## *DAFI: a single mode optical fibre interferometer for astronomy*

Chang, Mark Peter Joe-Ling

# DAFI:
# A Single Mode Optical Fibre Interferometer for Astronomy

Mark Peter Joe–Ling Chang

M.Sc., D.I.C.,

B.Sc., A.R.C.S.



**Department of Physics**

A thesis submitted to the University of Durham
in accordance with the regulations for
admittance to the Degree of Doctor of Philosophy.

# Symbols

| | | |
|---|---|---|
| **Electromagnetic** | $c$ | Speed of light |
| | $\lambda$ | Free space wavelength |
| | $k = \frac{2\pi}{\lambda}$ | Free space wavenumber |
| | $\omega = kc$ | Angular frequency |
| | $\nu$ | Frequency |
| | $\epsilon_0$ | Free space dielectric constant |
| | $\mu_0$ | Free space permeability |
| **Atmospheric** | $r_0$ | Coherence length scale/Fried parameter |
| | $\tau_0$ | Coherence time |
| | $\theta_0$ | Isoplanatic angle |
| **Interferometric** | $\gamma$ | Complex degree of spatial coherence |
| | $V = \|\gamma\|$ | Visibility amplitude |
| **Optical fibre waveguide** | $n_1$ | Core refractive index |
| | $n_2$ | Cladding refractive index |
| | $a$ | Core radius |
| | $\lambda_c$ | Single mode cutoff wavelength |
| | $\Delta$ | Refractive index profile height parameter |
| | $U$ | Transverse mode propagation parameter |
| | $V = ak\sqrt{n_1^2 - n_2^2}$ | Normalised frequency |
| | $W$ | Radial decay parameter |
| | $\beta$ | Propagation constant |
| | $b$ | Normalised propagation constant |
| | $\omega_0$ | Fundamental mode radius |
| | $\tan\Omega$ | Taper adiabaticity criterion |

# Chapter 1

# Seeing the Heavens clearly

Must get those old glasses of mine set right. *James Joyce: Ulysses*

## 1.1 The search for higher resolution

One of the first people to seriously attempt to see a more detailed picture of the night sky was Galileo. In his epic battle with opponents of the Copernican theory, he had to explain the fixed relative positions of the stars despite the motion of the Earth around the Sun. This awkward fact, he argued, could be accounted for if the stars were at extremely large distances as compared to the orbit of the Earth. In an effort to substantiate this, he attempted to measure the angular diameter of Vega by a parallax method. He suspended a fine silk cord vertically then, standing some distance away, measured the angle through which the cord obstructed his view of the star. He concluded from this experiment that Vega's diameter was less than $5''$. While we now know that this result is far too large, it was nevertheless a significant advance upon the then currently accepted value of $\sim 2'$.

His next contribution to astronomy was far more important. Learning of the invention of the telescope in Holland in 1608, he built himself one with the specific intention of studying the sky. He used it to great effect, being the first person to see the lunar mountains, Jovian satellites and of course a myriad other stars hidden from the unaided human eye.

If $\theta$ is the smallest angular detail which can be measured by an imaging instrument, then elementary theory says

$$\theta \approx \lambda/D \tag{1.1}$$

where $D$ is the physical aperture of the instrument and $\lambda$ is the wavelength of radiation. So, while the naked eye can discern only $\sim 40''$ ($D = 3$mm), a small 12cm telescope reduces this to $1''$.

The obvious next step to improving the angular resolving power is to construct larger telescopes. However the resolution gains hoped for were not achieved for well over 200 years. Despite the increasing telescope diameter (and so its theoretical resolution) over this time, the turbulence of the Earth's atmosphere could not be bypassed.

So, regardless of the telescope diameter, the image of a star[1] in the visible appears to be a roughly circular blur with a typical angular size of between 1–2$''$. In other words, once the instrument aperture has reached a diameter of about 12cm, any further increase in size is not rewarded by much improvement in the detail which can be seen.

## 1.2 The Fringe benefits of Interferometry

In 1803 Thomas Young demonstrated that wavefronts from a distant (point) source, when divided by two slits in an opaque screen, could be recombined to form a system of bright and dark bands — "Young's fringes". The interesting property of these fringes is that they can be used as a measure of the *spatial coherence* of the electromagnetic waves at the two slits. It is this spatial coherence which carries the information about the angular distribution of brightness over the source.

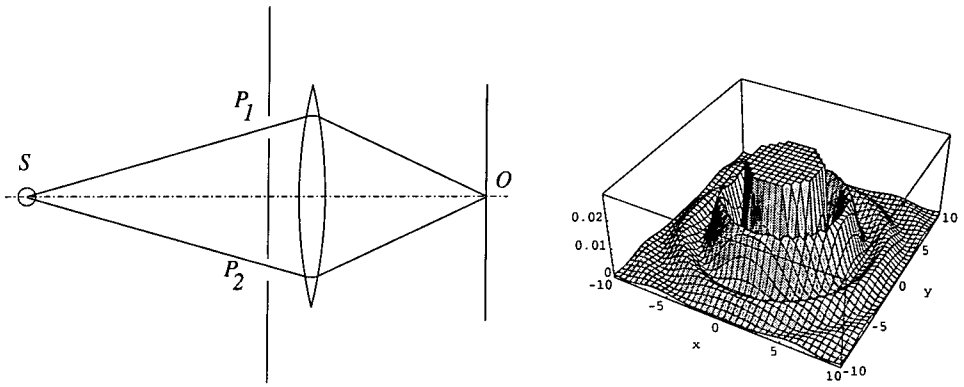



**Figure 1.1. Left:** Young's fringes produced by a point source with quasi–monochromatic radiation. $S$ represents the point source, $P_1$ and $P_2$ the two apertures and $O$ the centre of the recombination plane where the interference is observed. **Right:** Plot of the normalised visibility envelope, centred on $O$. The central disc has been truncated in intensity (vertical axis).

### 1.2.1 Interference fringes as a Spatial Coherence Measure

It is well known that the intensity function $I_O$ of light at a point $O$ in the recombination plane shown in Figure 1.1 is given by

$$I_O = I_1 + I_2 + 2\sqrt{I_1 I_2}\Re[\gamma(\tau)] \tag{1.2}$$

where $I_1$ and $I_2$ are the intensities arriving at $O$ from the respective slit and $\Re[\gamma(\tau)]$ is the (real part of the) *complex degree of spatial coherence* of the light at the two slits.

$$\gamma(\tau) \triangleq \frac{< \mathcal{V}_1^*(t)\mathcal{V}_2(t+\tau) >}{< \mathcal{V}_1^*(t)\mathcal{V}_1(t) >< \mathcal{V}_2^*(t)\mathcal{V}_2(t) >} \tag{1.3}$$

where $\mathcal{V}_1(t)$, $\mathcal{V}_2(t)$ are complex analytic signals[2] representing the incident radiation on the two slits and $\tau$ is the time delay between the radiation from each

[1]The most common bright objects in the sky are stars, and the angular extent of even the brightest (Sirius) is only $0.0068''$.

[2]In the scalar theory of light, the electromagnetic field is described fully by a *real* function of space and time $\mathcal{V}^{(r)}(\vec{x},t)$. Although purely real, it is mathematically more convenient to represent the field by a complex function $\mathcal{V}(\vec{x},t)$ called the *analytic signal*. This is defined in the following way:

$$\mathcal{V}(\vec{x},t) = \int_0^\infty v(\vec{x},\nu)\exp(-2\pi i\nu t)d\nu \tag{1.4}$$

where $v(\vec{x},\nu)$ is the *generalised* Fourier transform of $\mathcal{V}^{(r)}(\vec{x},t)$. Clearly, $\mathcal{V}(\vec{x},t)$ contains only positive frequencies — hence it is an analytic function. This also implies that the real and imaginary parts of $\mathcal{V}$ are Hilbert transform pairs.

slit reaching point $O$.

Following Michelson, define the fringe contrast or visibility as

$$V \triangleq \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{1.5}$$

then from Equation (1.3)

$$V = |\gamma(\tau)| \tag{1.6}$$

So the visibility of the fringes is directly proportional to the modulus of the spatial coherence of the light at the two slits.

The position of the fringes — their displacement relative to the centre of the system — is a measure of the phase angle of the complex degree of spatial coherence.

Therefore, at least in principle, it is possible to measure $\gamma$ for radiation arriving at two spatially separated points by considering the visibility and the position of the fringes formed by these two points.

If the spacing between the slits is now varied, and the variation of the spatial coherence as a function of this spacing is measured, then by the van Cittert–Zernike theorem[3], the angular distribution of brightness over the source can be found. In the far field (always the case for astronomical sources) this turns out to be the Fourier transform of the source brightness distribution. Thus if we can map the complex degree of coherence between points in a plane normal to a source, we can use that information to map the source.

Michelson put forward the idea that the resolving power of a large single aperture telescope can be synthesised by two small subapertures. If these are moved apart in a straight line, then the Fourier transform of the brightness distribution across the source in one dimension can be measured. Effectively we have a "map" of the source reduced to a rectangular strip where the brightness of any point on this strip is the integrated brightness along a line normal to the

[3]This theorem states that the correlations of vibrations at a fixed point $Q_2$ and a variable point $Q_1$ in a plane illuminated by an extended quasi–monochromatic source, is equal to the normalised complex amplitude at the corresponding point $Q_1$ in a certain diffraction pattern centered on $Q_2$ [1].

strip.

By moving the subapertures to different baselines with different separations and orientations, it is possible to build up a two–dimensional Fourier transform of the source. The two–dimensional brightness distribution can then be found through the Fourier inversion of data collected by this *aperture synthesis* technique.

### 1.2.2 Image forming Optical Interferometers

One of the main difficulties of Michelson's original stellar interferometer is that, unless the optical paths from the two small subapertures to the focal plane are very nearly equal, the fringe visibility is reduced. Consider the Wiener–Kintchine theorem[4],

$$V(\tau) = \int_0^\infty G(\nu) \exp(-2\pi i \nu \tau) d\nu / \int_0^\infty G(\nu) d\nu \tag{1.7}$$

for the fringe visibility of an unresolved source $V(\tau)$ when the relative time delay between the two paths through the instrument is $\tau$. $G(\nu)$ is the power spectrum of the light. In the trivial case where the optical passband is uniform over a width $\Delta\nu$,

$$|V(\tau)| = \sin \pi \Delta\nu\tau / \pi \Delta\nu\tau \tag{1.8}$$

It follows that for a bandwidth $\Delta\lambda = 100$nm at $\lambda = 500$nm, a path difference of $1\lambda$ will reduce $V(\tau)$ by 10%. This places an extremely stringent requirement on the mechanical stability and pointing accuracy of a large instrument.

A second problem is the presence of atmospheric perturbations (or *seeing*). Turbulence in the air creates inhomogeneities in the temperature and humidity which introduce random variations in the amplitude and phase of wavefronts from an extraterrestrial source. These changes are usually uncorrelated over distances greater than about 12cm. Their effect on an interferometer is to move the fringes about rapidly in the focal plane and to change their shape and

[4]This states that the autocorrelation function of a stationary variable is the Fourier transform of its power spectrum.

contrast on timescales dependent on the scale of atmospheric irregularities and on the wind speed.

Michelson's single vector baseline optical interferometer could not produce a true image of the sky because it yielded only the modulus of the visibility function and not its phase. The visibility phase is difficult to extract from a single baseline at optical frequencies due to a combination of atmospheric and instrumental effects resulting in quantum noise dominated signals. Multiple vector baselines in a synthetic aperture can circumvent the classical phase problem[5] by providing a relative phase datum point or *closure phase*. It is illustrated schematically in Figure 1.2. Essentially, this observable is constructed from the vector sum of the phases around the loop $C_{ABC} = \vec{\phi}_{AB} + \vec{\phi}_{BC} + \vec{\phi}_{CA}$. Provided that

1. the spatial scale of the atmospheric fluctuations does not strongly affect each subaperture $A$, $B$ and $C$ (for high fringe visibility across a single baseline),

2. the error in each baseline $\vec{AB}$, $\vec{BC}$ and $\vec{CA}$ is negligibly small (if they are not, the visibility phases will not "close") and

3. all the measurements are made instantaneously,

the closure phase will be "invariant" to subaperture errors, and will be dependent on the source alone. Using this quantity to calibrate the other data, it is possible to reconstruct images independent of atmospheric phase perturbations.

The technique of phase closure has been used by Baldwin *et al.* [2] to produce a diffraction limited image of Capella with a resolution of about $0.05''$ in the visible. Its principle is important to this study and will be discussed in more depth later.

[5]That is, how far is it possible to reconstruct an image uniquely from the modulus of the Fourier transform alone.

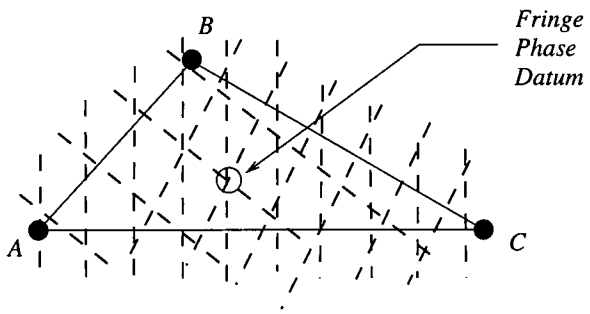



**Figure 1.2.** Idealised fringe signals produced by a multiple vector baseline $\vec{AB}$, $\vec{BC}$ and $\vec{CA}$. The peaks of the interference fringes are denoted by the dashed lines. A datum point can be assigned to the phases of the multiple vector baseline, but not the single baseline case.

## 1.3 Factors affecting Stellar Interferometers

The formation of interference fringes in the optical domain is affected by many different physical effects which degrade the fringe contrast. It is therefore of paramount importance to identify and analyse these problem areas. This is necessary to verify the feasibility of astronomical science objectives based on interferometry and also to identify the areas where the calibration of systematic errors is the most critical.

As has already been mentioned, the visibility is the prime observable of any interferometer. For the perfect instrument, it would be the direct measure of the complex degree of coherence of the incident beams. Unfortunately, this is not true in the practical case as the ideal visibility is always modified such that

$$V = T_i T_a |\gamma(\tau)| \tag{1.9}$$

where $T_i$ and $T_a$ relate to the instrumental and seeing induced transfer functions respectively.

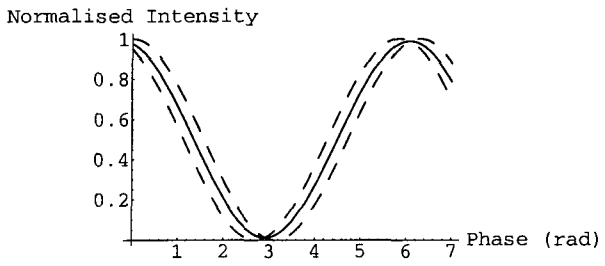



**Figure 1.3.** The effect of phase fluctuations over time (piston errors) on the visibility. The dashed lines represent instantaneous fringe postions at times $\tau$ and $\tau + d\tau$. The resultant signal is denoted by the solid curve.

### 1.3.1 Piston Errors ($T_a$)

The generic expression for the interference signal obtained by recombining two monochromatic beams of equal intensity $I_0$ from a perfect instrument is

$$I(\tau) = 2I_0 \left[1 + \gamma(\tau) \cos\left(k\tau + \phi\right)\right] \tag{1.10}$$

where $\tau$ represents the modulation delay and $\phi$ the phase difference between the two beams. $\tau$ can be imposed either temporally by modulating the optical path length of one of the beams, or spatially by using a 2–D array detector. In all cases the signal is measured in a finite exposure time $T$. If $\phi$ is not stable during this time, the resultant signal is a superposition of phase shifted instantaneous fringe patterns as illustrated in Figure 1.3. This time averaged signal can be expressed as

$$< I(\tau) >= \frac{1}{T} \int_{T_0}^{T_0+T} I(\tau) dt \tag{1.11}$$

The variation of $\phi$ occurs because the refractive index of the Earth's atmosphere varies over space and time (and wavelength). For our purposes, it is convenient to express the refractive index dependencies as

$$n(\vec{r}, t, \lambda) = n_0(\vec{r}, t, \lambda) + n_1(\vec{r}, t, \lambda) \tag{1.12}$$

where $n_0$ is the deterministic (non–random) portion of $n$ whereas $n_1$ denotes the random fluctuations about a mean value $< n >= n_0 \approx 1$. The deterministic changes in $n$ are generally slowly varying and macroscopic. We can approximate the physical situation by neglecting the slow temporal variation of $n_0$.

The random fluctuations $n_1$ arise from the presence of turbulence in the atmosphere. The wavelength dependence of these random fluctuations can usually be ignored, allowing us to write

$$n(\vec{r}, t, \lambda) = n_0(\vec{r}, \lambda) + n_1(\vec{r}, t) \tag{1.13}$$

We note that the typical values of $n_1$ are several orders of magnitude smaller than unity [3].

Since $n_1$ is small, we assume $\phi$ to be a Gaussian random variable. It then follows that

$$T_a =< e^{\phi} >= \exp\left(\frac{\sigma_\phi^2(T)}{2}\right) \tag{1.14}$$

with the variance (or piston errors) during one exposure time period being represented by $\sigma_\phi^2(T)$. This variance[6] can be computed for any $T$ from the power spectral density of the phase fluctuations $S_\phi$ by

$$\sigma_\phi^2(T) = \int_0^\infty S_\phi(\nu)\left[1 - \left(\frac{\sin(\pi T\nu)}{\pi T\nu}\right)^2\right] d\nu \tag{1.15}$$

where $\nu$ is the frequency of the light. For further discussion see, for example, Goodman [4].

### 1.3.2 Wavefront Errors ($T_i$)

The effect of wavefront errors is most easily understood by considering a pupil plane detection scheme. In the absence of such errors, the ideal visibility function would be seen in the recombination pupil. With wavefront errors, the normalised fringe contrast is lowered from unity since the beams are no longer

[6] $\sigma_\phi^2(T)$ can also be interpreted as the autocorrelation of phase fluctuations, assumed to be statistically homogeneous (so wide sense stationary) and isotropic (thus symmetric over all space).

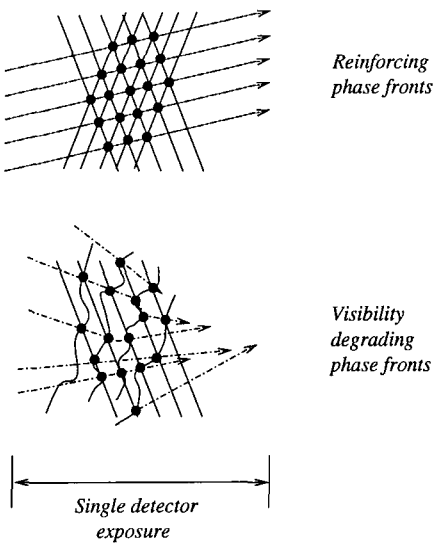



**Figure 1.4.** The effect of wavefronts with phase errors from one subaperture interfering with a flat wavefront, as compared to the case of no wavefront errors. The wavefronts with error are not spatially coherent, so the fringe visibility over one detector integration period is reduced.

spatially coherent. Figure 1.4 contrasts the ideal situation with a case where wavefront errors have occurred.

## 1.4 Attaining the diffraction limit

### 1.4.1 via Interferometry

The angular resolution of long exposure optical images taken with large telescopes, even at the best sites, is about 10–50 times worse than the theoretical diffraction limit. However, in 1986, Baldwin *et al.* demonstrated for the first time that the technique of *non–redundant masking* (NRM) could produce diffraction limited images in the visible.

The NRM method entailed placing an aperture mask with three or more holes, arranged in a non–redundant fashion, in the telescope pupil. The beams from the subapertures were focused onto a detector (e.g. a CCD) and the fringes recorded with a series of different exposure times.

The first diffraction limited images to be produced by this method were reported in 1987 by Haniff *et al.* (of the same group) [5]. Images of $\lambda$ Peg

and the binary star $\phi$ And were reconstructed from a four hole mask (i.e. 6 non–redundant baselines) on the 2.5m Isaac Newton Telescope, using the Image Photon Counting System (IPCS). The fringes were observed at $\lambda = 512$nm, $\Delta\lambda = 12$nm, with 16ms exposure times. The seeing was reported to be $1.2''$. The summed autocorrelation function of several thousand short exposure images were computed and Fourier transformed to obtain the centro–symmetric power spectrum. The spatial frequencies of each baseline could then be found from the peaks either side of the origin. Consequently, for each frame, a Fourier transform was performed at these spatial frequencies to measure the amplitudes and phases of the six fringe patterns present. The images were reconstructed by a maximum entropy method [6] from the measured visibility amplitudes and closure phases. An extremely good agreement with previously accepted values for the separation of $\phi$ And was found ($0.460''$).

The limitations of NRM arise from the small effective subaperture sizes (5.6cm for Haniff *et al.*) and the narrow bandwidths used. Larger subaperture sizes, which would be beneficial in terms of collecting area, are seriously affected by wavefront errors. Fortunately it is possible to remove these high frequency errors by spatial filtering with optical fibres as we will see in the next section. Wider bandwidths too allow more photons to be detected, but decrease the coherence length. The shorter the coherence length, the more susceptible the system is to piston errors, increasing the requirement for accurate path length compensation.

### 1.4.2 via Single Mode Fibre Interferometry

We have seen that large area apertures are affected by atmosphere–induced wavefront errors. It is therefore necessary to remove these errors before we can take advantage of the better collection efficiency of these large apertures for interferometry. Spatially filtering the wavefronts is the most immediate method available to remove the high frequency "noise".

In single mode (SM) dielectric waveguides, the normalised radiation output profile is independent of the input wavefront and the phase is constant across

the guided beam. On the other hand, the intensity of the guided radiation depends on the field amplitude distribution in the focal plane of the input coupling optics.

So, SM fibres essentially force the transverse coherence of the guided radiation into a fixed form. The wavefront phase aberrations are translated into intensity fluctuations of the field coupled into the fibres. Unlike wavefront perturbations, intensity fluctuations can be easily monitored and used to correct each interferogram individually if desired.

In summary, what SM waveguides offer, apart from simplifying beam transport, is the ability to spatially filter atmospherically perturbed wavefronts thereby *phasing* each interferometer subaperture.

Since optical fibre waveguides are chromatically dispersive, and dispersion of fibres is a problem for interferometry, beam transfer through metres of glass for any interferometer may appear bizarre. Fibres (primarily fused silica) are regularly fabricated for the communications industry with a nominally zero dispersion at a single wavelength. Concepts exist for flattening the dispersion (of fluoride fibres) over a relatively wide bandwidth [7]. However, what is important is the dispersion *difference* between interferometer arms, so if this can be minimised or even eliminated, there is no paradox in employing SM fibres in interferometry.

It is instructive to look at two examples of SM fibre interferometric systems intended for astronomical applications. While they are separate telescope systems, it is quite simple to envisage their individual inputs as subpupils of a single large telescope.

### FLUOR

The **F**iber **L**ink **U**nit for **O**ptical **R**ecombination (FLUOR) described by Coudé du Foresto [8] is a two telescope stellar interferometer using non–polarisation preserving fluoride fibres for mid–infrared ($\lambda = 2 - 4\mu$m) observing. Originally located at Kitt Peak Observatory, it has since moved to the Harvard–Smithsonian Center for Astrophysics' IOTA facility on Mount Hopkins, Ari-

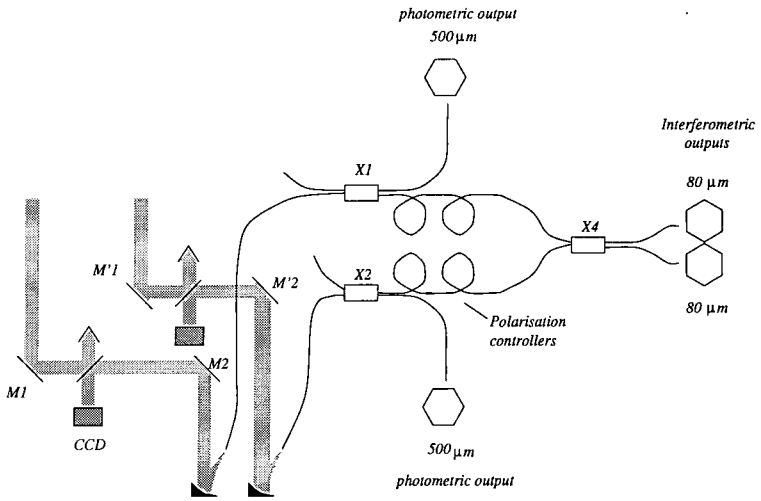



**Figure 1.5.** Schematic of the FLUOR interferometer. $M1, M'1$ and $M2, M'2$ are flat mirrors for beam alignment, $X1$ and $X2$ are pigtailed x–couplers of which only three pigtails are used. The beams from each arm are divided so that they can be monitored photometrically (at $500\mu$m). Polarisation control is achieved through bending birefringence in the arms before $X4$ where the beams are interfered.

zona.

As shown in Figure 1.5, FLUOR has two inputs and four outputs: two complementary interferometric signals and a photometric calibration signal for each telescope. The two infrared beams arriving from each collecting telescope (of IOTA) are incident on two fold mirrors $M1, M'1$ and $M2, M'2$. The beams are then injected into the waveguides by off–axis paraboloids. Between $M1, M'1$ and $M2, M'2$ are Michelson subsystems for ensuring that the stellar image is formed on the fibre core. This is done by backlighting the waveguides and ensuring the star and fibre images overlap. IOTA has an adaptive tip–tilt system (whose feedback loop uses the visible part of the starlight) that compensates for the transverse atmospheric perturbations of the star's image across the field–of–view.

$X1$, $X2$, $X4$ and the fibre sections between them are refered to as the *triple coupler*. Between the directional couplers are two chromatic dispersion compensating fibre segments. The polarisation controllers, which will be described in more detail in Section 4.5, are made up of two loops of fibre that can be

twisted with respect to each other.

Fringes are acquired in a temporal "long scan" mode. The optical path difference is scanned around the equal delay point by a smooth delay line. The optical path modulation therefore becomes the algebraic sum of the diurnial motion of the source and the scanning speed of the overall delay line. No fringe tracking is provided by the interferometer, so the scanning speed needs to be fast enough to "freeze" the differential piston mode of the atmosphere. The fringe signal is obtained, at present, through separate InSb photodetectors at each output.

### The 5 telescope fibre imaging interferometer

Shaklan [9] describes a laboratory experiment in interferometric imaging using a synthetic aperture of 5 optical fibres. This is shown schematically in Figure 1.6. The five non–polarisation preserving fibres had an $LP_{11}$ cutoff at 590nm.

Each input fibre (A1 to C1) was fed by a single lens telescope, resulting in a measured 50% coupling efficiency (although this can clearly be improved). The fibre arms fed down to a first stage of fused biconical directional couplers (x–couplers A to C). These x–couplers had a 3dB split at 630nm, with insertion losses of $< 7\%$. Their splitting ratios changed by 0.24% $nm^{-1}$. Five of the six connections (indicated by the numbers 1 to 6 on the Figure) were fusion splices, with losses ranging from 15–30% (though it is claimed to be possible to reduce this to some 3% with better splicing). The sixth was a GTE Elastomeric splice, tuned to give $< 10\%$ loss. The second stage (x–couplers D to F) was where all the input beams were combined. Fringe detection was achieved by means of an uncooled EMI 9658 PMT (with S20 photocathodes) for each of 3 output fibres. The remaining fibres were used to couple light backwards through the system for alignment and cophasing purposes.

The system was illuminated by linearly polarised light to avoid the need for either full polarisation control (discussed in more depth in Section 4.4) or a second set of detectors.

The input fibres were epoxied to a lever arm, which was actuated by a

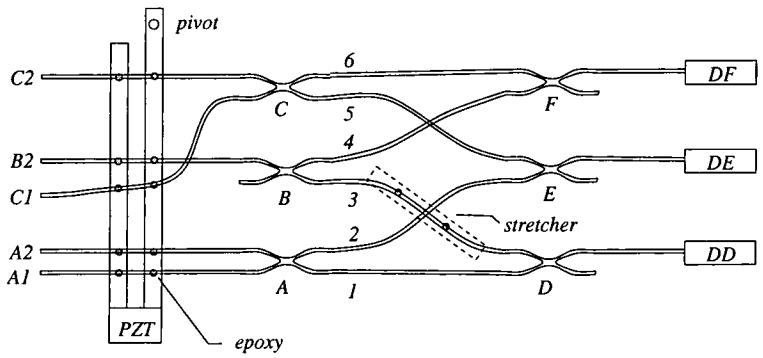



**Figure 1.6.** Schematic of the 5 telescope fibre imaging interferometer. A1 to C1: input single mode fibres from telescopes. PZT: Piezoelectric transducer. Pivot: Flexure hinge. A to F: 50/50 directional couplers. 1 to 6: "internal" single mode fibres. Stretcher: Translation stage and base. DD to DF: Photomultiplier tubes. Dots: Epoxy.

piezotransducer. This arrangement allowed a continuous scan temporal multiplexing technique to be employed for fringe detection, since the amount of stretch was proportional to the distance of the fibre from the pivot. The input fibres were spaced such that no fringe (temporal) frequency was repeated. The PMT output was digitised during the fringe scanning, and the amplitude and phase at the temporal frequency of particular pair of fibres was extracted from the signal's Fourier transform[7]. As with FLUOR, no fringe tracking is provided by the interferometer, so the scanning speed would need to be faster than the atmospheric piston variations in an on–sky system.

### Comments

One of the main points of both the aforementioned examples is their use of pigtailed x– (and y–) fibre couplers. While good in theory, in practice there are several objections to using them. The first is based upon component cost. While directional couplers are inexpensive and easily available at standard telecommunications wavelengths (1.3$\mu$m and 1.55$\mu$m), other wavelengths usually need to be custom made. Also they have limited useable bandwidths, although this is not as bad for non–polarisation preserving couplers ($\sim$ 200nm) as for the

[7]This method is analogous to the 1–D spatial multiplexing technique described later.

highly birefringent type ($\sim$ 100 to 150nm). Another cause for concern is their input coupling. While the maximum value is generally quite high (provided the pigtail endfaces are flat) the problem arises when joining (splicing) a fibre to the pigtail. Temporary splices, be they locking cams or merely grooved mounts, are extremely sensitive to dust (which is not a negligible issue at typical observing sites). It is not necessarily trivial to align the centres of the fibre cores either. Fusion splices are obviously better, being permanent joins, but require facilities that are not always available. Also combining more than two beams, as we have seen with the 5 telescope interferometer, increases the number of couplers and the complexity of the system substantially.

It is also to be noted that both instruments use temporal multiplexing to scan the interferograms arising from each baseline. Although this is entirely equivalent to 1–D spatial multiplexing in theory, the practical aspect is quite different. The simplest form of 1–D spatial multiplexing merely involves placing the fibre outputs side by side and recording the fringe signal from all the baselines on a multi–pixel detector. The problem again encountered when doing this is the error introduced by atmospheric piston (cf. FLUOR). One solution to this is to read out the detection system at frame rates faster than the atmospheric variation. However, this will result in low signal–to–noise fringes per frame. Another would be to have an active element upstream of the fibres which compensates for the piston as well as the tip–tilt variations. This would allow for longer exposure times per frame. A way of implementing the latter solution is described Section 1.6.

### Input Coupling to Optical Fibres

To maximise the coupling into optical fibres requires diffraction limited spots to be incident on the input fibre endface. Shaklan [9] calculated that the maximum coupling possible for the diffraction limited case (through a circular entrance pupil) into the fundamental mode of a SM fibre is $\sim$ 78% of the incident power. This value includes the losses due to Fresnel reflections at the optical fibre/air interface — if these could be removed by a suitable anti–reflection coating (not

a straightforward task), the efficiency would increase to $\sim 81\%$. In this diffraction limited case, the fall–off in coupling due to pointing errors approximately follows the profile of the waveguide fundamental mode which closely resembles a Gaussian curve.

In order to ease the alignment tolerances of the incident diffraction limited spots onto the optical fibre inputs, it is possible to use *tapered* SM fibres. These devices have a larger input aperture which taper slowly downwards to a normal single–mode waveguide. The details of such fibres are discussed further in Chapter 3.

In ground based stellar interferometry, there is also the issue of the turbulent atmosphere. Atmospheric perturbations perturb the incident wavefronts so that the instantaneous position of the diffraction limited spot may be offset from the optical axis. Clearly this tilt will reduce the amount of energy coupled into the waveguide. Higher order aberrations on the incident wavefronts also add to this reduction in coupling, to a lesser extent. Low order adaptive optics can introduce some degree of compensation for the wavefront tilts, as with the FLUOR instrument.

## 1.5 Enter: Adaptive Optics

Adaptive Optics (AO) technology provides a means for the real time optical correction of atmospheric aberrations over a small field–of–view, known as the isoplanatic angle. Typically this is between $2''$ and $10''$ in the visible and near infrared, depending on seeing conditions. The principle of AO is illustrated in Figure 1.7. The phase of a wavefront from a distant point source or *guide star* is distorted by the turbulent atmosphere of the Earth. These seeing induced phase variations[8] are approximately conjugated in real time by the action of a correcting element — for example a rapidly guided deformable mirror (DM). The beam is then divided so that some of it is incident on a wavefront sensing device (WFS), part of the mirror control loop. The remainder of the beam,

[8]In astronomical observations, the amplitude variations can usually be neglected [10, 11].

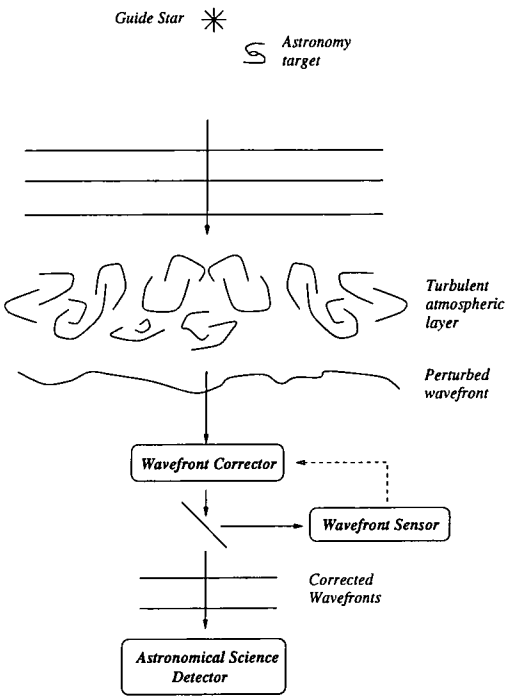



**Figure 1.7.** Schematic of an adaptive optics system.

referred to as the (astronomy) science arm, yields the data of interest.

There are a number of active devices that can conjugate the phase aberrations of incident wavefronts. In astronomy, these are generally macroscopic DMs which have either a continuous surface or are made up of a number of independent segments. The most common actuators of these DMs are piezoelectric transducers, which push/pull on a localised area of the DM.

There are also a number of different flavours of wavefront sensing devices. Current AO systems favour two types: the Shack–Hartmann and the Curvature Sensor. The former is, in essence, a classical Hartmann test which uses a lenslet array rather than a pupil mask. Therefore it senses the average tip–tilt (first moment) of the incident wavefront across a single lenslet. Ideally the Curvature Sensor measures the two near field extra–focal images simultaneously to give an estimate of (the second moment of) the wavefront.

If the DM is segmented, rather than a continuous facesheet, it is easy to see that each segment conjugates the tilt of part of the incoming wavefront. So by

matching a series of segments to SM fibres, we can construct a fibre interferometer with a maximum baseline that is the diameter of the DM. Additionally, if the segments are adjusted to match the wavefront piston (offset along the optical axis) the interferometric array will be *cophased*, in principle allowing longer integration times.

One may question the motivation for using aperture synthesis with an AO system, when adaptive optics itself is intended to remove seeing induced aberrations. To answer this, consider the characteristic scales of the atmosphere. Fried [10] derived the atmospheric coherence length scale (Fried parameter)

$$\begin{aligned} r_0 &= \left(0.423k^2 \sec(\zeta) \int_0^h C_n^2(h)dh\right)^{-3/5} \\ &\propto \lambda^{6/5} \end{aligned} \tag{1.16}$$

$k$ is the vacuum wavenumber, $\zeta$ the zenith distance and $C_n^2(h)$ is the altitude dependent structure constant of refractive index. $r_0$ varies with time and depends on geographical location[9].

The temporal behaviour of the disturbed wavefront is governed mainly by wind velocities at different heights in the atmosphere. Though in general this is highly complex, an idealised situation can be envisaged in which only one layer of the atmosphere causes these perturbations. Since the lifetime of a turbulent cell in this layer is longer than the time taken for the cell to cross a telescope aperture, the turbulence can be considered to be "frozen" [13]. Approximations to this situation have been observed by Gendron [14] and Racine & Ellerbroek [15]. The degree of temporal coherence can then be characterised by

$$\tau_0 = \frac{r_0}{v} \propto \lambda^{6/5} \tag{1.17}$$

where $v$ stands for the wind velocity of the layer.

The isoplanatic angle has already been mentioned. This is defined as the

[9]It is based on Kolmogorov's theory [12] which is only valid for very weak turbulence. This is because the Kolmogorov refractive index power spectrum is deduced from dimensional arguments rather than a physical picture.

radius in the field–of–view over which the wavefront perturbations can be considered identical (i.e. the change in Strehl ratio[10] is less than $1/e$). It is again given by Fried [16]

$$\begin{aligned} \theta_0 &= \left(2.905k^2 \sec^{8/3}(\zeta) \int_0^h C_n^2(h) h^{5/3} dh\right)^{-3/5} \\ &\propto \lambda^{6/5} \end{aligned} \tag{1.18}$$

It should be noted that this expression is strictly true only for the limiting (worst) case of an infinitely large aperture.

Since the parameters in Equations (1.16), (1.17) and (1.18) are all proportional to $\lambda^{6/5}$, it is clear that AO correction bandwidths (spatial and temporal) have to increase with decreasing wavelength[11]. With present technology, correction bandwidths are insufficient to produce diffraction limited images in the visible, so the full resolution (in a finite field–of–view) of large ground based telescopes is yet to be realised. We can see this by considering the number of spatial degrees of freedom $N$ required by an AO system to reach the V–band diffraction limit where we assume $r_0 = 10$cm. It is clear that $N \propto (D/r_0)^2 \propto \lambda^{-12/5}$ which means if

- $D = 4$m, $N \sim 1.6 \times 10^4$,
- $D = 8m$, $N \sim 6.4 \times 10^4$,
- $D = 10m$, $N \sim 10^5$.

The values of $N$ are approximately equal to the number of actuators required — a daunting task to implement.

The highest order AO systems that exist currently have correcting elements with several hundred degrees of freedom. This is enough to reach the diffraction

[10] This is the normalised measured Point Spread Function (PSF), where the normalisation is with respect to the diffraction limited PSF for the aperture.

[11] Another way to arrive at this conclusion is to realise that, to first order, the wavefront aberration introduced by a turbulent atmosphere is achromatic. Thus the number of microns of optical path retardance is constant, but the number of *wavelengths* in this optical path varies. It immediately follows that for shorter wavelengths, there are more waves of retardance so the distortion is greater.

limit (in the isoplanatic patch around the guide star) in the infrared region[12], at least for 4m class telescope apertures. So, to reach the diffraction limit at shorter wavelengths necessitates a different approach.

## 1.6 Fibre Interferometry and Adaptive Optics

Previously, the concept of 1–D spatial multiplexing to form the interferometric fringe signal was discussed. Here the plan for tip–tilt and piston compensator in the form of an AO system placed upstream of a fibre interferometer is described, leading to the **D**ilute **A**perture **F**ibre **I**nterferometer (DAFI).

Consider placing an AO system upstream of a SM fibre interferometer. Let the fibre subapertures of the interferometer be matched to areas of the deformable mirror's surface through coupling optics. When operating in a "closed loop" manner, it is straightforward for the AO system to conjugate the atmospherically induced tip–tilt variations on the incident wavefront. By doing so, assuming that the coupling optics introduce no static aberrations, the image of the astrophysical target can be kept registered with the fibre core.

Because of the nature of the WFS's mentioned in Section 1.5, the piston offsets of the mirror are not measured directly by the feedback loop. This would not be a problem for correcting elements which responded linearly to applied signals, since their positions would be entirely predictable from a given starting point. Unfortunately, the current generation of DM actuators suffer from hysteresis effects (see Chapter 2). So the conjugation of atmospheric piston errors with a standard DM is not as straightforward a task as for tip–tilt. A method of linearising the actuator hysteresis is described in the next Chapter. Given that the DM response can be linearised to a high degree, it is reasonable to consider cophasing the input subapertures of the fibre interferometer. To be able to cophase the inputs, a feedback signal is evidently required. This can be obtained by either reconstructing the wavefront from the WFS signal, which we discuss in more detail in Section 2.3 or by using a separate (DM) Figure Sensor

[12] where $r_0 \approx 40$cm.

to measure the mirror's surface profile. One example of this device is described in Section 2.2.3.

Since minimising light loss is a major constraint of this project, it is desirable to simplify the input coupling optics to the fibre as much as possible. To that end, we chose to use *adiabatically tapered single mode fibres*, whose larger cores (due to the tapering) slowed the fibre f–ratio sufficiently. This, in turn, relaxes their alignment tolerances considerably.

The arrangement of the fibres at the input plane should be easily variable, so that both non–redundant and redundant geometries are possible (see Chapters 6 and 7).

With the input coupling issues addressed, the next problem to be considered is the control of polarisation states through the optical fibre arms. If the beam polarisations from each fibre are allowed to vary randomly, the fringe signal with be degraded. Maximum visibility only occurs when the polarisations of interfering beams are equal. To understand how to solve this problem, the different methods of imposing control over the polarisation states in a fibre have to be examined. We will see that polarisation states depend on the waveguide's birefringence. The total birefringence is a sum of two quantities: the amount of birefringence intrinsic to the waveguide and the amount that is induced by external stressing.

The length mismatch of each optical fibre waveguide is also an issue. These length mismatches will induce chromatic dispersion, which is detrimental to the final fringe contrast for observations with wide bandwidths. Therefore, either a method for producing equal length optical fibres, or for actively controlling the optical path length in each fibre has to be devised. The former is the more economical option and so was pursued.

Lastly the beam recombining output plane has to be constructed. The 1–D spatial multiplexing technique has already been mentioned, in comparison with the temporal multiplexing used by the two example instruments. It is reasonable to imagine a series of optical fibre outputs laid side–by–side in a linear fashion to form an interference plane of straight (tilt) fringes. What

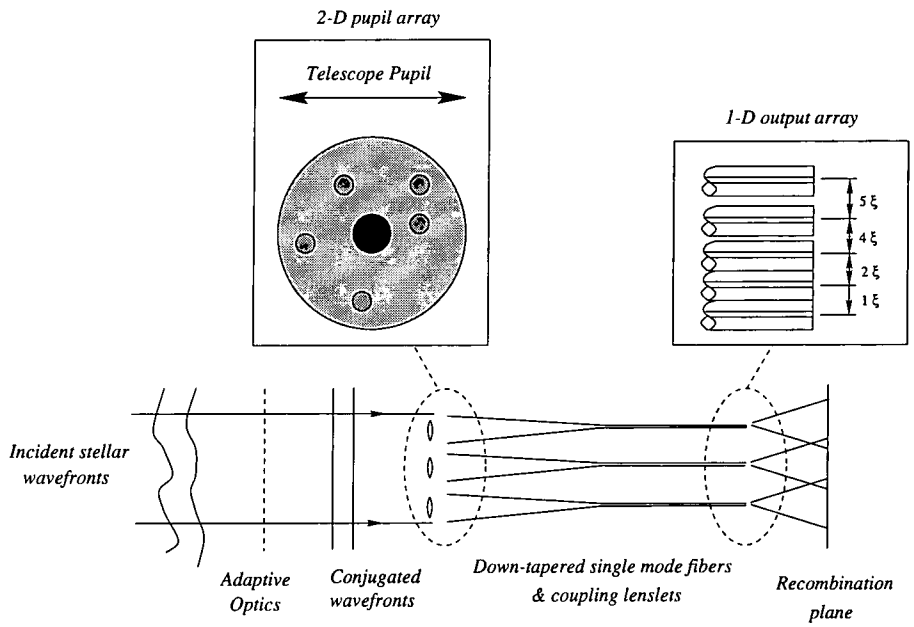



**Figure 1.8.** Schematic of DAFI.

is important at this stage is the ability to separate the fringe signals arising from each baseline. In the laboratory demonstration by Shaklan, the fringes were given a different temporal scanning frequency, which allowed for baseline identification. In the spatial case, if the fibres are arrange with non–redundant spacings, the baselines can be identified from their different spatial frequencies. Of course, the wider the fringes in the recombination volume, the easier they are to detect. Of the ways to increase the fringe widths, etching the fibre output diameters to place the cores closer together was the most straightforward.

The system is illustrated in Figure 1.8.

## 1.7 Thesis structure

DAFI was conceived and designed in Durham as a simple proof–of–concept system to test the feasibility of a multiple tapered SM fibre stellar interferometer. It is intended to be used with another Durham instrument, the **ELE**ment **C**ophasing and **TRA**cking (ELECTRA) AO system [17], which itself was undergoing development at the time.

The contents of this thesis are organised as follows:

- **Chapter 2** — ELECTRA is described and so is chiefly a system review. The AO system is intended to act as the feed for DAFI, conjugating the low order atmospheric perturbations. Additionally, having an active device upstream of DAFI relaxes many of the alignment tolerances of the interferometer's input. Results of measurements conducted on the performance of ELECTRA's deformable mirror are described. The impact of these results on the linearity of the device, and hence the implication for DAFI, is discussed.

- **Chapter 3** — The basics of single mode optical fibres are reviewed with relation to interferometry. Experimental tests of the behaviour of tapered SM fibres are reported. An attempt was made to manufacture tapered SM fibres with a high degree of intrinsic birefringence, the objective being to build some polarisation control into the fibres directly. The results of this prototype are described. An extension to the classical scalar field theory was developed as a consequence, to derive the design parameters of tapered high birefringence single mode fibres.

- **Chapter 4** — The requirements of single mode fibre interferometry is elucidated in this Chapter. The essence of interferometry is calibration, so the problems and solutions involved in calibrating the fringe visibility amplitude and phase are discussed. The concept of phase closure is introduced as a method of calibrating the visibility phase. The theory of the technique of polarisation state control, the motivation for equalising the optical path lengths and the reasoning behind the output plane geometry are explained; these being the steps to calibrating the visibility amplitude. The results of fibre cleaving tests are also given.

- **Chapter 5** — The experimental results for fibre etching, fibre length equalisation and output plane construction, resulting in a multi–fibre interferometer are described. The output of the interferometer is demonstrated pairwise on a point source.

- **Chapter 6** — The optical design of DAFI with ELECTRA is detailed. Measurements of DAFI's tapered fibres after processing and the implication for the final instrumental signal–to–noise are discussed.

- **Chapter 7** — Current science programmes with astronomical interferometers are reviewed. A scientific case appropriate to the DAFI + ELECTRA system is proposed as an example of what a fibre–based stellar interferometer can do.

- **Chapter 8** — We finish by summarizing the work done and looking towards the future.

# Chapter 2

# ELECTRA

But never will I cease from dirge and sore lament, while I look on the trembling rays of the bright stars ...

*Sophocles: Electra*

## 2.1 Introduction

ELECTRA is a sophisticated high order AO system, designed to operate at the GHRIL[1] Nasmyth platform of the 4.2m William Herschel Telescope (WHT) in La Palma, the Canary Islands. It was first commissioned during the period of 17th — 20th June 1997 (referred to as $E0$), during which time it closed loop in tip–tilt only mode[2]. Cophasing of the DM (through the linearisation of its actuator extensions) was achieved during the $E1$ observing run between 1st — 7th July 1998.

The purpose of this Chapter is to provide a brief overview of the key system components and discuss some of the characterisation measurements undertaken.

An AO system like ELECTRA is able to conjugate the low order wavefront aberrations (piston and tip–tilt) easily. As stated in Chapter 1.1, areas of the correcting element can be used to stabilise the incident beams on the input endfaces of DAFI's optical fibres. This will (a) ease the component alignment

[1] **G**round–based **H**igh **R**esolution **I**maging **L**aboratory.
[2] The deformable mirror segments were not cophased.

tolerances during setup, (b) conjugate the incident wavefront's tip–tilt aberration across the corresponding subaperture (fibre), so maximising the input coupling and (c) compensate for the piston errors between different subapertures, thereby cophasing the interferometer.

## 2.2 Optical layout

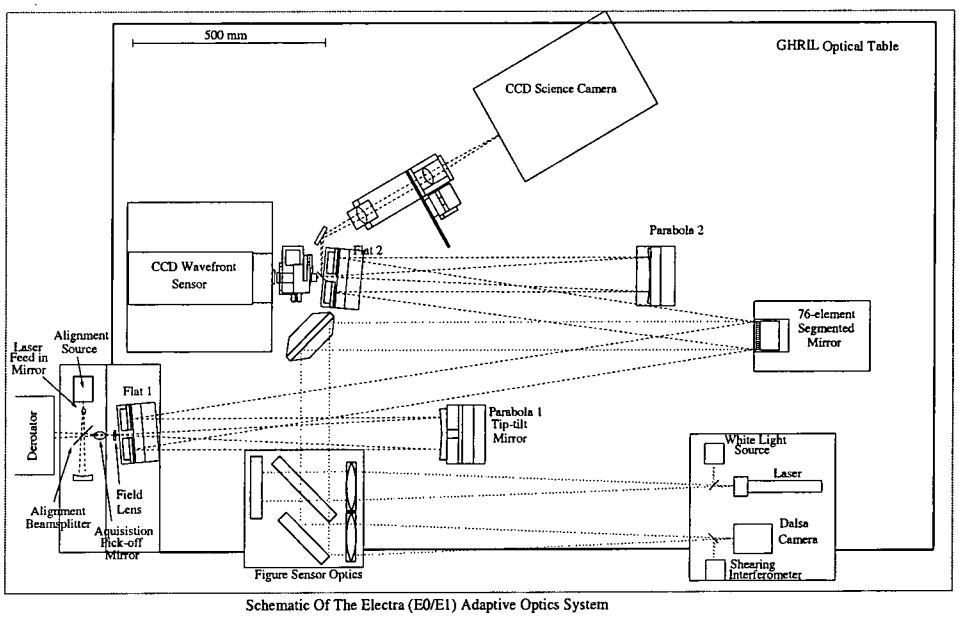



**Figure 2.1.** Schematic layout of ELECTRA optics. The main optical train is emphasised; the Figure Sensing path is also shown as a separate path. See main text for details.

Light from an unresolved (on–axis) reference star is directed on to the Nasmyth focus of the WHT which lies in the central aperture of an angled flat mirror *Flat 1*, as shown in Figure 2.1. The f/11 beam is initially incident on a parabolic mirror *Parabola 1*, which acts as a Newtonian collimator. It directs the beam back on to *Flat 1* which folds the light path to the deformable mirror (DM). The DM is positioned so that its conjugate focus is the telescope entrance pupil, though this can be varied to alter the size of the isoplanatic patch[3]. Downstream from this is a set of refocusing optics which reproduces

[3]This statement can be understood in the following manner: assume the limiting case of

the original $f_{\#}$; part of the refocused beam is incident upon the wavefront sensor (WFS) and the other part is diverted to form the science arm.

