult to

achieve. If, say, the reference star produces a seeing disc $\alpha_s = 0.5arcsec$ in diameter, the stability required ($0.02arcsec$) is 1/25th of the seeing disc size.

2.1.3 Spectrographs for Large Telescopes

Astronomical spectrographs are usually classified on the basis of their resolution. Low to medium resolution usually means $\mathcal{R} \leq 10000$, while high resolution refers to the range with $\mathcal{R} \geq 50000$ [26]. High resolution can be achieved only by cross-dispersed echelle spectrographs (and Fourier transform spectrometers). Most telescopes have both a high and a low dispersion spectrograph among their instruments. Of the 8-metre class telescopes, only the Keck is currently fully operational. To this date, it has a low resolution spectrometer (LRIS) for the Cassegrain focus [27, 28] and a high resolution instrument (HIRES) at the Nasmyth focus [29]. This configuration, with low resolution at Cassegrain and high resolution at Nasmyth (or at the Coude focus), is typical of many large telescopes. Instrument flexure plays a part in this because the spectrum of a cross-dispersed echelle is more sensitive to calibration errors and flexure shifts. High resolution instruments also tend to be larger, which makes the problem of flexure worse. In this section I will concentrate on the stability performance of Cassegrain spectrographs for 4 and 8 metre telescopes.

The case of the very large Low Resolution Imaging Spectrometer for the Keck telescope is particularly interesting because it is the first Cassegrain instrument for a 8-metre class telescope. It had a very ambitious design goal of $< 10\mu m$ image drifts for telescope movements of 60^o in elevation [27]. After commissioning however, it showed worst case flexure shifts of 110 to $140\mu m$ over the whole sky (elevation 20^o to 90^o). To my best knowledge, the telescope engineers are still trying to identify the main sources of flexure in the attempt to reduce it. It seems clear that the problems encountered in designing and constructing LRIS, using a traditional passive approach, and in particular the constraints of the mass budget and space envelope, can be expected at most of the Cassegrain instrumentation for 8-metre telescopes. This is one of the strongest evidence in support for the need of an innovative approach to the problem of flexure.

Even in spectrographs for smaller telescopes flexure has been a major issue. The echelle spectrograph for the McDonald Observatory [30], for example, showed more

than 1.0 pixel of flexure ($27\mu m$ pixel size) over 60^o, a value much greater than expected on theoretical grounds. It was only after several months of work which included adding stiffening supports, that the flexure could be reduced to 0.3-0.4 pixels for the same telescope motion. Other recent instruments have had fewer problems: the CFHT multi-object spectrograph (MOS/SIS) currently has around $10\mu m/hr$ spectrum drifts, and it is anticipated that should improve to $\sim 4\mu m/hr$ once the guiding system for the spectrograph is completed [31]. It is important to point out that this high accuracy guiding system is based on a tip-tilt active mirror, driven by guiding information collected from objects in the area around the imaging field. With pure passive structures, even some smaller instruments, like the LPL (Lunar & Planetary Lab) echelle spectrograph which is $1.22m$ long, have shown significant flexure: the LPL spectrograph has up to $25\mu m/hr$ drifts [32]. All these examples prove that flexure is at best a very difficult problem to solve with traditional methods and at worst a serious limitation to the spectrograph performance. It is for this reason that I started looking at active optics to find a possible solution.

2.2 Active Optical Systems

It is difficult to give a precise definition of active optics. In its most general description, active optics concerns with an optical component which is moved, deformed or controlled in any way by an "intelligent" system, which may be a computer, a sensor feedback or both. In practice, different scientific communities have a slightly different interpretation of the term. So, for example, astronomers distinguish it from *adaptive* optics, which refers to the specialized branch of optics that concerns with deforming a mirror surface to correct wavefront aberrations as measured by a separate wavefront sensor. The distinction is normally based on frequency response, with an adaptive system capable of much faster corrections (KHz) than an active one (Hz). In this respect the deformation of the surface of the primary mirror of a telescope is described as active, because of the slow response (typically around $0.03Hz$ [33]).

For the purpose of this discussion I prefer to stick to the general definition of active

optics, so that adaptive optics can be considered one of its sub-divisions. The essential point of active optics is that the properties or the position of an optical component is *actively* controlled in some way. Even if the distinction is based on frequency response, adaptive optics is obviously a subset of active optics in that a component that moves fast, can also operate at lower frequencies. Usually the active optical component in active optics is a mirror, even if any optical element can in theory be used. The control system is in most cases a combination of sensors and a computer.

One of the most important components of an active optics system, is therefore the active mirror. There is a great variety of active mirrors and applications, ranging from laser optics to ground based astronomy and space technology. These mirrors can be classified in different ways according to the definition chosen for active and adaptive optics. So we have fine steering mirrors, fast steering mirrors, adaptive tip-tilt mirrors, adaptive deformable mirrors, active deformable mirrors and so on.

2.2.1 Adaptive Optics

In the astronomical community the "adaptive" term is strictly limited to an optical component designed to compensate fast wavefront distortions introduced by the atmosphere of the Earth when observing an extraterrestrial object. The adaptive mirror can then be separated from the rest on the basis of its purpose. This is a somewhat arbitrary classification, but it is one that it is widely accepted. When the light from an astronomical object crosses the atmosphere to reach a telescope, the originally plane wavefront (the star being effectively at infinity) is distorted by turbulence in the air, blurring the star image (seeing). These wavefront distortions can however be measured and, if the system response is fast enough, corrected by an adaptive optics system, like the one shown schematically in figure 2.4. Adaptive optics is becoming increasingly popular among astronomers, to the point that nowadays almost every major ground-based telescope (existing or planned) has an adaptive optics program. This is because by correcting the image blur introduced by the atmosphere, an adaptive optics system can dramatically improve the angular resolution of the telescope. I showed in chapter 1 that modern large telescopes working in the visible and near infrared are always seeing limited. Adaptive optics reduces

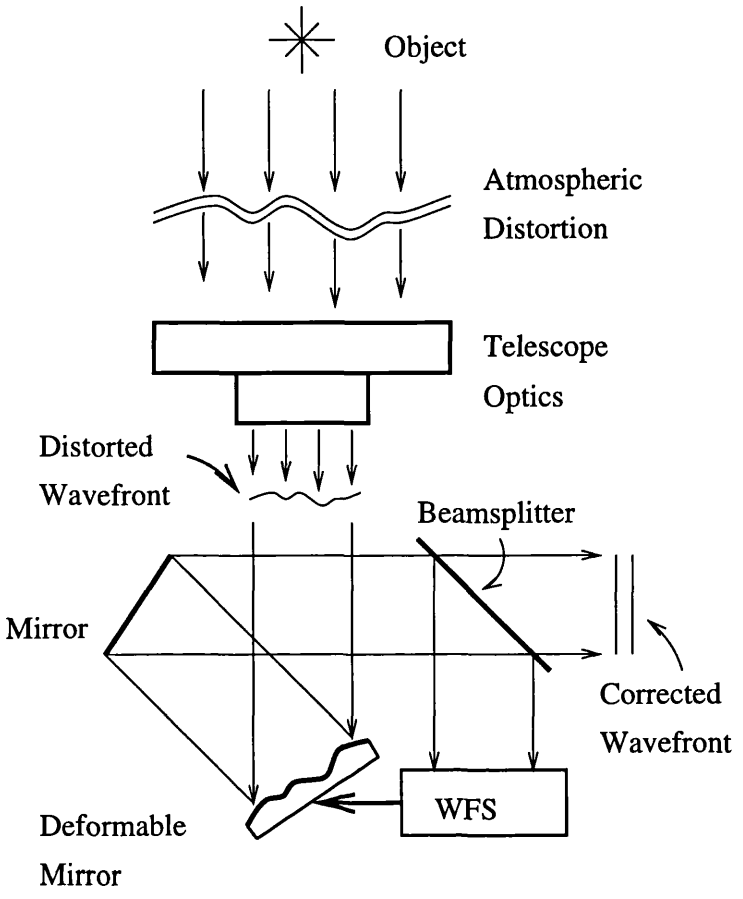


Figure 2.4: Schematic adaptive optics system. The light from an astronomical object is distorted by the turbulent layers of the Earth's atmosphere. The adaptive optics system detects the wavefront distortion with a wavefront sensor (WFS) and the information is used to modify a deformable mirror which compensates the distortion. The corrected wavefront at the output can then be fed to any instrument.

the size of the seeing disc, pushing the telescope performance towards the diffraction limit.

In theory an adaptive optics system could use a fraction of the light from the observed star to measure the wavefront and leave the rest for the astronomical observation. This is however undesirable because some of the light from the star is lost. The solution is to use a reference star for the wavefront measurement: if a star could be found close enough to the observed object, it is reasonable to assume that the light from that star has suffered the same wavefront distortions. The limitation is of course set by the chance of finding a nearby star for any given object. Once the

wavefront from a reference star is measured, the information about its distortion is fed back to the mirror, which continuously changes its shape so as to compensate for these distortions. The accuracy of the correction will depend on the response time of the system and on the accuracy with which the wavefront is measured and corrected. Higher levels of adaptive corrections can be achieved by mirrors with more degrees of freedom, that is by those with more actuators. In fact, the larger the number of actuators, the more closely the adaptive mirror can emulate the shape of the distorted wavefront and the better the correction. The accuracy of the adaptive correction will also depend on the wavefront sensor because of the limited sampling of the wavefront the sensor can provide. Moreover, an adaptive system needs a fast response time, because of the speed at which the turbulent processes in the atmosphere evolve.

Several adaptive optics systems are currently working on large telescopes and many more are planned or under construction. They usually differ on the type of adaptive mirror used, in a similar way as telescopes do. Some adaptive mirrors are of the thin deformable type, like the one used in the COME-ON system [34, 35] at the ESO La Silla Observatory. Others are made of several rigid tip-tilt segments, like MARTINI [36, 37] or just consist of a single rigid mirror for tip-tilt correction [38, 39].

2.2.2 Active Mirrors

Active optics is becoming widely used, for the primary mirrors for 4 and 8 metre class telescopes. As we saw in the previous chapter, the development of active primary mirrors is one of the main technological advances that made 8 metre telescopes possible. The concept of using active systems to maintain the surface accuracy of a large telescope primary mirror was first developed at the ESO New Technology Telescope (NTT) [40, 41, 42]. This is a thin meniscus mirror, $3.5m$ diameter supported on 78 points (figure 2.5). The idea behind this project was to produce a telescope that had the means not only to measure in real time its own errors, but also had the means to correct them without disturbing the observation. In practice, the NTT analyses the light from a guide star in the telescope field of view to measure the optical quality of its own optics and their alignment. This information is then used to actively control the primary mirror with the actuators at the back.

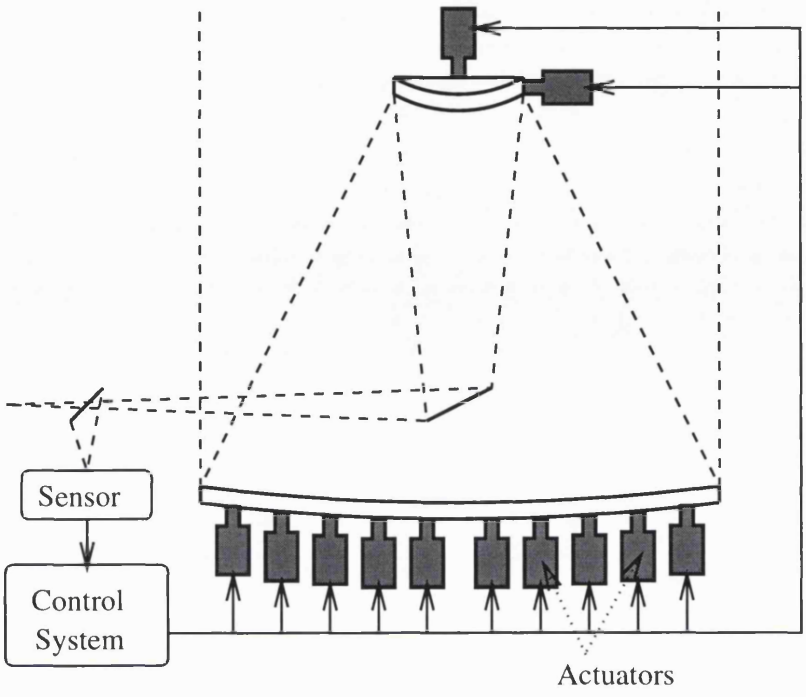


Figure 2.5: Schematic active optics system at the ESO New Technology Telescope. The primary mirror is supported on 78 actuators which control its surface accuracy. The secondary mirror is also actively controlled for focusing and alignment. Most active primary mirror supports on large telescopes resemble this pioneering design.

The success of this new concept, prompted several other large telescope projects to adopt a similar strategy. Since then, an active primary mirror has been implemented on the 10-metre Keck Telescope [43, 44] and proposed for several new telescopes [11, 45, 46]. As explained in chapter 1, active optics is the fundament of the new generation of ground-based telescopes. Active optics is applied to telescope secondary mirrors as well, both for alignment and chopping when observing in the infrared [47]. And active optics is not limited to ground based applications: in space technology the use of active mirrors has great advantages, allowing remote tuning and adjustments from Earth. Nor are active mirrors only employed in astronomy: in laser optics, for example, fine steering mirrors can be used for accurate pointing and stabilization of laser beams [48, 49].

All these different types of active mirror systems have in common the same basic structure: the mirror is controlled with several actuators, usually mounted at the

back of the optical surface, some sensors give indication of the mirror position or of the wavefront reflected by it and a computer system or electronic feedback adjust the mirror according to the information collected. The number of actuators can vary from two or three for simple tip-tilt motions to tens or hundreds for more complex surface deformations. Their type and characteristics varies widely, according to the specific application. So does the sensor type and how the information is processed. In the next section I will discuss how active optics can be used to reduce the effects of flexure in a Cassegrain spectrograph. In chapter 3 I will analyze the various characteristics and options for actuators and sensors for active mirrors. This review was part of the preliminary study for the feasibility of an active system for compensating flexure in the ISIS spectrograph.

2.3 Active Optics and Flexure

The problem of flexure in large Cassegrain spectrographs has been described in in section 2.1. This problem has become increasingly difficult to solve with traditional methods for the new generation of telescopes and their Cassegrain instruments. The use of active optics for reducing flexure in a spectrograph is a totally different and new approach to the problem. The idea in itself is fairly simple: an active element is introduced in the spectrograph and, provided that the flexure shifts are known, its position is actively controlled with a computer to keep the spectrum position on the detector stable. The idea of flexure compensation using an active mirror is illustrated in fig. 2.6

The idea of active flexure compensation has only recently been proposed by Bhatia [50] as a way of easing the pressure on the mechanical design, improving the image stability and even correcting optical aberrations. The proposed system is much more complex than the one I adopted for ISIS because it attempts to correct flexure and optical aberrations at the same time by using a deformable active grating. The engineering challenge in that approach is such, that the feasibility of the whole system within reasonable cost and time is questionable. To this date, the idea of an active deformable grating has not been developed beyond the conceptual level. The idea of a simpler flexure compensation system, however, is starting to grow.

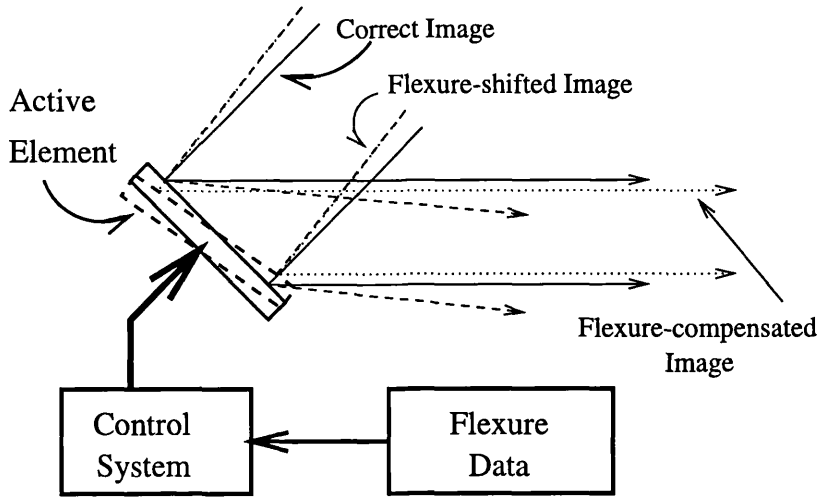


Figure 2.6: Compensation of flexure using active optics. An active element (in this case a tip-tilt mirror) is introduced in the optical path of the object's image. The tilt of the mirror compensates for the flexure-induced shifts and restores the image in the original position.

An active flexure compensation system is also currently being investigated for the Gemini Multi-Object Spectrograph [51], following a study of the problem for Gemini performed in this laboratory [52]. The growing support gained among spectrograph designers is also reflected by some very recent developments. A system similar to the one I tested in ISIS, based on an active collimator, is proposed for the DEIMOS spectrograph for the Keck II telescope [54] and possibly for the ESI spectrograph as well (on the same telescope) [55]. Before the experiment described in this thesis, active compensation of flexure had never been attempted on a Cassegrain spectrograph. Therefore, there were many the challenges and uncertainties in developing such system. It is encouraging to see, however, that this experiment has pioneered a trend in the design of Cassegrain instrumentation, hopefully leading to a new generation of "active" spectrographs.

Even if the basic concept of flexure compensation is straightforward, there are complex technical issues that need to be tackled to produce an effective working system. These issues include the choice of the active element and the actuation system, how to get the necessary information about spectrum shifts and what type of correction to apply to compensate the motion. Given the vast number of different types of

actuators and sensors available, I dedicated a separate section to this issue. I will discuss the remaining points in this section.

2.3.1 The Active Element

When the spectrograph flexes under its own weight, the optical alignment of the spectrograph is disrupted and the image of the slit on the detector drifts away from the original position. The purpose of an active optical system is therefore to detect or calculate this shift and to operate an active element to restore the proper alignment. In theory, there are several options for this active element:

1. An additional folding mirror or prism. An extra mirror or prism could be introduced in the optical path of the spectrograph and its position adjusted to correct the flexure shifts. The drawback in this option is that extra surface losses of light are introduced and the throughput of the instrument is reduced.

2. The dispersing grating. The advantage in this case is that the grating is normally in the pupil plane of the spectrograph and it is being hit by a parallel beam of collimated light. This would ease the stability requirements on the system. This option, however, would be problematic for a spectrograph with interchangeable gratings or with very large and heavy echelles. This was the option proposed by Bhatia in its paper [50].

3. The collimator. The tilting of the collimator could provide the necessary adjustment. The advantages are its simplicity and versatility: the collimator is normally present in all the spectrograph configurations and once it has been properly focused and aligned, its position does not need changing. A small tilt of the collimator would adjust the angle of the parallel beam hitting the grating and consequently the spectrum position in the camera focal plane. Also, the wealth of literature in active mirrors described earlier can be utilized for the mechanical design. This was indeed the option I chose for active flexure compensation in ISIS and the one proposed for DEIMOS [54].

4. The detector. The detector could be translated in the camera focal plane to restore the image in the correct position. This idea is intrinsically very simple.

However, interchangeable detectors and the use of cryogenically cooled CCDs, complicate matters considerably. Since moving the whole cryostat may be impractical, the CCD may need to be translated within the cryostat, making the actuator choice and interface more difficult. This is the option proposed for the Gemini Multi Object Spectrograph, which has only transmission optics, so that some of the other options would not be viable [53].

5. The Slit. In place of the detector, the slit could be translated instead. The slit position however, has important implications for the telescope acquisition and guiding system. It is also the main reference point for the spectrograph optics, so that, even if possible, it may not be desirable to change its position.

2.3.2 Flexure Detection: Open and Closed Loop

The active system needs to have information about the amount of spectrum shifts during tracking to be able to compensate them. There are essentially two ways in which this information can be given to the active element controller: it can be measured in advance, stored in a look-up table and retrieved during operation or it can be measured directly by propagating an optical reference through the spectrograph. When the flexure data is stored in a look-up table we talk about an *open loop* system because it relies on external information to derive the necessary corrections to the active element. When instead the system itself measures the flexure shifts via an optical reference, we talk about a *closed loop* system, since its operation can be totally independent from outside intervention. The principle of open and closed loop flexure correction is illustrated in fig. 2.7

In an open loop flexure compensation system, a model for calculating spectrum shifts for every telescope and spectrograph orientation needs to be produced. A simple way of doing this is to move the telescope and the spectrograph in steps to cover the whole sky and then to calculate flexure at any point by interpolation. In ISIS flexure is measured is by stepping the telescope in elevation for different positions of the spectrograph on the Cassegrain turntable. Alternatively the telescope position could be stepped in declination of the observed object for different values of the hour angle. This is an approach used, for example, at the Asiago Observatory [56]. The

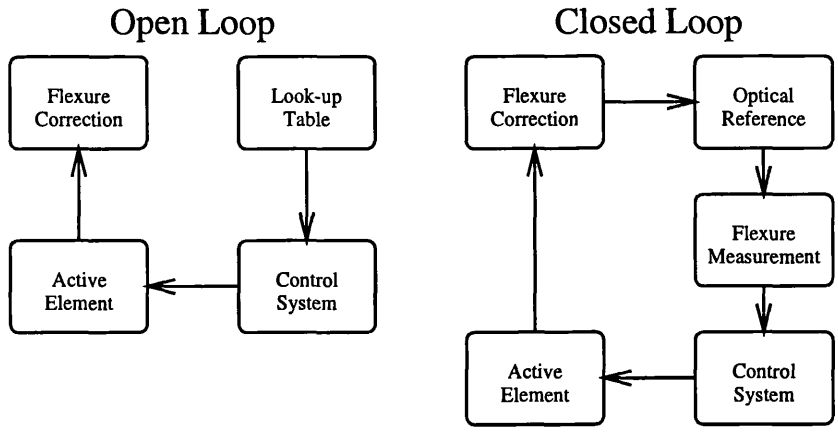


Figure 2.7: Open and closed loop flexure correction.

main advantages of an open loop system are its simplicity and versatility: different spectrograph setups and modes of operation can be handled by separate look-up tables. It also confines the active compensation system hardware to the active mirror and relative controller, making it easier to implement in a spectrograph, particularly an existing one. The key point concerning open loop correction is that hysteresis is the factor which limits the the precision with which the image motion can be compensated. If flexure has some irreproducible behaviour, this will translate into modelling errors and ultimately lower stability. Moreover, even if hysteresis can be reduced to a minimum, the long-term accuracy of a flexure model can be an issue, and frequent updates to the look-up table may be necessary.

In a closed loop system flexure is measured in real time during tracking. The system relies on an optical reference which propagates through the spectrograph and drifts in the same way as the science spectrum. By sensing the movements of this reference, the information can be fed back directly to the active element control system. Since the position of the reference image can in theory be measured extremely accurately, a closed loop flexure compensation system is intrinsically more accurate than an open loop one. Moreover, such a system can work in the presence of large hysteresis and even of some malfunctions in the spectrograph (e.g. a loose component). The key issues in a closed loop system are concerned with the optical reference. Usually a spectrograph has different setups and modes of operation, which makes extremely difficult to propagate the reference through the spectrograph for each of these options. A closed loop flexure compensation system contrasts directly

with the spectrograph versatility: the more the versatility, the more complex (and possibly less accurate) the compensation system. The optical reference in itself may be a problem, too because it can introduce stray light in the spectrograph. However, if these problems can be overcome, the potential spectrum stability achievable with a closed loop system, far superior to the open loop one, makes it very attractive.

2.4 Closed Loop Flexure Correction

Several unresolved problems are connected to a closed loop flexure compensation system. In this section I will try to review them and to propose some possible solutions. The major critical points are:

- The choice of the light source for the optical reference.
- How to inject the the light from this source into the spectrograph.
- How to overcome the dispersion of the light at the grating when this can be set for several different wavelengths.
- The choice of the type and the position of the detector that senses the reference.
- How to avoid stray light and contamination of the science data.
- Which direction to propagate the reference: from the slit to the camera or backwards from the camera to the slit.

The above points are essentially concerned with two major issues: the light path of the reference beam and the reference light source. A schematic diagram of a generic closed loop flexure compensation system is shown in fig. 2.8.

2.4.1 Light Path of the Reference Beam

The main consideration is that the reference should ideally be propagated through the *entire* spectrograph optical system, in order to compensate for motions of all surfaces. The light can be i) monochromatic (fixed wavelength e.g. laser, or programmable using a continuum lamp and monochromator), ii) emission line source,

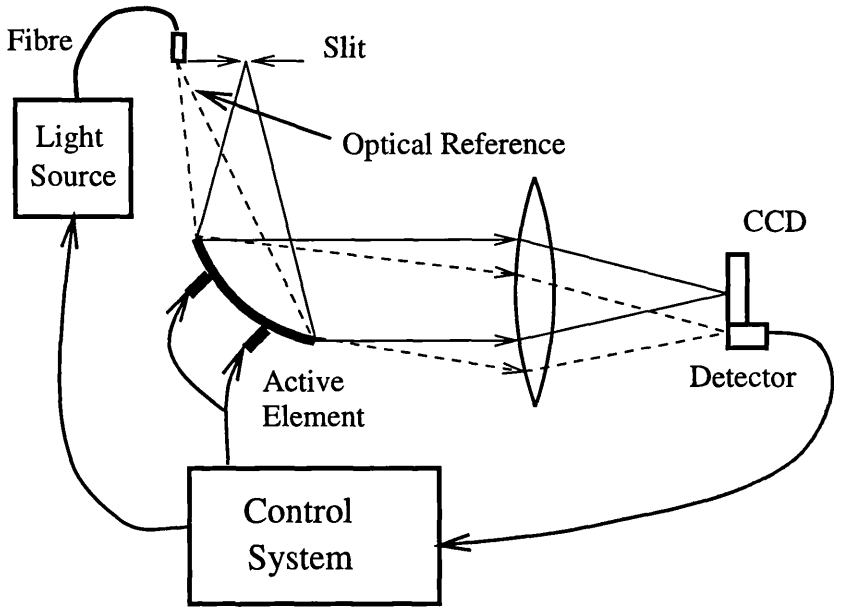


Figure 2.8: A closed loop flexure compensation system. In this case the optical reference is propagated through all the optical components. The injection is at the slit and detection near the CCD, but the system could also work backwards.

iii) continuum, or iv) continuum with emission lines superimposed from e.g. a Fabry-Perot etalon.

The reference beam can be propagated through the spectrograph as a narrow pencil of light (e.g. laser). Alternatively it can be a diverging cone to fill or partially fill the optics. In principle both methods should work. A laser beam through an optical system with a central obstruction would need to be propagated off-axis. Ghosting might be a problem since the optics are illuminated at different local angles than the science beam. More importantly, for an off-axis propagation, a focus shift is indistinguishable from an image shift. Also, the spot to be sensed will be $\sim mm$ rather than $\sim \mu m$ diameter, which impacts on the size of detector required. Overall, the filled-beam more closely emulates the science beam and therefore in principle is considered preferable.

The reference can be propagated forwards, with injection at the slit and detection in the camera focal plane, or backwards, with the reverse. The latter has an advantage in that the usual case of $F_{cam} < F_{col}$ conveniently magnifies the motion to be sensed.

The reference might conveniently be injected into the optical system as follows:

1. in the slit-plane (e.g. via a fibre), and sensed on a small CCD or quadrant photodiode adjacent to the science CCD.

2. at one end of the slit, and sensed on the science CCD. In this case the reference may be projected onto the $\sim$ 10 rows adjacent to the readout register. The procedure, which is repeated during the exposure is as follows:

   - close shutter
   - clock out 10 rows encompassing the reference
   - reverse-vertical-clock by 10 rows to restore science image to original pixels
   - centroid or cross-correlate reference and update active element
   - open shutter and continue integration

   This method is analogous to the photographic spectroscopic procedure of exposing an arc either side of the science spectrum, the difference being that we propose to detect the reference repeatedly *during* the exposure.

3. at the science CCD. In this case, the light source would be fibre fed, with the fibre output in the camera focal plane, but pointing towards the collimator and suitably baffled. The reference would propagate through the spectrograph backwards, being sensed on a small CCD or quadrant photodiode in the vicinity of the slit. In the case where it is impractical to shutter the CCD separately from the fibre source, care would need to be taken to analyse the ghosting of the optics for the *reverse* path in order to avoid contamination of the science image.

2.4.2 Reference Light Source

The spectrograph dispersing element(s) will deflect the reference beam depending on wavelength. If the reference were monochromatic, and the spectrograph configured for a different science wavelength (e.g. by rotating the grating), then the reference light may not reach a separate reference detector. An interesting – and extreme – possibility would be to choose an infrared wavelength for a visible spectrograph, to avoid stray-light contamination of the science data. In general, the problem of

grating deviation could be overcome by mounting a separate rotatable mirror on the grating cell, from which the reference would be reflected. However, image-motion due e.g. to grating mounting defects would not be compensated, and the mirror rotation might introduce its own problems. For very low resolution spectrographs, it might be possible to move the reference detector laterally, but thereby losing its fixed geometry with respect to the slit or science detector. These factors render the simple laser reference unattractive in most cases. For a prism cross-dispersed instrument such as Gemini HROS, the configuring for different wavelengths requires scanning the prism angles. The resulting change in prismatic deviation of the reference is even more difficult accommodate.

The laser can be replaced with an optical parallel-light projector, fed from a continuum source and tuneable monochromator. The monochromator could then be tuned according to the spectrograph setting to ensure that the reference beam always hits the detector. If the light from the monochromator is fibre-fed to the focal plane of the spectrograph, the reference could fill the optics and still be able to reach the detector for every grating and prism setup. In this case it is important that the wavelength output of the monochromator is stable.

An emission line ("arc") source (or mixture) can be chosen to provide emission lines on the reference detector whatever the spectrograph configuration. A continuum lamp and etalon provides convenient equally-spaced absorption features, but at the expense of filling the spectrograph with unnecessary (and stray-light inducing) continuum light. Hence the arc is in general the preferred source.

A special case is the method described earlier of using the science CCD with reverse-clocking as the reference detector. The arc would be fibre-fed at one end of the slit, giving a beam matching that of the telescope. With a single-dispersed spectrograph, the reference spectrum could be two or three pixels wide, contain many lines, and span the full width of the CCD. Each line could be very faint, with the precision of determining image motion being achieved by cross-correlation. However, with a cross dispersed instrument such as HROS this will not work, since arc spectra would bracket *every order*, reducing the (already pressed) inter-order spacing available for the science data.

2.5 Impact on Astronomical Observations

I will conclude this chapter discussing the advantages of a reduction in the spectrograph flexure from an astronomer's perspective. The impact of an active system for compensation of flexure on astronomical observations is more difficult to quantify than the impact on spectrograph design. For an astronomer, the relevance of active flexure compensation depends on how effective the system is and on the type of observation he is making. To estimate the benefits of active compensation, we need to start from what flexure means to an astronomer.

2.5.1 Flexure, Resolution and Spectral Calibration

The most important consequences of the spectrum motions on the detectors produced by flexure in the spectrograph can be summarized as follows:

- Reduction in spectral and spatial resolution.
- Spectral wavelength calibration errors.
- Errors in the measurement of line profiles.

In section 2.1 I described how flexure affects the instrument profile, in terms of shape, width and position. I demonstrated that for a perfect spectrograph, without aberrations, the full width at half maximum of the profile does not change as long as the flexure shift is less than half the projected slit width. In this first stage, flexure simply transfers photons from the peak to the wings of the profile. When aberrations are taken into account, the effect is similar, but the FWHM of the profile starts increasing for smaller flexure values. Figure 2.9 shows the effect of flexure on the slit profile when $20\mu m$ of aberrations are also present. The increase in the width of the profile is evident in the graph. For the astronomer, the change in the instrument profile translates into a change in resolution. In a spectrograph, the effective spectral resolution $\mathcal{R}_{eff}$ is defined in terms of the projected slit width on the detector [57]. As calculated in equation 1.38 (where the projected slit width was fixed $\Delta\omega = 2p$), the product of resolution and projected slit width on the detector

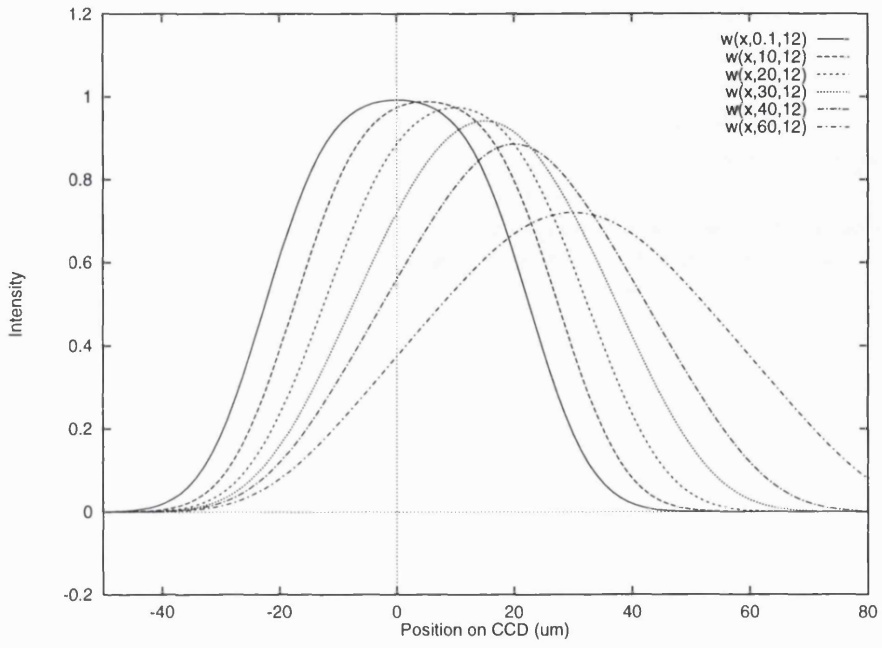


Figure 2.9: Effect of flexure on a projected slit width of $45\mu m$ in the presence of aberrations having a Gaussian distribution with FWHM $\Delta a = 20\mu m$. The graph shows the instrument profile as it would be recorded on the detector for different values of flexure Φ. The solid curve is the profile in absence of flexure, while the lowest curve corresponds to a flexure shift of $30\mu m$.

is approximately constant:

$$\Delta\omega\mathcal{R}_{eff} = F_{cam}\frac{\sin\vartheta_i + \sin\vartheta_r}{\cos\vartheta_r} \tag{2.8}$$

Since flexure increases the projected slit width, it also reduces the instrument resolution. Figure 2.10 shows the effect of flexure on resolution in the presence of different levels of aberrations. The figure shows that the effect of flexure on resolution becomes more relevant when aberrations are higher. In a real spectrograph, the aberrations are likely to be in the range between 20 and $30\mu m$ FWHM. Therefore, if we take a 5% reduction in resolution as a tolerable level, flexure shifts up to $10 - 15\mu m$ could be accepted. However, figure 2.10 also shows that the reduction in resolution is negligible when flexure is less than $\sim 5\mu m$. Therefore, even though the reduction in spectral resolution due to flexure is something the astronomers have learnt to live with, an active compensation system could eliminate the problem completely, provided it can limit image motion to less than $\sim 5\mu m$. In the case of

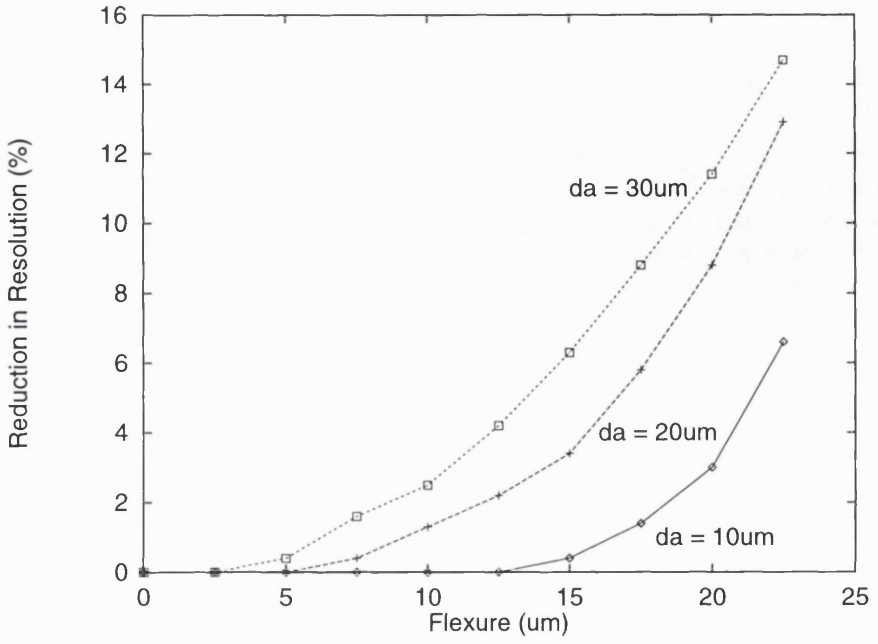


Figure 2.10: Percentage reduction in resolution due to flexure for different levels of aberrations. The projected slit width on the detector is $45\mu m$, the FWHM of the aberration function is da = 10, 20 and $30\mu m$.

ISIS flexure is usually in the $10 - 15\mu m$ range, so an effective flexure compensation system could increase the spectrograph resolution by 5%. In absence of flexure compensation, to achieve the same resolution the astronomer would have to use a slit 5% narrower, reducing the throughput by a similar amount. Therefore, to obtain the same signal to noise ratio, the exposure would need to be roughly 5% longer. A 5% saving of telescope time may seem a small improvement, but it is significant after one year of astronomical observations.

The main effect of flexure, however, is not the reduction in resolution, but the loss of calibration reference due to the shift of the spectrum on the detector. During a typical astronomical observation, the astronomer takes an arc spectrum before the scientific exposure to calibrate the scientific data and often one afterwards. If the spectrograph suffers from flexure, the observed spectrum will be shifted from the calibration arc spectrum, and wavelength calibration errors will be introduced. Some of these errors can be corrected during the data reduction process, but a residual calibration error always remains. For this reason the flexure rate and the

calibration accuracy required determine how frequently calibration spectra need to be taken. By reducing the flexure rate, active flexure compensation can affect both the length of the exposure and the frequency of calibration arc spectra.

Another important effect of flexure is the fact it changes the shape of the instrument profile. When observing narrow spectral lines, whose spectral width is comparable with the width of the instrument profile, the knowledge of this profile is very important. In these cases it is difficult to calculate the true profile of the observed line, because, as seen in equation 1.7, this is convoluted with the instrument profile. Normally the solution is to measure the instrument profile from the calibration spectrum and then deconvolve it from the astronomical data. In presence of flexure however, the instrument profile will depend on the exposure length, as fig. 2.9 clearly shows. This means the effective instrument profile for a long exposure will differ significantly from the "instantaneous" profile recorded during the short calibration exposure. If the latter is used in the deconvolution, errors will be introduced in the reduction of the astronomical data.

2.5.2 Flexure Compensation and Exposure Length

In a large Cassegrain spectrograph like ISIS, a flexure rate of 10 to $20\mu m/hr$ can be expected. When accurate spectral calibration is needed, it is recommended the exposure time to be limited to roughly $30min$ [58]. An active flexure compensation system can potentially reduce the flexure shifts to 1 or $2\mu m$ when tracking the same object. There is therefore a great potential for increasing the exposure time of an astronomical observation. I will examine the advantages and disadvantages of such a process.

- Advantages:
  - Many astronomical observations, particularly of faint objects, are sums of the equivalent of several hours (up to 10 hours or more) of exposure time, split into many short exposures. The same observation could then be split into fewer exposures, reducing the read-out noise from the detector.
  - With the same type of long observations, the time spent in reading-out the CCD would also be reduced, because the data would be read less

often. This is more important with large format CCDs proposed for the new generation of instruments, where the read-out time can be a few minutes.

- Some observations, like radial velocities measurements of a pulsating star, require accurate absolute wavelength calibration. This type of observation would benefit directly from the increased stability and indirectly, because longer exposures would give a better signal to noise ratio. The temporal scale of the event may, however, limit the exposure.

- Disadvantages:

  - Cosmic rays hitting the CCD can destroy information recorded in some pixels. The longer the exposure, the higher the number of pixels affected. Some algorithms exist to remove cosmic rays from the spectra and they are more effective with broad lines. With narrow, unresolved lines there is a danger of losing important data.

  - Long exposures have also an increased risk in case of system failure. If for some reason the telescope or spectrograph control system breaks down, the loss (in terms of telescope time) is greater with a long exposure.

  - In some observations there is a need to keep the slit vertical (pointing at zenith) and the turntable fixed. This is done to keep the slit parallel to the atmospheric dispersion, so that all the wavelengths enter the spectrograph. In these cases the telescope field of view rotates and exposures are usually kept short, to prevent the field rotation becoming too large.

It is difficult to estimate the point at which the advantages in lengthening the exposure are overtaken by the disadvantages. This decision will eventually rest on the astronomer's judgment, but it may prove to be a valuable instrumental flexibility. At the moment very few people take exposures longer than one hour and most astronomers take 30 minutes as a limit. With an effective flexure compensation system, this limit may be pushed to more than 1 hour. The limit is usually driven by cosmic rays hitting the detector, so there would be an incentive to develop detectors less sensitive to cosmic rays. We must be aware that not all astronomical observations would benefit in the same way. For some observations, like radial

velocities measurements of a close binary star, the exposure time can be limited by the temporal scale of the object. Nonetheless, a longer exposure time may be an advantage in some applications.

2.5.3 Flexure Compensation and Telescope Time

The greatest impact of active flexure compensation will be in saving telescope time. I have already explained how the reduction in spectral resolution can be translated into loss of observing time. Even more important is the loss due to frequent wavelength calibration exposures. At present the required frequent wavelength calibration exposures take a considerable amount of telescope time during a standard observing run. I have examined the observing logs for a typical night, when using ISIS at the William Herschel Telescope [59]. On that occasion the astronomers tracked the same object for 6 hrs 47 min, splitting the observation into one hour exposures. The total exposed time was 5 hrs 50 min and the wavelength calibration spectra had taken 37 min, that is 10.6% of the total time. This is a high percentage: when this time is added up at the end of one year, it amounts to several observing nights lost. With active compensation, particularly when tracking the same object, only one calibration spectrum should be necessary, taking an average 5 min. The percentage of time lost in during calibration in this type of observation would be reduced to only 1.4%, more than a factor 7 improvement. This nearly 10% increase in efficiency, combined with an increased resolution (shorter exposures for the same S/N), is likely to be the most important benefit from active flexure compensation. Overall, providing a stable instrument profile over several hours of telescope tracking is of great importance for the astronomer. The improvement in the astronomical data and in the efficiency of the observation would vary from case to case, but it would be a benefit for all.

Chapter 3

Actuators for Active Mirrors

3.1 Introduction

A review of the possible actuators and sensors for active mirrors was a necessary basis for the choice of the actuation system for ISAAC. In this chapter I will mainly analyze properties and characteristics of piezoelectric positioners (PZT), since they usually have many advantages on other actuators: high frequency response, no power dissipation in static conditions, little noise generation, compact size, strong pushing forces, very high accuracy when servo controlled and low cost. Moreover most of the characteristics and parameters of PZTs, can offer a basis of discussion for other actuators.

The theory of piezoelectricity is briefly reviewed, so that the physics behind all the technical parameters can be easily understood. Since electrostriction and magnetostriction are closely related to piezoelectricity, they will be briefly mentioned, too. All the most important parameters of PZTs are then reviewed for application in the Active Collimator project. An analysis of commercially available actuators is then carried out to find the products that match the scientific requirements.

3.2 Theory of Piezoelectricity

In physics of materials, all the effects in which electric energy is produced by conversion of another form of energy (mechanical, thermal) are closely correlated one

to each other and assume different names depending on the type of energy involved. Moreover for thermodynamical reasons, the converse effect is always possible (that is conversion of electrical energy to another form). Piezoelectricity refers to conversion of mechanical (elastic) energy into electric and vice versa.

3.2.1 Dipoles and Domains

The definition of piezoelectricity implies that in the crystal a polarization arises when a stress is applied or that application of an electric field produces strain in it. Polarization, electric field, stress and strain are the macroscopic variables that define the state of the material (together with temperature and entropy). From a microscopic point of view polarization is produced by electric dipoles in the primitive cell of the crystal. With the exception of quartz and few other materials, all the piezoelectrics are ferroelectrics, that is the microscopic electric dipole in the cell is present even in absence of external stresses or electric fields [61, 62] (more accurately ferroelectricity refers to the presence of a permanent dipole that can be reversed by an external field). The zero polarization arises only because of random distribution of those dipoles. Even if quartz has some important applications, all the ceramics used for micro-positioners are of ferroelectric type and we can thus limit ourselves to discussion of ferroelectricity (the piezoelectric effect is much smaller in non-ferroelectric materials). The most commonly used piezoceramic is the lead zirconate titanate $Pb(Zr, Ti)O_3$, from which the PZT denomination originally derived.

Above a critical temperature, called the Curie temperature ϑ_C, the material has a symmetric cubic structure and no microscopic dipoles are present [63]. Lowering the temperature induces a phase transition to a non centre-symmetric structure where microscopic dipoles in the unit cell appear. Immediately below that temperature the dipoles retain some mobility and they naturally dispose themselves in domains in which they are all in the same direction. Crystal axes give the natural orientation for the dipoles, leaving freedom for the direction (+ or −). In the material there are many of these domains and since they will tend to minimize the energy, they will orientate so that average polarization is still almost zero [64]. Further decreasing the temperature, the freedom of motion rapidly decreases too, and the dipoles remain almost fixed in their positions requiring stronger and stronger fields

for re-orientation.

In ceramic materials no preferential direction exists and a process called *poling* is necessary to produce a significant piezoelectric effect in the material. In that process a permanent polarization P_o is "frozen" in the material by applying an electric field above ϑ_C and lowering temperature below that point. This procedure is possible also in crystalline piezoelectrics. In this case however, crystal defects limit the degree of polarization achievable and it is found that in many of them, when the electric field is switched off, the domains tend to return in their original configuration. Even in the ceramics, a complete orientation of dipoles in a single domain is limited by defects and cannot exceed $80 - 90\%$ [65]. The above procedure is used in piezoelectric actuators to obtain large extensions. A strong field is applied until saturation is reached. A polarization $P = P_s$ is thus frozen in the ceramic. The piezoelectric ceramic then resembles a single crystal with a net macroscopic dipole [66].

When almost all the dipoles are in the same direction, and no external field is applied, surface charges will be present to screen polarization. If we apply an external stress in the same direction of polarization, this will change distances between ions in the cell and, proportionally, the magnitude of microscopic dipoles. Polarization will thus change and correspondingly an electric field will appear. It is found that the electric field scales linearly with the applied stress as would be expected from a single macroscopic dipole. In the same way generation of strain by an external electric field can be explained: the external field displaces the ions, producing strain in the material. Below Curie temperature, piezoelectrics exhibit a very high, temperature dependent susceptibility η [65]:

$$\eta = \frac{C}{\Theta - \vartheta_C} \tag{3.1}$$

This means that high polarizations can be obtained with relatively small electric fields ($\mathbf{P} = \eta \mathbf{E}$). As a consequence the piezoelectric constant is high and relatively small electric fields can generate high strains. Moreover, this also implies that high useful susceptibility (and consequently large piezoelectric effect) will persist for a wide range of temperatures.

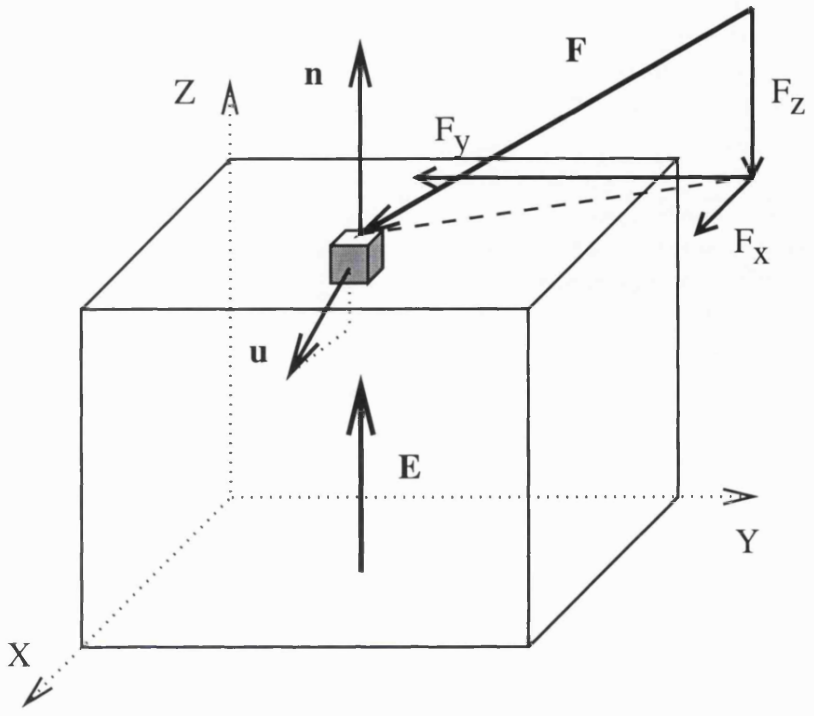


Figure 3.1: A layer of piezoelectric ceramic with applied electric field (along the z-axis) and a generic external force. The displacement of the material (**u**) introduces strain in the ceramic.

3.2.2 Constitutive Equations

Relations connecting the main macroscopic variables describing the system are called constitutive equations. In this treatment [67] hysteresis is not considered, and will be treated later. The typical geometry for a layer of piezoelectric ceramic with applied electric field and forces is shown in fig. 3.1. The equations derived below refer to the figure, but allow for any orientation of the field with respect to the layer surface. First of all it is necessary to define the stress and strain tensors. The Stress tensor T_{ij} is a measure of the force per unit area acting across a surface at a point; if **F** is the force and **n** the normal, then:

$$F_j = n_i T_{ij} \tag{3.2}$$

Stress has the dimension of $[N/m^2]$. The Strain tensor S_{ij} is defined in terms of the infinitesimal displacement **u** of a material point as follows:

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{3.3}$$

The strain tensor is adimensional, usually expressed in parts per million. The constitutive relations, using T and E as independent variables, can be derived from the Gibbs free energy of the system (here σ is the entropy) [68]:

$$G = U - \Theta\sigma - T_{ij}S_{ij} - E_k D_k \tag{3.4}$$

From the form of free energy we can derive the interactions between elastic and electric effects by differentiation. It is necessary to consider adiabatic or isothermal conditions, but these are good approximations since the thermal coupling is small. We obtain the following relations:

$$S_{ij} = s_{ijkl}T_{kl} + d_{kij}E_k \tag{3.5}$$

$$D_i = d_{ikl}T_{kl} + \varepsilon_{ik}E_k \tag{3.6}$$

Where:

- s_{ijkl} is the elastic compliance tensor
- d_{ijk} is the piezoelectric tensor
- ε_{ij} is the dielectric tensor

Thermal effects on strain are not small, however, and if the temperature is changed they will sensibly affect the strain. It is possible to consider this by adding a thermal term to equation 3.5:

$$S_{ij}^{(\Theta)} = \alpha_{ij}\Delta\Theta \tag{3.7}$$

where α_{ij} is the thermal expansion tensor. The above equations are valid in static conditions; they can however be applied also in quasi-static operation since s, d and ε have a small dependence on E and T. The fact that the dielectric tensor is a function of stress and particularly of electric field has to be considered, though. This is the one of the causes of non-linearity in the $S(E)$ curve (together with electrostriction) and it is a way of including hysteresis effects. Typical values of the most important piezoelectric constants are [69, 70, 93]:

$$d_{zzz} = 2 \text{ to } 6 \cdot 10^{-10} mV^{-1}$$

$$d_{zxx} = -1 \text{ to } -3 \cdot 10^{-10} mV^{-1}$$

$$d_{zzx} = 5 \cdot 10^{-10} mV^{-1}$$

Low-voltage PZTs exhibit the higher values of these constants. Typical values of the most important compliance constants are:

$$s_{zzzz} = 1.5 \text{ to } 2.0 \cdot 10^{-11} m^2 N^{-1}$$

$$s_{zzxx} = 1.2 \text{ to } 1.5 \cdot 10^{-11} m^2 N^{-1}$$

$$s_{zxzx} = 4.5 \cdot 10^{-11} m^2 N^{-1}$$

Again the higher values are for Low-voltage PZTs. The dielectric tensor has the following typical values [70]:

$$\varepsilon_{zz} = 1500 \text{ to } 3500$$

$$\varepsilon_{xx} = 1500 \text{ to } 3200$$

Typical value of the linear thermal expansion coefficient is:

$$\alpha = 1 \text{ to } 8 \cdot 10^{-6} K^{-1}$$

The thermal expansion coefficient of the piezoelectric material on its own is actually negative. The above values refer to a typical arrangement, where the actuator is constructed of many individual ceramic discs, stacked on top of each other (fig. 3.2) and α is determined by the properties of the ceramics and of the contact materials used. Another important parameter is the coupling factor k^2, that measures the fraction of electric energy converted into mechanical energy. It is a measure of the efficiency of the process and it is given by the ratio of mechanical and electrical internal energy ($k^2 = U_m/U_e$).

3.2.3 Hysteresis

Piezoelectrics, being ferroelectric, have significant hysteresis. The effect is similar to the one observed in ferromagnets. As explained above, application of an external field affects polarization by changing the size of microscopic dipoles. A residual amount of change, however, will be due to dipole re-orientation under the effect of the external field, and this can present hysteresis. Moreover, even if with the poling procedure a saturation polarization is frozen in the ceramic, when the poling field

is switched off, some depolarization occurs. When an external field is applied in the same direction of the permanent polarization, some re-orientation can still take place. Correspondingly the total polarization will have two contributions:

$$P_{tot} = \eta E + P_o \tag{3.8}$$

The subscript "o" indicates that this is the contribution due to orientation. The value of P_o will depend on the size and direction of the external field E and on the degree of poling and uniformity of the ceramic material. The process of further orientation presents the characteristic behaviour of hysteresis. When the field is increased, the dipoles begin to orientate parallel to it. If the field is then decreased the dipoles will tend to remain in their previous orientation since this is still parallel to the field. De-orientation will then occur because of thermal agitation of dipoles: the result is the typical hysteresis curve. As a consequence the permanent polarization is a non-linear function of the external field: $P_o = P_o(E)$. Hysteresis in polarization causes hysteresis in strain: a contribution due to dipole re-orientation will add to the linear relationship between strain and electric field. We can thus write:

$$S = dE + \mathcal{F}[P_o(E)] \tag{3.9}$$

Since the second term present hysteresis, an hysteresis curve is superimposed on the normal linear behaviour ($S = d \cdot E$).

3.3 Electrostriction

Electrostriction is related to inharmonic motion of crystal ions when an electric field is applied. The effect is quadratic in the field and therefore can exist even in an isotropic medium, with symmetry with respect to inversion. This implies that while reversing the field in a piezoelectric crystal will reverse also the strain, in an electrostrictive material the strain will be in the same direction. In our theoretical framework the electrostrictive constant is a fourth-rank tensor. Electrostriction is proportional to E^2 but it is usually masked by the much stronger piezoelectric effect. The materials that present the strongest electrostrictive effect are piezoelectrics above Curie temperature. The most commonly used is the lead magnesium niobate, $Pb(Mg, Nb)O_3$. It is possible to develop such materials with a relatively low Curie

temperature ($0°C$ or less), so that they can operate at room temperature. In the framework of constitutive equations, electrostriction adds a term in the Gibbs free energy of the form [68, 85]:

$$G^{(ES)} = -q_{ijkl}T_{ij}E_kE_l \tag{3.10}$$

The corresponding contributions to constitutive equations are:

$$S_{ij}^{(ES)} = q_{ijkl}E_kE_l \tag{3.11}$$

$$D_n^{(ES)} = 2q_{ijmn}T_{ij}E_m \tag{3.12}$$

These terms have to be added to the equations 3.5 and 3.6. When working above Curie temperature, the piezoelectric term will not be present and electrostriction will dominate.

3.4 Magnetostriction

The effect in which elastic strains are developed during magnetization of a ferromagnet is called magnetostriction. The effect, unlike piezoelectricity and electrostriction, is a consequence of rotation of magnetic dipoles towards the magnetic field direction. This effect has a complex dependence on the orientation of magnetization with respect to the crystalline axes and can be either quadratic or linear in magnetization [71]. Moreover the effect is essentially quadratic for small fields and reversing the field will not correspondingly reverse the displacement. However, in the simplest case, the linear magnetostrictive effect can be treated similarly to the piezoelectric effect. Constitutive equations analogue to the electric case can be written, with the magnetic induction B and the field H replacing electric displacement and electric field respectively. The equations will hold within the range of approximate linearity of the effect and they are not a true reflection of the physical processes behind them. We have [72, 63]:

$$S_{ij} = s_{ijkl}T_{kl} + \Lambda_{kij}H_k \tag{3.13}$$

$$B_i = \Lambda_{ikl}T_{kl} + \mu_{ij}H_j \tag{3.14}$$

Typically the the magnetostrictive constant of rare earths alloys (where magnetostriction is greatest) has a value of [73]:

$$\Lambda = 3 \cdot 10^{-5} Gauss^{-1}$$

It has to be noted that magnetostrictive effect is linear only in a short field range, where a jump in magnetization orientation produces very high strains. Further increasing the field, however will still increase the strain, due to further rotation of domains. In this latter region the relation is no longer linear [74]. As we already pointed out, in the very low field region, magnetostriction is almost absent and the effect is nonlinear, too.

3.5 Characteristics of Piezoelectric Actuators

3.5.1 Maximum Travel

The typical PZT element is a thin layer of material between two plane parallel electrodes at a distance L. In this simple geometry the elastic strain occurs mainly normal to the electrodes (z-axis) and in absence of stress can be approximated by:

$$S_{zz} = \frac{\Delta L}{L} = d_{zzz}\frac{V}{L} \tag{3.15}$$

We see that the expansion $\Delta L = d_{zzz}V$ is independent from the plate separation, so that to obtain high strains is necessary to add several of these stages one over the other. The expansion will be determined by the number of stages and the applied voltage ($\Delta L_{tot} = N\Delta L$) [76]. A standard stacked piezoelectric actuator is showed in fig. 3.2. This arrangement is typical of Low-voltage PZTs: with High-voltage types the very high electric field between electrodes usually does not allow this geometry. A cylindrical setup, with an electrode inside and the other on the surface is used in this case. The largest expansions however, are reached with a stacked assembly of High-voltage PZTs. In stacked High-voltage PZTs the layer has to be thicker to avoid dielectric breakdown and very good insulation of electrodes is also necessary. The strain values that can be achieved are typically:

$$S_{max} = 1.5 \cdot 10^{-3}$$

with an external field of about $E = 2 \cdot 10^6 V/m$ and a thickness of the layers ranging from $0.3mm$ to $1mm$. With these strain values, stacked actuators can normally expand by $\Delta L = 10 - 100\mu m$.

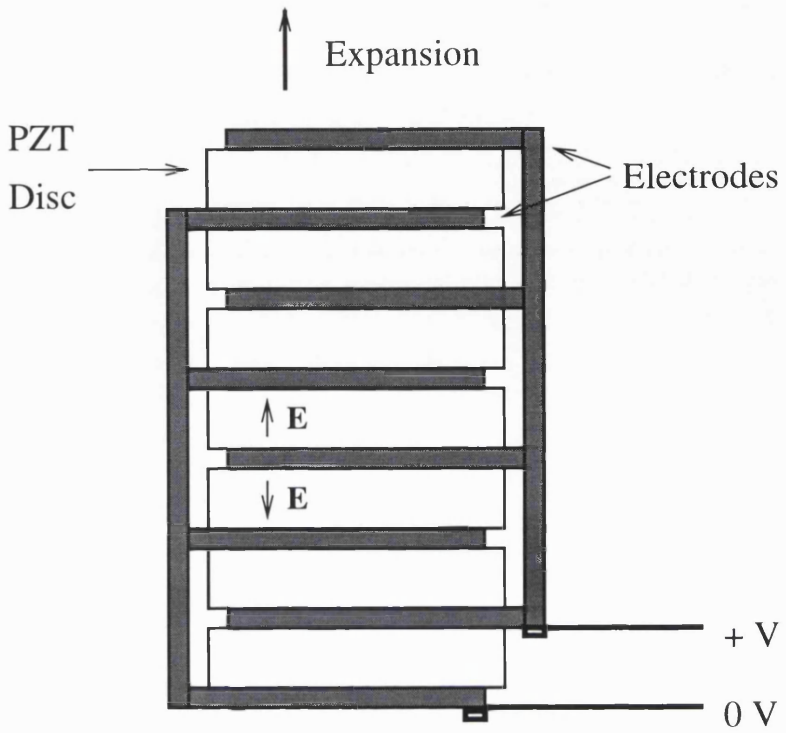


Figure 3.2: Stacked piezoelectric actuator. The thickness of the ceramic discs usually varies from 0.1 to $1mm$. The number of layers is much higher than the one shown in the figure. Voltages vary from tens to hundreds of Volts, expansions are usually in the range from 10 to $100\mu m$.

3.5.2 Pushing Forces

The range of pushing forces available is very wide, going from $50N$ up to $> 10000N$. The strongest forces can be obtained only with High-voltage PZTs (values greater than $3000N$). The amount of pushing forces achievable is related to the compliance tensor components s_{ijkl} of the material. The strain along the z-axis due to stress along the same axis, represented by a force F acting on a surface A, can be approximated by:

$$S_{zz} = \frac{\Delta L}{L} = s_{zzzz}\frac{F}{A} \tag{3.16}$$

The maximum pushing force, i.e. the blocking force, can be calculated from:

$$F_{max} = A\frac{S_{max}}{s_z} \tag{3.17}$$

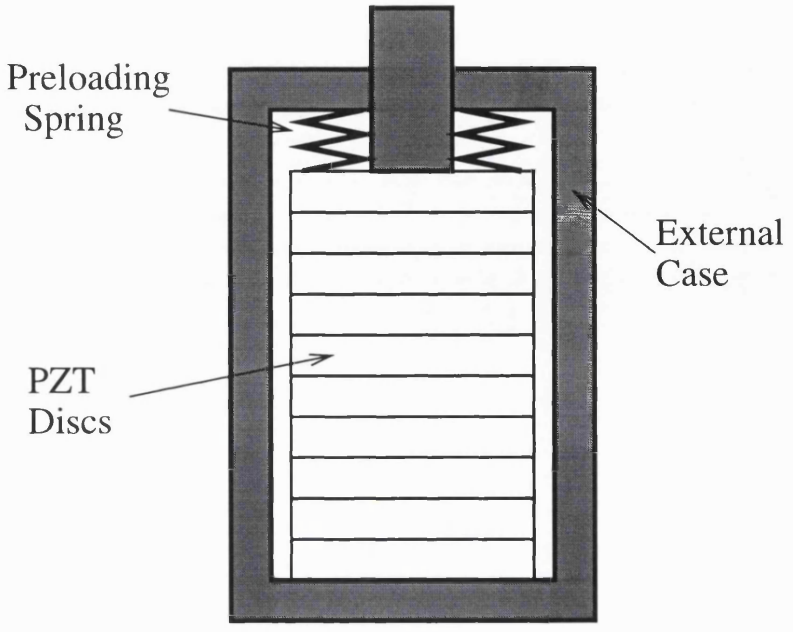


Figure 3.3: Preloaded piezoelectric actuator. The preloading spring provides the actuator pulling force. The actuator maximum expansion is however slightly reduced. The setup shown also prevents applying tensile forces directly on the PZT stack, which is very fragile.

From the previous equation it is evident that to have strong pushing forces we need low compliance and large cross sections A. A more practical form for calculating the effect of an external load is derived in the paragraph dedicated to stiffness.

3.5.3 Pulling Forces

Pulling forces are always smaller than pushing ones and only preloaded PZTs can develop such forces. This happens because the piezoelectric material is polarized in one direction and the electric field has to be in the same direction (a small negative field can be tolerated, but larger fields would depolarize the material). As a consequence, the actuator can only expand, and a contraction can be achieved only by providing a return force through an external spring preloading the material. The arrangement for a preloaded piezoelectric actuator is showed in fig. 3.3. The amount of the pulling force is a bit smaller than the preload (typically with a preload of $350N$ is possible to obtain a pulling force of $300N$). Preloaded PZTs

can usually produce pulling forces ranging from $20N$ up to $2000N$ in High-voltage PZTs. Preloading is also sometimes necessary because the compressive strength of piezoceramic material is more than one order of magnitude greater than its tensile strength. Therefore the preload protects the PZT from external tensile forces that could break the ceramic. The preload also affects other properties of the actuator, like maximum load and frequency response, because the PZT is actually pushing against an additional spring.

3.5.4 Shear Forces

PZTs are very fragile with respect to shear forces and torque moments: they should be stressed only axially. If a PZT has to bear shear forces, it is necessary to implement it with a mechanical mount on which those forces will act. No quantitative information to what extent PZTs can bear shear forces is available from manufacturers, although a value of $20N$ has been reported [77].

3.5.5 Accuracy

The accuracy of the PZT positioner depends on the type of sensor used to detect position. Some of the PZT manufacturers provide an optional or already included position sensor and give the corresponding accuracy. The intrinsic resolution in the expansion of the piezoelectric material is of the the order of few nanometers and it is never a limiting factor. The accuracy will be treated more extensively in the section dedicated to position sensors.

3.5.6 Hysteresis

Both High-voltage and Low-voltage PZTs present significant hysteresis, even if the problem is more serious in the latter. Hysteresis is typically 15% of the travel [87, 94], but it can be completely corrected in closed-loop operation. Drift, that is relaxation of position after a steep change in the applied voltage, is a form of hysteresis and can be corrected in the same way. Typically the PZT reaches only 93% of the final position $2sec$ after a voltage step is applied.

3.5.7 Linearity

PZTs have also a certain degree of non-linearity: this is typically 5% of the extension and can be corrected in a closed-loop system, too. This non-linearity has several causes. One is the fact that the piezoelectric tensor is not a constant, but a function of the applied field, and ultimately depends on the re-orientation process that is the cause of hysteresis. This will add terms with higher powers of the electric field. To a first approximation, we can consider only the quadratic term in the expression for piezoelectric strain. Since $S = dE$ we need to retain only the first term of the expansion of d. Electrostriction will produce a term of the same form in which the q coefficient will appear. Therefore, the non-linear term in the actuator expansion can then be approximated as:

$$\Delta L^{(NL)} = \left[q_{zzzz} + \left(\frac{\partial d}{\partial E} \right) \right] \frac{V^2}{L} \tag{3.18}$$

The actual linearity in positioning will depend on the position sensor when the actuator is used in a closed-loop setup in combination with a sensor. Normally this is the case in accurate positioning systems, because hysteresis and non-linearities are quite large in piezoelectric actuators.

3.5.8 Stiffness

The axial stiffness coefficient c_T of the PZT affects the behaviour of the actuator under a load, modifying the expansion according to:

$$\Delta L = \Delta L_o + F/c_T \tag{3.19}$$

Here ΔL_o is the expansion due to the piezoelectric effect. The stiffness coefficient is clearly related to the compliance tensor as approximated in equation 3.16 by:

$$c_T = \frac{A}{L s_z} \tag{3.20}$$

It is important to notice that the total expansion of the actuator does not change under a static load. The external force only shifts the zero point of the expansion range, as shown in equation 3.19 and in fig. 3.4. When a variable load, like a spring of elastic constant k_s, is applied, the total expansion will instead change according

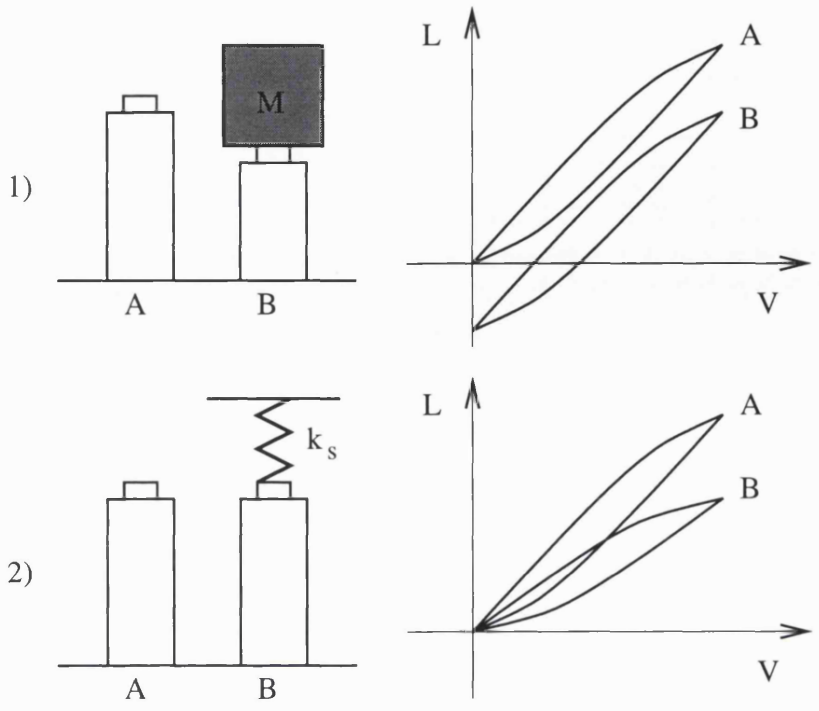


Figure 3.4: Effect of a load on a piezoelectric actuator. 1) When a static load is applied the maximum expansion of the actuator is not affected. The constant force $F = Mg$ simply shifts the zero point in the actuator expansion vs voltage curve. 2) When the actuator is loaded with a spring of constant k_s, the force depends on the expansion $F = k_s \Delta L$ and the maximum actuator travel is reduced.

to [75]:

$$\Delta L_o' = \Delta L_o \frac{c_T}{c_T + k_s} \tag{3.21}$$

Typically c_T ranges from $30N/\mu m$ to $400N/\mu m$ and it is proportional to the maximum pushing force the PZT can develop. These values will shift the zero point not more than $1\mu m$ for a $5Kg$ load. A higher stiffness will allow also a higher resonant frequency. The problem is that stiffness, according to equation 3.20, decreases with increasing the actuator length. Since the longer the travel, the longer the actuator length, long travel actuators will have relatively low stiffness and lower frequency response. We can also define a bending stiffness coefficient c_B [78], whose value for a cylindrical rod of diameter a is given by:

$$c_B = \frac{3\pi a^4}{64 L^3 s} \tag{3.22}$$

Because of the strong dependence on the diameter of the rod, which is relatively small, the bending stiffness coefficient is very small, producing large bending motions when a lateral force is applied. As with the axial stiffness, this problem is more serious with the long travel actuators.

3.5.9 Temperature Range

PZT positioners cannot be used at high temperatures because of the depoling of the ceramic material. High-voltage PZTs have a higher Curie temperature than Low-voltage ones: typical values are $\vartheta_C = 550K$ and $\vartheta_C = 420K$. PZTs can operate at very low temperatures (4 K), but, since the piezoelectric tensor d_{ijk} is temperature-dependent, the characteristics of PZTs can be very different. At $77K$ the expansion is almost 20% less than the value at room temperature. The different thermal coefficients of the materials in the assembly have to be considered when operating at very low temperatures, so manufacturers provide specialized PZTs for cryogenic operation. Otherwise the normal temperature range is $-40°C$ to $+80°C$ for High-voltage PZTs and $-20°C$ to $+50°C$ or $+80°C$ for Low-voltage PZTs.

3.5.10 Temperature Expansion

The value of the linear thermal expansion coefficient α was given earlier. The effect on expansion will be given by:

$$\Delta L = \alpha L \Delta\Theta \tag{3.23}$$

With a the total length of the PZT material typically between 2 and $15cm$, the effect of thermal expansion will be a displacement of 0.3 to $1.2\mu m/K$ for a High-voltage PZT and slightly less for a Low-voltage PZT. If not actively compensated, this would require an extremely accurate control of temperature. The compensation of temperature effects depends on the type of position sensor. The piezoelectric effect is also temperature-dependent, and this can affect the total expansion. The effect is only about 0.2% per K, producing a variation of the expansion of $\approx 1\mu m$ for a travel of $15\mu m$ and a temperature excursion of $40K$ [75].

3.5.11 Frequency Response

The frequency response of PZTs is limited by the resonant frequency and usually the positioner cannot be driven at frequencies higher than 80% of resonant frequency ν_o; therefore a value of $0.8\nu_o$ can be taken as an upper limit. The resonant frequency is given by:

$$\nu_o = \frac{1}{2\pi}\sqrt{\frac{c_T}{m_{eff}}} \tag{3.24}$$

Here $m_{eff} = m/2$, where m is the mass of the actuator. Low-voltage PZTs have a slightly higher frequency response. This can be seen considering equation 3.20: even if Low-voltage PZTs have a slightly higher compliance, the ratio L/A is much smaller because they are mounted in stacks of thin layers. The resulting stiffness coefficient is then higher and so is the resonant frequency. The upper limit usually ranges from $5KHz$ to $15KHz$ without a load. If a load M is added to the mass m of PZT, the frequency will change according to:

$$\tilde{\nu} = \nu\sqrt{\frac{m}{m+2M}} \tag{3.25}$$

The maximum frequency that is possible to obtain will depend also on the power supply, the limiting factor being the maximum output current of the supply. An approximate estimate is given by:

$$\nu_{max} = \frac{i_{max}}{2CV_o} \tag{3.26}$$

The limiting factor in the equation above is normally the large capacitance of piezoelectric actuators, which can push i_{max} to quite large values.

3.5.12 Input Voltage

As their name explains, the input voltage changes from High-voltage to Low-voltage PZTs. Usually Low-voltage PZTs are driven at a voltage of $0 - 100V$, whereas High-voltage PZTs work in the range $0 - 1000V$. Using a high voltage supply reduces jitter due to electrical noise but it is also more expensive. The piezoelectric effect is a function of the electric field and not of the voltage, therefore the attainable strains are almost the same for both types, since the smaller layer thickness in Low-voltage PZTs compensates the lower voltage.

Increasing the voltage will increase the field and correspondingly the strain. The upper limit for expansion is thus determined by the maximum field that will not damage the ceramic due to dielectric breakdown. It is also possible to apply negative voltages to increase the range of expansion: in this case the limit will come from depolarization. In fact, when a negative voltage is applied, the external field will be opposite to polarization, thus decreasing it through depoling. Further increasing the field in the negative direction will eventually exceed the coercive limit, after which the dipoles invert their direction and they have again the same orientation of the field. After this point the ceramic will start to expand again. Coercive fields for Low-voltage PZTs are typically [89]:

$$E_c = -0.6 \text{ to } -1.0 \cdot 10^6 V/m$$

While for High-voltage ones they are:

$$E_c = -1.4 \text{ to } -2.0 \cdot 10^6 V/m$$

The actuators, however, work far from the coercive limit and only small negative voltages can be applied. For most actuators, the maximum negative voltage we can apply is less than 20% of the voltage range.

3.5.13 Capacitance

The capacitance of a stacked actuator of N layers of thickness t with cross section A is given by:

$$C = \frac{N\varepsilon A}{t} \tag{3.27}$$

The capacitance of Low-voltage PZTs is usually one order of magnitude greater than High-voltage ones because of the thinner layers. Normal values are:

$$C = 0.5 \text{ to } 5\mu F \quad (\text{Low} - \text{voltage})$$

$$C = 20 \text{ to } 100nF \quad (\text{High} - \text{voltage})$$

This means Low-voltage PZTs will require more current than the High-voltage ones. Capacitance depends on extension because the dielectric constant will change with the electric field. The thickness change due to material deformation is not responsible for a significant capacitance change [87]. Values normally given are for zero

expansion (zero field), then the capacitance will increase almost linearly to reach a value that can be as much as two times greater at maximum expansion.

3.5.14 Power Dissipation

Since the resistance of a PZT is $R \simeq 10^{11}\Omega$, in a static configuration practically no power dissipation occurs. Only a small amount (5 to 8%) of the total energy stored $U = 1/2CV^2$ is dissipated during discharging. The effect is negligible in static operation but it must be considered when PZT is driven at a frequency ν. The dissipated power is:

$$P = 2\pi\nu CV^2 \tan\delta \tag{3.28}$$

The effect is a consequence of the hysteresis in polarization and can cause some heating of the PZT when driven at high frequencies. Since the dissipation factor or loss tangent $\tan\delta$, is 0.03 for Low-voltage and 0.005 for High-voltage PZTs, the dissipated power is almost the same: approximate values are $P \sim 4W$ for $\nu = 250Hz$.

3.5.15 Power Consumption

The electrical power required to operate at a frequency ν is given by:

$$P_{out} = CVV_o\nu \tag{3.29}$$

When operating in quasi-static conditions the power consumption is very low due to the high efficiency in converting electrical into mechanical energy. As I mentioned earlier, this is expressed by the coupling factor, which can be as high as $k^2 = 0.5$. A higher coupling factor will of course reduce the power consumption in high frequency operation. The power consumption will increase with the operating frequency and the capacitance, the latter being determined by the length and the diameter of the stack.

3.5.16 Lifetime

Aging in High-voltage PZTs due to depoling is virtually zero. Aging in Low-voltage ones can produce small effects in time periods over several days. This depoling

can be recovered though, by driving the actuator at the highest voltage ($100V$) for a short time. Due to this recovering procedure, lifetime of piezoelectric actuators is extremely long: $3.9 \cdot 10^9$ cycles without failures are reported [89]. Fatigue will depend also on the applied stress, but when an actuator has to exert only small forces this will not sensibly affect the performance [96]. The main source of failure in a PZT can be dielectric breakdown. For this reason another parameter that can affect lifetime is humidity. The electrodes of PZT stacks are moisture-sensitive and exposing them to relative humidity greater than 60-70% can cause damage when this is repeated over a long period of time. The use of long stacked actuators, with a great number of electrodes can increase the chances of failure.

3.5.17 Conclusions

It is possible to summarize all the previous characteristics in few parameters that can account for several effects. These are the parameters to look at when listing scientific requirements for the actuators:

1. Expansion ΔL. We can include in a single expression all the main effects that affect the total expansion of the piezoelectric stack:

$$\Delta L = N d_{zzz} V + N^2 \beta \frac{V^2}{L} + \frac{F}{c_T} + \alpha L \Delta\Theta \tag{3.30}$$

   Here β is a coefficient accounting for non-linear effects. Expansion is determined by the applied voltage, the load and the temperature, other effects are negligible. Most of these parameters do not affect the maximum travel, which is essentially determined by the applied voltage. A plot of the strain curve as a function of the applied voltage for a Low-voltage PZT is shown in fig. 3.5; the same curve for an electrostrictive actuator is also shown for comparison.

2. Push/Pull Forces. These parameters depend on the ceramic material, the mechanical mount and the value of preload and are normally provided by the manufacturer.

3. Absolute Accuracy. As already pointed out, this will depend on the position sensor used.

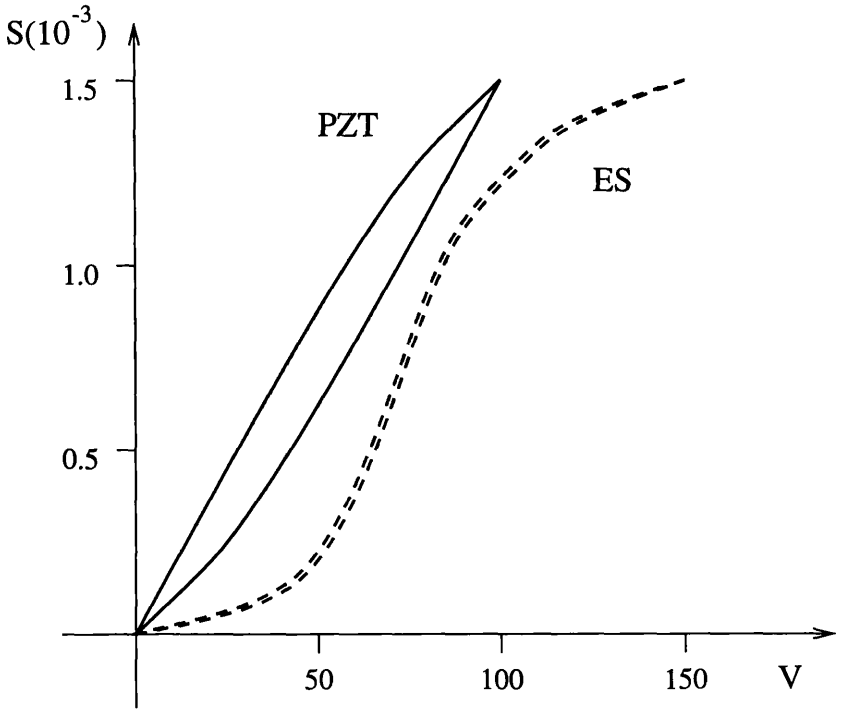


Figure 3.5: Strain vs Voltage curve for a Low-voltage PZT and an Electrostrictive (ES) actuator. The maximum strain attainable from both actuators is practically the same, but the ES actuator achieves that at higher voltages (higher electric fields). The ES actuator also shows much less hysteresis, barely detectable in the graph.

4. Temperature Range. Depends on the ceramic material and on the mechanical mount.

5. Upper Frequency. The upper frequency limit can be determined by the power supply. This is not a physical limit, however. The latter will approximately be given by:

$$\nu_{max} = \frac{0.8}{2\pi}\sqrt{\frac{2c_T}{m+2M}} \tag{3.31}$$

It will mainly depend on the stiffness coefficient and on the load mass.

6. Lifetime.

3.6 Electrostrictive Actuators

Electrostrictive actuators are based on a further electric-field deformation called electrostriction [79, 80, 81]. This effect is normally hidden by the larger piezoelectric effect, but above Curie temperature it can be observed separately. The electrostrictive materials are ferroelectrics in a disordered state: they have a Curie temperature below room temperature, normally around $270K$. Moreover they have a very high dielectric constant, usually an order of magnitude bigger than PZTs ($\varepsilon = 15000$) [87]. Most of the other characteristics of these materials are similar to the ones of piezoelectric ceramics. The main differences are:

1. Electrostrictive materials have a lower hysteresis (1 to 2%) and present no drift, resulting in better reproducibility. When operated in closed-loop control, this property produces no advantages. Moreover in electrostrictive materials hysteresis is strongly dependent on temperature and at the lowest temperatures it can be as much as 25% [84]. This implies that if the environmental temperature is not stable during operation, hysteresis can be a problem also in electrostrictive actuators.

2. Much higher capacitance for electrostrictive actuators (up to a factor 5), resulting in higher currents. This is a consequence of the combined effect of higher electric fields required by electrostriction, resulting in thinner layers (0.1 to $0.25mm$) and of the higher dielectric constant of these materials [88].

3. Temperature dependence of electrostriction may require a closed-loop feedback for high accuracy positioning. The temperature expansion is about $0.1\mu m/K$ in a narrow temperature range and then can be as high as $1\mu m/K$. The thermal expansion coefficient is instead smaller than PZTs an has a typical value of $\alpha = 1 \cdot 10^{-6} K^{-1}$.

4. Temperature range is usually very narrow for electrostrictive materials. Due to their low Curie temperature, their effectiveness is greatly reduced below $0°C$ [82, 87]. This happens because the maximum attainable strain decreases sharply with temperatures approaching the Curie temperature. Moreover, the strong dependence on temperature of electrostriction, limits the working range

on the high temperature side. The working temperature ranges usually from $+10°C$ to $+40°C$, although it is possible to shift the central temperature on the temperature scale. Some electrostrictive actuators are reported to work between $-10°C$ and $+70°C$ [88], but they cannot reach large extensions on such a wide temperature range. A plot of the maximum strain as a function of temperature for an electrostrictive material is shown in fig. 3.6. the curve is compared with the correspondent ones for piezoelectric and magnetostrictive materials.

5. Since electrostriction is quadratic in the field, there is no negative voltage displacement. Inverting the voltage produces an expansion in the same direction. This also implies smaller aging effects, because no depoling occurs in electrostrictive materials. The quadratic relationship, however, implies lower linearity in the actuator expansion.

3.7 Magnetostrictive Actuators

Magnetostrictive actuators are significantly different in design and performance from piezoelectric and electrostrictive stacks. The expanding element of a magnetostrictive positioner is a single rod of the active material. The common magnetostrictive material is Terfenol, an alloy of iron and rare earths ($Tb_{0.3}Dy_{0.7}Fe_{1.95}$). The magnetic field is provided by a solenoidal coil around the rod. Sometimes permanent magnets are also used to fix the zero position in the most effective region of the $S(H)$ curve (strain vs magnetic field curve, fig. 3.7). A schematic view of a magnetostrictive actuator is shown in fig. 3.8. The performance of these actuators must be separated from the one of PZTs, thus I will briefly review all the parameters I had already examined. The main differences came from the mechanical and electrical configuration and the different material properties.

- Maximum Travel. The travel will depend on the rod length: the maximum strains are up to $2.0 \cdot 10^{-3}$. This means that a travel of $300\mu m$ is achievable with a rod $20cm$ long. Due to the solid metallic core of these actuators this length is not excessive as for piezoelectric actuators. At the present time

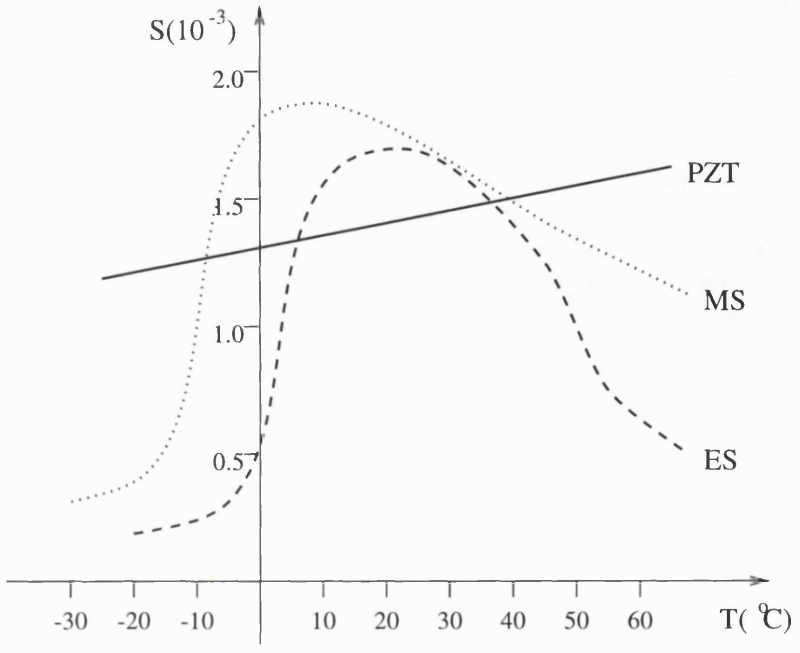


Figure 3.6: Maximum Strain vs Temperature for electrostrictive (ES), magnetostrictive (MS) and PZT materials. The curves have been derived from various sources and they are only indicative. The exact position and intensity of the maximum of each curve can be changed by manipulating the composition or the structure of the material. Piezoelectric materials show a fairly linear behaviour, while the largest strains can be obtained with magnetostriction.

actuators with travels up to $200\mu m$ are commercially available and rods with lengths up to $25cm$ are being developed. The actuator expansion will depend on the input current, rather than the voltage, and can be calculated as follows: for a thin solenoid of length L, with N coil windings and in which a current i is driven, the magnetic field is given by:

$$H = \frac{N^2 i}{L} \tag{3.32}$$

Consequently from equation 3.13 we obtain the expansion in absence of external stress:

$$\Delta L = \Lambda N^2 i \tag{3.33}$$

The maximum strain, is a function of the load and increases with it, thus, to obtain the largest strains, a preload of the actuator is necessary.

