ANALYSIS OF X-RAY IMAGES AND SPECTRA

A thesis submitted to the University of Leicester by R. Willingale for the degree of Doctor of Philosophy.

1979

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Four research projects are reported. They are all concerned with X-ray data analysis techniques and although each project is different there is a common theme which links all four. Part I considers the use of grazing incidence optics to produce two dimensional X-ray pictures with the emphasis on the data analysis and presentation. The mathematical development within this project forms a foundation for the subsequent reports. The research concentrates on viable methods of cleaning up blurred and noisy X-ray images using Fourier filtering and the Maximum Entropy Method including a practical implementation of the theory using digital computers. The major product of this project was a software package for the processing of large matrices and this is documented in Appendix I. This package was used to process astronomical X-ray data from a sounding rocket flight to yield a soft X-ray image of the Cygnus Loop supernova remnant and these results are presented to allow comparison of the analysis techniques developed.

The second project presented in Part II applies the deconvolution theory developed in Part I to the problem of decoding data from coded mask telescopes. The design of such devices is described and computer simulations of X-ray burst monitors are reported with analysis and comment to give a realistic estimate of the expected performance of proposed instruments and to compare the different methods of analysis available.

Part III reports a small project in which the possibility of analysing anode pulse height data from

proportional counters using the Maximum Entropy Method was investigated. A computer program was written to both simulate and analyse real data. The algorithm was used to analyse pulse height spectra from the Cygnus Loop observations.

The final project, concerned with the calibration of crystal spectrometers, was somewhat different from the other three and is presented in Part IV. The mathematical description of crystal spectrometers is shown to be very similar to that used for imaging devices but instead of studying data analysis methods which require an accurate description of the instrument response, the more fundamental problem of characterising and calibrating the response is addressed. Both theoretical and practical methods for finding crystal reflection parameters are discussed and then applied to three crystals; Langmuir-Blodgett lead stearate multilayers, gypsum 020 and beryl $10\bar{1}0$. Sophisticated theoretical calculations using an atomic model developed by other workers were used to predict the crystal response. Direct measurements of the response at a set of wavelengths through each crystal's range were made using a two crystal X-ray spectrometer. The combination of theory and measurement provide a nearly complete description of current pseudo lead stearate crystal production while the excellent agreement between theory and measurement for both gypsum and beryl demonstrates the power of both the theoretical and practical techniques employed. The results from all three crystal types provide excellent calibration data for use in subsequent spectral analysis using these crystals as

Bragg Analysers.

Measurement of the response across the sulphur k and aluminium k edges in gypsum and beryl respectively also provided direct experimental evidence of k electron resonance in these two atomic types.

Contents

PART I

THE ANALYSIS OF FOCUSED X-RAY IMAGES.

CHAPTER 1: THE NATURE OF THE IMAGING PROBLEM.

1.1 Introduction to Part I.

Optical plates and radio contour maps have been a major data source for astronomers for many years and their production and interpretation is well understood. Now X-ray astronomy stands on the brink of 2-D imaging with the development of hardware using grazing incidence optics to focus soft X-rays (0.1 to 5.0 KeV) into a conventional image and the first true 2-D X-ray maps of the sky are being made using sounding rocket payloads. The HEAO-B satellite, which was successfully launched on November 13th 1978, should provide a wealth of high quality 2-D image data to open a new era in X-ray astronomy.

Techniques for producing final X-ray maps of the sky are in their infancy and it will be some time before X-ray 'plates' approach the quality of their optical counterparts. This is partly due to present hardware limitations and partly because data processing techniques dedicated to astronomical X-ray images have not been developed.

A review of the problem facing X-ray astronomers is presented here and the results can be summarized by the following. Firstly the characteristics of the raw data must be fully studied and assessed. Secondly the surprisingly difficult task of deciding what is required of the final result must be tackled. This involves tricky subjective judgements concerning the 'information content' of the data and how it can best be represented. Finally, just as optical and radio astronomers have developed photographic, optical, Fourier analysis and many other

specialised methods, X-ray astronomers must draw on expertise in other fields to achieve a final result of high quality, worthy of their efforts.

1.2 The general characteristics of 2-D astronomical X-ray images.

The mean X-ray flux from the source $f(\alpha,\beta,t,E)$ photons $cm^{-2}st^{-1}sec^{-1}keV^{-1}$, where (α,β) is the angular position relative to the instrument, t is time and E is photon energy, is focused by conventional ray optics onto an image plane and is recorded by a position sensitive X-ray detector. The image is outputted as a set of events Jx_n,y_n,t_n,E_n; event n occurring at (x_n,y_n) in the counter position sensing plane at time t_n and having associated energy E_n (if available).

The measurement process can be modelled in two stages although this mathematical representation is by no means unique, merely convenient. Firstly the instrument performs an integral transform on the source distribution $f(\alpha,\beta,t,E)$ to give an image distribution $J(x,y,t',E')$ counts $cm^{-2}sec^{-1}keV^{-1}$

$$J(x,y,t',E') = \int k(\alpha,\beta,t,E,x,y,t',E') \, f(\alpha,\beta,t,E) \, d\alpha \, d\beta \, dt \, dE \qquad (1.1)$$

where k is known as the instrument kernel function. Secondly the detector measures a sample of J to provide the data set Jx_n,y_n,t_n,E_n:

$$Jx_n,y_n,t_n,E_n = S\{J(x,y,t',E')\} \qquad (1.2)$$

The discrete nature of the incident X-ray photon beam is

introduced in the sampling stage (1.2), whereas perhaps a more physical approach would trace the history of each photon through the instrument. Expressions (1.1) and (1.2) neatly divorce the instrument response from the statistical behaviour of the interaction of the X-rays with the telescope, providing a very powerful and useful representation of the behaviour of grazing incidence telescopes used in X-ray astronomy.

The form of the functions $S\{\}$ and k in (1.2) and (1.1) above must now be dealt with in detail. $S\{\}$ includes the statistical fluctuations in the data and the effect of non-photon background counts induced in the counter by cosmic rays which cannot all be discriminated out of the data set. The statistical fluctuation in the number of counts within an area of the detector sensing plane is governed by the Poisson distibution. For an area ΔA cm^2 sec keV at nominal position (x,y,t',E') in the instrument co-ordinates, with cosmic background count $C(x,y,t',E')$ counts $cm^{-2}sec^{-1}keV^{-1}$:

$$N(x,y,t',E') = \Delta A J(x,y,t',E') + \Delta A C(x,y,t',E') \quad \text{counts} \qquad (1.3)$$

and the fluctuations about the mean will have variance $\sigma^2 = N$ providing $N \geq 8$. For very low values of N a more accurate estimate of variance may be required (see reference 1), however this is not normally necessary. It can be seen from (1.3) that the statistics of the recorded image are dependent on the image function $J(x,y,t'E')$ which, of course, in turn is related to the source brightness distribution $f(\alpha,\beta,t,E)$. The noise content of

the image is therefore dependent on the form of $f(\alpha,\beta,t,E)$ (whether the source is a nebula or a star etc.) and on the behaviour of the instrument described by the function $k(\alpha,\beta,t,E,x,y,t',E')$.

A more detailed analysis of the instrument is now required to provide a breakdown of the instrument kernel. The first element of the telescope system is the mirror which can be arranged in a variety of geometries. All configurations involve the two grazing incidence reflections ($\leq 3^{o}$) off highly polished metal surfaces to give a focused image using conventional ray optics. The action of the mirror can be adequately described using but a few functions and parameters. The efficiency and image quality of the system can be handled separately, although they are not strictly separable, because the efficiency is only a slowly varying function of source position and the imaging performance is only weakly affected by photon energy.

The efficiency is conveniently expressed as a collecting area $A(E,\alpha,\beta)$ cm^2. For the Wolter Type I geometry the mirrors are hyperboloid and paraboloid with circular symmetry giving $A(E,\psi)$ cm^2, where ψ is the off-axis angle of the incoming X-ray beam. Note that $A(E,\psi)$ repr-sents an average efficiency over the complete proj cted aperture of the instrument. When using a bandwidth $\leq$ 1 keV, the collecting area can be expressed as the product of an energy dependent efficiency and a beam shape $A(E,\psi) = \mathfrak{z}_m(E)\ B(\psi)$. $\mathfrak{z}_m(E)$ will be a function of the mirror surface material and strongly modulated by absorption edges. $B(\psi)$ cm^2, as previously mentioned, is a slowly varying function of ψ defining a smooth beam

profile which is determined by the size and figuring of the surfaces.

The image quality of the mirror is best described by a point spread function, which for astronomical applications is the response to a parallel X-ray beam incident over the entire surface. The function $P_m(x_p'-x', y_p'-y', \alpha, \beta)$, describing the image of a point source at (α,β) as a function of image plane co-ordinates (x',y'), is a strong function of (α, β). $((x_p', y_p')$ is related to (α,β) by equation (1.4).) The figuring of the surfaces can give excellent on-axis performance which unfortunately degrades rapidly towards the edge of the field of view. For large Ψ the point spread is strongly asymmetric and can be affected by rogue, single reflection rays which cannot fully be stopped. The quality of the surface polish also affects the imaging performance. In general, it limits the on-axis performance by producing a scattering halo which envelops the image of a point source. This halo is not a strong function of (α, β) and is generally dominated by the figuring error at the extremities of the field of view.

The scale of the image in the focal plane is given by the focal length L mm. Using the optical axis as the origin of (α,β) and (x',y') and with an arbitrary rotation angle Θ:

$$(\alpha,\beta) = \{(x_\rho' \cos\Theta - y_\rho' \sin\Theta)/L, \quad (x_\rho' \sin\Theta + y_\rho' \cos\Theta)/L\} \qquad (1.4)$$

where x',y' are measured in mm and α,β are in radians. Expression (1.4) ignores any distortion introduced by

using a flat focal plane , which in practice is very small because α and β are both $\leq 2^{o}$ ($\equiv 3.5 \times 10^{-2}$ rads).

The mirror kernel can now be expressed using the efficiency and imaging parameters given above:

$$k_m = \zeta_m(E)\ B(\alpha,\beta)\ P_m(x_p'-x', y_p'-y', \alpha, \beta) \qquad (1.5)$$

In order that $\zeta_m(E)$ contains all the efficiency information $P_m(x_p'-x', y_p'-y', \alpha, \beta)$ must be normalised to 'conserve flux':

$$\int P(x_p'-x', y_p'-y', \alpha, \beta)\ dx'\ dy' = 1 \qquad (1.6)$$

Care must be taken in choosing the limits of integration in (1.6) so that all the blurred flux in the image plane is included.

The focused image is recorded by a specialised, position sensitive, X-ray detector placed at the focal plane. Two types of detector are currently available, both of which measure the position of a photon absorption event in, or near, the image plane as an (x,y) co-ordinate pair. In both devices - the imaging proportional counter and microchannel plate array - a charge is dumped onto a R-C line complex, the output pulses of which are used to derive the (x,y) for the initiating absorption event. The action of the R-C lines is very similar in the two cases, however the physical processes used to provide the initial charge are very different.

The imaging proportional counter uses a charge avalanche onto a grid of very thin anode wires to provide sufficient charge to drive the R-C lines and associated electronics. Ideally the photon absorption must occur at

the image plane to prevent blurring. However this is not possible because of the exponential absorption of the X-ray beam in the counter gas after penetration of the thin counter window. The resultant blurring is a strong function of energy dependent on the geometry of the mirror and the counter gas mixture and pressure. It is also a weak function of (α,β) since the entry angle of photons into the gas varies with source position. The 'gas blurring' can be expressed as a point response for the gas given by $P_g(x'',y'',x',y',E)$ (where (x'',y'') are absorption co-ordinates relative to the entry point in the image plane (x',y')), which is normalised in exactly the same way as P_m in equation (1.6) above. P_g represents the spatial response of the gas to a point source imaged by a perfect imaging mirror.

The small, compact charge cloud created by the photon absorption is then drifted onto the counter grid system which acts as an R-C line complex. It is sufficient to say that the overall response of the counter at this stage can be expressed by a new linear point response $P_1(x-x'',y-y'',E)$. The overall spatial response of the IPC is therefore given by the point response:

$$P_d(x-x',y-y',x',y',E) = \int P_g(x'',y'',x',y',E)\, P_1(x-x'',y-y'',E)\, dx''\, dy'' \qquad (1.7)$$

The efficiency of the counter is dependent on the window transmission and the absorption properties of the counter gas. At low energies the window dominates and provided that the thickness of the window is uniform, the efficiency will be independent of (x',y') and simply a

function of the material and thickness $\mathfrak{z}_d(E)$.

A proportional counter provides energy information on each detected photon. The response of the system takes the form of a non-linear convolution of the input system with a response function R(E-E',E). Hence given a spectrum G(E), the result has the form:

$$G'(E') = \int \mathfrak{z}_d(E)\ G(E)\ R(E-E',E)\ dE \qquad (1.8)$$

G'(E') is then sampled as individual photons by the function S{} as given by equation (1.2). The efficiency $\mathfrak{z}_d(E)$ must obviously be included to describe the actual pulse height spectrum which appears in the data.

The microchannel plate array consists of a honeycomb of microscopic glass tubes of diameter $\approx 15\mu$. The interior walls of the tubes are coated with MgF which acts as a photocathode. X-rays hitting the inner walls of the tubes release an electron. A high p.d. is placed across the plate to drift the electrons down the tubes creating secondary electrons on the way. After several steps the resulting total charge is dumped on an R-C line complex. Under proper working conditions the performance of the plate is limited by the tube diameter which bins the incident photon flux. The overall response can be expressed as a single linear point response $P_d(x-x',y-y')$ which is independent of the angle of entry of the photon relative to the plate normal.

The efficiency of the microchannel plate is dependent on the photon entry angle, however when used with a grazing incidence mirror the entry angle does not alter appreciably over the field of view and hence the

efficiency is given adequately by $\mathcal{J}_d(E)$ in a similar fashion to the IPC.

The overall detector response is therefore given by:

$$k_d = \mathcal{J}_d(E)\ P_d(x-x',y-y',x',y',E)\ R(E-E',E) \qquad (1.9)$$

where $P_d(x,y,x',y',E)$ reduces to $P_d(x-x',y-y')$ and $R(E-E',E)$ is unavailable for the MCP detector. Equations (1.5) and (1.9) combine to yield the complete instrument kernel:

$$\begin{aligned} k(\alpha,\beta,t,E,x,y,t',E') = {} & \mathcal{J}_m(E)\ \mathcal{J}_d(E)\ R(E-E',E) \\ & B(\alpha,\beta) \int \{P_m(x_p'-x',y_p'-y',\alpha,\beta) \\ & P_d(x-x',y-y',x',y',E)\}\ dx'\ dy' \end{aligned} \qquad (1.10)$$

The above analysis provides a full description of the instruments' action going from the initial surface brightness distribution $f(\alpha,\beta,t,E)$ to the measured data set Jx_n,y_n,t_n,E_n. Since any such instrument employed by X-ray astronomers will be aboard a space vehicle, the data set Jx_n,y_n,t_n,E_n will be transmitted to Earth by a telemetry link. The number of bits available per data element must be adequate so that the experiment is instrument performance rather than telemetry limited and the bandwidth must be sufficient to handle the data rate expected when imaging astronomical sources. If insufficient bits are used then the form of the above equations will be irrelevant and no amount of subtle post-processing of data can improve the image quality.

The explicit form of the functions used to describe the instruments' operation is needed to produce good

results and the next section deals with specific systems to provide some idea of what this entails.

1.3 The performance of existing X-ray telescope systems.

The general form of the instrument response is summarised by the functions $C(x,y,t',E')$, the cosmic background count, $\eta_m(E)$ the mirror efficiency, $B(\Psi)$ the beam shape of the mirror, $P_m(x_p'-x',y_p'-y',\alpha,\beta)$ the point response of the mirror, $P_g(x'',y'',x',y',E)$ the gas response (only present for gas detectors), $P_1(x-x'',y-y'',E)$ the detector R-C line point response, $\eta_d(E)$ the detector efficiency and $R(E-E',E)$ the energy response (only applicable to proportional counters) of the detector.

Examples are given here of these functional forms for existing systems. Whenever possible, measured values are given but this is not always easy because of measurement difficulties. Details of three specific payloads are given; the payload built by MIT/ Leicester University for launch on an Aerobee sounding rocket to image supernova remnants in soft X-rays, secondly the Leicester University/ MPI Germany collaboration to build a payload capable of imaging an X-ray dust scattering halo about a point source to be flown on a Skylark sounding rocket and thirdly the HEAO-B satellite built by a large consortium centred on SAO.

All three utilise the Wolter Type I mirror geometry and an IPC. Only HEAO-B, which is a high resolution instrument, employs a MCP as a detector to provide high quality ($\sim$2"), small field plates to complement the wider field, relatively low resolution ($\sim$1') plates. The

performance of a complete system, mirror plus detector, cannot be calculated directly from the individual responses because of the gas blurring effect, which depends both on the mirror geometry and the counter gas.

The MIT mirror.

The mirror assembly consists of a nested pair of mirrors, each consisting of a front paraboloid and rear hyperboloid giving a focal length of 1143 mm. The on-axis response was found experimentally to have the form:

$$P_m(0-x',0-y',0,0) = \frac{1}{2^{\frac{1}{2}}\pi^{\frac{3}{2}}\sigma_m} \frac{\exp\{-(x'^2+y'^2)/2\sigma_m^2\}}{(x'^2+y'^2)^{\frac{1}{2}}} \qquad (1.11)$$

where σ_m = 2.8'. Moving off-axis caused degradation to the symmetry of the response but σ_m can still be used to provide a good indication of performance. Figure 1 shows the theoretical and measured values of σ_m as a function of off-axis angle Ψ. Figures 2 and 3 show the measured response of the mirror to a slit collimated beam giving the line response of the mirror at two energies: 0.28 keV and 1.5 keV. Figure 4 shows the measured effective area of the mirror; $A(\Psi,E)$ for three spot energies E. The unnormalised beam shape is only a weak function of energy but the absolute value of $A(\Psi,E)$ is strongly modulated by $\mathfrak{Z}_m(E)$. Figure 5 shows the theoretical efficiency of the nickel mirror surface for a normal grazing incidence angle of $\langle \theta_g \rangle = 2^o$. The quality of the mirror polish limits the actual efficiency to about 30% of theoretical but leaves the gross absorption edge structure unaltered.

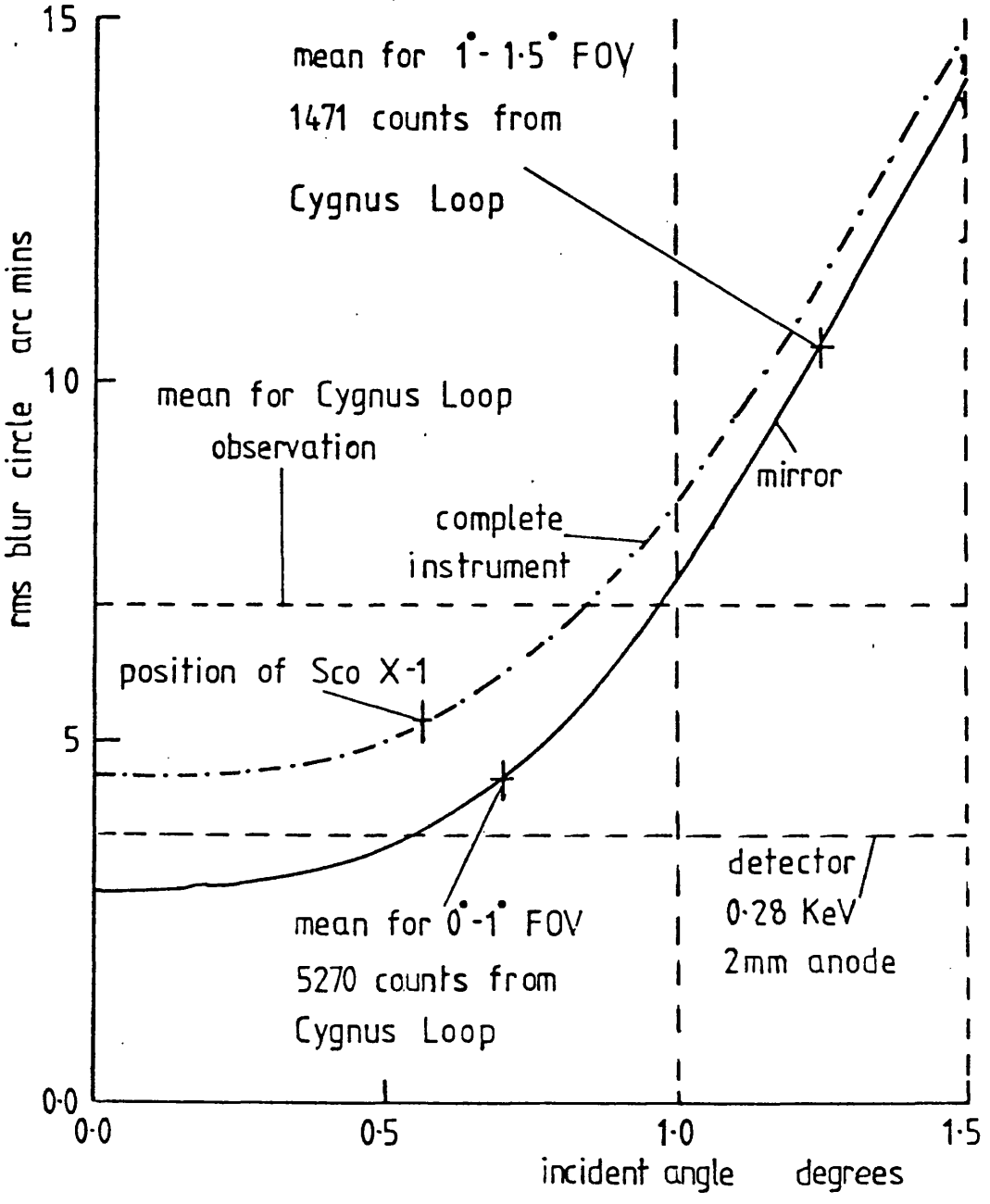


Figure 1. Resolution of the MIT/ Leicester payload used to observe the Cygnus Loop.

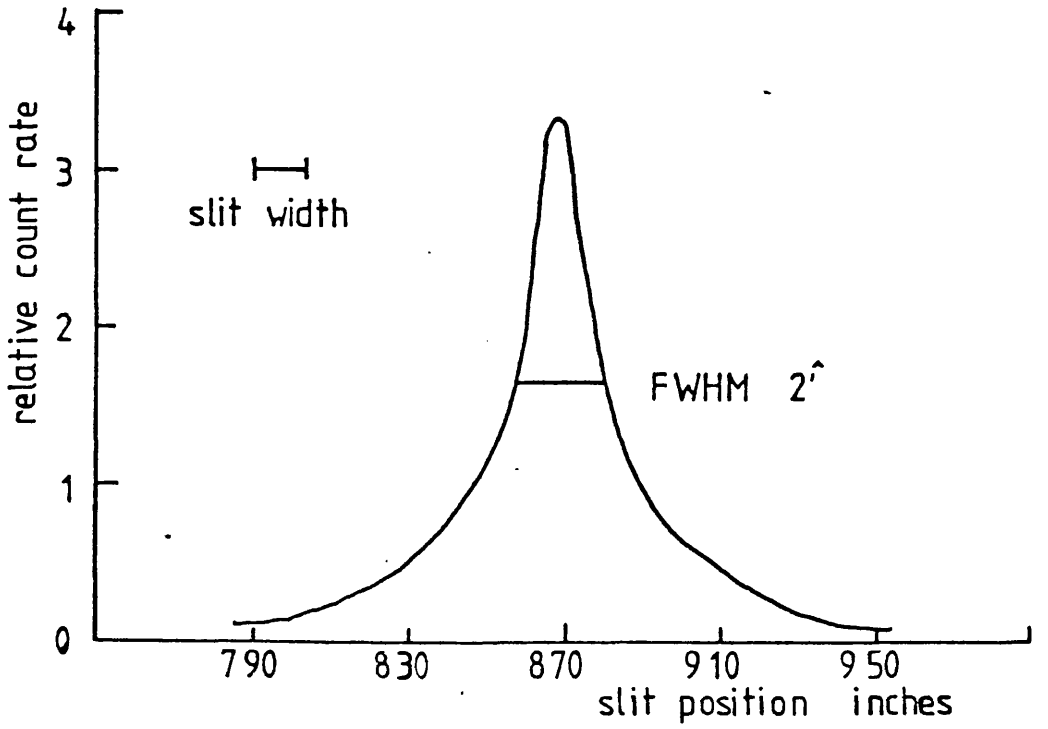


Figure 2. Measured line response of MIT/Leicester payload using 1 mm anode grid, Al_k 1.5 keV X-rays.

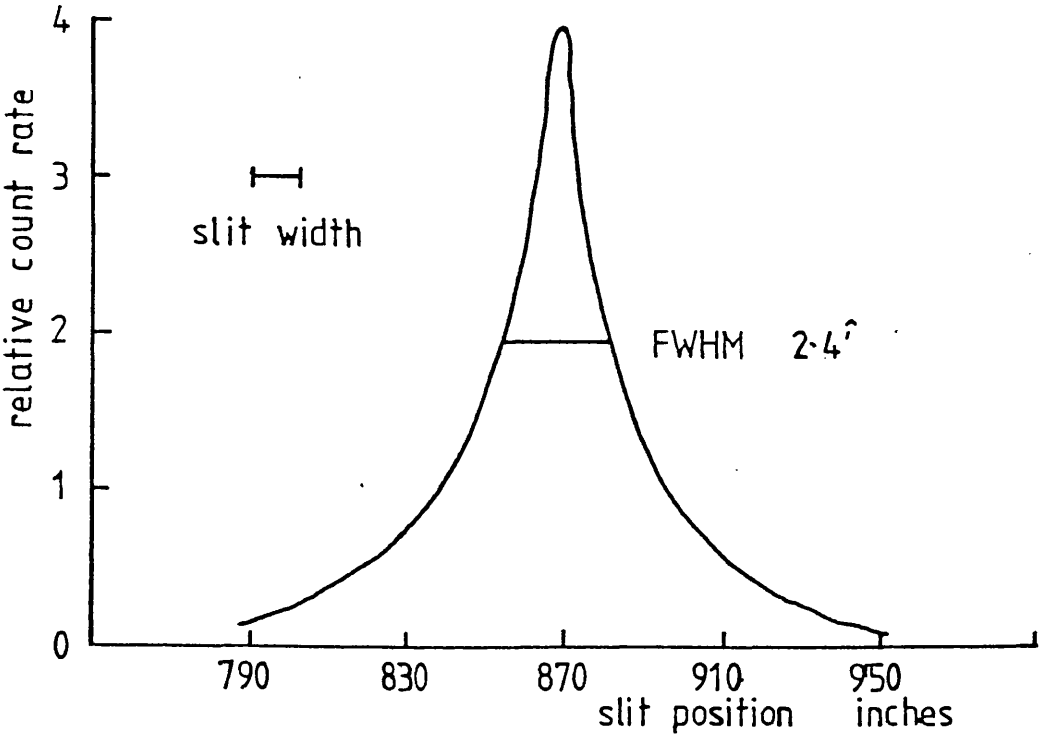


Figure 3. Measured line response of MIT/Leicester payload using 1 mm anode grid, C_k 0.28 keV X-rays.

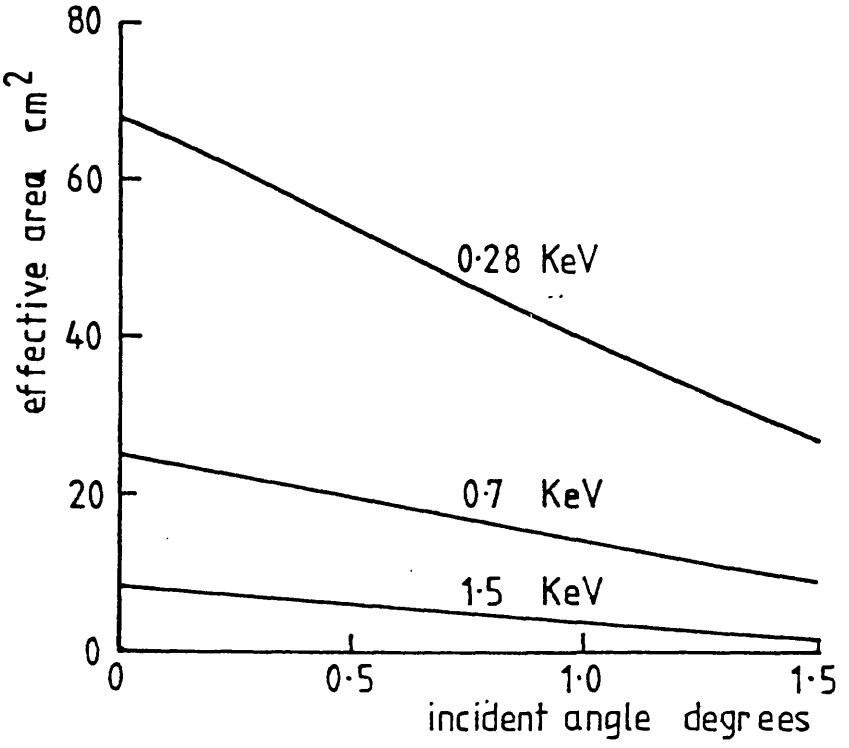


Figure 4. Effective area of the MIT/Leicester payload.

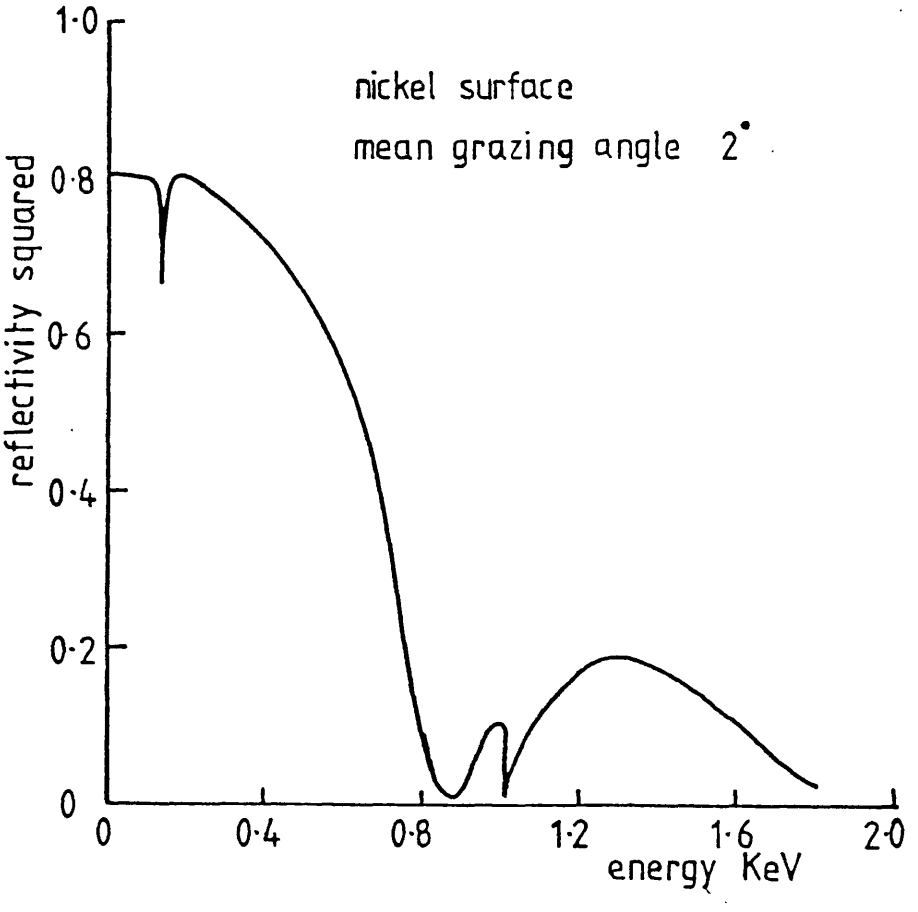


Figure 5. MIT mirror reflectivity (theoretical).

The MPI mirror.

The MPI mirror has a focal length of 1427 mm and was designed to have a very high image quality free from blurring due to figuring errors or a scattering halo due to polishing limitations. Figure 6 shows the expected performance based on theoretical figuring limitations and scattering measurements performed on optical flats polished to the mirror standard. The function is expressed as the fractional flux in an anulus of width $d\theta$ at radius θ from a point source on-axis. Curves 27 and 29 indicate the expected limits of the scattering halo. Off-axis aberrations affect the performance and the resulting degradation is summarised by figure 7. The effective area has a very similar functional form to the MIT mirror and theoretical curves for $A(\psi, E)$ and $\zeta_m(E)$ are given in figures 8 and 9. Again, the efficiency achieved is limited by surface quality and is about a factor of 3 below the theoretical curves.

The Leicester IPC.

Both the MIT and MPI payloads use the Leicester IPC as a detector. The gas mix used is A/ CH_4 in various possible proportions which must be optimised to provide adequate gain without breakdown. The X-ray flux focused by the mirror enters the detector in a cone, the shape of which is only weakly dependent on source position. The entry angle θ_e for the MIT mirror is $\approx 9^\circ$ while that for the MPI mirror is $\approx 6^\circ$. The X-rays suffer exponential absorption in the gas giving the response:

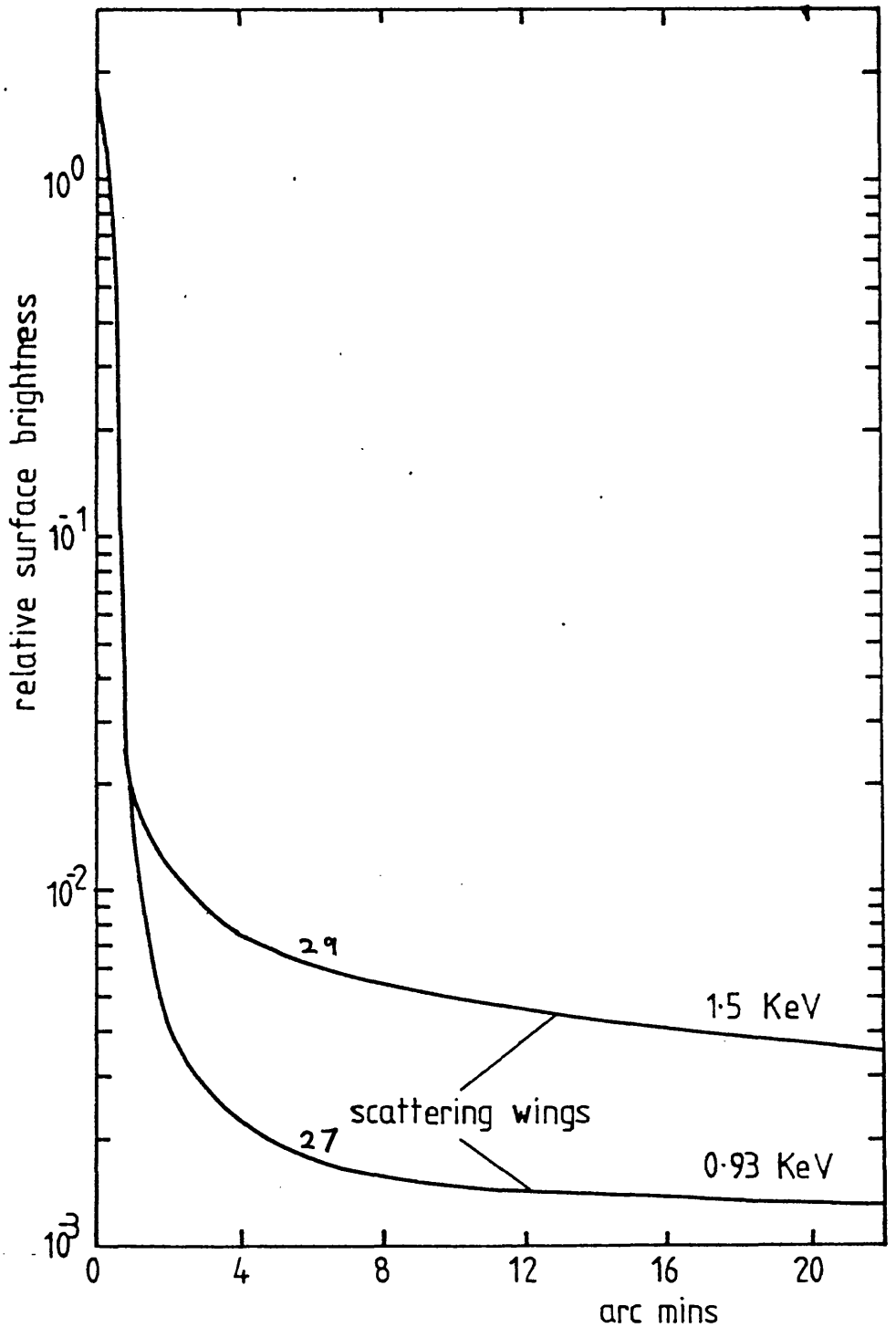


Figure 6. Point response of MPI mirror.

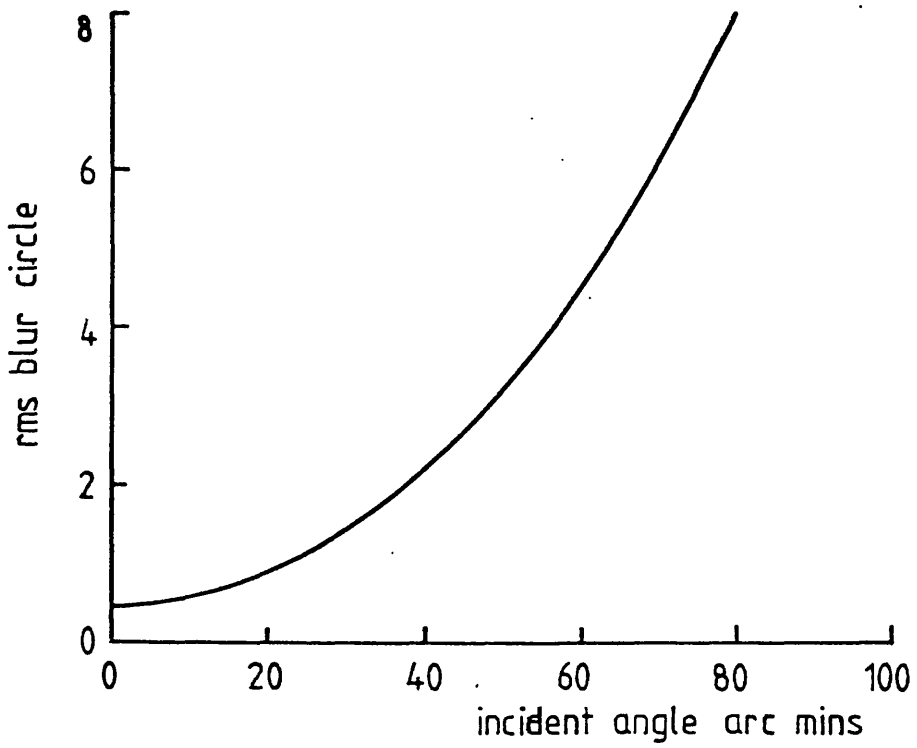


Figure 7. Resolution of the MPI mirror.

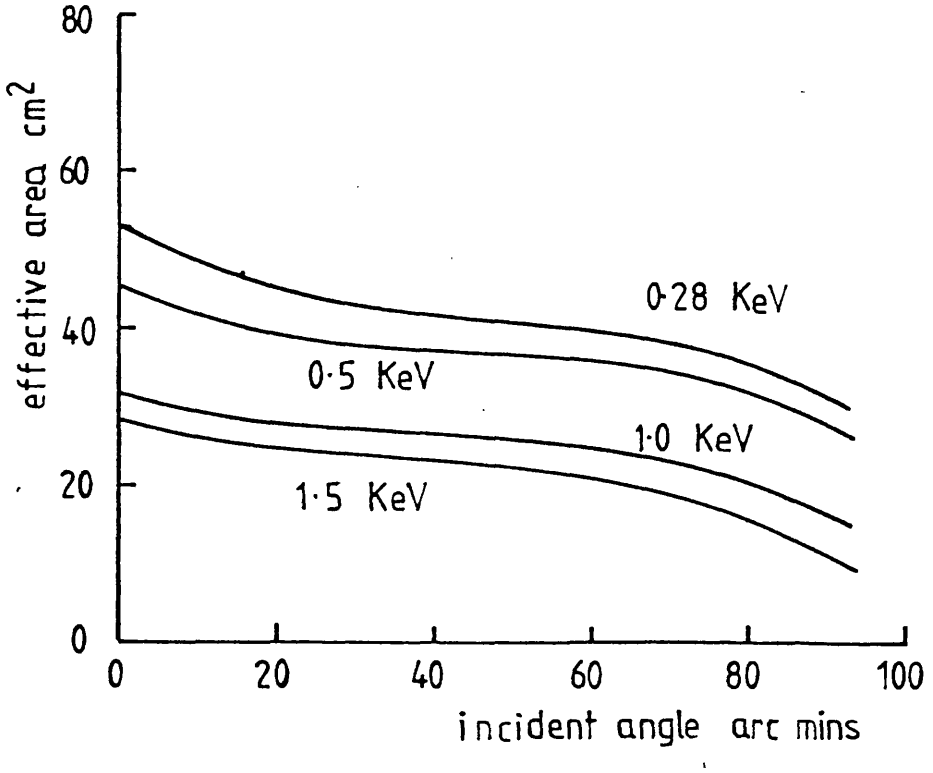


Figure 8. Effective area of the MPI mirror.

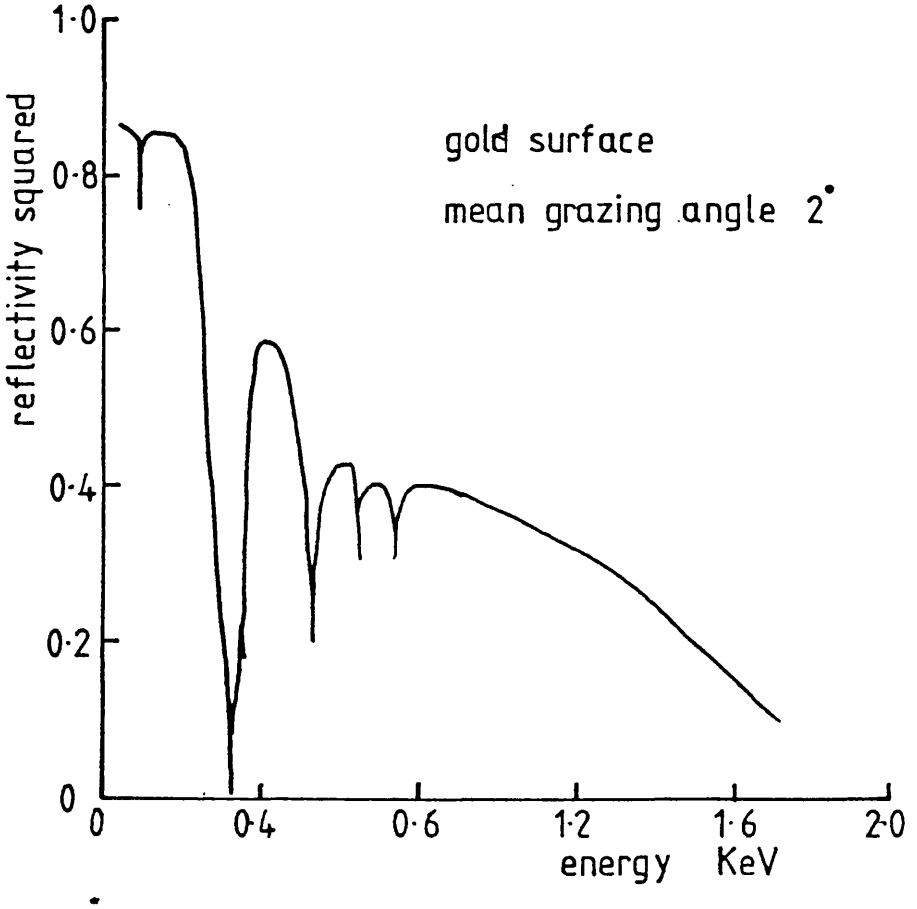


Figure 9. Theoretical MPI mirror reflectivity.

$$P_g(x'',y'',0,0,E) = \frac{1}{2\pi d_g(E)\ \sin\theta_e (x''^2+y''^2)^{\frac{1}{2}}} \qquad (1.12)$$

$$\exp\left\{-(x''^2+y''^2)^{\frac{1}{2}}/(d_g(E)\ \sin\theta_e)\right\}$$

where $d_g(E)$ is the 1/e depth for X-rays of energy E in the counter gas. The parameter $\sigma_g(E) = d_g(E) \sin\theta_e$ conveniently describes the gas blurring for a particular instrument as a function of energy and figures 10 and 11 show $\sigma_g(E)$ for the MIT and MPI payloads respectively. The contribution of the gas absorption to the overall blurring of the image clearly becomes important for E > 1 keV.

The R-C line response of the counter has been extensively measured both parallel and perpendicular to the anode wires and at different energies. The performance depends on the pitch of the anode grid and the counter gain and can be expressed in terms of the product of the responses parallel and perpendicular to the anode wires using a Gaussian form:

$$P_1(x-x'',y-y'',E) = \frac{1}{2\pi\sigma_x(E)\sigma_y(E)} \exp\left\{\frac{-(x-x'')^2}{2\sigma_x(E)^2} - \frac{(y-y'')^2}{2\sigma_y(E)^2}\right\} \qquad (1.13)$$

$\sigma_x(E)$ and $\sigma_y(E)$ characterise the R-C line response as a function of photon energy E and values for these parameters are given in figure 12.

The detector efficiency $\eta_d(E)$ is entirely dependent on the window transmission function including all supports etc. and not the gas since all photons entering the counter are trapped because of the very small absorption depth. Measured values for the thin

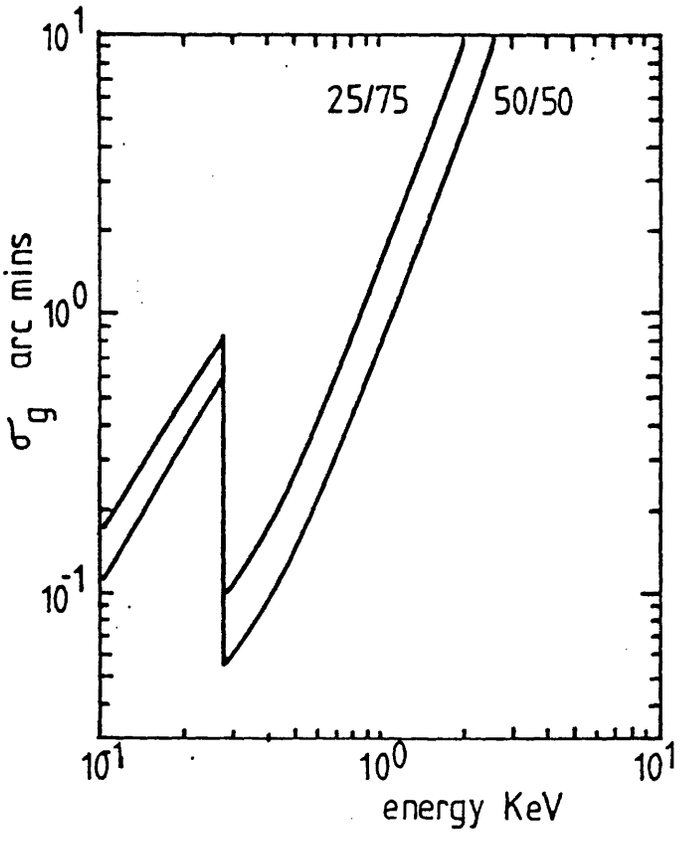


Figure 11. $\sigma_g(E)$ using the MPI mirror, A/CH_4 gas mix at 0.28 keV.

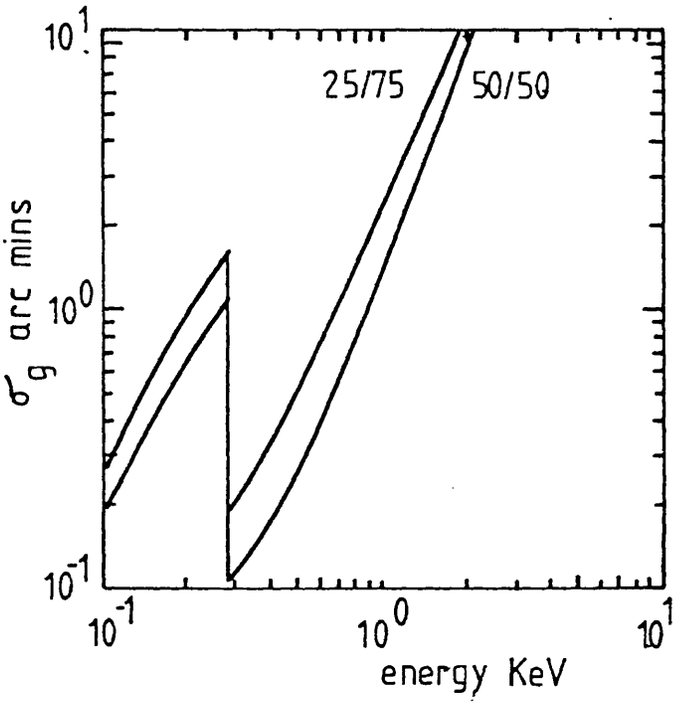


Figure 10. $\sigma_g(E)$ using the MIT mirror, A/CH_4 gas mix at 0.28 keV.

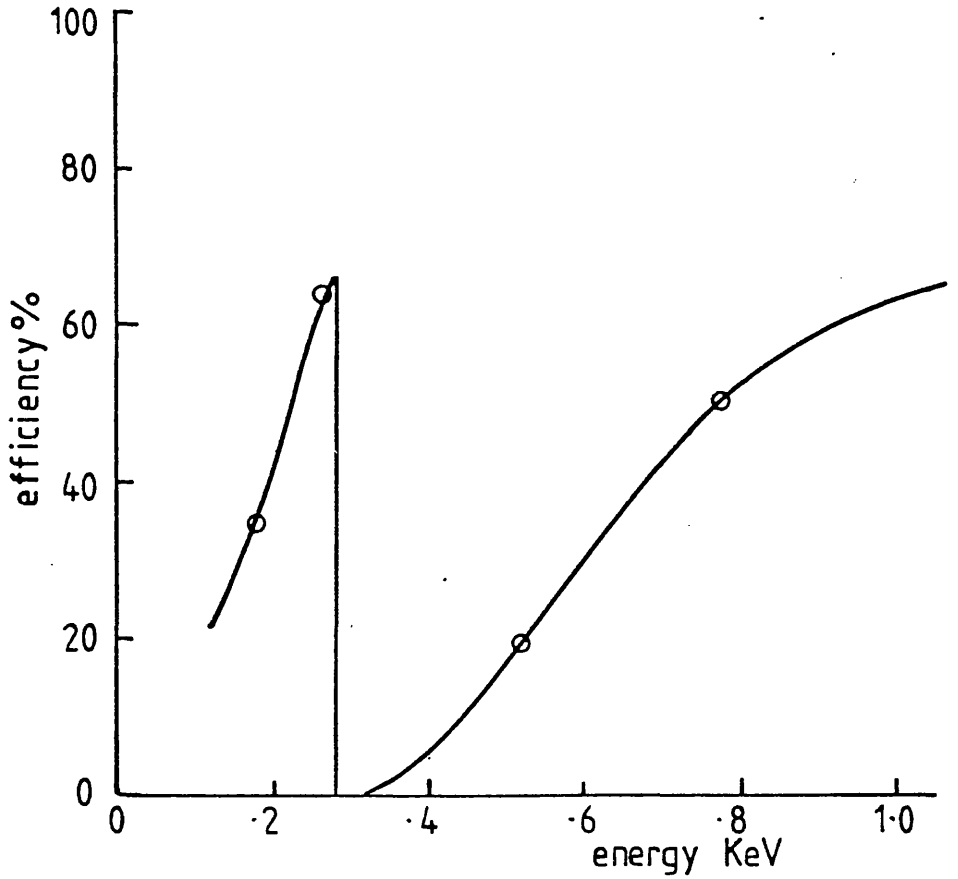


Figure 13. Efficiency of the Leicester IPC.

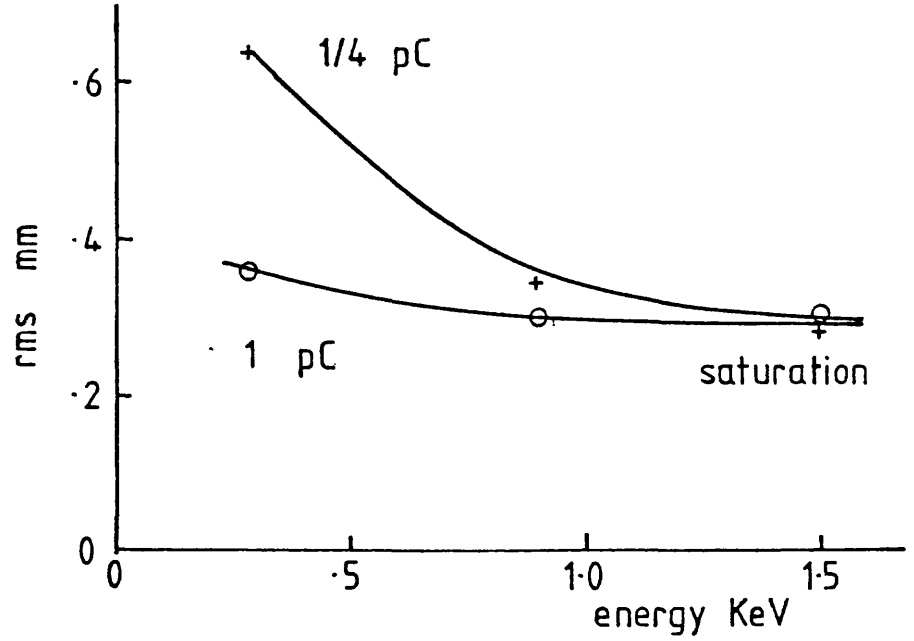


Figure 12. rms blur of the Leicester IPC, 1 mm grids, 25/75 A/CH_4.

polypropylene window covered in carbon dag used are given in figure 13, the major feature present being the C_K absorption edge at 0.283 keV.

The HEAO-B satellite.

The mirror consists of a nested set of four Wolter Type I geometry surfaces with a focal length of 2450 mm giving 1" ≡ 16.6μ in the focal plane (1 mm = 1 arc min.). Either an IPC or MCP detector can be used to provide large field/low resolution and small field/high resolution images respectively. When used with the MCP, the overall image quality is dominated by the mirror because of the very high performance of the MCP with a resolution of ≈ 1", limited only by the channel diameter. Therefore figure 14 shows the complete telescope response on-axis.

Figures 15 and 16 show the fractional effective area within a given radius, thereby showing the distribution of point source power and clearly showing the large scattering wings which are present out to very large radii. Figure 17 shows the full range of the point response, again highlighting the faint but extensive scattering wings. Figure 18 shows the blur circle degradation as a point source moves off-axis following the same pattern as demonstrated by figures 1 and 7 for the sounding rocket mirrors above. Figure 19 shows the beam $A(\psi,E)$ for the HEAO-B mirror.

The IPC used on HEAO-B has very similar characteristics to the Leicester version already described, giving a resolution of ≈ 1' ≡ 1 mm in the focal plane. As already mentioned the MCP device has very good resolution. The

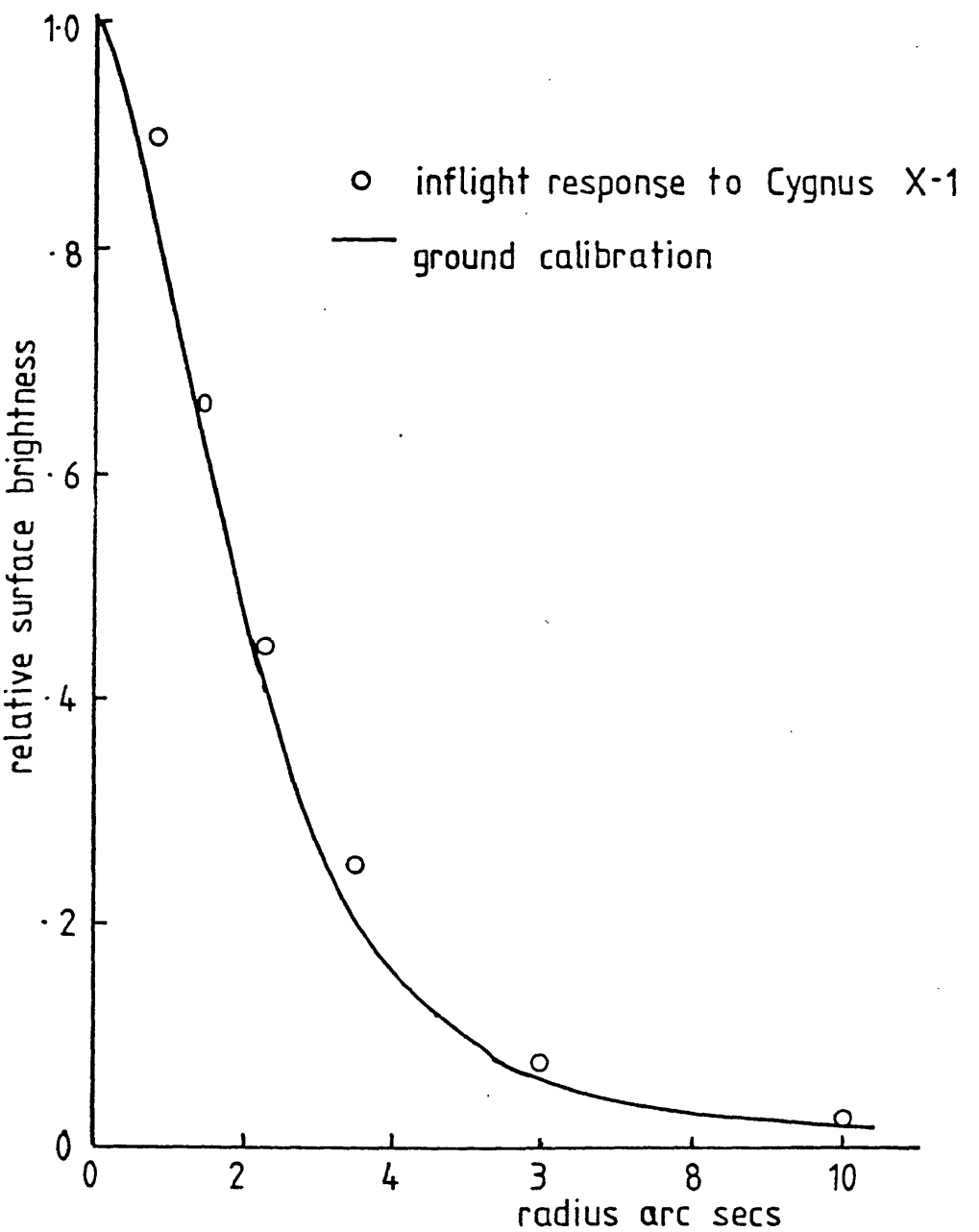


Figure 14. Measured resolution of HEAO-B mirror with HRI.

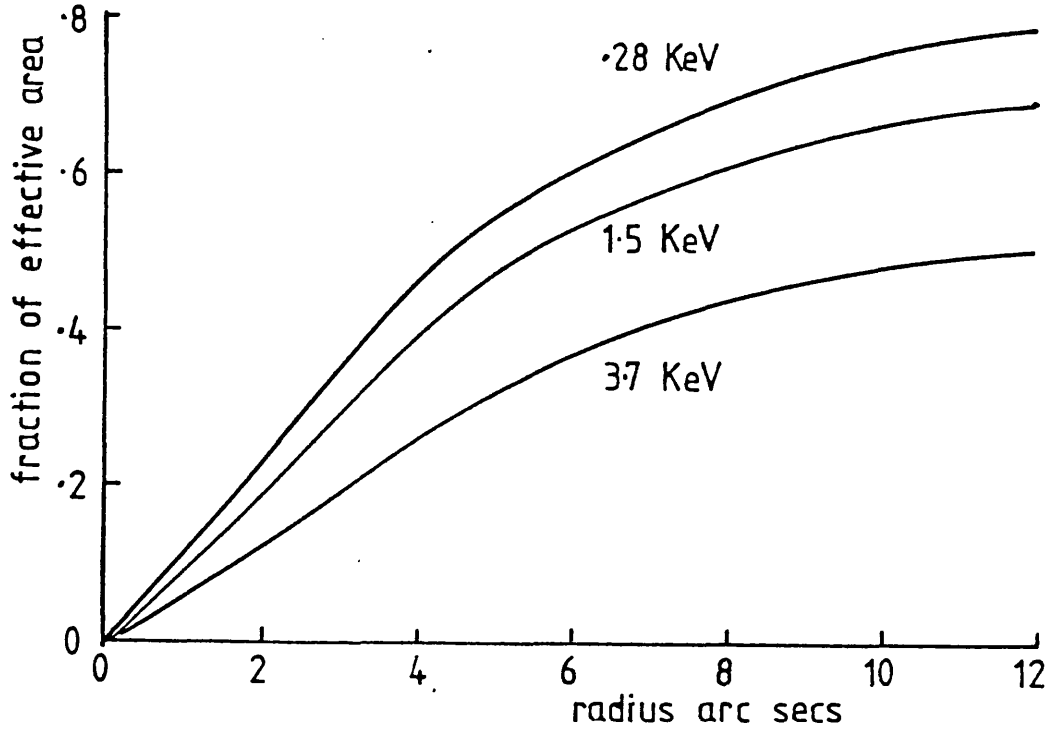


Figure 15. Scattering wings of the HEAO-B mirror, small angles.

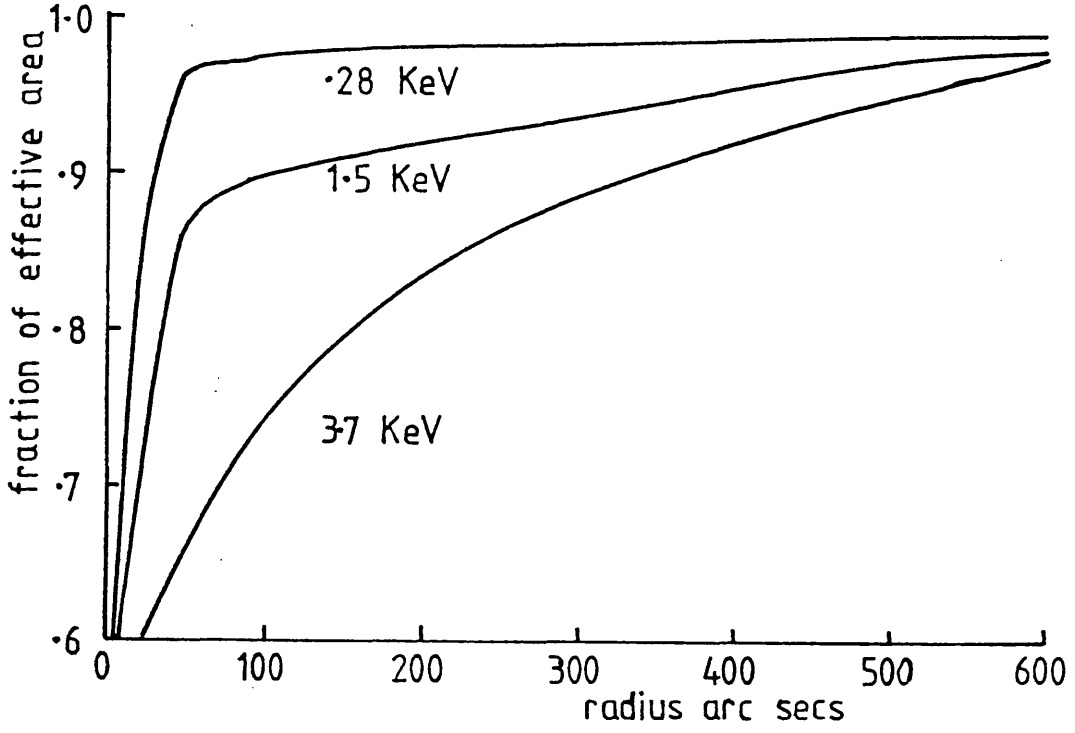


Figure 16. Scattering wings of the HEAO-B mirror, large angles.

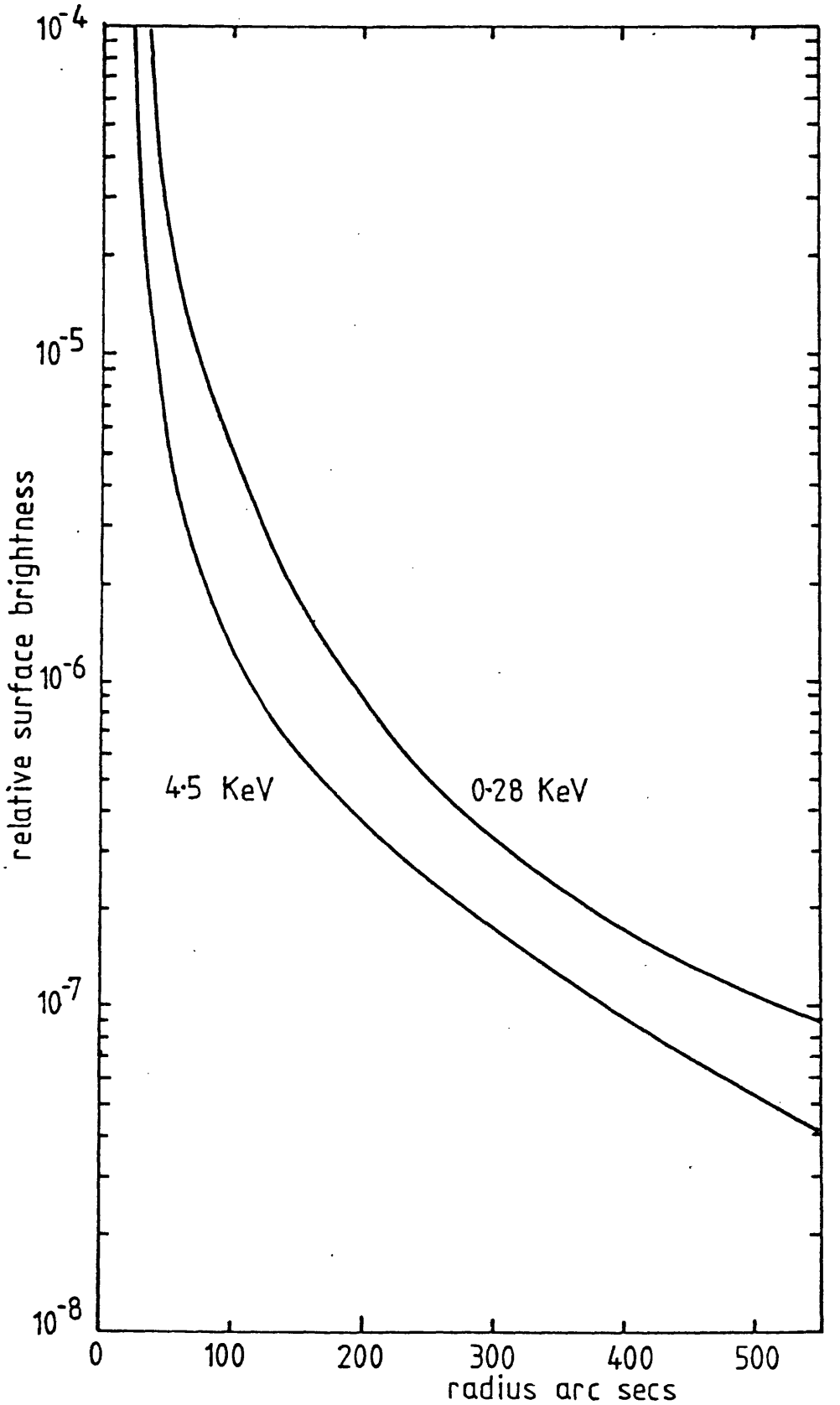


Figure 17. Point response of the HEAO-B mirror scattering wings at large angles.

