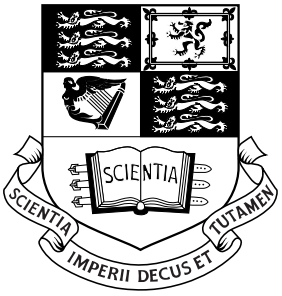
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

UNIVERSITY OF LONDON



Low-cost Bimorph Mirrors in Adaptive Optics

April, 1999

Edric Mark Ellis

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Supervised by Professor J C Dainty, Applied Optics, Department of Physics

Of making books there is no end;
and much study is a weariness of the flesh.

Ecclesiastes, chapter 12, verse 12

Abstract

This thesis describes the modelling and manufacture of bimorph mirrors as well as an example of their application. A general review of adaptive optics is provided, paying particular attention to those technologies which would allow adaptive optics to be employed in non-astronomical situations. The first stage of the theoretical analysis of the bimorph mirror involves developing a mechanical model of the laminated structure. This is carried out using standard thin plate laminate theory; this results in expressions for the overall sensitivity of a bimorph mirror. The mechanical model also leads to an equation describing the response of the surface of the mirror to a given applied voltage. An analytical solution is developed, and its implementation described. Using this analytical solution, the performance of a bimorph mirror under various operating conditions is analysed. Particular attention is paid to the case of correcting phase fluctuations that have a Kolmogorov spectrum. The process of manufacturing bimorph mirrors is described in some detail, and results of this procedure are presented. Results from various bimorph mirrors are analysed, and the comparisons with the theory are presented. A possible use of bimorph mirrors in laser scanner systems is described. A simple demonstration system is described, and results from its operation are presented. A detailed design for a laser printer using adaptive optics is presented, along with the associated deformable mirror requirements.

Acknowledgements

The work that comprises this thesis would not have been completed were it not for the efforts and contributions of a number of third parties. Firstly, I would like to thank Chris Dainty for offering me the opportunity to embark on a PhD course, and for giving me the guidance to help me see it through to the end. Chris's efforts also ensured that I was well funded during (and after) my PhD course. During the course of this work, funding has been received from several sources: EPSRC and Pilkington were my CASE sponsors throughout the course of my PhD. EPSRC also provided some of the money used in the building of practical devices. During the last 6 months of my time at Imperial, I was funded on a DERA project. I would like to thank Alan Greenaway at DERA Malvern for his part in arranging that funding, and his interest in my work. Thanks also to my various industrial supervisors at Pilkington: Steve Collier, Andrew Ramsbottom and Paul Chandley, who made a vital contribution to this work.

The practical aspect of this thesis would have been unthinkable were it not for the fine optical and mechanical workshop in the Applied Optics department. For all his efforts in polishing the assorted pieces of glass and piezo that I gave him, I wish to thank Martin Dowman—his contribution was fundamental to the success of the practical devices. Thanks also to Paul Brown, Dave Mountain and John Bean who bravely attempted to interpret my sketches, and provided me with many excellent mechanical components which enabled me to complete a practical campaign.

During my time at Imperial, I have made many good friends whose company I have been priveliged to enjoy: John Quartel who was foolish enough to share a flat with me for 3 years, and with whom I have had many late night philosophical discussions on diverse and ever fascinating topics. The eternal **Dr.** Víctor Ruiz Cortés, whose generosity of spirit has done much to make the Applied Optics department a pleasant working environment. Luis Diaz Santana Haro, who has also shared a flat with me for the last 2 years, and has been kind enough to shine a bright laser into my eye. The various residents of 612 deserve thanks for putting up with me: Tom Nicholls, Robin Rutherford, Liam Holohan, Valerie Peuré, Florence Rigal, Fredérique Decortiat, David Catlin. Thanks also to the other people in Applied Optics with whom it has been a pleasure to work: Ian Munro, Carl Paterson and Nick Wooder. Thanks to Mathias RIDOU for teaching me a variety of amusing dairy product related songs. Thanks to Chris and Heather for many nights sampling the beers at BierRex, and a great week on the boat.

I would like to thank my parents for being incredibly supportive and encouraging throughout. My sister Rachel also did her bit to keep my spirits up by sending me postcards from various exotic locations telling me what a wonderful time she was having.

Finally, I would like to thank Melanie, whose love and unique brand of cajoling encouragement helped me throughout my writing up, and who gave me a reason to get to the end.

Contents

List of Figures

List of Tables

1 Introduction

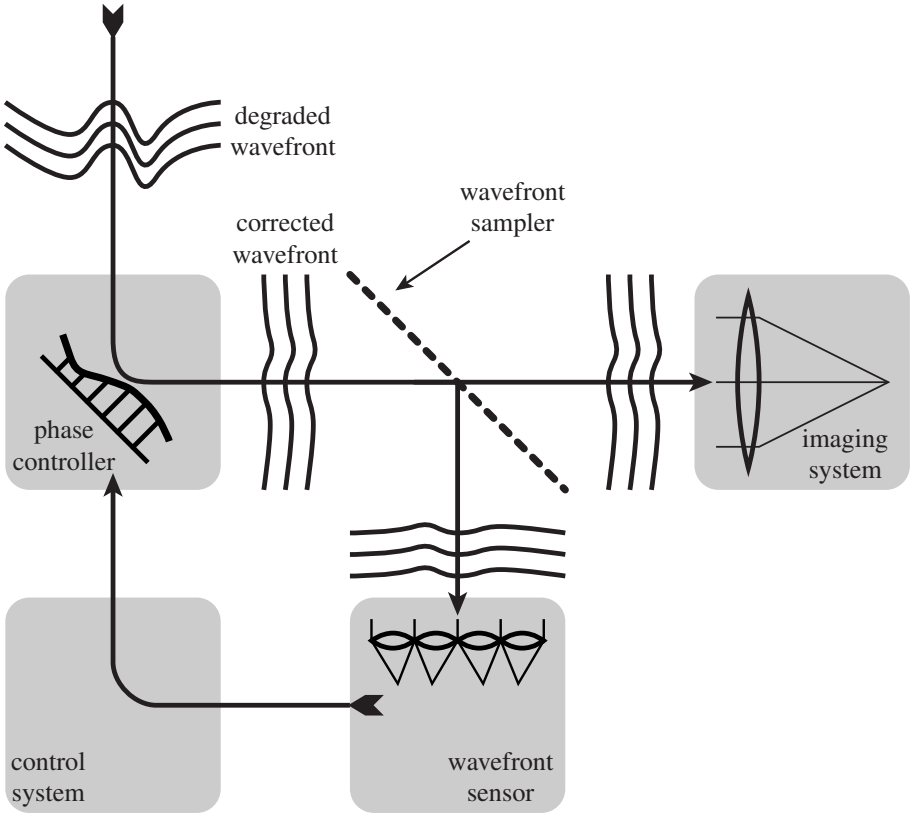
Adaptive optics (AO) is a discipline that was first proposed by Babcock in 1953 [1] as a way of improving the imaging properties of astronomical telescopes. The deleterious effect of the earth's atmosphere on the image produced by a telescope has been well known since the time of Newton, who in 1730 suggested that a solution may be to build a telescope "on the tops of the highest Mountains". This approach has reached its apotheosis with the construction of the Hubble Space Telescope (HST), entirely removed from the earth's atmosphere. Babcock's proposal, which has now been realised, does not incur the exceptional expense of a space-based telescope, and can be expected to produce higher resolution images than the HST. The method consists essentially of *sensing* the phase distortions introduced by the atmosphere, and then applying an appropriate *correction* to the phase, which results in a diffraction-limited image. In the absence of atmospheric influence, the resolution of a well made telescope is said to be "diffraction-limited", this is a fundamental limit on the resolution of a telescope, and depends only on the wavelength of interest and the diameter of the telescope's primary mirror. However, uncorrected atmospheric turbulence makes the imaging properties of a large telescope the same as those of a telescope of diameter around 10cm. Therefore, if AO can be used to negate the effects of the atmosphere, then by using it on a telescope larger than the HST, higher resolution imaging should be the result.

The hardware demands of the astronomical imaging problem have been considered in great detail since AO was first suggested. Advances have been made in the two essentially optical components of an AO system, namely the *deformable mirror* whose surface shape changes in a controllable way in response to a drive signal, and the *wavefront sensor* which measures the atmospherically introduced phase fluctuations across the optical beam in real time. The third key component in the AO system is the control computer; the increase in computer power in recent decades has greatly increased the scope of AO by permitting more and more channels to be run in parallel, and at higher speeds. The basic interaction of these key components is shown diagramatically in figure 1.1. This is the layout that is used in conventional astronomical adaptive optics.

1.1 Low Cost Adaptive Optics

Now that Adaptive Optics is becoming a relatively mature technology with proven benefits, there is growing interest in applying the developments in other areas of optics. The majority of these applications do not have the extremely large budgets available to astronomers building an AO system for a telescope. Therefore, the scope of the AO system must be reduced in the sense that the number of control channels of the deformable mirror and the number of wavefront sensor channels must be reduced. Reducing the number of channels in each of these subsytems has two main advantages. Firstly, the wavefront sensor and deformable mirror will almost inevitably be easier to construct (and hence less expensive) as the number of channels is reduced. Secondly, the amount of data that must be processed decreases as the number of sensor and control channels decreases. The reduction in control channels inevitably has an impact on the range of deformations that can be adequately

Figure 1.1 A simple conventional Adaptive Optics system



corrected by a low-cost system. For a potential low-cost system to be viable, the question of whether this low-order correction is adequate must be answered. Some of the situations where low order correction are viable are outlined below.

1.1.1 Rapid focussing sytems

This is one of the simplest applications using AO technology that can be conceived. There are many fields in which a rapidly varying focal length mirror could be used. One example, discussed in detail in chapter 6, is a laser scanner system. There are also applications in laser machining, where a beam must be kept in focus on a workpiece [2]. Imaging systems may also benefit from having a very rapidly (compared to a conventional automatic system) focal length element [3].

1.1.2 Small telescopes

The analysis presented in chapter 4 shows that a bimorph mirror with around 40 channels can provide adequate correction to correct the kind of phase fluctuations found in a telescope with a diameter of around 1m, in the visible region of the spectrum.

1.1.3 Imaging through turbulence

The general problem of imaging through turbulence can be somewhat more complicated than the astronomical case. When a star is imaged through the atmosphere, the region in which the phase distortions are introduced is rather small compared to the path length. This fact makes the problem tractable. Even in astronomy, the thickness of the atmosphere introduces a limit on the field of view of an AO system: this field of view cannot be any larger than the “isoplanatic patch”—the region

over which the atmospherically introduced aberrations are approximately constant. The size of the isoplanatic patch is determined partly by the strength of the phase fluctuations, but the distribution of the turbulence along the optical path has a much greater influence. The nearer the turbulence is to the source, the smaller the isoplanatic patch. Therefore, the problem of imaging ground-based objects is actually a considerably more difficult problem because of the desire for a larger field of view than is common in astronomy coupled with the fact that the object is likely to be immersed in the turbulent medium.

In some cases, the phase distortions may be known to be introduced in only one plane. In this case, the isoplanatism problem can be eliminated by making the deformable mirror element optically conjugate to the plane introducing the aberrations. This fact may enable AO to be applied succesfully in imaging the living human retina, for example, since it is believed that most of the significant aberrations are introduced by the cornea.

1.1.4 Laser correction

Away from the field of astronomy, the largest AO effort has been in improving the performance of laser systems. Adaptive techniques have been used to attempt to correct phase distortions introduced into a laser cavity [4]. The success of such schemes have been at best rather limited. With the advent of very large lasers, such as Vulcan in the UK, there is renewed interest in tackling the problem of intracavity AO. In the case of Vulcan, it is believed that the intensity at a target may be increased by up to a factor of 10 if the residual aberrations can be effectively cancelled. Also, the type of AO that must be used employed in such large laser systems is, in fact, simpler in principle than that that is required inside a typical laser cavity: in very large lasers, the light essentially traverses the "cavity" only once, and the situation shares several of the properties of the astronomical case. If AO is to be used in the cavity of a normal laser where the light undergoes many traversals of the cavity, then controlling the deformable mirror is a complex issue due to the way the transverse modes build up in such a device.

1.2 Thesis Synopsis

The main body of this work consists of a theoretical and practical investigation of the bimorph deformable mirror. Following on from this part of the work, the practical devices were then applied to the problem of using AO in a laser scanner system. A plan of the rest of this work, including comments on the originality of the content, is given below.

- **Chapter 2** presents a review of relevant optics theory, a review of available deformable mirror concepts and technologies, a description of various other AO components, a summary of modal AO techniques, a description of selected AO implementations and a review of the bimorph mirror.

- **Chapter 3** develops the theory of the bimorph mirror from mechanical principles. The analysis leading to the biharmonic equation is based on standard texts. The Green's function used in the solution is due to Timoshenko [5]; the analytical integration of this is original. Other solutions to the biharmonic equation for the bimorph problem do exist; this solution is presented in considerably more detail than those in the open literature.

- **Chapter 4** uses the theoretical solution derived in chapter 3 to perform numerical simulations of an AO system. Particular attention is given to simulations involving Kolmogorov

turbulence, the standard model of atmospherically induced phase fluctuations. The results presented in this chapter are entirely original.

- **Chapter 5** presents results obtained from practical bimorph mirrors. A detailed bimorph construction scheme is laid out. Although bimorph mirrors have been known for some 20 years, details of construction techniques have been scarce; the construction technique used was arrived at with some assistance from Alexis Kudryashov, who made helpful comments at an early stage of the construction programme. The use of a strain gauge to reduce hysteresis effects in a bimorph mirror is believed to be original. The use of pyrex as a substrate to minimise thermal effects is thought to be original. The level of detail of the performance evaluation is somewhat greater (as far as the quasi-static performance is concerned) than any published in the open literature.

- **Chapter 6** develops the concept of an AO-assisted laser scanner. This concept was conceived by the author, in conjunction with Andrew Ramsbottom and Paul Chandley at Pilkington. An analysis of various aspects of AO-assisted laser scanning is presented. Results from a simple practical laser scanner are presented. Also, a detailed optical design for a laser printer type scanner is presented. All the above items are believed to be completely original.

2 Principles of Adaptive Optics

2.1 Diffraction and Image Formation

The behaviour of a light beam can be described in several different ways. The simplest description considers light to be composed of rays which propagate in straight lines, unless bent by refractive index changes, in accordance with Snell's law. This approach fails to predict diffraction phenonmena such as Young's fringes. The next level of approximation is to consider light to be a wave phenomenon, in which the light waves are transverse and scalar. This is known as the scalar diffraction theory. In this level of approximation, we must assume that the angles at which beams of light are interacting with each other are not too great. If the angles of interaction are large, then the full vector-field solution to Maxwell's equations is required. Fortunately, for all situations encountered in this thesis, the scalar approximation is quite adequate. From Maxwell's equations, we can derive a scalar wave equation which describes the propagation of a beam of light.

A computationally efficient technique for performing diffraction calculations was reported by Sziklas and Siegman [6] in which the *fast Fourier transform* (FFT) is employed. To show how this is achieved, consider propagating an arbitrary light field from a plane $z = z_1$ to $z = z_2$, and using the geometry defined in figure 2.1, the *Rayleigh-Sommerfeld diffraction formula* can be written [7]:

$$\begin{aligned} u_2(x_2,y_2) &= \iint_{-\infty}^{\infty} u_1(x_1,y_1)\left(\frac{z_{12}}{j\lambda r_{12}^2}\right)\exp(-jkr_{12})\,\mathrm{d}x_1\,\mathrm{d}y_1 \\ &= \iint_{-\infty}^{\infty} u_1(x_1,y_1) \\ &\qquad \times \frac{\exp\left(-jkz_{12}\left[1+\left(\frac{x_2-x_1}{z_{12}}\right)^2+\left(\frac{y_2-y_1}{z_{12}}\right)^2\right]^{1/2}\right)}{k\lambda z_{12}\left[1+\left(\frac{x_2-x_1}{z_{12}}\right)^2+\left(\frac{y_2-y_1}{z_{12}}\right)^2\right]}\,\mathrm{d}x_1\,\mathrm{d}y_1\,, \end{aligned} \tag{2.1}$$

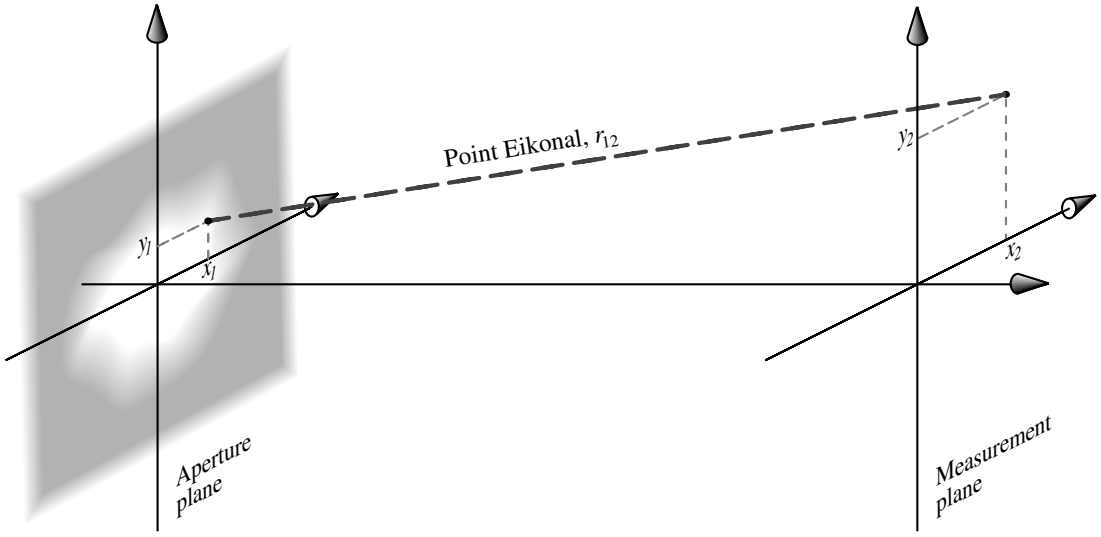
where $j=\sqrt{-1}$, and the functions u_i are the complex scalar-field amplitudes in the appropriate planes, and the appropriate expression for r_{12} has been substituted. We note that this expression has the form of a convolution, so if we define

$$h_{12}(x,y) = \frac{\exp\left\{-jkz_{12}\left[1+\left(\frac{x}{z_{12}}\right)^2+\left(\frac{y}{z_{12}}\right)^2\right]^{1/2}\right\}}{j\lambda z_{12}\left[1+\left(\frac{x}{z_{12}}\right)^2+\left(\frac{y}{z_{12}}\right)^2\right]}, \tag{2.2}$$

then we can write

$$u_2(x_2,y_2) = u_1(x_2,y_2)\otimes h_{12}(x_2,y_2)\,. \tag{2.3}$$

Figure 2.1 Geometry for the diffraction integral



The Fourier transform of Equation 2.2 can be found analytically, and is

$$H_{12}(\xi,\eta) = \exp\left(-jkz_{12}\left[1-\lambda^2(\xi^2+\eta^2)\right]^{1/2}\right), \tag{2.4}$$

so that finally we can write

$$U_2(\xi,\eta) = U_1(\xi,\eta)H_{12}(\xi,\eta). \tag{2.5}$$

This procedure lends itself well to numerical calculation because of the speed of the FFT sections of the calculation.

2.1.1 General FFT propagation

The above section shows how it is possible to perform diffraction propagations using two FFTs and a transfer function for the case of free space propagation. However, it is possible to use FFT propagation to perform propagation through any arbitrary $\mathcal{ABCD}$ system[1].

Firstly, we note that an $\mathcal{ABCD}$ matrix only has three independent elements because the determinant of the matrix is required to be equal to unity. However, for convenience, we make our equivalent system from four elements, *viz.*:

1. a lens of focal length f_1
2. a relay magnifier of magnification M
3. a free space of length L
4. a lens of focal length f_2.

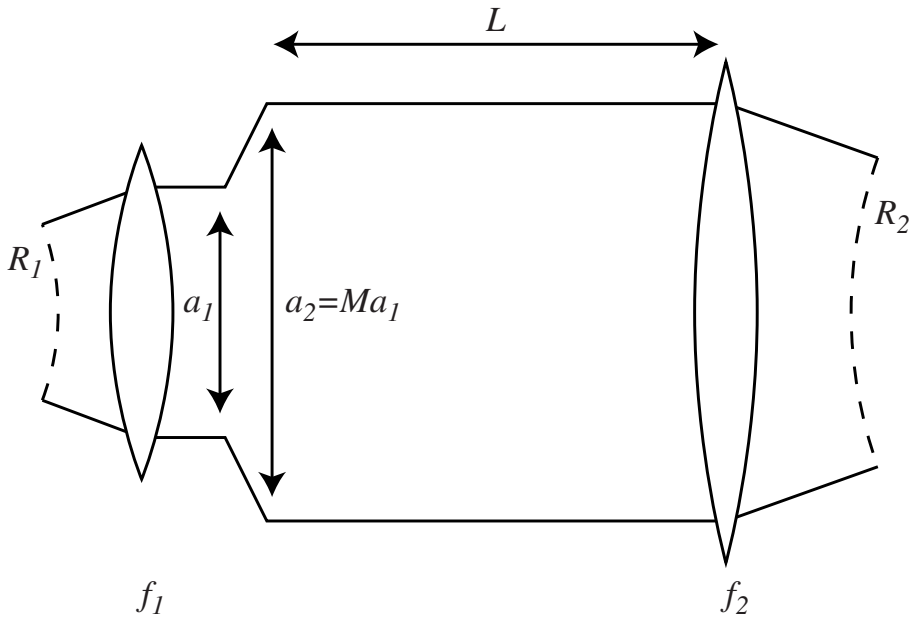
This equivalent system is shown in Figure 2.2. We need to find the $\mathcal{ABCD}$ matrix for this system[2]:

$$\begin{bmatrix}\mathcal{A} & \mathcal{B}\\ \mathcal{C} & \mathcal{D}\end{bmatrix} = \begin{bmatrix}1 & 0\\ -1/f_2 & 1\end{bmatrix}\begin{bmatrix}1 & L\\ 0 & 1\end{bmatrix}\begin{bmatrix}M & 0\\ 0 & 1/M\end{bmatrix}\begin{bmatrix}1 & 0\\ -1/f_1 & 1\end{bmatrix} \tag{2.6}$$

[1] For an introduction to $\mathcal{ABCD}$ systems, see, for example [8].

[2] N.B. In Siegman's book [9], the matrix used for the magnification stages is wrong, but the resultant is correct

Figure 2.2 An equivalent system to a general $\mathcal{ABCD}$ system



so that, in terms of the original $\mathcal{ABCD}$ matrix,

$$\mathcal{A} = M - \frac{L}{Mf_1} \tag{2.7}$$

$$\mathcal{B} = \frac{L}{M} \tag{2.8}$$

$$\mathcal{D} = -\frac{L}{f_2 M} + \frac{1}{M}. \tag{2.9}$$

The reason we allow ourselves an additional degree of freedom is so that now, knowing the overall curvature of the input field, we can correct this using the first lens. This stage defines f_1. Now, we can find the other parameters, L, M, and f_2:

$$M = \mathcal{A} + \mathcal{B}f_1 \tag{2.10}$$

$$L = \mathcal{B}M \tag{2.11}$$

$$f_2 = \frac{L}{1 - M\mathcal{D}}. \tag{2.12}$$

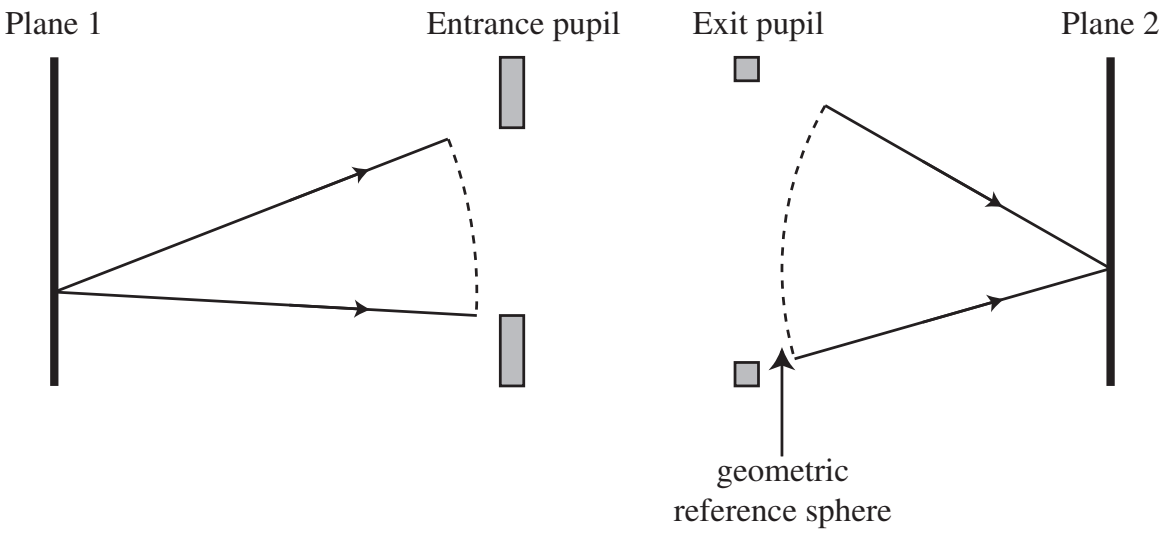
We have thus successfully reduced the problem of diffraction propagation through any system into one of scaling the coordinates and using computationally efficient FFT propagation.

2.1.2 Image formation

The theory of image formation is well described in many standard texts (e.g. reference 10); here it will suffice to summarise the main results. Figure 2.3 shows a generalised imaging system. If the object in plane 1 radiates coherent, monochromatic light (e.g. a partially transmissive plate illuminated from behind by a collimated laser beam), then the complex amplitude in plane 2, $u_2(x_2, y_2)$ is related to the complex amplitude in plane 1, $u_1(x_1, y_1)$ by:

$$u_2(x_2, y_2) = \iint u_1(x_1, y_1) K(x_1, y_1; x_2, y_2)\, \mathrm{d}x_1\, \mathrm{d}y_1 \tag{2.13}$$

Figure 2.3 Planes involved in image formation



where the integration is taken over the whole of plane 1. We can think of the function $K(x_1, y_1; x_2, y_2)$ as relating a point in plane 1 to a distribution of energy in plane 2: it is the impulse response of the optical system. These distributions then interfere with each other to form the whole image. The function $K(x_1, y_1; x_2, y_2)$ contains all the information about the image forming system, including aberrations, and how they vary with position in the field of view. For many optical systems, the impulse response (or *point-spread function*) is a constant over the field of view, and hence we can replace $K(x_1, y_1; x_2, y_2)$ by the simplified $K(x_2 - x_1, y_2 - y_1)$. This situation is known as *isoplanatism*, and in this case, the image formation clearly becomes a simple convolution of the object distribution with the impulse response. Furthermore, it can be shown that $K(x, y)$ is the Fourier transform of the optical "pupil function". This pupil function is just the complex amplitude in the exit pupil of the optical system due to a single point object, with an amount of spherical shape subtracted. The radius of the sphere that must be subtracted is calculated from geometrical considerations: this sphere is centred at the geometric focal point of the system.

In the case of an incoherent object in plane 1, we are interested in the relationship between the intensity there, $I_1(x_1, y_1)$ and that produced in plane 2, $I_2(x_2, y_2)$. This relationship can be found by recalling that each point in the object will independently form an image according to equation 2.13, but these will not interfere with each other. So, by taking the modulus squared of the impulse response, we can immediately say that:

$$I_2(x_2, y_2) = \iint I_1(x_1, y_1) |K(x_1, y_1; x_2, y_2)|^2 \, \mathrm{d}x \, \mathrm{d}y \tag{2.14}$$

Thus incoherent imaging is again a linear Fourier-filtering operation in the isoplanatic case. In this case, the filter is the *optical transfer function* (OTF), which is the Fourier transform of $|K(x, y)|^2$. Since K is the Fourier transform of the optical pupil function, it follows that the OTF is directly related to the auto-correlation of the pupil function.

As mentioned above, it is possible to include the effects of aberrations of the optical system into the impulse response function K. These aberrations can be inherent in the detailed design of the optical system, or they may arise from atmospheric disturbances. In either case, the treatment is the same.

2.1.3 The Strehl ratio

A common measure of imaging quality is the "Strehl ratio". This ratio describes how well a real optical system performs compared to an ideal one. The Strehl ratio is defined as the ratio of the real on-axis intensity to the ideal on-axis intensity that would be produced by an unaberrated (i.e. diffraction limited) system. In the limit of small aberrations, it is found that the precise form of the aberrations is not important, and an expression for the Strehl ratio can be found simply in terms of the variance of the phase of the aberrated wavefront. Therefore, writing the mean-square deformation, or variance, of the wavefront as:

$$\sigma^2 = \frac{\iint \left(\phi - \overline{\phi}\right)^2 \rho \, d\rho \, d\theta}{\iint \rho \, d\rho \, d\theta} = \overline{\phi^2} - \left(\overline{\phi}\right)^2 \tag{2.15}$$

where ϕ is the wavefront aberration expressed in terms of absolute distance, and the integrals are take over the optical pupil. This leads to Maréchal's expression for the Strehl ratio [10]:

$$S = 1 - \left(\frac{2\pi}{\lambda}\right)^2 \sigma^2 \tag{2.16}$$

where the $2\pi/\lambda$ factor simply rescales the aberrations to give a phase variance. In adaptive optics, it is common to refer to any wavefront variance (such as a fitting error of a deformable mirror) in terms of a Strehl ratio.

2.2 Gaussian Beams

Since the output from a laser often has a Gaussian transverse intensity profile, it is useful to consider how such a beam propagates. If a beam has initially uniform phase and Gaussian intensity profile, it is well known [8] that the form of the intensity profile does not change as the beam propagates, although the phase across the beam changes, and the width of the Gaussian shape changes. A region where the phase across the beam is uniform is known as a "beam waist": from here, the beam diverges in accordance with diffraction. If we consider the beam to be propagating in the z direction, and that the origin of this axis is coincident with the beam waist, then the subsequent spot radius to the point at which the intensity falls to $1/e^2$ of the on-axis value, $w(z)$ varies as:

$$w(z) = w_0 \left(1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2\right)^{1/2}, \tag{2.17}$$

where w_0 is the beam waist radius and λ is the wavelength. Other useful quantities to consider are the radius of curvature of the wavefront, $R(z)$, which is given by:

$$R(z) = z \left(1 + \left(\frac{\pi w_0^2}{\lambda z}\right)^2\right). \tag{2.18}$$

The "confocal beam parameter", z_0, describes the distance along the z axis away from the beam waist dividing the region where the beam behaves in a geometric way (the far-field region) from the region where the beam's behaviour is dominated by diffraction effects (the near-field region). At a distance from the waist of z_0, the beam radius is $\sqrt{2}$ times larger than at the waist; the value of z_0 is:

$$z_0 = \frac{\pi w_0^2}{\lambda}. \tag{2.19}$$

When the beam has propagated a distance $\sqrt{3}z_0$, then the beam radius is twice the beam waist—this is where two beams cease to be resolved, and so the effective depth of focus of a Gaussian beam is $\sqrt{3}z_0$.

A convenient formulation commonly used when dealing with Gaussian beams is to describe them in terms of the "complex curvature", $q(z)$, which is defined as follows:

$$\frac{1}{q} = \frac{1}{R} + \frac{j\lambda}{\pi w^2}, \tag{2.20}$$

where the real part represents the curvature, and the imaginary part measures the beam size. The advantage of this formulation is that the complex curvature can be calculated after a Gaussian beam has passed through an $\mathcal{ABCD}$ system easily:

$$q_2 = \frac{\mathcal{A}q_1 + \mathcal{B}}{\mathcal{C}q_1 + \mathcal{D}}. \tag{2.21}$$

2.3 Adaptive Optics Sytems

As mentioned in the introduction, adaptive optics (AO) is a technique first suggested by Babcock [1] in order to compensate for the effects of atmospheric turbulence on astronomical images. It is in this field that the technique has flourished, and largely due to efforts in this field, the concepts and components have been so developed that they are now finding wider applications. There are three essential components of an AO system: some sort of wavefront modifier (hereafter referred to a "phase controlling device", or PCD), capable of introducing a controlled phase change that varies across the optical pupil in real time to an optical beam; a wavefront sensor (WFS) which can measure in some way the correction to apply to the phase controller; and some intermediate control stage that conditions the output from the wavefront sensor into an appropriate form for the phase controller. Before entering a detailed description of the physical realisations of the various components, it is helpful to describe the overall system concepts in the various forms of AO, and which components are required in each case. Some complete practical systems will be described later in section 2.7.

2.3.1 Conventional AO

The most developed form of an AO system is that which is used in the field of astronomical imaging—this will be referred to as "conventional" AO. This type of system also covers some non-astronomical imaging applications, e.g. some high-power laser correctors operate in a way that is conceptually the same as the astronomical case.

A conventional AO system typically has a randomly distorted incoming optical beam. The goal of the AO system is to correct those distortions in such a way as to produce a high-resolution image. To this end, a system such as that shown in figure 1.1 is employed. In this scheme, the input beam encounters first the phase controller. After the phase correction has been applied, the beam is incident upon a wavefront sampler (usually some sort of beamsplitter). Part of the beam goes on to form the high resolution image, the rest goes to the wavefront sensor. The wavefront sensor measures the *residual* wavefront error—the system is operating in *closed loop*. This means that the combined effect of the distortion and correction is measured, and this is driven towards zero. The output from the wavefront sensor is then manipulated by the control system, and output to the phase controller.

Conventional astronomical AO

Since astronomy is where AO has been most developed, it is appropriate to discuss very briefly the problems typically encountered in this application, since some of them will apply in other non-astronomical applications. When conventional AO is applied to an astronomical imaging system, much attention is required on the subject of wavefront sensing. Most objects of interest in the sky are relatively faint. It has been shown that the number of photons coming from most objects is insufficient to drive a WFS with enough precision to obtain good correction—roughly 100 photons per sensor channel is an approximate criterion [11]; there is no sharp limit, the signal-to-noise ratio gradually becomes lower the fewer photons there are available. Therefore, it is almost always necessary to use an auxilliary source to drive the WFS. This assumes that the light from the so-called "guide star" has suffered the same phase aberrations as has the light from the object of interest. It can be shown that this is not the case, unless both objects lie within the same "isoplanatic patch" [12, 13]; anisoplanatism refers to different point-spread functions caused by light experiencing different aberrations. This patch is very small, typically a few seconds of arc. It is heavily dependent upon the distance between the observer and the turbulence plane: the maximum angular distance θ_0 between a reference star and an object for a single turbulent layer a distance $\bar{h}$ from the observer is given by:

$$\theta_0 \approx 0.3\frac{r_0}{\bar{h}} \tag{2.22}$$

Where r_0 is a measure of the strength of the turbulence (larger r_0 implies less strong turbulence). Thus we see that as the distance $\bar{h}$ increases, or the turbulence strength increases, the angular separation over which we can expect good correction decreases. There are very few astronomical objects bright enough to drive a many-channel AO system, and so the chances of one of these being within the same isoplanatic patch as the object of interest is essentially zero. Therefore, an artificial guide star is necessary. This consists (in one implementation) of a laser resonantly illuminating the layer of sodium that exists about 90km above the earth's surface [14]. This approach raises two significant problems: although it solves the sky coverage problem, it introduces another anisoplanatic effect, known as focal anisoplanatism, since the artificial star is at a height of 90km. There is also the "tilt determination" problem: if the full aperture of the telescope is used to launch the laser beam, then the overall wavefront slope introduced by the atmosphere is not sensed by the artificial guide star. Both of these problems are shown in figure 2.4.

2.3.2 Multidither AO

The principle of multidither AO was first proposed by O'Meara, and he reviewed the technique in reference 15. This type of AO is particularly well suited to an attempt to illuminate an object through an intervening turbulent atmosphere; it will be in this arrangement that the principle will be discussed here.

The general scheme of operation is shown in figure 2.5. The illuminating beam is divided into several regions. Each of these regions is "tagged" by having a small-amplitude, high-frequency phase perturbation known as a "dither" applied to it. Each region has a different frequency dither applied to it. This division and dithering can be achieved by, for example, using a segmented mirror and dithering the separate segments of the mirror to apply the dither signal.

After the beam has been tagged by the dither signals, it passes through a turbulent atmosphere, and arrives at the target. The system relies on a "glint" returning from the target—i.e. a bright reflection. This light then passes back through the turbulence to the launch system. By examining the various frequency components of the intensity of this return beam, the various regions can have

Figure 2.4 Laser guide stars in astronomical adaptive optics

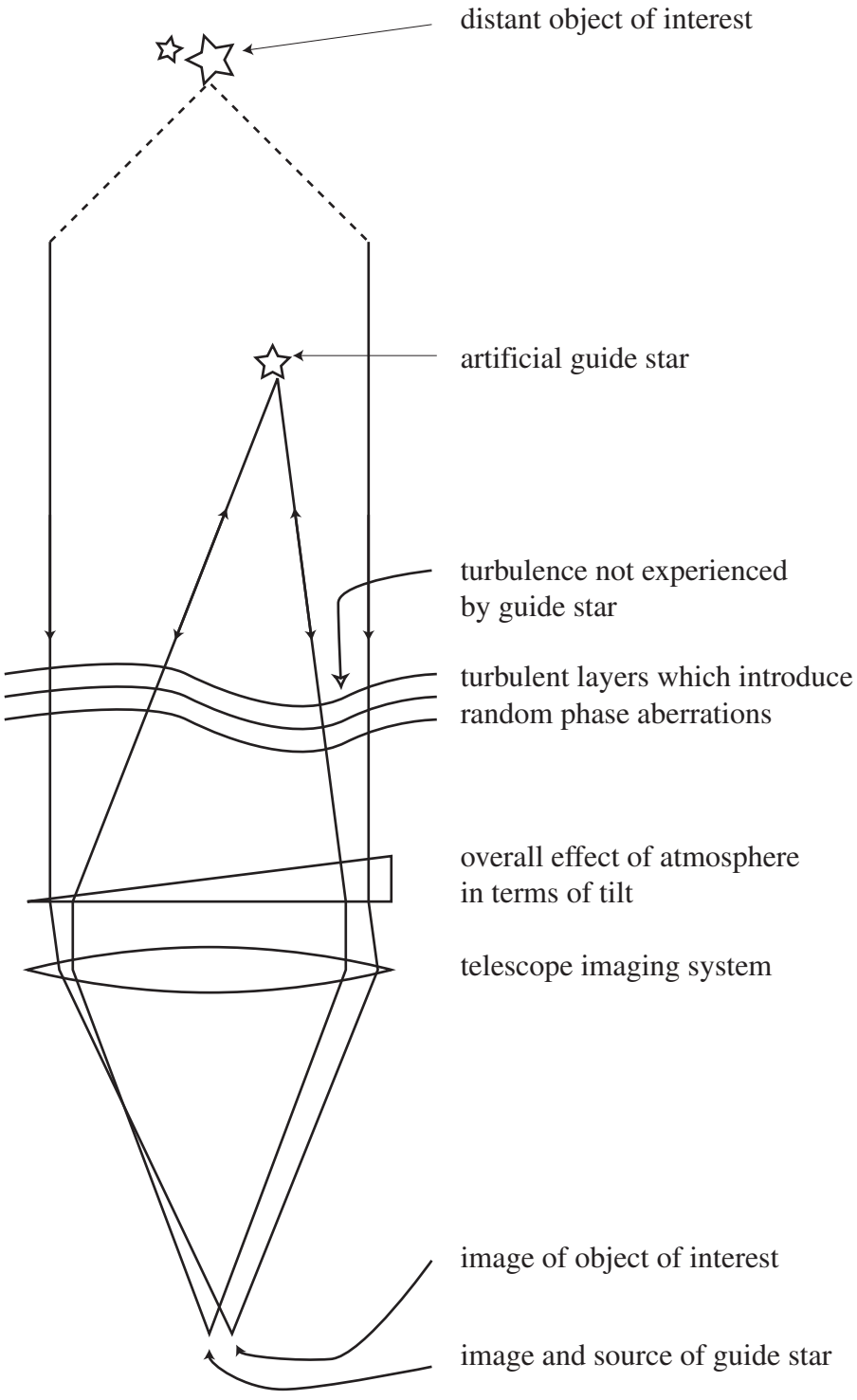
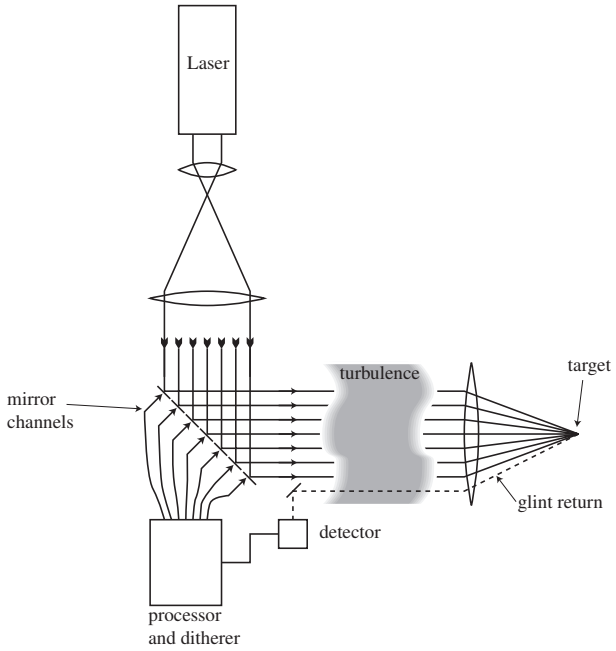


Figure 2.5 Schematic of a energy-projection multidither AO system



there overall phase position adjusted so as to maximise the intensity of this return beam. This correction scheme is essentially a type of "hill climbing" method.

This type of AO system has not flourished, and in fact is now rarely used. One of the main drawbacks of the method is the difficulty in finding and maintaining a reliable glint. Multidither adaptive optics was conceived at a time when performing conventional AO was impractical due to computing limitations—as these limitations have been gradually overcome, the use of multidither systems has declined.

2.3.3 Image sharpness AO

In a system with few degrees of freedom (in the PCD), then an image sharpness AO system may be appropriate. An image sharpening system analyses an image taken through turbulence, and attempts to optimise certain aspects of that image. This technique was first conceived by Muller and Buffington [16] as a means of overcoming the necessity of finding a bright "natural guide star". In this scheme, the adaptive optics system is essentially the same as that shown for conventional AO in figure 1.1, except that there is no WFS arm. Instead, the feedback to the mirror comes directly from the image of the star. Muller and Buffington proposed several measures of image sharpness that may be used, and proved for several of them that the quantity is only maximised if the phase aberrations introduced by the atmosphere are corrected by the PCD. An example of such a measure is:

$$S_1 = \int I^2(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{2.23}$$

where $I(x,y)$ is the intensity of the image at a point (x,y). In a sharpness correction scheme, each channel of the PCD is perturbed in turn, and the effect on the sharpness measure noted. Once the change in sharpness for a given change in a corrector channel is noted, the gradient is then used to maximise the sharpness measure.

More recently, this technique has been applied by Vorontsov *et al.* [17] as a method of correcting for a complicated image, whose phase distortions may not be well sensed by a conventional WFS. The approach used also involved a coherent optical method of estimating the sharpness criterion. A

spatial light modulator (SLM) was used in conjunction with a liquid crystal light valve (LCLV) to imprint a phase distortion on a coherent beam of light, where the phase distortion was proportional to the image-plane intensity. By Fourier transforming this beam (using a lens), and passing it through a rotating piece of frosted glass, a speckle field was created. The image sharpness was found by analysing the power spectrum of the intensity detected at a point in this speckle field.

2.3.4 Beam forming AO

Instead of correcting for degradation of an image, beam forming AO aims to change the form of an optical beam so that at some surface it has an arbitrary transverse intensity profile. The most common application of this is to reshape Gaussian-profiled laser beams to profiles more useful for industrial machining etc. This goal in the past has been achieved using holographic elements [18]. There are two AO techniques that have been developed to produce laser beams with a custom intensity profile.

The approach most similar to the holographic approach is to simply use a deformable mirror to apply the necessary phase change, resulting in the desired intensity profile. This has been demonstrated by Nemoto *et al.* [19], who used a 9-degrees-of-freedom continuous faceplate mirror. The drawback with this approach is that the phase changes required to reshape the beam are relatively large—the deformable mirror used in this study had to have its fine-positioning actuators mounted on micrometers to obtain the overall shape required.

Adding a deformable mirror inside a laser cavity can reshape the output beam profile. This has been done by Van Neste *et al.* [20], who used a diamond-turned graded-phase mirror to produce a "super-Gaussian" output intensity profile. The advantage of this technique is that the phase perturbation required to change the intracavity mode properties is relatively quite small (the mirrors produced for the experimental study had maximum departures from a spherical shape in the region of 2μm). This effect was also observed by Cherezova *et al.* [21], although in a less controlled manner.

2.4 Modal Analysis of AO Systems

One of the most important considerations when designing an AO system is the matter of *observability* and *controllability*. Observability describes how well the WFS can detect what state of the PCD is, and controllability describes which combinations of WFS measurement can be corrected by the PCD. The following discussion will make the preceeding definitions more quantitative; we will use the notation used by Furber and Jordan in reference 22, and we will also follow their reasoning.

Consider a conventional AO system consisting of a wavefront sensor with N_s sensor outputs, and a wavefront controller with N_a actuators (the number of inputs to the wavefront controller). When a signal is applied to the wavefront corrector, this can be expressed as a *control vector*, $\mathbf{x}$ which has N_a components. When a measurement is made with the wavefront sensor, its output is in the form of a vector, $\mathbf{y}$ which has N_s components.

The link between the PCD and the WFS is that the PCD controls the wavefront aberrations of the optical beam, and the WFS measures those aberrations. For each control channel on the PCD, there is a corresponding aberration introduced. That is, if all the elements of $\mathbf{x}$ are zero except for the i^{th}, then the aberration introduced by the PCD will be $Z_i(r,\theta)$. Assuming that the influence functions of the PCD add in a linear fashion we can say that the effect on an optical beam of a control vector $\mathbf{x}$ is:

$$\phi(r,\theta) = \sum_{i=1}^{N_a} x_i Z_i(r,\theta) \tag{2.24}$$

where $\phi(r,\theta)$ is the resulting phase of the optical beam. If the influence functions Z are calculated on a rectangular grid of N_g points, then we can build an influence matrix $\mathbf{Z}$ whose i^{th} column gives the i^{th} influence function. The matrix $\mathbf{Z}$ therefore has dimensions $N_g \times N_a$, and equation 2.24 can be written as:

$$\phi(r,\theta) = \mathbf{Zx} \tag{2.25}$$

Next, we consider the operation of the WFS. When an optical beam is incident, the WFS performs some operation to generate its output vector $\mathbf{y}$. If we call this operator S, then:

$$\mathbf{y} = S(\phi(r,\theta)) \tag{2.26}$$

Neglecting noise and any effects on $\phi(r,\theta)$ other than that of the PCD (for the moment), then we can find the relationship between the WFS and the PCD:

$$\mathbf{y} = S(\mathbf{Zx}) \tag{2.27}$$

If we use a spatially discrete and linear approximation, then the operator S becomes a matrix, and we write:

$$\mathbf{y} = \mathbf{SZx} \tag{2.28}$$

It is usual to combine the effects of the two matrices $\mathbf{S}$ and $\mathbf{Z}$ into what is referred to as the input-output matrix $\mathbf{H}$:

$$\mathbf{y} = \mathbf{Hx} \tag{2.29}$$

In a real AO system, as a calibration, we would normally measure the input-output matrix by applying signals to the actuators one at a time—this procedure picks out columns of the matrix $\mathbf{H}$ which appear as the WFS measurement $\mathbf{y}$. In a real AO system, we do not have direct access to the full matrices $\mathbf{S}$ and $\mathbf{Z}$, all our information is contained within the input-output matrix.

After we have measured the input-output matrix, the next stage is to try an control the WFS measurements using the PCD. To do this, we must invert the relationship given in equation 2.29. Generally, the number of sensors and actuators will not be the same, so we cannot perform a simple inverse. The technique known as singular value decomposition (SVD) gives a "pseudoinverse" matrix which has the property that when multiplied by a given measurement vector $\mathbf{y}$, the result is a control vector $\mathbf{x}$ which minimises the residual error in $\mathbf{y}$ in a least-squares sense. Denoting the pseudoinverse by the superscript $+$, this relationship is:

$$\mathbf{x} = \mathbf{H}^{+}\mathbf{y} \tag{2.30}$$

The SVD procedure does more than simply give us the pseudoinverse. The result of SVD in the general case gives us three matrices, the product of which gives us the original matrix:

$$\mathbf{H} = \mathbf{U\Sigma V}^{T} \tag{2.31}$$

The matrices $\mathbf{U}$ and $\mathbf{V}$ are both square; the matrix $\mathbf{\Sigma}$ has the same dimensions as $\mathbf{H}$, and only has non-zero elements along the diagonal. These matrices have a crucial rôle in the properties of the system as a whole.

The matrix $\mathbf{U}$ consists of N_s column vectors of N_s elements, each of which represents a "mode" of the wavefront sensor. These columns are all orthonormal to each other. The matrix $\mathbf{V}$ consists of N_a column vectors of N_a elements, where these columns represents a "mode" of the PCD. Again,

these modes are orthonormal to each other. The matrix of singular values, $\mathbf{\Sigma}$ contains the scaling information: that is, for the i^{th} mode of the PCD, the i^{th} singular value tells us how much of the i^{th} mode of WFS output we will measure. If we call the diagonal elements of $\mathbf{\Sigma}$ σ_i, then we can say, and we apply a control vector which is a column of $\mathbf{V}$, $\mathbf{x} = v_i$ then:

$$\mathbf{y} = \sigma_i u_i \tag{2.32}$$

where u_i refers to the i^{th} column of $\mathbf{U}$ etc.

The orthonormality properties of the matrices $\mathbf{U}$ and $\mathbf{V}$ are of great importance: since each is column-orthonormal, then each gives a complete mapping of the appropriate space. This means that the $\mathbf{U}$ matrix tells us about all the possible WFS measurements. The matrix $\mathbf{V}$ gives us an orthonormal basis of PCD commands. By using the singular values σ_i, we can quantify how well the various modes of the PCD are sensed (observability), and how well the various modes of the WFS can be matched by the PCD (controllability). If we consider the (usual) situation in which there are more sensor signals than control channels, then there are several conclusions that we can draw.

Firstly, we consider the singular values that are non-zero: there will be r of these, where $r \leq N_a$. These correspond to the modes that of the PCD that can be observed by the WFS, and also the modes of the WFS that can be controlled. Usually, there will be several very small, or perhaps zero singular values. These correspond to modes of the PCD that require a very large (or infinite) amplitude to create a measurable response on the WFS[3]. This can be seen from equation 2.32 which shows that the amount of PCD input required to give unit WFS output is $1/\sigma_i$. The remaining columns of the $\mathbf{U}$ matrix correspond to modes of the WFS that are completely uncontrollable by the PCD.

In addition to aiding system design, this analysis also suggests an improvement over using the simple pseudoinverse traditionally used to control an AO system. The simple pseudoinverse is found from the matrices discussed above:

$$\mathbf{H}^{+} = \mathbf{V}\,(1/\mathbf{\Sigma})\,\mathbf{U}^{T} \tag{2.33}$$

where the notation $1/\mathbf{\Sigma}$ means replace each of the non-zero singular values with its reciprocal; those elements that are zero are not changed. As we can appreciate from the above analysis, those modes that are poorly sensed by the WFS will have a large gain, since they have a small singular value. The net effect of this on the AO system is usually that these modes will quickly build up a large amplitude, until the limit of the PCD range is reached. In order to prevent this from occuring, it is useful to artificially set some of the mode gains to zero (equivalent to discarding some singular values). The criterion for discarding is usually based on the "condition number", which is the ratio of the largest to the smallest singular value. By setting a target for the condition number, we can choose those modes which are kept.

The formalism developed above for the relationship between sensor and actuator can equally be applied to the sensor or the actuator on its own, essentially by considering the case in which the other component is ideal. For example, in considering an ideal WFS, we could approximate this as measuring the wavefront distortion over a discrete grid. Therefore, to perform the analysis, we would calculate a matrix which is $N_g \times N_a$, where N_g is the number of grid points considered; each column is simply a PCD response function calculated over a discrete grid. Performing the SVD will give us $\mathbf{U}$, $\mathbf{\Sigma}$ and $\mathbf{V}$ matrices as above. In this case, the $\mathbf{U}$ matrix will tell us those modes that are reproducible by the PCD, regardless of the WFS used. These modes will also have an associated singular value, which will tell us the amplitude of this mode that can be reproduced before we reach

[3] A classic example of this effect is the "waffle mode" found in Fried geometry AO systems—this mode is completely unsensed by the WFS [23]

the limit of the drive signal. The $\mathbf{V}$ matrix will tell us which combinations of control signals result in large phase distortions and those which have a small effect. When applied to a WFS, the singular-value analysis will give us essentially similar information: it will tell us about wavefronts that are inherently not detectable, and those modes which are most easily detectable.

2.5 Adaptive Optics Components

It is useful to have an understanding of the many types of real components that are in use now, and have been used in the past. This description will endeavour to help place the current work in its appropriate context.

2.5.1 Phase controlling devices

There are many different options for dynamically controlling the phase of an optical beam. These devices are known generically as "phase controlling devices" (PCDs). Each device has strengths and limitations, and therefore for a given adaptive optics application, there will be a device which is in some way optimal. For a given situation, there will be constraints based on the following parameters:

- *Range of spatial correction*—for a given system, there will exist some spectrum of wavefronts that the device will be required to correct. This spectrum may be defined in a statistical way (as is the case for imaging through characterised turbulence), or it may be defined in a more deterministic way (e.g. an adaptive optical system that is require to match, say, up to a given amount of defocus and astigmatism).

- *Range of temporal response*—in many situations, the frequency characteristics required will be coupled to the spatial characteristics. For example, for the case of atmospheric turbulence, the bandwidth required to correct tip and tilt is lower than that for higher aberration orders

- *Wavelength characteristics*—some of the devices discussed are suitable only for monochromatic operation, others can cope with polychromatic light.

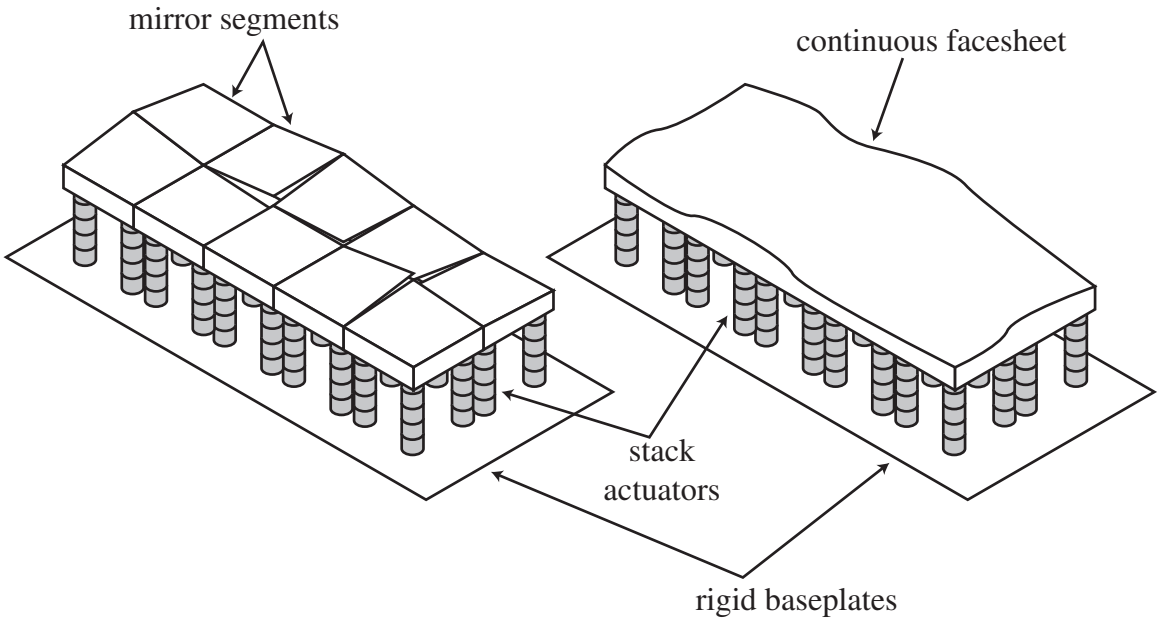
- *Coupling to wavefront sensor*—if the choice of wavefront sensor is restricted, then it may be necessary to consider how well the sensor used can detect the phase distortions introduced by the device.

- *Efficiency*—in cases where light is scarce (such as in astronomy, or retinal imaging), the amount of light scattered or absorbed may be a determining factor.

There are two ways of categorising phase controlling devices: they can be classified according to their physical realisation, or by the type of optical distortion that they can introduce. Considering first the optical classification, there are essentially two types of device. These are *modal* and *zonal* devices. A modal device can be defined as one which introduces a different phase distortion over the whole optical aperture for each input channel. A zonal device is one in which the spatial extent of the phase distortion for each control channel is limited to a small portion of the aperture.

There are many different physical realisations of both types of phase controlling device. There are two distinct types of device: those in which a portion of reflecting material is physically distorted (deformable mirrors), and those which operate by changing the refractive properties of some medium.

Figure 2.6 Monolithic and segmented deformable mirrors



Deformable mirrors

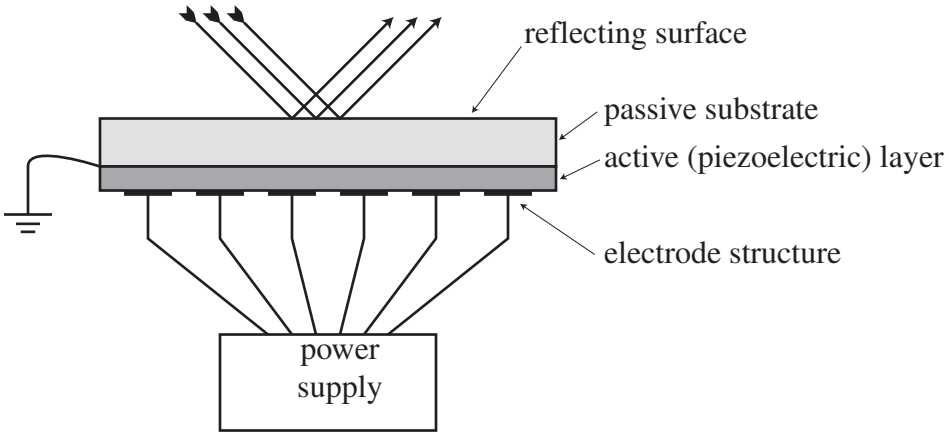
Deformable mirrors form the largest group of phase controllers. In all cases, a physical force is generated, and this moves a reflecting surface. The forces may be generated in a number of ways, but those which have been exploited with most success are the piezoelectric effect, the magnetostrictive effect and electrostatic attraction. In general, where piezoelectric stack actuators are used, in most cases a magnetostrictive actuator could equally well be used; the various properties of these types of actuators will not be discussed here.

A very common type of deformable mirror is the "segmented" mirror. This consists of several small, flat mirrors mounted on stack actuators in such a way that their position can be altered relative to some rigid backplane. Usually, each mirror has three degrees of freedom—namely, piston, tip and tilt. Some devices have been constructed in which only the piston of each mirror is controllable. These devices are therefore completely zonal in their operation, and typically require very many elements to achieve good correction. In terms of spatial frequency, those frequencies that a segmented mirror can reproduce are all reproduced with the same amplitude [24]. This type of deformable mirror couples well with slope-type sensors, although great care must be taken over the precise relative location of the slope-sensor subapertures and the mirror actuators, as discussed in section 2.4. In some geometries, it is necessary to have some sort of feedback on the relative pistons of the mirrors. In all segmented mirrors, there is the problem that light will be lost due to the gaps in between the segments. The segmented nature also means that light will also be lost due to diffraction.

A variation on the segmented mirror is the "monolithic" mirror, in which a matrix of stack actuators push the rear surface of a continuous reflective sheet. Each of these stacks causes a small region of the mirror nearby to be deflected—this is again a zonal type of mirror. The continuous reflective sheet means that the light losses from a monolithic mirror are typically somewhat smaller than from a segmented mirror. Both monolithic and segmented mirrors are depicted in figure 2.6

The monolithic and segmented mirrors are the types most commonly used for astronomy, largely because they scale up to many actuators easily; however they are usually rather expensive. The cost is due to the fact that stack actuators are used. This is a rather inefficient way of producing simply

Figure 2.7 Cross-section of a bimorph deformable mirror



displacement (there is no need to produce force in a deformable mirror, other than to move the reflecting surface). A more efficient way of using piezoelectric material is to make a bimorph-type actuator, as described below.

The bimorph concept has been around for many years [25, 26], and was first applied to adaptive optics by Steinhaus and Lipson in 1978. The operating principle is that a passive layer is glued to a layer of piezoelectric material. In response to a control voltage, the piezoelectric layer expands in the plane of the device, this causes a bending moment, and the layered structure bends. By applying a voltage to a region of the piezoelectric layer, local bending is caused. The rest of the device then deforms to match the boundary conditions. This principle is illustrated in figure 2.7, and is discussed in detail in chapter 3. The piezoelectric bimorph has many attractive features: it has a continuous facesheet, it is a modal device and is therefore suited to low-order applications, it may have a rather small component cost and the drive electronics require less power than an equivalent stack-actuator based system.

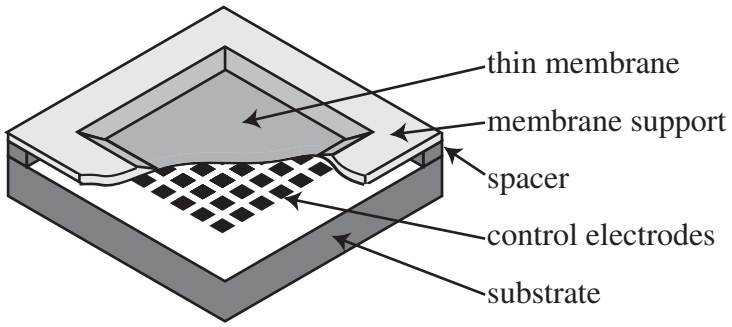
The final commonly used type of deformable mirror is the membrane mirror. This mirror is quite different in its construction to the other DM types. A cross-section of a typical membrane mirror is shown in figure 2.8. The thin membrane is moved towards the electrodes by electrostatic attraction. This mode of producing deformation has one immediate disadvantage: an electrode can only pull the membrane. This means that membrane mirrors must be pre-biased. The surface of a membrane mirror has been shown to take on a shape that is determined by Poisson's equation, where the pressure term is generated by electrostatic attraction. Therefore, membrane mirrors are an inherently modal device. The spatial frequency response is proportional to $1/k^2$, i.e. they have a lower amplitude response to higher spatial frequencies. The case of a rectangular membrane can be solved quite simply using a Fourier sine series, see for example reference 26.

Membrane mirrors have the potential to be effectively mass-produced, since they may be constructed using standard silicon fabrication techniques. Another advantage is the large stroke they are capable of reproducing—from flat to 4 Dioptres has been reported by Vdovin [3]. Unfortunately, membrane mirrors are not physically rugged (typically, a membrane might be 0.5μm thick), and are very easily damaged. Notwithstanding this fragility, membrane mirrors have succesfully been used to correct pulsed lasers with around 2.5J/pulse [27].

Liquid Crystal Devices

Liquid crystals are materials that have certain optical qualities that make them suitable for making PCDs. There are two types of LC material in common usage: nematic and ferroelectric. In both

Figure 2.8 Membrane deformable mirror



cases, the LC material has a birefringence that is dependent on the voltage applied across the layer. This means that a single layer of LC can imprint a phase change on one polarization that is dependent upon the applied voltage. By using two layers of LC arranged at right angles, it is possible to construct a polarisation-insensitive device [28]. LC devices are usually transmissive, although it is not uncommon to place a mirror behind the layer of LC to make a reflective device.