### 2.2.1 Deformable Mirror

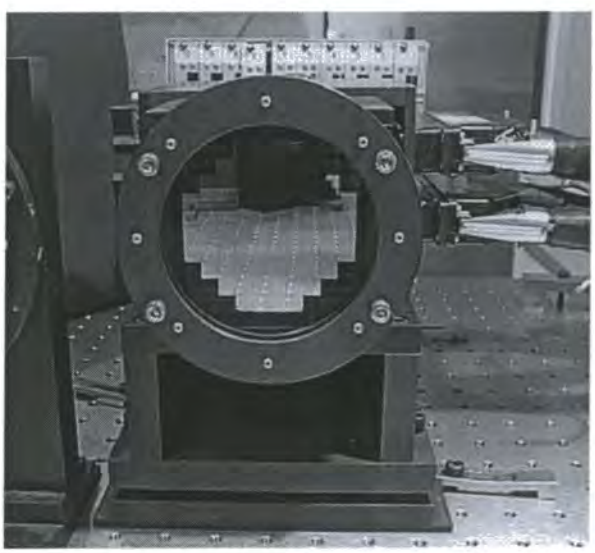


**Figure 2.2.** The segmented deformable mirror shown with its protective coverplate. On the right of the picture can be seen the high voltage actuator cables.

The DM is the compensator for the wavefront distortions. In fact, since global wavefront tilt is responsible for almost half the overall aberration, two components are employed for full correction. The first (*Parabola 1*) is the *fast steering mirror* (FSM) which minimises the wavefront tilt across the full aperture, and the second is the DM itself. The DM consists of 76 independent segments, each driven by an actuator such that it is adjustable in tip, tilt and piston. Precision positioning with nanometer resolution, a high degree of stiffness and fast frequency response are prerequisites for this application, so piezoelectric ceramics are used.

---

one layer of turbulence at some altitude $h$. By placing the DM at the conjugate position in the optical train, all the atmospherically generated aberrations (within the limit of the DM) will be corrected to give a (nearly) infinite $\theta_0$ within the field–of–view.

Each actuator is cylindrical, with three separate electrodes running along its length to provide three–point movement to its mirror segment, as shown in Figure 2.4. In total, therefore, the DM has 228 degrees of freedom.

The actuators, being ferroelectric materials, have an intrinsic non–linear response or hysteresis with respect to applied electric fields which limits their accuracy. They are also subject to drift over time. The maximum error due to hysteresis can be as much as 15% of the path traversed if they are run in an open loop fashion [18]. The microscopic explanation for the origin of hysteresis is still unclear, although it is attributed macroscopically to domain switching [19]. Each magnetic domain is made up of several dipoles; under the influence of externally applied electric fields the effective number of dipoles aligned parallel to the field varies. This alteration does not occur instantaneously and it is this delay that gives rise to the hysteresis loop.


**Figure 2.3.** A single segment mounted on a piezoelectric transducer (PZT). The three electrodes are mounted along the length of the actuator cylinder, separated by 120°. The strain gauge signal and PZT high voltage wires can be seen at the top of the photo.

Drift (or creep) can be caused by two mechanisms:

1. Follow–up polarisation of the ceramics, which is closely connected to the hysteresis behaviour.

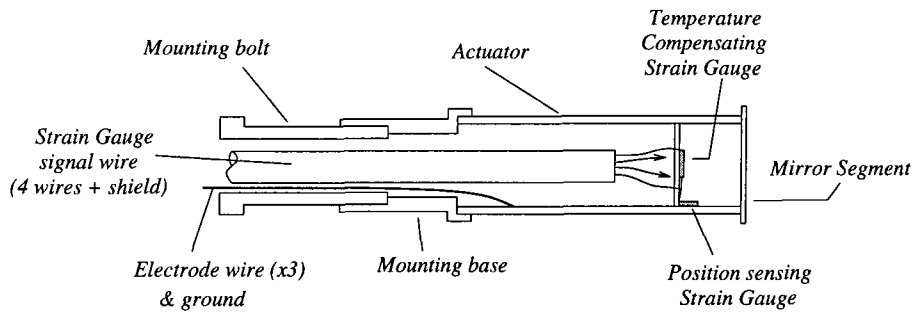



**Figure 2.4.** Cross–sectional profile of a piezoelectric actuator showing the electrodes and strain gauges. There are 6 strain gauges per segment — 3 position sensors and 3 temperature compensators. The high voltage drive electrodes run along the exterior wall of the cylinder.

2. Warm up of the actuator due to a prolonged input excitation.

This effect is only noticeable in the short period (from milliseconds to the order of seconds for large voltage steps) after positioning, so is not important when the DM is operating in a closed loop. It does affect the initial mirror flattening procedure, which is a non–trivial task.

Because ELECTRA's actuators are made of a single piece of piezoceramic, they do not need to be preloaded like equivalent stacked actuators. The preloading of piezoceramics produces a polarisation in the piezotransducer. This polarisation is superimposed onto that caused by the applied electric field. Therefore the hysteresis loop becomes distorted and asymmetric, generating the characteristic 'S' curve of displacement *vs* applied voltage. That is to say, there is no axis of symmetry between the descending trajectory and the ascending trajectory. Each actuator of ELECTRA's is free from this behaviour when its electrodes are pistoned, which is advantageous if predictive hysteresis compensation strategies are desired by using, say, a modified Preisach model [20].

To linearise the extension of the actuators, strain gauges have been placed inside the cylinders parallel to each electrode. These strain gauges have a linear resistance change with respect to extension, so provided that the bonding agent has a negligible effect, the extension of the actuator along the relevant electrode can be calibrated. External temperature variations affect the extension of the gauges. To allow for this, each position sensing gauge is connected in series

with a temperature compensation gauge. These compensators lie in a plane perpendicular to the electrodes forming a set of potential dividers (see Figure 2.11).

The linearity of the DM response is very important to cophase the mirror surface; if the subapertures are pistoned correctly so that they form a continuous sheet, it is clear that a higher degree of wavefront correction can be applied than when only the local tip and tilt variations are conjugated. For the specific case of a dilute pupil interferometer, linearising the positional response of the DM allows us to compensate for piston errors.

Despite the problems of trying to place these components inside the extremely small internal diameter of the piezoactuator cylinders, this is the most effective method of determining the piston ($z$ axis offset) position of each actuator. An alternative method for determining piston, which is utilised later for calibrating the hysteresis, is to use a Michelson interferometer illuminated by a quasi–monochromatic source. By examining the change in the visibility phase as a function of time, that is to say "counting the fringes", the rate of change of piston can be determined. This is not a trivial measurement in terms of the technological demands (e.g. the speed of the detector and pixels per fringe required by the detector) and does not give an absolute value for the piston, unlike the strain gauges.

### 2.2.2 Wavefront Sensor

The WFS unit provides the DM control loop with information about the wavefront distortions of the light from the guide star. It is critical, therefore, to have the highest signal–to–noise possible over the short integration periods required to sample the wavefronts. In essence, the WFS sets the bound on the overall sky coverage of the AO system by limiting the magnitude of the guide star.

The WFS used to probe the wavefront shape across the DM is a Shack–Hartmann type[4]. It operates by re–imaging the telescope pupil (where the DM

[4]The camera and electronic readout were made by Rutherford Appleton Laboratories Space Sciences Division.

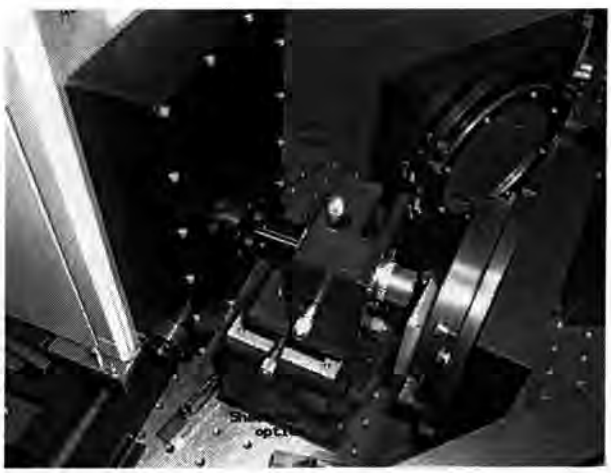


**Figure 2.5.** The wavefront sensing camera and Shack–Hartmann lenslet array.

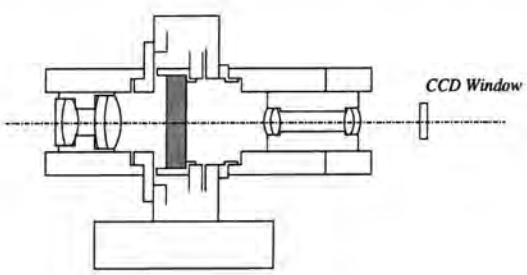



**Figure 2.6.** Shack–Hartmann optics schematic. The lenslet array is shaded and the WFS window is also shown. The optical assembly is mounted on a three–axis stage for alignment purposes.

is effectively located) onto a series of lenslets to form an array of point–like images. The positions of these images measure the local wavefront tilts across each lenslet.

The WFS consists of an 80 × 80 pixel square array CCD placed behind a microlens array. Each microlens maps a DM segment to an 8 × 8 pixel bin on the CCD. The wavefront sensing camera necessarily uses the visible passband, so the WFS lies downstream of a 40/40 optical beamsplitter[5] (behind *Flat 2* in Figure 2.1). It should be noted that due to the practically negligible wavelength

[5]This type of pellicle beamsplitter has a coating which gives approximately equal transmittance and reflectance between 400nm and 800nm, but unfortunately is also lossy.

dependence of the refractive index of air, the science arm could contain different wavelengths to the WFS arm (most notably infrared).

### 2.2.3 Figure Sensor

**Figure 2.7.** Typical Michelson Figure Sensor video frame of the DM.

A separate component exists with ELECTRA to independently characterise the DM surface figure. This *Figure Sensor* combines both Michelson and shearing interferometers. The classical Michelson interferometer is illuminated by low intensity HeNe laser light ($\lambda = 633$nm) and is used to extract the tip–tilt information of the subapertures. The shearing interferometer was intended to calibrate the relative piston positions of the segments, and so is illuminated by a white light source. However, it was found to be difficult to implement during the $E1$ run[6]. The method of finding the datum piston position of the DM is still under investigation.

The optical layout of the Figure Sensor subsystem is also shown in Figure 2.1.

Figure 2.7 illustrates a typical view of the unpowered DM surface with the Michelson interferometer. For any one segment, the higher the fringe spatial frequency, the more the segment is tilted with respect to the reference flat.

[6]The problem encountered was that the white light fringes, which only exist in a small region (about the width of one segment), ran diagonally across the viewing frame as a result of chromatic dispersion within the interferometer. More seriously, the position of the fringes could not easily be adjusted without affecting the angle and phase of the fringes, so potentially losing the piston reference.

## 2.3 System Control

At the most elementary level, there exist two feedback loops for ELECTRA. One processes the signals from the WFS in real time to construct the mirror command matrix. It finds the centroid positions of the Hartmann spots and formulates the initial command matrix to translate them back to the centre of each Hartmann subaperture. The other reads in the electrical strain gauge signals to sense the position of the DM electrodes. The positional offsets derived from the strain gauge signals are subtracted from the original matrix to provide the corrected command matrix. This is then applied to the DM.

More specifically, let $\Phi(x,y)$ be the phase of a wavefront incident on the system producing $m$ signals at the WFS. The measurement vector, $[\vec{s} \triangleq s_j : j = 1,\ldots m]$ is related to the control signal $\vec{c}$ by the command matrix[7] $\hat{M}$,

$$c_i = M_{ij} s_j \; ; \qquad i = 1,\ldots n, \; j = 1,\ldots m \tag{2.1}$$

The number of sensor signals $m \geq n$, where $n$ is the number of DM electrodes. For notational brevity the Einstein summation convention has been adopted. It is noted that when finding $\vec{c}$, a direct wavefront estimate in terms of Zernike or Karhunen–Loève coefficients is not explicitly measured in ELECTRA. In a noise–free situation, the sensor signals $\vec{s}'$ would ideally be

$$s'_j = H_{ij} c_i \; ; \quad H_{ij} \triangleq \left\{ \frac{\partial r_i}{\partial x_j}, \frac{\partial r_i}{\partial y_j} \right\} \tag{2.2}$$

for the $m/2$ Hartmann spots. The DM influence function vector is denoted by $[\vec{r} \triangleq r_i(x,y) : i = 1,\ldots n]$. In reality the operator $\hat{H}$ averages the pupil across each lenslet, which this definition does not take into account.

What is actually measured is $\vec{s}$. Let the difference vector $\vec{D}$ be

$$\vec{D} \triangleq \vec{s} - \vec{s}' = \vec{s} - \hat{H}\vec{c} \tag{2.3}$$

[7]Operators will be indicated by a circumflex above the letter.

The problem is now to find $\vec{c}$ such that the squared error $\epsilon^2$

$$\epsilon^2 = ||\vec{D}||^2 = \vec{D}\vec{D}^T \tag{2.4}$$

is minimum, with $||\ldots||$ denoting the norm and $\vec{D}^T$ is the transpose of $\vec{D}$. This is the standard *least squares* problem, and its solution is given by the *normal equation*

$$\hat{H}^T\vec{s} = \hat{H}^T\hat{H}\vec{c} \tag{2.5}$$

If $\hat{H}^T\hat{H}$ is non–singular, so its inverse exists, then the least squares solution is

$$\vec{c} = (\hat{H}^T\hat{H})^{-1}\hat{H}^T\vec{s} \tag{2.6}$$

The control matrix $\hat{M}$ is therefore

$$\hat{M} = (\hat{H}^T\hat{H})^{-1}\hat{H}^T \tag{2.7}$$

Since in practice, the operator $\hat{H}^T\hat{H}$ is often singular, an approximate inverse has to be found. Singular–value decomposition [21] methods can be used to define this. If $\vec{U}$ is defined to be the eigenmodes of $\hat{H}$ and $\lambda$ is the set of eigenvalues then it can be shown that

$$\hat{H}^T\hat{H} = \vec{U}\Lambda\vec{U}^T \tag{2.8}$$

$\Lambda$ is a diagonal matrix of the eigenvalues $\lambda$. The least squares solution for the control matrix can now be written as

$$\hat{M} = \vec{U}\vec{\Lambda}^{-1}\vec{U}^T\hat{H}^T \tag{2.9}$$

where $\vec{\Lambda}^{-1}$ is a diagonal matrix of $\lambda^{-1}$ when $\lambda \neq 0$ and zero when $\lambda = 0$.

$\lambda^{-1}$ denotes the gain of the system eigenmodes in the loop. If $\lambda = 0$ for a mode, the WFS is insensitive to that mode — the Shack–Hartmann being "blind' to wavefront piston.

This entire analysis assumes that the WFS–DM system is linear, which is a

very poor approximation for the real case. By measuring the $n$ electrode piston values through the strain gauge feedback loop, a matrix of offsets $[G \triangleq G_{ij} : i = 1, \ldots n; j = 1, \ldots m]$ can be created (through a lookup table, for example) which can then be added to $\hat{M}$ to linearise the DM.

A shortfall of this method is that the least squares solution for $\hat{M}$ minimises the *sensor measurement* residuals, rather than the wavefront phase residuals. Neither does it make use of any prior knowledge about the wavefront (e.g. the spatial statistics) nor incorporate the measurement noise/mirror influence function characteristics.

### 2.3.1 Control Hardware

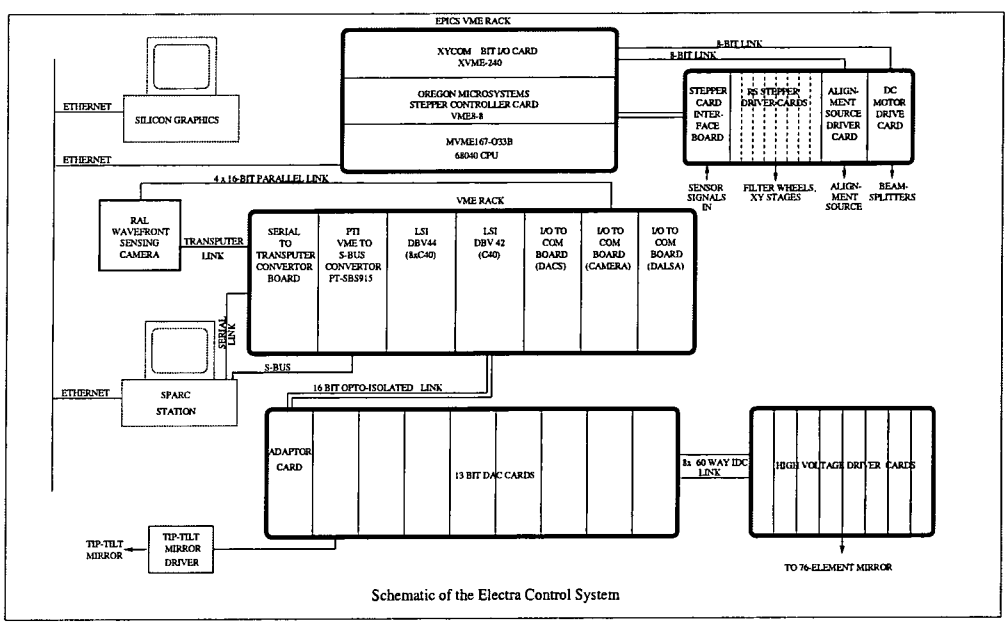



**Figure 2.8.** ELECTRA control system for *E0*.

The computing control devices used are a VME rack of TMS320C40 ("C40") digital signal processors[8], a Sun Sparc10 engineering console and a Silicon Graphics visualisation console. The C40s are connected to the DM, FSM and the WFS by C40 communication port links (commports), and so appear simply

[8]Produced by Loughborough Sound Images Ltd.

as 32–bit data streams. Internally the C40s are also connected via commport links. These links are high speed (20MB/sec), bi–directional, byte–wide interfaces designed to allow interconnection of multiple C40 processors without the need for any additional "glue" logic. The topology of the links can be altered by using front panel jumpers, but currently the C40 network consists of 8 C40s connected in a ring, with a ninth connected as a spur to the ring. The ninth processor acts as the diagnostics processor.

The C40 cards are connected to the Sun workstation over the VME bus via an SBUS–to–VME converter. The Sun has access to the VME bus which appears as a memory range in the Sun's memory space. The C40 motherboards have various interfaces to the VME space, the most important being the Link Interface Adapters (LIAs). These map a commport from each of 5 C40s onto register sets on the VME bus. By reading and writing these registers, the host system can send and receive datastreams to and from a C40 commport. Device drivers on the Sun make these appear as simple stream devices. An important function of these LIAs is to download C40 programs at boot time.

### 2.3.2 Control Software

The Sun runs Solaris, a non real–time operating system, and talks to the rest of the world via standard ethernet. The available bandwidth over this link is approximately 200Kb/sec (measured using DTM as the communications protocol). Connected over this link is a SGI Indigo–2 workstation, which has the (measured) ability to render 3–dimensional views of wavefronts at video rates ($> 10$ frames/sec). The intention is to use this as the main console of the system during normal operation.

The software architecture is schematically shown in Figure 2.9. Its design was bandwidth driven. Since the C40 network has the highest bandwidth data links, it was made the basis of the highest speed (nominally 1kHz) control loops — principally to control the DM using WFS data. The C40s are configured as a server architecture so they do not initiate actions but rather respond to requests from external client processes.

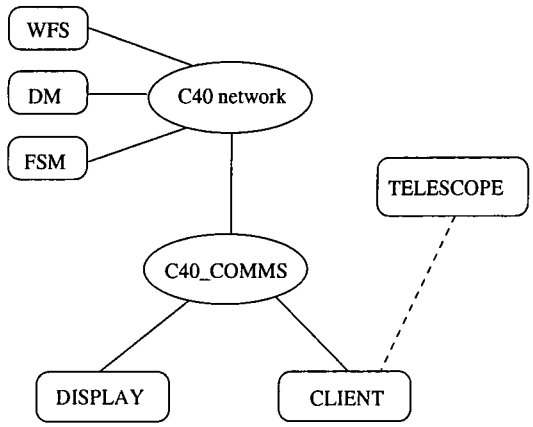



**Figure 2.9.** ELECTRA's control software architecture (simplified). The details of the various software loops within the C40 network have been omitted. The link to the main telescope control is not yet implemented.

The clients talk to the C40s via the LIAs. Since the LIAs are a global resource, and multiple clients need to talk to multiple C40s, a software layer is interposed which facilitates this. It consists of a messaging protocol that delivers messages between the C40s and between the Sun and the C40s, plus a central process running on the Sun called **c40Comms**, which multiplexes messages arriving on the Sun between multiple clients. The net result is that any client residing anywhere on the network can talk transparently to any C40, and vice versa.

For more details on the software architecture, the reader is referred to Reference [22].

### 2.3.3 The Mirror Mimic

DM commands can be applied manually via the Mirror Mimic GUI (Figure 2.10). This client application interfaces directly with the mirror control C40 through c40Comms. Individual actuator electrodes can be messaged so that each segment can be altered in tip, tilt or piston. The resolution of the mirror drive DACs is 13 bits.

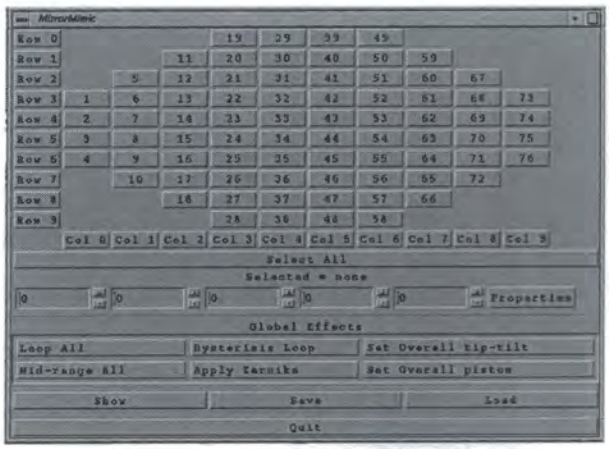



**Figure 2.10.** Mirror Mimic graphical interface.

## 2.4 Strain Gauges

In Section 2.2 the motivation for mounting strain gauges onto the piezoelectric transducers of the DM was discussed. A set of measurements have been carried out to assess the response of the strain gauges as postion sensors are described in this Section. With an upper bound on their non–linearities thus established, the hysteresis behaviour of the actuators was characterised. The conventional technique for minimising the magnitude of the hysteresis by means of decaying sinusoids was investigated. At the time (pre–$E1$), the strain gauge control loop had not been implemented so a comparison between the open and closed loop performance of the piezoactuators could not done.

The physical characteristics of the gauges follow

- Material : P–Type Silicon
- Resistance (R): 500Ω (±10% at 21° C)
- Gauge Factor (GF): $\frac{\Delta R/R}{\Delta L/L} = +130$ (±5% at 21° C)
- Thermal Coefficient of Resistance : $\frac{\Delta R/R\ \%\pm 3\%}{55°\ C} = +9\%$ (21° to 77° C)
- Thermal coefficient Gauge Factor : $\frac{\Delta GF\ \%\pm 3\%}{55°\ C} = -15\%$ (21° to 77° C)

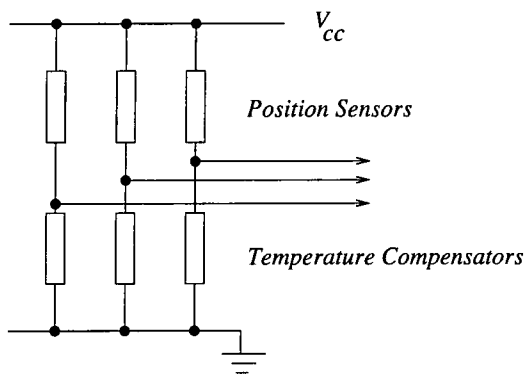



**Figure 2.11.** Circuit diagram of strain gauges within one actuator.

- Typical strain level[9] : $0.1\mu\epsilon$ (where $1\mu\epsilon \triangleq \Delta L/L \times 10^{-6}$ is the unit microstrain.)

### 2.4.1 Strain Gauge Amplifiers

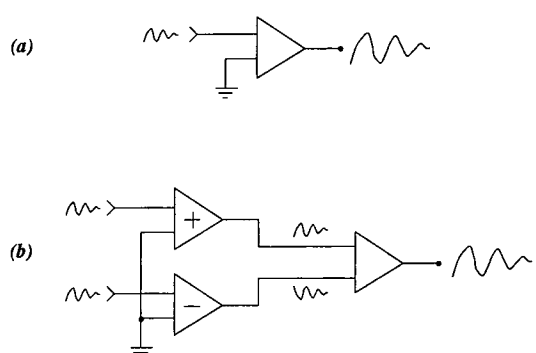



**Figure 2.12.** Schematic representation of **(a)** single-sided *vs* **(b)** differential amplification. The former geometry relies on the stability of the reference ground in order to provide a linear amplification stage. The latter amplifier is fed a positive and negative version of the original signal from two active transimpedence stages. The difference between the amplifier's inputs is multiplied by a gain factor. This method isolates the output from reference ground drifts.

The design of the strain gauge signal amplification stage naturally had the objective of minimising the degradation of the weak signal by amplifier "noise"

[9]Manufacturer's recommendation.

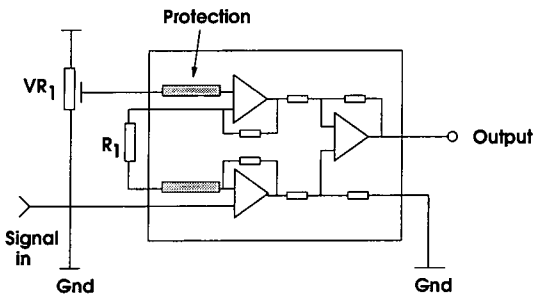



**Figure 2.13.** Amplification stage of strain gauge signal.

or external pickup. On the other hand, as this was a prototype system, to be tested in controlled surroundings, it was decided that the amplification stage could be powered by a single–sided (with respect to ground) power supply, rather than implementing a more robust differential drive system. The difference between the two systems is schematically illustrated in Figure 2.12. The single–sided amplification stage relies on the reference d.c. voltage (ground) to remain stable for linear amplification. We assume that the amplifier gain is itself linearised by some internal negative feedback circuit. The differential drive introduces two more active transimpedence devices before the amplification stage. These devices provide a positive and a negative version of the original signal (with respect to some common reference). The signals are then differentially amplified, so any reference voltage drifts are rejected. Although much more stable, the transimpedence devices require two separate voltage rails, usually chosen to be symmetric about a reference, so requiring a double–sided power supply.

As can be seen in Figure 2.13, the classical (monolithic) instrumentation amplifier was chosen to amplify each strain gauge signal independently [23]. This provides high differential gain and unity common–mode gain without the need for close resistor matching. It also has the advantage of permitting the d.c. offset to be trimmed at one of the input op–amps by manually adjusting $VR_1$. The gain is set by the resistor $R_1$.

### 2.4.2 Digital to Analog Converters

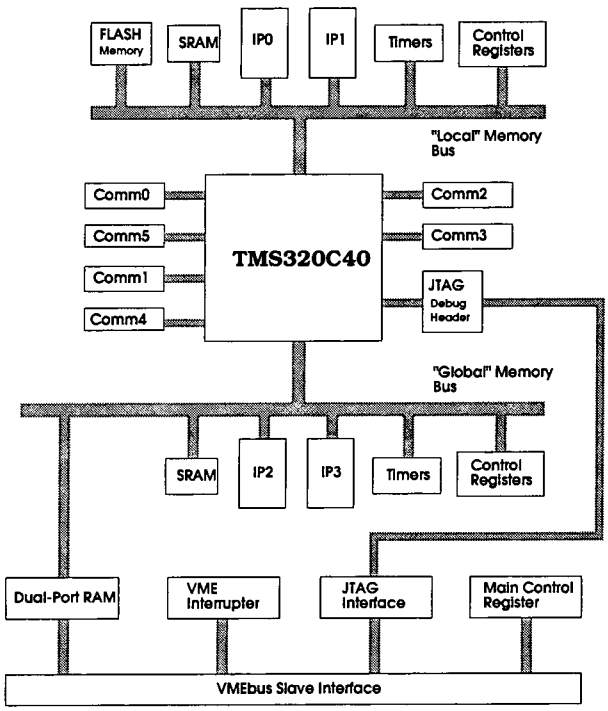



**Figure 2.14.** Hydra CommIO–IP block diagram.

The device used to read in the amplified strain gauge signals was an Ariel Hydra CommIO–IP processing engine, on which were sited four modular multichannel analog–to–digital convertors. The Hydra CommIO–IP, based on the C40 digital signal processor, in principle provided an I/O path for other C40 systems as well as connections to the VMEbus. It was initially investigated as a possible solution for the ELECTRA system, but unfortunately interfacing the Hydra to the C40 ring proved to be a non–trivial task. In addition to this problem, interfacing the strain gauge amplifiers to the Hydra would have required an extra post–amplification stage because of the length of the cables involved. In light of these considerations, the Hydra was not found to be suitable.

The IndustryPack (IP) modules are otherwise a versatile method of implementing a wide range of I/O, control, interface, slave processor, analog and digital functions. About the size of a business card, IPs mount parallel with the host Carrier board which provides the host processor and primary bus in-

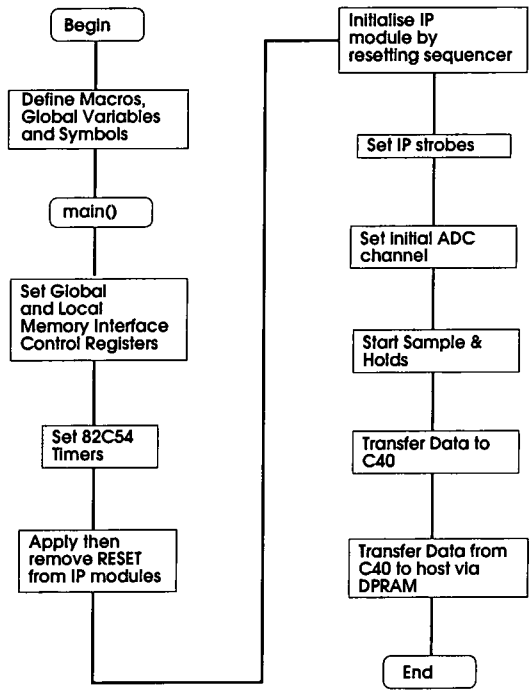



**Figure 2.15.** Hydra CommIO–IP program synopsis.

terfacing, as well as a mechanical interface to the outside world.

The Hydra (a double height VMEbus board) served as the host Carrier — its structure in block diagram form is shown in Figure 2.14. It is a dual port device, with two identical memory ports which are designated as the *Local bus* and the *Global bus*. Connected to the Local memory interface are

1. **Flash Memory** — A bank of 512 Kword (×8 bits) of locked page Flash memory which stores custom boot firmware.

2. **SRAM** — A bank of 256 Kword (×32 bits) of zero wait–state static memory.

3. **IP Sites 0 & 1** — Two single width, 16 bit IndustryPack sites.

4. **82C54 Programmable Timers** — Programmable timers available as sample clocks or event indicators for IPs 0 and 1.

5. **IP Site Control Registers** — A set of registers to select the IP operating modes and status.

The Global memory interface has a similar architecture as the Local bus for IPs 2 and 3, with an additional connection to the C40–to–VMEbus dual port memory (DPRAM).

The C40 itself[10] has six commport links, of which four are accessible from the front panel. While in principle this is convenient, in practice, due to the lack of a commport electrical standard, connectors linking the Hydra to the C40 ring were extremely unsatisfactory.

The IPs used were Wavetron IP–HiADC analogue–to–digital converters. Each IP–HiADC contained 16 sample and hold amplifiers organised as four groups of four. The channels could be triggered by the CommIO–IP timer, available on the IP–bus, a software strobe or the IP's onboard 8MHz timer. The input range of each channel was fixed at ±5 volts with an impedence of 1MΩ and the resolution was 12 bits with an accuracy of ±2 LSB. The ADC maximum conversion rate was 1.25MHz.

A synopsis of the software driver written for the ADCs is shown in Figure 2.15.

### 2.4.3 Minimising Hysteresis Effects in Open Loop

The initial condition necessary for cophasing the DM is a flat mirror surface, so providing a datum position. The flattening procedure is made easier by removing the local memory of the actuator. Two types of decaying sinusoid was applied to the piezoceramics to effect this.

The Global antihysteresis (AH) loop was applied to the whole DM with the intention of removing the long term history of trajectories. A Segment AH (SAH) loop, of slightly smaller amplitude and shorter duration than the AH loop, could be applied to each segment before stepping the electrodes in an attempt to minimise the trajectory error. These algorithms were used during the $E0$ run to help flatten the DM. They were not so useful once the strain gauge feedback loop was implemented during the $E1$ observing run.

[10]It should be noted that the C40 50MHz processor is actually run at 48MHz on the Hydra for synchronisation with the 8MHz IndustryPacks.

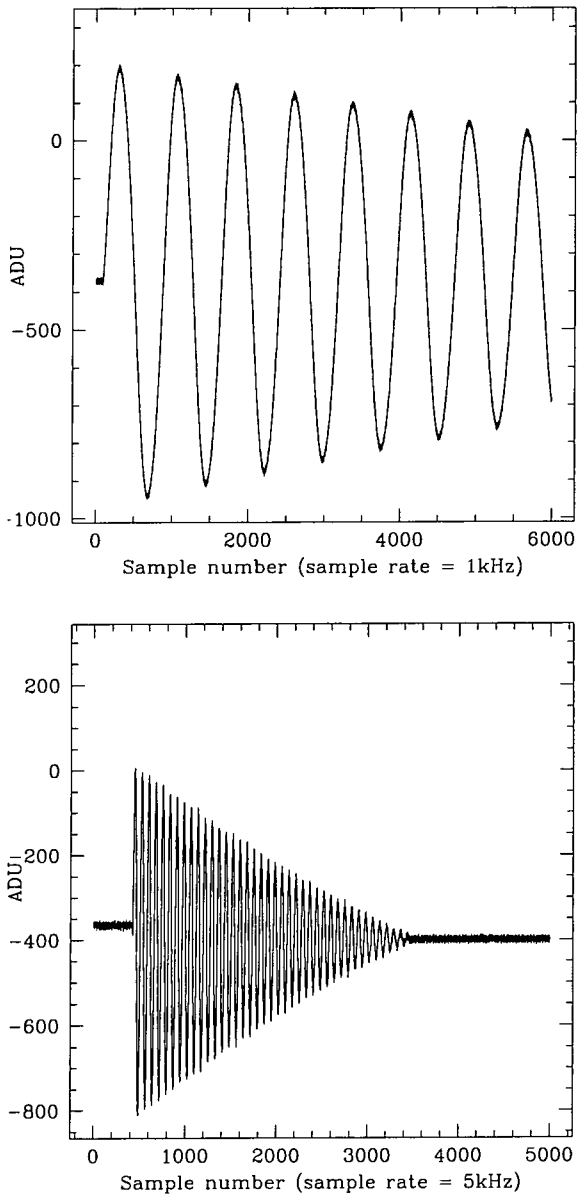



**Figure 2.16. Left:** Part of the Global antihysteresis loop, applied to the entire DM. **Right:** Segment antihysteresis loop, applied to all the electrodes of a single segment.

### 2.4.4 Actuator response

#### Strain gauge linearity

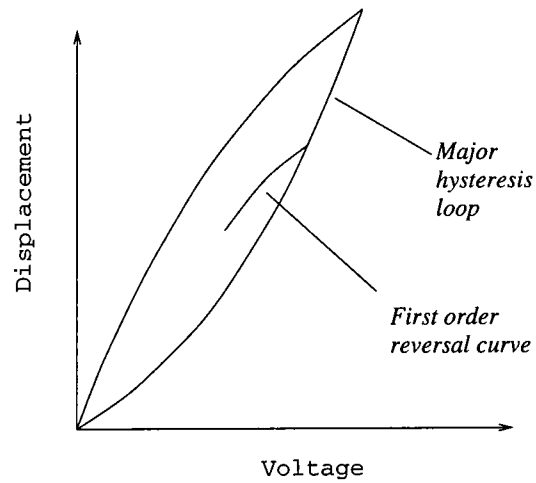



**Figure 2.17.** Sketch of the hysteresis response of a piezoactuator.

In order to calibrate the actuator displacement as a function of both applied DAC values and strain gauge ADC readings, it was necessary to optically measure the actuator piston steps. The Michelson Figure Sensor was used to do this. Starting with a sufficient number of tilt fringes (i.e. $> 4$) across a segment to extract the fringe visibility phase, the segments were pistoned with steps of less than $\lambda/4$. At each step, both an image frame from the Figure Sensor and the three position strain gauge readings were recorded.

The CCD used by the Figure Sensor had a low spatial resolution, so attempting to measure the behaviour of a single electrode optically (i.e. tilting the segment) was not possible over a large range of an actuator's movement due to aliasing problems.

The Global AH algorithm was applied to the DM after midranging all the actuators, prior to taking the readings. Since no attempt was made to impose a local trajectory history[11] the measurements trace out the leading order trajectory reversal curves. Figure 2.19 shows the hysteresis curve of a mirror segment

[11] i.e. Looping around a trajectory several times before making the measurement.

measured optically and through the strain gauges.

The Figure Sensor images were two dimensionally Fourier transformed to find their amplitude and phase. The Figure Sensor tilt fringes are essentially monochromatic, so each fringe image defines a single Fourier pixel in the 2D Fourier (or fringe visibility) plane[12]. It was therefore possible to extract each Fourier pixel coordinate in the visibility amplitude plane. These coordinates served to constrain the position of the visibility phase information. Since the visibility phase is the argument of the real and imaginary parts of the complex Fourier vector, it will suffer from a $2\pi$ ambiguity. Therefore the phases were unwrapped to give the fringe shift (in radians) as a function of segment displacement.

Local seeing effects were significant for such a long optical path, so it was necessary to calibrate the path variations by considering the temporal phase variation of adjacent static segments. The seeing cell sizes ranged between 2 to 3 segments ($r_0 \approx 9$ to 15cm) in diameter so it was possible to measure the fringe phase drift over time. It was found that the optically measured hysteresis loops could be recovered from the raw data by subtracting the phase drift from an adjacent segment.

Figure 2.18 shows the errors in the linearity of the optical *vs* ADC measurements[13] after calibrating for the seeing. The constant offset between the two graphs can be disregarded since this arises from different initial positions of the segment. We discuss each curve in turn:

**Without AH:** The best fit line is given by $y \propto 0.089x$, with an error in the gradient of 0.15% ($y$ is in mV and $x$ is in radians). The estimated variance of the data residuals is 2.46mrad$^2$.

The curve also exhibits a small amount of hysteresis, such that the downward

[12]Care should be taken to rearrange the four quadrants of the 2D signal before passing it through a discrete 2D FT, in order to recenter the phase origin. For a full discussion, see [24] for example.

[13]The strain gauge amplifiers had a gain of 500 over an input range of $\pm$5V giving 4.88$\mu$V/ADU.

trend is approximated by

$$y \propto 0.089x - 2.35 \times 10^{-5}x^2$$

The $x^2$ coefficient has a standard error $\sigma = 7.1 \times 10^{-6}$ (i.e. the coefficient is $3.3\sigma$), indicating that the curvature is a real effect. This is supported by the variance of this subset of residuals, which at $0.3\text{mrad}^2$ is 7.75 times smaller than the full dataset. Similarly, the upwards trend of the plot is approximated by

$$y \propto 0.089x + 4.40 \times 10^{-5}x^2$$

where the second order coefficient lies within $4.85\sigma$. The rms of the residuals in this subset is $0.56\text{mrad}^2$, or 4.4 times less than the full dataset.

**With AH:** This curve's best fit line is $y \propto 0.090x$ with an error of 0.07%. The variance of the data residuals is $1.3\text{mrad}^2$. Any remaining hysteresis lies within the noise level, since the second order coefficient of the downward and upward trajectories are both less than $3\sigma$.

The significance of these results are twofold: firstly, the strain gauge sensing method is linear to 10nm using this measurement method[14]. Secondly, a small amount of hysteresis remains. This may be a result of the measurement technique, since the method assumes that the actuator electrode gains are all equal. In the real situation, where the electrodes have unequal gains, a small degree of tilt is probably introduced onto the segment. This tilt would affect the phase measurement, and appear as the residual hysteresis. Also, since the position sensing strain gauge only samples a small part of the electrode, it is unlikely to be able to compensate for the hysteresis perfectly.

Having quantified the upper limit on the non–linearity of this position sensing method, we are now in a position to characterise the behaviour of the piezoactuators themselves.

[14]This value has been deduced from the widest separation (in mV) of data points from the graph without AH.

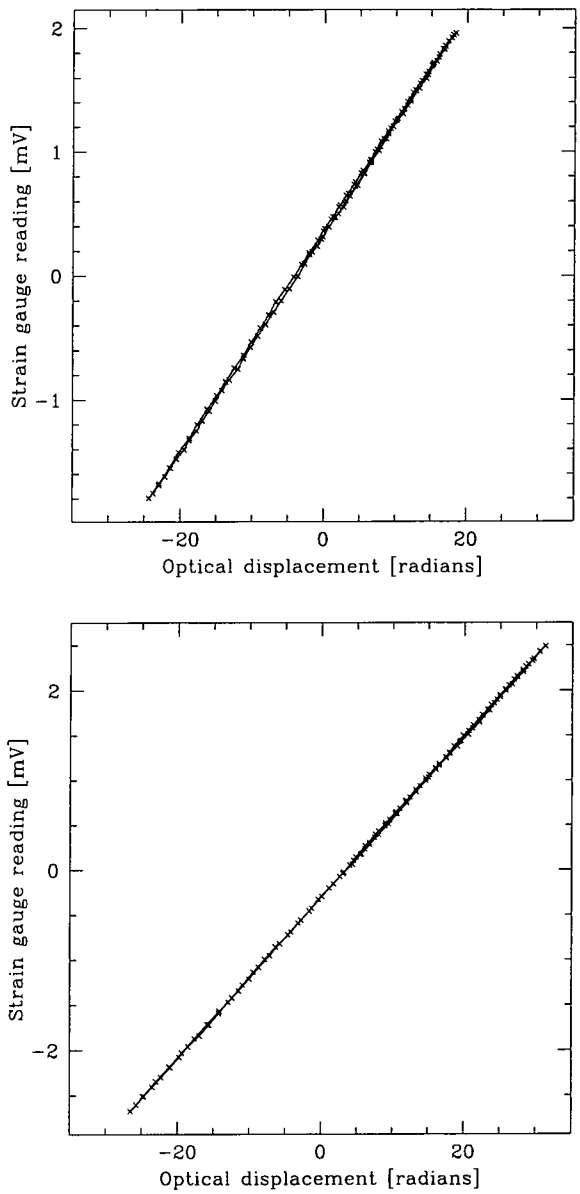



**Figure 2.18.** Errors on the linearity of measurements: optical vs strain gauge (**top**) without the SAH algorithm and (**bottom**) with the SAH algorithm. The d.c. offset between the two lines arises because the segment was not at the same physical displacement for both cases. It was also pistoned over a greater range for the case with SAH than without. The gradients of the best fit lines are $m = 0.090$ for the SAH loop and $m = 0.089$ without it. See text for discussion.

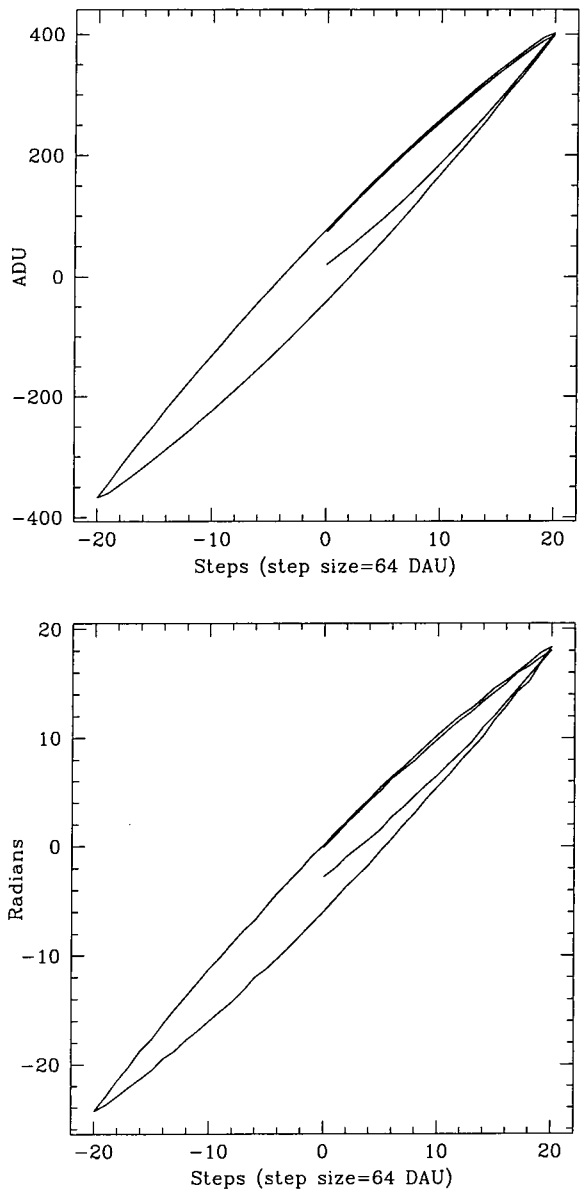



**Figure 2.19. Top:** Strain gauge measurements hysteresis curve of an actuator. **Bottom:** Same displacements measured optically.

### Actuator response time

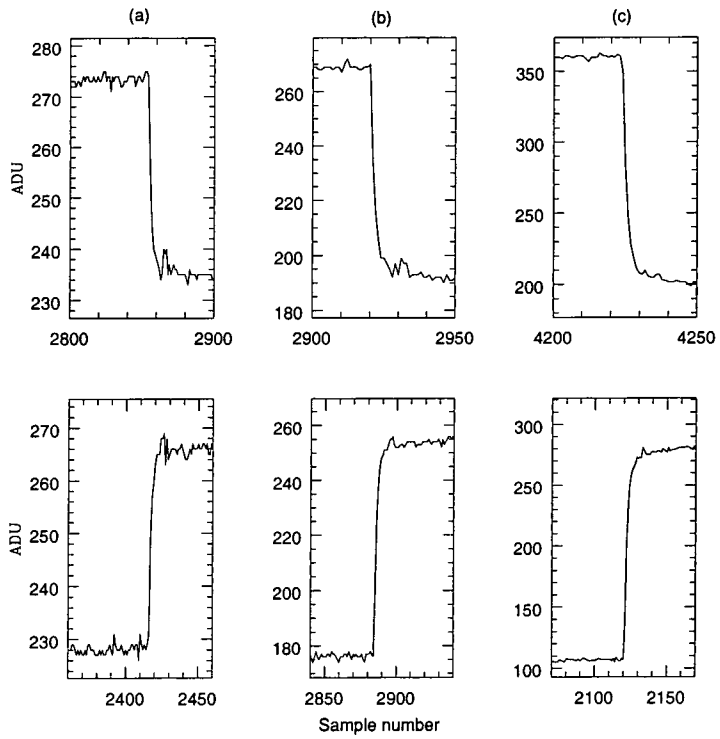



**Figure 2.20.** Step profiles for various step sizes (sample rate = 10kHz).

The actuator response time was measured, via the strain gauge sensors, for various step sizes as shown in Figure 2.20. The actuators have a mechanical resonance at 2kHz, so a 1kHz filter was put on the ouput of the drive circuit to supress this. The rise time was measured to be $\sim$ 1ms; this can be improved by removing the drive circuit filters and suppressing the ringing in software. It is also possible to see the creep effects which last between tens of milliseconds up to several seconds, depending on the step size, after moving the actuators.

### Piston

In Figure 2.21, two effects of the AH loops may be seen. Principally, the hysteresis effect is reduced by about 20% when measured across the greatest width of the curve.

A second result is that the actual displacement is increased by 36% ($\sigma$ = 1.1%) of the standard hysteresis loop when SAH is applied for a given DAU step.

The positioning error at the top endpoint of the hysteresis curve with SAH applied was $\sim$ 16.4nm, as compared to the standard curve $\sim$ 4.4nm, which at first glance is surprising. However, this is attributable to the SAH algorithm which, as can be seen in Figure 2.16, uses a maximum $\sim$ 20% of the available throw. The code was designed to prevent the SAH loop from exceeding full throw by modifying its amplitude. The measurements ranged over 31.25% of the full actuator motion, so the SAH amplitude limiter clearly affected the local history removal.

As a point of reference, define the major hysteresis loop as the trajectory followed by the actuators for maximum throw.

The changes in width and angle may be explained by considering these minor hysteresis loops with respect to the major hysteresis curve. The initial order reversal curves simply trace out a trajectory dependent on the initial position within the bounds of the major loop. Application of the Global and Segment AH algorithms attempt to displace this position toward the centre of the major loop, where the hysteresis effect is minimised. So, positive displacements, which are traced out by the ascending (lower half) of the hysteresis curve, move further upwards and vice versa for the negative displacements, which follow the descending (upper half) of the curve. This argument assumes that the hysteresis loop of piezoceramic actuators is prevented from attaining positive or negative saturation. This is quite valid, since the expansion of any such actuator is limited by the finite maximum field strength at which the actuator can be operated without being damaged by voltage sparkovers.

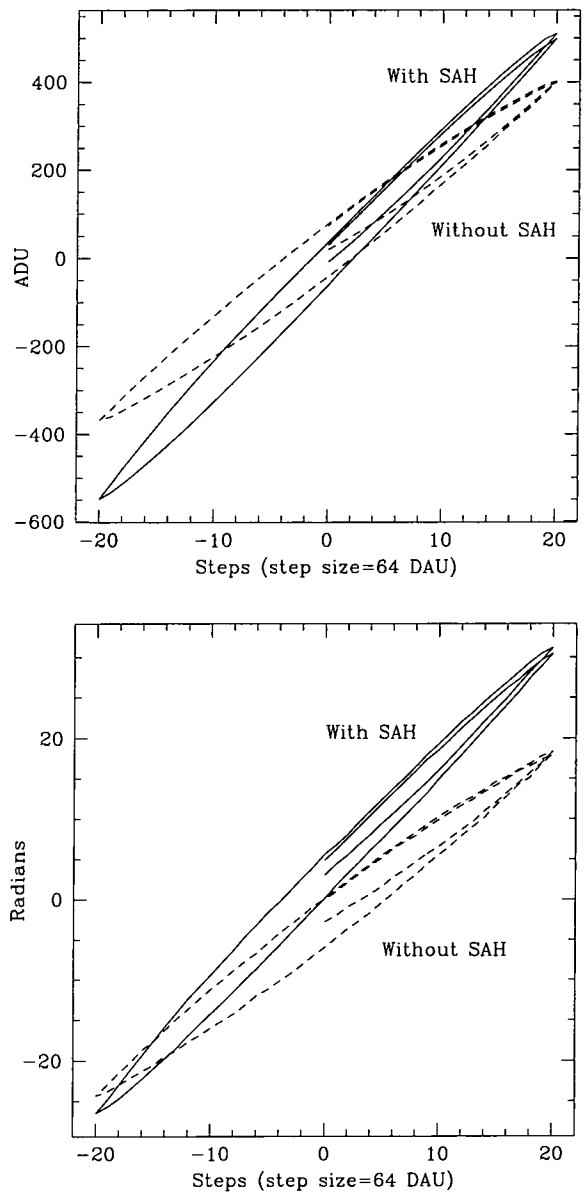



**Figure 2.21. Top:** Strain Gauge **Bottom:** Optical piston readings with (solid line) and without (dashed line) SAH. The initial positions lie at the centre of the loops; the first order reversal curves are traced out to the initial turning point. The effect of the SAH algorithm is to reduce the hysteresis effect from 15% to 12% of the full trajectory. It also alters the angle of the hysteresis curve and extends its length. (Note: 64 DAU = 3.3 V).

## 2.5 GHRIL environment

During the *E*0 run, the temperature variations in the GHRIL environment were measured, as shown in Figure 2.22. The solid curve in the GHRIL room plot is the temperature profile from a sensor placed next to the DM. The dotted curve is from a sensor placed near the science detector. These can be compared to the readings of a temperature sensor in the WHT dome (solid curve) and one at the primary mirror (dotted curve) [25] over some of this time period. The detectors were calibrated to within 0.1% accuracy of each other, so the d.c. offset is not attributable to systematic error. This was confirmed by a separate, hand held thermocouple at the start and end of both nights.

