UNIVERSITY OF SOUTHAMPTON
Faculty of Science
Department of Physics

The Optimisation of Discrete Pixel Coded Aperture Telescopes

by

Ian David Jupp

A Thesis submitted for the degree of Doctor of Philosophy

September 1996

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE

PHYSICS

Doctor of Philosophy

THE OPTIMISATION OF DISCRETE PIXEL CODED APERTURE TELESCOPES

by Ian David Jupp

The launch at the end of the last decade of NASAs Compton Gamma-Ray Observatory and the Franco-Soviet gamma-ray observatory GRANAT, signified the beginning of a new multi-wavelength approach to gamma-ray astronomy. Together with currently operating X-ray and high energy gamma-ray telescopes, a quasi-continuous coverage from keV to GeV energies has been established. The future goals of gamma-ray astronomy in the tens of keV to several MeV range must therefore be defined in the context of complementarity. The next generation of gamma-ray observatories will be led by ESAs INTEGRAL mission which will address the fine spectroscopy and accurate positioning of celestial gamma-ray sources in the 10 keV to 10 MeV energy range. The observational requirements will be achieved by a payload utilising the coded aperture imaging technique in combination with pixellated detector arrays for both high resolution imaging and spectroscopy. It is within the context of the INTEGRAL mission that the majority of research contained in this thesis has been conducted.

Following reviews of the astrophysical goals for the INTEGRAL imager and the developments which have made such a mission possible, the fundamental characteristics of coded aperture telescopes based on pixellated detector arrays are investigated in detail. The point source capability of such a telescope is examined first, resulting in a relationship describing the sensitivity dependence on the detector pixel/mask element sampling ratio. To complement this work, the sensitivity of the coded mask technique is then examined with respect to extended sources. It is shown that mask patterns optimised for high angular resolution imaging are less sensitive to extended sources. The situation can be improved at the expense of the point source capability by reducing the resolution of the mask pattern. However, several techniques including the use of wavelet transforms and dual-scale mask configurations, are shown to be capable of either extracting source-scale information or improving the extended source sensitivity. Throughout these chapters it has been assumed that a coded mask observation leads directly to a shadowgram which can then be deconvolved. On the contrary, it is shown that for a pixellated detector array, deconvolution can proceed only after some form of incident pixel reconstruction (IPR) has been applied to account for multiple pixel interactions. The optimum IPR algorithm is derived for the INTEGRAL imager EIDOS and the effects of IPR on telescope performance are discussed. Although it is demonstrated that in certain situations correlation analysis is the optimum method of image reconstruction, the reliability of correlation comes into question under extreme observing conditions. Consequently, the final chapters focus on maintaining the correlation image quality through the use of pre and post-processing techniques. The penultimate chapter outlines the theoretical development of these techniques and the final chapter covers the results of their application to a prototype coded aperture imager based on EIDOS detector technology.

Contents

List of Figures

PREFACE

The majority of the work in this thesis has been conducted under the auspices of the INTEGRAL Imager both under the EIDOS design and more recently, but to a lesser extent, the IBIS design. The original research which forms Chapters 3, 4, 5, 6 and 7 is entirely the author's own work. The design and construction of the prototype imager outlined in Chapter 7 was conducted in collaboration with Dr. A.J. Bird, Dr. T. Carter, Dr. A.R. Green and Dr. F. Lei. All software, both Monte Carlo and data analysis, unless otherwise indicated, has been developed solely by the author.

ACKNOWLEDGEMENTS

There are a number of people who deserve recognition now that this thesis is finally complete. First and foremost I must thank Helen for her continual encouragement and motivation throughout the last three years and in particular, for those early morning lifts into the lab which were really appreciated ?! Also, I would like to thank my parents and sister for their support over the last three years and for my sister's lunchtime visits that allowed me on the odd occasion to avoid going to lunch with a bunch of Astrophysicists ! I would like to thank my supervisor, Tony Dean, for his continual guidance throughout my Ph.D. and for the countless stimulating discussions that have made the research in this thesis possible. I must thank Andy Green for his support and advice in just about every aspect of my Ph.D., for proof reading my entire thesis (!), for his day-to-day guidance in conducting research, for his *always* constructive criticism and for his constantly bright and cheerful attitude to life in general.

For those fast(ish) in-swinging, body-line (head-line ?) deliveries that livened up those late summer afternoons and also for his help and advice when trying to set up the data aquisition system used for the work in Chapter 7, I must thank Tony Bird. For answering my seemingly endless list of questions on the intricacies of GEANT and for stopping a cross-court, forehand, top-spinning winner with his nose (resulting in a nose-bleed and set-point to myself and TC !) I must thank Fan Lei. For the easiest of introductions into Ph.D. life, for teaching me the fundamentals of coded aperture imaging, for the endless impersonations of Jimmy Saville and Les Dawson and for the odd good joke here and there, and for his continual support over the last three years, I must thank Kevin Byard.

Now I would like to thank those members of the group who have made the last three years so enjoyable both from a research perspective and of course from a social perspective : Jimmy (Tony Carter), Brian-fish (Graeme Hills), Donald Trusas (Ian Pleasants), Steve McCabe, Rochey (Paul Roche – for managing to get me on a three week observing run to Australia !), John Willoughby, Nutter-steve (Steve Pinfold), Chris Lee, Chris Everall, Chris Leach, Terry (Simon Clarke), Pablito (Pablo Reig), Nacho (Ignacio Negeruela), Esther Perez, Irini Sakelliou, Sue Wright, Max Palmer and Paul Durrant.

I must also mention a group that I'm X-tremely grateful to for making me do stupid things like jump out of planes, cycle 100 miles at night, cycle over mountain passes, freeze various bits off in the middle of Southampton Water for the sake of 60 seconds of planing and for making me run down a hill with what was supposedly a hang-glider attatched to my back for 30 seconds of air-time and a rather slippery landing in several fresh cow-pats ! For these experiences I'm forever indebted to the members of the X-treme Division of the Astronomy Group : founder members Planet-X (Grant Crossingham) and Antoine-X (Anthony Lawson), and late arrivals Rob-X (Rob Evans) and Chad-X (Graham Chadwick).

The Dutch contingent (Corr, Frank and Wauter) also deserve a special mention for making the evening of the 18th June 1996 so enjoyable (4-1) ! For introducing me to the delights of editing postscript, but more importantly for saving me many, many hours in producing some of the figures for this thesis, I thank Andy Newsom. For keeping goal rather well in the Group's successful 5-a-side football history (!), I thank Chris Lawton.

I would also like to thank the other members of the Astronomy Group who have not yet been mentioned : Malcolm Coe, Ian McHardy, Mike Merrifield and Tom Marsh for their advice and some interesting discussions over the years.

STARLINK provided the computing facilities used for this thesis and so a special thanks must go to Laurence Jones (where ever he is now), Mike Hill and Simon Harris for keeping the SUNS alive (and the VAX for my first few months).

Finally, and for several reasons, I would like to thank David Ramsden. Firstly, it was on completion of my 3rd year undergraduate project with David that I decided that research may not be so bad after all ! Secondly, if it wasn't for his persistence in tracking me down whilst I was travelling around Australia, I would not have had the opportunity to start this Ph.D. Finally, without his patience and financial support over the last few months, the completion of this thesis would have been considerably more difficult.

For Mum, Dad and Helen.

Chapter 1

The Gamma-Ray Sky

Overview

There are several fundamental imaging prerequisites for an astronomical gamma-ray telescope that ensure the basic scientific demands of the mission are met. The telescope must be able to identify and resolve individual point sources, often in crowded regions of sky. It must be able to characterise extended emission with angular scales that range from point-like hot-spots to emission that traces the large-scale structure of the Galaxy itself. In addition, both point and extended source capabilities must achieve their respective goals with high sensitivity for both broad band and line spectral coverage. Logically, it is the nature of the gamma-ray sky coupled with the scientific objectives of the telescope that determine how the imaging configuration is optimised. In this opening chapter, consideration is given to the first point, namely the nature of the gamma-ray sky in the 15 keV to 10 MeV wave-band, principally from an imaging perspective but since the two are inextricably linked, also from a spectroscopic perspective. The 15 keV to 10 MeV energy band which links the classical X-ray and relativistic particle gamma-ray regimes, has not been chosen arbitrarily for this review but reflects the fact that the majority of the research contained in this thesis was conducted under the auspices of the INTEGRAL imager. The INTEGRAL mission will beq the first of the next generation of gamma-ray observatories to operate in the 15 keV to 10 MeV wave-band.

1.1 Introduction

The field of gamma-ray astronomy encompasses a wide range of objects and physical processes both Galactic and extragalactic. The penetrating nature of photons at these energies renders the imaging telescope a powerful tool for probing the very heart of some of the most violent astrophysical phenomena known throughout the Universe. The launch of NASA's COMPTON Gamma-Ray Observatory (GRO) in 1992 following the launch of Franco-Soviet GRANAT satellite at the end of the last decade, hailed the beginning of a new era in high energy astrophysics.

With the successful operation of WATCH, ART-P and SIGMA onboard GRANAT, and of BATSE, OSSE and COMPTEL onboard CGRO the gamma-ray sky has already been mapped on several angular scales in the energy range of the proposed INTEGRAL imager, and a large number of new gamma-ray sources have already been discovered. One of the most important aspects of the GRANAT and CGRO missions is their continuous broad-band spectroscopic coverage over several decades of photon energy from a few keV to several MeV. Such broad-band coverage is essential for a better understanding of complex source mechanisms in which the physical processes of several different energy regimes are inextricably connected.

Both CGRO and GRANAT instruments have relied on sky maps, excellent timing capabilities and moderate to fair spectral resolution to achieve their scientific objectives. Consequently, although the observations to date have provided the community with a valuable insight into the morphological structure of celestial gamma-ray sources, it is still so often the case that the corresponding spectra are not 'deep' enough to significantly constrain the emission mechanism(s). The unprecedented capabilities of both GRANAT and CGRO have provided answers to many long standing problems as well as raising new astrophysical questions, but more importantly, they have highlighted the fact that many questions remain unanswered.

The goal in this chapter is to present an overview of the gamma-ray sky with reference to the imaging capabilities required to achieve the scientific goals, thus setting the scene for the imaging development work that forms the basis of this thesis. However, as a starting point, by considering the possible emission mechanisms constraints are placed on the physical environments that lead to the production of gamma-ray photons.

1.2 Gamma-ray Emission Mechanisms

The mechanisms by which gamma-ray photons can be created have been described in great detail by many authors using precise quantum mechanical treatments (for example, [1]), consequently only a qualitative description is given here.

1.2.1 Blackbody Radiation

An optically thick region of material in thermal equilibrium, radiates a continuous rather than discrete spectrum since the mutual Coulomb interactions between the atoms/ions in the region result in a multitude of adjacent quantum states indistinguishable from a continuous band of permitted transitions. Following Einstein's treatment of an optically thick material in thermodynamic equilibrium at temperature T, the emission spectrum can be shown to have the blackbody spectral form given by Planck's radiation formula (Eq. 1.1) :

$$B_\nu(T) = \left(\frac{2h\nu^3}{c^2}\right)\left(\frac{1}{exp\left(\frac{h\nu}{kT}\right) - 1}\right) \tag{1.1}$$

Wien's displacement law, which relates the maximum frequency of emission to the effective temperature of the radiating body, can be derived by maximising Eq. 1.1 to give $h\nu_{max} = 2.82\ kT$. Thus, in order for the peak blackbody emission to occur at energies greater than 15 keV, temperatures greater than 10^8 K are required.

1.2.2 Bremsstrahlung Radiation

Bremsstrahlung radiation is emitted by an electron as it is accelerated in the Coulomb field of an atom/ion. It should be noted that the collision of like particles (electron-electron, proton-proton) produces only 'virtual' bremsstrahlung photons and so only collisions between different particles need to be considered. For astrophysical processes, where bremsstrahlung photons are likely to come from a hot, optically thin plasma, it can be shown that a thermal distribution of electron speeds gives a total emission (per unit time per unit volume per unit frequency range) of :

$$\frac{dW}{d\omega\ dV\ dt} = \frac{32\pi e^6}{3mc^3}\left(\frac{2\pi}{3km}\right)^{1/2} T^{-1/2} Z^2 n_e\, n_i\, \overline{g}_{ff}(v,\omega) exp\left(\frac{-h\nu}{kT}\right) \quad , \qquad (1.2)$$

where m and v are the electron rest mass and velocity, and n_e and n_i are the electron and ion densities respectively. The factor $\overline{g}_{ff}(v,\omega)$ is a correction factor (the velocity averaged Gaunt factor) to ensure a smooth transition from the classical case when the impact parameter is large, and the quantum mechanical case when the impact parameter is small and the uncertainty principle becomes important.

The $exp\left(\frac{-h\nu}{kT}\right)$ term is simply a consequence of the initial electron velocity distribution – a non-thermal distribution will produce a different spectral profile. For $\frac{h\nu}{kT} \gg 1$, $\overline{g}_{ff} \sim 1$, and so the spectral profile cuts-off exponentially. For the range $10^{-4} < \frac{h\nu}{kT} < 1$, $\overline{g}_{ff}$ lies in the range 1 to 5, and so the spectral profile flattens out considerably. At lower photon energies the plasma will become optically thick and the spectral profile then simply follows the Raleigh-Jeans region of the blackbody spectrum. Thus for $\frac{h\nu}{kT} \ll 1$ the emission is proportional to ν^2. For bremsstrahlung emission above 15 keV, plasma temperatures greater than 10^7 K are required.

1.2.3 Magneto-bremsstrahlung Radiation

A charged particle moving in the presence of a magnetic field will gyrate around the field lines and emit radiation as a result of its radial acceleration. Generalising the equations of motion for relativistic particles in static magnetic fields gives Eq. 1.3 for the electromagnetic force on the particle :

$$\frac{d}{dt}(\gamma m\mathbf{v}) = q(\mathbf{v} \times \mathbf{B}) \qquad (1.3)$$

The velocity component perpendicular to the magnetic field determines the gyration frequency in that plane. Equating electromagnetic and centripetal forces, the gyration frequency can be determined :

$$\nu_g = \frac{q\ B}{\gamma\ 2\pi\ m} \quad , \qquad (1.4)$$

where γ is the Lorentz factor. The parallel velocity component of the particle will result in

a helical precession of the particle along the magnetic field lines. Again assuming electrons are the radiating particles, in the non-relativistic scenario this form of radiation is known as cyclotron radiation. The frequency of emission is simply given by the electron gyration frequency. As a result of the dipolar radiation pattern along the instantaneous direction of motion, the time dependence of the E-field as seen by an observer is sinusoidal. Thus the emission spectrum for single-speed electrons is simply a line at the gyration frequency. From Eq. 1.4 it follows that to observe a cyclotron line from the first harmonic above 15 keV, magnetic field strengths of 10^8 T are required.

As the Lorentz factor increases, several relativistic effects come into play : the previously dipolar radiation pattern is now more beamed in the direction of motion with an open angle $\sim \frac{2}{\gamma}$; the time dependence of the observed E-field is no longer sinusoidal but is now composed of higher harmonics of the fundamental gyration frequency as a result of the relativistic beaming ; the emission spectrum now consists of several lines at integer multiples of the gyration frequency. As γ increases further, harmonics at higher frequencies become significant, until $v \sim c$, at which point the emission spectrum consists of a large number of harmonic lines. It can be shown that as a consequence of both relativistic beaming and Doppler effects, the observed pulse width Δt_o is a factor of γ^2 shorter than the non-relativistic gyration period, or in terms of ν_g :

$$\frac{1}{\Delta t_o} \approx \nu_c \approx \gamma^3 \nu_g \tag{1.5}$$

where ν_c is some characteristic frequency of the resulting 'synchrotron' spectrum. If instead of a single-speed electron velocity distribution, a power-law distribution is present, the many cyclotron harmonics in the synchrotron spectrum form one continuous spectrum. If the power-law spectral index of the electron velocities is $-p$, it can be shown that the spectral profile in the optically thin scenario is proportional to $\nu^{-(p-1)/2}$. When the emitting region becomes optically thick, synchrotron self absorption occurs and the spectral profile tails of as $\nu^{5/2}$, which is steeper than the Raleigh-Jeans profile since the electron velocities are non-thermal. From Eq. 1.5 it is evident that synchrotron emission above 15 keV requires magnetic field strengths of $\frac{10^8}{\gamma^2}$ T.

1.2.4 Inverse Compton Scattering

Normal Compton scattering occurs when a stationary electron gains kinetic energy from a collision with a photon which in turn loses energy. For low photon energies, $h\nu \ll mc^2$, this reduces to the classical case of elastic Thomson scattering. However, whenever the electron has sufficient kinetic energy compared to the photon, net energy may be transferred from the electron to the photon, boosting it to a higher frequency. Inverse Compton scattering is most probable when, in the centre of momentum frame of the collision, the photon energy is much lower than the electron rest mass energy, i.e. $\gamma h\nu \ll mc^2$.

For relativistic velocities it can be shown using the Doppler shift formulas and the inverse equivalent of the Compton scattering equation, that for $\gamma h\nu \ll mc^2$, the photon energy is boosted by a factor of γ^2 on collision :

$$\nu_1 \approx \gamma^2 \nu_0 \tag{1.6}$$

Again the actual spectral profile arising from the Comptonisation process depends strongly on the velocity distribution of the scattering electrons and the spectral distribution of the seed photons. For a non-thermal distribution of electrons scattering seed photons of a single energy, the spectral profile can be calculated as [2] :

$$I(\nu)\, d\nu = \frac{3\sigma_T c}{16\gamma^4} \frac{N(\nu_0)}{\nu_0^2} \nu \left[2\nu\, ln\left(\frac{\nu}{4\gamma^2\nu_0}\right) + \nu + 4\gamma^2\nu_0 - \frac{\nu^2}{2\gamma^2\nu_0} \right] d\nu \ , \tag{1.7}$$

where ν_0 is the seed photon frequency, $N(\nu_0)$, is the photon number density, and σ_T is the Thomson cross-section. At low frequencies, the term in square brackets is a constant and hence the spectral profile is proportional to ν.

An important example of the Comptonisation process, is the up-scattering of synchrotron radiation. The power-law distribution of electron velocities that give rise to the characteristic $\nu^{5/2}$ and $\nu^{-(p-1)/2}$ synchrotron components, will also inverse Compton scatter the synchrotron photons to higher energies. The process is known as synchrotron self-Compton and is capable of producing MeV gamma-rays from optical seed photons for $\gamma \approx 10^3$. The synchrotron and synchrotron self-Compton profiles have the same spectral characteristics.

Decay Chain	Mean life (yrs)	Emission (MeV)
$^{56}Ni \rightarrow ^{56}Co \rightarrow ^{56}Fe$	0.31	e^+, 0.847,1.238, 2.599
$^{57}Co \rightarrow ^{57}Fe$	1.1	0.014, 0.122
$^{22}Na \rightarrow ^{22}Ne$	3.8	e^+, 1.275
$^{44}Ti \rightarrow ^{44}Sc \rightarrow ^{44}Ca$	$\sim$ 78	e^+, 0.068, 0.078, 1.156
$^{26}Al \rightarrow ^{26}Mg$	1.1×10^6	e^+, 1.809
$^{60}Fe \rightarrow ^{60}Co \rightarrow ^{60}Ni$	2.2×10^6	0.059, 1.173, 1.332

Table 1.1: Key isotopic decay chains from nucleosynthesis processes.

1.2.5 Pair Production and Annihilation

If a photon with enough energy ($E_\gamma > 2m_ec^2$) passes through the Coulomb field of an atomic nucleus *or* if two photons with enough energy collide, the photon(s) may be converted into an electron-positron pair with a total kinetic energy equal to the difference between E_γ and $2m_ec^2$. Similarly, an electron-positron pair may annihilate in either the para-positronium state in which two anti-parallel 511 keV photons are produced (25 % probability), or in the ortho-positronium state in which 3 photons are produced in a plane with a total energy less than or equal to 511 keV (75 % probability). The annihilation mode is strongly dependent on the surrounding environment and thus line or continuum annihilation emission has a powerful diagnostic potential.

1.2.6 Nuclear Gamma-ray Lines

Nuclear gamma-rays come from two principle mechanisms : the spontaneous decay of radio-isotopes or the induced decay of excited nuclei. The principle gamma-ray production isotopes of interest in astronomy are outlined in Table 1.1 [3]. As will be seen, such decay chains are expected from nova outbursts and supernova explosions, where in particular the $^{56}Ni \rightarrow ^{56}Co \rightarrow ^{56}Fe$ chain and the longer lived $^{44}Ti \rightarrow ^{44}Sc \rightarrow ^{44}Ca$ chain are of importance. Similarly, with respect to mapping the Galactic sites of nucleosynthesis, the $^{26}Al \rightarrow ^{26}Mg$ decay line at 1.809 MeV has considerable diagnostic potential.

Nuclear excitation followed by de-excitation and the emission at discrete gamma-ray energies, can occur in a number of ways : in radioactive decay chains, the daughter nucleus is very often produced in an excited state ; collisions between fast moving nuclei and the interstellar medium (ISM) will cause both the fragmented target nucleus and the

products to produce gamma-rays.

1.3 The Galactic Gamma-ray Sky

1.3.1 Compact Objects

One of the most compelling objectives of the next generation of gamma-ray telescopes in the 15 keV to 10 MeV band, is the study of stellar compact objects in both binary systems and in the isolated remnants of type II supernovae explosions.

Cataclysmic Variables

The degenerate cinders of supergiant stars, white dwarfs (WD), with surface temperatures less than 10^5 K, are merely the thermal remnants of the parent star destined to become cold degenerate matter when isolated from any source of power. On the contrary, WDs in binary systems are effectively deep gravitational wells capable of drawing material from the distended layers of a companion star that is filling its inner Lagrangian point (Roche Lobe overflow) and accelerating this material to shock temperatures greater than 100 keV. Since WD systems are at least 10^4 times more common in the Galaxy than neutron star systems, they are ideal for the study of relatively high energy accretion mechanisms. However, the density of sources suggests that as instrument sensitivities are driven higher, high resolving powers are essential for identifying the objects at other wavelengths.

In non-magnetic classes of cataclysmic variable (CV), the accreting material forms a disc around the WD. Study of the soft gamma-ray spectra/flux from these systems when in outburst (Novae) may shed light on the interactions between the inner edge of the accretion disc and the WD surface. Novae outbursts are thought arise from the ignition of the inner accretion disc material, a theory supported by the anti-correlation between the gamma-ray flux and the optical luminosity of the outburst. For strongly magnetised (10^3 T) CVs , the accretion flow is channelled directly onto the polar regions of the WD. To date there is no evidence that this type of WD system is responsible for Novae outbursts. However, the gamma-ray variability of these systems, linked to quasi-periodic oscillations (QPOs), will place constraints on the interactions between the accretion column and WD surface. Additionally, in both magnetic and non-magnetic systems, shock fronts

arising from the deceleration of the accreting plasma should produce a soft gamma-ray bremsstrahlung continuum. Reprocessing of the gamma-ray continuum by the WD photosphere suggests a soft X-ray blackbody component should be evident in CV spectra and is in fact observed [4].

An example of the necessity of a high resolution imaging capability was the apparent first ever detection at gamma-ray energies of the system identifiable at other wavelengths with the CV V1223 Sgr [5], in which a transient annihilation feature was seen. However, later analysis indicated that the emission was more likely to emanate from the nearby black hole candidate system 1H1822-37 [6].

Neutron Stars

Most neutron star (NS) binary systems show spectra at soft X-ray energies consistent with blackbody radiation from the accretion heated NS surface. Harder bremsstrahlung radiation from accretion shock fronts for either an accretion column in a highly magnetised NS, or for an accretion disc for the non-magnetic systems, is not expected to be visible as a hard tail because the inner accretion regions are optically thick at gamma-ray energies. Similarly, Comptonisation of the soft X-ray blackbody component by the energetic accretion flow is thought to be too efficient to produce a hard tail extending up to gamma-ray energies. Thus, the lack of hard tails in NS systems has been taken as a reliable discriminator between NS and black hole systems. However, contrary to this, SIGMA observations of the NS system KS1731-260 [7] have provided strong evidence of a hard tail emission component extending up to 150 keV. Similarly, a hard tail in GX354-0 has been detected by both GRIP [8] and SIGMA [9] up to 100 keV.

Models to justify the presence of a hard tail essentially focus on the transfer of relativistic bulk energy from the accretion flow to a hard tail through Comptonisation. If Comptonisation is too efficient and the accretion flow is low, the scattering plasma will rapidly cool and the hard tail will not persist. Similarly, at higher accretion rates the flow becomes optically thick to the up-scattered gamma-rays. Indeed, the anti-correlation between the soft X-ray flux and the hard tail implies that it is Comptonisation of the accretion flow that is responsible for the hard tail. The fact that the hard tails seen by SIGMA are at relatively low flux levels (10^{-5} ph cm^{-2} s^{-1} keV^{-1}) is a partial explanation as to why they have only recently been detected.

Cyclotron lines were first detected at gamma-ray energies from a NS system, namely Her X-1 [10]. The feature, which has since been confirmed by GINGA observations [11], was seen at around 40 keV suggesting (as in Section 1.2.3) the presence of a $\approx 10^8$ T magnetic field, which in turn suggests a considerable amount of accretion has already taken place. There is growing evidence, supported by GINGA observations, that a number of NS binary systems have cyclotron line features above 15 keV and hence super-strong magnetic fields.

The first compact objects detected at gamma-ray energies were thought to be isolated neutron stars and were identified at other wavelengths with the Crab and Vela radio pulsars by SAS-2. COS-B observations [12] showed that for both the Crab and Vela pulsars, the gamma-ray profiles were highly variable and in particular for Vela, the radio and gamma-ray pulses did not coincide. The latter discovery suggested that gamma-ray emission originates in a different region of the NS magnetosphere than the radio synchrotron. Consequently, the so called 'outer-gap' models have been developed [13] which essentially rely on the formation of electron/positron pair cascades through photon-photon collisions between soft X-ray photons from the NS surface and synchrotron gamma-rays from ultra-relativistic electrons caught in the NS magnetosphere. Once the pair cascade has been initiated, perhaps by chance accretion from the ISM, the cascade is self-sustaining through the conversion of the NS spin-down power into the re-creation and acceleration of pairs. Re-combination of pairs close to the NS surface may lead to 511 keV line emission, and indeed features around this energy have been reported previously [14, 15] as can be seen in Figure 1.1.

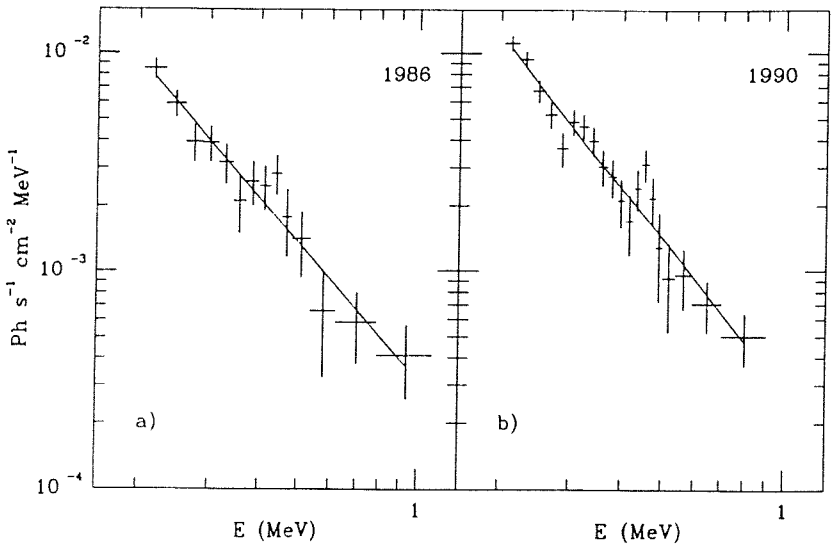


Figure 1.1: Photon spectra obtained for the Crab pulsar with the FIGARO balloon-borne instrument, indicating a line feature at ~440 keV.

The steady conversion of angular momentum into gamma-ray synchrotron radiation may explain the apparent evolution of Crab-like pulsars to Vela-like and then Geminga-like pulsars until ultimately the pair cascade process is quenched and gamma-ray emission is no longer possible.

It is essential that as many NS systems as possible are characterised in the 15 – 300 keV band in order to place reliable constraints on any distinguishing spectral features. For the detection and characterisation of a large sample of NS systems, the next generation of imaging gamma-ray telescopes will require a continuum sensitivity that is at least an order of magnitude greater than that of SIGMA, and with at least the same angular resolution. Similarly, observations of a large sample of gamma-ray pulsar spectra with higher sensitivity above a few MeV is crucial for the validation of outer-gap models and pulsar evolution theories since certain spectral breaks are predicted in this energy range. Ideally, spectral coverage from keV energies, where thermal emission dominates, to MeV energies, where relativistic mechanisms become significant is required to accurately pin-point the spectral breaks and their relationship with the pulsar rotation period.

Black Hole Candidates

As with NS systems, black hole (BH) systems are initially constrained through optical spectroscopic measurements of the radial velocity of the secondary star. However, modelling of the accretion flow into the BH has suggested that the emission processes will produce spectral features that can be directly linked to BH accretion. Early SIGMA observations of the BH candidate Nova Muscae [16] revealed a hard tail power-law component extending up to 300 keV and a transient emission feature around 500 keV, both of which supported HEAO-3 observations of another BH candidate, CYG X-1 [17]. For CYG X-1, in addition to the hard tail (which extended up to MeV energies) and annihilation feature, a transient 'MeV-bump' was also observed. This 'bump' may be explained in terms of an inner/outer accretion disc geometry. The outer disc is an optically thin, ion dominated region at temperatures of less than 10^9 K, radiating by bremsstrahlung below 400 keV. Conversely, the inner disc is an optically thick, often pair dominated region at temperatures of 5×10^9 K, responsible for the MeV-bump emission through Comptonisation of outer disc hard X-rays.

In the low-γ state, X-ray bremsstrahlung from the outer disc dominates the emission and

only the power-law emission is seen. Catastrophic changes in the accretion flow both super-heats and enlarges the inner accretion disc. At this point, Comptonisation of hard X-ray photons becomes significant although not efficient enough to quench the inner disc. Collisions between the seed and Comptonised X-ray photons results in the development of the inner disc pair plasma. Comptonised X-rays now dominate the emission and result in the MeV bump, whilst the expansion of the inner disc reduces the X-ray emission region, thus supporting the anti-correlation between X and gamma-ray luminosities. Radiation pressure from the inner disc has been shown both theoretically [18] and by Monte Carlo modelling [19] to be sufficient to promote the escape of positrons from the pair plasma. The positrons would rapidly slow down and annihilate in any cold and sufficiently dense medium (e.g. the secondary photosphere or stellar wind, the accretion stream, or the local ISM), thus producing a transient annihilation line correlated with the MeV-bump feature. Backscattered annihilation features, if seen, may help to place constraints on the accretion disc geometry [20].

As can be seen in Fig. 1.2, SIGMA spectra of Nova Muscae strongly resemble those of other BH candidates (for the prototype A0620-00 see [21]) such as CYG X-1 and 1E1740.7-2942, for which episodic excesses at around 500 keV have also been detected [22, 23, 24].

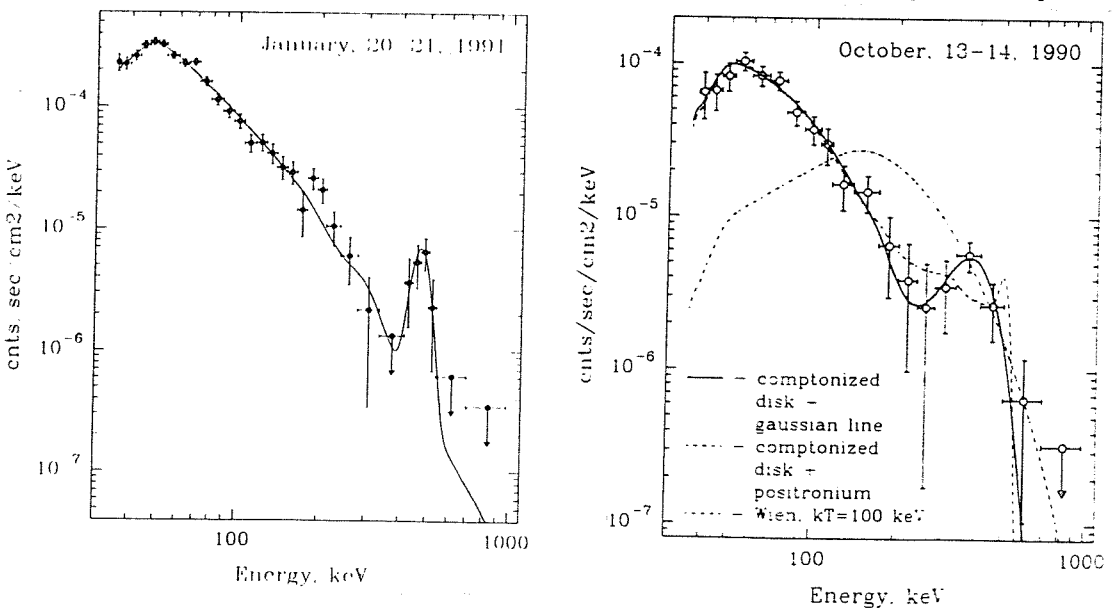


Figure 1.2: Comparison of Nova Muscae and 1E1740.7-2942 spectra.

Some of the more important issues that need to be addressed by the next generation of gamma-ray telescopes are : (1) establishing the correlation between narrow annihilation features and MeV emission for a larger sample of BH candidates; (2) utilising both spectral (line widths, peak energy shift) and temporal (correlation with other wave-bands) characteristics; and (3) long-term observations of the gamma-ray variability in order to

place further constraints on the cause of such long time-scale variability (months to years).

Quasi-periodic oscillations (QPOs), may provide an additional degree of discrimination between NS and BH systems when used in conjunction with the annihilation line and MeV-bump prerequisite. QPOs seen in quiescent NS systems all fall within the narrow 5 – 8 Hz range [25] and are thought to arise from a 'feedback' process between the NS magnetosphere and the in-falling material [26]. Since the accretion disc is the Comptonising medium for the gamma-ray hard tail, changes in the optical thickness of the disc by QPOs are imprinted on the gamma-ray flux.

After the discovery of QPOs in CYG X-1 [27] at a frequency of 0.04 Hz over the 1 – 20 keV (GINGA/EXOSAT) *and* 20 – 320 keV (SIGMA/BATSE) energy bands the search for QPOs in BH candidates was on. QPOs have now been discovered in a total of five BH candidates (CYG X-1, LMC X-1, GX339-4, Nova Muscae and Nova Persei), all at low frequencies. However, the fact that QPOs were present at all in BH systems demanded the revision of the standard QPO model with respect to magnetosphere-accretion flow interactions. The most promising of recent theories involves the superposition of a sequence of interacting flares in the disc, originating as accretion flow instabilities and indeed these models give rise to QPOs at around 0.04 Hz. A deep spectral/timing survey of both NS and BH systems is thus essential to validate the QPO frequency discriminator.

1.3.2 Nucleosynthesis in the Galaxy

Observations of gamma-ray lines from nuclear transitions of decaying radio-nuclei and the annihilation of positrons in sites of nucleosynthesis provides a direct means of validating models for both hydrostatic and explosive processes. Some of the more important decay chains for nucleosynthesis products were given in Table 1.1.

Supernovae

The occurrence of the type II supernova (core bounce explosion of a collapsing, ≥ 20 $M_{\odot}$ star) SN1987–A in the Large Magellanic Cloud (LMC), presented a once in a lifetime opportunity to study explosive nucleosynthesis. The detection of gamma-ray lines from the decay of ^{56}Co (847 and 1238 keV) in this event [28, 29, 30], provided for the first time the possibility of validating models of supernovae (SN) nucleosynthesis by studying line

profiles, energy shifts, fluxes and temporal evolution.

Even early observations of SN1987–A placed constraints on the underlying dynamics of the system. For example, the large ^{56}Co line widths observed [30] implied high expansion velocities which in turn suggested a large amount of turbulent mixing must have taken place during the explosion, dredging up the lower ejecta layers. This mixing of the inner nucleosynthesis products with the outer H/He envelope may explain the rapid turn-on of the X-ray flux from SN1987-A as observed by GINGA [31]. In addition, the actual line profiles were consistent with a transparent envelope of ejecta, in direct contradiction with the optical light-curve which relied on a significant portion of the lower ejecta gamma-ray flux being reprocessed in the outer layers. The apparent paradox was solved by dropping the then standard homologous model for SN ejecta – the ejecta must, in some cases, become clumpy, consisting of optically thick filaments and transparent regions in between. Current models for a type II event with a 15 – 20 $M_\odot$ progenitor, predict that possibly 10^{-4} $M_\odot$ of ^{44}Ti will be synthesised and ejected. Thus, with a half-life of 78 years, the 68 and 78 keV lines from ^{44}Ti in SN1987–A should still be detected by the next generation of gamma-ray telescopes.

In general, the ^{44}Ti lines are not likely to be as intense (by a factor of 10^{-3}) in type Ib and II SN (core bounce explosion of either a Wolf Rayet star or a normal massive star) as in type Ia SN (WD thermonuclear explosion) as a result of the massive amounts of ejecta in the former. However, with a predicted Galactic recurrence of ~50 years, the ^{44}Ti fluxes will be strongly indicative of the most recent sites of explosive nucleosythesis and of the occurrence rate for each type of SN.

The study of gamma-ray lines from type Ia SN has been shown to provide a direct means of measuring nucleosynthesis yields and ejecta dynamics [32], which with a larger sample of events, may help to constrain type Ia SN to either WD detonation [33] or WD deflagration [34] models. For example, branching ratios are sensitive to the mass and velocity distribution of the ejecta, as well as the degree of turbulent mixing that has taken place in the explosion. Line widths are also sensitive to the degree of mixing since a broad line suggests emission is from a broad range of ejection velocities. Line shapes contain information on the overall ejecta dynamics since at early times when the optical depth is greater than unity, only emission from the approaching edge of the SN is visible and all lines are blue-shifted. As the remnant expands and becomes more transparent, the line peak moves towards the centroid energy, and in addition, a more intense red-shifted

component is seen due to the look-back time for the receding edge of the remnant.

Novae

Classical novae outbursts have been modelled as the thermonuclear runaway of the accreted, H-rich envelope of a WD, in which nucleosynthesis occurs and the envelope is ejected from the WD by the radiation pressure of decay radiation and annihilation photons [35]. The predicted C→N→O chain reaction should synthesize elements up to and including ^{22}Na [36] which has a half-life of 3.75 years. If produced in sufficient quantities, ^{22}Na may be detected (1275 keV and annihilation line) as the accumulated debris of the many (tens per year) novae thought to occur throughout the Galaxy, although so far extended 1275 keV emission associated with the Galactic Bulge has not been detected.

The Remains of Nucleosynthesis

One of the fundamental questions in astronomy is the source(s) of heavy element production from the basic H/He primordial building blocks. It is widely believed that the main sources are massive main sequence stars, supernovae and novae. For a long term picture of the spatial distribution of nucleosynthesis sites in the Galaxy, ^{26}Al was recognised early on as a promising tracer of activity [37]. With a half-life of $\sim 10^6$ years, the ^{26}Al distribution was predicted to be an accumulation of ejecta from thousands of supernovae and millions of novae, and therefore expected to be a smooth extended structure along the Galactic plane with hot-spots from recent activity. In addition, the intrinsically narrow width of the 1809 keV ^{26}Al line makes for a powerful velocity diagnostic and indeed the line width ($\leq$3 keV) of ^{26}Al observed by HEAO-3 [38] agreed with predictions that emission is from within the ISM. Thus, any shifts in the centroid energy should reflect either random motions in the ISM or the Galactic rotation itself.

The first imaging observations of the 1809 keV line by COMPTEL [39] sought to place restrictions on the relative contributions from the possible sources by comparison with their distributions at other wave-bands. It is predicted that if a substantial amount of ^{26}Al is ejected into the ISM by each individual event, a 'clumpy' distribution should be observed. Thus, the relatively 'clumpy' distribution observed by COMPTEL initially ruled out smooth source contributions from high frequency events such as classical novae,

and in also restricted the total emission from ^{26}Al to only 1 – 2 $M_\odot$ of material [39]. Recent interpretations of the COMPTEL data have concluded that it is difficult to achieve the large scale 'clumpiness' of the observed distribution, even if W-R stars and supernovae were the sole contributors to the emission (novae emission is below the sensitivity limit of COMPTEL) [40]. However, there are several distinctive features in the COMPTEL images that have led to the latest interpretation : (1) the ^{26}Al is distributed along the Galactic plane *but* has a global antisymmetry about the Galactic Centre ; (2) several large-scale (a few degrees) hot-spots are seen near the Galactic Centre and at galactic longitudes of 32°, 263° (Vela), 280° (Carina), 310° and 345°, and also in the Cygnus region ; (3) the most prominent hot-spot is offset (south-east) from the Galactic Centre by (2°,-2°). Consequently, it has recently been proposed [41] that the large-scale hot-spots are in fact the geometrical result of looking *down* the spiral arms of the Galaxy.

In addition to ^{26}Al production, current stellar nucleosynthesis models predict that a few 10^{-5} $M_\odot$ of ^{60}Fe will also be produced in massive (15 – 40 $M_\odot$) stars by neutron capture in the core and shell He burning phase and find its way into the ISM [42]. A correlation between the distribution of ^{60}Fe (1.133 and 1.322 MeV lines) and ^{26}Al emission, would suggest that massive stars are the main source of ^{26}Al. The predicted 5×10^{-5} ph cm^{-2} s^{-1} fluxes from 0.5 $M_\odot$ of ^{60}Fe in the ISM, should be detectable by the next generation of gamma-ray telescopes.

For future imaging surveys of the ^{26}Al and ^{60}Fe lines it is essential to have a large-scale imaging capability coupled with a large field of view (FOV). In addition, a fine angular resolution may also be required to localise single event hot-spots. Spectral resolution of less than 1% would ensure the velocity diagnostic potential of ^{26}Al observations was realised.

1.3.3 Interstellar Processes

COMPTEL has detected intense gamma-ray emission in the 3 – 7 MeV band from the Orion complex which has been identified with the broad de-excitation lines of ^{12}C and ^{16}O at 4.44 and 6.13 MeV respectively [43]. Nuclear excitation of the ambient ISM can occur either by the bombardment of a heavy nuclei (C and O) gas by low energy cosmic-ray protons and α-particles, *or* by the inverse process whereby H and He nuclei are bombarded by heavier C and O cosmic-rays.

In principle the observed line shape provides a powerful diagnostic for discerning which is the dominant interaction process, since broad de-excitation lines would suggest the cosmic-rays are heavier nuclei. However, the first process can be ruled out since by assuming solar abundances of H and He, the required energy density for low energy cosmic-rays to produce the observed level of nuclear excitation would be three orders of magnitude larger than the energy density required to fully ionize the ISM – this is not observed at other wavelengths.

Thus, from the line fluxes and shapes, constraints may be placed on the origin of the heavy nuclei cosmic-rays. Modelling of massive (up to 60 $M_{\odot}$) stars suggests that their supernovae (type II) ejecta is dominated by ^{16}O and ^{12}C [44]. In the case of the Orion complex, it has been proposed that the C and O cosmic-rays result from the acceleration (up to several MeV per nucleon) of C and O ejecta by a combined stellar wind/multiple supernovae, shock front [45, 46]. Thus, restrictions can again be placed on the explosive nuleosynthesis models for type II supernovae events.

Perhaps the most interesting point from an imaging perspective is that different cosmic-ray species are expected to interact on different size-scales. For nuclei propagating into the dense ISM, the scale size of the line emission region will be determined by the stopping length due to Coulomb interactions. Heavier nuclei will be stopped in a shorter distance, whereas lighter nuclei will diffuse a greater distance, e.g. line emission from ^{56}Fe should be restricted to 2 – 3 times smaller angular scales than for ^{16}O.

1.3.4 The Galactic Centre

From radio through to infrared observations, the Galactic Centre (GC) is known to be the site of violent activity and indeed, gamma-ray observations of this region so far suggest the physical processes are no less extreme in the high energy domain. The earliest gamma-ray observations of the GC with balloon-borne spectrometers [47] revealed complex spectral features that required at least three components to fit the profile over the 70 – 700 keV band, all of which were extremely variable : (1) an underlying power-law continuum ; (2) an emission line near 511 keV ; (3) an additional continuum below 511 keV. With the advent of gamma-ray telescopes with a true imaging capability above 30 keV, the complexity of early GC spectra was attributed to a combination of the relatively high density of point-like sources in the region and extended emission covering several degrees.

Compact Emission from the Galactic Centre

The first imaging survey of the GC by the Space-Lab 2 coded aperture instruments revealed at least 9 point-like sources of emission and extended structure on scales of a degree in the 2.5 – 30 keV energy band [48]. At a few keV, emission was detected from Sgr A*, the strong radio source associated with the Galactic Centre and the putative 10^6 $M_\odot$ dormant black hole. With a series of ten observations taken over the 35 – 300 keV band, examples of which are shown in Fig. 1.3, SIGMA has gone a long way to simplifying the Galactic Centre scenario. In addition to transient sources occasionally detected in the survey, SIGMA has found at least eight persistent gamma-ray sources above 35 keV, 3 of which are located within 1^o of the radio source Sgr A* [49].

Four of the sources have been identified with different classes of neutron star binary, and the two hardest sources, 1E1740.7-2942 and GRS1758-258, are considered as black hole candidates by comparing their spectra with that of CYG X-1 (see Section 1.3.1). Thus the underlying power-law continuum seen by the earliest balloon-borne experiments and by HEAO-3 [50], was a superposition of several individual power-law spectra.

The remaining continuum and 500 keV line feature identified by Leventhal [47], certainly as far as compact emission is concerned, can be attributed to the pair plasma regions within the accretion discs of 1E1740 and possibly GRS1758. The transportation of positrons out of the accretion disc and into an annihilation region has already been covered in Section 1.3.1 [23]. However, the interesting point with 1E1740 is that it appears to have an additional mechanism for transporting positrons away from the accretion disc, namely diametrically opposed jets that have been imaged through their radio synchrotron emission [51]. The annihilation of positrons in the viscinity of the binary system (within one pc) via either the para-positronium state (25 % $\rightarrow$ 511 keV line) or the ortho-positronium state (75 % $\rightarrow$ continuum below 511 keV), can account for the remaining spectral features seen.

Extended Emission from the Galactic Centre

In addition to the compact annihilation line and continuum, there is also a diffuse component distributed along the Galactic Plane, associated with the ISM, and with a longitude extent of $\pm 40^o$. OSSE observations [52] suggest that the extended positronium

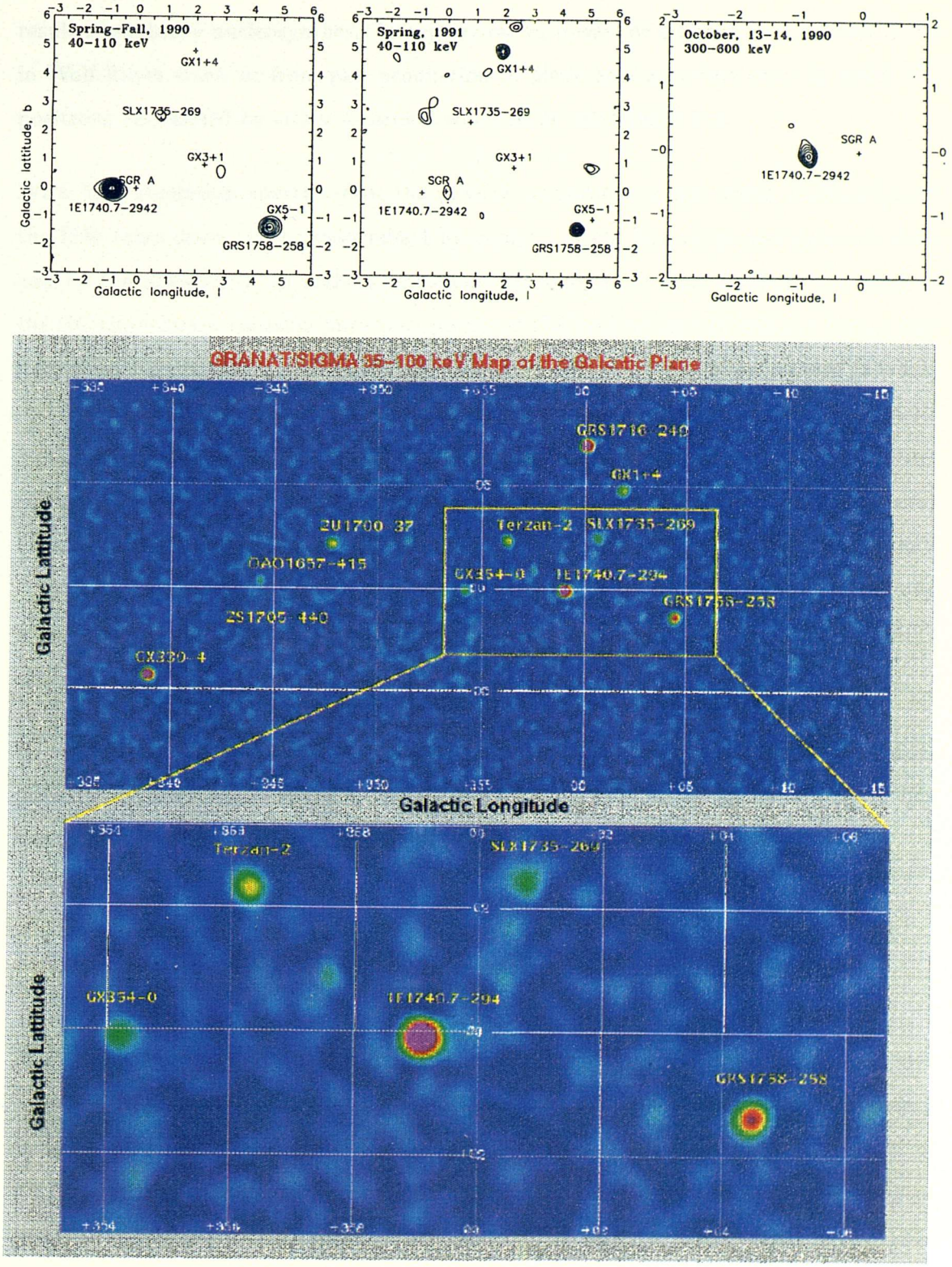


Figure 1.3: Images of the Galactic Centre region obtained with SIGMA.

emission has two components, one associated with the Galactic bulge and one with the disc. The origin of the positrons however, is still an open question. They may come from the spallation products of cosmic-ray/ISM interactions, from radioactive nuclei as a

result of explosive nucleosynthesis in supernovae or novae, or hydrostatic nucleosynthesis in Wolf-Rayet stars, or from pair production in black hole accretion discs in which the positrons are carried by either a coronal wind, or by relativistic jets.

With high resolution spectroscopy, the physical conditions under which annihilation in the ISM takes place can be determined by analysis of the line shape and centroid shift, and by the presence of an ortho-positronium continuum. Several diagnostic features of the GC annihilation emission have been predicted [53] such as the 511 keV line from a cold (100 K) H-I gas which should consist of two components : (1) a narrow feature (FWHM $\sim$ 1.5 keV) resulting from direct annihilation with atomic electrons and (2), a broader feature (6.5 keV) due to the formation of positronium and subsequent para-annihilation whilst still in motion. In addition, suppression of the ortho-positronium continuum may be used to infer the radiation environment in which annihilation occurs. Over the 10 keV - 10 MeV energy range, the observed emission is thought to be either bremsstrahlung from cosmic-ray electrons passing through the ISM, or a result of the Sunyaev-Zeldovich effect [54].

In conclusion, it is apparent that detailed mapping of the entire Galactic Bulge region, on both fine and coarse angular scales, and for both continuum and annihilation line energies, is essential to disentangle point and diffuse components.

1.4 Gamma-Ray Bursts

The observed properties of the enigmatic Gamma-Ray Bursts (GRBs) cover at least four orders of magnitude in duration (10^{-2}s - 10^{2}s), flux (10^{-7} - 10^{-3} erg cm^{-2}) and energy (4 keV - 200 MeV), and even after more than 20 years of observations, the origin of GRBs is still unknown. As such, GRBs are obvious targets for the next generation of gamma-ray telescopes. An important parameter for testing the validity of the many GRB models that have been proposed over the years is the measure of spatial homogeneity. An ideal measure of homogeneity, ideal in that only one observable is required, is the peak flux distribution. Specifically, it can be shown that the spatial homogeneity can be parameterised by the mean value of Eq. 1.8 :

$$V/V_{max} = \left(\frac{C_p}{C_{lim}}\right)^{-3/2} , \qquad (1.8)$$

where V and V_{max} are the volumes in space enclosed by the radii between the observer and the GRB, and the observer and a GRB at a distance such that it is just above the triggering threshold, C_{lim}. A homogeneous distribution produces a $< V/V_{max} > = 0.5$. On performing this type of analysis on a large sample of GRBs, for example, on the latest catalogue (3B) of BATSE GRB observations which contains over 1100 GRBs, the inhomogeneity of the distribution of GRBs is apparent immediately [55]. Current values of $< V/V_{max} >$ continue to support the spatial inhomogeneity hypothesis, for example, Fig. 1.4 shows a plot of log N – log P, where N is the number of GRBs with a peak luminosity P, which highlights the deficit of low intensity GRBs [56].

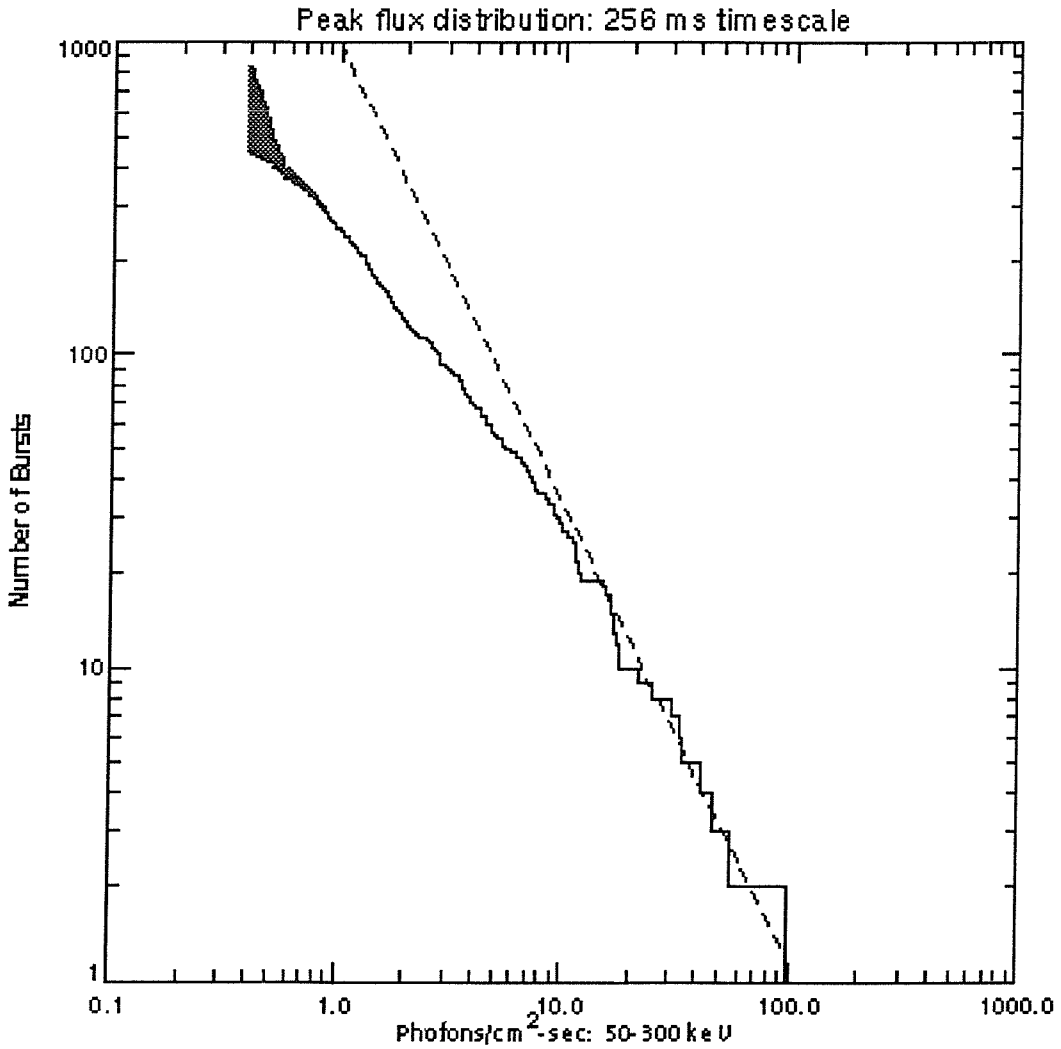


Figure 1.4: The continuing deficit of low intensity GRBs provides further support for the spatial inhomogeneity hypothesis.

The inhomogeneity was expected under the Galactic Plane hypothesis of GRBs, on the contrary however, the distribution of GRBs across the sky was quite unexpected. As can be seen in Fig. 1.5, within statistical fluctuations GRBs are isotropically distributed on

the sky [55, 57] which immediately rules out all Galactic Plane hypothesis (e.g. merging neutron star theories). The most plausible models in the circumstances involve either a Galactic Corona origin or an extragalactic origin. With over 1000 GRBs observed by BATSE, the distribution remains isotropic [58]. The observed distribution, combined with the lack of GRB enhancement in the direction of M31, places tight constraints on the Galactic Corona model [59]. In addition, GRB duration studies have shown that faint bursts are longer than bright bursts, possibly as a consequence of cosmological time dilation [60].

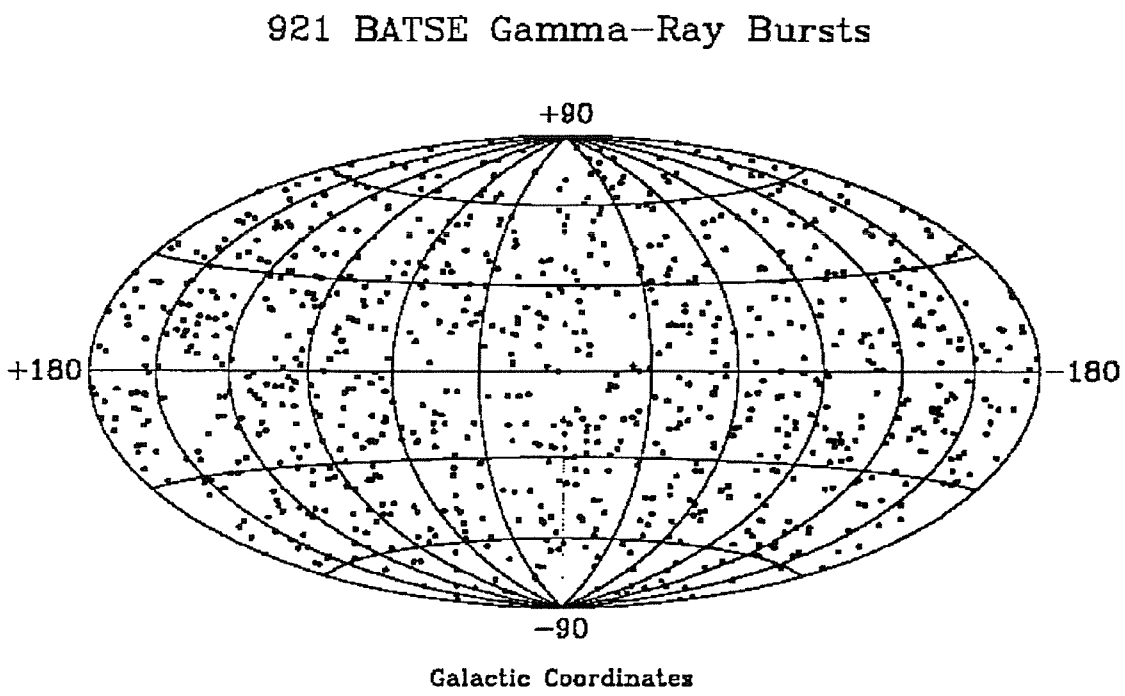


Figure 1.5: The spatial distribution of the first 921 GRBs detected by BATSE.

The most recent GRB hypothesis is based on the final transition of a spun-down neutron star from a neutron degenerate state to a strange quark state, which occurs when rotational forces can no longer balance gravitational forces [61]. During the phase transition, the radius of the neutron star reduces by ≈ 20 % releasing around 10^{52} ergs of energy. This model seems to be able to explain some of the more important features of GRB observations : (1) the energy release can account for bursts at cosmological distances ; (2) the phase transition and hence energy release can occur on time-scales from 0.05 s to a few minutes ; (3) the estimated burst frequency of 10^{-6} year^{-1} per galaxy is in very good agreement with current observations ; (4) the phase transition can occur only once, explaining the lack of recurrence of GRBs ; (5) the fact that the surface of the neutron star may be blown away by the super-Eddington radiation supports the power-law spectral

models which are attenuated by a Compton scattering medium.

Searches for counterparts at other wavelengths have to date proved unsuccessful, and the promise of future gamma-ray missions with high resolution optical transient cameras has long been anticipated. In addition, with the expected increases in FOV and sensitivity, future high resolution gamma-ray telescopes may expect to see several GRB per year. Continual coverage by a single instrument over the 15 keV to 10 MeV range may finally resolve the controversy surrounding various spectral components seen in current burst data. Deep survey observations with up to 10 days on a single pointing will provide excellent pre and post-GRB coverage, which is essential for the analysis of any GRB after-effects i.e. decay lines if explosive nucleosynthesis is involved.

1.5 The Extragalactic Gamma-ray Sky

1.5.1 Active Galaxies

Why are gamma-rays good for the study of AGN ? Firstly, their penetrating nature provides a direct means of probing the innermost regions of AGN. Secondly, current X and gamma-ray data imply that a substantial fraction of the total luminosity of AGN is emitted in the X and gamma-ray band, although as will be seen the gamma-ray characteristics differ dramatically between each AGN type. Thirdly, the extremely high luminosity and short variability time-scales imply that the X and gamma-ray emission region is very compact and thus high photon densities and pair production is expected. Finally, the intensity and spectral profile of Cosmic Diffuse Background (CDB) observations suggests that if AGN are a significant component of the CDB, their spectra cannot extend into gamma-ray energies with the same power-law slope observed in the 2 – 100 keV band.

Active Galactic Nuclei can be classified initially into either *radio-loud*, or the more common (between 90 and 97% of observed AGN) *radio-quiet* objects [62]. AGN belonging to the former class are generally divided further into compact (core dominated) sources labelled quasars, and extended (lobe dominated) sources associated with elliptical galaxies. Quasars can be further classified with increasing luminosity : BL Lacs exhibit a single, smooth synchrotron component from millimetre to X-ray wavelengths ; OVVs (optically violently variable quasars) exhibit rapidly variable, polarized optical continua,

very broad and strong optical emission lines and harder spectra above X-ray energies than BL Lacs ; blazars, initially introduced as a re-classification of BL Lacs and OVVs, now appear to be phenomenologically separate from BL Lacs and OVVs. The powerful lobe-dominated radio galaxies are designated either FR1 or FR2 sources (Fanaroff and Riley [63]), again with respect to increasing luminosity. FR1 sources have prominent two-sided jets and most of the radio brightness is centrally concentrated. On the contrary, in FR2 sources the most intense emission comes from hot-spots located in the radio-lobe extremities and in addition, the jets are often very faint and one-sided. From a unification perspective, the simple view is that radio-loud AGN possess either mildly relativistic jets, which account for FR1 galaxies and BL Lacs when coupled with an orientation factor, *or* ultra-relativistic jets, which account for FR2 galaxies and OVV quasars, again when coupled with an orientation factor [64]. For the observer aligned cases of both FR1s and FR2s i.e. BL Lacs and OVVs/blazars, the jet emission overwhelms the roughly isotropic continuum emission from the active nucleus.

Radio-quiet AGN are designated as either Seyfert galaxies or radio-quiet quasars (originally Quasi-Stellar Objects (QSOs)) and are all associated with spiral galaxies. The further classification of Seyferts depending on the observation of broad and narrow line emission (Seyfert 1), or just narrow line emission (Seyfert 2) in the optical, also relies on an orientation effect for unification. Coupled with differences in the total luminosity of the AGN, it is now almost possible to unify Seyferts and QSOs [65].

Thus, by following the unification philosophy, all AGN can be 'constructed' from a fundamental set of physical components. However, by considering the spectral features seen in both radio-quiet and radio-loud AGN data and the possible origins of these features, the limitations on our current understanding becomes apparent, as does the necessity for more sensitive observations of AGN over a broader spectral range.

Seyferts and QSOs

Seyfert and QSO spectra are dominated by a roughly isotropic, underlying power-law continuum. Analysis of spectra from many Seyfert 1 galaxies reveal a global power-law from 2 to 20 keV, with a mean spectral index of -1.9. At the centre of all AGN is believed to be a $10^6 - 10^8$ $M_\odot$ black hole into which material (with angular momentum) is accreting via a geometrically thin, optically thick disc. The accretion disc is thought to extend out

to 1 AU from the black hole and is thought to be the origin of the 'big blue bump' (BBB) seen in the UV via thermal emission of a 3×10^4 K optically thick plasma. The soft X-ray excess observed at keV energies and seen on top of the underlying power-law, is believed to be the tail of the 'big blue bump' [65].