Ferroelectric types are usually characterised by a response time of a fraction of a millisecond (commercial pixelated devices supplied by DisplayTech have a frame refresh rate of up to 2.5kHz), however they can only be used to imprint a phase change of 0 or π. Nematic types typically have much slower response times, but the phase change that they can imprint is continuous over a given region, typically up to 1μm.

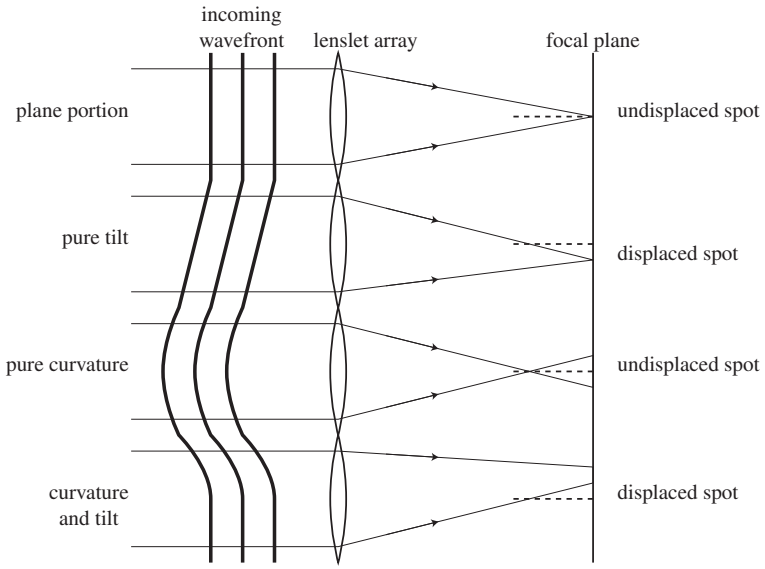
Pixellated devices are the most common implementation for both LC types, and these are available commercially. Meadowlark supply a 69 pixel nematic device, and this has been demonstrated in reference 29. This device was shown to have high optical quality, and static correction was demonstrated for aberrations with amplitudes up to around 1 wavelength at 633 nm. This device was also demonstrated in a simple closed-loop adaptive optics system [30]. In this work, the 69-element LC device was used in conjunction with a 19-element Shack-Hartmann wavefront sensor. The system attempted to correct for laboratory-generated turbulence (which approximately simulated a Kolmogorov spectrum), and succeeded in improving the Strehl ratio of the "star" image from 0.07 to 0.12. This system had a frame rate of around 20 Hz which is somewhat below that required for true atmospheric correction. The authors point out that there are significant gains to be made in their control system, which has not been optimised.

There are several important limitations of nematic LC devices; many of these are pointed out by Love in reference 29. If the polarisation problem is overcome as described above, there still remain the issues of bandwidth and stroke. The device described by Love has a bandwidth of only 26 Hz if one wave of stroke is required, and the total stroke is limited to 1μm. Love notes that advances in LC technology are likely to overcome the bandwidth limitations, and devices with bandwidths up to 200 Hz should be possible.

Ferroelectric pixellated devices may at first seem somewhat limited by their ability to change the phase by only 0 or π but if used correctly, they can be used to create a continuously varying phase. If, instead of using the device directly to imprint a phase, a phase hologram is written on the device, the diffraction orders can be so arranged that when only one of them is allowed to be Fourier transformed, a continously varying wavefront emerges [31].

A novel method of producing a continuously varying (i.e. not pixellated) phase change using a nematic LC device has been reported by Naumov [32]. By applying an AC signal to regions of one of the LC electrodes, a smoothly spatially varying phase change is produced. These devices are still in the developmental stage—all that has been demonstrated so far is that the principle of operation is

Figure 2.9 Shack-Hartmann WFS showing the effect upon the spot displacements of local tilts and curvatures



valid, and that devices with several control channels can be fabricated. One interesting point from an AO control point of view is that the modes produced by each actuator have two degrees of freedom: the amplitude of the response function is governed by the amplitude of the AC signal and the spatial extent is controlled by the frequency of the signal.

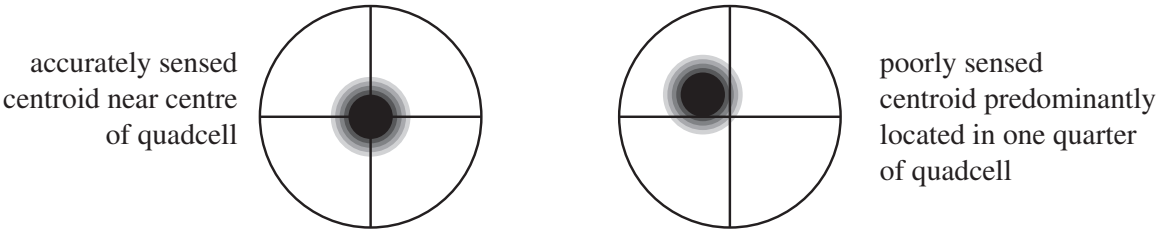
2.5.2 Wavefront sensors

There are many ways of measuring an optical wavefront. Many of these were developed with the aim of detecting defects in optical components. Examples include the Foucault knife-edge test, the star test and the multitude of interferometric techniques such as the Fizeau and Twyman-Green types [33]. None of these traditional optical techniques have found widespread use in AO sytems. This is partly because they either rely on finding a coherent, unaberrated reference beam (most interferometric techniques, except point-diffraction types), or they are simply not suited to high bandwidth applications. The two techniques most commonly employed are essentially geometric in nature, and therefore do not require a coherent reference. These are the Shack-Hartmann test and the curvature sensor, which shall be discussed in more detail below. Other sensors proposed for use in AO systems include shearing interferometers, of which there are several types.

The Shack-Hartmann WFS

This WFS technique is based on the Hartmann test, which was originally used to test large telescope optics. The modification that Shack proposed was to use a lenslet array instead of a perforated plate, therefore there is no loss of information. This lenslet array causes segments of an essentially collimated beam to be focused into an array of spots. If the portion of the beam illuminating one of the lenslets has an average tilt, then the corresponding spot is displaced. This is shown diagrammatically in figure 2.9. The centroid of the resulting spot can be shown to be displaced by an amount proportional to the average subaperture tilt [34]. Thus the Shack-Hartmann sensor measures a series of wavefront slopes. These slopes can be used to recreate the wavefront if necessary [35], or they can be used directly to control the PCD, as outlined in section 2.4.

Figure 2.10 Accuracy of quadcells used to determine centroid locations



There are two practical ways of measuring the centroid displacements. A quadcell can be used, the outputs of which can be translated into centroid position with minimal computing overhead. This has the disadvantage that the accuracy of the centroid changes as the spot moves [36], the limit being reached when the spot is wholly in one side of the detector, as shown in figure 2.10 One commonly used configuration to avoid this problem is to use a CCD camera array to record an image of the spots. The CCD array can be so arranged that each spot sits in its own region which covers many pixels. This approach has the drawback that the computing time required to compute the centroids accurately from the image data is large—especially when compared to the quadcell which requires almost no processing at all. With the advent of modern parallel processors, the advantages of the CCD approach are making this the method of choice in most situations. In addition to the constant spot-accuracy, using a CCD enables more sophisticated spot-location techniques to be employed [37], at the expense of yet more computation. A quadcell SHWS can only measure those aberrations that do not cause its lenslet's spot to move away from the quadcell into the neighbouring quadcell. Using a CCD array, even this limitation can be overcome [38], and wavefront distortions of several hundred waves may be measured with high precision.

The main advantages of the SHWS sensor are:

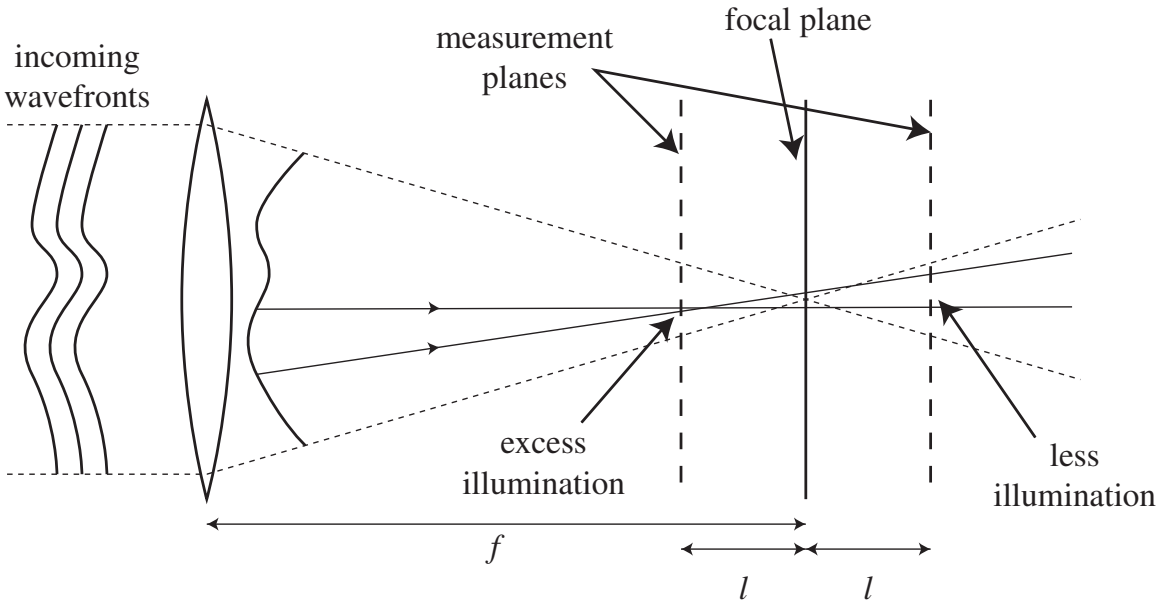
- Geometrical measurement technique means reduced vibration sensitivity, and works with white light
- With a CCD detector, high measurement accuracy and possibility of measuring large deflections
- Frame rates of up to 800Hz have been achieved using inexpensive, commercially available components [39].

Curvature Sensors

The curvature sensor was first proposed by Roddier [40] based on the observation that there is information to be obtained about the aberrations of an optical system from an extra-focal intensity measurement. The method he proposed was a quantitative version of this method. The incoming aberrated wavefront is brought to a focus. At planes either side of this focal point, measurements are made of the intensity distribution, as shown in figure 2.11. Roddier showed that inside the pupil, the difference in intensity at two corresponding points in the aperture was proportional to the local curvature. The intensity difference around the edge of the pupil gives slope information. Therefore, we can reconstruct the phase inside the pupil having both curvature information, and boundary conditions.

Referring to the dimensions given in figure 2.11, Roddier derived the sensor signal $S(\mathbf{r})$ as a

Figure 2.11 The curvature sensor



function of the wavefront distortion $\phi(\mathbf{r})$:

$$S(\mathbf{r}) = \frac{I_2(\mathbf{r}) - I_1(\mathbf{r})}{I_2(\mathbf{r}) + I_1(\mathbf{r})} = \frac{f(f-l)}{2l}\left[\frac{\partial}{\partial \mathbf{n}}\phi(f\mathbf{r}/l)\delta_c + \nabla^2\phi(f\mathbf{r}/l)\right] \tag{2.34}$$

where I_1 is the pre-focal intensity, and I_2 is the post-focal intensity distribution. The first term in the square brackets is that which measures the boundary conditions: the term δ_c is an impulse function around the edge of the optical pupil. This expression is valid in the geometrical approximation, and shows that this sensor measures the Laplacian of the incoming wavefront. This fact means that this type of sensor couples very well to both bimorph and membrane mirrors, although it has been shown that the control is not completely diagonal for the case of a bimorph mirror [41]. Notwithstanding the fact the control is not diagonal, numerical studies presented in section 4.5 show that the bimorph mirror is very well sensed by a curvature sensor.

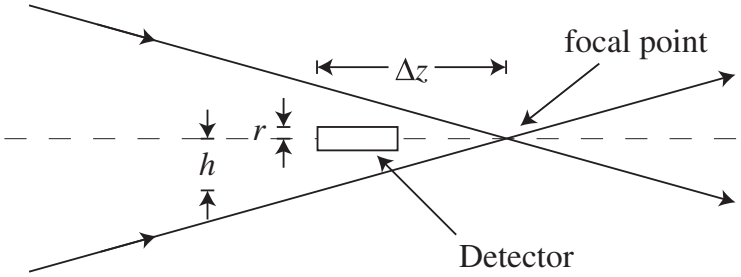
The technique of curvature sensing has other attractive features: as with the SHWS, it is mainly geometric in its operation, and so it copes well with an incoherent, white light source. Again, vibration sensitivity is not nearly so critical as in an interferometric technique. In contrast to interferometric techniques, long-exposure images (in both curvature sensors and SHWSs) give the average wavefront error—the effect of changing turbulence is averaged out. In an interferometric technique, if the turbulence induces changes in a fringe pattern of more than 1 fringe, then in a long exposure image, all the fringes are blurred and no wavefront information remains.

Roddier and Roddier [42] showed that this technique can be used as a high-resolution testing method for telescopes, in addition to being used as an element in an AO system. In this work, they showed how a "closed-loop" form of wavefront sensing could be implemented: they iteratively corrected the extra-focal images after calculating the low-order aberrations and then subtracting their effect from the images.

Other techniques

Although the two techniques described above have accounted for most of the systems so far implemented, there are other options that have been proposed. These techniques include an automated

Figure 2.12 Principle of a focus detector



Foucalt knife-edge test [43], and the various shearing interferometers [44]. Various schemes for retrieving the phase from a focused image are possible, but they are generally iterative in nature [45] and not really suited to real-time operation. Of more interest for a low-cost, low-order AO system are the various methods of sensing a focal error. Several of theses are described by Kuttner [46], and two types particularly suited to low-order AO use will be described here.

The pinhole focal sensor works on a very simple principle. Consider a beam of light being caused by a lens to form a focus. At that focus, we place a pinhole, and directly behind that pinhole, we place a photodetector. As the beam is defocused, the size of the beam at the pinhole increases, and so the intensity measured by the detector becomes lower. We are essentially measuring the current Strehl ratio of the beam.

Figure 2.12 shows the principle of operation, and it is easy to calculate the expected detector signal. We will consider the case of moving a detector of radius r a distance Δz from the focal point of a beam with a focal ratio F. At this point, the radius of the beam (in the geometrical approximation) is:

$$h = \frac{\Delta z}{F} \tag{2.35}$$

The intensity measured at the detector is simply the fraction of the beam intercepted, i.e. r^2/h^2. Therefore,

$$I = \frac{r^2 F^2}{(\Delta z)^2} \tag{2.36}$$

Inverting this expression gives us the distance Δz in terms of the measured intensity:

$$\Delta z = \frac{\sqrt{I}}{rF} \tag{2.37}$$

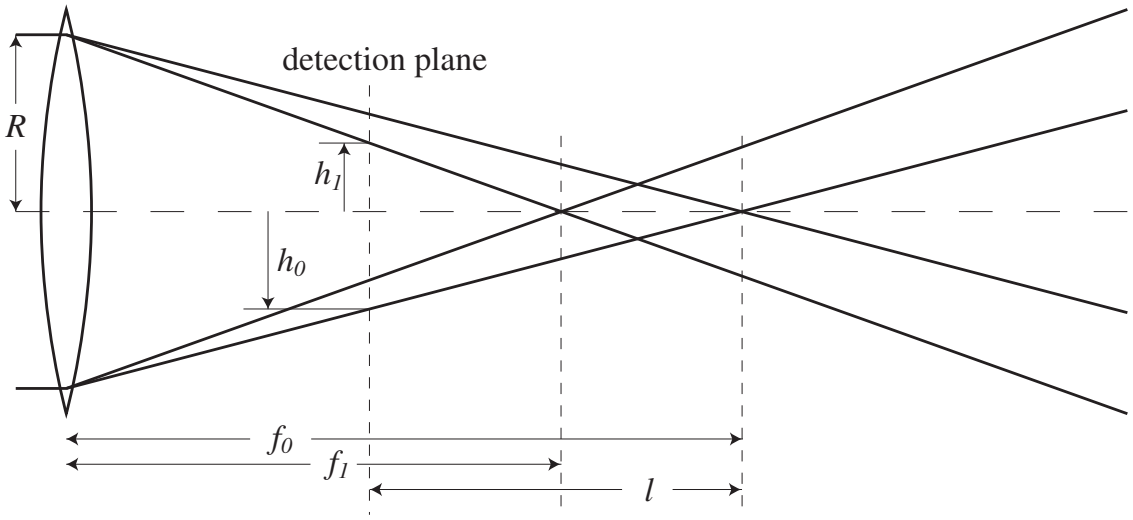
This expression is calculated assuming a constant focal ratio for the incoming beam. This is only true if the detector is moving, it is not the case when we are introducing a defocus with an adaptive mirror. This case is also simply dealt with: referring to the notation defined in figure 2.13, we note as before that the intensity within the illuminated region is approximately proportional to the ratio h_0^2/h_1^2. We can immediately say, by considering similar triangles that:

$$\frac{h_0 f_0}{l} = R = \frac{h_1 f_1}{l - f_0 + f_1} \tag{2.38}$$

In the thin-lens approximation, we simply add the focal powers of the original beam (i.e. that which produced the focus at a distance f_0). If we call the adaptive power introduced p_a, then

$$f_1 = \frac{f_0}{f_0 p_a + 1} \tag{2.39}$$

Figure 2.13 Principle of a pinhole focal detector



Substituting equation 2.39 in equation 2.38 and rearranging, we obtain:

$$I = \frac{h_0^2}{h_1^2} = \frac{l^2}{\left(l - {f_0}^2\, p_a + f_0\, l\, p_a\right)^2} \tag{2.40}$$

This can be solved to give the adaptive power in terms of the sensor signal:

$$p_a = \frac{l\left(1 \pm \frac{1}{\sqrt{I}}\right)}{f_0\,(f_0 - l)} \tag{2.41}$$

This shows that in this case, it is again the square root of the detected intensity that is important. The form of the detector signal is shown for typical parameter values in figure 2.14.

This type of sensor has been used throughout this this thesis as a simple yet effective method of measuring the focal length of an adaptive mirror. It has the advantage that it is sensitive enough to measure a different signal for the smallest change applicable by the mirror-driving hardware used in the experimental work. It can also operate at very high speed—the bandwidth is only limited by the photodetector and associated amplifier—these are typically orders of magnitude faster than even the fastest full-field wavefront sensors used in adaptive optics.

Another option for a simple focus detector is the astigmatic sensor. In this sensor, an astigmatic element is placed in a converging beam. This causes the beam to come to two line foci separated by some distance. These two line foci are at right angles to each other, and the beam changes shape smoothly between these two extremes. Figure 2.15 shows how the outputs of a quadcell can be used to measure the focal state. Referring to the sensor quadrants shown in the figure, the sensor signal is calculated as $(A+D)-(B+C)$. The sensitivity of this type of sensor is determined by the relative difference in focal lengths of the converging element and the cylindrical element.

2.6 Review of Bimorph Mirrors

Since the main topic covered in this thesis is the subject of bimorph deformable mirrors, a review of the literature concerning these devices will be useful.

The first practical bimorph mirror was reported by Steinhaus and Lipson [25]. This device was demonstrated both quasistatically and dynamically. The quasistatic interferograms were produced

Figure 2.14 Detected intensity in a pinhole focal sensor with $l = 15\text{mm}$ and $f_0 = 30\text{cm}$

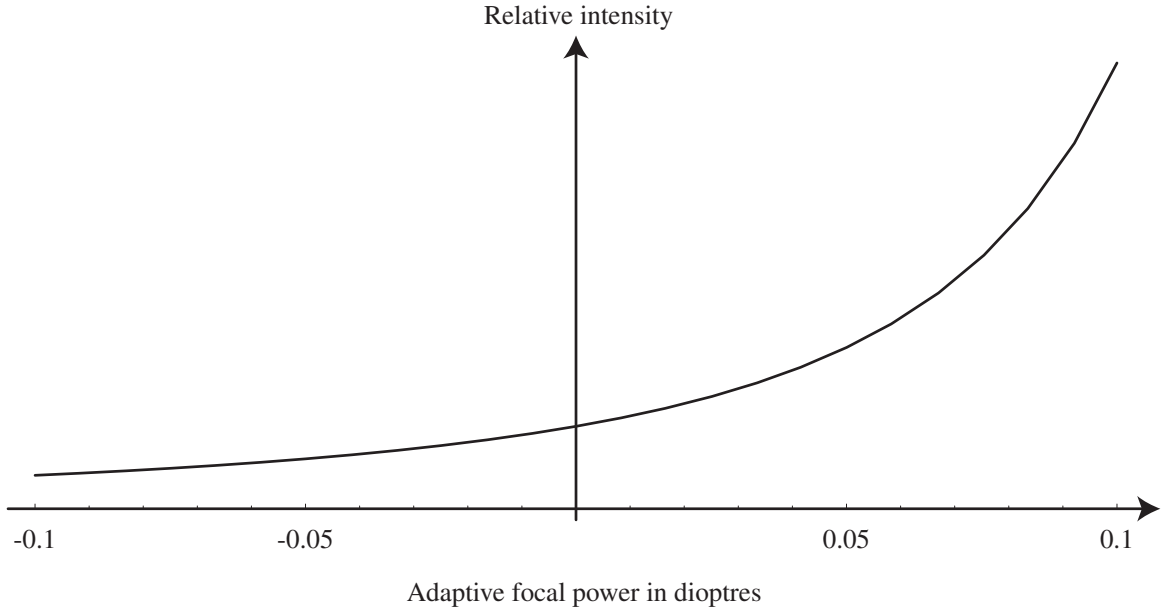
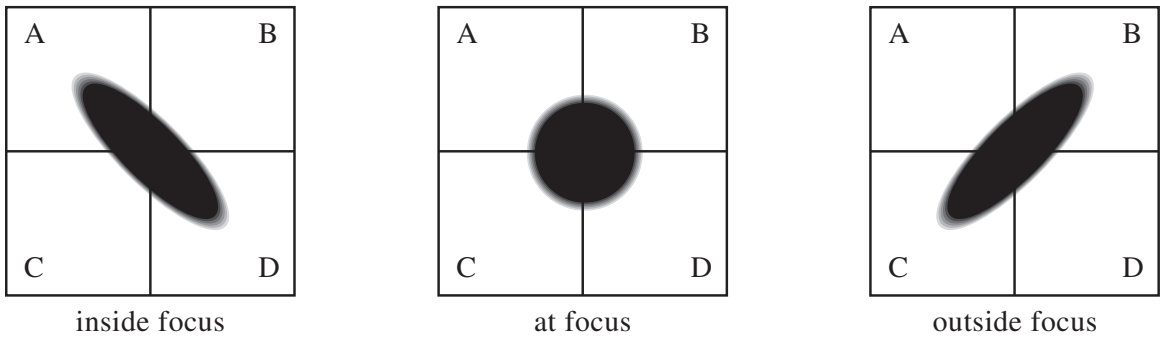


Figure 2.15 The different spot shapes in an astigmatic focal sensor



with a holographic interferometer, suggesting that the initial surface figure of the device was rather poor. Also, the sensitivity was rather low due to the use of a thick (1.5mm) layer of piezoelectric material. The device was found to have roughly constant sensitivity up to a frequency of around 1.3 kHz; above this frequency, the device became rather insensitive.

At around the same time as the Steinhaus and Lipson paper, Kokorowski [26] independently produced a theoretical analysis of a true bimorph structure (the theory is slightly different in detail to that presented in chapter 3 which deals primarily with a semi-passive bimorph, or "unimorph" structure). Several very important results are obtained, which are relevant even though most practical devices are semi-passive. The first of these results is that the bending sensitivity of the device is strongly dependent on the layer thicknesses. The second result is that the sensitivity of a bimorph is related to the reciprocal of the squared spatial frequency of the applied voltage.

Recently, a review paper was published by Safronov [47] which references many of the subsequent developments in the field of adaptive optics. Of particular interest is the table showing the various parameters of practically realised bimorph mirrors. The most interesting features of this table are the variety of substrate materials used, and the curvature sensitivities achieved.

Substrates used have included various glasses, copper (for CO_2 laser applications), silicon, quartz and molybdenum, as well as "true bimorph" devices which use 2 layers of piezoceramic, one of which is mirrored. Some devices have cooling channels incorporated into their structure [48] which is important if the devices are to be used in high power lasers: the cited example was used in the cavity of a 5kW CO_2 laser.

The diameters of practical bimorphs have generally been limited by the available diameters of thin piezoceramic plates: in order to limit the voltage requirements, thicknesses $\leq$0.5mm are generally preferable; this usually means that a piezo diameter of around 60mm is the upper limit. Safronov suggests in his review that larger bimorphs may be constructed by attaching a mosaic of piezoceramic plates to the back of a large substrate. In fact, he describes plans to build a 3.3m diameter bimorph mirror. This mirror would have a substrate thickness of 78mm and a piezoceramic thickness of 1mm—he notes that the ratio of thicknesses is far from optimal (as will be shown in chapter 3) however it should still produce enough deflection to be usable for correcting atmospheric turbulence.

The sensitivity of a bimorph mirror is intrinsically linked to the thickness of the device: simply stated, the thinner the device, the more sensitive it will be, provided the substrate and piezo layer are approximately the same thickness. The fundamental measure of bimorph mirror sensitivity is the deflection in the defocus mode (i.e. all electrodes actuated to the same voltage). Devices have been reported ranging from 17μm/kV for the cooled mirror described above up to around 100μm/kV for a device whose substrate was a thin (around 0.3mm) layer of silicon [49], however this device was not dynamically stable. Forbes [50] reports the construction of a "true bimorph" device whose defocus sensitivity is around 50μm/kV.

Ikeda and Sato [51] considered the possibility of constructing a bimorph structure from PVDF film, and succeeded in creating a practical device. PVDF film has the property that its expansion is anisotropic—it expands in one direction only, and so the shape produced is cylindrical rather than spherical. In the course of the construction of bimorph devices undertaken during the current work, it was found that most PVDF films are insufficiently mechanically rigid to allow them to be made into usable bimorph mirrors.

Kudryashov and Shmalhausen [52] describe the construction of low-order semipassive bimorph mirrors, and describe a theoretical solution to the equation governing the response functions of the mirrors. Experimental response functions are shown, although no detailed comparison of the form of the response functions and the form of the theoretical solution is given. Even though an expression

for the sensitivity of the bimorph structure in terms of the thicknesses and elastic properties of the layers was found, this was not discussed. This expression is compared later with the theory developed during the course of the current work in section 3.4.2.

2.7 Practical Adaptive Optics Systems

In order to put into context the practical work carried out, it is useful to describe AO systems that have been experimentally realised. Most working AO systems are still employed in the field of astronomical imaging; however this thesis is concerned with systems which could be built for a much lower overall cost than the very expensive systems in use on telescopes. Hence, such systems will receive only a cursory treatment. Of more interest are the systems involved with intracavity laser AO and retinal imaging AO—for these applications to be commercially viable, the component cost of the overall system must be very significantly lower than an astronomical system.

2.7.1 Astronomical systems

All major large telescopes are currently implementing adaptive optics systems. These systems are at varying stages of completion, and using a variety of deformable mirrors, wavefront sensors and artificial guide stars. The capabilities of two quite different systems of these will be outlined here to demonstrate the current state of the art.

On Mauna Kea, the Canada-France-Hawaii 3.6m telescope has an AO system called "Hokupa'a" [53]. This system is of particular interest because it is based around a 36 element bimorph deformable mirror and a 36 element curvature sensor. The curvature sensor uses a vibrating membrane to switch between defocused pupil images, and photon counting APD detectors are used which have excellent noise properties. The detector subapertures are mapped onto the bimorph deformable mirror. The mirror is an advanced type of "true bimorph": the device consists of 2 active piezoelectric layers, and by using epoxy replication techniques, a reflective surface was deposited directly on one of the piezoelectric layers. This should have the effect of allowing a great deal of sensitivity, or alternatively, lower the power requirements of the deformable mirror. The interaction matrix was demonstrated to be almost completely diagonal, and has the advantage that the signal-to-noise ratio is the same for each electrode. Also, this shows that there is no chance of hidden modes such as the "waffle mode" [23]. Results are shown from early observations, which show an improvement of a factor of 30 in peak intensity. These data were taken in the spectral region $\lambda = 0.936\mu$m while the Fried parameter had the value $r_0 = 17$cm at that wavelength (i.e. $D/r_0 \approx 20$); the corrected Strehl ratio was 0.3.

An illuminating comparison can be made with the Starfire Optical Range (SOR) 3.5m AO system [54]. This system is based around a zonal approach: the deformable mirror is a 746 element segmented type, and the wavefront sensor is a 684 subaperture Shack-Hartmann sensor, effectively using a "quadcell" of CCD pixels to obtain centroid measurements. The data processing for such a system is, of neccessity, a good deal more complicated than the system described above. The closed loop bandwidth was measured to be 50Hz, using a WFS frame rate of over 800Hz [55]. Preliminary results shown demostrated a peak intensity improvement of a factor of around 65, achieving a Strehl ratio of around 0.5. Also shown was a resolved binary star with a separation of less than 0.1 arcseconds. These data were taken in the I band ($\lambda \approx 0.85\mu$m), with a Fried parameter r_0 in the range 6–8cm, leading to a D/r_0 parameter of around 60.

The large difference between the severity of the turbulence experienced by the Hupaka'a system and the SOR system partly explains why the SOR needs on the order of 1400 slope measurements whereas the Hupaka'a systems operates giving broadly the same degree of correction with only 36

sensors. An approximate criterion is that the number sensor/actuator channels scales as $(D/r_0)^2$ [11]. Since the D/r_0 is some 3 times greater for SOR, they should need 9 times as many actuators to achieve the same correction as Hupaka'a. That they need 40 times as many to achieve only slightly better correction deserves some comment. The squared scaling ignores temporal fitting errors—if the turbulence is more rapidly varying at SOR, they would need more actuators to achieve good correction.

2.7.2 Intracavity laser AO

Away from astronomy, the problem of correcting for the aberrations of a laser gain medium is the one which has received most attention. This is a conceptually more difficult problem than that of imaging—because the process of mode formation inside a laser cavity is an inherently complex one, and apart from the simplest cases, must be modelled in an iterative, numerical way [6]. Therefore, most of the methods of correction attempted have been variations on some sort of hill-climbing.

In the early 1980s, a series of theoretical [56] and experimental papers [4,57–59] were published concerning the intracavity correction of an unstable CO_2 laser, using several optimisation parameters. These experiments demonstrated the difficulty of achieving good correction, as they were only able to correct for small amounts of artificially introduced aberration.

More recently, a group working at NICTL in Russia have worked with Nd:YAG lasers. This culminated in a paper by Cherezova *et al.* [21], which summarised their results. They succeeded in improving the divergence of the multimode radiation by a factor of 2. They also noticed that some deformable mirror configurations resulted in rectangular or triangular mode structures, although no deterministic control over these shapes was attempted.

Kudryashov and Samarkin [48] used a water-cooled bimorph mirror in the cavity of a high power CO_2 laser. By changing the focal length of this mirror, they were able to change the cavity parameters sufficiently to induce intensity modulation.

A very simple application of an adaptive mirror was developed by Bär *et al.* [2] in which the focal length of a mirror was changed to keep a CO_2 laser beam in focus over a workpiece. This work showed that it is not necessary to produce a complicated, many-degrees-of-freedom system in order to obtain a useful instrument.

2.7.3 Retinal imaging

The eye is an imperfect optical system, and experiments have shown that diffraction limited performance is obtained with a pupil diameter of about 3mm. If diffraction limited performance could be achieved at a pupil diameter of 7mm (around the limit for the eye), then individual photoreceptors on the retina could be resolved. This would be an invaluable clinical tool for assessing the development of these photoreceptors—without AO, they have never been imaged *in vivo*. This application requires much care: the most problematic aspect is wavefront sensing. Since there is no light source in the retina, it must be illuminated, and the amount of reflected light from the retina is very low. Also, the "double-pass" process involved incurs a loss of information [60] about the wavefront aberrations. Notwithstanding these difficulties, Liang *et al.* [61] have managed to resolve individual cones.

2.8 Kolmogorov Turbulence

As mentioned above, the light reaching us from stars is distorted by the effect of the atmosphere. This distortion has been much studied, both theoretically and experimentally, and a model for the

stastical properties of the distortion has been developed. Turbulence that is consistent with these statistical properties is known as Kolmogorov turbulence. A thorough discussion of Kolomogorov turbulence is beyond the scope of this thesis; a thorough review of the optical implications can be found in reference 62. For the purposes of this work, all that is required is a knowledge of how to simulate wavefronts which have been subject to the effects of the atmosphere. It will also be useful to consider briefly the underlying cause of the distortion, so that we can understand how to perform the simulations.

2.8.1 Physical basis

As a beam of light propagates through the atmosphere, it encounters regions with different refractive indices. If the refractive index of the air is not uniform across the beam, then the wavefront of the beam will suffer distortion. In general, the refractive index of the air will vary from place to place. This variation is primarily due to variations in the local air temperature. Temperature variations arise from the non-uniform heating of the earth's surface. This very large scale temperature fluctuations are dispersed by turbulent convection, as well as wind flow. The mixing of air at different temperatures cause turbulent eddies, which continually break down into eddies of smaller size. This movement of refractive index inhomogeneity to smaller and smaller scales allows us to measure a power spectrum of refractive index fluctuations, $\Phi(\kappa)$. This power spectrum tells us the relative strength of the fluctuations at spatial scales $l = 2\pi/\kappa$. The work by Kolmogorov lead to a form for power spectrum that contains three distinct regions. For scale sizes greater than some length L_0 known as the *outer scale*, the fluctuations are dominated by large-scale considerations such as weather, and the geographical location. The spatial frequency associated with this is $\kappa_0 = 2\pi/L_0$. In this region, an *adhoc* form is used for the power spectrum — this also prevents the power spectrum from becoming infinite as $\kappa \to 0$. In the *intertial subrange*, predictions can be made about how the energy is transferred between the different size eddies. This leads to a form for the power spectrum in this region [63]:

$$\Phi(\kappa) = 0.033 C_n^2 \kappa^{-11/3} . \tag{2.42}$$

The parameter C_n^2 is known as the *structure function* of the fluctuations. This parameter essentially defines the strength of the turbulence, and it varies from place to place and from time to time. When the eddies become smaller than a certain size l_0, known as the *inner scale*, they dissipate their energy in a viscous manner, rather than transferring it to yet smaller eddies. This causes a rapid drop in the power spectrum at the associated spatial frequency κ_m.

To account for the inner scale, the outer scale and the inertial range all together, a form for the structure function that is commonly used is that of von Kármán:

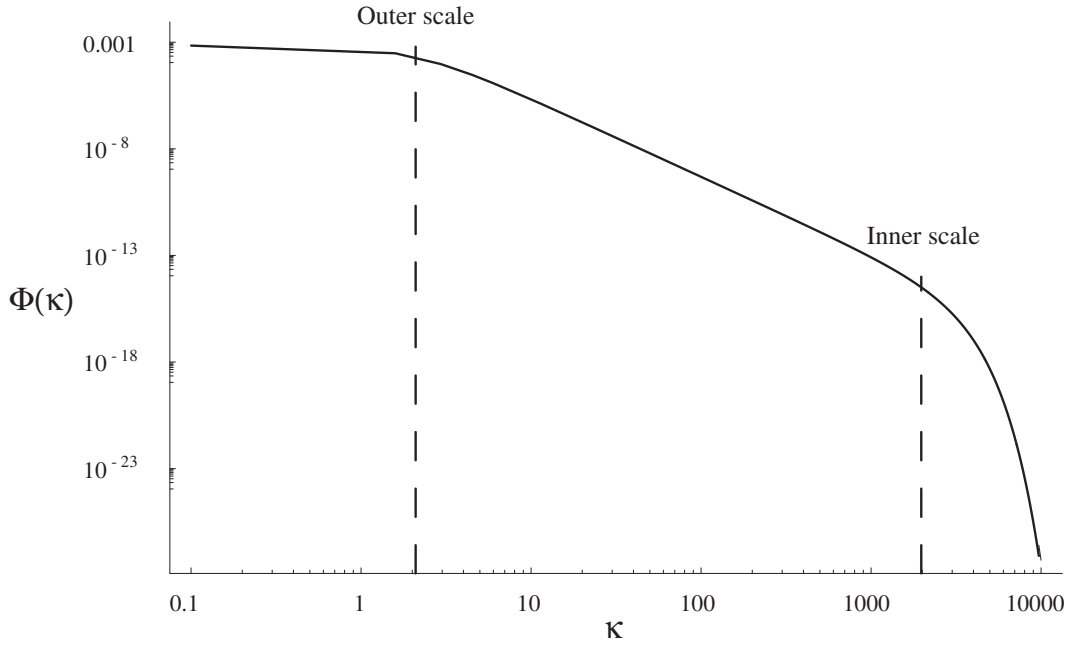
$$\Phi(\kappa, z) = \frac{0.033 C_n^2(z)}{\left(\kappa^2 + \kappa_0^2\right)^{11/6}} \exp\left(\frac{-\kappa^2}{\kappa_m^2}\right) , \tag{2.43}$$

where we have incorporated the possibility of the variation of C_n^2 with vertical displacement z. This structure function is shown in figure 2.16.

2.8.2 Optical effects of Kolmogorov turbulence

A full discussion of the effects of a turbulent atmosphere on a beam of light passing through it requires a careful discussion in terms of the propagation of the beam through the turbulent medium. However, for the purposes of this thesis, we will simply state the result which tells us that the power

Figure 2.16 A von Kármán power spectrum, with $C_n^2 = 1$.



spectrum of the phase in the pupil is related to the power spectrum of refractive index fluctuations by [64]:

$$\Phi_\phi(k) = (2\pi/\lambda)^2 \int_0^\infty C_n^2(z)\Phi(k)\cos^2\left(\frac{k^2 z\lambda}{4\pi}\right) \mathrm{d}z\,. \tag{2.44}$$

It is usual to ignore the cosine term in this expression, which is equivalent to setting the wavelength to zero, i.e. this is a geometric approximation. Once this is done, we are left with an expression in which only C_n^2 depends on z in the integral.

2.8.3 Simulating Kolmogorov wavefronts

To test the potential performance of an AO component, it is necessary to be able to simulate wavefronts that have the same statistics that the real system will encounter. Since most AO systems to date have been concerned with the problem of astronomical imaging, a great deal of effort has been expended trying to simulate realistic atmospherically degraded wavefronts. There exist two relatively straightforward approaches to these simulations: one is a zonal approach, the other is a modal approach. The method that will be used elsewhere in this thesis is the modal approach. This method is based on a Karhunen-Loève expansion, as described by Goodman [63]. This method was applied by N. Roddier [65] to the problem of simulating wavefronts with a Kolmogorov spectrum in terms of Zernike polynomials (Zernike polynomials are described in Appendix A). The approach used by Roddier will be outlined here.

Roddier notes that the Zernike coefficients of an atmospherically degraded wavefront will be zero-mean Gaussian random variables. However, since the covariance between the different terms is not zero, we cannot simulate wavefronts directly. The Karhunen-Loève (KL) method creates a new orthogonal basis with zero covariance between the various terms in the expansion. In order to perform the KL expansion, we need an expression for the covariance between the Zernike orders. This was found by Noll [66], and was corrected by Roddier. Using the resulting expression, a covariance matrix $\mathbf{C}$ can be calculated. This matrix is Hermitian; one property of a Hermitian

matrix is that it there exists a unitary matrix $\mathbf{U}$ where the product

$$\mathbf{U}\mathbf{C}\mathbf{U}^T \tag{2.45}$$

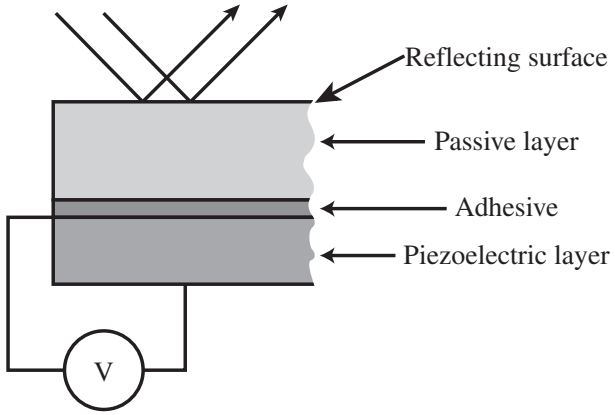
is diagonal. This matrix $\mathbf{U}$ can be found using SVD. The columns of this matrix are the Karhunen-Loéve basis functions in terms of Zernike polynomials, and their covariance is zero. The variances of the modes are given by the singular values of the covariance matrix. Therefore, all that is required to simulate an instance of a turbulence degraded wavefront is to produce a vector of Gaussian random variables whose variance is scaled by the appropriate singular value, and then pre-multiply this by $\mathbf{U}^T$ to generate its Zernike expansion.

This method of generating turbulence degraded wavefronts has one significant disadvantage: it does not simulate how a wavefront will evolve over time. However, once the $\mathbf{U}$ matrix has been calculated, the simulation of many independent wavefronts is very quick, requiring only a random input vector and a multiplication of this matrix by a vector.

3 Theory of Bimorph Mirrors

A bimorph mirror is essentially a laminated plate, consisting of several layers, see figure 3.1. Although this figure shows only two layers (neglecting for now the adhesive layer), we can de-

Figure 3.1 Cross section of a bimorph mirror



velop the theory in a more general way, allowing the consideration of any number of layers. The device deforms because of the piezoelectric effect: when there is an electric field across the layer of piezo ceramic, a strain is induced. This strain causes an expansion or contraction relative to the passive layer, and as a result the structure bends. Evaluation of this strain, and the subsequent stresses set up in the structure allow an equation to be derived relating the applied voltage to the deformation of the mirror surface.

3.1 Basic Introduction to Stress and Strain Analysis

Before we can derive the equilibrium deformation of a bimorph mirror, we need to establish the basic mechanical principles that govern the behaviour of the device.

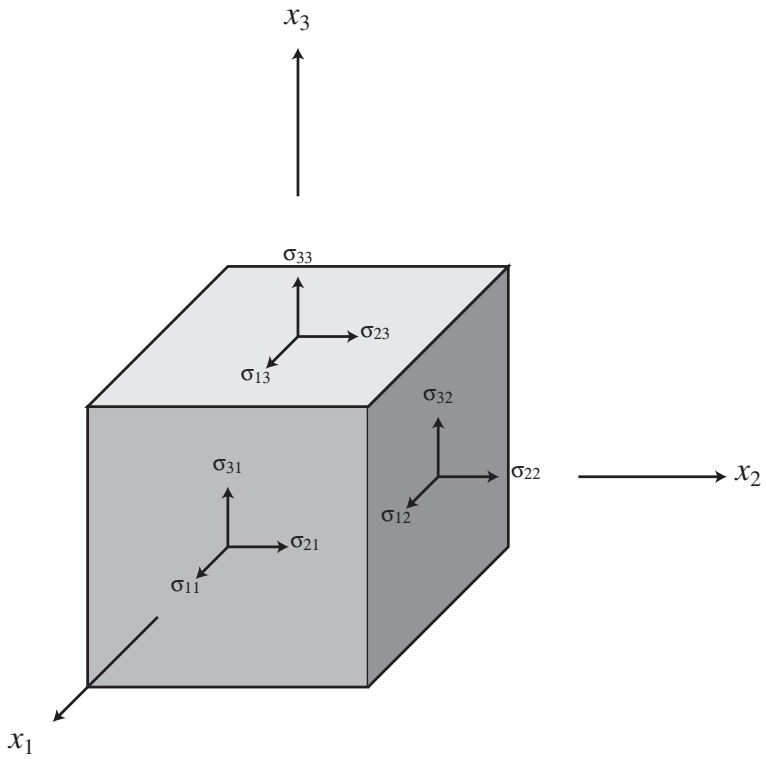
3.1.1 Stress

A stress is a force applied by one part of a body on a neighbouring part. This is distinct from a force which acts on a body as a whole, such as the effect of gravity. The stresses acting on an element are shown in figure 3.2. The force on a surface of an element is proportional to the product of the stress and the area of that surface. In three dimensions, the 9 components of stress form a tensor, σ. In the absence of body-torques this tensor is symmetric, and so only has 6 independent elements.

3.1.2 Strain

When a body is in a deformed state, we describe this deformation as a *strain*. To define an appropriate mathematical description of this phenomenon, consider a point P, at a position $\mathbf{OP} = \mathbf{x}$, in a

Figure 3.2 Stresses acting on a unit element



body which is then deformed so that the point moves to P'. Let the displacement vector **PP'** = **u**. In order to investigate the strain at this point, we first define the quantities

$$e_{ij} = \frac{\partial u_i}{\partial x_j}. \tag{3.1}$$

We may see the interpretation of these quantities with reference to figure 3.3 Consider the displacement of point Q, initially displaced by $\Delta\mathbf{x}$ from the point P. Under the deformation, the point moves to Q'. The new position of Q' can be found by considering derivatives of the displacement **PP'**, which is **u**. By considering components of the extra displacement along the 1 and 2 directions, we can write:

$$\Delta u_1 = \frac{\partial u_1}{\partial x_1}\Delta x_1 + \frac{\partial u_1}{\partial x_2}\Delta x_2 \tag{3.2}$$

$$\Delta u_2 = \frac{\partial u_2}{\partial x_1}\Delta x_1 + \frac{\partial u_2}{\partial x_2}\Delta x_2 \tag{3.3}$$

Using equation 3.1, we can write this relationship which shows that the elements of e_{ij} form a tensor:

$$\Delta u_i = e_{ij}\Delta x_j \tag{3.4}$$

Although we have found the tensor e_{ij} which describes the state of deformation of a body, this quantity is not consistent with what we would like to measure: the state of deformation of a body. If a body is rotated as a whole, we would like our strain to remain zero, although this does not happen with e_{ij}. It can be shown [67] that the pertinent part of the tensor is its symmetric part, given by

$$\varepsilon_{ij} = \frac{1}{2}(e_{ij} + e_{ji}). \tag{3.5}$$

The antisymmetric part of the tensor simply describes the state of rotation of the body. Since the strain tensor is symmetric, it shares with the stress tensor the property of having only 6 independent elements.

Figure 3.3 Demonstration of strain in 2 dimensions

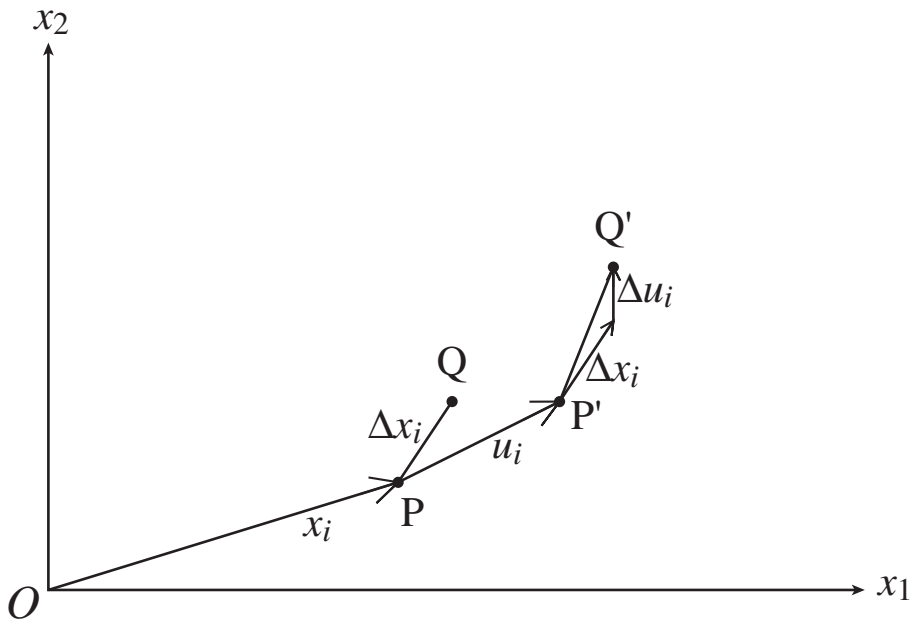


Table 3.1 Compact notation for stresses and strains

Stresses		Strains	
Tensor	Compact	Tensor	Compact
σ_{11}	σ_1	ε_{11}	ε_1
σ_{22}	σ_2	ε_{22}	ε_2
σ_{33}	σ_3	ε_{22}	ε_2
$\tau_{23} = \sigma_{23}$	σ_4	$\gamma_{23} = 2\varepsilon_{23}$	ε_4
$\tau_{31} = \sigma_{31}$	σ_5	$\gamma_{31} = 2\varepsilon_{31}$	ε_5
$\tau_{12} = \sigma_{12}$	σ_6	$\gamma_{12} = 2\varepsilon_{12}$	ε_6

3.1.3 Stress and strain compact notation

Since both stress and strain tensors have only 6 independent elements, it is convenient to use a compact notation, as shown in table 3.1. The alternative shear strains γ_{23} etc. are often referred to as "Engineering shear strains", and it must be remembered that these elements do not form a tensor, and so do not transform in the same way as a tensor.

3.1.4 Hooke's Law

A generalised form of Hooke's law relates stresses to strains. For an isotropic[1] material, this can be written as follows:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{11} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{12} & S_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2(S_{11}-S_{12}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 2(S_{11}-S_{12}) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2(S_{11}-S_{12}) \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} \tag{3.6}$$

The components of the S matrix can be related to the more usual elastic parameters of Young's Modulus E and Poisson's ratio ν as follows:

$$S_{11} = \frac{1}{E}, \tag{3.7}$$

$$S_{12} = -\frac{\nu}{E}. \tag{3.8}$$

Also, we note that the shear modulus, G is given by

$$G = \frac{E}{2(1+\nu)}. \tag{3.9}$$

Poisson's ratio defines the ratio of lateral strain to longitudinal strain under the condition of uniform longitudinal stress.

We can invert equation 3.6 to allow us to calculate stresses from strains. This gives us the matrix C_{ij}. Anticipating the reduction to an essentially 2-dimensional problem, we set $\sigma_3 = \sigma_4 = \sigma_5 = 0$, which gives us the following relationship:

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \tau_{12} \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & 0 \\ Q_{12} & Q_{11} & 0 \\ 0 & 0 & Q_{66} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \gamma_{12} \end{bmatrix} \tag{3.10}$$

where the Q coefficients are given by:

$$\begin{aligned} Q_{11} &= \frac{E}{1-\nu^2}, \\ Q_{12} &= \frac{\nu E}{1-\nu^2}, \\ Q_{66} &= \frac{E}{2(1+\nu)} = G. \end{aligned} \tag{3.11}$$

3.2 The Piezoelectric Effect

The *direct piezoelectric effect* is the name applied to the phenomenon whereby an electric field is produced across a crystal which is proportional to the applied stress. The physical basis of this effect is a crystal lattice with an embedded permanent dipole moment. The *converse piezoelectric effect* induces a strain in the crystal, causing a change in shape, under the influence of an applied voltage. The relationship between applied electric field to the induced strain can be written as follows:

$$\varepsilon_{jk} = d_{ijk}\mathcal{E}_i, \tag{3.12}$$

[1] Although a piezoelectric material must, of necessity, be anisotropic, the material consists of multiple domains, and so we can treat the mechanical properties as being isotropic. Also, the symmetry properties of the crystal structure lead to properties that are isotropic in the *xy* plane.

Table 3.2 Values of the d_{31} parameter for various piezolectric materials

Material	$d_{31} \times 10^{-12}\mathrm{mV}^{-1}$
PZT4D	-135
PZT5A	-171
PZT5H	-274
PIC151	-210

where the d_{ijk} are components of the piezoelectric modulus, and $\mathcal{E}_i$ are components of the applied electric field. It is more usual to use compact notation, so that this equation takes on the form

$$\varepsilon_j = d_{ij}\mathcal{E}_i. \tag{3.13}$$

Due to crystal symmetry properties, many of the components of d_{ij} are either equal, or zero. In a bimorph mirror, we use a disk of piezoelectric material, applying a field in the z direction. Therefore, the most important piezoelectric coefficient is d_{31}; typical values for this parameter are shown in table 3.2. These values are for materials supplied by Morgan Matroc (those starting with the letters PZT) and Physike Instrumente (those starting with the letters PIC). They show that typical strains of piezoelectric materials are a few millistrains for applied fields of up to a kilovolt per millimetre — the usual maximum field that can be applied.

3.3 Laminated Plates

A bimorph mirror can be treated as a laminated mirror made up of several layers. In the simplest approximation, a semi-passive bimorph (or "unimorph") mirror is just two layers - one of piezoelectric material, and a passive reflecting layer. The theory developed here will be general, for any number of layers, any of which may be piezoelectric and each may have a different coefficient of thermal expansion. This enables us, in principle, to model mirrors with more than one layer of piezo, and also to include effects due to the layers of adhesive that hold the other layers together.

In the "Classical Lamination Theory" (CLT), once we have found the general relations between the stresses and strains for the laminate, we can treat the whole structure as if it were one plate. Using CLT, we can derive the effective elastic properties of the laminate.

3.3.1 Assumptions of the Classical Lamination Theory

The principle simplifying assumptions used are as follows:

- The layers of the laminate are presumed to be perfectly bonded together by infintesimally thin bonds. We can, of course, treat an adhesive layer along with any other layer; however there can be no relative slipping of the various layers.
- A line that is originally perpendicular to the surface of the laminate is presumed to stay straight and perpendicular to that surface under bending. This is equivalent to setting $\gamma_{13} = \gamma_{23} = 0$, where the 3 direction is normal to the surface.
- Lines normal to the surface are presumed to have constant length. This is equivalent to setting $\varepsilon_3 = 0$.
- The plate is much thinner than it is wide or long.

These assumptions are known collectively as the *Kirchhoff hypothesis* for plates. These assumptions become a progressively worse approximation as the plate thickness becomes comparable to the length or width of the plate.

3.3.2 Incorporating the piezoelectric effect

In order to accommodate the piezoelectric expansion, we need to make a modification to the constitutive equation relating stresses to strains. The Duhamel-Neumann form is usually used for this [68], which changes equation 3.6 as follows:

$$\varepsilon_i = \sum_{j=1}^{6} S_{ij}\sigma_j + \varepsilon_i^P \,, \tag{3.14}$$

where the expansional strain components ε^P are due to the piezoelectric effect. It is also possible to incorporate thermal expansivities in this way, by including a thermally induced strain ε^T. We may also rewrite the stiffness constitutive relationship

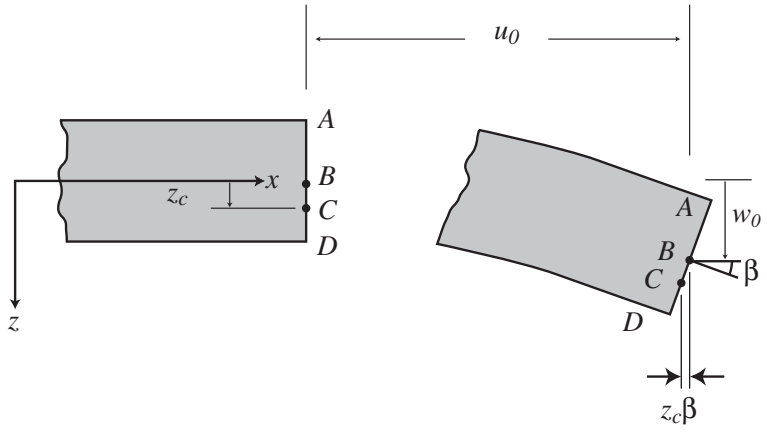
$$\begin{bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{bmatrix} = \begin{bmatrix} Q \end{bmatrix} \begin{bmatrix} \varepsilon_x - d_{31}\mathcal{E}_3 \\ \varepsilon_y - d_{31}\mathcal{E}_3 \\ \gamma_{xy} \end{bmatrix} \tag{3.15}$$

where the matrix Q is defined in equation 3.11. We also note that d_{31} is present, but not d_{32}. This is slight simplification is introduced because all the piezoelectric ceramic materials considered will have symmetry properties that ensures that $d_{31} = d_{32}$. When considering piezoelectric film materials, it will be necessary to consider the general case which allows $d_{31} \neq d_{32}$.

3.3.3 Derivation of the bending equations

Using the Kirchhoff hypothesis, we can relate the displacements (u, v, w) of a point to the stresses and strains in the laminate, following Jones [69]. Figure 3.4 shows an xz cross-section with a line ABCD through a laminate before and after deformation. The point B is in the "middle surface" of

Figure 3.4 Deflection of a laminate



the structure, that is the surface half-way through the laminate. We can relate the coordinates of point C, (u_C, v_C, w_C), to the coordinates of point B, (u_0, v_0, w_0). In the x direction, we can see that

$$u_C = u_0 - z_C\beta \,. \tag{3.16}$$

Since ABCD remains perpendicular to the middle surface, we can say

$$\beta = \frac{\partial w_0}{\partial x}, \tag{3.17}$$

so that, considering the y direction too, we can write

$$\begin{aligned} u &= u_0 - z\frac{\partial w_0}{\partial x}, \\ v &= v_0 - z\frac{\partial w_0}{\partial y}. \end{aligned} \tag{3.18}$$

The laminate strains that are non-zero in the Kirchhoff approximation are:

$$\begin{aligned} \varepsilon_x &= \frac{\partial u}{\partial x}, \\ \varepsilon_y &= \frac{\partial v}{\partial y}, \\ \gamma_{xy} &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}. \end{aligned} \tag{3.19}$$

Now we can write using equations 3.19 and 3.18

$$\begin{aligned} \varepsilon_x &= \frac{\partial u_0}{\partial x} - z\frac{\partial^2 w_0}{\partial x^2}, \\ \varepsilon_y &= \frac{\partial v_0}{\partial y} - z\frac{\partial^2 w_0}{\partial y^2}, \\ \gamma_{xy} &= \frac{\partial u_0}{\partial y} + \frac{\partial v_0}{\partial x} - 2z\frac{\partial^2 w_0}{\partial x \partial y}. \end{aligned} \tag{3.20}$$

Writing the middle surface strains as:

$$\begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \end{bmatrix} = \begin{bmatrix} \frac{\partial u_0}{\partial x} \\ \frac{\partial v_0}{\partial y} \\ \frac{\partial u_0}{\partial y} + \frac{\partial v_0}{\partial x} \end{bmatrix}, \tag{3.21}$$

and the middle surface curvatures as:

$$\begin{bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 w_0}{\partial x^2} \\ \frac{\partial^2 w_0}{\partial y^2} \\ 2\frac{\partial^2 w_0}{\partial x \partial y} \end{bmatrix} \tag{3.22}$$

allows us to write these relationships in a more compact form that will be of benefit when we come to analyse the whole laminated structure. This form is:

$$\begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{bmatrix} = \begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \end{bmatrix} + z \begin{bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix}. \tag{3.23}$$

We can now use equations 3.23 and 3.15 to calculate the stresses in any layer of our laminated plate as follows:

$$\begin{bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{bmatrix}_k = \Big[Q\Big]_k \left\{ \begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \end{bmatrix} - \begin{bmatrix} d_{31}\mathcal{E}_3 \\ d_{31}\mathcal{E}_3 \\ 0 \end{bmatrix} + z \begin{bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix} \right\} \tag{3.24}$$

Figure 3.5 Forces and moments acting on a portion of a laminate

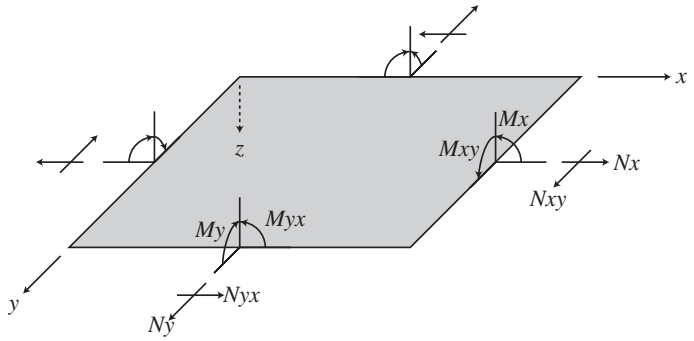
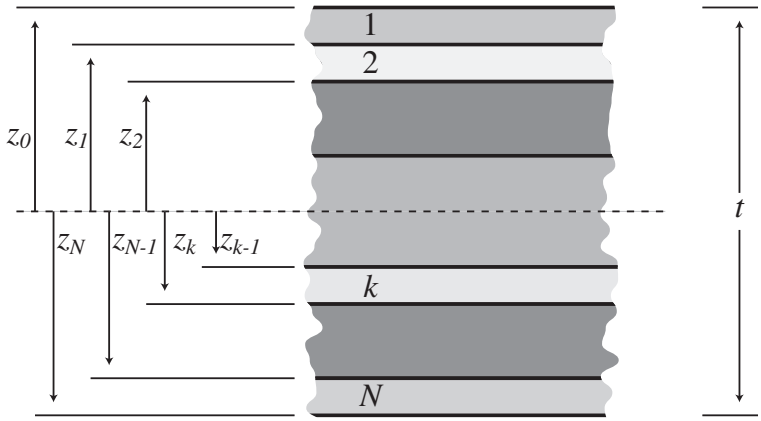


Figure 3.6 Vertical slice through a laminate showing the z coordinates of the various layers



With these expressions for stress, we can calculate the resultant forces and moments about the middle surface in the laminate by integrating over the laminate thickness. The forces, N_i are given by:

$$N_i = \int \sigma_i \, \mathrm{d}z \,, \tag{3.25}$$

and the moments M_x, M_y and M_{xy} are given by

$$M_i = \int \sigma_i z \, \mathrm{d}z \,. \tag{3.26}$$

These forces and moments are depicted in figure 3.5

In equations 3.25 and 3.26, the limits of the integration are the limits in z of the laminated plate, and the values of stress should be found using equation 3.24 as appropriate through the laminate. To calculate these forces and moments, we will need to set up a coordinate system in z. This system is depicted in figure 3.6. Under some circumstances, it may be more useful to define the origin of the z axis in such a way that it coincides with a plane that remains unstrained under bending. This plane is known as the *neutral surface* of the laminate.

We can now produce the main result of CLT: the overall stiffness relationships for our laminated

plate. Firstly, we need the complete expressions for the forces and moments.

$$\begin{bmatrix} N_x \\ N_y \\ N_{xy} \end{bmatrix} = \sum_{k=1}^{N} \Big[Q\Big]_k \left\{ \int_{z_{k-1}}^{z_k} \left\{ \left(\begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \end{bmatrix} - \begin{bmatrix} d_{31}\mathcal{E}_3 \\ d_{31}\mathcal{E}_3 \\ 0 \end{bmatrix} \right) + z \begin{bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix} \right\} \mathrm{d}z \right\} \tag{3.27}$$

$$\begin{bmatrix} M_x \\ M_y \\ M_{xy} \end{bmatrix} = \sum_{k=1}^{N} \Big[Q\Big]_k \left\{ \int_{z_{k-1}}^{z_k} \left\{ z \left(\begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \end{bmatrix} - \begin{bmatrix} d_{31}\mathcal{E}_3 \\ d_{31}\mathcal{E}_3 \\ 0 \end{bmatrix} \right) + z^2 \begin{bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix} \right\} \mathrm{d}z \right\} \tag{3.28}$$

Since the vectors of ε^0 and κ are the values of strain and curvature at the middle surface, they are not dependent on the vertical position in the laminate; they are independent of z and can thus be removed from under the integration. If we ignore for the moment the piezoelectric effect, we can rewrite equations 3.27 and 3.28 in the following manner:

$$\begin{bmatrix} N_x \\ N_y \\ N_{xy} \end{bmatrix} = \Big[A\Big] \begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \end{bmatrix} + \Big[B\Big] \begin{bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix} \tag{3.29}$$

$$\begin{bmatrix} M_x \\ M_y \\ M_{xy} \end{bmatrix} = \Big[B\Big] \begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \end{bmatrix} + \Big[D\Big] \begin{bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix} \tag{3.30}$$

The components of the matrices A, B and D are given by:

$$A_{ij} = \sum_{k=1}^{N} (Q_{ij})_k (z_k - z_{k-1}) \tag{3.31}$$

$$B_{ij} = \frac{1}{2} \sum_{k=1}^{N} (Q_{ij})_k (z_k^2 - z_{k-1}^2) \tag{3.32}$$

$$D_{ij} = \frac{1}{3} \sum_{k=1}^{N} (Q_{ij})_k (z_k^3 - z_{k-1}^2) \tag{3.33}$$

The matrix A is known as the extensional stiffness. This is because the components of A relate the normal force components N to the middle-surface strains ε^0. B is the coupling stiffness matrix — if its components are non-zero, then an extension implies an associated curvature. D relates the bending moments to the curvature, and is thus known as the bending stiffness matrix.