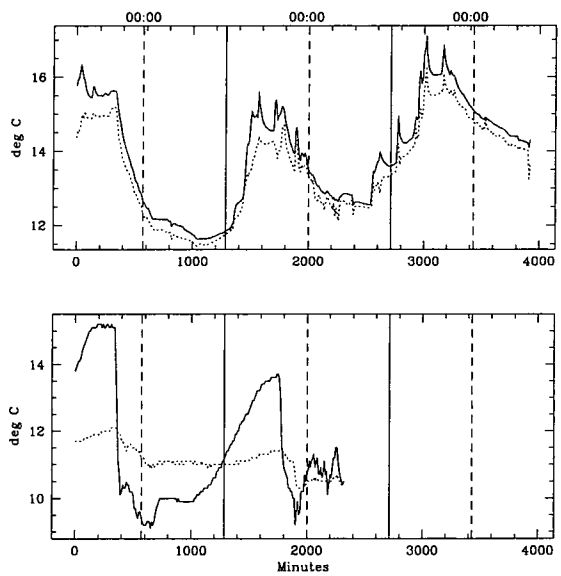



**Figure 2.22.** Temperature profiles **Top:** GHRIL **Bottom:** WHT internal dome and primary mirror, from 14:34 (18 June 1997) to 07:54 (21 June 1997) taken at 10 minute intervals. Midnight and midday are indicated by vertical dashed and solid lines respectively. See text for details.

The effect of opening the dome as soon as the Sun is down is evident. The internal temperature of the dome falls then stabilises quite quickly, but the GHRIL optics room requires some three hours. The temperature conditions in GHRIL differ between the initial night (18th) and the following (19th) mainly

because the extractor fans in the adjacent electronics room were switched off, although the roof hatch was opened to allow cooling. It appears that the fans are effective at maintaining temperature stability, which needs to be considered as ELECTRA's electronics are not cooled.

The fluctuations in temperature on the second night was potentially the cause of "dome seeing" within the GHRIL.

Thermal stability is of importance to the DM because of the large thermal coefficients of the strain gauges. Fortunately, the temperature within the GHRIL does not vary by more than half of one degree Celsius provided the extraction fans are on. So the repeatability of the DM flattening voltages should be quite good over the period of one night, though some strain gauge (and hence mirror segment position) drift is expected.

## 2.6 Summary

A brief description of ELECTRA has been given in this Chapter. The motivation for cophasing the DM surface and therefore the need to linearise the segment actuators' responses has been discussed. Having an AO system in front of DAFI simplifies the alignment of a fibre interferometer's input subapertures, since this allows the DM segments to compensate for any static errors. Therefore the energy coupling into the fibres should be maximised — one of the major objections to using fibre waveguides being their low coupling efficiency.

Linearising the DM is a necessary condition to correct for the piston errors that occur on the incident stellar wavefronts. By having piston compensation before the fibre arms, the fringe visibility degradation should be minimal.

We have seen that response of the strain gauges, as compared to optical measurements, is highly linear. This is a very positive result, since the mechanism for calibrating the DM actuators' pistons rely on the strain gauges producing a linear output. Electrical noise was also a concern, but during the $E1$ commissioning run at the WHT, it appeared that the strain gauge noise was as low as 1.5ADU (as compared to $\sim$ 6ADU as seen in the laboratory in Durham). The application of decaying sinusoids prior to moving an actuator has been shown

to remove some of the long term memory of the PZTs. However, ELECTRA's actuators do demonstrate quite a long memory since the improvement was only some 20%.

Environmental data taken during the $E0$ observing run has also been presented. Although limited, the data indicate that preserving the DM white light flat over the course of one night should be possible given the current conditions of the GHRIL. This seemed to be confirmed during the $E1$ run. In order to guarantee that the DM datum flat is restorable over a night (and perhaps between nights), considerable effort should be devoted to cooling (and controlled heat removal from) the local electronics so as to allow neither large temperature gradients across the room, nor the more serious temperature variations over time.

# Chapter 3

# Optical fibre waveguides

Mathematicians are like Frenchmen: whatever you say to them, they translate it into their own language, and forthwith it means something entirely different.
*Johann Wolfgang von Goethe*

## 3.1 Introduction

In this Chapter we review the basic theory of light propagation through single mode fibre waveguides. Because it is not a trivial task to calibrate an interferometric system, a good understanding of the various systematic errors that can be introduced by the waveguides is necessary. The two principal effects within a waveguide are (a) polarisation change through fibre birefringence and (b) wavelength dependent (chromatic) dispersion. These are discussed in Sections 3.3 and 3.7 respectively.

However, the most important factor in determining the final SNR of a fibre based instrument is the fibre throughput. The selection of the fundamental mode, as will be seen in the review of fibre theory (Section 3.2), requires that the waveguide be physically small. This characteristic means that diffraction effects are predominant at the input/output of a fibre, leading to fast $f_{\#}$'s. Typical SM fibres $f_{\#}$'s range between f/1 and f/5; evidently it is not a straightforward (or economical) task to design coupling optics to match this. In Section 3.4, the use of *adiabatically tapered* SM fibres to slow the $f_{\#}$'s is considered, to simplify the

input coupling issue. Some elementary throughput trials were conducted on two test fibres. The first of these was a tapered, circular core SM fibre which behaved as predicted by the theory. The second was an attempt at manufacturing a tapered SM fibre with a high degree of intrinsic birefringence. The intention was to control the state of polarisation of the transmitted light beam over the entire fibre length. This fibre did not behave as desired, prompting a brief study into the design parameters, described in Section 3.6. The result is shown to be a useable, if weak, condition for designing the fibre taper angle.

## 3.2 Single mode fibre basics

The study of dielectric waveguide physics is a well established field, such that there are numerous texts describing the solutions of the Helmholtz equation (e.g. [26]) for a number of different boundary conditions. However, a complete description of the guided and radiation modes of optical fibres is non–trivial since the full modal expressions require six component hybrid fields. As a result, its application to real (non–idealised) fibres is not practical. A simplification in the modal description is made possible by realising that most real fibres have core–cladding refractive index differences which are extremely small. This approximation reduces the six component field to a four component field. This is further simplified by employing Cartesian coordinates, rather than the cylindrical coordinates that appear so much more suitable to the waveguide geometry[1]. It has been demonstrated that the simplified modes can be obtained by superposition of two nearly degenerate modes of the exact solution of the boundary value problem. The validity of this *weakly guiding* approximation[2] has been studied by Marcuse [27].

This section briefly reviews the theoretical foundation, main results and their physical interpretation of weakly guiding dielectric waveguides pertinent to fibre stellar interferometry.

[1] A rectangular coordinate system allows us to set one of the transverse ($x$ or $y$) vibrations to zero.

[2] Optical fibres are referred to as "weakly guiding" when the difference between refractive indices of the core and cladding $\Delta n << 1$.

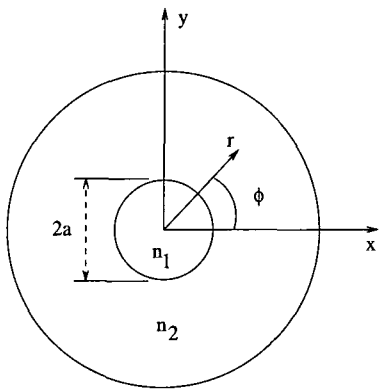



**Figure 3.1.** Circular core step–index fibre section definitions.

Referring to Figure 3.1, region 1 (the core) has a refractive index $n_1$ and region 2 (the cladding) has a refractive index $n_2$, where $n_2 < n_1$. This idealised structure is often called the *step–index matched–clad* fibre. The simplified analysis that follows assumes an infinitely extended cladding, which is justified in that the guided modes have exponentially decaying fields outside the core. As long as the cladding radius is large enough, the guided mode fields will have decayed to insignificant amplitudes at the outer cladding boundary. It is only when radiation modes are of interest that the infinite cladding assumption is questionable, since the solutions of the boundary value problem extend out to infinity so are most certainly affected by the cladding surface and its imperfections. Although the cladding's finite radius does trap some of the radiation field, thereby giving rise to cladding modes with a discrete frequency spectrum, as long as the number density of the cladding modes is effectively a continuum when compared to the guided modes' spectrum, the infinite cladding assumption is still valid.

### 3.2.1 Guided modes of the step–index optical fibre

A complete description of the modes within an optical fibre results from examining Maxwell's equations for non–magnetic materials, which may be written

in the form

$$\begin{aligned} \nabla \wedge \vec{H} &= \epsilon_0 n^2 \frac{\partial \vec{E}}{\partial t} \\ \nabla \wedge \vec{E} &= -\mu_0 \frac{\partial \vec{H}}{\partial t} \end{aligned} \tag{3.1}$$

where $\epsilon_0$ and $\mu_0$ denote the dielectric permittivity and magnetic permeability of vacuum respectively. Only strictly time harmonic fields, whose time and space dependencies are conventionally expressed as

$$\exp \mathrm{i}(\omega t - \beta z) \tag{3.2}$$

are considered here. The eigenvalue $\beta$ is referred to as the propagation constant and is equal to $nk$ in an homogeneous medium of refractive index $n$. $k \triangleq 2\pi/\lambda$ denotes the free space circular wavenumber. In the weakly guiding model, the guided modes lie within

$$n_2 k < \beta < n_1 k \tag{3.3}$$

so the condition $\beta = nk$ where $n \approx n_1 \approx n_2$ can be used.

The simplified guided modes are derived by first finding the longitudinal field components $E_z$ and $H_z$. The transverse components may be derived from them via Maxwell's equations. $E_z$ and $H_z$ are solutions of the reduced Helmholtz equation, which is

$$(\nabla^2 + \kappa_j^2)\psi = 0 \tag{3.4}$$

with

$$\kappa_j^2 = n_j^2 k^2 - \beta^2 \tag{3.5}$$

The index $j$ labels the appropriate waveguide region. To remain consistent with the notation used by Marcuse, the normalised wavenumbers will be relabelled so that $\kappa \triangleq \kappa_1$ and $\gamma \triangleq -\kappa_2$.

It is convenient to define a few universal parameters at this point

$$U \triangleq \kappa a \tag{3.6}$$

$$W \triangleq \gamma a \tag{3.7}$$

$$V \triangleq (U^2 + W^2)^{1/2} \tag{3.8}$$

$U$ describes the transverse mode propagation and $W$ its radial decay. $V$ is termed the normalised frequency; with Equation (3.5) it can also be written as

$$V \triangleq ak\sqrt{n_1^2 - n_2^2} \tag{3.9}$$

$a$ being the core radius. The waveguide numerical aperture is defined to be

$$NA \triangleq \sqrt{n_1^2 - n_2^2} \tag{3.10}$$

We can also define

$$\Delta \triangleq (n_1 - n_2)/n_1 \tag{3.11}$$

so that $V$ becomes

$$V = akn_1\sqrt{2\Delta} \tag{3.12}$$

in the weak guidance regime.

The field expressions which satisfy Equation (3.4) with the appropriate boundary conditions can be either Bessel or Neumann cylinder functions. Since they must remain finite at $r = 0$, they must be Bessel functions $J_\nu(x)$, of order $\nu$, in the core. In the cladding region, they are Hankel functions of the first kind $H_\nu^{(1)}(x)$, dictated by the condition that the field vanishes at $r \to \infty$. The general expressions are quite long, so we present only the fields of the fundamental modes to illustrate their form,

$$E_{y,x} = \pm \frac{Z_0}{n_2} H_{x,y} = E_0 \times \begin{cases} \dfrac{J_\nu(Ur/a)}{J_\nu(U)} & 0 \le r \le a \\ \dfrac{H_0^{(1)}(\mathrm{i}Wr/a)}{H_0^{(1)}(\mathrm{i}W)} & r \ge a \end{cases} \tag{3.13}$$

$$E_z = -\mathrm{i}\frac{E_0}{kan_2}(\sin\phi, \cos\phi) \begin{cases} \dfrac{UJ_1(Ur/a)}{J_0(U)} & 0 \le r \le a \\ \dfrac{-WH_1^{(1)}(\mathrm{i}Wr/a)}{\mathrm{i}H_0^{(1)}(\mathrm{i}W)} & r \ge a \end{cases} \tag{3.14}$$

$$H_z = -\mathrm{i}\frac{E_0}{kaZ_0}(\cos\phi, \sin\phi) \begin{cases} \frac{UJ_1(Ur/a)}{J_0(U)} & 0 \le r \le a \\ \\ \frac{-WH_1^{(1)}(\mathrm{i}Wr/a)}{\mathrm{i}H_0^{(1)}(\mathrm{i}W)} & r \ge a \end{cases} \tag{3.15}$$

Either $E_x$ or $E_y$ can be taken as 0. The first of $(\sin\phi, \cos\phi)$ for $E_z$ and of $(\cos\phi, \sin\phi)$ for $H_z$ holds if $E_x = 0$. The converse is true if $E_y = 0$. $Z_0$ is the vacuum impedance.

By equating the $E_z$ coefficients either side of the core boundary, we obtain the eigenvalue equation[3] of the simplified modes:

$$U\frac{J_{\nu+1}(U)}{J_\nu(U)} = -\mathrm{i}W\frac{H_{\nu+1}^{(1)}(\mathrm{i}W)}{H_\nu^{(1)}(\mathrm{i}W)} \tag{3.16}$$

The continuity of the transverse fields across the core interface is automatically provided for by the weak guidance condition. This condition also means that the longitudinal components can almost be neglected, such that any single mode can be considered as transversely polarised with a linear polarisation.

Equation (3.16) is extremely simple, especially when compared to the exact representation. However, this simplicity comes at a price. The exact equation has twice as many solutions as the simplified form. The propagation constants of the exact modal solutions[4] are nearly degenerate, becoming exactly the same in the limit $n_1 \to n_2$. The simplified (LP) modes can actually be shown to be a superposition of $HE_{\nu+1,\mu}$ and $EH_{\nu-1,\mu}$ modes [27] so that the total number of modes is still the same. So, although the simplified modes are a good description of the behaviour of fibre waveguides owing to the near degeneracy of the exact solutions, it is nevertheless prudent to remember that they are not modes in the usual sense of the word.

[3]Since we are following Marcuse, the radially decaying functions are Hankel functions of the first kind, $H_\nu^{(1)}(x)$. It should be noted that many authors choose to use the modified Bessel function [28], $K_\nu(x) \triangleq (\pi/2)\mathrm{i}^{\nu+1}H_\nu^{(1)}(\mathrm{i}x)$ instead.

[4]These are denoted as $HE_{\nu\mu}$ and $EH_{\nu\mu}$. The notation for the simplified modes, $LP_{\nu\mu}$ introduced by Gloge [29] is meant to be suggestive of "linearly polarised" modes. The subscript $\nu$ denotes the integer of the circular functions, $\cos\nu\phi$ and $\sin\nu\phi$. $\mu$ labels the roots of the eigenvalue equation for a given $\nu$.

### The fundamental mode

The point at which the various $LP$ modes cease propagating in a step–index profile can be calculated from (3.16). If we set $W = 0$ , the roots of (3.16) are the values [29] of $U$; by Equation (3.6), $U = V$.

So when $0 \leq V < 2.405$, we find that only the $LP_{01}$ mode can be maintained, thus defining the so–called "single mode" regime. Physically this condition means that the maximum guiding angle from Snell's law must be smaller than some diffraction angle proportional to $\lambda/a$. It can be used to define the single mode cutoff wavelength $\lambda_c$ from

$$V = 2.405\frac{\lambda_c}{\lambda} \tag{3.17}$$

The value of $\beta$ for the $LP_{01}$ mode is the largest of all the possible modes.

The eigenvalue equation (3.16), being transcendental, usually has to be solved by numerical techniques. However, an approximate closed–form solution is available near and far from cutoff for each mode.

Far from cutoff, $\gamma r >> 1$, so by considering the large argument approximation of the Hankel function

$$H_\nu^{(1)}(\mathrm{i}\gamma r) = \left(\frac{2}{\pi \mathrm{i}\gamma r}\right)^{1/2} \exp\left\{-\mathrm{i}\left[\nu\left(\frac{\pi}{2}\right) + \left(\frac{\pi}{4}\right)\right]\right\} \exp\left(-\gamma r\right) \tag{3.18}$$

it can be seen for large $\gamma r$ the field outside the core is proportional to $\exp(-\gamma r)$. The parameter $\gamma$ can therefore be used as an estimator of the mode radius, where the field intensity has decayed to $1/e$ of its core interface value. Defining this to be

$$\omega_0 \triangleq a + \gamma^{-1} \tag{3.19}$$

we can find the value of $\omega_0$ for a given $V$ value. $\gamma$ is available from the radial decay parameter $W$. It is useful to have an expression for the fundamental mode field radius near the cutoff condition.

Near cutoff $\gamma r << 1$, so we can use the approximations of the Hankel

functions for small arguments. For $\nu = 0$

$$H_0^{(1)}(\mathrm{i}\gamma r) = \frac{2\mathrm{i}}{\pi} \ln(\Gamma\gamma r/2) \tag{3.20}$$

with $\Gamma = 1.781672$. For $\nu \neq 0$ the approximation

$$H_\nu^{(1)}(\mathrm{i}\gamma r) = -\frac{\mathrm{i}(\nu-1)!}{\pi}\left(\frac{2}{\mathrm{i}\gamma r}\right)^\nu \tag{3.21}$$

holds. Substituting these expressions into the eigenvalue equation (3.16) yields

$$\gamma r = \frac{2}{\Gamma}\exp\left(\frac{-1}{\kappa r}\frac{J_0(\kappa r)}{J_1(\kappa r)}\right) \tag{3.22}$$

for $\nu = 0$. Hence, we find the near cutoff approximation for the $\nu = 0$ modes

$$W = \gamma a = \frac{2}{\Gamma}\exp\left[-\frac{1}{V}\frac{J_0(V)}{J_1(V)}\right] \tag{3.23}$$

For an optical fibre with a core radius $a = 2\mu$m, a numerical aperture of 0.1, transmitting $\lambda = 0.633\mu$m, the mode radius[5] $\omega_0 = 2.06\mu$m.

A useful relation between $\omega_0$ and $a$ is found when $V = 2.4$ in Equations (3.23) and (3.19),

$$\omega_0 = 1.1076a \tag{3.24}$$

### Gaussian approximation for the fundamental mode

The shape of the fundamental ($LP_{01}$) mode is similar to a Gaussian curve. This suggests an appropriate approximation of the exact field distribution. Additionally, a Gaussian profile is the exact field shape of the $LP_{01}$ mode in a non–truncated parabolic index profile. Marcuse [31] approximated the field to better than 1% accuracy by

$$\begin{aligned}\frac{\omega_0}{a} &= 0.65 + 0.434\left(\frac{\lambda}{\lambda_c}\right)^{3/2} + 0.0149\left(\frac{\lambda}{\lambda_c}\right)^6 \\ &= 0.65 + 1.619V^{-3/2} + 2.879V^{-6}\end{aligned} \tag{3.25}$$

[5]The approximations $J_0(x) \approx 1 - 0.25x^2 + 0.025x^3$ and $J_1(x) \approx 0.5815\sin(\pi x/3.7)$ have been used [30].

with $\omega_0$ being the mode field radius. For $V = 2.4$ this reduces to

$$\omega_0/a \approx 1.1 \tag{3.26}$$

which agrees with Equation (3.24). The most restrictive part of this approximation is the evanescent field far from the core–cladding interface, where the field is underestimated by the Gaussian form.

### 3.2.2 Radiation modes

The number of guided modes supported by a waveguide core is always finite. Therefore other solutions of Maxwell's equations exist that satisfy the boundary conditions in order to provide a complete set of orthogonal modes. A physical argument is useful in determining the nature of the various types of modes. If there is a source of radiation outside the waveguide core (as in Figure 3.2), it is trivial to see that such a source cannot contribute to the guided modes by the application of Snell's law. The energy emitted by the source is reflected and refracted at the core–cladding boundary without being trapped by the core. It would require imaginary angles to inject a ray from the outside into the core in such a situation. That is to say, a ray refracted into the core must always exceed the critical angle. Imaginary geometric angles correspond to evanescent fields, which can tunnel into the core from the outside. However, they themselves must be evanescent field tails due to total internal reflection, so clearly cannot result from the source.

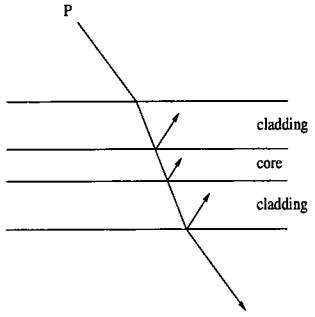



**Figure 3.2.** Source above waveguide core

A plane wave incident on the core from above will have some of its energy reflected above the core and some transmitted through to the other side. Above the core, the total field becomes a standing wave. The resulting radiation field is a solution of Maxwell's equations with the appropriate boundary conditions. It qualifies as a mode, but unlike the guided modes, it reaches undiminished to infinite distances in the direction normal to the core. The propagation constants $\beta$ of such radiation modes are not constrained to a discrete set of values, since they are related to the (arbitrary) angle of the incident wave. They form a continuum; hence an alternative name for them is continuum modes.

Naturally, radiation phenomena around the core region are described by these modes. We know that radiation originating inside the core results in travelling wave fields in the cladding and outer regions. On the other hand, we have seen that the radiation modes must be standing waves, at least on one side of the waveguide core. This seeming paradox is resolved by recalling that the radiation modes form a continuum. It is impossible to excite one single radiation mode with a source inside the waveguide core[6]. Any mechanism that excites a continuum mode always excites infinitely many continuum modes in the surrounding $\beta$ space. Only in the unrealistic case of an infinitely extended source positioned at infinity could a pure radiation mode be excited. So, if radiation is excited by a waveguide imperfection, it will consist of infinitely many radiation modes which interfere in such a way that the incoming parts of the standing wave are eliminated by destructive interference.

### 3.2.3 Graded Index fibres

Manufacturing imperfections can introduce some grading of the core–cladding interface and/or some index depression at the centre of the core[7]. Evaluation of the transmission characteristics of such fibres can be undertaken by *equivalent*

[6]There are no normalisable solutions of Maxwell's equations satisfying the boundary conditions at the core interface while forming only travelling waves outside the core.

[7]In the communications industry, the main driver for the production of SM optical fibres, index gradients in the core are sometimes intentionally induced to obtain specific characteristics such as longer zero chromatic dispersion wavelengths; specifically shifting from 1300nm to 1550nm.

*step–index* (ESI) methods, instead of directly solving the scalar wave equation for the general case as treated by [32, 33].

Consider a general weakly guiding index profile of the form

$$\begin{aligned} n^2(r) &= n_2^2\left[1 + 2\Delta\, h\left(\frac{r}{a}\right)\right] \\ &= n_2\left[n_2 + 2\Delta n(r)\right] \end{aligned} \tag{3.27}$$

where

$$\begin{aligned} h\left(\frac{r}{a}\right) &= 1 - \left(\frac{r}{a}\right)^g, \qquad 0 \le r \le a \\ &= 0, \qquad\qquad\quad r > a \end{aligned} \tag{3.28}$$

This corresponds to a rounding of the core–cladding interface. Values of $g > 8$ can be considered as typical defects in the modified chemical vapour deposition (MCVD) manufacturing process.

We require an ESI fibre with parameters $a_{\mathrm{ESI}}, \Delta_{\mathrm{ESI}}$ and $V_{\mathrm{ESI}}$ which has an $LP_{11}$ mode cutoff wavelength $\lambda_{c(\mathrm{ESI})}$, $LP_{01}$ propagation constant $\beta_{\mathrm{ESI}}$ and mode spot radius as close as possible to the real fibre parameters — at least over a restricted wavelength range such as $0.8 \le \lambda/\lambda_c \le 1.5$.

Jeunhomme [30] gives the parameters as

$$\frac{V_{\mathrm{ESI}}}{V} = \left(\frac{g}{g+2}\right)^{1/2} \tag{3.29}$$

$$\frac{a_{\mathrm{ESI}}}{a} = \frac{g+2}{g+3} \tag{3.30}$$

where the expression for $V$ is as (3.9) but with $n_1$ replaced by $n_2(1 + \Delta)$.

The waveguide characteristics of a graded core fibre can therefore be approximated fairly well by ESI methods. This approximation fails in describing waveguide dispersion, since the graded profile is too far from a step over a large wavelength range. It should be noted that central index dips, another common feature of communications fibres, have little influence on the $LP_{11}$ mode cutoff, but strongly increase the mode field radius compared to the step–index fibre. This effect increases the microbending loss sensitivity of fibres.

| g | Exact $V_c$ [32] | Eqn (3.29) |
|---|---|---|
| 1 | 4.381 | 4.166 |
| 2 | 3.518 | 3.401 |
| 3 | 3.181 | 3.105 |
| 4 | 3.000 | 2.946 |
| 5 | 2.886 | 2.846 |
| 8 | 2.710 | 2.689 |
| 10 | 2.649 | 2.635 |
| 20 | 2.527 | 2.522 |
| $\infty$ | 2.405 | 2.405 |

**Table 3.1.** Normalised cutoff frequencies $V_c$ for index profiles $h(x) = 1 - x^g$.

The guided modes described exist provided that the waveguide structure is free of imperfections. If the dielectric material is lossy (through absorption) , the propagation constant $\beta$ becomes complex. However, for reasonably low loss where the core and cladding losses are the same[8], one can still use the solutions derived thus far.

## 3.3 Birefringence properties

Up to now, the single mode fibre has been studied assuming it to be perfectly circular with a circularly symmetric refractive index distribution. Real fibres, though, exhibit some core ellipticity and/or anisotropy in the refractive index distribution due to anistropic stresses. This results in each polarisation component acquiring a different $\beta$, leading to perturbations of the state of polarisation (SOP) of the transmitted light.

Any interferometric device requires that the interacting beams have identical polarisations for efficient operation; if the two beams are orthogonally polarised, complete fading of the interference signal will result. Control of the SOPs is crucial; a large modal birefringence is required for this. The presence of such birefringence breaks the degeneracy of the various modes, preventing them from mixing in a random fashion.

[8]so that the weakly guided mode losses are equal to plane wave loss.

To this end, we now briefly discuss both intrinsic and induced birefringence in optical fibre waveguides. For intrinsically birefringent weakly guiding fibres, we demonstrate that the solutions to the scalar Helmholtz equation provide a sufficiently accurate analytic description without recourse to the full vector field treatment.

### 3.3.1 Intrinsically birefringent fibres

We have seen that so–called "single mode" fibres are in fact bimodal, propagating two nearly degenerate modes with orthogonal polarisations; $HE_{11}^{x}$ and $HE_{11}^{y}$. The principal axes are determined from symmetry elements of the cross–section, so the larger the cross–sectional anisotropy, the greater the difference between $\beta_x$ and $\beta_y$. If the fibre cross–section is independent of length $z$, then the fibre acts as a birefringent medium with a modal birefringence $B$ given by [34]

$$B = \frac{(\beta_x - \beta_y)}{k} \tag{3.31}$$

Polarisation states travelling along one of the principal axes will be retained for all $z$. Field vectors polarised at an angle $\theta$ with respect to the x–axis will pass through various states of elliptic polarisation as the phase retardation

$$\Phi(z) = (\beta_x - \beta_y)z \tag{3.32}$$

varies with length, assuming that the two normal mode components maintain phase coherence. The length $z$ that corresponds to $\Phi(z) = 2\pi$ is the beat length between polarisations

$$z_b = \lambda/B \tag{3.33}$$

This is directly observable by means of Rayleigh (dipole) scattering from the fibre. The radiation pattern of a dipole is minimum along the axis and maximum normal to the axis, so a fibre viewed along the direction of the incident polarisation will exhibit a regular variation in intensity with a period $L$. It is possible to determine $B$ directly from the observed beat length.

Modal birefringence $B$ results from two components, a geometric component

$G$ and an effective material birefringence $M$ due to strain. These contributions may have the same or opposite signs, and the effect of $M$ is an average, weighted by the square of the mode eigenfunction, over the material birefringences of the core and cladding.

If a well–defined linear polarisation is required, it may be achieved by maximising $B$, i.e. maximising $M, G$ or both $M$ and $G$.

A case of particular interest analytically is the elliptical fibre (Figure 3.3) of infinite parabolic refractive index profile. Although unphysical since $n^2(x,y) \to \infty$ as $(x,y) \to \infty$ it does lead to simple expressions for modal quantities. The weak guidance approximation is accurate here provided that modal power is concentrated close to the fibre axis.

The infinite parabolic profile for an elliptical fibre is

$$n^2(x,y) = n_1^2 \left\{ 1 - 2\Delta \left[ \frac{x^2}{a_x^2} + \frac{y^2}{a_y^2} \right] \right\} \; ; \; -\infty < x, y < \infty \tag{3.34}$$

where $a_x$ and $a_y$ are length scales along the respective coordinate axis. Contours of constant index form a set of concentric ellipses, so that every ellipse has the same eccentricity

$$e = (1 - a_y^2/a_x^2)^{1/2} \; ; \; a_x > a_y \tag{3.35}$$

Since there are two orthogonal length scales, it is necessary to generalise the universal waveguide parameters $V \to V_x, V_y$ ; $U \to U_x, U_y$ ; $W \to W_x, W_y$. From Equations (3.6) and (3.12) these are

$$\begin{aligned} V_{x,y} &= a_{x,y} k n_1 \sqrt{2\Delta} \\ U_{x,y} &= \left( \frac{V_{x,y}^2}{2\Delta} - a_{x,y}^2 \beta^2 \right)^{1/2} \\ W_{x,y} &= \left( a_{x,y}^2 \beta^2 - \frac{V_{x,y}^2}{2\Delta n_1^2} \, n_2^2 \right)^{1/2} \end{aligned} \tag{3.36}$$

Finding the exact propagation constants for the $x$ and $y$ polarisations would

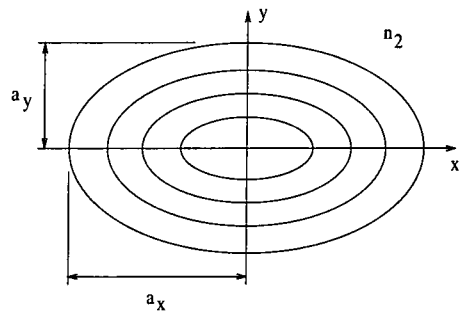



**Figure 3.3.** Elliptical fibre of infinite parabolic profile. Contours of constant refractive index are shown, as defined by Equation (3.34).

require solving the homogeneous vector wave equations

$$(\vec{\nabla}_t^2 + n^2k^2 - \beta^2)\vec{E} = -(\vec{\nabla}_t + \mathrm{i}\beta\hat{\mathbf{z}})\vec{E}_t \cdot \vec{\nabla}_t \ln n^2 \tag{3.37}$$

$$(\vec{\nabla}_t^2 + n^2k^2 - \beta^2)\vec{H} = \{(\vec{\nabla}_t + \mathrm{i}\beta\hat{\mathbf{z}}) \wedge \vec{H}_t\} \wedge \vec{\nabla}_t \ln n^2 \tag{3.38}$$

The subscript $t$ indicates the transverse field components and $\hat{\mathbf{z}}$ is the unit vector along the fibre axis. This is non–trivial due to the vector Laplacian operator, since it involves computing an infinite set of Mathieu functions of the second kind [35, 36] to account for the field in the cladding. This raises the problem of how to terminate this infinite set without introducing substantial numerical error. However, the $\vec{\nabla}_t \ln n^2$ term is small for weakly guiding fibres, so simple perturbation methods can be used to find the exact values of $\beta_x$ and $\beta_y$ from the solutions to the scalar Helmholtz equation. In Cartesian form this is

$$\left\{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + (k^2n^2(x,y) - \beta^2)\right\}\psi = 0 \tag{3.39}$$

Substituting Equation (3.34) into (3.39), we can make a change of variables and express the solution in separable form

$$\psi = F(\alpha_x)G(\alpha_y)\exp\left(-\frac{\alpha_x^2 + \alpha_y^2}{2}\right) \;;\quad \alpha_x = \frac{x}{a_x}\sqrt{V_x},\ \alpha_y = \frac{y}{a_y}\sqrt{V_y} \tag{3.40}$$

This leads to a pair of ordinary differential equations

$$\frac{d^2F}{d\alpha_x^2} - 2\alpha_x\frac{dF}{d\alpha_x} + \left(\frac{U_x^2}{2V_x} + c - 1\right)F = 0$$
$$\frac{d^2G}{d\alpha_y^2} - 2\alpha_y\frac{dG}{d\alpha_y} + \left(\frac{U_y^2}{2V_y} - c - 1\right)G = 0 \tag{3.41}$$

where $c$ is the separation constant. The transverse scalar solutions which ensure $\psi$ is bounded as $|x|, |y| \to \infty$ are the Hermite polynomials $H_n$ [37]. The multiplying factors for $F$ and $G$ must be even positive integers or zero. The associated eigenvalue equations are found by eliminating $c$. We have then

$$\begin{aligned}\psi_{mn} &= H_{m-1}\left(\frac{x}{a_x}V_x^{1/2}\right)H_{n-1}\left(\frac{y}{a_y}V_y^{1/2}\right) \\ &\times \exp\left\{-\frac{1}{2}\left(\frac{V_xV_y}{a_xa_y}\right)^{1/2}\left(\frac{x^2}{a_x} + \frac{y^2}{a_y}\right)\right\} \qquad (3.42)\\ U_xU_y &= (2m-1)V_y + (2n-1)V_x \qquad (3.43)\end{aligned}$$

where $m, n = 1, 2, ...$ and $H_n$ are the Hermite polynomials. Substituting Equation (3.42) into (3.39) gives the scalar value of $\beta$. Polarisation corrections $\delta\beta_{x,y}$ can be added to this to obtain the exact values.

The order of magnitude of $\delta\beta_{x,y}$ can be estimated by considering the relationship between $\beta$, $V$ and $U$,

$$\beta = \frac{1}{a}\left(\frac{V^2}{2\Delta} - U^2\right)^{1/2} \tag{3.44}$$

Denoting the variables relating to the exact eigenfunctions by the subscript $e$ and rearranging,

$$\beta_e = \frac{V}{a\sqrt{2\Delta}}\left(1 - 2\Delta\frac{U_e^2}{V^2}\right)^{1/2} \tag{3.45}$$

We can expand this binomially

$$\beta_e = \frac{V}{a}\left(\frac{1}{\sqrt{2\Delta}}\right) - \frac{U_e^2}{aV}\left(\frac{\Delta}{2}\right)^{1/2} - \frac{U_e^4}{aV^3}\left(\frac{\Delta}{2}\right)^{3/2} \tag{3.46}$$

Since $\Delta << 1$, it is possible to assume an expansion of $U_e$ such that

$$U_e(V, \Delta) = U + \Delta U' + \ldots \tag{3.47}$$

where $U_e = U$ when $\Delta = 0$. Using this with the approximation

$$\Delta \approx V^2/2U^2 \tag{3.48}$$

which is valid because of the weak guidance, allows us to write

$$\beta_e = \frac{V}{a}\left(\frac{1}{\sqrt{2\Delta}}\right) - \frac{U^2}{aV}\left(\frac{\Delta}{2}\right)^{1/2} - \frac{U^4 + 4UV^2U'}{aV^3}\left(\frac{\Delta}{2}\right)^{3/2} \tag{3.49}$$

Similarly, for the scalar propagation constant,

$$\beta = \frac{V}{a}\left(\frac{1}{\sqrt{2\Delta}}\right) - \frac{U^2}{aV}\left(\frac{\Delta}{2}\right)^{1/2} - \frac{U^4 - 4UV^2U'}{aV^3}\left(\frac{\Delta}{2}\right)^{3/2} \tag{3.50}$$

Thus the magnitude of the polarisation corrections is

$$\delta\beta = \beta_e - \beta \sim \Delta^{3/2} \tag{3.51}$$

### 3.3.2 Induced birefringence

In real circular core fibres, the output state of polarisation drastically depends on external stresses imposed by bends, twists and clamping[9]. It is known that bending a fibre induces linear polarisation retardation and that twisting induces circular birefringence (or optical activity).

Bending causes three effects :

1. CORE ELLIPTICITY

   To first order, a circular core becomes more elliptical, the ellipticity being given by $e = \nu a/R$. Poisson's ratio[10] $\nu$ is listed in the literature as 0.16

[9]We consider the effects of temperature, hygrometry and atmospheric pressure to be negligible in this study.

[10]The absolute ratio of (a) the transverse to (b) the longitudinal normal strains in a strain field produced by a uniaxial stress along the longitudinal axis. The value is always between 0.0 and 0.5 for materials having isotropic mechanical properties.

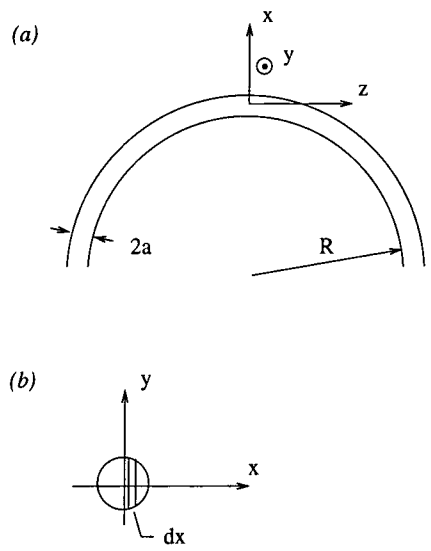



**Figure 3.4.** Variable and coordinate definitions for the bending geometry of a fibre.

for silica. The retardation per metre $B_e$ expected for a fibre with $V = 2.4$ is given by [38]

$$B_e = 0.125(e^2/r)(2\Delta)^{3/2} \tag{3.52}$$

This constant is very small, so this is not a major effect.

2. WAVEGUIDE GEOMETRY CHANGE

The non–uniform geometry lifts the degeneracy of the two $HE_{11}$ modes, so that the modal propagation constants vary with the inverse square of the bend radius $R$. The magnitude of linear retardation has not been calculated but is considered negligible because of the last effect,

3. STRESS

This modifies the core refractive index through the elasto–optic effect and is the principal source of the uniaxial negative birefringence induced by bending [39, 40]. It results from the lateral, compressive stress $-\sigma_x$ that builds up in a bent fibre under "large" deformations. The first order stress component, given by $\sigma_z = Ex/R$ where $E$ is Young's modulus and $R$ is the radius of curvature, does not contribute directly to the birefringence

due to the fourfold symmetry of the core deformation.

Ulrich et al. calculated lateral stress in a cross–section $z = 0$, assuming an elastically homogeneous and isotropic fibre in the limit of weak curvature, $r/R >> 1$. The layer $dx$ in Figure 3.4 is bent with a radius of $(R + x)$. The tension $\sigma_x(x)$ contributes to the lateral stress component $\sigma_x$ with

$$\sigma_x(x + dx) - \sigma_x(x) = \sigma_z(x)dx/(R + x) \tag{3.53}$$

with $|\sigma_x| << |\sigma_z|$. Neglecting $|x| << R$, we obtain

$$\frac{\partial \sigma_x}{\partial x} = \frac{Ex}{R^2} \tag{3.54}$$

Integrating this expression along the $x$ axis with $\sigma_x(\pm r) = 0$ at the fibre surfaces,

$$\sigma_x(x) = \frac{E}{2R^2}(x^2 - r^2) \tag{3.55}$$

The bending induced birefringence is given by $B_b = \beta_x - \beta_y \approx k(\delta n_x - \delta n_y)$. The elasto–optic index changes $\delta n_x$ and $\delta n_y$ are evaluated from the stress at the fibre axis $\sigma_x(x = 0)$ by [41]

$$\delta n_i = -\frac{n^3}{2}\Sigma p_{ij}\epsilon_j \tag{3.56}$$

where the $p_{ij}$ are the strain–optical coefficients and the $\epsilon_j$ are the components of strain.

Using Poisson's ratio $\nu$, we have

$$\epsilon_x - \epsilon_y = \frac{(\sigma_x - \sigma_y)(1 + \nu)}{E} \tag{3.57}$$

Therefore the constant of birefringence is given by

$$\begin{aligned} B_b &= 0.25kn^3(p_{11} - p_{12})(1 + \nu)\left(\frac{a}{R}\right)^2 \\ &= -0.135k\left(\frac{a}{R}\right)^2 \text{ radians/m} \end{aligned} \tag{3.58}$$

for pure fused silica[11]. Although this effect is small in comparison to crystal birefringence, because it is integrated along the fibre it becomes quite significant.

Twisting a fibre around its axis induces a shear stress that leads to a circular birefringence [42]. Three typical situations may be distinguished depending on the relative magnitudes of the twist rate $\Upsilon$ and the intrinsic linear birefringence $B$. In the case of weak twist, $|\Upsilon| << |B|$ so linear birefringence dominates, thus the polarisation is altered geometrically. At medium twist, $|\Upsilon| \approx |B|$ so the shear strain in the silica fibre gives rise to a circular birefringence proportional to the twist $B_t = (p_{11} - p_{12})\Upsilon/2$. So a rather complicated elliptical polarisation evolution results from both the linear and circular birefringence effects. Finally the situation simplifies again for $|\Upsilon| >> |B|$, as the effect of optical activity dominates.

## 3.4 Adiabatically tapered SM fibres

As mentioned in the introduction to this Chapter, adiabatically tapered fibres can be a useful tool in simplifying the issue of input coupling to SM waveguides. It is evidently more difficult to align an incident beam accurately with normal SM fibres, whose core radii range between 1 and 3$\mu$m, than with larger core fibre of 10$\mu$m radius, say.

The favoured method of describing tapered waveguides is to use local normal modes. This assumes that the mode at any point along the taper is equal to the equivalent mode of a straight fibre with the same core diameter.

Using this assumption we can postulate that, for a perfect taper, the mechanism for energy loss from the fundamental mode is up–conversion to the next closest guided mode. This mode has, therefore, the next closest valued propagation constant to the fundamental.

We can define a coupling or "beat" length $z_b$ between the local modes $LP_{01}$

[11] We take $n = 1.46, (p_{11} - p_{12}) = -0.15$ at $\lambda = 633$nm.

and $LP_{11}$ (the first harmonic) of a fibre section [43]

$$z_b = \frac{2\pi}{(\beta_{01} - \beta_{11})} \tag{3.59}$$

Let us define the local taper length scale $z_t$ as the height of a cone which has its base coincident with the local core cross–section. If $z_t >> z_b$, mode conversion and loss will be minimal. Conversely, significant energy coupling will occur if $z_t << z_b$. A delineation between the two regimes can be expressed as

$$\tan\Omega = \frac{da(z)}{dz} \leq \frac{a(z)}{z_b} \tag{3.60}$$

Of course, the power loss from the fundamental mode due to tapering is not quantified, so this delineation can only be an approximate criterion.

Hence, it is possible to reach any level of beam expansion theoretically provided that no other perturbations are introduced. This means that the taper can act as a self–aligned, low loss microlens for a SM fibre, reducing the number of degrees of freedom in a fibre–based instrument.

For such a fibre, the waveguide numerical aperture $NA$ relates to the spatial parameters through the conservation of energy (system étendue). So

$$A_i\Omega_i = A_o\Omega_o \tag{3.61}$$

where $A_{i,o}$ represent the input and output aperture areas respectively and $\Omega_{i,o}$ are the acceptance solid angles. We can see from this that

$$\Omega_{i,o} = n_{i,o}^2 \sin^2(\theta_{i,o}) = (NA_{i,o})^2 \tag{3.62}$$

and

$$A_i/A_o = R^2 \tag{3.63}$$

where $R$ is the taper core diameter ratio, so

$$NA_o/NA_i = R \tag{3.64}$$

| Fibre i.d. | Standard core diameter $[2a\mu m]$ | Tapered core diameter $[2a_T\mu m]$ | Taper length [m] |
|---|---|---|---|
| TAPS18/6/1/3 | $6.5 \pm 1$ | $19 \pm 2$ | 1.0 |
| TAPH18/6/1/3 | $5 \pm 1.5$ | $18 \pm 2$ | 1.0 |

**Table 3.2.** Manufacturer's data for tapered fibres. TAPS refers to the standard taper, whereas TAPH refers to the high birefringence (SMTE) taper. The cutoff wavelength for the TAPS fibre was $\lambda = 585 \pm 10$nm.

## 3.5 Experimental tests

To begin with, it was necessary to test the robustness of the preceding analysis of tapered fibres experimentally, in terms of the mode profiles (from both ends) and the overall throughputs. The latter property has a very strong dependence on the input coupling efficiency, which is dependent on the fibre mode profile.

Two sample fibres were obtained from Oxford Electronics Ltd, their (manufacturer provided) details being shown in Table 3.2. Fibre TAPS was a standard circular core tapered fibre with a 3:1 taper ratio. Fibre TAPH was the first attempt (to the author's knowledge) at producing an adiabatically tapered fibre with a high degree of intrinsic birefringence. No information regarding the refractive index profiles was available.

The approximate $f_{\#}$'s of the single mode end and the tapered end of the waveguides can be deduced. The waveguide $NA$ can be found through Equation (3.9) and the values in Table 3.2. $n_1$ is assumed to be 1.459 (the value for fused silica at 585nm) and $V$ is taken to be 2.7 (see Section 3.2.3). Since

$$f_{\#} \approx \frac{1}{2NA} = \frac{ak}{2V} \tag{3.65}$$

we have $f_{\#} \approx 6$. At the tapered end, using Equation (3.64), we have that the $f_{\#} \approx 18$.

The two fibres were characterised prior to deciding on the components for DAFI's interferometer arms.

The first tests of the fibres was to examine their field profiles. The fibres were illuminated by a 15mW, $\lambda = 633$nm HeNe laser, with a microscope objective as

the input coupling lens. In order to maximise the input coupling, the required demagnification was calculated with the following expression,

$$M = f/160 \tag{3.66}$$

where $M$ is the objective glass' magnification and $f$ (in mm) is found from the fibre $f_{\#}$, together with the HeNe beam diameter[12].

The width of the HeNe beam was measured by simply moving a knife edge in steps across the front of a p–i–n diode (which was in series with a resistor to provide a linear signal) and recording the detected intensity per step. By differentiating the measured function and fitting a Gaussian to the result (the fringing being at a level low enough to be neglected), a diameter of 0.8mm was estimated. So, the ideal objective glass (O.G.) magnifications required were $M = 13.3$ for the tapered end and $M = 40$ for the standard end of fibre TAPS. The closest commercially available O.G. for the tapered end was $M = 10$; the non–tapered end was not a problem as $M = 40$ is a standard magnification. It was not so clear as to whether the value of 13.3 was applicable to fibre TAPH since the fibre's fundamental mode profile was essentially elliptical, whereas the incident field had a circular Gaussian cross section.

### 3.5.1 TAPS18/6/1/3

**Figure 3.5. Left:** The output profile of the standard (f/6) end. **Centre:** An overexposed image of the tapered (f/18) end's beam profile. The first diffraction ring from the cladding modes is clearly evident. **Right:** The same profile with a longer exposure time. The second diffraction ring is just visible.

[12]The value of 160mm is the standard microscope tube length.

The fibre was illuminated and the field profiles from both ends were recorded photographically. The results are shown in Figure 3.5. The cladding modes are strongly evident, showing up as diffraction rings and slightly broadening the central disc. It was straightforward to remove these modes by immersing the coiled section of the fibre in glycerol[13]. The $f_{\#}$'s were confirmed (by eye) to be in agreement with the predicted values.

The throughput of this fibre was measured with a Newport power meter. The fibre was illuminated through its standard end, which was loosely coiled with a radius of 40cm. 1.5m from its output was kept as straight as possible, to avoid losses from the taper. Ten measurements were made, giving a throughput of 91% ±5%. When illuminated through its tapered end, the throughput was measured to be 88% ±5%, the lower value resulting from an imperfect match of the coupling O.G.

The throughput tests on this fibre demonstrated the single mode nature of the taper and its ablility to efficiently couple circular Gaussian beams.

### 3.5.2 TAPH18/6/1/3

**Figure 3.6. Left:** An overexposed image of the standard end beam profile. The profile is quite clearly elliptical, with an ellipticity of approximately 0.6. **Centre:** The beam profile from the taper end. The image was overexposed in an attempt to bring in the cladding modes, which do not appear. **Right:** The same profile seen through a polarisation analyser. The axis of the analyser is horizontal (across the image).

The throughput of this fibre, when illuminated through its standard core, was measured to be 65% ±5%, using the same method as described for fibre

[13]Glycerol acts as a cladding mode stripper because its refractive index is close to that of silica.

TAPS. Illumination from the tapered end did not result in a high throughput ($\sim$ 5%) despite attempts at polishing and recleaving the non–tapered (output) end. This is attributed to an extreme mismatch between the incident beam profile (circular Gaussian) and the waveguide modal profile in addition to the non–ideal microscope objective magnification. Figure 3.6 shows the exit beam profiles of the fibre. We deduce from the profile measurements of the standard end that the intrinsic birefringence is created by making an elliptical core with an approximate eccentricity of 0.63.

A polarisation analyser was placed in front of the output beam and a sequence of images were photographed with the analyser oriented in different directions. It appeared that the polarisation varied slightly across major axis of the beam profile, although this was not quantifiable. The beam profile from the tapered end was shown not to arise from the cladding modes by immersing the coiled fibre into glycerol. The fibre was also illuminated by an off–axis beam, producing the same profile; if the nulls had been caused by the presence of *cladding* modes, an off–axis beam should have excited them more strongly than the core mode(s).

The most probable cause of this non–fundamental[14] mode output would be incorrect taper angles. It is believed that the tapering of the fibre along one axis (the minor axis of the fibre) was not adiabatic, with respect to the fundamental mode, allowing a higher order mode to be supported in that direction in addition to the fundamental mode. This would suggest that the mode structure is non–single mode. In order to experimentally confirm this, it would be necessary to illuminate the fibre waveguides with a range of (longer) wavelengths until no nulls in the output field profile are seen. This was not done as the appropriate equipment was not available. Also, since the field profile seemed to be composed of more than the fundamental mode, it was decided that the standard SM taper would be used in the interferometer arms. This is necessary to avoid the mixing of modes referred to in Section 3.3.

[14] "Fundamental mode" in this case means that the wave function does not exhibit a null in either the $x$ or $y$ direction.

## 3.6 Adiabatically tapered elliptical SM fibres

Since the experimental observations of fibre TAPH did not produce the expected results, SM tapered elliptical (SMTE) fibres were studied, to see if any design parameters for the taper angle could be identified. SMTE fibres would be advantageous as a way of controlling the SOP without having to use external devices.

From Equations (3.42) it is readily verified that the fundamental mode solution is given by

$$\psi_{11} = \exp\left\{-\frac{1}{2}\left(\frac{V_x V_y}{a_x a_y}\right)^{1/2}\left(\frac{x^2}{a_x}+\frac{y^2}{a_y}\right)\right\} \tag{3.67}$$

The first harmonic solutions are

$$\psi_{21} = 2\left(\frac{x}{a_x}\sqrt{V_x}\right)\exp\left(-\frac{1}{2}\left(\frac{V_x V_y}{a_x a_y}\right)^{1/2}\left(\frac{x^2}{a_x}+\frac{y^2}{a_y}\right)\right) \tag{3.68}$$

$$\psi_{12} = 2\left(\frac{y}{a_y}\sqrt{V_y}\right)\exp\left(-\frac{1}{2}\left(\frac{V_x V_y}{a_x a_y}\right)^{1/2}\left(\frac{x^2}{a_x}+\frac{y^2}{a_y}\right)\right) \tag{3.69}$$

Substituting these solutions into Equation (3.39) provides the expressions for the respective $\beta$'s,

$$\beta_{11} = kn_1\left\{1-\frac{2\Delta}{\sqrt{V_x V_y}}\left[\left(\frac{a_y}{a_x}\right)^{1/2}+\left(\frac{a_x}{a_y}\right)^{1/2}\right]\right\} \tag{3.70}$$

$$\beta_{21} = kn_1\left\{1-\frac{2\Delta}{\sqrt{V_x V_y}}\left[3\left(\frac{a_y}{a_x}\right)^{1/2}+\left(\frac{a_x}{a_y}\right)^{1/2}\right]\right\} \tag{3.71}$$

$$\beta_{12} = kn_1\left\{1-\frac{2\Delta}{\sqrt{V_x V_y}}\left[\left(\frac{a_y}{a_x}\right)^{1/2}+3\left(\frac{a_x}{a_y}\right)^{1/2}\right]\right\} \tag{3.72}$$

From Equation (3.51), the polarisation corrections are small since $\Delta \approx 10^{-3}$ typically. They can therefore be neglected for these purposes.