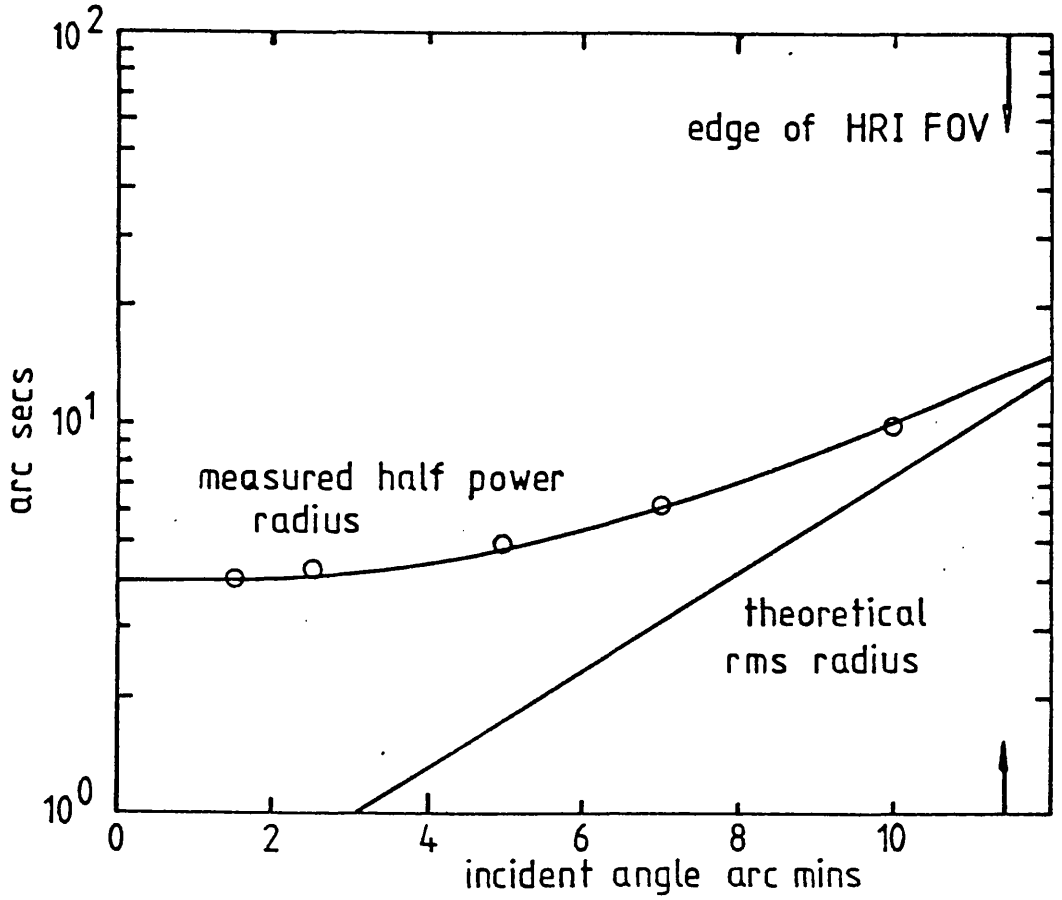


Figure 18. Resolution of the HEAO-B mirror + HRI.

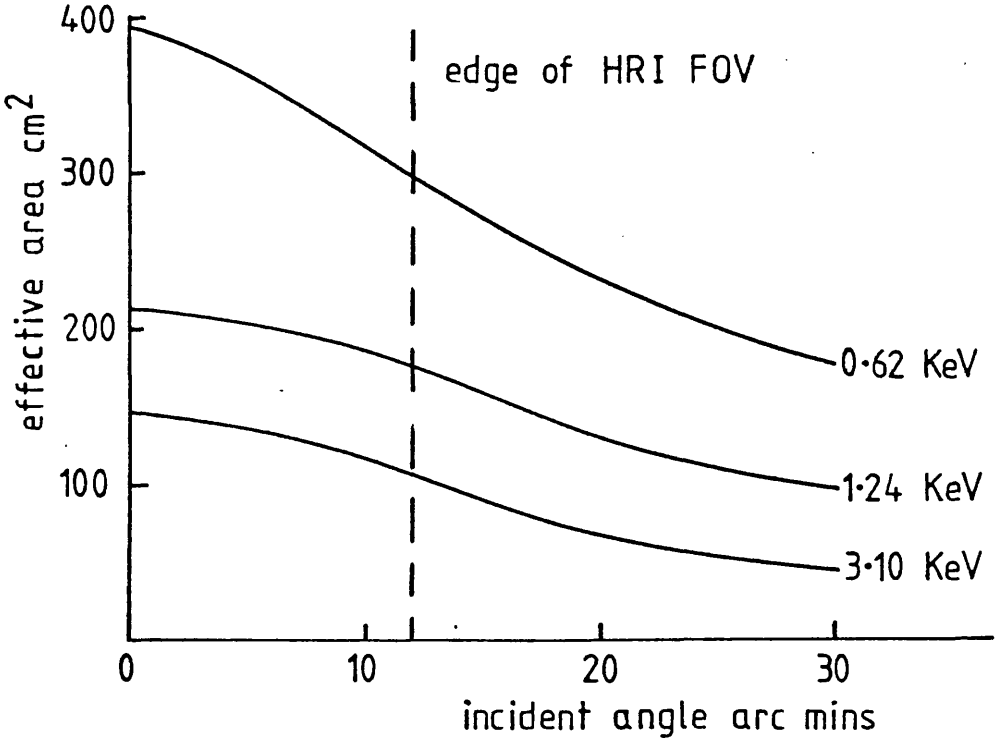


Figure 19. Vignetting function of the HEAO-B mirror.

efficiency of the MCP $\eta_d(E)$ is given by figure 20, which demonstrates how the collecting area of the complete telescope drops dramatically with increasing energy. This can be compared to the IPC efficiency used with the mirror given by figure 21.

Background counts in the detector.

The function $C(x,y,t',E')$ depends on the structure of the instrument and external environment and cannot be accurately estimated before flight. However the background counts are normally fairly isotropic in (x,y) with a flat spectrum in E'. $C(x,y,t',E')$ reduces to a count/mm^2/sec/keV independent of position in the field of view. Typical values are 4×10^{-3} counts mm^{-2} sec^{-1} keV^{-1} for an IPC and 3×10^{-3} counts mm^{-2} sec^{-1} for a MCP detector over its complete bandpass as shown by figure 20.

1.4 The quality of astronomical X-ray image data.

The form of the data is an event set Jx_n, y_n, t_n, E_n and the quality of the raw image is determined by three distinct phenomena. These are the blurring introduced by the instrument response, statistical fluctuations inherent to the X-ray beam and unrejected background counts from the detector. All processing on the event set Jx_n, y_n, t_n, E_n must be in sympathy with the form and quality of the data to provide the best image.

In general, the source surface brightness $f(\alpha, \beta, t, E)$ photons cm^{-2} st^{-1} sec^{-1} keV^{-1} is low and unless long exposure times are used the total count will be very low, typically a few thousand counts. A raw image can be

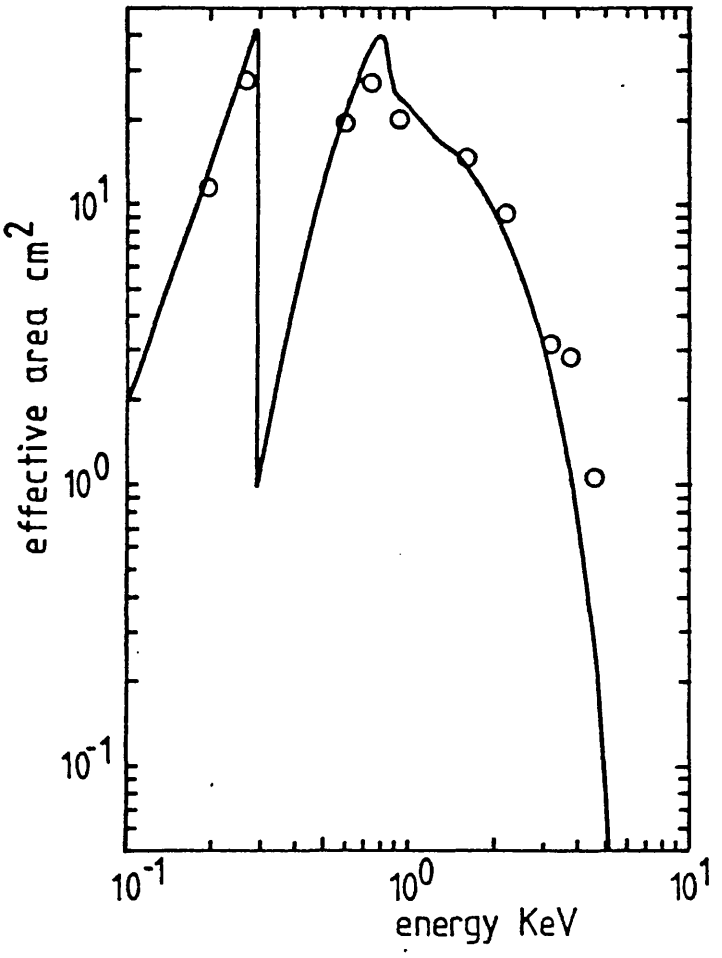


Figure 20. Effective area of HEAO-B HRI-2.

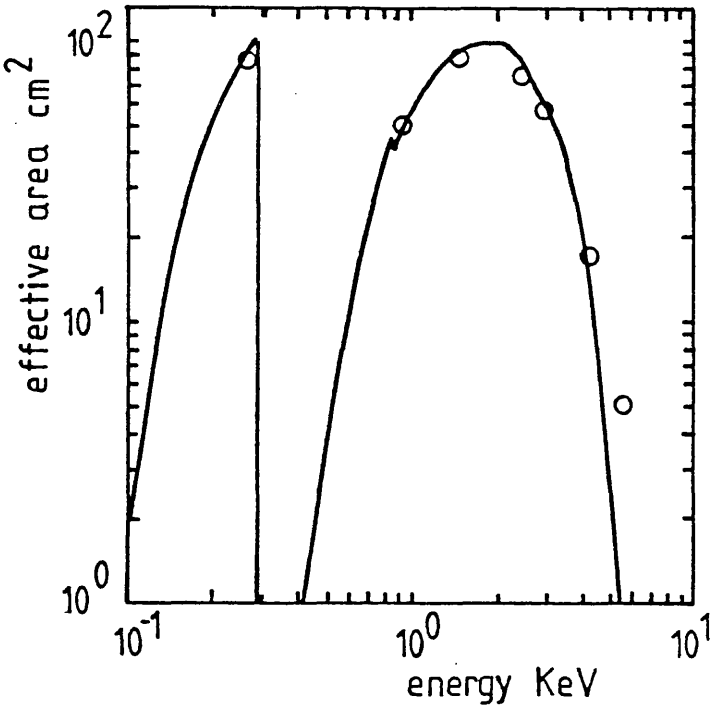


Figure 21. Effective area of HEAO-B IPC-1.

obtained from Jx_n, y_n, t_n, E_n by plotting the individual positions (x_n, y_n) from a chosen energy and time band using crosses, dots or some other suitable mark. However this is inconvenient and does not present an easily assimilable representation of the required distribution.

A better approach is binning the data in cells $x \to x+\Delta x$ and $y \to y+\Delta y$ producing a matrix of numbers representing the count within each cell. This can readily be done by truncating the binary representation of the positions (x_n, y_n) either by raster display hardware or computer software. Since the events are telemetered to the ground by digital techniques, the data is already binned at the bit limit of the telemetry. The choice of this inherent bin size and any subsequent imposed binning must be made with care since it potentially limits the image quality. The final presentation of $\hat{f}$ must, as far as possible, be free from distortion and degradation introduced by the digital processing as well as from the blurring and statistics described above.

Binning is a two stage process; firstly the data selected from chosen ranges of E_n and t_n expressed as the function $J(x,y)$ is convolved with the top hat function $T = 1$, $-\Delta x/2 < x < \Delta x/2$, $-\Delta y/2 < y < \Delta y/2$ and $T = 0$ otherwise. Secondly the resulting function is sampled by a regular grid of delta functions to give the image matrix J_{ij}:

$$J_{ij} = \delta(x-i\Delta x, y-j\Delta y) \int J(x',y')\, T(x-x',y-y')\, dx'\, dy' \qquad (1.13)^a$$

The sampling function $T(x,y)$ imposed by least significant

bit truncation is necessarily a top hat but general forms e.g. a Gaussian sampling function could be imposed using computer software. The top hat form is by far the easiest to apply in practice, although it is not necessarily the best.

The production of the image matrix J firmly places the data into a digital environment and marks the end of the imaging stage. Any subsequent data handling and display techniques can be considered to be post-processing. However this distinction is arbitrary since both analog and digital methods are required to obtain J_{ij} and the image is still in an undeveloped stage analogous to a freshly exposed photographic emulsion. In contrast to the highly sophisticated photographic methods required to produce good quality optical plates, the processing of J_{ij} must be digitally biased and relies on large computers utilising many techniques which are still in their infancy.

Figure 22 provides a summary of the stages of the imaging process in the form of a flow diagram.

1.5. The introduction of discrete notation.

The action of a digital processor is best stated using vector, matrix and tensor analysis. The continuous function analysis used so far must be carefully transcribed into discrete notation without destroying the validity of the statements. The discrete description can then be used to prescribe the digital processing.

Using the sampling function $T(x,y,t,E)$ defined by

$T = 1$ for $-\Delta x/2 < x < \Delta x/2$, $-\Delta y/2 < y < \Delta y/2$,

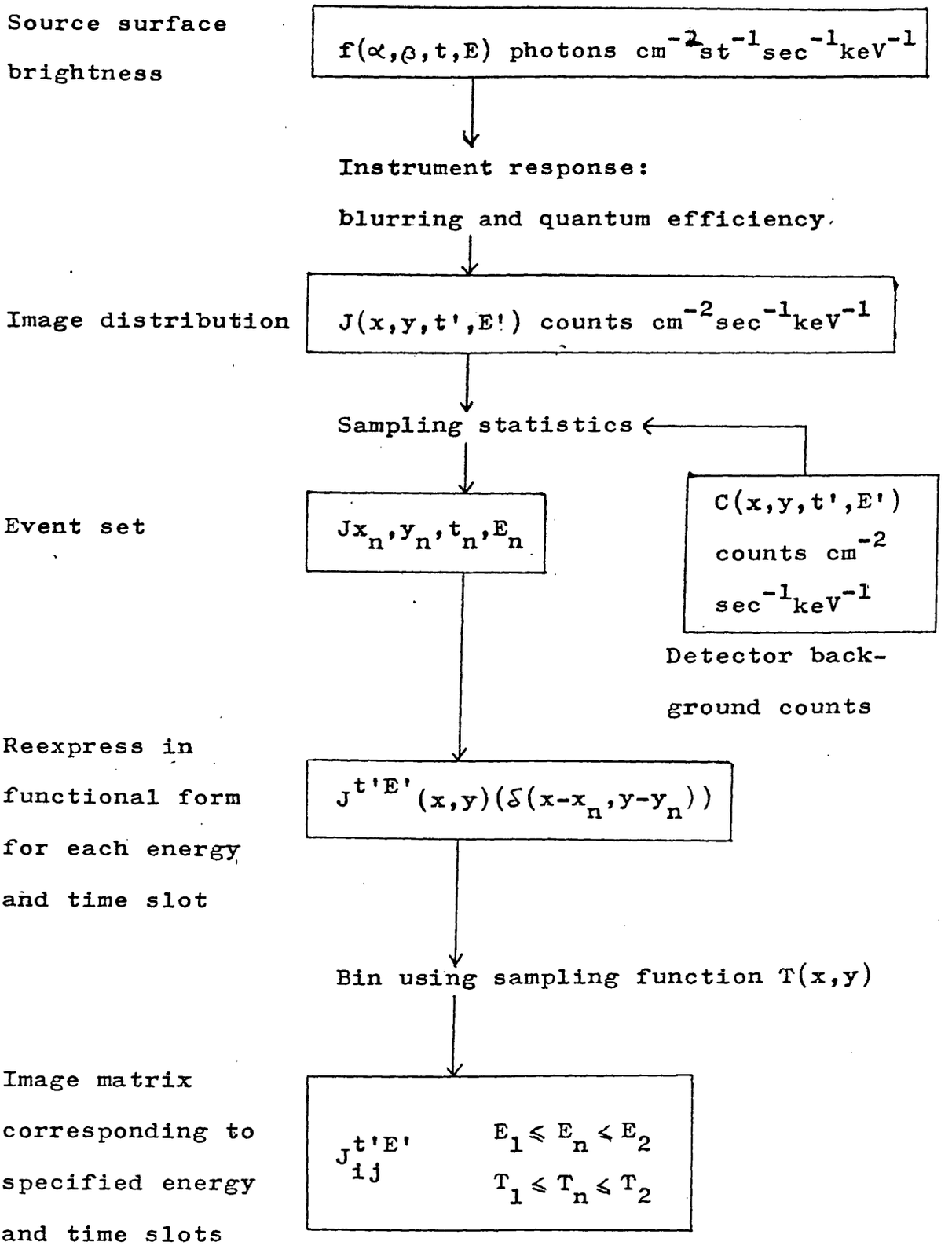


Figure 22. Flow diagram of the imaging process.

$-\Delta t/2 < t < \Delta t/2$, $-\Delta E/2 < E < \Delta E/2$ and $T = 0$ otherwise, each of the functions in figure 22 can be sampled to give a tensor:

$$Y_{ijkl} = \delta(x-i\Delta x, y-j\Delta y, t-k\Delta t, E-l\Delta E) \int Y(x',y',t',E')\ T(x-x',y-y',t-t',E-E')\ dx'\ dy'\ dt'\ dE' \tag{1.14}$$

Equation (1.14) can be conveniently expressed in operator form:

$$Y_{ijkl} = V\{Y(x,y,t,E)\} \tag{1.15}$$

where V is the sampling operator. Providing the increments Δx, Δy, Δt and ΔE are small enough the following equation can replace the continuous analysis summarised in figure 22:

$$J_{ijkl} = \sum_{\alpha\beta tE} k_{ijkl\alpha\beta tE}\ f_{\alpha\beta tE} + N_{ijkl} \tag{1.16}$$

The first term is the sampled image distribution $V\{Jx,y,t',E'\}$ and the second term N_{ijkl} introduces the statistics from both the source beam and the detector background. Each image element J_{ijkl} represents the total count made in the sampling volume $\Delta A_{ijkl} = \Delta x\ \Delta y\ \Delta t\ \Delta E$ at position $(i\Delta x, j\Delta y, k\Delta t, l\Delta E)$. This count will suffer statistical fluctuations governed by the Poisson Distribution and previously stated in equation (1.3). The noise process N_{ijkl} will have a mean of C_{ijkl}. the background count, and a variance given by:

$$\sigma_{ijkl}^{\ 2} = \sum_{\alpha\beta tE} k_{ijkl\alpha\beta tE}\ f_{\alpha\beta tE} + C_{ijkl} \tag{1.17}$$

the mean count, signal plus background, received in the

sampling volume ΔA_{ijkl}.

The general equation (1.16) is very cumbersome and since the total count is only likely to be a few thousand, the image tensor J_{ijkl} will be very sparse with most elements set to zero. Except for a weak spatial response dependence on energy shown by the IPC, as shown by figures 10, 11 and 12, the spatial, temporal and energy variables are separable and equation (1.16) can be written:

$$J_{ijkl} = \sum_{\alpha\beta tE} k'_{ij\alpha\beta}\, \delta_{kt}\, k''_{lE}\, f_{\alpha\beta tE} + N_{ijkl} \qquad (1.18)$$

where

$$k'_{ij\alpha\beta} = V\{B(\alpha,\beta) \int P_m(x_p'-x', y_p'-y', \alpha, \beta) P_d(x-x', y-y', x', y')\,dx'\,dy'\}$$

$$k''_{lE} = V\{\Im_m(E)\ \Im_d(E)\ R(E-E', E)\}$$

and the Kronecker delta δ_{kt} has been used for the temporal response since this is only limited by the detector and telemetry 'dead time' which is very small and only seriously affects very high count rates.

Summations over slots $i_1 < i < i_2, j_1 < j < j_2$, $k_1 < k < k_2$ and $l_1 < l < l_2$ give the contracted image forms:

$$J^{kl}_{ij} = \sum_{kl} J_{ijkl} = \sum_{kl} \sum_{\alpha\beta tE} k'_{ij\alpha\beta}\, \delta_{kt}\, k''_{lE}\, f_{\alpha\beta tE} + \sum_{kl} N_{ijkl} \qquad (1.19)$$

$$J^{ijk}_{l} = \sum_{ijk} J_{ijkl} = \sum_{ijk} \sum_{\alpha\beta tE} k'_{ij\alpha\beta}\, \delta_{kt}\, k''_{lE}\, f_{\alpha\beta tE} + \sum_{ijk} N_{ijkl} \qquad (1.20)$$

$$J_k^{ijl} = \sum_{ijl} J_{ijkl} = \sum_{ijl} \sum_{\alpha\beta tE} k'_{ij\alpha\beta}\, \delta_{kt}\, k''_{lE}\, f_{\alpha\beta tE} + \sum_{ijl} N_{ijkl} \qquad (1.21)$$

where the range of the space, energy and time summations define the sampling slots corresponding to each contracted image. If these slots are large enough, the statistics of J_{ij}, J_l and J_k will be good. The source distribution can be contracted in three ways:

$$f_{\alpha\beta}^{kl} = \sum_{kl} \sum_{tE} \delta_{kt}\, k''_{lE}\, f_{\alpha\beta tE} \qquad (1.22)$$

$$f_E^{ijk} = \sum_{ijk} \sum_{\alpha\beta t} k'_{ij\alpha\beta}\, \delta_{kt}\, f_{\alpha\beta tE} \qquad (1.23)$$

$$f_t^{ijl} = \sum_{ijl} \sum_{\alpha\beta E} k'_{ij\alpha\beta}\, k''_{lE}\, f_{\alpha\beta tE} \qquad (1.24)$$

where the range of summations in ijkl define the spatial, temporal and energy slots in image co-ordinates spanned by f_k^{ijl}, f_E^{ijt} and $f_{\alpha\beta}^{kl}$. Because of the forms of the space, time and energy kernels, these slots will not represent independent slots in source distribution co-ordinates and there will be crosstalk between these slots due to the instrument response.

Equations (1.19), (1.20) and (1.21) can be rewritten as:

$$J_{ij}^{kl} = \sum_{\alpha\beta} k'_{ij\alpha\beta}\, f_{\alpha\beta}^{kl} + N_{ij}^{kl} \qquad (1.25)$$

$$J_l^{ijk} = \sum_{E} k''_{lE}\, f_E^{ijk} + N_l^{ijk} \qquad (1.26)$$

$$J_k^{ijl} = \sum_{t} \delta_{kt}\, f_t^{ijl} + N_k^{ijl} \qquad (1.27)$$

Equations (1.25), (1.26) and (1.27) are the foundation of the digital image processing. J^{kl}_{ij} is a matrix of binned counts selected using a specified time slot ($k_1 \prec k \prec k_2$) energy slot ($l_1 \prec l \prec l_2$ if available) and has exactly the same form as $J^{t'E'}_{i\ j}$ in figure 22. J^{ijk}_{l} is a vector of binned counts selected using a specified time ($k_1 \prec k \prec k_2$) and spatial ($i_1 \prec i \prec i_2$, $j_1 \prec j \prec j_2$) slot and J^{ijl}_{k} is a vector of binned counts selected using a specified spatial and energy slot ($l_1 \prec l \prec l_2$ if available).

1.6. The structure of the instrument kernels $k'_{ij\alpha\beta}$ and k''_{lE}

The kernels $k'_{ij\alpha\beta}$ and k''_{lE} have a structure dictated by the instrument response as given by equation (1.1). It is convenient to express $k'_{ij\alpha\beta}$ and equation (1.25) in 'stacked form' so that all three equations (1.25), (1.26) and (1.27) have the same matrix-vector form rather than the more awkward 4-tensor-matrix form. A matrix is 'stacked' into a column vector by placing successive columns of the matrix end to end, thereby producing a single long column vector. The sub-matrices of a 4-tensor can be 'stacked' into a matrix in an analogous fashion. The underlying structure of tensors is unaltered by the stacking operator although care must be exercised to avoid disruption of the index relationships in the final stacked equation. Reference 2 provides an excellent exposition of this operation. After application of the stacking operator, equation (1.25) becomes:

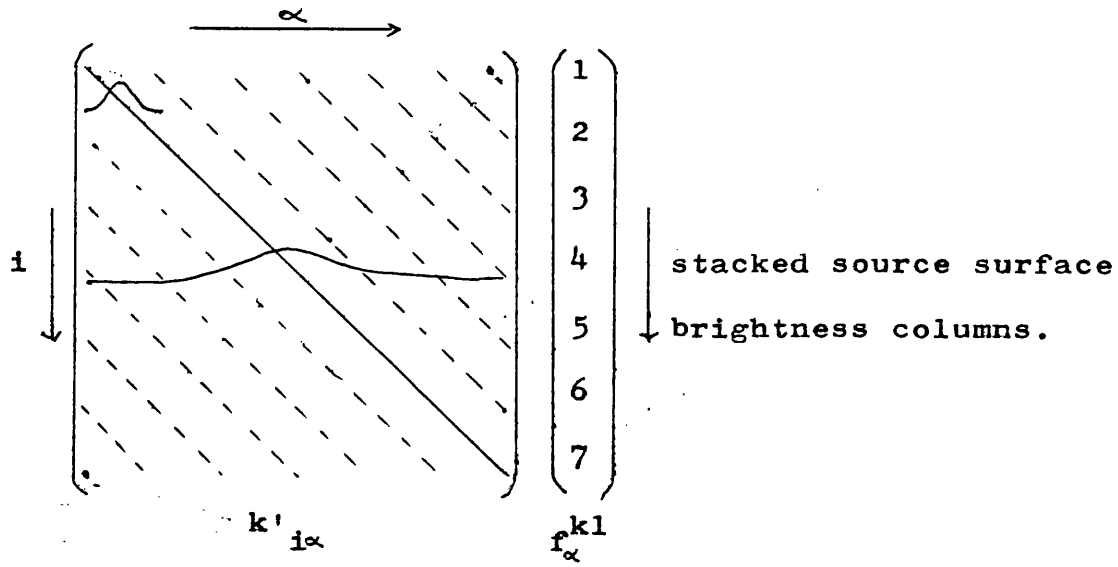
$$J_i^{kl} = \sum_{\alpha} k'_{i\alpha} f_{\alpha}^{kl} + N_i^{kl} \qquad (1.28)$$

where

$$\left.\begin{aligned} J_{ij}^{kl} &= J_{i+Nj}^{kl} \\ k'_{ij\alpha\beta} &= k'_{i+Nj,\alpha+N\beta} \\ f_{\alpha\beta}^{kl} &= f_{\alpha+N\beta}^{kl} \\ N_{ij}^{kl} &= N_{i+Nj}^{kl} \end{aligned}\right\} \quad \text{N is the number of rows in the image matrix } J_{ij}^{kl}$$

The ith element of the image column vector J_i^{kl} (energy and time contracted) is given by the multiplication of the source vector f_{α}^{kl} by the ith row of the kernel matrix $k'_{i\alpha}$. The summation involved represents the blurring or crosstalk between source elements in the final image. The result is then disrupted by the addition of the noise process N_i^{kl}.

Since the position of a point source in the image plane is related to its sky position by equation (1.4), the position of the non-zero elements of a row of the kernel $k'_{i\alpha}$ depends on the row index i. Increasing the index i shifts the centre of the response across the matrix giving a band structure. Because the kernel is in stacked form, each row must contain the crosstalk along a given column of the image. This introduces a side band structure to the kernel $k'_{i\alpha}$, each band representing the crosstalk between elements at a specified number of rows apart.



The 'profile' across a single band is determined by the blurring along a column of the source matrix and the profile across the bands is determined by the blurring along a row of the source matrix. If the instrument's spatial response is good, with no blurring, the side bands disappear and the primary band reduces to a diagonal of 1's. The kernel is then the unit matrix and the image distribution before noise disruption is a faithful reproduction of the original source surface brightness.

Since the degradation of a point source is position dependent, the band and inter-band profiles will depend on the position along the diagonal. The centre of the kernel matrix will be sparse corresponding to the good on-axis response while the degradation at the edges will be coarse, introducing many non-zero elements at the top and bottom.

The structure of k''_{1E}, the energy kernel, is very similar to the spatial counterpart described above. However since the stacking operation is not required, no side bands are present to give the correlation between columns. The kernel k''_{1E} has non-zero terms on and near the main diagonal and because the energy resolution of a

proportional counter degrades at low energies, the diagonal band will be wider at the low energy end.

The kernel matrix as presented above also contains an 'edge discontinuity' at the end of each row. This provides a natural edge to the image J_{ij}^{kl}, but leads to the loss of some of the source flux outside the boundaries of the image.

The overall band structure and edge effects present in the instrument kernels have important consequences in the digital analysis. The success of the digital processing methods used to produce the final image from the image matrix J_{ijkl} relies on the understanding of the properties of the response and noise matrices used in the digital analysis of the imaging system.

1.7. The processing problem.

The image J_{ijkl} must be used to provide an image or representation of the source brightness distribution $f(\alpha,\beta,t,E)$. The major difficulty is the processing problem of allowing for the blurring, noise and digitization degrations present in J_{ijkl} so that the final result is free from both systematic and random errors. Equation (1.28) holds the key to tackling this problem since it relates the image J_{ijkl} to the source distribution f_{α}^{kl}.

Unfortunately it is impossible to simply solve equation (1.28) because $k'_{i\alpha}$ and N_i^{kl} are not fully determined. Measurements of $k'_{i\alpha}$ are somewhat scanty and N_i^{kl} is a stochastic process characterised by a set of moments or averages rather than a fully determined number set. However, given the information that is available about the response and noise processes it is possible to

estimate $\hat{f}_{\alpha}^{kl}$ from J_i^{kl}. Some form of criterion is required to define a 'best' estimate of the source brightness $\hat{f}_{\alpha}^{kl}$ so that an 'optimum' solution to equation (1.28) can be found.

The form of equation (1.28):

$$J_i^{kl} = \sum_{\alpha} k'_{i\alpha} f_{\alpha}^{kl} + N_i^{kl} \qquad (1.28)$$

is not peculiar to the present situation and in fact occurs in many data processing fields. A solution is required from a set of linear equations (in this case expressed in matrix notation) without a complete knowledge of all the coefficients. The success of methods used to solve such underdetermined problems depends on the choice of optimisation criteria and the structure of the kernel $k'_{i\alpha}$. Equation (1.28) is the discrete variable counterpart or approximation to an integral equation and the processing involved to find $\hat{f}_{\alpha}^{kl}$ amounts to solving the integral equation representing the action of the instrument using discrete analysis. It must be stressed that it is impossible to 'beat the system' and remove all degradations present in the measured data. However data processing can be an integral part of the experiment rather than a fancy piece of software and degradation can be so bad that without digital processing the true result of the experiment cannot be realised.

The solutions to the processing problem fall into two categories; direct and indirect matrix inversion or diagonalisation. The feasibility of a given method depends on both finding an analytic solution and programming a computer to provide the answer. The analytic answer may only exist under certain conditions or approximations and

the computing power may not be available to enable the solution to be calculated. Consequently the number of viable methods is restricted.

CHAPTER 2: THE THEORETICAL APPROACH TO THE PROBLEM.

2.1 Matrix inversion and diagonalization.

A set of linear equations expressed in matrix notation can be solved by finding the inverse matrix.

If:

$$Y_i = \sum_I M_{ij} X_j \qquad (2.1)$$

then: (indices will be dropped for clarity)

$$[M]^{-1} Y = [M]^{-1}[M] X = [I] X \qquad (2.2)$$

where $[M]^{-1}[M] = [I]$, the unit matrix. If the set of equations is independent and complete (fully determined), then M^{-1} will exist and the solution can be found by matrix inversion. However, direct calculation of the inverse matrix will require $O(N^2)$ operators when N variables are involved. Another approach is matrix diagonalization. If a similarity transform exists which diagonalizes the matrix then:

$$[M] = [W][\Lambda][W]^{-1} \qquad (2.3)$$

where Λ is a diagonal matrix. Substituting in equation (2.3) gives:

$$Y = [W][\Lambda][W]^{-1} X \qquad (2.4)$$

Multiplying on the left by W^{-1} gives:

$$[W]^{-1} Y = [W]^{-1}[W][\Lambda][W]^{-1} X \qquad (2.5)$$

Putting $[W]^{-1} Y = \bar{Y}$, $[W]^{-1} X = \bar{X}$ and $[W]^{-1} [W] = [I]$ gives:

$$\bar{Y} = [\Lambda]\,\bar{X} \tag{2.6}$$

Since $[\Lambda]$ is diagonal, the matrix multiplication of equation (2.1) has been reduced to a direct product in the transform domain and $\bar{X}$ can be calculated by element to element division of $\bar{Y}$ by the diagonal elements of Λ. The form of the transform W is easily derived by considering the eigenvector equation:

$$M_{ij}\, Z_j^n = \Lambda^n\, Z_j^n \tag{2.7}$$

where Z_j^n is the nth eigenvector and Λ^n is the corresponding eigenvalue. Substituting for $[M]$ from equation (2.3) gives:

$$[W][\Lambda][W]^{-1}\, Z^n = \lambda^n\, Z^n \tag{2.8}$$

If W is constructed by packing the column vectors Z^n into a matrix:

$$[W] = [Z^1\ Z^2\ Z^3\ \ldots\ Z^n\ \ldots.] \tag{2.9}$$

and $[\Lambda]$ is diagonal with elements λ^n:

$$\Lambda_{ii} = \lambda^i \tag{2.10}$$

then equation (2.8) is satisfied since:

$$[W][\Lambda][W]^{-1}\,[W] = [\Lambda][W] \tag{2.11}$$

Providing the transformation $[W]$ satisfying equation (2.3) can be found, the matrix transform is reduced to a direct product and the inverse operation is easy. However the transforms $\bar{X}$ and $\bar{Y}$ must be calculated which will also require $O(N^2)$ when N variables are involved. Real imaging systems are normally illconditioned because there are zero

elements in the diagonal of the matrix [Λ]. The matrix [M] is then said to be singular since these zero elements are singularities in the system response and the system cannot provide any information concerning the eigenvectors corresponding to the zero eigenvalues. The inverse matrix M^{-1} does not exist for such a system and progress can only be made by using a pseudo-inverse which avoids the singularities. Direct inversion of [M] is impractical because of the large amount of computation involved and in most instances is impossible because of singularities in M. Diagonalization provides a method for calculating pseudo-inverses but the transformation W must be found and the computing power must be available to cope with the calculations required.

The validity of any inversion scheme hinges on equation (2.3), the diagonalization of the matrix [M] using the transformation W. Rather than considering arbitrary forms of [M], the available forms of W must be used and the corresponding forms of [M] adapted to suit real imaging systems.

The instrument kernels that occur in practice approximate to one of the following distinct types.

1. Separable space invariant point spread function:

$$P(x,y,x'',y'') = a(x - x'')\, b(y - y'')$$

The matrix representation of such a system involves two Toeplitz matrices [A] and [B]. Toeplitz matrices have the property:

$$T_{jk} = T_{mn} \qquad \text{if } j - k = m - n \qquad (2.12)$$

This property has already been described as the banded

structure characteristic of instrument kernels and equation (2.12) introduces the added restriction that each row (or column) is identical to its neighbours apart from a linear shift. This ensures that the system is space invariant (also called linear or isoplanatic). The system is represented by the equation:

$$[G] = [A][F][B] \tag{2.13}$$

where [F] is the source matrix and [G] is the blurred image.

2. Separable space variant point spread function.

$$P(x,y,x'',y'') = a(x,x'')\ b(y,y'')$$

The system can again be represented by two blur matrices but because of the non-linearity the Toeplitz structure is lost. Equation (2.13) represents the discrete system.

3. Non-separable space invariant point spread function.

$$P(x,y,x'',y'') = k(x - x'',y - y'')$$

Since the kernel is non-separable, the stacked form must be used as described above:

$$G = [K]F \tag{2.14}$$

The kernel matrix [K] has the banded structure previously described and because the system is shift invariant, the matrix is <u>block Toeplitz</u>.

$$\text{stacked columns} \downarrow \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ \cdot \\ \cdot \\ \cdot \\ g_{j+m-1} \end{bmatrix} = \begin{bmatrix} [K_0] & & & \\ [K_1] & [K_0] & & \\ \cdot & & [0] & \\ \cdot & & & \\ [K_{j-1}] & \cdot & \cdot & [K_0] \\ & & & & \cdot \\ & & & & & [K_0] \\ & & [0] & & & [K_{j-1}] \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_{m-1} \end{bmatrix} \qquad (2.15)$$

where the submatrices $[K_i]$ are found from the sampled point spread function:

$$[K_i] = \begin{bmatrix} P_{i0} & & & \\ P_{i1} & P_{i0} & 0 & \\ \cdot & & & \\ \cdot & & & \\ P_{i,j-1} & \cdot & \cdot & \cdot \; P_{i0} \\ & & & \\ 0 & & P_{i,j-1} & P_{i0} \end{bmatrix} \qquad (2.16)$$

Each submatrix $[K_i]$ is of Toeplitz form and they are arranged in Toeplitz form to give the full block Toeplitz matrix $[K]$.

4. Non-separable space variant point spread function.

$$P(x,y,x",y") = k(x,y,x",y")$$

Again, the stacked form must be used giving equation (2.14) as in case 3. However the form of k, the blur matrix or kernel, is arbitrary although it still possesses the basic band structure.

Adequate direct inversion methods are at present only available for blur matrices with a Toeplitz structure

and direct methods are therefore restricted to dealing with the space invariant or linear systems. Indirect methods can be applied to the more general banded matrices but they are restricted to problems of small dimensionality because of the vast amount of computation required.

Toeplitz matrices are very closely related to circulant matrices. A ciculant matrix has the property that each row is a circular right shift of the row above it. A Toeplitz matrix can be converted into a circulant by changing zero elements into the appropriate circulant elements. In fact, as the size of the Toeplitz matrix gets very large compared to the width of the non-zero bands that construct it, the Toeplitz form approaches that of the circulant and the differences only affect the ends or edges of the transformed vector or matrix (equations (2.13) and (2.16)). The elements which have to be added to the Toeplitz matrix correspond to the flux or data which is lost off the edge of the recorded image. In fact the complete data set required to invert the blurring process is not recorded and the inverse to the Toeplitz matrix does not exist. However information is only normally lost at the edges and using the circulant form which assumes no loss of data allows us to proceed.

Circulant matrices and block circulant matrices are diagonalized by the discrete Fourier Transform which, fortunately, has a fast transform algorithm. Circulants are therfore easily diagonalized as described above using the FFT algorithm and the solution to the circular shift-invariant system can be found quickly, even when a very large number of discrete variables are present. In order

to invert image blur, the effect of the circulant approximation to the Toeplitz structure of the blur matrix must be allowed for and the resulting edge errors reduced to a minimum.

More practical details about diagonalization of blur matrices using a digital computer will be dealt with elsewhere. Now the direct inversion methods involving the Discrete Fourier Transform will be described.

2.2 Statistical model for the source and noise processes.

Apart from singularities in the system response, the major difficulty involved in the processing problem is the noise vector $N_i^{kl} = N$. The complete image formation can be viewed as a stochastic process which can be characterised by an underlying probability density function. Moments or expectation values are used to represent the processes involved and in discrete analysis, the two most important are the covariance matrix $[\phi_f]$ and the mean value vector U_f. They are defined as follows:

$$U_f = E\{f\} \tag{2.17}$$

$$[\phi_f] = E\{(f - U_f)(f - U_f)^t\} \tag{2.18}$$

where $E\{\}$ is the expectation value over the ensemble governed by a particular probability density function and f is a stacked matrix. U_f describes the mean of the vector f as a function of position, hence if $f \equiv N$, U_N would be the mean noise as a function of position in the image. If f represents the source field, then U_f will represent the underlying structure in the source. $[\phi_f]$ describes the correlation between picture elements (pixels) as

a function of position and inter-pixel distance. In general, $[\phi_f]$ characterises an image type whereas U_f conveys structure specific to a particular image.

Processes f characterised by U_f and $[\phi_f]$ fall into two distict classes, stationary and non-stationary. A two dimensional stationary model has the properties that:

$$E\{f\} = U_f = U_c \text{ (a constant vector)} \tag{2.19}$$

and $[\phi_f]$ is partitioned by the stacking of columns into matrices:

$$\left.\begin{aligned} &[\phi_{mn}] = E\{(f_m - U_m)(f_m - U_m)^t\} \\ &\text{for which:} \\ &[\phi_{pq}] = [\phi_{rs}] \quad \text{when } p-q = r-s \\ &\text{and;} \\ &[\phi_{pq}]_{jk} = [\phi_{pq}]_{mn} \quad \text{when } j-k = m-n \end{aligned}\right\} \tag{2.20}$$

Equations (2.20) define a Block Toeplitz matrix and the one dimensional counterpart would be the simple Toeplitz matrix. If ϕ_f is not a Toeplitz form then the model is non-stationary. If the mean is known but not constant, it can be subtracted and the model can still be stationary. A stationary image model cannot be expected to characterise more than the gross or 'average' properties of a class of images whereas non-stationary models are far more specific and can accommodate the peculiarities of particular images.

Closely related to the covariance matrix is the correlation matrix $[R_f]$:

$$[R_f] = E\{ff^t\} \tag{2.21}$$

If f is a stacked matrix, $[R_f]$ will be partitioned in exactly the same way as previously described (equation (2.15)). Since $R_{ij} = R_{ji}$, R_f is symmetric. The blocks $[R_{ij}]$ along the diagonal represent the correlation of each column with itself while off-diagonal blocks describe inter-row correlation with separation $|i - j|$. $[R_f]$ and $[\phi_f]$ are simply related by the equation:

$$[R_f] = [\phi_f] + [U_f \; U_f^t] \qquad (2.22)$$

Typically, images possess no correlation for distances greater than 20-30 pixels and $[\phi_f]$ will have a banded structure very similar to the instrument kernels derived above. Furthermore if inter-pixel correlation is dependant only on the difference $|i - j|$, the covariance matrix will have the Toeplitz structure characteristic of a stationary process.

The reduction of $[\phi_f]$ to a Toeplitz form enables the FFT to be exploited in computation by using the circulant approximation to $[\phi_f]$. However this is not the same as using a circulant to approximate $[R_f]$ because of the effect of the mean U_f in equation (2.22). If $U_f = 0$, then $[R_f] = [\phi_f]$ and no problem arises. However this is physically impossible since $U_f \neq 0$ for images. If $U_f = U_c$, a constant vector, then the statistical model gives random fluctuations about a constant mean and the constant cannot be affected (other than a scaling) by the action of the blur matrix $[K]$. Therefore having $[R_f] = [\phi_f]$ and neglecting the mean has no effect on the final result; the necessary mean shift being provided by the data. Finally, however, if $U_f \neq U_c$, the model is non-stationary and if

$[\phi_f]$ is Toeplitz, fluctuations occur about a non-uniform distribution. This mean will be affeeted by $[K]$ and in this case $[R_f]$, including the mean (equation (2.22)), must be used. $[U_f\ U_f^t]$ will not be Toeplitz and the FFT cannot be utilized to alleviate the computation burden.

2.3 Discrete filtering.

The matrix inversion technique introduced in section 2.1 are sometimes referred to as discrete filtering. Equation (2.1) is in general solved by multiplication on the left by a matrix $[Q]$:

$$[Q]\,Y = [Q][M]X \tag{2.23}$$

$[Q]$ is known as a discrete filter and is in general related to the matrix M. If the system is noiseless and non-singular, then $[Q] = [M]^{-1}$ and $[Q]$ is known as the inverse filter. However in practice, use of the inverse filter results in gross distortion due to noise and singularities and the filter $[Q]$ must be chosen to reduce these errors.

The filtering can be reduced to a direct product by using a diagonalising transformation as described above and for space invariant systems, discrete filtering can be performed in the Discrete Fourier Transform Domain. Equation (2.6) would then refer to a direct product operation in the Discrete Fourier Transform Domain and the filter $[\bar{Q}]$ expressed in that domain would be closely related to $[\Lambda]$.

It must be remembered that the processing problem of finding an estimate $\hat{f}_{\sim}^{kl}$ from J_i^{kl} by equation (1.27) includes the noise vector N_i^{kl}. The construction of $[\bar{Q}]$

must allow for the singularities in $k'_{i\alpha}$ (which are recognised as zeros in the diagonal matrix $[\Lambda]$) and the corruption caused by the noise vector. In short, components of the transform of $x(\bar{x}_i)$ corresponding to small or zero components of the instrument response Λ_{ii} will be vulnerable to noise and excessive amplification of noise at these spatial frequencies must be avoided. The suppression of these frequencies will limit the final performance of the system according to the deficiencies in the system point response and the noise level.

The consequences of structure in the noise covariance matrix and the response kernel are very similar. The correlation between noise elements is analogous to the spreading of point sources, both giving rise to an overall shape or function in the DFD. Correlation of the noise amplitude is analogous to a space variant point response which introduces off-diagonal elements in the DFD description of the system (equation (2.6)). Filtering is no longer a direct product operation in the DFD for such non-isoplanatic systems and exact diagonalisation is not possible. Effective use of direct product DFD filters is therefore restricted to systems with space invariant noise and blurring processes.

The detailed structure of $[\bar{Q}]$ will depend on the criteria used to define the 'optimum' result, however most filters have the general form:

$$\bar{Q}_{pq} = \frac{\bar{k}^*_{pq}}{\bar{k}_{pq}\,\bar{M}_{pq} + Z_{pq}} \qquad (2.24)$$

where p and q index in the two dimensional DFD, $\bar{M}_{pq}$ is the

DFT of the blur matrix M_{ij} (equation (2.1)), $\bar{k}_{pq}$ is related to $\bar{M}_{pq}$ and is often the complex conjugate $\bar{M}^*_{pq}$ and Z_{pq} is a function of the signal and noise covariance matrices. If the noise is negligable, Z_{pq} can be set to zero and if $\bar{k}_{pq} = \bar{M}^*_{pq}$, $\bar{Q}_{pq}$ reduces to $1/\bar{M}_{pq}$ which is the inverse filter. When the noise process dominates, Z_{pq} is designed to act as a moderator preventing the blowing up of pure noise components when $\bar{M}_{pq}$ is small.

2.4 The Wiener filter and Fourier filtering.

Methods for constructing discrete filters from knowledge of the blur matrix and noise covariance matrix appropriate to a system vary in effectiveness and sophistication. One approach, to be described here, uses the Principle of Least Squares to arrive at the final general form given by equation (2.24) and results in the Wiener filter, named after N. Wiener (1949) who first applied the principle to the filtering of time series (reference 3). The application of the technique to image processing was introduced by C. W. Helstrom (1967), see reference 4, and has been developed by many other workers in recent years.

Equation (1.28) can be rewritten as:

$$J = [k]f + N \qquad (2.25)$$

An estimate of the original source distribution $\hat{f}$ is required which minimises the mean square error for both blurring and noise. The error estimate is defined as:

$$\mathcal{E} = f - \hat{f} \qquad (2.26)$$

and the positive quantity $\varepsilon^t\varepsilon$ provides the mean square error to be minimised:

$$\text{minimum} \quad E\{\varepsilon^t\varepsilon\} = E\{Tr(\varepsilon\ \varepsilon^t)\} \tag{2.27}$$

where $E\{\varepsilon^t\varepsilon\}$ is the expectation value of the product $\varepsilon^t\varepsilon$. An operator $[Q]$ is required such that:

$$\hat{f} = [Q]\ J \tag{2.28}$$

which yields $E\{\varepsilon^t\varepsilon\}$ where $[Q]$ is chosen to be linear for practical reasons. Substituting for ε in equations (2.27) and (2.28) and using the correlation matrices

$$E\{f\ f^t\} = [R_f]$$
$$E\{N\ N^t\} = [R_N]$$
$$E\{f\ N^t\} = E\{N^t\ f\} = [0] \quad \text{signal independent noise}$$

gives:

$$E\{Tr(\varepsilon\ \varepsilon^t)\} = Tr\{[R_f] - 2[Q][k][R_f] + [Q][k][R_f][k]^t[Q]^t + [Q][R_N][k]^t\} \tag{2.29}$$

Differentiating equation (2.29) with respect to $[Q]$ and setting the derivative equal to zero yields the mean-square-error solution:

$$[Q] = [R_f][k]^t\ ([k][R_f][k]^t + [R_N])^{-1} \tag{2.30}$$

The information or model required to construct $[Q]$ from equation (2.30) is in the form of the noise and signal correlation matrices $[R_N]$ and $[R_f]$, hence the assumptions are purely statistical in nature. More sophisticated procedures known as constrained least squares estimations can incorporate further information about the image

processes involved (for example symmetry constraints).

The filter $[Q]$ equation (2.30) can only be successfully implemented in image processing applications if a quick method is available for matrix inversion. The only method available is diagonalisation using the DFT but this is only valid if $[k]$, $[R_f]$ and $[R_N]$ can be approximated using circulants as previously explained. Equation (2.30) can be rewritten as:

$$[Q] = [\phi_f][k]^t ([k][\phi_f][k]^t + [\phi_N])^{-1} \qquad (2.31)$$

providing $U_N = 0$, which is usually reasonable, and neglecting the mean source flux since it is unaffected by the point response. If $[\phi_f]$ $[\phi_N]$ and $[k]$ can be approximated by the circulants $[C_f]$, $[C_N]$ and $[C_k]$, equation (2.29) becomes:

$$[Q] \simeq [C_f][C_N]^t ([C_k][C_f][C_k]^t + [C_N])^{-1} = [Q_c] \qquad (2.32)$$

where $[Q_c]$ is the circulant approximation of the minimum mean square error filter. Since circulants are diagonalised by the DFT with transform matrix $[W]$, $[Q_c]$ can be written as:

$$[Q_c] = [W][\Lambda_f][\Lambda_k^*] ([\Lambda_k \Lambda_f \Lambda_k^*] + [\Lambda_N])^{-1} [W]^{-1} \qquad (2.33)$$

where * denotes the complex conjugate and $[\Lambda]$'s are the diagonal eigenvalue matrices. Substituting this form of the filter into equation (2.28) gives:

$$[W]^{-1}\hat{f} = [\Lambda_f][\Lambda_k^*]([\Lambda_k][\Lambda_f][\Lambda_k^*] + [\Lambda_N])^{-1} [W]^{-1} J \qquad (2.34)$$

Equation (2.34) is a relation between diagonal matrices, If $[W]^{-1}\hat{f} = \bar{\hat{f}}$ and $[W]^{-1} J = \bar{J}$, this equation can be written as a direct operation involving the individual discrete Fourier samples of the blur matrix and covariance matrices of the source and noise processes:

$$\bar{\hat{f}}_{pq} = \frac{\bar{k}^*_{pq} \bar{J}_{pq}}{|\bar{k}_{pq}|^2 + [P_{n\,pq}/P_{f\,pq}]} \tag{2.35}$$

where, by definition, the DFT of the covariance matrices $[\phi_f]$ and $[\phi_N]$ are the power spectra $[P_f]$ and $[P_N]$. The explicit form of $[W]$ and the DFT will be considered when describing the implementation of Discrete Filtering using a digital computer.

The major drawback of the Wiener filter is that it is based on linear assumptions about the image processes. Most imaging systems are non-linear with space variant noise and minimum mean square estimations tend to be too smooth, especially when the signal to noise ratio is small. The stationary assumption using only the covariance information and not the mean signal and noise information tends to limit the performance of the Wiener filter in real situations where the variation of the mean is of prime importance.

Although Fourier Domain Filtering is essentially linear, it is possible to adapt the above model to treat signal dependent and independent multiplicative noise. The resulting filters have the same general form as equation (2.24) and according to reference 5 show no marked improvement over the Wiener result. Reference 5 also provides an excellent and thorough discussion on all

aspects of linear filtering and formed the basis of the above exposition.

2.5 Algebraic methods.

Discrete filtering can be thought of as a direct method for image restoration, relying heavily on matrix inversion and diagonalisation to construct the 'optimum' filter as well as to speed up the computation. It is essentially a linear technique centred around the DFT although some refinements can be made to handle non-stationary noise and signal processes. The stationary assumption which assigns global properties to the image model rather than image specific properties, which are required for good restoration, is felt to be a major limitation of discrete filtering.

The algebraic or indirect approach can provide a more flexible basis for solving the processing problem, but it is not without its difficulties. Most algebraic methods rely on iteration to arrive at the optimum solution $\hat{f}$ to equation (2.25). (This can be costly on central processing unit time, especially for non-linear algorithms which do not converge quickly.) Successive guesses at the required result are made $\hat{f}^i$ and an associated guess at the data set $\hat{J}^i$ is calculated by using the equation:

$$\hat{J}^i = [k] \quad \hat{f}^i \qquad (2.36)$$

$\hat{J}^i$ is then compared to the actual image data vector J and the next guess $\hat{f}^{i+1}$ is made using the comparison as a guide. Such a scheme requires a strict criterion to define the optimum $\hat{f}$ and a carefully chosen method for updating

$\hat{f}^i$ to give $\hat{f}^{i+1}$, giving fast convergence to that optimum.

It is not a prerequisite of algorithm techniques that they involve iteration but simply that the specified solution cannot be found by direct computation. Only one approach, the maximum entropy method, will be discussed in detail despite the many varied approaches that exist (reference 5) because this formulation has the ingredients necessary for successful algebraic restoration. The label 'maximum entropy method' is perhaps misleading because the term 'entropy' is borrowed from other fields. 'MEM' is an apt title but is inadequate in a similar way to the expression 'Fourier filtering' because it shrouds the philosophy behind the technique. While Fourier filtering is concerned directly with the mechanics of the blurring process and treats noise in a global fashion avoiding the ill-conditioned nature of the imaging process, MEM brings the statistics of image formation to the fore.

2.6 The quantum statistics of image formation.

Rather than using covariance matrices to categorise the noise processes involved (see section 2.2), the maximum entropy method considers the quantum statistics of the image formation. Although the basis of the MEM has been in existence for some time (B.R. Frieden (1972) in reference 6 seems to have originated the method, but independent attempts using slightly different approaches were made subsequently e.g. J. G. Ables (1974) reference 7) it has only recently enjoyed active research because of the ever increasing capability of modern digital computers. A paper by R. Kikuchi and B. H. Soffer (1977) reference 8

has also provided a much needed foundation for the method which previous publications did not give in rigorous detail and the following discussion uses reference 8 to a large extent.

The object distribution $f(\alpha, \beta, t, E)$ is represented in discrete form by $F^{kl}_{\alpha\beta}$, a matrix of elements (α, β) giving the flux from each pixel of area $\Delta\alpha\Delta\beta$ on the sky. The number of photons observed from each pixel will be given by:

$$n^{kl}_{\alpha\beta} = F^{kl}_{\alpha\beta}\, tA \qquad (2.37)$$

where t is the observing time and A is an aperture area for the instrument. In an ideal instrument, all photons observed from a given sky cell would be positioned correctly by the instrument without blurring, but as previously described this is unfortunately not the case in practice and the process is in general described by equation (1.28). The processing problem is to find the most probable distribution matrix $[n]$. To simplify the analysis, the bandwidth will be assumed to be small and the object monochromatic. This is not essential but merely aids the discussion of coherence volume and degrees of freedom for photons which is to follow.

Each pixel at (α, β) of area $\Delta\alpha\Delta\beta$ is considered to have $z_{\alpha\beta}$ degrees of freedom or quantum states which is proportional to $\Delta\alpha\Delta\beta$, the bandwidth $\Delta\gamma$, collecting area or aperture A and the observing time t. The $n^{kl}_{\alpha\beta}$ photons from the pixel (α, β) can be distributed in these $z_{\alpha\beta}$ degrees of freedom using Bose-Einsteim statistics (since photons are Bose particles) and each degree of freedom can

have multiple occupancy. The number of macroscopically indistinguishable ways $q_{\alpha\beta}$ that any arrangement within a pixel can be formed is given by the familiar Bose-Einstein formula:

$$q_{\alpha\beta}(n_{\alpha\beta}) = \frac{(n_{\alpha\beta} + z_{\alpha\beta} - 1)!}{n_{\alpha\beta}!\ (z_{\alpha\beta} - 1)!} \tag{2.38}$$

The factor $n_{\alpha\beta}!$ occurs in the denominator because the photons are supposed to be indistinguishable and the factor $(z_{\alpha\beta} - 1)!$ occurs because the quantum states or degrees of freedom are also assumed indistinguishable. Formula (2.38) assumes a single frequency ν and more properly a product should be taken over all frequencies ν using a different z for each ν.

Each of the $q_{\alpha\beta}$ arrangements is a quantum mechanical state of the $n_{\alpha\beta}^{kl}$ photons in the object cell (α,β) and it can be postulated that each of the $q_{\alpha\beta}$ different (that is microscopically distinguishable but macroscopically indistinguishable) states occurs with equal a priori probability. $q_{\alpha\beta}$ can then be interpreted as a weighting factor, degeneracy or probability for the intensity $n_{\alpha\beta}^{kl}$. It is mathematically convenient to take the logarithm when dealing with factorials, introducing the function $s_{\alpha\beta}(n_{\alpha\beta})$:

$$s_{\alpha\beta}(n_{\alpha\beta}) = \ln\ q_{\alpha\beta}(n_{\alpha\beta}) \tag{2.39}$$

Expanding equation (2.39) gives:

$$s_{\alpha\beta}(n_{\alpha\beta}) = \ln\ (n_{\alpha\beta}+z_{\alpha\beta}-1)! - \ln\ (n_{\alpha\beta}!) - \ln\ (z_{\alpha\beta}-1)! \tag{2.40}$$

Stirling's approximation for large x:

$$\ln x! \simeq x \ln x - x \qquad (2.41)$$

can be used to simplify equation (2.39):

$$\begin{aligned} s_{\alpha\beta}(n_{\alpha\beta}) &- (n_{\alpha\beta}+z_{\alpha\beta}-1) \ln (n_{\alpha\beta}+z_{\alpha\beta}-1) - (n_{\alpha\beta}+z_{\alpha\beta}-1) \\ &- n_{\alpha\beta} \ln n_{\alpha\beta} + n_{\alpha\beta} - (z-1) \ln (z_{\alpha\beta}-1) \\ &+ (z_{\alpha\beta}-1) \\ = (n_{\alpha\beta}&+z_{\alpha\beta}-1) \ln (n_{\alpha\beta}+z_{\alpha\beta}-1) - n_{\alpha\beta} \ln n_{\alpha\beta} \\ &- (z_{\alpha\beta}-1) \ln (z_{\alpha\beta}-1) \qquad (2.42) \end{aligned}$$

Three limiting cases of equation (2.42) are of interest;

a) $z_{\alpha\beta} = 1 \rightarrow q_{\alpha\beta} = 1 \rightarrow s_{\alpha\beta} = 0$. This is reasonable since all photons must be occupying the same state and they are indistinguishable, so the degeneracy is unity.

b) When $1 < z_{\alpha\beta} \ll n_{\alpha\beta} \rightarrow z_{\alpha\beta}/n_{\alpha\beta}$ is small and can be neglected in (2.42) to give:

$$s_{\alpha\beta}(n_{\alpha\beta}) \simeq (z_{\alpha\beta}-1) \ln n_{\alpha\beta} - (z_{\alpha\beta}-1) \ln (z_{\alpha\beta}-1) \qquad (2.43)$$

c) If $z_{\alpha\beta} \gg n_{\alpha\beta}$ then $n_{\alpha\beta}/z_{\alpha\beta}$ can be neglected in equation (2.42) giving:

$$s_{\alpha\beta}(n_{\alpha\beta}) \simeq n_{\alpha\beta} \ln z_{\alpha\beta} - n_{\alpha\beta} (\ln n_{\alpha\beta} - 1) \qquad (2.44)$$

Expression (2.44) is the classical (Maxwell-Boltzmann) result for $n_{\alpha\beta}$ distinguishable particles as expected for bosons in the classical limit $z_{\alpha\beta} \gg n_{\alpha\beta}$.

So far, only a single pixel (α,β) has been considered but the complete set making up the image can now be handled. What is the total number of ways that $n_{\alpha\beta}$ comes from (α,β), where (α,β) takes all values present in the image? Assuming each pixel to be independent of all other

pixels, the complete number of ways that a particular distribution can occur will be given by the product of all the $q_{\alpha\beta}$'s:

$$Q(n_{11}, n_{12} \ldots) = \prod_{\alpha\beta} q_{\alpha\beta}(n_{\alpha\beta}) \qquad (2.45)$$

Again, the logarithm corresponding to three cases can be considered.

a) $z_{\alpha\beta} = 1 \rightarrow s_{\{\alpha\beta\}} = 0$ as before, and hence the number of arrangements is independent of $\{n_{\alpha\beta}\}$ and thus all are equally likely.

b) $1 < z \ll n_{\alpha\beta} \rightarrow s_{\{\alpha\beta\}} = \sum s_{\alpha\beta}$

$$s_{\{\alpha\beta\}} \simeq \sum_{\alpha\beta} [(z_{\alpha\beta}-1) \ln n_{\alpha\beta} - (z_{\alpha\beta}-1) \ln (z_{\alpha\beta}-1)] \qquad (2.46)$$

If $z_{\alpha\beta}$ is independent of (α,β) then this simplifies to give:

$$s_{\{\alpha\beta\}} \simeq (z-1) \sum_{\alpha\beta} [\ln n_{\alpha\beta} - \ln (z-1)] \qquad (2.47)$$

c) $z_{\alpha\beta} \gg n_{\alpha\beta}$ gives:

$$s_{\{\alpha\beta\}} \simeq \sum_{\alpha\beta} [n_{\alpha\beta} \ln z_{\alpha\beta} - n_{\alpha\beta} (\ln n_{\alpha\beta}-1)] \qquad (2.48)$$

and again if $z_{\alpha\beta} = z$ this reduces to:

$$s_{\{\alpha\beta\}} \simeq \ln z \sum_{\alpha\beta} n_{\alpha\beta} - \sum_{\alpha\beta} n_{\alpha\beta} (\ln n_{\alpha\beta}-1) \qquad (2.49)$$

The logarithm of degeneracy is normally called the entropy and therefore equations (2.47) and (2.49) above define $s_{\{\alpha\beta\}}$, the configurational entropy (reference 9) of the image $[n]$. The above entropy expressions have been derived using a well defined model for the image process evoking Bose-Einstein statistics, degrees of freedom or quantum states and assuming equal a priori probability for

the arrangements of bosonsin the quantum states. The two extremes of many bosons per degree of freedom b) and many degrees of freedom per boson c) lead to different entropy expressions; b) containing a $\sum_{\alpha\beta} \ln n_{\alpha\beta}$ term and c) containing a $-\sum_{\alpha\beta} n_{\alpha\beta} \ln n_{\alpha\beta}$ term. In order to determine when each of these cases is applicable, the concept of degree of freedom must be studied in greater depth. The apparant anomaly of case a) when z = 1 giving $s_{\{\alpha\beta\}} = 0$, independent of the distribution $[n]$ can be resolved by considering the image to be characterised by average counts per pixel $\bar{n}_{\alpha\beta}$. The statistics of the image $[\bar{n}]$ can be analysed using an ensemble of images. This analysis will be given later because it completes the present model and intimately connects the configurational entropy $s_{\{\alpha\beta\}}$ with the more familiar statistical description of the image process.

From the above it can be seen that the ratio n/z dictates the form of the configuration entropy expression. One degree of freedom is approximated by the coherence volume for a photon since it is impossible to distinguish photons by interference experiments within one degree of freedom or coherence volume. In the direction of propagation, the coherence length is given by:

$$l = c\tau \qquad (2.50)$$

where $\tau = 1/\Delta\nu$, the coherence time. In the transverse direction, the coherence area σ depends on the distance from the source R and the area of the source $\Delta\alpha\Delta\beta$ (pixel area on sky):

$$\sigma \simeq R^2 \lambda^2/\Delta\alpha\Delta\beta \qquad (2.51)$$

This expression is given by the van Cittert-Zernike Theorem and is discussed in reference 10, pages 13-15. The coherence volume is taken as the product of 1 and σ:

$$V_{coh.} = \frac{c^3 R^2}{\nu^2 \Delta\nu \Delta\alpha \Delta\beta} \qquad (2.52)$$

The number of degrees of freedom z can now be defined in terms of $V_{coh.}$ and the observing instrument parameters. If the observing time is t, then the photons will be detected in t/τ coherence lengths of degrees of freedom. Using a collecting area or aperture A will give a further introduction A/σ to the number of degrees of freedom. The total z will correspond to the number of times $V_{coh.}$ is contained in the 'detection volume' ctA:

$$z = \frac{ctA}{c\tau\sigma} = \left(\frac{t}{\tau}\right)\left(\frac{A}{\sigma}\right) \qquad (2.53)$$

Kikuchi and Soffer distinguish between the spatial component (A/σ) and the temporal component (t/τ) and point out that the description of degrees of freedom can also be made in phase space using the conjugate variables time/ frequency, length/ spatial frequency.

In all the above discussion, the photon distribution from the source is assumed to be 'closely mirrored' by the photoelectron distribution which gives rise to the actual detected distribution. Detailed analysis of detection processes throughout the electromagnetic spectrum will reveal deficiencies in this assumption but they are of little consequence since it is an estimate of the ratio n/z which is of importance in deciding which entropy

expression is appropriate to a given observation.

Given the brightness of a source in $B(\nu)$ (watts m^{-2} steradians^{-1} Hz^{-1}) the ratio n/z is given by:

$$n/z = B(\nu)\ c^2/2h\nu^3 \qquad (2.54)$$

Using equation (2.54), the following values arise:
Cygnus A (3C403) at 960 MHz $n/z \simeq 10^6$
Centaurus A (CTA39) at 178 MHz $n/z \simeq 10^4$
All optical astronomical objects give $n/z \ll 1$
All X-ray astronomical objects give $n/z \ll\ll 1$
In short, for astronomical observations at frequencies less than infra-red $n/z > 1$ and at frequencies above infra-red $n/z < 1$, although the actual dividing line will depend on the specific object and waveband under consideration. Fortunately Frieden (reference 6) working in the optical region chose $-n_{\alpha\beta} \ln n_{\alpha\beta}$ for his entropy measure and Wernecke (1977, reference 11) working with radio observations chose $\ln n_{\alpha\beta}$ and both are therefore consistant with reference 8. The two cases b) and c) correspond to the classical wave limit and the classical particle limit to the radiation field under observation.

So far the analysis has considered the number of photons actually detected in each pixel (during a given observation) and the entropy expressions derived relate to the probability of the various arrangements of these observed photons within the image matrix. It is legitimate to ask what happens if the average number of photons $\bar{n}_{\alpha\beta}$ rather than a specific observed number $n_{\alpha\beta}$ is considered. The number of pixels within the ensemble that contain n photons is written as Mf_n. The function f_n has the form of

a probability density function and is normalised:

$$\sum_{n=0} f_n = 1 \tag{2.55}$$

The complete set of f_n's, $\{f_n\}$, then represents a particular state of the ensemble. In parallel to the previous analysis the number of ways, $P\{f_n\}$, an ensemble with a given $\{f_n\}$ can be constructed is the product of two distinct terms. The first arises from the number of ways a distribution $\{f_n\}$ can be constructed over the ensemble:

$$\Omega\{f_n\} = \frac{M!}{\prod_n (Mf_n)!} \tag{2.56}$$

The second term takes account of the degeneracy factor of the photons within the pixel:

$$Q\{f_n\} = \prod_n q(n)^{Mf_n} \tag{2.57}$$

$q(n)$ is the a priori weight or degeneracy factor for each pixel (see equation (2.38). The total number of arrangements $P\{f_n\}$ is given by the product:

$$P\{f_n\} = \frac{M!}{\prod_n (Mf_n)!} \prod_n q(n)^{Mf_n} \tag{2.58}$$

The most probable distribution $\{f_n{}^p\}$ when $n = \bar{n}$ will be the $\{f_n\}$ distribution that maximises expression (2.58). Again, it is convenient to deal with the logarithm $s\{f_n\}$:

$$s\{f_n\} = \ln \Omega\{f_n\} + \ln Q\{f_n\} \tag{2.59}$$

In order to find $\{f_n{}^p\}$, $s\{f_n\}$ must be maximised. Since both terms, the combinatorial and a priori contributions, must be maximised, expression (2.59) for $s\{f_n\}$ is

considered to be the entropy by Kikuchi and Soffer. Using Stirling's approximation in equation (2.41) gives (assuming that M is very large):

$$s\{f_n\} \simeq M \ln M - M - \sum_n (Mf_n \ln(Mf_n) - Mf_n) + \sum_n Mf_n \ln q(n)$$

$$\frac{s\{f_n\}}{M} = \ln M - 1 - \sum_n (f_n \ln f_n) + \sum_n f_n \ln q(n) - \sum_n (f_n \ln M - f_n)$$

$$= -\sum_n f_n \ln f_n + \sum_n f_n \ln q(n) \qquad (2.60)$$

in which the normalisation equation (2.57) has been used to simplify the expression for the entropy.