It is believed that most of the gravitational energy of the accretion flow is released in the coronal surface of the accretion disc forming a cloud of hot electrons and ion species. It is likely that this surface corona powers an out-flowing accretion disc wind. Indeed, below 10 keV, various absorption features are used to infer the presence of the coronal wind. The presence of the disc corona also provides the ideal medium for producing the power-law continuum observed in Seyfert 1 AGN. There are currently two competing models for the power-law emission, both of which involve Comptonisation of inner accretion disc UV photons by a hot disc corona [66] and differ mainly in the assumed velocity distribution of the coronal electrons – one is thermalised, whereas the other is relativistic. The Compton scattering of UV disc photons is isotropic in both cases and so approximately half of the emission is reprocessed by the disc to produce the frequently observed spectral hardening above 10 keV [67] (the Compton reflector) and an Fe Kα fluorescence line from the inner disc. The Compton reflection component produces a broad hump between 10 and 100 keV peaking around 30 keV. Photon-photon scattering in the corona results in pair production and/or pair-cascades, after which annihilation may occur in varying degrees, but is not expected to be seen as an annihilation line in thermal models because of extreme broadening.

Gamma-ray observations may provide for the first time unification of the underlying physical processes in both types of Seyfert. However, the lack of broad-band gamma-ray coverage of a large sample of Seyferts has meant that there has been no convincing evidence to suggest that the optical differences between Seyferts 1 and 2 are simply an orientation effect. For example, SIGMA observations of NGC 4151 [68] (now thought to be uncharacteristic of Seyfert 1s) claim to show evidence of a spectral break at 50 keV possibly supported by early non-thermal corona-disc models [69] which predicted spectral breaks above ~30 keV (in addition to breaks above a few MeV and an annihilation line feature). These observations were in direct contradiction with MISO data [70] which showed a hard power-law recovery extending to several MeV and to some extent with MIFRASO observations [71] which failed to show a spectral break below 200 keV. OSSE observations of NGC 4151 apparently show a soft spectral state leading to the hypothesis that Seyferts in general have hard and soft states above 30 keV [72]. Recent joint observations between

ROSAT and OSSE [73] of the Seyfert 1 IC 4329A can be fitted by *both* thermal and modified non-thermal corona-disc models [66] and show weak evidence for a spectral break between ~250 and ~1700 keV. Further confusion is added to the situation by the lack of COMPTEL or EGRET detections of Seyferts above MeV energies [74].

FR1/FR2 Radio Galaxies and BL Lacs/OVV Quasars

The dominant emission mechanism from radio to gamma-ray energies in all radio-loud AGN is thought to be non-thermal in origin, either directly through the beamed emission from observer aligned jets or indirectly through the deceleration of jets in the inter-cluster medium. One possible mechanism for creating electromagnetic jets postulates that a large-scale poloidal magnetic field threads the accretion disc and the black hole. The disc, being a good conductor, will convect and wind up the magnetic field lines, driving a hydromagnetic wind which is collimated by the toroidal magnetic field [75].

The relativistic outflow from the accretion disc is thought to be sustained by the efficient transfer of spin-energy from the black hole to the magnetic flux [76]. Such models provide a qualitative explanation for the differences between FR1 and FR2 sources : in sources with either slowly spinning or weakly magnetised black holes, the relativistic core will be weak and the resulting jet will be slow moving - an FR1 source ; when a powerful relativistic core is produced, the resulting jet is relativistic - an FR2 source.

From a unification perspective, there is growing support both observational and theoretical that FR1 radio-galaxies are the parent population of BL Lac objects [77]. The underlying galaxies of the FR1s and the BL Lacs are qualitatively similar (giant ellipticals) as are the extended radio fluxes. Similarly, unification between FR2s and OVV Quasars has found support in recent years [78]. Indeed, although BL Lacs and OVV quasars have in the past been categorised into the same radio-loud AGN subclass, observational evidence suggests they are two distinct classes of object with different parent populations. For example, VLBI polarisation studies, indicate that the magnetic field lines associated with the jet flow from BL Lacs are generally tangled with strong components transverse to the jet. The transverse magnetic field components are associated with shock regions in the jet, which in turn are indicative of slower bulk motions than in OVV jets. For OVVs, the magnetic field lines are observed to lie along the jet direction, tending to support the idea that in the absence of shocks, the underlying magnetic field is aligned with the jet

and *relativistic* jet flow prevents the formation of shocks.

Observationally, radio galaxy spectra extend up to MeV energies with possible breaks in the power law at around 150 keV as observed in Cen-A spectra [79]. For Cen-A, correlations between gamma-ray and radio emission suggests the same mechanism, possibly synchrotron self Compton in the relativistic core region, is responsible for producing both extremes of the e-m spectrum. Blazar spectra appear to be composed of at least two components : a low energy component with luminosity peaking in the IR-UV range, and a high-energy component with luminosity strongly dominated in the gamma-ray regime.

There are currently two competing models which address the double component nature of blazar spectra, namely synchrotron self-Compton [80] jet models (SSC) and external Compton jet models [81, 82] (EC). However, it has recently been shown that SSC models are difficult to fit for most blazar spectra [83]. For the EC model to be efficient, the external radiation field must first interact with the jet well outside the radiation source in order to prevent absorption of inverse Compton gamma-rays via pair production. The logical source of soft-X/UV photons is the accretion disc, which with the addition of the coronal wind as a Thomson scattering medium, produces a diffuse source of seed photons well away from the accretion disc. The EC models have been elaborated upon to include pair cascades which reprocess the external radiation field further, and are capable of producing GeV gamma-rays [84]. Both SSC and EC models produce relativistically beamed and Doppler boosted emission in the direction of the jet, which thus leads to an elegant unification between radio galaxies and blazars.

Broad band gamma-ray observations of a large sample of radio-loud AGN will undoubtedly provide the opportunity to address some of the fundamental questions facing the understanding of AGN : studies of the putative beamed (blazars) and unbeamed (radio galaxies) populations will provide space densities and hence an estimate of the opening angle to the cone in which beaming occurs ; similarly, relativistic beaming should flatten a steep luminosity distribution of the parent population [77] ; all EGRET detections to date have been associated with compact radio sources, where in contrast, OSSE and COMPTEL detections have been of radio-quiet AGN [85] as well as compact radio sources, supporting the pre-CGRO view that radio-loudness is a direct consequence of the formation and collimation of relativistic jets ; future observations will identify the specific spectral characteristics that not only verify the presence of relativistic jets, but also place

constraints on the specific jet gamma-ray emission mechanism(s) ; continuous spectral coverage linking X-ray and hard gamma-ray regimes should constrain the differences between, for example, OVVs and radio-loud quasars – COMPTEL results suggest a break in the power law at 1 MeV ; variability studies on, for example Cen-A for which the 5σ detection limit should be as low as 20 seconds for the future gamma-ray telescopes, will place constraints on the mass of the putative black hole – the light travel time across the inner edge of an accretion disc for a 10^7 $M_\odot$ black hole is on a time-scale of minutes.

1.5.2 Galaxies and Galaxy Clusters

Gamma-ray emission from normal galaxies is expected, for example, via individual compact binary systems, as discussed in Section 1.3. To date, non-AGN gamma-ray detections have been restricted to a few objects which include SN1987-A, LMC X-1 and diffuse emission from the Large Magellanic Cloud (LMC). The increased sensitivity of future gamma-ray telescopes should be of immense benefit to the study of extragalactic sources in normal galaxies, especially when coupled with a high angular resolution capability and a large FOV. CYG X-1 type objects i.e. black hole candidates, in Local Group galaxies such as M31 will then be in range and when summed with similar objects in the remainder of Local Group galaxies, will provide the basis for comparative studies with our Galaxy.

Although clusters of galaxies have been extensively observed in the 1 – 10 keV energy band, there are in fact very few observations above 20 keV [86]. The detection of hard emission may yield new astrophysical information on the intra-cluster environment. Specifically, observations above 30 keV are essential to distinguish between true intra-cluster medium (ICM) emission via the Sunyaev-Zeldovich effect [87] and unresolved AGN emission [88].

1.5.3 The Diffuse Gamma-ray Background

A catalogue of detailed gamma-ray AGN spectra is essential for understanding the true nature of the Cosmic Diffuse Background (CDB). In the last decade or so, the CDB has been studied extensively in the few to 100 keV range [89] and has been attributed to several mechanisms including bremsstrahlung models below 60 keV. Recently however, the few

to 500 keV CDB has been explained entirely in terms of unresolved AGN [90, 91, 92], although the situation above a few hundred keV is much less clear due to the lack of detailed spectral information.

Models for the 500 keV to 10 MeV CDB have included inverse Compton scattering of 2.7 K photons by cosmic ray electrons and bremsstrahlung from non-thermal electron interactions with intergalactic matter [93], although the most favoured hypothesis again has centred around unresolved AGN [94]. For unresolved AGN to produce the CDB spectrum, however, certain constraints must be placed on the mean AGN spectrum. In particular, a mean spectral break between 100 keV and a few MeV is essential and is in fact possible with current models of both radio-loud and radio-quiet AGN, as described in Section 1.5.1. Early CDB observations [94] apparently revealed a spectral 'bump' around MeV energies which is actually difficult to explain in terms of AGN spectra [95] and more recent studies both theoretical [96] and observational [97] appear to convincingly disprove the existence of the 'MeV bump'.

Complete models, however, for the 15 keV to 10 MeV CDB remain tenuous. The study of a large sample of AGN by future gamma-ray telescopes will be able to define the average gamma-ray spectra for the various types of AGN, determine their luminosity functions and define their cosmological evolution for the 15 keV – 10 MeV range. From such observations it will then be possible to determine the relative contributions of each class of AGN to the CDB and whether AGN alone are sufficient to explain the observed background radiation.

1.6 Conclusions

This chapter has briefly outlined the types of gamma-ray source both Galactic and extragalactic, that have either been, or are likely to be, detected in the 15 keV to 10 MeV regime. Figure 1.6 summarizes the imaging capabilities required to achieve a few of the spectroscopic observations described in this chapter [98].

The underlying limitation which appears to pervade *all* observational facets of gamma-ray astronomy to date, is the genuine lack of continuous broad band spectroscopic coverage over this energy range. Although with the advent of the GRANAT and CGRO missions, the situation has improved immeasurably, it should come as no surprise however, that

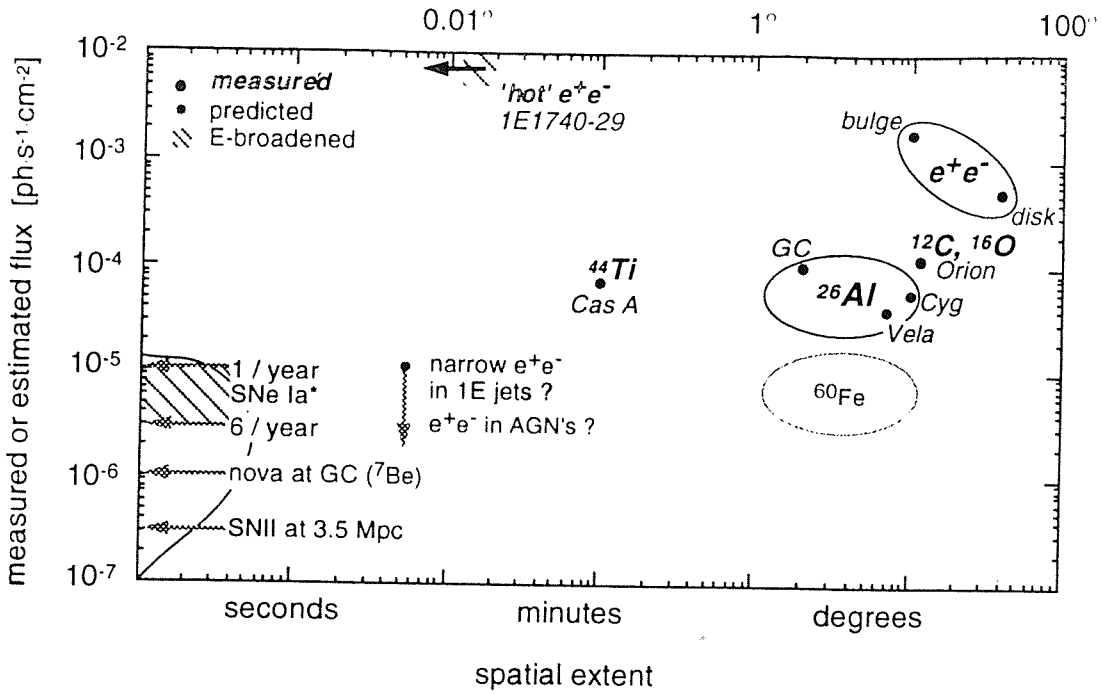


Figure 1.6: The sensitivity and imaging capabilities required to achieve a few of the spectroscopic goals outlined in this chapter.

the technological options for achieving high angular/spectral resolution, high sensitivity coverage from 15 keV to 10 MeV are far from being exhausted. The key demands placed upon future gamma-ray missions in order to achieve the above goals are now summarised :

- The energy coverage of any one instrument should be as wide as possible allowing spectral models to be tested directly without recourse to re-normalised data from other instruments. Similarly temporal variability may be compared to temporal behaviour at other energies to provide insight into production mechanisms and source characteristics. If the mission payload is comprised of more than one instrument. the overlap between instruments should be sufficient to allow continuous extended spectral coverage.

- As has been outlined in this chapter, the various γ- ray production models provide predictions of the characteristic spectral features for each class of astronomical source. To test these theoretical predictions extensively, the energy resolution of an instrument dedicated to spectroscopy must be at least an order of magnitude less than the most narrow width predicted. An instrument dedicated to imaging requires sufficient energy resolution to correlate high angular resolution imaging of a source with the high spectral resolution of the complimentary spectrometer.

- To increase the likely scientific yield of the mission and improve the chances of serendipitous source detection, it would be desirable to observe as large a FOV as possible, although at lower energies an increased FOV results in an increased background level from unresolved sources (see Section 1.5.3).

- Imaging studies from previous missions provide an upper limit to the angular resolution required for unambiguous identification of sources in highly populated regions. However, perhaps the most important reason for implementing a high resolving power is to enable optimisation of the available sensitivity.

- The instrument sensitivity essentially determines the scientific returns possible with the mission. Indeed, for satellite-borne missions improvements in sensitivity are invaluable as they essentially lengthen the lifetime of the mission.

The following chapter provides a review of the imaging concept that has revolutionised gamma-ray astronomy over the last two decades by virtue of its ability to produce high angular resolution images of the gamma-ray sky. The technique in question, namely coded aperture imaging, coupled with the high spectral resolution of the latest detector technologies, as will be seen, is capable of meeting the key demands outlined above in the form of the INTEGRAL imager.

Chapter 2

Coded Aperture Imaging in Gamma-ray Astronomy

Overview

The penetrating nature of gamma-ray photons which enables their use as a probe of the most compact and violent astrophysical phenomena known, also renders gamma-ray imaging a complex art from an instrumentation perspective. Above photon energies of 10 keV, traditional focusing methods for imaging the sky are only partially viable. This chapter reviews the basic concepts of what is generally regarded as the optimum method of high resolution imaging in the 15 keV to several MeV range, namely coded aperture imaging. This review includes the historical development of the coded mask and the techniques for 'reconstructing' the source distribution. The chapter concludes with a brief overview of the INTEGRAL mission paying particular attention to the imager, the coded aperture instrument around which the research in this thesis is based.

2.1 Introduction - Why Coded Aperture Imaging ?

Relative to other branches of astronomy, there are two fundamental problems with producing high angular resolution images of the gamma-ray sky : (1) reflecting/focusing becomes increasingly difficult and eventually impossible with increasing photon energy ;

(2) background counting rates at gamma-ray energies can be three orders of magnitude larger than the source flux. There are many options for imaging in the soft X-ray to hard gamma-ray regimes, however, since the research in this thesis is based specifically upon an imaging telescope, the following discussion concentrates purely on the optimum complementary *imaging* techniques either side of the 15 keV – 10 MeV range. In considering these techniques 'optimum' refers to the overall imaging performance with the of highlighting the fact that these techniques are not viable at higher/lower energies.

Grazing Incidence Telescopes

In recent years a general consensus appears to have been reached on the effectiveness of focusing techniques above 10 keV, in particular that high resolution imaging can only be accomplished by coded aperture techniques. Focusing of hard X-rays through grazing incidence reflection was until recently thought to be technically feasible only for energies below 10 keV [99]. Contrary to popular belief however, there now exists several focusing techniques that extend up to energies of 100 keV and consequently may be of importance in the design of the next generation of large X/gamma-ray observatories beyond the Advanced X-ray Astrophysics Facility (AXAF) and the X-ray Multi-mirror Mission (XMM).

Although the proposed AXAF telescope only reaches conventional X-ray focusing energies (10 keV), the 0.3 arcminute point spread function [100] does highlight the potential angular resolution advantage over coded aperture telescopes in the 10 – 100 keV range. In addition, since the area required to detect the focused photons is considerably smaller (four orders of magnitude) than that for a coded aperture system with an equivalent bulk area, the gain in sensitivity due to the reduced detector background is about an order of magnitude [101]. Figure 2.1 shows schematically some of the optical arrangements used in current grazing incidence systems.

However, as promising as focusing techniques may sound, there are fundamental limitations : (1) the field of view (FOV) is restricted severely by the grazing incidence angle to about 1^o ; (2) to date, no instruments of significant size have been built and tested up to 100 keV energies ; (3) there remains the physical 100 keV limit beyond which grazing incidence reflection becomes impossible ; (4) they suffer from unfavourable aperture/mass ratios.

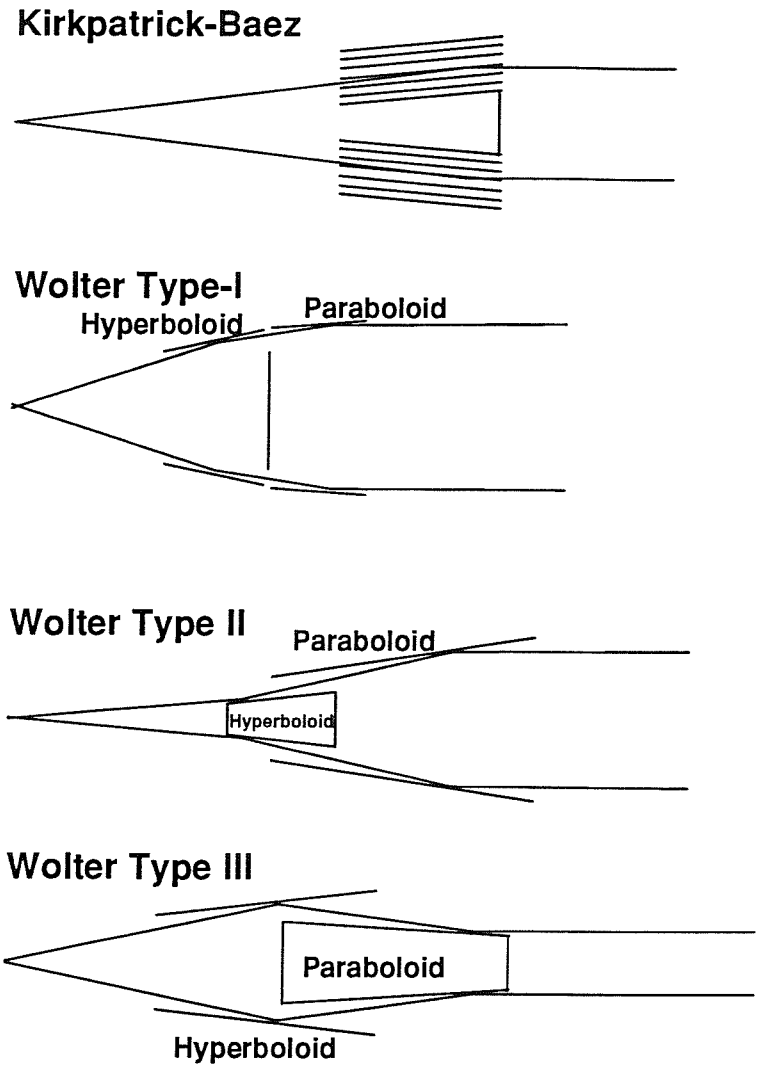


Figure 2.1: Optical arrangements for grazing incidence telescopes.

2.1.1 Spark Chambers

In parallel with the development of grazing incidence optics where the primary goal was to increase the high energy threshold, the development of techniques for imaging in the hard gamma-ray regime (above tens of MeV) was being fuelled by the recent advances in nuclear physics and understanding of cosmic-rays. Thus, at the opposite extreme, above energies of 10 MeV, imaging techniques evolved which also provided coverage down into the MeV regime.

The earliest space-borne gamma-ray telescopes optimised for gamma-ray detection above tens of MeV, relied on pair production kinematics to provide an imaging capability. The most successful and proliferate instrumentation developed during this period were

the spark chamber telescopes [102, 103]. However, as with the development of grazing incidence optics, attempts to increase the energy coverage by extending in this case, the low energy threshold, were hampered by fundamental physical limitations. Figure 2.2 shows schematically the basic elements of the spark chamber.

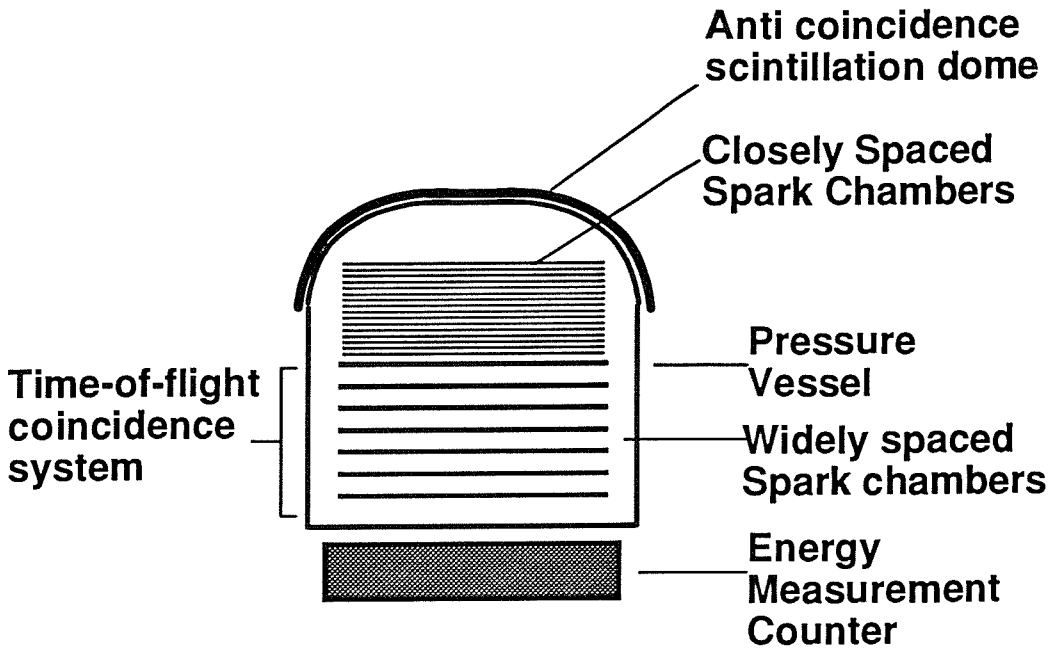


Figure 2.2: The basic construction of a spark chamber.

The limitations encountered can be understood by considering the construction of a spark chamber. Thin high-Z plates convert incoming gamma-rays into electron-positron pairs. These plates are interleaved with spark detection cells of only a few millimetres in width usually filled with Neon gas and operated at a voltage just below the break-down voltage. Thus on passing through a cell, an electron or positron will induce break-down via Coulomb interactions, the position of which can be measured with a multi-wire readout system. The accuracy of pair tracking and hence calculation of the incident photon direction, is dependent on the number of detection cells the pairs pass through.

Thus, although pair production is energetically possible above 1.022 MeV, for the photon direction to be accurately determined the pairs must have sufficient kinetic energy to pass through several detection cells. Consequently, in addition to a low energy threshold that is considerably higher than 1.022 MeV, there is also a trade-off between detection efficiency and angular resolution. A recent example of a spark chamber telescope is the Energetic Gamma-Ray Telescope Experiment (EGRET) onboard CGRO which has a mean detection efficiency of $\sim$25% and an angular resolution of 1.6^o at 100 MeV, reaching

0.2° at 2 GeV. The low energy threshold of EGRET is ~20 MeV and its FOV of ~ 40°×40° is far superior to that of any grazing incidence telescope [104].

2.1.2 Compton Telescopes

Motivated by the need to fill the gap between grazing incident telescopes, which at the time were capable of reaching up to only a few keV, and spark chambers which even now have low energy thresholds above 10 MeV, the late 1970s saw the development of telescopes for the 1 – 30 MeV regime [94]. Gamma-ray interactions in the 1 – 30 MeV range are dominated by Compton scattering in most materials and it is possible for telescopes in this range to provide an imaging capability through the use of Compton scattering kinematics.

The basic Compton telescope requires two position sensitive detection planes positioned one above the other relative to the pointing direction of the telescope. The aim is to produce a single Compton scatter event followed by a photoelectric event in upper and lower detection planes respectively. Consequently, the upper detector must have a high Compton interaction cross-section for the photon energy of interest. When such an event occurs the incident photon angle can be determined simply from the Compton kinematics equation, which when reconstructed via back-projection describes an event circle on the 'sky'. In reality the event circle is in fact an annulus, the width of which is determined by the propagation of the two energy deposit and position errors. Figure 2.3(a) shows an example of a Compton telescope, and (b) the simple back-projection kinematic reconstruction process.

The necessity of Compton scattering as the first interaction restricts the low energy threshold to around MeV energies. Although lower-Z upper detector planes would provide Compton scattered events at lower energies, the scatter efficiency would be negligible at higher energies and so there is a trade-off between low energy threshold and detection efficiency. COMPTEL onboard CGRO has a low energy threshold of 1 MeV and a mean detection efficiency of between 0.5 and 1%. The effective angular resolution of the Compton telescope is determined by the mean width of the combined event annuli and for COMPTEL this is around 2.1° at 1 MeV and 0.7° at 20 MeV [105]. As with spark chambers, the Compton telescope has a large FOV (~ 60° × 60° for COMPTEL).

From the above discussion it is apparent that there is a definite need for an imaging

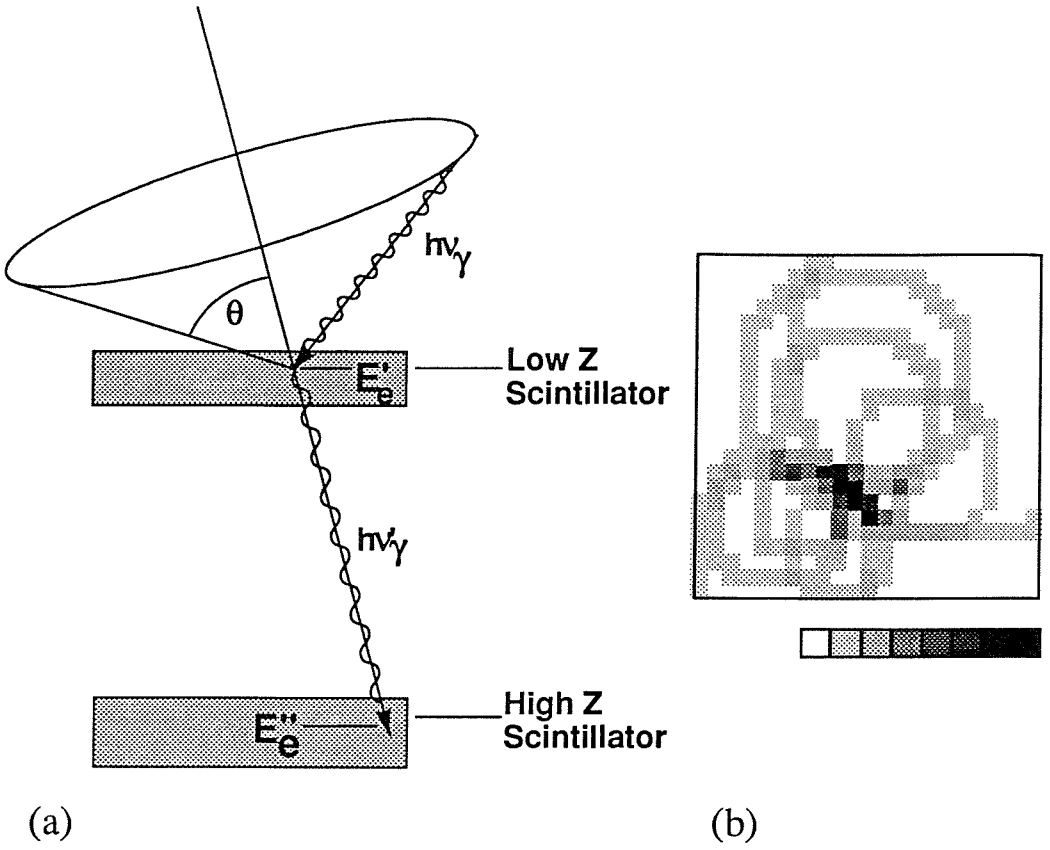


Figure 2.3: (a) The fundamental principle of the Compton telescope and (b) image reconstruction by back-projection.

system that can provide the high angular resolution of grazing incidence techniques, the large FOV of both spark chambers and Compton telescopes and a broad band *continuous* spectral coverage to link the X-ray band with the tens of MeV gamma-ray band.

2.2 Multiplexing Techniques

To bridge the gap between the X-ray and MeV gamma-ray band, the last two decades have seen the development of a class of telescope that is based neither on the wave nature of light (grazing incidence reflection) nor the particle nature of light (spark chamber, Compton telescope), but on simple geometric optics. All geometric optic telescopes have two things in common : (1) they attempt to encode the source strength and direction by means of source modulation, either temporally or spatially ; (2) the encoding stage must be followed by a decoding stage to extract the source distribution.

2.2.1 Temporal Multiplexing

The most sophisticated implementation of the temporal modulation technique is the Rotation Modulation Collimator (RMC), which has evolved from a concept originally conceived by Mertz [106] and adapted for X-ray astronomy by Schnopper [107], to the present day RMC, examples of which are the WATCH all sky monitors onboard GRANAT [108, 109]. RMCs were made partially obsolete with respect to precise source positioning by the development of grazing incidence telescopes. However, as WATCH has demonstrated, with an energy range of 10 – 200 keV RMCs can be useful as a low cost, technically simple, wide-field survey system which in theory can image up to several MeV.

An RMC consists of two identical sets of opaque and transparent bars, positioned one above the other (bi-grid collimator). A source in the FOV will project a shadow which is a combination of the two sets of bars onto a non-position sensitive detector which collects a source flux commensurate with that source position. The spatial information of the source is then encoded by rotation of the entire telescope about the pointing axis, producing a temporal modulation pattern that is unique, within some redundancy, for each source position. Figure 2.4 shows a schematic diagram of the WATCH RMCs.

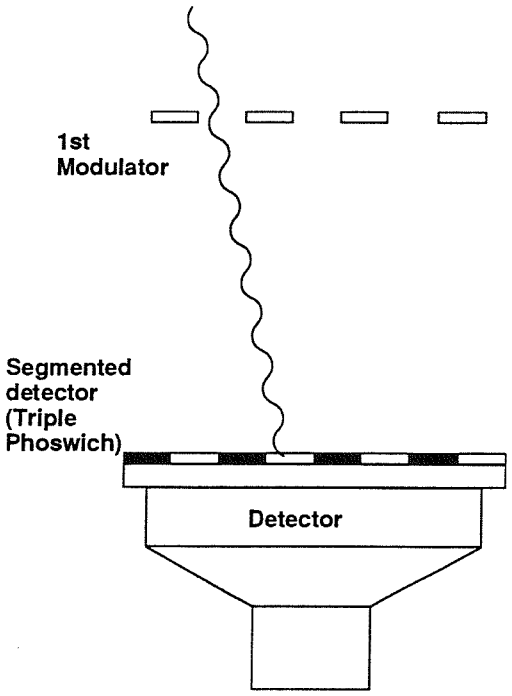


Figure 2.4: A schematic diagram of the WATCH RMCs onboard GRANAT.

It has been shown that the modulation pattern obtained is a partial sampling of the spatial frequencies describing the source distribution and as such, a 'reconstruction' can be

obtained by taking the Fourier transform of the modulation pattern [107]. A more effective way of reconstructing the source distribution can be achieved with cross-correlation [110] and indeed this will be shown later to be the 'optimum' reconstruction method for certain coded aperture systems. Cross-correlation for the RMC simply means correlating the detected temporal response function with all possible temporal response functions within the RMC FOV. However, the system point spread function (SPSF) of the RMC is far from ideal and is given by the zeroth order Bessel function (see Fig. 2.5).

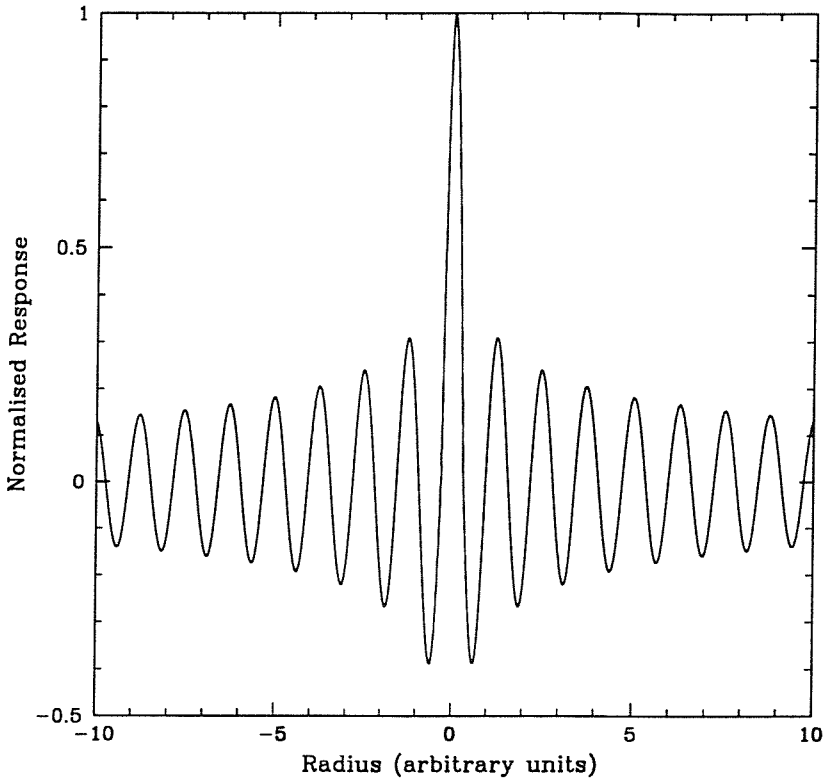


Figure 2.5: The SPSF of the WATCH RMCs.

Bi-grid collimator RMCs are however, not suited to imaging gamma-ray sources that are temporally variant since interference will occur between the source flux and source position time signatures. This problem can be overcome by replacing the lower collimator and detector with a detector comprised of parallel and independent detector segments (triple phoswich), forming effectively two anti-phase RMCs. Not only does this enable removal of any telescope independent time signature by reconstructing the difference between the two modulation responses, but also doubles the sensitivity.

However, the reconstruction artifacts that arise from the inherent undersampling of the spatial frequencies of the source distribution [111] (shown in Fig. 2.5), will conceal weak sources and even degrade the flux and position estimates of strong sources. The

simplest solution to this problem is to employ either some CLEAN-type algorithm [112] to iteratively remove the artifacts associated with each significant peak, or an iterative non-linear reconstruction process such as the Maximum Entropy Method (MEM). Although when coupled with post-processing or advanced reconstruction techniques RMCs are capable of producing wide FOV images (up to 130^o open angle - WATCH) with 20 arcminute resolution, by the very nature of the imaging principle they are unsuitable for variability observations.

2.2.2 Spatial Multiplexing

As opposed to temporal modulation techniques, spatial modulation techniques require a position sensitive photon detector and utilise the spatial multiplexing that can be obtained by using either pairs of widely separated collimator grids, or arrays of opaque and transparent elements (coded masks). The first class of multiplexing systems are known as Fourier Transform imagers [113], whilst the second class refers to Coded Aperture imagers.

Before continuing, it should be noted that RMCs are in fact the temporal equivalent of Fourier Transform telescopes, since they rely on sampling particular Fourier components of the source spatial frequencies. The Fourier Transform (FT) telescope achieves the same sampling of the source spatial frequencies via spatial modulation [114] and thus requires a position sensitive detector and unlike the RMC, the telescope is kept in a stationary position. As Fig. 2.6 demonstrates, a bi-grid collimator if configured correctly, can sample in one dimension one Fourier component (a single point in (μ,ν) space) of the source distribution.

By using collimators with a variety of spacings and orientations, the source angular distribution can be sampled at a variety of points in (μ,ν) space. Once a suitable number of Fourier components has been sampled, reconstruction can proceed with the superposition of the inverse Fourier transforms of each detected response. The main advantage of FT telescopes over other spatial multiplexing techniques is the less stringent requirements on the resolution of the position sensitive detector, as will be seen Chapter 3. Firstly, the detector only needs resolution in one direction. Secondly, the spatial frequency information is encoded by a modulation pattern whose period is of the order of the detector width and so large discrete samples of this modulation pattern will not severely degrade

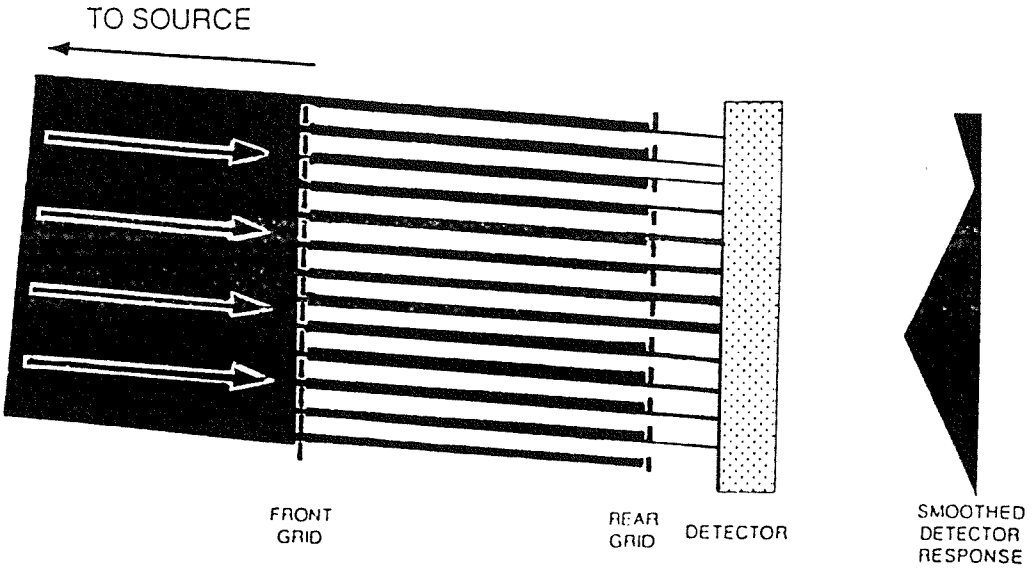


Figure 2.6: The bi-grid collimator of a Fourier Transform Telescope.

the angular resolution as determined by the pitch and separation of the collimators. However, with recent developments in pixellated detector technology, this advantage has become less significant.

In addition, without a sufficient number of collimator orientations and spacings, artifacts in the reconstructed image from undersampling the spatial frequencies of the source distribution, will severely reduce the sensitivity to weak sources. Again, as with RMCs, the situation can be improved by employing an iterative post-processing or reconstruction technique.

The second class of spatial modulation telescope, namely coded aperture systems, as will be seen can circumvent the problems of spatial frequency undersampling and produce images that are virtually artifact free. In addition, coded aperture systems are relatively simple to implement requiring no moving parts and only a simple mask of opaque and transparent elements which is considerably lighter than the several bi-grid collimators of the FT telescope. In the remainder of this chapter, the basic principles of coded aperture imaging, its historical development and methods for reconstructing the source distribution are discussed. The chapter concludes with an overview of the next generation of coded aperture based observatories, namely the INTEGRAL mission.

2.3 The Coded Aperture Concept

A coded aperture system designed to image in the photon energy band E_L to E_H, as mentioned above is a two stage process relying on simple geometric optics. The two basic components of the telescope are : (1) a coded aperture comprised of discrete elements which are either opaque or transparent to photons over the operational energy range, E_L – E_H. Generally, all mask elements are equal in size and are arranged on a regular grid ; (2) a position sensitive detector with sufficient spatial resolution to resolve the individual mask elements and sensitive to photons over the operational energy range, E_L – E_H.

The principle of the coded aperture telescope is relatively straightforward. Photons from a certain direction on the sky project a shadow of the mask onto the detector. The projection has the same coding as the mask pattern but is shifted according to the direction of the photon source. In a single observation, the detector collects a 'shadowgram' from each source in the FOV, whereby the source information is encoded in terms of the sky position according to the shadowgram shift, and an intensity according to the shadowgram strength. The observation stage must then be followed by a reconstruction stage in order to determine the source fluxes and positions. For optimum performance, the fundamental requirement of a coded aperture system is that every resolved sky position is encoded by the mask pattern in a unique way.

In general, three different geometric arrangements of the mask–detector system are practical and are shown schematically in Fig. 2.7. The simplest arrangement is to have the mask and detector of equal size (a), although such an arrangement suffers from the fact that the entire FOV, apart from the on-axis direction is only partially coded. The partial coding of a source can introduce artifacts into the reconstructed image when the simplest of reconstruction algorithms are used. However, the problems introduced by partial coding can be overcome by either employing a more advanced reconstruction algorithm, or simply by employing a detector that is larger than the mask (b), thus giving a wider fully coded FOV (FCFOV). The configuration most favoured in gamma-ray astronomy (c), employs a detector matched with the basic mask pattern, with the mask extended to provide the same FCFOV and sensitivity as in (b) but with a significantly reduced detector area.

Since a coded aperture telescope is a linear imaging system, the shadowgram formation process can be defined in terms of a convolution of the source distribution function with the aperture transmission function. If $\underline{\mu}$, $\underline{\nu}$ and $\underline{\omega}$ are vectors in the source plane, the

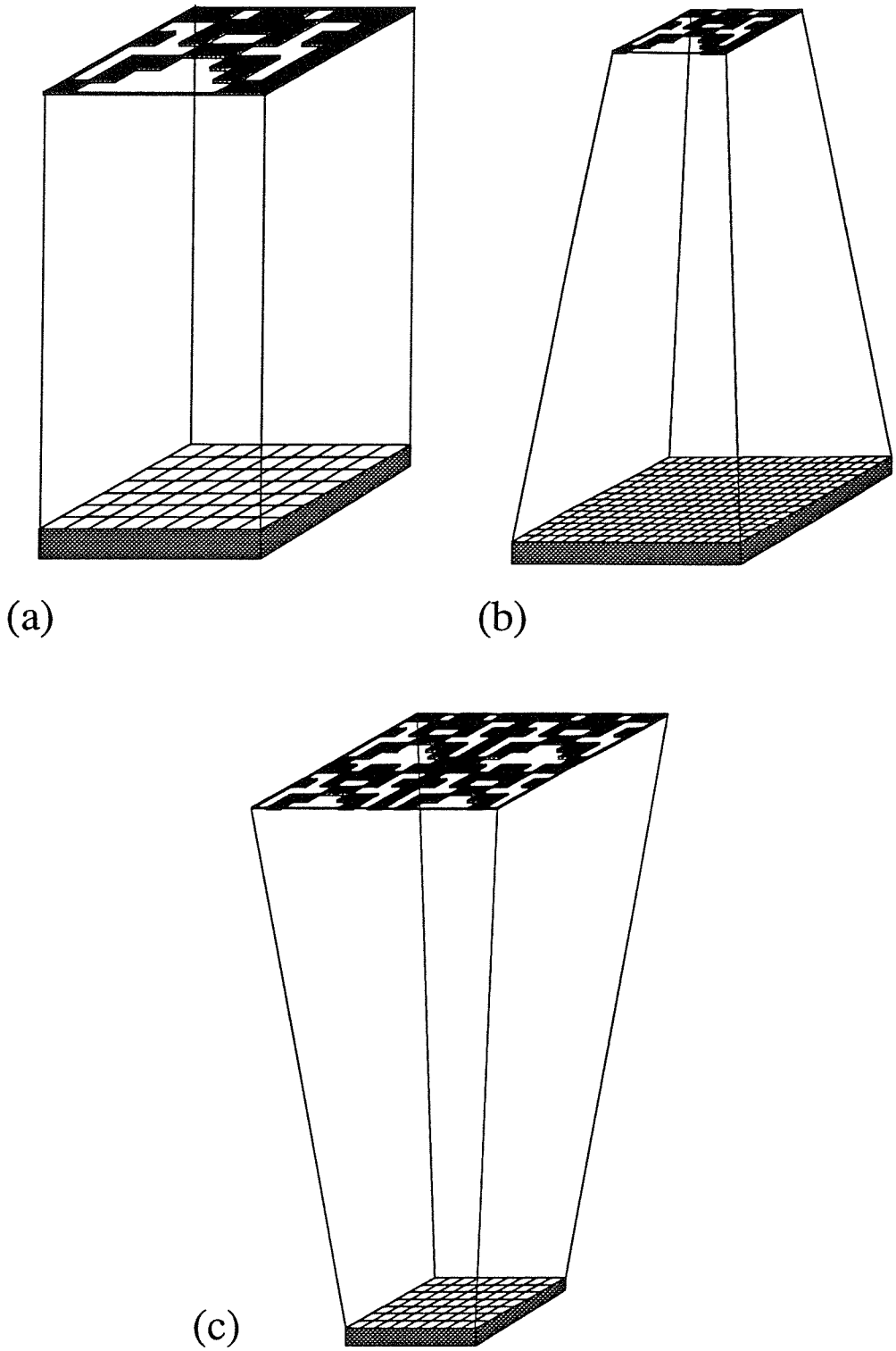


Figure 2.7: The most commonly used optical configurations for a coded aperture telescope. (a) The simple box camera arrangement, (b) a fully coded field of view system and (c) the 'optimum' configuration.

aperture plane and the detector plane respectively, the shadowgram convolution equation in shorthand integral notation is given by Eq. 2.1 :

$$P(\underline{\omega}) = F(\underline{\mu}) * R(\underline{\nu}) + B(\underline{\omega}) + N(\underline{\omega}) \ , \tag{2.1}$$

where $P(\underline{\omega})$, $F(\underline{\mu})$ and $R(\underline{\nu})$ are the detected shadowgram, the source distribution and aperture response respectively, $B(\underline{\omega})$ is the detector background component unmodulated by the coded aperture and $N(\underline{\omega})$ is the statistical noise associated with both background and source counting. One important point apparent immediately from Eq. 2.1 is that the statistical fluctuations from one source in the FOV will be induced in *all* other reconstructed sky pixels, a direct contrast to focusing systems where the statistical fluctuations in a single pixel are dependant only on the photon flux in that pixel.

Reconstruction methods for retrieving the source distribution are covered in detail in Section 2.6. However, prior to a discussion on the development of coded masks there is one important fact to note about image reconstruction : *all* linear methods of image reconstruction are equivalent to multiplying the recorded shadowgram data $[P]$ with a suitable reconstruction matrix $[G]$. Re-writing Eq. 2.1 in terms of matrix multiplication (Eq. 2.2), and then 'deconvolving' by use of a suitable reconstruction matrix, gives Eq. 2.3.

$$[P] = [R][F] + [B] + [N] \tag{2.2}$$

$$\begin{aligned} [\hat{F}] &= [G][P] \\ &= [G][R][F] + [G][B] + [G][N] \ . \end{aligned} \tag{2.3}$$

Since the fundamental requirement of reconstruction is that $[\hat{F}]$ is as close to $[F]$ as possible to within the statistical errors, Eq. 2.3 can be used to place restrictions on the required properties of the mask pattern. Equation 2.3 suggests that to achieve a unique aperture response for every resolved element of the sky, $[G][R]$ should equal the identity matrix $[I]$, or in terms of deconvolution, the correlation $G(\underline{\nu}) * R(\underline{\nu})$ must be equal to a δ-function – a central peak with perfectly flat side-lobes. Although not necessarily an indication of ideal imaging properties, it is apparent from the above discussion that

a δ-function auto-correlation function (ACF) is a good starting point when choosing or designing a mask pattern. The equivalent characteristic function in frequency space is the modulation transfer function (MTF) which is simply the Fourier transform of the ACF. In the ideal case the MTF has a single value for all spatial frequencies. The following section charts the development of coded aperture design with respect to the optimisation of image quality.

2.4 Development of Coded Masks

In addition to a δ-function correlation between the decoding array $[G]$ and the response function $[R]$, for gamma-ray astronomy it is desirable to maximise the detected source flux by ensuring the open area of the aperture is as large as possible. Thus, the following discussion is largely based on meeting these two fundamental mask pattern prerequisites.

2.4.1 Fresnel Zone Plates

The idea of coded aperture imaging was first conceived with the implementation of the Fresnel Zone Plate (FZP) as a coded mask by Mertz and Young [115] and although this type of mask was not applied to astronomical imaging, it was successfully implemented in medical imaging [116]. A FZP is characterised by the aperture function given in Eq. 2.4 :

$$M(r) = cos(r^2) \quad , \tag{2.4}$$

where, $r = (x^2 + y^2)^{1/2}$. The motivation behind the use of the FZP was the fact that it possessed both a δ-function ACF and 50 % transparency. However, practical implementation of the FZP requires digitizing the continuous response of Eq. 2.4 into a binary aperture transmission function, a restriction which degrades the ACF such that the side-lobes are no longer flat but contain concentric rings about the central peak. As will be seen in Section 2.6, the problems of a non-ideal mask ACF may be overcome to some extent with the use of alternative reconstruction techniques.

2.4.2 Random Pinhole Arrays

The imaging properties of the pinhole camera are ideal since it has a δ-function ACF. However, from a sensitivity perspective the small aperture is far from ideal and improvements in sensitivity can only be achieved at the expense of the angular resolution. The logical extension to the simple pinhole camera, was the random pinhole array proposed independently by Dicke [117] and Ables [118].

The random pinhole array consists of a large number of randomly spaced pinholes having an overall transmission of $\approx$ 50 %, and hence the sensitivity is increased in proportion to the number of pinholes, whilst the angular resolution of the single pinhole is maintained. In addition, the discrete nature of this type of aperture has the advantage of being simple to implement in comparison to the FZP. However, although the ACF of an infinite random array has the required δ-function for artifact-free imaging, when restricted to finite aperture sizes the ACF is degraded to a central peak with non-zero, randomly fluctuating side-lobes. Figure 2.8 shows practical examples of a FZP and a random pinhole mask along with their respective ACFs.

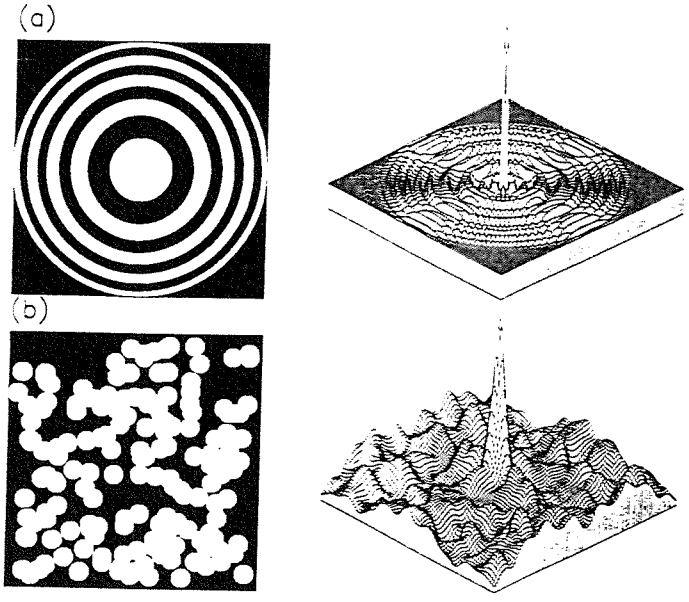


Figure 2.8: The FZP and random pinhole mask along with their respective ACFs.

Although imaging with random pinhole arrays has been applied to X-ray astronomy [119], later work on the development of coded aperture imaging concentrated on finding binary patterns that possessed either δ-function ACFs *or* δ-function SPSFs (the mask pattern correlated with a different decoding array gives a δ-function) when physically implemented. The research contained in this thesis concentrates on imaging with the former type of coded mask, thus, in the following section only patterns that possess

δ-function ACFs are discussed.

2.4.3 δ-function ACF Binary Arrays

Gunson and Polychronopulos were the first to realise that the sequences of numbers defined by cyclic difference sets (CDS) had the fundamental imaging properties required by a coded aperture telescope [120]. A CDS parameterised by the coefficients (ν, κ, λ) is defined [121] as the set of κ residues taken modulo ν, such that any non-zero difference between residue pairs taken modulo ν, occurs exactly λ times. To each CDS it is possible to assign a binary sequence, $\{\alpha_i\}$, of length ν where α_i takes the value 1 if i belongs to the CDS, and 0 otherwise. Baumert [122] found that the cyclic ACF of CDS binary sequences is *always* of the form :

$$\begin{aligned} ACF_i &= \sum_{j=0}^{\nu-1} \alpha_j \, \alpha_{j'} \qquad j' = (j+i)_{mod\ \nu} \\ &= \kappa \qquad if \;\; j = 0 \;\; (mod\ \nu) \\ &= \lambda \qquad if \;\; j \neq 0 \;\; (mod\ \nu) \\ &= (\kappa - \lambda)\, \delta_{jj'} + \lambda \ , \end{aligned} \tag{2.5}$$

and consists of a single peak on a perfectly flat background. The possible mapping of the 1-D CDS sequence onto a 2-D array was mentioned first by Gunson and Polychronopulos and later described explicitly by Procter [123]. The bi-valued δ-function ACF of both 1-D and 2-D arrays is a direct consequence of the uniform redundancy of the CDS sequence. The vector spacing between pairs of 1s in the array occurs exactly λ times and for this reason such binary sequences or arrays are known as Uniformly Redundant Arrays (URAs). There are two subclasses of CDS which are particularly suitable for generating mask patterns for coded aperture imaging. These are the Hadamard [124] and Singer difference sets [125, 122] :

- Hadamard sets are characterised by the parameters $\nu = 4t - 1$, $\kappa = 2t - 1$, and $\lambda = t - 1$, where t is an integer [126]. The known Hadamard difference sets may

be classified further according to their construction method, of which at least three are known :

1. Quadratic Residues (QR). If ν is prime the CDS is given by the residues $(mod\ \nu)$ of the squares of the first $(\nu+1)/2$ integers. However, since ν is prime, mapping onto a rectangular 2-D array is impossible. All patterns in the QR family (the skew-Hadamard sets) can be mapped onto hexagonal lattices [127]. Although not all of these arrays have hexagonal symmetry, they do all possess 180^o rotation pattern antisymmetry. A sub-classification of QR sets, parameterised by $\nu = 12t + 7$, give skew-Hadamard sets that not only possess hexagonal symmetry but also 60^o rotation pattern antisymmetry. The importance of this property with respsect to background subtraction is described in detail in Chapter 6.
2. Twin Primes (TP). Classified by $\nu = p \times q$, where p and q are prime and $|p - q| = 2$. It is possible to generate 2-D TP mask patterns from two QR sequences of order p and q without generating the ν-length binary sequence [128].
3. Pseudo Noise or M-Sequences (PN). In this case ν takes the value $2^m - 1$ where m is an integer > 1, and in fact these sets are complementary to the Singer sets with $t = 2$. The basic ν-length binary sequence can be generated using shift-register algorithms [129], which can then be mapped onto 2-D arrays using a number of procedures [130, 123, 131].

- Singer Sets are characterised by the parameters $\nu = (t^{m+1} - 1)/(t - 1)$, $\kappa = (t^m - 1)/(t - 1)$, and $\lambda = (t^{m-1} - 1)/(t - 1)$, where t is prime. It was this type of CDS that was first used to construct a 2-D coded mask [123]. Singer sets have a transmission that approximates to $1/t$ for $t \gg 1$, and so are not suited to gamma-ray imaging in the 15 keV to 10 MeV band where sensitivity can only be optimised by maximising the open area of the aperture. It can be shown that the optimum open fraction approaches 0.5 when the unmodulated detector background dominates the source flux.

Figure 2.9 shows some examples of optimum open fraction URAs.

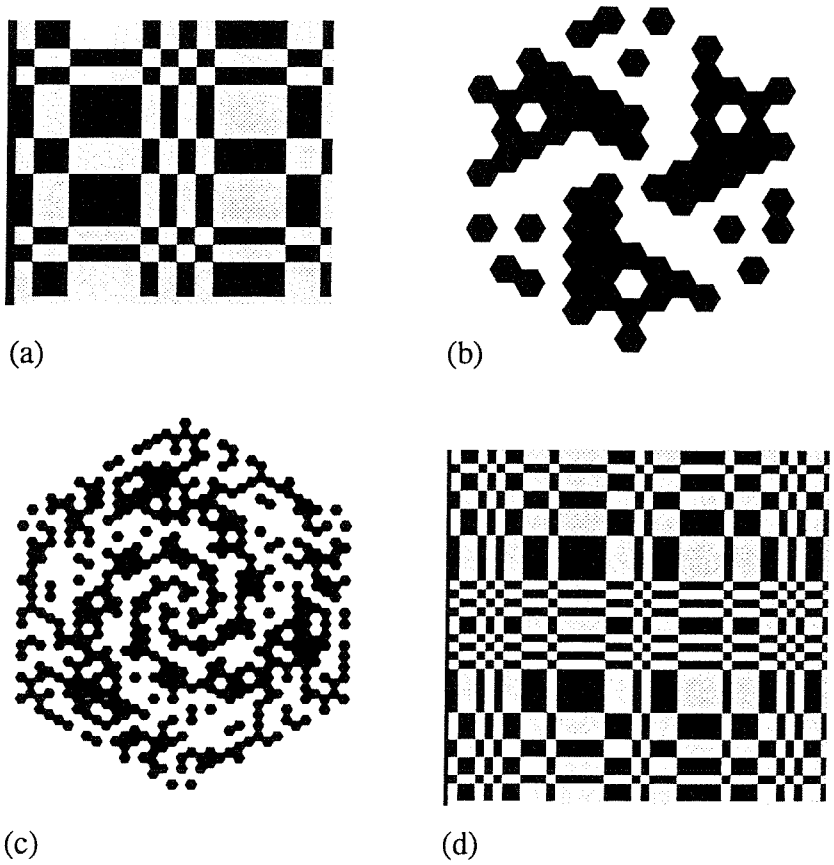


Figure 2.9: Examples of optimum open fraction URAs.

2.4.4 Alternative Coded Masks

In addition to patterns that have ideal ACF properties, there are a compendium of other mask designs that have been proposed for use, or have been used in coded aperture imaging. Common to the majority of these alternative designs is the fact that an ideal SPSF can be obtained by correlating the recorded mask shadow with a decoding array that is similar, but not identical to the mask pattern. Many of these designs were developed to provide, for example, patterns which had 90° rotational antisymmetry such as the modified URAs (MURAs) [132], patterns which were self-supporting such as the Pseudo-Noise Product (PNP) masks [133] or the generalised MURA-PN/MURA-MURA masks [134], or patterns with considerably lower open fractions such as the URA-MURA masks [135]. A thorough compilation of the various mask patterns proposed since the coded aperture concept was first introduced, including those mentioned above, has been produced by Skinner and Rideout [136].

2.5 Detector Requirements

There are several position sensitive detector systems suitable for coded aperture applications above energies of a few keV, although not all systems span the 15 keV – 10 MeV range, some are optimised for X-ray detection whereas others are optimised for mid to hard gamma-ray detection. To name but a few, in ascending order of low energy threshold, these include : proportional counter arrays [137] ; multi-wire proportional counters (MWPC) [138, 139, 140] (including the high energy extension configuration [141, 142, 143]) ; gas scintillation proportional counters (GSPC) operated in either the Anger Camera configuration [144, 145, 146], or the position sensitive photomultiplier tube (PSPMT) configuration [147] ; semi-conductor detectors such as silicon (Si), germanium (Ge) and cadmium telluride (CdTe) arranged in discrete pixel arrays [148] ; scintillation detectors operated in either the Anger Camera configuration [149], the PSPMT configuration, the position sensitive bar configuration [150], or the discrete pixel configuration with photodiode readout [151, 150].

2.5.1 General Properties

The general conclusions regarding the suitability of each of the above detector systems in the 15 keV to 10 MeV regime are as follows :

- Gas based detector systems are limited to operation at relatively low photon energies as a result of their poor detection efficiency above ~100 keV. They can however, provide good spectral resolution (~5%) over the 40 to 100 keV range, and good position resolution (0.5 mm for the MWPC).
- Semiconductor detectors have a fundamental advantage over other types of detection system with respect to their spectroscopic capabilities (1.5% at 100 keV for Ge). Combined with a high density and atomic number and thus high detection efficiency (2.9 cm attenuation length at 1 MeV), this type of detector is suitable for use over the 15 keV to 10 MeV regime. Fabricated in discrete pixel form, semiconductor arrays can provide position resolution on scales of several millimetres.
- The robust and technically simple nature of scintillation crystal detectors has resulted in the widespread implementation of this type of detector in both clinical and astronomical applications. Scintillation detectors can provide spatial resolution

on millimetric scales, over a broad spectral range (60 keV – 10 MeV) and with good spectral resolution (8% at 511 keV).

2.5.2 Spatial Resolution

Whatever detector configuration is chosen for a coded aperture system, it should be capable of producing an accurate recording of the source shadowgram and should meet a number of minimum requirements with respect to its response. Since both the INTEGRAL imager and spectrometer will use pixellated detector planes to achieve the spatial resolution required for optimised imaging, for the remainder of this discussion only pixellated detector designs are considered.

In the ideal case, the detected shadowgram is a perfect projection of the mask pattern for each source in the FOV. However, the pixellated nature of the detector response results in a blurring of the shadowgram. The degree of pixellation and therefore the accuracy with which the shadowgram position for a particular source can be measured, determines the source location accuracy (SLA). Similarly, although the nominal angular resolution of the telescope is determined by the angular size of a mask element at the detector plane, the pixellated response means nominally resolved sources may be blurred together and the angular resolution reduced. In addition, the degree of pixellation determines the percentage of shadowgram flux blurred between detector pixels and therefore the percentage degradation in sensitivity.

Thus, the finite detector resolution has a three-fold effect on the performance of the telescope. Chapter 3 examines in detail the effect of a pixellated detector resolution on the sensitivity of a coded aperture telescope. Similarly, Chapter 5 gives a qualitative example of the angular resolution degradation that occurs with increasing pixel size in the context of the INTEGRAL imager.

2.5.3 Flat-field Response

Spatial variations in the underlying detector response, as will be demonstrated in Chapter 6, can have a detrimental effect on the quality of the reconstructed image. Variations in the flat-field response can be attributed to either non-uniformity in the pixel-to-pixel characteristics or large-scale uniformity in the background counting rate. Non-

uniformity on pixel scales propagates through to the reconstructed image as an increase in the random background noise, whereas large-scale non-uniformity results in large-scale and structured artifacts in the reconstructed image. Chapter 6, Section 6.4 discusses in more detail the causes of detector non-uniformity along with techniques which enable the partial suppression of such non-uniformities. In conclusion, it is sufficient to say that for optimum sensitivity it is preferable for the flat-field response to be as uniform as possible.

2.6 Data Analysis for Coded Aperture Telescopes

Several techniques exist for analysing coded aperture data. The specific technique chosen depends on the goal of the observation (e.g. flux measurements of several known point-like sources, morphological studies of extended structure, or unknown source detection), the instrument configuration and the available processing power. This section gives an overview of some of the most commonly used analysis algorithms in coded aperture imaging. For each technique the general advantages and disadvantages are highlighted as well as any specific applications to which they are suited.

From Section 2.3 it can be seen that the shadowgram formation process can be expressed in matrix form in the following way, where R is the response matrix, F is the sky vector representing either single sources or a collection of sky pixels, and P is the detected shadowgram :

$$\begin{pmatrix} \\ P \\ \\ \end{pmatrix} = \begin{pmatrix} \\ R \\ \\ \end{pmatrix} \begin{pmatrix} \\ F \\ \\ \end{pmatrix} + \begin{pmatrix} \\ B \\ \\ \end{pmatrix} + \begin{pmatrix} \\ N \\ \\ \end{pmatrix}$$

Note that in this formulation, the unmodulated (by the mask) background component is given by a separate vector, where it could have just as easily been included in the response function and the source vector as an extra term. To clarify the precise form of R, consider a telescope based on the 'optimum' configuration defined in Fig. 2.7(c). If the detector plane employed provides sampling of the mask pattern on a 1:1 basis, the matrix R would have rows comprised of cyclic shifts of the basic pattern and bi-valued matrix elements, the precise values of which depend on a combination of the transmission of the

mask and the response of the detector. All of the following techniques aim to provide a solution to F from the observed data P and a knowledge of the response function, R. It should be noted that the presence of noise in the shadowgram formation process removes the possibility of a unique solution to the source vector, consequently all reconstruction techniques attempt to find the best solution according to some selection criteria.