3.3.4 Special cases of the laminate stiffness matrices

Under the conditions of having only isotropic laminae, there are several general statements that can be made about the stiffness matrices A, B and D. If we allow the z components of the laminae to have a constant offset δ, and assume that each layer has the same Poisson's ratio ν, then we can re-write equation 3.32 as

$$B_{ij} = \frac{1}{2} \sum_{k=1}^{N} (Q_{ij})_k \left((z_k - \delta)^2 - (z_{k-1} - \delta)^2 \right) , \tag{3.34}$$

$$B_{ij} = \frac{1}{2} \sum_{k=1}^{N} (Q_{ij})_k \left(z_k^2 - 2z_k\delta - z_{k-1}^2 + 2z_{k-1}\delta \right) . \tag{3.35}$$

Then, using the form of Q_{ij} given in equation 3.11, we can see that if

$$B_{11} = \Sigma, \tag{3.36}$$

then

$$B_{22} = \Sigma, \tag{3.37}$$

$$B_{12} = B_{21} = \nu\Sigma, \tag{3.38}$$

and

$$B_{66} = \frac{1-\nu}{2}\Sigma. \tag{3.39}$$

We can find δ by solving $\Sigma = 0$, where Σ is given by:

$$\Sigma = \frac{1}{2}\sum_{k=1}^{N}(Q_{11})_k\left((z_k-\delta)^2-(z_{k-1}-\delta)^2\right). \tag{3.40}$$

The consequence of this is that we can always set $B_{ij} = 0$ if all the layers are isotropic, and have a common value for Poisson's ratio. This, in turn, implies that in this case, we can always find a surface that we can apply a transverse force N_x to which will induce a stretch, but no bending.

We also note that if the Poisson's ratio is the same for all layers, then the D_{ij} coefficients take on certain relationships, i.e.

$$D_{11} = D_{22} = \frac{D_{12}}{\nu} = \frac{D_{21}}{\nu} = \frac{2}{1-\nu}D_{66}. \tag{3.41}$$

3.3.5 Re-introduction of the piezoelectric effect

Now that we have this relationship between forces and moments and strains and curvatures, we are ready to reintroduce the effect of externally induced strains, namely thermal strains or piezoelectric strains. These induced strains create induced forces and moments, which, referring to equations 3.27 and 3.28 can be written as [69, 70]:

$$\begin{bmatrix} N_x^P \\ N_y^P \\ N_{xy}^P \end{bmatrix} = \int \left[Q\right]_k \begin{bmatrix} d_{31}\mathcal{E} \\ d_{31}\mathcal{E} \\ 0 \end{bmatrix}_k \mathrm{d}z \tag{3.42}$$

$$\begin{bmatrix} M_x^P \\ M_y^P \\ M_{xy}^P \end{bmatrix} = \int \left[Q\right]_k \begin{bmatrix} d_{31}\mathcal{E} \\ d_{31}\mathcal{E} \\ 0 \end{bmatrix}_k z\,\mathrm{d}z \tag{3.43}$$

As before, we can calculate the overall N^P and M^P vectors by performing the integrations over the layers. Assuming that there is no variation of $\mathcal{E}$ or d_{31} across a layer and that the layer is isotropic (its Q matrix is of the form given in equation 3.11), then we can say:

$$\begin{aligned} N_x^P = N_y^P &= \sum_{k=1}^{N}\mathcal{E}t_k\frac{E_k(d_{31})_k}{1-\nu_k} \\ N_{xy} &= 0 \end{aligned} \tag{3.44}$$

$$\begin{aligned} M_x^P = M_y^P &= \sum_{k=1}^{N}\mathcal{E}t_k\bar{z}_k\frac{E_k(d_{31})_k}{1-\nu_k} \\ M_{xy}^P &= 0 \end{aligned} \tag{3.45}$$

where t_k is the thickness of the k^{th} layer, and $\bar{z}_k$ is the mean position of that layer, i.e. $\frac{1}{2}(z_k - z_{k-1})$.

We can write the overall equation governing the behaviour of the laminate structure, incorporating the piezoelectric effect as follows:

$$\begin{bmatrix} N_x \\ N_y \\ N_{xy} \\ M_x \\ M_y \\ M_{xy} \end{bmatrix} = \begin{bmatrix} A & B \\ B & D \end{bmatrix} \begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \\ \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix} - \begin{bmatrix} N_x^P \\ N_y^P \\ 0 \\ M_x^P \\ M_y^P \\ 0 \end{bmatrix} \tag{3.46}$$

Our final goal is to find the values of κ, which tells us how sensitive our structure is, i.e. the amount of deflection per applied voltage we can expect. In order to find these values, we must invert equation 3.46 as follows:

$$\begin{bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \\ \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix} = \begin{bmatrix} A & B \\ B & D \end{bmatrix}^{-1} \left(\begin{bmatrix} N_x \\ N_y \\ N_{xy} \\ M_x \\ M_y \\ M_{xy} \end{bmatrix} + \begin{bmatrix} N_x^P \\ N_y^P \\ 0 \\ M_x^P \\ M_y^P \\ 0 \end{bmatrix} \right) \tag{3.47}$$

Under the conditions of no externally applied forces or moments, all the components N_i and M_i are zero, and so we have everything we need to calculate values of κ for any laminated structure with piezoelectric layers.

3.3.6 Thermal bending of laminates

The extension of this approach to incorporate thermal effects is simple: we must calculate N^T and M^T in the same way that we calculated N^P and M^P. If we assume that there are no residual strains when the laminate is at the temperature at which it was glued together, then the thermal moments and forces are given by:

$$\begin{bmatrix} N_x^T \\ N_y^T \\ N_{xy}^T \end{bmatrix} = \int \left[Q\right]_k \begin{bmatrix} \alpha_x \\ \alpha_y \\ \alpha_{xy} \end{bmatrix}_k \Delta T \, \mathrm{d}z \tag{3.48}$$

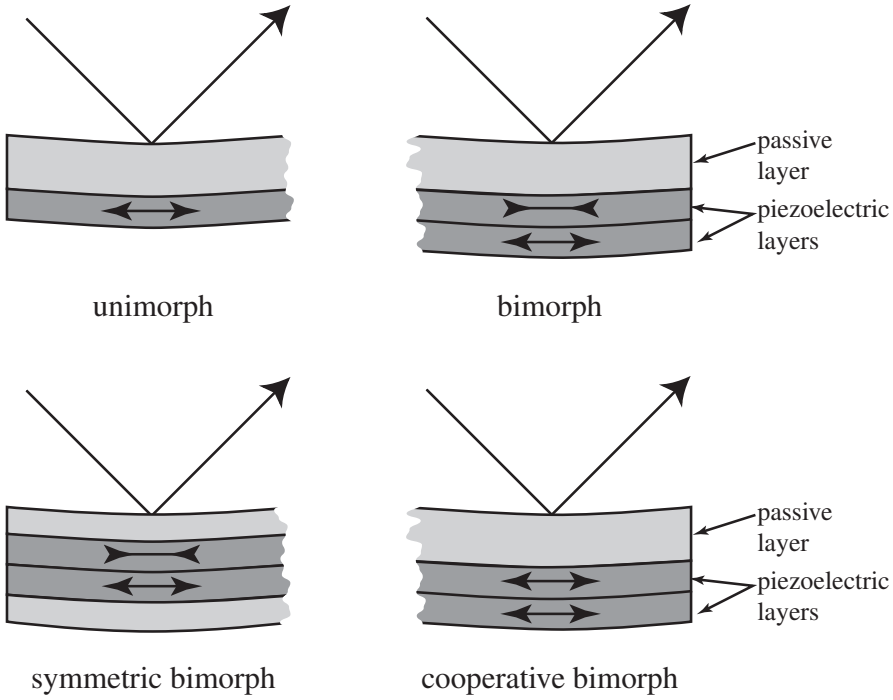
$$\begin{bmatrix} M_x^T \\ M_y^T \\ M_{xy}^T \end{bmatrix} = \int \left[Q\right]_k \begin{bmatrix} \alpha_x \\ \alpha_y \\ \alpha_{xy} \end{bmatrix}_k \Delta T z \, \mathrm{d}z \tag{3.49}$$

where ΔT is the difference in temperature from the adhesion temperature. In an isotropic material, $\alpha_x = \alpha_y$, and $\alpha_{xy} = 0$.

3.4 Evaluating Laminate Curvatures

We are now in a position to calculate the expected curvature of various possible bimorph-type devices, and hence find their voltage sensitivity. In all the cases that follow, the values of κ are calculated from the theory developed above. The algebraic manipulations were all performed using the *Mathematica* system, and hence none of the intermediate stages are shown, since these tend to be somewhat unwieldy, and are, in any case, of little interest. Initially, four cases we shall consider are the unimorph and bimorph devices, the "cooperative" bimorph and the symmetric bimorph. These

Figure 3.7 Unimorph and bimorph mirrors in cross section



arrangements are shown in figure 3.7. A unimorph consists of an active piezo layer bonded to a passive layer. The passive layer then becomes the mirror surface. A bimorph consists of two layers of piezo acting in opposition. There may or may not be a third layer in a bimorph which acts as the mirror surface. The "cooperative" bimorph has two layers of piezo material acting in the same direction — doubling the effective thickness of the piezo material without doubling the maximum applicable voltage. The symmetric bimorph has two layers of piezo sandwiched between two passive layers. This arrangement does not bend when its temperature is changed.

Using equation 3.47, we can calculate the curvatures for the each structure. The matrix inversion is quite involved, and in some circumstances, it is useful to note the inversion procedure detailed in [69], section 4.4.1.

3.4.1 Calculated curvatures

The curvature for a unimorph actuator is found to be:

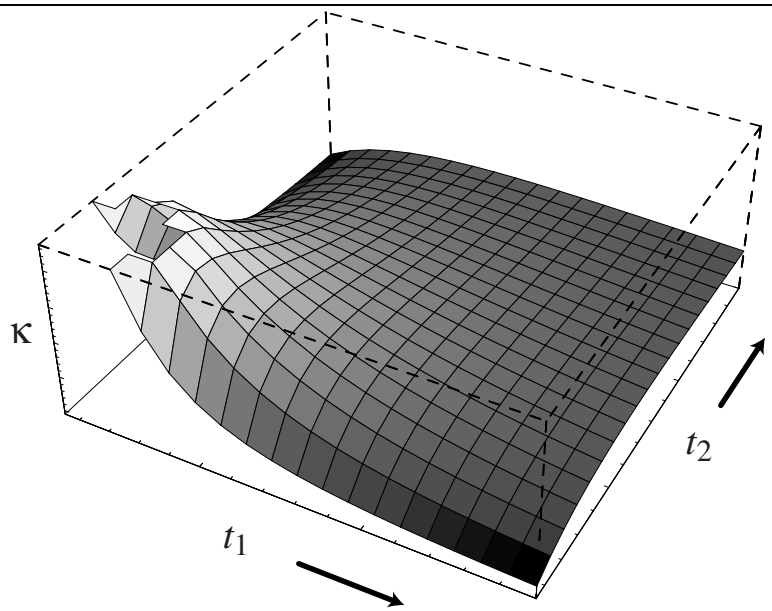
$$\kappa_{\text{uni}} = \frac{-6 d_{31} V E_1 E_2 t_2 \,(t_1 + t_2)\,(\nu_1 - 1)\,(\nu_2 - 1)}{E_2{}^2 t_2{}^4 (\nu_1 - 1)^2 + 2 E_1 E_2 t_1 t_2 \,(2 t_1{}^2 + 3 t_1 t_2 + 2 t_2{}^2)\,(\nu_1 - 1)\,(\nu_2 - 1) + E_1{}^2 t_1{}^4 (\nu_2 - 1)^2}\,, \tag{3.50}$$

where V is the applied voltage, the E_i are Young's moduli, the t_i are layer thicknesses, d_{31} is the relevant piezoelectric constant and the ν_i are Poisson's ratios; the subscript 1 refers to the active piezo layer and the subscript 2 refers to the passive layer. This expression can be simplified somewhat if we assume (as is usually the case) that the Poisson's ratios are the same for both materials, i.e. $\nu_1 = \nu_2$:

$$\kappa_{\text{uni}} = \frac{-6 d_{31} V E_1 E_2 t_2 \,(t_1 + t_2)}{E_1{}^2 t_1{}^4 + 4 E_1 E_2 t_1{}^3 t_2 + 6 E_1 E_2 t_1{}^2 t_2{}^2 + 4 E_1 E_2 t_1 t_2{}^3 + E_2{}^2 t_2{}^4}\,. \tag{3.51}$$

From equation 3.51, we can see that the curvature is proportional to the applied voltage and also to the d_{31} parameter of the particular piezoelectric material. Therefore, once a piezoelectric material is

Figure 3.8 Variation of κ with t_1 and t_2 for a unimorph with $E_1 = E_2$



chosen, it is the thickness of the layers and their Young's moduli that determine how much curvature is obtainable, and it is these parameters that we must optimise when designing a unimorph device. The behaviour of κ_{uni} is shown in figure 3.8, which shows the main features of the sensitivity of a unimorph structure.

In order to see more clearly the dependance upon the parameters of the structure, we can express the curvatures in terms of the ratio of the thicknesses, $r = t_2/t_1$ and the ratio of the Young's moduli, $k = E_2/E_1$, as in reference 71. In terms of these ratios, and also upon application of the maximum voltage $V_{\max} = t_1 \mathcal{E}_{\max}$, equation 3.51 becomes:

$$\kappa = \left(\frac{d_{31}\mathcal{E}_{\max}}{t_1}\right) \times \frac{6kr(1+r)}{1+k^2r^4+2kr(2+3r+2r^2)}. \tag{3.52}$$

For a cooperative bimorph, we note that we have effectively doubled the applied voltage and the piezo thickness, and so if the thickness of each layer of piezo is t_1, then the curvature is given by:

$$\kappa_{\text{coop}} = \left(\frac{d_{31}\mathcal{E}_{\max}}{t_1}\right) \times \frac{12kr(2+r)}{16+k^2r^4+8kr(4+3r+r^2)}. \tag{3.53}$$

For a standard bimorph, we can calculate the curvature to be:

$$\kappa_{\text{bi}} = \left(\frac{d_{31}\mathcal{E}_{\max}}{t_1}\right) \times \frac{12(2+kr)}{16+k^2r^4+8kr(4+3r+r^2)}. \tag{3.54}$$

Finally, we can calculate the curvature of a symmetric bimorph to be:

$$\kappa_{\text{sym}} = \left(\frac{d_{31}\mathcal{E}_{\max}}{t_1}\right) \times \frac{3}{2+2kr(3+3r+r^2)}. \tag{3.55}$$

In each of the above cases, the factor $d_{31}\mathcal{E}_{\max}/t_1$ is an overall scaling, which shows us that a thinner structure will always bend more. The remaining factors are plotted in figure 3.9 for several values of k. We can see that in the case $k \approx 1$, unless the ratio of thicknesses is greater than about 3 or less than about 0.5, then the best choice of structure is the unimorph. If a ratio of less than 0.5 is acheivable, then the bimorph is clearly more sensitive (this implies a very thin passive layer, which may be difficult to achieve in practice). However, if the value of k is small, we can see that the bimorph,

and even the symmetric bimorph may be considered. A small value of k implies that the passive layer is much more easily deformable than the active layer, and thus may not be easily polishable — most glasses have Young's moduli that are approximately equal to that of piezoceramic, in the range around $80\,(GPa)$. If the passive layer is relatively very stiff, then it can be seen that the unimorph or even cooperative bimorph are the most suitable structures. This case for k may exist if piezoelectric film is used, as these materials have very low Young's moduli.

3.4.2 Comparison with published curvatures

It is interesting to compare these results with those previously published for such devices. Several authors have published either expected curvatures, or scale factors for the differential equation governing surface deflection. Comparisons will be made between the results obtained here, and those found in references 41 (SRL), 52 (KS), 51 (IS) and 72 (Ro).

SRL only considers the case of a simple unimorph, and obtains the following result for the unimorph curvature:

$$\kappa_{SRL} = \frac{12 V d_{31} (t_1 + t_2)}{t_1^3 k} \tag{3.56}$$

$$k = 4 + 6\left(\frac{t_2}{t_1}\right) + 4\left(\frac{t_2}{t_1}\right)^2 + \frac{E_2 t_2^3 (1-\nu_1)}{E_1 t_1 (1-\nu_2)} + \frac{E_1 t_1 (1-\nu_1)}{E_2 t_2 (1-\nu_2)}. \tag{3.57}$$

This result is a simple modification of that in Ro, except that it has been modified to account for piezoelectric rather than thermal expansion. This can be rearranged into the same form as equation 3.50, which confirms that the theory developed here is correct.

KS also only considers the case of a simple unimorph, with Poisson's ratio the same for each layer. The proportionality constant ζ they derived is:

$$\zeta = \frac{d_{31} E_1 (2\Delta_1 t_1 - t_1^2)}{2D'(1-\nu)}, \tag{3.58}$$

where

$$D' = \frac{E_2}{1-\nu^2}\left(\frac{\Delta_1^3}{3} + \frac{\Delta_2^3}{3} - \Delta_1^2 t_1 + \Delta_1 t_1^2 - \frac{t_1^3}{3}\right) + \frac{E_1}{1-\nu^2}\left(\Delta_1^2 t_1 - \Delta_1 t_1^2 + \frac{h_1^3}{3}\right), \tag{3.59}$$

$$\Delta_1 = t_1 + t_2 - \Delta_2, \tag{3.60}$$

$$\Delta_2 = \frac{E_2 h_2^2 + E_1\left((t_1+t_2)^2 - t_1^2\right)}{2\left(E_2 t_2 + E_1 t_1\right)} \tag{3.61}$$

After much manipulation, we finally arrive at the following expression:

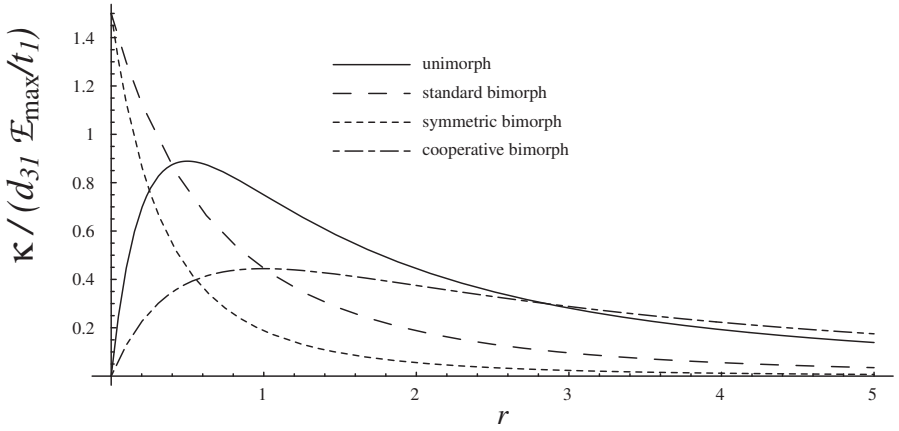
$$\zeta = \frac{6\,(1+\nu)\,d_{31}\,E_1 t_1\,(t_1+t_2)\,(E_1\,(t_1-t_2)+E_2 t_2)}{{E_2}^2 {t_2}^4 + 2E_1 E_2 t_1 t_2\,(2{t_1}^2 + 3t_1 t_2 + 2{t_2}^2) + {E_1}^2\,(4{t_1}^4 - 6{t_1}^2{t_2}^2 + 3{t_2}^4)} \tag{3.62}$$

This expression is in agreement with the proportionality constant given below in equation 3.93, providing the z coordinate is offset by the correct amount[2].

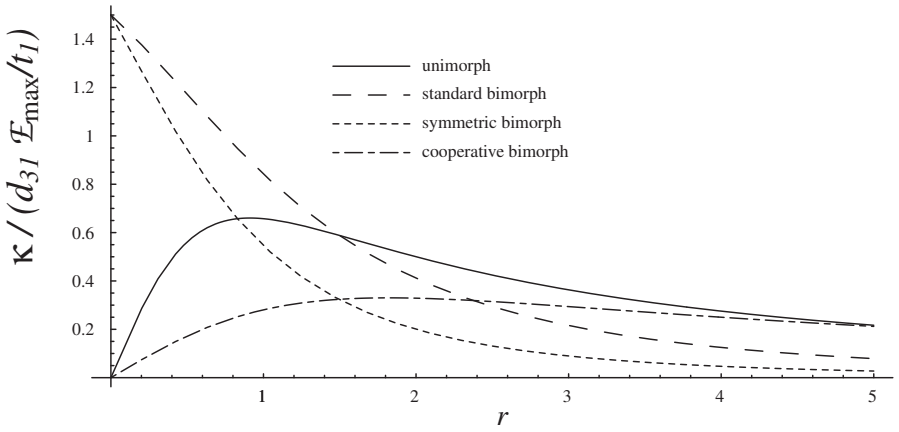
IS considers the cases of ideal unimorph and bimorph actuators, and also the case of a device with a non-negligible adhesive layer. The results derived in this work are presented in a rather

[2]Offsetting the z coordinate does not affect the calculated curvature, although it does affect the strain references ε_0

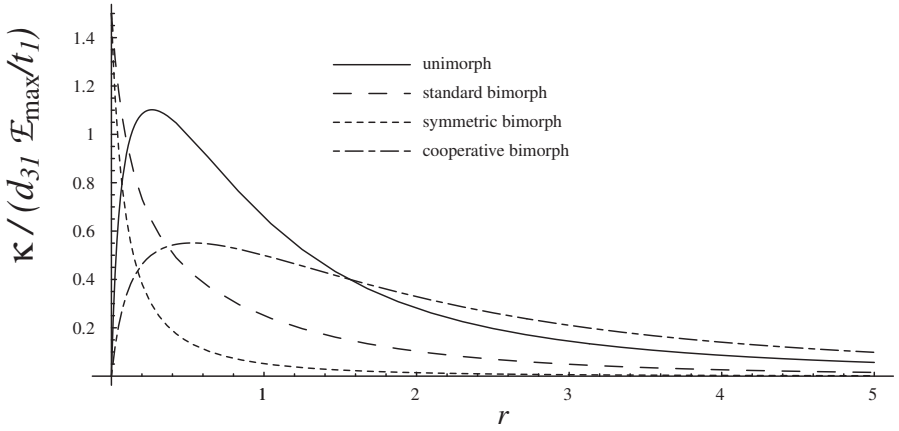
Figure 3.9 Variation of the maximum curvature scaling parameter as a funcition of the thickness ratio r for several values of the Young's moduli ratio k, for each type of deformable structure



(a) $k = 1$



(b) $k = 1/4$



(c) $k = 4$

different form, so a comparison of only one result will be made. The derived value for the radius of curvature for a unimorph device found, in the paper's original notation is:

$$R_{\mathrm{IS}} = \frac{(G_1+E_1)(G_2+E_2)-(\nu_g G_1+\nu_e E_1)(\nu_g G_2+\nu_e E_2)}{(G_2+E_2)M-(\nu_g G_1+\nu_e E_1)M}, \tag{3.63}$$

where

$$M = VEd_{31}t\left(s-a+\frac{t}{2}\right), \tag{3.64}$$

$$G'' = \frac{G}{1-\nu_g^2}, \tag{3.65}$$

$$E'' = \frac{E}{1-\nu_g^2}, \tag{3.66}$$

$$G_i = (G''/3)\left((s-a)^3+a^3\right), \tag{3.67}$$

$$E_i = (E''/3)\left((s-a+t)^3-(s-a)^3\right). \tag{3.68}$$

The distance a depends upon the location of the neutral surface, and is found to be

$$a = \frac{G's^2+E'(t^2+2st)}{2(G's+E't)}, \tag{3.69}$$

where

$$G' = G/(1-\nu_g), \quad E' = E/(1-\nu_e). \tag{3.70}$$

To convert this to standard notation, we must make the following transpositions: $t \to t_1$, $E \to E_1$, $G \to E_2$, $\nu_e \to \nu_1$, $\nu_g \to \nu_2$, $s \to t_2$. Making these substitutions, replacing voltage V for electric field V, and after some manipulation, we arrive at the following expression for the curvature (reciprocal of the radius of curvature):

$$\kappa_{\mathrm{IS}} = \frac{-6d_{31}E_1E_2t_2\,(t_2+t_1)\,V\,(-1+\nu)}{E_2^2t_2^4+E_1^2t_1^4+2E_1E_2t_2t_1\,\left(2t_2^2+3t_2t_1+2t_1^2\right)} \tag{3.71}$$

Here, in order to make the discrepancy clear, the Poisson's ratios have been made equal — this expression still depends upon ν, whereas it should not, according to equation 3.50. The cause of this discrepancy is believed to be that IS used an incorrect form for the moment caused by the expansion of the piezoelectric layer. If we compare equation 3.64 with equation 3.45, we can see that a factor of $1/(1-\nu)$ has been omitted — this is precisely the discrepancy noted above. This error is introduced by failing to account fully for the 2-dimensional effect of the piezoelectrically induced strain.

3.4.3 Maximising sensitivity

There are essentially two situations we need to consider when designing a unimorph device. The first is when we have a given thickness of piezoelectric material, and we must choose the thickness of the passive layer. The second case is when the overall thickness of the device is constrained, and we must choose the ratio of thicknesses of the two layers. Due to polishing considerations, this second case is the more practical case when considering constructing a unimorph mirror — it becomes difficult to polish to optical flatness a device whose aspect ratio (diameter:thickness) is much greater than about 10. We would like to find the maximum possible curvature, and this happens when the voltage is the maximum that can be applied to the layer of piezo. This voltage depends upon the thickness of the layer, $V_{\max} = \mathcal{E}_{\max}t$. Therefore, we need to solve

$$\frac{\partial(\kappa t_1)}{\partial\alpha} = 0, \tag{3.72}$$

where α is the free variable in the optimisation. So, for a unimorph in which both layers have the same Poisson's ratio, the two cases are:

- $\alpha = t_2$, i.e. fixed thickness of piezo. In this case, the complete expression for the maximum curvature is too unwieldy to reproduce here; the condition on t_2 is:

$$t_2 = \frac{1}{2E_2}\left\{ \left(-E_2^2(E_2-2E_1)t_1^3 + 2\sqrt{E_1E_2^4t_1^6(E_1-E_2)} \right)^{1/3} + E_2t_1 \left(\frac{E_2t_1}{\left(-E_2^2(E_2-2E_1)t_1^3 + 2\sqrt{E_1E_2^4t_1^6(E_1-E_2)} \right)^{1/3}} - 1 \right) \right\} \quad (3.73)$$

  In the case where $E_1 = E_2$, this relationship reduces to $t_2 = t_1/2$, and the curvature becomes $\kappa_{\text{uni,max}} = -\frac{8d_{31}\mathcal{E}_{\max}}{9t_1}$.

- $\alpha = t_1 + t_2$, i.e. the overall thickness of the device is constrained. In this case, we find much simpler expressions both for the optimum curvature and the optimum condition:

$$E_1t_1^2 = E_2t_2^2\,, \quad (3.74)$$

$$\kappa_{\text{uni,max}} = \frac{-3\,d_{31}\sqrt{E_1}\,\mathcal{E}_{\max}}{2\left(\sqrt{E_1}+\sqrt{E_2}\right)t_2}\,. \quad (3.75)$$

The optimisation of a bimorph device (either standard type or symmetric type) is much simpler. The maximum curvature occurs when the thickness of the passive layer is zero — the curvature always increases as the thickness of the passive layer decreases. In this case, the curvature of the structure is given by:

$$\kappa_{\text{bi,max}} = -\frac{3d_{31}\mathcal{E}_{\max}}{2t_1} \quad (3.76)$$

3.4.4 Structures with non-negligible adhesive layers

The theory developed above can also be applied to structures with an adhesive layer in between the piezo and the passive substrate. The full expression for a unimorph actuator with a layer of adhesive in between the piezo and the substrate can be found analytically but this expression is too unwieldy to reproduce here. Figure 3.10 shows the curvatures of an ideal unimorph, and one with a thin layer of adhesive as a function of substrate thickness. The parameters of the unimorph are given in table 3.3. The piezo and substrate parameters are those of some samples used in the later experimental investigations, and the parameters of the adhesive layer are taken from reference 51.

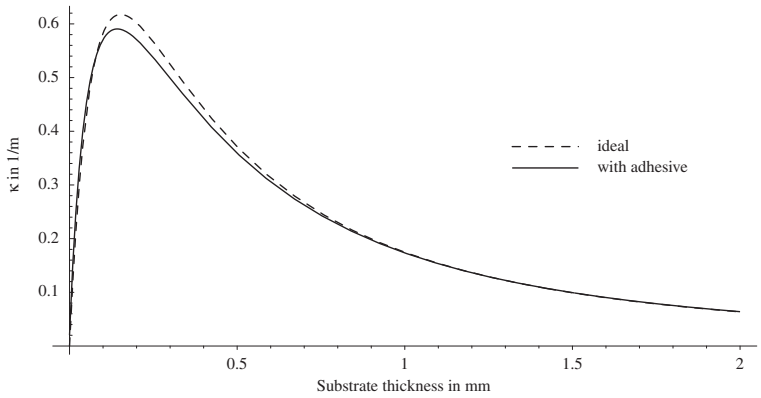
As can be seen from figure 3.10, the effect of this adhesive layer is barely noticeable at large values of substrate thickness, and it's effect is small even when the substrate is comparable in thickness to the layer of piezo.

Although the adhesive layer may seem to play a fairly insignificant rôle in determining the device sensitivity, the theory is misleading in this case. This is because some of the the fundamental assumptions of the theory are no longer good approximations. If the Young's modulus is much lower than that of the other layers (which it may well be), then the approximation that a straight line perpendicular to the surface remains straight under bending ceases to be true. In general, an adhesive with a low modulus will tend to absorb the expansion of the piezoelectric layer, without passing the effect on to the other layers, and thereby reducing sensitivity. It is for this reason that a relatively stiff, thin layer of adhesive is desirable.

Table 3.3 Parameters of unimorphs in figure 3.10

Parameter	Value
E_1	70×10^9 Pa
E_2	64×10^9 Pa
E_a	3×10^9 Pa
t_1	0.3 mm
t_a	30 μm
V	300 V
d_{31}	-210×10^{-12} mV^{-1}

Figure 3.10 The effect of an adhesive layer on κ



3.5 Temperature Dependent Curvatures

The temperature curvatures of bimorph and unimorph devices are derived in exactly the same way as the piezoelectric bending coefficients. The only difference is that the important parameter is now the difference in thermal expansion coefficient α between the piezoceramic layers and the passive layers.

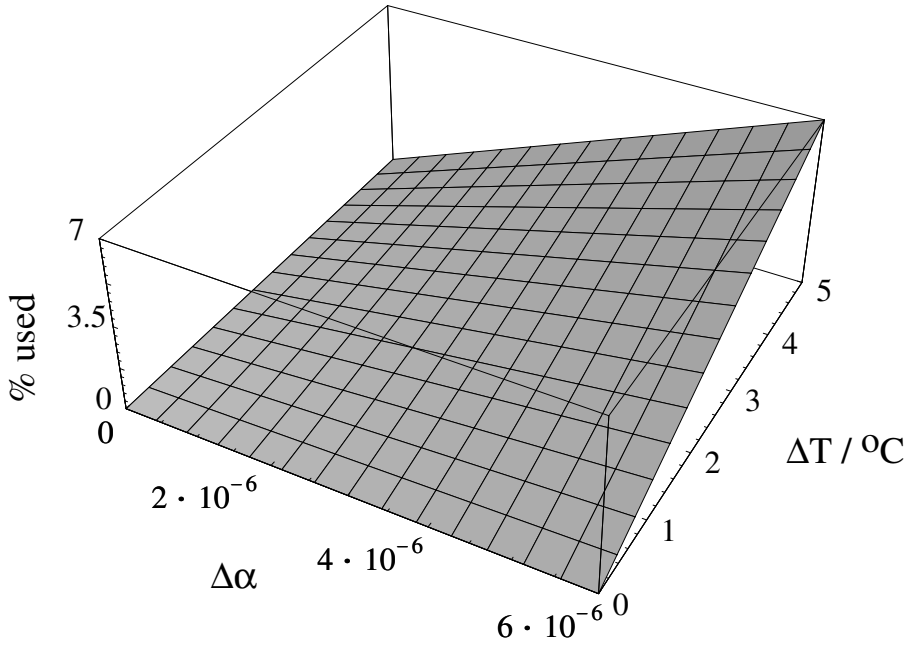
If we denote the (isotropic) thermal expansion coefficient of the active layer as α_1, and α_2 for the passive layer, then we can simply arrive at the thermal analogue of equation 3.51 by replacing $d_{31}V$ by $(\alpha_1 - \alpha_2)t_1\Delta T$. The extra factor of t_1 occurs because the piezoelectrically induced strain depends upon the electric field present. The temperature difference ΔT is simply the difference between the applied temperature, and the temperature at which the device is flat — usually the temperature at which the layers were originally bonded together. For the cases of cooperative and standard bimorphs, the thermal bending coefficients are, of course, the same. In this case, we must make the same replacement as before in equation 3.53.

The curvature of the symmetric bimorph remains identically zero when its temperature is changed from the temperature at which it was bonded. Under certain circumstances, this may be a very useful property, however the lack of sensitivity compared to other arrangements will usually render this arrangement inappropriate.

An alternative consideration to the amount of curvature induced by a given change in temperature is the amount of voltage required to nullify such a change. This may be calculated quite simply by considering the piezoelectric expansion required to compensate for the differential expansion of the two layers. Equating these quantities, we find:

$$\frac{d_{31}V_{\Delta T}}{t_1} = (\alpha_2 - \alpha_1)\Delta T\,, \tag{3.77}$$

Figure 3.11 Percentage of actuator range used in compensating for a shift in temperature



from which we see that the voltage required is simply:

$$V_{\Delta T} = \frac{t_1}{d_{31}}(\alpha_2 - \alpha_1)\Delta T\,. \tag{3.78}$$

We can express this result in terms of the maximum stroke of the mirror by considering the ratio of this voltage to the maximum voltage that can be applied. This maximum voltage depends on the maximum field, $\mathcal{E}_{\max}$ and is simply $V_{\max} = t_1\mathcal{E}_{\max}$. So, from equation 3.78, we obtain:

$$\frac{V_{\Delta T}}{V_{\max}} = \frac{(\alpha_2 - \alpha_1)\Delta T}{d_{31}\mathcal{E}_{\max}} \tag{3.79}$$

The fraction of the dynamic range thereby used is simply half the modulus of this expression, since voltages can be applied in either direction.

Typical values for α_1 for piezoceramic materials such as PZT are in the range 2×10^{-6} to 4×10^{-6}, whereas for the passive layer, the expansivity may be anything from 3×10^{-6} for pyrex, to around 9×10^{-6} for typical glass, or even as low as 0.4×10^{-6} for fused silica [73]. The percentage of the maximum voltage used in compensating for a temperature change is shown in figure 3.11, for typical piezoceramic parameters $d_{31} = 200\times10^{-12}\,\mathrm{mV}^{-1}$ and $\mathcal{E}_{\max} = 10^6\,\mathrm{Vm}^{-1}$. It is easily seen that with a poor choice of passive material (i.e. one whose thermal expansivity differs significantly from that of the active layer), a temperature shift of just 5^oC could cause 7% of the dynamic range to be used.

3.6 Resonant Frequency of a Plate

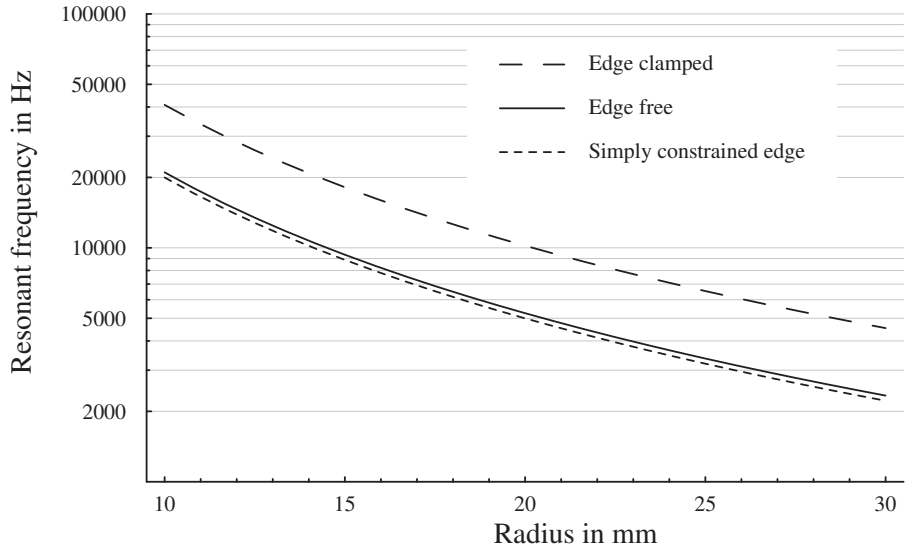
The derivation of the relationships governing the resonant frequencies of a plate under varying mounting arrangements is beyond the scope of the current work, and the result will simply be quoted from Roark [72]. The general form is this:

$$f = K\frac{1}{2\pi}\sqrt{\frac{Dg}{wr^4}}\,, \tag{3.80}$$

Table 3.4 Values of the scaling constant K which for different boundary conditions

Boundary conditions	K
Simply constrained	4.99
Edge clamped	10.2
Edge free	5.25

Figure 3.12 Resonant frequency for a piezo-on-glass unimorph plate, as a function of actuator radius



where D is the effective elastic constant for the plate, K is a constant that depends on the particular boundary conditions, g is the acceleration due to gravity and w is the loading per unit area of the plate, including its own weight. In the case of no other loading, this becomes simply:

$$w = g \sum \rho_k t_k \,, \tag{3.81}$$

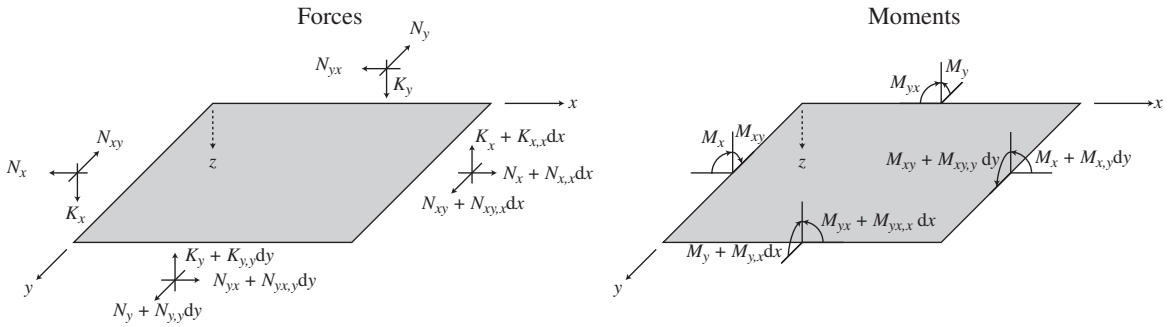
where ρ_k is the density of each layer. Values of the constant K are shown in table 3.6. For typical actuator dimensions of $t_1 = 0.3\,\text{mm}$ and $t_2 = 1.5\,\text{mm}$, and material parameters typical for a piezo-on-glass unimorph, figure 3.12 shows the first resonant frequency as a function of mirror radius for each of the boundary conditions mentioned above.

3.7 Derivation of the Biharmonic Equation

To find the relationship between the vertical displacement $w(x,y)$ of the surface of a unimorph or bimorph structure, we need to consider the equilibrium state of an element of the laminate. Referring to figure 3.13, we see the forces and moments acting on our element; we use the notation that subscripts after the comma denote differentiation with respect to that variable. We also have a load term, p, per unit area acting vertically downwards (i.e. in the positive z direction). We note that in order to derive the equilibrium conditions, we need to be aware that previously, we have ignored the transverse shear stresses, τ_{xz} and τ_{yz}, however we need now to consider their integrated resultants, K_x and K_y in order to balance the loading force. These apparently contradictory approximations are commonly used in thin plate theory, and are known to give good results [5, 69]. Firstly, we can consider forces acting in the z direction:

$$K_y\,\mathrm{d}x + K_{y,y}\,\mathrm{d}y\,\mathrm{d}x + K_x\,\mathrm{d}y + K_{x,x}\,\mathrm{d}x\,\mathrm{d}y + p\,\mathrm{d}x\,\mathrm{d}y = K_x\,\mathrm{d}y + K_y\,\mathrm{d}x\,, \tag{3.82}$$

Figure 3.13 All forces and moments acting on a portion of laminate in equilibrium



hence

$$K_{x,x} + K_{y,y} + p = 0\,. \tag{3.83}$$

We can also take moments about the x and y axes. Considering first the x axis:

$$M_y\,\mathrm{d}x + M_{xy}\,\mathrm{d}y + M_{xy,x}\,\mathrm{d}x\,\mathrm{d}y = M_{xy}\,\mathrm{d}y + M_y\,\mathrm{d}x + M_{y,y}\,\mathrm{d}y\,\mathrm{d}x + K_y\,\mathrm{d}x\,\mathrm{d}y\,, \tag{3.84}$$

where we have neglected terms such as the moment of the force p as a small quantity. Canceling terms, we have

$$M_{y,y} - M_{xy,x} = K_y\,. \tag{3.85}$$

Similarly, taking moments about the y axis gives us

$$M_{yx}\,\mathrm{d}x + M_{yx,y}\,\mathrm{d}y\,\mathrm{d}x + M_x\,\mathrm{d}y + M_{x,x}\,\mathrm{d}x\,\mathrm{d}y = M_x\,\mathrm{d}y + M_{yx}\,\mathrm{d}x + K_x\,\mathrm{d}y\,\mathrm{d}x\,, \tag{3.86}$$

whence

$$M_{yx,y} + M_{x,x} = K_x\,. \tag{3.87}$$

We know that stress is a symmetric tensor (see section 3.1.1), so we know that $\tau_{xy} = \tau_{yx}$. From the definitions of M_{xy} and M_{yx} (equation 3.26), and referring to figure 3.2, we can see that $M_{xy} = -M_{yx}$. This allows us to differentiate equations 3.85 and 3.87, and substitute into equation 3.83 to give us

$$M_{x,xx} + M_{y,yy} - 2M_{xy,xy} = -p\,. \tag{3.88}$$

Now we must relate the bending moments to the deflection of the plate.

Referring to equation 3.46, we can see that the moments depend upon the B_{ij} coefficients, the D_{ij} coefficients, the curvatures κ, the strains ε and the piezoelectrically induced forces and moments, N^P and M^P. If we make the assumptions that the Poisson's ratios are the same for all layers of the laminate, then we can simplify the D_{ij} coefficients, and make the B_{ij} coefficients zero, as described in section 3.3.4. Therefore, we have the much simplified relationships:

$$\begin{aligned} M_x &= D\kappa_x + \nu D\kappa_y - M_x^P\,, \\ M_y &= \nu D\kappa_x + D\kappa_y - M_y^P\,, \\ M_{xy} &= \frac{1-\nu}{2} D\kappa_{xy}\,. \end{aligned} \tag{3.89}$$

Using the expressions for the components of κ in terms of the deflection of the plate $w(x,y)$ (equation 3.22), we can say that

$$\begin{aligned} M_{x,xx} &= -Dw_{,xxxx} - \nu Dw_{,xxyy} - M^P_{x,xx} \\ M_{y,yy} &= -\nu Dw_{,xxyy} - Dw_{,yyyy} - M^P_{y,yy} \\ M_{xy,xy} &= -\frac{(1-\nu)}{2} D(-2w_{,xxyy}) \end{aligned} \tag{3.90}$$

where $D = D_{11}$. Combining equations 3.90 with equation 3.88, we can say that

$$D(w_{,xxxx} + w_{,yyyy} + 2w_{,xxyy}) = p - M^P_{x,xx} - M^P_{y,yy}\,. \tag{3.91}$$

All that remains to be done now is to substitute expressions for the derivatives of the piezoelectrically induced moments. These moments are defined in equation 3.45. If we assume that the electric field is a function of x and y only, and there is no applied force p, then we can say that

$$\nabla^4 w(x,y) = \zeta \nabla^2 \mathcal{E}\,(x,y)\,. \tag{3.92}$$

The operator on the left hand side of this expression is known as the *biharmonic* operator, and so the surface shape of a bimorph or unimorph mirror takes on the shape of a solution of the biharmonic equation, with the forcing term being the Laplacian of the applied electric field. The proportionality constant ζ is calculated to be:

$$\zeta = \frac{6\,(1+\nu)\,d_{31}\,e_1\,t_1\,t_2}{e_2\,t_2\,(3{t_1}^2 + {t_2}^2) + e_1\,t_1\,({t_1}^2 + 3{t_2}^2)} \tag{3.93}$$

These curvatures can be used to scale the solutions to equation 3.92 if necessary.

3.8 Solution of the Biharmonic Equation

Having demonstrated that the surface of a bimorph mirror is related to the applied voltage by the biharmonic equation, we need to construct a solution to this equation that satisfies the relevant boundary conditions. The deflection of the bimorph plate will be the same as that for a simple plate but with an effective elastic constant D and Poisson ration ν which can be calculated from the material parameters of the two layers of the bimorph structure [72]. In each case, we will use the same series solution [5, 74] to calculate the response of the surface under the influence of a force applied at a point. Following Timoshenko [5], we can write the series solution in cylindrical polar coordinates as:

$$w(r,\theta) = \sum_{n=0}^{\infty} w_n \cos\left(n(\theta - \phi)\right) \tag{3.94}$$

where (r,θ) is the point on the surface, ϕ is the angular position of the applied load and the w_n are given by

$$w_0 = A_0 + B_0 r^2 + C_0 \ln(r) + D_0 r^2 \ln(r), \tag{3.95}$$

$$w_1 = A_1 r + B_1 r^3 + C_1 r^{-1} + D_1 r \ln(r), \tag{3.96}$$

$$w_n = A_n r^n + B_n r^{n+2} + C_n r^{-n} + D_n r^{(n-2)}. \tag{3.97}$$

The A_n etc. are all constants of integration, and need to be fitted to the boundary conditions. To create a complete solution for the surface of the mirror, we need to split it into 2 regions, that within and that without the radius at which the load is being applied. This point of loading will be at (ρ,ϕ). We will use a tilde over the constants of integration for the outer region to distinguish them from those of the inner region.

3.8.1 Boundary conditions

As mentioned above, we need a series of boundary conditions in order to fit the constants of integration. In fitting the constants, the orthogonality of the cosine function allows us to consider only one value of n at a time. There are several possible boundary conditions that must be considered at the inner and outer boundaries of the mirror corresponding to various mounting arrangements of the mirror. For the inner edge, the two possibilities are:

- No special condition, other than the response function must remain finite and continuous
- Clamped inner edge, i.e. the value of the function and its first radial derivative must be zero at some inner clamping radius

For the outer edge, there are three possible options:

- Clamped outer edge, i.e. the value of the function and its first radial derivative are both zero at the outer edge
- "Free" outer edge – this is the Kirchhoff boundary case
- Simple boundary condition – in this case, the deflection of the surface is zero and there is no radial moment applied at the edge.

The Kirchhoff boundary conditions for a free edge are:

- That there are no bending moments about the free edge,
- the twisting moment along the free edge is balanced by the shear force.

These conditions can be expressed in polar coordinates in terms of the surface deflection as [5]:

$$M_r = 0 \tag{3.98}$$

$$Q_r = \frac{\partial M_{rt}}{r\partial\theta} \tag{3.99}$$

where M_r is the radial part of the moment, and is given by

$$M_r = -\frac{D}{a^2}\left[(1-\nu)\frac{\partial^2 w}{\partial r^2} + \nu\nabla^2 w\right], \tag{3.100}$$

where D is the effective elastic constant of the plate, a is the radius of the plate and ν is the effective value of Poisson's constant. The shear force Q_r is given by

$$Q_r = -\frac{D}{a^3}\frac{\partial}{\partial r}(\nabla^2 w) \tag{3.101}$$

and M_{rt}, the twisting moment along the edge is given by

$$M_{rt} = \frac{D}{a^2}(1-\nu)\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial w}{\partial\theta}\right) \tag{3.102}$$

At the loading circle, $r = \rho$, we require that the deflection of the surface and its first and second radial derivatives are continuous. We can write this condition as:

$$w(\rho,\theta) = \tilde{w}(\rho,\theta), \tag{3.103}$$

$$\frac{dw}{dr} = \frac{d\tilde{w}}{dr}, \tag{3.104}$$

$$\frac{d^2 w}{dr^2} = \frac{d^2\tilde{w}}{dr^2}. \tag{3.105}$$

The last of these equations is a consequence of considering the balance of the shear forces on either side of the loading circle. All that remains is to take account of the force being applied. The shear force in a plate is continuous everywhere, except at the point of loading. We represent the load as a Fourier series:

$$\frac{P}{\pi\rho a}\left(\frac{1}{2}+\sum_{n=1}^{\infty}\cos(n(\theta-\phi))\right), \tag{3.106}$$

and the equation for shear force balance becomes:

$$\left[Q_r-\tilde{Q}_r\right]_{r=\rho}=\frac{P}{\pi\rho a}\left(\frac{1}{2}+\sum_{n=1}^{\infty}\cos(n(\theta-\phi))\right). \tag{3.107}$$

3.8.2 Application of boundary conditions

Now that we have all these equations, we can calculate all of the constants of integration by considering each azimuthal order (i.e. each value of n) separately, as mentioned above. We will consider three different mounting arrangements for the bimorph mirror:

1. Outer edge clamped,
2. outer edge free and
3. outer edge simply supported.

For each of these cases, there is a different set of constants of integration. These constants are found by solving three sets of 8 simultaneous equations – one for $n=0$, another for $n=1$ and a final set for the general case $n\geq 2$. Some of the sets of simultaneous equations will be common to more than one set of boundary conditions. Some of these simultaneous equations can be compared to those presented by Mehta in reference [74]. Once we have a set of constants of integration, we have everything that we need to calculate the surface deflection due to a point load for the set of boundary conditions.

Inner region continuous

In all the cases considered, there is no central hole, so we require that the solution remains finite as $r\to 0$. This immediately implies that:

$$C_n=D_n=0 \quad \forall \quad n. \tag{3.108}$$

Continuity at the loading circle

There are three equations which apply at $r=\rho$, the radius at which the load is being applied. Introducing the notation $\widetilde{\mathrm{LHS}}$ as referring to the left hand side of the equation, except with A_n replaced by $\tilde{A}_n$ etc., Equation 3.103 leads directly to the following conditions:

$$A_0+B_0\rho^2+C_0\log(\rho)+D_0\rho^2\log(\rho)=\widetilde{\mathrm{LHS}} \tag{3.109}$$

$$A_1\rho+B_1\rho^3+C_1r^{-1}+D_1\rho\log(\rho)=\widetilde{\mathrm{LHS}} \tag{3.110}$$

$$A_n\rho^n+B_n\rho^{n+2}+C_n\rho^{-n}+D_n\rho^{2-n}=\widetilde{\mathrm{LHS}} \tag{3.111}$$

Equation 3.104 leads to:

$$2B_0\rho + C_0\rho^{-1} + D_0\rho\,(1+2\log(\rho)) = \widetilde{\mathrm{LHS}} \tag{3.112}$$

$$A_1 + 3B_1\rho^2 - C_1\rho^{-2} + D_1(1+\log(\rho)) = \widetilde{\mathrm{LHS}} \tag{3.113}$$

$$A_n n\rho^{n-1} + B_n(2+n)\rho^{n+1} - C_n n\rho^{-n-1} + D_n(2-n)\rho^{1-n} = \widetilde{\mathrm{LHS}} \tag{3.114}$$

Finally, equation 3.105 leads to:

$$2B_0 - C_0\rho^{-2} + D_0(3+2\log(\rho)) = \widetilde{\mathrm{LHS}} \tag{3.115}$$

$$6B_1\rho + 2C_1\rho^{-3} + D_1\rho^{-1} = \widetilde{\mathrm{LHS}} \tag{3.116}$$

$$A_n n(n-1)\rho^{n-2} + B_n(1+n)(2+n)\rho^n + C_n n(1+n)\rho^{-2-n} + D_n(n-1)(n-2)\rho^{-n} = \widetilde{\mathrm{LHS}} \tag{3.117}$$

Shear balance at the loading circle

Using equation 3.107, we can find another condition at the loading circle:

$$\frac{P}{2\rho\pi a} = -\frac{4D(D_0 - \tilde{D}_0)}{a^3\rho} \tag{3.118}$$

$$\frac{P}{\rho\pi a} = \frac{2D}{a^3\rho^2}\left((D_1 - \tilde{D}_1) - 4\rho^2(B_1 - \tilde{B}_1)\right) \tag{3.119}$$

$$\frac{P}{\rho\pi a} = -\frac{4Dn\rho^{-1-n}}{a^3}\left((1+n)\rho^{2n}(B_n - \tilde{B}_n) + (n-1)(D_n - \tilde{D}_n)\right) \tag{3.120}$$

Edge clamped

If the centre of the plate is clamped at a radius $\rho = 1$, then we obtain the following conditions:

$$w(1,\theta) = 0\,, \tag{3.121}$$

$$\left.\frac{\partial w(r,\theta)}{\partial r}\right|_{r=1} = 0\,. \tag{3.122}$$

From the first condition (eq. 3.121), the equation we obtain is simply equations 3.95–3.97 set to zero. The second condition (eq. 3.122) leads to:

$$2B_0 + C_0 + D_0 = 0 \tag{3.123}$$

$$A_1 + 3B_1 - C_1 + D_1 = 0 \tag{3.124}$$

$$A_n n + B_n(2+n) - C_n n + D_n(2-n) = 0 \tag{3.125}$$

Edge free (Kirchhoff boundary conditions)

For a free edge, there are two boundary conditions, as expressed in equations 3.98 and 3.99. The first of these gives us:

$$2\tilde{B}_0(1+\nu) - \tilde{C}_0(1-\nu) + \tilde{D}_0(3+\nu) = 0 \tag{3.126}$$

$$2\tilde{B}_1(3+\nu) - 2\tilde{C}_1(\nu-1) + \tilde{D}_1(1+\nu) = 0 \tag{3.127}$$

$$\tilde{A}_n n(n-1)(\nu-1) + \tilde{B}_n(1+n)(n\nu - 2\nu - n - 2) + \tilde{C}_n n(1+n)(\nu-1) + \tilde{D}_n(n-1)(n(\nu-1) + 2(1+\nu)) = 0 \tag{3.128}$$

The second of these gives us:

$$\tilde{D}_0 = 0 \tag{3.129}$$

$$2\tilde{B}_1(3+\nu) - 2\tilde{C}_1(\nu-1) + \tilde{D}_1(\nu-3) = 0 \tag{3.130}$$

$$\tilde{A}_n n(n-1)(\nu-1) + \tilde{B}_n(1+n)(n(1-\nu)-4) + \tilde{C}_n n(1+n)(\nu-1) + \tilde{D}_n(n-1)(n(\nu-1)-4) = 0 \tag{3.131}$$

Simple Edge constraint

In this case, we have the conditions that the deflection at the edge is zero, and the radial moment is zero. The first of these conditions gives us:

$$\tilde{A}_0 + \tilde{B}_0 = 0 \tag{3.132}$$

$$\tilde{A}_1 + \tilde{B}_1 + \tilde{C}_1 = 0 \tag{3.133}$$

$$\tilde{A}_n + \tilde{B}_n + \tilde{C}_n + \tilde{D}_n = 0 \tag{3.134}$$

The second condition is the same as the condition on a free edge that leads to equations 3.126–3.128.

Edge clamped

This condition constrains the deflection and its first radial derivative to be zero at the edge. The first condition leads to equations 3.132–3.134. The second condition leads to:

$$2\tilde{B}_0 + \tilde{C}_0 + \tilde{D}_0 = 0 \tag{3.135}$$

$$\tilde{A}_1 + 3\tilde{B}_1 - \tilde{C}_1 + \tilde{D}_1 = 0 \tag{3.136}$$

$$n\tilde{A}_n + \tilde{B}_n(2+n) - n\tilde{C}_n + \tilde{D}_n(2-n) = 0 \tag{3.137}$$

3.8.3 Piezoelectric loading

Now that we have a solution for loading at a point, we can integrate this over whatever loading we wish to consider. If our loading-response function is $f(\rho,\phi;r,\theta)$, where (ρ,ϕ) is the loading point and (r,θ) are the position at which the surface deflection is evaluated, then for a distributed load $s(\rho,\phi)$ we can find the total surface deflection $z(r,\theta)$ from:

$$z(r,\theta) = \iint f(\rho,\phi;r,\theta)s(\rho,\phi)\rho\,\mathrm{d}\rho\,\mathrm{d}\phi \tag{3.138}$$

In the case of a unimorph or bimorph actuator, we know that the applied force takes the form of the Laplacian of the applied optics (equation 3.92). The usual case that we shall consider is that the voltage is only applied to a region bounded by two radii at angular positions θ_1 and θ_2, and by two arcs at radii ρ_1 and ρ_2, where $\rho_2 > \rho_1$. In order to evaluate the surface deflection, we need the Laplacian of this voltage distribution,

$$V = H(\rho-\rho_1, \rho_2-\rho, \phi-\phi_1, \phi_2-\phi)\,, \tag{3.139}$$

where $H(x_1,x_2,\dots)$ is the Heaviside function, which is 1 for $x_1 > 0 \cup x_2 > 0 \dots$, and zero otherwise. The Laplacian of this can be found in terms of the Heaviside function, and Dirac delta functions and

its derivatives, and in its full form is:

$$\begin{aligned}\nabla^2 V = \left(\frac{-1}{\rho}\right)\delta(\rho_2-\rho)H(\rho-\rho_1,\phi-\phi_1,\phi_2-\phi)+\\ \left(\frac{1}{\rho}\right)\delta(\rho-\rho_1)H(\rho_2-\rho,\phi-\phi_1,\phi_2-\phi)+\\ (-2)\delta(\rho-\rho_1)\delta(\rho-\rho_2)H(\phi-\phi_1,\phi_2-\phi)+\\ \delta'(\rho-\rho_1)H(\rho_2-\rho,\phi-\phi_1,\phi_2-\phi)+\\ \delta'(\rho_2-\rho)H(\rho-\rho_1,\phi-\phi_1,\phi_2-\phi)+\\ \left(\frac{-2}{\rho^2}\right)\delta(\phi-\phi_1)\delta(\phi_2-\phi)H(\rho-\rho_1,\rho_2-\rho)+\\ \left(\frac{1}{\rho^2}\right)\delta'(\phi-\phi_1)H(\rho-\rho_1,\rho_2-\rho,\phi_2-\phi)+\\ \left(\frac{1}{\rho^2}\right)\delta'(\phi_2-\phi)H(\rho-\rho_1,\rho_2-\rho,\phi-\phi_1)\end{aligned} \quad (3.140)$$

Some of these terms are zero, and so we arrive at our final expression for the applied load:

$$\begin{aligned}\nabla^2(V) = H(\rho-\rho_1,\phi-\phi_1,\phi_2-\phi)\left(\delta'(\rho_2-\rho)-\frac{\delta(\rho_2-\rho)}{\rho}\right)+\\ H(\rho_2-\rho,\phi-\phi_1,\phi_2-\phi)\left(\frac{\delta(\rho-\rho_1)}{\rho}-\delta'(\rho-\rho_1)\right)+\\ \frac{\delta'(\phi-\phi_1)}{\rho^2}H(\rho-\rho_1,\rho_2-\rho,\phi_2-\phi)+\\ \frac{\delta'(\phi_2-\phi)}{\rho^2}H(\rho-\rho_1,\rho_2-\rho,\phi-\phi_1)\end{aligned} \quad (3.141)$$

This load is a combination of Dirac deltas and derivatives around the edge of the region to which a voltage is applied.

The next stage in finding the response function is to find the integrals of an arbitrary point-response function when the loading function is of the form of equation 3.141. As a slight simplification, we note that the solution is a sum of terms of the form $f(r,\rho)g(\theta-\phi)$. Therefore, we need to find, for each of these terms:

$$I_1 = \int_{\rho_1}^{\infty}\int_{\phi_1}^{\phi_2}\left(\delta'(\rho_2-\rho)-\frac{\delta(\rho_2-\rho)}{\rho}\right)f(r,\rho)g(\theta-\phi)\,\rho\,d\rho\,d\phi \quad (3.142)$$

$$I_2 = \int_{0}^{\rho_2}\int_{\phi_1}^{\phi_2}\left(\frac{\delta(\rho-\rho_1)}{\rho}-\delta'(\rho-\rho_1)\right)f(r,\rho)g(\theta-\phi)\,\rho\,d\rho\,d\phi \quad (3.143)$$

$$I_3 = \int_{\rho_1}^{\rho_2}\int_{0}^{\phi_2}\frac{\delta'(\phi-\phi_1)}{\rho^2}f(r,\rho)g(\theta-\phi)\,\rho\,d\rho\,d\phi \quad (3.144)$$

$$I_4 = \int_{\rho_1}^{\rho_2}\int_{\phi_1}^{2\pi}\frac{\delta'(\phi_2-\phi)}{\rho^2}f(r,\rho)g(\theta-\phi)\,\rho\,d\rho\,d\phi \quad (3.145)$$

Evaluating the integrals, we find:

$$I_1 = \left.\frac{\partial f(r,\rho)}{\partial \rho}\right|_{\rho=\rho_1} \int_{\phi_1}^{\phi_2} g(\theta-\phi)\,\mathrm{d}\phi \tag{3.146}$$

$$I_2 = -\left.\frac{\partial f(r,\rho)}{\partial \rho}\right|_{\rho=\rho_2} \int_{\phi_1}^{\phi_2} g(\theta-\phi)\,\mathrm{d}\phi \tag{3.147}$$

$$I_3 = \left.\frac{\partial g(\theta-\phi)}{\partial \phi}\right|_{\phi=\phi_1} \int_{\rho_1}^{\rho_2} \frac{f(r,\rho)}{\rho}\,\mathrm{d}\rho \tag{3.148}$$

$$I_4 = \left.\frac{\partial g(\theta-\phi)}{\partial \phi}\right|_{\phi=\phi_1} \int_{\rho_1}^{\rho_2} \frac{f(r,\rho)}{\rho}\,\mathrm{d}\rho \tag{3.149}$$

So, for each term in the expansion, we need to find the integral and the derivative of the appropriate expressions for $f(r,\rho)$ and $g(\theta-\phi)$.

All the terms $g_n(\theta-\phi)$ are of the form $\cos(n(\theta-\phi))$, and so the integrations and differentiations are straightforward:

$$\frac{\partial g_n(\theta-\phi)}{\partial \phi} = n\sin(n(\theta-\phi)) \tag{3.150}$$

$$\int_{\phi_1}^{\phi_2} g_n(\theta-\phi)\,\mathrm{d}\phi = \frac{1}{n}\left\{\sin\left(n(\theta-\phi_1)\right) - \sin\left(n(\theta-\phi_2)\right)\right\} \tag{3.151}$$

and when $n = 0$,

$$\int_{\phi_1}^{\phi_2} g_0(\theta-\phi)\,\mathrm{d}\phi = \phi_2 - \phi_1 \tag{3.152}$$

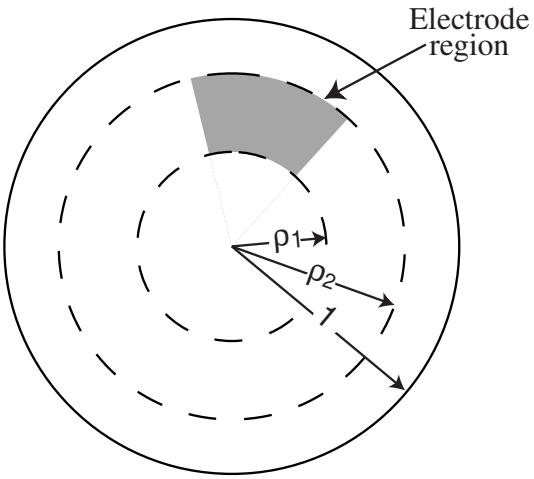
The integrations over the loading point, ρ, depend upon the detailed form of the solutions to the simultaneous equations derived from the boundary conditions. In all cases, the integrations are straightforward, since $f(r,\rho)$ only ever depends upon ρ in a simple way, usually as a polynomial series.