Using the local normal mode assumption, we can postulate that, for a perfect taper, the mechanism for energy loss from the fundamental mode is up-conversion to the next closest guided mode. This mode has therefore, the next closest valued propagation constant to the fundamental. We define two compo-

nent "beat lengths" between the fundamental and the first overtone along each axis,

$$z_{bx} = \frac{2\pi}{(\beta_{11} - \beta_{21})} \tag{3.73}$$

$$z_{by} = \frac{2\pi}{(\beta_{11} - \beta_{12})} \tag{3.74}$$

As in Section 3.4, define an adiabaticity criterion for each axis

$$\tan \Omega_x = a_x(z)/z_{bx} \tag{3.75}$$

$$\tan \Omega_y = a_y(z)/z_{by} \tag{3.76}$$

where $\Omega_{x,y}$ is the taper angle along the $x$ or $y$ coordinate axis.

Equation (3.35) set the $x$ coordinate axis as being parallel to the semi–major axis. Taking $V_{x,y} < 2.405$ as the single–mode constraint gives both the maximum taper ratio $R$ and the maximum allowable refractive index difference, $\Delta$. Our elementary argument shows that the permitted values of $\Delta$ are governed by the semi–major length $a_x$. Figure 3.7 plots the range of maximum $\Delta$ allowed for the fibre to remain monomodal at $\lambda = 0.585\mu$m, $\lambda = 0.633\mu$m and $\lambda = 1\mu$m as the value of $a_x$ changes.

The variations of the permitted taper angles $\Omega_x$ and $\Omega_y$ with their respective linear dimensions are plotted in Figure 3.8 for an eccentricity of 0.63. Since $\Omega_y$ is always smaller than $\Omega_x$ it can be deduced that the maximum taper angle is solely dependent on the semi–minor length $a_y$ rather than $a_x$. This result is consistent with the heuristic argument drawn from the experimental tests of fibre TAPH.

The conclusion based on this analysis is that a tapered elliptical fibre of given core length scales $a_x$ and $a_y$ must satisfy the condition

$$\tan \Omega_y < a_y(z)/z_{by} \tag{3.77}$$

in order to be single mode. This is true for any point below the curves for $\Omega_y$ in Figure 3.8 at the relevant wavelength.

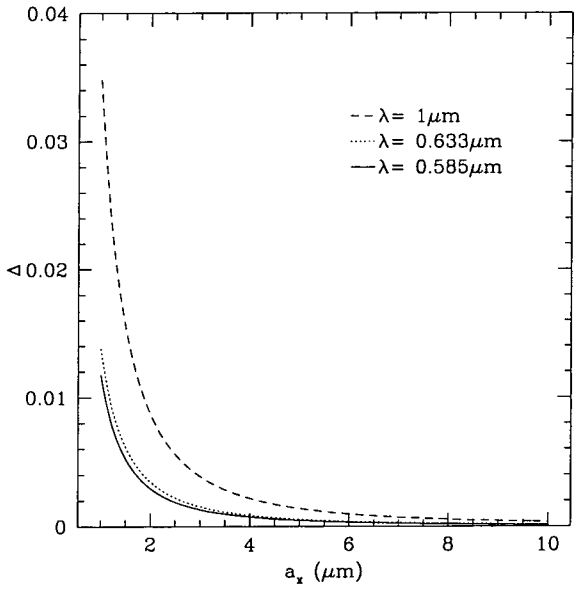



**Figure 3.7.** Range of maximum permissible $\Delta$ for $V_x(\lambda) < 2.405$ with various cutoff wavelengths.

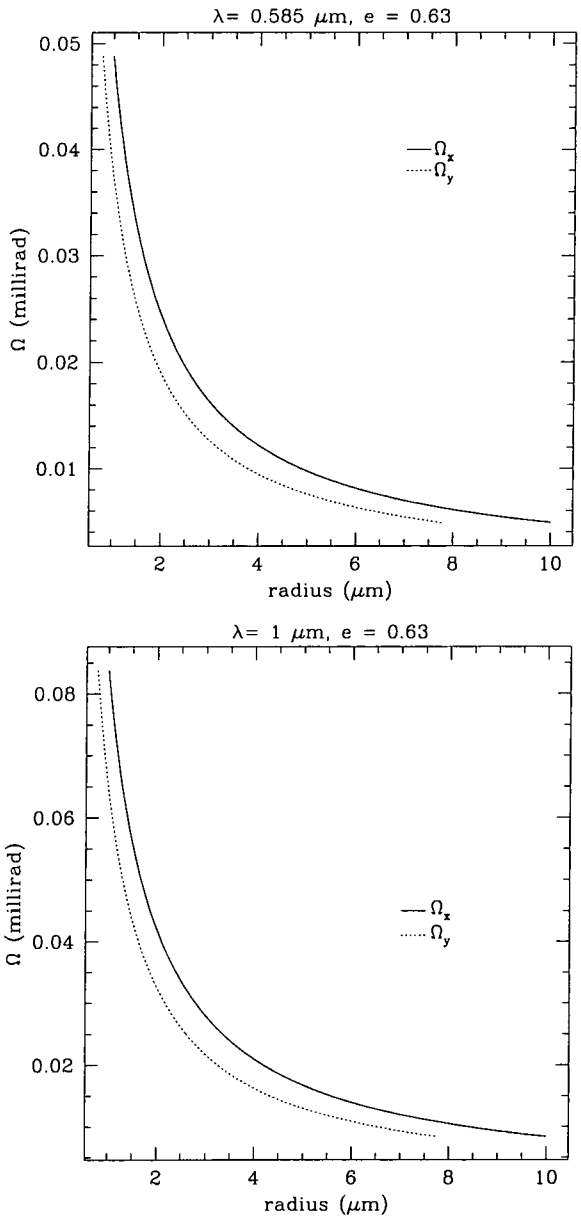



**Figure 3.8.** Variation of core taper angle $\Omega$ with respect to the radius ($a_x$ or $a_y$) for various cutoff wavelengths. The values of $\Delta$ given in Figure 3.7 were used in the calculations, imparting the curvature onto the plots. $\Omega_y < \Omega_x$ in all cases, implying that the taper adiabaticity is determined by $a_y$.

By extension of the preceding argument, the taper angle of fibre TAPH should lie somewhere between the curves of $\Omega_x$ and $\Omega_y$ for $\lambda = 585$nm. The manufacturer's data in Table 3.2 gives a lower and upper limit on the core diameters $2a$. Since the taper length is 1m, we can determine the values of the taper length scales $a_x^T$ and $a_y^T$ from the taper angles and compare it to the quoted taper radius $a^T$:

| | $a$ [$\mu$m] | $\Omega_x$ [mrad] | $\Omega_y$ [mrad] | $a_x^T$ [$\mu$m] | $a_y^T$ [$\mu$m] | $a^T$ [$\mu$m] |
|---|---|---|---|---|---|---|
| Low | 1.75 | 0.028 | 0.021 | 29.75 | 22.75 | 8 |
| Mid | 2.5 | 0.0195 | 0.015 | 22 | 17.5 | 9 |
| High | 3.25 | 0.015 | 0.012 | 18.25 | 15.25 | 10 |

The figures for both $a_x^T$ and $a_y^T$ are larger than the given values for $a^T$. In other words, fibre TAPH should still be well within the single mode regime. This result indicates that the analysis, as it stands, is not a good description of the behaviour of the taper. However, if the value of the *diameter* is used instead (so that $2a \Rightarrow a$), we have

| | $a$ [$\mu$m] | $\Omega_x$ [mrad] | $\Omega_y$ [mrad] | $a_x^T$ [$\mu$m] | $a_y^T$ [$\mu$m] | $a^T$ [$\mu$m] |
|---|---|---|---|---|---|---|
| Low | 3.5 | 0.014 | 0.011 | 17.5 | 14.5 | 16 |
| Mid | 5 | 0.010 | 0.008 | 15 | 13 | 18 |
| High | 6.5 | 0.008 | 0.006 | 14.5 | 12.5 | 20 |

The result for the lower bound is in agreement with the heuristic argument, although the middle and upper bound values are a little too low. The implication is that the geometrical adiabaticity criterion is a weak delineation parameter — a description in terms of the waveguide mode overlap integrals is probably required for a more accurate scalar model.

## 3.7 Dispersion effects in single mode fibres

Having discussed the effects that optical fibre waveguides have on monochromatic light, we must now turn our attention to the effects of broadband sources.

Broadband light, when propagating in any medium, inevitably suffers from the effect of dispersion.

Dispersion in a waveguide can be considered as the behaviour of the modal propagation constants with respect to frequency. Its effect is to limit the bandwidth that can be accepted by a single mode fibre. The principal sources are

- material — due to the $\lambda$ dependence of refractive index, $n(\lambda)$,
- waveguide — due to the $\lambda$ dependence of the fundamental mode,
- profile — due to the $\lambda$ variation of the refractive index profile.

When considering the dispersive effects of weakly guiding fibres, it is convenient to introduce the parameter

$$\begin{aligned} b(V) &\triangleq \frac{W^2}{V^2} = 1 - \left(\frac{U}{V}\right)^2 \\ &= \frac{(\beta^2 - k^2 n_2^2)}{(k^2 n_1^2 - k^2 n_2)} \end{aligned} \tag{3.78}$$

This can be interpreted as a normalised $\beta$, which varies between 0 (for mode cutoff) and 1 (for zero wavelength). Since $\Delta << 1$, another expression involving $\beta$ and $b$ is

$$\beta = k n_2 (1 + b\Delta) \tag{3.79}$$

$\beta$ is related to the guided wave phase velocity by

$$v = \omega/\beta \tag{3.80}$$

where $\omega = 2\pi\nu$ is the angular frequency in radians. More importantly, this defines the group velocity

$$v_g \triangleq \frac{d\omega}{d\beta} = c\frac{dk}{d\beta} \tag{3.81}$$

At optical frequencies, the group velocity is the only observable quantity. Related to this is the group delay, which links the pulse transit time through a

waveguide of length $L$ to the propagation constant.

$$\tau_g = L\frac{d\beta}{d\omega} = \frac{L}{c}\frac{d\beta}{dk} = \frac{LV}{ck}\frac{d\beta}{dV} \tag{3.82}$$

The total group delay can be expressed as a sum of the various dispersion terms. However, the full derivation is long and tedious [44] and the final expressions are not as easy to interpret as the more elementary argument presented here.

Let

$$\tau_g = \tau_m + \tau_w + \tau_p \tag{3.83}$$

be written as the sum of three terms, where the subscripts $m, w$ and $p$ refer to the material, waveguide and profile components respectively. Since $\tau_p$ arises from the variation of $n$ with respect to wavelength, we have

$$\tau_p \sim \frac{d\Delta}{d\lambda} \tag{3.84}$$

Equation (3.11) states $\Delta << 1$, implying that its derivative is small enough that it may be neglected to first order. Hence there are only two dispersion components that need to be considered.

Using Equation (3.79) and (3.82), the delay characteristic of the material becomes

$$\tau_m = \left(\frac{L}{c}\right)\frac{dn_2}{dk} \tag{3.85}$$

which shows that it is independent of the particular mode. Similarly, the waveguide delay is

$$\begin{aligned} \tau_w &= \left(\frac{LV}{ck}\right)\frac{d(b\Delta n_2 k)}{dV} \\ &= \left(\frac{LV}{ck}\right)\sqrt{\Delta}\,\frac{d(bn_2 k\sqrt{\Delta})}{dV} \\ &= \left(\frac{LV}{ck}\right)\left[\frac{\sqrt{\Delta}}{a\sqrt{2}}\right]\frac{d(Vb)}{dV} \end{aligned} \tag{3.86}$$

where the refractive index derivatives are again ignored. Gambling deduced an

expression for the derivative in the waveguide delay $\tau_w$

$$\frac{d(Vb)}{dV} = b\left[1 - \frac{2J_0^2}{J_{-1}(\kappa a)J_1(\kappa a)}\right] \tag{3.87}$$

The explicit $b$ dependence indicates that $\tau_w$ is different for each mode. For pure silica waveguides $\tau_w$ is negligible compared to $\tau_m$. Therefore, our treatment of dispersion effects within silica fibres in the following Chapters is justified in considering only the material delay component.

## 3.8 Summary

A review has been made of the theory of weakly guiding dielectric waveguides, with emphasis on the single mode regime. The most relevant analytical description derives from scalar theory, which lends itself to simple design equations. It must be emphasised that polarisation birefringence and control of the SOP is extremely important in a real fibre based imaging system. An approximate theoretical description of adiabatically taper SM fibres was given and found to be in agreement with a real taper. The results of Oxford Electronic's attempt to fabricate a tapered SM fibre with a high intrinsic birefringence proved negative in that more complex modes than the fundamental were seen to propagate at the tapered end. This motivated a study of tapered elliptical fibres through the scalar field description. The theory was shown to support the heuristic conclusions drawn from the experimental results. However, when applied to the known data for fibre TAPH, the adiabaticity criterion used was found to be weak. Despite this, the basis of the arguments still imply that it should be possible to design and produce intrinsically high birefringence tapered waveguides that are spatially monomodal. These waveguides would be superior to the standard tapered SM fibres for interferometry, in that they would constrain the SOP, removing another degree of freedom from the system.

# Chapter 4

# Single Mode Fibre Interferometry

Remember folks. Traffic lights timed for 35 mph are also properly calibrated for 70 mph.

## 4.1 Introduction

The issues to be tackled when attempting to define an interferometric imaging system based on SM fibres can be summarised as **calibration**

... of the visibility phase. It has already been mentioned that the optical phase cannot be extracted directly from a single baseline interferometer in the presence of atmospheric turbulence. However, with three or more baselines, provided that the error in measuring the baselines themselves is negligible, it is possible to find a good observable through phase closure.

... of the visibility amplitude. A perfect fringe contrast from each baseline is the ultimate objective, so the problems need to be addressed from the outset. In order to maximise the fringe signal–to–noise ratio (SNR), we have to calibrate

- the input coupling mechanism to the SM fibres. Interferometric sensitivity is wholly dependent on the throughput of the fibres, so max-

imising the coupling efficiency in turn maximises the final SNR.

- the polarisation effects of the interferometer arms. Differential retardation (the effect of linearly retarding one polarisation component with respect to the other) between two arms of an interferometer can completely destroy the visibility information.
- the optical path differences between the interferometer arms. Mismatches in the optical path lengths lead to chromatic dispersion effects that are seriously detrimental to the fringe visibility.
- the recombination plane fringe spacing. This arises because the atmospheric coherence timescale $\tau_0$ limits the permitted detector integration time. A fast readout is required, so a number of broad fringes across as few pixels as is necessary would be the desired situation.

We will discuss each item in turn, as a precursor to presenting the experimental results from a multi–fibre interferometer (Chapter 5) and the work leading up to the construction of DAFI (Chapter 6).

## 4.2 Synthetic Aperture Calibration

Visibility phase calibration is a highly non–trivial aspect of the operation of a dilute aperture interferometer, and is an important factor in determining the quality of the final reconstructed image.

The visibility $V'_{ij}$ observed at a time $t$ on the $ij$th baseline is related to the true visibility $V_{ij}$ generally by

$$V'_{ij}(t) = g_i(t)g_j^*(t)G_{ij}(t)V_{ij}(t) + \epsilon_{ij}(t) \tag{4.1}$$

The multiplicative factors $g_i(t)$ and $g_j(t)$ represent the effects of the complex gains of the subapertures $i$ and $j$. $G_{ij}(t)$ is the non–factorable part of the complex gain on the $ij$th baseline. $\epsilon_{ij}(t)$ is a pure, zero–mean noise term representing "thermal" noise. The effect of $G_{ij}$ is assumed to be negligible.

Jennison [45] made use of the fact that an appropriate sum of visibility

phases around a closed loop of baselines is free of subaperture related errors. Taking the phase part of Equation (4.1)

$$\varphi_{ij}(t) = \phi_{ij}(t) + \theta_i(t) - \theta_j(t) + \text{noise term} \tag{4.2}$$

where $\theta_i(t) = \arg g_i(t)$. Summing over a closed loop of say, three baselines gives the observed closure phase

$$\begin{aligned} C'_{ijk} &= \varphi_{ij}(t) + \varphi_{jk}(t) + \varphi_{ki}(t) \\ &= \phi_{ij}(t) + \phi_{jk}(t) + \phi_{ki}(t) + \text{noise} \\ &= C_{ijk}(t) + \text{noise} \end{aligned} \tag{4.3}$$

thus the closure phase is a good observable, being insensitive to all phase errors. The noise term can be neglected as we are operating well within the photon limited regime. For an array composed of $N$ elements, there are $\frac{1}{2}N(N-1) - (N-1)$ independent closure phases or constraints on the true sky intensity distribution.

A *closure amplitude* $\Gamma_{ijkl}$ can be similarly defined for any 4 element loop,

$$\Gamma_{ijkl}(t) = \frac{V'_{ij}(t)V'_{kl}(t)}{V'_{ik}(t)V'_{jl}(t)} \tag{4.4}$$

Since the complex gains cancel out of these ratios, the observed and true closure amplitudes should be identical, apart from noise. There are $\frac{1}{2}N(N-1) - N$ such closure amplitudes.

### 4.2.1 Errors on the Closure Phase

When attempting to measure the closure phase, it is of course desirable to maximise the SNR. The exposure time, the aperture size and the number of apertures can be varied separately to do this.

### Exposure time $t_{\text{exp}}$

At high light levels, the error on any closure phase will be a monotonically increasing function of the exposure time. However, work by Readhead [46] and Buscher [47] conclude the slightly surprising result that the closure phase error can remain below 1 radian up to exposure times of $3.4\tau_0$, if the only source of noise is atmospheric.

At low light levels, the Poisson statistics that govern the detection of individual photons become important. As a result, a balance has to be struck between the temporal smearing of the data from a finite exposure, and the increased photon count provided by a longer exposure. The optimum time for $t_{\text{exp}}$ has been found by Buscher to be $2\tau_0$, which results in a $\sim 90\%$ fringe visibility (neglecting variations in the seeing cell sizes).

### Number of subapertures $n_s$

The number of subapertures $n_s$ obviously affects the efficiency with which the data can be used. This efficiency can be quantified by looking at the number of closure data as a fraction of the number of data available if full calibration were possible. For the phase data, the fraction is [24]

$$\frac{\frac{1}{2}(n_s-1)(n_s-2)}{\frac{1}{2}n_s(n_s-1)} = \frac{(n_s-2)}{n_s} \tag{4.5}$$

For the amplitude data, the fraction is

$$\frac{\frac{1}{2}n_s(n_s-3)}{\frac{1}{2}n_s(n_s-1)} = \frac{(n_s-3)}{n_s-1} \tag{4.6}$$

The relationships are plotted in Figure 4.1.

This reasoning seems to point towards the use of as many apertures as possible to extract the optimum amount of information. However, it does not account for the "total" $\text{SNR}_{\text{tot}}$ as measured by an interferometer. Let this quantity be defined as

$$\text{SNR}_{\text{tot}} \triangleq \text{SNR}_{\text{cp}}\sqrt{N_{\text{cp}}} \tag{4.7}$$

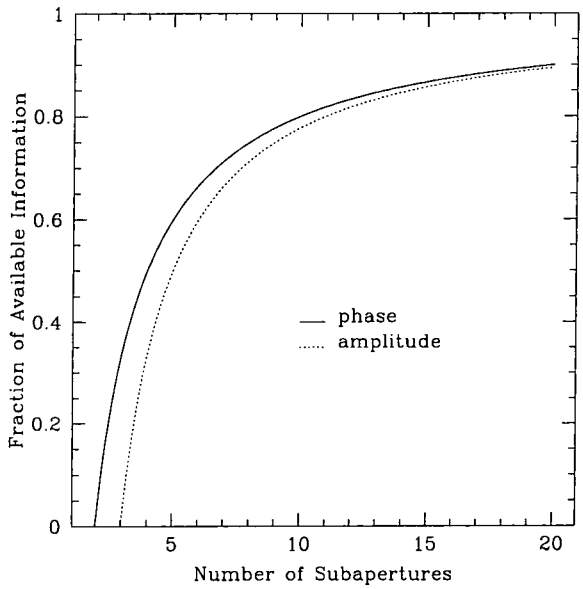



**Figure 4.1.** Visibility data obtained through adaptive calibration as a fraction of that available from a fully calibrated array of subapertures.

with $\mathrm{SNR_{cp}}$ representing the mean SNR of the closure phases and $N_{\mathrm{cp}}$ is the number of (statistically) independent closure phases measured.

At high light levels, the SNR of a measured closure phase is proportional to the SNR of the visibilities on the baselines forming the closed loop [48]. The number of independent closure phases obtained is $N_{\mathrm{cp}} = (n_s - 1)(n_s - 2)/2$. The quadratic dependence of $N_{\mathrm{cp}}$ upon $n_s$ increases faster than any decrease of $\mathrm{SNR_{cp}}$. One way the $\mathrm{SNR_{cp}}$ decreases can arise is from a reduction in fringe visibility as more $n_s$ are added. So the optimum number of subapertures when the photon flux is high is still "as many as possible", this statement being subject to the constraint that $\mathrm{SNR_{cp}} >> 1$.

At low light levels, the $\mathrm{SNR_{cp}}$ has been shown to vary as the cube of the visibility amplitude SNR [48]. Also, the value of $N_{\mathrm{cp}}$ is different to the previous situation. A useful $\mathrm{SNR_{tot}}$ can be found only if the average of a large number of noisy closure phase estimates is taken, which ensures that the full set of closure

phases becomes statistically independent. Thus,

$$\lim_{\text{photons}\to 0} N_{\text{cp}} \to \frac{n_s(n_s-1)(n_s-2)}{6} \tag{4.8}$$

If photon noise is the dominant noise source then $\text{SNR}_{\text{cp}} \propto n_s^{-3/2}$. Again we find that $\text{SNR}_{\text{tot}}$ increases monotonically with $n_s$, reaching an asymptote as $n_s \to \infty$. This result also seems to advocate the use of as many subapertures as possible. However, the issue of whether a large amount of low SNR data is actually useable needs to be addressed. This is related to the method used for inverting the optical closure phase and visibility amplitude data and will depend on its stability in the presence of noisy data. Another obvious constraining factor comes from the fact that the total observing time required to reach a final SNR> 1 cannot exceed one night.

### Subaperture size $D$

AMPLITUDE: We consider the case where no spatial filtering is employed in the subapertures of interferometer. It has been shown [47] that the loss in visibility from moderately sized interferometer subapertures ($D > 3r_0$) is considerable. Most of the instantaneous phase fluctuations across the subapertures, though, is due to a mean tilt across the wavefront which shifts the centre of the image. By correcting this tilt, the SNR of the fringe visibility is improved.

In the atmospheric noise dominated regime (high light level), the results of simulations produced by Buscher supported the statement that tilt correction improves the fringe SNR, by an order of magnitude. As $D/r_0 \to \infty$, the SNR was found to be fall below unity, an interesting result that has not yet been explained.

The optimum subaperture diameter for photon noise dominated signals with tilt correction was shown to be, unsurprisingly, infinitely large. In practice, subapertures greater than $3r_0$ in diameter fare little worse than an infinite diameter. Tilt corrected subapertures are distinct, though, in that baseline elements of $3r_0$ diameter show a significant advantage in SNR terms than larger

subapertures.

CLOSURE PHASE: Further simulations have shown that the atmospheric triple product phase error increases almost logarithmically with subaperture diameter. With the requirement that the closure phase error be less than 1 radian, the optimum apertures should be less than $2r_0$ for uncorrected subapertures, and less than $7r_0$ for tilt corrected subapertures. With low light levels, the simulated results predicted the optimum aperture diameters to be $1.2r_0$ without correction, and about $2.8r_0$ with.

When spatial filtering is included, the parameter important to the design is the energy coupling into the spatial filter. Shaklan [9] simulated the coupling of light into SM optical fibres (using a normalised frequency $V = 2.2$) as a function of $D/r_0$. The maximum coupling for the case with tilt removal occured when $D = 4r_0$, where the coupling was 18% efficient. Without tilt compensation, the expected result of "as large as possible" for the subaperture diameter was demonstrated. The graph of coupling efficiency *vs* subaperture size increased asymptotically to about 6% in this case.

So, bearing all these points in mind, five non–redundant subapertures were chosen for DAFI, yielding 60% of the phase and 50% of the amplitude data (from Figure 4.1). This was a compromise between simplicity of design and a maximisation of information. It was decided that the effective subaperture sizes should be around $4r_0$ to maximise the input energy coupling, which is the most significant parameter to arise from the discussion in this Section.

Having briefly discussed the issue of visibility phase calibration, we now consider the obstacles (and solutions) to producing fringe signals with a high contrast with a fibre interferometer.

## 4.3 Practical Fibres

Optical fibres are made from plastics, crystals or glasses, the latter being the most mature technology. The glass of most interest is that formed by fusing molecules of silica ($SiO_2$). To obtain the different refractive indices dopants are added, to either core or cladding. Examples of dopants are titanium, thallium,

germanium, boron and fluorine.

Glass intrinsic absorption is very strong in the short wavelength ultraviolet region, owing to strong electronic and molecular transition bands. This effect diminishes as wavelengths increase, to the point where it is considered negligible in the visible. In the infrared, stretching vibrations of the chemical bonds such as between silicon and oxygen limit the transition bandwidth longward of about 1.6$\mu$m. Impurities are the major source of loss within practical fibres, especially the hydroxyl (OH) and transition metal ions which absorb strongly in the visible.

There are basically three types of silica fibre available for use in the visible bandpass: high OH (wet), low OH (dry) and hydrogen doped, low OH. The wet fibres contain OH impurities leading to poor red and near–IR transmission properties, although they are the best option for the UV regime where losses approach the Rayleigh scattering limit. Dry fibres are purified of OH, so have excellent transmission between 0.7 and 2$\mu$m. Unfortunately there exists material defect absorption blueward of 0.63$\mu$m due to broken Si–O bonds, created by the drawing process. In wet fibres, OH aids the annealing of the broken bonds so that this defect is suppressed. It has been found that H doping the dry fibres also helps to anneal the glass bonds [49] so extending the blue performance to essentially that of high OH fibres.

For mechanical strength and handling, fused silica optical fibres are usually coated with a buffer layer. The most common buffer coatings are acrylate, which has a translucent appearance, and polyimide, which is strikingly yellow. The latter material is more easily controllable so the diameter tolerances are stricter than the former. Also a greater thickness of acrylate is required compared to polyimide to achieve the "same" protective quality. However, polyimide is very difficult to remove, potentially increasing fibre wastage prior to the cleaving process.

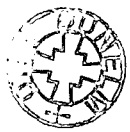
### 4.3.1 Fibre cleaving

An important precursor to efficiently coupling light into a fibre waveguide is the endface preparation. The principle is straightforward: scratch (scribe) the glass surface to produce a concentration of stress, then subject the fibre to tension (and possibly bending) so that the stress is maximum at the scratch point. This causes the material to fail.

A good cleave is one that produces a completely flat end with an angle between the cleaved plane and a plane normal to the fibre axis of less than one degree. The common defects of cleaved endfaces are defined in the following way:

- Lip : Sharp protrusion at the edge of the fibre.
- Rolloff : Complement of Lip.
- Chip : Localised fracture on the end of a cleaved fibre.
- Hackle : Multiple surface irregularities across fibre surface.
- Mist : Similar but less severe form of Hackle.
- Spiral (Step) : Endface conditions which feature abrupt changes in the endface topology.
- Indent : Deliberately introduced flaw from which the controlled crack propagates.

The cleave quality is a function of indent depth, angle and the amount of tension and torsion applied. The simplest form of cleaver is a blade of ceramic or tungsten carbide that is used to manually score the fibre before tensioning by hand until it cleaves. This is understandably a poor technique in terms of cleave repeatability. The next level of tool is one that is often hand held and scores the fibre by pushing a blade of tungsten carbide, sapphire or diamond against the fibre, which is supported by an anvil. Tension is then applied, either along or away from the fibre axis to cause a crack to propagate. These tools



are highly dependent on the skill of the user. It is suggested that shock waves reflected from the anvil can cause problems with cleave quality.

The cleave tool used during this work is a York Technology FK11. This device holds the fibre between two low torsion clamps while a user defined amount of tension is applied. An ultrasonic vibrating diamond blade mounted on damped bearings is then brought at a controlled speed towards the bare fibre until cleaving occurs [50]. The blade oscillation frequency[1] is approximately 60kHz with an amplitude of $1\mu m$.

Although the fracture dynamics of brittle materials is a poorly understood, non–linear pattern forming phenomenon, it is possible to make some statements that can help in producing a good cleave. Assuming that the fibre is clean and free from static cling, the most important factor influencing the fracture propagation is its interaction with its own sound energy. In general, the amount of sound energy emitted by the crack is much smaller than the surface energy required to allow the crack to propagate. However, it is well known from dynamical theory that a small perturbation may produce large changes on the dynamics of an unstable system. Thus, we cannot neglect the effects of sound on the crack.

With the York cleaver, we can assume that the tension applied to the fibre is constant. So the only variable is the amount of buffer removed; since the buffer can damp sound waves within the fibre, it is desirable to minimise the amount of damping[2] by removing 10mm of the buffer around the region to be cleaved. By careful control of the applied tension, it is possible to produce perfectly smooth cleaved ends with an undamaged blade. Figure 4.2 show some 633nm laser illuminated fibre endfaces. The problem with illuminating the fibres in this manner is that it is not trivial to align the endface exactly along the illuminating beam's optical axis. Aberrations such as coma and astigmatism are easily introduced as a consequence.

Figure 4.3 show two cleaved $GeO_2$ fibres as measured by a scanning electron

[1]York Technology private communication.

[2]This assumption has been made since there is currently no data on crack dynamics in fused silica.

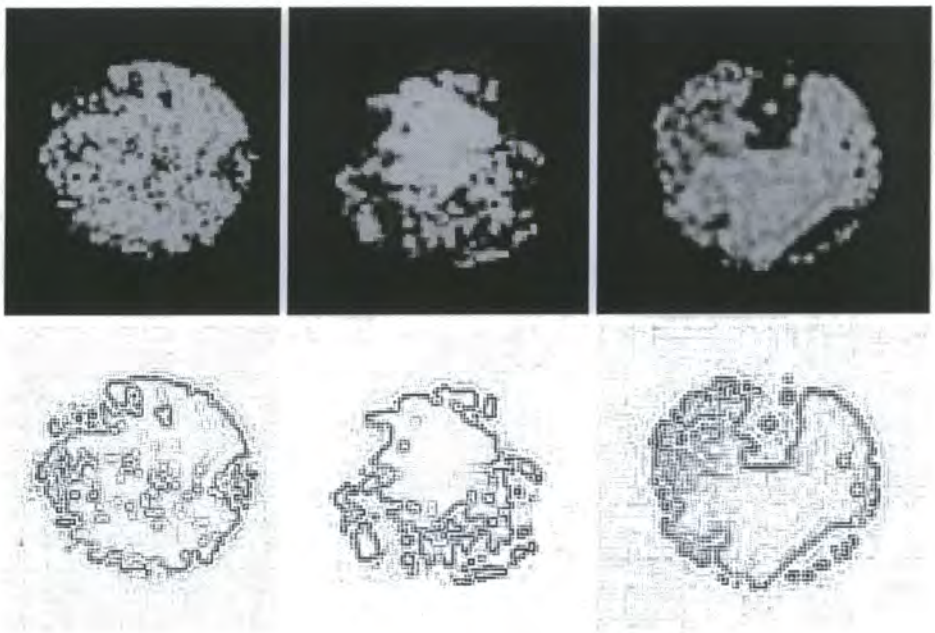


**Figure 4.2. Top**: These images show the endfaces of three fibres after cleaving. The image on the left shows a good cleave face, the central image illustrates the effect of mist and hackle and the image on the right suffers from rolloff. **Bottom**: The same images enhanced by a 50% differential edge detection algorithm.

microscope (SEM). The SEM source beam wavelength was set at 49.5pm for all the SEM photomicrographs shown in this work. The samples were coated with a few tens of nanometres of gold to provide a conductive surface.

It should be noted that a good overall endface quality does not necessarily guarantee an acceptable beam profile. Only the local (to the core region) endface quality is important. This is evident from Figure 4.4.

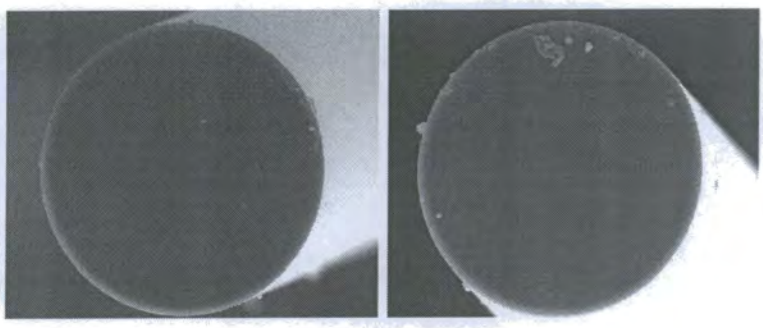


**Figure 4.3.** Fractographs of well cleaved endfaces, imaged with a scanning electron microscope. The cleave tool indent can just be seen at the perimeter of the fibres. (Magnification ×700.)

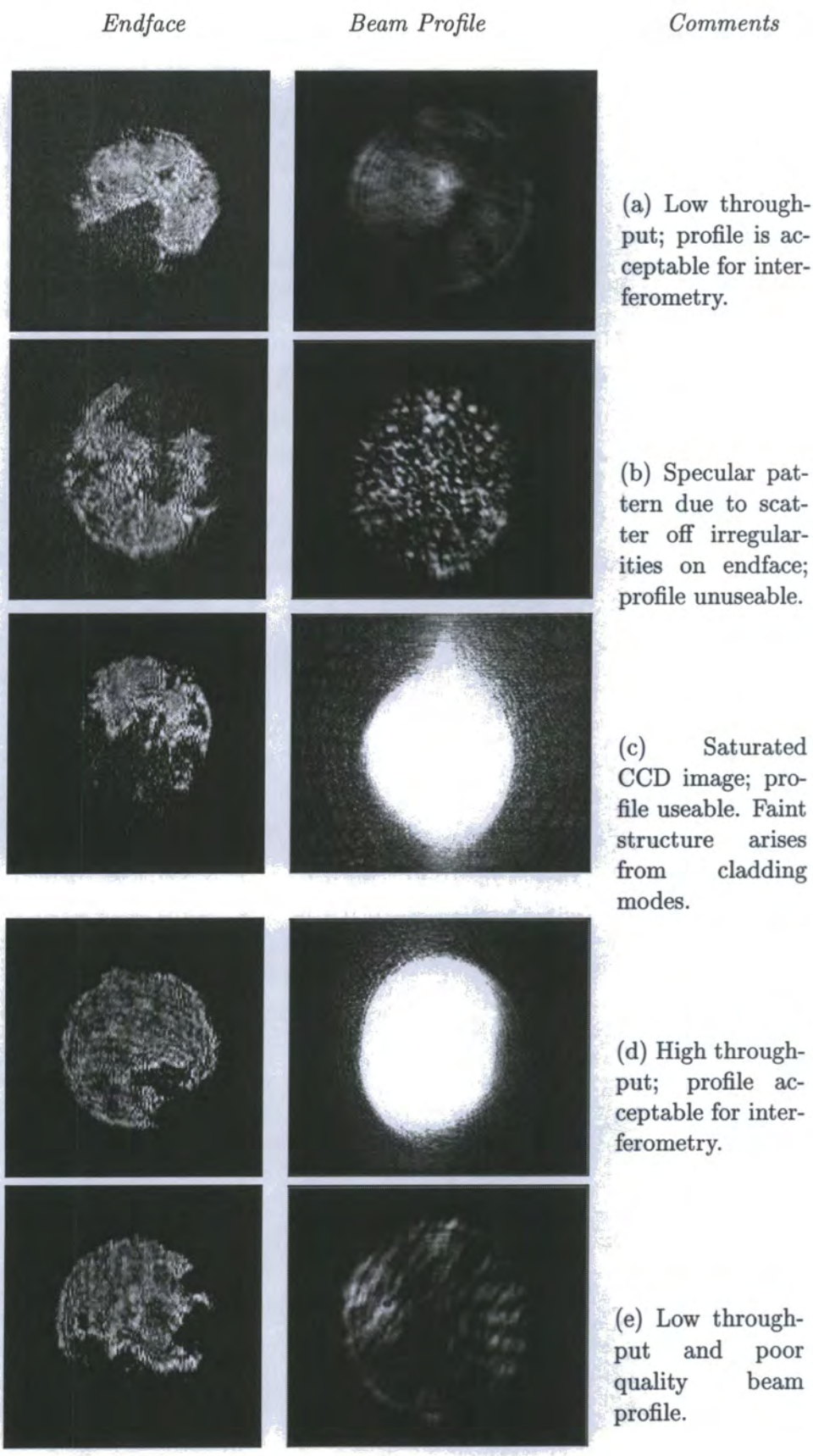



**Figure 4.4.** Fibre endfaces and beam profiles. The greyscale is the same for all the beam profile images.

In conclusion, it is straightforward to cleave 125$\mu$m silica fibres with consistently flat endfaces. This result very much simplifies the method of producing the fibre arms of the interferometer. We now turn our attention to the problem encountered when attempting to interfere quasi–monochromatic beams.

## 4.4 Polarisation in Interferometers

The wave nature of light manifests itself as transverse oscillations of the electromagnetic field. The nature of the vibration of the electric displacement vector in the plane normal to the direction of wave propagation defines the state of polarisation of a light beam.

In a completely polarised beam[3], the vibration may be either linear in any azimuth at right angles to the propagation direction, or elliptical, with the major axis at any azimuth. The ratio of axes of the ellipse can have any value and the sense of the ellipse may be right or left handed. The two limiting cases of elliptic vibrations are linear and circular vibrations. Correspondingly, the light beam would be said to be linearly, elliptically or circularly polarised.

A general state of polarisation can thus be described by two quantities: (a) the orientation of the major axis of the ellipse, which may be specified by the angle $\lambda$ which it makes with a given direction in the wavefront and (b) the ratio of the axes of the ellipse ($b/a, b < a$). The sense of the ellipse can be specified by making the axial ratio positive for left–rotating (laevorotatory) ellipses and negative for the opposite (dextrorotatory) case. The sense of rotation is with respect to an observer looking towards the source of light, where clockwise rotation is righthanded.

The orientation of the major axis of the ellipse (the *azimuth*) is given by the angle $\lambda$ which it makes with the horizontal ($\vec{x}$) coordinate axis measured in the laevorotatory sense, as shown in Figure 4.5. The *ellipticity* is defined by another angle $\omega$, given by $\tan\omega = b/a$. These two angles uniquely specify the SOP of a light beam.

[3]Full descriptions of unpolarised and polarised beams of light can be found in [51, 1].

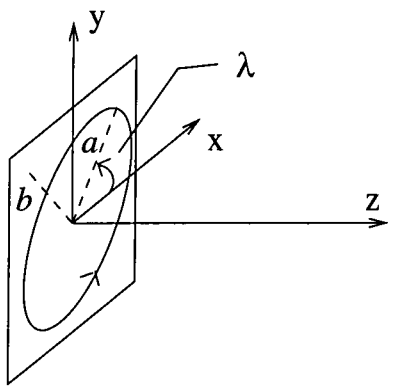



**Figure 4.5.** Definitions for polarised light.

Of the various formalisms for describing the SOP of a light beam, we prefer to use the Poincaré sphere (and its group theoretic equivalent [52], the Jones matrices) to analyse the propagation of SOPs through optical systems.

In general, a beam of polarised light can be represented on the Poincaré sphere by a point $P$. A device which produces light of polarisation state $P$ will be termed a "polariser $P$"; the complementary device which transmits light of polarisation state $P$ completely will be called an "analyser $P$".

### 4.4.1 Interference of two partially coherent beams

A light beam in any state of polarisation can always be decomposed into two completely polarised component beams in any given state. However, each component will in general be partially coherent. This means that there exists only a partial correlation between the fluctuations in the absolute phases and intensities of the beams. For example, even unpolarised light can be resolved into two non–orthogonal polarised beams which will then be partially coherent with one another.

With this in mind, consider the interference of two completely polarised light beams with SOPs $P_1$ and $P_2$, of intensities $I_1$ and $I_2$, which are partially coherent. We have already seen that the degree of correlation can be represented by $\gamma$. It is now convenient to introduce a parameter $\delta$ for the effective phase advance of the first beam $P_1$ over the second $P_2$. These quantities are defined

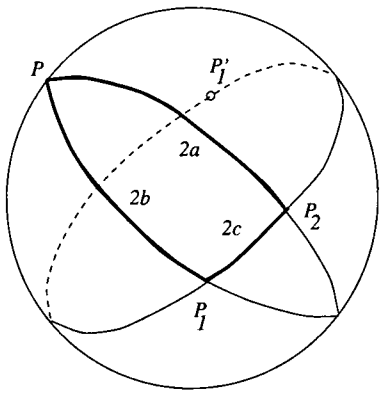



**Figure 4.6.** Coherent addition of polarised beams as represented on the Poincaré sphere.

through

$$< |\sqrt{i_1 i_2}| \exp(\mathrm{i}\delta_t) >= |\gamma| \sqrt{I_1 I_2} \exp(\mathrm{i}\delta) \tag{4.9}$$

Here $< a >$ stands for the time averaged value of $a$ and $i_{1,2}$ are the instantaneous intensities of the beams with states $P_{1,2}$. $\delta_t$ is the instantaneous phase advance of the first vibration over the second.

In order to find the resultant intensity of the two combined beams, we need to average the instantaneous intensities. With reference to Figure 4.6, the states $P_1$ and $P_2$ combine to give state $P$. Denote the arcs $\overset{\frown}{PP_2}$, $\overset{\frown}{PP_1}$ and $\overset{\frown}{P_1P_2}$ by $2a$, $2b$ and $2c$ respectively. Let $P_1'$ be the state opposite to $P_1$. Resolve beam 2 into states $P_1$ and $P_1'$. The intensities of the components of beam 2 will be

$$P_1 \text{ component} = i_2 \cos^2 \frac{1}{2} \overset{\frown}{P_1P_2} = i_2 \cos^2 c \tag{4.10}$$

$$P_1' \text{ component} = i_2 \cos^2 \frac{1}{2} \overset{\frown}{P_1'P_2} = i_2 \sin^2 c \tag{4.11}$$

The resultant instantaneous intensity at $P_1$ is then

$$i = i_1 + i_2 \cos^2 c + 2\sqrt{i_1 i_2} \cos c \cos \delta \tag{4.12}$$

At state $P_1'$ the component intensity is

$$i_{P_1'} = i_2 \sin^2 c \tag{4.13}$$

Since the component beams are orthogonal, the sum of the instantaneous intensities gives the intensity of the momentary resultant independent of the phase between them. Averaging the momentary resultant over time gives

$$I = I_1 + I_2 + 2|\gamma|\sqrt{I_1 I_2}\cos c \cos \delta \tag{4.14}$$

From this equation, it is seen that the visibility of fringes obtained by the interference of two completely polarised beams is affected by two factors. These are $|\gamma|$ and $\cos c$, where the latter specifies the difference in the SOPs of the two beams.

These two factors must be carefully distinguished; for example, two orthogonally polarised beams can never give rise to interference in intensity yet may be completely coherent (yielding an elliptic vibration). Similarly two beams may be in the same SOP but be completely incoherent. In general, the beams can be tested for partial coherence after transmission through an analyser which resolves them into the same state. It is convenient to adjust the analyser so that the intensity of the transmitted beams are equal. If the transfer function of the optical system is known, then $|\gamma|$ is readily found from the fringe visibility, remembering that $\cos c = 1$.

If the beams are not resolved through an analyser, $|\gamma|\cos c$ is determined. Since $|\gamma|$ can be found by the previous experiment, the non–orthogonality factor may be separated out[4].

### 4.4.2 Visibility losses

If the SOPs in an interferometer are not well controlled, complete fading of the fringe signal can result. The various effects due to polarisation mismatches on the fringe visibility are described below.

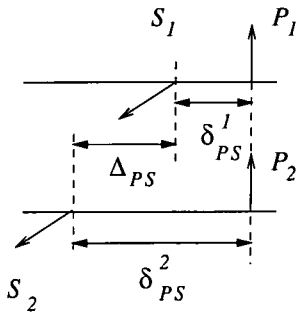



**Figure 4.7.** The effect of linear retardation of polarisation.

### Linear retardation of polarisation

Figure 4.7 shows how linear retardation affects the visibility. Upon passing through an optical system, the two orthogonal polarisations of an incident light beam can be phase shifted with respect to each other by $\delta$. If two beams from a single baseline are affected by the same linear retardation, the fringes produced by the s and p polarisations will be superimposed. If a differential retardation $\Delta = \delta^1_{PS} - \delta^2_{PS}$ exists, the two fringe signals will be shifted with respect to each other. For unpolarised light, these two patterns add incoherently on the detector so that the final visibility function is affected by

$$T_{lr} = \left|\cos\left(\frac{\Delta}{2}\right)\right| \tag{4.15}$$

### Rotation of polarisation reference frame

The incident polarisation frame of reference can experience rotation when propagating through optical systems. If the rotation in each arm of an interferometer is not exactly the same, the situation illustrated in Figure 4.8 occurs after recombination. Each polarisation can be projected on the two reference axes. If the incident light is unpolarised (or linearly polarised) the only components generating fringes are $P_1$ with $P_2 \cos\theta$ and $S_1$ with $S_2 \cos\theta$ for a frame rotation angle $\theta$. The other two components add constant intensity vectors $(P_2)^2 \sin^2\theta$

[4]In essence, we have interpreted $|\gamma|^2$ to be that fraction of the intensity of beam 1 which is completely coherent with beam 2, where beam 2 leads beam 1 by a phase $\delta$.

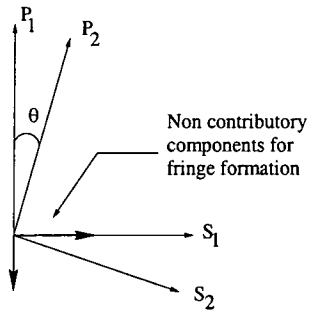



**Figure 4.8.** The effect of rotating the polarisation reference frame.

and $(S_2)^2 \sin^2\theta$. The visibility is then degraded by

$$T_{rot} = |\cos\theta| \tag{4.16}$$

### Polarisation attenuation

The relative attenuation of one polarisation with respect to the other leads to visibility loss in a similar manner to unequal beam intensities. For an unpolarised source, the visibility factor is [1]

$$T_{pp} = \frac{2(1+p)}{(3+p^2)} \tag{4.17}$$

where $p$ is the degree of polarisation affecting one arm of a pair. For a 10% polarisation (i.e. $1-p^2 = 0.1$), only a 0.07% visibility loss occurs.

## 4.5 Polarisation control

In Chapter 3 it was shown that the $LP_{01}$ mode of a SM fibre can be decomposed into two orthogonal polarisation states. One of the advantages that DAFI will have over other fibre–based stellar interferometers will be its tapered SM waveguides, which should increase its overall sensitivity as compared to standard waveguides. The results of an attempt that was made to build a degree of birefringence into the tapered fibres have already been described in Section

3.6. Since this was unsuccessful, SOP control through induced birefringence was studied instead.

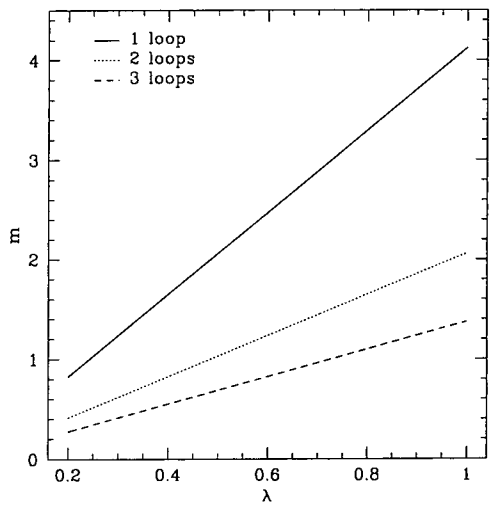



**Figure 4.9.** Equivalent fractional waveplates for 125$\mu$m diameter silica fibres formed into Mickey ears of 1.06inch (= 26.924mm) diameter. $\lambda$ is shown in units of $\mu$m.

The output state of polarisation drastically depends on the random external stresses applied to the fibre, in the form of bends, twists and clamping. SM fibre fractional wave devices [39, 40] can be created by forming loops of fibre. These *Lefèvre loops* or *Mickey ears* induce a phase delay $\Delta\Phi$ which can be related to the number of loops of bend radius $R$.

$$\Delta\Phi = |B_b|NR \tag{4.18}$$

$B_b$ comes from Equation (3.58). $\lambda\Delta\Phi$ is the delay expressed as a fraction of a wave, more conventionally written as $\lambda/m$. For pure silica fibres, we have

$$\frac{\lambda}{m} = 0.135\left(\frac{r}{R}\right)^2 2\pi NR \tag{4.19}$$

So a Mickey ear is analoguous to a classical birefringent plate. Figure 4.9 plots the wave fraction $m$ as a function of wavelength $\lambda$ for the set of commercially

available holders[5] used for this work.

It should be noted that this theoretical analysis does not account for the buffer coating; since its stress–optic constants are generally small, it seems reasonable to neglect the buffer's contribution.

It is clear from Equation (4.19) that the chromaticity ($dm/d\lambda$) of the Mickey ears is inversely proportional to $N$, so as $N \to \infty$, $dm/d\lambda \to 0$. In the optimal situation, it is desirable to reduce the dependence on $\lambda$ away from the central wavelength of the linear retardance — this can be achieved by increasing the number of loops.

Of course, bending birefringence is only associated with the linear retardation of polarisation. To compensate for circular retardation, a means of introducing optical activity is needed. This can be done by simply creating a series of Mickey ears which can then be positioned at various angles to each other (thereby torturing Mickey?). The effect of twisting the intervening fibre sections alters the SOP ellipticity.

If the number of twists in the optical fibre is kept to a minimum prior to reaching the Mickey ear array (which is assumed to be near the fibre output) then the effect of circular birefringence can be ignored. By constraining the perturbations to bends, we find that only an unspecified degree of linear retardance is necessary to control the SOPs. It is possible to create a linear retarder whose delay varies between 0 and $2\pi$ through the geometric phase[6] [54, 55].

Pancharatnam showed that a combination of three waveplates can behave as a single linear retarder of delay $2\delta$ whose fast axis is inclined at an angle $c_1$ to the first plate, where $2\delta$ and $c_1$ are given by

$$\cos\delta = \cos 2\delta_1 \cos\delta_2 - \sin 2\delta_1 \sin\delta_2 \cos 2c \tag{4.20}$$

[5]Produced by Thorlabs.