2.6.1 Inverse Filtering

Equation 2.1 describes the shadowgram formation process in terms of a convolution between the source distribution and telescope response function with the addition of an unmodulated background function and a noise function. Since the required 'deconvolution' processes can be reduced to simple divisions in Fourier space, the imaging formulation given in Eq. 2.1 lends itself to a straightforward method of image reconstruction, namely inverse Fourier filtering. Providing the Fourier transform of each function exists, an estimate of F can be found from Eq. 2.6 :

$$\hat{F} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(P)}{\mathcal{F}(R)}\right) \quad , \tag{2.6}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are Fourier and inverse-Fourier transform operations respectively. The obvious problem with this formulation is that any zero or near-zero amplitude spatial frequencies in the MTF will result in amplification of the random source and background noise. In this case distortion and deterioration of the reconstructed image becomes dramatic even in relatively high signal to noise situations (SNR$\leq$100). The low or zero MTF amplitudes are characteristic of mask patterns that do not posses the ideal ACF discussed in Section 2.3 and are indicative of their 'ill-conditioned' nature. The problems with the MTF can be compensated for by the use of an additional filter, known as the Weiner filter $W(\omega)$, in the deconvolution process. The goal is to find $W(\omega)$ such that when evaluating

$$\hat{F} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(P)\mathcal{F}(W)}{\mathcal{F}(R)}\right) \quad , \tag{2.7}$$

$\hat{F}$ is as close to F as possible *in the least squares sense.* Thus, application of the least squares criteria between $\hat{F}$ and F leads to an optimum filter of the form :

$$\mathcal{F}(W) = \frac{|\mathcal{F}(R)|^2}{|\mathcal{F}(R)|^2 + r} \quad , \tag{2.8}$$

where r is the ratio of the noise power density to the source power density and $*$ denotes the complex conjugate. A generalised version of this type of filtering has been applied recently to coded aperture imaging within the matrix formulation of the imaging process [152]. Such generalised minimum-mean-square-error (MMSE) [153] versions of Weiner filtering make this technique applicable to a wide range of coded aperture systems where far from ideal mask patterns are employed. The MMSE filtering techniques however, are limited in that they *can not* compensate for imperfections in the shadowgram recording process caused, for example, by detector pixel failures [154].

2.6.2 Inverse Matrix

Continuing with the matrix formulation of the imaging process, the next logical solution to Eq. 2.3 would be to multiply the observed data P by the inverse of R to give :

$$[R]^{-1}[P] = [R]^{-1}[R][F] + [R]^{-1}[B] + [R]^{-1}[N] \quad . \tag{2.9}$$

The success of matrix inversion relies heavily on several essential characteristics of the telescope configuration. The response function R must be square and non-singular, that is to say the rows of R are not dependent or nearly dependent (ill-conditioned) on other rows. The consequence of such row degeneracy is elements of R^{-1} that are large enough for the $R^{-1}B$ term in Eq. 2.9 to dominate the reconstructed source term and degrade the image irretrievably. Even if these requirements are satisfied, the inversion process may itself be impractical from a computing perspective. For the majority of astronomical applications, the engineering and payload restrictions are such that a configuration optimised for matrix inversion reconstruction is impracticable. It should be noted however, that if R is ill-conditioned or not square, it is still possible to define a solution to Eq. 2.9 by use of the 'generalised inverse' techniques [155] such as Singular Value Decomposition (SVD) [156]. The use of SVD for pixellated coded aperture telescopes is discussed in more detail in Chapter 3, Section 3.6, but an important point to note again is that SVD attempts to find a solution to F from a least-squares perspective, as with MMSE filtering.

2.6.3 Cross Correlation

The third method of direct decoding suggested by both Eq. 2.1 and Eq. 2.2 and shown in Eq. 2.3 is by correlation of the recorded shadowgram P with a decoding function G. Since the correlation can be performed either in the Fourier domain using direct multiplication, or in the spatial domain using cyclic matrix multiplication, the similarities between inverse filtering, inverse matrix and correlation reconstruction now become apparent. It should therefore come as no surprise that for the ideal configuration, in which a δ-function ACF mask pattern is employed cyclically and sampled completely by the detector plane, inverse filtering, matrix inversion and cross-correlation produce identical images. Although at first the idea of correlating the recorded shadowgram with a decoding matrix appears somewhat arbitrary, it will be demonstrated in Chapter 3, Section 3.4 that correlation reconstruction has a sound statistical basis and as with inverse filtering and matrix inversion, is essentially a least squares fit to the observed data. Since the research contained in the following chapters is based mainly upon the implementation of ~50% transparent δ-function ACF mask patterns, only correlation reconstruction for this type of mask is considered from herein.

Considering the observed shadowgram P, as a vector of length N, where N, for simplicity is the number of sky pixels as well as the number of detector pixels, then the reconstruction of the ith sky pixel by correlation can be expressed by Eq. 2.10 :

$$\begin{aligned} \mathbf{P} &= \sum_{j=1}^{N} \mathbf{R_j} F_j + B\mathbf{u} + \mathbf{n} \\ \hat{F}_i &= \mathbf{G_i}.\mathbf{P} \\ &= \sum_{j=1}^{N} \mathbf{G_i}.\mathbf{R_j} F_j + \mathbf{G_i}.\mathbf{u}B + \mathbf{G_i}.\mathbf{n} \ , \end{aligned} \tag{2.10}$$

where $\mathbf{u}$ is the vector $(1, 1, ..., 1, 1)$ and as in Eqs. 2.1 and 2.2, $\mathbf{n}$ describes the statistical fluctuations in the source and background fluxes. Since correlation analysis attempts to utilise the ideal ACF properties of the URA, in general the decoding array $\mathbf{G_i}$ is based directly on the mask pattern and is usually of the following form :

$$\mathbf{G_i} = (\alpha + \beta)\,\mathbf{M_i} - \beta\mathbf{u} \quad , \tag{2.11}$$

where, $\mathbf{M_i}$ is a vector describing the mask pattern sequence and α and β are two free parameters which can be adjusted to alter the properties of the correlation image. Substituting Eq. 2.11 into Eq. 2.10 and using Eq. 2.5 for the definition of the ACF, the ith reconstructed sky pixel is thus :

$$\begin{aligned} \hat{F}_i = \; & \kappa\alpha F_i + \textstyle\sum_{j\neq i}((\alpha+\beta)\lambda - \beta\kappa)F_j \\ & +((\alpha+\beta)\kappa - \beta\nu)B \\ & +(\alpha+\beta)\mathbf{M_i}.\mathbf{n} - \beta\mathbf{u}.\mathbf{n} \quad , \end{aligned} \tag{2.12}$$

where for simplicity R_i and M_i are considered identical. By adjusting α and β the correlation image may be manipulated to achieve certain desirable properties : (1) the level shift introduced into the image by a uniform detector background is given by $((\alpha+\beta)\kappa - \beta\nu)B$ and may be eliminated if $(\alpha+\beta)\kappa = \beta\nu$; (2) the level shift introduced into the image by other sources in the FOV is given by $\sum_{j\neq i}((\alpha+\beta)\lambda - \beta\kappa)F_j$ and may be eliminated if $(\alpha+\beta)\lambda = \beta\kappa$; (3) it may be desirable for the value of the reconstructed pixel measured above the local background to be an unbiased estimate of the source flux, in which case (ignoring other sources in the FOV) it is required that $\kappa = \kappa\alpha - (\alpha+\beta)\lambda + \beta\kappa$.

Satisfying condition (3) with the choice of $\alpha = 1$ and hence $\beta = (\kappa - 1)/(\nu - \kappa)$, also satisfies condition (2) and for ~50% transparent URAs where $\kappa = (\nu + 1)/2$, leads to simple decoding arrays with $\alpha = 1$ and $\beta = 1$ [157]. Similarly, the choice $\alpha = (\nu - 1)/\nu$ and hence $\beta = \kappa(\nu - 1)/\nu(\nu - \kappa)$ satisfies condition (1) as well as (3) and corresponds to the 'balanced deconvolution' of Fenimore and Cannon [128]. In the following chapters, unless stated otherwise, all correlation image reconstruction is performed with 'balanced deconvolution'.

As mentioned above, the fact that the correlation operation can be performed with an $N^2/N \log_2 N$ speed advantage over the point-by-point cross-correlation, makes it a favourable deconvolution choice. In some circumstances however, the performance of correlation reconstruction may be severely degraded from a sensitivity perspective, particularly when the shadowgram is affected by modulations independent of the aperture

as with background non-uniformities. The sensitivity may be partially restored in these situations through the use of pre and post-processing techniques such as those analogous to the CLEAN algorithm frequently used in radio-astronomy [112]. A more detailed discussion of the pre-processing techniques is reserved for Chapter 6 in which they are used to suppress the effects of non-ideal observing conditions.

2.6.4 Iterative Maximum Likelihood Fitting

If direct decoding is not possible by one of the three linear methods outlined above, then the image reconstruction must proceed iteratively. Although iterative reconstruction techniques may be applied to the optimum coded mask configurations mentioned above, in general they are implemented only when the telescope response function is non-ideal and when attempts at direct linear deconvolution result in a severely degraded reconstruction. It should be noted that the generalised inverse techniques such as SVD mentioned above, are a form of iterative maximum likelihood fitting (MLF) where the iteration in this case is necessary to determine the inverse response function.

Most forms of MLF involve the iterative optimisation of a trial source solution vector F_T, via a χ^2 test statistic [158]. The success of any iterative procedure depends on the method adopted to update the trial solution using the differences between the model prediction and the actual observation and the method used to measure the 'goodness of fit' between the two.

Standard algorithms such as the Levenberg-Marquardt Method [159] iterate towards the minimum χ^2 by varying smoothly between the extremes of the inverse Hessian method (when the solution vector is far form the minimum) and the steepest descent method (as the minimum is approached). An alternative iterative procedure is embodied in the Richardson-Lucy algorithm [160, 161] which iterates towards the minimisation of the residuals $\mathcal{R}$, between the trial solution F^n and the recorded data via the following equation :

$$F^{n+1} = F^n \left(\frac{F^n R + \mathcal{R}}{F^n R} R^T \right) \quad . \tag{2.13}$$

Robust modifications to this simple algorithm include the addition of noise suppression through an intermediate step between iterations that uses wavelet transform analysis to

extract only the structurally significant residuals [162].

As has been mentioned in the previous sections a non-ideal response function will degrade the performance of the direct reconstruction techniques. In addition and perhaps more importantly, although the generalised inverse and iterative methods provide *a single* minimum χ^2 solution to the source intensities in this situation, there will exist *an infinite* set of solutions that will produce *the same* value of χ^2. Solutions in this infinite set differ from one another at the unmeasured spatial frequencies in the MTF by arbitrary amounts, and at the measured spatial frequencies in the MTF by amounts consistent with the noise. In the following section a image reconstruction technique is discussed, namely the Maximum Entropy Method (MEM), which has received much attention in recent years and aims to produce a single solution in the situations described above.

2.6.5 The Maximum Entropy Method

The Maximum Entropy Method (MEM) was first introduced into X-ray astronomy and then specifically into coded aperture imaging by Willingale [163, 154], but again as an iterative reconstruction process is extremely demanding of computer time. Despite this fact, the MEM has gained widespread favour in many different image processing areas as a tool for restoring degraded data [164, 165, 166, 167, 168, 169, 170, 171].

Essentially, the MEM non-linear iterative fitting scheme, constrains the infinite solution set mentioned in the previous section, by iterating towards the solution which contains the minimum amount of information as measured by its entropy. The entropy of the current iteration can be defined by the following equation :

$$\mathcal{S} = \int\int h[F(x,y)]dxdy \ , \tag{2.14}$$

where h is a suitably selected function. The most commonly implemented formulations of the entropy function (although the discussion in the literature as to which formulations are 'correct' is ongoing), are $h_1(F) = lnF$ and $h_2(F) = -FlnF$. There are various arguments in favour of either function (see for example [168]), although both posses the important properties of positivity enforcement and maximum entropy for a uniform image with constant intensity. The second property suggests that the MEM is non-committal with respect to unmeasured data and as a result produces images that are as featureless

as possible.

To see how the MEM differs from the standard iterative MLF routines, it is useful to consider the reconstruction problem from a Bayesian perspective. The act of observation with a coded aperture telescope, or indeed any instrument, can be considered as a filtering of the complete set of images corresponding to all possible intensity distributions. Only those images that satisfy the observed data to within the noise are then considered as possible solutions. The maximum entropy constraint results in the natural selection of the image that could have arisen in the maximum number of ways. This approach to statistical inference was suggested as early 1763 by Bayes [172] and can be summarized in terms of the following probabilities :

$$\mathcal{P}(F|P) = \frac{\mathcal{P}(F)\mathcal{P}(P|F)}{\mathcal{P}(P)} \quad , \tag{2.15}$$

where F is the unknown image and P is the recorded observation. The quantity $\mathcal{P}(F|P)$ is the probability of obtaining the image F given the measurements P and is the quantity which should be maximised in finding F. The right hand side of Eq. 2.15 comprises $\mathcal{P}(F)$ the *a priori* distribution of images, related to the entropy $\mathcal{S}$ of the image by

$$\mathcal{P}(F) \propto e^{\mathcal{S}(F)} \quad . \tag{2.16}$$

$\mathcal{P}(P|F)$ is the probability of obtaining the recorded data P for a given image F and for the simple case of uncorrelated Gaussian measurements, is given by

$$\mathcal{P}(P|F) \propto \prod_i exp\left(-\frac{\left(\sum_j F_j R_{ij} - P_i\right)^2}{2\sigma_i^2}\right) \quad , \tag{2.17}$$

an equation which will be met again in Chapter 3. The denominator $\mathcal{P}(P)$ is independent of F and ensures correct normalisation. Maximising Eq. 2.15 is equivalent to, after taking the logarithm, maximising the linear combination of the trial image entropy and a quantity measuring the discrepancy between the data and the predictions of the model from the trial image, namely the χ^2 merit function :

$$ln\mathcal{P}(F|P) = \mathcal{S}(F_j) - \sum_i \frac{\left(\sum_j F_j R_{ij} - D_i\right)^2}{2\sigma_i^2} \quad . \tag{2.18}$$

Again, as with MLF procedures, there are many iterative algorithms that through the maximisation of Eq. 2.18, produce the updating procedure for iterating from one trial solution to the next. These algorithms range from the simplest that are derived from a direct maximisation of Eq. 2.18 [165, 163, 154], to the most recent fully Bayesian approaches [173, 174, 175, 176, 177].

With regards to the computing effort required, all MLF and MEM techniques require a minimum of two convolutions per iteration : the first produces the trial recorded data through a convolution of the trial solution with the instrument response function, whilst the second updates the trial solution through a convolution of some residual vector with the (stable) transpose of the instrument response function.

It should be noted that the continuing controversy over the implementation of the MEM within the Bayesian framework is not centred on the fundamental Bayesian equation (Eq. 2.15), but on the choice of the *a priori* distribution, $\mathcal{P}(F)$. Even recent discussion of the MEM as applied to data from the Hubble Space Telescope [175] indicates that a general consensus has yet to be reached. One major advantage of the MEM lies in the fact that it appears to be a powerful tool for recognizing extended structures as well as point sources [178, 179].

2.7 The Next Generation of Coded Aperture Telescopes - INTEGRAL

The INTEGRAL proposal [180] was submitted to the European Space Agency (ESA) in response to a call for new mission proposals issued in 1989 in the second selection cycle (M2) of the Horizon 2000 long-term plan. The mission will be dedicated to the fine spectroscopy and imaging of celestial gamma-ray sources in the 15 keV to 10 MeV energy band through the implementation of a Ge spectrometer and a CsI(Tl)/CdTe imager both of which will use discrete pixel technology.

The majority of the research presented in this thesis was conducted under the auspices of the INTEGRAL mission, consequently this final section outlines the instrumentation

Parameter	Imager	Spectrometer
Energy Range	15 keV to 10 MeV	
Detector Area	2500 cm^2	327 cm^2
Spectral Resolution (E/ΔE @ 1 MeV	~25	~500
FOV - FC - PC - FWHM	3.2° 22° 6° (< 300 keV)	5.6° 13° 10°
Angular Resolution	17' FWHM	1.4°
PSLA - 20σ source	1'	10'
Continuum Sensitivity 3σ in 10^6 s @ 1 MeV, with ΔE = 1 MeV	3×10^{-8} ph cm^{-2} s^{-1} keV^{-1}	6×10^{-8} ph cm^{-2} s^{-1} keV^{-1}
Line Sensitivity 3σ in 10^6 s @ 1 MeV	1.2×10^{-5} ph cm^{-2} s^{-1} (20 keV FWHM line)	1.5×10^{-6} ph cm^{-2} s^{-1} (2 keV FWHM line)
Timing Accuracy (3σ)	0.1 ms	0.1 ms

Table 2.1: Key performance parameters of the INTEGRAL imager and spectrometer.

onboard the INTEGRAL observatory. The mission overview is based on the above mentioned INTEGRAL proposal, whereas references to the imager and in particular the research contained in the following chapters is based on the EIDOS scientific and technical plan [148] and where mentioned on the current INTEGRAL imager design, IBIS.

2.7.1 The INTEGRAL Observatory

Designed to build on the achievements of the CGRO and GRANAT missions, the INTEGRAL observatory offers improved sensitivity, angular resolution and energy resolution over a wide spectral band. At the time the research in this thesis was conducted, the two main instruments onboard the INTEGRAL payload (shown in Fig. 2.10) had the following baseline configurations : a cooled Ge spectrometer providing a high spectral resolution (2 keV FWHM at 1 MeV), a high line sensitivity and a 1.4° imaging capability ; a CdTe/CsI(Tl) imaging telescope providing a high angular resolution (nominal 17' FWHM), a fine point source location accuracy, a high continuum sensitivity and a good broadline sensitivity. Table 2.1 outlines the key performance parameters of these two instruments. A more detailed description of the INTEGRAL imager, EIDOS, is saved for Chapter 5 in which the kinematic incident pixel reconstruction capabilities are examined in detail.

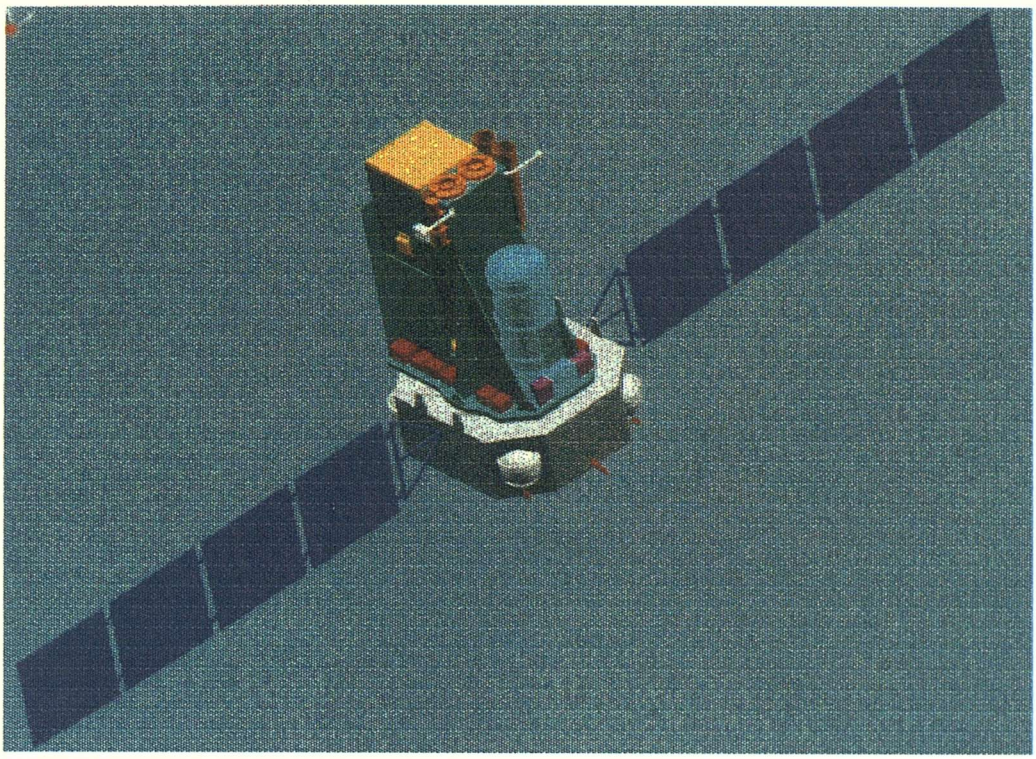


Figure 2.10: The deployed INTEGRAL spacecraft.

The concluding remarks of Chapter 1 suggest that an important factor in the success of next generation observatory missions is the degree of complementarity between different instruments. Indeed, the success of both the GRANAT and CGRO missions owes a great deal to the complementary nature of the instruments onboard which achieve the goal of simultaneous coverage in several overlapping wave-bands. The INTEGRAL mission has naturally adopted this fundamental philosophy and as well as the two main instruments outlined above also supports two monitor instruments, an Optical Transient Camera (OTC) optimised for the detection and location of the optical counterparts of GRBs, and an X-ray Monitor (XRM), also a coded aperture instrument, which will extend the coverage of the two main instruments down to a few keV for both imaging and spectroscopy.

2.7.2 Mission Overview

It is planned that the INTEGRAL observatory will be launched into a Highly Eccentric Orbit (HEO) with a relatively high inclination, although the specific characteristics of the orbit will depend on the type of launcher used. Important parameters in the orbit selection include a high apogee in order to avoid the radiation belts and to provide a constant

background level for long periods in the orbit. In addition, high apogee orbits reduce the occultation percentage from the Earths disk. Furthermore, high orbital inclinations allow good orbital coverage from Northern hemisphere ground stations which can be extended further by synchronisation of of the satellite orbit with the Earths rotation.

As with any space-borne mission, strict limits are imposed upon the payload mass and power requirements. For INTEGRAL these limitations restrict the total instrumental mass to ~2000 kg and the total payload power to 568 W. These limits arise from the fact that the spacecraft module developed for the ESA X-ray Multiple Mirror (XMM) cornerstone mission will also be used for INTEGRAL. The planned launch date for the INTEGRAL observatory is April 2001 and with a nominal mission lifetime of 2 years with a possible extension to 5 years, the INTEGRAL mission will carry gamma-ray astronomy well into the next millennium.

2.8 Conclusions

This chapter has outlined the motivation behind the development of the coded aperture technique for high angular resolution imaging above the classical X-ray regime. Through the advancements of pixellated detector technology, the coded aperture technique will take gamma-ray imaging into the next generation under the auspices of the INTEGRAL mission. Through the implementation of the innovative multi-layer detector configuration, the INTEGRAL imager is expected to be an order of magnitude more sensitive than SIGMA with an equally fine imaging capability and with the advantages of a wider operating energy range.

In the following two chapters the imaging capability of a coded aperture telescope based on pixellated detector technology is considered in terms of both the point source and extended source capabilities. Chapter 5 examines in detail the operational performance of the multi-layer detector geometry of EIDOS and develops the techniques necessary to retain good positional resolution over the entire energy range. Chapter 6 then examines the performance of a pixellated coded aperture telescope under extreme, but nevertheless possible observing conditions. Finally, in Chapter 7 the design and construction of a prototype imaging system based on EIDOS detector technology is outlined, through which the imaging procedures developed in Chapters 5 and 6 are then verified.

Chapter 3

The Point Source Sensitivity of a Pixellated Coded Aperture Telescope

Overview

The finite spatial response of a coded aperture telescope reduces the point source sensitivity to below the statistical limit given by the observed fluxes alone. The research in this chapter begins with an assessment of the sensitivity losses that occur when a detector with a pixellated spatial response is used and when correlation analysis is used to reconstruct the source distribution. The sensitivity loss is first investigated for square geometry coded aperture telescopes, the results of which are then also shown to be directly applicable to hexagonal telescope geometries. In the light of the EIDOS detector geometry, the work is then extended to include a more realistic detector configuration, where it is no longer possible to employ the ideal 'integer sampling' arrangement. However, it is shown that there are advantages in having a so called 'non-integer sampling' ratio for producing a more uniform sky response, especially at low mask element/detector pixel area ratios. The agreement between Monte Carlo simulations and theoretical predictions lead to the conclusion that the absolute sensitivity of a pixellated coded aperture telescope is independent of the deconvolution technique and depends only on the area ratio between mask element and detector pixel.

3.1 Introduction

The success of coded aperture imaging systems relies primarily on the existence of sets of numbers that when implemented as a tessellation of opaque and transparent elements have an autocorrelation function that consists of a central spike and perfectly flat side-lobes *i.e.* a δ-function. Consequently, if the shadow cast by such a mask from a gamma-ray source is sampled perfectly by the detector, and is thus an identical cyclic shift of the unit pattern (according to the source position), the shadowgram will contain all of the necessary spatial information describing the source distribution. Image reconstruction by correlating the recorded data with a level shifted/weighted version of the mask pattern will produce images that are artifact free, with flat side-lobes (excluding statistical fluctuations).

Such a system is obviously only an ideal. In practice the position sensitive detector by which the mask shadow is detected, may be far from perfect, and will always have a relatively finite spatial resolution. For an astronomical coded aperture system, the detector can take on a number of forms, including the Anger camera [181], discrete pixel devices [182], and multi-wire gas proportional counters [183].

Another possible detector system which has recently received much attention is the vertically stacked array of discrete pixel detectors [184, 180]. The advantages of such a detector system over the other devices include thicker detection modules enabling high energy photon detection, and a large area to mass ratio in comparison to the Anger camera. In a stacked array of discrete pixel detectors, a photon interaction can be detected as having occurred within a single pixel, thus providing the position resolution with which the source shadowgram is sampled. However, because there is no position information within the cross-sectional area of each pixel, the spatial resolution is equal to the distance between pixel centres.

The unavoidable consequence of the finite detector response is a 'blurring' of the projected source shadow. The resulting 'shadowgram' will no longer consist of simply a pattern of source plus background counts from the open elements and background counts only from the closed elements, but will also have intermediate flux values according to the response of the detector. The correlation function of the mask pattern with the recorded data is no longer equivalent to the ideal autocorrelation function of the mask, which as will be shown, reduces the telescope sensitivity for a given baseline geometry.

The research contained in the following chapter addresses the increasingly important situation in which coded aperture telescopes are employing pixellated detector arrays. The IBIS coded aperture telescope design [185] is the most recent example of the implementation of discrete pixel technology. The IBIS design illustrates how the sensitivity losses incurred by a non-optimised sampling configuration can be avoided by a careful selection of mask element and detector pixel size.

The latter part of the chapter begins to address some of the image analysis procedures required for the next generation of telescopes. For the particular detector geometry under consideration, accurate flux estimates and spectra, as well as detection significances, can no longer be determined directly from the correlation image, although a modified cross-correlation algorithm is shown to be accurate in determining source positions prior to a more rigorous analysis.

In such a situation a solution to the true sky distribution can only be found via the implementation of a simulated model of the telescope response. Thus, our ability to determine the source distribution to within the telescope sensitivity limits is now dependent on our success at simulating the response of the telescope.

3.2 Telescope Sensitivity vs Sampling Ratio - Square Geometries

Fenimore has previously quantified the expected imaging performance of the uniformly redundant array (URA) mask patterns [186] for an optimum imaging situation. In a later report Fenimore and Weston [187] showed using a brief qualitative example that the signal to noise ratio (SNR) of a point source can be severely diminished with a pixellated detector plane, if the source flux from a particular mask element is divided between detector pixels. This leads to what they term a *phasing error*, and also illustrated that using a fine sampling technique mitigates the phasing error. The SNR is defined here as the ratio of the reconstructed peak height above the surrounding plateau, to the statistical fluctuations of the surrounding plateau. Thus if the detector pixels were made smaller than the mask elements by an integer factor then the point source SNR could be partially restored closer to the maximum value, where the maximum value is obtained using a detector with the ideal response. Similarly, Caroli *et al* [157] have modelled the sensitivity

loss resulting from a Gaussian sampling of the mask, typical of detectors employing an Anger Camera arrangement. More recently, Finger has provided a thorough analysis of both the sensitivity and point source location accuracy (PSLA) dependence on sampling ratio for the Anger Camera arrangement [188]. Similarly, work by Stephen on the GRASP telescope, the predecessor to the INTEGRAL imager concept, has already quantified the effects of a discrete pixel detector on the PSLA [189].

Each of these independent studies reaches the same conclusion : the finite nature of the detector spatial response degrades both the PSLA and the sensitivity, such that the finer the resolution, the smaller the performance degradation. Since the PSLA has already been quantified for a discrete pixel detector [189], the work in this chapter concentrates on the sensitivity aspect only.

3.2.1 Image SNR Dependence on Source Position

The theoretical SNR in the presence of a phasing error can be calculated combinatorially for a two dimensional coded aperture system based on a square pixel geometry and using a URA as the mask. Assume that for each mask element there are an integer number of detector pixels in both the x and y directions, equal to n_x and n_y respectively. Such an arrangement allows reconstruction of the source distribution via a straight forward finely sampled cross-correlation of the detector array with a suitable decoding function [187]. This method will be shown later as the optimum form of reconstruction for a specific observation scenario (see Section 3.4). A schematic diagram of the type of sampling arrangement under consideration is shown in Fig. 3.1.

The centre to centre angular separations of the deconvolved sky pixels are b_x and b_y and are equal to the angles subtended by the detector pixels over the mask to detector separation. When a point source lies at the very centre of a sky pixel, the edges of the individual mask element shadows will coincide with the edges of the detector pixels in both directions, and no phasing error will occur.

In this case the deconvolved flux above the mean background level at the source position is equal to S_T. This gives the maximum possible SNR value, denoted by SNR_c. If the source does not lie at the centre of a sky pixel, there will be a phasing error. For a source at angular distances β_x and β_y from the nearest sky pixel centre (where $\beta_x \leq b_x/2$ and $\beta_y \leq b_y/2$), the decoded peak flux above the mean background now becomes

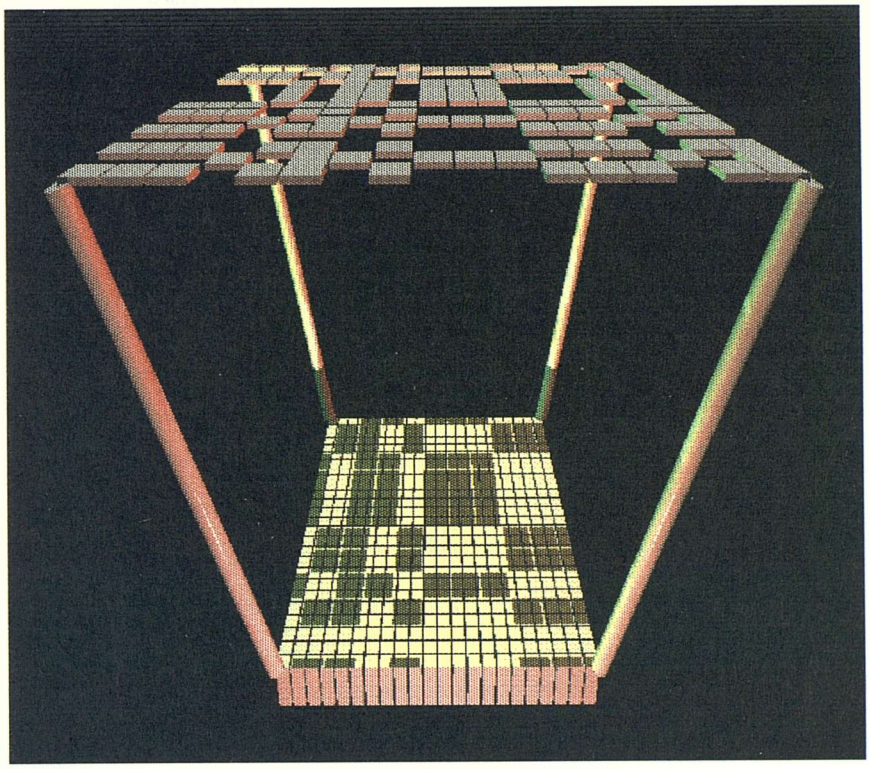


Figure 3.1: Example of a finely sampled coded aperture system. There are $v = 323$ aperture elements in this unit pattern, and the sampling ratio is $n_x = n_y = 2$ detector pixels per mask element.

$$F(\beta_x, \beta_y) = S_T \frac{1}{n_x} \left(n_x - \frac{\beta_x}{b_x} \right) \frac{1}{n_y} \left(n_y - \frac{\beta_y}{b_y} \right). \tag{3.1}$$

Assuming a high detector background rate, as is always the case for astronomical γ-ray (above 20 keV) observations, this diminished peak height leads to a correspondingly reduced SNR value, given by Eq. 3.2.

$$SNR(\beta_x, \beta_y) = SNR_c \frac{1}{n_x} \left(n_x - \frac{\beta_x}{b_x} \right) \frac{1}{n_y} \left(n_y - \frac{\beta_y}{b_y} \right). \tag{3.2}$$

Equation 3.2 predicts how the SNR of the deconvolved image should vary with source location for different sampling values. Assuming square detector pixels and mask elements, $n_x = n_y = \sqrt{n}$, Table 3.1 quantifies the predictions made by Eq. 3.2 for the two worst possible cases, namely at the sky pixel edge and corner, and shows that as the sampling increases the SNR tends towards a constant value inside a single sky pixel. Conversely, as the sampling decreases the SNR values have a larger range, with the SNR reduced to only 25% at the sky pixel corner for the worst case when $n = 1$.

Mask element area / Detector pixel area	SNR_{edge}	SNR_{corner}
1	0.50	0.25
4	0.75	0.56
9	0.83	0.69
16	0.86	0.77
25	0.90	0.81
36	0.92	0.84
49	0.93	0.86
64	0.94	0.88
81	0.94	0.89
100	0.95	0.90

Table 3.1: SNR values as a percentage of ideal SNR for increasing fine sampling values, and for sources positioned at the edge and corner of a sky pixel respectively.

Since the correlation map is used to determine source fluxes and positions either directly (when the sampling arrangement allows), or indirectly prior to a more rigorous maximum likelihood analysis, it is important that the correlation map is as uniform in response as possible. From the above analysis it is evident that for a single pointing observation, sources may be effectively 'lost' if they happen to lie at unfavourable sky pixel positions. Consequently, because of the significance of these results, computer simulations of a gamma-ray telescope have been performed in order to verify quantitatively the results in Eq. 3.2.

3.2.2 Monte Carlo Simulations and Results

The simulations have been performed using observational parameters typical for a satellite-borne gamma-ray imaging telescope. In general the input telescope parameters were based on those of the INTEGRAL imager design, initially for a square detector pixel/mask element geometry and in the following section (Section 3.3), for hexagonal geometries more representative of EIDOS. The observation parameters used in all simulations are given in Table 3.2 [180]. As shown, the observational energy range used was 100 keV to 10 MeV, this was the original baseline before the introduction of ISGRI – the Integral Soft Gamma-Ray Imager. This energy range, coupled with the predicted background count [180] (see Fig 3.2) and source flux [190] profiles, give the flux values shown in Table 3.2. In each of the following simulations the data points have been calculated from the mean values of 50 simulations per observation.

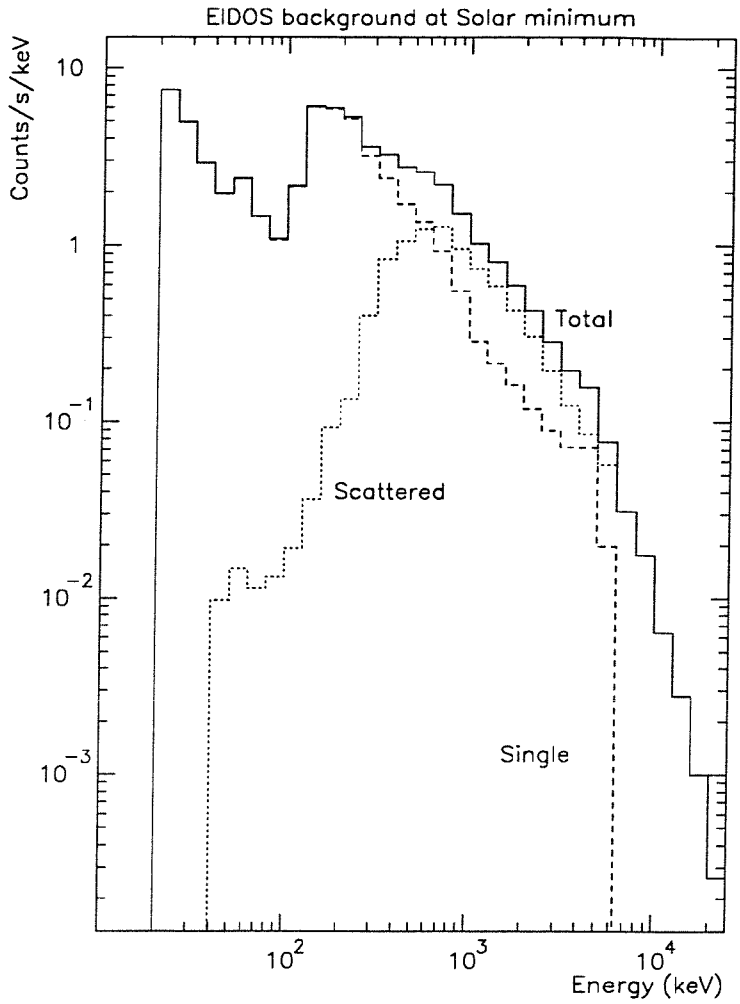


Figure 3.2: Predicted background counting rates at solar minimum by energy and mode for EIDOS.

3.2.2.1 Image SNR vs Fine Sampling

An initial analysis of the SNR of a point source with respect to the degree of discrete fine sampling employed was performed in order to quantify the absolute sensitivity of the correlation map in this situation. Mask unit pattern orders were chosen depending on the parameters of a particular study, and square detector pixels and mask elements are assumed ($n_x = n_y = n$), thus making the sky pixel dimensions a and b equal.

For the initial study an 11×13 unit pattern twin prime URA [128] was used. For each fine sampling value n, deconvolved SNR values were measured for a point source lying at three positions relative to the sky pixel geometry: (i) at a sky pixel centre; $\beta_x = \beta_y = 0$, (ii) centrally at the edge of a sky pixel; $\beta_x = b/2$, $\beta_y = 0$, (iii) at a sky pixel corner; $\beta_x = \beta_y = b/2$. These source positions were chosen as they represent extreme cases in

Detector area	2500 cm^2
Mask - detector separation	370 cm
Photon detection efficiency	100 %
Closed mask element opacity	100 %
Open mask element transparency	100 %
Observation time	10^5 s
Source flux ($\equiv$ 100 mCrab)	0.002 ph cm^2 s^{-1}
Background count	0.7 cts cm^2 s^{-1}
Observation energy range	100 keV - 10 MeV

Table 3.2: Observation parameters used in all simulations.

terms of the decoded SNR as predicted by Eq. 3.2.

Figure 3.3 shows the simulation results for the SNR dependence on sampling area ratio (n) for integer fine sampling values up to $n = 64$. Theoretical curves from the model in Eq. 3.2 and one-sigma errors for each data point are also shown. For all three sets of data the fit with their respective theoretical curves is good ('poorest' fit $\chi^2 = 10.72$ with 8 d.o.f.), thus verifying Eq. 3.2 for these particular source positions.

All other telescope parameters being equal, these results demonstrate the trade off between angular resolution and SNR in a discrete pixel coded aperture telescope, since for the same detector pixel size the nominal angular resolution[1] is equal to the angle subtended by a mask element over the mask to detector separation [157]. The following section investigates this trade off in greater detail using two specific telescope parameters, employing 1:1 and 4:1 sampling ratios respectively.

3.2.2.2 SNR vs Source Position - Specific Analysis

Again, the overall telescope parameters used were as outlined in Table 3.2. Following current methods for the optimisation of CsI(Tl)-PIN photodiode detectors [191], the cross-

[1]Note that the actual resolution for a particular sampling ratio will be worse than the nominal resolution, such that the lower the sampling ratio, the poorer the actual resolution (see Chapter 5, Section 5.6.2.2).

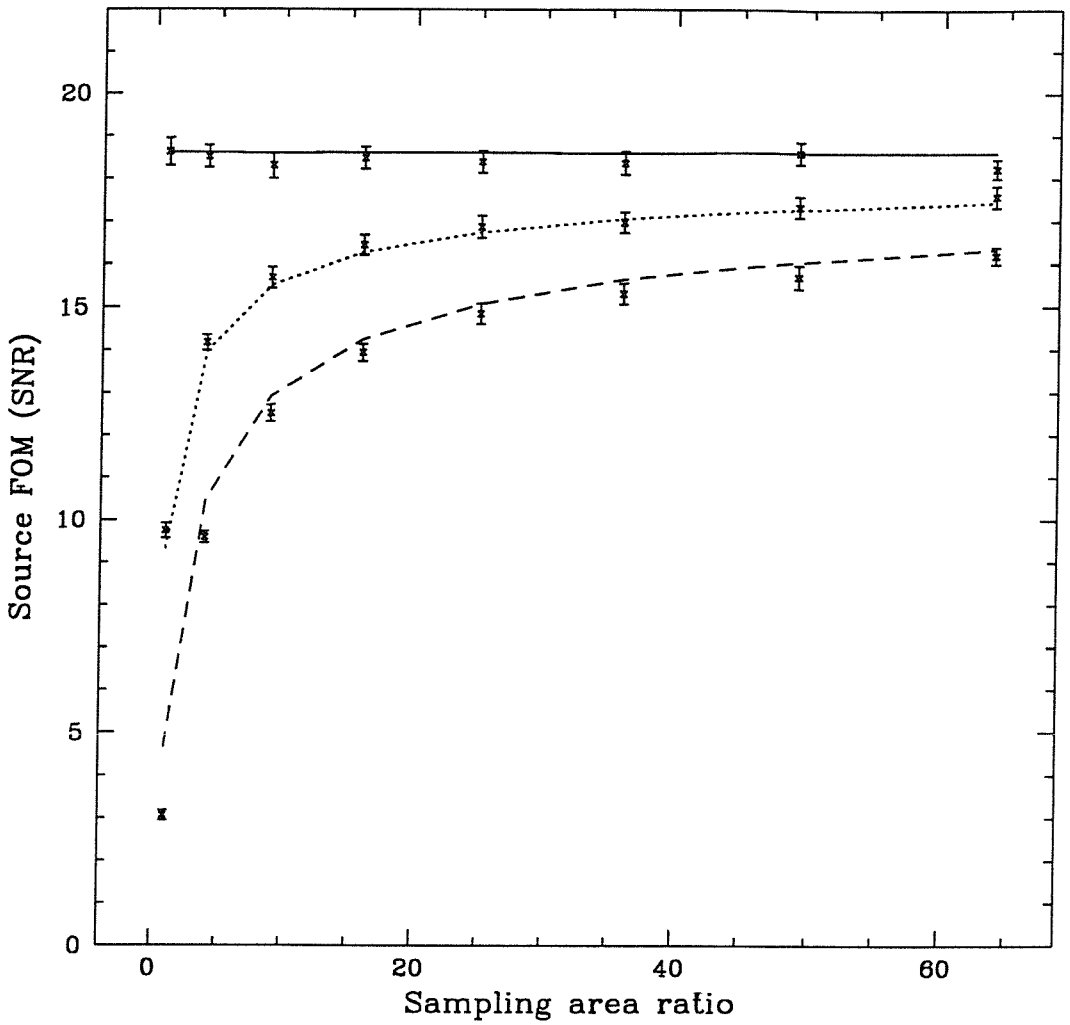


Figure 3.3: Image SNR vs sampling ratio for three different source positions within a sky pixel: centre of sky pixel (solid line); central edge of a sky pixel (dotted line); and the corner of a sky pixel (dashed line). The theoretical profiles are given by Eq. 3.2.

sectional area of the simulated detector pixels was restricted a minimum of approximately 1 cm^2, with slight adjustment to fit with the particular telescope configuration.

The fine sampling values investigated were $n = 1$ and $n = 16$. In the case of $n = 1$, a 41×43 element unit pattern URA was used. This pattern, coupled with the detector area of $2500cm^2$, gave square detector pixels 1.19cm across and a nominal angular resolution of 11 arcminutes. For the sampling value $n = 16$, an 11×13 element unit pattern URA was used, giving a detector pixel of side 1.05cm for the same detection area and an angular resolution of 39 arcminutes. Figure 3.4 shows the variation in the deconvolved SNR of a 100 mCrab source as the source position moves (a) from the centre of a sky pixel towards the edge of the sky pixel, and (b) from the centre to the corner of the sky pixel. In all cases the simulated data is plotted with the corresponding theoretical curves derived from

Eq. 3.2.

All data sets closely follow their respective theoretical curves ('poorest' fit $\chi^2 = 12.23$ with 11 d.o.f.), and so verify the theoretical model in Eq. 3.2. The trade off between angular resolution and SNR is clear from a comparison of the two graphs. From an SNR value of approximately 18σ when the source lies at the sky pixel centres, the SNR drops to around 4.5σ at the sky pixel corner for $n = 1$ (barely visible above the background noise), and to the better value of 14σ for $n = 16$.

A qualitative illustration of the improvement offered by the use fine sampling is given in Fig. 3.5. This figure shows, for each fine sampling value, a deconvolved image of three 100 mCrab sources within the field of view. In both images the source at the left is in the centre of a sky pixel, the middle source lies centrally at the edge of a sky pixel, and the right hand source is at the corner of a sky pixel.

3.2.3 Discussion

For a coded aperture system which uses a discrete pixel detector plane, both theory and computer simulations show that a serious sensitivity problem is encountered when the source shadowgram falls across detector pixel boundaries. When the source lies at the centre of a sky pixel the SNR of the reconstructed image is at a maximum. However as the source location moves towards the boundaries of two or more sky pixels, a phasing error is introduced causing a decrease in the SNR of the reconstructed object. The SNR reduction becomes more severe the smaller the sampling ratio n. In the worst cases this SNR degradation can be so large that a weak source may be lost in the deconvolved background noise.

The minimum SNR occurs when a source lies at a sky pixel corner, and the centres of the mask element shadows fall exactly across the corners of the detector pixels. This minimum SNR value is shown to increase by increasing the fine sampling parameter n, thus reducing the phasing error. For a fine sampling parameter $n = 1$ the reconstructed image SNR falls to 25% of the maximum when the source is located at the corner of a sky pixel. Increasing the fine sampling to $n = 16$ gives a minimum SNR of 77% of the maximum value. Because the phasing error is mitigated by the use of fine sampling then, if detector pixel sizes are fixed, a trade off between angular resolution and sensitivity results.

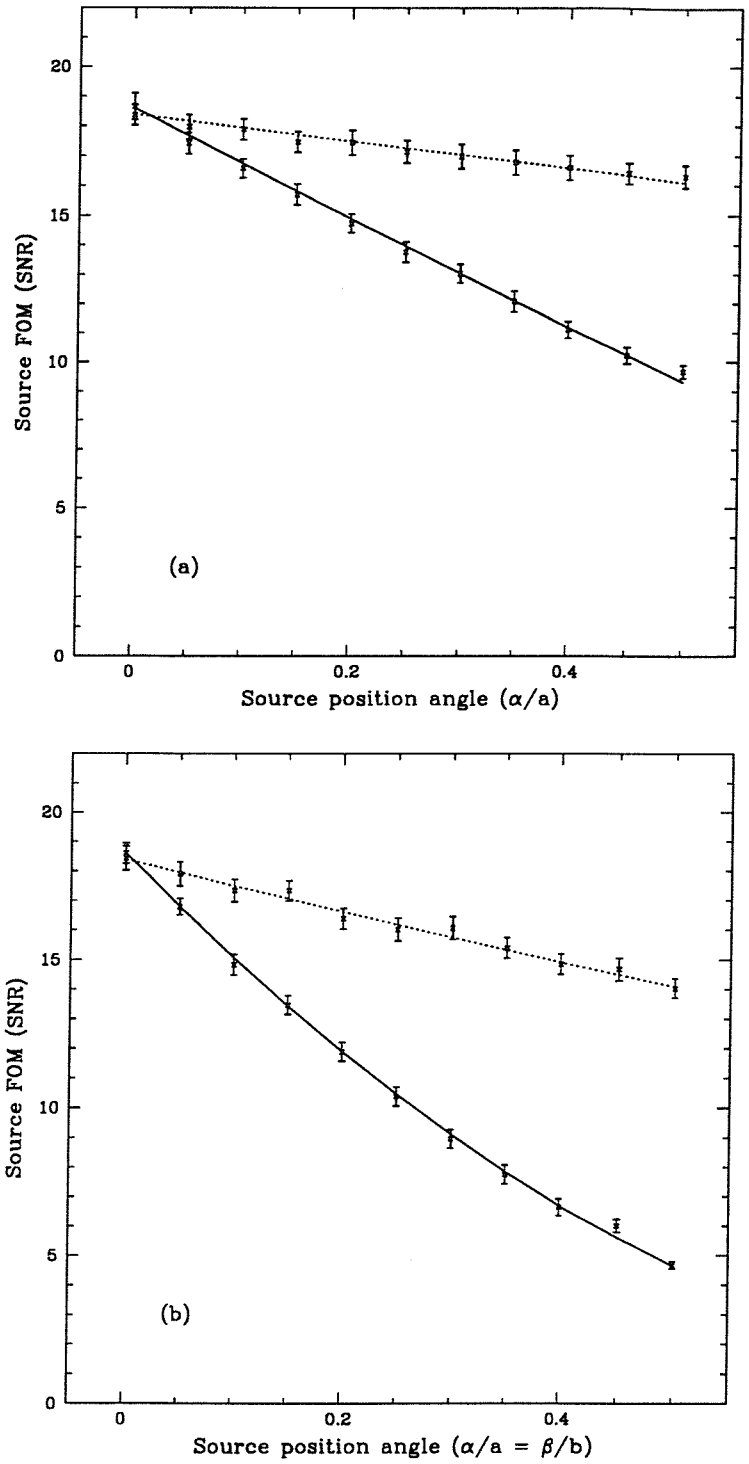


Figure 3.4: Image SNR variation with source position for two fine sampling values: solid line, 1:1, with angular resolution 11 arcmin; dotted line, 16:1, with angular resolution 39 arcmin. The simulations towards the central edge (a) assume $\beta_y = 0$, similarly the simulations towards the corner (b) assume $\beta_x = \beta_y$. The theoretical curves were derived from Eq. 3.2.

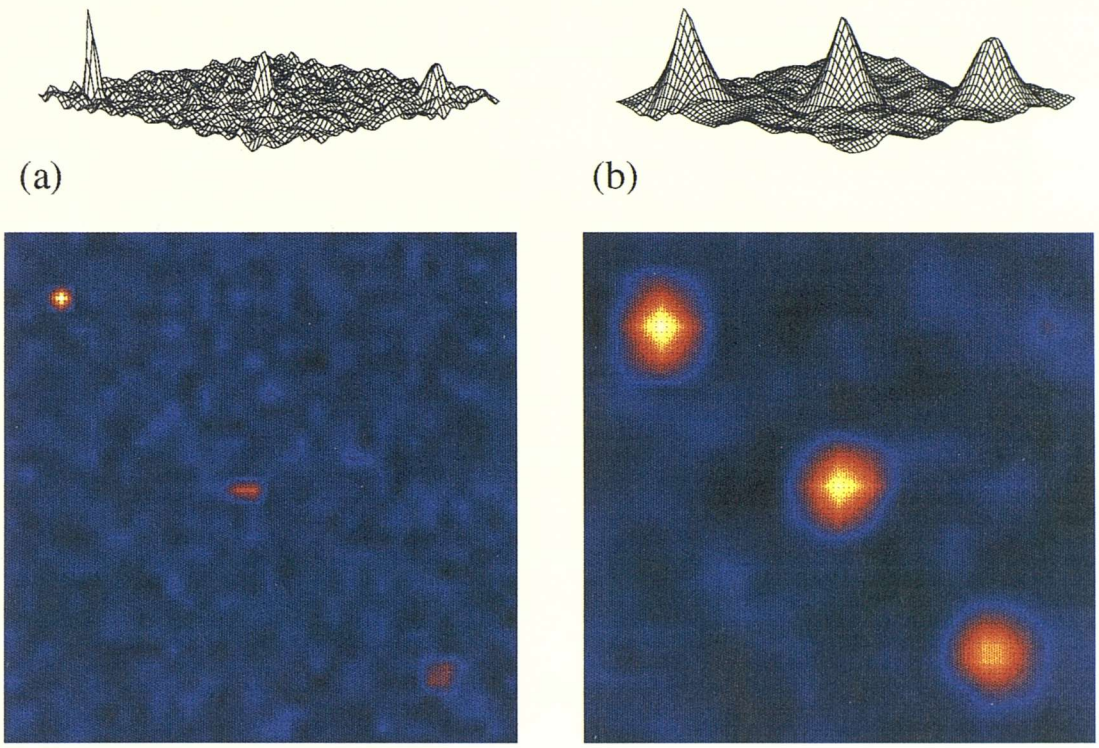


Figure 3.5: Deconvolved images of three 100 mCrab sources. The left hand source lies at a sky pixel centre, the central source lies centrally at the edge of a sky pixel, and the right hand source lies at a sky pixel corner, (a) $n = 1$, (b) $n = 16$.

The verification of the theory by the data enables the mean value of the SNR to be calculated versus the fine sampling parameter by using Eq. 3.2. This mean SNR is calculated by

$$SNR_m = \frac{\int_0^{b/2} \int_0^{b/2} SNR(\beta_x, \beta_y)\, d\beta_x\, d\beta_y}{\int_0^{b/2} \int_0^{b/2} d\beta_x\, d\beta_y} \tag{3.3}$$

$$= SNR_c \left(0.0625 \left(\frac{1}{n_x} \right)^2 - 0.5 \left(\frac{1}{n_x} \right) + 1 \right) \tag{3.4}$$

a model which is shown in Fig. 3.6 for a unit flux source.

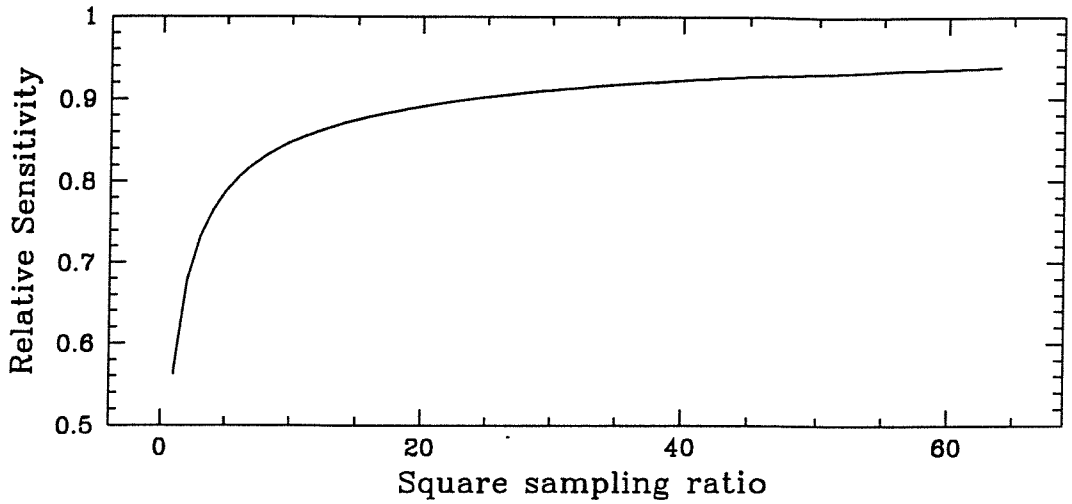


Figure 3.6: Graph of mean image SNR over a single sky pixel ($= SNR_m$) versus the fine sampling parameter n for a unit flux source.

3.3 Telescope Sensitivity vs Sampling Ratio - Hexagonal Geometries

The concluding remarks of the previous section were that for integer fine sampling configurations where an exact number of detector pixels sample each mask element, the phasing error can be suppressed and a more uniform response achieved by increasing the sampling ratio. Problems arise however if there are restrictions on the minimum detector pixel size - employing a higher sampling ratio results in a loss in angular resolution. In the framework of the EIDOS imager, and for completeness, the validity of Eq. 3.4 is now investigated for hexagonal geometries on the assumption that the telescope sensitivity is governed solely by the mask element to detector pixel area ratio, and is independent of the precise mask/detector geometry.

3.3.1 Image SNR Dependence on Source Position

Byard has suggested [192] that the use of hexagonal element uniformly redundant arrays (HURAs) with discrete pixel detectors introduces an intrinsic systematic coding noise resulting in non-zero side-lobes in the system point spread function (SPSF) unless specific sampling configurations are employed. These specific sampling orientations (see Fig. 3.7) are equivalent to the integer fine sampling configurations used by Fenimore and

Weston [187] with square geometries. Although these orientations enable the fast finely sampled cross-correlation algorithm to be used, and produce images free from the coding noise inherent at other mask orientations, the severe degradation in SNR towards sky pixel boundaries is still present. The specific mask orientations described by Byard also suffer from a second loss in sensitivity regardless of the source position resulting from the slight overlap errors that can be seen in Fig. 3.7 which culminate in a degraded peak height.

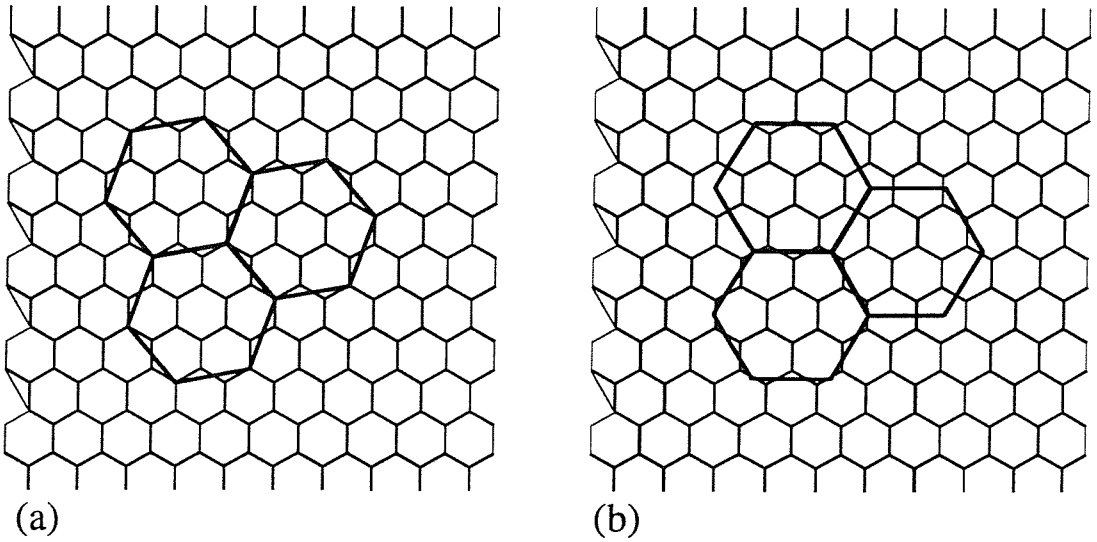


Figure 3.7: (a) The sampling orientations identified by Byard that produce zero side-lobe artifacts in the SPSF. However, the fractional overlap seen along mask element boundaries will propagate through to the correlation image as a degraded peak height. (b) For comparison the standard orientation is also shown.

In order to produce an analytical expression for the sensitivity degradation by following similar arguments to those used in Section 3.2.1, hexagonal geometry configurations were considered in which the mask elements were an exact integer number of detector pixels in size, thus retaining the flat side-lobes in the SPSF, whilst also avoiding the degraded peak height of the orientations used by Byard. This was achieved by simply tessellating a HURA onto a base element composed of an integer number of detector pixels, as in Fig. 3.8. The sensitivity degradation with respect to the integer sampling ratio, is now analysed as in Section 3.2.1.

For a source located at the very centre of a sky pixel, the shadowgram boundaries coincide with the edges of the detector pixels, and hence no phasing error occurs. Again, as in Section 3.2.1 as the source position moves towards the edge of a sky pixel the reconstructed

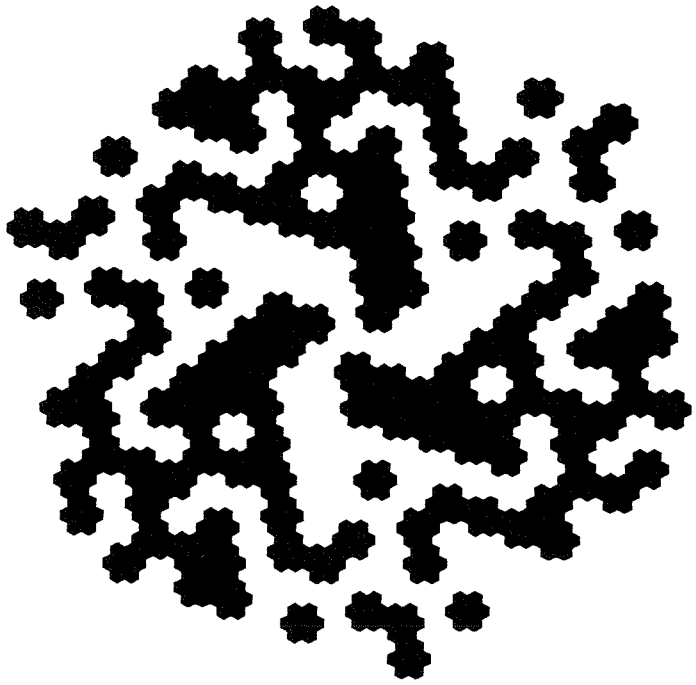


Figure 3.8: An example of a HURA tessellated onto a basic element composed precisely of an integer number of detector pixels. The mask shown has a sampling area ratio of 7:1.

peak height is degraded since the flux is now distributed between a greater number of sky pixels. Working in the hexagonal coordinate basis, the phasing error percentage for the 1:1 sampling case can be derived geometrically for one direction as :

$$\Phi(\beta_{hx}) = \frac{1}{3}\,(4\,\beta_{hx} - {\beta_{hx}}^2) \tag{3.5}$$

where β_{hx} is the source position in hexagonal sky pixel units away from the sky pixel centre. For two dimensions (but still in hexagonal unit vectors), and generalising for any integer sampling ratio, it is found that the reduction in peak height defined within a single sky pixel (all possible phasing errors that can occur are defined within a single sky pixel), is given by Eq. 3.6.

$$SNR(\beta_{hx},\beta_{hy}) = SNR_c\left(1 - \frac{N_{hx}}{3N}(4\,\beta_{hx} - {\beta_{hx}}^2)\right)\left(1 - \frac{N_{hy}}{3N}(4\,\beta_{hy} - {\beta_{hy}}^2)\right), \tag{3.6}$$

where $N_{hx} = N_{hy}$ is the number of detector pixels across the mask element and N is the sampling area ratio (number of detector pixels per mask element). Again, in order to verify

in a controlled way the validity of Eq. 3.6, Monte Carlo simulations of telescopes based on hexagonal geometries using the above 'tessellation' mask patterns were performed, and are now described.

3.3.2 Monte Carlo Simulations and Results

The observational and baseline telescope parameters as outlined in Table 3.2 were used for the following hexagonal geometry Monte Carlo simulations. The two specific sampling geometries simulated in Section 3.3.2.1 were chosen so as to retain a similar detector pixel size in both cases, and although the exact detector shape is different in each, the total detector area remains constant at 2500 cm^2. Complete sampling of each element in the unit pattern was provided in both arrangements, ensuring the SNR calculations taken from the correlation image were unaffected by SPSF artifacts.

3.3.2.1 SNR vs Source Position - Specific Analysis

The hexagonal fine sampling values investigated were 1:1 for the fine angular resolution mask, and 7:1 for the coarse angular resolution mask. The first arrangement (fine mask) uses a HURA with a 919 element unit pattern, thus requiring 919 detector pixels to supply full sampling. When coupled with the baseline detector area of 2500 cm^2, this gives a detector pixel size of 1.77 cm across flats (AF). The second arrangement (coarse mask) uses a HURA with a 127 element unit pattern, and so employs 889 detector pixels at 1.80 cm AF for the same detection area. With all other parameters remaining equal, the nominal angular resolutions are thus 15.2 and 46.4 arcminutes respectively.

As in Fig. 3.4, Fig. 3.9 shows the variation in the deconvolved SNR of a 100 mCrab source as the source position moves (a) from the centre of a sky pixel centrally to the edge of the sky pixel, and (b) from the centre to a sky pixel vertex and again the simulated data is shown with the corresponding theoretical curves, as derived from Eq. 3.6. The Monte Carlo results closely follow the respective theoretical curves ('poorest' fit $\chi^2 = 13.56$ with 11 d.o.f.), and thus verify the theoretical model for hexagonal sampling configurations as defined by Eq. 3.6. Figure 3.10 shows the effects of phasing errors on hexagonal geometry images for the two sampling ratios discussed above.