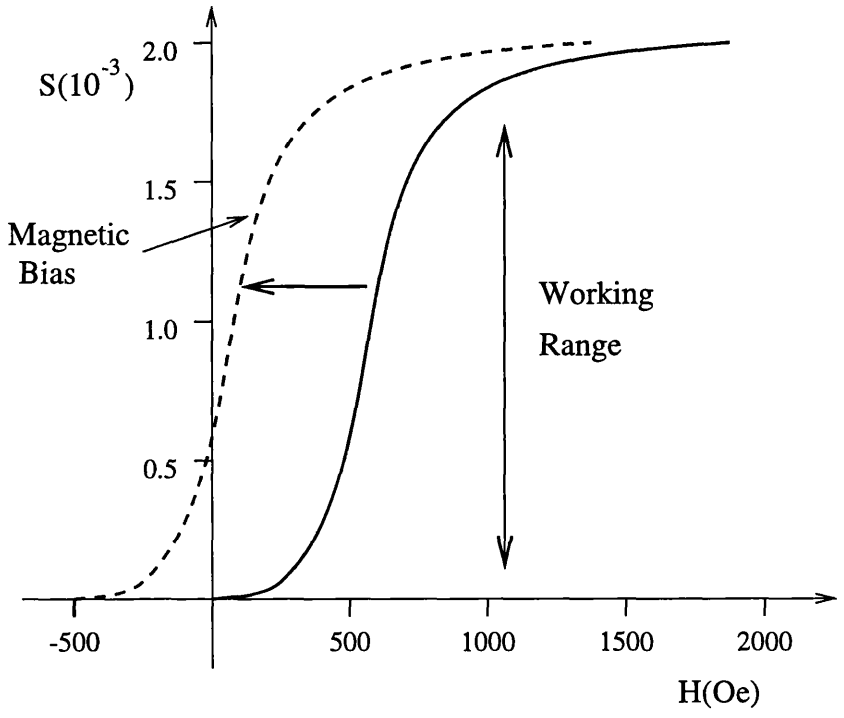


Figure 3.7: Strain vs magnetic field in a Terfenol rod. The working range is typically the steep linear section of the curve, where large strains can be obtained with a relatively small field. Often a permanent magnetic bias is introduced, to reduce the amount of current required. In this case the magnetic bias is $H_o = 500Oe$.

- Push/Pull Forces. This is not a critical parameter, the actuators can easily bear loads of hundreds of Newtons. Moreover, the attainable strain initially increases with increasing the load [83], so the actuator maximum expansion is not reduced by loading the actuator. Magnetostrictive actuators are also stiffer with respect to tensile and shear forces. Tensile and compressive strengths of Terfenol are $28MPa$ and $700MPa$ respectively.

- Hysteresis. Magnetostriction presents hysteresis due to its ferromagnetic nature: the usual value is 10% to 15% of the travel but it can be much larger if the full range of expansion of the actuator is used. Hysteresis is mainly a consequence of re-orientation of microscopic dipoles, as for piezoelectric materials. Here however, unlike piezoelectricity, the main contribution to the expansion comes from reorientation, making hysteresis a direct consequence of magnetostriction.

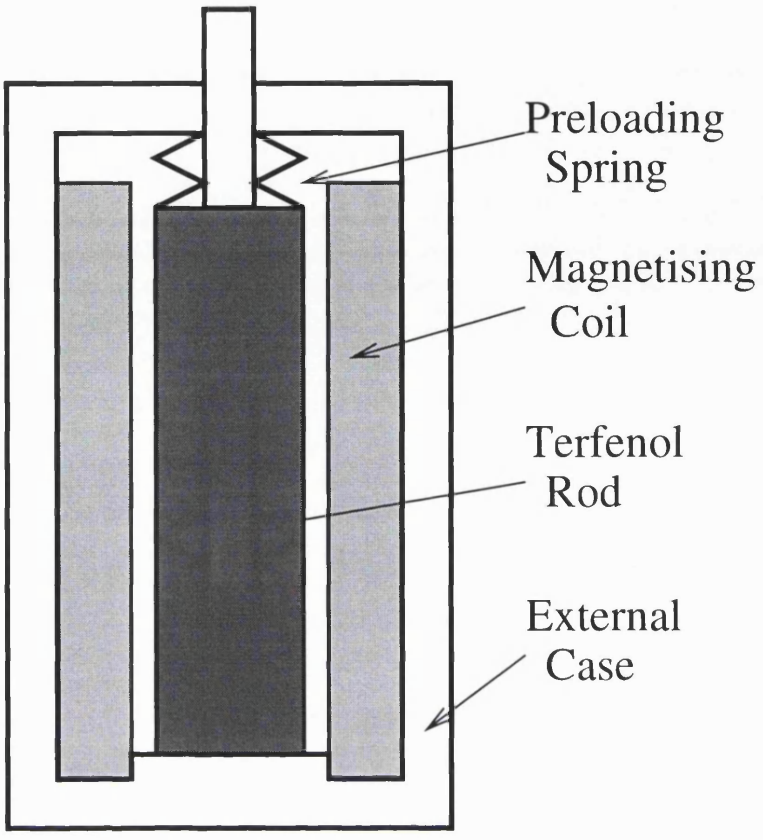


Figure 3.8: A magnetostrictive actuator. The preload is necessary to obtain pulling forces, since magnetostriction always produces a positive expansion. The coils around the Terfenol rod provide the driving magnetic field.

- Linearity. The effect is linear in the magnetic field only in a narrow range. In the strain vs. field curve we can separate three different regions [74]: a) At very low fields magnetostriction is almost zero. b) Above a critical field magnetization "jumps" towards the direction of the field and a giant magnetostriction is observed. In this region the Λ coefficient has its higher value and the relation is almost linear. c) At high fields magnetostriction is no longer linear in the field and approaches a saturation value. A plot of the magnetostrictive effect as a function of the magnetic field generated by the coil windings is shown in fig. 3.7. The curve shifting due to a permanent magnetic bias is also shown.

- Stiffness. The axial stiffness coefficient c_T of Terfenol usually ranges from 40 to $250 N/\mu m$ [83]. The value is comparable with that of piezoelectric actuators.

The Young's Modulus has a value of about $Y = 3 \cdot 10^{10} N/m^2$.

- Temperature Range. Curie temperatures of more than $650K$ allow operation over a wide range above room temperature. The lower limit is more critical: the magnetostrictive strain has a steep fall below $-10°C$ [74] therefore operation is usually limited to temperatures above $-15°C$. A plot of maximum strain versus temperature is shown in fig. 3.6. The temperature peak is dependent on the proportion of Tb and Dy in the alloy, therefore the temperature range is slightly tunable [86].

- Temperature Expansion. The linear thermal expansion coefficient of Terfenol is

$$\alpha = 1.2 \cdot 10^{-5} K^{-1}$$

  This means the expansion of a long rod of Terfenol is highly temperature sensitive and can be as much as $2.5\mu m/K$ for a rod $20cm$ long.

- Frequency Response. The frequency is upper limited by the resonant frequencies of the Terfenol rod and the driving solenoid, but generation of eddy currents in the magnetostrictive material is usually the limiting factor. However the maximum value achievable is still about $5KHz$ for the longest rods and these actuators can easily operate up to $2KHz$ [90]. For shorter actuators the frequency response is even higher and can be as much as $20KHz$.

- Input Current. Currents are quite high in these actuators: to reach the biggest strains currents up to $5A$ are needed. By incorporating permanent magnets though, the actuators can operate with currents of $500mA$. The corresponding fields are about $50Gauss$ ($40KAm^{-1}$).

- Power Dissipation. The coil resistance ranges from 5 to 10Ω, thus power dissipation in static conditions is not very high. In dynamic operation the dissipated power is higher, but it has been estimated to be less than $10W$ [73].

- Power Consumption. The magneto-mechanical coupling factor is $k^2 = 0.5$: this means 50% efficiency in converting magnetic energy. The estimated power consumption is similar to the dissipated power.

The main advantages of magnetostrictive actuators are large expansions, large pushing forces and high mechanical strength in combination with wide frequency bandwidth and wide temperature range. The main problems instead seem to be the ohmic heating of the solenoidal drive and the relatively high thermal expansion coefficient. The relatively large mass and dimensions of the actuator can produce further problems in the mechanical design of the actuation system when little space is available.

3.8 Leadscrew Actuators

The main feature of these actuators is a high precision leadscrew driven by either a DC motor (Newport PM500-A1) or a stepped motor (Melles Griot Nanomover). Several turns of the motor are reduced to a single turn of the leadscrew through gearboxes. Both these actuators operate in a closed-loop system the main purpose of which is to eliminate driving errors due to leadscrew pitch errors, backlash and friction. According to the given characteristics, they are able to reach an accuracy of $1.5\mu m$ on a travel of $300\mu m$. Leadscrew actuators can exert push/pull forces of about $100N$ and can bear shear forces of $50N$. One major drawback is that they have a very low frequency response. Commercial leadscrew actuators usually have a specified bandwidth of less than $8Hz$. Moreover they have relatively high power dissipation, even in static conditions when driven by a DC motor. Using a stepper motor will reduce power consumption only in static operation. Using the actuators at their upper frequency limit will worsen these thermal effects.

3.9 Voice Coil Actuators

Voice coil actuators utilize a permanent magnet field and a coil winding to produce a force proportional to the current applied to the coil. The electromechanical conversion mechanism is based on the Lorentz Force. For N conductors in series of length L, through which flows a current i in a magnetic field B, the corresponding force is given by [91]:

$$F = kBNLi \tag{3.34}$$

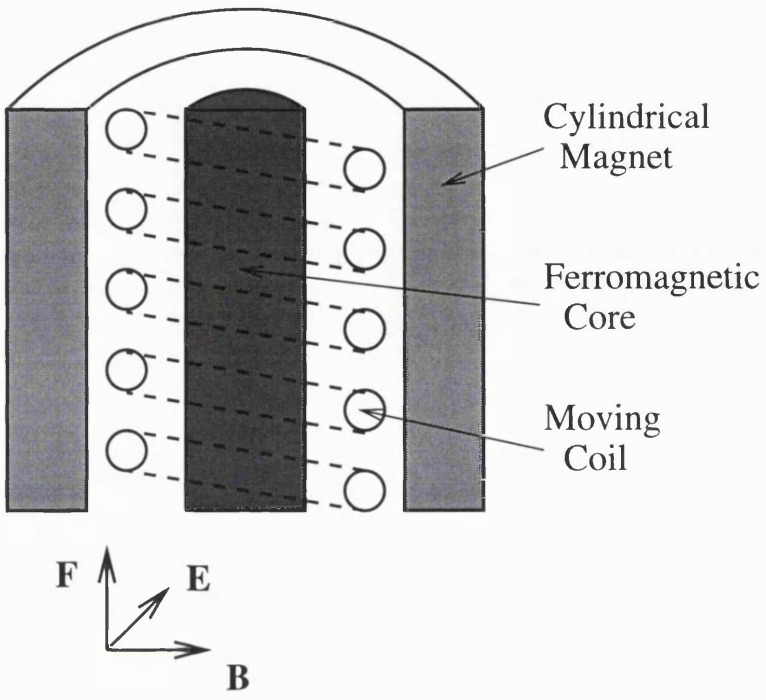


Figure 3.9: Working principle of a voice coil actuator. The cylindrical magnet produces a radial magnetic field, which interacts with the electric field in the coil to produce an axial force in the coil itself.

In the actuator, NL is the length of the coil and k is a proportionality constant. The direction is determined by the vector product of electric and magnetic fields, thus reversing the current will also reverse the force. A schematic view of a voice coil actuator is shown in fig. 3.9. In the usual configuration a cylindrical magnet with a ferromagnetic core is used to produce a radial magnetic field in the cylinder. The coil is placed between the magnet and the core, so that, when a current flows through the coil, a force is generated parallel to the coil axis. A rare-earth magnet can be used to increase the magnetic field up to a value of $11KGauss$ [91]. A parameter that characterizes the performance of these actuators is the Force Constant $K_F = F/i$: the stronger the force constant, the higher the forces that can be generated. The maximum force available is usually the limiting factor with voice coil actuators, however, since they are not based on an expanding crystal or ceramic, the maximum travel can be as much as $30mm$.

Voice coil actuators can produce forces up to $300N$, but they are more competitive

with other actuators when the force is less than $20N$. If the maximum force is kept small, these actuators have resonant frequencies far beyond 1 KHz and very low masses (few hundreds or even tens of grams). When a high force is needed instead, they start becoming large, massive and with a low resonant frequency. Power consumption is a function of the maximum force, too: the characteristic parameter is the motor constant K_m, which relates the two quantities through the following equation:

$$P = \left(\frac{F}{K_m}\right)^2 \tag{3.35}$$

Typically K_m ranges from 2 to $10NW^{-1/2}$ and power consumption for a device that can produce $40N$ is about $20W$.

3.10 Sensors

A micro-positioning system, besides actuators, includes also position sensors and a control system. Earlier I mentioned the importance of the position sensor in determining the accuracy with which the actuator can be moved. In this section I will review the performance and the technical characteristics of high accuracy position sensors. A schematic diagram of the most common position sensors is shown in fig 3.10

3.10.1 Strain Gauges

Strain gauges are commonly used in combination with actuators because they can be easily packaged inside the housing of the expanding element and they have good resolution, linearity and the possibility of compensation of thermal effects. Conceptually the basic principle of a strain gauge is very simple: when a conducting material is subjected to tensile stress it increases its length l and reduces in cross section A. Since the resistance of that material is given by (ρ is the resistivity of the material):

$$R = \frac{\rho l}{A} \tag{3.36}$$

The change in resistance can be used to detect the change in linear dimension (dR is proportional to dl). If the strain gauge is solidly connected to an actuator, for

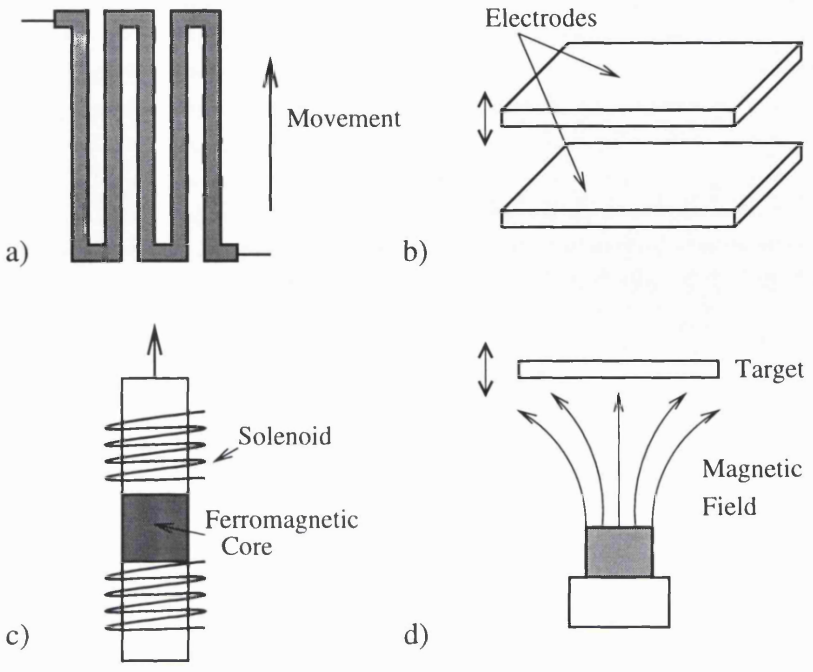


Figure 3.10: High accuracy position sensors: a) Strain gauge. The resistor strip is usually folded several times to increase the length l in the direction of movement. b) Capacitance sensor. The capacitor is sensitive to changes in the gap between electrodes. c) Linear Variable Differential Transformer (LVDT). The position of the ferromagnetic core with respect to the two solenoids can be calculated from the current change. d) Eddy current sensor. The magnetic field produced by the sensor is influenced by the position of a metallic target.

example, the measurement of the expansion of the actuator can be reduced to a measurement of the strain gauge resistance. The simplest way of doing this is to put the gauge in a Wheatstone bridge configuration, together with three normal ("dummy") resistors having the same value R of the gauge when the strain is zero (fig. 3.11). The output voltage of the bridge will then be proportional to the change in resistance of the gauge and therefore to the gauge expansion [92]. This configuration however, is highly temperature sensitive and the output of the bridge remains constant only if it is kept at constant temperature. The solution is to put four gauges in a "full" bridge configuration: one of the gauges is linked to the expanding element and gives the position measurement while the others compensate the temperature effects. Accuracy of position sensing is given as a percentage of the travel: strain

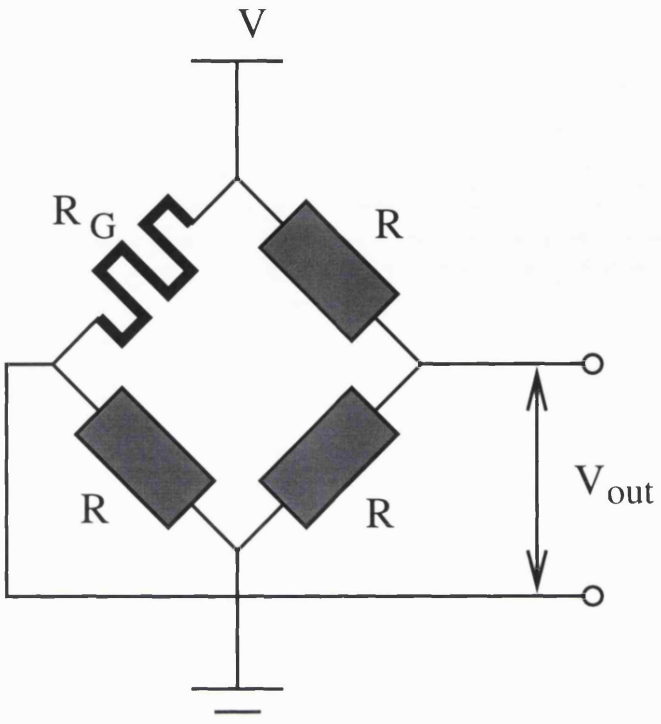


Figure 3.11: Wheatstone bridge configuration for measuring the resistance R_G of a strain gauge. The other resistors in the bridge have the same resistance R and they are chosen so that when the gauge is at zero expansion $R = R_G$ and the bridge is balanced ($V_{out} = 0$). When the gauge expands, the bridge becomes unbalanced and produces a voltage output.

gauges can give an accuracy of 0.2%.

3.10.2 Capacitance Sensors

Capacitance sensors measure the variations in a capacitor due to displacements of the electrodes. In a simple plane parallel capacitor, the capacitance is varied by changing either the plate separation or the position of an insulator between them. In a cylindrical capacitor this is obtained shifting the electrodes along the axis. In the simplest configuration, with two plane-parallel electrodes of area A and separation h, the capacitance is given by [92]:

$$C = k\varepsilon\frac{A}{h} \tag{3.37}$$

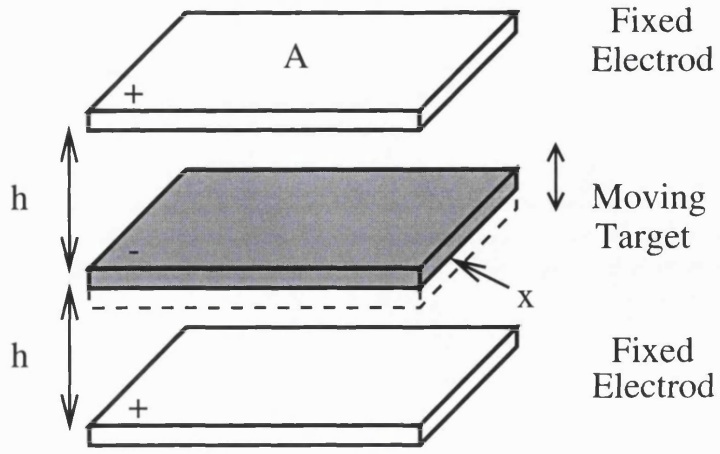


Figure 3.12: Two parallel plate capacitors in a differential configuration. The moving part of the sensor is chosen so that when it moves by a small distance x, it produces an output on both capacitors. The resulting differential signal can be measured more accurately than the change in capacitance in an individual sensor.

In the equation ε is the dielectric constant of the medium between electrodes and k is a proportionality constant. To achieve better performance, the position sensor is often composed of two capacitors in a differential configuration like the one shown in fig. 3.12. As for the strain gauges, a position measurement is converted into a voltage measurement. This is done by arranging the sensors into a capacitance bridge in which the differential change in capacitance unbalances the bridge and produces an AC output voltage. If the bridge is driven by a sinusoidal voltage of amplitude V_o and the target electrod moves by a distance x, the corresponding output will be [92]:

$$V = \frac{1}{2}\left(\frac{x}{h}\right) V_o \tag{3.38}$$

Where x is the motion to be measured. The accuracy usually achievable with these sensors is 0.02% of the travel, which is an order of magnitude better than strain gauges. Thermal effects in capacitance sensors are small, but not negligible: indeed this may be the main source of error if the temperature is not controlled. The thermal drift of Queensgate capacitance sensors is [97]:

$$\alpha_C = 7 \cdot 10^{-5} K^{-1}$$

These position sensors have the advantage of presenting no hysteresis and a fairly high frequency bandwidth: the upper frequency limit is $5KHz$. Moreover they are

made of two separate parts, so that one can be fixed to the moving object and the other to the reference position. Tilting of the plates has to be avoided, though, and a mechanical mount to keep them parallel is necessary.

3.10.3 Differential Transformers

Linear Variable Differential Transformers (LVDT) work on a principle similar to capacitance sensors. Here differential variation of inductance are measured. The inductance is changed by moving a non-magnetic rod with a ferromagnetic armature inside two solenoids. The position of the ferromagnetic core is determined by the effect on the currents inside the solenoids. LVDTs can reach very high accuracy: Physik Instrumente [108] produces sensors with $10nm$ accuracy over travels bigger than $1mm$. Their drawbacks are a comparatively low frequency response, the upper limit being approximately $100Hz$ and fairly large nonlinearity which sometimes requires accurate calibration.

3.10.4 Eddy Current Sensors

These sensors use the principle of impedance variation caused by eddy currents that are induced in a conductive target (usually aluminium) by a sensor coil. The coil is driven by a high frequency oscillator so as to generate an electromagnetic field that couples with the target. The gap between the sensor and target affects the strength of this coupling, thus changing the coil impedance. The sensor often uses two coils arranged in a differential configuration in a bridge, in a way similar to LVDTs and capacitance sensors. Kaman Instrumentation gives the following parameters for the differential eddy current sensors: Measuring Range $2mm$ or more, Frequency Response $22KHz$, Resolution of $3nm$ at $1KHz$, Temperature Range $-20°C$ to $+60°C$, Long Term Stability $120nm$ per month. The temperature drift is about 0.02% per Kelvin or $60nm/K$ for the longest travels [98].

3.11 Commercial Positioners

3.11.1 PZTs vs. Other Actuators

In order to make a selection among the wide range of commercially available micro-positioners, it is first necessary to choose the type of product that best matches the requirements. The first choice is between PZTs and other actuators. On a general basis, piezoelectric actuators have several advantages:

- High frequency response. The piezoelectric actuators have a wide frequency bandwidth allowing fast corrections of position. Only electrostrictive materials have a comparable bandwidth.
- Low power dissipation. Consumption of power is practically zero in quasi-static conditions and it is still low on high frequency operation. With leadscrew positioners driven by a DC motor, there will be a significant consumption of power even in static conditions. The same is true for magnetostrictive actuators.
- Low power consumption. PZT positioners can be driven with relatively low currents, resulting in low power consumption. Magnetostrictive actuators will instead require quite high currents.
- High positioning accuracy. PZTs have a resolution limited only by the position sensor and $0.1\mu m$ accuracy or less can be easily reached. This accuracy is instead at the limit of the possibilities of leadscrew motors, resulting in much higher cost and lower reliability.
- High pushing forces in a compact design. PZTs can produce forces of hundreds or thousands of Newtons from an actuator a few centimeters long and even smaller diameter. Voice coil actuators, for example, cannot produce forces larger than few hundred Newtons and they need a bulky design.
- Wide temperature range. PZTs can easily operate far below $-10°C$ and to temperature up to $+80°C$ almost without any side effect. Electrostrictive materials cannot operate at low temperatures and even leadscrew actuators can have some problems due to thermal expansion.

- Low cost. With few exceptions PZTs are the cheapest among all high precision positioning systems. Even when the cost of the closed-loop servo control is included, PZTs are still competitive.

The only serious disadvantage of using PZTs is the fact they can bear very small shear forces (this problem is in common with electrostrictive actuators). If it is possible to overcome this difficulty, a piezoelectric positioner is the best choice as a high accuracy, small range actuator.

3.11.2 High-voltage vs. Low-voltage PZTs

If a PZT is chosen, a further possible choice is between High-voltage and Low-voltage PZTs. Characteristics of those two materials are very similar and both of them can satisfy the technical requirements for the active collimator. I will summarize the properties of High-voltage PZTs which distinguish them from the Low-voltage type:

- Stronger push/pull forces. This is not a critical parameter since even if Low-voltage PZTs cannot reach the performance of High-voltage ones, they can still provide up to $3000N$ push and $700N$ pull.

- Hysteresis and linearity. High-voltage PZTs present lower hysteresis and higher linearity, but these advantages are almost useless when the actuator is used in a closed-loop system with a position sensor.

- High input voltage. This is one of the major differences. Using a high voltage supply reduces electrical noise, but again in closed-loop operation this noise does not affect accuracy. On the other hand high voltage power supplies are much more expensive than low ones.

- Cost. High-voltage and Low-voltage PZTs have almost the same price, but there is big difference in the cost of the related electronics.

- Lifetime. High-voltage PZTs have a negligible depoling, but since depoling in Low-voltage PZTs can be recovered, this is not a major advantage. Aging effects due to humidity will occur in both actuators, even if Low-voltage ones can be more exposed due to their multi-layer structure.

It is evident that Low-voltage PZTs are the preferred option. I will then briefly compare the products of different manufacturers.

3.11.3 Manufacturers and Products

I contacted and received information from the following manufacturers of piezoelectric positioners:

- Burleigh Instruments Inc.
- Dr. Lutz Pickelmann Piezomechanik Optik GmbH
- Physik Instrumente GmbH
- Queensgate Instruments Ltd.

In table 3.1 the characteristics of the equivalent products from the above companies are compared. The proposed solution was then a piezoelectric actuator system in closed-loop operation, using a short-travel Low-voltage PZT as actuator and a strain gauge position sensor. Among manufacturers P.I. offered the most competitive solution for the complete system. Moreover the performance of their actuators is superior to most of other commercial actuators in a given product range.

	P.I. P-840.10	**Lutz Pick.** 150/5/15	**Queensgate** MTP 15	**Burleigh** PZL-015
Max Travel	$15\mu m$	$15\mu m$	$15\mu m$	$10\mu m$
Push Force	$1000N$	$300N$	$840N$	$550N$
Pull Force	$100N$	$30N$	$50N$	$40N$?
Stiffness	$55N/\mu m$	$20N/\mu m$	$56N\mu m$	$10N/\mu m$
Temp Range	$-20° + 80°C$	$-40° + 80°C$	$-20° + 60°C$	-
Temp Expansion	$0.1\mu m/K$	$< 0.2\mu m/K$	$0.1\mu m/K$	$0.2\mu m/K$
Frequency Band	$14KHz$	$20KHz$	$11KHz$	$3.5KHz$
Input Voltage	$-20 + 120V$	$-30 + 200V$	$0 - 120V$	$0 - 100V$
Capacitance	$1.8\mu F$	$1.8\mu F$	$1.6\mu F$	$0.8\mu F$
Length	$32mm$	$21mm$	$30mm$	$34mm$
Cost	£345	£220	£288	£310
With Sensor	£591	NO	£1664	NO

Table 3.1: Comparison of similar products from some major manufacturers of piezoelectric actuators. The most relevant characteristics of the actuators are listed and compared (from the company product catalogues).

Chapter 4

Defining the ISAAC Project

4.1 A Flexure Compensation System for ISIS

ISIS (Intermediate-dispersion Spectroscopic and Imaging System) is a double beam spectrograph designed and built by the Royal Greenwich Observatory and has been in operation at the Cassegrain focus of the $4.2m$ William Herschel Telescope (WHT) for several years. Systematic measurements of the position of the spectrum as a function of zenith distance and rotation angles have shown evidence of flexure in ISIS structure and probably in its supporting flange system. At the start of this research project ISIS was one of the largest instrument of that type and, not surprisingly given its size, was affected by significant spectrum drifts at the detector, up to $20\mu m/hr$ [58]. ISIS was therefore an ideal instrument to prove the concept of active flexure compensation in large Cassegrain spectrographs.

The ISIS Spectrograph Automatic Active Collimator (ISAAC) project was conceived as an open-loop system for active compensation of flexure in the ISIS spectrograph. The choice of the active element fell on the collimator for two main reasons: it is easy to control the mirror with actuators mounted on its back and to steer it when applying active corrections, and given its position in the spectrograph, it can be accessed and replaced easily. Moreover the collimator is a natural choice for flexure compensation, being in place for every spectrograph setup and producing significant image shifts on the detector with relatively small tilts. The latter is an important point in avoiding introducing optical aberrations by modifying too much the optical

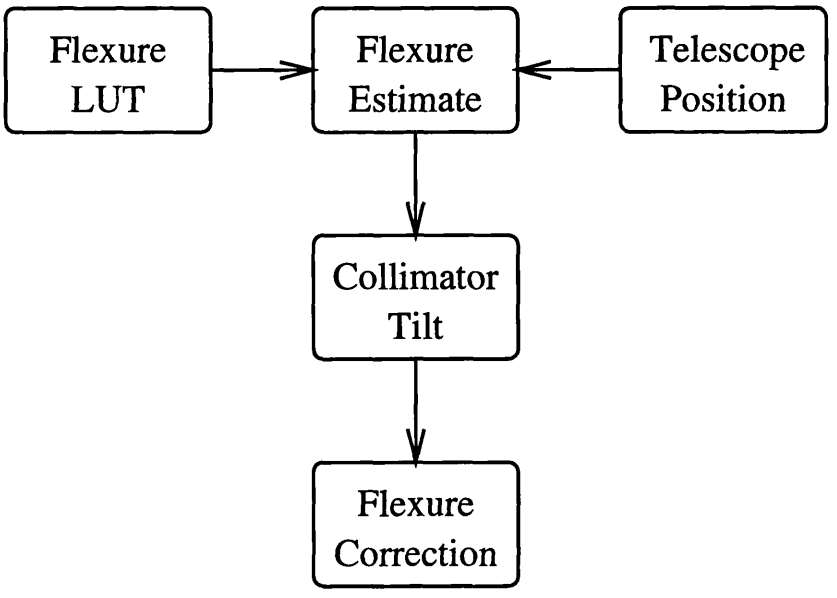


Figure 4.1: Schematic functional diagram of the ISAAC flexure compensation system. The system works in open loop, with a flexure look-up table (LUT) from which the spectrograph flexure is estimated (knowing the telescope position). This information is used to drive the collimator tilt that compensates for flexure.

path. The blue collimator was chosen for its larger flexure shifts and because the mirror is aluminium-coated, a process that we could reproduce more easily in our laboratory.

This is the background that led us to define the scope of the ISAAC project. The active collimator system of flexure compensation is based on a simple concept: when the spectrum position on the detector drifts during tracking, the collimator is slightly tilted, thus deflecting the optical path of the light from the telescope and restoring the image in the original position. Figure 4.1 shows a schematic drawing of the ISAAC system illustrating the functional concept of the system. The collimator is finely steered with three piezoelectric actuators pushing against its back surface. The amount of tilt required is calculated on the basis of the spectrograph geometry and of a flexure map, which gives the detector shifts for a set of different spectrograph positions. The orientation in space of the spectrograph is calculated from the celestial coordinates of the observed object, the sidereal time at the start of the exposure and the position angle of the slit on the sky. Once the position of the spectrograph is known, the control software interpolates the flexure map to derive the

corresponding shifts, translates these into actuator voltages and tilts the collimator to keep the spectrum stable.

The project is conceived as a first step towards a closed loop flexure compensation system and throughout the project the open and closed loop options have always been seen as complementary rather than alternative options. Moreover, with the new generation of 8 m telescopes coming up, we envisaged that this experiment would be highly relevant for all the Cassegrain instrumentation for these telescopes and in particular for the Gemini High Resolution Optical Spectrograph (HROS).

4.2 Overview of ISIS Spectrograph

ISIS is the most popular instrument of those available at WHT telescope; it a double spectrograph, with the blue arm optimized for the spectral region from 3000 to 5500$\AA$ and the red arm for the region from 5000 to 10000$\AA$ and it is situated at the f/11 Cassegrain focus of the telescope. It is possible to use both arms at the same time by inserting a dichroic mirror after the entrance slit. If this is removed, only the red arm will receive light, while substituting the dichroic with a flat mirror directs the light into the blue arm. An exploded view of the optical layout of ISIS is shown in fig. 4.2 [99]. Each of the two arms of ISIS is a conventional spectrograph with interchangeable reflection gratings. I derived most of the information about ISIS listed below from the telescope manuals and from discussions with telescope staff and people at the Royal Greenwich Observatory [100]. Unfortunately, to the best of my knowledge no scientific publication concerning the spectrograph and its design is available.

4.2.1 Slit area

The slit width can be varied from $30\mu m$ to $5mm$ and the spectrograph geometry is such that $1mm$ projects to $4.51arcsec$ on the sky. A dekker just before the slit controls its length by masking a suitable fraction of the slit jaws; the narrowest dekker projects to $1.2arcsec$ on the sky.

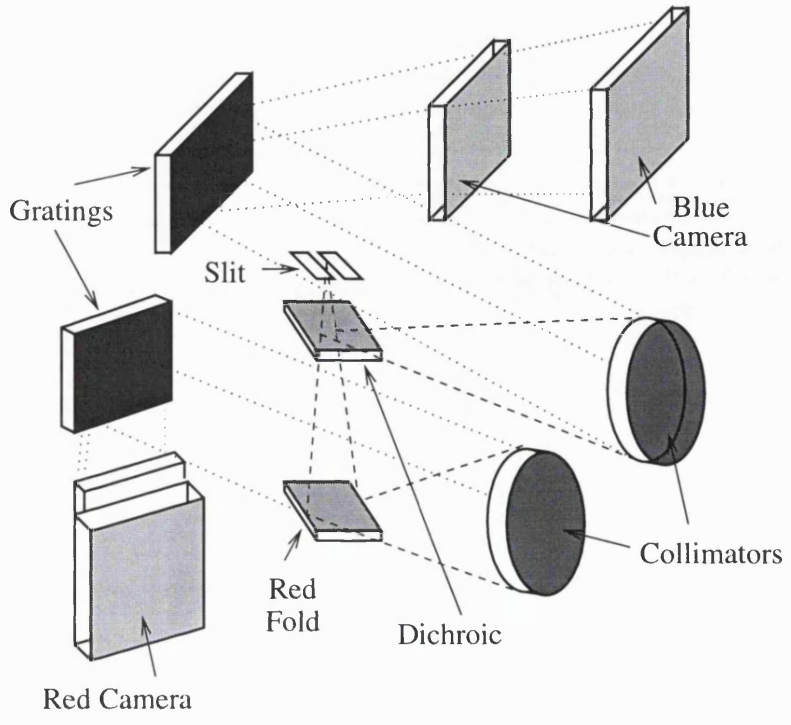


Figure 4.2: Exploded view of ISIS spectrograph showing the light paths through the components. The dichroic splits the light between the red and blue arms of the spectrograph and it can be substituted by a folding mirror if only the blue arm is used. Pre-slit optics, filters and shutters are not shown. Diagram copied from the ISIS Users Manual

4.2.2 Collimators

Both collimators are off-axis paraboloids (nominal off-axis angle 5°) with a focal length $F_{col} = 1650mm$ and they accept a beam diameter of $B = 150mm$ (f/11 focal ratio). They are housed in an aluminium cell which can be moved with a stepper motor controlled by an LVDT sensor. The collimator assembly for both arms is shown in fig. 4.3. The collimator mirror is focused by moving the cell along parallel bars on special carriage, moved by the stepper motor. When the mirror has reached the focus position a spring system locks the carriage and the cell with it, on the parallel bars. The collimator is then aligned by moving the cell with respect to the carriage with two fine capstan screws; the cell can then be clamped on three different hinges which gradually restrict the freedom of motion. The existing collimator is kept in a fixed position within the cell by a diaphragm on the back and several

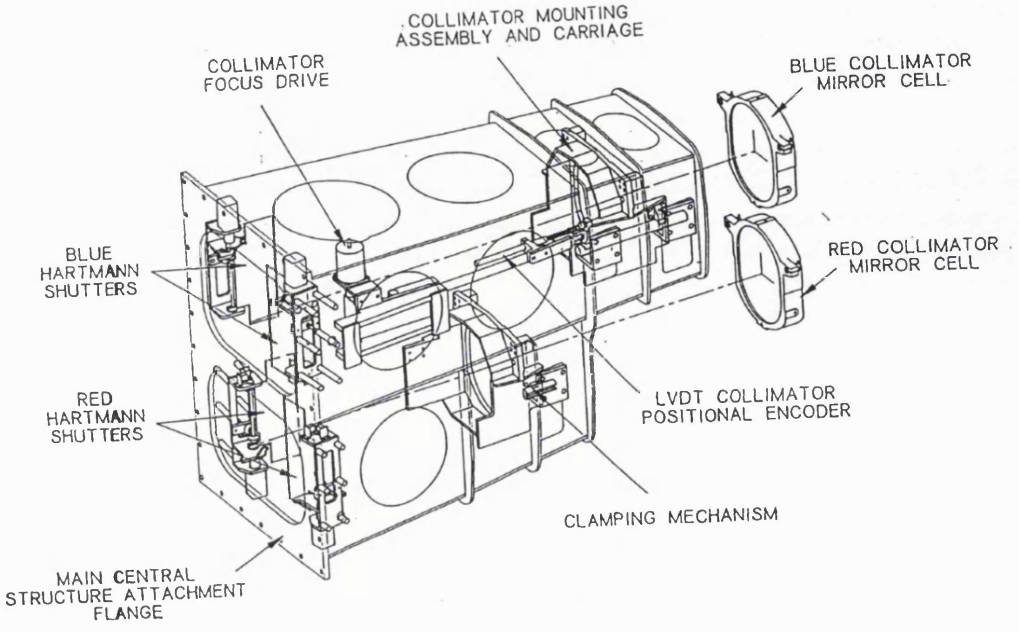


Figure 4.3: The ISIS collimator assembly. The collimator cells are clamped to the collimator carriage with a kinematic mount. A stepper motor and LVDT sensor are used for focusing the mirrors.

screws on the sides.

4.2.3 Gratings

The spectrograph has a set of interchangeable gratings for both arms, with ruling varying between $158 lines/mm$ and $1200 lines/mm$. A holographic grating with $2400 lines/mm$ is also available on the blue arm. In the configuration in which they are used, most of these gratings have a low level of anamorphism. The anamorphic factor $\xi = \cos\vartheta_r / \cos\vartheta_i$ increases with the grating dispersion and with the wavelength, but for most cases it remains close to 1.0. At $4000\mathring{A}$ it ranges between 1.0 for the $158 lines/mm$ ruling and 1.2 for the $1200 lines/mm$. The holographic grating is the exception, with an anamorphism of 1.5 and rapidly increasing with the wavelength, so that at $6000\mathring{A}$ it is over 2.5 [99]. The collimator to camera angle in ISIS is nominally 40^o.

4.2.4 Cameras and Detectors

The ISIS blue and red cameras are of a folded Schmidt design, with a focal length $F_{cam} = 500mm$. The cameras are usually equipped with cryogenically cooled CCDs, the most commonly used being the Tek-1024, a 1024×1024 array of $24\mu m$ square pixels or the EEV-88300, a 1152×1242 array of $22.5\mu m$ square pixels. Typical cosmic ray event rates for these CCDs are $0.2sec^{-1}$ events over the entire array for the EEV chip and $0.4sec^{-1}$ events for the Tektronix chip.

4.2.5 Calibration Lamps

The spectrograph has an extensive choice of calibration lamps which can be mounted in a specific module. The standard ones, however, are the Cu-Ar lamp, which better covers the blue end of the spectrum, and the Cu-Ne lamp, with coverage mostly in the red. A continuum tungsten lamp is also provided as a standard in the module. Two eight-position filter wheels are placed at the output of the lamp system with both colour and neutral filters of various intensity.

4.2.6 Control System

The control of ISIS is arranged as a distributed system: low-level instrumentation control is managed by a local processor placed on the spectrograph itself and this communicates with the main computers in the telescope control room via an Ethernet link. The main control software is written as D-tasks in the ADAM environment which are run as sub-processes of the general observer process. The user interface is an interactive console where the spectrograph status is displayed and the user can type in specific commands to the instrument.

4.3 A Study of Flexure in ISIS

The ISIS spectrograph was originally designed to have maximum flexure shifts of $5\mu m$ over one hour tracking time. The instrument however, has shown more flexure than predicted and at the start of this project displayed up to $20\mu m/hr$ drifts

with an average of $10\mu m/hr$. Moreover, occasionally some components inside the spectrograph become loose and the situation gets much worse, since the spectrum can suddenly shift by a large amount when the component flips between two positions. While it is relatively easy to track down a loose component and fix the problem, identifying the source of the residual flexure and trying to correct it, has been extremely difficult. At the time of writing no satisfactory solution, or indeed explanation for the relatively large spectrum shifts has been found. In this section I report the results of a preliminary study of flexure in ISIS which I performed with the purpose of gaining essential information about the behaviour of the spectrograph with respect to flexure and to estimate the effectiveness of a flexure compensation system. The study utilizes some early flexure data measured by R. Rutten for routine maintenance of the spectrograph (1993-1994) and some data I measured in my first visit to the observatory in May 1994. In section 7.4 I report the results of more extensive flexure measurements I took during the commissioning phase of the active collimator in July 1995.

4.3.1 Generation of the Flexure Model

Flexure is measured by recording the shifts of a point source on the CCD detector for different telescope and spectrograph positions. An approximate point source is created by observing a spectral line through a narrow dekker. In practice a whole spectrum of a calibration lamp is recorded, so that each line of the spectrum appears as an almost circular spot. The profile produced by the dekker in spatial direction is not exactly Gaussian, but it is a good enough approximation for the software used to analyze it. By cross-correlating, both in spectral and spatial direction, two spectra corresponding to different spectrograph orientations, it is possible to measure drifts in two orthogonal directions. IRAF software has been used for the cross-correlating algorithm: this technique gives an accuracy of better than 1/20 of a pixel in the measurements.

The spectrograph position is defined by only two parameters: the zenith distance z of the telescope (or elevation $E = 90^o - z$) and the position angle (PA) on the Cassegrain turntable. To avoid confusion with the parallactic angle q, which is sometimes also indicated with PA, in the future we will use ψ to indicate the angular

position of the Cassegrain rotator. It is easy to see why only two angles are needed to define the position of the spectrograph: the third angle corresponds to an azimuth rotation of the whole telescope, which obviously has no effect on the flexure inside the spectrograph. A flexure map that covers the whole range of telescope and spectrograph positions can be generated by moving the telescope in steps from zenith to 15° above horizon (mechanical limit) with the turntable rotator at different angles. At each step the spectrum shifts relative to the zenith position (and a chosen turntable angle) are measured. The number of points for the flexure map needs to be balanced against the time it takes to collect them: I decided that 15° steps in the zenith distance and 45° in turntable position angle was a reasonable compromise.

4.3.2 Flexure Map Characteristics

A typical flexure map produced in this way is shown in fig. 4.4; the graph is for ISIS red arm and so are all of the next few. This is because at the beginning of the project, when I started investigating the characteristics of flexure in ISIS much more data was available for this arm and it was not yet entirely clear on which arm the active collimator would be placed. However, the few data concerning the blue arm I had at the time was fairly similar to the red arm data and there was no reason to believe that the blue arm would behave differently with respect to flexure. Indeed, when I finally managed to get a good set of flexure data for the blue arm during the on-site experiment, this assumption proved to be correct.

Several conclusions can be drawn from this flexure map: it is evident that the *direction* of flexure shifts depends mostly on the turntable position, while the *amount* of movement is strongly influenced by the telescope elevation. This is consistent with image motions caused mainly by elastic deformation of the spectrograph structure under gravity. When the telescope is lowered from zenith to horizon, with the Cassegrain rotator in a fixed position, the structure flexes and the spectrum drifts. By changing the turntable angle, the drift occurs in a different direction, but it is of a similar magnitude. The graph is also consistent with the average $10\mu m/hr$ drifts observed during tracking. The average tracking speed for the telescope is around 12 degrees/hr along zenith distance (this can be derived from the equations converting equatorial into altazimuthal coordinates, see section 4.4.3.1) and two

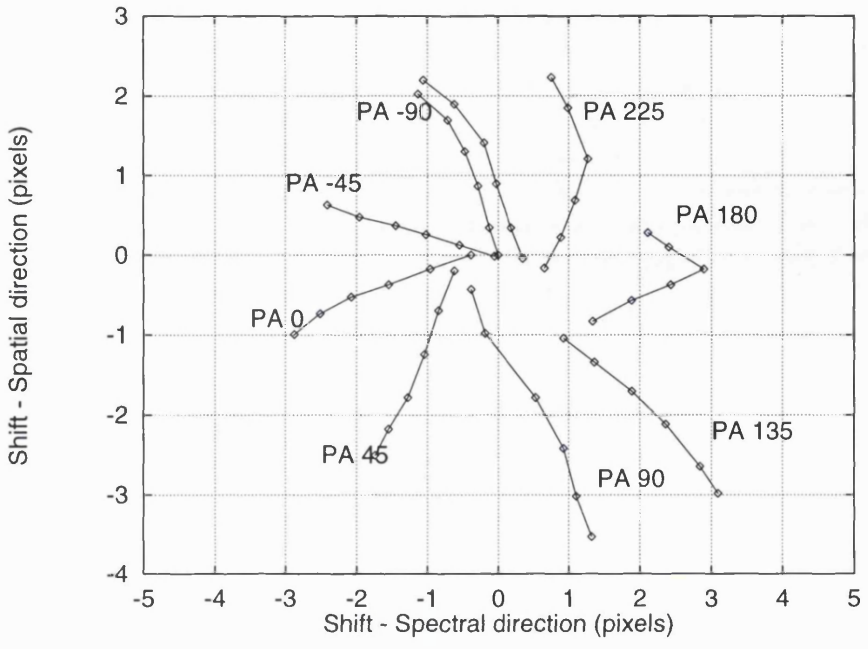


Figure 4.4: Flexure map for the ISIS red arm. The shifts are displayed as they would appear on the CCD detector, each line corresponding to a different turntable position and each point within the line indicating the zenith distance (15° steps); the pixel size is $24\mu m$.

successive points on the map (which are therefore separated by slightly more than 1 hour time interval) are around half a pixel apart or $12\mu m$.

A relevant feature of the graph is the fact is that even if the the tracks for different turntable angles all start from zenith, the points corresponding to that position do not coincide and are displaced from the origin. These points are expected to coincide because when the telescope is pointing at zenith a rotation of the Cassegrain turntable does not change the orientation of the spectrograph with respect to the gravity vector. This is a clear indication of hysteresis in the spectrograph structure and a consequence of the way the data is taken. The procedure for taking a flexure map starts with the telescope pointing at zenith and Cassegrain rotator at a given position (always $\psi = -90°$). The telescope is moved in elevation in steps down to $E = 15°$ and then back to zenith before the turntable is rotated. Due to hysteresis in the spectrograph flexure, the starting point for the new track (with a different rotator angle) is displaced from the origin. If the map is recorded using a different

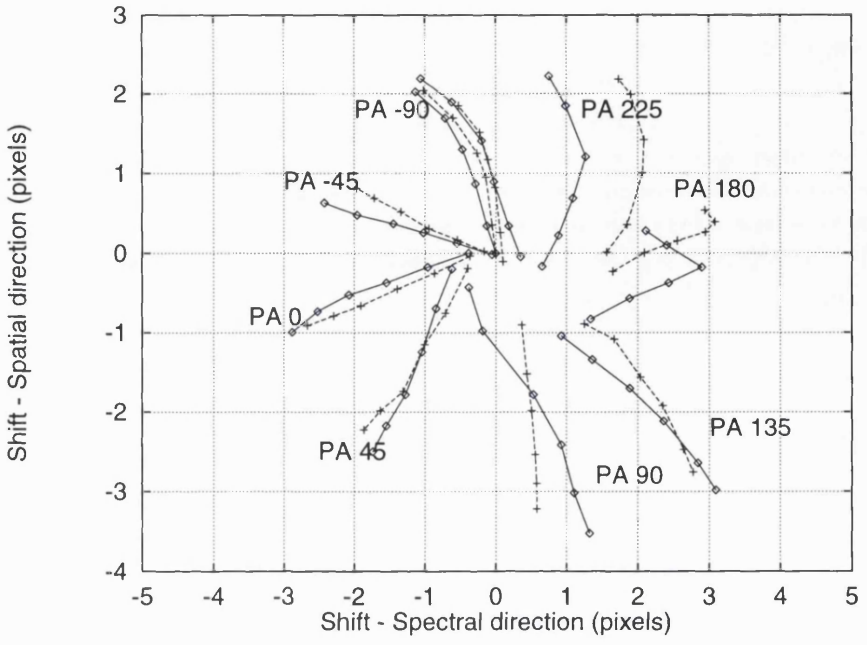


Figure 4.5: Two flexure maps taken only two days apart but with slightly different procedures. It emphasizes the presence of hysteresis in the ISIS structural distortions under gravity.

procedure, the result will be also different. In fig. 4.5 the previous map is compared with another one taken only two days earlier, but following a procedure in which the turntable is rotated *before* the telescope is raised back to zenith. Both sets of data were taken during my earlier visit to the observatory on May 1994.

Further proof that the observed behaviour is due to hysteresis is the fact that when the telescope is pointing at zenith and the Cassegrain turntable is rotated, hardly any image shift is detected. During one of such tests the maximum drift observed was 0.03 pixels in spectral direction and 0.06 pixels in spatial direction [101]. In comparison, the starting points of the tracks in fig. 4.4, in theory corresponding to the same spectrograph movement, can be as far apart as 1 pixel or more. It is not clear what causes this hysteresis: in theory the metallic structure should be almost perfectly elastic. Junction points within the structure, bolts and supports for the optical components may be responsible, or there might be stresses generated when connecting the spectrograph to the Cassegrain rotator. Hysteresis has been measured also in other Cassegrain spectrographs like LDSS-2, even if it was not as

large [102]. The presence of hysteresis has serious implications when trying to model flexure for an open loop system of active compensation, which are discussed in the next section.

It is important to stress that the above data and most of the following flexure maps were taken with the same spectrograph setup and in particular with the $158 lines/mm$ grating and the Tektronix detector. Later we observed that flexure is affected by changing the grating and therefore the same setup is necessary for the consistency of the results and conclusions. A possible source of flexure that I have not considered so far is a temperature gradient inside the spectrograph. This can produce differential expansion in the structure and consequently image shifts; experience however, has shown that, if present, this is a minor factor compared to gravitation and hysteresis. Moreover this is not expected to affect the data collection for a flexure map, since the whole process takes only about two hours.

4.3.3 Modelling Flexure

An open loop flexure compensation system requires accurate modelling of the flexure shifts in the spectrograph. The accuracy of the correction relies on how closely the flexure map models the spectrograph real performance. This task is complicated enormously by the presence of hysteresis, because of its intrinsically irreproducible nature. Figure 4.5 clearly shows the serious impact of hysteresis on the reliability of this flexure measurement as a model for the compensation system. Also, the fact that rotating the turntable with the telescope pointing at zenith does not affect flexure is in direct contrast with the data in the measurement. These considerations brought me to the conclusion that the raw data coming from the flexure measurements cannot be used as a reliable look-up table. Moreover, in this raw flexure map, the hysteretic effects add from one track to the next eventually producing large errors. During normal telescope tracking, the telescope motion, and so the hysteresis, is much smaller. Also, since hysteresis depends on the previous history in the telescope movements, the above errors will add differently with a different history.

To estimate the errors resulting from using a raw flexure map like the one in fig. 4.4 we can plot the difference between this data set and the one taken two days earlier of fig. 4.5 as detector shifts (fig. 4.6). One set of data is subtracted from the other

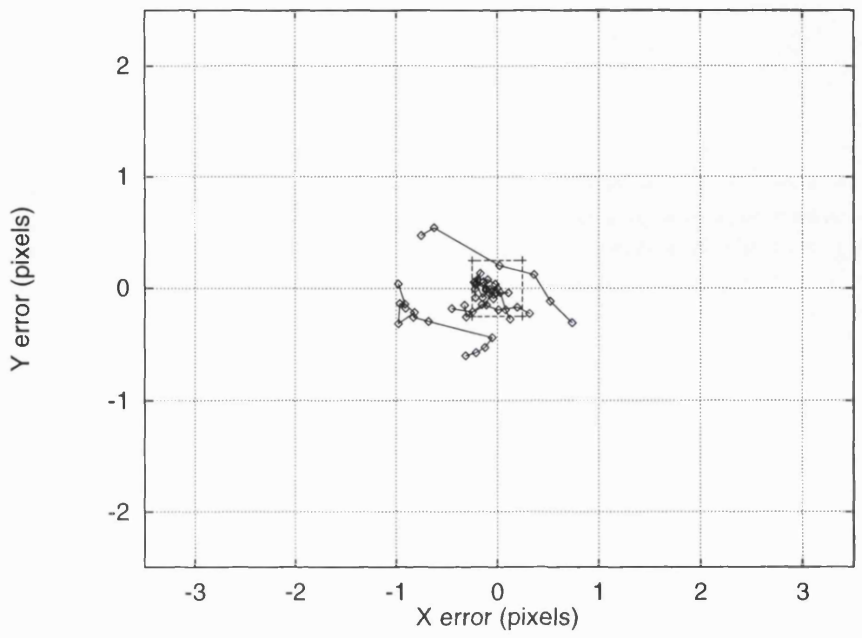


Figure 4.6: Compensation errors resulting from using a raw flexure map to correct flexure from a map taken with a different procedure. Since the telescope was moved in a different way, relatively large errors are produced by hysteresis in the measurements. The square in the centre is half a pixel in size and contains 56% of the points.

point-by-point; this simulates the performance of an active compensation system that uses the first data set to correct the shifts in the second (i.e. a different telescope motion). In most cases the stability is better than half a pixel, but sometimes the error can go up to 1-2 pixels; in the graph 56% of the points are inside a half pixel square centered on the origin. This situation however, can be improved by refining the flexure model and making some useful assumptions. Firstly, we can notice that during a normal exposure the telescope does not make such extreme movements as when the flexure data is recorded and so the hysteresis is expected to be much less significant. Secondly, we already know that there is negligible hysteresis when the turntable is rotated with the telescope pointing at zenith. Therefore I assumed that flexure could be better modelled by a map where each track is translated in the detector plane so that all the starting points would coincide with the origin; figure 4.7 shows this modified flexure map.

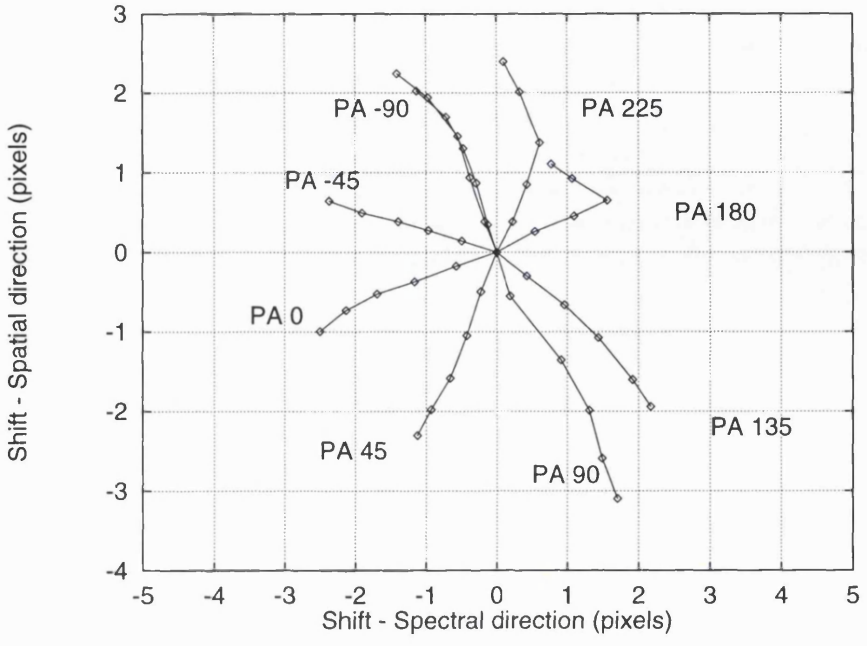


Figure 4.7: Modified flexure map, obtained by translating every track with different rotator position angle so that the starting points ($z = 0$) coincide with the origin. This procedure reduces the impact of hysteresis on the flexure model.

A similar test to the one described above (with point by point subtraction of two maps taken with different telescope motions), can be repeated with the two modified flexure maps. This already produces a significant improvement: now 70% of the points are inside the half pixel square. The test however overestimates the error induced by hysteresis because of the large telescope movements involved. To better estimate the effect hysteresis may have on the accuracy of the flexure compensation system, I made a slightly different simulation. In this case, one modified look-up table is still used to correct the other but I considered only points within 15° in zenith distance and 45° in the rotator position angle from the starting point. Moreover the subtraction is not done point by point, but between the corresponding vectors joining adjacent points in the two tables. The result is shown in fig. 4.8: in this case the position error due to hysteresis is less than 0.25 pixels for 75% of the cases. Note that in the figure the points inside the half pixel square shown are only 0.25 pixels away from the starting point (centre of the square). In the previous simulations the starting and ending points could be any couple inside the square, that is as much as half a pixel apart.

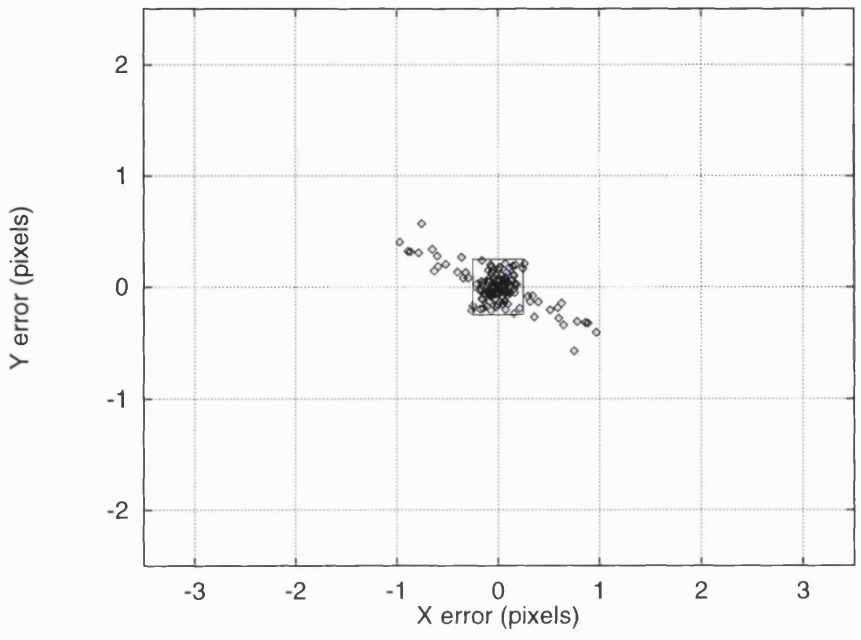


Figure 4.8: Simulation of flexure correction errors due to inaccuracies in the flexure look-up table. The errors are produced by residual hysteresis in the modified flexure model. In the simulation, the points are within 15^o in zenith distance and 45^o in the rotator position angle from the starting point (center of the square). This is to better simulate a normal astronomical observation, when the telescope moves only by relatively small angles.

The above simulations proved that the modified flexure map is a much better model for estimating the flexure shifts in the spectrograph. This model relies on reducing the impact of hysteresis on the raw data coming from the flexure measurements and assumes that hysteresis has little impact during normal telescope tracking motions. The first hypothesis has support from experimental observations, namely the fact that the modified flexure map is more reproducible than the original set of data. The second hypothesis is based on the observation that, in general, hysteresis is greater when the direction of motion is inverted, while during tracking the telescope continuously moves in the same direction. I made no attempt to include hysteretic effects into the model. Apart from being an incredibly difficult task, it is dubious whether such a model could ever work. Hysteresis can be modelled when the history of the movements follows a simple pattern, but any model would fail for random motions. This was proved, for example, for hysteresis correction models for piezo-

electric actuators [103], where good results were obtained only when the actuator was moved cyclically.

This study had the aim of better understanding the characteristics and behaviour of flexure in the ISIS structure, so that I could model it more accurately and also understand the limitations that hysteresis poses to an open loop compensation system. After ISAAC was finally completed, I had a chance of testing it on site on the spectrograph. During these tests I could also collect more information concerning flexure in ISIS and do some specific tests. The results of this further investigation into flexure are discussed later. The important result of this preliminary study is that hysteresis is likely to be a limiting factor in the accuracy of ISAAC. Even so the errors should be less than 0.25 pixels, greatly improving the spectrograph performance.

The conclusion is that flexure in ISIS can be modelled quite accurately, provided that some precautions are taken. The flexure data needs to be recorded always with the same procedure: this process is now automated with computer program that moves the telescope and saves the exposures in a standard (and quick) way. The raw data from this flexure map needs then to be modified to reduce the influence of hysteresis as much as possible. I showed how, by simply translating the telescope tracks in the detector plane, the compensation accuracy could be significantly increased. With these precautions and modifications the flexure map becomes much more reproducible and reliable. I need to stress however, that flexure tends to change with time, particularly if the spectrograph is taken off the telescope or if its setup is changed considerably. Therefore, it will be necessary to take frequent flexure measurements to make an open loop active compensation system really effective.

4.4 Active Collimator Specification

The requirements and specifications for the active collimator system mainly concern the actuator performance, the cell design and, to a lesser extent, the electronics and the computer hardware. I included the environment in which ISAAC will operate among the requirements for the impact it has on the actuator choice and, marginally on the electronics and mechanical design.

4.4.1 Mechanical Requirements

The active collimator for ISIS was designed from the beginning to be integrated as much as possible with the spectrograph and its potential as a future common-user instrument always kept in mind. The first requirement was that the collimator cell should be interchangeable with the existing one and that mechanical installation should be relatively easy. This puts some severe constraints on the mechanical design which are discussed in detail in chapter 5. It was also imperative that the collimator cell and the mirror movements should have hysteresis reduced to a minimum. Some elastic flexure could be tolerated, but we tried to minimize that as well, to avoid introducing some new external sources of flexure in the spectrograph. This means that the mirror needs to be mobile and the motion free of hysteresis, but it also needs to keep the position firmly after the movement.

Mechanically the collimator mirror is a solid block of Zerodur, a near-zero thermal expansion glass, about $20cm$ diameter, $4.0cm$ thick and has a weight of $5Kg$. The mirror will be tilted by small angles, but in every direction (that is, independently around two orthogonal axes). To be able to achieve this freedom of movement I decided to support the mirror on three points (which identify a plane). The force requirement for the actuators can then be derived on the basis that they need to sustain at least the weight of the mirror ($50N$). The mirror however, has to be kept firmly in place by pushing it against some spring-like element on the other side, so a higher force is required. I concluded that the actuators should apply the maximum force that does not introduce significant stresses in the mirror. Finite element analysis of the mirror (performed by Mark Dryburgh) showed that $100N$ at each actuator could be safely applied, so this was chosen as minimum specification.

4.4.2 Stability Requirements

The level of image stability at the detector is the main driving factor in the specification for the mechanical design and for the actuators. The amount of spectrum shift that can be tolerated is a matter of open discussion and it is driven by the astronomical observations to be performed on the spectrograph: I discussed these matters in chapter 2. For the purpose of drawing the specifications for the ISAAC

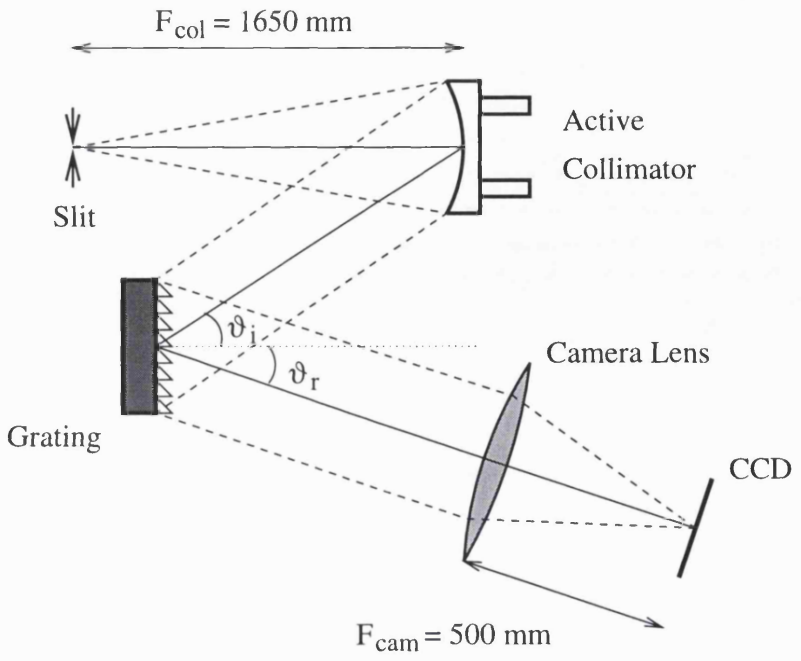


Figure 4.9: Schematic diagram of the ISIS spectrograph that shows the relation between collimator, grating and the image on the detector (not in scale). When the active collimator is tilted, the angle of incidence on the grating is changed and the spectrum on the detector moves.

experiment, I chose 0.1 pixel as a reasonable target for the spectrum stability. This value would be maintained as long as flexure remains within the correcting range of the collimator. With a pixel size of $22.5\mu m$ or $24.0\mu m$, this requirement translates into around $2\mu m$ error on the detector. This value for spectrum stability is of the same magnitude as the one required for the Gemini Multi Object Spectrograph (GMOS): 0.1 pixel per hour with $15\mu m$ pixels [53]. The requirement for the High Resolution Optical Spectrograph (HROS) for the Gemini telescope is also $2\mu m/hr$ (1/20 of the resolution element) [104].

Considering the time gap between updates and the fact that $2\mu m$ was the maximum error I wanted to achieve, I chose $\Delta x = \pm 1\mu m$ as the target position accuracy on the CCD. The accuracy of the flexure look up table was a secondary factor in this choice, since, as I explained earlier, the active collimator was designed keeping the closed loop option in mind, too.

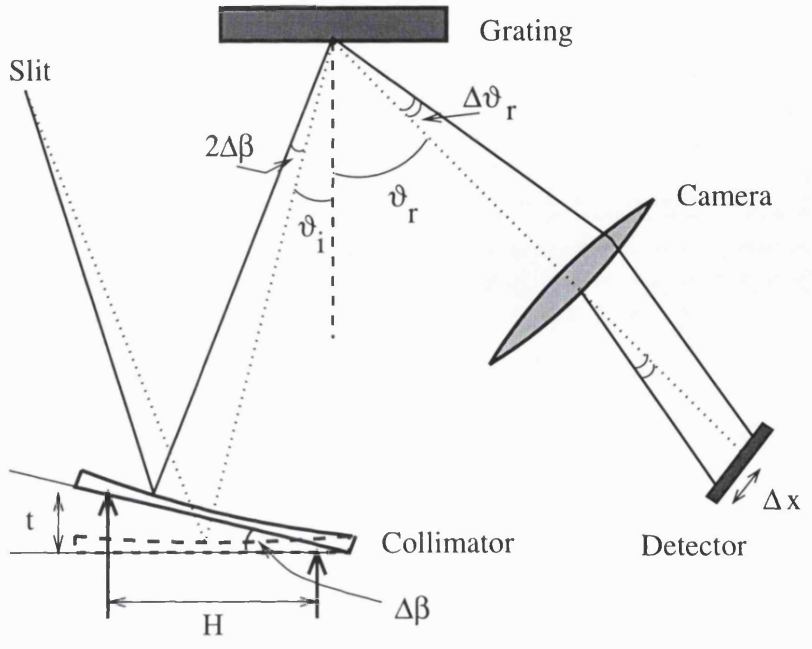


Figure 4.10: Effect of a tilt of the collimator on the position of the spectrum in the camera focal plane. In the figure the collimator is tilted by expanding one actuator by t and keeping the other actuator at zero expansion.

4.4.2.1 Collimator Positioning Accuracy

To calculate how the tilts of the collimator mirror translate into shifts on the detector, we need to look at the layout of the spectrograph, as shown schematically in fig. 4.9. The diagram is only shown to explain the formulae below and has no resemblance to the real spectrograph. When the collimator is tilted by a small angle $\Delta\beta$, the central ray coming from the slit is deflected by twice that angle, due to the reflection on the collimator. At this point we need to distinguish between tilts in the grating dispersion plane and tilts orthogonal to it: if the tilt is in the dispersion plane, it will be affected by the anamorphism of the grating. The effect of a collimator tilt is illustrated by the diagram of figure 4.10. The incident parallel beam on the grating is in fact tilted by the same angle $2\Delta\beta$, so that if ϑ_i was the original angle of incidence on the grating we have:

$$\vartheta_i \rightarrow \vartheta_i + 2\Delta\beta \tag{4.1}$$

The beam reflected by the grating is connected to ϑ_i through the grating equation:

$$\frac{m\lambda}{a} = \sin\vartheta_i - \sin\vartheta_r \tag{4.2}$$

By differentiating equation 4.2 with respect to ϑ_i we have:

$$\frac{\Delta\vartheta_r}{\Delta\vartheta_i} = \frac{\cos\vartheta_i}{\cos\vartheta_r} \tag{4.3}$$

That is, the anamorphism of the grating introduces a further multiplying factor (called anamorphic factor) $\xi = \cos\vartheta_r/\cos\vartheta_i$ in the beam hitting the camera, so that the total deflection of the beam leaving the grating is:

$$\Delta\vartheta_r = 2\Delta\beta\frac{\cos\vartheta_i}{\cos\vartheta_r} \tag{4.4}$$

If the collimator tilt is not in the dispersion plane, then the anamorphism is absent. We have therefore to distinguish between detector shifts in the spectral direction Δx and shifts in the spatial direction Δy, which are obtained by multiplying the beam tilt times the camera focal length, having:

$$\Delta x = 2F_{cam}\Delta\beta\frac{\cos\vartheta_i}{\cos\vartheta_r} \tag{4.5}$$

and

$$\Delta y = 2F_{cam}\Delta\beta \tag{4.6}$$

The anamorphic factor ξ, however, is not far from 1 for most of the ISIS gratings, as I pointed out in section 4.2. Thus, for the purpose of this discussion, it can be safely neglected. Given the stability requirement of $\Delta x = \Delta y = \pm 1\mu m$ and neglecting anamorphism, we have:

$$\Delta\beta \simeq \frac{\Delta x}{2F_{cam}} \tag{4.7}$$

This gives a value of $\Delta\beta = \pm 0.2 arcsec$ ($\pm 10^{-6} rad$) as the tilting accuracy required for the collimator mirror. The accuracy in positioning the actuators will depend on this and on the distance H between the two actuators. When the actuator expands by a given amount t, the mirror is tilted by an angle $\beta = t/H$ and using equation 4.7 we obtain:

$$t \simeq \frac{H}{2F_{cam}}\Delta x \tag{4.8}$$

This equation relates the actuator movement to the shift in the camera focal plane; if both these movements are expressed in microns and we take an actuator distance

$H = 16cm$, we get a the scale factor given by the ratio $2F_{cam}/H = 6.25$. This means the actuator movements are magnified on the CCD detector by a factor $R = 6.25$. When calculating the accuracy required for a single actuator, however, we need to consider that the expansion t that produces a collimator tilt is in fact the difference of two actuator expansions $t = t_1 - t_2$, therefore the error on t is:

$$\Delta t = \Delta t_1 + \Delta t_2 = 2\Delta t_a \tag{4.9}$$

I indicated with Δt_a the required position accuracy for a single actuator. Combining this with equation 4.8 we finally obtain:

$$\Delta t_a = \Delta x/2R = 80nm \tag{4.10}$$

This is a very stringent requirement and it has important consequences in the choice of the actuators, since not many can reach this level of accuracy; this is discussed in detail in the next chapter.

Incidentally, equation 4.8 also gives us the *maximum* amount of correction we can apply with the active collimator. Taking an actuator maximum expansion of $15\mu m$, the corresponding shift at the detector is $\Delta x = 94\mu m$, or nearly 4 pixels (considering $24\mu m$ pixels); this is more than enough to correct the worst case flexure. The maximum tilt of the mirror would be $\beta = 20arcsec$ $(10^{-4}rad)$.

4.4.2.2 Mirror Radial Stability

The stability goal at the detector also affects the amount of lateral motions of the mirror that can be tolerated. In other words, it affects the specified accuracy for the radial location of the collimator. When the mirror is shifted by Δr in the plane orthogonal to the optical axis, the slit moves by the same amount in the collimator focal plane. Thus, the parallel beam leaving the collimator will be deflected by:

$$\vartheta = \frac{\Delta r}{F_{col}} \tag{4.11}$$

Again neglecting the anamorphism of the grating, the corresponding displacement at the detector will be:

$$\Delta x \simeq \frac{F_{cam}}{F_{col}}\Delta r \tag{4.12}$$

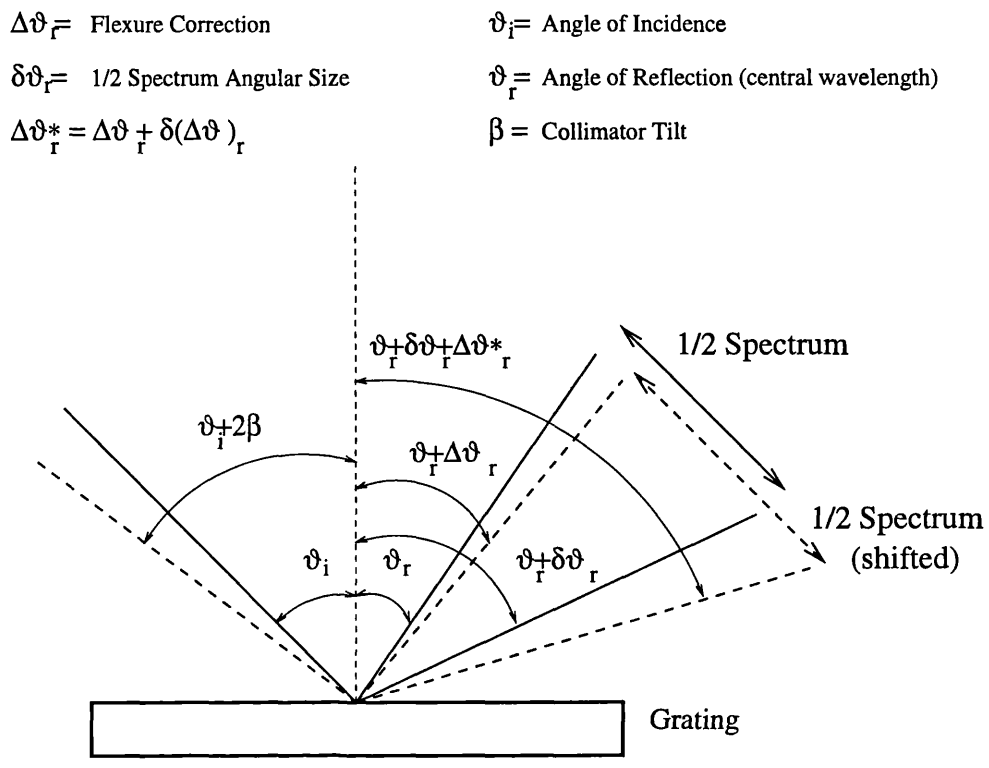


Figure 4.11: Changes in the angles of incidence and reflection on the grating due to flexure compensation. As a result of the flexure correction, the spectrum is not only shifted, but also slightly stretched.

That is the two displacements are related by the ratio of collimator and camera focal lengths. Given the numerical value $F_{cam}/F_{col} \simeq 1/3$ for the ISIS spectrograph, a lateral stability of about $\pm 3\mu m$ is needed to keep motions at the detector within specification.

4.4.2.3 Flexure Correction and Grating Dispersion

A tilt β of the collimator changes the angle of incidence on the grating, albeit by a very small amount. It is necessary, therefore, to calculate the effect on the spectrum at the edges of the CCD when compensating for the motion at the centre, so that the flexure compensation works for the whole spectrum. This situation is shown schematically in fig. 4.11 When the centre of the spectrum is reflected by the grating at an angle ϑ_r, the edge will be reflected at a different angle $\vartheta'_r = \vartheta_r + d\vartheta_r$, where $d\vartheta_r$ is half the angular size of the CCD. To calculate the effect on ϑ'_r by

the collimator tilt, we need to derive how the change in the angle reflected by the grating $\Delta\vartheta_r$ due to flexure correction, depends on the angle of reflection itself ϑ_r. This relation can be obtained by differentiating equation 4.4 with respect to ϑ_r.

$$\frac{d}{d\vartheta_r}(\Delta\vartheta_r) = 2\beta\cos\vartheta_i \frac{d}{d\vartheta_r}\left(\frac{1}{\cos\vartheta_r}\right) \tag{4.13}$$

Therefore:

$$\frac{d}{d\vartheta_r}(\Delta\vartheta_r) = \Delta\vartheta_r \cdot \tan\vartheta_r \tag{4.14}$$

Or:

$$d(\Delta\vartheta_r) = d\vartheta_r \cdot \Delta\vartheta_r \cdot \tan\vartheta_r \tag{4.15}$$

The above equation tells us that while the centre of the spectrum is shifted by an angle $\Delta\vartheta_r$ when the collimator is tilted by β, the edge is shifted by $\Delta\vartheta_r + d(\Delta\vartheta_r)$, thus stretching the spectrum. Combining equation 4.5 and equation 4.4 we have an expression for the flexure correction that depends on the shift on the CCD (rather than on β):

$$\Delta\vartheta_r = \frac{\Delta x}{F_{cam}} \tag{4.16}$$

Also, because ϑ_r' corresponds to the edge of the CCD, the angle $d\vartheta_r$ is half the angular dimension of the CCD. If we indicate with $2L$ the CCD length, then:

$$d\vartheta_r = L/F_{cam} \tag{4.17}$$

Combining the last 3 equations together and indicating with $\Delta x_{edge} = F_{cam} d(\Delta\vartheta_r)$ the motion of the spectrum at the edge of the CCD relative to spectrum at the center, we obtain:

$$\Delta x_{edge} = \frac{L\Delta x}{F_{cam}} \tan\vartheta_r \tag{4.18}$$

For the ISIS Tex detector $L = 500$ pixels $(12mm)$ and almost always $\tan\vartheta_r < 1$. Therefore, even for flexure corrections of $\Delta x = 4$ pixels, the 'stretching' of the spectrum at the edge of the detector is very small:

$$\Delta x_{edge} < 0.1 pixel \tag{4.19}$$

This result could obviously be calculated more accurately by ray-tracing the entire spectrograph. Given the above result however, I considered that was not necessary. The specifications for the active collimator and its actuator are summarized in table 4.1.

Collimator Max Tilt	$\beta = \pm 20 arcsec$
Collimator Tilt Accuracy	$\Delta\beta = \pm 0.2 arcsec$
Collimator Radial Stability	$\Delta r = \pm 3 \mu m$
Actuator Max Expansion	$t = 15 \mu m$
Actuator Position Accuracy	$\Delta t = \pm 80 nm$

Table 4.1: Summary of the specifications in positioning the active collimator and of the corresponding specifications for the actuators.

4.4.3 The Environment on the Telescope

The active collimator is designed for operation on the ISIS spectrograph in La Palma, there it will be exposed not just to the weather conditions on the mountain, but also to the telescope motions and accelerations. Both these factors need to be taken into account in drawing the specifications for the ISAAC system.

4.4.3.1 Telescope Tracking

The telescope motion for an altazimuthal mount is determined by the spherical coordinates transformations that convert the equatorial coordinates of the object (declination δ and right ascension α) into altazimuthal coordinates (elevation E or zenith distance $z = 90^o - E$ and azimuth A). The two spherical coordinate systems are shown in fig. 4.12. The zenith distance can be easily related to α and δ given the sidereal time S_t and the latitude Φ of the telescope site. The equation is the following [105]:

$$\cos z = \sin\Phi \sin\delta + \cos\Phi \cos\delta \cos(S_t - \alpha) \tag{4.20}$$

The WHT telescope at La Palma has a latitude $\Phi = 28.8^o$. The telescope motion in elevation is shown in fig. 4.13 as a function of the hour angle $h = S_t - \alpha$ for a star having declination $\delta = +15^o$. The speed of tracking is clearly not constant and also depends on the declination of the target, but an estimate of the worst cases can be done by taking the steepest section of the graph. In this way we can estimate that the highest speed in elevation during tracking is around 15 degrees/hr (the average speed for that star would be only a bit less, around 12 degrees/hr). This telescope

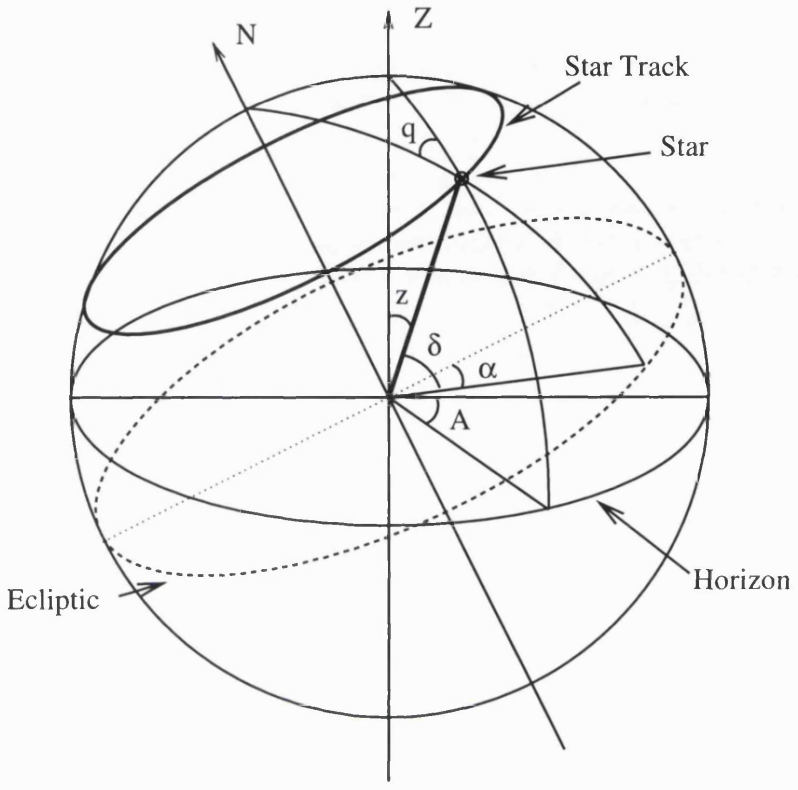


Figure 4.12: Relation between equatorial and altazimuthal coordinates systems. Both equatorial (declination δ, right ascension α) and altazimuthal (zenith distance z, azimuth A) coordinates of a star are shown. The parallactic angle q is the angle between the zenith (Z), the star and the celestial North pole (N).

motion is very slow and does not pose any severe constraint on the update rate for the flexure compensation system.

The azimuth motion is not relevant from the point of view of flexure, but the angle of rotation of the Cassegrain turntable ψ is. This motion compensates the rotation of the telescope's field of view typical of altazimuthal telescopes. When tracking a star across the sky the other stars in the field of view rotate about the center. While the stars in the field keep a fixed orientation with respect to the celestial North Pole, the north pole direction in the field rotates because the telescope system of reference always points at zenith. The field rotation is therefore determined by the parallactic angle q, which is the angle between the celestial North pole, the object's position in the sky and the zenith. If the turntable rotation equals the speed of the

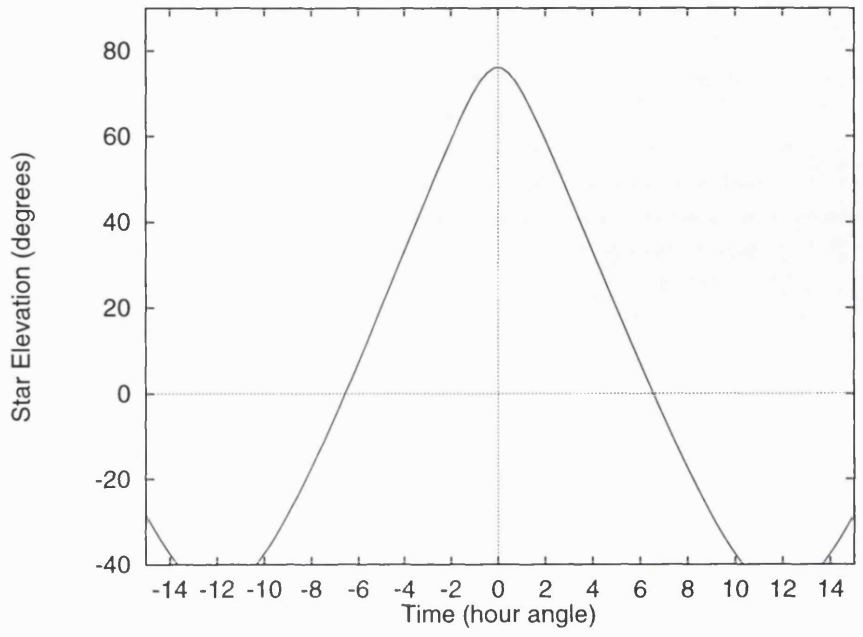


Figure 4.13: Elevation of a star having declination $\delta = +15^o$ as a function of the hour angle $h = S_t - \alpha$ when observed at the latitude of the La Palma Observatory ($\Phi = 28.8^o$). The graph shows the altitude of the telescope when tracking the star. The rotation of the Cassegrain turntable can be calculated in a similar way.

parallactic angle, this prevents field rotation. The equation for the parallactic angle is the following [105]:

$$\tan q = \frac{\cos\Phi \sin(S_t - \alpha)}{\sin\Phi\cos\delta - \cos\Phi\sin\delta\cos(S_t - \alpha)} \tag{4.21}$$

During normal astronomical observations, however, the Cassegrain rotator is not usually placed at the parallactic angle, even though it moves at the same rate. This is because the turntable position ψ determines the orientation of the slit when projected onto the sky. Usually the observation requires the slit to be placed at a given angle σ with respect to the North pole direction (required sky PA). Provided that the instrument slit is aligned with the zero position on the Cassegrain rotator, the turntable position is given by [106]:

$$\psi = \sigma - q \tag{4.22}$$

The required slit angle in the sky only acts as an offset, while the minus sign for the parallactic angle is a consequence of the way angles are measured on the sky

and on the turntable. It is possible to demonstrate that the motion is fast when observing close to the zenith and that it slows down when close to the horizon. This happens when in equation 4.21 the denominator in the fraction approaches zero and the parallactic angle suddenly changes by nearly 180^o. Fortunately when the telescope is close to the zenith flexure shifts are at their minimum, and this fast motion does not have much effect on the flexure compensation system. Away from zenith the turntable rotation is even slower than the elevation motion.

