some new techniques for faint object spectroscopy in astronomy

Ian R. Parry

Some New Techniques For Faint Object Spectroscopy In Astronomy

-by-

Ian R. Parry, B.Sc., M.Sc.

A thesis submitted to the University of Durham for the Degree of Doctor of Philosophy.

Being an account of work carried out at the University of Durham during the period October 1982 to February 1986.

Abstract

Astronomers require large amounts of spectroscopic data for faint astronomical sources if they are to successfully confront today's most important cosmological and astrophysical problems. However, until recently such data has been particularly difficult to acquire and the supply of telescope time available has fallen well short of demand. This thesis describes spectroscopic techniques of high efficiency that allow the data to be obtained using a minimum of telescope time. The technical aspects discussed include instrumentation, observing practices and data reduction. In particular, a new faint object spectrograph for the La Palma Observatory and an automated multi-fibre spectrograph coupler for the Anglo-Australian Telescope are described. It is now feasible for extensive spectroscopic surveys to be carried out at very faint magnitudes.

To Ishbel

Contents

Contents

Contents

Contents

Chapter 1

The need for faint object spectra

1.1 Introduction

Optical spectroscopy remains today, as it has been for many decades, the most powerful tool available for providing important astrophysical data. However, good spectroscopic observations have always been difficult to obtain, especially for the faintest of objects. Yet these faint observations are required in abundance if we are to answer some of the most challenging and exciting questions concerning the universe in which we live. The scant data that are available at present are fraught with complicated selection effects and only give us a tantalisingly incomplete glimpse of the full picture. Consequently, theoretical models are largely unconstrained by the observations and we are all too often forced to resort to assumptions, speculations and guesswork.

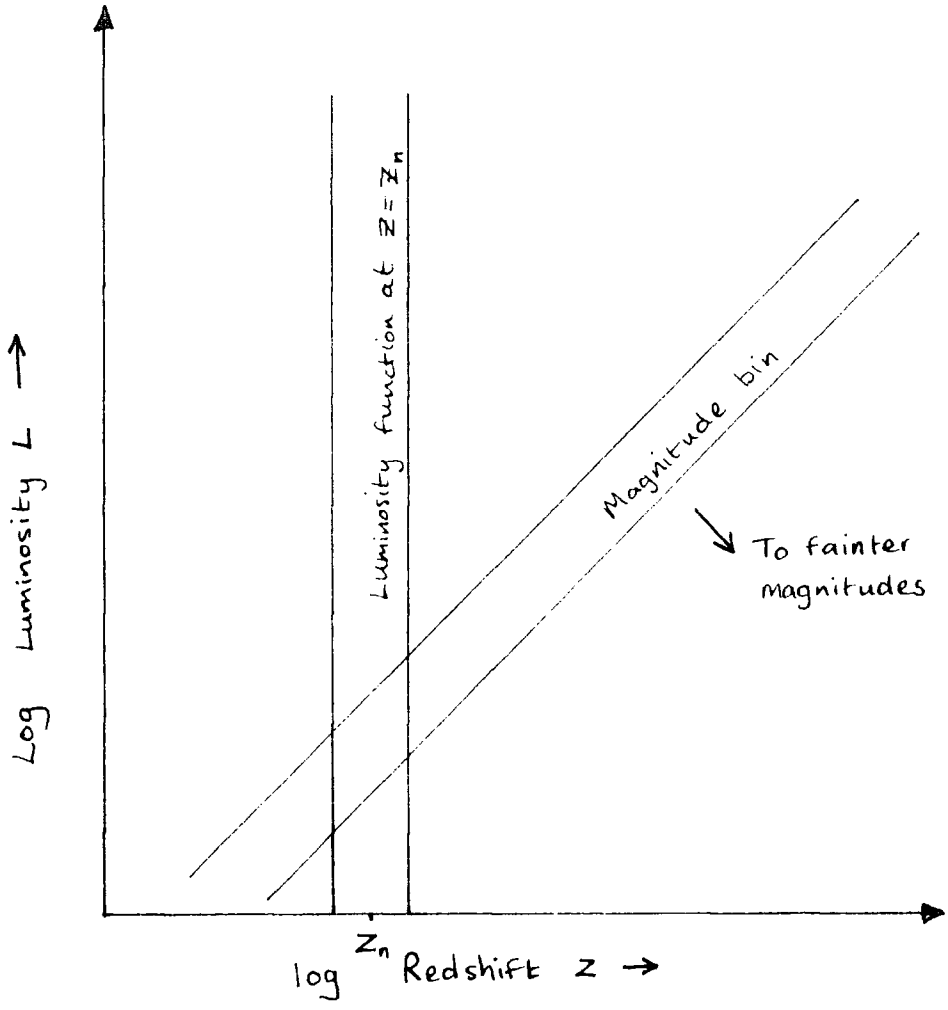
Faint object spectra are required for extra galactic work because we need to observe objects with long look-back times in order to study the universe at significant fractions of its present age. The cosmological interpretation of redshifts (which can only be determined via optical spectroscopy) enables us to determine the epoch at which the objects being studied lie. Similarly, for galactic work we need to study stars at significant distances from the solar neighbourhood and also

near-by intrinsically faint stars. In the next section I describe (with extreme brevity) some of the key research areas that require faint object spectra.

1.2 Research areas requiring faint object spectroscopy

Evolution*. There is considerable evidence that the properties of galaxies have evolved since the Universe was half its present age. The evidence for the evolution of the radio luminosity function is very strong (see for example, Wall et al 1980 & 1981, Peacock & Gull 1981, Djorgovski & Spinrad 1985) though the precise form of the epoch dependant radio luminosity function is unclear. The evidence for normal galaxy evolution is not so conclusive (see for example, Ellis 1984, Dressler & Gunn 1983 & 1985, Butcher & Oemler 1984, Longair 1984, Shanks et al 1984) though it does seem that blue spirals show significant evolution. Number counts and multi colour photometry of clusters of known redshift have been used in attempts to unravel the problems but these are no substitute for spectra which yield the nature and redshift of the galaxies in question. Figure 1.1 shows the luminosity redshift plane. In this plane the luminosity function at a given epoch (redshift) is represented by the population in a vertical column. A clear indication of evolution would be the variation of the form of the luminosity function with epoch. To populate this plane with galaxies (for a particular world model) we need the apparent magnitudes (photometry) and redshifts (spectroscopy) of a complete sample. A magnitude bin is represented by a diagonal strip. Number counts tell us the number of galaxies in each strip. However, because a galaxy

* of galaxies



The Luminosity - redshift plane

Figure 1.1

without a measured redshift can lie anywhere in its magnitude strip (i.e. it could be intrinsically faint and close or luminous and distant) then clearly, number counts alone cannot yield conclusive results with regard to evolution.

The dynamics of the Universe as a whole. We can in principle determine H_o and q_o* from observations of standard candles. In particular, the determination of q_o requires redshifts at very faint magnitudes as the variation of the rate of expansion of the universe can only be detected over large time scales (e.g. see Rees 1978 for an account of the methods of determining H_o and q_o). The accurate evaluation of these parameters is however very complicated in practice and a good understanding of galaxy evolution is required beforehand because the observations are very sensitive to evolutionary effects.

The large scale structure of the Universe. Using redshifts as distance indicators we can determine the space distribution of objects on the sky (e.g. Shanks et al 1983) This type of work requires large numbers of faint object spectra in order to probe the universe out to large distances and also to give us as complete a picture as possible of the clustering of nearby objects. Observations of this type can also tell us a great deal about the early Universe because different models predict different types of galaxy clustering (see Peebles 1980).

* the Hubble constant and the deceleration parameter

Active galaxies. Activity (i.e. galaxies with active nuclei*) was much more common in the past (see paragraph on evolution above) and therefore large numbers of these highly luminous objects can be seen to enormous distances. Faint object spectroscopy is essential if we are to understand these enigmatic objects. Furthermore, the extremely high luminosities associated with QSOs offers the possibility of observing a redshift cut-off provided we have sufficiently large and complete redshift samples. Such data, when available, will have strong implications regarding the epoch of galaxy formation.

Galaxy clusters. The study of galaxy clusters can tell us much about how galaxies interact with their environment. Spectroscopy allows us to determine the stellar mix of the galaxies concerned. An interesting new result has been the discovery of "starburst" galaxies in which massive bursts of star formation have been triggered (Dressler & Gunn, 1983). Spectroscopy can also tell us about the dynamics of the cluster members and hence the cluster mass and the amount of dark matter.

Galactic work. The study of the dynamics and populations of stars in our own galaxy will lead to an increased understanding of stellar evolution and galaxies in general. Via faint object spectroscopy we can study a large volume of space, significantly beyond the solar neighbourhood. Also, the faint end of the disk luminosity function can be defined more accurately and should gives us clues to the amount and nature of the dark matter in our galaxy.

* e.g. QSOs, Seyferts and radio galaxies

1.3 The complete spectroscopic survey.

Most spectroscopy can be classified as "follow-up" spectroscopy, i.e. an object has attracted attention to itself via some other form of observation before valuable telescope time is devoted to observing it spectroscopically (e.g. a object that is suspected as being a radio source from a radio position). The reason for this is, of course, that astronomers are reluctant to waste telescope time on an object that has a low probability of being "interesting". However, this approach has lead to a collective spectroscopic data set that is riddled with complicated selection effects and therefore quite useless for much of the science discussed above.

Clearly, a complete spectroscopic survey is desperately needed so that we can tackle these exciting problems. The keyword here is "complete" by which I mean that all objects within a specific area of the sky brighter than a certain magnitude limit* should be observed spectroscopically. The selection effects for such a sample are therefore as simple and well defined as possible. Each individual observation might not be all that interesting in itself but statistically every single observation would be important. Also every effort should be made to ensure that each spectrum has adequate signal-to-noise so that it can be unambigously classified. Most spectroscopic surveys to date attempt to observe a particular group of objects (e.g. QSOs) and therefore many potential spectroscopic targets are rejected (e.g. by morphology or colour) in advance of doing the spectroscopy. These surveys are therefore subject to additional selection

* Ideally total magnitudes should be used but for low surface brightness objects these are very difficult to determine. The magnitude determination procedure will therefore lead to complex selection effects.

effects which depend on the selection criteria used. Since all types of object are of interest to somebody or other it is better to observe everything and divide the observations into classes afterwards than have several small independant surveys each of which is aimed at a particular type of object.

The task of obtaining such a sample of spectroscopic data requires a considerable amount of organisation. Firstly there is the need to devote a very substantial slice of observing time to the project. The SERC working group who reviewed the observing time allocation (ROTA) process state in their 1983 report, as one of their main recommendations, that "At least 30% of the time on SERC allocated facilities should be made available for large (10 4-metre equivalent nights) statistical surveys of outstanding scientific merit". The panel for the allocation of telescope time (PATT) has also recently defined a "long term status" project as one which is guaranteed a total of 15 nights of clear telescope time. Unfortunately, however, PATT are understandably reluctant to grant projects long term status as it puts the proposers in an extremely responsible and privileged position. Futhermore, the full scientific interpretation of the observations I am proposing here (which would require more than 15 nights to complete) could not be delivered by a small section of the UK astronomical community. In short, I believe that our present time allocation procedure would block a large spectroscopic survey. The second organisational problem is that of data reduction. A considerable amount of observing data would be generated and it could not be interpreted until it is properly reduced and presented. Thus the data management aspects of the survey are

certainly not trivial.

1.4 Technical aspects

Above I have argued that astronomy is crying out for large amounts of faint object spectra. But is it technically feasibile that considerable quantities of faint spectra can be acquired with reasonable efficiency? In this dissertation I shall argue that the answer to this question is yes because the high efficiency of modern spectrographs can keep the required integration times to a minimum and developments in multi-object spectroscopy allow us to observe very many objects simultaneously. In chapter 2 I discuss the signal-to-noise of spectroscopic observations and the implications for the design of a faint object spectrograph and observing and data reduction practices. Chapter 3 describes the on-line data reduction package for the Faint Object Spectrograph (FOS) on the Isaac Newton Telescope on La Palma and chapter 4 describes the FOS itself including an account of its performance. The various methods of multi-object spectroscopy are described in chapter 5 which includes a comparison of the multi-slit and multi-fibre methods. Finally, chapter 6 describes an automated fibre-optic coupler (AUTOFIB) which will make the acquisition of large spectroscopic samples relatively straightforward.

Chapter 2

Faint Object Spectroscopy

Spectroscopy is concerned with measuring the intensity of the light from some astronomical object at different wavelengths. Each individual measurement is subject to random errors introduced by counting statistics and instrumental noise. Faint object spectroscopy can be defined as spectroscopy for which the photon counting error dominates because of a strong contaminating background sky signal which has to be evaluated and subtracted off. In addition, for each spectral channel there is a systematic error due to atmospheric absorption, losses in the telescope, losses in the spectrograph and the efficiency of the detector. For the detection of features in faint spectra the sky subtraction process is essential. Precise calibration of the systematic variation of efficiency with wavelength is less important as long as the variation is smooth on scales larger than the features that one is looking for. This is not true in the near infra-red where atmospheric absorption features are present. Of course, if accurate spectrophotometry is required to provide an insight into the physical conditions of the source, accurate calibrations must be determined to correct the systematic errors.

2.1 Sky Subtraction

2.1.1 Signal-to-noise ratio using a long slit

The random error (due to the detector and photon shot noise) associated with the signal in a spectral channel can be derived as follows. Poisson statistics, which relate to photon counting measurements, tell us that if over a long time period, T, we detect N photons then in a short time period, t, we expect to detect $Nt/T = n \pm \sqrt{n}$ photons, i.e. for many measurements each of duration t, the mean and the variance will be n. When several measurements are added together the error associated with the result is

$$\sqrt{(error_1)^2 + (error_2)^2 + \ldots\ldots}$$

Consider a spectral channel from a longslit observation that corresponds to a small wavelength interval. Figure 2.1 shows an example of the profile along the slit, in a single channel, for a small faint object. If there are m pixels that contain the object signal then the co-added sum of the signal from these pixels and its associated error is

$$\sum_{i=1}^{m} Q_i + mAS \pm \sqrt{\sum_{i=1}^{m} (Q_i + AS + R^2)}$$

$$= Q_{TOT} + mAS \pm \sqrt{Q_{TOT} + m(AS + R^2)} \qquad -(1)$$

where R is the root-mean-square readout/instrumental noise in photo-electrons, S is the number of sky photons detected per square arc-second, A is the area of sky which contributes to one pixel in the channel, Q_i is the number of photons detected in pixel i and Q_{TOT} is the total number of object counts in the

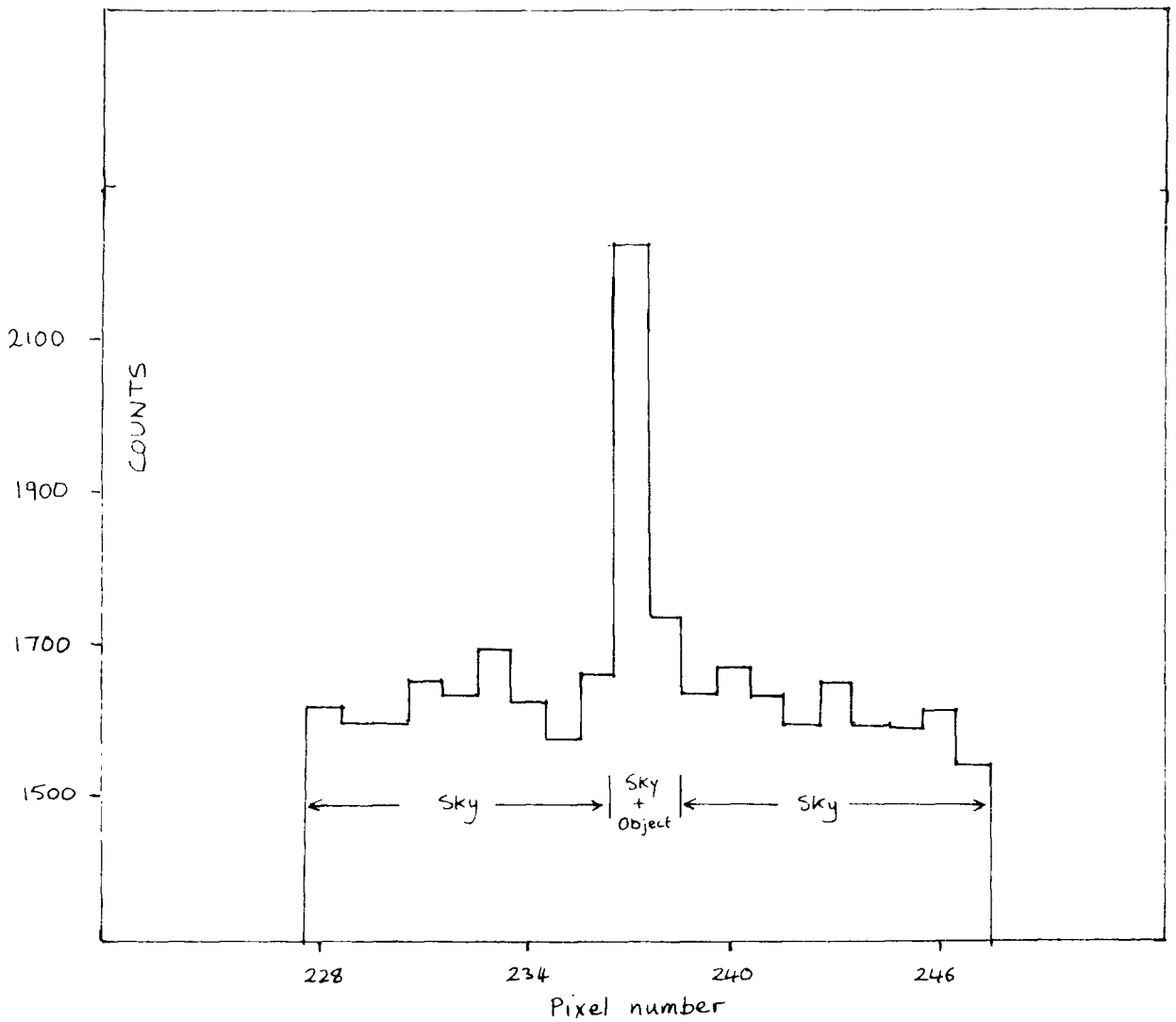


Figure 2.1 Example of data profile along the slit.

channel.

Similarly for the part of the channel illuminated only by sky, if n pixels are used to determine the sky for sky subtraction we have a co-added signal and its associated error of

$$nAS \pm \sqrt{\sum_{i=1}^{n} (AS + R^2)}$$

$$= nAS \pm \sqrt{n(AS + R^2)}$$

Multiplying the sky signal by m/n gives

$$mAS \pm \sqrt{\frac{m^2}{n}(AS + R^2)} \qquad -(2)$$

To perform the sky subtraction we subtract (2) from (1) and get

$$mAS + Q_{TOT} - mAS \pm \sqrt{[Q_{TOT} + m(AS+R^2)] + [\frac{m^2}{n}(AS+R^2)]}$$

$$= Q_{TOT} \pm \sqrt{Q_{TOT} + m(1 + \frac{m}{n})(AS + R^2)}$$

and so the signal-to-noise (S/N) is

$$\frac{S}{N} = \frac{Q_{TOT}}{\sqrt{Q_{TOT} + m(1+\frac{m}{n})(AS+R^2)}} = \frac{Q_{TOT}}{\sqrt{Q_{TOT} + X}}$$

where $X = m(1 + m/n)(AS + R^2)$

This result tells us several interesting things

a) when $Q_{TOT} >> X$ then $S/N = Q/\sqrt{Q}$ i.e. for bright objects the signal-to-noise is simply that due to the shot noise from the object itself.

b) The slit width is directly proportional to AS wheras Q_{TOT}

does not vary with slit width at all when the slitwidth is wider than the object. Thus the slit should be closed down so that AS is kept to a minimum without significantly reducing Q_{TOT}. In practice, a slitwidth marginally larger than the seeing disk maximises the signal-to-noise ratio for well resolved features (see Robertson, 1983).

c) The number of object pixels, m, is proportional to the seeing disk size and as AS is also proportional to the size of the seeing disk if the slitwidth is matched to the seeing then X is proportional to the square of the seeing (assuming R is negligible). Thus improvements in the seeing significantly improve the signal-to-noise ratio.

d) So that (1 + m/n) should be of the order of 1, about 10 times as many sky pixels as object pixels should be available in each spectral channel. On the other hand, as n tends to infinity only a small increase in S/N is obtained (and systematic errors increase, see section 5.6).

e) If R is a readout noise (i.e. it is introduced when the photons or photo-electrons are counted rather than while they are gathered) then for a given total integration time the fewer individual exposures there are the better. This is particularly true for CCD work at wavelengths where the system efficiency is poor or high dispersion is being used. Furthermore, the more photo-electrons gathered, the less detrimental are the effects of charge transfer and again minimising the number of exposures helps. On the other hand, CCD's suffer from cosmic ray events* which are best dealt with by comparing separate exposures against each other. When using a CCD it is therefore best to break the total exposure into

* Cosmic Ray Event is hereafter abbrieviated to CRE

just two or three individual runs.

f) Q and S are proportional to the exposure time t so we have two possibilities for the variation of S/N with time. Firstly,

$$\frac{S}{N} \propto \sqrt{t}$$

when shot noise limited (i.e $R^2 << AS$ or $Q_{TOT} >> X$) and secondly

$$\frac{S}{N} \propto t$$

when readout noise dominates (i.e. $R >> AS$ and $Q_{TOT} << X$).

2.1.2 Interpolative sky subtraction

In practice, the measured sky intensity along the slit for a spectral channel will have some systematic variation due to instrumental effects. In addition, for CCD work, the presence of CREs will introduce enormous errors in the sky estimate used for sky subtraction. To allieviate these problems it is best to interpolate the sky "under" the object using a low order polynomial least squares fit (typically one can use a parabola). To ensure that the fit is not mislead by any CREs that might be present, the sky measurements that were used to make the fit are compared with the fitted curve and any anomolous pixels are rejected. A new curve fit is made and the process is continued until no anomolous pixels are present. Note that this procedure removes all significant anomolies whatever their origin (e.g. pixel defects, other objects etc.). The signal-to-noise expression for interpolative sky subtraction is the one derived in the previous section but the value of n is less than the number of curve fitting pixels used as the fit at a point depends more strongly on the nearest

pixels used to make the fit. The actually value of n used depends on the order of the polynomial fit. For a polynomial of order zero, n is equal to the number of fitting points because the fit is equal to the average.

2.1.3 Combining spectral contributions using weights

The signal-to-noise of section 2.1.1 is valid when the pixel values that contribute to a channel are co-added i.e. if each pixel has a total count of C_i which is made up of Q_i counts from the object and S_i counts from the sky then the coadded result for a spectral channel is R_c

$$R_c = \sum (C_i - S_i)$$

The signal-to-noise associated with R_c, however, is not optimum (it cannot be as the signal-to-noise for a single pixel may well be greater). The optimum signal-to-noise is obtained using a weighted combination, R , given by

$$R_w = \frac{\sum w_i (Q_m/Q_i)(C_i - S_i)}{\sum w_i}$$

Note that the weights are not applied directly to the data but to the data scaled by Q_m/Q_i where Q_m is the object signal for the pixel that receives the maximum illumination. The reason for this is that the weighted combination technique is applicable only to the combination of several estimates of the same value which have different variances. In this well known technique the weights used are the inverse variances so in our case, where we wish to apply weights to scaled values, the weights themselves are the scaled inverse variances.

$$w_i = \left[\left\{ \frac{Q_m}{Q_i} \sqrt{C_i} \right\}^2 \right]^{-1} = \left(\frac{Q_i}{Q_m} \right)^2 \frac{1}{C_i}$$

In practice, we do not know Q_m or Q_i for individual channels so we use many channels summed together in the wavelength direction to get a good representation of the object profile along the slit. This gives weights and scaling factors that are applicable on average to the individual channels (see Robertson, 1983). Successfully applied weighting can increase the signal-to-noise ratio by something like 20% to 30%.

Because the data has to be summed along the wavelength direction to establish the weights and the scaling factors, the technique cannot be applied to data in which the object profile is not sampled in the same way for each channel. In other words, to apply the technique the spectrum must either be very straight and precisely lined up with the detector's pixel grid or it must be sufficiently over-sampled (which adds noise with CCDs) so that it can be resampled to make it appear to be straight and lined up with the pixel grid. For a spectrograph such as FOS, neither of these conditions is met as the pixel size is matched to the seeing and the spectra are curved by an order-separating cross-disperser. For curved spectra the weighting technique could in principle be applied if the appropriate weights, the scaling factors Q_m/Q_i, and an accurate scaling factor to increase R_ω (which is an estimate of Q_m) to the total count $Q_{TOT} = \Sigma Q_i$ can be estimated for each channel. However, in practice these are difficult to estimate accurately enough and the resulting spectrum suffers severe systematic errors. Errors in assigning the weights will yield a less than optimum signal-to-noise ratio but this could still be an improvement over co-adding. Errors in the determination of the scaling factors Q_m/Q_i and Q_{TOT}/Q_m are much more serious as they

lead to a systematic modulation of the spectrum which has quite sharp discontinuities that could easily be interpreted as spectral features. The discontinuities appear every time the point where the object is brightest lies exactly between two pixels. To calculate the scaling factors accurately we need to know the profile of the object along the slit precisely, in terms of its position and shape, for every channel. This information simply cannot be deduced from the observation of a faint object. Furthermore, an observation of a bright object cannot be used to deduce this information for another observation as its position and profile on the slit are bound to differ.

2.1.4 Beam switching

In this method of sky subtraction two separate apertures are used; one for the object and the other for the sky only. The object is switched between the two apertures so that it spends exactly half the total exposure time in each aperture. For one half of the exposure we have a total signal in one spectral channel of

$$C_{1A} = Q_1 + S_1$$

in aperture A and

$$C_{1B} = f S_1$$

in aperture B. The factor f is introduced to account for the fact that the efficiency through the two apertures may be different due to their sizes, the detector response, the spectrograph optics etc. Q denotes the object signal as before

and we are assuming that the sky is intrinsically the same in both apertures. Although this assumption is essential for all sky subtraction methods it may not be a very good one for apertures that are well separated on the sky (see section 5.6). For the other half of the exposure we have similar expressions for the detected signals

$$C_{2A} = S_2 \; ; \qquad C_{2B} = f(Q_2 + S_2)$$

If we calculate $C_{1A} + C_{2B} - C_{1B} - C_{2A}$ from our four measurements we get

$$\begin{aligned} R_{BS} &= C_{1A} + C_{2B} - C_{1B} - C_{2A} \\ &= [Q_1 + S_1] + f[Q_2 + S_2] - fS_1 - S_2 \\ &= Q_1 + fQ_2 + fS_2 - fS_1 + S_1 - S_2 \\ &= Q_1 + fQ_2 + S_2(f-1) - S_1(f-1) \end{aligned}$$

$$R_{BS} = Q_1 + fQ_2 + (f-1)(S_2 - S_1)$$

If we have perfect instrumentation then f=1 and $R_{BS} = Q_1 + Q_2$, which is the desired result. On the otherhand if $S_1 = S_2$, either because the observing conditions were photometric or the object was switched often and accurately between the apertures to average out temporal variations, then $R_{BS} = Q_1 + fQ_2$ which only depends on the object signal and the variation of f from channel to channel due, mainly, to the detector (the contribution to f from the apertures will be the same for all channels).

However in practice none of these conditions are likely to hold. Perfect instruments do not exist and photometric conditions are rare if not non-existent. Also, the rapid beam switching method which averages out the temporal variations is not practical for optical spectroscopy and requires a detector with no readout noise. Even if the conditions could be met the result would be subject to excessive noise as it corresponds to the m=n longslit case. The technique is, however, potentially useful when the apertures contain information on spatial structure which one wishes to preserve beyond the sky subtraction stage (e.g. for objects with extended emmission regions).

2.2 Systematic effects

The number of counts detected in a spectral channel depends on the total system efficiency at the channel's wavelength. The total system efficiency depends on the individual efficiencies associatied with the atmosphere, the telescope, the slit or fibre, the spectrograph and the detector.

2.2.1 Atmospheric losses

The continuum transmission of the atmosphere is shown in figure 2.2 (using data from Allen's Astrophysical Quantities). The precise form of the curve varies with height above sea-level, zenith-distance and weather conditions (including high altitude dust). Figure 2.2 shows the transmission at the zenith for average weather conditions at sea level. Objects

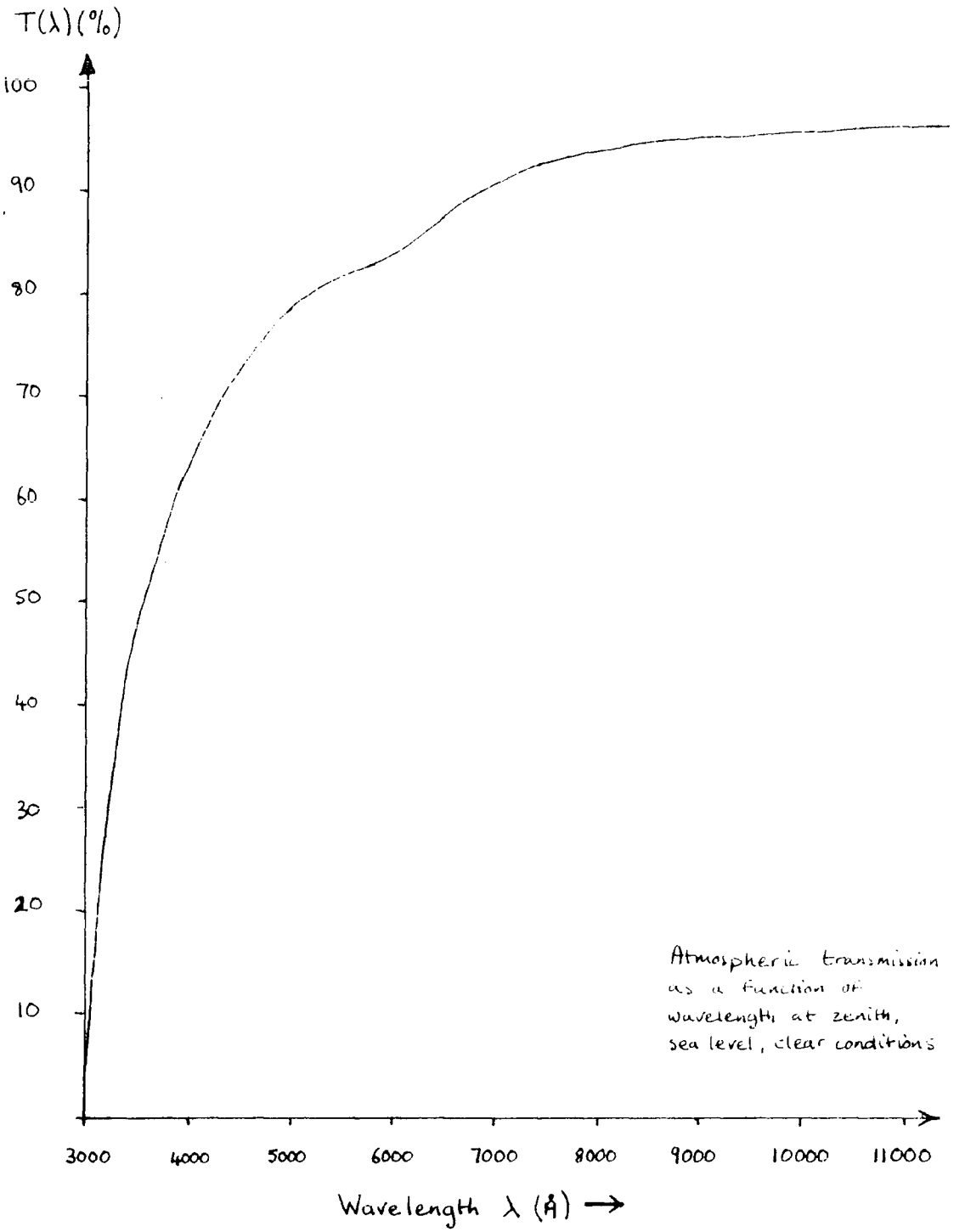

T(λ)(%)
100
90
80
70
60
50
40
30
20
10
3000
4000
5000
6000
7000
8000
9000
10000
11000
Wavelength λ (Å) →
Atmospheric transmission as a function of wavelength at zenith, sea level, clear conditions

Figure 2.2

appear redder as their zenith-distance increases (see figure 2.3). In addition to the continuum losses, which are mainly due to scattering, there are some absorption bands in the near infra-red due to oxygen and water molecules (see figure 2.4). The strengths of the O_2 features relative to the continuum absorption are quite stable. However, the H_2O features can vary significantly with time.

To accurately determine an object's flux at the top of the atmosphere one must also observe a flux standard under as similar conditions as possible. If, on the other hand, accurate spectrophotometry is not essential and only the removal of atmospheric features is required, observations of any bright featureless object will provide the calibration information. When studying faint objects, observations should be planned so that targets are acquired whilst crossing the meridian to ensure that atmospheric losses, particularly in the blue, are kept to a minimum.

2.2.2 Telescope losses

The efficiency of a reflecting telescope is fairly uniform throughout the wavelength region 3,000A to 10,000A. Typically the reflection efficiency of a bare aluminium mirror coating is 90% although there is a slight fall in efficiency at about 8,000A. The overall efficiency of an aluminium coating falls slightly as the metal tarnishes with time and hence the mirrors are annually re-aluminised. Of course, the fact that a reflecting telescope has a central obscuration means that the actual collecting area of the telescope is less than a circular

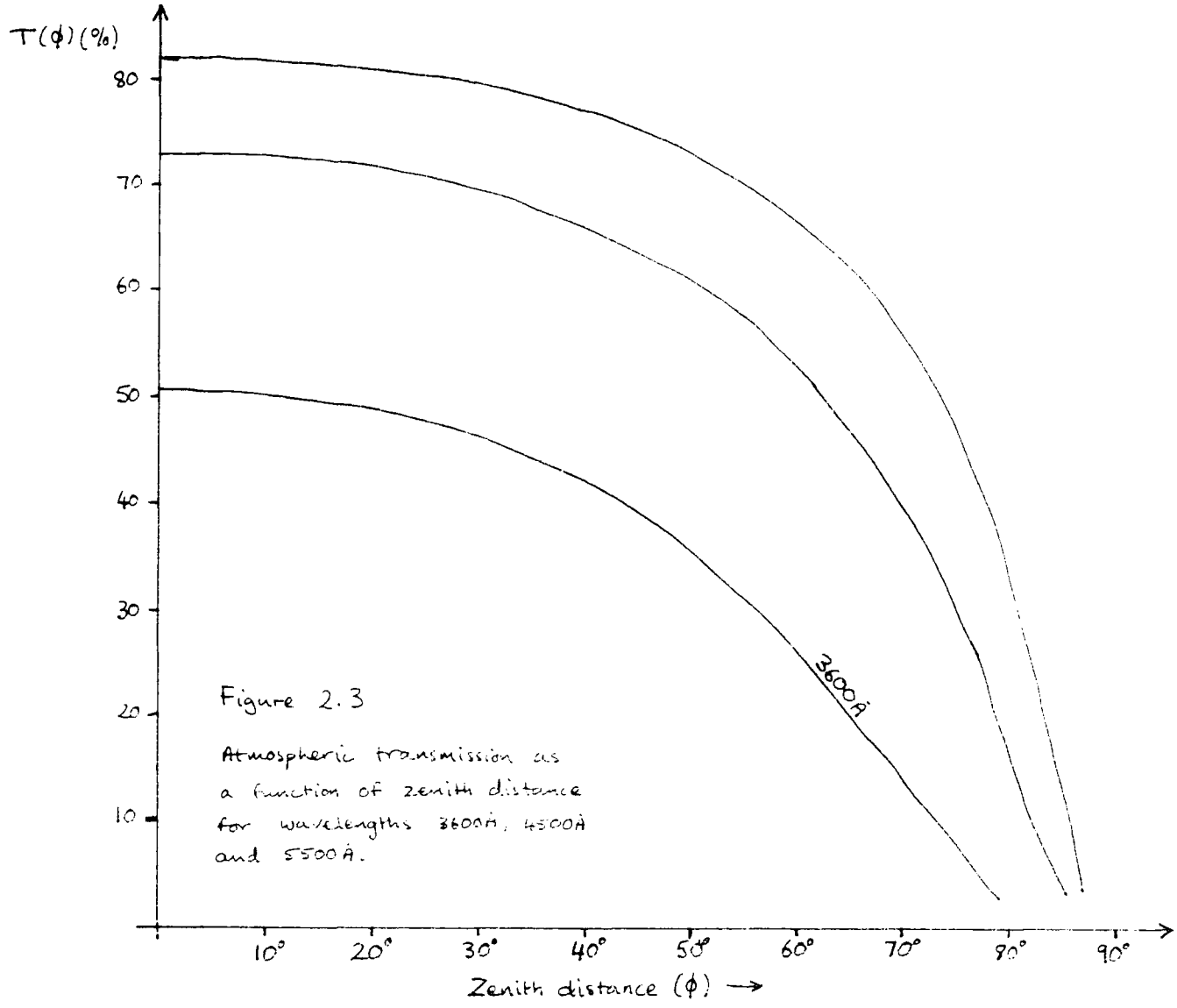


Figure 2.3

Atmospheric transmission as a function of zenith distance for wavelengths 3600Å, 4500Å and 5500Å.

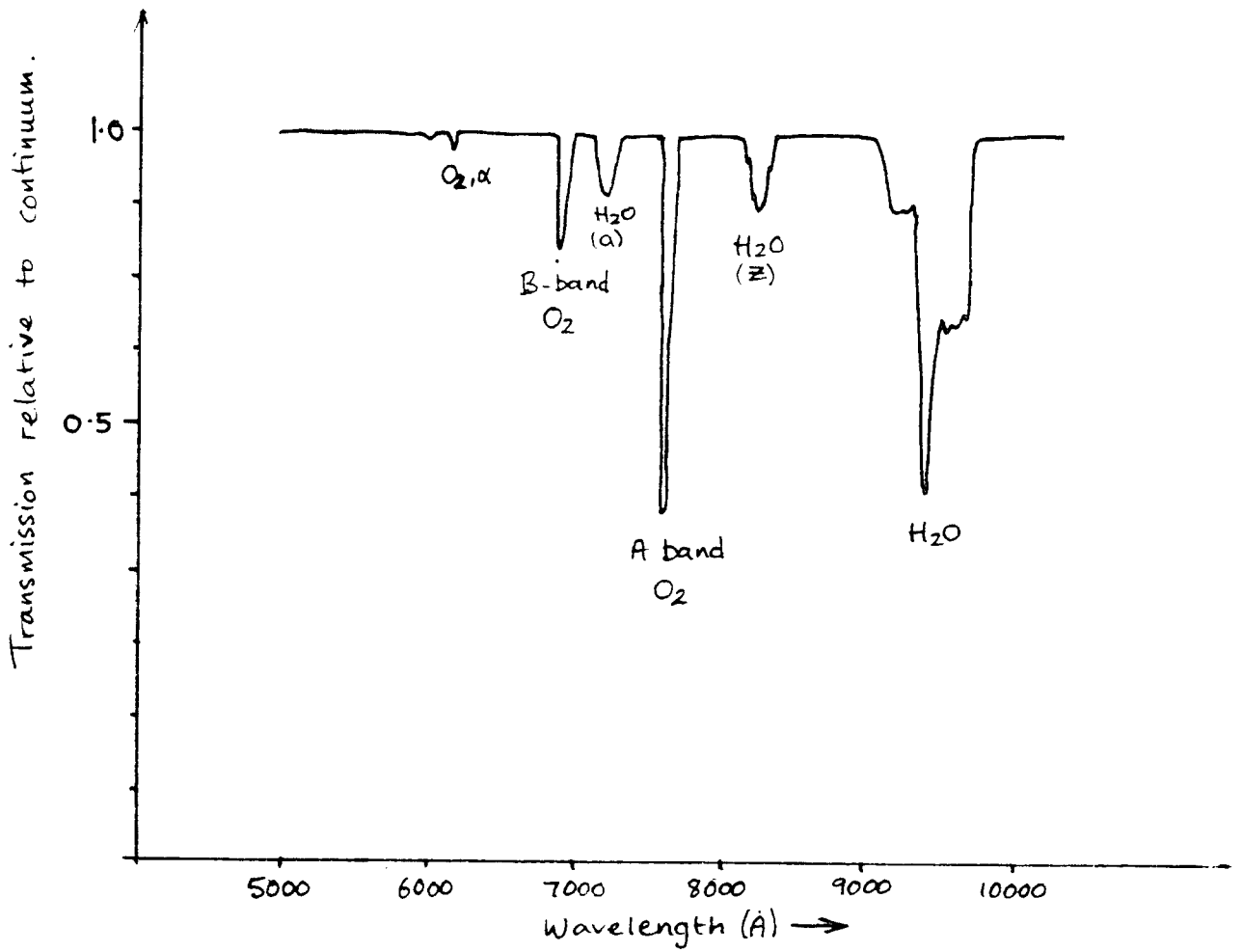


Figure 2.4 Atmospheric absorption lines

area of diameter equal to that of the primary mirror. The effective collecting area is typically 85% of the primary's total area. If one considers this as a loss then the overall efficiency of a telescope with two reflecting elements is about 70%.

2.2.3 Entrance Aperture losses

The spectrograph entrance aperture normally consists either of a slit or an optical fibre. The amount of light blocked by a slit depends on the seeing, the guiding accuracy and the slit width being used. The attenuation can also have a significant wavelength dependance at large zenith-distances ($zd > 40^{\circ}$) if the slit is not alligned along the direction of atmospheric dispersion. In this case a dispersed image falls across the slit at some angle and some wavelengths are vignetted more than others. Dispersion along the slit contributes to the total cross-dispersion of the spectrograph and can introduce spectral curvature for a system which otherwise produces straight spectra. When a large wavelength coverage is being used and the slit width matches the seeing, the slight dependance of the size of the seeing disk on wavelength (it gets smaller towards the red) makes the slit vignetting slightly wavelength dependant.

An optical fibre can suffer light losses due to focal ratio degradation (FRD), positioning errors (at both ends of the fibre), reflection losses and absorption within the core material. If the f-ratio of the input beam is larger (slower) than that of the output beam then the fibre is said to suffer

FRD. FRD introduces losses into a spectrographic system when it causes the fibre output beam to over-fill the spectrograph's entrance pupil (which is usually designed to accept the f-ratio of the fibre's input beam). FRD losses can be eliminated by using short, unstressed, quality fibres, fast fibre input beams and faster spectrograph collimators (or micro-lenses). Errors in positioning the ends of a fibre introduce vignetting (see section 5.6) and angular misalignments between the fibre's input end and the input beam introduce FRD. The output beam of the fibre also has to be accurately "aimed" at the spectrograph entrance pupil. Reflection losses can be minimised using anti-reflection coatings. Absorption losses are significant in the blue part of the spectrum and, for blue work especially, the fibres should be kept as short as possible. As an example, figure 2.5 shows the transmission, due to core absorption, for two types of fibre as a function of wavelength (from Powell, 1983 and Gray & Sharples, 1985). Quartz and Silice of France all silica (QSF-AS) fibre is available in two types - all silica dry (AS) or all silica wet (ASW) depending on the amount of OH present after the fibre has been drawn. The ASW type appears to have better blue transmission. Fibres are affected by atmospheric dispersion in the same way as slits except there is no prefered fibre orientation. Also any dispersion across the fibre aperture is completely scambled within the fibre and no extra cross-dispersion can be introduced. Seeing effects are also comparable for a slit and a fibre.

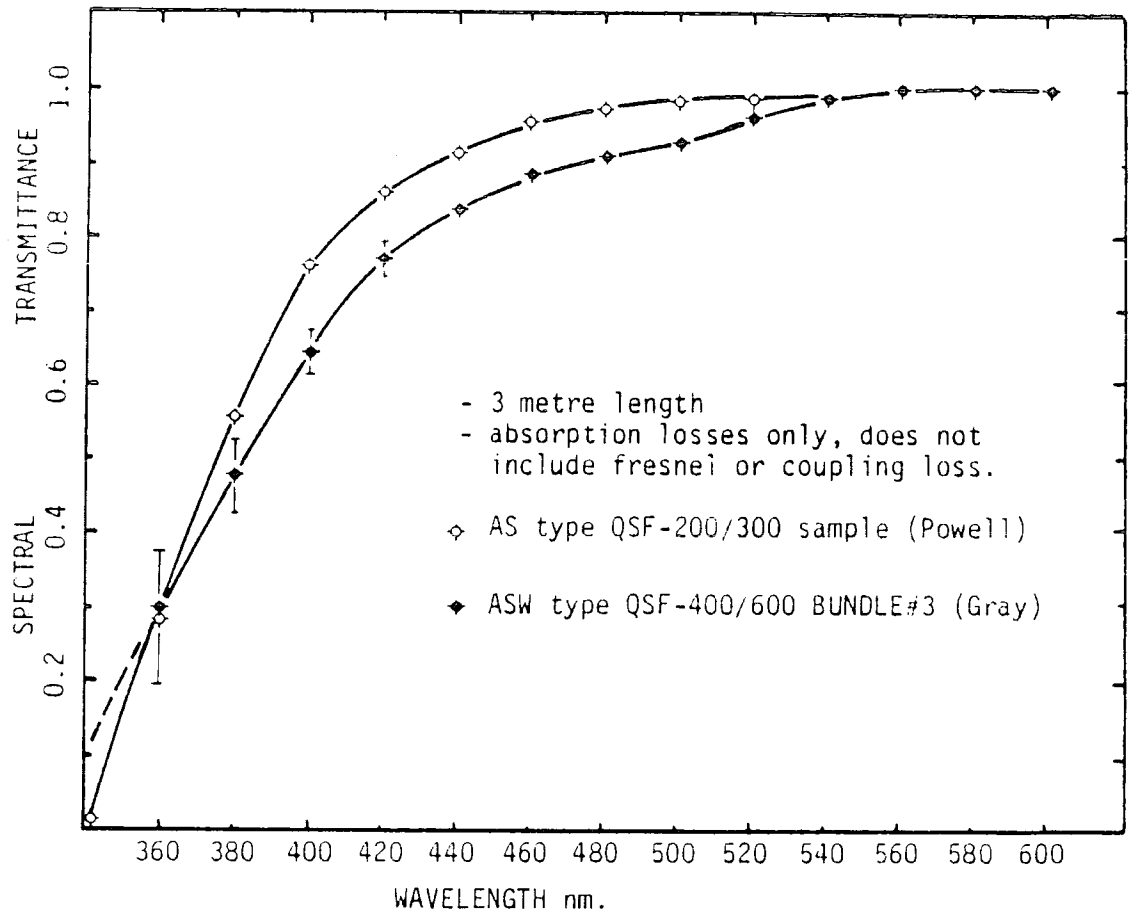


Figure 2.5

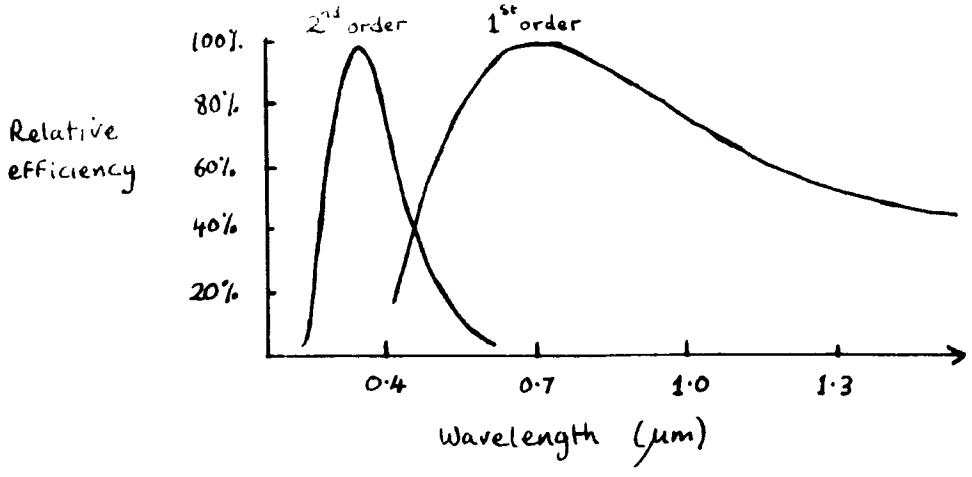


Fig 2.6. Efficiency of a grating as a fuction of wavelength (relative to blaze efficiency).