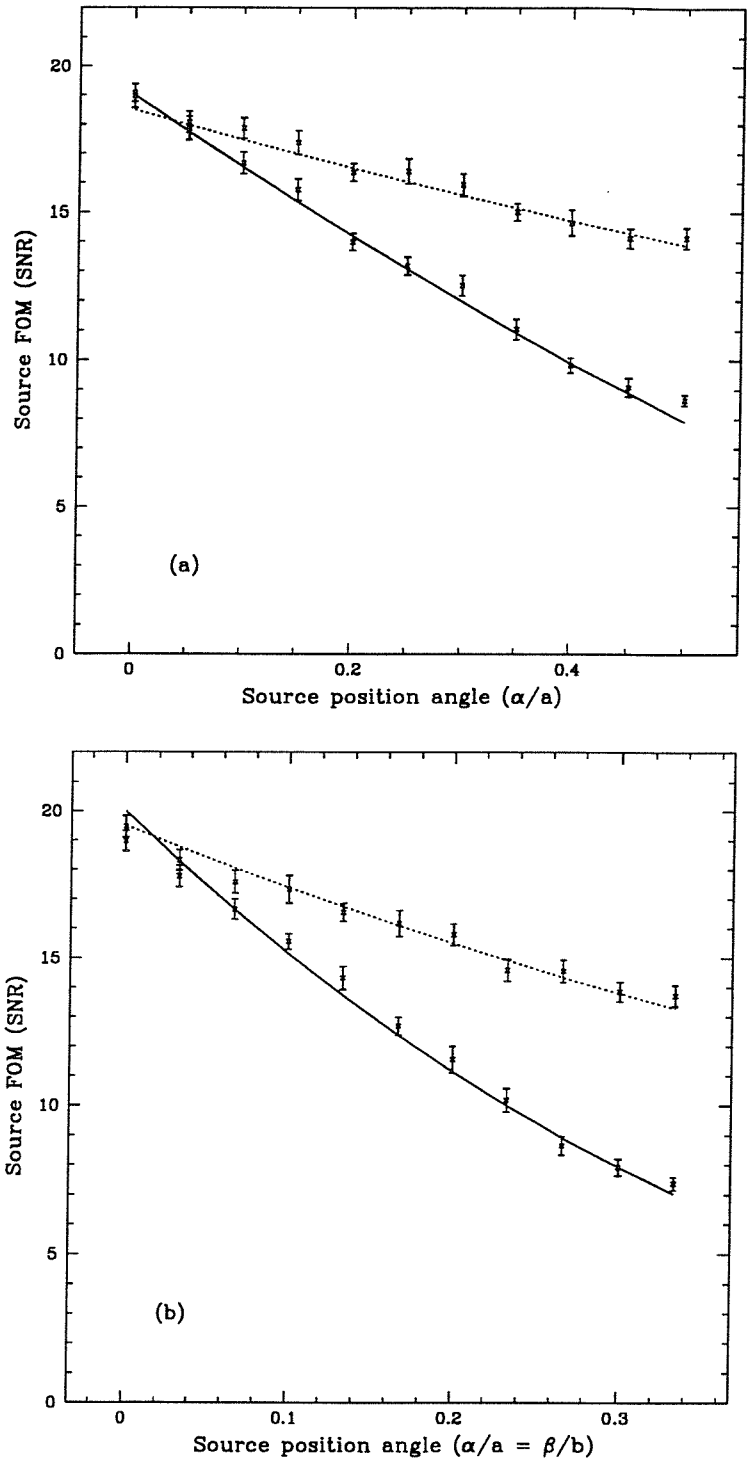


Figure 3.9: Image SNR variation with source position for two fine sampling values: solid line, 1:1, with angular resolution 15.2 arcmin; dotted line, 7:1, with angular resolution 46.4 arcmin. The simulations towards the central edge (a) assume $\beta_{hy} = 0$, similarly the simulations towards the corner (b) assume $\beta_{hx} = \beta_{hy}$. The theoretical curves were derived from Eq. 3.6.

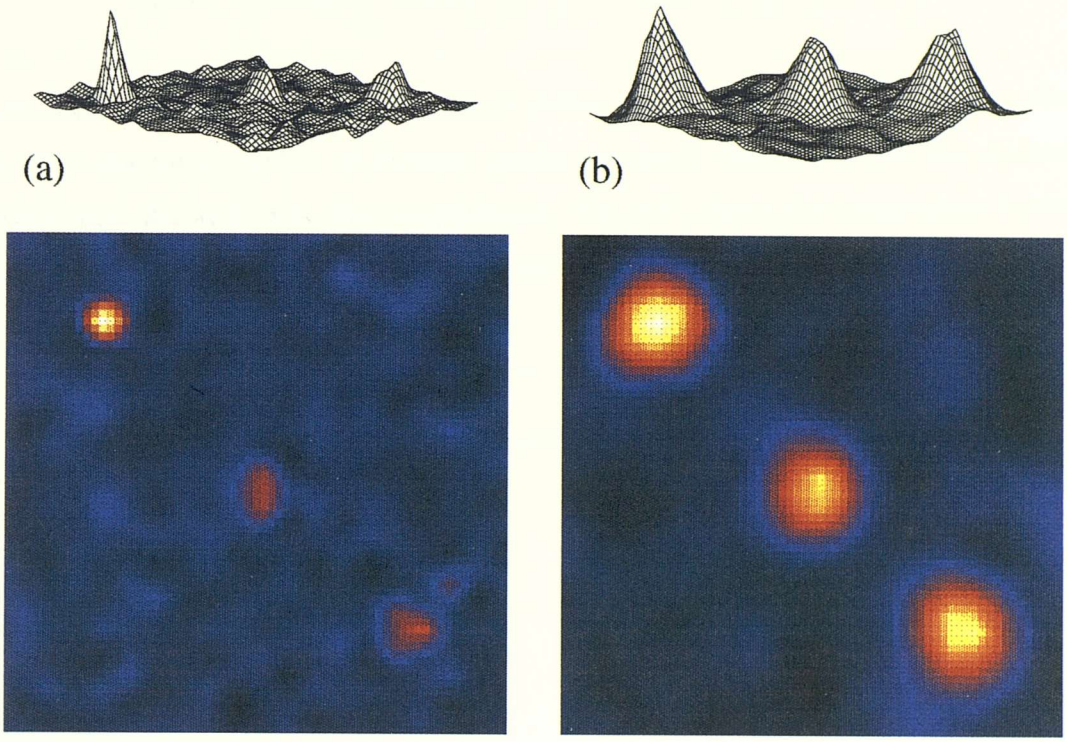


Figure 3.10: Deconvolved images of three 100 mCrab sources. The left hand source lies at a sky pixel centre, the central source lies centrally at the edge of a sky pixel, and the right hand source lies at a sky pixel vertex, (a) $n = 1$, (b) $n = 7$.

3.3.3 Discussion

In the previous section it has been demonstrated that similar phasing errors occur for hexagonal geometries as for square geometries causing a 'blurring' of the shadowgram. When limited by detector technology to pixel dimensions of the order of 1 cm^2, from the perspective of achieving uniformity in the SPSF, whilst retaining good angular resolution, it is apparent that a so called 'perfect' sampling arrangement is undesirable. For hexagonal geometries the worst possible phasing error occurs at every hexagonal sky pixel vertex where the correlation image SNR drops to 33% of the maximum SNR. Increasing the sampling seven-fold raises this value to 68% of the maximum theoretical value, but as before at the expense of the angular resolution.

Integrating Eq. 3.6 over the sky pixel area containing all possible SNR values (in a similar manner to Eq. 3.4) gives the mean reconstructed image SNR with respect to the sampling ratio :

$$SNR_m = SNR_c \left(0.054 \left(\frac{N_{hx}}{N} \right)^2 - 0.483 \left(\frac{N_{hx}}{N} \right) + 1 \right) \qquad (3.7)$$

where N_{hx} is the number of detector pixels across the mask element (as defined earlier), and N is the total number of detector pixels per mask element. Note the similarity between Eqs. 3.4 and 3.7. For completeness, Fig. 3.11 shows the comparison between hexagonal and square geometry sensitivity curves.

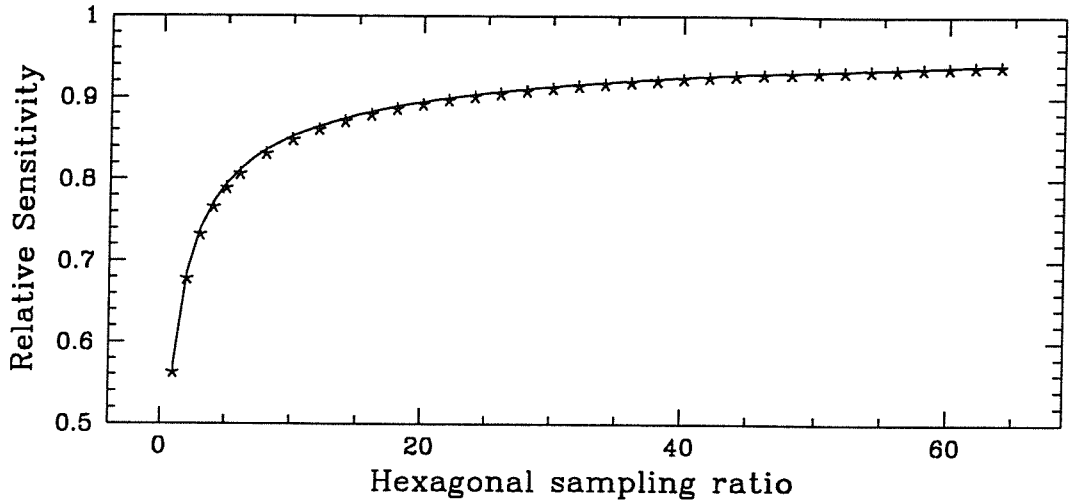


Figure 3.11: Graph of mean SNR for hexagonal geometries, for the same observation and telescope parameters used in Fig. 3.6. Continuous line for hexagonal geometries, data points for square geometries.

Both Eq. 3.4 and the above equation have been derived primarily for integer sampling configurations. Both suggest that care should be taken when selecting detector element sizes for discrete pixel, position sensitive arrays, or sensitivity may be compromised severely. But, is this also true from an absolute statistical perspective ? Equations 3.4 and 3.7 imply that for the same detector pixel size, and with all other telescope and observation parameters equal, increasing the mask element size allows fainter objects to be detected, a proposition that is not intuitively apparent.

In the following section, through a least squares approach for defining the SNR (as opposed to a figure of merit definition), it is shown that the absolute statistical significance of a point source is dependent *only* on the area ratio between the mask element and detector pixel, and is effectively independent of the pixel geometry or whether there are an integer number of detector pixels sampling each mask element.

3.4 Maximum Likelihood Analysis

Thus, with the aim of validating from a statistical viewpoint the analytical expressions derived in the previous sections, the quantification has also been performed by a least squares approach. This should not only provide an expression for Eq. 3.4 in terms of the absolute statistical sensitivity of a particular telescope geometry, but also confirm that the sensitivity degradation is independent of the deconvolution technique and intrinsic only to the sampling ratio of the system.

3.4.1 Source Significance vs Sampling Ratio

The analysis proceeds by considering a source arbitrarily positioned at known coordinates in the fully coded field of view (FCFOV). For practical telescopes, the response of the detector to that particular source is no longer the ideal two valued (1 or 0) function corresponding to whether the photon passed through an open or closed mask element. Instead, the pixellated response of the detector introduces intermediate values into the shadowgram response function, which are added to by the fact that the mask elements are never 100 % transparent/opaque over the entire energy range.

Although the shadowgram is no longer simply a combination of source plus background flux pixels and background flux pixels only, the data is still linearly dependent on the response function. Assuming this linearity, the significance of a point source detection can be determined from a maximum likelihood fit to the detector data/response function curve. This fitting procedure is now followed for a single point-like source in the FCFOV. An example of a typical set of observation data is shown against the corresponding response function values in Fig. 3.12.

The recorded shadowgram consists of a sample extracted from a parent distribution which in the simplest case can be defined with just two parent coefficients, the source flux F_0 and the total detector background rate B_0. The precise dependence between the observed detector pixel counts P_i and the detector response function R_i is then given simply by :

$$P_i = F_0 \; R_i + B_0 \tag{3.8}$$

The above equation assumes that the background counting rate is constant in all

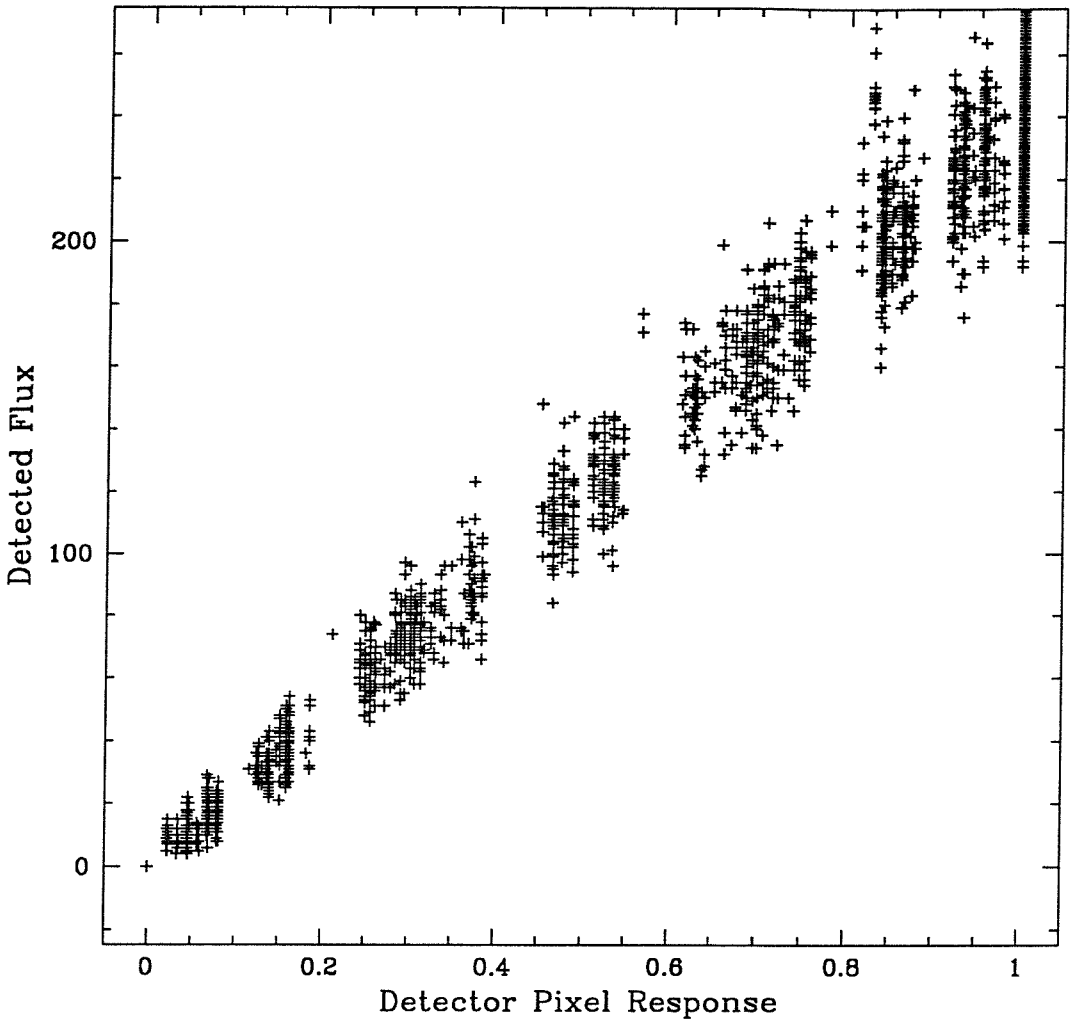


Figure 3.12: A plot of the recorded detector pixel values versus the response of the particular pixel to the sky. The data points shown were for a simulated observation of a 25 mCrab source and using the same general observation and telescope parameters as outlined in Table 3.2, although detector background has not been added.

pixels. It should be noted however that in the observational situations encountered so far [193, 194, 195, 180], the background rate has been found to vary at best, smoothly with position across the detector and at worse, from pixel to pixel as expected for the INTEGRAL imager. Thus ultimately Eq. 3.8 needs to include also a model of the detector background variation, a problem to which solutions are considered in Chapter 6. For a pixellated detector the probability of obtaining the observed set of pixel values during a particular observation for a sky model of $F\ R_i\ +\ B$, is simply the product of the of the individual probabilities for making the pixel to pixel observations of P_i with that model [159] :

$$Prob(F,B) = \prod_{i=1}^{N} \left(\frac{1}{\sigma_i \sqrt{2\pi}}\right) exp \left(- \frac{1}{2} \sum_{i=1}^{N} \left(\frac{\Delta P_i}{\sigma_i}\right)^2 \right) \tag{3.9}$$

assuming the measured P_i are taken from a Gaussian distribution (the detector background counting rate is sufficiently large to be governed by Gaussian statistics), and where the $\sigma_i = \sqrt{P_i} = \sigma_P$ are the standard errors on the P_i (assumed equal for each data point since it is the $B \gg F$ scenario that is being examined and so the covariance terms in the error propagation equations can be ignored). The products and sums are performed over the N detector pixels. The $\Delta\ P_i$ datum are the differences between the recorded pixel fluxes and the modelled parent distribution. The maximum likelihood estimates for F and B are thus those values that minimise the quantity χ^2, where

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{P_i - F\ R_i - B}{\sigma_i} \right)^2 \tag{3.10}$$

which expanding becomes

$$\chi^2 \equiv \frac{1}{\sigma_p^2} \left(\sum P_i^2 - 2\ F \sum P_i\ R_i - 2\ B \sum P_i + F^2 \sum R_i^2 + 2\ F\ B \sum R_i + \sum B^2 \right). \tag{3.11}$$

At the minimum value of the function χ^2, both partial derivatives $\partial\chi^2/\partial F$ and $\partial\chi^2/\partial B$ are equal to zero, thus yielding a pair of simultaneous equations for the parameters F and B. The least squares solution for the two parameters is thus :

$$F = \frac{1}{\Theta} \left(N \sum R_i\ P_i - \sum R_i \sum P_i \right) \tag{3.12}$$

$$B = \frac{1}{\Theta} \left(\sum {R_i}^2 \sum P_i - \sum R_i \sum R_i\ P_i \right) \tag{3.13}$$

$$\Theta = N \sum {R_i}^2 - \left(\sum R_i \right)^2 \tag{3.14}$$

It should be noted from Eq. 3.12 that the maximum likelihood solution to the source flux is given by a correlation of the recorded data P_i with the response matrix R_i, which is just a level shifted and weighted version of the mask pattern. Thus correlation analysis

is found to be the optimum deconvolution technique for flux estimates (and positions) in the background dominated scenario.

The uncertainty in the estimates of fluxes F and B are determined by considering the propagation of statistical fluctuations from the recorded shadowgram through to the deconvolved flux whether by fitting to the correlation image or by fitting to the raw data. Since the standard deviation σ_z for an arbitrary parameter z is given by the error propagation formula as :

$$\sigma_z{}^2 = \sum_{i=1}^{N} \left(\sigma_i{}^2 \left(\frac{\partial\, z}{\partial\, P_i} \right)^2 \right) \tag{3.15}$$

the resulting error on the source flux estimate is given by :

$$\sigma_F{}^2 = \frac{N\ \sigma_P{}^2}{\Theta} \tag{3.16}$$

and thus the detection significance of the source can be defined as :

$$SNR = \frac{N \sum R_i\, P_i \ - \sum R_i \sum P_i}{(\ N\ \Theta\ \sigma_P{}^2)^{1/2}} \tag{3.17}$$

where the function Θ is defined as above. At this point, it is the effect of the sampling area ratio *only* on the detection significance that is of interest. Thus, assuming all other parameters are equal in Eq. 3.17, i.e. the numerator remains constant for the same detector characteristics but different mask arrangement, the only term dependent on the response of the detector to the mask shadow is the function Θ. Closer inspection of the form of Θ reveals it is in fact equal to $N^2\ R_{rms}^2$ where R_{rms} is the root-mean-square (rms) deviation of the detector response R_i from the mean detector response $\overline{R}$ for that particular source position.

As discussed earlier the detection significance is affected severely at specific source positions in the FOV, producing a SNR modulation across the sky. From the above analysis it is seen that the sensitivity is directly proportional to the rms value of the instrument response in that direction, and so to determine the absolute mean sensitivity for a particular telescope geometry, the rms response for the FCFOV must first be evaluated.

The rms response for a particular pointing direction is derived from the combination of mask element and detector pixel overlaps that occur for that source position. This evaluation can be performed analytically and is done so below for the classical case in which the mask pattern is a square geometry URA with 100 % opaque closed elements and 100 % transparent open elements. The detector plane employs square pixels which are less than or equal to the mask element size. Using the standard definition of the rms value of a function, the detector response rms is given by

$$R_{rms}\,(x,y) = \left(\frac{1}{N} \sum_{p=1}^{N} R_p^2\,(x,y) - \frac{1}{N^2} \left(\sum_{p=1}^{N} R_p\,(x,y) \right)^2 \right)^{1/2} \tag{3.18}$$

where $N = Vn^2$, V is the mask unit pattern order and n is the ratio of mask element/detector pixel size. The second term in Eq. 3.18 is simply the square of the mean transparency of the mask, which for the URA pattern under consideration is $((V+1)/2\,V)^2 \approx 0.25$. Thus to determine the rms function analytically, the form of $\sum R_p^2$, the first term in Eq. 3.18 has to be solved implicitly. This is trivial to solve for the ideal case and leads to an expression for the sensitivity (RS) relative to the ideal geometry :

$$RS = \frac{SNR_T}{SNR_I} = \frac{R_{rms_T}}{R_{rms_I}} = 2\,R_{rms_T} \tag{3.19}$$

where SNR_T and R_{rms_T} are the signal to noise ratios and rms response functions respectively for the particular telescope, T, under consideration. Similarly, for the ideal telescope geometry, I. Note that R_{rms} reduces to 0.5 in the ideal case. The problem of evaluating R_{rms_T} is simplified by expressing the sum of all pixel responses as a combination of the set of responses directly beneath a particular mask element and the sum of this resultant over all mask elements in the unit pattern :

$$\sum_{p=1}^{N} R_p^2\,(x,y) = \sum_{k=1}^{V_k} \sum_{l=1}^{V_l} \left(\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} R_{k,l,i,j}^2 \right) \tag{3.20}$$

Figure 3.13 shows graphically the pixel to pixel responses for a single mask element. In the geometry considered here, the sampling ratio is of integer value and the source is off-axis by angles β_x and β_y. For an integer sampling ratio it is evident from Fig. 3.13,

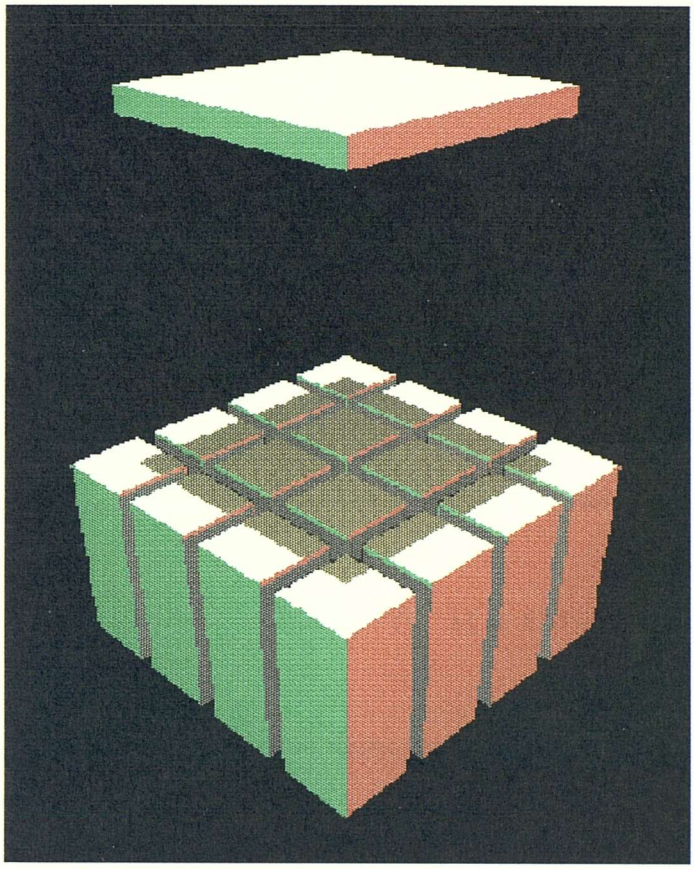


Figure 3.13: The sampling geometry under consideration.

that the response beneath a particular mask element is composed of 4 regions : the pixel R_c whose response is a combination of 4 different mask elements; the set of pixels R_{ex} and R_{ey} with a 2 mask element combination which occur $n_y - 1$ and $n_x - 1$ times respectively where n_x and n_y are the linear sampling ratios in the x and y directions; the set of $(n_x - 1)(n_y - 1)$ pixels R_m with a single mask element response. Thus the response of the subset of detector pixels below the k, l th mask element is given by :

$$R^2_{k,l} = R^2_c + R^2_{ex}\,(n_y - 1) + R^2_{ey}\,(n_x - 1) + R^2_m\,(n_x - 1)\,(n_y - 1)\ , \quad (3.21)$$

$$R^2_c = \left(\frac{\beta_x\,\beta_y}{b_x\,b_y}\,M_{k-1,l-1} + \frac{\beta_x b_y - \beta_x\beta_y}{b_x\,b_y}\,M_{k-1,l} + \frac{\beta_y b_x - \beta_x\beta_y}{b_x\,b_y}\,M_{k,l-1} \right.$$
$$\left. + \frac{b_x b_y - \beta_x b_y - \beta_y b_x + \beta_x\beta_y}{b_x\,b_y}\,M_{k,l} \right)^2 , \quad (3.22)$$

$$R^2_{ex} = \left(\frac{\beta_x}{b_x}\,M_{k-1,l} + \left(1 - \frac{\beta_x}{b_x}\right)\,M_{k,l} \right)^2 , \quad (3.23)$$

$$R_{ey}^2 = \left(\frac{\beta_y}{b_y} M_{k,l-1} + \left(1 - \frac{\beta_y}{b_y}\right) M_{k,l} \right)^2 , \tag{3.24}$$

$$R_m^2 = M_{k,l}^2 . \tag{3.25}$$

When the above expression is simplified and the second sum of Eq. 3.20 is performed using the following conditions,

$$\sum_{k=1}^{V_k} \sum_{l=1}^{V_l} M_{k1,l1} M_{k2,l2} = \frac{V+1}{2}, \quad if \quad k1 = k2 \quad and \quad l1 = l2 \tag{3.26}$$

$$\frac{V+1}{4}, \quad if \quad k1 \neq k2 \quad or \quad l1 \neq l2, \tag{3.27}$$

an expression for the rms response function dependent only on the position of the source and the telescope sampling ratio is found :

$$R_{rms_T}(x,y) = \frac{1}{2\,n} \left(2x^2 - 2x + n \right)^{1/2} \left(2y^2 - 2y + n \right)^{1/2}. \tag{3.28}$$

The substitution $x = \beta_x/b_x$ and $y = \beta_y/b_y$ has been made and it is assumed that $n_x = n_y = n$. Figure 3.14 shows the values of the relative sensitivity function given in Eq. 3.19 within a single sky pixel for the sampling ratio $n = 1$.

Finding the mean sensitivity across the sky pixel and hence the mean for that sampling ratio involves integrating Eq. 3.28 over a single sky pixel. No simple expression was found as a result, but the shape of the function can be found easily by numerical methods, the results of which, expressed in terms of the relative sensitivity are shown in conjunction with the previous model for square geometries in Fig. 3.15.

3.4.2 Discussion

The behaviour of a coded aperture telescope in the near ideal situation where the detector positional resolution is at least an order of magnitude smaller in size than the coding elements of the mask, is well documented and relatively well understood. The previous

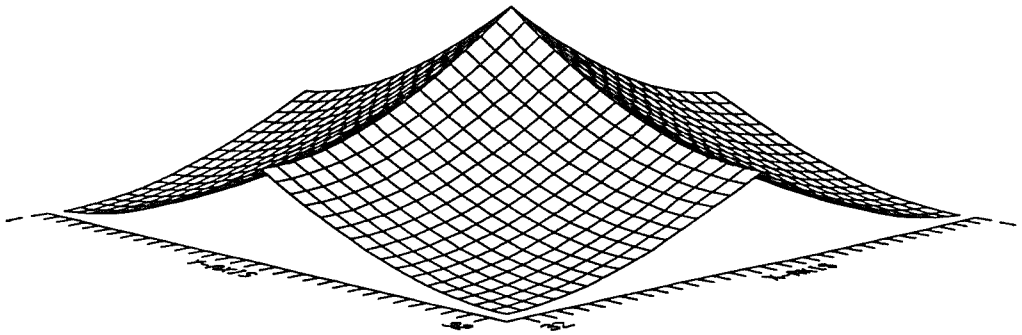


Figure 3.14: Possible detection significances within a sky pixel for a 1:1 sampling ratio. The peak significance, when the shadowgram boundaries are aligned with the detector pixels, is equal to the maximum sensitivity. The lowest significance, in the corner of a sky pixel, is equal to half the maximum sensitivity.

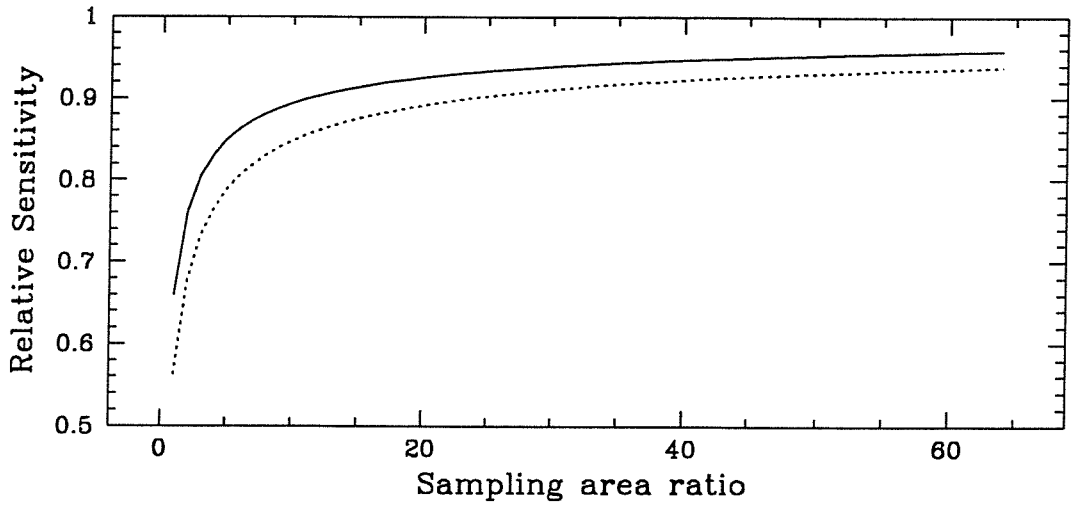


Figure 3.15: Source statistical significance (true SNR) with respect to discrete sampling ratio. For comparison the previous model for the image SNR as defined in Section 3.2.3 is also shown (dotted line).

sections have demonstrated however that careful consideration is required if the potential telescope sensitivity is to be realised. Outlined now is a method of improving the sampling configuration in this situation, providing a more uniform response to all sky positions without significant loss in angular resolution.

3.5 Non-Integer Fine Sampling

A non-integer fine sampling arrangement where the detector pixel repetition is out of phase by a fractional value with the mask element repetition has been investigated on the assumption that such an arrangement will suppress any particular phasing error. For any given source location, only a small proportion of the mask element shadows will be unfavourably matched with the detector pixels, whereas for integer fine sampling the shadow is either completely in phase with the detector sampling, or out of phase at all shadow boundaries. Thus non-integer sampling involves deliberately mismatching the mask element and detector pixel pitch.

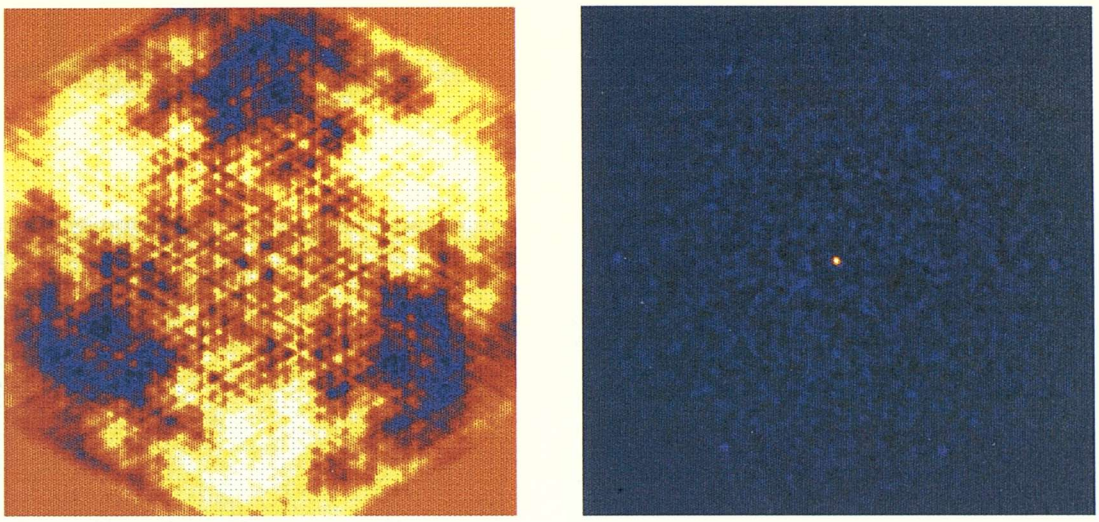
3.5.1 Deconvolution

The immediate consequence of this mismatch is that if a correlation analysis is to be used to obtain the first estimate of the source distribution, both the deconvolution and raw data arrays must be modified so that they are both sampled at the same frequency. Ensuring this provides compatibility for cross-correlation whether performed sequentially or with the use of Fast Fourier Transforms (FFTs).

On modifying the deconvolution array the perfect autocorrelation properties (central peak with perfectly flat side-lobes) of the original coding pattern will be lost. This degradation in the SPSF has been previously highlighted for hexagonal geometries where all sampling configurations bar a few specific orientations (see Section 3.3.1) produce 'noisy' SPSFs [192]. If however it is possible to suppress the coding noise and hence obtain clean correlation images, the mismatched geometry would be favourable with respect to both the response uniformity and the mask alignment since no special orientation would be necessary.

Although the coding noise in the SPSF side-lobes is seemingly insignificant for these 'non-optimum' orientations (largest noise peak is less than 0.1 % of the peak height), the presence of background counting rates which were expected with EIDOS (see Fig. 3.2) severely magnify any SPSF non-uniformities. The outcome is that even for fully coded sources, the coding noise is large enough to render the source undetectable as demonstrated in Fig. 3.16(a). However, as will be seen in Chapter 6, it is possible to suppress the coding noise artifacts with the use of several pre and post-processing

strategies, such as mask-antimask imaging which is used to produce Fig. 3.16(b).



(a) (b)

Figure 3.16: (a) The magnification of coding noise artifacts by the high background counting rate. The image is of a 10 mCrab source with a background counting rate as given by Fig. 3.2, for a 10^5 second integration time. The observational energy range is 15 keV - 10 MeV. (b) Elimination of coding noise artifacts with the use of a mask-antimask strategy.

As far as the uniformity of the response is concerned, non-integer sampling is favourable over integer sampling as can be seen in Fig. 3.17. Although the image SNR remains modulated with source position, the variation is negligible in comparison to the likely background fluctuations thus providing a smoother instrument response across all sky pixel positions than for integer sampling. Since the dominant parameter governing the sensitivity for a pixellated detector configuration is the mask element to detector pixel area ratio, non-integer fine sampling will not improve the overall sensitivity of an instrument, it will however ensure that the correlation map provides a most likely (in the least squares sense) 'a priori' estimation of the source distribution without the potential 'source loss' found with integer sampling.

3.5.2 Conclusions

The trade-off between angular resolution and point source sensitivity has serious implications for the reconstructed image SNR of a source which lies close to sky pixel boundaries, lowering the mean SNR within a sky pixel when the angular resolution is increased. The sensitivity across sky pixel boundaries can be restored to a uniformly high

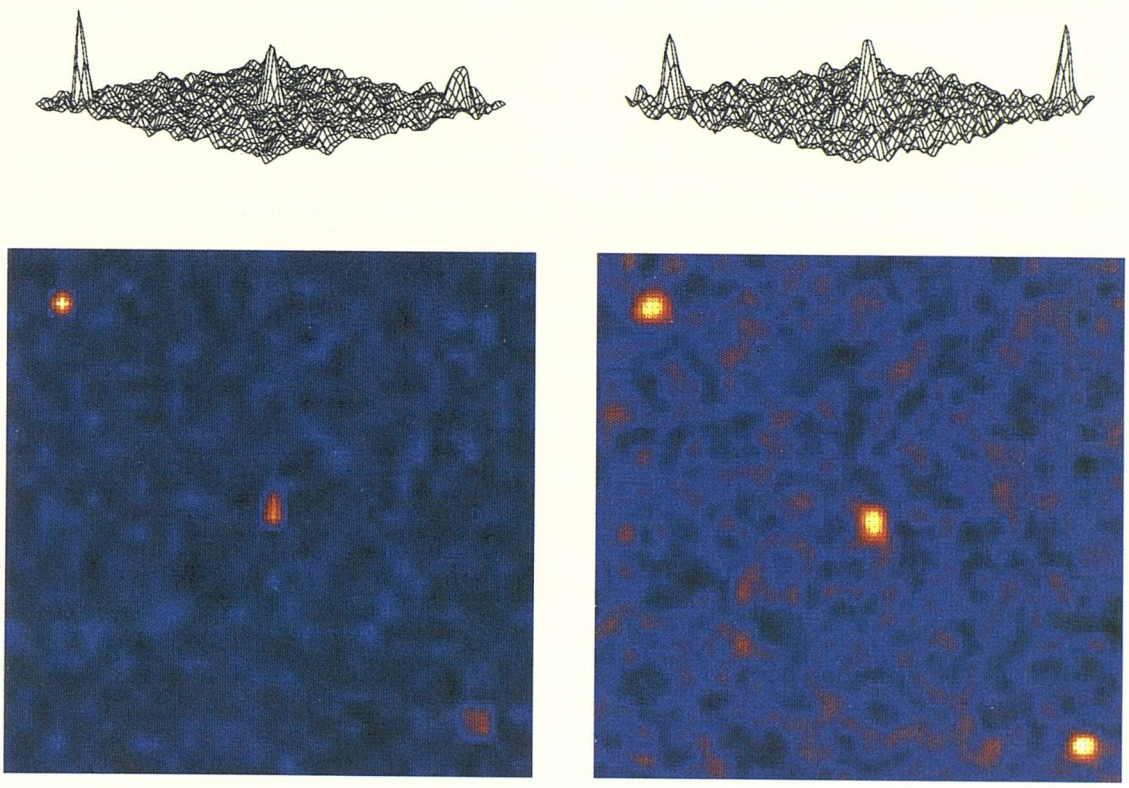


Figure 3.17: The suppression of phasing errors with non-integer fine sampling.

level without drastically compromising the telescope angular resolution with the use of non-integer fine sampling. Ultimately, however, the absolute point source sensitivity is determined by the area ratio between mask elements and detector pixels. In the following section, the first steps are outlined in the development of an imaging sequence for pixellated coded aperture telescopes. The minimum requirements of this imaging sequence, are (1) the suppression of coding noise in reconstructed images, (2) reliable source positioning, and (3) maximum likelihood flux estimates from the raw shadowgram data. In addition, from a processing time perspective, it is preferable to avoid iterative techniques at any stage in the analysis.

3.6 Fast Flux and Spectral Retrieval for Pixellated Coded Aperture Telescopes

For optimally coded systems, based on the use of cyclically employed URAs, correlating the recorded shadowgram with a level-shifted/weighted version of the mask pattern will produce a SPSF with perfectly flat side-lobes. In the ideal case, the only ambiguous

features are ghost sources which arise from the finite extent of the mask pattern and detector plane. The strength of the ghost sources is dependent on the extent to which the pattern is repeated.

It was demonstrated in Section 3.4 that correlation reconstruction not only produces images free from artifacts in the ideal case, but, in the weak source limit where the variances of each detector pixel count are equal and (virtually) uncorrelated, images that are most likely in the least squares sense. As this will nearly always be the case for EIDOS, correlation reconstruction may still provide a reliable and fast option for extracting source fluxes and spectra, even in the light of recent advances in Maximum Entropy reconstruction techniques. If iterative techniques are to be avoided, what are the options for direct image reconstruction ?

3.6.1 Alternative Methods - A Brief Review

For mask patterns with perfect imaging properties, as defined by a δ-function autocorrelation function (ACF) and consequently a perfectly flat modulation transfer function (MTF), correlating the ideal shadowgram with the mask pattern is identical to multiplying the data by the generalised inverse of the mask pattern/response matrix or applying an Optimal (Weiner) fllter to the data (see Chapter 2, Section 2.6).

In the ideal case, which for matrix inversion requires the number of sky pixels to be approximately equal to the number of data points, and for Weiner filtering requires that there is no loss of data due to undersampling of the shadowgram, both methods are perfectly valid. For EIDOS, the collimator FOV extends to cover the PCFOV, which consequently gives around 8 times as many sky pixels as detector elements - clearly an underdetermined problem !, which even if the inverse matrix could be calculated, would produce too large a solution set to be quantifiable. Discrete filtering using a Weiner filter compensates only for imperfections in the mask pattern, which for EIDOS is not too far removed from ideal. Detector support structure and/or dead pixels are a source of information loss which may modify the system response to such an extent that a true inverse filter cannot be found. However, as mentioned in Chapter 2, Section 2.6.2, there does exist a powerful technique for identifying the components in the response matrix responsible for singularities and produces a stable fit to the parameters without noise amplification. The technique is known as Singular Value Decomposition (SVD) and will

be discussed further in the next section.

3.6.2 Cross-correlation and Least Squares Analysis

3.6.2.1 'Quick-look' Correlation

The third and most commonly used non-iterative alternative is correlation analysis, the reliability of which has already been demonstrated from a least squares fit perspective in Section 3.4. The simplicity of this deconvolution method combined with the fact that the correlation operation can be performed with ease in the Fourier domain with the use of FFT algorithms, makes correlation analysis a viable option with a distinct process time advantage over iterative techniques. As with matrix inversion and discrete filtering, a correlation map will be susceptible to coding noise - the collective term for ghost sources and side-lobe artifacts.

Possible sources of coding noise are discussed in detail in Chapter 6, Section 6.4, specifically for EIDOS, which has led to the development of contingency plans to maintain optimum performance for a range of severe observation scenarios. However, it is sufficient to say that it *is* possible to produce correlation images from which reliable source positions can be identified prior to a more rigorous analysis to retrieve flux estimates and spectra.

3.6.2.2 Maximum Likelihood Fitting

It has already been suggested that matrix inversion for the entire FOV is not a viable option because of the underdetermined nature of the problem. However, solving the source distribution for a limited number of point sources, or for a an extended source region *is* viable for sources identified in the 'quick-look' correlation map. The problem can be expressed as :

$$\begin{pmatrix} \\ R \\ \\ \end{pmatrix} \begin{pmatrix} \\ F \\ \\ \end{pmatrix} = \begin{pmatrix} \\ P \\ \\ \end{pmatrix}$$

where R is now the response matrix covering only those sky positions identified in the correlation image, F is the source solution vector, and P is the observed data set. The solution to this type of linear problem can be found using a technique known as singular value decomposition (SVD) [196]. Although SVD is just one of many techniques for solving linear problems, it is a powerful technique for dealing with matrices that are either singular or numerically very close to singular, precisely the situation for EIDOS. SVD methods for solving inverse problems are based on the following theorem of linear algebra :

Any $M \times N$ matrix whose number of rows M is greater than or equal to its number of columns N can be written as the product of an $M \times N$ column-orthogonal matrix U, an $N \times N$ diagonal matrix W with positive or zero elements, and the transpose of an $N \times N$ orthogonal matrix V. Thus, expressed in matrix form :

$$\begin{pmatrix} \\ R \\ \\ \end{pmatrix} = \begin{pmatrix} \\ U \\ \\ \end{pmatrix} \begin{pmatrix} w_1 & & \\ & w_2 & \\ & & w_N \end{pmatrix} \begin{pmatrix} \\ V^T \\ \\ \end{pmatrix}$$

Since both U and V are orthogonal, their inverses are trivial to calculate and are simply equal to their transposes. Similarly since W is a diagonal matrix, it has an inverse that is equally simple to calculate, and is given by the diagonal matrix with elements that are the reciprocal of the values w_j. Thus for the overdetermined set of simultaneous equations being considered here, in theory it should be possible to construct the solution vector x using the inverse of R as follows :

$$\begin{pmatrix} \\ F \\ \\ \end{pmatrix} = \begin{pmatrix} \\ V \\ \\ \end{pmatrix} \begin{pmatrix} 1/w_1 & & \\ & 1/w_2 & \\ & & 1/w_N \end{pmatrix} \begin{pmatrix} \\ U^T \\ \\ \end{pmatrix} \begin{pmatrix} \\ P \\ \\ \end{pmatrix}$$

The only possible problem with this construction is identical to the problem faced by *every* inverse matrix solution, namely singularities, caused by either underdetermined or overdetermined and degenerate experimental scenarios. These singularities manifest within W as zero or near zero (to within the processor roundoff errors) elements w_j. Thus fortuitously, the matrix W provides a diagnosis of the singularity of R. If any of the w_j's equal zero, then R is said to be *singular.* If any of the w_j's are zero to within roundoff errors, then R is said to be *ill-conditioned.*

The SVD technique therefore provides a means of identifying and eliminating both types of singularity, which can be done so in a strictly least squares sense. The important factor making SVD favourable in either the singular or ill-conditioned scenario, coupled with the consequences of 'noisy' data in which P is moved outside the range of R, is that in dealing with the singularities, a solution vector F is always found which minimises the residual :

$$\chi^2 = | R.F - P |^2$$

In the following chapters, unless stated otherwise, it can be assumed that the data analysis of all simulated observations has been performed using the SVD approach.

3.7 Conclusions

In this chapter, the point source capability of coded aperture telescopes based on pixellated detector technology has been objectively quantified. This has led to a thorough understanding of the trade-off between point source sensitivity and angular resolution, and consequently the sampling requirements to ensure optimum use of the available detector

area. It has also been shown that, regardless of the sampling geometry ultimately used, correlation analysis produces images that are reliable from a least squares perspective. Although practical observing conditions may severely degrade the correlation image to the extent that flux estimates can no longer be obtained directly from this image, it will be demonstrated in Chapter 6 that source positions can still be identified accurately from the correlation image. As a result, the final sections of this chapter have outlined a possible technique for extracting flux estimates from the raw shadowgram data.

Chapters 5 and 6 examine the primary stages of deconvolution as applied to EIDOS. This includes the reconstruction of the raw, multiple-site events to determine the pixel of incidence and hence produce the initial shadowgram for a particular energy bin. From there the first correlation map is produced, and if necessary the appropriate contingency plans are applied to suppress the effects of coding noise. Sufficiently cleaned of spurious sources and artifacts, the correlation map provides source positions for which the maximum likelihood flux and detection significance can be determined with ease using the SVD technique.

However, since this chapter has concentrated on the point source capability of a pixellated coded aperture telescope, it now seems logical to examine the multi-scale imaging capability of such a telescope. The following chapter is therefore concerned with extended source sensitivity and techniques for imaging both extended and point sources simultaneously.

Chapter 4

Coded Aperture Imaging on Multiple Angular Scales

Overview

In this chapter, the performance of the coded aperture technique has been studied in order to quantify objectively the sensitivity to extended structures. A coded aperture telescope is demonstrated to be relatively insensitive to extended structure on scales greater than the angular size of the SPSF. By examining the reconstruction of an extended source by a simple correlation technique, theoretical predictions for the loss in sensitivity have been made and are verified by the results of Monte Carlo simulations. The large-scale sensitivity of a particular telescope configuration is found to be improved by employing either a coarser mask pattern (poorer angular resolution), or a lower throughput mask pattern (open fraction less than 0.5). With the aim of improving the sensitivity to extended structure without compromising the point source capabilities, both a mathematical and a technological solution are investigated. The mathematical solution involves the use of a new multi-scale image processing technique, namely Wavelet Transform analysis, whereas the technological solution involves the use of a new technique developed specifically for multi-scale imaging, namely the Composite Mask technique.

4.1 Introduction

In the previous chapter, the performance of a coded aperture telescope based on pixellated detector technology has been objectively quantified. As outlined in Chapters 1 and 2, the motivation behind the development of future gamma-ray telescopes has been the necessity to maximise sensitivity, energy range and energy resolution, field of view (FOV) and of most relevance to this chapter, angular resolution. As with imaging in other wave-bands, the goal in gamma-ray astronomy has been to drive the angular resolution of instruments ever smaller with the result that the next generation of telescopes such as the INTEGRAL imager, will have unprecedented angular resolutions. This improvement in angular resolution is essential if the nature of gamma-ray sources, for example in regions of overcrowding such as the Galactic Centre, is to be understood. The spectacular angular and energy resolution of the INTEGRAL imager/spectrometer combination over the 300 keV to 1 MeV band should finally tie down the source(s) of para-positronium (511 keV) and ortho-positronium (continuum) emission from the Galactic Centre region. However, the quest for improved point source imaging should not always be the absolute goal since an increase in angular resolution, as will be shown in this chapter, is accompanied by a decrease in the sensitivity to extended sources.

A particular telescope baseline (angular resolution, number of open elements in the mask pattern) is sensitive only to a narrow range of source angular scales, beyond which the sensitivity is severely compromised. A telescope configured for high angular resolution imaging will be relatively insensitive to large scale structure. Conversely, by employing a mask pattern with a larger element size, the extended structure sensitivity can be improved but only at the expense of the point source capability. Although the relationship between angular resolution and extended source sensitivity has been implied and acted upon several times [197, 198], no objective analysis has been performed. In the following sections it is shown that for optimised sensitivity the angular resolution of the telescope must be greater than or equal to the angular extent of the proposed target source.

4.2 Extended Source Sensitivity of the URA

4.2.1 Theoretical Overview

As discussed in Section 3.2.2, a space-borne coded aperture telescope operating in the 15 keV to 10 MeV energy range may be subject to detector background fluxes that dominate the source flux by a few orders of magnitude [148]. It is well established [128] that in the weak source limit, optimum sensitivity is achieved by maximising the open fraction of the mask pattern. For optimally coded systems based on the use of URA mask patterns, the optimum open fraction tends towards 0.5 (equal numbers of open and closed mask elements). A combination of both sensitivity and angular resolution requirements has resulted in the widespread implementation of high order URAs for both space and balloon-borne coded aperture telescopes. For a sparsely populated FOV containing point-like objects, the performance of the URA based telescope can reach the statistical limit determined by the observed fluxes alone. However, when observing objects with an angular extent beyond the angular size of the system point spread function (SPSF), the URA based coded aperture telescope begins to lose sensitivity.

The objective here is to make theoretical predictions on the performance of a given telescope geometry to extended emission. To achieve this some initial assumptions have been made about the observation scenario and telescope configuration, and then correlation analysis has been followed to reconstruct the source distribution. First however, the shadowgram formation process is considered for an extended source as this will provide some insight into where the sensitivity loss to large-scale structure begins.

The mask pattern of a coded aperture telescope is effectively a collection of pinholes each projecting an image, in a similar manner to a pinhole camera, onto a position sensitive photon detector. Consider the projected image of an extended source from a mask with a single open element - a pinhole camera. If the extended source has an angular flux distribution given by $S(\theta, \phi)$ and the open area response of the mask element is given by $A(\theta, \phi)$ (as seen by the detector plane), then the contribution to the projected image from the source region $d\theta\ d\phi$ is given by Eq. 4.1.

$$dI_{\theta,\phi} = \frac{1}{A_m}\ S(\theta, \phi)\ A(\theta, \phi)\ d\theta\ d\phi \tag{4.1}$$

Integrating Eq. 4.1 over the entire source distribution results in Eq. 4.2 which describes the projected image distribution, $I(x,y)$, across the detector plane (converting from position angle to detector position coordinates),

$$I(x,y) = \frac{D^2}{A_m} \int_{-X}^{+X} \int_{-Y}^{+Y} S(x,y)\, A(x,y)\, dx\, dy\ , \tag{4.2}$$

where A_m is the mask element area, D is the mask-detector separation and $S(x,y)$ would be the detector image obtained with a 'perfect' pinhole camera. $A(x,y)$ is the aperture response at the detector plane with respect to offset angle from the straight through position, and X and Y describe the projected source size on the detector plane. Thus, a single unresolved source in the FOV will simply project the shape of the open element onto the detector plane. If instead of a single open element mask, a URA pattern is used, there will be multiple projections of the source onto the detector plane. For a point source, the combined projection from all open elements, or the *shadowgram*, is simply a copy of the mask unit pattern to within the positional resolution of the detector and shifted according to the source position. In this case it has been shown (Chapter 3, Section 3.4) that a simple correlation between the shadowgram and a version of the mask pattern is the optimum image reconstruction method from a least squares perspective. But what happens when the source is extended ?

An extended object in the FOV will project an extended image (given by Eq. 4.2) onto the detector plane and do so from every open element of the mask. From a sensitivity perspective, the most important fact to note about the subsequent extended source shadowgram is that the percentage of flux projected *straight through* from each open mask element is *linearly* dependent on the solid angle of an open element. Consequently, the source 'flux' reconstructed at this position in the correlation image is also expected to depend on the open element solid angle.

Figure 4.1 illustrates the dependence of the reconstructed 'flux' on the mask element solid angle for two different telescope configurations imaging the same extended source under identical observing conditions. The general telescope and observing parameters used are given by Table 4.1 in Section 4.3.1. Since both the detector area and observation time used in the simulations was fixed, the actual source flux detected by each telescope was identical. The profile in Fig. 4.1 shows the simulated Gaussian source extended by 2.5 degrees at FWHM. The two mask patterns used to observe the source are then shown.

The fine-scale mask pattern had a 0.24 degree angular resolution, whereas the coarse-scale mask pattern had a 1.2 degree angular resolution. Reconstructing the source distribution by correlation analysis (see Fig. 4.1(d) & (e)) demonstrates that the coarse resolution mask is more sensitive to the extended source than the fine resolution mask.

An important point to note is that the statistical fluctuations in the reconstructed background are *at the same level* in both images and are equal to the Gaussian fluctuations expected from the detected source and background fluxes. If the simulations are performed without any statistical fluctuations on the detected fluxes, as in Fig. 4.2, it becomes apparent where the large-scale sensitivity difference arises. The source 'flux' reconstructed in the coarse resolution image is greater than in the fine resolution image and by a factor equal to the ratio between mask element solid angles. However, any normalisation introduced to ensure the source flux reconstructed per sky pixel is equal in both cases, also results in normalisation of the background noise. Consequently, the ratio of the reconstructed peak height above the surrounding statistical fluctuations will *always* be larger for the coarse mask configuration.

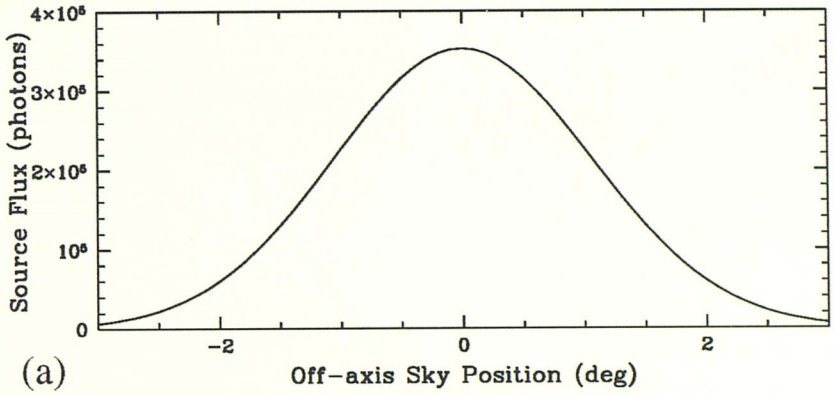
4.2.2 Theoretical Analysis

Since the ratio of the source scale to the angular resolution of the mask pattern appears to be an indication of the telescope sensitivity to that angular scale, a parameter to be known as the f-fraction is now defined in order to quantify theoretically the telescope sensitivity. For a given source extent and mask element size, the f-fraction is the same for *all* open mask elements (neglecting edge effects) and is given in terms of integrations over the open element size (OE) and the projected source size (PS) according to Eq. 4.3.

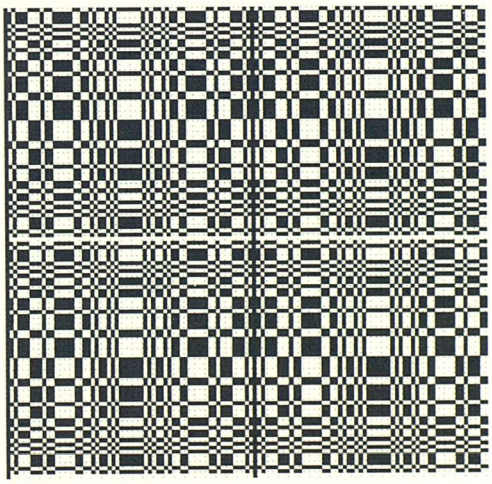
$$f = \frac{\int\int_{OE} I(x,y)\, dx\, dy}{\int\int_{PS} I(x,y)\, dx\, dy} \quad . \tag{4.3}$$

Thus, f is simply the ratio between the source flux projected at the 'straight through' position (with respect to the source centroid) and the total source flux incident on the mask element. Figure 4.3 shows the f-fraction as a function of source size for the mask patterns in shown Figs. 4.1(b) & (c) and for a circular source with a uniform surface brightness (constant flux $\mathrm{cm}^{-2}\ \mathrm{s}^{-1}\ \mathrm{deg}^{-2}$).

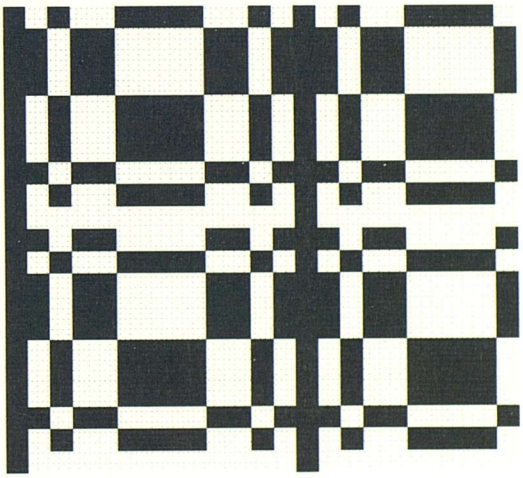
In order to link the f-fraction with a measure of the telescope sensitivity, the process



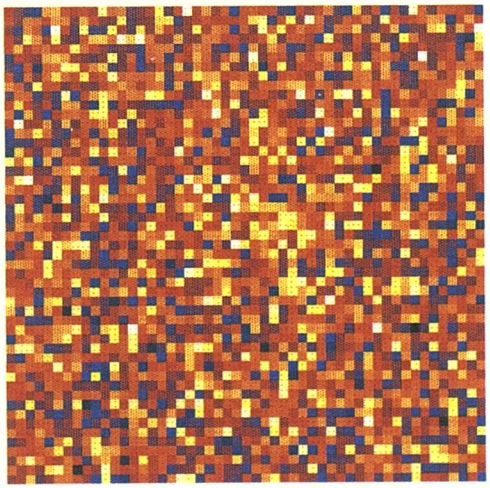
(a)



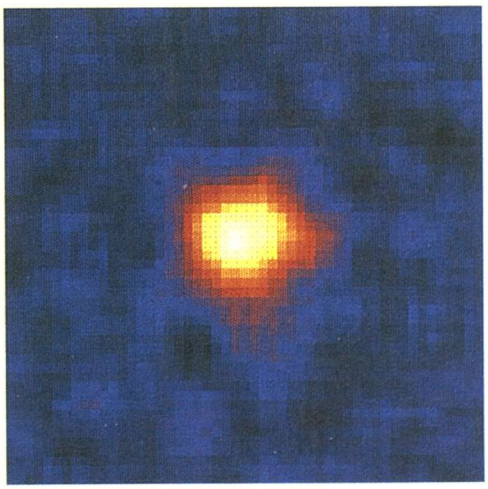
(b)



(c)



(d)



(e)

Figure 4.1: (a) The simulated source profile. (b & c) The fine and coarse mask patterns used to image the source. (d & e) The correlation reconstructions for fine and coarse telescope geometries.

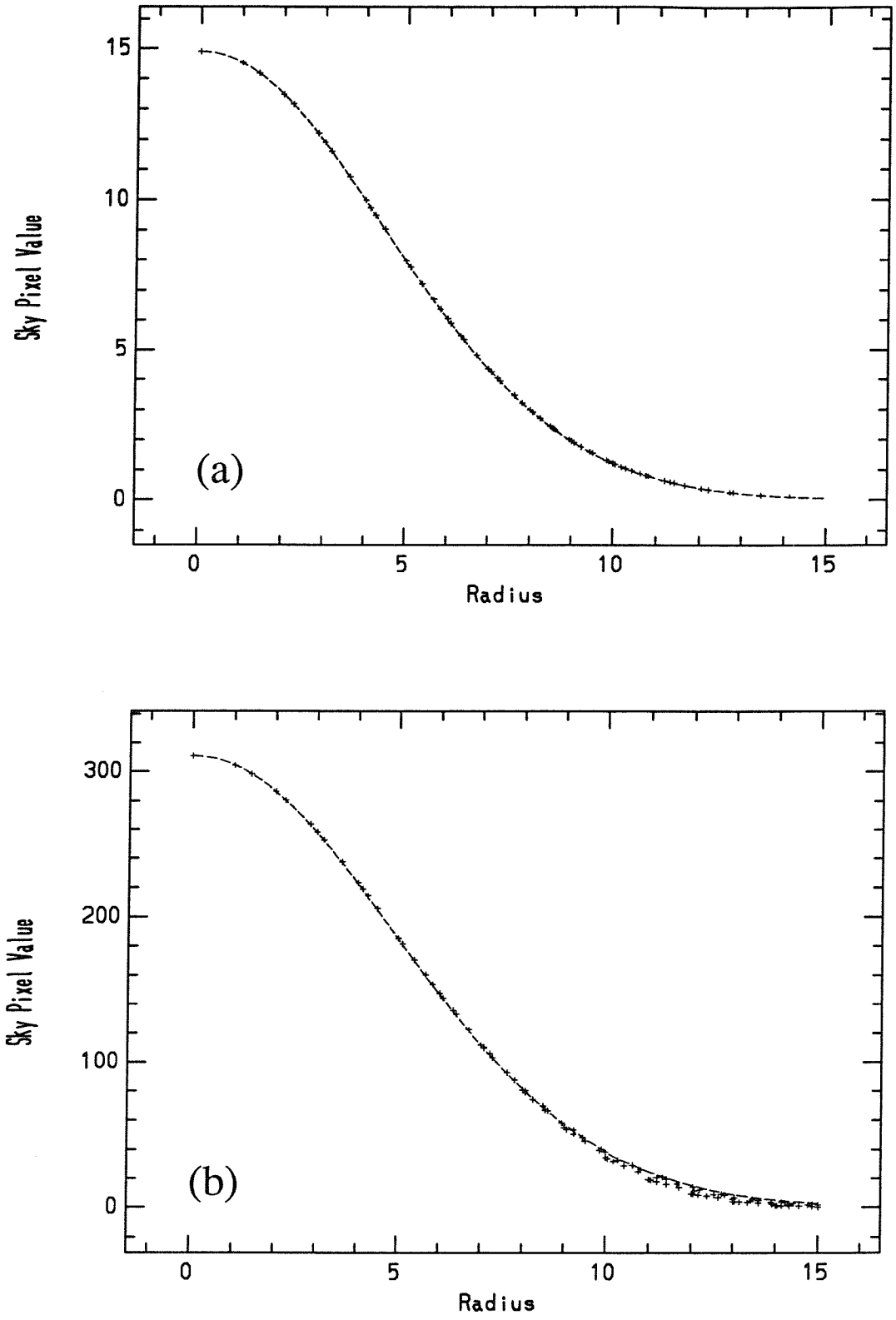


Figure 4.2: The reconstructed profiles for non-statistical simulated observations. (a) The fine-scale reconstructed profile. (b) The coarse-scale reconstructed profile.

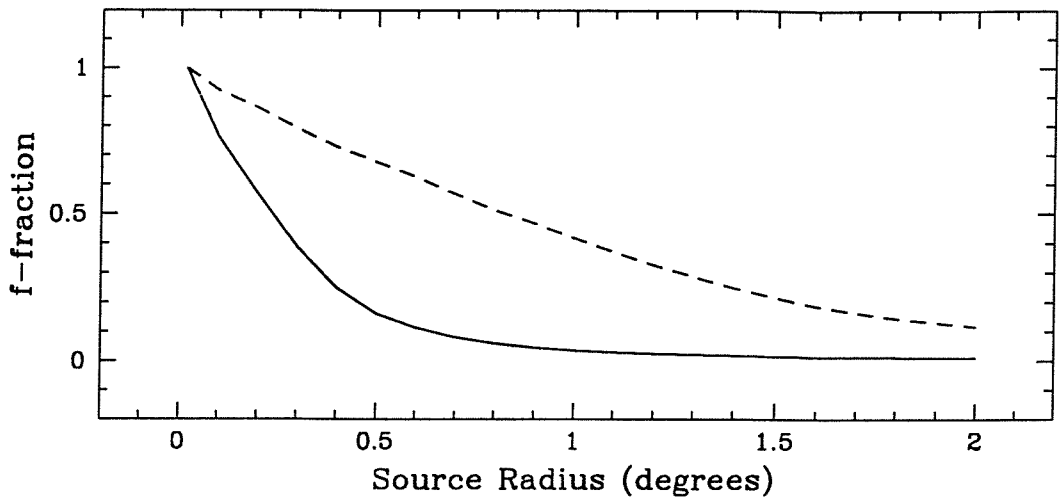


Figure 4.3: The f-fraction for the mask patterns shown in Fig. 4.1 and a circular extended source with a constant surface brightness. Solid line denotes the fine mask response, whereas the dotted line denotes the coarse mask response.

of image reconstruction by correlation is now examined for an extended source. The observation scenario under consideration comprised the following components :

1. An extended source region of constant surface brightness, F_Ω photons cm^{-2} s^{-1} deg^{-2}, positioned on-axis, with an extent of Ω_s deg^2.