One difficulty remains before we can calculate the deflection of a surface under the influence of an electrode, and this is that the solution is split into a form that is valid outside the loading circle, and inside the loading circle. That means that the integrated solutions need to be used differently in the three different regions: wholly inside the electrode area, between the inner and outer regions of the electrode, and wholly outside the electrode area. These regions are depicted in figure 3.14. In the inner and outer regions, we simply use the appropriate solution for the integrated inner and outer functions respectively. In the middle region, we effectively divide the electrode into 2 parts, one part outside the current evaluation radius and one part inside, and add these two parts together. In this way, we can construct the solution for the entire surface.

3.8.4 Example surface evaluations

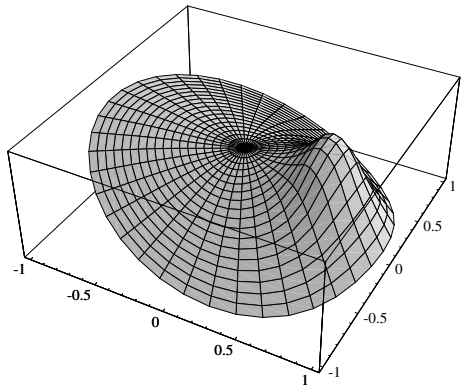
The solution found in the preceeding sections has been implemented for the cases of free edge, simply constrained edge and clamped edge; all with no central constraint. The implementation is in the form of a *Mathematica* program, which sets up the simultaneous equations, and then finds analytical solutions to them, performs the analytical integrations, and finally adds all the terms together. In the implementation, only the first 80 terms in the infinite summation are considered, since the infinite summation cannot be computed analytically. As an example, surfaces are plotted for an electrode region between $r = 0.5$ and $r = 0.75$, and an angular extent of $\pi/4$, for each of

Figure 3.14 The three regions of the solution for an electrode region extending between $\rho = \rho_1$ and $\rho = \rho_2$

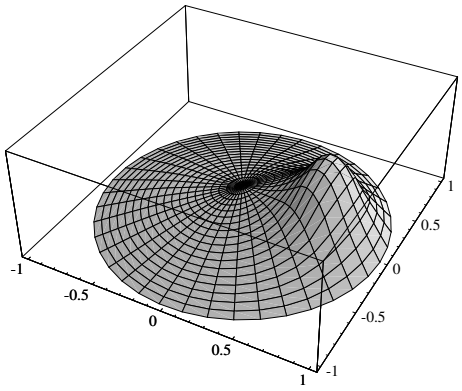


the three cases. Also plotted is the summation of the terms $n = 41$ to $n = 80$, to demonstrate the magnitude of the error in only considering the first 40 terms. For the three surfaces, the vertical scale is the same for each; in the case of the remainder surface, the vertical scale is increased by a factor of 1000.

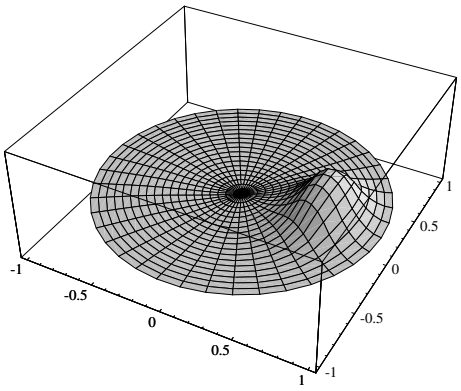
Figure 3.15 Deflections for a voltage applied to a region $0.5 \leq r \leq 0.75$ and $-\pi/8 \leq \theta \leq \pi/8$ for three different boundary conditions



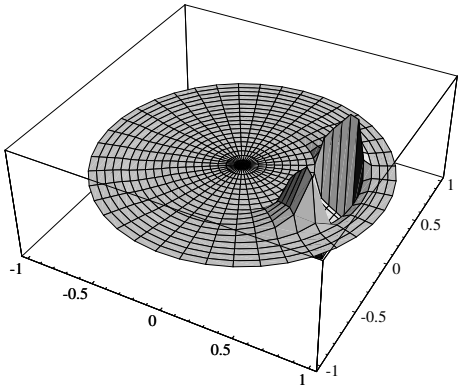
(a) Edge free



(b) Simply constrained edge



(c) Clamped edge



(d) Terms 41 to 80

4 Numerical Simulations of Bimorph Mirrors

Using the theory developed in chapter 3, we can perform numerical simulations of the predicted performance of bimorph mirrors under various conditions. In all the simulations that follow, we will use the series solution of the biharmonic equation described in section 3.8. We will investigate various electrode geometries, and how well they satisfy the constraints of different applications. Also, using these simulations, it is possible to simulate in a controlled way the interactions between bimorph mirrors and various types of wavefront sensor. Of particular interest is the relationship with curvature sensors.

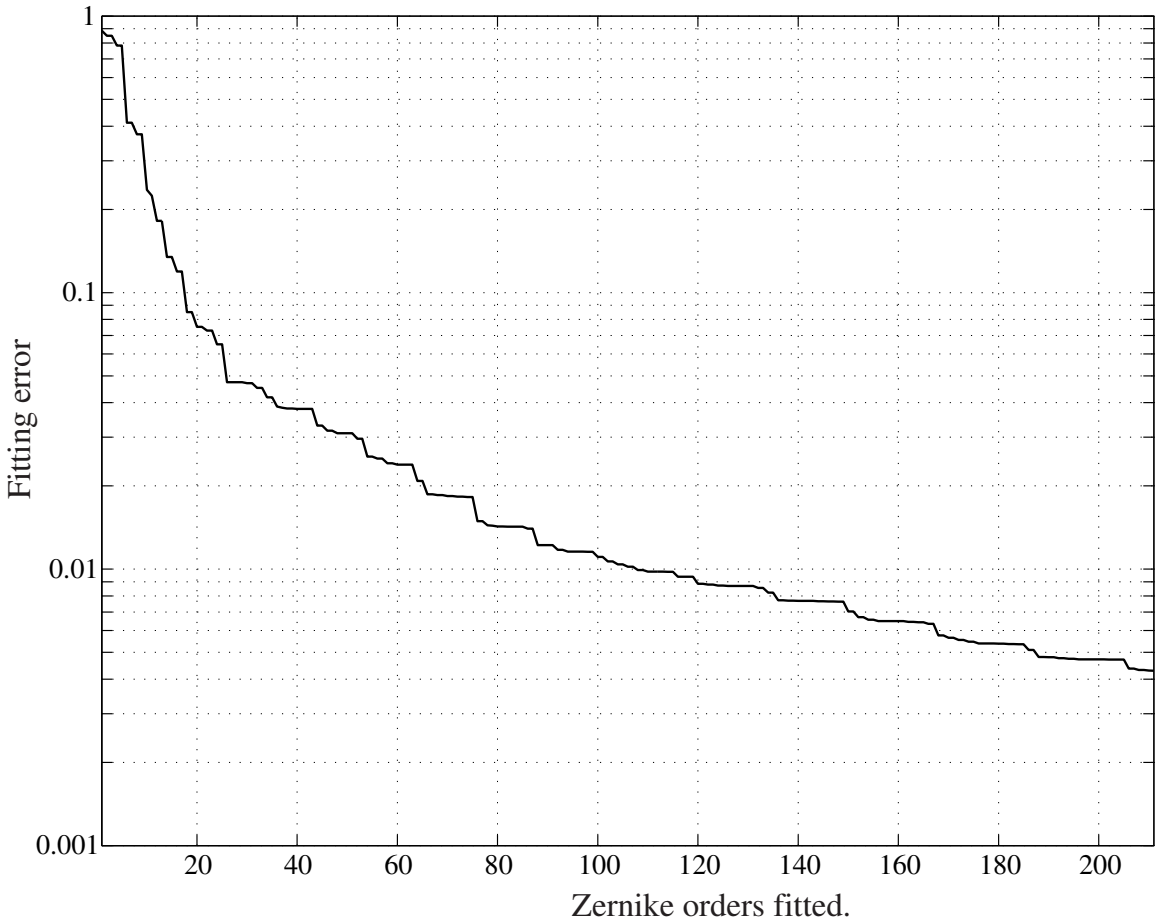
4.1 Preparation for Quasistatic Optimistation

There are two different aspects of a bimorph mirror that must be designed, namely the thicknesses of the various layers, and the layout of the electrodes. The choice of the thicknesses of the layers influences only the amount of curvature that can be achieved, and so does not directly affect the choice of electrode structure. However, when choosing the electrode structure, the constraint on the curvature achievable must be borne in mind. For example, it may be the case that by applying extreme curvatures, a good fit to some desired shape may be obtained whereas a more holistic design considering the effect that the curvatures needed has upon the laminate structure would perhaps sacrifice some amount of fit to aid construction of the practical device. In the following optimisations, it will be useful to consider some practically achieved curvatures, and use these as guides to design. We will consider a curvature equivalent to a defocus sensitivity of 50μm/kV comfortably achievable, 100μm/kV the upper limit, and 25μm/kV possible if very good initial surface figure is required. These assumptions will be used to normalise the output from the numerical simulations, whose vertical scaling is otherwise difficult to extract. Also, in the following optimisations, it will generally be assumed that a separate tip-tilt mirror is part of the AO system, since bimorph mirrors do not produce tip-tilt aberrations very efficiently.

4.1.1 Data calculation and Zernike estimation

In the following simulations, the mirror surface shape is calculated from the exact solution over a rectangular grid with 64 points on each side, using the first 80 azimuthal terms in the expansion. A solution of this size takes about 30 seconds on a Pentium II processor clocked at 400MHz. The numerical simulations produce a large amount of data as output. It is not a simple matter to deal with this data directly. Rather, it is preferable to expand the generated surfaces in terms of Zernike polynomials (described in detail in Appendix A). The expansion uses a least-squares fitting technique described in section A.2. By precalculating most of the matrices in the Zernike fitting procedure, the time required to fit 211 Zernike coefficients is around a third of a second, on the same proces-

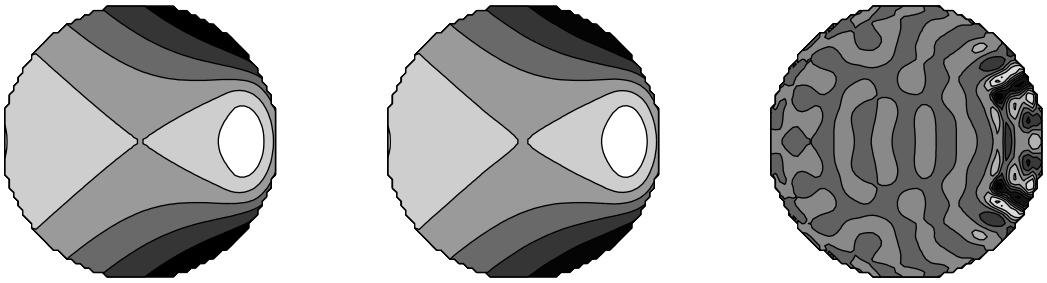
Figure 4.1 Relative fitting error as a function of the number of Zernike orders fitted



sor as above. The fitting error for fitting a typical electrode response function as a function of the number of Zernike terms used to fit is shown in figure 4.1. The fitting error shown is the variance of the difference between a surface reconstructed from the fitted Zernike coefficients and the actual surface, divided by the variance of the original surface shape. This graph shows that the relative amount of fitting error is less than 1 part in 200 when we consider 211 Zernike terms, corresponding to a Strehl ratio of 0.995 when the overall phase variance is 1 radian^2. In the following simulations, we will consider only the first 211 terms, since this gives us a good compromise between accuracy of fit and execution time. The original surface to be fitted, the best fit, and their difference are shown in figure 4.2. As can be seen from the graph of the difference, the largest error is around the edge of the electrode region, which is where there is a discontinuity in the high-order derivatives.

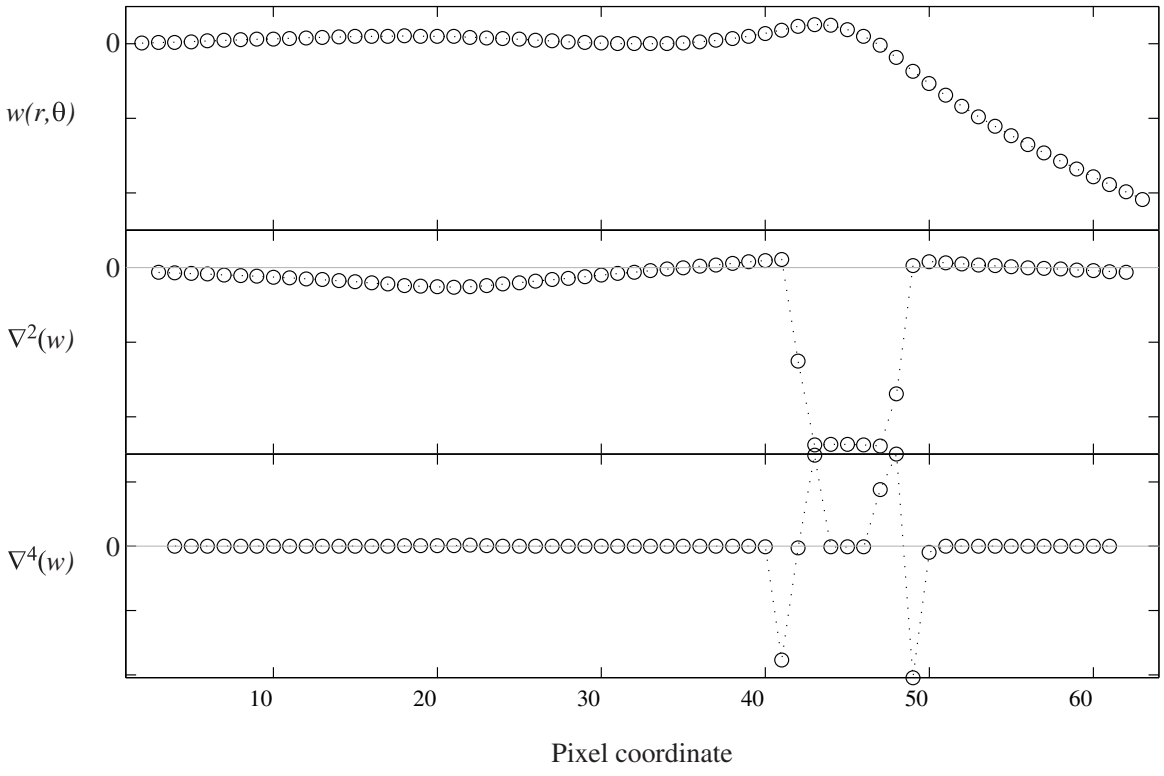
Before we go on to the optimisations, it is interesting to observe the behaviour of the surface shape when operated on by the Laplacian operator. Figure 4.3 shows a slice through a calculated surface, the Laplacian, and the Laplacian of the Laplacian (i.e. the biharmonic) of this surface. We can clearly see from this graph that there is some curvature even in the region away from the electrode, but the biharmonic behaves as expected. In calculating these graphs, a simple discrete numerical Laplacian operator was used.

Figure 4.2 A surface, its best Zernike fit, and the difference. The vertical scale for the difference is 87 times greater than for the other two.



Original surface Fitted surface Difference

Figure 4.3 Cross section through a surface, its Laplacian and its biharmonic



4.1.2 Degrees of freedom

When designing the electrode layout of a bimorph mirror, there are very many degrees of freedom that must be chosen. Even in the case where we have decided on the total number of electrodes, there are still many degrees of freedom remaining. These are:

- Number of rings of electrodes
- Number of electrodes in each ring
- Angular distribution of electrodes
- Radii of the rings
- Fraction of the mirror used for the optical pupil

So, for a given number of electrodes, N_e, we can choose the number of rings to be $1 \leq N_r \leq N_e$. For a given N_r, there are many ways of arranging the electrodes between these rings. We can enumerate these by observing that we have $N_e - 1$ positions in which we can place $N_r - 1$ boundaries. Therefore, the number of arrangements of electrodes (not counting the radii of the rings) is:

$$\sum_{N_r=1}^{N_e} \frac{(N_e-1)!}{(N_e-N_r)!(N_r-1)!} \tag{4.1}$$

This summation is equal to 2^{N_e-1}; clearly there are very many arrangements for even a relatively small number of electrodes.

In order to make the problem tractable, the ring radii are constrained to take on values incrementing only in a step of one twentieth of the radius. The number of rings is constrained to lie between 2 and 4, and the number of electrodes in a ring constrained to be fewer than 18. Since Kolmogorov turbulence is spatially isotropic, in the first instance, only electrodes equally spaced in angle are considered. This constraint could be relaxed if we had some *a priori* knowledge that suggested that this would be expeditious, e.g. if the mirror were to be used at non-normal incidence, it may again be useful have non-equal angular spacing of the electrodes. Also, the fraction of the pupil used is only allowed to shrink to half the area of the mirror, or 71% of the diameter. These restrictions on their own still leave very many possible configurations; for the total electrode numbers considered, only some arrangements are considered and optimised.

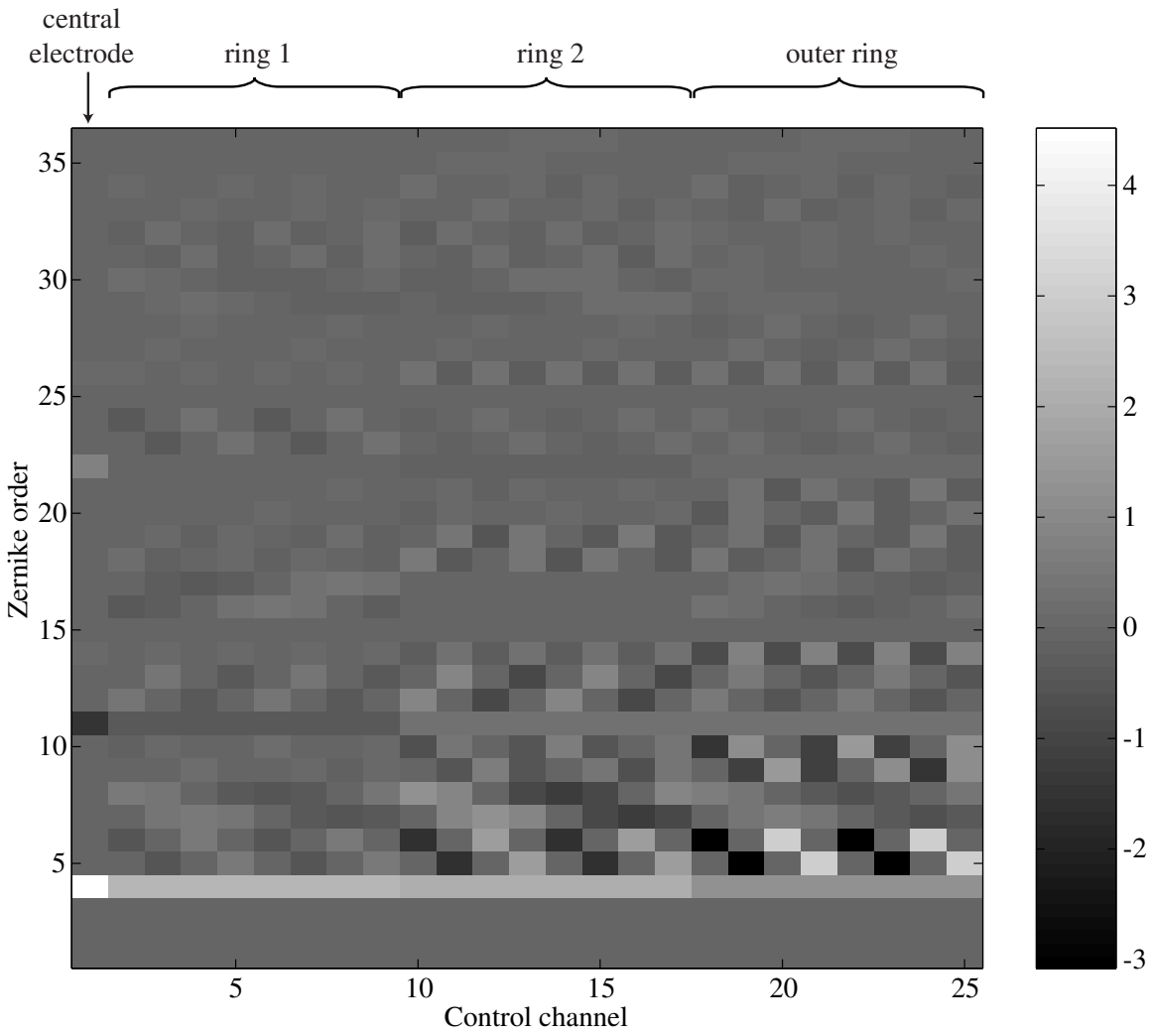
4.1.3 Creation of the Zernike transfer matrix

As described elsewhere (section 2.4), the transfer matrix is of fundamental importance when discussing the mode properties of a deformable mirror. To clarify the following sections, we shall now define the vectors and matrices used. The vector of Zernike coefficients is $\mathbf{a}$, with coefficients a_i. The number of Zernike coefficients considered is N_a. The vector of control voltages is $\mathbf{e}$, with coefficients e_i. The number of control signals used is N_e. The matrix relating these vectors is $\mathbf{H}$, where:

$$\mathbf{a} = \mathbf{H} \times \mathbf{e} \tag{4.2}$$

Therefore, $\mathbf{H}$ has N_a rows, and N_e columns. To calculate this matrix, firstly the surface maps of an electrode from each ring are decomposed into a Zernike vector. This vector is then "rotated" to produce the Zernike decompositions of the electrodes in the same ring. This rotation is described in Appendix A; the process produces vectors which are precisely as accurate as the input vector, and

Figure 4.4 Transfer matrix **H** for a 1–8–8–8 mirror



there is no computationally expensive rotation of the surface map matrix. An example **H** matrix is shown in figure 4.4. Only 36 Zernike orders are shown for clarity. In this example it is clearly seen that piston, tip and tilt are forced to zero. Also, it is clear that the radially symmetric orders remain constant for a given ring, as expected.

4.2 Effective Degrees of Freedom

Now that we have the necessary data (i.e. the Zernike transfer matrix for a very wide variety of electrode layouts), we are in a position to perform various optimisations. In the absence of any detailed knowledge about the form of the correction that is required, we can still perform some layout optimisation by examining the orthonormal modes of the mirror. By performing the singular value decomposition of **H**, we are given the three matrices, **U**, Σ and **V**. Using these matrices, we can establish certain properties about the expected performance of the mirror.

One simple measure of the performance is to take the ratio of the first singular value to the last singular value: this is known as the "condition", μ. The singular value of a given mode is tells us the variance of that mode's corresponding surface, when the sum of the squared electrode voltages ($\sum e^2$) is 1. So, the condition number tells us the ratio of the value of $\sum e^2$ between the most and

least easily reproduced modes, when the same surface variance is applied.

It is instructive to examine an example of the singular value decomposition of a mirror's transfer function. This is demonstrated graphically in figure 4.5. The electrode configuration chosen is one found in the sections following this which optimise the condition number. The configuration is 1–4–4; the ring radii are 0.25 and 0.6 respectively, and the pupil fraction used is 0.95. In the figure, the top row shows contour plots of the 3 computed mirror response functions, clipped at a pupil radius of 0.95. The 3 rows beneath that show the normal modes, along with the corresponding singular value. The condition number for this situation is 11.7. The final row shows modes produced in the SVD procedure which cannot be reproduced at all by this mirror.

4.3 Fitting to a Kolmogorov Turbulence Spectrum

As mentioned in section 2.8, we know that the effect that a turbulent atmosphere has on an optical beam can be characterised by a Kolomogorov power spectrum of phase fluctuations. We have also shown how to use the Karhunen-Loève method to produce a vectors of Zernike coefficients compatible with Kolmogorov turbulence. This method can be extended to give information about how well a deformable mirror will fit a Kolomogorov spectrum. In the following sections, a method will be developed for calculating the expected residual variance of a wavefront after correction by a deformable mirror, and also the expected control signals that must be applied in order to achieve this correction—all this can be calculated directly, without recourse to a time-consuming Monte-Carlo procedure. However, in order to verify the arguments here propounded, specimen results will be scrutinised using a Monte-Carlo procedure. Also, if any non-linear effects are to be considered, then a Monte-Carlo procedure must be necessary.

4.3.1 Expected residual variance

The variance of the phase of an optical wavefront tells us how well that wavefront can be imaged. The Strehl ratio is commonly used to quantify imaging fidelity; as described in section 2.1.3, this is the ratio of the intensity at the centre of the unaberrated PSF to the maximum intensity of the aberrated PSF, it is related to the phase variance σ^2 in the following way:

$$S = 1 - \sigma^2 \tag{4.3}$$

providing $\sigma^2 \ll 1$.

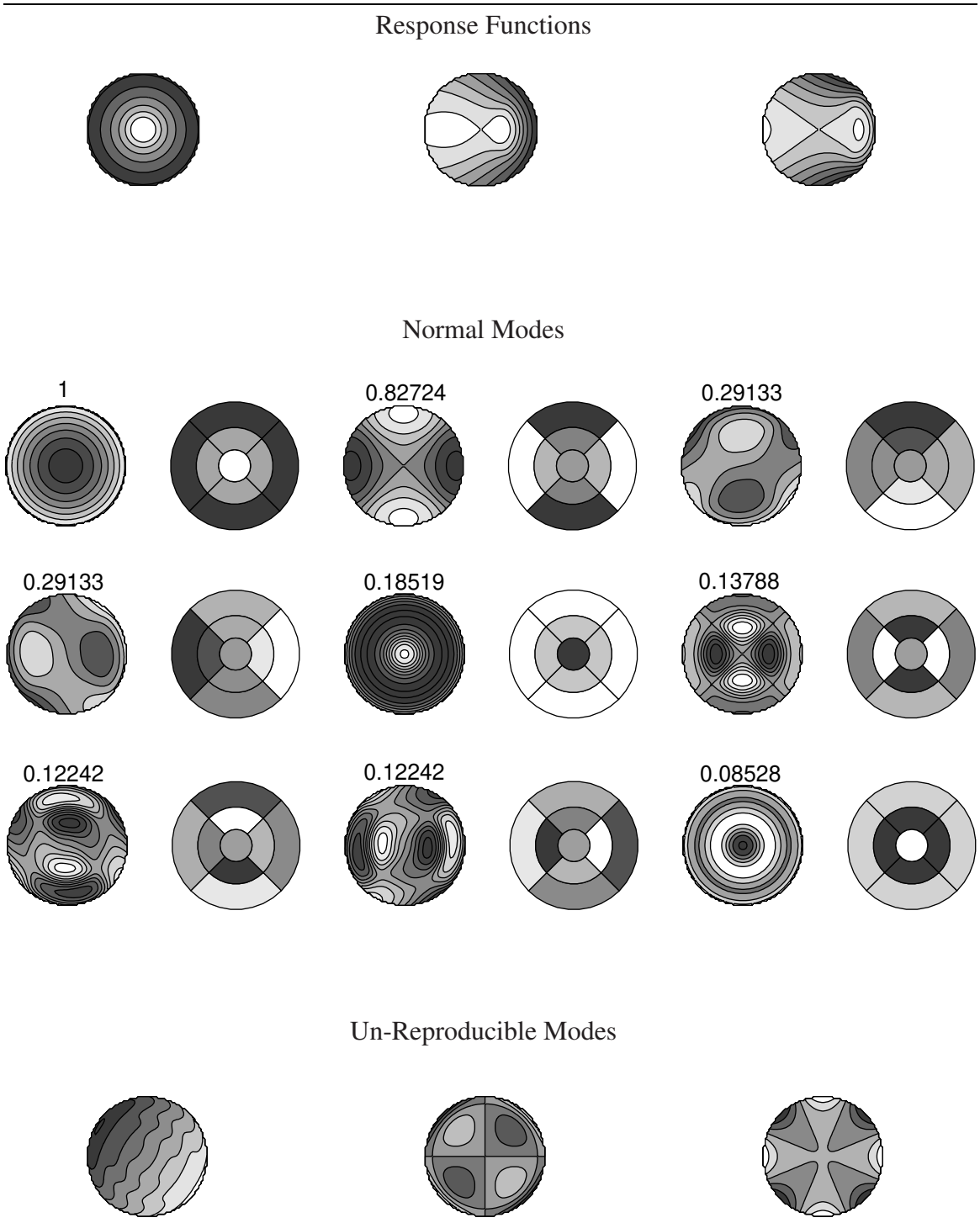
In order to calculate the expected residual variance, we proceed as follows. Firstly, we calculate the Zernike transfer matrix as described above. Then, we convert this to a Karhunen-Loève (KL) transfer matrix. This step is achieved by pre-multiplying the Zernike transfer matrix by $\mathbf{U}^T$, where $\mathbf{U}$ is that calculated from the SVD of the covariance matrix (NB: for this section, it is useful to ignore piston). If the result of this multiplication is $\mathbf{H}_{KL}$, then we require its SVD matrices:

$$\mathbf{H}_{KL} = \mathbf{U}_{KL}\Sigma_{KL}\mathbf{V}_{KL}^T \tag{4.4}$$

As described in section 2.8.3, this $\mathbf{U}_{KL}$ matrix has columns which are the orthonormal modes of the deformable mirror, expressed in KL modes. Since the KL modes are a diagonal basis, we can then calculate the variance of each column of $\mathbf{U}_{KL}$ by simply summing the KL mode variance weighted by the square of that mode's coefficient. The nett result is that the variance removed can be calculated:

$$\sigma^2_{\text{rem}} = \sum\sum(\mathbf{U}_{KL} \times \mathbf{U}_{KL}^T) \cdot \Sigma \tag{4.5}$$

Figure 4.5 The normal modes of a bimorph mirror



where Σ is a diagonal matrix of the weights of the KL modes, found from the SVD of the Zernike covariance matrix. We know the total piston-removed variance of a Kolmogorov spectrum, and so we can calculate the total residual phase variance:

$$\sigma^2_{\text{resid}} = 1.04035 - \sigma^2_{\text{rem}} \tag{4.6}$$

The numerical offset in the above expression is found by summing the values in Σ: this is the average wavefront variance for the generated wavefronts[1]. The actual atmospheric correction will be slightly worse than this: if we consider only the first 211 Zernike orders, then the total variance that is not simulated can be found by subtracting the sum of Σ from known value for the piston-removed variance of a Kolmogorov wavefront. The resulting unsampled variance is $0.0029 \times (D/r_0)^{5/3}$. Providing our correction is considerably worse than this, as it will be if we only have a small number of control channels, then this correction is small. We can also calculate the maximum value of D/r_0 that is well sampled by considering 211 Zernike orders: if our criterion is that the Strehl ratio of the fit should be ≥ 0.8, then the maximum value of D/r_0 is given by:

$$(D/r_0)_{211} = \left(\frac{0.2}{0.0029}\right)^{3/5} \approx 12 \tag{4.7}$$

So, we must restrict our attention to systems that correct for turbulence less severe than this limit. If we were to consider higher order systems, then more Zernike orders would be required in the transfer matrix. For example, 466 Zernike orders allow us to examine systems whose limiting value of D/r_0 is around 18.

4.3.2 Expected control signal variance

Likewise, it is useful to calculate the variance and standard deviation of the control signals—this will give us a useful indication of how much of the deformable mirror's available stroke we need to use. In order to calculate this, we again need the transfer function in terms of KL modes, $\mathbf{H}_{KL}$ and its pseudoinverse, $\mathbf{H}^*_{KL}$. A vector of control voltages is calculated by multiplying $\mathbf{H}^*_{KL}$ by a vector of Gaussian-distributed random numbers whose variances are scaled by the diagonal values from the Σ matrix found from the SVD of the covariance matrix. Since the KL modes have zero covariance, calculating the voltage variances is a simple matter of squaring the inverse matrix (element-by-element), and multiplying by the appropriate KL mode variance. This gives us a vector of voltage variances, σ^2_V:

$$\sigma^2_V = (\mathbf{H}^*_{KL} \cdot \mathbf{H}^*_{KL}) \times \Sigma \tag{4.8}$$

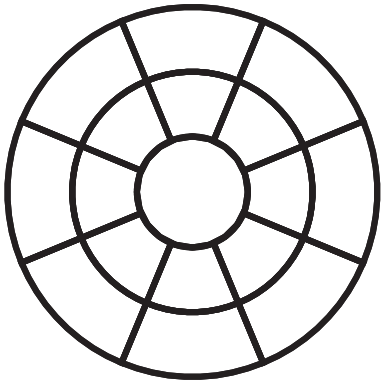
4.3.3 Comparison of expected values with Monte-Carlo simulations

In order to verify that the above relationships hold true, it is useful to perform a sample complete Monte-Carlo simulation. The stages of this simulation are outlined below:

- Construct a (correctly scaled) Zernike transfer matrix $\mathbf{H}$, and its pseudoinverse $\mathbf{H}^*$.
- Generate 2000 Kolmogorov wavefronts using the KL technique. These are stored in a matrix $\mathbf{K}$ where each column corresponds to a wavefront.
- Calculate $\mathbf{E}$, the control signals required. This is found from $\mathbf{E} = \mathbf{H}^* \times \mathbf{K}$.

[1] The offset often quoted is 1.03—the discrepancy here is due to the precise calculated values of the covariance, which use the form given by Roddier [65]

Figure 4.6 Electrode layout for Monte-Carlo comparisons



- Calculate $\mathbf{K}' = \mathbf{H} \times \mathbf{E}$, the deformable mirror response
- Calculate $\Delta = \mathbf{K} - \mathbf{K}'$, the residual wavefront expressed in terms of a Zernike expansion.

These matrices constitute all the data we need to calculate the residual wavefront variance, and the control signal standard deviations. This procedure was carried out for a range of values of D/r_0, for a 17-electrode mirror; the electrode arrangement is shown in figure 4.6. A graph showing the residual wavefront variance is shown in figure 4.7. The solid line is the value calculated according to $\sigma^2_{\text{resid}} \times (D/r_0)^{5/3}$, the circles represent the numerically calculated values (in this example, the residual variance is that relative to the 211 generated terms, as in equation 4.6). Figure 4.8 shows the control signal standard deviation for the inner electrode, a middle ring electrode and an outer ring electrode. Again, the discrete points represent the results of the Monte-Carlo simulation, and the smooth lines are the direct calculation.

These figures show that there is very good agreement between the Monte-Carlo estimations and the direct calculations. Therefore, we will rely on the direct calculations when considering different mirror configurations, as these calculations can be done much more quickly than a Monte-Carlo simulation.

4.4 Optimisations for Kolmogorov Turbulence

Now that we know how to calculate the expected performance of a bimorph mirror when correcting for turbulence with a Kolmogorov spectrum, we can go on to attempt to optimise the parameters of a given mirror configuration.

4.4.1 Optimising a low-order corrector

In this section, we will show how a well-designed bimorph mirror can obtain good atmospheric correction whilst using only a very small number of control channels. A convenient number of control channels to consider is 16: this is a common upper limit on the number of analogue outputs one can easily achieve from a standard PC. Working within this constraint, we need to choose the arrangement of electrodes.

Considering first the 3-ring case, we still have 3 independent parameters after our choice of electrodes-per-ring, namely the two ring boundary radii, and the fraction of the pupil that is used. The contour plot in figure 4.9 shows the results of the optimisation. In these pseudocolour plots, two degrees of freedom have been removed, leaving only the choices for the ring radii. The removal was done by choosing, for each pair of radii values, the best electrodes-per-ring arrangement and also

Figure 4.7 Comparison of Monte-Carlo and direct residual variation calculation. Circles are Monte-Carlo estimates, solid line is the direct calculation

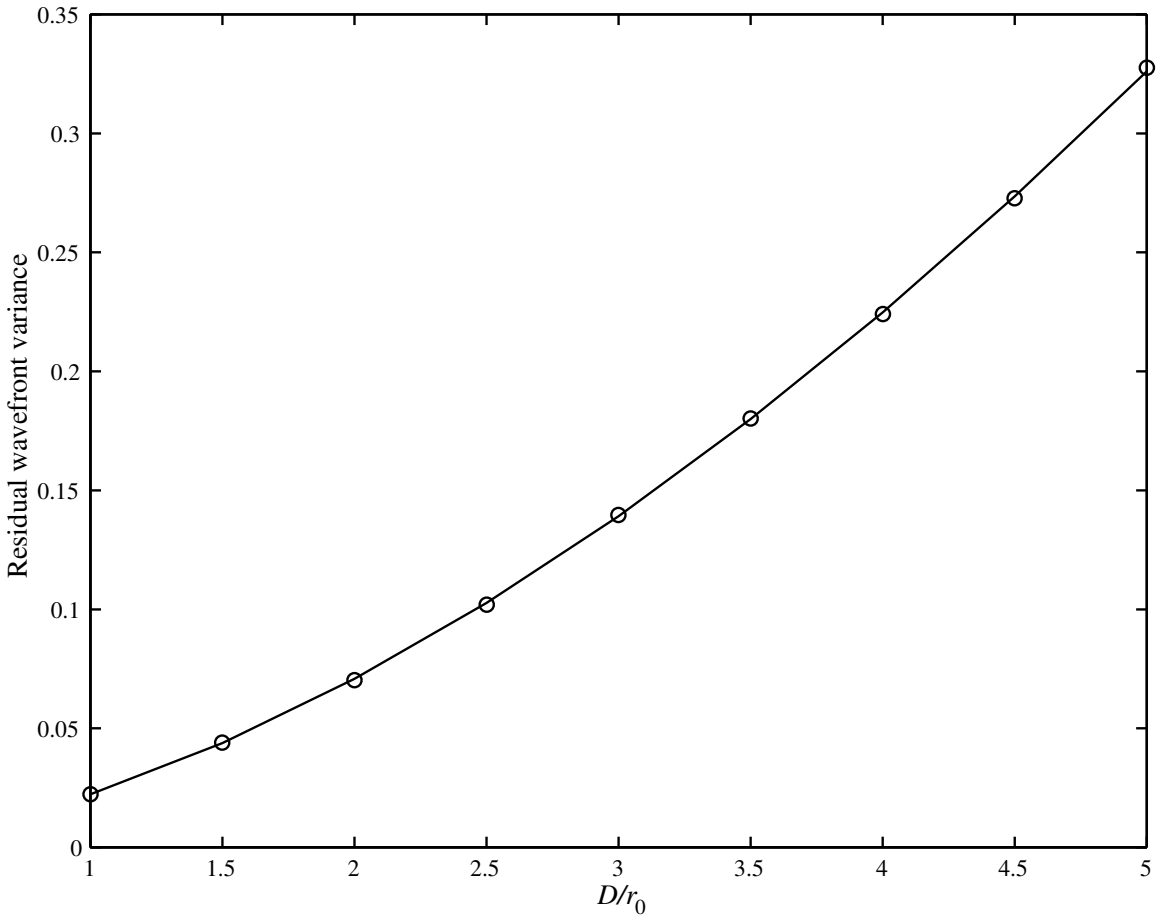


Figure 4.8 Comparison of Monte-Carlo and direct calculation of control signal standard deviation. Discrete points are Monte-Carlo estimates, the smooth lines are the direct calculations

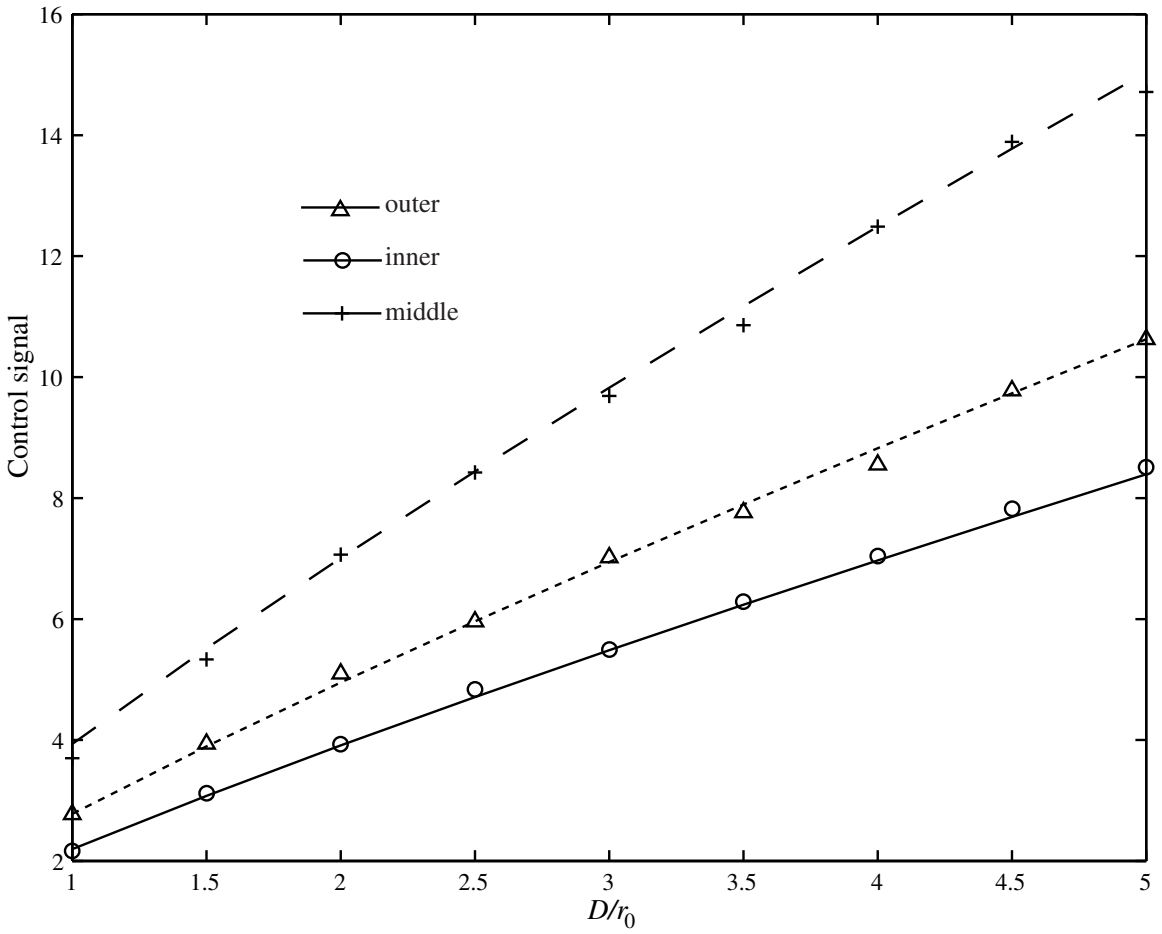
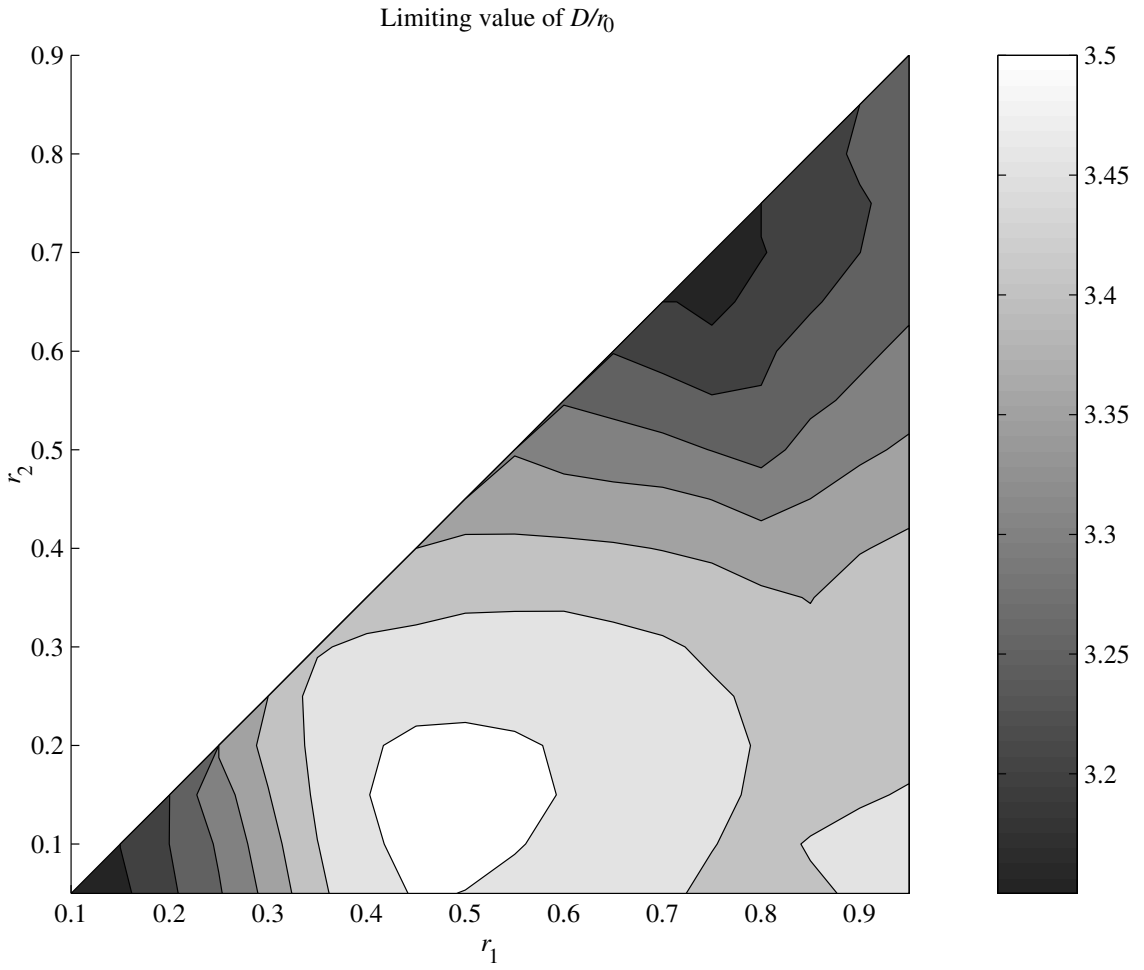


Figure 4.9 Results from 16 electrode Kolmogorov spectrum optimisation



the best pupil fraction to use. This plot shows the value of D/r_0 that can be corrected to a Strehl ratio of 0.8. This is found simply from:

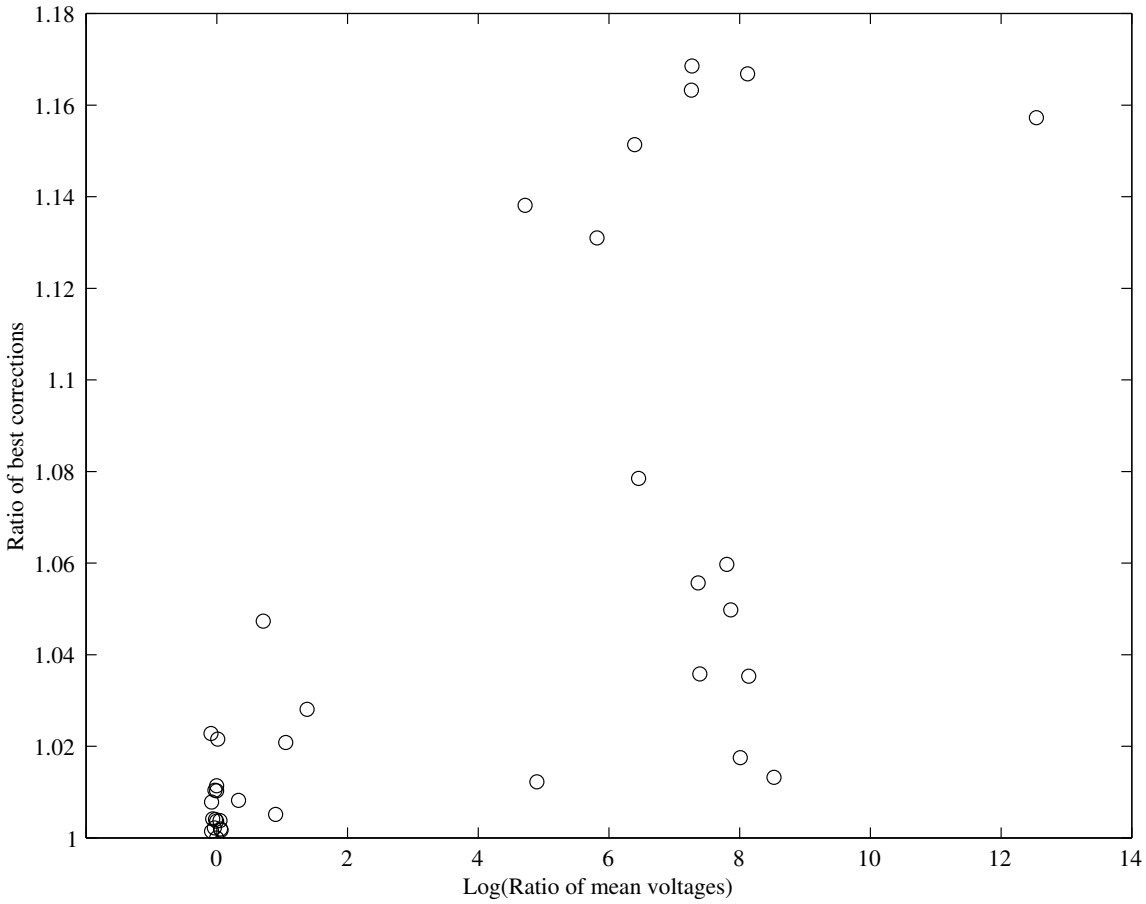
$$(D/r_0)_{\text{limit}} = \left(\frac{0.2}{\sigma^2_{\text{resid}}}\right)^{3/5} \tag{4.9}$$

as in equation 4.7. Sometimes, it is useful to correct only to a Strehl ratio of 0.6, and in this case, we need to multiply the previously found limiting D/r_0 by a numerical factor of $(0.4/0.2)^{3/5} \approx 1.5$.

4.4.2 Optimising a 3-ring corrector

We can now go on to optimise the more general case of a corrector with three rings of electrodes. In this case, we again have the same 3 degrees of freedom as above for each arrangement of electrodes. Instead of concentrating on only a small number of possible layouts, it is possible to consider many arrangements, and discover the best parameters for each such layout. From earlier optimisations, it was discovered that it is best to have only one electrode in the inner ring, and this was chosen to be the case for all the arrangements considered here. In the course of this optimisation procedure, for each layout, the optimal value of the condition number μ, and the best value of D/r_0 that could be corrected were found, along with the associated parameters. It was discovered that the optimisation for μ was consdirably more succesful. For all arrangements considered, the best value of μ did not

Figure 4.10 Comparison of optimisation for μ and optimisation for D/r_0 for a 3 ring corrector



occur with the same parameter values as for the best value of D/r_0. However, the value achieved was always within 10%. It was found that there was a large price to pay for this slight improvement in performance: the voltages required to achieve this better correction were incredibly large. Figure 4.10 demostrates this. In this scatter plot, the ordinate is the logarithm of the ratio of mean RMS voltages required, and the abscissa shows the ratio of values of D/r_0 achieved. It is clear that in the cases where the voltages required are at all comparable, the correction achieved is almost identical. Therefore, optimisation for μ is the only sensible option.

Figure 4.11 shows the values of D/r_0 obtained from an optimisation of μ, against the total number of electrodes. The legend refers to the number of electrodes in the outermost ring—this clearly seems to be an important trend. Figure 4.12 shows the mean value of the RMS voltages required for the correction. Table 4.1 shows the actual values found. In this table, ρ is the fraction of the mirror diameter used for correction.

4.4.3 Optimising a 4 ring corrector

In the same manner as the preceeding section, we can find the best configuration of electrodes arranged in 4 rings. In this case, we have one extra degree of freedom, and so the amount of computing time to find the radial parameters for a given arrangement of electrodes-per-ring increases from around 20 minutes to around 4 hours. We can decrease this amount of time by adding a constraint that the "thickness" of each ring cannot decrease below 0.1 (this roughly halves the time

Figure 4.11 Values of D/r_0 obtained with a 3 ring corrector

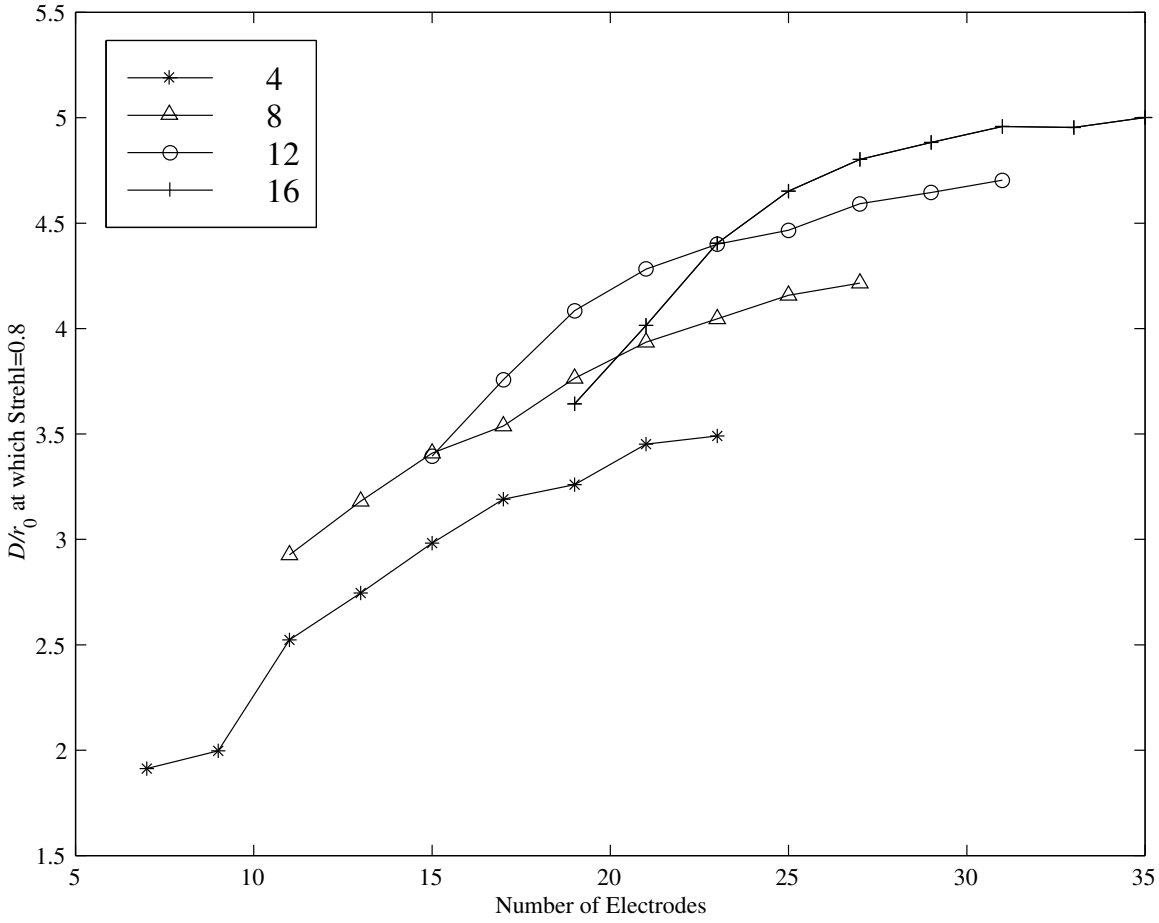


Figure 4.12 Mean RMS voltage required for the optimised arrangement

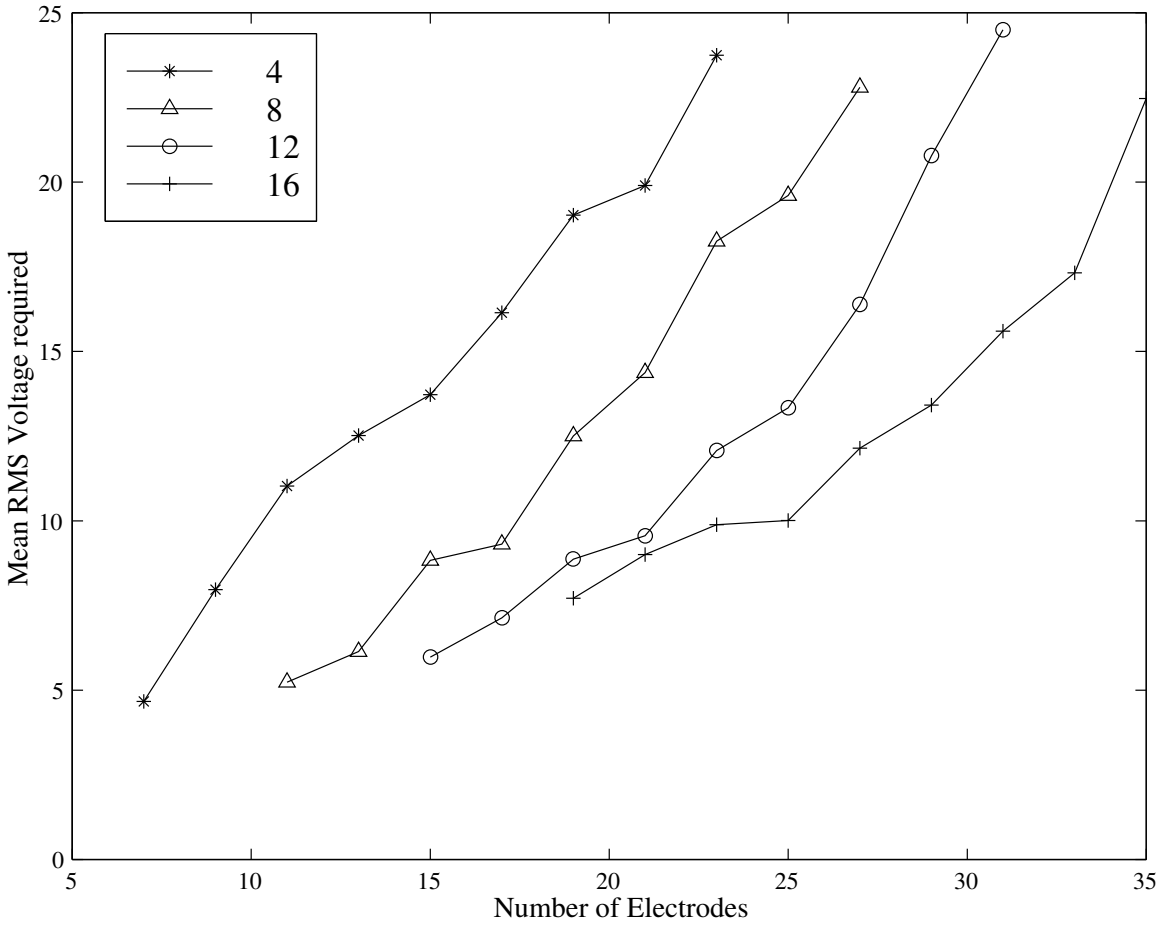
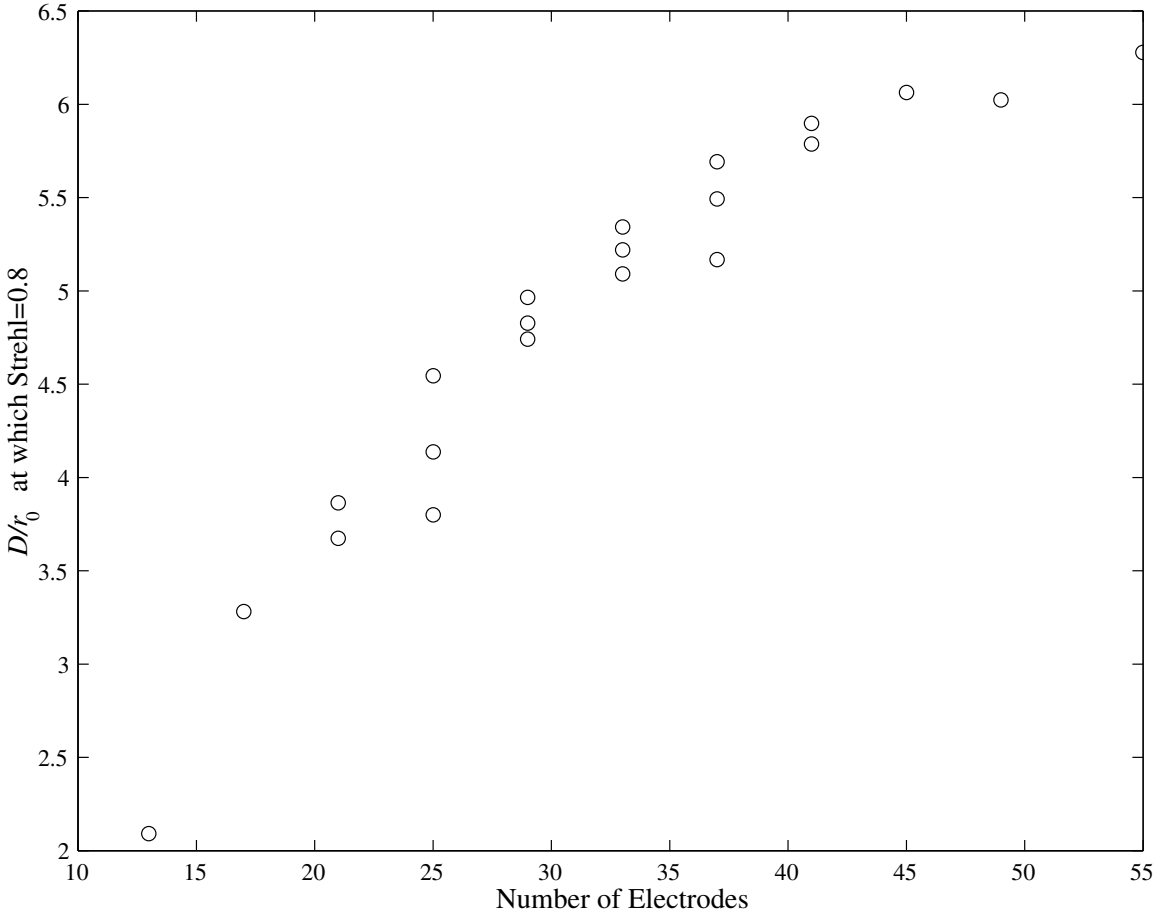


Table 4.1 Results from the optimisation of a 3 ring corrector

Input Parameters			Optimised Parameters			Results		
N	n_2	n_3	r_1	r_2	ρ	μ	D/r_0	$\bar{V}$
7	2	4	0.25	0.5	0.8	11.8	1.91	4.67
9	4	4	0.25	0.6	0.95	11.7	2	7.97
11	2	8	0.2	0.4	0.7	11.8	2.93	5.23
11	6	4	0.25	0.65	1	11.6	2.52	11
13	4	8	0.2	0.5	0.8	11.8	3.18	6.14
13	8	4	0.2	0.65	0.95	11.9	2.75	12.5
15	2	12	0.2	0.4	0.75	11.9	3.4	5.98
15	6	8	0.2	0.5	0.85	11.7	3.41	8.83
15	10	4	0.2	0.75	1	14.3	2.98	13.7
17	4	12	0.2	0.45	0.8	11.8	3.76	7.14
17	8	8	0.15	0.55	0.8	13.5	3.54	9.31
17	12	4	0.15	0.75	0.95	17	3.19	16.1
19	2	16	0.2	0.35	0.8	13.8	3.64	7.71
19	6	12	0.2	0.5	0.85	11.7	4.08	8.87
19	10	8	0.15	0.65	0.9	14.7	3.76	12.5
19	14	4	0.15	0.8	1	22	3.26	19
21	4	16	0.2	0.4	0.8	13	4.02	9
21	8	12	0.15	0.55	0.8	13.7	4.28	9.56
21	12	8	0.15	0.65	0.85	17.5	3.94	14.4
21	16	4	0.15	0.8	0.95	24.6	3.45	19.9
23	6	16	0.15	0.45	0.8	12.5	4.41	9.88
23	10	12	0.15	0.65	0.9	16.5	4.4	12.1
23	14	8	0.15	0.75	0.95	22.6	4.05	18.3
23	18	4	0.15	0.85	1	30.3	3.49	23.7
25	8	16	0.15	0.5	0.75	13.4	4.65	10
25	12	12	0.15	0.65	0.85	20	4.47	13.3
25	16	8	0.15	0.7	0.85	25.1	4.16	19.6
27	10	16	0.15	0.6	0.85	16.9	4.8	12.1
27	14	12	0.15	0.7	0.9	22.9	4.59	16.4
27	18	8	0.15	0.75	0.9	31.3	4.22	22.8
29	12	16	0.15	0.6	0.8	20.3	4.88	13.4
29	16	12	0.1	0.65	0.8	26.7	4.64	20.8
31	14	16	0.15	0.65	0.85	25.1	4.96	15.6
31	18	12	0.1	0.7	0.85	30.9	4.7	24.5
33	16	16	0.15	0.6	0.75	30.4	4.95	17.3
35	18	16	0.1	0.65	0.8	32.1	5	22.5

Figure 4.13 Values of D/r_0 obtained with a 4 ring corrector. Circles with crosses are the points for each total number of electrodes where the correction is best.



required). This is a reasonable constraint to impose, since previous experience has shown that a good condition number never occurs if the rings are too thin. Because of the time taken to examine each arrangement, only 21 example arrangements were considered. A graph showing the limiting value of D/r_0 achievable for the various configurations is shown in figure 4.13. The mean RMS voltages required are shown in figure 4.14, and the results are summarised in detail in table 4.2.