2.2.4 Spectrograph losses

The efficiency of a spectrograph depends on reflection/absorption losses in the optical components, off-axis vignetting, and the efficiency of the dispersing element. General optical losses in a well designed spectrograph are quite low and usually depend on the number of optical elements used. Reflecting elements have good efficiency throughout the optical wavelength range. Transmitting elements need to be anti-reflection coated and tend therefore to be designed for a particular wavelength range. Off-axis vignetting can be eliminated by either using a field lens or arranging the collimator so that it images the telescope pupil onto the spectrograph pupil (usually the grating). For multi-fibre spectroscopy off-axis vignetting can also be eliminated by "aiming" the slit ends of the fibres at the entrance pupil so that the fibres are fish-tailed.

The most significant optical losses are due to the grating. Figure 2.6 shows the general form of the variation of efficiency with wavelength. The peak efficiency occurs at the blaze wavelength and the efficiency falls off steeply on the blueward side and somewhat less steeply on the redward side. As most, if not all, of the other losses discussed here hit hardest in the blue part of the spectrum it is best to choose a blaze angle that makes the blue end of the required spectral range equal to the blaze wavelength. This policy is backed up by the fact that many astronomical sources have a flux distribution that increases with increasing wavelength and as $E=h\nu$ the number of photons rises even more steeply with

creasing wavelength (remember the signal-to-noise increases with the number of detected photons).

The use of a prism as the dispersing element instead of a grating has the advantage of providing a large free spectral range with good efficiency throughout. The disadvantages are that the dispersion is non-uniform (typically about 7-8 times higher at 4000A than at 10000A) and only low dispersions and low resolutions can be achieved. However, for some faint survey applications these disadvantages are not very significant and the use of a prism should perhaps be revived.

2.2.5 Detectors

For faint object work high quantum efficiency imaging detectors, which have a good dynamic range and allow convenient digitisation, are essential for sky subtraction and good signal-to-noise. Photographic plates are therefore unsuitable and the best detectors available at present are photon counting systems (which use a photocathode, a high gain image intensifier and discriminatory electronics to ensure that only photon events are recorded) and CCDs. Figure 2.7 shows the quantum efficiency curves of several useful detectors. The Tektronics CCDs which will be available in both small and large format seem ideal but at the time of writing are unproven in performance on a telescope (Blouke et al 1985). Increasing the sensitivity of the thick GEC CCD, by the use of a special fluorescent coating which converts blue photons to red ones which then have a greater chance of penetrating the polysilicon electrodes, is another promising development which has recently

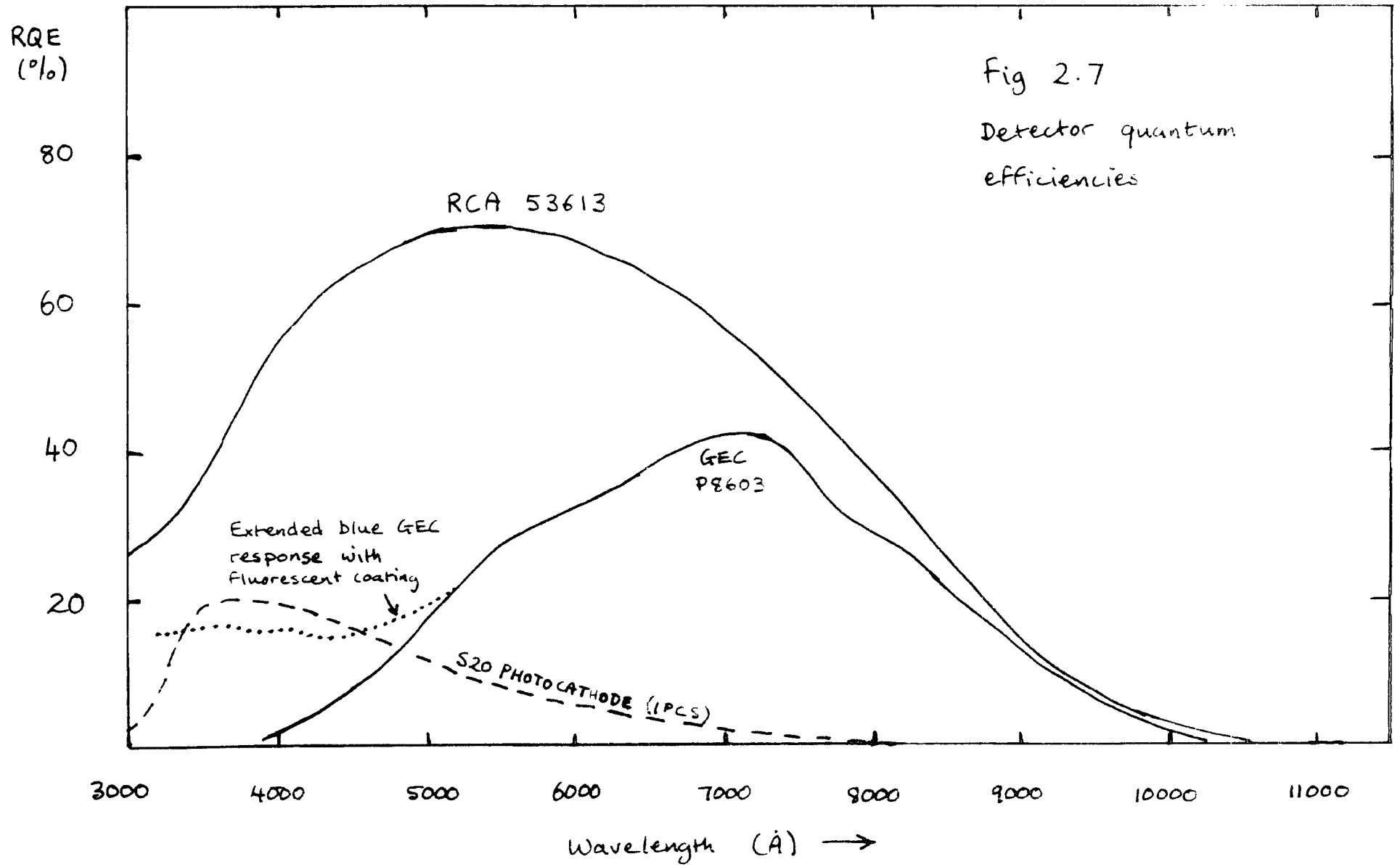


Fig 2.7

Detector quantum efficiencies

been tested on an astronomical spectrograph (Cullum et al, 198). A similarly coated GEC CCD will be installed in the LPO FOS during 1986. The only thin, blue-sensitive CCD to have been available to UK astronomers is the RCA type which unfortunately differs from the GEC and Tektronics CCDs in that its readout noise is poor (70 photo-electrons rms compared with less than 10 for the GEC and Tektronics chips). US astronomers have also been able to use a thinned Texas Instruments device. Thinned CCDs like the Tektronics and the RCA also suffer from fringing due to very small thickness variations in the silicon wafer. Presently available photocathodes are not as sensitive as CCDs although they have very good blue response. The IPCS (Image Photon Counting System) also has the advantage of little or no readout noise which is useful at high dispersion.

2.3 The background sky spectrum

It is the intensity of the sky spectrum that determines the limiting magnitude for faint object spectroscopy. Thus the deepest work is done when the sky background is weakest. By far the greatest factor affecting the sky brightness is the presence of the moon and the intensity increases with its proximity on the sky and with the lunar phase. Figure 2.8 shows two FOS spectra - one from an observation at new moon and the other at full moon.

In addition to the continuum background there are many emmision features present, particularly in the near infra-red where they are the main constituent of the airglow. The intensities of these lines can vary with time during the night,

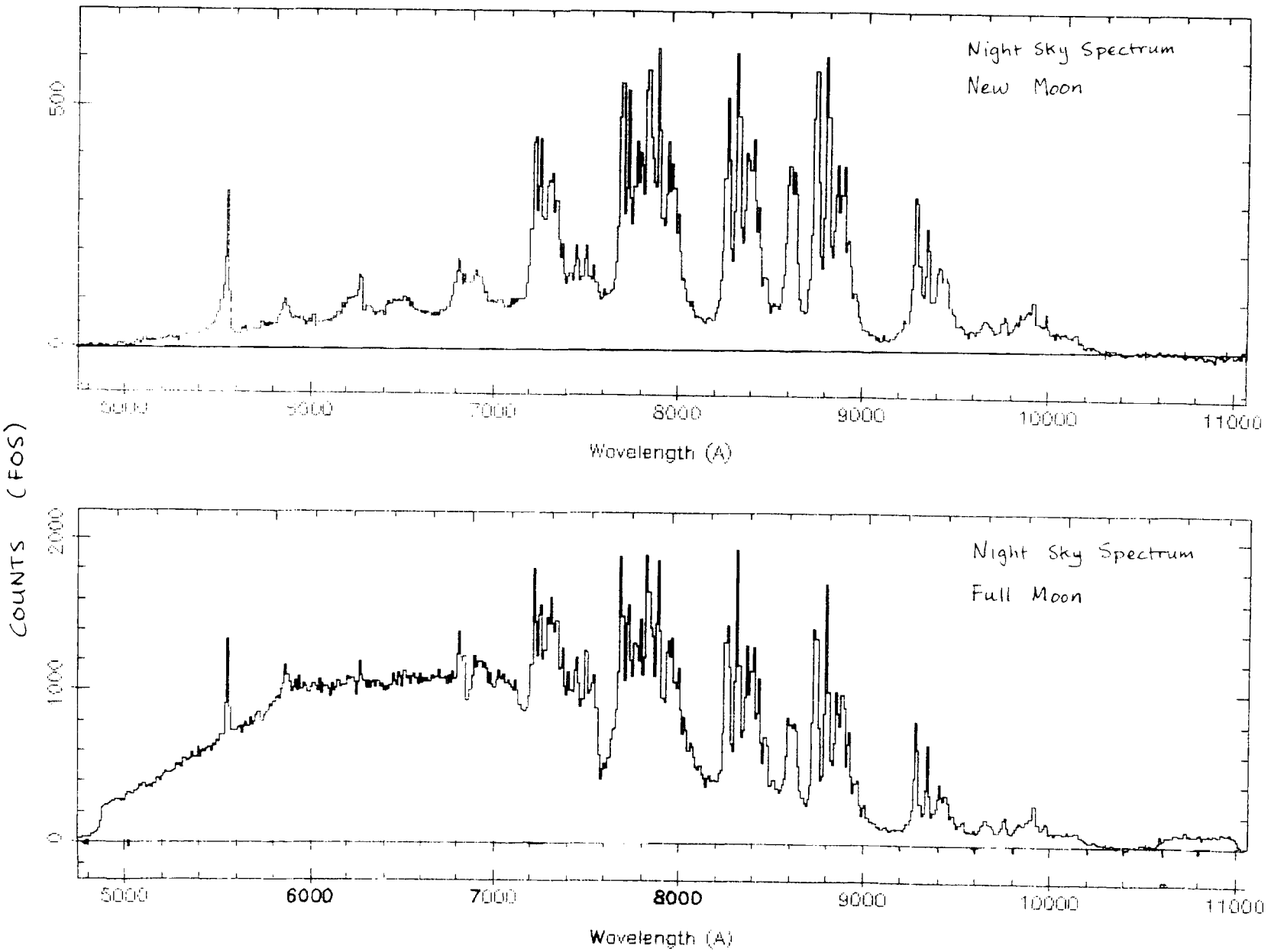

Night Sky Spectrum
New Moon
500
0
5000
6000
7000
8000
9000
10000
11000
Wavelength (A)
COUNTS (FOS)
Night Sky Spectrum
Full Moon
2000
1000
0
5000
6000
7000
8000
9000
10000
11000
Wavelength (A)

Figure 2.8

the season, the solar cycle and the observer's lattitude. For a comprehensive review of the night airglow see Silverman, 1970, and for an atlas of the emmission features throughout the entire optical window see Broadfoot & Kendall, 1968. The blue part of the optical window (3500A - 5000A) is virtually free from emission features. Figure 2.9 shows the night sky spectrum in the wavelength range 5000A to 10000A taken very early, on a moonless night. This wavelength region is dominated by OH emmission lines (Meinel 1950) which originate at an altitude of around 100km and are not dependant on sunspot activity or lattitude. Figure 2.10 shows the variation of intensity of several emmission lines during a relatively short, moonless, May night. FOS data from La Palma was used to derive the results. The measured intensities have been used directly without subtracting off the faint continuum which was constant throughout the dark part of the night. The results have been normalised to account for variations in the slit width and exposure times used.

There is clear evidence that the OH emmission fades as the night grows older. On the night in question the intensity drops to about two thirds of its initial value in eight hours. Other workers have also studied the nocturnal variation of the OH emmision and found a similar, fading trend (Fiocco et al, 1970). The different OH bands are also strongly coupled. There is some evidence, from the December 1984 FOS commissioning run, that the OH emmission continues to fade further on the longer winter nights and the emmission appeared to start off with an intensity similar to that seen in May. The variations of the night sky are, however, very erratic and

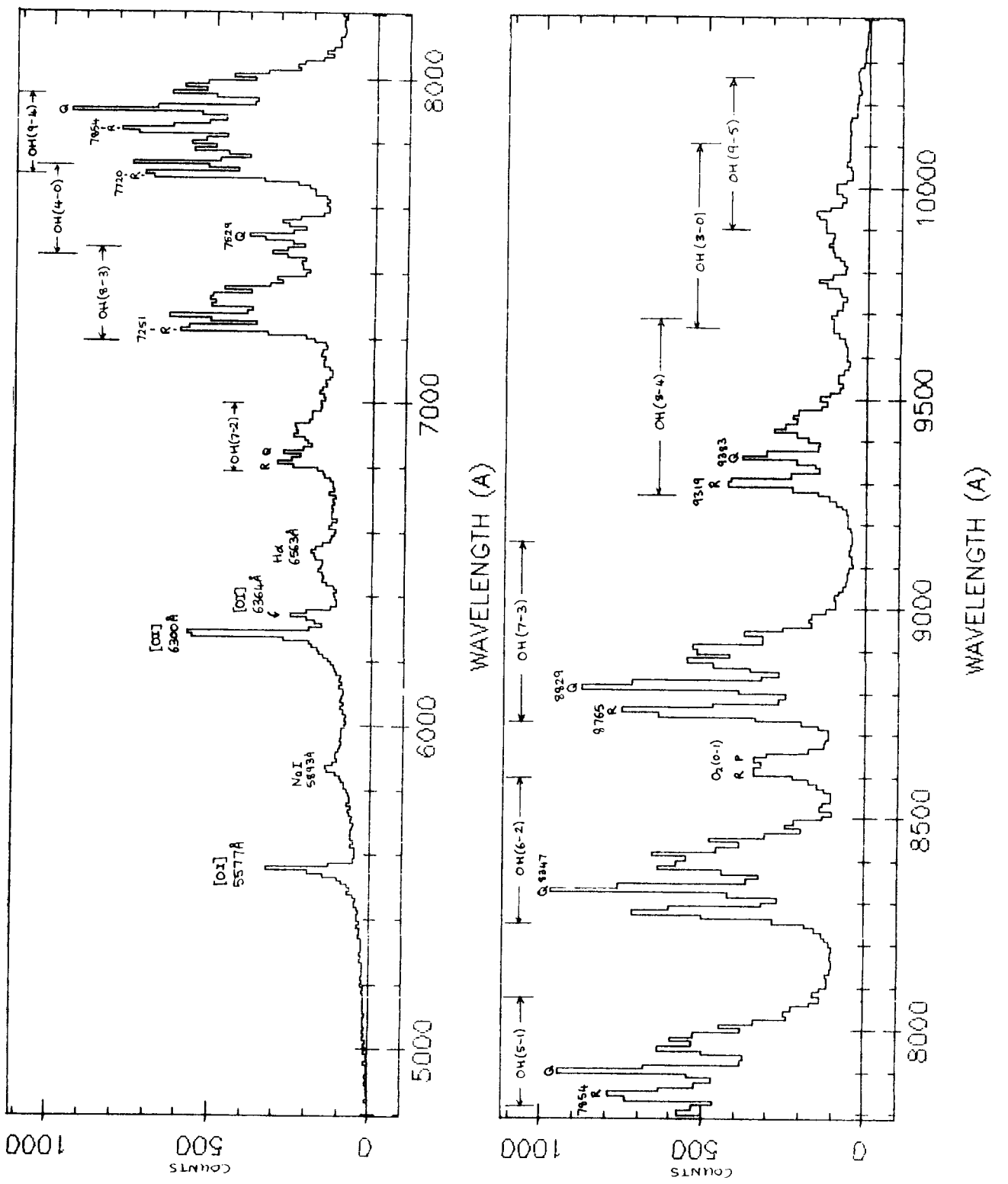


Figure 2.9 Night sky spectrum

Intensities are counts detected with FOS in 1000 seconds per pixel for a 1".5 slit width. Spectrum was taken soon after evening twilight.

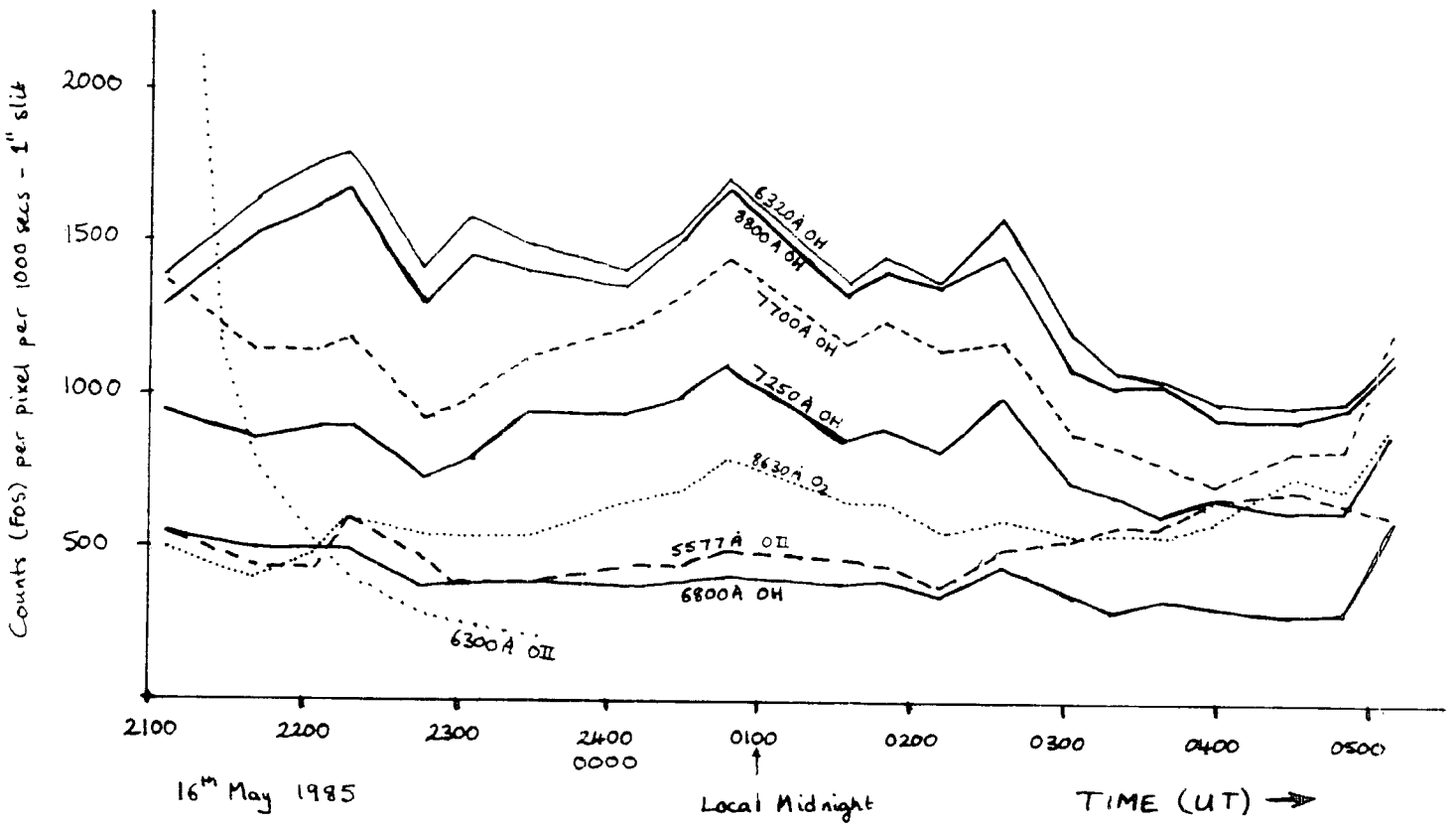


Figure 2.10 Temporal variation of sky emissions

large numbers of observations need to be studied to determine the underlying trends. As OH dominates longward of 7000A it is best to do faint work in this wavelength region just before morning twilight and therefore first quarter grey time is better than third quarter grey time.

No obvious trend appears in the variation of the strong OI line at 5577A but the OI lines at 6300A (and 6364A) decrease very significantly just shortly after evening twilight. Finally, the molecular oxygen emmision at 8645A shows no obvious nocturnal trend.

2.4 Conclusions for the design and use of a FOS

The ideal faint object spectrograph would cover the full optical spectral range from 3500A to 10000A, have an excellent system throughput throughout this wavelength region, a low sky background and insignificant readout noise in all wavelength channels. Unfortunately this is impossible to achieve in practice.

The full spectral range is more than one free spectral range (octave) and therefore either a cross-dispersed grating or a prism must be used. Cross dispersion rules out the use of a weighted sky subtraction algorithm for improved signal-to-noise and reduces the number of objects that can be observed simultaneously when employing a multi-object technique. The non-linear wavelength scale and low dispersion provided by a prism are somewhat inconvenient but perhaps the best for very faint low dispersion work. Another alternative is to use a dichroic beam-splitter behind the slit to send the

red and the blue halves of the full spectral range to two seperate spectrographs. Unfortunately, dichroics have very uneven (though reasonably high) efficiency in both beams and there is a narrow waveband between the red and blue halves which has very poor efficiency. On the other hand, the two spectrographs have to work in a smaller wavelength region than a single spectrograph and can be designed with better efficiency.

Atmospheric losses can be reduced by observing at low zenith distances from very high mountains or better still from a satellite observatory. The broadband efficiency of a reflecting telescope is already very good and little room for improvement exists other than the direct use of prime focus instead of the more commonly used Cassegrain focus. Slit apertures are simple devices and little, if any, design improvement is possible also. In contrast optical fibre feeds would benifit from further design improvement both in terms of positional allignment methods (see chapter 6) and inherent fibre quality. Efficient spectrograph optics (cameras and collimators) are presently available which use the minimum number of reflecting/transmitting surfaces. However, gratings cannot provide very high efficiency throughout the entire optical region and for a very large wavelength interval a prism might be more effective. Finally, low noise, thinned, large-array CCD detectors (e.g. the Tektronics chip) offer excellent quantum efficiency throughout the optical spectral range and room for further improvement is now becoming hard to find (photon counting systems can be superior at high dispersions but such dispersions are not very useful at very

faint magnitudes). It may be possible to deposit customised fluorescent coatings onto thinned CCDs used in fixed format spectrographs (i.e. optimise the coating at each location on the chip for the wavelength of light it will receive).

The observer should take care to get the most from his observing time by making observations at small zenith distances with a slit width matched to the seeing to reduce the sky background. If a large zenith distance has to be used then the slit should be set at the paralactic angle, along the direction of atmospheric dispersion. Very faint work can only be done under moonless conditions and the sky is generally faintest at the end of the night, just before the onset of morning twilight. When using a CCD the number of exposures for an object should be kept to a minimum to reduce the effects of readout noise and charge transfer inefficiency although at least two exposures are required to deal with CREs.

Many aspects of the data reduction process are important if one is to get the best from one's data. Also, it is very useful to be able to reduce the data whilst observing so that the best use is made of the observing time and facilities. Sky subtraction is the most critical data reduction process. Interpolation to determine the sky and a weighting technique should be used wherever possible. Also, the more sky used to determine the sky spectrum, the better the results will be. When weighting is not used one should be careful not to co-add sky subtracted pixels that do not contain significant object signal. Likewise one should ensure that all pixels that contain significant object signal are co-added. It is

important to remove (or at least recognise) features, such as CREs and atmospheric absorption lines, that have nothing to do with the object spectrum. Flatfielding is important if the flatfield noise is comparible to the shot noise (not often true for faint object spectra).

Chapter 3

The FOS on-line data reduction package

3.1 Introduction

The Faint Object Spectrograph (FOS) on the 2.5 metre Issac Newton Telescope (INT) would be an incomplete instrument if the astronomer could not assess the quality of his data as he observed. Rapid data inspection ensures that his allocated telescope time is used with maximium efficiency because exposures that are either too long or too short can be avoided. The software package that enables the astronomer to do this should also allow him to carefully reduce his data during the daylight hours to a very high standard so that its full information content is revealed before he returns to his home institution.

The package that was developed consists of a set of reduction modules that operate on the data sequentially. For rapid inspection purposes some of these can be left out and "safe" program parameters used. For full data reduction, all reduction stages are employed and each stage may require that the operator iterates to optimise the program parameters.

The package was developed on the Durham starlink VAX 11/750 in FORTRAN using the ARGS display (a lexidata display is used on La Palma) and the interim starlink environment for handling bulk data. Making the programs work on the La Palma PE 3220 was non-trivial as the bulk data i/o (the PE3220 is not a virtual memory machine like the VAX) and the image display

interactions are completely different. Some experience at converting code to run on the PE3220 was gained during a couple of visits to the Royal Greenwich Observatory (RGO) while FOS was still under construction, but the bulk of the package was put together and debugged during three FOS commissioning runs in July '84, December '84 and May '85, after which the package was officially handed over to the La Palma software team. In particular, all the lexidata interaction code was created during the fortnight-long La Palma visits as the PE3220's at RGO do not have lexidatas. Another problem associated with running the programs on the PE3220 is that it has no general purpose mathematical subroutine libraries, such as NAGG, so the programmer has to write his own. Therefore it was necessary to write a general purpose curve fitting subroutine to deal with the curved FOS spectra. In fact three routines were written, one to set up the normal equations (including the use of pivots to reduce arithmetical errors) for a least squares fit, a routine to solve simultaneous equations to solve the normal equations and obtain the polynomial coefficients and a simple routine to accurately give the value of a polynomial given an argument and the coefficients. Polynomial fits to 20th order are possible. The development code which exists on the VAX is not as polished as the La Palma programs and the task of releasing a FOS reduction package for STARLINK users remains to be done.

The fixed format nature of FOS allows a tailor-made package to be developed which makes many (hopefully sensible) assumptions about the data. It also helps the user to reduce his data quickly and enables standard data formats to be

adopted. The package has also been written so that reduced FOS data can be analysed using other standard spectral reduction packages like SPICA on STARLINK.

The reduction package is run from within the ADAM (Astronomical Data Acquisition Monitor) environment on the PE3220 which provides other data display and reduction programs and lets the astronomer control his instrumention. It is also possible to write ADAM procedures to string together several of the reduction programs, and thereby streamline the data reduction task.

3.2 The form of the data provided by FOS

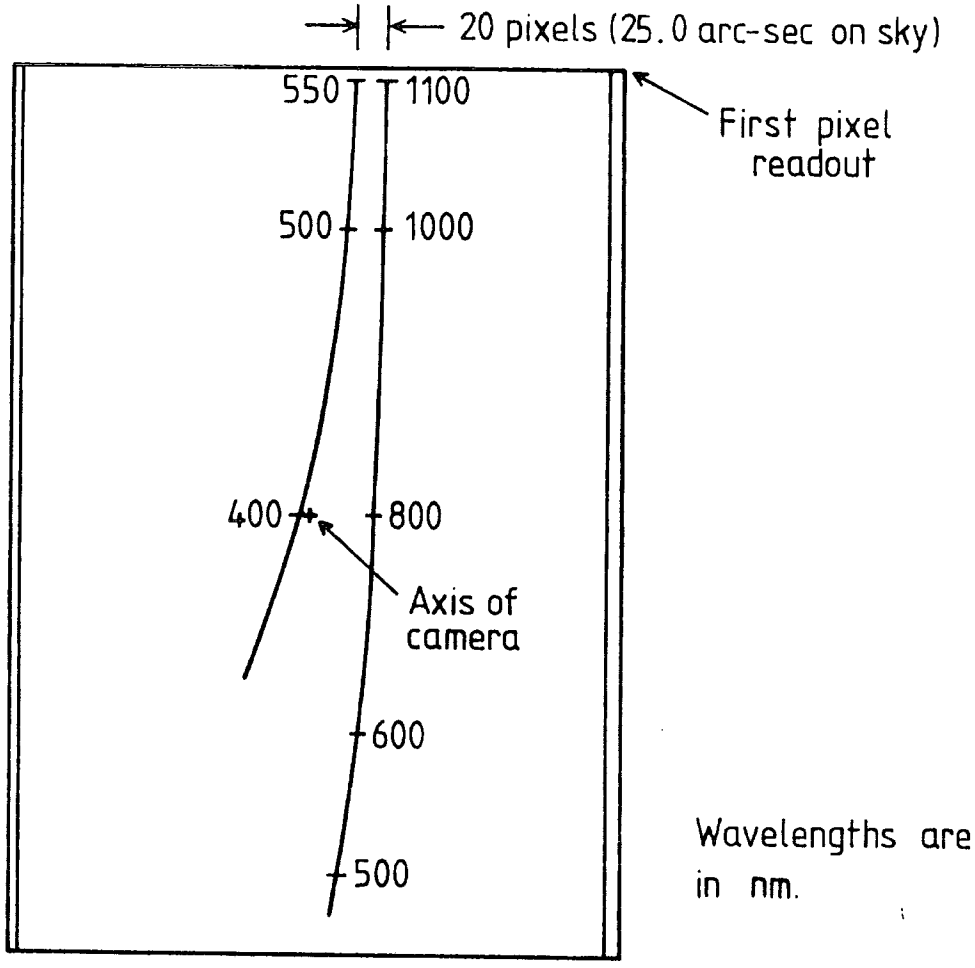
The data files produced by FOS are stored as bulk data frames (BDFs) on the PE 3220's fast winchester as the CCD is read out. Once there, it is usually automatically displayed as an image on the lexidata system and may also be copied to tape as a safety measure. These "raw" FOS data files are always 400 columns by 590 rows in size with each pixel being stored as a 2-byte integer (0.472 Mbytes per image). The appearance of the raw data depends on the slit, dekker and filter options selected and also on the source of illumination. The data will also be affected by readout noise, CCD defects, cosmic ray events (CREs) and a bias level generated by the CCD electronics.

In the absence of illumination a dark exposure will show the DC bias level (about 150 counts) with its characteristic noise (about 7 electrons rms), possibly some bright column defects and if the exposure time is long enough some CREs.

These CREs appear as bright spots, often somewhat elongated, of a few pixels extent. CREs of upto a few thousand counts can occur and significant events (more than 50 counts) occur at a rate of about 17 per 1000 seconds per image. Note the CREs are confined to the light sensitive area of the CCD which is only 385 columns by 578 rows. Not all the columns are light sensitive because the CCD's readout register is wider than the light sensitive area and the first 11 and the last 4 elements do not receive photo-electrons. The last 12 rows of data are not photo-sensitive either, because they are merely the contents of the readout register after the photo-electrons have been clocked out. Any illumination effects seen in the overscan regions are caused by either charge transfer inefficiency or charge flooding. The 12 overscan rows are read out to determine the bias level.

A point continuum source at the centre of the telescope field will produce the characteristic first and second order, curved spectra (see figure 3.1). The curvature is mainly due to the cross dispersing element of the FOS grism which cannot be altered, but atmospheric dispersion along the slit, which varies with zenith distance, can slightly modify the total amount of cross dispersion in the system and hence the curvature of the spectra. For zenith distances less than 30 degrees, or for a slit at right-angles to the atmospheric dispersion, the effects on the curvature are negligible.

The selection of various slit lengths (via dekker masks) and below slit filters gives rise to the various data modes which are shown in figure 3.2. Mode A is obtained by selecting



Shape of FOS spectra

Fig 3.1

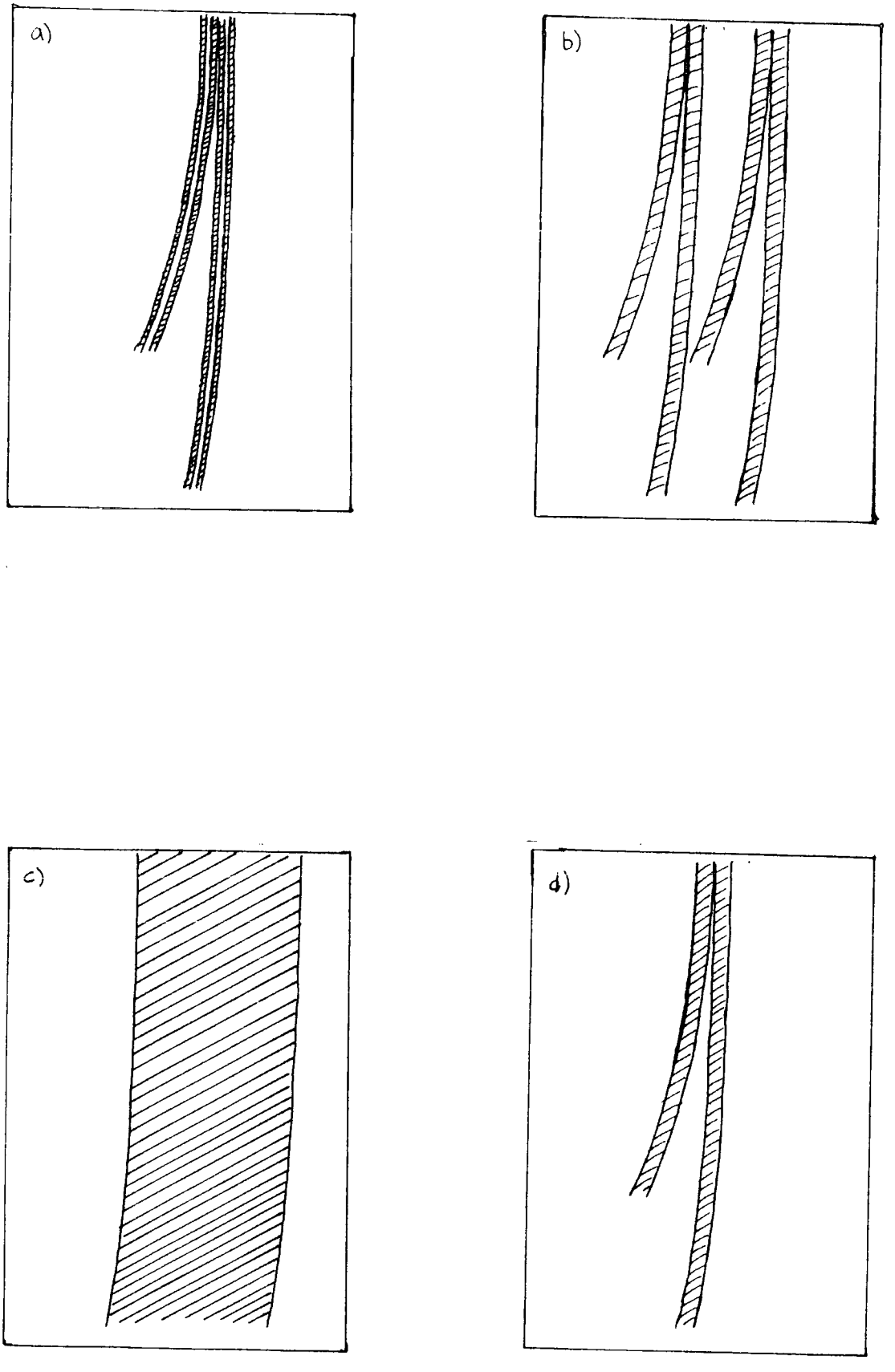


Figure 3·2

FOS data modes

a dekker which gives 2 short slit lengths with their centres 12.5 arc-seconds apart. Three such dekkers are available with slit lengths of 5, 2, and 1.2 arc-seconds. The separation of 12.5 arc-seconds was chosen because the closest approach of the first and second order spectra of a point source is 25 arc-seconds and therefore the spectra are equispaced where they most are densely packed (see figure 3.1). Figure 3.2a shows the mode A data format. In practice, when using this mode, a point source is placed in one dekker slot and the other is used for sky only. After half the total required exposure time the roles of the two slits are reversed, the idea being that differences between the two halves will cancel out and accurate sky subtraction will be simple to achieve. Although, this "beam switching" technique is popular with astronomers it cannot be recommended for the study of very faint sources where accurate sky subtraction is essential (see chapter 2.1.4).

Mode B is another beam switching mode but here the slit lengths are 25 arc-seconds - any longer and the first and second order spectra from the slit would overlap. To ensure that the spectra from the separate slits do not overlap they are placed 42.5 arc-seconds above and below the field centre (see figure 3.2b). A field lens is required to prevent vignetting along the slit direction.

Mode C - the "longslit" mode - is selected by removing the dekker slide altogether (see figure 3.2c). As in mode B a field lens is needed to ensure equal instrumental sensitivity along the slit. Because the slit is longer than 25 arc-seconds (4 arc-minutes in fact) there will be some spectral overlap

unless second order is filtered out using a GG495 field lens. Such overlap is ruinous to the second order spectrum for faint work as its faint signal is swamped by the strong night sky spectrum between 8000A and 10000A in first order. The effect of the very weak second order sky spectrum on first order is much less and is particulary negligible during dark time. When several objects are observed with a long slit it is possible that different orders of different objects will interfere with each other. In this case the observer has to decide whether second order provides useful additional information or merely contaminates the first order spectrum. Longslit mode is also useful for extended objects.

Mode D is a single slit of 25 arc-second length placed on axis to provide maximum throughput when a field lens is not used (see figure 3.2d). For single, small (less than 5 arc-second diameter) objects where the entire 4000A to 10400A wavelength range is required, mode D is ideal. As with mode B the 25 arc-second slit length ensures that there is no overlap between the two spectral orders. It also provides adequate sky information for accurate sky subtraction provided the object is small. The vignetting along the slit is small because the slit is short and can be removed entirely by the use of a field lens. However, although this makes the sensitivity along the slit uniform, the non-uniform sensitivity obtained without the field lens is higher and is recommended since the data reduction software can handle this very effectively.

3.3 Description of the data reduction package

Details of the data reduction programs, and how they work, are given in this section. In subsections 3.3.2 and 3.3.3 the programs are listed in alphabetical order.

3.3.1 Overview of the data reduction processes

The signal in each pixel in the raw data is the result of many contributing factors. For an astronomical spectrum obtained with FOS these can include;

1) The pixel's relative sensitivity (flatfield noise).
2) The total system throughput from above the atmosphere.
3) The object illumination.
4) The sky background illumination.
5) CREs.
6) Readout noise.
7) CCD defects, charge traps, charge injection, dirt etc.
8) Charge transfer and flooding.
9) Thermally generated electrons.

The aim of the data reduction software is to provide a good estimate of the number of photons (or the energy flux) received from the object as a function of wavelength, above the earth's atmosphere. As the primary function of FOS is to obtain redshifts for faint extra-galactic objects, it is more important that the software should enhance the visibility of genuine spectral features than provide an accurate photometric result. Of course, spurious features of terrestrial origin should be removed altogether.

The FOS on-line data reduction package

The full data reduction process involves calibration, background removal and an assessment of which pixels contain what sort of imformation. Some of the reduction operations are more significant than others therefore the software has been written so that the processes that only marginally improve the data's appearance can be left out if a quick look-see approach is required. It is useful to refer to all the programs that one might use whilst observing as the mainstream programs. In addition to these there are some secondary programs that are used to provide instrumental calibrations and assist with setting up the instrument. As FOS is a fixed format device, visiting observers should find FOS already setup when they arrive with all the calibration information they require, so they may not have to concern themselves with the secondary programs.

The mainstream data reduction options are shown as a flow diagram in figure 3.3. Note that the only two operations deemed essential are those of spectrum extraction (including optional background subtraction) and spectrum plotting (which includes wavelength calibration). The data frame size drops from 400x590 pixels to 590x2 pixels, for first order, or 300x2, for second order, at the extraction stage. The extracted spectra frames are two-dimensional because they each contain 2 spectra; the spectrum with background removed and the background spectrum itself. The data is also reversed in the wavelength direction so that wavelength increases with channel number in the reduced data. There are two extraction alternatives; EXTRIC for mode D data and TRIM + LEXEXT for all types of data. All other mainstream programs preserve the

Main on-line FOS data reduction options

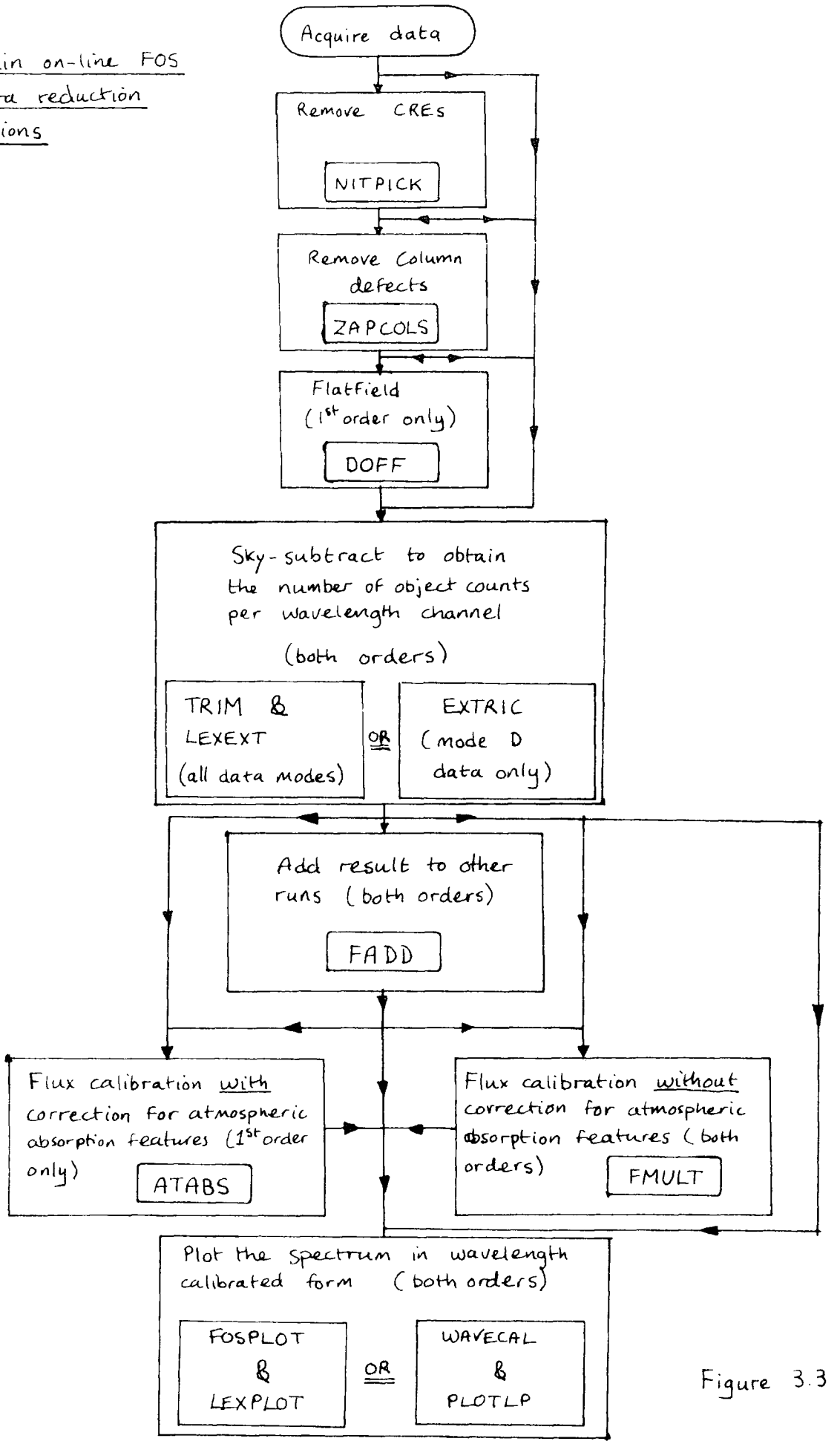


Figure 3.3

input frame size on output so that they can be left out of the reduction process if desired.

There are several types of calibration file associated with the reduction programs and these should be available by default at all times since FOS is a fixed format instrument. There are three shape-files which contain the coefficients of least-squares polynomial fits to the curvature of the first and second order spectra; FRSTCOFS.DAT for first order, SCNDCOFS.DAT for second order and ORDSEPCF.DAT for the separation between the orders. These contain, in the following order, the order of the fit (always 6), the start and end row numbers for which the fit was made and the 21 polynomial coefficients. Other calibration information includes a standard normalised flatfield image (see notes on the programs DOFF and FFNORM), and wavelength calibration data (see notes on the programs FOSPLOT and WAVECAL).

The raw FOS data frames come complete with numerous descriptors giving lots of useful data about the exposure including the exposure time, the object name, the run number, the telescope coordinates etc. So that the final reduced data can be associated with its origin and the plotted spectrum automatically shows what the plot is about, the descriptors NUMRUN (run number) and OBJECT (object name) are passed through to the output data by all the mainstream programs.

3.3.2 Description of the main stream programs

ATABS converts spectral data from detected counts per channel at the bottom of the atmosphere to relative flux at the top of the atmosphere. In other words it flux calibrates the data and corrects for atmospheric absorption. To make this conversion ATABS requires in addition to the spectrum it converts, the flux calibration spectrum for first order and the sky subtracted spectrum of a standard featureless star. ATABS is only applicable to first order spectra as there is no significant atmospheric line absorption in second order. The convertion for each channel is represented by the simple equation

$$F(\lambda) = C(\lambda).f(\lambda).A(\lambda)$$

where F(lambda) is the result, C(lambda) is the detected count for the object, f(lambda) is the fixed instrument plus atmospheric continuum correction and A(lambda) is the atmospheric absorpion line correction factor. The flux correction spectrum f(lambda) is a default calibration which was established on the Durham starlink VAX using SPICA for an observation made at a zenith-distance of 27 degrees. Objects observed at greater zenith-distances will appear slightly redder than they ought to and likewise objects observed closer to the zenith will appear slightly bluer. These discrepencies are, however, very small in first first order except for very large zenith-distances. No facilities exist at La Palma to generate f(lambda) from the observation of a flux standard.

A(lambda) is determined by ATABS from the standard star spectrum and the wavelength region over which A(lambda) is applied is specified by the user. Outside this region A(lambda) is set to unity. A(lambda) is determined by first multiplying the standard's spectrum by the flux correction function F(lambda) and then fitting a smooth polynomial curve through the result using channels known not to be affected by atmospheric absorption lines. The smooth, resultant spectrum is then divided by the spectrum it was derived from to give A(lambda). The standard is multiplied by F(lambda) initially beause the result is easier to fit a polynomial to than the characteristically bell-shaped, unfluxed spectrum. Note that A(lambda) will exhibit the shot noise of the standard star and standards should be observed so that plenty of signal is recorded. Likewise the CCD's fixed pattern noise should be removed from the standard by flatfielding for the best results. The standard has to be inherently featureless only in the wavelength region where the atmospheric absorption features are seen (6800A to 10000A). The majority of stars with spectral types earlier than G-type are therefore fairly suitable. Ideally, standards should be observed at the same time and zenith-distance as the main object of interest so that the atmosphere affects both standard and object as equally as possible.