2. A detector background component, unmodulated by the mask, of b counts per detector pixel per second and an observation time of T_i seconds.

Upon reconstructing the ij th sky pixel (the source centroid pixel), a correlation of the recorded shadowgram with the open mask elements yields the flux :

$$D_{ij,open} = (f\ F_m\ N_o\ +\ N_o\ A_r\ B\ +\ (1-f)\ F_m\ N_o\ t) \quad cts \tag{4.4}$$

where $F_m = F_\Omega\ A_m\ T_i\ \Omega_s$ is the total source flux incident on each open element of the mask and $B = b\ T_i$ is the total background count per pixel per observation. N_o is the number of open elements in the mask pattern, A_r is the detector sampling ratio (mask element area/detector pixel area) and t is the open fraction of the mask. For the same sky pixel, the closed mask element correlation yields the flux :

$$D_{ij,closed} = (N_c A_r B + (1-f) F_m N_o (1-t)) \quad cts \tag{4.5}$$

where N_c is the number of closed elements in the mask pattern. Normalising Eqs. 4.4 and 4.5 with the corresponding open and closed fractions and subtracting Eq. 4.5 from Eq. 4.4 gives the reconstructed source flux for sky pixel ij[1] :

$$D_{ij} = \frac{f F_m N_o}{A_r t} \tag{4.6}$$

The Gaussian noise at sky pixel ij (and that of the entire image) is derived from the statistically independent counts given by Eqs. 4.4 and 4.5 as :

$$\sigma_{ij} = \left(\frac{f F_T}{A_r^2 t^2} + \frac{B_T}{A_r^2 t} + \frac{B_T}{A_r^2 (1-t)} + \frac{(1-f) F_T}{A_r^2 t} + \frac{(1-f) F_T}{A_r^2 (1-t)} \right)^{1/2} \tag{4.7}$$

where $F_T = F_m N_o$ is the total source flux detected over the entire observation, and similarly, $B_T = B N_o A_r/t$ is the total detector background flux for the entire observation. In order to assess image quality, a figure of merit (FOM) for the reconstructed source has been defined as follows :

$$\begin{aligned} FOM &= \frac{D_{ij}}{\sigma_{ij}} \\ &= \frac{f F_T}{\left(f F_T + \frac{(1-f)}{(1-t)} t F_T + \frac{1}{(1-t)} t B_T \right)^{1/2}} \end{aligned} \tag{4.8}$$

It should be noted that this is not intended as a measure of the absolute statistical significance of the extended region, but simply a measure of the 'visibility' of the source extent above the background noise. However, in Section 4.3.2 it is demonstrated that the FOM *is* a reliable indication of the absolute statistical significance of the source. For 50% transparent URAs and for negligible detector background, Eq. 4.8 simplifies as :

$$FOM = f F_T^{1/2} \tag{4.9}$$

[1] A_r is simply the sampling ratio i.e. the number of detector pixels sampling each mask element and only has a scaling effect on the reconstructed image.

Thus for the same detector area, coded aperture telescopes with larger mask element sizes (larger f) will be more sensitive to extended structure, but only at the expense of the angular resolution. However, since the FOM is dependent on both the f-fraction and the open fraction t, Eq. 4.8 also suggests that high angular resolution imaging may be achieved as well as good extended source sensitivity by reducing the mask open fraction, but this time at the expense of the point source sensitivity in the background dominated case. In order to verify Eq. 4.8, Monte Carlo simulations of a coded aperture telescope were performed with both the fine and coarse-scale mask patterns shown in Fig. 4.1 for a range of extended source sizes. The results of these simulations are the subject of the following section.

4.3 Simulated Extended Source Imaging

4.3.1 Correlation Image Reconstruction

The simulated telescope geometry comprised of a pixellated detector array with a 2500 cm^2 sensitive area, a 0.8 cm pixel size (e.g. CdTe solid state detectors and/or CsI(Tl) crystal-photodiode detectors) and was sensitive over the 15 keV to 10 MeV energy range. The coded masks were positioned 200 cm above the detector. The simulated observation comprised of an on-axis, circular extended structure with a uniform intensity profile (i.e. flux per unit solid angle is radially constant) and an unmodulated detector background component commensurate with a CsI(Tl)-photodiode array over the 15 keV – 10 MeV energy range. The telescope and observational parameters used for the simulations are summarized in Table 4.1. The source diameter was increased from zero to 2.5 degrees, at which point it was approximately an order of magnitude greater than the fine mask angular resolution. The FOM was calculated from the correlation reconstruction of each source.

The resulting FOM curve (normalised with the ideal FOM) is plotted as a function of the ratio between source diameter and telescope SPSF (FWHM) in Fig. 4.4 for both fine and coarse-scale mask patterns. The solid curve in Fig. 4.4 is the theoretical sensitivity as predicted by Eq. 4.8. Both fine and coarse-scale data-sets are seen to fit the curve well and thus verify the validity of Eq. 4.8.

	COARSE MASK	FINE MASK
Mask type	11x13	59x61
Detector area	2500 cm^2	
Detector pixel size	0.834 × 0.834 cm	
Mask-detector sep.	200 cm	
Angular Resolution	1.2 degrees	0.24 degrees
Source flux	10^{-3} ph cm^{-2} s^{-1} deg^{-2}	
Energy range	0.1-10 MeV	
Observation time	10^5 seconds	

Table 4.1: Telescope and observation parameters used in the extended source sensitivity simulations. Twin-prime URA masks were used in both configurations.

4.3.2 Maximum Likelihood Image Reconstruction

A true estimate of the telescope sensitivity to an extended structure should include all reconstructed source pixels in the calculation, whereas the FOM estimate only includes the central pixel of the source. Therefore, in order to provide a more meaningful estimate of sensitivity, a maximum likelihood fitting (MLF) procedure between a model of the source distribution and the observed data was adopted. Figure 4.1 has demonstrated that the correlation image is severely compromised if the source is extended beyond the telescope SPSF. However, if there exists some technique for extracting morphological information regardless of the telescope resolution, it might reasonably be assumed that in the ideal case, fitting the extracted source profile with the recorded data would reveal a source detection significance that is independent of the telescope resolution. Thus, the question is, would a different conclusion be reached with respect to the multi-scale sensitivity of the coded aperture technique if the *all* the source information is included in the sensitivity assessment ?

To answer this question, the MLF procedure was taken a step further and the observed fluxes (the raw shadowgram data) were fitted directly with a model shadowgram derived from the *true* source profile itself. Using the same telescope configurations as in the previous simulations, the detection significance of a circular extended source with a constant total flux was measured. Note however, that for these simulations a more realistic

Gaussian radial intensity profile and a constant total source flux (as opposed to a constant flux per unit solid angle) was used. If the predictions of Eq. 4.8 are fundamental to the coded aperture technique, then the maximum likelihood source significance for the coarse-scale mask is expected to be higher than for the fine-scale mask for a given source extent. Figure 4.5 shows the source detection significance as a function of source diameter for both fine and coarse-scale masks, and for both FOM and MLF measurement techniques.

The important point to note from Fig. 4.5 is that even by fitting the recorded data with the *known* source profile, a fine-scale observation of an extended structure is *less* significant than the equivalent coarse-scale observation. The FOM estimation of telescope sensitivity is thus a meaningful indication of the true detection significance.

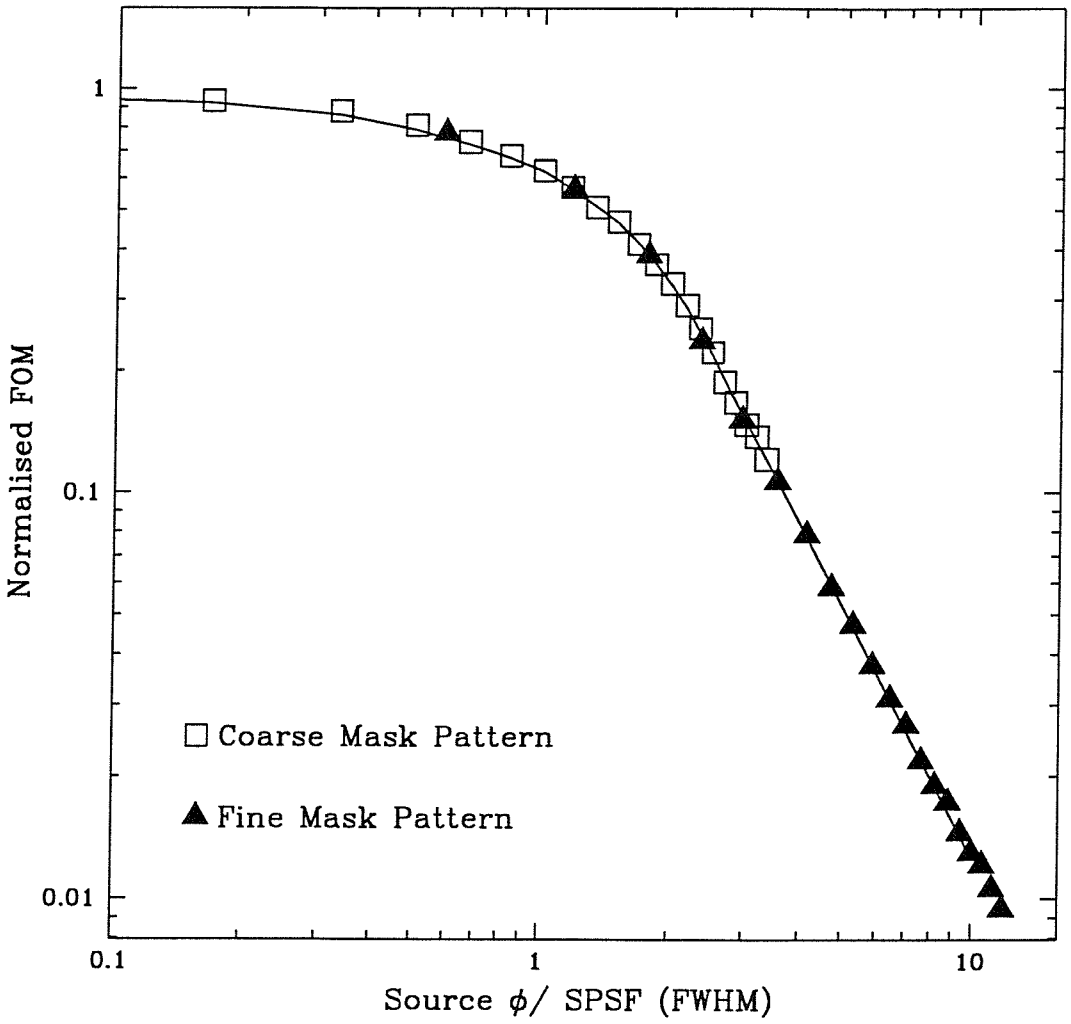


Figure 4.4: The sensitivity (FOM) dependence on the source diameter to SPSF (FWHM) ratio for the fine resolution mask (triangles) and the coarse resolution mask (squares). The solid curve is taken from Eq. 4.8.

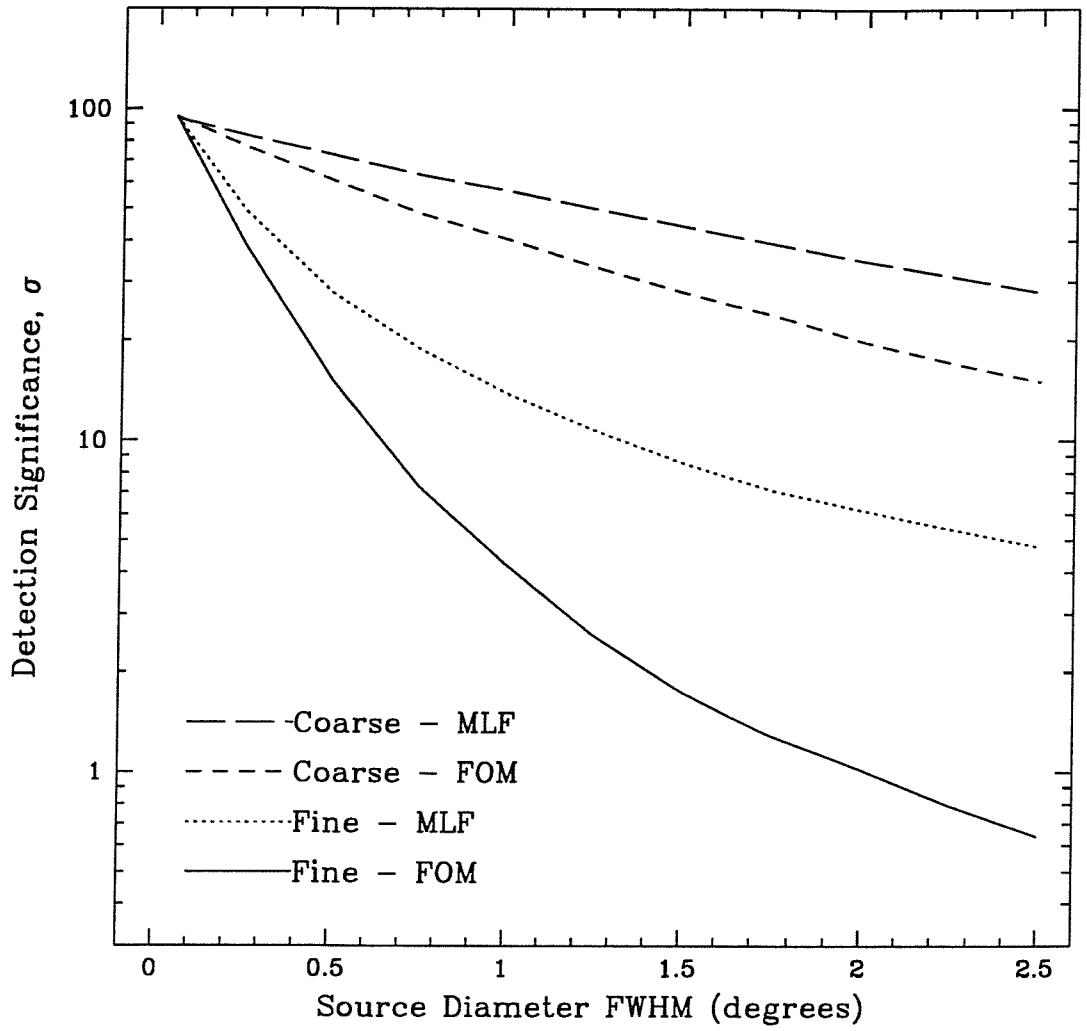


Figure 4.5: Comparison of the FOM and MLF sensitivity estimates for a constant flux, Gaussian profile extended source.

4.3.3 Discussion

When imaging extended structure with a telescope optimised for point source imaging, sensitivity is severely compromised if the source scale exceeds the angular resolution. A mask pattern with larger mask elements will partially restore the extended source sensitivity but at the expense of the angular resolution and hence the point source capability. It has been demonstrated that a coarse-scale mask pattern (broad SPSF) has greater success at modulating the lower spatial frequencies of an extended object than a fine-scale mask pattern (fine SPSF). This suggests that the observed data is effectively 'smoothed' by increasing the SPSF width. Thus, in the light of the relationship between the extended source sensitivity and angular extent of the telescope SPSF, the following

section explores the use of a relatively new multi-scale 'smoothing' technique, namely wavelet transform analysis, for extracting extended structure on several angular scales. Since the problem in the case of the correlation image is one of source identification and characterisation, the objective is to retrieve morphological information on extended structure from high resolution images that would otherwise reveal only point like objects.

4.4 Wavelet Transform Analysis

Originally developed for the analysis of seismic data [199], the wavelet transform analysis technique has since been successfully applied to a diverse range of fields in physics : the study of fractal structures [200] ; the investigation of clustering and sub-clustering of galaxy distributions in Astronomy (see [201] and references therein). More recently, wavelet transform analysis has found a way into coded aperture imaging as a multi-scale flat-fielding technique for telescopes without a mask-antimask capability [202]. In this section the wavelet transform analysis technique is introduced for the identification of extended structure in coded aperture images.

4.4.1 The Wavelet Transform

The complete wavelet transform of a coded aperture image is a decomposition of the image onto a contiguous set of wavelet 'images', each with a discrete characteristic scale size. The wavelet image at a particular scale size, a, can be defined by a correlation between the image, $i(x,y)$, and the analysing wavelet $\omega(\frac{\alpha_x}{a}, \frac{\alpha_y}{a})$. Thus, at a particular measurement position (α_x, α_y) the wavelet coefficient at scale a is given by the correlation integral :

$$w(\alpha_x, \alpha_y, a) = \frac{1}{a} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} i(x,y)\, \omega \left(\frac{\alpha_x}{a}, \frac{\alpha_y}{a} \right) dx\, dy \quad , \tag{4.10}$$

where ω in this case is a real isotropic analysing wavelet. The complete wavelet transform is therefore comprised of a set of wavelet images, each of which is comprised of $m \times n$ wavelet coefficients, where $m \times n$ is the area of interest in the original image $i(x,y)$. Transforming into the Fourier domain we see that the wavelet image at the scale size a, is simply $I(\beta_\nu, \beta_\upsilon)$ filtered with the dyadic passband filters $\Omega(a\,\beta_\nu, a\,\beta_\upsilon)$:

$$W(\beta_\nu, \beta_\upsilon, a) = \; a \; I(\beta_\nu, \beta_\upsilon) \; \Omega(a \; \beta_\nu, a \; \beta_\upsilon) \tag{4.11}$$

where W, I and Ω are the Fourier transforms of the wavelet image, the original image, and the analysing wavelet respectively. From Eq. 4.11, the process of decomposition may appear to be just the iterative application of translated and dilated filters as in a standard Fourier analysis. There is, however, an important difference between using a 'wavelet' and, for example, the dyadic filter $\Pi(\beta_\nu, \beta_\upsilon)$ often used in Fourier analysis, for the decomposition. A correctly chosen 'wavelet' will suppress any Gibbs' or $sinc(\alpha_x, \alpha_y)$ patterns in real space that occur when using filters that have discontinuities in Fourier space. Additionally, if the scale factor between each wavelet image is chosen correctly, a contiguous spectral band is extracted and no information is lost. Consequently, the original image, $i(\alpha_x, \alpha_y)$, can be reconstructed entirely from its set of wavelet images. The multi-scale analysis ability of the wavelet transform arises from the fact that the wavelet coefficient at position (α_x, α_y) provides a description of the image surrounding (α_x, α_y) in image space and $(\frac{\beta_\nu}{a}, \frac{\beta_\upsilon}{a})$ in Fourier space for the scale size a. The wavelet coefficient for the centroid of an extended structure will only become 'significant' when the analysing scale size is of the order of the source scale size.

4.4.2 The Wavelet Method for Coded Aperture Images

Many functions exist that satisfy the equivalent of the Parseval theorem for Wavelet Transforms and can as a result be classified as wavelets. The so called Mexican Hat wavelet given by Eq. 4.12 and shown in Fig. 4.6, was the analysing wavelet chosen for this work mainly because of its ability to remove flat components and gradients (the integral of $\omega(\frac{\alpha_x}{a}, \frac{\alpha_y}{a})$ is zero), and because it had already been successful in quantifying extended structure in X-ray images [201]. The Mexican Hat wavelet is derived from the second derivative of a Gaussian function.

$$\omega \left(\frac{\alpha_x}{a}, \frac{\alpha_y}{a} \right) = \left(2 - \frac{r^2}{a^2} \right) e^{-\frac{r^2}{2a^2}} \quad , \tag{4.12}$$

where $r = x^2 + y^2$.

In performing the discrete wavelet analysis, several additional factors must be considered

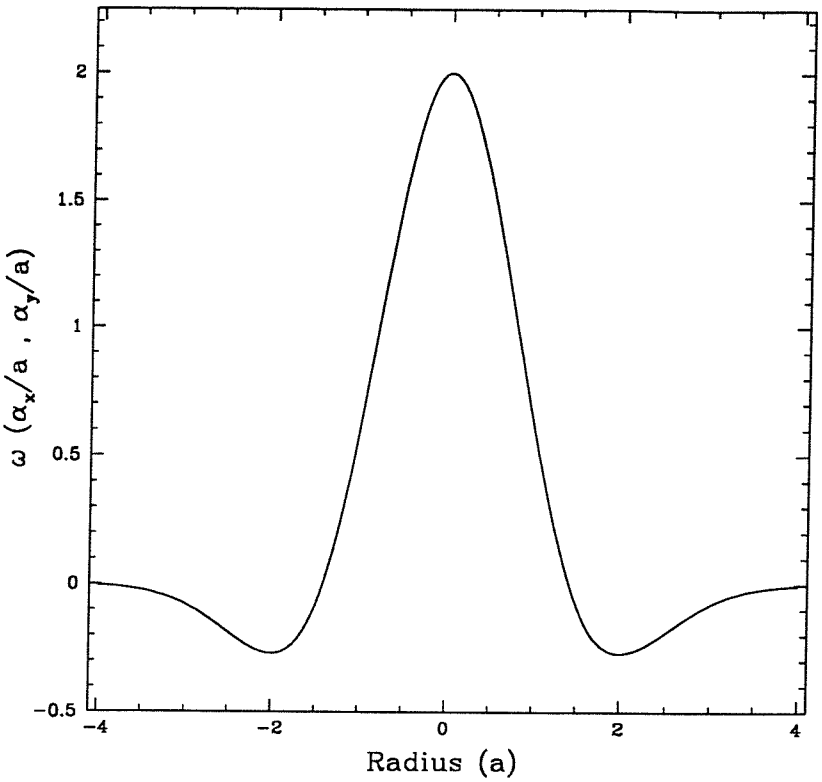


Figure 4.6: The so called Mexican Hat wavelet.

: the Mexican Hat wavelet can only be truncated at $r \approx 4\ a$, implying that the region of interest must be at least 8 times the analysing scale-size in dimensions to avoid edge effects ; to ensure artifacts are not introduced at small analysing scales, the SPSF (and hence the wavelet) must be well sampled by the sky pixel grid ; when using the Mexican Hat wavelet, complete sampling in wavelet space can be achieved with a factor of $\approx \sqrt{2}$ between analysing scales [203]. The task of identifying which wavelet coefficients are significant at each analysing scale however, is not trivial. Producing a significance map at each scale is complicated by the fact that the wavelet image distribution is not Gaussian, even if the original data is. Several methods have been proposed to assess significance levels, which include the extensive use of Monte Carlo modelling of random fields [204], and the use of a significance parameter, Q, defined as the ratio of the wavelet coefficient of the image to the wavelet coefficient of the image noise [205]. The image noise is derived by assuming that each sky pixel is statistically independent and Gaussian distributed, thus the error for the ijth sky pixel containing F counts is assumed to be $\sqrt{F}$.

The important difference between optically focused and coded aperture images lies in the fact that each sky pixel of a coded aperture image is an ensemble of the same set of

fluxes recorded by the detector plane – Gaussian fluctuations from one particular source in the FOV are induced at every other point in the image. Thus, the Gaussian error on a sky pixel count is the same for all sky pixels, regardless of the reconstructed flux in each pixel. Consequently, the wavelet coefficient significance can not be determined from the wavelet transform of the image noise alone. Since computing overheads will already be considerable for image reconstruction in many (a few hundred) spectral and/or time bins, it is undesirable to increase overheads further with extensive Monte Carlo simulations of random fields to calibrate absolute significance levels. As a result of these complications, in the following section a method is described for determining significance levels of wavelet coefficients from the wavelet image itself.

4.4.3 The Application to Simulated Coded Aperture Images

The telescope geometry used in the following simulations is similar to the fine angular resolution geometry outlined in Table 4.1. The only modification is a reduction in the detector pixel size by a factor of 4 to enable reasonable sampling of the wavelet at small analysing scales. The fully coded FOV of the arrangement was $\sim 14^o \times 14^o$, sampled by 232×232 sky pixels. In Section 4.3.1 it was shown that for the fine resolution geometry there are no significant structures visible in the correlation image once the source extent is above $\sim$1 degree. Thus, the objective is to quantify source morphology for sources extended beyond the telescope SPSF. The observational parameters were as outlined in Section 4.3.1 and the source FWHM was again incremented from zero to 2.5 degrees.

For each source, wavelet coefficients were obtained for 10 analysing scales corresponding to $a \approx 0.12$, 0.18, 0.24, 0.36, 0.48, 0.72, 0.96, 1.40, 1.91, and 2.87^o. In order to produce wavelet images at the largest scale sizes, it was necessary to increase our region of interest and include, in addition, a percentage of the partially coded FOV (PCFOV). The PCFOV is inherently noisier with respect to background magnified coding noise and to such an extent that in these simulations it was necessary to employ the plateau removal technique (see Chapter 6, Section 6.2.2) prior to deconvolution to suppress this noise. With the PCFOV extension, the region of interest was increased to 432×432 sky pixels. However, at the largest analysing scales the wavelet image was still disrupted by edge effects, the importance of which will become apparent in the next section.

The Gaussian profiles of both the source and the SPSF, when combined produce a

Gaussian reconstructed source profile with a width given by the true source and SPSF widths added in quadrature. On this assumption, it possible to perform analytically the correlation given by Eq. 4.10 for a Gaussian source profile. Since both the analysing wavelet and the theoretical source profile are real and even functions, the correlation integral shown in Eq. 4.10 is equivalent to the convolution integral and an analytical expression for the theoretical wavelet coefficients at the source centre can be derived :

$$w(\alpha_x, \alpha_y, a) = \frac{2\,I}{a}\left(1 + \frac{\sigma^2}{a^2}\right)^{-2} , \tag{4.13}$$

where I is the total source flux and σ is the source width.

Since the wavelet image noise distribution, as shown in Fig. 4.7 was only approximately Gaussian, 'significant' wavelet coefficients were defined at a high threshold level of $4\sigma_w$, where σ_w is the deviation of the wavelet image background noise as derived from the wavelet image itself. However, because the noise distribution was Gaussian at the 10% level, it was possible to use the standard χ^2-minimisation technique for parameter and parameter confidence level estimation [206] for Eq. 4.13. On the contrary, at the largest analysing scales edge effects produced a wavelet noise distribution that was far from Gaussian and rigorous simulations were required to verify that the wavelet image at these scales was still reliable for source identification. When fitting Eq. 4.13 to the wavelet coefficient data, the interest was focused mainly on the source extent σ since the source flux I can be derived with reliable Gaussian errors from the recorded shadowgram using the MLF technique described earlier.

To determine quantitively whether a source was extended or not, a comparison was made between the value of χ^2 for a 2 parameter fit and the value of χ^2 for a 1 parameter fit with the source extent fixed at the SPSF width. Thus, for the 10 scales analysed, both 9 degree of freedom (dof) and 8 dof fits to the data were obtained, where the value $\Delta\chi^2 = 9\chi^2_9 - 8\chi^2_8$ is distributed as χ^2 with 1 dof [206]. Because of the relationship between the reduced χ^2 and the ratio of the parameter S (the fitting variance) to the actual data variance for a 1 dof fit, it follows that $\sqrt{\Delta\chi^2}$ gives the significance of the source being extended in units of the standard deviation of a normal distribution [159]. The results of fitting to wavelet data for extended structures ranging from 0.25 to 2.5 degrees FWHM are given Table 4.2. Figures 4.8 and 4.9 show the χ^2_8-fit between Eq. 4.13 and wavelet data for a few of the sources simulated.

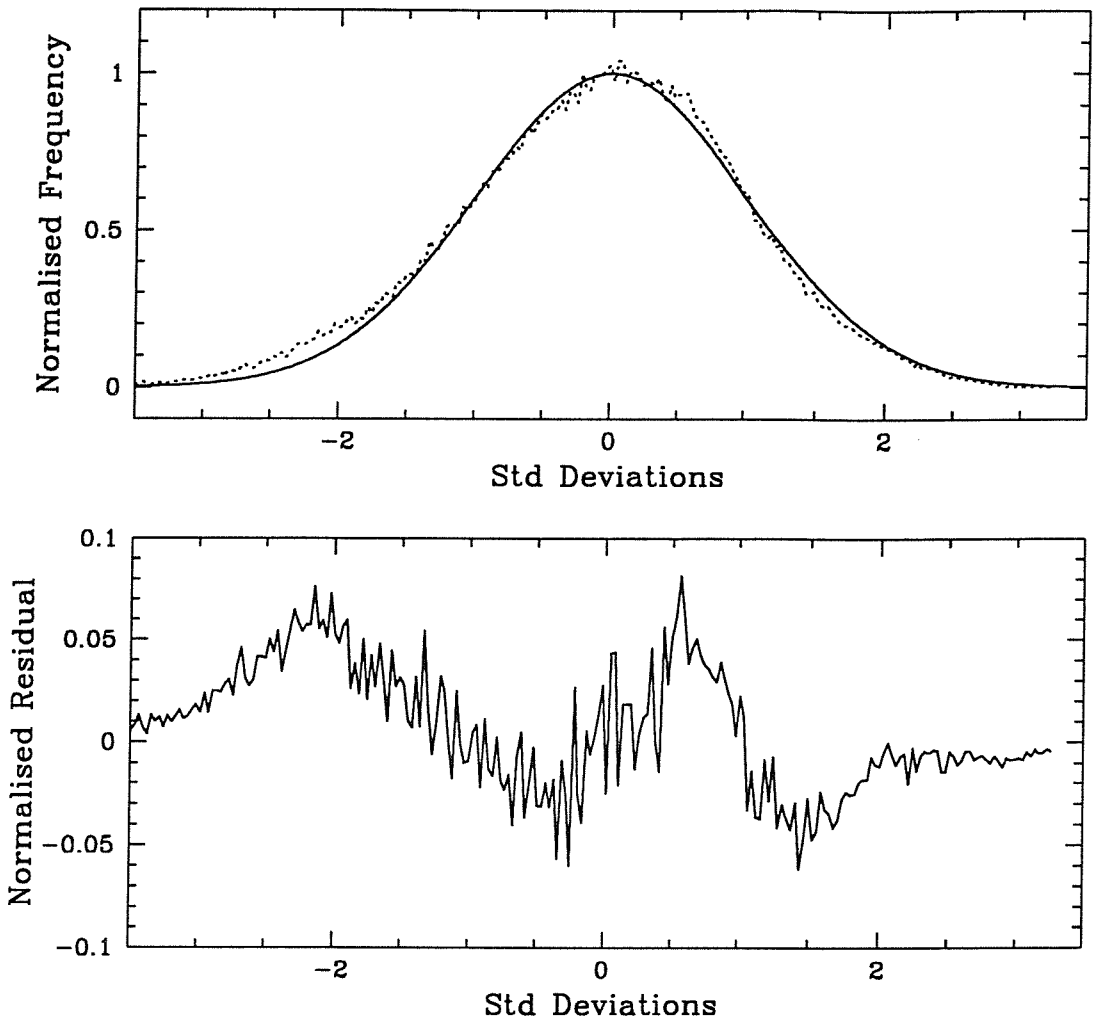


Figure 4.7: The noise distribution of the wavelet image at $a = 8$, where the original image was a Gaussian distributed random field. The solid line represents the best fit Gaussian - the residuals show that the fit is to within 10 %.

As an example of the ability of the wavelet technique to characterise structure that is not immediately apparent in the original image, Fig. 4.10 shows a comparison between the unprocessed correlation image and the wavelet image for a source extended to ~ 8 times the SPSF. The wavelet image is shown at the $a = 24$ sky pixels analysing scale, which for this particular source size gives the highest source centre wavelet coefficient. In total, each simulated observation of a particular source extent and the subsequent wavelet analysis for all possible scales ($a \approx 0.12$, 0.18, 0.24, 0.36, 0.48, 0.72, 0.96, 1.40, 1.91, and 2.87°) was repeated 10 times to reduce error bars. Consequently, it was possible to quantify the reliability of the wavelet coefficient data, an important exercise especially for the larger scale sources where edge effects degrade the precision of the wavelet data. Figure 4.11

True Size	W-A Size	Size Error	χ^2_8	χ^2_9	Signif.
0.25	0.234	6.656e-3	7.744	477.554	21.675
0.50	0.465	1.174e-2	7.419	1045.80	32.242
0.75	0.758	2.580e-2	3.313	911.075	30.129
1.00	1.184	5.088e-2	2.998	811.253	28.430
1.25	1.359	6.871e-2	1.359	714.656	26.708
1.50	1.501	0.111	7.771	255.044	15.725
1.75	1.380	0.104	3.110	263.782	16.145
2.00	1.814	0.181	1.847	154.220	12.344
2.25	1.919	0.162	2.554	241.188	15.448
2.50	2.146	0.319	3.288	82.378	8.893

Table 4.2: Wavelet analysis (W-A) data from one set of simulations.

shows a comparison between the true source size and the mean source size derived from the wavelet analysis of all simulated sources. The results of all simulations and wavelet

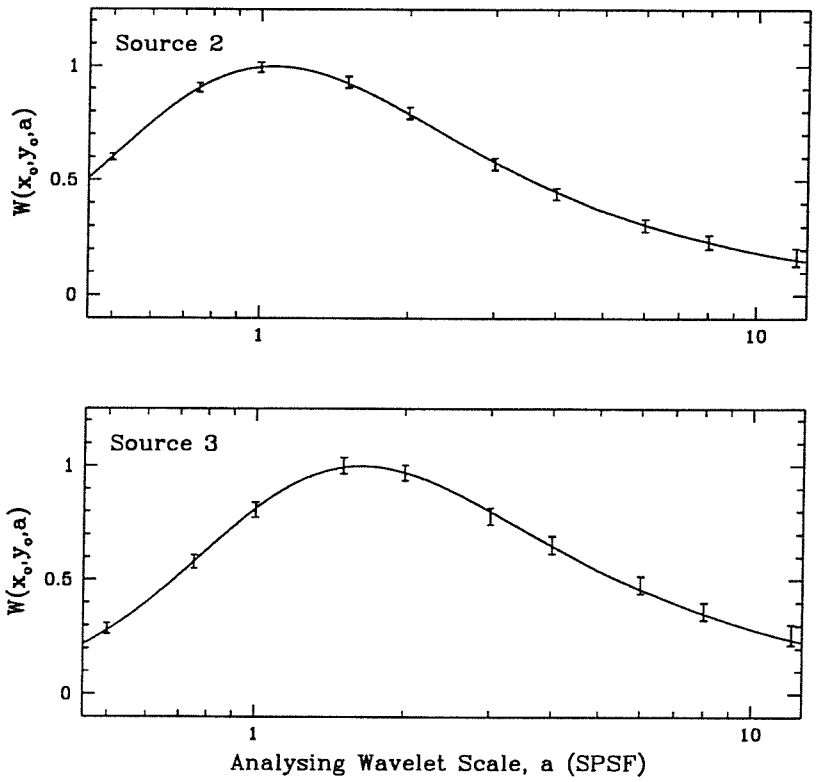


Figure 4.8: Wavelet analysis data and fit for source FWHM of 0.25 and 0.5 degrees.

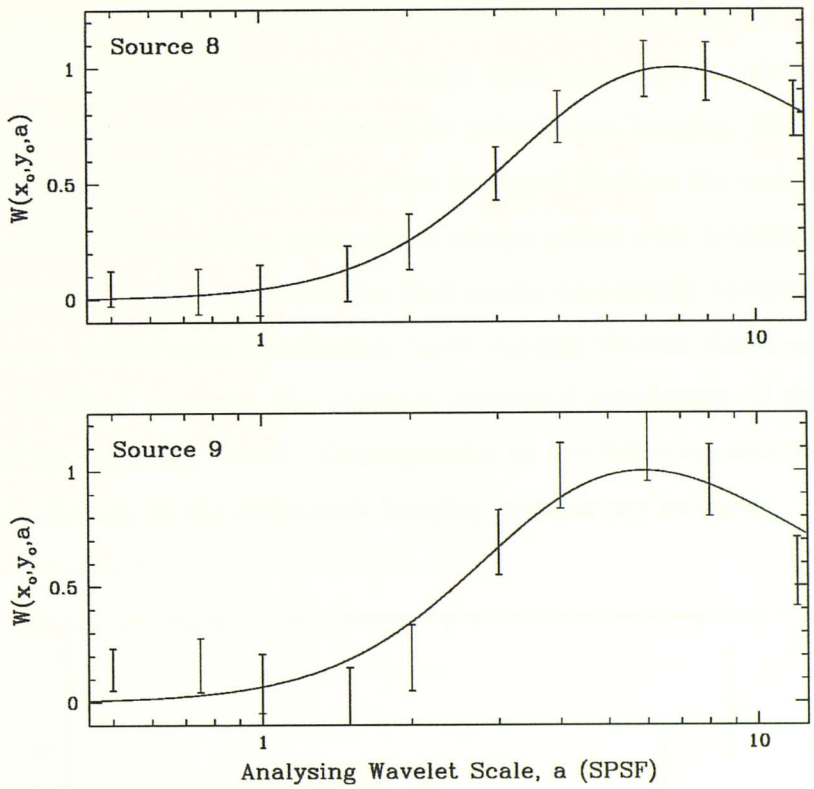


Figure 4.9: As Fig. 4.8 but for source FWHM 1.75 and 2.0 degrees.

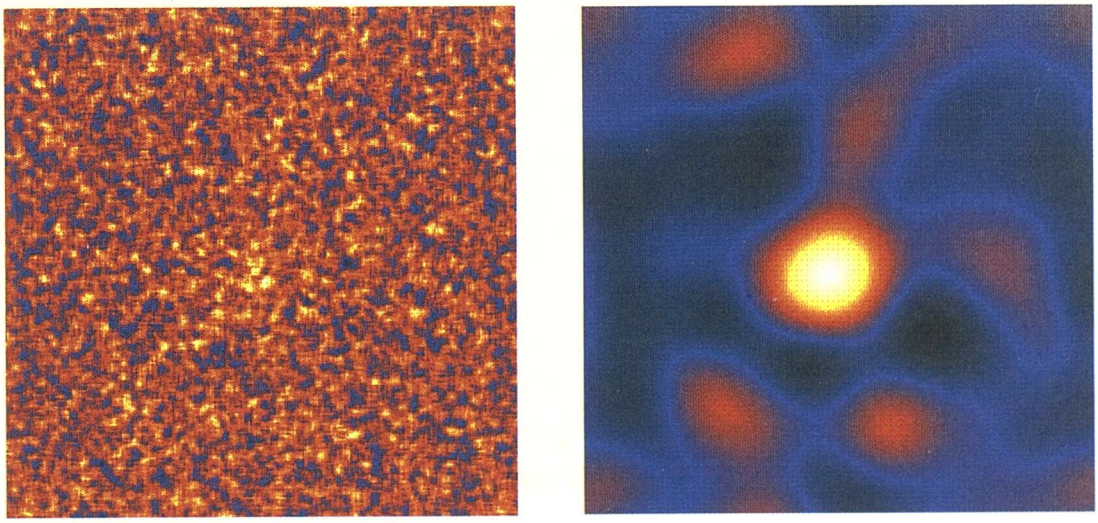


Figure 4.10: (a) The correlation image of a 2.0 degree FWHM Gaussian extended source, and for comparison, (b) the wavelet image at $a = 24$.

analysis are shown in Fig. 4.11.

4.4.4 Conclusions

From Figs. 4.10 and 4.11 it is apparent that wavelet analysis may be a powerful new technique for the initial identification of large scale structure, particularly in coded aperture images that are primarily optimised for point source imaging. However, Fig. 4.11 also highlights the fact that edge effects place an upper limit on the scale sizes that can be analysed whilst still providing estimates of source extent with meaningful error bars. Additionally, for the range of source scales that *can* be successfully analysed by assuming a near Gaussian wavelet noise distribution, there remains the fact that even if the source structure is identified precisely, the absolute statistical significance of the detection is still governed by the SPSF width. Consequently, in the following section a number of technological solutions to the multi-scale imaging problem are outlined.

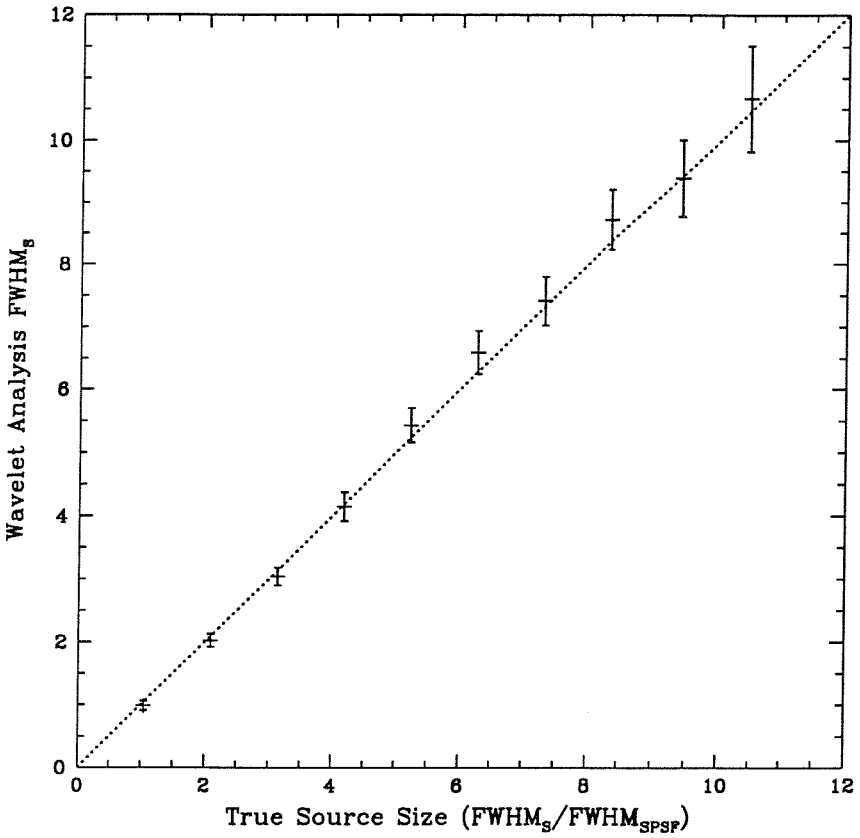


Figure 4.11: Source morphology as extracted by the wavelet analysis of the 10 simulated extended sources. Each data point is the mean of 10 simulations. Error bars on the last 3 data points are unreliable because of problems with edge effects. The size of the error bars is expected to increase with source size however, because the source flux was held constant for all source sizes.

4.5 Technological Solutions to Multi-scale Imaging

The simplest solution to imaging on more than one angular scale would be to use two co-aligned telescopes mounted side by side as used on the Spacelab 2 instrument, XRT [197]. For a fixed detector area, the dual telescope (DT) configuration, although providing simultaneous coarse and fine coverage, will only provide dedicated imaging on one angular scale at half the possible sensitivity. It would be useful to be able to utilise both fine and coarse detector planes for either fine or coarse scale imaging, as well as simultaneous dual-scale coverage if necessary. Three alternative telescope geometries are now outlined, all of which have a dual-scale imaging capability and employ a single detector plane only.

4.5.1 Adjacent Masks Telescope (AMT)

Retaining the same mask configuration as the DT but replacing the two independent detector planes with a single detector is the logical step forward. This system enables dedicated single scale imaging at full sensitivity if necessary, but because there is only a slight FOV overlap, the majority of observations would be sequential with respect to the dual-scale coverage. A mask-antimask capability (see Chapter 6, Section 6.2.4) would be necessary for both masks in order to remove cross-talk between the coarse and fine scale shadowgrams. A typical imaging sequence could therefore proceed as follows :

Sequence	Fine Mask Position	Coarse Mask Position
1	0°	0°
2	0°	60°
3	60°	60°
4	60°	0°

$$\left.\begin{matrix}1-2\\4-3\end{matrix}\right\} \Rightarrow \textit{Coarse Shadowgram Only.}$$

$$\left.\begin{matrix}1-4\\3-2\end{matrix}\right\} \Rightarrow \textit{Fine Shadowgram Only.}$$

4.5.2 Co-aligned Masks Telescope (CoMT)

The next logical step in the evolution towards the optimum dual-scale system would be to position one mask directly beneath the other. This would restore the simultaneous imaging capability but would not enable dedicated imaging on a single scale. The same observing sequence used for the AMT would be required.

4.5.3 Composite Mask Telescope (CMT)

It was demonstrated in Chapter 3, Section 3.3.1, that the quadratic residue difference sets used to generate HURA coded masks can be mapped onto any tesselating basic element. Consequently, it was possible to employ a finely sampled detector-mask configuration with an ideal SPSF for hexagonal as well as square telescope geometries. The ability to re-bin the HURA precisely into sub-elements without degrading the SPSF, coupled with the rotational anti-symmetry of the HURA, has led to an innovative solution to the dual-scale imaging problem, namely the Composite Mask Telescope, which will now be described.

Consider a relatively coarse HURA, for example a 127 element unit pattern, which is finely sampled with a 7:1 ratio. To provide a fine imaging capability whilst retaining the coarse pattern the next logical step might be to tessellate a second HURA with a unit pattern approximately equal to the coarse unit pattern multiplied by the sampling ratio (e.g. a 919 HURA), onto the open elements of the coarse HURA. The problems with this arrangement are immediately obvious : the sensitivity is reduced by a factor of 2 because of the lower mask transparency ; the fine mask is only 50 % sampled and so SPSF artifacts would dominate ; there is no way of separating fine and coarse-scale spatial information. The Composite Mask solution to these problems relies on the decomposition of the coarse and fine-scale mask patterns into a complementary pair of sub-masks.

Again, assuming there would be no technological restrictions to implementing the mask-antimask technique in the space-borne environment, the sub-mask patterns would be positioned one above the other and would rotate independently. The innovation of the Composite Mask is the way in which the coarse and fine-scale masks are decomposed to complement each other precisely. The first tessellation in the construction of the Composite Mask pair is of the fine HURA onto the closed element configuration of the

coarse HURA which produces the upper mask pattern. The second tessellation is of the remaining anti-mask section of the fine HURA onto the open element configuration of coarse HURA which produces the lower pattern of the composite pair. The rotational anti-symmetry and tessellation properties of both coarse and fine HURAs mean that for sequential 60 degree rotations of the upper pattern followed by the lower pattern (or alternatively sequential 30 degree anti-rotations between each pair), the combined mask pattern will alternate between the coarse and fine HURAs and also provide mask-antimasking on both scales. Figure 4.12 shows the decomposition of the fine and coarse masks into the complementary pair for a 127 HURA coupled with a 919 HURA. The resulting mask-antimask imaging sequence is shown in Fig. 4.13.

It should be noted that the 919-127 composite pair shown in Fig. 4.12 is not a special case and that dual-scale imaging can be achieved for any 'hexagonal' ratio (7, 19, 37 etc) between fine and coarse SPSFs, as long as two unit patterns exist that cover approximately the same number of sub-elements. For example, a 2467-127 composite pair with a 19:1 fine sampling of the coarse mask would provide imaging on 2 angular scales separated by a factor of $\sqrt{19}$ and indeed this complementary pair is used in the next section in a demonstration of the capabilities of the Composite Mask technique. Table 4.3 summarizes all possible technological solutions to dual-scale imaging and highlights the fact that the new Composite Mask imaging technique is clearly the most elegant solution from a sensitivity and implementation perspective.

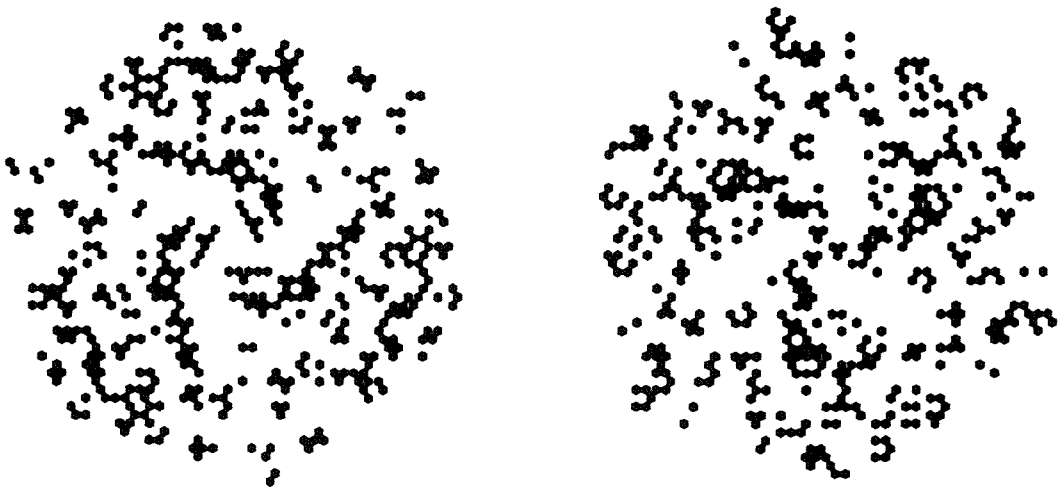


Figure 4.12: The complimentary mask pair for the Composite Mask based on 127 and 919 HURAs.

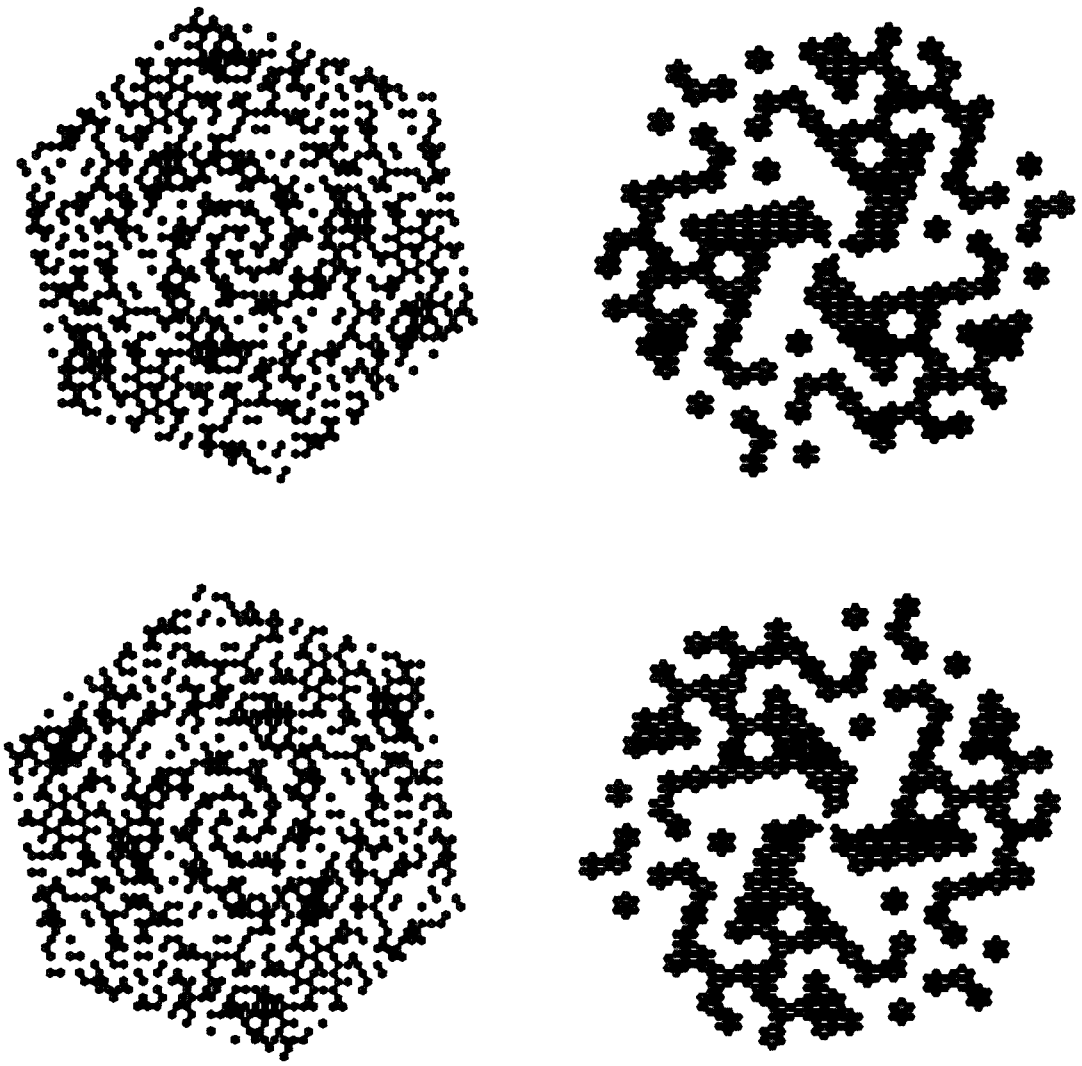


Figure 4.13: The 4 stage imaging sequence for the Composite Mask. The sequence provides mask-antimask imaging, over 2 angular scale ranges simultaneously.

Design	Simultaneous Image Collection	Dual-scale Sensitivity		Dedicated Sensitivity
		Fine	Coarse	
DT	Y	0.5	0.5	0.5
AMT	N	0.5	0.5	1.0
CoMT	Y	0.5	0.5	0.5
CMT	Y	0.5	0.5	1.0

Table 4.3: Comparison of the technological solutions to multi-scale imaging.

4.5.4 Imaging with the Composite Mask Telescope

At any one stage in the Composite Mask imaging sequence source information is gathered on one angular scale only. The rate of dual-scale coverage of a source will be determined by

the rotation rate of the masks. As a result of this sequential coverage, the problem of cross-talk between fine and coarse shadowgram data that limits other telescope arrangements, will not occur. Consequently, the fine and coarse shadowgram data can be deconvolved directly with the appropriate deconvolution array without any pre-processing.

However, since the coarse correlation image will contain initially sources on all angular scales up to the extended source limit including unresolved sources, it is possible that extended structure may be missed as a result of bright point sources in the FOV. It now becomes apparent that there is an additional advantage to having a dual scale sensitivity. The fine scale image, being sensitive to a more limited range of source scales, can be used to identify the positions and fluxes of any unresolved sources in the FOV. Once flux estimates and positions for these sources have been determined, an iterative post-processing technique similar to the CLEAN algorithm can be used to remove both the point sources and if necessary, the corresponding SPSF artifacts, from the coarse scale correlation image. This should reveal the coarse scale structure only which can also be CLEANed to remove any spurious structure due to SPSF artifacts.

As a token example of the capabilities of the Composite Mask technique, a simulated observation was performed of a region consisting of both unresolved and extended structure. As mentioned above, for this simulation a 2467-127 HURA pair was used, the fine and coarse SPSFs for which are shown in Fig. 4.14. The simulated region comprised of three Crab-like, unresolved sources each with a flux of 1×10^{-3} ph cm^{-2} s^{-1}, and an elliptical extended region 2° wide and 12° long with a flux of 1×10^{-3} ph cm^{-2} s^{-1} deg^{-2}, for the 100 keV to 10 MeV energy range. The total background counting rate simulated was 0.7 cts cm^{-2} s^{-1} and the observation time was 2×10^5 seconds. Figure 4.15(a) shows the fine resolution correlation image obtained from the simulation and as expected only the point sources appear significant. However, using the position and flux information obtained from the fine image (performing a maximum likelihood fit to the observed shadowgram), the unresolved sources were iteratively removed from the coarse image to leave only the large-scale structure which is clearly visible in Fig. 4.15(b).

4.6 Conclusions

From an observational perspective, it has been demonstrated that an increase in the resolving power of a coded aperture telescope does not necessarily improve the overall

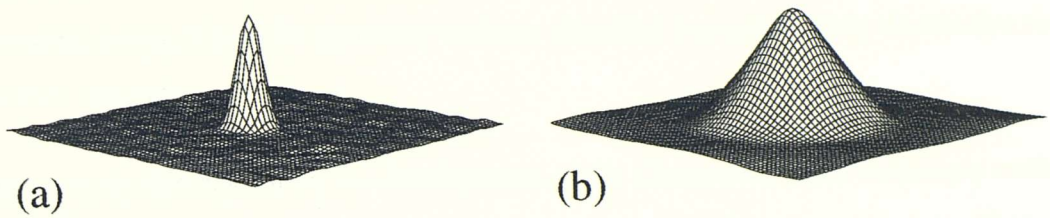


Figure 4.14: (a) The SPSF for the fine resolution mask, and (b) for the coarse resolution mask.

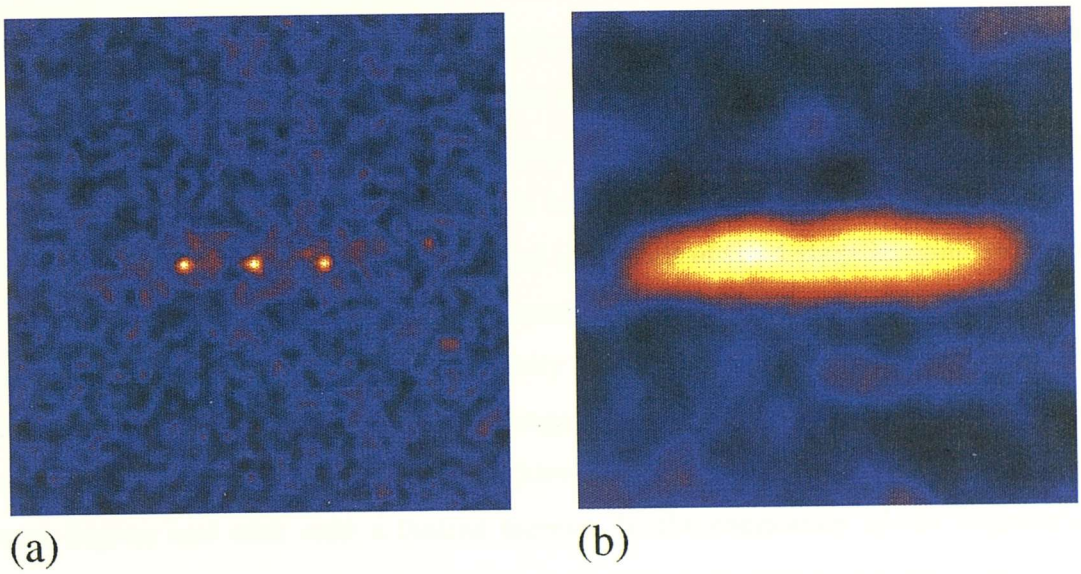


Figure 4.15: (a) The fine-scale correlation image showing predominantly unresolved sources. (b) The coarse-scale correlation image after the unresolved sources have been removed. The large-scale structure is seen clearly.

performance of the instrument. In Chapter 3 it was concluded that for the present pixellated detector technology there are limitations on the maximum achievable angular resolution, beyond which any resolution improvements are accompanied with a loss in the point source sensitivity. Similarly, in this chapter the converse trade-off has been highlighted, in which a loss in the sensitivity to extended sources occurs when the angular resolution is increased.

New analysis techniques such as the Wavelet Transform method implemented here may prove to be invaluable for the identification and characterisation of extended structure in fine resolution coded aperture images. Certainly, with the next generation of instruments pressing towards higher and higher angular resolutions, extensions to Wavelet Transform methods which also involve the use of iterative Bayesian reconstruction techniques [162]

may become necessary for *all* coded aperture data analysis. However, if the increased complexity of the analysis techniques are not balanced with significant improvements in processing power, it may be necessary to focus attention towards finding alternative coded aperture concepts.

Consequently this has led to the development of the Composite Mask technique outlined in Section 4.5.3. This novel method of simultaneous imaging on more than one angular scale over the same energy range, has been successfully demonstrated for a realistic observational scenario comprised of extended as well as unresolved sources. The near perfect imaging properties of both individual mask patterns are retained with the Composite Mask and so reconstruction can proceed by simple correlation methods. Similarly, because coarse and fine resolution data is collected separately, it is not necessary to pre-process each shadowgram to remove cross-talk between scales sizes. The mask-antimask technique and the benefits that accompany it (see Chapter 6, Section 6.2.4), are automatically implemented as a result of the 4 stage imaging sequence. The complementary nature of each sub-mask ensures that dedicated imaging on a single angular scale with optimum sensitivity is also possible. Thus, the Composite Mask technique can increase the effective angular sensitivity range of a coded aperture telescope from the fractions of a degree achieved with a single mask instrument, up to several degrees and with only a limited increase in the complexity of the engineering requirements, requirements that for the next generation of space-borne instruments will be technologically viable.

Chapter 5

Incident Pixel Reconstruction for Pixellated Coded Aperture Telescopes

Overview

Chapter 3 has highlighted the importance of detector position resolution with respect to point source sensitivity, in addition to the obvious benefits to source location accuracy. Prior to this chapter, it has been assumed that all incident photons can be detected with a spatial resolution equal to the detector pixel dimensions. However, as will become apparent, determining the pixel of incidence for a single incident photon can be a complicated task. This chapter considers the detection and what will be known as IPR (Incident Pixel Reconstruction) efficiencies of multiple layer CsI(Tl)-photodiode/cadmium telluride detector arrays in the framework of the EIDOS design. Initially, the objective is to quantify the ability with which the pixel of incidence can be reconstructed from a sequence of multiple bar and/or multiple layer energy deposits, all from a single incident photon. The necessity of IPR for EIDOS means that image reconstruction is no longer a two stage process but requires an initial 'construction' of the shadowgram prior to deconvolution.

5.1 Introduction

One of the major areas of advancement in the design of high energy astronomical telescopes has been in the ability to simultaneous collect high resolution images whilst providing coherent spectroscopic coverage over several decades of energy. The next generation of imaging telescopes such as the recently proposed imager onboard the ESA M2 mission INTEGRAL, will utilise current developments in pixellated detector technology to provide the simultaneous coverage. Chapter 3 has demonstrated that in order to optimise the available sensitivity (and as will become apparent, angular resolution) the detector plane of a coded aperture telescope must provide good sampling of the mask pattern.

The major technological challenge in attaining such a fine position resolution has been in the design of a detector that provides good sampling over not just a few hundred keV, but from tens of keV up to several MeV, whilst retaining a good spectroscopic capability and high detection efficiency. Such were the basic scientific demands on the proposed INTEGRAL imager, EIDOS. The solution to the problem has been the development of pixellated, multi-layered detectors where the upper layers are sensitive to low energy photons and where the added depth of the lower layers ensures good stopping power for high energy photons.

5.2 The Need for Efficient IPR

Even for high-Z materials such as CsI(Tl) and CdTe, at energies above several hundred keV an incident photon is more likely to Compton scatter into a second pixel rather than be absorbed photoelectrically in a single pixel. For a multiple layer pixellated detector, a series of pixels will be triggered by a single incident photon that has undergone Compton scattering. If the pixel in which the first interaction occurred can not be determined correctly, a significant reduction in sensitivity will be seen since the detector has effectively lost position resolution. EIDOS was designed to detect photons efficiently up to several MeV where there will be a preponderance of multi-pixel events and it is thus essential that the methods of IPR derived are optimised for these multi-pixel interaction sequences.

Since the light travel time (typically tens of picoseconds in arrays of this dimension) of the

interacting photon is well beyond the timing capabilities of both the scintillating material and the processing electronics, 'time-tagging' events is not a realistic way of determining the first interaction position. The only remaining option is to produce a semi-empirical solution to event reconstruction via detailed modelling of the detector. The following sections present the results of simulations of the proposed EIDOS detector plane with respect to its IPR capabilities.

5.3 Monte Carlo Simulations of EIDOS

5.3.1 The Modelled EIDOS Detector Geometry

The detector modelled in this section, as mentioned above, was based very closely on the initial instrument proposed in response to the recent ESA Announcement of Opportunity (AO) for the design of the INTEGRAL imager. The EIDOS detector plane contained three sensitive layers, each consisting of 24 triangular modules. Within each of these modules there were 120 hexagonal detector cells, which on the upper layer were divided again into 6 independent detector elements.

Prior to the INTEGRAL Phase-A study [180], the upper detector layer of the imager was composed of CsI(Tl)-photodiode pixels. This was changed however, prior to the AO such that the imager, as EIDOS, employed a 17280 element array of Cadmium Telluride (CdTe) pixels (6 CdTe detectors per hexagonal cell). This allowed the energy range of the instrument to be extended down to 15 keV (compared with 80 keV for the preceeding CsI(Tl)-photodiode array). As well as the improvement in the low energy threshold, the introduction of the CdTe array also brought with it improved position resolution which, as was demonstrated in Chapter 3 will improve both the sensitivity and as will be seen, the angular resolution of the instrument.

The middle and lower layers were identical in dimensions and consisted of 2880 hexagonal CsI(Tl)-photodiode pixels (120 pixels in each module) with each layer 3 cm in depth. The stack of three planes was surrounded by a BGO scintillator veto shield which was operated in anti-coincidence with the CsI detector elements. A tantalum tube collimator across the entrance aperture was employed to reduce the effects of the diffuse gamma-ray background. The general construction of the detector is shown in Fig. 5.1.