[6]The geometric phase in polarisation optics is a particular example of the Berry phase [53]. This is a phase factor due to an adiabatic transport of a system around a closed path in the parameter space. In polarisation optics the closed path traces the evolution of SOP on the Poincaré sphere. The value of the geometric polarisation phase factor is given by half the solid angle subtended by the area on the Poincaré sphere contained within the closed path. It should be distinguished from the dynamical phase factor, which arises (say) from the effective wavelength change when light crosses the boundary from one medium to another.

and

$$\cot 2c_1 = (\sin 2\delta_1 \cot \delta_2 + \cos 2\delta_1 \cos 2c) \csc 2c \tag{4.21}$$

The quantity $2\delta_1$ is the common retardation of the first and last waveplates and $2\delta_2$ is that of the central plate. The first and third plates are aligned with their fast axes parallel so $c$ is the angle between the fast axis direction of the central plate and the other two (see Figure 4.10). We shall refer to this combination of three waveplates as a Pancharatnam triplet.

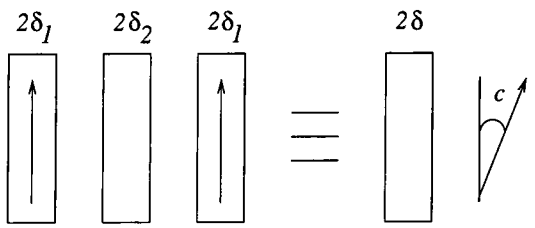



**Figure 4.10.** Equivalence of a variable linear retarder composed of three waveplates. The fast axes of the waveplates are indicated by the arrows.

A continuous variation from 0 to $2\pi$ is available if 2 quarterwave plates are chosen for the first and third plates and a halfwave plate for the centre to form a "QHQ". The principal planes of the QHQ will always be inclined at an angle $c = 45°$.

### 4.5.1 Achromaticity

To make the Pancharatnam triplet achromatic over a finite range of wavelengths, we define two wavelengths $\lambda'$ and $\lambda''$ either side of the waveplate $\lambda$ such that

$$\lambda' \triangleq (1 - f)\lambda$$
$$\lambda'' \triangleq (1 + f)\lambda$$

where $f$ is a system dependent factor. Their respective retardations through the first and last plate then become

$$\delta_1' = (1-f)\delta_1$$
$$\delta_1'' = (1+f)\delta_1$$

Pancharatnam impose the constraint that $2\delta' = 2\delta'' = 2\Delta$, where $2\Delta$ is the required retardation of the achromatic triplet. By setting $\delta_2 = \pi/2$ he found the condition for $\delta_1$,

$$\sin(2\delta_1 f) = \frac{\sin\left(\frac{\pi}{2}f\right)}{\cos\Delta}\sin(2\delta_1) \tag{4.22}$$

which is transcendental, so has to be solved iteratively. An expression for the orientation angle was also found

$$\cos(2c) = -\frac{\tan\left(f\frac{\pi}{2}\right)}{\tan(2\delta_1 f)} \tag{4.23}$$

To illustrate these concepts, consider an optical fibre of 125$\mu$m diameter, at $\lambda = 633$nm. The result for a quarterwave Mickey ear is

$$\frac{N}{R} = 4.77 \times 10^{-5}\ [\mu\mathrm{m}^{-1}] \tag{4.24}$$

A halfwave Mickey ear for the same fibre has

$$\frac{N}{R} = 9.55 \times 10^{-5}\ [\mu\mathrm{m}^{-1}] \tag{4.25}$$

remembering that $N$ has to be an integer number. The larger the value of $N$, the more a Mickey ear tends towards being achromatic. If the bend radius of the interferometer fibre arms can be specified accurately, it should be possible to predict the amount of retardation $2\Delta$ required for the fibre equivalent of the Pancharatnam triplet.

The design expressions for polarisation controllers which employ bending–induced birefringence have been reviewed in this Section. These controllers have

the advantage of being easy to implement, requiring only mechanically stable mounts. Of course, the mounts should also be tolerant of the temperature changes that are experienced at a telescope, so it is preferable to make them out of a low expansion plastic (such as Delrin) rather than a metal.

## 4.6 Optical path difference effects

This Section is devoted to solving the problems that occur when observing sources with a (large) finite bandwidth.

From Equation (1.10), it can be seen that as soon as the group delay ($\tau$) is not zero, the fringe intensity assumes a dependence on wavelength. For finite spectral bandpasses this can become quite problematic if the optical path difference between interferometer arms is not nearly zero. The excess lengths of silica essentially limit the instrumental bandwidth. Active compensation schemes [56, 57] are a solution, whereby the fibres are stretched or heated, but none of these methods exactly cancels the effect of the excess waveguide section, especially if it is a significant quantity. Clearly the most desirable starting point is to equalise the length of the waveguides in each arm.

For a high contrast fringe signal, when observing sources with wide bandwidths (eg. $\Delta\lambda = 300$nm) at short wavelengths, the tolerance on the relative length difference between fibre arms can be severe.

In Subsection 4.6.1, we deduce an expression relating the length difference $\Delta L$ to the fringe modulation as a function of the observed frequency and bandwidth. The results given by expression are in agreement with numerical simulations produced by Shaklan [9] and serve as a guide to the tolerances required to achieve a given fringe modulation depth.

### 4.6.1 Tolerance on length difference

Consider an optical fibre waveguide illuminated by a time dependent signal from an astronomical source. The incident scalar field, written as a Fourier expansion in terms of angular frequency $\omega$, is

$$E_{in}(t) = \int_{-\infty}^{+\infty} E(\omega) \exp(\mathrm{i}\omega t) d\omega \tag{4.26}$$

The output field, after propagation through the waveguide, can be expressed as

$$E_{out}(t) = \int_{-\infty}^{+\infty} E(\omega) \exp(\mathrm{i}\omega t + \Phi(\omega)) d\omega \tag{4.27}$$

The induced phase change $\Phi(\omega)$ can be expressed as the sum of a dispersive material component and an air path component. The material contribution $\Phi_m$ can approximated by a Taylor expansion

$$\Phi_m \approx \beta L + (\omega - \omega_0)\frac{d\beta}{d\omega}L + \frac{(\omega - \omega_0)^2}{2}\frac{d^2\beta}{d\omega^2}L \tag{4.28}$$

with $\omega_0$ the centre frequency and $L$ the length of the fibre. The propagation constant of the fibre waveguide is the conventional $\beta \triangleq kn$. This expression is valid for fused silica SM fibres operating in the visible region, where the effect of waveguide and refractive index profile delays are negligible. The air path contribution $\Phi_a$ can be written as

$$\Phi_a = \omega[x_v + x_a\{1 + N_a(\omega)\}] \tag{4.29}$$

where $x_v$ is the path in vacuum, $x_a$ is the path length in air and $\{1 + N_a(\omega)\}$ is the index of refraction of air.

We now refer to the schematic two fibre, single baseline interferometer shown in Figure 4.11. The coupling is assumed to be perfect for both waveguides and their material dispersion functions are taken to be exactly the same. The light is guided to a recombination plane where a detector produces a signal linearly proportional to the intensity of the recombined beam. In the shorter arm, the

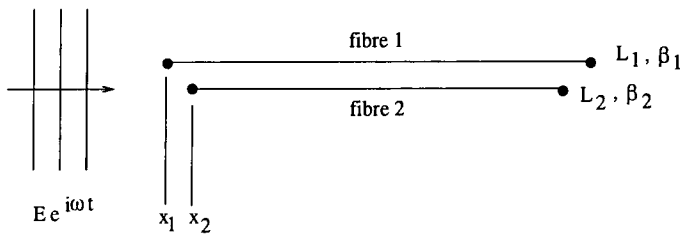



**Figure 4.11.** Idealised two SM fibre interferometer illuminated by a point source.

missing thickness of dispersive material is compensated for by an additional air path. For a zero group delay position, the linear parts of the material and air contributions to the induced phase cancel. We are then left with a second order induced phase shift

$$\Delta\Phi = \frac{(\Delta\omega)^2}{2}\frac{d^2\beta}{d\omega^2}\Delta L + \text{constant} \tag{4.30}$$

where $\Delta\omega$ is the angular frequency bandwidth and $d^2\beta/d\omega^2$ is the chromatic dispersion. It should be noted that we neglect path fluctuations caused by air turbulence and any wavelength dependence of the visibility phase. It is also assumed that field–of–view path length differences are unimportant.

It is generally assumed that the condition $\Delta\Phi \leq \pi$ should be satisfied to produce a signal modulation at the output. That is to say, the recombining fields have a phase lag of less than half a wave. This gives the maximum tolerance for standard silica fibres operating in the astronomical V–band ($\lambda_c = 0.55$ $\mu$m, $\Delta\lambda/\lambda = 0.16$) to be 71.85$\mu$m.

A more stringent and practical tolerance can be derived from the following considerations. It is well known that the character of an interferogram formed from a point source is determined by the complex degree of coherence $\gamma(\tau)$. This is the normalised form of the self coherence (autocorrelation) function $\Gamma(\tau)$; this is the same as the fringe contrast $V$, for a system with unit transmission. For a stationary random process $\gamma(\tau)$ is a Fourier Transform (FT) of the normalised

source power spectral density $G$

$$\gamma(\tau) = \int_0^\infty G(\nu) \exp(-2\pi i \nu \tau) d\nu \tag{4.31}$$

While this expression is true for a perfect instrument, in all practical cases the ideal visibility is always modified by the instrumental transfer function $T_i(\nu)$,

$$\gamma(\tau) = \int_0^\infty T_i(\nu) G(\nu) \exp(-2\pi i \nu \tau) d\nu \tag{4.32}$$

We assume that the shape of the normalised source power spectrum is approximately Gaussian so it may be written as

$$G(\nu) = \frac{2A}{\Delta\nu\sqrt{\pi}} \exp\left( - \left( A \frac{\nu - \nu_0}{\Delta\nu} \right)^2 \right) \tag{4.33}$$

$\nu_0$ is the central frequency, $\Delta\nu$ denotes the frequency bandwidth and $A = 2\sqrt{\ln 2}$. Equations (4.28) and (4.29) indicate that the transmission function can be written (in terms of $\nu$) as

$$\begin{aligned} T_i(\nu) &= \exp(-i\Phi) \\ &= \exp\Bigg( - i\Delta L \Big( \beta + (\nu - \nu_0)\beta_\nu \\ &\qquad + \frac{(\nu - \nu_0)^2}{2} \beta_{\nu\nu} + \nu x_a [1 + N_a(\nu)] \Big) \Bigg) \end{aligned} \tag{4.34}$$

where $\beta_\nu = d\beta/d\nu$ and $\beta_{\nu\nu} = d^2\beta/d\nu^2$. It is assumed that the additional air path in this interferometric system is very much less than a metre; since $x_v$ and $x_a$ typically differ by less than 300$\mu$m/m of path length it is valid to neglect the vacuum geometric path[7]. Adjusting the system air paths to remove the linear terms and considering the zero group delay position (for the highest

[7]It has recently been brought to the author's attention that it is possible to correct the vacuum path errors by means of temperature and strain control within the restriction of a single polarization state [58].

fringe visibility) allows us to write

$$\gamma(\tau) = \frac{2A}{\Delta\nu\sqrt{\pi}} \int_0^\infty \exp\left( -\left\{ \frac{A^2}{\Delta\nu^2} + \frac{\mathrm{i}\Delta L}{2}\beta_{\nu\nu} \right\} (\nu - \nu_0)^2 + 2\pi\mathrm{i}(\nu - \nu_0)\tau \right) d\nu \tag{4.35}$$

We find

$$\gamma(\tau) = \left(\frac{A}{\Delta\nu}\right) \frac{1}{(\xi + \mathrm{i}\eta)^{1/2}} \exp\left(\frac{-(\pi\tau)^2}{(\xi + \mathrm{i}\eta)^{1/2}}\right) \tag{4.36}$$

where $\xi \triangleq (A/\Delta\nu)^2$ and $\eta \triangleq (\Delta L/2)\beta_{\nu\nu}$. The modulus of Equation (4.36) is

$$|\gamma(\tau)| = \left(\frac{A}{\Delta\nu}\right) \frac{1}{(\xi^2 + \eta^2)^{1/4}} \exp\left(-(\pi\tau)^2 \Re\left\{\frac{\xi - \mathrm{i}\eta}{\xi^2 + \eta^2/4}\right\}\right) \tag{4.37}$$

where $\Re$ shows that the real part of the expression should be taken. We are concerned with the variation of the visibility function at the zero group delay position ($\tau = 0$), so

$$V = \left(\frac{A}{\Delta\nu}\right) \left[ \left(\frac{A}{\Delta\nu}\right)^4 + \left(\frac{\Delta L \beta_{\nu\nu}}{2}\right)^2 \right]^{-1/4} \tag{4.38}$$

It is now possible to impose the condition of minimum visibility $V_{\min}$. Therefore $V \geq V_{\min}$ which gives

$$\Delta L \leq \frac{2}{\beta_{\nu\nu}} \left[ \left(\frac{A}{\Delta\nu V_{\min}}\right)^4 - \left(\frac{A}{\Delta\nu}\right)^4 \right]^{1/2} \tag{4.39}$$

The calculated values for the constraints on the optical path difference of single mode silica fibres are shown in Figures 4.12 and 4.13.

Let $V_{\min} \geq 0.8$, which is equivalent to the classical Strehl criterion of diffraction limited imaging. In this case the relative fibre length tolerance for a source of 100nm bandwidth, centered on 633nm is $\Delta L \leq 410\mu$m. For a source of $\Delta\lambda = 300$nm centered on 633nm, the tolerance is $\Delta L \leq 41\mu$m.

For a more relaxed constraint $V_{\min} \geq 0.5$, the maximum allowable values of $\Delta L$ are $1325\mu$m and $132\mu$m respectively, with the same sources as before.

The two plots in Figure 4.14 illustrate the degradation in fringe visibility as a function of $\Delta L$ for two specific wavelengths, $\lambda = 633$nm and $\lambda = 1\mu$m.

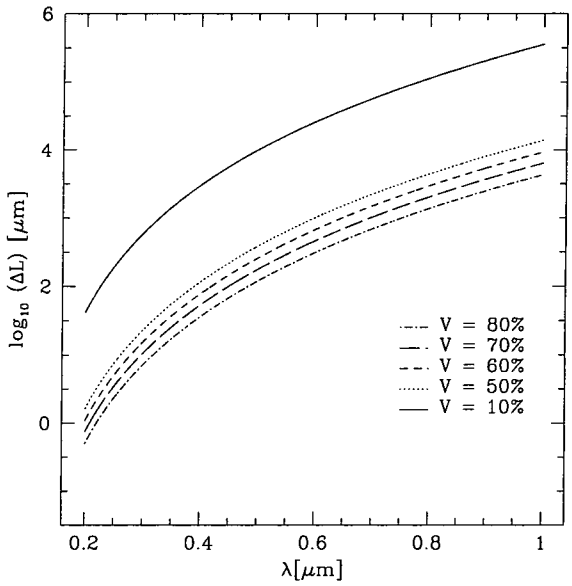



**Figure 4.12.** Tolerance on fibre lengths for a source bandwidth of $\Delta\lambda = 0.1\mu$m as a function of wavelength, for various visibility $V$ criteria.

### 4.6.2 Fibre Induced Phase

The relationship between fibre dispersion and interferometric phase measurements is quantified here. By doing so, we show that the length difference between any two fibres may be found from the fringe visibility curve.

We have seen that the chromatic dispersion can be expressed as

$$D \triangleq \frac{d^2\beta}{d\omega^2} = \frac{1}{c^2}\frac{d^2\beta}{dk^2} \tag{4.40}$$

Recall Equation 4.30 for the phase difference between fields in the recombination plane. $\Delta\Phi(k)$ is not a useful observable since it is non–trivial to calibrate for the constant term. Considering the first differential with respect to wavenumber $d\Delta\Phi(k)/dk$ is also ambiguous since it too includes a constant relating to a linear translation of the interferogram. The quantity which is of use is the second order differential with respect to $k$. This can be regarded as the phase

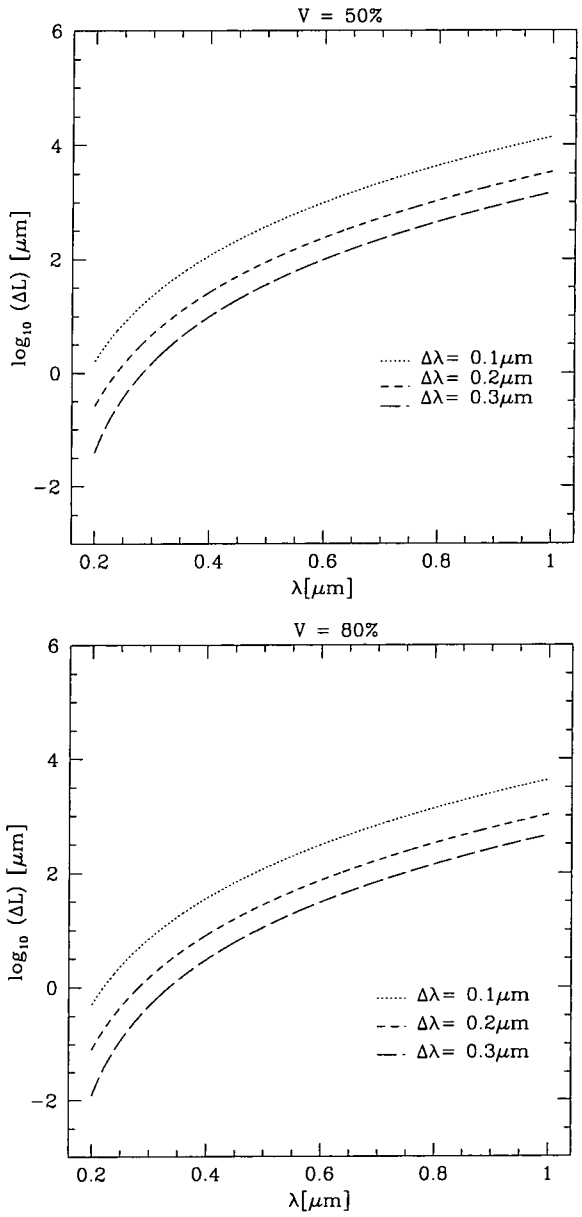



**Figure 4.13.** Tolerance on fibre lengths for (top) $V = 0.5$ and (bottom) $V = 0.8$ for source bandwidths $\Delta\lambda = 0.1\mu$m, $\Delta\lambda = 0.2\mu$m and $\Delta\lambda = 0.3\mu$m.

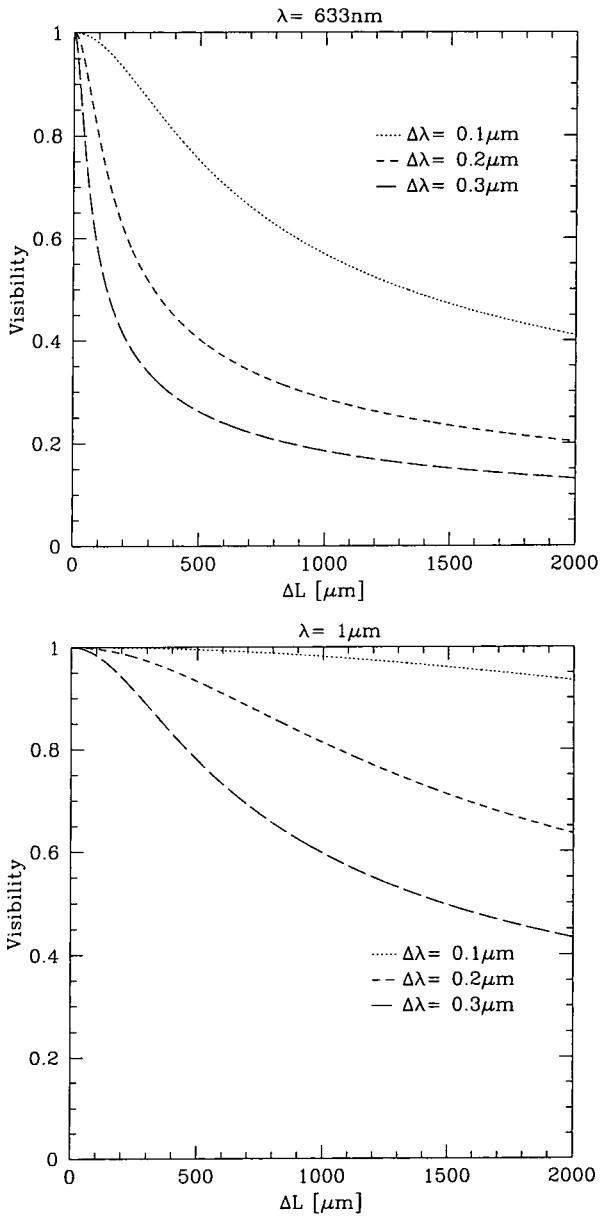



**Figure 4.14.** Fringe visibility as a function of fibre length difference $\Delta L$ (top) for $\lambda = 0.633\mu$m and (bottom) $\lambda = 1\mu$m.

curvature and is found to be

$$\frac{d^2\Delta\Phi(k)}{dk^2} = \frac{d^2\beta}{dk^2}\Delta L \tag{4.41}$$

The higher order terms are not relevant to this approximation so have been neglected.

We now have

$$\frac{d^2\Delta\Phi(k)}{dk^2} = c^2\Delta LD \tag{4.42}$$

or equivalently

$$\Delta L = \Delta\Phi_{\lambda\lambda}/\beta_{\lambda\lambda} \tag{4.43}$$

which means that the length difference between fibre arms can be deduced from the second differential of induced phase with respect to wavelength, provided that the fibres are made of the same material. The abbreviations $\beta_{\lambda\lambda} = d^2\beta(\lambda)/d\lambda^2$ and $\Delta\Phi_{\lambda\lambda} = d^2\Delta\Phi(\lambda)/d\lambda^2$ have been used. The functional form of $\beta_{\lambda\lambda}$ is

$$\beta_{\lambda\lambda} = \frac{4\pi}{\lambda^3}n - \frac{4\pi}{\lambda^2}\frac{dn}{d\lambda} + \frac{2\pi}{\lambda}\frac{d^2n}{d\lambda^2} \tag{4.44}$$

The refractive index function is given by the Sellmeier equation

$$n^2 - 1 = \frac{a_1\lambda^2}{\lambda^2 - b_1} + \frac{a_2\lambda^2}{\lambda^2 - b_2} + \frac{a_3\lambda^2}{\lambda^2 - b_3} \tag{4.45}$$

so that

$$\frac{dn}{d\lambda} = -\frac{\lambda}{n}\left[\frac{a_1b_1}{(\lambda^2 - b_1)^2} + \frac{a_2b_2}{(\lambda^2 - b_2)^2} + \frac{a_3b_3}{(\lambda^2 - b_3)^2}\right] \tag{4.46}$$

and

$$\frac{d^2n}{d\lambda^2} = -\frac{4\lambda^2}{n}\left[\frac{a_1b_1}{(\lambda^2 - b_1)^3} + \frac{a_2b_2}{(\lambda^2 - b_2)^3} + \frac{a_3b_3}{(\lambda^2 - b_3)^3}\right] + \frac{1}{\lambda}\frac{dn}{d\lambda} - \frac{1}{n}\left(\frac{dn}{d\lambda}\right)^2 \tag{4.47}$$

As the dopants of SM fibres do not affect the refractive indices significantly, the difference between the core and cladding being typically $\sim 10^{-3}$, the Sellmeier constants for pure fused silica were used [59]. Table 4.1 lists these coefficients as well as those for various typical dopant concentrations. Equation (4.43) thus

| Dopant (mole %) | $a_1$ | $a_2$ | $a_3$ |
|---|---|---|---|
| None | 0.6961663 | 0.4079426 | 0.8974994 |
| $GeO_2$ (3.1) | 0.7028554 | 0.4146307 | 0.8974540 |
| $GeO_2$ (3.5) | 0.7042038 | 0.4160032 | 0.9074049 |
| $GeO_2$ (5.8) | 0.7088896 | 0.4206803 | 0.8956551 |
| $GeO_2$ (7.9) | 0.7136824 | 0.4254807 | 0.8964226 |
| | $b_1$ | $b_2$ | $b_3$ |
| None | 0.004679148 | 0.01351206 | 97.934002 |
| $GeO_2$ (3.1) | 0.005295810 | 0.01306644 | 97.934000 |
| $GeO_2$ (3.5) | 0.002646230 | 0.01668230 | 97.933900 |
| $GeO_2$ (5.8) | 0.003709450 | 0.01573806 | 97.934020 |
| $GeO_2$ (7.9) | 0.003808952 | 0.01614969 | 97.934010 |

**Table 4.1.** Sellmeier constants for fused silica with various dopant concentations.

served as the principle of the fibre length experiments described in the following Chapters.

The main simplification in the above model has been to neglect the effects of polarisation upon the fringe visibility. These effects and the issue of controlling the SOPs in the system have already been dealt with.

Although the core/cladding diameters and doping levels of real SM fibres do vary, within manufacturing tolerances, it is not necessary to account for these effects. This is apparent if we recall Equation 3.79. The product $b\Delta << 1$, so any variations will be negligible from the point of view of dispersion.

## 4.7 Output plane and Beam recombination

A major advantage afforded by fibres in an interferometric system is the ability to spatially remap the two dimensional input plane to a linear output plane. This greatly simplifies the problem of matching baselines to fringe signals encountered in non–fibre dilute aperture systems. This concept is shown in Figure 4.15, for a single large telescope interferometer.

The number of fringes $n_f$ formed in the recombination plane is a function of the separation between the fibre outputs (really the fundamental mode field radii) and the numerical aperture of the individual outputs. This can be found

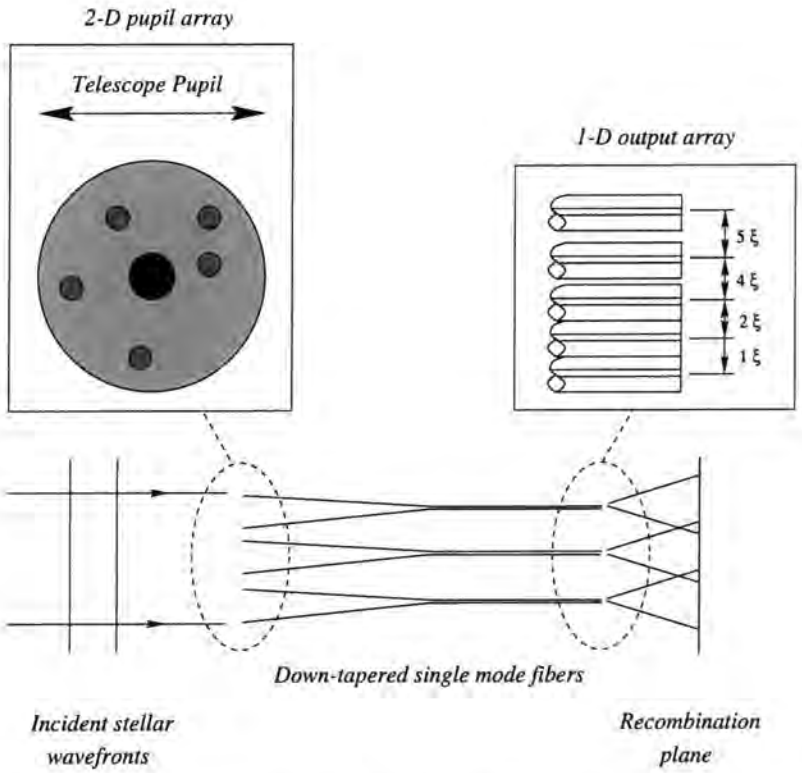



**Figure 4.15.** Spatial multiplexing of a 2-D telescope pupil plane to a 1-D interferometer recombination plane.

geometrically to be

$$n_f = \frac{2d\tan\theta}{\lambda} - \frac{d^2}{\lambda F} \tag{4.48}$$

where $\theta$ is the numerical aperture semi-angle, $d$ is the distance between the fibre outputs and $F$ is the distance from the outputs to the detection plane.

A small number of fringes is desired for each fringe to have a large width. This is so that a low resolution (few pixel) detection system can be used. Having fewer pixels naturally gives a faster readout, permitting short exposure times. From (4.48), either having the outputs extremely close together or decreasing their numerical apertures (NA) can be seen to give fewer fringes in the recombination plane. The latter method is the less desirable since it involves light loss. A possible strategy for decreasing the NA would have been to match the fibre outputs to microlenses. Alternatively single mode up-tapers could have been spliced onto the outputs, although the technology required for fusion splicing

was not available to us at the time of writing.

For example, with Equation (4.48), assuming $\lambda = 633$nm and a fibre NA= 0.1, when the core spacing $d = 125\mu$m (this being the standard cladding diameter of silica SM optical fibres) about 40 fringes will be found if the recombination plane is 10mm from the subapertures. In contrast, when $d = 52\mu$m (a standard spacing of commercially available V–grooved positioning templates) approximately 16 fringes will be formed.

### 4.7.1 Cladding etching

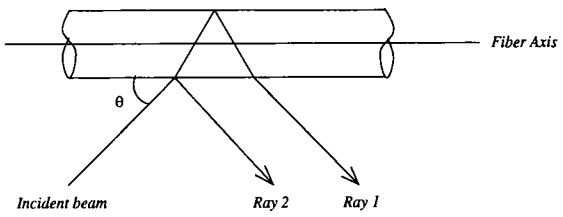



**Figure 4.16.** Collimated laser beam incident on fibre.

In light of the results of the fringe width estimations and the considerations mentioned at the beginning of this Section, it was decided that etching the outer diameters of the fibres might be a reasonable way to construct a useable output plane. Having the fibre outputs closer together would produce wider fringes and therefore require fewer pixels in the detector.

By immersing a small length of the silica fibres in 48% dilute hydrofluoric acid, it was possible to alter the fibre diameters with a reasonable amount of control[8]. The results of the trials are described in Section 5.2.

The method employed to accurately determine the cladding diameter was an elementary one: the fibre was illuminated by a HeNe beam at oblique incidence (Figure 4.16). The reasons for using this technique instead of a calibrated microscope are convenience and accuracy. Lengths of fibre can be awkward to handle under a microscope, especially when there is limited space, so may limit

[8]The rate of etching is a diffusion limited process, so agitation of the solution accelerated the etching process. Therefore some care had to be taken when calibrating the etch times.

the measurement precision.

Assuming the laser to be collimated, which is valid to first order, we see constructive interference when

$$m = \frac{2d}{\lambda}\sqrt{n^2 - \cos^2\theta} - \frac{1}{2} \tag{4.49}$$

$\theta$ is the angle of the ray to the fibre axis, $n$ is the refractive index and $m$ is an integer referring to the diffraction order. The number $N$ of maxima located between $\theta_1$ and $\theta_2$ is

$$\begin{aligned} N &= m_1 - m_2 \\ &= \frac{2d}{\lambda}\{\sqrt{n^2 - \cos^2\theta_1} - \sqrt{n^2 - \cos^2\theta_2}\} \end{aligned} \tag{4.50}$$

which gives $d$ as

$$d = \frac{N\lambda}{2(\sqrt{n^2 - \cos^2\theta_1} - \sqrt{n^2 - \cos^2\theta_2})} \tag{4.51}$$

Naturally $\theta_1$ and $\theta_2$ must be less than a critical angle $\theta_{crit}$ for fringes to appear. This is evident if it is realised that the fibre is acting as a cylindrical lens, in the paraxial approximation. When the tilt angle is equal to $\theta_{crit}$, the collimated beam is focused by the first surface onto the second surface, so the fibre acts as a back-reflector. Thus only a bright spot at infinity is seen. For a tilt angle greater than $\theta_{crit}$, two different rays are reflected from the second glass-air interface, giving a complicated three beam interference pattern, which is not considered further here.

The precision of the measurements is sensitive to the accuracy to which $\theta_1$ and $\theta_2$ and $n$ are known. When compared to microscope determinations, this *side-beam illumination* method was accurate to approximately 1$\mu$m. This implies that most of the error derives from the assumption that $n = 1.46$ at 633nm.

## 4.8 Summary

In this Chapter we have discussed the individual issues that need to be addressed when constructing a SM fibre interferometer. The concept of visibility phase calibration through a closure phase was introduced, supported by a brief discussion of the effects of certain variables on the SNR. The key results for a tilt corrected interferometric system based on SM dielectric waveguides were seen to be that

1. the optimum exposure time $t_{\text{exp}}$ is about $2\tau_0$ for photon starved observations.

2. the optimum number of fibres for a test-bed system such as DAFI is (arguably) five.

3. the optimum effective diameter of DAFI's input subapertures is around $4r_0$.

We showed that high quality flats on fibre endfaces can be reproduced easily through controlled cleaving of a number of 125$\mu$m diameter silica fibres.

Polarisation in interferometry was reviewed and a solution for SOP control was discussed in terms of bending birefringence.

A field theoretic method was given for calculating the effects of optical path differences in an interferometer based on weakly guiding dielectric waveguides on the fringe contrast. The results of applying this formalism to silica SM optical fibres demonstrated that broadband visible fibre interferometry is viable, assuming that the input coupling problem is resolved. At visible wavelengths, even over a 300nm bandwidth, a fibre length difference of about 100nm still produces 80% contrast (assuming that chromatic dispersion is the only significant degrading effect on the fringe visibility).

Finally, the output array concept was presented requiring the etching of standard diameter fibres. A simple method for measuring the fibre diameters was described.

# Chapter 5

# Making a multi–fibre interferometer

Rules for Successful Experiments:

- Experiments should be reproduceable - They should all fail in the same way.
- Always keep a record of your data. It indicates that you have been working.
- Do not believe in miracles - Rely on them.

## 5.1 Introduction

In this Chapter the steps leading to the construction of a multi–fibre interferometer are discussed. We are concerned with three main issues to achieve a high SNR, which are (in order of importance)

1. Fibre endface quality.
2. Residual length differences between fibres.
3. Output endface positions.

A technique was developed to cleave a series of silica fibres to high relative accuracy and to measure their residual length differences; this will be described in Section 5.3. As discussed in Section 4.7.1, it is desirable to have as large a fringe spacing as possible in the recombination volume. Therefore, tests were

conducted to examine the possibility of etching the output plane fibres. The effects of etching the fibre endfaces was a concern; the results are detailed in Section 5.2. The method of producing the output assembly is given in Section 5.5 together with the final position errors. Finally, the measured results from the multi–fibre interferometer are given in Section 5.6.

## 5.2 Cladding etching — preliminary trials

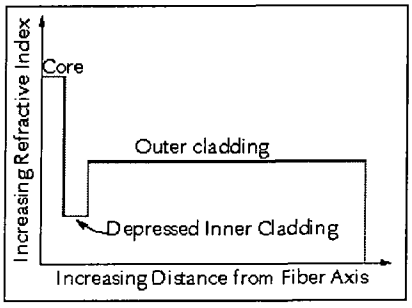



**Figure 5.1.** Typical doubly–clad, single–mode $GeO_2$ core doped fibre refractive index profile.

To investigate the effects of acid etching on the endface quality of the fibres, preliminary trials were conducted.

Two sets of test samples were prepared, one set being step–index $GeO_2$ doped SM fibres ($\lambda_c = 633$nm), where the dopants were added to the core. In the absence of any real refractive index profile data, it is assumed that the fibre is similar to the standard doubly clad profile, illustrated in Figure 5.1.

The second set consisted of step–index matched–clad F doped fibres (offcuts of the tapered fibres). The refractive index profile of their preform is shown in Figure 5.2. The core region is pure silica, surrounded by a lower index cladding fabricated by adding fluorine dopants to silica. External to that is a pure silica mechanical support which has no direct bearing on the waveguide structure. The reason for utilising such a structure was convenience since the preform manufacturing process had already been established[1].

[1]The structure had been originally designed for a high radiation environment. In such

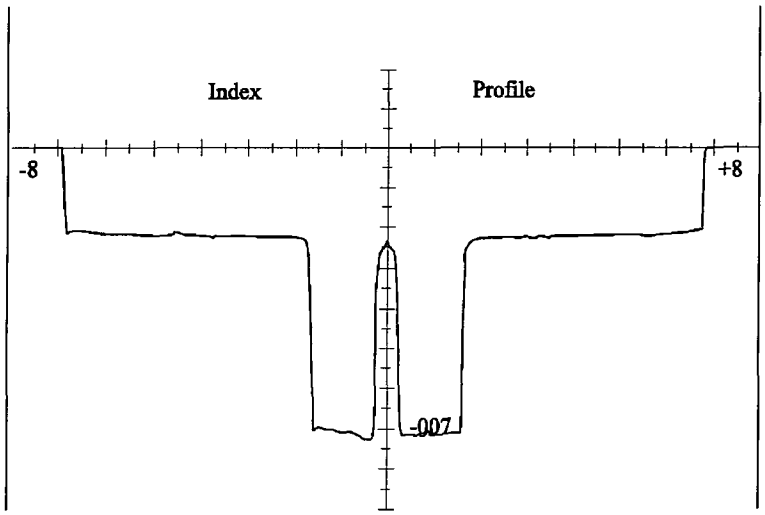



**Figure 5.2.** Tapered fibre preform refractive index profile measured at $\lambda = 633$nm. The step size is $20\mu$m and the refractive index difference between the core and cladding $\Delta n \sim 4.5 \times 10^{-3}$. The index rise at the edges of the plot is due to the immersion of the preform in index–matching fluid. Courtesy of Oxford Electronics.

They were prepared by cleaving to produce a flat endface, then stripping back 4cm of the acrylate buffer coating with a buffer stripping tool. 2cm of the bare fibre was immersed in 48% dilute hydrofluoric acid (HF) and left for a prescribed time without agitation. The fibre diameters were then measured by the method of side–beam interference, described in Section 4.7.1.

Since unbuffered HF is hydrophilic, the fibres may be etched, rinsed, dried then re–etched without a significant variation in the etched diameter change *vs.* time, when compared with samples that are etched once for the same total time. Figure 5.3 illustrates the results of etching as measured by the side–beam interference technique. The non–linear behaviour exhibited by the standard fibres below $\sim 50\mu$m diameter is due to the diffusion limit of the etchant.

The Figures 5.4 and 5.5 show photomicrographs of fibre endfaces taken with a SEM. The inconsistency in the diameters of the $GeO_2$ fibres may be attributed to the fact that the buffers were also immersed in the acid, thus affecting the

surroundings, high purity silica fibres are more resistant to irradiation defect absorption than less pure glasses.

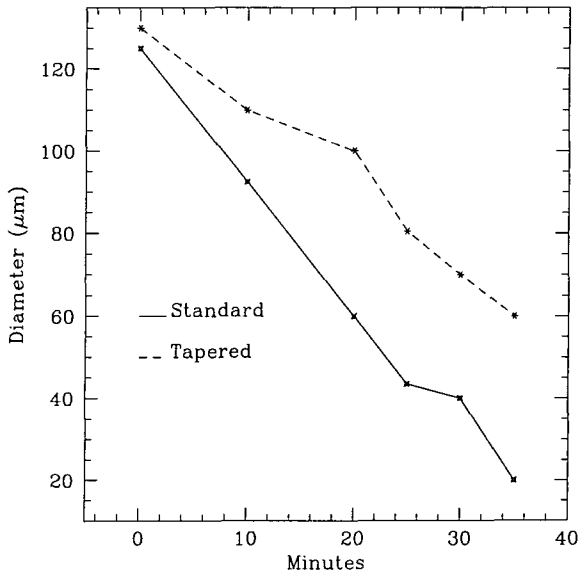


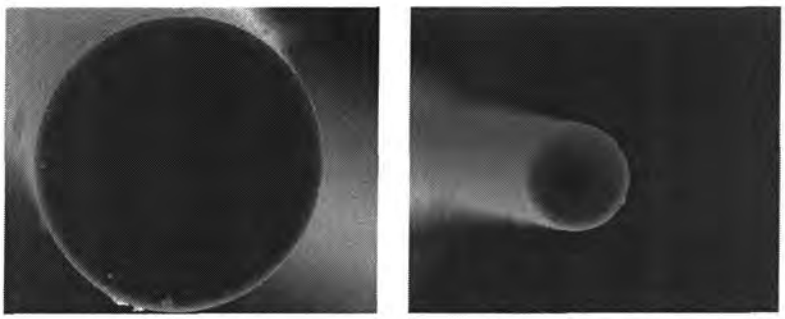

**Figure 5.3.** Fibre etch rate in 48% solution HF with no agitation. The deviation from linear in the tapered fibre etch rate is due to experimental error. See main text for details.

concentration of $F^-$ ions in the vicinity of the glass. The F fibres, of which only a subset are shown, were highly consistent in terms of diameter. Dust particles are clearly visible on the smaller endfaces, underlining the importance of cleanliness for the fibres.

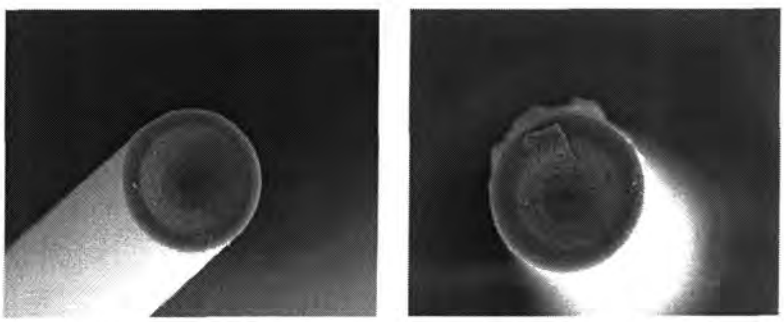
It is interesting to note that the differential etch rates have essentially created core–sized microlenses on the fibre endfaces. This undoubtedly alters the $f_{\#}$ of the fibre and needs to be characterised further.

A Twyman–Green was also used to inspect sample endfaces before and after etching. The fibre endfaces were magnified with a $\times 10$ microscope objective, the results being shown in Figure 5.5. The correlation with the SEM photomicrographs in Figure 5.6 is quite clear. The uneven illumination is due to misalignments of the fibre endfaces with the incident beam optical axis.

Post–etching beam profiles of the fibres shown in Figure 4.4 are given in Figure 5.7. One can see that the effects of mist and hackle, evident on the pre–etched beam profile of fibre (b), may be removed through etching. Fibre (c) was possibly damaged upon insertion into the fibre holder, resulting in its



Diameter = 125$\mu$m Diameter = 41$\mu$m



Diameter = 56$\mu$m Diameter = 68$\mu$m

**Figure 5.4.** $GeO_2$ fibre endface SEM photomicrographs. The 125$\mu$m endface shows a sample immediately after cleaving. The other three fibres were immersed in 48% HF for 25 minutes. The inconsistency in the diameters may be due to the fact that the buffers were also immersed in the acid. (Magnification ×700.)

poor quality output.

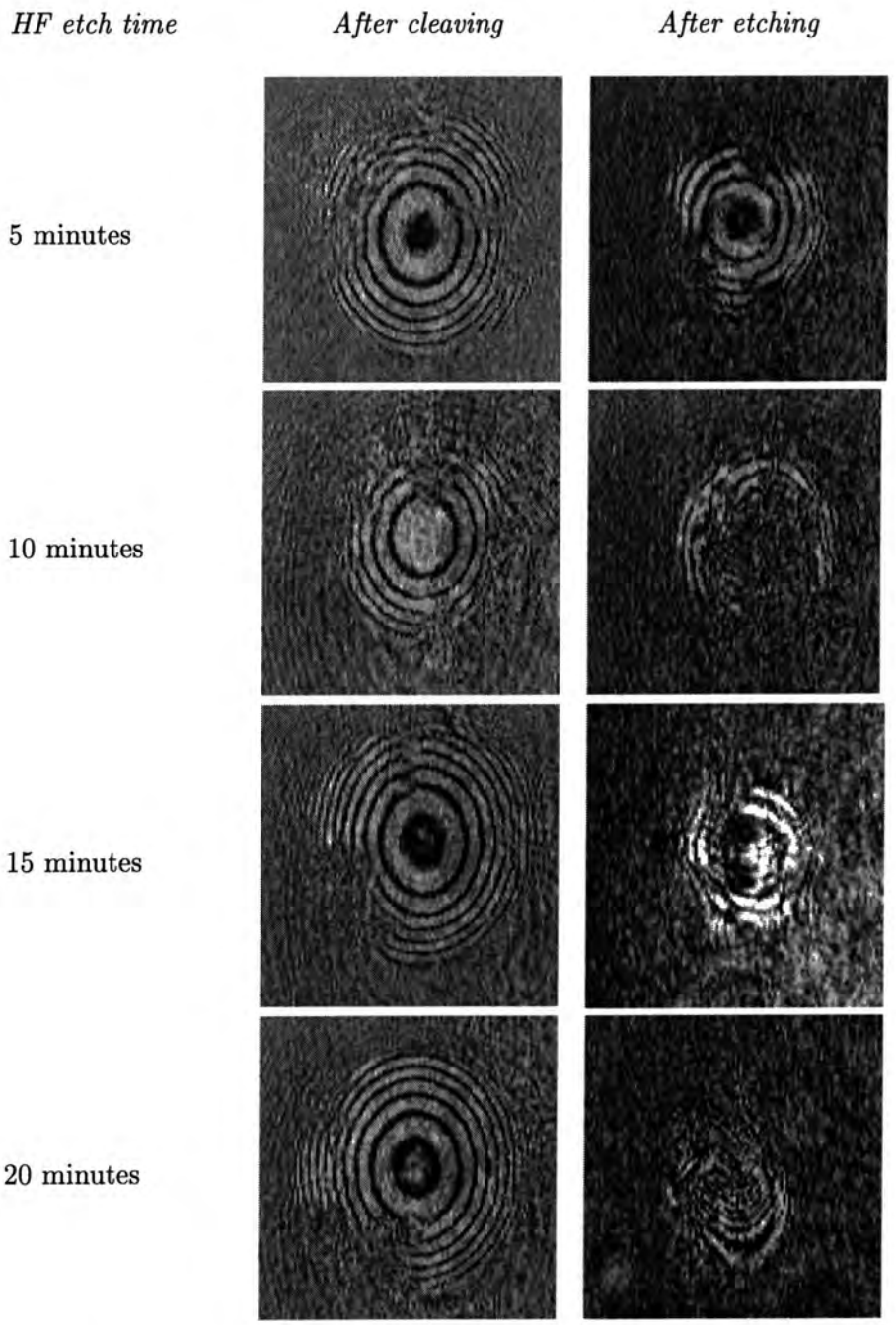



**Figure 5.5.** F doped fibre endfaces as inspected by a Twyman–Green interferometer (with background subtraction). Compare with Figure 5.6.

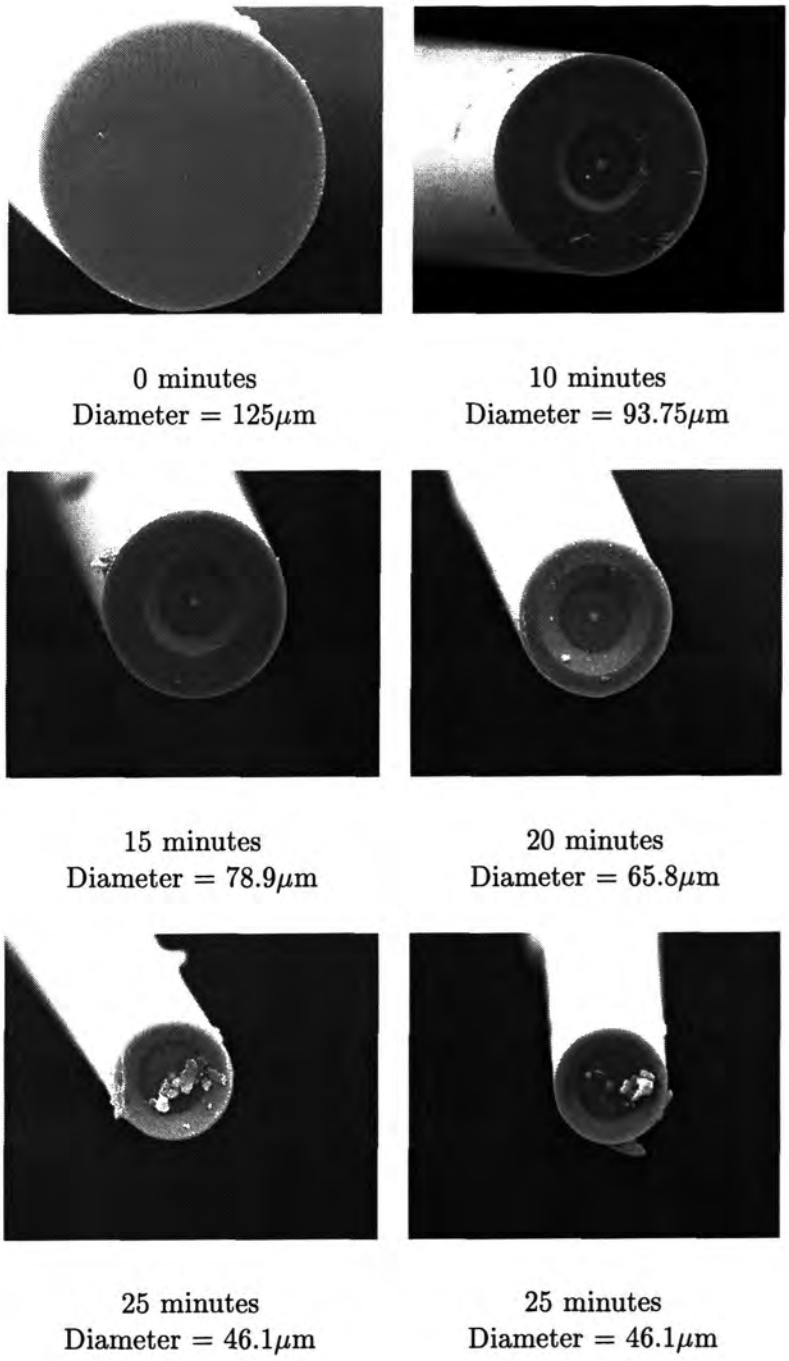



**Figure 5.6.** F doped fibre endface SEM photomicrographs. (Magnification ×700.)

(a) Good beam profile.

(b) Good beam profile.

(c) Unusable beam profile.

(d) Good beam profile.

(e) Good beam profile.

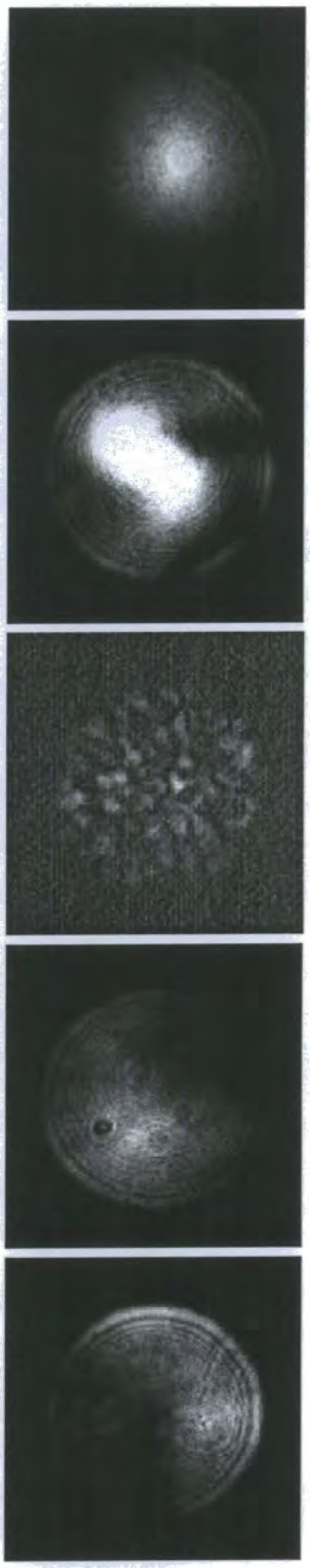


**Figure 5.7.** Beam profiles of fibres shown in Figure 4.4 after etching.

## 5.3 Fibre length equalisation

Previous work by Burnett and Jones [60] has demonstrated that it is possible to produce short (nominally 1.5m) lengths of fibre to a relative accuracy of 10.8$\mu$m. For a source centered on $\lambda = 500$nm with $\Delta\lambda = 300$nm, this translates to a fringe visibility $V = 0.8$ so is very good for interferometry.