Expression (2.60) for the entropy of the ensemble must be maximised under constraints in (2.55) and the following:

$$\bar{n} = \sum_n n f_n \qquad (2.61)$$

to yield the $\{f_n^P\}$ distribution. Using Lagrangian multipliers λ and u gives:

$$\text{maximise} \quad -\sum_n f_n \ln f_n + \sum_n f_n \ln q(n) + u(\bar{n} - \sum_n n f_n) + \lambda(1 - \sum_n f_n) \qquad (2.62)$$

Differentiating with respect to f_n and setting the result equal to zero for maximum gives:

$$f_n^P = \exp\{-u_n - \lambda - 1\} q(n) \qquad (2.63)$$

Substituting equation (2.40) and using the constraints

(2.55) and (2.61) yields the Lagrangian multipliers:

$$e^{\lambda} = \frac{1}{(1 - e^{-u})^{z}} \quad , \quad u = \ln \frac{\bar{n} + z}{\bar{n}}$$

$$\longrightarrow \lambda = z \ln \left(\frac{\bar{n} + z}{z} \right)$$

which reduces the most probable distribution to:

$$f_n^{\,p} = \left(\frac{z}{\bar{n} + z} \right)^{z} \left(\frac{\bar{n}}{\bar{n} + z} \right)^{n} \frac{(n + z - 1)!}{n!(z - 1)!} \qquad (2.64)$$

Using expression (2.64), the three cases cited above can be reanalysed in the ensemble context.

a) $z = 1$ reduces the entropy to $- \sum_n f_n \ln f_n$ and the most probable distribution becomes:

$$f_n^{\,p} = \frac{1}{\bar{n} + 1} \left(\frac{\bar{n}}{\bar{n} + 1} \right)^{n} \qquad (2.65)$$

Substituting equation (2.65) into the entropy expression yields:

$$\frac{s\{f_n^{\,p}\}}{M} = 1 + (\bar{n} + 1) \ln(\bar{n} + 1) - \bar{n} \ln \bar{n}$$

$$\simeq 1 + \ln \bar{n} \quad \text{if } \bar{n} \gg 1 \qquad (2.66)$$

The distribution (2.65) is known as the exponential distribution. The expression for entropy (2.66) is closely related to that used by Shannon in his information theory, reference 12. The derivation of Shannon's result uses a set of samples, each corresponding to a given state i.e. $z = 1$. The summation $\sum_n f_n \ln f_n$ is introduced as an averaging procedure for the information content.

b) If $z \ll \bar{n}$ then:

$$\frac{s\{f_n^p\}}{M} = z \ln \bar{n} + z - z \ln z \qquad (2.67)$$

c) when z is large compared to n then using exactly the same approximation as before, expression (2.64) approximates to:

$$f_n^p = e^{-\lambda - un} \; z^n/n! \qquad (2.68)$$

where $u \simeq \ln(z/\bar{n})$ and $\lambda \simeq \bar{n}$. Substituting the Lagrangian multipliers gives:

$$f_n^p = \frac{e^{-\bar{n}} (\bar{n})^n}{n!} \qquad (2.69)$$

The corresponding entropy expression has the form:

$$\frac{s\{f_n^p\}}{M} = \bar{n} \ln z - \bar{n}(\ln \bar{n} - 1) \qquad (2.70)$$

It can be seen that the entropy expressions in the cases b) and c) closely resemble the previously described versions concerning the actual number of photons n rather than the average $\bar{n}$. The ensemble model also provides a result for case a), the classical wave field limit of Bose-Einstein statistics, governed by the exponential distribution. The particle limit c) is found to be governed by the Poisson distribution (2.69) which agrees well with counting statistics experiments. It should be noted that the Poisson distribution only results because the a priori degeneracy terms are included and the degeneracy within each member of the ensemble is significant. Using the correspondances $n_{\alpha\beta} = M\bar{n}$ and $z_{\alpha\beta} = Mz$ reveals an

equivalence between (2.43), (2.67) for $M \gg 1$ and (2.44), (2.70), hence the entropy of a given pixel is equivalent to the entropy of the ensemble constructed from that pixel. The entropy of the entire image will follow in exactly the same manner whichever model is chosen.

It is important to realise that the form of the entropy expression depends on the problem in hand and the correct expression can only be derived by setting up a representative model for the system. Maximising the entropy leads to two important results. Firstly it generates the most probable distribution f_n^p. Secondly it can provide the most probable count distribution for the entire image. f_n^p was derived by maximising the entropy under constraints and it will be seen that the most probable image $[n^p]'$ can also be found by maximising the configurational entropy $s\{\alpha_c\}$ under constraints imposed by the observed image data.

There is much suspicion and controversy about the definition and use of the concept of entropy. The Principle of Maximum Entropy originated in thermodynamics and the statistical interpretation was introduced by J. W. Gibbs and developed by J.C. Maxwell, L. Boltzmann and M. Planck. C. Shannon connected the idea of entropy with that of information and it is this concept which causes the problem. In order to apply information theory to the image processing, or indeed any other, problem it is necessary to introduce 'cells and quanta' but once this is done information theory takes over and gives Shannon's expression for entropy. Some authors like to think of the maximum entropy method as being founded on information

theory without concern for the inital quantization which is necessary in order to apply it. They argue that image restoration cannot possibly have anything to do with thermodynamics. Yet the two disciplines are linked as clearly indicated by Brillouin in reference 12 and it is the quantization which links them. The nature of the quantization has been considered above, providing a basis for the use of the MEM in image analysis.

2.7 Maximum entropy restoration.

There are several algebraic methods which are in current use that are based on the stochastic representation described briefly in section 2.2. They all have a very similar form to the basic discrete filtering case of section 2.3 with additional subtleties to improve the performance and tailor the response to particular imaging situations. Instead of developing the discrete filtering model and adapting it to the algebraic form, there is a completely different approach which might bear more fruit; that is to discard the stochastic model of section 2.2 and adopt in its place the quantum statistical model presented in section 2.6. This approach promises greater flexibility than discrete filtering but unfortunately introduces a possibly prohibitive computation burden into the image processing problem.

Maximisation of entropy is a very well tried technique in the field of statistical physics but it is usually only applied to theoretical problems. The so-called maximum entropy method attempts to apply the concept of entropy maximisation to the data analysis. The first

essential step is to derive the correct entropy expression for the problem given by the logarithm of the degeneracy and reference 8 makes it clear that the detailed form of the expression will depend on the problem in hand. Section 2.6 gives the analysis of the image formation case offered by Kikuchi and Soffer (reference 8). The analysis provides the probability distribution that governs the statistics on each pixel and the configurational entropy $s\{\alpha\beta\}$ that describes the degeneracy of the complete pixel set or image $[n]$.

The observed image must, in some way, act as a constraint on the restored image. In many cases this is conveniently achieved by using the Principle of Least Squares setting up a statistic of the form:

$$\chi^2 = \sum_{\alpha\beta} \frac{(J_{\alpha\beta} - \hat{J}_{\alpha\beta})^2}{\sigma_{\alpha\beta}^2} \qquad (2.71)$$

If the processes involved were normal, then χ^2 would be the chi-squared statistic but without that restriction it becomes a weighted least square error function, which must be minimised to achieve the most probable solution. The estimate $[\hat{J}]$ must be derived from the data using an image transformation equation of the form:

$$[\hat{J}] = [A][\hat{f}][B] \qquad (2.72)$$

where $[A]$ and $[B]$ define the instrument kernel and $[\hat{f}]$ is an estimate of the source distribution. The transforms $[A]$ and $[B]$ can have many varied forms and need not be a blur matrix. In radio astronomy applications using aperture synthesis $[A] = [B]^{-1}$ would be a Discrete Fourier

Transform (reference 11).

It is assumed above that the errors or 'noise' in each pixel are uncorrelated. This will be true in many imaging experiments, however if the errors are correlated a more general form of χ^2 must be used - χ^2_g given by:

$$\chi^2_g = (J - \hat{J})^t [W] (J - \hat{J}) \qquad (2.73)$$

$[W]$ is a weight matrix related to the covariance matrix of J by:

$$[\phi_J] = \sigma^2 [W]^{-1} \qquad (2.74)$$

σ^2 is simply a scale factor. If there is no correlation $[W]$ is diagonal with elements given by the reciprocal of the variance $\sigma^2_{\alpha\beta}$ of the individual data points and χ^2_g reduce to the simpler form equation (2.71).

If the data J has a non-zero mean noise component, this must be allowed for in $\hat{J}$ by introducing the mean noise level into equation (2.72):

$$[\hat{J}] = [A] [\hat{f}] [B] + [u_f] \qquad (2.75)$$

Since the minimisation of χ^2 is dependent on the normalisation or scaling of $[\hat{J}]$ the estimate $[\hat{f}]$ must be correctly normalised. Normalisation can be achieved by introducing the extra constraint that the sum of the total flux in $[\hat{f}]$ is consistent with the measured data flux.

In order to maximise the entropy under the data constraints an 'objective function' is set up having the form:

$$O_f = s_{\{\alpha\beta\}} - \lambda \chi^2 + u(\sigma_T - \sum_{\alpha\beta} \hat{f}_{\alpha\beta}) \qquad (2.76)$$

where λ and u are Lagrange multipliers and n_T is the observed photon flux. The expression used for the configurational entropy $s_{\{\alpha\beta\}}$ will depend on the observation. The detailed form of the objective function is of the utmost importance to the maximum entropy method and if it is not properly compiled, the resulting solution will not have the desired properties.

Taking X-ray imaging as an example and using equation (2.49) for the configurational entropy, equation (2.76) becomes:

$$O_f = \sum_{\alpha\beta} \hat{f}_{\alpha\beta} \ln z_{\alpha\beta} - \sum_{\alpha\beta} \hat{f}_{\alpha\beta} (\ln \hat{f}_{\alpha\beta} - 1) - \lambda \sum_{\alpha\beta} \frac{(J_{\alpha\beta} - \hat{J}_{\alpha\beta})^2}{\sigma_{\alpha\beta}^2} + u(n_T - \sum_{\alpha\beta} \hat{f}_{\alpha\beta}) \qquad (2.77)$$

Differentiating with respect to $f_{\alpha'\beta'}$ gives:

$$\frac{\partial O_f}{\partial \hat{f}_{\alpha'\beta'}} = \ln z_{\alpha'\beta'} - \ln \hat{f}_{\alpha'\beta'} + \lambda \sum_{\alpha\beta} A_{\alpha'\beta}^{t} \frac{(J_{\alpha\beta} - \hat{J}_{\alpha\beta})}{\sigma_{\alpha\beta}^2} B_{\alpha\beta'}^{t} - u \qquad (2.78)$$

Setting the derivative to zero to find the maximum gives:

$$\hat{f}_{\alpha'\beta'} = z_{\alpha'\beta'} \exp\{-u\} \exp\{\lambda \sum_{\alpha\beta} A_{\alpha'\beta}^{t*} \frac{(J_{\alpha\beta} - \hat{J}_{\alpha\beta}) B_{\alpha\beta'}^{t*}}{\sigma_{\alpha\beta}^2}\} \qquad (2.79)$$

u acts as a scaling factor and must be chosen to satisfy the observed count constant:

$$\sum \hat{f}_{\alpha\beta} = n_T \qquad (2.80)$$

λ controls the χ^2 of the solution $[\hat{f}]$ and setting $\lambda = 0$ removes the χ^2 constraint and gives the solution:

$$\hat{f}_{\alpha'\beta'} = z_{\alpha'\beta'} \exp\{-u\} \qquad (2.81)$$

The most probable solution in the absence of positional data is simply the normalised [z] matrix. This would be expected since from equation (2.53), z is seen to be proportional to the response tA for the observation.

Unfortunately equation (2.79) is transcendental and there is no analytic solution $[\hat{f}]$, the most probable solution. Therefore an iterative search must be used to find the solution. The exponential:

$$\exp\left\{\lambda \sum_{\alpha\beta} A_{\alpha\beta}^{t} \frac{(J_{\alpha\beta} - \hat{J}_{\alpha\beta})}{\sigma_{\alpha\beta}^{2}} B_{\alpha\beta}^{t}\right\} \qquad (2.82)$$

contains the major burden of the computation because it requires two sets of matrix multiplications, firstly equation (2.72) to give $[\hat{J}]$ and then the conjugate form (a crosscorrelation) to calculate expression (2.82). In the case of image blur, (2.82) performs the blurring of the instrument and the result is the cross-correlation of the weighted difference $(J_{\alpha\beta} - \hat{J}_{\alpha\beta})/\sigma_{\alpha\beta}^{2}$ with the point response. The design and structure of an algorithm to solve equations like (2.79) will be discussed in detail later.

In section 2.6 the statistics applicable to the photon counting situation were shown to be governed by the Poisson distribution and the greater the number of photons received from a particular feature in the field of view, the greater the significance of that feature. However in applying the MEM, correlation is introduced between features close together because of the blurring of the instrument.

The use of the χ^2 statistic ensures that all the data points are given their correct statistical significance in the final result, providing a reasonable χ^2 value is found. λ controls the value of χ^2 and interpretation of solutions with different χ^2 values will be discussed in section 3.6.

CHAPTER 3: THE PRACTICAL IMPLEMENTATION OF PROCESSING THEORY.

3.1 General processing requirements and machine restrictions.

Since all the processing described in Chapters 1 and 2 can be expressed in matrix notation, it is convenient to regard the processing software required as a system for handling large matrices. The system must include input and output in a convenient format and must be able to perform multiplication, transposition etc.. Because of the large number of elements in even the smallest images (1024 in a 32 x 32 matrix), 'processing by hand' is impractical and all the operations must be performed within a digital computer's central processing unit (CPU).

The power of the CPU in terms of speed and storage restrict the capability of the software. Using the CDC CYBER 72 computer of the University of Leicester Computer Department it is impossible to load a matrix larger than about 100 x 100 into the CPU along with associated software. Therefore in order to handle large matrices, the operations must be performed sequentially, a row or column at a time. Almost all operations required in image processing can be carried out sequentially and the only drawback to this approach is the channel time needed for reading and writing data between magnetic disc and the CPU. Although the storage capacity ($\sim$100,000 words for CYBER 72) poses a software problem, it can be overcome by careful software design and does not remain a fundamental machine limit. However the speed of the

CPU does limit the type of processing which can be achieved.

Performing a matrix multiplication such as a circular convolution requires a great deal of CPU time. Direct computation with 128 x 128 matrices, including channel time for accessing data, would require ~ 6 minutes and repeated submission of such demanding jobs would be very costly and turnround would be very slow. If possible, fast algorithms for calculating such multiplications must be utilised. Fast algorithms only exist for transformation matrices which can be factorised into a set of very sparse matrices. A detailed exposition on fast transforms was given in reference 13 and the structure of the FFT and similar algorithms will not be discussed here. A major difficulty is providing matrix output in a compact and easily assimilated form. Since the major purpose of the software is image processing, a method of display using light intensity levels is obviously advantageous and meaningful. Various modes of displaying the image matrices were tried with varying degrees of success using the full range of peripheral output devices available.

3.2 Large matrix processing software.

The software was developed over a period of two years in order to cope with processing problems as they arose. Matrices are held on magnetic disc files, one record per row of data, each matrix element occupying one Cyber word (60 bits). All processing within the CPU uses square matrices with a maximum dimension of 1024 x 1024. Non-

square matrices are easily handled by padding them out with zeros to form a square equivalent. Fast transformations are restricted to dimensions equal to powers of two for convenience. This is not strictly necessary since mixed radix transformations could be used for any factorisable dimension, however the power of two restriction does not reduce the power of the software since matrices can always be padded out to a power of two using zeros.

Complex matrix data as found within the DFT domain is stored sequentially in records, one record for the real part of a row and a following record for the imaginary part of a row. Since the quadrants of the DFT domain exhibit a conjugate relationship, only one half of the complex values needs to be stored (reference 14, page 118). This redundancy within the DFT domain saves both magnetic disc storage space and processing time.

A detailed breakdown of all the routines available and how to use the software is given in Appendix I.

The processing system has been used to simulate experimental data of two rocket payloads. The MIT/ Leicester imaging payload for surveying super-nova remnants and the MPI/ Leicester imaging payload originally designed to see a scattering halo due to cosmic dust around a point source. The parameters for both these payloads were given in section 1.3.

Following on from simulations, the system was used for the processing of flight data from the MIT/ Leicester payload providing images of Sco X-1, Cygnus Loop, Puppis A, IC443 and several update stars. These images had to be

exposure corrected to remove the beam profile distortion and noise filtered because of the counting statistics. Although the statistics were poor, deblurring was attempted by the method described in Chapter 2 and despite the poor noise characteristics, some 'feature extraction' was achieved.

3.3 The MIT/ Leicester rocket flight data.

The form of the event set was exactly described in Chapter 1 except that the axis of the instrument was scanned across the sky during observation. Therefore the transformation from detector co-ordinates to sky co-ordinates included a time factor as well as a scale factor. Using the equation of motion in sky co-ordinates of the instrument axis, the event co-ordinates were transformed into sky co-ordinates. This procedure is described in more detail in reference 15. The event set was then binned to form an image matrix.

The first flight payload which looked at the Cygnus Loop had an on-axis resolution of ~15' FWHM (see figure 1) and a bin size of 4' x 4' was chosen to give reasonable definition across the point response. After background rejection and removal of suspect counts mispositioned by an electronic fault, the total count from the loop was 6741 in the energy range 0.15 - 1.12 KeV (as defined by the calibrated anode pulse height). Using 4' x 4' bins gave a largest bin count of 19. The observation was obviously 'photon limited'. Using 4' x 4' bins, the total exposed area of sky conveniently fitted onto a 64 x 64 matrix, the edges of which received very little exposure

(see later) and provided a clear border to the image so that any 'edge effects' would not be troublesome in the linear processing (see section 2.1). Figure 23 is the raw count image of Cygnus Loop.

The degradations present in figure 23 are as follows. The exposure is not constant over the matrix but a rather complicated function of the instrumental beam profile and the equations of motion of the two scans which were made to give complete coverage of the loop. The image has been blurred by the instrument point response, which was unfortunately non-linear. However because of scanning (North to South and then South to North), the non-linearity of the response has been largely removed and a single linear blur matrix can be used to describe this degradation adequately. The non-photon background count is fairly low since most cosmic ray events could be discriminated against using anticoincidence signals. Assuming the antico cathode to be 70% efficient, the estimated residual background was 428 counts - a mere 0.1 counts per pixel. The observation is clearly source count limited and because the object is diffuse, the total 6741 counts are widely distributed over the pixels with an average of 1.6 counts/pixel. It is believed that errors in unfolding the event set using the derived equations of motion were small ($\prec 1$ pixel) and do not distort the image to any great extent compared with the other degradations.

The exposure of each pixel must be calculated in order to correct for apparant brightness variations caused by the exposure alone. Fortunately although the magnitude of the instrument efficiency was a strong function of

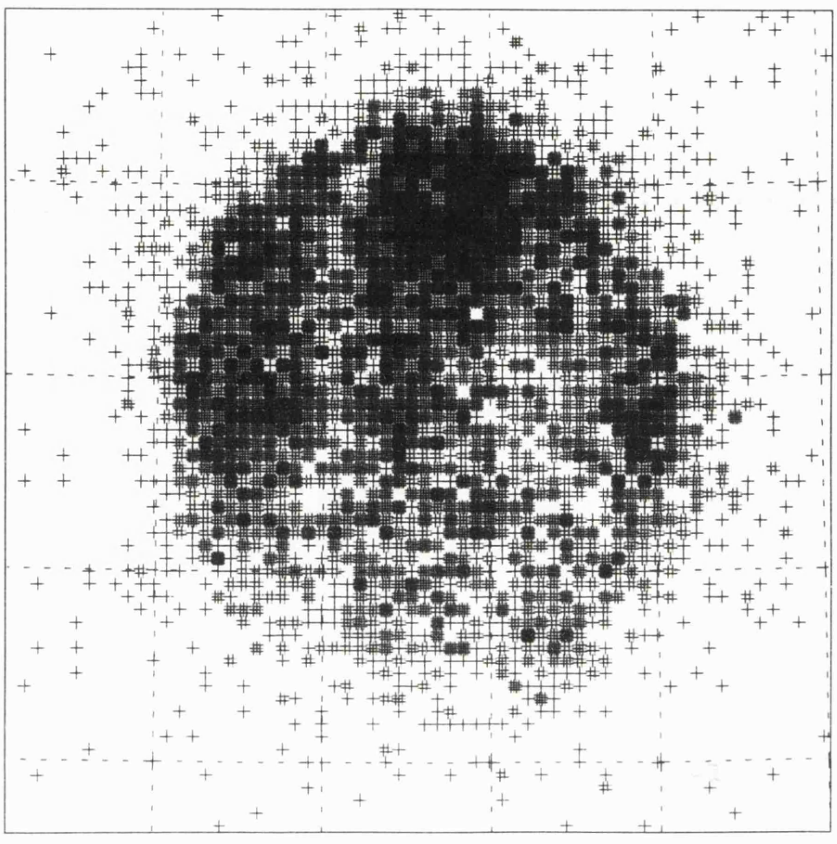


Figure 23a. Raw count image of the Cygnus Loop.
One cross per count, pixels 4' x 4',
energy range 0.15 - 1.12 KeV.

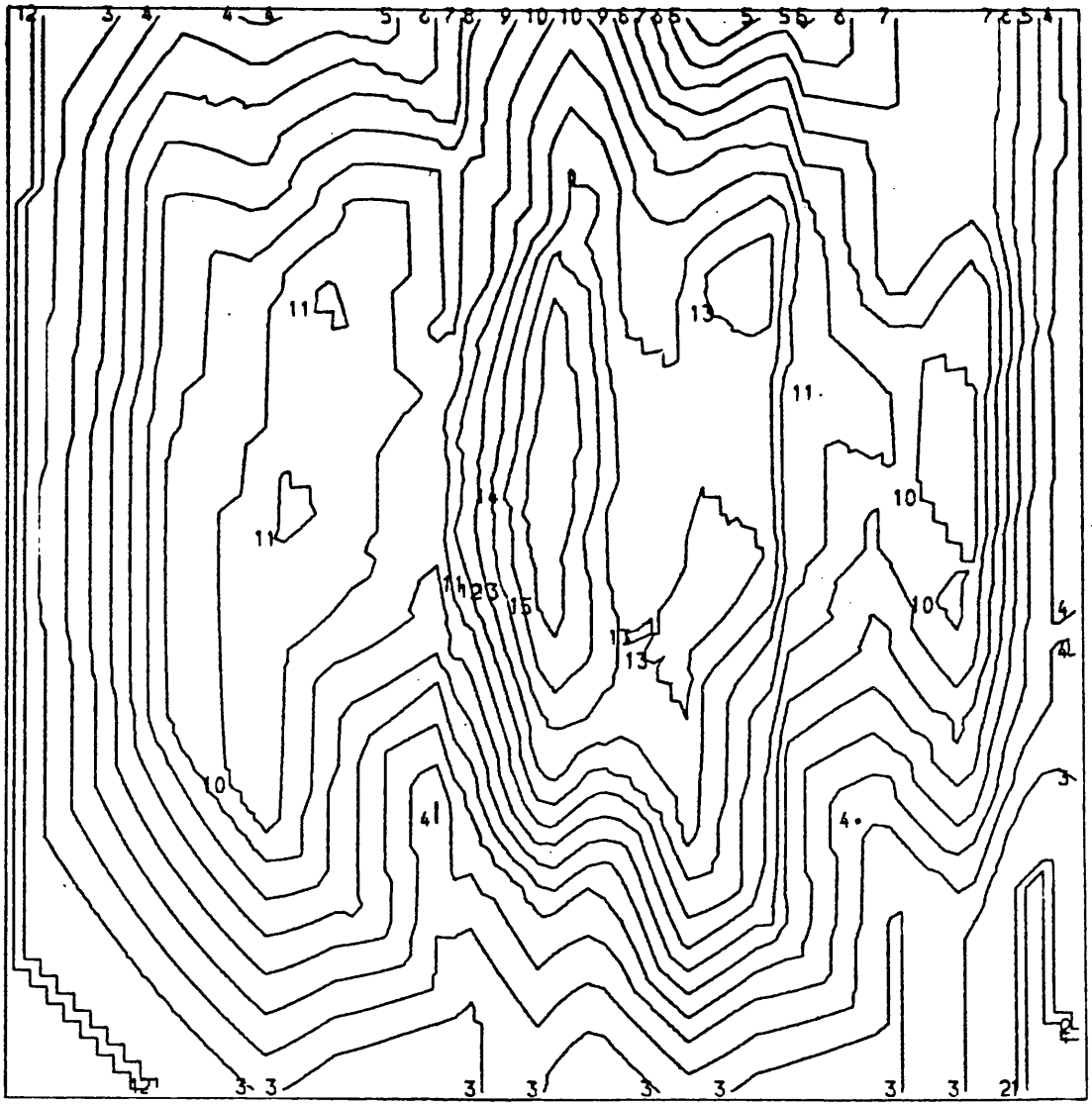


Figure 23b. A contour map of the exposure for the Cygnus Loop observation. The contour interval is approximately 40 cm^2 sec and the average exposure is about 240 cm^2 sec.

energy, the beam profile was a very weak function of energy (see figure 4). The relative exposure suffered by each pixel is therefore dependent upon only a single beam function and the equations of motion:

$$E(\alpha,\beta) = \int B(\alpha - f_1(\alpha,t),\ \beta - f_2(\beta,t))\, dt \qquad (3.1)$$

where f_1 and f_2 are the equations of motion. Integral (3.1) was calculated for each pixel (α,β) by a simple numerical integration technique, using a beam matrix to represent the sampled beam profile. Fortunately f_1 and f_2 were linear and the beam profile smooth so that the final exposure matrix was smooth and a sophisticated integration method was not needed to give good results. Figure 23bis the exposure matrix corresponding to the raw image in figure 23.

In order to correct for exposure difference, each pixel must be weighted bythe reciprocal of the exposure it received. However the result must also be normalised in some way so that the least distortion of scale occurs. The least mean square difference criterion applied between the uncorrected and corrected map yields the following result:

$$C\ J_{ij}\ E_{ij} = J'_{ij} \qquad (3.2)$$

C is chosen such that:

$$\sum \xi_{ij} = \sum (J'_{ij} - J_{ij})^2 \quad \text{is a minimum}$$

$$\frac{\partial \xi_{ij}}{\partial C} = 2C\ J_{ij}^{\,2}\ E_{ij}^{\,2} - 2J_{ij}\ J_{ij}\ E_{ij} \qquad (3.3)$$

For a minimum it is required that:

$$\sum C\left[J_{ij}^2 E_{ij}^2 - E_{ij} J_{ij}^2\right] = 0$$

$$C = \sum J_{ij}^2 E_{ij} \Big/ \sum E_{ij}^2 J_{ij}^2 \qquad (3.4)$$

That is, if C satisfies (3.4), then the minimum mean square difference is achieved. Unfortunately in regions which suffered very low exposure and consequently have low signal to noise, the exposure correction will lead to abnormally large fluctuations that are due almost entirely to the counting statistics noise. The only way to avoid this is to rolloff the exposure correction E_{ij}, avoiding very small exposure values which would otherwise degrade the image. This distortion of the exposure matrix has very little effect on the validity of the result since only very underexposed and therefore low count regions are affected. (In practice only about 4 counts in figure 23 had to be suppressed in this way and they occured at the very edges of J_{ij}.) The exposure correction obviously destroys the Poisson nature of the statistics but not to a large extent since the dynamic range of the exposure matrix, figure 23, is not too large.

The presence of fairly severe image blur with low signal to noise sets a difficult problem in processing. The two degradations are to some extent complimentary, since suppressing the noise tends to increase the blurring and sharpening up the image inevitably increases the noise. The methods discussed in Chapter 2 were carefully tried and tested using the Cygnus Loop data as the prime example of a degraded astronomical image.

Information about the point response or blur matrix

associated with figure 23 was drawn from three sources. Firstly, the payload calibration data and theoretical instrument response as already presented in section 1.3. Secondly, the in-flight response of the payload to Sco X-1 and thirdly, modulations present in the unfolded data matrix, figure 23, indicating the upper limit of the blurring present. Figure 24 is an illustrated comparison of these sources of information. The final response to the predominantly soft X-rays of the Cygnus Loop was probably ~15' FWHM and certainly no worse than 20' FWHM. The detailed shape of the response was not Gaussian and thus was not as peaked as the theoretical response, however the precise shape is not important since the observation was severely count limited and enhancement of detail was therefore not possible.

3.4 Simple noise suppression for the Cygnus Loop data.

Simple noise suppression can be achieved by correlation of adjacent pixels using a convolution with a suitable function of the appropriate width. Perhaps the simplest form is the top hat function, as defined above in equation (1.13)a. The discrete convolution has the form:

$$J'_{ij} = \sum_{x,y} J_{x,y}\, T_{i-x,j-y} \qquad (3.5)$$

If:

$$\sum_{ij} T_{ij} = 1 \qquad (3.6)$$

then:

$$\sum_{ij} J'_{ij} = \sum_{x,y} J_{x,y} \qquad (3.7)$$

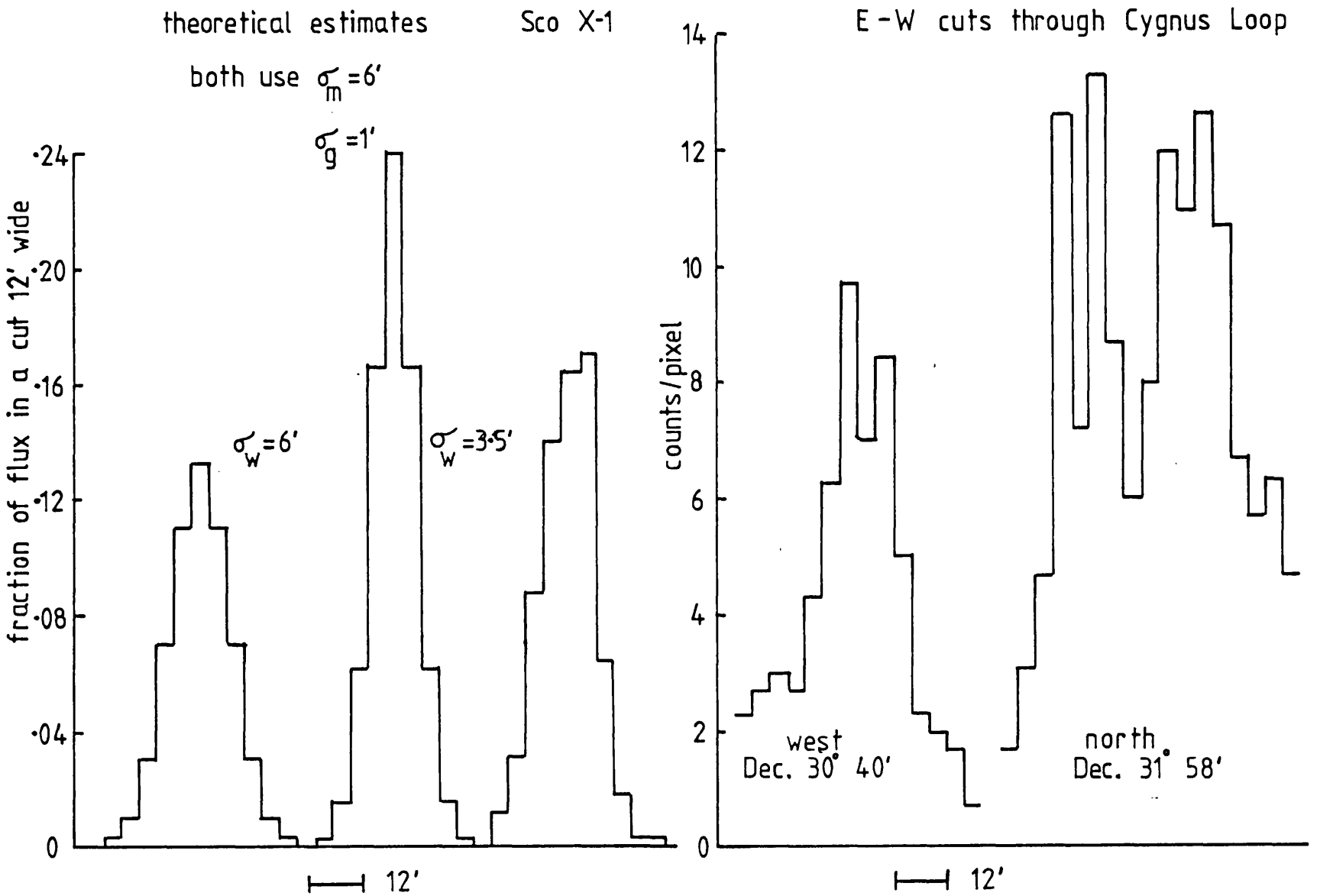


Figure 24. Comparison of various point response data for the Cygnus Loop observation.

and the total count contained in J will be unaltered. Since equation (3.5) is a linear operation, the variance of element J'_{ij} is simply given by the sum of the weighted covariances involved in the convolution:

$$\sigma'^2_{ij} = \sum_{x,y} \sigma_{xy}^2 \, T^2_{i-x,j-y} \qquad (3.8)$$

When the statistics are Poisson, $\sigma_{xy}^2 = J_{xy}$, so that the variance matrix σ'^2_{ij} corresponding to J'_{ij} is easy to calculate using equation (3.8). The change in mean signal to noise is given by:

$$\frac{\sum_{ij} J_{ij}}{\sum_{ij} \sigma_{ij}^2} = 1 \qquad \xrightarrow{\text{convolution with } T}$$

$$\frac{\sum_{ij} J'_{ij}}{\sum_{ij} \sigma'^2_{ij}} = \frac{N}{\sum_{ij}\sum_{xy} \sigma_{xy}^2 \, T^2_{i-x,j-y}}$$

Providing the convolution limits of x,y are large enough and no edge effects are present, the denominator can be simplified:

$$\longrightarrow \quad \frac{N}{N \sum_{ij} T_{ij}^2}$$

The signal to noise has been improved by a factor:

$$\mathfrak{Z} = \frac{1}{\sqrt{\sum_{ij} T_{ij}^2}} \qquad (3.9)$$

If the function T is very narrow then:

$$T_{ij} = \delta_{ij} \qquad (3.10)$$

and $\mathfrak{Z} = 1$ as would be expected. As T is made wider, $\sum T_{ij}^2$

will diminish and the corresponding noise suppression will increase. However the operation (3.5) which produces the noise suppression also blurs the image matrix J_{ij} and a corresponding loss of resolution results.

The summation (3.5) can be calculated directly, but if T is not a sparse matrix then diagonalisation using the Discrete Fourier Transform can be used (see section 2.1) to reduce the number of operations required. The expression of the convolution as a direct product in the DFT leads naturally to a more general Fourier filtering approach which will be dealt with in the following section.

The Cygnus Loop data (figure 23) was subjected to a 'top hat filter' of 12' x 12' (3 x 3 pixels) chosen to be within the extent of the point response so as not to degrade the resolution by too much. The result was then exposure corrected as described above, giving the image in figure 25. The noise suppression achieved is obviously considerable but the corresponding loss of resolution is made apparant by figure 26, which compares the Sco X-1 image before and after application of the 12' x 12' 'top hat filter'.

3.5 Fourier filtering of the Cygnus Loop data.

Since the linear filtering operation described in section 2.4 can be reduced to a direct product operation in the DFT, a parallel description of the operation of the filter can be made in the Fourier domain. Such a linear operation is naturally global in effect and its effect is described in terms of 'means' or averages. In

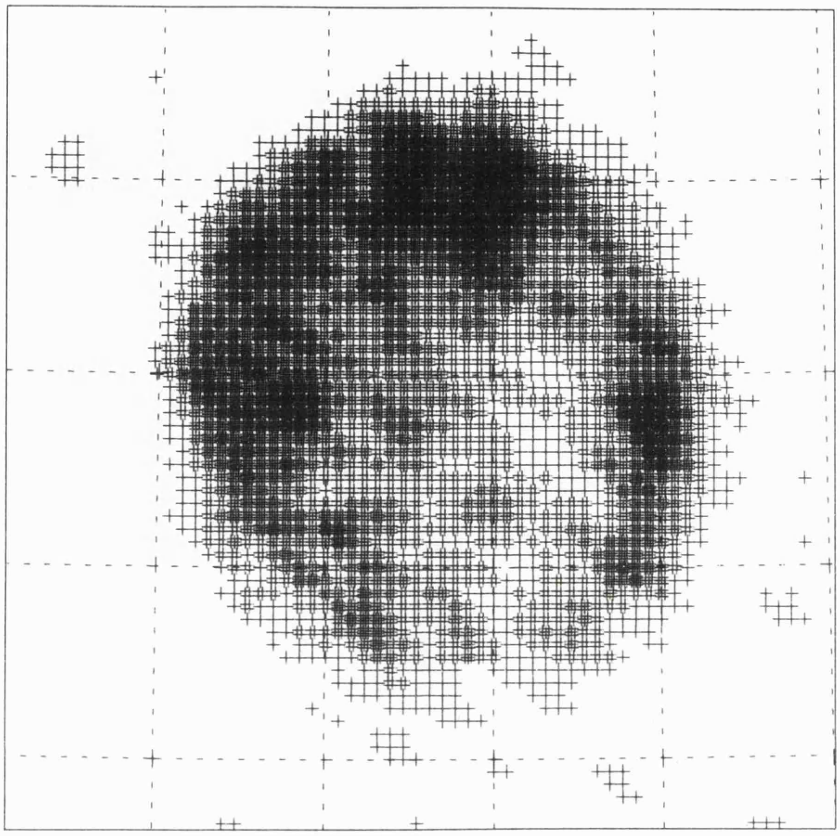


Figure 25. Top hat filtered and exposure corrected Cygnus Loop image.

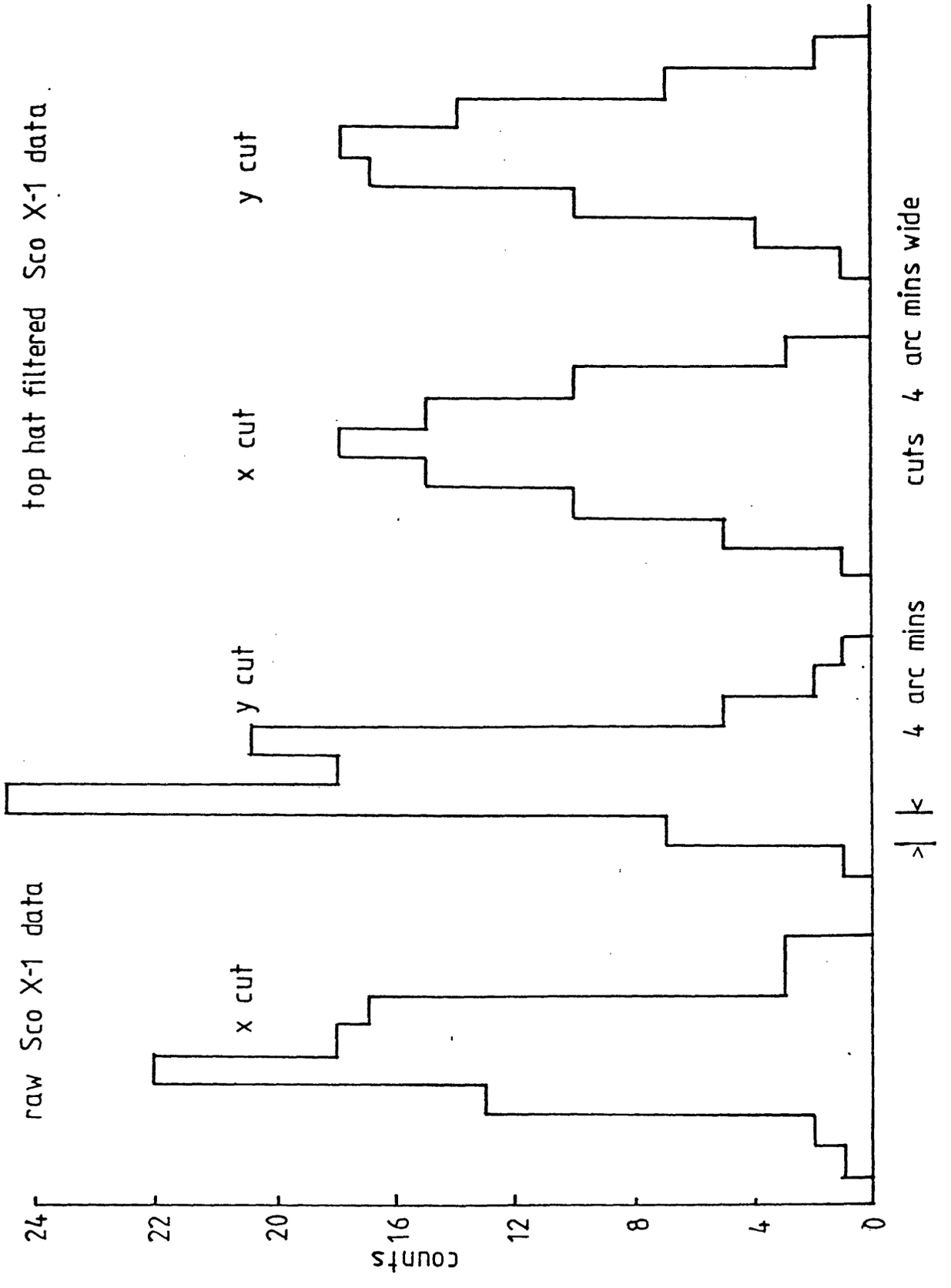


Figure 26. Comparison of the Sco X-1 profile before and after Top hat filtering.

the detailed presentation of section 2.4, the ratio of the signal power spectral density to the noise power spectral density was shown to be of key importance. The first step is therefore to estimate this ratio for the data matrix in figure 23.

If both signal and noise processes are assumed white, which is reasonable for the noise but not justified for the signal, then:

$$P_{npq} = P_n = \sum_{ij} J_{ij} \tag{3.11}$$

$$P_{fpq}' = P_f' = \sum_{ij} J_{ij}^2 - \sum_{ij} J_{ij} \tag{3.12}$$

The summations are very easy to calculate and in the case of Cygnus Loop give $P_n = 6741$, $P_f = 23537$. The signal to noise is therefore about 3.5. However the signal has suffered a power loss due to the action of the point response and this represents the image signal to noise rather than the source signal to noise. Therefore the source signal to noise required is probably something like 5. This estimate is obviously very rough and can be improved a good deal by a study of the modulation transfer function of the instrument $\bar{k}_{pq}$ (see section 2.4.) and the precise form of the image spectral density function.

The MTF of the instrument is shown in figure 27, corresponding the the ~15' FWHM response to Sco X-1. Variations present in the FWHM obviously change the cutoff frequency to some extent but have surprisingly little effect on the general form of figure 27. The 2-D power spectrum of the observation was calculated using

the FFT and in order to improve the statistics, the averages around annuli of constant spatial frequency amplitude were calculated yielding a radial power spectral density function plotted in figure 28.

The spectrum consists of two distinct components, the (as expected) flat noise spectrum extending to high spatial frequencies and a low frequency peak due to the Cygnus Loop. The MTF of the instrument is also shown in figure 28 for comparison. In order to find the source radial power spectrum, the noise component $\hat{\phi}_n$ was estimated using the high frequency region and this was subtracted from each sample. The resulting image spectrum is shown in figure 29. This image spectrum was then divided by the MTF to give an estimate of the original source spectrum before it was filtered by the instrument response. This is also shown in figure 29.

A Wiener filter in the Fourier domain (equation (2.34)) was constructed in two ways using the above information. Firstly using the crude signal to noise ratio estimate giving a constant value of about 5 (also shown in figure 29 for comparison) and secondly using a 2-function analytic fit to the estimated radial power spectral density (again shown in figure 29). The resulting images were then exposure corrected and plotted along with an image of Sco X-1, which was fitted in exactly the same way, see figures 30 and 31. Both images are clearly noise suppressed; the cruder signal to noise estimate causes amplification of mid-frequency noise. The 2-function analytic fit produced a very smooth result which appears to be very conservative towards detail. This effect is due

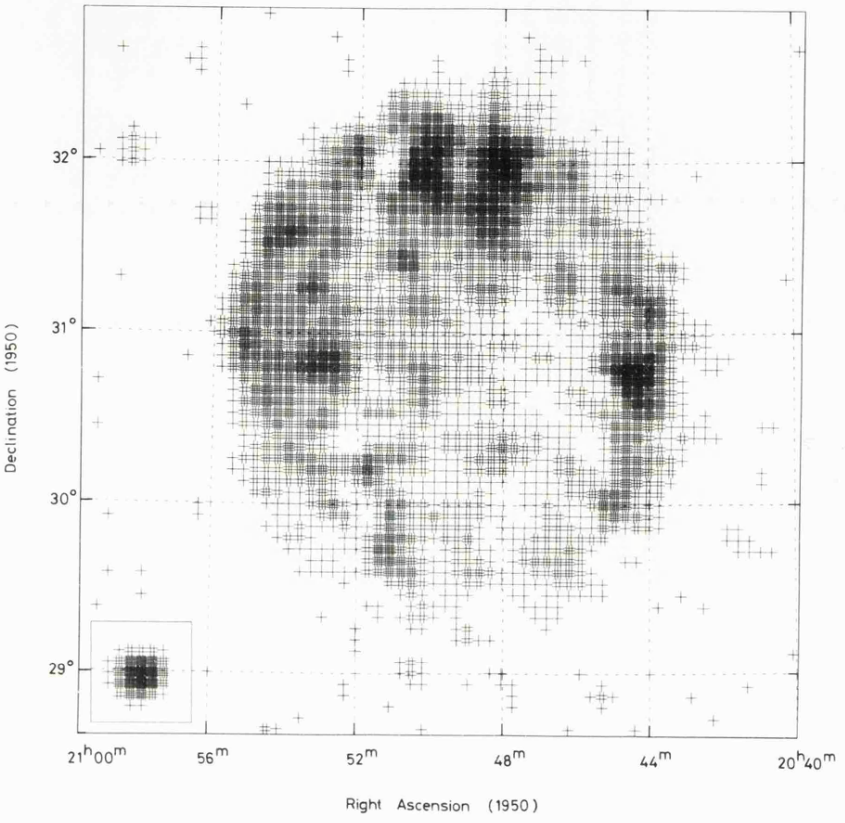


Figure 30. The Wiener filtered and exposure corrected Cygnus Loop image using a 'white' estimate for the signal to noise power spectral density ratio. The inset shows Sco X-1 filtered in the same way.

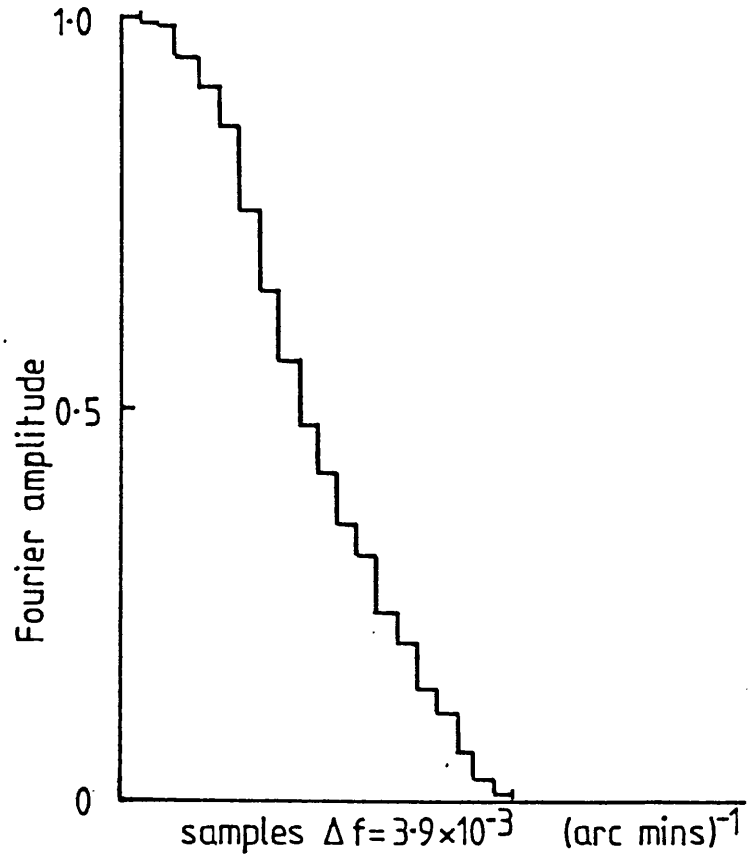


Figure 27. MTF for Cygnus Loop observation.

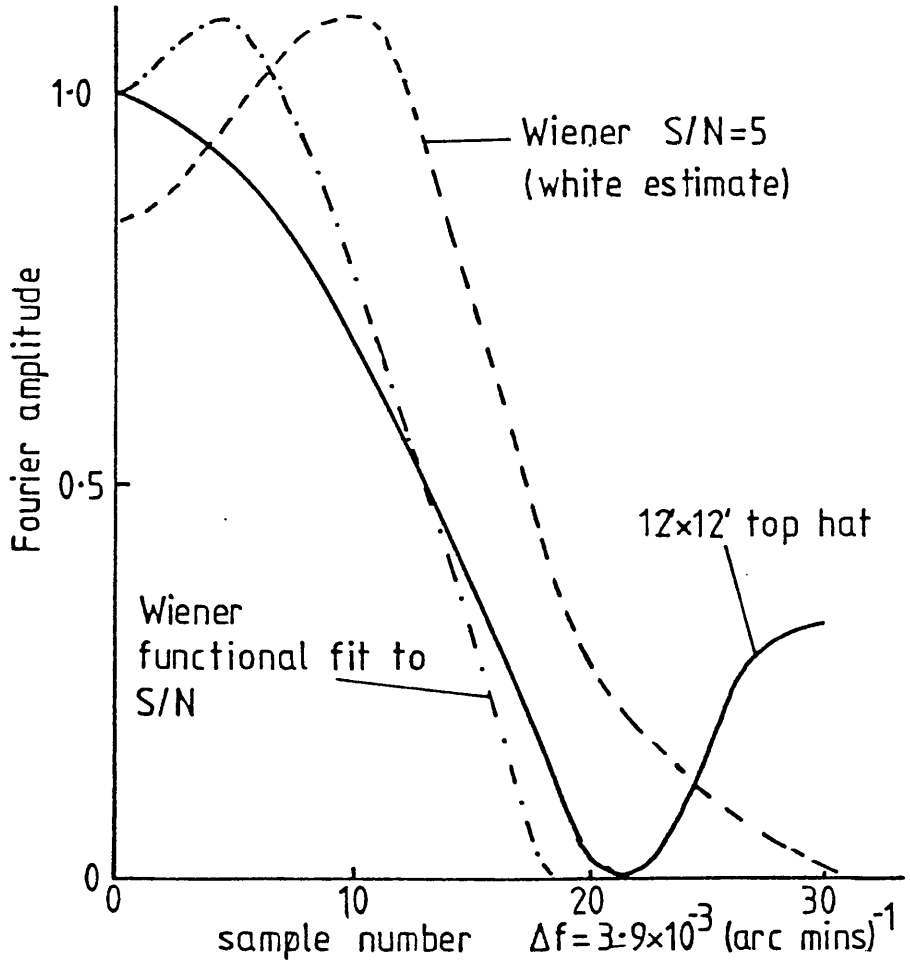


Figure 32. MTF's of filters used on the Cygnus Loop.

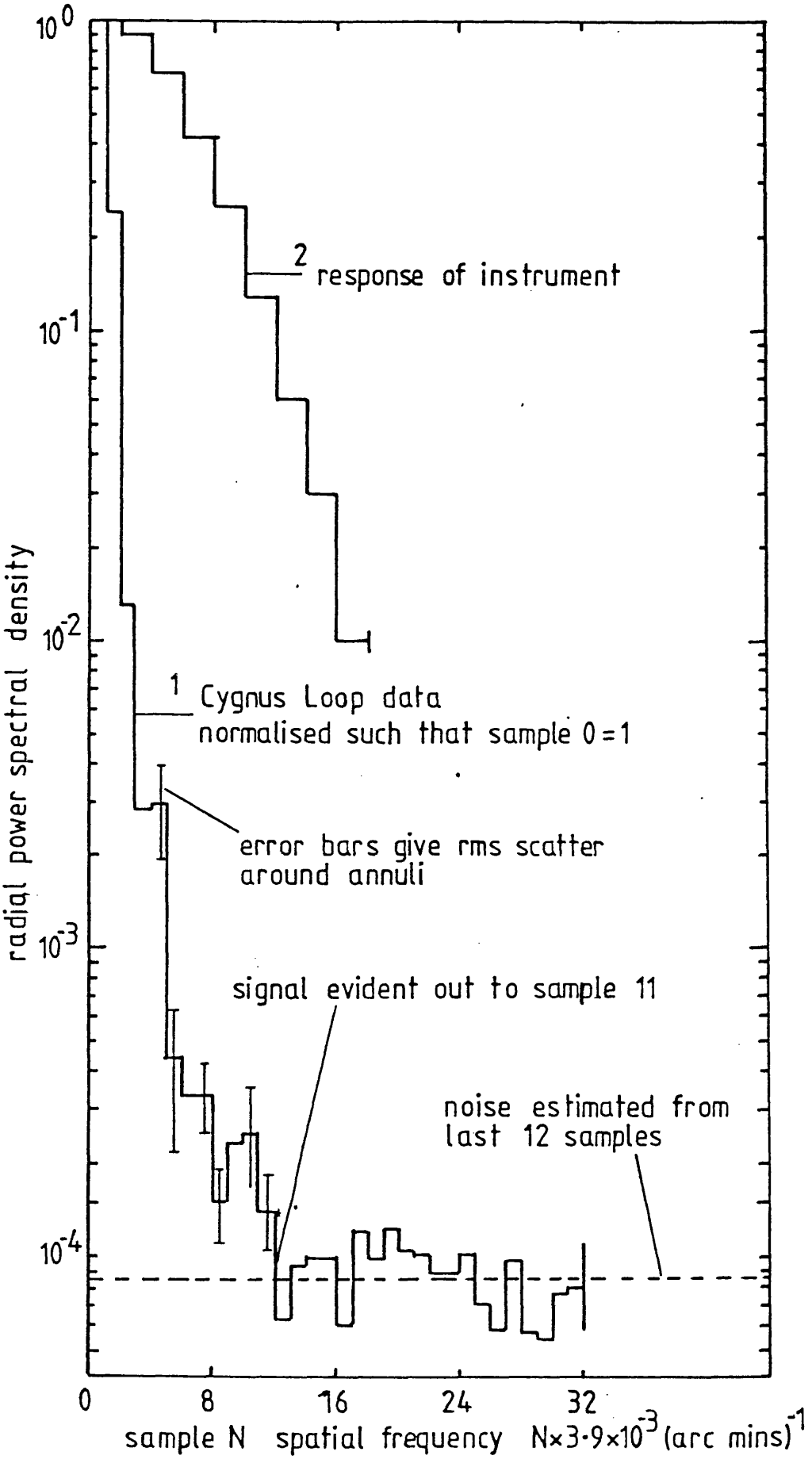


Figure 28. Radial power spectrum analysis of the Cygnus Loop data.

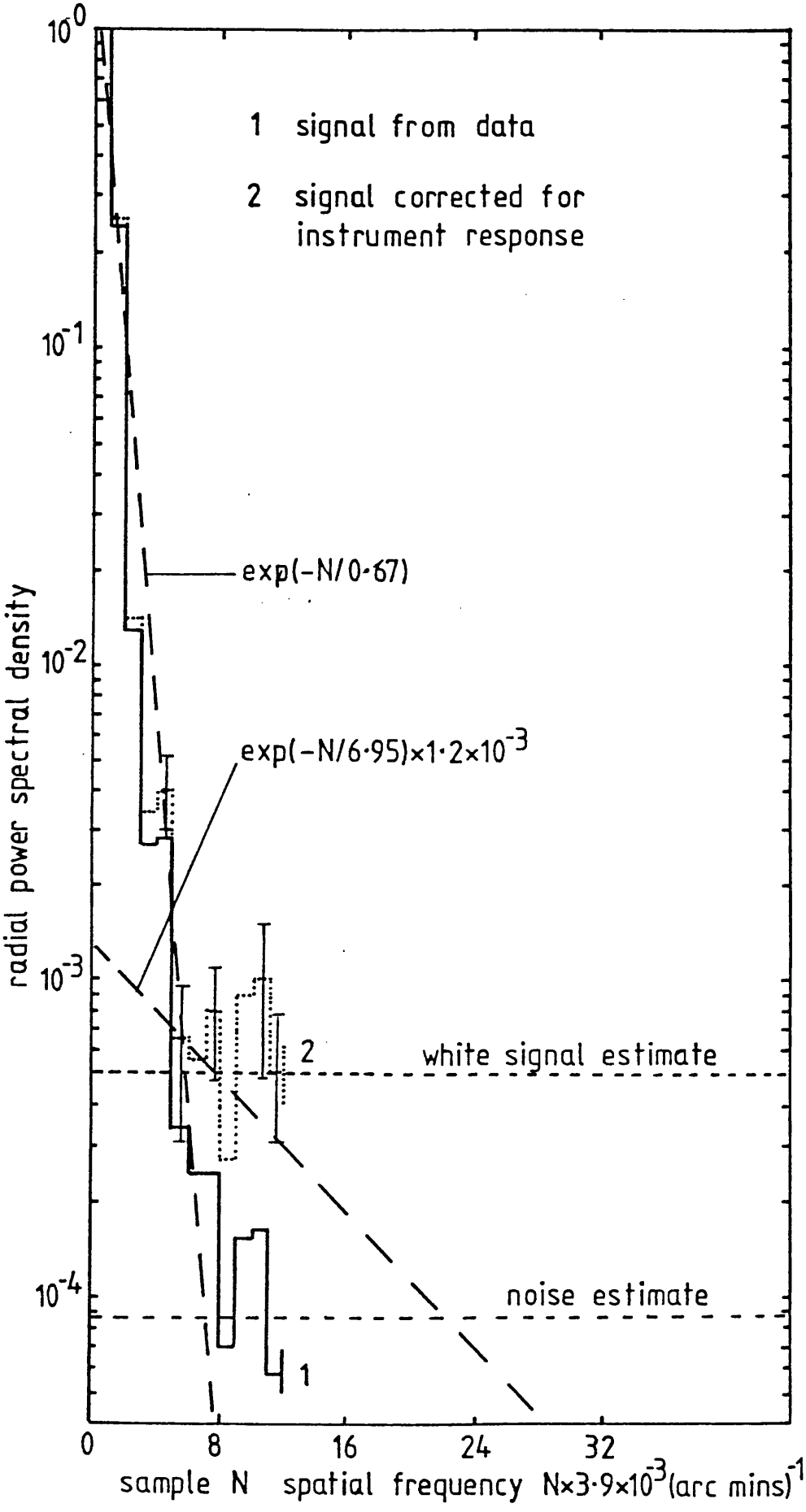


Figure 29. Estimated source power spectral density of the Cygnus Loop.

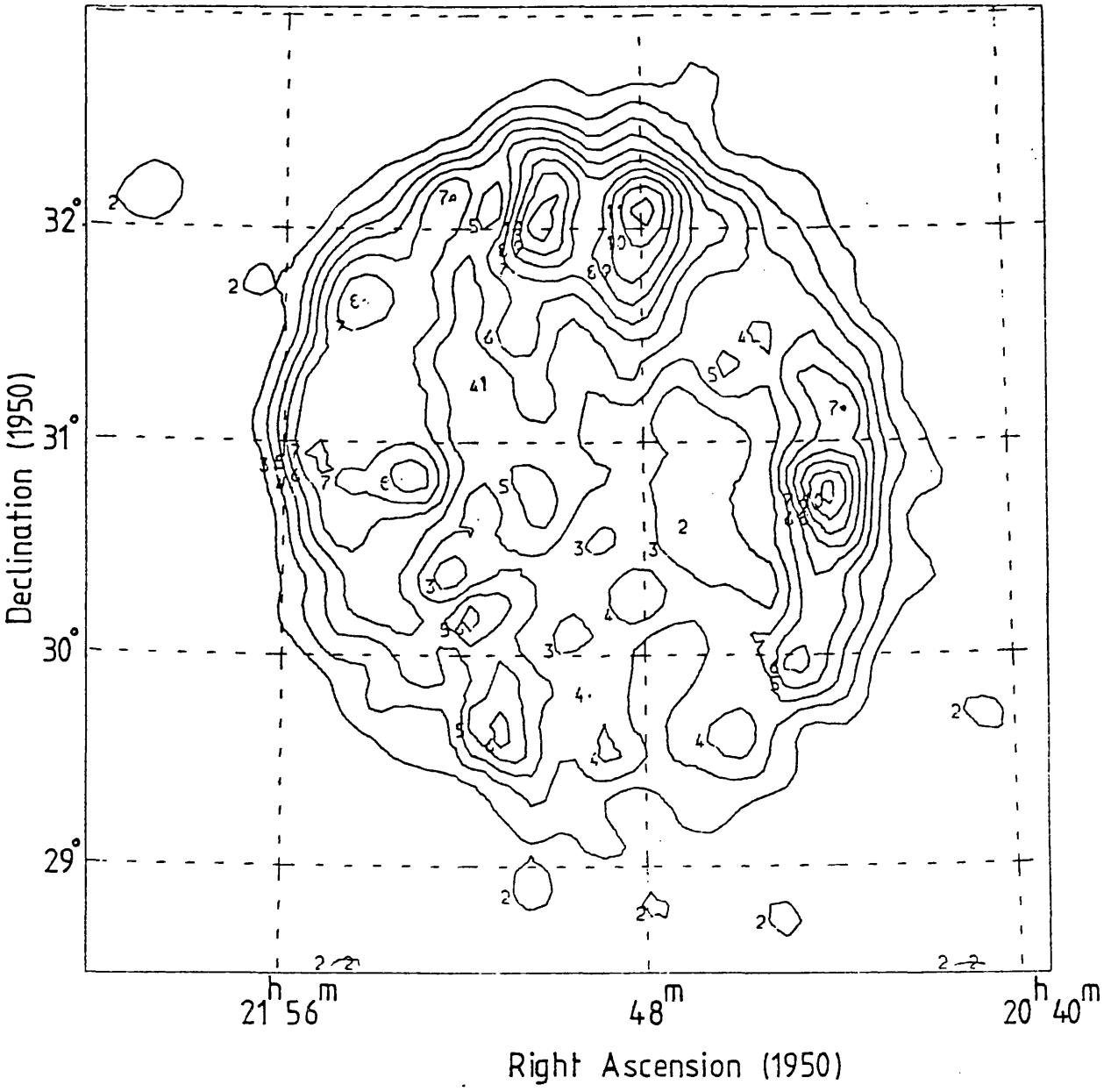


Figure 31a. Contour plot of the Wiener filtered and exposure corrected Cygnus Loop image using a two function analytic fit for the signal to noise power spectral density ratio. Contour interval 1 count/4' x 4'.

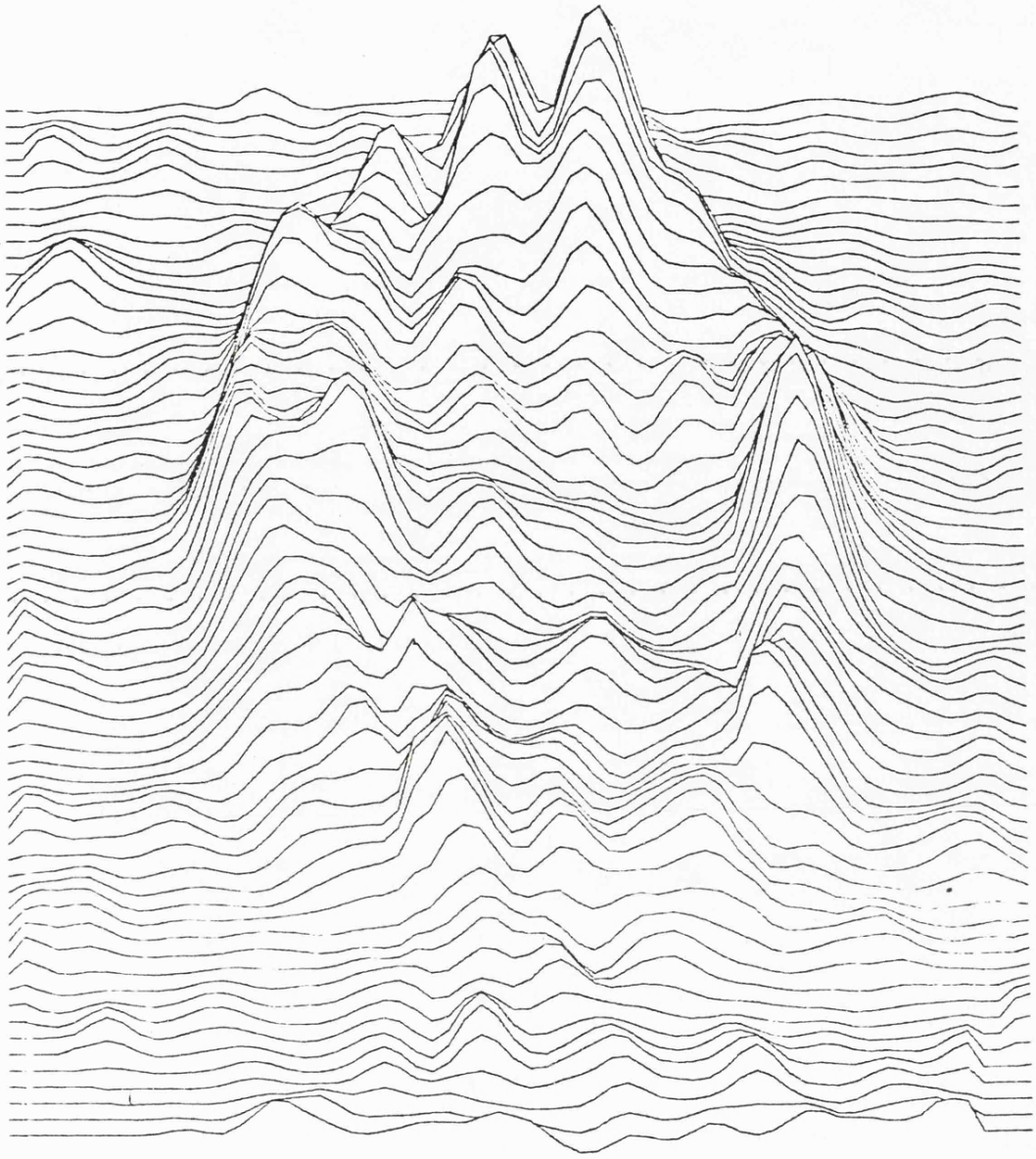


Figure 31b. Isometric projection of the image in figure 31a. Vertical scale 0.3 counts/4' x 4'/mm.

to the global nature of the filter which suppresses regions of good signal to noise (high count).

Figure 32 compares the MTF's of the top hat, white Wiener filter and 2-function analytic fit Wiener filters. The top hat suppresses low frequency noise but allows some leak through of noise at high spatial frequencies. The white Wiener filter gives optimistic feature extraction and poor low frequency response while the 2-function analytic fit Wiener filter provides excellent high spatial frequency noise suppression with good low frequency restoration.

3.6 The application of the Maximum Entropy Method to the Cygnus Loop data.

An algorithm was required to solve equation (2.79) with $[A]$ and $[B]$ as the blur matrices desribing the point response detailed above, $[\sigma^2]$ as the variance matrix corresponding to figure 23, the matrix $[z]$ proportional to the exposure matrix (see equation (2.53)) and λ and u to be found to satisfy constraints (2.80) and χ^2 (2.71) a minimum. The algorithm had to be made to converge quickly because of the large amount of computation involved and after much trial and error and with educated guess work, this was achieved!

An iteration formula based on equation (2.79) was used with a 'damping factor' included to prevent divergence and oscillation.

$$\hat{f}^{i+1}_{\alpha'\beta'} = \frac{1}{\gamma} z_{\alpha'\beta'} \exp\{-u\} \exp\left\{\lambda \sum_{\alpha\beta} A^{t}_{\alpha'\beta} \frac{(J_{\alpha\beta} - \hat{J}_{\alpha\beta}) B^{t}_{\alpha\beta'}}{\sigma^{2}_{\alpha\beta}}\right\} - \hat{f}^{i}_{\alpha'\beta'} \frac{(1-\gamma)}{\gamma} \qquad (3.13)$$

where:

$$\hat{J}_{\alpha\beta} = \sum_{\alpha'\beta'} A_{\alpha\beta'} \hat{f}^{i}_{\alpha'\beta'} B_{\alpha'\beta} \qquad (3.14)$$

It was found that $\gamma = 2$ worked well when $0 < \lambda \leq 2$ but not for larger λ. (If $\gamma = 2$, then (3.13) effectively takes the arithmetic mean between the new and old estimates.) Using $\gamma = 3$ gave a well controlled convergence for the Cygnus Loop data and also in other applications to be discussed later. The parameter u was not entered explicitly but introduced by normalising to the total count $\sum J_{\alpha\beta}$ in each pass when calculating χ^2. The solution seemed to be fairly weakly dependent on λ which only needed to be incremented slowly if the solution converged with too large a value for χ^2. Convergence was quickest if the initial estimate was the normalised exposure matrix corresponding to the least solution for $\lambda = 0$. The final value of λ which gives a small χ^2 obviously depends on the severity of the blurring and the signal to noise. The better the signal to noise and the worse the blur, the larger λ must be to yield a small χ^2. For data like the Cygnus Loop observation, $\lambda \approx 5$ was adequate and the solution at $\lambda \approx 2$ was not very different from the final solution.

The variance matrix used had to be carefully chosen. For non-zero pixels the best estimate for σ_{ij}^{2} was obviously J_{ij} but this was useless for zero elements. In order to find an estimate of the average count in a zero

pixel the image was crosscorrelated with the point response and zero pixels set equal to the new value. The exact form of the function used to find the average was found to have little effect on the final solution shape and a top hat would do just as well.