DOFF is used to apply a normalised flatfield to a raw data image in order to remove the CCD's fixed pattern noise. Only first order flatfielding is provided as the sensitivity of the GEC CCD is so poor in second order that shot noise always

dominates. The normalised flatfield that is applied should be available as a standard calibration image but if not it can be generated using COMBINE and FFNORM. DOFF is a fairly simple program as most of the hard work is done in advance by COMBINE and FFNORM. The default normalised flatfield is the result of adding together many flatfield exposures to reduce shot noise, removing the bias and normalising along the wavelength direction so that the gross counts versus channel-number relation is flat with a mean of 20000 counts. Also the pixels which do not receive adequate illumination are given a value of exactly 20000 counts. To apply this flatfield to a data frame DOFF uses the equation

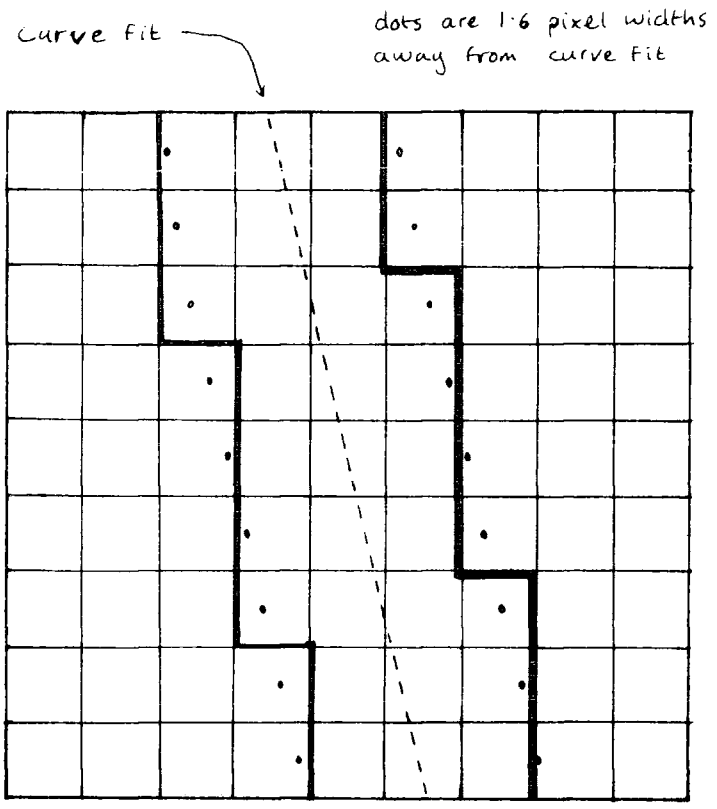
Output pix value = (input pix value - bias).20000/f + bias

where f is the flatfield pixel value. Notice that f is largest for the most sensitive pixels. DOFF uses the median of the input frame's last 2 overscan lines as the bias value because, unlike the mean, it is totally insensitive to bright columns which flood charge into the overscan lines. DOFF puts the bias back on after the flatfield has been applied so that the lexidata display parameters can be left unaltered when comparing the raw data with the flatfielded data. The presence of the bias does not have a detrimental effect on any subsequent processing.

EXTRIC is one of the two extraction plus background subtraction options available. The program is called EXTRIC (short for extricate) because a program called "extract" already existed on the PE 3220 before FOS arrived and also

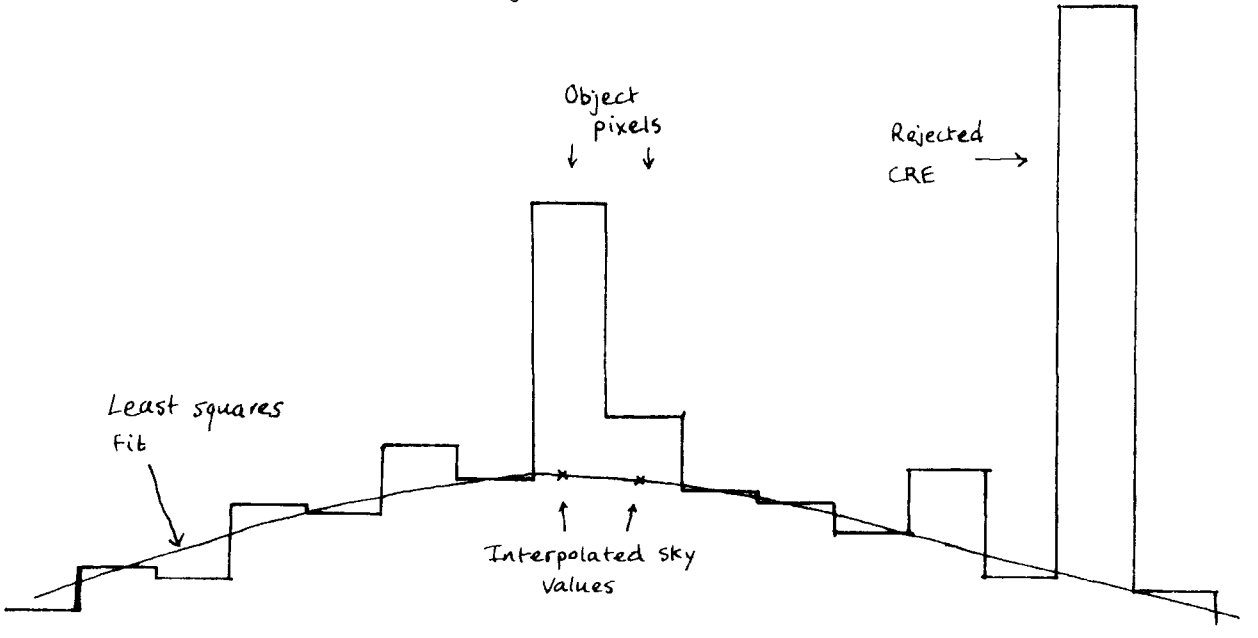
because it pulls the data out of a tricky spot. EXTRIC was written specifically with mode D data in mind and its use with other data modes, although possible, is not recommended. EXTRIC automatically determines the object and sky areas of the data without the need to use the lexidata. Only one spectrum is extracted at a time so if both orders are required from the data, EXTRIC has to be run twice.

After being prompted for the name of the file to be processed and the spectral order to be extracted the user can either let EXTRIC determine the positions of the object and sky spectra automatically or specify these areas himself by giving one point on the object spectrum and its width. When automatic area selection is adopted, curved windows of the same shape as the spectra are set up using the shape-file polynomial coefficients stored on disk. The specific details of these windows depends on which order is being extracted and whether the program is searching for the sky spectrum or the object spectrum. These curved windows are scanned along the slit direction to find the position for which the integrated signal in the window is maximum. When searching for the sky spectrum the window is moved a full pixel at a time and the window's width is 20 pixels (25 arc-seconds). Once the sky spectrum has been located, a narrower window is moved about within the sky area to determine the position of the object. This window has a width supplied by the user and scans for the signal maximum with a resolution of a tenth of pixel i.e the curve which is used to generate the window is moved by a tenth of a pixel between integrations. Figure 3.4 shows exactly how such windows are defined by including pixels with centres within a



EXTRIC window definition

Figure 3.4



Interpolative sky subtraction

Figure 3.5

specified distance of the curved centre line. This means that the window only changes in some of the channels when the centre curve is moved by a tenth of a pixel. Once the object signal has been maximised the program defines the window which gave this maximum as the object data area and the maximised sky window, without the object area, as the sky signal area.

Once the sky and object regions have been established the data in each channel can be extracted independantly to give the sky subtracted object spectrum and the sky spectrum with the bias removed. EXTRIC uses interpolative sky subtraction which accurately deals with non-uniform sky illumination along the slit whilst ignoring CREs in the sky areas. In this method, the sky background "under" the object signal for a single channel is estimated by fitting a parabola to the sky data that lies either side of the object signal (see figure 3.5). To ensure that anomolous sky pixel values are ignored (e.g. because of a CRE) the residual from the fitted curve of each point is calculated and if it is greater than a certain threshold the point is rejected and a new curve is fitted with the remaining points. The curve fitting procedure is continued until none of the pixel values used to make the fit is rejected. The polynomial can then be used to give estimates of the sky background for the pixels in the object area. The resultant value for the object spectrum in a channel is simply the sum of all the object area pixels with their individual sky estimates subtracted off first. The output sky spectrum is the average sky estimate (i.e. the sum of all the sky estimates in the object area divided by the number of pixels in the object area for each channel) with the image bias removed. The image

bias adopted is the median value of the data in row 590.

FADD is used to add two, 4 byte, real, BDFs together. The BDFs should have the same dimensions and the output BDF will also be the same size. FADD was written for the specific task of adding together reduced spectra of the same object from different runs. The first input BDF specified has the object name descriptor passed through to the output BDF and the run number descriptor of the output BDF is always set to "Sum of several runs".

FMULT is similar to FADD except that the result is the product of the input files and the run number and object name descriptors of the first (object) BDF are passed through to the output BDF. The second input should be a calibrating spectrum e.g. a flux calibration that converts from counts to relative flux per channel. FMULT can be used as an alternative way of flux calibrating a spectrum without correction for atmospheric absorption lines. As ATABS is only used for first order, FMULT is the only way of flux calibrating second order spectra.

FOSPLOT is an interactive plotting program for displaying FOS spectra graphically, on the lexidata, in the conventional form of intensity versus wavelength. Figure 3.6 shows an example plot actually obtained whilst obervations were being made. FOSPLOT is actually a version of the La Palma software group's program PLOT modified by Dr Alan Purvis. FOSPLOT gives a wavelength calibrated plot which displays the run number and object name descriptors. It also allows the user to display

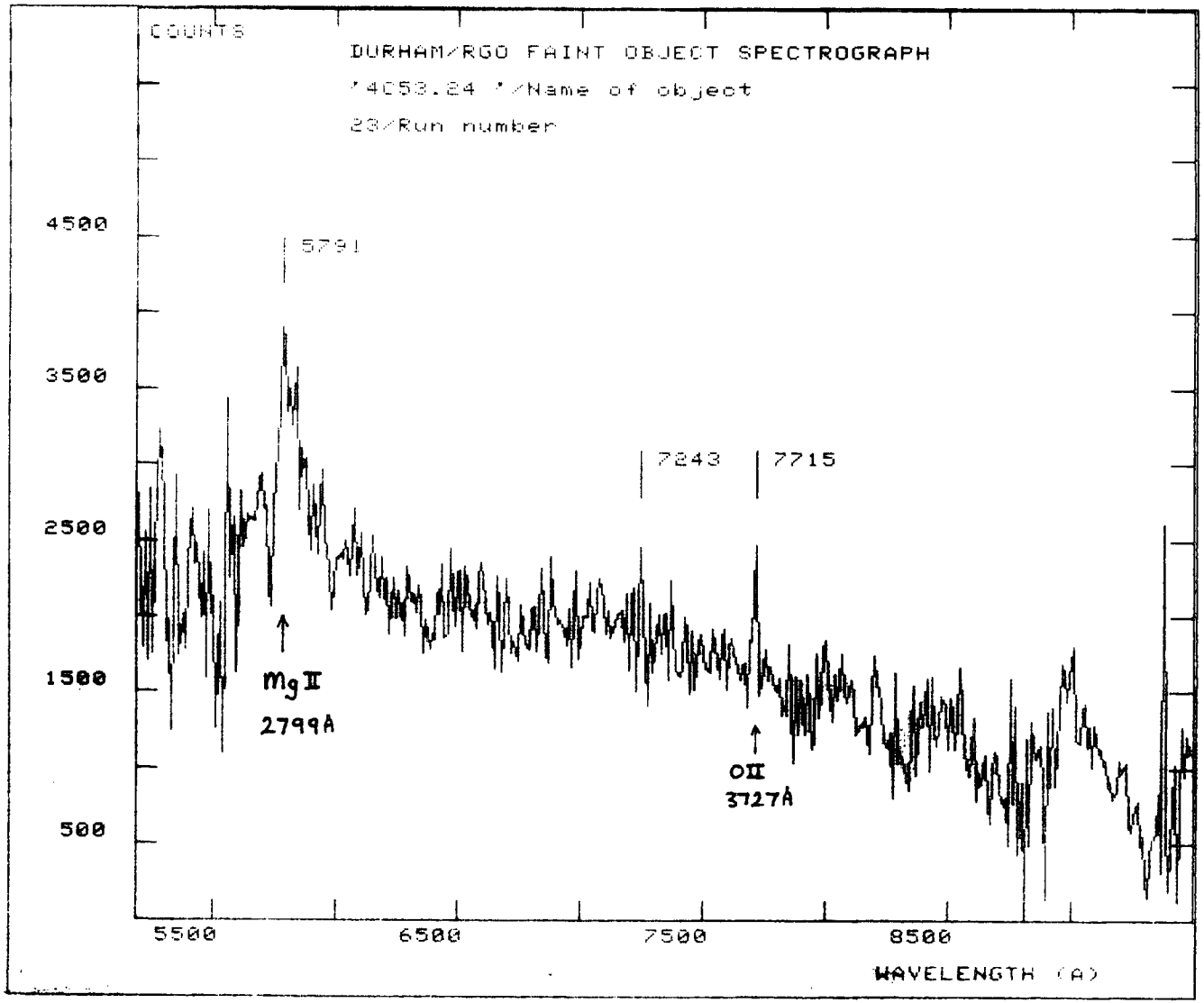

COUNTS
DURHAM/RGO FAINT OBJECT SPECTROGRAPH
'4C53.24 '/Name of object
23/Run number
4500
3500
2500
1500
500
5791
7243
7715
MgII
2799Å
OII
3727Å
5500
6500
7500
8500
WAVELENGTH (A)

Figure 3.6

comments and label features with their observed wavelength. Once a suitable plot has been created a hardcopy can be obtained on the printronix printer using the utility program LEXPLOT.

LEXEXT is used in conjunction with TRIM to obtain extracted and sky-subtracted spectra. LEXEXT allows the user to interactively define sky and object areas on a straight spectrum using the lexidata joystick module. The BDF input file must of course be displayed (with zooming and panning to taste) before LEXEXT is run. LEXEXT is a very general purpose program and can not only handle all modes of FOS data after they've been straightened by TRIM but also cope with data directly from the IDS when used with the CCD. If a background area is defined by the user it is appropriately subtracted from the object area spectrum, the subtraction being interpolative if possible. It is also possible to use LEXEXT without the lexidata, if for some reason it is unusable, but this method of operation is not the prefered one.

When using the lexidata the user must use the joystick module to specify two columns which define the object area inclusively and upto two columns to define the background area(s). In order to see exactly where it is best to place these lines the display parameters must be chosen to give good object to sky contrast and the image should be zoomed up using, for example, LEXPAN. When the lexidata is unavailable the object and background columns must be input numerically. As an aid to choosing these numbers correctly LEXEXT determines the median of several channel which it assumes is the background

for those channels. It then looks for columns that are more than a user specified threshold over the background and lists them as contigious blocks along with their strength relative to the background.

When no background columns are given by the user (or zero column numbers are given) the object area has only the bias level removed. When one background column is defined the area from it (inclusive) to the nearest object column (exclusive) is used for background subtraction. When two background columns are defined that form an area separate from the object area, the area inclusive of the background columns is used for background subtraction. Finally when two background columns are defined which lie either side of the object area the area between and including the background columns except for the object area is used for background subtraction. In this case, therefore, we have two separate background areas and if the total width of the background area is at least 5 columns then interpolative background subtraction will be use exactly as employed by EXTRIC. In all other cases where a background spectrum is subtracted off, the mean background spectrum for a single column is subtracted from each object column before they are added up to give the object spectrum. Note that the bias is not removed from the background spectrum until after it has been subtracted off the object area. In this way the object spectrum that results does not depend on the bias being determined accurately.

NITPICK compares two data files and removes CREs in one by replacing them with the corresponding pixel values of the other. The two images should be as alike as possible in terms of the exposure parameters (e.g. slit width, sky brightness, object position on slit etc.) but not necessarily of the same exposure time. Each pixel in the image to be cleaned is compared with its corresponding pixel in the comparison image. The expected rms error (sigma) associated with the comparison image pixel is determined and if the other image pixel is more than N sigma (where N is specified by the user) greater than the comparison pixel value the output pixel is taken from the comparison frame, otherwise it is taken from the frame that is being cleaned. A map image is also output which shows where the anomolously high pixel values where found. In this map image the swapped pixels are given a value of 1 and the clean pixels a value of 0. To clean two images one simply applies NITPICK twice with the roles of the two input images being reversed for the second run.

TRIM is used to straighten out FOS spectra and remove the uninteresting areas of data. Both spectral orders can be trimmed out with one call to TRIM. The user specifies which spectral order(s) he wishes to trim out and specifies two points to define the region to be trimmed out for each order required. These points are specified using the lexidata display and joystick module and the image being trimmed must be displayed (without zooming) before TRIM is employed. TRIM draws a polynomial through each point specified using the appropriate shapefile coefficients and the region in between

the curves is written to another file. The curves drawn can only be moved along the slit direction. For first order the region trimmed out has 590 rows whereas the second order output has only 300 rows. The width of the output files depends on the positions specified by the user. TRIM does not resample the data as it is undersampled to start with. Instead, TRIM uses integer pixel shifts to straighten out the curved regions and consequently a straightened point source spectrum will have a rather jagged appearance.

There is also an automatic mode of operation in which TRIM determines the positions of dekker edges for itself. To do this it chooses a channel where a bright sky line is expected and it passes a median filter along it. The use of a median filter of 20 pixels width removes features smaller than 10 pixels width and is well suited to finding the sky spectrum in FOS data. The channel chosen is number 350 which corresponds to a wavelength of 7332A and is therefore unaffected by second order which only affects first order longward of 8000A. The filtered channel will have a minimum value corresponding to the CCD bias and a peak corresponding to the sky level. By passing along this filtered channel and looking for jumps between these two levels TRIM determines the dekker edges for that channel. If there are only two such points TRIM simply uses them to trim the data. If there are four dekker edges (e.g. with mode B data) the extreme end ones are used. If there are more than four edges found TRIM plays safe and trims a very large region to ensure that all the illuminated data area is included. The dekker search is always conducted using channel 350 in first order, even if second order only is required, because in

automatic mode the second order area is chosen on the basis of the position of the first order area.

ZAPCOLS is used to cosmetically remove column defects from CCD images. The user is prompted for the start and end column numbers of a region to be "zapped" and this process continues until a negative column number is entered. When a region is zapped it is replaced with columns which are the mean of the two adjacent columns. If the region contains one edge of the data frame the region is replaced with the bordering data column and if both edges of the data frame are specified by the user the entire frame is "zapped" with the value zero.

ZED provides a table of spectral features, including their rest and observed wavelengths on the basis of a redshift supplied by the user. There are three feature files available on disk which contain lists of common spectral features and their rest wavelengths. ZED picks up one of these files, prompts the user for a redshift and outputs the feature file complete with predicted, observed redshifts. The three files available are GAL which contains typical absorption lines found in galaxies, LZQSO which contain emmission features found in low redshift active galaxies and HZQSO which contains features found in high redshift quasars. ZED is used after FOSPLOT to test out possible redshift values by trial and error.

3.3.3 Description of the secondary programs

COMBINE takes several input images (which must all be the same size and of two byte integer format) and outputs the mean image. Its main purpose is for combining raw flatfields to produce a flatfield with reduced shot noise. Ideally, the final shot noise should be significantly less than the CCD's fixed pattern noise even for wavelengths where few counts are registered. Upto 5 input images are allowed. If more images are to be combined they should be combined as groups of equal number and the results finally combined into a single image afterwards. No data descriptors are passed to the output image.

CONVERT creates a 2 byte integer BDF from a 4 byte real BDF so that it can be written to a FITS tape. CONVERT was written so that the astronomer can take his reduced data home (FITS does not support 4 byte real data). The real data has to be scaled into the range -32,000 to 32,000. The scaling is described by the equation

$$P(out) = P(in) \cdot 32000 / Q$$

where P(out) is the output integer pixel value (nearest integer), P(in) is the real input pixel value and Q is the maximum pixel value (ignoring sign) in the input BDF. The standard FITS descriptors BSCALE = 32000/Q and BZERO = 0 are also written to the output file. BZERO is set to zero so that the converted data can be plotted without it showing a false zero level.

DCSUB is used to subtract off the DC offset bias from an image. The median pixel value in the last two overscan lines (rows 589 and 590) is used as the bias level because, unlike the mean, it is very insensitive to spurious effects such as charge flooding in defective columns. The output image is, of course, the input image with every pixel value reduced by the bias value. The most important use of DCSUB is to remove the bias from a flatfield before it is normalised by FFNORM.

FFNORM normalises flatfields. Each pixel value in the normalised flatfield indicates the sensitivity of the pixel with respect to all the other pixels for the wavelength that it receives in FOS. Thus it contains no information regarding the variation of sensitivity of each pixel with wavelength due to the spectrum of the tungten lamp, the throughput of the FOS optics and the intrinsic properties of the CCD itself. The averagely sensitive pixel will have a value of 20000 counts whereas more sensitive pixels will have larger values and less sensitive one smaller values. Normalisation occurs in the wavelength direction only; flatness along the slit has to be ensured by using an integrating sphere and a fieldlens. If the profile along the slit is non-uniform application of the flatfield will still remove the short scale fixed pattern noise and therefore enhance the definition of spectral features, but can also alter the total number of counts in the final spectrum. This is only important for spectrophotometric work.

When used it is necessary to specify where the edges of the fully illuminated region are by providing two points. The first order shapefile FRSTCOFS.DAT provides the information for FFNORM to define the illuminated and non-illuminated areas from these two points. The non-illuminated areas are set to 20000 counts exactly. An integrated spectrum of the illuminated area is derived and heavily smoothed to remove the short scale structure. The illuminated area of the normalised flatfield is then obtained for each pixel using the relation

$$P(out) = 20000 \, . \, N \, . \, P(in) \, / \, S$$

where P(out) is the result, P(in) is the pixel value in the raw flatfield, N is the width of the illuminated area in pixels and S is the value of the smoothed integrated spectrum for the same channel (row) as the input and output pixels. Notice that P(in) does not have the bias removed from it by FFNORM and this should therefore be done using DCSUB before applying FFNORM.

FOSCOFS is used to generate the polynomial shapefiles that describe the spectral curvature exhibited by FOS. It does this using an observation of a bright star. The curvature of the spectrum depends slightly on the zenith-distance of the star and the orientation of the slit. Firstly FOSCOFFS locates first order in the rows where there is no second order spectrum (shortward of 8000A) and then follows it to both longer and shorter wavelengths, establishing centroids of the spectrum as it goes. A sixth order least squares fit polynomial is then fitted through these centroid points and the coefficients are

stored on disk in FRSTCOFS.DAT. The whole process is repeated in the region to the side of the first order spectrum, where the second order spectrum is expected, and the second order polynomial coefficients are written to the file SCNDCOFS.DAT. Finally the separation between the spectral orders is fitted with a polynomial and the coefficients written to the file ORDSEPCF.DAT.

A spectrum is found by locating the highest pixel value in each row and looking for a number of these maxima (specified by the user) that lie contigiously. The spectrum is then followed by looking in the 7 nearest pixels in the next row and centroiding on the brighest pixel and the two either side. The background is subtracted off first to improve the accuracy of the centroiding. If

a) the maximum pixel value is greater than twice the background noise
b) the maximum pixel value is within a factor of ten of the last maximum value
c) the total integrated signal is non-zero
d) the centroid position is within two pixels of the last one

the centroid is deemed acceptable. Otherwise, the last ten centroid values are used to extrapolate the centroid in the row. Testing for these conditions helps prevent FOSCOFFS from going wrong due to CREs, absorption lines and the spectrograph's wavelength cutoffs. In particular, because the areas where the spectra fade away are assigned extrapolated centroids, the ends of the spectra are fitted accurately by the polynomial and ringing is reduced.

RESTCOFS restores the default shapefiles should some recently updated versions created by FOSCOFS be no longer required. It simply takes the contents of the files FRSTCOFS.BAK, SCNDCOFS.BAK and ORDSEPCF.BAK and puts them in the files FRSTCOFS.DAT, SCNDCOFS.DAT and ORDSEPCF.DAT, overwriting their previous contents. The default shapefiles were created using FOSCOFS on the combined spectra (the program COMBINE was used) of an M star and a white dwarf observed with a pinhole slit. This produced a very good result as there was plenty of signal at both the blue and red extremes of the spectrum. The use of a pinhole slit on objects near the zenith also means that the default shapefiles give the spectral shape in the absense of atmospheric cross dispersion.

SLICE enables the user to obtain centroids, profiles and line widths of emission lines in the raw data. Its main use is to assist with the focusing of FOS by giving precise line widths and Hartmann shifts. The position of the centre of a box is prompted for along with the box dimensions. The box must not be more than 25 by 25 pixels in size and will have odd numbered dimensions equal to twice the half-size given by the user plus one. Profiles can be obtained in either the vertical or the horizontal directions. The rows or columns at right angles to the slice direction are averaged to give a one dimensional profile and the 5 pixels centred on the peak are centroided to give the line position and width. The areas outside these five pixels are used to determine the background and so suitable box dimensions should be chosen to ensure that they are not contaminated by other lines. To determine

precisely where the box should be put, the program PHOTOM can be used and SLICE will use the cursor coordinates given to ADAM as the default when it prompts the user.

VIEW prints the pixel values of a 20 by 30 area from a two byte integer BDF. The user has to supply the coordinates of the centre of this area. VIEW is useful when the exact numbers are required, for example, to check that the CCD electonics has all its data bits working. The data is automatically printed on the printronix printer when VIEW is run.

WAVECAL writes standard FITS wavelength calibration descriptors to a reduced FOS spectrum so that it can be plotted with a wavelength scale by PLOTLP. The wavelength calibration data is obtained from a file whose name is supplied by the user. The three parameters which define the linear calibration are CRPIX (the number of a pixel with a known wavelength), CRVAL (the wavelength of pixel number CRPIX in metres) and CDELT (the dispersion in wavelength per pixel). Thus for FOS's first order one might have a file which contains the values

0 4746E-10 10.7305E-10

Default wavelength calibration files are provided but the user can create his own using SLICE on an arc spectrum (remember that wavelength increases with channel number in the reduced data but decreases with row number in the raw data).

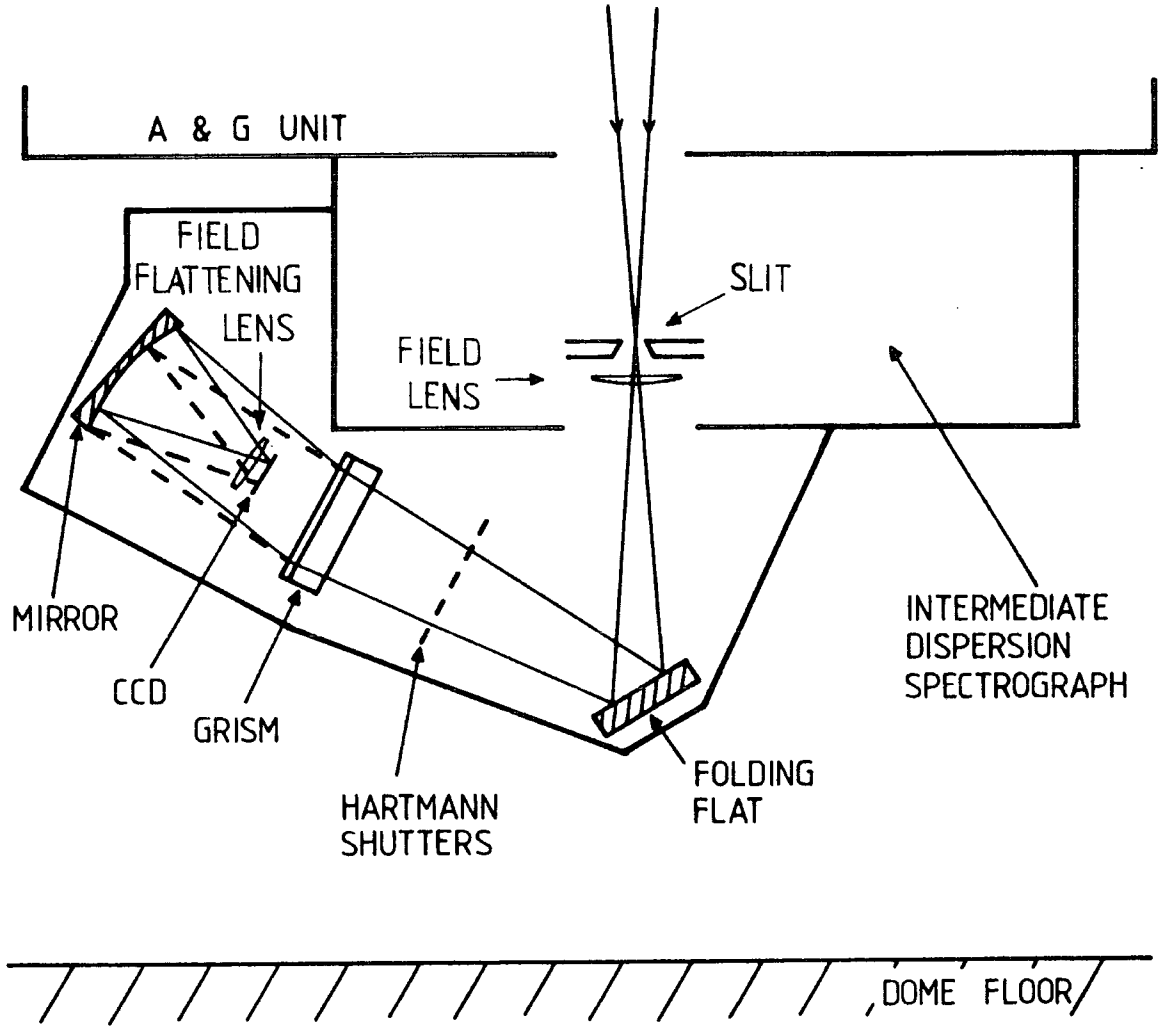
Chapter 4

The 2.5m Faint Object Spectrograph

4.1 Description

The faint object spectrograph (FOS) was built as collaboration between the University of Durham and the Royal Greenwich Observatory (RGO). It was installed on the 2.5 metre Isaac Newton Telescope (INT) on La Palma in July 1984 and subsequently commissioned during three observing runs in July 1984, December 1984 and May 1985. It is a common-user instrument and is now in regular use. FOS is a fixed format, highly efficient, low dispersion spectrograph offering a large wavelength coverage from 4000A to 10,400A in a single exposure. The large wavelength range is achieved by utilising both first and second order of FOS's transmission grating, the orders being separated by a cross-dispersing prism. A CCD detector (GEC P8603 type) is used because of the high quantum efficiency offered. The optical design is due to Charles Wynne (see Wynne 1982) who originally conceived the instrument for the 4.2 metre telescope on La Palma which will be commissioned in 1987. His design created much interest and it was decided that a prototype should be built for the INT. For a comprehensive description of FOS see Breare et al 1985, 1986b, Martin 1986 and Waltham 1986.

The overall optical layout of FOS is shown in figure 4.1 and the camera and cryostat are shown in figure 4.2. To match the f/15 seeing image to the GEC pixel size, a slit-to-detector reduction factor of about 10 is required. In other words, a



FOS OPTICS

FIG 4.1

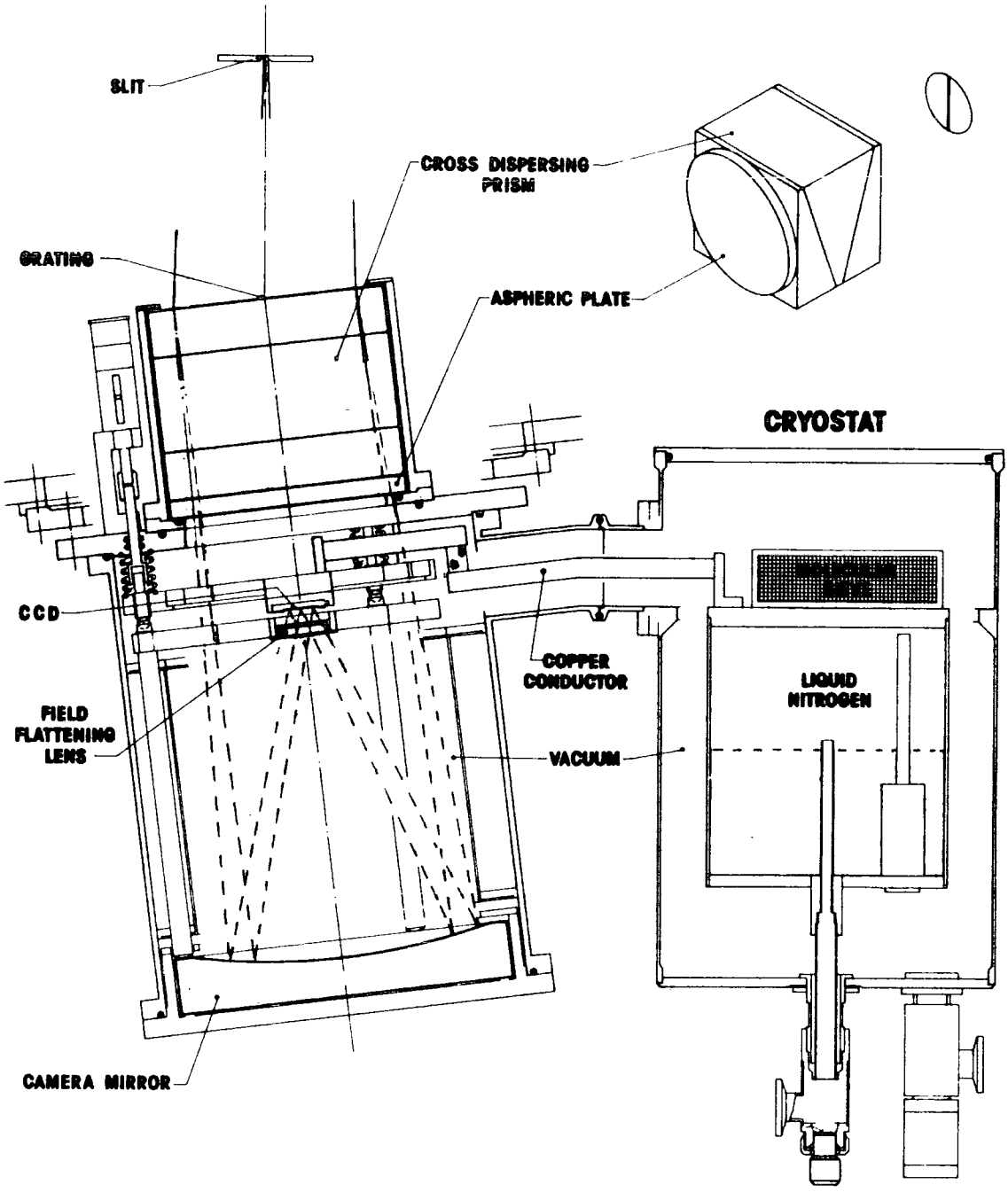


FIG. 4.2 Camera and cryostat

fast camera with an f-number of about f/1.5 is needed. FOS employs an f/1.4 Schmidt camera with the CCD held at the focus by a low profile "spider". The small size of the CCD conveniently allows it to be placed within the shadow of the telescope secondary and so very little vignetting is introduced. The number of air-glass surfaces has been kept to a minimum by combining the grating, the cross-dispersing prism and the aspheric corrector into a single composite element. The use of a transmission grating is also advantageous at FOS's low dispersion because transmission gratings are slightly more efficient than reflection gratings and because it allows the camera optics to be placed very close to the grating. Furthermore, Wynne has ingeniously dispensed with the need for a collimator by correcting the aberations, introduced by placing the grating in a diverging beam, in the subsequent optical elements. However, because there is no collimator, a field lens (optional) has to be used to prevent off-axis vignetting and a folding mirror has to be used because the observing floor prevents FOS from being used in the straight-through position.

It is, of course, necessary to cool the CCD to 150K to eliminate the generation of thermal electrons. This is achieved by evacuating the camera and mounting the CCD on a cold copper block which is linked by a copper finger to a liquid nitrogen cryostat.

The CCD spider is mounted on kinematic seats to allow precise focusing across the chip and a manually operated rotation mechanism allows the CCD's pixel grid to be aligned

with the grism rulings. Rotational alignment of the CCD with respect to the slit is achieved by rotating the entire camera body. The slit, shutter, above and below slit filter slides, the dekker slide and the calibration lamps used by FOS belong to the intermediate dispersion spectrograph (IDS). FOS has its own Hartmann shutters for focusing. The entire FOS optical train is rigidly held together by a 3mm steel plate framework which obtains increased stiffness from the IDS to which it is mounted. Image shifts on the CCD due to flexure are barely detectable.

4.2 Performance

The most important figures of merit for FOS are its throughput and resolution. Via the throughput we can predict the likely signal-to-noise performance.

4.2.1 Resolution

The resolution of FOS in first order is 13A FWHM when a narrow slit width is used and is only degraded slightly to 15A by opening the slitwidth to 2 arc-seconds. This is because the resolution is dominated by the imaging properties of the unconventional optical system which is optimised to give a point spread function consistent with the CCD pixel size at all wavelengths and positions along the slit. In practice the resolution varies between 15A to 20A depending on how features are sampled by the 10.73A pixels (FOS is slightly undersampled). For second order the dispersions and resolutions are approximately twice that of first order.

4.2.2 Throughput

Figure 4.3 shows the measured and predicted throughput values for FOS. The throughput shown is the total system throughput, i.e. it is due to the combined effects of the continuum (ignoring absorption lines) atmospheric losses at a zenith-distance of 27 degrees, losses in the telescope, losses in the FOS optics and the efficiency of the detector.

The experimentally determined curve was derived as follows. A flux standard was observed with a very wide slit width to ensure that no light was lost at the slit. The standard used was VMa 2, without a field lens, for an exposure time of 30 seconds. The flux distribution of this star has been accurately determined (Oke 1974) from 3260A to 10640A. The observed spectrum was sky-subtracted making sure that all the observed counts were co-added for each channel. The data was wavelength calibrated and then input to the SPICA routine STANDARD (SPICA is a general purpose spectroscopic data reduction package available on STARLINK) which established the relation between the observed counts and the flux in each channel. STANDARD does this by deriving the flux spectrum from Oke's measurements which are stored in permanent STARLINK files. The "fluxing" function generated was then applied to the observed spectrum to produce a spectrum for which the units were ergs/sec/cm^2/channel. The number of photons from VMa 2 passing through an area, A (=49087 cm^2 for the INT), in a time, t=30 sec, at the top of the atmosphere was calculated by multiplying every channel in the flux spectrum by Atλ/hc where λ is the wavelength of the channel (in cms),

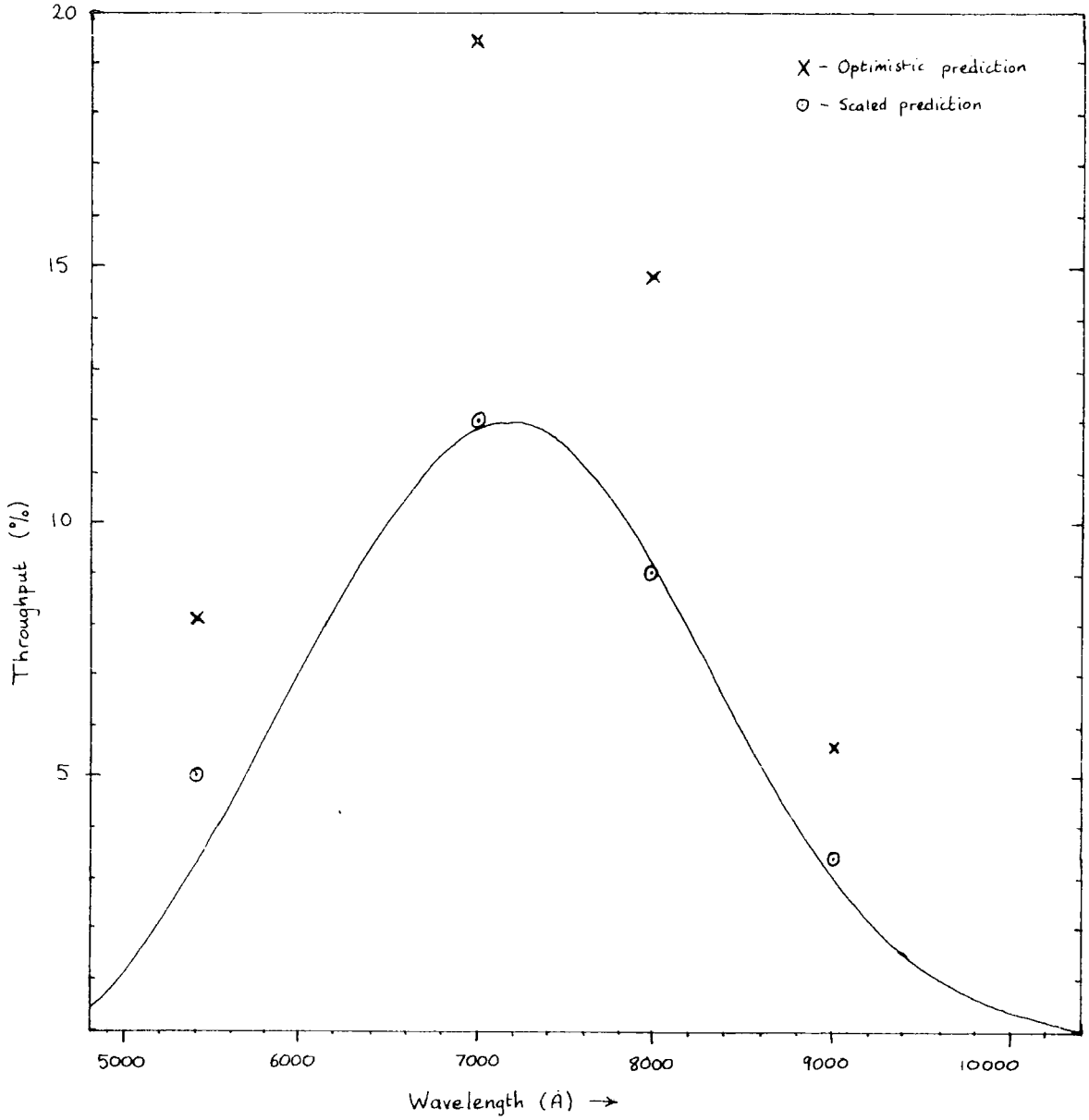


Figure 4·3 (a) FOS system throughput, Measured and predicted, for FOS + INT + atmosphere at zenith-distance of 27°. First order.

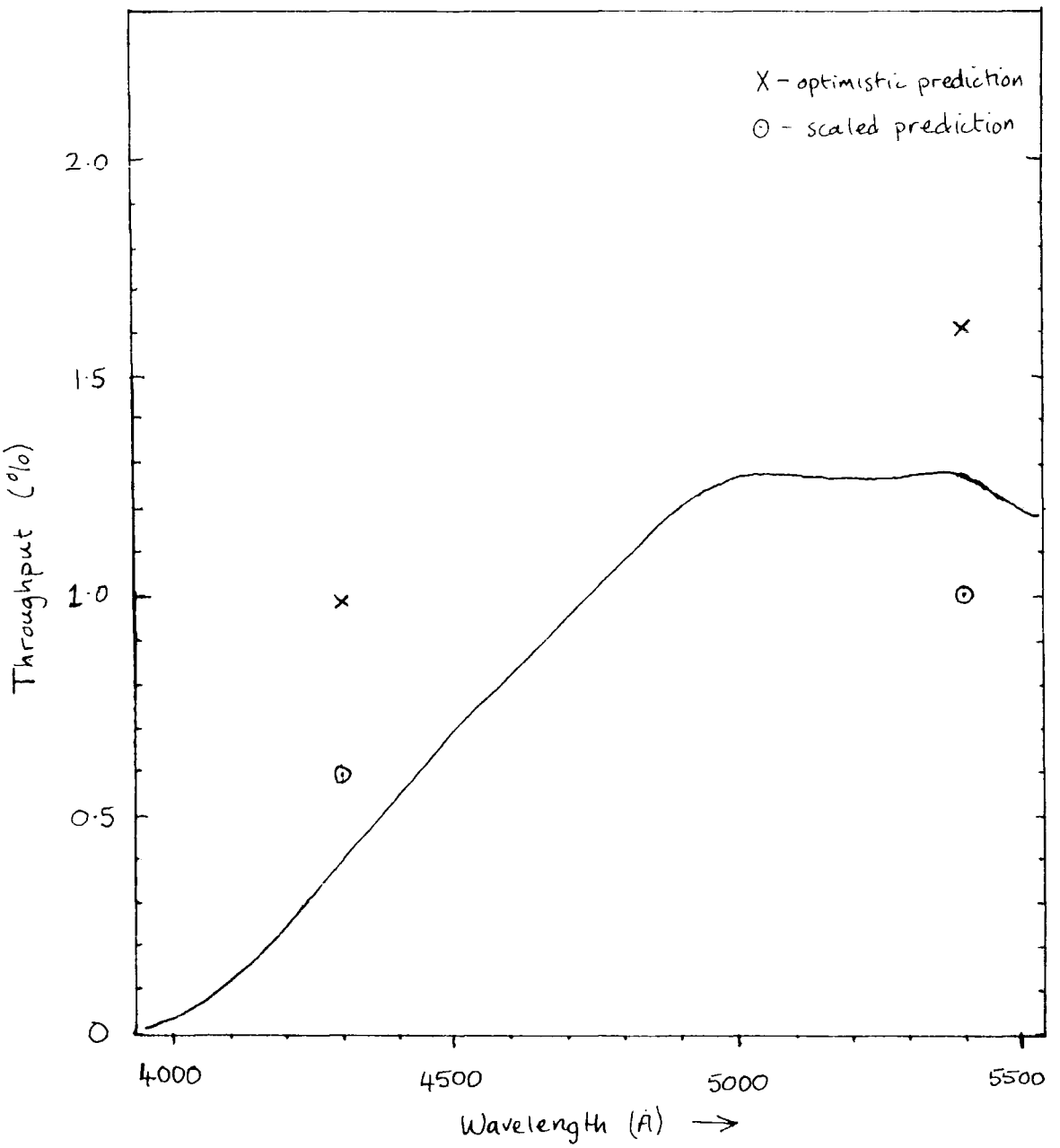


Figure 4·3 (b) FOS system throughput, measured and predicted, for FOS + INT + atmosphere at zenith-distance of 27°.

2nd order.

$c=3.0 \times 10^{10}$ cms/sec is the speed of light and $h=6.6 \times 10^{-27}$ erg seconds is Planck's constant. Smooth polynomials were then fitted to both the observed counts and the photon counts above the atmosphere and the throughput curve was obtained by dividing the former by the latter. Note that for FOS each digitisation unit corresponds almost exactly to one photo-electron. A similar procedure was followed for second order using a 60 second exposure on W1346 at a zenith distance of 3 degrees.

Table 4.1 shows the predicted values of the throughput at several different wavelengths. The results are consistently larger than the measured values by a factor of 1.63 which presumably indicates that at least some of the figures in table 4.1 are somewhat optimistic*. However, the relative variation of throughput with wavelength seems correct. The FOS optical efficiency (without the grating) of 0.8 is based on 2 reflections of 95% efficiency and 4 air-glass surfaces of 97% efficiency. Using 90% and 95% instead for these efficiencies, respectively, gives an optical efficiency of 0.66. Furthermore, I have not accounted for absorption in the transmitting elements or vignetting, but these should be small. Another possibility is that the CCD has below average quantum efficiency (a variation is sometimes seen in devices from the same manufacturer). Finally, a blaze efficiency of 67% in conjunction an optical efficiency of 66% would explain the factor of 1.63. In view of the uncertainties, the theoretical and measured efficiencies are in reasonable agreement.

* On the other hand it is more likely that the measured values are below average because the observations were made through high altitude dust from the Sahara desert with a primary mirror that had not been aluminised for 18 months.