4.4.4 Comparison between 3 and 4 ring correctors

By plotting the best correction achievable for each total number of electrodes, we can compare directly whether a 3 or 4 ring mirror gives the best correction for the same total number of electrodes. This plot is shown in figure 4.15. The lines are unweighted least-squares fits to the *ad hoc* fitting function:

$$(D/r_0)_{0.8} = C_1 + C_2 \log(N_e) \tag{4.10}$$

This can be seen to give a reasonable fit, and the fitting parameters are shown below:

Number of rings	Parameter C_1	Parameter C_2
3	-2.241	2.119
4	-5.096	2.952

Figure 4.14 Mean RMS voltage required with a 4 ring corrector. Circles with crosses are the points for each total number of electrodes where the correction is best.

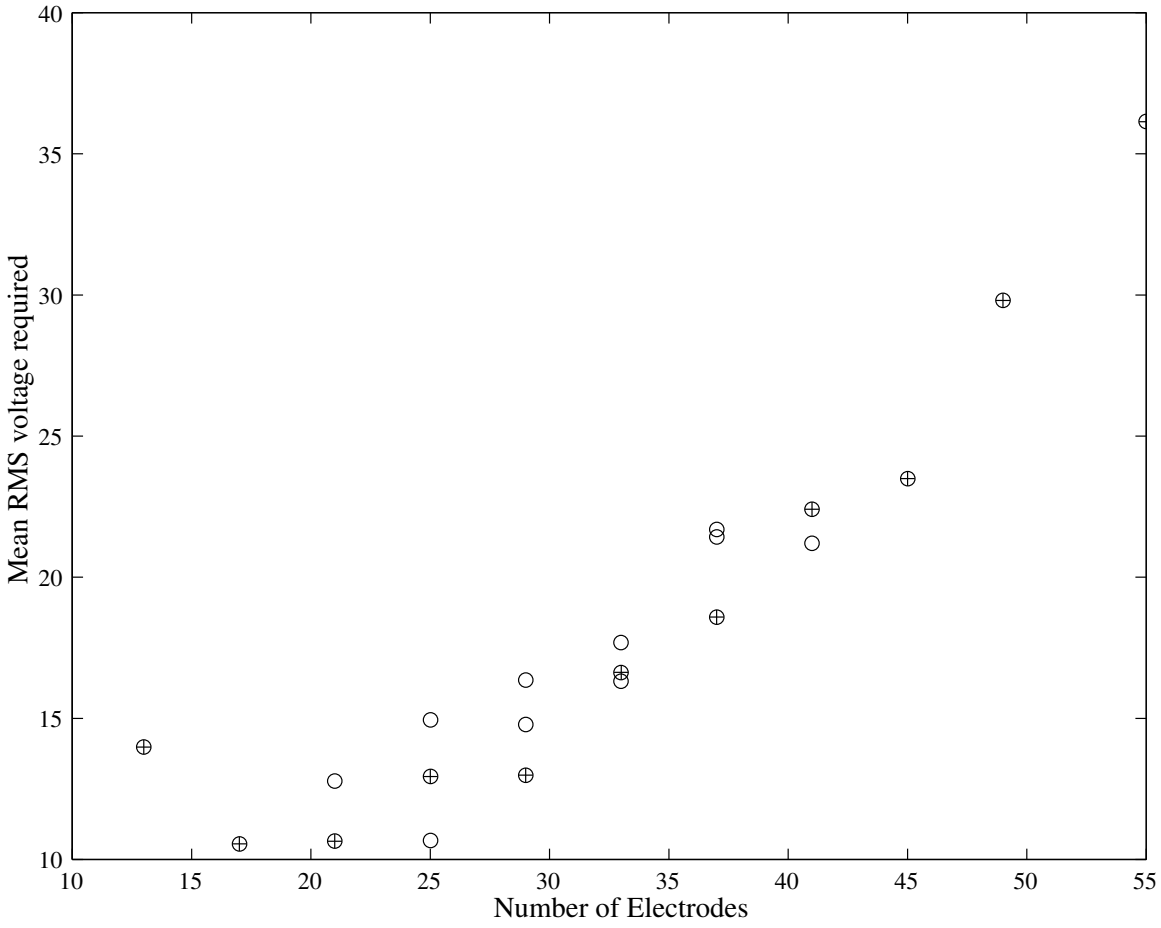
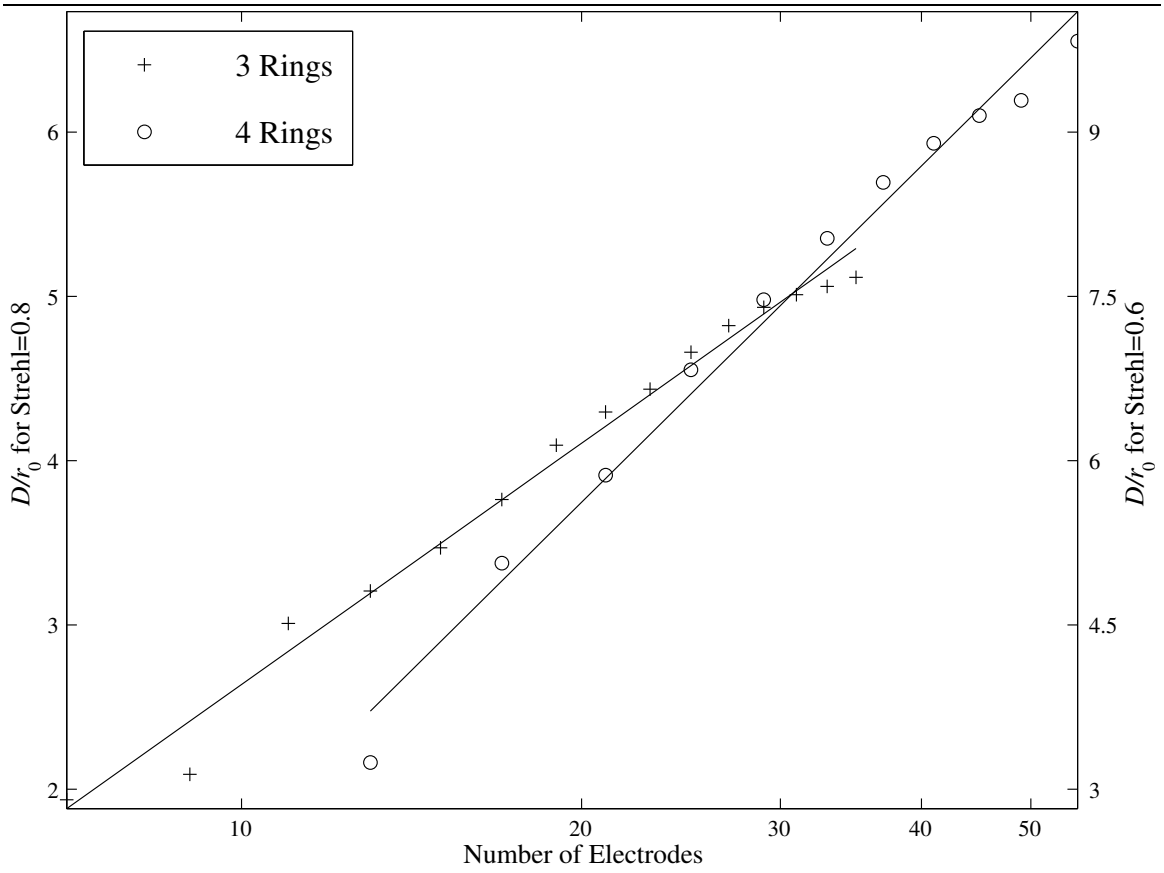


Table 4.2 Results for a 4 ring corrector

Input Parameters				Optimised Parameters				Results		
N	n_2	n_3	n_4	r_1	r_2	r_3	ρ	μ	D/r_0	$\bar{V}$
13	4	4	4	0.2	0.45	0.7	0.95	20.8	2.09	13.98
17	4	4	8	0.2	0.4	0.6	0.85	21.31	3.28	10.55
21	4	4	12	0.15	0.35	0.55	0.8	21.63	3.86	10.64
21	4	8	8	0.2	0.4	0.65	0.9	21.29	3.67	12.78
25	4	4	16	0.15	0.35	0.5	0.75	21.39	4.14	10.67
25	4	8	12	0.15	0.35	0.6	0.85	21.68	4.54	12.94
25	8	8	8	0.15	0.5	0.7	0.9	25.39	3.8	14.94
29	4	8	16	0.15	0.35	0.55	0.8	21.37	4.97	12.99
29	4	12	12	0.15	0.35	0.65	0.9	22.33	4.83	16.36
29	8	8	12	0.15	0.45	0.65	0.85	25.95	4.74	14.78
33	4	12	16	0.15	0.35	0.65	0.9	22.53	5.34	16.63
33	8	8	16	0.1	0.4	0.6	0.8	27.47	5.22	16.33
33	8	12	12	0.15	0.45	0.65	0.85	26.23	5.09	17.68
37	4	16	16	0.15	0.3	0.7	0.9	29.06	5.49	21.69
37	8	12	16	0.1	0.4	0.6	0.8	27.87	5.69	18.59
37	12	12	12	0.1	0.5	0.7	0.85	38.45	5.17	21.43
41	8	16	16	0.15	0.45	0.75	0.95	30.54	5.9	22.41
41	12	12	16	0.1	0.45	0.65	0.8	39.02	5.79	21.2
45	12	16	16	0.1	0.45	0.65	0.8	39.1	6.06	23.5
49	16	16	16	0.1	0.55	0.7	0.85	54.15	6.02	29.81
55	18	18	18	0.1	0.6	0.75	0.9	65.17	6.28	36.14

Figure 4.15 Comparison between 3 and 4 ring correctors



This is a purely empirical fit; the expected variation of D/r_0 with number of electrodes would probably be expected to require that number of electrodes squared would be proportional to D/r_0—this does not fit well with the results found here.

4.4.5 Optimising for non-Kolmogorov turbulence

The procedure outlined above could equally well be applied to any given power spectrum of turbulence, not just Kolmogorov-type turbulence. All that is required is a new covariance matrix for the Zernike coefficients in the presence of this turbulence. Optimisations for specific surface shapes (e.g. for a system with a pre-calculable gamut of possible aberrations) could, of course, be performed; in the absence of any sensible specific shapes, an optimisation for low-order Zernike modes has been carried out for demonstration purposes.

The design goal used in the Zernike optimisation was a simple one: it was required that the lowest 21 Zernike orders should be corrected as efficiently as possible. The performance metric used was the sum of the residual phase variances for fitting one radian^2 of the Zernike orders individually. This can be calculated very simply from the Zernike transfer matrix and its pseudoinverse:

$$\text{metric}_{21} = \sum_{i=1}^{21} \left[\sum \left(\mathbf{H} \times \mathbf{H}^{+} \right) - \mathbf{I} \right] \tag{4.11}$$

where the inner summation is taken over rows of the resulting matrix.

This optimisation was performed for a series of 3-ring mirrors, and the best values of the metrics found are shown in figure 4.16. An examle of the residual variance for fitting the first 21 Zernike

modes is shown in figure 4.17—clearly the modes are matched very well indeed by this 33 electrode mirror. The disadvantage of optimising simply for the fitting error is that the voltages required to reproduce the modes are very high indeed; for most modes, it would be only just possible to achieve one wavelength of surface displacement in the visible region. This is almost certainly unacceptable in any real application.

Figure 4.16 Zernike fitting errors: best values of metric_{21}

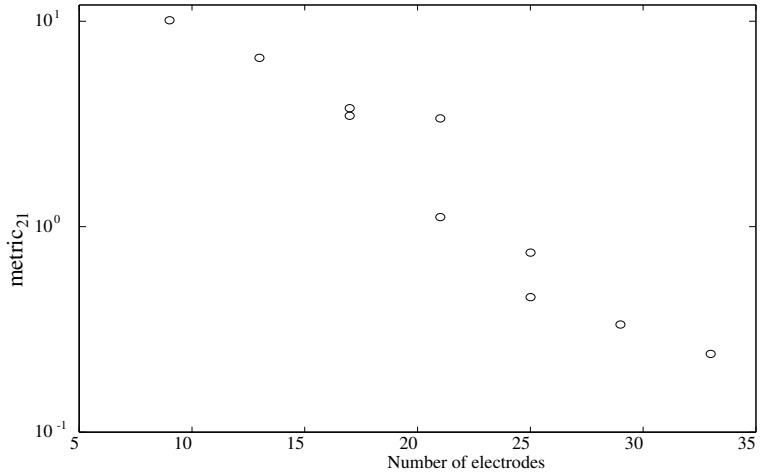
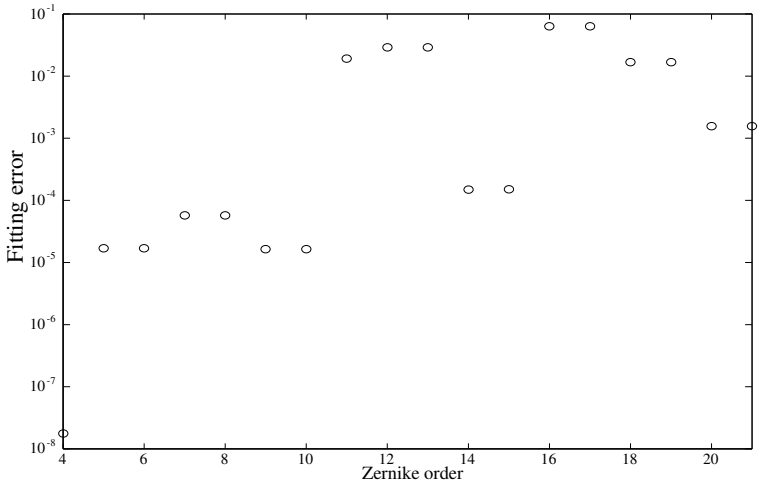


Figure 4.17 Zernike fitting errors for a 33 electrode mirror

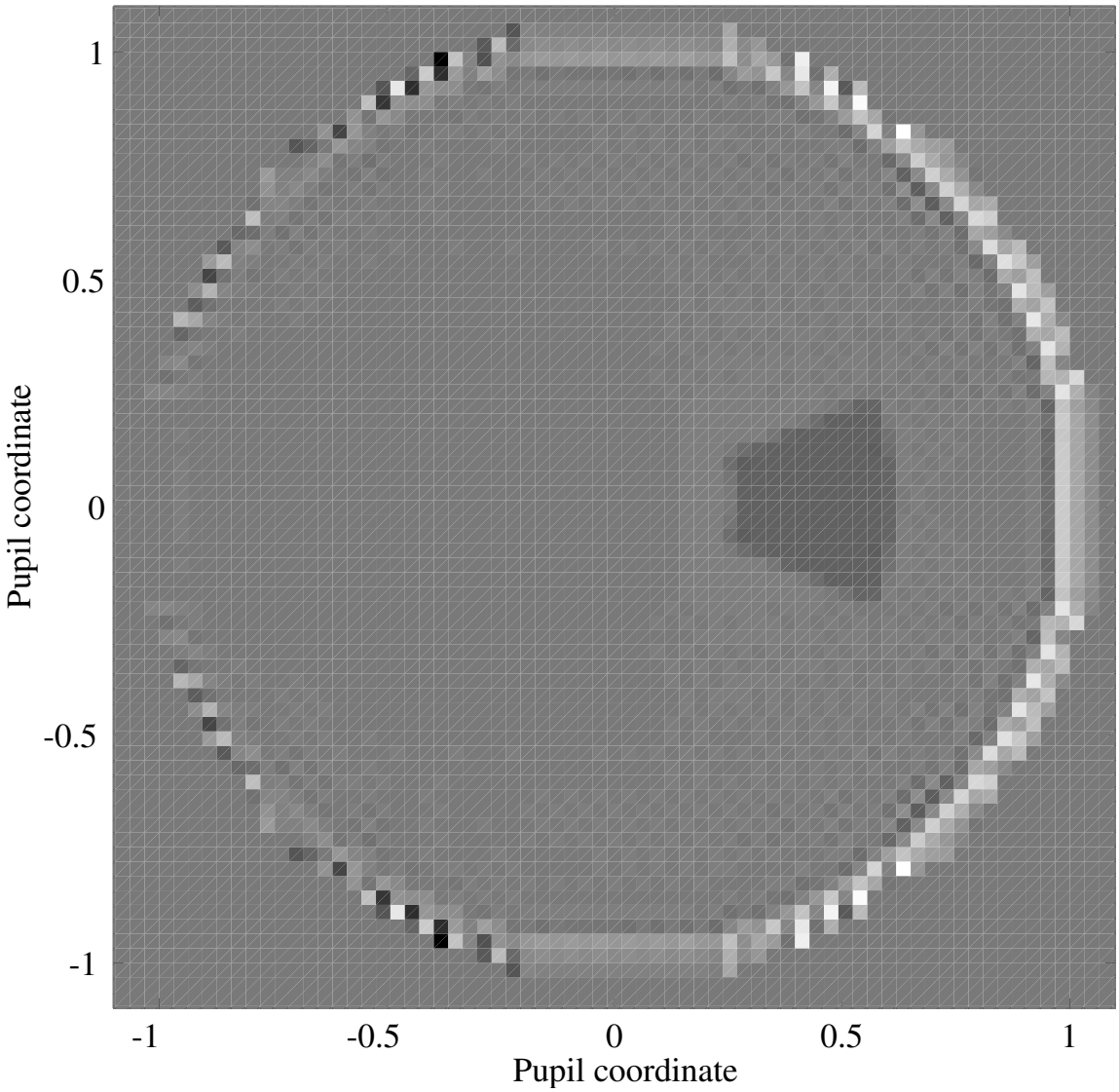


4.5 Interaction with Curvature Sensors

Now that we have some mirror designs that have been optimised for correcting Kolmogorov-type, we can test how well these mirrors would be sensed by a curvature sensor. In this section, only a brief overview the interaction will be covered; the behaviour of the curvature sensor is quite complex and nonlinear, and a thorough treatment is beyond the scope of this thesis.

The curvature sensor is modelled by performing Fresnel diffraction calculations. As described in chapter 2, any series of ideal optical components is equivalent to a transverse scaling, the addition of an average phase curvature and a free-space propagation. Clearly, it is the free-space propagation that is of paramount importance; the other stages will have no effect on the measured curvature

Figure 4.18 Example curvature sensor output

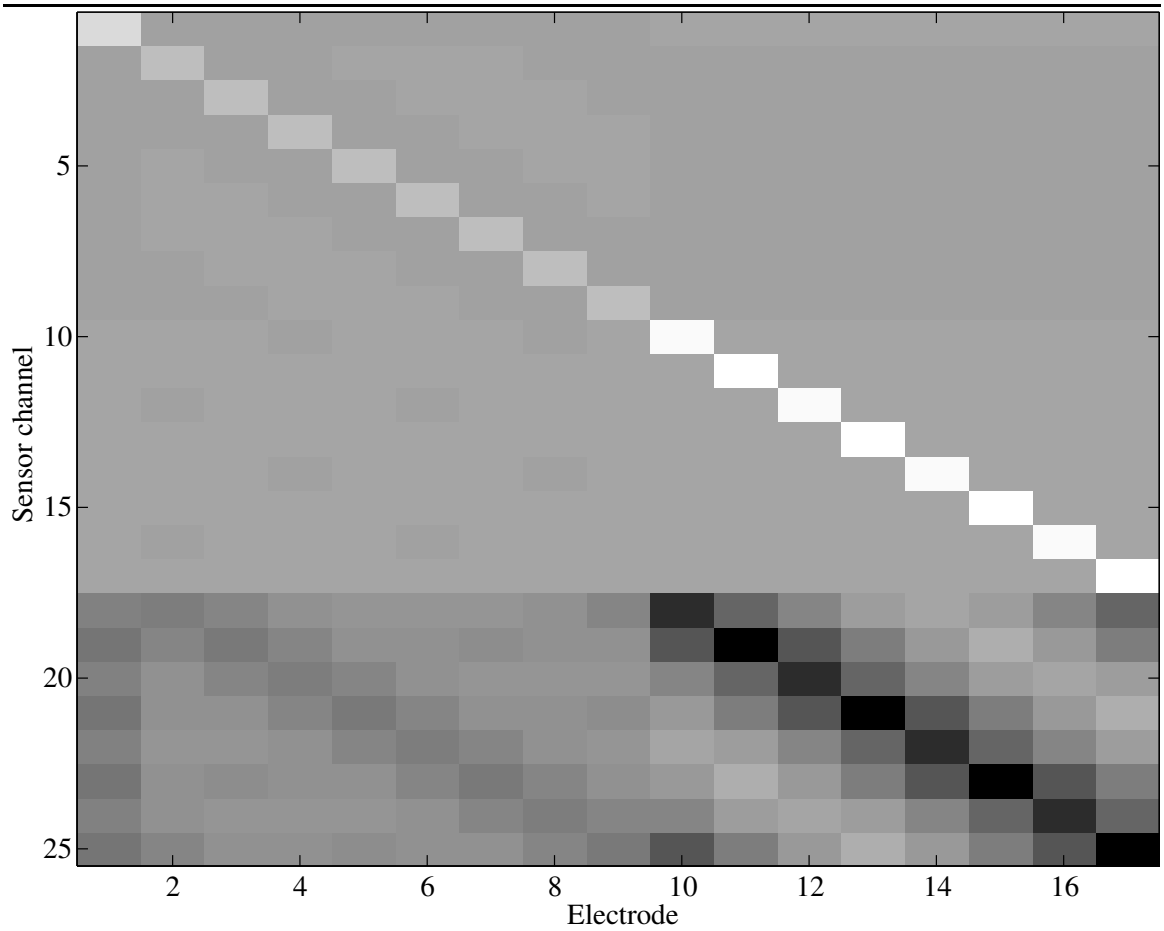


sensor signal. In the Fresnel propagation, the important parameter is the "Fresnel number", which is defined as $N = a^2/z\lambda$, where a is the width of the beam, z is the distance that the beam has propagated. The amount of phase excursion applied by the mirror is also very important. As mentioned by Rigaut *et al.* [75], the curvature sensor does not have a linear response. In this brief examination of the technique, we will attempt to stay in the linear portion of the sensor's response.

An example of a curvature sensor signal is shown in figure 4.18. This was created using an electrode response function that extended from $r = 0.25$ to $r = 0.6$, and its angular extent was $\pi/4$. In fact, the region of the electrode is clearly visible in the pseudocolor plot. The other parameters involved in making this plot were: Fresnel number $N = 2000$; input surface variance 1 radian^2.

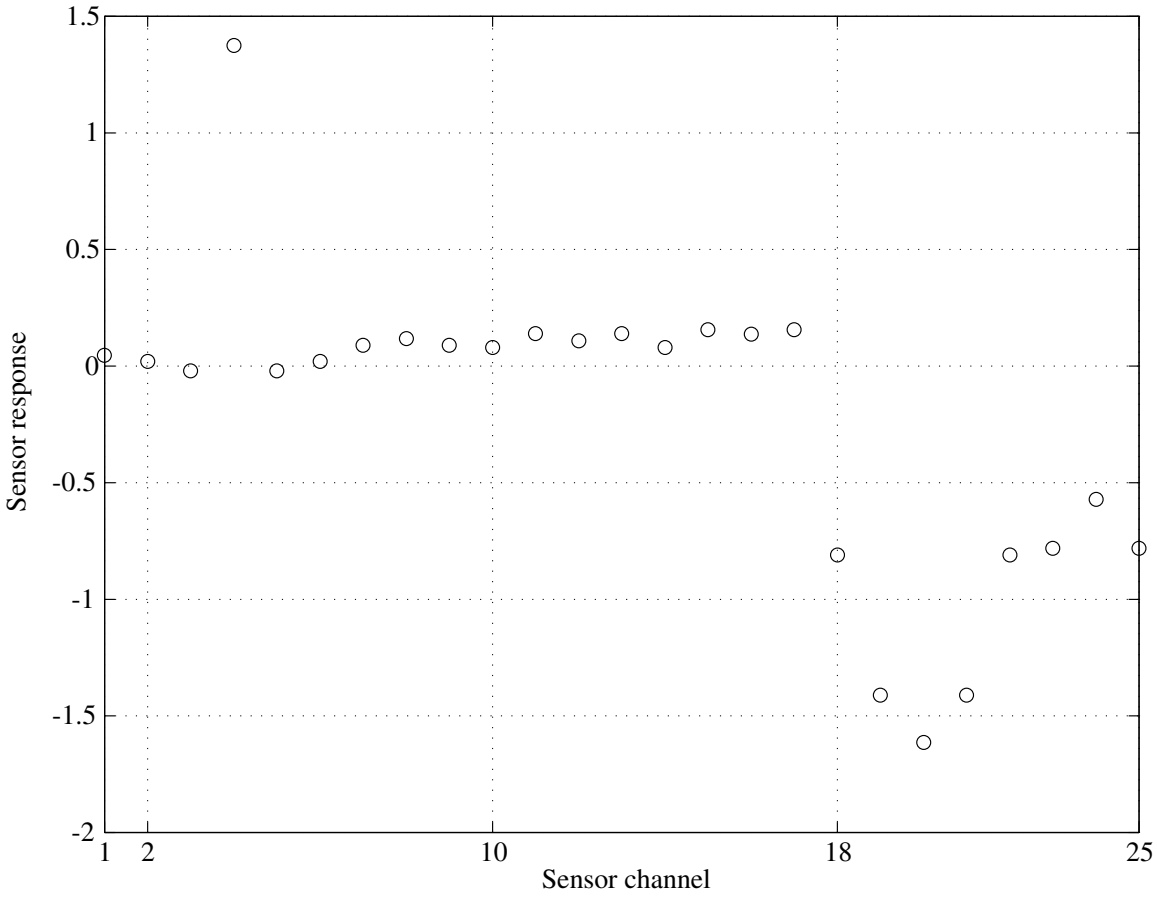
Using a Zernike transfer matrix created as in the preceeding sections, we can build up a mirror-sensor interaction matrix. However, rather than use the whole image (as in figure 4.18), it is preferable to divide the pupil into separate integrating regions. This was done in an empirical manner. A configuration that was found to give good results (in the sense of giving a good condition number) was to treat each electrode as an integration area, with a slight amount of "padding" between the

Figure 4.19 Mirror-curvature sensor transfer matrix



areas, and also to have an extra ring of "edge sensors", to sense the boundary conditions. A transfer matrix built in this way is shown in figure 4.19. This is clearly an almost-diagonal transfer matrix, and it is also clear that each mode is well sensed by the first 17 sensors—those within the optical pupil. A sample electrode response function (one of the columns of the transfer matrix of figure 4.19) is shown in figure 4.20. This shows more clearly than the pseudocolor plot does the non-diagonality of the response. The condition number of this transfer matrix is 14.9, which is slightly better than that of the mirror alone (i.e. the mirror–Zernike transfer matrix), whose condition number is 19.5. This implies that a curvature sensor does indeed sense well a bimorph mirror. However, the question of whether or not a curvature sensor senses atmospherically induced phase fluctuations efficiently is somewhat moot: it is known [75] that the curvature sensor is considerably more sensitive to modes whose Laplacian is non-zero.

Figure 4.20 A single electrode curvature-sensor response



5 Practical Bimorph Mirrors

Although bimorph mirrors have been known in the open literature for some 20 years now, there are still very few commercially available devices. Those that can be purchased are rather expensive, costing somewhere in the region of £1000 per channel. This cost may be prohibitive for many applications, such as thermal lens correction inside a Nd:YAG laser cavity. Therefore, a series of experiments was undertaken to attempt to construct bimorph mirrors, with the intention of constructing them using inexpensive, commercially available components.

This chapter will describe the fabrication process used, and the results achieved using this approach. In addition, the results obtained will be analysed and compared to the theory developed in the preceeding two chapters.

5.1 Practical Design and Fabrication Procedure

When designing a practical device, one must recall the fundamental equations derived in the preceeding chapters. As mentioned in the introduction to chapter 4, there are two somewhat independent design choices to be made: the thicknesses and also the electrode layout.

When choosing the thicknesses for a practical device, one must bear in mind the limitations imposed by the whole of the fabrication process. In order to design the mirror, the order in which the various operations are performed must be considered. The order found to give best results is:

1. Acquire piezo and substrate discs
2. Ensure substrate is adequately flat
3. Glue discs together
4. Attach mirror mount
5. Attach control wires
6. Polish mirror until desired flatness achieved
7. Apply reflective coating

Each of the above points requires some discussion as to how it affects the design.

5.1.1 Acquire piezo and substrate discs

There are many varieties of piezo ceramic commercially available. They each have a different set of electromechanical properties. A choice must be made as to which is the best for the purposes of constructing a bimorph mirror. Referring to equation 3.51 on page 56, we see that the total amount of curvature achieved is directly proportional to the d_{31} value of the piezo ceramic used. Since the devices are to be used at relatively low frequencies (typically $\leq$ 1kHz), then we are free to choose

Table 5.1 Piezo ceramic properties

Property	PIC 151	PIC 155	PZT5H
d_{31} (10^{-12}mV^{-1})	-210	-140	-274
Curie Temperature (°C)	250	320	195
Elastic Constant (S_{11}^E in 10^{-12} m^2 N^{-1})	15.0	13.2	16.4
Dielectric Constant ($\varepsilon_{33}/\varepsilon_0$)	1700	2100	3400
Supplier	Physike Instrumente		Matroc

the ceramic with the highest d_{31} value (the other ceramics that are available with smaller d_{31} values are designed to be used in high-power ultrasound applications). Table 5.1 shows the properties of some commercially available piezo ceramics. In all the results presented below, PIC 151 was used, as this is available in disks 0.3mm thick at 45mm diameter.

The choice of substrate is governed by four concerns:

1. How well the thermal expansion coefficient matches that of the piezo ceramic
2. How well the mechanical properties match that of the piezo ceramic
3. How easily the material can be polished
4. How easily the material can be made reflective at the wavelength of interest

In this work, most of the devices have been made from pyrex. This has a thermal expansion coefficient that is very close to that of piezo ceramics. It is also well matched mechanically. It is not the most easily polished glass, but it is one that is readily available: it may certainly be possible to find a glass with the right thermal coefficient that is easy to polish. Pyrex can be coated in the normal way: sputter coating. If the device were to be used at CO_2 wavelengths, then copper may be a more appropriate choice as substrate material.

5.1.2 Ensure substrate is adequately flat

The substrate does not need to be perfectly flat before gluing to the piezo layer. This is because the gluing will almost inevitably introduce some strain into the structure, and so a substrate that is initially flat will almost certainly no longer be flat after gluing: since a final polish is required in any case, it makes sense not to expend too much effort in requiring the substrate to be initially extremely flat.

5.1.3 Glue discs together

The adhesive layer is the single most crucial factor in the production of a bimorph mirror. Its thickness, uniformity and elastic properties can all have an overwhelming effect on the overall performance of the device.

Dealing with the elastic properties first: most adhesives are nowhere near as mechanically rigid as either glass or piezo ceramic. This means that there will be some "give" in the glue layer. This situation represents a "shear deformation" of the glue layer in the laminate, and is thus not modelled by the non-shear deformable theory developed in chapter 3.

To minimise the shear deformation effects, it is imperative to use as thin a layer of adhesive as is reasonably possible. This usually means that a low-viscosity adhesive is preferable. Another important property of an adhesive is that it should shrink as little as possible on curing, otherwise

Table 5.2 Adhesive Properties

Supplier	*Product*	*Type*	*Comments*
Permabond	UV7349	UV-curing	Fairly low-viscosity.
Permabond	QB5002	Toughened Acrylic	Relatively high viscosity, high strength
Summers Optical	P-92	UV-curing	Zero shrinkage-on-cure, but relatively elastic
Summers Optical	SK-9	UV-curing	Low-shrinkage, higher Young's modulus than P-92
Holdtite	ST3500	UV-curing	Very low viscosity possible

the structure will be severely deformed and it may not be possible to polish the structure back to flat. Heat curing of an adhesive is a distinct disadvantage, since no matter how well the thermal properties of the substrate and the piezo are matched, there will still be some differential expansion, and so there will be residual stresses once the temperature of the device returns to room temperature. If the device is to perform reliably, the adhesive should have constant properties with time, after the initial cure.

The overall effect of the above considerations is that we are left with a few possible classes of adhesive that may give good results. Epoxy resins have not been used, as these are often rather brittle, and often do not age well. Some manufacturers do supply specialised, high-performance epoxies; these have not been investigated in any great detail. Most attention has been focused on UV-curing 1 part adhesives. These have the advantage of generally being quite low in viscosity (or can be made so by some suppliers). Also, toughened acrylic two-part adhesives have been used—these are particularly useful if there is no possibility of using a UV cure (e.g. in between two layers of piezo). The properties of some of the adhesives used are summarised in table 5.2. Each of these adhesives has been tried.

5.1.4 Mirror mounting

There are various options for mounting the mirror once the structure has been glued together. The mounting effectively determines the boundary conditions. The options considered in this work have been:

- Loosely held at the edge
- Attached to a central post
- Firmly clamped at the edge

Although holding the device loosely at the edge seems at first to offer the opportunity of allowing the edge to deform, it was found to be difficult to obtain uniform "loose clamping". Thus, this option was not investigated in any great detail.

Most of the mirrors produced have been mounted by a central post, which is attached by a slightly flexible adhesive. This flexibility allows the central portion to change its curvature. The boundary conditions that this mounting method imposes are zero slope at the centre of the device. This is not likely to be much of a disadvantage since it is usual to use a deformable mirror in series with a dedicated tip-tilt mirror.

If the mirror radius is small, then it may be advantageous to constrain the mirror by gluing it rigidly to a ring. This approach gives an device with better overall mechanical rigidity. There are two

Table 5.3 Resulting resonant frequencies of various tip-tilt stages with bimorph mirror attached

Bimorph characteristics		Tip-tilt stage	f_0	I_0	f'
Glass thickness	1.5 mm	S226	9 kHz	215 g mm ^{2}	2.8 kHz
Glass density	2.23 g cm^{-2}	S330	3.3kHz	1530g mm^2	2.1 kHz
Piezo thickness	0.3 mm	S340	1.4 kHz	18000 g mm^2	1.3 kHz
Piezo density	7.8 g cm^{-2}				
Radius	22.5 mm				
Post height	5 mm				
Moment I_m	1460 g mm ^{2}				

disadvantages to constraining the gradient to be zero at the edge: firstly, the effective pupil that can be used will be considerably smaller than the size of the substrate. This is important since it is the aspect ratio of the device that is limited by the polishing procedure. The second disadvantage is that the stroke of the device is somewhat reduced. This can easily be understood when one considers that, since the gradient is constrained at the edge, the bending energy is used bending the mirror in two opposite directions. Numerical simulations of a response function for free-edge and clamped-edge are shown, with the same vertical scaling, in figure 3.15 on page 74.

In any mounting method, there is the possibility of mounting the whole structure on a tip-tilt stage. The main requirement that would enable this is that the overall mass of the device should be low. Physike Instrumente supply several fast tip-tilt stages specifically designed for optical use, and they also give simple expressions that enable the expected resonance frequency to be evaluated. Firstly, one must find the moment of inertia of the mirror; this can be found from:

$$I_m = m\left[\frac{3r^2+h^2}{12} + \left(\frac{h}{2}+t\right)^2\right] \tag{5.1}$$

where r is the mirror radius, h is the mirror thickness, m is the mirror mass and t is the distance from the mirror to the axis pivot. The resulting resonant frequency is calculated from the resonant frequency of the unloaded tip-tilt stage, f_0, as follows:

$$f' = \frac{f_0}{\sqrt{1+\frac{I_m}{I_0}}} \tag{5.2}$$

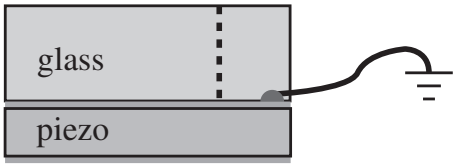
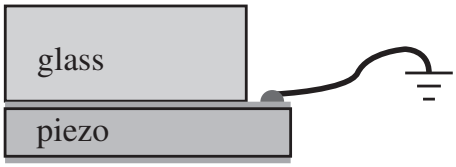
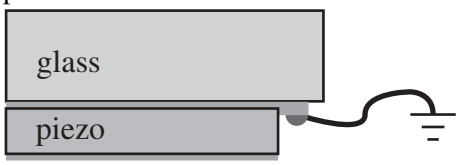
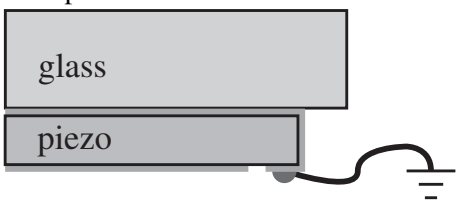
where I_0 is the moment of inertia of the tip-tilt platform. Typical values for the parameters and resulting resonant frequencies are given in table 5.3. In each case, the mirror's moment of inertia is calculated using the sum of the post height and the distance from the tip-tilt stage surface to its pivot point. The results here show that mounting a bimorph mirror on a tip-tilt stage could give a resonant frequency that is easily sufficiently high to correct for atmospheric turbulence.

5.1.5 Attach control wires

Although the attachment of the electrode wires seems a trivial point, it is worth pointing out a few matters found to be of practical importance when constructing the devices.

Firstly, it is advisable to use a small amount of conductive epoxy resin to attach the wires, since the amount of heat applied to the piezo ceramic during soldering is not easily controllable, and could easily lead to localised loss of piezoelectric sensitivity. Also, it is preferable to glue the wires to the electrode regions before the final polishing stage—the regions where the glue is applied are

Table 5.4 Attaching earthing wires to a bimorph structure

Attachment technique	Comments
Wire attached through slot cut in (upper) glass layer. 	This has the advantage that the piezo layer does not protrude beyond the glass, however the slots cut in the glass have a rather detrimental effect on the surface shape of the mirror.
Piezo layer protruding beyond glass. 	Although better than the above technique, in order to protect the piezo layer from buffeting in the polishing machine, a large disc must be attached to the mounting post. This loads the mirror unevenly during polishing.
Rear electrode drawn out with silver paint. 	In this case, the glass protrudes beyond the piezo layer, so the structure can be polished more easily. Achieving a good contact is sometimes difficult—the contact can only be tested after the layers have been glued together.
"Wrap-around" electrode 	In this case, a small portion of the "control" side of the piezo disc is electrically connected with silver paint to the ground electrode. This has the advantage that the contact can be tested before gluing, and has minimal impact on the glue layer. This technique is due to Morgan Matroc.

noticeably affected. It is imperative that these effects are minimised, since they will inevitably be uncorrectable by the device, since their spatial extent is smaller than an electrode.

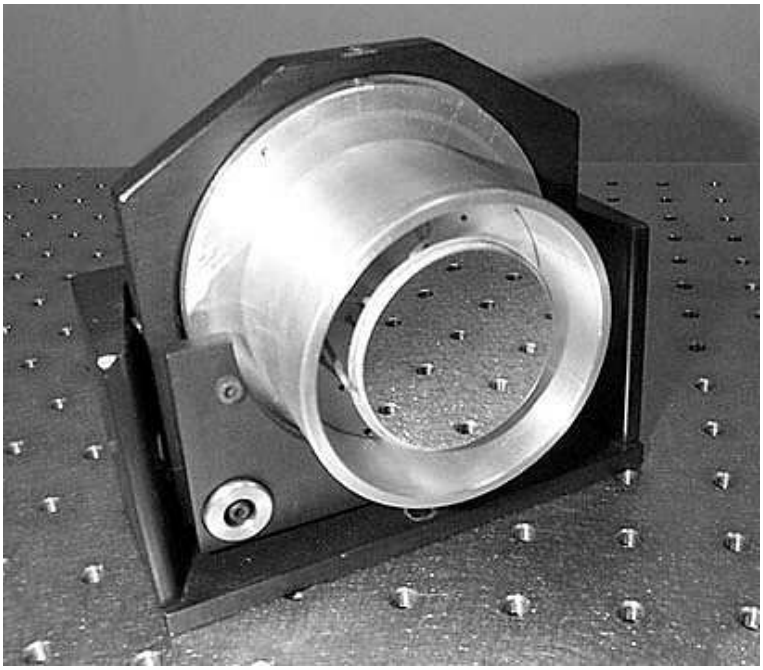
Attaching the wire to the common electrode of a bimorph device requires some thought. Various methods have been tried, with varying degrees of success. These strategies are shown in table 5.4, along with a discussion of their success.

Once the control wires have been succesfully attached to the electrode structure, it is desirable to have some sort of strain-relief stage in between the mirror and the final electrical connection. This ensures that connecting the cable to the high power amplifier will have a minimal effect of the alignment of the mirror in an optical system.

5.1.6 Final polishing

As mentioned above, it is greatly desirable to perform all possible stages of construction and assembly before the final polish. Bimorph mirrors are very thin devices, and almost any processing is sufficient to distort the surface figure. Certainly, it is imperative that any mounting structure that is to be glued directly to the mirror (such as a central post, or and edge-mounting ring) should be attached before polishing. In the final polish, it is desirable to achieve as flat a surface figure as possible. Even though the mirror will be able to remove some of the distortions in operation, this

Figure 5.1 Photograph of a bimorph mirror



will use up part of its dynamic range. Also, any distortions that are of a higher spatial frequency than the electrode distribution will be unremovable.

5.1.7 Apply reflective coating

Once the mirror has the desired surface flatness, it must be coated with a suitable reflective layer. Unfortunately, many coating processes involve heating of the substrate. This is undesirable for two reasons. Firstly, if the temperatures exceed the Curie temperature of the piezoceramic (usually around 250° C), it will become depolarised. This depolarisation is reversible, however the repolarisation procedure will have an unpredictable effect on the "off" position of the piezoceramic. Secondly, any heating of the bimorph structure at all will cause permanent deformation. This is because most adhesives will change their properties somewhat at elevated temperatures, and allow some slipping of the layers. Therefore, when the device returns to room temperature, some deformation will be "frozen in". These problems do not preclude reflective coating: devices made in the course of this work have succesfully been coated with reflective layers of silver and aluminium.

5.2 Bimorph Amplifier

Throughout the practical studies of bimorph mirrors, the same high-voltage supply was used. This device had certain limitations that made some of the experimental investigations (especially those relating to the temporal properties of bimorph mirrors) difficult. The amplifier was a custom-made device, supplying ± 150 V in 17 channels. The amplifier was controlled from a PC via the parallel port. This meant that the channels had to be addressed serially, and there was a limitation on the maximum rate at which one could address the channels. This rate was approximately 250 Hz, implying a "frame" refresh rate of around 36 Hz. The voltage range was divided up into 256 discrete levels (i.e. 8 bits of information per channel). The quasistatic performance of the amplifier was evaluated by measuring, for each channel, the output voltage for a given input signal, over a range

Figure 5.2 Amplifier calibration of channel 0

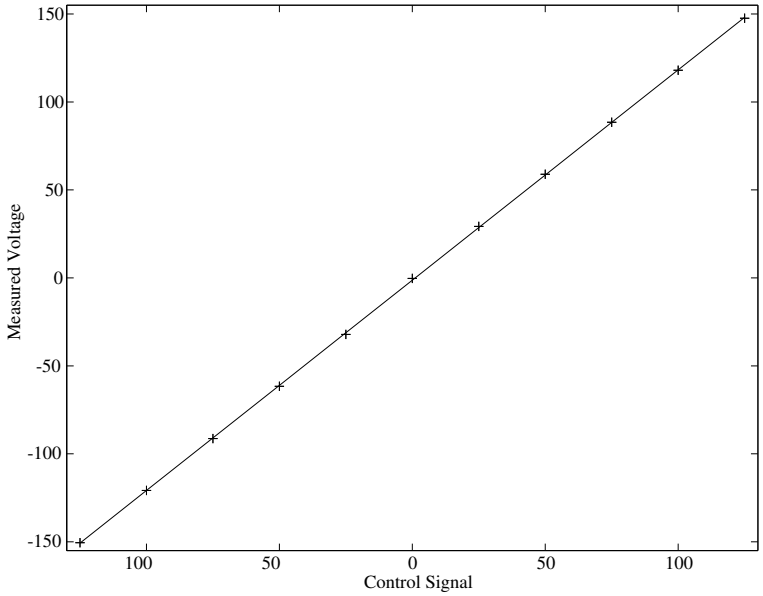


Table 5.5 Least-square fitted amplifier calibration

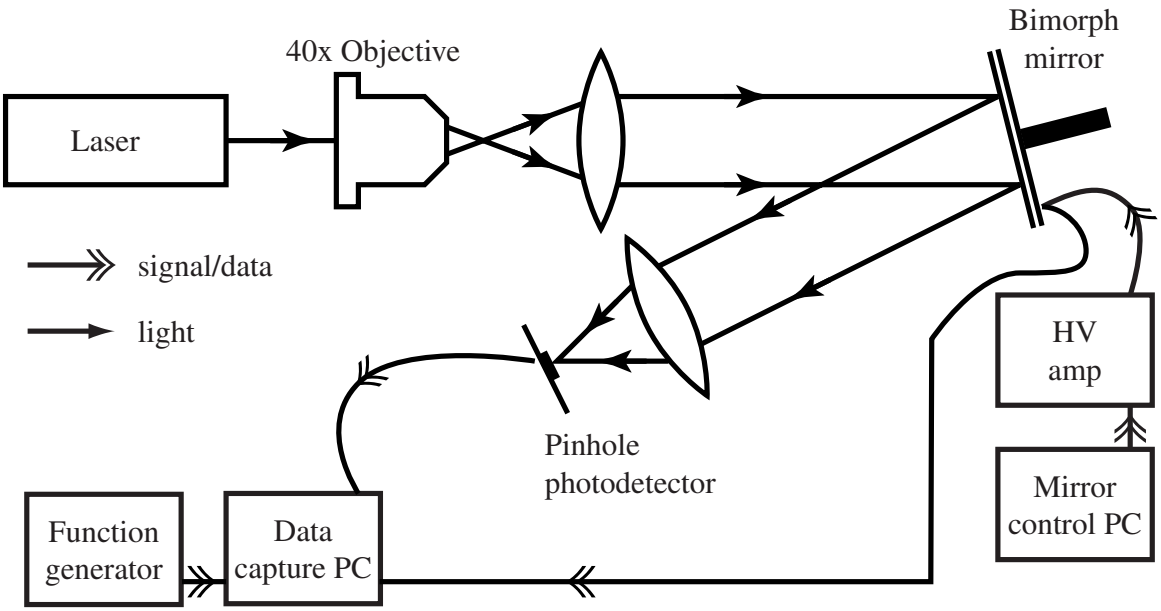
Channel	a_i	b_i	Channel	a_i	b_i
0	-1.3	1.196	8	-2.1	1.195
1	-1.4	1.195	9	-1.4	1.207
2	1.0	1.200	10	0.9	1.205
3	0.8	1.195	11	0.8	1.195
4	-0.4	1.205	12	1.5	1.200
5	0.6	1.199	13	1.5	1.191
6	-1.2	1.192	14	1.4	1.190
7	0.1	1.192	15	0.7	1.195
			16	-2.2	1.196

of input signals. A typical channel's response, along with a best-fit line, is shown in figure 5.2. The line was fitted to the expression:

$$V_i = a_i + b_i U_i \tag{5.3}$$

where U_i is the signal applied to the i^{th} channel. Error bars are not shown: the voltages were measured several times, and were found to be the same to within 0.1 V in each instance. The fitted parameters are given in table 5.5. For convenience, for the rest of this chapter, except where explicitly specified, signals will be measured in amplifier units. Where comparison with actual voltages are required, then the appropriate conversion can be performed using table 5.5. These data clearly show that the gradient is almost identical for each channel; the zero offsets vary slightly from channel to channel.

Figure 5.3 Practical layout for an all-optical measurement of a bimorph dynamic response



5.3 Dynamic Response

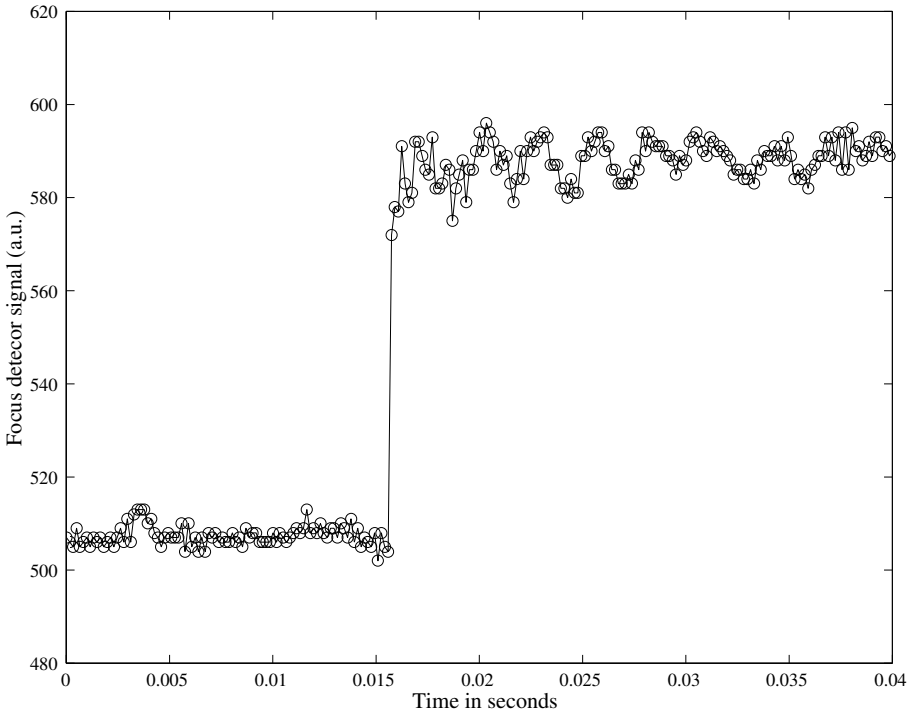
There are several possible ways of estimating the dynamic characteristics of a bimorph mirror. It was not possible to attempt many of these due to equipment or amplifier limitations. For example, it was not possible to perform any measurements that involve large amplitude smoothly varying sinusoidal waveforms, as these could not be produced with the available high-voltage amplifier described above. Two different methods were chosen: one optical, and one purely electrical.

5.3.1 Optical measurement

The experimental arrangement for a purely optical measurement of the response time is shown in figure 5.3 . The bimorph mirror was illuminated by a slightly converging HeNe laser beam. The incident radiation was slightly off-axis at the bimorph mirror—this factor was not important because the detector used was a simple pinhole detector. The pinhole detector was used because of its inherent simplicity, and the facility with which it can produce high-speed measurements. The bimorph mirror was connected to the high-voltage amplifier, which was controlled by a PC. Another PC, equiped with a multi-channel analogue data acquisition card, was used to measure the pinhole detector signal, the actual voltage applied and also a reference sine wave of known frequency. The reference sine wave was required to provide an accurate time reference because the data acquisition card's on-board clock chip was somewhat unreliable. The frequency of the reference sine wave was measured with a multimeter which had a dedicated frequency-measurement feature. As a separate time measurement, the time difference as calculated by the system clock was used, as this was found to give a reasonably accurate indication of the time interval between measurement points, as compared to the reference sine wave. The time measurement given by the system clock is accurate to one thirteenth of a second, leading to an uncertainty of about 4% over a 2 second run.

This method had the disadvantage that the mirror had to be driven in a defocus mode to obtain a signal that was distinguishable from the noise. This implies that large signals need to be sent to the

Figure 5.4 Measured pinhole-detector for a single electrode bimorph mirror to a 24 volt step



whole mirror in order to get a good signal-to-noise ratio. Sending large signals to a one-electrode mirror is problematic: the capacitance of a 45mm diameter device with a 0.35mm thick disc of piezoceramic is around 100nF. Therefore, a very large supply current is required to charge up this capacitance quickly. Unfortunately, the high voltage amplifier used in the experiments was incapable of supplying enough current to give a true test of the mirror step response to a large step input. Nonetheless, even with the high noise levels, it is possible to see from the optical experiment that the mirror responds extremely quickly to low-voltage step impulses. A response to a 24 volt (20 amplifier units) step is shown in figure 5.4. The time taken to measure the data run of 10000 points was measured to be 1.64 ± 0.04 seconds, so the sample frequency was 6100 ± 150Hz. It is clear from the graph that the mirror responds to the voltage command within one sample time, although there is some considerable lower frequency oscillation. This oscillation was estimated using a Fourier transform to be around 400Hz. This frequency is considerably lower than the anticipated first resonance of the device, which can be calculated to be 4861 Hz. Therefore, given the instantaneous response, it is likely that the low frequency oscillation is caused by the inadequately secure mounting of the device.

5.3.2 Non-optical measurement

Given the problems in the above optical measurement, an attempt was made to measure the resonance frequency in a purely electronic way. The method, described by Koryabin [76], is simply this: a signal is applied to one electrode of a multi-electrode device, and the induced signal is measured on a different electrode. This voltage is small—in the experiments carried out, the induced voltage was around $1/200$ the magnitude of the excitation. However, it is a clean signal, and the influence of mounting method is neither expected nor observed. Also, in exciting only a small region of the mirror, the capacitance that must be driven is correspondingly much smaller, and so the problems of

Table 5.6 Measured and calculated resonant frequencies for various bimorph mirrors

Piezo thickness (mm)	Glass thickness (mm)	Measured Frequency (Hz)	Calculated Frequency $n = 1$ (Hz)	$n = 2$ (Hz)
0.35	3.45	12500 ± 1000	9000	16000
0.35	1.8	8100 ± 200	5600	9700
0.2	0.16	2950 ± 100	730	1260

the power limitation of the high-voltage amplifier are not present.

A typical response to a square wave is shown in figure 5.5. The excitation was a square wave of amplitude 120V (100 amplifier units). This figure clearly shows the "AC-coupled" nature of the response. This is not unexpected—the voltage induced in the capacitative piezo ceramic simply leaks away. Also, it is clear from this figure that the mirror responds more quickly than can be measured by the data acquistion PC. By applying a smaller signal, and using an oscilloscope, a resonance was found, the period of which was $80 \pm 5\mu$s. This corresponds to a resonant frequency of 12500 ± 1000Hz. The calculated resonant frequency can be found using equation 3.80 on page 63. The same procedure was carried out for mirrors with a variety of thicknesses, and the results are shown in table 5.6. These data show that the resonance found is at a frequency in between the calculated fundamental mode and the next mode. In addition, low amplitude sinusoidal voltages of varying frequencies from a signal generator were applied to an electrode, and the amplitude of the response measured. These curves are shown in figure 5.6. The peaks in these curves match well the frequencies found from the square-wave method described above. Note the different horizontal scales in each plot. These graphs show that the response to low frequencies is fairly flat, which is desirable. Above the first resonance, there are generally many subsidiary resonances. Therefore, it is desirable that any high-voltage bimorph amplifier should be bandlimited to ensure that these resonances are not excited.

5.4 Thermal Response

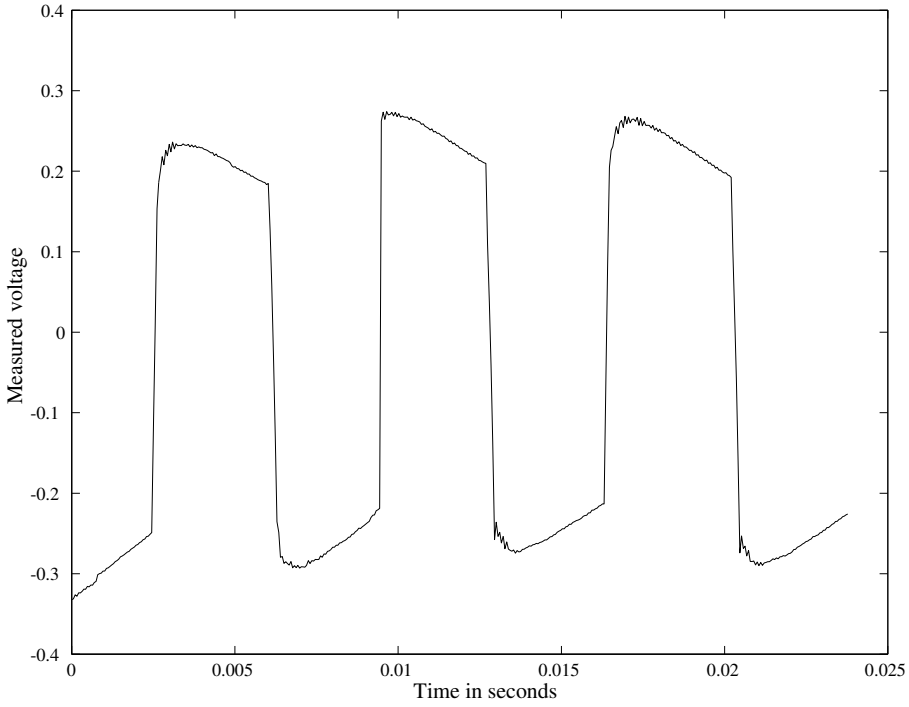
As mentioned in section 3.3.6, unless the thermal expansion coefficients of the layers in a laminate are exactly the same, there will be some bending as the laminate is removed from the temperature at which it was constructed. The bimorph mirrors constructed as part of this work were made from pyrex, as this has a thermal expansion coefficient which has the best match to that of piezo ceramic. As a test, mirrors with nominally identical parameters but using BK7 as the substrate material were constructed. To test the thermal stability of a device, the procedure used was as follows. The mirror was examined with a digital phase-shifting Fizeau interferometer, described below.

5.4.1 Digital phase-shifting interferometer

A digital phase-shifting interferometer takes several interferograms, with a phase shift added to the reference surface in between each interferogram. This phase shift moves the fringe pattern in such a way that it is possible to recover unambiguous wavefront information by considering all the fringe patterns together.

The commercial Möller-Wedel interferometer was connected to a controlling PC, with the necessary software to interpret the phase-shifted wavefronts. The accuracy of the wavefront information is (in this case) limited by the flatness of the reference surface, which is specified as better than $\lambda/20$ RMS deviation at $\lambda = 633$nm. This RMS deviation is for the full aperture of the reference surface

Figure 5.5 Measured induced signal when exciting a neighbouring electrode



(whose diameter is 10cm); the mirrors to be tested had a diameter of 42mm, and so we expect the RMS error in the surface measurement to be no worse than $\lambda/20$, and possibly better. In any case, all the wavefronts recorded were subtracted from a reference measurement, and so the combined effect of the non-flatness of the reference surface and the initial shape of the mirror are accounted for. Thus we are left with the inherent measurement error of the interferometer from sources such as the digitisation of the fringe patterns in the CCD camera. This error is estimated in the following section.

5.4.2 Measurement stability

In order to estimate the approximate errors in the fitted Zernike coefficients, 20 frames from each type (pyrex and BK7) of mirror were taken, and their Zernike coefficients estimated. The standard deviations of the estimated coefficients are plotted in figure 5.7. This clearly shows that the defocus coefficient is the most variable—this is unlikely to be measurement error, rather it is almost certainly due to slight temperature changes in between the measurements. This can be stated since it was seen that while all of the other coefficients vary in a more random fashion, the defocus coefficient varies in nearly monotonic way, as shown in figure 5.8. Also plotted in this figure is the "Strehl ratio" of the fit— the correlation between the two is marked. The mean Strehl ratios of the fitted coefficients are shown in table 5.7. This shows that, apart from the thermally influenced focus coefficient, the other coefficients are fitted with very good accuracy.

5.4.3 Experimental procedure

To measure their temperature, the mirrors had a K-type thermocouple attached. This enabled a good estimate of the temperature of the mirror to be obtained at the instant that the interferogram was recorded; the electronic thermometer readout gave one decimal place. However, the real uncertainty

Figure 5.6 Amplitude of response as a function of frequency when exciting a neighbouring electrode

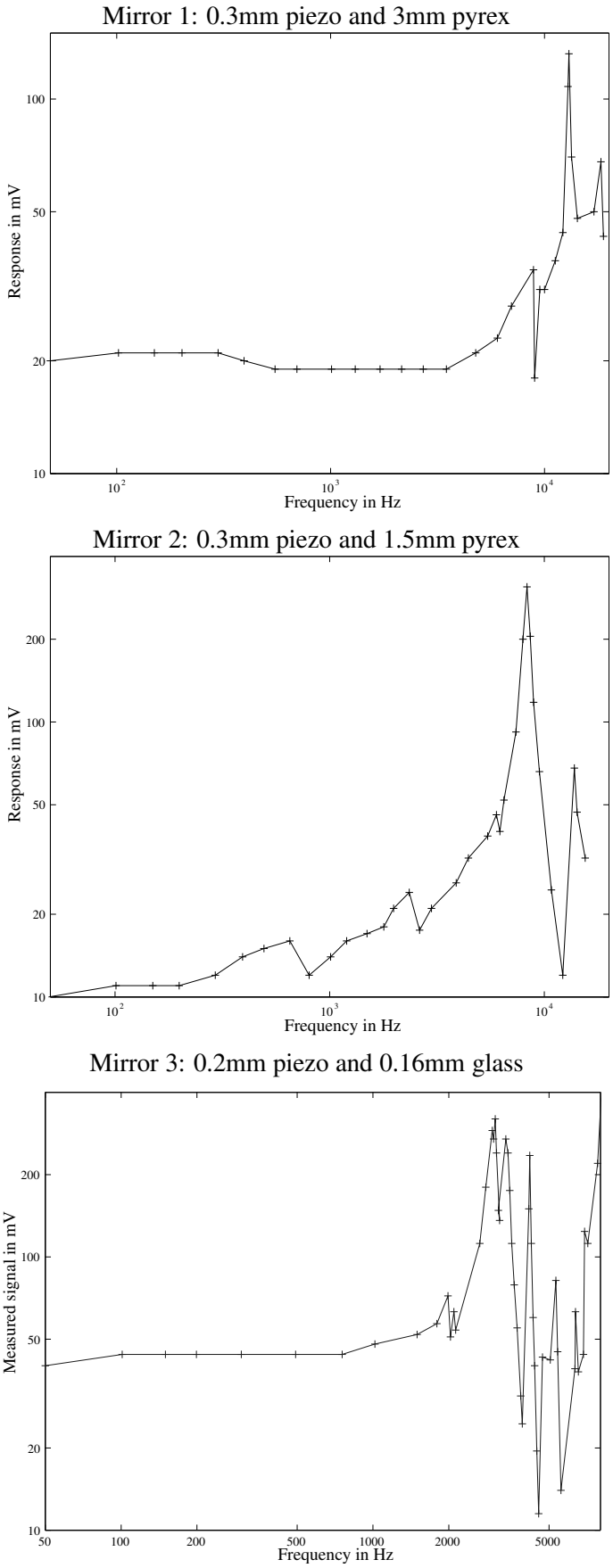


Figure 5.7 Standard deviations of Zernike coefficients for 20 frames

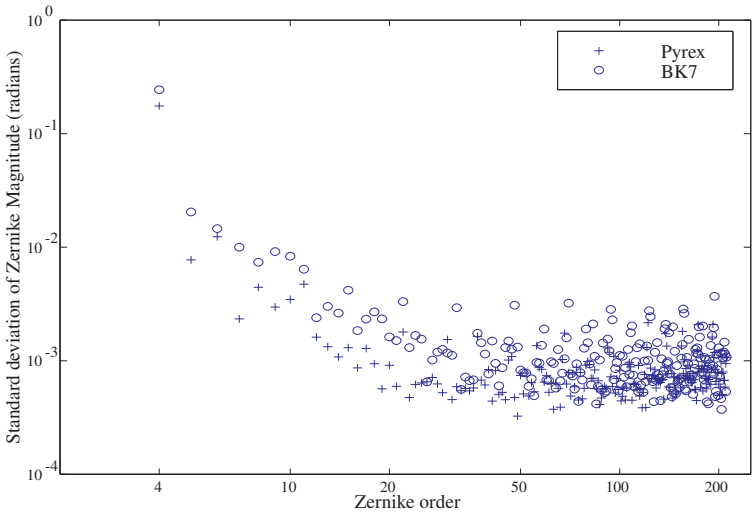


Figure 5.8 Strehl Ratio and amount of defocus for BK7 mirror

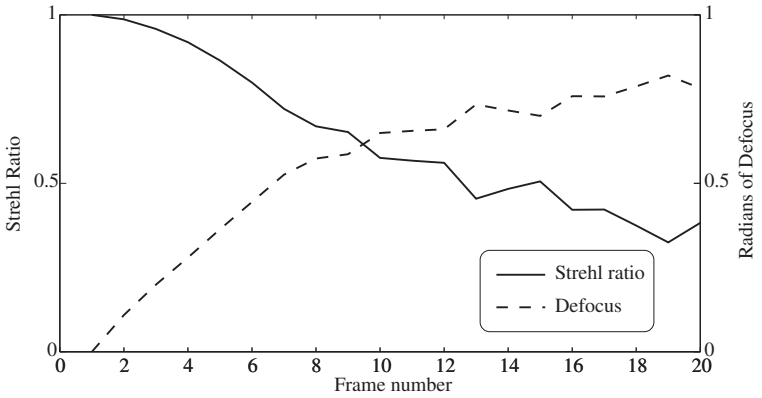


Table 5.7 Mean Strehl ratios of fitted Zernike coefficients

Mirror Substrate	Overall Mean Strehl Ratio	Focus-removed Strehl Ratio	Focus-removed RMS wavefront error
BK7	0.63	0.997	$\lambda/110$
Pyrex	0.96	0.999	$\lambda/210$

in the temperature of the whole device must be larger than this since the temperature was only measured on one surface—the error is of the order of $\pm 0.1\,^{\circ}\mathrm{C}$.