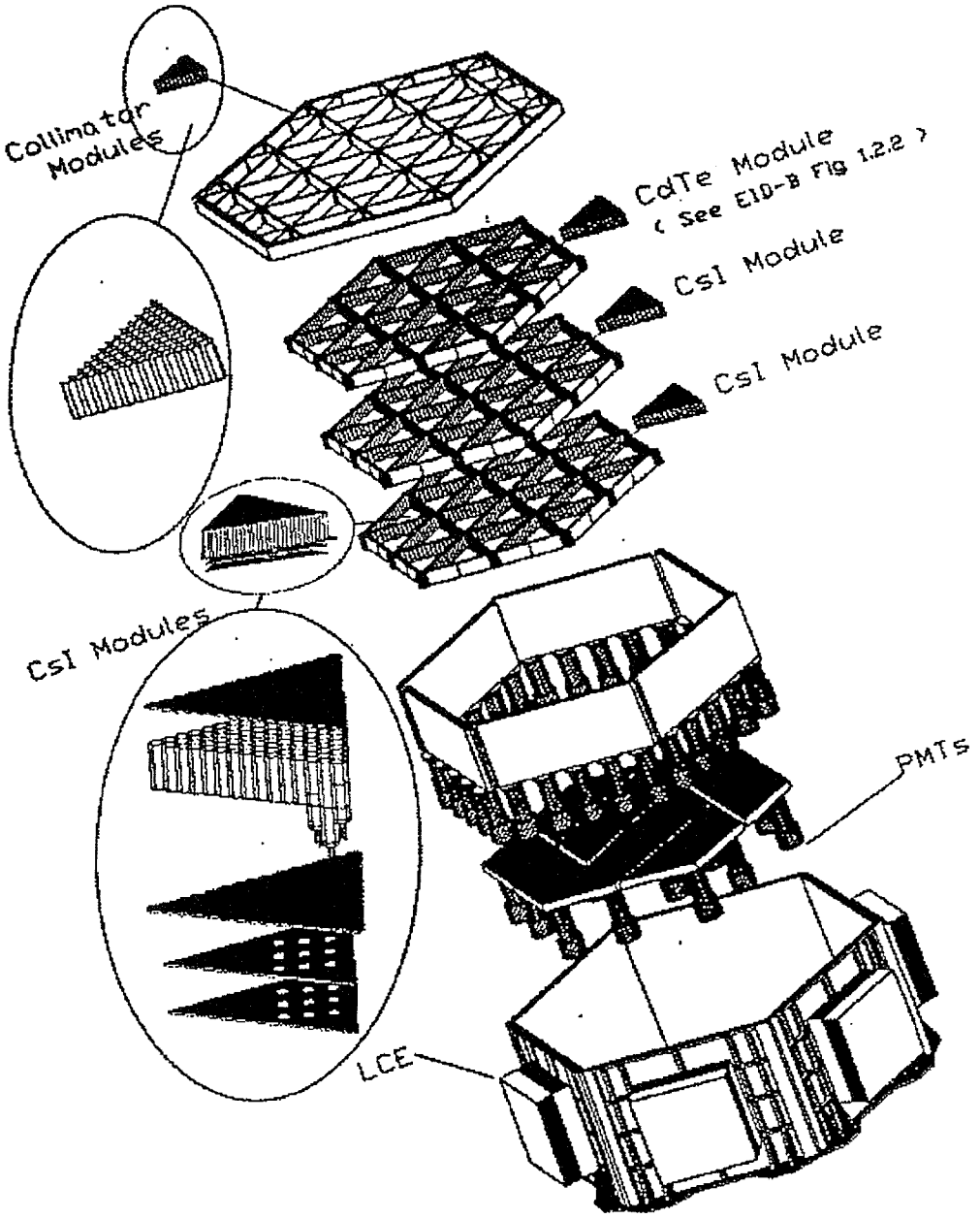


Figure 5.1: A three dimensional view of the EIDOS detector plane.

The division into three layers provides a way of tracking in 3 dimensions photons which scatter and interact in more than one element. This additional 3-d resolution capability allows discrimination between true source photons and background events that have intrinsically different scatter sequences at different energies. As shall be seen in the following Monte Carlo simulations, at higher energies (above 500 keV) true source interactions are more likely to be multiple site events. On the contrary, background events resulting from β decays in the CsI crystals (induced by cosmic ray interactions) are more likely to be single site events and can be rejected to improve the signal to noise ratio at

higher energies.

5.3.2 The Simulation Philosophy

An extensive model of the EIDOS geometry, which simulates all major components and includes most of the finer details such as the passive inter-crystal material, was generated with the GEANT-3 Monte Carlo simulation package [207].

To determine how efficient a particular IPR algorithm was for each type of interaction mode, knowledge of the incident pixel was required. The term 'IPR efficiency' used throughout this chapter refers specifically to the ratio of the number of events reconstructed in the correct pixel, to the total number of events incident on that particular pixel. The term 'pixel' refers to the 6 CdTe elements considered as a single element, along with the 2 CsI(Tl)-photodiode elements directly beneath. Since with the GEANT package it is possible to tag any interaction in any of the active materials, the first option for testing IPR efficiency involved distributing photons with a flat (logarithmic) energy spectrum uniformly across the entire detector plane. However, this would include firing photons into the passive detector support structure, photons which can not be reconstructed if they subsequently trigger any of the detector elements. Since it is debatable whether these events should be included in the IPR efficiency measurements, a second option for investigation was devised.

The second option involved firing a beam of photons, again with a flat energy spectrum (in log space), into each individual detector pixel contained within a single 120 element module, thus avoiding as much as possible the surrounding passive material. However, as will be seen in Section 5.4.3.1 the surrounding environment still has an affect on the IPR efficiency of a particular pixel. Thus into each pixel 12800 photons were fired with energies between 14 keV and 10 MeV distributed such that approximately equal numbers of photons were in each logarithmic decade of energy.

5.3.3 Data Processing

Upon firing a photon into the model, all interactions in any material defined as being active (CdTe, CsI(Tl) and BGO) were registered. Consequently an event, possibly containing several interactions, was characterised by a sequence of triggers each comprising two

numbers. The first number contained information on both the interaction layer and the particular pixel in which the energy deposit occurred, and the second number gave the precise energy deposit.

The raw data from each GEANT simulation is ideal in that all deposited energies are precise and no energy thresholds (lower and/or upper) are applied. Thus prior to any IPR analysis it was necessary to spectrally broaden each interaction according to known resolution functions and apply energy thresholds, below which any interactions were rejected and above which an out of range flag was triggered. Additionally, the BGO shield was operated in anti-coincidence with the detector pixels, thus if an event included a BGO interaction, the event was rejected in the IPR sequence. Initially however, the IPR response with respect to deposited energy was investigated for the idealised geometry without spectral broadening and thresholding.

5.3.4 Detection Modes

For photon interaction energies above a few hundred keV, multiple site interaction modes account for a substantial percentage of detected events, implying it is necessary to investigate the efficiency of a particular IPR method with respect to the interaction mode as well as the total event energy. Consequently events were categorised into the following 5 interaction modes, of which the first 3 are sub-classified into events with specific interaction characteristics. All interaction modes subsequently referred to are defined in Table 5.1. Some examples are shown graphically in Fig. 5.2.

Mode Type	Interaction
1a	single site - top layer
1b	single site - middle layer
1c	single site - bottom layer
2a	two site - top layer
2b	two site - middle layer
2c	two site - bottom layer
3a	> two site - top layer
3b	> two site - middle layer
3c	> two site - bottom layer
4	two site - different layers
5	> two site - different layers

Table 5.1: Event mode interaction characteristics.

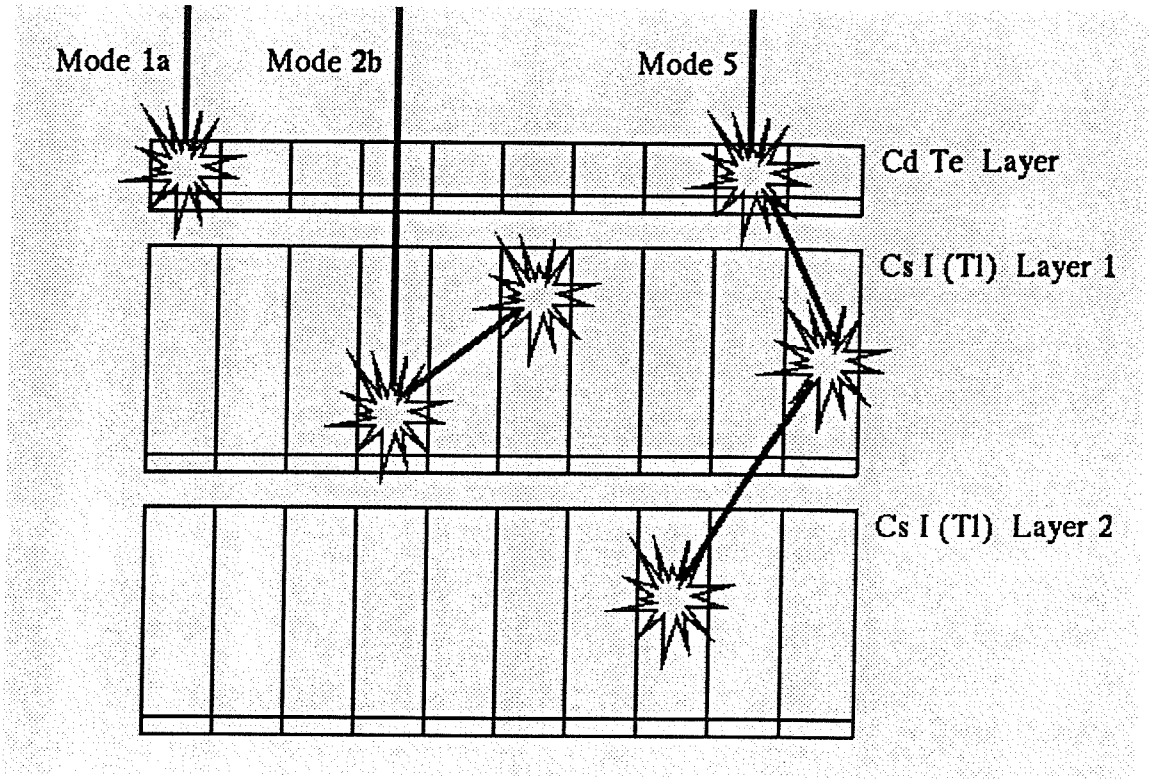


Figure 5.2: A few of the event mode interaction characteristics.

As mentioned above, as the incident photon energy increases above 500 keV, Compton scattering becomes the dominant first interaction mechanism and remains so for CsI(Tl) up until around 6 MeV and so multiple site 'showers' were expected to be the most frequent interaction mode. On the contrary, below 500 keV photoelectric absorption is the dominant interaction mechanism and thus single site events were prevalent. Figure 5.3 shows the percentage of events in each interaction mode as a function of the total event energy for the combined 120 pixel data set. The dominance of single site events up to 500 keV is clear, as is the prevalence of multiple site events above this energy.

5.4 IPR Algorithm Development

The methods of IPR examined here are derived directly from the detector geometry and are based on a combination of both deposited energy and interaction depth.

5.4.1 Methods of IPR

In accordance with the 3-layer EIDOS detector geometry, the 6 logical methods of IPR selected for investigation are given in Table 5.2. The description specifies how the incident

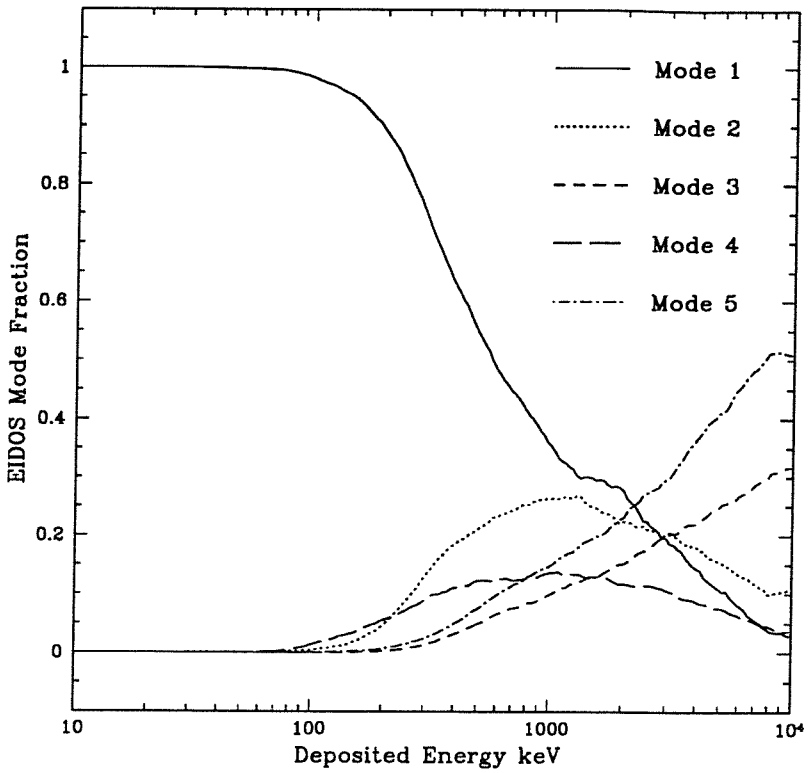


Figure 5.3: Interaction modes as a function of energy

pixel was found. It should be noted that for IPR methods using both the penetration depth and deposited energy as selection criteria, it was the penetration depth that was selected first. If more than one pixel was triggered in this layer, the incident pixel was then selected by the deposited energy value. The efficiency of a particular IPR method in the following analysis is given by the ratio of the number of events assigned to the correct incident pixel, to the total number of events in that mode per energy bin.

IPR Method	Description
1 MAXE	Maximum Energy
2 MINE	Minimum Energy
3 MAXD-MAXE	Maximum Depth + Maximum Energy
4 MIND-MAXE	Minimum Depth + Maximum Energy
5 MAXD-MINE	Maximum Depth + Minimum Energy
6 MIND-MINE	Minimum Depth + Minimum Energy

Table 5.2: The 6 IPR methods tested.

5.4.2 Ideal IPR Efficiencies

Using the event data from all 120 pixels as outlined in Section 5.3.3 which gave a total of 1536000 incident photons between 14 keV and 10 MeV, the efficiency of the 6 IPR methods was determined. In the initial investigation the effects of spectral broadening and energy thresholding were ignored. Figure 5.4 shows the IPR efficiency for each of the 6 techniques as a function of the *measured* energy (as opposed to the input energy).

Intuitively and evidently judging from Fig. 5.4, no single IPR method is capable of reconstructing all events correctly at all measured energies and it is clear that the efficiency may be optimised by selecting different IPR methods at different energies. This suggests that some methods are particularly suited to certain interaction modes, although at first glance, it would seem that the optimum efficiency is achieved simply by applying which ever method dominates at a particular energy. If this were so, then the only requirement for optimum reconstruction would be the total deposited energy regardless of interaction mode.

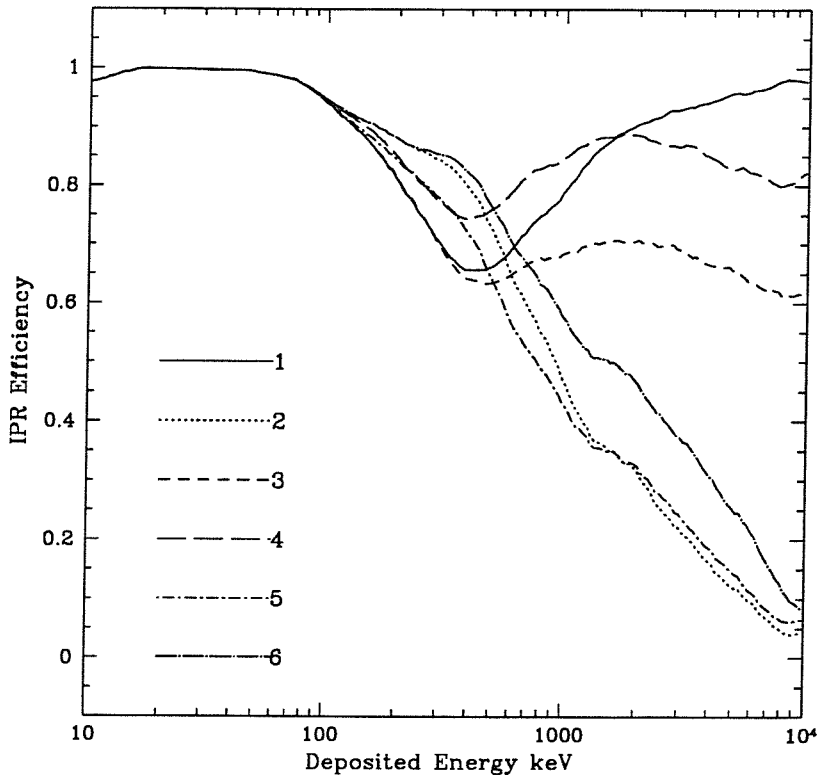


Figure 5.4: IPR efficiencies obtained by applying the same method to all modes of interaction. IPR methods are labelled as in Table 5.2.

To understand precisely the response of each IPR method and whether optimum

reconstruction can be achieved using only the three dominant methods in Fig. 5.4, a more rigorous approach was adopted. This involved determining IPR efficiencies for all 6 methods with respect to each mode. Figures 5.5 – 5.15 show the results of this mode by mode approach.

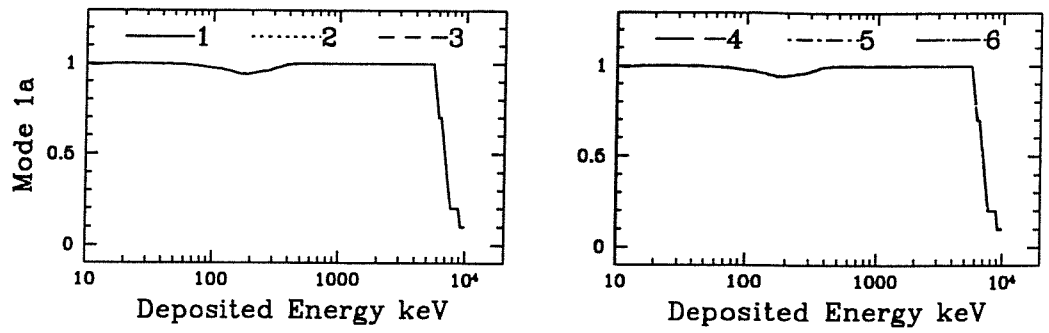


Figure 5.5: Mode 1a IPR efficiencies. Spectral broadening and energy thresholding have not been applied.

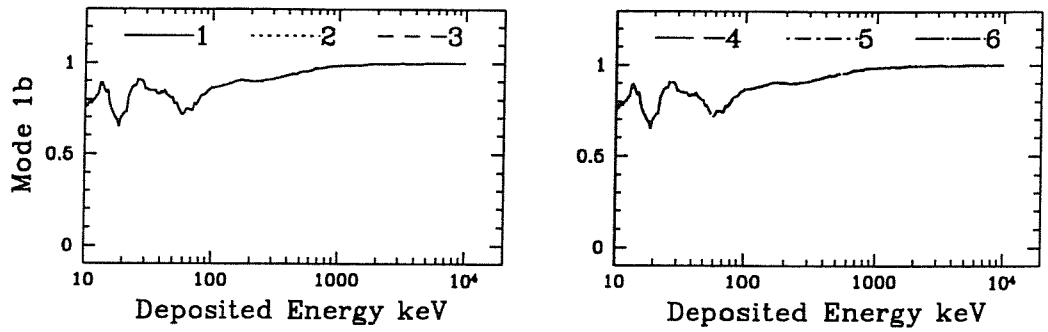


Figure 5.6: Mode 1b IPR efficiencies.

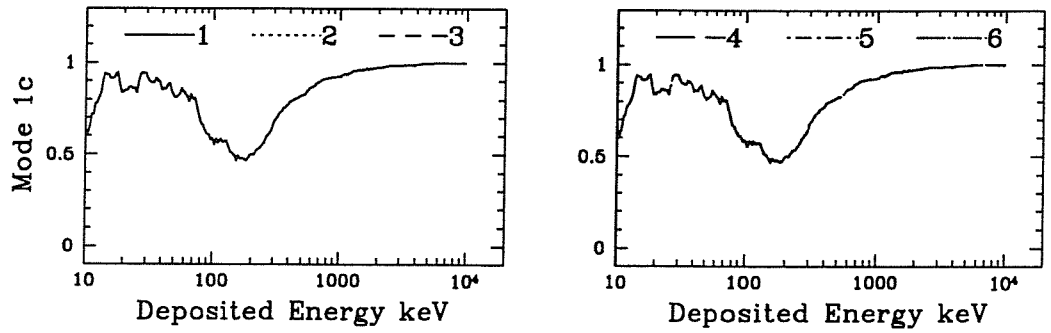


Figure 5.7: Mode 1c IPR efficiencies.

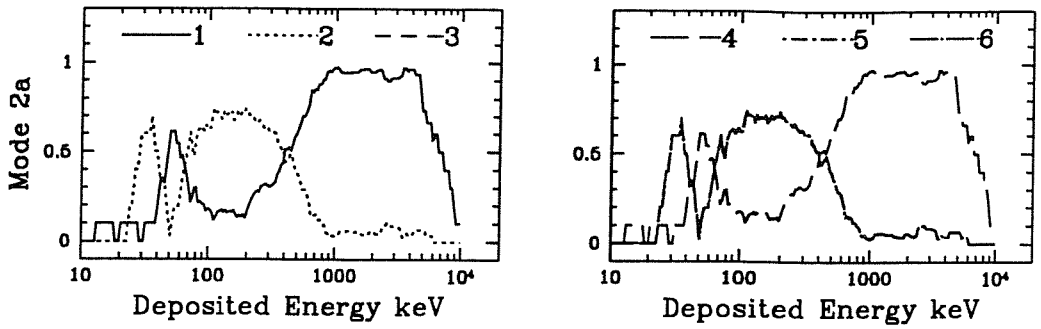


Figure 5.8: Mode 2a IPR efficiencies.

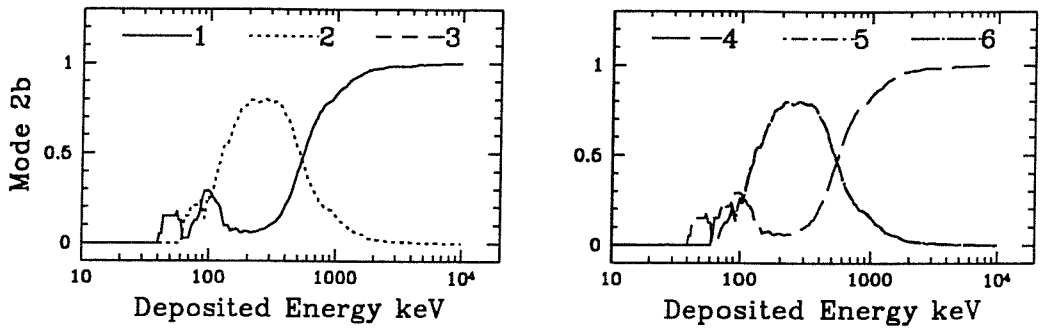


Figure 5.9: Mode 2b IPR efficiencies.

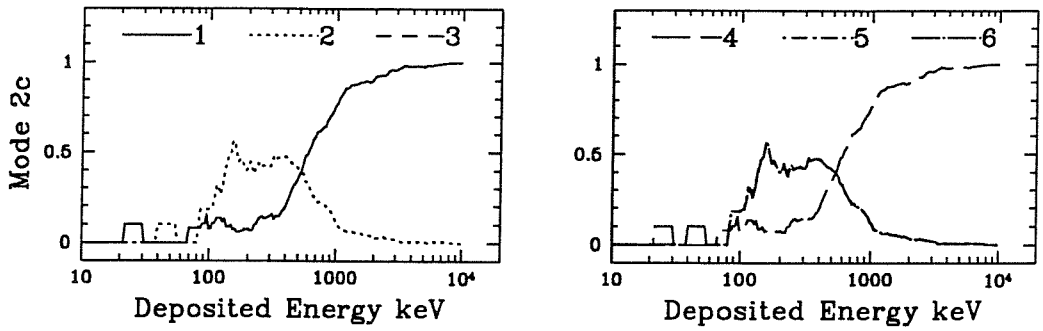


Figure 5.10: Mode 2c IPR efficiencies.

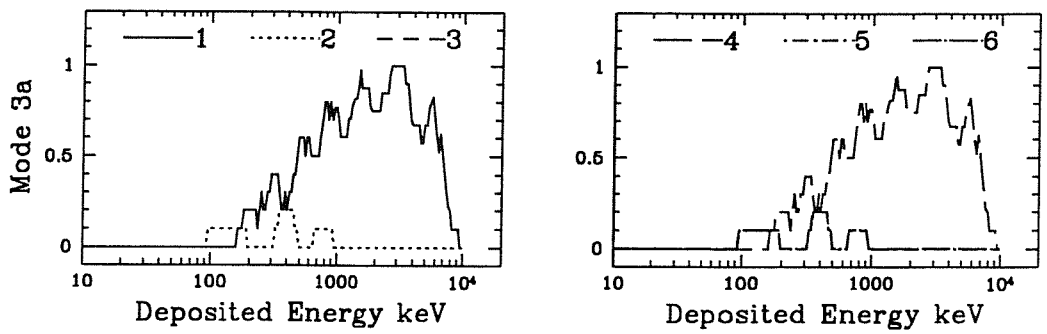


Figure 5.11: Mode 3a IPR efficiencies.

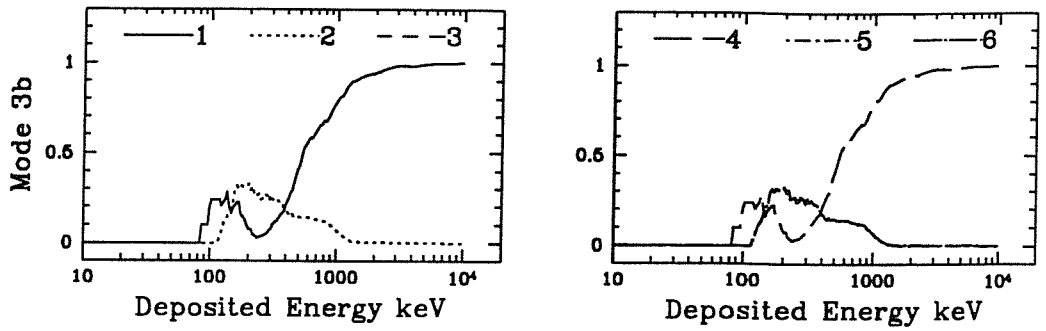


Figure 5.12: Mode 3b IPR efficiencies.

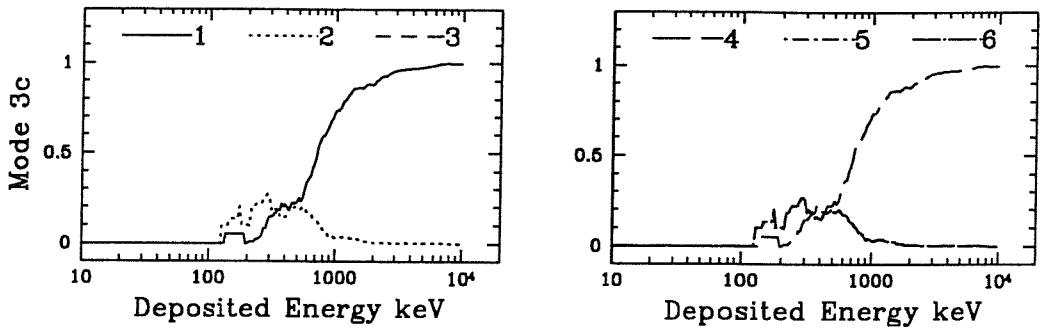


Figure 5.13: Mode 3c IPR efficiencies.

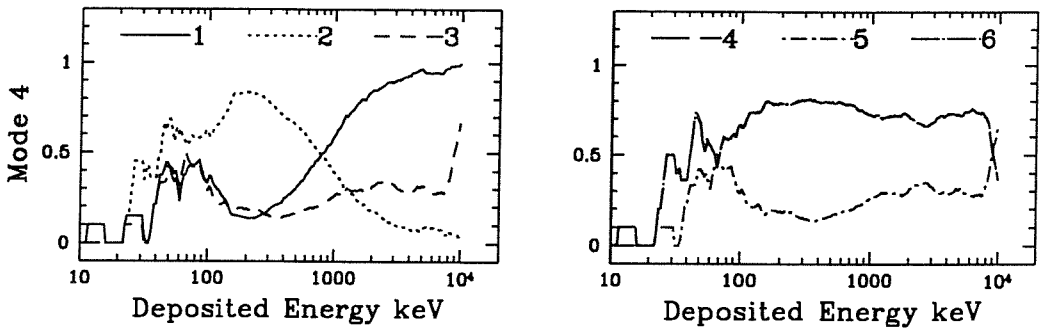


Figure 5.14: Mode 4 IPR efficiencies.

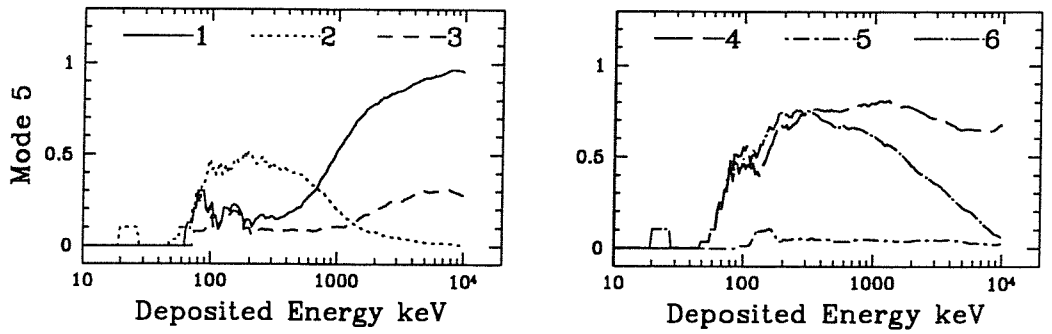


Figure 5.15: Mode 5 IPR efficiencies.

5.4.3 Discussion

5.4.3.1 Mode 1 Interactions

Figures 5.5 – 5.7 show the trivial result that all 6 IPR methods are equally efficient at reconstructing single site events. An interesting feature however, of all single site efficiencies (modes 1a, 1b, and 1c) is the decrease in efficiency between 80 keV and 800 keV, reaching a minimum of around 85% between 200 and 300 keV. If the photons, as explained above, are only incident on a single pixel, why is a drop in efficiency observed for these modes ?

The feature can be explained by considering the interaction of incident photons with the passive detector support structure. The inter-pixel structure seen in Fig. 5.1 is composed almost entirely of aluminium (Al) and there are also Al plates in front and behind the 3 detector layers. For any single site event to occur in a pixel other than the pixel into which the photons were fired, Compton scattering must have taken place in the surrounding passive material. According to the integrated Klein-Nishina formula (Eq. 5.1) [208] (integrated over all angles), the probability of Compton scattering in Al reaches a maximum at between 100 and 300 keV.

$$\sigma = ZN\frac{\pi r_0^2}{\alpha}\left(\left(1-\frac{2(\alpha-1)}{\alpha^2}\right)ln(2\alpha+1) - \frac{1}{2(2\alpha+1)}^2 + \frac{1}{2} + \frac{4}{\alpha}\right) \qquad (5.1)$$

where σ is the linear Compton absorption coefficient, r_0 is the classical electron radius, Z and N are the atomic number and the number of atoms per unit volume in the scattering material, and $\alpha = h\nu/m_0c^2$.

Thus in the upper detector layer a small percentage of events between 100 and 300 keV are expected to be incorrectly reconstructed as a direct consequence of scattering from the Al face-plate (at angles > 80 degrees) into the adjacent CdTe pixels. This effect is seen in mode 1a shown in Fig. 5.5. Sub-pixellation of the CdTe pixels into 6 triangular elements requires a substantial amount of additional Al support in the top layer. Consequently the probability of secondary Compton scattering in passive material will be high which suggests that a considerable fraction of lower energy single site deposits in the upper CsI(Tl) layer are also expected to be incorrectly reconstructed.

Alternatively, following the first Compton scatter in the Al face-plate, the scattered photon may then interact in the lower CsI(Tl) layer only, again in an adjacent pixel. However, the percentage of single site events in the upper CsI(Tl) layer at energies below 100 keV, and in the lower CsI(Tl) layer at energies below 400 keV was under 10%, and so the primary contribution to the loss of IPR efficiency is from Compton scattering prior to deposits in the CdTe.

5.4.3.2 Mode 2 Interactions

All mode 2 events (Figs. 5.8 - 5.10) whether mode 2a, 2b, or 2c, show a distinct cross over point between IPR methods at around 511 keV. The two competing methods for these interactions are MAXE and MINE (all other IPR methods give the same efficiency as either MAXE and MINE since they all include either MAXE or MINE as one of the selection criteria).

By considering the interaction depth of the initial photon as well as the deposited energy, the observed cross over point can be explained. Consider for example, two bar interactions in the middle CsI(Tl) layer : an incident photon with energy below the 511 keV cross over is more likely to interact in the upper part of the crystal. Since only forward scattering photons (and a small fraction of photons scattered through angles $\geq$ 90 degrees) are likely to interact again in the same layer, the second energy deposit will be larger and thus Mode 2 events below 511 keV can be reconstructed by selecting the minimum of the 2 energy deposits.

At incident photon energies above 511 keV, the first interaction is more likely to occur in the lower regions of the layer and thus for a second interaction to occur in the same layer, backscattering from the first interaction is required. This implies that the first deposit will be the largest and the event will be reconstructed correctly using MAXE.

At around 511 keV, both MAXE and MINE give the same result which suggests that both forward and backward scattering probabilities (integrated over the solid angles that restrict the second interaction to a different bar) are approximately equal. This is demonstrated graphically in Fig. 5.16. This is expected since the layer has been specifically optimised for 511 keV photon detection. This has been achieved by optimising both the trade-off between energy resolution and bar length, and also the trade-off between detection efficiency and bar length.

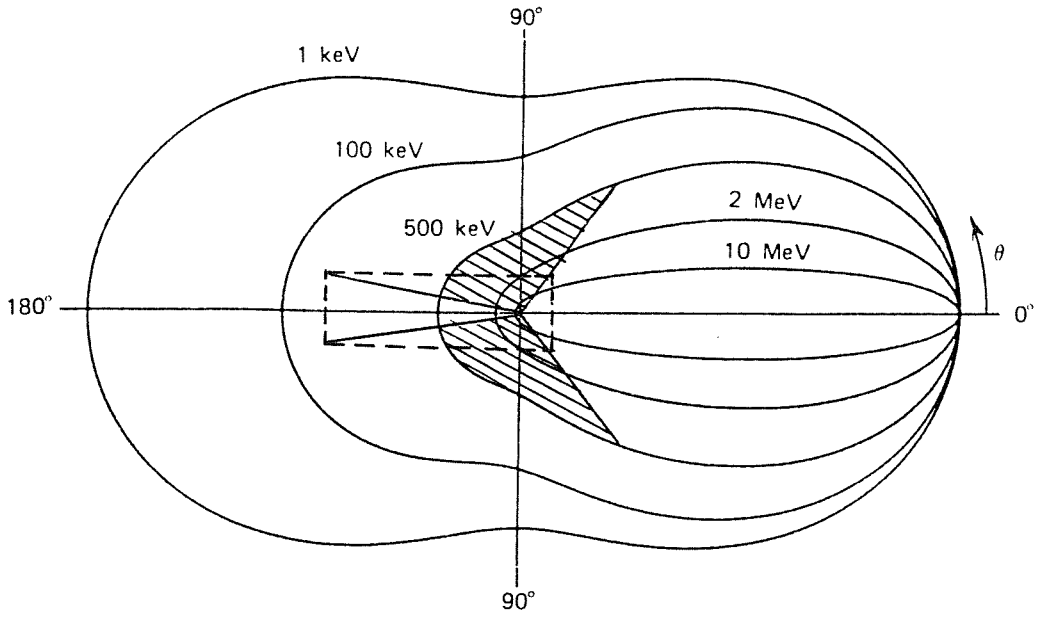


Figure 5.16: A polar plot of the number of photons (incident from the left) Compton scattered into a unit solid angle at the scattering angle θ. The highlighted areas show the fractions of forward and backward scattered Mode 2 events. These are approximately equal.

At deposited energies of ~4 MeV where pair production becomes the dominant interaction mechanism over Compton scattering, MAXE will continue to be the most efficient reconstruction algorithm *regardless* of the interaction mode. This is because for a pair production interaction, all of the incident photon energy apart from possibly 511 or 1022 keV, will be deposited in a single bar and will always be larger than the annihilation photon deposits. Similar arguments to those used above can be used to explain the cross over point for Mode 3 interactions shown in Figs. 5.11 – 5.13.

5.4.3.3 Mode 4 Interactions

Figure 5.14 shows the IPR performance for Mode 4 interactions (2 layer events). As this figure suggests, it is only for the multiple layer/bar interactions that a difference in performance between the remaining 4 IPR methods is seen. Again there is a cross over point between MINE and MAXE algorithms such that at energies below around 800 keV, MINE is more efficient than MAXE. This is indicative of the fact that most low energy Mode 4 detections, where the first interaction is in the upper layers, are necessarily forward scattered events.

Similarly at energies above 800 keV, Mode 4 detections will be predominantly

backscattered events and so the first interaction is the one with the highest energy deposit. The curious dominance of the MIND-MAXE (simply MIND in this case) algorithm between 280 and 1250 keV is best explained in terms of the ability of MIND to reconstruct both the low energy forward scattered events (normally recovered by MINE) and a significant percentage of the remaining events in which the second Compton scatter results in a lower energy deposit than the first. Above 1 MeV the MAXE algorithm begins to take effect since the majority of Mode 4 events at these energies are backscattered lower layer Compton interactions, as well as pair production events at higher energies.

5.4.3.4 Mode 5 Interactions

Figure 5.15, suggests that only two reconstruction algorithms are needed for Mode 5 events. High energy deposits (above 2 MeV) from the first Compton interaction, are more likely to occur in the lower layer(s) of the detector. Consequently Mode 5 events must follow on from either backscattered Compton interactions or pair production events in which the annihilation photons have triggered several other bars. In both situations, and this applies really to all interaction modes at these energies, the first energy deposit is the largest. Below 2 MeV, back in the Compton region, both MIND algorithms fair better than the remaining MAXD algorithms (which incidentally, are redundant algorithms), with MIND-MAXE the most efficient down to around a few hundred keV, at which point MIND-MINE may be marginally better. As Mode 5 events can encompass any combination of multiple layer/bar triggers it seems pointless trying to generalise the specific interaction modes in order to explain the dominance of MIND-MAXE over other IPR methods at these energies. However, from the evidence in Fig. 5.15, it is worth noting that around 1 MeV energies, Mode 5 events are expected to contain a significant fraction of forward scattered events with the first interaction in the upper CsI(Tl) layer.

Thus, Figs. 5.5 – 5.15 demonstrate the suitability of each IPR method to certain interaction characteristics. In the following section the 4 apparently necessary IPR algorithms are implemented in the form of a complete IPR formula from which any interaction mode with a deposited energy within the EIDOS observation range, can be reconstructed with the maximum possible efficiency.

5.4.4 The Complete IPR Formula

Table 5.3 shows the complete IPR formula for the ideal case where no spectral broadening or energy thresholding has been applied. The cross-over points between each of the dominant IPR methods have been taken from Figs. 5.5 – 5.15. Applying the above solution to the original flat continuum spectrum gives the total IPR efficiency for the ideal case. This is shown in Fig. 5.17 along with the IPR efficiency obtained by using the most dominant 3 IPR methods from Fig. 5.4 (MIND-MINE, MIND-MAXE and MAXE).

Mode	Optimum IPR Method
1a,b,c	MAXE for all E_D
2a,b,c	$E_D \leq 511$ keV $\rightarrow$ MINE
	$E_D > 511$ keV $\rightarrow$ MAXE
3a	MAXE for all E_D
3b	$E_D \leq 400$ keV $\rightarrow$ MINE
	$E_D > 400$ keV $\rightarrow$ MAXE
3c	$E_D \leq 200$ keV $\rightarrow$ MINE
	$E_D > 200$ keV $\rightarrow$ MAXE
4	$E_D \leq 150$ keV $\rightarrow$ MIND-MINE
	150 keV $< E_D \leq 280$ keV $\rightarrow$ MINE
	280 keV $< E_D \leq 1250$ keV $\rightarrow$ MIND-MINE
	$E_D > 1250$ keV $\rightarrow$ MAXE
5	$E_D \leq 300$ keV $\rightarrow$ MIND-MAXE
	300 keV $< E_D \leq 2000$ keV $\rightarrow$ MIND-MAXE
	$E_D > 2000$ keV $\rightarrow$ MAXE

Table 5.3: The complete IPR formula.

However, from Fig. 5.17 it would appear that an IPR formula based on only the three dominant IPR methods in Fig. 5.4 is as efficient as the optimum IPR formula given in Table 5.3. How can such extreme simplifications be made in the light of the above IPR requirements ? Essentially, the complexity of the IPR algorithm results from an attempt to reconstruct events at energies and modes that have only a marginal effect on the final efficiency curve. The simplified IPR algorithm is shown in Table 5.4.

Mode	Optimum IPR Method
All	$E_D \leq 511$ keV $\rightarrow$ MIND-MINE
	511 keV $< E_D \leq 1722$ keV $\rightarrow$ MIND-MAXE
	$E_D > 1722$ keV $\rightarrow$ MAXE

Table 5.4: The simplified IPR formula.

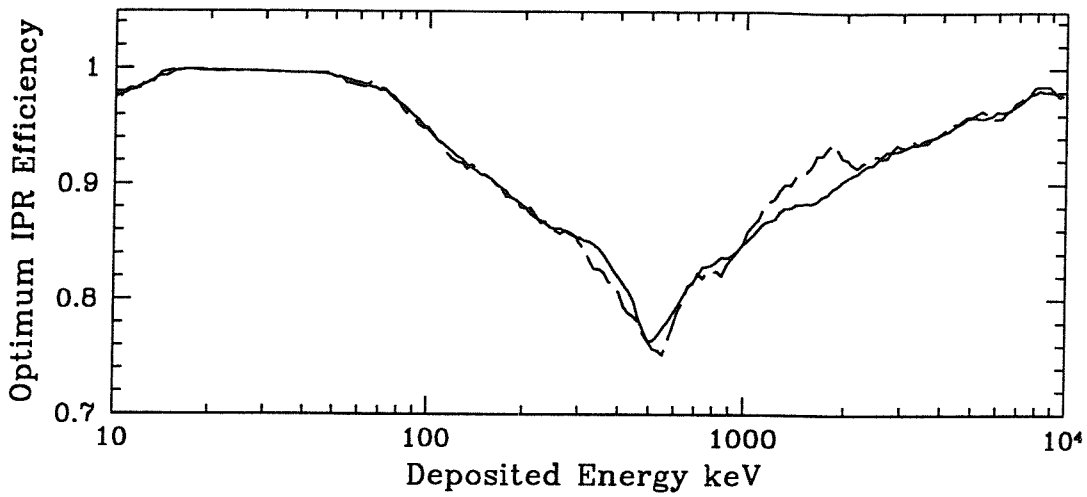


Figure 5.17: The ideal IPR efficiency as a function of deposited energy obtained with the IPR formula of Table 5.3 (dotted line), and for comparison, the IPR efficiency obtained by using the most dominant 3 IPR methods in Fig. 5.4 (solid line).

In reality, the performance of the EIDOS detector will be restricted by energy thresholds and spectral broadening. As will be demonstrated in the following sections, the IPR efficiency is expected to be significantly altered by introducing threshold and spectral broadening factors.

5.5 Energy Thresholds and Spectral Resolution

5.5.1 Low Energy Thresholds

As will be seen in Chapter 7 (Section 7.2.4), the low energy threshold (LET) of a particular detector channel is a physical limitation determined primarily by the electronic noise present in the detector itself and in the subsequent processing electronics. For EIDOS, the CdTe solid state detectors were expected to have an LET of around 15 keV [209, 148], whilst the CsI(Tl)-photodiode detectors were expected to have an LET of around 120 keV [148]. On applying these thresholds to the simulated data, the efficiency of each IPR method was expected to decrease in some energy bins and increase in others. A change in IPR efficiency should occur anywhere that on applying the LETs, a primary or secondary interaction is removed from the event. To illustrate this point, Figs. 5.18 – 5.22 show the resulting IPR efficiencies after LETs have been applied. As before the IPR analysis is performed mode by mode, although in Figs. 5.18 – 5.22 only the most efficient IPR

method is shown at a particular energy. For comparison the ideal IPR data is also shown.

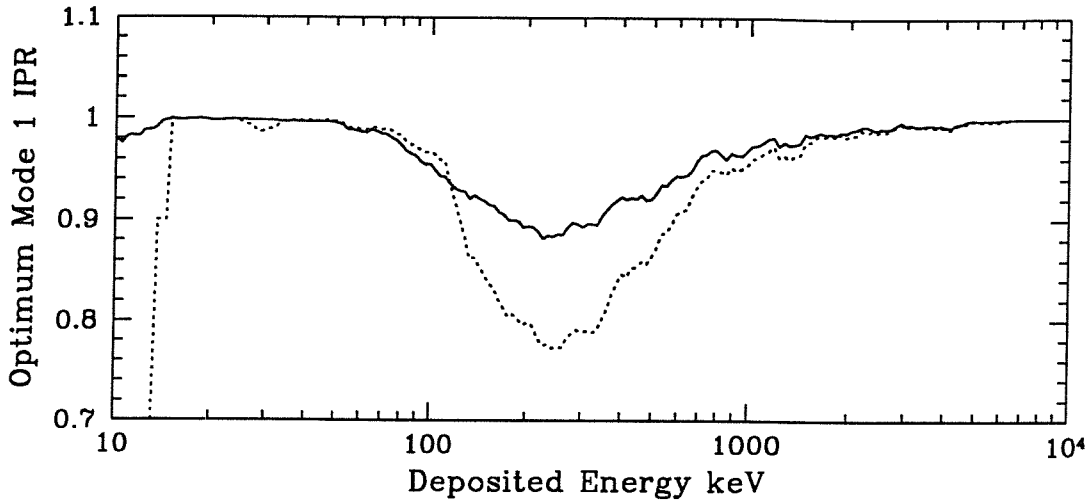


Figure 5.18: Most efficient IPR efficiencies for all Mode 1 events for both the ideal case (solid line) and with the addition of layer to layer LETs (dotted line).

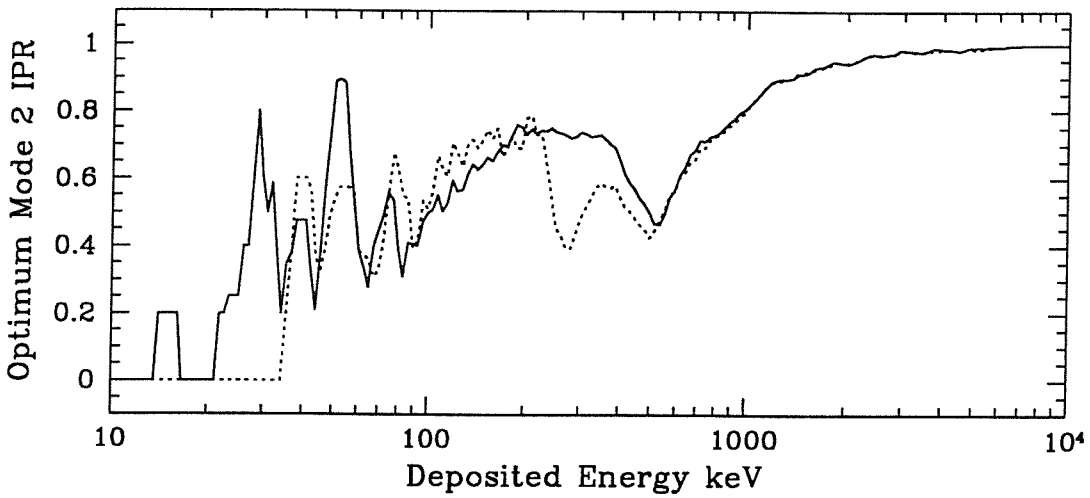


Figure 5.19: As in Fig. 5.18 but for Mode 2 events.

The changes in IPR efficiency seen in these figures must be weighted against the dominance of that mode (Fig. 5.3) over the energy range that the change has occurred in order to determine the net effect on the final efficiency. For all figures, the major differences between the two IPR curves occur between 100 keV and 750 keV. Since the most dominant interaction modes over this energy range were Modes 1, 2 and 4, the most significant change in total IPR efficiency was expected to result from changes in these modes. The largest contribution to the overall drop in efficiency was found to arise from incorrectly reconstructed Mode 1 events which accounted for more than 50% of all events in this energy range.

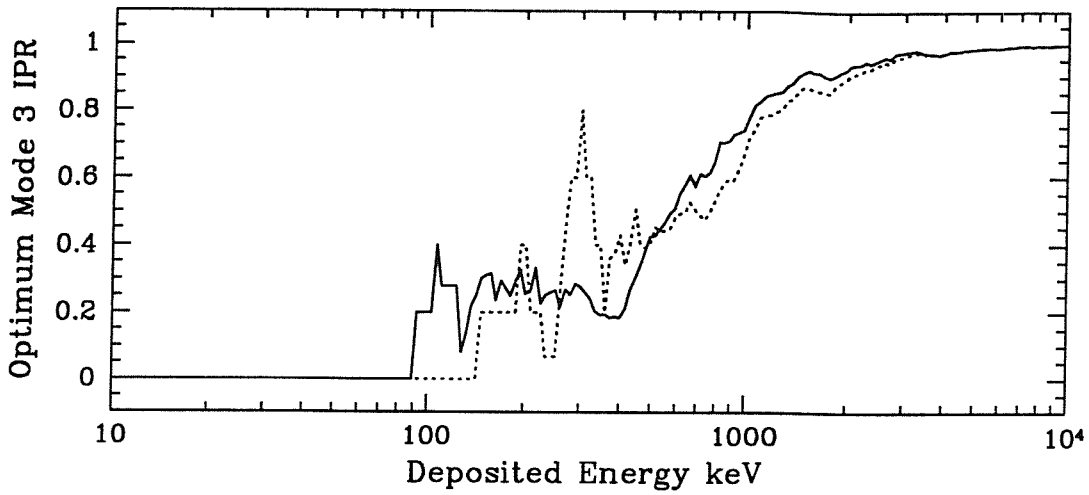


Figure 5.20: As in Fig. 5.18 but for Mode 3 events.

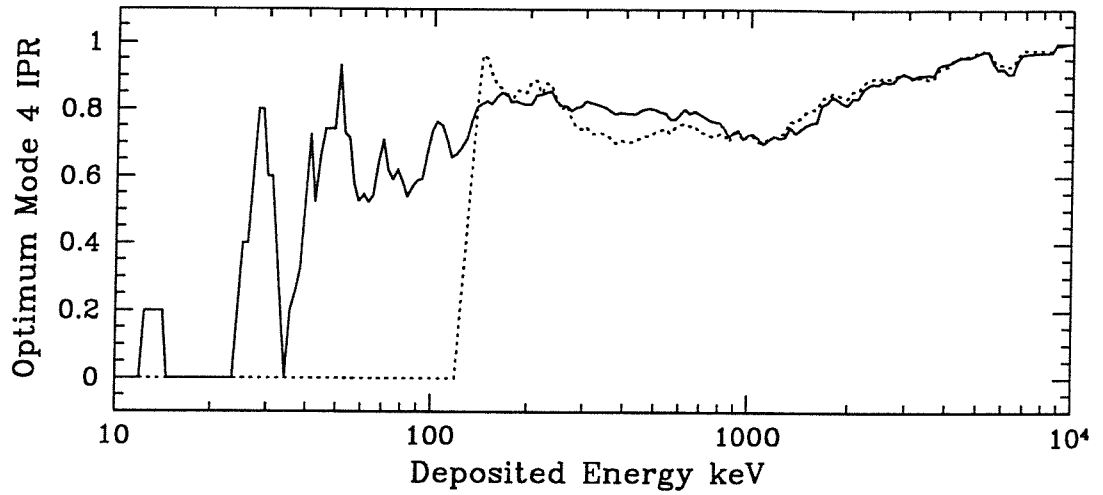


Figure 5.21: As in Fig. 5.18 but for Mode 4 events.

The drop in IPR efficiency is best explained by considering which type of events give rise to Mode 1 interactions when the LET is applied. Considering the case where only a single interaction is lost upon thresholding, it is possible for new and incorrectly reconstructed Mode 1b/c events to be created out of events previously classified as either Mode 2b/c, or Mode 4 events.

The drop in IPR efficiency reaches a maximum at around 250 keV (Fig. 5.18) implying the previous Mode 2 and/or Mode 4 events can not have deposited an energy of more than around 370 keV. Assuming this energy is close to the full energy peak of the incident photon and not part of the Compton tail of higher energy photons, several scenarios can be envisaged that would lead to an increase in the number of incorrectly reconstructed Mode 1b or 1c events :

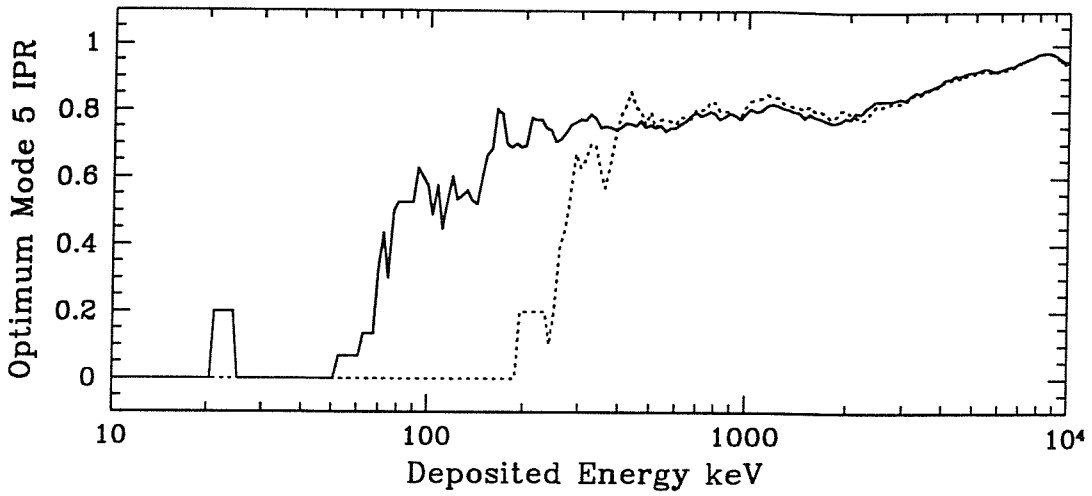


Figure 5.22: As in Fig. 5.18 but for Mode 5 events.

- If an incident photon interacts for the first time in the upper CsI(Tl) layer, it is more likely to occur in the upper part of the crystal (see discussion in Section 5.4.3.2). Consequently a second interaction in the same layer is more likely to occur after a forward scatter of the incident photon. This implies $E_1 < E_2$ with respect to the deposited energies. On applying the LET, E_1 maybe lost, leaving E_2 as a Mode 1b event that can not be correctly reconstructed.

- Similar arguments can be applied to create new Mode 1c events from Mode 2c events.

- If the forward scattered photon described above reaches the lower CsI(Tl) layer before interacting a second time (Mode 4), this implies again that $E_1 < E_2$ and that E_1 maybe lost after thresholding to leave behind a Mode 1c event.

From the total IPR curve for all modes (Fig. 5.23) it is evident that the detector LETs seriously degrade the instrument sensitivity, reducing the IPR performance by around 10% at 300 keV.

5.5.2 High Energy Thresholds

High energy thresholds (HETs) do not affect the IPR response quite so seriously as LETs as can be seen in Fig. 5.24, simply because although an out of range energy deposit can not be measured, it can be registered as an interaction with an 'out of range' flag. Consequently, although this capability does not allow these events to be reconstructed, it

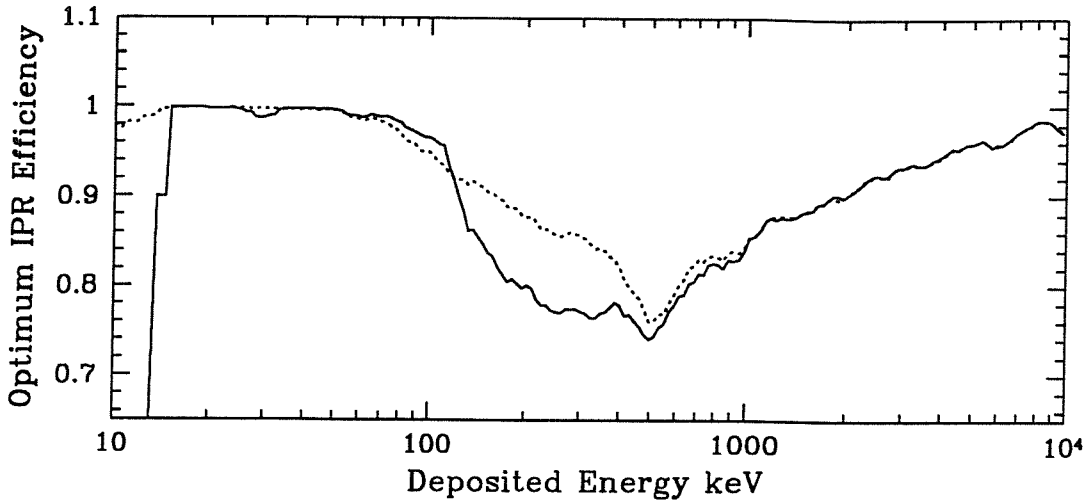


Figure 5.23: The IPR Efficiency as a function of deposited energy with the LETs applied. For comparison, the ideal IPR efficiency is shown (dotted line).

does allow them to be rejected and thus eliminates what would otherwise be an additional contribution to the background count at lower energies.

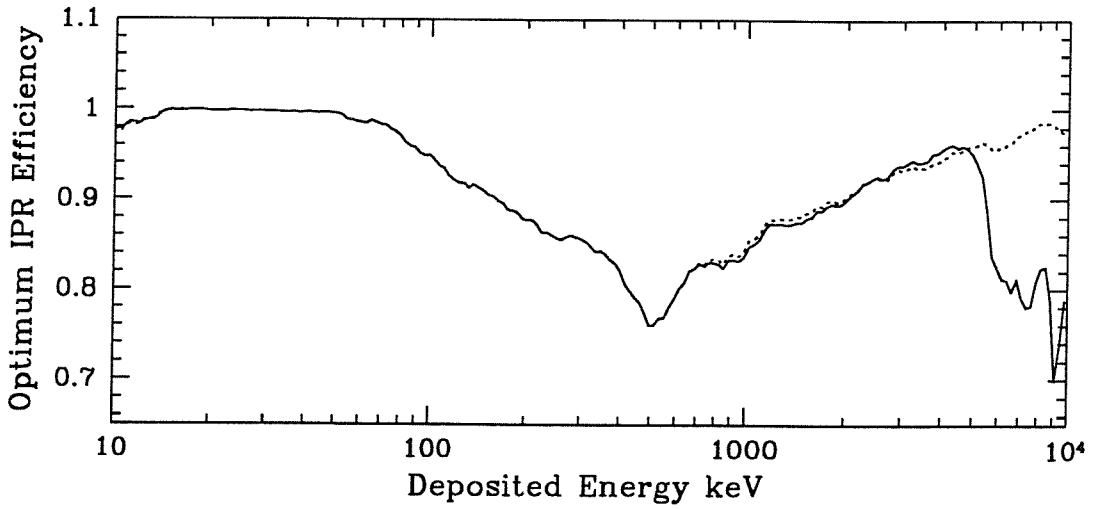


Figure 5.24: As Fig. 5.23 but with the HETs applied.

5.5.3 Spectral Broadening

The successful application of each IPR method to the observed data depends on the accuracy with which *all* interactions in an event can be measured. As will be discussed further in Chapter 7, (Section 7.2) in reality the detector will have a finite spectral resolution the width of which can be determined from the following contributions :

$$W_T^2 = W_D^2 + W_C^2 + W_N^2 \ . \tag{5.2}$$

W_D represents the inherent statistical fluctuation on the number of charge carriers created in the detector readout, either from the scintillation light in the CsI(Tl) detectors, or directly by the incident radiation in the CdTe detectors. W_C represents the contribution from the charge collection variances, and W_N is the electronic noise component inherent to the photodiode and subsequent processing electronics. Fitting to experimental data from prototype EIDOS detector elements, both CdTe and CsI(Tl) [148], has led to the following expressions for the spectral resolution (FWHM) dependence on energy :

$$\text{FWHM}_{CdTe}(\%) = 0.026 + \frac{3.354}{E_D} \tag{5.3}$$

$$\text{FWHM}_{CsI}(\%) = 72.84\ E_D^{-1.0303} \tag{5.4}$$

If spectral broadening does have a detrimental effect on the total IPR efficiency, it is expected to be most significant at the cross-over point between IPR methods. The simplified IPR formula given in Table 5.4 has only two cross-over points at ~511 keV and at ~1722 keV. In addition, spectral broadening should primarily only affect the IPR of high multiplicity events where the quadrature addition of the errors for each interaction may be sufficient to change the IPR method selected. However, on balance it was expected that there would be approximately equal numbers of events lost and gained through the misassignment of IPR methods at the cross-over points and indeed there were no significant (above a few percent) changes in the total IPR efficiency when spectral broadening was introduced.

5.5.4 The Complete IPR Response for EIDOS

Combining the effects of both energy thresholding and spectral broadening, and by using the IPR formula given in Table 5.4, the total instrument IPR efficiency was determined and is shown in Fig. 5.25.

In addition to the sensitivity losses incurred through the IPR process, the sensitivity is reduced further at higher energies by the decrease in detector stopping power. Combining

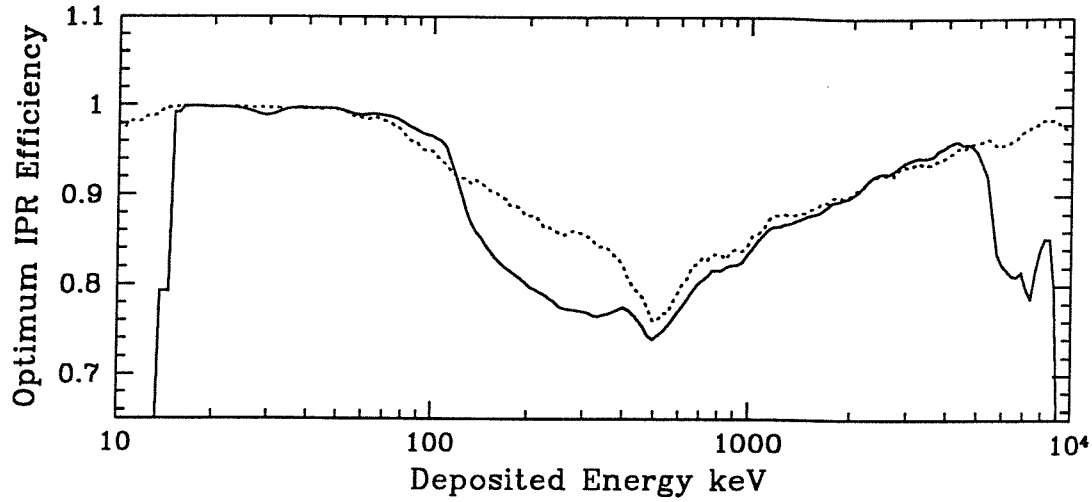


Figure 5.25: The EIDOS IPR Efficiency as a function of deposited energy, including upper and lower thresholds and spectral broadening. The input spectrum was a E^{-1} power law as described in Section 5.3.2 and vetoing has been applied prior to analysis. The dotted line shows the IPR efficiency without thresholds.

the IPR efficiency with the detection efficiency (the percentage of incident photons detected at each energy) gives the IPR detection efficiency shown in Fig. 5.26.

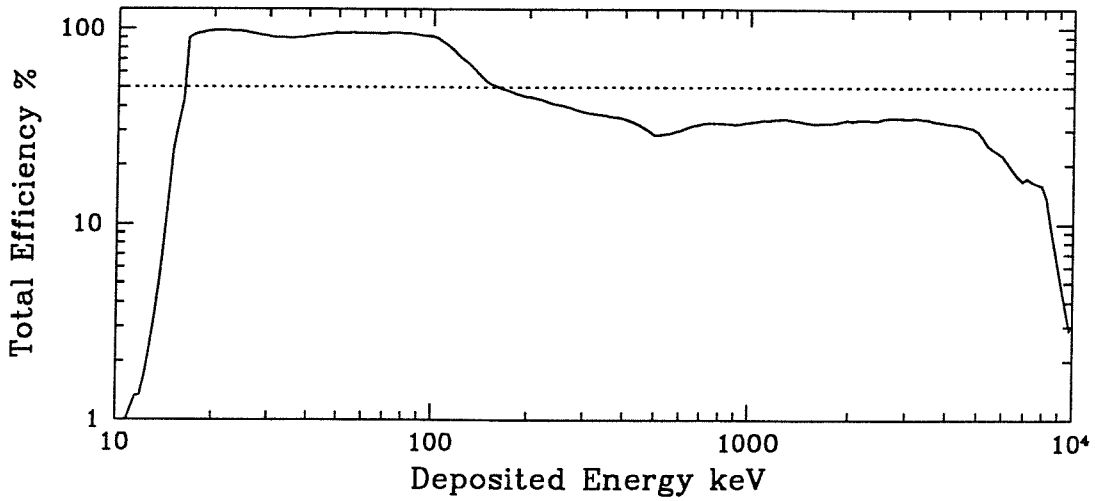


Figure 5.26: The total EIDOS Efficiency as a function of deposited energy, including upper and lower thresholds and spectral broadening.