Table 4.1 - predicted FOS efficiency

Order	wavelength (A)	efficiency atmosphere	efficiency telescope	efficiency grating	efficiency optics	efficiency CCD	total efficiency	0.615xtotal efficiency
2	4300	0.745	0.7	0.477	0.799	0.05	0.01	0.0061
2	5400	0.85	0.7	0.126	0.799	0.27	0.016	0.01
1	5400	0.83	0.7	0.648	0.799	0.27	0.081	0.050
1	7000	0.90	0.7	0.900	0.799	0.43	0.195	0.120
1	8000	0.92	0.7	0.837	0.799	0.34	0.147	0.090
1	9000	0.93	0.7	0.765	0.799	0.14	0.056	0.034

4.2.3 Signal-to-noise and exposure times for various magnitudes

Figure 4.4 shows the signal-to-noise obtained in 10.73A intervals at 4300A, 5400A, 7000A and 8000A, as a function of magnitude for a 1000 second exposure. The curve for 4300A corresponds to the sum of 2 adjacent channels in second order. The signal-to-noise is calculated from

$$\frac{S}{N} = \frac{Q}{\sqrt{Q + 6(S+R^2)}}$$

where Q is the total number of object counts detected, S is the number of sky counts per pixel and R is the rms readout noise. This expression is obtained by assuming that 4 pixels are co-added to get the object signal and 8 sky pixels are used to determine the sky background for sky-subtraction (see section 2.2.1). Q was determined from the observed throughput, $T(\lambda)$ given in figure 4.3 using the formula

$$Q = \int_\lambda A\, t\, \Delta\lambda\, T(\lambda) \frac{\lambda}{hc} = 2.66 \times 10^{16} \int_\lambda T(\lambda)\, \lambda(\text{Å})$$

where f_λ is the flux at a wavelength of λ in ergs/sec/cm^2/A. f_λ is given by

$$f_\lambda = K \cdot 10^{-(0.4\,m + 8.43)}$$

where m is the magnitude at a given wavelength and K is the ratio (f_λ / f_{5500}) for an A0 star (the magnitude of an A0 star is constant at all wavelengths by definition). The values used are shown in table 4.2. The values of S given are "average" values obtained from a moonless night's spectrum obtained using a 1 arc-second slit. A readout noise of R=7 e^- rms has been used throughout. It can be seen that absorption lines (which can usually be identified with a signal-to-noise ratio of about

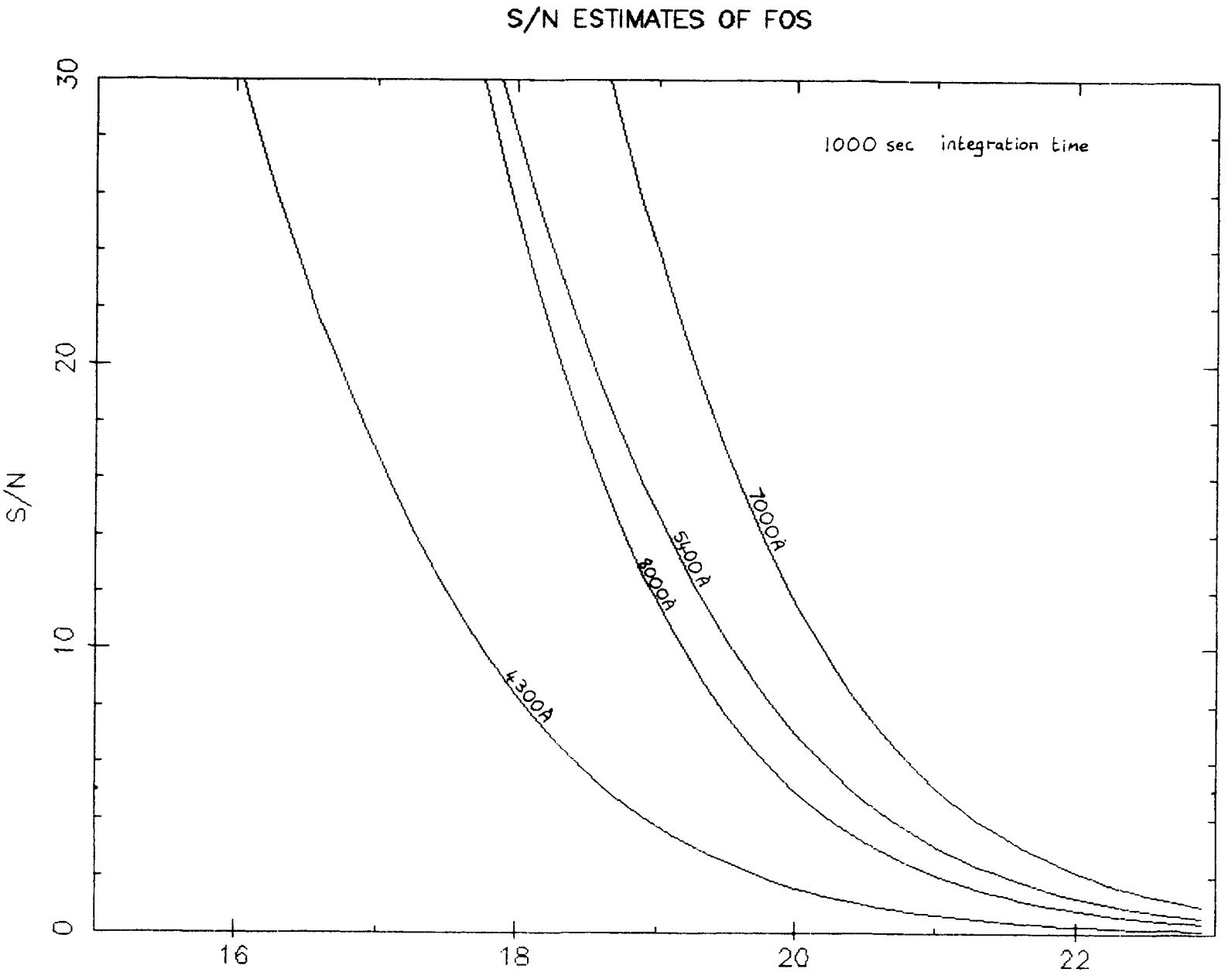

S/N ESTIMATES OF FOS
1000 sec integration time
4300Å
5400Å
8000Å
7000Å
S/N
M
0
10
20
30
16
18
20
22

FIG 4.4

Table 4.2
performance curve parameters

wavelength (A)	throughput	K	S photons in 1000 secs	r_s photons/sec
4300	0.004	1.773	4	-
5400	0.033	1.0	25	0.025
7000	0.120	0.511	100	0.1
8000	0.90	0.34	300	0.3

15) will be seen for objects brighter than about 19th magnitude and emission lines (with a signal-to-noise of 5) for objects brighter than about m=20.5 in a 1000 second exposure.

It is also useful to plot the integration time required to give a certain signal-to-noise, against magnitude. Plots for signal-to-noise ratios of 5 and 15 are shown in figure 4.5. From the signal-to-noise expression given above we get

$$\frac{S}{N} = C = \frac{r_{obj}\, t}{\sqrt{r_{obj} t + 6\,(r_s t + R^2)}}$$

where r_{obj} is the total rate of photons detected from the object and r_s is the rate per pixel from the sky. This expression can be rearranged to give a quadratic in exposure time t

$$r_{obj}^2\, t^2 - C^2 (r_{obj} + 6 r_s) t - 6 C^2 R^2 = 0$$

r_{obj} is given by

$$r_{obj} = 2.66 \times 10^{13}\; f_\lambda\, T(\lambda)\, \lambda(\text{Å})$$

and r_s is given in table 4.2. The curves in figure 4.5 were simply obtained by solving the above quadratic for t. It is interesting to note that the readout noise for FOS is significant at 5400A and 7000A for integration times upto about 10000 seconds (indicated by a steepening of the curve when completely sky-limited). The plots indicate that it should be possible to obtain absorption line redshifts upto 21st magnitude, and for emission line objects upto a magnitude of 22.5, in 10000 seconds.

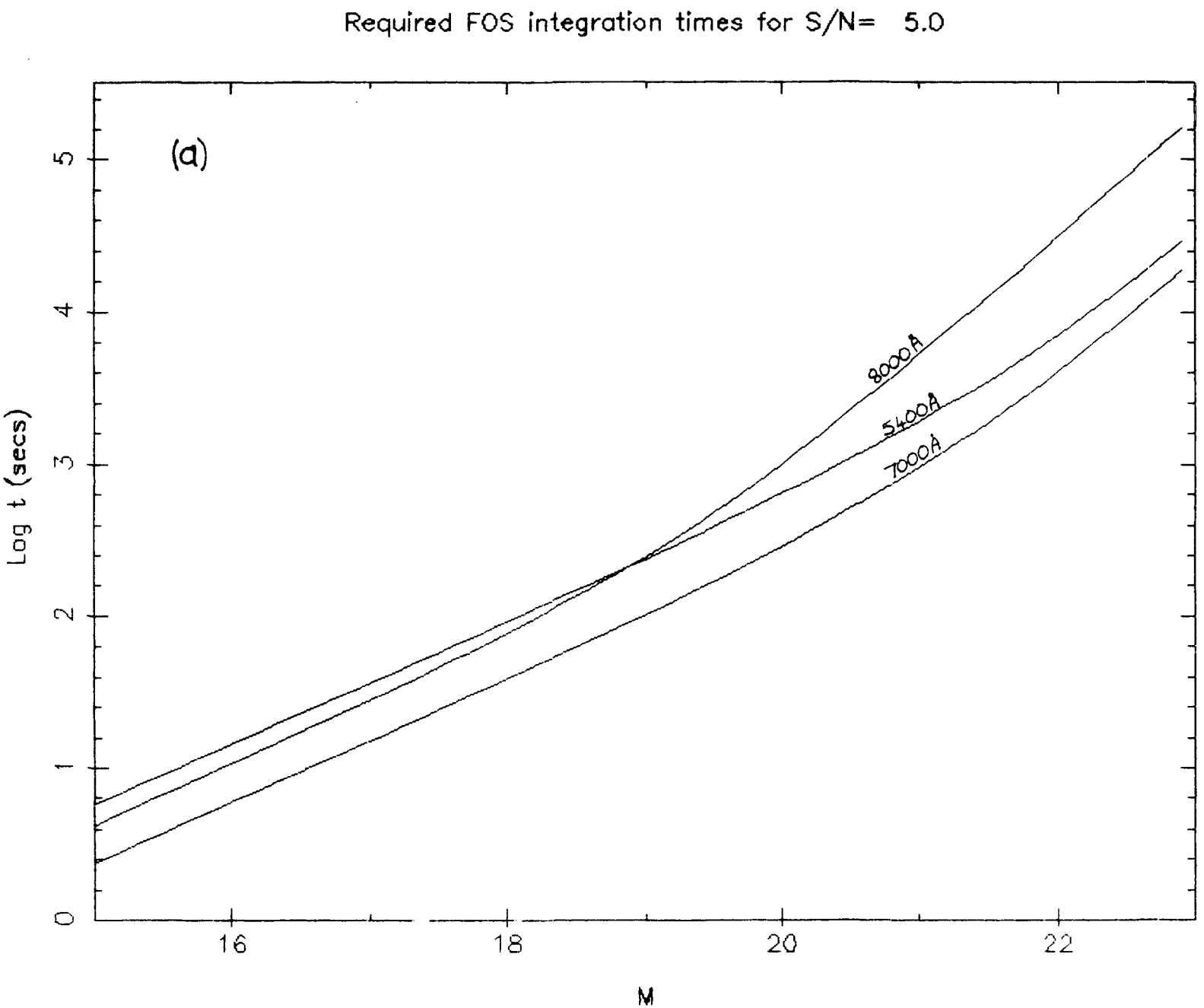

Required FOS integration times for S/N= 5.0
(a)
Log t (secs)
0
1
2
3
4
5
16
18
20
22
M
8000Å
5400Å
7000Å

FIG 4.5 (a)

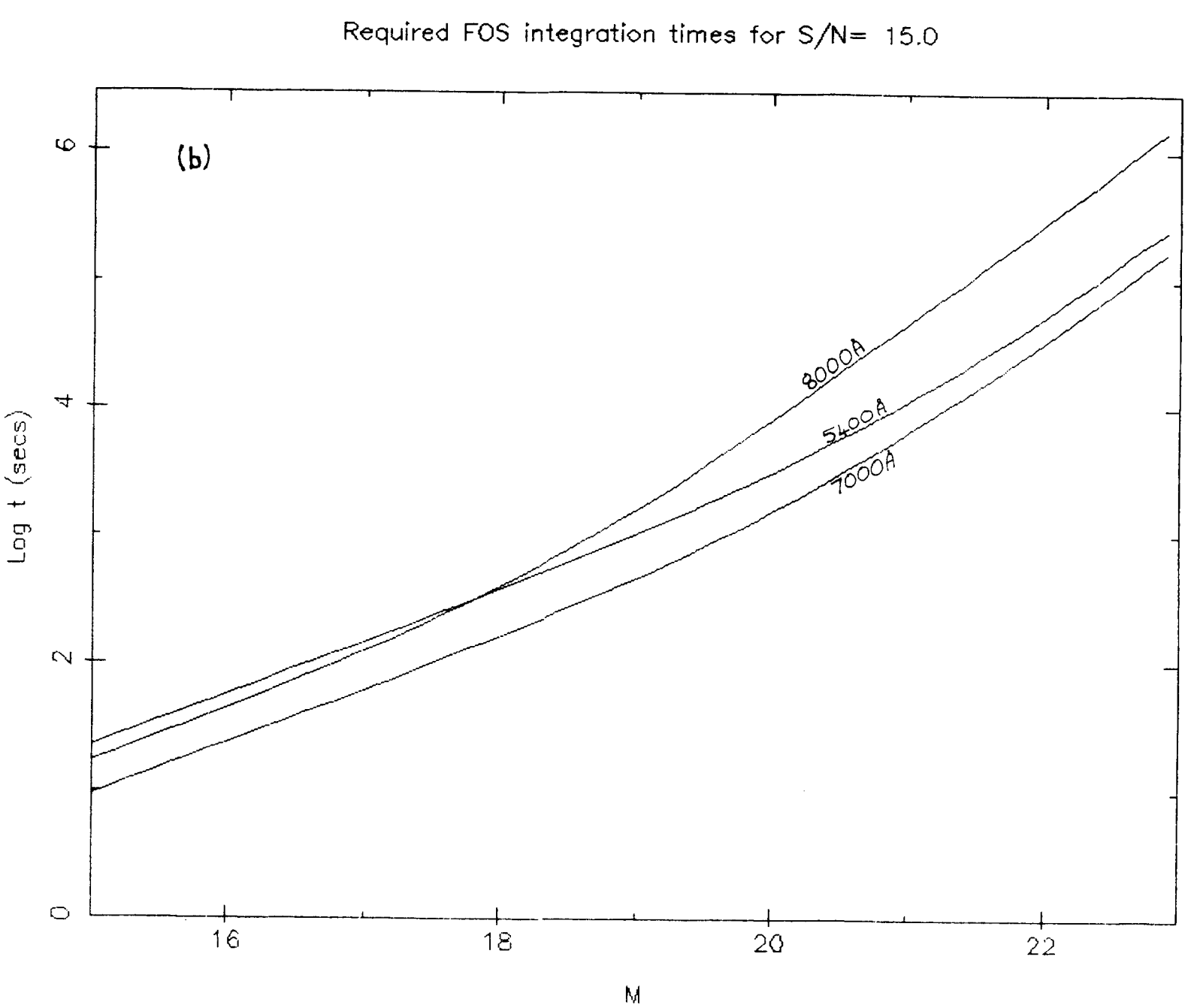

Required FOS integration times for S/N= 15.0
(b)
8000Å
5400Å
7000Å
Log t (secs)
0
2
4
6
16
18
20
22
M

FIG 4.5 (b)

4.3 FOS in action

The acquisition of the redshifts of very faint extragalactic objects is FOS's primary function. The features in such objects are generally broad enough to be detected with low resolution and the high throughput, large wavelength coverage and low readout noise of FOS make it the ideal instrument for redshift determination. But FOS has other uses too. It is excellent for rapidly obtaining wide wavelength coverage spectra for classification purposes. Not only can features be seen but the spectral energy distribution of an object can be clearly deduced by flux calibration. FOS is also therefore an excellent instrument for doing accurate spectrophotometry (because of the linear CCD response) - accurate flux measurements can give valuable insight into the physical processes occuring in the source.

In order to extract the full information from the raw data it is necessary to carefully apply some corrections and calibrations. The on-line data reduction programs are described in full in chapter 3 but before presenting some spectra to illustrate the use of FOS I shall outline the data reduction process that has to be applied before serious interpretation can begin. Figure 4.6 shows the spectrum of the radio source PKS2128-12 (obtained in an exposure time of 800 seconds) at various stages during data reduction. Figure 4.6 (a) shows the observed counts including the sky contamination. This 16th magnitude object is significantly brighter than the background sky yet many of the features seen at this stage are of terrestrial origin. For faint objects one

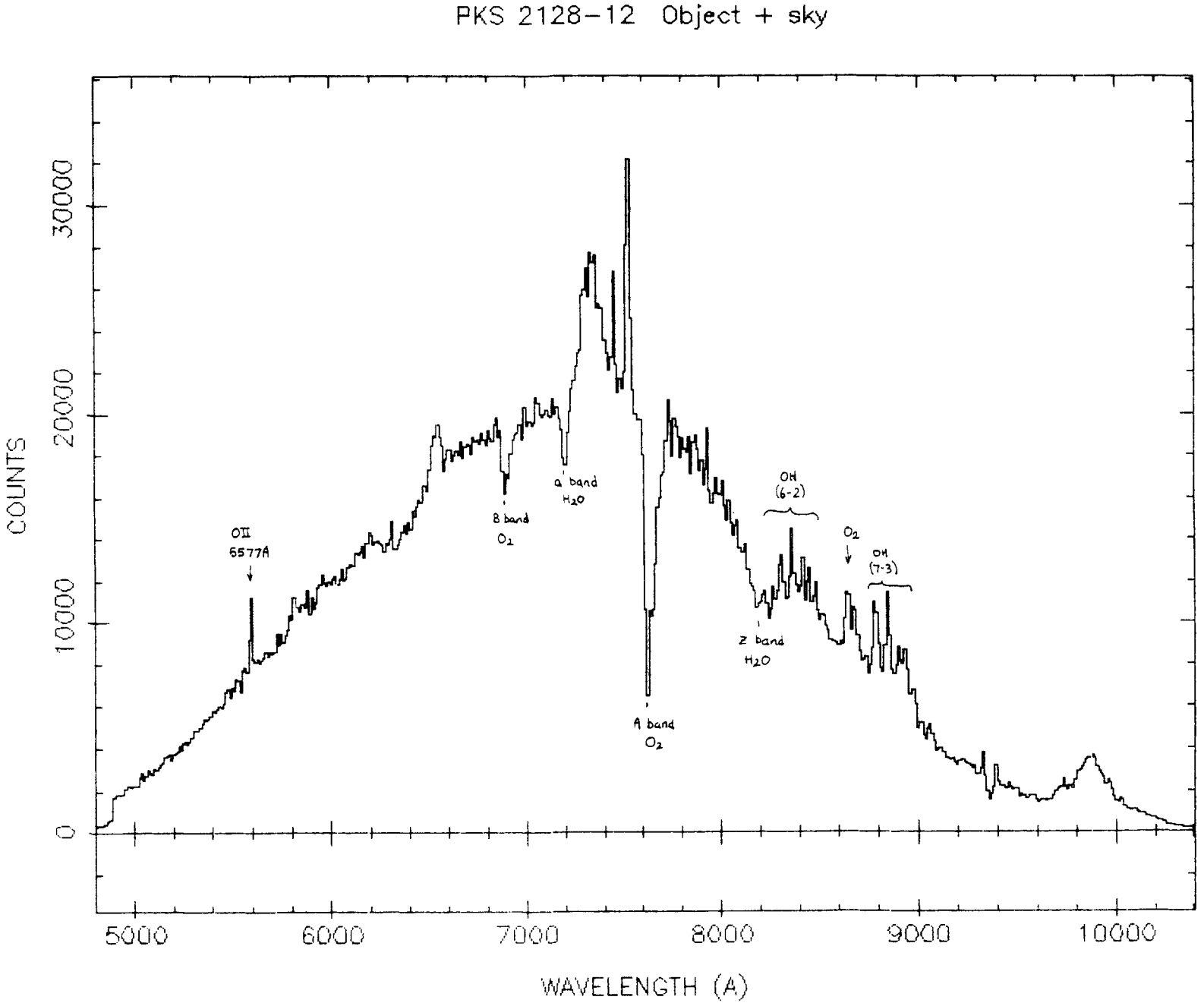

PKS 2128−12 Object + sky
COUNTS
0
10000
20000
30000
5000
6000
7000
8000
9000
10000
WAVELENGTH (A)
OII
6577Å
B band
O₂
a band
H₂O
A band
O₂
Z band
H₂O
OH
(6-2)
O₂
OH
(7-3)

FIG 4,6(a)

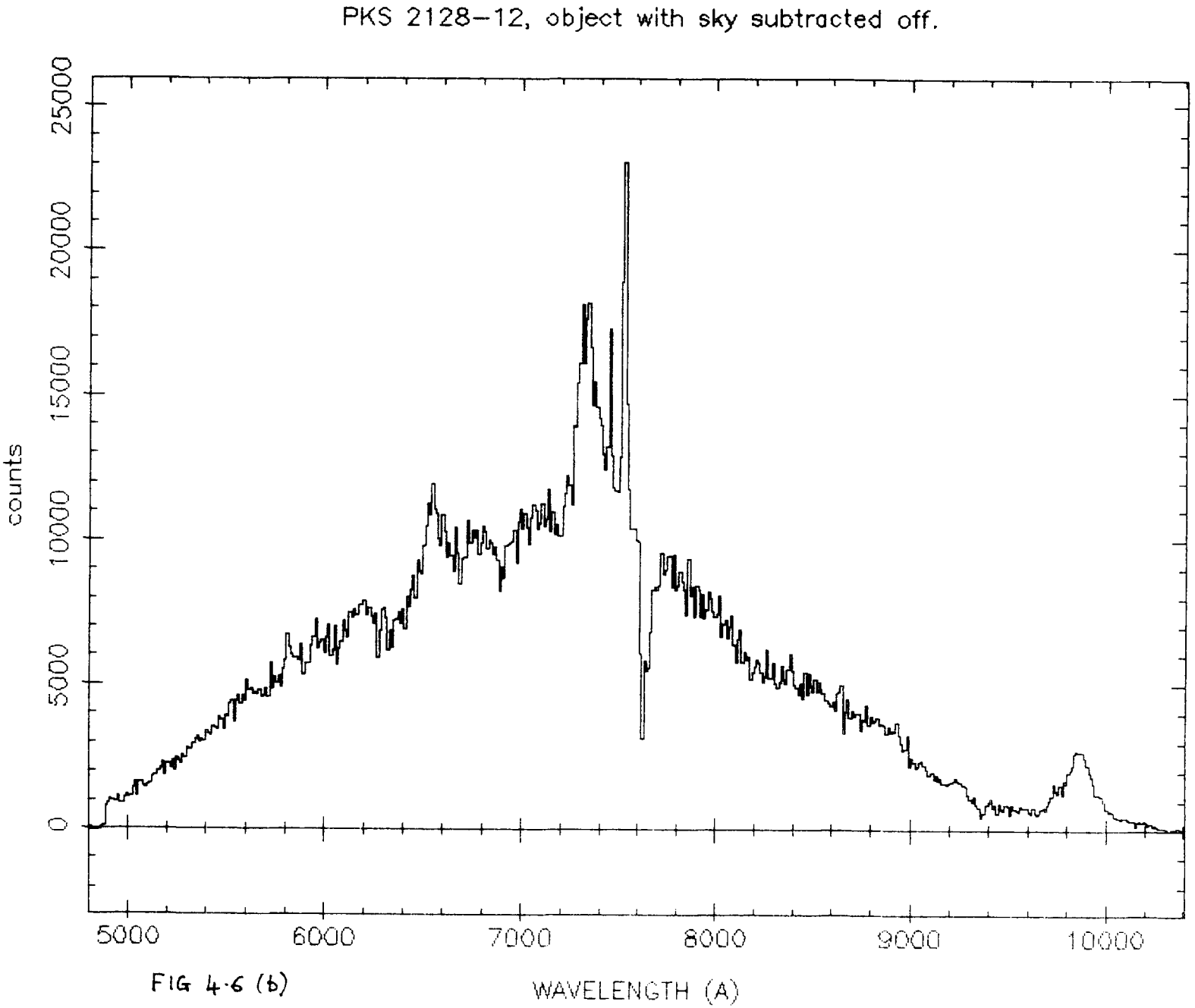

PKS 2128−12, object with sky subtracted off.
counts
0
5000
10000
15000
20000
25000
5000
6000
7000
8000
9000
10000
WAVELENGTH (A)

FIG 4·6 (b)

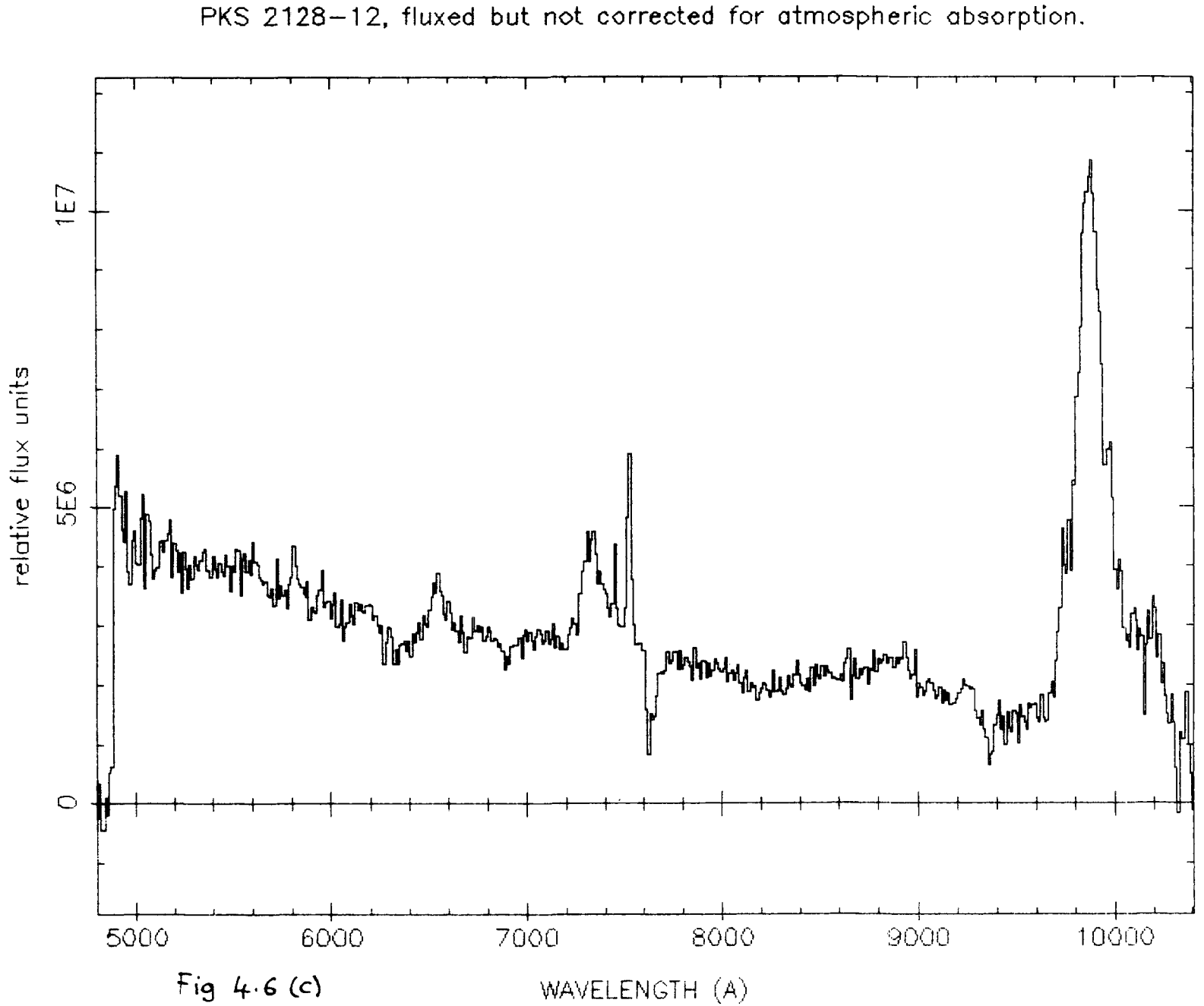

PKS 2128−12, fluxed but not corrected for atmospheric absorption.
relative flux units
1E7
5E6
0
5000
6000
7000
8000
9000
10000
WAVELENGTH (A)

Fig 4.6 (c)

PKS 2128−12, fluxed & corrected for atmospheric absorption.

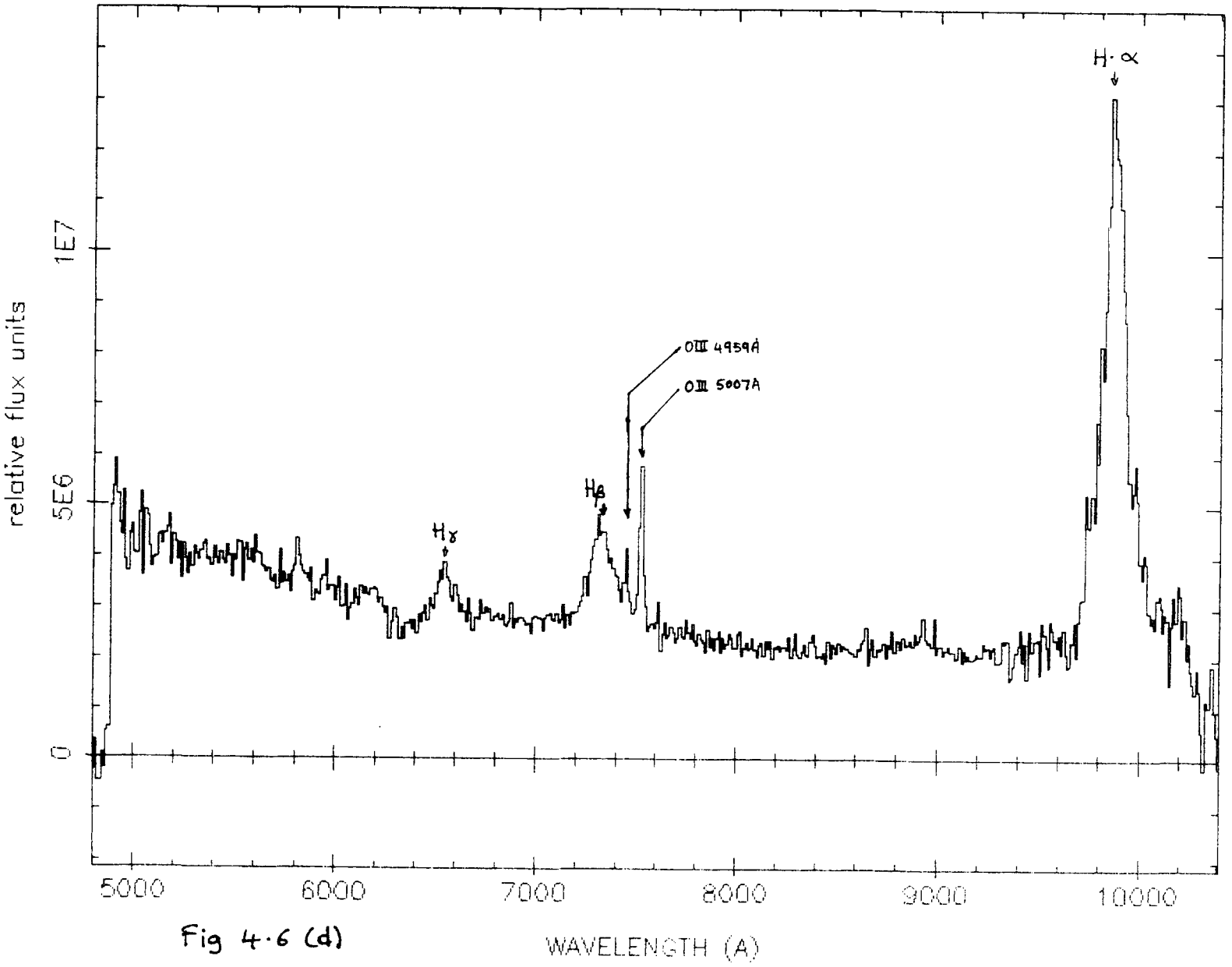


Fig 4·6 (d)

cannot even hope to recognise object features without sky-subtraction. Figure 4.6 (b) shows the counts observed from the object alone i.e. after sky subtraction. Note that atmospheric absorption features still remain and also that the general shape of the spectrum is mainly due to the system throughput rather than PKS2128-12 itself. Figure 4.6 (c) shows the spectrum on a relative flux scale. This is obtained by correcting for the known system throughput. We can now clearly see that the continuum of PKS2128-12 is blue. Also notice how the true intensity of the H-alpha line near 10000A and the Balmer decrement are clearly shown after this correction is made. Finally, figure 4.6 (d) shows the spectrum after the atmospheric absorption lines have been corrected. The correction is determined by using an observation of a bright, intrinsically featureless object. The only significant features left in the spectrum at this stage are due to PKS2128-12 and interpretation is made much easier. We see clearly that PKS2128-12 is a type I Seyfert galaxy with a redshift of 0.501. Bright, emission line objects such as PKS2128-12 are easy meat for FOS.

4.3.1 Redshift determinations

Faint object redshifts are not as easy to get as for PKS2128-12 discussed above. Long exposures are required and the faint spectrum has to be carefully seperated from the considerable sky and CRE contamination. Objects with emission lines are easier to get redshifts for than absorption line galaxies because emission features can be quite strong and stand out clearly from the noisy continuum. Figure 4.7 shows

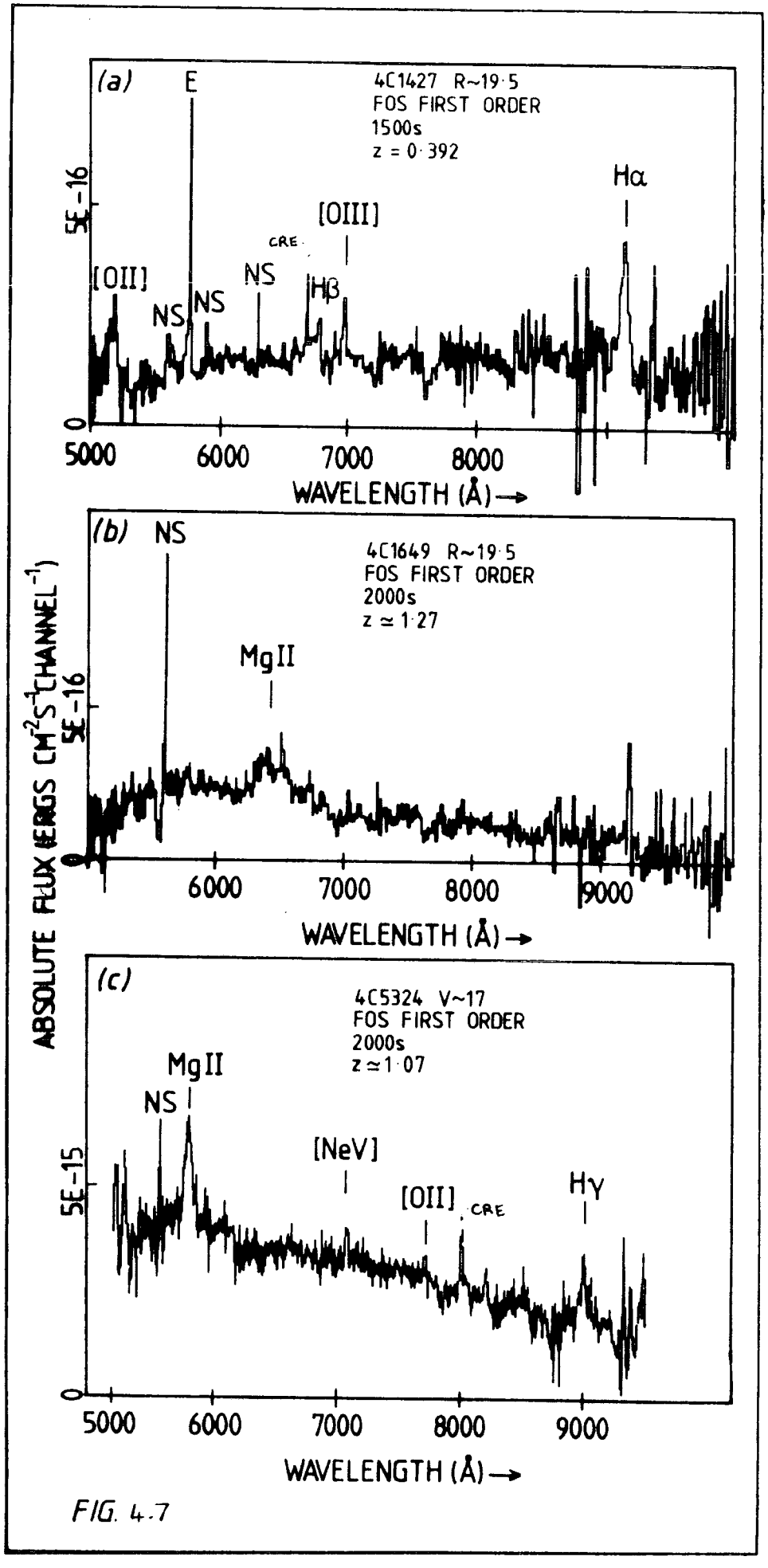

(a)
E
4C1427 R~19·5
FOS FIRST ORDER
1500s
z = 0·392
Hα
[OIII]
CRE
[OII]
NS
NS
NS
Hβ
5E-16
0
5000
6000
7000
8000
WAVELENGTH (Å)→
(b)
NS
4C1649 R~19·5
FOS FIRST ORDER
2000s
z ≃ 1·27
MgII
ABSOLUTE FLUX (ERGS CM⁻²S⁻¹CHANNEL⁻¹)
5E-16
0
6000
7000
8000
9000
WAVELENGTH (Å)→
(c)
4C5324 V~17
FOS FIRST ORDER
2000s
z ≃ 1·07
MgII
NS
[NeV]
[OII]
CRE
Hγ
5E-15
0
5000
6000
7000
8000
9000
WAVELENGTH (Å)→

FIG. 4.7

the spectra of three radio sources obtained with FOS which have provided three new redshifts. The spectrum of 4C14.27 is essentially a noisy version of the spectrum of PKS2128-12 shown in figure 4.6 (d) and the uncertainty in the redshift interpretation is virtually non-existant. The spectrum of 4C53.24 was taken during a full moon and shows that FOS is capable of doing moderately faint work (V=17) during bright time (it was in fact difficult to see it on the acquisition TV). The proposed redshift of 1.27 for 4C16.49 is based on a single, but clearly seen, broad emission feature. The uncertainty of this interpretation leaves a little to be desired but it is commonly accepted that the MgII 2799A feature is often seen in isolation.

FOS should prove a useful tool for discovering high redshift QSOs as it can detect Lyman-alpha in first order for redshifts in the range 3.1 to 7.2 (the highest redshift known to date is 3.78). As an example of the ease with which FOS can get the spectra of high redshift quasars figure 4.8 shows the spectrum of the well known quasar OQ172 which has a redshift of 3.6.

For absorption line redshifts a greater signal-to-noise is required. Figure 4.9 shows the spectrum of the bright nearby galaxy NGC5813 in both spectral orders. This shows just how weak the features can be. Apart from the well known features Mgb 5174A, NaI 5892A and H-alpha 6563A there are molecular bands in the near infra-red due to K stars (c.f. figure 4.11). The second order spectrum also reveals that such weak features are more easily detected with higher resolution. When a

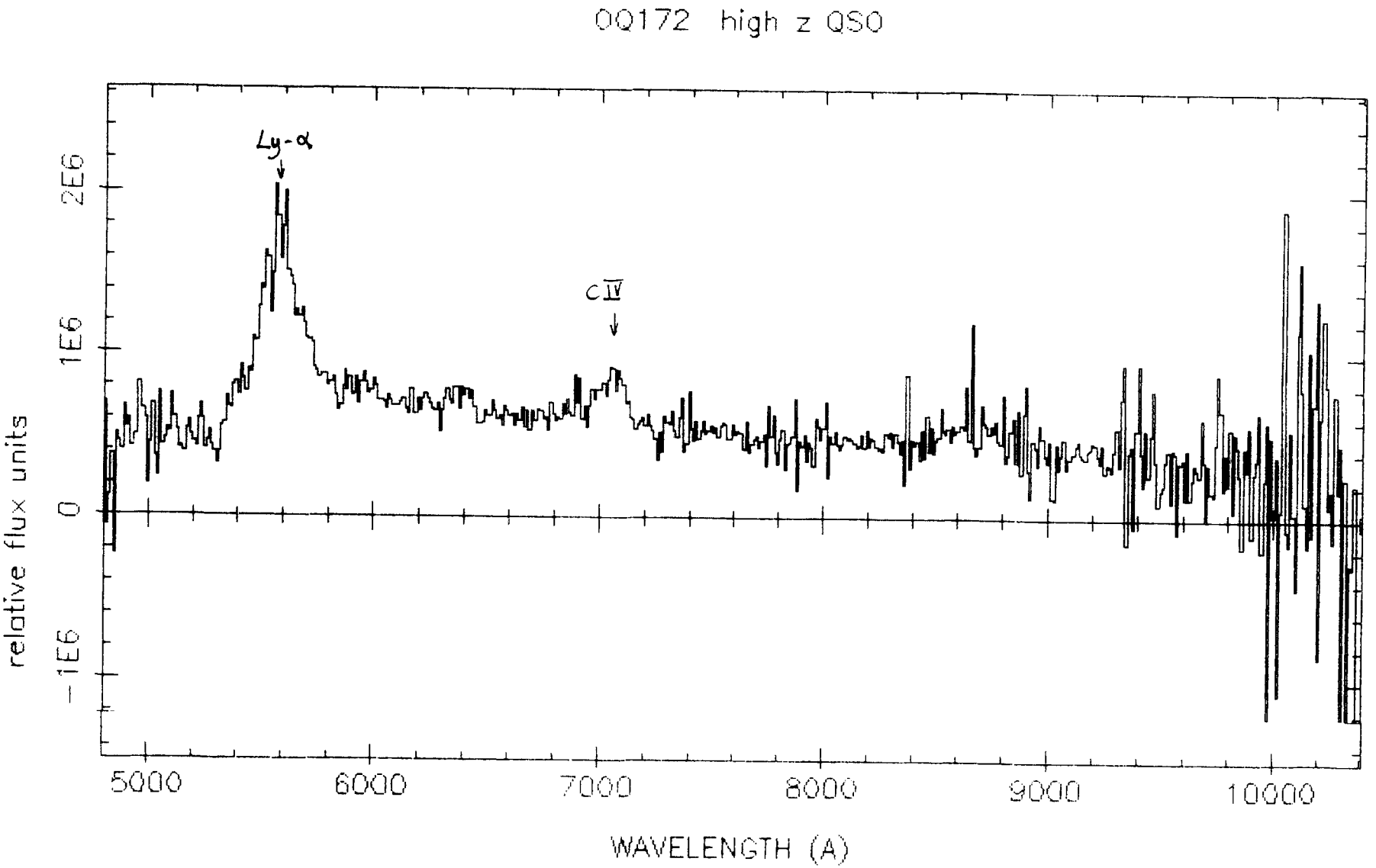

0Q172 high z QSO
Ly-α
CIV
relative flux units
2E6
1E6
0
-1E6
5000
6000
7000
8000
9000
10000
WAVELENGTH (A)

FIG 4.8

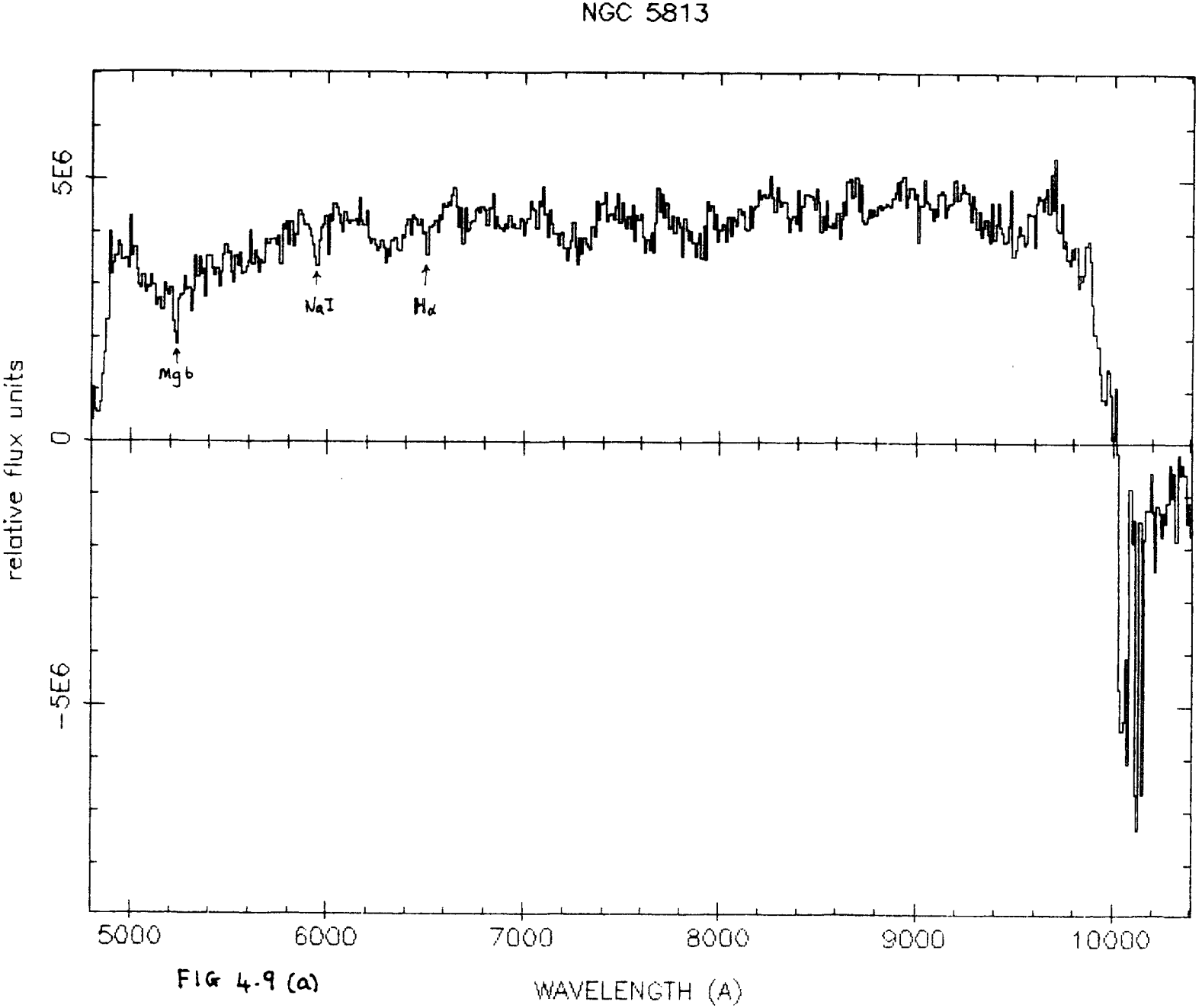

NGC 5813
5E6
0
-5E6
relative flux units
Mg b
NaI
Hα
5000
6000
7000
8000
9000
10000
WAVELENGTH (A)

FIG 4.9 (a)

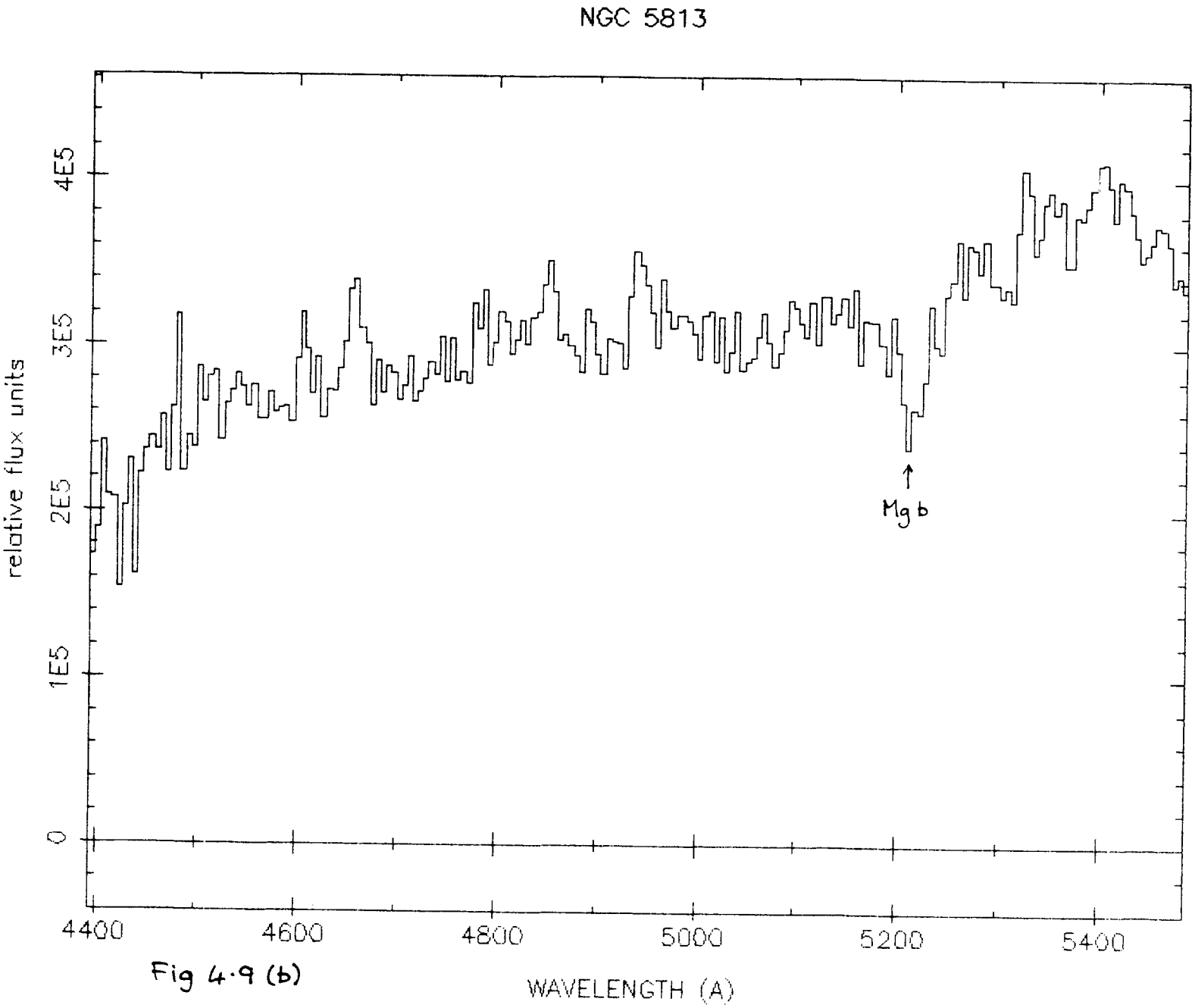

NGC 5813
relative flux units
0
1E5
2E5
3E5
4E5
Mg b
4400
4600
4800
5000
5200
5400
WAVELENGTH (A)

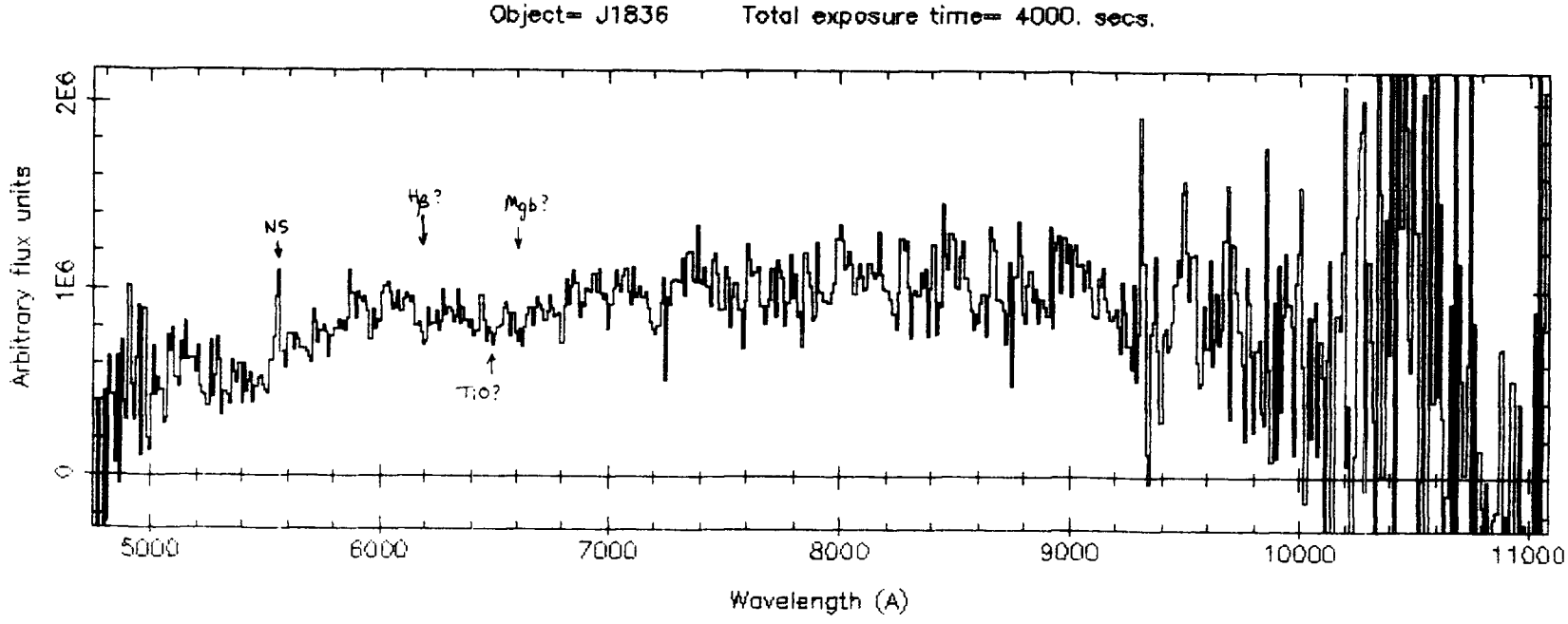
Fig 4.9 (b)

spectrum of this type is redshifted these features can be pushed into the atmospheric emission/absorption "jungle", longward of 7200A but other strong features such as the H & K lines of CaII and the Balmer break come in from the blue. Figure 4.10 shows the spectrum of a faint galaxy in the suspected poor, low redshift cluster J1836.10RC. A redshift of 0.275 has been proposed*. This example is indicative of the difficulties associated with faint galaxy work.