The algorithm could not be said to converge completely but the solution changed very little after about 6 iterations. Increasing λ to try and reduce χ^2 still further was not very effective because of the very flat χ^2 against λ curve, as illustrated by figure 33. In fact, the interpretation of χ^2 in this application must be made with care. If $\chi^2 \leqslant$ the number of data points N, then the 'fit' of the solution to the data is obviously good and if $\chi^2 > N$, the 'fit' is obviously poor. However χ^2 can also be a measure of how significant features in the solution are. If $\chi^2 \rightarrow 0$ indicating an excellent fit, then fluctuations due to noise will have penetrated into the final solution. When χ^2 is large ($\gg N$), any features which appear must be way above the noise level and hence very significant. A feature need not be a single pixel but is more likely to be a group of pixels and the significance of such a feature will depend on the total count it contains. If the blurring is severe and the signal to noise is good, a $\chi^2 < N$ will be difficult to obtain. However a solution with $\chi^2 > N$ need not be rejected since the large χ^2 merely reflects the inadequacy of the instrument resolution and features in the solution will be resolution limited.

Hence the χ^2_{min} is a measure of the limitation of the observation rather than just the 'fit' to the observed

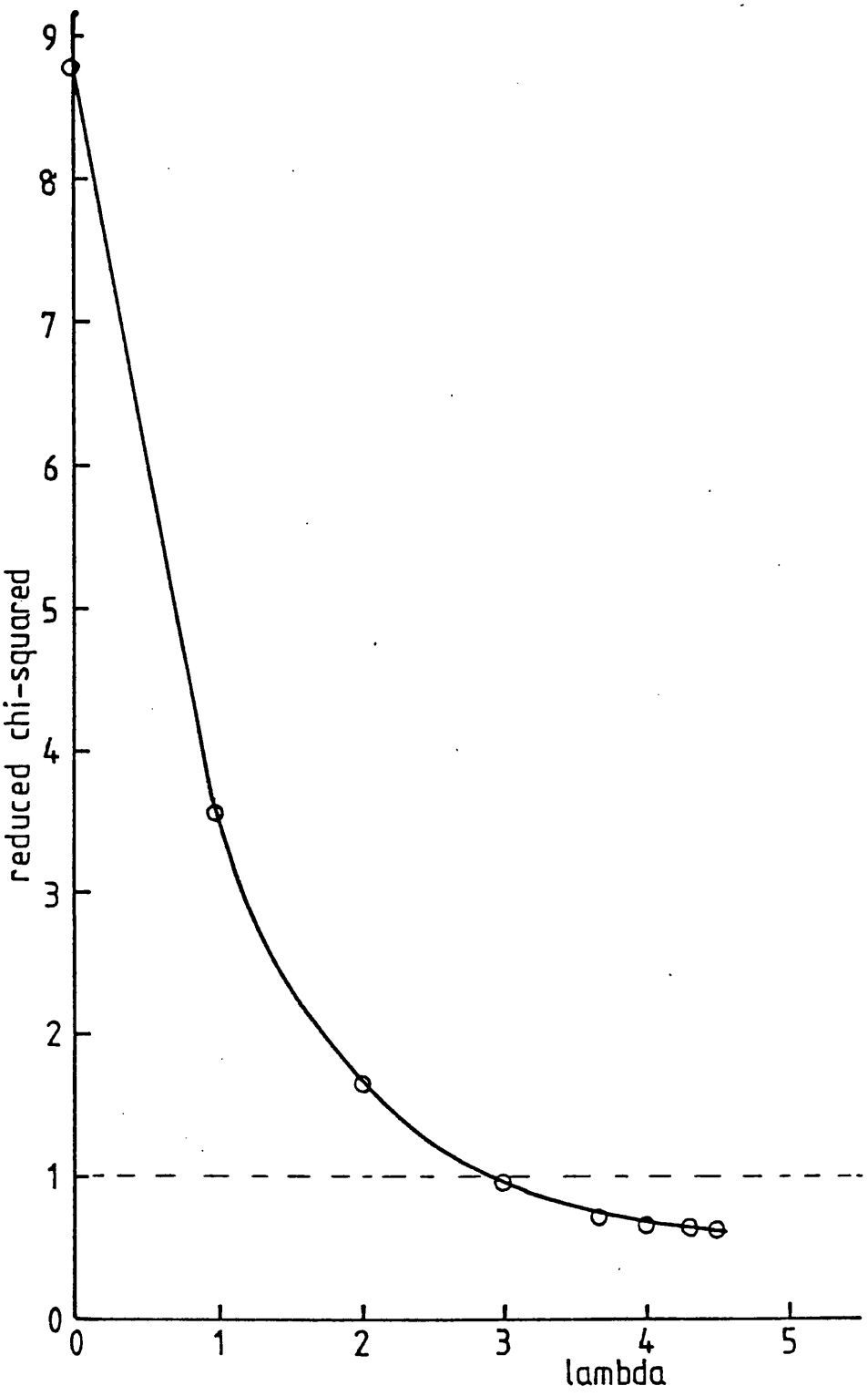


Figure 33. χ^2_y v λ from the Maximum Entropy algorithm applied to the Cygnus Loop data.

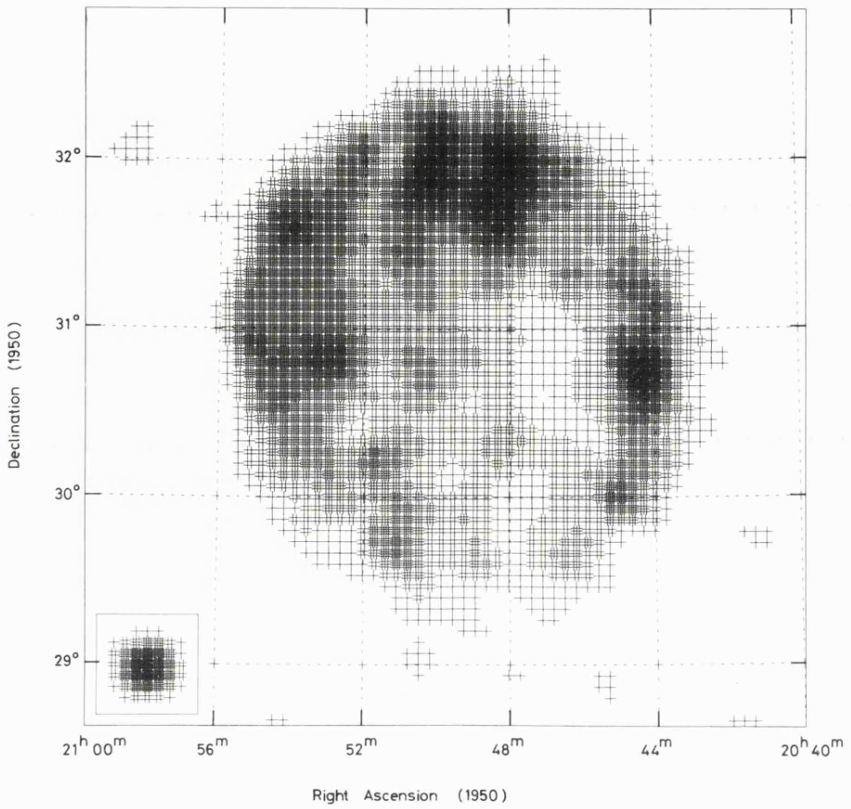


Figure 34a. The MEM reconstruction of the Cygnus Loop including exposure correction. Inset shows the top hat filtered response to Sco X-1.

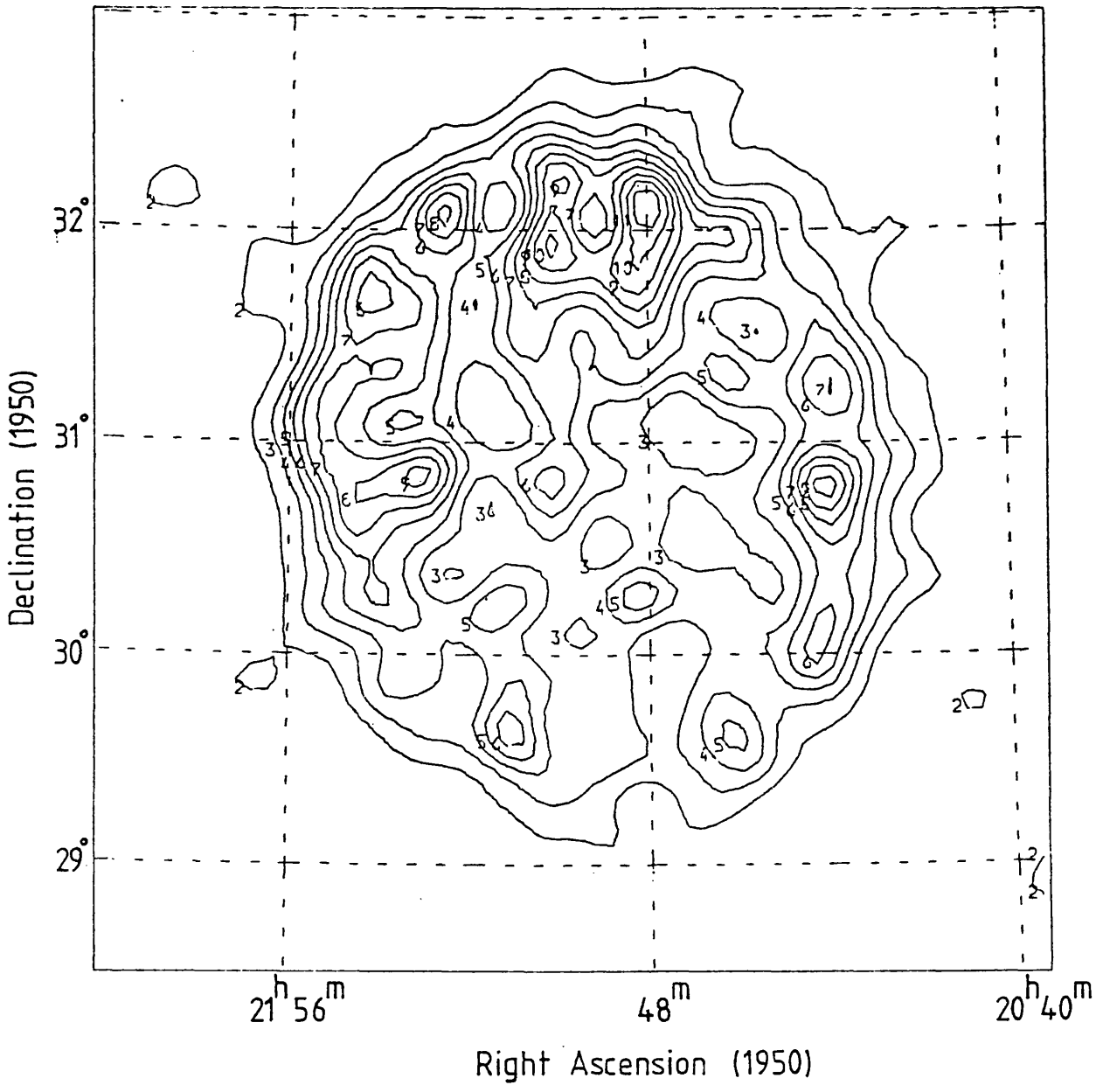


Figure 34b. Contour plot of the image in figure 34a.

Contour interval 1 count/4' x 4'.

data. This characteristic of χ^2 used in this way will be bourne out by other applications of the maximum entropy method later in this thesis (section 5.4). The parameter which determines the feature extraction achieved is the point response assumed. This must be chosen to be consistent with the observation if the method is to be successful. It should be noted that if the point response was very good and no blur was present, then the maximum entropy solution would be the data itself since it would indeed be the best estimate of the original photon distribution from the source! Although the MEM does consider the noise present, it is essentially a deconvolution procedure with good noise suppression characteristics rather than a noise filter. Its all-important feature is that it cannot produce features due entirely to counting statistics, which are more significant than is indicated by the count within the features. Unlike Fourier filtering, it is therefore 'safe'.

Figure 34 shows the maximum entropy solution to the Cygnus Loop data including an exposure correction. It is very similar to the Wiener filter solutions (figures 25, 30, 31) and exhibits good noise suppression when compared to figure 23.

The behaviour of the maximum entropy algorithm towards noise is obviously very important to the success of the method. In order to test the apparant noise immunity of the MEM solution indicated in the above discussion, a flat field observation was simulated using the same exposure and total count as the Cygnus Loop observation. The simulated data set is shown in figure 35.

This data was given to the algorithm and the solution after exposure correction is shown in figure 36. A reduced chi-squared of 0.8 was achieved with a completely flat field and $\lambda = 0$, but in order to test the noise immunity, six iterations were allowed to bring the reduced χ^2 down to 0.72 and some leak through of noise fluctuations is to be expected. The patchiness indicates a modulation in the solution, however the striking appearance of the patches is due to the choice of the grey scale, since the modulation only ranges from 1.5 to 2.1 counts and if a grey level of 1.5 counts had been used, a completely uniform display would have resulted. The algorithm clearly exhibits good noise immunity as was expected.

Although further work and experience is needed to clarify the detailed behaviour of the maximum entropy solution, especially with respect to χ^2, the method is clearly a safe one provided that the degradations and statistical nature of the observed data are well understood and calibrated.

3.7 Conclusion to Part I.

The research described in Chapters 1, 2 and 3 represents an attempt to apply and develop existing data processing techniques to the specific problem of producing good astronomical X-ray images from the raw event set provided by grazing incidence mirrors used in conjunction with position sensitive detectors. In order to achieve this, texts such as reference 5 had to be relied on heavily to provide the basic theory behind existing image processing methods. Software was developed to solve the

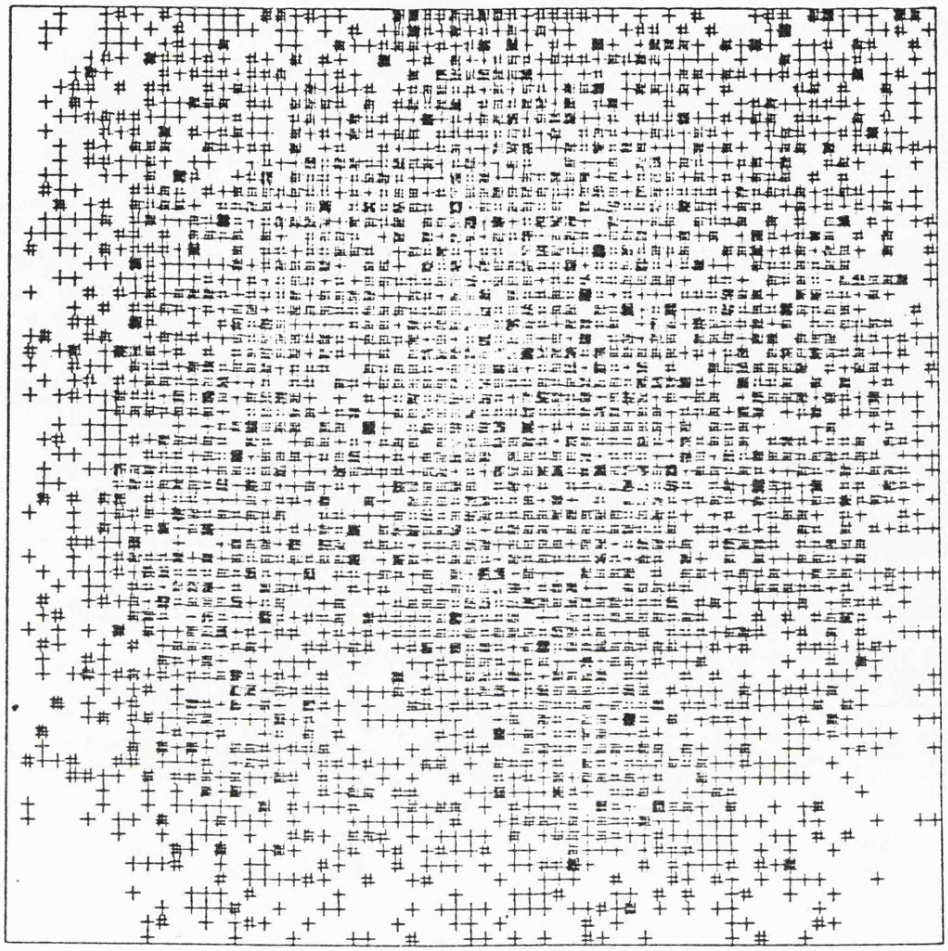


Figure 35. Simulated constant surface brightness observation using the exposure and count corresponding to figure 23.

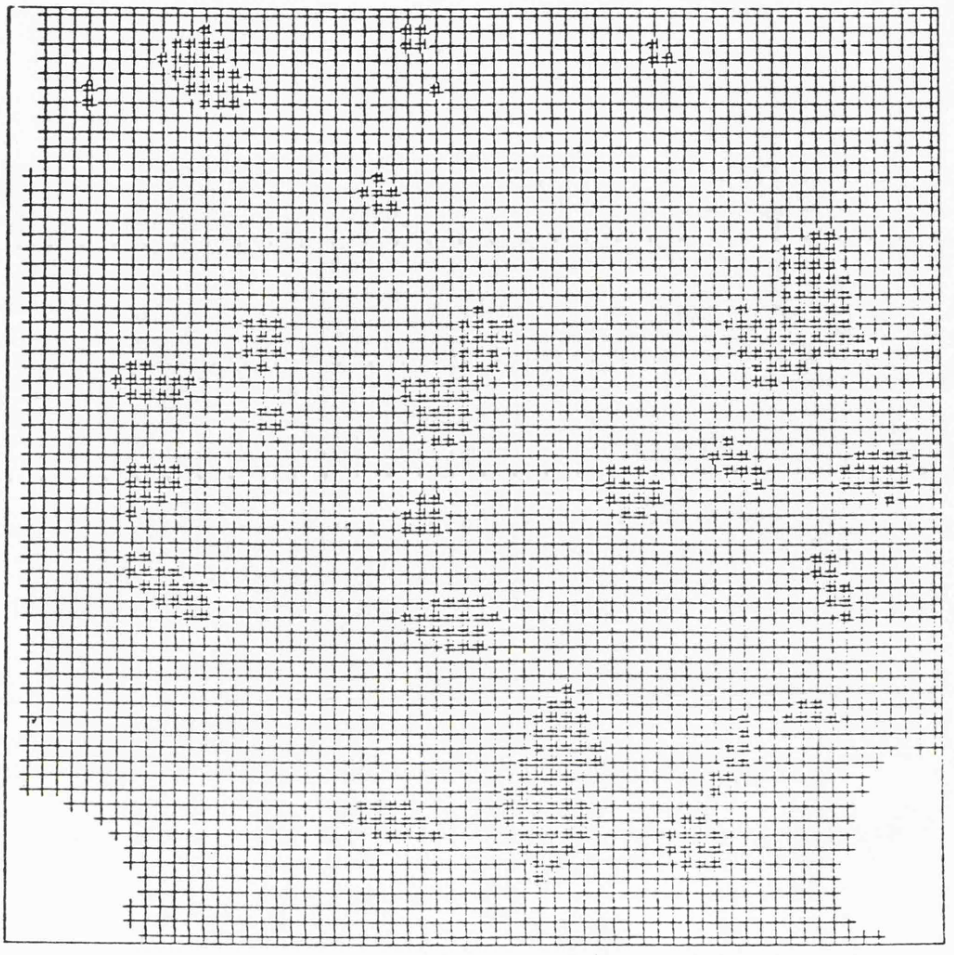


Figure 36. The MEM solution from the simulated constant surface brightness data. χ^2_ν = 0.72, minimum = 1.5, maximum = 2.1.

equations resulting from the theory, partly to investigate the various methods available but more importantly to apply them to real data. The Cygnus Loop data from the MIT/ Leicester rocket payload was an excellent test bed for the whole project and now the reflight data including images of Puppis A and IC443 super-nova remnants are being analysed using the techniques developed. It is hoped that future projects, such as Leicester's involvement in the HEAO-B data analysis, will benefit from and utilise the basic techniques tested in this research.

As with all research projects, a good deal of effort was expended on problems which cannot be included in the main body of the account. Therefore Appendix I provides documentation of the computer software developed and the methods of displaying data which were tried are demonstrated in figures throughout this thesis.

PART II

THE ANALYSIS OF SHADOWED X-RAY HOLOGRAMS

CHAPTER 4: DECONVOLUTION METHODS FOR CODED MASK X-RAY TELESCOPES.

4.1 Introduction to Part II.

Grazing incidence telescopes are limited by the upper cutoff of the reflection efficiency as illustrated by figures 5, 9, 20 and 21 in Part I. Above about 5 keV some other method of imaging X-rays must be used. Various devices using shadowing rather than reflection of X-rays have been proposed and developed, to try and provide an imaging instrument at energies > 5 keV which has the equivalent of the focusing advantage in signal to noise that grazing incidence telescopes have at lower energies.

The simplest device is a slat collimator placed in front of a detector limiting the solid angle which can be seen by the detector. Unfortunately the beam of the instrument must be scanned around the sky to produce an image and if high resolution is required, the slats must be made very deep and narrowly spaced. With no focusing advantage, the incoming signal is far more easily swamped by background counts in the detector.

If a position sensitive detector is available then the 'pin hole camera' principle can be used but again there is no focusing advantage to combat the detector background. The sensitivities of the slat collimator and pinhole camera are similar and only differ because of construction differences. The slat collimator loses by a factor of $\sqrt{2}$ because of the triangular response of the collimator. With a detecting area A cm^2 and observing time T secs, the sensitivity to which an N element image, with

each element Ω steradians, can be made with a detector background B counts/cm^2/sec is:

$$n_\sigma = \frac{3SAT\Omega}{N}\sqrt{\frac{N}{BAT}} \qquad (4.1)$$

where S is the source strength in photons/cm^2/sec/steradian in bandpass of detector and 3 is the photon detector efficiency.

The essential difference between the slat collimator and pinhole camera is their mode of operation. The slat collimator has been extensively used for sky surveying, in which the sky image is slowly built up after many scans. The pinhole camera however, can monitor a fixed area of sky continuously and could be useful for finding transient phenomena.

The sensitivity of a non-focusing instrument can only be improved in two ways. The detector background B must be reduced by technological development or the area of detector visible to a given sky element must be increased without increasing the fundamental pixel size Ω. It may be noted that a further complication arises when there is a diffuse component in the source distribution. The detection of a point source is then dependent on the pixel size Ω. In the diffuse background limited case:

$$n_\sigma = \frac{3SAT\Omega}{N}\sqrt{\frac{N}{DA\Omega T3}} \qquad (4.2)$$

where D is the diffuse background in photons/cm^2/sec/steradian. In case (4.1) an increase in the area of sky visible to a detector element can apparently give a

potential increase in sensitivity but in case (4.2) such an increase will have less effect because of the diffuse background component is also affected. Various methods have been proposed and tried to increase the area of sky visible to the detector while keeping the fundamental pixel size constant. They all consist of coding or multiplexing the incoming signal, either as a time series or a spatial pattern and any signal to noise advantage gained by such a procedure is nomally called a multiplex advantage (reference 16). Such a multiplex advantage is definitely different from the focusing advantages enjoyed by grazing incidence telescopes because it involves the use of a code.

The ground work for the following research was presented in reference 17, in which several multiplex methods were discussed and one chosen as being particularly promising and worthy of further investigation. Unfortunately the nomenclature of multiplex devices is still in turmoil with new variations being dreamt up all the time and the class of device discussed here has been called multiplex pinhole cameras, transform telescopes, shadow cameras, coded mask telescopes; the list is endless. However the underlying principle is the same in them all. The original idea stems from two independent papers by R.H. Dicke and I.G. Ables (1968), references 18 and 19. Subsequently other authors have provided developments and deeper understanding of the technique, notably T.M. Palmieri (reference 20).

4.2 The principle of the coded mask telescope.

The incident beam of photons is intercepted by a plane mask perforated with a pattern of holes. Photons which penetrate the holes are then detected by a planar position sensitive detector placed parallel and at a distance d below the mask. A point source at infinity will produce a shadow of mask pattern on the detector plane displaced from the centre by a distance proportional to the off-axis position of the source. When many sources are present, many such patterns will be shadowed onto the detector producing a 'hologram' or pattern characteristic of the source distribution. Figure 37 is a schematic diagram of the coded mask telescope. It is easy to show that the hologram has the form of a convolution:

$$h(x,y) \approx \int\int_{-\infty}^{\infty} f(\alpha,\beta)\ m(x+d\ \tan\alpha\ ,y+d\ \tan\beta\ )\ d\alpha\ d\beta \qquad (4.3)$$

where $f(\alpha,\beta)$ is the source brightness distribution and $m(x,y)$ is the mask pattern. Providing the mask is thin and there are no deep window support members on the detector, equation (4.3) is an accurate description of the shadow pattern present in the detector plane. Comparison with equation (1.1) reveals a close similarity between the form of the hologram distribution $h(x,y)$ and the focused image distribution, the only real difference being that the point response or instrument kernel is a pattern of holes instead of a single peaked response function. If only one small hole is present then the device is obviously a pinhole camera. The hologram distribution is sampled in exactly the same fashion as the focused image

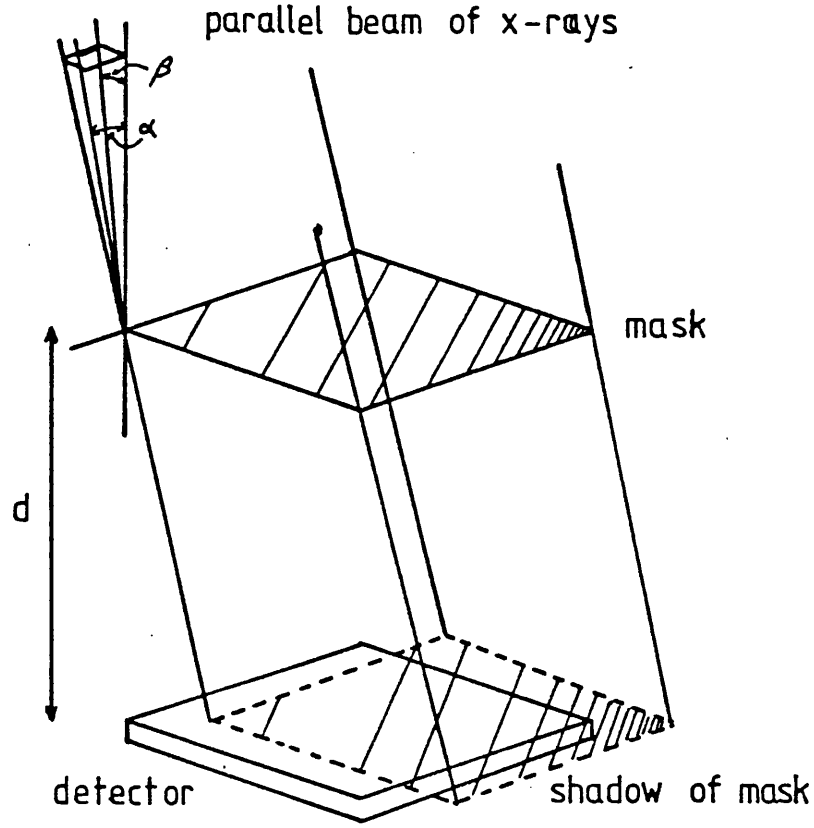


Figure 37. Schematic view of a coded mask telescope.

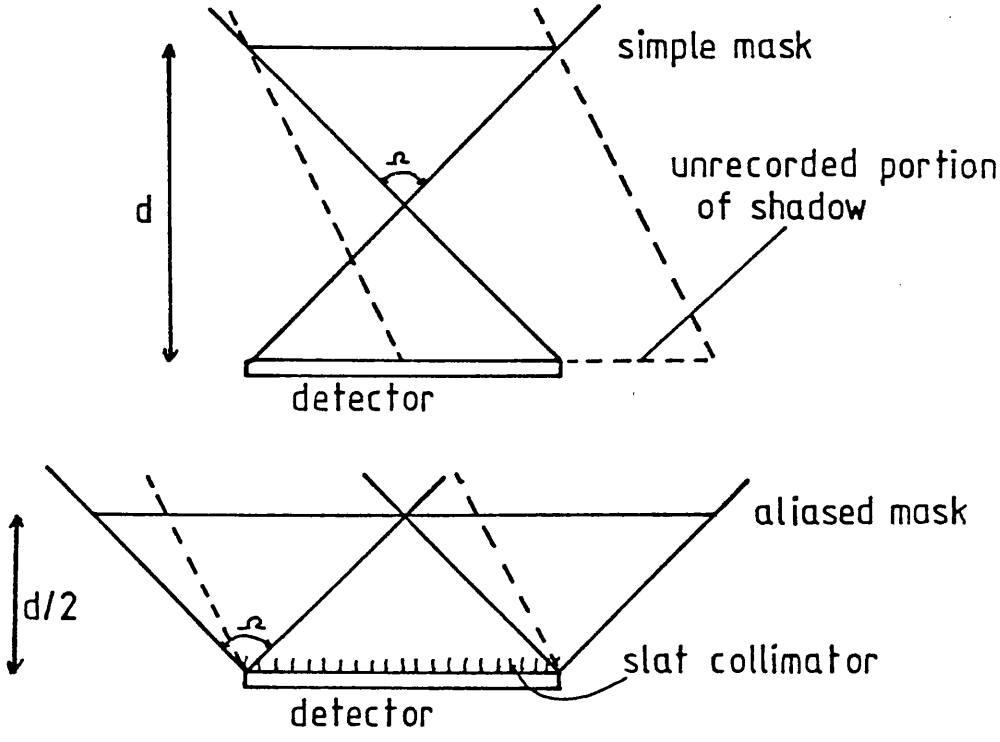


Figure 38. A comparison of the simple and aliased mask forms monitoring the same field of view Ω.

yielding an event set x_n', y_n', t_n', E_n'. The only difference in the recording processes is brought about by the fact that the photons are not focused into a cone but are travelling in the same direction as they were before reaching the instrument. The effect of window supports and obstructions will therefore be different, as will also the gas spreading or blur due to the finite absorption if a gas counter is used.

It is unnecessary to restate the analysis given in section 1.2, since the difficulty in coded mask imaging obviously lies in the form of the hologram given by equation (4.3) rather than in the minor degradations introduced by the detector. Whereas processing of focused event sets is designed to improve the estimate of the source distribution, the shadowed hologram is not a true image at all and it must be decoded before anything like a true image can be obtained. Since the coding or multiplexing process has the form of a convolution, the decoding can be called a deconvolution. The theory already presented in Chapter 2 can be used to provide decoding methods for coded mask telescopes and this research is centred about such methods. However before proceeding, the theory and design of coded mask telescopes needs to be considered in greater detail. The coding has the form of the integral equation (4.3) with imposed physical restrictions. The mask pattern is constructed of holes and therefore $m(x,y)$ only takes the values 0 or 1. Although the shadow distribution will be of infinite extent only a small portion can be recorded by the detector. The range of the (α,β) integration will depend on the extent of the mask

pattern or the form of the object distribution $f(\alpha,\beta)$. Decoding is a matter of solving the integral equation using a recorded sample of $h(x,y)$ and a detailed knowledge of the mask pattern $m(x,y)$, where $m(x,y)$ must be chosen to give a code which is as complete as possible.

If the object distribution and detector are assumed finite in extent, then the digital form of the transformation (4.3) is Toeplitz (see section 2.1). Unfortunately a Toeplitz matrix cannot be diagonalised and therefore perfect coding is not apparantly possible. However the closely related circulant matrix can be diagonalised by the DFT and careful design can provide a coded mask telescope with perfect coding. The Toeplitz form arises from the fact that the shadow of the mask pattern is bound to be cast over the edge of the detector by off-axis sources. The part of the shadow pattern not recorded is not therefore available for decoding the hologram. This lost information can be recorded by using a periodic or aliased mask pattern. The recorded pattern is then a circular shift of a single period. Using such an aliased mask converts the Toeplitz form into a circulant form and enables complete coding to be achieved. This has been suggested by many authors, notably references 21 and 22. Unfortunately there is one further physical restriction which must be allowed for. In practice, the object distribution $f(\alpha,\beta)$ is not finite and using a periodic mask leads to ambiguities, since the same pattern can be shadowed by at least two distinct sky elements. In order to pre vent this, a slat collimator must be used to restrict the field of view of each individual detector element. The

two possible coding systems based on the Toeplitz form with a non-periodic mask and the circulant form with a periodic mask are illustrated by figure 38. Obviously decoding by the circulant approximation is bound to be more successful in the periodic mask case but the necessary introduction of a slat collimator to avoid ambiguities reduces the sensitivity and therefore the multiplex advantage (MA).

There must be a trade off between sensitivity (MA) and coding and the balance is determined by the use to which the telescope is to be put. If the object field is sparse and coding deficiencies are not likely to be too troublesome, then the non-periodic or simple mask system will be preferred. Hence for detecting rare, transient sources in a large field of view with full multiplex advantage, the simple system should be used. In contrast, for imaging faint, relatively compact but nebulous sources, such as clusters of galaxies, the aliased system offers high resolution imaging without distortion and with a possible MA.

Having established the basic form of the coded mask, the detailed form of the pattern must now be considered.

4.3 The choice of mask pattern for coded mask telescopes.

If the coding of the hologram is to be unique, then the patterns formed by all possible object configurations must be different and distinguishable. The pattern must therefore be aperiodic (except for the aliasing to give a circular convolution). The circulant matrix describing the action of the instrument must have an inverse or pseudo-

inverse so that the hologram can be decoded. The system must have good signal to noise immunity, which in real space means that the chance of noise fluctuations emulating a point source must be as small as possible for all possible sources. This implies that in the DFT domain, all spatial frequencies of the object must be adequately represented in the hologram so that noise at poorly sampled frequencies will not dominate.

$$[h] = [A][f][B] + [N] \tag{4.4}$$

Equation (4.4) is the discrete representation of the hologram formation. $[A]$ and $[B]$ are the blur matrices representing the mask pattern and $[N]$ is a noise matrix. Using the stacked form of the object and hologram, this can be rewritten as:

$$h = [M]\ f + n \tag{4.5}$$

where $[M]$ is the block Toeplitz matrix corresponding to the mask pattern. Using the circulant approximation for $[M]$, equation (4.5) is diagonalised by the DFT $[W]$:

$$h = [W][\Lambda][W]^{-1}\ f + n \tag{4.6}$$

Substituting $[W]^{-1}\ f = \bar{f}$, $[W]^{-1}\ h = \bar{h}$ and $[W]^{-1}\ n = \bar{n}$ for the DFT's of the sampled source hologram and noise processes gives:

$$\bar{h} = [\Lambda]\ \bar{f} + \bar{n} \tag{4.7}$$

Since $[\Lambda]$ is a diagonal matrix, decoding is possible by a direct product operation in the DFT domain. However eqaution (4.7) also indicates what properties are

desirable in the mask pattern. The diagonal elements $[A]$ are formed from the elements of the DFT of the stacked mask pattern. Decoding at all spatial frequencies can only be achieved if all those diagonal elements are non-zero. Furthermore the presence of $\bar{n}$ limits the choice to one in which all the terms of $[A]\,\bar{f}$ will dominate $\bar{n}$ and hence the diagonal elements of $[A]$ must be as large as possible. The amplitude of the DFT of the mask pattern must therefore be non-zero and as large as possible at all spatial frequencies to give good noise immunity to the coding process. The real mask pattern is also limited to taking the values 0 or 1 (transmitting or opaque). In short, the mask must have a flat, extended spatial frequency power spectrum which, evoking the Wiener-Khinchine theorem, implies that the autocorrelation function of the mask pattern must be strongly peaked (the Wiener-Khinchine theorem states that the Fourier transform of the autocorrelation function equals the power spectrum (amplitude squared)).

The choice of mask pattern has been tackled by many authors, including references17, 20, 21 and 22. Using the aliased system in which the instrument performs a circular convolution, it is possible to construct a 'perfect' mask with ideal properties using pseudo-noise sequencies which are designed to be self orthogonal under circular cross-correlation. Both one and two dimensional masks can be constructed using such sequences or using the related Hadamard waveforms, see reference 22. Mask patterns which have the desired properties are random in the sense that they show no self correlation under circular shifts. Small sections taken in isolation do, in fact, look random but

often viewed as a whole they exhibit definite structure. Some examples of mask patterns are given in figure 39.

4.4 Decoding holograms from coded mask telescopes.

The mathematical construction of the hologram equations (4.4) and (4.5) is exactly the same as the linearly blurred image formulation (1.28). The same techniques described to solve the processing problem presented in Part I can be easily adapted to unravel the hologram data.

In the aliased mask pattern case, the use of a circulant is not a gross approximation since the only non-circular effects are relatively minor degradations introduced by the detector response and necessary physical support structures. Using an 'ideal' mask provides an easy way of decoding, simply circularly crosscorrelating the hohlgram pattern with the mask pattern (reference 21). This is so apart from a D.C. shift introduced by the 1,0 rather than 1,-1 character of the physical mask, the inverse of the instrument's circulant is in fact the transpose of the circulant itself and the circulant M is said to be orthogonal (real unitary), reference 13 section 1.15. Apart from the noise corruption, the hologram is therefore an orthogonal transformation of the object field and exact reconstruction is possible.

A far more severe decoding problem is offered by the simple mask configuration. The blurring now has a Toeplitz form which, unlike the grazing incidence case, is not well approximated by padding out with zeros to form a circulant because the blurring of any point can extend over all of the detecting area and 'edge effects' will in fact degrade

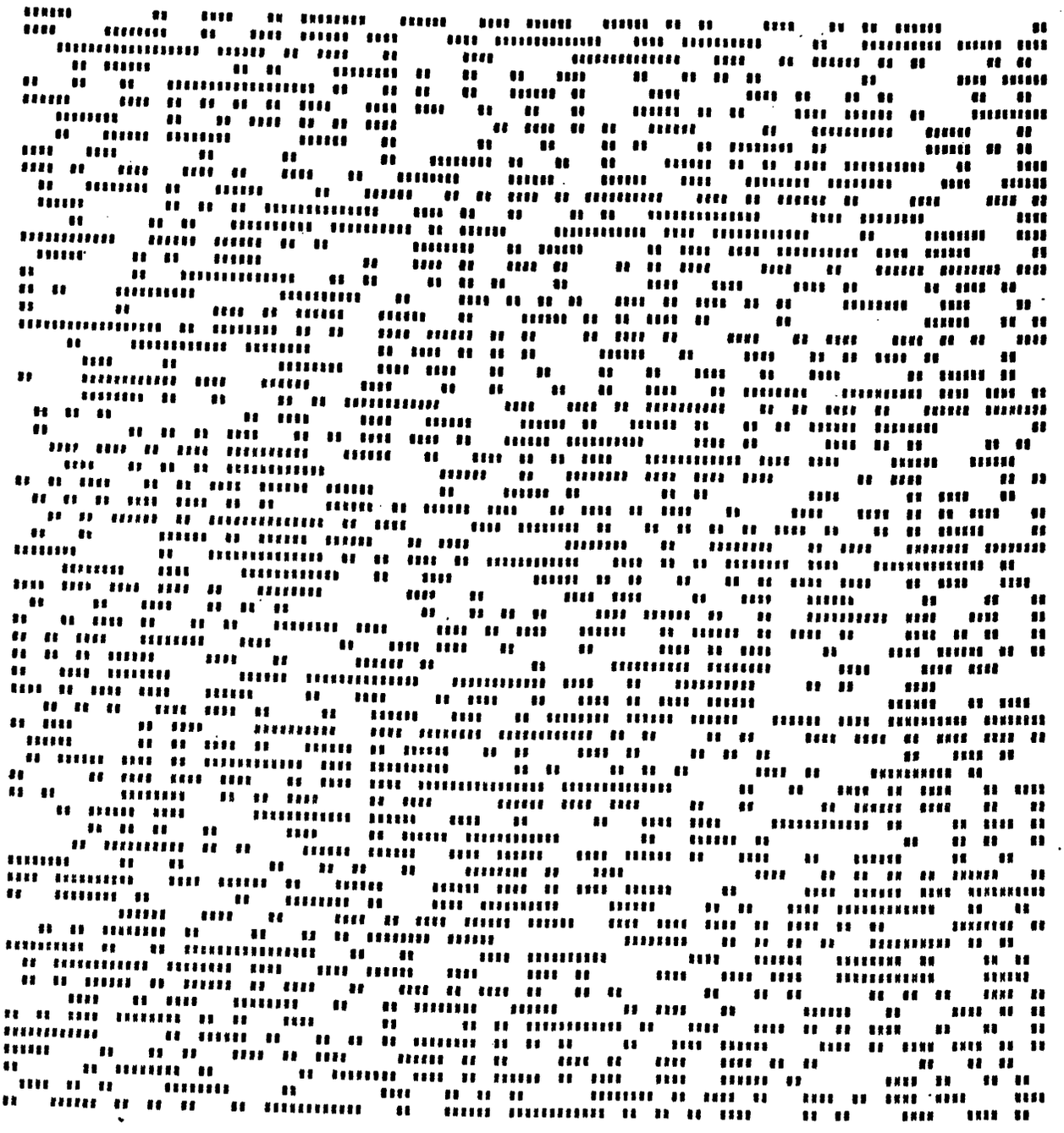


Figure 39. A random mask pattern, 64 x 64 elements, produced by a random number generator.

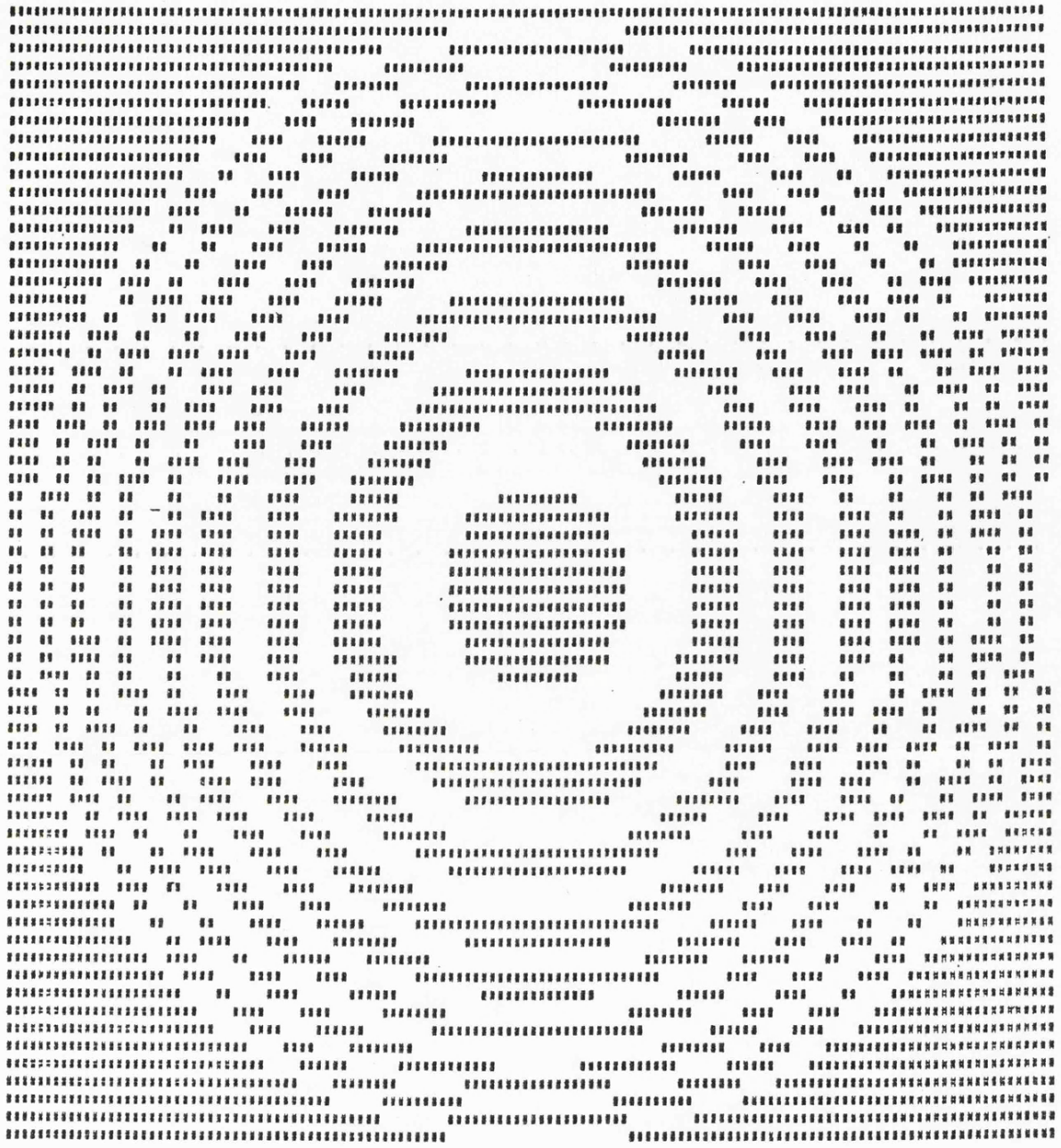


Figure 39 cont.. A Fresnel zone plate with 8 rings sampled on a 64 x 64 grid. Note the outer ring just touches the edge of the grid and is approximately one sample wide.

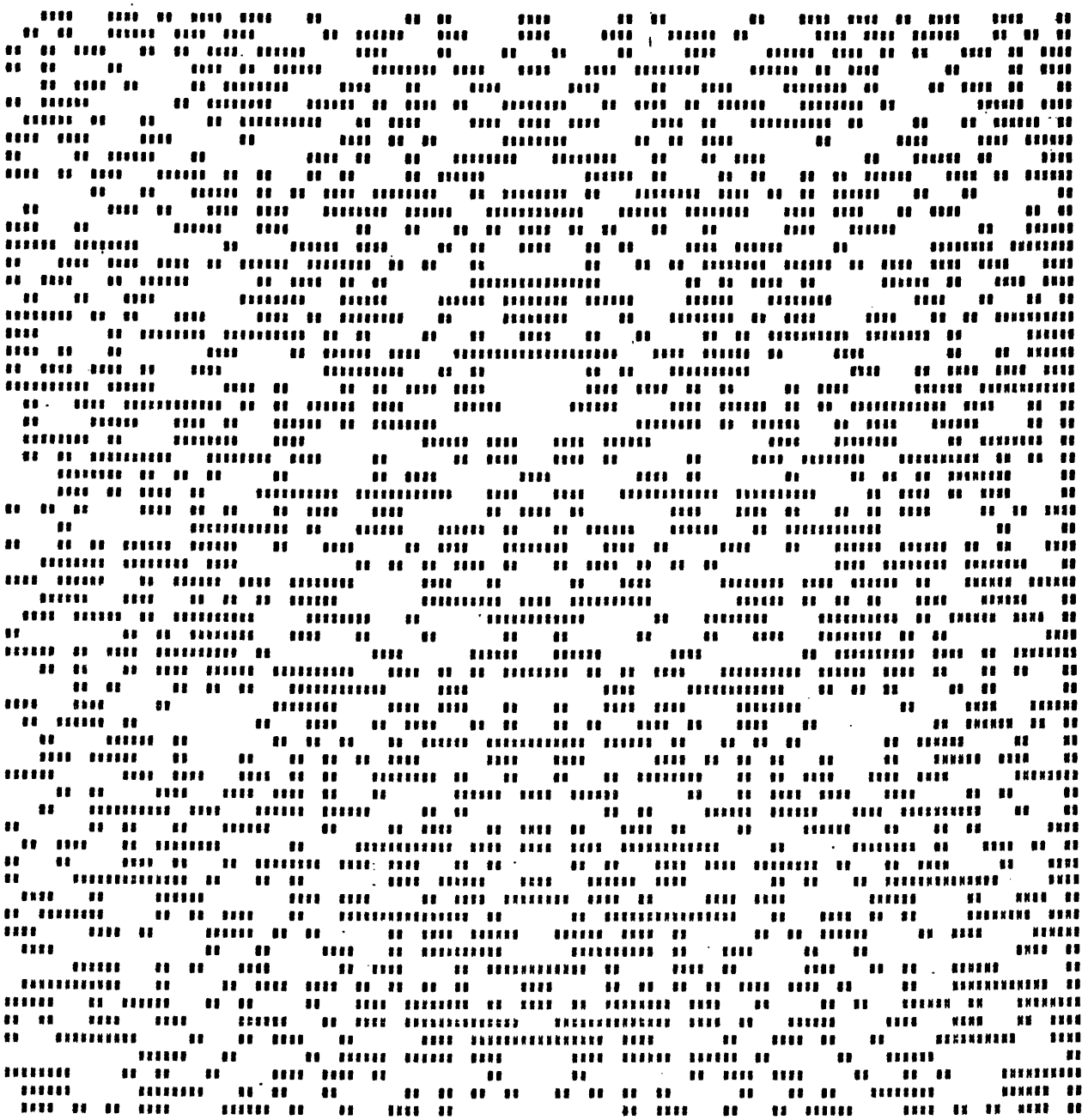


Figure 39 cont.. A pseudo-random mask, 65 x 63, generated using the algorithm given in reference 56.

the entire field of view. Even without noise corruption there is no unique solution to the deconvolution because the image coding or transformation performed by the instrument is incomplete.

Despite the coding deficiency, the simple mask type has a sensitivity advantage over and above the aliased type and providing the object field is not too complicated (i.e. sparse) this advantage will compensate for the coding errors. It is therefore worth studying methods of decoding in the simple mask case which minimise the decoding error or crosstalk which is inherent in the system.

Discrete filtering using a direct product filter in the DFT domain provides a potentially fast method for decoding. Using the circulant approximation, the hologram function can be expressed in the DFT domain as in equation (4.7):

$$\bar{h}_w = \Lambda_w \bar{f}_w + \bar{n}_w \qquad (4.8)$$

where Λ_w are the components of the DFT of the mask pattern and $\bar{h}_w$ is the DFT of the complete hologram pattern, including the portions lost beyond the boundaries of the detector. The instrument behaves as a spatial frequency filter with response Λ_w. If the mask pattern is well chosen, then $|\Lambda_w| \simeq C$ (a constant) and the inverse filter will be given by $Q_w = \Lambda_w^* / |\Lambda_w|^2$ where $|Q_w| \simeq 1/C$. The complex conjugation results in a transposition in the real domain and hence the real domain operation is simply a crosscorrelation by the mask pattern, apart from the normalisation factor. If the noise is assumed to be white, then $|\bar{n}_w|^2$ = a constant and the inverse filter cannot be

improved upon to give better signal to noise performance when the mask is ideal. However using the simple mask system, an ideal mask cannot be found and $|\Lambda_w|$ will not be constant. For spectral frequency samples w where $|\Lambda_w|$ is small, the noise $\bar{n}_w$ will dominate the signal and such frequencies can be suppressed by using a more sophisticated filter. The general form (2.24) can be utilised:

$$Q_w = \frac{\Lambda_w^*}{\Lambda_w \Lambda_w^* + z_w} \tag{4.9}$$

Equation (4.9) is the Wiener Filter constructed in the DFT domain, where z_w is the ratio of the power spectral densities of the noise and signal processes. Unfortunately it can be quite tricky to find z_w for the hologram data h. The difficulty lies in estimating the signal power spectral density function $|\bar{f}_w|^2$. In some applications such as imaging star fields, the assumption that the object field is 'white' is a reasonably good one and this can be used to estimate z_w:

$$z_w = z \approx \frac{\sum h_i \sum |\Lambda_w|^2}{\{\sum h_i^2 - \frac{[\sum h_i]^2}{N} - \sum h_i\} N} \tag{4.10}$$

Equation (4.10) gives an approximation to z (constant for all w) using the first and second moments of the hologram data. $\sum h_i$ is the total noise power due to counting statistics and $\{\ \}$ contains the estimated total signal power in the hologram. The factor $\sum|\Lambda_w|^2/N$ is the average mask spatial frequency filter weight used to correct the signal power in the hologram to a source power before

the filtering action of the mask takes place. This factor bears out the inherent 'coding disadvantage' of coded mask telescopes compared to, say, the pinhole camera. The angular source distribution suffers an approximately uniform spatial frequency attenuation $\sum|\Lambda_w|^2/N$ on passing through the mask and being recorded as a pattern by the detector. Unfortunately the noise does not suffer the same attenuation and hence if the observation is source photon limited there is no MA.

It must be remembered that discrete filtering as described above uses the linear circulant approximation to the Toeplitz form of the instrument response and this causes decoding errors over and above the statistical noise. In fact since the coding is incomplete and a unique solution does not exist, an optimum or most likely solution must be sought. The problem becomes more acute as the object field exhibits more structure and a decoding scheme which is more in sympathy with the Toeplitz nature of the response is desirable. The maximum entropy method offers just such an approach.

The application of the maximum entropy method to decoding holograms is exactly the same as that described for the deblurring of grazing incidence telescope images in section 2.7. The solution is given by:

$$\hat{f}_{\alpha'\beta'} = z_{\alpha'\beta'} \exp(-u) \exp\left\{-\lambda \sum_{\alpha\beta} A^t_{\alpha'\beta} \frac{(h_{\alpha\beta} - \hat{h}_{\alpha\beta})}{\sigma^2_{\alpha\beta}} B^t_{\alpha\beta'}\right\} \quad (4.11)$$

where:

$$[\hat{h}] = [A][\hat{f}][B] \quad (4.12)$$

[A] and [B] are the blur matrices representing the mask pattern response, [z] is the angular response or vignetting function of the instrument which will be dominated by the triangular response of the basic box system, $[\epsilon^2]$ is the variance matrix which will be dependant on the counting statistics of the hologram [h] and u and λ are constants which must be found to satisfy the constraints:

$$\sum_{\alpha\beta} \hat{h}_{\alpha\beta} = \sum_{\alpha\beta} h_{\alpha\beta} \qquad (4.13)$$

$$\sum_{\alpha\beta} \frac{(h_{\alpha\beta} - \hat{h}_{\alpha\beta})^2}{\epsilon_{\alpha\beta}^2} = \text{a minimum} \qquad (4.14)$$

The algorithm used for finding the maximum entropy solution to the blurred image problem described in section 3.6 can be directly applied to solving the decoding problem presented here.

4.5 The multiplex advantage of coded mask telescopes.

It is pertinent at this juncture to inquire whether or not coded mask telescopes really do provide a multiplex advantage over slat collimators or pinhole cameras as described by equations (4.1) and (4.2). Unfortunately the significance of features in the decoded image will depend on the deconvolution method used, but since the ultimate sensitivity is set by the statistics of the hologram rather than the deconvolution, it is only necessary to consider one decoding method to reveal the essential nature of the limiting sensitivity. The easiest case to analyse is the aliased system for which an ideal mask

pattern exists and which can be deconvolved by direct crosscorrelation with the mask pattern. Gunson and Polychronopalus deal with this in reference 21 and the following is based on their treatment.

When the hologram is crosscorrelated with the mask pattern at a position corresponding to a source, all source counts will contribute to the peak and a fraction of approximately r of all other counts will contribute, where r is the mask transmission. At positions off a source only r of the source counts will contribute plus again r of all other counts. Using the symbols introduced in section 4.1 we have:

$$\text{Peak strength } P_s = r_3 SAT_\Omega + r^2_3 D_\Omega AT + rBAT \tag{4.15}$$

$$\text{Plateau strength } P_p = r^2_3 SAT_\Omega + r^2_3 D_\Omega AT + rBAT \tag{4.16}$$

D includes all sources except the source under inspection. The significance of a mask P_s is governed by fluctuations in the background plateau P_p. The total background count $r_3 D + B$ has a probability r of contributing and 1-r of not contributing and the distribution has the binomial form, the variance of which is given by:

$$\sigma^2 = r(1 - r)(r_3 SAT_\Omega + r_3 D_\Omega AT + BAT) \tag{4.17}$$

The signal to noise of a peak in terms of standard deviations above background is given by:

$$n_\sigma = \frac{r_3 SAT_\Omega + r^2_3 DAT_\Omega + rBAT - (r^2_3 SAT_\Omega + r^2_3 DAT_\Omega + rBAT)}{[r(1 - r)(r_3 SAT_\Omega + r_3 DAT_\Omega + BAT)]^{\frac{1}{2}}} \tag{4.18}$$

For the case of $r = \frac{1}{2}$, this simplifies to give:

$$n_{\epsilon} = \frac{\frac{1}{2}\zeta SAT\Omega}{\left[\frac{1}{2}\zeta SAT\Omega + \frac{1}{2}\zeta DAT\Omega + BAT\right]^{\frac{1}{2}}} \qquad (4.19)$$

Expression (4.19) is simply the ratio of the total count detected from a source to the square root of the total recorded count. The device will in general have a multiplex advantage over a pinhole camera since if there are N holes in the mask, the total source count will be increased by N. However since no decoding is necessary for the pinhole camera the background per pixel will be reduced by a factor M, the number of pixels in the detecting area A. For an aliased system with $r = \frac{1}{2}$, $M = 2N$. The diffuse background contribution will also be increased by a factor N in the multiplexed case. Taking two extremes,
a) the detector background limited case $MA = \sqrt{N/2}$ and
b) the diffuse background limited case $MA = 1$.
The factor of $\sqrt{2}$ arises because of the statistical correlation inherent in the crosscorrelation process and the rather surprising diffuse background limited result arises because the total diffuse count is reduced by a factor N for the single pinhole and no crosscorrelation is necessary for the pinhole camera. Although there is no MA for the diffuse background limited case, the multiplexing intsrument has a far greater chance of detecting photons from a faint source and this gives it a 'photon limit advantage' in practice as will be seen later.

The significance of a peak is also dependent on the number of pixels in the field of view since the greater the number of possible source positions, the greater the

likelihood of a false peak occurring due to statistical fluctuations. This was first analysed by Palmieri, reference 20. If $P(n_\sigma)$ is the probability that a fluctuation exceeds a significance n_σ, then:

$$[1 - P(n_\sigma)]^M \simeq 1-MP(n_\sigma); \quad nP(n_\sigma) << 1 \qquad (4.20)$$

For 99% confidence, $NP(n_\sigma) \simeq 10^{-2}$ and using the normal approximation to the Binomial distribution, when the total count is large, the significance levels for different values of M can be found from statistical tables, for example:

$M = 512 \qquad n_\sigma = 4.2$

$M = (512)^2 \qquad n_\sigma = 5.5$

The significance is clearly a very weak function of M.

Formula (4.19) above must be modified to allow for any slat collimator included to prevent ambiguities and this will reduce the sensitivity to off-axis sources. (4.19) has been derived for an aliased mask system in which the mask pattern is superimposed on the background count with a cyclic shift for off-axis sources. The simple mask, on the other hand, shadows the pattern with a linear shift and as the source moves towards the edge of the field of view, the proportion of the pattern recorded reduces and hence the total count recorded drops. If the two configurations, simple and aliased, use the same detecting area and the mask-detector distance is made half as large for the aliased case, the field of view and sensitivity in terms of source counts received will be identical for both types (see figure 38). However the pattern cast by the simple mask will only cover part of

the total hologram for off-axis sources and therefore these sources will be competing against smaller background fluctuations in the simple mask configuration. Equation (4.19) will describe the on-axis sensitivity and on moving off-axis, the total count in the denominator will have to be modified to include only counts covered by the mask pattern.

Since the detector-mask distance is smaller for the aliased system, looking at the same field of view, the angular resolution of the aliased system (given by $\Delta x/d$ with Δx as the detector resolution and d as the detector-mask distance) will be twice that of the simple system. Providing the coding errors introduced by the simple system can be suppressed, the inherent sensitivity and resolution advantage of the simple system will be realised. However this is only likely to be possible when viewing sparse fields of point sources rather than large, nebulous structures.

4.6 Computer simulations of coded mask telescopes.

There are many practical applications of the coded mask telescope principle but they divide quite strongly into two classes. Firstly, hard X-ray imaging devices for studying compact clusters or nebulous sources, providing pictures at hard energies inaccessible to grazing incidence mirrors. The diffuse X-ray background starts to dominate over detector background when each element in the detector can see greater than $\sim$10 square degrees, so that using a field of view comparable to a grazing incidence mirror will give observations which are either background

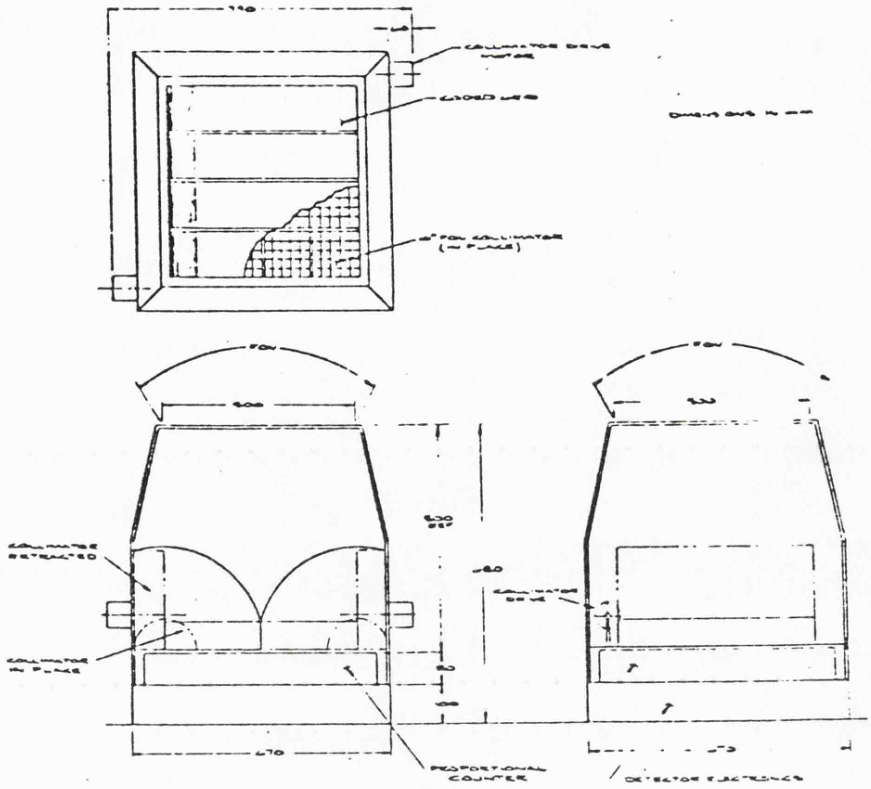
or source photon limited. Since an image of a possibly extended source is wanted, it is best to opt for an aliased system in which coding error can be eliminated. Such a device is described in reference 21, including a one-dimensional simulation. The second class of instrument is burst and transient monitors in which the multiplexing is used to provide continuous coverage of a very large area of the sky with high position resolution and sensitivity. The simulations to be described here come into this category and were carried out because the simplified analysis presented above is unable to assess the true performance of such instruments.

The simple non-aliased system is probably better for a burst monitor because it provides about four times the sky coverage at the same sensitivity and angular resolution. This is so because a collimator is not needed to avoid ambiguities. Unfortunately the M sky resolution elements observed are only represented by M/4 detector bins and so the coding is underdetermined by a factor of 4. However providing the field is sparse with the number of sources $T \ll M$, there should be sufficient hologram information to adequately reconstruct the source field. It has already been pointed out in section 4.4 that the choice of mask pattern, decoding method and statistical behaviour are clearly specified for an aliased mask system but the same is not true for the simple mask system. A mask which is ideally suited to the aliased system does not necessarily work adequately for the burst monitor application because vital information necessary for even rough decoding may be lost off the detector's edge. Neither

crosscorrelation nor more sophisticated Fourier filtering allow for the Toeplitz form of the hologram and some more suitable decoding procedure may be necessary to give adequate reconstruction.

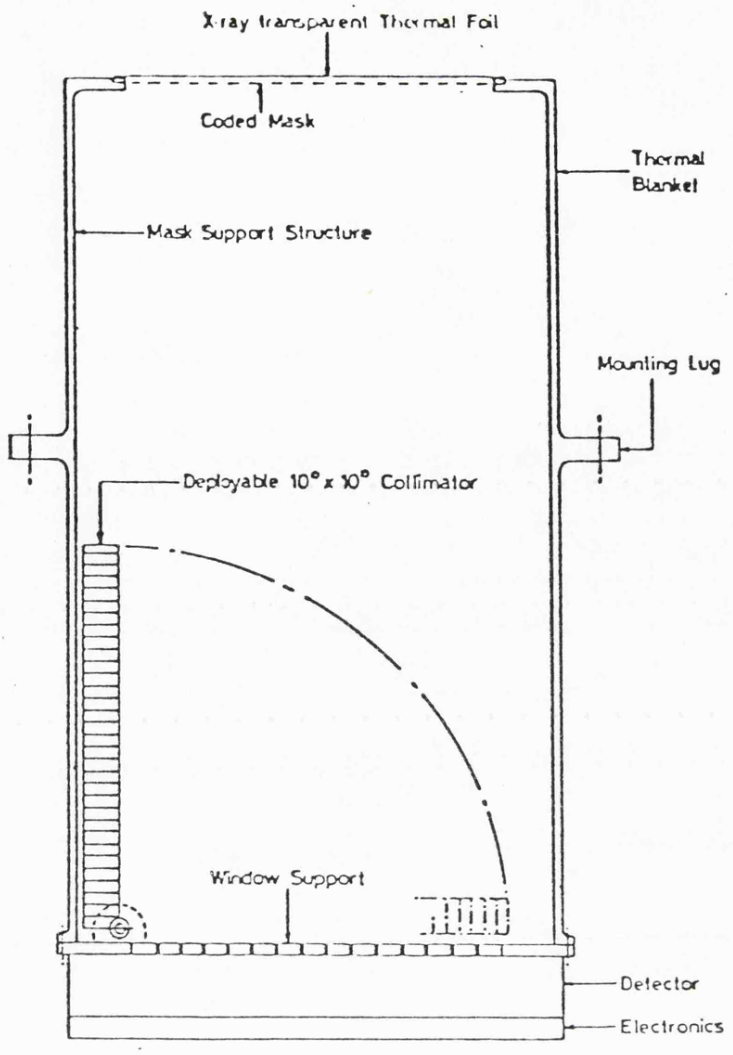
Two instruments proposed to NASA, reference 23, were simulated to test their expected capabilities. The first was a crossed pair of one-dimensional simple coded mask cameras designed to provide positions accurate to a few arc minutes of X-ray bursts and transients. The use of two 1-D systems made the construction of a working detector easier and reduced the telemetry bandwidth necessary for transmitting back the data. Figure 40 provides a schematic design and instrument characteristics for this proposal. The second instrument was more sophisticated, using a 2-D detector and mask arrangement as originally proposed by Dicke (1968), reference 18. The potential power of this instrument lies in the possibility of surveying relatively complicated areas of sky as well as monitoring bursts and transients. Figure 41 is a diagram of one module accompanied by the relevant instrument parameters. A full description of the capabilities and characteristics of both proposed instruments is presented in reference 23, but very little of that data concerns this research. The object of simulating the performance was to check that a reasonable mask pattern could be found, that the statistical analysis was correct and to see whether the crosscorrelation decoding method was good enough.

The 1-D system was simulated in the wide field mode without the 10^{o}x 10^{o} collimator deployed. Window support



Configuration	Four 1/4 square meter Shadow Cameras
Combined FOV	45° x 110° FWHM
Effective Area per pair of Cameras	
Positioning	1000 cm^2
Timing (X and Y data summed)	2000 cm^2
Energy Range	
Positioning	2-20 keV, 4 channels
Timing	2-60 keV, 5 channels (250 μsec) 16 channels (0.1 sec)
Time Resolution	
Positioning	0.1 sec
Timing	250 μsec
Angular Resolving Power	3 arc min.
Error Box Area (below 8 keV)	
Burst 1 CFU for 1 sec	On Axis 0.3 arc min^2, at 23° 1.2 arc min^2
Burst 1 CFU for 10 sec	On Axis 0.1 arc min^2, at 23° 0.4 arc min^2
Source at limit of sensitivity (0.003 CFU for 6000 sec)	On Axis 10 arc min^2
Sensitivity: (detectable source intensity at 6 standard deviations)	
0.1 sec duration	On Axis 0.3 CFU, at 23° 0.6 CFU (with collimator, 0.05 CFU)
100 sec duration	On Axis 0.01 CFU, at 23° 0.02 CFU (with collimator, 0.005 CFU)
6000 sec (1 orbit)	On Axis 0.003 CFU, at 23° 0.006 CFU (with collimator, 5 10^{-4} CFU)
Count Rates	
1 CFU = Crab Nebula Flux Unit	3100 counts sec^{-1} per detector
Particle Background (Escaping Guard)	56 counts sec^{-1} per detector
Diffuse Background	6120 counts sec^{-1} per detector
(With 10° x 10° Collimator)	207 counts sec^{-1} per detector
Total Count (including Steady Sources) when Observing Galactic Center	15400 counts sec^{-1} per detector
(With 10° x 10° Collimator)	1830 counts sec^{-1} per detector

Figure 40. Proposed 1-D instrument, reference 23.