On the basis of the speed of the telescope motion, both in terms of elevation and rotator position, and of the extent of flexure shifts observed, I concluded that an update rate of once every minute was more than enough to keep the spectrum stability within specification. Even with the worst case flexure of $20\mu m/hr$ the image shift in a minute would be only $0.3\mu m$, well within our target. The update rate has also an effect on the lifetime requirement for the actuator: expecting the flexure compensation system to work once a minute, for 10 hours a night, means the actuator needs to work around $2 \cdot 10^5$ cycles/year.

Finally, all the equations for the telescope motions are slightly inaccurate because the positions are not corrected for precession, nutation, aberration and atmospheric refraction. The total error when we input the catalogue coordinates of the object is expected to be around 1^o [106]. This error is acceptable for flexure modelling purposes, however it could be drastically reduced if we input the apparent coordinates of the object (i.e. accounting for precession, nutation, aberration). The total error would be reduced to a few minutes of arc, mainly due to atmospheric refraction.

4.4.3.2 Meterological Conditions

The physical environment inside the telescope dome also affects the working condition for the flexure compensation system. The following are estimates of the mean temperature and humidity levels *inside* the dome of the telescope on the Roque de los Muchachos in La Palma [107]

1. Average humidity 30%

   Maximum humidity 80% (after which the telescope enclosure is closed).

2. Minimum temperatures during the night in winter:

Mean -5^oC Lowest -10^oC

Temperatures during the night in summer:

Maximum $+23^oC$ Minimum $+8^oC$

Thermal excursions during the night are less than 10^oC both in summer and in winter.

All the above values refers to conditions inside the telescope enclosure, but they are expected to be similar to those inside the spectrograph. It has to be considered though, that there is up to 1 KW dissipation at the Cassegrain focus where the spectrograph is situated. This can slightly alter the conditions inside the spectrograph, but the effect, if present, would only improve conditions, by reducing humidity and increasing the temperature.

The temperature range of operation that can be derived from the above data is roughly between -10^oC and $+30^oC$. This is a fairly ordinary range for most electronics and actuator system. Humidity levels can be a bit high, but inside the spectrograph I would expect them to be somewhat lower.

Chapter 5

Building the Instrument

5.1 Mechanical Design

The mechanical design of ISAAC was focused on the collimator cell, which is the active element of the system. The cell needed to be interchangeable with the existing one, but also to allow the tilting of the mirror without introducing any hysteresis in the movement. Achieving all this within the very tight volume limitation of the collimator area, presented a serious challenge in the design. The mirror was also reproduced as an exact replica of the existing one and was polished and figured entirely at OSL by D. Brooks. A picture of the collimator mirror after being polished and aluminized is shown in fig. 5.1. The final design of the cell structure was the product cooperation between myself and M. Dryburgh, who is a design engineer at OSL. The result is a design that integrates original drawing of the existing mirror cell with the needs and specifications of actuation. The cell design is shown in the drawings of fig. 5.2 and fig. 5.3 and it is discussed in detail in the next sections. The main challenge we encountered in producing the active collimator cell was the very limited space available, which explains the very compact design shown.

5.1.1 Collimator Actuation

The mirror is mounted on three actuators placed at the vertices of an isosceles triangle with a basis of $160mm$ and a height of $172mm$. I chose to use three actuators

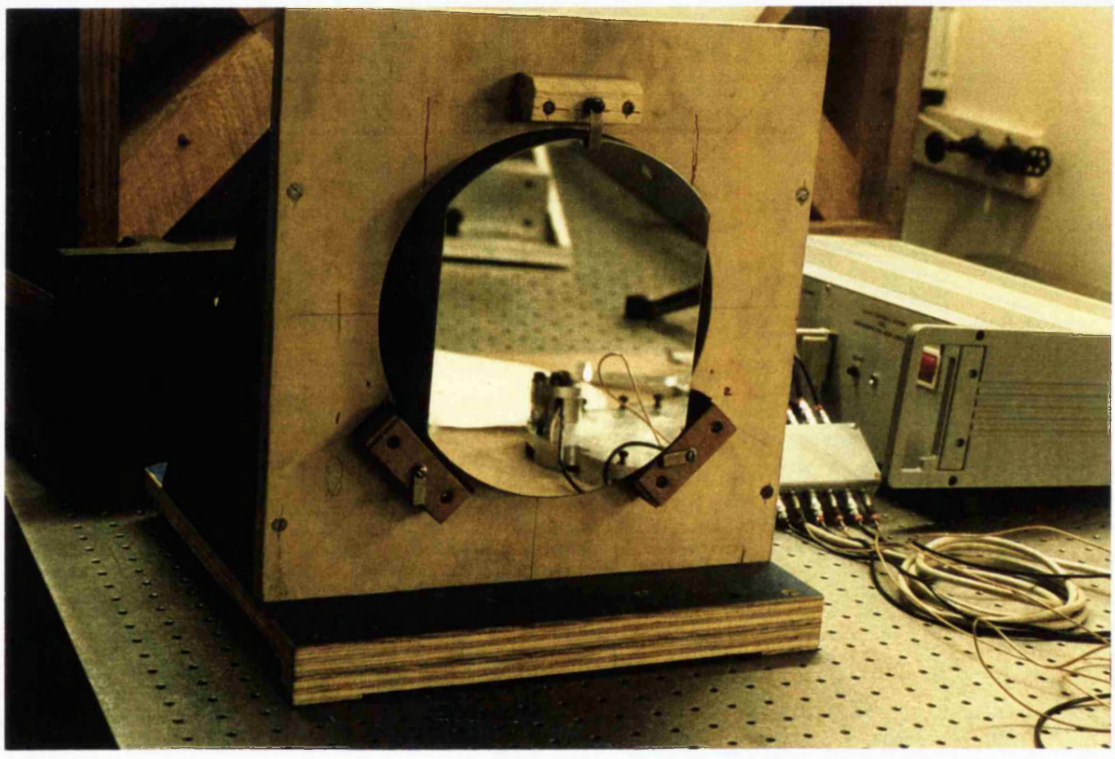


Figure 5.1: Picture of the collimator mirror outside the cell, after being polished, cut and aluminized.

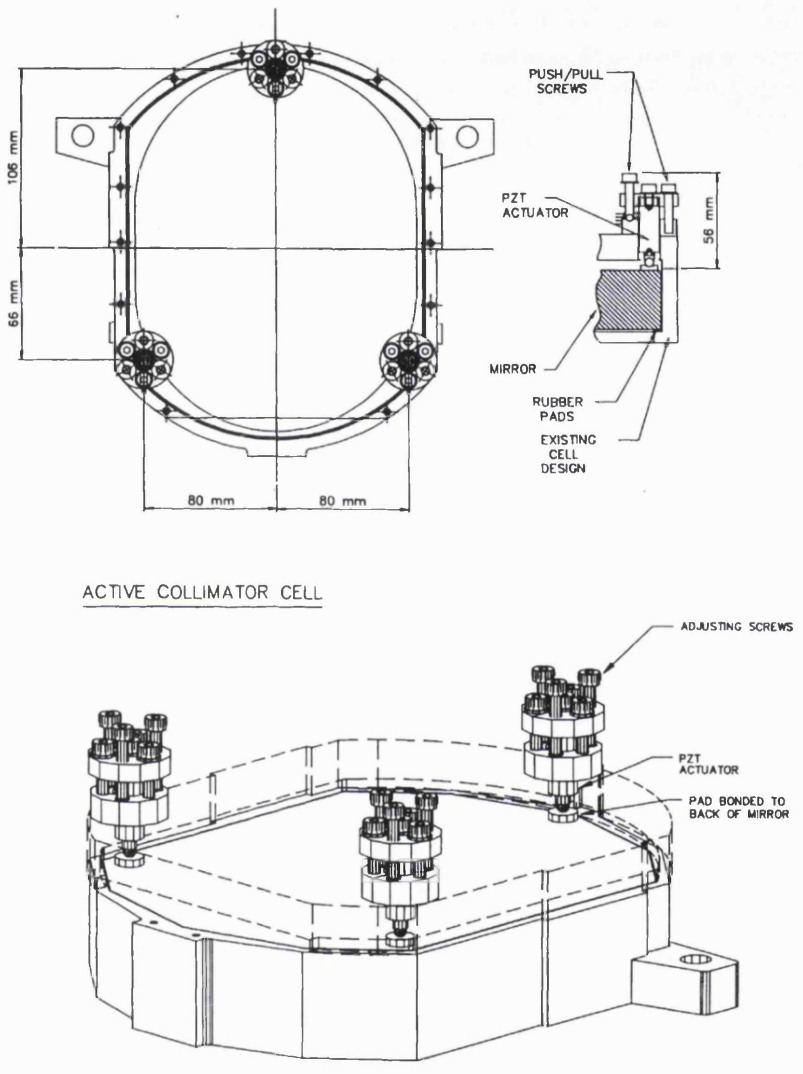


Figure 5.2: Drawing of the active collimator cell as seen from the back of the mirror. A detail of the actuator mounting system is also shown. The actuators are fitted with a spherical tip.

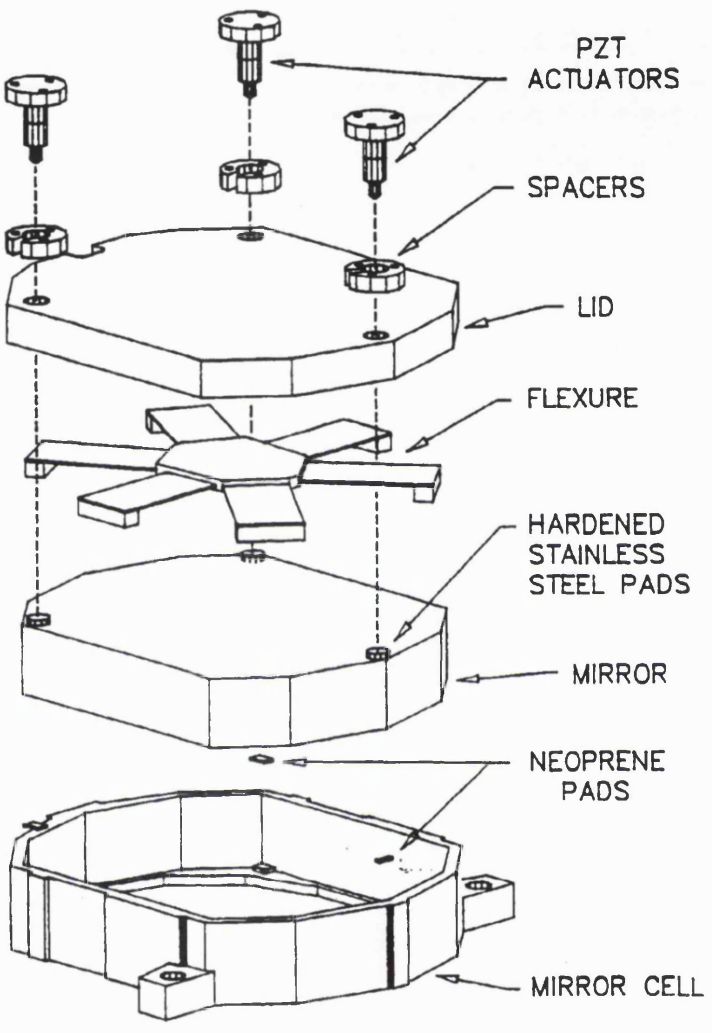


Figure 5.3: Exploded view of the active collimator. The mirror cell at the bottom is a replica of the existing one. The flexure and the lid at the back were added to allow for active control, the flexure provides radial support, while the lid houses the piezoelectric actuators, which support the mirror axially.

(instead of two or four) for two main reasons: three points identify a plane, so the collimator is automatically located once it is compressed against the hard stop of the actuators, and this position is not affected by temperature changes because the actuators have all the same thermal characteristics. With only two actuators and a fixed pivot, we would have the additional problem of matching its thermal expansion with that of the actuators. Moreover, the actuators experience different thermal expansion when operated with a position sensor and when simply switched off. Therefore the fixed pivot could never match the expansion of the actuators in all conditions. A practical result of using three actuators is also that the maximum tilting angle of the mirror is increased by a factor two, even if the actuator travel remains the same. The actuators are not equidistant because of the shape of the mirror cell itself, which is not circular. A picture of the active collimator cell is shown in fig. 5.4.

As actuator type I decided to use piezoelectric stacks with strain gauge position sensors for closed-loop positioning. The choice was to use one of Physik Instrumente (P.I.) products, the actuator PI-841.10, which is already equipped with a strain gauge sensor inside its metal housing. The actuator has a nominal maximum travel of $15\mu m$ and in closed loop operation, according to P.I., it will be possible to position the actuators with an accuracy of $75nm$ for the whole travel range and for temperature excursions up to $20°C$. In absence of temperature changes, the specified positioning accuracy is $30nm$, well beyond our specifications. The basic properties of this actuator are summarized in table 5.1.

The actuator requires an $100V$ DC supply, but at very low currents, so that power dissipation is practically negligible. The total length of the actuator is $32mm$ and can produce pushing forces up to $1000N$; it is preloaded, so it can also pull up to $100N$. The operating temperature range is from $-20°C$ to $+80°C$ and this also fits well within the requirements. The lifetime of the system is expected to be very long: P.I. actuators have been tested up to 10^9 cycles without any degradation in performance. Strain gauges are specified with a lifetime of 10^8 cycles if the expansion is limited to 0.1% of their working range, which is our case [108]. At our rate of operation of only $2 \cdot 10^5$ cycles/year, the lifetime issue becomes irrelevant. The only possible reason of concern is the the humidity inside the telescope dome;

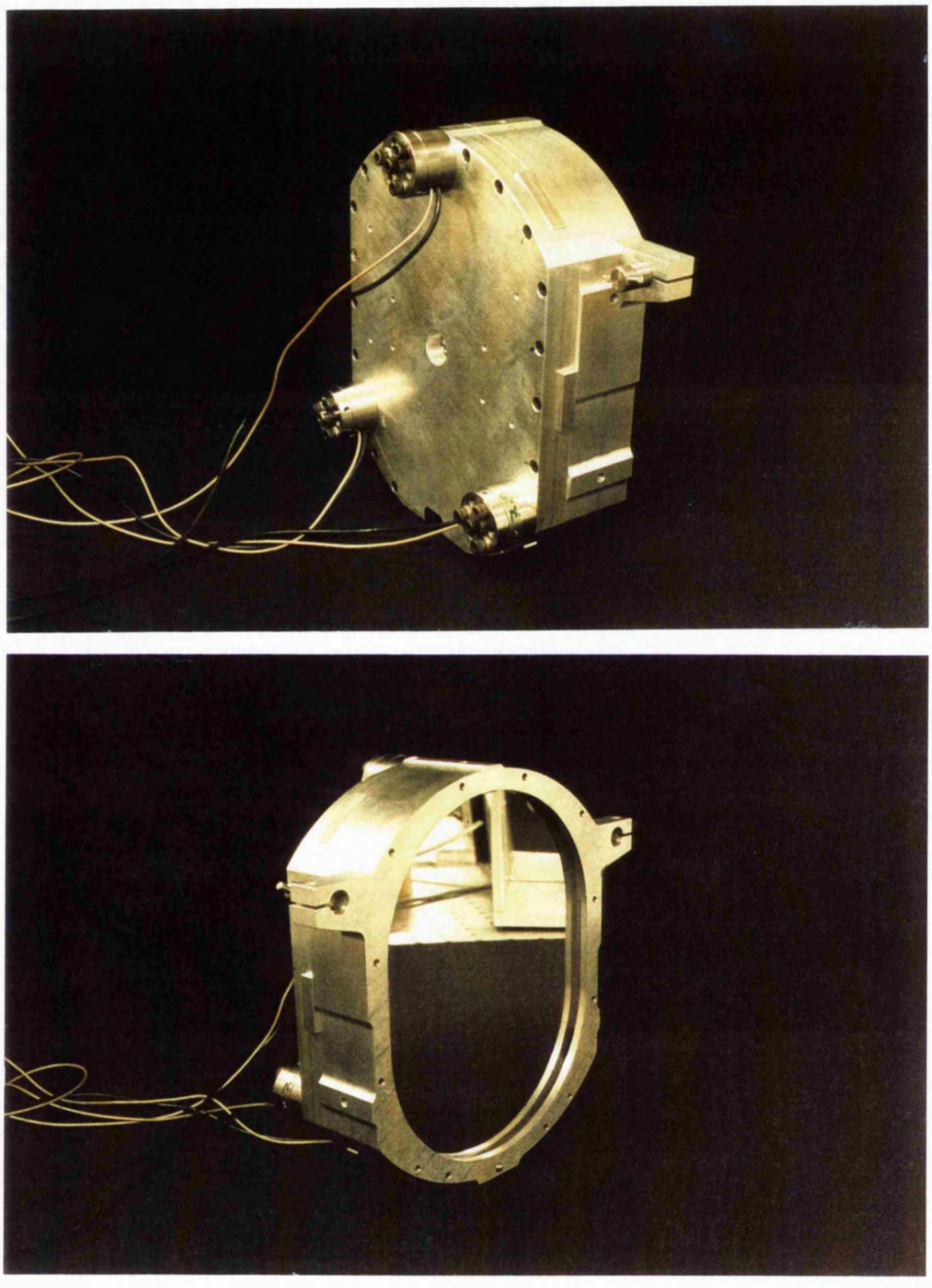


Figure 5.4: Picture of the collimator mirror in its cell.

	P.I. P-841.10
Max Travel	$15\mu m$
Nominal Accuracy	$30nm$
Pushing Force	$1000N$
Pulling Force	$100N$
Stiffness	$55N/\mu m$
Temp Range	$-20° + 80°C$
Temp Expansion	$0.1\mu m/K$
Max Frequency	$14KHz$
Input Voltage	-20 to $+120V$
Capacitance	$1.8\mu F$
Length	$32mm$

Table 5.1: Summary of the physical properties of the actuator P-841.10 from Physik Instrumente.

prolonged exposure to humidity above 60-70% can damage the electrical contacts of the actuators. However, considering that the telescope site is usually very dry and that the actuators are safely encased in the cell inside the spectrograph, this should still allow for a very long operational life.

The strain gauges sensors have a nominal accuracy of 0.2% of the travel, that is $30nm$ in our case; this is what determines the overall accuracy of the actuator. These values are in absence of temperature changes, therefore the system accuracy can be increased, if necessary, by the introduction of temperature sensors. A review of the options and of the reasons that convinced me to choose PZTs and this particular actuator is discussed in detail in chapter 3.

5.1.2 Mirror Axial Support

In our mechanical design the collimator cell is left unchanged apart from the additional surface at the back where the actuators are mounted. Each actuator pushes

the mirror against a small rubber pad on the front. The actuator/rubber pad clamping substitutes the diaphragm of the original design. The actuators are held from the bottom with a screw that connects them to a small disk which serves as main actuator support. This support disk is connected to the lid at the back of the cell with some disk spacers and three screws. The disk spacer between the actuator support and the cell determines the compression of the rubber pads on the other side and consequently the force applied to the mirror; the spacer also helps to keep the actuator orthogonal to the mirror surface. We had a set of spacers of different thickness with which it was possible to vary the clamping force until the required value was reached.

The axial support for the collimator is provided by the three actuators pushing against metal pads glued to the back of the mirror and by the corresponding rubber pads on the front. These pads act as a spring under compression and they are placed as close as possible to the (axial) line of action of the actuators, to minimize the torque moment applied to the mirror. The actuators are fitted with a spherical tip, to avoid the PZT suffering lateral (shear) forces which could damage the piezoelectric material. In this design, it is important that the compression force on the rubber pads is larger than the weight of the mirror, to avoid any movements of the collimator when the spectrograph moves.

5.1.3 Mirror Radial Support

The radial support of the mirror was a more complicated issue because the system used in the original design proved to be impractical. The previous collimator was supported laterally with several screws pushing on the sides of the mirror. The pressure applied was carefully controlled to avoid introducing stresses on the collimator. This system however, is designed to keep the mirror firmly in position and restricts the freedom of movement in axial direction as well. My original idea was that the tilt of the collimator was so small that this radial support could still work. Early tests instead showed that the lateral screws would not prevent the tilting, but they were a major source of hysteresis in the movement. Clearly the actuators were strong enough to move the mirror against the resistance of the screws, but the friction in the movement produced hysteresis.

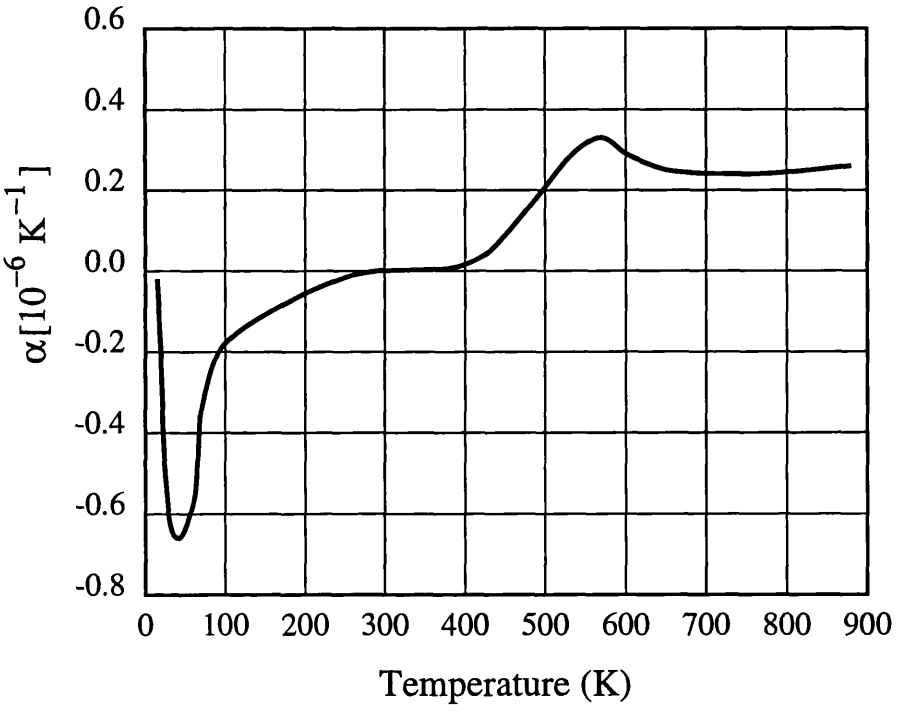


Figure 5.5: Linear thermal expansion coefficient of Zerodur (from Schott GmbH technical brochure). The material has a nearly zero expansion coefficient between $0°C$ and $100°C$.

The radial support had to be redesigned from scratch, and it had to fit in the very tight space limits inside the cell. We abandoned the idea of holding the mirror from the sides and decided to use a flexural element connected to the back of the collimator. The flexure would be strong in radial direction and weak in axial direction to allow the tilting and reduce hysteresis. The problem was how to connect the flexure to the mirror, maintaining high radial stiffness and avoiding stresses due to differential thermal expansion in the cell, the flexure and the collimator. The mirror is made of Zerodur, which is a glass-ceramic with an extremely low linear thermal expansion coefficient. The coefficient varies with temperature, but, as it can be seen from the data provided by the manufacturer (fig. 5.5) [109], around room temperature $\alpha_Z < 10^{-8}K^{-1}$.

The flexure instead, is made of aluminium to match the thermal expansion of the cell and has therefore a relatively large thermal expansion coefficient: $\alpha_{Al} = 23 \cdot 10^{-6}K^{-1}$ [110]. The usual solution in these cases is to use a flexible adhesive, which will easily

adsorb the stresses between the two surfaces. This is not an option for this project because of the stringent requirements in radial position accuracy. It was decided to have a high precision Invar insert in the mirror to which the flexure could be screwed on. Invar was chosen because it has a much smaller thermal expansion coefficient than aluminium and matches more closely the expansion of Zerodur: $\alpha_{Inv} \leq 1.5 \cdot 10^{-6} K^{-1}$. The insert was machined very accurately to fit exactly in the hole in the mirror: we estimated the gap between the two to be less than $0.5 \mu m$. In this way no adhesive between the Invar and the Zerodur mirror was required. The diameter of the insert can also be reduced to minimize the stresses induced in the mirror. A diameter of $10mm$ provides enough stability without compromising the mirror figure. For $30°C$ temperature changes the insert will expand by about $0.5 \mu m$ with respect to the Zerodur mirror and this does not affect the reflecting surface of the mirror. The components of the active collimator cell are shown in fig. 5.6. Extensive testing was necessary to prove the correctness of the design and of these theoretical calculations: the tests are described in chapter 6. The solution was then to provide radial support by a flexural element connected to the cell through six thin metal strips and bolted with a screw to a small Invar cylinder inserted in the back of the mirror. As explained earlier, the Invar insert was not cemented to the mirror because the tight mechanical fit proved to give enough accuracy in the radial direction.

5.2 Electronics

Most of the electronics for the active collimator consists in the drive and the control for the actuators. Only a small amount of electronics is needed for the computer system and the connection with the telescope and the control room. All the electronic circuits were designed by G. Nixon and H. Jamshidi and manufactured at OSL by myself. The decision to make the electronics ourselves was taken because the very low frequency of operation for the actuators would allow us to greatly simplify the circuitry. Every electronic component was first developed on a prototype board and then converted into a standard Eurocard PCB format, to fit into a 19 inch electronic rack. Our electronics is totally independent from the one controlling the telescope and the spectrograph.

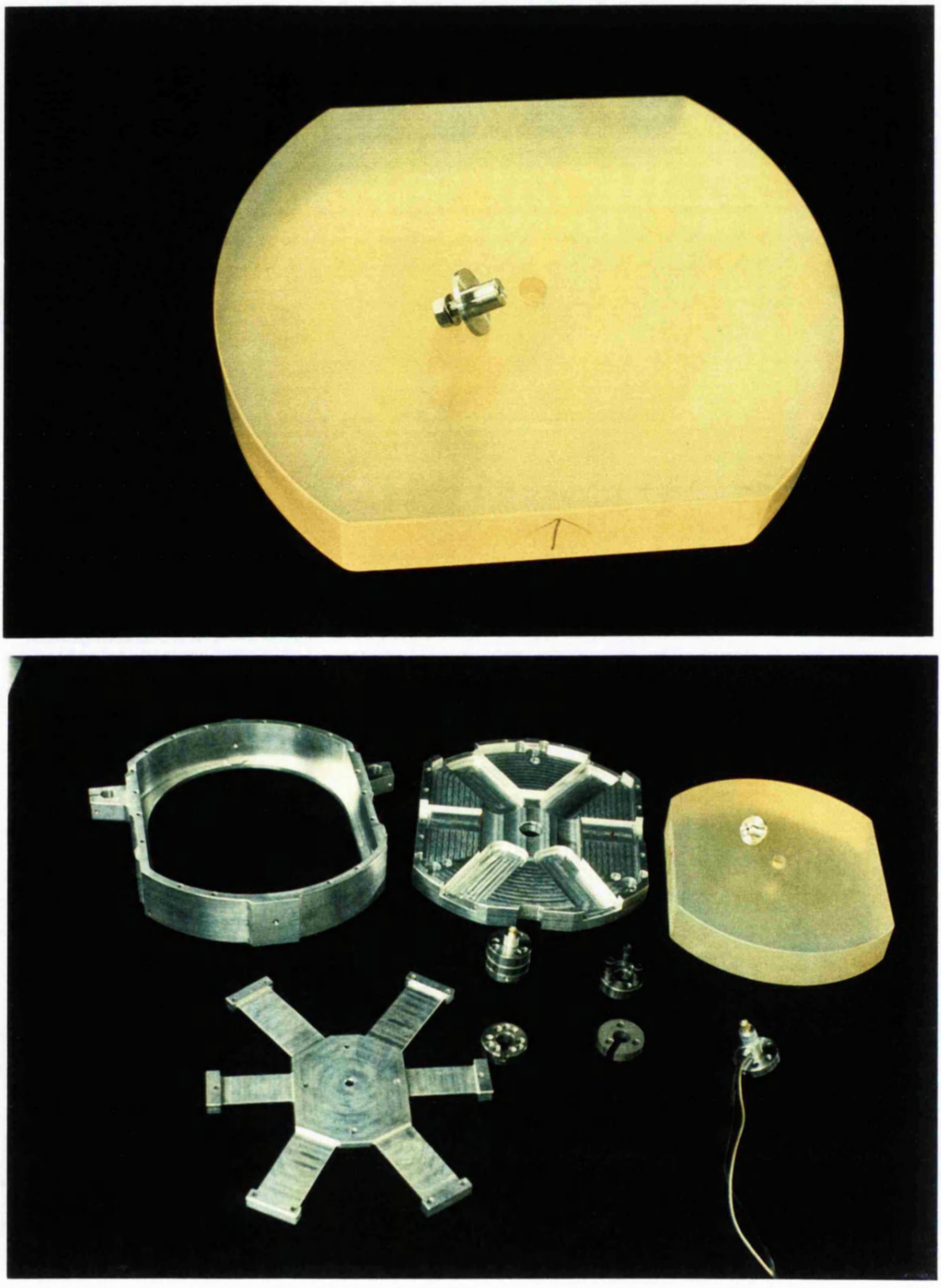


Figure 5.6: Picture of the components of the active collimator cell showing the flexure and the metal insert for radial support.

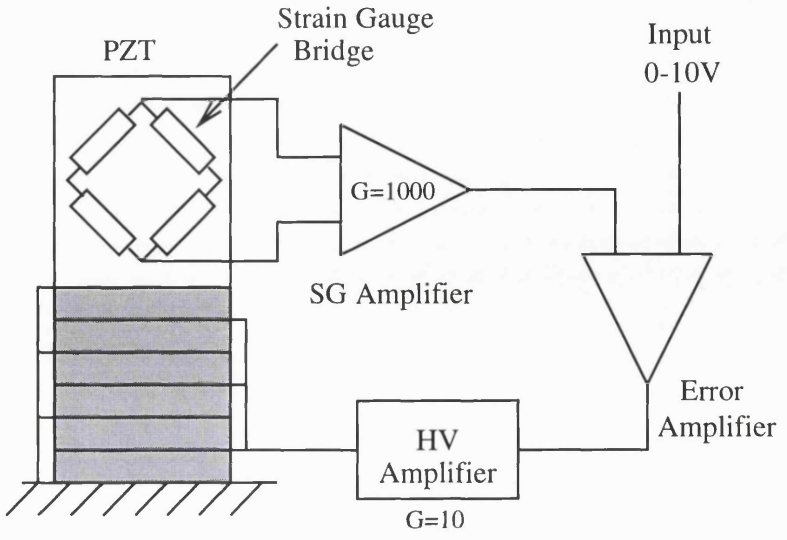


Figure 5.7: Closed loop control of the actuators with strain gauge sensors. The error amplifier assures that the output from the strain gauge sensor matches the input from the DAC by adjusting the input to the high voltage amplifier that drives the actuators.

5.2.1 Closed Loop Actuator Control

The actuator expansion is controlled by a negative feedback loop including the strain gauge sensors for high positioning accuracy. The feedback loop ensures that the actuator actually reaches the position entered by the user (or by the control computer). The schematic electronic diagram of the feedback loop is shown in fig. 5.7. The circuit works on the same principle of any other negative feedback loop: the control processor inputs a given actuator voltage through the DAC interface (in our case the input is between 0 and 10 Volts). This voltage is amplified by the High Voltage amplifier by a factor 10 so that the output range of the amplifier matches the input voltage of the PZT (0-100 Volts). When the actuator expands, some residual errors may be present (e.g. hysteresis) so that it actually does not reach the desired position. This positioning error is sensed by the strain gauge which, when properly calibrated, will differ slightly from the input coming out of the DAC. The error amplifier then compares the signal from the sensors with the one at the input and corrects the voltage applied to the actuators. The closed loop ensures that this process is automatic and neither the user nor the control computer need

to intervene to correct the position. All the important electronic components are shown in fig. 5.7: in the next sections I will analyze each one in detail. The circuitry for the actuator drive was designed by G. Nixon at OSL, after consultation with myself to determine the required performance.

5.2.2 High Voltage Supply

The high voltage supply provides the $100V$ input for the amplifier to drive the actuators. The characteristics of the supply are determined by the maximum slewing rate Σ, which is defined as the speed at which the voltage needs to be changed, that is:

$$\Sigma = \frac{dV}{dt} \tag{5.1}$$

The maximum slewing rate then drives the maximum output current that the supply has to provide to charge up the actuator, which is essentially a large capacitor with electric capacitance C, according to:

$$i_{max} = C\Sigma_{max} \tag{5.2}$$

In specifying the characteristics of the the supply I took the conservative value of $1/10$ of a second as the maximum time needed to charge the actuators to $100V$; this implies a slewing rate of $\Sigma_{max} = 1000V/sec$. Given the actuator capacitance $C = 1.8\mu F$, the maximum current required is still only $i_{max} = 1.8mA$. The reduced power requirement for the supply was used to improve the stability of the output, which is instead very important. The solution was to increase the voltage output of a standard low power ($5W$), low voltage ($\pm 15V$) DC power supply. The circuit diagram is shown in fig. 5.8.

The high voltage supply is based on an oscillating circuit that feeds a transformer with an alternated current. The oscillator is made out of two inverters contained in a 74HCT04E integrated circuit and works at $50KHz$. The two outputs of the oscillator are square wave signals inverted between each other and they are sent through two transistors to the transformer. The transformer primary winding has 15 turns between each output and the $+15V$ supply. The alternate current in the primary winding induces a current in the secondary which is rectified into a DC output by the full bridge of diodes of the secondary circuit. The voltages across the

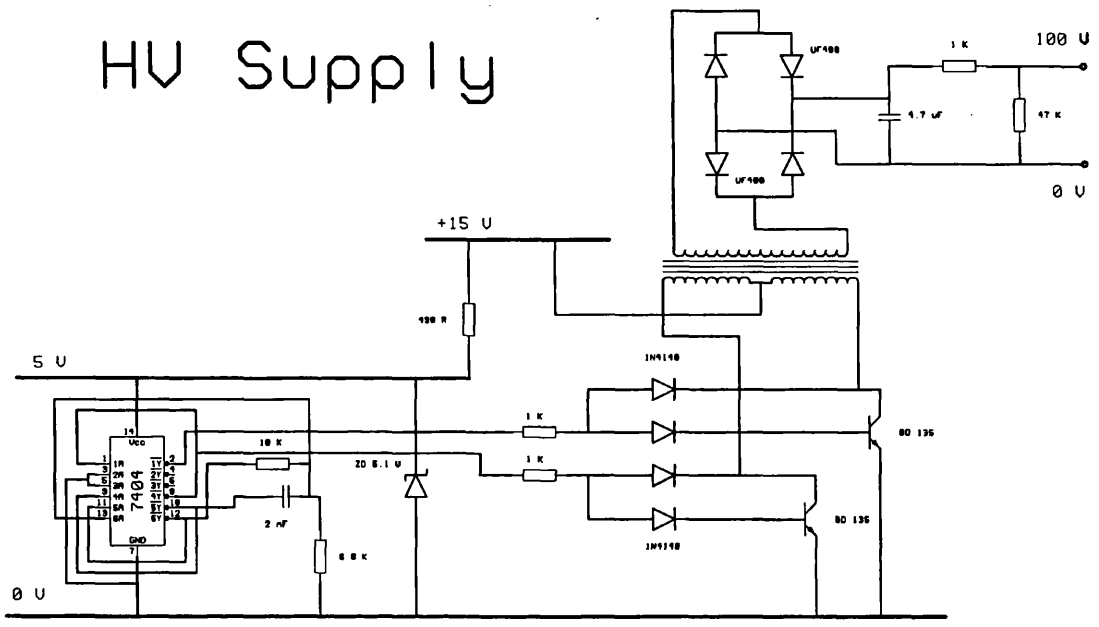


Figure 5.8: Circuit diagram for the HV Supply.

primary and secondary windings are directly proportional to the number of turns in the winding according to:

$$\frac{V_s}{V_p} = \frac{n_s}{n_p} \tag{5.3}$$

The number of turns in the secondary winding, originally 100, was increased to 118 to account for the high load of driving two actuators simultaneously. The supply is fairly insensitive to the load, producing $109V$ when all the three amplifiers have zero output (minimum load), reduced to $98V$ when two amplifiers have maximum output (maximum load). A large resistor was put in parallel with the output to prevent the voltage from raising too much when the supply is working at zero load. This resistor fixes the the working range of the supply in the flat part of the Voltage vs Load curve, so that the output is less sensitive to the load.

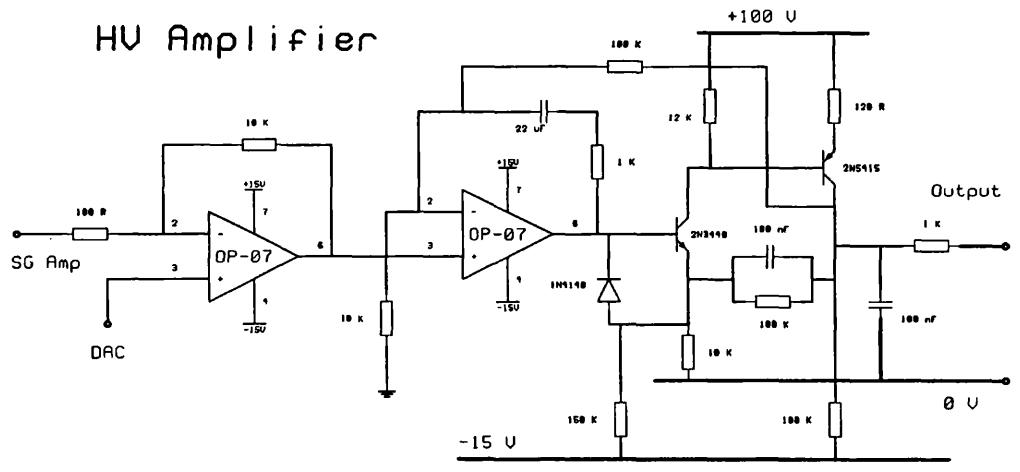


Figure 5.9: Circuit diagram for the HV amplifier and error amplifier. The comparator receives inputs from the Strain Gauge amplifier and the DAC which range between 0 and 10 V.

5.2.3 High Voltage Amplifier and Comparator

The HV amplifier was designed to have a gain $G = 11$, producing an output between 0 and 110 V with an input between 0 and 10 V. Again output stability and linearity was preferred to frequency bandwidth as the parameter to optimize. The circuit diagram is shown in figure 5.9. The amplifier showed excellent linearity up to the limit of the HV power supply, as fig. 5.10 shows. The data in the graph was taken when the HV supply had only 100 turns in the secondary winding of the transformer and could provide only $80V$ output; as I explained earlier, the output was later increased to $100V$. The capacitors in the circuit were introduced to prevent the output from oscillating when operating the actuators in closed loop and to reduce the high frequency noise in the output. The large $22\mu F$ capacitor and the 1 $K\Omega$ resistor provide a local feedback at the operational amplifier for the high frequency components. To protect the PZT actuators from damage a resistor with $R = 1$ $K\Omega$ is placed in series with the output. Since the PZTs are essentially large capacitors,

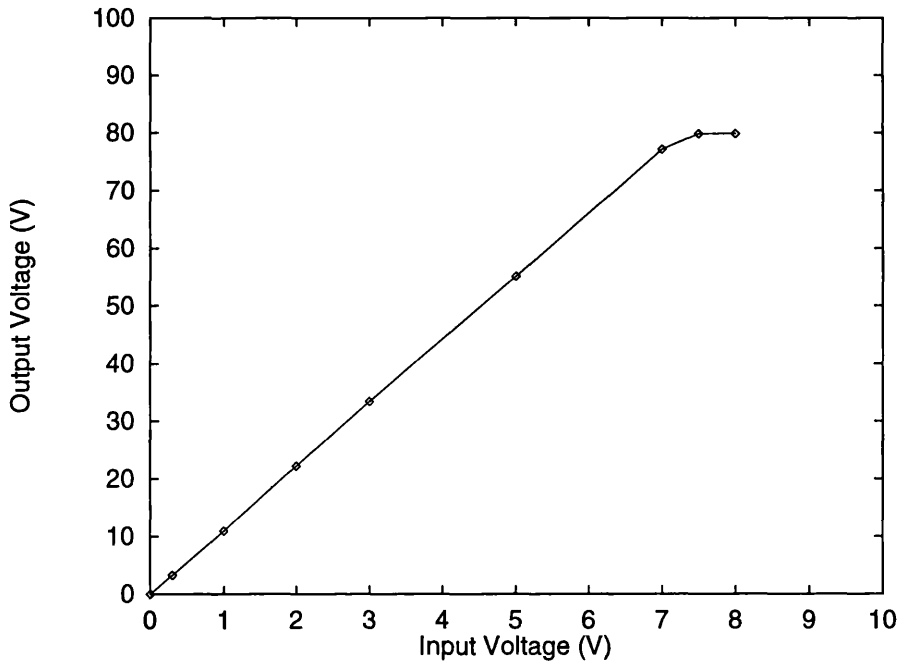


Figure 5.10: Linearity test for the HV amplifier. The curve shows the gain $G = 11$ of the amplifier and its excellent linearity up to the limits of the voltage supply (when the amplifier was tested the supply could produce only $\sim 80V$).

shortening the inputs while the actuator is charged would produce a rapid discharge and a consequent contraction shock that may destroy the device. The resistor slows down this process by providing a load on which the actuator can discharge; in this case the time constant is $\tau = RC \simeq 2ms$.

The comparator consists simply of an operational amplifier with a local feedback having a gain of 100. It was necessary to reduce the gain of the comparator to prevent saturation and avoid oscillation in closed loop operation. In the end a low-frequency ($100Hz$) oscillation still remains, but its amplitude at the output of the HV amplifier is small enough ($0.2V$) to be neglected. The required accuracy of about 0.5% in positioning the actuators ($0.08\mu m$ over $15\mu m$ total travel) translates into 0.5 V at the amplifier output.

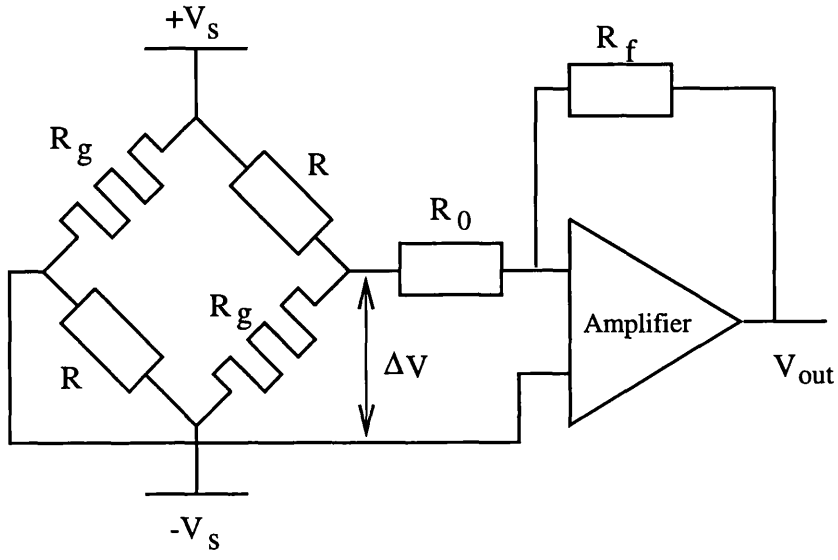


Figure 5.11: Schematic circuit diagram for the strain gauge bridge and amplifier. The circuit is a simplified version of the real one and explains the principles on which the strain gauge circuitry works. The voltage difference ΔV is introduced when the actuator expands and the strain gauge resistance changes.

5.2.4 Strain Gauge Bridge and Amplifier

The actuators I used have an already built-in position sensor; the sensor consists of two strain gauges in a Wheatstone bridge configuration. The two gauges are cemented directly on the piezoelectric stack inside the metal housing, while the two other resistors that complete the bridge need to be added to the control electronics for the actuators. A schematic diagram of the strain gauge bridge and amplifier is shown in fig. 5.11. The real circuit is significantly more complex than the one shown, because it includes a low noise, high stability amplifier specially designed for driving strain gauge bridges bought from RS Components Ltd.

The strain gauge is essentially a resistor that changes its resistance R when a strain $S = \Delta L/L$ is applied to it (see chapter 3), so that:

$$\frac{\Delta R}{R} = K_G \frac{\Delta L}{L} \tag{5.4}$$

Here K_G is the gauge factor, which in our case has a value $K_G \simeq 2$ [112]. If we consider the bridge configuration of fig. 5.11, we can calculate the output voltage of the bridge ΔV as a function of the strain (or of the actuator expansion, which is

the same thing). For the bridge to be in equilibrium the current in both arms (A and B) needs to be the same, thus:

$$I_A = I_B = \frac{2V_s}{R_G + R} \tag{5.5}$$

The voltage drop in A is $V_A = 2V_s - R_G I_A$ while $V_B = 2V_s - RI_B$, therefore:

$$\Delta V = V_B - V_A = 2V_s \frac{R_G - R}{R_G + R} \tag{5.6}$$

Considering that at zero strain the resistance of the strain gauge has to match that of the external resistor R we can write $R_G = R + \Delta R$, where $\Delta R \ll R$ represent the change in resistance due to actuator expansion. The previous equation then simplifies to:

$$\Delta V = 2V_s \frac{\Delta R}{2R + \Delta R} \simeq V_s \frac{\Delta R}{R} \tag{5.7}$$

Or, in terms of the strain:

$$\Delta V = V_s K_G \frac{\Delta L}{L} \tag{5.8}$$

Equation 5.8 shows the principle on which a strain gauge sensor works: when the actuator expands, the bridge gives a voltage output proportional to the movement. In practice the circuit we used provides both the voltage supply for the strain gauge bridge and the amplifier with an adjustable output and gain. the bridge supply was set to $V_s = \pm 5V$ and the gain to $G = 1000$; the output of the amplifier could then be adjusted so that it ranged between 0 and 10 V to match the input to the comparator from the DAC. Since the output of the strain gauges depends on temperature, by adjusting the zero offset we set the temperature at which the actuators have zero expansion when driven in closed loop. The two strain gauges attached to the PZT have a resistance of 700Ω, and the bridge was completed using two high precision resistors of the same value; the resistors were chosen because of their low thermal coefficient.

5.2.5 Digital to Analog Converter

I already mentioned in the previous sections that the input signal for the actuator feedback loop is provided by the computer via a Digital to Analog Converter (DAC). The accuracy for the DAC is determined by the need to match the best possible performance of the positioning system. Since the strain gauge sensors have a nominal

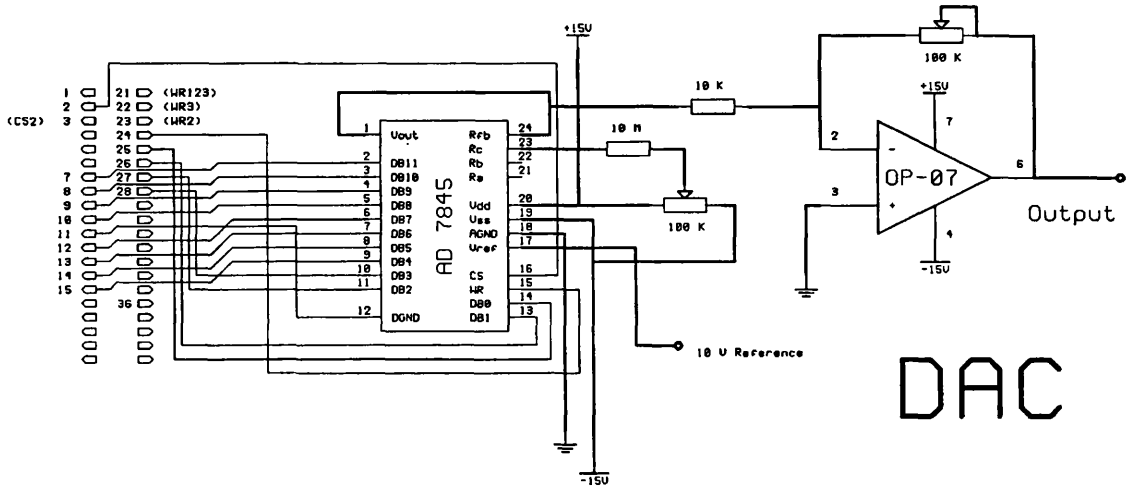


Figure 5.12: Circuit diagram for the Digital to Analog Converter (DAC).

accuracy of 0.2% or 1/500, a 10-bit or even a 9-bit DAC would be enough. In practice however, I decided to use some standard 12-bit converters for the DAC card. The circuit, which includes a 40-way connection with the instrument processor, is shown in fig. 5.12 and was designed by H. Jamshidi at OSL.

5.2.6 Temperature Effects

All the electronics was designed to work at temperatures well below $-10°C$, which is the specified minimum operating temperature for the instrument. However, it is also important that the performance of the electronic components remains fairly insensitive to temperature changes and, in particular, that the accuracy in positioning through the actuator feedback loop is still within specification. The piezoelectric material of the actuators has a relatively large linear thermal expansion coefficient $\alpha = 3 \cdot 10^{-6} K^{-1}$ [111], therefore an actuator $3cm$ long expands at a rate of $0.1 \mu m/K$ when temperature changes. However, when used in closed loop operation with strain gauge sensors, this expansion is sensed and corrected, so that no position error is introduced.

Still, the other components in the control loop do suffer from temperature drifts which are not sensed and produce real position errors. In absence of temperature measurements, these errors cannot be corrected. The components involved are the strain gauge sensors, the strain gauge bridge and amplifier and the error amplifier. The temperature drifts in the strain gauge are usually expressed in terms of the temperature coefficient of the gauge factor, which in our case has the following value [112]:

$$\frac{1}{K_G}\frac{dK_G}{dT} = 1.04 \pm 0.1 \cdot 10^{-4} K^{-1} \tag{5.9}$$

We can see that this is a very small drift when compared with our accuracy requirement of $80nm$ over $15\mu m$ (one part in $5 \cdot 10^{-3}$). The stain gauge amplifier was specifically chosen for its high thermal stability: the maximum input voltage offset drift (as quoted in the RS catalogue) is only $1\mu V/K$. However, since the amplifier receives very small signals from the strain gauges ($10mV$ max), the resulting thermal drift, as a fraction of the full scale value, is of the same order of the strain gauge: $1 \cdot 10^{-4} K^{-1}$. Finally, the comparator has typical voltage drifts of $20\mu V/K$, which translate into a fractional contribution of $2 \cdot 10^{-6} K^{-1}$. This value is negligible when compared with the effects in the strain gauge sensors and amplifier. The overall thermal drift in the loop is therefore:

$$\alpha_{loop} \simeq 2 \cdot 10^{-4} K^{-1} \tag{5.10}$$

The corresponding drift in the actuator position due to temperature changes is $3nm/K$, which means that the maximum error for $20K$ temperature difference is only $60nm$, well within the requirements.

5.3 Controls and Communications

The instrument hardware consists of three separate subsystems: the mirror cell with the driving actuators is connected through coaxial cables to the actuator control electronics which sits in an instrumentation rack; this in turn communicates via a serial link with a standard lap-top 486 PC. The whole system is shown schematically in fig. 5.13. The main control software runs on the lap-top PC; the user is required to input on the keyboard the parameters necessary for starting active compensation.

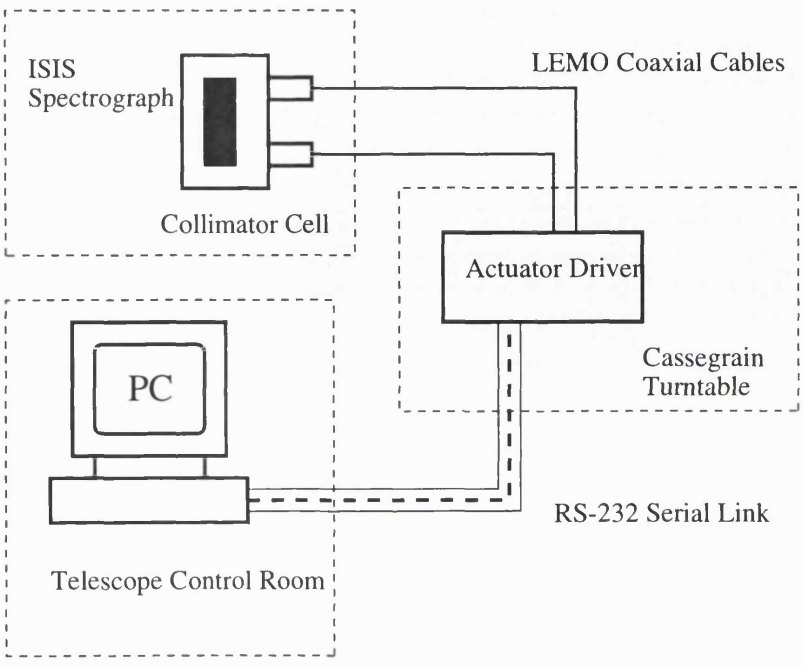


Figure 5.13: Schematic view of ISAAC hierarchical control system and of the routes of communication. The user interfaces with the PC where the main software runs; the local processor is only responsible for driving the actuators.

These include the object's equatorial coordinates, the sidereal time at the beginning of the exposure and the slit position angle on the turntable. The computer performs all the necessary calculations to convert these data into three actuator voltages and sends this information down the serial link to the local processor on the spectrograph. The processor confirms receiving the signal and feeds this to the DAC which generates three voltages for the actuator driver.

The actuator control has already been described in the previous sections. The actuators and the strain gauges sensors are part of a negative feedback circuit which guarantees accurate positioning. Six LEMO coaxial cables carry the signal for the 3 actuators and the 3 sensors to and from the electronic rack with the voltage supplies and the amplifiers. It is particularly important that the signal from the sensors is carried on coaxial cables to minimize sources of disruption: these signals are extremely small, only a few mV.

The local processor is a Flight 86 control board and it is based on a 8086 Intel processor. The board has two parallel I/O ports for communication with the digital

to analog converter and a RS-232 serial port through which it is connected to the control PC. This serial link allows remote operation of ISAAC from the telescope control room by using a spare serial connection in the telescope cable wrap. The local processor has indeed a minor role in the system control and very little processing is required from it. Its presence is however necessary to check the correct flow of information from the PC to the actuators. Moreover, when the system was being built, it was still not clear which share of the control software should go on the local processor and the processor chosen gave us the necessary flexibility. The lap-top PC was the final choice for the user interface because it was decided that the priority was that ISAAC should be a completely stand alone system.

5.4 Software

The main task for the instrument control system is to drive the three actuators on which the collimator mirror is mounted according to the orientation in space of the telescope and the spectrograph. This position, as pointed out earlier, is identified by two parameters: the zenith distance z of the telescope and the rotation on the Cassegrain turntable ψ (which is linked to the parallactic angle of the observed object). The voltages for the actuators can be determined when these parameters are known by using a look-up table with the stored flexure information. The control software therefore, needs to know the values of these two parameters at any time during the astronomical observation, to translate this information into voltages for the actuators and to send the voltages to the active cell through a digital to analog converter. Since the ISAAC experiment was designed to be completely stand-alone, I wrote the software in "C" to run on a 486 lap-top PC. In the final operating version, the software will have to run on the existing computer network at the telescope and to be fully integrated with the software which is already there.

In writing the software I allowed for a testing mode of operation to facilitate the calibration of the instrument and possible modifications or failure checks. In closed loop operation instead, the active collimator will be operated continuously without need of interaction with the user except at the start of the exposure. A schematic diagram of the software structure is shown in fig. 5.14. I wrote most of the software

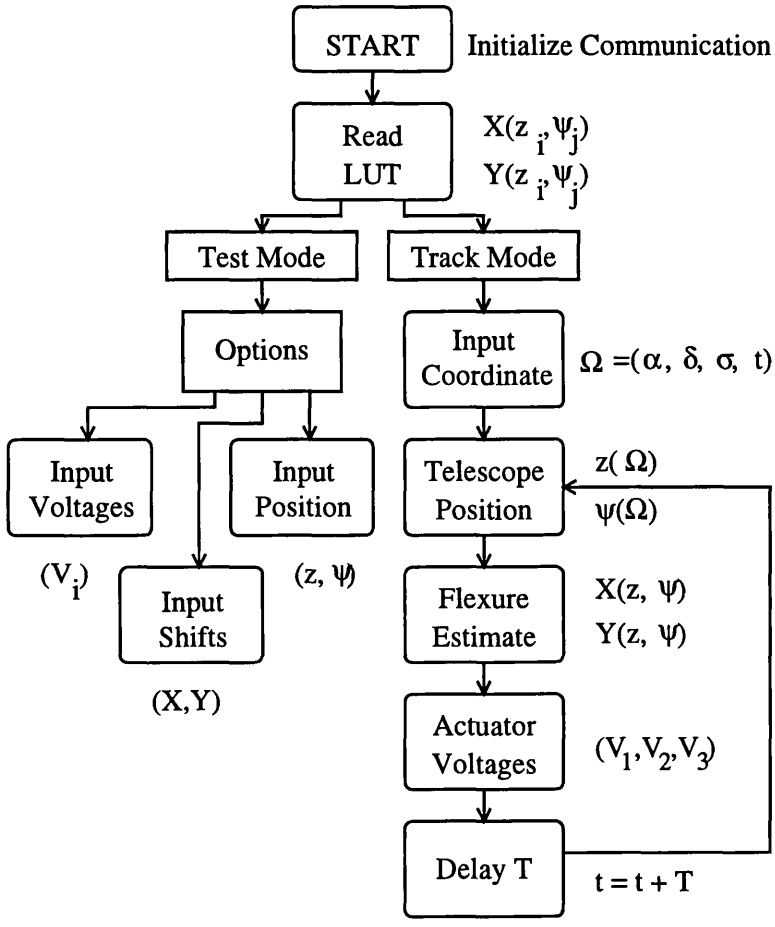


Figure 5.14: ISAAC Software diagram. The program, which runs under DOS on a standard PC, has two modes of operation: engineering (Test Mode) and automatic compensation (Track Mode). In the tracking mode, the user inputs the object's coordinates (Ω) and the computer estimates the flexure shifts on the detector (X, Y). This information is used to drive the active collimator by applying voltages to the actuators (V_i) at constant intervals (T).

myself, but some communication routines and the software for the local processor were written by H. Jamshidi at OSL.

When it starts, the control software initializes the serial link for communication with the local processor on the spectrograph. The initialization routine sends a string of characters to the Flight 86 and receives confirmation that the status of the serial link and the local processor is correct. Then it sets all the actuator voltages to zero by resetting the output ports to the DAC on the Flight board. Once the system is

initialized, the software reads some free calibration parameters from a file and the flexure data from a separate file; both files are stored in the memory of the lap-top PC. The look-up table contains the X and Y shifts at the detector (measured in pixel units) for a set of telescope and spectrograph positions. The data is stored in two matrices relating the position coordinates (z, ψ) to the absolute shifts in the plane of the detector (X, Y):

$$X_{i,j} = X(z_i, \psi_j) \tag{5.11}$$

$$Y_{i,j} = Y(z_i, \psi_j) \tag{5.12}$$

The zenith distance ranges between 0^o (telescope pointing at zenith) and 75^o sampled at intervals of 15^o; the turntable rotation covers 360^o with steps of 45^o. The X and Y shifts are converted into microns by multiplying the data by the detector pixel size ($22.5\mu m$ for the EEV or $24.0\mu m$ for the Tek CCD). The program displays a Main Menu where it is possible to choose between tracking and engineering mode of operation.

5.4.1 The Engineering Mode

This testing mode allows manual operation of the collimator, either remotely or through an engineering interface at the spectrograph. First the initial position of the telescope is requested, then three different options are provided:

- Voltages. The actuator voltages are given directly as an input.
- Detector Shifts. The user inputs relative shifts at the detector with respect to the starting point. The software converts these into actuator voltages without using the stored flexure data.
- Telescope Position. A new telescope position is given and the collimator is moved so as to compensate the flexure shift relative to the original starting point.

This mode facilitates the calibration of the instrument and any possible engineering work.

5.4.2 The Tracking Mode

In the tracking mode the program calculates the telescope and spectrograph position from the coordinates of the observed object and it does not need any input from the user after the exposure started. The tasks the software performs can be summarized as follows:

- Start. Initialization of communication links.
- Read LUT. The data stored in the flexure look-up table is loaded into memory.
- Input Object Coordinates. The user inputs the object's Right Ascension and Declination (α, δ) in degrees (hours), minutes and seconds.
- Input Sidereal Time. The sidereal time is given directly; soon afterwards the program starts the computer timer to compute the elapsed time.
- Input Slit Position. The desired slit orientation in the sky is given as an angle σ on the Cassegrain rotator; zero means the turntable follows the parallactic angle.
- Calculate Initial Position. Calculates the initial spectrum shifts on the detector (with respect to zenith) $X(0), Y(0)$ and then starts the closed loop operation.
- Calculate Elapsed Time. Reads the timer to update the sidereal time.
- Calculate Telescope Position. Using spherical trigonometry formulae determines zenith distance and turntable angle (z, ψ) from the object's coordinates (α, δ, S_t).
- Calculate Detector Shifts. Interpolates the LUT data to determine the absolute position at the detector $X(t), Y(t)$ at that time. Then the displacements relative to the starting position are calculated.
- Calculate Actuator Voltages. Using the optical and geometrical characteristics of the spectrograph and the expansion vs voltage relation of the actuators, the relative shifts are converted into three voltages $(V1, V2, V3)$.

- Write/Read Serial Port. These three data are coded and sent through the serial link to the local processor, which in turn acknowledges the reception of the signal by sending a reply.
- Delay. A delay of around one minute is introduced.
- Escape. The loop is closed by repeating the last 6 points indefinitely. When the program is running it is however possible to pause the operation or to escape from the loop from the user interface.

5.4.3 From Object's Position to Actuator Voltages

The tasks of the software for the active collimator mainly consist in performing changes in coordinates systems. The orientation in space of the spectrograph in the altazimuthal system is calculated from the object's equatorial coordinates using spherical coordinate transformations. This position is translated into displacements (X,Y) in the plane of the detector through the flexure map stored in the look-up table. The detector shifts are in turn converted into tilting of the collimator plane (which I will call the (j,k) plane) with respect to the optical axis. The angle of tilting is finally translated into voltages (expansions) for the actuators using the geometry of the cell and the collimator support.

5.4.3.1 Calculating Spectrum Shifts

The program starts with the equatorial coordinates of the object, the slit position and the sidereal time $(\alpha, \delta, \sigma, S_t)$. Using the equations 4.20 and 4.21 for telescope tracking, it calculates zenith distance and Cassegrain rotator position angle according to:

$$z = \arccos\left[\sin\Phi\sin\delta + \cos\Phi\cos\delta\cos(S_t - \alpha)\right] \tag{5.13}$$

$$\psi = \sigma - \arctan\left[\frac{\cos\Phi\sin(S_t - \alpha)}{\sin\Phi\cos\delta - \cos\Phi\sin\delta\cos(S_t - \alpha)}\right] \tag{5.14}$$

When the spectrograph position is known, the shifts on the detector relative to the starting point can be calculated by interpolating the data in the flexure look-up table. Since the data in the flexure map is measured at finite angular steps, an interpolation is necessary to find the flexure for an arbitrary telescope position (z,

ψ). The angular steps in the look-up table are $\Delta z = 15^o$ and $\Delta\psi = 45^o$. These values may seem too large, but the interpolation algorithm can give a good approximation, provided that flexure has a fairly linear behaviour in that range of spectrograph motions. The interpolation routine finds the closest values in the stored matrix for the given telescope position ($z_i < z < z_{i+1}$ and $\psi_j < \psi < \psi_{j+1}$). It then calculates the departures from these values:

$$c_z = \frac{z - z_i}{\Delta z} \tag{5.15}$$

$$c_\psi = \frac{\psi - \psi_j}{\Delta\psi} \tag{5.16}$$

The interpolation is made by weighing the four shifts in the grid $X_{i,j} = X(z_i, \psi_j)$ according to the above coefficients so that:

$$X = (1 - c_z)(1 - c_\psi)X_{i,j} + c_z(1 - c_\psi)X_{i+1,j} + c_\psi(1 - c_z)X_{i,j+1} + c_z c_\psi X_{i+1,j+1} \tag{5.17}$$

This is simply an extension in two dimensions of the usual linear interpolation, but I found it can provide enough accuracy even with the relatively large steps with which the measures are made.

5.4.3.2 Conversion of Shifts into Voltages

The active collimator is designed to compensate for the shifts at the detector relative to the starting position and not the absolute ones, thus decreasing the scale of collimator motion required. In fact, at the beginning of the exposure, a reference line in the spectrum will be, in general, displaced from the arbitrary zero position by the quantities $X(0)$ and $Y(0)$. During the exposure the image will then move from the initial position according to:

$$\Delta X = X(t) - X(0) \tag{5.18}$$

$$\Delta Y = Y(t) - Y(0) \tag{5.19}$$

The relative displacements during a single exposure are significantly smaller than the "absolute" ones because the tracking speed of the telescope is usually small and the exposure time rarely exceeds one hour. The relative shifts in the detector have to be compensated by tilts of the collimator mirror, which can be identified by two

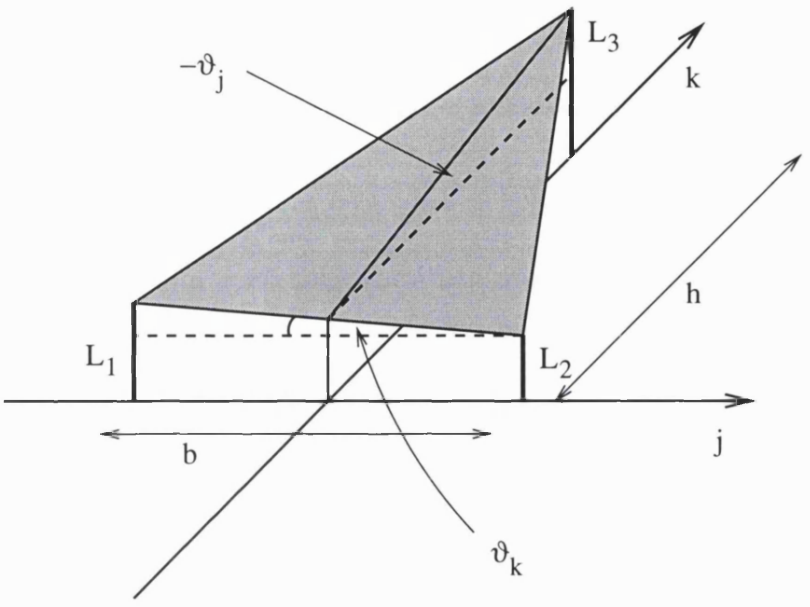


Figure 5.15: Schematic view of a generic tilt of the collimator mirror in its own system of reference. The (j,k) plane coincides with the one identified by the actuators at zero expansion ($L_1 = L_2 = L_3$). The figure shows three arbitrary positions for the actuators and the corresponding collimator tilts.

angles of tilt about two orthogonal axes in the collimator plane (j,k). The collimator system of reference used for the following equations is shown in fig. 5.15 and it is defined by the plane formed by the actuators at zero expansion (j,k) and any line normal to it. I chose this system of reference so that actuators one and two are aligned along the j-axis, while the k-axis is identified by the orthogonal line passing through actuator three. The (j,k) plane can be considered as parallel to the (X,Y) plane on the detector with the axes rotated by an angle γ. In other words, this means that by tilting the collimator with respect to the j-axis, the image in the camera focal plane will move following a straight line which forms an angle γ with the detector frame of reference. The mirror will be parallel to this collimator plane when the tilts are both zero. The angle γ is determined by the geometry of the spectrograph and by the optical components in the light's path. I decided it was not worth trying to calculate it and that it was better to measure it experimentally on the spectrograph. Assuming that the tilting angles are small enough to be a good approximation of their sine, and indicating with ξ the anamorphic factor of

the grating we obtain (the formal calculation is given in appendix A):

$$\vartheta_j = \frac{1}{2F_{cam}}(\xi \Delta X \sin\gamma + \Delta Y \cos\gamma) \tag{5.20}$$

$$\vartheta_k = \frac{1}{2F_{cam}}(\xi \Delta X \cos\gamma - \Delta Y \sin\gamma) \tag{5.21}$$

In our conventional system, these angles are then related to the actuator expansions (L_1, L_2, L_3), given the base $b = 160mm$ and the height $h = 172mm$ of the triangle formed by the actuators. The relationship between these variables can be easily derived from the geometry of the tilt shown in fig. 5.15 and it is given by the solutions of the following system:

$$\begin{cases} 2h\vartheta_j = L_1 + L_2 - 2L_3 \\ b\vartheta_k = L_1 - L_2 \end{cases} \tag{5.22}$$

The control software needs to invert the above system to calculate the actuator expansions from the collimator tilts. Since there are three variables for the actuator lengths and only two for the tilts, one of the actuator positions is arbitrary. To achieve maximum tilting range and avoid negative expansions, the actuator with the lowest position (number 2 in fig 5.15) can be set to have zero expansion. The lowest actuator can be determined from the sign of ϑ_j and ϑ_k and the value of their ratio. Once the lowest actuator is found and its expansion is set to zero, the system can be solved and the two remaining expansions calculated. Since the voltage versus expansion curve for the actuators is linear (thanks to the closed loop operation with the strain gauge sensor), the actuator voltages are given by:

$$V_i = K_i L_i \tag{5.23}$$

Here K_i is simply a scale factor, which I decided to leave as a free parameter to measure experimentally.

Chapter 6

Laboratory Tests

6.1 Actuator Tests

When the driving electronics for the actuators was completed, it was necessary to test it in combination with the actuators. I decided to test the piezoelectric actuators in open loop first, to measure the hysteresis and compare this with the data provided by the manufacturer. In this experiment the actuator was driven by the high voltage amplifier and the signal out of the strain gauge sensor recorded. To construct an hysteresis curve, the input voltage is increased to the maximum and brought back to zero in steps; the result is shown in fig. 6.1. The hysteresis is defined as the ratio of the maximum difference between the upper and lower curves and the total span of the curve and it is usually measured as a percentage. In the graph it amounts to around 12%, in good agreement with the data from the manufacturer, which quotes a value of 11.5% [113]. The contribution to the hysteresis from the strain gauges is extremely small, the nominal value being less than 0.2%. This implies that when operated in closed loop the actuators should show very little hysteresis. The actuator accuracy in closed loop could be measured directly, but since this implies using a position sensor independent from the strain gauges, I decided to accept the values from the manufacturer and measure the actuator movements *combined* with the collimator displacements within the cell. The accuracy in the mirror movement rather than in the actuator movement is the critical factor for the success of the ISAAC project. The measurements of the mirror movements are described in the

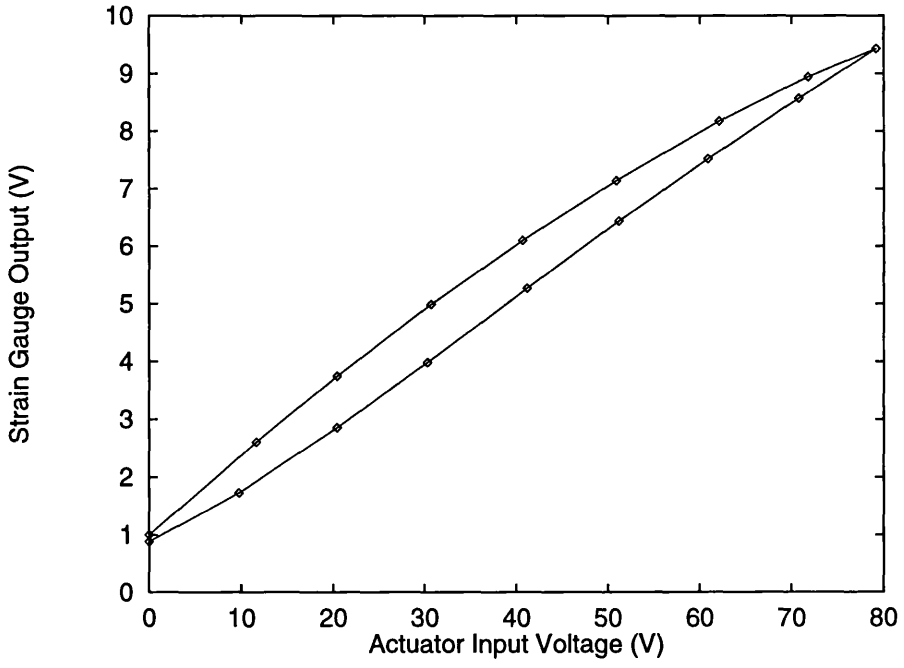


Figure 6.1: Piezoelectric actuator hysteresis curve as measured by the strain gauge sensor. The hysteresis is defined as the ratio of the maximum difference of upper and lower curves and the total output; in this case around 12% .

next sections.

I also tested the frequency response by driving the actuator with a sinusoidal wave and measuring the corresponding displacement. By gradually increasing the frequency of the sinusoidal signal, it is possible to measure the speed of the actuator amplifier. I found that the system could work at frequencies in excess of $1.0Hz$, which was my original target. Figure 6.2 shows the expansion of the actuator when driven by a sinusoidal signal at $1.0Hz$. In the measurement the actuator was connected to a Heidenhain probe which sensed the actuator expansion. The probe is based on a Moire fringe sensor, where two finely ruled gratings, placed in front of each other, are illuminated by a white source. When the probe moves the gratings are shifted with respect of each other and change in the fringe pattern is detected by a photodiode. The probe has an accuracy of $\pm 0.5\mu m$, which was enough to show that the actuator could follow the sinusoidal input exactly. The limiting factor in the frequency response is the actuator driving electronics, not the actuator itself: piezoelectric actuators can work at several KHz.

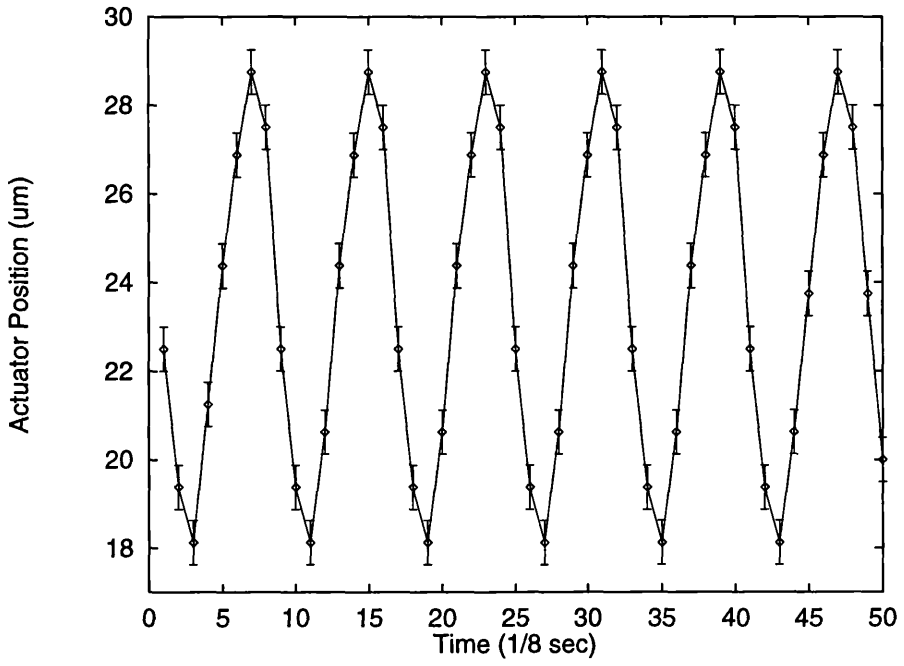


Figure 6.2: Frequency response of a piezoelectric actuator as measured by a Moire-fringe sensor. The actuator received a sinusoidal input between 0 and $8V$ at $1.0Hz$, producing an expansion amplitude of $12\mu m$. The accuracy of the measuring system was $\pm 0.5\mu m$. The graph shows the actuator is clearly capable of following the input, reproducing the sinusoidal signal.

The last test on the piezoelectric actuator involved measuring the expansion as a function of temperature. As I mentioned in chapter 3, the piezoelectric stack has a positive linear thermal expansion coefficient, so the absolute position of the actuator varies with temperature even if the input voltage is maintained constant. In open loop configuration this drift is not compensated and the absolute accuracy of the actuator is significantly reduced. To measure this effect I decided to monitor the room temperature over several hours and to register the corresponding output of the strain gauge amplifier. The actuator expansion is compared with the temperature change in fig. 6.3. This is not an independent measurement of the thermal expansion of the actuator because includes possible temperature effects in the strain gauge bridge. However, I was confident that the strain gauge sensor was quite insensitive to temperature and that would not affect the results of the experiments. In fact, when the temperature changes and the actuator expands, the strain gauge will have a different resistance and will unbalance the bridge, giving a voltage output. Only

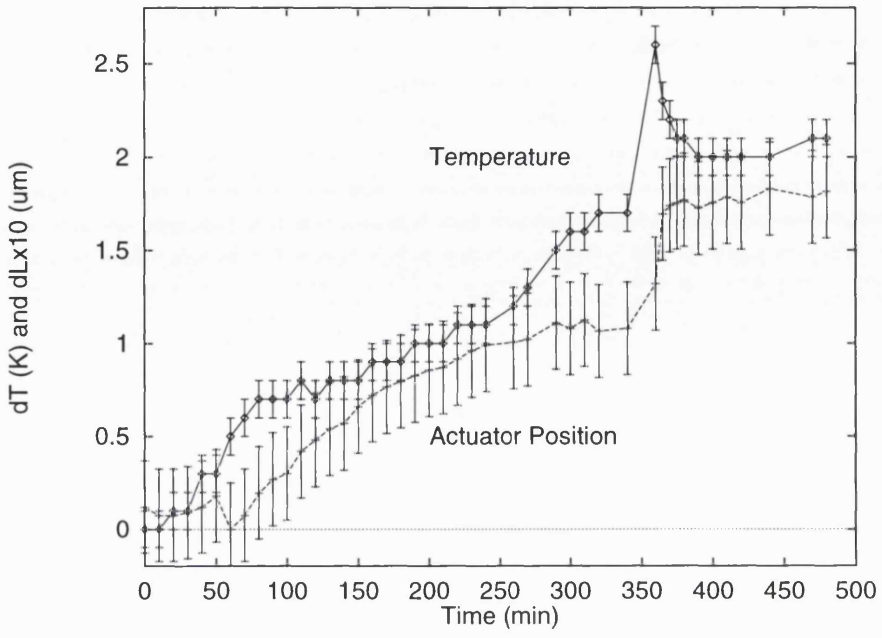


Figure 6.3: Actuator expansion compared to room temperature changes when no feedback from the strain gauge sensor is applied. The temperature change measured in Kelvin and the actuator expansion in tenths of μm share the same Y-axis to facilitate comparison. Due to this x10 magnification, 1 on the expansion axis (dL) corresponds to $0.1\mu m$. The graph clearly shows that after a time delay, the actuator expansion follows the temperature change. The effect is significant, since nearly $0.2\mu m$ of expansion are produced my a mere $2K$ temperature change. The closed loop operation of the actuator eliminates this thermal drift.

changes in the gauge coefficient with temperature may affect the sensor calibration, but these effects are quite small.

As fig. 6.3 shows, the actuator is very sensitive to temperature: even a change as small as 1^oC in the room, produces around $0.1\mu m$ expansion in the actuator (in the graph this occurs after $\sim 200min$). This value of $0.1\mu m/K$ is consistent with the linear thermal expansion coefficient of the PZT material and the length of the actuator (see chapter 3). Note the time delay of 20-30 minutes before the change in the room temperature is transferred to the actuator. This means the PZT has a fairly large heat capacity or it has some insulation from the external environment. When operated in closed loop the thermal drift of the actuator disappears because,

as explained before, the sensor measures the change in position and compensates for it by modifying the input voltage.

6.2 Mechanical Tests of the Mirror Cell

The mechanical design of the collimator cell had to be tested to verify that the motion accuracy and the location of the mirror inside the cell conformed to requirements. This involves testing the correctness of the axial and radial support. It was therefore necessary to measure the hysteresis in the motion, the lateral stability, the reliability of the neoprene pads and of the axial clamping.

6.2.1 Hysteresis and Radial Support

I already mentioned that our original design did not include a flexure for the radial support and relied to plastic screws on the sides of the collimator. Since these tests were conceived to check the correctness of the design and not to measure the final performance, I decided that it was not necessary to use the real collimator, and to substitute it with a dummy made of aluminium. Aluminium has a density very close to the one of Zerodur so the dummy could be made of a similar shape and weight of the real collimator. The use of the dummy simplified testing because I could use mechanical instead of optical techniques.

In order to measure the position accuracy and the hysteresis in the motion of the dummy mirror, I used three LVDT sensors (Linear Variable Differential Transformers). The LVDTs were rigidly clamped to the mirror cell pushing on the front of the mirror, almost exactly opposite the actuators on the back. The experimental setup is shown in fig. 6.4; this test can be done only using a dummy, because involves the sensors pushing against the *front* of the mirror, which would not be acceptable with an accurately polished mirror. The LVDT sensors I used had a nominal accuracy of $0.1\mu m$. The effective maximum accuracy, however, was only $0.3\mu m$, due to the electronics used to drive them (more specifically the DAC used was not accurate enough). Even so, the test was able to gave me important information concerning the shortcomings of the initial design.

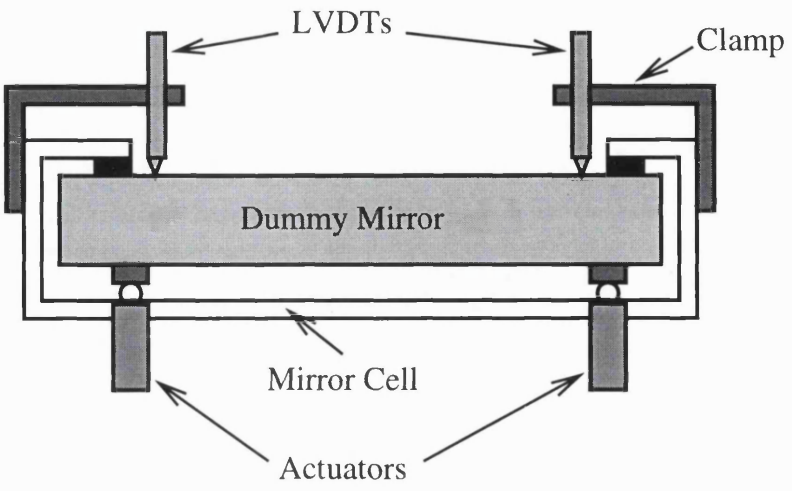


Figure 6.4: Experimental setup for the mechanical test of the mirror motions. LVDT sensors measure the motions of a dummy mirror with respect to the collimator cell. The purpose of the test was to check the validity of the mechanical design, so a dummy could be used. The mirror tilt is small enough to consider the motion entirely along the axis of the LVDT.

The experiment essentially consisted in pushing one actuator to maximum expansion and back in several steps, to produce a hysteresis curve. The readings from the LVDTs would give the total travel, the hysteresis and the influence the motion of one actuator had on the position of the whole mirror. I performed several tests in which different combinations and positions of the actuators were used (one actuator moving alone, two in tandem, one actuator always at maximum expansion, etc.). All these tests had a similar pattern of which fig. 6.5 is a typical example. The test proved that the initial choice for the radial support was incorrect. The graph shows an hysteresis in the movement of around 10% of the total travel, too high for my requirements and not attributable to the actuators, which were operated in closed loop with the strain gauge sensors. Moreover, the total travel of the mirror is only $\sim 10\mu m$, while the actuator expands by $15\mu m$. Having excluded the actuators as a possible source of hysteresis, the obvious culprit was the friction between the lateral screws and the dummy mirror. This was confirmed by the fact that the mirror occasionally jammed and, after the force on the actuator was increased, it suddenly jumped to the new position. The radial support with lateral screws was therefore

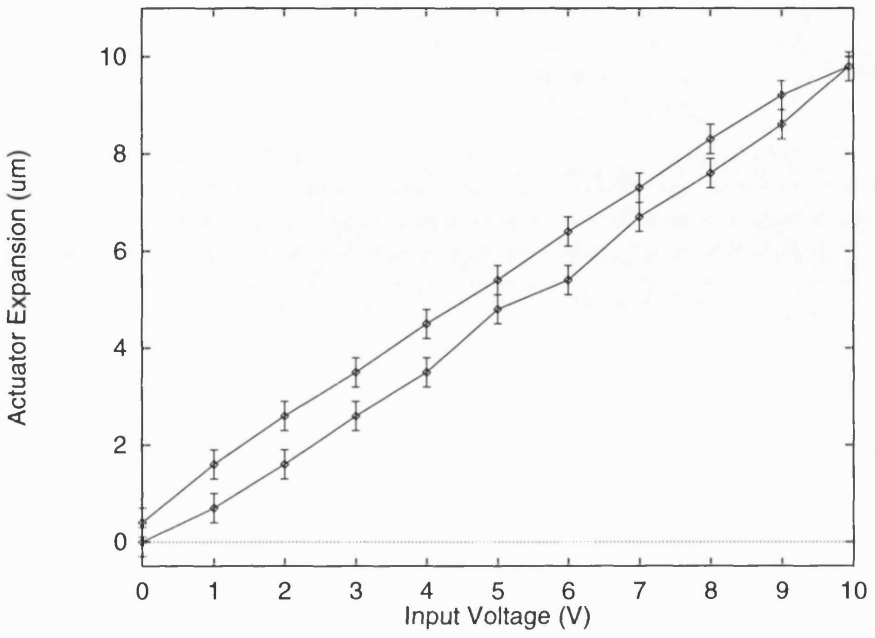


Figure 6.5: Hysteresis curve for the motion of the dummy mirror inside the cell. The radial support is provided by lateral nylon screws, which we found were responsible for the relatively large hysteresis (around 10%) shown.

rejected and substituted by a flexural element.