Once the mirror under test was aligned in the interferometer, it was heated using a hand-held hot air gun. Then, as it was cooling, interferograms were taken as quickly as possible, and the temperature at the instant the frames were recorded was noted[1]. For each mirror, several cooling cycles were recorded. These data files were analysed as follows: firstly, the appropriate data scaling was extracted using the technique describe in appendix B. Then, using the technique described in appendix A, the surface data were converted to vectors of 211 Zernike coefficients.

5.4.4 Results and comparison with theory

The behaviour of the Zernike modes with temperature were examined. As expected, order 4 (defocus) was the only one with any substantial variation. The variation of defocus with temperature for the three mirrors tested is shown in figures 5.9 and 5.10. The Zernike magnitude has been rescaled to give the vertical scaling of *microns of surface deflection*. In these graphs, the data are split into individual cooling curves. It is clear from these data that there is some movement of the the curves between the runs—this is most likely to be due to some slipping of the adhesive, and highlights how important it is not to expose bimorph mirrors to high temperatures. However, each run is well fitted by a straight line. The mean gradient, and standard deviation of each gradient is shown in table 5.8, along with the mirror parameters and the values predicted by theory. The thermal expansion coefficients for the glasses are taken from references 73 and 77. The thermal expansion coefficient for piezoceramic was taken to be 2×10^{-6}; this is only an approximate value since, in fact, the thermal expansion coefficient for most piezoceramics varies with time, temperature, and poling.

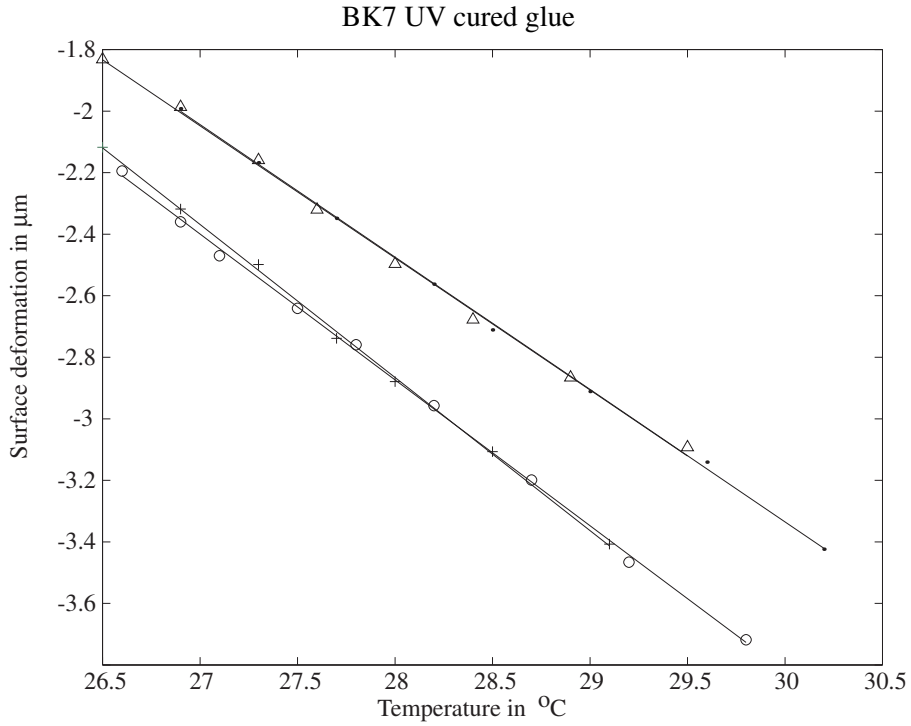
Table 5.8 Mirror parameters, and measured and calculated thermal deformation coefficients

Substrate material	Substrate thickness mm	Adhesive type	Thermal coefficient $\times 10^6$	Poisson's ratio	Measured deformation ($\mu\mathrm{m}\,^{\circ}\mathrm{C}^{-1}$)	Calculated deformation ($\mu\mathrm{m}\,^{\circ}\mathrm{C}^{-1}$)
BK7	1.8	Non-UV	8	0.25	-0.56 ± 0.05	-0.58
BK7	1.8	UV	8	0.25	-0.46 ± 0.03	-0.58
Pyrex	2.0	UV	3.3	0.20	-0.12 ± 0.01	-0.10

The closeness of the experimentally measured and predicted thermal deformations in table 5.8 is remarkable, and possibly coincidental. The agreement means that the effect of the adhesive layer on the deformation of the device is negligable, since the adhesive properties are not considered in the predicted value. However, it must be recalled that some of the parameters (especially the thermal coefficient of the piezo ceramic) are known only approximately, and any change in this parameter away from the assumed value will have a noticeable effect on the calculated values. It is clear from these results that the thermal performance of the pyrex mirror is superior to that of the BK7 mirror, as expected. Also, we can see that the non-UV curing glue performs slightly better than the UV curing glue, even though it is considerably more viscous and thus more difficult to spread into a thin layer.

[1] The frame acquisition takes around 2 seconds; the rest of the processing takes around 30 seconds

Figure 5.9 Thermal deformation of bimorph mirrors with different adhesives



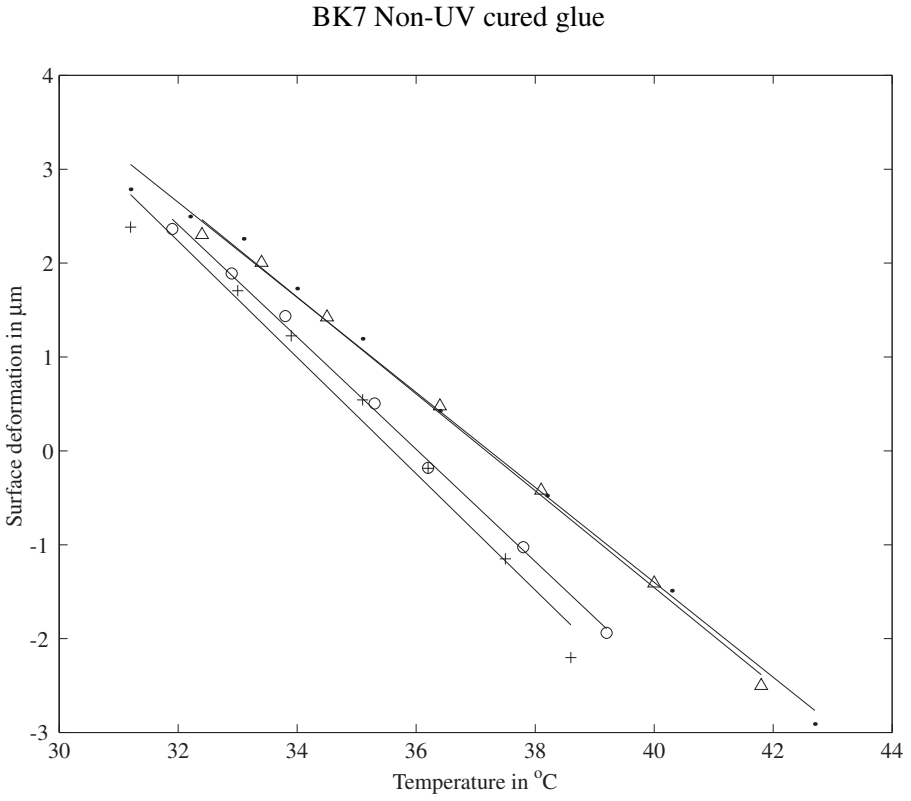
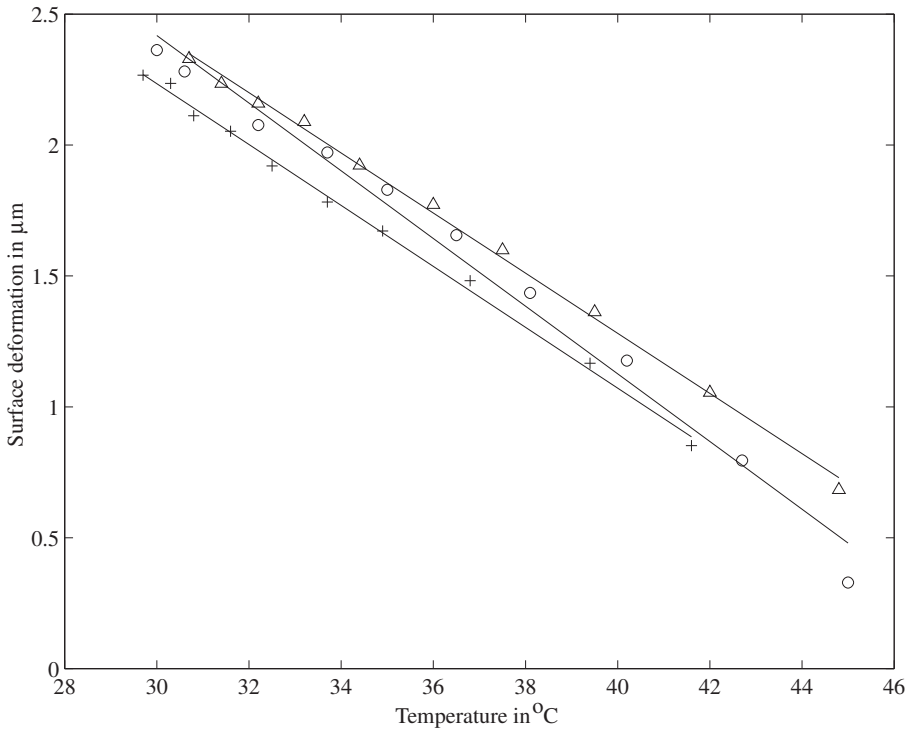


Figure 5.10 Thermal deformation of a bimorph mirror with pyrex substrate



5.5 Focal Sensitivity

The parameter that indicates how well the laminated structure of a bimorph mirror is performing is its focal sensitivity. This is analogous to measuring the thermal sensitivity as was carried out in the preceeding section. In order to measure the focal sensitivity, the bimorph mirror under test was once again mounted on the digital Fizeau interferomter, and interferograms recorded for a variety of voltage positions. It is well known that piezoceramics suffer from considerable hysteresis. In order to minimise this effect, a hysteresis compensation scheme suggested by Kudryashov and Shmalhausen [52]. In this scheme, a sinusoidally decaying voltage is applied to the mirror in an attempt to return it to a known position. Then, a voltage change is applied and the resulting surface deformation measured. Unfortunately, due to the very large numbers of fringes produced, it was not possible to test the mirrors over their whole range of deflections. Therefore, for each mirror, several voltages were chosen which gave as wide a range of deformations as were measureable by the interferometer. This range is determined by the maximum local slope of the wavefront; when measuring a mirror which has a changing focal length, it was not possible to measure surfaces which had a spherical shape which had a peak-to-valley surface deformation of around 3μm.

For each mirror, around 40 surface maps were taken. These were then processed in the usual way to extract Zernike information. Straight lines were fitted to each Zernike coefficient in turn; i.e. the result are fitted vectors **a** and **b** so that the overall surface deformation is described by:

$$\phi(r,\theta) = \sum_{i=1}^{211} (a_i + b_i V)\, Z_i(r,\theta) \tag{5.4}$$

In this part of the experiments, only single electrode mirrors were used, and so V is simply the voltage difference between the surfaces of the piezo disk.

5.5.1 Results: sensitivity

Once we have the vector of Zernike coefficient gradients, $\mathbf{b}$, we can calculate both the sensitivity of the device, and how accurately spherical the device remains when a uniform voltage is applied.

Considering first the sensitivity, we can convert the Zernike magnitude for the 4th order into a curvature using the scaling outlined in section A.6. The measured curvature sensitivities expressed in terms of curvature per volt, and also in deflection per kilovolt are given in table 5.9. The errors quoted are derived from the fitting error of the best fit line. The data points are shown along with the best fit line in figure 5.11. For comparison, data taken from the thinner pyrex mirror one year before those shown in figure 5.9 are plotted in figure 5.12. These older data were taken using the digital Fizeau's own Zernike coefficient outputs. These Zernike coefficients are somewhat unreliable since it is not possible to know precisely the circle over which they are being fitted. The offset and gradient are compared in table 5.10. The errors quoted on the parameters are simply the fitting errors—the uncertainty in the older data in reality are much larger than this due to the uncertainty in the radius over which the Zernike polynomials are fitted. The uncertainty in these parameters due to a radial uncertainty of 2% are shown in the table in parentheses. This shows that the sensitivity has become slightly worse in that year—unsurprising since the piezoelectric effect is known to decay gradually in time. Of more concern is the fact that the "off" position of the mirror has moved by 0.22μm, although this can be remedied with a signal of around 4 Volts.

Table 5.9 Focal sensitivity for three different mirrors

Substrate		Adhesive	Sensitivity		Calculated
Material	Thickness (mm)	Type	in μm/kV	in m^{-1}/kV	Sensitivity in m^{-1}/kV
BK7	1.8	Non-UV	61.2 ± 0.5	0.308 ± 0.003	0.26
Pyrex	2.0	UV	49.5 ± 0.3	0.249 ± 0.002	0.22
Pyrex	3.45	UV	13.4 ± 0.2	0.059 ± 0.001	0.091

Table 5.10 Comparison of data taken one year apart for the same mirror

Date	Focal offset (μm)	Focal sensitivity (μm/kV)
2/4/1998	$1.478 \pm 0.007 (\pm 0.06)$	$50.5 \pm 0.3 (\pm 2)$
7/4/1999	$1.70 \pm 0.01 (\pm 0.07)$	$49.5 \pm 0.3 (\pm 2)$

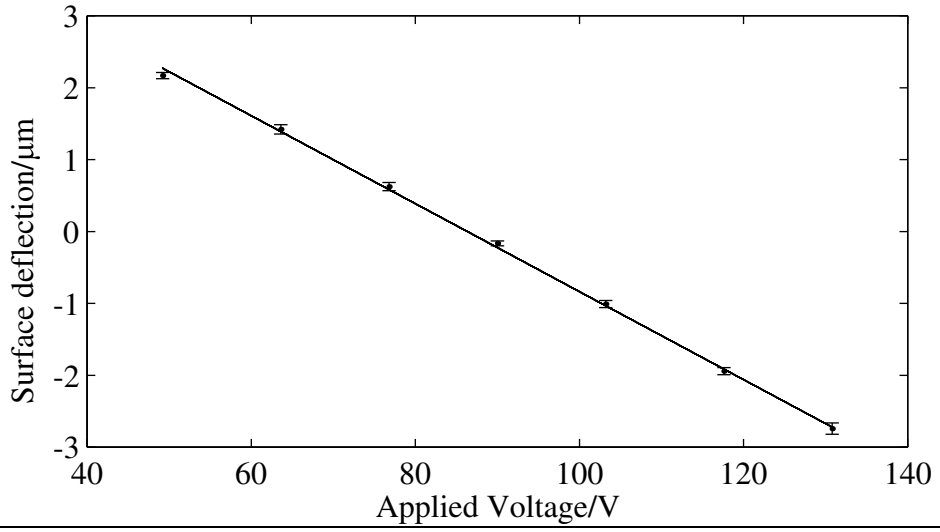
The apparent discrepancy between the predicted and measured values in table 5.9 requires some comment. Firstly, it should be stressed that many of the parameters required to perform the theoretical calculations are somewhat uncertain: the piezoelectric coefficient d_{31} is known to vary with time. Also, the radius over which the Zernike coefficients are evaluated has a considerable effect on the calculated curvature. The performance of the thicker of the pyrex mirrors considered in this section was repolarised after the piezoelectric effect had been lost due to overheating, and so the coefficient d_{31} must certainly be somewhat different to its original value. However, it is still somewhat surprising that the overall amount of curvature is some 20% greater than predicted by theory.

5.5.2 Results: sphericity

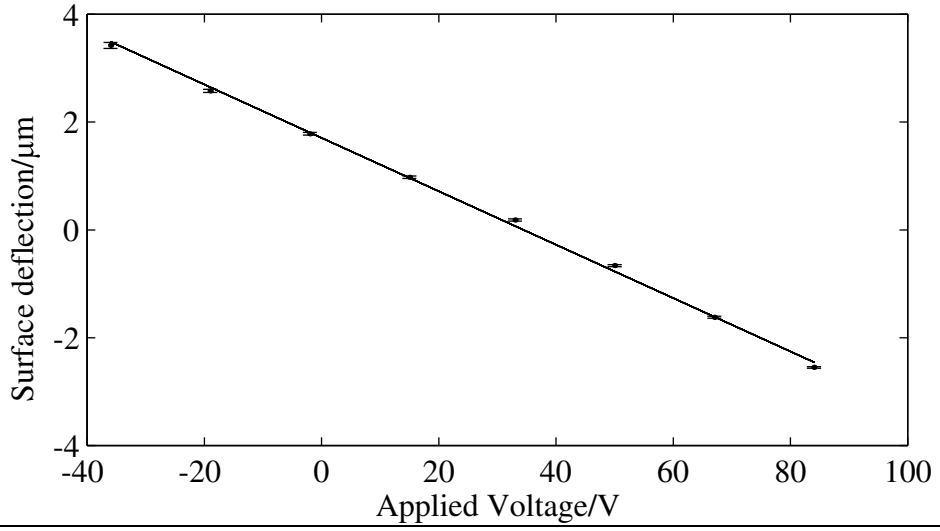
Using Zernike expansion gradients as found above (see equation 5.4), we can see how accurately the devices deform to a parabolic shape. The gradients for the first 40 Zernike terms for the pyrex

Figure 5.11 Focal deflection for three different mirrors

1.8mm BK7, Non-UV curing adhesive; sensitivity=61.2 $\pm$ 0.5μm/kV.



2.0mm Pyrex, UV-curing adhesive; sensitivity=49.5 $\pm$ 0.3μm/kV.



3.45mm Pyrex, UV-curing adhesive; sensitivity=13.4 $\pm$ 0.2μm/kV.

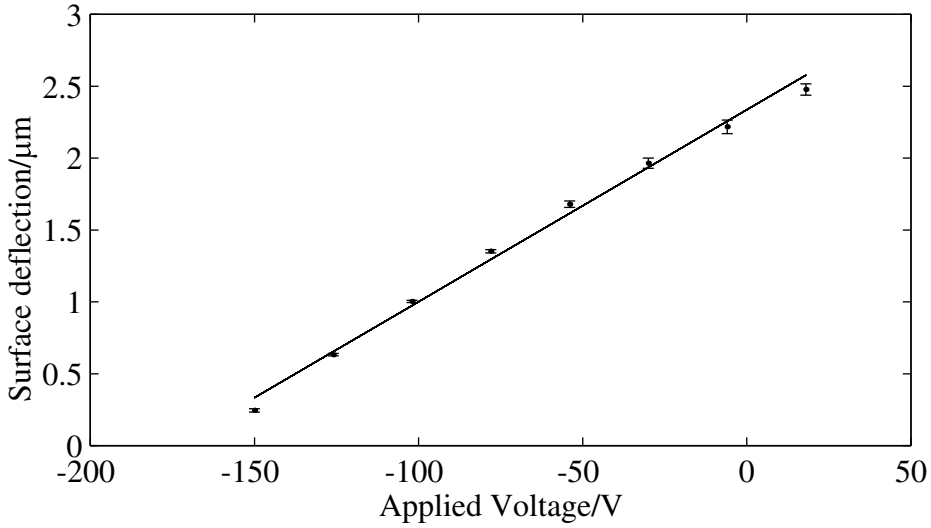
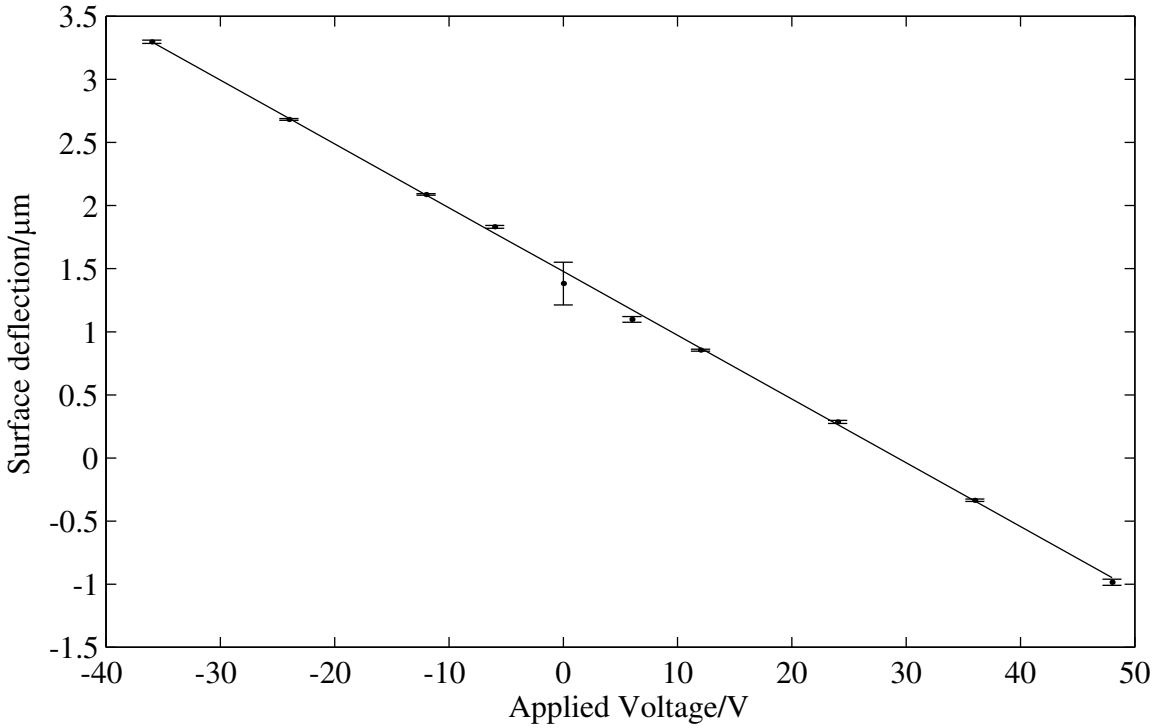


Figure 5.12 Data from the thinner pyrex mirror taken one year before that shown in figure 5.11



2.0mm device are shown in figure 5.13. The gradients are normalised so that $b_4 = 1$; tip and tilt have been ignored for the moment. It is clear from this figure that the magnitudes of all the orders other than defocus are very small indeed. This quality of fit can be expressed in a more quantitative way as follows: since the entire spectrum simply scales with applied voltage, we can find the maximum amount of defocus that can be applied whilst still retaining a Strehl ratio of greater than 0.8. If we scale the spectrum by a factor γ, then non-defocus residual variance, σ_r^2, is given by:

$$\sigma_r^2 = \sum_{i=5}^{\infty} \left| \gamma \frac{b_i}{b_4} \right|^2 = \gamma^2 \sum_{i=5}^{\infty} \left| \frac{b_i}{b_4} \right|^2 \tag{5.5}$$

The quantity $\sum_{i=5}^{\infty} |b_i/b_4|^2$ is a constant for a given mirror which we shall refer to as σ_0^2. So, the Strehl ratio is

$$S = 1 - \gamma^2 \sigma_0^2 \tag{5.6}$$

where σ_0^2 is expressed in radians. We can find the scale factor γ at which the Strehl ratio is reduced to 0.8 by requiring $\gamma^2\sigma_0^2 = 0.2$. Then, we can convert γ to a deflection in microns, or an optical power. To convert to microns, we must multiply by $\lambda/2\pi$. Then, the focal length is related to the edge deflection by $f = r^2/4h$ where r is the mirror radius, taken as 21mm, and h is the edge deflection. The results from performing these calculations are shown in table 5.11; where a wavelength is required, it is taken as 633nm. This shows that a one electrode mirror can reproduce a reasonable range of defocus with good accuracy. However, this does require that the initial figure of the mirror is flat (or contains only defocus). The considerably poorer performance of the thicker pyrex mirror is due to the fact that this mirror was constructed with 8 notches in the edge of the glass substrate to allow connection of the common electrode. This result is a graphic demonstration as to why this method is not desirable.

Figure 5.13 Relative Zernike magnitudes for a one electrode mirror

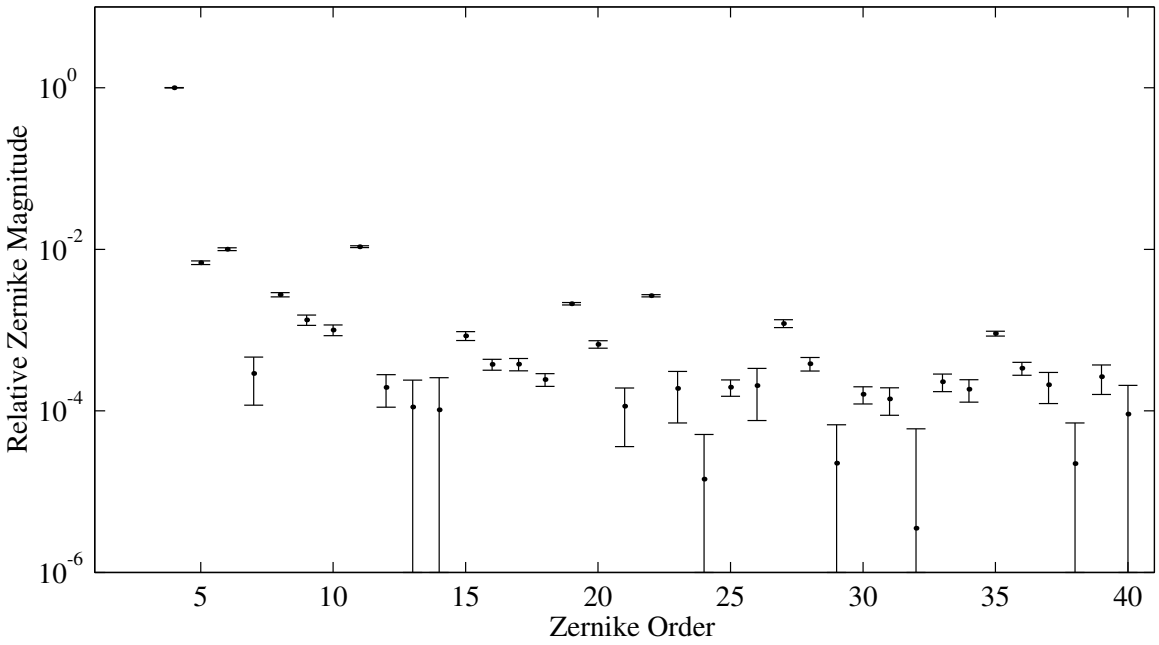


Table 5.11 Amount of defocus that can be applied accurately by a one electrode mirror

Mirror	$\sigma_0^2 \times 10^5$	$h_{\max}$ in λ	f in m
Pyrex 2mm	29	4.2	42
Pyrex 3.5mm	1800	0.5	100
BK7 1.8mm	50	3.2	55

5.6 Strain Gauge Hysteresis Compensation

As mentioned in the preceeding section, hysteresis is present in all piezoceramic materials. In most adaptive optics applications, it is necessary to have optical feedback in order to control the mirror to a desired state. Even when optical feedback is available, it is often slow, and requires much digital processing. Therefore, there is a need for a fast, non-optical method of driving a deformable mirror to a commanded position in between receiving new command instructions from the wavefront sensor. This has already been implemented on some segmented mirrors using strain gauges attached to the piston controllers [78]. This technique can be extended to a bimorph corrector. Using a strain gauge with a multi-electrode bimorph mirror is quite complex—when one electrode is actuated, strains will be induced in all other strain regions. Therefore, the signals from these strain gauge sensors are not diagonal, and would require some processing. However, in the simpler case of a single-electrode focus-only corrector, one strain gauge is sufficient since in the defocus mode of operation, a uniform strain is produced over the outside surfaces of the bimorph device. These strains can be predicted using equations 3.47 and 3.23.

5.6.1 Experimental procedure

In the experiments, a strain gauge was attached to the rear (piezoceramic) side of a single-electrode bimorph mirror. The strain gauge used was a standard RS strain gauge. This was connected to an RS strain gauge amplifier, with the gain set at 10000. Unfortunately, with such a high gain, the bandwidth of the amplifier was rather low at around 30Hz. It was hoped that the mirror could then be mounted in the digital Fizeau interferometer and an accurate measurement of the focal length made at the same time as the voltage and strain gauge measurements were made. This was not possible due to the limitations of the mounting of the strain gauge. It was found that the strain gauge reading drifted by a significant amount over a time scale of around 30 seconds. This precluded any meaningful static results from being taken.

By using the mirror in a pinhole-focal detector arrangement (see figure 5.4), measurements of applied voltage, strain gauge signal and focal position could all be made in a sufficiently short time that the strain gauge signal would not drift. The mirror used for this experiment was the Pyrex mirror, whose construction details may be found in table 5.8. An approximately sinusoidally varying voltage was applied to the mirror, and the three data streams recorded using the data capture equipment described in section 5.3.1. The frequency of the sine wave was around 2 seconds.

5.6.2 Results

Graphs of applied voltage against strain, applied voltage against focus signal and focus signal against strain are shown in figure 5.14. The strain was calculated using the standard equation for a quarter-bridge strain gauge:

$$\varepsilon = \frac{4V_{\text{out}}}{V_{\text{in}}k} \tag{5.7}$$

where V_{out} is the (unamplified) output voltage from the quarter bridge, and V_{in} is the bridge supply, and k is the "gauge factor", the value of which is 2.1 for the type of strain gauge used. The calculated induced strain is $0.37\mu\varepsilon\,\text{V}^{-1}$. The approximate gradient from the graph of voltage against strain gives a value of about $0.3\mu\varepsilon\,\text{V}^{-1}$. This graph clearly shows that the gradient is dependent upon the state of the mirror. The graphs also show that there is a unique relationship (i.e. almost no hysteresis) in the relationship between the strain gauge and the focus detector. Also evident is a large amount of noise in the strain gauge signal when extreme voltages are applied—the cause of this is uncertain, although

the signal is clearly visible through the noise. The amount of hysteresis between the various signals can be quantified using the following method: firstly, two points are located which correspond to a given signal's "off" position (i.e. the equilibrium position), and occur when the voltage is varying in a different direction. Then, given these two times $t_{i,1}$ and $t_{i,2}$ where i refers to the current comparison signal, the hysteresis on the signal $f_j(t)$ is calculated as:

$$\text{hysteresis} = \left| \frac{f_j(t_{i,2}) - f_j(t_{i,1})}{\max(f_j) - \min(f_j)} \right| \tag{5.8}$$

This enables us to build up a hysteresis matrix, which is given in table 5.12

Table 5.12 Hysteresis matrix for the strain gauge hysteresis compensation

	Focus detector	Strain gauge	Applied voltage
Focus detector	0%	3%	12%
Strain gauge	3%	0%	14%
Applied voltage	13%	16%	0%

5.6.3 Conclusions

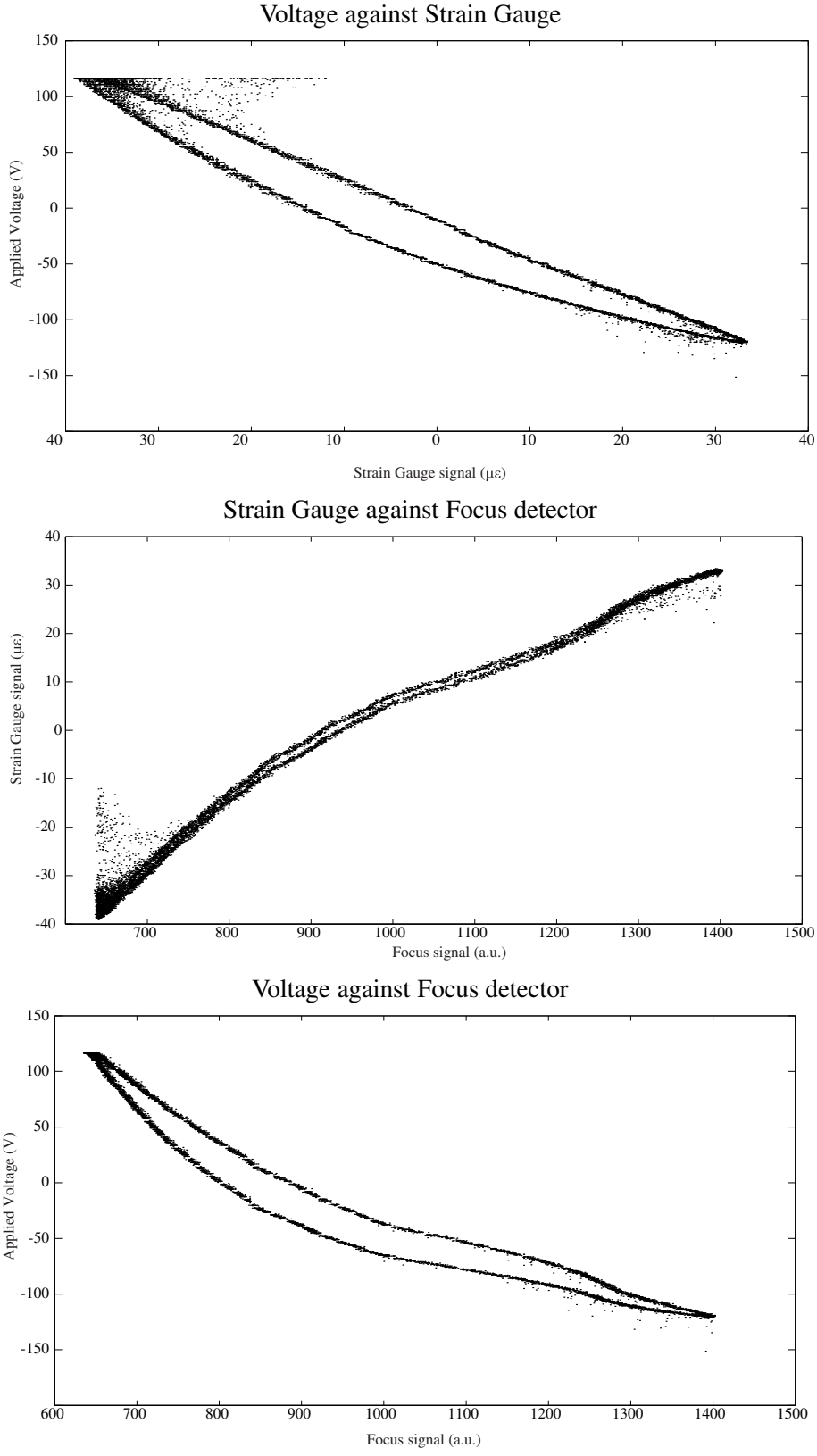
From the results presented above, it is clear that the use of strain gauges with bimorph mirrors is a viable means of reducing the effects of piezoceramic hysteresis. Although the experiments above only succeeded in proving the concept, it is anticipated that with correct handling of the extremely sensitive strain gauges, and construction of a tailored amplifier, it will be possible to obtain a static non-optical measurement of the focal length of a bimorph mirror. This would widen the possible field of applications of bimorph mirrors by enabling a focal position of the mirror to be "dialed in" with no optical feedback required. This type of operation would widen the potential field of applications of bimorph mirrors.

5.7 Multi-Electrode Bimorph Mirror

All of the mirrors in the preceeding sections have consisted of only one electrode. Although these defocus-only mirrors are undoubtedly useful, to perform more general adaptive correction, multi-electrode mirrors are required. By chemically etching tracks into the rear electrode of a piezo disc before gluing to a pyrex substrate, a 17 electrode mirror was made. This device was constructed before all the results from chapter 4 were available, so the electrode layout chosen was not optimised. Due to problems with the chemical etching, one of the electrodes was not functional, and two of the electrodes were joined together, so the mirror had only 15 functioning electrodes.

The experimental procedure used was essentially the same as that used in the preceeding sections where interferometric surface measurements were made. For each electrode, 5 different voltages were applied, ranging from -120V to 120V in steps of 50V. These raw surface measurements were then converted into vectors of 211 Zernike orders, whose uncertainties are assumed to be the same as those found in section 5.4.2.

Figure 5.14 Results from strain gauge hysteresis compensation



5.7.1 Measured response functions

Once the response functions for each electrode have been estimated in terms of Zernike coefficients, the average gradient and its error are calculated. Surfaces reconstructed from these average gradients are shown in figure 5.15. The "vertical" scale for the contour lines is the same for each response function; they are laid out so as approximately to resemble their position in the device.

5.7.2 Linearity of response functions

One of the most important properties a deformable mirror ought to possess is response function linearity. That is, the shape of wavefront deformation produced should remain constant as the signal to a given channel is varied; only the amplitude should change. We can see how well the mirror satisfies this condition by plotting various estimated Zernike parameters as a function of applied voltage. Figure 5.16 shows, for three typical electrodes, the several estimated Zernike coefficients as a function of applied voltage V. The Zernike coefficients shown are reference subtracted—that is, they have the $V = 0$ position subtracted. Due to hysteresis, it is necessary to have a different reference position for each electrode. Each column in this figure corresponds to a particular electrode: where this electrode occurs in the device is noted at the head of the column. Each row corresponds to a different Zernike coefficient. The graphs in each row are scaled so that the range of the ordinate is the same in each case (i.e. identical gradients will appear at the same angle, although there is a slightly different offset). The errorbars are those calculated in section 5.4.2. The text within each plot gives the gradient and its error for the fitted line. The units are Zernike coefficients multiplied by 10000, calculated in waves at $\lambda = 633$nm, with the Zernike expressions given in appendix A. The common names for the Zernike modes chosen are defocus, astigmatism, coma and spherical aberration. It is clear from these graphs firstly that the response for each channel is linear to a good approximation. Also, we note that the central electrode has very small non-azimuthally symmetric coefficients—as expected. In each case, the variation of the coefficient is seen to be well approximated by a linear fit. The error quoted on the gradient is calculated according to the prescription given by Topping [79].

5.7.3 Self-consistency of the response functions

A good indicator of how uniformly the device has been constructed is to observe how self-similar the response functions that ought to be identical (up to a rotation) are. In order to perform this comparison more easily, the "polar" Zernike expansion will be used—this is described in section A.3. In the polar expansion, it is the magnitude that is of most interest, since this should be the same for each electrode in a given ring. Figure 5.17 shows the mean polar Zernike expansion for the middle ring and outer ring of electrodes. The error bars represent the standard deviation for each coefficient. These graphs show that there is each electrode in a given ring behaves in almost the same way as each other. The central electrode's behaviour is not shown as there is nothing with which it can be compared.

5.7.4 Superposition of response functions

In order to perform another crucial check on the superposition of the response functions, the following simple test was carried out. The response functions of the various electrodes were measured by the interferometer. In this experiment, to simplify the procedure, the Zernike spectrum produced directly by the interferometer was used[2]. These Zernike spectra were used to build up a transfer matrix, which was then inverted. The "off" position Zernike spectrum was measured, and the voltages

[2]This was not used generally since it only supplies 36 terms of the expansion

Figure 5.15 The measured response functions of a multi-electrode bimorph mirror

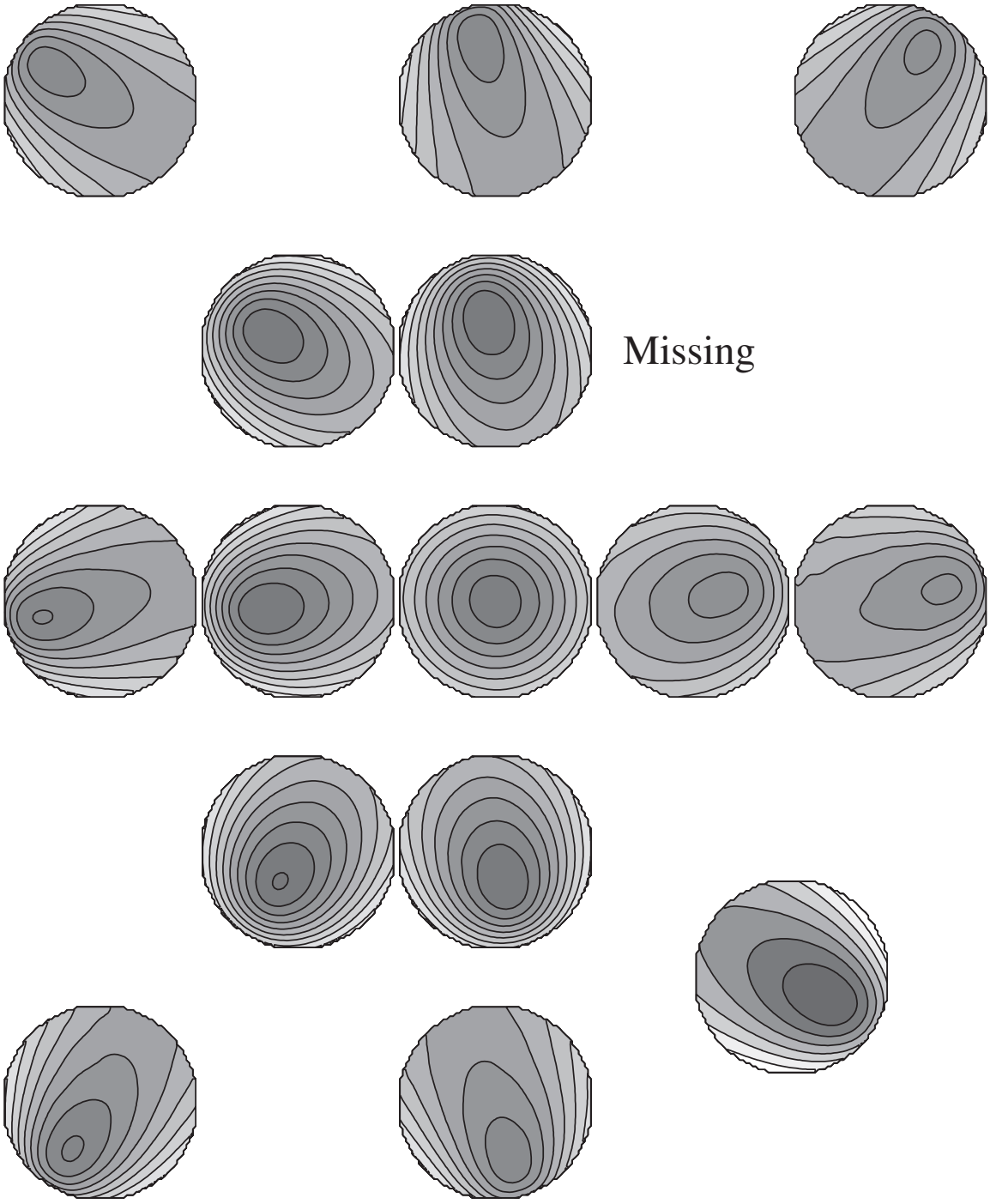


Figure 5.16 The behaviour of various Zernike coefficients as a function of applied voltage and electrode position. Zernike coefficients are in units of λ.

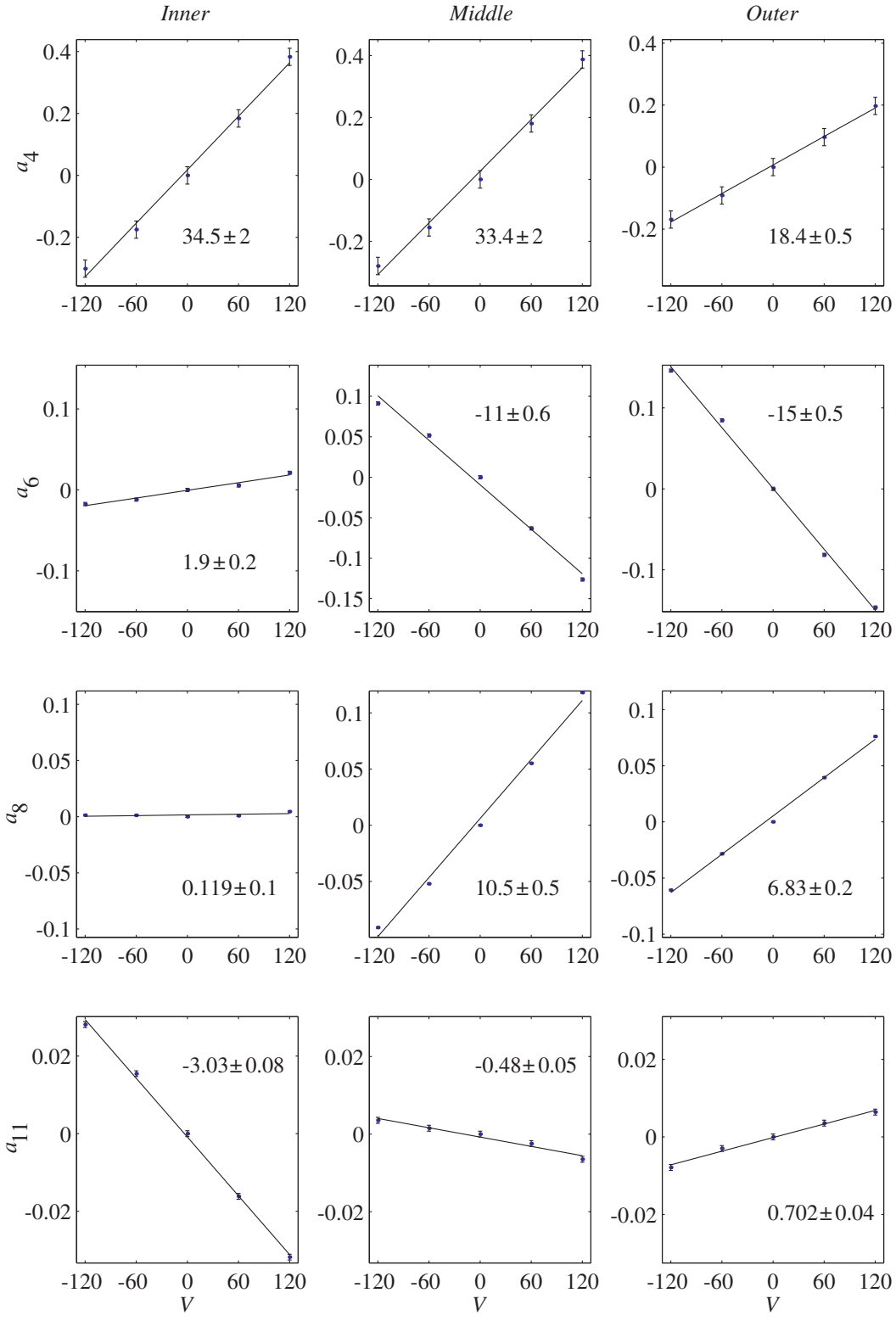


Figure 5.17 Polar Zernike coefficients for the middle ring and outer ring of electrodes (Units of λ)

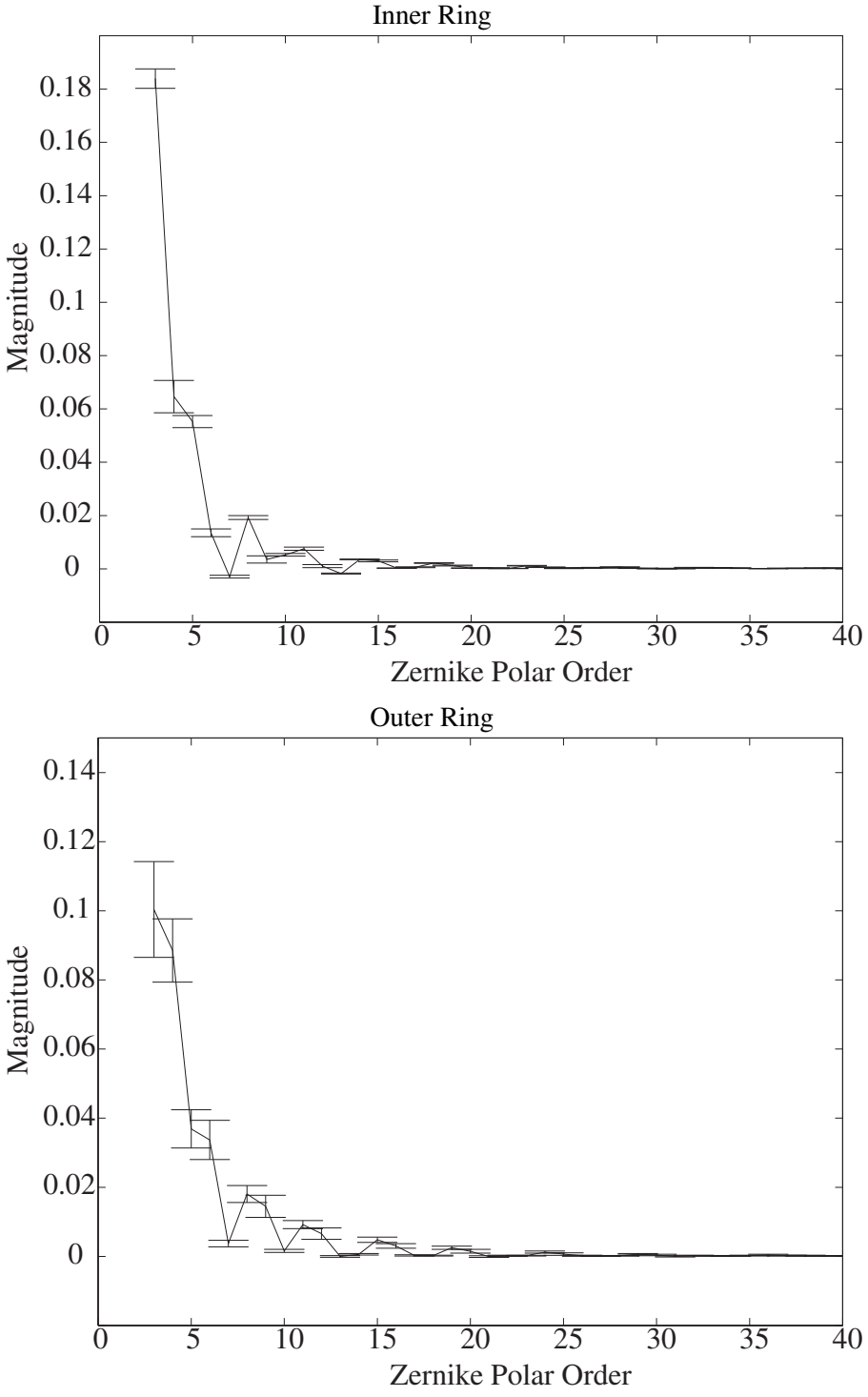
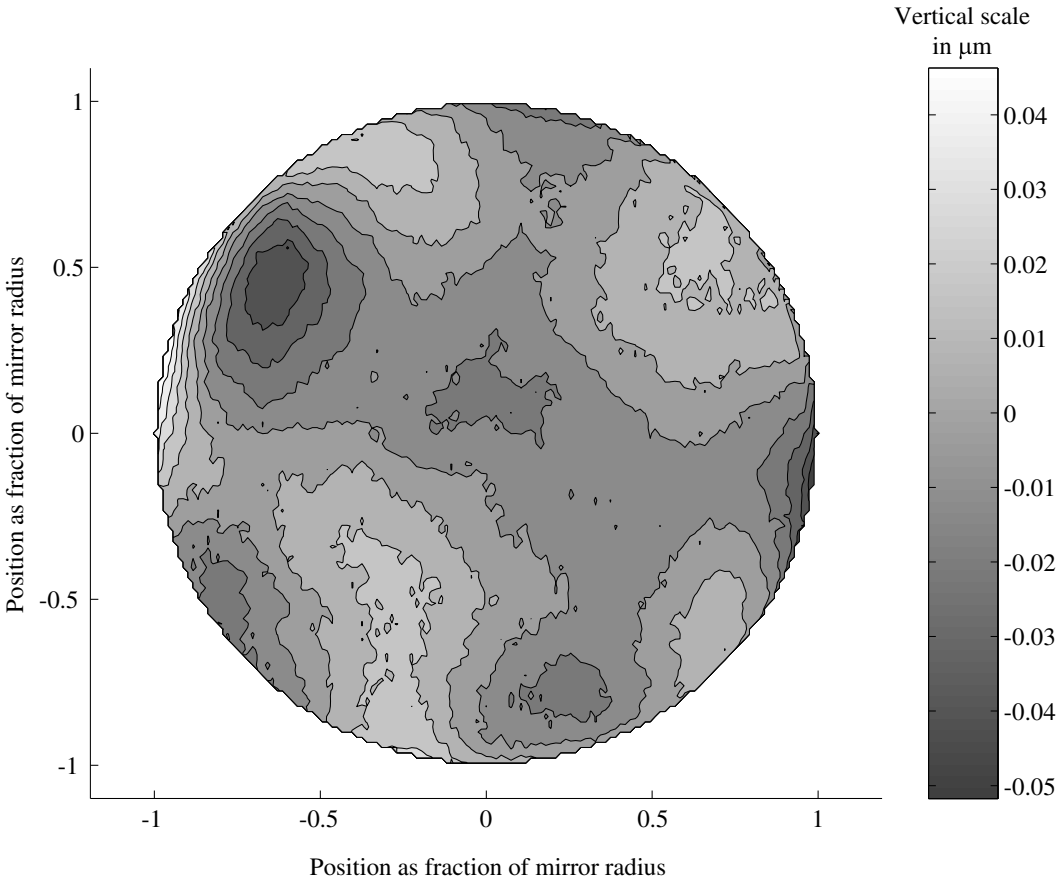


Figure 5.18 Contour plot of best flat mirror surface

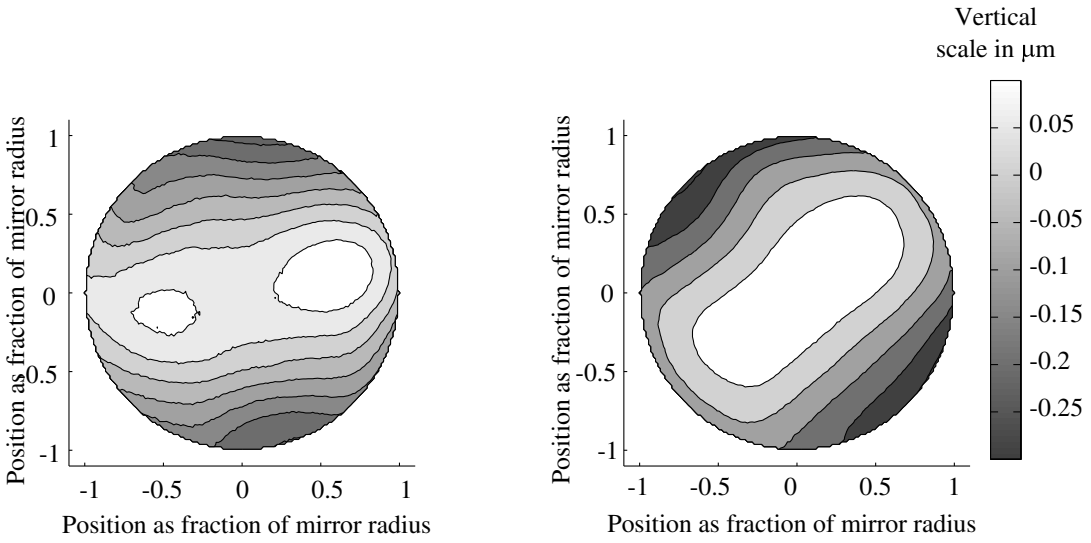


calculated to drive the mirror flat were applied. This procedure was iterated 3 times, until the shape given in figure 5.18. This figure shows a contour plot of the raw data file—no reference position is subtracted. The vertical scale is microns of wavefront deformation (i.e. twice the surface deformation). The RMS wavefront deformation is calculated to be 0.01μm, which corresponds to a Strehl ratio (at 633nm) of 0.98. This is very flat indeed compared to the "off" position of the mirror, which has a peak-to-valley deformation of about 2λ. The fact that such flatness can be achieved implies that the transfer matrix works very well, and therefore that the response functions do superpose in an additive way. As another demonstration of the principle, the mirror was commanded to take on two different cylindrical shapes. The results this are shown in figure 5.19. Again, the contour maps are not reference-subtracted. The accuracy of the fit is evaluated as follows: firstly, the shapes are re-scaled to have a variance of 1 radian^2. Then, the Strehl ratio is calculated for that surface relative to the commanded surface. For the left-hand shape, the resulting Strehl ratio is 0.94. For the right-hand shape, the resulting Strehl ratio is 0.95.

5.7.5 Comparison with theory

We can compare the modelled surfaces with the experimentally realised electrode response shapes. A quantitative way of doing this is to superimpose the modelled surface shapes on figure 5.17. The surface shapes were calculated in exactly the same way as described in chapter 4, and the polar Zernike coefficients were estimated. Figure 5.20 shows the superimposed polar Zernike spectra. In the case of the middle and outer electrodes, the plots are scaled to be the same at polar order 4, which corresponds to astigmatism. The central electrode ought to introduce no astigmatism, so in

Figure 5.19 Contour plot of cylindrical mirror surface



this case number 3, the focus term, was used. For the middle and outer electrodes, the error bars correspond to the variation between the electrodes in the same ring, for the central electrode, the errorbars show the error in the fitted gradient. For a less quantitative comparison, contour plots of the measured and predicted response functions are shown in figure 5.21. In these plots, the focus is set to zero (Zernike order 4) to show the residual shape more clearly.

It is also possible to carry out the same kind of analysis with the real response functions as was carried out with the simulated surfaces in chapter 4. In this case, the expected limit on D/r_0 was found to be 3.4 and a condition of 110, the simulated values were found to be 3.3 and 41 respectively. Thus, the correction achievable in principle is roughly the same as the simulation predicts. However, since the condition number is somewhat worse, the voltages that will be required will also be noticeably higher. In order to keep the voltages within the applicable limits, it is useful only to utilise the most prominent modes in the SVD decomposition for applying correction. When the number of modes was restricted to 12, the limit on D/r_0 dropped to 3.0, with a corresponding drop in the average RMS excursion from over 400 volts to 90 volts (the average RMS voltage for the simulated surfaces was 14 volts). This clearly shows that the loss of the higher spatial components in the mirror response functions does not cause too much harm to the amount of turbulence that can be corrected, but the voltages required to perform the correction are considerably higher.

It is clearly seen that the match between experiment and theory is not quantitatively correct, although the overall shapes are essentially as predicted. The main discrepancy is a progressive loss of the higher order Zernike terms. The most likely cause of this smoothing is the inadequate chemical etching of the laminated structure. The procedure used in the manufacture of this device was somewhat unsatisfactory, leading to substantially rounded corners. Also, even if the electrode shapes did posses sharp corners, the electric field set up in the piezo would be somewhat smoothed. This effect could not be incorporated into the calculated surface due to the limitations on the shapes of electrode that can be calculated. Other contributory factors include the thickness of the laminate, and shear deformation of the adhesive. The adhesive used for this mirror was the UV curing type, which was found (see section 5.4) to be a slightly worse adhesive than the non-UV type. The etching problem has since been solved, although results from practical devices made with the new etching method were not available at the time of writing.