Their method required that the fibres be clamped at their midpoints on a stable mounting block. With the cleave tool situated at one end of the optical table, a fibre was stretched across the tool and tensioned by a small mass. Upon cleaving, the clamp was rotated and the other end of the fibre was prepared in the same way. The accuracy is therefore dependent on the reproducibility of both the fibre catenary (from clamp to cleave tool) and the clamp positioning.

Our technique is based on providing a datum position for one endface and cleaving the other side of the fibre. Therefore the principal variable is the reproducibility of the catenary. For standard 125$\mu$m fibres whose outer diameters (even including the buffer) are constant to within a few microns along the length of the fibre, the catenaries are easy to reproduce accurately. Tapered fibres, due to the manufacturing process, have larger outer diameters at the tapered endface than at the normal end. This can lead to a decrease in flexibility of the fibre, so requiring more care when tensioning the fibre.

### 5.3.1 Preliminary trials

The optical fibres used for the tests and prototyping were standard telecommunications fibres. These are step–index silica, where the cladding was lightly doped with germanium dioxide. The cladding diameter was 125$\mu$m with a mean acrylate buffer diameter of $\sim 250\mu$m, provided by Oxford Electronics.

One (flat) endface of the fibre served as the test component in a Twyman–Green interferometer, illustrated in Figure 5.8. The fibre was clamped in PTFE jaws, so as not to permanently deform the acrylate buffer coating, since this ultimately deteriorates the transmission performance.

The interferometer was illuminated by a spatially filtered, collimated HeNe laser, and the light focused through a $\times 10$ microscope objective onto the fi-

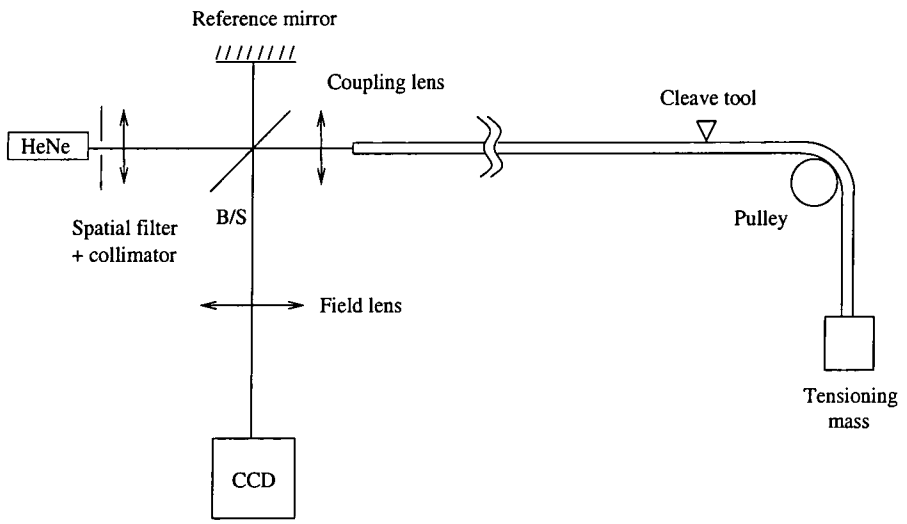



**Figure 5.8.** Cleaving method. The datum end is a classical Twyman–Green interferometer illuminated by a spatially filtered HeNe laser. It also served to measure the fibre endface quality.

bre endface. The detector was a standard television CCD. A small mass was attached to the other end of the fibre, which was stretched over the cleaving unit and a pulley. A 3cm section of the monomer acrylate buffer was removed[2] from that end prior to cleaving. The fibre was thoroughly wetted to counter any static charge.

After finding the interference fringes in the recombination plane, the reference end was positioned such that the Airy disk filled the detector's field–of–view. This is clearly the focal point of the objective and acted as the datum for all the fibres (see Figure 5.9). If the fibre endface is not aligned on axis with respect to the microscope objective, a highly comatic fringe pattern can be produced.

The fibre was clamped down in the cleaver carefully, so as not to significantly alter the shape of its catenary. Misalignments of the fibre position from the pulley to the datum also had to be kept to a minimum. The mass was removed

[2]The acrylate was stripped with a soldering iron. Alternative methods of buffer removal from the middle of a fibre proved to be less satisfactory. The heat from the iron can set up stresses in the glass, rendering it more brittle, so care must be taken.

**Figure 5.9.** Twyman–Green interferograms, moving through the focus of the microscope objective. The central image occurs when the fibre endface is at the focal plane of the Twyman–Green.

after clamping the fibre in the cleave tool, then the cleave tool tensioner was engaged[3].

### OPD measurement

A Mach–Zehnder interferometer illuminated by a reference laser diode and a tungsten source was used to determine the length difference between any given fibre pair, as shown in Figure 5.10.

Initially the inputs of the fibre pair were aligned with a $\times 3$ travelling microscope such that the image of one endface overlapped the position of the other. Since the positional tolerance of the core within the cladding is very high, this alignment was sufficient to locate the cores along the transverse ($x$–$y$) directions. The interferometer was illuminated with the 0.3mW laser diode, operating above its threshold current. Prior to overlapping the output beams, a polariser $P_1$ was placed in front of the laser source and an analyser $P_2$ in front of the detector. They were oriented with their polarisation axes perpendicular to each other. The polarisation controllers $PC_1$ and $PC_2$ were individually adjusted by altering the angle of each Mickey ear to minimize the throughput. Although for a wide bandwidth source there is some residual polarisation

[3]Prior calibration of the tensioning force applied to the fibre is vital to achieving a good cleave. The diamond blade of the York FK11 can be damaged if the fibre fails to cleave, so tension adjustment during calibration should **always** be decreased from a high value to avoid this.

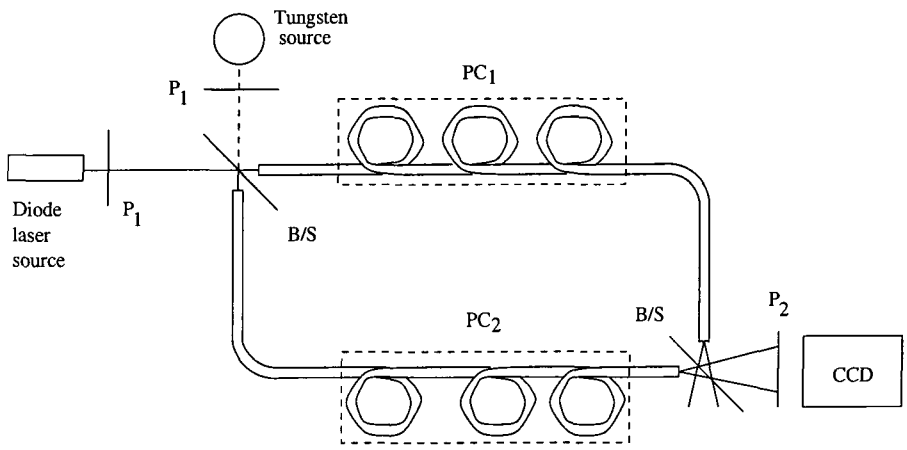



**Figure 5.10.** The fibre lengths were measured with a white light fibre Mach–Zehnder interferometer. Polarisation control was achieved by looping the fibres to form Mickey ears ($PC_{1,2}$). Polarisers $P_1$ and $P_2$ were oriented such that their axes were parallel during the measurements. See text for details.

leakage, this is not an issue since it is less than 1%.

With the axes of $P_1$ and $P_2$ parallel, the fibre outputs were aligned to form tilt fringes at the recombination plane. Reducing the current to just below threshold increased the spectral bandwidth (to the maximum width of the laser gain envelope), so decreasing the coherence length to about 100$\mu$m. This helped to locate the white light fringe position in the transverse ($z$) direction.

The broadband fringes formed with a tungsten source were recorded with a Cohu CCD camera[4] attached to a frame grabber. CCD images of the laser source provided the reference wavenumber. To improve the signal–to–noise ratio, good background subtraction was necessary. The background for the laser source was measured by coadding ten frames with a fringe signal. The visibility phase was offset by a small amount between frames by moving a fibre output along the $z$ axis. The white light background was obtained by offsetting an output along the $z$ axis until the fringes were out of the field–of–view.

The visibility curve of the background subtracted laser interference pattern was calculated by an inverse 2D Fourier transformation. By identifying its

[4]The CCD read out at standard television rates.

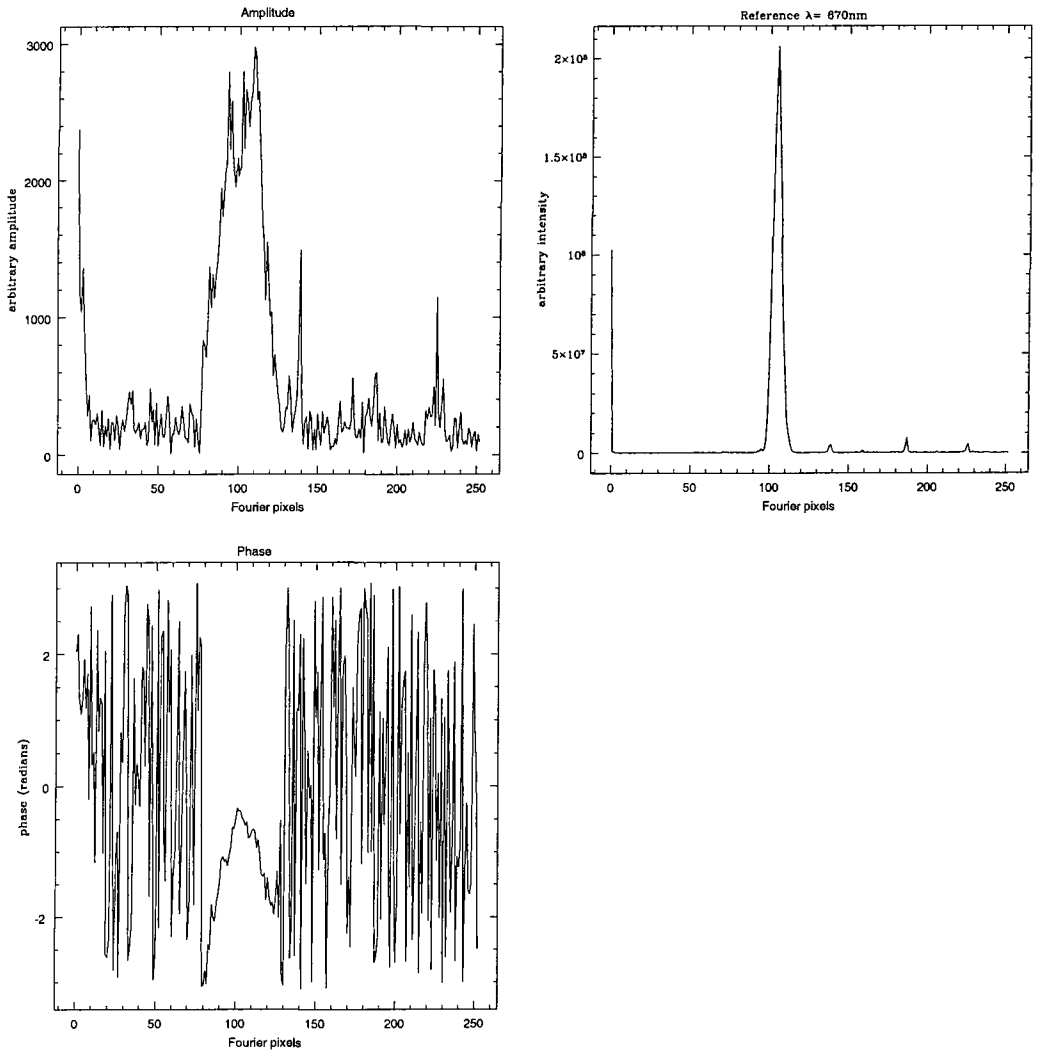



**Figure 5.11.** Interferogram amplitude and phase from the white light fibre Mach–Zehnder and its corresponding wavelength reference power spectrum.

| Fibre pair | Bandwidth (endpoints) [$\mu$m] | $\beta_{\lambda\lambda}$ [$\mu m^{-3}$] | $\Delta\Phi_{\lambda\lambda}$ [$\mu m^{-2}$] | $\Delta L$ [$\mu$m] |
|---|---|---|---|---|
| 1,2 | 0.318 (0.452 to 0.771) | 87.917 | -143.059 (-163.842) | -1.63 (-1.69) |
| 1,3 | 0.300 (0.462 to 0.762) | 87.0162 | -228.388 (-227.858) | -2.62 (-2.35) |
| 1,4 | 0.379 (0.466 to 0.845) | 70.9102 | -145.098 (-124.675) | -2.05 (-1.29) |
| 1,5 | 0.415 (0.383 to 0.798) | 103.915 | -148.887 (-194.094) | -1.43 (-2.00) |
| 1,6 | 0.429 (0.322 to 0.750) | 145.544 | -457.414 (-336.058) | -3.14 (-3.47) |
| 2,3 | 0.346 (0.432 to 0.778) | 92.1566 | -337.044 (-334.508) | -3.66 (-3.45) |
| 2,4 | 0.355 (0.394 to 0.749) | 111.551 | -149.824 (-113.699) | -1.34 (-1.17) |
| 2,5 | 0.437 (0.379 to 0.816) | 101.328 | -178.072 (-141.738) | -1.76 (-1.46) |
| 2,6 | 0.335 (0.412 to 0.747) | 105.289 | -303.544 (-344.210) | -2.88 (-3.55) |
| 7,8 | 0.276 (0.434 to 0.709) | 107.484 | -119.327 (-141.704) | -1.11 (-1.46) |

**Table 5.1.** Results of Fibre cleaving trials. Fibres 1 to 6 were cleaved to be nominally 1.13m, while 7 and 8 were 2.75m in length. The values in parentheses of $\Delta\Phi_{\lambda\lambda}$ and $\Delta L$ use the bandwidth common to all measurements (0.466 to 0.709$\mu$m) where $\beta_{\lambda\lambda} = 96.971\mu m^{-3}$.

peak position, the wavenumber calibration in terms of the Fourier plane could be extracted.

The amplitude and phase of the background subtracted white light visibility curve was then calculated. Clearly, the phase produced can have a $2\pi$ ambiguity since it is the argument of the real and imaginary parts of the FTs. For the more highly dispersed interferograms it was necessary to unwrap the phase information. The full width at half maximum of the amplitude data was required to constrain the region of phase data which was unwrapped. A second order polynomial fit was then applied to the phase data. From this was obtained the mean curvature $\Delta\Phi_{\lambda\lambda}$ over the detected bandwidth.

The value of $\beta_{\lambda\lambda}$ was calculated by fitting a second order polynomial to $\beta$ to the measured bandwidth. $\Delta L$ was then found using Equation (4.43).

| Fibre triple | | | $\Delta L_{ij}$ | $\Delta L_{ik}$ | $\Delta L_{jk}$ | $\epsilon$ |
|---|---|---|---|---|---|---|
| $i$ | $j$ | $k$ | [$\mu$m] | [$\mu$m] | [$\mu$m] | [$\mu$m] |
| 1 | 2 | 3 | -1.63 (-1.69) | -2.62 (-2.35) | -3.66 (-3.45) | -2.67 (-2.79) |
| 1 | 2 | 4 | -1.63 (-1.69) | -2.05 (-1.29) | -1.34 (-1.17) | -0.92 (-1.57) |
| 1 | 2 | 5 | -1.63 (-1.69) | -1.43 (-2.00) | -1.76 (-1.46) | -1.96 (-1.15) |
| 1 | 2 | 6 | -1.63 (-1.69) | -3.14 (-3.47) | -2.88 (-3.55) | -1.37 (-1.77) |

**Table 5.2.** Results of "closed loop" measurements for Table 5.1. The loop error is defined as $\epsilon = (\Delta L_{ij} - \Delta L_{ik}) + \Delta L_{jk}$. The values in parentheses refer to the common bandwidth results.

6 fibres of nominally 1.13m in length were produced of which 9 pairs were measured as described. The results are given in Table 5.1. The mean accuracy for cleaving fibres of nominal length 1.13m can be seen to be -2.28$\mu$m, with a standard deviation $\sigma$ of 0.78$\mu$m. The bandpass endpoints and thus the bandwidth variations shown are due to a combination of signal–to–noise considerations and calibration errors. Fibres of nominally 2.75m were also produced to high accuracy, the maximum length being limited only by the optical table.

Of course, the accuracy of the wavelength calibration affects the accuracy of the bandwidth determination. Even assuming the calibration is perfect, it is natural to expect that the detected bandwidths for different pairs of fibres will not be equal for a given signal–to–noise ratio. This is because of the transmission bandwidth limitation imposed by chromatic dispersion effects.

To check the calibration accuracy, we recalculated $\Delta L$ for the fibres using the value of $\beta_{\lambda\lambda}$ for the detected bandwidth common to all the 1.13m fibre pairs (0.466 to 0.709$\mu$m). The results have a mean accuracy of -2.27$\mu$m, with $\sigma = 0.93\mu$m, which is consistent with the result found previously.

A series of "closed loop" comparisons were also made to determine the measurement error. That is to say, for fibres $i, j$ and $k$, $\Delta L_{ij}$, $\Delta L_{ik}$ were measured to deduce the value of $\Delta L_{jk}$. This was compared to measurement (Table 5.2), resulting in a mean closed loop errors of -1.73$\mu$m, $\sigma = 0.65\mu$m for the full detected bandwidth calculation and -1.82$\mu$m, $\sigma = 0.6\mu$m for the common bandwidth determination. The residual closed loop errors show that the measurements are still noise dominated, since they are not much smaller

than the individual length differences.

An accuracy of 2.28$\mu$m is within tolerance for sources of $\Delta\lambda = 0.3\mu$m centered on $\lambda_c \approx 0.4\mu$m, even if the strictest criterion of $V \geq 80\%$ is imposed. So even broadband interferometric imaging in the visible is possible with short (< 5m) optical fibres.

Active compensation schemes can be used to account for the residual $\Delta L$ should the need arise, although the residuals of the cleaving trials indicate that path length mismatches will not be the main source of error.

## 5.4 Interferometer fibres

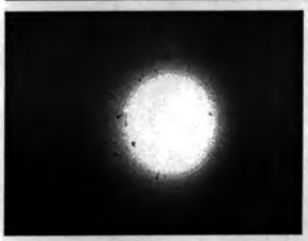
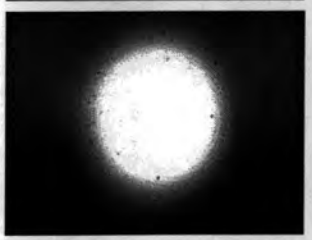
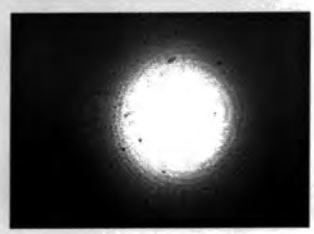
To produce the prototype multi–fibre interferometer, ten fibres were cleaved to the same lengths. The fibre beam profiles were inspected (Figure 5.12). It is clear that the removal of the buffer layer with a soldering iron prior to cleaving sets up stresses within the silica since the variation in the cleave quality is larger than expected. The output endfaces were then etched down to 50$\mu$m. The endfaces which emit poor quality beam profiles can be corrected by etching, so the higher quality ends were always selected to be the inputs.

Because of the high level of illumination, it is possible to see the cladding modes, which appear in the background as a faint Airy diffraction pattern.

The post–etching output beam profiles were again inspected; four were selected to make the prototype multi–fibre interferometer. Their profiles are shown in Figure 5.13. The length differences of these fibres, as measured post–etching, are given in Table 5.3.

| Fibre pair | Bandwidth (endpoints) [$\mu$m] | $\beta_{\lambda\lambda}$ [$\mu$m$^{-3}$] | $\Delta\Phi_{\lambda\lambda}$ [$\mu$m$^{-2}$] | $\Delta L$ [$\mu$m] |
|---|---|---|---|---|
| 1,7 | 0.369 (0.670 to 1.039) | 30.314 | 30.991 | 1.02 |
| 1,9 | 0.335 (0.558 to 0.893) | 86.924 | 51.973 | 0.60 |
| 1,10 | 0.838 (0.558 to 1.396) | 33.776 | 22.000 | 0.65 |

**Table 5.3.** The length differences of the multi–fibre interferometer fibres. The third measurement is clearly subject to some wavelength calibration error, although this is not significant since an exact value for $\Delta L$ is not necessary.

| *Fibre & comments* | *Endface A profile* | *Endface B profile* |
|---|---|---|
| (1) |  |  |
| (2) A: Hackle on core causing scatter. |  |  |
| (3) |  |  |
| (4) |  |  |
| (5) A: Mist on cladding. |  |  |

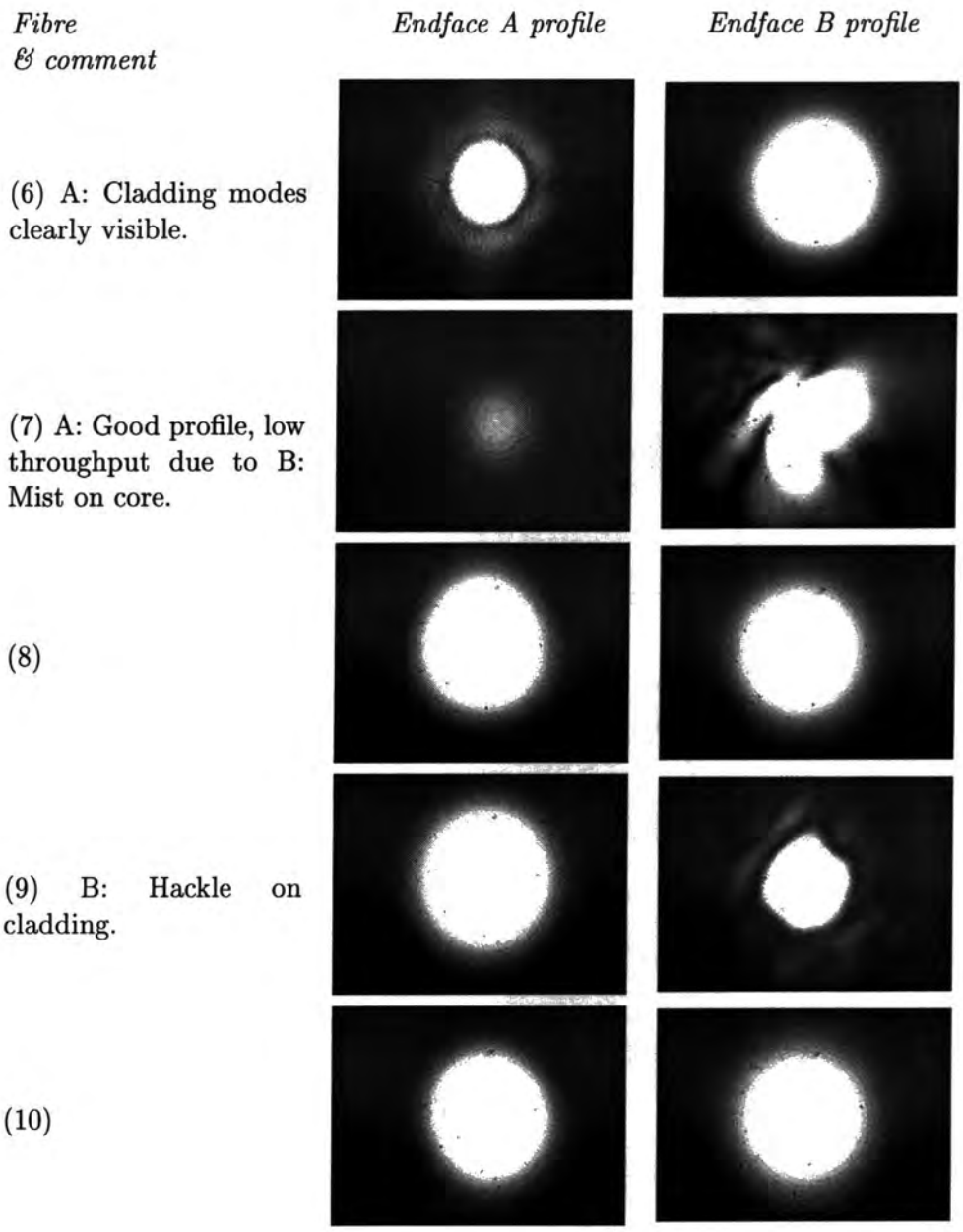



**Figure 5.12.** Pre–etched beam profiles of the ten $GeO_2$ SM fibres.

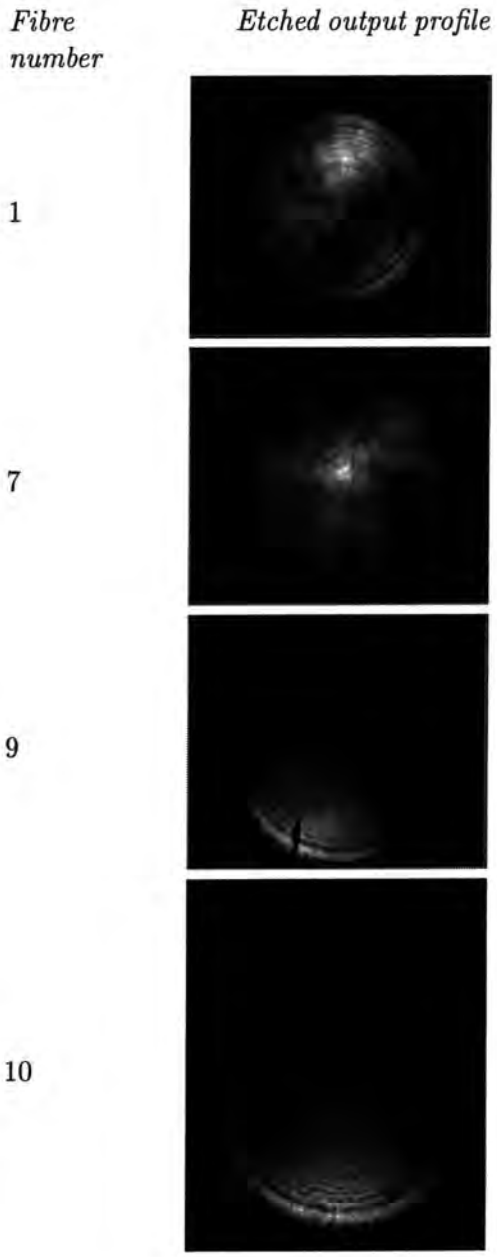



**Figure 5.13.** The uneveness of illumination is an artifact of the magnifying objective glass in front of the CCD, which was not aligned along the fibre output axes.

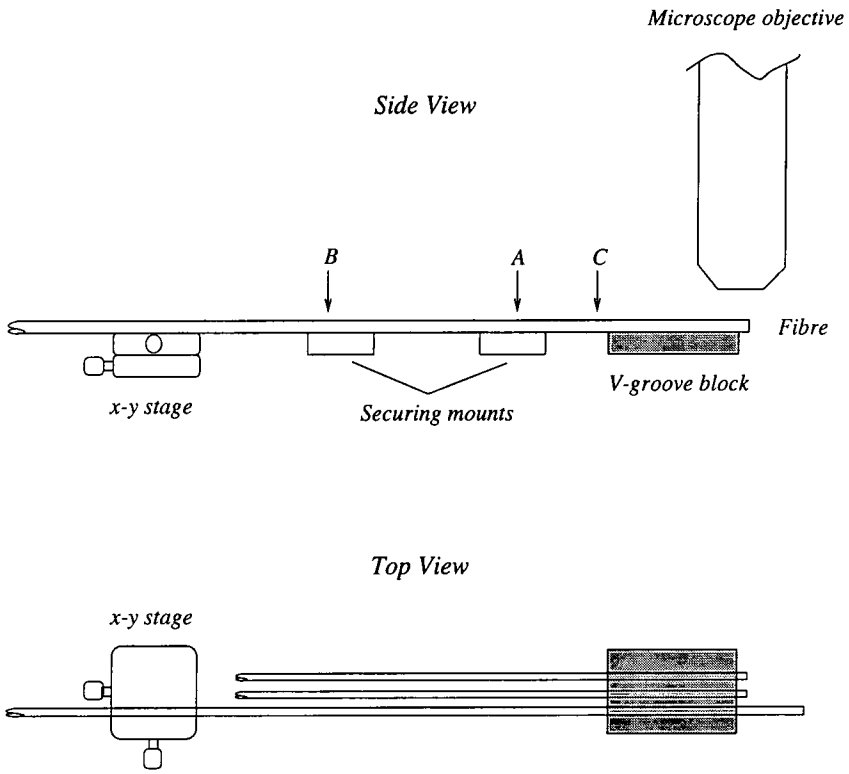



**Figure 5.14.** Schematic of output plane assembly procedure. The final ground glass mounting block, which was placed on top of the fibres above the V–grooves, is not shown.

## 5.5 Interferometer output plane

The objectives in constructing this prototype output plane are (a) to demonstrate that it can form high visibility fringes, (b) to demonstrate and characterise the non–redundancy of the output positions in the transverse direction and (c) to characterise the longitudinal tolerance when positioning the fibres.

The assembly method of the multi–fibre interferometer output plane is illustrated schematically in Figure 5.15. A commercially available 52$\mu$m spacing V–groove block[5] was used as the mounting template for the etched fibres.

Initially the etched ends of the optical fibres were degreased by immersing them in a bath of ethanol. When dry, they were laid across the V–groove block, with an inter–fibre spacing of about 3 grooves. The fibres were then

[5]produced by Optometrics.

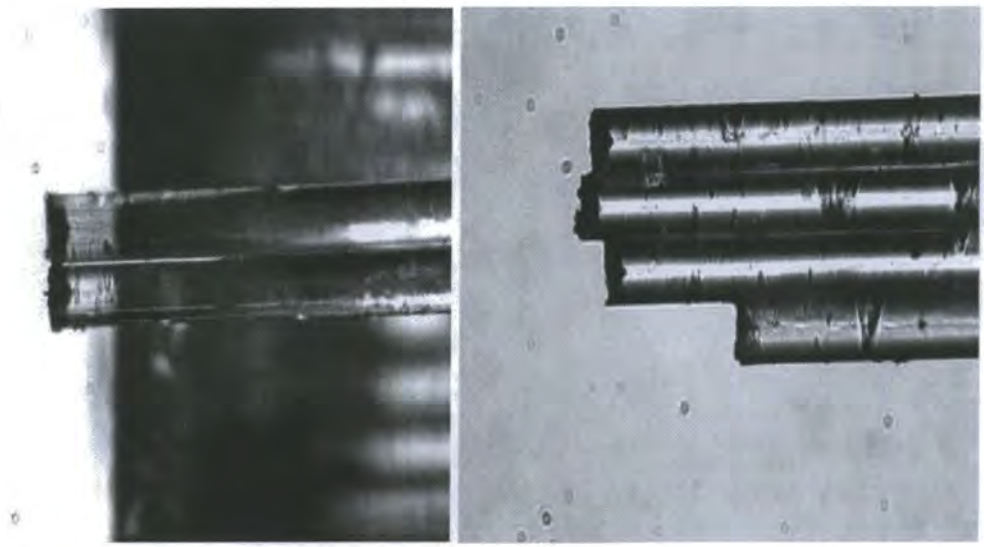


**Figure 5.15. Left :** Trial two–fibre assembly. The alignment V–grooves can be seen (defocused) in the background. **Right :** Four–fibre assembly before glueing. (Compare with Figure 5.18.)

gently secured in position (at point A) with a strip of adhesive (masking) tape. One fibre was chosen to act as a position reference. The other fibres were affixed in turn to the $x-y$ stage and adjusted in the $z$ direction until they were satisfactorily aligned with the reference fibre.

The fibres were then straightened and secured at point B, as a strain relief for A. A small drop of water was put at point C, using a syringe, to bring the fibres together with surface tension. A ground glass mounting block was carefully laid atop the fibres to ensure that they were flat.

A drop of adhesive was then placed at point C with a syringe and left to advance between the fibre surfaces and the mounting block by capillary action. The adhesive properties had to be such that

- It was relatively low in viscosity so that it wetted the fibre surfaces completely. It also had to be hydrophilic, given the method of pulling the fibres together.

- It had a short set and cure time, once the transition process from fluid to solid was initiated.

These requirements pointed to a one–part liquid, specifically from the family

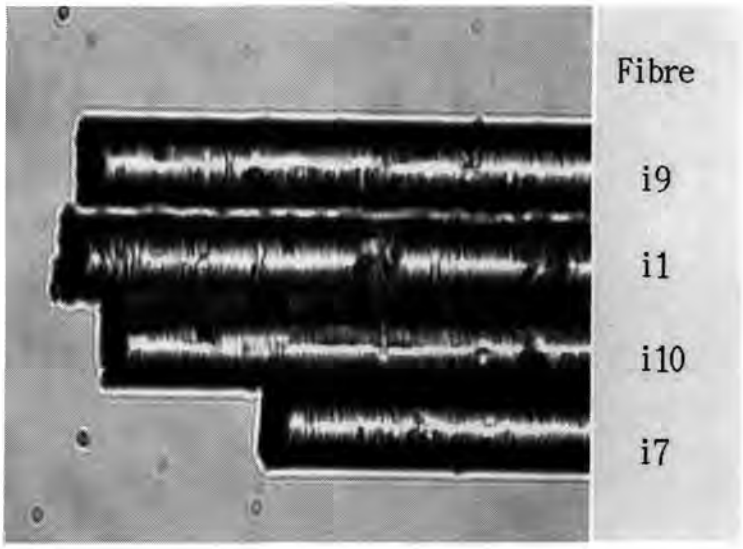



**Figure 5.16.** Microscope photo of output plane assembly after glueing. The top fibre separated from its neighbour by about 6.4$\mu$m after removal of the V–groove template.

of alkyl 2-cyanoacrylate monomers (Superglue). The strong bonds are formed rapidly as a result of anionic polymerization. This polymerization is initiated by weakly basic species — essentially water or alcohol, at ambient temperature. The bond setting time is rapid (< 1min).

When cured, the V–groove block was removed and the fibres potted with an additional layer of cyanoacrylate for mechanical reinforcement.

Figures 5.16 and 5.18 show the assembled output plane. The former photograph shows the $z$ displacements of the fibres with respect to $i9$ are 17$\mu$m for $i1$, $-28\mu$m for $i10$ and $-122\mu$m for $i7$, where the negative values indicate a distance behind the reference datum. These $z$ axis alignment errors are simple to compensate for by adjusting the relative positions of the input endfaces.

The tolerance for the $z$ axis errors can be deduced by means of "depth–of–focus" arguments. This is defined to be the maximum movement away from the reference "focus" position (fibre endface) that may be made without causing serious deterioration to the fringes. In Figure 5.19, $AP$ is a spherical wave diverging from the geometrical source B. If a similar source is placed at C, points on AP are no longer equidistant from C so that disturbances do not

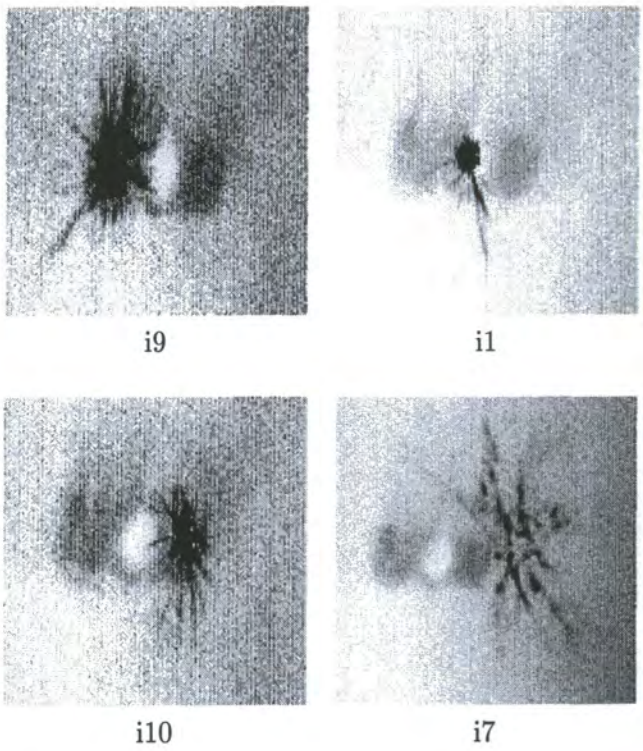



**Figure 5.17.** Front view of the output plane assembly with each fibre illuminated in turn. The image greyscale has been negated so that black is high intensity.

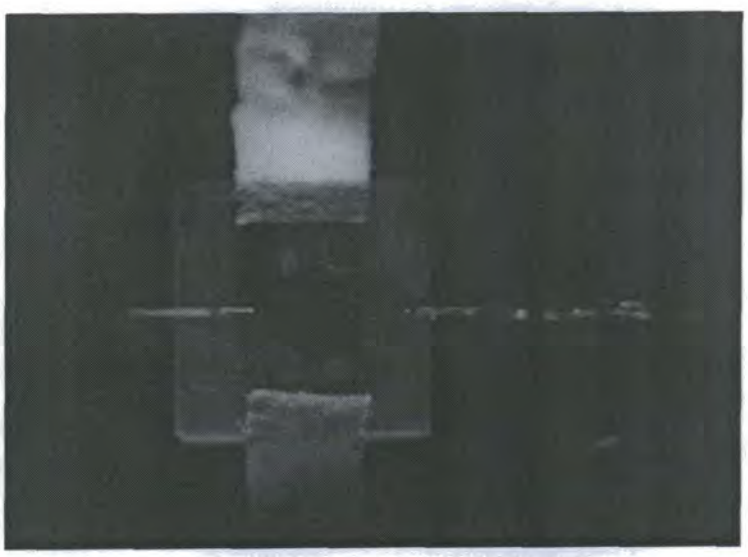


**Figure 5.18.** Entire output plane assembly. The output plane is on the left hand side.

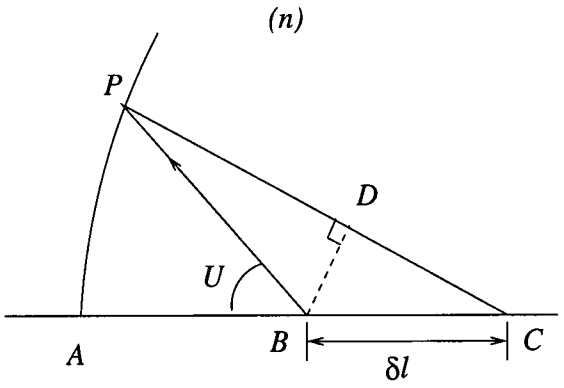



**Figure 5.19.** Definitions of variables for the "depth of focus" estimation. $n$ is the refractive index of air.

arrive in phase. Therefore, the interference pattern is reduced in visibility. Let us assume the Rayleigh criterion, so the maximum phase difference tolerable is $\pi/2$.

For a "depth–of–focus" $\delta l$, and $\delta l << AB$, then

$$PD = PB = AB \tag{5.1}$$

The optical path difference is given by

$$\begin{aligned} n(AC - PC) &= n(BC - DC) \\ &= n\ \delta l(1 - \cos U) \\ &= 2n\ \delta l \sin^2(U/2) \\ &= \frac{\lambda}{4} \end{aligned} \tag{5.2}$$

when $\delta l$ is the depth–of–focus. Hence

$$\delta l = \pm \frac{\lambda}{8n \sin^2(U/2)} \tag{5.3}$$

If $\lambda = 0.6\mu$m, $\sin U = 0.1$ and $n \approx 1$, we find that $\delta l = 30\mu$m. So of the fibres in the prototype, $i7$ is out of tolerance. This result also implies that once a fibre is with the $z$ tolerance, the effects of shadowing by its neighbour(s) should

be nearly negligible. This can be argued more concisely if we approximate the output of a fibre by a spherical Gaussian beam. Gaussian beams have an effective region of collimation around the beam waist, which in this case is the exit of a fibre, called the Rayleigh range $z_R$. Using the relevant Equation (given in Appendix A), we see that $z_R = 47\mu$m which is larger than $\delta l$.

## 5.6 Results of the multi–fibre interferometer

To test the quality of the output plane assembly, the interferometer was illuminated pairwise as shown in Figure 5.21. The input coupling mechanism to the fibres was not optimised since the main purpose of this prototype assembly was to test the output plane concept practically. The light source used was the 3mW laser diode run below its threshold current.

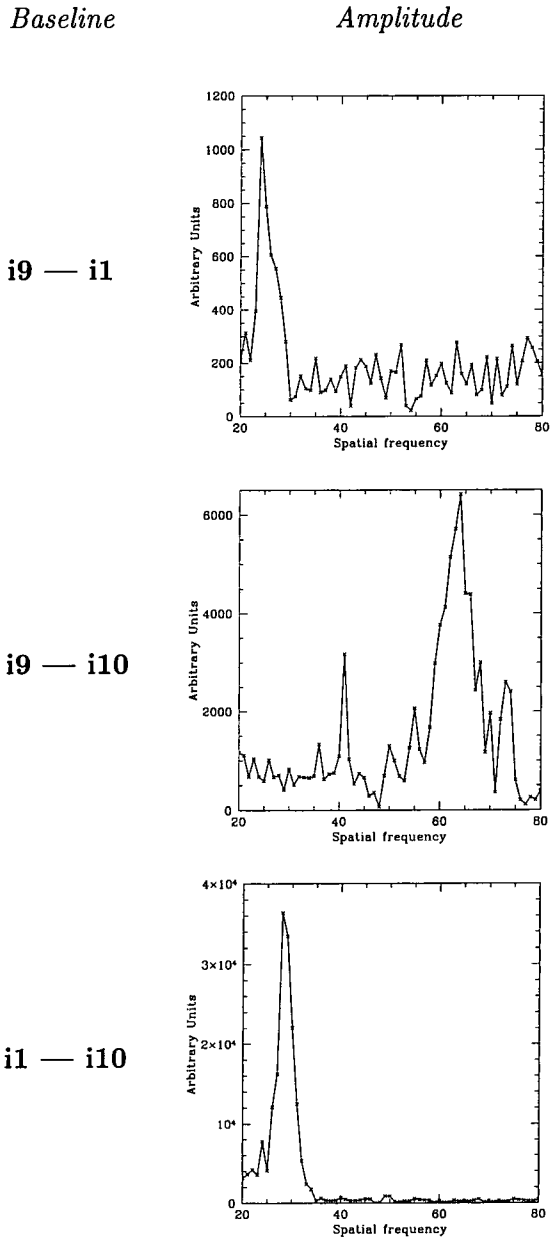



**Figure 5.20.** Spatial frequency amplitudes and phases of fringes generated from the 3mW diode laser run under its threshold current. The background has been subtracted from all except the *i1 – i10* baseline. The peak spatial frequencies of the fibre pairs can be seen to be 24, 64 and 28 Fourier pixels, so the non–redundancy is clear. The post–processing SNR of each baseline's amplitude is 2.5, 3 and 60 respectively.

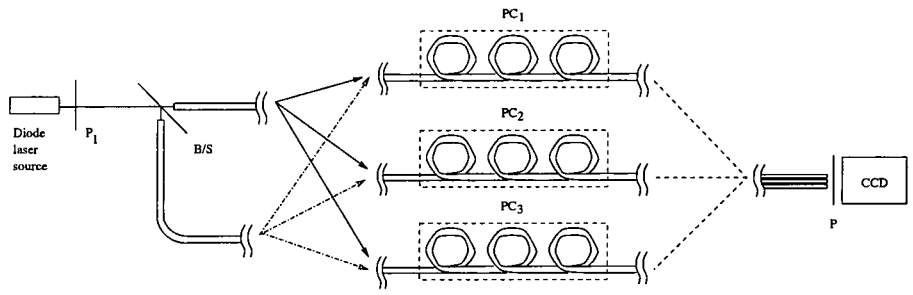



**Figure 5.21.** Schematic of the interferometer illumination geometry.

It is evident from the photographs in Figure 5.17 that fibre $i7$ did not exhibit an ideal profile or throughput. This is attributed to damage while handling the fibre before assembly. It has also been shown to be some $92\mu$m out of tolerance. As a consequence of these considerations this fibre was disregarded.

With the pairwise input configuration it was not practical to attempt a phase closure measurement across the three baselines. Therefore we present the fringe signal that was obtained from each fibre pair. The signals were background subtracted where necessary, then 2 dimensionally Fourier transformed. The amplitudes of the spatial frequencies are shown in Figures 5.20. The non–redundant spacing between the fibre outputs can be seen. This confirms the fact that it is straightforward to construct a linear array of non–redundant fibres[6]. The background was obtained by merely offsetting one of the fibre inputs along the $z$ direction until the straight fringes moved out of the field–of–view.

The resultant SNR of each baseline depended almost entirely on the input coupling of light into the fibres since the optical path length differences were negligible and the output endfaces showed no detectable scatter. The fringe visibility was good ($\sim 68\%$ for the highest contrast fringe) as can be seen from Figure 5.22. A cross–section through this (non–background subtracted) image is also presented.

It was also possible to find the white light fringes from the tungsten source. This acted as a test of the input coupling of the baselines; however the quality

[6]In fact it is difficult to place the outputs redundantly!

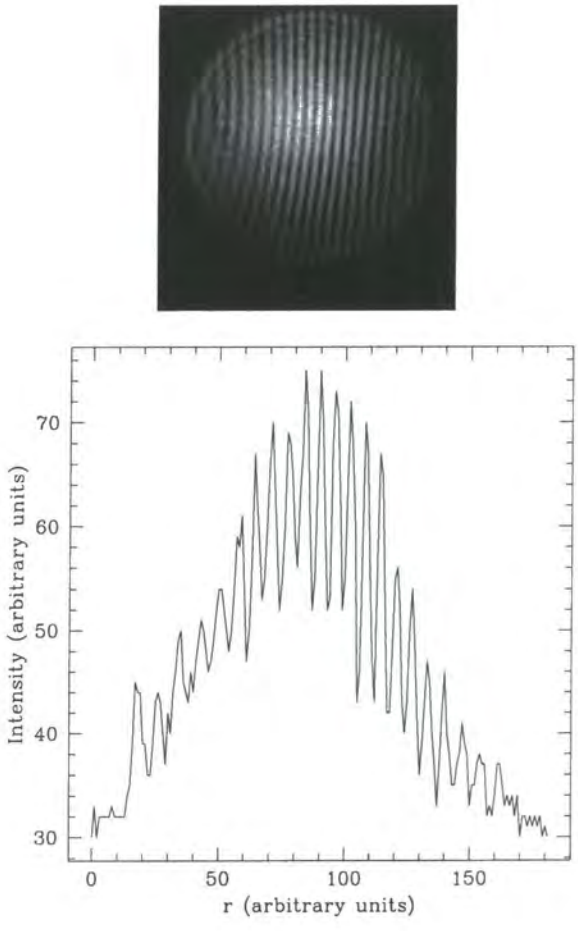



**Figure 5.22. Top:** Fringes formed by a short coherence length source, using fibres $i1$ and $i10$. **Bottom:** Cross–section through the image. The fringe visibility is $\sim 68\%$ for the best contrast fringe.

of the input plane for this prototype was not of interest.

Another point to note about this prototype is the fragility of the mounting, since the fibres are merely potted with adhesive. The final instrument will need to be rugged enough to survive the journey to and from the WHT, so the output array mount will have to be extremely well protected.

## 5.7 Summary

In this Chapter, the construction of a multi–fibre interferometer with standard $GeO_2$ doped SM optical fibres was described. The results of the output plane were positive, showing that the instrument concept is workable in practice. The main issue underlined by the results reported in this Chapter is the importance of high input coupling and throughput, as the optical path length differences have been minimised from the outset.

# Chapter 6

# DAFI

To guess is cheap. To guess wrong is expensive.

*Old Chinese Proverb*

## 6.1 Introduction

The **D**ilute **A**perture **F**ibre **I**nterferometer (DAFI) is intended to be a visible ($0.6 - 0.8\mu$m) waveband interferometer based on adiabatically tapered single mode optical fibres.

This Chapter describes the original optical design of the instrument which unfortunately was incorrect for the $f_{\#}$ of the tapered fibres fabricated. The resulting loss in coupling is characterised, which demonstrates that the design is still useable albeit rather inefficiently. A corrected design for the coupling lenses is calculated, although not implemented due to a lack of time.

Finally the stages in the construction of DAFI to date are described, with an outline of the procedures needed to complete the interferometer.

## 6.2 Original Optical Design

Our principal objective in the design stage is to maximise the light coupled into the optical fibres to achieve the highest possible SNR. It is assumed that the quality of the fibre arms is perfect (ideal endfaces, no optical path mismatches)

and that there is no difficulty in implementing the polarisation controllers.

This defined the requirement that each fibre input be matched to a segment of ELECTRA's deformable mirror, which would eliminate the low-order tip-tilt drift of the stellar wavefront across the subaperture. It so happens that one segment of the DM is about $4r_0$, which is ideally suited for the input coupling to a fibre (see Section 4.2.1).

It also motivated the desire to minimise the number of separate transmissive coupling elements before the fibre inputs, to eliminate alignment errors. Theoretically this can be achieved by fabricating a microlens directly on the fibre endface by, for example, laser machining[1]. In the absence of such a possibility, adiabatically tapered (circular core) SM fibres were selected, as has already been discussed, and the fibre coupling lenses were designed to be single component elements.

### 6.2.1 Conceptual Design

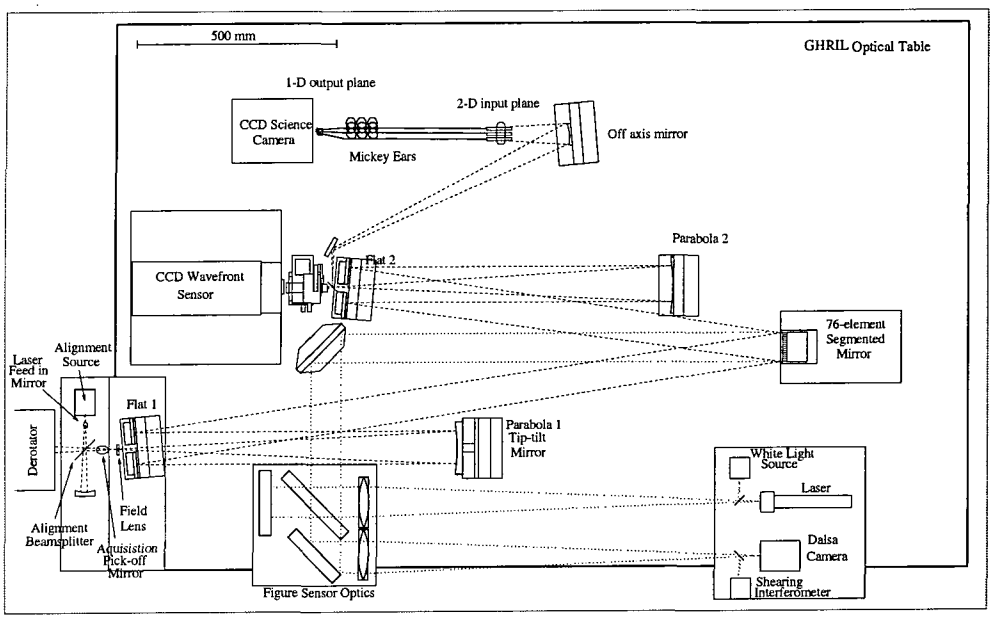



**Figure 6.1.** Conceptual layout of the Electra system with DAFI. The fibre interferometer components are not drawn to scale.

[1] Opto & Laser Europe — issue 44, October 1997.