6.2.2 Test of the Flexural Radial Support

To prove that the flexural support would produce the necessary stability and positioning accuracy, I decided to test it with the dummy mirror first. The design of the flexure used is shown in chapter 5, the only difference being that when using the dummy aluminium mirror the flexure could be screwed directly on it. The same test as described above could then be performed and the result is shown in fig. 6.6. I again made several tests for each actuator, repeating the cycle more than once and changing the positions of the other actuators, with results very similar to the one shown in fig. 6.6 with respect to the level of hysteresis and the maximum movement. As the graph shows the introduction of the flexure has reduced hysteresis dramatically. In fact, hysteresis has been reduced to a point that it is not measurable any more within the accuracy of the LVDT sensors. I could only place an upper limit on it by taking twice the LVDT error ($0.6\mu m$) as the maximum

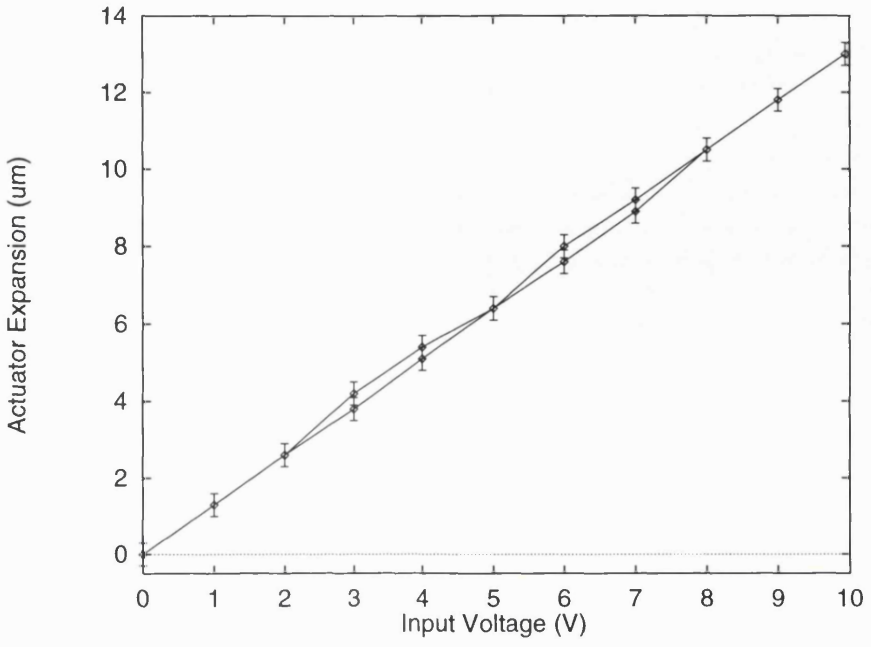


Figure 6.6: Hysteresis curve for the motion of the dummy mirror inside the cell, with the radial support provided by a flexural element. The hysteresis is much reduced when compared to the previous system and it is actually beyond the accuracy of the measuring system and cannot be evaluated.

hysteresis in the cycle, obtaining an hysteresis of less than 4.5%. Since I was now confident that the real value would be actually significantly less than that, I decided the flexural option for the radial support was worth pursuing to the level of more accurate optical tests. The same mechanical tests were repeated both with the cell laying horizontal (gravity vector orthogonal to the mirror surface) and in a vertical position (gravity parallel to the mirror surface), giving the same results. When the cell was laying vertical I also verified that a rotation about a horizontal axis did not affect the measurements. These further tests proved that the clamping mechanism was strong enough to hold the mirror firmly in place, whatever its orientation with respect to gravity.

Incidentally, these tests gave also the maximum collimator motion achievable with the combination of the actuators and the mirror support. As the graph shows, the motion is slightly less than the expected $15\mu m$. An average from several tests (in different orientation or with a different actuator combination) produced a value

$t_1 = 13.0 \pm 0.3\mu m$ for the actuator in position #3 and $t_2 = 12.5 \pm 0.3\mu m$ for the other two. We have to consider that the real displacement of the mirror can be up to $0.5\mu m$ bigger than the above values due to the fact that, for practical reasons the LVDT could not be placed exactly above the PZT actuator. Even so, the average motion is around $13\mu m$, $2\mu m$ less than the expansion of the actuators. It is not clear why this happened, because the large force the actuators can produce should overcome any resistance in the motion. I concluded that some of this force must be used for an elastic deformation of the whole mirror cell which does not displace the collimator within it. There is no proof for this happening, but, as long as the motion is reproducible and free from hysteresis, there was no need for investigating this matter even further.

6.2.3 Collimator Radial Stability

As a further test of the cell design I tested the radial stability of the dummy mirror within the cell. To measure lateral movements I again used LVDTs rigidly clamped to the cell. This time however, the LVDTs were placed parallel to the mirror surface and pushing against some L-shaped metal beams bolted onto the front of the dummy mirror. The cell could then be moved around and the mirror movements measured. The test however, proved to be less conclusive than expected, because I found that the cell was flexing under its own weight and the results depended on the way it was supported during the movements. Nonetheless, the mirror radial shifts I measured were very small: at worst they were $2\mu m$ in every direction. These shifts probably were partially due to the bending of the cell and partially to the mirror movement. However, the specified target for radial stability was $\pm 3\mu m$, that is above the measured values. I concluded that this test, albeit inaccurate (due to the overestimate of the lateral shifts), was good enough to prove that the radial support was meeting the specifications.

6.2.4 Elastic Properties of the Rubber Pads

The mechanical design relies on three rubber pads on the front of the mirror to provide the return spring against which the actuators will push. It is important to

know the elastic characteristics of the rubber to control the amount of stresses introduced in the mirror. The force applied to the mirror can be controlled by varying the thickness of the cylindrical spacers between the actuators and the back of the cell, as explained in chapter 5. There were two possible candidates for the rubber pad: neoprene and natural rubber. Neoprene has better resistance to environmental conditions and shows very little aging and temperature effects, while natural rubber supposedly has better elastic properties. I decided to use neoprene for the first mechanical tests in view of its better durability. If the tests were successful, neoprene would be the preferred choice. The tests proved that on the scale of the actuator movements (a few microns), neoprene has also excellent elasticity and little hysteresis, so neoprene was chosen as the material for the rubber pads. Before testing the performance of neoprene as a return spring however, I needed to measure the elastic constant of neoprene, so that the necessary compression could be calculated.

6.2.4.1 Rubber Compression Theory

The geometry of a rubber pad, like those used for the axial support of the active collimator is illustrated in figure 6.7. The stiffness of a block of rubber in compression depends upon the shape factor S, defined as the ratio of one loaded area to the total force-free area. For a standard rectangular block ($L \times B$) of thickness h, this is given by [114]:

$$S = \frac{LB}{2h(L+B)} \tag{6.1}$$

From the shape factor and the Young's modulus Y_o, the compression modulus Y_c of the rubber block can be derived:

$$Y_c = Y_o(1 + 2kS^2) \tag{6.2}$$

Where k is a numerical factor which depends on the hardness of the rubber and has a value between 0.5 (hard) and 1 (soft). The compression stiffness K_c is defined in analogy with the elastic constant of a spring as the ratio of the force F and the compression x. Its value is related to the compression modulus by:

$$K_c = \frac{F}{x} = \frac{Y_c A}{h} \tag{6.3}$$

Where $A = LB$ is the cross-sectional area. The compression stiffness can be easily measured experimentally and used to determine the elastic modulus of the rubber.

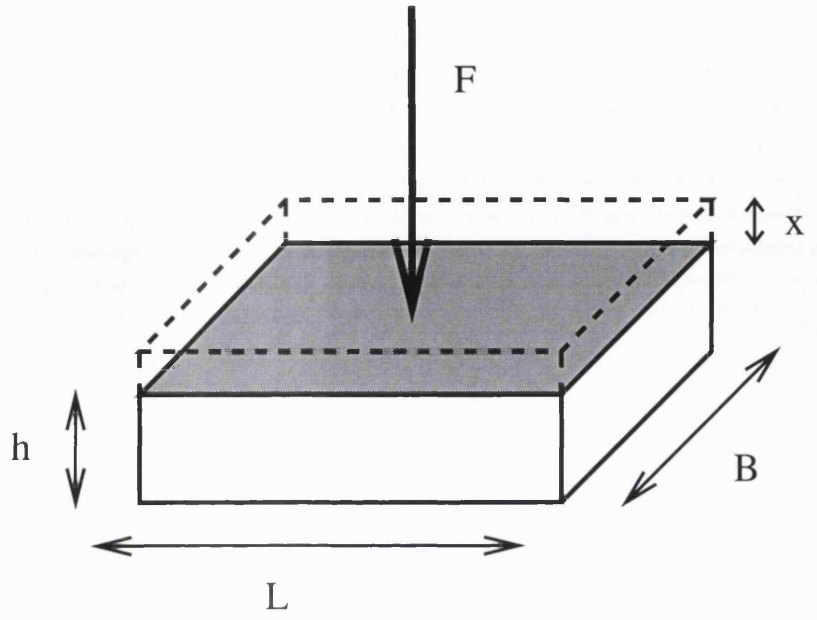


Figure 6.7: Geometry of a rubber pad. In the experiment to measure the compression stiffness of the pad, the compression x was varied and the force F necessary to produce that compression was measured.

Knowing also the expansion Δx of the actuators we can calculate the additional force introduced by the actuation process.

6.2.4.2 Measurement of the Compression Stiffness

The compression stiffness of the rubber was determined by gradually increasing the load on a single pad and measuring the corresponding compression. This was achieved by varying the thickness of the spacers holding the actuators: this would produce a gap of known size between the actuator holder an the cell. Several weights were then added on top of the actuator until the gap was closed. This gives the force necessary to compress the rubber by a given amount x. By choosing a set of different compressions and measuring the force necessary to achieve them, it is possible to draw a graph of the force versus compression, as shown in fig. 6.8. The graph displays a clear hysteresis curve and therefore inelastic behaviour. The effect, however, is characteristic of the large compressions we applied (up to about $1.5mm$ or 1000 times the motion range of the actuators). I have already mentioned that this

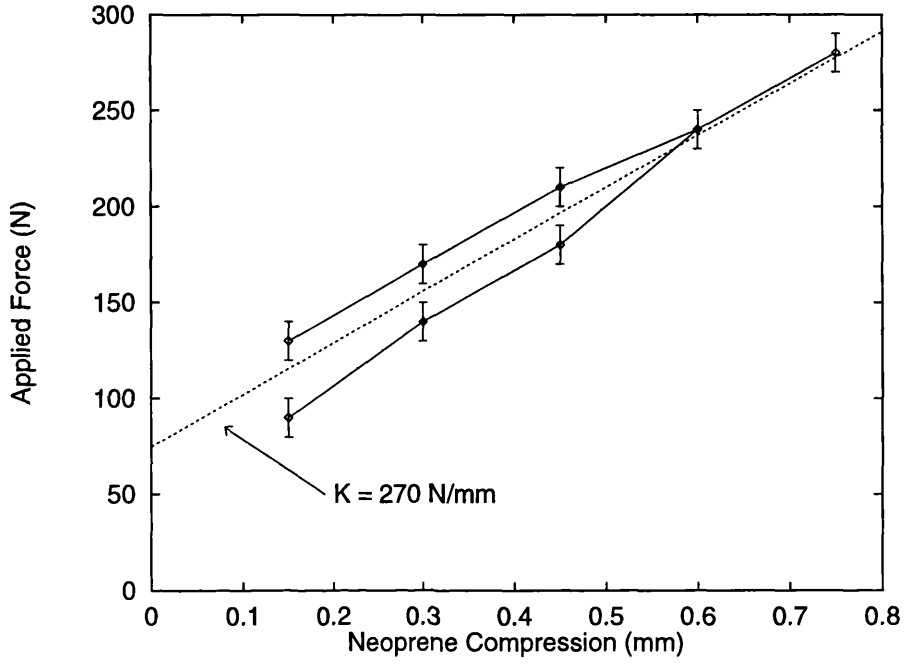


Figure 6.8: Compressibility curve of a neoprene pad inside the cell. The compression is relative to an arbitrary (and non-zero) preload, which however does not affect the result. The curve also shows a fitting line used to estimate the elastic constant of neoprene and to avoid considering the significant hysteresis that neoprene has on this scale of compressions.

hysteresis was not observed on much smaller compressions. An approximate value for the compression stiffness can still be extrapolated from the graph, and used to estimate the force applied on the mirror by just putting the actuators in place. The compression stiffness for the neoprene pad was measured as:

$$K_c = 270 \pm 40 N/mm \tag{6.4}$$

This value is important to decide the amount of force applied in the axial clamping of the collimator. This experiment shows that in order to apply around $100N$ at each actuator (as specified) the rubber pad needs to be compressed by $\sim 0.4mm$. The additional compression introduced by moving the actuator to tilt the collimator ($15\mu m$ expansion) adds only $4N$ to the total force applied and can be neglected. To verify that the above measurements were correct, I derived an estimate of Young's modulus from equations 6.2 and 6.3 and compared the result with known values. The shape factor S for the rubber pad can be easily calculated: the pad had $L = 10mm$,

$B = 8mm$ and $h = 2.5mm$, therefore $S = 0.9$. For neoprene this produces a value of $Y_o = 4.3N/mm^2$ ($k = 0.6$), in agreement with typical values for rubber compounds [114].

6.3 Temperature Effects and Radial Support

After the mechanical tests described in the previous sections, it became evident that an aluminium flexure was the best choice for the radial support system. This however, created the additional issue of how to connect the flexure to the Zerodur mirror and avoid introducing stresses due to differential thermal expansion. In chapter 5 I mentioned that these two materials have very different linear thermal expansion coefficients and that the flexure needed to be made of aluminium to match the thermal expansion of the cell. Solutions already tested, like the radial support system for the segments of the primary mirror of the Keck telescope [115], were impractical due to the space restrictions within the cell. Therefore we had to design and test a new system that would keep the radial position and overcome the thermal issues. The solution, described in chapter 5, was to have an Invar cylinder inserted in a hole bored in the back of the mirror. I had however to estimate the possible stresses introduced and to test these predictions before the collimator mirror could actually be modified.

6.3.1 Experiment with a Pyrex Mirror

To test the correctness of our idea and gain some experience in glass-metal bonds, I did a few experiments using a small flat mirror made of Pyrex. The mirror is circular, $6.0cm$ diameter and roughly $2cm$ thick. Pyrex has a linear thermal expansion coefficient $\alpha_P = 3 \cdot 10^{-6} K^{-1}$ [110], much larger that is, than the one of Zerodur. Therefore, to create thermal stresses between the mirror an the metal insert I decided to use a steel insert, which would have a relatively large thermal expansion ($\alpha_{steel} = 13 \cdot 10^{-6} K^{-1}$). The purpose of the test was to measure the effect these stresses would have on the mirror surface and then extrapolate the effects on the collimator. A hole $20mm$ diameter was drilled in the back of the Pyrex mirror and a metal cylinder was machined to fit exactly in the hole. The fit was very good: the

air gap between the hole and the insert was estimated to be $2-3\mu m$ only. In fact, warming the steel cylinder by simply holding it in the hand, was enough to close the gap and make the fit so tight that the insert could not be removed unless it was cooled down.

The Pyrex mirror did not have the flatness I needed to measure surface distortions due to thermal effects, so I had to polish it flat again. Figure 6.9 shows the pyrex mirror during polishing and the final result with the steel cylindrical insert. The flatness of the mirror can be easily tested by placing it on a reference flat and illuminating the air gap between the two surfaces with a low-pressure sodium lamp. With two perfectly flat mirrors, the illumination of the thin air wedge in between produces a series of parallel fringes of equal thickness (Newton fringes) [116]. The sodium lamp is needed to produce the necessary fringe contrast. To polish the Pyrex mirror, I prepared a flat pitch by compressing it against the reference mirror for around one hour. The pitch is placed on a rotating table and the Pyrex mirror is polished by sliding it on the rotating pitch. Sufficient randomness in the motion is guaranteed by describing a W on the pitch with the mirror. During the polishing I frequently tested the progress with the interferometric test described above.

After a few hours work the mirror was polished flat with a maximum departure of around half a wavelength or $0.3\mu m$ ($\lambda_{Na} = 589nm$). This happens when a single Newton fringe covers the whole mirror surface in the above test. To measure thermal distortion of the mirror surface due to differential expansion, I set up a knife-edge test for the Pyrex mirror using the ISAAC collimator to produce a parallel beam. The knife-edge test is described in detail in section 6.5, the experimental setup is shown in fig. 6.10. I used a thermocouple probe in contact with the metal insert to measure the temperature changes. Initially I used a hair drier to uniformly heat both the steel cylinder and the mirror, but the test with the Newton fringes showed that even by heating the mirror on its own surface distortions could be produced. Therefore I decided to heat only the metal with a soldering iron and take advantage of the low thermal conductivity of glass with respect to steel ($k_{steel}/k_P \sim 30$) to reduce the heating of the mirror to a minimum.

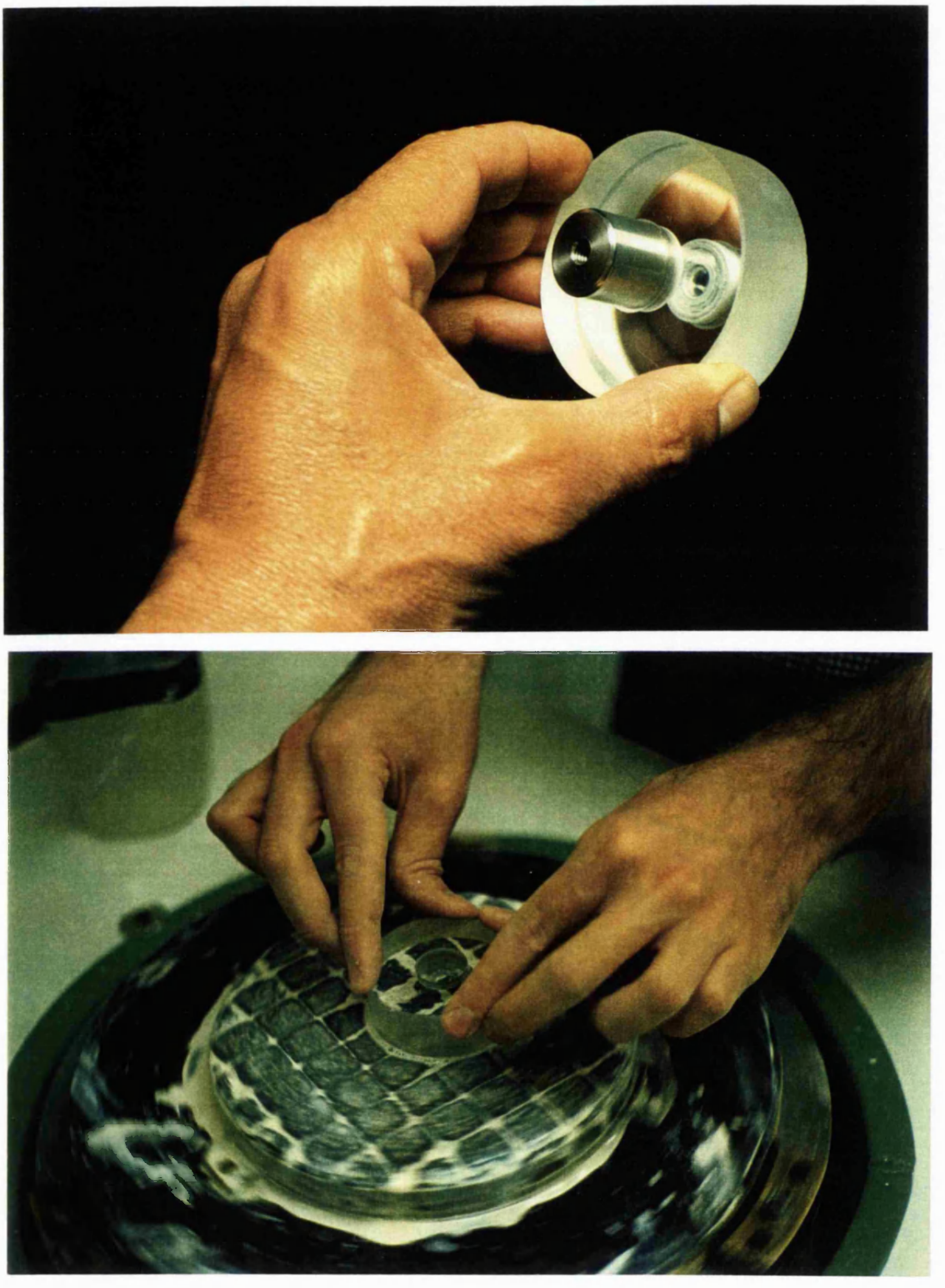


Figure 6.9: Polishing of the pyrex mirror and the mirror after polishing with the metal insert.

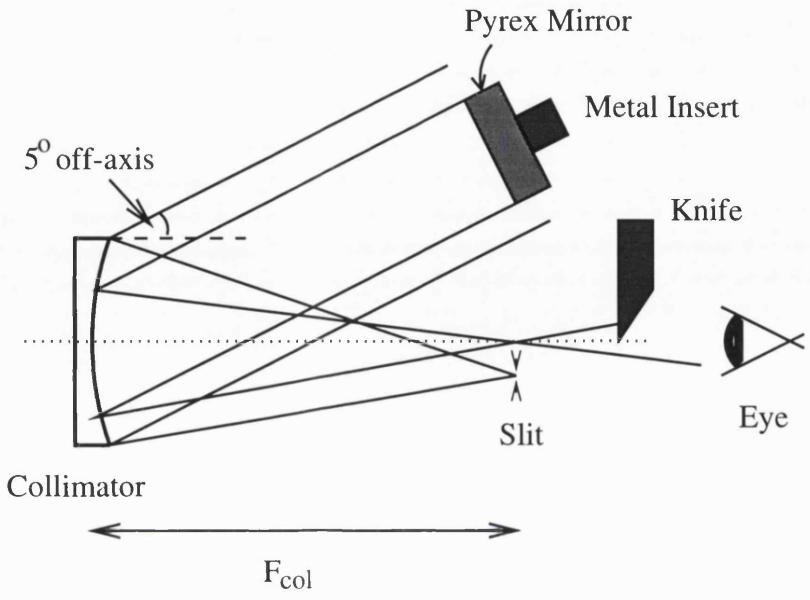


Figure 6.10: Experimental setup for testing surface distortions of a Pyrex mirror due to thermal expansion of a steel cylinder inserted in a hole in the back. The knife-edge test reveals both asymmetric distortions and changes in the focus (curvature) of the mirror. A thermocouple probe in contact with the steel was used to monitor the temperature.

6.3.2 Test of the Thermal Stresses

The experiment consisted in heating the metal and at the same time to observe the effect on the mirror surface with the knife-edge to check for asymmetric distortions and changes in curvature (focus). The test did not show any distortion of the knife image, indicating that the surface deformations were symmetric. The knife-edge test also allowed me to detect changes in the focus position with the heating. For this purpose the knife was mounted on an X-Y micrometric stage so as to measure small curvature changes in the mirror. Since the soldering iron has a temperature of around $300°C$, the steel insert can be heated fairly quickly. A graph of the focus change as a function of temperature is shown in fig. 6.11 The metal insert was heated for approximately $15min$ and left to cool down for the same time. The experiment showed that the change in focus follows the temperature change fairly quickly, suggesting that this is produced by the pressure exerted by the metal on the glass rather than by the thermal distortion of the glass itself. Assuming this assumption is true,

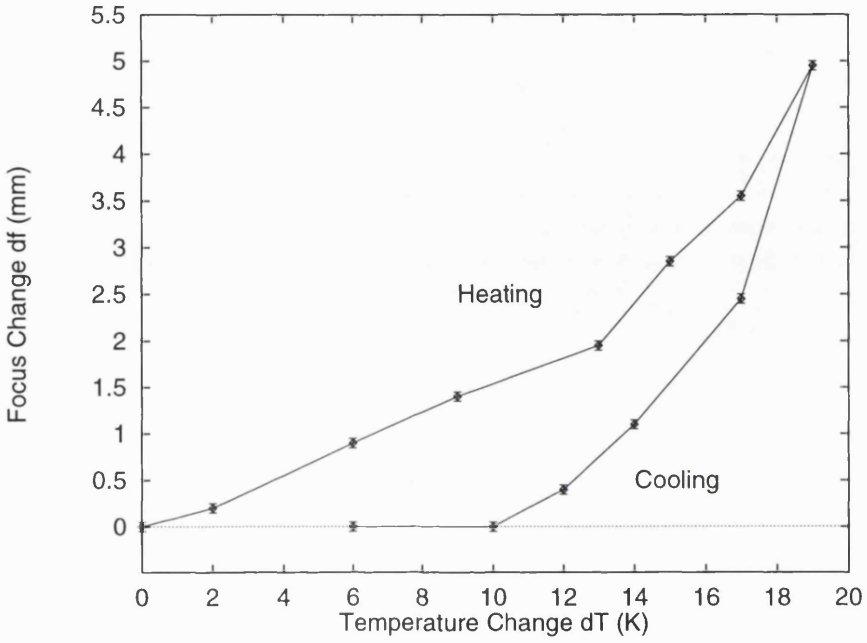


Figure 6.11: Changes in the focus position in the knife-edge test due to temperature changes of the metal cylinder inserted in the back of the pyrex mirror. The heating time and the cooling time are approximately $15min$.

it is possible to evaluate the effect of differential thermal expansion on the surface of the Pyrex mirror.

If we indicate with f the combined focal length of the system (Pyrex mirror + collimator) and with f_P the focal length of the Pyrex flat then:

$$\frac{1}{f_P} = \frac{1}{f} - \frac{1}{F_{col}} = \frac{F_{col} - f}{f F_{col}} \tag{6.5}$$

Since the Pyrex mirror is almost flat, $f_P \simeq \infty$ and the combined focal length almost equals the collimator focal length. Thus:

$$\frac{1}{f_P} \simeq \frac{\Delta f}{F_{col}^2} \tag{6.6}$$

The mirror can be seen as a very shallow spherical mirror with radius of the generating sphere r, radius of the mirror $a = 30mm$ and focal length $f_P = r/2$. It is easy to see that the deviation from flatness at the edges of the mirror is approximately given by $\delta \sim a^2/r$, therefore:

$$\delta \simeq \frac{a^2 \Delta f}{2F_{col}^2} \tag{6.7}$$

Given $F_{col} = 1650mm$ and $\Delta f = 5.0mm$ (for $20°C$ temperature change), the maximum distortion at the edges of the Pyrex mirror is roughly $\delta \simeq 0.8\mu m$.

Now the problem arises of how to extrapolate the effect of an Invar insert in the back of the Zerodur collimator. The important parameters are the diameter of the insert l, the temperature change and the difference in linear thermal expansion coefficients $\Delta\alpha$. For the Pyrex/steel bond $\Delta\alpha = 10 \cdot 10^{-6}K^{-1}$ while for the Zerodur/Invar is less than $\Delta\alpha = 1.5 \cdot 10^{-6}K^{-1}$, almost 10 times smaller. We also decided to have an insert $l = 10mm$ diameter, half the size of the one for the Pyrex. There is almost a factor 20 reduction in the expected effects, just taking the above factors into account. The collimator mirror is also twice as thick as the Pyrex, suggesting that the effect should be even smaller. If we assume a scale factor of 20, we can expect thermal distortions less than $40nm$ on temperature changes of more than $20°C$. This value needs to be compared with the mirror polishing accuracy of 1/10 of a wavelength or $60nm$. Therefore I concluded that this radial support system could be safely used for the active collimator.

6.4 Optical Tests: Mirror Motion

The target accuracy for the tilts of the collimator is so high that the mechanical tests described in section 6.2 were not able to measure it. Moreover, the active collimator needed to be tested with the real mirror inside. An optical test was necessary to measure the mirror motion and accuracy. The goal for tilting accuracy was extremely stringent: I wanted to control mirror tilts as small as $\beta = 0.2arcsec$ ($10^{-6}rad$) and the optical test had to be able to detect such small tilts. It soon became clear that shining a laser onto the collimator and measuring the movements of the reflected beam was not a satisfactory option. If L is the distance between the collimator and the detector, a tilt of an angle β would produce a movement on the detector by $2\beta L$. Given the accuracy in tilting I required, with a distance of, say, $L = 4m$ the motion would be only $8\mu m$. On such distances however, the cross-sectional spread w of a laser due to the divergence of the Gaussian beam is [117]:

$$w(L) \simeq \frac{\lambda L}{\pi w_0} \tag{6.8}$$

For a standard He-Ne laser with $w_0 = 0.4mm$, and the above distance L, the spot would have a radius $w \simeq 2000\mu m$. This means we would have to detect motions as small as $8\mu m$ of a spot more than $4000\mu m$ across. For this reason I set up a different testing apparatus in which the motion of an image in the focal plane of the collimator is measured.

6.4.1 Test with Slit and Microscope

In this test the collimator mirror is placed in auto collimation with a reference flat mirror (flatness better than $\lambda/10$ at $632.8nm$). A slit, illuminated with white light, is in the focal plane of the collimator and it is re-imaged in the same plane a few millimetres apart from the source after being reflected back by the flat mirror. The motion of the image in the focal plane is then amplified by a convex achromatic lens by roughly a factor 9.6 (the amplification is determined by the ratio of the conjugate points of the lens: $p = 120cm$ and $q = 12.5cm$). The amplified motion of the slit image is finally measured with a microscope placed onto a micrometric xy stage. The microscope is at the conjugate focus of the magnifying lens. The experimental setup for the test is shown in fig. 6.12. When the mirror is tilted by an angle β, because of multiple reflections (twice on the collimator and once on the flat), the converging beam coming from the collimator is tilted by an angle 4β. This can be derived as follows: the first reflection on the collimator introduces a tilt 2β, the flat folds the beam back on the collimator which introduces another additional tilt 2β, bringing the total to 4β. Therefore the displacement d of the slit image in the collimator focal plane is given by:

$$d = 4\beta F_{col} \tag{6.9}$$

And the amplified motion d' at the microscope is:

$$d' = \left(\frac{p}{q}\right) d \tag{6.10}$$

Since the tilting angle of the collimator is related to the expansion t of the actuator by $\beta = t/H$ (where H is the distance between two actuators), we can calculate the scale factor of our test system:

$$d' = 4\left(\frac{p}{q}\right)\frac{F_{col}}{H} \cdot t \tag{6.11}$$

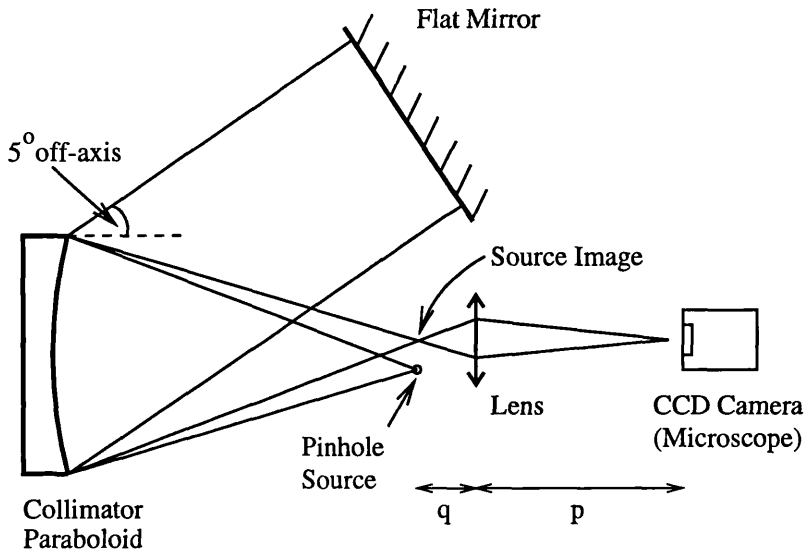


Figure 6.12: Optical test of mirror movements: experimental setup. The drawing shows the auto-collimation setup which was used for the slit + microscope test and for the pinhole + CCD camera test. The setup is identical in both cases, only the source and the detector change. The 5^o off-axis angle of the collimator is greatly exaggerated. The lens magnifies the motion of the image on the CCD detector by the ratio p/q. The pinhole was $25\mu m$ in diameter, but a direct measurement of the scale factor through the spot size on the CCD was not possible due to the image spread introduced by optical aberrations.

By substituting numerical values we obtain $d' \simeq 3960 \cdot t$. When one actuator is driven to full expansion, $t = 13\mu m$, therefore we expect a maximum displacement of $d' \simeq 5.1mm$; this agrees very well with the measured value of $5.00mm$ (see fig. 6.13). The scale factor is such that $1mm$ slit motion corresponds to about $2.5\mu m$ in the actuator expansion. Once the measuring system was properly aligned, the test of the mirror movement was analogue to the one performed with the LVDTs: the actuators are pushed to maximum expansion and back to produce an hysteresis curve; the result is shown in fig. 6.13.

During the test one actuator was moved in steps to tilt the collimator. This displaces the silt image in the focal plane of the magnifying lens, so I had to move the microscope with the micrometric screws to re-center the slit in the microscope eyepiece. The detection of the movement was a combination of the visual centering

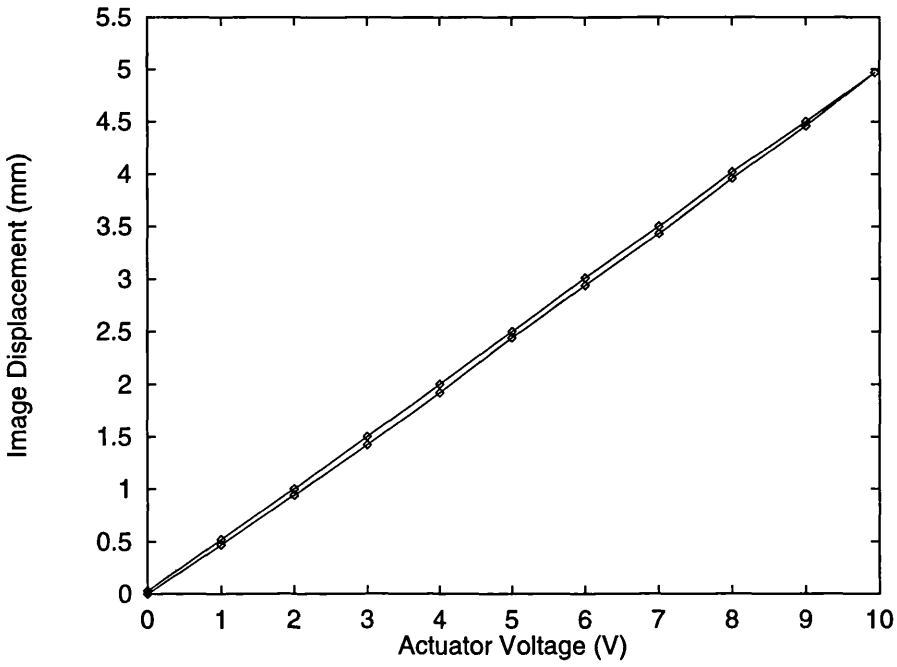


Figure 6.13: Hysteresis in the collimator movements with the optical test using a slit source and a microscope for detection. The actuator is moved cyclically to produce the hysteresis curve. The detection is visual through the microscope.

of the slit and the position on the micrometric screw. A picture of the setup for this test is shown in fig. 6.14. The estimated error in the measurements is $\pm 0.02mm$ because the image of the slit was sometimes a bit fuzzy and difficult to position accurately in the microscope field. Changing the actuator produces a very similar curve. The experiment in some cases showed the presence of backlash in the microscope movement when the direction of motion was reversed. Nonetheless, I could further lower the upper limit for hysteresis with respect to the measurements made with the LVDTs. The hysteresis, measured for two different actuators, resulted to be $1.6 \pm 0.8\%$ for actuator 2 and $2.6 \pm 0.8\%$ for actuator 1. The errors are unfortunately large (for the reasons explained before), but this hysteresis is still significantly less than the previous upper limit measured with LVDTs. Movements of the slit image could be detected, even if not measured, for changes in applied voltage as small as $0.05V$; that corresponds to an actuator movement of $75nm$. This suggested that the accuracy of the system was probably better than the test equipment could measure, so I decided to try to improve the accuracy of the testing apparatus by substituting the slit with a pinhole and the microscope with a CCD camera.

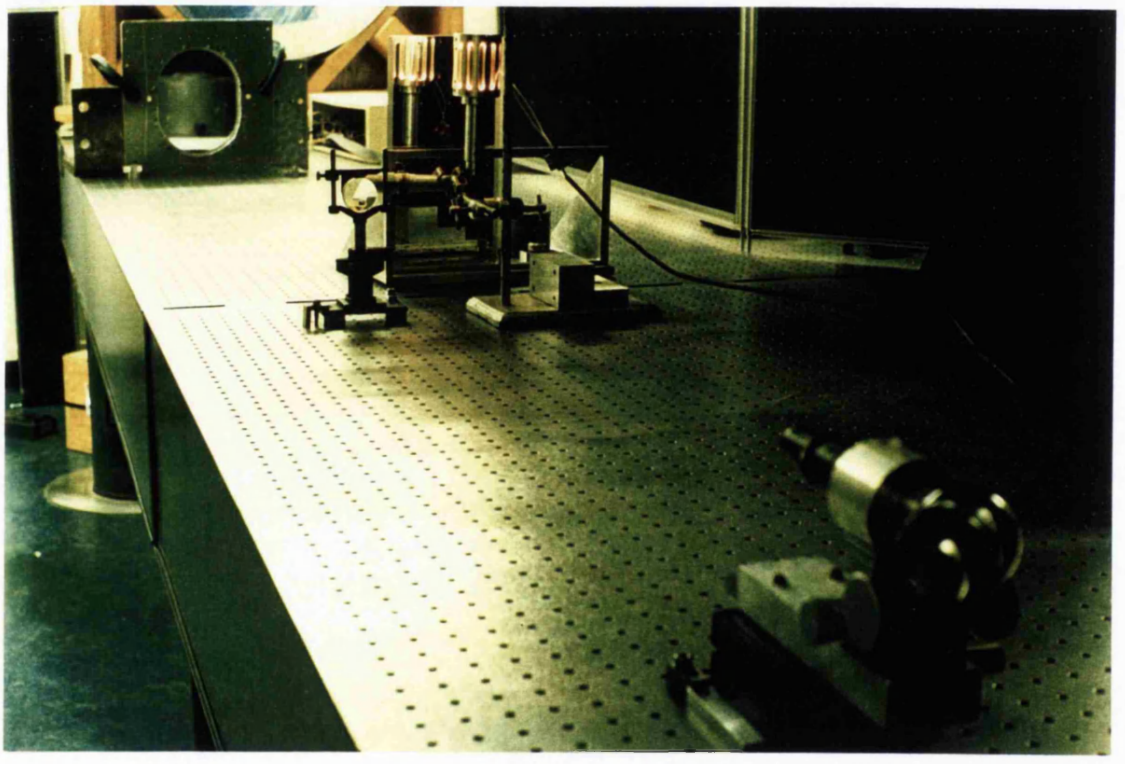


Figure 6.14: The setup for testing the mirror motions on the optical bench at OSL. In a later test the microscope was substituted with a CCD camera.

6.4.2 Test with Pinhole and CCD Camera

This tests uses the same optical setup as the precedent one, shown in fig 6.12, with the difference that a pinhole substitutes the slit source and the CCD camera the microscope as a detector. The pinhole, $25\mu m$ diameter, provides a smaller image and allows to detect both horizontal and vertical motions. The advantage in using the CCD camera is twofold: the camera doesn't need moving, eliminating any possible problem of backlash, and the image motion can be measured with centroiding techniques, rather than by visual estimate of the image center, improving the measurement accuracy. The only limitation is the camera field of view, determined by the CCD size, which sets the maximum voltage we can apply to the actuators to $6V$. Note that the camera objective was removed and the pinhole was imaged directly on the CCD. This setup virtually reproduces the spectrograph layout, with the grating substituted by a flat mirror. The theoretical value of the image displacement can be calculated in the same way as in the previous section.

The pinhole image on the CCD is very sensitive to the optical alignment of the experimental setup and can show severe optical aberrations if some optical component (the magnifying lens in particular) is out of place. After careful alignment, the image can be reduced to an almost circular spot with a Gaussian-like intensity profile. The three-dimensional intensity of the pinhole image on the CCD is shown in fig. 6.15, the intensity map in fig. 6.16. The CCD data is recorded using IRIS Frame Grabber and converted into a rectangular matrix containing the intensity data of each pixel (I_{mn}). This set of data can then be centroided to provide a very accurate measurement of the pinhole position on the CCD. The frame grabber has an 8-bit accuracy (1 in 256), therefore the saturation level is at 255 ADU (Analogue Digital Unit).

Since the purpose of this test was to measure positioning errors in the collimator movement, I was not interested in determining an absolute scale factor for the motion. The task of relating the absolute actuator movements to CCD shifts is made more difficult by the uncertainties in the magnifying ratio p/q and by the optical aberrations in the pinhole image. However, the accuracy of the motion in terms of position reproducibility, can be determined as a percentage of the total actuator motion ($13\mu m$) or of the maximum input voltage ($10V$). This allows us to translate

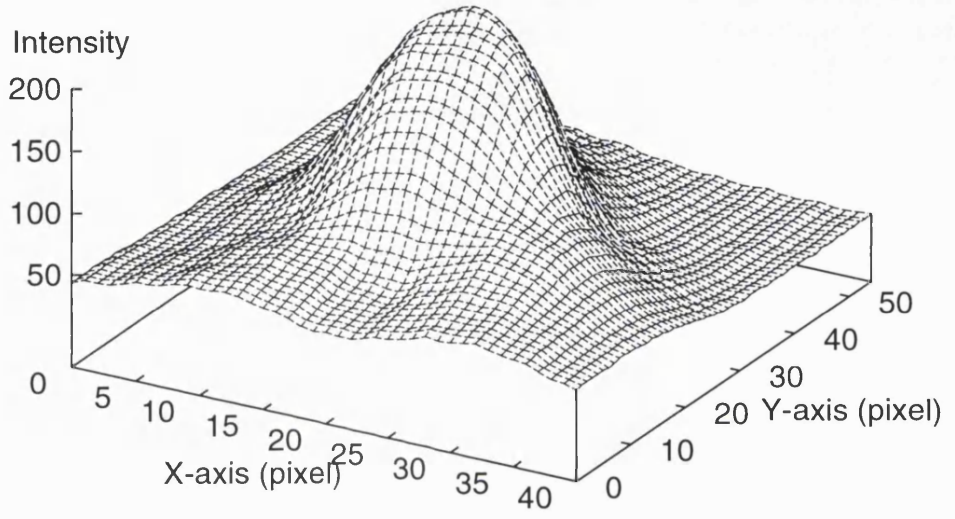


Figure 6.15: 3-D image of the pinhole as recorded by the CCD camera. Some residual optical aberrations can be seen in the form of low-intensity smearing of the image profile. The image intensity is however much higher than the background noise, giving a good sampling area and facilitating background subtraction.

accuracy in pixels on the CCD into accuracy in microns of actuator motion, without having to measure the scale factor.

6.4.2.1 Centroiding Algorithm

The center of the image of fig. 6.15 (in pixel units) can be determined from the image intensity matrix I_{mn} with a simple centroiding algorithm which weights the position of each pixel with its intensity. If we take the x axis along the CCD rows and the y axis along the columns and we indicate with $(\bar{x}, \bar{y})$ the intensity-weighted image center, we obtain:

$$\bar{x} = \frac{\sum_{mn} m I_{mn}}{\sum_{mn} I_{mn}} \tag{6.12}$$

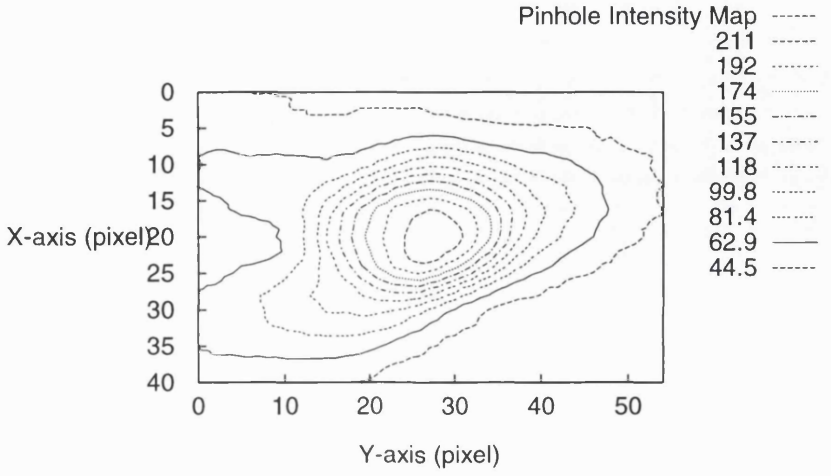


Figure 6.16: Intensity map of the pinhole image on the CCD. The spot is not exactly circular, but some distortion is introduced by the frame grabber, which converts a 385×288 CCD format into a 512×512 square frame. Note the two weak "tails" in the profile introduced by coma and other optical aberrations in the measuring system.

and

$$\bar{y} = \frac{\sum_{mn} n I_{mn}}{\sum_{mn} I_{mn}} \tag{6.13}$$

This is, in a way, equivalent to calculating the centre of gravity of a two dimensional body. In fact, from my point of view, it was not important to find the exact centre of the pinhole image. Provided that the shape of the spot remains approximately constant, all I needed was a reliable procedure to find a particular point in each image. I decided that this 'centre of gravity' was the easiest option. Then I could relate the positions on the CCD of two different images by simply considering the position of the two 'centre' points. In practice, I did not need to analyze the whole CCD frame, and I could limit myself to a window big enough to contain most of the pinhole image, greatly reducing the amount of calculations the centroiding program requires. It is important however, that the background signal from an image matrix like the one shown in fig. 6.15 is removed before the centroiding algorithm is applied.

The reason for this is that if the image is faint with respect to the background, the algorithm above would hardly notice any image movement, even if the displacement is quite large. Thus, subtracting the background increases the sensitivity of the centroiding process. When the background is subtracted, most of the signal on the CCD is produced by the imaging of the pinhole. In this experiment the image spot produced maximum counts of around 235 ADUs, of which approximately 40 were from the background. The spot itself, after background subtraction was nearly 30 by 30 pixels.

This simple centroiding algorithm does not take into account the effect of pixel to pixel variations in sensitivity. These variations, however, are quite small: for the CCD used in the experiment, the manufacturer quotes a value of $\pm 3\%$. I calculated that even in the worst cases the error introduced by these variations is still of the same order of magnitude, or $\pm 3\%$. Thus, for 1 pixel image shift, the error is still only ± 0.03 pixels. Moreover, this is the error resulting from particularly unfavourable situations, and the average error due to pixel to pixel variation would be much smaller than that. The error produced by the centroiding itself can be calculated from equations 6.12 and 6.13 by differentiation. For example, if we take $\bar{x}$, the error in its value derives from the differential according to:

$$d\bar{x} = \sum_{mn} \left| \frac{\partial \bar{x}}{\partial I_{mn}} \right| dI_{mn} \tag{6.14}$$

Where:

$$\frac{\partial \bar{x}}{\partial I_{mn}} = \frac{m \sum_{mn} I_{mn} - \sum_{mn} m I_{mn}}{\left(\sum_{mn} I_{mn}\right)^2} \tag{6.15}$$

More accurately, given the statistical nature of the algorithm (due to the relatively large number of points involved), the summation of absolute errors of eq. 6.14, should be substituted by the square root of the summation of the squares of the errors [118]. The intensity error dI_{mn} is produced by the fact that every CCD pixel (m,n) is essentially a photon-counting system. Therefore, assuming a Poissonian distribution of counts, the error on the counting will be simply the square root of the count itself, that is:

$$dI_{mn} = \sqrt{I_{mn}} \tag{6.16}$$

The total error on the image position can then be written as:

$$d\bar{x} = \sqrt{\sum_{mn} \left(\frac{\partial \bar{x}}{\partial I_{mn}} \right)^2 I_{mn}} \tag{6.17}$$

Both the centroids and the centroiding errors can be easily calculated with a computer program to give us the pinhole motion on the CCD. Typical errors from the centroid resulted to be of the order of ± 0.02 pixels. This error is comparable with the maximum errors from pixel to pixel variations, so an overall accuracy of ± 0.05 pixels was assumed.

6.4.2.2 Experiment Results

In the test I once again moved the actuators to produce an hysteresis curve as in the test with slit and microscope. After each movement a CCD frame is recorded with the frame grabber and the data saved into a matrix. By centroiding each of these frames I calculated the X and Y shifts on the CCD. The resulting curve, for one of the actuators and for the X axis, is shown in fig.6.17. To have an idea of the scale of motions with respect to the pinhole image size, fig. 6.18 shows the CCD 3-D intensity plots of the data from which the points in the previous figure are calculated. Each plot in the figure corresponds to a one volt actuator step ($1.3\mu m$ actuator motion). From the plot of figure 6.17 the hysteresis is calculated by measuring the maximum distance between upper and lower curves and comparing it with the total motion. For each actuator we get two graphs corresponding to x and y displacements and two values for hysteresis, which are expected to be the same within the errors. Actuator 2 showed lower hysteresis with values $1.47 \pm 0.04\%$ (x axis) and $1.51 \pm 0.06\%$ (y axis). For actuator 1 we got values $2.16 \pm 0.06\%$ (x axis) and $2.05 \pm 0.08\%$ (y axis). Note that these results are compatible with the ones from the previous experiment, but much more accurate. I concluded that some residual hysteresis, with an average value of around 2%, was still present in the collimator motions. It is difficult to identify the source of this residual hysteresis: the piezoelectric actuators and strain gauge sensors may contribute to it, but other causes, like the neoprene rubber pads, the flexural support and the mechanical axial support can all add some hysteresis.

However, the positioning error resulting from this is very small. For a total motion of the mirror of $13\mu m$, the corresponding error due to hysteresis is between 0.2 and $0.3\mu m$. Note that this is the error in the position of the *collimator* and not the position of a single actuator, for which the accuracy requirement was $80nm$. Also this is an error based on hysteresis, and it is expressed as a percentage of the amount

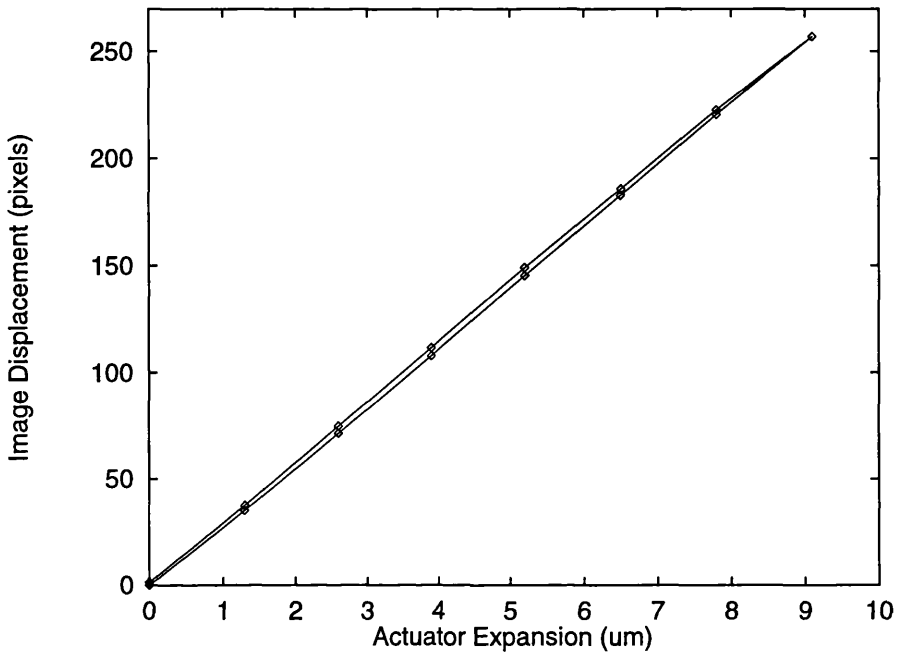


Figure 6.17: Hysteresis in the collimator movements with the optical test using a pinhole source and a CCD camera for detection. Each point is calculated by centroiding the image spot from the CCD frame. The errors are much smaller than the size of the symbol used for each point. The residual hysteresis shown is around 1.5%.

of collimator motion. Therefore a position error of $0.26\mu m$ refers to the maximum collimator tilt. For smaller tilts the error would decrease accordingly. So if, for example, the collimator tilts by half the range, the error would be only $0.13\mu m$. For very small tilts this is obviously not true, and other factors, like the strain gauge accuracy, need to be considered. To test the smallest movements, I applied voltages as small as $0.05V$ and I was still able to measure the image shift on the detector. An actuator voltage of $0.05V$ corresponds to $65nm$ expansion. After moving the collimator by $65nm$ and $130nm$, I brought the voltages back to zero: the collimator returned within $20nm$ of the original position. This is consistent with the strain gauge nominal accuracy of $30nm$ at constant temperature.

The performance of the collimator fits well within the position accuracy requirement. This can be better seen by translating the above values into positioning accuracy on the CCD detectors of the ISIS spectrograph. Considering the scale factor between

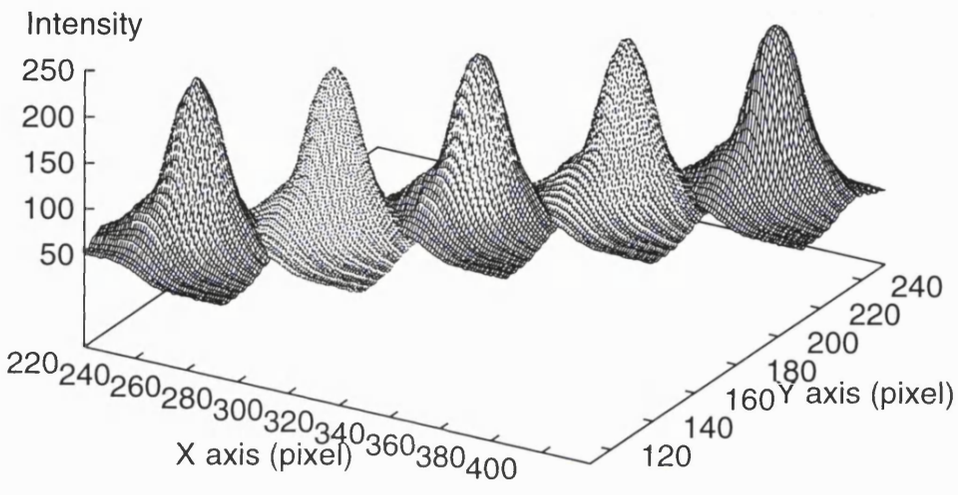


Figure 6.18: CCD image of the pinhole image moving across the frame when the collimator is tilted. Each 3-D intensity plot corresponds to one volt step in the actuator voltage. The background is not subtracted.

collimator motions and detector shifts in the focal plane of the ISIS camera (eq. 4.5) and a collimator error of 0.2-0.3μm, the resulting error at the detector for the largest movements will be between 1.2 and 1.9μm. These position errors are still less than the target value of 1/10 of a pixel (2.4μm). Average values can actually be expected to be significantly less than that because the collimator will be normally tilted by much smaller angles (not by its full range).

6.5 Optical Tests: Mirror Quality

Some of the optical tests of the collimator were designed to measure the optical characteristics of the mirror and to check that the mechanical support of the collimator was not affecting the accuracy of the optical surface. Therefore the collimator had

to be tested *inside* the cell and the results compared with those previously obtained with the mirror on its own.

6.5.1 Knife-Edge Test

The knife-edge test [119] is based on a simple principle: a luminous source is re-imaged by the optical element that needs testing. When the sharp edge of a knife is used to block some of the rays in the illuminating beam that forms the image, a shadow pattern appears over the otherwise uniformly illuminated optical surface. The shadow pattern carries information about the aberrations and irregularities of the surface under test. In the case of an off-axis paraboloid collimator, the setup for a knife-edge test is analogous to the one in fig. 6.10. The difference in this case is that the surface under test is the collimator, and not the flat mirror. The knife-edge test can be used i) to measure the curvature of the mirror or ii) as a null test, for checking that all the zones of a supposedly spherical wavefront have the same radius of curvature.

In the first case (figure 6.19), the knife position is varied with respect to the focal plane of the mirror. When the knife edge is introduced inside the focus of a 'perfect' mirror, the shadow pattern consists of a dark and a bright region, sharply separated along a line. When the knife is outside the focus, the same pattern is repeated, but the dark and bright areas are swapped between each other. When the knife is introduced at the focus, the mirror darkens suddenly, without any apparent motion of the shadow pattern [122]. This allows to accurately measure the focus position. As a null test, the knife is inserted exactly at the focus: if the mirror surface has some irregularities, the shadow pattern will not be completely dark, but will have dark and bright areas, depending on the slope of the mirror at each point. A knife-edge null test was performed on the collimator within the cell to check for surface distortions. The knife image, when compared with the one measured with the collimator outside the cell, did not show any detectable changes.

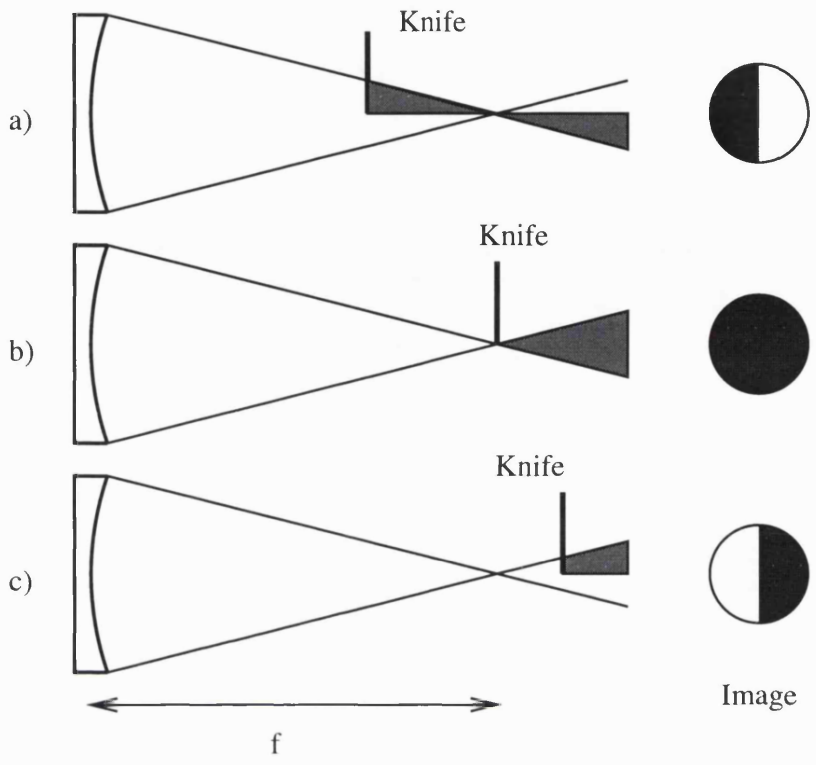


Figure 6.19: Using a knife-edge tests to determine the focus position of a parabolic mirror. When the knife is placed before the focus (a), only half of the mirror is seen illuminated, while when the knife is after the focus (c), the other half or the mirror is bright. When the knife is introduced *at* the focus (b), the image suddenly darkens.

6.5.2 Scatter Plate Interferometric Test

The scatter plate interferometric test differs from the knife-edge test in that it can quantitatively measure errors and distortions of the mirror surface. The scatter plate interferometer belongs to the family of common-path interferometers, so called because the reference wavefront and the one under measurement follow the same optical path. This is very important for the accuracy of the test: in this setup the common path ensures that air turbulence, vibrations and other environmental effects do not change the fringe pattern. Both wavefronts, following the same path, will be affected in the same way, so that these unwanted effects will cancel out when the fringe pattern is formed. The experimental setup for the scatter plate test is shown in fig. 6.20.

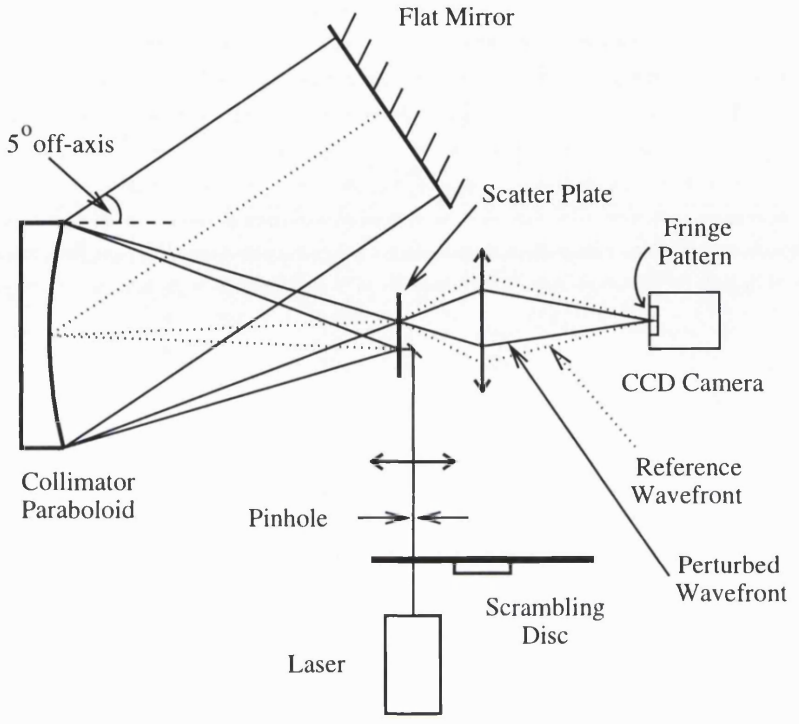


Figure 6.20: Experimental setup for the scatter plate interferometer. A CCD camera was used to record the fringe pattern. The interference is produced after the second passage through the scatter plate by a reference wavefront (generated by scattering of a laser beam) and the wavefront under measurement. The scrambling disc is introduced to eliminate laser speckle, which would disrupt the fringe pattern. A camera objective is used to focus the fringe pattern on the CCD.

6.5.2.1 Scatter Plate Interferometry

In the scatter plate interferometer, a laser beam is directed onto a transmitting scattering plate. In my experiment the scatter plate consisted in a photographic image of laser speckle. This produces a randomly distributed set of intensity maxima within the emulsion of the plate which can be used as scattering centres [120]. When the laser hits the scatter plate a fraction of the light is simply transmitted, while the rest is scattered. The purpose of the scatter plate is therefore to generate a spherical wavefront which uniformly illuminates the optical surface under measurement. The wavefront generated at the scatter plate then hits the surface under test (the collimator) and is slightly distorted by the surface irregularities. At the

same time, a fraction of the laser beam propagates unperturbed as a narrow pencil of light through the scatter plate and the rest of the optics. After the reflection from the flat mirror, the testing wavefront is refocused by the collimator. Provided that the scatter plate is in the focal plane of the collimator, the testing wavefront will be refocused on the scatter plate.

The laser beam will also hit the scatter plate again. This is the basic idea behind the common-path interferometer: the laser follows the same path as the testing beam and provides a reference wavefront when it hits the scatter plate a second time. The scatter plate is made of two specular photographic images of the same speckle pattern, so if the system is properly aligned, the laser beam will be scattered by the same scattering points to produce a reference spherical wavefront. This reference wavefront will be identical to the testing wavefront but for the fact that it has not been distorted by irregularities in the mirror surface. The two wavefronts can then interfere with each other and generate a fringe pattern. The interference fringes are produced by defocus and tilt between the two spherical wavefronts and they form a circular pattern. In the test they appear straight and parallel, but they are just concentric circles with a very large radius. Only when the scattering point is exactly in the focal plane the defocus is removed and the fringes are truly parallel.

In this test the interference pattern is generated by the the light which is scattered only once by the scatter plates. The light which is not scattered by either passage through the plates gives rise to a bright spot in the centre of the aperture. The light which is scattered twice simply adds to the background level and decreases the contrast in the fringe pattern [121]. In the setup for the test of the collimator paraboloid, the surface under test is used in double pass, so the testing accuracy is effectively doubled. The generation of fringes from the interference of the two wavefronts can be calculated. The intensity of the reference (unperturbed) spherical wavefront (R) can be expressed in complex notation as follows [122]:

$$f_R(r) = ae^{ikr} \tag{6.18}$$

Here $k = 2\pi/\lambda$ is the wave vector normal to the wavefront, r is the radius of the wavefront and a the amplitude. If we indicate with $w(x, y)$ the deviation from an ideal spherical surface due to the distortions on the testing surface and tilt and

defocus between the wavefronts, then the testing wavefront (T) can be written as:

$$f_T(r,x,y) = be^{ik[r+w(x,y)]} \tag{6.19}$$

The intensity of the fringe pattern will be:

$$I(x,y) = |f_R + f_T|^2 \tag{6.20}$$

Consequently:

$$I(x,y) = 1 + \gamma \cos kw(x,y) \tag{6.21}$$

Where:

$$\gamma = \frac{2ab}{a^2 + b^2} \tag{6.22}$$

The interference produces a pattern of alternating bright and dark fringes. The parameter γ is called fringe visibility, and it is maximum when the amplitudes of the two wavefronts is the same. In this discussion the contribution from the light scattered twice by the plates is not considered. The function $w(x,y)$ will determine the shape of the fringes, so fringe analysis software can reconstruct the mirror surface directly from the fringe pattern.

6.5.2.2 Experimental Setup

In the practical setup of fig. 6.20 I used a low power He-Ne laser source. I introduced a translucent scrambling disc to eliminate laser speckle and further reduce the laser intensity. The intensity of the laser beam is an important issue because the double transmission of the beam through the scatter plate produces a very bright spot at the center of the fringe pattern. The size and intensity of this spot need to be minimized because it destroys the fringe pattern in the centre of the aperture. The pinhole placed after the scrambling disc has exactly this purpose. A small lens is used to focus the beam on the scatter plate. The scatter plate itself is mounted on a micrometric carriage for fine adjustment of the position. The collimator is placed in the standard auto-collimation setup in which the parallel beam reflected off the collimator hits a flat mirror, is reflected back and refocused by the collimator. The ISIS collimator is an off-axis paraboloid with a angle of 5°, so the the angle between the flat, the collimator and the scatter plate needs to be set to this value. I found that the scatter plate test was extremely sensitive to this angle and that great care

was needed to accurately match it with the off-axis angle. If the angle was only a fraction of a degree bigger or smaller than the correct value, optical aberrations (astigmatism) were introduced.

I used a reference laser beam to finely adjust the position of the collimator and of the flat in the setup. I found that the vertical (i.e. orthogonal to the plane of the off-axis angle) tilt of the collimator was also a critical parameter. A CCD camera recorded the fringe pattern which was then saved as a screen video file with a frame grabber. Wyko Phase II software was used to analyze the fringe pattern and to determine the mirror shape. The use of the CCD camera and the Phase II software for recording the fringe pattern allowed me to increase the fringe contrast by subtracting some of the background light. The software can also eliminate the central area from the fringe pattern, thus reducing the impact of the central bright spot in the image. This limits the surface measurement to a ring, rather than the whole aperture. The central obscuration however, is quite small (its diameter is roughly 1/5 of the full aperture) and most of the mirror surface is included in the measurement.

Once the system is properly aligned and the scatter plate suitably tilted, the fringe pattern consists of a series of parallel fringes equally spaced. The density of fringes can be varied by changing the tilt of the two wavefronts; this is done by moving the position of the scatter plate. I found that by reducing the number of fringes in the image the accuracy of the analysis software could be improved. This is because the software analyzes the pattern to find the center of each fringe, so when the fringe is thicker, the error is smaller. The collimator was used in double pass, which meant each bright fringe is effectively separated by a quarter of a wavelength in terms of of surface inaccuracies (Fringes are normally separated by half a wavelength, since the maxima of $I(x,y)$ correspond to $w = \lambda/2$). In fact, in this case the function $w(x,y)$ is determined by two reflections at the mirror, thus a surface feature that stretches from one bright fringe to another corresponds to a bump or a hole a quarter of a wavelength deep ($158.2nm$ at $\lambda = 632.8nm$).

6.5.2.3 Collimator Surface Measurements

The collimator mirror was tested for surface accuracy both inside and outside the cell. The measurements outside the cell give the figuring accuracy of the mirror

and a reference to which the measurements inside the cell could be compared. A typical fringe pattern from the scatter plate interferometer is shown in fig. 6.21. The fringe pattern is analyzed via software to determine surface inaccuracies. The software produces a 3-D map of surface deviation from the paraboloid shape (or more precisely, the deviation from an ideal spherical wavefront in the testing beam). Wyko fringe analysis system was the software tool used: the program gives a peak to valley and r.m.s. error in surface accuracy after subtracting wavefront tilts. For the mirror outside the cell, the software produced a peak to valley error of $0.077 \pm 0.018\lambda$ or $49 \pm 11nm$. This is better than the specified value of 1/10 of a wavelength. The surface error map is shown in fig. 6.22. The result I got for the surface accuracy of the mirror outside the cell is also consistent with the measurements made by D. Brooks one year earlier after finishing the figuring: he obtained a peak to valley error of 0.067λ. I then put the collimator back in the cell and I repeated the test above with the actuators at zero expansion. The result is shown in fig. 6.23. The peak to valley error amounts to $0.063 \pm 0.014\lambda$ or $40 \pm 9nm$, which is, within the errors, the same as obtained with the mirror outside the cell. It is also interesting to note that the shape of the distorted wavefront is very similar to the one of fig. 6.22. These results provide a convincing proof that no additional stress is introduced on the collimator as a result of the mechanical mounting inside the cell.

As a further test I checked that by pushing the actuators to the maximum expansion the surface accuracy remained the same. I tried both pushing all three actuators at the same time or just one to maximum expansion (figure 6.24 shows the surface errors when $V1 = 8V$). The problem is that, particularly in the latter case, the optical alignment of the test is disrupted by the change in position of the collimator. In the first test, when all the actuator are expanded by the same amount, the collimator surface is simply translated parallel to the optical axis. Nonetheless, even this simple movement slightly disrupts the optimization of the scatter plate test. The change was small enough to be able to measure the surface accuracy without readjusting the setup, however. In this case I obtained a peak to valley error of $0.104 \pm 0.016\lambda$ or $66 \pm 10nm$. This result is only slightly worse than the one for the actuators at zero expansion and might simply be due to the fact that the testing setup was not re-optimized for the new collimator position. Even so, the accuracy is still within specification and since the alignment procedure is very time consuming, I did not

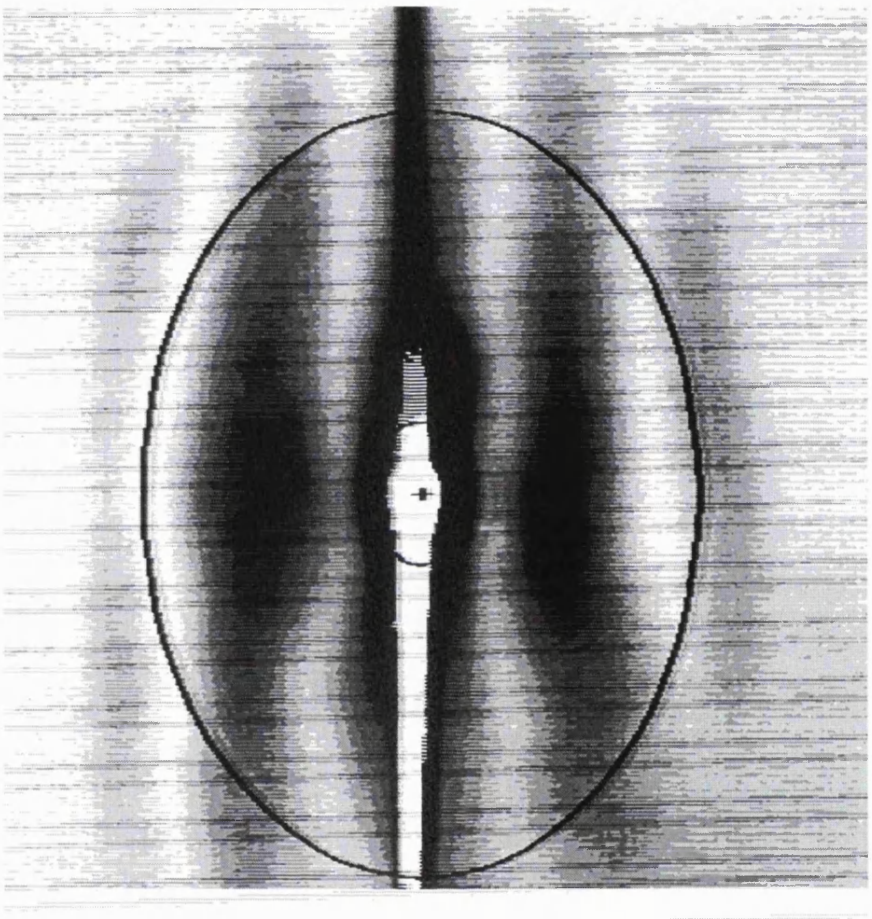


Figure 6.21: Video image of the fringe pattern from the scatter plate interferometric test. The data was recorded using a CCD camera and analyzed with Wyko fringe analysis software. Each fringe is effectively separated by a quarter of a wavelength or $158.2nm$ in terms of surface measurements.

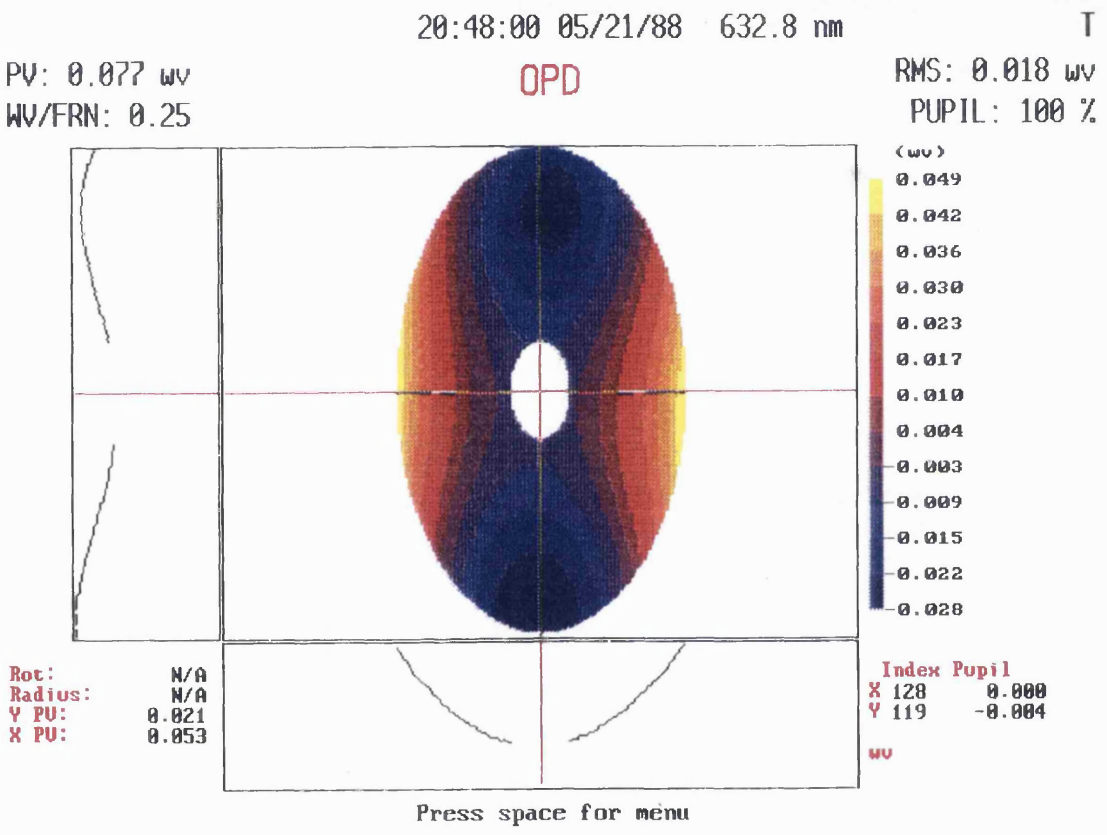


Figure 6.22: Map of the optical path differences over the collimator surface when the mirror is tested outside the cell. The peak to valley error over the whole surface is 0.077λ or $49nm$. This is consistent with earlier measurements soon after polishing.

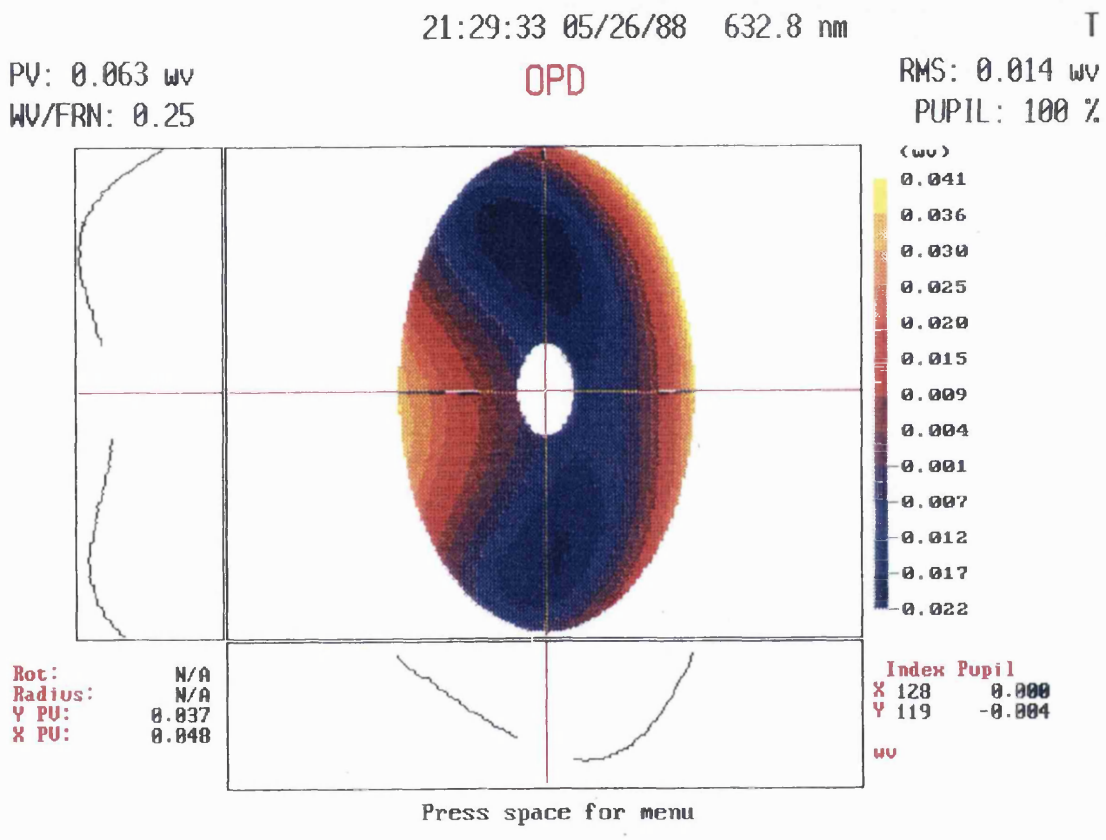


Figure 6.23: Map of the optical path differences over the collimator surface when the mirror is tested inside the cell. The actuators are all set at zero expansion ($V_1 = V_2 = V_3 = 0$). The peak to valley error over the whole surface is 0.063λ or $40nm$.

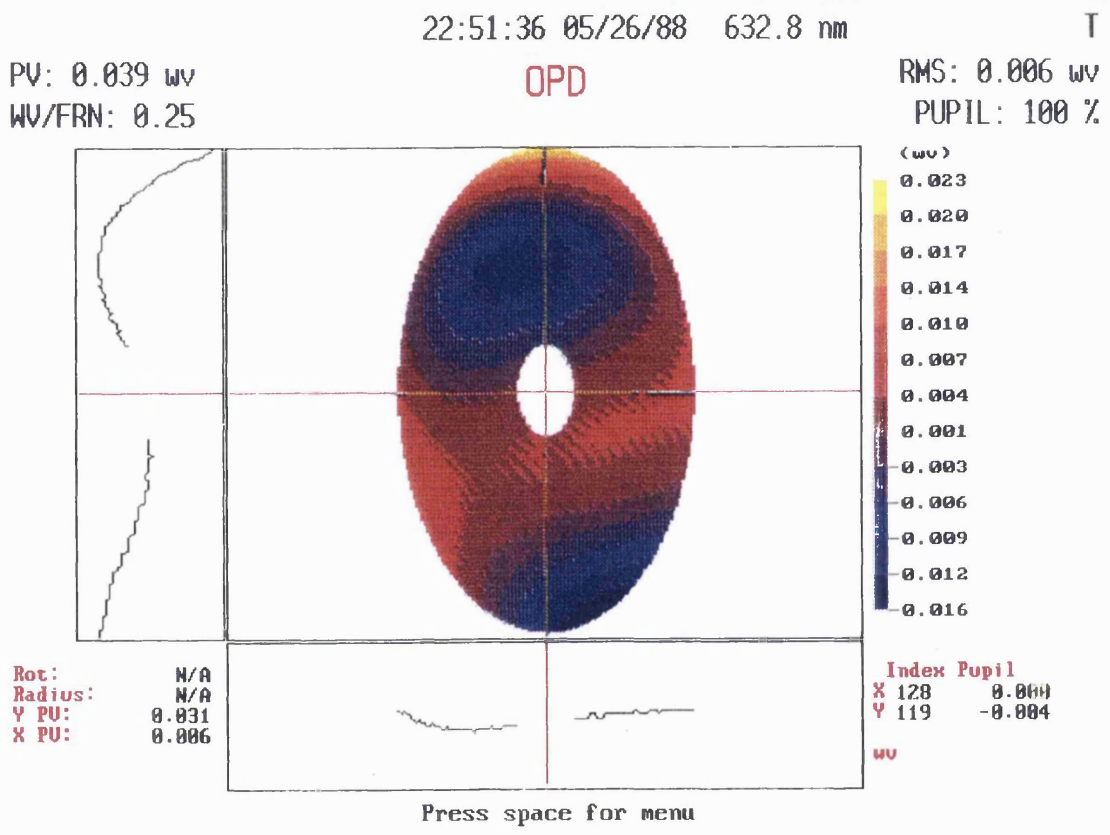


Figure 6.24: Map of the optical path differences over the collimator surface when the mirror is tested inside the cell with one actuator pushed to maximum expansion ($V_1 = 8V,\ V_2 = V_3 = 0$). The peak to valley error over the whole surface is 0.039λ or $25nm$.

consider necessary to try to further improve on this result. In the test when only one actuator is expanded, the change in the alignment is such that re-optimization is unavoidable. The fringe pattern actually disappears after the actuator moves by around $4\mu m$, corresponding to a tilt of the mirror of $5arcsec$. This is also a more interesting test, because of the asymmetric movement of the collimator within the cell. Figure 6.24 shows the surface accuracy when one actuator is driven by $V = 8V$ and the others are kept to zero. The peak to valley error in this case was reduced to $0.039 \pm 0.006\lambda$ or $25 \pm 4nm$. This error is lower than the one taken with actuators at zero expansion, but it is probably due to a slightly better optical alignment.

6.5.3 Measurement of the Off-Axis Angle

The off-axis angle was measured by shining a laser normal to the collimator and detecting the reflected beam at long distance (to improve accuracy). To be sure the collimator and the laser were in the correct positions (that is the beam was normal to the surface), I used the auto-collimation setup of fig. 6.20 to check that beam was really reflected at the off-axis angle. Once the collimator is in auto-collimation, the knife-edge test can be used to check the astigmatism produced by the misalignment. By minimizing astigmatism, it is ensured the mirror is placed at the true off-axis angle. I then removed the flat mirror that was folding the beam back and I measured the distance between incident and reflected beams $2.8m$ away from the collimator. The off-axis angle proved to be $\alpha = 5°10' \pm 2'$, which is within the specified value.

6.6 Cold Room Tests

As a final laboratory test of the collimator performance, I wanted to check how the system responded to changes in the room temperature, to simulate the environmental conditions on the telescope in La Palma. In particular, since the temperatures on the mountain can be as low as $-10°C$ (see chapter 4), I needed a cold room for the experiment. I was allowed to use the facilities at the Royal Greenwich Observatory, where a large cold room is in operation. The test was aimed at verifying that the expansion of the actuators and the functioning of the electronics was not affected by the low temperatures. For practical reasons, it was not possible to test the mirror

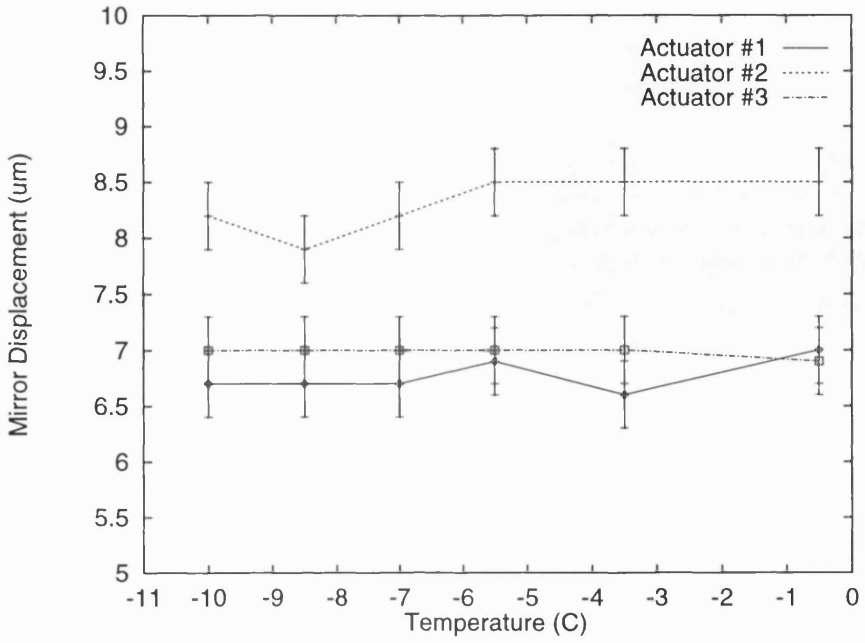


Figure 6.25: Variation (in μm) of the dummy mirror movement for temperatures between $0°C$ and $-10°C$. LVDT sensors were used to measure the motion of the dummy mirror inside the cell. A thermometer/hygrometer monitored the temperature and relative humidity changes.

optically inside the cold room, so I had to rely on the less accurate mechanical test with the dummy aluminium mirror and LVDT sensors. The test was therefore more a qualitative rather than a quantitative one. Nonetheless, the test was significant in that no measurable change in performance was detected over more than $10°C$ temperature change. The variation in the dummy mirror movement inside the cell as a function of the room temperature is shown in fig. 6.25. A thermometer/hygrometer monitored the temperature and relative humidity changes. The graph shows that over the temperature range covered the motion remained practically unaffected. Unfortunately I could not extend the range even further because of the time it takes for the cold room to change its temperature. It is worth noticing that the relative humidity levels were quite high: between 50% and 60%. This did not affect the performance of the actuation system either.

Chapter 7

Flexure Compensation in ISIS

7.1 Introduction

In July 1995 we were given a week of telescope daytime to test ISAAC at the William Herschel Telescope (WHT). During this week the active collimator needed to be installed and calibrated, reducing the time available for flexure compensation testing. Moreover telescope maintenance work had priority over the testing, limiting our time even further. The whole experiment, including installation and calibration had to be carefully planned in advance to avoid wasting time. For these reasons, the instrument was designed in a modular format (collimator cell, electronics rack, etc.) with simple connections between each component. The software was also written so as to be easy to modify, with a very simple user interface and with key calibration parameters stored in separate files. In my visit to the telescope I was accompanied by Mark Dryburgh, the mechanical engineer and, a few days later, by Dr. David Walker, head of OSL.