MODULE CHARACTERISTICS

Size	40cm x 40cm x 80cm
Mass	29kg
Full Field of View (FWHM)	22.5° x 22.5°
Field of View with Deployed Collimators	10° x 10°
Mask Area	30cm x 30cm
Mask-Detector Window Separation	72.5cm
PSPC Effective Area	380cm^2

DETECTOR CHARACTERISTICS

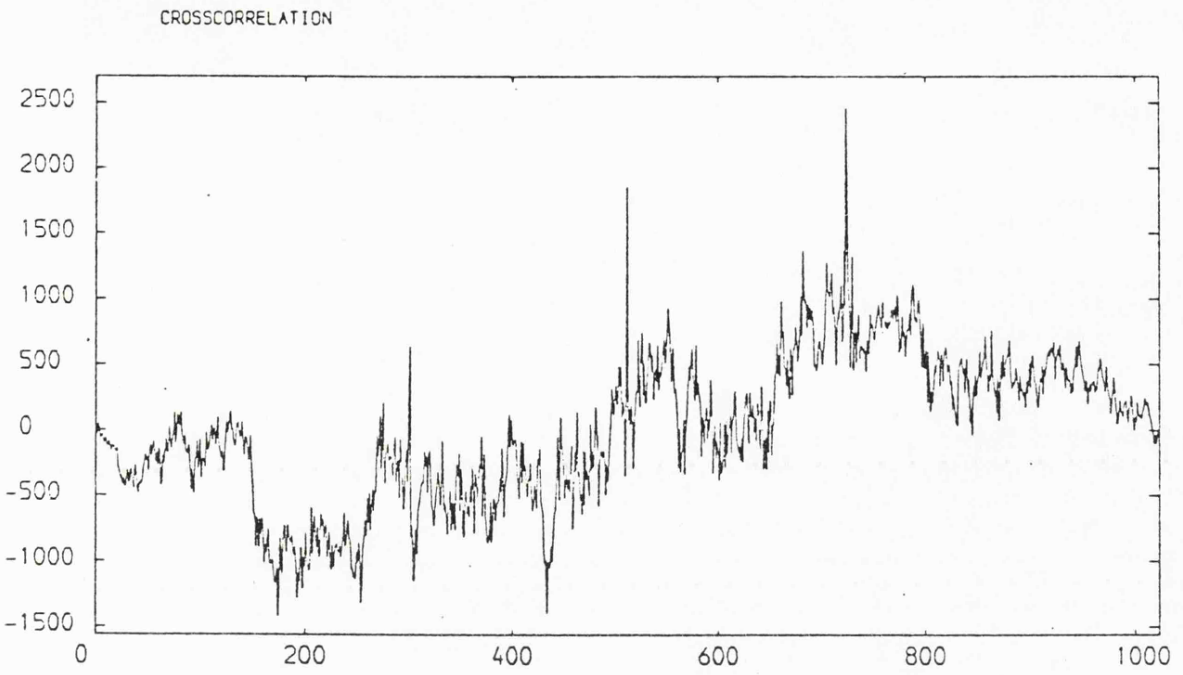
Window size			30cm x 30cm
Window Material			Beryllium
Gas Mixture			Xenon-methane (90:10)
Gas Pressure			2 atmospheres
Energy Resolution			20% FWHM at 5.9 keV
Spatial Resolution			FWHM = 0.6cm
	Anode	Cathodes	Anti-Co Anode
Wire Pitch (cm)	1.0	1.0	10.0
Wire Diameter (μm)	20	50-125	20
Interelectrode Gaps:			
Window-Cathode			15.0cm
Cathode-Anode			5.0cm
Guard Cell			15.0cm

Particle background 0.01 cts/cm^2/sec.

Sky background 8.8 cts/cm^2/sec/ster

Figure 41. Proposed 2-D instrument, reference 23.

structure, including off-axis shadowing, was included so that all major expected sources of coding errors were present. The data was in fact simulated using a program written by M. Sims of the X-ray Astronomy Group at Leicester University. The simulated data presented here represents three Crab-like sources in the field of view integrated for 1 sec with the appropriate detector and including diffuse sky background. The mask used was 511 elements long, generated using a pseudo-noise sequence. The data was processed in three ways; crosscorrelation with the mask pattern replacing zeros by -1's to remove the overall triangular response, using a Wiener filter constructed using a white noise and signal estimate taken directly from the data as described above and finally using a maximum entropy algorithm exactly the same as that described in section 3.6. All convolution sums were calculated using the FFT algorithm and counting statistics were simulated using a random number generator. The resulting estimates of the original source vector are plotted in figure 42. Two of the three sources are clearly visible in the crosscorrelation estimate but the third is rather weak and the background is strongly modulated by small spatial freqency terms. A statistical analysis of the background fluctuations reveals that the significance of the peaks is approximately half that indicated by equation (4.19). This is probably due to the gross variation in the average background introduced by the loss of pattern off the edge of the detector. The Wiener filter result shows better average control but no marked improvement in the significance of the three peaks. Unlike the



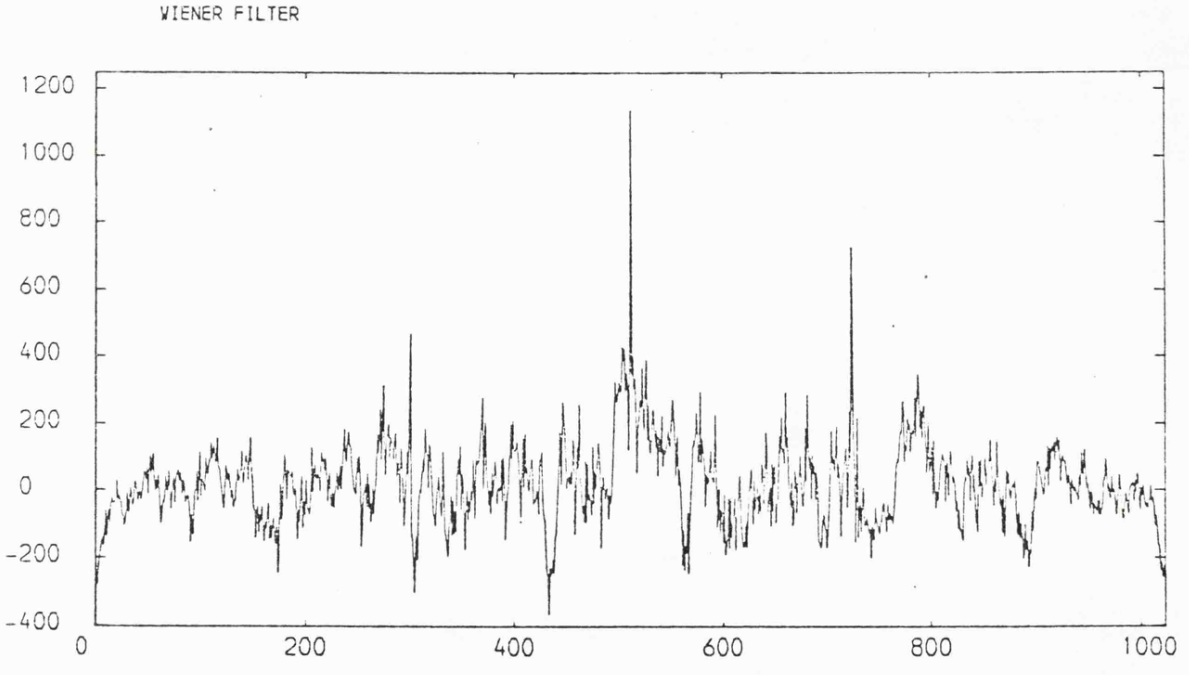


Figure 42. The results of the simulations of the proposed 1-D instrument. The mask used was a pseudo-noise sequence. The total count recorded was 20830. Detector window support structure was included. The Wiener filter produced a slightly more controlled background but no marked signal to noise advantage.

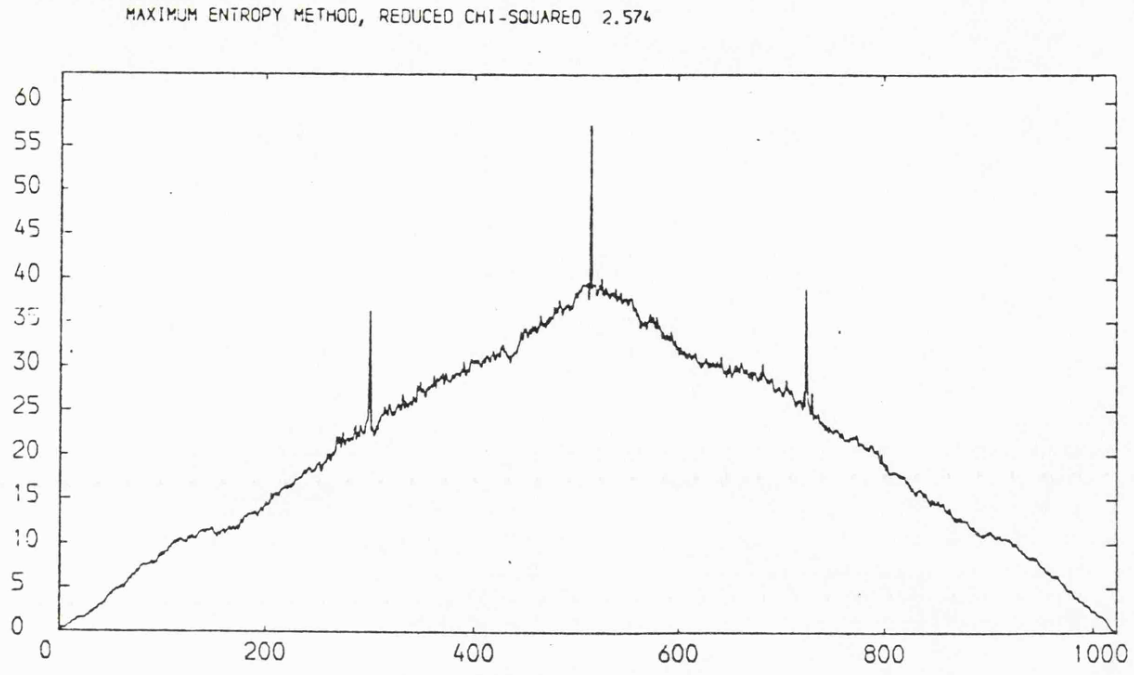


Figure 42 cont.. The triangular response to the background count is clearly visible in the MEM reconstruction.

Note. The plots show estimated recorded counts and have not been corrected for the off-axis response function.

crosscorrelation result, the total sum of all elements equals the actual count. The MEM reconstruction is that achieved after 5 iterations. Unfortunately the reduced χ^2 is rather large at 2.57 but the solution had converged to a large degree after 5 passes and a significant reduction in χ^2 would have required a large number of extra passes. Despite the large χ^2, the three peaks are very clear, standing out of a triangular background response. The total CYBER 72 CPU time required for all three methods was only 23 seconds, including plotting the results.

These rather simple simulations demonstrate that the 1-D system does indeed work at a sensitivity comparable (within a factor of 2) with the predicted performance. All three deconvolution methods were successful, crosscorrelation and Wiener filtering producing fairly similar reconstructions and the iterative MEM providing a significant inprovement in image quality. Further analysis of the 1-D system could have been made but the greater potential and aesthetic appeal of the 2-D proposal was too large and so all efforts were expended on simulating the 2-D system using the image processing routines originally written for Part I and documented in Appendix I.

The two-dimensional instrument proposed by Pounds et al (1978), reference 23, was simulated with a reduced number of detector elements so that it could be accomodated easily on existing facillities. All other parameters were used as given in figure 41. Two masks were tried, the pseudo-noise sequence mask as described by Procter et al (1978), reference 56, and a Fresnel zone plate, The scale of the zone plate was chosen so that the outer ring

was approximately the same width as a single sampling bin. It was shown in reference 17 that for a 63 x 63 matrix, a Fresnel zone plate containing 8 rings with the outer cicumference just touching the edges of the array gave a good spatial frequency response. In fact, a N x N (N = odd) matrix sampling at $\Delta s = 1$ and using the zeros of cos (kr) to generate the zone plate, with r as the radius vector and k as a scaling factor, will provide an optimum pattern if:

$$k = \frac{2\pi}{2N - 3} \tag{4.21}$$

This will give a zone plate with p zero crossings given by:

$$p = \tfrac{1}{2}\left[1 + \frac{(N - 1)^2}{(2N - 3)}\right] \tag{4.22}$$

The window support structure and minor detector aberrations were not included in the instrument response in order to simplify the calculations. These were considered to be minor problems in comparison with the major coding error.

Existing X-ray source catalogues were used to provide source field distributions for testing the instrument performance. Holograms were simulated including all detector and sky background using a random number generator. All three deconvolution methods described above were used to produce source field reconstructions from the holograms. These solutions were then subjected to a global statistical analysis to find the rms of the background fluctuations. Obvious large peaks corresponding to sources were not

included in the analysis to prevent biasing the results. Peaks in the images were then assigned significance values using the rms estimations. An aliased system using the same detector and looking at the same field of view was simulated with Procter et al's mask pattern so that the performance of the simple and aliased configurations could be compared. The results of some of the simulations are summarised in figure 50. Other source distributions from the Sco. region were tried and they gave comparable results, indicating that the two sets presented are not biased by the distributions chosen. Different statistical samples of the same observation were also taken along with slightly different pointing directions to further check that the results presented were truely an unbiased representation of the performance. These tests clearly demonstrated that the figures given in figure 50 were unbiased and not freak results.

Comparison of simulations 1 and 2 demonstrates that the pseudo-noise sequence mask gives a better signal to noise performance than the Fresnel zone plate. However the difference is not large and use of the Wiener filter and MEM improves the Fresnel mask performance as shown by figures 44 and 45. The Fresnel mask image also exhibits slight spreading of point sources but the associated loss in resolution is small. Crosscorrelation in simulation 2 produced an rms value considerably larger than the theoretical value at the centre of the field of view. The increase was due to coding errors. This is clearly demonstrated by comparison of simulations 4 and 6 since all fluctuations in 6 are due to coding errors. The MEM

reconstruction for simulations 1 and 2 largely accounted for these errors and greatly improved the sensitivity of the observation. This result for simulation 2 is dramatically illustrated by figure 46 in which all the visible peaks correspond to real sources. Some real peaks of high significance found by the statistical analysis are too small to be seen on the diagram! Three false peaks of > 5 rms are also present but not visible and are probably present because of the rather large χ^2_ν value. Unfortunately a very large amount of computer time would have been required to suppress these peaks further. The stability of this solution against statistical fluctuations was tested using a different statistical sample of the hologram. The solution was practically identical. A more rigorous test could not be carried out because of the very large amount of CPU time required, so the true significance of peaks in the solution could not be found. Simulation 3 is an aliased system simulation for comparison, in which the rms value agrees fairly well with theoretical prediction and in which, as shown in figure 47, the background fluctuations have a constant amplitude over the entire field of view, as expected.

The Coma region simulations provide a test in which the instrument is severely sky background limited. There was no multiplex advantage at all but a large photon detection limit advantage because of the large detecting area presented to a large part of the field of view. The deconvolution methods were able to pull out two sources from an enormous background count. The rms value achieved by crosscorrelation in the simple mask case was much lower

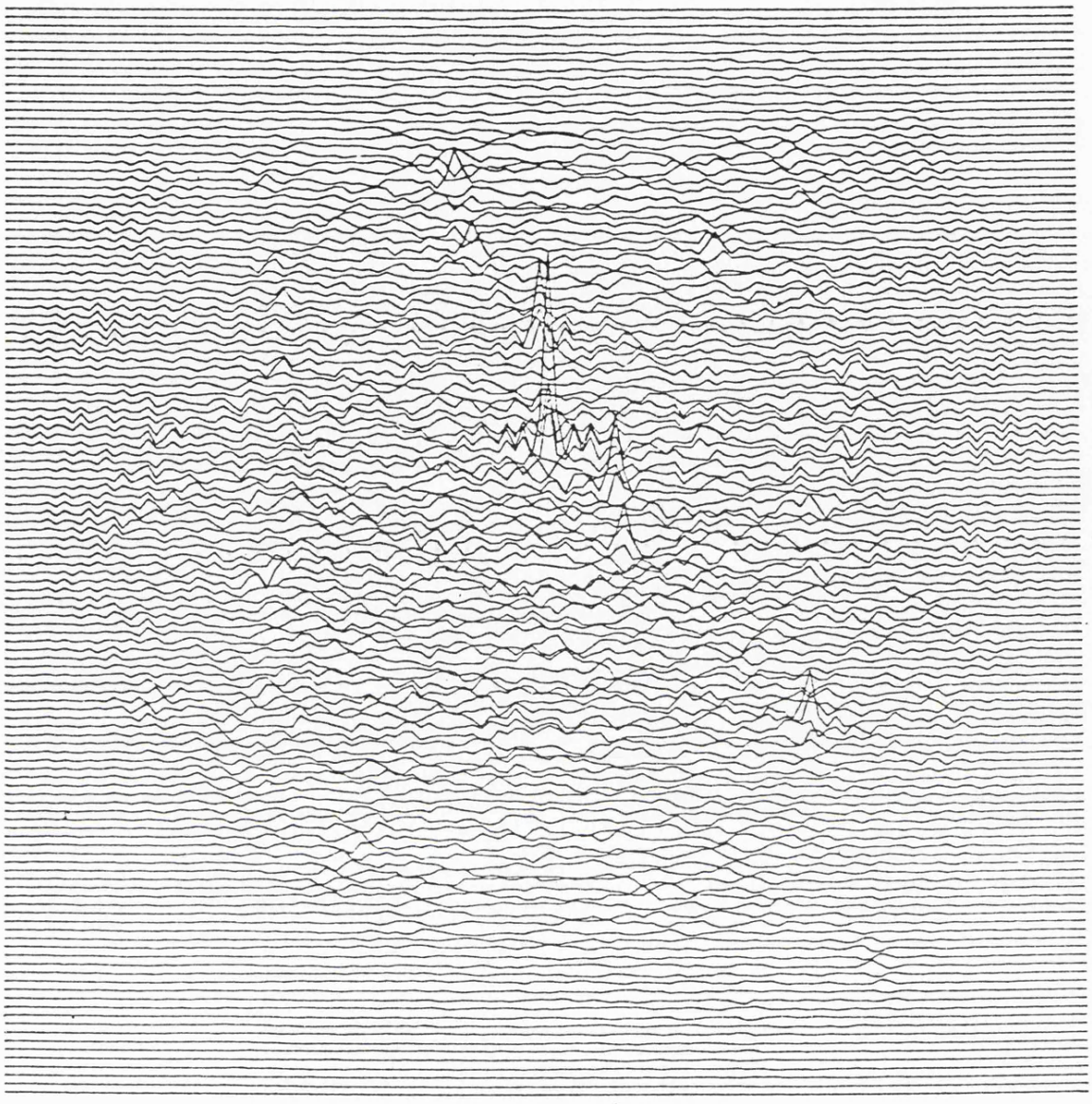


Figure 44. Simulation 1 using a Fresnel mask and crosscorrelation. The circular side lobes are clearly visible. Vertical scale 2000 cts/mm.

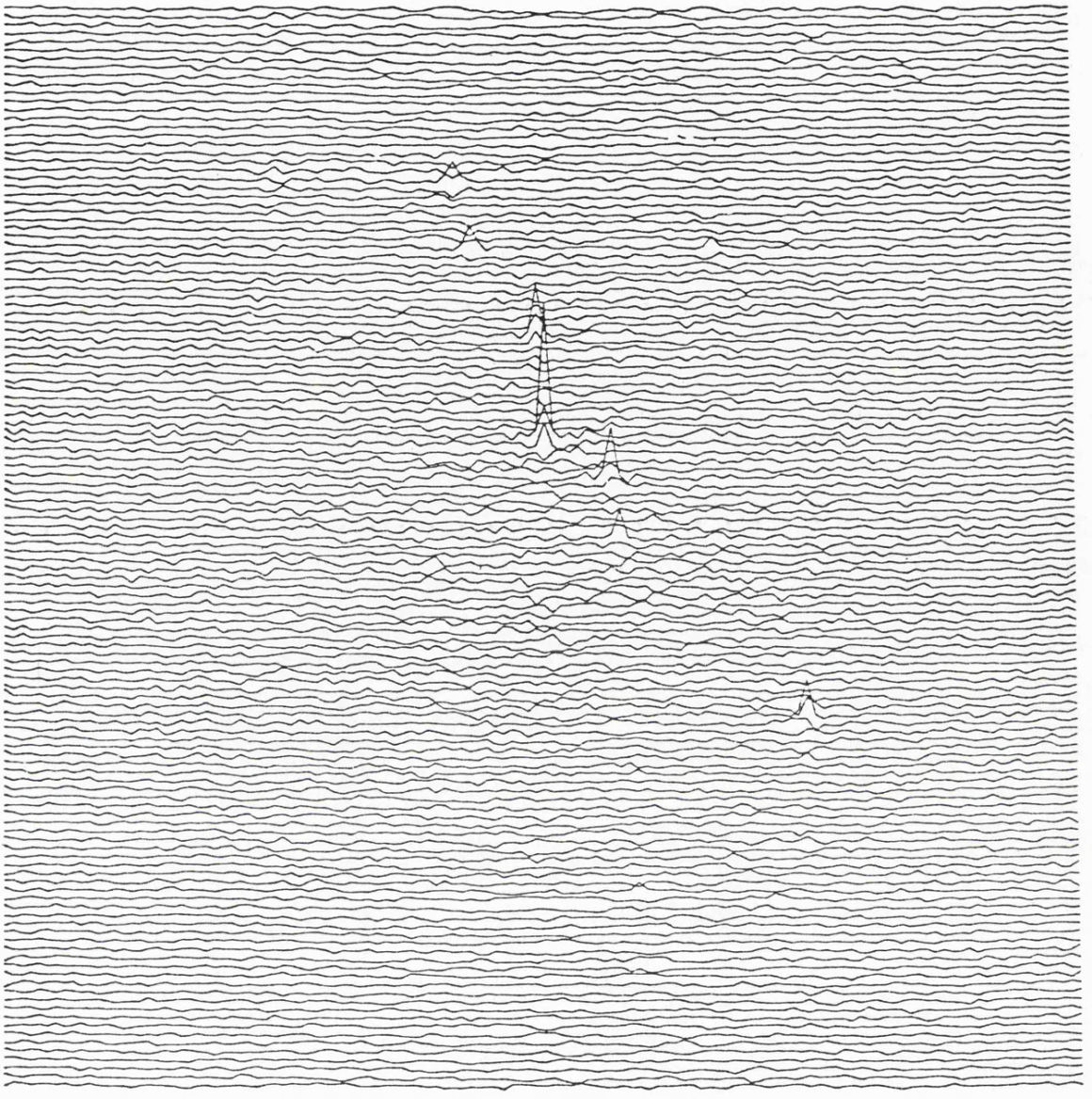


Figure 45. Simulation 2 using a Fresnel mask and Wiener filter. The side lobes visible in figure 44 have been eliminated. Vertical scale 2000 cts/mm.

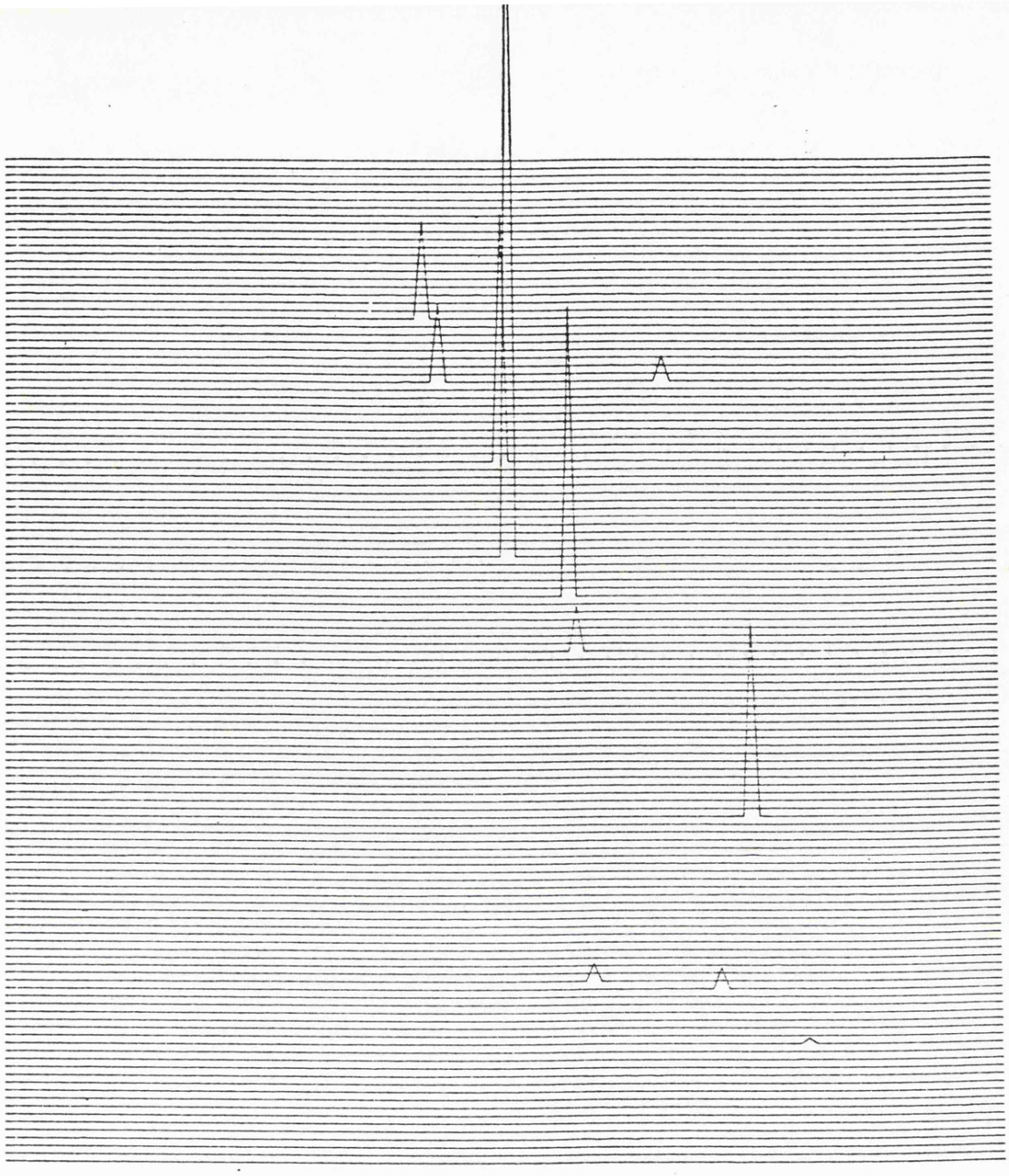


Figure 46. Simulation 2 using a pseudo-noise sequence mask and the MEM. Even the strongest false peaks are too small to be seen. Vertical scale 500 cts/mm.

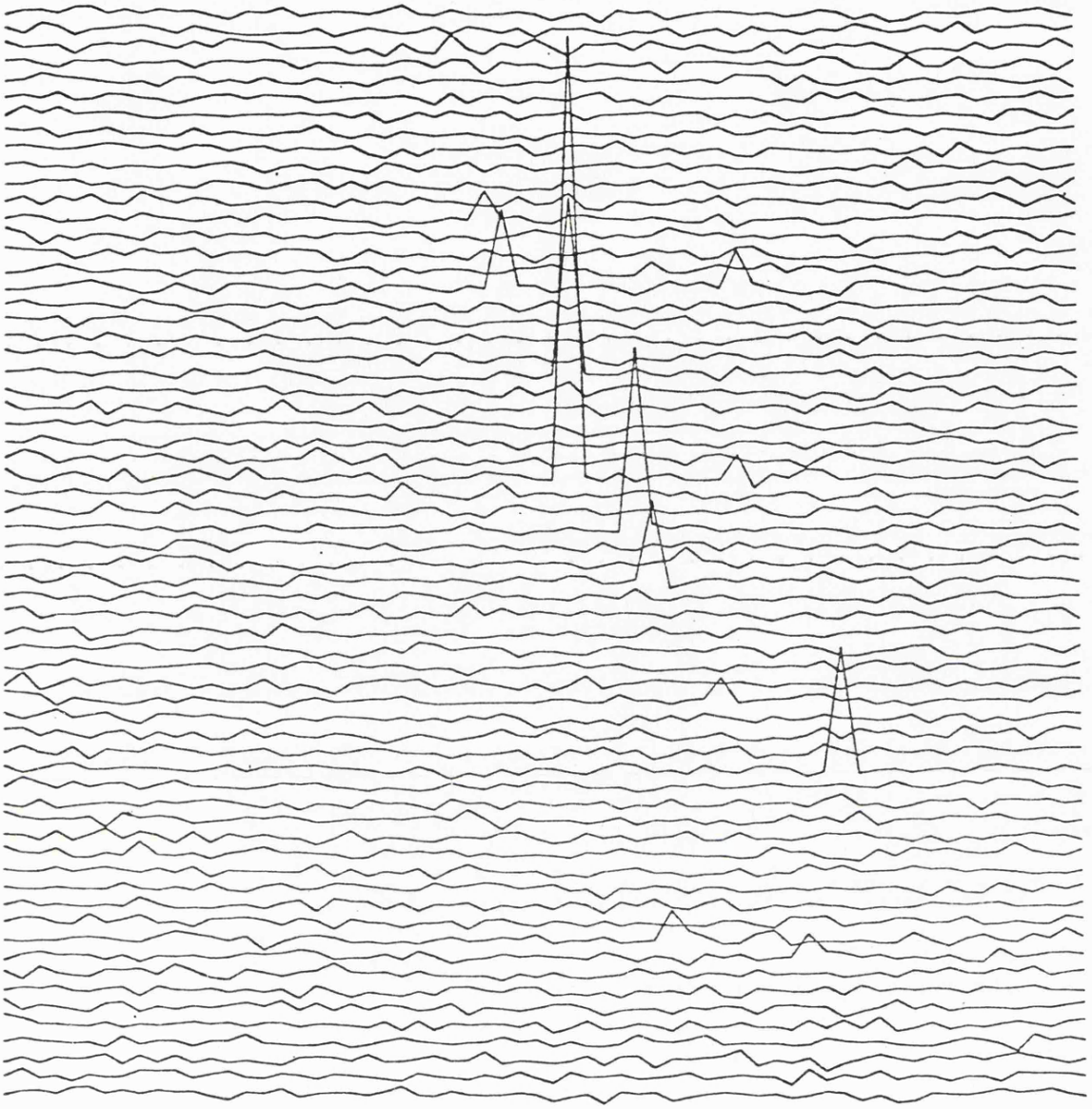


Figure 47. Simulation 3 using the aliased system. Note that the rms of the background fluctuations is roughly constant over the whole field of view and that the resolution is reduced by a factor of 2. Vertical scale 500 cts/mm.

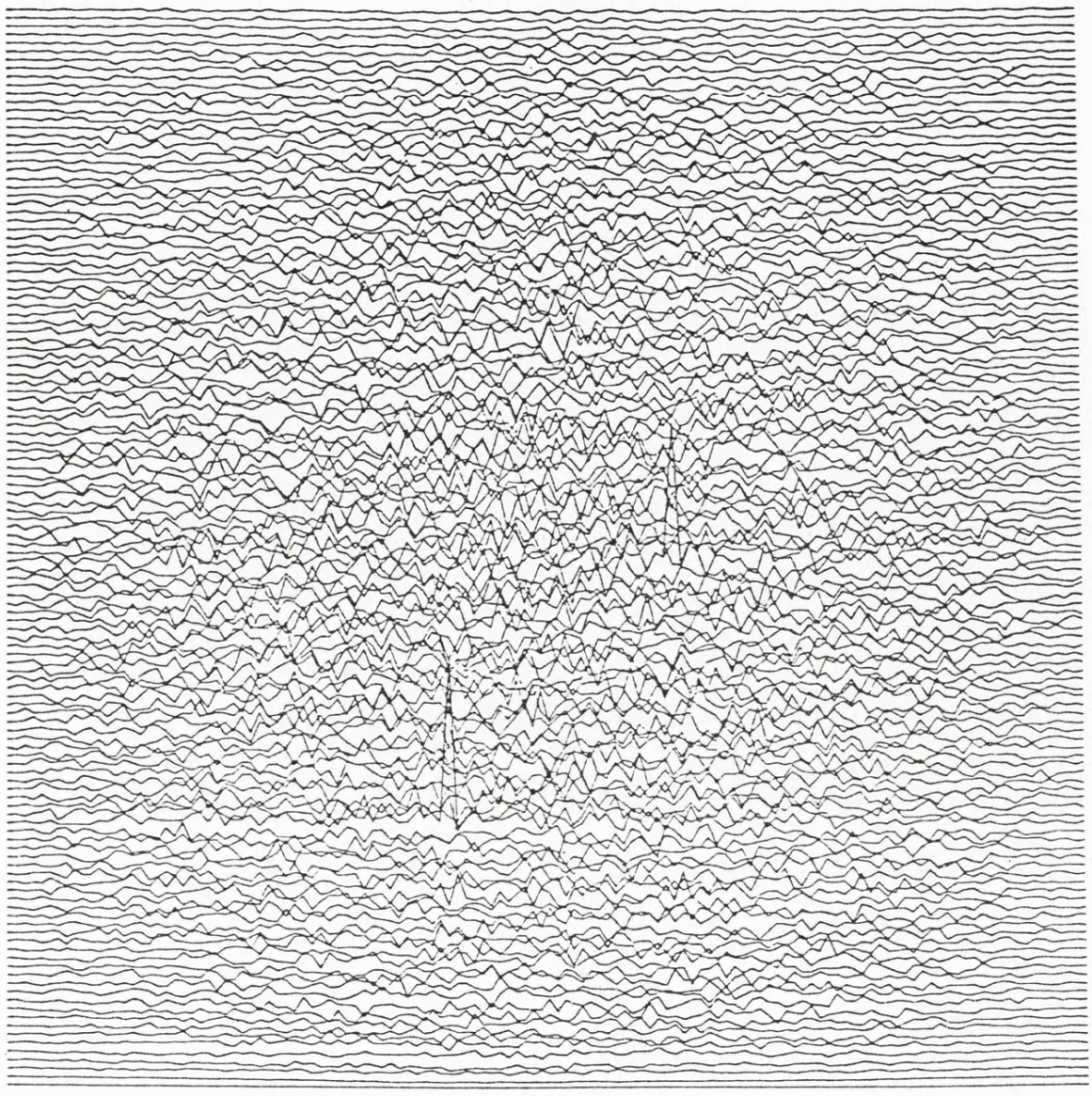


Figure 48. Simulation 4 using a pseudo-noise sequence mask and cross-correlation. Two sources are visible while the third is completely obscured by the background. The diminution of the background fluctuations towards the edges of the field of view is particularly marked in this simulation. Vertical scale 1000 cts/mm.

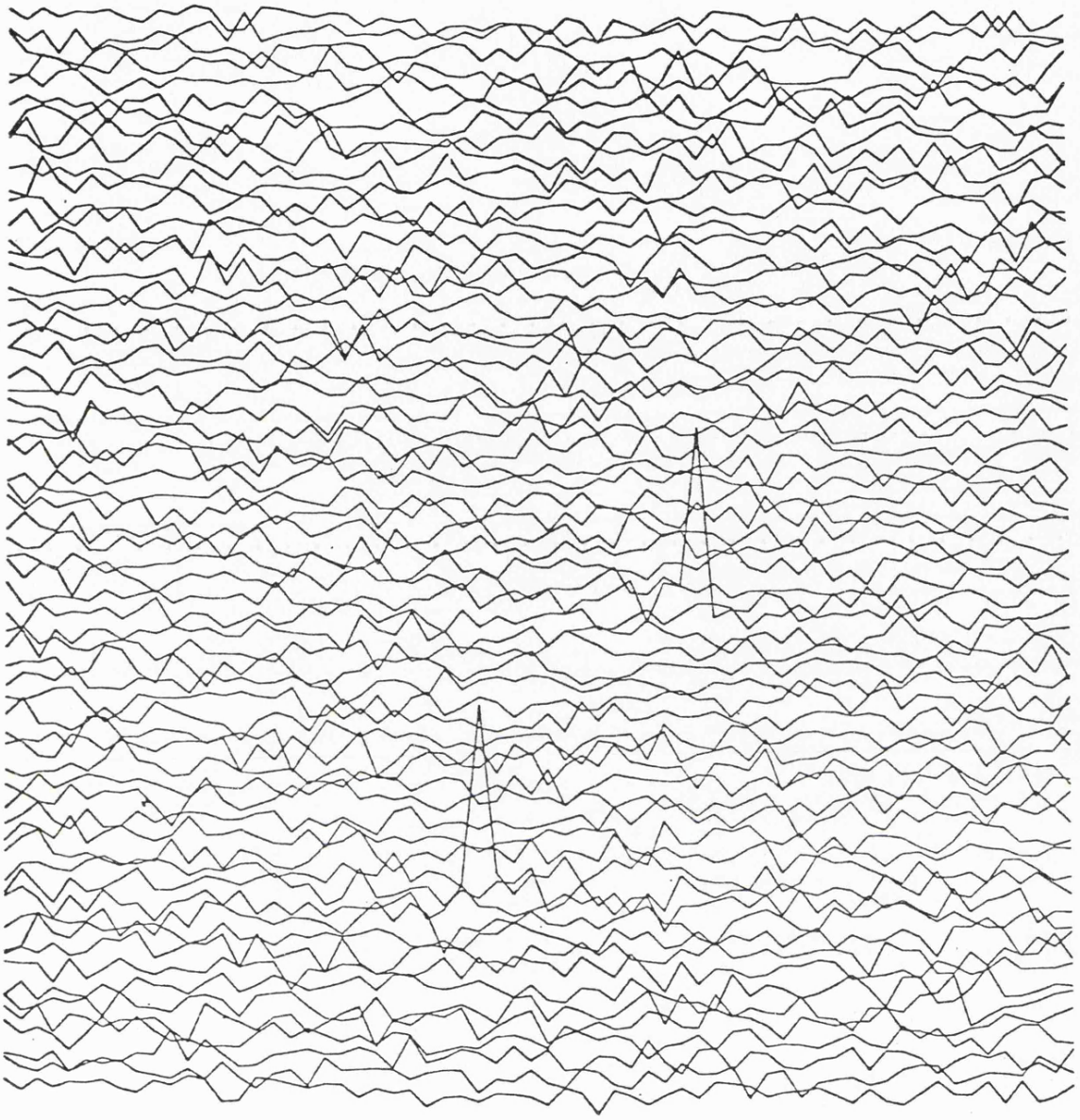


Figure 49. Simulation 5 using the aliased system. Note the reduced resolution and constant background when compared to figure 48. Vertical scale 500 cts/mm.

SIMULATION NUMBER	INSTRUMENT TYPE	SOURCE FIELD 45° BY 45°	INTEGRATION TIME SECS.	DETECTOR PARTICLE BACKGROUND COUNT	SKY BACKGROUND COUNT	DECONVOLUTION METHOD	THEORETICAL RMS FLUCTUATION ON AXIS USING CROSSCORRELATION	RMS FLUCTUATION OF PIXELS EXCLUDING OBVIOUS SOURCES	SIGNIFICANCE OF LARGEST FALSE PEAK OR PEAKS > 5 RMS	NUMBER OF SOURCES ≥5 RMS	NUMBER OF SOURCES > LARGEST FALSE PEAK	SIGNIFICANCE OF A TYPICAL SOURCE (N RMS)
1	FRESNEL MASK WITH 8 RINGS	GALACTIC CENTRE CENTRE RA 270.00 DEC-30.00 ALL SOURCES≥10M CRABS (2-18) KEV)	100	380	19000	CROSSCORRELATION	393	660	4.62	8	8	6.17
						WIENER FILTER		402	3.82	9	12	6.32
						MEM (χ^2_ν=4.19)		3.09	33.0 13.3 12.4	14	10	252.4
2	PSEUDO NOISE SEQUENCE MASK		100	380	25550	CROSSCORRELATION	435	638	4.09	10	11	9.46
						WIENER FILTER		544	4.71	11	11	8.36
						MEM (χ^2_ν=3.44)		3.13	18.3 8.52 6.00	14	12	306.0
3	CYCLIC PSEUDO NOISE SEQUENCE		100	380	25550	CROSSORRELATION	435	406	3.86	12	14	10.95
4	PSEUDO NOISE SEQUENCE MASK	COMA REGION CENTRE RA 190.00 DEC 40.00 ALL SOURCES	5000	19000	1277500	CROSSCORRELATION	1163	643	5.00	2	2	30.7
						WIENER FILTER		452	5.14	2	2	28.0
						MEM (χ^2_ν=1.16)		71.7	3.28	2	2	7.26
5	CYCLIC PSEUDO NOISE SEQUENCE		5000	19000	1277500	CROSSCORRELATION	1163	1117	3.32	2	2	17.7
6	PSUEDO NOISE SEQUENCE MASK	COMA REGION AS ABOVE WITHOUT STATISTICS	5000	19000	1277500	CROSSCORRELATION	—	295	5.59	3	3	65.4

Figure 50. Results of the 2-D burst monitor simulations.

than predicted. This was as expected because, as shown in figure 48, the fluctuations are greatly reduced near the edges of the field of view. The rather large rms value obtained by the MEM algorithm was dominated by the pyramidal background mean variation rather than random fluctuations. The two sources were in fact seen at a $\gtrsim 30$ rms in the MEM solution of simulation 4.

The computer time required for deconvolution is very important in the application of the methods described here. Crosscorrelation and Wiener filtering both require one two-dimensional convolution while the MEM needs two such convolutions for one iteration. Typically, ten passes will be needed to arrive at a reasonable solution. The time required for the convolutions using different dimensions are given in Appendix I and they indicate that although the above simulations were limited by computing power, a real system could be made viable providing dedicated array processing hardware was used.

4.7 Conclusion to Part II.

A brief resumé of the theory of coded mask telescopes was given with the emphasis on the form of the hologram obtained. The deconvolution theory presented in Part I was used to provide methods for decoding the holograms. The characteristics of the simple, non-aliased system as originally described by Dicke was studied in depth, so that the performance of two instruments proposed to NASA could be simulated and assessed.

It was demonstrated that the simple system has distinct advantages over the aliased system for imaging

sparse, wide source fields, providing that a suitable decoding method is employed. The MEM was shown to perform well under simulation and promises to yield such a deconvolution procedure.

The choice of mask pattern for imaging sparse fields is not critical and indeed using a circularly symmetric mask such as a Fresnel zone plate has the advantage that overlaying of hologram data can be achieved, even if a rotation about the instrument axis has occurred.

The aliased system is more difficult to construct, since a slat collimator of fine pitch must be used and the entire detecting area must be free of support structure if the 'coding advantage' is to be retained.

Although the MEM is of greatest use for the simple configuration, it can be used to allow for coding errors introduced by inadequacies of the hardware in both cases. More work is needed to clarify the performance of the MEM algorithm with respect to noise fluctuations, but it is clearly very powerful. Part III, to follow, provides more insight into this problem.

The Wiener filter form was shown to provide a better performance when the mask pattern was below optimum because it carefully allowed for the exact form of the modulation transfer function of the mask. Perhaps a final data processing system for decoding coded mask telescope data would include many deconvolution methods to cater for particular circumstances, trading off speed for image quality, looking for single transient events or surveying complicated, static fields of point sources etc.

PART III

PROPORTIONAL COUNTER ANODE PULSE HEIGHT DISTRIBUTION ANALYSIS.

CHAPTER 5. THE MAXIMUM ENTROPY METHOD OF DECONVOLVING X-RAY ENERGY SPECTRA.

5.1 Introduction to Part III.

The theory and techniques developed in Parts I and II for the analysis of X-ray images can be applied to the problem of X-ray energy or spectrum analysis. The general form of the response of a proportional counter was given in equation (1.8):

$$G'(E') = \int \mathcal{Z}_d(E)\, G(E)\, R(E-E',E)\, dE \qquad (1.8)$$

where $G(E)$ is the source energy spectrum, $R(E-E',E)$ is the counter response function and $\mathcal{Z}_d(E)$ is the detector efficiency. The distribution $G'(E')$ is sampled to give an event set G_{En}.

$$G_{En} = S\{G'(E')\} \qquad (5.1)$$

The response of a proportional counter is not linear since although the mean pulse height is proportional to the X-ray energy, the spread of pulse heights described by $R(E-E',E)$ is a function of E. The linear techniques centred on the fast Fourier transform cannot be utilised and in fact direct product filtering is only possible if the transformation which diagonalises the instrument's integral transform can be found.

The most common approach currently used is the setting up of an analytical model function for the source spectrum $\hat{G}(E)$, applying the instrument transformation (1.8) to give $\hat{G}'(E')$ - an estimate of the pulse height distribution, and comparing this with the data set G_{En}

using the chi-squared statistic. The success of such a procedure relies on the choice of an appropriate modelling function. If the physics of the source is well understood then a good fit to the free parameters of the system can be obtained, although it is difficult to estimate the systematic errors involved in the fitting process. It is very easy to be optimistic about the instrument's performance and introduce more degrees of freedom than the instrument is capable of giving. A method which is as free as possible from any preconceptions about the source and which carefully considers the action and limitations of the instrument is required.

The use of the maximum entropy method described in sections 2.6 and 2.7 is tempting. It is capable of handling the non-linear nature of the response and the statistics of the event set. Unfortunately, the statistical significance of the maximum entropy estimate is not fully understood although the solution is very safe when the blurring and noise are both bad (see section 3.6). The following sections describe the application of the maximum entropy method to the deconvolution problems presented by the proportional counter.

5.2 Formulation of the Maximum Entropy Method for X-ray spectral analysis.

Although section 2.6 concentrated on image formation, the theory is directly applicable to spectral analysis. The physical interaction of photons with the instrument can be described in exactly the same way and the entropy expressions will be identical. The only modification

necessary is the interpretation of z, the number of degrees of freedom, given by equation (2.53) which is reproduced below:

$$z = \frac{ctA}{c\tau\sigma} \tag{2.53}$$

ctA is the detection volume and $c\tau\sigma$ is the coherence volume. For the case of a 2-D X-ray imaging system, the only strongly varying function over the data field was the exposure, tA, set by the mirror vignetting and equation of motion of the instrument. That is, the only a priori weighting used was the response of the instrument and otherwise all pixels were treated identically. However the coherence volume $c\tau\sigma$ can be expressed as in equation (2.52):

$$V_{coh.} = \frac{c^3 R^2}{\nu^2 \Delta\nu \Delta\alpha \Delta\beta} \tag{2.52}$$

The z appropriate to spectral data includes a $1/\nu^2$ term or a priori weighting associated with a sample of width $\Delta\nu$ at frequency ν. The function $S_{\alpha\beta}(n_{\alpha\beta})$ given as expression (2.42) can be rewritten in terms of frequency samples $\Delta\nu$ at frequency ν rather than position samples $\Delta\alpha\Delta\beta$ at position (α,β):

$$S_\nu(n_\nu) = (n_\nu+z_\nu)\ \ln(n_\nu+z_\nu) - n_\nu \ln n_\nu + z_\nu \ln z_\nu \tag{5.2}$$

The approximation $z_\nu \approx z_\nu - 1$ has been made since z_ν is so large for X-rays.

The entropy expression for the system of photons is

given by the mean of $S_\nu(n_\nu)$ over all samples:

$$S_{\{\nu\}} = \sum_\nu (n_\nu+z_\nu)\ \ln(n_\nu+z_\nu) - \sum_\nu n_\nu \ln\ n_\nu - \sum z_\nu \ln\ z_\nu \tag{5.3}$$

$S_{\{\nu\}}$ is the spectral entropy of the photon field rather than the configurational entropy, $S_{\{\alpha\beta\}}$, used in the imaging case. The number of degrees of freedom z_ν is given by:

$$z_\nu = \frac{t_\nu\ A_\nu\ \nu^2\ \Delta\nu\ \Delta\alpha\ \Delta\beta}{c^2\ R^2} \tag{5.4}$$

Just as the exposure tA was a function of (α,β) in the imaging case, it is a function of ν in the spectral case.

Suppose there is only one constraint, that the total energy is constant:

$$\sum h\nu\, n_\nu = U\ (\text{constant}) \tag{5.5}$$

The 'objective function' for the spectral case corresponding to equation (2.76) in the imaging case is:

$$O_{n_\nu} = S_{\{\nu\}} + \beta(U - \sum h\nu\, n_\nu) \tag{5.6}$$

where β is acting as a Lagrange multiplier.

Differentiating with respect to n_ν and finding the maximum by setting $dO_{n_\nu}/dn_\nu = 0$ gives:

$$\ln\left(1 + \frac{z_\nu}{n_\nu}\right) - \beta h\nu = 0 \tag{5.7}$$

Rearranging expression (5.7) yields:

$$n_\nu = \frac{z_\nu}{\exp(\beta h\nu) - 1} \tag{5.8}$$

(5.8) is the maximum entropy solution under the single constraint (5.5). Substituting for z_ν from equation (5.4) gives:

$$n_\nu = \frac{t_\nu A_\nu \nu^2 \Delta\nu \Delta\alpha \Delta\beta}{c^2 R^2} \frac{1}{\exp(\beta h\nu) - 1} \qquad (5.9)$$

This photon number distribution yields the corresponding energy distribution:

$$U_\nu = n_\nu h\nu = \frac{t_\nu A_\nu h \Delta\alpha \Delta\beta}{c^2 R^2} \frac{\nu^3}{\exp(\beta h\nu) - 1} \qquad (5.10)$$

Expression (5.10) is essentially Planck's radiation equation, reference 24 (sections 13-12, 13). The multiplier β is usually shown to be related to the thermodynamic temperature T of the source by the relation:

$$\beta = 1/\ kT \qquad (5.11)$$

In the above example of the application of the maximum entropy method, an analytic solution exists in the form of equation (5.9). If the source was a black body then the single constraint (5.5) is sufficient to yield the energy distribution characterised by a single parameter, the temperature of the black body. In order to introduce an instrument dependent in the event set G_{En}, it is convenient to express equation (1.8) in discrete form which will be a good approximation providing the sampling rate $\Delta\nu$ is small enough:

$$G' = [R]\ G \qquad (5.12)$$

G' and G are the column vectors formed by sampling G'(E')

and $G(E)\ \zeta_d(E)$ respectively (G represents the detected photon distribution before blurring). [R] is the instrument kernel matrix. [R] will have the banded structure described in section 2.1 but because the system is non-linear, it will not have the convenient Toeplitz structure.

Given an estimate of the source spectrum $\hat{G}$, The corresponding counter distribution $\hat{G}'$ is simply given by direct application of (5.12). The event set G_{En} must be binned into a vector using bins of width $\Delta\nu$ to correspond to the sampling rate in G' and G:

$$D_i = \delta(\nu - i\Delta\nu) \int G'(\nu')\ T(\nu - \nu')\ d\nu' \qquad (5.13)$$

where D_i is the data vector and $T(\nu)$ is a top hat function given by $T = 1$ for $-\Delta\nu/2 < \nu < \Delta\nu/2$ and $T = 0$ otherwise (c.f. equation (1.13) for 2-D image counterpart).

The comparison between the estimated vector $\hat{G}'$ and the data vector D can be made using a chi-squared statistic using σ_i^2, the variance associated with data element D_i:

$$\chi^2 = \sum_i \frac{(D_i - \hat{G}_i')^2}{\sigma_i^2} \qquad (5.14)$$

The data constraint can then be introduced into the objective function using the Lagrange multiplier λ:

$$O_G = S_{\{i\}} + \beta\left(U - \sum_i hi\Delta\nu\hat{G}_i\right) - \lambda \sum_i \frac{(D_i - \hat{G}_i')^2}{\sigma_i^2} \qquad (5.15)$$

with $$S_{\{i\}} = \sum_i (\hat{G}_i + z_i)\ \ln\ (\hat{G}_i + z_i) - \sum_i \hat{G}_i\ \ln\ \hat{G}_i - \sum_i z_i \ln\ z_i \qquad (5.16)$$

Equation (5.15) is the spectral counterpart of equation

(2.77) used for the objective function O_f in the 2-D image case. Differentiating with respect to elements $\hat{G}_j$ and setting the derivative equal to zero gives:

$$\ln\left(\frac{1+z_j}{G_j}\right) - \beta h_j \Delta\nu + \lambda \sum_i \left\{ R_{ji}^t \frac{(D_i - \hat{G}_i')}{\sigma_i^2} \right\} = 0 \qquad (5.17)$$

Notice, as before, that $\{\ \}$ contains a crosscorrelation of the weighted difference with the response matrix because of the transposition. The complex conjugation has no effect since the operators are real in this case. Providing z_j/G_j is large, which has been shown to be true for all astronomical X-ray observations in section 2.6, then (5.17) can be approximated and rearranged as:

$$\ln \hat{G}_j = \ln z_j - \beta h_j \Delta\nu + \lambda \sum_i \left\{ R_{ji}^t \frac{(D_i - \hat{G}_i')}{\sigma_i^2} \right\} \qquad (5.18)$$

The maximum entropy solution is therefore given by the transcendental equation :

$$\hat{G}_j = z_j \exp\{-\beta h_j \Delta\nu\} \exp\{\lambda \sum_i R_{ji}^t \frac{(D_i - \hat{G}_i')}{\sigma_i^2}\} \qquad (5.19)$$

where β and λ must be chosen to satisfy the constraints χ^2 minimum or small and energy conservation:

$$\sum_i h_i \Delta\nu \hat{G}_i = U \qquad (5.20)$$

The normalisation condition imposed by (5.20), is more difficult to achieve than the corresponding constraint (2.80) in the 2-D image case:

$$z_j \exp\{-\beta h_j \Delta\nu\} = \frac{t_j A_j j^2 \Delta\nu^3 \Delta\alpha \Delta\beta}{c^2 R^2} \exp\{-\beta h_j \Delta\nu\} \qquad (5.21)$$

The 'normalisation function' (5.21) is an explicit function of j and finding the parameters β and R (or an equivalent normalising constant) to satisfy equation (5.20) is not simple. Furthermore, the total energy U of the detector photons is not known with the same certainty as n_t, the total count, was in the imaging case.

The presence of the function (5.21) ensures that if $\lambda \to 0$, i.e. the data constant is very weak, then the spectrum has the most likely form corresponding to a black body source with temperature given by $T = 1/k\beta$ and total luminosity determined by the angular size $\Omega = \Delta\alpha\Delta\beta/R^2$. However as λ is increased and χ^2 decreases, the data will dominate over the a priori black body function. If χ^2 is to be small then constraint (5.20) must be reasonably well satisfied. An approximate solution ignoring the explicit j dependence in (5.21) and normalising to the total count in the data n_t:

$$n_t = \sum_i D_i \qquad (5.22)$$

should produce a reasonable result despite the fact that the a priori weighting is different. Such an approximation eases the computation difficulty of normalisation imposed by (5.20) without seriously degrading the resulting estimate $\hat{G}$, providing that the proportional counter is reasonably well behaved and the overall pulse height distribution D is fairly representative of the source spectrum distribution $\mathcal{J}_d(E)\ G(E)$ which is blurred and sampled to give D.

The approximate solution has the form:

$$\hat{G}_j = t_j A_j \lambda \exp\{\lambda \sum_i R_{ji}^t \frac{(D_i - \hat{G}_i')}{\sigma_i^2}\} \qquad (5.23)$$

where λ is chosen to satisfy constraint (5.22). Equation (5.23) can be solved using exactly the same algorithm used to process the image data described in section 3.6; the only differences being the contraction to 1-D which eases the time restriction on the number of iterations possible and the form of [R], the response matrix. The matrix multiplications have to be carried out directly, without the use of a fast transformation. However since the problem is only one dimensional this imposes no real restriction on the viability of the algorithm.

5.3 The application of the Maximum Entropy Method to real and simulated proportional counter anode pulse height data.

This project was undertaken because pulse height data was available for the Cygnus Loop data already described in section 3.3.

The information content of the data was obviously very limited by the performance of the instrument but a hardness map was constructed by splitting the full energy range map, figure 23, into two separate maps using the nominal energy ranges 0.15 - 0.57 keV and 0.57 - 1.12 keV to select the counts for binning. Using 4' x 4' bins produced very poor counting statistics for any comparison between the two maps and so the bin size was increased to 32' x 32'. The hardness map was calculated by direct division of the high energy matrix by the low energy

matrix:

$$H_{ij} = \frac{J_{ij}''}{J_{ij}^2} \qquad (5.24)$$

The hardness matrix $[H]$ was then subjected to a significance test to suppress elements which were dominated by counting statistics errors. The ratio of variance due to counting statistics to hardness squared, H_{ij}^2, was calculated for each element of $[H_i]$:

$$R = \frac{\sigma_{ij}^2}{H_{ij}^2} = \frac{\sigma_{ij}^{L^2}}{J_{ij}^{L^2}} + \frac{\sigma_{ij}^{H^2}}{J_{ij}^{H^2}} = \frac{1}{J_{ij}^{L}} + \frac{1}{J_{ij}^{H}} \qquad (5.25)$$

Equation (5.25) is the appropriate combination of errors formula for the ratio quantity H_{ij}. $1/\sqrt{R}$ was used as a measure of significance of the hardness H_{ij}. The resulting hardness map, plotting only those elements for which $1/\sqrt{R} \geqslant 3$ (i.e. 3 sigma significance), is shown in figure 51. It can be seen that using 32' x 32' bins and a 3σ significance produced a map with good coverage and which was reasonably smooth, without an obviously noisy appearance. Furthermore the map displays an unmistakable trend of hardness variation from the bottom left to the top right.

The success of this crude assessment of spectral information in the data promoted further investigation. The pulse height distribution corresponding to the two hardest (bottom left) and two softest (upper right) bins were plotted and are shown in figure 52. They are plotted as distributions in 0.05 keV bins using the calibration values of gain to convert from the actual anode pulse

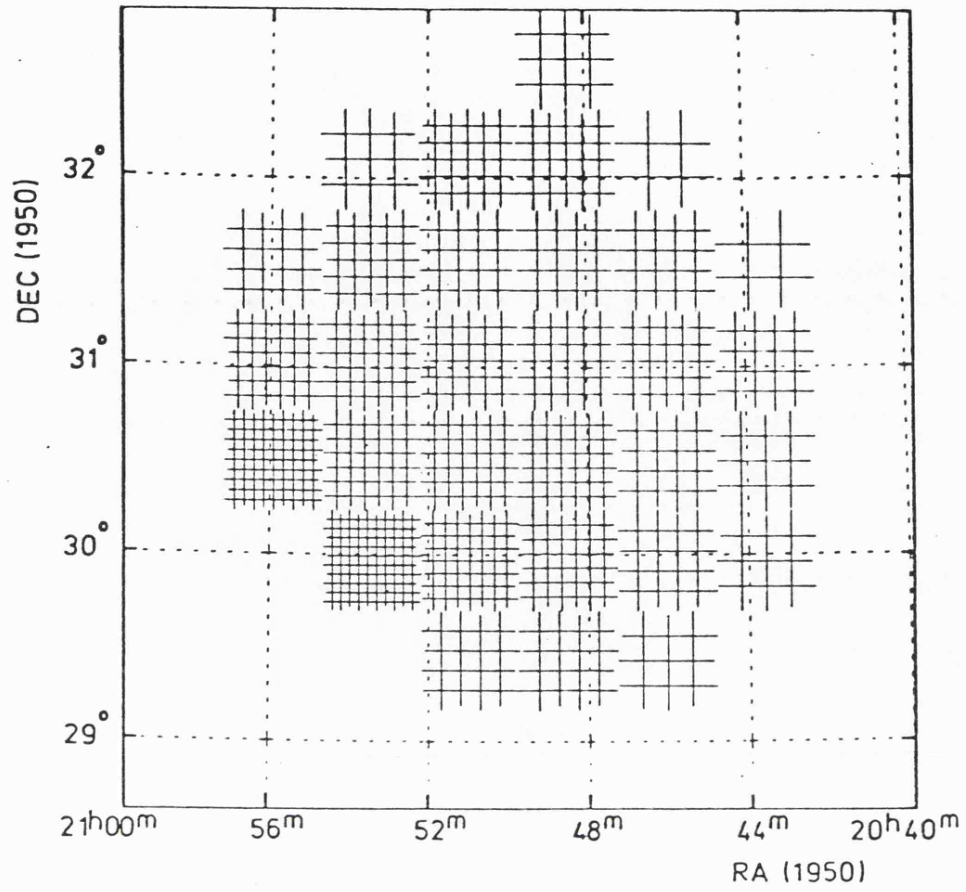
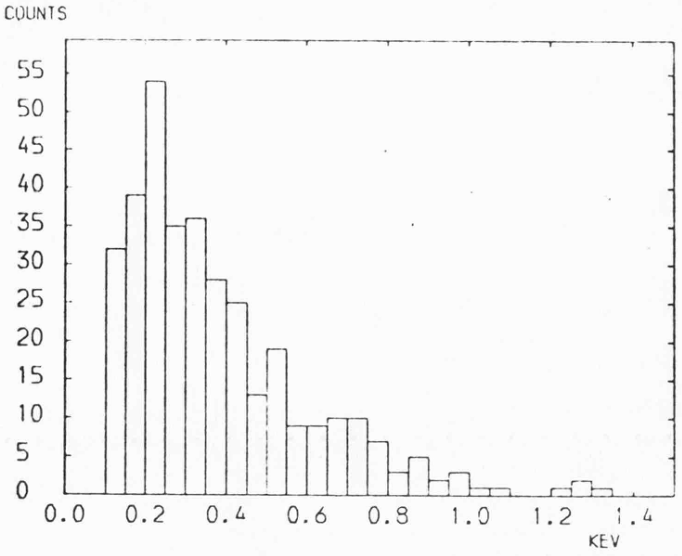
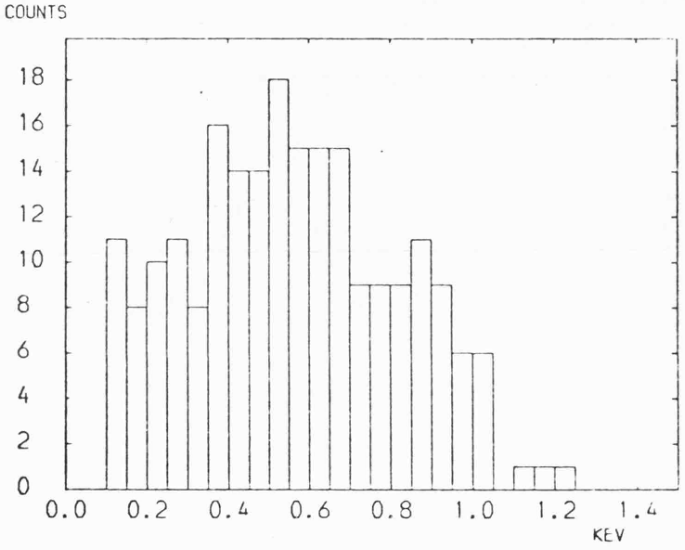


Figure 51. The hardness map of the Cygnus Loop. Pixel size 32' x 32'. A systematic change from soft in the North West to hard in the South East is to be noted.

Softest region in the North West, 345 counts.



Hardest region in the South East, 217 counts.



Western limb, 877 counts.

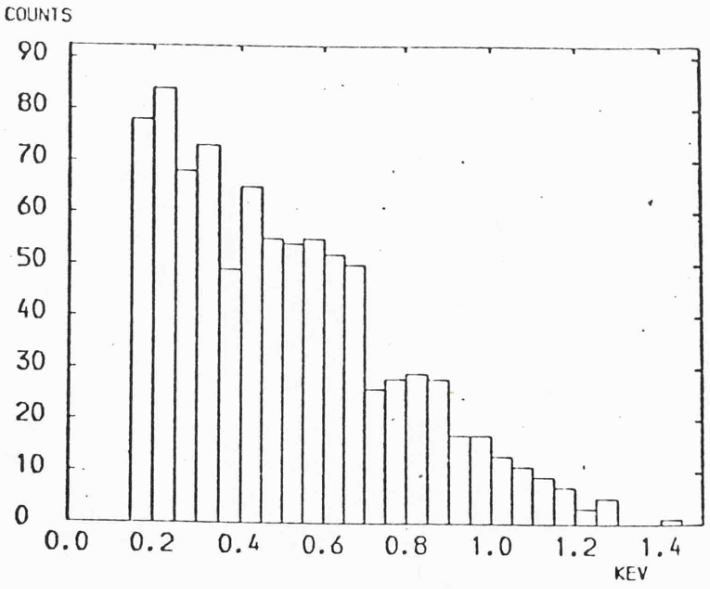


Figure 52. Anode pulse height data from the Cygnus Loop.

height voltages. The hardness difference indicated by figure 51 is borne out by the differences in the pulse height distributions.

In order to analyse the pulse height data further, a good description of the instrument response functions $\mathfrak{Z}_d(E)$ and $R(E-E',E)$ was required. Using the preflight calibration data, the detection efficiency including the mirror reflectivity and the counter window transmission for on-axis X-rays was calculated. The absolute value of this efficiency is impossible to calibrate to within about a factor of 2 but the relative variation as a function of incident X-ray energy is quite well understood, with measurement and theory in reasonable agreement. Figure 53 shows the overall efficiency as a function of energy for the MIT/ Leicester payload used to collect the Cygnus Loop data. The resolution of the counter was well calibrated before launch. The response to a single energy was Gaussian, with the peak position proportional to the energy and the width fitted by:

$$\sigma(E) = 0.21 \times \sqrt{E} \qquad (5.26)$$

where:

$$R(E-E',E) = \frac{1}{\sqrt{2\pi}\,\sigma(E)} \exp\left\{\frac{(E - E')^2}{2\sigma^2\,(E)}\right\} \qquad (5.27)$$

The maximum entropy algorithm was tested with simulated data using the calibration values for the counter performance. This was felt to be necessary since the response was very non-linear compared to the previous imaging applications reported above. Figures 54 and 55

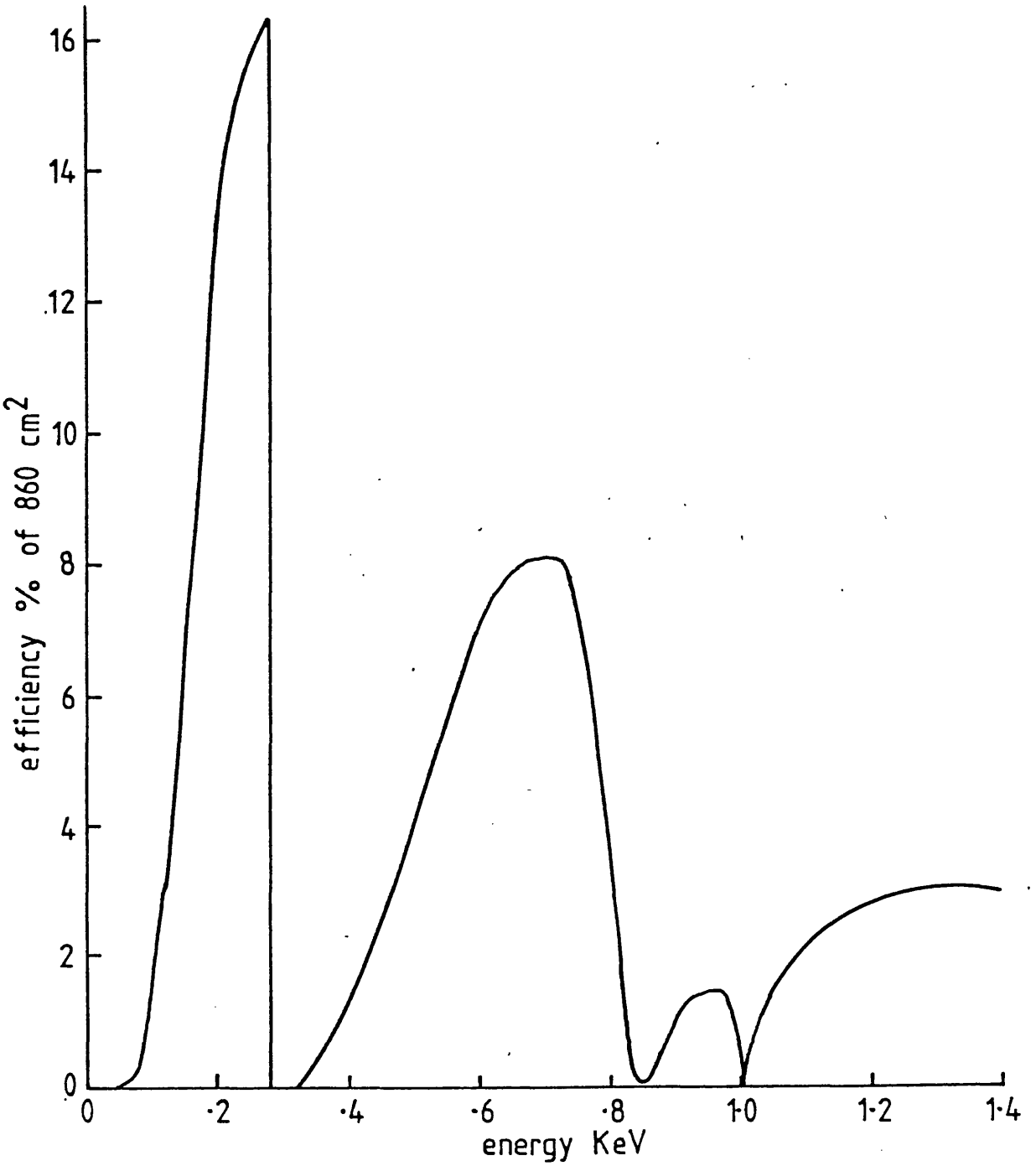


Figure 53. The efficiency of the MIT/Leicester payload used to observe the Cygnus Loop expressed as a percentage of the geometric area. The absolute value of this efficiency is only known to within a factor of 2.