The concept for DAFI is illustrated in Figure 6.1. An off–axis (non–flat) mirror refocuses the diverging f/11 beam from ELECTRA towards the 2 dimensional input plane of DAFI. This plane is made up of 5 subapertures, each subaperture being a tapered SM fibre with an associated small coupling lens. We distinguish between the diameter of the input plane and the diameter of each subaperture by referring to the former as the full pupil and the latter as a subaperture pupil. The fibres are held straight for the full length of the taper, then pass through the Mickey ear polarisation controllers before feeding to the 1 dimensional output array. Depending on the stability of the polarisation controllers, their achromaticity and the accuracy of their correction, it may or may not be necessary to include an analyser between the output array and the detector.

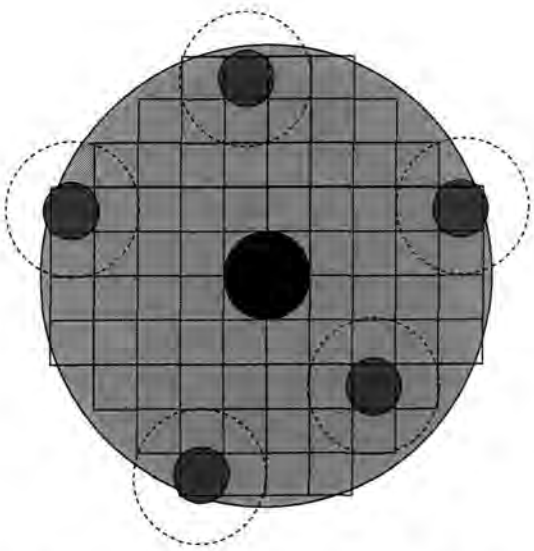


**Figure 6.2.** Schematic diagram of DAFI's input geometry. The telescope entrance pupil is shown superimposed upon ELECTRA's 76 segment DM. The telescope secondary is represented by the black circle in the centre. The interferometer subapertures are the five grey circles, arranged in a non–redundant fashion. The dotted circles are the estimated minimum diameters of the subaperture holders.

The input plane geometry is shown in Figure 6.2. The first simplification made was to deliberately oversize each input subaperture's pupil with respect to its corresponding DM segment. This eases the manufacturing and alignment

tolerances considerably. It is assumed that the smallest diameter for the subaperture mounts is on the order of three segment widths within the full pupil. It is immediately clear that the input subaperture arrangement can be remapped as desired, permitting either non–redundant, redundant or a combination of input baselines to be set.

It was decided that a fixed field plate in a tip–tilt adjustable holder was the most straightforward method of mounting the fibres with their input coupling lenses.

### 6.2.2 System

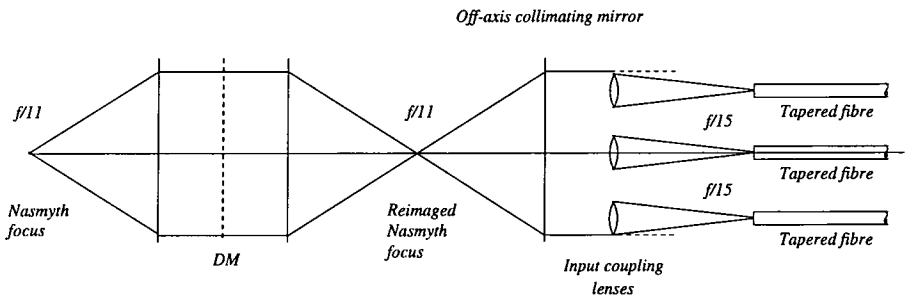



**Figure 6.3.** Optical design diagram for DAFI's input plane.

The design of the optical system was based on the assumption that the SM fibres would have 3:1 tapers, converting from a f/15 beam at the tapered end to f/5 at the standard end. The taper ratio was selected on the basis that the sample fibres were fabricated to this specification and were seen to behave as predicted.

Obviously, the diameter of the input coupling mini–lenses is a compromise between minimising the mini–lens' focal length and accounting for the restrictions imposed by the mechanical mountings. Given the tooling available in–house at the time the components were being designed, the smallest component lenses that could easily be accomodated (and handled) would have a diameter of 4mm. With a mini–lens of $d = 4$mm (clear aperture), a focal length of $f = 60$mm is necessary to obtain an f/15 beam. Since each lens is matched to

| Curvatures (mm) | Thicknesses (mm) | Glass Types |
|---|---|---|
| 44.3269 | 1.7920 | BK7 |
| -24.6356 | 2.1940 | SF1 |
| -55.3018 | | |

**Table 6.1.** Coupling lens design parameters. The full lens diameter is 4.4mm with its clear aperture being 4mm.

a DM segment, this fixes the diameter of ELECTRA's pupil to 40mm.

To steer the beam from the reimaged Nasmyth focus (f/11 for the WHT), an off–axis collimating mirror was required. Although the best option for this case is an off–axis paraboloid, this is an extremely costly option. An off–axis spherical mirror was decided upon in lieu of the ideal shape. Naturally this imparts astigmatism (see Figure 6.7) across the whole pupil. However the reduced diameter of a subaperture pupil can minimise this effect to an acceptable level.

### Coupling lenses

A single component coupling lens was desirable to make the alignment as straightforward as possible. On the other hand, to obtain such a slow $f_{\#}$ requires a thick lens (since the idealised "thin" lens approximation breaks down), with a very small degree of curvature for its surfaces.

The slow f–ratio of the lens aperture effectively removes all the aperture and some field dependent aberrations[2] leaving only transverse chromatism, distortion and astigmatism. Distortion is negligible for a good lens, which is trivial to produce at this size. Astigmatism is also strongly field dependent, so is minimal in the case of this almost zero aperture system. Therefore a thick singlet of small diameter would principally result in too much chromatic aberration.

Chromatic correction can be implemented by using two glasses (crown and flint). The resulting secondary spectrum can be reduced to the well known optical materials' limit ($\Delta f \sim f/2000$ where $f$ is the focal length) by a judicious choice of the two types of glasses and surface curvatures.

[2]That is to say, spherical, coma and longitudinal chromatism for the former, and field curvature for the latter.

The coupling lenses were designed with the Zemax ray–tracing package. The sole constraint imposed was that the image at focus be diffraction limited at f/15 over the wavelength range $600 \leq \lambda \leq 900$. The design model included the optics for the WHT up to the Nasmyth focus and the off–axis spherical collimating mirror. The Nasmyth f/11 beam is conserved through ELECTRA, so (to a good approximation) there was no need to include the AO system explicitly. An off–axis spherical fold mirror was also included in the ray–trace, positioned so that it essentially collimated the beam. The mirror details were based on an aluminium mirror that was available:

- Focal length $f_m = 440$mm
- Diameter $D_m = 50$mm

The lens parameters are given in Table 6.1. Figures 6.4 and 6.5 show the polychromatic spot diagrams at and through focus. The wavefront (for $\lambda =$ 600nm) at focus is shown in Figure 6.7. No optical coatings were included in the modelled system.

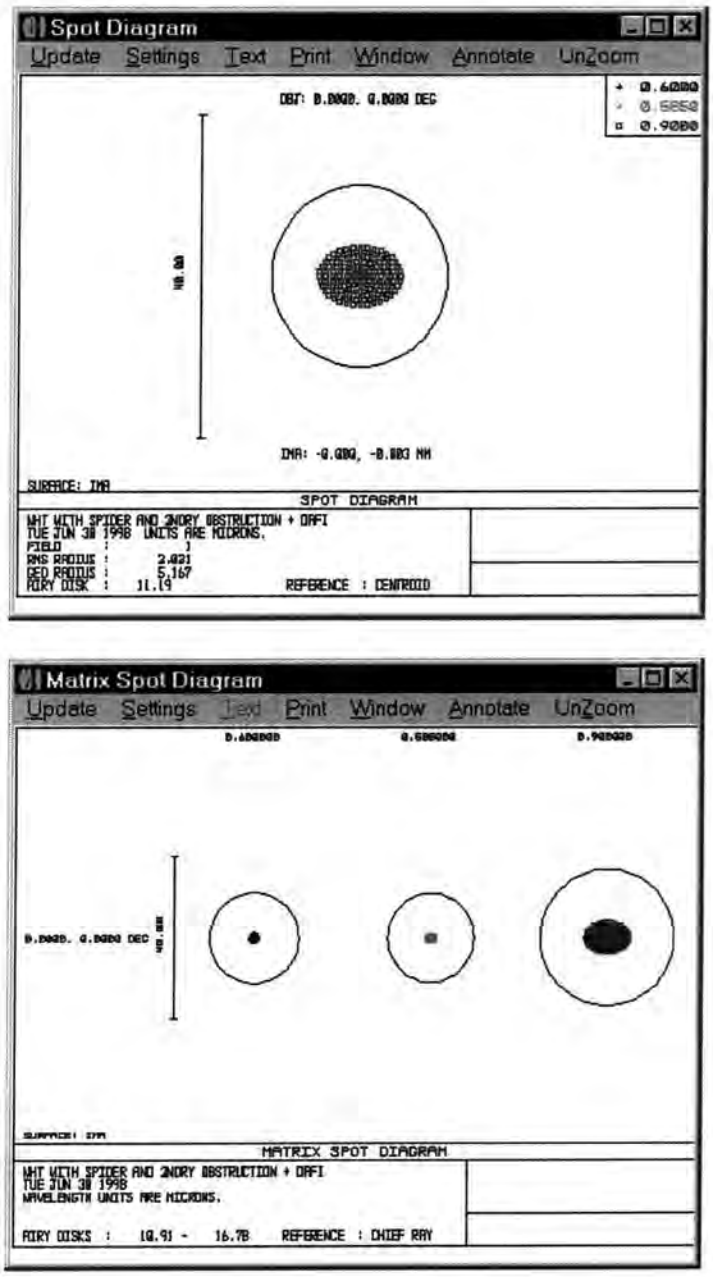



**Figure 6.4. Top:** Polychromatic spot diagram at focus for the f/15 coupling lens. The circle represents the Airy disk (radius = 11.19$\mu$m), so it can be seen that the ray aberrations are within the diffraction limit. **Bottom:** Matrix of spot diagrams for each wavelength with their appropriate Airy disks.

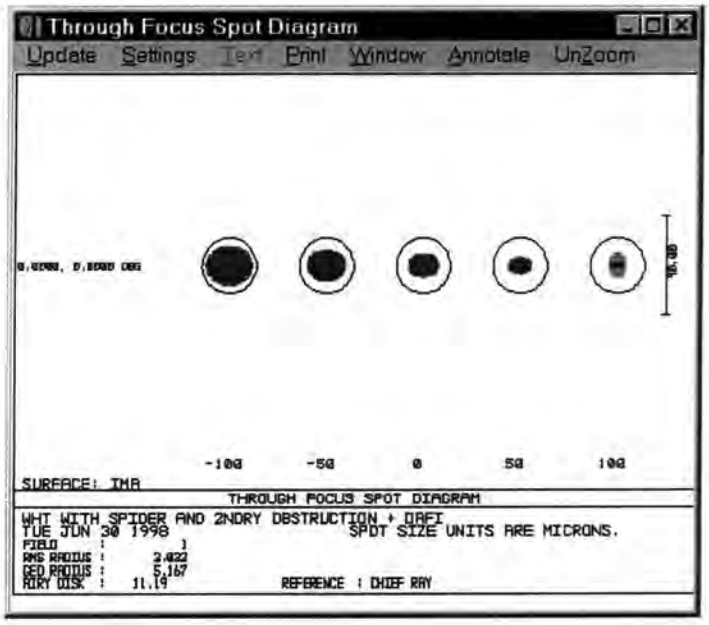



**Figure 6.5.** Through focus polychromatic spot diagrams for the f/15 coupling lens. Again the circles represent the Airy disk.

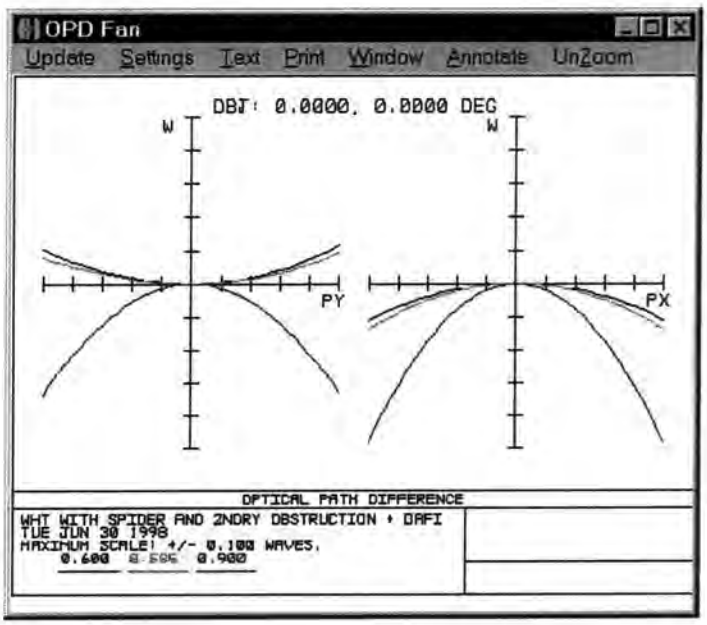



**Figure 6.6.** Optical path differences for each wavelength. The scale labelled $W$ is the number of waves of offset (maximum $\pm 0.1$ waves). The scales marked $PY$ and $PX$ are the pupil diameters across the $x$ and $y$ coordinate axes.

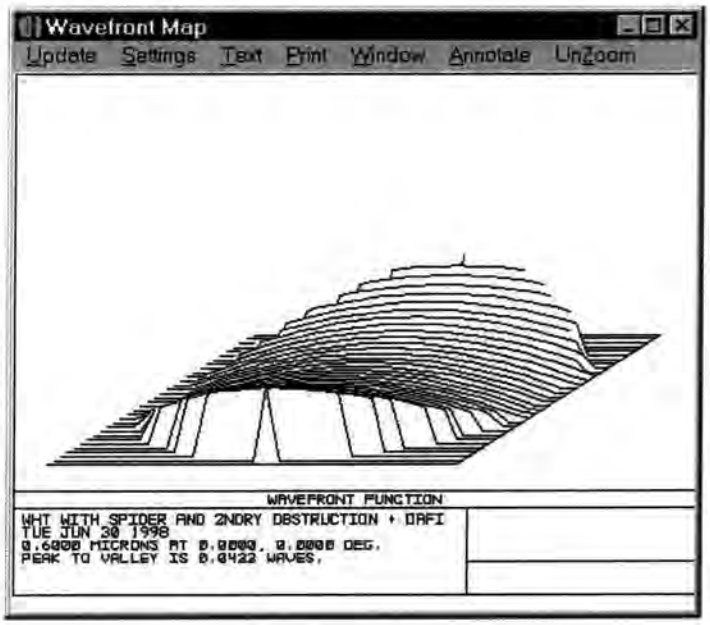



**Figure 6.7.** Wavefront at focus for the f/15 coupling lens ($\lambda$ = 600nm). The residual astigmatism is 0.0422$\lambda$ peak–to–valley.

A six layer broadband ($\lambda$ = 400 to 700nm, mean reflectance $\leq 0.5\%$) anti-reflection coating, was put on the final lenses[3].

### 6.2.3 Components

| | |
|---|---|
| Input radius $a_t$ | 10$\mu$m ($\pm$1.5%) |
| Output radius $a$ | 3$\mu$m ($\pm$1.5%) |
| $\Delta(\lambda = 633\text{nm})$ | $3.1 \times 10^{-3}$ |

**Table 6.2.** Tapered SM fibres' physical parameters. Errors are manufacturer's quoted values.

The physical parameters of the fibres incorporated into DAFI are given in Table 6.2. The subscript $t$ refers to the taper endface. From this data, the fibre parameters were calculated to be:

- output $f_{\#} = 2\sqrt{2\Delta} = 6.35$.
- input $f_{\#t} = Rf_{\#} = 21.17$.

[3]Manufactured by Achro Optics.

- $\lambda_{\text{cutoff}} = 617\text{nm}$.
- $LP_{01}$ mode radius at standard end $r_m = 3.3\mu\text{m}$
- $LP_{01}$ mode radius at taper end $r_{mt} = 11\mu\text{m}$

Unfortunately, the fibres had a different taper $f_{\#t}$ than designed for originally due to an error in specification.

The coupling loss per fibre due to the $f_{\#}$ mismatch, assuming that the image diameter matches the diameter of the fundamental mode field, can be calculated by subtracting the areas[4] of the circular cones subtended. We deduce that 43.75% of the light incident on the coupling lens will be lost. Although this is significant, the current components are still useable. Of course, this figure does not take into account the losses due to the atmosphere. The presence of atmospheric perturbations means that larger apertures are not necessarily better for coupling, since more seeing cells (of diameter $r_0$) are included with increasing aperture size.

In Figures 6.10, 6.11 and 6.12 the design drawings are given for the input plane field plate, lens and fibre mounts. As can be seen in the photographs of the completed unit (Figures 6.8 and 6.9), the five lens and fibre mounts are seated in the field plate. The fibre mounts are intended to hold the fibre taper, which will be fixed permanently in position with UV curing adhesive. The fibre mount then fits into the lens mount as indicated, in such a way that the fibre endface can be adjusted to sit at the coupling lens' focus. The coupling lens itself is seated at the front of the lens mount, where it may be adjusted with the three grub screws to centre the incoming subaperture pupil on the fibre endface. The method of alignment of the input plane will be discussed in Section 6.3.1.

The decision to use brass, despite the fact that it has a less smooth surface finish than steel (so results in relatively high stiction interfaces) is motivated by concerns about rusting over time. The design tolerances are relatively high, especially the concentricity requirement, so stainless steel is not suitable by virtue of its hardness which makes it difficult to machine. The obvious choice of

[4] $\pi r^2$ where $r$ is the radius of the base of the cone.

tool grade aluminium was rejected immediately as it suffers even more strongly from stick–slip friction than brass.

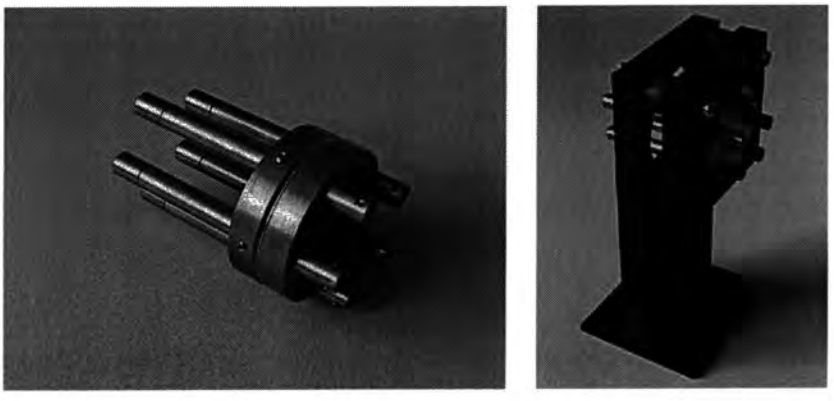


**Figure 6.8.** DAFI input plane. **Left:** The field plate and the five lens and fibre mounts. **Right:** In its tip–tilt holder.

**Figure 6.9.** DAFI lens and fibre mount.

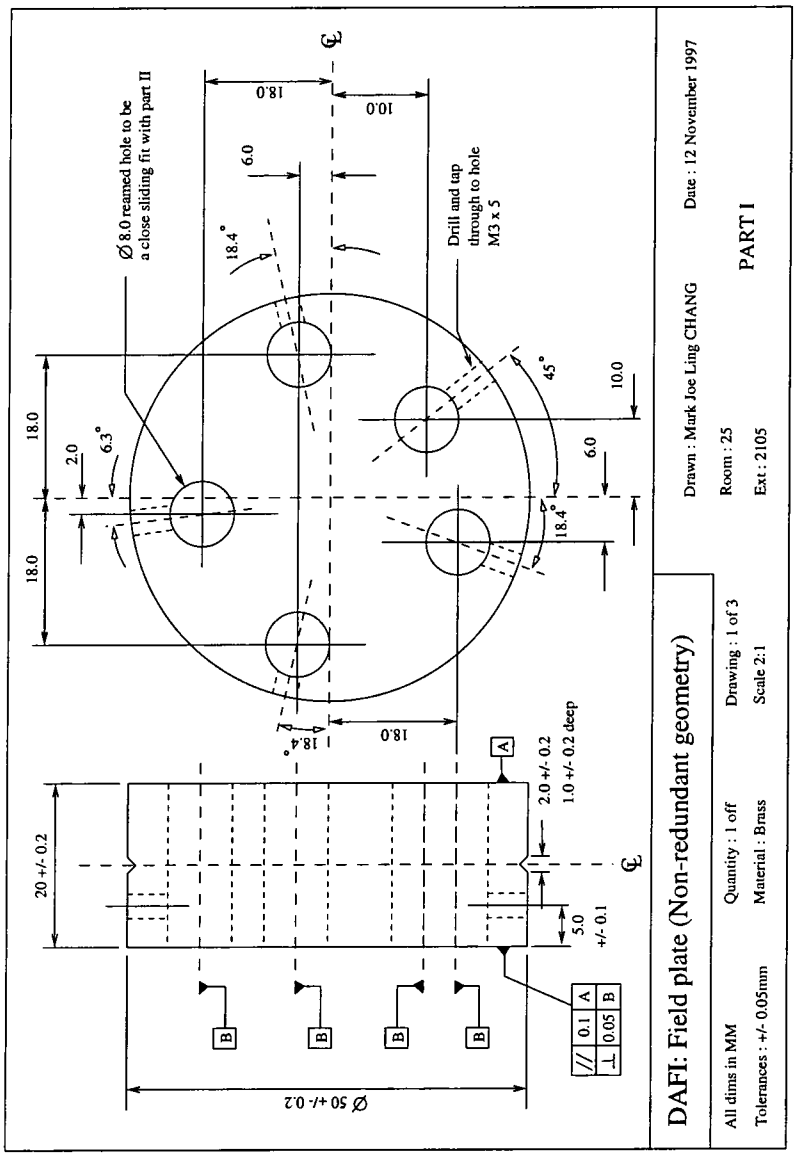



**Figure 6.10.** DAFI input plane drawing: Field plate.

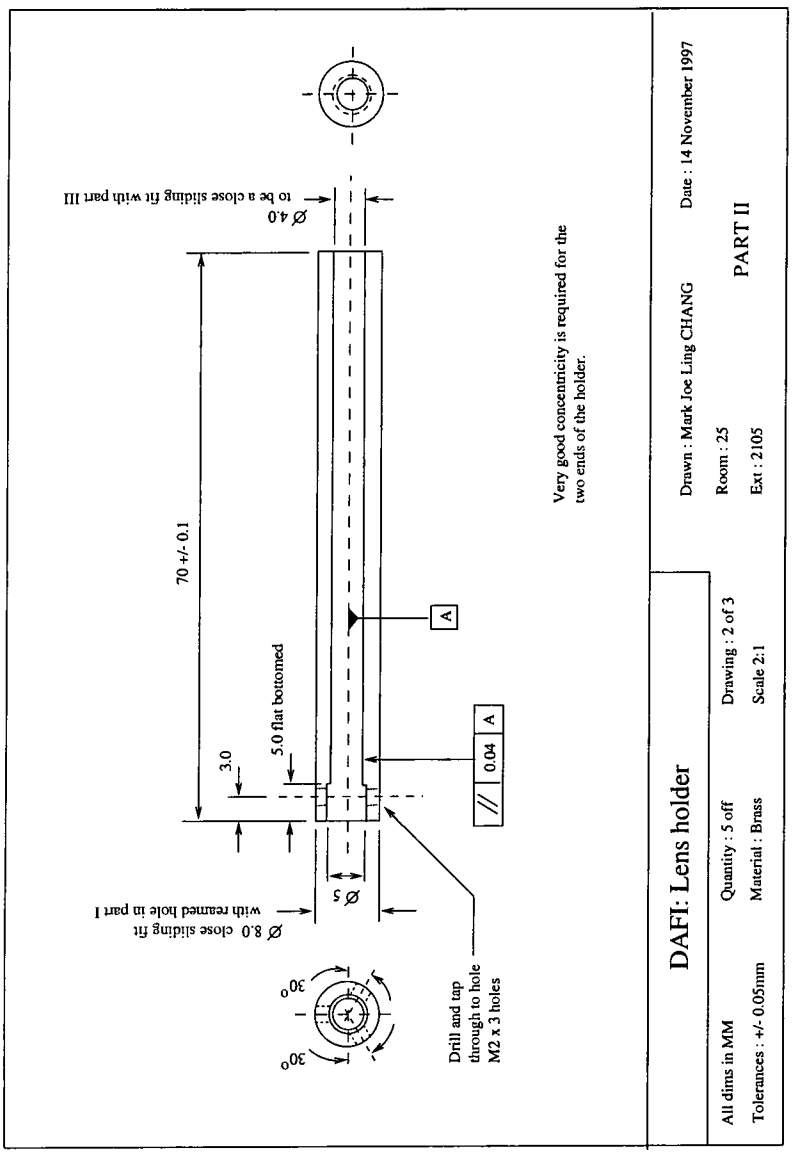



**Figure 6.11.** DAFI input plane drawing: Lens mount.

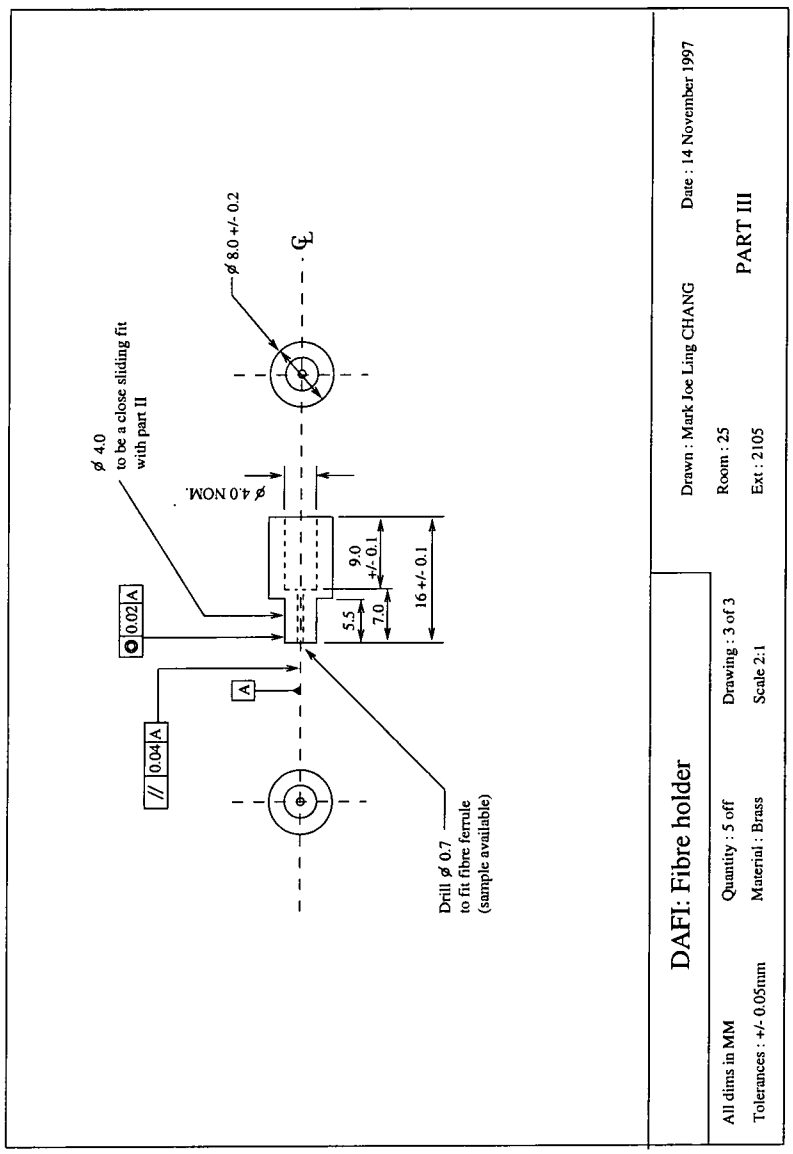



**Figure 6.12.** DAFI input plane drawing: Fibre mount.

| Curvatures (mm) | Thicknesses (mm) | Glass Types |
|---|---|---|
| 45.9760 | 2.0 | BK7 |
| -44.0212 | 4.0 | SF1 |
| -147.9863 | | |

**Table 6.3.** Corrected coupling lens parameters. Again the full lens diameter is 4.4mm with its clear aperture being 4mm.

### 6.2.4 Alternative coupling lens design

To improve upon the collection efficiency of the previous design, an attempt was made to match the $f_{\#}$'s of the coupling lens to the fibre taper. The resulting design, given in Table 6.3, gives a diffraction limited spot at a focal length of 84.68mm with a clear aperture of 4mm diameter. The spot size at focus is slightly oversize, with a radius of 15.98$\mu$m. This is not a concern as there is still an overlap of 88.8% (again modelling the $LP_{01}$ mode as a Gaussian). It is assumed that these lenses will also be coated with a broadband anti–reflection coating.

Because of the slower beam, the field dependent aberrations are less pronounced for this lens than the f/15 design.

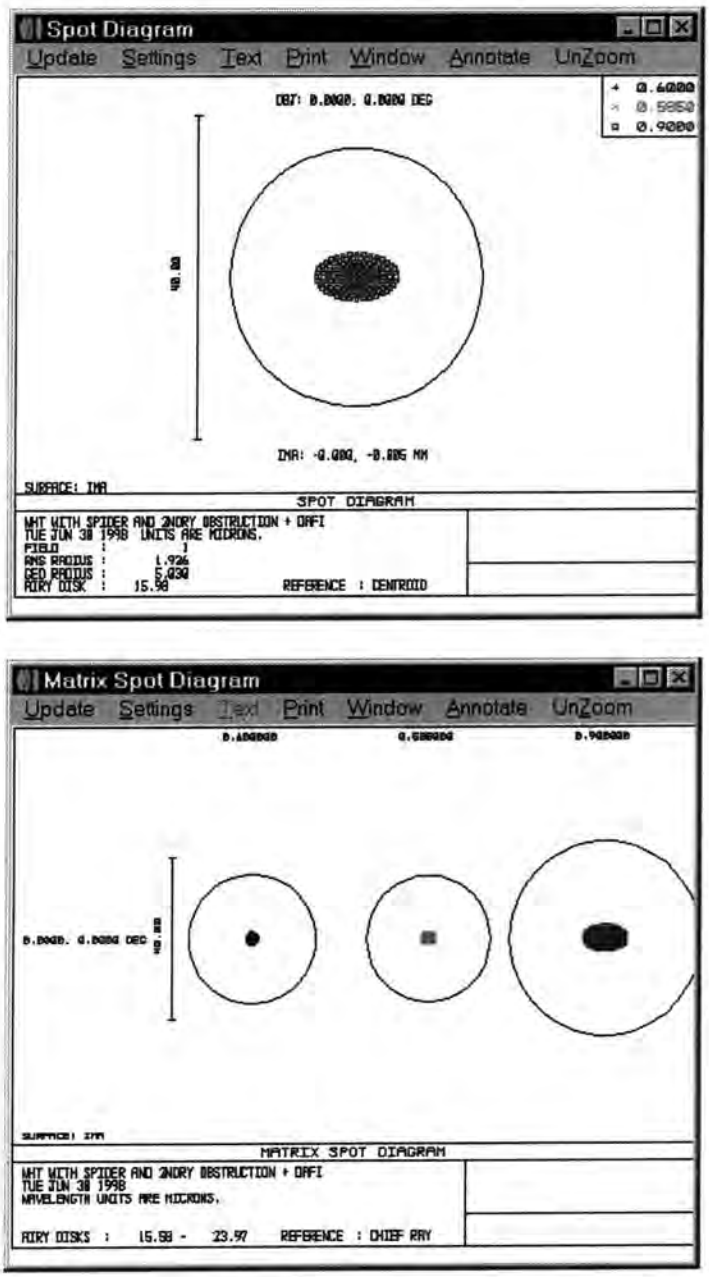



**Figure 6.13. Top:** Polychromatic spot diagram at focus for the f/20 lens. The Airy disk diameter is found to be 15.98$\mu$m. **Bottom:** Matrix of focal plane images at different wavelengths.

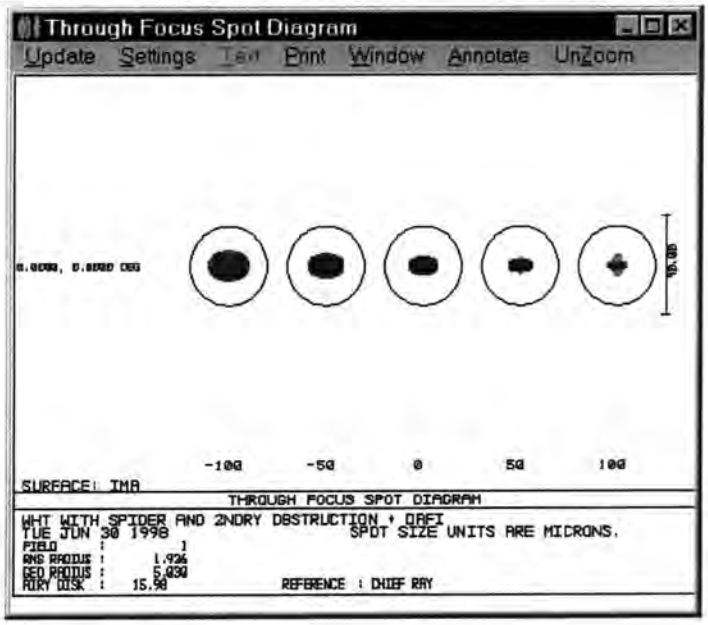



**Figure 6.14.** Through focus ($\pm 100\mu$m) polychromatic spot diagrams for the f/20 lens. Again the circles represent the Airy disk.

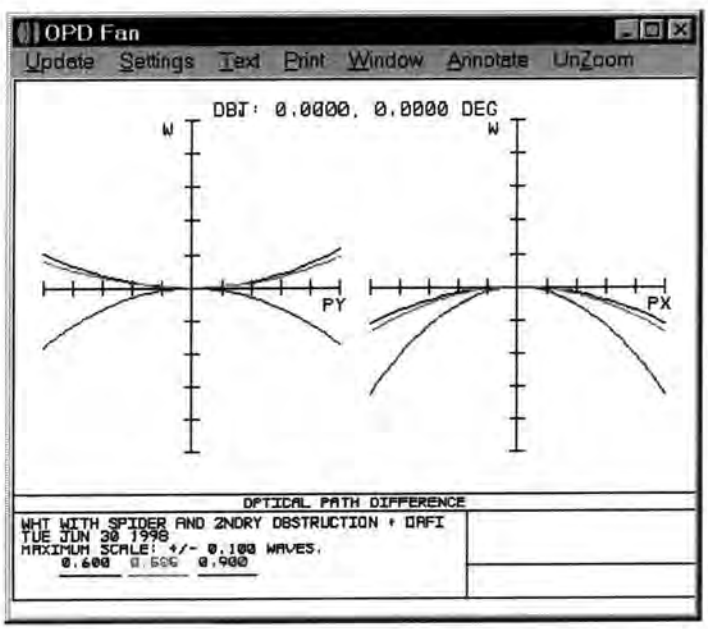



**Figure 6.15.** Optical path differences across the f/20 lens aperture.

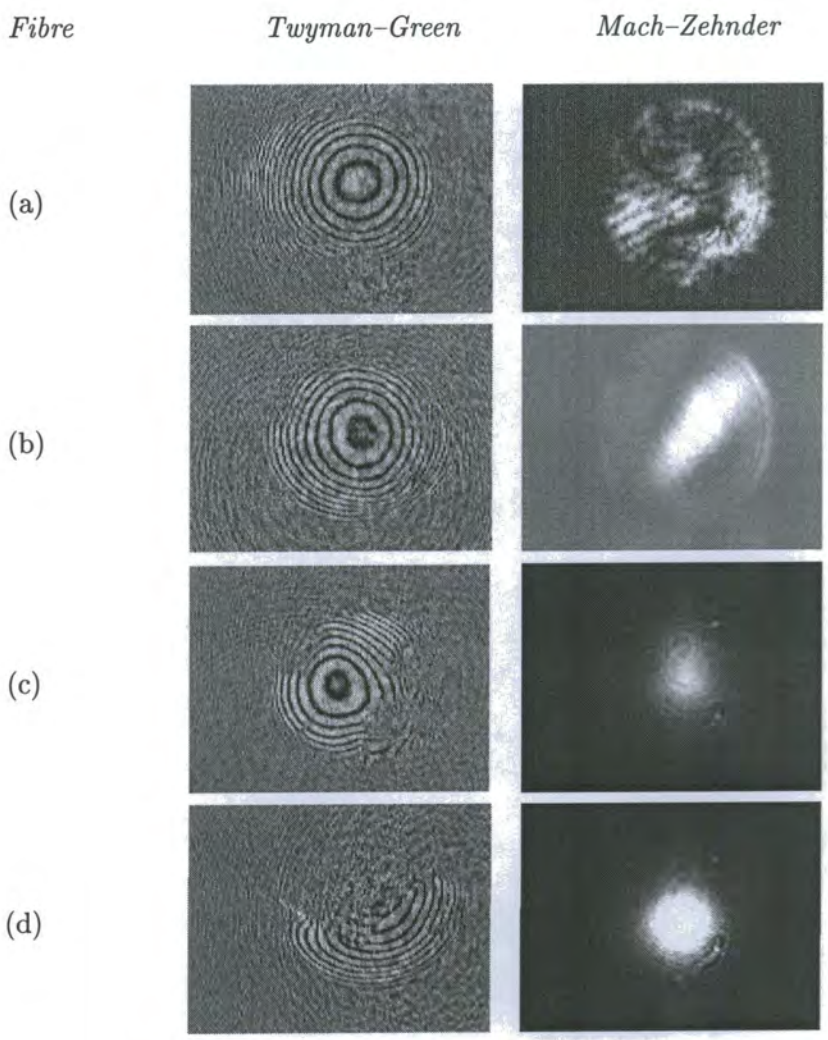
| Fibre pair | Bandwidth (endpoints) [$\mu$m] | $\beta_{\lambda\lambda}$ [$\mu m^{-3}$] | $\Delta\Phi_{\lambda\lambda}$ [$\mu m^{-2}$] | $\Delta L$ [$\mu$m] |
|---|---|---|---|---|
| (g),(a) | 0.402 (0.369 to 0.771) | -220.6 | 116.1 | -0.53 |
| (g),(b) | 0.134 (0.637 to 0.771) | -209.1 | 54.28 | -0.26 |
| (g),(c) | 0.160 (0.479 to 0.638) | -467.2 | 111.45 | -0.24 |
| (g),(f) | 0.174 (0.524 to 0.699) | -698.8 | 83.95 | -0.12 |
| (g),(e) | 0.379 (0.350 to 0.728) | -287.3 | 194.17 | -0.68 |
| (g),(f) | 0.168 (0.586 to 0.754) | 865.9 | 63.33 | 0.07 |

**Table 6.4.** Residual length differences of the tapered fibres after cleaving.

## 6.3 DAFI's fibres

Seven tapered fibres manufactured by Oxford Electronics Ltd were cleaved to the same length. Naturally, the tapered endface of each fibre was used as the datum for cleaving. The residual length differences were measured using the Mach–Zehnder technique. The tapers were clearly used as the inputs. Due to their longer $f_{\#}$'s than the standard fibres used up to now, it was less trivial to align the inputs such that the fundamental mode fields overlapped. It was necessary to illuminate the Mach–Zehnder in reverse with a diode laser so that, at the input end, an Airy diffraction pattern could be seen exiting from each tapered endface. The full diffraction pattern resulted from the cladding modes which were not totally lost over the short fibre length. The fields were then overlapped producing interference fringes. The fibre cores could then be aligned by searching for the central circular fringe and ensuring that this fringe filled the entire field of view. The conventional set–up could then be resumed and the length measurements taken. The results are presented in Table 6.4.

The beam profiles from the fibres were recorded as were the Twyman–Green images of the output endfaces. These are shown in Figure 6.16. Again, some imperfection appears in the profiles as a result of the buffer being removed by heating prior to the fibre being cleaved. Fortunately all the beam profiles are useable for interferometric purposes. The measured $f_{\#}$'s of the standard and tapered ends agreed with the theoretical parameters.


Fibre
Twyman–Green
Mach–Zehnder
(a)
(b)
(c)
(d)

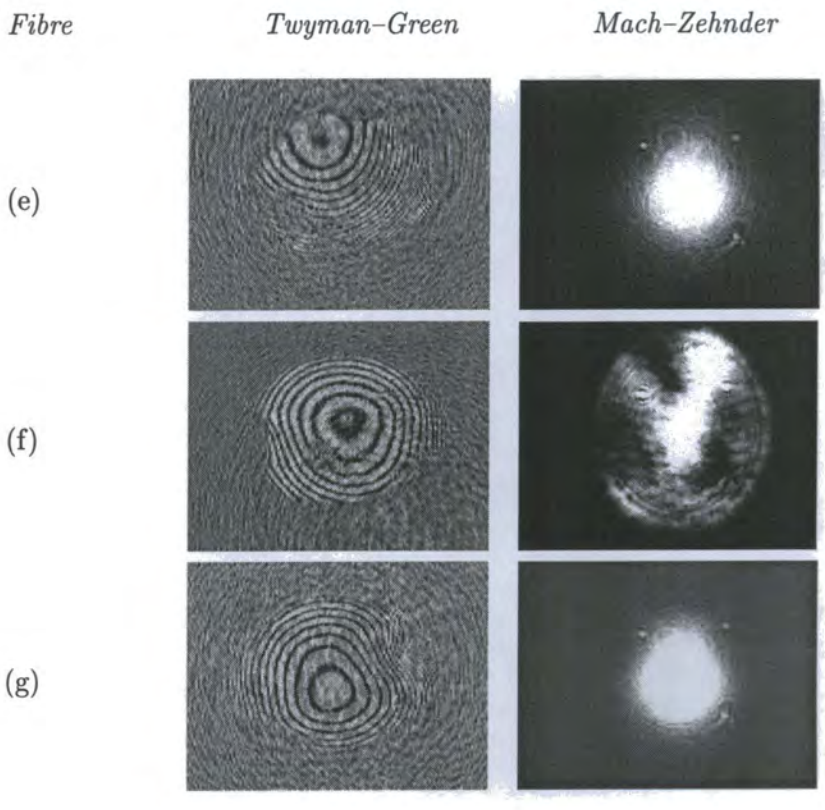



**Figure 6.16.** Ouput endfaces and beam profiles of the tapered fibres.

### 6.3.1 To completion

The obvious next step is to construct the output plane of DAFI by etching and mounting the standard ends of the fibres. The post–etching output profiles and $f_{\#}$'s need to be characterised. This is in order to determine the effect of the core diameter microlenses made by the etching process. It is expected that the etching induced curvatures will not affect the original $f_{\#}$ greatly. The decision to correct the input array by making the new lenses (and also redesigning the fibre & lens holders) will need to be made before continuing.

The output array will then be mounted; it will then be necessary to assemble and align the input plane. Assuming that the coupling lenses are oriented correctly, positioning the taper endface at the focus of the lens should be a straightforward task. The anticipated procedure is to use the classical Twyman–Green, again with a spatially filtered, collimated laser source. Once at focus, the fibre mount can be glued in place with UV–curing adhesive. The alignment of the pupil follows the focusing of the fibre endface. The plan is to illuminate the fibres of the interferometer in reverse. At the 2 dimensional input plane of the interferometer, each subaperture pupil will be visibly defined, provided that the input coupling is high, so that the lenses may be adjusted by means of the grub screws until the pupils are aligned with the DM segments. A similar technique has been described by Haynes [61] for mounting the input and output microlenses of the Spectroscopic Multi–object Infrared Fibre System (SMIRFS).

For the detection of the (straight) fringes, one can conceive of aligning each fringe along a few columns of CCD detector pixels. It would then be feasible to perform on–chip binning, collapsing the two dimensional image to a line where each element of that line contained a sum of charge from its corresponding column. This would increase the SNR per exposure and perhaps enable a faster readout time. On the other hand, the fringes can be collapsed optically (with a cylindrical lens) and read out on a linear array of pixels. The latter method would definitely lead to a faster readout as compared to a two dimensional detector array.

## 6.4 Summary

In this Chapter, the first stages towards making DAFI have been discussed. The original optical design has been shown to be useable, with a 56.25% collection efficiency, despite the $f_{\#}$ mismatch with the supplied tapered fibres. A corrected design was calculated, although the components were not made. The tapered fibres were cleaved and their length differences estimated by the Mach–Zehnder technique described earlier in this work.

Lastly the plan for completion of the interferometer was put forward; DAFI was unfinished due to a combination of a lack of time and the absence of ELECTRA, which was necessary for the input plane assembly task.

# Chapter 7

# Astronomy with interferometers

"Gee Brain, what do you want to do tonight?"

"The same thing we do every night Pinky. Try to take over the world."

*Pinky and The Brain*

## 7.1 Introduction

This Chapter begins with a review of the status and astronomy programmes of a selection of current interferometric instruments as they stand at July 1998. A brief description of each instrument and its published/planned science programmes is given, with emphasis given to the observations (as opposed to the astrophysical models deduced from data). Notably excluded from this list is the celebrated Hanbury–Brown–Twiss intensity interferometer (SUSI) at Sydney University, since it operates on the principle of second order quantum correlations rather than the traditional semi–classical wave theory adhered to by DAFI.

To demonstrate how DAFI can be used, a possible astronomy case is discussed briefly. A comparison in terms of the predicted signal–to–noise ratios is made between DAFI and the method of NRM (with and without tilt correction).

## 7.2 Review of current programmes

### 7.2.1 FLUOR/IOTA

#### The Instrument

At the present time, the only stellar interferometer based on guided optics in full operation is the **F**iber **L**ink **U**nit for **O**ptical **R**ecombination (FLUOR) which is part of the instrumentation of the **I**nfrared–**O**ptical **T**elescope **A**rray (IOTA) facility located at the Smithsonian Institution's Whipple Observatory, Mount Hopkins, Arizona[1].

This Michelson–type instrument, based on non–polarisation preserving, single mode fluoride glass waveguides ($2a = 6.5\mu$m, $NA = 0.23$, $\lambda_c = 1.95\mu$m.) has already been described in Chapter 1.1. We recall that it operates in the near–IR (K–band, 2 - 2.4 $\mu$m) region of the electromagnetic spectrum. It is only able to sample a source object along one baseline, so the closure phase cannot be calculated. It has two 45cm telescopes (each comprised of a siderosat and a 10× beam compressor) that can be located at different stations on the arm of an L–shaped array (15m × 35m).

Polarisation control is achieved through the use of Mickey Ears, which limits the loss of K–band fringe contrast due to polarisation mismatch [8] to less than 20%. The fringe signals are presently detected by InSb photodiode based photometers; these may be replaced by single NICMOS3 infrared array, which is undergoing testing.

#### Astronomy

The first published results of the FLUOR/IOTA combination (G. Perrin *et al.* [62]) extends the effective temperature scale of the faint, cool M giant stars from M6 to M8. Knowledge of the effective temperature gives the radiometric properties of the stellar atmosphere which feeds in as a constraint for stellar atmosphere models. Direct measurements of the effective temperature of a

[1]Website address: http://despa.obspm.fr/fluor/

star may be found from two observables: the apparent bolometric flux and the apparent angular diameter. The bolometric fluxes were determined from wide band photometry and published literature, taking into account interstellar reddening.

They measured the diameters of 9 reference stellar objects in K, ranging between -2.2 to -0.16 in magnitude. The diameters ranged between $5.94 \pm 0.30$ and $16.19 \pm 0.23$mas — these values were compared against the scale of stellar diameters given by Dyck *et al.* [63] giving a dispersion between the two trends of less than 5%.

The visibilities of 8 test sources were measured and the data fitted with a uniform disk model. The results claimed range between $10.08 \pm 0.48$ and $20.20 \pm 0.08$mas. With the bolometric fluxes, the conclusion drawn was that the coolest giant has an effective temperature around 2921 $\pm$110K. This claim is supported by recent work [64, 65] establishing that many carbon stars and long period variables have significantly cooler effective temperatures.

### 7.2.2 COAST

#### The Instrument

The **C**ambridge **O**ptical **A**perture **S**ynthesis **T**elescope (COAST)[2] is a coherent array of four telescopes (with a fifth coming online presently) sited at Lord's Bridge Observatory, west of Cambridge. It operates in the red and near infrared (0.65 to 1$\mu$m and 1.3 to 2.2$\mu$m), using Michelson interferometry on baselines of up to about 100m. 50cm siderostats feed fixed, horizontal 40cm, f/5.5 Cassegrain telescopes. These are set out in a Y configuration, with one telescope on each arm (moveable to a series of fixed positions) and a telescope in the centre of the Y. The fifth telescope will be used as an alternative to the telescope on the same arm, allowing rapid changes to be made between the two [66]. The faintest object observed with COAST is Lalande 21185, a nearby star whose apparent magnitude in the V–band is +7.6.

[2]http://www.mrao.cam.ac.uk/telescopes/coast/

### Astronomy

The objective of the astronomy programme for COAST is to provide very high (milliarcsecond) resolution images for a wide variety of stellar systems down to a red magnitude of 10. This will allow the study of stellar surfaces, the envelopes of pre-main sequence stars, pulsating variables, circumstellar shells, compact planetary nebulae and close binary and multiple systems.

The first images by aperture synthesis in the optical (830nm) were made with three elements of COAST in September 1995 [2]. The object was the binary star Capella ($\alpha$ Aurigae), which was observed with milliarcsecond resolution[3]. The individual components themselves were unresolved however.

More recently, results of studies of the surface structure and limb darkening profile of Betelgeuse ($\alpha$ Orionis, the brightest M supergiant with the largest apparent angular diameter of any star apart from the Sun). This is the first resolved image of a star from an optical aperture synthesis array [67]. At the time of the observations, no brightness asymmetries ("hot spots") in the stellar flux were noted, which is an unusual result when compared to previous work on the object (see Section 7.3). As a consequence of this source symmetry, it was possible to show that the stellar disk intensity distribution is flat except near the edge, where it falls off in a Gaussian manner. This means that the star is larger than previously measured, a fact that will allow a more refined estimate of the effective temperature to be made.

COAST has also studied the long–period (Mira) variable star R Leo [68] to obtain unbiased estimates of the angular extent of its emission. This type of object has characteristic cyclic variations in V–band intensity of up to 11 magnitudes. The periods of these $\sim 1$ $M_{\odot}$ stars range between 100 to 500 days. They are also rather large objects, having apparent angular diameters in the order of tens of milliarcseconds. Since theoretical models of Mira variables are as yet unsatisfactory in predicting observed properties, studies of this kind provide important constraints on the allowed physics. R Leo was observed over a period

[3]The restoring beam shape was an ellipse of 30 by 18mas.

of two years, spanning two pulsation cycles of the star, at 4 wavelengths.

### 7.2.3 NPOI

#### The Instrument

The **N**avy **P**rototype **O**ptical **I**nterferometer (NPOI), sited at the Lowell Observatory near Flagstaff, Arizona[4] is a decendent of the Mark I through to Mark III series of phase–tracking interferometers built by Shao *et al.* [69, 70, 71] on Mount Wilson. It has a synthetic aperture of six 50cm siderostats feeding 12cm diameter subapertures with baselines from 19m to 38m. The interferometer observes in 32 spectral channels covering 450-850nm so that each baseline produces 32 fringe visibility measurements at spatial frequencies ranging over a factor of 1.9. The anticipated V–band apparent magnitude limit for the instrument is $m_V \approx +7$.