The WHT telescope has a $4.2m$ primary mirror, $f/11$ focal ratio, with an altazimuthal mount. ISIS sits at the Cassegrain focus, connected with a flange to a rotating turntable at the back of the primary mirror support. The picture in fig. 7.1 shows ISIS mounted on the telescope. Both ISIS and the telescope are controlled from Vax workstations. Their operation is fairly intuitive and, in the case of ISIS, it is facilitated by an interactive ("Mimic") display. The display shows the position and status of all the major components in the spectrograph and in the acquisition

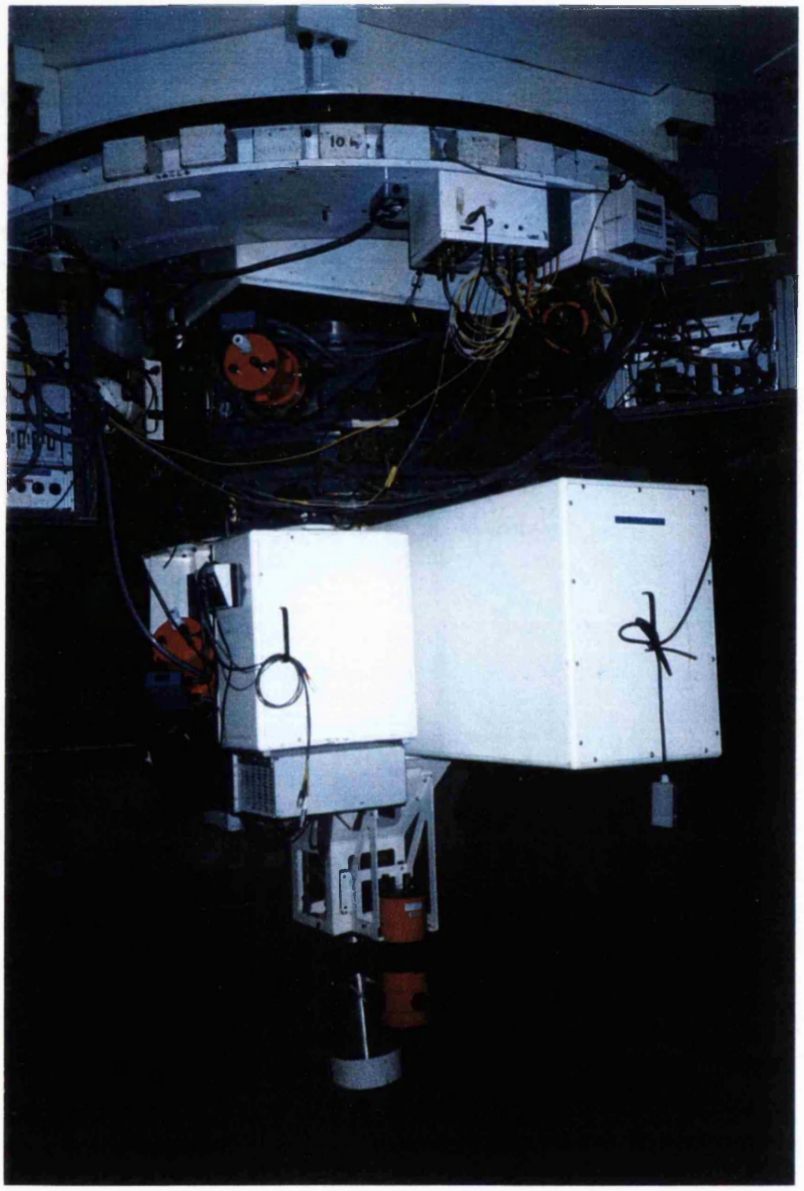


Figure 7.1: Picture of ISIS mounted at the Cassegrain focus of the William Herschel Telescope. The white rectangular box in the front on the right is the external case containing the two collimator arms of the spectrograph.

and guiding area. By simple direct commands it is possible to configure and operate the spectrograph from a single console. After getting acquainted with this software, Mark and I were able to operate both the telescope and the spectrograph without assistance.

7.2 Installation of ISAAC

From the point of view of mechanical installation, ISAAC consists of two main parts: the collimator cell and the electronics box. In both cases the installation on the spectrograph was fairly straightforward. The mirror cell could fit in the mount for the existing cell and it did not require any particular mechanical adjustment. A picture of the housing of the two collimators in ISIS is shown in fig. 7.2. The electronics is housed in a standard "19 in" rack-size case, so it could be placed in an empty slot in the existing racks on the Cassegrain turntable. Some care is needed in controlling the weight of this case not to unbalance the weight distribution on the turntable. However, after discussion with the telescope staff, it became clear that was not a problem either. To be able to run connecting cables from the collimator cell to the control electronics and to avoid modifying the existing spectrograph external case, we also built a new cover panel for the collimator arms. A small connection box was inserted in this panel, so that the cover panel itself could be detached from both the collimator cell and the control electronics. Six LEMO coaxial cables carried the signals for the actuators and the strain gauge sensors to the control electronics. This in turn is connected to the mains supply on the telescope turntable and to the computer room via a RS-232 serial link. During the experiment on the telescope, we used a spare connection normally reserved for CCD control electronics.

7.2.1 Alignment and Focusing

The normal procedure for aligning the collimator position involves adjusting the detector and collimator tilts with fine threaded screws and then the focus with the Hartmann shutters. However, since the collimator and the detector were already aligned before putting the active collimator in place, no adjustment of the CCD tilt was necessary. Therefore, after the collimator was replaced, I observed the

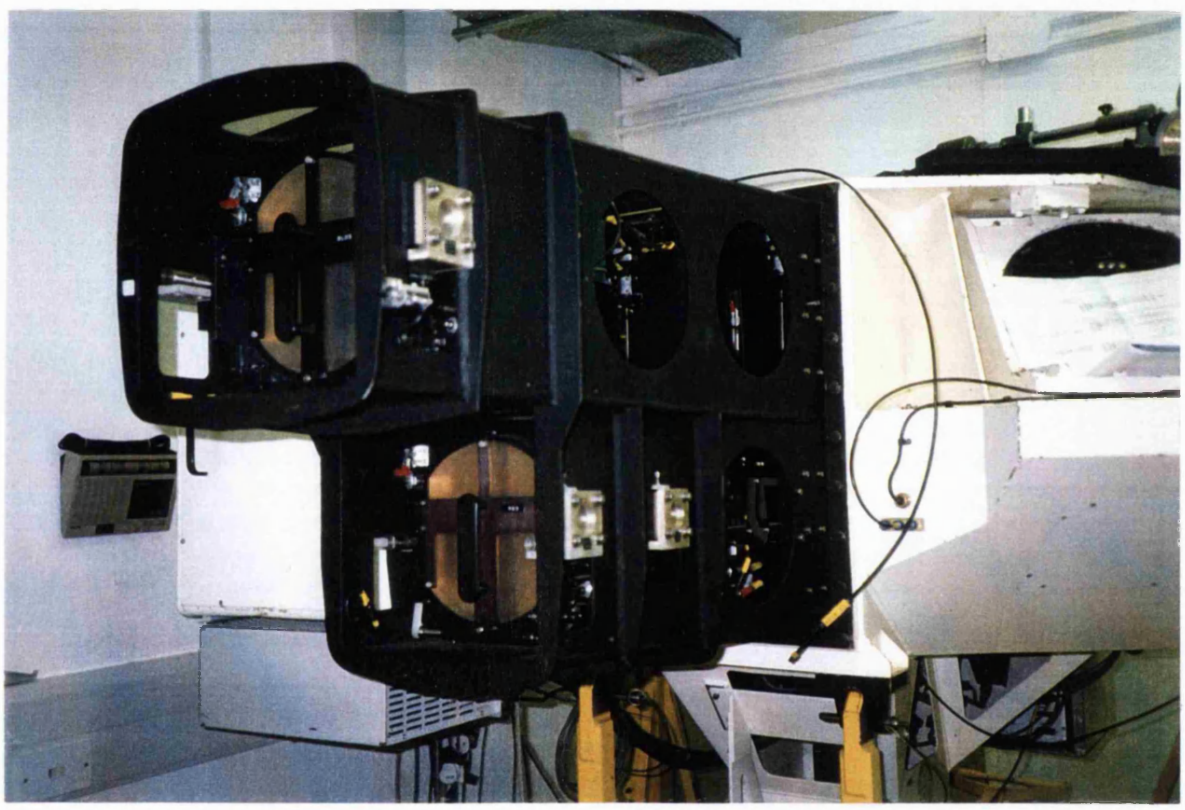


Figure 7.2: Picture of the support system for the collimators in ISIS.

spectrograph slit illuminated by a continuum tungsten lamp. The match between the active collimator cell and the previous one was good enough so that the slit was already imaged on the CCD, even if near one corner. This facilitated the alignment of the collimator, since it was possible to tilt the mirror cell with the appropriate screws and immediately check the effect on the CCD image. After a few adjustments, the image of the full slit was roughly centered on the CCD (that is the detector was almost uniformly illuminated).

Then, a finer centering of the slit on the detector was performed separately in spectral and spatial direction. In the former case, which corresponds to left/right tilts of the collimator cell, the central wavelength of the grating was set to coincide with a strong emission line of a CuNe lamp. The position of the collimator could then be changed until the line was in the center of the CCD. For the spatial direction centering (i.e. top/bottom collimator tilt) the slit was observed through a narrow dekker and illuminated with the tungsten continuum lamp. This produces a narrow bright line along the grating dispersion, which can be centered in the same way as a spectral line.

Once the collimator is properly aligned, it can be focused by moving the whole cell parallel to the optical axis with stepper motors. The exact focus position can be measured with a Hartmann test. The principle of focusing with an Hartmann shutter is illustrated in fig. 7.3. In this test one (or several) spectral line is imaged on the CCD detector; a shutter then covers half the collimator alternatively and the position of the line is recorded. If the collimator is out of focus, the position of the same line will be different in the two cases. If instead, the collimator is at the exact focus, the line will not move. This happens because the rays contributing to one wing of the line are blocked by the shutter, and the centre of the line is effectively displaced. By measuring the shift of the line, the error in focus can be calculated. On ISIS this is done automatically by a software program that receives the left and right Hartmann shutter images as an input and it gives the Hartmann shifts and the required collimator position in the output. The program requires three spectral lines on which to calculate the shifts. After a few iterations, I obtained Hartmann shifts of $X_H = 0.04$, 0.05 and -0.02 pixels on the three lines. These shifts are small enough to consider defocusing negligible. The Hartmann test can also calculate the

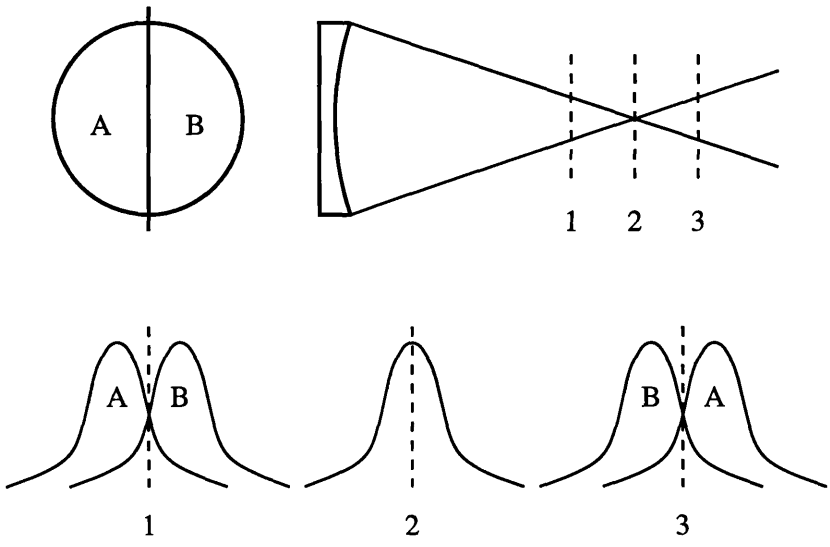


Figure 7.3: Hartmann shutter procedure for finding the focus position of the spectrograph collimator. The shutter covers alternatively half the mirror while observing a spectral line (positions A and B). If the collimator is not in focus, the line will be shifted by switching from A to B.

collimator tilt (in spectral direction) by comparing the shift of a line at the center of the detector with one at the edge. This test measured a left/right tilt of 0.072^{o} ($4arcmin$). This result could also be produced by a curvature of the measured wavefront. However, I considered this tilt to be small enough to be neglected.

When the collimator is aligned and focused, the spectrum of a calibration lamp with a narrow slit gives an indication of the optical quality of the system. Alternatively, if the imaging quality of the spectrograph is known, the spectrum shows focusing errors. To check the correctness of the Hartmann tests and to compare the optical quality of the two collimators, I took a 20-second exposure of CuAr and CuNe calibration lamps with a slit $116\mu m$ wide ($0.5arcsec$ projected on the sky); the result is shown in fig. 7.4. If a spectral line is observed out of focus or the collimator introduces significant optical aberrations, the line will be broadened. For the ISIS spectrograph the minimum line width for a slit around $100\mu m$ wide, is 2 pixels [99]. As shown in fig. 7.4, the results with the active collimator are consistent with this value, indicating both that the collimator is not optically worse than the previous one and that it has been properly aligned.

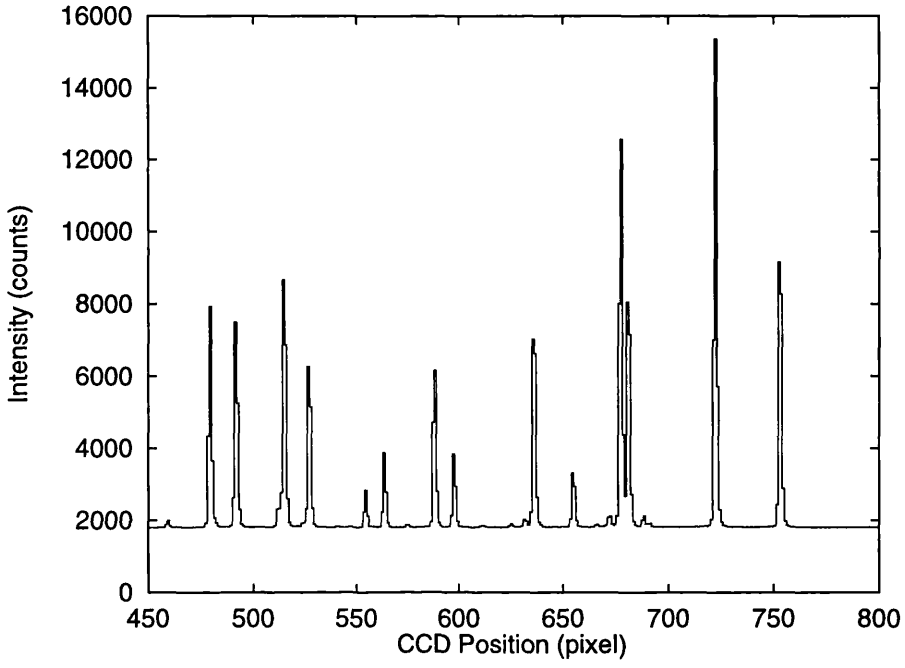


Figure 7.4: Spectrum of CuAr and CuNe calibration lamps with a narrow slit $116\mu m$ wide and the active collimator. The average line FWHM is around 2 pixels, as expected for a correctly focus spectrograph with the other collimator.

7.3 Calibration

The ISAAC system, as currently implemented, is designed to compensate flexure in open loop, using information from a flexure map. In an open loop system the motion of the collimator needs to be calibrated very accurately because, unlike the closed loop, it has not an optical reference which might detect slight calibration errors (in practice even a closed loop system also needs accurate calibration because of the finite time interval between updates). Therefore, as explained in chapter 5, I left a few calibration parameters to be determined experimentally. For the calibration I chose the same spectrograph setup as the one used for most of subsequent tests. The data was recorded with a $150\mu m$ entrance slit ($0.70arcsec$ on the sky) and with the low dispersion ($158lines/mm$) grating centered around $\lambda = 7000\mathring{A}$. This central wavelength was chosen to provide a good S/N ratio in the calibration lamp spectra (and so reduce the exposure time). All the data were recorded using the same 1024×1024 Tektronix CCD with $24\mu m$ pixels (La Palma Tek1 CCD).

The reference spectra were generated by observing an arc spectrum through a narrow

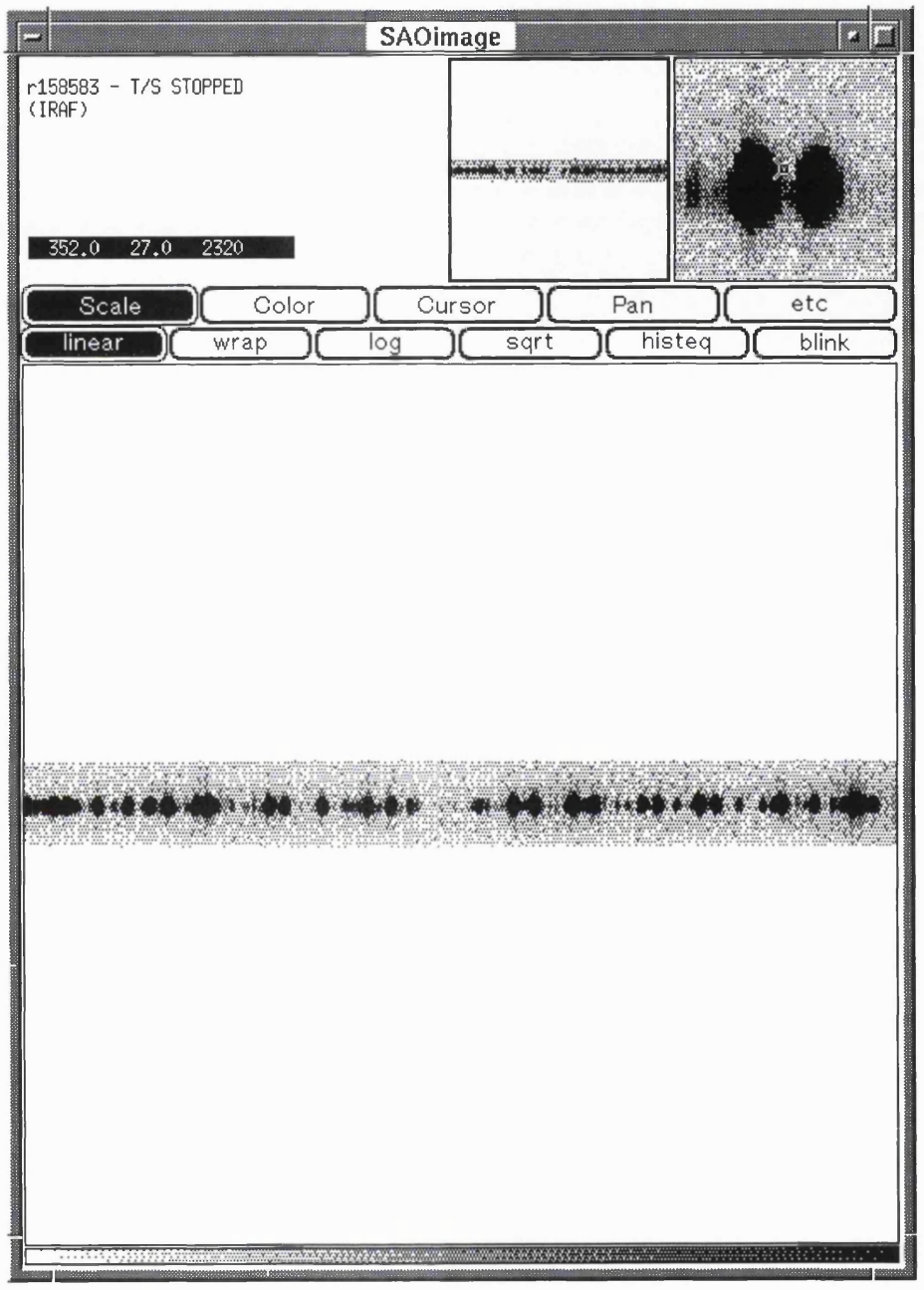


Figure 7.5: CCD image of a calibration lamp spectrum observed through a narrow dekker. Only a fraction of the slit is shown in the image. The dekker limits the image extension in spatial direction.

dekker, so that each spectral line would produce an almost circular spot. An example of such spectra, as recorded by the CCD, is shown in fig. 7.5. The detector shifts (in spectral and spatial direction) were calculated by centroiding a single line in the spectrum. I checked this procedure with other lines in the spectrum which gave the same result within 0.05 pixels, confirming that this system was as effective as cross-correlating the whole spectrum (and much quicker). The centroiding technique itself gives an error of ± 0.025 pixels on the shift measurements. The spectra were analyzed using IRAF data reduction software.

7.3.1 Movement Calibration

The calibration of collimator movements has two main aims: to ensure that the tilting axes of the collimator coincide with the spectral and spatial directions on the CCD and to adjust the scale factors so that, in each direction, the motion produced by the tilt is of the right amount. The relation between detector shifts and actuator movements was calculated in chapter 5. When calibrating the instrument, the inverse relations are needed because the detector shifts are measured and the correct values for the actuator voltage vs expansion coefficients need calculating. Inverting equations 5.20 and 5.21, we obtain:

$$\xi X = 2F_{cam}(\vartheta_j \sin\gamma + \vartheta_k \cos\gamma) \tag{7.1}$$

$$Y = 2F_{cam}(\vartheta_j \cos\gamma - \vartheta_k \sin\gamma) \tag{7.2}$$

Then combining these with equations 5.22 and 5.23 and using the same notation, the relation between actuator input voltages and detector shifts becomes:

$$X = \frac{2F_{cam}}{\xi b}\left(\frac{V_2}{K_2} - \frac{V_1}{K_1}\right)\cos\gamma + \frac{F_{cam}}{\xi h}\left(\frac{V_1}{K_1} + \frac{V_2}{K_2} - \frac{2V_3}{K_3}\right)\sin\gamma \tag{7.3}$$

$$Y = -\frac{2F_{cam}}{b}\left(\frac{V_2}{K_2} - \frac{V_1}{K_1}\right)\sin\gamma + \frac{F_{cam}}{h}\left(\frac{V_1}{K_1} + \frac{V_2}{K_2} - \frac{2V_3}{K_3}\right)\cos\gamma \tag{7.4}$$

The free parameters are the scale factors K_1, K_2 and K_3 and the rotation angle γ. To measure them it is necessary move the collimator by a known amount and see how this translates on the detector. The angle γ can be determined by applying a voltage V only to actuator #3 and measuring the (experimental) shifts X_o, Y_o on the detector:

$$\tan\gamma = \xi X_o / Y_o \tag{7.5}$$

The scale factor for the third actuator can be determined, too; the equation is derived by imposing the distance from the origin to equate the calculated value (approximating the anamorphic factor $\xi = 1$), that is:

$$\sqrt{X_o^2 + Y_o^2} = \sqrt{X^2 + Y^2} = \frac{2F_{cam}V}{hK_3} \tag{7.6}$$

We can therefore calculate K_3 as:

$$K_3 = \frac{2F_{cam}V}{h\sqrt{X_o^2 + Y_o^2}} \tag{7.7}$$

The other two scale factors can be determined in a similar way by applying voltages to the other two actuators. From the practical point of view, this implies a somewhat iterative process: first a few CCD frames were taken with the actuators in different positions, the data is analyzed and the free parameters calculated, then I took new exposures to check the result and repeated the same procedure until I obtained a satisfactory result. To check the accuracy of the calibration, the control software had a mode of operation in which I could input the coordinates on the detector directly. In this way the input value can be compared with the measured shift immediately, facilitating a finer adjustment of the calibration parameters. The calibration parameters I chose were $\gamma = 0$ and the following values for K_i:

$$\begin{cases} K_1 = 0.958 \\ K_2 = 0.764 \\ K_3 = 0.879 \end{cases} \tag{7.8}$$

The above values were those that gave the most accurate (although not perfect) calibration. To have an idea of the accuracy of this calibration, in table 7.1 the measured shifts are compared with the input values. The data in the table show there was a residual error I could not compensate with the calibration. When moving the image along the X axis (spectral direction), a small positive shift along the Y axis (spatial direction) was also introduced. Since this happened for both positive and negative motions on the X axis, it could not be corrected with a simple rotation of the XY axes. The reason for the residual error is difficult to establish. One possibility is that it may be linked to optical layout of the spectrograph, since the collimator tilt changes the angle of incidence on the grating and the off-axis configuration. This may introduce optical aberrations that shift the centre of the

Input Values (μm)		Measurements (μm)	
X	Y	X	Y
+40	0	$+39.6 \pm 0.6$	$+3.4 \pm 0.6$
−40	0	-39.4 ± 0.6	$+3.8 \pm 0.6$
0	+40	$+0.7 \pm 0.6$	$+41.8 \pm 0.6$
0	−40	-0.5 ± 0.6	-41.0 ± 0.6

Table 7.1: Calibration of the movements of the active collimator on the ISIS spectrograph. The X axis corresponds to spectral direction, Y axis to spatial. The input values from the user are compared with the measured shifts on the CCD detector.

line in spatial direction. I excluded the mechanical support of the collimator as a possible explanation, because the laboratory tests did not show such errors.

To compensate for these small errors implies increasing the motion of the actuator at the top. However, this in turn implies having an excessive motion when displacing the image along the Y axis. A compromise was therefore necessary: I decided to have some residual error and a slightly larger displacement along the Y axis (as table 7.1 shows). Moreover, the errors were small enough not to compromise the active collimator performance: the maximum error was just 0.16 pixels for nearly two pixels of image motion and being almost entirely in spatial direction, it is even less important.

7.3.2 Tracking Model Calibration

The tracking model calculates the spectrograph position during exposure given the equatorial coordinates of the observed object, the sidereal time and the required sky PA. The model is based on spherical coordinates transformations and it should not need calibration. However, apart from the necessity to check the correct functioning of the model, one parameter still missing to the model was the angle on the rotator corresponding to the slit direction (i.e. parallel to the slit length). The equation $\psi = \sigma - q$ that relates the turntable position angle ψ, the slit position angle on the

sky σ and the parallactic angle q, assumes the slit is aligned with the position 0^o on the rotator. As it happens, there is an offset Ω between the two positions, so that:

$$\psi = \sigma - \Omega - q \tag{7.9}$$

The value of this offset becomes apparent when, for example, the sky position angle is zero and the rotator position is compared with the parallactic angle. In this way I was able to determine that $\Omega = 91°37'$.

7.4 Flexure Measurements in ISIS

The first step towards the testing of ISAAC on the telescope was to study flexure in ISIS and to test the accuracy of the flexure model I developed. These measurements completed the ones I took on my earlier visit to the observatory and tested the hysteresis in the spectrograph. An automatic procedure for producing a flexure map that covers the whole range of telescope and spectrograph possible positions had already been developed by R. Rutten in La Palma. The procedure for obtaining a flexure map requires little interaction from the user: a program moves the telescope in elevation in 15^o steps from zenith to $z = 75^o$, then moves the telescope back to zenith, rotates the turntable by 45^o and repeats the elevation stepping. This is repeated for 360^o of turntable motion. A spectrum is acquired at every elevation/rotation setting. This not only speeds up the process considerably, but also assures that the data is taken always in the same way (with the same telescope motion and therefore same hysteresis).

7.4.1 Test of the Flexure Model

When the active collimator was properly aligned and its motion calibrated, I decided to take a flexure test. My flexure data was more than a month old, and was not a reliable basis for the flexure model. Figure 7.6 shows the raw data from the flexure test which I used to build the model for the look-up table. Figure 7.7 is the modified flexure map obtained with the procedure described in chapter 4. As explained there, the raw data coming from the flexure test is not a reliable model, because of the large hysteresis introduced by the "unnatural" telescope movements during

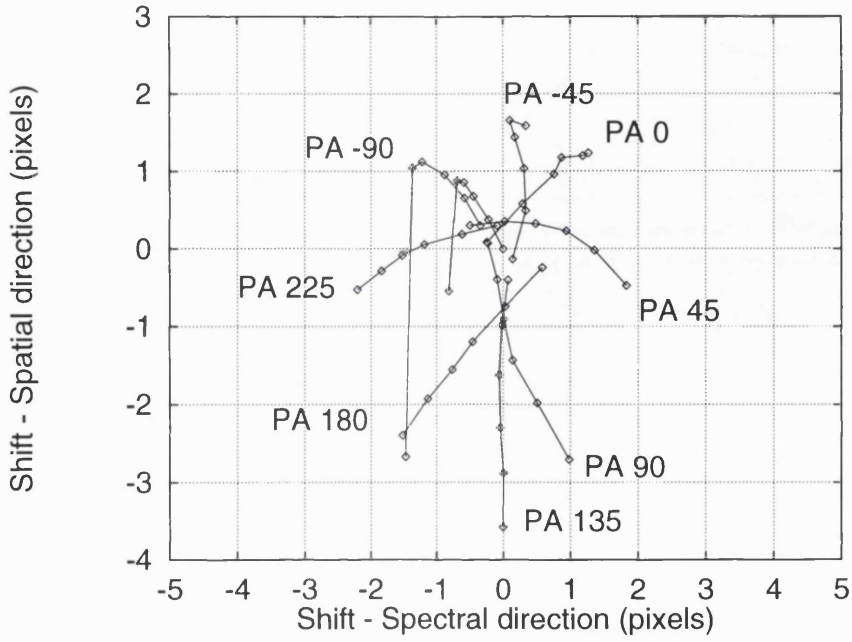


Figure 7.6: Flexure test on the ISIS spectrograph. Data from the blue arm with the $158 lines/mm$ grating at $\lambda = 7000\mathring{A}$. Each "arm" in the graph corresponds to a different turntable position angle (PA), while the points in one arm correspond to different telescope elevation. Note that the points corresponding to the telescope pointing at zenith (and different PA) do not coincide. This is attributed to hysteresis in the flexure measurement procedure.

the tests. During the testing at the William Hershel Telescope, it became clear that my assumptions concerning the spectrograph behaviour with respect to flexure were correct. In particular, I found evidence that the raw data coming from the flexure test could not be used directly for the active corrections because of the large hysteresis present. Fig. 7.8 shows three sets of data taken on three different days. Obviously flexure shows a reproducible pattern, even if translated in the CCD plane, but the data set, as it is, cannot be used for a flexure model. Moreover, since every "arm" of the map has a starting point at zenith (with a different turntable angle), those point would be expected almost to coincide, given that when the telescope is pointing at zenith a rotation of the Cassegrain turntable does not change the orientation of the spectrograph with respect to the gravity vector. The fact that they do not coincide can be attributed to hysteresis during the flexure measurement, since previous measurements [101] have shown that simply rotating the turntable

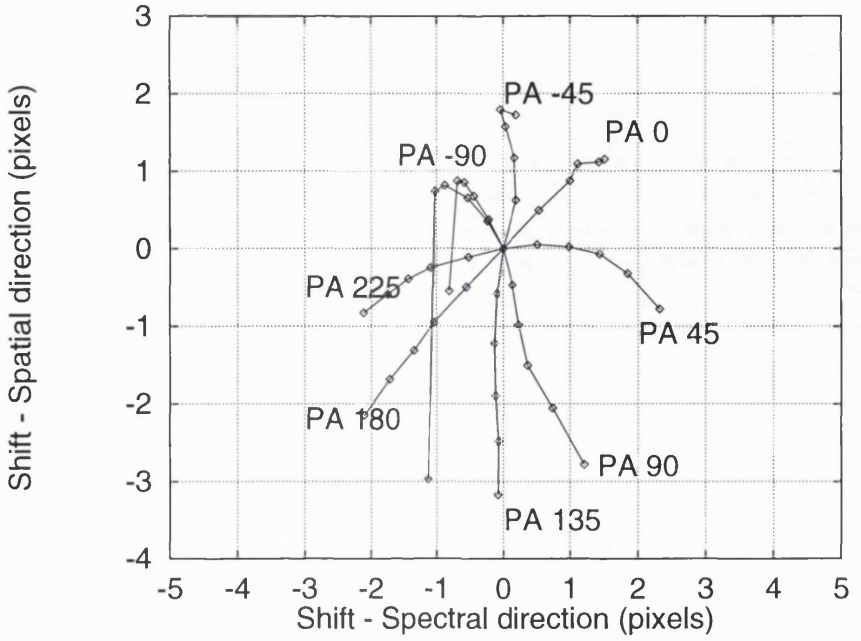


Figure 7.7: Flexure map used in the model for open loop flexure compensation. This map was obtained from the flexure measurement data by translating each arm of the graph so that the zenith positions coincide with the origin. This procedure greatly reduces the impact of hysteresis on the flexure model. The track for $PA = -90^o$ shows a sudden jump when elevation 15^o is reached. This is probably due to flipping over of a loose component in ISIS.

with the telescope pointing at zenith produces negligible flexure shifts.

The other important result I derived from the tests in La Palma is that the modified flexure map I had envisaged gives a much more reliable flexure model and eliminates the greatest part of the hysteresis introduced by the large movements of the telescope during the flexure test. Fig. 7.9 shows the equivalent graph to fig. 7.8 for the modified flexure maps. The figure shows that the modified flexure map is highly reproducible on a timescale of around one week. The only exception is the turntable position angle of -90^o where there was clearly something wrong in the ISIS structure, possibly a loose optical component which could suddenly flip from one position to another. Even if the reproducibility of the modified maps is quite remarkable if compared with the raw data from which they are derived, the data still show residual hysteresis which ultimately limits the accuracy we can expect from open loop corrections. On

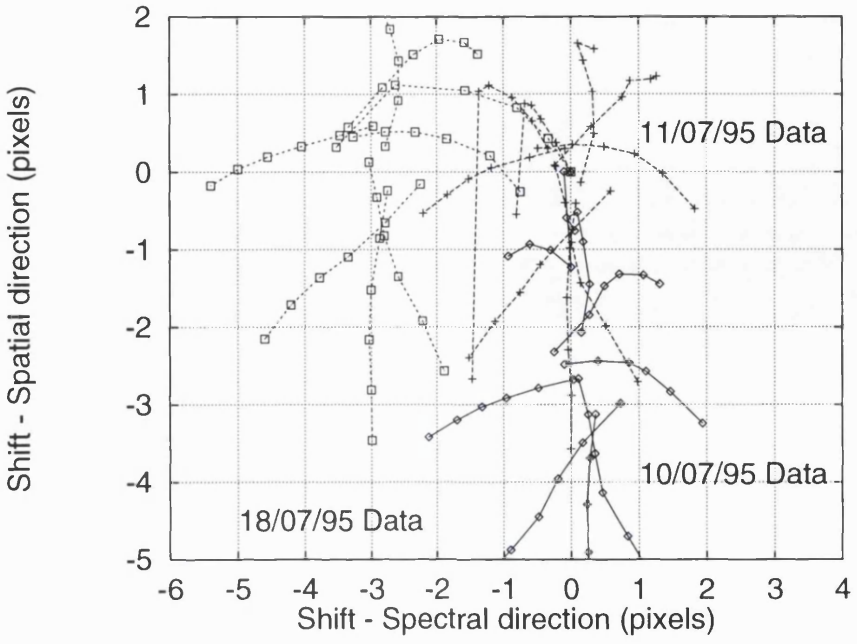


Figure 7.8: Raw data from flexure tests on three different days. The 10/07 and 11/07 data sets have exactly the same spectrograph setup, in the 18/07 set a different grating was used.

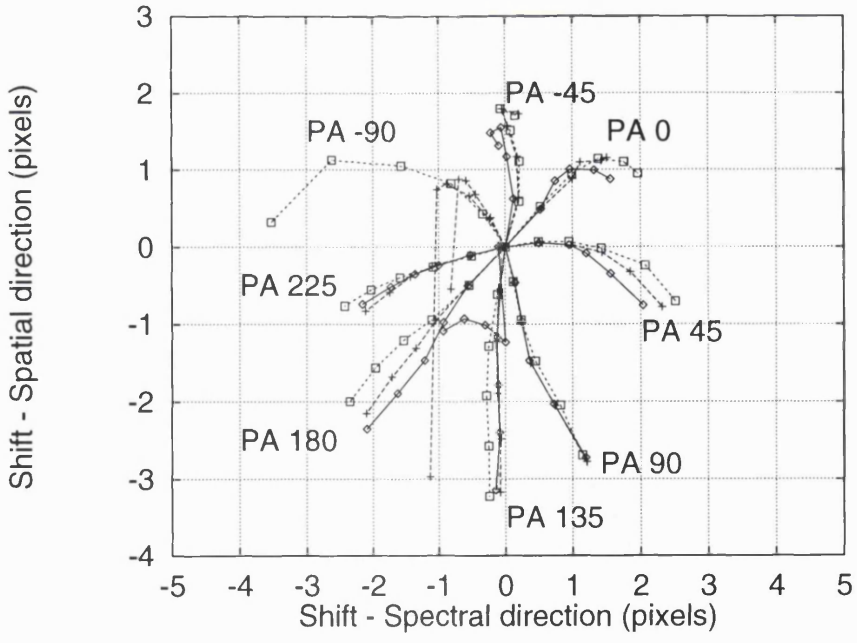


Figure 7.9: Changes in the flexure model after flexure tests in three different days. Apart from the turntable position angle of $-90°$ where there was clearly something wrong in the ISIS structure, the three maps show a remarkable reproducibility.

the other hand the choice of the ISIS setup seems less important than expected: the data set taken on 18th of July had a different grating and spectrograph setup, but the corresponding flexure map is still similar to the other two.

7.4.2 Model Analysis and Long-Term Stability

It is difficult to evaluate the information contained in a flexure map in a quantitative way and even such things as average flexure cannot be easily determined. In order to extract some information from the graphs, I produced some simulation software that allows to compare a map with another. The flexure map consists of a series of vectors in the plane of the detector identified by:

$$V_{ij} = (X_{ij}, Y_{ij}) \tag{7.10}$$

The index i identifies the telescope elevation (or zenith distance) and the index j the turntable position, X and Y are the spectral and spatial directions on the CCD. When the telescope tracks a star, the spectrum moves between these points. To extract a rough estimate of a mean flexure from the graph, I considered the four nearest points to each point in the model (see figure 7.10) and assumed the spectrum would move between these points. This motion does not correspond to a real star track, but it is useful to quantify flexure. The four nearest points are identified by the indices $(i+1, j)$, $(i-1, j)$, $(i, j+1)$ and $(i, j-1)$. Therefore, the spectrum shift between the starting point and, say, the first of the above points is another vector given by:

$$\mathbf{u_1} = \mathbf{V_{i+1,j}} - \mathbf{V_{i,j}} \tag{7.11}$$

By calculating the vectors u_k for all the points in the flexure map (excluding $z = 0^o$ and $z = 75^o$ as a starting point for simplicity), it is possible to produce an histogram with the number of occurrences of a set of spectrum shifts. The graph of fig. 7.11 shows one of such histograms for the map recorded on the 11/07/95 (figure 7.7). The graph simulates the effect of stepping the the telescope by 15^o in elevation and 45^o in the Cassegrain rotator angle. In this particular case the mean flexure shift produced is $\bar{u} = 0.73$ pixels ($\bar{u} = 17.5\mu m$), but 65% of the points are below mean. This value is close enough to the average flexure in ISIS in one hour of tracking (we measured an average $12\mu m/hr$), to give an indication of the flexure the spectrograph was

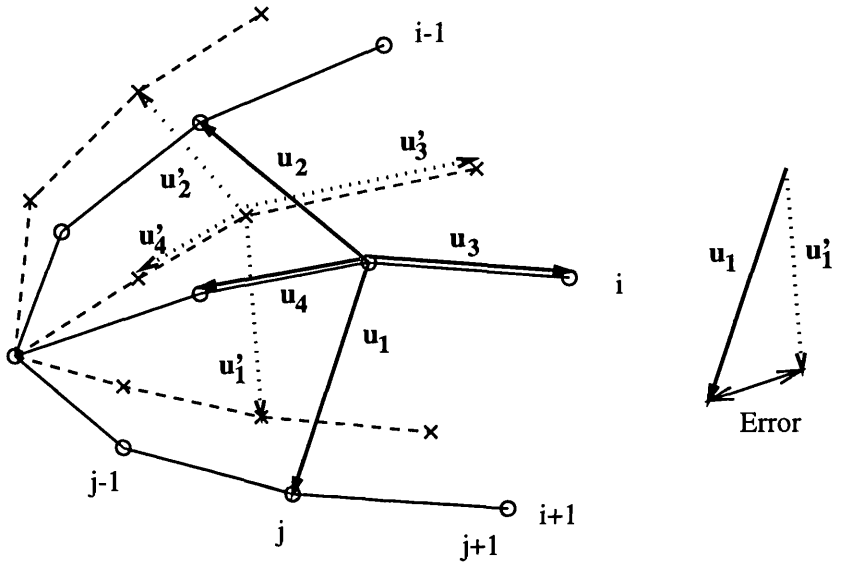


Figure 7.10: Mathematical model for a flexure map. Two slightly different maps are shown, together with the residual error introduced by using the first map to correct the flexure in the second. Each point in the map is identified by the indices (i,j) which define the turntable PA (i) and the telescope zenith distance (j). For the simulation four vectors $\mathbf{u_k}$ for each point were used to quantify the differences between the two maps. These vectors simulate a flexure shift when the telescope is tracking a star.

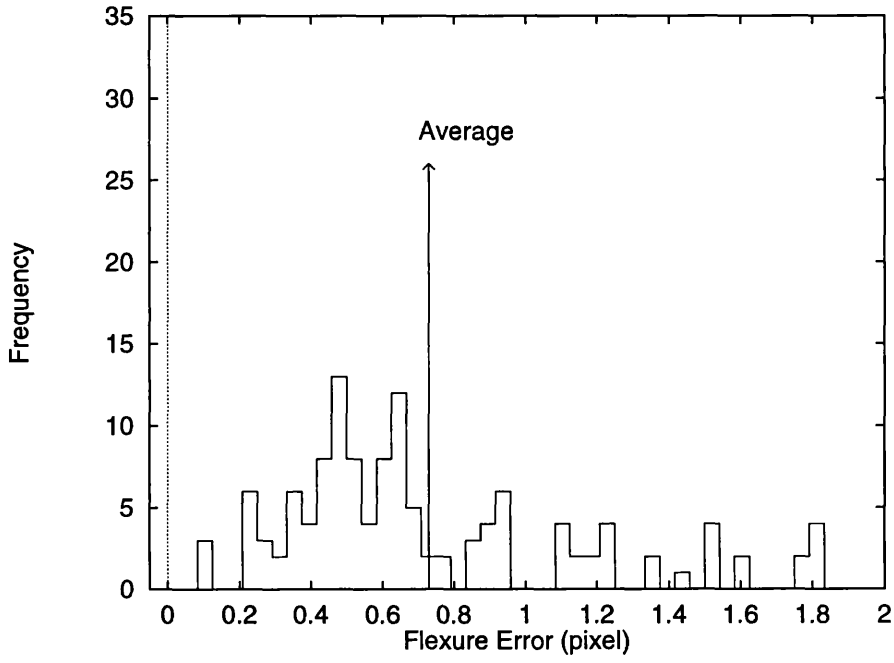


Figure 7.11: Distribution of flexure shifts obtained by stepping the telescope by 15^o in elevation and 45^o in the Cassegrain rotator angle. This is a way of estimating the level of flexure errors that can be expected from the spectrograph. The flexure model used was the one taken on 11/07/95, after modification. In the graph the mean flexure is 0.73 pixels ($17.5\mu m$), but 65% of the points are below mean. The equivalent telescope tracking time that would produce such flexure is $\sim 90 min$.

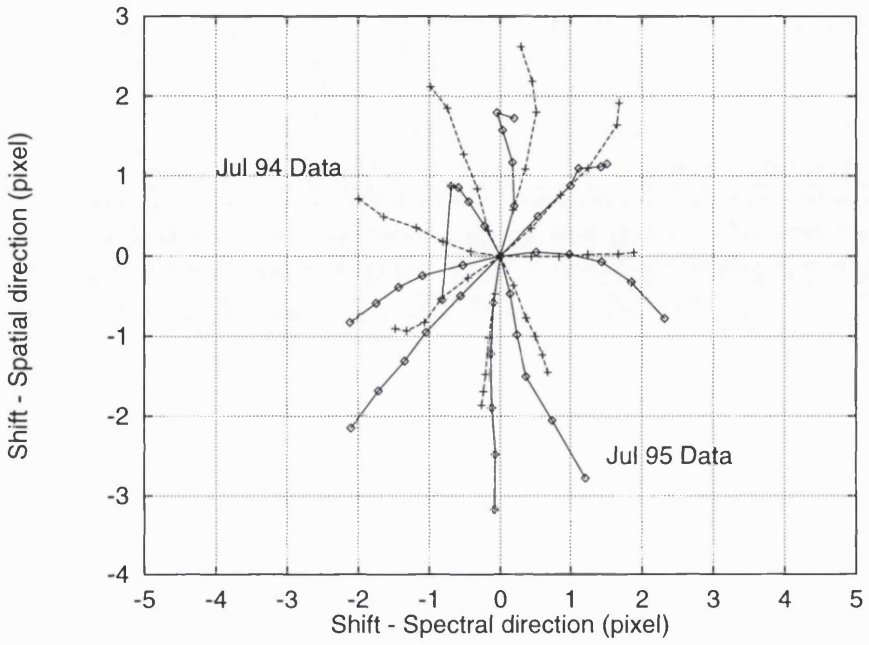


Figure 7.12: Modified flexure map taken on 11/07/95 compared with the one taken one year earlier. The two models are visually still quite similar, but a flexure simulation is needed to quantify the similarities.

experiencing when the flexure map was recorded. The equivalent telescope tracking time that would produce such flexure shifts is $\sim 90min$. The type of distribution of figure 7.11 is typical of ISIS in normal operation and it is useful to estimate the long-term stability of the flexure model.

Since the simulation gave an encouraging result, I decided to use this system to estimate the level of flexure shifts ISIS was experiencing in the past and to evaluate the effect of using an old flexure model for active compensation corrections (instead of using a model recorded only a few days earlier). As an example, I compared the flexure map taken on 11/07/95 with the measurement of the previous year taken on 31/07/94 by R. Rutten, who was managing the ISIS spectrograph at the time. Figure 7.12 compares the two modified look-up tables. The data from one year earlier seem to show slightly less flexure, and indeed this is confirmed if we apply the same procedure described above to the older data. A similar distribution of shifts as in fig. 7.11 gives a mean flexure $\bar{u}' = 0.65$ pixels ($\bar{u}' = 15.6\mu m$).

To estimate the level of flexure compensation we have to consider the effect of using

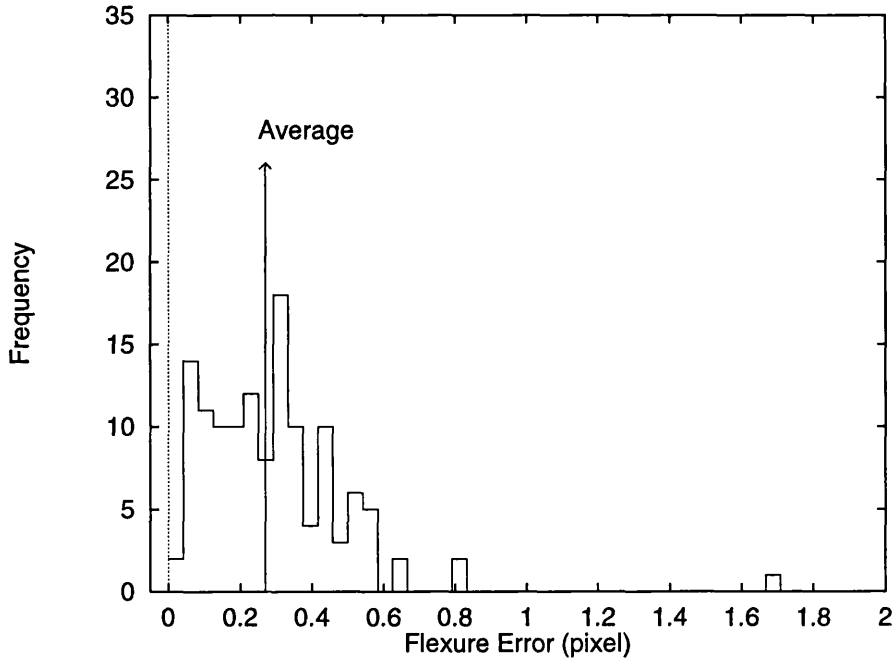


Figure 7.13: Distribution of flexure residual errors after applying active compensation, but using a look-up table one year old. The mean residual is $\bar{\delta u} = 0.27$ pixels. The flexure maps from 11/07/95 and 31/07/94 were used. In the distribution 49.3% of the data is below mean.

a look-up table which does not represent the flexure in the spectrograph at the time. We can, for example, use the 11/07/95 map to correct flexure represented by the 31/07/94 map (see fig 7.10). In our mathematical analysis of the flexure model, this corresponds to evaluating the differences between the vector shifts u_k in the first map (i.e. the shifts applied) to the shifts u'_k in the second (i.e. the real spectrum shifts). The residual errors δu after flexure compensation will be:

$$\delta \mathbf{u_k} = \mathbf{u_k} - \mathbf{u'_k} \tag{7.12}$$

A distribution of flexure shifts can be produced in this case, too, to obtain the histogram of fig. 7.13 The graph shows that flexure has been significantly reduced: the mean flexure shift is now $\bar{\delta u} = 0.27$ pixels ($\bar{\delta u} = 6.5\mu m$). Compared to the original $\bar{u}' = 0.65$, flexure has been reduced by a factor 2.5. If we take two flexure maps at a short time interval, the improvement is even more marked. We can repeat the above procedure using two maps only one week apart, that is the 11/07/95 data and the 18/07/95 data. The distribution of residual errors in this case is shown in

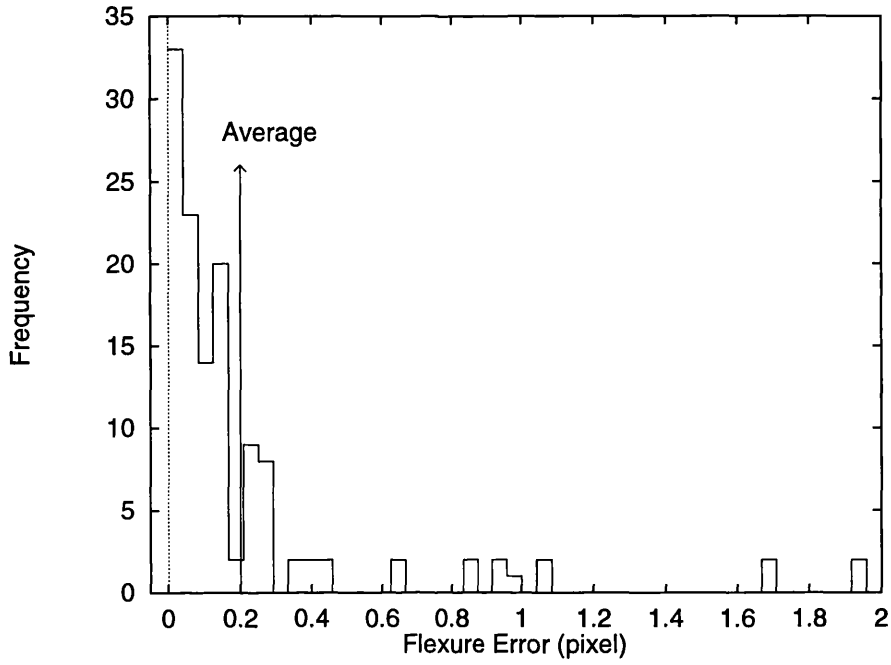


Figure 7.14: Distribution of flexure residual errors after applying active compensation, but using a look-up table one week old. The mean residual is $\bar{\delta u} = 0.20$ pixels. The flexure maps from 11/07/95 and 18/07/95 were used. In this distribution 71.7% of the data is below mean, indicating that apart from a few exceptional cases, the expected level of flexure is significantly lower than 0.2 pixels.

fig 7.14. The mean flexure shift is now $\bar{\delta u} = 0.20$ pixels ($\bar{\delta u} = 4.8\mu m$), but even more importantly, the distribution is highly asymmetric, with 71.7% of the data below mean. This indicate that open loop flexure compensation can improve flexure in ISIS by a factor three or four. Indeed, it is interesting to observe how close the above estimates are to real residual errors measured during our experiments described later in the chapter. It is also clear that provided nothing goes wrong in the spectrograph, it is possible to achieve significant improvement in image stability even with a look-up table several months old.

7.4.3 Hysteresis in Flexure

In order to better evaluate the amount of hysteresis present in the spectrograph, I decided to produce a hysteresis curve by moving the telescope from zenith to horizon and back while keeping the turntable position fixed. By taking a few flexure

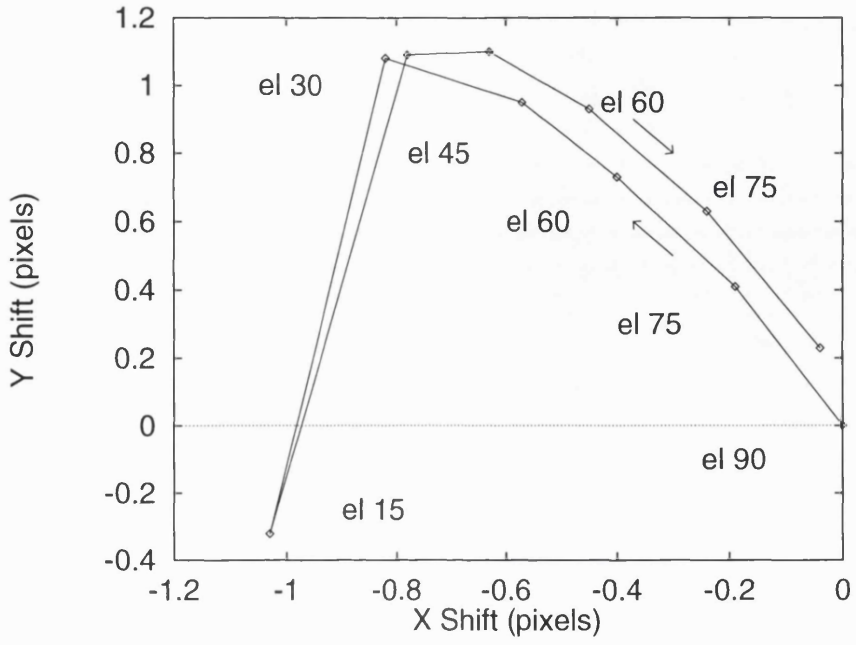


Figure 7.15: Hysteresis in flexure shifts on the detector as a function of telescope elevation for turntable position angle -90^o. The telescope was moved starting from zenith, down to elevation 15^o and back to zenith in angular steps of 15^o. The hysteresis shown amounts to around 15% of the total motion.

measurements along the path followed by the telescope and comparing the positions on the detector, the amount of hysteresis can be estimated. Figure 7.15 shows the flexure shifts on the CCD detector for the rotator position angle of -90^o. The graph definitively demonstrates the presence of a significant hysteresis in the deformations of the spectrograph structure under gravity. The hysteresis in the figure is approximately 15% of the maximum flexure shift. The point corresponding to elevation 15^o suggests the presence of a loose component in the spectrograph that apparently flips suddenly between two distinct positions. When the telescope is lowered below a certain elevation (between 15^o and 30^o), the spectrum suddenly jumps by more than a pixel. However, if the telescope is again lifted above the critical point, the spectrum reverts back to almost the same position.

During our testing at the telescope, we had time to produce only another hysteresis curve, for the rotator position angle -45^o. The result is shown in fig. 7.16. This graph is more difficult to interpret than the previous one, because the curve is

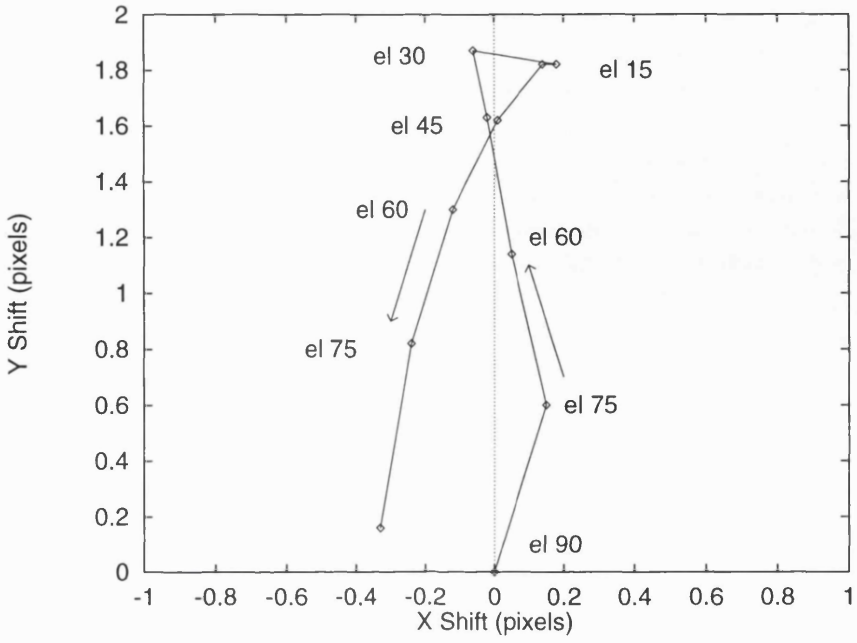


Figure 7.16: Hysteresis in flexure shifts on the detector as a function of telescope elevation for turntable position angle -45^{o}. The telescope was moved starting from zenith, down to elevation 15^{o} and back to zenith in angular steps of 15^{o}.

twisted. However, hysteresis is still clearly present, probably in a similar amount (15% or more). Hysteresis in flexure shifts can result from a variety of causes, including stresses and non-uniformity in the spectrograph structure, joints and loose bearings. Hysteresis has been measured in other Cassegrain instruments, like the Low Dispersion Survey Spectrograph (LDSS-2) for the William Herschel Telescope [102], even if the total flexure in that case was much less. Hysteresis is the main limitation for an open loop flexure compensation system.

7.5 Open Loop Flexure Compensation

In order to test the performance of ISAAC in open loop, it was necessary to simulate the tracking of a star with the telescope (I only had daytime access to the telescope). To avoid spending a long time tracking, I chose to calculate the telescope position at 20 minutes intervals when following the star motion in the sky. The telescope could then be stepped from one point to the next and at each point active compensation

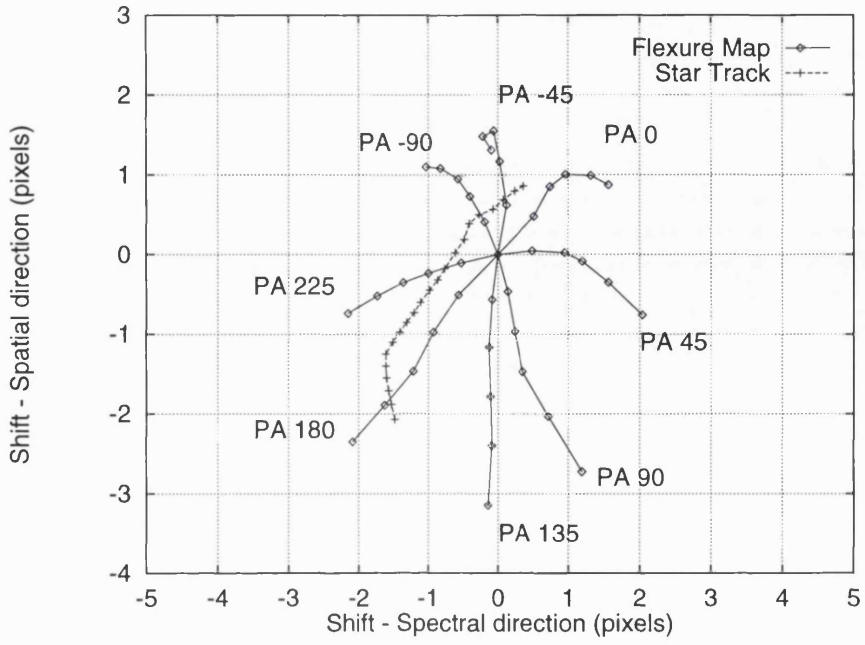


Figure 7.17: Image motion of an object with declination $\delta = 45^o$ on the CCD detector. The points in the track are 20 minutes apart and the whole track covers 7 hours of telescope motion. The corresponding flexure map used for the simulating is shown as well. The object describes a track on the CCD detector because of flexure in the spectrograph. The track in the graph has been simulated from the flexure model shown, knowing the telescope position when tracking such a star.

applied. As a target for the telescope, I chose an object with declination $\delta = 45^o$, for no particular reason apart from the fact that such an object produces a long track on the sky with large movements in telescope elevation and turntable rotation. Figure 7.17 shows the simulated image motion on the detector when following such an object, superimposed to the flexure map used to correct for it. The track was generated by calculating the telescope position when tracking the star and converting this into spectrum shifts using the shown flexure map.

7.5.1 All-Sky Flexure Model

The experiment consisted in stepping the telescope along a star track with and without active compensation of flexure applied, and to compare the results. It was decided that it was better to take the two measurements in successive runs, to avoid

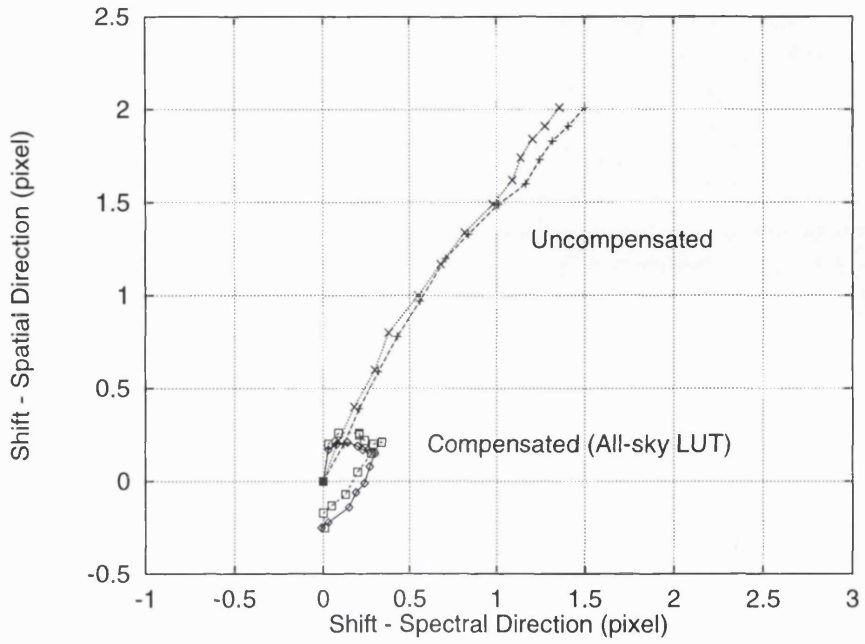


Figure 7.18: Reduction of image motion on the detector by applying active compensation with ISAAC. The flexure model used a look-up table covering the whole sky. Clearly, when active compensation is used, the motion of the image is limited to a much smaller area of the detector.

having to move the active collimator back and forwards at each telescope step. This was done to emulate the behaviour of the active collimator in real astronomical observation and to reduce the hysteresis in the mirror movements. Therefore, I first took a series of measurements along the track with active compensation of flexure operational, then the collimator was reset to the rest position (with all the actuators at zero expansion) and the same set of measurement was repeated. The active compensation used an all-sky look-up table, deriving the corrections from the flexure map covering the whole sky accessible with the telescope. The result of applying active compensation on the image motion on the detector and for spectral and spatial directions separately, are shown in figs. 7.18, 7.23 and 7.24.

The graphs clearly show a significant improvement in the spectrum stability, particularly for long tracking times. Over more than 4 hours of tracking time, flexure was reduced from around 1.5 pixels ($36\mu m$) to less than 0.3 pixels ($7.2\mu m$) in spectral direction. In spatial direction the improvement was even greater, reducing flexure from 2.0 pixels to 0.25 pixels. This residual error of 20% of the total flexure is

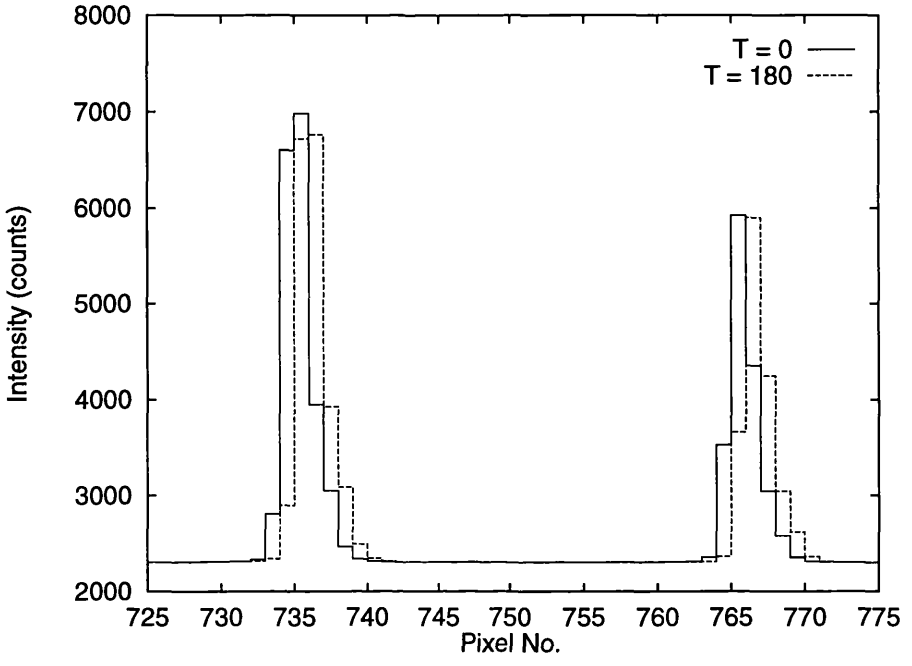


Figure 7.19: Displacement of two spectral lines of a CuAr + CuNe calibration lamp due to flexure. The two spectra are relative to two telescope positions which are separated by $180min$ of tracking. The shift of the lines is more than one pixel.

consistent with the 15% hysteresis measured earlier and with the expected level of inaccuracies in the flexure model. Indeed, the analysis of the behaviour of flexure in ISIS I performed before testing at the telescope (chapter 4), and the further analysis described in this chapter, indicated maximum compensation errors to be around 0.25 pixels. The error on each measurement is given by the centroiding technique used to calculate the flexure shifts. This varies slightly with the intensity of the spectrum, but I estimated the algorithm produces the correct value within at least ± 0.025 pixels ($0.6\mu m$).

The direct effect of active compensation on the stability of the image on the detector can be better appreciated by showing the displacement of a spectral line on the CCD due to flexure, as in fig. 7.19. The graph shows the displacement on the CCD of two spectral lines of a combination of CuAr and CuNe calibration lamps. The equivalent time of tracking for the telescope motion involved is three hours, which produces a shift of more than one pixel. This tracking time is unrealistic for real astronomical observation, but I chose it to magnify the effect of flexure on the line position. On

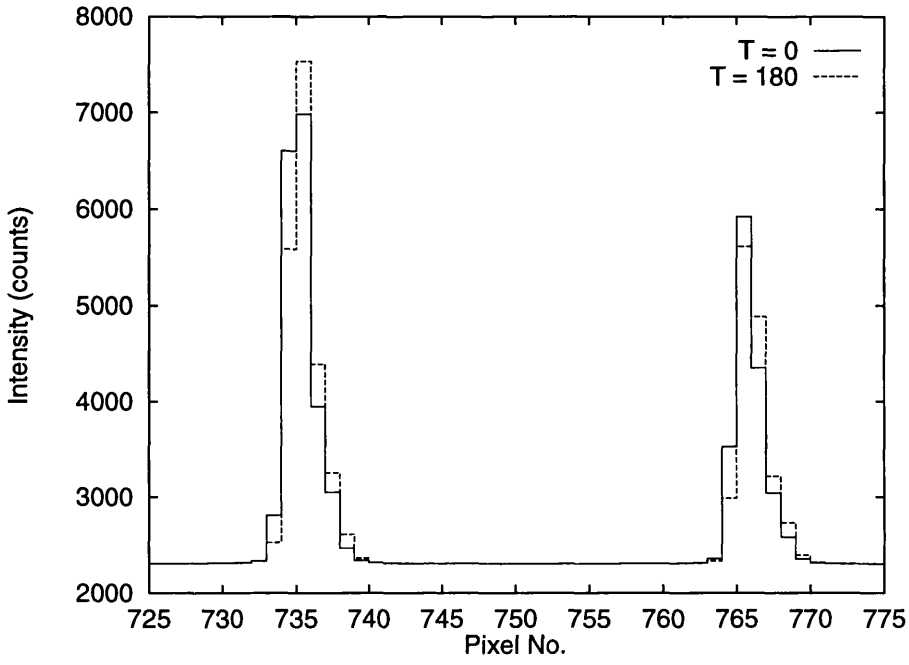


Figure 7.20: Displacement of two spectral lines of a CuAr + CuNe calibration lamp after active compensation of flexure is applied. The two spectra are relative to two telescope positions which are separated by $180min$ of tracking. The shift of the lines is around 0.2 pixels.

a shorter tracking time, the flexure shift is still astronomically very relevant, but it is more difficult to detect visually. When active compensation is applied, the two lines hardly move, as figure 7.20 shows. The shift of the lines can be measured to be 0.2 pixel with centroiding techniques.

This first experiment proved that spectrum stability can be improved with open loop active compensation by almost a factor 5 over several hours of telescope tracking. Over shorter tracking times, however, the improvement is less dramatic. During, say, one hour exposure, the total amount of instrument flexure to correct is much smaller, of the order of 0.5 pixels ($12\mu m$), and the inaccuracies in the flexure model become more significant. This is already evident from fig. 7.18, but to have a better sampling, I repeated the first hour of the track stepping the telescope every 5 minutes of tracking time. The comparison between non-compensated and compensated tracking is shown in figs. 7.21 and 7.22 with the image drifts in spectral and spatial directions respectively. The graphs clearly show that in this case the improvement

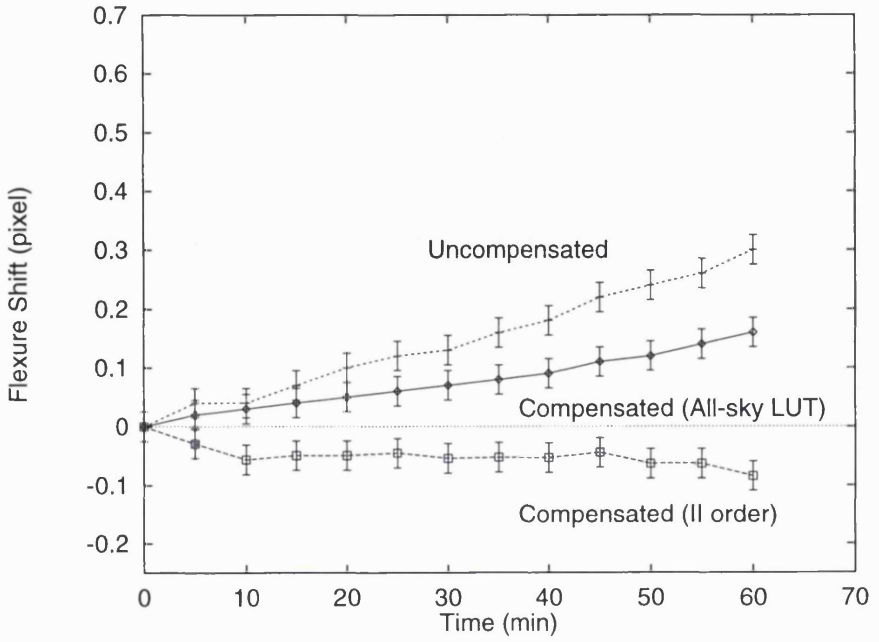


Figure 7.21: Flexure compensation in spectral direction with an all-sky flexure map over one hour of telescope tracking. The second order correction is shown as well.

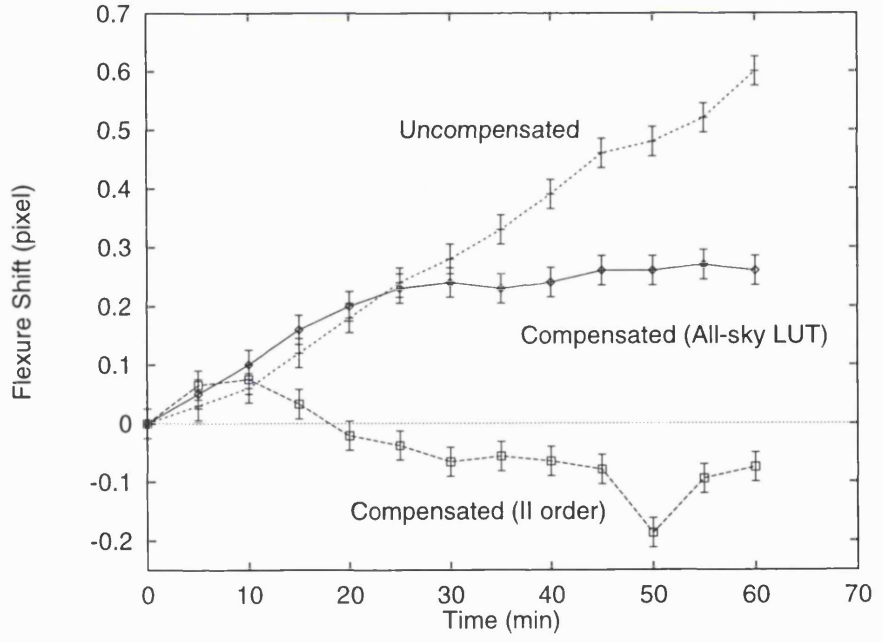


Figure 7.22: Flexure compensation in spatial direction with an all-sky flexure map over one hour of telescope tracking. The second order correction is shown as well.

in stability with respect to the total shift is much less, the reduction being a factor 2 instead of 5. Nonetheless, the maximum spectrum drift with active compensation is around 0.15 pixels in spectral direction and 0.2 pixels in spatial. This second experiment indicated that the all-sky flexure model contains some intrinsic errors. Some of these errors are probably a consequence of the different type of telescope motion between tracking and taking a flexure map. The flexure model may then be improved if the elevation/rotator position grid is replaced by a declination/right ascension one, where the flexure map is constructed by moving the telescope along star tracks of different declinations. The following tests were designed to estimate the modelling errors and to explore the possibility of improving the pointing model.

7.5.2 Modelling Errors

To test the significance of the residual errors in the above experiments, I decided to repeat the same telescope track the next day, with exactly the same spectrograph setup. This is a direct way of separating the contributions from the errors in the model and those from hysteresis in the spectrograph. Figures 7.23 and 7.24 show the differences between the two runs, in spectral and spatial directions, for both the non-compensated flexure and the residual errors. Several conclusions can be drawn from these graphs: firstly, on a timescale of one or two days, the flexure in the spectrograph is highly reproducible, with maximum errors of less than 0.15 pixels (10% of the total). This is in good agreement with the 15% hysteresis measured, if we consider that in this case the telescope was moving in the same way. The high reproducibility of the instrument flexure implies that the greater part of the residual errors with active compensation is not due to the intrinsic behaviour of the structure. Even more interesting is the fact that also the residual errors after active compensation are highly reproducible. Since the error appears to be systematic, it indicates that the flexure model is the most likely cause and *not* the motion of the collimator within the cell. Indeed, if significant hysteresis was present in the collimator motions, the residuals would not be so accurately reproducible. Therefore, the clear result coming from this experiment was that the limiting factor in image stabilization with open loop active compensation was the accuracy of the flexure model.

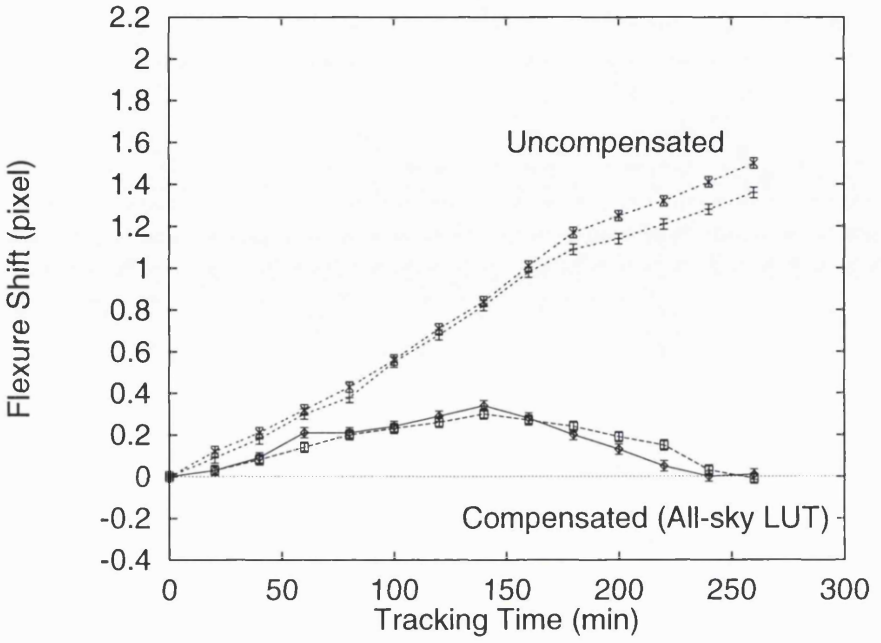


Figure 7.23: Reproducibility of flexure shifts and of the residual errors after active compensation in spectral direction. The data were recorded on two consecutive days with telescope and spectrograph in the same conditions.

The inaccuracies of the flexure model are easy to understand: the data in the all-sky flexure map were recorded by moving the telescope in a completely different way compared to tracking. The raw data was modified to reduce hysteresis, but no attempt was made to account for the different telescope motion. I suggest that a possible improvement on the model would be to produce a flexure map by scanning the sky along star tracks of different declinations. The number of points in the model is also a limitation, although not a serious one, considering that the linear interpolation provides a satisfying approximation of flexure between points in most cases.

7.5.3 Star-Track Flexure Model

To improve the flexure model, I decided to use the measured values of flexure shifts along a star track to build a look-up table for that particular track. This experiment was designed to overcome the inaccuracies of the model originating from the fact that when recording the data for the look-up table the telescope is moving in an

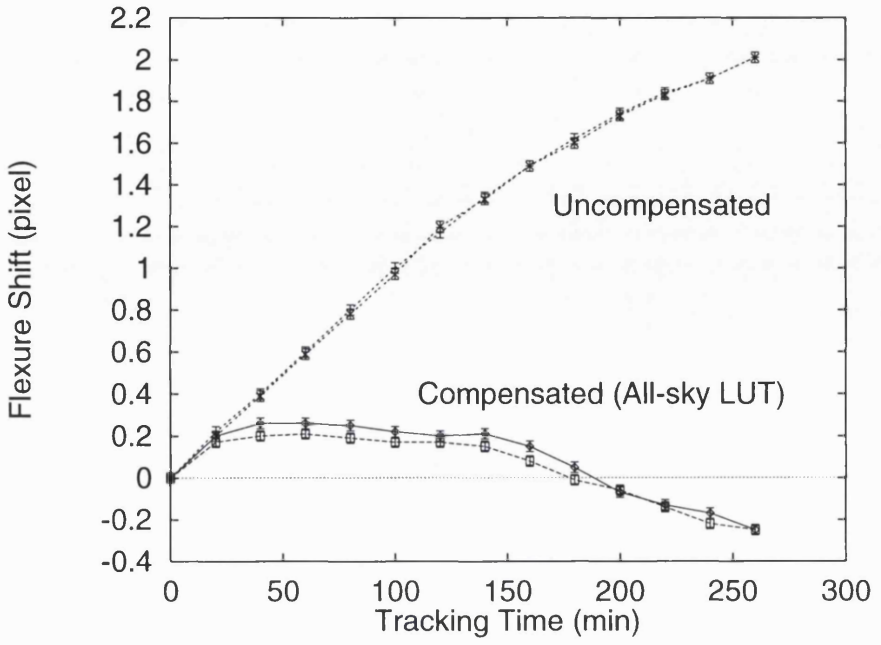


Figure 7.24: Reproducibility of flexure shifts and of the residual errors after active compensation in spatial direction. The data were recorded on two consecutive days with telescope and spectrograph in the same conditions.

"unnatural" way, keeping the Cassegrain rotator fixed and moving up and down in elevation. Using the flexure along the star track for the model, implies the look-up table can be used only for that particular object. However, the model could be extended to cover the whole sky by generating several tracks, stepping in the object declination and time, instead of telescope elevation and turntable angle. There was obviously no time to create this all-sky model and I thought following a single track would give me enough information. Figures 7.25 and 7.26 compare the result with the star track flexure model, with the all-sky model and without compensation, both in spectral and spatial direction. As expected, the star track flexure model improves the stability of the spectrum significantly when compared with the all-sky model. The residual image drift was 0.15 pixels ($3.6\mu m$) over 1.5 pixels of flexure in spectral direction and 0.3 pixels over 2.0 pixels in spatial direction. The reduction of spectrum motion was down to 10-15% of the non-compensated value. These errors are closer to the intrinsic limit to open loop flexure compensation, as calculated from the levels of hysteresis in the ISIS structure. More importantly, it seemed that the accuracy of the collimator itself was far higher than the one of the model,

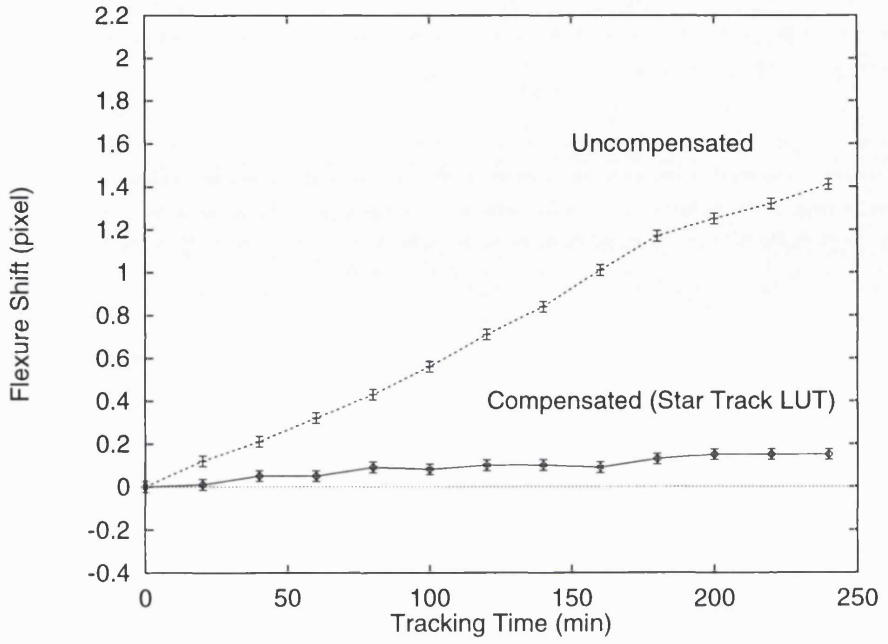


Figure 7.25: Active compensation using a look-up table modelled for a specific track compared with the standard all-sky flexure model (spectral direction).

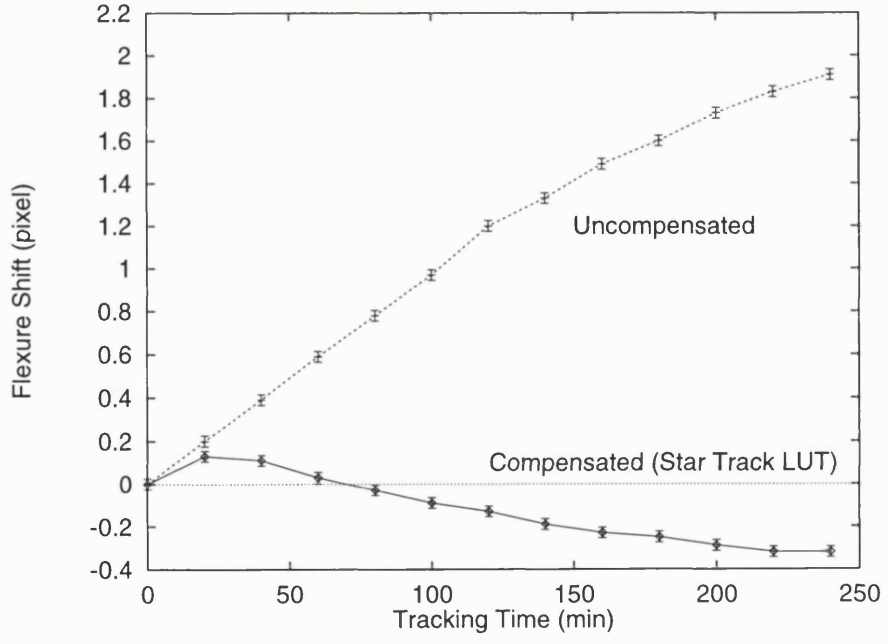


Figure 7.26: Active compensation using a look-up table modelled for a specific track compared with the standard all-sky flexure model (spatial direction).

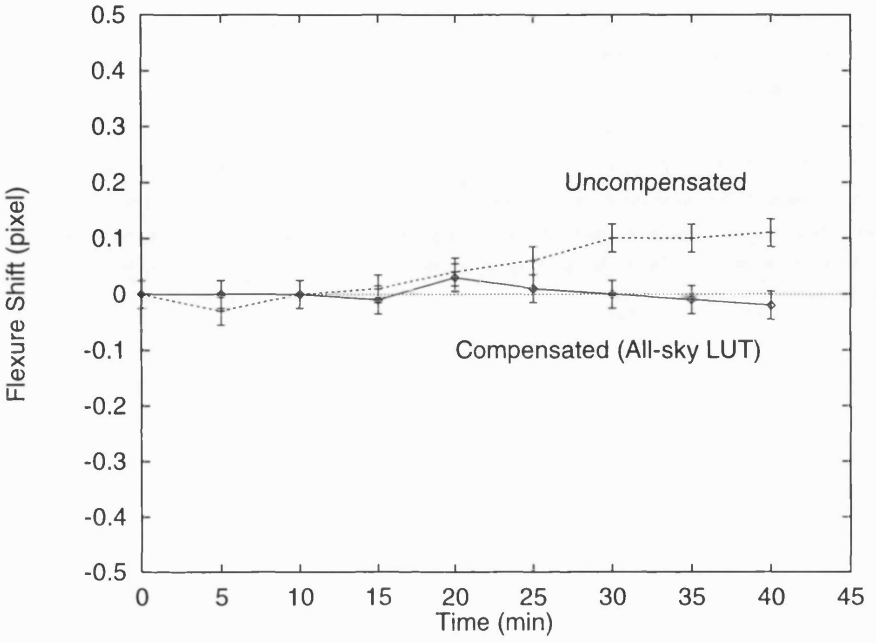


Figure 7.27: Active compensation during normal telescope tracking (spectral direction. The spectrum shift was checked every 5 minutes by tacking a 20 second exposure of the calibration lamps.

consistently with the laboratory tests.