4.3.2 Classification work

Very faint stars can be classified using FOS by measuring the shape of their continuum spectrum. Figure 4.11 shows a sequence of 6 stellar spectra; 5 from the cluster NGC6664 and the M star Wolf 1040. The change in continuum slope from early to late type is very clear. Also molecular bands are clearly evident for the two late type stars. The nature of most objects can be determined from a FOS observation and a knowledge of the morphology. Figure 4.12 shows the spectrum of the radio source candidate 1342-016 in both orders. The plots shown are in fact the actual plots obtained whilst observing. The spectrum appears to be that of an early-intermediate type star. Because the reduced data could be examined whilst observations were being taken, the telescope time was used efficiently to determine the nature of the object. Another interesting example of FOS being used to investigate the nature of an object is illustrated in figure 4.13. The object was investigated with FOS because of its unusual, extreme red colour (it is only visible on red, 4-metre plates). The object is designated FFRT21 (funny faint red thing number 21). An

* The proposed redshift is very uncertain and further work is required.


Object= J1836 Total exposure time= 4000. secs.
Arbitrary flux units
2E6
1E6
0
NS
Hβ?
Mgb?
TiO?
5000
6000
7000
8000
9000
10000
11000
Wavelength (A)

8b

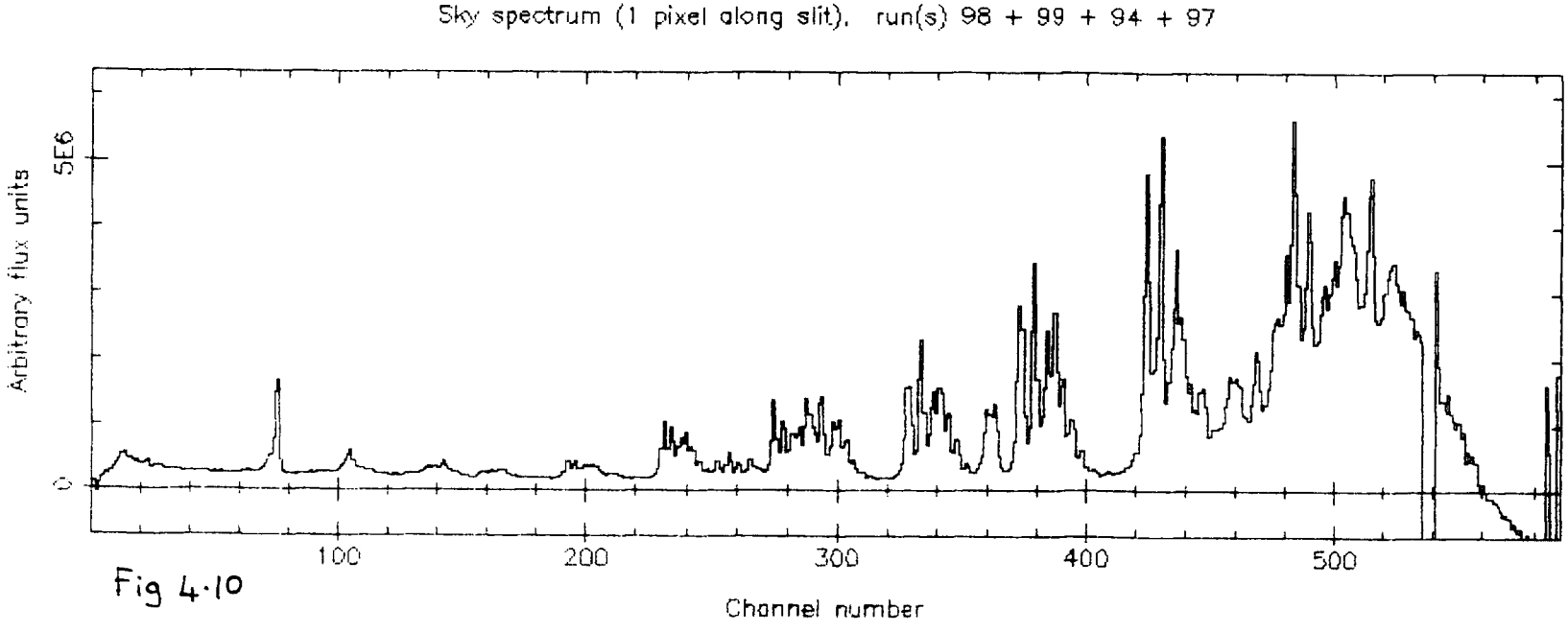

Sky spectrum (1 pixel along slit). run(s) 98 + 99 + 94 + 97
Arbitrary flux units
5E6
0
100
200
300
400
500
Channel number

Fig 4.10

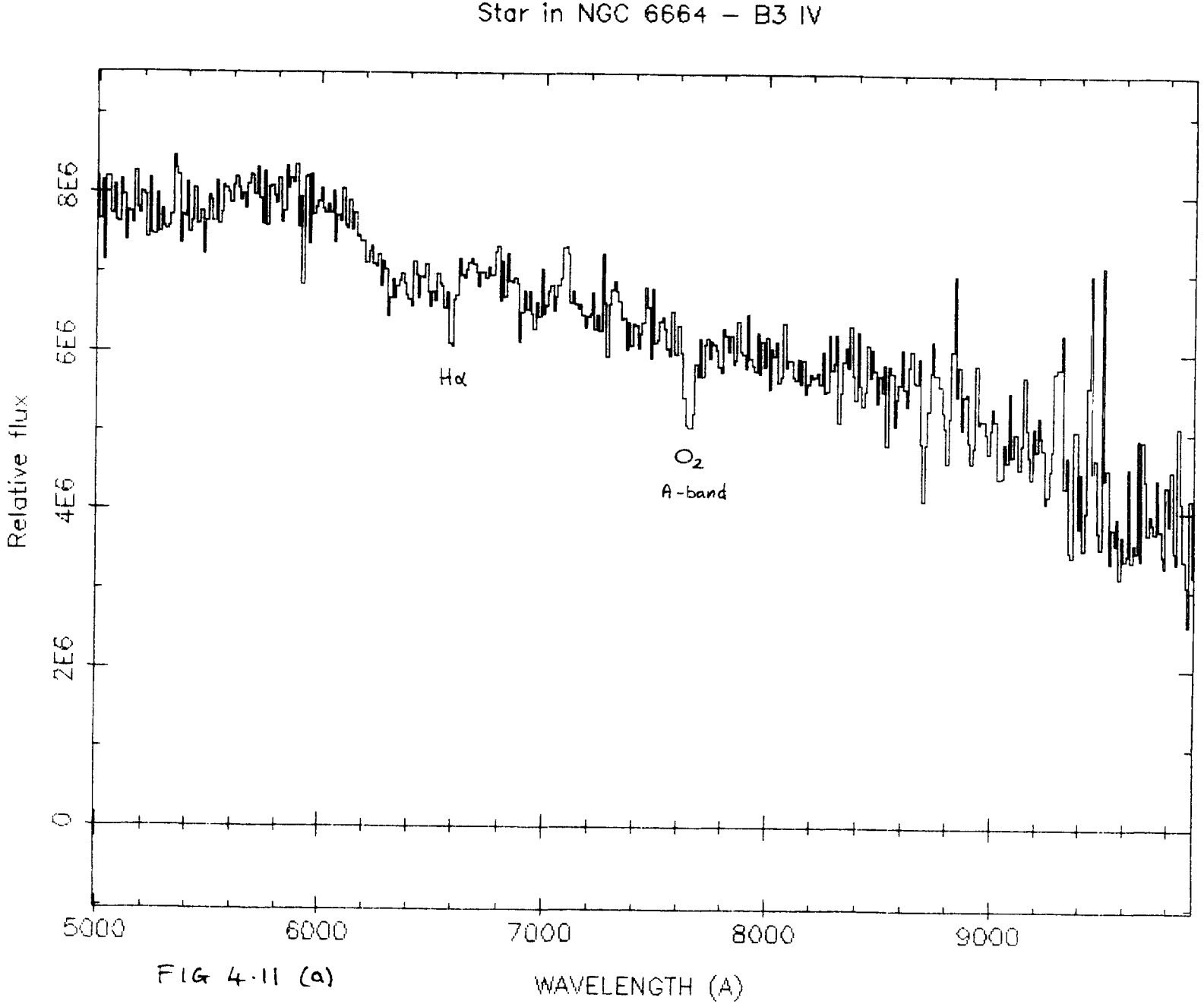

Star in NGC 6664 – B3 IV
Relative flux
8E6
6E6
4E6
2E6
0
Hα
O2
A-band
5000
6000
7000
8000
9000
WAVELENGTH (A)

FIG 4.11 (a)

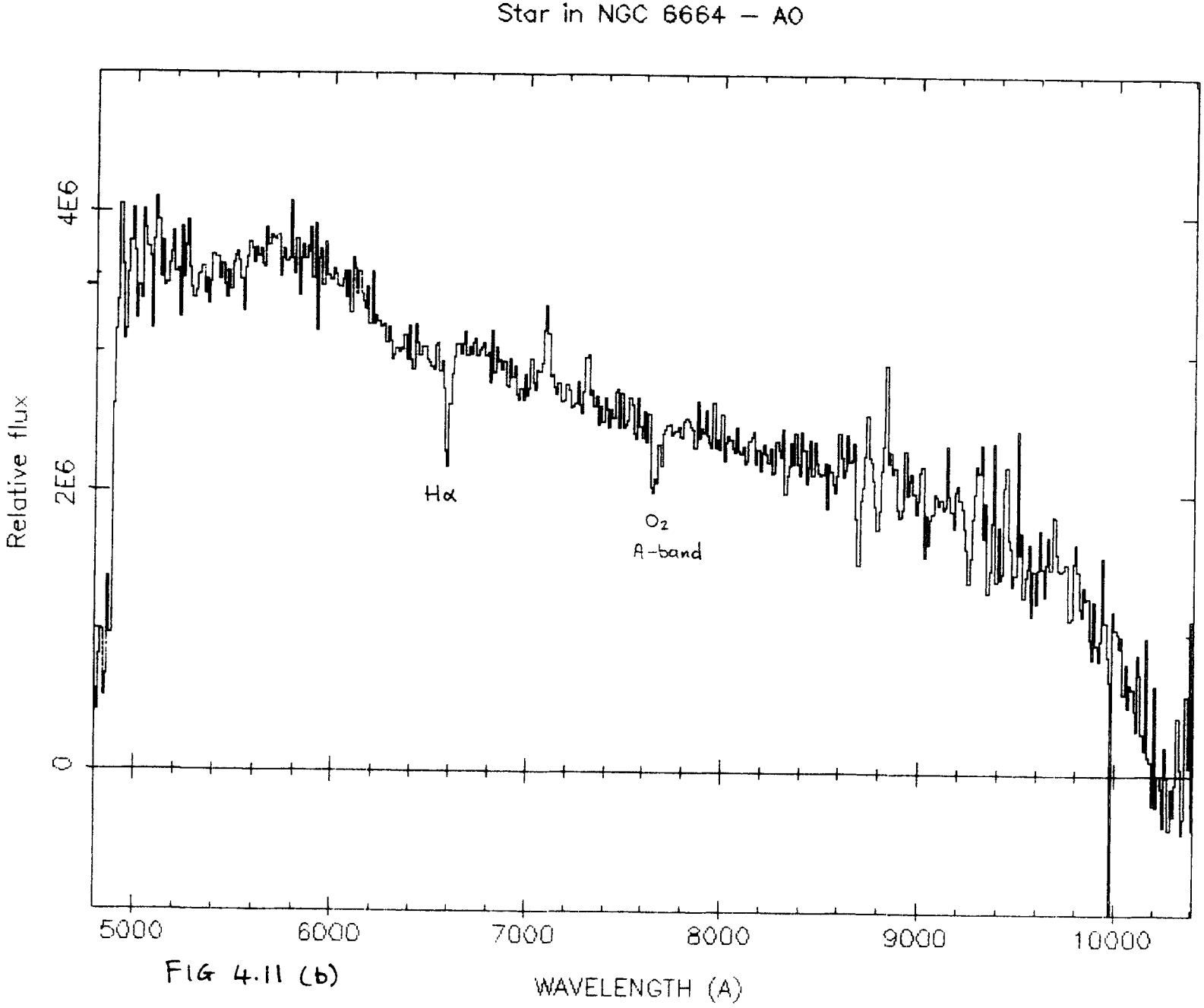

Star in NGC 6664 – A0
Relative flux
4E6
2E6
0
Hα
O2
A-band
5000
6000
7000
8000
9000
10000
WAVELENGTH (A)

FIG 4.11 (b)

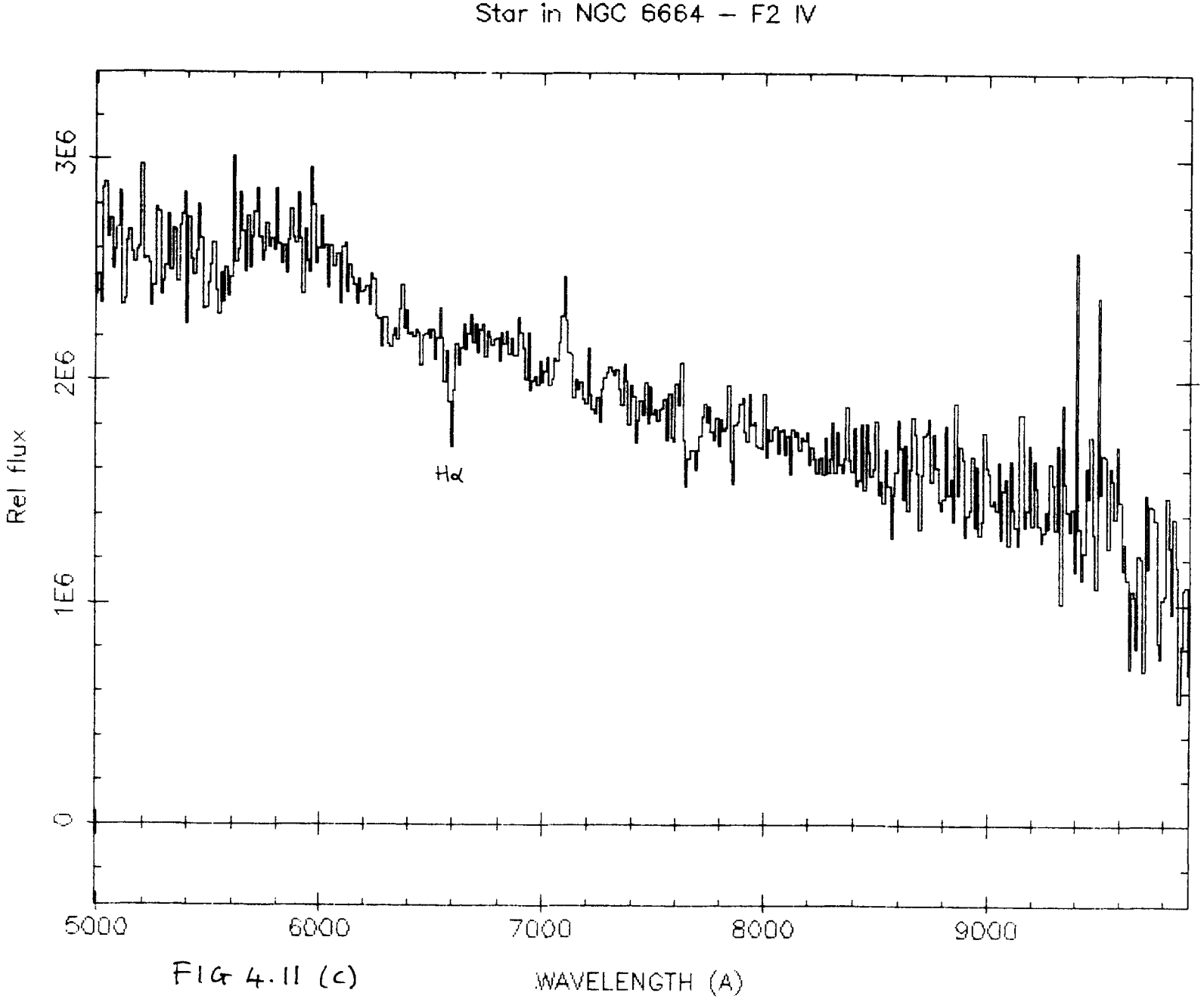

Star in NGC 6664 – F2 IV
Rel flux
3E6
2E6
1E6
0
Hα
5000
6000
7000
8000
9000
WAVELENGTH (A)

FIG 4.11 (c)

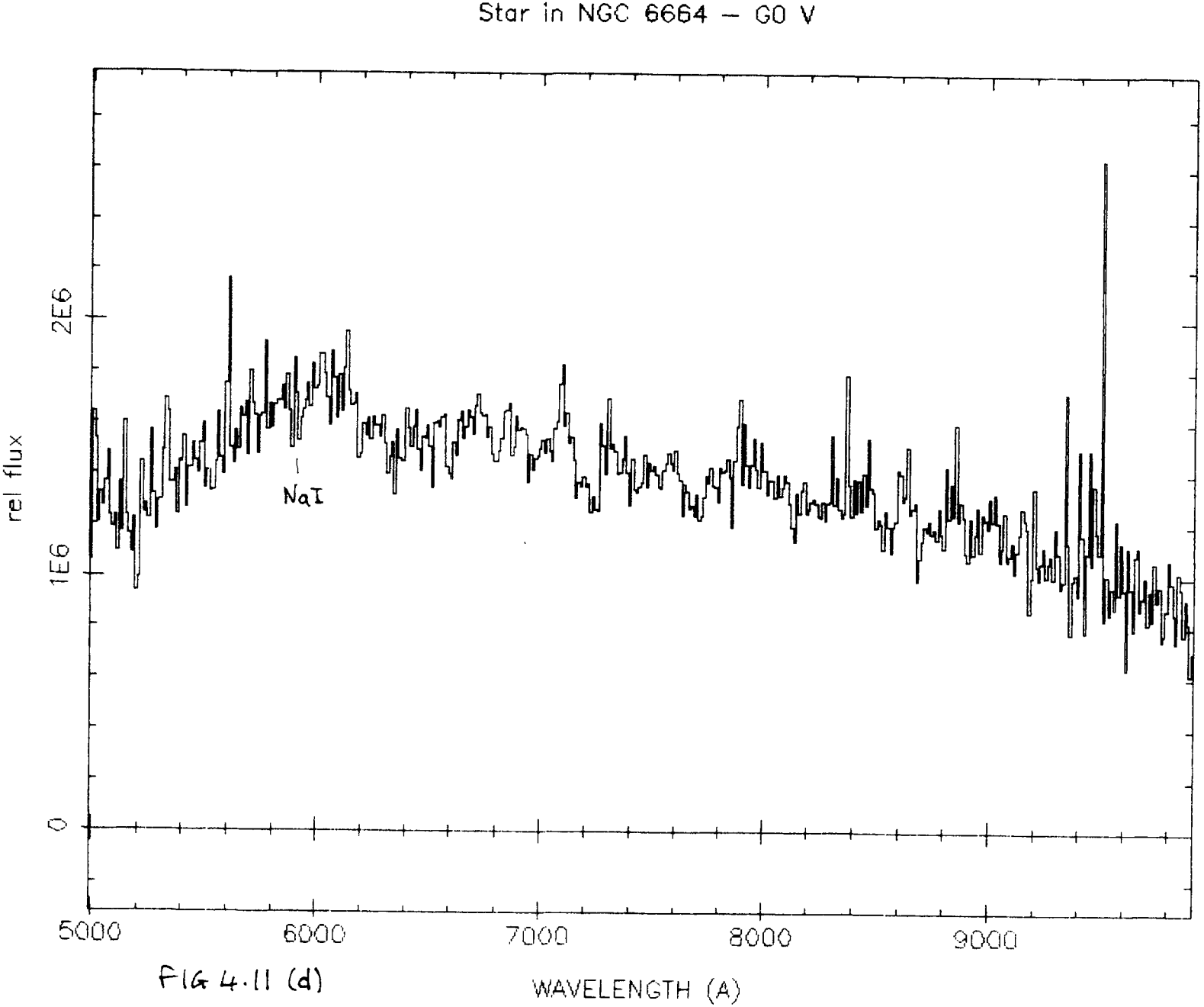

Star in NGC 6664 – G0 V
rel flux
2E6
1E6
0
NaI
5000
6000
7000
8000
9000
WAVELENGTH (A)

FIG 4.11 (d)

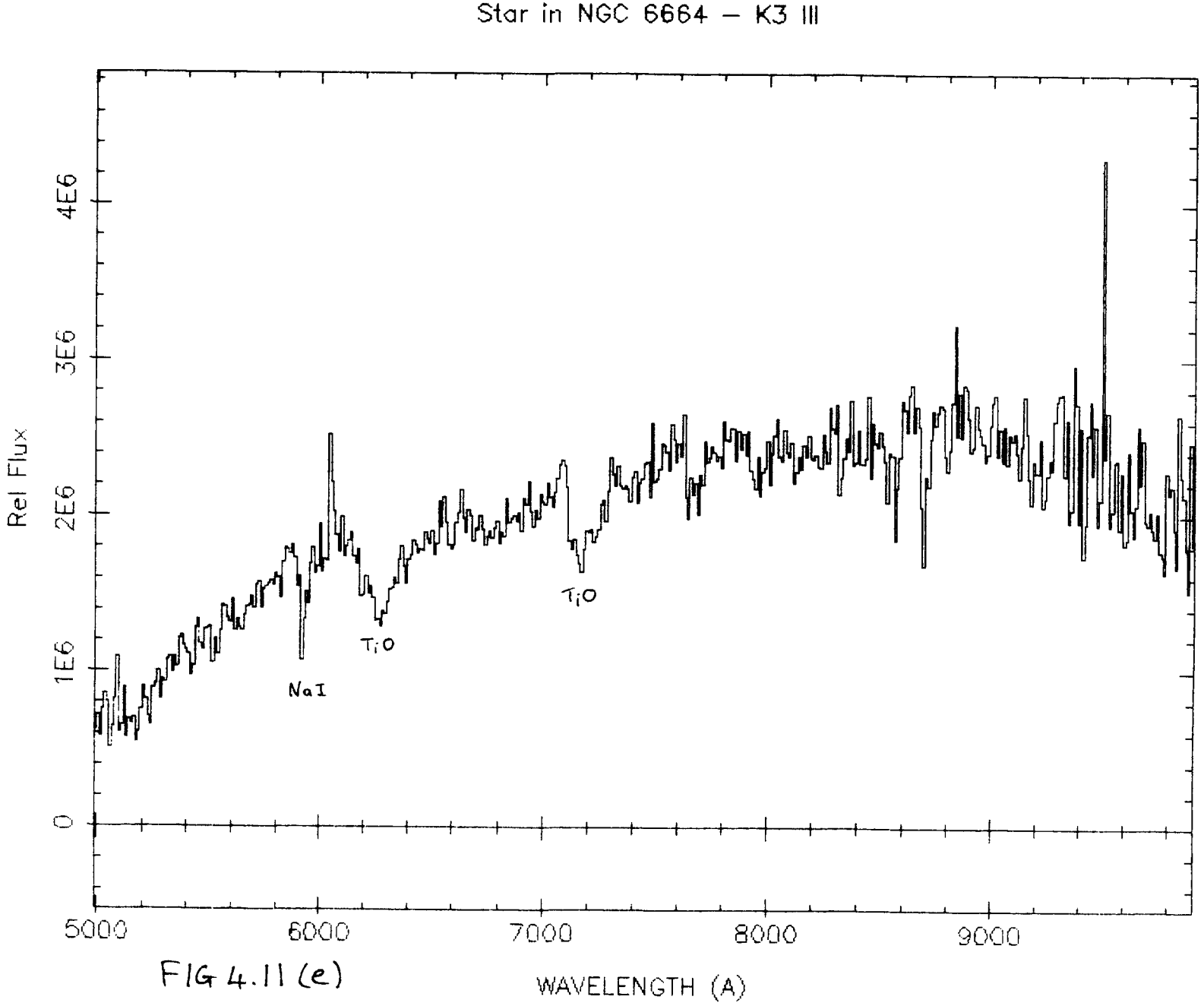

Star in NGC 6664 – K3 III
Rel Flux
0
1E6
2E6
3E6
4E6
NaI
TiO
TiO
5000
6000
7000
8000
9000
WAVELENGTH (A)

FIG 4.11 (e)

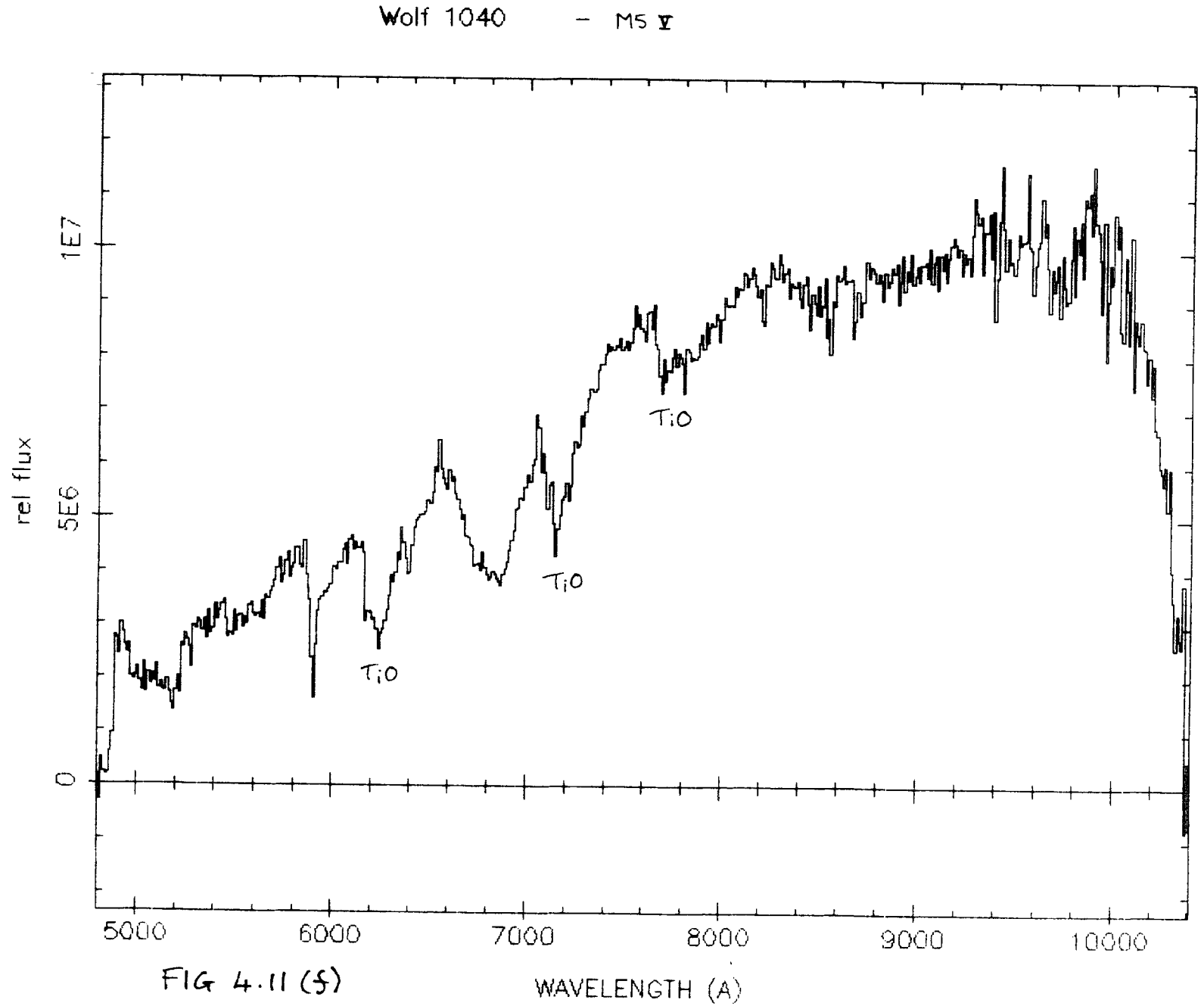

Wolf 1040 - M5 V
rel flux
1E7
5E6
0
TiO
TiO
TiO
5000
6000
7000
8000
9000
10000
WAVELENGTH (A)

FIG 4.11 (5)

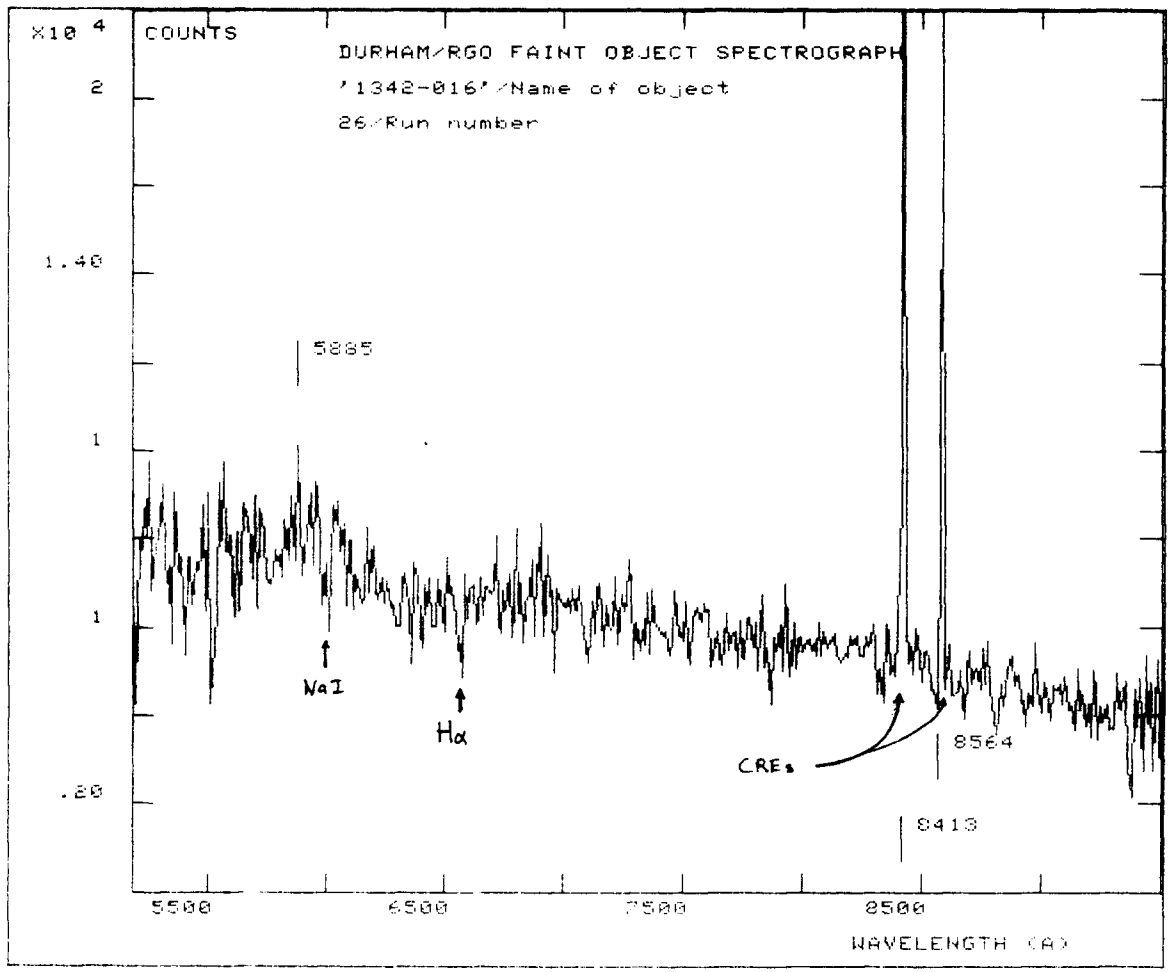

X10 4
COUNTS
DURHAM/RGO FAINT OBJECT SPECTROGRAPH
'1342-016'/Name of object
26/Run number
2
1.40
1
1
.80
5885
NaI
Hα
CREs
8564
8413
5500
6500
7500
8500
WAVELENGTH (A)

Figure 4.12 (a)

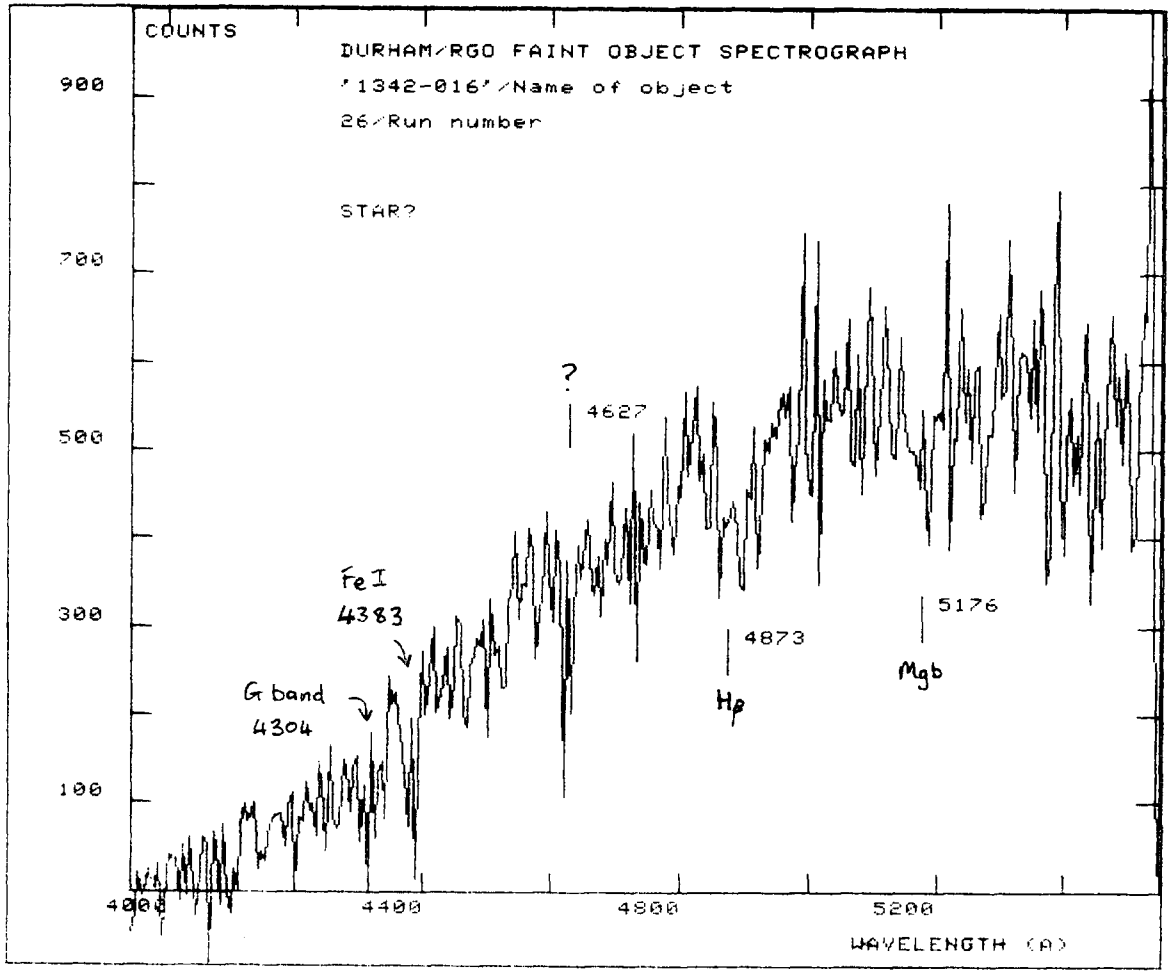

COUNTS
DURHAM/RGO FAINT OBJECT SPECTROGRAPH
'1342-016'/Name of object
26/Run number
STAR?
900
700
500
300
100
?
4627
FeI
4383
G band
4304
4873
Hβ
5176
Mgb
4000
4400
4800
5200
WAVELENGTH (A)

Figure 4·12 (b)

Anomolously red object – faint M star.

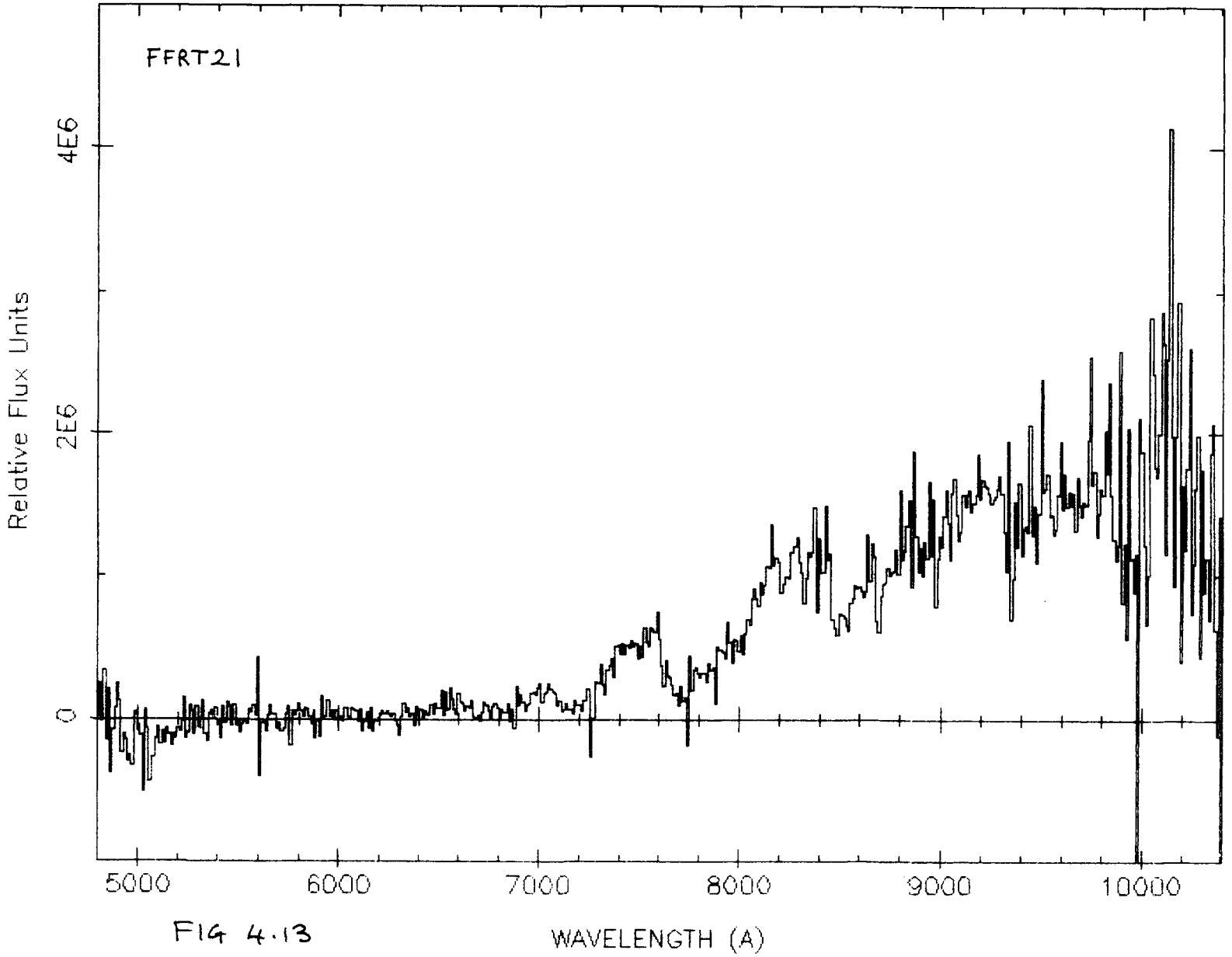


FIG 4.13

exposure of 1000 seconds was enough to determine that the object is infact an M star (compare it with the spectrum of Wolf 1040 in figure 4.11). It had been predicted beforehand that FFRT21 might be a primordial galaxy at high redshift and interest was lost in it as soon as it was discovered to be an M star (several others were discovered too). However, FFRT21 is very red even by M star standards and merits further investigation.

4.4 Future developments

The most obvious improvement that can be made to FOS is to increase its sensitivity in the blue by furnishing it with a thinned or coated CCD (see section 2.2.5). This will make FOS a very powerful instrument indeed as good signal-to-noise spectra in the relatively clean blue part of the atmospheric window would then be available. This blue region is also very important because many prominent galaxy features are seen in the blue for redshifts less than 0.5.

Another exciting FOS enhancement is the manual multislit unit which was developed by the RGO and will be commissioned in dark time in February 1986. This will be particularly suitable for studying faint distant galaxy clusters like J1836.1ORC discussed above. The multislit unit replaces the conventional slit module and consists of 12 mechanical slit slides which are free to move in a direction normal to the slits themselves. The slits have to be manually arranged in the correct observing positions using a measuring machine. The slit mechanism can be removed and put back on the telescope in a matter of minutes

and in due course several will be available. A field of 3 by 4 arc-minutes can be accessed with the slits. A GG495 field lens is required to cut out second order. With the present GEC CCD this is no great loss but when a blue sensitive device is available it will be a considerable drawback. A similar device, operated under computer control is being developed for FOS in Durham.

The fact that the FOS for the 4.2 metre telescope will be very similar to the existing FOS is clear evidence of how good the design is for single object work. However, as a multi-object spectrograph FOS has its limitations. Only a small number of fibres can be used if both spectral orders are required. Alternatively, the use of the full slit length with fibres means that one spectral order must be sacrificed and so part of the most useful spectral range between 3700A to 7400A must be lost. Similar arguments hold for multislit work and there is the addition problem that the accessible field is very small. For multi-fibre work an uncross-dispersed system with a single order, free spectral range from 3700A to 7400A and a conventional size, blue-sensitive CCD would be more suitable. An even larger spectral range might be obtained using a prism instead of a grating but one would then have to live with a dispersion that changes rapidly with wavelength.

Chapter 5

Techniques of Multi-object Spectroscopy

5.1 Introduction

This chapter describes and compares the various methods of multi-object spectroscopy (MOS). A MOS technique is a spectroscopic technique in which the light from several objects is detected simultaneously. MOS techniques have been developed because many important astrophysical and cosmological problems cannot be properly addressed without considerable amounts of good quality spectroscopic data at faint magnitudes (see chapter 1). However, sufficient data is not available at present because MOS techniques have yet to make their full impact and the traditional method of obtaining spectra one at a time is far too inefficient given the integration times required, even on the World's largest telescopes (e.g. a required integration time of 10,000 seconds is typical on a 4-metre class telescope). MOS techniques rely on the fact that all telescopes have fields of view that image many faint, potentially interesting, objects simultaneously. But for single object spectrocopy only a tiny fraction of the potential information is gathered. It is interesting to speculate on how many objects of world shattering significance have been illuminating the back of the slit whilst light from some object of lesser importance has been entering the spectrograph. Furthermore, the additional costs of upgrading to a multi-object capability are a mere fraction of the cost of the telescope itself.

5.2 Slitless spectroscopy

In this technique the light is dispersed before it is imaged by the telescope. Thus all objects in the field have dispersed images and as the field is normally quite large a photographic plate is used as the detector. In terms of the number of simultaneous spectra the slitless technique is unbeatable. However, because there is no slit mask each element of the spectrum, which is only a fraction of the total object signal, lies on a very strong sky background which is undiminished by dispersion. Examples of the slitless technique include the enormous field coverage offered by the UK Schmidt objective prism and grisms for 4-metre class telescopes which have smaller fields of view but can go somewhat fainter. The use of low quantum efficiency photographic plates and the high sky background mean that the technique is restricted to low resolution and cannot be employed at the faintest magnitudes. There is also evidence (Savage & Peterson, 1983) that the response of the photographic emulsion varies with wavelength on scales similar to the width of detectable spectral features and this has introduced strong selection effects. Finally, calibration arcs cannot be taken and wavelength calibration has to be based on the observed spectral cut offs.