5.6 The IPR Formula Performance

An important point derived from Fig. 5.25 is that although some events are not reconstructed correctly they are still assigned however, to a pixel. In the following

section the spread in reconstructed photon positions with respect to deposited energy is examined and in addition, the effect that this decrease in position resolution has on both the telescope sensitivity and angular resolution is examined.

5.6.1 Reconstructed Detector Point Spread Functions

In order to assess the effect of IPR on the spatial resolution of the detector the original continuum data was used to generate detector point spread functions (DPSFs) over several energy bands. Specifically, the DPSF was defined as the distribution of reconstructed events about the correct pixel of incidence. Each DPSF was thus determined by binning the incorrectly assigned events onto a 37 element hexagonal grid with the correct pixel of incidence at the centre. Figure 5.27 shows the DPSFs for 5 bands over the EIDOS energy range (0–100 keV, 150–20 keV, 461–561 keV, 1050–1150 keV and 3950–4050 keV).

As expected the DPSFs are at their worse over the 100 keV to 1 MeV range which corresponds to the drop in IPR efficiency seen in Fig. 5.27 over this range. Table 5.5 gives the percentage of incident photons assigned to each of the 'nearest-neighbour' ring of pixels from the central pixel outwards as derived from Fig. 5.27.

Energy Band	P_0 %	P_1 %	P_2 %	P_3 %
1	99.343	0.052	0.018	0.007
2	84.591	1.376	0.375	0.147
3	71.256	2.347	0.772	0.299
4	84.316	1.473	0.348	0.152
5	94.668	0.597	0.101	0.031

Table 5.5: Correct and 'nearest-neighbour' IPR percentages from the DPSFs shown in Fig. 5.27.

5.6.2 The Effect on Instrument Sensitivity and Angular Resolution

5.6.2.1 Sensitivity

Chapter 3 has shown that in order to optimise the available sensitivity, the detector resolution element should, if possible, be around an order of magnitude smaller than the mask element area. Although this is not possible for EIDOS, Section 3.5 describes the technique of 'non-integer sampling' for ensuring that the telescope response is reasonably

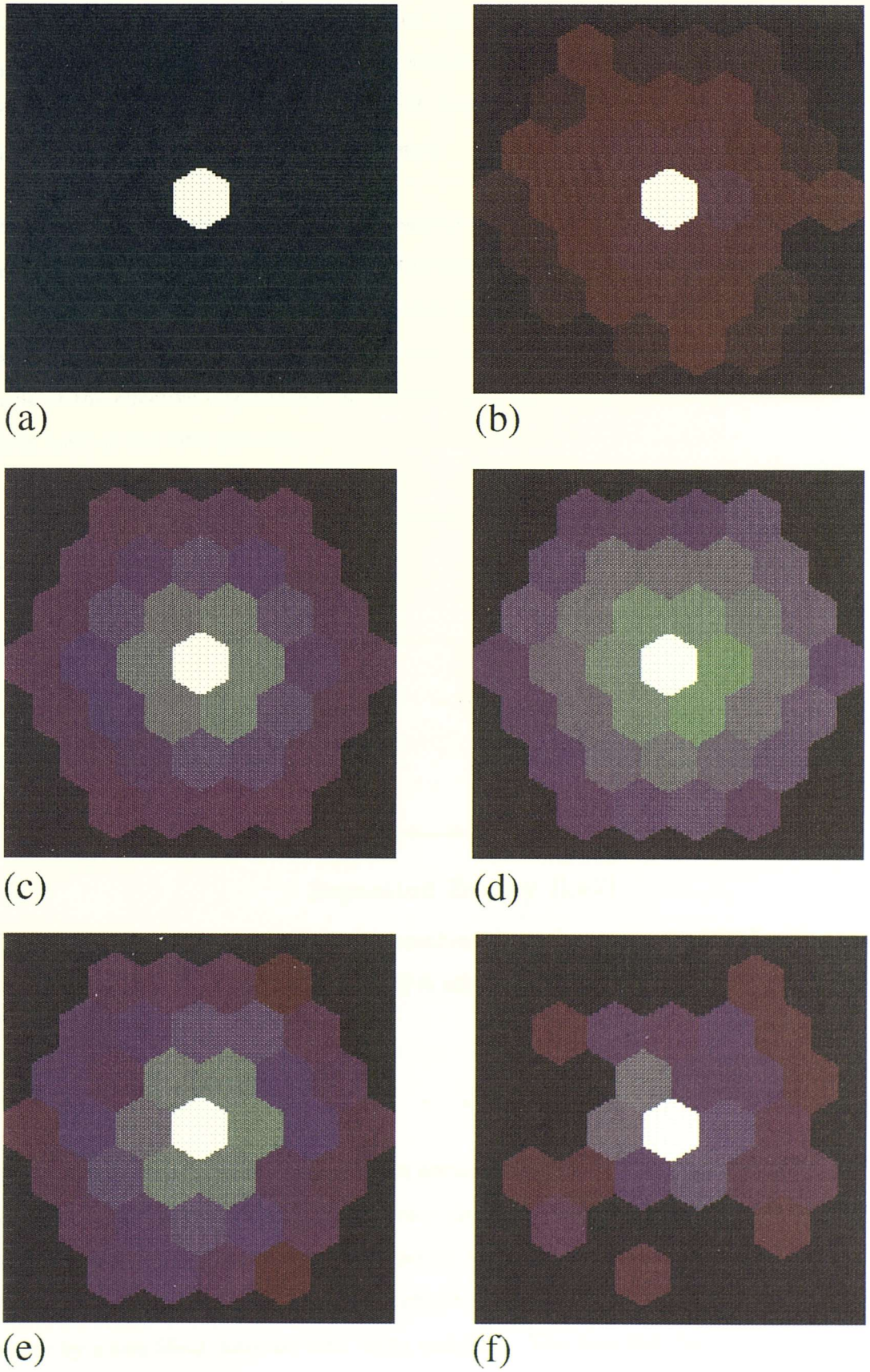


Figure 5.27: DPSFs for the energy bands (b) 0 - 100 keV, (c) 150 - 250 keV, (d) 461 - 561 keV, (e) 1050 - 1150 keV and (f) 3950 - 4050 keV. The ideal response is also shown (a).

uniform. Since the absolute sensitivity of the telescope is dependent on the sampling area ratio only, any increase in the detector resolution size, for example with energy, will degrade the sensitivity further. Thus, the varying DPSF shown in Fig. 5.27 suggests that there will be an additional energy dependent sensitivity variation, as well as the sensitivity drop due to the changing detection efficiency.

The effect of a varying DPSF on the sensitivity was quantified with respect to the reconstructed image figure of merit (FOM) (see Chapter 3, Section 3.2) using a simple Monte Carlo simulation of EIDOS. For the 5 energy bands described above, the relative sensitivity plot shown in Fig. 5.28 is obtained. Note that this loss in sensitivity is a direct result of the effective degradation in detector position resolution as opposed to photons being 'lost' in the IPR process.

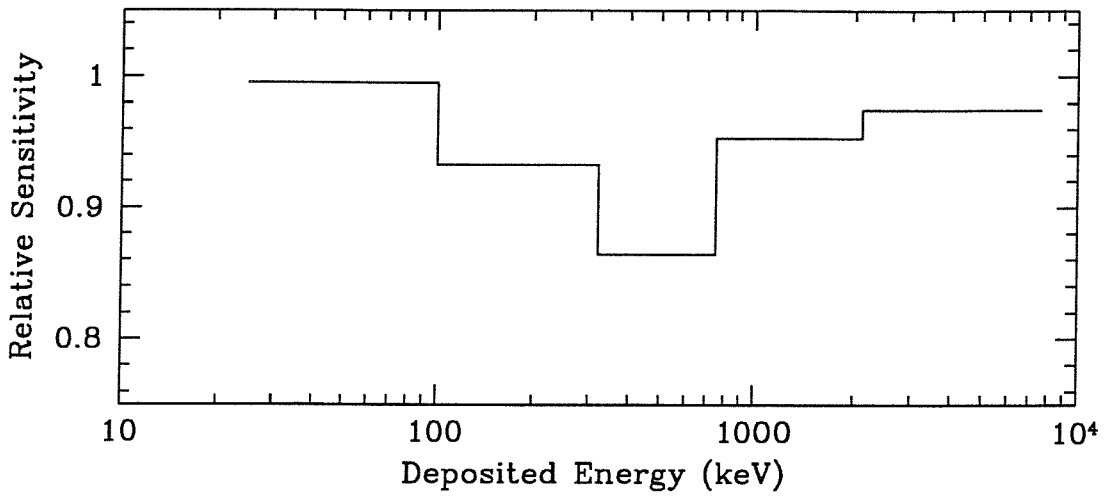


Figure 5.28: The decrease in sensitivity resulting from changes in the DPSF with energy. A sensitivity of 1 is achieved for 100 % IPR efficiency.

5.6.2.2 Angular Resolution

Section 3.2.2.1 briefly mentioned that the nominal angular resolution, determined by the projected mask element size at the detector plane, can only be achieved with an infinite sampling ratio. This can be demonstrated by following the Fourier operations that lead to the reconstruction of an image. Consider the shadowgram recorded from a point source by a non-ideal detector with finite sampling. The recorded data can be expressed as a *convolution* between the ideal shadowgram (obtained with infinite sampling) and the (pixellated) response function of the detector. Subsequent deconvolution involves a *correlation* of the shadowgram with the mask pattern to produce an image. Since both

convolution and *correlation* are commutative the following can be equated :

$$(D * R) \star M = (D \star M) * R \tag{5.5}$$

where D, R and M are the ideal shadowgram, the detector response and the deconvolution functions respectively, and where $*$ denotes *convolution*, and $\star$ denotes *correlation*. Equation 5.5 implies that the system point spread function (SPSF) of the non-ideal telescope is equal to the ideal SPSF ($D \star M$), convolved with the detector response function. Thus, a finite detector resolution should have a detrimental affect on the angular resolution. Figure 5.29 shows the normalised SPSF for each of the energy bands outlined above and as can be seen from the fitted Gaussian profiles, the resolution is poorest at around 511 keV which is a direct result of the poor DPSF at this energy.

5.7 Conclusions

This chapter has demonstrated that for a multiple layer, pixellated detector array combined with a carefully selected IPR algorithm, it is possible to maintain a high position resolution over several decades of energy. Figure 5.25 shows that the total IPR efficiency is always above 74%. The main consequences of a reduction in the IPR efficiency are a loss in sensitivity and a loss in angular resolution, both a direct result of the decrease in position resolution. These effects are summarised in Table 5.6, as a function of IPR efficiency.

Energy Band	Relative Sensitivity	Relative Resolution
0	1.000	1.000
1	0.995	1.076
2	0.933	1.112
3	0.864	1.159
4	0.953	1.112
5	0.975	1.098

Table 5.6: The effects of a changing DPSF on the relative sensitivity and angular resolution.

In Section 3.6 of Chapter 3, a possible approach to the analysis of data from a discrete pixel system was outlined. This approach avoided the use of time consuming iterative

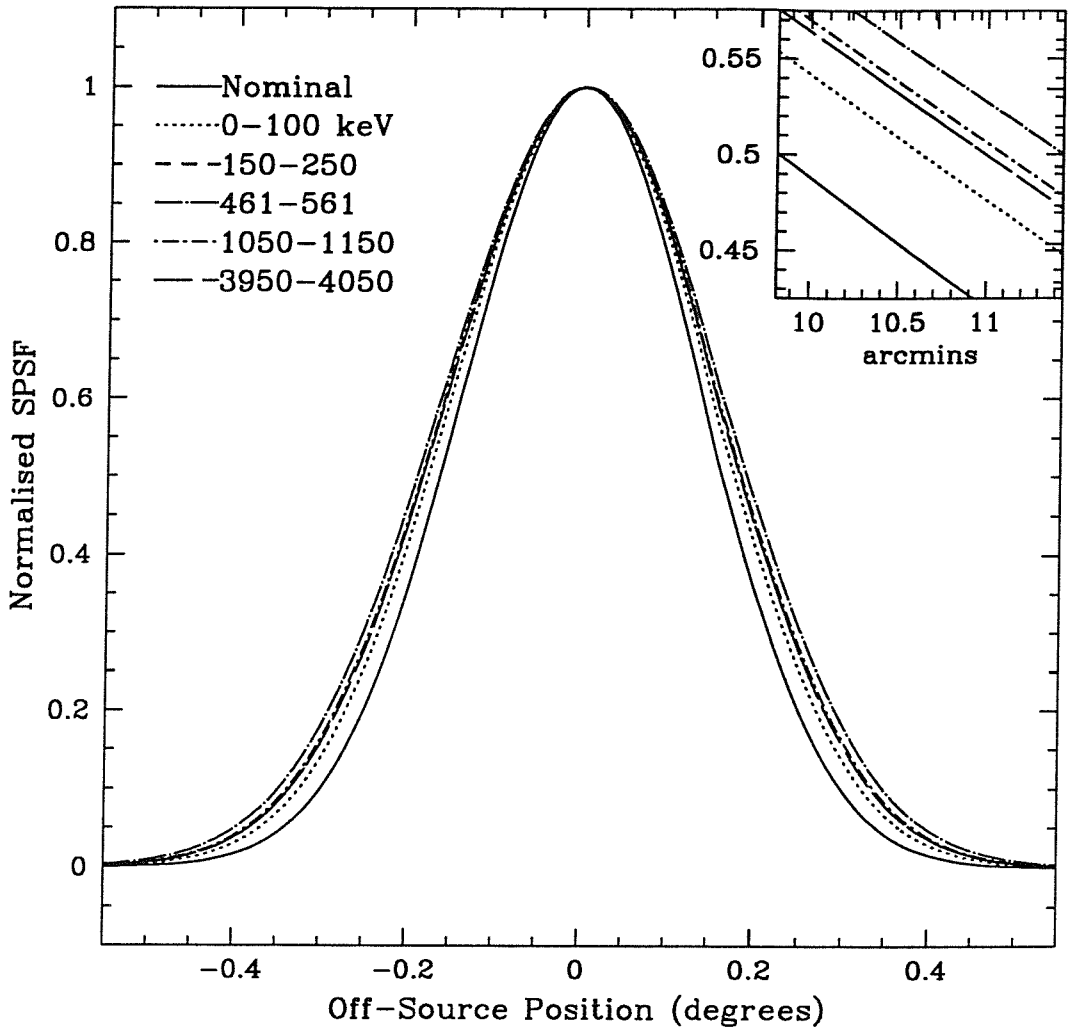


Figure 5.29: Normalised SPSFs for the DPSF energy bands defined above. The nominal angular resolution is also shown. Inset figure shows the HWHM for each energy band.

reconstruction techniques, yet was still capable of producing accurate images in the 'least squares' sense and promoted statistical and spectral analysis that used the raw shadowgram data as opposed to the deconvolved image data. This chapter has outlined the first stage in this imaging sequence, namely the IPR stage. The second stage of analysis is the actual image reconstruction and identification of sources in the FOV, whether point-like or extended in nature. However, as will be seen in the following chapter, the operational environment of a space-borne coded aperture telescope may be such that the sensitivity of the correlation image is dramatically reduced. The question therefore arises, is the correlation image viable in a practical imaging scenario ?

Chapter 6

Imaging Strategies for Extreme Observing Conditions

Overview

The concluding section of the previous chapter raised the question of whether correlation analysis remains a viable option under practical observing conditions. Consequently, in this chapter the performance of a space-borne, pixellated coded aperture telescope has been studied under extreme observation conditions. It is demonstrated that images reconstructed by a simple correlation technique without the use of a mask-antimask strategy, can be degraded severely by the systematic effects introduced under certain conditions. As a result, alternative imaging strategies have been developed that do not rely on a mask-antimask capability. The performance of these strategies has been studied in a controlled way using the GEANT-3 Monte Carlo simulation of EIDOS. The results show that for a range of severe conditions, both observational and telescope based, it is still possible to reconstruct a reliable map of the sky with correlation analysis. On the contrary, not all strategies have a solely beneficial effect on the correlation image, and in the concluding sections of this chapter the side effects of one particular strategy are considered within the framework of the current INTEGRAL imager, IBIS.

6.1 The Need for Alternative Imaging Strategies

For optimally coded systems based on the use of URA coded masks, correlating the recorded shadowgram with a modified version of the mask pattern can produce an artifact free SPSF. The SPSF is simply the practical ACF and can be thought of as the response of the instrument (including image reconstruction) to a non-statistical point source. The only ambiguous features in the SPSF are ghost sources which arise from the cyclic nature of the mask pattern which is repeated to provide an extended or fully coded field of view (FCFOV). In the weak source limit, where the statistical deviations of each detector pixel count are equal and (virtually) uncorrelated, it has been shown that correlation analysis also produces images that are optimum from a minimum χ^2 perspective (see Chapter 3, Section 3.4).

However, in practice the telescope design has to comply with precise engineering guidelines and correlation analysis is not always readily implemented or may be degraded by what is termed 'coding noise' in the SPSF. For EIDOS [148], an attempt to optimise both sensitivity and angular resolution resulted in a mismatch between detector pixel and mask element pitches. Consequently, to enable correlation to be performed, re-binning of both the recorded data and decoding arrays was required prior to deconvolution. Since the re-binning process is only approximate, it may introduce coding noise into the SPSF in the form of non-zero side-lobe fluctuations (see Chapter 3, Fig. 3.16). Although the coding noise is actually only at the 1% level in the SPSF, it can cause considerable problems once the telescope is operational since high background counting rates may dramatically magnify any coding noise.

Thus, in reality coding noise artifacts are unavoidable and may arise from a number of other 'imperfections' in the telescope configuration : large-scale spatial variation in the detector background can produce non-uniformities in the reconstructed image ; pixel to pixel low energy threshold variations can introduce an additional 'random' noise component in the correlation image ; complete detector pixel failure can also introduce systematic artifacts, all of which may appear as significant structures or even conceal weak sources.

In view of the limitations of the basic correlation technique under extreme but nevertheless plausible observing conditions, there are three ways to proceed with image reconstruction. Many solutions to 'deconvolution' problems involve the use of either matrix inversion or

inverse filtering techniques. However, as with all inverse deconvolution methods, even if the inverse transform exists, degeneracy of the response matrix leads to instability problems with the reconstruction. The remaining options can be classified as either iterative reconstruction techniques, or a technique specific to coded aperture imaging, namely Mask-Antimask imaging [210].

The iterative techniques include all forms of iterative maximum likelihood reconstruction, the Maximum Entropy Method (MEM) and a technique analogous to the CLEAN technique used in radio astronomy [211]. The major disadvantage with *all* iterative methods is that they are demanding of computer processing time, by no means a trivial overhead when dealing with spectra that are composed from a few hundred ($\sim$256) images, each containing possibly in excess of $\sim 10^6$ sky pixels. In addition, MEM techniques have the added disadvantage of being intrinsically non-linear [212].

Returning to the basic correlation technique, because the discrete correlation integrals can be performed with speed in the Fourier domain using Fast Fourier Transforms (FFTs), the processing time required for a single image is insignificant in comparison to iterative reconstruction techniques (seconds as opposed to hours). If it was possible to regain the image quality of the ideal correlation image by introducing a method of coding noise suppression prior to image reconstruction, then the correlation technique would clearly be favourable over *all* iterative techniques. This is where the mask-antimask technique, as described below, provides the solution. The mask-antimask technique relies on the 60 degree rotational antisymmetry of the HURA, and enables the suppression of several causes of coding noise prior to deconvolution.

However, failure of the mechanism for rotating the coded mask between the zero and antisymmetric positions is possible, regardless of the level of inbuilt redundancy. The task of producing a reliable correlation image without mask-antimasking, however, is considerably more complicated. The motivating factor behind the following Monte Carlo simulations was the importance of developing alternative imaging strategies for retaining sensitivity both in the event of mask rotation failure for EIDOS, or as routine for a fixed mask telescope. The need for alternative imaging strategies has recently become apparent with the latest iteration of the INTEGRAL imager, IBIS. The IBIS design does not include a mask-antimask capability and so observations may be susceptible to the potential sensitivity losses induced by coding noise. In Section 6.4, possible contingency plans derived to cope with mask rotation failure are applied to a GEANT-3 [207] Monte

Carlo simulation of EIDOS. In Section 6.5, the positive and negative impacts of one possible imaging strategy have been assessed specifically for IBIS.

6.2 Options for Correlation Analysis

Four variants of the basic correlation sequence have been implemented as contingency plans in the following Monte Carlo simulations with the aim of quantifying the performance of each method under extreme observation conditions, however unlikely. Three of the four methods have been proposed for, or implemented in the space-borne telescope environment prior to this work. The fourth option, to be known as Dithered Imaging, was recently proposed for use on the balloon-borne γ-ray telescope, EXITE2 [202]. However, the original concept of dithered imaging was developed as a contingency plan in the event of mask rotation failure for GRASP [198], the predecessor to the INTEGRAL mission concept. The dithered imaging technique has been refined here for use as either a possible contingency plan, or as a routine imaging strategy in extreme observing conditions.

6.2.1 Unprocessed Reconstruction

The simplest option available is a direct correlation of the recorded data with a modified version of the mask pattern. For the ideal geometry and for a well defined observation scenario this is the optimum reconstruction technique (see Chapter 3, Section 3.4.1). However, as the telescope geometry moves away from the ideal case, the system response will be degraded, and the correlation image will become susceptible to coding noise.

6.2.2 Plateau Removal Imaging

Plateau removal prior to correlation has been shown to be effective in suppressing 'coding noise' magnification by high 'dc' background levels [192]. The optimum plateau to be removed can be expressed in terms of both background and source flux values, but in the weak source limit involves simply the removal of the mean pixel level across the detector plane. As would be expected, the success of this method relies on the spatial uniformity of the detector background level and as demonstrated in Fig. 6.4, is severely degraded by

even the smallest (<5 %) non-uniformity.

6.2.3 Flat Field Imaging

Flat field imaging uses non-local observations of 'empty' fields to estimate background profiles for background subtraction. To ensure that the flat field itself adds little in the way of statistical fluctuations to the final image, flat fields are created from observations that are at least an order of magnitude greater in exposure time than the object field. However, the integration time for flat fields can be reduced with the use of smoothing techniques [213]. Although smoothing is beneficial if the detector background has a large scale spatial distribution, for a pixellated detector where pixel to pixel background variations may be significant, there will be little advantage in smoothing the flat field. Recent developments of wavelet smoothing techniques may overcome some of these problems [202] by enabling smoothing on a range of scales, but again for a pixellated detector there may be difficulties in suppressing pixel to pixel variations. For the flat-fielding strategy employed in Section 6.4, the flat field integration time used was an order of magnitude larger than the source observation time.

6.2.4 Mask-antimask Imaging

Mask-antimask imaging, as mentioned in Section 6.1, is capable of removing all background non-uniformities so long as they are temporally invariant with respect to the mask rotation time-scale. This includes the removal of pixel to pixel variations, and also the partial suppression of dead pixel artifacts. The success of this technique relies solely on the fact that the effects to be removed are unmodulated by the mask pattern. Thus, an observation shadowgram is composed from mask-antimask shadowgram pairs, the first taken with the mask in the 'zero' position, the second with the mask rotated to the antisymmetric position (the mask patterns used here are 60 degree rotationally antisymmetric). When the second shadowgram is subtracted from the first, background non-uniformities are removed and only the source modulation remains.

EIDOS was designed to make use of mask-antimask pairs to remove the effects of systematic background variations. For the proposed Highly Eccentric Orbit (HEO) of INTEGRAL [148] any temporal background variations are expected to occur on

comparatively long time-scales. Analysis of background data from the SIGMA spacecraft in a similar orbit suggests that this is indeed the case (excluding times of solar activity). Thus, with mask rotation repeated on a 10 minute time-scale, which is well within the estimated time-scale of background variations, most systematic background effects can be removed.

6.2.5 Dithered Imaging

The dithered imaging procedure involves making N observations of the source distribution, each performed with a slightly different telescope pointing. Assuming that any image artifacts are a consequence of factors that are independent of the telescope attitude, a superposition of all shadowgrams will produce a detector map or flat-field of the systematic artifacts. In essence, by adding the shadowgrams from several pointings an attempt is made to demodulate the source flux. The pseudo flat-field is then normalised and subtracted from each of the individual pointings to remove the dominant 'dc' background level and any spatially dependent background variations. Each individual pointing is then deconvolved to produce a sub-image, which when shifted according to the telescope attitude and superposed with the remaining $N-1$ sub-images, yields the final 'artifact free' image. Figure 6.1 demonstrates graphically the dithered imaging process.

For a limited number of pointings the flat-field will still contain residual modulation in the source flux and this will be subtracted in addition to the background level from each pointing (standard method). Thus, with a limited pointing strategy a degradation in the deconvolved peak height is expected that is inversely proportional to the number of pointings used. However, for strategies using a low number of pointings, the degradation in the peak height can be avoided by ensuring that the flat-field to be subtracted from observation P_i is constructed without using P_i (optimum method). Additionally, the statistical fluctuations in the final image are also expected to be dependent on the number of pointings since each reconstructed image is the result of combining, subtracting and re-combining pixel values, all of which have Gaussian fluctuations on their count rates. Equations 6.1 and 6.2 were derived for the degraded peak height and the increase in noise respectively as a function of the number of pointings employed, for both standard (S) and optimum (O) methods of flat-field construction. For these derivations it was assumed that the source was kept within the FCFOV for all pointings, and the pointing step size was greater than or equal to the FWHM of the SPSF :

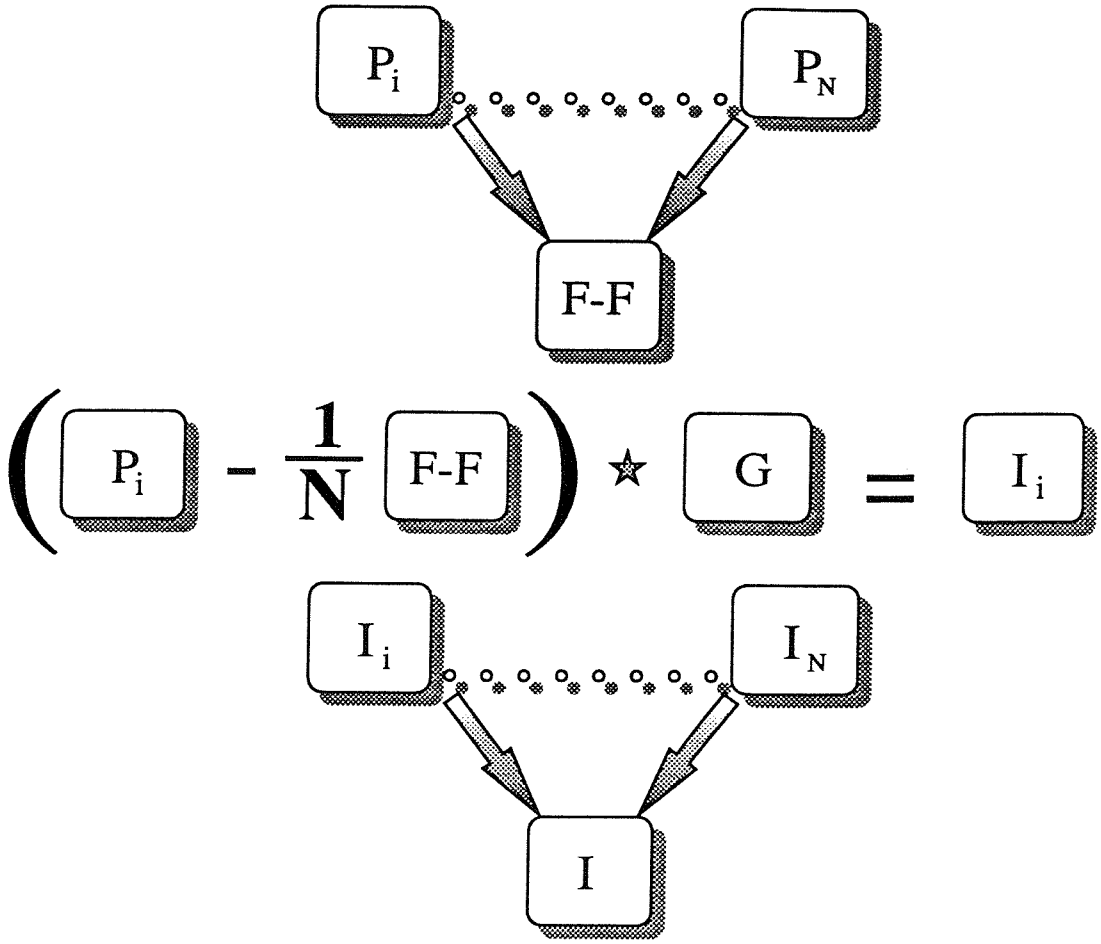


Figure 6.1: A schematic diagram showing the stages in the dithered imaging strategy. N observations, P_i...P_N, of the source(s) are taken, each with a slightly different pointing. The N observations are then combined *before* deconvolution to produce a flat-field (F-F). The normalised (with N) flat-field, and any systematic effects, are subtracted from each pointing, and then deconvolved (denoted by *) with the decoding array, G. This produces N images, I_i...I_N, of the source(s) with the systematic effects suppressed. Because the pointing strategy is known, the images can be combined to produce the final artifact-free image of the source(s), I.

$$\begin{aligned} \text{Peak height, } P_S &= \left(1-\frac{1}{N}\right)(v-k)S \\ P_O &= (v-k)S \end{aligned} \tag{6.1}$$

$$\text{Total noise, } \sigma_S = \left(1+\frac{1}{N}\right)^{1/2}((v-k)S+vB)^{1/2} \tag{6.2}$$

$$\sigma_O = \left(1+\frac{1}{N-1}\right)^{1/2}((v-k)S+vB)^{1/2}$$

$$\text{thus, relative sensitivity, } R_S = \frac{\left(1-\frac{1}{N}\right)}{\left(1+\frac{1}{N}\right)^{1/2}} \tag{6.3}$$

$$R_O = \frac{1}{\left(1+\frac{1}{N-1}\right)^{1/2}} \ .$$

N is the number of pointings, $v - k$ is the number of open elements in the mask pattern, S is the source flux per detector pixel and B is the background per detector pixel. It should be noted that if the number of pointings, N equals the order of the mask and the pointing step size is equal to the SPSF FWHM, the resulting flat-field will be ideal in the sense that the source flux will have been completely demodulated. For the simulations in the following sections, the standard method of dithered imaging was employed in order to provide an absolute minimum sensitivity prediction. Figure 6.2(a) shows the degraded peak height and noise curves respectively for the standard method of flat-field subtraction, which when combined (Eq. 6.3) give the relative sensitivity curve shown in Fig. 6.2(b). For comparison, the relative sensitivity obtained using the optimum method of flat-field construction is also shown. Additionally, it should be noted that the pointing strategy *does not* have to coincide with the telescope geometry. For example, a square geometry telescope can quite successfully employ a hexagonal dither strategy, so long as each individual pointing shift is at least the SPSF FWHM in size.

In reality, more than simply the number of pointings needs to be considered and in fact there are two additional factors that will degrade the dithered imaging sensitivity. These are (1) a loss of effective detector area as the source moves off-axis by the factor $\cos\theta \cos\phi$, where θ and ϕ are the off-axis source angles, and (2) a loss in 'coding' equal to the percentage of flux, fully coded by the mask, lost if the pointing sequence requires imaging outside the FCFOV. In addition, unlike the mask-antimask technique, dithered imaging will not remove artifacts induced by spacecraft shadowing from strong off-axis sources, or indeed any artifacts that are dependent on the telescope attitude.

The simulations described in the following sections used a 7 point hexagonal dither strategy with a 1 degree step size (outside the SPSF) and with all pointings in the FCFOV. The $\cos\theta \cos\phi$ factor was negligible for this dither strategy and so was not included.

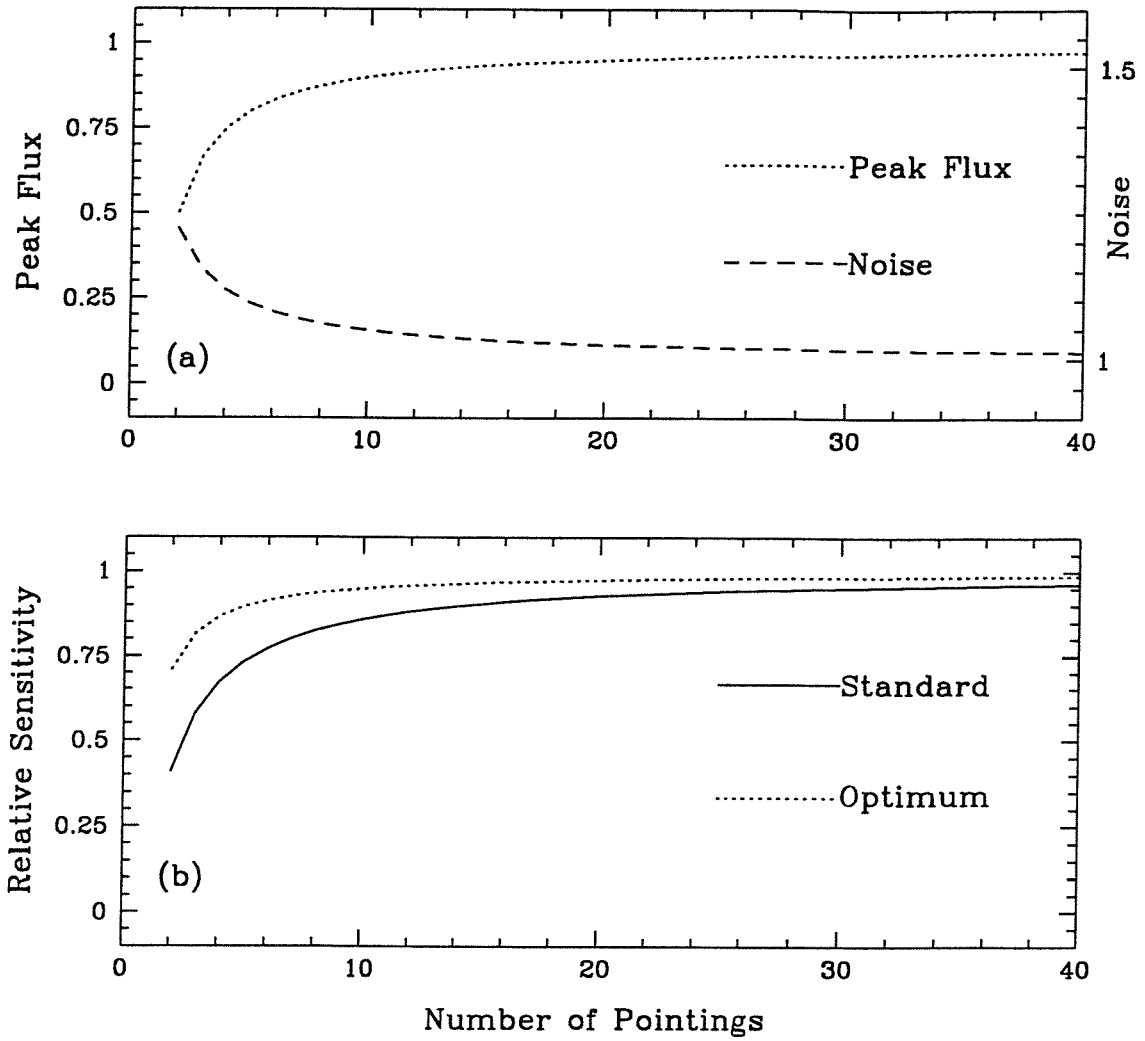


Figure 6.2: (a) The increase in reconstructed peak height and decrease in Gaussian noise in the final image as a function of the number of pointings for the standard dithered imaging strategy. (b) The combined effects of dithered imaging on the relative telescope sensitivity for standard (solid line) and optimum (dotted line) flat-field construction methods.

6.3 The Simulated Observations

The GEANT-3 [207] model of EIDOS was used to simulate observations of a 10 mCrab point source, positioned on-axis. The integration time for all simulations was 10^5 seconds and the observational energy range was 15 keV to 10 MeV. The source spectral profile was based on the $dF/dE = 0.24\ E^{-2.3}$ ph cm^{-2} s^{-1} keV^{-1} model of the total Crab emission [214]. The background spectral profile was taken from earlier modelling of the telescope for a known orbital environment [148] (see Chapter 3, Fig. 3.2). However, since generating the background profile with GEANT is immensely time consuming,

background counts were added after image convolution. Background modelled in this way retained the correct spectral profile and also provided a convenient means of introducing any number of systematic effects into each simulated observation. With the basic observation scenario set in place, a number of extreme observing conditions were then introduced, the effects of which on each of the above correlation methods are discussed in the following sections.

6.4 Extreme Observing Conditions

The image quality in the following simulations was assessed with a figure of merit (FOM) related to the true signal to noise ratio (SNR) of the reconstructed source in the ideal case. This approach also allowed the degradation of the reconstructed image to be objectively quantified in the presence of coding noise artifacts. Specifically, the FOM was defined as the ratio of the deconvolved source flux above the standard deviation of the surrounding Gaussian (and subsequently systematic) background fluctuations :

$$FOM = \frac{s_p}{\frac{1}{N}\sum_{i=1}^{N} s_i^2 - \frac{1}{N^2}\left(\sum_{i\neq p}^{N} s_i\right)^2} \quad , \tag{6.4}$$

where s_p is the peak reconstructed pixel flux, N is the number of reconstructed sky pixels outside the source region, and the denominator is simply the rms deviation of the reconstructed background noise.

6.4.1 Coding Noise Artifacts

Fig. 6.3 shows the results of applying all correlation methods to the recorded data for observations made with detector background levels increasing from zero to the expected space-borne levels. The loss in sensitivity for the unprocessed image reconstruction is a direct consequence of the magnification of SPSF artifacts. All imaging strategies are successful at removing the 'dc' background component and therefore also the magnified coding noise, restoring the source to the expected level of significance. It should be noted that the difference in FOM between the near-ideal mask-antimask and dithered imaging strategies (a factor of 1.25) agrees with the predictions of Eq. 6.3. Figure 6.3 suggests that without the mask-antimask capability, a simple plateau removal prior to deconvolution

can suppresses the coding noise artifacts. However, if the spatial distribution of the detector background is more complex, this simple pre-processing technique may not be so effective.

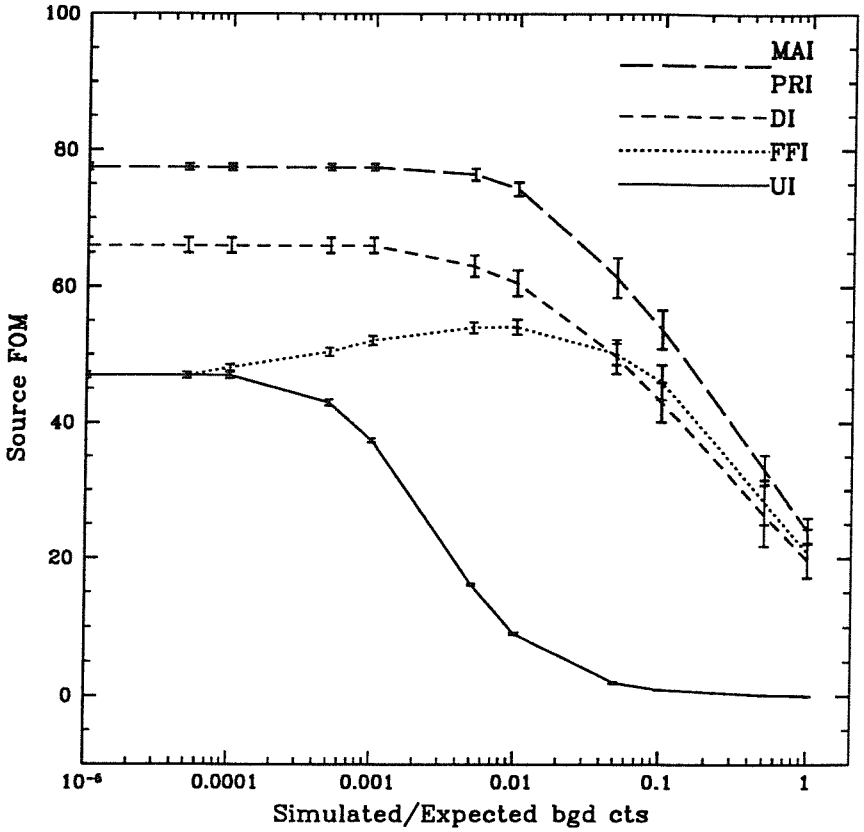
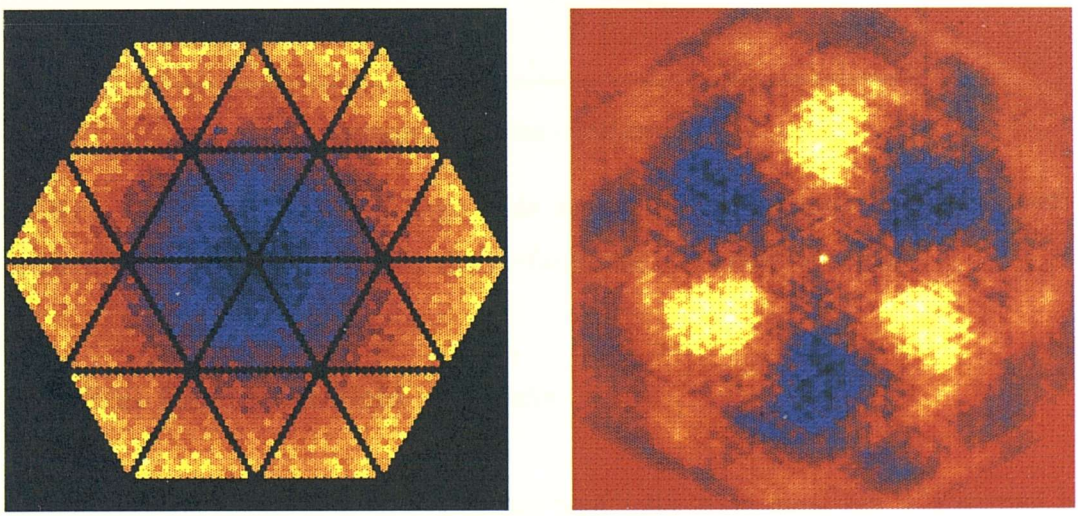


Figure 6.3: The performance of each correlation sequence for background magnified SPSF artifacts. (PRI) plateau removal imaging, (FFI) flat field imaging, (DI) dithered imaging, (MAI) mask-antimask imaging, and (UI) unprocessed imaging.

6.4.2 Background Non-uniformity Artifacts

Just as the fluctuating side-lobes of the telescope SPSF result in a non-uniform reconstructed background level, any initial spatial variation in the background count will also propagate through to the correlation image. Here, simple plateau removal can no longer rectify the image non-uniformity, which is present in both the partially coded (PC) and FCFOV. Without the use of mask-antimasking, the only alternative technique that has been applied successfully to a space-borne coded aperture telescope is the technique of non-local flat-fielding as used on SIGMA [213]. However, as will be discussed further in Section 6.4.5, this technique may be inappropriate for coded aperture telescopes employing pixellated detector planes.

The primary background components in both the Low Earth Orbit (LEO) and Highly Eccentric Orbit (HEO) environments can be divided into two categories, namely gamma-ray induced and particle induced background. The gamma-ray induced component can be categorised further into (a) cosmic diffuse background, (b) atmospheric gamma-rays and (c) locally produced gamma-rays. The particle induced component can be classified according to the 'parent' particles and is either (a) proton induced (spallation background) or (b) neutron induced. All of these background components will be anisotropic to some extent, both temporally and spatially. Consequently, according to the telescope configuration and attitude (pointing and orbital), anisotropy in the external background components will ultimately result in detector background non-uniformity. For example, in the proposed HEO environment of INTEGRAL, a significant cause of background non-uniformity may be the non-uniform distribution of passive material around the detector plane. In the following simulations to represent the background non-uniformity induced by the passive material distribution, a low frequency radial sinusoidal background profile has been used. Figure 6.4(a) shows the detector background profile simulated and Fig. 6.4(b) shows the effect it has on the reconstructed image.



(a) (b)

Figure 6.4: (a) A simulated shadowgram showing the spatial profile of the detector background for the simulations in section 6.4.2. (b) The effect on the plateau removed, deconvolved image.

Figure 6.5 shows the reconstructed source FOM with respect to the background non-uniformity for each of the four correlation techniques. The percentage of non-uniformity is defined as the ratio of the background flux variation to the mean detector background rate. The simulations were repeated for different frequency spatial variations, from which

it was concluded that low frequency variations are most disruptive to the correlation map. However, as Fig. 6.5 suggests, both dithered imaging and flat-field imaging are capable of producing artifact free correlation images in the presence of large-scale background variations.

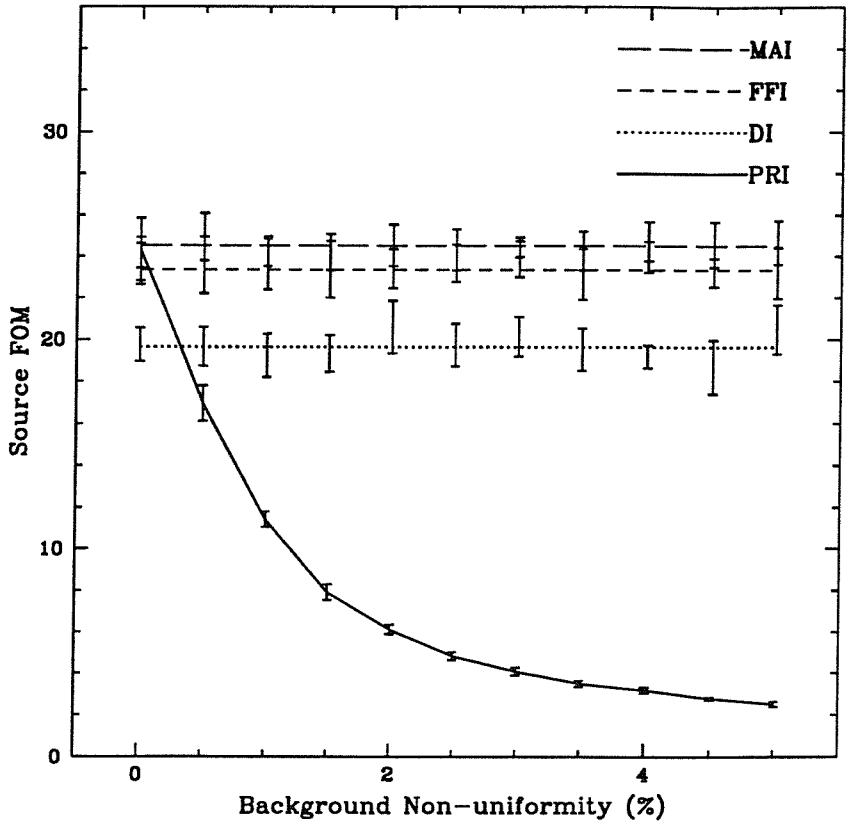


Figure 6.5: As Fig. 6.3 but for large scale spatial non-uniformities in the detector background. The dotted curve shows the performance of plateau removal.

6.4.3 Pixel-Pixel Threshold Artifacts

The low energy threshold of both scintillator and solid state detectors is governed by two factors. The first is the electronic noise component, intrinsic to the detector and subsequent processing electronics. The second is the overall conversion efficiency (OCE) of the crystal/detector, which is a measure of the number of electron-hole pairs produced in the detector per keV of detected radiation. Thus the threshold is simply given by the ratio of the electronic noise to the OCE at some confidence level. For EIDOS, channel to channel variation in electronic noise, coupled with pixel to pixel variation in OCE would have resulted in threshold variations across both the CdTe and CsI(Tl) layers. For the expected EIDOS background spectrum, the variations in threshold were expected to

contribute significantly to the pixel to pixel background variations. For the simulations performed here only the CdTe threshold variations have been included and consequently the effect of pixel-to-pixel non-uniformities on the correlation image is conservative.

Figure 6.6 shows the effect of a Gaussian spread in the 15 keV CdTe threshold on the reconstructed image for each correlation method. Again a straight plateau removal of the 'dc' background level is unsuccessful for all but the smallest of background variations. For larger pixel to pixel variations either flat field or dithered imaging is required to produce a reliable correlation image.

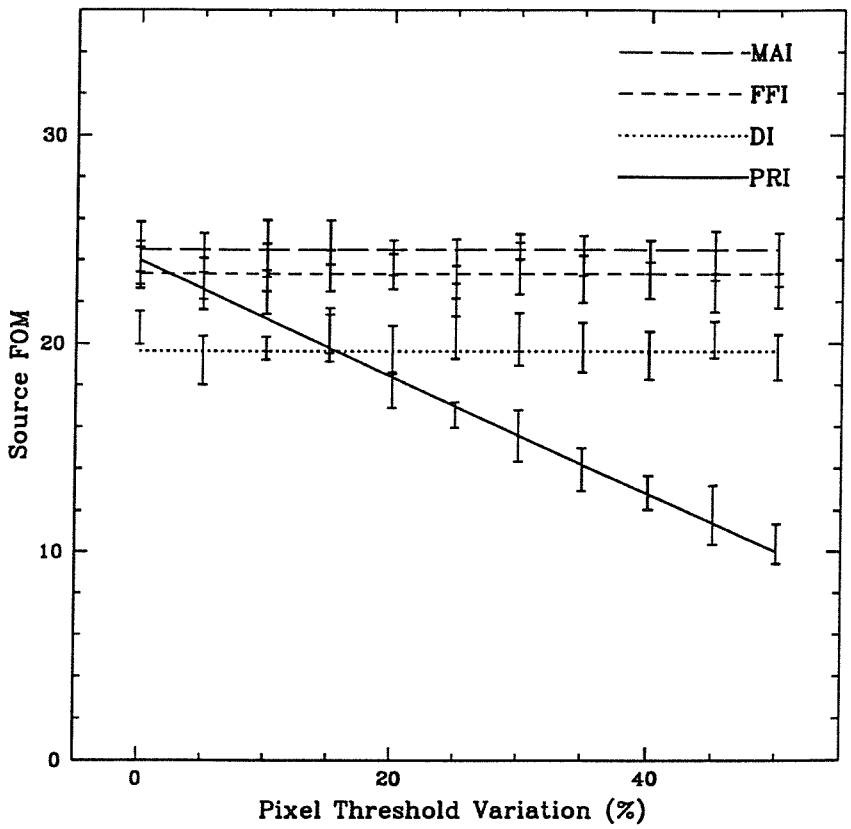


Figure 6.6: As Fig. 6.3 but for pixel to pixel variations in the detector background.

6.4.4 Dead Pixel Artifacts

Complete detector pixel failure, whether random or localised in position, will affect the correlation image by introducing systematic artifacts into the SPSF. The resultant undersampling of the mask pattern will be evident in the modulation transfer function [1] as

[1] The MTF is simply the Fourier transform of the SPSF - the ideal SPSF, a δ-function, has a perfectly flat MTF.

'holes' at those spatial frequencies lost in the undersampling. The inter-module structure of the EIDOS detector plane already degrades the SPSF in this manner, here the objective is to determine the effect of both random and localised pixel failures on the correlation image.

6.4.4.1 Random Pixel Failures

Completely random pixel failures were not expected to degrade the correlation image significantly for three main reasons : (1) with respect to increased systematic artifacts in the SPSF, the size of the detector array (2880 CsI(Tl) pixels) used in these simulations suggests that for single pixel failures, the induced coding noise will be negligible in comparison with the reconstructed source peak ; (2) artifacts of greater significance should be expected if the dead pixel percentage is large enough, although the random nature of the failures suggests the artifacts will 'smooth out' in the SPSF ; (3) although magnification of the SPSF artifacts by high background levels will almost certainly disrupt the unprocessed correlation image, it has been demonstrated (Fig. 6.3) that the four alternative correlation methods are all capable of suppressing these artifacts. Thus, the only major loss in image quality will be a direct result of the loss in sensitive area, which is proportional simply to the dead area percentage. This is supported by the simulation results shown in Fig. 6.7, where a factor of $\sim$1.25 is lost in FOM for a 20 % pixel failure. Note the additional loss in FOM for both flat-field and dithered imaging strategies as a result of the additional Gaussian noise from the flat-field.

6.4.4.2 Localised Pixel Failures

The EIDOS detector plane is divided into 24 modular units, each containing 120 hexagonal pixels arranged in a triangular array. Complete failure of one or more of these detector modules may potentially be more serious for a given dead pixel percentage than for random pixel failures. The systematic effects from each failure in a localised group of dead pixels will all have a similar strength and structure in the deconvolved image. Consequently, the combined artifacts will be stronger than if the same number of dead pixels were randomly distributed. However, as with random dead pixel artifacts, the dead module artifacts will only be significant in the correlation image if no attempt is made to remove the 'dc' background level.

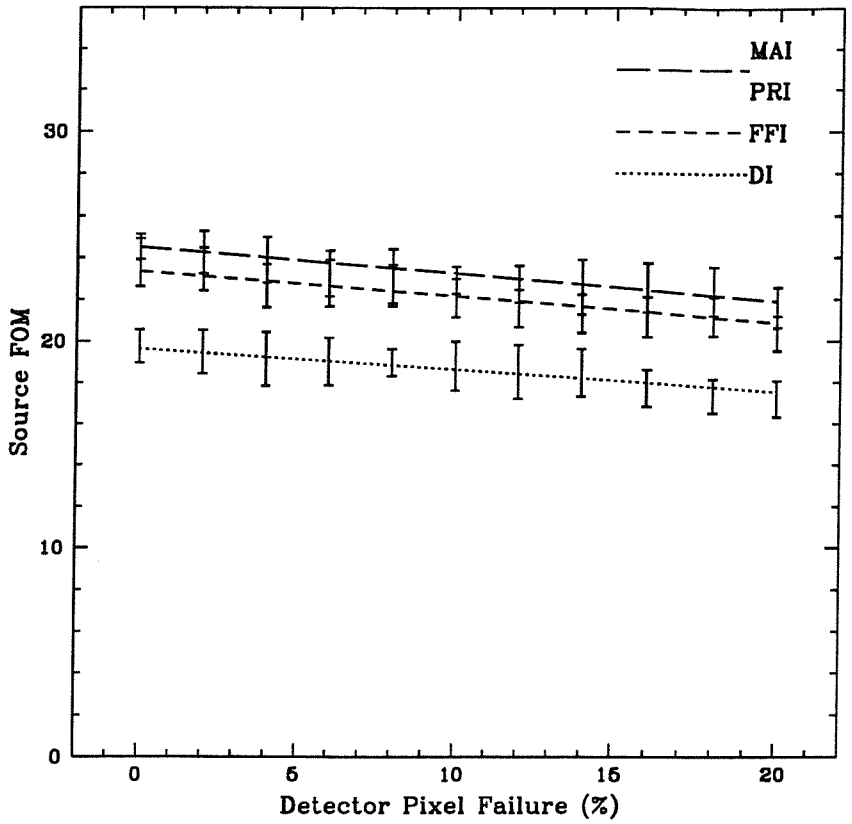


Figure 6.7: As Fig. 6.3 but for the artifacts induced by randomly dead pixels.

There is, however, one situation where dead pixel artifacts may become significant in the correlation image, even after contingency plans have been applied. This occurs when the source flux is large enough for the systematic artifacts to be comparable to the background statistical fluctuations. Thus, the presence of a strong source in the FOV can reduce sensitivity by concealing weak sources below the systematic artifacts. Similar SPSF artifacts are introduced if the source lies in the PCFOV. Again, the artifacts arise from undersampling the mask pattern, but can be removed as mentioned in Section 6.1, by CLEANing the correlation image.

6.4.5 Discussion

The above simulations have demonstrated that the extreme observing conditions encountered in the space-borne telescope environment may severely degrade the sensitivity of a pixellated coded aperture telescope. Without a mask-antimask capability, it is clear that in certain situations the only viable options for image reconstruction are flat-field imaging and dithered imaging. Artifact removal by direct flat-fielding is effective in that

if the integration time is long enough, the subtraction contributes little to the statistical noise of the final image. The obvious disadvantage is that in building up the flat-field, observing time is lost. There are ways around this which include smoothing a lower integration time flat-field, or more recently, wavelet smoothing to retain flat-field features on several spatial scales [202]. A second disadvantage of all non-local flat-field methods is that temporal background variations on time-scales equal to the source integration time will not be removed. Finally, but perhaps of most importance, the combination of increased sensitivity and FOV for the next generation of coded aperture telescopes suggests it may be increasingly difficult to actually find an 'empty' field to observe.

For a carefully selected pointing strategy, the dithered imaging option thus has two main advantages over the flat-field option. Firstly, observing time is not wasted in building up flat-field data, since source and background data are collected simultaneously. Secondly, where temporal background variations have similar time-scales to the observation time, the dithered flat-field will provide a more accurate spatial profile of the background than a non-local flat-field. Figures 6.3 to 6.7 have demonstrated that dithered imaging is capable of suppressing the artifacts induced by either defects in the SPSF, or non-uniformities in the detector background profile.

6.5 The INTEGRAL Spacecraft Dithering – Negative Impacts on the Performance of the Imager, IBIS

The previous sections have highlighted the beneficial effects of a dithered imaging strategy on the performance of EIDOS in the event of mask rotation failure. In this section the effects of dithered imaging on sensitivity are discussed specifically for the current INTEGRAL imager baseline (IBIS), where in addition to the proven advantages of dithering, the adverse effects outlined in Section 6.2.5 are elaborated upon.

The configuration of the spectrometer instrument (SPI) onboard the INTEGRAL spacecraft means that dithered pointings of a source are necessary to accurately reconstruct the sky distribution. Consequently, a number of pointing strategies have already been suggested which enable artifact-free imaging with SPI [215]. Fortunately, as the previous sections have demonstrated, the performance of a fixed mask telescope may be enhanced with respect to the removal of systematic coding noise by using a dithered

imaging strategy. Thus, if the INTEGRAL spacecraft is to be dithered, it is important that any negative effects of such a strategy are quantified. In Section 6.2.5 two factors that may degrade the off-axis sensitivity were highlighted, namely, the loss in source coding outside the FCFOV, and the geometric loss in sensitive area. Figure 6.8 shows for the IBIS baseline [185], the loss in coding and the loss in sensitive area for off-axis sources. However, as will now be discussed, for IBIS there may be a further loss in sensitivity for off-axis pointings.

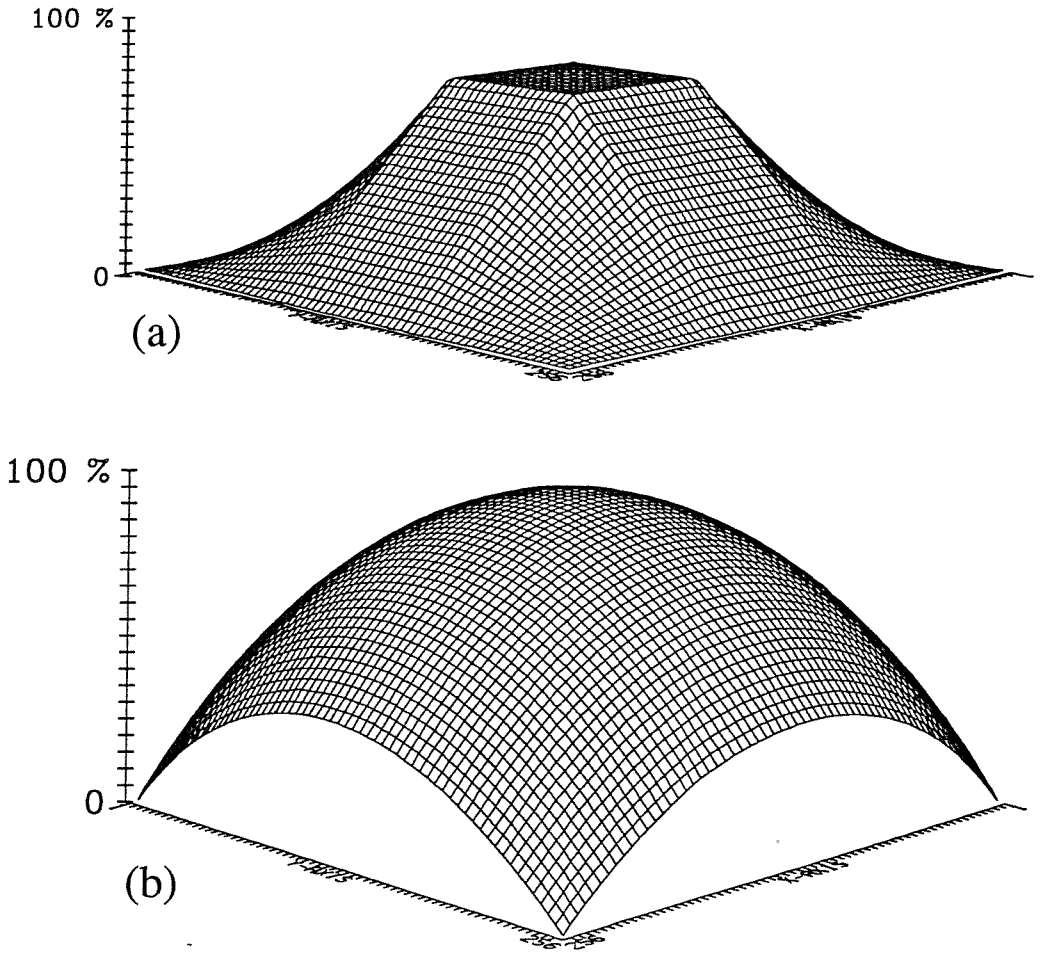


Figure 6.8: The loss in off-axis sensitivity due to the loss in coding (a), and the loss in sensitive area (b).

The current IBIS design employs a tantalum collimator for the reduction, to an acceptable level, of the Cosmic Diffuse Background (CDB) contribution, which would otherwise overwhelm other background components at low energies. There are currently two collimator options for IBIS. The first option is a standard 'cell' collimator, 6 cm tall and 0.92 cm in pitch (equal to the CsI(Tl) pixel pitch), whereas the second option is a single 'tube' collimator which extends from the outside of the detector plane to the

outside of the mask. Both collimators, in addition to reducing the aperture background flux, will also reduce the source flux reaching the detector if the source is off-axis. Thus, at low energies (below 200 keV), IBIS will be subject to an additional loss in sensitivity for off-axis sources according to the response of the collimator. Figure 6.9 shows the standard and tube collimator responses.

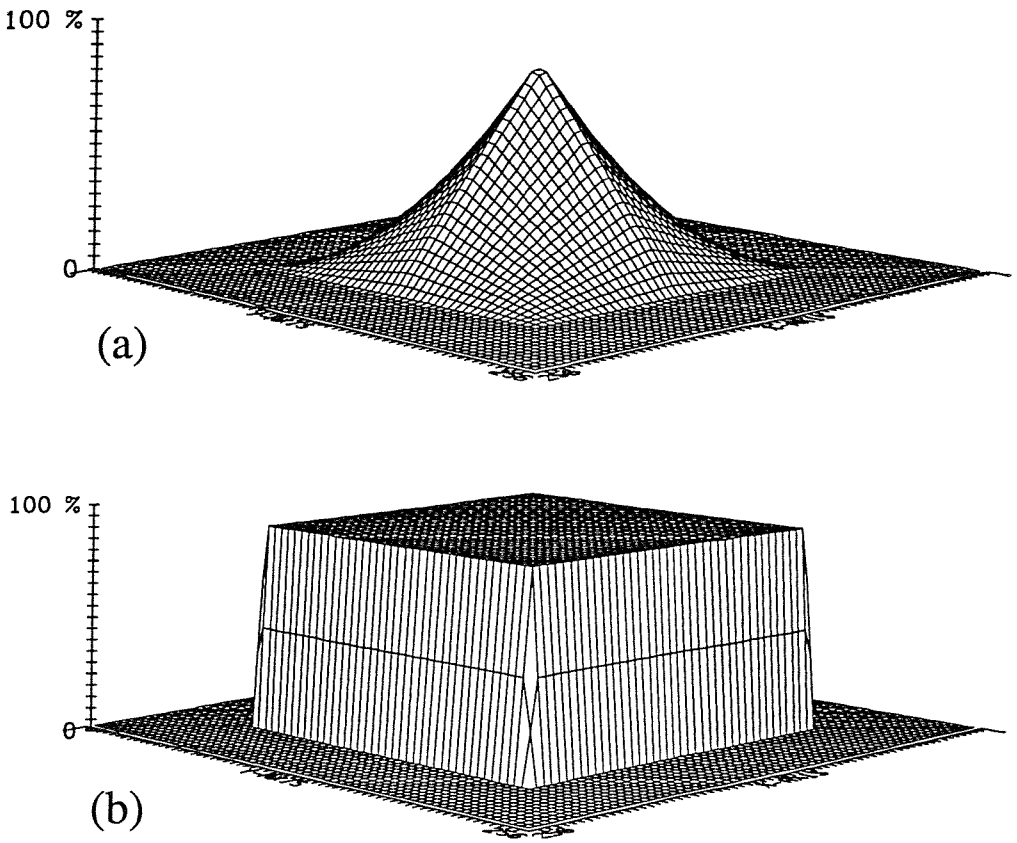


Figure 6.9: The loss in off-axis sensitivity due to the cell collimator response (a), and the tube collimator response (b).

As mentioned, several pointing strategies have been suggested to ensure the sensitivity of SPI is optimised. There are currently seven dithering strategies grouped into three categories depending on the objective of the SPI observation (single point source, multiple point sources, extended emission). In view of the above sensitivity losses, the relative sensitivity of IBIS was assessed for each of these proposed dithering strategies and for both standard and tube collimators. In addition to the coding, geometric area and collimator response losses, these calculations also include the dithering sensitivity loss given by Eq. 6.3. Table 6.1 describes the pointing strategies employed and the resulting sensitivities for each collimator option for a source initially positioned on-axis.