7.5.4 Compensation During Tracking

As a separate test, I also tried the effect of compensation during normal telescope tracking (instead of stepping the position and then applying compensation). In this case the telescope was moving continuously tracking a star and the active collimator was updating the compensation every minute. This is closer to the normal mode of operation of ISAAC during real astronomical observations, but for reasons of time the total tracking was limited to only 40 minutes. Again I found some improvement in the stability of the spectrum, even if, due to the short exposure time, it was less significant when compared to the non-compensated value. The result, in spectral and spatial directions is shown in figs. 7.27 and 7.28. During tracking, the shift of the spectrum was checked by taking 20 second exposures of the arc lamps every 5 minutes. Both graphs show an improvement on the non-compensated flexure, particularly towards the end of the exposure. Note that this time the telescope was

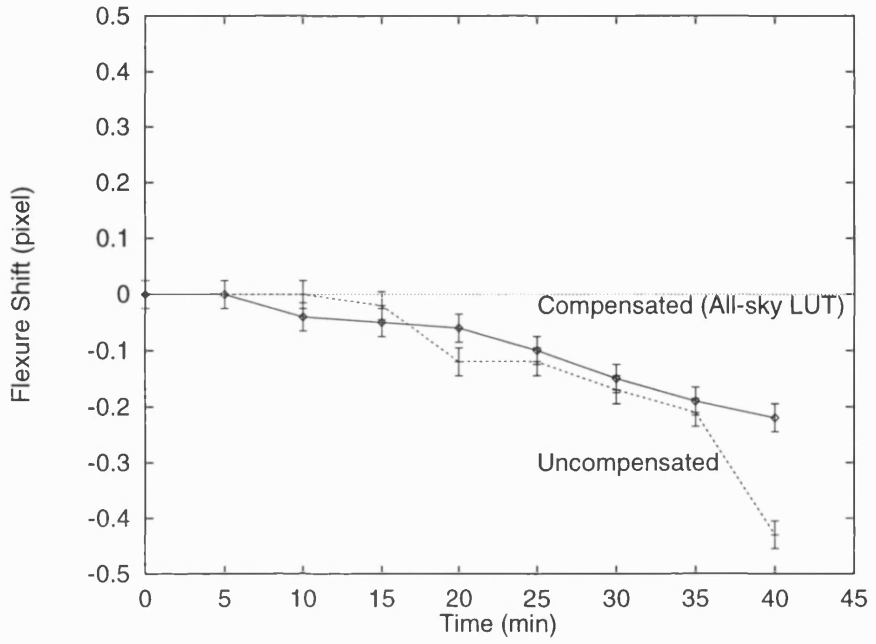


Figure 7.28: Active compensation during normal telescope tracking (spatial direction. The spectrum shift was checked every 5 minutes by tacking a 20 second exposure of the calibration lamps.

tracking a different object, with declination $\delta = 20^o$, so a direct comparison with the previous measurements is not possible.

7.6 Improving on the Open Loop Flexure Model

The experiments in the previous section were concerned with exploring the possibilities and the accuracy of an open loop flexure compensation system. Since all the tests showed that the model was always the limiting element in the level of stability achievable, we conceived a way of testing the capabilities even further. The idea was to take the residual errors in flexure compensation and feed them back in the collimator control system. With this simple procedure, it was possible to reduce the modelling errors to a minimum. Moreover this would eliminate any systematic error coming from the collimator or the spectrograph. In practice, this meant having a double look-up table: the total correction is the sum of the the motion extrapolated from the standard flexure model and the position error derived from a previous

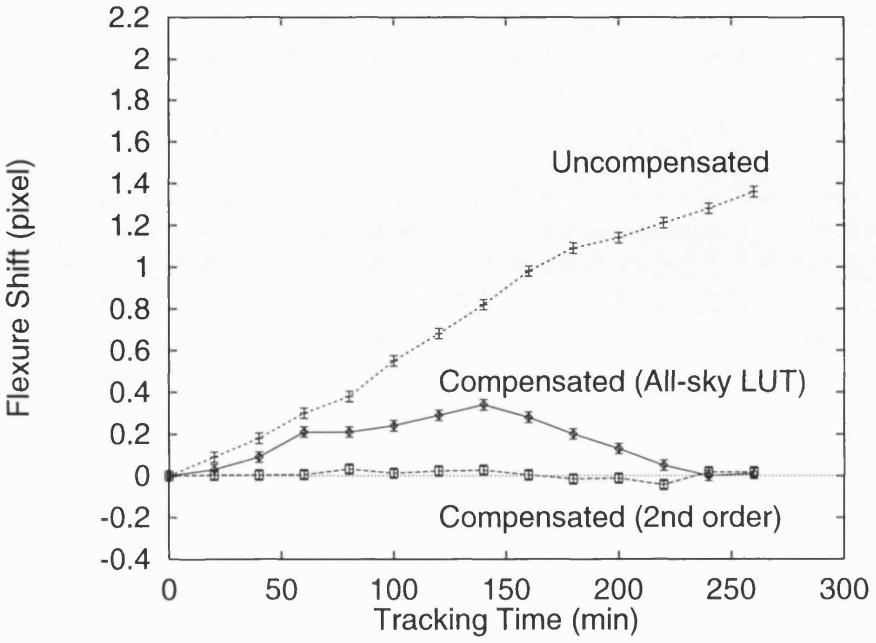


Figure 7.29: Flexure compensation using an improved model which accounts for position errors of the standard all-sky flexure map by introducing second order corrections (spectral direction).

compensation experiment.

This type of second-order correction is obviously not practical for an open loop system, due to the extreme complexity in deriving the corrections for any possible target. It is however relevant, because it simulates more closely the performance of a closed loop system, where the compensation errors are sensed directly, instead of being recorded in a second look-up table. Figures 7.29 and 7.30 compare the second order flexure compensation technique with the standard all-sky model for both spectral and spatial direction.

As the graphs show, in this case the improvement is truly dramatic. The maximum errors are reduced to 0.04 pixels in spectral direction ($1.0\mu m$) and 0.15 pixels in spatial direction, over 4 hours of telescope tracking. Also, the slightly larger errors in spatial direction are produced in the last hour of tracking, so that the image motion remains well within 0.1 pixels in both directions for over three hours. The improvement in image stability with second order flexure corrections is even more apparent when spectral and spatial shifts are plotted together, as in fig. 7.31. This

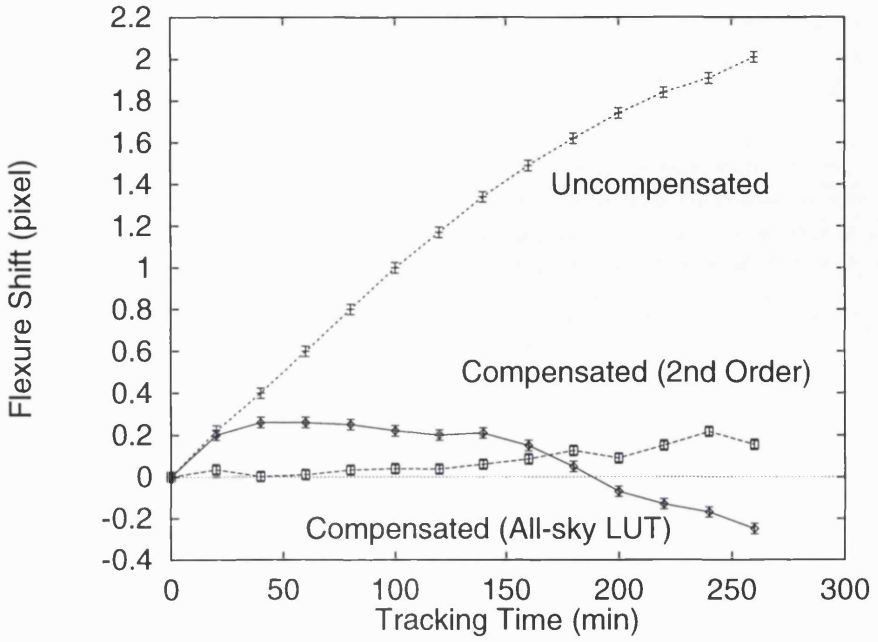


Figure 7.30: Flexure compensation using an improved model which accounts for position errors of the standard all-sky flexure map by introducing second order corrections (spectral direction).

test demonstrates the obvious advantage of a closed loop flexure compensation system. More importantly, however, it clearly shows that the active collimator in itself performed exceptionally well. In fact, the laboratory tests showed it was possible to control the collimator position so as to correct detector shifts as small as $0.4\mu m$ or 0.02 pixels. It seems the achieved stability of 0.1 pixels over 3-4 hours could be improved even further. Second order flexure compensation was tested also on a shorter tracking time of one hour. The result is shown in figs. 7.21 and 7.22. Again it was possible to limit the total image motion within 0.1 pixel. The graphs seem to show some residual systematic errors which a real closed loop feedback would eliminate. The conclusion from these experiments is that the limit performance of the collimator has not yet been reached. Only a real closed loop experiment on ISIS would allow to measure this limit. The open loop system however, has shown the potential for great improvement in the spectrum stability in ISIS. Unfortunately the accuracy of the flexure model degrades slowly with time, and it is sensitive to changes in the spectrograph setup.

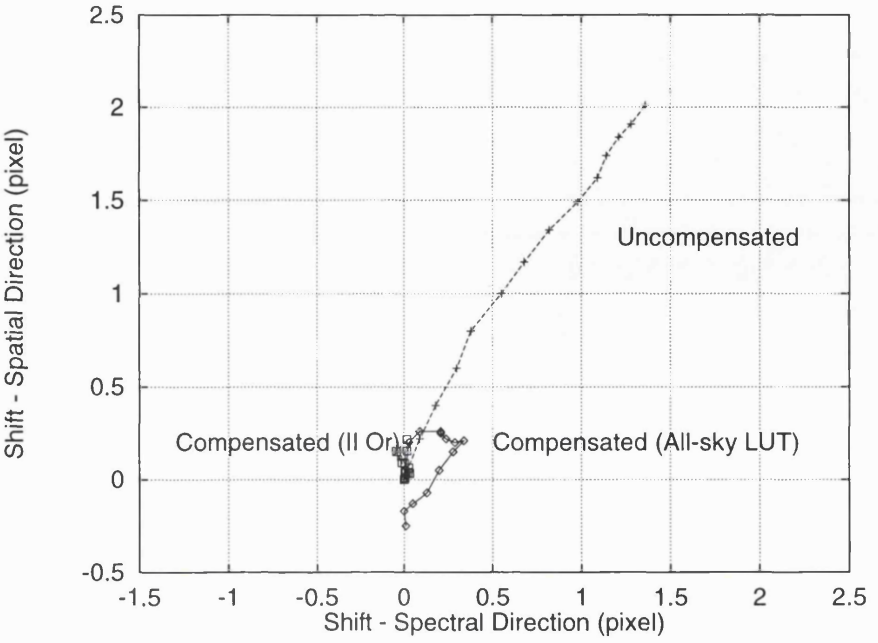


Figure 7.31: Displacement of the spectrum on the CCD detector with flexure compensation using a model with second order corrections. The motion of the image is almost entirely eliminated when compared with the non-compensated flexure shift.

Towards the end of the testing of ISAAC at the telescope, I took a set of measurements with both simple open loop active compensation and second-order corrections after a significant change in the spectrograph setup. I wanted to test how the flexure look-up table relative to a different spectrograph setup would cope with these changes. The changes involved changing the grating to the $600 lines/mm$ ruling and the central wavelength from $7000\mathring{A}$ to $4500\mathring{A}$. The observed object was also different, having a declination $\delta = -16^{o}$. The experiment showed that the flexure model has some problems in coping with this new spectrograph setup, particularly in spectral direction. This is to be expected because a grating change is a substantial modification of the setup. Nonetheless, active compensation was still effective in reducing flexure and the spectrum stability was still significantly better when compensation was applied. As expected, the second order compensation coped very well, because the position errors are fed back to the collimator control system. Again, in theory, a closed loop compensation should at least match the performance of the second order correction.

7.7 Conclusions

This experiment has proved that active compensation is a viable and effective way of reducing flexure in Cassegrain spectrographs. It has also shown that the collimator can be controlled with very high accuracy and it is a suitable element for applying active corrections. Another result is that ISAAC is capable of reducing flexure in ISIS significantly. During the time we spent at the telescope ISIS was showing an average spectrum drift of over $40\mu m$ in 4 hours tracking time. By applying active compensation it was possible to reduce it to an average $4.8\mu m$ over 4 hours with an all-sky flexure map and to less than $2.4\mu m$ by applying second-order corrections. These results have several implications on the astronomical data the spectrograph can provide, not just in terms of better line profile measurements and reduced radial velocity errors. The reduction in flexure achieved is such that the time wasted in taking regular calibration arcs can be cut drastically. Moreover, this low flexure can allow longer exposure times and improved S/N ratio. The limiting factor in the length of exposures would be effectively shifted from flexure drifts to cosmic ray events rate.

As I already pointed out earlier, the experiment has also shown that the reduction in flexure we achieved was limited by the accuracy of the flexure model and not by the accuracy in the movement of the collimator. The laboratory tests had demonstrated that the collimator could handle movements as small as $0.4\mu m$ on the ISIS CCD, so it is conceivable that that flexure compensation could be improved even further with a more accurate model. Indeed, some other ways of modelling spectrograph flexure have been suggested [56], but these rely strongly on the absence of hysteresis in the spectrograph. In fact, another result was to prove the existence of a relatively large hysteresis in the spectrograph distortions. This is likely to put some severe constraints on the accuracy a flexure model can ultimately achieve. Moreover, we have to take into account the fact that to collect the data necessary for an accurate flexure map takes considerable telescope (day) time, which is often needed for engineering and maintenance. ISAAC using an all-sky flexure map is therefore a compromise solution, balancing performance in image stability versus time required to achieve that.

It is clear however, that to exploit fully the potential of the active collimator itself,

a different approach is needed. By closing the loop with an optical reference beam propagated through the spectrograph it is possible to avoid the problem of hysteresis altogether. The flexure shifts would be sensed in real time while the spectrograph is tracking and directly fed back into the collimator control. It also would avoid the need of producing several different flexure models for corresponding different spectrograph setups. How this may be achieved in practice on a new or an existing spectrograph, will be the subject of the next chapter.

Chapter 8

Future Developments

8.1 ISAAC Open Loop

The ISAAC experiment has been successful in reducing flexure in ISIS and the La Palma Observatory, in cooperation with the Royal Greenwich Observatory, is now examining the possibility of having ISAAC permanently installed on the spectrograph and commissioning to OSL the construction of a second active collimator for the red arm. The prototype active collimator has already proven its effectiveness, but a permanent flexure compensation system can have an even better performance. In a recent meeting (July 1996) with Mick Johnson, the instrumentation manager at the Royal Greenwich Observatory, the first stage of this project was finally agreed. This stage will mainly involve integrating the active collimator control software with the telescope and spectrograph control system and improving the open loop modelling of flexure in ISIS.

The first advantage of a common user active collimator is its being completely integrated with the telescope control system. Then, the collimator control system can access the telescope position directly, instead of relying on input from the user when changing the target. This saves time and may allow, for example, to continue compensating flexure between exposures, when observing nearby objects in the sky (for objects further apart, flexure may be too large and hysteresis becomes significant). This would reduce the need of calibration spectra exposures even further. The recorded flexure data can be processed into a format accessible by the

collimator control software, limiting the effort required from the telescope servicing staff. The procedure for producing a flexure map is already highly automated, so updating the flexure look-up table could be reduced to simply starting a computer routine for the telescope control system. This would allow much more frequent flexure measurements and consequently better accuracy in the flexure corrections.

The way flexure data is recorded can be modified as well. I explained earlier that stepping the telescope along ideal star tracks should reduce hysteresis in the flexure measurements themselves, producing a more accurate model. The whole sky could be covered by following tracks in, say, 15° steps in declination and 1 hour steps in hour angle. Such a routine would be slightly more complex than the existing one, but it would take a similar time to complete the measurement.

The operational scenario for an integrated flexure compensation system for ISIS is quite simple. One of the attractive features of such system is that the user does not need to know the details about its operation. A simple power on/off option in the ISIS control display could allow switching between active compensation to normal operation. The collimator control software would take care of everything else: synchronization with the start of the exposure, active corrections, resetting the system when changing target. The observing astronomer need only to be aware of the need to take new calibration spectra when changing target (to an object far away from the previous one). Since the collimator control would be integrated with ISIS, it would know the spectrograph setup, and choose the appropriate look-up table accordingly.

The servicing required from the telescope staff is also quite straightforward, even if somewhat time-consuming. Regular flexure measurements need to be taken, to keep the open loop performance to its best. When image stability is of particular importance for an astronomical observation, a new flexure map should be taken the same day or a few days earlier. This is not unreasonable, because an automated flexure measurement would take only 2 or 3 hours of telescope day time. In other cases, a look-up table a few weeks old would still achieve significant improvement. It would be advisable, however, to take a new measurement when the spectrograph setup is changed considerably, or after ISIS had been off the telescope for a while.

The prototype I tested on ISIS was able to reduce flexure by a factor 2 or more

over one hour, and the average residual flexure was around 0.2 pixels or $\sim 5\mu m$. The original stability specification for ISIS was indeed $5\mu m/hr$ and ISAAC was able to meet that target. An integrated system, with a better flexure model and regular servicing, can be expected to improve that performance significantly, possibly reducing errors to 0.1 pixels on average. I believe the spectrograph would have large benefits form an open loop flexure compensation system permanently installed. A first stage in this development has already been approved by the Royal Greenwich Observatory and it is due to start soon. In this stage the control software for the active collimator would be transferred to the telescope control system, while the control electronics would be kept separate. In a second stage the electronics would need to be integrated as well.

8.2 ISAAC Closed Loop

While an open loop active collimator can significantly improve the image stability of the ISIS spectrograph, a few shortcomings have also been identified. The open loop system cannot handle hysteresis or occasional malfunctions, thus limiting the level of stability achievable. The goal of 0.1 pixels ($2.4\mu m$) of average error is probably the maximum that such system can achieve. The mechanical tests of the active collimator showed however that, when installed on ISIS, it could potentially produce residual errors smaller than $0.4\mu m$ on the spectrograph CCDs. Even taking $1\mu m$ as a more reasonable goal, This is more than a factor two improvement on what an open loop system may achieve and a factor 5 better than what ISAAC did. Stability at the level of 0.05 pixels is a reasonable target for a closed loop system.

An important advantage of a closed loop system is that it would drastically reduce the time spent in obtaining adequate flexure data, even if the closed loop was implemented together with a lower accuracy open loop system. Moreover, the closed loop would even handle sudden jumps in the spectrum due to malfunctions, possibly helping to pinpoint the cause of such jumps and providing diagnostic information at a later stage, when the malfunction needs to be identified. Therefore, the closed loop compensation system, besides providing higher image stability, could also be used to measure and monitor flexure in the spectrograph, helping the understanding

and the maintenance of the mechanical structure of ISIS.

After the tests of the active collimator on ISIS, David Walker conceived a simple experiment to proof the feasibility of a closed loop flexure compensation system for ISIS. The idea is to use one or more arc lamps (CuAr, CuNe) and to use these as the optical reference during the exposure. The light from these lamps would be injected in the spectrograph through an optical fibre which would terminate at one end of the slit, thus illuminating the first few (e.g. 10 or so) lines of the CCD close to the output register. The lamps would produce a narrow calibration spectrum at one end of the CCD. During the exposure the following process would be repeated:

- close CCD shutter
- read out $\sim$ 10 lines of data with the reference arc
- reverse the CCD vertical clocks for the same $\sim$ 10 vertical transfers to restore the science data in the original position
- cross-correlate the arc with the one taken at the beginning of the exposure to derive the X-Y image shifts
- calculate the collimator actuator positions and update them
- open the shutter and continue the exposure

Staff at the Royal Greenwich Observatory have estimated the dead time for this procedure to be approximately $0.5sec$. Therefore an update rate of one every 1 or 2 minutes would not significantly affect the exposure time. This simple system would allow closed loop flexure compensation and would require only minimal modifications to the ISIS hardware to inject the optical reference. Moreover the top few lines of standard CCDs used on ISIS are hardly used for recording scientific data, being at the edges of the camera field of view, so the data collection capability would not be compromised. The largest CCDs are actually bigger than the camera field of view, so for a future common user instrument some modifications in the ISIS structure would be necessary (e.g. decentering the CCD cryostat).

This experiment would be a crucial test for a closed loop system, and it would give essential information for the design of a flexure compensation system for new

spectrographs, like HROS. A grant application has been submitted for this experiment and the Royal Greenwich Observatory has already expressed willingness to cooperate with OSL in this project. The proposal has not yet been approved.

8.3 HROS and Gemini

The High Resolution Optical Spectrograph is a cross-dispersed echelle spectrograph for the Cassegrain focus of the 8-metre Gemini telescope in Cerro Pachon, Chile. Once completed, it will probably become the most important instrument at the southern telescope, providing high resolution spectroscopy at optical wavelengths. The instrument performance has been defined by the Gemini Science Committee as follows [104]:

- Detector: 4096x4096 CCD array, $15\mu m$ pixels.
- Resolution: two modes having $\mathcal{R} = 50000$ and $\mathcal{R} = 120000$.
- Sampling: 2.5 pixels per resolution element.
- Stability: $2\mu m/hr$ during tracking at Cassegrain.

Following these guidelines, the Optical Science Laboratory has produced a preliminary design for the instrument [123]. It was decided that the spectrograph would have an echelle immersed in fused silica and that it would use prisms for cross dispersion. A schematic diagram of the spectrograph is shown in fig. 8.1. Here are the most relevant parameters in the design:

- Collimator: off-axis paraboloid, $F_{col} = 2560mm$.
- Cross dispersion: 3 prisms, 60^o apex angle.
- Immersed Echelle: ruled at $100lines/mm$, blaze angle 63.5^o.
- Short camera: $F_{cam} = 470mm$ giving $\mathcal{R} = 50000$.
- Long camera: $F_{cam} = 1135mm$ giving $\mathcal{R} = 120000$.
- Spectral coverage: $300nm$ to $1100nm$.

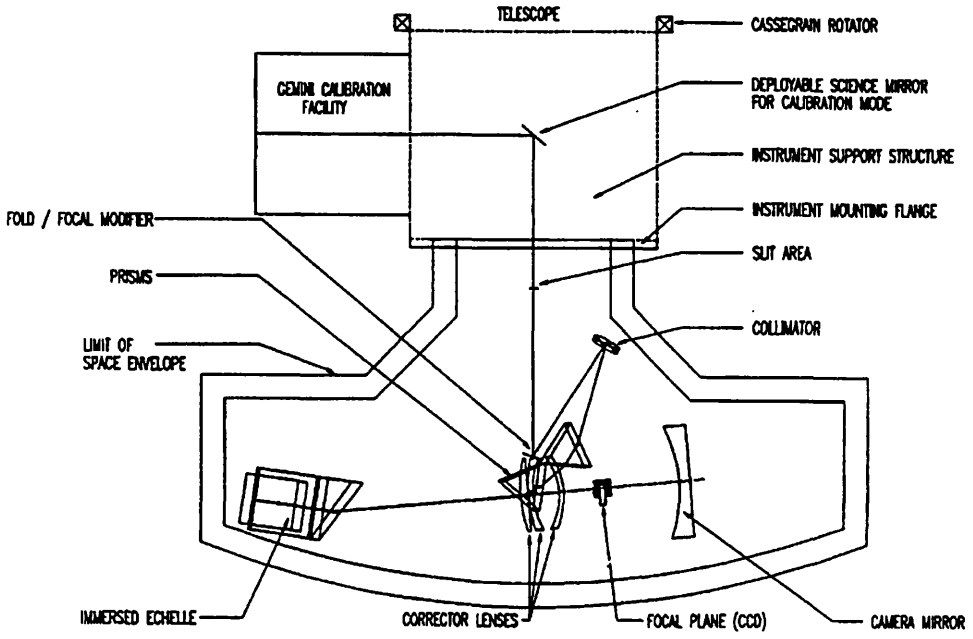


Figure 8.1: Schematic diagram illustrating the main optical components in the optical layout of the High Resolution Optical Spectrograph. Prisms are used for cross dispersion while an echelle immersed in fused silica gives the main spectral dispersion. The diagram does not reproduce the true geometry of the spectrograph or the components within.

Given the stringent requirement on the instrument stability of $2\mu m/hr$, it was also decided that HROS will need a system for active compensation of flexure. The stability requirement is the driving force in this decision, particularly if we compare this target with the stability in other Cassegrain spectrographs of similar size, like the Keck LRIS or ISIS itself. The average flexure in ISIS is currently $12\mu m/hr$, in the LRIS the median motion is quoted to be $13.5\mu m$, with no indication on the timescale [28]. The maximum motion (as specified by Gemini) is much worse in both cases.

It is still not clear whether HROS will have only an open loop flexure compensation system or a closed loop one as well. However, the results from the ISAAC experiment on ISIS suggested that a simple open loop system may not be adequate, both in

terms of achieving the target and of having a reasonable maintenance requirement. It is therefore necessary to investigate the implications of a closed loop system on HROS.

8.4 HROS Closed loop

With a closed-loop system, an optical reference is sensed directly, processed and the information used to control the active element. The analogy with the telescope control is the 'autoguider'. In a spectrograph, a reference light source would propagate through the optical system, be spatially sensed, and the error signal fed back to correct the active element. Again, calling upon telescopes, the most precise correction is likely to be achieved by using the best available flexure model to remove the bulk of flexure, and closed-loop to provide second-order correction of residuals due to non-linear and non-repeatable effects. This is indeed our target for the HROS image stabilization system.

Ideally, for better accuracy, the optical reference should be propagated through the *entire* spectrograph optical system, in order to compensate for motions of all surfaces. The reference can be propagated forwards, with injection at the slit and detection in the camera focal plane, or backwards, with the reverse. In the latter case, the echelle spectrum is magnified by the ratio of collimator and camera focal lengths. For HROS the magnification achieved in backward propagation is actually detrimental, because it makes the echelle spectrum size and the range of flexure shifts very large. Forward propagation is therefore a preferred option if the camera obstructions can be reduced to a minimum. This is the case if the science CCD can be used for detection of the optical reference. In this case the optical reference would be injected at the slit with an optical fibre and imaged on the few lines of the science CCD close to the output register. These CCD rows would be read regularly to provide the flexure measurement, in a similar way to that described for ISAAC in closed loop. There are however reasons why this may not be possible in HROS, in which case backward propagation would be necessary. The first problem to solve in designing a closed loop system for HROS is what to use for the light source to overcome the echelle dispersion.

8.4.1 Scale Factors and the HROS Echelle Spectrum

As explained for ISIS, the scale factor between collimator tilts and image motion in the camera focal plane is essentially determined by the camera focal length. The relation is:

$$\Delta x = 2F_{cam}\Delta\beta\frac{\cos\vartheta_i}{\cos\vartheta_r} \tag{8.1}$$

We need therefore to distinguish between the short and long camera options in HROS. The short camera has a focal length slightly shorter than the ISIS camera ($F_{cam} = 500mm$ in ISIS), so for a given tilt the total motion is also smaller. However, the maximum collimator tilt is not critical, and can easily be extended by modifying the actuator system. The long camera poses more serious problems in terms of actuator absolute accuracy. Since the long camera has $F_{cam} = 1135mm$, more than twice the ISIS camera, to achieve the same position accuracy on the detector, the actuators need to be more than twice as accurate. The requirement for ISIS was $\Delta t = \pm 80nm$, the corresponding value for HROS, keeping the different anamorphism into account, is $\Delta t = \pm 23nm$.

The other major difference between ISIS and HROS, from the point of view of flexure correction, is the spectral format. HROS produces an echelle spectrum in the camera focal plane, which has a characteristic shape and size. The size, in particular, depends on the camera focal length. Figure 8.2 shows the full spectrum between $300nm$ and $1100nm$ and the corresponding size of the CCD detector. The CCD cannot cover the whole echelle spectrum in one exposure, so a tilt of the echelle is required to cover the whole spectral range. As explained earlier, the long camera magnifies the spectrum (and the flexure shifts) by the ratio of the two focal lengths:

$$M = \frac{F_L}{F_S} = 2.4 \tag{8.2}$$

The scale between detector and echelle spectrum in this case is therefore multiplied by a factor 2.4, the spectrum becoming larger. In fig. 8.2 the CCD detector for the long camera is shown smaller to maintain proportions, even if it is the echelle spectrum which varies in size. Again it is the long camera that creates more problems for a flexure compensation system. The spectrum size and the high resolution impose severe constraints on the type of optical reference that can be used. Also, as the figure shows, relatively large tilts of the echelle are required to cover the whole

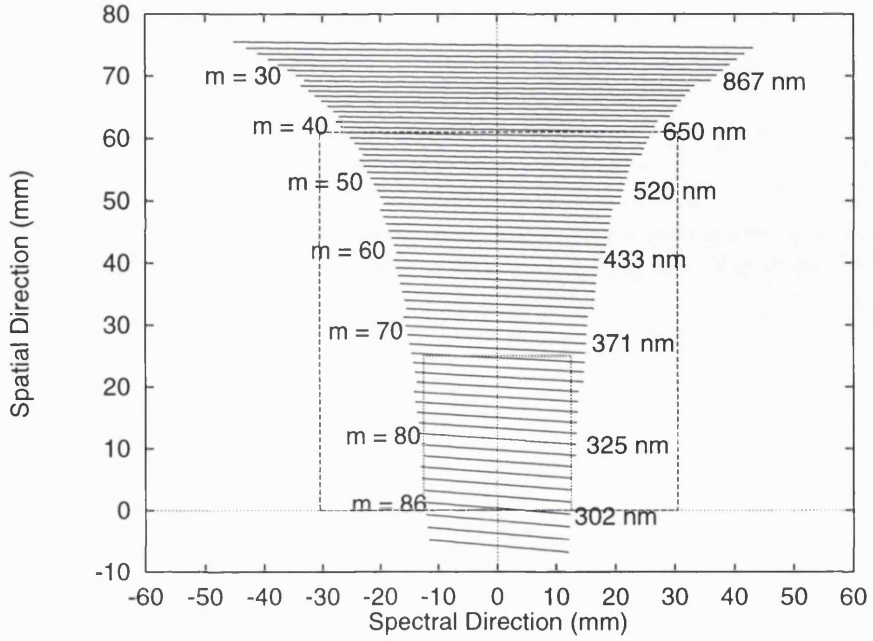


Figure 8.2: HROS echelle spectrum between $300 - 1100nm$ and CCD detector in the focal plane of the short camera ($F_{cam} = 470mm$). The CCD detector (larger square) is 4096x4096 pixel array, its side $61.4mm$ long. Some order numbers and the central wavelength for these orders are also shown. The smaller CCD square shows the scale factor for the long camera ($F_{cam} = 1135mm$). In this case the dimensions on X and Y axes need to be multiplied by a factor 2.4, since it is the spectrum, and not the CCD that changes size.

echelle spectrum in this case. As an example, when changing between observing at the blue end of the spectrum and observing at the red end, the spectrum needs to move by $120mm$ in spatial direction. To cover in full the longest orders (spectral direction), the motion is even larger: $156mm$ from one end to the other of the free spectral range.

8.4.2 A Reference Light Source for HROS

The spectrograph dispersing element(s) will deflect the reference beam depending on wavelength. If the reference were monochromatic, and the spectrograph configured for a different science wavelength, then the reference light may not reach a separate reference detector. In general, the problem of grating or echelle deviation could be

overcome by mounting a separate rotatable mirror on the grating cell, from which the reference would be reflected. However, image-motion due e.g. to grating mounting defects would not be compensated, and the mirror rotation might introduce its own problems. These factors render the simple monochromatic laser reference unattractive in most cases. For a prism cross-dispersed instrument such as Gemini HROS, the configuring for different wavelengths requires scanning the prism angles. The resulting change in prismatic deviation of the reference is even more difficult accommodate.

There are several options for a reference light source for a closed loop flexure compensation system. Many of these ideas were suggested in a study commissioned to OSL by Gemini [52]; the main possibilities can be summarized as follows:

1. Laser operating at a single wavelength (or a few).
2. Continuum lamp + monochromator.
3. Emission line source + monochromator.
4. Tunable laser.
5. Continuum lamp + Fabry-Perot.

The main problem each of these options have to solve is how to overcome echelle and prism dispersion to ensure a reference spot of the right size hits the detector in the right place. It is also necessary to avoid contamination of the science data by the optical reference. We need to start from the analysis of the echelle spectrum with the long camera. Here are the most relevant parameters:

- Spectrum length (300 to $1100nm$): $180mm$ in spatial direction.
- Longest orders: $216mm$ in spectral direction ($FSR = 40nm$).
- Shortest orders: $60mm$ in spectral direction ($FSR = 3.5nm$).
- Maximum order separation: $4.8mm$ at $300nm$.
- Minimum order separation: $2.4mm$ at $1100nm$.

It is also worth mentioning that the CCD covers the wavelength range from $300nm$ to $350nm$ when at the blue end of the spectrum and it covers (only partially) the range from $510nm$ to $1100nm$ when at the red end. By comparison, with the short camera it is possible to cover the region between $300nm$ and $650nm$ in the blue and between $330nm$ and $1100nm$ in the red. I will now analyze each option in detail.

8.4.2.1 Single-mode Laser

When using a source at a fixed wavelength, the direction of the beam after going through the prisms and the echelle, can vary by a very large amount. Suppose using a wavelength $\lambda = 400nm$, which is in the middle of the spectrum (in cross dispersion terms) in the order $m = 65$. The reference spot will be shifted by $\pm 60mm$ when the echelle spectrum is scanned in cross dispersion. In theory, it is be possible to compensate for these shifts by changing the angle of incidence (and consequently the deflection) on the cross dispersing prisms. This can be done by moving the reference in the slit plane. If we displace the source by s in the focal plane of the collimator, the beam leaving the collimator will be deflected by an angle $\Delta\sigma$ given by:

$$\Delta\sigma = \frac{s}{F_{col}} \tag{8.3}$$

With a camera focal length $F_{cam} = 1135mm$, the shift in its focal plane of $\pm 60mm$ subtends an angle of $\Delta\epsilon = \pm 3^o$. The relation between the angle of incidence and the deflection of a prism is approximately linear near minimum deviation. Therefore it is necessary to change the angle of incidence on the prism by roughly the same amount, to account for the different spectrograph setup. Thus we can equate $\Delta\sigma \sim \Delta\epsilon$ to obtain:

$$s \sim F_{col}\Delta\epsilon \simeq \pm 140mm \tag{8.4}$$

The amount of motion needed to ensure the reference hits the detector in the correct position is clearly too large. It is therefore impractical to use a single fixed monochromatic source as optical reference. An improvement could be to use more than one source or a laser that works at more than one wavelength.

A possibility is to use a Nd:YAG laser with second and third harmonic generator. This laser has a fundamental output wavelength of $1064nm$, which can be reduced to $532nm$ with frequency doubling and to $355nm$ through frequency tripling. The laser

has a linewidth of $0.08\mathring{A}$ at $1064nm$, which becomes narrower at shorter wavelengths ($0.04\mathring{A}$ at $532nm$). The three laser wavelengths are well suited for imaging at the edge of the detector, provided that reading from both ends of the CCD is possible. In this case the maximum motion required is probably only half the CCD size, or $\pm 30mm$. The motion in the slit plane would be reduced to:

$$s \simeq \pm 70mm \tag{8.5}$$

This range of motion is easier to accommodate in the slit area. Moreover, the use of two sources at either end of the slit may reduce the total motion even further, so this option may be viable. The fibre need not to be positioned extremely accurately, an error of one pixel on the CCD can be tolerated, when the reference needs to be within a 10 pixel range. This translates to an error in the slit plane of $\Delta s \simeq \pm 40\mu m$, which is a reasonable target. The main problem in this option may be obstruction from the post-slit components (filters, polarimetry, etc.) and the cost of the laser.

8.4.2.2 Continuum lamp + Monochromator

A variable monochromatic source can be produced by filtering a white light lamp through a monochromator. The continuum lamp can be, for example, a Quartz Tungsten Halogen (QTH) lamp and can be, if necessary, very luminous (up to $1000W$). The main problem in this option is the resolution of the monochromator compared to the echelle. With the long camera, HROS reaches a resolution of 120000 and in the blue a whole order, covering $60mm$ (4000 pixels) can be only $3.5nm$ wide in spectral terms. Also, the order extends $4.8mm$ in spatial direction (that is 320 pixels) because of the tilting introduced by cross dispersion. A commercial monochromator using a $1200lines/mm$ grating and with an output slit only $10\mu m$ wide has a bandpass of $0.1nm$. This means that in the focal plane of the HROS camera it would produce a spot of 115 pixels in spectral direction and 9 pixels in spatial direction. Using a double monochromator would halve the above values, but it would also reduce the throughput.

It is not clear if a 9x115 pixel (or a 5x57 pixel) spot is viable as an optical reference. Centroiding techniques can determine small shifts of very large images, provided the shape remains stable. If the science detector is used to record the reference position,

at least 20 CCD rows need to be reserved for this purpose. The monochromator throughput may be critical, too, particularly if we consider that it will need to be coupled with an optical fibre to deliver the source to the slit plane. Ultimately, however, the main objection to this option is the stability of the monochromator output. Since its resolution is much smaller than the one of the echelle, thermal effects or even flexure within the monochromator become extremely critical. Since the output slit width of $10\mu m$ correspond to 115 pixels, a shift of 0.1 pixels on the HROS CCD corresponds to only $9nm$ at the output slit, which is probably impossible to achieve.

To overcome the resolution gap, an etalon could be used in series with the monochromator. This, however, makes the optical source very complicated and it is likely that the stability between the monochromator and etalon position would become a main source of errors. Most of the discussion above is still valid even if the optical reference is imaged on the red end of the CCD, because at $350nm$ the FSR is still $4.8nm$ and the order separation is only slightly less.

8.4.2.3 Arc lamp + Monochromator

The use of an emission line lamp in place of the continuum source would solve the problem of the monochromator stability because the emission lines would provide the reference, rather than the monochromator bandpass. A wider output slit could be used, because the reference spectrum could cover a whole echelle order. It is possible to obtain a bandpass between 3.5 and $5.0nm$ with an output slit width between 0.5 and $1.5mm$. Moreover the position of the arc lines would not be affected by temperature and flexure within the monochromator. These reasons make this option very attractive.

The problem in this case is to find a spectral source with sufficiently dense emission lines. Most arc lamps have a very low emission at the blue end of the spectrum, so the detection on the HROS CCD (or on a separate detector) may have to occur on the top red end side or even on both sides. I will assume detection either on the science CCD or on a separate detector with similar pixel size, roughly 10 pixel by 3-4 thousand. In this case the spectral source will have to cover at least the range between $350nm$ and $1100nm$ (red end) or between $300nm$ and $550nm$ (blue end)

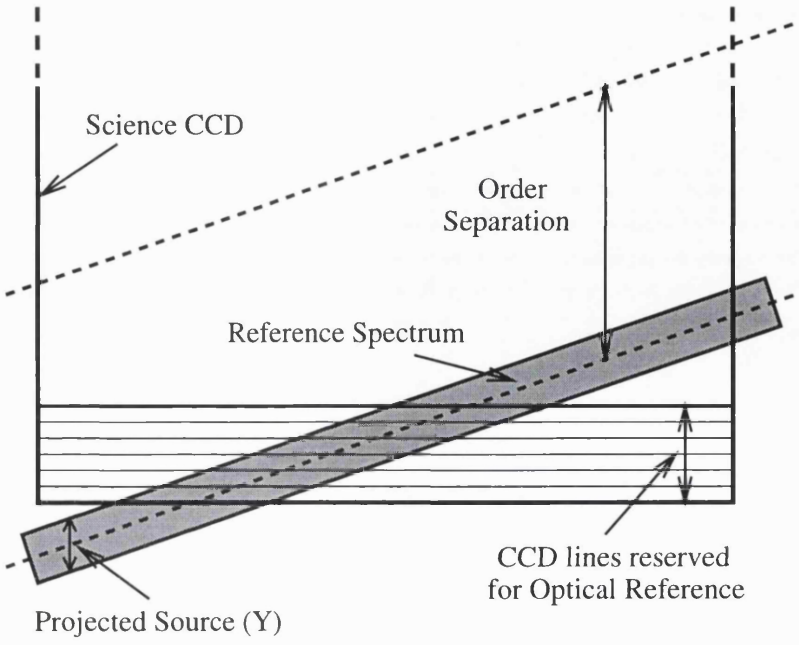


Figure 8.3: Reference spectrum projected on few lines close to the output register of the science CCD (or on a separate detector of similar size). Since only a small number of lines can be used and the orders are tilted with respect to the CCD, only a fraction of the echelle order can be used for the reference. The reference is assumed to project onto Υ pixels in cross dispersion direction. This graph assumes the CCD output register is almost parallel to the echelle orders. This may actually be not true in the case of the HROS CCDs, so a separate detector would be necessary.

with very dense lines. If the red end is chosen, the critical point is likely to be the density of lines around $350nm$. The order separation at that point is still about $4.0mm$ or 260 pixels. Therefore, if 10 pixels are reserved for the optical reference, they will cover a small fraction of the order, corresponding to only $1.8\mathring{A}$ for an ideal point source (figure 8.3). In practice however, the source will extend over Υ pixels in cross dispersion direction, the exact value depending on the size of the input fibre. For an extended source, projecting onto Υ pixels in cross dispersion direction, the range would be only slightly larger:

$$\Delta\lambda(\Upsilon) \simeq \frac{10p + \Upsilon}{10p}\Delta\lambda(10p) \tag{8.6}$$

If we take $\Upsilon = 2$ pixels, then the range is extended to $2.2\mathring{A}$. This is the worst case because it uses the maximum order separation and the maximum angular dispersion

at the blue end of the spectrum. At the red end the wavelength dispersion is much smaller both in the echelle and the cross dispersion directions (roughly by a factor 3 and 2 respectively). Nonetheless, approximately at least one spectral line every 5-6$\AA$ is required if only the first 10 pixels of the CCD can be used for the optical reference. This is not impossible to achieve, however, if a Th-Ar arc lamp is used as a source at the red end of the spectrum (probably a detector at both ends would be needed). The Th-Ar lamp has very dense emission lines, but some of them are very weak, so the throughput of the system becomes critical. The mean line separation for a Th-Ar lamp is 2$\AA$ and the maximum is 25$\AA$ in the UV [124], so the chance of having at least one line on the detector should be fairly high. The Th-Ar lamp feeds the monochromator at the input slit and this coupling should be designed for maximum throughput. At the output the monochromator feeds an optical fibre, which should also be chosen for maximum transmission. Overall, this option looks viable, even though further research is necessary.

In the case of an arc lamp and monochromator *backward propagation* is also an attractive possibility. True, the echelle spectrum is magnified by more than a factor two in the slit plane, but in this case a larger detector can be used. In fact, this may be the best option for detection of the optical reference in HROS. This is because the use of the science CCD seems an increasingly unlikely option for detecting the reference. Following a conversation with T. Boroson from the Gemini project, I found out that the current design for the HROS detectors is a mosaic of two 2048 $\times$ 4096 pixels CCDs with the output register on the 2K side. This means that the blind area at the junction between the two CCDs would be across the echelle orders if the output register needs to be parallel to them. Moreover, with the separation between the two CCDs parallel to the echelle orders it is possible to have a CCD optimized for blue light and one for the red, improving the spectrograph efficiency. Therefore, a separate detector for the optical reference would be needed. Since a separate detector increases the camera obstruction, the backward propagation becomes more competitive.

The full size of the echelle spectrum in the slit plane when the source is injected in the camera focal plane is shown in fig. 8.4. The spectrum is more than twice the size of the one in forward illumination. The magnification M is given by the ratio of

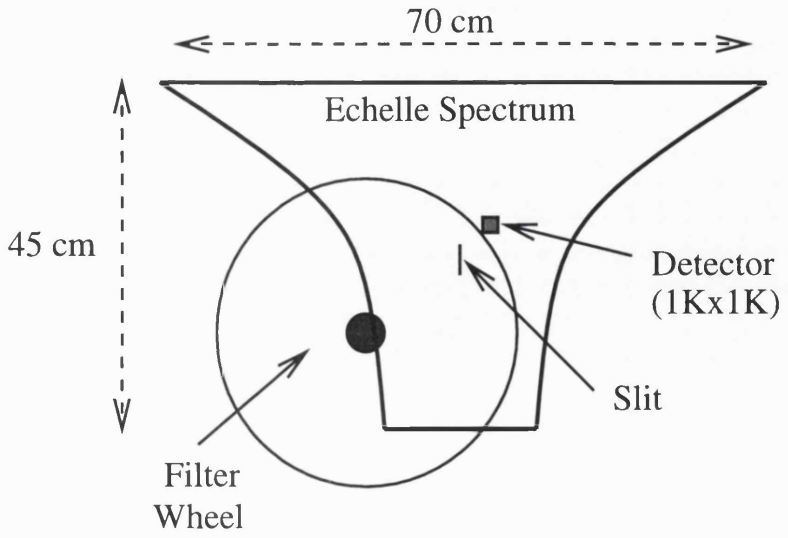


Figure 8.4: Echelle spectrum size in the slit plane when the optical reference for flexure compensation is propagated backwards in the HROS spectrograph. The filter wheel that obstruct some of the slit area is also shown. The figure shows the full echelle spectrum, but some of it may be cut off due to obstructions in the spectrograph. A detector 1000 pixels square ($15\mu m$ pixels) is shown for comparison.

collimator and camera focal length, which for the long camera gives $M = 2.3$. The order separation would then vary between 5.5 and $11mm$ and the angular dispersion would also be increased. The implication is that a fixed detector at the slit would have to be fairly large to ensure that at least one or two spectral lines of the Th-Ar lamp can be detected. The advantage here is that the CCD can be placed parallel to the echelle orders, maximizing the spectral coverage. A detector 1000 pixels square with $15\mu m$ pixels would have a sensitive area $15 \times 15mm$. This will ensure that at least one order is imaged on the detector and would cover between 4 and $9\mathring{A}$ in spectral direction (depending on the wavelength). This should guarantee at least one spectral line can be used for reference. This detector could also be moved in the slit plane to increase the chance of a good sampling of the reference spectrum.

8.4.2.4 Tunable Laser

The use of a tunable laser as a variable monochromatic source has the advantage of the very narrow spectral output with respect to the continuum lamp and monochro-

mator combination. The laser output is usually very narrow in spectral terms, so it would not spread over several pixels like the monochromator output. The main problems with the tunable laser are the spectral range and the spectral stability. Most lasers are tunable only over a narrow wavelength range, making it impossible to use them as a single source for every spectrograph setup.

Dye lasers can cover a wide range of wavelengths, but they require tens of different dye compounds to achieve that coverage. The laser will need to be tunable between $300nm$ and $550nm$ if it is to operate in the blue edge of the CCD. The range on the red edge is even wider, being between $350nm$ and $1100nm$. The dye laser option may be still viable if it is possible to use a dye mixture that covers the whole range. The only other broadly tunable laser is a Ti:Sapphire type which covers the range between 695 and $950nm$ in first harmonic and between 350 and $470nm$ in second harmonic. The difficulty in both cases is to find a single laser that can be tuned on a broad wavelength range. Using more than one tunable laser makes the system too complicated and expensive. Therefore this option too, seems impractical, mainly because remote operation of the source of reference looks very difficult to achieve and the cost of such system would also be very high.

8.4.2.5 Continuum source + Etalon

In this option a continuum source could be used together with a Fabry-Perot etalon to produce narrow emission bands from the continuum spectrum. This is the least attractive option for two main reasons. Firstly, the system would still require filtering, probably through a monochromator, to avoid contamination of the echelle spectrum and minimize stray light. Secondly, the etalon spectrum stability is very sensitive to temperature, while a powerful lamp close to the etalon would dissipate a lot of heat. For these reasons this option seems impractical.

Conclusion

The work exposed in this thesis was based on the innovative concept of using an active optical system to try to solve the long-standing problem of flexure in large astronomical spectrographs at the Cassegrain focus of a telescope. This project, which was subsequently called ISAAC (ISIS Spectrograph Automatic Active Collimator) involved designing, building and testing one of such active systems for the the ISIS spectrograph at the William Herschel Telescope.

Since this concept was almost completely new, I had to start from scratch: looking at other applications of active optics, reviewing modern actuators and position sensors and defining the goals and requirements for an active flexure compensation system. The characteristics and behaviour of flexure in ISIS had also to be investigated. My earlier visit to the La Palma Observatory was also designed to help me identify the technical challenges and to define the performance targets of the active collimator.

Then the system had to be designed and built, and here the challenge was not only from the stringent requirements for the position accuracy of the tip-tilt collimator mirror, but also from the need to be compatible and easily interchangeable with the existing collimator in ISIS. During this thesis I was responsible for the entire project and for most of the experimental work, including the overall system design. As project manager I delegated some aspects to specialists and in this respect, the mechanical design done by Mark Dryburgh, the mirror polishing by David Brooks and the electronics design by Gil Nixon and Heshmat Jamshidi has been invaluable help. I was also helped by David Brooks and Francisco Diego with the alignment and design of the optical test bench and with the polishing of the flat mirror test sample and by Heshmat Jamshidi with the writing of the control software. The physics mechanical workshop produced most of the metal work, while Brian Humm helped with the mirror cutting and aluminizing.

Once the prototype was completed, I had to thoroughly test it in the laboratory, both in its optical and mechanical performance, to verify that it would meet the

specifications. These tests needed not only to convince myself, but also the technical staff at the La Palma Observatory that the active collimator was fully reliable and performing as expected. Finally, on July 1995, I had a chance to test ISAAC at the William Herschel Telescope. The experiment was a very successful: active compensation of flexure proved to be a viable and effective way of reducing flexure in large Cassegrain spectrographs.

The success of the experiment in fact opened the way for several new potential developments, including having a flexure active compensation system permanently installed on ISIS and one included in the design of HROS from the beginning. The Gemini Multi-Object Spectrograph (GMOS) [53] and the DEIMOS and ESI spectrographs for the Keck II telescope [54, 55] will also probably include an active flexure compensation system in their design. It seems that the idea of active compensation of flexure is gaining increasing support among spectrograph designers. This is probably because active flexure compensation makes the design of a spectrograph much simpler, and the instrument cheaper and lighter.

I hope that this thesis will contribute to the construction of better instruments. The research described in the thesis is however far from concluded: I am currently investigating the possibility of a closed loop compensation system, which would improve the performance of a simple open loop one. In the last year I also examined various possibilities for a closed loop flexure compensation system for HROS, which would greatly simplify its mechanical design and would allow for the first time high resolution spectroscopy at the Cassegrain focus of a large telescope. I believe the understanding of flexure in Cassegrain spectrographs has been increased, by some detailed measurements on ISIS I performed in cooperation with Rene' Rutten at the La Palma Observatory. This information can be very useful to the design of future Cassegrain instrumentation.

I hope that the work in this thesis would add some contribution to the design of modern instrumentation for new ground-based telescopes and, in doing so, ultimately to our knowledge of the universe.

Acknowledgements

I wish to thank all the people at OSL for their friendship and support, starting with Dr. David Walker who gave me the opportunity to take a Ph.D. course in the laboratory and Dr. Francisco Diego, who supervised my work. Everyone at OSL was a great help during the last three and a half years, and everyone contributed in their own way. Therefore, I prefer to mention their names rather than their individual contribution: Bruce Bigelow, Richard Bingham, David Brooks, Andy Charalambous, Mark Dryburgh, Lee Hubbard, Brian Humm, Heshmat Jamshidi, Yalda Jamshidi, Do Hyung Kim, Ee Eul Kim, Ho Soon Kim, Sug Whan Kim, Young Soo Kim, Jun Ho Lee, Gil Nixon, Nigel Pugh, Alan Radley, David Rees, Trevor Savidge, Ron Szumski. Among the people at UCL whom I also wish to mention are: Ian Crawford, John Deacon, Orsola De Marco, Adrian Fish, Ian Howarth, Linda Smith, Serena Viti. I also wish to thank the people at the La Palma Observatory, in particular Rene Rutten; Mick Johnson and Sue Worswick at the Royal Greenwich Observatory. Finally I wish to thank my parents and my sister for their love and support. I must thank Yalda Jamshidi once again for helping me writing these acknowledgements.

Appendix A

Calculation of Actuator Expansions

The software controlling the active collimator needs to calculate three actuator expansions (L_1, L_2, L_3) from two shifts $(\Delta X, \Delta Y)$ on the CCD detector of the ISIS spectrograph. The software initially translates these two shifts into collimator tilts $(\vartheta_j, \vartheta_k)$. Following the notation of section 5.4.3.2, I will identify the collimator plane with (j, k) and assume that this is parallel to the detector plane when the actuators are at zero expansion. If we indicate with γ the angle between the cartesian axes (j, k) and (X,Y), then it is possible to calculate the shifts $(\Delta J, \Delta K)$ by a simple transformation of coordinates. The anamorphism of the grating ξ is accounted for by substituting the shift ΔX with the "anamorphized" $\xi \Delta X$. Thus:

$$\Delta J = \xi \Delta X \cos\gamma - \Delta Y \sin\gamma \tag{A.1}$$

$$\Delta K = \xi \Delta X \sin\gamma + \Delta Y \cos\gamma \tag{A.2}$$

The collimator tilts can then be derived by imposing that the tilt should restore the image to the starting position (the origin). The tilting angle is determined by the camera focal length F_{cam} and the shifts in the (j, k) plane. Since the angles are very small and I indicated with ϑ_j the tilt *about* the j-axis, we obtain:

$$\vartheta_j = \frac{\Delta K}{2F_{cam}} \tag{A.3}$$

$$\vartheta_k = \frac{\Delta J}{2F_{cam}} \tag{A.4}$$

The factor 2 in the equations above is due to the fact that the light coming from the spectrograph slit is reflected by the collimator and if this were tilted by an angle β, the beam would be deflected by an angle 2β. Combining these equations with the expression for ΔJ and ΔK we obtain:

$$\vartheta_j = \frac{1}{2F_{cam}}(\xi \Delta X \sin\gamma + \Delta Y \cos\gamma) \tag{A.5}$$

$$\vartheta_k = \frac{1}{2F_{cam}}(\xi \Delta X \cos\gamma - \Delta Y \sin\gamma) \tag{A.6}$$

When the collimator tilts are known, it is possible to calculate the actuator expansions which would produce that tilt. Following the conventions of fig. 5.15, we can see that:

$$\vartheta_j = \frac{(L_1 + L_2)/2 - L_3}{h} \tag{A.7}$$

$$\vartheta_k = \frac{L_1 - L_2}{b} \tag{A.8}$$

We obtain therefore a system for the three unknowns L_1, L_2 and L_3:

$$\begin{cases} 2h\vartheta_j = L_1 + L_2 - 2L_3 \\ b\vartheta_k = L_1 - L_2 \end{cases} \tag{A.9}$$

Since there are three unknowns and only two equations, we have freedom in choosing one of the unknowns. I chose to set to zero the expansion of the actuator with the lowest position. Which actuator has the lowest position will depend on the sign and value of the two tilting angles. Following again fig. 5.15, we can calculate this explicitly. First we notice that $L_1 = L_2$ when $\vartheta_k = 0$. Also we have:

$$L_1 = L_3 \quad \Rightarrow \quad \vartheta_k = -\frac{2h}{b}\vartheta_j \tag{A.10}$$

$$L_2 = L_3 \quad \Rightarrow \quad \vartheta_k = \frac{2h}{b}\vartheta_j \tag{A.11}$$

Therefore we can identify three different cases:

1. If $\vartheta_j \leq 0$ and $\vartheta_k \leq 0$ or if $\vartheta_j > 0$ and $\vartheta_k < -\frac{2h}{b}\vartheta_j$ the actuator #1 will be the one with the lowest position, so we can impose $L_1 = 0$.

2. If $\vartheta_j \leq 0$ and $\vartheta_k > 0$ or if $\vartheta_j > 0$ and $\vartheta_k > \frac{2h}{b}\vartheta_j$ the actuator #2 will be the one with the lowest position, so we can impose $L_2 = 0$.

3. In every other case the actuator with the lowest position would be #3, so we can impose $L_3 = 0$.

For each one of the above cases the system is then reduced to two unknowns and two equations and can be solved to derive the actuator expansions. Note that since the expansion of the actuator with the lowest position is set to zero, the expansions of the other two actuators will be always positive. This is important because it ensures that only positive voltages are applied to the actuators.

Appendix B

Publications

B.1 List of Published Papers

The papers I published during this thesis are included in this appendix. Below is a complete list of these publications.

- B.C. Bigelow, D.D. Walker, R.G. Bingham, P. D'Arrigo, "A deformable secondary mirror for adaptive optics," Proc. ICO-16 Satellite Conference on Active and Adaptive Optics, Garching/Munich, Germany, 2-5 August 1993, pp. 261-266.

- P. D'Arrigo, F. Diego, D.D. Walker, "Active Compensation of Flexure on ISIS," *Spectrum,* Newsletter of the Royal Observatory, No. 9, March 1996, pp. 25-27.

- D.D. Walker and P. D'Arrigo, "On the stability of Cassegrain spectrographs," *Mon. Not. R. Astron. Soc.* **281**, 673-678 (1996).

- P. D'Arrigo, F.Diego and D.D. Walker, "Active compensation of flexure on the WHT ISIS spectrograph," *Mon. Not. R. Astron. Soc.* **281**, 679-686 (1996).

- P. D'Arrigo, F. Diego, D.D. Walker, "Active compensation of flexure on the High Resolution Optical Spectrograph for Gemini," Proc. of the SPIE/ESO Symposium on *Optical Telescopes of Today and Tomorrow,* Landskrona, Sweden, May 29 - June 2 1996 (in press)

A DEFORMABLE SECONDARY MIRROR FOR ADAPTIVE OPTICS

Bruce C. Bigelow [1], David D. Walker [1], Richard G. Bingham [1,2], Paolo D'Arrigo [1]

[1] Optical Science Laboratory, University College London
[2] Royal Greenwich Observatory, Madingley Rd, Cambridge

Abstract

The design of a proposed deformable Cassegrain secondary mirror for adaptive optics is presented. The 1-m diameter thin meniscus mirror is supported on 90 actuators, and is designed for use in a Gemini 8-m telescope. In addition to the adaptive function, the mirror design also provides the capability for vibration balanced IR chopping within the unit. The mirror geometry, materials, actuators, and finite element analysis are described.

1 Introduction

Adaptive optical systems are rapidly changing from R&D exercises to indispensible facilities for modern telescopes. Most systems currently described in the literature are of equal or greater complexity than many instruments on a telescope, involve reflective losses at many surfaces, and do not provide a corrected wavefront to most telescope focal planes. Applying an adaptive correction at the secondary mirror has been suggested by several authors (Dunn, 1991, and Beckers, 1993), and was proposed in the Technical Description for the NOAO (Gemini) 8-m Telescopes[5], 1989. There are many advantages to this approach: the correction is made at an existing surface, precluding the need for extra mirrors and reflective losses; Tip/tilt and low-order corrections are provided at one surface; adaptive correction is provided for all Secondary foci; and existing telescopes can be retrofitted. The design also provides chopping for IR sky subtraction as an extension of the tip/tilt capability. A companion paper (this symposium) describes the implications of adaptive correction at the secondary in terms of wavelength, guide stars, and image improvements. This paper describes the conceptual design and analysis of a 1-m thin meniscus deformable secondary mirror, suitable for the Gemini 8-m telescopes, and discusses the mirror specifications, geometry, materials, actuators, detail design, failure modes, and finite element analyses of the mirror substrate.

2 Performance Specifications

The objective was to devise a realistic option to the traditional chopping secondary mirror, which could also provide partial adaptive correction for the Gemini Telescopes. Consequently, the parameters of the mirror and its performance were optimized for that project. The dimensions of the Gemini f/16 IR Secondaries[6] (1992) are as follows: radius of curvature = 4193.0685 mm, clear aperture = 1023.6 mm (outer), 180.7 mm (inner).

The adaptive requirements are: tip/tilt mirror angular throw: ± 1 Arc-sec (for 90% of atmospheric tilt); tip/tilt frequency response: 20-30 Hz; higher order frequency response: 100Hz minimum, 1KHz goal; vibration control: 95% vibration balanced.

Chopping for IR sky subtraction is a low-frequency rigid body rotation of the mirror, with the angular throw limited by the travel of the actuators. The following specifications were taken from the Gemini Science Requirements[7]: Chopping Mode 1: mirror angular throw: ± 30 arc-sec; frequency response: 10 Hz; Duty Cycle: 80%. Chopping Mode 2: mirror angular throw: ± 60 arc-sec; frequency response: 5 Hz Duty Cycle: 80%

3 Mirror Geometry

One figure of merit for chopping mirrors is the angular moment of inertia, which determines how much energy is required to vibrate the mirror, and places practical limits on the duty cycle. For a thin flat disc, the moment of inertia scales directly with mass and quadratically with the radius and thickness. For a meniscus secondary mirror, approximated by a thin disc of fixed diameter, the mass and thickness can be minimized in order to minimize the power requirement and maximize the duty cycle. Higher order corrections require that the mirror be sufficiently flexible that it can be rapidly deformed to the desired shape. This flexibility also allows the mirror figure to be updated many times during each chopping cycle, given a reference. Both the low angular inertia and flexural requirements are well suited by thin meniscus construction. Typically, a mirror needs to be stiff enough to hold its figure during grinding and polishing. However, as this mirror is inherently deformable, smoothness is probably more important than the unsupported figure. In this case, a 10-mm thick mirror was analysed and found to combine sufficient stiffness under gravity, with acceptable flexibility for adaptive correction.

4 Mirror Materials

Traditional deformable mirrors have been small (on the order of 100 to 200 mm in diameter) and have used materials such as silicon or molybdenum, as well as more common optical glasses such as ULE, fused silica, and Zerodur. A variety of issues drive the selection of the substrate material, such as elastic strength (modulus of elasticity), thermal properties (coefficients of expansion and conductivity), durability (resistance to fracture) fatigue life (critical for an adaptive mirror that will experience billions of cycles per year), ease of fabrication (optical and mechanical), surface finish quality, and resistance to corrosion.

A thin deformable mirror will thermalize quickly, and its figure can be corrected, via the wavefront sensor, for errors due to thermal gradients. Consequently, the thermal properties of the mirror appear to be relatively unimportant. Stress-cycling fatigue life, or endurance limit, is more critical. A deformable mirror operating at 500 Hz, 8 hours per day, 300 days per year, will experience 4.3 billion stress cycles annually, but endurance limit testing typically specifies allowable stresses for 50-500 million cycles. However, results from finite element analyses indicate peak stresses 10 times less than the fatigue limits of the candidate materials, so although fatigue life is important, it does not appear to dictate the choice of material. The final choice will probably hinge on the ease of fabrication. Figuring and polishing of aluminum mirrors is well understood, and aluminum could be used in nearly all the mirror mechanical components. In terms of mechanical simplicity, a metallic mirror, which can be be tapped for threaded fasteners, and which is relatively resistant to damage or breakage, seems to provide the best balance of features. However, attachment points are commonly bonded to glass mirrors, and large optics must be handled carefully regardless of material, so Zerodur or ULE are viable alternatives.

5 Actuators

The actuators will provide support under gravity, and displacement for adaptive correction and chopping. These functions place differing requirements on the travel, force, and frequency response of the actuators. Chopping places the most stringent displacement requirement on the actuators. Chopping of 7.5 and 15 arc-sec at the focal plane corresponds to secondary mirror tilts of $\pm$ 30 and $\pm$ 60 arc-sec respectively. Given those tilts, the resulting displacement ranges for the extreme edge of the mirror are $\pm 72\mu$m and $\pm 145\mu$m. Results from the finite element analyses indicate that the required actuator forces are less than 3 N for gravity support and 40 N for 2.2μm of adaptive deformation. The force values required for chopping are dependent on

the duty cycle, which defines the acceleration required to move a given mass element supported by one actuator at the edge of the mirror. These forces are less than 1 N for both chopping modes. The frequency response requirement will be determined by the adaptive correction, previously mentioned to be 100Hz minimum (1KHz goal), for higher order corrections. This value could increase substantially without causing a problem for the actuators, most of which can be operated at kilohertz rates.

Historically, **Piezo-electric** actuators have been the standard choice for adaptive mirrors. The high hysteresis and temperature sensitivity of PZT materials can be effectively controlled by the addition of displacement sensors, which will be required regardless of actuator type. Although standard models with 100μm of travel are common, no PZT actuator has been found which can meet the 300μm stroke requirement. The longest travel available in commercial PZT actuators appears to be 180μm, which corresponds to a chopping angle of $\pm$ 37 arc-sec. Lifetime is a concern for PZT actuators, as it is counted in millions, rather than billions of cycles required for a year of mirror operation. The limitations in stroke and lifetime appear to eliminate PZT actuators.

Electro-Strictive Actuators in variety of stroke and force combinations are commercially available. However, there do not appear to be standard ES actuators with strokes as high as $\pm$145μm, and the long strokes are limited to operation above 10C. Below 10C, the electro-strictive effect is dropping while hysteresis is rising. The primary advantage of ES actuators over PZT's is lower hysteresis, but this can be compensated with position sensors in either case. For this application, there is no advantage in using ES actuators.

Two types of **Magneto-Strictive** (MS) actuators have been identified, each of which has the potential to meet all of the actuator specifications. The first is a MS "Elastic Wave Motor" (EWM) with a range of **100 mm**, manufactured by DynaMotive[3]. This actuator can generate long travel, low frequency motion as a stepper motor, and short travel, high frequency motion as a simple actuator. The actuator has high static and dynamic axial stiffness, together with low mass and relatively small dimensions. Its dual mode operation provides long travel and low power dissipation. The second type is a conventional actuator, available from Hull University[8] and Edge Technologies Inc. (ETI)[4]. According to ETI, "Terfenol-D does not fatigue", and they have tested actuators to 10^{10} cycles. In both cases, resolution is mainly limited by the displacement sensor used, making MS actuators the preferred type for the deformable secondary.

6 Mirror Detail Design

Support of thin mirrors has been a challenge for large telescopes such as the VLT and the two Kecks, which use high aspect ratio primaries. Nelson et al.[9] examined plate deflection on point supports, derived expressions for the efficiency of hexagonal, square, and triangular support grids, and optimized the locations of support points for various numbers of supports on round plates. They conclude that a modified triangular grid provides the lowest RMS deflection, and that deflections are extremely sensitive to the location of the support points. For the case of a 1 meter diameter, 10mm thick, aluminum (or Zerodur) mirror, and a 30nm maximum allowable gravity deflection, a solution utilizing 90 support points has been found. The layout of the support points, though not mathematically optimized, was developed following Nelson's guidelines and iterated via FEA models to the 30nm specification.

The adaptive secondary mirror concept scales up traditional adaptive mirror construction, and consists of a flexible face-sheet supported on a structured aluminum mounting plate via an array of 90 actuators (Figure 2). Counter-balances on the opposite side of the mounting plate are supported by additional actuators, located in line with the connections to the telescope top-end ring. The counter-balance control signals can be derived from computations based on the signals sent to the mirror actuators, or from high frequency load cells incorporated in the telescope attachment, which would allow vibration to be sensed and compensated directly.

Chopping is accomplished by driving the actuators simultaneously, so as to rotate the mirror as a rigid body about its center of gravity. Linear momentum is conserved and angular momentum compensation is provided by driving the reaction masses about their composite center of gravity. This arrangement provides several advantages. First, the low mass meniscus reduces the weight of the chopping mirror, and allows for higher duty cycles and lower power consumption. Second, the moment of inertia for the meniscus is much lower than that of a monolithic mirror, thus reducing the resultant torque to be compensated by the reaction mass(es). Reducing the weight of the reaction masses also reduces the force requirement for their actuators, providing further weight and cost savings. The adaptive secondary with less mass, smaller moment of inertia, and many actuators, is inherently easier to chop than a monolithic mirror, and less subject to figure errors due to chopping forces. With an adaptive correction running at a higher frequency than the chopping function, the wavefront could be corrected many times during each chop cycle.

7 Failure Modes

For graceful loss of functionality after failure, three capabilities should be present: first, the secondary must operate with a failed actuator, and its loss should only marginally reduce performance; second, no actuator or control error should be able to damage the mirror; and finally, replacement of failed actuators must be simple, avoiding both disassembly of the secondary mirror system and removal from the telescope. The loss of an actuator will affect the ability of the mirror to maintain its figure, but as each actuator drives roughly 1% of the mirror, a single failure should have a minimal effect on performance. Nevertheless, concepts have been developed for a clamp mounting for the actuators (mechanical or electro-mechanical), which will permit an actuator to be released from the mirror. This will not eliminate errors caused by the loss of the actuator, but could minimize the effect until the actuator could be replaced. The greatest threat to the mirror would be a control error which caused one actuator to generate maximum stroke or force, while its neighbors were driven to the opposite extreme. This has been examined for a 180μm stroke, using the finite element model, and found to produce stresses sixty times smaller than the yield stress of aluminum (6061-T6). Nevertheless, a force-limiting assembly utilizing a spring concept is proposed, which would absorb excess stroke or force in the event of a control failure. This assembly acts as a stiff link up to a given preload, after which the springs compress, absorbing further travel. This link will be carefully designed in order to provide a margin of safety, while avoiding reduction of the resonant frequency of the actuator. For reliability, actuator failures should not require a complete disassembly of the secondary mirror. Various methods for removing and replacing actuators have been described, including the destruction of a flexural link[2], and separable magnetic attachments[10]. The previously mentioned clamps should allow the actuators to be installed on the mirror without excessively deforming it, but this would require a test, possibly using the signal from the wavefront sensor, to verify the quality of the power-off figure.

8 Finite Element Analyses

Approximately 25 finite element models were analysed, for estimating static deflection, actuator forces, optimizing support locations, and predicting resonant frequencies and mode shapes. The models used 1500 4-node shell elements, with 9000 degrees of freedom. All modelling was performed using the ANSYS finite element analysis package. A summary of the results indicates that an aluminum (or Zerodur) meniscus, supported on 90 points, will have static deflections (zenith pointing) on the order 40nm P-V, and 17nm RMS (Figure 1). Deflections on horizon pointing are a factor of 10 less. Actuator forces of 3.7 N generate a displacement of 0.20μm at

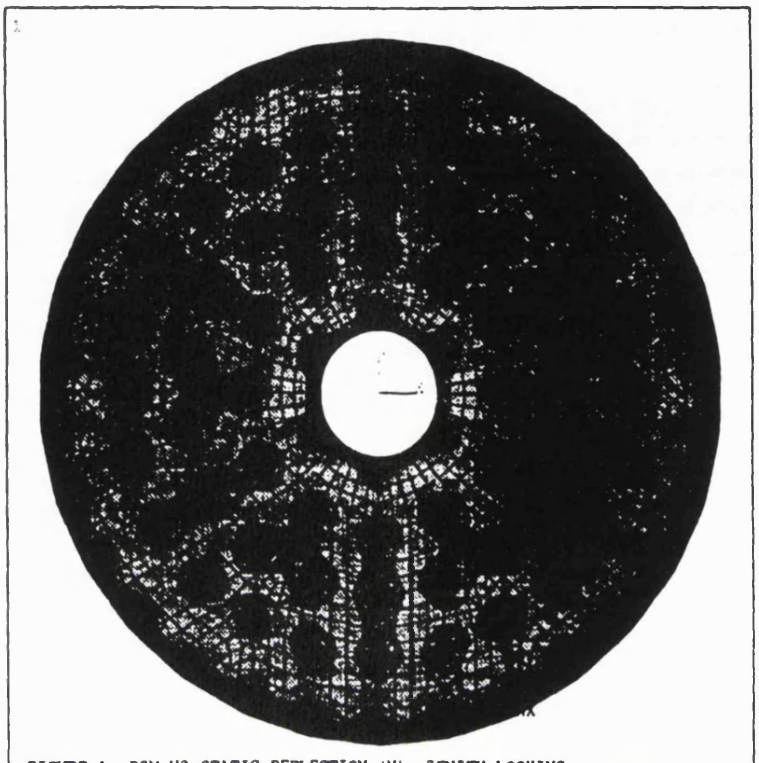


FIGURE 1. DSM V2 STATIC DEFLECTION (M), ZENITH LOOKING

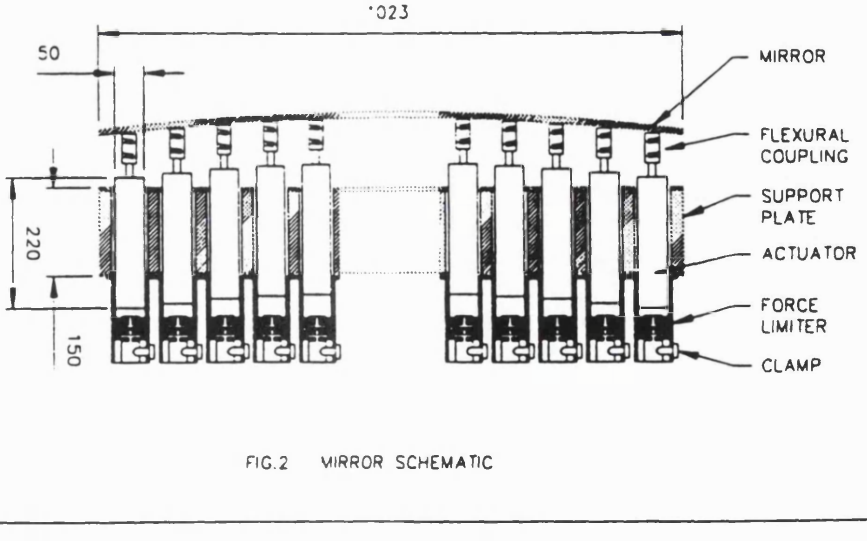


FIG.2 MIRROR SCHEMATIC

any actuator and scale linearly. The maximum stresses at 2.0μm deflection are less than 1 MPa. Dynamic analyses show that the first mode looks like the inter-actuator sag under gravity, with a resonant frequency in excess of 1200 Hz. To conclude, the finite element analyses indicate that modest forces are required to deform the meniscus for adaptive correction, bending stresses are negligible, and the resonant frequencies of the mirror are at least 20% higher than the estimated peak operating frequency.

9 Conclusions

This work has shown the mechanical feasibility of providing adaptive wavefront correction and IR Chopping with a deformable cassegrain secondary mirror. A survey of mirror substrate materials suggests that aluminum, ULE, and Zerodur are possible choices, with aluminum preferred for simplicity of attachment and mechanical durability. Magneto-strictive actuators appear to provide the best combination of travel, stiffness, frequency response, and durability. Finite element analyses show that static deflection, actuator forces, and dynamic response are all well within achievable values. The next two tasks in this work will be to fit Zernike polynomials to the FEA results, in order to define the optical limits of adaptive correction with the deformable secondary, and to begin detailed preparations for building a mirror for an 8-m telescope.

10 Aknowledgments

Much of this work was completed under a feasibility study contract with the Gemini project, UK office at Oxford, and we gratefully acknowledge their support.

References

[1] J. M. Beckers, "Adaptive Optics for Astronomy: Principles, Performance, and Applications", Annual Review of Astronomy and Astrophysics, Vol. 31, 1993.

[2] R. B. Dunn, G. W. Streander, W. Hull, L. Wilkins, "The NSO/Sac Peak Continuous-Face-Plate Adaptive Mirror", Proceedings of the SPIE, Vol. 1543, p. 88, 1991

[3] DynaMotive Corporation, 3760 Wesbrook Mall, Vancouver, B.C. V65 2L2 Canada

[4] Edge Technologies Inc., 306 S. 16th, Ames, Iowa 50010 USA

[5] The NOAO 8-M Telescopes. Technical Description Vol. II, a Proposal to the National Science Foundation, AURA, Sept. 1989.

[6] Gemini Drawing, 85-GP-1000-001, Sept. 1992

[7] Gemini Science Requirements, Ver. 1.1, 11 Nov. 1992

[8] Rod Greenough, University of Hull, UK, Private Communication

[9] J. E. Nelson, J. Lubliner, T. S. Mast, "Telescope Mirror Supports: Plate Deformations on Point Supports", Proceedings of the SPIE, Vol. 332, p.212, 1982

[10] C. D. Swift, J. W. Bergum, E. S. Bliss, F. A. House, M. A. Libkind, J. T. Salmon, C. L. Weinzapfel, "Zonal deformable mirror for laser wavefront control", Proceedings of the SPIE, Vol. 1543, p. 107, 1991.

Active Compensation of Flexure on ISIS

ISAAC (ISIS Spectrograph Automatic Active Collimator) was developed at the Optical Science Laboratory, University College London, as an experimental system for active compensation of flexure. Whilst the basic motivation was a proof-of-concept for the Gemini High Resolution Optical Spectrograph (HROS), it was clear that ISIS exhibited flexure at a level where it could benefit from the active collimator approach we had proposed.

ISAAC is based on the concept of active compensation, where spectrum drifts due to the spectrograph flexing under the effect of gravity are compensated by the movement of an active optical element. The active component in ISAAC is a fine steering tip-tilt collimator mirror, optically an exact replica of the existing one. It was successfully tested on the spectrograph on July 1995. The experiment shows that active compensation can reduce flexure to less than 3μm over four hours of telescope motion, dramatically improving the spectrograph performance.

The traditional method of coping with flexure, which involved building bulkier and stiffer structures, is reaching its limits. In recent years flexure has proven to be an issue difficult to solve in Cassegrain spectrographs. Active compensation offers a new approach in which the image shifts on the detector are constantly corrected through motions of an active element within the spectrograph. Flexure compensation can be achieved either with an open-loop system, where the drifts are accurately measured and stored in a look-up table or with a closed-loop one, where spectrum shifts are sensed directly using an appropriate optical reference and fed back to the active element control.

The ISAAC project was conceived as an open-loop system for active compensation of flexure in ISIS. To achieve this the collimator is finely steered with three piezoelectric actuators pushing against its back surface. The open-loop system involves controlling and sensing the collimator position, knowing the orientation in space of the spectrograph, and correcting the corresponding flexure shifts using the data stored in the look-up table. The instrument, as well as being the first step towards a closed-loop system, is also the baseline solution for the Gemini High Resolution Optical Spectrograph (HROS).

Experiment on ISIS

On July 1995 we obtained access to ISIS for daytime tests at the William Herschel Telescope. The aim was to study flexure in ISIS, monitor it effectively and then prove that we could compensate it with the active collimator. The ISAAC collimator was installed on the ISIS blue arm and its focus and tilt were adjusted using the standard Hartmann shutter. The optical quality was checked by imaging a narrow slit illuminated by CuAr and CuNe calibration lamps. The results were at least as good as those obtained with original collimator.

The fine calibration of the mirror movements was performed with the collimator optically aligned in the spectrograph. The three actuators were moved in turn

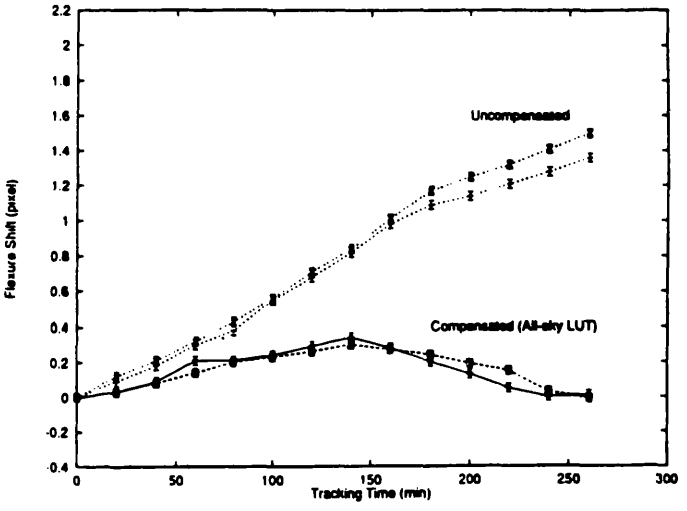


Fig 1 – Flexure shift, spectral direction.

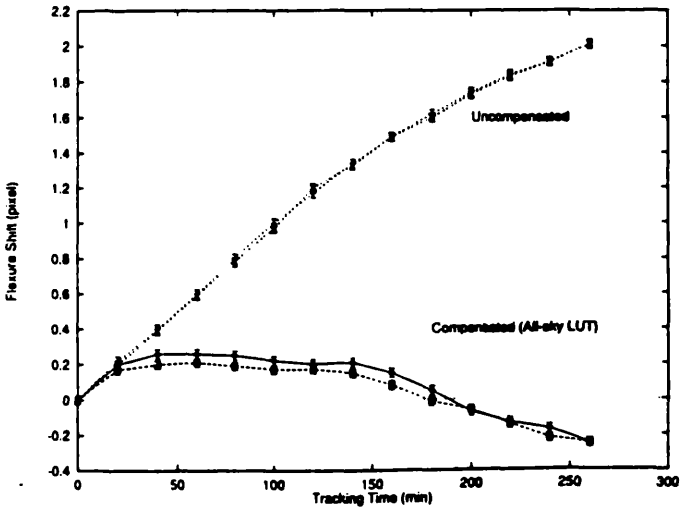


Fig 2 – Flexure shift, spatial direction.

and in different combinations and the corresponding detector shifts recorded. From this the parameters relating the actuator voltages to detector shifts were calculated. The reference spectra were generated by observing an arc spectrum through a short dekker, so that each spectral line would produce an almost circular spot, which could be centroided to give the spectrum shifts (in spectral and spatial direction).

The flexure look-up table was produced using an automated procedure developed by René Rutten to monitor flexure in ISIS. The resulting map covers the whole range of possible telescope elevations and spectrograph positions on the turntable. The flexure map thus obtained was reformatted to reduce the impact on the flexure model of hysteresis in the telescope movements. The flexure measurements and all the tests for active compensation were performed during the daytime, using the calibration lamps to feed the spectrograph.

In our first experiment we simulated a star track by moving the telescope in 20 arcminute steps following the motion of a virtual star on the sky. In this way we managed to compress more than four hours of effective tracking into less than one hour of measurements. We compensated for the flexure shifts using the all-sky look-up table and then repeated the same track without compensation. We also repeated the same measurements the following day to check how reproducible were these results. The compensated and uncompensated spectrum shifts are shown in figures 1 and 2 for the spectral and spatial directions.

Several conclusions can be drawn from the graphs. Firstly, that on average, active compensation of flexure can reduce spectral drifts to within 0.2 pixels (worst case 0.3 pixels) over more than four hours tracking. We must stress here that the exposure time during astronomical observations is anyway limited to around one hour by accumulation of cosmic ray events on the CCD. Our experiment of compensating flexure over several hours effective tracking proves that large amount of flexure can be corrected with the same accuracy. This is highly relevant both for HROS, which may experience such large spectrum shifts, and for ISIS, by potentially reducing the need of frequent calibration arcs between exposures.

The image stability is a significant improvement on the existing situation, with typical drifts of 1.5 – 2.0 pixels during the same tracking time. The graphs also show that the improvement is more relevant for longer tracking time. This is because the uncompensated flexure usually increases fairly smoothly (at an average rate of 10μm/hr), whereas the inaccuracy in compensation depends mostly on residual errors and very little on the amount of flexure to correct.

Finally, the graphs prove that when the uncompensated flexure is reproducible, the curve representing the residual errors is also highly reproducible indicating that these errors are mostly due to inaccuracies in the flexure model and not in the collimator movements.

This last point induced us to refine the look-up table, to see if we could improve image stability even further. For the same track, we calculated the residual errors and fed them back into the active collimator control system as second order corrections. In practice the proce-

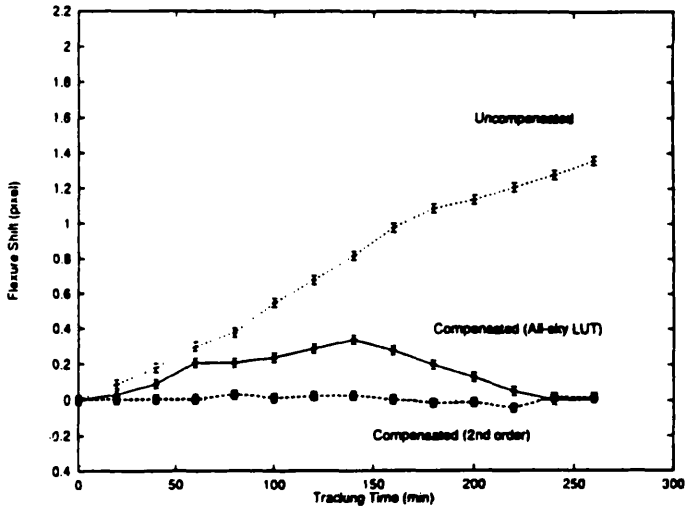


Fig 3 – Second order compensation, spectral direction

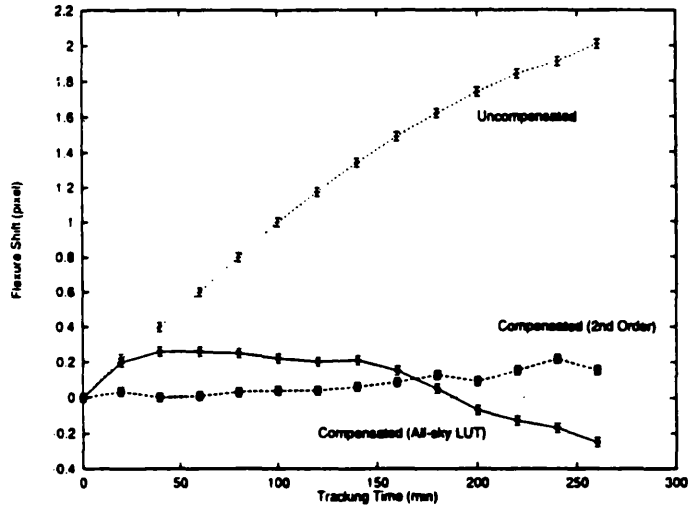


Fig 4 – Second order compensation, spatial direction

dure involved stepping along a track with active compensation on, record the residual shifts, convert them into actuator voltage corrections and then run along the same track again applying the corrected voltages. Results using these second order corrections are shown in figures 3 and 4 compared with the standard all-sky compensation. The improvement is quite remarkable. Once the errors in the flexure model are corrected, the spectrum drifts can be contained within 0.1 pixel for nearly four hours of effective tracking time (more than four hours in the spectral direction).

Future developments

This experiment has proved that active compensation is a viable and effective way of reducing flexure in ISIS. It has also shown that the collimator can be controlled with very high accuracy and it is a suitable element for applying active corrections. The improvement in imaging stability is indeed significant: during our experiment ISIS was showing an average spectrum drift of over 40μm in four hours tracking time. By applying active compensation we were able to reduce it to an average 4.8μm over four hours with an all-sky flexure map and to less than 2.4μm with a flexure look-up table modelled for that particular observation.

These results have several implications on the astronomical data the spectrograph can provide, not just in terms of increased spectral resolution and reduced radial velocity errors. The reduction in flexure we achieved is such that the time wasted in taking regular calibration arcs can be cut drastically. Moreover, this low flexure can allow longer exposure times and improved signal-to-noise ratio. The limiting factor in the length of exposures would be effectively shifted from flexure drifts to cosmic ray events rate. Our laboratory tests had demonstrated that the collimator could provide movements as small as 0.4μm on the detector and we could expect that flexure compensation could be improved even further with a more accurate model. Another result of our experiment was to prove the existence of a relatively large hysteresis in the spectrograph distortions. This is likely to put severe constraints on the accuracy a flexure model can ultimately achieve. Moreover, we have to take into account that to collect the data necessary for an accurate flexure map takes considerable telescope (day) time, which is often needed for engineering and maintenance. ISAAC in combination with an all-sky flexure map is therefore a compromise solution, balancing performance in image stability versus time required to achieve that. Even so, it is clear that ISIS can benefit from having ISAAC permanently installed as a common user instrument. We are currently discussing this possibility with staff at the Royal Greenwich Observatory and at the Isaac Newton Group of telescopes.