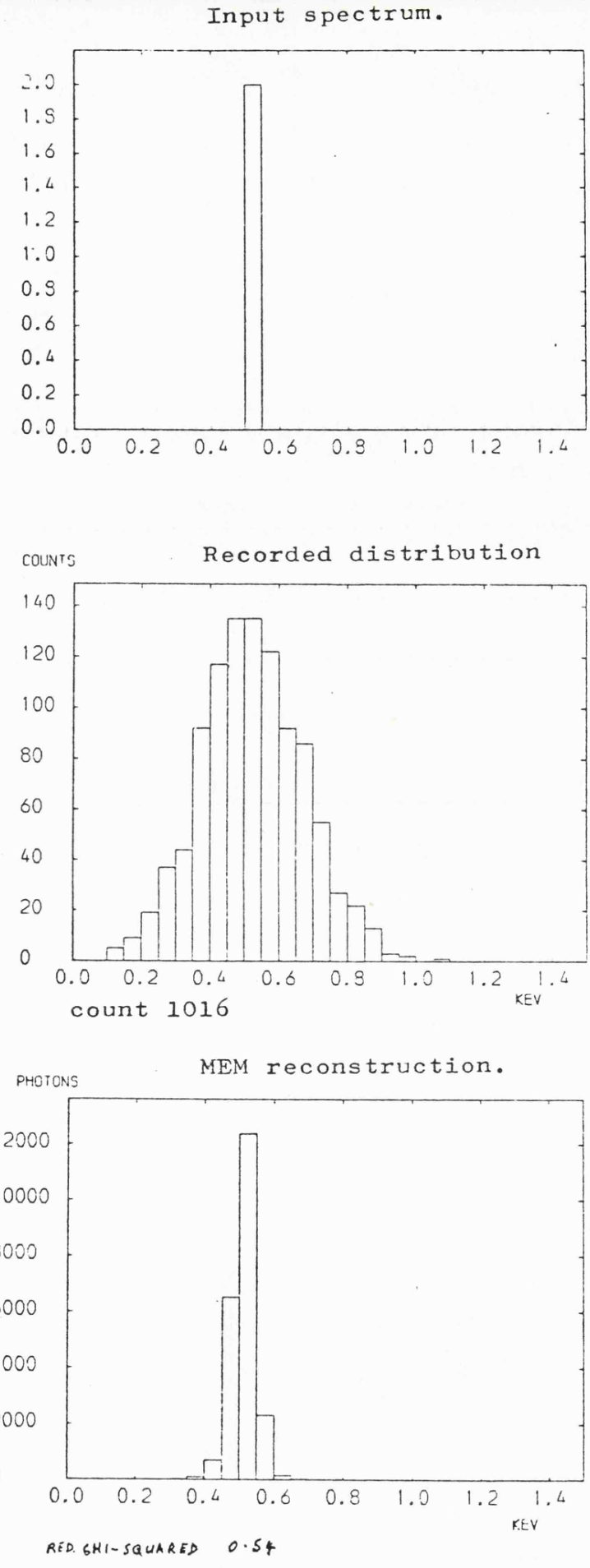


Figure 54. Simulation of a single line spectrum with subsequent MEM reconstruction.

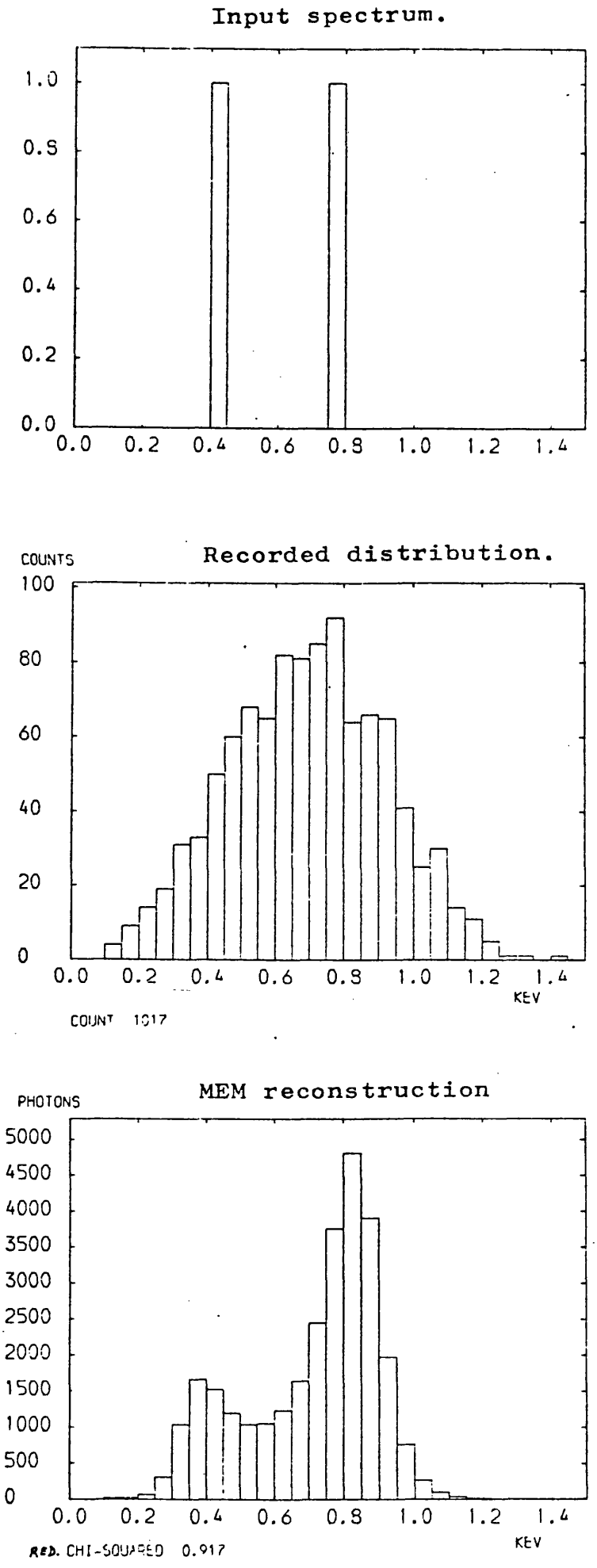


Figure 55. Simulation of a 2 line spectrum with subsequent MEM reconstruction. The lines are resolved but the response is non-linear because of the very strong modulation of the C_k absorption edge..

demonstrate the success of the simulations and the algorithm is clearly capable of handling very badly degraded data. The resulting distributions are well controlled and never contained unreal artifacts due to noise. However the sensitivity is necessarily non-linear because of the strong efficiency modulation. The relative strength of features in the final distribution is therefore affected by the efficiency of the instrument, even after correction. This is demonstrated by the flat field simulation, figure 56, which clearly yields a non-flat reconstruction. However the initial flat distribution used as the first guess gave a $\chi^2_\nu = 1.22$, indicating that there could not be any strong features in the data. Because the maximum entropy algorithmis non-linear, the maximum entropy reconstruction is necessarily non-linear. In fact any attempt to 'fit' a source spectrum to the data will also suffer from the severe degradation caused both by the absorption edges and by the instrument's resolution and care must be exercised in interpretation of results.

The simulations described above were obviously not exhaustive, but do demonstrate that the algorithm produces meaningful results and does achieve quite a high degree of deconvolution. The real data distributions obtained from the Cygnus Loop data were therefore given to the algorithm. The resulting source distributions are plotted in figure 57 and should be compared to the pulse height distributions in figure 52. The source spectra are plotted as photons incident on the geometric collecting area over the exposure time appropriate for two 32' by 32' bins. By using figure 53, the efficiency profile of the instrument,

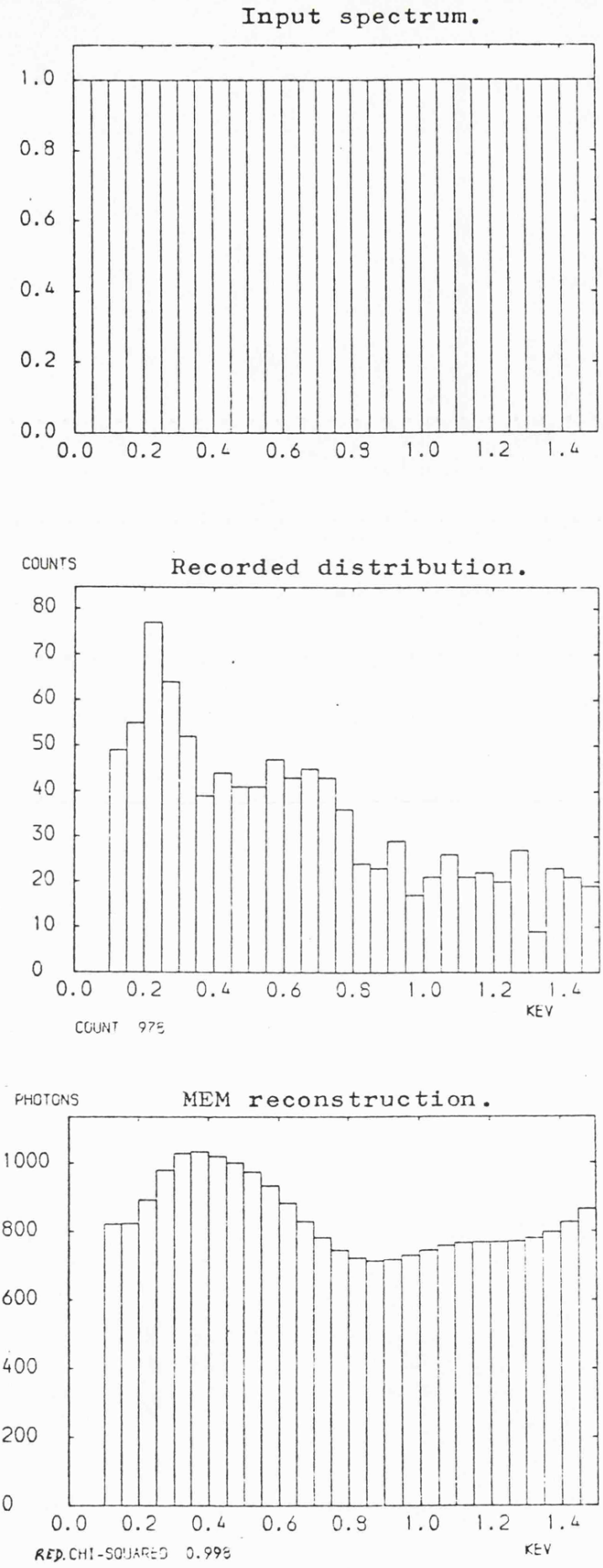


Figure 56. Simulation of a flat continuum spectrum with MEM reconstruction. Although modulation is present, it is well within the statistical fluctuations expected from the estimated recorded count.

the best estimate of the number of counts received from each bin can be calculated and hence the approximate significance of each bin can be gauged. Despite the obvious limitations of the spectra obtained, there is a distict difference between the hard and the soft regions of the Loop.

The maximum entropy solutions are, in a real sense, the best estimates of the source spectra that can be coaxed from the data without making any assumptions about the physics of the source. The small peak at low energies present in the hard spactrum is probably due to noise counts and possibly uv which has 'leaked' into the counter and this 'peak' is probably also present in the soft spectrum. Apart from that anomalous feature, both spectra are characterised by a single peak which is shifted by about 0.15 keV in the harder region. Any physical interpretation of the hardness map need only account for this simple behaviour since there is no evidence for finer structure in the data from the regions used.

Further studies made by Kayât, reference 15, showed that smooth spectra resulted from most regions of the hardness map. However the western limit exhibited a rather large bump at about 0.9 keV, illustrated on figure 57. Adjustment of the instrument response calibration figures within reasonable limits had very little effect on the appearance of this bump and it was decided to test its significance using Monte Carlo techniques. The data set shown on figure 52 was perturbed using a random number generator and the MEM algorithm used to produce another estimate of the source spectrum. This was carried out

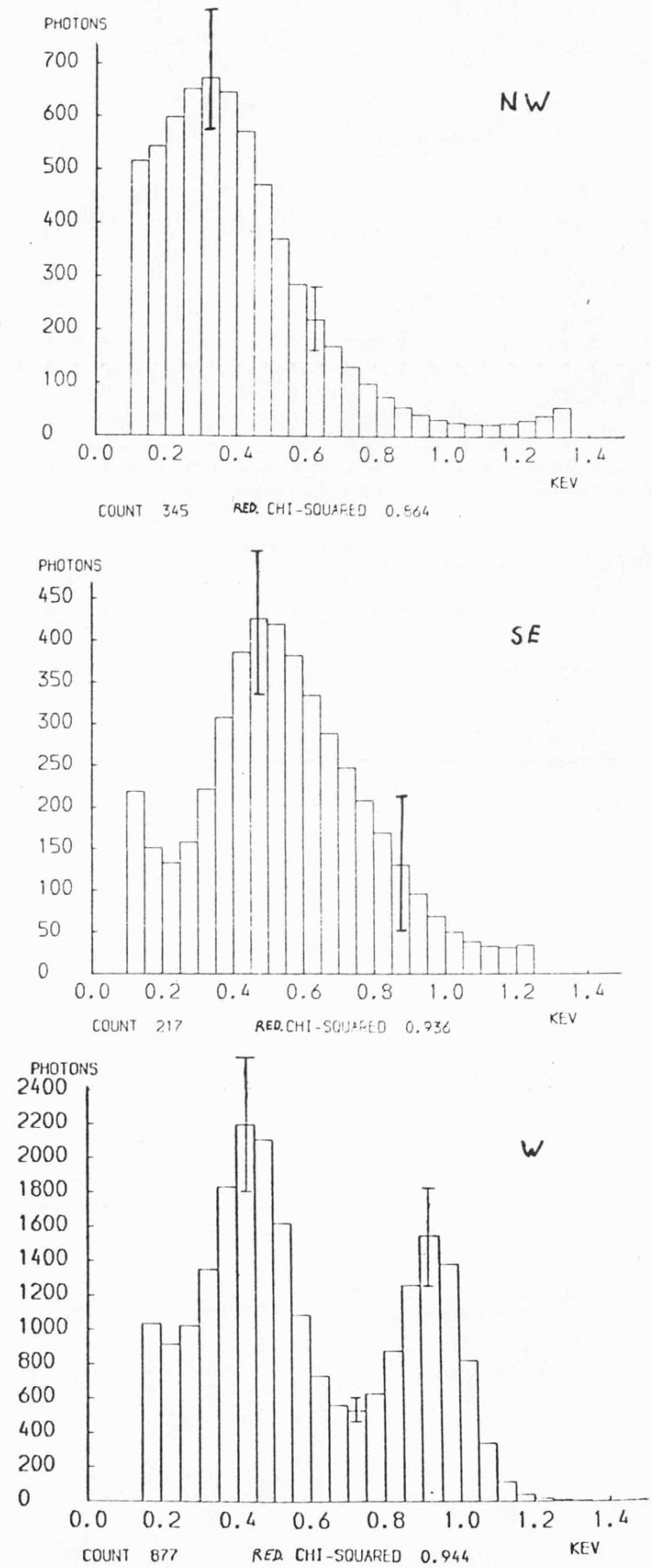


Figure 57. MEM reconstruction corresponding to figure 52. Error bars show rms fluctuations calculated from 12 independent perterbations.

using 12 independent statistical samples and the mean and rms fluctuation of bins in the solution calculated. The results are given in figure 57. The feature at 0.9 keV is clearly significant under this test and the derived rms values are slightly smaller than a crude estimate of the rms fluctuations derived from the number of counts which correspond to features in the solution, that is $\sigma_j^2 \simeq \hat{G}_j$.

5.4 Conclusion to Part III.

Soft X-ray spaectral analysis using proportional counters is hampered by the poor resolution due to insufficient charge, the strongly modulated efficiency function and the fact that the pulse height distribution is related to the source spectrum by a non-linear integral transform. An analysis scheme which allows for all instrument degradations and is not dependent on a source spectrum model is presented here, based on the maximum entropy method.

Parts I and II concentrated on linear systems analysis but the general theory extends to handling non-linear systems. The MEM can never 'beat ' the counting statistics limit and the solutions can be analysed safely providing the counting statistics are kept in mind. If a solution is found with a reduced chi-squared $\simeq 1$, then pure noise features are not expected in the solution. However, as indicated by the flat field simulation, remnants of the instrument response can remain and can be misleading. The estimate of the count from the j th bin $\hat{G}_j$ can be used as the statistical σ_j^2 value when

fitting model spectra to $\hat{G}_j$. This assertation was tested by Monte Carlo techniques and $\hat{G}_j$ was found to be a slightly pessimistic estimate of the significance of features, presumably because of correlation between adjacent bins.

The way in which the MEM algorithm handled the very severe efficiency functions of the MIT/ Leicester payload, including C_k and Ni_1 absorption edges, was quite impressive. The solutions $\hat{G}$ were smooth across these absorption features although a residual modulation did remain in the flat field simulation.

The MEM algorithm developed was used by Kayât (1979), reference 15, for detatiled spectral analysis of the Cygnus Loop. This revealed an anomolous emission feature on the western limit at about 0.9 keV. Exhaustive tests were carried out to see whether this feature could be an artifact of the instrument calibration or the MEM algorithm, but it refused to go away.

Although further development of the method described here is required, especially with regard to the statistical stability of the solutions, it promises to be another powerful tool in pulse height spectrum analysis.

PART IV

THE CALIBRATION OF BRAGG CRYSTAL X-RAY SPECTROMETERS.

CHAPTER 6. THE INSTRUMENT RESPONSE OF CRYSTAL SPECTROMETERS.

6.1 Introduction to Part IV.

The projects so far reported have centred on the theory of imaging instruments in X-ray astronomy, including a slight digression towards the spectral information provided by proportional counters in Part III. Crystal spectrometers based on the Bragg law of X-ray diffraction by crystals:

$$n\lambda = 2d_n \sin \Theta_n \qquad (6.1)$$

have been used in X-ray analysis since the Bragg spectrometer was introduced by Bragg and Bragg (1915), reference 25. The theory of spectrometers based on equation (6.1) has reached a high degree of sophistication and it is impossible to present every facet of such instruments here. However, as was made apparant in the imaging projects, a full understanding of instrument response is necessary if the best use is to be made of data obtained.

It is hoped that a spectrometer will measure the energy spectrum G(E) of a small portion of X-ray emitting sky with high energy resolution and good photometric calibration. A good estimate of G(E) will only be obtained if the instrument response is fully understood and accurate calibration data is available. The work reported here forms part of a long term project being carried out within the X-ray astronomy group at the University of Leicester to provide calibration data for crystals that have or will be used in astronomical observation. The work consists of both a theoretical study of crystal response

and measurement of crystal diffraction parameters for comparison with theoretical predictions.

6.2 The Bragg spectrometer.

An excellent analysis of the Bragg spectrometer is given by J.S. Thomsen in reference 26 and only the bare essentials will be included here. The angular dispersion of the X-ray spectrum is found by differentiating equation (6.1):

$$\frac{d\Theta_n}{d\lambda} = \frac{\tan \Theta_n}{\lambda} \qquad (6.2)$$

By setting $d\Theta_n = \hat{W}_n$, the width of the instrument window in the n th order reflection, an estimate of the resolution can be made:

$$\frac{\lambda}{d\lambda} = \frac{\tan \Theta_n}{\hat{W}_n} \qquad (6.3)$$

For good resolution a small window W_n is required. If the slit system defining the divergence of the X-ray beam in the dispersion plane is narrow, then $\hat{W}_n$ will be dominated by the crystal window but if the beam has a large angular divergence, then this will dominate the resolution. The precise form of the instrument window function must be considered to gain further insight into the spectrometer's performance.

Considering a Bragg spectrometer to be set for a nominal wavelength of $\bar{\lambda}$, the power output will be represeted by the function:

$$F(\bar{\lambda}) = \int_0^\infty W(\bar{\lambda} - \lambda, \lambda)\ J(\lambda)\ d\lambda \qquad (6.4)$$

where $J(\lambda)$ is the input spectrum power $ster^{-1}$ unit λ^{-1} on a wavelength scale and $W(\bar{\lambda}-\lambda,\lambda)$ is the window function, which is a function of the difference $\bar{\lambda}-\lambda$ and generally a slow function of λ. The window function W is the kernel of the instrument transformation relating the output spectrum to the input spectrum and because the results are initially obtained in terms of angles rather than wavelengths, it is easier to express equation (6.4) using angles. The angular position of the crystal can be denoted by β, where $\beta = 0$ corresponds to the peak wavelength λ_o. Then z can be defined as:

$$z = \left(\frac{d\beta}{d\lambda}\right)_{\lambda_o} (\lambda-\lambda_o) \simeq \beta_\lambda \qquad (6.5)$$

where β_λ is the setting for a wavelength λ. Using equation (6.1), $\lambda_o \rightarrow \theta_{no}$ and $\lambda \rightarrow \beta_\lambda$ and equation (6.4) becomes:

$$F(\beta) = \int_{-\infty}^{\infty} W(\beta-z)\, J(z)\, dz \qquad (6.6)$$

The low integration limit has been set to $-\infty$ for convenience with very little effect due to the fast convergence of the integral. The specrometry equation (6.6) is analogous to the imaging equation (1.1) presented in Part I but the direction of research in this spectrometry project is different. In imaging, the concern was how the data was related to equation (1.1) and how, given the form of the instrument kernel, the best use could be made of the data. The accent now is on studying the form of the window function itself in order to obtain the best instrument and to provide accurate calibration of the window function for use in X-ray spectral analysis.

As mentioned above, the window function has two components. Firstly the angular intensity distribution incident on the crystal, which is set by the slit system in front of the crystal and the angular and/or spatial distribution of the source. The function $G(\alpha,\Psi)$ defines this distribution where α is the angle with the reference direction θ_{no} in the dispersion plane and Ψ is the angle of divergence perpendicular to the dispersion plane. Equation (6.6) can then be written as:

$$F(\beta) = \iiint P_n(\beta-z-\alpha+\varepsilon(\Psi,\delta_1))\ G(\alpha,\Psi)\ J(z)\ d\alpha\ d\Psi\ dz \qquad (6.7)$$

where $\varepsilon(\Psi, \delta_1)$ is an error correction introduced by divergence of the beam perpendicular to the dispersion plane and the tilt error of the crystal δ_1 (The angle between the normal to the reflecting planes within the crystal and the dispersion plane). The function $P_n(\theta)$ is the crystal window function and for unpolarised incident radiation it represents the average for the two component radiations. As originally expressed in equation (6.4), $P_n(\theta)$ is also a function of θ_{no} (or λ_o), the reference position, but over a small range $P_n(\theta,\lambda_o)$ will not vary very much. However over the entire angular range $P_n(\theta,\lambda_o)$ for most crystals will vary quite considerably. P_n is purely a function of the crystal diffraction process and not of instrument geometry. Therefore study of P_n for a variety of crystals can provide calibration data for use in many spectrometer configurations. The effect of the $\varepsilon(\Psi,\delta_1)$ term will not be considered here since it is important in alignment rather than calibration. At present

it is sufficient to state that $G(\alpha,\Psi)$ can usually be adequately represented by the product $g(\alpha)\ h(\Psi)$ and that providing $h(\Psi)$ is narrow and δ_1 is small, then the modification to the window function and peak position shifts can be kept small compared with other instrument effects.

When a single emission line of wavelength λ_o is observed, the total energy recorded will be the integral of $F(\beta)$ over a suitable range. This integration is usually achieved by rotating the crystal at constant velocity through the position of the line in the dispersed spectrum. $J(z)$ will contain a single, very narrow peak at z_o. Integrating equation (6.7) over z will give:

$$F(\beta) = I_{\lambda_o} \iint P_n(\beta - z_o - \alpha + \varepsilon(\Psi,\delta_1))\ g(\alpha)\ h(\Psi)\ d\Psi\ d\alpha \qquad (6.8)$$

Providing $h(\Psi)$ is narrow, integration over Ψ will yield:

$$F(\beta) = I_{\lambda_o}\ \bar{h} \int P_n(\beta - z_o - \alpha + \bar{\varepsilon}(\delta_1))\ g(\alpha)\ d\alpha$$

where $\bar{h}$ is a constant of the slit system and $\bar{\varepsilon}(\delta_1)$ is a correction term due to tilt. Further, providing the function $g(\alpha)$ is much wider than the crystal window, equation (6.8) can be integrated in β:

$$E_{\lambda_o} = I_{\lambda_o}\ \bar{h} \iint P_n(\beta - z_o - \alpha + \bar{\varepsilon}(\delta_1))\ d\beta\ g(\alpha)\ d\alpha \qquad (6.9)$$
$$= I_{\lambda_o}\ \bar{h}\ \bar{g}\ R_c^{n}(\lambda_o)$$

Equation (6.9) holds providing the integration range of β $\theta_1 \rightarrow \theta_2$ includes all the energy from the emission line. The output power E_{λ_o} is clearly related to $\bar{h}\ \bar{g}$, which is purely a function of the slit system geometry and $R_c^{n}(\lambda_o)$ which

is a function of the crystal and is known as the integrated reflectivity:

$$R_c^n(\lambda) = \int_0^{\pi/2} P_n(\theta,\lambda)\, d\theta \qquad (6.10)$$

To summarise, the Bragg spectrometer response is characterised by two separable functions, the slit system or collimator $G(\alpha,\psi)$ and the crystal window function $P_n(\theta,\lambda)$. Crystal alignment and divergence perpendicular to the dispersion plane affect the response but only to a small degree if the instrument is well aligned. Full calibration of a spectrometer can only be achieved if $G(\alpha,\psi)$ and $P_n(\theta,\lambda)$ are known accurately and this research concentrates on the form of $P_n(\theta,\lambda)$ for three crystals; Langmuir-Blodgett lead stearate multilayers, gypsum and beryl.

CHAPTER 7: THE THEORETICAL FORM OF CRYSTAL WINDOW FUNCTIONS.

7.1 The results of the dynamical theory of diffraction by perfect crystals.

Soon after the discovery of the diffraction of X-rays by crystals a detailed theory for the interaction was developed. Three physicists were responsible for the work; C.G. Darwin (1914), reference 27, P.P. Ewald (1916), reference 28, and M. Von Laue (1931), reference 29. The combined efforts of these men using different approaches produced the Dynamical theory, which is very powerful and satisfying. It has inspired many subsequent workers in the field and now supports a vast body of research, however only the results relevant to this project will be quoted here since many excellent texts exist on the subject, for example R.W. Janes (1962), reference 30 and L.V. Azároff (1974), reference 31.

The most important result of the theory as far as this report is concerned is known as the Prins Function. J.A. Prins (1930), reference 32, adapted Darwin's theory to predict the window function for perfect crystals including atomic absorption. Although Prins used Darwin's approach, the resulting expression is the same adapting Ewald's method of solving Maxwell's equations for an electromagnetic wave travelling in a medium with periodic refractive index. The Prins function has the form:

$$\frac{|\underline{E}_H|^2}{|\underline{E}_0|^2} = |b| \left(\zeta \pm (\zeta^2 - 1)^{\frac{1}{2}}\right)^2 \qquad (7.1)$$

with $$b = \frac{\gamma_0}{\gamma_H}$$

and $$\zeta = \frac{b\Delta\theta \sin 2\theta + \frac{1}{2}\Gamma F_0(1-b)}{\Gamma |P| \ (F_H \ F_{\bar{H}})^{\frac{1}{2}} \ |b|^{\frac{1}{2}}}$$

The left hand side is the ratio of the diffracted power to the incident power expressed using the electric wave vectors $\underline{E}_H$ and $\underline{E}_0$, H denotes the Miller indices of the crystal planes involved in the reflection, γ_0 and γ_H are the directional cosines of the incident and reflected rays referred to the surface normal (surface assumed planar), θ is the Bragg angle as given by equation (6.1) and $\Delta\theta$ is the direction from θ, P is the polarization constant which equals 1 for the σ state in which vectors $\underline{E}_0$ and $\underline{E}_H$ are perpendicular to the corresponding wave vectors $\underline{k}_0$ and $\underline{k}_H$ and equals cos 2θ for the π state in which $\underline{E}_0$ and $\underline{E}_H$ lie in the plane defined by $\underline{k}_0$ and $\underline{k}_H$. F_0, F_H and Γ describe the structure and composition of the crystal. The charge density of the crystal is assumed to be periodic and therefore expressible in Fourier series form within a unit cell:

$$\rho(\underline{r}) = \frac{1}{V} \sum_H F_H \exp(-2\pi i \ \underline{R}_H \cdot \underline{r}) \qquad (7.2)$$

V is the volume of the unit cell and R_H is a reciprocal lattice vector defined by H, the Miller indices of the reflection. F_H is known as the structure factor. Using the properties of Fourier series:

$$F_H = \int_V \rho(\underline{r}) \exp(2\pi i \ \underline{R}_H \cdot \underline{r}) \, d\underline{r} \qquad (7.3)$$

Assuming that the atoms within the crystal occupy well

defined 'points', the integral form of (7.3) can be reduced to a summation over all atoms within the unit cell:

$$F_H = \sum_m f_m \exp\left(2\pi i \, \underline{R}_H \cdot \underline{r}_m\right) \qquad (7.4)$$

the n th atom being situated at $\underline{r}_m$. f_m is known as the atomic scattering factor. F_H is therefore composed of a scattering term from each atom f_m, which is a function of the atomic species and not the crystal structure, weighted by a phase factor dependent on the Miller indices H and the position of the atom within the unit cell. Finally, Γ is given by:

$$\Gamma = \frac{e^2}{\varepsilon_0 \, m \, w^2 \, V} \qquad (7.5)$$

where e is the electron charge, ε_0 is the permitivity of free space, m is the electronic mass and w is the angular frequency.

7.2 Theoretical calculation of the Prins Function.

Although the Prins Function (7.1) is complicated, most of the parameters are easy to obtain and the only difficulty is the calculation of the F's as given by the summation equation (7.4). The calculation of the structure factors is very much a bootstrap affair since crystalographers use X-ray reflections to determine atomic positions within the unit cell. The position and relative strength of many reflections off many planes can yield an electron density map (or its Fourier domain counterpart) of the unit cell. By using the known composition of the crystal and theoretical estimates of the atomic scattering factors

f_m, the positions r_m can be estimated. Any calculation of F_H for prediction of the crystal window function must take these calculated positions and use equation (7.4). Because the atomic positions are found using many reflections, the atomic scattering factors need not be predicted very accurately, especially if the crystal is fairly simple. Most analyser crystals are fairly simple and nobody would dream of using RNA for X-ray spectroscopy! When calculating the window function for a particular reflection as a function of wavelength, the atomic scattering factors will invariably be the stumbling block since all the other parameters are provided by crystalographers.

The physics of the photon-crystal interaction is contained in the atomic scattering factor f and a great deal of involved theory is required to yield useful expressions for f. However it is not necessary to expound the complete development to understand the structure of the theory. The atom behaves as if f 'free' or Thomson electrons are coincident with the atomic nucleus. A correct description of the scattering process can only be achieved using quantum mechanics and so the notion of a classical charge density distribution used in equations (7.2) and (7.3) is superceded by wave functions and associated quantum mechanical relations. Since the practical application of the theory rather than any theoretical development is intended here, only the relevant results will be quoted and the reader is referred to the many excellent texts on quantum mechanics to provide a complete derivation.

If an atom, initially in a state described by the

wave function Ψ_0, is perturbed by radiation of angular frequency w_i which is much greater than w_{0m}, the angular frequency corresponding to the energy difference between states Ψ_0 and Ψ_m, then the ratio of the total scattered intensity, incoherent and coherent, to the intensity scattered by a free electron is given by:

$$\frac{I}{I_T} = \sum_m \left| \int \Psi_m^* \Psi_0 \sum_k e^{[(2\pi i/\lambda)\underline{s}.\underline{r}_k]} \, d\underline{r} \right|^2 \qquad (7.6)$$

where the summation over m includes all possible wave-functions or final states of the atom and k denotes the individual electrons. $\underline{s}$ is a vector normal to the lattice planes performing the reflection and has magnitude 2 sin θ where 2θ is the scattering angle. The reciprocal lattice vector R_H has magnitude $1/d_H$ where d_H is the inter-planar distance for planes with Miller indices H and using the Bragg equation (6.1), the correspondence between (7.3) and (7.6) is apparant. Equation (7.6) can be simplified using the orthogonality and normalisation of the wave functions Ψ_0 and Ψ_m:

$$\frac{I}{I_T} = \int |\Psi_0|^2 \left| \sum_k e^{[(2\pi i/\lambda)\underline{s}.\underline{r}_k]} \right|^2 d\underline{r} \qquad (7.7)$$

The scattering factor for coherent scattering alone is given by:

$$f_0 = \sum_k \int |\Psi_0|^2 \, e^{[(2\pi i/\lambda)\underline{s}.\underline{r}_k]} \, d\underline{r}_k \qquad (7.8)$$

The quantity $|\Psi_0|^2$ is directly analogous to $\rho(\underline{r})$ in the classical formulation and quantum mechanics provides, in principal, a method for calculating the electron density

function for the atom. Unfortunately solving the wave equation to find the ground state wave function Ψ_0 is, in general, impossible and approximation methods must be used, but before these are discussed a further complication must be dealt with; namely anomalous dispersion.

Equation (7.8) is a good approximation for f providing the photon frequency is much greater than any absorption edge frequencies due to the bound nature of the electrons but an extra frequency-dependent term is significant for all incident energies below and just above absorption edges:

$$f = f_0 - \sum_k g_{k0} \frac{w_{k0}^2}{w_{k0}^2 - w_i^2} \qquad (7.9)$$

w_{k0} is the angular frequency corresponding to the energy difference between the ground state Ψ_0 and excited state Ψ_k, w_i is the incident frequency and the summation over k ensures all possible excited states are accounted for. The coefficient g_{k0} corresponds to the classical oscillator strength and is given by:

$$g_{k0} = \frac{2m}{3\hbar w_{k0}} w_{kn}^2 |r_{0k}|^2 \qquad (7.10)$$

where the co-ordinate matrix element r_{0k} is given by:

$$r_{0k} = \int \Psi_k^* \, \underline{r} \, \Psi_0 \, d\underline{r} \qquad (7.11)$$

and an average over all polarisations has been made assuming spherical symmetry, thereby introducing the factor of 1/3 in equation (7.10). The reduction to the matrix element (7.11) uses the dipole approximation, in which the wavelength λ is assumed much greater than the atomic

dimensions.

The major contribution to the summation in (7.9) will be from unbound electron states rather than unoccupied bound states and since the unbound states form a continuum, the summation will be an integral and g_{k0} will represent the oscillator density between w_{k0} and $w_{k0}+dw_{k0}$. The integral has a singularity when $w_{k0} = w$ but this can be dealt with by introducing a decay time for the initial state. In the absence of scattering, the time dependence will be of the form exp $(-i\ w_0\ t/h)$. This will be modified to exp $(-i(w_0-i\delta)\ t/h)$, where w_0 is the energy of the ground state and δ is a very small number, which gives rise to the natural width of the emission line corresponding to the transition $0 \rightarrow k$ and has the same effect as a damping constant in classical oscillator resonance theory. The part of the resulting contour integral close to and just above the singularity will yield an imaginary term:

$$i\Delta f''(w_i) = i\pi w_i\ g_{k0} \tag{7.12}$$

where k_i is the continuum state which gives $w_{k_i 0} = w_i$. This term will be independent of δ unless $w_i - w_{k0} \lesssim$ natural line width. The remaining part of the integral will give rise to the real term:

$$\Delta f'(w_i) = P \int_0^\infty \frac{g_{k0}\ w_{k0}^2}{w_{k0}^2 - w_i^2}\ dw_{k0} \tag{7.13}$$

where the integration will cover all continuum and unoccupied bound states and P indicates the principal part of the integral (neglecting the singularity).

The atomic scattering factor f including anomalous

dispersion has the form:

$$f = f_0 + \Delta f'(w_i) + i\Delta f''(w_i) \qquad (7.14)$$

with the terms on the right hand side defined using the wave functions of the electrons as indicated above. f_0 contains only the coherent scattering terms and not the total as expressed by equation (7.6) since crystal diffraction is only concerned with the coherent component. The incoherent processes will be present but in nearly all crystal diffractions they introduce an insignificant correction into the linear absorption coefficient. The imaginary term gives rise to absorption and is in fact responsible for the photoelectric cross-section, the relation between the two being given by:

$$\sigma(\hbar w) = \Delta f'' \frac{4\pi e^2}{mcw} \qquad (7.15)$$

The cross-section $\sigma(\hbar w)$ is therefore directly expressible in terms of the oscillator density g_{k0}, which provides a very useful method for comparing the theoretical calculations of the scattering factor with experimental results other than crystal diffraction measurements.

Since f_0, $\Delta f'$ and $\Delta f''$, the components of the scattering factor f, are all directly expressible in terms of the atomic wave functions, the crux of calculating f is finding a good enough estimate of the wave functions for all the electrons which have an appreciable effect on the scattering. Various sets of numerical wave functions have been computed for many atomic species using the self-consistent field method. For outer orbitals it is found that methods

which try to allow for electron exchange using a Slater type exchange potential give considerably different results from calculations which neglect the necessity for an anti-symmetric wave function Ψ_0. Furthermore, relativistic effects become appreciable when the atomic number is large and since many of the crystals under study contain heavy atomic species, the use of relativistic calculations is possibly desirable. An interesting comparison between the f_0's calculated using four different atomic models was made by Cromer (1965), reference 32, but for this research the f_0 values calculated by Cromer and Waber (1965), reference 33, and $\Delta f' + i\Delta f''$ values calculated by Cromer and Liberman (1970), reference 34, were used. Reference 33 uses Dirac-Slater wave functions relying on the Slater (1951) exchange potential $\sim\rho(\underline{r})^{\frac{1}{3}}$, reference 35. Unfortunately Kohn and Shan (1965), reference 36, have given evidence that this potential should be $\frac{2}{3}$ as great as Slater's original estimate. This adjustment expands the atom and therefore decreases the scattering factor. The relativistic corrections included in D-S wave functions contract the atom and for $z \gtrsim 55$ they appreciably increase the scattering factor. The decision to use scattering factors from reference 33 instead of the many alternatives was taken to give consistency between f_0 and the anomalous terms calculated in reference 34. Inspection of equation (7.8) shows that f_0 is a function of $\sin\theta/\lambda$ and since crystal diffraction is governed by Bragg's equation (6.1), f_0 is conveniently independent of wavelength. However it is a function of diffraction order n. Reference 33 provided a convenient analytic fit using 9 coefficients:

$$f_0(\sin\theta/\lambda) = \sum_{i=1}^{4} a_i \exp(-b_i \sin^2\theta/\lambda^2) + c \qquad (7.16)$$

This was programmed and used to generate the f_0 terms for all atomic types.

Reference 34 uses the relativistic quantum theory to develop expressions for $\Delta f'$ and $\Delta f''$. However apart from minor correction terms, the results are precisely the same as those given by the non-relativistic approach. The cross-section of every orbit of elements up to Cf ($z = 98$) are given at 10 energies, 5 of which are well used crystalographic lines, the other 5 being specified by the Gauss-Legendre quadrature used to perform the necessary integration in equation (7.13) for $\Delta f'$. A FORTRAN program was provided by the authors to calculate $\Delta f'$ and $\Delta f''$ at any energy within the crystalographic range 2.29 - 0.599 Å. The cross-section at a specified energy was found by interpolation of the spot energies given and $\Delta f''$ calculated directly. The integral for $\Delta f'$ was then performed using the 5 predetermined energies and the interpolated photo-electric cross-section at the specified wavelength. Since the complete range of orbitals was covered by the cross-section data, the program was modified to cover any incident wavelength in the range 100 - 0.599 Å. The interpolation error is larger for the longer wavelengths but acceptable for the purposes of this research. The orbital cross-sections provided were calculated using Dirac-Slater-Kohn-Shan wave functions utilising the modification to Slater's exchange potential mentioned above and therefore the complete scattering factor produced by the combination of references 33 and 34 is pleasingly derived

from the same atomic model apart from the slight exchange modification.

The computer routines based on the work of Cromer et al as described above were interfaced with existing crystal diffraction programs which were developed at Leicester by Lewis and Underwood, reference 37, and Maksym, reference 38. The set of routines was modified so that anomalous dispersion calculations of various types could be used and the results compared. The main output of the calculations is the Prins function and its integral $R_c(\lambda)$, but many associated parameters and functions are also available relating the theoretical calculations to the measurable quantities which will be described in the next section.

Before the inclusion of the scattering factor calculation method described here, the anomalous terms were estimated using a sem-empirical method described by James, reference 30, and developed by Parratt and Hempstead (1954) reference 39 in which the photoelectric absorption coefficient as a function of energy is approximated by:

$$\sigma(\hbar w) = \left(\frac{w_{k0}}{w_i}\right)^n \sigma(\hbar w_{k0}) \qquad (7.17)$$

where the index n depends on the type of orbital under consideration. This power law fit is adequate for inner orbitals but very poor for higher orbitals in heavy elements. This is demonstrated by figure 58 which shows the absorption edge profiles of various orbitals calculated by Cromer and Liberman. Although the Dirac-Slater atomic model contains many approximations, the

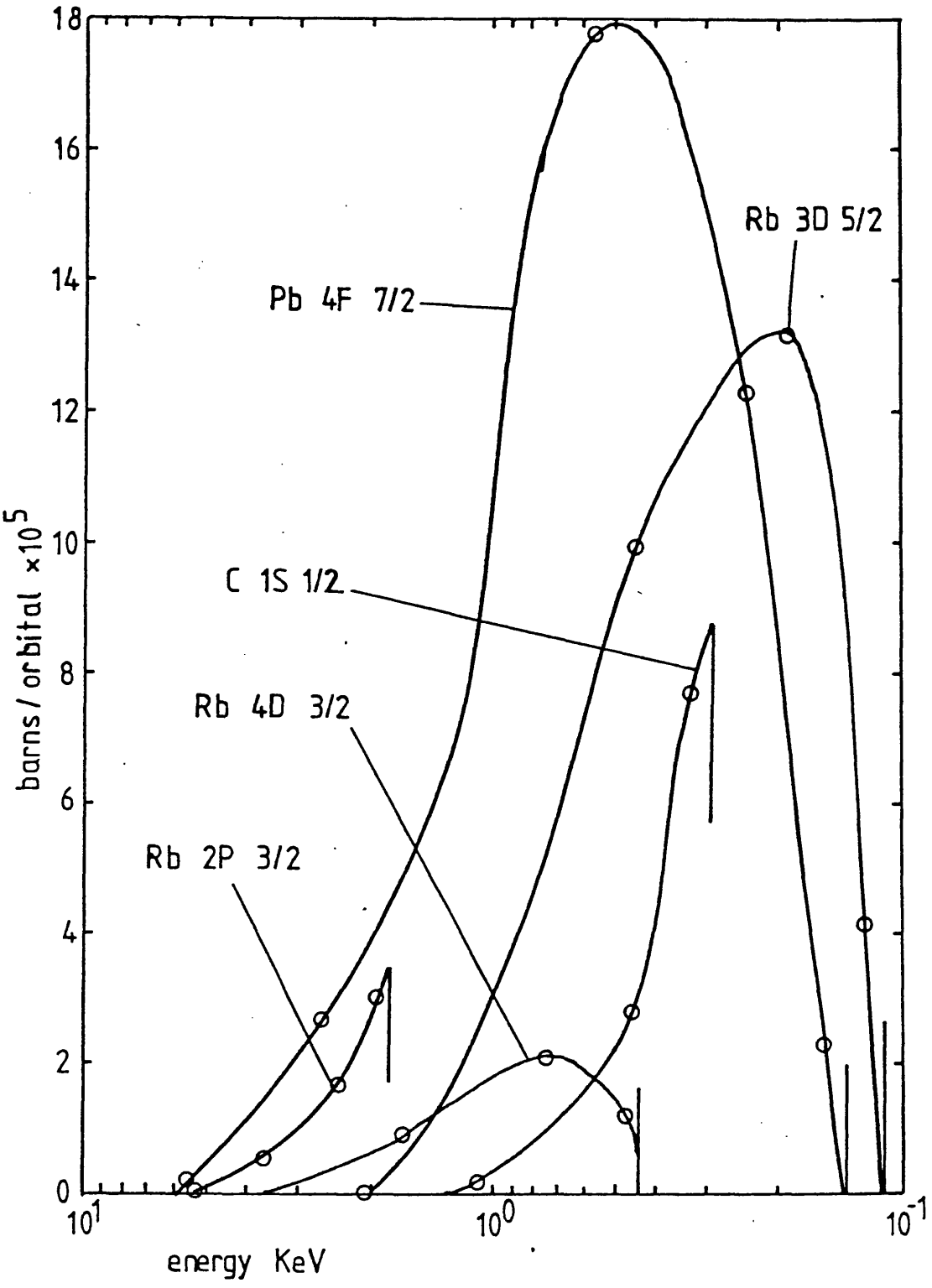


Figure 58. Absorption edge profiles calculated by Cromer and Liberman.

scattering factors derived from it are independent of empirical methods and have the advantage of being complete, covering all atomic species and a large wavelength range. Theoretical calculations for specific crystals are given in Chapter 8 but before presenting the results of both theory and experiment, the experimental techniques used must be described.

7.3 Measurement of the Prins Function.

Sucessful measurement of a crystal window function relies on careful control of the divergence and spectral content of the X-ray beam. By far the best way to achieve reliable measurements of both the shape and integral of the Prins function is the use of two crystal reflections, the first to provide a monochromatic beam which is reflected off the second, test crystal. The only drawback with this procedure is the polarization caused by the monochromator but this can be corrected for.

The best configuration is known as the 1-1 mode, which is illustrated by figure 59. Both crystals are made of the same material and mounted in two tables with parallel axes. A constant X-ray beam is reflected off the first to give a monochromatic beam of well defined wavelength λ_0. The second crystal is placed in this beam and slowly rotated at constant angular velocity through the position where the two sets of crystal planes are parallel. The complete beam off the second crystal is detected and the count rate recorded as a function of time giving a 'two crystal rocking curve' as shown in figure 59.

The two crystal rocking curve can be written in

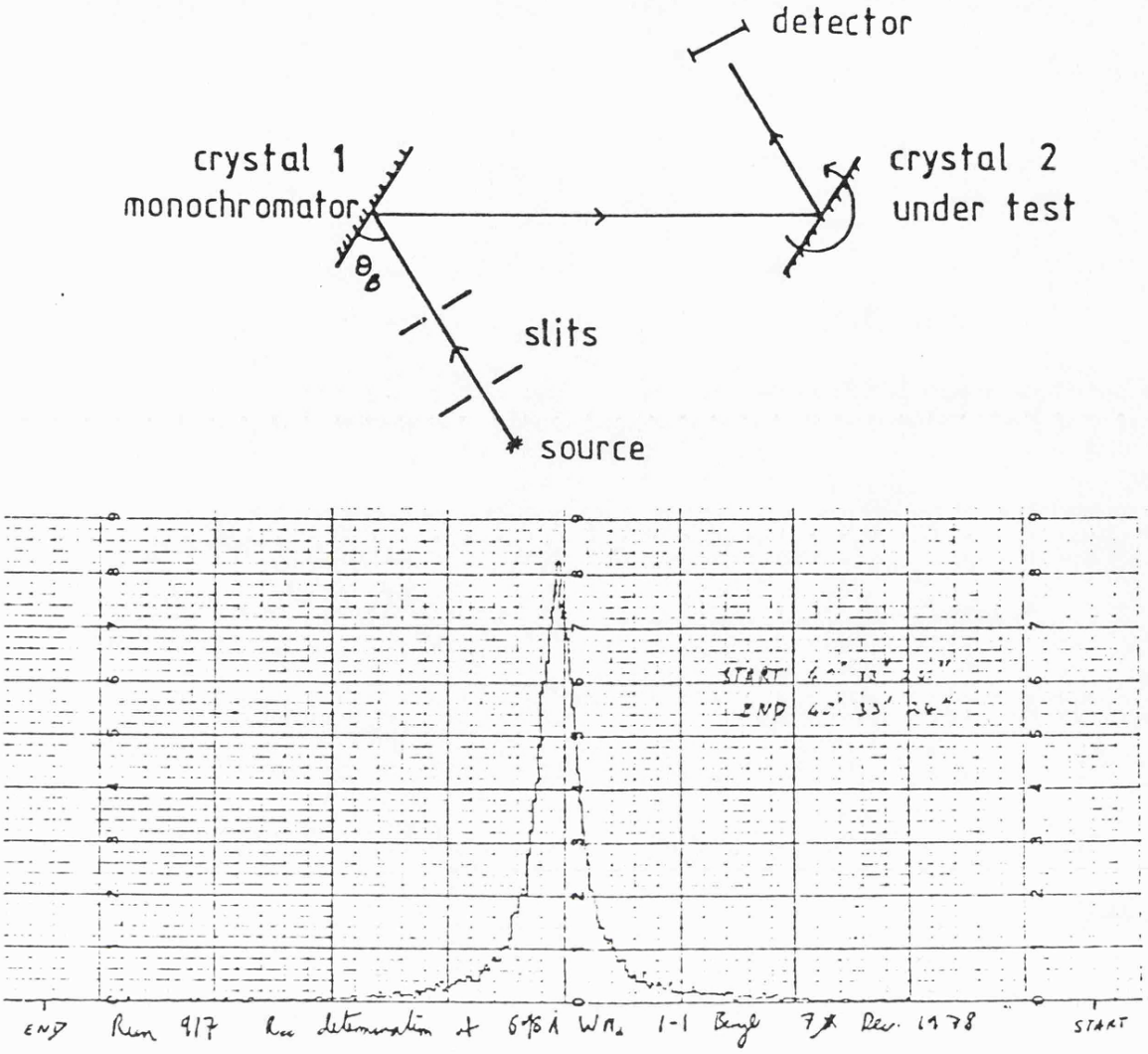


Figure 59. The 1-1 mode with a typical rocking curve.

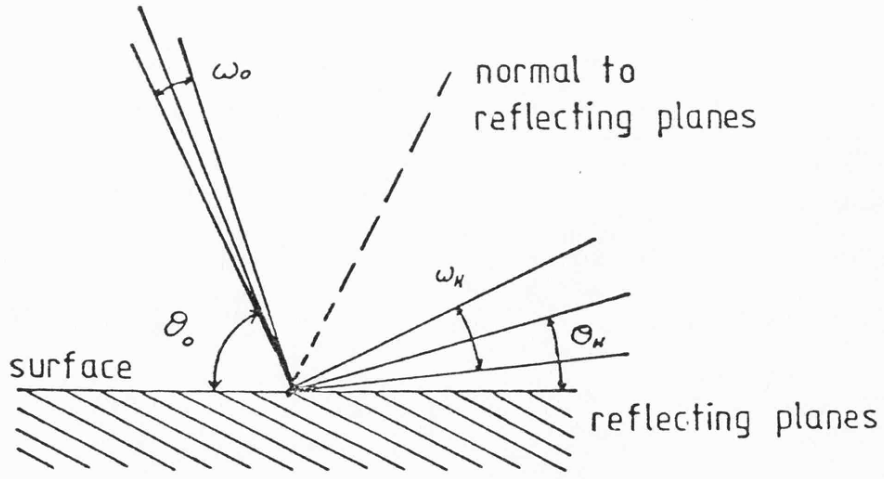


Figure 60. An asymmetric Bragg relection.

integral form (see reference 26 for derivation) like equation (6.7) for the single crystal case:

$$f_2(\beta) = \tfrac{1}{2} \int\int\int_0^{\pi/2} \sum_p P_n^p(\xi_1 - z - \alpha)\ P_n^p(\beta - z - \alpha + \xi_2)$$

$$g(\alpha,\Psi)\ J(z)\ d\alpha\ d\Psi\ dz \qquad (7.18)$$

The choice of limits is arbitrary providing they are far enough away from peak. $\sum_p$ indicates a sum over both polarizations, β is the angle rotated by the second crystal away from the reference position, (α,Ψ) represents the position of a ray with the beam relative to the reference ray and z is defined by equation (6.5). The range of angles over which the response is appreciable is normally very small and the angular spread of the beam can be very large compared to it. $g(\alpha,\Psi)$ can therefore be made constant over the entire range of the integral which contributes to $f_2(\beta)$ and $g(\alpha,\Psi)$ can be approximated by $g(\Psi)$. Transforming the variables for the integration over the beam in the dispersion plane gives:

$$f_2(\beta) = \tfrac{1}{2} \int J(z) \int g(\Psi) \int_0^{\pi/2} \sum_p P_n^p(v)\ P_n^p(\beta + v - \xi_1 + \xi_2)$$

$$dv\ d\Psi\ dz \qquad (7.19)$$

If the divergence perpendicular to the dispersion plane is kept small ($g(\Psi)$ narrow) and providing the tilt errors are small :

$$f_2(\beta) = \tfrac{1}{2} \int J(z)dz \int g(\Psi)\ d\Psi \int_0^{\pi/2} \sum_p P_n^p(v)\ P_n^p(\beta + v)dv \qquad (7.20)$$

Therefore the rocking curve has the form of the convolution of the crystal window function scaled by the spectrum

J(z) and the slit function $g(\Psi)$, providing the divergence of the beam defined by $g(\alpha,\Psi)$ is large in α compared to the width of the Prins function and small in Ψ and the crystal tilts are kept small. In fact if both crystals are tilting by exactly the same amount in the same direction, the error terms causing broadening of the rocking curve cancel and the remaining terms are independent of Ψ and only cause a shift in the peak position. (For verification of this see reference 26 and the formulae given in section 7.4.)

The 1-1 mode therefore provides the self-convolution of the single crystal window function free from spectral distortion. This convolution is easily calculated from the theoretical form of the Prins function and the two crystal rocking curve can then be compared with the theoretical result. Integrating the rocking curve to give the total energy received over all angles gives:

$$\int_{\Theta_1}^{\Theta_2} f_2(\beta)d\beta = \tfrac{1}{2}\int J(z)dz \int g(\Psi)d\Psi \int_{\Theta_1}^{\Theta_2} \int_0^{\pi/2} \sum_p P_n{}^p(v)$$

$$P_n{}^p(\beta+v)\ dv\ d\beta \qquad (7.21)$$

The order of integration can be reversed if the integration range $\Theta_1 \rightarrow \Theta_2$ goes well outside any appreciable contribution to the integral and comparison with equation (6.10) reveals that the β and v integration simply yield $R_c{}^p(\lambda_0)$, the integrated reflectivity for polarization P:

$$E_{2\lambda_0} = \tfrac{1}{2}\int J(z)dz \int g(\Psi)d\Psi \sum_p R_c{}^p(\lambda_0)\ R_c{}^p(\lambda_0) \qquad (7.22)$$

where λ_0 is the wavelength determined by the setting of the first crystal relative to the source through Bragg's

equation. But the beam power of radiation hitting the second crystal can be expressed using equation (6.9) in terms of the integrated reflectivity:

$$E_{\lambda_0} = \tfrac{1}{2}\int J^{\sigma}(z)dz \int g(\psi)d\psi \; R_c^{\sigma}(\lambda_0) + \tfrac{1}{2}\int J^{\pi}(z)dz \int g(\psi)d\psi \; R_c^{\pi}(\lambda_0) \tag{7.23}$$

Dividing (7.22) by (7.23) and assuming the source to be unpolarized, ($J^{\sigma}(z) = J^{\pi}(z)$) gives:

$$\frac{E_{2\lambda_0}}{E_{\lambda_0}} = \frac{R_c^{\sigma}(\lambda_0)^2 + R_c^{\pi}(\lambda_0)^2}{R_c^{\sigma}(\lambda_0) + R_c^{\pi}(\lambda_0)} \tag{7.24}$$

The left hand side is denoted $R_{cc}(\lambda_0)$. Using the polarization ratio defined by:

$$k(\lambda_0) = \frac{R_c^{\pi}(\lambda_0)}{R_c^{\sigma}(\lambda_0)} \tag{7.25}$$

and knowing that the integrated reflectivity for unpolarized radiation is:

$$R_c(\lambda_0) = \frac{R_c^{\pi}(\lambda_0) + R_c^{\sigma}(\lambda_0)}{2} \tag{7.26}$$

equation (7.24) reduces to:

$$R_{cc}(\lambda_0) = \frac{2R_c(1 + k(\lambda_0)^2)}{(1 + k(\lambda_0))^2} \tag{7.27}$$

The two crystal rocking curve is actually measured by scanning the second crystal at a constant angular velocity w giving a curve of count rate as a function of angle. The total power of the beam hitting the second crystal is easily measured as I counts/sec using the same detector. This count rate can be converted into a count per radian

using the angular velocity w rads/sec; $E_{\lambda_0} = I/w$. $E_{2\lambda_0}$ is the total count under the rocking curve C and therefore:

$$R_{cc}(\lambda_0) = \frac{C\ w}{I} \tag{7.28}$$

Using the theoretical form of the Prins function $P(\theta,\lambda_0)$, the two crystal rocking curve can be predicted and compared to the experimental curve using three parameters; the integrated reflectivity $R_{cc}(\lambda_0)$ as given by equation (7.28), the peak reflectivity $P_{cc}(\lambda_0)$ and the full width at half maximum $W_{cc}(\lambda_0)$. $P_{cc}(\lambda_0)$ is simply found by dividing the count rate recorded at the peak of the rocking curve by the incident count rate I. Making measurements at various angular positions corresponding to emission lines on the source will build a complete picture of the performance of the two crystals as a function of λ_0. Providing there is good agreement between theory and measurement, the performance of a single crystal can be deduced from the two crystal rocking curves using theory. There are two ways in which theory can be further tested to instill confidence in this extrapolation. The form of the polarization constant (7.25) can be changed by using a monochromator of much larger 2d spacing than the crystal under test. The analysis of the resulting rocking curve can be carried out as above using two different Prins functions. The two crystal integrated reflectivity measured, $R_{ab}(\lambda_0)$, is related to $R_b(\lambda_0)$ using the polarization constants for the two crystals $k_a(\lambda_0)$ and $k_b(\lambda_0)$:

$$R_{ab}(\lambda_0) = \frac{2R_b(\lambda_0)\,(k_a(\lambda_0) + k_b(\lambda_0))}{(1 + k_a(\lambda_0))(1 + k_b(\lambda_0))} \qquad (7.29)$$

This is fully explained by Evans, Leigh and Lewis, reference 40, but in short the use of a large 2d monochromator reduces the Bragg angle at the first reflection and therefore reduces the polarization in the beam hitting the test crystal. The polarization correction is therefore small and an estimate of R_b can be made which is not very sensitive to the theoretical parameters $k_a(\lambda_0)$ and $k_b(\lambda_0)$.

The use of a monochromator with 2d different from the test crystal introduces dispersion into the rocking curve formula and the response is no longer free from spectral distortion. To obtain further information on the shape of the single crystal rocking curve, an asymmetric Bragg reflection must be used in which the crystal surface is cut at an angle to the crystal planes. The dynamical theory predicts that the reflection profile from an asymmetrically cut crystal is scaled depending on the angle of cut. Figure 60 shows such a reflection in which the incident and reflected beams make angles Θ_0 and Θ_H with the surface. The parameter $\Delta\Theta$ in equation (7.1) is the difference between the incident beam and the Bragg angle. The dynamical theory predicts that:

$$\Delta\Theta_H = \frac{\sin\Theta_0}{\sin\Theta_H}\,\Delta\Theta_0 \qquad (7.30)$$

and the directional cosine parameter b will be affected by asymmetric cutting. The scales of the incident and diffracted curves from the configuration illustrated in figure

60 will therefore differ, the incident beam being narrower than the reflected beam. Replacing the second beam in figure 59 by the asymmetric reflection in figure 60 will have the effect of scanning the response of the first crystal with a very narrow beam. The rocking curve convolution integral is then dominated by the angular response of the first, symmetric reflection. Several workers have used such an arrangement to get an explicit mapping of the single crystal rocking curve, notably Kohra, reference 41. Unfortunately a particular asymmetrically cut crystal is only usable over a very narrow wavelength range and for most purposes the symmetric two crystal rocking curve provides enough information for comparison with theory. In most crystals, imperfections in the lattice structure dominate the angular response if a large area of surface is used. The Prins function then represents the limiting performance that can be expected from a particular crystal type and the two crystal rocking curve will indicate perfection of the crystal lattice.

7.4 The crystal spectrometer research programme within the X-ray astronomy group, Leicester.

Sections 7.2 and 7.3 form the basis of the research. A comprehensive set of computer programs is available to provide not only single crystal parameters but also theoretical two crystal rocking curve parameters for direct comparison with experiment. The Prins based calculations are inherently concerned with ideal or perfect crystal lattices. Real crystals are not genereally anywhere near perfect and the W_{cc}, P_{cc} parameters measured can

be in disagreement with theory. The limiting case of a very imperfect crystal is modelled by the so-called mosaic crystal, in which it is imagined that the crystal is composed of many, very small, perfect crystals lying slightly misaligned to each other so that they are decoupled from each other when diffraction takes place. The integrated reflectivity of such a mosaic crystal as defined by equation (7.28) is derived by James, reference 30, and can be written using the same notation as used above for the Prins function:

$$R_{cc}(\lambda_0)_{\text{mosaic}} = \frac{e^2\pi^2 |F|^2 V}{\mathcal{E}_0 w^2 m \sin\theta} \frac{(1+\cos^2\theta)}{2\sum_j \Delta f_j''} \left[2 \frac{1+\cos^4 2\theta}{(1+\cos^2 2\theta)^2} \right] \tag{7.31}$$

where θ is the Bragg angle corresponding to λ_0, F is the structure factor for the unit cell and $\sum_j$ scans over all contributions to the imaginary part of the scattering factor. The $1+\cos^2\theta/2$ term is a polarization correction to $R_c(\lambda_0)$ and the term in [ ] is a polarization correction for the two crystal case (unpolarized source assumed). The integrated relectivity for the mosaic model is easily calculated along with the Prins result and it acts as a good guide to the perfection of the crystal lattice, independent of the functional shape of the rocking curve which can be distorted by beam divergence and the source spectrum. The important parameters produced by computer calculations are $W_{cc}(\lambda_0)$, $P_{cc}(\lambda_0)$, $R_{cc}(\lambda_0)$, $R_{cc}(\lambda_0)_{\text{mosaic}}$ and $R_c(\lambda_0)$ for any order reflection required.

Experimental two crystal rocking curves are obtained

using a two crystal X-ray spectrometer which has been described by B. Leigh, reference 42. and more recently by R. Hall, reference 43. The later reference gives full details of the alignment procedure carried out before the measurements reported here were made but for completeness the final alignment achieved is given below.

The table bearings were checked for tightness and the axis set to the gravity vector to within ± 1 arc sec. The control table carrying the crystal drive micrometers was set perpendicular to the gravity vector to ± 10 arc secs. The crystal face defined by 3 mounting balls was set to contain the rotation axis on both tables using an optical flat and autocollimator. The tolerance achieved was about ± 20 μm translation and ± 2 arc secs tilt. The slits defining the beam were aligned to intersect the axis to ± 0.1 mm and set orthogonal to the axis to ± 10 arc mins. The source and detector drive and phase monitor systems were set to achieve twice the angular velocity of the crystal tables centred about a reflected ray defined by the slit system. The source target distribution was studied using pinhole photography to ensure the illumination of the beam would be reasonably uniform.

All the above are general alignments, common to all crystals. Further adjustments which are crystal specific need to be made before good quality rocking curves can be obtained. The geometrical window caused by divergence and tilt errors is given in reference 26:

$$W(t) = \int g(\psi)\, \delta\left(t+\tfrac{1}{2}(\delta_1^2-\delta_2^2)\tan\theta - \frac{2\delta_1^2}{\cos\theta} - 2\delta_1\delta_2\tan\theta + \frac{\psi\delta_1}{\cos\theta} + \frac{\psi\delta_2}{\cos\theta}\right)d\psi$$

The terms independent of ψ cause a shift of the peak and are unimportant for this work. The remaining two terms yield a broadening of the peak due to divergence and tilt error. If W_b is the slit function FWHM, the geometric window will have a FWHM W_g given by:

$$W_g = W_b \frac{(\delta_1 + \delta_2)}{\cos\theta} \qquad (7.32)$$

Because the crystals face one another, $\delta_1 + \delta_2$ is in fact the difference between the tilts. Equation (7.32) is used below to assess the beam tolerance that is required to achieve good results. Except for cleaved crystals, the crystal planes are not necessarily parallel to the crystal surface. Any discrepancy must be removed so that the planes of the two crystals are parallel and reasonably aligned to the rotational axis, otherwise the two crystal rocking curve will be broadened. This is achieved by measuring a series of rocking curves with different positions of the second tilt micrometer. The parallel position is indicated by the setting which yields the largest peak reflectivity and smallest FWHM. In practice, the turning point is shallow and the tilt setting is not critical if the rocking curves are > 10 arc secs FWHM. It is easy to achieve a tilt error of less than $\pm$ 10 arc mins. The geometric window then becomes important unless very fine tilt adjusments are made. As explained in section 7.3, the slits defining the beam must be set intelligently if good results are required. Using slits of 1 mm height (defining beam perpendicular to the dispersion plane) and assuming 10 arc mins tilt error, a geometrical window of

1 arc sec at $\theta_B = 30^o$ will be produced. This is well within the required tolerance for most crystals. The divergence of the beam in the dispersion plane must be set much wider than the rocking curve. The third slit, which prevents the beam missing the second crystal, is the major problem. Setting all slits to 1 mm width produces a W_b of 4' 55". In general the slit must be widened until the beam is just contained on the second crystal so that accurate R_c measurements can be made. For most crystals it is possible to achieve a beam width about 10 times the W_{cc} value and in fact W_{cc} is not affected until $W_b \leq 3\ W_{cc}$.

In practice, a compromise must be struck between geometric distortions and beam power. In some cases a distortion-free rocking curve must be sacrificed in order to obtain good integrated reflectivity measurements, which are totally insensitive to misalignments providing all the beam is contained on the second crystal. The errors introduced by alignment etc. in the following results will be discussed as they are reported. For many cases, the statistical errors involved in the measurements introduced by photon counting are far in excess of any systematic errors. The ratio of the efficiency of the counter when measuring the incident beam to when measuring the reflected beam off the second crystal is significant and has been measured at various spot energies over the entire instrument energy range; 0.1 - 5.0 KeV. The intensities can therefore be corrected for this. The only other major problem is the detector background count, mostly due to cosmic rays. Accurate determination of the background was made before and after every run to try and keep the

background errors below the inherent statistics in the signal. The procedures used for making the two crystal rocking curve measurements are well tried and adequately documented in references 42 and 43.

There is a continuous interplay between the practical and theoretical sides of the research programme. The three crystals under study have helped improve the theoretical calculation and encouraged a careful study of the experimental set-up. The results now being produced are considered to be of excellent quality.

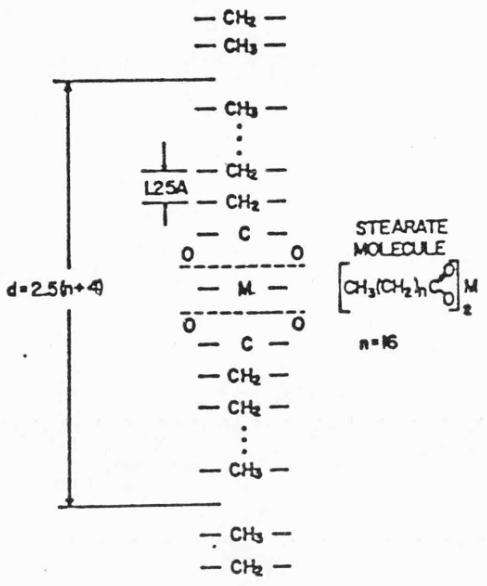
CHAPTER 8: EXPERIMENTAL AND THEORETICAL RESULTS FOR LANGMUIR-BLODGETT LEAD STEARATE MULTILAYERS, GYPSUM 020 AND BERYL $10\bar{1}0$.

8.1. Langmuir-Blodgett lead stearate multilayers.

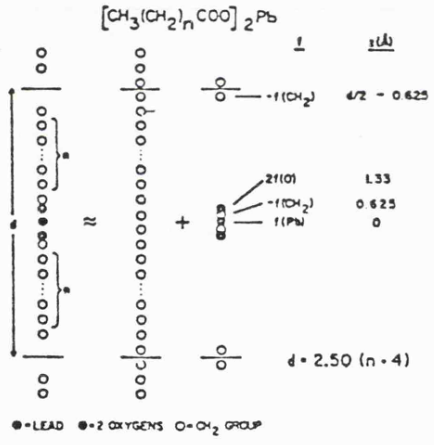
The spectral region 1 - 100 Å bridges the gap between the domain of grating spectrometers used in the XUV band and Bragg analyser crystals used in X-ray spectrometry. Within the ultra-soft X-ray band gratings are difficult to make and use while Bragg analysers must be found with large lattice plane spacing d if crystal spectrometers are to function. Langmuir-Blodgett multilayers of lead stearate have a d-spacing of 50 Å and are therefore potentially useful over the ultra-soft X-ray region.