5.3 Imaging with filters

This is included because many objects are observed simultaneously and spectral information is obtained. However, unlike all the other techniques discussed in this chapter, multi-colour photometry does not record the signal in each

wavelength channel simultaneously. The spectral resolution is fairly low by spectroscopic standards as it relies on the bandpass of the filters. Also far more sky information than is needed is recorded though this does not introduce unneccessary contamination of the object measurements. The technique is generally employed to determine gross spectral energy distributions because the resolution and the sometimes incomplete wavelength coverage mean that line identification work is not practical. When observing it is therefore necessary to have either perfect photometric conditions or several standards in the field. Spectral information on many objects can be obtained with quite good efficiency, despite the need to take a number of exposures, but the data is of limited use and follow-up spectroscopy is inevitable.

The above paragraph was written with discrete point like objects in mind. If objects which extend in two dimensions on the sky are being studied then narrow band imaging techniques are essential. In particular, Fabry-Perot imaging spectroscopy such as with TAURUS is particularly valuable if higher spectral resolution is required. In the Fabry-Perot technique however, different parts of each individual image arise from different wavelenghts and the etalons have to be scanned through a complete free spectral range before all parts of the field are recorded at all wavelengths in the spectral range.

5.4 Long slit spectroscopy

Instrument rotators and spectrographs with long slit capability are commonly available at major telescope facilities and a limited amount of multi-object work can be done with such equipment. The total number of objects that can be placed on the slit depends on the field density of the objects (which depends on the limiting magnitude and the astronomical programme) and how lucky the observer is at finding objects that lie in a straight line. Thus faint galaxy cluster studies are a practical possibility and upto about half a dozen or so objects may be simultaneously acquired. The full range of standard spectrographic dispersions and low noise detectors are available and the use of a slit means that the resultant spectra are subject only to Poisson type errors (see section 2.1.1). Thus the long slit technique allows the very faintest of objects to be investigated and is only inefficient in terms of the total number of simultaneously observable objects. Also, spatial structure is recorded and it is possible, for example, to examine the data for evidence of extended emmision regions.

5.5 Multi-slit spectroscopy

This technique employs an aperture plate with precisely engineered slits that allows the light from the objects (and some sky) to enter the spectrograph. The principle is much the same as for conventional slit spectroscopy and can be likened to the objective prism technique with the sky background reduced to a minimum. However, simply replacing existing slit

units with multi-slit devices is not ideal as most existing spectrographs have small fields due to their poor off-axis imaging properties (i.e. aberations and vignetting are significant). Thus, ideally, multi-slit devices are best used with purpose built spectrographs that have a wide field capability. Once such a spectrograph is available one then has to find a suitably large detector since the spectrograph camera simply forms a dispersed image (with some demagnification) of a sizable portion of the telescope field. The currently available generation of CCDs (typically about 500 pixels square) are therefore unsuitable and multi-slit work will not reach its full potential until the next generation of large CCDs (about 1500 - 2000 pixels square) are available. The beauty of the multi-slit technique is that the limiting magnitude is set by Poisson statistics (as for conventional single slit observations) and provides a significant multi-object advantage without compromising the liming magnitude to which one can work. This is particularly important for cosmological programmes where long look-back times are essential. The faint magnitude limit also means that there is no shortage of targets in the field, even when clustered objects are not being studied, and so full use of the instrumentation's capability is always provided.

At the time of writing, the most outstanding* multi-slit system conceived is the Low Dispersion Survey Spectrograph (LDSS) which will be first used on the AAT in the summer of '86 before being used on the 4.2 metre telescope on La Palma in 1988. The specifications for this instrument, based on the arguments of the previous paragraph, were set out by Richard

* Other multi-slit systems such as EFOSC and the KPNO Cryogenic camera have smaller fields of view.

Ellis, Keith Taylor and Paul Atherton and an acceptable optical design was provided by Charles Wynne. LDSS (see figure 5.1) will be able to obtain spectra of objects in a field of 12 arcmin diameter. The modular, fully transmiting design allows it to be used with various dispersions and detectors (since the ideal detector is not yet well established) and replacement of the collimator allows it to be used on different telescopes. The spectral resolution is limited not by the grism but by the imaging properties of the collimator and camera and if high dispersions are used the size of the detector will limit the wavelength coverage of objects at the sides of the field. Thus LDSS is limited to resolutions of less than 10A but at very faint magnitudes this is ideal for cosmological work and higher resolutions would require considerably longer integration times.

Unlike the MOS techniques described above, multi-slit observing requires a considerable amount of planning and preparation so that a multi-slit configuration can be accurately and quickly set up when it is required. This preparation consists of astrometry of the target field followed by the precise design of the optimum slit pattern so that the greatest number of interesting object spectra (each with adequate sky sampling in its individual slit) are acquired on the spectrograph detector without there being any overlaps. The optimisation process therefore has to arrive at a list of which objects are to be observed, the positions (including width and length) of the slits and the position angle of the instrument rotator. The multi-slit devices themselves are either pre-manufactured aperture plates or re-usable mechanical

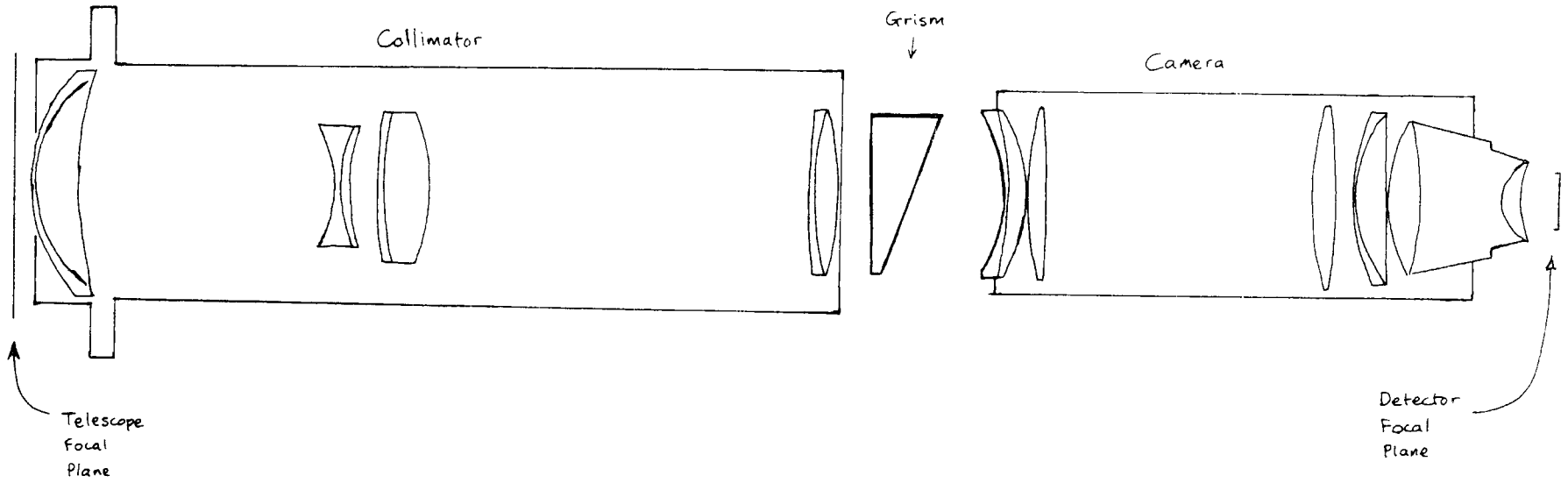


Figure 5.1 LDSS

devices (Breare et al, 1986a). The use of liquid crystal plates has been proposed but they are not ideal because the non-transmitting state leaks light and the transmitting state absorbs light. LDSS will use pre-manufactured aperture plates mounted in a carousel for quick field changeovers whereas two mechanical multi-slit devices (one manually set up, the other under computer control) are being developed for the small 3 by 4 arcmin field of the 2.5 metre FOS (see section 4.4).

LDSS is at the limit of present optical and detector technology - particularly detector technology where the ideal detectors are as yet unproven. To observe yet more objects simultaneously one needs either several LDSS's at the telescope focus or wider field spectrographs equiped with even bigger detectors. Wider field optics would probably be rather large and cumbersome but not impossible. However, to utilise the full field one would have to use either a relatively insensitive photographic plate or a mosaic of CCDs. Using several LDSS type spectrographs is not really very practical either as LDSS fills up the space behind the telescope focus and an ingenious (perhaps impossible) arrangement of folding mirrors would be required (note that LDSS will only utilise 1/9th of the full field of the AAT!). Finally, as multi-slit work seems unlikely to be able to address the full field of present day telescopes the technique does not demand that the next generation of optical telescopes have even wider fields (fields of about 3 degrees across are being discussed).

5.6 Multi-fibre spectroscopy

If one end of an optical fibre is placed in the focal plane of a telescope it will sample the light from a small circular area of the field and the light collected will emerge from the other end of the fibre. In multi-fibre spectroscopy the input ends of a set of fibres are used to sample sky areas and objects in the field and the output ends are arranged in a line to form a psuedo-slit which feeds the spectrograph. The technique can therefore make use of existing long-slit spectrographs and well established, small-format detectors because the fibre-slit is changeless and totally independant of the field arrangement. Furthermore, objects from all over the telescope's field can be observed and all the spectral resolution options of traditional spectroscopy are available.

If the number of simultaneously observable objects is limited by the spectrograph (typically 50 to 100 spectra per spectrograph) then the greedy astronomer can simply use extra spectrographs. This presents no great technical difficulties as the telescope and the spectrographs do not need to be aligned in any particular way. Indeed, the engineering of off-the-telescope spectrographs is made considerably easier in the absense of differential flexure and standard optics laboratory equipment can be used. The total number of spectra that can be acquired at the same time can therefore be very large (of the order of 1000 say) depending on the telescope field, the number of spectrographs and the density with which the fibres can be placed in the field.

The technique requires considerable planning before the observations are made, just as for multi-slit work, no matter which method of placing fibres in the focal plane is adopted (see chapter 6). This preparation involves astrometry, assignment of fibres to targets and the process of actually setting-up the fibres into the required configuration. It is also important that the fibres are placed accurately on the objects. Figure 5.2 shows the variation of fibre efficiency as a function of positioning error assuming a gaussian seeing disk. The transmission shown is based purely on the geometry, i.e. it is essentially the vignetting of a circular aperture.

The theroretical S/N ratio that can be achieved, using fibres, based on photon statistics, can be derived for a spectral channel as follows. Assume that each fibre illuminates p detector pixels along the slit direction, Q_{TOT} is the total number of detected photons from the object and S_{fib} is the total number of sky photons detected for a single fibre. Thus for an object fibre we have a total signal and error of (using Poisson statistics)

$$Q_{TOT} + S_{fib} \pm \sqrt{Q_{TOT} + S_{fib} + pR^2} \qquad -(1)$$

where R is the readout noise in photo-electrons associated with a single pixel. If there are q suitable sky fibres (i.e. suitable in the sense that they sample the <u>same</u> parent sky population) then the total signal and its error from these is

$$qS_{fib} \pm \sqrt{qS_{fib} + pqR^2} \qquad -(2)$$

To obtain the sky subtracted signal and its associated error we divide expression (2) by q and subtract it from expression (1)

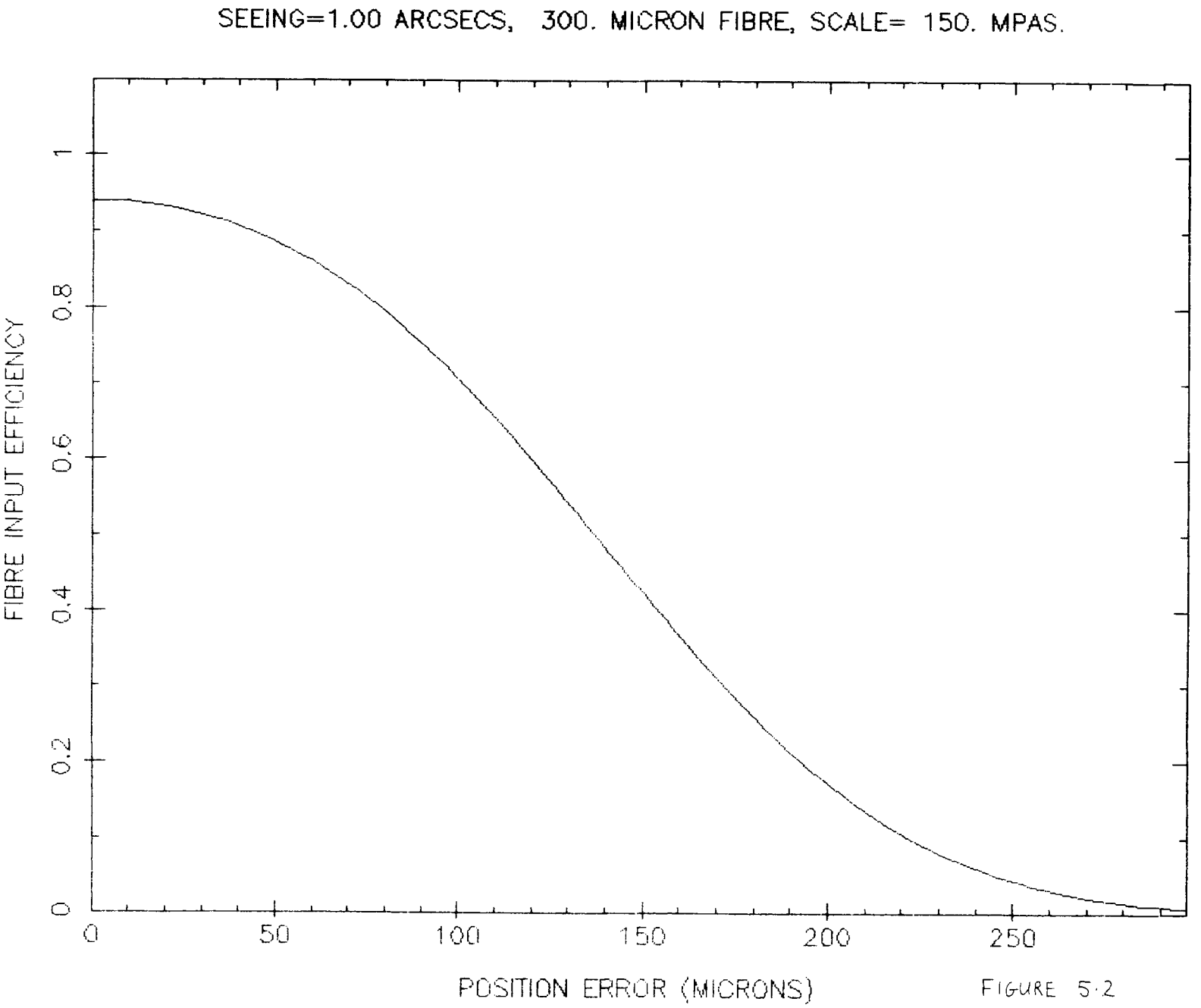

SEEING=1.00 ARCSECS, 300. MICRON FIBRE, SCALE= 150. MPAS.
FIBRE INPUT EFFICIENCY
0
0.2
0.4
0.6
0.8
1
0
50
100
150
200
250
POSITION ERROR (MICRONS)

FIGURE 5·2

to get

$$Q_{TOT} + S_{Fib} - S_{Fib} \pm \sqrt{\left[\tfrac{1}{q} S_{Fib} + \tfrac{p}{q} R^2\right] + \left[Q_{TOT} + S_{Fib} + pR^2\right]}$$

$$= Q_{TOT} \pm \sqrt{Q_{TOT} + (1/q + 1) S_{Fib} + \left(p + \tfrac{p}{q}\right) R^2}$$

$$= Q_{TOT} \pm \sqrt{Q_{TOT} + (\tfrac{1}{q} + 1)(S_{Fib} + pR^2)}$$

$$= Q_{TOT} \pm \sqrt{Q_{TOT} + p(\tfrac{1}{q} + 1)(S_{Fib}/p + R^2)}$$

Putting $S_{fib} = A_{fib} S$ where A_{fib} is the area of the fibre aperture in square arcsecs and S is the number of sky photons detected per arcsec for the channel concerned we get

$$\frac{S}{N} = \frac{Q_{TOT}}{\sqrt{Q_{TOT} + p\,(\tfrac{1}{q} + 1)(\tfrac{A_{Fib} S}{p} + R^2)}}$$

Comparing this with the equivalent expression for a slit (see section 2.1.1) we see (hardly suprisingly) that the result is basically the same. i.e. p=m, 1/q=m/n and $A_{fib}/p=A$. Thus if readout noise and photon statistics are the only source of error, fibre spectroscopy can go as deep as multi-slit spectroscopy. The longer integration times required because of fibre losses (say twice as long) are more than made up for by the fact that more objects can be observed simultaneously (perhaps 20 times more than with multi-slits).

However, there is some doubt as to what the actual limiting magnitude of fibre optic spectroscopy is in practice, due to the presence of systematic errors associated with the sky-subtraction process. Firstly, unlike slit work, the sky "under" the object signal is generally not estimated from the sky adjacent to the object. The S/N expression derived above assumes that the sky sampled by the sky fibres comes from the same parent population as the sky from the object fibre(s).

Does this assumption become less and less valid as the distance between the object and sky fibres increases? To answer this question a longslit FOS expoure was analysed to see whether the error in subtracting one sky spectrum from another increased with the distance between them. The analysis was carried out upto a separation of 168 pixels which corresponds to about 130 arc seconds on the sky. As there was a slight variation in the number of detected sky photons along the slit, the sky being subtracted off was normalised so that the net signal in the result was exactly zero. The sky spectrum at one end of the slit was successively sky-subtracted using all the other sky spectra in turn and the rms noise for each result was calculated. The analysis was performed twice, subtracting from the spectra at opposite ends of the slit.

Figure 5.3 shows the standard deviation of all the spectral channels in the full, first order wavelength range for the sky-subtracted sky spectra, as a function of the distance between the sky spectra. The large spikes are due to CREs or objects contaminating the sky spectrum being subtracted off. The important thing to look at is the lower continuum of the curve which clearly indicates that the sky subtraction process gets noisier as the distance between the sky areas gets larger. This trend was also evident when the analysis was performed in the reverse direction. Whether the result is caused by varying background light to do with the instrument or is caused by a genuine variation of the sky background is not clear, but the fact that sky subtraction gets worse with separation is ominous for fibre work whatever the cause. Another, more likely explanation is that the sampling of the sky spectrum varies

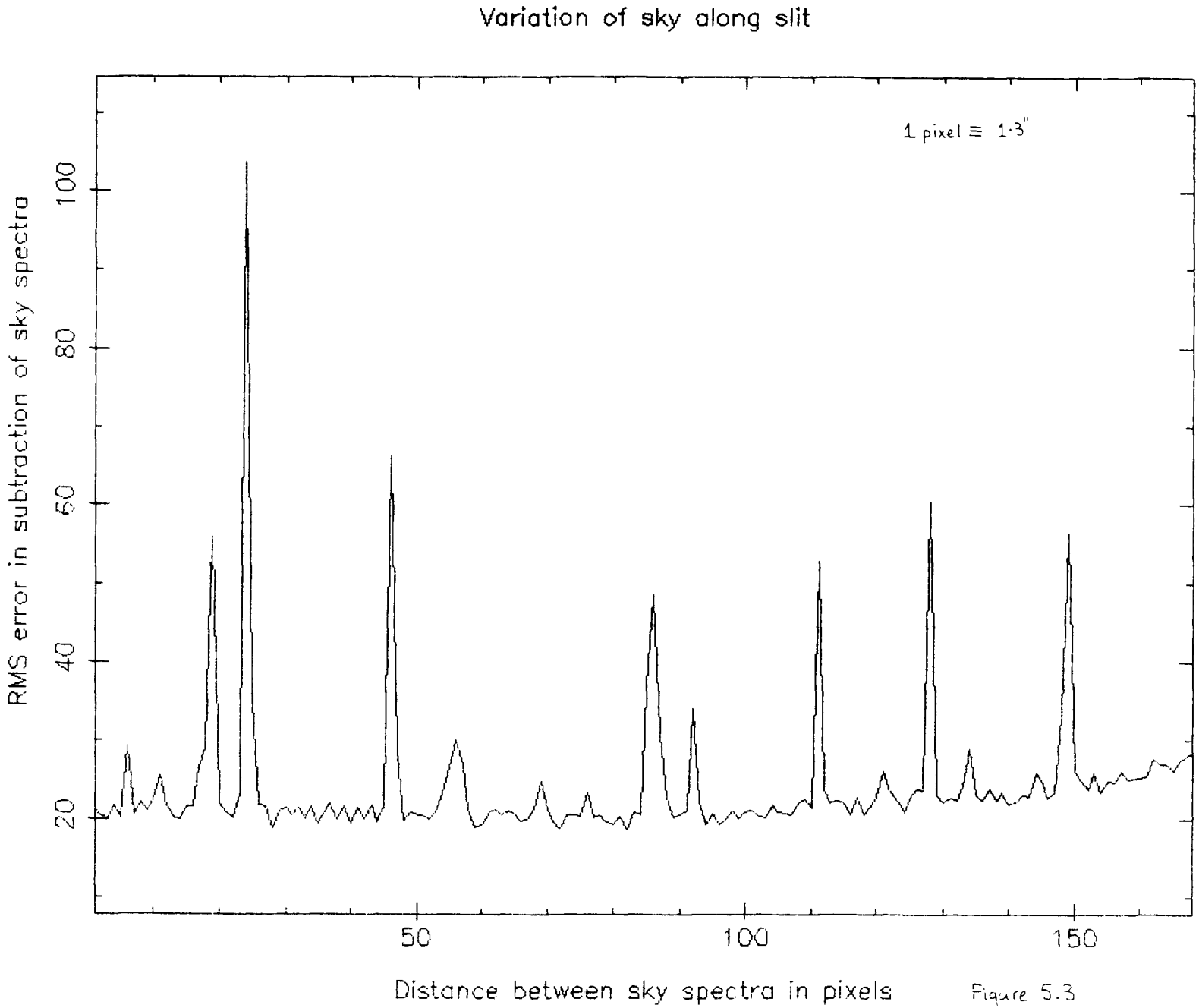

Variation of sky along slit
1 pixel ≡ 1·3"
RMS error in subtraction of sky spectra
20
40
60
80
100
50
100
150
Distance between sky spectra in pixels

Figure 5.3

along the slit even though the slit and the detector grid are parallel to a high degree of accuracy for the FOS data used to derive figure 5.3. This effect is likely to be worse for fibre "slits" which may not lie accurately on a straight line as well as being non-parallel with the detector.

The second problem associated with fibre sky-subtraction is that the overall transmission varies from fibre to fibre. In principle this can be determined for a given fibre arrangement by uniformly illuminating the telescope field (e.g. using twilight sky). For very faint work the relative overall transmissions of the fibres being used has to be estimated very accurately because the object signal is only a small fraction of the sky signal. If the subtracted sky is only slightly wrong it can contribute significantly to the result and what is more such errors do not get smaller as the integration time increases. Further work is required to investigate methods of accurately determining relative fibre transmissions and to ensure that adequate uniform lighting conditions are available. Also, improvements in the design and manufacture of fibre optic couplers might ultimately lead to a fibre transmission uniformity comparable to the transmission uniformity along a slit (see chapter 6).

Finally, the use of relative overall transmissions assumes that the relative wavelength transmission is constant from one fibre to the next. Although some laboratory work has been done which supports this assumption (Peter Gray, private communication) further investigation is desirable because of the highly accurate fibre calibrations needed for very faint

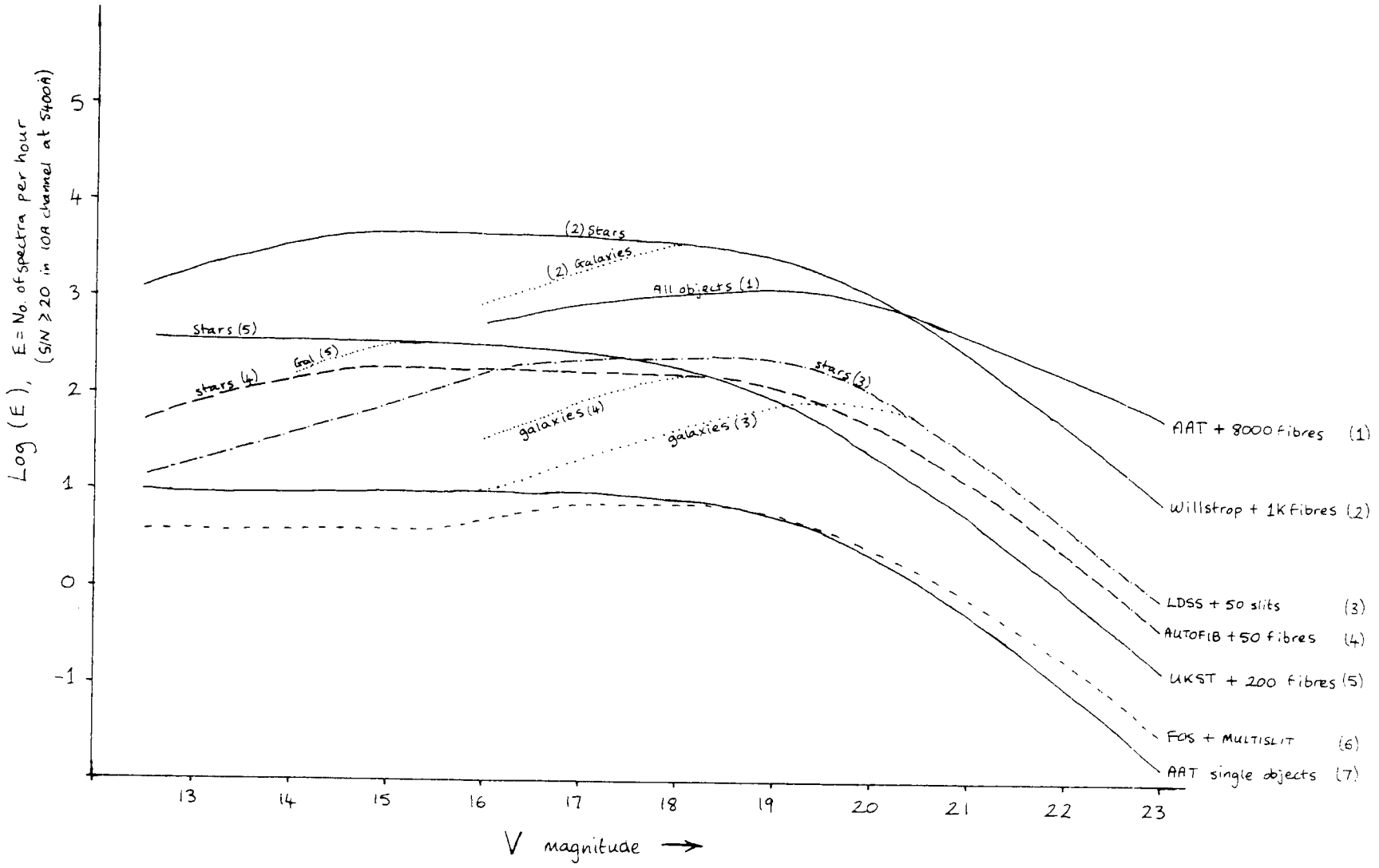

Log (E), E = No. of spectra per hour (S/N ≥ 20 in 10A channel at 5400A)
5
4
3
2
1
0
-1
13
14
15
16
17
18
19
20
21
22
23
V magnitude →
(2) Stars
(2) Galaxies
All objects (1)
Stars (5)
Stars (4)
Gal (5)
Stars (3)
galaxies (4)
galaxies (3)
AAT + 8000 Fibres (1)
Willstrop + 1K Fibres (2)
LDSS + 50 slits (3)
AUTOFIB + 50 fibres (4)
UKST + 200 fibres (5)
FOS + MULTISLIT (6)
AAT single objects (7)

Figure 5.4

work.

5.7 Discussion

Of the above five methods, only the last two (multi-slits and multi-fibres) stand out as being able to provide large numbers of reasonably high resolution spectra (e.g. for redshift determination) at faint magnitudes (B > 17). Figure 5.4 shows the efficiency of various multiobject facilities as a function of limiting magnitude, for spectra with a signal-to-noise not less than 20 in a 10A channel at 5500A. The efficiency parameter E is defined as the number of spectra that can be obtained per hour with a signal to noise C for objects brighter than a limiting magnitude V. In deriving the E - V curves the following formulae and assumptions were made. The flux at the top of the atmosphere f_{55} for an object of magnitude V is

$$f_{55} = 10^{-(0.4\,V + 8.43)}$$

where f_{55} is the flux at 5500A and is in ergs/sec/cm^2/A. The rate at which photons are detected from the object, R_{obj}, in a 10A spectral channel at 5500A is given by

$$R_{obj} = 8.64 \times 10^{14}\ f_{55}\ D^2$$

where D is the diameter of the telecope in metres. This expression was derived assuming that 4% of the available photons at the top of the atmosphere are actually detected. This is in fact the figure achieved in practice for the FOS (see chapter 4) and includes atmospheric losses, telescope central obscuration, reflection losses, spectrograph losses and the

detector's quantum efficiency. If we assume that the sky magnitude in the V band is V=21.1 per square arc second and that 3 square arc seconds worth is mixed up with the object signal (a 3x1 arc second slit or a 0.977 arc second diameter fibre) we get

$$R_{sky} = 0{\cdot}035\ D^2$$

where R_{sky} is the number of photons detected per second from the sky. A signal-to-noise ratio C will be obtained in t seconds given by

$$C = \frac{R_{obj}\ t}{\sqrt{R_{obj}\ t + 2R_{sky}\ t}}$$

using S/N expressions for slits and fibres (see sections 2.1.1 and 5.6) with m=n=p=q=1 and R=0. Conversely an integration time of T_{int} hours is needed to get a S/N ratio of C given by

$$T_{int} = \frac{C^2\,(R_{obj} + 2R_{sky})}{R_{obj}^{\,2}} \Big/ 3600$$

When observing a field of objects the elapsed time from the start of observations of one field to the next is simply

$$T_{tot} = T_{int} + T_{dead}$$

where T_{dead} is the time spent changing the fibre configuration or multi-slit aperture plates and acquiring the new field etc. in hours. Finally, we define E by

$$E = \frac{NA}{T_{tot}}$$

where N is the number of objects available to limiting magnitude V per square degree and A is the accessible field area in square degrees. Note that E does not make any

allowance for the fact that telescope time is easier to come by on telescopes of smaller aperture. Also, NA must lie between 1 and N_{max} where N_{max} is the maximum number of simultaneously available spectra due to practical constraints such as the number of fibres available or the detector size etc. NA is otherwise based on the number counts of Tyson and Jarvis, 1979. The curves in figure 5.4 were derived using the following instrumental parameters:

Equipment	T_{dead}	Field $(^\circ)^2$	Nmax	D (m)	Thruput 5500A
AAT all objects (8K fibres)*	0.25	0.35	8000	3.9	0.04
Willstrop + 1000 fibres	0.20	7.0	1000	3.9	0.04
AAT + LDSS	0.10	0.04	50	3.9	0.08
AAT + AUTOFIB	0.25	0.35	50	3.9	0.04
UKST + 200 fibres	0.50	28.0	200	1.2	0.04
INT + FOS + multislit unit	0.25	0.0033	10	2.5	0.04
AAT single object mode	0.10	-	1	3.9	0.08

Note that the throughput of the AAT with both LDSS and single slits was set to 8% to allow for the absence of fibre losses.

Generally speaking figure 5.4 tells us that wide telescope fields are useful for bright magnitude limits, especially for stellar work, and that for faint surveys the aperture of the telescope and the total number of fibres or slits govern the efficiency because even small fields have plenty of potential targets. It is interesting to note that the proposed Willstrop telescope** with its enormous 7 square degree field can be matched in efficiency with 1000 fibres by the AAT for magnitude

* 8000 objects available in AAT field with $V \lesssim 23$
** Willstrop R. V., MNRAS 216, 411, 1985.

limits fainter than V=20. In other words, existing telescopes provide much more potential targets than can be acquired at faint limits and wider fields would require enormously powerful MOS instrumentation to be justified. On the other hand, for brighter magnitudes ($16 < V < 20$) and for selective surveys where a particular class of object has a low sky density, the Willstrop telescope would be very useful. It is also interesting to compare the LDSS and AUTOFIB curves. For galaxies LDSS is more efficient than AUTOFIB for $V > 19.5$ but this is only because LDSS has twice the throughput of AUTOFIB. Therefore, if AUTOFIB can provide more than 100 object fibres it will be more efficient than LDSS.

In conclusion, if the problem of inaccurate sky-subtraction can be solved, multi-fibre spectrocopy is clearly the most powerful technique of MOS at all magnitudes as it can exploit the full telescope field, provide the highest potential for the number of simultaneously acquired spectra and allow the astronomer a great range of dispersion and wavelength coverage options. However, at present the sky-subtraction problem has not been solved and the multi-slit method seems the most powerful at faint magnitudes where small field coverage and low dispersion are not particulary disadvantageous.

Chapter 6

Fibre optic couplers and the development of AUTOFIB

This chapter is about an astronomical instrument which at the time of writing is being built as a collaboration between Durham University and the Anglo-Australian Observatory (AAO). The instrument is an automated fibre optic coupler for multi-object spectroscopy and has been designated AUTOFIB. This chapter describes the development of AUTOFIB from its initial conception right through to the final design and includes references to other instruments which have influenced the design process. AUTOFIB will be available at the AAT's Cassegrain focus by the end of 1986.

6.1 Why do we need to automate?

Having established that multifibre spectrocopy is a worthwhile technique one then has to decide exactly how the objects in the field of the telescope are acquired with a set of fibres. A simple technique is to plug the fibres into the holes of a suitable aperture plate, mounted at the telescope focus. An excellent example of this method is the AAT's FOCAP system which is shown in Figure 6.1. The plate is machined some time in advance of observing using the measured co-ordinates of the field objects and the image scale of the telescope. Another method* is to use a photographic plate of the target field taken at the same telescope focus as is used for fibre observations and to prepare a glass contact negative on which the objects are burnt out and therefore transparent.

* e.g. the UKST fibre system. See Watson F. G. and Dawe J. A., ROE preprint # 17/84

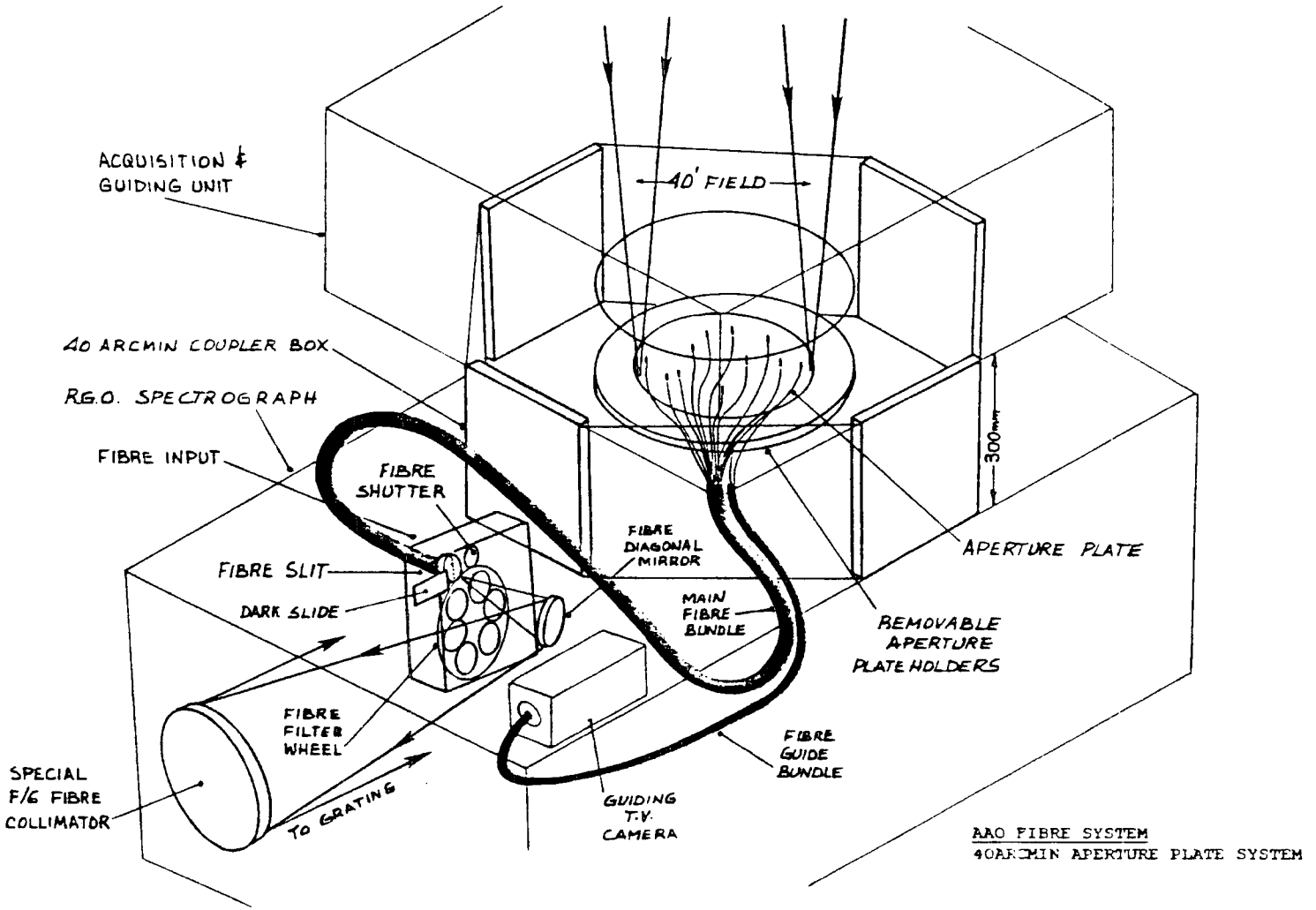


Figure 6·1 - FOCAP

The fibres can then be glued directly on to these images and the plate mounted on the telescope without the need for astrometric measurements. Both methods are relatively easy to get started and allow the scientific goals to be achieved. However, each observed field requires that a certain amount of manufacturing be done and so these methods of observing continually cost time and money and the heavy handling received by reused fibres means the quality of such fibre systems is often less than optimal. Clearly, a method of positioning fibres that involves a high degree of automation is worthwhile simply on the grounds of economy and convenience as long as it does not compromise the scientific potential:- the counterclaim that automation is unnecessary because it does not open up any new scientific possibilities is much like saying "Why build a boat when you can swim the channel?"

The AAO have been leaders in the field of multifibre spectroscopy for several years now and the acquisition of AUTOFIB is seen as a natural step forward, building on from their aperture plate system, FOCAP. FOCAP can employ either the auxilliary Cassegrain focus or the full 40 arcmin diameter main Cassegrain focus by lowering the RGO spectrograph (which normally sits directly under the A&G unit) by 300 mm and inserting a box which supports the spectrograph in its new position and houses a removable aperture plate holder. By having 2 fibre bundles, each of which feeds 50 fibres to the spectrograph, 2 guide bundles which feed the acquisition TV and 2 aperture plate holders, it is possible to prepare one aperture plate whilst observing with another and so keep the field change-over time to a minimimum. It takes about 15 to 20

minutes to park the telescope at the zenith, enter the cage, exchange aperure plates and acquire the new field. A record of which fibre is used for which object has to taken manually at the same time as the fibres are plugged into the aperture plate. For further details see Gray 1983 and Gray & Sharples 1985.

I have already justified the need for an automated fibre coupler in general terms. Now I compare the relative merits of two specific systems. I make the comparison as an astronomer - not an engineer - for it is the quality of the obtainable astronomical data that determines the better technique.

a) AUTOFIB is far cheaper to run than FOCAP. It costs the AAO A$170 on average for a 40 arcmin aperture plate and something like A$12,000 per annum. Furthermore specialist staff are required at the telescope when FOCAP is used and the system's popularity, and therefore usage, continues to rise. On the other hand AUTOFIB's running costs and staff requirements will be the same as for conventional slit spectroscopy. The major costs involved with AUTOFIB are those of building and commissioning which should not amount to more than £20,000 (£1=A$2 at time of writing).

b) From the astronomer's point of view AUTOFIB will be much easier to use than FOCAP. Target lists need not be prepared so far in advance and a complete observing log including the vitally important information as to which spectrum corresponds to which object will be provided automatically by the AAO VAX.

c) The quality of the fibre bundles should be better with AUTOFIB. With FOCAP the fibres are constantly being plugged into and pulled out of aperture plates. Variations in the hole diameters inevitably leads to the occasional fibre ferrule fitting too tightly and an undesirably large force has to be used to unplug it. Fibres are therefore broken every so often particularly since the staff concerned are working long late hours. With AUTOFIB however not only will the fibres maintain their original quality due to lack of wear and tear but because they are more protected greater care can be afforded in their manufacture making their initial quality as high as possible. In addition, the minimal stressing that the fibres will receive will lead to less focal ratio degradation and therfore better overall throughput and less fibre to fibre variation.

d) AUTOFIB will position the fibres onto the objects with higher precision than FOCAP. There are two ways in which a fibre can be misaligned; by positional misalignment, laterally away from the incoming light cone, and by angular misalignment of the fibre's axis with respect to the axis of the incoming light. For the purpose of comparison I am ignoring astrometric errors since they are common to both fibre systems. For FOCAP the positional alignment is dependant on the accuracy with which the holes are drilled ($\pm$ 30 microns) and the concentricity of the fibre and its ferrule ($\pm$ 15 microns). The total misalignment can therefore be as much as 45 microns. The angular misalignment can be as much as 3 degrees for a loosely fitting ferrule and if the aperture plate is stressed so that its shape approximates the focal surface of the AAT the total misalignment can be as much as 4 degrees. This much

misalignment reduces the throughput into the spectrograph entrance pupil by a factor of 0.75. AUTOFIB will be able to position the fibre apertures to within 20 - 30 microns and with angular misalignments of less than 0.5 degrees (see section 6.4.2). The reduction of these misalignments, in particular the angular one, will provide better sky subtraction for faint objects and therefore clearer results.

e) AUTOFIB offers much more flexibility whilst observing than FOCAP. The posibilities include real time correction of zenith-distance dependant field distortions and swapping a fibre on a sufficiently well observed bright object to another whilst continuing the exposure on the other fainter objects.

f) Bright object surveys are far easier with AUTOFIB. This is simply because field change-overs are easier. For example, a project which involves a dozen or so bright object fields would be easy with AUTOFIB but would be exhausting for the observer and his support staff using FOCAP because of the non-stop activity.

g) Finally, as AUTOFIB is so easy to use it opens up possibilities for the future such as remote and service multi-object observing.

On the other hand FOCAP has some advantages over AUTOFIB though on balance these do not in my opinion outweigh the disadvantages. Also, one must bear in mind that AUTOFIB has yet to be built and the above comparison is based on its expected, as apposed to its actual, performance. One advantage of FOCAP is that the positioning of fibres onto objects is less

restricted, in terms of the general field arrangement, and fibres can be placed as close as 19 arcsecs apart (centre to centre) compared with 28 arcsecs for AUTOFIB. Another advantage is that FOCAP's aperture plates can potentially handle greater numbers of fibres than AUTOFIB although at present the maximum number of fibres that can be used is limited by the length of the spectrograph slit. Finally, FOCAP is much simpler and therefore less likely to go wrong and easier to put right if something does go wrong. In order to keep all options open AUTOFIB has been designed so that rapid change-overs to the aperture plate system are possible because it is inevitable that it will be more suitable for a small percentage of projects. This flexibility will also be used as a backup when AUTOFIB is being commissioned but not generally as the cost of making aperture plates for backup purposes only is not justifiable.

6.2 Methods of automation

Having established that multi-fibre spectrocopy would benifit greatly from automation, we are faced with the problem of inventing an instrument that not only offers user convenience but at least matches the data gathering power of an aperture plate system. This problem had been addressed by other workers prior to the summer of 1983 when I first gave it serious consideration. Their methods all involved the use of a precision positioning device per fibre to both deliver and hold it in the required position. Some of these schemes offered very limited choice in terms of object selection because the rather small number of fibres available were each confined to

their own non-overlapping patrol areas (e.g. Cohen et al 1981). By far the best scheme suggested upto that time was that of John Hill and his collaborators at the Steward Observatory (Hill et al, 1980 & 1983; Hill, 1984) which uses 32 probes to position 32 fibre pairs. This arrangement has been likened to anglers sitting around a pond with their fishingrods extended over the pond (see figure 6.2). Their system, which they have named "the MX spectrometer", has now been constructed and at the time of writing is undergoing tests on the Steward Observatory's 2.3 metre telescope. Although their scheme seemed a reasonable one to adopt I felt that a better alternative must exist, if only on economic grounds, for MX is quite expensive.

My first thought was that the cost would be considerably less if there was only one positioning system which delivered all the fibres individually to their designated positions. This means that the fibres have to have some other mechanism to hold them in place once they had been set in position. This approach is actually the same as is used with aperture plates - a technician delivers the fibres and the aperture plate makes sure they stay there. However, it is not practical to replace the technician with an equally intelligent robot and it is difficult to construct an aperture plate in which the holes move about! A reusable fibre holding scheme that can be set up by a simple pick-and-place robot is required. The obvious positioning system that came to mind was an X-Y carriage which carries a manipulator head that can engage and disengage individual fibres. Thinking up a means of holding the fibres in place proved much harder. After considering everything from

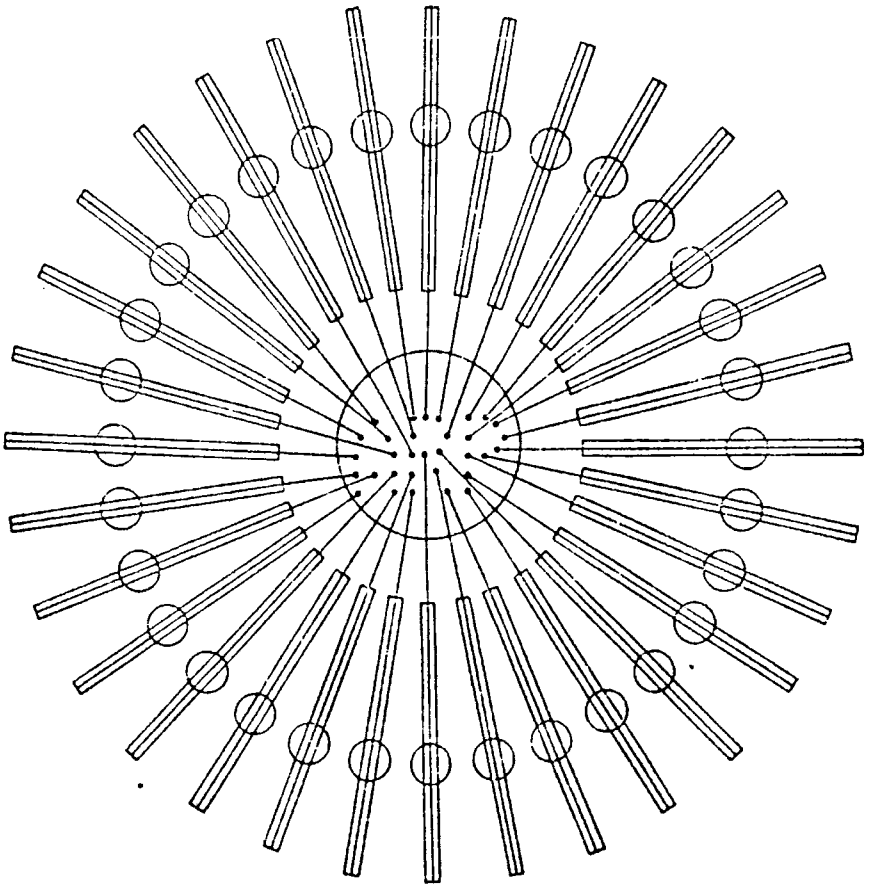


Figure 6·2 – MX concept

fibres held onto a glass plate by suction, to fibres poked through Blue Tak sheets, I eventually came up with the idea of lying the fibre along a steel plate, putting a tiny prism on the end of the fibre so that it accepted light traveling normal to the plate and securing the fibre tip to the plate by means of a small permanent magnet (see figure 6.3). It is of course necessary to package the fibre, prism and magnet in some way so that the positioner can pick it up in precisely the same way every time. The package (or button) type used in AUTOFIB is made of a ferromagnetic material such as steel and has a conical hole in it to both accept light from the telescope and to enable it to be located precisely on the pickup head by a conical spigot. The pickup head itself is an electromagnet that can grip the button with more force than the button can grip the field plate. The complete arrangement of fibres, positioner and spectrograph is shown schematically in figure 6.4. Incidentally, fibre probes of this type not only lend themselves to automation but also can be set up manually using any standard X-Y coordinate measuring machine such as is available at most observatories. This would then remove the need for expensive aperture plates and an expensive robot positioning system and is probably the cheapest way of doing multi-fibre spectroscopy of all.