It is clear however, that to exploit fully the potential of the active collimator itself, a different approach is needed. By closing the loop with an optical reference beam propagated through the spectrograph it is possible to avoid the problem of hysteresis altogether. The flexure shifts would be sensed in real time while the spectrograph is tracking and directly fed back into the collimator control. It also would avoid the need of producing several different flexure models for corresponding different spectrograph setups.

We have submitted a grant proposal for a closed loop experiment on ISIS. The optical reference would be provided by a calibration lamp source, injected with an optical fibre at one end of the entrance slit, so as to be imaged at one end of the science CCD. The few lines close to the output register of the CCD where the reference is imaged would be read periodically during an exposure (say every 1 – 2 minutes) with the shutter closed. After reading these few lines, the vertical clocks of the CCD would be reversed to restore the charge in the science image to its original position. The exposure would be continued.

We can conclude that our recent experiment has shown that active compensation of flexure is a real option in designing a new spectrograph or in improving the image stability of existing Cassegrain instruments. The open loop approach can drastically improve the stability of the spectrum on the detector even with its limitations. We believe that a closed loop system can improve this even further, possibly opening the way for new astronomical research with Cassegrain spectrographs.

Paolo D'Arrigo, Francisco Diego and David Walker,
University College London

RGO Preprints

The following RGO preprints have appeared since the last issue of *Spectrum*

243
RA Laing
Brightness and polarization structure of decelerating relativistic jets
Energy Transport in Extragalactic Radio Sources

244
CA Jackson and JV Wall
Space densities for powerful radio sources in the light of unification
Robert Laing
Large-scale structure: Jets on kiloparsec scales
JV Wall
Space distribution of radio-source populations
Energy Transport in Extragalactic Radio Sources

245
Linda Smith, Max Pettini, David Ling and Richard Hunstead
The chemical evolution of galaxies at high redshift
The Impact of Stellar Physics on Galaxy Evolution

Mon. Not. R. Astron. Soc. 281, 673–678 (1996)

On the stability of Cassegrain spectrographs

D. D. Walker and P. D'Arrigo

Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT

Accepted 1996 February 21. Received 1996 February 16; in original form 1995 December 18

ABSTRACT

Gravity-induced flexure has been a long-standing challenge in Cassegrain spectrographs at 4-m class telescopes; it is more so on the scale of 8-m telescopes. This is of particular concern for the Gemini High Resolution Optical Spectrograph, which will be Cassegrain-mounted for its routine mode of operation. In this paper we address the general flexure problem and make specific recommendations. In a companion paper we present results on experimental compensation for flexure in a specific Cassegrain spectrograph – ISIS on the 4.2-m William Herschel Telescope (WHT).

Key words: instrumentation: spectrographs – telescopes.

1 INTRODUCTION – THE FLEXURE PROBLEM

An important problem that afflicts large astronomical instruments mounted at Cassegrain foci of modern telescopes is structural flexure. The instrument and its supporting structure flex under their own weight while the telescope is tracking a star and the optical alignment is disrupted. The problem is potentially serious because flexure produces spectrum shifts at the detector which degrade the quality of the data. The issue is becoming more critical as telescopes, and spectrographs, grow in size. The gravitational bending of a structural beam of length L is proportional to L^3 (Venkatraman & Patel 1970) and flexure can be expected to follow a similar scaling law. Our work is the development of the High Resolution Optical Spectrograph (HROS) which has emphasized the flexure problem and led to the ideas presented in this paper.

As a telescope tracks, a Cassegrain spectrograph is subject (in its local coordinate frame) to a varying gravity vector. In the case of an alt-azimuth mounting, this results from the combination of altitude motion and Cassegrain turntable rotation. The spectrograph will also be subject to second-order effects owing to the varying accelerations of all three telescope motions, which add small vector components to pure gravity. The result on image motion of these effects will be different, depending on whether the slit is maintained at the parallactic angle (i.e. pointing at the zenith), or at a fixed position angle (PA) on the sky.

The spectrograph structure and components within it must at some level deform as the telescope tracks. Also, some internal components may move bodily and suddenly, as the sign of the gravity vector in the component frame of reference reverses. For example, this may occur as the plane of a mirror surface passes through the zenith and the mirror 'flips' in its mount; or a bearing-mounted mechanism may rock from one side to the other owing to bearing play. Clamping moveable components can stress parts of a mechanism or structure, and the stresses can suddenly and unpredictably be relieved as the local gravity vector changes.

Deformations in the overall instrument structure, or the subunits within it, may be elastic or non-elastic. Non-elastic deformations (i.e. hysteresis) in the structure can result from a variety of causes, from microscopic motions and rubbing within bolted joints, through subtleties of the crystal structure of the structural material, to the build up and release of structures with it.

The combination of all the non-linear processes has the effect that the current image position on the detector depends not only on the current instrument orientation in space but on its recent history as well. 'History' includes the direction, speed of motion and acceleration to which the instrument was subject in reaching its current orientation, and also includes temperature changes.

In general, the above linear and non-linear effects contrive to produce some or all of the following: a secular (smooth) image motion on the detector, modulated by higher frequency 'wobbles' and sudden 'jumps'. The results of image motion include (i) reduction in spectral and spatial resolution, (ii) degradation of line profiles and radial velocities, and (iii) imperfect registration of flat fields. This last effect degrades the effectiveness of flat-field division in

compensating for CCD pixel-to-pixel sensitivity variations, and for fringing in the red.

If the spectrum drifts on the detector during the exposure by 0.1 resolution element (typically 0.2 pixels), then the spectrum is smeared and the spectral resolution is reduced by approximately 10 per cent. Consider a perfect spectrograph with a uniformly illuminated slit. To achieve the same (degraded) resolving power as the flexing spectrograph, the slit could have been opened by 10 per cent, admitting 10 per cent more light! [In the case of a star image (seeing disc) the additional light admitted will be < 10 per cent.] Therefore, flexure can be expressed as a loss in throughput as well. If the spectrograph flexes in the spatial direction, then not only is spatial resolution of extended objects (or crowded fields) impaired, but the contrast against the sky of a compact source is also reduced, degrading the signal-to-noise (S/N) ratio in the sky-subtracted spectrum.

The case of the very large Low Resolution Imaging Spectrometer (LRIS) for the Keck telescope is particularly interesting. It had a very ambitious design goal of < 10 μm image drift for telescope movements of 60° in elevation. After commissioning it showed worst case flexure shifts of 110 to 140 μm over the whole sky (elevation 20° to 90°) (see Oke et al. 1994, 1995). We believe that the problems encountered in designing and constructing LRIS using a traditional passive approach, in particular the constraints of the mass budget and space envelope, can be expected for most of the Cassegrain instrumentation for 8-m telescopes.

Even in spectrographs for smaller telescopes flexure has been a major issue. The echelle spectrograph for the McDonald Observatory, for example, showed more than 1.0 pixel of flexure (27-μm pixel size) over 60°, a value much greater than expected on theoretical grounds. It was only after several months of work, which included adding stiffening supports, that the flexure could be reduced to 0.3–0.4 pixel for the same telescope motion (McCarthy et al. 1993). Other recent instruments have had fewer problems: the CFHT multi-object spectrograph (MOS/SIS) currently has around 10 μm h^{-1} spectrum drifts, and it is anticipated that this should improve to ~ 4 μm h^{-1} once the guiding system for the spectrograph is completed (Le Fevre et al. 1994). Even some smaller instruments, like the LPL (Lunar & Planetary Lab) echelle spectrograph which is 1.22 m long, have shown significant flexure: the LPL spectrograph has up to 25 μm h^{-1} drifts (Hunten et al. 1991).

2 A PHILOSOPHY FOR FUTURE INSTRUMENTS

The challenge of controlling flexure in a spectrograph is remarkably similar to that of a telescope, albeit on a smaller scale. Therefore, it behoves us to learn from, and apply, the methods successfully applied to telescopes. With many telescopes, there is a hierarchical approach, embodying (i) passive compensation in the Serrurier trusses, (ii) application of an open-loop pointing model and (iii) closed-loop compensation of residuals using an autoguider. With large modern telescopes weighing tens of tonnes, this has proved spectacularly successful. We advocate applying the same ideas to Cassegrain spectrographs.

2.1 Passive compensation

The classical Serrurier truss maintains to first order the parallelism of its end rings whilst flexing under gravity. The truss therefore introduces decentration without tilt. The primary and secondary supporting trusses of a conventional telescope are designed to decentre equally with respect to the telescope centre cube, thereby giving passive compensation and maintaining the primary and secondary mirrors in alignment. [Note that Gemini departs from this; the truss is unbalanced and the secondary-mirror centration is actively servoed to compensate (Raybould et al. 1994).]

A conventional parallel Serrurier truss in an instrument could lend itself to the support of optics which are flat (e.g. grating, folding mirrors). In general, decentration in the plane of the optical surface has, to first order, no effect on final image position, since there is no tilt error introduced. In the special case of illumination with parallel light, translation in any direction without rotation has no effect either. We now present an extension of this. If a truss were instead made non-parallel (i.e. converging or diverging), it can be designed so that, when under gravity deflection, it rotates an optical element about an arbitrary point in space. Fig. 1 shows a schematic non-parallel truss. The geometry of the truss is such that it rotates (by small angles) about a virtual centre at a distance R from the mirror, when

$$\sin \alpha = \frac{r}{R}, \tag{1}$$

where α is the convergence of the truss and r the radius of the mirror. For support of a concave mirror (e.g. camera or collimator mirror), this point would be chosen as the centre of curvature of the optical surface. Flexure is then passively compensated to first order as the surface rotates around its centre.

Another relevant factor is the interface between telescope and spectrograph. The custom of mechanically referencing the slit, to the interface by which the top of the spectrograph is attached to the telescope, may not be ideal. The whole weight of the spectrograph hangs off the interface, maximizing the motion of the slit with respect to the spectrograph optics. Since the spectrum basically comprises monochromatic images of the slit, the image motion at the detector will tend to be maximized. We advocate referencing the slit separately to the part of the structure which

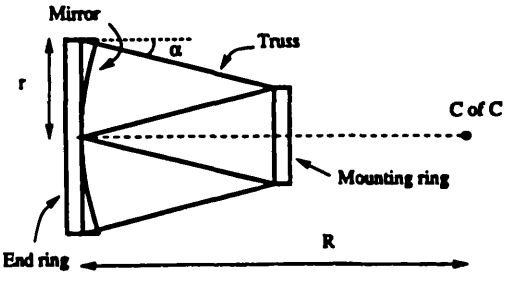


Figure 1. Exaggerated view of a diverging truss supporting a concave mirror.

supports the optics. The slit/optical system may then move bodily with respect to the mounting interface, but this can be taken out by guiding the telescope. Note that there are then implications on the acquisition and guiding system of the telescope, and the telescope pointing model, which have to handle the resulting offsets. However, this should be considered preferable to drifts of the spectrum on the detector, which degrades the science. In the case of Gemini, each instrument will include a wavefront sensor. By referencing this mechanically to the spectrograph slit, the advocated configuration will naturally compensate the telescope offsets (off-slit guiding has the same effect).

An alternative strategy would be to adopt a Cassegrain mounting method which has been used on the Palomar 5-m telescope. In this case, the *centre* of the instrument is mounted in trunnions fixed to the underside of a massive pair of forks attached to the Cassegrain turntable (Fig. 2). The instrument is balanced about the trunnions, and the top of the instrument is free (and accessible), except for an arm which prevents rotation in the trunnions. This concept immediately reduces flexure compared with the same instrument mounted from the top. This is because the scale-length of the structure with respect to flexure is effectively halved, and flexure tends to go as the third power of scale-length. A spectrograph based on this concept might resemble a classical telescope, with the spectrograph 'centre cube' mounted off the trunnions, and the upper and lower Serrurier trusses supporting the slit and collimator respectively. Since the grating is in parallel light, it would naturally be mounted off the upper ring. We therefore achieve passive compensation for the slit–collimator–grating system. This can be extended to include the camera, if its spherical mirror were supported by another Serrurier truss, mounted off the cube, with the truss diverging to give rotation about the centre of curvature of the mirror.

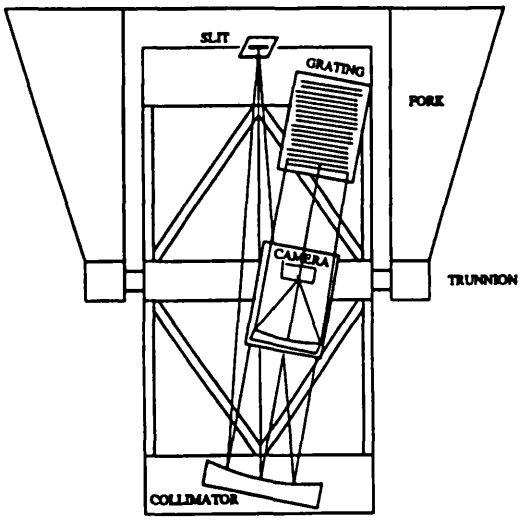


Figure 2. Schematic Serrurier truss spectrograph in fork mounting at Cassegrain.

The above hypothetical example illustrates the types of approach which may be directed towards spectrograph structures. In reality, the entire structure must of course be modelled; the trusses themselves must be supported from some other structure such as the central cube, which itself will be subject to flexure. However, if the truss–cube interface deflects by angle θ, then the angle of convergence or divergence of the truss can in principle be chosen to offset the effect at least to first order.

2.2 The principle of open loop correction

We contend that it is probably inherently impossible to achieve adequate stability of a large Cassegrain spectrograph for an 8-m telescope with purely passive means and within a reasonable mass, space and cost budget; also, to maintain this for different configurations, and through maintenance cycles over the instrument lifetime. In an active system, an adjustable element within the spectrograph corrects the flexure shifts by deflecting the optical path of the light from the telescope during the exposure. Active control is a novel approach to flexure, which has recently been envisaged for the Gemini Multi Object Spectrographs ('GMOS') (Allington-Smith & Davidge 1994), following our study for Gemini (Walker, Dryburgh & Bigelow 1992) and our work on HROS. Some kind of active compensation involving correction at the grating has also been proposed (Bhatia 1994) as a way of improving performance and reducing the cost of astronomical spectrographs for the new generation of telescopes.

Following the analogy with a telescope, the next level of compensation after passive is that of the open-loop pointing model. That is, the telescope is moved to cover the sky in steps (only elevation and Cassegrain rotation for an alt-azimuth). An image of an arc spectrum using a very narrow Dekker (or circular aperture) is recorded on the detector for each step. The resulting spectra are cross-correlated (or centroided) to give the two orthogonal components of image motion, for each of the steps. The results are embodied in a look-up table, and subsequently interpolated to provide flexure predictions for any orientation of the spectrograph. An element of the spectrograph is servoed during the exposure to compensate for the image motion.

The active motion could be provided, for example, by (i) tilting the grating, collimator, a folding mirror, or a wobbler (plane-parallel) plate; (ii) *translating* the slit, image slicer, a fibre feed, a lens, or the detector itself. All of these, with the exception of translating the slit or image slicer, give the capability to compensate for flexure in two orthogonal directions. The baseline for Gemini HROS is to tilt the collimator mirror; for GMOS, which has transmission optics, to translate the CCD (Gemini Multi-Object Spectrograph 1995). The collimator has proved to be an effective and simple way of compensating spectrograph flexure in our experiment on ISIS at the William Herschel Telescope (WHT: see D'Arrigo, Diego & Walker 1996, hereafter Paper II). However, a necessary check when the correction is applied at the collimator is that the correction has a minimal effect on the spectrum format geometry.

The tilting of the collimator, to compensate for flexure, changes the angle of incidence on the grating, albeit by a very small amount. It is necessary, therefore, to check that

the correction at the edges of the CCD is practically the same as the one in the centre, so that the flexure compensation works for the whole spectrum. This can be done by calculating how the small change in the angle diffracted by the grating $\Delta\vartheta_r$ due to flexure correction, depends on the angle of diffraction itself, ϑ_r. The correction in the diffracted angle is given by twice the collimator tilt φ, times the anamorphic factor of the grating, that is

$$\Delta\vartheta_r = 2\varphi \frac{\cos\vartheta_i}{\cos\vartheta_r}. \tag{2}$$

The effect on the spectrum format is obtained by differentiating the previous equation with respect to ϑ_r.

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta_r}(\Delta\vartheta_r) = 2\varphi\cos\vartheta_i \frac{\mathrm{d}}{\mathrm{d}\vartheta_r}\left(\frac{1}{\cos\vartheta_r}\right). \tag{3}$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta_r}(\Delta\vartheta_r) = \Delta\vartheta_r \tan\vartheta_r, \tag{4}$$

or

$$\mathrm{d}(\Delta\vartheta_r) = \mathrm{d}\vartheta_r \Delta\vartheta_r \tan\vartheta_r. \tag{5}$$

The flexure correction can be expressed in terms of the spectrum motion on the detector Δx as

$$\Delta\vartheta_r = \frac{\Delta x}{F_{cam}}, \tag{6}$$

where F_{cam} is the camera focal length. If we want to calculate the effect of the flexure correction at the edge of the detector, then the angle $\mathrm{d}\vartheta_r$ is half the angular dimension of the CCD. If we indicate by $2L$ the CCD length: then,

$$\mathrm{d}\vartheta_r = L/F_{cam}. \tag{7}$$

Combining the last three equations and indicating by $\Delta x_{edge} = F_{cam}\,\mathrm{d}(\Delta\vartheta_r)$ the motion of the spectrum at the edge of the CCD relative to spectrum at the centre, we obtain

$$\Delta x_{edge} = \frac{L\Delta x}{F_{cam}} \tan\vartheta_r. \tag{8}$$

For the ISIS Tex detector $L = 500$ pixel (12 mm), $F_{cam} = 500$ mm and almost always $\tan\vartheta_r < 1$. Therefore, even for large flexure corrections of $\Delta x = 4$ pixel, the 'stretching' of the spectrum at the edge of the detector is very small:

$$\Delta x_{edge} < 0.1 \text{ pixel}. \tag{9}$$

For HROS we have $L = 30$ mm and $\vartheta_r = 57°.4$. For the $R = 50\,000$ mode with the short camera ($F_{cam} = 450$ mm) $\Delta x = 4$ pixel produces $\Delta x_{edge} \sim 0.4$ pixel. Hence we can still achieve a factor of 10 improvement in flexure performance using an active collimator. With the long camera ($F_{cam} = 1080$ mm) the corresponding motion is $\Delta x_{edge} \sim 0.17$ pixel, giving a factor of 24 improvement. Since the long focal-length camera effectively magnifies the spectrum and its motion by 2.4 times with respect to the short one, the effective correction in both cases is the same. In principle, image motion at the detector centre, and the first order of differential motion, can both be compensated using two active elements (e.g. collimator and grating). In practice this level of refinement is unlikely to be required and would unduly complicate analysis of the feedback data.

2.3 The impact of hysteresis and other considerations

The key point concerning open-loop correction is that *hysteresis*, not the raw flexure, is the factor that limits the precision with which image motion can be compensated. *Low hysteresis must be considered as the primary goal* for future open-loop active spectrographs. This will impact on spectrograph design at all levels, from the telescope interface, through the structural design, to the individual mechanisms and mounts. Some very general conclusions from experience may be drawn as follows.

(i) If structural units must be bolted, then the bolting interfaces should be carefully relieved to minimize the build up of stresses and microscopic 'stick–slip'.

(ii) Welded fabrications are likely to be superior to bolted structures.

(iii) Structures with continuous, uniform internal crystalline structure with low internal stresses are likely to give lowest hysteresis. Alloys should be selected for low hysteresis, and heat treated (annealed) to minimize internal stresses. It may be worth using forged alloys, as developed for the substrates of large and stable telescope metal mirrors (Ruch 1992; Morette, Dupont & Rozelot 1992).

(iv) Mechanisms utilizing flexures rather than rolling-element bearings are almost always superior, assuming that sufficient motion can be obtained from a flexure. Optics can be mounted in cells using cemented flexures rather than mechanically clamped ones.

(v) For motions requiring large angular or linear range (particularly grating turrets or slides) we propose that the cells for the optics be compliantly mounted on a simple interchange mechanism to which no particular attention is given to minimize flexure. The selected subassembly is then pushed forwards by a separate mechanism against a reference embodied within the spectrograph structure. This desensitizes the turret or slide, which can be lightweight, and which no longer requires high-precision bearings carrying large masses.

This was the solution proposed by UCL for the Keck LRIS grating turrent. The selected grating cell would be pushed against a single grating-angle mechanism on the structure. This mechanism and associated clamps could be built much more robustly (and expensively!) than the five separate adjusters required on a rotating turret. This philosophy was not adopted by the project because of the critical space constraints and additional complexity of interfaces between turret and structure. The turret as built was a 'classical' rotating structure some 1 m in diameter, and indeed contributed flexure to the spectrograph.

What can be learnt is the need to consider spectrograph stability as a 'systems' problem, not a collection of functional modules bolted together. If hysteresis can be reduced to low levels, then other sources of instabilities can be compensated actively and by open-loop from a pointing model or look-up table. However, there are complications. The

look-up table will in general be different for different instrument configurations, disturbed by maintenance and thermal drifts etc., and may require frequent updating. This has led us to the third level of flexure compensation described below.

3 THE FUTURE – CLOSED LOOP FLEXURE COMPENSATION

The third-level analogy with telescope control is the 'autoguider'. In a spectrograph, a reference light source would propagate through the optical system, be spatially sensed, and the error signal would be fed back to correct the active element. Again, calling upon telescopes, the most precise correction is likely to be achieved by using the best available flexure model to remove the bulk of flexure, and closed-loop to provide second-order correction of residuals due to non-linear and unrepeatable effects.

A simple and highly effective closed-loop experiment was performed by Lowne & Bingham (private communication) on the coude spectrograph of the 2.5-m Isaac Newton Telescope when it was located at Herstmonceux. This spectrograph suffered large thermal drifts. Lowne & Bingham propagated a He–Ne laser through the spectrograph optics, sensed the laser spot by a quadrant photodiode, and used the offset signals to control a transmission wobbler plate behind the slit.

3.1 The light path of the reference beam

We recommend that the reference should be propagated through the *entire* spectrograph optical system, in order to compensate for motions of all surfaces. The light can be (i) monochromatic (fixed wavelength, e.g. later, or programmable using a continuum lamp and monochromator); (ii) emission-line source; (iii) continuum; or (iv) continuum with absorption lines superimposed from, e.g., a Fabry–Perot etalon.

The reference beam can be propagated through the spectrograph as a narrow pencil of light (e.g. laser). Alternatively it can be a diverging cone to fill or partially fill the optics. In principle both methods should work. A laser beam through an optical system with a central obstruction would need to be propagated off-axis. Ghosting might be a problem since the optics are illuminated at different local angles from the science beam. More importantly, for an off-axis laser beam, a focus shift is indistinguishable from an image shift. Also, the spot to be sensed with be ~mm rather than ~μm in diameter, which impacts on the size of detector required. Overall, the filled beam more closely emulates the science beam and therefore in principle is considered preferable.

The reference can be propagated forwards, with injection at the slit and detection in the camera focal plane, or backwards, with the reverse. The latter has an advantage in that the usual case of $F_{cam} < F_{ccd}$ conveniently magnifies the motion to be sensed. A reference might be injected into the optical system as follows.

(i) In the slit plane (e.g. via a fibre), and sensed by a small CCD or quadrant photodiode adjacent to the science CCD.

(ii) At one end of the slit, and sensed by the science CCD. In this case the reference may be projected on to the ~10 rows adjacent to the readout register. The procedure, which is repeated during the exposure, is as follows:

close shutter;
clock out ~10 rows encompassing the reference;
reverse vertical clock by 10 rows to restore science image to original pixels;
centroid or cross-correlate reference and update active element;
open shutter and continue integration.

This method is analogous to the photographic spectroscopic procedure of exposing an arc either side of the science spectrum, the difference being that we propose to detect the reference repeatedly *during* the exposure. This method has been proposed as a closed-loop extension of the open-loop experiment conducted on the ISIS spectrograph, as described in Paper II.

(iii) At the science CCD. In this case, the light source could be fibre fed, with the fibre output being in the camera focal plane, but pointing towards the collimator and suitably baffled. The reference would propagate through the spectrograph backwards, being sensed on a small CCD or quadrant photodiode in the vicinity of the slit. In the case where it is impractical to shutter the CCD separately from the fibre source, care would need to be taken to analyse the ghosting of the optics for the *reverse* path in order to avoid contamination of the science image.

3.2 Reference light source

The spectrograph dispersing element(s) will deflect the reference beam depending on wavelength. If the reference were monochromatic, and the spectrograph configured for a different science wavelength (e.g. by rotating the grating), then the reference light may not reach a separate reference detector. An interesting – and extreme – possibility would be to choose an infrared wavelength for a visible spectrograph, to avoid stray-light contamination of the science data. In general, the problem of grating deviation could be overcome by mounting a separate rotatable mirror on the grating cell, from which the reference would be reflected. However, image-motion owing, e.g., to grating mounting defects would not be compensated, and the mirror rotation might introduce its own problems. For very low-resolution spectrographs, it might be possible to move the reference detector laterally, but thereby its fixed geometry with respect to the slit or science detector will be lost. These factors render the simple laser reference unattractive in most cases. For a prism cross-dispersed instrument such as Gemini HROS, the configuration for different wavelengths requires scanning the prism angles. The resulting change in prismatic deviation of the reference is even more difficult to accommodate.

The recommended solution is to use an optical light projector, fed from a continuum source or arc lamps, and a tuneable monochromator. The monochromator could then be turned according to the spectrograph setting to ensure that the reference beam always hits the detector. If the light from the monochromator were fibre fed to the slit plane of the spectrograph, the reference could fill the optics and still

be able to reach the reference detector for every grating (and prism) setup.

One or more arc lamps can be chosen to provide emission lines on the reference detector whatever the spectrograph configuration. Alternatively, a continuum lamp and etalon provides convenient equally spaced absorption features, but at the expense of filling the spectrograph with unnecessary (and stray-light inducing) continuum light. Hence the 'arc' is in general the preferred source.

A special case is the method described in the previous section of using the science CCD with reverse-clocking as the reference detector. The arc would be fibre fed at one end of the slit, giving a beam matching that of the telescope. With a single-dispersed spectrograph, the reference spectrum could be 2 or 3 pixels wide, contain many lines, and span the full width of the CCD. Each line would be very faint, with the precision of determining image motion being achieved by cross-correlation. However, with the cross-dispersed instrument such as HROS this will not work, since arc spectra would bracket *every order*, reducing the (already pressed) inter-order spacing available for the science data. In this case two possible solutions are envisaged: (i) use of a white light source and tuneable monochromator, producing one bright spot at the top of the science CCD near the amplifier; or (ii) use of an arc and monochromator, producing a limited range of fainter lines spanning in wavelength no more than one free spectral range of the echelle. The latter is currently the favoured option since instabilities in the monochromator are of less consequence.

4 CONCLUSION

Image motion in large Cassegrain spectrographs is a long-standing problem which has become acute at the scale of 8-m telescopes. A 10 per cent degradation in resolution is significant in terms of data science content, yet this translates to a mere ~ 3 μm of image motion during an exposure which may be one hour or more long. We have emphasized the difficulty in achieving even this modest scientific goal for a large instrument, due to flexure and non-elastic effects. The analogy with telescope engineering is of profound significance and leads us to propose a new hierarchical approach of passive, open-loop active and closed-loop compensation in future spectrographs. We recommend this should be adopted for all major new spectrographs. In Paper II, we demonstrate how open-loop active compensation can be successfully applied also to existing spectrographs.

ACKNOWLEDGMENTS

We thank the Gemini IPO for the study contract under which many of the ideas were developed, and PPARC grant GR/H 34708.

REFERENCES

Allington-Smith J., Davidge T., 1994, in Crawford D. L., Craine E. R., eds, Proc. SPIE 2198, Instruments in Astronomy VIII, p 201
Bhatia R., 1994, in Crawford D. L., Craine E. R., eds, Proc. SPIE 2198, Instrumentation in Astronomy VIII, p. 554
Gemini Multi-Object Spectrograph, Conceptual Design Review documents for the Gemini project office, 1995, Durham
D'Arrigo P., Diego F., Walker D. D., 1996, MNRAS, 281, 679 (Paper II, this issue)
Hunten D. M., Wells W. K., Brown R. A., Schneider N. M., Hilliard R. L., 1991, PASP, 103, 1187
Le Fevre O., Crampton D., Felenbok P., Monnet G., 1994, A&A, 282, 325
McCarthy J. K., Sandiford B. A., Boyd D., Booth J., 1993, PASP, 105, 881
Morette M., Dupont M., Rozelot J. P., 1992, in Bingham R. G., Walker D. D., eds, Proc. SPIE 1931, Metal Mirrors, p. 144
Oke J. B. et al., 1994, in Crawford D. L., Craine E. R., eds, Proc. SPIE 2198, Instrumentation in Astronomy VIII, p. 178
Oke J. B. et al., 1995, PASP, 107, 375
Raybould K., Gillett P., Hatton P., Pentland G., Sheehan M., Warner M., 1994, in Stepp L. M., ed., Proc. SPIE 2199, Advanced Technology Optical Telescopes V, p. 376
Ruch E., 1992, in Bingham R. G., Walker D. D., eds, Proc. Spie 1931, Metal Mirrors, p. 112
Walker D. D., Dryburgh M., Bigelow B. C., 1992, Control of Flexure in Gemini Instrumentation, Design study report commissioned by the Gemini Tucson office
Venkatraman B., Patel S. A., 1970, Structural Mechanics, McGraw-Hill, p. 176

Mon. Not. R. Astron. Soc. 281, 679–686 (1996)

Active compensation of flexure on the WHT ISIS spectrograph

P. D'Arrigo, F. Diego and D. D. Walker

Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT

Accepted 1996 February 21. Received 1996 February 16; in original form 1995 December 18

ABSTRACT

ISIS Spectrograph Automatic Active Collimator (ISAAC) was developed at the Optical Science Lab, University College London, as an experimental system to improve the image stability on the detectors of the Intermediate-dispersion Spectroscopic and Imaging System (ISIS) at the William Herschel Telescope (WHT). It is based on the concept of active compensation, where spectrum drifts, owing to the spectrograph flexing under the effect of gravity, are compensated by the movement of an active optical element. ISAAC is a fine-steering tip-tilt collimator mirror; it was tested on the spectrograph in 1995 July. Here we report the details of the design and the laboratory tests together with the results of the experiment on ISIS at the WHT. The experiment shows that active compensation can reduce flexure down to less than 3 μm over 4 h of telescope motions, dramatically improving the spectrograph performance. A way of implementing ISAAC as a common user facility is briefly discussed, together with a study of the characteristics of flexure in ISIS.

Key words: instrumentation: spectrographs – telescopes.

1 INTRODUCTION

As described in the companion to this paper (Walker & D'Arrigo 1996, hereafter paper I), flexure in spectrographs is an important issue, and one that particularly concerns the designers of instruments for very large telescopes. The approach we adopted was to employ active optics as a way of solving this problem. With active compensation of flexure the image shifts on the detector are repeatedly compensated through motions of an active optical element within the instrument.

An active flexure compensation system is currently being investigated for the Gemini Multi-Object Spectrograph (GMOS) (Allington-Smith & Davidge 1994), following a commissioned study of the problem performed under contract to Gemini in this laboratory (Walker, Dryburgh & Bigelow 1992). In this case, since GMOS has transmission optics, the detector motion was chosen as the active element. Recently this concept has also been proposed by Bhatia (1994) as a way of easing the pressure on the mechanical design, improving the instrument performance and even reducing optical aberrations. Bhatia, however, proposes to apply higher order corrections and at much higher rate, to improve the image *quality* and not just the *stability*. This implies that the active corrections need to be applied with a thin deformable grating. The technical challenge in producing such a system seems to be the major obstacle in this approach. Before the experiment described in this paper, active compensation of flexure had never been attempted on a Cassegrain spectrograph.

Active flexure compensation can be achieved either with: (i) an open-loop system, where the drifts are accurately measured and stored in a look-up table from which the active element is controlled; or (ii) with a closed-loop system, where an optical reference is sensed directly and fed back to the control of the active element. Having decided that active flexure compensation was the right way forward, we saw an opportunity of developing an active system for the Intermediate-dispersion Spectroscopic and Imaging System (ISIS) spectrograph at the 4.2-m William Herschel Telescope (WHT). Whilst the basic motivation was a proof of concept for the Gemini High Resolution Optical Spectrograph (HROS), it was clear that ISIS exhibited flexure at a level where it could benefit from the active collimator approach we had proposed.

2 ISIS AND THE ISAAC PROJECT

ISIS is a double-beam spectrograph designed and built by the Royal Greenwich Observatory and has been in opera-

tion at the Cassegrain focus of the WHT for several years. Systematic measurements of the position of the spectrum as a function of zenith distance and rotation angles of the Cassegrain turntable have shown evidence of flexure in the ISIS structure and probably in its supporting flange system (Rutten, private communication). ISIS is one of the largest Cassegrain instruments currently in operation on a 4-m class telescope, and therefore one where imaging stability is most critical. It is also a very popular instrument in the astronomical community for its versatility and overall performance. These reasons make it an almost ideal choice for the testing of our novel system of image stabilization.

The ISIS Spectrograph Automatic Active Collimator (ISAAC) project was originally conceived as an open-loop system for active compensation of flexure. The choice of the active element fell on the collimator for two main reasons: it was easy to control with rear-mounted actuators to provide tip/tilt active corrections, and given its position in the spectrograph, it could be accessed and replaced easily. Moreover, the collimator is a natural choice for flexure compensation, being in place for every spectrograph setup and producing significant image shifts on the detector with relatively small tilts. The last is important in avoiding optical aberrations which could be introduced if the correction significantly disturbed the optical geometry. The blue arm was chosen because, when we started this project, it was experiencing larger flexure.

Flexure compensation with an active collimator is based on a simple concept: when the spectrum position on the detector drifts during tracking, the collimator is slightly tilted, thus deflecting the reflected beam and restoring the image to the original position. The active collimator replaces the existing collimator in the spectrograph and it is operated remotely with a computer. Fig. 1 shows a schematic drawing of ISAAC illustrating the major components and communication links in the system. The collimator is finely steered with three piezoelectric actuators pushing against its back. The amount of tilt required is calculated from the spectrograph geometry and a previous set of flexure measurements. The orientation in space of the spectrograph is calculated from the equatorial coordinates of the observed object, the sidereal time at the start of the exposure, the elapsed time and the position angle of the slit on the sky.

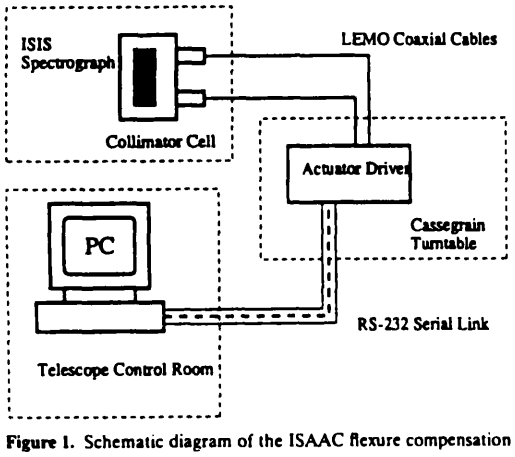


Figure 1. Schematic diagram of the ISAAC flexure compensation system.

3 MECHANICAL DESIGN AND CONSTRUCTION

The mechanical design of the active collimator was driven by two major objectives: to make it compatible and easy to interchange with the existing collimator for the ISIS blue arm, and to minimize hysteresis. The latter is required to avoid introducing new sources of image shifts during the correction process. Hysteresis in the collimator movements would affect the accuracy with which the spectrum position on the detector can be controlled and could make the active compensation ineffective. The resulting design is shown in Fig. 2 is an exploded view. Three piezoelectric actuators are fitted with strain-gauge position sensors. This combination of actuator and sensor allows a very fine control of the collimator displacement, with a nominal accuracy better than 75 nm over temperature excursions up to 20° (Physik Instrumente GmbH, private communication). The actuators have a total expansion of 15 μm and a lifetime of more than 10^9 cycles. The rear actuators in combination with three neoprene pads on the mirror front edge also provide

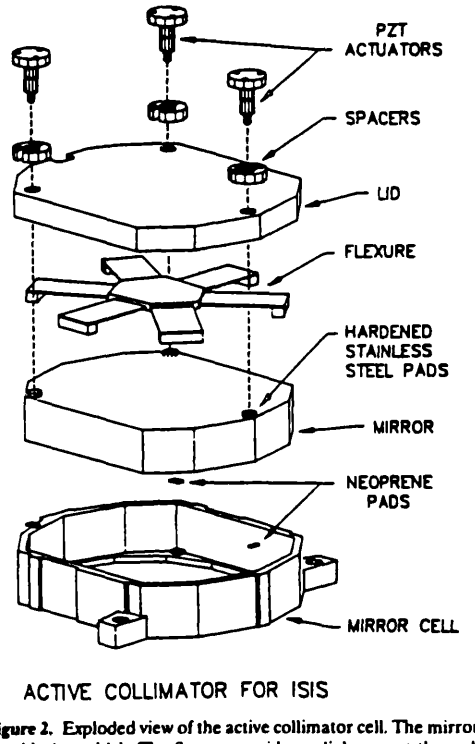


Figure 2. Exploded view of the active collimator cell. The mirror is roughly 4 cm thick. The flexure provides radial support through a small Invar cylinder inserted in the back of the mirror (not shown).

detector as small as 0.4 ± 0.1 μm (corresponding to a tilt of only 0.4 μrad).

The wavefront from the collimator mirror was measured using a scatter-plate interferometer with Wyko Corp. Phase II fringe analysis software. Measurements were made with the mirror both inside and outside the cell. Peak-to-valley error was 0.067λ (He–Ne 632.8 nm) outside (unstressed) and 0.077λ inside, showing that no significant stress was introduced. Even when one actuator was maximally extended, the peak-to-valley error was only 0.097λ. The actuator performance was tested in variable environmental condition. By operating the actuators in a cold room with temperatures as low as -10°C, we were able to prove that no measurable degradation of performance occurs. The actuators were also working correctly with relative humidity levels as high as 70–80 per cent.

6 THE EXPERIMENT ON ISIS

Only 1995 July we obtained access to ISIS for day-time tests at the WHT. The aim was to study flexure in ISIS, monitor it effectively and then prove that we could compensate it with the active collimator. The ISAAC collimator was installed on the ISIS blue arm and its focus and tilt were adjusted using the standard Hartmann shutter. The optical quality was checked by imaging a narrow slit illuminated with Cu–Ne and Cu–Ar calibration lamps. The results were at least as good as those obtained with the original collimator. Most of our subsequent tests were performed with a 150-μm slit and with the low-dispersion (158 line mm^{-1}) grating centred around $\lambda = 7000$ Å. This value was chosen to provide a good S/N ratio in the calibration lamp spectra. All the data were recorded using the same 1024×1024 Tektronix CCD with 24×24 μm^2 pixels (La Palma Tek1 CCD).

The fine calibration of the mirror movements was performed with the collimator optically aligned in the spectrograph. The three actuators were moved in turn and in different combinations, and the corresponding detector shifts recorded. From this the parameters relating the actuator voltages to detector shifts were calculated. The reference spectra were generated by observing an arc spectrum through a short dekker, so that each spectral line would produce an almost circular spot. The detector shifts (in spectral and spatial direction) were calculated by centroiding a single line in the spectrum. We checked this procedure with other lines in the spectrum that gave the same result to within 0.05 pixel, confirming that this system was as effective for this application as cross-correlating the whole spectrum (and much quicker).

To achieve effective open-loop compensation, flexure needs to be measured and modelled very accurately. For the purpose of flexure measurements the spectrograph orientation in space is defined by only two parameters: the zenith distance z of the telescope (or the elevation $= 90° - z$) and the position angle of the Cassegrain turntable on which ISIS is mounted. The telescope azimuth rotation does not change the position with respect to the gravity vector and is irrelevant. Flexure is measured by recording the shifts on the CCD detector of a point-source for the different orientations in space. A 'point'-source is created by observing a spectral line through the short dekker. Flexure shifts are measured by centroiding spectra taken in different instrument positions.

Rutten (private communication) has generated an automatic procedure for producing a flexure map that covers the whole range of telescope and spectrograph possible positions. A program moves the telescope in elevation in 15° steps from zenith to $z = 75°$, then moves the telescope back to zenith, rotates the turntable 45° and repeats the elevation stepping. This is repeated for 360° of turntable motion. A spectrum is acquired at every elevation/rotation setting. The automation not only speeds up the process considerably but also assures that the data is taken always in the same way (with the same telescope motion and therefore same hysteresis). A typical flexure map thus produced is shown in Fig. 4.

We found that the raw data coming from this measurement were not a reliable model for our active corrections because of the large hysteresis present. Fig. 5 shows three sets of data taken on three different days. Flexure shows a reproducible pattern, although translated in the CCD plane.

Moreover, since every 'arm' of the map has a starting point at zenith (with a different turntable angle), those points would be expected to coincide, given that when the telescope is pointing at the zenith a rotation of the Cassegrain turntable does not change the orientation of the spectrograph with respect to the gravity vector. The fact that they do not coincide can be attributed to hysteresis during the flexure measurement procedure, since previous measurements (Rutten, private communication) have shown tht simply rotating the turntable with the telescope pointing at zenith produces negligible flexure shifts.

For these reasons, we modified the raw flexure map by translating every arm so that the zenith points coincide, and used this as our flexure model. The result is shown in Fig. 6. With this simple artifice the flexure map becomes easier to

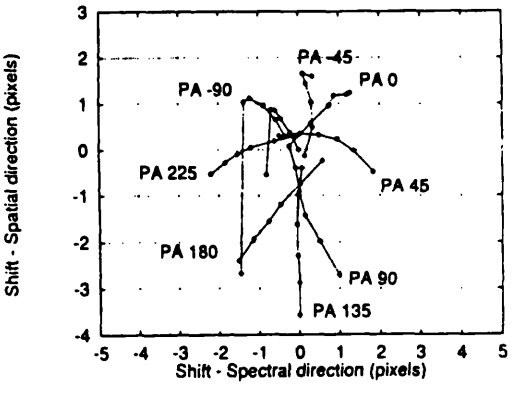


Figure 4. Map of the flexure shifts on the blue detector of ISIS. Each 'arm' represents a different turntable position angle (PA), while the points in one arm correspond to different telescope elevation (from zenith to 15° above horizon in 15° steps). Note that the points corresponding to the telescope pointing at zenith (and different PA) do not coincide. This is attributed to hysteresis in the flexure measurement procedure.

the axial support for the mirror. These pads are located exactly opposite the actuators to avoid introducing bending moments, and act as springs to produce the necessary return force for the small tilting of the mirror. The overall movement is very small: the maximum tilting angle is about 17 arcsec (10^{-4} rad) when the actuators are at full travel. This corresponds to image motions of 80 μm or 3.4 pixel at the camera focal plane (24-μm pixels). The radial support is provided by a flexural element connected to the external cell and bolted to a small Invar block inserted in a blind hole in the back of the Zerodur mirror. This system was chosen to minimize thermal stresses between the mirror and the mirror cell.

The mirror, an off-axis paraboloid, is an exact replica of the existing ISIS blue collimator. The surface was polished and figured by D. Brooks at the Optical Science Laboratory (OSL) to better than one-tenth of a wavelength (at 628 nm) peak-to-valley, and its focus was measured to be within 2 mm of the target value. The elastic characteristic of neoprene pads was also measured, in order to estimate the optimal level of compression that would hold the mirror firmly in place without introducing distortions on its surface. Finite element analysis showed that by applying a force of 100 N with each actuator, the mirror surface distortion was smaller than the polishing accuracy (60 nm). This force is more than sufficient to keep the mirror, which weighs around 50 N, in its axial position within the cell for all orientations of the spectrograph while in service.

4 CONTROL SYSTEM

The instrument hardware consists of three separate subsystems. The mirror cell and actuator system are connected through coaxial cables to the actuator driver which sits in the instrumentation rack on the telescope. This in turn is connected via a serial link to a standard lap-top 486 PC in the control room, which runs the main software (Fig. 1). During our experiment the user interface was the lap-top PC. This was completely separate from the ISIS control system, as demanded by the La Palma Observatory as a condition for conducting the experiment. The actuator driver contains the power supplies and amplifiers for the actuators and the sensors. It also contains a local microprocessor which communicates with the PC in the control room and sends signals to the 12-bit digital-to-analog converter which controls the drivers for the piezoelectric actuators.

In its main mode of operation, the software takes the coordinates of the observed object, the required slit position angle on the sky and the sidereal time at the start of the exposure, as keyboard input. It calculates the spectrograph orientation in space during tracking (in terms of telescope zenith distance and Cassegrain rotator position angle). Alternatively, the orientation in space can be given directly. The spectrograph position is used to predict the flexure shifts from flexure data in a look-up table. The flexure data are stored in two matrices (spectral and spatial direction shifts) with rows and columns representing telescope elevation (or zenith distance z) and Cassegrain turntable position angle (PA). Given the spectrograph position, the software interpolates the look-up table to calculate the required correction at the collimator. This correction is translated into actuator voltages and sent (in digital form) to the local processor.

5 LABORATORY TESTS

The laboratory tests verified that the movement accuracy and repeatability for the collimator conformed to specification. Our goal was to correct detector shifts within $\Delta x = 0.1$ pixel, which for the 24-μm Tek CCD in ISIS, means 2.4 μm at the camera focal plane. The corresponding collimator tilt (neglecting anamorphism) is given by

$$\Delta\omega \simeq \frac{\Delta x}{2F_{cam}}. \qquad (1)$$

Given the camera focal length $F_{cam} = 500$ mm, this translates to collimator tilts as small as 0.5 arcsec or 2 μrad.

The mirror was set up in autocollimation against a reference flat, and used to form an image of an illuminated 25-μm pinhole. A re-imaging lens projected a magnified ($\times 10$) image of the pinhole on to a CCD. When the collimator is tilted, the pinhole image moves on the CCD and the shift is measured by centroiding or cross-correlating the two CCD images.

To check accuracy and repeatability we drove the actuators to near maximum expansion and then back to zero, to produce the hysteresis curve of Fig. 3. Note that the figure represents the total hysteresis in the motion of the active collimator, not just the actuator hysteresis. The residual hysteresis shown in Fig. 3 is the major source of inaccuracy in the system. Even so, the effect is very small: if this is measured as a percentage of the total motion, our worst case result was a hysteresis of 2 per cent of the travel. By applying these results to the performance of the instrument on ISIS we can see that in terms of detector shifts this means that over a total image motion of 72 μm (3 pixel), the maximum error is only 1.44 μm on the detector (0.06 pixel), better than our target value. The test also showed that the collimator can be controlled and resolve movements at the

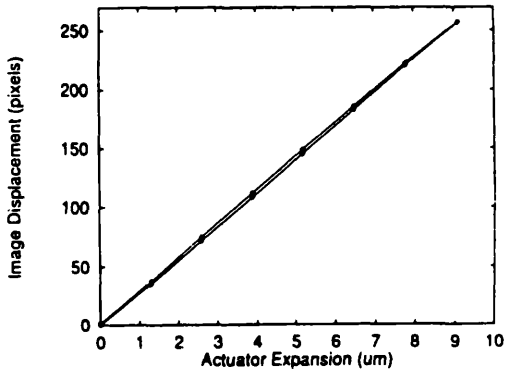


Figure 3. Residual hysteresis in the collimator movements within the cell. A single actuator was expanded in steps and then brought back to zero expansion to produce a hysteresis curve. The graph shows the sum of actuator and mirror support hysteresis. Hysteresis is measured as a percentage of the total motion.

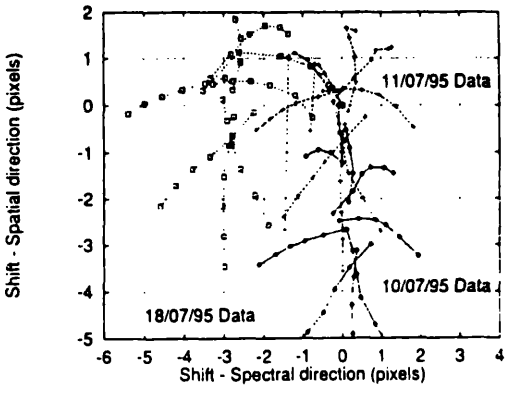


Figure 5. Comparison of flexure maps taken on three different days.

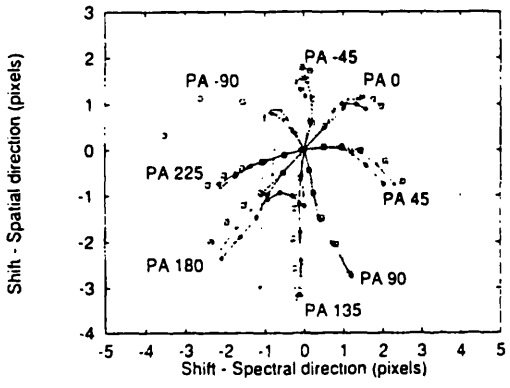


Figure 6. Modified flexure map. Comparison of three modified maps used for modelling flexure in ISIS. This flexure model clearly is more reproducible than the data in Fig. 5. The maps were produced by translating the tracks corresponding to different turntable positions so that the starting points (telescope at zenith) for each track coincide. This reduces the impact of hysteresis on the flexure model. The track for PA = −90° is repeated. The sudden jump in this track for an elevation of 15° is probably due to flipping of a loose component in ISIS.

understand: the *amount* of flexure (relative to zenith) is essentially a function of telescope elevation, while the *direction* of flexure mainly depends on the turntable position. Moreover, the map becomes much more reproducible: the figure shows the three different maps of Fig. 5, after they have been modified in this way.

Even so, hysteresis is intrinsic in the spectrograph structure and mechanisms, and is difficult to eliminate. A similar hysteretic effect has been observed in other Cassegrain spectrographs (Allington-Smith et al. 1994). The presence of hysteresis is proved by the graph of Fig. 7, where the telescope was moved from zenith to horizon and back keeping the rotator position angle (PA) fixed. A classical hysteresis curve was produced.

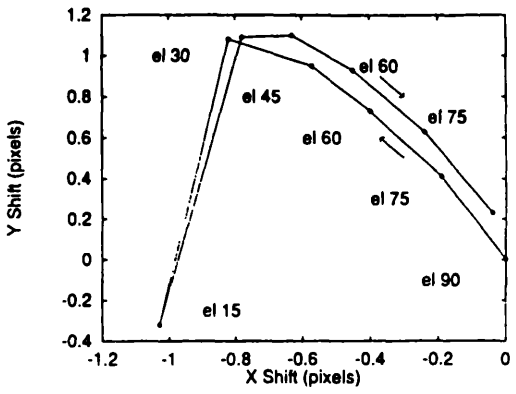


Figure 7. Hysteresis in the flexure shifts in ISIS. The curve was produced by stepping the telescope from zenith to horizon and back while keeping the turntable position fixed. The hysteresis shown is roughly 15 per cent of the total motion.

The flexure model developed increases the reproducibility of the flexure shifts. We need to stress, however, that the flexure map is quite sensitive to changes in the spectrograph settings (grating, central wavelength, presence of dichroic, etc.). During our tests we tried the effect of changing the grating ruling from 158 line mm^{-1} to 600 line mm^{-1} and using red and blue arms at the same time. We found that if we compensated flexure without updating the look-up table, we could still improve on the uncompensated data, but the improvement was reduced by nearly a factor of 2. Moreover, flexure changes with time, so that over one or two months, the data (even if taken with the same settings) can look appreciably different. As a reference, Fig. 8 shows a flexure map taken one month earlier with the previous blue collimator. The graph shows that flexure in ISIS was clearly larger in June than when we took our measurements in July, so if the look-up table were not updated, flexure could be overestimated (or underestimated). However, it is necessary to mention that earlier in the spring of 1995 ISIS was taken off the telescope and dismantled to have some of its optics re-coated. It is possible that this may explain the larger flexure observed in June. The next earlier flexure measurement available was taken one year earlier in 1994 July. In this case the amount of flexure is of the same order of magnitude as the one observed in 1995, but the tracks in the model are somewhat different, so compensation errors can still be expected.

To evaluate the long-term stability of flexure in ISIS, regular measurements would be necessary. The maintenance work on the spectrograph should also be taken into account. With the extremely limited data available, our rough estimate is that the instrument flexure could still be reduced by a factor between 1.3 and 2 with a look-up table of a few months old. This may also be improved if the data for the model were taken in a different way, as explained in Section 8. However, since the procedure for taking a flexure map takes only 2–3 h of telescope daytime, we would recommend that whenever flexure is a sensitive issue for a particular astronomical observation, a new set of data should be recorded.

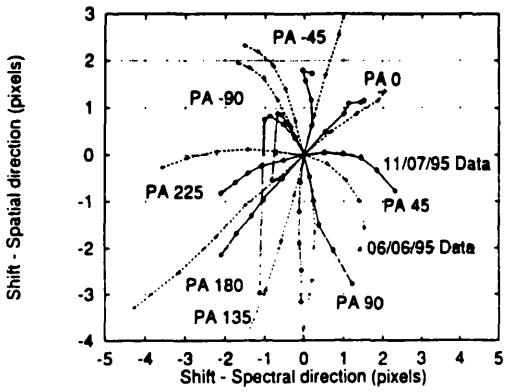


Figure 8. Changes in the flexure model over one month. The data taken during our tests in La Palma is compared with the flexure map obtained from data taken one month earlier by R. Rutten with the old collimator.

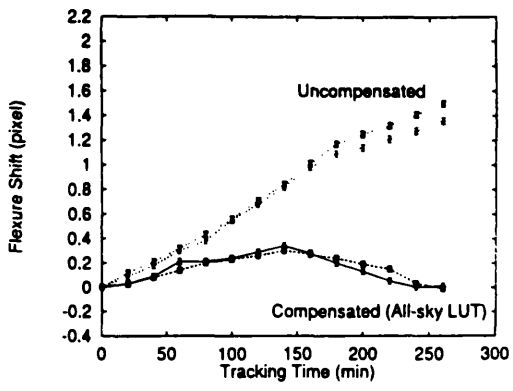


Figure 9. Flexure compensation with all-sky look-up table (spectral direction). The same star track was repeated twice both with and without flexure compensation to check the reproducibility.

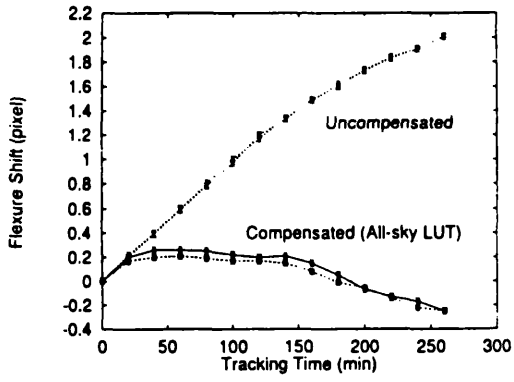


Figure 10. Flexure compensation with all-sky look-up table (spatial direction). The same star track was repeated twice both with and without flexure compensation to check the reproducibility.

7 ACTIVE COMPENSATION OF FLEXURE

The tests for active compensation of flexure were performed during the daytime, using the calibration lamps to feed the spectrograph. In our first experiment we simulated a star track by moving the telescope in 20-min steps following the motion of a 'virtual' star on the sky. In this way we compressed more than 4 h of tracking into less than 1 h of measurements. We compensated for the flexure shifts using an all-sky look-up table obtained with the procedure described earlier, and then we repeated the same track without compensation. We also took the same measurements the following day to check how reproducible these results were. The compensated and uncompensated spectrum shifts are shown in Figs 9 and 10 for the spectral and spatial directions.

Several conclusions can be drawn from these graphs. First, that on average, active compensation of flexure can reduce spectral drifts to 0.2 pixel or less (0.3 pixel worst case) for more than 4 h of tracking. We must stress here that the exposure time during astronomical observations is, anyway, limited to around 1 h by accumulation of cosmic ray events on the CCD. Our experiment of compensating flexure over several hours proves that large amounts of flexure can be corrected with the same accuracy. This may not have much impact on the length of exposures, but it is relevant in that it shows that active compensation can potentially avoid the need for frequent calibration arcs between exposures, saving telescope time. It is also highly relevant to new instruments for 8-m class telescopes.

The image stability we achieved is a significant improvement on the existing situation, with typical drifts of 1.5–2.0 pixel during the same time. The graph also shows that the improvement is more relevant for longer tracking times. This is because the uncompensated flexure usually increases fairly smoothly (at an average rate of 10 μm h^{-1}), whereas the inaccuracy in compensation depends mostly on residual errors and very little on the amount of flexure to correct. Finally, the graphs prove that when the uncompensated flexure is reproducible, the curve representing the residual errors is also highly reproducible indicating that these errors are mostly due to inaccuracies in the flexure model and *not* to collimator movements.

This last point induced us to attempt to refine the look-up table, to see if we could improve image stability even further. For the same track, we calculated the residual errors and fed them back into the active collimator control system as second-order corrections. In practice the procedure involved stepping along a track with active compensation on, recording the residual shifts, converting them into actuator voltage corrections and then running along the same track again applying the corrected voltages. These second-order corrections, for spectral and spatial directions, are shown in Figs 11 and 12, compared with the standard all-sky compensation. The improvement is quite remarkable: once the errors in the flexure model are corrected, the spectrum drifts can be contained within 0.1 pixel for nearly 4 h of equivalent tracking time (more than 4 h in spectral direction).

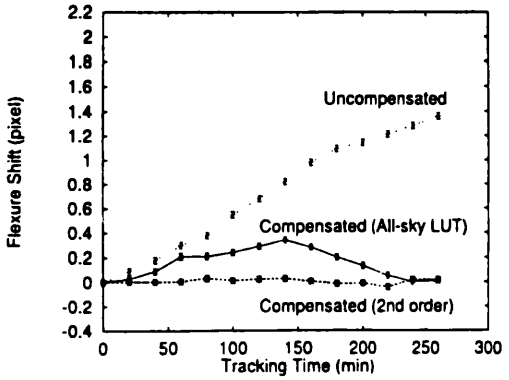


Figure 11. Second-order flexure compensation (spectral direction). The second-order corrections were obtained by feeding back to the control system the residuals from the all-sky compensation run.

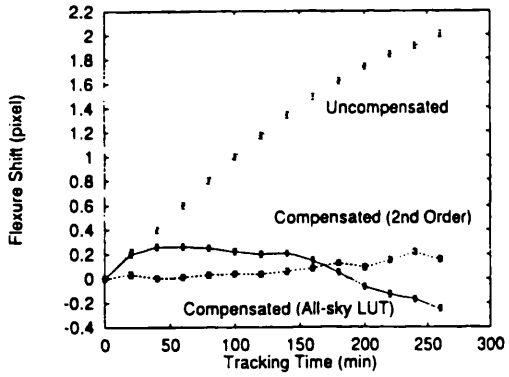


Figure 12. Second-order flexure compensation (spatial direction). The second-order corrections were obtained by feeding back to the control system the residuals from the all-sky compensation run.

8 DISCUSSION AND FUTURE DEVELOPMENTS

This experiment has proved that active compensation is a viable and effective way of reducing flexure in Cassegrain spectrographs. It has also shown that the collimator can be controlled with very high accuracy and it is a suitable element for applying active corrections. More importantly, ISAAC proved capable of reducing flexure in ISIS significantly. During our experiment in 1995 July ISIS was showing an average spectrum drift of over 40 μm in 4 h of tracking time. By applying active compensation we were able to reduce it to an average 4.8 μm over 4 h with an all-sky flexure map and to less than 2.4 μm with a flexure look-up table modelled for that particular observation.

These results have several implications on the astronomical data that the spectrograph can provide, not only in terms of increased spectral resolution and reduced radial velocity errors. The reduction in flexure we achieved is such that the time wasted in taking regular calibration arcs can be cut drastically. Moreover, this low flexure can allow longer exposure times and improved S/N ratio: the limiting factor in the length of exposures would be effectively shifted from the flexure drifts to the cosmic ray events rate.

This experiment has also shown that the reduction in flexure we achieved was limited by the accuracy of the flexure model and not by the accuracy in the movement of the collimator. Our laboratory tests had demonstrated that the collimator could provide movements as small as 0.4 μm on the ISIS CCD. We may therefore anticipate that flexure compensation could be improved even further with a more accurate model. This may include a different procedure for recording flexure data, for example by recording several star tracks at different declinations until the whole sky is covered. Indeed, some other ways of modelling spectrograph flexure, based on a polynomial fit of the data, have been suggested (Munari & Lattanzi 1992), although relying strongly on the absence of hysteresis in the spectrograph.

Another result of our experiment was in fact the confirmation of the large hysteresis in the spectrograph. This is likely to place some severe constraints on the accuracy a flexure model can ultimately achieve. Moreover, to collect the data necessary for an accurate flexure map takes considerable telescope (day) time, which is often needed for engineering and maintenance. ISAAC, using an all-sky flexure map, is therefore a compromise solution, balancing performance in image stability versus time required to achieve that. We are currently discussing with staff at the Royal Greenwich Observatory and at the ING group of telescopes the possibility of implementing our system as a common user instrument for ISIS. The plan is to have an active collimator on the red and blue arms and to integrate the relative software and hardware with the control system for ISIS.

It is clear, however, that to exploit fully the potential of the active collimator, a different approach is needed. By closing the loop with an optical reference beam propagated through the spectrograph, it is possible to avoid the problem of hysteresis altogether. The flexure shifts would be sensed in real time while the spectrograph is tracking and directly fed back into the collimator control. It also would avoid the need for producing several different flexure models for corresponding different spectrograph setups.

For these reasons we have proposed a closed loop experiment for ISIS. The experiment is discussed in more detail in Paper I and involves injecting an optical reference into the spectrograph and using the CCD with reverse clocking to sense the reference motion. We can conclude that our experiment has shown that active compensation of flexure is an attractive option in designing a new spectrograph and in improving the image stability of existing Cassegrain instruments. The open-loop approach can drastically increase the stability of the spectrum on the detector, even with its limitations. We believe that a closed-loop system can improve this even further.

The experiment reported here, as well as being the first step towards a closed-loop system, has also proved the key to the Gemini High Resolution Optical Spectrograph (HROS) (Walker et al. 1992). The HROS project is currently in the conceptual design phase and it is clear that the instrument will need some form of active flexure compensa-

tion. At this moment, the plan is to build an active collimator, based on the experience with ISIS.

ACKNOWLEDGMENTS

We thank all the staff at the La Palma Observatory and in particular R. Rutten for his support and advice to the project. We also thank the Royal Greenwich Observatory staff, including M. Johnson and S. Worswick for providing advice and information concerning ISIS and access to the cold room facility in Cambridge. We thank the staff at the Optical Science Laboratory, particularly D. Brooks who figured the mirror and helped with the optical tests, M. Dryburgh who designed the collimator cell and tested ISAAC with us at the WHT. H. Jamshidi and G. Nixon who designed the electronics. The original idea of active compensation was proposed by A. Charalambous at UCL. The development of the ideas in this paper started under internal OSL support, was progressed under PPARC grant GR/H 34708 and further developed under a study contract from Gemini IGPO. We gratefully acknowledge these sponsors.

REFERENCES

Allington-Smith J., Davidge T., 1994, in Crawford D. L., Craine E. R., eds, Proc. SPIE 2198, Instrumentation in Astronomy VIII, p. 201

Allington-Smith J. et al., 1994, PASP, 106, 983

Bhatia R., 1994, in Crawford D. L., Craine E. R., eds, Proc. SPIE 2198, Instrumentation in Astronomy VIII, p. 554

Munari U., Lattanzi M. G., 1992, PASP, 104, 121

Walker D. D., D'Arrigo P., 1996, MNRAS, 281, 673 (Paper I, this issue)

Walker D. D., Dryburgh M., Bigelow B. C., 1992, in Control of Flexure in Gemini Instrumentation, Design study report, Gemini Tucson office, Tucson, AZ

Active compensation of flexure on the High Resolution Optical Spectrograph for Gemini

P. D'Arrigo, F. Diego, D.D. Walker

University College London, Department of Physics and Astronomy,
Gower Street, London WC1E 6BT

ABSTRACT

Gravity-induced flexure has been a long-standing challenge in Cassegrain spectrographs at 4-metre class telescopes; it is the more so at the scale of 8-metre telescopes. This is of particular concern for the Gemini High Resolution Optical Spectrograph, which will be Cassegrain-mounted for its routine mode of operation. In this paper we address the general flexure problem, and how to solve it with the use of active optics. We also present the results of an experimental active flexure compensation system for the ISIS (Intermediate-dispersion Spectroscopic and Imaging System) spectrograph on the 4.2m William Herschel Telescope (WHT). This instrument, called ISAAC (ISIS Spectrograph Automatic Active Collimator), is based on the concept of active correction, where spectrum drifts, due to the spectrograph flexing under the effect of gravity, are compensated by the movement of an active optical element (in this case a fine steering tip-tilt collimator mirror). The experiment showed that active compensation can reduce flexure down to less than 3μm over four hours of telescope motions, dramatically improving the spectrograph performance. The results of the experiment are used to discuss a flexure compensation system for the High Resolution Optical Spectrograph (HROS) for the 8m Gemini telescope.

Key Words: Spectrograph, Active Optics, Flexure, Cassegrain.

1. INTRODUCTION

As a telescope tracks, a Cassegrain spectrograph is subject to a varying gravity vector. The spectrograph structure and components within it must at some level deform as the telescope tracks. Also, some internal components may move bodily and suddenly, as the sign of the gravity vector in the component frame-of-reference reverses. For example, this may occur as the plane of a mirror surface passes through the zenith and the mirror "flips" in its mount; or a bearing-mounted mechanism may rock from one side to the other due to bearing-play. Clamping moveable components can stress parts of a mechanism or structure, and the stresses can suddenly and unpredictably be relieved as the local gravity vector changes. The results of image motion include i) absolute calibration errors ii) degradation of line profiles and radial velocities, iii) reduction in spectral and spatial resolution and iv) imperfect registration of flat fields. Flexure has therefore a significant impact both on the spectrograph performance and on the time wasted in frequent calibration arcs. The issue is becoming more critical as telescopes, and spectrographs, grow in size. The gravitational bending[1] of a structural beam of length L is proportional to L^3 and flexure can be expected to follow a similar scaling law. In its current design, the High Resolution Optical Spectrograph (HROS) for the Gemini telescope, is going to be one of the largest Cassegrain spectrographs in the world and it is expected to experience significant flexure shifts.

2. THE CONCEPT OF ACTIVE COMPENSATION OF FLEXURE

The challenge of controlling flexure in a spectrograph is remarkably similar to that of a telescope, albeit on a smaller scale. With many telescopes, there is a hierarchical approach, embodying i) passive compensation in the Serrurier trusses, ii) application of an open-loop pointing model and iii) closed-loop compensation of residuals using an auto-guider. With large modern telescopes weighing tens of tonnes, this has proved spectacularly successful. We advocate applying the same ideas to Cassegrain spectrographs. We contend that it is probably inherently impossible to achieve adequate stability of a large Cassegrain spectrograph for an 8-metre telescope with purely passive means and within a reasonable mass, space and cost budget. Also to maintain this for different configurations, and through maintenance cycles over the instrument lifetime. An active optical system offers a possible way forward. In an active system, an adjustable element within the spectrograph corrects the flexure shifts in real-time by deflecting the optical path of the light from the telescope during the exposure.

An active flexure compensation system is currently being investigated for the Gemini Multi-Object Spectrograph (GMOS),[2] following a commissioned study of the problem performed under contract to Gemini in this Laboratory.[3] In this case, since GMOS has transmission optics, the detector motion was chosen as the active element. Recently this concept has also been proposed by Bhatia[4] as a way of easing the pressure on the mechanical design, improving the instrument performance and even reducing optical aberrations. Bhatia, however, proposes to apply higher order corrections and at much higher rate, to improve the image *quality* and not just the *stability*. This implies the active corrections need to be applied with a thin deformable grating. The technical challenge in producing such a system seems the major obstacle in this approach. Before the experiment described in this paper, active compensation of flexure had never been attempted on a Cassegrain spectrograph.

Following the analogy with a telescope, open loop flexure correction is based on data from a pointing model. That is, the telescope is moved to cover the sky in steps and flexure is measured at each point. The results are embodied in a look-up table, and subsequently interpolated to provide flexure predictions for any orientation of the spectrograph. An element of the spectrograph is servoed during the exposure to compensate for the image motion.

In a closed loop system flexure is measured directly during an exposure, so an optical reference that propagates through the spectrograph needs to be added to the system. The flexure correction is not calculated with the look-up table, but it is measured by the shifts of the optical reference as the exposure progresses. The active component can be the same for both systems. The active motion could be provided, for example, by i) *tilting* the grating, collimator, a folding mirror, or a wobbler (plane-parallel) plate: or ii) *translating* the slit, image-slicer, a fibre-feed, a lens, or the detector itself. All these, with the exception of translating the slit or image-slicer, give the capability of compensating for flexure in two orthogonal directions. The baseline for Gemini HROS, following our experiment on the ISIS spectrograph at the WHT, is to tilt the collimator mirror. As described below, the collimator has proved to be an effective and simple way of compensating spectrograph flexure.

The key point concerning open-loop correction is that *hysteresis* is the factor which limits the precision with which image-motion can be compensated, not the raw flexure. *Low hysteresis must be considered as the primary goal* for future open-loop active spectrographs. For a closed loop correction system the stress would be more on reducing the total amount of flexure in the spectrograph, so open and closed loop systems do not exclude each other, but they are ideally suited to work together.

3. FLEXURE COMPENSATION SYSTEM FOR ISIS

Having decided that active flexure compensation was the right way forward, we saw an opportunity of developing an active system for the ISIS (Intermediate-dispersion Spectroscopic and Imaging System) spectrograph at the 4.2m William Herschel Telescope (WHT). Whilst the basic motivation was a proof of concept for the Gemini

High Resolution Optical Spectrograph, it was also clear that ISIS exhibited flexure at a level where it could benefit from the active collimator approach we had proposed.

Systematic measurements of the position of the spectrum as a function of zenith distance and rotation angles of the Cassegrain turntable have shown evidence of flexure in the ISIS structure and probably in its supporting flange system. ISIS is one of the largest Cassegrain instruments currently in operation on a 4-m class telescope and therefore one where imaging stability is most critical. It is also a very popular instrument in the astronomical community for its versatility and overall performance. These reasons made it an almost ideal choice for the testing of our novel system of image stabilization.

The ISIS Spectrograph Automatic Active Collimator (ISAAC) project was originally conceived as an open-loop system for active compensation of flexure. The choice of the active element fell on the collimator for two main reasons: it was easy to control with rear mounted actuators to provide tip/tilt active corrections, and given its position in the spectrograph, it could be accessed and replaced easily. Moreover, the collimator is a natural choice for flexure compensation, being in place for every spectrograph setup and producing significant image shifts on the detector with relatively small tilts. The last is important in avoiding optical aberrations which could be introduced if the correction significantly disturbed the optical geometry.

Flexure compensation with an active collimator is based on a simple concept: when the spectrum position on the detector drifts during tracking, the collimator is slightly tilted, thus deflecting the reflected beam and restoring the image to the original position. The collimator is finely steered with three piezoelectric actuators pushing against its back. The amount of tilt required is calculated from the spectrograph geometry and a previous set of flexure measurements. The orientation in space of the spectrograph is calculated from the equatorial coordinates of the observed object, the sidereal time at the start of the exposure, the elapsed time and the position angle of the slit on the sky.

The instrument hardware consists of three separate sub-systems. The mirror cell and actuator system are connected through coaxial cables to the actuator driver which sits in the instrumentation rack on the telescope. This in turn is connected via a serial link to a standard lap-top 486 PC in the control room, which runs the main software. During our experiment the user interface was the lap-top PC. This was completely separate from the ISIS control system, as demanded by the La Palma Observatory as a condition for conducting the experiment.

3.1. Mechanical design and construction

The mechanical design of the active collimator was driven by two major objectives: to make it compatible and easy to interchange with the existing collimator, and to minimize hysteresis. The latter is required to avoid introducing new sources of image shifts during the correction process. Hysteresis in the collimator movements would affect the accuracy with which the spectrum position on the detector can be controlled and could make the active compensation ineffective. The resulting design is shown in fig. 3.1. in an exploded view.

Three piezoelectric actuators are fitted with strain gauge position sensors. This combination of actuator and sensor allows a very fine control of the collimator displacement with a nominal accuracy better than $75nm$ over temperature excursions up to 20^{o}. The actuators have a total expansion of $15\mu m$ and a lifetime of more than 10^{9} cycles. The rear actuators in combination with three neoprene pads on the mirror front edge, also provide the axial support of the mirror. These pads are located exactly opposite the actuators to avoid introducing bending moments, and act as springs to produce the necessary return force for the small tilting of the mirror. The overall movement is very small: the maximum tilting angle is about $17arcsec$ ($10^{-4}rad$) when the actuators are at full travel. This corresponds to image motions of $80\mu m$ or 3.4 pixels at the camera focal plane ($24\mu m$ pixels). The radial support is provided by a flexural element connected to the external cell and bolted to a small Invar block inserted in a blind hole in the back of the Zerodur mirror. This system was chosen to minimize thermal stresses between the mirror and the mirror cell.

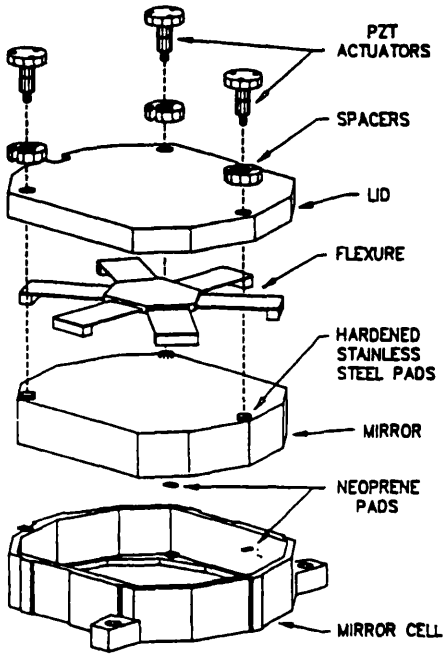


Figure 1: Exploded view of the active collimator cell. The mirror is roughly 4*cm* thick. The flexure provides radial support through a small Invar cylinder inserted in the back of the mirror (not shown).

3.2. Flexure measurements on ISIS

On July 1995 we used ISIS for tests at the William Herschel Telescope. The aim was to study flexure in ISIS, monitor it effectively and then prove that we could compensate it with the active collimator. The ISAAC collimator was installed on the ISIS blue arm and its focus and tilt were adjusted using the standard Hartmann shutter. The optical quality was checked by imaging a narrow slit illuminated with Cu-Ne and Cu-Ar calibration lamps. The results were at least as good as those obtained with the original collimator. Most of our subsequent tests were performed with a $150\mu m$ slit and with the low dispersion ($158 lines/mm$) grating centered around $\lambda = 7000\AA$. This value was chosen to provide a good S/N ratio in the calibration lamp spectra. All the data were recorded using the same 1024 x 1024 Tektronix CCD with 24 x $24\mu m$ pixels (La Palma Tek1 CCD). The reference spectra were generated by observing an arc spectrum through a short dekker, so that each spectral line would produce an almost circular spot. The detector shifts (in spectral and spatial direction) were calculated by centroiding a single line in the spectrum. We checked this procedure with other lines in the spectrum which gave the same result within 0.05 pixels, confirming that this system was as effective for this application as cross-correlating the whole spectrum (and much quicker).

To achieve effective open loop compensation, flexure needs to be measured and modelled very accurately. For the purpose of flexure measurements the spectrograph orientation in space is defined by only two parameters: the zenith distance z of the telescope (or the elevation = $90^\circ - z$) and the position angle PA of the Cassegrain turntable on which ISIS is mounted. The telescope azimuth rotation does not change the position with respect to the gravity vector and is irrelevant. Flexure is measured by recording the shifts on the CCD detector of a point source for the different orientations in space. R. Rutten at the La Palma Observatory has generated an automatic procedure for producing a flexure map that covers the whole range of telescope and spectrograph possible positions. A program moves the telescope in elevation in 15° steps from zenith to $z = 75^\circ$, then moves the telescope back to zenith, rotates the turntable 45° and repeats the elevation stepping. This is repeated for 360° of turntable motion. A spectrum is acquired at every elevation/rotation setting. The automation not only speeds up the process considerably, but also assures that the data is taken always in the same way (with the same telescope motion and therefore same hysteresis).

We found that the raw data coming from this measurement were not a reliable model for our active corrections because of the large hysteresis present. Therefore, we modified the raw data by translating every track corresponding to a different turntable position angle so that the zenith points coincide, and used this as our flexure model (fig. 2). With this simple artifice the flexure map is easy to understand: the *amount* of flexure (relative to zenith) is essentially a function of telescope elevation, while the *direction* of flexure mainly depends on the turntable position. Moreover, the map becomes much more reproducible: fig. 3 shows three maps taken in three different days.

However, hysteresis is intrinsic in the spectrograph structure and mechanisms and difficult to eliminate. A

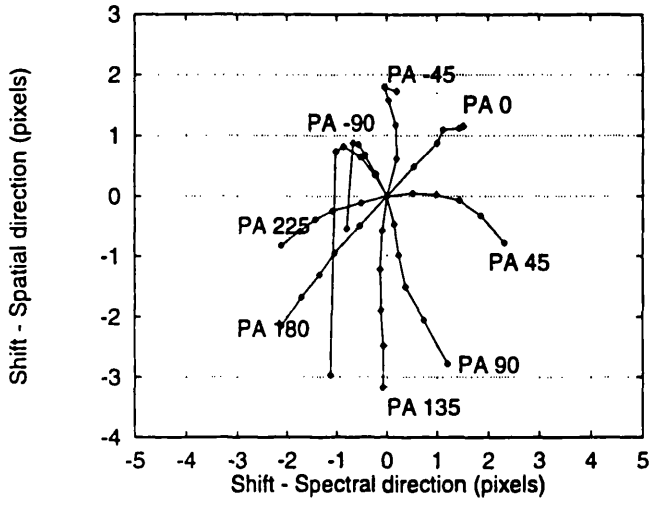


Figure 2: Modified flexure map. Each "arm" represents a different turntable position angle (PA), while the points in one arm correspond to different telescope elevation (from zenith to 15° above horizon in 15° steps). The map was produced by modifying the raw data so that the starting points (telescope at zenith) for each track coincide. This reduces the impact of hysteresis on the flexure model. The track for $PA = -90°$ is repeated. The sudden jump in this track for el = 15° is probably due to flipping of a loose component in ISIS.

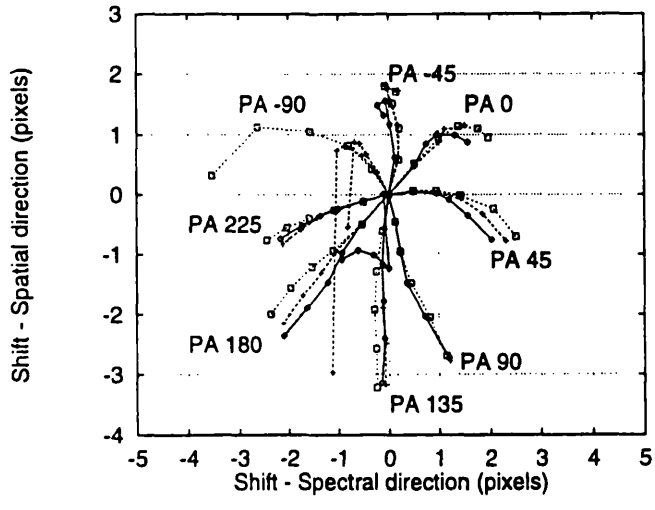


Figure 3: Comparison of three modified flexure maps used for modelling flexure in ISIS. This flexure model can be reproduced fairly well by repeating the measurement after a few days.

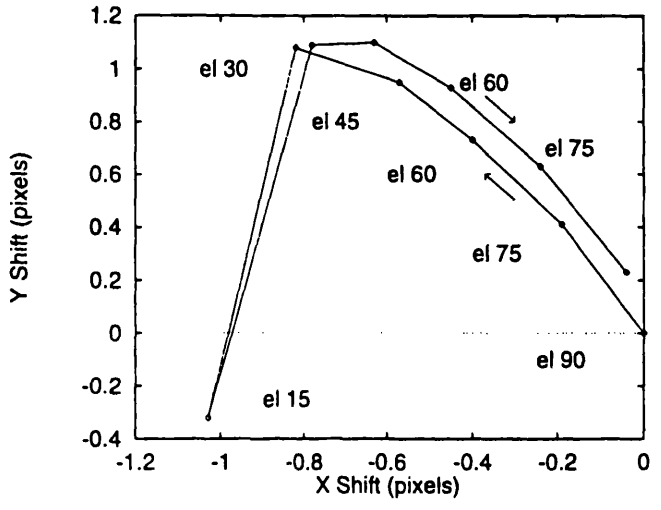


Figure 4: Hysteresis in the flexure shifts in ISIS. The curve was produced by stepping the telescope from zenith to horizon and back while keeping the turntable position fixed. The hysteresis shown is roughly 15% of the total motion.

similar hysteretic effect has been observed in other Cassegrain spectrographs.[5] The presence of hysteresis is proved by the graph of fig. 4, where the telescope was moved from zenith to horizon and back keeping the rotator position angle (PA) fixed. A classical hysteresis curve was produced.

The flexure model developed increases the reproducibility of the flexure shifts. We need to stress, however, that the flexure map is sensitive to changes in the spectrograph settings (grating, central wavelength, presence of dichroic, etc.). Moreover flexure changes with time, so that over one or two months, the flexure map (even if taken with the same settings) can look appreciably different. Even so, with the few data at our disposal we estimated[6] that an improvement of a factor between 1.3 and 2 is still possible with a model several months old.

3.3. Active compensation experiment

The tests for active compensation of flexure were performed during the daytime, using the calibration lamps to feed the spectrograph. In our first experiment we simulated a star track by moving the telescope in $20min$ steps following the motion of a "virtual" star on the sky. In this way we compressed more than $4hrs$ tracking into less than one hour of measurements. We compensated for the flexure shifts using an all-sky look-up table obtained with the procedure described earlier and then we repeated the same track without compensation. We also took the same measurements the following day to check how reproducible these results were. The compensated and uncompensated spectrum shifts are shown in fig. 5 for the spectral and spatial directions as a function of tracking time. The same shifts are shown as they would appear on the CCD detector in figure 6

Several conclusions can be drawn from these graphs. Firstly, that on average, active compensation of flexure can reduce spectral drifts to 0.2 pixels or less (0.3 pixels worst case) for more than 4 hours tracking. The exposure time during astronomical observations is usually limited to around one hour by accumulation of cosmic ray events on the CCD and not just by flexure. However, our experiment of compensating flexure over several hours proves that large amount of flexure can be corrected with the same accuracy. This may not have much impact on the

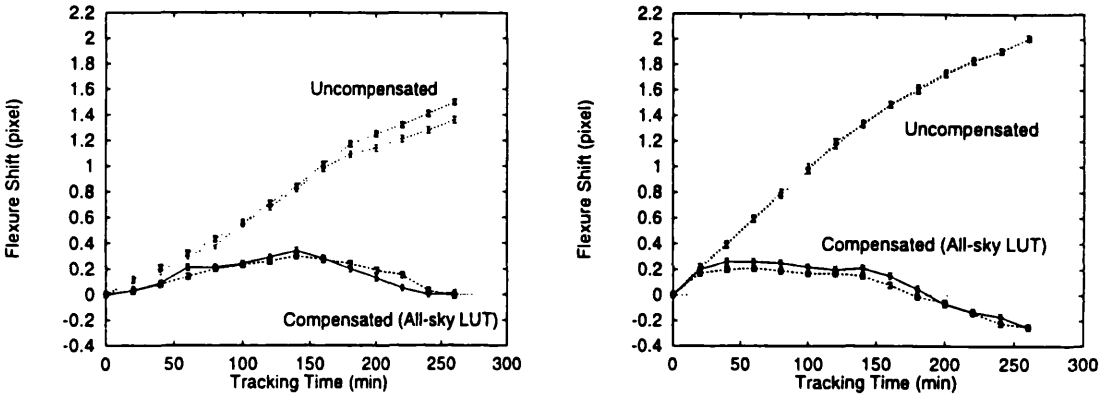


Figure 5: Flexure compensation with all-sky look-up table: spectral direction (left) and spatial direction (right). The same star track was repeated twice both with and without flexure compensation to check the reproducibility.

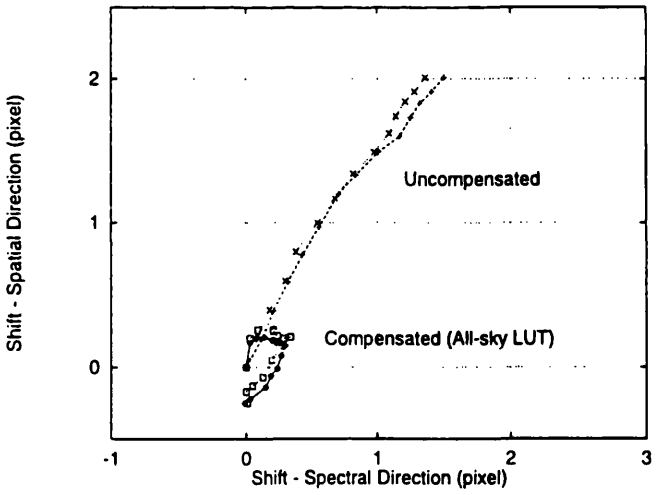


Figure 6: Flexure compensation with an all-sky look-up table. X and Y shifts are plotted together, as they would appear on the CCD detector.

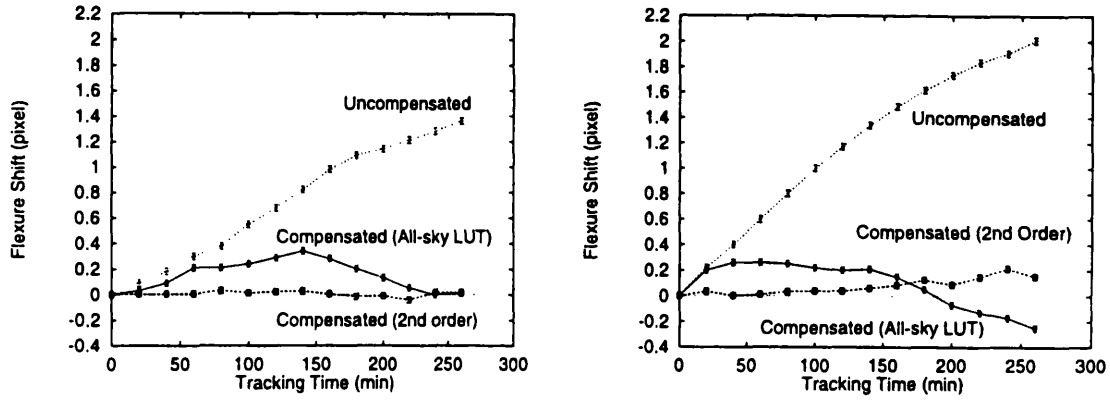


Figure 7: Second-order flexure compensation in spectral direction (left) and spatial direction (right). The second order corrections were obtained by feeding back to the control system the residuals from the all-sky compensation run.

length of exposures, but it is relevant in that it shows that active compensation can potentially avoid the need for frequent calibration arcs between exposures, saving telescope time.