Figure 5.20 Comparison of theory with experiment: Polar Zernike spectra in units of λ

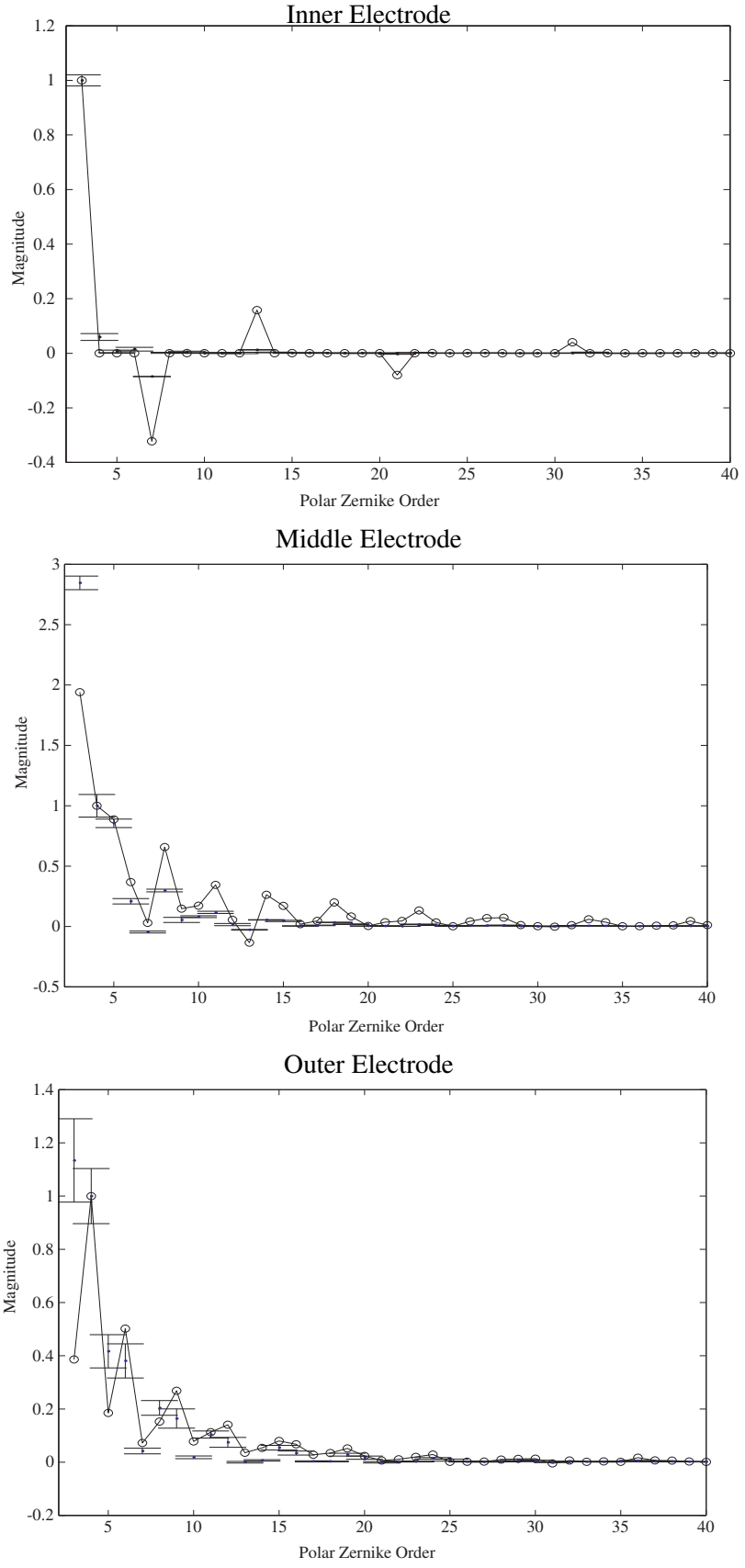
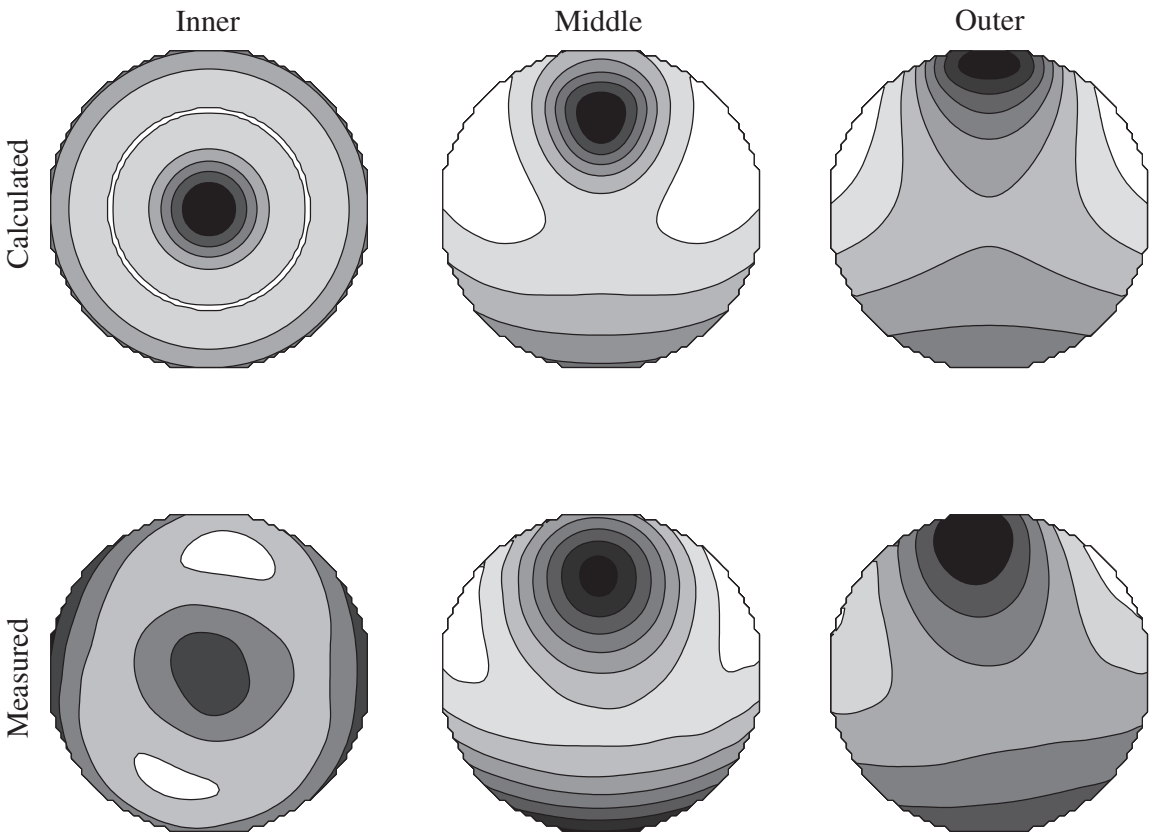


Figure 5.21 Comparison of theory with experiment: Focus removed response functions



6 Adaptive Laser Scanner

There are many traditional optical systems that have the potential to benefit from the introduction of adaptive optical components and techniques. The laser scanner was selected as being particularly appropriate for further investigation because it posseses several features that facilitate the application of a simple adaptive scheme, namely:

- No field of view — although a scanning system scans across some plane of interest, it does so in a time resolved manner, and hence the rather difficult problem of anisoplanatism is avoided.
- There is the possibility in a scanning system to precompute the optical aberrations, and hence the design of the adaptive component can be optimised more efficiently than if only statistical aberration information is available.
- Scanning optical systems are of considerable commercial interest; there are many situations in which recovering the full scan resolution, or simplifying the optical system would be beneficial.

6.1 Conventional Laser Scanners

Before we go on to discuss how adaptive optics can be used to improve laser scanning systems, we shall briefly discuss conventional scanning systems, covering nomenclature and applications.

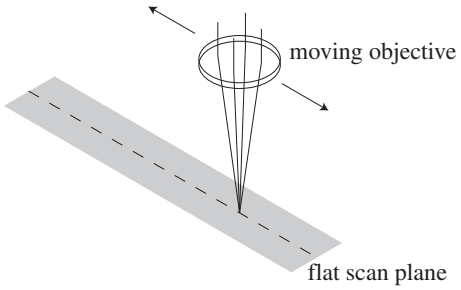
Stated briefly, a laser scanner is an optical system for inspecting or marking a workpiece. A point source is imaged onto the workpiece. In between the point source and the image is some means of deflecting the optical beam so that the image source translates over the workpiece. This translation may be in one or two dimensions—for many common applications, the workpiece is translated mechanically in a direction normal to the optical scan, e.g. the photosensitive drum in a laser printer.

6.1.1 Scanner classifications

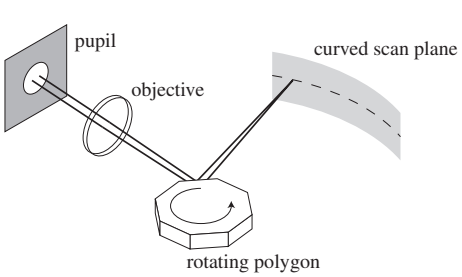
There are two main ways of classifying laser scanners: the way in which the beam is deflected, and in what part of the optical system the deflection takes place. For the latter classification, we use the notation introduced by Beiser [80], which describes three general locations of the scanning mechanism relative to the image-forming optical component:

- *Preobjective* scanners have the deflection mechanism located before the image-forming element. This configuration is characterised by a flat scan-plane, and a relatively large objective lens, which is often used in a roughly telecentric manner.
- *Objective* scanners comprise a moving head, the objective lens and its illumination are scanned over the workpiece.

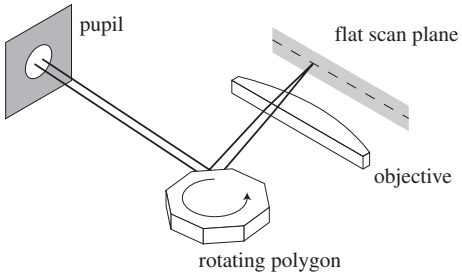
Figure 6.1 Three laser scanner configurations



(a) Objective



(b) Postobjective



(c) Preobjective

- *Postobjective* scanners deflect the beam after the objective lens. This configuration generally has a physically smaller objective lens than the preobjective scanner. The scan locus for this type of scanner is curved, the precise nature of the locus is determined by the deflection means.

These arrangements are depicted in figure 6.1. Pre- and postobjective scanners are the types most commonly encountered; the demands of the particular application will decide which type is used. The choice is usually dependent on whether or not a flat scan plane is required or not.

There are essentially five types of deflection means in common usage:

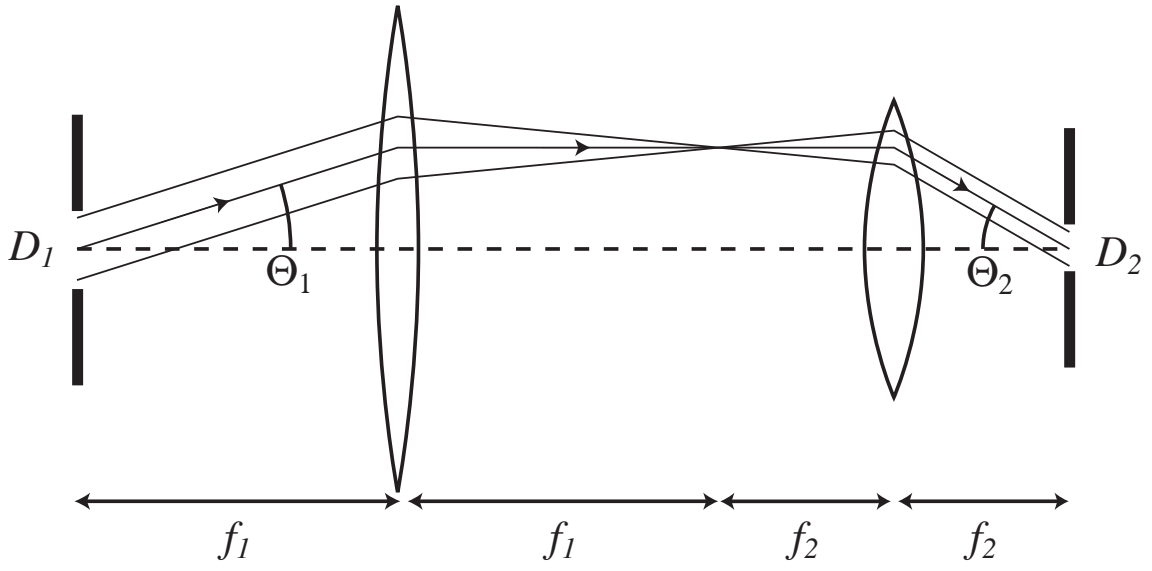
- *Translational:* in this case, an optical head is translated over the scan subject; this is an objective scanner.

- *Rotating Polygon:* a reflective, rotating element in the shape of a polygonal prism deflects the optical beam. These devices typically can support fast, wide scans and are economical in production

- *Acousto-optic deflector:* Bragg scattering from a diffraction grating set up by acoustic waves within an elastic medium causes deflection of an optical beam. These deflectors can be used in a "random access" mode—the deflection angle does not have to follow a periodic pattern as for a polygon. The range of scan angles possible is somewhat limited; the number of resolved points is limited to around 2000.

- *Galvanometer:* a reflective plate is suspended on an axis, and is deflected magnetically. Depending on the damping force, the rotation can be resonant. The scan angle for non-resonant galvanometers is limited, but they can be used in a random access mode. Resonant galvanometers have the property that their deflection angle varies sinusoidally in time.

Figure 6.2 Demonstration that the quantity ΘD is invariant



- *Holographic:* a rotating disc imprinted with a hologram deflects the beam. As with the polygon scanner, these devices can support wide scan angles.

The choice of deflection means is influenced by the overall number of scanned spots required, the speed at which these spots must be scanned, and whether or not random access is required.

6.2 Scanner Resolution

Several important relationships have been derived regarding the diffraction limited performance of a scanning optical system [80]. One of the most fundamentally important results is that it is the *number of resolved points* that is the crucial parameter. This turns out to be invariant under any combination of diffraction-limited optical components, and is in fact, a form of the Lagrange invariant. The invariant quantity is ΘD, where Θ is the full scan angle, and D is the beam diameter at the point of deflection. This invariance is readily seen for the case of telescopic demagnification, as shown in figure 6.2, following Beiser.

We can calculate how many resolved points there are in a scan line by considering how many times the resolvable spot size fits across the scan angle. Beiser expressed the number of resolved points as:

$$N = \frac{\Theta}{\Delta\Theta}, \tag{6.1}$$

where $\Delta\Theta$ is the resolvable angle, related to the spot size δ by

$$\Delta\Theta = \frac{\delta}{f}, \tag{6.2}$$

where f is the distance from the deflection point to the focal point. In the diffraction limited case, we can say that:

$$N = \frac{\Theta D}{a\lambda} \tag{6.3}$$

where a is correction factor depending upon the type of illumination. We can generalise equation 6.1 to the case where the spot size is variable over the scan:

$$N = \int_0^{\Theta} \frac{\mathrm{d}\theta}{\Delta\Theta(\theta)}, \tag{6.4}$$

where $\Delta\Theta(\theta)$ is now the spot size as a function of the current scan angle θ. We can immediately see that if the spot size is constant, then equation 6.4 reduces to the form derived by Beiser in equation 6.3.

We can now calculate how the number of scanned spots is affected by the offset of the scan medium from the scan plane. If we have an incident Gaussian beam, then we can calculate the relative change in spot size as a function of offset in terms of the confocal beam parameter (see section 2.2):

$$\delta = 2w(z) = 2w_0\sqrt{1+\left(\frac{z}{z_0}\right)^2}. \tag{6.5}$$

The amount of offset z is a function of scan angle, and we can express this as a series:

$$\frac{z(\theta)}{z_0} = \zeta(\theta) = \zeta_0 + \zeta_1\frac{\theta}{\Theta} + \zeta_2\frac{\theta^2}{\Theta^2} + \cdots \tag{6.6}$$

It will be more convenient to use the scaled angular parameter $\phi = \theta/\Theta$ for the following discussion. Now we can calculate the unaberrated number of resolved spots, when $\zeta(\phi) = 0$:

$$N_0 = \int_0^1 \frac{\Theta\,\mathrm{d}\phi}{2w_0/f} = \frac{\Theta f}{2w_0}. \tag{6.7}$$

If we define the resolution ratio $R = N/N_0$, then we can see that:

$$R = \frac{2w_0}{\Theta f}\int_0^1 \frac{\Theta\,\mathrm{d}\phi}{2w_0/f\sqrt{1+(\zeta(\phi))^2}} = \int_0^1 \frac{\mathrm{d}\phi}{\sqrt{1+(\zeta(\phi))^2}}. \tag{6.8}$$

In the case where we retain only the offset and tilt terms in the series expansion of ζ, then there is a simple analytical form for R:

$$R_{0,1} = \int_0^1 \frac{\mathrm{d}\phi}{\sqrt{1+(\zeta_0+\zeta_1\phi)^2}} = \frac{\sinh^{-1}(\zeta_0+\zeta_1) - \sinh^{-1}(\zeta_0)}{\zeta_1} \tag{6.9}$$

In the case where we retain terms up to order 2 in the series expansion of ζ, an analytical expression is again possible, but the complexity of the expression means that it is much quicker (in terms of computing time) to calculate the integral numerically using Mathematica[1]. If we retain more terms than 2, then the integration must be carried out numerically.

This expression for the number of resolved points can be used to calculate the relative number of resolved points for an uncorrected post-objective scanner compared the same scanner with correction. For a post-objective scanner, scanning a flat workpiece, the distance away from the scan plane can be expressed:

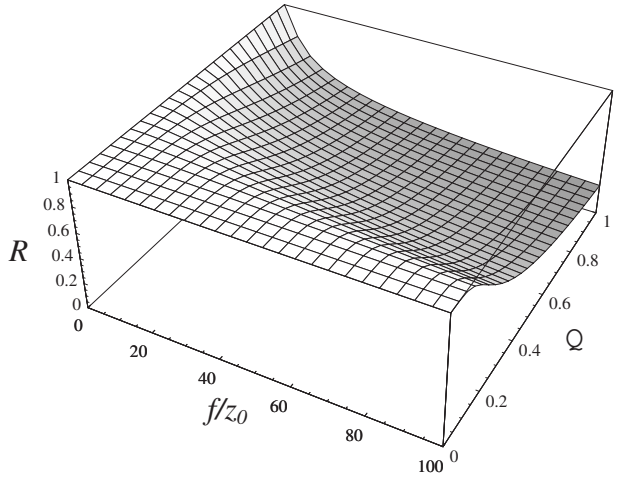
$$z(\theta) = f\left(1 - \frac{1}{\cos\theta}\right) \tag{6.10}$$

[1] Mathematica calculates numerical integrals using adaptive Gaussian quadrature with error estimation based on evaluation at Kronrod points [81]

where f is the radius of the scan.

Using equations 6.8 and 6.5, we can see that the two important parameters are the ratio of the scan radius to the confocal distance, f/z_0, and the angular extent of the scan Θ. A plot of the predicted resolution ratio is shown in figure 6.3. From this, we can see that if the scan radius is much larger than the confocal distance, as would normally be the case in a high resolution scanner, an uncorrected post-objective scanner performs very badly.

Figure 6.3 Resolution for an uncorrected post-objective scanner



6.3 AO Scanner Motivation

There are many current commercial applications of scanning optical systems, ranging from infrared imaging systems to supermarket barcode readers. The systems that are most likely to benefit from the introduction of adaptive optics techniques are those in which diffraction limited operation is desired, but not achieved. This failure may have many causes: it may be inherent in the optical design, it may be due to incorrect system alignment, the object to be scanned may not be in best plane or it may be curved. In each of these cases, adaptive optics may be able to offer a solution.

When correcting for aberrations inherent in an optical design, there is the potential advantage of being able to make the rest of the optical system to become less complicated, or less expensive to manufacture. In principle, a high resolution optical system may be made using only those components that are economical to manufacture (e.g. no aspheric components), with a single adaptive element that can correct for the residual aberrations. The limitations of this approach will depend largely upon the adaptive element used: for example, a bimorph mirror would be able correct for a relatively large amount of defocus (around 100 wavelengths of HeNe is possible with mirrors made as part of this PhD project), however the amount of spherical aberration that is correctable will be considerably less.

There are two distinct modes of operation that are possible when trying to correct for the mislocation of the object to be scanned. In the simplest case, the adaptive optical element would be used to refocus (and perhaps change other aberrations accordingly) the system as well as possible, without any correction depending on the scan angle being applied. This correction scheme would put considerably less taxing demands on the adaptive optical element—it need only change shape at the rate at which the scanned object changes, which would normally be slower than the rate at which the object is scanned. The more challenging correction scheme would be to correct for aberrations across the scan. This would enable a curved or tilted object to be scanned with the same resolution

that a flat object could be scanned, without any redesign of the optical system. Both of these modes have been demonstrated, and results from these experiments will be presented later in this chapter.

6.4 Wavefront Sensing in a Laser Scanner

Most adaptive optics systems require some form of wavefront sensor in order to function. This is certainly true for systems where the correction to be applied is unpredictable—such as in the case of astronomical imaging, where the aberrations evolve randomly and unpredictably. In the case of a laser scanner, however, the aberrations to be corrected may evolve in a predictable or even calculable way, and it is possible that, in the case of a low-hysterisis corrector, no wavefront sensor is required during operation. Even in this case, however, it is likely that some form of wavefront sensing would be desirable to evaluate precisely what correction is required to give the desired effect in the scan plane. For example, in the case where it is known that the potential movement of the scanned object is such that it has no detrimental effect upon the scan resolution (such as the situation in a laser printer, where the distance between the photosensitive drum and the scanning optics is essentially fixed), then the correction needs only be measured once. If the scanning element is a polygon, then it may be necessary to account for manufacturing errors in the polygon by measuring the correction once for each facet of the polygon.

Before we go on to discuss introducing a wavefront sensor into our system, it is helpful to clarify exactly what it is that we hope to measure. Our goal is to produce a series of diffraction limited spots in some scan locus. Therefore, for each point in the scan locus, we need to measure a difference in wavefront between that exists, and that which would produce a diffraction limited spot at that point.

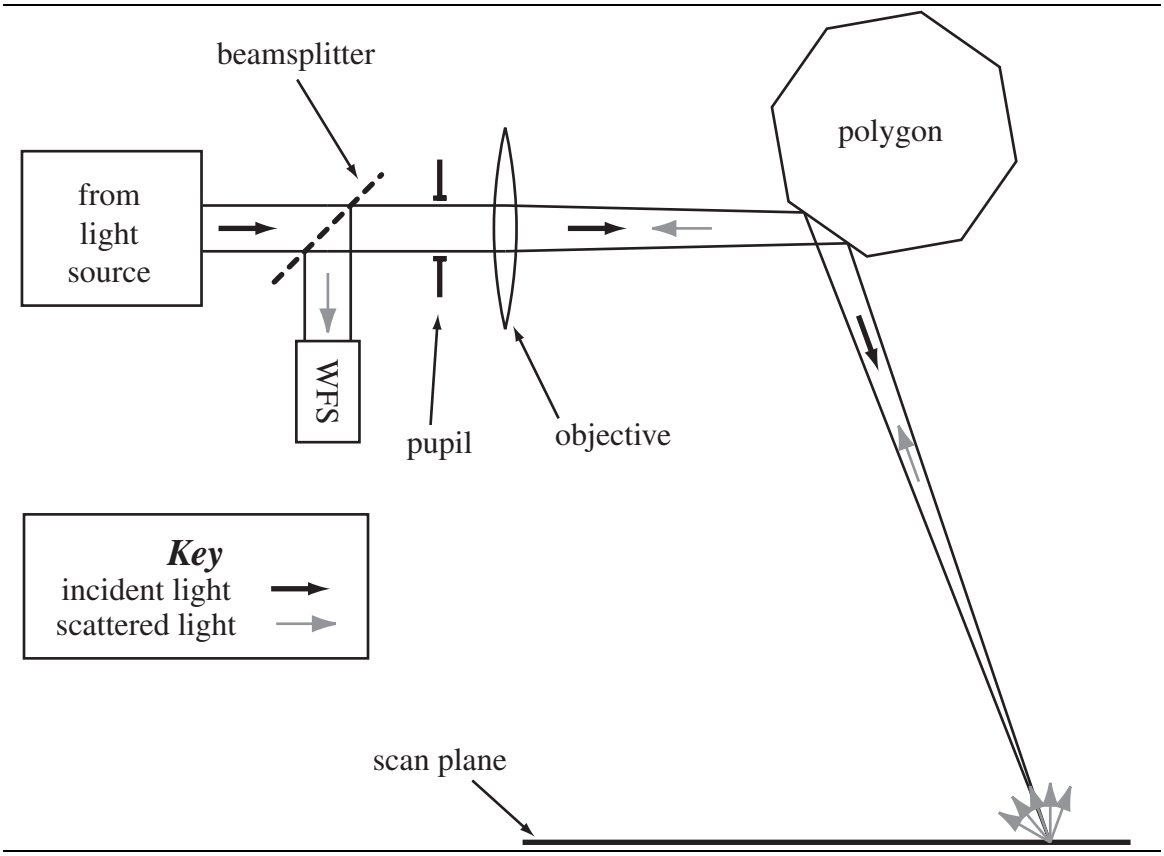
Integrating a wavefront sensor into a scanning optical system is not always a simple task. In the case of a scanner used in transmission, not only will the aberrations of the illuminating optics (i.e. that part of the optical system which produces the spots in the scan locus) vary with scan angle, but so will the aberrations of the collecting optics. This problem means that it is probably impossible to estimate which part of the accumulated aberrations are due to the illumination optics—this is the only aberration that we are trying to correct. A possible solution to this problem is to have a series of point sources in the scan plane which will give a callibration of the wavefront sensor incorporating the effects of the collection optics as a function of scan position.

Another option to try and derive a wavefront sensor signal is to use a portion of the retroreflected light from the scan plane. Providing that the reflection is completely incoherent with respect to the incoming light, then wavefront sensing can be carried out as depicted in figure 6.4. If the reflection is a coherent one, then this method will suffer from the well-known "double-pass" problem, where information relating to aberrations with odd symmetry is lost [60]. Light scattered from a rough surface in a direction other than the specular reflection direction can be used as this will act as an essentially incoherent source [10].

Once we have a reasonable wavefront sensor for our scanner, as proposed above, we must tackle the issue of a reference signal for our wavefront sensor. In many wavefront sensors, it is not possible *a priori* to know what sensor signal corresponds to ideal correction, i.e. they are not self-referenced. In the case of a Shack-Hartmann sensor, due to our imperfect knowledge of the precise location of the lenslets with respect to the spot-locator, we have no way of knowing what sensor signal corresponds to ideal correction. Therefore, in most adaptive optics systems, a well-corrected reference signal is required.

One possibility for a partial reference signal is to place a point source somewhere in the desired scan plane, and observe the wavefront sensor signal. This has the obvious disadvantage that we require the scan mirror to be in exactly the right orientation. If this is not the case, then we will

Figure 6.4 A wavefront sensor for a scanning optical system employing incoherent reflection in the scan plane



measure not only a large amount of tilt, but also incorrect aberrations, as the light reaching the pupil will have traversed the wrong part of the optical system. This problem is illustrated in figure 6.5.

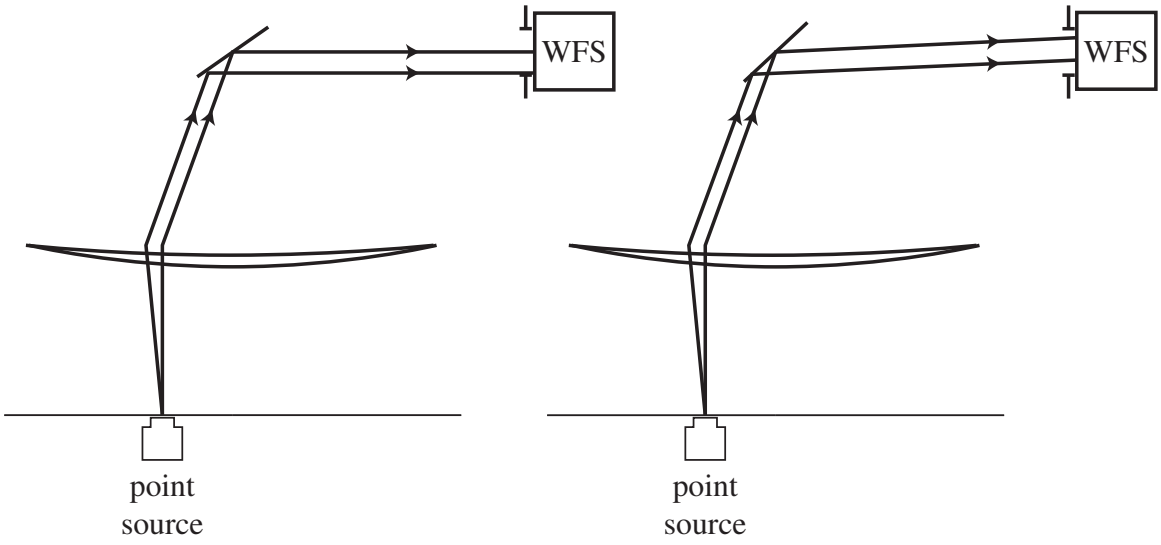
In the position on the left, the light from the point source illuminates the scan mirror in such a way as to illuminate the pupil with light that has essentially no tilt, although it will contain other aberrations. The scan mirror then rotates slightly in an anticlockwise direction to give us the situation on the right, where the pupil is illuminated differently. The rays illuminating the scan lens are the same as those shown on the left, and we can see that this aberration is sensed in a different part of the WFS pupil, and there is an additional tilt term.

Not all adaptive optics systems need a wavefront sensor signal to drive the correction. As outlined in section 2.3.3, resolution criteria may provide a suitable signal which determines whether correction has been achieved. An example of this approach was used in the practical scanner described below.

6.5 A Practical AO Laser Scanner Based on a Resolution Measurement

To demonstrate the principle of using adaptive optics to enhance the performance of a laser scanner, a simple practical system was constructed and evaluated. The intention in building this system was not to create a very high resolution scanner; rather the aim was to build a very simple system in which it would be easy to demonstrate correction across a scan in a variety of situations.

Figure 6.5 The reference source problem in a scanning system

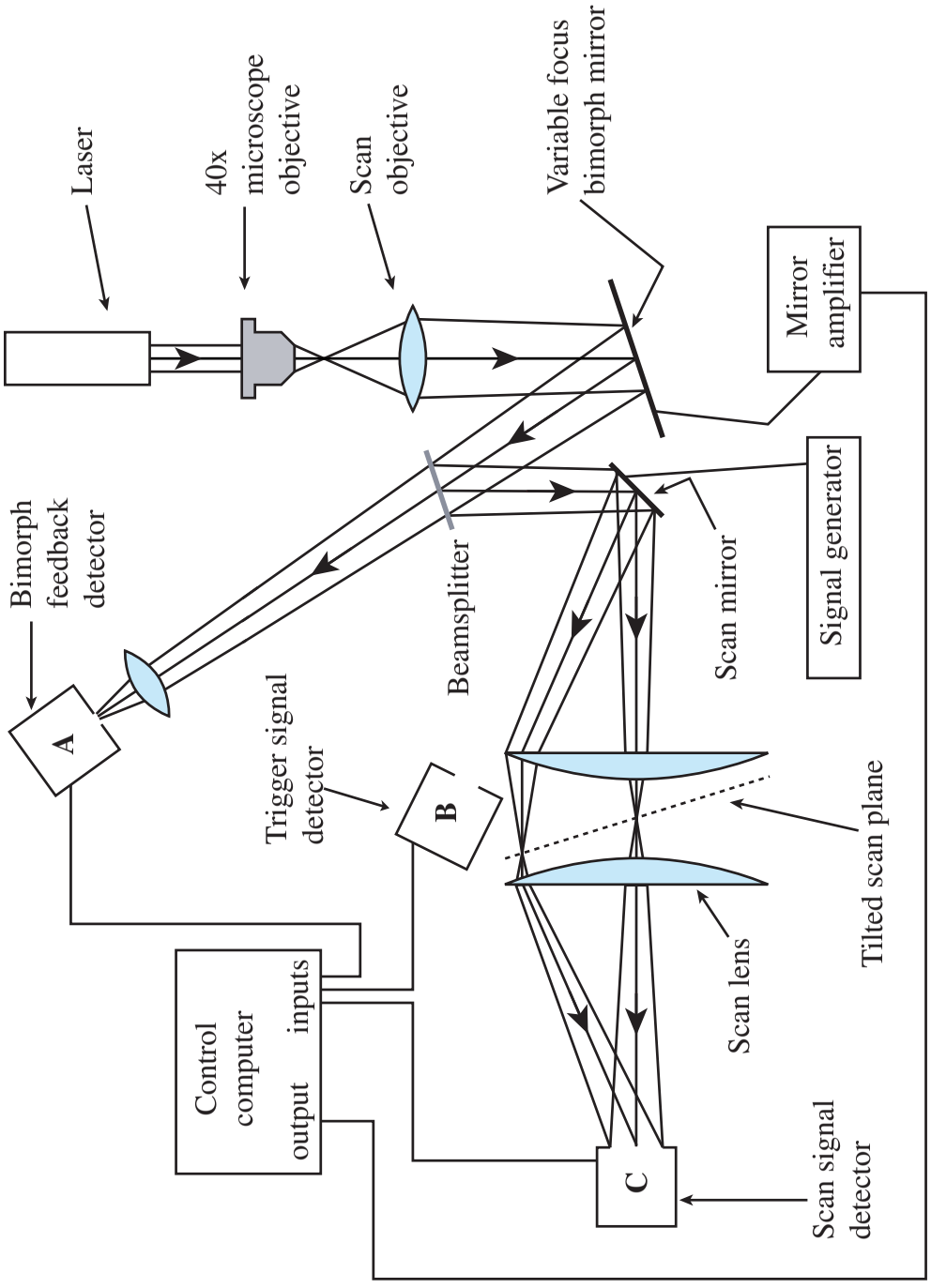


6.5.1 Optical layout

The basic form of the scanner is a pre-objective type scanner, to be used in transmission, to inspect some object in the scan plane. A diagram showing the relevant details of the system is shown in figure 6.6. As can be seen from the diagram, the laser beam is first approximately collimated using a microscope objective and a lens, then after the beam reflects from the bimorph mirror, it is incident upon a beamsplitter. The purpose of splitting off a part of the beam is so that it can be used as a feedback sensor on the mirror position—this simple sensor is described in section 2.5.2. The beam is then passes through a pupil to ensure that the scanning mirror, which is of the galvanometer type, is underilluminated. The scanned beam is then incident upon a large $f-\theta$ lens, which collimates the scan. The axis of the scan is coincident with the focal point of this lens. There is a second, identical lens placed some distance away, and this has the effect of bringing the light back to an image of the scan mirror. An additional lens was used to make the image of the scan mirror smaller so that it would fill the detector. Providing the system is well aligned, the image of the scan mirror does not move when the scan angle is varied. The beam converges to a focus somewhere in the region in between the two large lenses. It is in this region that the object to be scanned is placed.

There are three photodetectors employed in this arrangement. The first (A) is to sense the state of the bimorph mirror. The second (B) is a trigger sensor. At the beginning of each scan, the beam is deflected through a large enough angle to miss the first scan lens. A photodector is placed so that it detects this light. This signal serves as a trigger signal for timing the rest of the scan. The final detector (C) is the one that is conjugate to the scan mirror. In the absence of any scanned object, this picks up a uniform intensity for the duration of the scan. When an object is placed in the scan plane, its variation in transmissivity is recorded by this detector. In the case where a resolution target is used as a feedback mechanism, this detector provides the intensity information from which the resolution is derived. This principle of this method of sensing the focal length of a mirror is described in section 2.5.2, and practical matters concerning its operation are discussed in section 2.5.2.

Figure 6.6 Layout of a practial AO laser scanner



6.5.2 Operating parameters

The important operating parameters of the scanner are shown in table 6.1. Correctable focal range refers to the total change in position of the resolution target that could be compensated for with the adaptive mirror used. An iris was used to ensure that the scan mirror was underilluminated. All the data was read into a PC controlling computer using a 12 bit analogue to digital converter. The computer was also equiped with a 1MHz clock chip, which allowed accurate timing information to be gathered. The mirror could be controlled to 256 digital levels over a range of $\pm$150V. The focus feedback sensor A was arranged so that the noise in the system corresponded to a voltage change on the mirror of less than 4 levels .

Table 6.1 Operating parameters of the practical laser scanner

Parameter	Value
Scan lens focal length	30cm
Scan lens diameter	10cm
Scan mirror diameter	10mm
Scan lens separation	10cm
Correctable focal range	about 5cm
Resolution target	about 3 line pairs/mm
Scan frequency	up to 70 Hz
Adaptive mirror type	cooperative bimorph
Adaptive mirror diameter	50mm

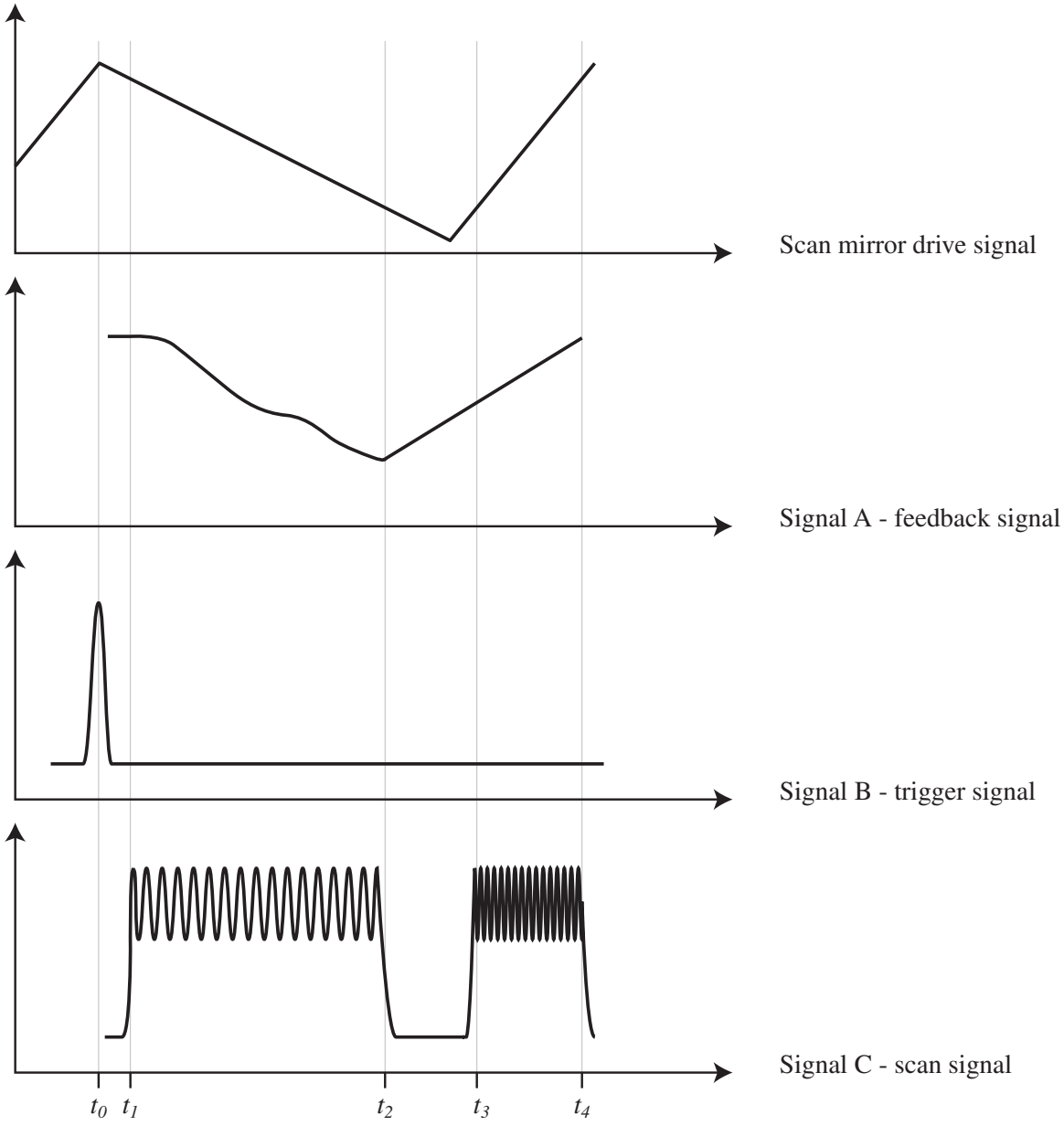
6.5.3 Scanner correction scheme

When the scanner is operational, there are several stages that must be completed before correction can start. The first thing that must be established is the relationship between the signal applied to the mirror and the mirror-feedback signal. This was carried out as follows: the mirror was driven to the maximum and then the minimum voltages, and the signals noted at each point. This was carried out 3 times, and an average gradient calculated. This assumes that the relationship between the applied voltage and the signal is linear. Even though this is not the case, we can still use this technique as the signal will always be driven in the correct direction (providing that the sensor is set up correctly), but it will be driven by the wrong amount. This means that, in general, it will take more steps than necessary to reach our goal—our effective correction bandwidth is reduced.

Once the mirror-sensor relationship has been characterised, the various timing parameters of the scan are measured. That is, the amount of time between the end of the trigger signal and the start of the scan signal is measured; and the duration of the scan is measured. The start and duration of the scan are derived from signals from photodetector C. The drive signal sent to the scan mirror is an asymmetric sawtooth wave, only the slowly varying part of the scan is used in this experiment. The forms of these various signals are represented in figure 6.7.

The active part of the scan is now divided into a number of discrete regions in time. These regions define parts of the scan over which a given correction will be applied. The next stage of the process involves taking a series of scans with a different signal applied to the deformable mirror for each scan. For each of these scans, the mean and the mean squared signal level (measured by detector C) are calculated for each of the regions in time. Also, for each of these scans, the signal at the feedback detector A is noted. For each region in time, the feedback signal measured at A which gives the largest ratio of mean squared signal to mean signal at C is selected. This is then the signal

Figure 6.7 The forms of the photodetector signals in the AO scanner



that A will be driven towards during the appropriate portion of the scan. This process is depicted in figure 6.8.

When the scan is underway, the control system used is a simple integrator. That is, the change in voltage applied is proportional to the difference between the measured signal at A and the desired signal at A. The gain used in the experiments was between 0.7 and 0.9; there were no significant problems with instability or oscillation using these values.

6.5.4 Results

Using the correction scheme outlined above, the scanner was used with the resolution target in various positions. The results presented here show three different arrangements, using different numbers of scan regions and different numbers of control points across the scan. The graphs shown in figures 6.9–6.11 show data recorded from detector C. Due to computing limitations, it was only possible to record from this detector when the mirror was not being driven (the mirror-drive routine was supplied written in such a way that it required complete control of the processor while the signal was being sent to the amplifier). Therefore, when the correction is being applied, there are gaps in the measured signal corresponding to the signals being sent to the mirror.

For each scan, there are 4 graphs shown. These show an uncorrected scan signal, a corrected scan signal, and the roughness. The vertical scale on the corresponding corrected / uncorrected graphs is the same in each case.

The roughness in the figures was calculated in the following way: firstly, a rolling mean-value was calculated by convolving the signal with a rectangular 30-point window. Then, the following quantity was calculated:

$$\sqrt{\frac{(I-\bar{I})^2}{\bar{I}^2}} \tag{6.11}$$

Table 6.2 gives the details of each scan.

Table 6.2 Scan details for figures 6.9–6.11

Property	Scan 1	Scan 2	Scan 3
Scan time	0.1s	0.1s	0.02s
Scan frequency	11.7Hz	11.7Hz	49.9Hz
Plane distortion	Curved and tilted	Curved and tilted	Curved
Amount of curvature	2cm	2cm	2cm
Number of correction zones	2	15	5

These results show that it is indeed possible to correct for a curved, tilted or displaced scan object. It has been demonstrated that an image sharpness-type criterion can be employed in order to derive a correction signal. The low repetition rate achieved can be largely attributed to the mirror drive electronics, and the control computer used.

6.6 Feasability of a Post-Objective Scanned A4 Laser Printer

One of the largest markets for optical scanners is the A4 laser printer. Current systems typically utilise several large optical components, thereby increasing manufacturing costs and system weight. The aim of this section is to demonstrate that, by using adaptive optics, one could construct a laser scanner that has only one lens, an adaptive mirror and a scanning mechanism. The scanner layout

Figure 6.8 Determining the appropriate correction across a scan based on a resolution criterion

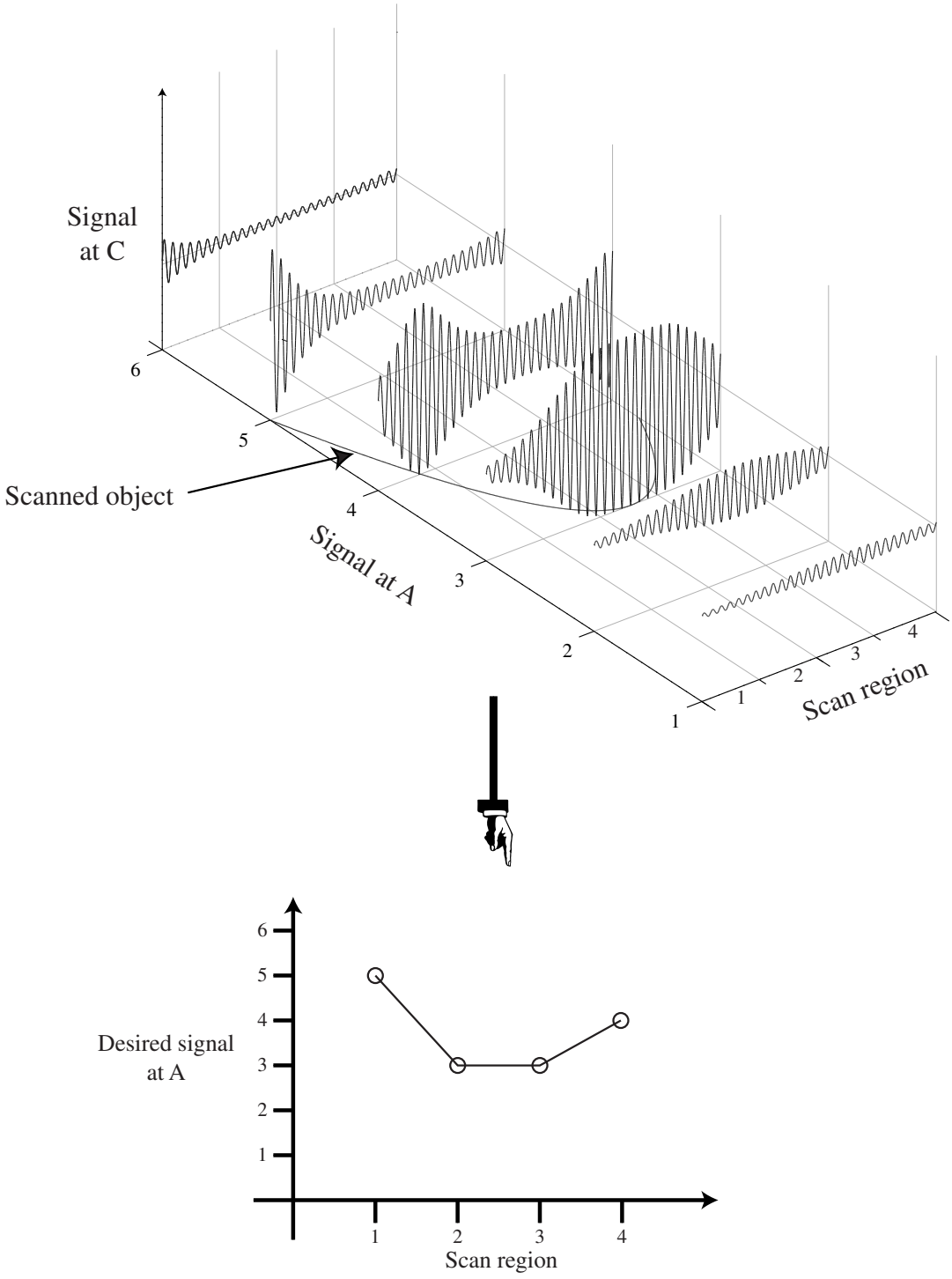


Figure 6.9 Results for scan 1

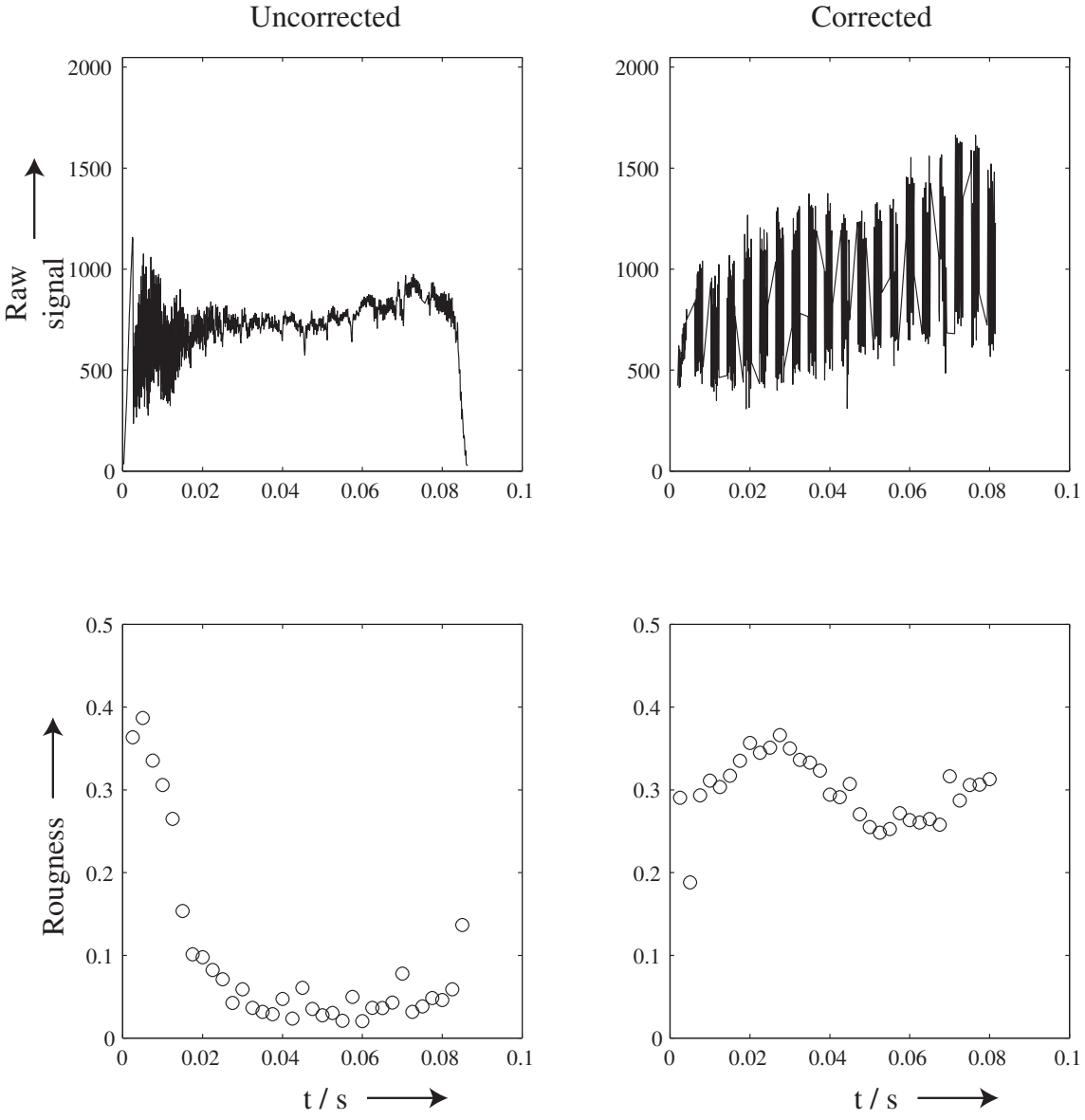


Figure 6.10 Results for scan 2

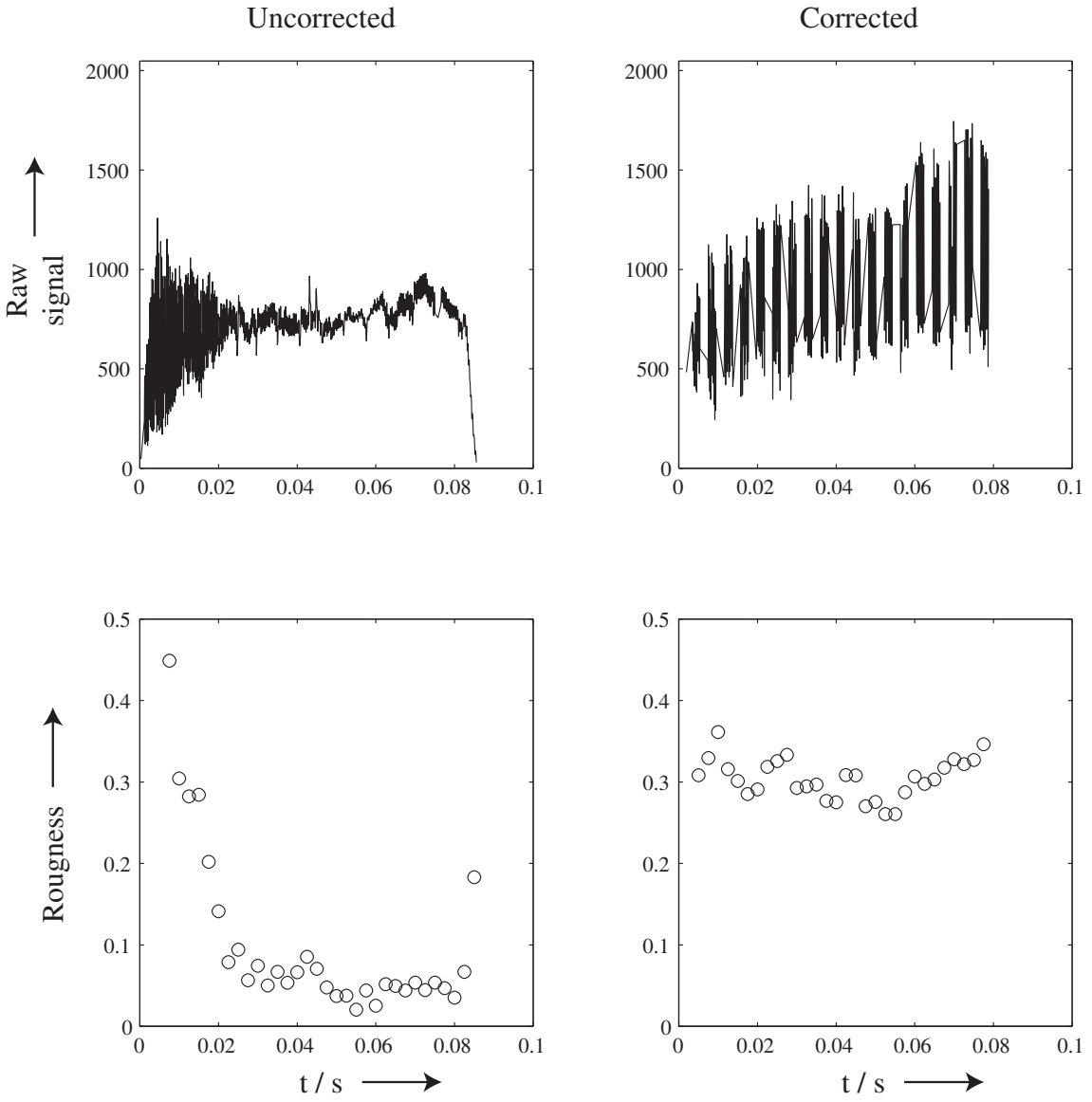
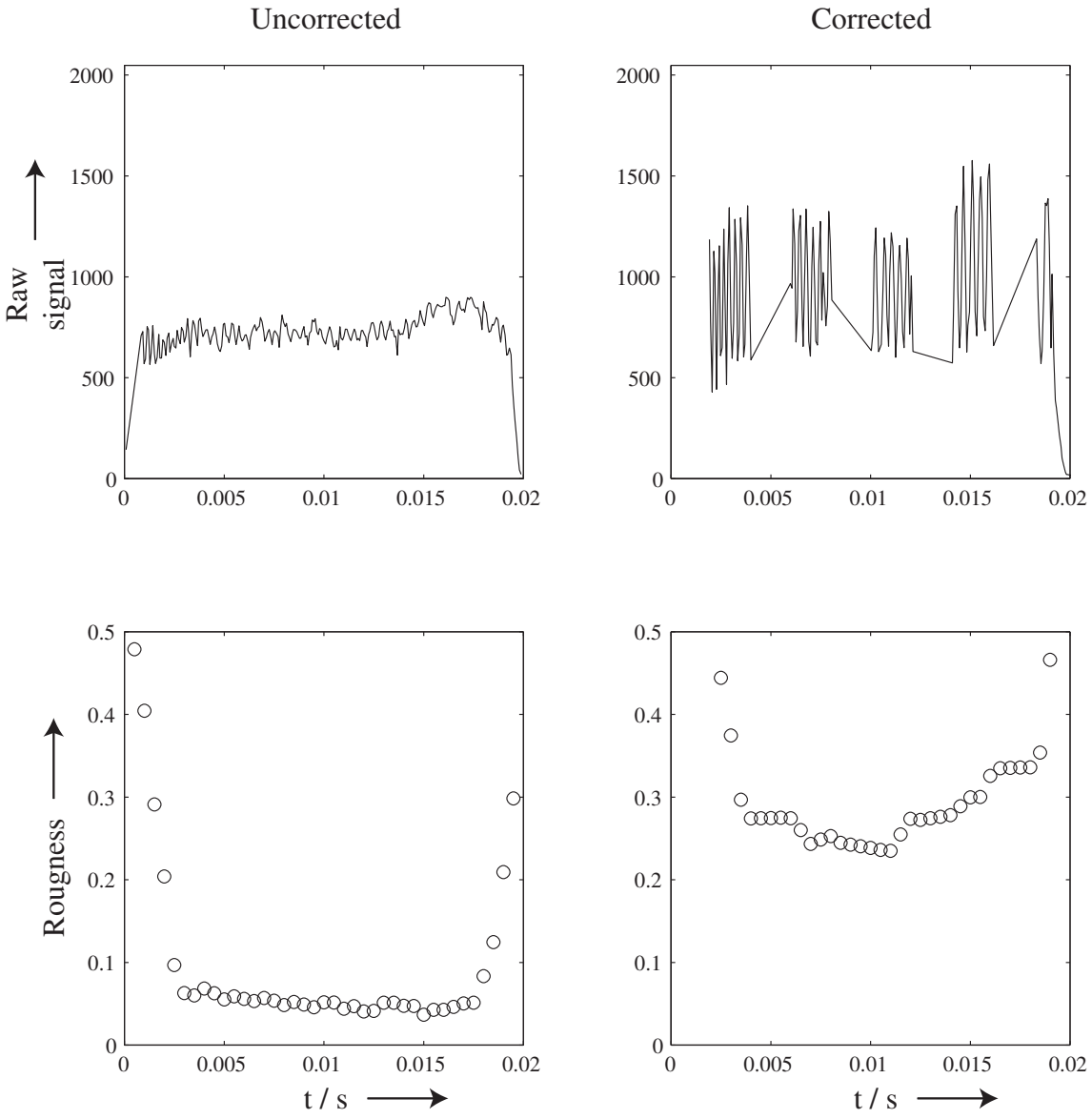


Figure 6.11 Results for scan 3



will be a post-objective type so that no physically large optical elements are required. Also, in this arrangement, the question of the double-pass problem is not of great importance as the only aberrations introduced will be those that result from using the optical system at conjugates other than those for whcih it was initially designed. These aberrations can only be of azimuthal order zero (for an objective that initially has only aberrations of azimuthal order zero), and hence will be sensed by either coherent or incoherent reflection in the scan plane (the aberration measured by a coherent reflection will be twice as great as that measured by an incoherent reflection). If we use the adaptive mirror at non-normal incidence, then there is the possibility of introducing higher azimuthal order aberrations. These aberrations will still be symmetric about the axis through which the mirror is rotated.

6.6.1 Design goals

The resolution required for a laser printer is usually specified in dots per inch (dpi)—currently, the best systems operate at up to 2000 dpi; 600 dpi is common for general purpose printing. This design shall attempt to achieve 600 dpi. The scan length is well defined: the width of a sheet of A4 paper is 210mm; most laser printers can print on a width of about 200mm. Therefore, the total number of spots across the scan is given by:

$$N = 600 \times \frac{200}{25.4} \approx 4700\,. \tag{6.12}$$

Therefore, we will design our scanner to be able to resolve 5000 points. If we use a laser with a Gaussian intensity profile and a wavelength of 650nm (a typical laser diode wavelength), then we can say from equation 6.3:

$$\Theta D = 5000 \times 1.27 \times 650 \times 10^{-9} \approx 4\text{mm rad}^{-1} \tag{6.13}$$

i.e. for a scan full-angle of 1 radian, a pupil diameter of 4 mm is required.

We can specify the required rate by noting that a normal A4 laser printer can print at a rate of about 6 pages per minute. Therefore, given that the length of an A4 sheet is 297mm, we can calculate the scan rate, r in lines s^{-1} as follows:

$$r = \frac{\text{lines per sheet}}{\text{seconds per sheet}} \tag{6.14}$$

$$= \frac{600 \times (297/25.4)}{10} \tag{6.15}$$

$$= 700\text{lines s}^{-1} \tag{6.16}$$

An additional requirement is that the scanned spot should ideally have a constant velocity. If this is not the case, the computation required to modulate the intensity of the write beam will be considerably more complex. A typical specification on the scan linearity is that the spot velocity should not vary from its average value by more than 1%.

6.6.2 Fundamental relationships

In order to meet the design goals, we must consider how the components available can be used to meet them. Firstly, we must assume that our deformable mirror can be deflected quickly enough. This should indeed be the case, as 700Hz should be well within the range of the two most likely options for the mirror: the bimorph and membrane mirrors. However, care must be taken when operating bimorph mirrors at high stroke coupled with high bandwidth. There are internal losses in

the piezoelectric material that can cause sufficient heat to depolarise the material. For both mirrors, the amount of stroke needed is found to be easily possible. We will also assume that the objective lens can give diffraction limited performance at the aperture used. Eventually, a detailed raytrace will show that this assumption is reasonable.

The main constraint in the optical design is the range of focal power that the adaptive mirror can reproduce. We are interested in the change in focal length that we can achieve, as this will determine how far from the scanning element the scan plane must be in order for it to be made flat from its initially circular shape. As a simple approximation, we will assume that the objective lens and adaptive mirror are in the same plane, and that they behave as thin lenses: this implies that the focal powers simply add. This will serve as a first approximation—exact raytracing can be carried out later for fine tuning as appropriate. The change in focal length is therefore:

$$\Delta f = \frac{1}{p_o + p_a} - \frac{1}{p_o} = \frac{-f_o}{1 + (p_o/p_a)} \tag{6.17}$$

where p_o is the power of the objective lens, p_a is the power of the adaptive mirror and f_o is the focal length with no adaption.

Now that we know the change in focal length that we can achieve, we need to find the change in focal power required for a given scan radius. Simple geometry tells us that if R is the distance from the scan mirror (assuming that the mirror rotates about an axis that is in the plane of the mirror—this is generally true for a galvanometer type mirror, and is a useful first approximation for the polygon case) to the scan plane; and ΔR is the extra distance to the scan plane at the edge of the scan. At the edge of the scan, the beam has been deflected through an angle of $\Theta/2$ from the central position, so:

$$\Delta R = R \times \frac{1 - \cos(\Theta/2)}{\cos(\Theta/2)} \tag{6.18}$$

We know that this distance must be the amount by which the deformable mirror can change the focal length, so we can construct a relationship between R, Θ, f_o and p_a:

$$R \times \frac{1 - \cos(\Theta/2)}{\cos(\Theta/2)} = -\frac{f_o}{1 + 1/(f_o p_a)} \tag{6.19}$$

If we say that there is a distance d between the objective lens and the scan mirror, then $R + d$ is equal to f_o.

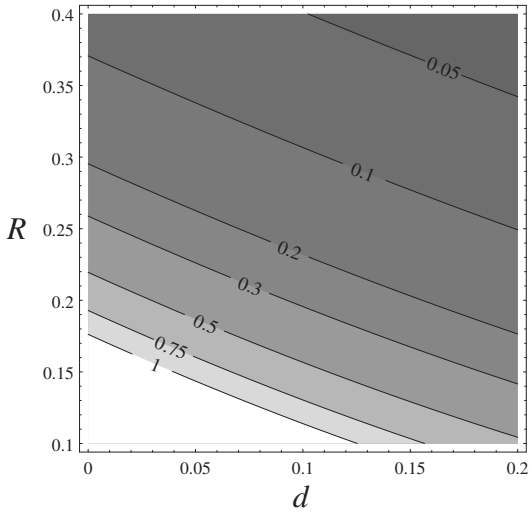
For the current scanner, we know that the length of the scan is fixed at 20cm, so we can eleminate Θ by noting that

$$\tan(\Theta/2) = 0.1/R \tag{6.20}$$

Using this expression in conjunction with equation 6.19, we can plot the required focal power as a function of the two distances d and R; this is shown as a contour plot in figure 6.12. In order to keep the size of the scan mirror (which, as shown in equation 6.13, depends only on the scan angle Θ, and therefore only on the scan radius R) as small as possible, we need to keep the scan radius as small as possible. To keep the scanner's footprint small, we must also keep the distance d small, although providing that it is smaller than the radius R, this should not add to much to the overall size of the scanner if the optical elements are folded suitably.

If we are using a galvanometer-type scanner, then we must consider the linearity of the spot velocity. Resonant galvanometers have a deflection that varies sinusoidally in time. We can show that this can be used to give a linear velocity in the scan plane. If the mirror deflection angle is θ, then the beam is deflected through 2θ, and the spot displacement is given by $R\tan(2\theta)$. For a

Figure 6.12 Focal power required in a post-object scanner as a function of scan radius R and offset d



galvanometer, θ can be expressed as $\theta_0 \sin(\omega t)$. This means that the spot displacement x can be written:

$$x = R\tan(2\theta_0 \sin(\omega t)) \tag{6.21}$$

We are interested in the normalised spot velocity, $\dot{x}/\dot{x}_0$. If we call this normalised velocity v, then from equation 6.21, we obtain:

$$v = \cos(\omega t)\sec^2(2\theta_0 \sin(\omega t)) \tag{6.22}$$

This expression allows us to optimise the velocity linearity by choosing θ_0, relative to the amount of the scan that we use. Figure 6.13 shows several plots of v for various choices of θ_0, for the first quarter cycle of the scan. This graph shows that if we choose θ_0 in the region of 0.75 we can find a region of nearly constant spot velocity in the central portion of the scan. This is achieved at the expense of duty cycle. We can use this formula to choose a galvanometer amplitude. The compromise that we make will be in the duty cycle and linearity of the scan.