My scheme of automation as employed in AUTOFIB has several advantages over the MX system;

a) AUTOFIB is less expensive than MX. AUTOFIB's total component costs will be about £15,000 whereas MX's material costs were about US$60,000 out of a total budget of US$250,000

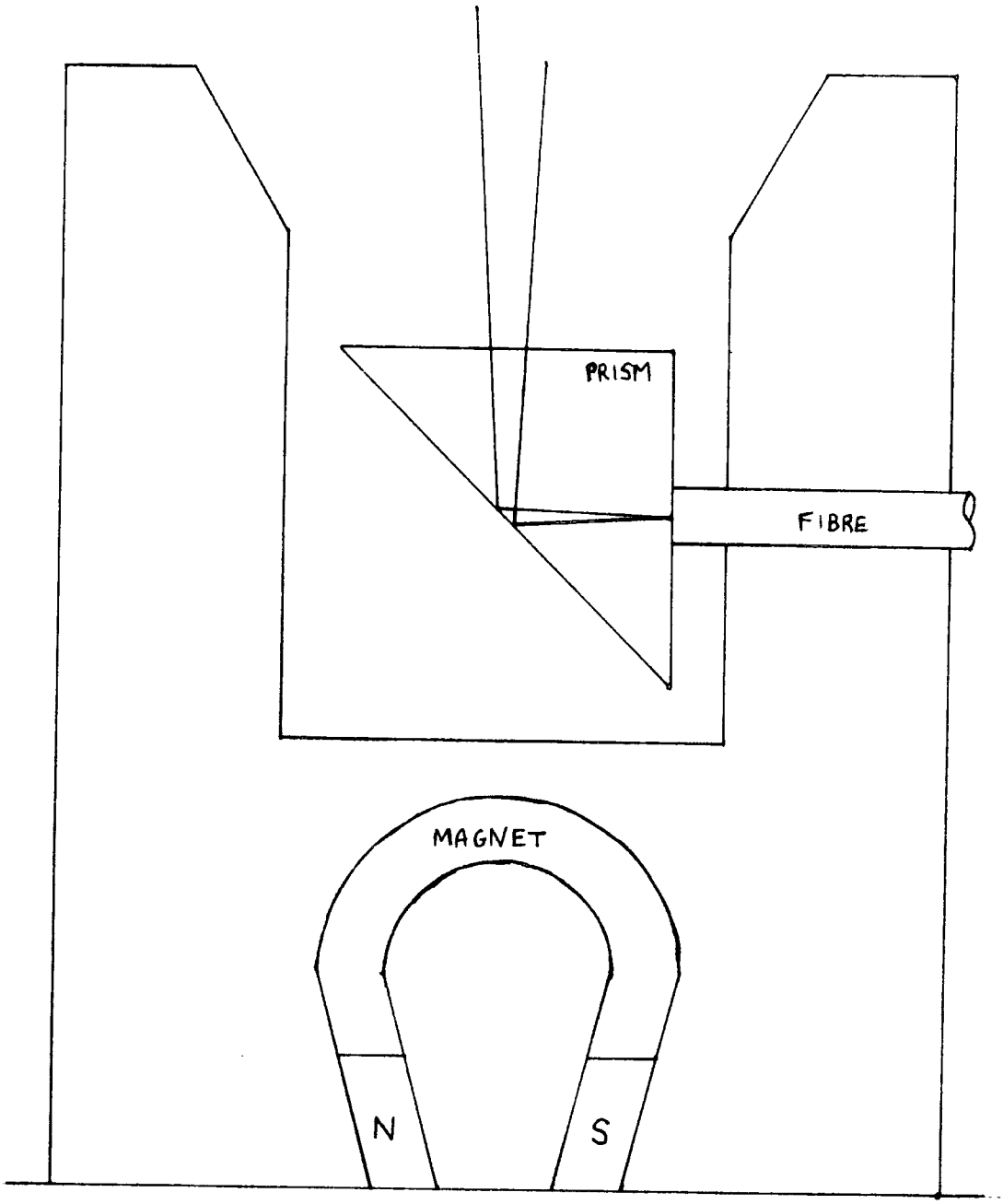


Figure 6.3 - Fibre button concept

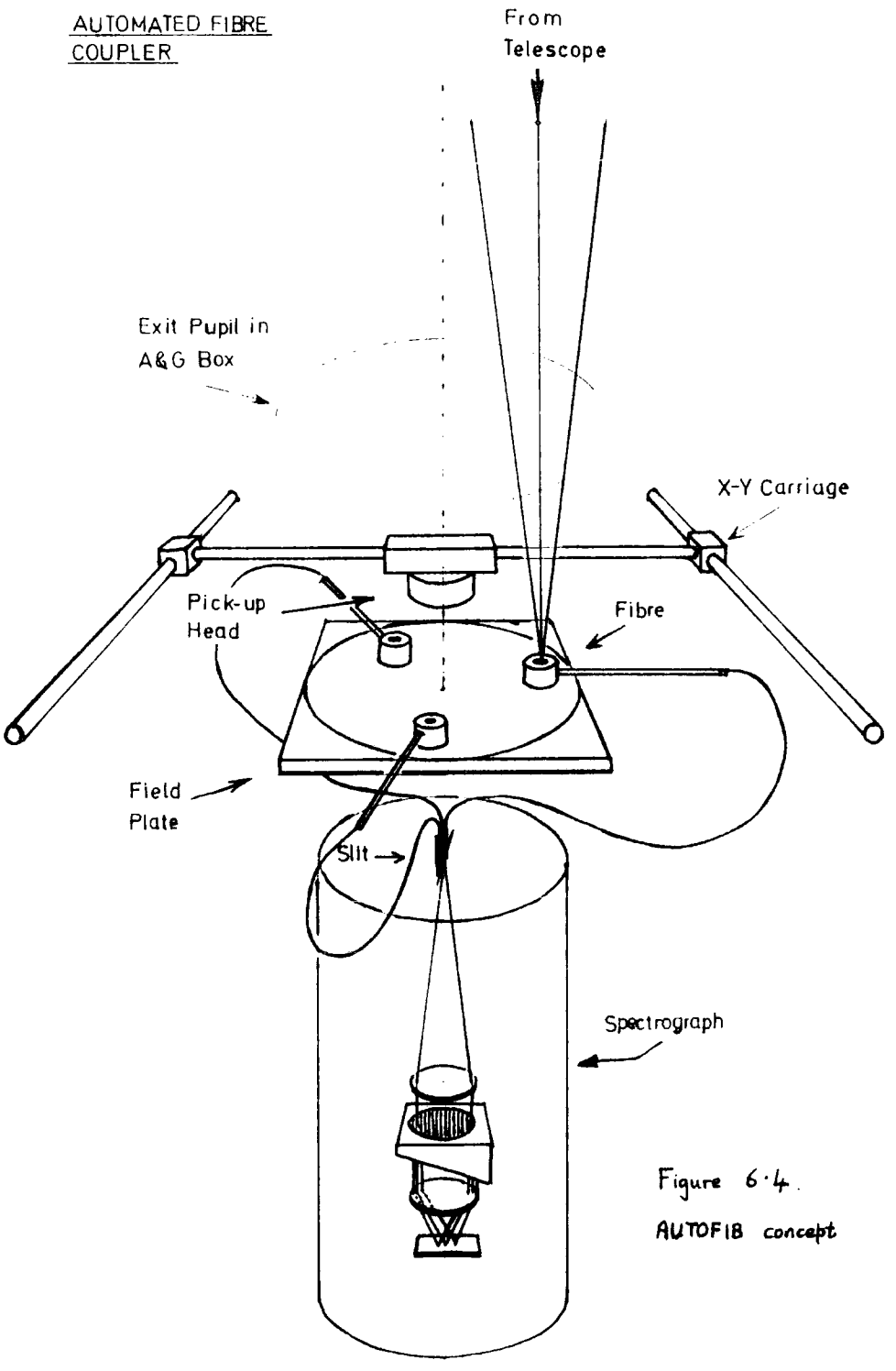


Figure 6·4.
AUTOFIB concept

(£1 = US$1.4 at time of writing). Also for these costs AUTOFIB will be built using components of the highest quality with very little compromise whereas the quantities involved with MX have forced Hill and his collaborators to cut corners in order to stay within their budget (e.g. MX does not use encoders but employs open loop step counting for positioning).

b) AUTOFIB's positional accuracy is superior. Each MX probe has two degrees of freedom; radial motion and angular rotation about a pivot outside the field. The accuracy of an MX probe is therefore strongly dependant on its position in the field because the angular positional error is proportional to the radial position and the tip of the supporting rod that holds the fibre flexes as the cube of the radial position. Hill has used 9 mm thick rods to reduce rod flexure to a minimum and anti-backlash mechanisms to eliminate play. The main sources of error are the non-linearity of the gear teeth and the eccentricity of the cogs. Ideally all the probes are identical but this is unlikely in practice. It is possible that a probe can be misplaced by as much as 50 microns (0.5 arc-seconds on the Steward 2.3 m telescope) at the centre of the field. The accuracy of AUTOFIB's X-Y carriage will be $\pm$ 10 microns.

c) MX's probes have to be orchestrated so that they all position their fibres in the same coordinate frame. In principle this is just an inconvenient calibration procedure that has to be applied but in practice the calibration will depend on telescope attitude (flexure) and temperature. Furthermore, the calibration has to be made with the instrument

on the telescope. Any inaccuracies in establishing this will lead to positioning errors in addition to those decribed in (b) above. At the time of writing it is this difficulty that is causing problems for Hill and his collaborators. The problem of orchestrating AUTOFIB is trivial of course as there is only one positioner and flexure is far less of a problem. AUTOFIB even has the safety net of being able to position all the fibres with the telescope pointing at the zenith which will eliminate differential flexure entirely.

d) The approach adopted in AUTOFIB can deal with focal plane curvature far more easily than the MX approach. This is again due to the fact that AUTOFIB has only one positioner. Incorporating two extra degrees of freedom into the positioner to accomodate the tilting required to deliver a fibre onto a curved field plate along a line normal to the plate is much easier than adding three extra degrees of freedom (position along the telescope axis and two tilts) to each and every MX probe.

6.3 The Prototype Coupler

Once the basic idea had been concieved, my collaborators and I were keen to see the instrument built and installed on a telescope. The AAT was the obvious telescope to use because a great deal of practical experience in using fibres for multi-object spectroscopy had already been gained there. Also, the AAT has a large aperture, a wide and physically large field of view and excellent pointing and tracking. After two proposals had been submitted to ACIAAT (the Advisory Committee

for Instrumentation on the Anglo-Australian Telescope) in February and August of 1984, including a visit to the AAO by Richard Ellis and I to finalise the second proposal with Peter Gray. ACIAAT decided that some preliminary laboratory work was necessary, to demonstrate the feasibility of the idea, before they could fund the project. ACIAAT agreed to fund the prototype. Since the bulk of the cost of a fully working instrument is in the computer controlled X-Y positioning system and ACIAAT's doubts were confined to the other, less common-place, aspects of the proposed system, it was possible to construct a low cost prototype system to investigate these grey areas. In particular it was important to establish that the fibre probes could be manufactured successfully, that the magnetic forces involved were adequate, that the pick up head operated with suficient precision, that the fibre probes could be handled in close proximity to each other without detremental interference and that the operational speeds involved were fast enough. The prototype also provided invaluable help when it came to designing AUTOFIB for the AAT. The following paragraphs describe the prototype and the results obtained.

6.3.1 The prototype fibre probes.

Figure 6.5 is an exploded view of a complete fibre assembly. It consists of a 2.5 metre length of QSF-AS 200/300 fibre with one end polished and terminated in two short hypodermic tubes, a 1 mm prism, a machined silver steel button, a SmCo (samarium cobalt) disc magnet and a 400 mm long length of hypodermic tube. The geometrical requirements of this assembly are that the virtual image of the fibre aperture is

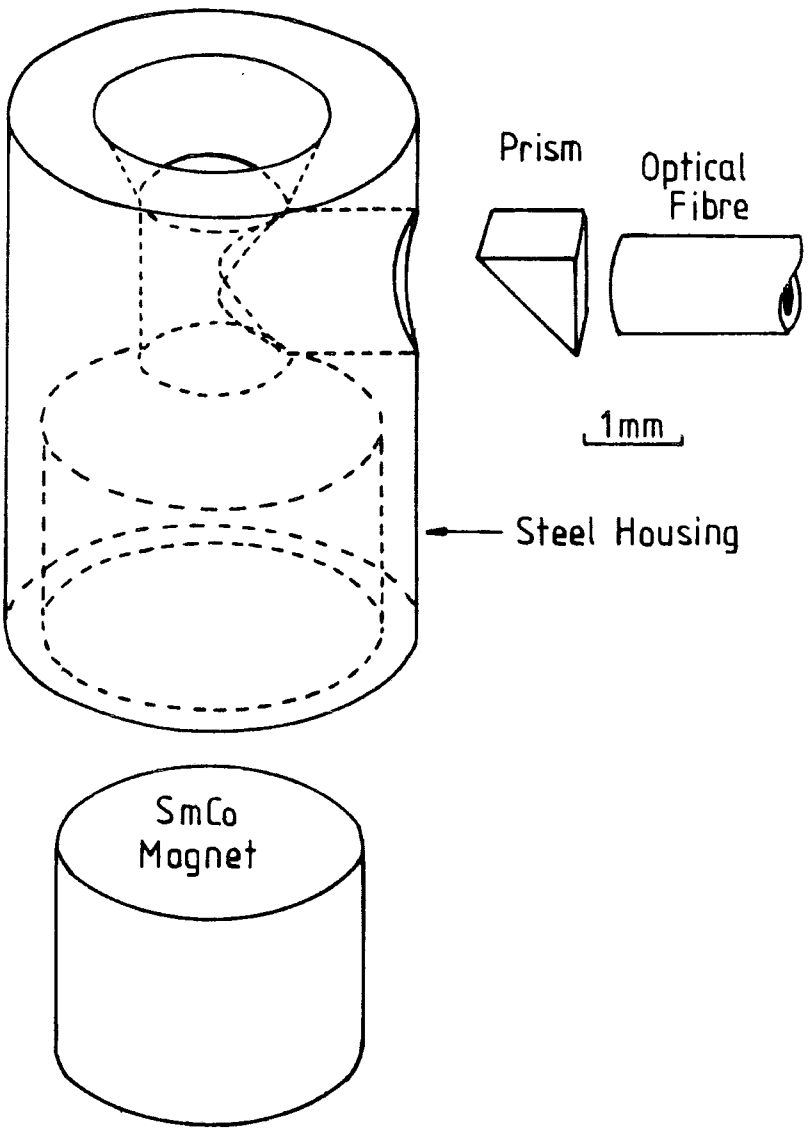


Figure 6·5

Fibre button design.

concentric with the conical hole of the button and lies parallel to the bottom surface of the button at a fixed distance from it. Further requirements are that the button should be attracted strongly to any flat steel surface on which it is placed, exhibit no play when on the spigot of the electromagnet and be very strongly attracted to it when current flows in the electromagnet's coil. Several design variations were considered and manufactured and ultimately the design shown in figure 6.5 was arrived at, which met the operational requirements and yet was relatively simple to mass produce.

Silver steel, as a corrosion-resistant, magnetic material conveniently available in 4 mm rod form was chosen for the construction of the buttons. Firstly, the rod was parted off into 5.1 mm lengths. Then all the end faces of the resultant cylinders were ground flat which brought the length of the cylinders down to 5.0 mm. Next each cylinder was mounted in a simple jig in the jaws of a lathe chuck and a standard centre drill, mounted in the tail-stock, was brought up until it met a micrometer end-stop. This stop was to ensure that the hole was drilled to a precise depth as any error at this stage would have lead to fibre positioning errors. The micrometer end-stop was adjusted to the exact setting required by examining the first few attempts with a microscope and measuring the diameter of the top of the tapered hole with the microscope's micrometer stage. When this dimension exactly matched that of the spigot on the electromagnetic pick-up head it was safe to proceed with the drilling of a large batch. Afterwards, it was still necessary to examine all the drilled buttons in order to reject those that were not to standard. In the batch made, about

20% had to be rejected. Note that the shape of the centre drill provided the tapered hole and the recess for the prism in one drilling operation. The next operation was to turn out a flat bottomed hole in the base of each button to accept the SmCo magnets. This was also done in batch mode using an end-mill mounted in the tail-stock. There was no need for high precision at this stage and so no rigid quality control procedure was required. The SmCo magnets were then epoxied in and when firmly set within their holes the buttons were returned to the grinding machine to finish off their bottom surfaces. This operaton was required because the thickness of the SmCo discs, as supplied by the manufacturer, varies by $\pm$ 0.2 mm. The final machine operation was to drill a clearance hole in the side of each button to accept a fibre. Another simple jig and an end-stop were enough to ensure adequate quality control.

Having manufactured the buttons it was then necessary to terminate the fibres and incorporate them into the buttons. In actual fact only one complete fibre was made (others were made without prisms) to test out the manufacturing scheme and provide a fibre to carry out positioning tests with. Firstly, the protective plastic cladding of the fibre was removed and the core was epoxied into a short length of gauge 23 hypodermic tube which in turn was epoxied into another slightly longer piece of gauge 19 tube. This brought the outside diameter of the end up to 1.1 mm which provided sufficient surface area to take a 1 mm prism. The end was then polished using a standard fibre optic polishing kit to give a good optical face, normal to the fibre's length. The prism was then mounted on this face

using optical cement and a pair of tweezers. A low power microscope aided inspection at this stage. Finally, the fibre and the button were brought together in a specially made jig mounted under a microscope by making the virtual image of the fibre's entrance aperture precisely concentric with the tapered locating hole of the button. The button and the fibre did not actually touch whilst being manipulated in the jig. Not until a small amount of epoxy resin was introduced into the gap between the side hole of the button and the fibre was there any contact. The fibre and the button were left in the jig until the epoxy had set. A relatively fast setting epoxy was used (with a work time of about 15 minutes) and it was difficult to make it go exactly where it was required with the microscope in its usual orientation. The jig was therefore removed from the microscope stage to apply the epoxy. However, this meant that the precise location of the fibre within the button could not be monitored during the application of the epoxy and although the fibre had been centred accurately beforehand it was found to be off centre by 80 microns in a direction normal to the fibres length when the epoxy had set. When the epoxy was applied great care was taken to ensure that none ended up on either the top or the reflecting surfaces of the prism (an unsilvered prism which relies on total internal reflection was used). UV curing resins would have been much more suitable for this purpose. To complete the fibre a 400 mm long, gauge 17 tube was slipped along the fibre until it passed over the gauge 19 tube.

The special jig used to align a fibre within a button was made in our workshop and essentially consisted of an X-Y-Z displacement unit which could be stood on the X-Y micrometer stage of a microscope. The performance of the dovetail slides was adequate but an improved version could have been made from optical bench accessories. A concentric circle eyepiece graticule proved a useful aid when setting the X and Y position of the fibre and the small depth of focus of the microscope was used to set the Z (height) position. The rotation angle of the prism about the fibre axis was set by ensuring that the top of the prism was always in focus as the microscope stage was traversed perpendicularly to the fibre's length. However, the jig incorporated no fine control to adjust this rotation. Precise alignment of the fibre to within 2-3 microns of the button's axis could be achieved in less than 10 minutes.

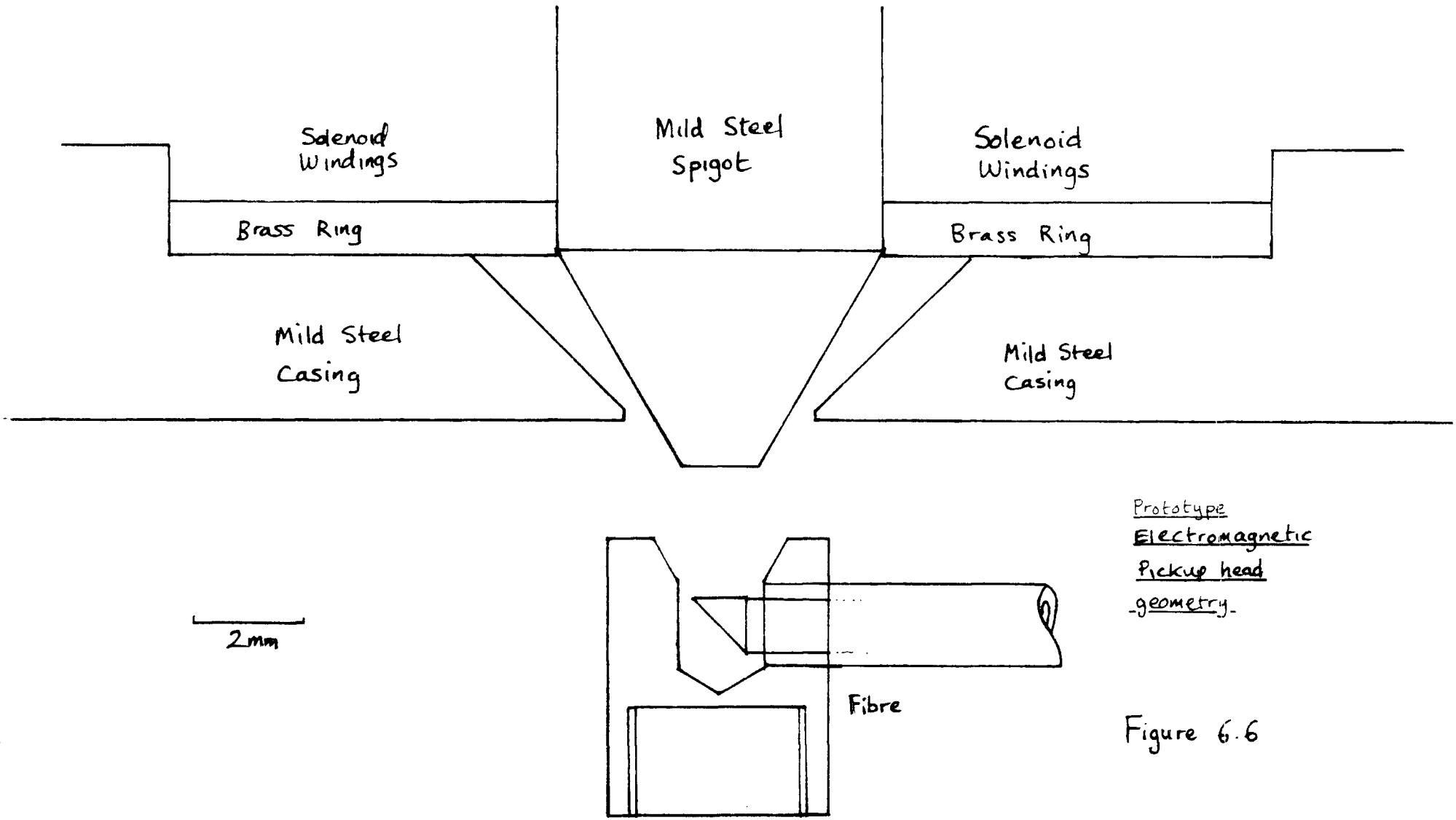
When the button end of the completed fibre was placed on a set of scales and the other end of the 400 mm tube was lightly supported, a weight of 1.5 grams was registered. On the other hand when a button was placed on a steel surface it was found that a force of 500 grams, or more, was required to remove the button from the surface and a lateral force of over 120 grams was required to make it slide.

6.3.2 The prototype electromagnetic pick-up unit.

The main requirements of the electromagnetic pick-up device are that it always overcomes the attractive force between the button and the steel field plate and that it cleanly releases the button without any lateral drift.

Therefore, a relatively powerful electromagnet had to be designed which could be switched off without leaving any residual magnetic forces. The design also incorporates a conical spigot to centre the button onto the electromagnet and hence ensure that it is precisely released onto the field plate. Other important features of the unit are that it should transport the fibres high enough above the field plate to avoid collisions with other fibres and that a button should be able to rotate with respect to the X-Y carriage as it is moved about.

It was found that the strength of the attractive force between a button and the pick-up head was strongly dependant on the geometry of the pick-up head itself (see figure 6.6 for details of the finally adopted design). Also important were the number of windings of the solenoid and the current through it. However, despite the fact that these were constrained by space limitations and the danger of overheating, the magnetic circuit was saturated at the operating voltage (see section 6.4.5). Many variations of the detailed geometry of the pick-up head were tried and eventually a satisfactory design was determined. Mild steel was used as the magnetic material and brass (or air) as the non-magnetic material because of their ready availability. However, more suitable magnetic alloys are available and would lead to an improved device if incorporated in the design. The prototype electromagnet grips a button with a force of about 600 grams and thus only just overcomes the attraction of the button to the field plate (about 500 grams). Nevertheless, during all the tests carried out, the pick-up head always managed to remove the buttons from



Prototype Electromagnetic Pickup head geometry.

Figure 6.6

the field plate. It is also important that the electromagnet exhibits no field when there is no current flowing in the solenoid or else residual attractive forces would cause the button to cling to the spigot as it went away and some lateral sliding of the button would result. This problem was overcome by applying an exponentially decaying sinusoidal voltage to the electromagnet's solenoid on switch-off. A simple electronic circuit was developed for this purpose by G. C. Cox and it proved to be a very successful technique.

The raising and lowering of the electromagnet was achieved by mounting it on precision linear bearings and driving it via a DC motor and screw thread (see figure 6.7). Limit switches were used to turn off the motor at the top and bottom of its travel. The drive from the thread to the pick-up head is via two compression springs so that the pick-up head can stop moving as soon as it meets an obstacle, i.e. before the DC motor stops turning. This ensures that a button is gently but firmly clamped between the pick-up head and the field plate when the pick-up head descends and prevents damage occuring if a button is accidentally placed on top of another. Rotary bearings were incorporated into the prototype unit to allow the electromagnet to rotate during fibre transport. However, this proved unnecessary in practice as the buttons easily rotated about the pick-up head spigot. All aspects of the unit, and in particular the bearings, had to be completely free from play to prevent there being any play at the spigot itself. In the prototype unit the centre of the spigot was displaced by 50 microns from the centre of rotation of the bearings, although in practice this offset was of no concern since the bearings

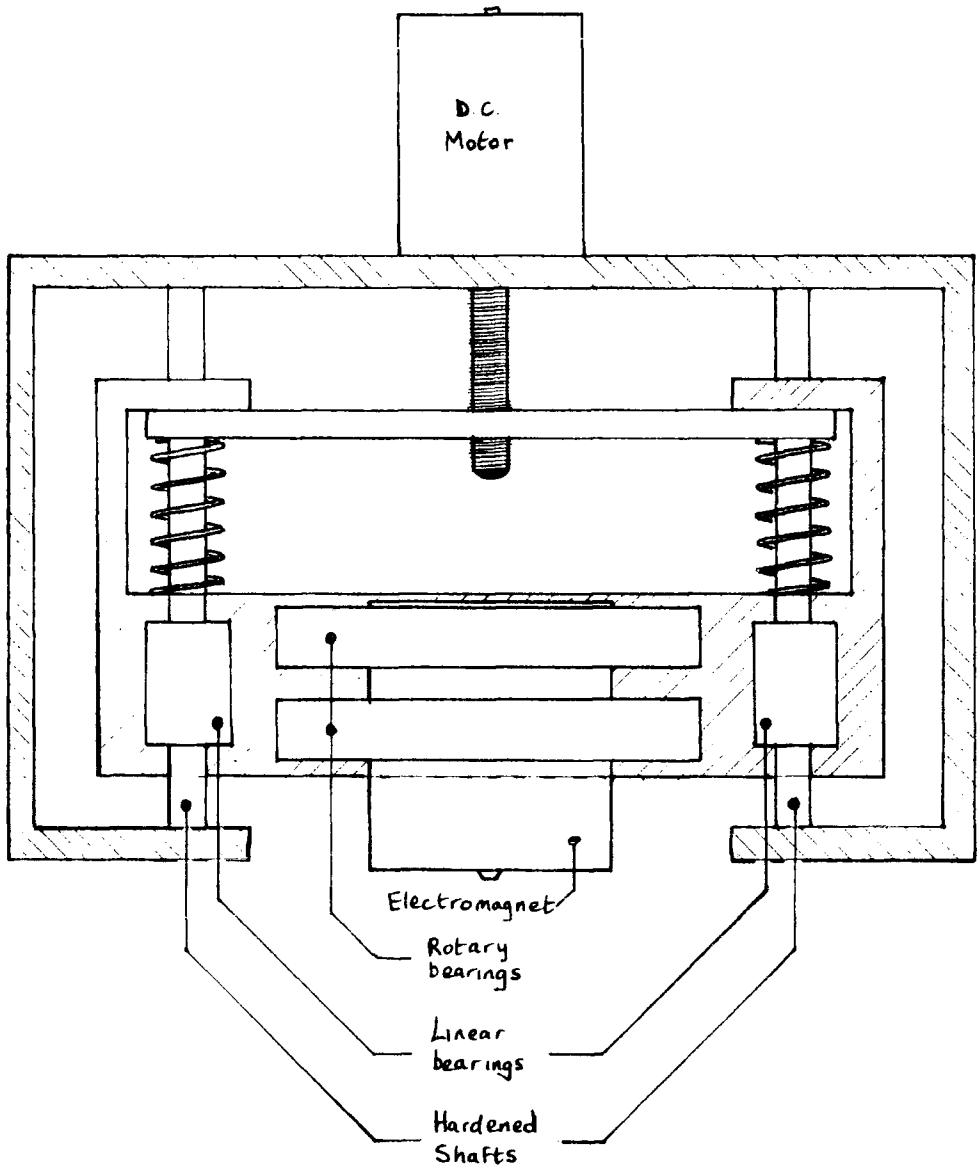


Figure 6.7 Prototype
Electromagnetic Pick-up Head Assembly

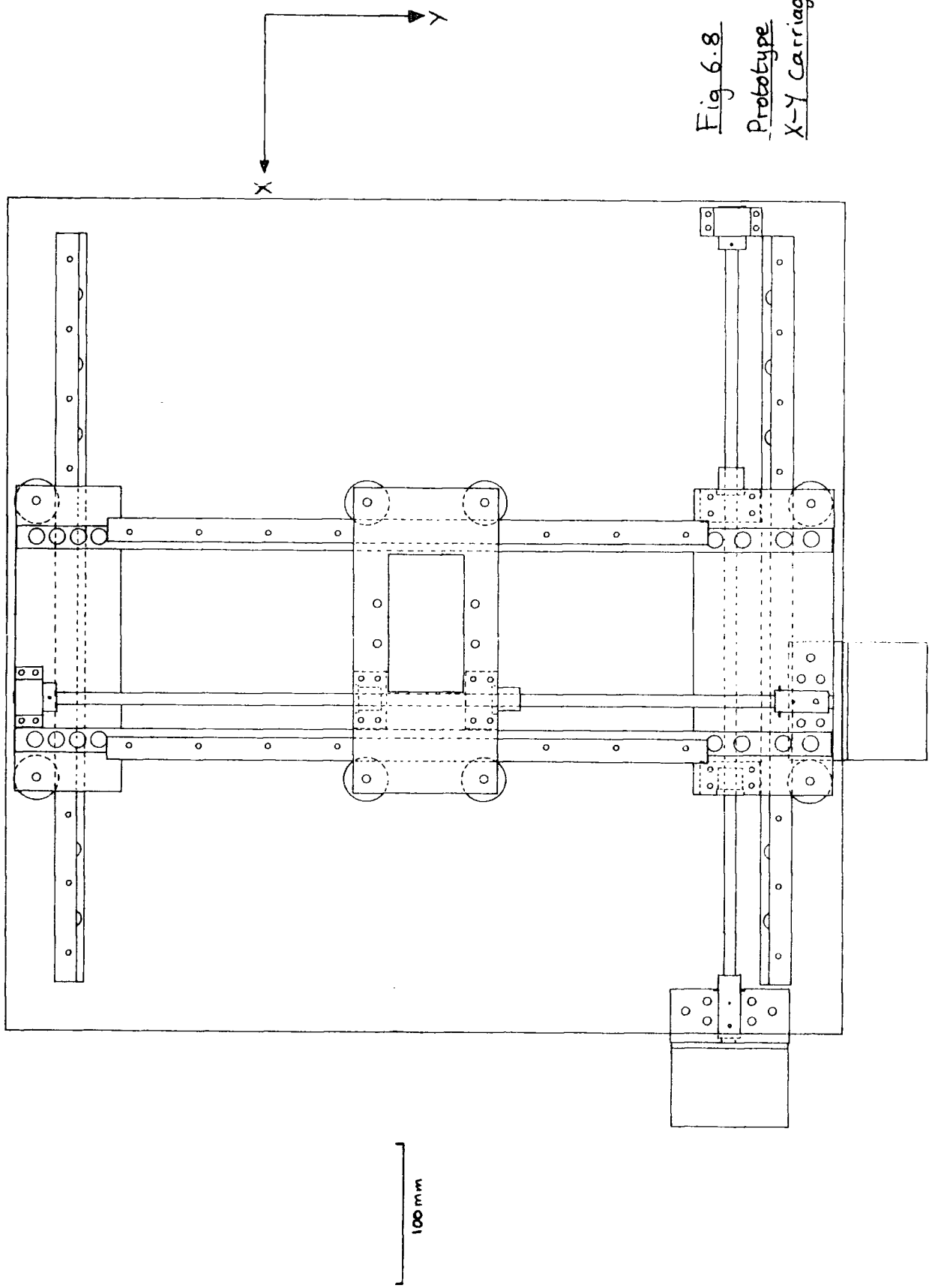
were never allowed to rotate during the tests.

6.3.3 The prototype X-Y carriage.

The X-Y carriage used in the prototype suffered from many deficiencies because of its low cost. The linear motions were facilitated by low cost rails and wheels. The Y-motion slide and the saddle were driven by stepper motors via low cost rolled ballscrews (see figure 6.8). The end of the Y-motion slide not driven by a ballscrew was free to flex and consequently introduced some backlash into the X-motion. The whole unit was fixed to a 1/4 inch thick steel plate. The electromagnetic pick-up unit was mounted through a large hole in the saddle plate. There was no encoding in the system and positioning was achieved by open-loop step counting. No attempt was made to accurately set the X and Y slides at precisely 90 degrees to each other, the slides are unlikely to be precisely linear (with perhaps 100 microns of bow) and the manufacturers quote a targeting accuracy of 100 microns for the ballscrews over the total travel (about 280 mm square). However, there was very little backlash in the ballscrews (less than 5 microns) and it was possible to operate the carriage without miscounting steps. In short, the carriage did not give high absolute positional accuracy but it offered good positional repeatability.

The carriage had to be operated at very slow speeds when good repeatability was required over diagonal paths (i.e. with both steppers running simultaneously). Individually, a stepper could operate at 1000 steps/sec (12.5 mm/sec) but because the

Fig 6.8
Prototype
X-Y Carriage



stepper pulses were generated in software (to save money) a stepper might receive pulses at an irregular rate during a diagonal movement. For example, if the X motor receives 2 pulses for every 3 received by the Y motor it has to perform in a very jerky fashion (on on off on on off ..) and is likely to miss steps. More sophisticated software (difficult) or electronics (expensive and time consuming) could have been used to cure the problem but the easy way out was taken by simply slowing down the stepping rates to a point were irregular pulse trains had no adverse affect on the steppers' performance. The fastest speed that could be tolerated was about 190 steps/sec or 2.4 mm/sec along one of the carriage axes. This was about 83 times slower than AUTOFIB's carriage (see section 6.4.4).

Despite the deficiencies of the carriage it is worth stressing that it was suitable for the main task it was designed for, i.e. to allow the operational properties of the pick-up head and fibres to be studied. Furthermore, experience gained with it proved helpful when it came to designing a carriage suitable for use on the AAT (see section 6.4.4).

6.3.4 The prototype control computer and electronics.

A PDP-11/23 was employed (via CAMAC) to control the functions of the prototype. The software was mainly written in FORTRAN with a few routines written in assembler. The CAMAC interface used was a general purpose module which handles 2, 16 bit TTL words, one each for input and output. The bits used and their functions are listed below.

Bit	Function and meaning.
o1	Step X motor
o2	X motor direction; 0=forwards, 1=backwards
o3	Step Y motor
o4	Y motor direction; 0=forwards, 1=backwards
o5	Move electromagnet; 0=raise, 1=lower
o6	Electromagnet current; 0=on, 1=off
i1	Magnet position sensing; 1=magnet up, 0=magnet not up
i2	Magnet position sensing; 1=magnet down, 0=magnet not down

The system is shown schematically in figure 6.9.

Two similar programs were written, one which accepted commands interactively and the other which obtained its commands from a file. The interactive commands accepted are MOVE(X,Y), PICK-UP, and DEPOSIT, whereas the non-interactive program accepts these plus PAUSE, REPEAT FROM THE BEGINING and STOP.

6.3.5 The operational tests.

The accuracy of the release operation was investigated by repeatedly depositing a fibre at a given place and examining the exact position of the fibre aperture by means of a transfer lens and microscope (see figure 6.10). The transfer lens introduced a magnification of 2.5x to aid measurement and the position of the magnified fibre image was measured using a 50 micron X-Y graticule mounted in a micrometer X-Y displacement unit and a microscope. The micrometers could be read to about 5 microns which corresponded to 2 microns at the fibre itself.

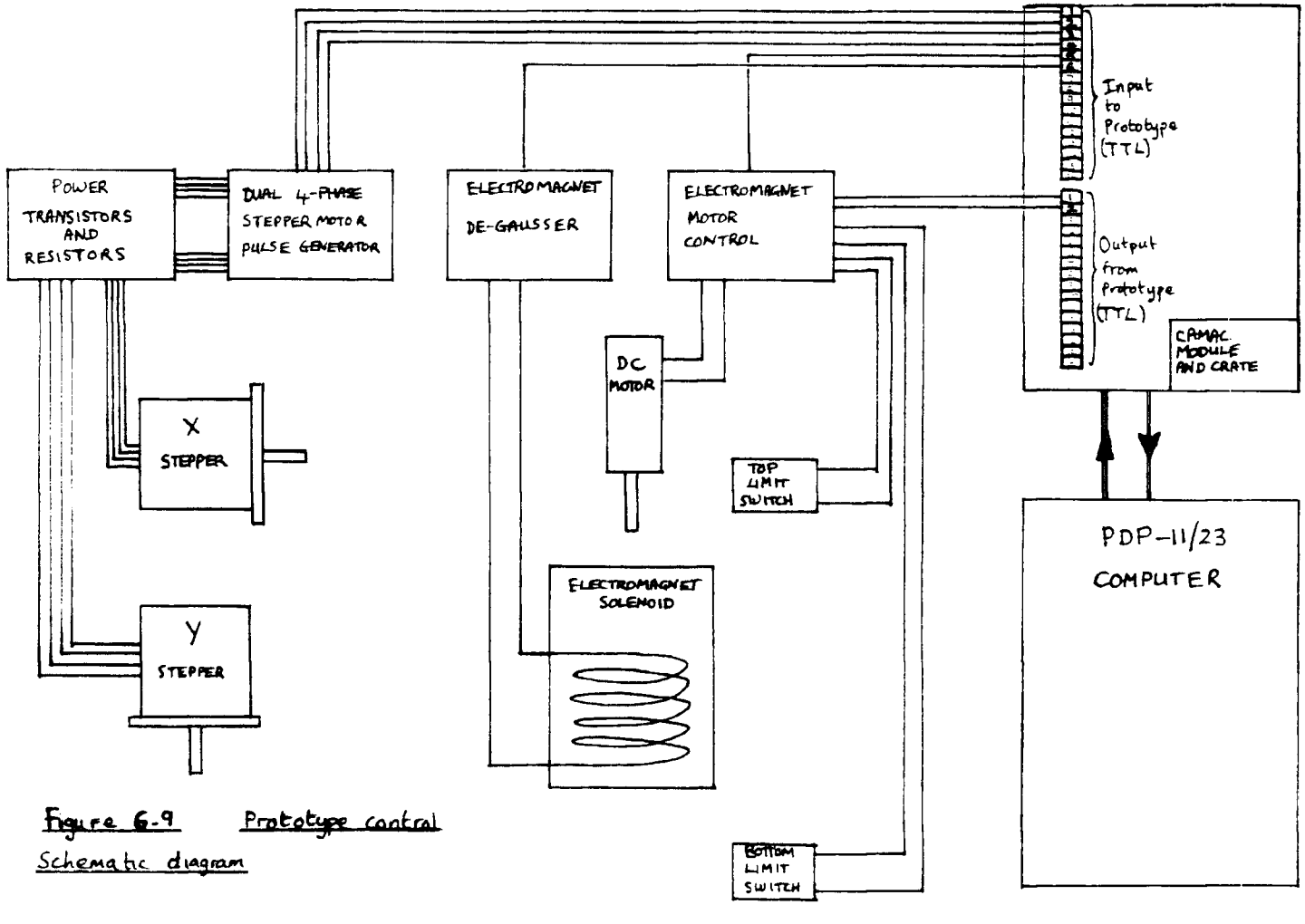


Figure 6-9 Prototype control Schematic diagram

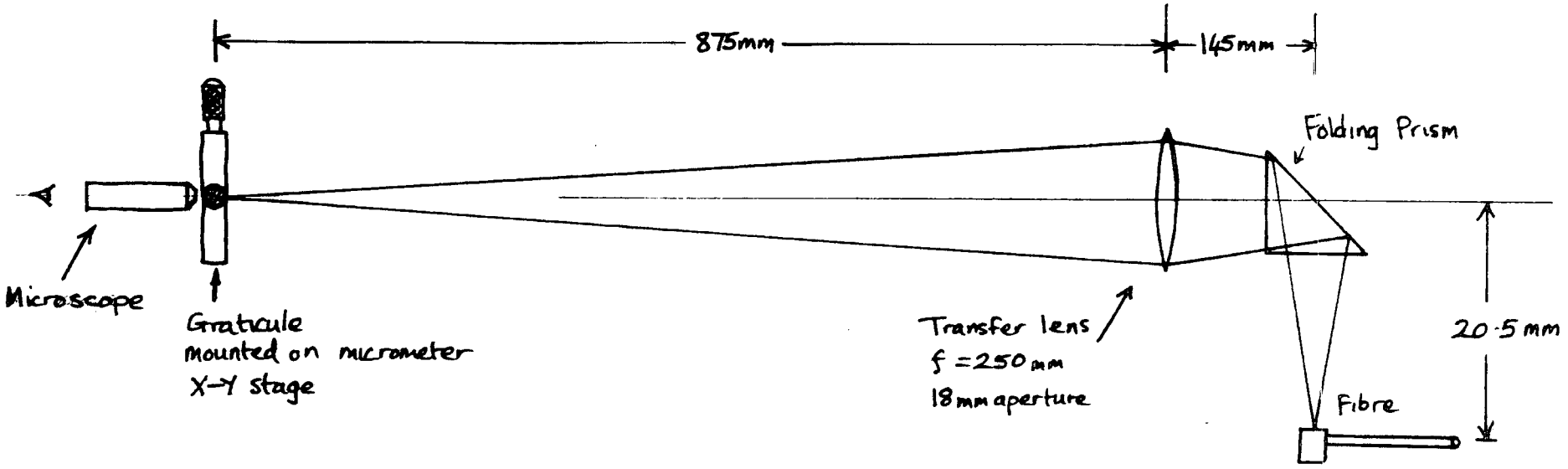


Figure 6.10

Optical equipment used to evaluate positioning accuracy of prototype coupler

Great care was taken to ensure that other sources of positional error, such as backlash of the X-Y carriage or movements of the optical bench, were kept to an absolute minimum. A second graticule with 10 micron divisions was mounted on the X-Y carriage saddle and regularly monitored using another microscope. The control software was instructed to do the following:

Move to fibre at 1000,10000 (coordinates in motor steps)
Pause - check X-Y carriage
Pick up fibre
Move to 1000,11000
Deposit fibre
Pick up fibre
Move to 1000,10000
Deposit fibre
Pause - check X-Y carriage
Move to carriage park position at 1000, 15000
Pause - measure position of fibre
Repeat from the top.

Note that the X-motor was not used for this particular test because of backlash problems. It was found that the position of the fibre was always within a circle of 15 microns radius over 50 trials. The test was repeated with the device on its side (Y-motion vertical) and the positions fell within a circle of 18 microns radius (see figures 6.11 and 6.12 for scatter plots of the results). There was some uncertainty as to where in the mechanism the scatter arose and it seems likely that the X-Y carriage contributed to some extent. There was

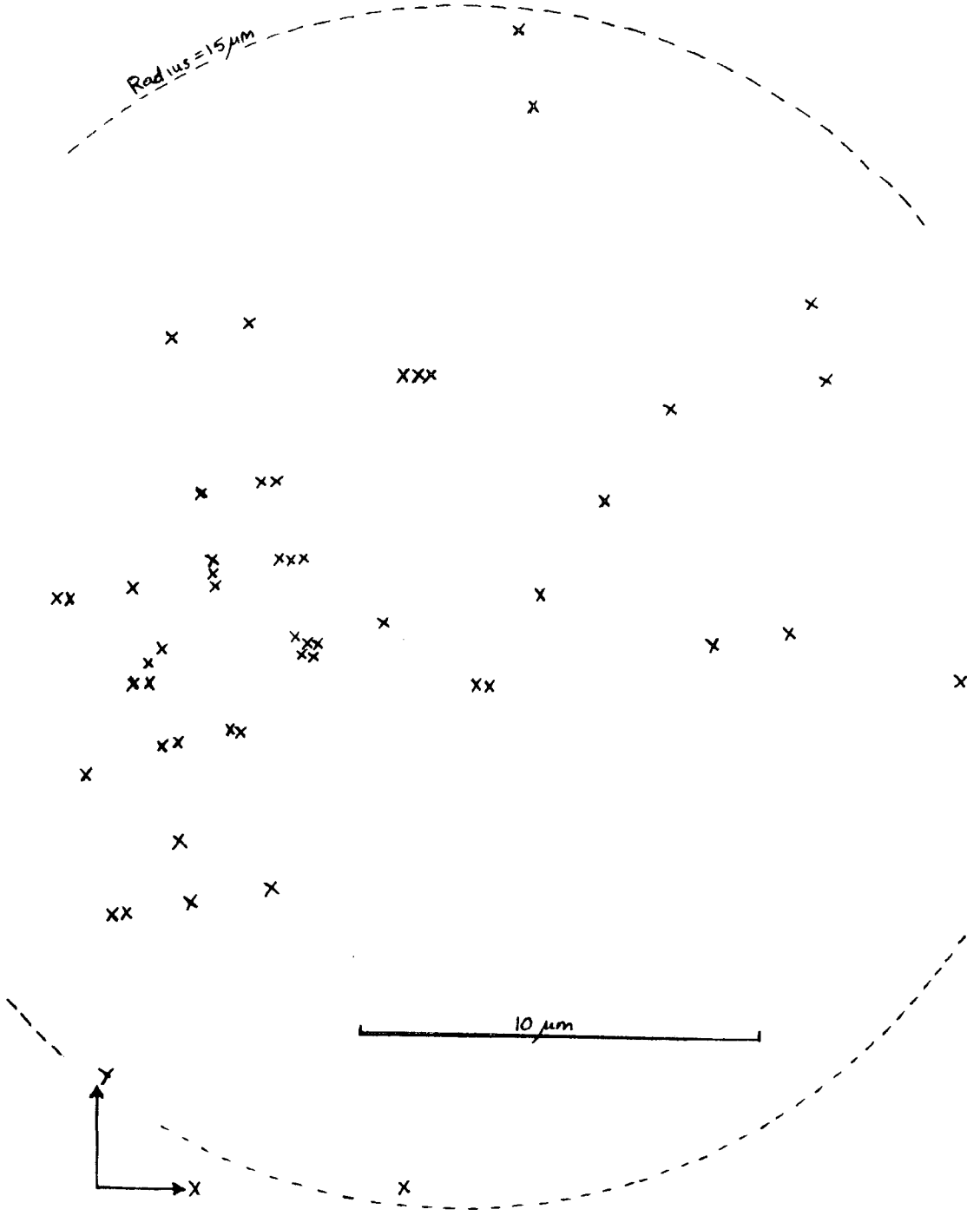


Figure 6.11 Positions of fibre over 50 trials taken with the X-Y carriage in a horizontal position.