The image stability we achieved is a significant improvement on the existing situation, with typical drifts of 1.5 – 2.0 pixels during the same time. The graph also shows that the improvement is more relevant for longer tracking time. This is because the uncompensated flexure usually increases fairly smoothly (at an average rate of $10\mu m/hr$), whereas the inaccuracy in compensation depends mostly on residual errors and very little on the amount of flexure to correct. Finally, the graphs prove that when the uncompensated flexure is reproducible, the curve representing the residual errors is also highly reproducible indicating that these errors are mostly due to inaccuracies in the flexure model and *not* to collimator movements.

This last point induced us to attempt to refine the look-up table, to see if we could improve image stability even further. For the same track, we calculated the residual errors and fed them back into the active collimator control system as second order corrections. In practice the procedure involved stepping along a track with active compensation on, recording the residual shifts, converting them into actuator voltage corrections and then running along the same track again applying the corrected voltages. These second-order corrections, for spectral and spatial directions, are shown in fig. 7 compared with the standard all-sky compensation. The improvement is quite remarkable: once the errors in the flexure model are corrected, the spectrum drifts can be contained within 0.1 pixels for nearly 4 hours of equivalent tracking time (more than 4 hours in spectral direction). Figure 8 shows the drift of the spectrum on the CCD with and without second-order flexure corrections.

The high level of stability achievable with second order flexure compensation indicates that the main limitation of an open loop system is the accuracy of the flexure model. In fact, when applying second-order corrections, the compensation system relied on two look-up tables for estimating the spectrograph flexure. The first, from the all-sky flexure model, provided the bulk of the correction required, while the second, derived from the measurements of the residual errors, gave the finer adjustments. This type of flexure model would be probably too complicated for practical use, but the residual errors could be measured in real time during the astronomical observation with an optical feedback, and this could operate in conjunction with the simpler flexure model. This type of closed loop system seems the best way to achieve high spectral stability in a Cassegrain instrument.

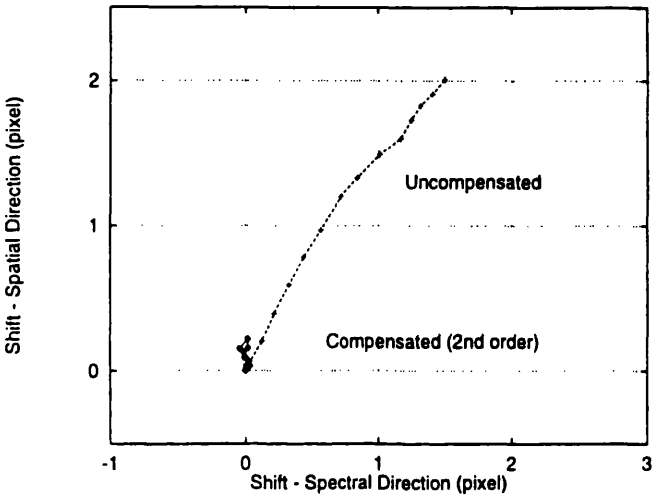


Figure 8: Second order flexure compensation. X and Y shifts are plotted together, as they would appear on the CCD detector. The spectrum shift is hardly detectable in this case.

4. FLEXURE COMPENSATION SYSTEM FOR HROS

4.1. The experience from ISAAC

This experiment has proved that active compensation is a viable and effective way of reducing flexure in Cassegrain spectrographs. It has also shown that the collimator can be controlled with very high accuracy and it is a suitable element for applying active corrections. More importantly, ISAAC proved capable of reducing flexure in ISIS significantly. These results have several implications on the astronomical data the spectrograph can provide, not only in terms of increased spectral resolution and reduced radial velocity errors. The reduction in flexure we achieved is such that the time wasted in taking regular calibration arcs can be cut drastically. Moreover, this low flexure can allow longer exposure times and improved S/N ratio: the limiting factor in the length of exposures would be effectively shifted from flexure drifts to cosmic ray events rate.

From the point of view of developing an active compensation system for HROS, here are the main conclusions that can be drawn from the ISAAC experiment:

1. Open loop flexure compensation is very effective in reducing flexure in the spectrograph, particularly over long tracking times. In ISIS flexure was reduced by roughly a factor 2 over 1 hour (from $\sim 16\mu m$ to $\sim 7\mu m$ combining X and Y shifts) and by nearly a factor 8 over 3 hours (from $\sim 46\mu m$ to $\sim 6\mu m$).

2. The residual error after open loop correction does not depend on the amount of flexure (i.e. on the tracking time) or on the opto-mechanical performance of the active collimator, but it is a consequence of modelling errors. Our laboratory tests had demonstrated that the collimator could provide movements as small as $0.4\mu m$ on the ISIS CCD. Most of the residual errors proved to be systematic and we were able to eliminate the greatest part of them by providing the collimator with additional modelling information (2nd order corrections). After the model was refined, the residual flexure was reduced to $\sim 3\mu m$ after 1 hour and to $\sim 4\mu m$ after 3 hours.

3. The errors in the flexure model are essentially caused by hysteresis in the flexure shifts. The large telescope movements involved in the measurement of a flexure map covering the whole sky suffered from large hysteresis. We estimated this to be about 15% of the total movement. As a consequence the raw data from a flexure measurement needed to be modified to try reducing the effect of hysteresis. Some residual errors obviously remained.

4. The flexure model may be improved by producing a map that covers the whole sky moving the telescope along star tracks of different declination. This model should simulate more closely the telescope movements and the flexure shifts during tracking. Hysteresis will still be present however, and because it depends on the previous history of telescope movements, it will be difficult to eliminate. Indeed, some other ways of modelling spectrograph flexure, based on a polynomial fit of the data, have been suggested,[7] although they relied on the absence of hysteresis in the spectrograph.

5. The flexure model also changes with time, so there would be no point in refining this too much, since it would soon become obsolete and a new, rougher model would provide a better flexure correction. We simulated the consequences of an aging model by quantifying the differences between flexure maps. We found that after 1 week the additional error would be less than $5\mu m$ in 70% of the cases. After 1 year we estimated an error of less than $7\mu m$ in 50% of the cases (the shape of the distribution changes with time).

6. A serious limitation of an open loop flexure model is the impossibility to account for malfunctions in the spectrograph. For example, it is fairly common for the spectrograph to exhibit flexure shifts larger than usual after being taken off the telescope. Since it may take weeks to track down and correct the problem, an open loop compensation system would give poor correction, unless the flexure map is updated.

Summarizing the above points, the experience with ISIS tells us that an open loop flexure compensation system would greatly improve the spectrum stability of the spectrograph and it is an attractive option for the design of new spectrographs. It would however leave some residual errors, which it would be difficult to eliminate. These errors are essentially due to modelling errors in the flexure map. To minimize these errors it would require frequent updates to the map. A closed loop system used in conjunction with an open loop one would clearly change the above conclusions, because it could account for any residual errors in the flexure model. By closing the loop with an optical reference beam propagated through the spectrograph, it is possible to avoid the problem of hysteresis altogether. The flexure shifts would be sensed in real time while the spectrograph is tracking and directly fed back into the collimator control. It also would avoid the need of producing several different flexure models for corresponding different spectrograph setups.

4.2. Closing the loop

With a closed-loop system, an optical reference is sensed directly, processed by the control system and the information used to drive the active element. We believe that, for better accuracy, the optical reference should be propagated through the *entire* spectrograph optical system, in order to compensate for motions of all surfaces. The light source for this reference can be i) monochromatic (fixed wavelength e.g. laser, or programmable using a continuum lamp and monochromator or a tunable laser), ii) emission line source, iii) continuum, or iv) continuum with absorption lines superimposed from e.g. a Fabry-Perot etalon. The reference beam can be propagated through the spectrograph as a narrow pencil of light (e.g. laser). Alternatively it can be a diverging cone to fill or partially fill the optics. In principle both methods should work. Overall, however, the filled-beam more closely emulates the science beam and therefore is considered preferable.

The reference can be propagated forwards, with injection at the slit and detection in the camera focal plane, or backwards, with the reverse. The latter has an advantage in that the usual case of $F_{cam} < F_{col}$ conveniently magnifies the motion to be sensed. A reference might be injected into the optical system as follows:

1. in the slit-plane (e.g. via a fibre), and sensed on a small CCD or quadrant photodiode adjacent to the science CCD.

2. at one end of the slit, and sensed on the science CCD. In this case the reference may be projected onto the $\sim$ 10 rows adjacent to the readout register. The procedure, which is repeated during the exposure is as follows: i) close shutter ii) clock out $\sim$ 10 rows encompassing the reference iii) reverse-vertical-clock by 10 rows to restore science image to original pixels iv) centroid or cross-correlate reference and update active element v) open shutter and continue integration. This method is analogous to the photographic spectroscopic procedure of exposing an arc either side of the science spectrum, the difference being that we propose to detect the reference repeatedly *during* the exposure. This method has been proposed as a closed loop extension of the open loop experiment conducted on the ISIS spectrograph, as described elsewhere.[8]

3. at the science CCD. In this case, the light source could be fibre fed, with the fibre output in the camera focal plane, but pointing towards the collimator and suitably baffled. The reference would propagate through the spectrograph backwards, being sensed on a small CCD or quadrant photodiode in the vicinity of the slit.

The spectrograph dispersing element(s) will deflect the reference beam depending on wavelength. If the reference were monochromatic, and the spectrograph configured for a different science wavelength, then the reference light may not reach a separate reference detector. Therefore the simple monochromatic laser reference is unattractive in most cases. For a prism cross-dispersed instrument such as Gemini HROS, the configuring for different wavelengths requires scanning the prism angles. The resulting change in prismatic deviation of the reference is even more difficult accommodate. Some of the technical challenges in developing a closed loop system on a spectrograph like HROS are discussed in the next paragraph.

4.3. HROS closed loop system

The details of the optical layout of the HROS spectrograph are described elsewhere in this conference proceedings.[9] Here it is sufficient to mention that the spectrograph is designed to work with two different cameras: a short camera having focal length $F_{cam} = 480mm$ and giving a resolution $R = 50000$ and a long camera having $F_{cam} = 1135mm$ and giving $R = 120000$. The stability requirement for HROS at the Cassegrain port has already been defined by the Gemini Science Committee[10] and it is a very ambitious target of $2\mu m/hr$. This is much more than similar Cassegrain instruments, like LRIS and ISIS,[11,12] are achieving to this date. Therefore, we believe that active compensation of flexure becomes essential to meet this specification.

The solution we propose for HROS is a dual system, including both open and closed loop corrections. This system would highly increase the level of stability achievable and would overcome the shortcomings of the open loop system mentioned before. Hysteresis would not be an issue any more because the closed loop control would correct the residual errors in the open loop model. For the same reason, frequent updates in the look-up table would not be necessary, since larger modelling errors can be tolerated. Moreover, the shifts of the reference beam constitute an additional flexure measurement, and could be used to improve the open loop model. Finally, this dual option provides a backup system, which would still give some flexure correction if one of the two systems fails.

We are still refining the characteristics of a practical closed loop referencing system. At this stage, an arc spectrum filtered through a monochromator to cover a single echelle order, seems to be the best option. A small linear detector (say 10 x 1000 pixels) on top of the science CCD or the science CCD itself, with the method of reverse-clocking described earlier, could be used to record the optical reference. The monochromator could then be tuned according to the spectrograph setting to ensure that the reference beam always hits the detector. The challenge is to achieve sufficient throughput from the arc lamp and monochromator to allow relatively short integration times for the optical reference (30-60 seconds). High throughput is needed due to the high dispersion of the spectrograph. If the light from the monochromator were fibre-fed to the slit plane of the spectrograph, the reference could fill the optics and still be able to reach the reference detector for every grating (and prism) setup.

5. ACKNOWLEDGMENTS

We wish to thank all the staff at the La Palma Observatory and in particular R. Rutten for his support and advice to the ISAAC project. We also thank the Royal Greenwich Observatory staff, including M. Johnson and S. Worswick for providing advice and information concerning ISIS and access to the cold room facility in Cambridge. We thank the all the staff at the Optical Science Laboratory and the the Gemini IGPO for their funding of part this research. The Gemini 8-m Telescopes Project is managed by the Association of Universities for Research in Astronomy, for the National Science Foundation and the Gemini Board, under an international partnership agreement. Part of the work was also supported under PPARC grant GR/H 34708.

6. REFERENCES

[1] B. Venkatraman and S.A. Patel, *Structural Mechanics*, p. 176, McGraw-Hill, New York, 1970.

[2] J. Allington-Smith and T. Davidge, "Proposed medium-low resolution optical spectrograph for GEMINI," *Instrumentation in Astronomy VIII*, D.L. Crawford and E.R. Craine ed., Proc. SPIE **2198**, 201-212, March 1994.

[3] D.D. Walker, M. Dryburgh, B.C. Bigelow, *Control of Flexure in Gemini Instrumentation*, Design study report commissioned by th Gemini Tucson office, November 1992

[4] R. Bhatia, "Active spectrographs: a new way to improve their quality," *Instrumentation in Astronomy VIII*, D.L. Crawford and E.R. Craine ed., Proc. SPIE **2198**, 554-558, March 1994.

[5] J. Allington-Smith, M. Breare, R. Ellis, D. Gellatly, K. Glazebrook, P. Jorden, J. MacLean, P. Oates, G. Shaw, N. Tanvir, K. Taylor, P. Taylor, J. Webster, S. Worswick, "A Low-Dispersion Survey Spectrograph (LDSS-2) for the William Hershel Telescope," *Publ. Ast. Soc. Pac.* **106**, 983-991 (1994).

[6] P. D'Arrigo, F. Diego, D.D. Walker "Active Compensation of Flexure on the WHT ISIS Spectrograph," *Mon. Not. R. Astron. Soc.*, **281**, 679-686 (1996).

[7] U. Munari, M.G. Lattanzi, "Flexures of Conventional Cassegrain-Fed Spectrograph," *Publ. Ast. Soc. Pac.* **104**, 121-126 (1992).

[8] D.D. Walker, P. D'Arrigo, "On the Stability of Cassegrain Spectrographs," *Mon. Not. R. Astron. Soc.*, **281**, 673-678 (1996).

[9] F. Diego et al. "The High Resolution Optical Spectrograph for the Cassegrain focus of the Gemini-south Telescope," these Proceedings.

[10] F. Gillett, "Project Scientist's Outlook," *Gemini Newsletter*, No. 10, 5-7, June 1995.

[11] J.B. Oke, J.G. Cohen, M. Carr, A. Dingizian, F.H. Harris, R. Lucinio, S. Labreque, W. Schaal, S. Southard, "The Low Resolution Imaging Spectrometer for the Keck Telescope," *Instrumentation in Astronomy VIII*, D.L. Crawford and E.R. Craine ed., Proc. SPIE **2198**, 178-184, March 1994.

[12] J.B. Oke, J.G. Cohen, M. Carr, J. Cromer, A. Dingizian, F.H. Harris, S. Labreque, R.Lucinio, W. Schaal, H. Epps, J. Miller "The Keck Low-Resolution Imaging Spectrometer," *Publ. Ast. Soc. Pac.* **107**, 375-385 (1995).

Appendix C

HROS Technical Notes

C.1 List

During my Ph.D. work I also wrote or contributed to the writing of Technical Notes on various subjects concerning the conceptual design of the High Resolution Optical Spectrograph. Here follows a list of the Notes included in this appendix.

- Technical Note # 06. Flexure Aberrations and Instrument Profile.
- Technical Note # 07. Error Budget for HROS.
- Technical Note # 17. Flexure Compensation System Performance.

HROS Technical Note #6
Flexure, Aberrations and Instrument Profile

P. D'Arrigo

Version 5

May 2, 1996

1 Introduction

The purpose of this technical note is to determine the effect of flexure and optical aberrations on the instrument profile. This will allow us to specify targets for the mechanical and optical design. Starting form the general requirements for stability and resolution defined by the Gemini Science Committee, the implications for the HROS design are estimated. This note is mainly concerned with the instrument profile, which is the image of the slit as observed on the CCD detectors when observing a narrow (unresolved) spectral line. The treatment of flexure and aberrations is only approximated, since only a full ray-tracing of the entire spectrograph would give the exact answers.

2 Define Symbols

ω projected slit width on CCD

w theoretical slit profile on CCD

s half-width of projected slit profile on CCD

α angle of incidence at echelle in air

β angle of diffraction at echelle in air

ϕ angle of collimator tilt for flexure correction

F_{cam} camera focal length

x CCD spectral direction

y CCD spatial direction

X CCD length

Δx spectrum shifts on CCD due to flexure

Φ flexure rate of motion on CCD (flexure shift)

a optical aberration function

Δa FWHM of the aberration function

3 Requirements

The specification for the image quality and stability on the detectors of HROS originates from the requirement of a resolution element of 2.5 pixels at $\mathcal{R} = 50000$ [1]. In fact, this sampling was subsequently increased to 3 pixels to overcome an otherwise impractically fast camera. In absence of flexure and other aberrations, the projected slit width on the detector would then be $\omega = 3$ pixels, which, with $15\mu m$ pixels translates into $\omega = 45\mu m$. There are several factors that may contribute to change the shape and width of the slit profile. We will assume to observe an unresolved spectral line, whose intrinsic width (when calculated in microns on the detector) is much smaller than the projected slit width. In this case the theoretical slit profile on the detector $w(x)$ is a rectangular step, with a width of three pixels. The effective slit profile on the detector $\omega(x)$ will in general depend on:

- Theoretical slit profile $w(x)$
- Optical aberration function $a(x)$
- Flexure shifts Φ
- Diffraction
- Internal seeing

In practice the last two items are negligible in a spectrograph like HROS, but flexure and optical aberrations contribute significantly. The effect of flexure is however of a different type and it needs to be considered separately from aberrations.

4 Flexure Contribution

The maximum level of flexure shift that can be tolerated has been specified as $\Delta x_f = 2\mu m$ [1] over one hour of telescope motion. Since HROS will have an active flexure compensation system, this is not the requirement for the mechanical stability of the entire spectrograph structure. Active compensation can theoretically reduce flexure at the center of each order to zero, the limiting factor being the accuracy with which flexure shifts can be measured. However, even when flexure is perfectly compensated in the center, the spectrum can still move at the edges of the echelle orders. This happens because the correction, applied through a tilt of the collimator, changes the angle of incidence on the grating.

First, we consider a tilt of the collimator in isolation. When the collimator is tilted by a small angle o, the central ray coming from the slit is deflected by twice that angle, due to the reflection on the collimator. At this point we need to distinguish between tilts in the grating dispersion plane and tilts orthogonal to it: if the tilt is in the dispersion plane, it will be affected by the anamorphism of the grating. The incident parallel beam on the grating is in fact tilted by the same angle $2o$, so that if α was the original angle of incidence (in air) we have:

$$\alpha \rightarrow \alpha + 2o \tag{1}$$

The angle β at which the beam reflected by the grating is connected to α through the grating equation, which introduces an anamorphic factor. It can be demonstrated that

this factor is unchanged by the immersion of the grating so that for small changes in α we have:

$$\frac{\Delta\beta}{\Delta\alpha} = \frac{\cos\alpha}{\cos\beta} \tag{2}$$

Thus the total deflection due to flexure is:

$$\Delta\beta = 2\phi\frac{\cos\alpha}{\cos\beta} \tag{3}$$

If the collimator tilt is not in the dispersion plane, then the anamorphism is absent. We have therefore to distinguish between detector shifts in the spectral direction Δx and shifts in the spatial direction Δy, which are obtained by multiplying the beam tilt times the camera focal length, having:

$$\Delta x = 2F_{cam}\phi\frac{\cos\alpha}{\cos\beta} \tag{4}$$

and

$$\Delta y = 2F_{cam}\phi \tag{5}$$

These equations give us the relation between collimator tilts and the shift of the spectrum on the detector. To compensate for a flexure shift $\Delta x = 1$ pixel, we need to tilt the collimator by an angle $\phi = 24\mu rad$ ($5arcsec$) when using the short camera ($F_{cam} = 480mm$); the tilt with the long camera would be smaller.

Second, we consider how a tilt of the collimator affect the echellogram on the camera detectors. A tilt ϕ of the collimator changes the angle of incidence on the grating, albeit by a very small amount. It is necessary, therefore, to calculate the effect on the spectrum at the edges of the CCD when compensating for the motion at the centre. When the centre of the echellogram is reflected at an angle β, the edge will be reflected at a different angle $\beta' = \beta + d\beta$, where $d\beta$ is half the angular size of the echelle order. To calculate the effect on β' by the collimator tilt, we need to derive how the change in the angle reflected by the grating $\Delta\beta$ due to flexure correction, depends on the angle of reflection itself β. This relation can be obtained by differentiating equation 3 with respect to β.

$$\frac{d}{d\beta}(\Delta\beta) = 2\phi\cos\alpha\frac{d}{d\beta}\left(\frac{1}{\cos\beta}\right) \tag{6}$$

Therefore:

$$\frac{d}{d\beta}(\Delta\beta) = \Delta\beta\cdot\tan\beta \tag{7}$$

Or:

$$d(\Delta\beta) = d\beta\cdot\Delta\beta\cdot\tan\beta \tag{8}$$

The above equation tells us that while the centre of the spectrum is shifted by an angle $\Delta\beta$ when the collimator is tilted by ϕ, the edge is shifted by $\Delta\beta + d(\Delta\beta)$, thus stretching the spectrum. Combining equation 4 and equation 3 we have an expression for the flexure correction that depends on the shift on the CCD (rather than on ϕ):

$$\Delta\beta = \frac{\Delta x}{F_{cam}} \tag{9}$$

Also, for an angle β' corresponding to the edge of the CCD, the angle $d\beta$ is half the angular dimension of the CCD. If we indicate with X the CCD length, then:

$$d\beta = X/2F_{cam} \tag{10}$$

Combining the last three equations together and indicating with $\Delta x_{edge} = F_{cam} d(\Delta\beta)$ the motion of the spectrum at the edge of the CCD relative to the one at the center, we obtain:

$$\Delta x_{edge} = \frac{X \Delta x}{2F_{cam}} \tan\beta \tag{11}$$

In the equation, the total width of the CCD X can be replaced by the order length when the order does not cover the whole CCD. Δx is the flexure correction applied in the centre. For HROS we have $X = 61mm$ and $\beta = 57.4°$. For the $\mathcal{R} = 50000$ mode with the short camera ($F_{cam} = 480mm$), $\Delta x = 1$ pixel produces $\Delta x_{edge} = 0.10$ pixels. That means there is a factor 10 between the shift at the edge and the flexure correction applied. If we require the whole spectrum (including the edges) to be stable to less than $2\mu m/hr$, the maximum allowable correction will be ten times that. Therefore, the mechanical requirement for the stability of the spectrograph can be set to:

$$\Delta x_{mec} = 20\mu m/hr \tag{12}$$

This is the goal for the maximum level of flexure in absence of active compensation. The value is specified over one hour of telescope motion, to follow the overall stability requirement (i.e. including flexure compensation). When active compensation is applied, the above requirement will ensure that the maximum shift over the whole spectrum does not exceed the specification:

$$\Delta x_f = 2\mu m \tag{13}$$

The above value will be taken as the total flexure contribution to the spectrum spreading on the detector. The use of the long camera does not affect the above calculations, because although the motion at the edges is reduced by the ratio of the camera focal lengths, the total motion is magnified by the same factor, thus producing the same result.

Effect on Slit Profile. To calculate the effect of flexure on the theoretical slit profile $w(x)$, we will assume that flexure will shift the spectrum always in the same direction and at a constant rate Φ. The unperturbed slit profile will be a rectangular step, as shown in fig. 1, having a mathematical expression of the form:

$$w(x) = \frac{I}{2}[\mathrm{sgn}(x + s_o) - \mathrm{sgn}(x - s_o)] \tag{14}$$

Due to flexure, this profile will be shifted while the detector accumulates photons during the exposure, thus at a time t:

$$s(t) = s + \Phi t \tag{15}$$

A the same time, the intensity also increases: $I = I_o t$. We can therefore write the time-dependent slit profile as:

$$w(x,t) = \frac{I_o t}{2}[\mathrm{sgn}(x + s - \Phi t) - \mathrm{sgn}(x - s - \Phi t)] \tag{16}$$

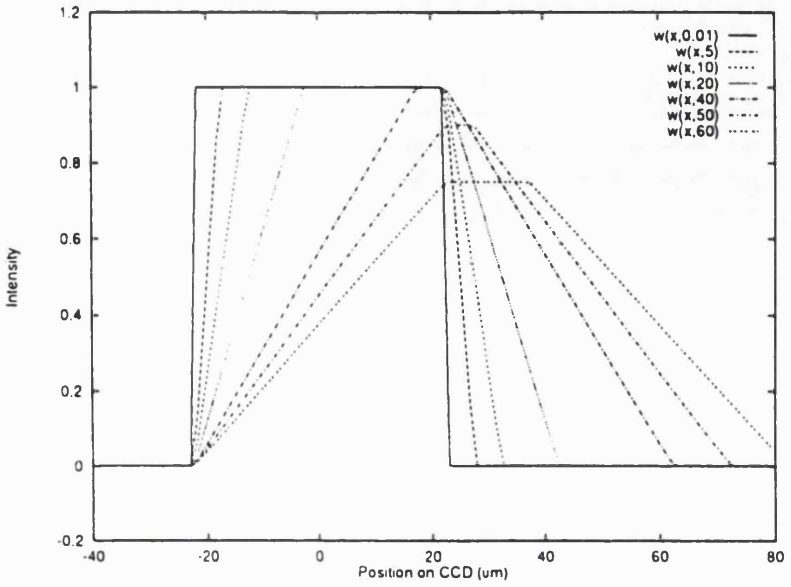


Figure 1: Convolution of the theoretical projected slit profile with several constant-rate flexure shifts. The projected slit width is $45\mu m$, the convoluted function has still the same FWHM as long as $\Phi \leq 45\mu m$.

The resulting slit profile, as recorded by the detector is obtained by integrating over t:

$$\omega(x) = \frac{1}{T}\int_0^T w(x,t)dt \tag{17}$$

The time of integration T is arbitrary and we can take $T = 1$. In this case we can consider Φ as the total flexure shift during the exposure. We can also normalize the intensity to 1, so that the above integral becomes:

$$\omega(x) = \frac{1}{2\Phi}\left(\mid x - s - \Phi \mid - \mid x + s - \Phi \mid + \mid x + s \mid - \mid x - s \mid\right) \tag{18}$$

The above function is shown for different values of Φ in fig 1. The figure shows that as long as $\Phi \leq 2s$, both the peak and the FWHM of the profile is still unchanged. The spectrum as a whole is however shifted: the centre of the projected slit can be calculated to be shifted by:

$$\Delta x_f = \frac{\Phi}{2} \tag{19}$$

We can therefore see that, in absence of significant aberrations and except for very large values (Φ greater than the projected slit width), flexure does not affect resolution, but it changes the calibration of the spectrum. This is true if we take the standard definition of resolution as $\mathcal{R} = \lambda/d\lambda$ and we define $d\lambda$ as the FWHM of a spectral line. This definition does not apply to the resolution of two close spectral lines, where Rayleigh's criterion gives a better estimate. It is possible to demonstrate that in this case the effect of flexure is slightly more significant. In practice however, we can for the moment neglect the effect of flexure on the projected slit width and on resolution.

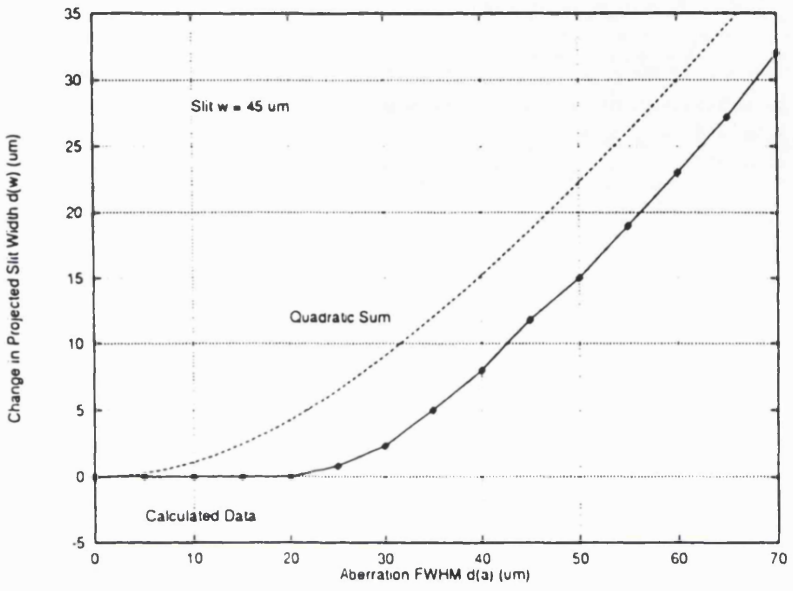


Figure 3: Change in the projected slit width (from original width $w = 45\mu m$) as a function of the FWHM of the Gaussian aberration function. A quadratic sum of slit and aberration widths is also shown for comparison.

We require this width not to be significantly larger than the theoretical width, $\Delta w = 45\mu m$. In order to quantify this, we can impose that the real slit width is no more than a fraction ϵ wider than Δw, that is:

$$\Delta\omega = (1 + \epsilon)\Delta w \tag{25}$$

Since the convolution process will increase the theoretical slit width by an amount $\Delta\omega_a$ due to aberrations, we can combine the two previous equations to get the condition:

$$\Delta\omega_a = \epsilon\,\Delta w \tag{26}$$

Taking $\epsilon = 0.05$ we obtain:

$$\Delta\omega_a = 2.25\mu m \tag{27}$$

The corresponding width of the aberration function cannot be calculated algebraically. The relation between this and the change in projected slit width is represented by the graph in fig. 3. When the convolution is taken into account, the maximum width of the aberration function becomes:

$$\Delta a \simeq 30\mu m \tag{28}$$

This is the maximum level of (Gaussian) aberration that, combined with the flexure requirement, would not perturb the slit profile by more than 5%. We saw in the previous paragraph that in a first approximation, flexure does not affect this calculation. However, to be more accurate, the convolution of aberration and slit profile should be taken as the basis for calculating the effect of flexure.

6 Flexure and Aberrations

The combined effect of flexure and optical aberrations on the slit profile is more difficult to calculate. In this case the time-dependent profile is derived from the expression for $c(x)$ as:

$$c(x,t) = I_o\left[\mathrm{erf}\left(\frac{x+s-\Phi t}{a}\right) - \mathrm{erf}\left(\frac{x-s-\Phi t}{a}\right)\right] \tag{29}$$

The above equation is integrated as before over t, to obtain the effective slit profile (with intensity normalized to one):

$$\omega(x) = \frac{a}{\sqrt{\pi}\Phi}\left[A(x) + B(x)\right] + \frac{1}{\Phi}\left[C(x) + D(x)\right] \tag{30}$$

Where:

$$A(x) = \exp\left[-\left(\frac{x-\Phi-s}{a}\right)^2\right] - \exp\left[-\left(\frac{x-\Phi+s}{a}\right)^2\right] \tag{31}$$

$$B(x) = \exp\left[-\left(\frac{x+s}{a}\right)^2\right] - \exp\left[-\left(\frac{x-s}{a}\right)^2\right] \tag{32}$$

$$C(x) = (x-\Phi-s)\mathrm{erf}\left(\frac{x-\Phi-s}{a}\right) - (x-\Phi+s)\mathrm{erf}\left(\frac{x-\Phi+s}{a}\right) \tag{33}$$

$$D(x) = (x+s)\mathrm{erf}\left(\frac{x+s}{a}\right) - (x-s)\mathrm{erf}\left(\frac{x-s}{a}\right) \tag{34}$$

The resulting graphs, for different values of Φ and a are shown in a series of figures: 4, 5, 6, 7, 8, 9, 10, 11. As with the case where only flexure was present, the FWHM of the curve is practically unchanged as long as $\Phi \leq\sim 2s$ and the width of the aberration function is small compared with $2s$. According to equation 13, the corresponding spectrum shift on the detector is:

$$\Delta x_f = s = 22.5\mu m \tag{35}$$

From the point of view of spectrograph resolution, flexure has a significant effect only when it is relatively large, of the order of 3 pixels. Only when aberrations start to become significant, the width of the effective slit profile is affected by flexure. This result has an important effect on the specification for the spectrograph stability.

7 Flexure and Resolution

As I mentioned earlier, flexure affects resolution through the effect it has on the width of the instrument profile. In fact, in a spectrograph the product of the projected slit width $\Delta\omega$ and the effective resolution $\mathcal{R}$ is approximately constant, that is:

$$\Delta\omega\mathcal{R} = k \tag{36}$$

Therefore any increase in the projected slit width proportionally reduces the spectrograph effective resolution. Since the projected slit width depends weakly on flexure, so will the change in resolution. In figure 12 the effective resolution is plotted as a function of the instrument flexure Φ, for different levels of spectrograph aberrations. The graph clearly demonstrates that even in the presence of significant aberrations ($30 - 40\mu m$ FWHM),

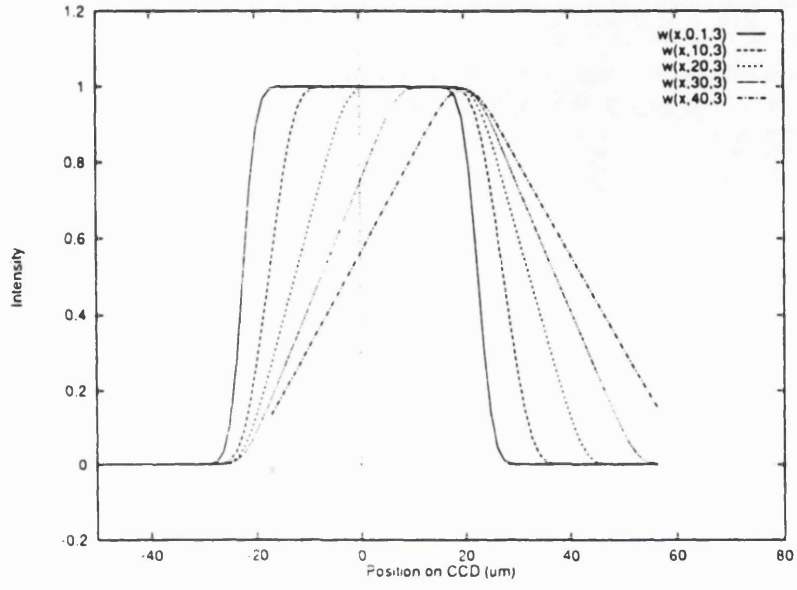


Figure 4: Effect of flexure on a slit profile in the presence of aberrations having FWHM $\Delta a = 5\mu m$. The theoretical slit width is $45\mu m$.

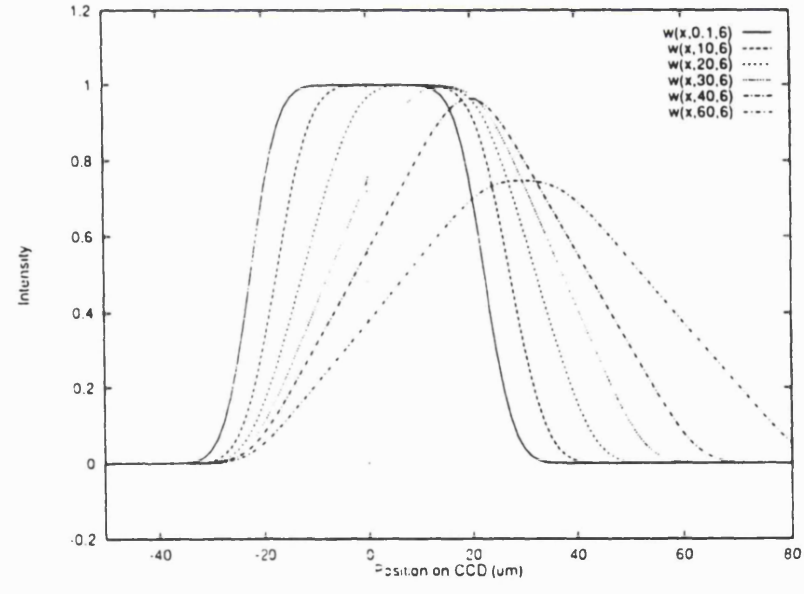


Figure 5: Effect of flexure on a slit profile in the presence of aberrations having FWHM $\Delta a = 10\mu m$. The theoretical slit width is $45\mu m$.

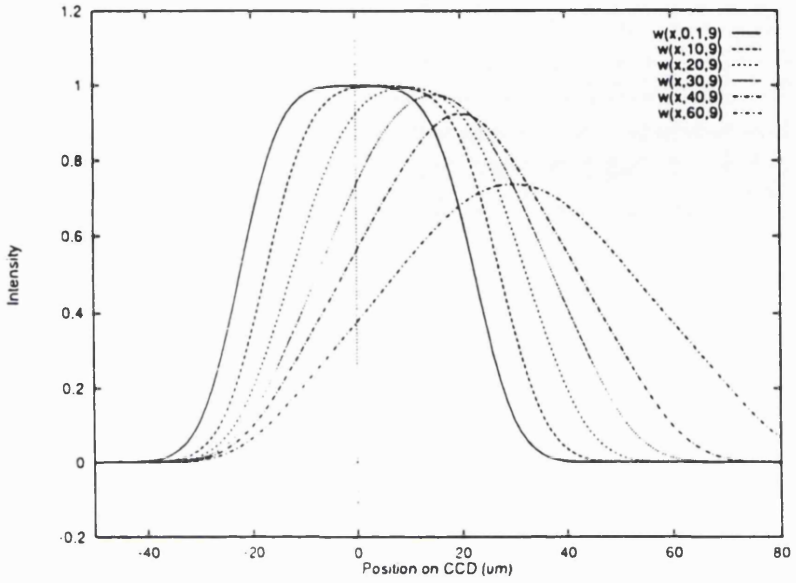


Figure 6: Effect of flexure on a slit profile in the presence of aberrations having FWHM $\Delta a = 15\mu m$. The theoretical slit width is $45\mu m$.

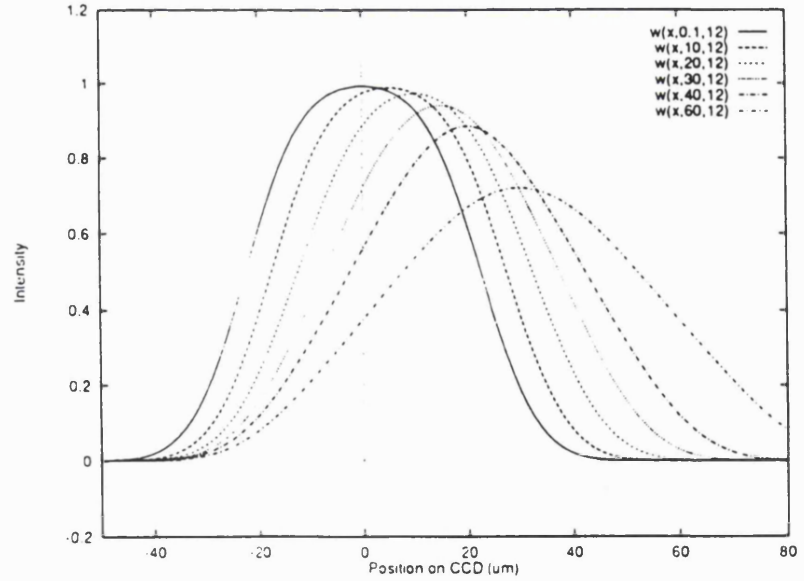


Figure 7: Effect of flexure on a slit profile in the presence of aberrations having FWHM $\Delta a = 20\mu m$. The theoretical slit width is $45\mu m$.

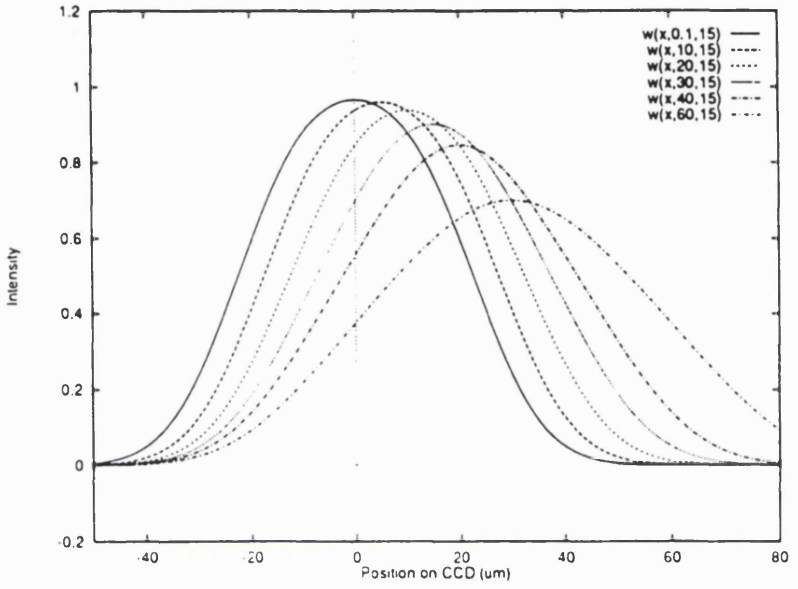


Figure 8: Effect of flexure on a slit profile in the presence of aberrations having FWHM $\Delta a = 25\mu m$. The theoretical slit width is $45\mu m$.

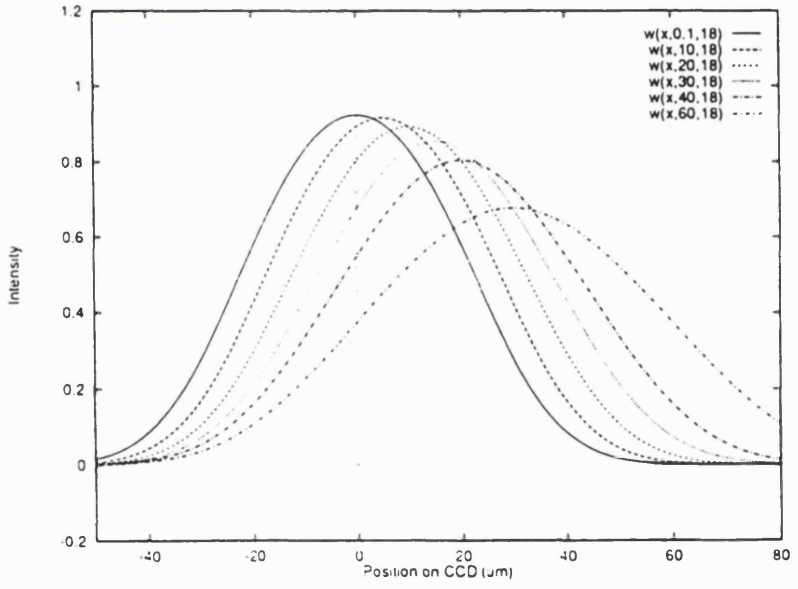


Figure 9: Effect of flexure on a slit profile in the presence of aberrations having FWHM $\Delta a = 30\mu m$. The theoretical slit width is $45\mu m$.

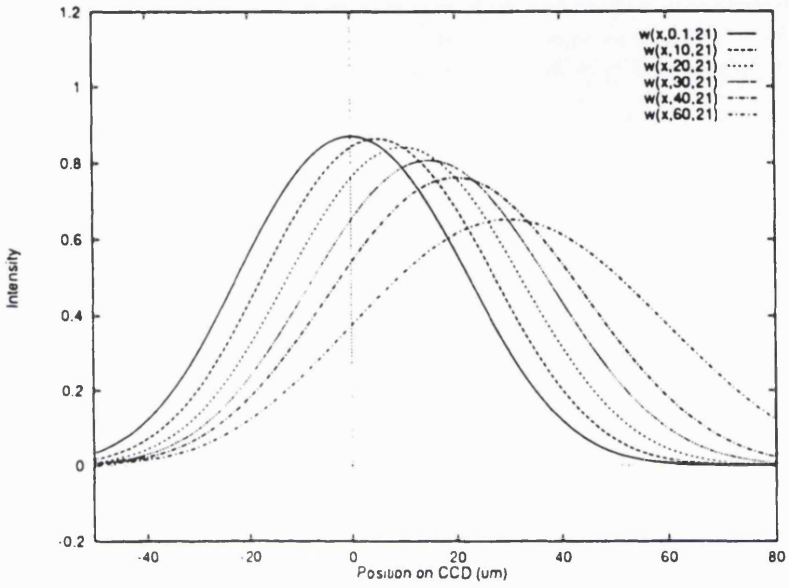


Figure 10: Effect of flexure on a slit profile in the presence of aberrations having FWHM $\Delta a = 35\mu m$. The theoretical slit width is $45\mu m$.

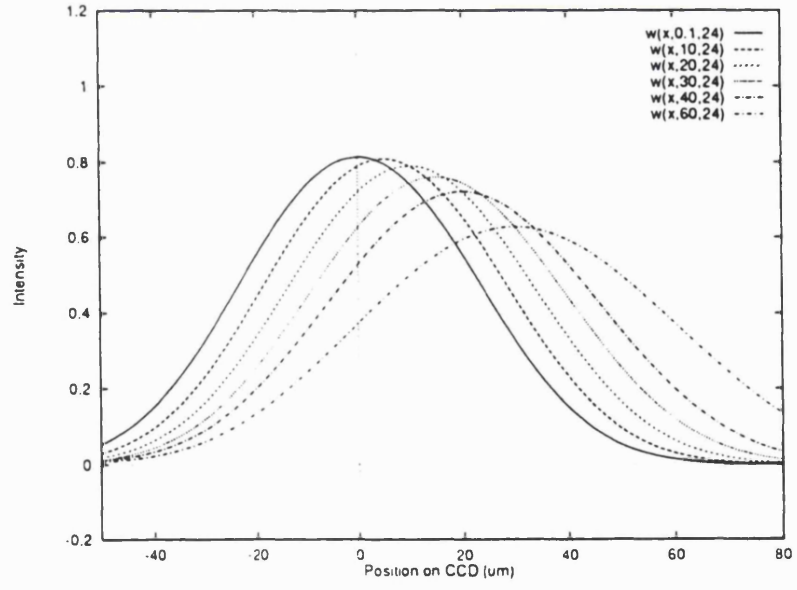


Figure 11: Effect of flexure on a slit profile in the presence of aberrations having FWHM $\Delta a = 40\mu m$. The theoretical slit width is $45\mu m$.

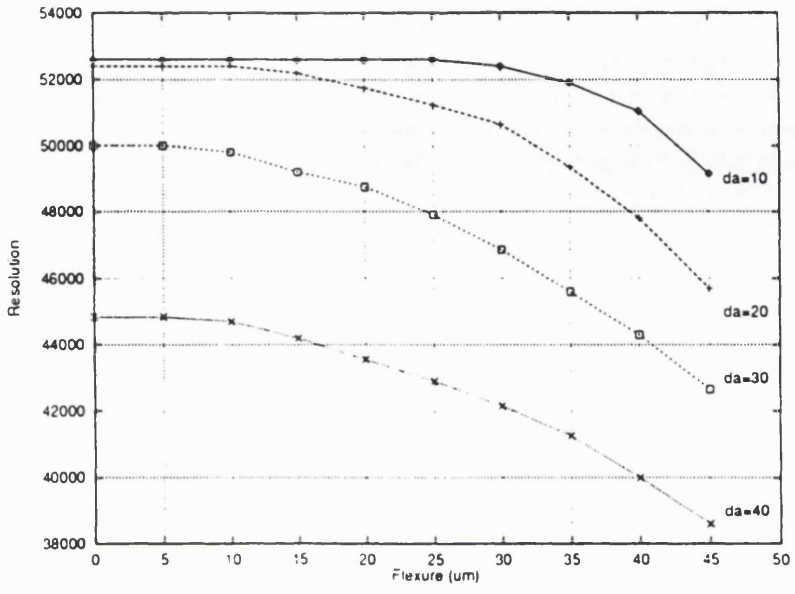


Figure 12: Change in the effective spectrograph resolution as a function of the instrument flexure Φ for different levels of aberrations. The aberrations *da* are in μm and measured as FWHM of an (approximate) Gaussian profile.

flexure has a small effect on the resolution. Therefore the stability requirement of $2\mu m/hr$ is an overkill as far as resolution is concerned. If we take the specified target of $30\mu m$ FWHM as the level of aberrations, by examining figure 12 we can conclude that up to $\sim 25\mu m$ of flexure would produce less than 5% degradation in resolution. Even if we take a consevative limit of 1% degradation in resolution, the amount of flexure that can be tolerated is:

$$\Phi \simeq 10\mu m \tag{37}$$

The $2\mu m/hr$ specification is therefore driven by the calibration accuracy of the spectrum, and not by the need to preserve resolution. This is actually much more sensitive to the level of aberrations, as fig. 12 shows.

8 Conclusions

In this note we showed that flexure has a small effect on the spectrograph resolution, when compared with the effect of optical aberrations. The main consequence of flexure is the error in calibration, so the target for a flexure compensation system is driven by the stability requirement ($2\mu m/hr$) and not by resolution. We showed that an active compensation system can meet the target stability with mechanical (passive) flexure levels of $20\mu m/hr$, which is therefore the goal for the mechanical design. We also estimated the effect of optical aberrations on the spectrograph resolution. Choosing the arbitrary target of a maximum 5% reduction in resolution (defined in terms of the width of the instrument profile), we showed that aberrations up to $\sim 30\mu m$ FWHM could be tolerated.

In view of the above results, however, it is questionable whether the specified $2\mu m/hr$ stability target is a reasonable goal. Dismissing the effect on resolution as negligible, we must ask ourselves what calibration errors can we tolerate and over what timescale. I would argue that maybe keeping the spectrum stable within $5 - 10\mu m$ over $3 - 5$ hours is a better target. A flexure active compensation system would have an intrinsic residual error which is almost independent from the telescope tracking time. Therefore it would be easier (and more useful) to achieve a lower stability in the short term, but a longer time over which that stability level can be maintained. For example, the specification for the mechanical design could be relaxed from $20\mu m/hr$ to $50\mu m$ in 2.5 hours, and the flexure compensation system would still be able to keep the spectrum within $5\mu m$ in the first 2.5 hours. Extending the telescope tracking time over which the spectrum is kept stable is extremely important, because it saves telescope time spent for calibration arcs.

References

[1] F. Gillett, Gemini Project Scientist, in *Gemini Newsletter* No. 10, p. 5, June 1995.

HROS Technical Note #7
Error Budget for HROS

P. D'Arrigo

Version 2

July 30, 1996

1 Introduction

In the HROS Technical Note No.6 we calculated how the specified targets for image stability and spectral resolution could be translated into meaningful parameters for an error budget. In the note we identified flexure and optical aberrations as sources of image degradation. More importantly, we found that for all but the largest flexure shifts, flexure *does not* affect resolution significantly. As a consequence, two separate error budgets can be constructed: one for the maximum allowable flexure in absence of flexure correction, and another for the total optical aberrations. We decided to specify the former as a flexure shift Δx_{mec} in microns on the detector after one hour of tracking and the latter as the FWHM Δa of a Gaussian aberration function projected on the detector. We found that the bottom line for these parameters is:

$$\Delta x_{mec} = 20\mu m \tag{1}$$

And:

$$\Delta a = 30\mu m \tag{2}$$

These maximum allowable errors have to be spread among all the components in the spectrograph that can contribute to the total figure. Since both the flexure and the aberration contribution of a component can add or subtract to the total, we chose to add these contributions with a quadratic sum:

$$\Delta = \sqrt{\sum_n \Delta_n^2} \tag{3}$$

Where Δ_n is the contribution from the n-th spectrograph component. To simplify matters, the spectrograph has been divided into areas, each one of which contains several components. As a first approximation, the contribution to the error budget has been spread equally among areas and among the components in the area. Table 1 contains the error budget for the flexure shifts, while table 2 has the ones for the optical aberrations. In both cases the long camera has been used as the basis for the estimate. The short camera is assumed to give a similar contribution. In the table with the aberration error budget the camera geometric aberrations have been given much greater weight than the other components, also the camera is considered as a whole and the individual components within it are not separated.

FLEXURE ERROR BUDGET			
Area	**Item**	Area EB	Item EB
SLIT		3.78	
	Periscopes		1.89
	Image Slicer		1.89
	Polarizers		1.89
	Filters		1.89
COLLIMATOR		3.78	
	Focal Modifier		1.69
	Folding Prism		1.69
	Collimator Support		1.69
	Collimator Focus		1.69
	Collimator Interchange		1.69
PRISMS		3.78	
	Prism Clamping		2.67
	Prism Support		2.67
ECHELLE		3.78	
	Echelle Support		2.67
	Echelle Tilt		2.67
CAMERA	(LONG)	3.78	
	Lens 1(L)		1.69
	Lens 2(L)		1.69
	Lens Interchange		1.69
	Mirror (L)		1.69
	Mirror Interchange		1.69
DETECTOR		3.78	
	Field Flattener		2.67
	CCD		2.67
STRUCTURE		3.78	
			$20.0\mu m$

Table 1: Error Budget for flexure shifts during one hour of tracking and *without* active flexure compensation. The numbers in the table are flexure shifts expressed in microns on the detector. Active corrections will reduce the total flexure to less than $2\mu m$.

ABERRATION ERROR BUDGET			
Area	Item	Area EB	Item EB
SLIT		7.42	
	Periscopes/Image Slicer		4.28
	Polarizers		4.28
	Filters		4.28
COLLIMATOR		7.42	
	Focal Modifier		4.28
	Folding Prism		4.28
	Collimator Mirror		4.28
PRISMS		7.42	
	Prism 1		5.24
	Prism 2		5.24
ECHELLE		7.42	
	Prism 3		5.24
	Echelle		5.24
CAMERA	(LONG)	25.00	
	Geometric Aberrations		23.00
	Surface Errors		6.93
	Alignment Errors		6.93
DETECTOR		7.42	
	CCD Flatness		7.42
			$30.0\mu m$

Table 2: Error Budget for optical aberration levels. The contribution of each optical component is expressed as the FWHM of the aberration function when projected on the detector (in microns).

HROS Technical Note #17
Flexure Compensation System Performance

P. D'Arrigo

Version 2

July 24, 1996

1 Introduction

The stability of the spectrum on the detectors of HROS is a critical part of the instrument design. At the HROS workshop of November 6-9 1995 it was proposed that an active system would correct flexure shifts in HROS. In this note we define the specifications for the active collimator and try to estimate its performance on the basis of the experience with ISIS.

2 Define Symbols

H actuator distance at collimator

S collimator tilt/CCD shift scale factor

L collimator actuator position

α angle of incidence at echelle in air

β angle of diffraction at echelle in air

ϕ angle of collimator tilt for flexure correction

F_{cam} camera focal length

x CCD spectral direction

y CCD spatial direction

X CCD length

Δx spectrum shifts on CCD due to flexure

Φ flexure rate of motion on CCD

$\mathcal{R}$ spectrograph resolution

3 Requirements

The specification for the image stability on the detectors of HROS originates from the requirement of a resolution element of 2.5 pixels at $\mathcal{R} = 50000$ [1] and a goal of 1/20th of the resolution element for the spectral stability over 1 hour. This produces a design goal of $\Delta x_f = 2\mu m$ maximum flexure shifts, considering $15\mu m$ pixel size and 1 hour exposure time. The requirement in spatial direction is less stringent, so the $2\mu m/hr$ limit can be taken as the goal for both directions.

4 Flexure Compensation in HROS

In specifying the design goals for the active collimator we will assume that guiding errors are not a significant source of flexure in good seeing conditions. This may require an accurate instrument wavefront sensor to provide the guiding reference.

4.1 Collimator Position Accuracy

In the Technical Note # 6 we calculated how the tilts of the collimator mirror relate to the shifts of the spectrum in the camera focal plane. If the collimator is tilted by an angle ϕ, the corresponding shift in spectral direction is (note that the immersion of the echelle does not change the anamorphic factor):

$$\Delta x = 2F_{cam}\phi\frac{\cos\alpha}{\cos\beta} \tag{1}$$

In HROS we have $\alpha = 69.4°$ and $\beta = 57.4°$, the anamorphic factor is then $\xi = \cos\alpha/\cos\beta = 0.65$. For the $\mathcal{R} = 50000$ mode with the short camera, $F_{cam} = 480mm$, while the $\mathcal{R} = 120000$ mode uses a long camera having $F_{cam} = 1135mm$. We can define a scale factor S between collimator tilts and detector shifts:

$$\Delta x(\mu m) = S\phi(\mu rad) \tag{2}$$

For the short camera $S_s = 0.63\mu m/\mu rad$ while for the long camera $S_l = 1.48\mu m/\mu rad$. To meet the $2\mu m$ stability goal, the collimator tilting error must produce shifts of less than $\Delta x = \pm 1.0\mu m$, meaning a tilting accuracy of:

$$\Delta\phi = \pm 0.67\mu rad \quad (\pm 0.14 arcsec) \tag{3}$$

The above specification is driven by the long camera, the corresponding value for the short one being $\Delta\phi = \pm 1.6\mu rad$. The accuracy required from the actuator system is very high, because of the very small angles involved. If we indicate with ΔL the accuracy in the actuator position and with H the distance between the two actuators providing the tilt, we have:

$$\Delta L = \frac{H\Delta\phi}{2} \tag{4}$$

The additional factor 2 in the equation is due to the fact that two actuators are necessary to tilt the collimator, so the accuracy of each need to be twice as high. The value of the position accuracy depends on the distance H over which the collimator is tilted.

To estimate the order of magnitude of this value, we can assume a nominal distance $H = 200mm$ between the actuators, so that:

$$\Delta L \simeq 70nm \tag{5}$$

This is a very stringent requirement for the actuators, but it is close to the specification for the ISIS active collimator ($80nm$). Therefore, the experience gained during the experiment on ISIS where piezoelectric actuators were used, not only proves that this is possible, but also indicates a way in which such requirements can be met.

4.2 Collimator Tilting Range

The tilting range of the collimator is a less critical issue because commercial piezoelectric actuators are available with a wide range of expansions. The requirement for the tilting range is driven by the size of the flexure shifts on the CCD. In Technical Note # 6 we showed how a maximum of $20\mu m/hr$ of flexure could be compensated with negligible effect on the dispersion across the echellogram format. A passive stability performance of $20\mu m/hr$ can reasonably be taken as the baseline for our calculations: this is a value consistent with spectrographs of a scale similar to HROS. The compensation system however, needs to account for several hours of tracking, in case the observing astronomer wants to take several exposures without calibration arcs in between. Indeed, one of the major advantages of an active flexure compensation system, can be the time saved in reducing the number of calibration spectra required per night.

A good observing night can include up to 8 hours of tracking, particularly during longer winter nights. At the specified rate of $20\mu m/hr$ the total motion on the CCD would be $x_o = 160\mu m$. Equation 2 will then give the corresponding collimator tilt. Since we are looking for the maximum range, the scale factor for the short camera should be used in this case giving:

$$\phi_o \simeq 255\mu rad \quad (53arcsec) \tag{6}$$

It is again possible to estimate the actuator specification, by taking $H = 200mm$ and using:

$$L_o = H\phi_o \tag{7}$$

The corresponding maximum expansion required from the actuators is then:

$$L_o \simeq 50\mu m \tag{8}$$

In the current design, however, two actuators are used to tilt the collimator in each direction. This means the above expansion can be halved, reducing the range specification for each actuator to $L_o = 25\mu$.

5 Open Loop Flexure Compensation System

Open loop flexure correction is based on data from a flexure model. This means X and Y flexure shifts are measured beforehand and saved in a look-up table. The control system uses these data to estimate the spectrograph flexure and to drive the active element. In the case of HROS, an active collimator seems the best choice. The experience with ISAAC (ISIS Spectrograph Automatic Active Collimator) showed that fine steering of the

collimator can produce very accurate motions in the camera focal plane, potentially at the sub-micron level.

The critical part of an open loop compensation system is the flexure model because the control of the active collimator relies on it to calculate the necessary correction. There are several point an ideal flexure model should fulfil:

- Cover the whole sky accessible by the telescope. This is obviously a fundamental requirement, but the implication is that a dense sampling of flexure shifts over such a wide range of telescope motions would take a long time to measure. During a normal astronomical exposure instead, only a small fraction of the sky covered by the model would be tracked by the telescope. One hour exposure covers around 15° in zenith distance and (normally) $25 - 30°$ on the Cassegrain turntable, roughly only 2% of the area covered by the model.

- Reproduce flexure accurately. There are two main difficulties in achieving this: one is the sampling problem mentioned above, and the other is hysteresis. Hysteresis in the telescope/spectrograph movements implies the flexure shifts are not determined univocally by the telescope position. The telescope motion before reaching a point in the sky also contributes and that cannot be accounted for in a simple point model. The poor sampling instead, can be significantly improved with interpolation techniques. Therefore, reducing hysteresis to a minimum is a priority for open loop flexure correction.

- Frequently update the flexure model. Flexure in a spectrograph is not a static, systematic error in the measurements. Changes in the spectrograph and even simple servicing can modify the behaviour of the structure, mechanisms and optical mounts with respect to flexure. The flexure model should be updated as often as possible, to maintain high accuracy and account for occasional malfunctions in the instrument.

- Require minimum effort from telescope staff. The procedure for updating the flexure model should be relatively simple and quick. The staff at the telescope are often fully committed to maintenance work, so unless the flexure measurement does not add significantly to the workload, frequent updates would be impossible. This means the model itself needs to be simple: there is no advantage in creating a sophisticated model that takes several days of measurements to produce.

6 Open Loop Performance

This discussion on the performance of an open loop flexure compensation system relies on the experience gained with the active collimator experience on ISIS at the WHT [2]. The experiment is described in detail in the cited paper, so here it is enough to show a graph that summarizes the results of the open loop experiment. Figure 1 shows the amount of flexure in spectral direction as a function of telescope tracking time. The existing flexure in ISIS is compared with the residuals after open loop compensation and after second-order corrections, produced by feeding the above residuals into the control system as a second look-up table. Several conclusions can be drawn from that experiment:

1. Open loop flexure compensation is very effective in reducing flexure in the spectrograph, particularly over long tracking times. In ISIS flexure was reduced by roughly

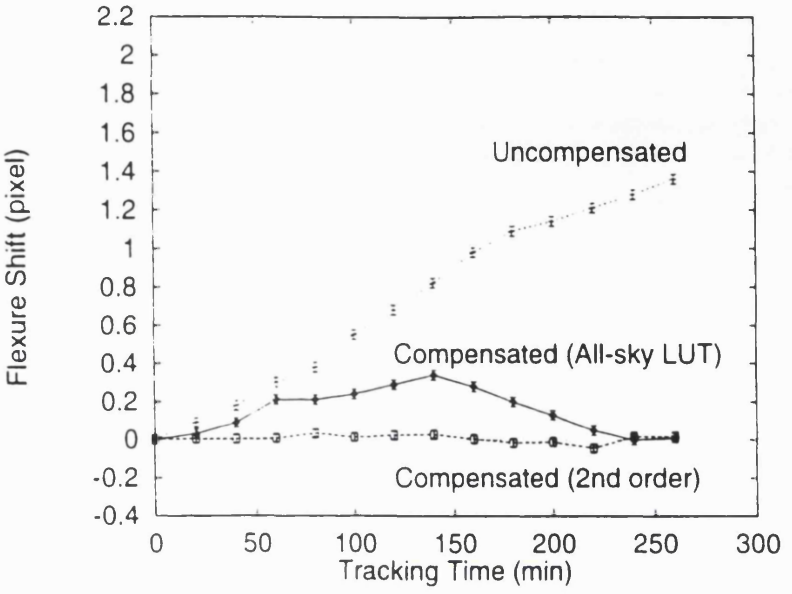


Figure 1: Open loop flexure compensation in ISIS (spectral direction). The uncompensated flexure shifts are compared with compensation with an all-sky flexure model and with second order corrections, using two look-up tables optimized for that particular telescope track. The pixel size is $24\mu m$, the track was covered in steps to save time.

a factor 2 over 1 hour (from $\sim 16\mu m$ to $\sim 7\mu m$ combining X and Y shifts) and by nearly a factor 8 over 3 hours (from $\sim 46\mu m$ to $\sim 6\mu m$).

2. The residual error after open loop correction did not depend on the amount of flexure (i.e. on the tracking time) or on the opto-mechanical performance of the active collimator, but it was a consequence of errors in the model. Most of the residual errors proved to be systematic and we were able to eliminate the greatest part of them by providing the collimator with additional model information (2nd order corrections). After the model was refined, the residual flexure was reduced to $\sim 3\mu m$ after 1 hour and to $\sim 4\mu m$ after 3 hours.

3. The errors in the flexure model were essentially caused by hysteresis in the flexure shifts. The large telescope movements involved in the measurement of a flexure map covering the whole sky, suffered from large hysteresis. We estimated this to be about 15% of the total movement. As a consequence, the raw data from a flexure measurement needed to be modified to try reducing the effect of hysteresis. Some residual errors obviously remained.

4. A suggested improvement on the flexure model used for ISIS, would be to produce a map that covers the whole sky by moving the telescope along star tracks of different declination (In ISIS the flexure model covered the sky by moving the telescope in elevation and rotating the turntable). This model should simulate more closely the telescope movements and the flexure shifts during tracking. Hysteresis will still be present however, and because it depends on the previous history of telescope

of the spectrograph flexure. It is assumed that such a closed loop system would operate in conjunction with an open loop look-up table, which would provide the bulk of the flexure correction. This system would then be equivalent to the second-order compensation in the ISAAC experiment on ISIS, with the second-order corrections provided by the optical reference.

There are three limiting factors in the accuracy of a closed loop flexure compensation system: the precision with which flexure is measured, the movement resolution of the active element and the uniformity of the correction in the camera field of view. The last two points affect also an open loop system: we will examine each point separately:

1. Flexure measurement. In the case of backward propagation, the closed loop measures the flexure shifts on a detector in the slit plane. Therefore, every movement of the image in the camera focal plane is magnified by the ratio M of camera and collimator focal lengths ($M = 2.3$). If the optical reference is an arc spectrum, the accuracy in the flexure measurement would be determined by the precision of a centroiding or cross-correlating algorithm that compares the relative position of two spectra. This will in turn depend on the number of pixels each spectral lines occupies and on the pixel size. I will assume the detector in the slit plane has $15\mu m$ pixels. The area illuminated by a spectral line of the Th-Ar lamp will depend on the size of the fibre which injects the light in the camera focal plane. For a $50\mu m$ fibre, the corresponding spot in the slit plane is $115\mu m$ or 7-8 pixels. Simple cross-correlation or centroiding techniques can usually achieve accuracies of less than 0.05 pixels ($0.75\mu m$), which would translate into less than 0.02 pixels ($0.3\mu m$) in the camera focal plane.

2. Active collimator resolution. The positional accuracy of an active collimator controlled by piezoelectric actuators can be extremely high. The relation between collimator tilts and spectrum shifts on the detector was discussed in a previous section. Here is enough to say that sub-micron positioning on the camera CCDs is within the possibility of such a system. The experience with ISIS showed that a tilting accuracy $\Delta o = \pm 1 \mu rad$ could be achieved. This is very close to the target value for HROS, so we are confident that with more accurate position sensors the goal of $\Delta x = \pm 1 \mu m$ can be met.

3. Uniformity in the correction. By choosing the collimator as the active element for the corrections, it is assumed that a small non-uniformity in the correction can be tolerated. The only way of having a completely uniform correction is to shift the CCD. The estimated effect of a collimator tilt was calculated in Technical Note No. 6 and it was proven to be very small. The spectrum at the edge of the CCD would still move even when flexure in the centre is perfectly compensated. The effect is however 10 times smaller than the compensation in the centre and it affects only a small portion of the total spectrum. This means with $20\mu m$ of flexure correction, the shift would be only $2\mu m$ at the very edge of the CCD.

8 Conclusions

This note shows how an active flexure compensation system can meet the stringent stability requirements for the High Resolution Optical Spectrograph working at the Cassegrain

port of the Gemini telescope. The system discussed is based on an active collimator providing the correction and on a combination of open and closed loop flexure measurement giving the necessary information to the control system. The use of the collimator is recommended for its simplicity and versatility, even if the motion of the CCD cannot be ruled out. The system based on the collimator has also the advantage of being tested on ISIS and and proved to work. The case for the closed loop system is based on the level of accuracy required. It is estimated that a simple open loop system would be pushed hard to meet the requirement for spectral stability.

References

[1] F. Gillett, Gemini Project Scientist, in *Gemini Newsletter* No. 10, p. 5, June 1995.

[2] P. D'Arrigo, F. Diego, D.D. Walker, *Active Compensation of Flexure on the WHT ISIS Spectrograph*, Mon. Not. R. Astron. Soc., **281**, 673 (1996).

Bibliography

[1] J.H. Crocker, H.C. Ford, G.F. Hartig, R. Jedrzejewski, in *Instrumentation in Astronomy VIII*, Proc. SPIE **2198**, 1170 (1994)

[2] R.C. Smith, *Observational Astrophysics*, Cambridge University Press, Cambridge 1995, p. 16

[3] M. Born, E. Wolf, *Principles of Optics*, 6th Ed. Pergamon Press, Oxford 1980, p. 397

[4] R. Racine, D. Salmon, D. Cowley, J. Sovka, *Publ. Ast. Soc. Pac.* **103**, 1020 (1992)

[5] G. Walker, *Astronomical Observations*, Cambridge University Press, Cambridge 1987

[6] B.V. Barlow, *The Astronomical Telescope*, Wykeham Publications, London 1975, p. 160

[7] D. Enard, A. Marechal, J. Espiard *Rep. Prog. Phys.* **59**, 601 (1996)

[8] B. Olbert, J.R.P. Angel, J.M. Hill, S.F. Hinman, in *Advanced Technology Optical Telescopes V*, Proc. SPIE, **2199**, 144 (1994)

[9] M.J. de Jonge, in *Advanced Technology Optical Telescopes V*, Proc. SPIE, **2199**, 22 (1994)

[10] R. Muller, H. Honess, H. Morian, H. Loch, in *Advanced Technology Optical Telescopes V*, Proc. SPIE, **2199**, 164 (1994)

[11] L. Stepp, E. Huang, in *Advanced Technology Optical Telescopes V*, Proc. SPIE, **2199**, 223 (1994)

[12] E.-D. Knohl, in *Advanced Technology Optical Telescopes V*, Proc. SPIE, **2199**, 271 (1994)

[13] A. Sasaki, I. Mikami, N. Shimoyama, K. Nishiguchi, W.R. Powell, M.J. Edwards, H. Ando, M. Iye, in *Advanced Technology Optical Telescopes V*, Proc. SPIE, **2199**, 156 (1994)

[14] J.E. Nelson, P. Gillingham, in *Advanced Technology Optical Telescopes V*, Proc. SPIE, **2199**, 82 (1994)

[15] P. Lena, *Observational Astrophysics*, Springer-Verlag, Berlin 1988, p. 275

[16] A.P. Thorne, *Spectrophysics*, Chapman and Hall, London 1974, p. 256

[17] L.H. Aller, *Atoms, Stars, and Nebulae*, Cambridge University Press, Cambridge 1991, p. 87

[18] M. Born, E. Wolf, *Principles of Optics*, 6th Ed. Pergamon Press, Oxford 1980, p. 180

[19] A.P. Thorne, *Spectrophysics*, Chapman and Hall, London 1974, p. 136

[20] P. Bosquet, *Spectroscopy and its Instrumentation*, Adam Hilger, London 1971, p. 49

[21] Bausch and Lomb inc. *Diffraction Grating Handbook*, p. 25

[22] D.J. Schroeder, *Astronomical Optics*, Academic Press, San Diego 1987, p. 225

[23] D.J. Schroeder, *Astronomical Optics*, Academic Press, San Diego 1987, p. 248

[24] B. Venkatraman, S.A. Patel, *Structural Mechanics*, McGraw-Hill, 1970, p. 176

[25] J. Vernin, C. Munoz-Tunon, *Astron. Astrophys.*, **284**, 311 (1994)

[26] V. Trimble, in *Publ. Ast. Soc. Pac.*, **107**, 1012 (1995)

[27] J.B. Oke, J.G. Cohen, M. Carr, A. Dingizian, F.H. Harris, R. Lucinio, S. Labreque, W. Schaal, S. Southard, in *Instrumentation in Astronomy VIII*, Proc. SPIE **2198**, 178 (1994)

[28] J.B. Oke, J.G. Cohen, M. Carr, J. Cromer, A. Dingizian, F.H. Harris, S. Labreque, R.Lucinio, W. Schaal, H. Epps, J. Miller *Publ. Ast. Soc. Pac.* **107**, 375 (1995)

[29] S. Vogt, S. Allen, B. Bigelow, L. Bresee, B. Brown, T. Cantrall, A. Conrad, M. Couture, C. Delaney, H. Epps, D. Hilyard, D. Hilyard, E. Horn, N. Jern, D. Kanto, M. Keane, R. Kibrick, J. Lewis, C. Osborne, J. Osborne, G. Pardeilhan, T. Pfister, T. Ricketts, L. Robinson, R. Stover, D. Tucker, J. Ward, M. Wei, in *Instrumentation in Astronomy VIII*, Proc. SPIE **2198**, 362 (1994)

[30] J.K. McCarthy, B.A. Sandiford, D. Boyd, J. Booth, *Publ. Ast. Soc. Pac.* **105**, 881 (1993)

[31] O. Le Fevre, D. Crampton, P. Felenbok, G. Monnet, *Astron. Astrophys.* **282**, 325 (1994)

[32] D.M. Hunten, W.K. Wells, R.A. Brown, N.M. Schneider, R.L. Hilliard, *Publ. Ast. Soc. Pac.* **103**, 1187 (1991)

[33] N. Hubin, L. Noethe, *Science*, **262**, 1390 (1993)

[34] P. Kern, P. Lena, P. Gigan, J-C. Fontanella, G. Rousset, F. Merkle, J-P. Gaffard, in *Active Telescope Systems*, Proc. SPIE, **1114**, 54 (1989)

[35] F. Rigaut, G. Rousset, P. Kern, J.C. Fontanella, J.P. Gaffard, F. Merkle, P. Lena, *Astron. Astrophys.* **250**, 280 (1991)

[36] R.M. Myers, A.P. Doel, C.N. Dunlop, J.V. Major, R.M. Sharples, A.J.A. Vick, in *Active and Adaptive Optics*, Proc. ICO-16 Satellite Conference, Munich 1993, p. 101

[37] R.M. Sharples, A.P. Doel, C.N. Dunlop, J.V. Major, R.M. Myers, A.J.A. Vick, A. Glindemann, in *Active and Adaptive Optics*, Proc. ICO-16 Satellite Conference, Munich 1993, p. 475

[38] L.M. Close, D. Wittman, D.W. McCarthy in *Adaptive Optics in Astronomy*, Proc. SPIE, **2201**, 447 (1994)

[39] E. Pitz, R.R. Rohloff, H. Marth, in *Active and Adaptive Optics*, Proc. ICO-16 Satellite Conference, Munich 1993, p. 267

[40] M. Tarenghi, R.N. Wilson. in *Active Telescope Systems*, Proc. SPIE, **1114**, 302 (1989)

[41] R.N. Wilson, *Contemporary Physics*, **32**, No. 3, 157 (1991)

[42] R.N. Wilson, F. Franza, L. Noethe, G. Andreoni, *J. Mod. Optics*, **38**, No. 2, 219 (1991)

[43] R. Cohen, T. Mast, J. Nelson, in *Advanced Technology Optical Telescopes V*, Proc. SPIE, **2199**, 105 (1994)

[44] J. Nelson, T. Mast in *Progress in Telescope and Instrumentation Technology*, Proc. ESO Conference No. 42, Garching 1992, p. 3

[45] L. Noethe, X. Cui, S. Stanghellini, in *Progress in Telescope and Instrumentation Technology*, Proc. ESO Conference No. 42, Garching 1992, p. 3

[46] F. Bortoletto, D. Fantinel, W. Gallieni, G. Giudici, R. Ragazzoni, R. Tommellieri, P. Vanini, in *Progress in Telescope and Instrumentation Technology*, Proc. ESO Conference No. 42, Garching 1992, p. 3

[47] K.R. Lorell, J-N. Auburn, G.J. Feher, E.O. Perez, D.F. Zacharie, P.J. Reshatoff, in *Adaptive Optics in Astronomy*, Proc. SPIE, **2201**, 821 (1994)

[48] L. Germann, J. Braccio, *Opt. Eng.* **29** (11), 1351 (1990)

[49] G.C. Loney, *Opt. Eng.* **29** (11), 1360 (1990)

[50] R. Bhatia, in *Instrumentation in Astronomy VIII*, Proc. SPIE, **2198**, 554 (1994)

[51] J. Allington-Smith, T. Davidge, in *Instrumentation in Astronomy VIII*, Proc. SPIE, **2198**, 201 (1994)

[52] D.D. Walker, M. Dryburgh, B.C. Bigelow, *Control of Flexure in Gemini Instrumentation*, Design study report commissioned by th Gemini Tucson office, November 1992

[53] Gemini Multi-Object Spectrograph, *Conceptual Design Review*, 6-7 June 1995, Durham.

[54] D.J. Cowley, E. James, "DEIMOS: A Wide-Field, Faint-Object Spectrograph" in *Optical Telescopes of Today and Tomorrow*, ESO/SPIE Symposium, Landskrona, Sweden, May-June 1996, in press.

[55] B.M. Sutin, "ESI: A New Spectrograph for the Keck II Telescope" in *Optical Telescopes of Today and Tomorrow*, ESO/SPIE Symposium, Landskrona, Sweden, May-June 1996, in press.

[56] U. Munari, M.G. Lattanzi, *Publ. Ast. Soc. Pac.* **104**, 121 (1992)

[57] J. Meaburn, *Detection and Spectrometry of Faint Light*, D. Reidel Publishing Company, Dodrecht, Holland, 1976, p. 3.

[58] R.E.S. Clegg, D. Carter, P.A. Charles, J.S.B. Dick, C.R. Jenkins, D.L. King, R.A. Laing, *ISIS Astronomers' Guide*, User Manual No. XXII, August 1992

[59] L.J. Smith, University College London, *personal communication.*

[60] Gemini Science Requirements, *Version* 1.1 (Nov. 1992)

[61] J.R. Christman, *Fundamentals of Solid State Physics*, John Wiley & Sons, New York, 1988, p. 304

[62] W.G. Cady, *Piezoelectricity*, McGraw-Hill, New York, 1946

[63] M.A. Ealey, in *Active and Adaptive Optical Components*, Proc. SPIE **1543**, 346 (1991)

[64] E. Fatuzzo, W.J. Merz, *Ferroelectricity*, North Holland, Amsterdam, 1967

[65] Ferroelectrics (*Encyclopedia of Chemical Technology*) in C.Z. Rosen, B.V. Hiremath, R. Newnham, *Piezoelectricity*, 1992

[66] B. Jaffe, W.R. Cook, H. Jaffe, *Piezoelectric Ceramics*, Academic Press, London, 1971, p. 16.

[67] IEEE Standard on Piezoelectricity, in C.Z. Rosen, B.V. Hiremath, R. Newnham, *Piezoelectricity*, 1992

[68] T. Ikeda, *Fundamentals of Piezoelectricity*, Oxford Science Publications, 1990

[69] Burleigh Instruments (UK) Ltd, 9 Allied Business Centre, Coldharbour Lane, Harpenden, Herts. AL5 4UT, *The Piezo Book*

[70] Piezo System Inc, 186 Massacusetts Avenue, Cambridge Massachusetts 02139 *Product Catalog.*

[71] S. Chikazumi, *Physics of Magnetism*, John Wiley & Sons, New York, 1964

[72] S. Krupička, J. Šternberk, *Elements of Theoretical Magnetism*, Academia, Prague, 1968

[73] R. Greenough, University of Hull, *Personal Communication.*

[74] A.E. Clark, in *Active and Adaptive Optical Components*, Proc. SPIE **1543**, 374 (1991)

[75] Physik Instrumente GmbH, Siemensstrasse 13-15, W-7517 Waldbronn, Germany, *Piezo Guide.*

[76] S. Takahashi, *Ceramic Bulletin*, **65**, 1156 (1986)

[77] C.D. Swift et al. in *Active and Adaptive Optical Components*, Proc. SPIE **1543**, 107 (1991)

[78] M.A. Ealey, J.A. Wellmann in *Active and Adaptive Optical Components*, Proc. SPIE **1543**, 36 (1991)

[79] K. Uchino, S. Nomura, L.E. Cross, R.E. Newnham, S.J. Jang, *J. Mater. Sci.* **16**, 569 (1981)

[80] K. Uchino, L.E. Cross, *Jpn. J. Appl. Phys.* **19**, L171 (1980)

[81] N. Uchida, T. Ikeda, *Jpn. J. Appl. Phys.* **6**, 1079, (1967)

[82] J. Galvagni, B. Rawal in *Active and Adaptive Optical Components*, Proc. SPIE **1543**, 296 (1991)

[83] M.J. Goodfriend, K.M. Shoop in *Active and Adaptive Optical Components*, Proc. SPIE **1543**, 301 (1991)

[84] H. Marth, M. Donat, C. Polhammer, in *Active and Adaptive Optical Components*, Proc. SPIE **1543**, 248 (1991)

[85] G.H. Blackwood, P.A. Davis, M.A. Ealey in *Active and Adaptive Optical Components*, Proc. SPIE **1543**, 422 (1991)

[86] A.E. Clark, J.P. Teter, M. Wun-Fogle, *J. Appl. Phys.* **69**, 5771 (1991)

[87] E.H. Anderson, D.M. Moore, J.L. Fanson, M.A. Ealey, *Opt. Engr.* **29**, (11), 1333 (1990)

[88] M.A. Ealey, P.A. Davis, *Opt. Engr.* **29**, (11), 1373 (1990)

[89] C.G. O'Neill, C.H. O'Neill, *Opt. Engr.* **29**, (11), 1383 (1990)

[90] M. Goodfriend, *Machine Design*, **63**, No. 6, 147 (1991)

[91] Bei Motion Systems Co, 150 Vallecitos de Oro, San Marcos, CA 92069, USA, *Applications Guide.*

[92] R. Loxton, P. Pope, *Instrumentation*, The Open University Press, 1986, p. 290

[93] J.K. Thornley, M.E. Preston, T.G. King, *I Mech E*, Eurotech Direct-Machine Systems, paper C414/057, p. 115 (1991)

[94] T.G. King, M.E. Preston, B.J.M. Murphy, D.S. Cannell, *Precision Engineering*, **12**, No. 3, 131 (1990)

[95] Dr. Lutz Pickelmann Piezomechanik Optik, 13 St.-Cajetan Strasse, D-8000 Munich, Germany, *Product Catalog.*

[96] T. Tanimoto, K. Okazaki, *Jpn. J. Appl. Phys.* 1, **30**, No. 9B, 2264 (1991)

[97] Queensgate Instruments Ltd. Silwood Park, Ascot, Berkshire SL5 7PW, England. *Product Catalog.*

[98] G. Ames, Kaman Scientific Corp. 1500 Garden of the Gods Rd. Colorado Springs, CO 80993-7463, USA, *Personal Communication.*

[99] D. Carter, C.R. Benn, R.G.M. Rutten, J.M. Breare, P.J. Rudd, D.L. King, R.E.S. Clegg, V.S. Dhillon, S. Arribas, J-L. Rasilla, A. Garcia, C.R. Jenkins, P.A. Charles, *WHT-ISIS Users' Manual*, User Manual No. XXIV, October 1993

[100] S. Worswick, Royal Greenwich Observatory, Madingley Rd. Cambridge, U.K. *Personal Communication*

[101] R.G.M. Rutten, Isaac Newton Group of telescopes, Roque de los Muchachos, La Palma, Canary Islands, *Personal Communication*

[102] J. Allington-Smith, M. Breare, R. Ellis, D. Gellatly, K. Glazebrook, P. Jorden, J. MacLean, P. Oates, G. Shaw, N. Tanvir, K. Taylor, P. Taylor, J. Webster, S. Worswick, *Publ. Ast. Soc. Pac.* **106**, 983 (1994)

[103] E. Kibblewhite, M.F. Smutko, F. Shi in *Adaptive Optics in Astronomy*, Proc. SPIE, **2201**, 754 (1994)

[104] F. Gillett, Gemini Project Scientist, in *Gemini Newsletter*, No. 10, June 1995, p. 5.

[105] W.M. Smart, *Spherical Astronomy*, Cambridge University Press, Cambridge 1971, p. 25-57

[106] P. Wallace, *Starlink Project Manager*, Rutherford Appleton Laboratory, Chilton, Oxfordshire OX11 0QX, *Personal Communication.*

[107] M. Breare, Physics Dept. Durham University, South Road, Durham, U.K. and Royal Greenwich Observatory, Madingley Road, Cambridge, U.K. *Personal Communication.*

[108] Physik Instrumente GmbH, Siemensstrasse 13-15, W-7517 Waldbronn, Germany, *Personal Communication.*

[109] Schott Glass Technologies Inc., 400 York Avenue, Duryea, PA 18642, *ZERODUR Technical Brochure.*

[110] G.W.C. Kaye, T.H. Laby, *Tables of Physical and Chemical Constants*, 14th edition, Longman Ltd, London 1985

[111] Physik Instrumente GmbH, Siemensstrasse 13-15, W-7517 Waldbronn, Germany, Piezo Guide, Micropositioning Catalogue.

[112] Hottinger Baldwin Messtechnik UK, Harrow Weald Lodge, 92 Uxbridge Road, Harrow, Middlesex HA3 6BZ, *Personal Communication.*

[113] Physik Instrumente GmbH, Siemensstrasse 13-15, W-7517 Waldbronn, Germany, *PZT Performance Test Document* PI 40556, October 1994.

[114] P.B. Lindley, *Engineering Design with Natural Rubber*, NR Technical Bulletin, Malaysian Rubber Producers' Research Association, Hertford 1978.

[115] The Design of the Keck Observatory and Telescope, *Keck Observatory Report No. 90*, J.E. Nelson, T.S. Mast and S.M. Faber editors, January 1985, p. 8-33.

[116] M. Born, E. Wolf, *Principles of Optics*, 5th ed. Oxford, Pergamon, 1975, p. 289

[117] A. Yariv, *Optical Electronics*, Holt-Saunders ed., Tokyo 1985, p. 30.

[118] M.L. Boas, *Mathematical Methods in the Physical Sciences*, J. Wiley & Sons, New York 1983, p. 734

[119] W.J. Smith, *Modern Optical Engineering*, McGraw-Hill, New York 1990, p. 502

[120] R.M. Scott, *Appl. Opt.* **8** No. 3, 531 (1969).

[121] A.H. Shoemaker, M.R.V.K. Murty *Appl. Opt.* **5** No. 4, 603 (1966).

[122] D. Malacara, *Optical Shop Testing*, J. Wiley & Sons, New York 1978

[123] HROS Design Workshop, London 6-9 November 1995.

[124] S. Ramsey-Howat, Royal Observatory Edimburgh, Blackford Hill, Edimburgh EH9 3HJ, *personal communication.*