Anticipating the use of a scan radius R of 0.2m, we can plot the scan distortion as function of position. The scan distortion is the difference between the actual spot position and a linearly predicted spot position (i.e. based on the assumption that the spot position is a linear function of time). Inverting equation 6.21, we can write the time as a function of the actual spot position:

$$t = (1/\omega)\arcsin\left(\frac{\arctan(x/R)}{2\theta_0}\right) \tag{6.23}$$

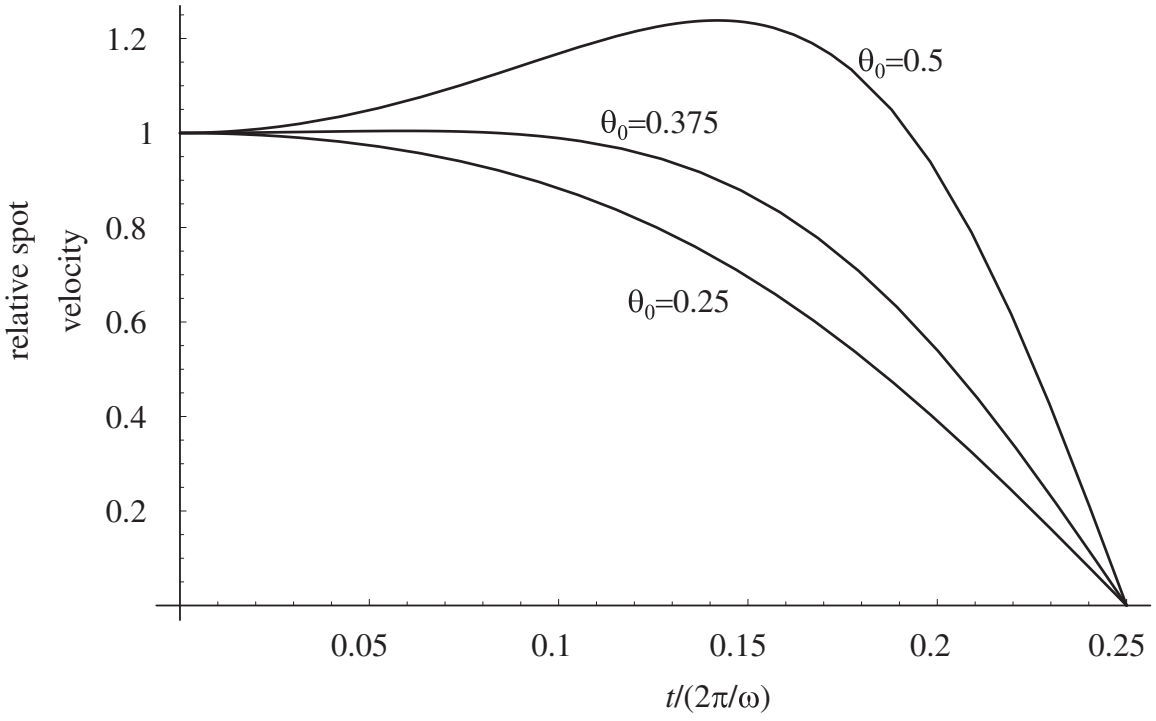
We can use this expression to find an expression for the error in assuming that the spot is travelling at the small-displacement velocity which means that the spot displacement is simply:

$$x_{\text{small}} = 2R\theta_0\omega t \tag{6.24}$$

Substituting equation 6.23 in equation 6.24 gives us an expression for the actual spot displacement in terms of the small-signal spot displacement.

Using this expression, we can plot the difference in actual scan position and predicted scan position as a function of actual scan position. If we use equation 6.24 on its own, we will find that

Figure 6.13 Normalised spot velocity as a function of time for the first quarter scan cycle



there will be apparently large amounts of distortion—some of this is real, but by adding a small scaling factor, we can reduce the amount of distortion. This is equivalent to choosing a different "reference velocity". An example showing that this technique can lead to very small amounts of distortion is shown in figure 6.14. In this example, θ_0 is chosen to be 0.365; R is 0.2m, and the reference spot position is $1.994R\theta_0\omega t$. The maximum distortion is just more than 0.1%, or just more than 0.2mm. This extremely small distortion is achieved at the expense of a duty cycle of only 40%. Unfortunately, even by allowing the maximum distortion to increase to 1%, the duty cycle only goes up to around 60%.

A final consideration is the accuracy with which the focal length of the mirror must be controlled. This is dependent on the depth of focus of the beam at the scan plane. The number of levels that the mirror needs to be able to reproduce (from which an estimate of the number of bits of data needed to drive the mirror can be made) can be found from

$$\text{Number of levels} = \frac{\text{Total focal change required}}{\text{Depth of focus}} \tag{6.25}$$

6.6.3 Detailed design

Now that we have covered the principles involved in designing our desired scanner, we can go on to choose a possible realisation of the system, and evaluate its performance in detail. The overall layout of the design is shown in figure 6.15. This shows all the dimensions that must be chosen.

For our example scanner, the nominal values chosen are shown in table 6.6.3. This data was entered into the Zemax ray-tracing program. For the purposes of optimisation, 5 angular positions were used, equally spaced across the scan. The objective lens used was based on a standard Melles Griot lens. During the optimisation process, the curvatures of the lens were allowed to vary, as was d_1. This correction was necessary because spherical aberration was introduced as a result of using the lens at conjugates different to those for which it was designed. This optimisation was

Figure 6.14 Scan distortion for a resonant galvonometer

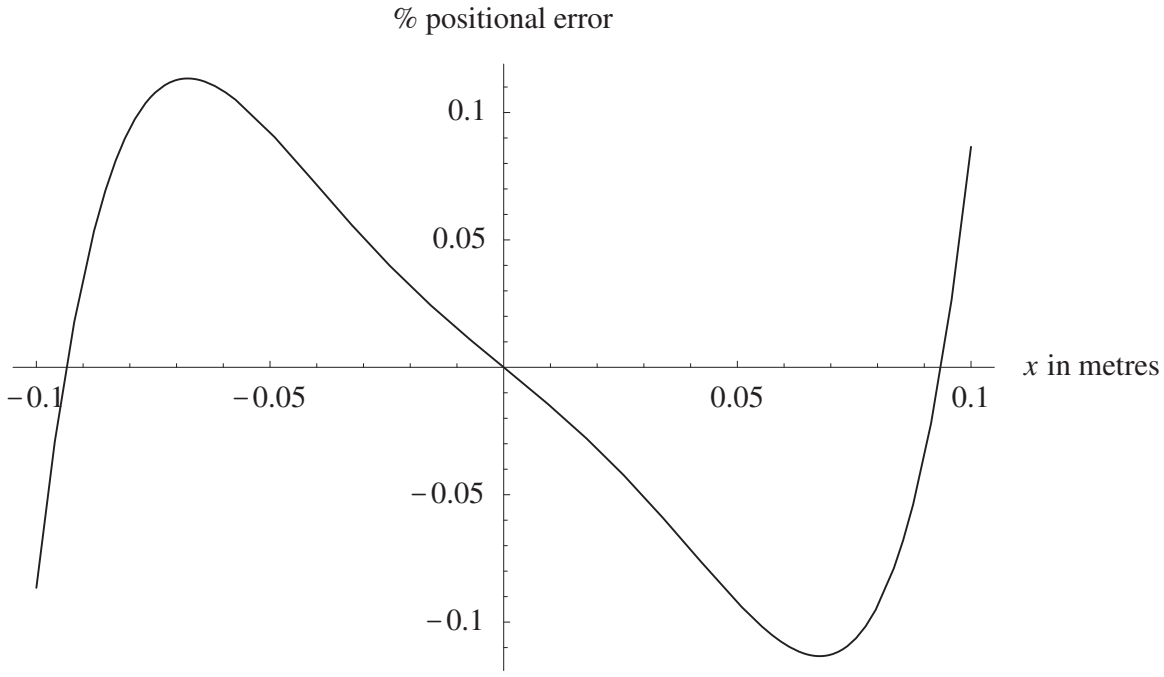
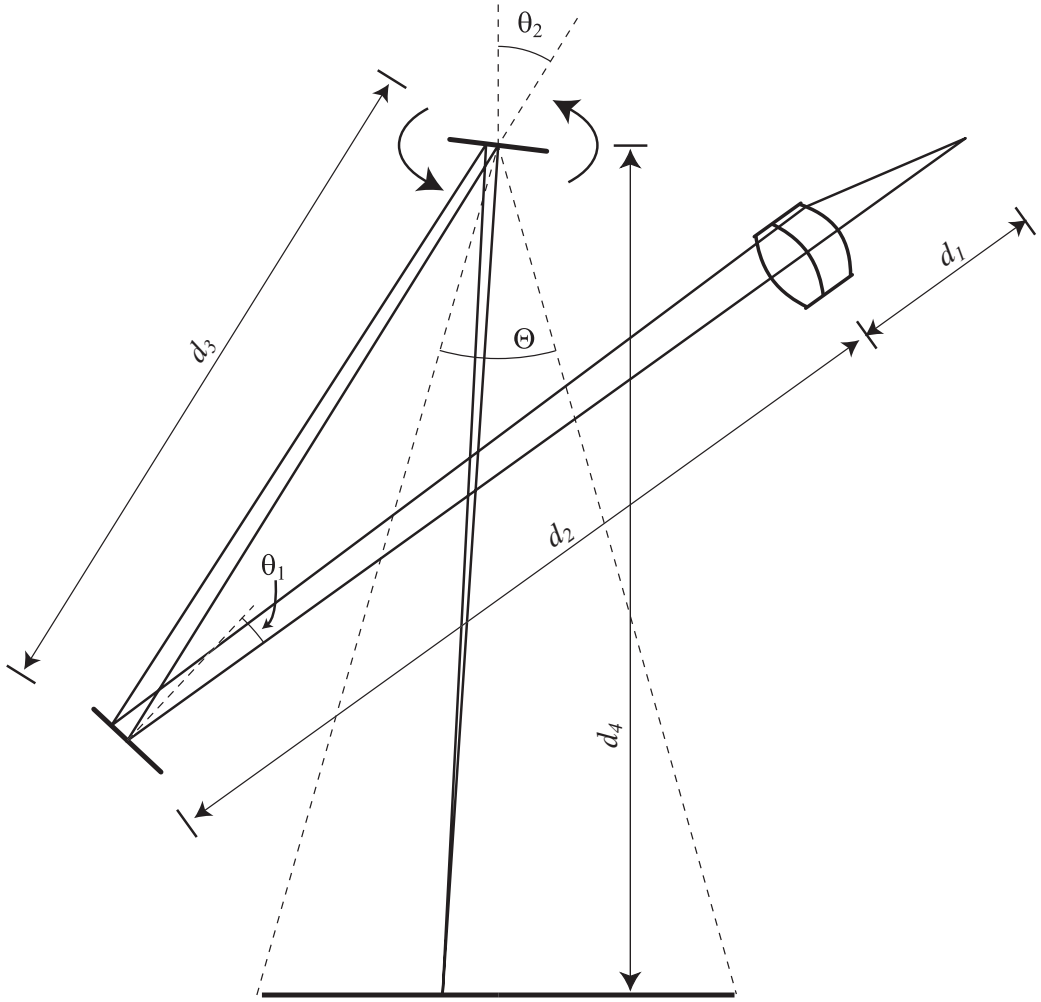


Table 6.3 Nominal values for A4 laser scanner

Dimension	Value
d_1	60mm
d_2	100mm
d_3	100mm
d_4	200mm
θ_1	10°
θ_2	18°
Θ	53°

Figure 6.15 Layout of an AO post-objective scanned A4 laser printer



done for the centre of the scan. After this central scan position had been optimised, the system was permitted to optimise the curvature of the deformable mirror for each of the scan positions in order to minimise the spot size. After this optimisation was complete, the spot sizes in the scan plane all had geometric radii of less than 15μm, and they were all therefore essentially diffraction limited. A graph of encircled energy is shown in figure 6.17 for one of the scan points—all points have essentially the same encircled energy profile. The detailed raytracing information is shown in table 6.4, and an accurate scale drawing is shown in figure 6.16.

Table 6.4 Raytracing information for a preobjective scanned A4 laser printer

Surface	Radius of Curvature (mm)	Radius (mm)	Thickness (mm)	Glass
Object	0	0	62.87 (d_1)	
1	66.6013	7	3.7	SSK4A
2	-15.4402	7	1.5	SF8
3	-51.1959	7	0	
Aperture	0	7	100 (d_2)	
DM	$1/\kappa$	7	-100 (d_3)	Mirror
Scan mirror	0	7	R (d_4)	Mirror
Scan plane				

The curvatures of the deformable mirror and the distance from the scan mirror to the scan plane depend upon the scan angle. These values are shown in table 6.5 From this table we can see that

Table 6.5 Variable parameters in the laser scanner

Position	Angle	κ mm^{-1}	R mm
1	-12°	5.0862×10^{-5}	218.9
2	-6°	-2.306×10^{-5}	204.5
3	0°	-4.768×10^{-5}	200.0
4	6°	-2.307×10^{-5}	204.5
5	12°	5.0732×10^{-5}	218.9

the maximum amount of focal power required is ± 0.05 dioptres, which is well within the range of bimorph and membrane mirrors.

We can calculate the accuracy with which we must control the mirror as follows. We know that a focal change of 0.0984 dioptres corresponds to a focal shift of 18.9 mm. We also know that the f number of the beam is roughly 400/12 = 33.3. The depth of focus of a diffraction limited beam is given by [80]:

$$\delta = \frac{\lambda}{2(\mathrm{N.A.})^2} \tag{6.26}$$

From this we can see that the vertical spot position must be controlled to better than 0.4mm. This means that the deformable mirror must have at least 50 control levels across the scan (if it is to be controlled digitally), and be controlled 100 times across the scan, since the vertical position varies from 18.9mm to zero and back again across the scan. This would be quite a difficult control task, were it not for the fact that the change in focus required is very smooth, and so an accurate analogue signal would be more appropriate.

Figure 6.16 Detailed layout for a preobjective scanned A4 laser printer at 3/4 scale

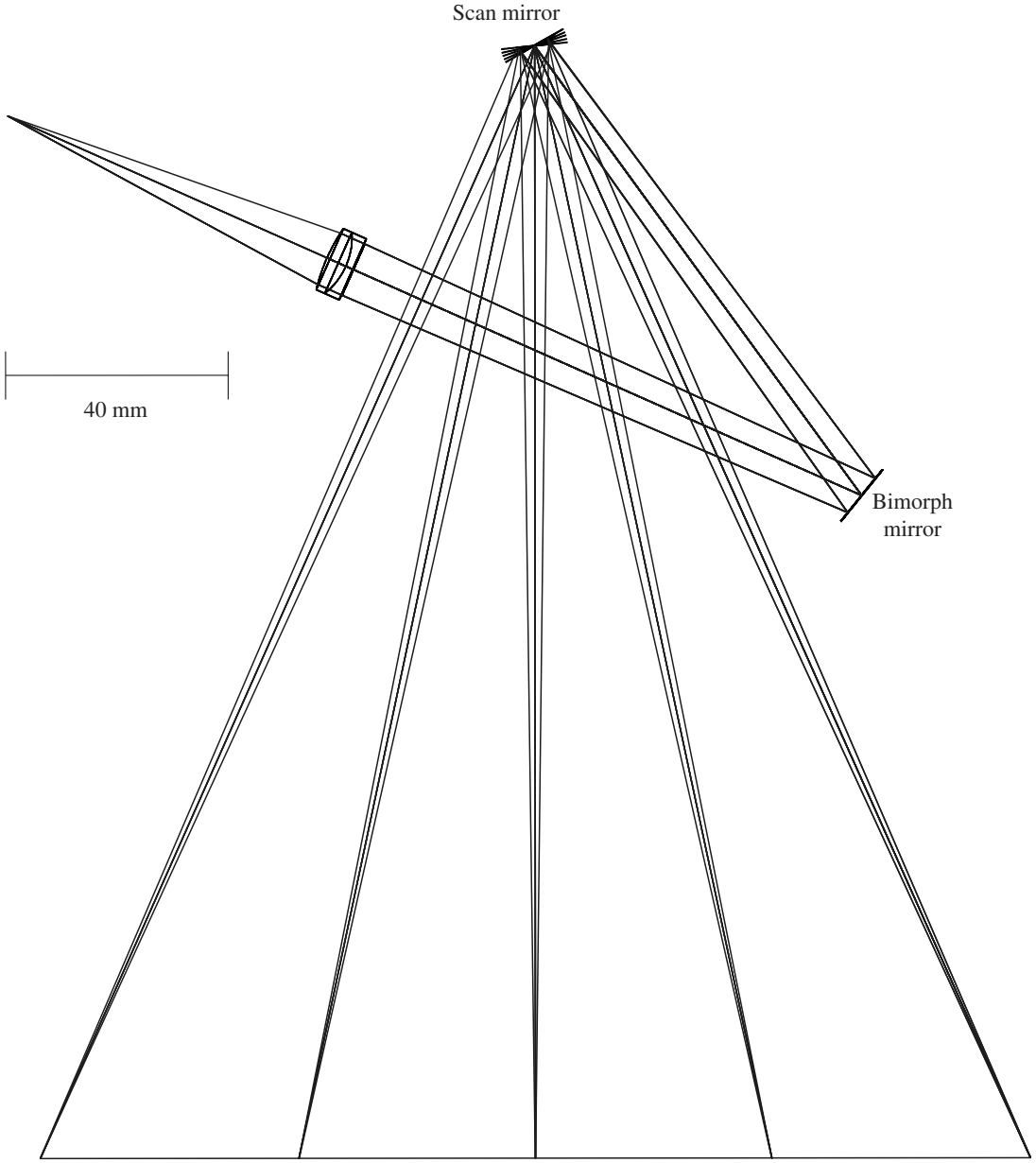
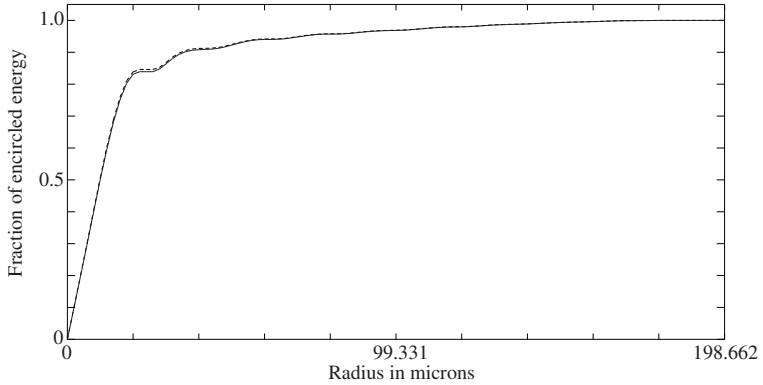


Figure 6.17 Encircled energy for postobjective scanner. Dotted line represents the diffraction limit



6.6.4 Conclusions from the detailed raytrace

The raytrace carried out above has demonstrated that it is indeed possible to make a scanner with high enough resolution to function as an A4 laser printer using a post-objective type scanner. Not only does this design allow for scanning at 600dpi, by simply changing the amount of defocus applied, it would also permit scanning at lower resolution. This would allow, for example, much faster "draft" mode printing. It would also allow large areas of flat colour to be shaded more quickly (this technique is used in some bubblejet-type printers).

7 Conclusions

This thesis has presented a detailed theoretical and practical analysis of the bimorph mirror. Firstly, a review of adaptive optics was presented in order to provide a context in which a bimorph mirror would be used. This review also described the various competing deformable mirror technologies such as the segmented mirror, and the membrane mirror. Once the necessary background of optics and adaptive optics had been described, some practical AO systems were described to demonstrate the current range of applications of AO.

With the preliminary review completed, a mechanical model of a bimorph mirror based on the classical laminate theory was presented. This developed expressions both for the overall curvature sensitivity of the bimorph mirror, and also the detailed surface shape produced when a voltage is applied to only one part of the device. With the aid of a computer-based analytical calculation program, a Green's function solution to the biharmonic equation was found. This solution was then used to perform detailed numerical simulations of the performance of a bimorph mirror in the presence of Kolmogorov turbulence. Optimised designs were produced for a variety of electrode layouts, and an approximate relationship was discovered between the number of electrodes and the strength of turbulence that could be adequately corrected was found. A brief consideration of the interaction between a bimorph mirror and a curvature wavefront sensor was presented, and it was found that a bimorph mirror is well sensed by a curvature sensor, although the relationship is not entirely diagonal.

A detailed recipe for the succesful construction of a bimorph mirror was given, based on the experience gained in the construction programme undertaken during the course of this work. Results from the mirrors constructed were then presented. The mirrors were found to have sensitivities up to 60μm/kV, which compares well with the highest presented in the literature. The measured sensitivity was compared to the calculated value, and a degree of consitency was observed. It was also shown how the use of a substrate with carefully chosen thermal properties could alleviate the thermal instabilities inherent in bimorph structures. A method for overcoming the hysteresis problems inherent in all piezoelectric materials was applied for the first time to a bimorph mirror, and good results were observed when the dynamic behavious was studied. It is believed that with more careful application of the strain gauges used, and with a custom strain gauge amplifier, that the useful lower limit on the operating frequency of the strain measurements may approach zero, thus enabling a focal length simply to be "dialed in" without the need for any optical feedback. The surface shapes produced when a voltage was applied only to a region of the mirror were measured and presented. Again, these were compared with the theory. Although the surface shapes agreed quite well qualitatively, it was found that the higher spatial frequencies were somewhat underrepresented. The causes for this spatial filtering were considered.

A new field of application for AO techniques was considered, namely that of AO-assisted laser scanning. Many practical details such as wavefront sensing and scan linearity were considered. Results from a practical AO scanner were presented. The scanner used a "image-sharpness" type correction criterion, and it was found that the AO scanner could could correct for misalignments of

the simple resolution target. A detailed optical design for an A4 laser printer was presented, using only one small lens, a defocus-only deformable mirror and a resonant galvo-type scanner. This was found to be capable of reaching resolutions up to 600dpi, without the need for any large optics at all.

7.1 Creating Commercial Bimorph Mirrors

One of the principal goals of this work was to acquire the knowledge and expertise necessary to model and create bimorph mirrors, and arrive at an assembly technique sufficiently robust to supply devices in a semi-commercial manner. While the theoretical side of the project is adequate for modelling low-order devices, the bimorph mirrors created during the course of this work, although succesful, could not be described as being truly ready to be marketed commercially. In order that this status may be achieved, there are still issues remaining that have not been answered due to time and equipment limitations.

Firstly, a more satisfactory method of patterning the electrodes must be found. Work currently proceeding at IC using more appropriate etching techniques has shown promising early results, although at the time of writing, none of these etched disks of piezoceramic have been assembled into functioning bimorph mirrors.

The question of reliability is one which needs more detailed investigation. Although data sets taken one year apart showed almost no difference in mirror performance, more detailed reliability measurements ought to be made. These should include regular measurements of the detailed "off" position of the mirror, in addition to regular measurements of the form and sensitivity of the various electrodes. In between these tests, the mirror should be subjected to regular usage, as might be anticipated in an application such as astronomical use.

Mounting of the devices has not always proved to be a simple matter. In order that the range of deformations achievable might not be too severly restricted, only relatively "loose" mountings have been used. This may allow an unacceptable amount of wobble. In this regard, edge mounting may well prove to give a more sturdy support, at the expense of reduced sensitivity and response functions that have a smaller spatial extent.

Creating the power supplies for bimorph mirrors can be somewhat problematic, and expensive. The requirements on the high-voltage amplifiers are quite demanding. The capacitance per channel of a typical bimorph mirror would normally be in the region of 5–10nF. If this is to be driven at high speed (i.e. up to 2kHz) and at the full range of the piezoceramce (normally $\pm$300V), then a large amount of electrical power must be expended. If a bimorph mirror is to be used in an astronomical (or similar) situation, it is the power supply limitations, rather than the total stroke achievable, that will demand that the device has a high sensitivity.

7.2 Future Directions for Low-Cost Adaptive Optics

There is still a paucity of the use of AO outside the traditionally high-budget arenas of large telescopes and large lasers. At the inaugural "International Conference on Adaptive Optics for Industry and Medicine" held in 1997, many of the talks described *potential* applications for low-cost AO. In the two intervening years, despite more published [3] and patented examples of potential applications, there is yet to emerge an application of low-cost AO that has become in any sense successful.

Although "low-cost" does not necessarily imply "low-order", I believe that it is in the lowest order applications that low-cost AO might find success. It is futile to hope that an AO system based on currently available technologies, even at the highest cost, could ever correct a useful field-of-view for ground-based imaging. During the course of this work, I have attempted to identify applications

in which AO could make a useful contribution to an instrument without incurring a prohibitive cost. One such example is the laser scanner, described in detail in chapter 6. Another application not described elsewhere in this work is thermal lens correction in high-power Nd:YAG lasers. It is well known that the thermal stresses in Nd:YAG laser rods cause a large change in focal length as the pump power changes. This means that the laser cavity must be realigned each time the pump power is changed—an undesirable situation. By incorporating a defocus-only mirror in the laser cavity, it may be possible to run the laser at a range of pump powers with minimal intervention.

By pursuing applications that are somewhat limited in scope, there is the opportunity for low-cost AO to become more widely used. Then, as the technology evolves still further, it may become feasible to consider more ambitious projects.

A Zernike Polynomials

A.1 Definition

The definition of Zernike polynomials used throughout this work is that given by Noll [66]. That form is:

$$Z_j = \begin{cases} \sqrt{2(n+1)}R_n^m(r)\cos m\theta & m \neq 0 \wedge j \text{ is even}, \\ \sqrt{2(n+1)}R_n^m(r)\sin m\theta & m \neq 0 \wedge j \text{ is odd}, \\ \sqrt{n+1}R_n^0(r) & m = 0\,, \end{cases} \tag{A.1}$$

where

$$R_n^m(r) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s!\,((n+m)/2-s)!\,((n-m)/2-s)!} r^{n-2s}\,. \tag{A.2}$$

With this definition, valid for $r \leq 1$, we note that some of the terms may be related to their traditional names as shown in table A.1. A graphical representation of this table is given in figure A.1. For clarity, only one of the azimuthal versions is shown.

The most important property of the Zernike polynomials is that they are orthonormal over the unit circle.

A.2 Least Squares Expansion

We follow the usual procedure for a least squares expansion to convert a sampled surface map into a vector of Zernike coefficients. The first stage in the fitting procedure is to convert the two space coordinates (e.g. (x,y) or (r,θ)) into a single indexed coordinate so that we can use directly the usual notation. If the points are not equally spaced in area, then we need to assign them different weights, otherwise the fit will be biased towards those regions where the calculated points are densest. In the calculations performed as part of this work, only regularly spaced rectangular coordinates were used, so each point has the same weight. The first thing we need to write down is the model that we are using to fit our data points. This can be written [82]:

$$\phi(x,y) = \sum_{k=1}^{M} a_k Z_k(x,y) \tag{A.3}$$

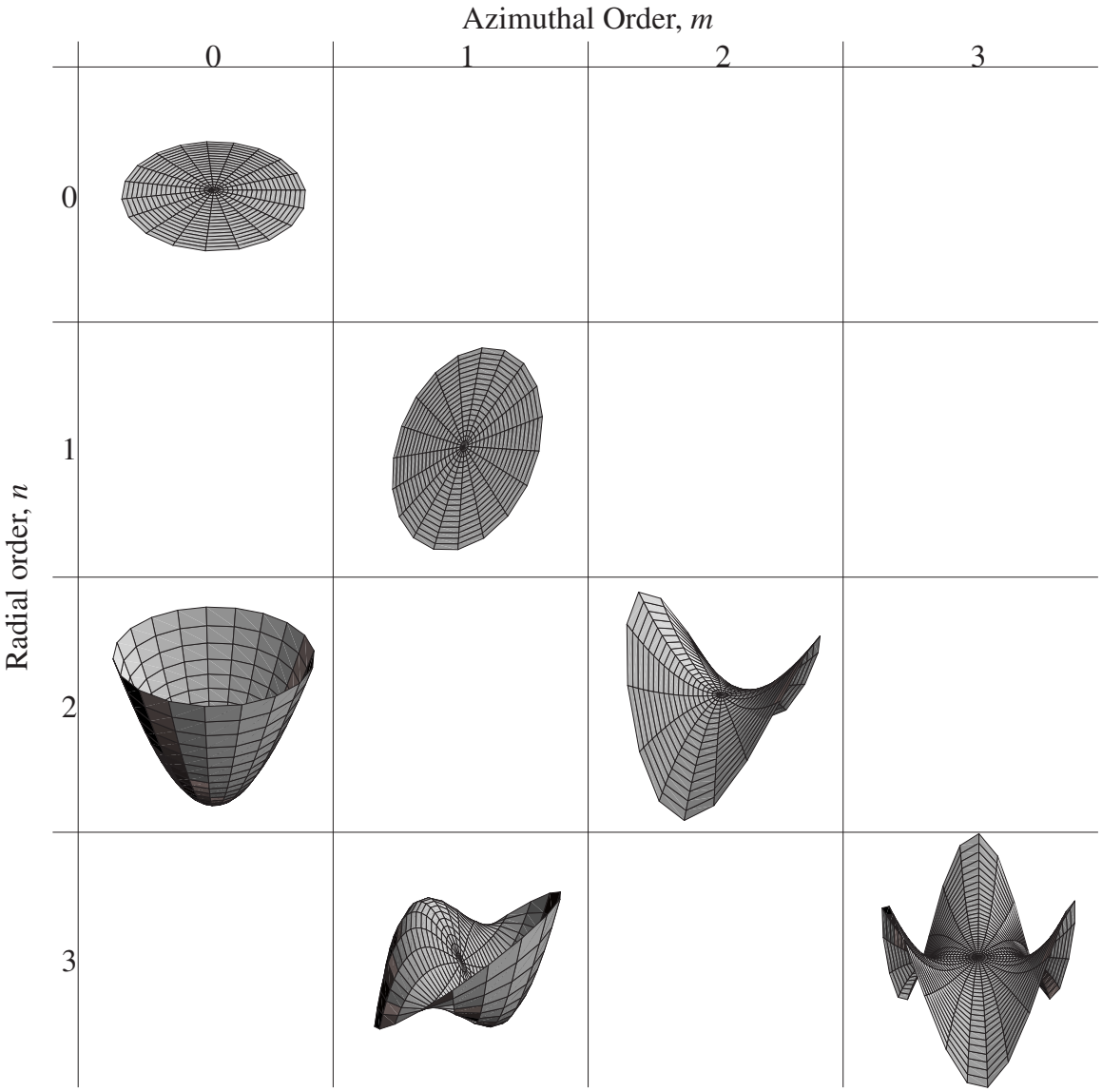
where M is the number of Zernike polynomials to be fitted, and a_k is the weight of that order. If i is the index of the N data points y_i, then the general unweighted least squares fitting parameter is:

$$\chi^2 = \sum_{i=1}^{N} \left[y_i - \sum_{k=1}^{M} a_k Z_k(x_i, y_i) \right]^2 \tag{A.4}$$

Table A.1 Zernike polynomials and their common names

n	m			
	0	1	2	3
0	$Z_1 = 1$ Piston			
1		$Z_2 = 2r\cos\theta$ $Z_3 = 2r\sin\theta$ Tip and Tilt		
2	$Z_4 = \sqrt{3}(2r^2 - 1)$ Defocus		$Z_5 = \sqrt{6}r^2\sin 3\theta$ $Z_6 = \sqrt{6}r^2\cos 3\theta$ Astigmatisms	
3		$Z_7 = \sqrt{8}(3r^2 - 2r)\sin\theta$ $Z_8 = \sqrt{8}(3r^2 - 2r)\cos\theta$ pure coma		$Z_9 = \sqrt{8}r^3\sin 3\theta$ $Z_{10} = \sqrt{8}r^3\cos 3\theta$ zero-curvature coma

Figure A.1 Surface plots of the first few Zernike terms



Next, we define the "design matrix", $\mathbf{A}$ whose components A_{ij} are given by:

$$A_{ij} = Z_j(x_i, y_i) \tag{A.5}$$

In terms of this design matrix, we can rewrite equation A.4 as follows:

$$\chi^2 = |\mathbf{A} \times \mathbf{a} - \mathbf{y}|^2 \tag{A.6}$$

The pseudoinverse of $\mathbf{A}$ as found by SVD minimises χ^2, and so our least-squares fitted parameters $\mathbf{a}$ are found:

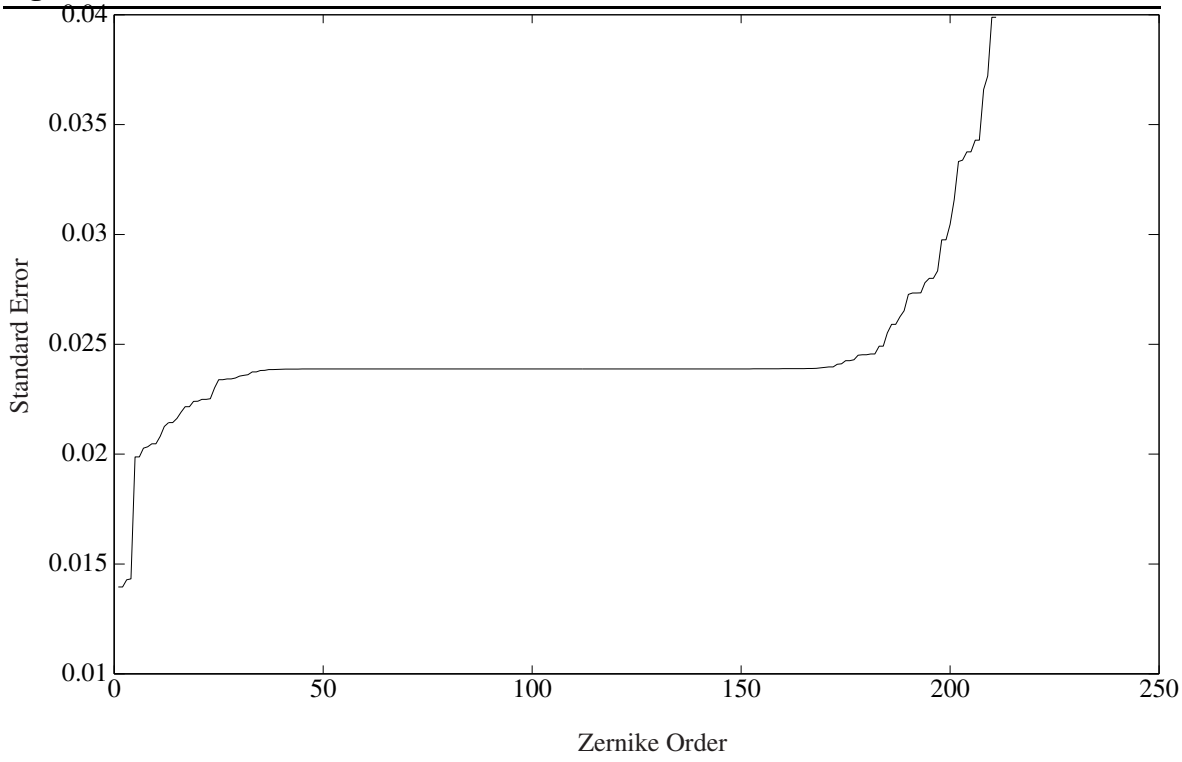
$$\mathbf{a} = \mathbf{A}^* \times \mathbf{b} \tag{A.7}$$

We can also find an estimate of the errors in the fitted parameters:

$$\sigma^2(a_j) = \sum_{i=1}^{M} \left(\frac{V_{ji}}{w_i} \right)^2 \tag{A.8}$$

where the w_i are the singular values from the SVD of $\mathbf{A}$. Hence, we can see that the standard deviation of the fitted parameters is constant if we have equally weighted points in our surface map, the standard errors of the fitted parameters will always be the same. These standard errors are plotted in figure A.2 for the case when 211 Zernike orders are considered, calculated over a 64 points square grid.

Figure A.2 Standard error in the fitted Zernike coefficients



A.3 Rotation of a Zernike Spectrum

In the construction of a transfer matrix for a simulated mirror, it is useful to have a method of rotating a Zernike spectrum (i.e. effectively fitting a spectrum to a rotated surface). This can be done without

compromising accuracy, as would happen if the surface map itself were rotated—this would involve an interpolation stage. This rotation is simple to perform, if we note that each Zernike term is of the form:

$$Z_j(r,\theta) = f_{n,m}(r) \begin{cases} \sin(m\theta) \\ \cos(m\theta) \\ 1 \end{cases} \tag{A.9}$$

Obviously, the terms which do not depend on θ require no modification. If we have a surface $\phi(r,\theta)$ expanded in terms of a vector of coefficients a_i, then we can immediately say that:

$$\phi(r,\theta-\psi) = \sum a_i Z_i(r,\theta-\psi) \tag{A.10}$$

We need to find a new vector of expansion coefficients a_i' for this rotated surface. We can say that:

$$a_i' = \iint \phi(r,\theta-\psi) Z_i(r,\theta)\,\mathrm{d}\mathbf{r} \tag{A.11}$$

$$= \iint \left(\sum a_j Z_j(r,\theta-\psi)\right) Z_i(r,\theta)\,\mathrm{d}\mathbf{r} \tag{A.12}$$

Using the trigonometric identies:

$$\sin(m(\theta-\psi)) = \cos(m\psi)\sin(m\theta) - \sin(m\psi)\cos(m\theta) \tag{A.13}$$

$$\cos(m(\theta-\psi)) = \cos(m\psi)\cos(m\theta) + \sin(m\theta)\sin(m\psi) \tag{A.14}$$

we can write (assuming that indices i and $i+1$ are a pair which share the same values of n and m):

$$a_i f_{n,m}(r)\sin(m(\theta-\psi)) = a_i f_{n,m}(r)\left[\cos(m\psi)\sin(m\theta) - \sin(m\psi)\cos(m\theta)\right] \tag{A.15}$$

$$a_{i+1} f_{n,m}(r)\cos(m(\theta-\psi)) = a_{i+1} f_{n,m}(r)\left[\cos(m\psi)\cos(m\theta) + \sin(m\theta)\sin(m\psi)\right] \tag{A.16}$$

Using the orthogonality properties of Zernike polynomials, we can immediately write down the result:

$$a_i' = a_i\cos(m\psi) + a_{i+1}\sin(m\psi) \tag{A.17}$$

$$a_{i+1}' = a_{i+1}\cos(m\psi) - a_i\sin(m\psi) \tag{A.18}$$

We can also use the above approach to convert a vector of Zernike coefficients into a "polar" representation, which stores the angular orientation and magnitude of each Zernike term. To do this, we can solve our above equations for ψ if we constrain $a_{i+1}' = 0$. Then, equation A.18 gives us:

$$\tan(m\psi) = \frac{a_{i+1}}{a_i} \tag{A.19}$$

Squaring both equations A.17 and A.18 gives us, after simplification:

$$a_i' = \sqrt{a_i^2 + a_{i+1}^2} \tag{A.20}$$

This representation is useful for fitting to experimental data, since the angular orientation of the measured data is usually known only approximately.

A.4 Radial Remapping of a Zernike Spectrum

In addition to rotations, we can perform an analytical remapping equivalent to choosing a new radius over which to perform the expansion. That is, if we have some function $\phi(r,\theta)$ which has a Zernike expansion:

$$\phi(r,\theta) = \sum a_i Z_i(r,\theta) \tag{A.21}$$

then we seek an expansion

$$\phi(r,\theta) = \sum b_i Z_i(r/R,\theta) \tag{A.22}$$

for the region $0 < r \leq R$, where $R \leq 1$. For this region, we can equate the right hand sides of equations A.21 and A.22. Next, we multiply each side by $Z_j(r/R,\theta)$, and integrate over the region $0 < r \leq R$, which gives us:

$$\int_{-\pi}^{\pi}\int_{0}^{R} \left(\sum a_i Z_i(r,\theta)\right) Z_j(r/R,\theta) r\,\mathrm{d}r\,\mathrm{d}\theta = \int_{-\pi}^{\pi}\int_{0}^{R} \left(\sum b_i Z_i(r/R,\theta)\right) Z_j(r/R,\theta) r\,\mathrm{d}r\,\mathrm{d}\theta \tag{A.23}$$

By the orthogonality of the Zernike polynomials, the term on the right hand side is only non-zero when $i = j$, and in this case, the integral is $b_j \pi R^2$. Recalling that the Zernike polynomials are separable into the form given in equation A.9, we can immediately see that only terms with the same azimuthal form will be non-zero. Using the form $Z_i = f_{n_i,m_i}(r) g_{m_i}(\theta)$, we can re-write the integral on the left-hand side of equation A.23:

$$b_j \pi R^2 = \sum_{i=1}^{\infty} \left(\left[\int_{-\pi}^{\pi} g_{m_i}(\theta) g_{m_j}(\theta)\,\mathrm{d}\theta\right] \left[\int_{0}^{R} f_{n_i,m_i}(r) f_{n_j,m_j}(r/R) r\,\mathrm{d}r\right] \right) \tag{A.24}$$

Because of the orthogonality of the harmonic form of g, only the terms with $m_i = m_j$ have the possibility of being non-zero. In these cases, the integral is simply 2π when $m = 0$, and π when $m \neq 0$. Then, to calculate the remaining integral, we need to find all the radial forms $f_{n,m}$ which share the value of m_j. Unfortunately, a closed form for the overall result is not possible. We proceed, therefore, by calculating a transfer matrix between the a_i and b_j, by writing

$$b_j = \sum a_i A_{ij} \tag{A.25}$$

where A_{ij} is the product of the integrals in equation A.24. To calculate the matrix of A_{ij} values, we proceed by first selecting only those combinations of i and j which give a non-zero result. Then, for those interacting terms, the integral is most easily performed by calculating the product of the radial polynomials, given in equation A.2. This results in a product-polynomial, which can be integrated analytically with ease.

A.5 Strehl Ratio from Zernike Coefficients

The Strehl ratio is a measure of imaging quality: it is defined as the ratio of the on-axis intensity of an aberrated point-spread function to the on-axis intensity of the point-spread function that would be produced if there were no aberrations present, see section 2.1.3. This quantity is directly related to the variance of the wavefront (in the limit of small variance), and hence is often used to express how well a given wavefront matches another. The Strehl ratio is given [10]:

$$S = 1 - \sigma_\phi^2 \tag{A.26}$$

providing ϕ is expressed in radians of phase. If ϕ is expressed in terms of distance, then a factor of $2\pi/\lambda$ must be incorporated.

Since the Zernike polynomials are orthonormal over a unit circle, it is straightforward to calculate the Strehl ratio directly from them. Firstly, we need the wavefront variance:

$$\sigma_\phi^2 = \frac{\iint \left(\phi(\mathbf{r}) - \bar{\phi}(\mathbf{r})\right) \, \mathrm{d}A}{\iint \mathrm{d}A} \tag{A.27}$$

Since each term of the Zernike expansion integrates to the same value over a unit circle, this expression becomes simply:

$$\sigma_\phi^2 = \sum a_i^2 \tag{A.28}$$

A.6 Scaling of the 4th Term

In some cases, it is useful to rescale the Zernike magnitude of the 4th term into either a curvature, or a radius of curvature. The most common case encountered in this work is that the Zernike magnitude is expressed in terms of a wavelength. Consider a spherical cap of radius r, radius of curvature R and centre-to-edge deflection h. Using the intersecting chords theorem, we can say:

$$2Rh = r^2 \tag{A.29}$$

This can be re-arranged to give:

$$R = \frac{r^2}{2h} \qquad \text{or} \qquad \kappa = \frac{1}{R} = \frac{2h}{r^2} \tag{A.30}$$

Next, we need to find h from a_4. From the definition in equation A.1, we see that the centre-to-edge deformation is:

$$h = \left(2\sqrt{3}\lambda\right) \frac{a_4}{2} \tag{A.31}$$

where the term in brackets accounts for the centre-to-edge deformation of the Zernike term, and the division by two accounts for the fact that it is the wavefront distortion (which is twice as large as the surface distortion) that is measured by the interferometer.

$$R = \frac{r^2}{2\sqrt{3}\lambda a_4} \qquad \text{or} \qquad \kappa = \frac{1}{R} = \frac{2\sqrt{3}\lambda a_4}{r^2} \tag{A.32}$$

The focal length of a spherical mirror is simply half of its radius of curvature; and its optical power is the reciprocal of the focal length.

B Ellipse Fitting

When Zernike polynomials are to be fitted to a measured data file, there is a need to match the coordinates over which the Zernike polynomials are evaluated to the natural coordinates of the data file. In the experiments conducted in this thesis, data were taken from circular mirrors, and measured with a camera with non-square pixels. Therefore, there is a need to fit an ellipse to the measured data. Since we know that the major and minor axes are constrained to lie along the axes of the data matrix, our task of fitting an ellipse is made somewhat simpler than the general case.

The algorithm developed proceeds as follows:

- Firstly, read in a matrix of data points.
- Make all those points within in the pupil take the value 1, and all those outside the pupil take the value 0.
- Calculate the sum along rows and columns of the data
- Square this vector element-by-element

The result of this procedure is two vectors, each of which contains information about the radii and centres of the ellipse. The vectors contain a series of values of h^2. We can say:

$$\frac{(x-x_c)^2}{r_x^2}+\frac{(y-y_c)^2}{r_y^2}=1 \tag{B.1}$$

Now, taking the example where we have performed the summation in the y direction, equation B.1 becomes:

$$\frac{(x-x_c)^2}{r_x^2}+\frac{h^2}{4r_y^2}=1 \tag{B.2}$$

This can be rearranged to give:

$$h^2=x^2\left(-\frac{4r_y^2}{r_x^2}\right)+x\left(\frac{8x_cr_y^2}{r_x^2}\right)+\left(4r_y^2-\frac{4r_y^2x_c^2}{r_x^2}\right) \tag{B.3}$$

The three bracketed expressions are the parameters that will be fitted when a parabola is fitted by least squares to h^2. Referring to the parameters as A_2, A_1 and A_0 as the coefficients of x^2, x^1 and x^0 respectively, we find:

$$x_c=\frac{A_1}{-2A_2} \tag{B.4}$$

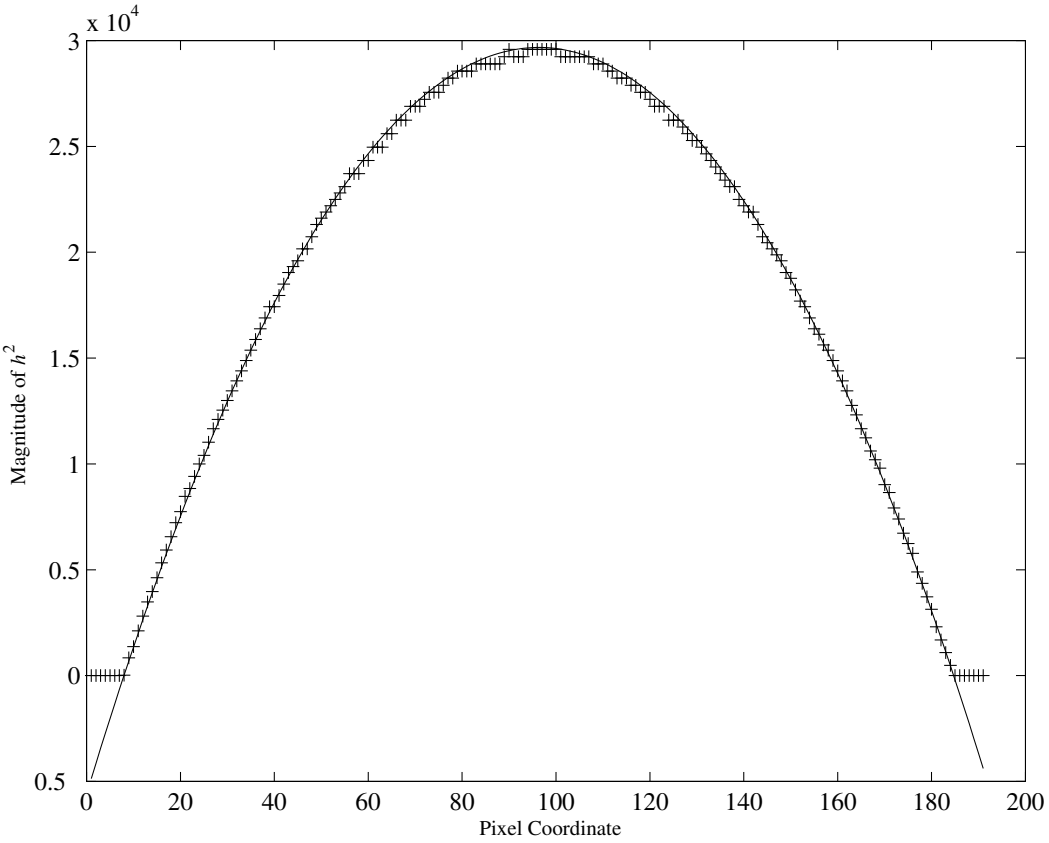
$$r_y=\frac{1}{2}\sqrt{A_0-\frac{A_1^2}{4A_2}} \tag{B.5}$$

$$r_x=\frac{2r_y}{\sqrt{-A_2}} \tag{B.6}$$

Similar expressions with x replaced by y and *vice versa* can be obtained by performing the summation in the other direction. Therefore, each summation gives us an estimate of r_x, r_y and one of the centre coordinates. If necessary, the errors in the derived gradients could be propagated through: this has not been done because, in practice, it is found that the circles fit exceedingly well, and any errors associated with the coordinate mismatch would be swamped by inherent uncertainty in the interferometer measurements

An example of a real h^2 vector, and the least-squares fit is shown in figure B.1.

Figure B.1 Fitting a circle: crosses are data points, solid line is fit



Bibliography

[1] H. W. BABCOCK. "The possibility of compensating astronomical imaging." *Publ. Astron. Soc. Pac.*, **65**, pp. 229, 1953.

[2] K. BÄR, B. FREISLEBEN, C. KOZLIK, and R. SCHMIEDL. "Adaptive optics for industrial CO_2 laser systems." *Lasers in Engineering*, **4**, pp. 233–242, 1995.

[3] G. VDOVIN. "Quick focusing of imaging optics using micromachined adaptive mirrors." *Opt. Comm.*, **140**, pp. 187–190, 1997.

[4] J. SPINHIRNE, D. ANAFI, R. FREEMAN, and H. GARCIA. "Intracavity adaptive optics. 1: Astigmatism correction performance." *Applied Optics*, **20**(6), pp. 976–984, 1981.

[5] S. TIMOSHENKO and S. WOINOWSKY-KRIEGER. *Theory of Plates and Shells*. McGraw-Hill, 1959.

[6] E. SZIKLAS and A. SIEGMAN. "Mode calculations in unstable resonators with flowing saturable gain. 2: Fast Fourier transform method." *Applied Optics*, **14**(8), pp. 1874–1889, 1975.

[7] J. GASKILL. *Linear Systems, Fourier Transforms, and Optics*. John Wiley and Sons, 1978.

[8] A. GARRARD and J. BURCH. *Introduction to Matrix Methods in Optics*. John Wiley and Sons, 1975.

[9] A. SIEGMAN. *Lasers*. University Science Books, 1986.

[10] M. BORN and E. WOLF. *Principles of Optics*. Pergamon Press, sixth (corrected) edition, 1980.

[11] R. K. TYSON. *Principles of Adaptive Optics*. Academic Press, second edition, 1997.

[12] D. L. FRIED. "Anisoplanatism in adaptive optics." *J. Opt. Soc. Am.*, **72**(1), pp. 52–61, 1982.

[13] F. RODDIER and C. RODDIER. "National Optical Astronomy Observations (NOAO) Infrared Optics Program II: modelling atmospheric effects in adaptive optics systems for astronomical telescopes." In *Advanced Technology Optical Telescopes III*, volume 628 of *Proc. SPIE*, pp. 298–304. 1986.

[14] R. A. HUMPHREYS, C. A. PRIMMERMAN, L. C. BRADLEY, and J. HERRMAN. "Atmospheric-turbulence measurements using a synthetic beacon in the mesospheric sodium layer." *Optics Letters*, **16**(18), pp. 1367–1369, 1991.

[15] T. O'MEARA. "The multidither principle in adaptive optics." *J. Opt. Soc. Am.*, **67**(3), pp. 306–315, 1977.

[16] R. A. MULLER and A. BUFFINGTON. "Real-time correction of atmospherically degraded telescope images through image sharpening." *J. Opt. Soc. Am.*, **64**(9), pp. 1200–1210, 1974.

[17] M. A. VORONTSOV *et al.* "Image quality criteria for an adaptive imaging system based on statistical analysis of the speckle field." *J. Opt. Soc. Am. A*, **13**(7), pp. 1456–1466, 1996.

[18] C.-Y. HAN *et al.* "Reshaping collimated laser bemas with Gaussian profile to uniform profiles." *Applied Optics*, **22**, pp. 3644–3647, 1983.

[19] K. NEMOTO *et al.* "Transformation of a laser beam intensity profile by a deformable mirror." *Optics Letters*, **21**, pp. 168–170, 1996.

[20] R. VAN NESTE, C. PARÉ, R. LACHANCE, and P.-A. BÉLANGER. "Graded-phase mirror resonator with a super-gaussian output in a CW-CO_2 laser." *IEEE J. Quantum Electron.*, **30**, pp. 2663–2669, 1994.

[21] T. CHEREZOVA, L. N. KAPSTOV, and A. V. KUDRYASHOV. "CW industrial rod YAG:Nd^{3+} laser with an intracavity active bimorph mirror." *Applied Optics*, **35**(15), pp. 2554–2561, 1996.

[22] M. E. FURBER and D. JORDAN. "Optimal design of wavefront sensors for adaptive optical systems: part 1, controllability and observability analysis." *Optical Engineering*, **36**, pp. 1843–1855, 1997.

[23] M. D. OLIKER. "Sensing waffle in the Fried geometry." In D. Bonaccini and R. K. Tyson, editors, *Adaptive Optical System Technologies*, volume 3353 of *Proc. SPIE*, pp. 964–971. 1998.

[24] R. TYSON. "Adaptive optics system performance approximations for atmospheric turbulence correction." *Optical Engineering*, **29**, pp. 1165–1173, 1990.

[25] E. STEINHAUS and S. LIPSON. "Bimorph piezoelectric flexible mirror." *J. Opt. Soc. Am.*, **69**, pp. 478–481, 1978.

[26] S. A. KOKOROWSKI. "Analysis of adaptive optical elements made from piezoelectric bimorphs." *J. Opt. Soc. Am.*, **69**, pp. 181–187, 1979.

[27] F. DRUON *et al.* "Wave-front correction of femtosecond terawatt lasers by deformable mirrors." *Optics Letters*, **23**(13), pp. 1043–1045, 1998.

[28] G. D. LOVE. "Liquid-crystal phase modulator for unpolarized light." *Applied Optics*, **32**(13), pp. 2222–2223, 1993.

[29] G. D. LOVE. "Wavefront correction and production of Zernike modes with a liquid-crystal spatial light modulator." *Applied Optics*, **36**, pp. 1517–1524, 1997.

[30] J. GOURLAY *et al.* "A real-time closed-loop liquid crystal adaptive optics system: first results." *Opt. Comm.*, **137**, pp. 17–21, 1997.

[31] M. A. A. NEIL, M. J. BOOTH, and T. WILSON. "Dynamic wave-front generation for the characterization and testing of optical systems." *Optics Letters*, **23**(23), pp. 1849–1851, 1998.

[32] A. F. NAUMOV and G. VDOVIN. "Multichannel liquid-crystal-based wave-front corrector with modal influence functions." *Optics Letters*, **23**, pp. 1550–1552, 1998.

[33] D. MALACARA, editor. *Optical Shop Testing*. JohnWiley and Sons, Inc., second edition, 1992.

[34] T. W. NICHOLLS. *Atmospheric Turbulence Measurements Relevant to Adaptive Optics*. Ph.D. thesis, Imperial College, 1996.

[35] D. L. FRIED. "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements." *J. Opt. Soc. Am.*, **67**(3), pp. 370–375, 1976.

[36] R. TYSON. *Principles of Adaptive Optics*. Academic Press, 1991.

[37] J. M. RUGGIU, C. J. SOLOMON, and G. LOOS. "Gram-Charlier matched filter for Shack-Hartmann sensing at low light levels." *Optics Letters*, **23**(4), pp. 235–237, 1998.

[38] J. PFUND, N. LINDLEIN, and J. SCHWIDER. "Dynamic range expansion of a Shack-Hartmann sensor by use of a modified unwrapping algorithm." *Optics Letters*, **23**(13), pp. 995–997, 1998.

[39] I. MUNRO, M. RIDOU, C. PATERSON, and J. C. DAINTY. "Locado system." The Locado system is a breadboard AO system built at Imperial College, London.

[40] F. RODDIER. "Curvature sensing and compensation: A new concept in adaptive optics." *Applied Optics*, **27**(7), pp. 1223–1225, 1988.

[41] C. SCHWARTZ, E. RIBAK, and S. LIPSON. "Bimorph adaptive mirrors and curvature sensing." *J. Opt. Soc. Am. A*, **11**(2), pp. 895–902, 1994.

[42] C. RODDIER and F. RODDIER. "Wave-front reconstruction from defocused images and the testing of ground-based optical telescopes." *J. Opt. Soc. Am. A*, **10**(11), pp. 2277–2287, 1993.

[43] D. G. KOCHER. "Automated Foucalt test for focus sensing." *Applied Optics*, **22**(12), pp. 1887–1892, 1983.

[44] J. PRIMOT and L. SOGNO. "Achromatic three-wave (or more) lateral shearing interferometer." *J. Opt. Soc. Am. A*, **12**(12), pp. 2679–2685, 1995.

[45] J. FIENUP. "Phase retrieval algorithms: a comparison." *Applied Optics*, **21**, pp. 2758–2769, 1982.

[46] P. KUTTNER. "Optics for data storage." In Marshall [83], chapter 7.

[47] A. G. SAFRONOV. "Controllable bimorph optics and the principles according to which they can be further developed." *J. Opt. Technol.*, **65**, pp. 4–12, 1996.

[48] A. KUDRYASHOV and V. SAMARKIN. "Control of high power CO_2 laser beam by adaptive optical elements." *Opt. Comm.*, **118**, pp. 317–322, 1995.

[49] S. LIPSON, E. RIBAK, and C. SCHWARTZ. "Bimorph deformable mirror design." In *Adaptive Optics in Astronomy*, volume 2201 of *Proc. SPIE*, pp. 703–714. 1994.

[50] F. F. FORBES. "Bimorph PZT active mirror." In *Active Telescope Systems*, volume 1114 of *Proc. SPIE*, pp. 146–151. 1989.

[51] O. IKEDA and T. SATO. "Comparison of deformability between multilayered deformable mirrors with a monomorph or a bimorph actuator." *Applied Optics*, **25**, pp. 4591–4597, 1986.

[52] A. KUDRYASHOV and V. SHMALHAUSEN. "Semipassive bimorph flexible mirrors for atmospheric adaptive optics applications." *Optical Engineering*, **35**, pp. 3064–3073, 1996.

[53] J. E. GRAVES, M. NORTHCOTT, F. RODDIER, C. RODDIER, and L. CLOSE. "First light for Hokupa'a 36 element curvature AO system at UH." In D. Bonaccini and R. K. Tyson, editors, *Adaptive Optical System Technologies*, volume 3353 of *Proc. SPIE*, pp. 34–43. 1998.

[54] D. LEE, B. ELLERBROEK, and J. CHRISTOU. "First results for the Starfire Optical Range 3.5m telescope adaptive optics system: point spread functions and tracking performance." In D. Bonaccini and R. K. Tyson, editors, *Adaptive Optical System Technologies*, volume 3353 of *Proc. SPIE*, pp. 1080–1091. 1998.

[55] J. M. SPINHIRNE *et al.* "The Starfire Optical Range 3.5m telescope adaptive optical system." In D. Bonaccini and R. K. Tyson, editors, *Adaptive Optical System Technologies*, volume 3353 of *Proc. SPIE*, pp. 22–33. 1998.

[56] K. OUGHSTUN. "Theory of intracavity adaptive mode control." In *Adaptive Optics Systems and Technology*, volume 365 of *Proc. SPIE*, pp. 54–65. 1982.

[57] D. ANAFI, J. SPINHIRNE, R. FREEMAN, and K. OUGHSTUN. "Intracavity adaptive optics. 2: Tilt correction performance." *Applied Optics*, **20**(11), pp. 1926–1932, 1981.

[58] J. M. SPINHIRNE, D. ANAFI, and R. H. FREEMAN. "Intracavity adaptive optics. 3: Hsuria performance." *Applied Optics*, **21**, pp. 3969–3982, 1982.

[59] K. OUGHSTUN, J. SPINHIRNE, and D. ANAFI. "Intracavity adaptive optics. 4: Comparison of theory and experiment." *Applied Optics*, **23**(10), pp. 1529–1540, 1984.

[60] L. DIAZ SANTANA HARO and J. C. DAINTY. "Single-pass measurements of the wave-front aberrations of the human eye by use of retinal lipofuscin autofluorescence." *Optics Letters*, **24**(1), pp. 61–63, 1999.

[61] J. LIANG, D. R. WILLIAMS, and D. T. MILLER. "Supernormal vision and high-resolution retinal imaging through adaptive optics." *J. Opt. Soc. Am. A*, **14**(11), pp. 2884–2892, 1997.

[62] F. RODDIER. "The effects of atmospheric turbulence in optical astronomy." In E. Wolf, editor, *Progress in Optics*, volume XIX. North-Holland, 1981.

[63] J. W. GOODMAN. *Statistical Optics*. John Wiley and Sons, 1985.

[64] D. M. WINKER. "Effect of a finte outer scale on the zernike decomposition of atmospheric turbulence." *J. Opt. Soc. Am. A*, **8**, pp. 1568–1573, 1991.

[65] N. RODDIER. "Atmospheric wavefront simulation using Zernike polynomials." *Optical Engineering*, **29**, pp. 1174–1180, 1990.

[66] R. NOLL. "Zernike polynomials and atmospheric turbulence." *J. Opt. Soc. Am.*, **66**, pp. 207–211, 1976.

[67] J. F. NYE. *Physical properties of crystals*. Oxford, 1957.

[68] R. C. WETHERHOLD. *Composites Engineering Handbook*, chapter 5. Marcel Dekker, Inc., 1997.

[69] R. M. JONES. *Mechanics of composite materials*. McGraw-Hill, 1975.

[70] J. M. WHITNEY. *Structural Analysis of Laminated Anisotropic Plates*. Technomic Publishing Company, Inc., 1987.

[71] M. R. STEEL, F. HARRISON, and P. G. HARPER. "The piezoelectric bimorph: An experimental and theoretical study of its quasistatic response." *J. Phys. D: Appl. Phys.*, **11**, pp. 979–989, 1978.

[72] W. C. YOUNG. *Roark's formulas for stress and strain*. McGraw-Hill, sixth edition, 1989.

[73] G. W. C. KAYE and T. H. LABY. *Tables of Physical and Chemical Constants*. Longmans, thirteenth edition, 1966.

[74] P. MEHTA. "Moment actuator influence functions for flat circular deformable mirrors." *Optical Engineering*, **29**, pp. 1213–1222, 1990.

[75] F. RIGAUT *et al.* "The Canada-France-Hawaii adaptive optics bonnette II: simulations and control." In *Adaptive Optics in Astronomy*, volume 2201 of *Proc. SPIE*, pp. 149–160. 1994.

[76] A. KORYABIN and J. DAINTY. "Low order adaptive optical systems." Internal report at Imperial College.

[77] N. BANSAL and R. DOREMUS. *Handbook of Glass Properties*. Academic Press, Inc., 1986.

[78] M. CHANG *et al.* "Hysteresis correction of a piezoelectrically actuated segmented mirror." In *Adaptive Optical System Technologies*, volume 3353 of *Proc. SPIE*, pp. 864–871. 1998.

[79] J. TOPPING. *Errors of observation and their treatment*. Chapman and Hall, third edition, 1962.

[80] L. BEISER. *Laser Scanning Notebook*. SPIE Optical Engineering Press, 1992.

[81] S. WOLFRAM. *The Mathematica Book*. Cambridge University Press, third edition, 1996.

[82] W. H. PRESS *et al.* *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.

[83] G. F. MARSHALL. *Laser Beam Scanning*. Marcel Dekker, Inc., 1985.