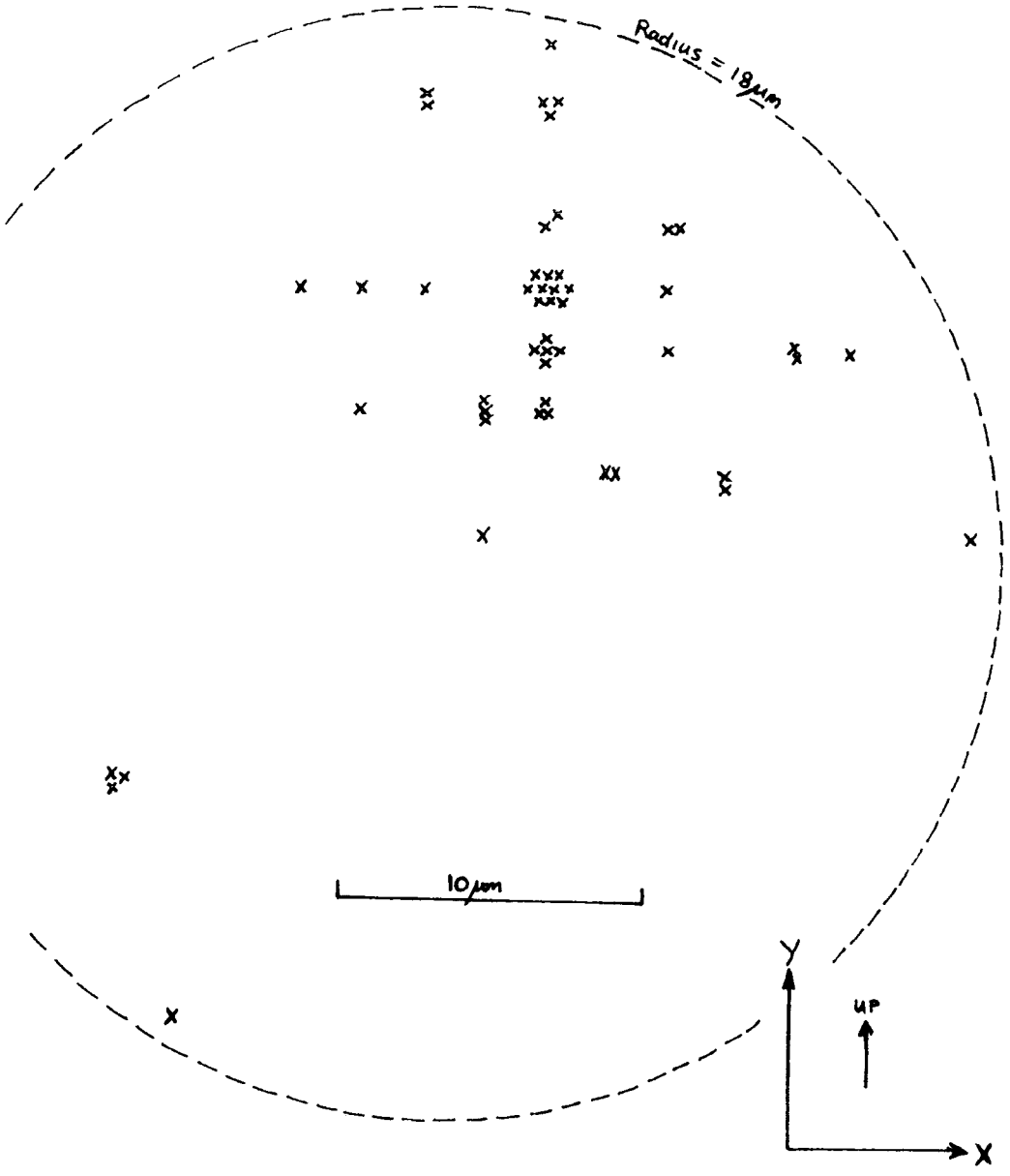


Figure 6·12 Positions of fibre over 50 trials taken with the X–Y carriage in a vertical position

also some evidence of settling - the X value tended to decrease with time in the horizontal test whereas the Y value tended to increase with time for the vertical test. These tendencies both correspond to the fibre image rising with time at the microcope focus (see figure 6.10). This was not too suprising since all the test equipment, including the measuring system, was resting on a wooden bench. However, the results were encouraging especially in view of the fact that the AAT's Cassegrain focus has an image scale of 150 microns/arc-second and that the smallest fibres in use there have 200 micron diameter apertures.

As a test of the entire prototype, 8 fibres were manipulated simultaneously over a long period of time. This demonstrated the device's ability to move the fibres about without tangling or dropping them. The prototype was instructed to pick them up, place them in a plausible field configuration and then return them to their park positions again. This cycle was repeated until the program was interrupted by the user. Figure 6.13 shows the configuration used. Each fibre was threaded through a pivot point on a circle of 240 mm radius. These pivots were 7.2 degrees apart which corresponds to 50 pivots on a full circle. The park positions were on a circle of 180 mm radius which corresponds to the full 40 arcmin Cassegrain field of the AAT. Each fibre was 1.5 metres long and was as described in section 6.3.1 except that no prisms were incorporated. Altogether, 38 successful cycles were performed including one continuous 15 cycle run. Each cycle lasted about 30 minutes, involved 18 pick-ups, 18 put-downs (a ninth button with no fibre attached

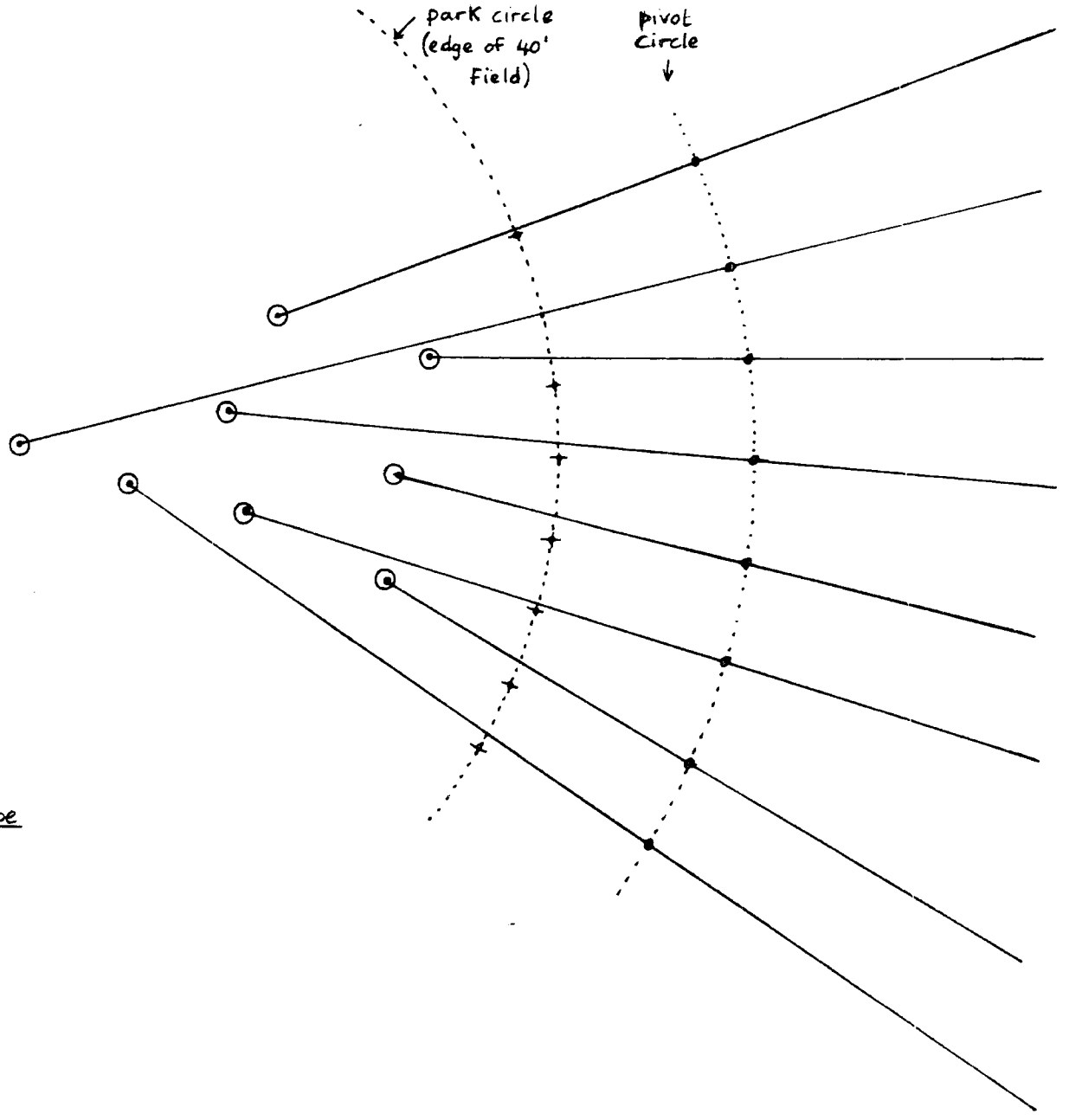


Figure 6.13

Field configuration used to conduct prototype performance tests

was thrown in for good measure) and 40 diagonal carriage motions of various lengths. The long time taken for a cycle is because of the slow carriage speed (about 28 minutes of the cycle was taken up by carriage movements). This test demonstrated that the device was capable of operating many fibres without dropping any fibres in transit, disturbing the positions of any of the other fibres or getting the fibres tangled up.

6.3.6 Conclusions drawn from the prototype.

Despite the use of low quality components in its constructon the prototype was seen to perform very well and all the indications were that the original concepts incorporated in the design could be made to work to high specification on the AAT.

No great difficulty was encountered with the construction of any of the component parts though the construction process itself lead to many improvements of the original design concepts.

The accuracy of the pick-up/put-down operation was well within specification. When repeatedly delivering the fibre to the same location its position always fell within a circle of 18 microns radius even with the device on its side (i.e. with the X-Y carriage in a vertical plane). Furthermore, it is suspected that this scatter will be even less for AUTOFIB.

Good operational reliability was achieved. The electromagnetic pick-up head manipulated the fibres in a very positive manner and never dropped any. This reliability was attributable to the relative magnitudes of the forces that acted during the manipulation of the fibres. The weight of a fibre tip is about 1.5 grams whereas it takes a force, acting normally, of more than 500 grams to pull a fibre away from the steel surface and a lateral force of more than 120 grams to slide it along the surface. A force of over 600 grams is required to remove a fibre from the electromagnetic pick-up head. These forces remained constant over many hundreds of pick-up and put-down operations. Also, the simultaneous handling of 8 close-packed fibres demonstrated that the fibres could be manipulated without becoming tangled up. This particular experiment also tended to suggest that it would be possible to operate more than 50 fibres.

Field set-up times for the prototype were poor on account of the low cost X-Y carriage employed. On the other hand, the operational speed of the pick-up head itself was satisfactory with a pick-up operation taking 2 seconds and a put-down operation taking 3 seconds. From the experience gained with the prototype and a knowledge of commercially available engineering components it was estimated that a coupler could be built that can reset 50 fibres from one field configuration to the next in under 12 minutes via a park position. Even faster set-up times are possible if the need to use the park position as an intermediate step is dispensed with by means of an optimising computer program. (see section 6.4.7)

The practical work carried out made no attempt to address the problems of flexure or field curvature. These were considered to be extra engineering complications to be dealt with when the fully operational instrument was designed (see sections 6.4.2 and 6.4.6).

6.4 AUTOFIB

Although AUTOFIB does not exist as a working instrument at the time of writing it exists as a design and therefore I shall describe it in the present tense. The design was finalised by Peter Gray and myself during Peter's visit to Durham in August 1985. The major difference between AUTOFIB and the prototype is the X-Y carriage which is much faster and more accurate for AUTOFIB. This is mainly due to the use of high quality engineering components and sophisticated electronics and microprocessors. Peter Gray's excellent idea that the fibre probe bundles and the field plate should comprise an interchangable unit has also been incorporated into AUTOFIB's design (see figure 6.14). This allows many different fibre options to be readily available to the observer. Many aspects of the FOCAP fibre bundles are the same for AUTOFIB (e.g. the slit units and the guide fibres). It is even possible to mount FOCAP aperture plates into AUTOFIB. The spectrograph support structure has also been copied from the FOCAP system. The fibre probes are the same as for the prototype except that 300 micron core-diameter fibre is used. The electromagnetic pickup (or manipulator) unit is very similar to the prototype one in principle. Finally, it was decided to include some extras to make the operation of AUTOFIB as simple and as pleasant as

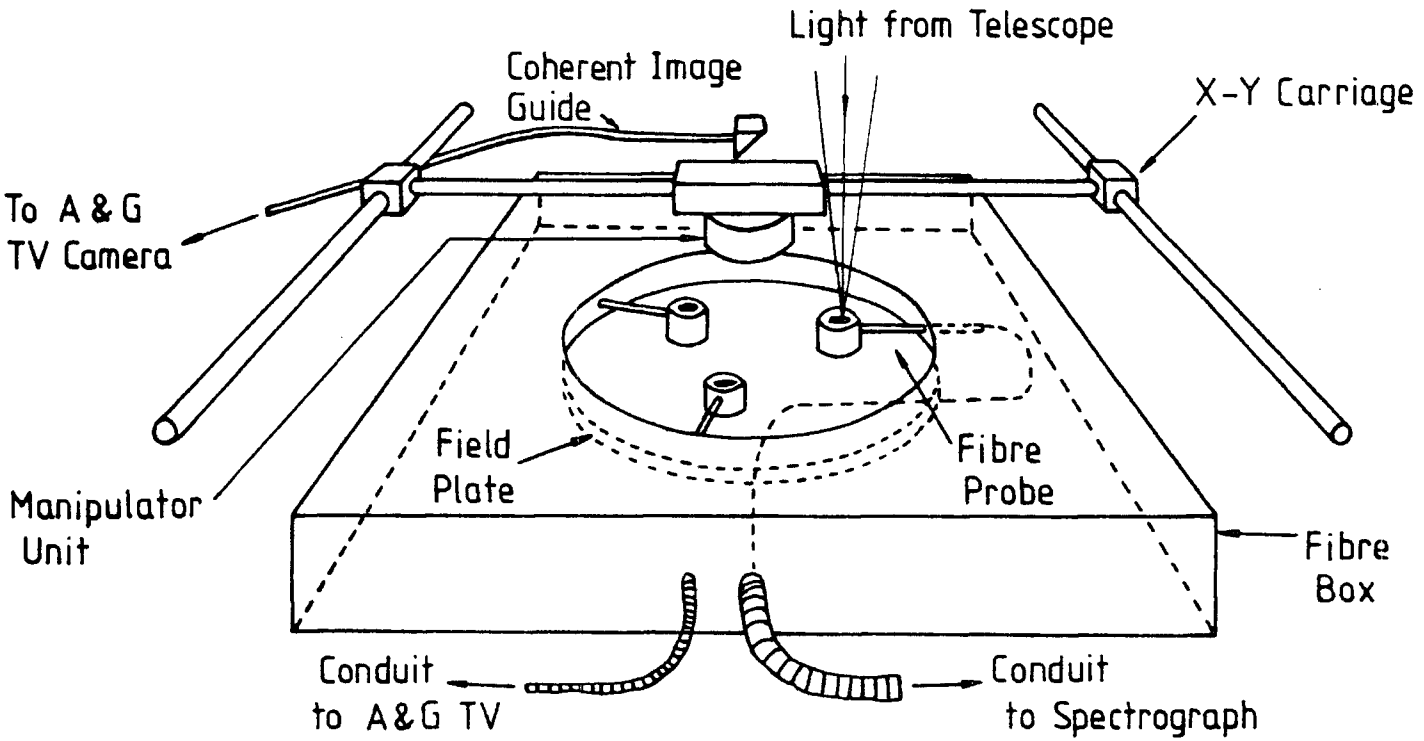


Figure 6.14 AUTOFIB concept incorporating removable fibre module.

possible. These extras include a coherent fibre-optic image guide mounted on the pickup unit to provide direct astrometric field viewing and a mimic display in the control room to let the astronomer see exactly what is going on.

6.4.1 The fibre probes.

The spectrograph probe bundle consists of 64 fibre probes which feed a spectrograph slit unit. These probes are the same as the prototype ones described above (see section 6.3.1) except that silvered prisms are used. This makes the task of epoxying the fibre into the button both easier and more reliable*. Each complete probe has the 300 mm length of fibre adjacent to the button encased by a thin hypodermic tube to prevent it bending** in the field. The fibres are fed to the slit unit via a protective conduit.

The guide bundle consists of probes very similar to those used to feed the spectrograph but employing seven fibres per probe. These probes are positioned at the locations of bright field stars and used for acquisition as well as guiding. The entrance faces of the seven fibres are arranged with one central fibre surrounded by the other six. Fibres with 200 micron cores are used and their centres are 300 microns apart due to the 300 micron cladding diameter. A similar, but wider spaced, arrangement feeds the acquisition TV camera. Eight guide probes are available with each fibre module but only two or three of these are required at any one time.

* because epoxy on the reflecting surface makes the finished probe stronger but would prevent total internal reflection for an unsilvered prism.
** The hypodermic tubes are straight only when unstressed.

6.4.2 The field plate.

Each fibre probe tip rests on the circular field plate and the hypodermic tube portion slides through a pivot at the plate's edge. These pivots are equispaced about the field plate's circumference and suspend it from the top of a thin aluminium sheet box which has a large hole in it to allow light from the telescope to illuminate the fibre probes. There are also three kinematic seats on the field plate's circumference. When the fibre module is mounted on the telescope the field plate has to be manually screwed up against a kinematic seating within the positioning unit to ensure that it is held rigidly with respect to the telescope focus. The ball ends of the kinematic seats are part of the fibre module as these are adjustable and can be preset once and for all for each module. The field plate is flat except that the central area is lower by 2 mm. This compensates adequately for the AAT's curved focal plane but makes the small anular area around the step inaccessible. If the surface of best focus was orthogonal to the incoming light at all points in the AAT's focal plane a curved field plate (and a manipulator with a delicate wrist action) could be used. However, this is not generally the case for Ritchey-Chretien telescopes. For the AAT the surface orthogonal to the incoming light (surface of best pointing) has a radius of curvature of 30.7 metres whereas the focal surface has a radius of curvature of only 4.37 metres. Both surfaces are concave like the primary mirror. Thus there is no smooth surface upon which a fibre probe can be placed and be in focus and point accurately at the same time. Therefore a flat, stepped field plate has been adopted which ensures that the

fibre apertures are well focused and point to within 20 arc-minutes of the axis of the incoming light cone. The use of a flat plate also means that the fibre probes can be manipulated by a robot with only three degrees of freedom.

6.4.3 The fibre module box.

A sheet metal box completely encloses the flexible parts of the fibres and prevents any snagging from occuring during fibre set-up. The box is empty except for the field plate and the fibres and its interior is completely smooth. There are two output conduits, one each for the spectrograph and guide bundles. When several modules are available it will be useful to have a computer-readable label attached to the exterior of the box so that the AUTOFIB microprocessor can recognise it.

6.4.4 The X-Y carriage.

The X-Y carriage is both fast and precise reaching speeds of 200 mm/sec, or more, and yet delivering the manipulator unit to within 10 microns of an absolute target position anywhere in its 380mm square patrol area. The carriage design is shown in figure 6.15. The accuracy is achieved by using recirculating linear bearings of the highest accuracy for both axes. The X-axis bearing is longer than the Y-axis one and the two axes are coupled together at precisely 90 degrees using a stiff triangular plate. Also, both axes are driven using zero-backlash ballscrews and encoded using extremely accurate linear transducers ($\pm$ 2 microns absolute accuracy) which give pulses every 1 micron. Note that unlike the prototype carriage

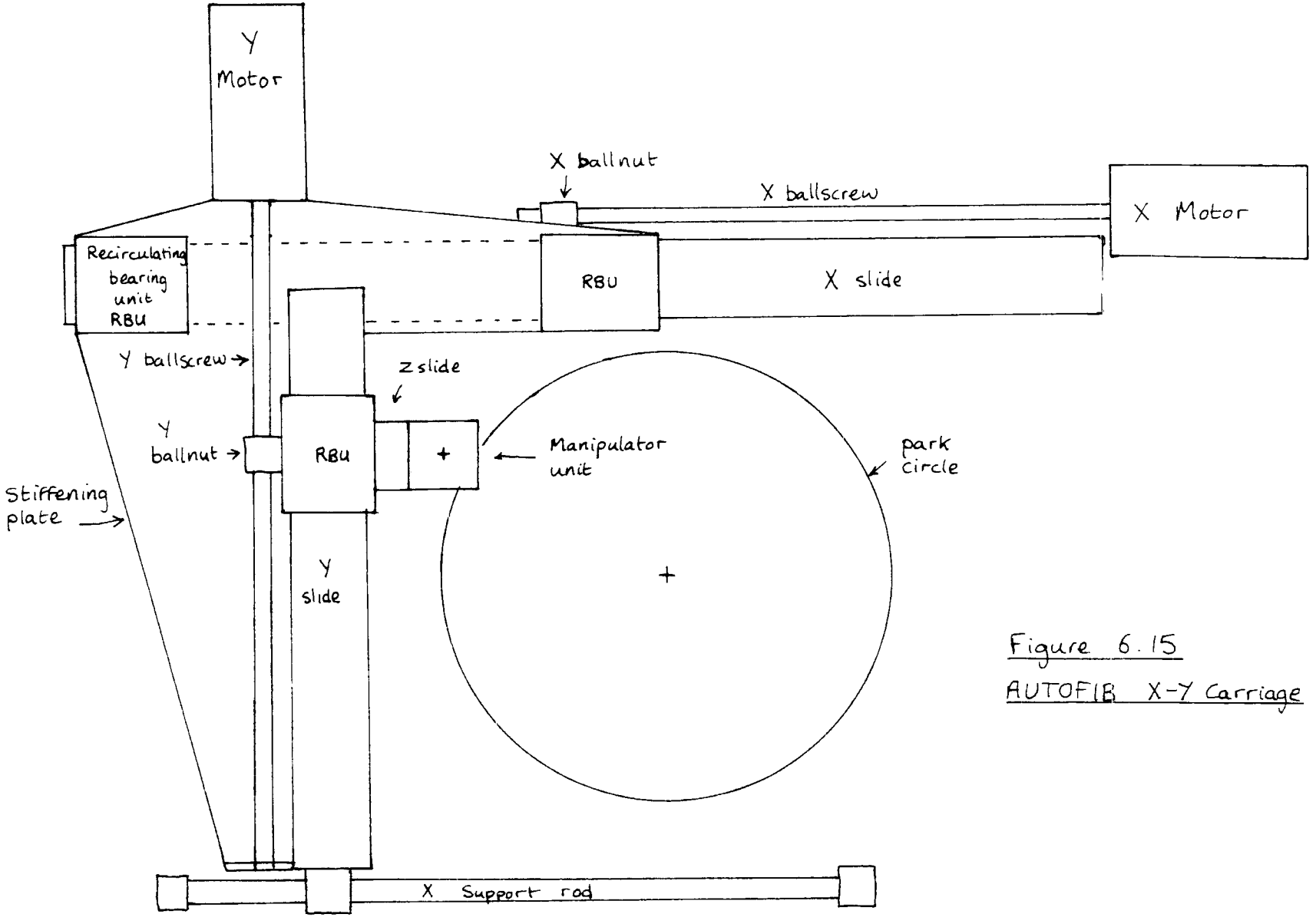


Figure 6.15
AUTOFIB X-Y Carriage

the ballscrews are never used to determine the carriage position but only to move it. It therefore does not matter if the screws have substantial pitch errors as long as they do not exhibit any backlash at the ballnut or the end bearings. Any such backlash will be detected by the linear encoders but is undesirable as it will make the positioning of the carriage difficult to control.

The X-Y carriage has been designed so that it can jump 100 mm in 1 second, starting and stopping at rest. If it does this by accelerating constantly for the first 0.5 seconds and then decelerating constantly for the final 0.5 seconds then it will reach a maximum speed of 200 mm/sec and accelerate at 0.4 metres/sec . Using a ballscrew pitch of 5 mm per turn the maximum rotational speed is 2400 RPM. The torque required to produce an acceleration, a, is given by

$$T = \frac{Map}{2\pi}$$

where M is the mass to be moved and p is the screw pitch. The X-axis of the carriage has to move the entire Y-slide and the pickup unit. If a mass of 20kg is assumed for these components then a torque of 0.006 Nm is required when $a=0.4$ m/sec^2 (horizontal motion) and a torque of 0.16 Nm is needed when $a=0.4 + 9.8 = 10.2$ m/sec^2 (vertical motion). These estimates do not allow for friction or the moment of inertia of the ballscrew itself but these are fairly low for the engineering components chosen. The standard engineering practice of multilplying the calculated torque by an "idiot factor" of about 3 or so leads to a torque requirement of 0.5 Nm available upto speeds of 2400 RPM. The Y-axis requirement is less but it

was decided that a motor identical to the X-axis one should be used so that only one spare motor needs to be kept in case of breakdown. Very large stepper motors would be required to meet these torque requirements at high speeds so it was necessary to use DC servo motors. To control these motors commercially available servo amplifiers and intelligent microprocessor controlers are used which together with the linear encoders and the integral motor tachometers close the two servo loops.

6.4.5 The manipulator unit.

The electromagnetic manipulator unit for AUTOFIB is very similar to the prototype in its operating principles. The main differences are that a standard short linear bearing is used instead of ball bushings, the geometry of the electromagnetic spigot allows close packing of the fibre probes, sliprings have been incorporated for the electromagnet's power supply and the rotary bearings are smaller and situated above the electromagnet. Figure 6.16 shows the unit schematically.

The Z-motion is obtained by using a small DC motor and a leadscrew with a compression spring coupling the leadnut to the electromagnet in the downward direction. Thus any resistance met in this direction is taken up by the spring to protect the mechanism. Limit switches are used to stop the motion in the up and down directions. The down limit switch is triggered by a preset compression of the spring so that the motion in the downward direction is load-limited, i.e. it stops when a preset resistance is met rather than at a preset height. This ensures that a probe is held onto the field plate with a known

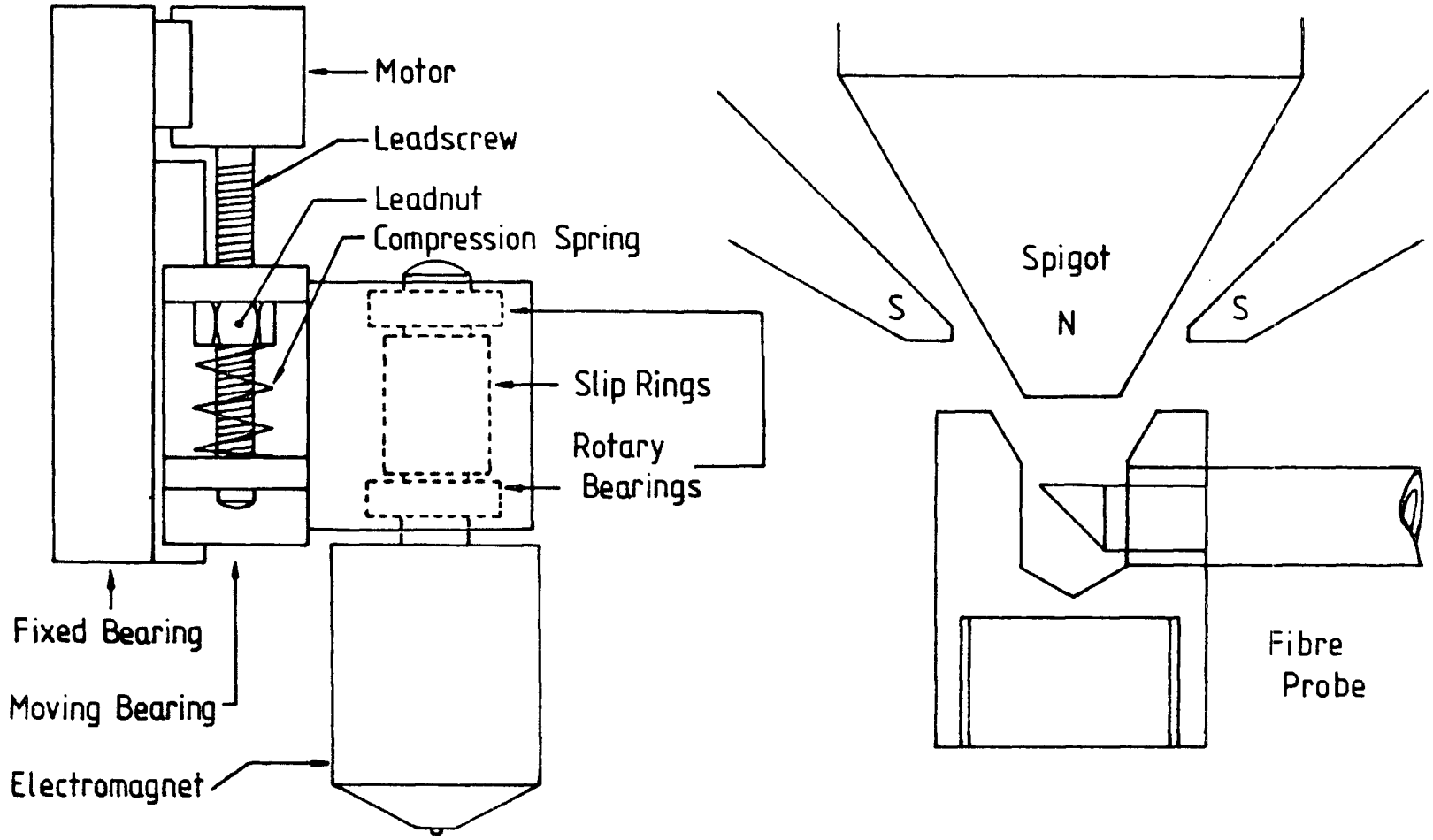


Figure 6.16 Schematic drawings of AUTOFIB manipulator mechanism and electromagnet geometry.

force when it is engaged by the manipulator even if the field plate has steps in it. It also prevents damage should one probe be placed on top of another. A rotary potentiometer encoder and a rack and pinion are used to measure the Z position so that the controling microprocessor (or the astronomer via the mimic display) can check that the height on put-down is sensible. Smaller rotary bearings are easier to rotate and the part of the mechanism that protudes into the space immediately above the field plate should be as small as possible. Therefore, the mechanism that allows the electromagnet to rotate while a probe is being transported across the field is placed above the electromagnet. This part of the mechanism also includes sliprings to feed power to the electromagnet. The conical spigot and the bearing shaft therefore have to be precisely concentric and this is achieved by machining both together while the work remains in the lathe chuck. A demagnetising circuit identical to that used with the prototype is employed to ensure that the probes are cleanly released. A pick-up operation (lower - switch on electromagnet - raise) takes 2 seconds whereas a release operation (lower - switch off and de-gauss - raise) takes 3 seconds as for the prototype.

The strength (holding force) of the electromagnet can be estimated using the formula (derived in most general electrical engineering texts)

$$F = \frac{B^2 A}{2\mu_0}$$

where F is the force in Newtons, B is the flux density through the object being held in Tesla, A is the cross section in

square metres through which the flux passes and $\mu_o = 4\pi \times 10^{-7}$ weber/amp-m is the permeability of free space. Clearly we wish to maximise B and A. However, we wish to keep the probes, and therfore A, as small as possible so that objects close together on the sky can be studied. For mild steel B saturates at 2.2 Tesla and the best material possible, an alloy of cobalt and iron in equal proportions (permendur), is not much better saturating at 2.4 Tesla. Using B=2.2 and A=6 mm for a button we get F=11.6 N. This is slightly higher than found in practice and maybe due to flux leakage or a lower saturation value for the silver steel buttons. It is likely that the core of the electromagnet is saturated as H=37 amps/cm for the electromagnet and this indicates saturation on the B-H curve for mild steel; H=NI/L where N is the number of coil turns (7000), I is the current (0.064 amps at 30V) and L is mean magnetic path length (12 cms). Thus it seems that significant improvement on the measured holding force of 0.9 kgF is difficult to achieve without compromising other aspects of the design. However, a 900 gram holding force is adequate.

6.4.6 The mechanical structure.

The mechanical structure not only supports the positioner and the fibre module but also carries the RGO spectrograph. The structure was therefore carefully designed to ensure that the large weight of the spectrograph is carried without putting large amounts of stress on the X-Y positioner and fibre module. Also the main plate which carries the X-Y carriage has stiffening members incorporated to keep the flexure of the positioning system to a minimum. The spectrograph is supported

by four shear walls which transfer the spectrograph load to the AG unit. This arrangement is very similar to the spectrograph suport structure employed by FOCAP (see figure 6.1). The positioner is mounted onto these walls at the points where they are mounted onto the AG unit and the positioner therefore benefits from its stiffness. Also, this ensures that a minimal amount of the spectrograph load is transfered to the positioner. One of the shear walls can be safely hinged to one side while the telescope is parked at the zenith, to interchange fibre modules.

6.4.7 Computer control

A dedicated microprocessor is used to control the positioning unit. The full control system is shown schematically in figure 6.17. The software is ROM based and resides in the 6809 processor card which acts as the master processor. The two servo-motor controllers are standard industrial units (Galil DCM100) which come complete with software and are used as slave processors. These controllers allow complete control of the velocity profile during motion and receive their high level instructions from the master processor which ensures that each movement is in a straight line. A purpose built interface card, which includes the de-magnetising circuit, is required to control the manipulator unit. A battery back-up memory card is encorporated to enable the system to recover from power failures and a general purpose i/o interface is used to monitor various telemetry signals. Finally, a graphics card is used to output information on the mimic display monitor in the telescope control room so that

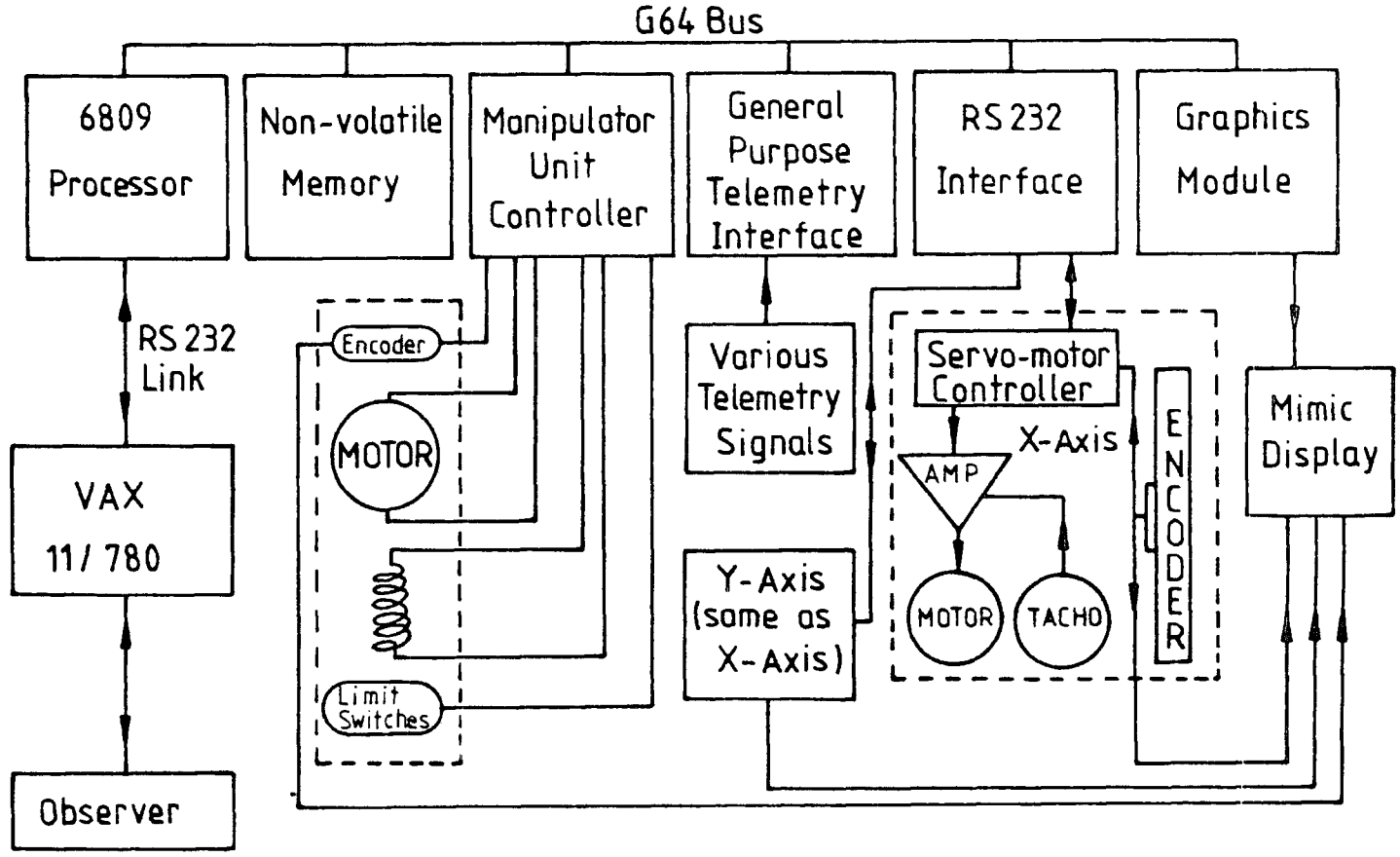


Figure 6.17 AUTOFIB control.

observers can see exactly what is going on. This display cabinet also receives information directly from the encoders to give a real time display without large processor overheads. Observers communicate with the microprocessor via the AAT's instrumentation computer (a VAX 11/780) and an RS232 link.

The microprocessor can receive and execute sequences of commands from the VAX. The current system configuration is stored and updated in the microprocessor's non-volatile memory and each new command is tested to ensure that no mechanical conflict (e.g. probe collision) will occur when it is executed. The mimic display, which shows the system configuration, is also updated whenever anything is changed. The VAX software allows the user to prepare an acceptable configuration for his list of field targets and generate and communicate the appropriate microprocessor commands. There are therefore four distinct software items; the microprocessor software, the configuration software (VAX), the command generation program (VAX) and the user interface (VAX).

The microprocessor program executes each command in turn (although it can receive and store blocks of commands). On power-up the micro "knows" where the fibres are from its battery back-up memory and it generates a graphical representation of the current system status on the mimic display. It can also compare its representation with that of the VAX and update itself if necessary. It then has to establish the zero point of the coordinate system by moving both axes towards the centre of the field to locate fiducial marks on the linear encoders. It can then execute commands.

Each command is tested to ensure that it can be performed without causing mechanical conflict. Commands are also tested to make sure they are consistent with the current configuration (e.g. if commanded to pick up a probe at a specific location, there has to be a probe at that position). Only three high-level commands are required; pick-up, put-down and move to position (X,Y). Lower level commands will also be available. If a command fails the tests an error message is sent to the VAX, otherwise it is executed. If, during the execution of a command, the microprocessor detects an inconsistency (e.g. when putting down a fibre the electromagnet gets closer to the field plate than would be expected) a warning message is sent and the microprocessor pauses. When a command is successfully executed the microprocessor's representation of the current configuration and the mimic display are updated. The microprocessor also informs the VAX so that it too has an up-to-date representation of the fibre configuration.

The configuration program (available throughout STARLINK as well as at the AAO) produces an allowable fibre configuration from a list of target objects for a given fibre module (see figure 6.18). This includes assignment of sky fibres and guide probes. The input target list provides the celestial coordinates, magnitude and priority of the target objects and is, of course, supplied by the astronomer. The outputs are a list of fibre positions, given as X-Y coordinates in microns from the centre of the field and the required instrument rotation angle. The configuration program therefore has to deal with all the required astrometric coordinate transformations and corrections. This program can (and should)

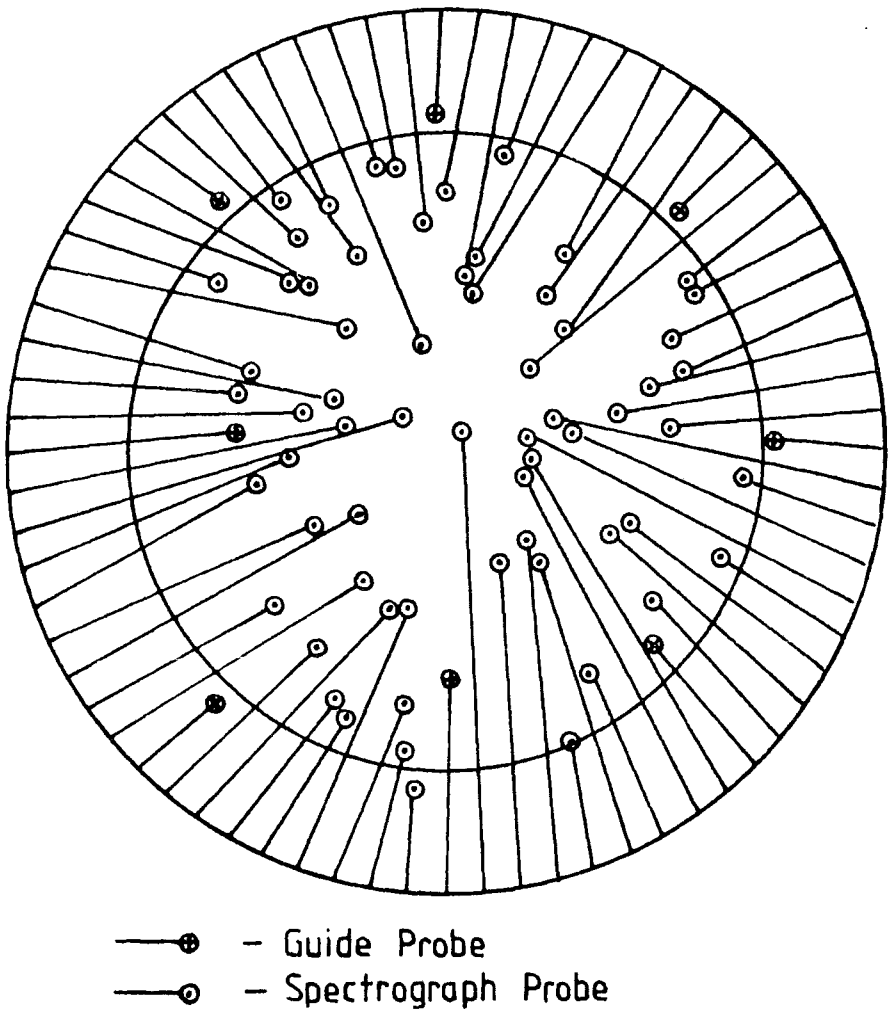


Figure 6.18 Example AUTOFIB configuration.

be run by the astronomer well in advance of visiting the AAT. Basically the configuration program does the following. Firstly it assigns a fibre to an object near the field centre. Then for the next fibre (moving clockwise) it finds the object f round from the pivot in the anti-clockwise direction that satisfies the mechanical constraints (conflict with fibres that have already been placed, angle at pivot with respect to the field centre and distance of object away from pivot). If no such object exists the fibre is parked or assigned to a patch of sky. It continues until it has tried to assign every fibre. In practice, this scheme is complicated because object priorities have to be taken into account, a specified number of sky fibres are required and guide probes have to be assigned.

The command generation program generates a sequence of commands to set up one configuration starting from another. This is optimised so that the total set up time is kept to a minimum. Sequences generated are such that illegal intermediate configurations are avoided. This program is run at the telescope when it is known what the next configuration is.

The user interface allows the astronomer to send commands (directly or from files) to the microprocessor, monitor its progress and run the configuration and command generation programs.

6.5 Future developments

As more and more experience is gained with AUTOFIB useful modifications and improvements will no doubt become obvious. Aspects where possible improvements are already apparent are those of the number of fibres used, the set-up time and the closest approach of two fibres. However, the basic AUTOFIB concept is very versatile in terms of making improvements and some ideas already exist. For example, to increase the number of fibres that can be simultaneously used for observations (which at present is limited by the spectrograph detectors as well as the fibre coupler) it may be possible to dispense with the hypodermic tubes and the constraint that fibres are not allowed to cross.

The concept also lends itself well to applications on other telescopes. An attractive variation is to keep the positioner off the telescope entirely and use it to set up fibre modules before they are mounted on the telescope. With two or more modules telescope dead-times can be kept very short even when using hundreds of fibres. Also the engineering of an off-the-telescope positioner would be easier and its operation could be much more closely monitored by the observer. For some telescopes, where space is at a premium, (such as on a Schmidt telescope) this is in fact the only option. Other possibilities include remote observing which includes being used with satellite telescopes.

Finally, I must point out the enormous scientific potential that would be offered by a future generation large aperture telescope equipped with an automated multifibre spectrographic survey facility.

Acknowledgements

My thanks are due to a great number of people who I have worked with in connection with the Durham/RGO Faint Object Spectrograph and AUTOFIB. In particular I would like to thank my supervisor, Mike Breare, for encouraging me to complete this work. Many thanks are also due to Richard Ellis, Peter Gray, Graham Cox, Don Morton, Graham Martin, Nick Waltham, Alan Purvis and John Webster for their encouragement and assistance.

Finally, I acknowledge the financial support of SERC throughout the period of this work.

References

Blouke, M. M., Heidtmann, D. L., Corrie, B., Lust, M. L., & Janesick, J. R., Proc SPIE no. 570, 1985.

Breare, J. M., Ellis, R. S., Parry, I. R., Purvis, A., Waltham, N., "The FOS user Manual", 1985.

Breare, J. M., Ellis, R. S., Purvis, A., Webb, D. A., Miller, W., Proc SPIE no. 627, in press, 1986a.

Breare, J. M., et al, MNRAS, in preparation, 1986b.

Broadfoot, A. L., Kendall, K. R., Journ Geophys Res, 73, 426, 1968.

Butcher, H. R. & Oemler, A. jnr., Nature, 310, 31, 1984

Cohen, J. G., Goss, W. C., Tubbs, E. F., Caltech application for funding, 1981.

Cullum, M., Deiries, S., D'Odorico, S., Reiss, R., Ast. & Astrophys, [illegible], 198[illegible].

Djorgovski, S. & Spinrad, H., Ap. J., 1985.

Dressler, A. & Gunn, J. E., Ap. J., 270, 7, 1983.

Dressler, A. & Gunn, J. E., Ap. J., 1985.

Ellis, R. S., in "Spectral Evolution of Galaxies", ed. P. M. Gondhalekar, RAL Publications RAL-84-008, pp 122-144, 1984.

Fiocco, G., Visconti, G., Congeduti, F.,
Nature, 228, 1079, 1970.

Gray, P. M., Proc SPIE no. 445, 57, 1983.

Gray P. M. & Sharples, R.,
AAO fibre system user guide, AAOUM-18, 1985.

Hill, J. M., Ph.D. dissertation, Univ. of Arizona, 1984.

Hill, J. M., Angel, J. R. P., Scott, J. S., Lindley, D.
& Hintzen, P., Ap. J. 242, L69, 1980.

Hill, J. M., Angel, J. R. P., Scott, J. S., Lindley, D.
& Hintzen, P., Proc SPIE no. 331, 279, 1983.

Longair, M. S., Nature, 310, 13, 1984.

Martin, G. P., Ph.D. thesis, Univ. of Durham, in prep., 1986.

Meinel, A. B., Ap. J., 111, 555, 1950.

Oke, J. B., Ap. J. Sup. Ser., 27, 21, 1974.

Peacock, J. A. & Gull, S. F., MNRAS, 196, 611, 1981.

Peebles, P. J. E., "The large scale structure of the Universe",
Princeton University press, 1980.

Powell J. R., Proc SPIE no. 445, 77, 1983.

Rees, M., in "Observational Cosmology", Saas Fee 8th advanced
course of the Swiss Soc. of Ast. & Astrophys., 1978.

References

Robertson, J. G., AAOUM-11, 1983.

Robertson, J. G., AAO preprint, 1986.

Savage, A. & Peterson, B. A., IAU symp no. 104, "The early evolution of the Universe and its present structure", eds Abell, G. O., & Chincarini, G., 1983.

Shanks, T., Bean, A. J., Efstathiou, G., Ellis, R. S., Fong, R. & Peterson, B. A., Ap. J., 274, 529, 1983.

Shanks, T., Stevenson, P. R. F., Fong, R. & MacGillivray, H. T., MNRAS, 206, 767, 1984.

Silverman, S. M., "Night air-glow phenomenology", Space Science Rev, 11, 1970.

Tyson, J. A., & Jarvis, J. F., Ap. J., 230, 1979.

Wall, J. V., Pearson, T. J. & Longair, M. S., MNRAS, 193, 683, 1980.

Wall, J. V., Pearson, T. J. & Longair, M. S., MNRAS, 196, 597, 1981.

Waltham, N., Ph.D. thesis, Univ. of Durham, in prep., 1986.

Wynne, C. G., Optica Acta, 29, 137, 1982.