The construction and structure of Langmuir-Blodgett multilayers are extensively covered in the literature. An excellent introduction to the field is given by Charles (1968), reference 44. B.L. Henke has published a great deal on the construction and use of such 'crystals', notably reference 45 for the construction and the technical report by Henke, Perera and Ono, reference 46, which provides a theoretical treatment with experimental results. Unfortunately the methods used for obtaining the measured integrated reflectivities are not fully explained and therefore it is impossible to gauge the accuracy of the measurements. However the theoretical model proposed by Henke et al provides a useful basis for this work.

Henke's model for the unit cell is illustrated by figure 61. It is based on experimental evidence from many sources and is not known to be contradicted by any available



A schematic for the molecule of the metal salt of a straight-chain fatty acid as is involved in the formation of Langmuir-Blodgett multilayers.



The one-dimensional structural model for the molecular "unit cell" for a multilayer formed by the salt of a straight-chain fatty acid. The scattering factor contributions have been divided into that part due to a continuous distribution of CH_2 groups and that due to the cation, M, two pairs of nearby oxygen atoms, and the negative contribution of two pairs of CH_2 groups, each spaced a d/2 distance apart--(the effect of which pairs cancel for odd-order diffractions).

Density 1.26.

Unit cell volume 1000 $Å^3$.

2d 100 Å.

Figure 61. The unit cell of Langmuir-Blodgett lead stearate multilayers, reference 46.

data. Theoretical crystal window parameters were calculated using this cell model and the dispersion calculations previously described. The crystals provided by Quartz et Silice through Nuclear and Silica Products Ltd. were known to have been made using 100 dipping cycles. This should have deposited 200 monolayers of lead stearate and therefore 100 d spacings because of the back to back nature of the structure (reference 44). The crystals were therefore not infinately thick and the results of the dynamical theory quoted above are not directly applicable to the experimental situation. However Henke et al, reference 46, give a very simple exposition of the modifications to the Prins function when there are only a finite number of layers. They use the Darwin-Prins argument but a similar result is yielded from the Ewald approach by putting in extra boundary conditions at the rear face of the crystal. Explicit expressions for the functions derived by this method are given by Zachariasen (1945). Zachariasen's equation 3.139 (reference 47, Dover edition) is rather complicated and unmanageable to say the least, but he does show that it reduces to the familiar Prins form given by(7.1)above. On the other hand,Henke et al find a rather more simple formula by making an approximation. They get the result:

$$I_N = I_\infty \left| 1 - e^{-Nx} \right|^2 \qquad (8.1)$$

where I_N is the intensity of the diffracted beam due to N layers and I_∞ is the normal Prins results for an infinite number of layers. They give x in terms of the normal Prins parameters but these can be converted to the form used

above giving:

$$x = \frac{\pi^2}{\sin^2\theta_0}\Gamma \sqrt{|P|\, F_H\, F_{\bar{H}} + \left(\frac{\Delta\theta \sin 2\theta_0 \Gamma}{4\pi} - F_0\right)^2} \qquad (8.2)$$

Result (8.1) will hold providing $x \ll 1$. Considering expression (8.2), the square root for $\Delta\theta = 0$ is simply a measure of the scattering power of a unit cell. $\pi^2\Gamma/\sin^2\theta_0$ can be rewritten using the Bragg equation:

$$\frac{\pi^2\Gamma}{\sin^2\theta_0} = \frac{\pi^2 r_0 \lambda^2}{V4\pi^2}\left(\frac{4d^2}{\lambda^2 n^2}\right) = \frac{r_0\, d^2}{Vn^2} \qquad (8.3)$$

where r_0, the classical radius of an electron, equals 2.82×10^{-5} Å. The volume per electron will always be much larger than the product $r_0 d^2/n^2$ for any small order n and spacing d, therefore the approximation $x \ll 1$ holds good.

The computer routines were modified to incorporate the finite thickness correction given above and a set of parameters obtained for comparison with experimental results. The correction necessary for the mosaic model integrated reflectivity, when only a finite number of layers are used, is given by James, reference 30, as:

$$R_c(\lambda_0, t) = R_c(\lambda_0, \infty)\left(1 - \exp\left(\frac{-2u(\lambda_0)\rho t}{\sin\theta_0}\right)\right) \qquad (8.4)$$

where $u(\lambda_0)$ is the mass absorption coefficient of the crystal, ρ is the density and t the thickness. $u(\lambda_0)$ is easily expressible in terms of the imaginary part of the atom scattering factors and is computed along with the other parameters:

$$u(\lambda) = 2 \frac{e^2}{\varepsilon_0 me^2} \frac{1}{V} \sum_j \Delta f_j''(\lambda) \qquad (8.5)$$

This correction was not incorporated into the programs but calculated directly. It is interesting to note that the mosaic correction depends only on the absorption terms because the elemental crystals are decoupled whereas the perfect case in which the incident and diffracted beams are coupled throughout the crystal volume gives rise to an intensity profile which is modified by all the scattering terms of the atoms.

The calculated single crystal integrated reflectivity curves are shown in figure 62. Mosaic values are marked m and the Prins values P. The absorption feature of C_k electrons 44.7 Å and O_k electrons 23.3 Å are very prominant and the finite thickness corrections are important over the entire range of the crystal. These calculations predict a lower $R_c(\lambda)$ than reference 46 by a factor of ~ 0.6 which must be due to the dispersion terms, since the model used here is identical in all other respects. Figure 63 shows the peak reflectivities for single and two crystal rocking curves. The correction for 100 layers is very large below the carbon edge, indicating that the crystals are likely to be thickness rather than absorption limited in this region. Before presenting the widths of the crystal window functions it is instructive to consider a far simpler model of the diffraction process. A multilayer crystal with N layers acts like a diffraction grating with N elements. Ignoring the extinction and absorption, the diffraction profile obtained from such a

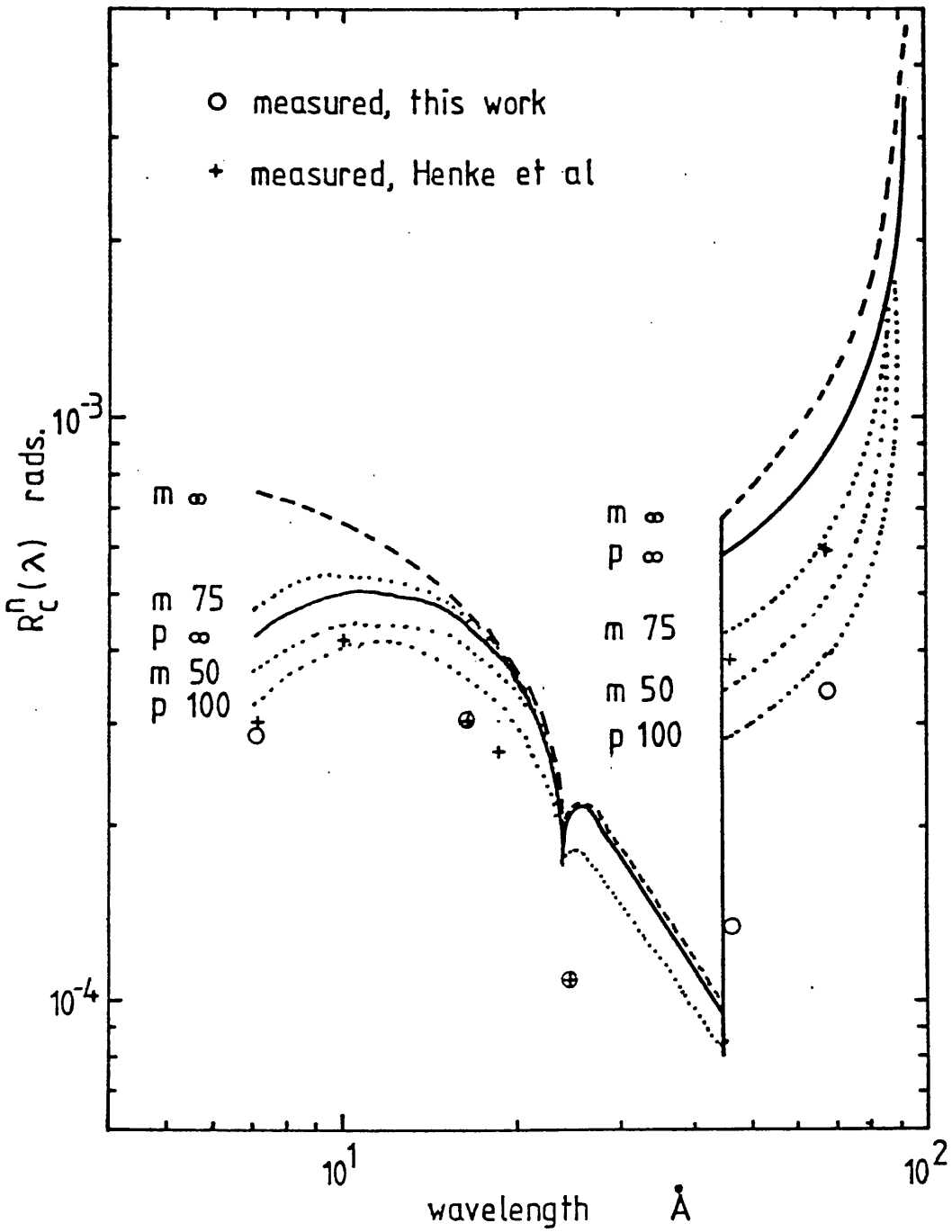


Figure 62. Lead stearate single crystal integrated reflectivity.

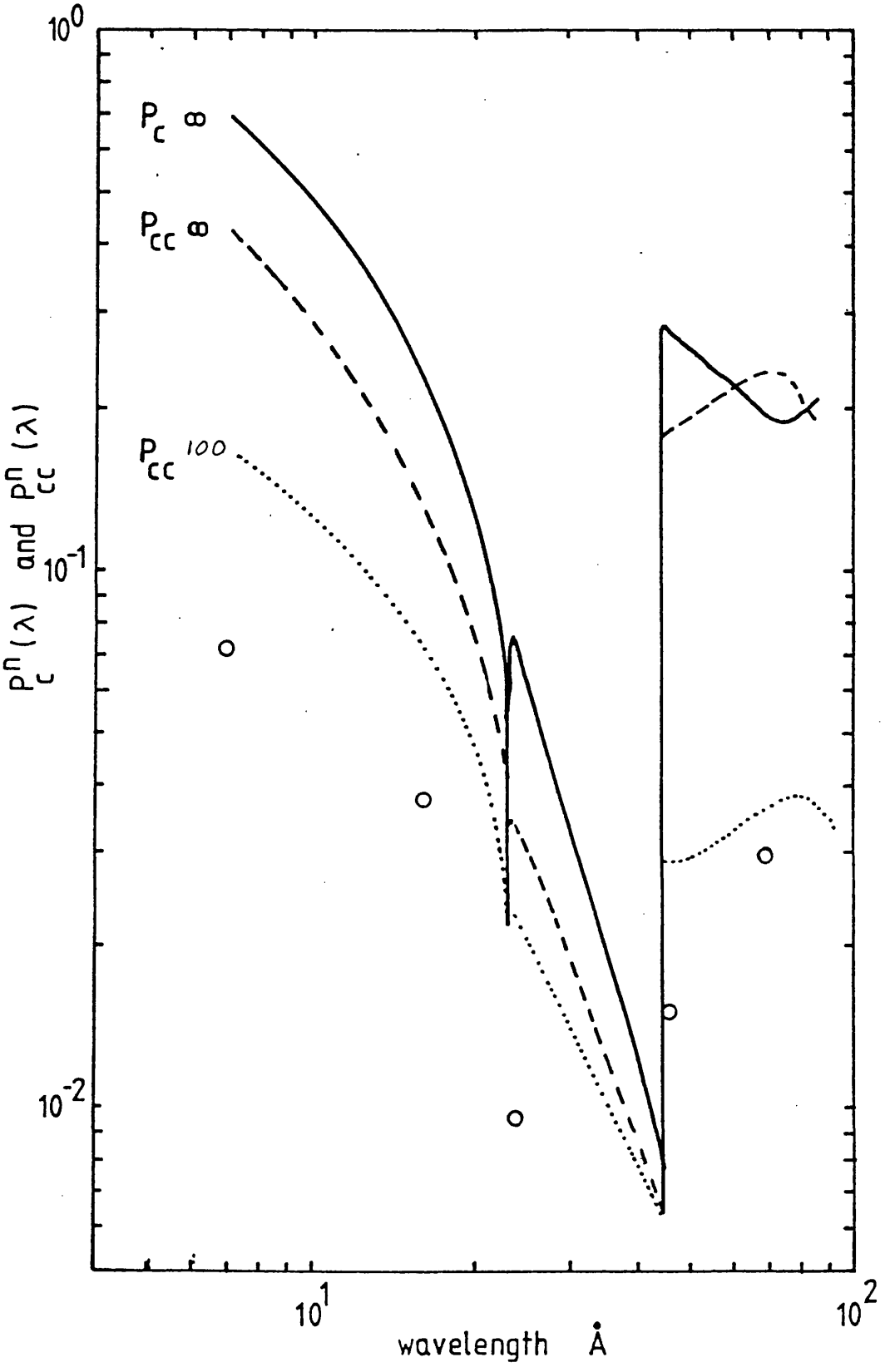


Figure 63. Lead stearate peak reflectivities.

grating has the well known form:

$$I(x) = \left(\frac{\sin Nx}{\sin x}\right)^2 \qquad (8.6)$$

where $x = \Delta\theta\ 2\pi d \cos\theta_0/\lambda$. The width is therefore approximated by the first zero at $\Delta\theta = \lambda/2Nd \cos\theta$. Using the dispersion equation (6.2) to substitute for $\Delta\theta$ gives:

$$\frac{\Delta\lambda}{\lambda} = \frac{\lambda}{2Nd \cos\theta \tan\theta} = \frac{1}{N} \qquad (8.7)$$

Equation (8.7) defines the resolution of an N element grating. If the crystal was infinitely thick, the resolution would be limited by the beam penetration set by absorption and extinction due to scattering into the diffracted beam. Ignoring the extinction, the resolution would be limited by the number of layers penetrated before the intensity dropped to 1/e of the initial value. The so-called 1/e depth is given by the linear absorption coefficient which is easily calculated from the mass absorption equation (8.5). Figure 64 shows the two crystal rocking curve FWHM and again the finite thickness correction is considerable. Figure 65 shows the single crystal FWHM for various models. Comparison of the Prins curves W_c and W_c 100 with the simple grating model curves 1/100 no-absorption layer limit and e^{-1}, the absorption limit, shows that below the carbon edge it is extinction limited. Furthermore there is considerable gain in using $N > 100$ at all wavelengths. Figure 66 presents the same information in terms of resolution $\Delta\lambda/\lambda$ so that the performance can be related to the quality of the output spectrum. Finally

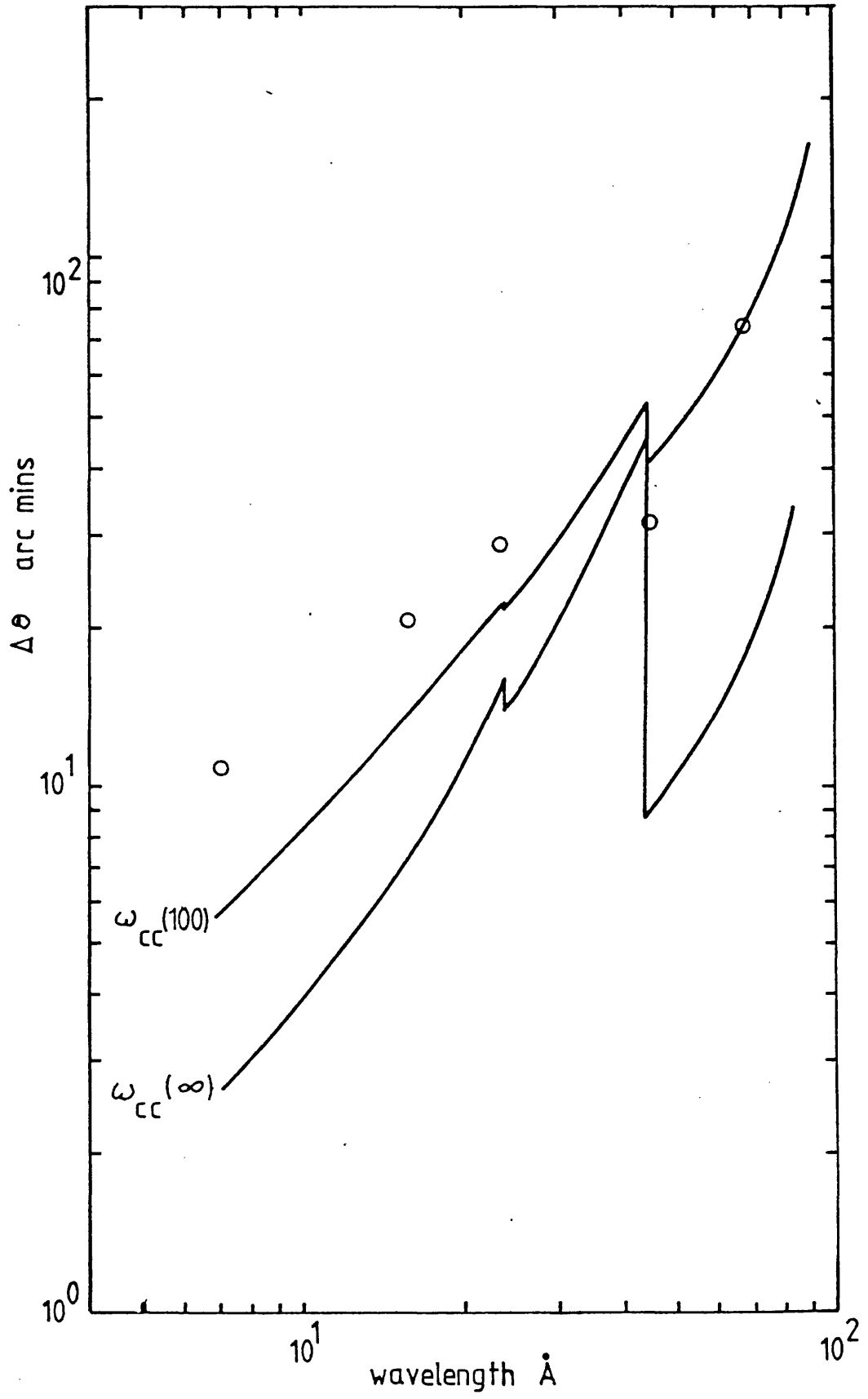


Figure 64. Lead stearate two crystal FWHM.

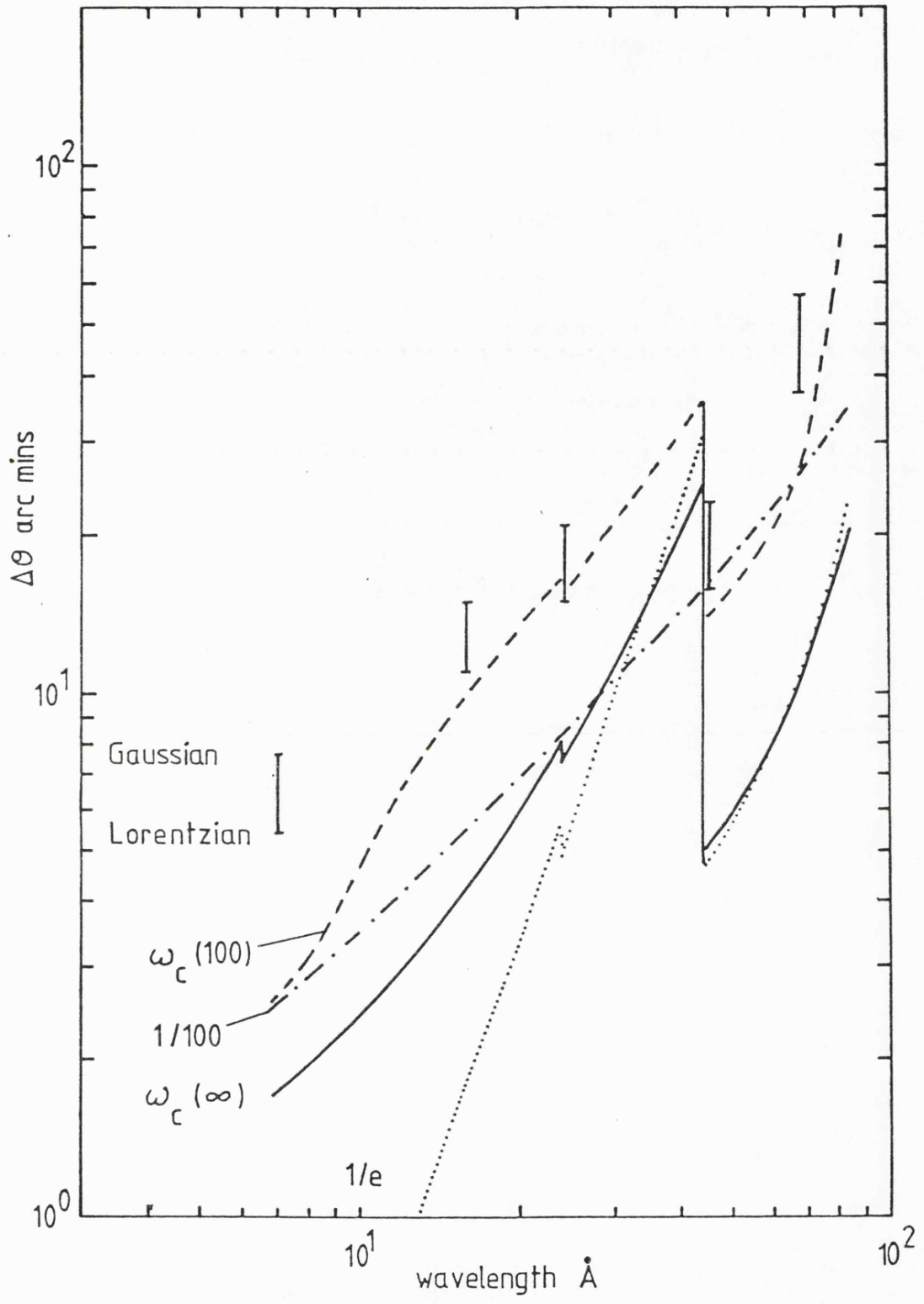


Figure 65. Lead stearate single crystal FWHM.

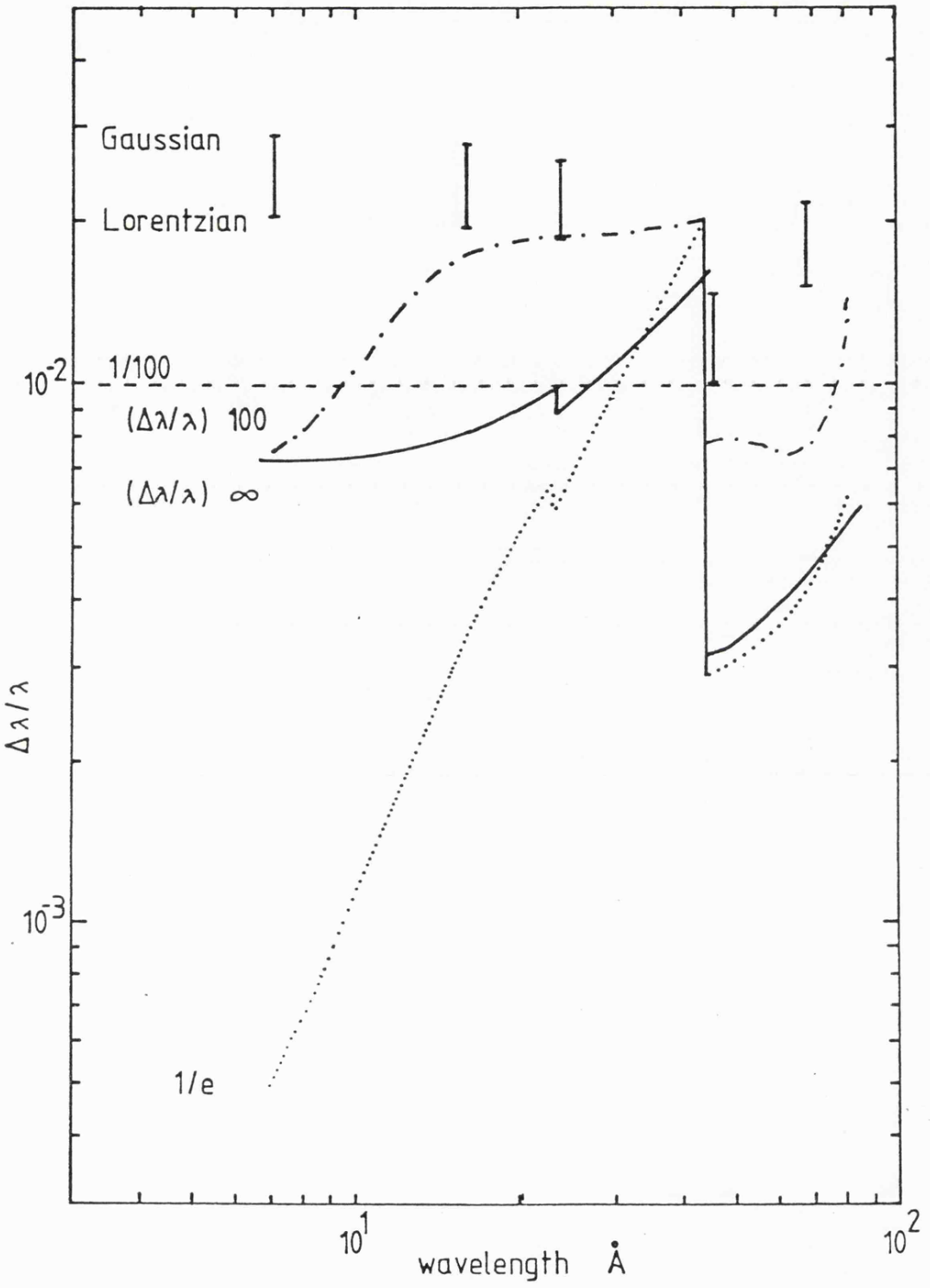


Figure 66. Lead stearate single crystal resolving power.

the single crystal integrated reflectivity for an infinite lattice in the first four orders is given in figure 67 so that the effect of higher order reflections can be estimated. It is interesting to note that there is an odd-even modulation, the odd numbers being far stronger than the even reflections. This is due to the gap in the unit cell model between the ends of the CH_2 chains. Henke et al measured the relative intensity of odd and even modes and found such a modulation, suggesting that such a gap did exist.

Two crystal rocking curve measurements were made at five emission lines using the Leicester facility. Unfortunately the largest dispersion plane divergence that can be achieved in the dispersion plane is about $\frac{1}{2}^{0}$ and therefore referring to figure 62 it is clear that $g(\alpha,\psi)$ will not be constant over a sufficiently large angle to prevent spectral broadening. The widths are therefore expected to be too large and the peaks somewhat depressed. However the integrated reflectivity measurements should be unaffected. Working at ultra-soft X-ray energies presented a few problems with the detector. The gain and discriminator levels had to be set very carefully to ensure that all the flux was included and background contamination was kept to a minimum. All fluxes were wavelength checked to ensure there was no second order contamination.

The $R_{cc}(\lambda)$ results were corrected using the Prins polarization correction calculated theoretically to yield $R_c(\lambda)$ values. The statistical errors on all measurements were all below $\pm$ 5%, including the background correction. If a mosaic correction had been used, there would be very

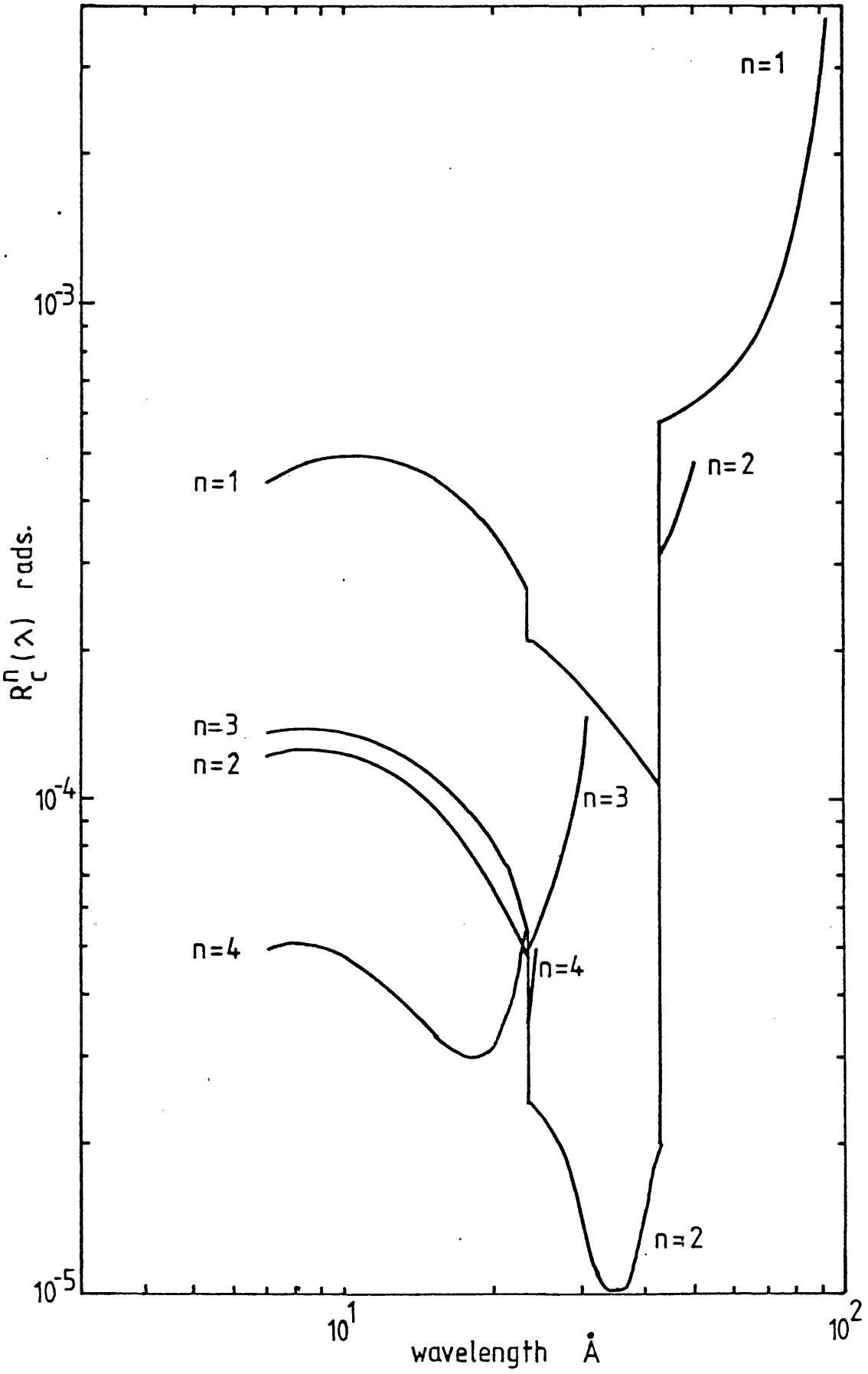


Figure 67. Lead stearate single crystal integrated reflectivities for the first 4 orders.

little difference because the crystals are so absorption dominated. The results are shown in figure 62, along with measurements by Henke et al, reference 46, using a 75 d layer crystal. Above the oxygen edge, all the experimental points agree reasonably well with the P 100 curve. The O_k result agrees with Henke's value but they are both well down on theoretical prediction for a perfect lattice. Below the C_k edge, Henke's values are strangely much larger than the present results. Both sets show anomolous behaviour below the oxygen edge and the present results also show anomolous behaviour just below the carbon edge. The B_k 67.6 Å and all results above the O_k edge suggest that the crystal is fairly well modelled by a perfect lattice, although it could be a mosaic with low efficiency due to inadequate substitution of Pb^{2+} ions in the acid when the monolayers were deposited.

The P_{cc} (λ) values obtained are plotted in figure 63. Although they are considerably smaller than $P_{cc}100$ theoretical values, they do show a strong absorption edge modulation as expected. The FWHM are plotted in figure 64. They are remarkably close to the $W_{cc}(100)$ theoretical curve when it is considered that the beam divergence was expected to cause a broadening. This suggests that the layers have been reasonably well deposited and are acting as a coherent grating rather than a decoupled mosaic.

The measured two crystal W_{cc} values were converted to equivalent single crystal W_c values assuming two functional models for the single crystal window functions; a Lorentzian profile which is known to be a very good approximation to the Prins function in the limit of high

absorption and a Gaussian which might be expected if the crystal was a mosaic. The resulting values are plotted as bars in figure 65, the top of each bar being the Gaussian value and the bottom the Lorentzian value. The values are in excellent agreement with the theoretical W_c 100 curve, especially considering that a broadening error is still present. The corresponding values are plotted in figure 66. Figure 69 provides a tabulated summary of the lead stearate multilayer results.

The overall agreement between the theoretical model and the measured results is fairly good. Although systematic errors do exist in the measurements, they are probably $< 10\%$ and in many cases less. The discrepancy at the two absorption edges is almost certainly real and could be due to three faults in the model. Firstly the anomolous dispersion calculation may be inadequate and fine structure of the edges may be giving the low results. This will be discussed in connection with the results from gypsum and beryl. Secondly there may be an incoherent scattering contribution which has not been allowed for, since coherent scattering is known to show anomalous behaviour at absorption edges. However this is unlikely. Thirdly the unit cell model may be incorrect. It was found that the shape of the $R_c(\lambda)$ curve just below the carbon edge was sensitive to the relative positions of the lead ions and the stearate chain. This does not, unfortunately, explain the vast difference between Henke's result and the present C_k value. There could be a genuine difference in the crystals used by Henke and those supplied by Quartz et Silice, but this also seems unlikely. The method of

measurement used here is generally free of such mosaic systematic errors and Henke's method is unfortunately not fully explained. The difference remains!

The results do have a brighter side since there is considerable evidence that the performance is limited by the number of layers deposited. A crystal subjected to several hundred dipping cycles should give better resolution and efficiency, especially below the carbon edge, provided the quality of the layers does not deteriorate. The combination of Henke's crystal model, Cromer and Liberman's anomalous dispersion calculations and the two crystal rocking curve measurements form a reasonably complete picture of currently available Langmuir-Blodgett lead stearate multilayers used as Bragg analysing crystals. In order to get a feel for their performance in practice, several single crystal scans were carried out using a beam divergence of 17' FWHM, similar to a commercial X-ray spectrometer used in the ultra-soft X-ray region. The source target was coated with a cocktail and the first and third slits set at 3.36 mm wide, 5 mm high. Figure 68 shows shows three scans through the O_k edge and C_k edge region. Numerous emission lines are visible and the change in crystal window width is clearly demonstrated although there is, of course, beam divergence and spectral line broadening present. The absorption edges are not visible although conditions were not optimum for finding the edges. Unfortunately even with a 'clean' target and narrow slits, the detection of the edges would be hindered by C_k and O_k contamination lines which are always present and the inherent mediocre resolution of the diffraction profile.

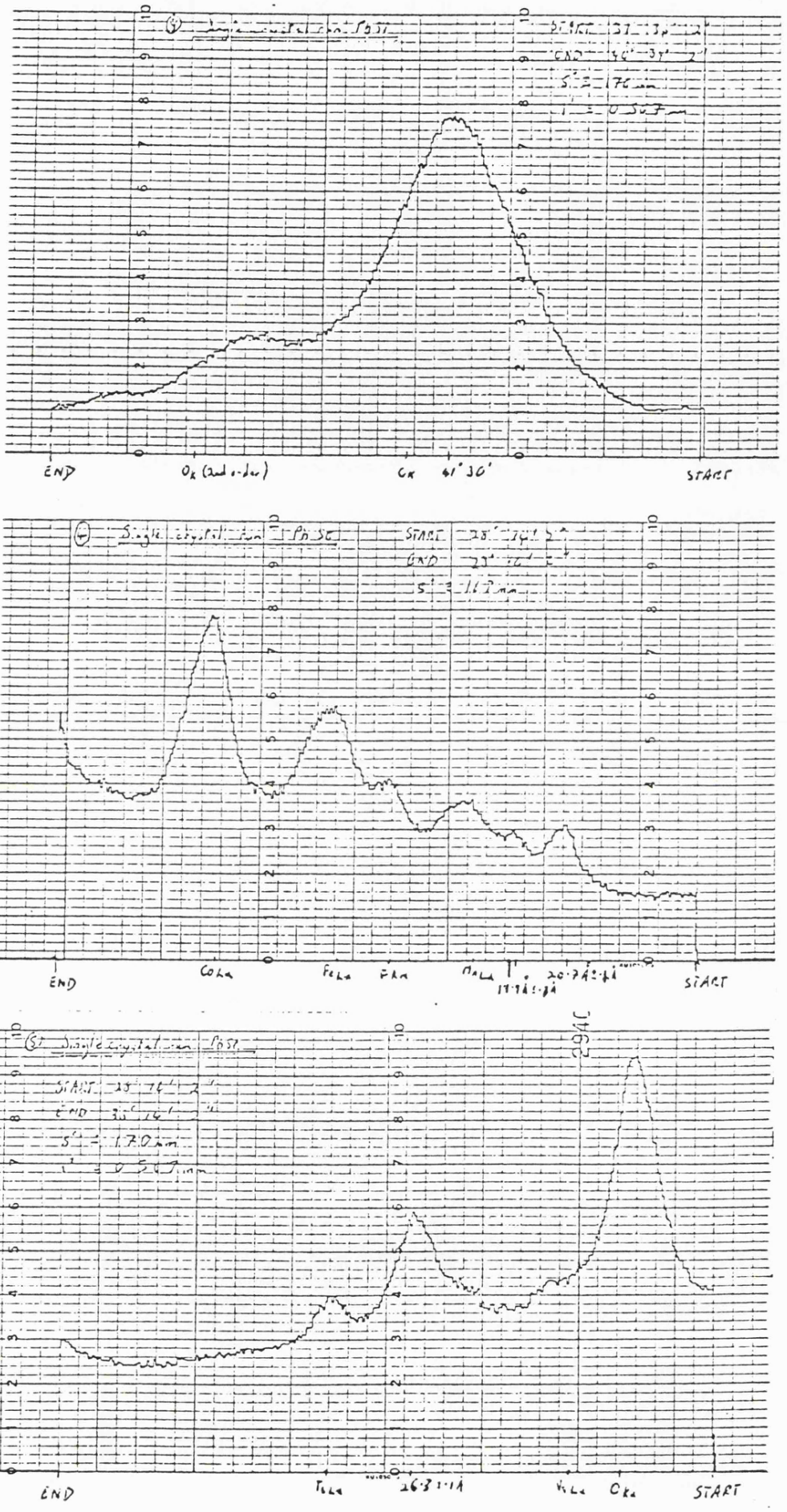


Figure 68. Single crystal scans using a lead stearate monochromator and collimation with 17' FWHM.

LINE	LAMBDA A	RCC RADS*E+4	POLARIZATION CORRECTION	RC RADS*E+4	WCC ARC MINS	PCC %
O K	23.62	1.07 ±.01	1.0136	1.06 ±.01	29 ±2	9.7 ±.1
CO L	15.94	3.02 ±.01	1.0028	3.01 ±.01	21 ±1	3.7 ±.3
C K	44.7	1.67 ±.05	1.219	1.37 ±.05	38 ±2	1.5 ±.1
B K	67.6	6.80 ±.1	1.97	3.50 ±.1	75 ±4	2.8 ±.2
SI K	7.13	2.83 ±.04	1.0001	2.83 ±.04	11 ±4	7.0 ±.1
C K	44.7	1.68 ±.05	1.219	1.38 ±.05	34 ±4	1.5 ±.4

FIGURE 69. SUMMARY OF THE LEAD STEARATE RESULTS.

ERRORS QUOTED DERIVED FROM COUNTING STATISTICS.

Henke has reported evidence for O_k edge structure but this investigation did not persue that any further. Such structure has been shown by the single crystal scans to be unimportant as far as the commercial use of lead stearate multilayers is concerned and the analysers do seem to perform usefully. Further investigation is required to get a more detailed picture of the edge structure, especially the apparantly low reflectivity at the O_k and C_k emission lines just below their respective absorption edges.

8.2 Gypsum 020.

Gypsum is a naturally occurring form of calcium sulphate. It has a monoclinic structure which was determined during the early stages of the development of crystalography. The structure is given by Wyckoff (1968), reference 48, and is summarised in figure 70. The 2d of the 020 planes is 15.15 Å and the planes are therefore potentially useful well into the soft X-ray region. Fortunately the crystal cleaves well parallel to the 010 planes and good quality Bragg analysers are very easy to produce. Gypsum has been used extensively for spectral analysis but unfortunately it deteriorates in a vaccuum due to efflorescence of water of crystalisation, limiting its use to short duration sounding rocket experiments. Nevertheless, fine spectra of the coronal emission of the sun have been obtained, for example Pye, Evans and Hutcheon (1977), reference 49.

Calibration of the efficiency of gypsum 020 has been carried out before, notably Leigh, reference 42, but

GYPSUM, CASO4.2(H2O), MONOCLINIC

DENSITY 2.32
2D OF 020 PLANES 15.15 A
VOLUME OF UNIT CELL 494.59 CUBIC A

ATOMIC COORDINATES

CALCIUM	X	Y	Z
	0.000000	.420000	.250000
	0.000000	.580000	-.250000
	-.500000	.920000	.750000
	-.500000	.080000	-.750000

SULPHUR	X	Y	Z
	0.000000	-.078200	.250000
	0.000000	.078200	-.250000
	.500000	.421800	.750000
	-.500000	-.421800	-.750000

OXYGEN	X	Y	Z
	.966500	.135200	.550600
	-.966500	-.135200	-.550600
	-.466500	.364800	-.050600
	.466500	.635200	.050600
	.966500	-.135200	.050600
	-.966500	.135200	-.050600
	-.466500	.635200	.449400
	.466500	.364800	.550600
	.760200	.022200	.673500
	-.760200	-.022200	-.673500
	.260200	.522200	.173500
	-.260200	.477800	-.173500
	.760200	-.022200	.173500
	-.760200	.022200	-.173500
	.260200	.477800	.673500
	-.260200	.522200	.326500
	.379700	.183100	.450000
	-.379700	-.183100	-.450000
	.879700	.683100	.950000
	.120300	.316900	.050000
	.379700	-.183100	.950000
	-.379700	.183100	-.950000
	.879700	.316900	.450000
	.120300	.683100	-.450000

HYDROGEN	X	Y	Z
	.236000	.165500	.489000
	-.236000	-.165500	-.489000
	.236000	-.165500	.989000
	-.236000	.165500	-.989000
	.736000	.665500	.989000
	.254000	.334500	.011000
	.736000	.334500	.489000
	.254000	.665500	-.489000
	-.410000	.246900	.488000
	.410000	-.246900	-.488000
	-.410000	-.246900	.988000
	.410000	.246900	-.988000
	.910000	.746900	.988000
	.090000	.253100	.012000
	.910000	.254000	.488000
	.090000	.746900	-.488000

FIGURE 70. UNIT CELL OF GYPSUM (REFERENCE 48).

provided with the power of the theoretical calculations now available, coupled with the Department's two crystal machine, a full theoretical calibration with a new set of experimental results was worth obtaining. Unlike the lead stearate multilayers, the application of the computer programs was straightforward and the results are summarised in figures 71-75. The absorption features of S_k 5.02 Å and Ca_k 3.07 Å are clearly visible in all the functions and there is the distinctive 45° dip which for gypsum occurs at ~10 Å. Figure 74 of the resolutions shows that it is reasonably constant over the entire range at about 3×10^{-4}.

A large number of gypsum samples were available for calibration but the best were two Quartz et Silice samples marked 'A' and 'B'; size 22 mm x 45 mm. The surfaces were clear and only slightly scratched at the edges where they had been previously mounted in the machine. The pair were used for a preliminary R_c run which showed that they were indeed in good condition. Although the crystals were cleaved, several tilt curves were produced to check the alignment of the crystals. This showed that the nominal parallel settings were indeed correct to within the sensitivity of the measurements. An investigation of the effect of slit widths was made to determine if there was any spectral interference in the rocking curves at high energies due to inadequate divergence. Unfortunately the two crystal rocking curves at high energies displayed a marked asymmetry but they remained fairly constant after the beam was narrowed to less than about 6'. The results are summarised by figure 76, including two R_{ab} measurements

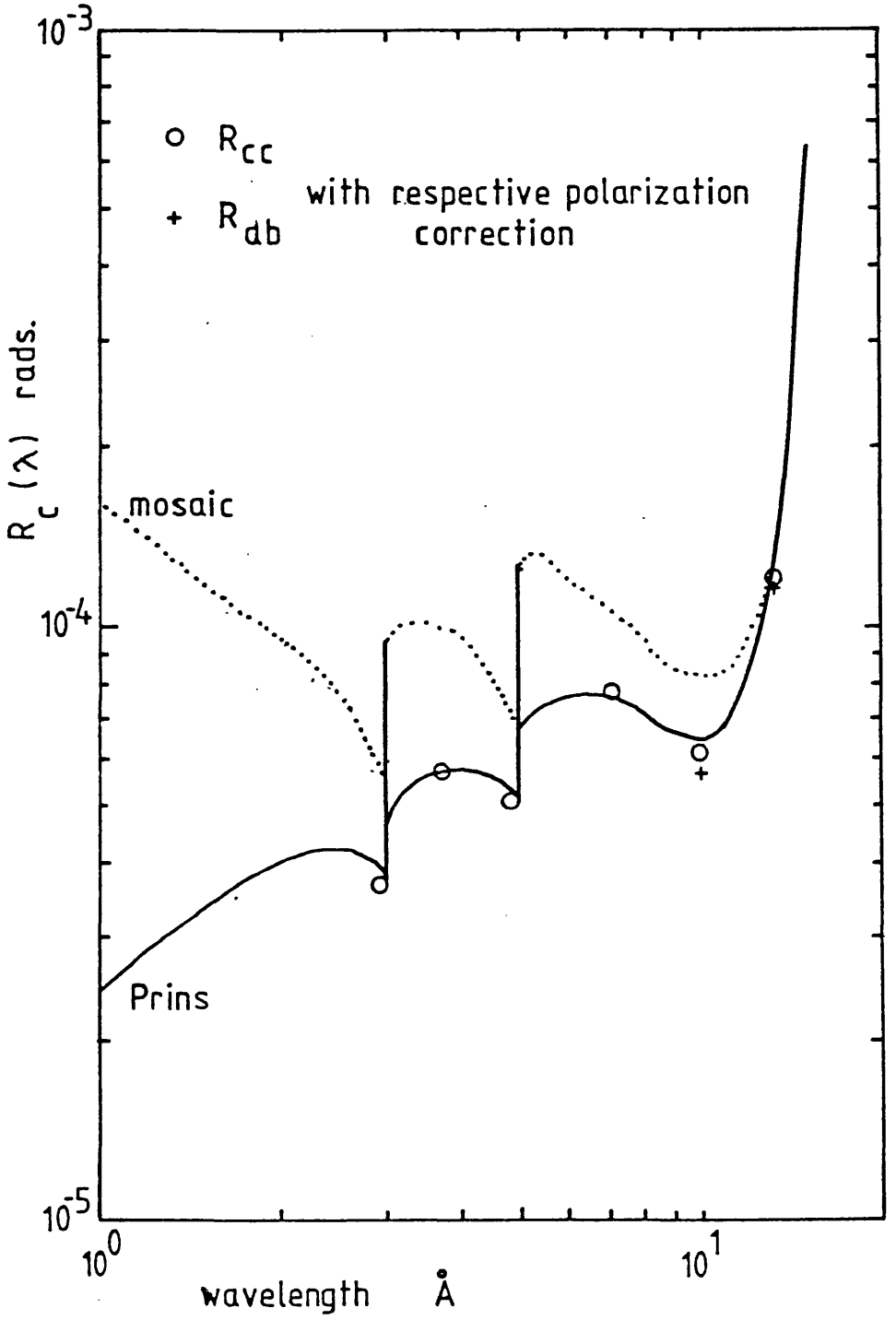


Figure 71. Gypsum 020 single crystal integrated reflectivity.

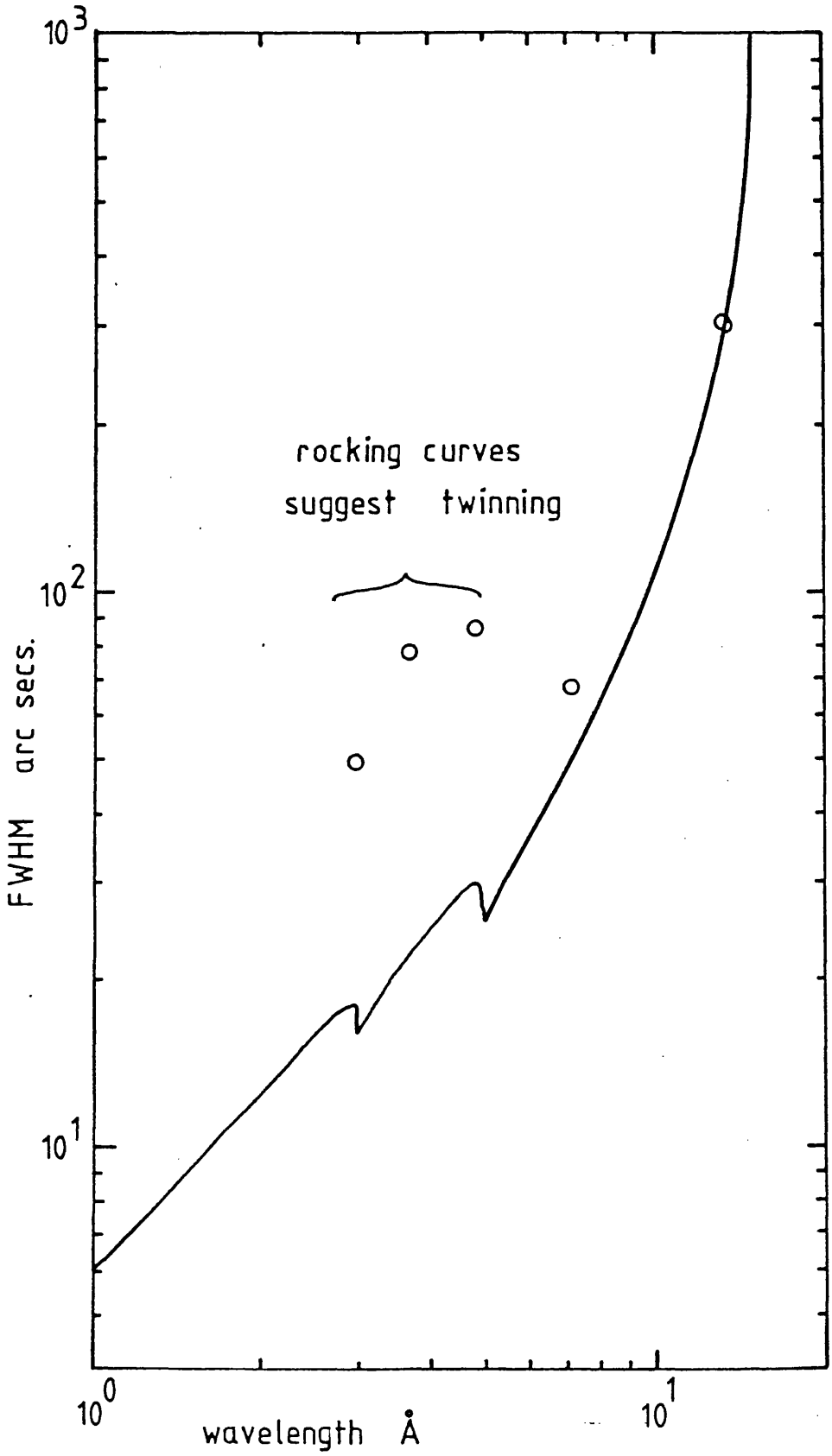


Figure 72. Gypsum 020 two crystal rocking curve FWHM.

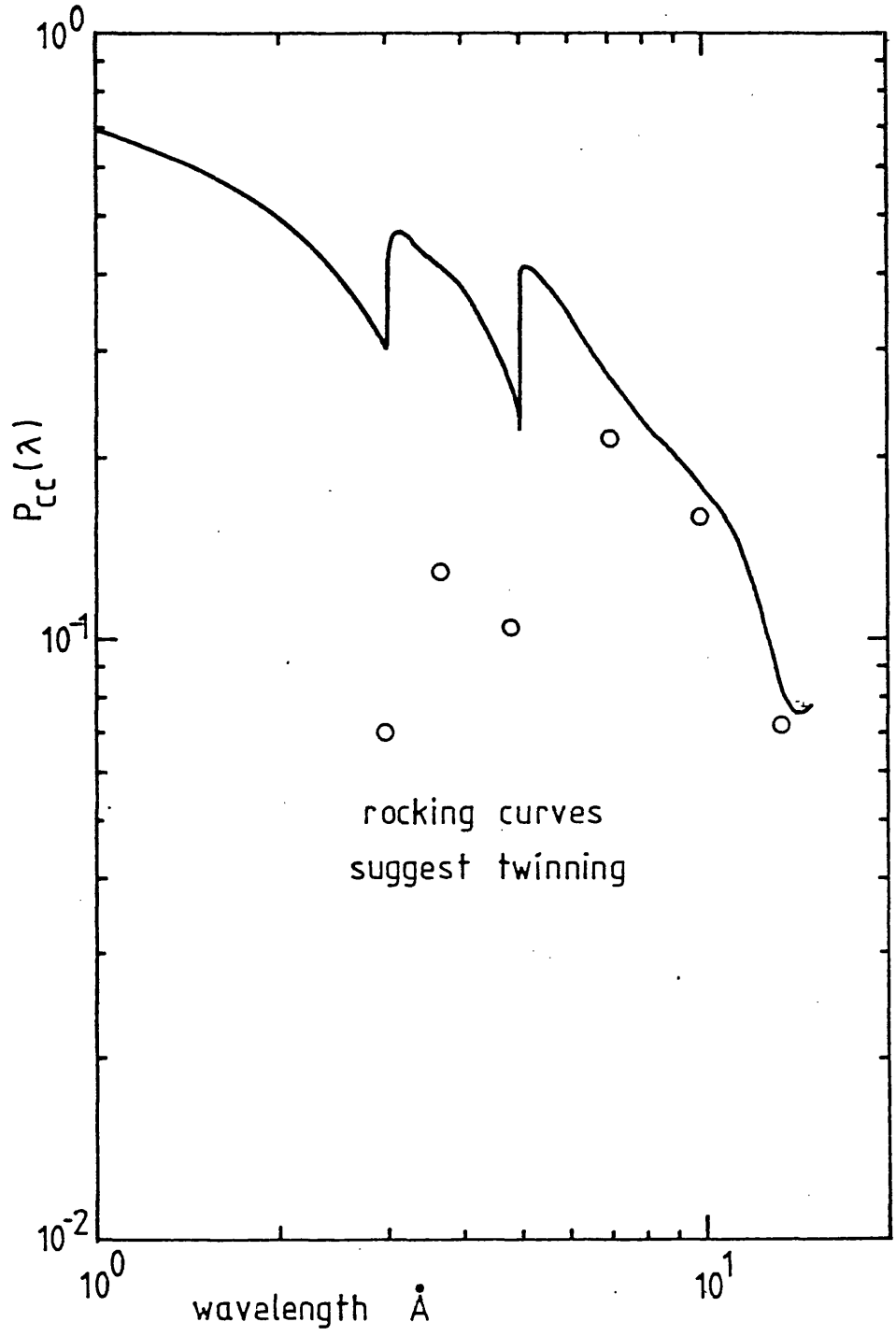


Figure 73. Gypsum 020 two crystal peak reflectivity.

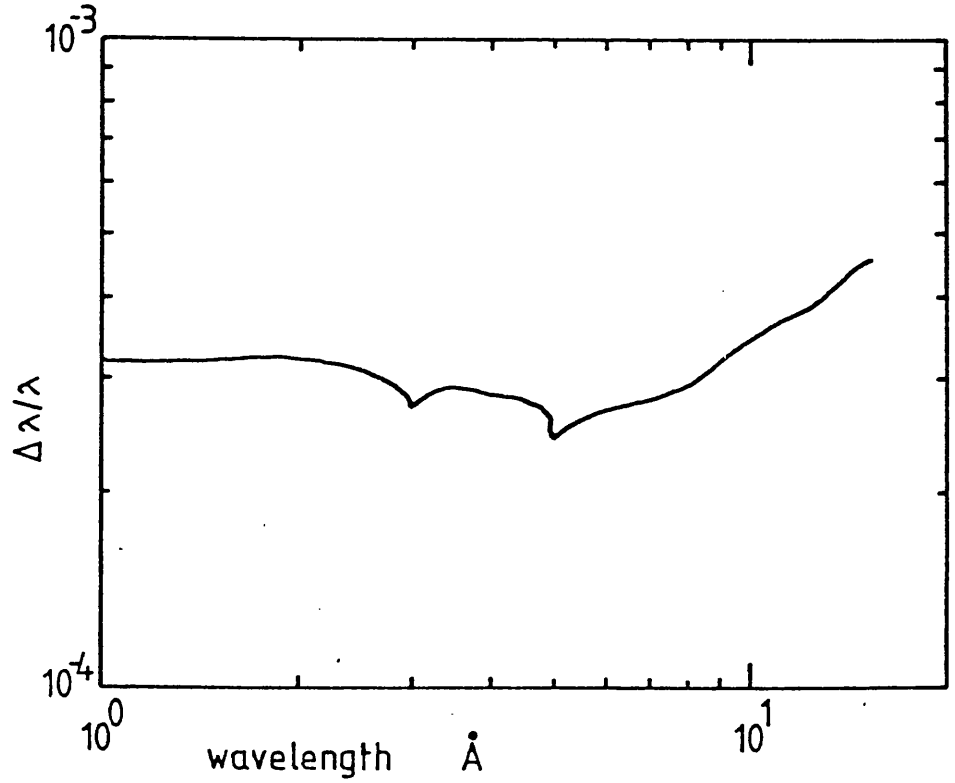


Figure 74. Gypsum 020 single crystal resolution derived from the theoretical FWHM.

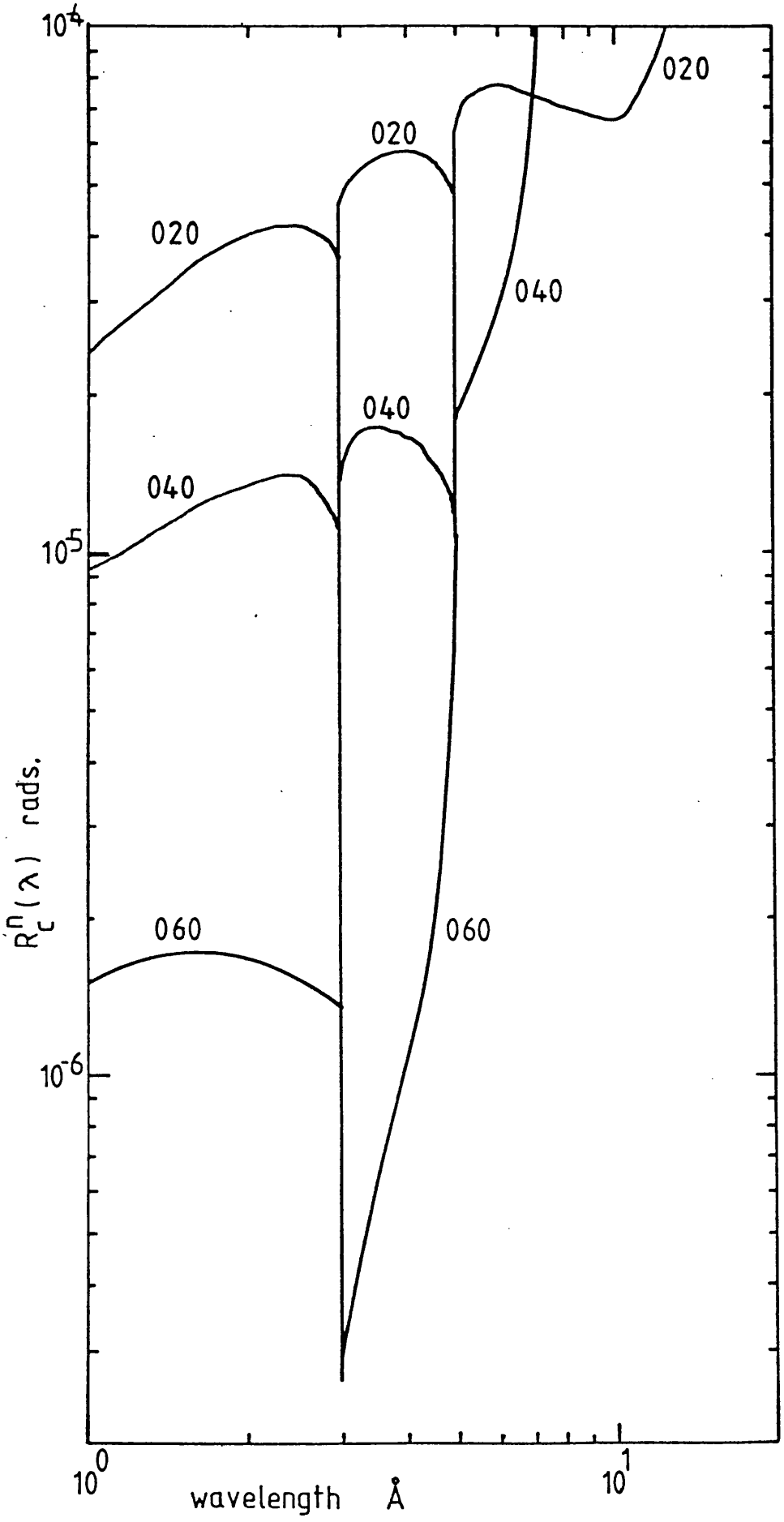


Figure 75. Gypsum 020, 040 and 060 single crystal integrated reflectivities.

LINE	LAMBDA A	RCC RADS*E+5	POLARIZATION CORRECTION	RC RADS*E+5	WCC ARC SECS	PCC %	RAB RADS*E+5	POLARIZATION CORRECTION	RC RADS*E+5
SI K	7.13	9.39 ±.2	1.1938	7.9 ±.2	68 ±2	20.6 ±.3			
SC K	3.03	3.70 ±.2	1.0043	3.70 ±.2	50 ±1	7.0 ±.2			
K K	3.74	5.70 ±.1	1.0084	5.70 ±.1	79 ±2	13.0 ±.5			
RU L	4.85	5.30 ±.2	1.0330	5.10 ±.2	86 ±4	10.5 ±.5			
MG K	9.89	11.6 ±.03	1.8980	6.10 ±.03	112 ±4	15.9 ±.5	6.20 ±.2	1.0090	6.10 ±.2
CU L	13.34	14.7 ±.09	1.2619	1.20 ±.1	306 ±20	7.2 ±.5	11.9 ±.05	1.0180	1.17 ±.05

FIGURE 76. SUMMARY OF THE GYPSUM 020 RESULTS.

ERRORS QUOTED DERIVED FROM COUNTING STATISTICS.

made using one of the lead stearate crystals as a monochromator. The results are included on the theoretical curves. The agreement with theory is seen to be excellent for the integrated reflectivity, both from R_{cc} and R_{ab} measurements, indicating that the lattices were reasonably perfect rather than mosaic. The width and peak values taken at the lower three emission lines show reasonable agreement but for the upper three values the deviation is enormous. However the rocking curves were very distorted, with a definite hump on the side, suggesting that one of the crystals was twinned; being formed by two perfect crystals slightly misaligned with each other. This effect had previously been reported by B. Leigh, who probably tested the same pair of crystals. This finding warns that if the resolution response of a crystal is important it must be tested rigorously using X-rays rather than mere visual appraisal. Undoubtedly if a sample without a twinning fault was tested, the width and peak values would follow the theoretical prediction and the theoretical values are good enough to be used as a calibration provided spot checks are made on the samples to be used.

In order to study the absorption edges, the source target was etched with a concentrated acid cocktail to remove as much contamination as possible. Single crystal scans of the source spectrum were then made using a beam divergence of 4' 55" (slits 1 and 3 1 mm) FWHM. The resulting traces are shown in figure 77. The calcium edge at 3.07 Å is clearly visible as a clean step, the depth of which is the same as indicated by the theoretical curve in figure 71. The sulphur k edge 5.02 Å also shows up clearly

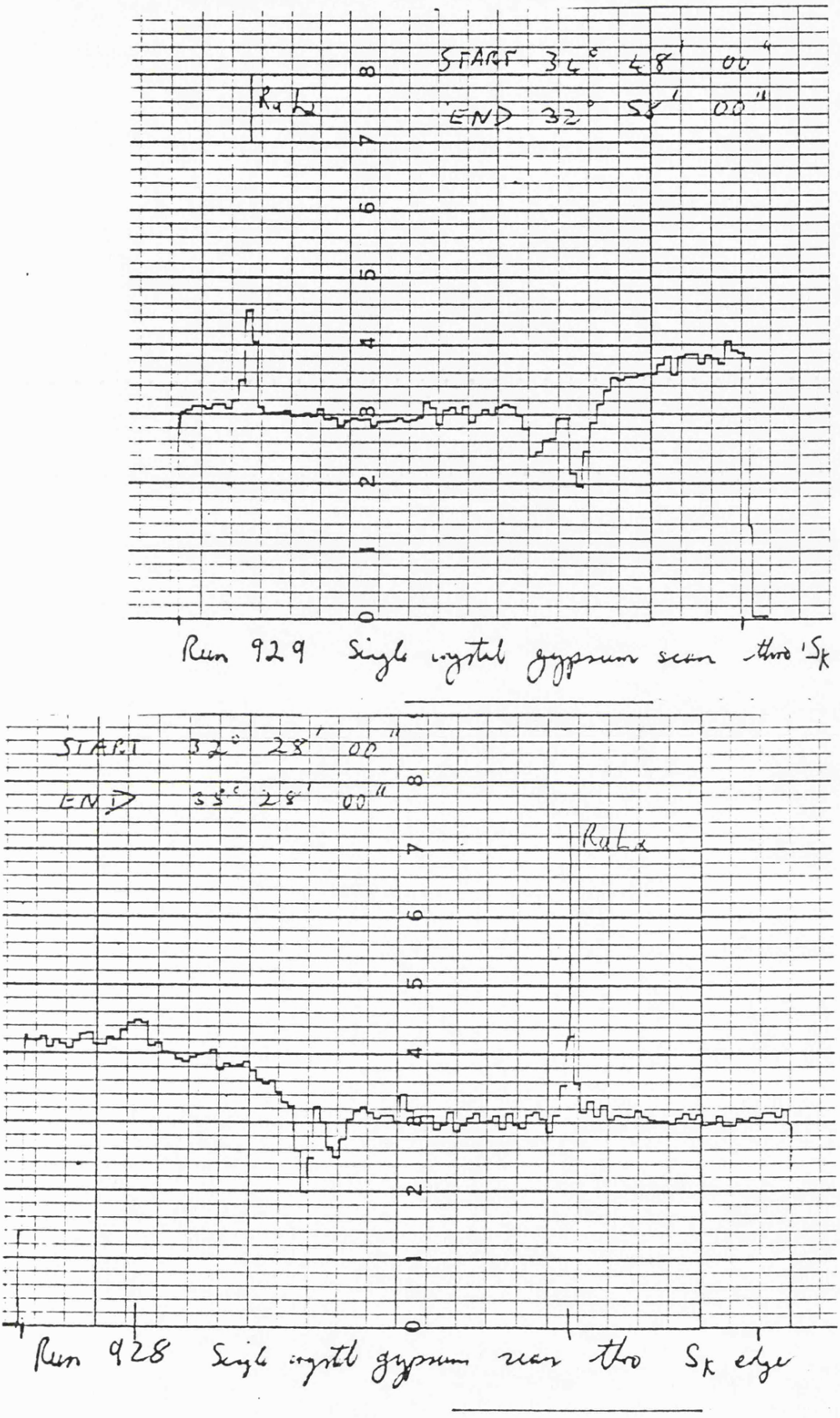


Figure 77. Gypsum 020 single crystal scans across the S_k absorption edge revealing enhanced reflectivity just above the edge energy.