#### Astronomy

Recent results from the NPOI have been related to observations of stellar limb darkening [72] on two K giant stars stars ($\alpha$ Aries and $\alpha$ Cassiopeiae, where $m_V = 2.0$ for the former and $m_V = 2.2$ for the latter) using the technique of "phase bootstrapping" [73]. Hummel [74] has also reported microarcsecond resolution observations of spectroscopic binaries — specifically of the double star Mizar A ($\zeta^1$ Ursae Majoris), Matar ($\eta$ Pegasi) and Alnitak A ($\zeta$ Orionis). In addition to the spectroscopic binaries, speckle binaries (with separations between 65 and 95mas) have been reconstructed from NPOI data.

### 7.2.4 GI2T

#### The Instrument

The **G**rand **I**nterféromètre à **2** **T**élescopes (GI2T) is a Michelson interferometer consisting of two Cassegrain–Coudé telescopes, moveable on north–south

[4] http://aries.usno.navy.mil/ad_home/npoi/

tracks[5]. The telescopes have a unique "boule" design [75] with diameters of 1.52m. The 13m baseline interferometer is sited on the Plateau de Calern, near Grasse in the south of France.

#### Astronomy

The research programme of the GI2T, up to now, has mainly been concerned with the circumstellar environment of hot active stars (Be stars) [76, 77, 78, 79, 80, 81]. These Be (non supergiant) stars are of spectral type B and are found on the main sequence, or between 0.5 and 1 magnitudes just beneath it. These characteristics exclude the principal pre–sequence stars: T–Tauri and Herbig Ae/Be. The P–Cygni type objects are also not members of this category since they are Class I supergiants. Much of the research has been focused on $\gamma$ Cassiopeiae ($m_V = 2.25$), one of the best studied Be stars in the northern hemisphere.

The GI2T has also been employed to determine the angular diameter variations of $\delta$ Cephei ($m_V = 3.48 - 4.34$) [82]. This resulted in a semi–empirical result for the distance to this object which, of course, is an important step in the Cosmological Distance Ladder.

### 7.2.5 PTI

#### The Instrument

The **P**alomar **T**estbed **I**nterferometer is a near–IR (K–band) stellar interferometer located near the 5m Hale Telescope at Palomar Observatory, California, USA[6]. It has three fixed 40–cm apertures which can be combined pairwise to provide baselines to 110 m; its active delay lines have a range of ±38m [83].

As with the NPOI, PTI is a development based on the (visible region) Mark III interferometer which used active fringe tracking for wide–angle astrometry and parametric imaging.

[5] http://wwwrc.obs-azur.fr/fresnel/gi2t/gi2t.htm
[6] http://huey.jpl.nasa.gov/palomar/

### Astronomy

The remit of the PTI, part of NASA's Origins search programme, is to search for exoplanets (planets orbiting neaby stars). The astrometric signature of a Sun–Jupiter system at 10pc is easy to deduce from the following argument. The mass ratio of the Sun to Jupiter is about 1000:1, so the planet's motion is about 1000× that of the stellar motion. From the definition of a parsec, a Jupiter sized planet at 5AU from a Solar type star would move 1″ peak–to–peak. Therefore the star's signature would be 1mas peak–to–peak. It is therefore necessary for the PTI to have at least 0.1mas resolution in order to identify such astrometric signatures unambiguously.

Boden *et al.* [84] have recently studied the star 51 Pegasi (HD 217014) in more detail, since the inference of a planetary–mass gravitational companion to the star was made by Mayor and Queloz [85] spectroscopically. The apparent radial velocity only sets the lower mass limit, so the possibility that the companion was a low–mass star still remained. The results of the study showed no evidence of a luminous companion, allowing an upper limit on the difference in magnitude (in K) of the companion with respect to 51 Peg. This upper limit $\Delta K = 4.27$ is claimed with a confidence level of 99%, implying that the companion object is not brighter that $m_K = 7.30$.

The angular diameters of 80 giant and supergiant stars (between G6 and M6, with one B7 for the giants; between F5 and M5 for the supergiants) were also determined [86] and compared to the IOTA interferometer datasets, with no statistically significant differences being found.

## 7.2.6 ISI

### The Instrument

The **I**nfrared **S**patial **I**nterferometer (ISI) is a two-telescope Michelson interferometer, operated by the Space Sciences Laboratory of the University of California at Berkeley[7]. It observes in the mid–infrared at a wavelength of 11

[7] http://ssl.berkeley.edu/isi_www/isi_core.html

microns.

The subapertures each consist of a 65inch (1.65m) parabolic mirror and an 80inch (2m) flat mirror. The parabolic mirror is fixed while the flat mirror acts as a siderostat, making this a Pfund–type telescope system. The baselines extend from 4 to 35m giving a maximum angular resolution of 30mas.

The construction of a third telescope has recently been planned, with site expansions allowing 75m (east–west) and 65m (north–south) baselines. This will increase the maximum angular resolution to 15mas.

#### Astronomy

The programme at ISI has concentrated on the study of evolved stars. These stars are further along in their life cycle compared with the Sun and typically return some fraction of their mass to the circumstellar environment. There, by cooling off, this material condenses into dust grains that absorb visible star light and re–radiate this energy in the infrared region of the spectrum. A survey of 13 stars was made [87] determining the distance the dust lies from the stellar photosphere.

Astrometric measurements of Betelgeuse and Antares ($\alpha$ Scorpii) were made — the first determinations of stellar diameters in the mid–IR. At these wavelengths it was found that limb darkening is a negligible effect. On–going dust formation was seen around Antares, to about $1''$ around the star.

### 7.2.7 Summary

This brief overview of the various programmes undertaken by a selection of the large interferometers operating at present highlights the amount of effort invested into observations in K–band and longward. This bias indicates the potential of high angular resolution imaging at shorter wavelengths such as V–band.

Another major difference between these instruments and DAFI is the size of the baselines employed. The objective of DAFI is to reach the diffraction limit of a single large telescope, so is constrained in terms of its ultimate angular

resolution by the size of the telescope primary. The systems reviewed in this Chapter are all separate telescope systems, with baselines generally extending over 10m. So part of the attraction of DAFI is its lack of complexity as compared to the other multiple baseline interferometric arrays.

## 7.3 Astronomy case

At this point we examine a specific science proposal in detail, to demonstrate the potential of DAFI.

The work of Wilson *et al.* [88] and Buscher *et al.* [89] on the M supergiant star $\alpha$ Ori (Betelgeuse) revealed bright, unambiguous, small–scale asymmetric features in the intensity profile through the method of Non–Redundant Masking (mentioned in Chapter 1.1) . These features appear as small unresolved spots on top of the stellar disc contributing as much as 20% of the total emission from the star.

Using the Faint Object Camera (FOS) of the Hubble Space Telescope (HST), Gilliland & Dupree [90] have also detected a bright unresolved feature offset from the centre of Betelgeuse in two UV continuum passbands (2550/236 Å and 2800/316 Å). Evidence for as many as three separate features occur in the data; the changes in position and brightness occur on short time–scales (about one month) and no correlation between images taken several months apart has yet been inferred.

Despite the frequency with which these features have been reported, their nature and physical properties have remained unclear. The early observations were exclusively focused on $\alpha$ Ori; recent work by Tuthill *et al.* [91] has provided examples of these "hotspots" on two other nearby M supergiants: $\alpha$ Sco (Antares) and $\alpha$ Her (Rasalgethi).

As yet, the least understood property of the surface hotspots seen on M supergiants is their lifetime. The first observations by Buscher and collegues placed an upper limit of 23 months on the evolutionary timescale of these features. Wilson later reported a reduced upper limit of 9 months based on observations of $\alpha$ Ori in January and September of 1991. Tuthill obtained data at

700nm in January, September and December of 1993 which traced the development of a single feature at three epochs over 12 months. Observations at 710nm (TiO molecular band) exhibited identical behaviour confirming the effects were source–based. These results indicated that the characteristic timescales of the optical phenomenæ is on the order of months.

### 7.3.1 Stellar diameters

The deviations from circular symmetry seem to be a common, time–varying feature of these stars. This raises a number of issues regarding attempts to measure their apparent sizes. It is known that interferometric angular diameter measurements of late–type giants and supergiants show a larger scatter [92, 93]. This has been attributed to the use of inadequate models for the centre–to–limb brightness distributions and the failure to treat correctly the strong wavelength–dependant opacity in the cool stellar atmosphere. A further problem has now been encountered with the evidence for localised features on the stellar surfaces. If such features which strongly contribute to the total flux from the source can appear, evolving in brightness and location, it is inevitable that they will introduce a source of systematic error.

Of course, reliable measurements of stellar diameters *can* be made if no features are detected during an observing epoch, as with the work by Burns *et al.* [67].

### 7.3.2 Physical models

The (pre–NRM and COAST) observed variations in the photometric, radial–velocity and polarimetric signatures of M supergiants have motivated a number of attempts at explaining these effects. These range from the presence of companions, stellar rotation effects, magnetic activity[8], radial/non–radial pulsation and transient convective cells on the stellar surface. The first of these models is not supported by the observational evidence presented separately by Buscher, Wilson or Tuthill; the multiple features evolved irregularly on short

[8]The Swiss Army knife of astrophysics!

timescales which is not consistent with the more regular and longer–period behaviour expected for systems with stellar companions. The irregularities do not support the idea that these phenomenæ have their origins in the object's rotation. Smith *et al.* [94] derived a characteristic rotational period for Betelgeuse of $\sim$ 1200 years from angular momentum arguments. So, regardless of differences in mass, composition, spectral type and the details of angular momentum transport, these arguments lead to rotational timescales which are orders of magnitude larger than the timescales observed. Similarly, although radial pulsation appears to be a common property of late–type supergiants [94, 95], it can be ruled out as the cause of the hotspots because of the timescales involved (about a year).

Evidence for strong magnetic activity through X–ray fluxes is not convincing. Total X–ray flux limits have been reported by Vaiana *et al.* and Maggio *et al.* [96, 97], which appear to be comparable to late–type dwarfs. However, the surface fluxes of M supergiants must be very low, so the possibility that locally enhanced magnetic fields are giving rise to starspots is not very high.

The most plausible explanation for these hotspots seems to be convection. The features may be optical manifestations of large convective cells in the stellar photosphere, along the lines of red supergiant supergranulation.

### 7.3.3 Observing efficiency

To gain a better understanding for the underlying processes driving these hotspots, it would necessary to undertake a series of observations to track the lifetime of one or more features. It may even be possible to relate the surface activity with mass loss if the mechanism of convective cells could be established.

We can calculate the photon rate in V ($\lambda = 0.55\mu$m, $\Delta\lambda/\lambda = 0.16$) at each subaperture of DAFI from the sources studied by Tuthill.

The incident photon flux determination proceeds as follows: We have the definition of astronomical magnitudes $m_i$

$$m_2 - m_1 = -2.5\log(f_2/f_1) \tag{7.1}$$

| Band | $\lambda_c$ [$\mu$m] | $\Delta\lambda/\lambda$ | Zero point [Jy] |
|---|---|---|---|
| U | 0.36 | 0.15 | 1810 |
| B | 0.44 | 0.22 | 4260 |
| V | 0.55 | 0.16 | 3640 |
| R | 0.64 | 0.23 | 3080 |
| I | 0.79 | 0.19 | 2550 |
| J | 1.26 | 0.16 | 1603 |
| H | 1.60 | 0.23 | 1075 |
| K | 2.22 | 0.23 | 667 |

**Table 7.1.** Some photometric bands, taken from Bessel [98] and Campins *et al.* [99].

where $f_i$ is the flux outside the Earth's atmosphere in Janskys[9].

The photometric datum point for the V–band is given in Table 7.1, so if we know the object's magnitude $m_2$ we immediately have its flux $f_2$.

Of course, the objects are intrinsically variable so all the values are only correct in the order of magnitude[10].

ELECTRA has 5 surfaces and a 40/40 beamsplitter, while DAFI has 1 surface (the off–axis spherical), the fibres with their coupling lenses upstream of the detector. All these components must be accounted for when predicting DAFI's SNR — the analyser being the only component that may not be necessary (discussed later). We assume (pessimistically) that the 6 surfaces have a reflectivity of 98%, giving a total reflectivity of about 88.6%. The reflectivity of the telescope primary, secondary and Nasmyth flat are not included. Reflection losses at the anti–reflection coated lenses are assumed negligible, which is a reasonable approximation. We will consider two situations: one where the lenses provide only a 56.25% overlap of (diffraction limited) image with the $LP_{01}$ mode, and one where there is an 88.9% overlap (see Chapter 6). Provided the low order phase perturbations obey Kolmogorov statistics, we can say that a maximum of 18% of the fluence is coupled into a fibre [9].

So there are two optical systems — (a) one with non–ideal coupling lenses,

[9] $1\mu$Jy = 15.1 photons $s^{-1}$ $m^{-2}$ $neper^{-1}$, where 1 neper $\triangleq d\lambda/\lambda$.

[10] The values used were taken from the Bright Star Catalog, a limited online version being available at http://www.sternwarte.uni-erlangen.de/www/bsc.html.

having a throughput of 3.6%, and (b) the other with the redesigned lenses, having a throughput of 5.7%. From the experimental result of Chapter 5 it seems reasonable to assume visibilities of about 68% with a fibre interferometer. This is, of course, a pessimistic value to use which will give the worst case scenario.

It is noted that no analyser has been included in the system, since it is assumed that the Mickey ears are achromatic over the whole observing bandwidth.

On a 4m telescope, such as the WHT, each of DAFI's subapertures presents a circular entrance pupil of 40cm ($\sim 4r_0$) diameter. We assume an atmospheric coherence time $\tau_0 = 10$ms in the V–band. The SNR of the interferometric signal from one baseline is calculated from a simplified version of Dainty and Greenaway's expression [100]

$$\mathrm{SNR} = \frac{(2S)V^2\sqrt{N_{\mathrm{exp}}}}{\sqrt{1+(2S)V^2}} \tag{7.2}$$

where $S$ is the number of photons from one subaperture, $V$ is the fringe visibility (ranging between 0 and 0.5), $t$ is a single exposure time, $N_{\mathrm{exp}} = T/t$ is the number of incoherently coadded exposures for a total exposure time $T$. The atmospheric noise term and the double frequency noise term of the full expression have been neglected. The former is reasonable if the exposure time is short and the tilts are well compensated by ELECTRA. The double frequency term can be neglected for mid to high spatial frequencies as it lies outside the beam combiner optical transfer function. We will use an optimum $t$ of $2\tau_0$ (cf. Chapter 4) and vary the number of coadded exposures $N_{\mathrm{exp}}$ for interferometric imaging over the full V–band.

It is not simple to include the noise introduced by a detector into the formalism. Therefore no detector model has been used in this brief analysis. This means that the absolute values of the signal–to–noise are not immediately useful. However, for a relative figure–of–merit they are still valid. Therefore we can use the SNR values to compare both optical coupling systems (a) and (b)

to assess the impact of the imperfect design.

The tapered fibres of DAFI have a single-mode cutoff wavelength of $\lambda = 633$nm, so that the standard photometric bands cannot be applied to describe the performance of the interferometer exactly. However, they do provide a reasonable guide to the difference in SNR of the two optical systems that can be expected, given a source magnitude.

**$\alpha$ Ori:** Betelgeuse (HR 2061) has a listed $m_V = 0.5$. The incident flux is therefore 697282 photons subaperture$^{-1}$ $\tau_0^{-1}$.

DAFI's signal for each of the optical systems described previously is then:

| Case | Counts subaperture$^{-1}$ $\tau_0^{-1}$ | $T$ [sec] | SNR |
|---|---|---|---|
| (a) | 25016.1 | 60 | 13171.2 |
| (b) | 39536.6 | 60 | 16558.9 |

**$\alpha$ Sco:** Antares (HR 6134) has a listed $m_V = 0.96$, leading to a flux of 456466 photons subaperture$^{-1}$ $\tau_0^{-1}$. The detected flux and SNRs become:

| Case | Counts subaperture$^{-1}$ $\tau_0^{-1}$ | $T$ [sec] | SNR |
|---|---|---|---|
| (a) | 16376.4 | 60 | 10656.3 |
| (b) | 25882.1 | 60 | 13397.3 |

**$\alpha$ Her:** Rasalgethi (HR 6406) has a listed $m_V = 3.48$ so we arrive at 44813.4 photons subaperture$^{-1}$ $\tau_0^{-1}$. The detected flux and SNR values are:

| Case | Counts subaperture$^{-1}$ $\tau_0^{-1}$ | $T$ [sec] | SNR |
|---|---|---|---|
| (a) | 1607.8 | 60 | 3334.9 |
| (b) | 2541.0 | 60 | 4194.5 |

From these V-band photon rate calculations we can see that the design mismatch of the coupling optics in case (a) does not reduce the final SNR by more than 20%.

### 7.3.4 Relative magnitude limits

A comparison between the magnitude limits estimated using this SNR argument between DAFI and the closest competitive technique, NRM, is undertaken here. The SNR values are found subject to the constraint that a maximum of 8 hours of observing time is available in one night. Again a pessimistic value for DAFI's fringe visibility of 68% is assumed.

The model for standard NRM (without tilt correction) assumes a collimating mirror of 98% reflectivity, a pupil mask with $1.2r_0$ diameter holes, a 50% transmissive analyser and a (lossless) refocusing lens in front of the detector. Buscher's calculation of the r.m.s. visibility as a function of aperture diameter (Chapter 3 of Reference [47]) indicated that apertures of $1.2r_0$ in diameter would produce fringes with a contrast of about 50%.

Tilt corrected NRM (NRM + AO) is modelled by including ELECTRA, a 98% reflective collimating mirror, a pupil mask with $2.8r_0$ holes and, as before, a 50% transmissive analyser in front of a lossless lens before the detector. Again the fringe visibility has been calculated to be 50%. It should be noted that these NRM models are highly idealised and will give quite optimistic results.

The Fried parameter $r_0$ is taken to be 10cm. If we assume that a useable value for the total SNR is around 10, we have:

| Case | $T$ [hours] | $m_V$ | SNR |
|---|---|---|---|
| (a) | 8 | 16 | 13.9 |
| (b) | 8 | 16.5 | 13.8 |
| Standard NRM | 8 | 18.5 | 15.3 |
| NRM + AO | 8 | 18.5 | 12.9 |

So, this (detectorless) model shows that DAFI's maximum imaging depth in the V-band is constrained by the available flux to be some 2 magnitudes less deep than the limit available to the idealised NRM model. Note that the NRM + AO model is less efficient than standard NRM. This is due to the number of extra (lossy) surfaces introduced by ELECTRA. However, although the system

throughputs are greater for NRM than for the DAFI, the actual number of photons contributing to the fringe signal is is lower. So DAFI uses the photons it receives more efficiently. This last point is more important than the relative imaging depths since it implies that the visibility is more easily calibrated in comparison to NRM.

### 7.3.5 Redundant geometry at input

The spatial multiplexing capability of a fibre system such as DAFI has already been discussed in earlier Chapters. One of the main advantages of using redundant baselines is the fact that it is easier to employ the technique of phase bootstrapping. Briefly, this means that the shorter baselines are used to fringe track, allowing observations to be made by the longer baselines. If the baseline redundancy is applied to these longer baselines, improving the SNR, interferometric imaging at higher spatial frequencies will be made easier. This is discussed in more detail in Reference [73].

Another way to utilise the multiplexing advantage is to add redundant baselines at the input plane while still keeping the non–redundancy of the output plane. This cannot be easily be done with NRM — more baselines sampling the same spatial frequency at the input means a higher SNR for that spatial frequency can be found. Of course, the complexity of the instrument increases, since the Fourier plane coverage is still essential. However, this is not so trivial since there are issues of calibrating the random phase offsets that exist between baselines. At low SNR this is rather difficult to do — on the other hand if the SNR is high enough to calibrate the phase errors, there is no reason to use the redundancy at the input plane to increase the SNR.

Finally, the most important issue to note when comparing DAFI to NRM is the visibility calibration advantage of SM fibres interferograms. The effect of spatially filtering the subapertures means that the visibilities will be easier to calibrate than the equivalent, unfiltered NRM fringe signals. This means that images of a higher dynamic range can be reconstructed.

## 7.4 Conclusion

We have briefly reviewed the status and current programmes of some of the large interferometers in operation at present. The conclusion drawn from this survey is that interferometric imaging in the visible has not been a focus of large, multi–telescope interferometers up to now, due to the paucity of published astronomical results. It is likely, therefore, to yield a substantial scientific return even with a test–bed interferometric system. A specific science programme was devised to provide a quantitative idea of the parameters involved.

In principle, it is now quite possible to attain the diffraction limit of any single large telescope in the visible regime through a number of techniques. A comparative study between the competing methodologies of NRM, tilt corrected NRM and DAFI was conducted. DAFI was found not be as sensitive as the NRM methods, a not unexpected result given the extra optical components. The models used for comparing the systems did not include the detector. Although this is clearly a major component to leave out, there is no well known formalism for including the effects of detector noise at low light levels on the fringe visibility. Therefore no attempt was made at predicting the object magnitude limit of the instrument.

By employing redundant input baselines, DAFI could be made more sensitive — although because of the complexity of this technique is unlikely to be used for this reason. The use of redundant baselines for phase bootstrapping should be a major area of investigation in the future however.

It must be emphasised that the calibration of the visibilities will be simplified by using SM waveguides as compared to NRM techniques, although the improvement has not been quantified here.

Finally, it should also be quite straightforward to extend DAFI so that it produces channeled spectra, i.e. multiple–wavelength imaging of the fringes, which would greatly increase its versatility.

# Chapter 8

# Summary ... and the future?

The best thing about the future is that it comes only one day at a time.

*Abraham Lincoln*

## 8.1 Then ...

In this final Chapter all that has been done in laying the groundwork towards the realisation of the Dilute Aperture Fibre Interferometer is summarized. Since DAFI was always intended to be a bolt–on instrument for ELECTRA, it seemed natural to start with a description of the AO system, albeit briefly. The hysteresis of the segments of the DM was characterised and reported in the first Chapter. The results showed that the strain gauge method for determining the pistons of the actuator electrodes is extremely linear. Initial tests with tapered fibres motivated a study into the design parameters required to produce such components. It was then necessary to enter into the theoretical background concerning single mode fibres in order to understand the variables of importance for a fibre–based interferometer. These turned out to be

1. Input coupling
2. Chromatic dispersion
3. Polarisation
4. Output array positioning

Of these, the first is the principal limitation — after all, without any light, what use is the rest? An attempt was made to find the design parameters for single mode tapered fibres which also had a large degree of intrinsic birefringence.

We calculated the allowable tolerances on the fibre arm differences analytically. A method was developed for cleaving short lengths of fibres to an extremely high relative accuracy, and for measuring these length differences.

Control of the polarisation states in a circular core, single mode fibre was found to be possible through bending birefringence[1] and the ill–treatment of a well known children's character.

The concept of constructing a useable output plane by etching the cladding of 125$\mu$m fibres down to around 50$\mu$m was proved viable in Chapter 5 and a (rather fragile) multi–fibre interferometer was constructed.

The next Chapter were concerned with designing the important input plane for DAFI and its associated coupling optics. Unique to this fibre interferometer are its adiabatically tapered inputs, which should improve the instrumental throughput with well–designed coupling lenses.

The current science done with modern large interferometers was reviewed, showing the potential for visible wavelength diffraction limited imaging, even on the angular resolution scales of a single large telescope. A possible programme was put forward, which would provide more information on the structure of M supergiants.

Finally, a comparison (in terms of SNR) between DAFI and the closest rival technique of NRM was made. Being a fibre based system, DAFI has the ability to introduce redundant baselines at its input quite easily without compromising its non–redundant output plane. This is not so simple to do with NRM. It will be interesting to see if the introduction of such redundancy will make it easier to make interferometric observations at the highest available spatial frequencies through phase bootstrapping.

[1] It should be noted that one can achieve the same result with the Kerr effect, which can introduce a linear delay if the applied electric field is sufficiently strong. Optical activity can arise through the Faraday effect too, as an alternative to twisting the fibres.

## 8.2 ...and now?

The work recorded in this thesis is but a small part of a very much larger effort — to see every possible detail of the Universe. With the completion of DAFI, another technique will be available to produce images of extraterrestrial objects at the theoretical resolution limit of single large telescopes. It may well prove to be more flexible than pupil masking methods and is certainly worth considering in its own right.

Presently, there is a limitation in the sky coverage possible with natural guide star AO. This, unfortunately, feeds down to DAFI. One solution is available in the form of laser guide stars[2] however, especially sodium beacons. These guide stars can be placed in front of sources of interest for the low–order AO system to use, while the interferometric source observations can be undertaken at wavelengths away from the beacon itself. Another way is to look toward the development of fast detectors with sub–electron read noise for wavefront sensing. With improvements in amplifiers for CCDs, this should be realised come the new millenium.

We are at an exciting time: for the astronomer there is the next generation of 8 and 10m class telescope to look forward to, after a lull of many years. Adaptive optics has appeared — for the first time there is a way to null the effects of the atmosphere within a finite field of view. Alongside these two is stellar interferometry, which has undergone a development unseen since the time of Michelson and Fizeau. Who can predict what we will see?

[2]See, for example, Max *et al.* [101]

# Appendix A

# Fringe Spacing Approximation

Accurate knowledge of the fringe spacing from any one baseline of a non–redundant recombination plane interferometer is required in order to realise the maximum signal-to-noise. Although the result of the geometrical model is sufficient as a first order estimator, a more accurate analysis is desirable. By knowing the variation of the fringe spacing with baseline, the appropriate fibre diameters can be specified (and consequently, any necessary mounting component dimensions). The Marcuse approximation, given in Equation 3.25, of the $LP_{01}$ mode suggests that paraxial Gaussian beams may be an appropriate approximation for the fibre waveguide fundamental mode outputs.

Kogelnik and Li [102] give the Gaussian spherical beam expression as

$$\vec{E}(x,y,z) = \sqrt{\frac{2}{\pi}}\frac{1}{w(z)}\exp\left[\frac{r^2}{w^2(z)}\right]\exp\left[-\mathrm{i}k\left(z-\frac{r^2}{R(z)}\right)+\mathrm{i}\phi\right] \qquad \text{(A.1)}$$

where $w(z_i)$ and $R(z_i)$ denote the beam spot radius and wavefront phase curvature respectively. The beam coordinate system is represented by $x_i, y_i$ and $z_i$ with its origin at the waist of beam $i$. The $z$ dependent complex phase shift is abbreviated by

$$\begin{aligned}\phi &\triangleq \arctan\left(\frac{\lambda z}{\pi w_0^2}\right)\\ &= \arctan\left(\frac{z}{z_R}\right)\end{aligned} \qquad \text{(A.2)}$$

where $w_0 = w(z = 0)$. The Rayleigh range $z_R$ is defined as

$$z_R \triangleq \frac{\pi w_0^2}{\lambda} \tag{A.3}$$

so that it is the point where the spot size $w(z)$ has increased by a factor of $\sqrt{2}$ compared with the value $w_0$ at the waist. The phase curvature $R(z)$ is obtained from

$$R(z) \triangleq z + \frac{z_R^2}{z} \tag{A.4}$$

and the spot radius

$$w(z) = w_0 \left[1 + \left(\frac{z}{z_R}\right)^2\right]^{1/2} \tag{A.5}$$

Consider the geometry depicted in Figure (A.1). The volume of the intersection of the two beams crossing with a half angle $\theta$, is defined in terms of an $x, y, z$ coordinate system. The origin of these coordinates lies at the intersection of the two beam centrelines. The appropriate transformations from the individual beam coordinates to the volume of intersection coordinates are

$$\begin{aligned}
x_1 &= x\cos\theta + z\sin\theta \\
y_1 &= y \\
z_1 &= -(x - x_{w_1})\sin\theta + (z - z_{w_1})\cos\theta \\
x_2 &= x\cos\theta - z\sin\theta \\
y_2 &= y \\
z_2 &= (x - x_{w_2})\sin\theta + (z - z_{w_2})\cos\theta
\end{aligned} \tag{A.6}$$

Interference fringes occur within the volume of intersection when the phase differences are a multiple of $2\pi$

$$\Delta\Phi = \Phi_1(x, y, z) - \Phi_2(x, y, z) = 2\pi n \tag{A.7}$$

where $n$ is an integer.

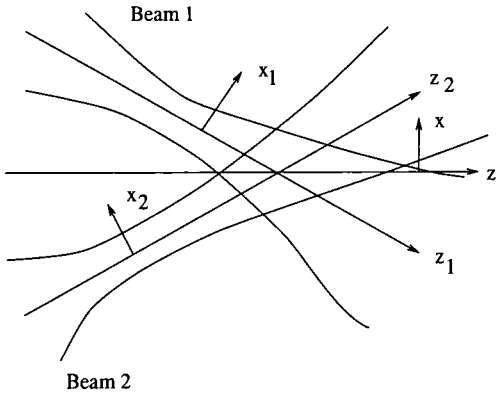



**Figure A.1.** Definition of the coordinate system for fringe geometry. The beam waists are at coordinates $x_{w_i}, z_{w_i}$.

The phase of each beam is seen to be

$$\Phi_i(x,y,z) = -k\left(z_i - \frac{x_i^2+y_i^2}{2R(z_i)}\right) + \phi(z_i) \tag{A.8}$$

from Equation (A.1). The phase difference is therefore

$$\Delta\Phi = -k(z_1-z_2) + \phi(z_1) - \phi(z_2) - \frac{k}{2}\left[\frac{x_1^2+y_1^2}{R(z_1)} - \frac{x_2^2+y_2^2}{R(z_2)}\right] \tag{A.9}$$

This describes the interference fringes as the fringes are surfaces of constant phase difference. The rate of change of this phase difference gives the inverse fringe spacing

$$\frac{1}{L} = \frac{d\Delta\Phi}{dx} \tag{A.10}$$

The full expression for the fringe spacing, $L = dx/d\Delta\Phi$ is

$$\begin{aligned} L &= \frac{1}{2k\sin\theta} \\ &\div \left[1 - \frac{1}{2\tan\theta}\left(\frac{x_1 z_1}{z_1^2+z_{R_1}^2} - \frac{x_2 z_2}{z_2^2+z_{R_2}^2}\right)\right. \\ &- \left(\frac{z_{R_1}}{2k(z_{R_1}^2+z_1^2)} + \frac{z_{R_2}}{2k(z_{R_2}^2+z_2^2)}\right. \\ &+ \left.\left.\frac{1}{4}(x_1^2+y_1^2)\frac{z_1^2-z_{R_1}^2}{(z_1^2+z_{R_1}^2)^2} + \frac{1}{4}(x_2^2+y_2^2)\frac{z_2^2-z_{R_2}^2}{(z_2^2+z_{R_2}^2)^2}\right)\right] \end{aligned} \tag{A.11}$$

In the divisor, the terms

$$\frac{z_{R_i}}{2k(z_{R_i}^2 + z_i^2)} \tag{A.12}$$

arise from the $z$ dependent phase shift, where $i = 1, 2$. The last two terms

$$\frac{1}{4}(x_i^2 + y_i^2)\frac{z_i^2 - z_{R_i}^2}{(z_i^2 + z_{R_i}^2)^2} \tag{A.13}$$

derive from the transverse phase derivative with respect to the phase curvature $R(z_i)$ plus changes in $R(z_i)$ with respect to $x$.

The fringe field can therefore be computed precisely throughout the intersection volume for arbitrary sizes and positions of the two beam waists. The beam waist positions model the output apertures of the optical fibres. It is useful to explore the magnitudes of the terms of Equation (A.11). The bounds of the $z$ dependent phase shift terms can be deduced from looking at the maximum value. As $z_i \to 0$ (i.e. the fibre outputs get closer to the recombination volume) we find

$$\frac{z_{R_i}}{2k(z_{R_i}^2 + z_i^2)} \to \frac{1}{2kz_{R_i}} = \left(\frac{\lambda}{2\pi w_0}\right)^2 \tag{A.14}$$

The final two terms of Equation (A.11) are bounded by the maximum beam radius $w(z_i)$ in the recombination volume.

$$x_i^2|_{max} = y_i^2|_{max} \approx w^2(z_i) \tag{A.15}$$

So, from Equation (A.5), the terms become

$$\frac{1}{4}w_{0_i}^2\left[1 + \left(\frac{z_i^2}{z_{R_i}^2}\right)^2\right]\frac{z_i^2 - z_{R_i}^2}{(z_i^2 + z_{R_i}^2)^2} = \left(\frac{\lambda}{2\pi w_{0_i}}\right)^2\left(\frac{z_i^2 - z_{R_i}^2}{z_i^2 + z_{R_i}^2}\right) \tag{A.16}$$

Therefore their absolute values are similar to the phase shift terms in that they are bounded by $(\lambda/2\pi w_0)^2$.

So these terms have very small magnitudes which may be neglected for a

first approximation. Equation (A.11) can then be simplified and written as

$$L = \frac{1}{2k\sin\theta}\left[1 - \frac{\left(\frac{x_1 z_1}{z_1^2+z_{R_1}^2} - \frac{x_2 z_2}{z_2^2+z_{R_2}^2}\right)}{2\tan\theta - \left(\frac{x_1 z_1}{z_1^2+z_{R_1}^2} - \frac{x_2 z_2}{z_2^2+z_{R_2}^2}\right)}\right] \tag{A.17}$$

Equation (A.17), as with Equation (A.11) can be used with the coordinate transforms of Equation (A.6) to compute the fringe spacing at arbitrary locations within the recombination volume.

Quantitatively, if we take $\lambda = 633\text{nm}$, $\omega_0 = 2\mu\text{m}$, $NA = \sin\theta = 0.1$ and assume the following parameters,

$$\begin{aligned} x_1 &= 25\mu\text{m} \\ x_2 &= -25\mu\text{m} \\ z_1 &= z_2 = -10 \times 10^3 \mu\text{m} \end{aligned}$$

we then find that the Rayleigh ranges are

$$z_{R_1} = z_{R_2} \approx 20\mu\text{m} \tag{A.18}$$

The fringe spacing (or width) is therefore $L \approx 42.3\mu\text{m}$. On the other hand, if

$$\begin{aligned} x_1 &= 100\mu\text{m} \\ x_2 &= -100\mu\text{m} \end{aligned}$$

then $L \approx 0.55\mu\text{m}$.

# Bibliography


[1] M. Born and E. Wolf, *Principles of Optics*, sixth ed. (Pergamon Press, 1993).

[2] J. Baldwin *et al.*, "The first images from an optical aperture synthesis array: mapping of Capella with COAST at two epochs," Astron. Astroph. pp. L13–L16 (1996).

[3] S. Clifford and J. Strohbehn, IEEE Transact. Antennas Propagation **AP-15,** 416 (1967).

[4] J. Goodman, *Statistical Optics* (Wiley–Interscience, 1985).

[5] C. Haniff, C. Mackay, D. Titterington, D. Sivia, J. Baldwin, and P. Warner, "The first images from optical aperture synthesis," Nature **328,** 694–696 (1987).

[6] D. Sivia, W. David, K. Knight, and S. Gull, "An introduction to Bayesian Model selection," Physica D **66,** 234–242 (1993).

[7] V. C. du Foresto, G. Perrin, and M. Boccas, "Minimizing fiber dispersion effects in double Fourier stellar interferometers," In *Fiber Optics in Astronomical Applications*, S. Barden, ed., Proc. SPIE **2476,** 86–95 (SPIE (Bellingham), 1995).

[8] V. C. du Foresto, G. Perrin, J.-M. Mariotti, M. Lacasse, and W. Traub, "The FLUOR/IOTA fiber stellar interferometer," In *Integrated Optics For Astronomical Interferometry*, P. Kern and F. Malbet, eds., pp. 115–125 (Grenoble, Observatoire de Grenoble, France, 1996).

[9] S. Shaklan, Ph.D. thesis, The University of Arizona, 1989.

[10] D. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," J. Opt. Soc. Am. **56,** 1372–1379 (1966).

[11] J. Goodman, in *Statistical Optics* (Wiley–Interscience, 1985), Chap. 8.

[12] A. Kolmogoroff, "The local structure of turbulence in incompressible viscous fluid for very large Reynold's numbers," Comptes Rendus (Doklady) de l'Acad. des Sciences de l'URSS **30,** 301–305 (1941).

[13] Taylor, In *Proceedings of the Royal Society of London.*, A **151,** 429–454 (1935).

[14] E. Gendron and P. Léna, "Single layer atmospheric turbulence demonstrated by adaptive optics observations," Astrophysics and space science **239,** 221–228 (1996).

[15] R. Racine and B. Ellerbroek, "Profile of night–time turbulence above Mauna Kea and isoplanatism extension in adaptive optics," In *Adaptive Optical Systems & Applications*, R. T. . R. Fugate, ed., Proc. SPIE **2534,** 248–257 (SPIE (Bellingham), 1995).

[16] D. Fried, "Anisoplanatism in adaptive optics," J. Opt. Soc. Am. **72,** 52–61 (1982).

[17] R. Myers, A. Longmore, R. Humphreys, G. Gilmore, B. Gentles, M. Wells, and R. Wilson, "The UK Adaptive Optics Programme," In *Adaptive Optical Systems and Applications*, R. Tyson and R. Fugate, eds., Proc. SPIE **2534,** 48–52 (SPIE (Bellingham), 1995).

[18] *Catalogue: Products for micropositioning*, Physik Instrumente GmbH & Co., Siemenstrasse 13-15, W-7517, Waldbronn Germany.

[19] P. Chen and S. Montgomery, "A macroscopic theory for the existence of the hysteresis and butterfly loops in ferroelectricity," Ferroelectr. **23,** 199–207 (1980).

[20] P. Ge and M. Jouaneh, "Modeling hysteresis in piezoceramic actuators," Precision Engineering **17,** 211–221 (1995).

[21] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical recipes* (Cambridge University Press, Cambridge, U.K., 1988).

[22] D. Buscher, "The ELECTRA realtime software architecture,", Internal document http://elsparc.dur.ac.uk/electra/architecture/global/globalArchitecture.html, 1996.

[23] P. Horowitz and W. Hill, *The Art of Electronics*, 2nd ed. (Cambridge University Press, Cambridge, U.K., 1990).

[24] A. Thompson, J. Moran, and G. Swenson, *Interferometry and Synthesis in Radio Astronomy* (Krieger Publishing Company, Malabar, Florida, U.S.A., 1994).

[25] C. Benn, "William Hershel Telescope environment data,", private communication.

[26] D. Marcuse, *Theory of Dielectric Optical Waveguides, Quantum Electronics – Principles and Applications* (Academic Press, 1974).

[27] D. Marcuse, *Light Transmission Optics* (Van Nostrand Reinhold, Princeton, New Jersey, 1972).

[28] G. Arfken, in *Mathematical Methods for Physicists* (Academic Press, 1985), Chap. 11.

[29] D. Gloge, "Weakly guiding fibres," Appl. Opt. **10,** 2252–2258 (1971).

[30] L. B. Jeunhomme, *Single-Mode fiber optics* (Marcel Dekker, Inc, 270 Madison Avenue, New York, New York, 1990).

[31] D. Marcuse, Bell Sys. Tech. J. **56,** 703–718 (1977).

[32] W. Gambling and H. Matsumura, "Propagation in radially inhomogeneous single–mode fibres," Opt. & Quantum Electron. **10,** 31–40 (1978).

[33] W. Gambling, H. Matsumura, and C. Ragdale, "Wave propagation in a single–mode fibre with dip in the refractive index," Opt. Quantum Electron. **10,** 301–309 (1978).

[34] I. P. Kaminow, "Polarization in Optical Fibers," IEEE J. Quan. Elec. **QE-17,** 15–22 (1981).

[35] C. Yeh, "Elliptical Dielectric Waveguides," J. Appl. Phys. **33,** 3235–3243 (1962).

[36] T. Do-Nhat and F. A. Alhargan, "Exact eigenvalue equations of modes in elliptical fibers of step–index profile," Radio Science **32,** 1337–1345 (1997).

[37] L. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw–Hill, Singapore, 1986).

[38] A. Smith, "Birefringence induced by bends and twists in single–mode optical fiber," App. Opt. **19,** 2606–2611 (1980).

[39] H. Lefevre, "Single–mode fibre fractional wave devices and polarisation controllers," Elec. Lett. **16,** 778–780 (1980).

[40] R. Ulrich, S. Rashleigh, and W. Eickhoff, "Bending–induced birefringence in single–mode fibers," Opt. Lett. **5,** 273–275 (1980).

[41] J. Nye, *Physical properties of crystals*, 1st ed. (Clarendon Press, Oxford, U.K., 1957).

[42] R. Ulrich and A. Simon, "Polarization optics of twisted single–mode fibers," Appl. Opt. **18,** 2241–2251 (1979).

[43] J. Love, W. Henry, W. Stewart, R. Black, S. Lacroix, and F. Gonthier, "Tapered single–mode fibes and devices Part1: Adiabaticity criteria," IEEE Proc-J **138,** 343–354 (1991).

[44] W. Gambling, H. Matsumura, and C. Ragdale, "Mode dispersion, material dispersion and profile dispersion in graded-index single-mode fibres," Microwaves, Optics and Acoustics **3,** 239–246 (1979).

[45] R. Jennison, "A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent," Mon. Not. Roy. Astron. Soc. **118,** 276–284 (1958).

[46] A. Readhead, T. Nakajima, T. Pearson, G. Neugebauer, J. Oke, and W. Sargent, "Diffraction limited imaging with ground–based optical telescopes," A. J. **95,** 1278–1296 (1988).

[47] D. Buscher, Ph.D. thesis, The University of Cambridge, 1988.

[48] T. Cornwell, "Radio–interferometric imaging of weak objects in conditions of poor phase stability — the relationship between speckle masking and phase closure methods," Astron. Astroph. **180,** 269–274 (1987).

[49] J. Brodie, R. Donnelly, H. Epps, M. Radovan, and W. Craig, "Efficiently mating fibers to spectrographs," In *Instrumentation in Astronomy VIII*, D. Crawford and E. Craine, eds., Proc. SPIE **2198,** 21–30 (SPIE (Bellingham), Kona, Hawaii, 1994).

[50] K. Johnson, "High–quality cleaves for better fiber splices," Photonics Spectra (1987), in the section "Fiber Optic Trends".

[51] G. Ramachandran and S. Ramaseshan, in *Crystal Optics/Kristalloptik*, Vol. XXV/1 of *Encyclopedia of Physics/Handbuch der Physik*, S. Flügge, ed., (Springer–Verlag, Berlin, Germany, 1961).

[52] C. Eckart, "The application of group theory to the quantum dynamics of monotonic systems," Rev. Mod. Phys. **2,** 305–380 (1930).

[53] M. Berry, "Quantal phase factors accompanying adiabatic changes," Proc. R. Soc. Lond. **A392,** 45–57 (1984).

[54] S. Pancharatnam, "Achromatic combinations of birefringent plates," Proc. Indian Acad. Sci. XLI (1955), section A.

[55] M. Martinelli and P. Vavssori, "A geometric (Pancharatnam) phase approach to the polarization and phase control in the coherent optics circuits," Opt. Comms. **80,** 166–176 (1990).

[56] D. Jackson, R. Priest, A. Dandridge, and A. Tveten, "Elimination of drift in a single–mode optical fiber interferometer using a piezoelectrically stretched coiled fiber," Appl. Opt. **19,** 2926–2929 (1980).

[57] B. White, J. Davis, L. Bobb, H. Krumboltz, and D. Larson, "Optical fiber thermal modulator," IEEE J. Lightwave Technol. **LT5,** 1169–1175 (1987).

[58] J. Burnett, A. Greenaway, R. McBride, and J. Jones, "Balancing optical path lengths in broadband fiber interferometers," Applied Optics **34,** 2194–2201 (1995).

[59] S. Kobayashi, S. Shibita, N. Shibita, and T. Izawa, "Refractive–Index Dispersion of Doped Fused Silica," In *Proc. International Conference on Integrated Optics and Optical Fiber Communication*, pp. 309–312 (1977).

[60] J. Burnett and J. Jones, "Cutting optical fibers to equal lengths for broadband stellar interferometry," Appl. Opt. **31,** 2977–2978 (1992).

[61] R. Haynes, Ph.D. thesis, The University of Durham, 1995.

[62] G. Perrin, V. C. du Foresto, S. Ridgway, J.-M. Mariotti, W. Traub, N. Carleton, and M. Lacasse, "Extension of the effective temperature scale of giants to types later than M6," Astron. Astrophys. **331,** 619–626 (1998).

[63] H. Dyck, J. Benson, G. van Belle, and S. Ridgeway, "Radii and effective temperatures for K–giant and M–giant and supergiants," A. J. **111,** 1705 (1996).

[64] H. Dyck, G. van Belle, and J. Benson, "Angular diameters and effective temperatures of carbon stars," A. J. 112 (1996).

[65] G. van Belle, H. Dyck, J. Benson, and M. Lacasse, "Angular size measurements of 18 Mira variable–stars at 2.2$\mu$m," A. J. **112,** 2147–2158 (1996).

[66] J. Baldwin, R. Boysen, C. Haniff, P. Lawson, C. Mackay, J. Rogers, D. St-Jacques, P. Warner, D. Wilson, and J. S. Young, "Current status of COAST," In *Astronomical Interferometry*, R. Reasenberg, ed., Proc. SPIE 3350 (SPIE (Bellingham), Kona, Hawaii, 1998), in press.

[67] D. Burns *et al.*, "The surface structure and limb-darkening profile of Betelgeuse," Mon. Not. R. astr. Soc. **290,** L11–L16 (1997).

[68] D. Burns *et al.*, "Large amplitude periodic variations in the angular diameter of R Leonis," Mon. Not. R. astr. Soc. **297,** 462–466 (1998).

[69] M. Shao and D. Staelin, "First fringe measurements with a phase–tracking stellar interferometer," Appl. Opt. **19,** 1519–1522 (1980).

[70] M. Shao, M. Colavita, D. Staelin, K. Johnston, R. Simon, J. Hughes, and J. Hershey, "Application of interferometry to optical astrometry," A. J. **93,** 1280–1286 (1987).

[71] M. Shao *et al.*, "The Mark–III Stellar Interferometer," Astron. Astrophys. **193,** 357–371 (1988).

[72] A. Hajian *et al.*, "Direct Confirmation of Stellar Limb Darkening with the Navy Prototype Optical Interferometer," A. J. **496,** 484–489 (1998).

[73] F. Roddier, In *High Resolution Imaging by Interferometry II*, F. Merkle, ed.,**29,** 565 (Garching, Germany, 1988).

[74] C. Hummel, "Images of Spectroscopic Binaries with the Navy Prototype Optical Interferometer," In *Astronomical Interferometry*, R. Reasenberg and J. Breckinridge, eds., Proc. SPIE 3350 (SPIE (Bellingham), Kona, Hawaii, 1998), in press.

[75] D. Mourard, I. Tallon-Bosc, A. Blazit, D. Bonneau, G. Merlin, F. Morand, F. Vakili, and A. Labeyrie, "The GI2T Interferometer on Plateau de Calern," Astron. Astrophys. **283,** 705–713 (1994).

[76] D. Mourard, I. Bosc, A. Labeyrie, L. Koechlin, and S. Saha, "The rotating envelope of the hot star gamma Cassiopeiae resolved by optical interferometry," Nature **342,** 520–522 (1989).

[77] P. Stee and F. de Araujo, "Line profiles and intensity maps from an axisymmetric radiative wind model for Be stars," Astron. Astrophys. **292,** 221–238 (1994).

[78] P. Stee, F. de Araujo, F. Vakili, D. Mourard, L. Arnold, D. Bonneau, F. Morand, and I. Tallon-Bosc, "gamma Cas revisited by spectrally resolved interferometry," Astron. Astrophys. **300,** 219–236 (1995).

[79] P. Stee, "On the kinematics of the envelope of gamma Cassiopeiae," Astron. Astrophys. **311,** 945–950 (1996).

[80] K. Rousselet-Perraut, F. Vakili, D. Mourard, F. Morand, D. Bonneau, and P. Stee, "An attempt to detect polarization effects in the envelope of gamma Cas with the GI2T interferometer," Astron. Astrophys. Suppl. **123,** 173–177 (1997).

[81] P. Stee, F. Vakili, D. Bonneau, and D. Mourard, "On the inner envelope of the Be star gamma Cas," Astron. Astrophys. (1998), in press.

[82] D. Mourard, D. Bonneau, L. Koechlin, A. Labeyrie, F. Morand, P. Stee, I. Tallon-Bosc, and F. Vakili, "The mean angular diameter of delta Cephei measured by optical long baseline interferometry," Astron. Astrophys. **317,** 789–792 (1997).

[83] M. Colavita *et al.*, "The Palomar Testbed Interferometer," Ap. J. , submitted.

[84] A. Boden *et al.*, "An Interferometric Search for Bright Companions to 51 Pegasi," Ap. J. Lett. , in press.

[85] M. Mayor and D. Queloz, "A Jupiter–mass companion to a solar–type star," Nature **378,** 355–359 (1995).

[86] G. van Belle *et al.*, "Radii And Effective Temperatures For G, K and M Giants And Supergiants," A. J. , submitted.

[87] W. Danchi, M. Bester, C. Degiacomi, L. Greenhill, and C. Townes, "Characteristics of dust shells around 13 late–type stars," A. J. **107,** 1469–1513 (1994).

[88] R. Wilson, J. Baldwin, D. Buscher, and P. J. Warner, "High-resolution imaging of Betelgeuse and Mira," Mon. Not. R. astr. Soc. **257,** 369–376 (1992).

[89] D. Buscher, C. Haniff, J. Baldwin, and P. Warner, "Detection of a bright feature on the surface of Betelgeuse," Mon. Not. R. astr. Soc. **245,** 7p–11p (1990).

[90] R. Gilliland and A. Dupree, "First image of the surface of a star with the Hubble–Space–Telescope," Ap. J. **463,** L29 (1996).

[91] P. G. Tuthill, C. A. Haniff, and J. E. Baldwin, "Hotspots on late-type supergiants," Mon. Not. R. astr. Soc. **285,** 529–539 (1997).

[92] A. Cheng, E. Hege, E. Hubbard, L. Goldberg, P. Strittmatter, and W. Cocke, "Diameter and limb–darkening measures for alpha–Orionis," Ap. J. **309,** 737–744 (1986).

[93] M. Scholz and Y. Takeda, "Model study of wavelength–dependent limb–darkening and radii of M–type giants and supergiants," Astron. Astrophys. **186,** 200–212 (1987).

[94] M. Smith, B. Patten, and L. Goldberg, "Radial–velocity variations in alpha–Ori, alpha–Sco, and alpha–Her," A. J. **98,** 2233–2248 (1989).

[95] A. Dupree, S. Baliunas, L. Hartmann, G. Nassiopoulos, E. Guinan, and G. Sonneborn, "Periodic photospheric and chromospheric modulation in alpha–Orionis (Betelgeuse)," Ap. J. **317,** L85–L89 (1987).

[96] G. Vaiana *et al.*, "Results from an extensive Einstein stellar survey," Ap. J. **245,** 163 (1981).

[97] A. Maggio, G. Vaiana, B. Haisch, R. Stern, J. Bookbinder, F. Harnden, and R. Rosner, "Einstein–observatory magnitude–limited X–ray survey of late–type giant and supergiant stars," Ap. J. **348,** 253–278 (1990).

[98] M. Bessel, "*UBVRI* photometry II: the Cousins *VRI* system, its temperature and absolute flux calibration, and relevance for two–dimensional photometry," Pub. Ast. Soc. Pacific **91,** 589–607 (1979).

[99] H. Campins, G. Rieke, and M. Lebofsky, "Absolute calibration of photometry at 1 through 5 $\mu$m," A. J. **90,** 896–899 (1985).

[100] J. Dainty and A. Greenaway, J. Opt. Soc. Am. **69,** 786 (1979).

[101] C. Max *et al.*, "Image improvement from a sodium-layer laser guide star adaptive optics system," Science, **277,** 1649–1652 (1997).

[102] H. Kogelnik and T. Li, "Laser beams and resonators," Appl. Opt. **5,** 1550 (1966).