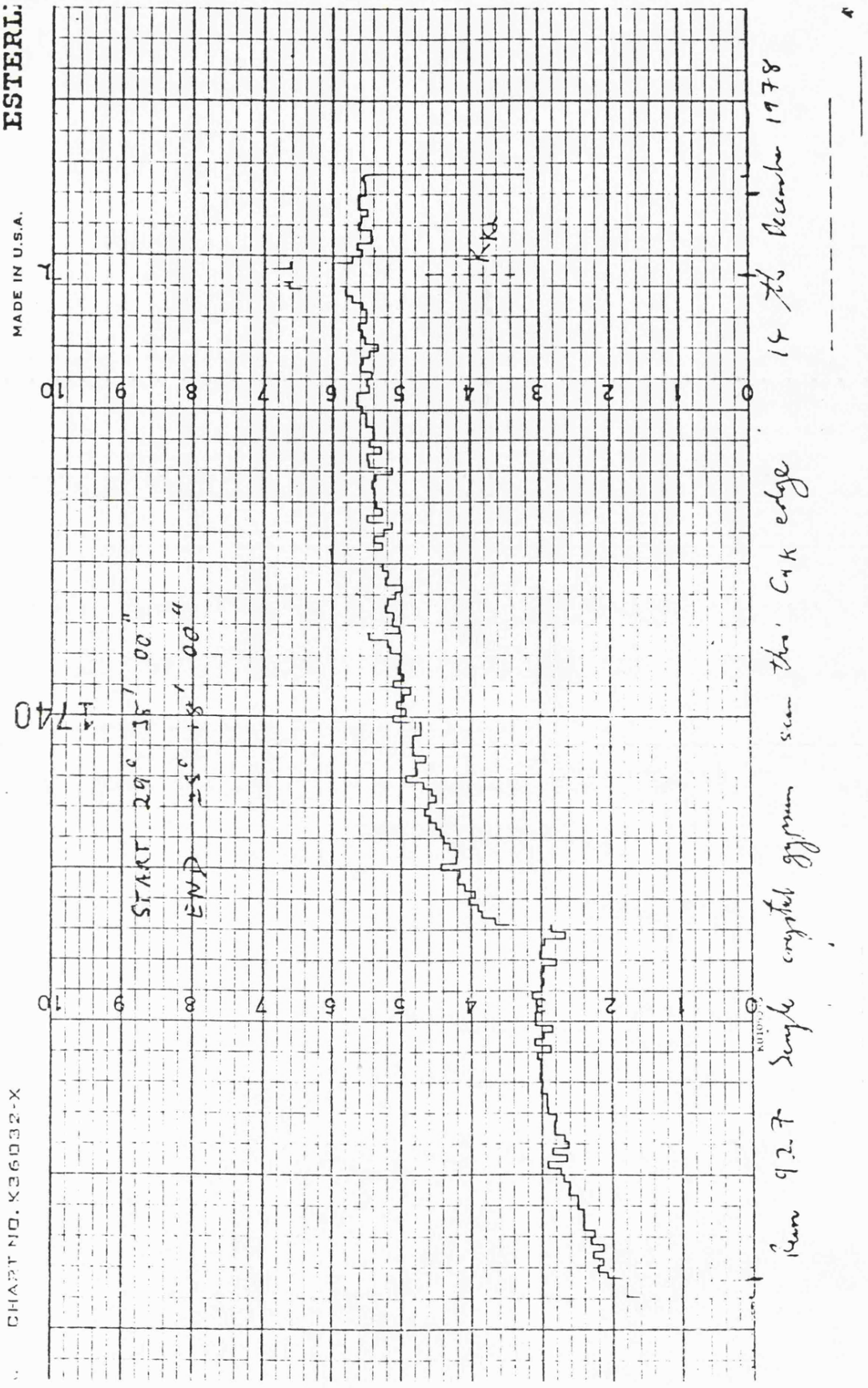


Figure 77 cont.. Gypsum 020 single crystal scan through the Ca_k edge producing a clean step.

but it displays a certain amount of structure with a peak occurring just above the energy of the absorption edge. This result was repeated to confirm this enhancement of the reflectivity and the result was identical. In fact such a peak is to be expected from the resonance of the electrons very close to the edge energy. Unfortunately the data provided by Cromer and Lieberman is too coarsly sampled to reveal such fine structure in the oscillator density and the nature of the integral for $\Delta f'$ means sharp peaks within the photoelectric absorption coefficient have very little effect anywhere other than in their immediate vicinity.

The results of both theory and experiment for gypsum are in satisfactory agreement and clearly demonstrate the power of the theoretical model. The agreement between R_{cc} and R_{ab} measurements provides confidence in the polarization correction, so important to successful application of two crystal reflection measurements in calibrating single crystal response. Single crystal scans across a continuum spectrum give direct evidence of an S_k electron resonance a few eV above the absorption edge at 5.02 Å.

8.3 Beryl $10\bar{1}0$.

Beryl is a gemstone and a major ore of beryllium. It has the formula $Be_3\ Al_2\ [Si_6\ O_{18}]$ although many impurities have been reported. Its colour differs depending on the exact composition, varying between pale blue, yellow and green. The lattice structure was first determined by Bragg and West (1926), reference 50, and subsequent measurements have refined the atomic positions to give those stated by

Wyckoff (1968), reference 48. The hexagonal strucure of the unit cell is illustrated by figure 78. The $10\bar{1}0$ planes are indicated and they have a 2d of 15.9549 Å.

M. Hayes and A. Samson ran the computer calculations for beryl in an M.Sc project (1978) and the results are fully presented in their report, reference 51. The results obtained using the anomalous dispersion calculation described above are summarised in figures 79-81. The response is remarkably similar to gypsum, the major difference being the absorption features due to Si_k 6.74 Å and Al_k 7.95 Å electrons. Hayes and Samson compared these calculations with the experimental data available from other workers including Korringa et al (1957), reference 52, who concentrated on two crystal rocking curve peak and width measurements. They were in reasonable agreement with the present theoretical calculations, demonstrating the excellence of their experimental technique. Corresponding integrated reflectivity measurements from Sawyer et al (1962), reference 53, are in very poor agreement and are consistently low by a factor of about 3, with a good deal of scatter. These experimental results are not included in the diagram presented here to avoid confusion.

The incentive to produce a complete set of experimental values using the University of Leicester's two crystal spectrometer was therefore high, especially since the theoretical calculations were thought to be of high quality. A pale green, hexagonal beryl crystal about 3 cm x 1.5 cm was available in the spectrometry laboratory. It appeared to be resonably clear inside and free from faults, except for one inclusion about 4 mm long running

BERYL, BE3.AL2.(SIO3)6, HEXAGONAL

DENSITY 2.641
2D OF 1010 PLANES 15.955 A
VOLUME OF UNIT CELL 6767 CUBIC A

ATOMIC COORDINATES

OXYGEN

X	Y	Z
.294	.242	0.0
.052	.294	0.0
.242	-.052	0.0
.242	.294	0.5
.294	.294	0.5
-.052	.242	0.5
.499	.143	.138
.143	-.356	.138
.356	.499	.138
.499	.143	-.138
.143	-.356	-.138
.356	.499	-.138
.143	.499	.638
.499	.356	.638
-.356	.143	.638
.143	.499	.638
-.499	-.356	.638
.356	-.143	.638
-.294	-.242	0.0
-.052	-.294	0.0
-.242	-.294	0.0
-.242	.052	-0.5
-.294	-.294	-0.5
.052	-.242	-0.5
-.499	-.143	-.138
-.143	.356	-.138
-.356	-.499	-.138
-.499	-.143	.138
-.143	.356	.138
-.356	-.499	.138
-.143	-.499	-.638
-.499	-.356	-.638
.365	-.143	-.638
.143	.499	-.638
.499	.356	-.638
-.356	.143	-.638

BERYLLIUM

X	Y	Z
0.5	0.5	0.75
0.5	0.0	0.75
0.0	0.5	0.75
-0.5	-0.5	-0.75
-0.5	0.0	-0.75
0.0	-0.5	-0.75

ALUMINIUM

X	Y	Z
.333	.667	.25
.667	.333	.25
-.333	-.667	-.25
-.667	-.333	-.25

SILICON

X	Y	Z
.382	.118	0.0
.264	.382	0.0
.118	-.264	0.0
.118	.382	0.5
.382	.264	0.5
-.264	.118	0.5
-.382	-.118	0.0
-.264	-.382	0.0
-.118	.264	0.0
-.118	-.382	-0.5
-.382	-.264	-0.5
.264	-.118	-0.5

FIGURE 78. UNIT CELL OF BERYL (REFERENCE 48).

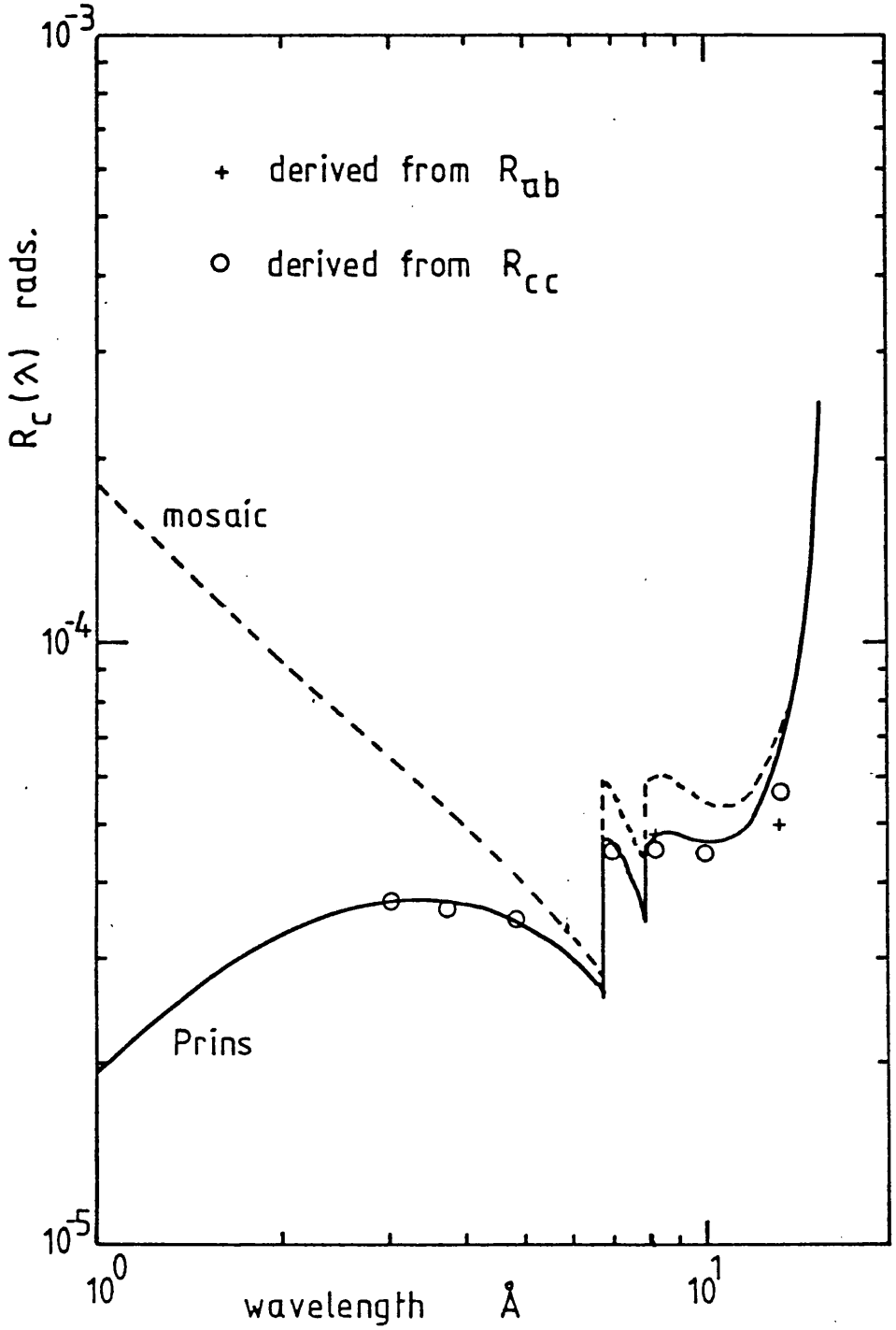


Figure 79. Beryl $10\bar{1}0$ single crystal integrated reflectivity.

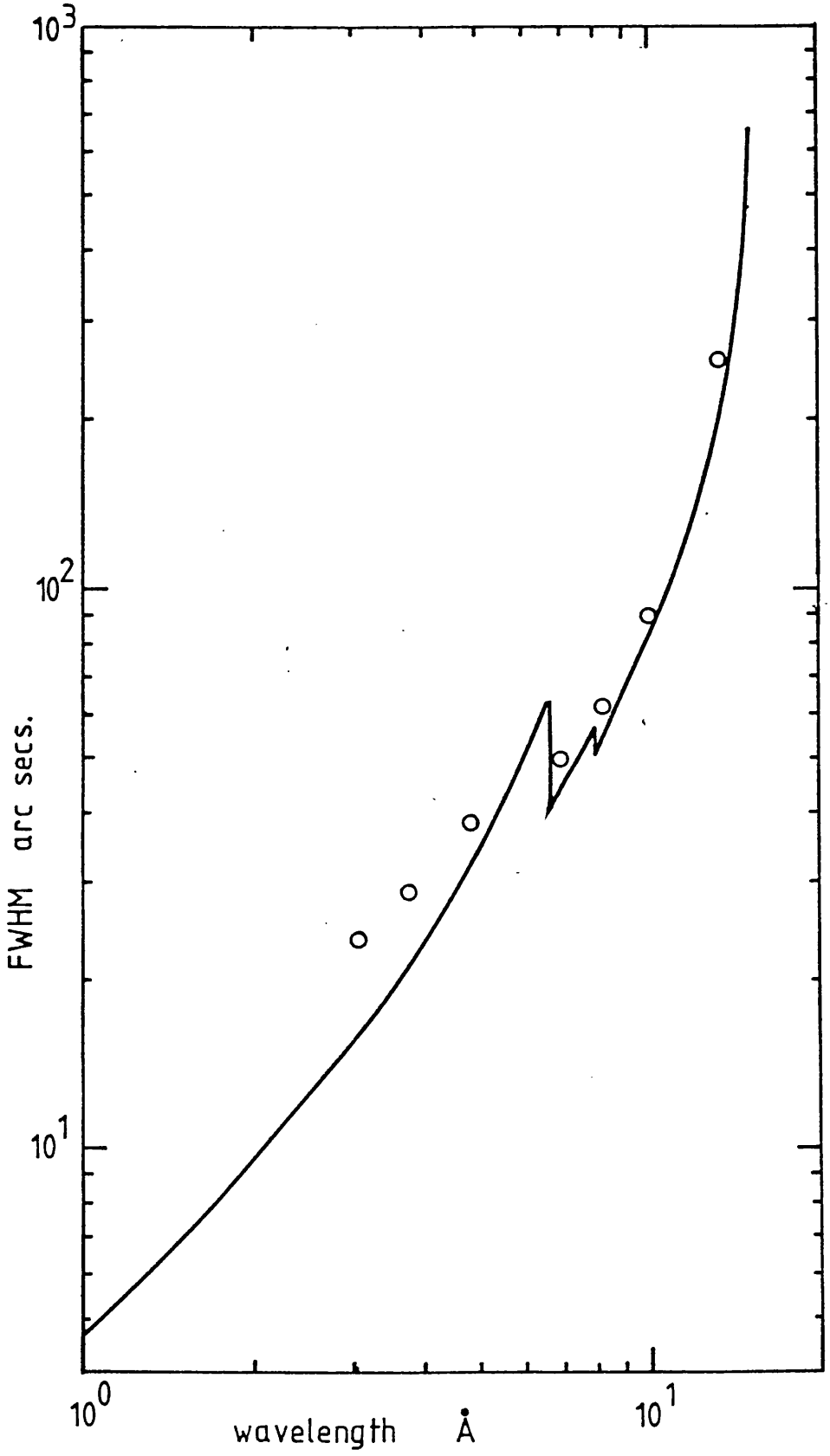


Figure 80. Beryl $10\bar{1}0$ two crystal FWHM.

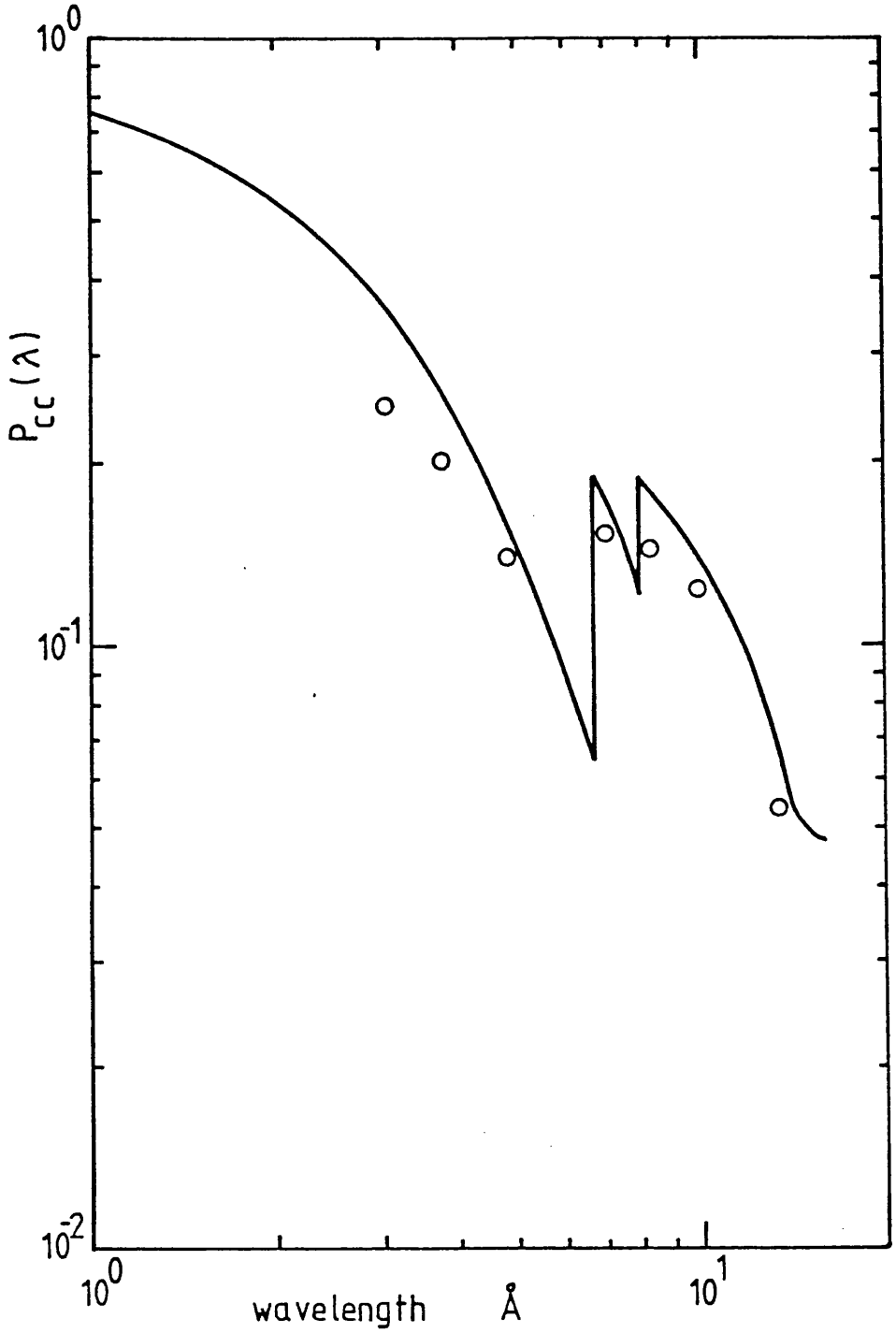


Figure 81. Beryl $10\bar{1}0$ two crystal peak reflectivity.

parallel to the sides. Slices of about 2 mm thickness were cut parallel to one of the hexagonal faces using a diamond saw in the Geology Department. The accuracy of cutting was estimated to be about $\pm$ 30' to the crystal face. Unfortunately gnonimetric grinding equipment was not immediately available to produce a crystal surface parallel to the reflection planes. The crystals were therefore not suited to spectroscopic use, however they were adequate for the purpose of calibration. The best two slices were selected and one surface of each ground and polished using diamond paste down to $\frac{1}{4}$ μm diameter. Finally they were polished to give an optically clean surface with no sign of scratches when inspected using a hand lens. Both crystals were then etched in 40% hydrofluoric acid for ~24 hours to remove any traces of polishing material and crystal debris, hopefully laying bare a surface representative of the internal lattice structure. The final quality of the crystals looked good with a usable surface of about 1 cm x 2 cm on each, although one was slightly narrower because it had been cut from nearer the original crystal face.

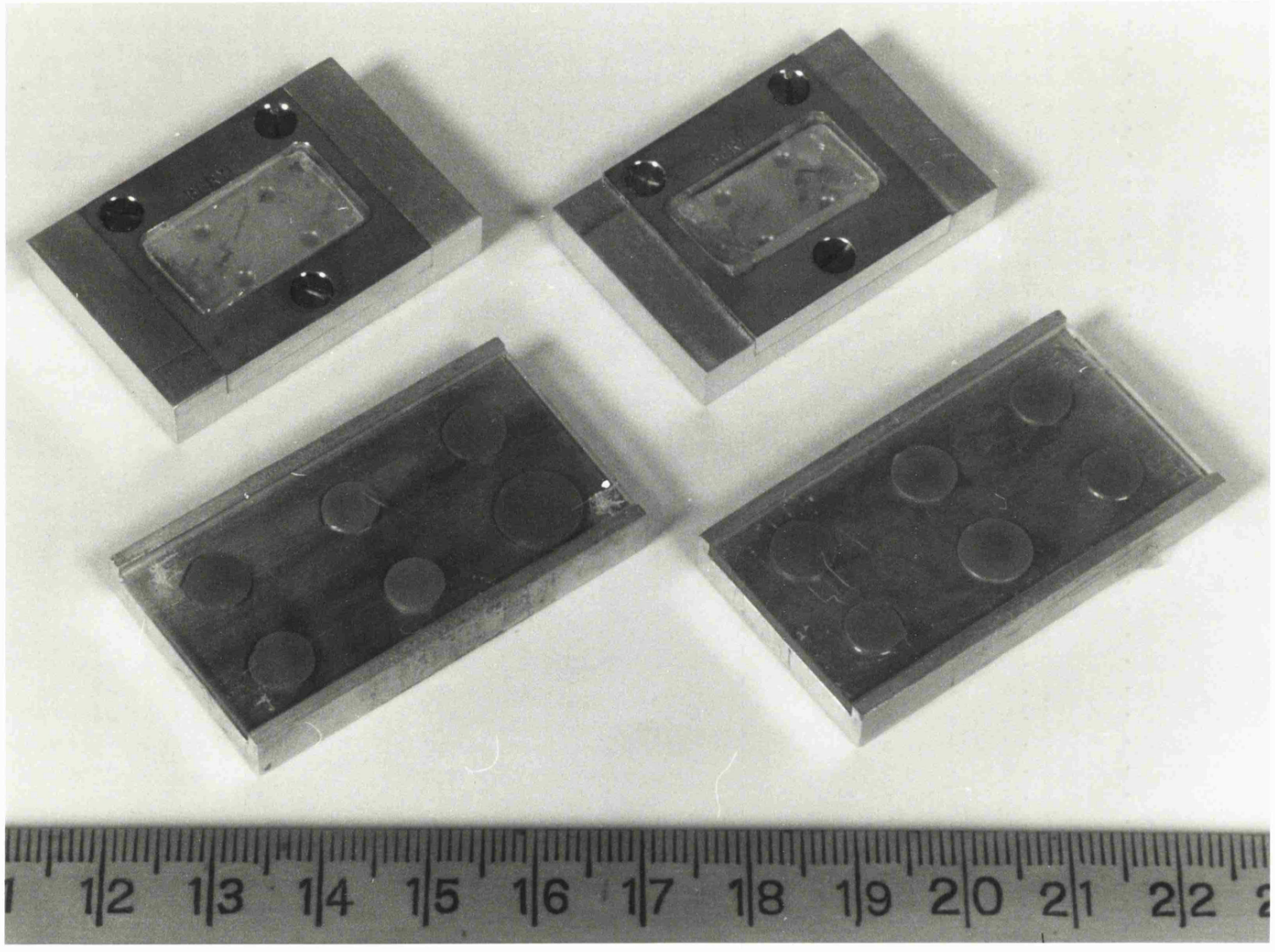
The crystals were far smaller than the standard supplied by Quartz et Silice and in order to utilise the normal crystal mounts, they were set in small aluminium frames. The front face of the frames were ground flat using diamond paste and then polished. The frames were then placed face down on a flat glass block and the crystals placed inside them, with their polished side against the same glass surface. The polished frame front and crystal surface were then coincident. The crystals were then

cemented to the frames using araldite epoxy. The resulting set crystals are shown in figure 82 alongside the gypsum samples 'A' and 'B'. When mounted in the machine, this mode of setting had the advantage that all the strain was taken by the aluminium frame leaving the crystals stress-free.

Because the crystals were small and inaccurately cut, they had to be positioned and aligned with extreme care. A single crystal reflection off the first crystal was set up as $Ru_{L\alpha}$, 4.85 Å. The crystal was then tilted until the position of maximum flux was located. This was in fact ~1° away from the nominal zero tilt setting, indicating a rather larger error in cutting than was hoped for. In order to achieve this tilt setting, the crystal holder had to be shimmed out using broken pieces of glass slide! At this maximum the relecting planes must have been reasonably perpendicular to the dispersion plane defined by the slit system. A series of two crystal rocking curves were then executed using varying tilts of the second crystal. Eventually a position that yielded a symmetric rocking curve was located and the exact maximum of the peak reflectivity as a function of tilt found by drawing a smooth curve through the various reflectivity values. The maximum was broad and the setting of the tilt micrometer was therefore not critical.

Before commencing a series of 1-1 runs the slit system and crystal positions were checked optically to ensure no part of the X-ray beam could miss the second crystal. This was tricky because the optical face of the first crystal was not perpendicular to the dispersion

Figure 82. The small beryl crystals mounted in aluminium frames compared with the much larger gypsum samples supplied by Quartz et Silice.


12 13 14 15 16 17 18 19 20 21 22

plane, in fact the tilt had to be offset for the alignment and reset afterwards. It was found that at all Bragg angles, slits 3 mm high and 1.5 mm wide restricted the beam adequately. These slit settings represent a dispersion plane divergence of 443" which was just adequate for the complete range of the crystals (see figure 80). A set of two crystal rocking curves was obtained at 7 emission lines and the results are summarised in the table in figure 83 and are plotted with the theoretical curves in figures 79-81. Two R_{ab} measurements were also made using a lead stearate monochromator and converted to R_c values using the Prins polarization correction supplied by the computer programs.

The R_c results are gratifyingly close to the theoretical curve and are a vast improvement on reference 53. The R_{ab} measurements agree fairly well with the R_{cc} results. The major error in the R_{ab} values is in fact likely to be loss of signal in the wings rather than statistical, because the lead stearate response profile is too wide to be comfortably accommodated on a single 5^o scan, which is the maximum the machine can handle. Both results at $Cu_{L\alpha}$ 13.34 Å are well down on the theoretical prediction. Purely statistical errors on signal and background put the theoretical curve at 5-6 standard deviations above the experimental points. This rather annoying discrepancy could be due to a fault in the crystal model, especially the anomolous dispersion calculation, however this is unlikely and a systematic measurement error was probably the cause.

The W_{cc} and P_{cc} points agree well with the theoretical

model and are comparable to the results presented by Korringa et al, reference 52. The small wavelength, small grazing angle results are about 8" larger than theoretical and the longer wavelength results lie less than 10% above the Prins limit. These discrepancies are probably due to residual alignment errors, divergence of the beam and imperfection in the surface and within the lattice of the crystals. The sensitivity of the rocking curves to the dispersion plane slits was tested at Al_k 8.34 Å by doubling the width to 3 mm. The beam was optically dangerously close to the edge of the second crystal, making the R_{cc} value uncertain but the rocking curve width and peak remained the same to within the estimated errors. This was considered to be convincing evidence that the beam divergence was having very little influence on the results. Since all measurements were made using the same divergence, the variation in rocking curve width was definitely due to the crystal and any geometrical window broadening should have been reasonably constant over the full range.

To complete the programme, single crystal scans across the two absorption edges were performed in exactly the same way as for gypsum using the same beam divergence of 4' 55". The resulting traces are shown in figure 84. Unfortunately a very strong tungsten line $W_{M\beta}$ 6.757 Å occurs directly on top of the Si_k 6.738 Å and completely obscures the actual edge, although the change in efficiency from below to above the edge by a factor of about 2 agrees with the predicted step. The Al_k 7.948 Å edge was extremely clear and displays definite structure with a distinct hump at a few eV higher than the edge itself. This undoubtedly has

LINE	LAMBDA A	RCC RADS*E+5	POLARIZATION CORRECTION	RC RADS*E+5	WCC ARC SECS	PCC %	RAB RADS*E+5	POLARIZATION CORRECTION	RC RADS*E+5
RU L	4.85	3.62 ±.07	1.026	3.53 ±.07	39 ±1	13.9 ±.2			
K K	3.74	3.72 ±.06	1.0137	3.67 ±.06	29 ±1	20.1 ±.2			
SC K	3.03	3.82 ±.04	1.0028	3.81 ±.04	24 ±1	24.9 ±.1			
AL K	8.34	6.27 ±.06	1.376	4.55 ±.05	62 ±1	14.3 ±.1	4.88 ±.07	1.008	4.84 ±.07
W M	6.98	5.27 ±.05	1.149	4.57 ±.05	50 ±3	15.2 ±.2			
CU L	13.34	8.69 ±.2	1.512	5.7 ±.2	255 ±13	5.4 ±.3	5.00 ±.4	1.0258	4.90 ±.4
MG K	9.89	8.0 ±.1	1.777	4.5 ±.1	90 ±4	12.4 ±.4			

FIGURE 83. SUMMARY OF THE BERYL $10\bar{1}0$ RESULTS.

ERRORS QUOTED DERIVED FROM COUNTING STATISTICS.

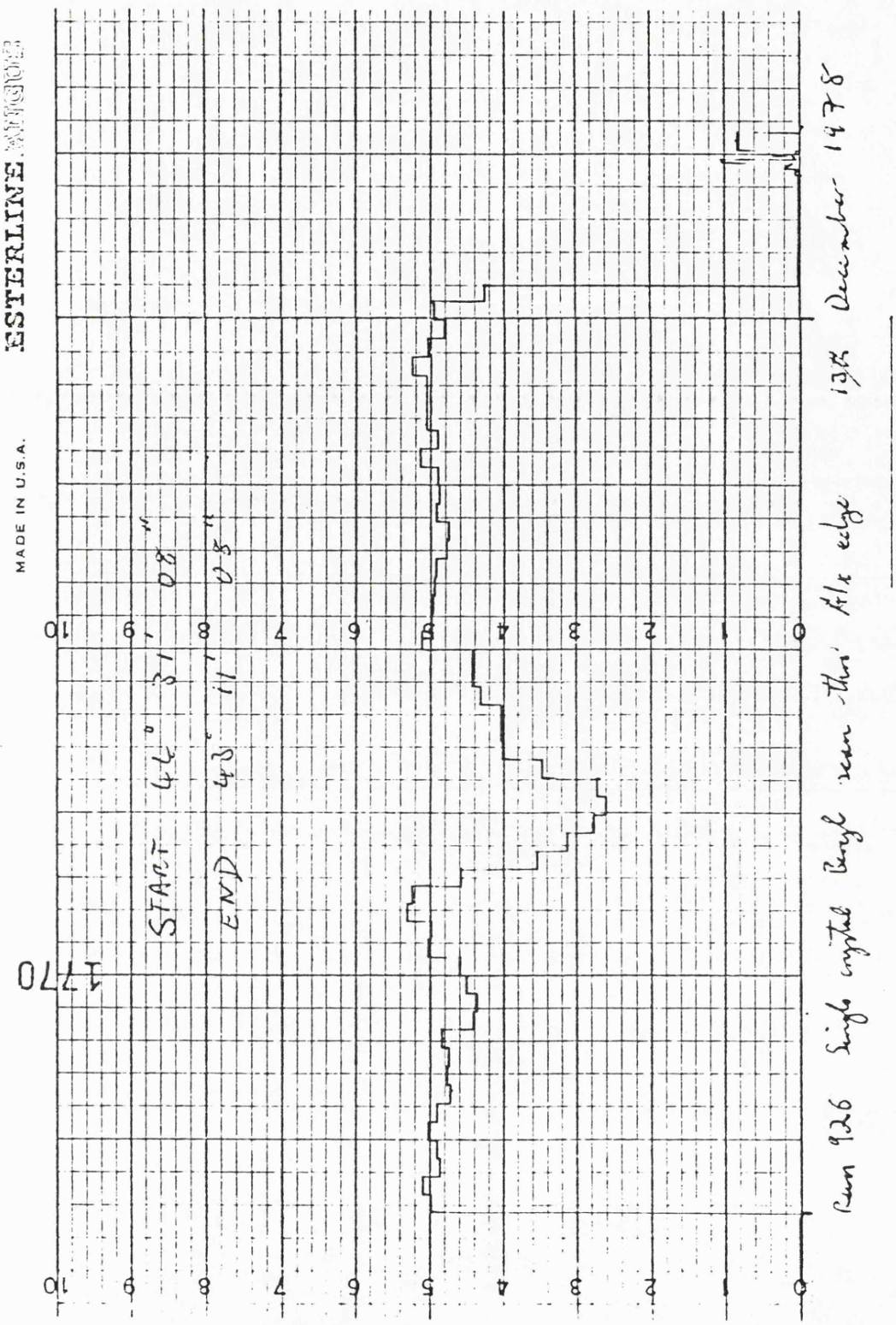


Figure 84. Beryl $10\bar{1}0$ single crystal scan through the Al_K absorption edge revealing enhanced reflectivity just above the edge energy.

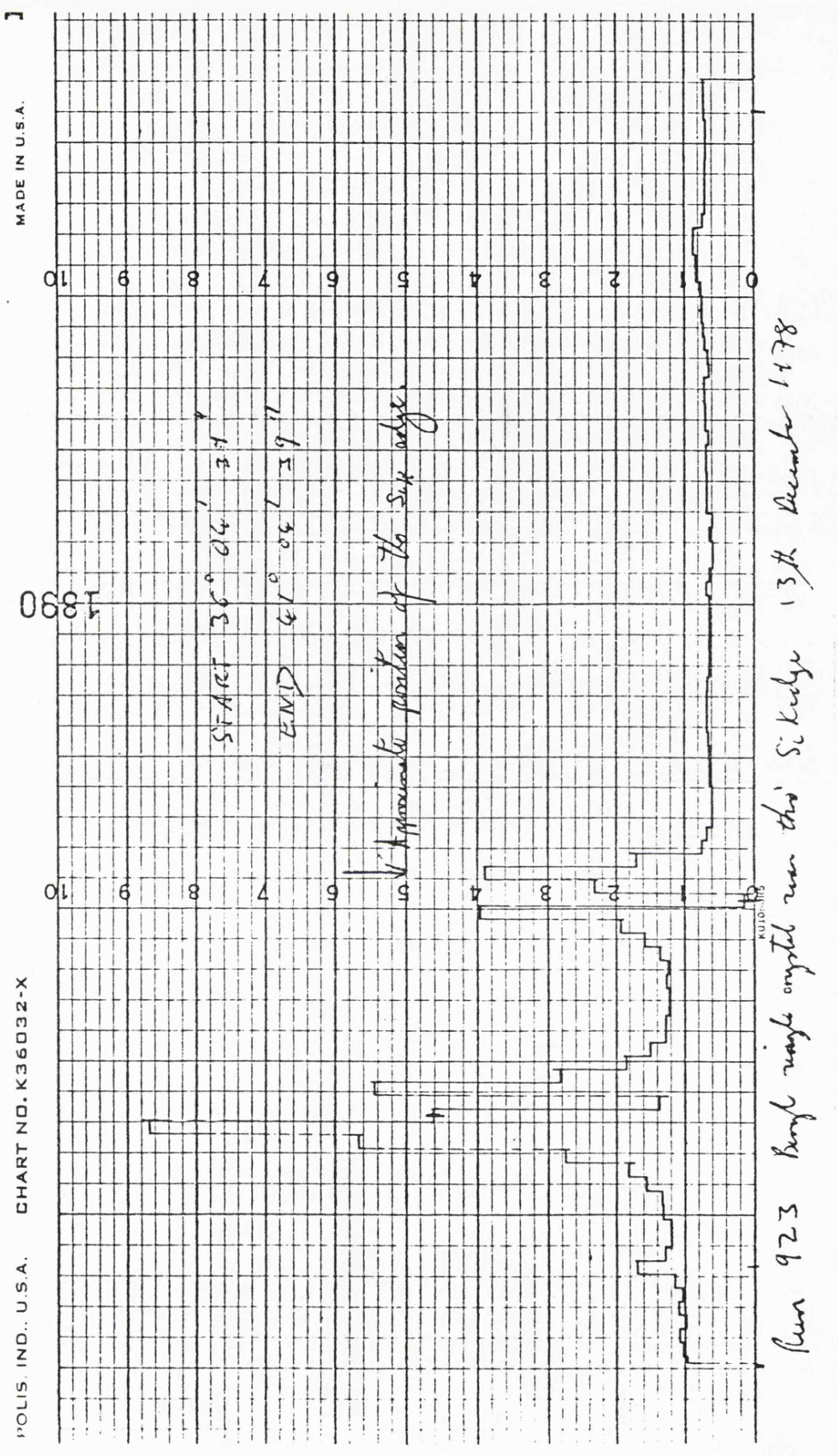


Figure 84 cont.. Beryl $10\bar{1}0$ single crystal scan through the Si_k absorption edge. Unfortunately the edge is obscured by a very strong W line and therefore only the step of the edge can be seen.

the same cause as the enhancement recorded in the S_k edge of gypsum as previously described, demonstrating a resonance of the Al_k electrons.

The study of beryl used as a Bragg analyser presented here is probably the most complete to date. The results agree well with the theoretical model and since impurities were not included in the model, they presumably have very little effect on the crystal's performance. The response of beryl $10\bar{1}0$ is very similar to gypsum 020 but the durability of beryl makes it ideally suited for use in space.

It is proposed that a spectrometer on the Solar Maximum Mission will utilise a beryl crystal and it is hoped that this calibration will be of use in interpretation of the results.

8.4 Conclusion to Part IV.

Part IV is rather different from the other three in that it concentrates on a different aspect of data analysis, instrument calibration. The theory of Bragg spectrometers was introduced in the light of the theory presented for imaging instruments. Attention was drawn to the crystal window functions which are at the heart of spectrometer calibration. A purely theoretical calculation was presented, based on a sophisticated atomic model and crystalographic data. Experimental results were obtained from three crystal types using a two crystal spectrometer in the 1-1 mode and these were compared to the theoretical predictions.

The combination of theoretical and experimental results gave a satisfyingly complete picture of the

behaviour of Langmuir-Blodgett lead stearate multilayers, providing calibration data much needed for this crystal type. The theoretical model is still not in absolute agreement but is better than any other known attempt to date.

The results for gypsum confirmed previous measurements and gave a satisfying demonstration of the power of the theoretical calculation. The run across the S_k absorption edge revealed electron resonance structure.

The beryl calculations provide a more complete picture of this crystal than previously reported, especially for the integrated reflectivity. Again the agreement with theory was excellent, proving the power of the calculation and giving confidence to the measured values. This calibration should prove especially useful in X-ray astronomy and solar experiments in the future. The run across the Al_k edge also revealed electron resonance which has not been observed before.

APPENDIX I

LARGE MATRIX PROCESSING SOFTWARE.

1. Introduction.

The software consists of a set of FORTRAN SUBROUTINES for handling large matrices up to 1024 x 1024. All processing uses square matrices and non-square matrices can only be handled by padding out to square matrices using zeros. A large central memory is not required since rows (or columns) are processed sequentially, using disc files to hold the matrix data rather than loading the entire matrix into the central memory at once. Simple operations are therefore limited by the channel time required for the unformatted READ and WRITE used to access the matrix data, however this has proved to be a rather weak restriction on the software in comparison to calculation time needed for most operations.

All the matrix data is held on disc files using 1 record for each matrix row and 1 word per matrix element (60 bits in CYBER 72). Three forms of binary image file are used. The first includes a header record to provide information about the matrix including its dimensions so that matrix data can be saved permanantly. This form does not have to be a square matrix. The second is a square matrix without a header, the file having N records each containing N words. This form is used for processing. The third is a complex matrix in the discrete Fourier domain in which two records are used for each matrix row; the first for the real part and the second for the imaginary part. Each record contains N words and there are N/2 + 1 pairs of records to completely specify the DFD.

2. Basic matrix file handling.

Matrix files are referenced by using the TAPE or unit number declared in the PROGRAM card. If the file TAPE IT holds a matrix it is referenced in the program by the integer IT. A permanent matrix file must be changed into a processing file before processing can begin. In order to do this a matrix dimension must be set up by a CALL to STARTP(N), where $2 \leqslant N \leqslant 1024$. N must be chosen to suit the permanent matrix file and if fast transformations are required, N must be a power of 2. A CALL to ENTER(IT,IO) then places the matrix IT into IO. If N is too small, elements with high row and column indices will be lost and if N is too large, the high index rows and columns will be padded out with zeros. If the dimensions of IT are unknown then a CALL STARTP(0) followed by CALL ENTER(IT,IO) will cause an automatic choice of N to be made so that no data is lost.

It is often useful to be able to enter a permanent matrix file into the processing matrix field set up by STARTP(N) including a circular shift. This can be accomplished by a CALL ADZEROS(IT,IO,NX,NY) where NX and NY specify the circular shift required. NX = NY = 0, implying no shift, makes ADZEROS similar to ENTER. There is no automatic choice of N available with ADZEROS and N must be large enough to accommodate IT. A CALL to ADZEROS also causes a corruption of the file IT. ADZEROS is rather slow and if a small matrix $\leqslant 16$ in either dimension needs to be entered with a circular shift, a CALL SHENTER(IT,IO,NX,NY) will have the same effect as ADZEROS but will be much faster if N is large.

Permanent matrix files are created in two ways, either from a FORTRAN array DATA(NX,NY) by a CALL to MAKFILE(DATA, NX,NXL,NXH,NY,NYL,NYH,IO) or from a processing file IT by CALL FSAVE(IT,IO,NXL,NXH,NYL,NYH). In both cases the sub-matrix NXL to NXH, NYL to NYH is written to TAPEIO. The header information is accessed via the labelled common block COMMON/HEADIT/I(8),NX,NY,R(10). NX and NY are the permanent matrix dimensions (NX = NXH - NXL + 1 and NY = NYH - NYL + 1) and are calculated by MAKFILE and FSAVE. I(8) and R(10) can be defined by the user <u>just before</u> a call to create a permanent matrix file. /HEADIT/ holds header information from the file just entered after a CALL to ENTER, ADZEROS or ENTER.

Processing matrix data can be copied from one file to another by a CALL COPY(IN,IO). This will <u>not</u> work for complex processing image files which cannot be copied.

3. General remarks about processing.

Throughout, the software matrices are referenced by their TAPE integer in the SUBROUTINE CALL and all TAPE's referenced must be declared in the PROGRAM card along with INPUT and OUTPUT. The routines available can be categorised:

General operations on matrices.

Fast transformations.

Direct product filtering in a transform domain.

Reading matrix elements into FORTRAN variables.

Output and display of matrix data.

Specialised operations on matrices.

COMMON//R(4069) is used for storage of matrix element data

by all the SUBROUTINES.

The following sections consist of a list of SUBROUTINES with an explanation of their function.

4. General operations on matrices.

DIFF(I1,I2,I3) Subtracts I2 from I1, resulting matrix I3.

ADD(I1,I2,I3) Adds I1 to I2, resulting matrix I3.

DIVIDE(I1,I2,I3) Kronecker (element to element) division of I1 by I2, resulting matrix I3. If $I2_{ij} = 0$ then $I3_{ij} = 0$.

KROPROD(I1,I2,I3) Kronecker product of I1 and I2, resulting matrix I3.

SUM(I1,TOTAL) Returns sum of all elements of I1 in TOTAL.

SUMIT(I1,NXL,NXH,NYL,NYH,TOTAL) Returns sum of all elements of sub-matrix NXL to NXH,NYL to NYH in TOTAL.

INVROW(I1,I2) Reverses the column order of I1, result I2.

TSPOSE(I1,I2) Transposes I1, result I2.

INVCOL(I1,I2) Reverses the row order of I1, result I2.

UNITM(I1) Sets up the unit matrix $I1 = \delta_{ij}$.

CONSTM(I1,CONST) Sets up the matrix $I1_{ij}$ = CONST for all ij.

CHECKM(I1) Sets up the matrix $I1_{ij} = (-1)^{(i+j-2)}$

MULTP(I1,I2,ES) Calculates $I2_{ij} = I1_{ij} \times ES$.

5. Fast transformations.

The multiplication of two large matrices requires a very large number of operations, N^2 (single multiplication and single addition) which would take an extremely long time to perform when N is large. However some unitary matrices used for orthogonal transformations of matrices

can be factorised into a set of sparse matrix factors (with most elements equal to zero) which results in a reduction in the number of operations to 2N ln N, giving considerable savings when N is large. Two such fast transformations are coded in the software; the 2-D Fast Fourier Transform (FFT) and the 2-D Fast Walsh Transform (the sequency ordered form of the Hadamard Transform), both of which would require N^4 multiplications and additions (complex in the FFT case) without the fast transform algorithm of Cooley and Tukey (1965), reference 54, which reduces the number of operations to $4N^2$ ln N. Only real matrices are catered for.

FTF(I1,I2,I3) Takes the 2-D DFT of I1, result I2. I3 used for temporary storage. I1 can be the same as I2. I2 will be complex.

FTR(I1,I2,I3) Takes the 2-D reverse DFT of I1, result I2. I3 used for temporary storage. I1 can be the same as I2. I1 must be complex.

WTF(I1,I2,I3) Takes the 2-D DWT of I1, result I2. I3 used for temporary storage. I1 can be the same as I2.

WTR(I1,I2,I3) Takes the 2-D reverse DWT of I1, result I2. I3 used for temporary storage. I1 can be the same as I2.

The speed of such transformations is important and the CPU seconds required for various values of N are given in figure 85 for the FFT. The FWT is somewhat faster because of the simpler form of the operations. Although the timings include channel time overheads, the major time consumer is the complex multiplication operation. The CDC CYBER 72

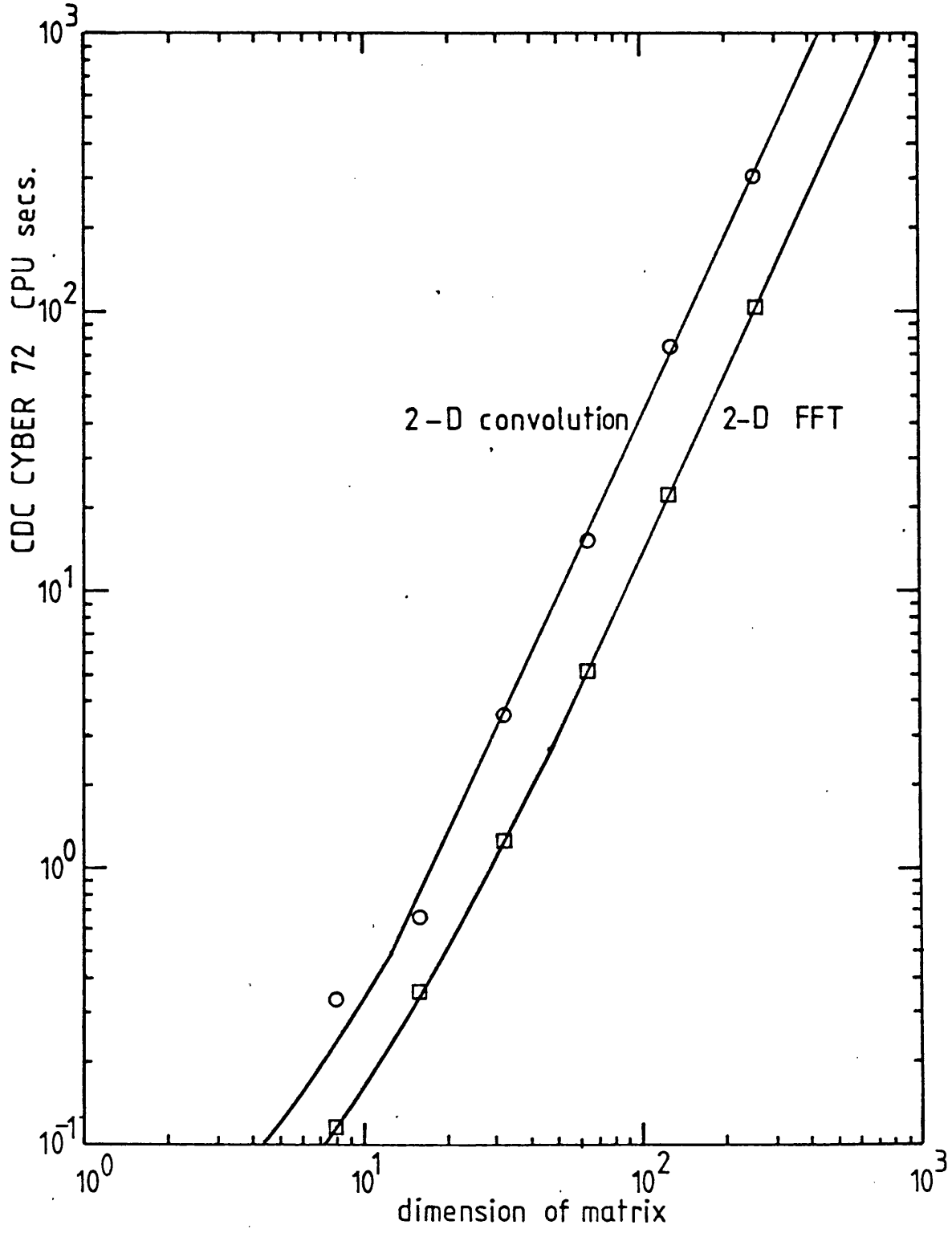


Figure 85. The speed of 2-D convolutions using a CDC CYBER 72.

is not very fast and a more powerful machine might be more restricted by channel time.

6. Direct product filtering in a transform domain.

Discrete filtering normally expressed in terms of a matrix multiplication can be reduced to a direct product operation in an orthogonal transform domain which dialgonalises the discrete filter matrix. Since a circular convolution is diagonalised by the DFT, circular convolutions can be calculated faster by using the FFT algorithm.

CCONV(I1,I2,I3) Performs the circular convolution I1 ⊛ I2 = I3 in the DFD; all three matrices are complex.

CCROSS(I1,I2,I3) Performs the circular crosscorrelation of I1 by I2 = I3 in the DFD; all three matrices are complex.

WFILT(I1,I2,I3,SN) Performs the circular Wiener Filtering of I1 using the response matrix I2 and the noise:signal spectral power ratio SN to give I3; all three matrices are complex. Since SN is independent of spectral frequency both the noise and signal are assumed 'white'.

WIENER(I1,I2,I3,PSDR,RCUT,MPSDR,HRCUT) is similar to WFILT except the signal to noise is entered as a functional fit assuming circular symmetry in the DFD.

$$\left.\begin{aligned} Sw^2/Nw^2 &= PSDR \times EXP(-Nw^2/RCUT) \text{ for w small} \\ Sw^2/Nw^2 &= HPSDR \times EXP(-Nw^2/HRCUT) \text{ for w large} \end{aligned}\right\} \begin{aligned} &RCUT < HRCUT \\ &PSDR > HPSDR \end{aligned}$$

where w is the DFT index, $w = \sqrt{i^2 + j^2}$. The parameters

RCUT, PSDR, HRCUT, HPSDR must be estimated from matrix I1 and SUBROUTINES to aid this are given later.

All the above routines operate in the DFD and matrices must be transformed before calling them. The calling sequence has the form:

```
CALL FTF(I2,I2,ITEMP)   transforms the filter.
.
.
.
CALL FTF(I1,I1,ITEMP)   transforms the data.
CALL CCONV(I1,I2,ITEMP) performs direct product
                        filtering.
CALL FTR(ITEMP,I1,I3)   reverse transformation to give
                        result I1 filtered by I2.
```

In many applications the filter matrix will remain the same and the first operation only needs to be carried out once, providing I2 is not overwritten. Therefore the time taken for the last three calls is most important. Timings for various N are given in figure 85. They are appropriate for all the filtering operations described above. It will be noted that the complete operation takes about 3 times the transformation time and since the direct product operation requires about N^2 multiplications, the transformation is obviously performing about the same number of operations as expected.

Direct product operations in the Walsh Transform Domain have not been coded explicitly, since they can be performed using KROPROD. Such a direct product would perform a cyclic dyadic convolution (shift operations performed modulo 2) of the real matrices. This is also called 'sequency filtering' by analogy with 'frequency filtering' carried out in the Fourier domain.

7. Reading matrix elements into FORTRAN variables.

It is often desirable to process matrix elements separately as normal FORTRAN variables within a program. Four routines allow this possibility:

FINDXY(I1,IX,IY,V) Returns V = $I1_{ij}$, where i = IX and j = IY.

ROW(I1,NR) Places the NR th row of I1 in the first N elements of blank COMMON.

COLUMN(I1,NC) Places the NC th column of I1 in the first N elements of blank COMMON.

CUTIT(I1,XL,YL,XH,YH,NC) Places NC elements of a 'cut' through I1 starting at XL, YL in the matrix domain, the direction specified by the position XH, YH, in the first NC elements of blank COMMON. (X,Y) of the first matrix element is taken to be (1,1). Interpolation is not used and 'nearest element' values are taken for samples along the cut at intervals of ΔX (= ΔY). The major use of this routine is to find the profiles through features in a digital image at arbitrary angles to the x,y axis defined by the image matrix.

8. Output and display of matrix data.

PRINTIT(I1,GLEV) prints the matrix I1 at the line printer as integers using FORMAT 64I2 scaling $I1_{ij} \longrightarrow I1_{ij}/GLEV$, i.e. GLEV is used to scale the numerical output so that it lies in the range $-9 \leqslant I1_{ij}/GLEV \leqslant 99$. If an

element lies outside this range, ** is printed (FORMAT unable to handle data). The position of the output on the listing page must be set by the calling routine.

The CALCOMP electro-mechanical plotter can be used to give a variety of output forms useful for image display. GHOST routines are utilised to drive the plotter, reference 55.

GHOSTIT(I1,NXL,NXH,NYL,NYH,MAXDEN,IPROM) plots a 'density map' of the sub-matrix NXL to NXH, NYL to NYH of I1 spanning the defined PSPACE. IPROM is a function used to convert the REAL matrix values to integer values and must be declared EXTERNAL. The functional transformation can be written by the user. MAXDEN sets the maximum density of crosses to be used for a single element. If an element exceeds the MAXDEN set it will be red and if an element is negative it will be green. The size of the plot can be set by a CALL to PLOTDIM(XSIZE,XSITE) which CALLS PSPACE to give a density map XSIZE mm by XSIZE mm when using default paper dimensions (maximum 300 by 300). GHOSTIT was used to plot the Cygnus Loop data in Part I.

LINEIT(I1,NXL,NXH,NYL,NYH,SCALE) produces an isometric plot of the sub-matrix NXL to NXH, NYL to NYH of I1 using a vertical scale of SCALE pixels/matrix units. The plot is projected off the PSPACE, which can be set by a CALL to PLOTDIM. The simulations presented in Part II used this routine.

CONTIT(I1,NXL,NXH,NYL,NYH,HT,N1,N2) produces a contour plot of sub-matrix NXL to NXH, NYL to NYH of I1 using a contour interval HT. Only heights $N1*HT \leq H \leq N2*HT$ are included and negative values are not catered for.

NXH - NXL + 1 and NYH - NYL + 1 must be ≤ 64. The plot spans the defined vector window which will be the same as the plotter space if WINDOW is not called explicitly.

The GOC microprocesser built by SIGMA ELECTRONICS can be used to give a colour raster display of an image or matrix.

SGPACK(I1,I2,NXL,NXH,NYL,NYH,IPROM) packs the sub-matrix NXL to NXH, NYL to NYH of I1 into hexadecimal form, including GOC commands, onto TAPE I2 using the real to integer conversion function IPROM which must be declared EXTERNAL in the calling routine.

SGSEND(I1,IPIX,IX,IY) will send the packed matrix I1 created by SGPACK to the GOC and it will be displayed on the screen using a pixel size IPIX by IPIX GOC pixels with the bottom left hand corner starting at GOC pixel (IX,IY), where $0 \leq IX \leq 255$ and $0 \leq IY \leq 255$. If the matrix fails to fit on the screen, SGSEND RETURNs control to the calling routine without producing a displayed image.

Three REAL to INTEGER conversion functions are provided for use with GHOSTIT and SGPACK:

IPROM(X) Returns IFIX(X) value between 0 and 15. If ≤ 0 then = 0, if ≥ 15 then = 15.

IPROML(X) Returns IFIX(ALOG(X+1)) and tests for range as IPROM, so result is between 0 and 15.

IPROMNB(X) Same as IPROM but only uses the NO BLINK HEX characters specified by the GOC microprocesser.

9. Specialised operations on matrices.

A number of dedicated routines are available for specialised operations, mostly concerned with image

processing. It is relatively easy to write new routines to perform rather specialised operations so that the task of expanding the current processing ability is easy.

NORM(I1,I2,CONST) normalises the sum of the elements of I1 to CONST to form the resulting matrix I2.

POWER(I1,B,PN,PS) calculates the sum PN and sum of the squares PS of the elements of I1. B is a constant which can be used to offset PS (not very useful!). This can be used to determine the parameter SN for use in WFILT.

PSPEC(I1,I2) calculates the power (amplitude squared) of the elements of the DFD matrix I2 giving the power spectrum matrix I2. The matrix I2 is real and an ordinary processing matrix containing one quadrant of the power spectrum matrix (since the other three quadrants contain exactly the same information it is pointless calculating them).

REBIN(I1,I2,NBINX,NBINY) rebins the matrix I1 by summing the elements of I1 over sub-matrices NBINX by NBINY. The result I2 will, of course, contain zeros for elements (NX,NY) where NX > N/NBINX and NY > N/NBINY.

SCALE(I1,I2,BMAX) scales the elements of I1 so that the largest element has the value BMAX, producing a scaled matrix I2.

ENTROPY(I1,ENTPY) calculates the configurational entropy ENTROPY = $\sum_{ij} I1_{ij} \ln I1_{ij}$ of matrix I1. Values $I1_{ij} \leq 0$ are ignored.

CHISQD(I1,I2,I3,CHIS) returns CHIS as the reduced χ^2 between matrices I1 and I2 using the variance matrix I3:

$$CHIS = \sum_{ij} (I1_{ij} - I2_{ij})^2 / I3_{ij}$$

NOISE(I1,I2) simulates counting statistics using a random

number generator taking $I1_{ij}$ as the average count producing $I2_{ij}$, a sample of the statistical distribution about the mean.

ZTRANS(I1,I2,F) performs a Z transformation on matrix I1 specified by the function F, which must be declared EXTERNAL by the calling routine:

$$I2_{ij} = F(I1_{ij})$$

DFTAMP(I1,I2,I3) calculates the amplitudes of the complex matrix I1 using I3 for temporary storage and creates an amplitude matrix I2 with ordinary processing form.

RADPSD(I1,SPEC) returns the radial density profile of matrix I1 in array SPEC(N/2+1). This routine is used in conjunction with PSPEC to find the radial power spectral density function of an image for use in calculating the signal to noise parameters for WIENER.

SYMMAT(I1,I2,F) produces a matrix I2 using I1 for temporary storage. I1 represents the radial density function F centred on element $I1_{11}$ and aliased to give a digital representation of the radially symmetric function specified by F.

FRESNEL(I1,BN,ITYPE) creates a matrix representation of a Fresnel zone plate with BN rings. If ITYPE = 2 then centre transmitting, otherwise not. This routine was used for the simulations in Part II. The routine used for generating the pseudo-random mask pattern was adapted for use with this software system from reference 56.

POINTM(I1,NX,NY) sets up the matrix $I1_{ij} = 1$ if NX = i and NY = j. $I1_{ij} = 0$ otherwise.

Note: This software is still under development for use in HEAO-B image analysis but the underlying structure has been finalised.

Acknowledgements.

I thank the following people for the stated reasons; Dr. Kenton Evans for his continuous encouragement and criticism, Mike Kayât for putting up with my criticism and allowing me to mess about with his data, Dan Rolf for his help in the image analysis and especially for reluctantly becoming a magnetic tape king, Ray Hall for teaching me the art of staying sane while using the two crystal spectrometer and Margaret Lewis for allowing me to destroy and rebuild her amazing crystal diffraction software. Mark Sims for his help in the burst monitor simulations and Professor Ken Pounds for tolerating the research I was pursuing. Finally Hilary Willingale for typing these acknowledgements and the preceding thesis.

I thank the U.K. SRC and Silca and Nuclear Products for jointly financing the CASE award which supported my research for two years.

REFERENCES

1. V.H.REGENER, PHYS.,REV(LETT),84,,161,1952.
2. P.W.HAWKES,OPTIK,VOL.40(NO.5),539,1974.
3. N.WIENER,"THE EXTRAPOLATION, INTERPOLATION AND SMOOTHING OF STATIONARY TIME SERIES",P84,JOHN WILEY & SONS, INC. N.Y.,1949.
4. C.W.HELSTROM,J.OPT.SOC.AMER.,VOL.57,NO.3,279,1967.
5. H.C.ANDREWS AND B.R.HUNT,"DIGITAL IMAGE RESTORATION",PRENTICE HALL SIGNAL PROCESSING SERIES, ED.A.V.OPPENHEIN,1977.
6. B.R.FRIEDEN,J.OP.SOC.AMER.,62,511,1972.
7. J.G.ABLES,ASTRON.ASTROPHYS.SUPPL.,15,383,1974.
8. R.KIKUCHI AND B.H.SOFFER,J.OP.SOC.AMER.,VOL.67,NO.12,1977.
9. S.F.GULL AND G.J.DANIELL,NATURE,272,636,1978.
10. J.R.KLAUDER AND E.C.G.SUDARSHAN,"FUNDAMENTALS OF QUANTUM OPTICS" W.A.BENJAMIN,INC.,1968.
11. S.J.WERNECKE AND L.R.D'ADDARIO,IEEE TRANS.COMPUTERS,C-26,351,1977.
12. L.BRILLOUIN,"SCIENCE AND INFORMATION THEORY",ACADEMIC PRESS INC.N.Y. 2ND ED.,1962.
13. R.WILLINGALE, M.SC. DISSERTATION UNIV. OF LEICESTER,1976.
14. H.C.ANDREWS,"COMPUTER TECHNIQUES IN IMAGE PROCESSING",ACADEMIC PRESS,1970.
15. M.A.KAYAT,PH.D. THESIS UNIV. OF LEICESTER,1979.
16. P.B.FELGETT,PH.D. THESIS CAMBRIDGE UNIV.,1951.
17. R.WILLINGALE AND T.CARTER,M.SC. PROJECT REPORT UNIV. OF LEICESTER,1976.
18. R.H.DICKE,AP.J.,VOL.153,L101,1968.
19. I.G.ABLES,PROC.ST.SOC.AUSTRALIA,1,NO.4,1968.
20. T.M.PALMIERI,ASTROPHYS. AND SPACE SCI.,20,431,1974.
21. J.GUNSON AND B.POLYCHRONOPULOS,MON.NOT.R.ASTR.SOC,177,485,1976.
22. S.MIYAMOTO,OSAKA-ATOM-1-7701,1977.
23. MIT/UNIV. OF LEIC. PROPOSAL TO NASA AO-OSS-2-76,DEC.1976.
    MIT/UNIV. OF LEIC. PROPOSAL TO NASA AO-OSS-2-78,NOV.1978.
24. M.W.ZEMANSKY,"HEAT AND THERMODYNAMICS",MCGRAW-HILL,1968.

25. W.H.BRAGG AND W.L.BRAGG,"X-RAYS AND CRYSTAL STRUCTURE",CHAP.3,5TH ED., LONDON:BELL,1925.

26. L.V.AZAROFF(ED.),"X-RAY SPECTROSCOPY",MCGRAW-HILL,INC.,1974.

27. C.G.DARWIN,PHIL.MAG.,27,315,1914,(KINEMATICAL THEORY) PHIL.MAG.,43,800,1914,(DYNAMICAL THEORY).

28. P.P.EWALD,ANN.PHYSIK,49,1,1916. ANN.PHYSIK,49,117,1916.

29. M.VON LAUE,NATURWISS,10,133,1931.

30. R.W.JAMES,"THE OPTICAL PRINCIPLES OF THE DIFFRACTION OF X-RAYS" 5TH ED.,G.BELL & SONS,LTD.,LONDON,1962.

31. L.V.AZAROFF ET AL,"X-RAY DIFFRACTION",MCGRAW-HILL,INC.,1974.

32. D.T.CROMER,ACTA CRYST.,19,244,1965.

33. D.T.CROMER AND J.T.WABER,ACTA CRYST.,18,104,1965.

34. D.T.CROMER AND D.LIBERMAN,LASL REPORT LA-4403,1970.

35. J.C.SLATER,PHYS.REV.81,385,1951.

36. W.KOHN AND L.J.SHAN,PHYS.REV.,140,A1133,1965.

37. M.LEWIS AND D.UNDERWOOD,M.SC. PROJECT REPORT UNIV. OF LEICESTER,1975.

38. P.A.MAKSYM,M.SC. PROJECT REPORT UNIV. OF LEICESTER,1976.

39. I.G.PARRATT AND C.F.HEMPSTEAD,PHYS.REV.,VOL.94,NO.6,1954.

40. K.D.EVANS,B.LEIGH AND M.LEWIS,X-RAY SPECTROSCOPY,VOL.6,NO.3,1977.

41. K.KOHRA,J.PHYS.SOC.JAPAN,17,589,1962.

42. B.LEIGH,PH.D. THESIS UNIV. OF LEICESTER,1974.

43. R.HALL,PH.D. THESIS UNIV. OF LEICESTER,1979.

44. M.W.CHARLES,PH.D. THESIS UNIV. OF LEICESTER,1968.

45. B.L.HENKE,ADVANCES IN X-RAY ANALYSIS(PLENUM PRESS N.Y.),VOL.7,460,1964.

46. B.L.HENKE,R.C.PERERA AND R.H.ONO,TECHNICAL REPORT AFOSR,72-2174,1974.

47. W.H.ZACHARIASEN,"THE THEORY OF X-RAY DIFFRACTION IN CRYSTALS", DOVER PUBLICATION INC.N.Y.,1967,(FIRST PUBLISHED 1945).

48. R.G.WYCKOFF,"CRYSTAL STRUCTURES",JOHN WILEY AND SONS N.Y.,1968.

49. J.P.PYE,K.D.EVANS AND R.J.HUTCHEON,MON.NOT.R.ASTR.SOC.,178,611,1977.

50. W.L.BRAGG AND J.WENT,PROC.R.SOC,IIIA,691,1926.

51. A.SAMSON AND M.HAYES,M.SC. PROJECT REPORT UNIV. OF LEICESTER,1978.

52. J.KORRINGA,E.L.JOSSEM,R.LIEFELD,R.E.KVARDA,C.H.SHAW,REPORT NO.7 CONTRACT N6ONR-22521 NR 017 606,OHIO STATE UNIV.,COLUMBUS,OHIO.

53. G.A.SAWYER,F.C.JAHODA,F.L.RIBE AND T.F.STRATTON,J.QUANT.SPECTOSC. RAD.TRANSFER,2,467,1962.

54. J.W.COOLEY AND J.W.TUKEY,MATH.COMP.VOL.19,297,1965.

55. UNIV. OF LEICESTER COMPUTER LAB.,"CULHAM GHOST USERS MANUAL".

56. R.J.PROCTOR,G.K.SKINNER AND A.P.WILLMORE, SUBMITTED TO MON.NOT.R.ASTR.SOC.,1978.

'Analysis of X-ray Images and Spectra' - R. Willingale 1979

Four research projects concerned with X-ray data analysis are reported. Part I considers methods of deblurring and noise suppression for improving the quality of images produced by grazing incidence optics. Two well established techniques, Fourier filtering and the Maximum Entropy Method, are studied in detail and digital computer software for implementing these methods, along with other image processing tasks, is documented in Appendix I. This software package was used to process astronomical X-ray data from a sounding rocket flight that looked at the Cygnus Loop supernova remnant and the results are presented in Part I as a demonstration of the techniques developed. Part II applies the deconvolution methods developed in Part I to decoding data from coded mask X-ray telescopes. Computer simulations of proposed X-ray burst monitors are reported, giving a realistic assessment of the usefulness of such instruments and providing a comparison between the various decoding methods available. Part III reports an attempt to apply the Maximum Entropy Method to the analysis of proportional counter anode pulse height data. The algorithm developed was used to analyse the pulse height spectra from the Cygnus Loop observation and the results of this analysis are included. Part IV is concerned with the calibration of Bragg crystal spectrometers. Theoretical calculations and experimental measurements of the response of three crystal types - Langmuir-Blodgett lead stearate multilayers, gypsum and beryl - are presented. The agreement between theory and measurement is good and the results provide excellent calibration data for use in subsequent spectral analysis using these crystals as Bragg analysers. The results of scans across the sulphur k and aluminium k absorption edges in gypsum and beryl respectively are reported, providing experimental evidence of k electron resonance in these two atomic types.