Strategy number	Pointing grid	Grid pitch	Pointings	Sensitivity STD Coll'	Sensitivity Tube Coll'
0	–	–	1	1	1
1a	Hexagonal	2^o	19	0.566	0.922
1b	Hexagonal	2^o	7	0.617	0.801
1c	Hexagonal	2^o	37	0.437	0.908
2a	Rectangular	2^o	81	0.216	0.735
2b	Rectangular	2^o	16	0.542	0.908

Table 6.1: The sensitivity of IBIS for five of the dither strategies proposed for use with the INTEGRAL spectrometer, SPI.

It is clear from Table 6.1 that if the standard collimator is used then the overall sensitivity of IBIS is severely compromised, reaching a minimum of 22 % for dither strategy 2a. On the contrary, if the tube collimator is used instead, the sensitivity never drops below 70 % of the single pointing sensitivity.

6.6 Conclusions

In the event of mask rotation failure for EIDOS, or for a fixed mask system such as the current INTEGRAL imager IBIS, it is apparent from the above simulations that dithering the spacecraft attitude would have a positive impact on the performance of the telescope. Dithering enables the removal of systematic effects from the detector plane which might otherwise severely degrade the imaging performance. However, for a practical telescope design which includes the use of a collimator to reduce the CDB, dither strategies need to be selected carefully in order to avoid severe sensitivity losses at low energies. The need for verification of dithered imaging in a real imaging scenario has led to the construction of a prototype imager based on the CsI(Tl)-photodiode layers of EIDOS. The detector used for the prototype has many of the characteristics of the EIDOS detector including the systematic defects that can degrade the correlation image. The design, construction and results of laboratory observations with the prototype imager are the subject of the following chapter.

Chapter 7

Imaging with a 37-element Prototype Detector

Overview

Theoretical analysis and Monte Carlo simulations have led in the previous chapters to a thorough characterisation of the performance of a pixellated coded aperture telescope. In addition this work has resulted in the development of the analysis procedures required to extract in an optimum way source flux and spectral information from the raw data. Many of these analysis techniques, for example, the IPR techniques developed in Chapter 5 and the imaging strategies developed in Chapter 6, have been successfully demonstrated on simulated data but have yet to be applied to a real pixellated imaging system. In order to evaluate both the engineering and scientific obstacles encountered in the construction of and imaging with a pixellated detector array, a single layer hexagonal geometry imager has been designed and constructed using technology similar in performance to the CsI(Tl)-photodiode layers proposed for use with EIDOS. With the use of GEANT Monte Carlo simulations, an IPR algorithm for the prototype detector was developed. The addition of a HURA coded mask gave the prototype system an imaging capability which was suitable for the implementation and verification of the imaging strategies outlined in Chapter 6.

7.1 Introduction

The good spectral match between the optical emission from CsI(Tl) scintillators and the spectral response of silicon PIN-photodiodes has lead to the widespread application of these detectors in systems where a broad band spectroscopic capability and fine position resolution are a necessity. Optimisation of both the reflective wrapping of the crystals and the coupling between crystal and photodiode, combined with low noise read-out electronics have resulted in a low energy threshold for these detectors of below 100 keV. In addition, when arranged in a multiple layer pixellated array, such detectors are capable of stopping photons up to several MeV. The combination of durability and compactness of this type of detector will prove invaluable for satellite-borne gamma-ray astronomy applications. The INTEGRAL imager will be the first satellite-borne telescope to use both scintillation crystal-photodiode and solid-state detector technology [180] to achieve its fine resolution imaging and broad band spectroscopic capabilities. However, the implementation of pixellated detector technology in a coded aperture imaging system is not necessarily as straightforward as might be assumed.

In Chapter 6, several factors that could have moved the EIDOS configuration away from the ideal response were highlighted with respect to both the SPSF and the observation scenario. These factors included restrictions imposed by engineering requirements, detector element failure both singular and modular, and extreme observing conditions, all of which resulted in systematic errors in the correlation image. It was concluded in Chapter 6, (Section 6.6) that if the correlation image is to provide a reliable solution to the source distribution, it is necessary to suppress *all* forms of coding noise with data pre and post-processing techniques. This has led to the development of alternative imaging strategies to be implemented either as part of the standard imaging sequence, or as contingency plans under extreme conditions. The objective in this chapter is to implement all of the necessary observation strategies in order to evaluate their performance in a real imaging situation. The prototype imager now described has thus served as an invaluable tool for testing not only the basic image and spectral reconstruction sequence outlined in Chapters 3 and 5, but also the imaging strategies/contingency plans developed in Chapter 6 for imaging under extreme observing conditions.

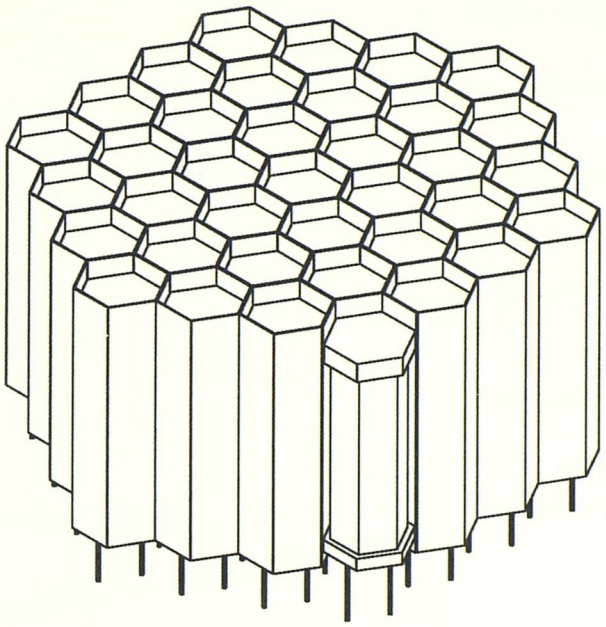
7.2 Instrument Design and Construction

Since the design of the prototype imager has evolved under the auspices of EIDOS, the prototype detector plane, shown in Fig. 7.1, was designed with a hexagonal pixel geometry. The complete detector array comprised of 37 discrete CsI(Tl)-photodiode detectors housed in a spark eroded aluminium honeycomb structure. The purpose of this structure was to provide both lightweight mechanical support and optical isolation for each detector element. The basic pixel design is shown in Fig. 7.2 and consisted of a 30 mm by 11.54 mm A/F hexagonal CsI(Tl) crystal coupled to a standard Hamamatsu S2620 photodiode with a 36 mm^2 sensitive area.

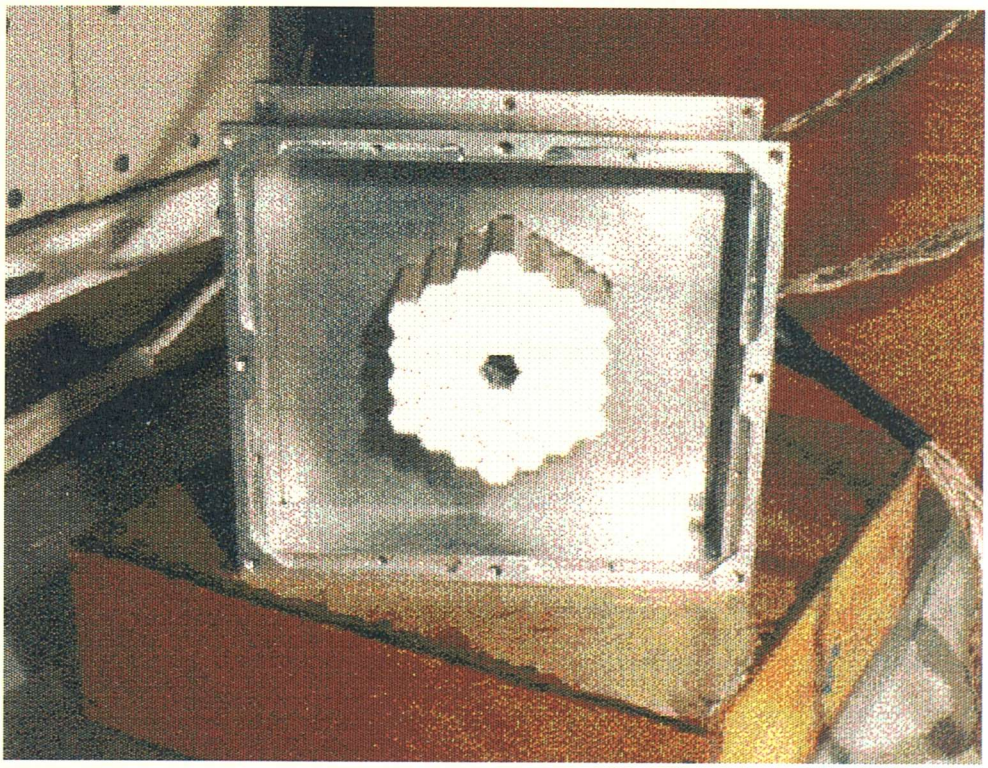
One of the many factors that can severely compromise the performance of a coded aperture imaging system is non-uniformity in the background or 'off-source' counting rate. The 'off-source' counting rate includes the unmodulated (by the mask) detector background flux, as well as contributions from partially coded and completely uncoded sources. As outlined in Chapter 6, non-uniformity in the detector background counting rate may arise for a number of reasons, although for a pixellated detector array, a significant contribution is from pixel to pixel variations in the low energy threshold (LET). The effects of pixel to pixel changes in LET on the reconstructed image were discussed in Chapter 6 (Section 6.4.3). In the following sections the individual components of each detector pixel are considered with specific reference to the uniformity in LET.

7.2.1 CsI(Tl) Scintillation Crystals

The depth of CsI(Tl) crystal used for the prototype was identical in dimensions to the middle and lower layer crystals proposed for EIDOS and provides significant detection efficiency of photons up to 5.0 MeV. The doping of pure CsI with small amounts of impurities (Tl in this case) is required to shift the scintillation emission away from the characteristic 370 nm of pure CsI which is likely to be re-absorbed and is also poorly matched with the spectral response of the photodiode. The addition of Tl shifts the scintillation emission to 570 nm, for which re-absorption is less likely. For scintillation crystals cut from the same boule, the variation in scintillation light yield per keV of incident photon energy as a function of the Thallium concentration is negligible. However, for crystals with unusually low Tl concentrations (100 ppm), the dominant contribution to the scintillation yield will be at 370 nm and so the light collection efficiency (LCE)



(a)



(b)

Figure 7.1: A schematic diagram of the 37 element prototype detector array (top), and the practical implementation of this design (bottom).

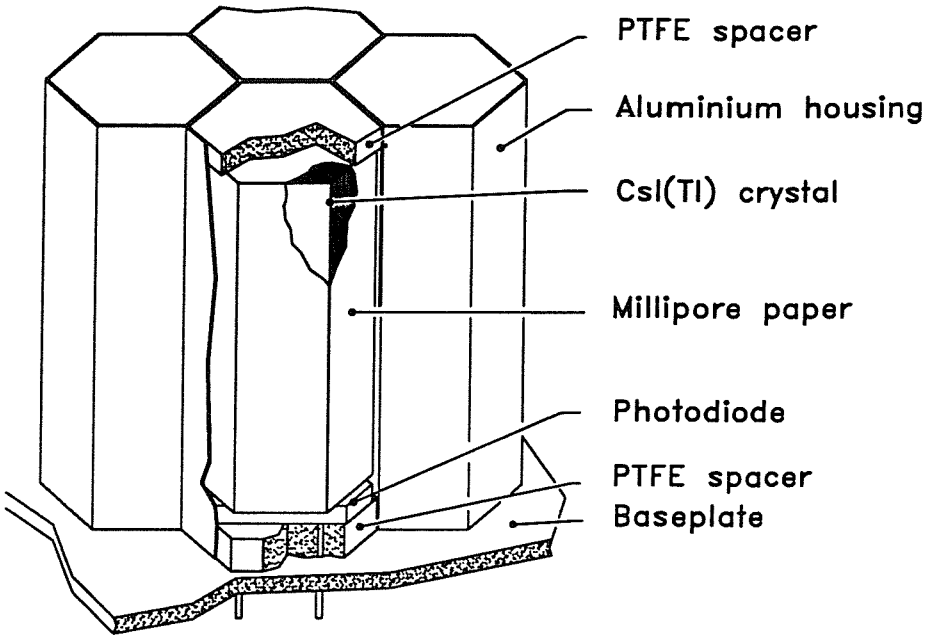


Figure 7.2: The construction and housing of the individual detector pixels.

will be reduced. Similarly, at extremely high Tl concentrations (2000 ppm), the optical properties of the crystal are altered, which again will lower the LCE. Tests on a wide range of crystals from the same boule (see Fig. 7.3) with different Tl concentrations, have shown a variation in light yield of at most a few percent. It is therefore expected that any non-uniformities in the pixel to pixel LETs will arise in the crystal preparation, wrapping and photodiode coupling stages, assuming that the pixel to pixel photodiode performance remains reasonably constant.

The effect of crystal surface preparation on the overall conversion efficiency (OCE) (the number of electron-hole pairs produced in the photodiode per keV of detected gamma-ray energy) of each detector has been studied extensively by Carter [150]. It has been demonstrated that the surface treatment is a critical component in maximising LCE. Further studies have indicated that the optimum surface finish, regardless of the crystal geometry, is obtained by polishing all sides except the exit face to be coupled to the photodiode, which should be roughened. EIDOS however employs almost 6000 identical CsI(Tl) crystals, suggesting it may be impractical to implement such a preparation philosophy whilst expecting a reasonable uniformity range. A more realistic solution would be to roughen all sides except the end-face which should be polished. This preparation performs to within 3 % of the optimum technique. One of the main goals

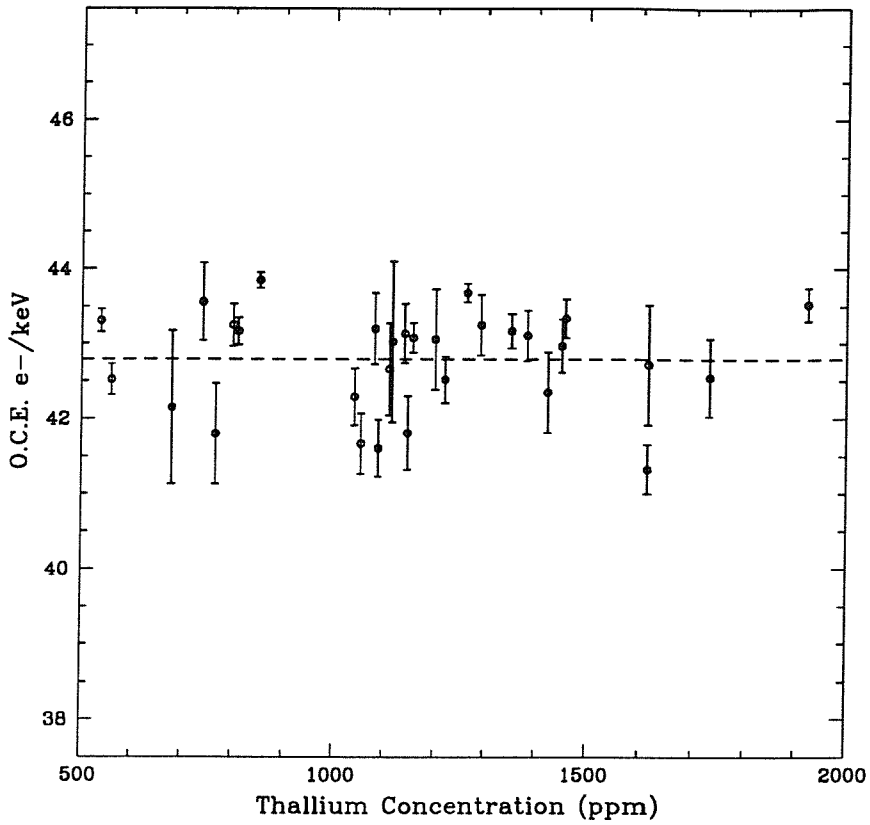


Figure 7.3: The spread of scintillation light yield with Thallium concentration.

of the prototype was to assess the performance of each imaging strategy developed in Chapter 6 when dealing with detector non-uniformities. Consequently, during the prototype construction stage, the opportunity was taken to employ several crystal preparation techniques.

In order to minimise the percentage of scintillation photons escaping from the surface of the crystal and being absorbed in the aluminium housing, it is necessary to either coat or wrap the crystal in a suitably reflective material. The LCE of CsI(Tl) crystals with different reflecting surface materials has been investigated independently for both vacuum deposited metallic coatings [216], and reflecting filter papers [217]. The results conclude that fine pore filter papers such as Millipore HAWP00010 give by far the best LCE of all the materials tested. Millipore HAWP00010 was thus chosen as the wrapping material for the prototype.

The crystal-photodiode interface has also been identified as a critical component for optimising the LCE. Because CsI(Tl) has a low absorption coefficient, light losses in the detector element will generally occur at interfaces and surfaces. Monte Carlo

simulations [150] have shown that the LCE is optimised by using a coupling compound with a refractive index greater than that of CsI(Tl) (1.5). Since the module presented an ideal opportunity for testing the long term stability of different coupling compounds, three different adhesives were used to bond crystal to photodiode : Epotek 301-2, Norland 68, and Dow Corning DC 92-500. High failure rates were seen with the Epotek and Norland adhesives (90 % and 100 % respectively), with the majority of failures occurring between the crystal-adhesive interface. The Dow Corning glue, which incidentally is the only space qualified adhesive of the three, was more successful with no failures seen once the detector had been inserted into the aluminium support.

7.2.2 Photodiodes

Readout of the EIDOS scintillators would have been provided by custom made hexagonal photodiodes. Initially, it was planned that the prototype detector should also make use of the first iteration of this photodiode technology, but at the time of construction these were unavailable. Instead, commercially available Hamamatsu S2620 photodiodes were coupled to the crystals, initially as a temporary solution, but as a result of their satisfactory performance were made permanent. In addition, the small diode dimensions (6 mm $\times$ 6 mm) allowed the complete pixel to be inserted into the honeycomb structure. This arrangement is non-optimal with respect to the area matching between scintillator and diode (30 % coverage) [191], although the degradation in signal-to-noise ratio as a result is not as serious as may be expected since the electronic noise is also reduced with a smaller diode collection area. The optimum photodiode area for this particular geometry would be achieved by using a photodiode that covers $\sim$65 % of the crystal output surface area [191]. Variations in the electronic noise from diode to diode and in the subsequent shaping and readout electronics will produce variations in the pixel to pixel LET. In the space environment the dominant factor affecting the electronic noise per channel is the variation in detector temperature. In order to evaluate the variation in electronic noise with temperature for the photodiodes used, each channel was tested using a standard Aurel PAC-L3 pre-amplifier (low noise) with appropriate shaping and post-amplification, after which the signal was processed using conventional ADCs.

The fact that X-rays interact directly with the photodiode silicon provides a convenient and direct means of calibrating each photodiode. The dominant interaction process of $\sim$keV photons in silicon is the photoelectric effect. Thus, an incident X-ray with energy

E_X produces $\frac{E_X}{\epsilon_{Si}}$ electron-hole (e-h) pairs, where ϵ_{Si} is the energy required to produce a single e-h pair. The position of the full energy peak in the resulting ADC bin spectrum enables calibration of the energy scale. The Gaussian spread of the full energy peak is a consequence of both the photoelectron statistical fluctuations and the photodiode noise. Since the magnitude of the statistical fluctuations are known, the photodiode noise or equivalent noise charge (ENC) can be defined in terms of the line spread and line energy (for a particular shaping time, in this case $5\mu s$) :

$$ENC = \left(\left(\frac{\delta E}{2.355\ \epsilon_{Si}} \right)^2 - \frac{E_X}{\epsilon_{Si}} \right)^{\frac{1}{2}} \tag{7.1}$$

Figure 7.4 shows the resulting variation in electronic noise (ENC) with temperature for all photodiodes. The important point to notice is that over the predicted EIDOS operating temperature range (0 ± 10oC), the curve is fairly flat suggesting that global changes in the mean LET and hence the detector background level will not be temperature induced. Each data-point and associated error bar has been derived from all 37 photodiodes.

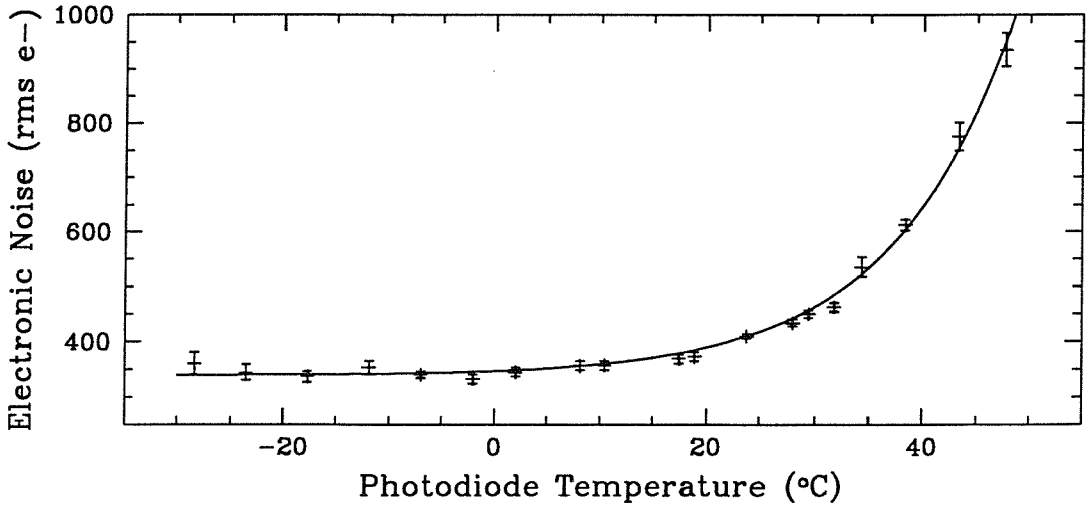


Figure 7.4: Variation in photodiode electronic noise with respect to temperature.

7.2.3 Readout Electronics

Readout of each photodiode was initially provided by three custom-built, low noise pre-amplifiers, each capable of handling 16 channels of detector output. The noise figure for these pre-amplifiers ($1\sigma \approx 650$ electrons) was similar to that expected from the EIDOS ASICs (Application Specific Integration Circuits) and in comparison with Fig. 7.4 it

can be seen that the pre-amp noise will dominate over the photodiode contribution at the expected operational temperatures. Subsequent stages of amplification, shaping and discrimination were achieved by using custom-built hybrid circuits, which provided a shaping time of 5 μs and an adjustable gain of up to 1500. After this stage, the signal was processed using conventional ADCs and associated electronics. Unfortunately the primary iteration of the data-acquisition system produced several unforseen technical difficulties which ultimately resulted in the design and construction of an alternative method of data acquisition. The following section discusses the calibration and resulting performance with respect to the LET for this first iteration.

7.2.4 Detector LET Performance and Calibration

Operational temperatures for the prototype detector were stabilised with the aid of a Class 10000 Clean Room environment. Initial measurements of the pixel to pixel LET however, were performed at room temperature. The LET of a particular pixel is governed by two factors which are the OCE, the efficiency of converting the incident photon energy into electron-hole pairs in the photodiode, and the ENC, the standard deviation in the electronic noise at the pre-amp input. An energy deposit in the scintillation crystal will only be 'detected' as a significant event once the number of electron-hole pairs produced in the photodiode are are above some threshold level in units of the ENC. The detection significance D, is thus defined by Eq. 7.2.

$$D = \frac{E_\gamma \, OCE}{ENC} \quad , \tag{7.2}$$

where E is the energy deposited in the crystal. Conversely, the minimum energy deposit that can be detected is simply :

$$E_{LE} = \frac{D \, ENC}{OCE} \quad . \tag{7.3}$$

Measuring the OCEs for each detector element in the module and using the values of ENC from Fig. 7.4, it was possible to calculate the LET for each detector. The values are shown graphically in Fig. 7.5 and have a mean of 160 $\pm$ 13 keV. The wide range in LET is a result of both the unavoidable production non-uniformities and those introduced

deliberately in the form of different surface finishes and optical couplings. It should be noted that the mean module LET was higher than the proposed 120 keV LET of the EIDOS CsI(Tl) layers, a direct consequence of the large area mismatch between the crystal and photodiode.

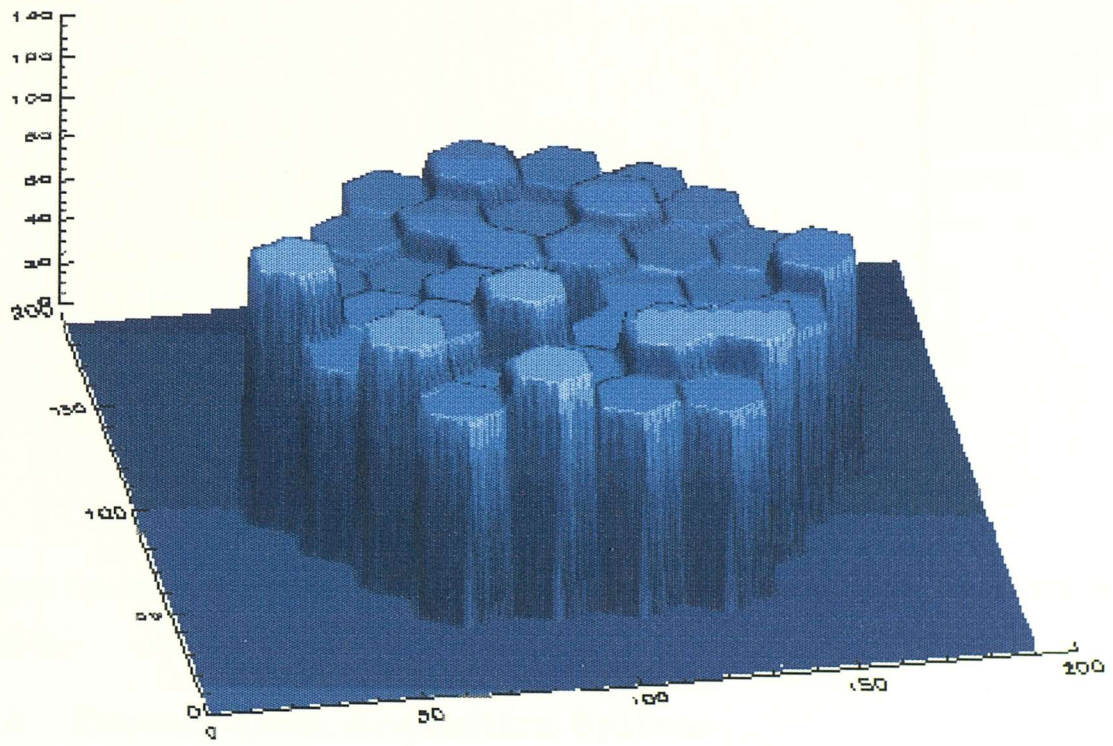


Figure 7.5: The values of LET across the module detector plane.

Channel by channel energy calibration was performed using a ^{22}Na source, producing an offset and gain look-up-table to ensure that there were no systematic contributions to the spectral broadening. By way of an example of the spectral capabilities of the prototype detector, Fig. 7.6 shows a calibrated spectrum of a ^{22}Na source. The energy resolution at 511 keV is ~15 % (FWHM), which is poorer than the 12 % predicted by Eq. 5.4 (Chapter 5) for the EIDOS CsI(Tl) layers. However, the poor energy resolution is not surprising given the test-bed nature of the prototype construction process. In the following sections, the current data acquisition system is outlined along with the IPR sequence which is developed using the same Monte Carlo approach as that used in Chapter 5 to develop the EIDOS IPR algorithm.

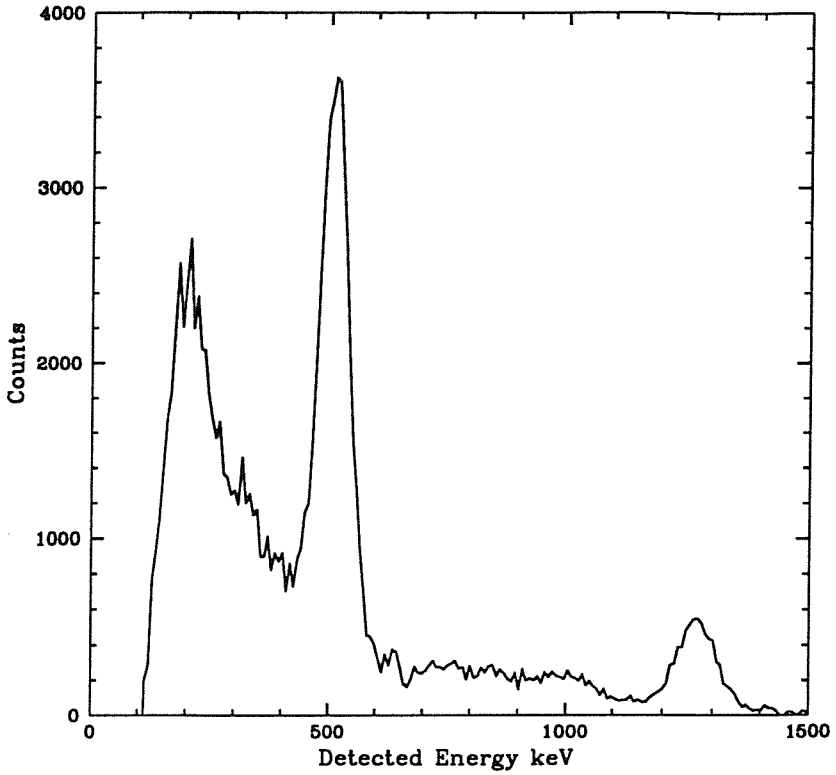


Figure 7.6: An energy calibrated spectrum of ^{22}Na. The energy resolution at 511 keV is ~15 %.

7.3 Current Data Acquisition System

Several fundamental flaws in the design of the original data acquisition system, for example the 16-channel pre-amps were not compatible with the the shaping amplifiers and were also considerably noisier than expected, led to a complete overhaul of the system and subsequently its re-design. The complete imaging sequence, from observation through to image reconstruction, including the current data acquisition system, is shown schematically in Fig. 7.7.

7.4 Incident Pixel Reconstruction for the Prototype Module

During the design phase of the prototype module, it was highlighted that a second identical layer of detectors could be added to produce a configuration that was directly representative of the lower layers of the EIDOS detector. This additional layer would

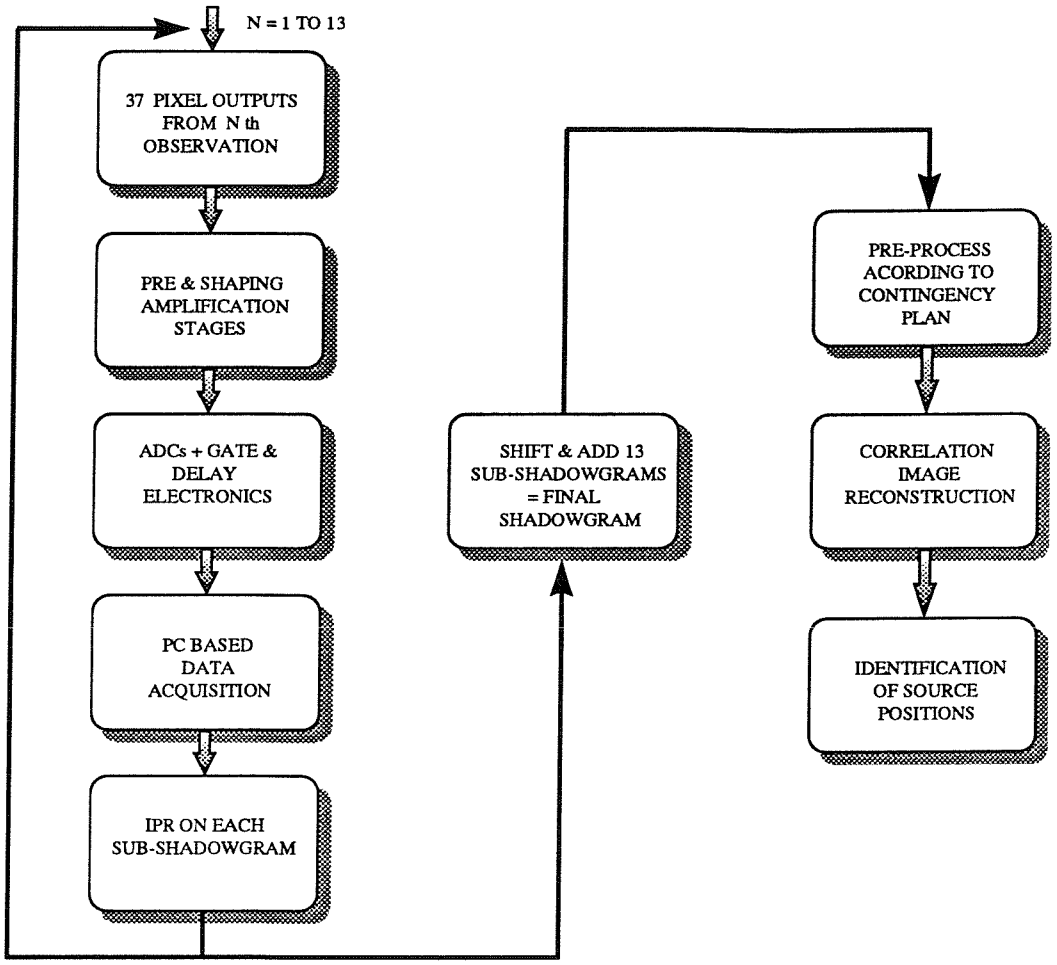


Figure 7.7: The complete imaging sequence, from the N sub-observations (see Section 7.5) to image reconstruction.

have allowed the implementation of the IPR techniques developed specifically for EIDOS in Chapter 5. However, at the time construction of a second layer was not viable and consequently a direct verification of the EIDOS IPR algorithm was not possible. For a single layer geometry however, it is preferable to implement an IPR capability to some extent since there is still a significant scattered event component in the detected photon count. To determine the significance of the scattered event component and hence the importance of an IPR capability, a GEANT-3 Monte Carlo simulation was used to model the response of the prototype detector over the 100 keV to 2 MeV energy range. The results of these simulations are now discussed.

7.4.1 The GEANT-3 Simulation

The GEANT-3 model of the prototype detector was virtually identical to the design outlined in Section 7.2. The only modifications were the addition of an Al detector housing, a support structure and an 'optical' bench, in an attempt to model the surrounding laboratory environment. As in Chapter 5, Section 5.3.2, detection modes and subsequent IPR efficiencies were obtained for a beam of photons fired into a single (central) pixel, but this time with a flat spectral profile over the 0 keV to 5 MeV energy range.

7.4.2 Detection Modes

For a single, pixellated layer of detectors, the possible modes of interaction are considerably less complex than for EIDOS and can be characterised simply by the event multiplicity. Figure 7.8 shows the percentage of events in each of the first 3 interaction modes (single, double, or triple site events) as obtained with the GEANT-3 Monte Carlo model. Energy thresholds and spectral broadening were not included.

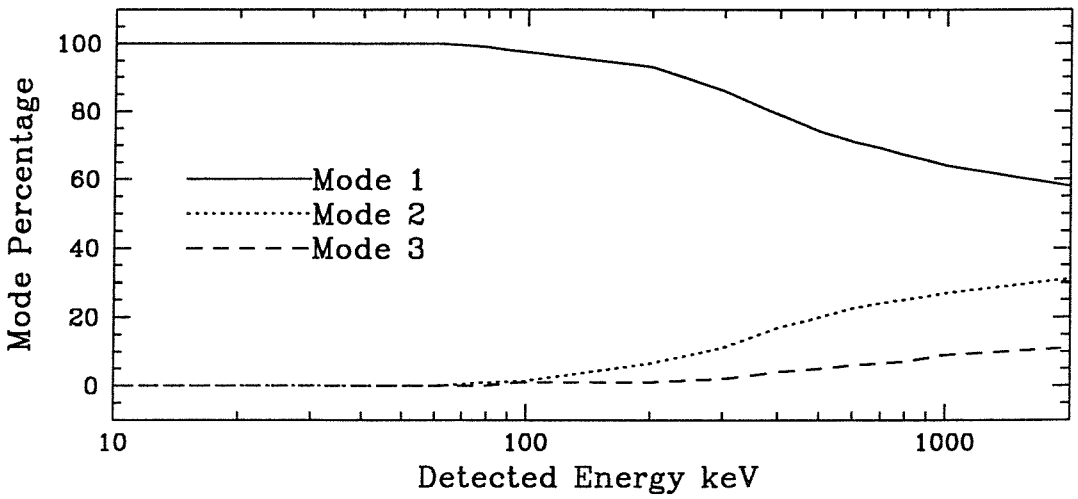


Figure 7.8: Detection modes for the prototype module. M1 – single site, M2 – double site, M3 – triple site events.

As expected, single site events, whether through photoelectric absorption at lower energies, or single Compton scatters at higher energies, dominated the entire prototype energy range. However, at mid to high energies (> 300 keV), multiple site interactions made a significant contribution to the total number of events detected which suggested

that some form of event selection was required at higher energies in order to determine the correct incident pixel.

7.4.3 IPR Efficiencies

In accordance with the detector geometry, only 2 methods of IPR methods could be defined. The correct pixel of incidence for a multiple site event would be given by either the interaction with the highest energy deposit or by the interaction with the lowest energy deposit. Thus, in terms of the methods of IPR defined in Chapter 5, (Section 5.4.1), the choice was between the MAXEN and MINEN methods. Figure 7.9 shows the efficiency of both IPR methods as a function of detected energy for the three most probable event modes. Energy thresholds and spectral broadening have been applied to the simulated data prior to IPR.

The most notable feature in Fig. 7.9 is the cross-over point at $\sim$ 500 keV for double site events. A similar cross-over point was seen in the EIDOS IPR efficiencies for these two IPR methods with double site events, and indeed the explanation for the cross-over point at $\sim$ 500 keV in Chapter 5 (Fig. 5.8), also applies to the cross-over point in Fig. 7.9. From Fig. 7.9 the most efficient IPR algorithm can be summarised as follows : MAXEN for *all* modes of detection at *all* energies, except for two site events below 511 keV for which MINEN should be used. Applying this algorithm to the simulated data reveals the total IPR efficiency expected from the prototype over the 120 keV to 2 MeV energy range (see Fig. 7.10).

7.5 Imaging Configuration for the Prototype Detector

Construction of the prototype detector has allowed the evaluation of a large scale implementation of EIDOS detector technology. However, the principle objective was the assessment of the imaging strategies developed in Chapter 6 in a real imaging scenario. Although the size of the prototype detector array was more than adequate for studying for example, the distribution of LETs when adhering to a precise construction strategy, the implementation of an array this size in a coded aperture system has severe limitations. These limitations arise from the fact that for a 37-element array, there are only two HURAs for which optimum matching between the unit pattern and the detector plane can

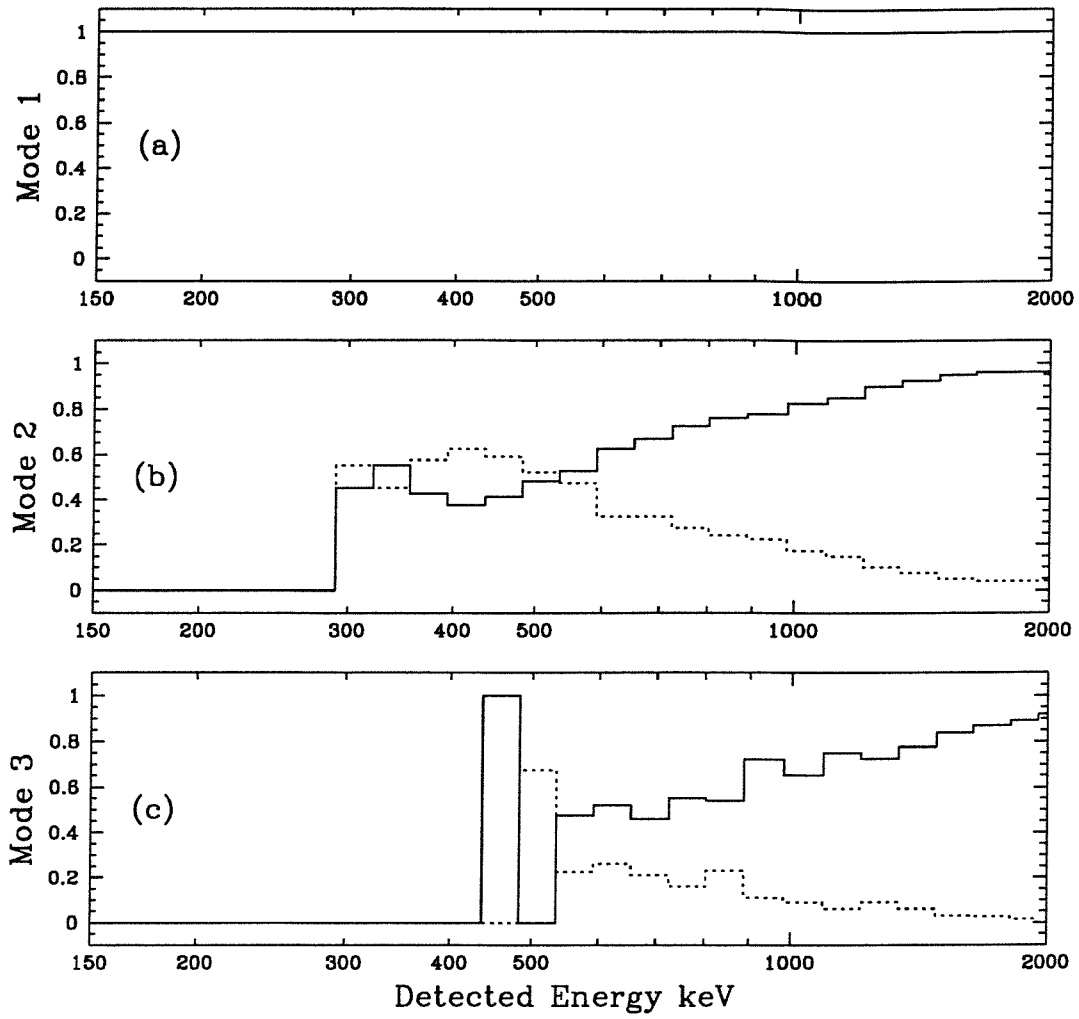


Figure 7.9: (a) MAXEN (solid line) and MINEN (dotted line) efficiencies for single site events, (b) for double site events, and (c) for triple site events.

be achieved. The optimum match occurs when the detector provides sampling of the unit pattern only, with a sampling frequency of at least one detector pixel per mask element. Thus, in order to increase the number of mask patterns that could be implemented in the 'optimum' configuration, it was necessary to increase the area sampled by the detector *without* modification of the array size.

The solution was simple : the effective detector area was increased by combining the incomplete shadowgrams recorded at several detector positions. Thus, a single observation entailed sequentially shifting the 37 pixel detector through a predetermined pattern in order to provide a complete sampling of the mask pattern. Using a sequence of 13 detector positions, a pseudo-detector array containing 331 independent detector pixels was produced covering an area of approximately 380 cm^2. The increased detector coverage

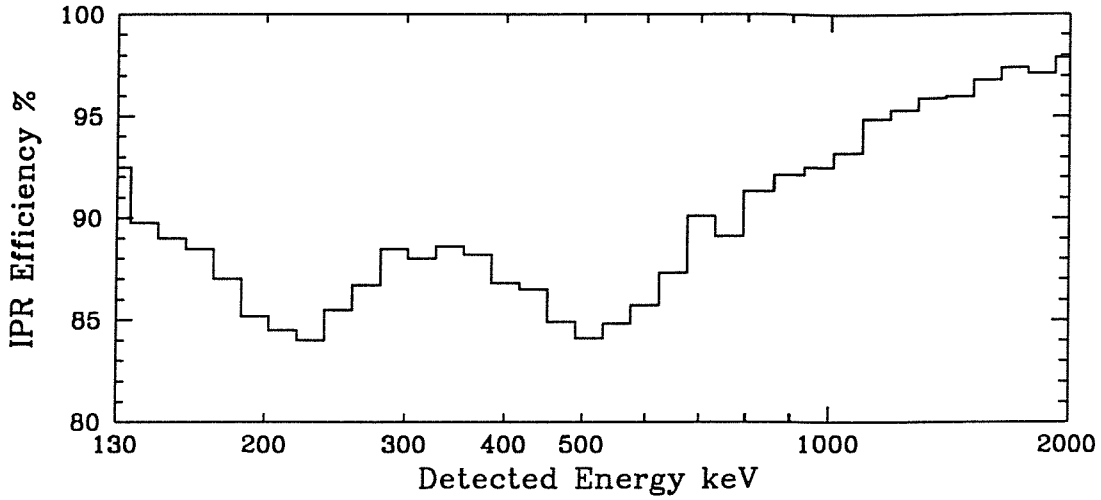


Figure 7.10: The total IPR efficiency using only the MAXEN IPR method.

allowed the use of all HURA orders up to and including the 331 unit pattern. Figure 7.11 shows schematically the mask/detector configuration and the area covered by the pseudo-detector array.

7.6 Imaging Strategies Applied to the Prototype Module

The motivation behind the development of alternative imaging strategies for correlation image reconstruction was outlined in Chapter 6. The conclusion from extensive Monte Carlo simulations of the INTEGRAL imager, for both the EIDOS and IBIS configurations, was that fast image reconstruction by correlation techniques *is* viable under extreme observing conditions. The aim in this section is to apply the same set of imaging strategies to the prototype imager in order to verify the conclusions of the previous chapter.

7.6.1 Sources of Systematic Defect

The 'test-bed' nature of the prototype detector construction has meant that systematic effects such as pixel to pixel non-uniformities in the LET are readily available. Similarly, realistic coding noise artifacts from the re-binning of the recorded data and decoding arrays will be induced in the SPSF (see Chapter 6, Section 6.4.1). In addition, it might be reasonable to expect some large-scale non-uniformities in the detector background purely because of the sequential nature in which each shadowgram is recorded. However, with reference to Fig. 7.5, pixel to pixel variations are expected to be the most dominant

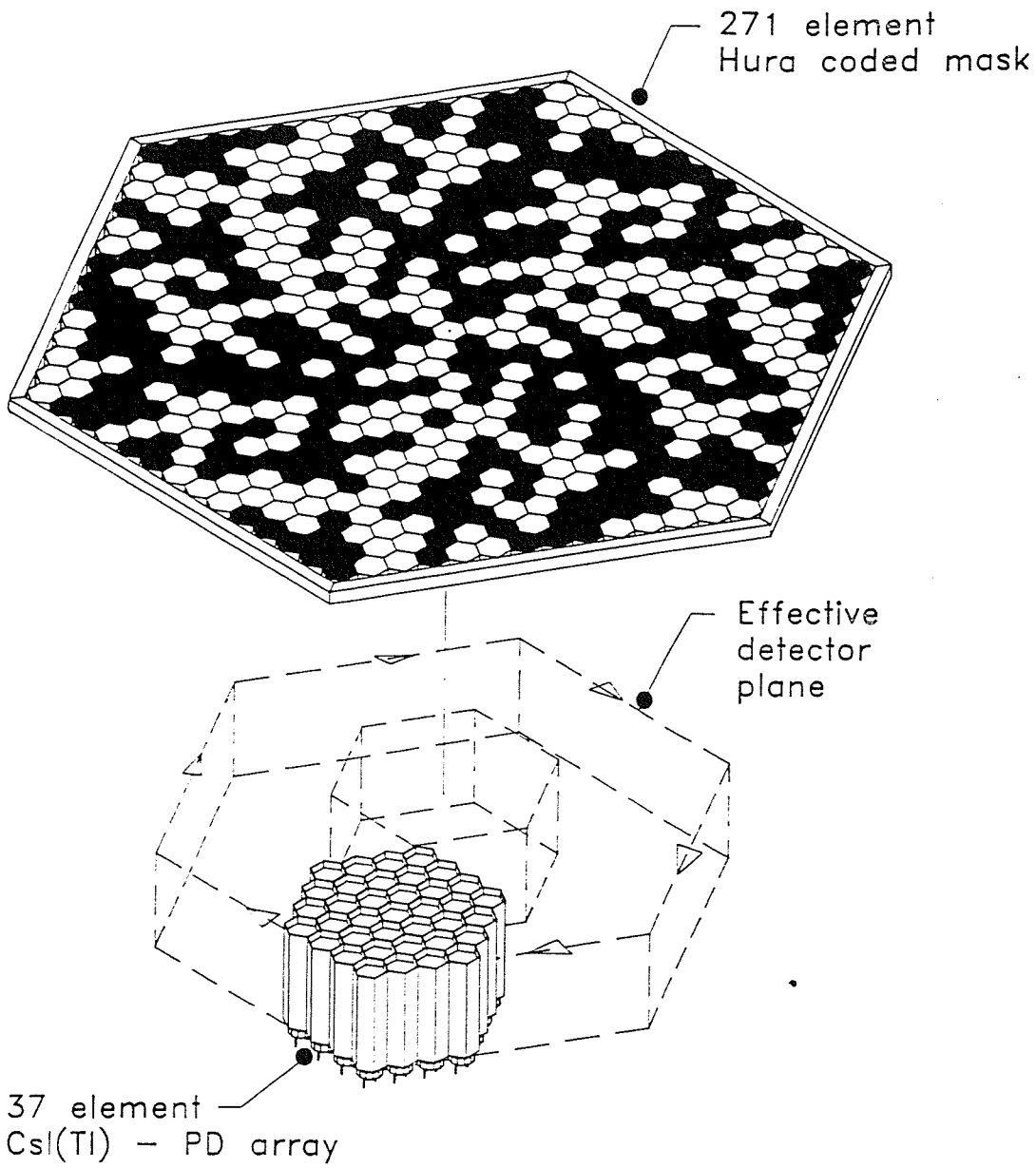


Figure 7.11: The imaging system based on the prototype detector module. The mask pattern shown, and used in most of the observations is the 331 HURA. The dotted lines shows the outline of the pseudo-detector array, once the prototype module has been moved through the 13 position sequence.

systematic effect.

An idea of the severity of the pixel-to-pixel non-uniformities was obtained by examining a typical background spectrum observed by the prototype module. Figure 7.12 shows a background spectrum taken in the Clean Room environment with a ~22 hour integration time. Such background observations were used as the flat-field data for the flat field imaging strategy. Consequently, to account for any large-scale spatial variation induced by the 13 position imaging sequence, the 22 hour background observation was divided

into 6000 second sub-observations. In addition to the underlying continuum, prominent features visible in the background spectrum include the ^{40}K line at 1.461 MeV originating in the concrete Clean Room walls, and other laboratory sources stored within the Clean Room. It should be noted that these additional sources were present throughout *all* observations and so the resulting additional background component was temporally and spatially invariant. The underlying power law continuum observed is thought to arise from reprocessed cosmic-rays through interactions in the atmosphere and the immediate environment. A crude fit to the spectrum ($B = 2000\ E^{-2.6}$) suggests that the 8 % fluctuations in the LET shown in Fig. 7.5, will lead to a 15 % variation in the pixel to pixel background level.

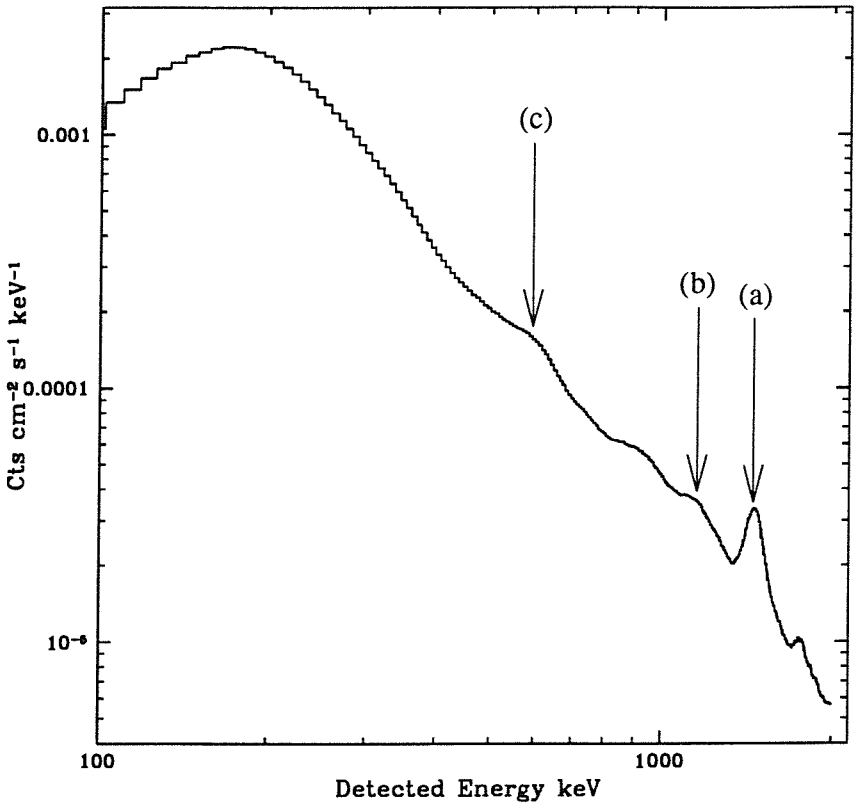


Figure 7.12: A 22 hour observation of the laboratory background spectrum showing (a) the expected ^{40}K 1.461 MeV line from the surrounding concrete and other sources ((b)^{60}Co and (c)^{137}Cs) located in the Clean Room.

Figure 7.13 shows the total 6000 second pseudo-flat field obtained from the 22 hour background observation and demonstrates that the detector background is far from uniform on pixel scales.

However, once imaging studies had commenced, a systematic effect was discovered that also produced background variations on pixel to pixel scales. The cause of this additional

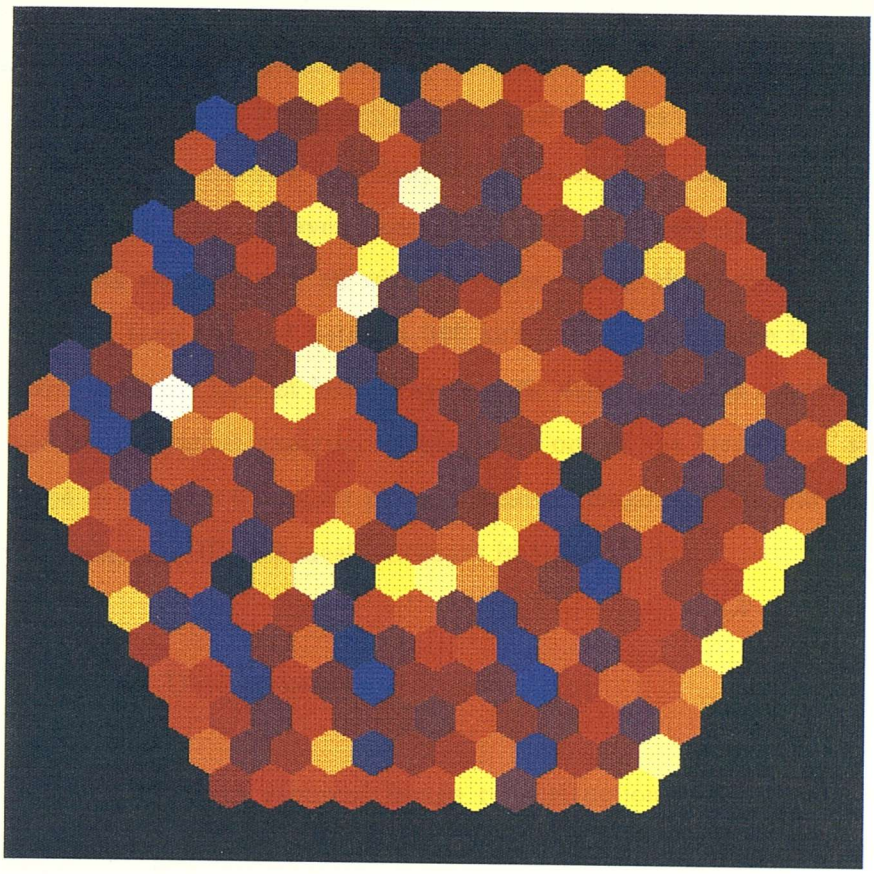


Figure 7.13: The 22 hour background observation shown in Fig. 7.12, binned according to event position with respect to pseudo-detector array coordinates, thus producing a 6000 second pseudo-flat-field.

'noise' was finally tracked down to faulty shielding in the connectors carrying the pre-amp signals from the detector to the shaping-amps, which as a consequence were picking up radio-frequency (RF) interference. The magnitude of the RF interference was sufficient to produce false triggers that were 'detected' as low energy background events and the temporal instability of the interference complicated attempts at coding noise suppression. Overall, this additional systematic effect was regarded as *the* dominant source of coding noise in the resulting correlation images.

7.6.2 Controlling the Defect Level

Limitations on the processing power of the data acquisition system meant for example, that it was not possible to introduce large scale variations in the background level in a controlled way. However, since *all* systematic defects observed were energy dependent,

as well as spatially and temporally dependent, it was possible to control the defect level by adjusting the energy band over which images were reconstructed. To simulate astronomical imaging of a point-like source, observations were made of a 10 μCi ^{133}Ba source (four strongest lines at 276.4, 302.8, 356.0 and 383.8 keV) for a total integration time of 600 seconds (per detector position) with the source positioned approximately 2 m from the detector plane. Eleven energy bands with increasingly significant levels of systematic defects were defined, all with an upper energy threshold of 450 keV. The lower threshold was then stepped between 300 keV and 0 keV, where the 0 - 450 keV shadowgram contained the highest level of systematic defects. Image reconstruction was then performed by correlation analysis, coupled with one of the four 'noise' suppression methods outlined in Chapter 6, namely plateau removal imaging (PRI), flat-field imaging (FFI), dithered imaging (DI) and mask-antimask imaging (MAI).

7.6.3 Results and Discussion

The unprocessed and plateau removed image reconstructions were performed in the standard way with all correlations, regardless of the imaging strategy, performed in the Fourier domain using FFTs. The flat field imaging sequence employed the non-local flat-field shown in Fig. 7.13 which had an integration time an order of magnitude larger than the observation time. For dithered imaging, a 9-position square pointing strategy was employed for which it was ensured that each individual pointing was greater than the SPSF width. A 331 element unit pattern HURA positioned ~30 cm from the detector plane was used for all observations and provided a mask-antimask capability with its 60 degree rotational antisymmetry.

Image quality for each energy band and imaging strategy was assessed using the figure of merit (FOM) definition of the reconstructed source significance :

$$FOM = \frac{s_p}{\frac{1}{N}\sum_{i=1}^{N} s_i^2 - \frac{1}{N^2}\left(\sum_{i\neq p}^{N} s_i\right)^2} \quad , \tag{7.4}$$

where s_p is the reconstructed peak flux, N is the number of reconstructed sky pixels in the image, and the denominator is simply the rms deviation of the reconstructed background noise obtained from all pixels excluding those associated with the source. Figure 7.14 shows the FOM of the reconstructed ^{133}Ba source for all correlation methods and energy

bands. For comparative purposes, the mask-antimask reconstruction is shown by the solid curve.

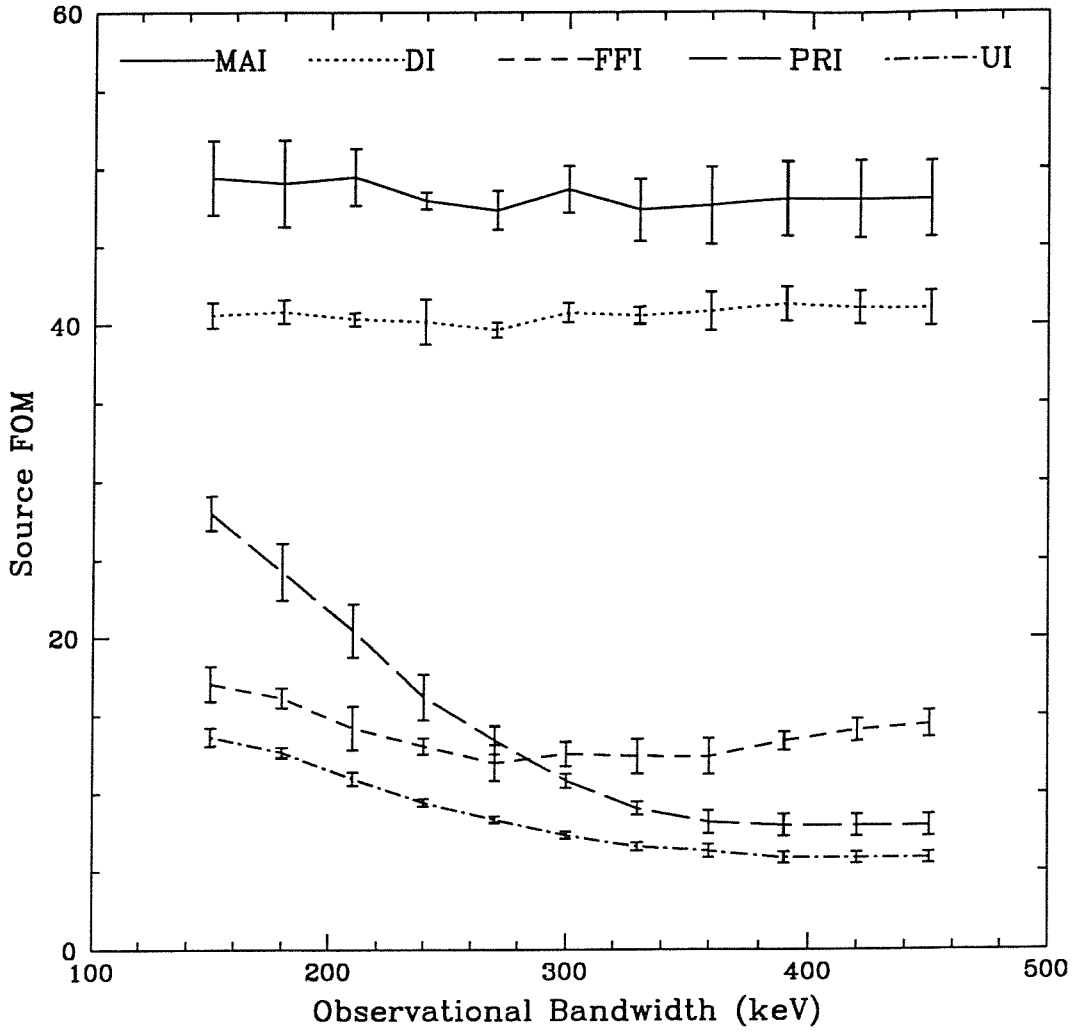


Figure 7.14: The performance of all imaging strategies for suppressing both pixel to pixel and SPSF coding noise artifacts in the prototype imager.

As expected, the mask-antimask imaging strategy effectively removes *all* systematic defects in *all* energy bands and can therefore be considered as the near-ideal FOM for these observing conditions. Figure 7.15(a) shows the mask-antimask image of the ^{133}Ba source for the highest level of systematic 'noise'. At the other extreme, the unprocessed image was severely degraded by both background magnification of SPSF coding noise and by the pixel to pixel background variations, and in almost all energy bands (Fig. 7.15(b)). The plateau removed reconstruction was reasonably successful at removing the 'dc' background component and consequently the background magnified coding noise, but failed to suppress the artifacts induced by pixel to pixel variations (Fig. 7.15(c)). The poor response of the flat-field technique (Fig. 7.15(d)) highlights one of the major disadvantages

of this method. The flat-field taken pre or post-observation is only effective if the background is temporally independent. When flat-field observations with the prototype imager were being taken, changes in observing conditions were noticed which were due primarily to small temperature fluctuations in the Clean Room environment. Although these fluctuations were small (at most a few degrees), they were sufficient to render the flat-field strategy ineffective in the suppression of background artifacts. From Fig. 7.14 it was concluded that in the event of mask rotation failure, the optimum imaging sequence would be dithered imaging. A comparison between the correlation images produced by unprocessed and dithered image reconstruction (see Fig. 7.16) reveals a seven-fold improvement in detection significance. Thus, both Monte Carlo simulations and now laboratory observations suggest that the dithered imaging strategy is a reliable method for removing the systematic defects induced by a non-ideal telescope/observation scenario.

7.7 Conclusions

For a pixellated coded aperture telescope without a mask-antimask capability, whether through mask rotation failure or by design, it is apparent from both simulations and now laboratory images, that dithering the spacecraft attitude would have a positive impact on the performance of the telescope in extreme observing conditions. For a reasonably well selected pointing strategy, dithered imaging enables the removal of systematic effects from the detector plane which might otherwise severely degrade the imaging performance. For observations made with the prototype imager a seven-fold increase in the image quality is achieved when a dithered pointing strategy is employed as opposed to an unprocessed correlation reconstruction. Thus, the successful demonstration of dithered imaging on data from a pixellated coded aperture system, is the last step in the image deconvolution sequence.

It has been shown that image reconstruction by correlation analysis is reliable under many severe observing conditions with the use of dithered imaging. Additionally, since the reconstruction sequence is performed in the Fourier domain, the image processing time is still kept to a minimum. So what else does the correlation image now provide ? From a maximum likelihood perspective Chapter 3 has shown that correlation analysis is reliable and therefore source positions identified in the correlation map are also reliable. Is it now possible to make flux estimates from the correlation map ? It was suggested in

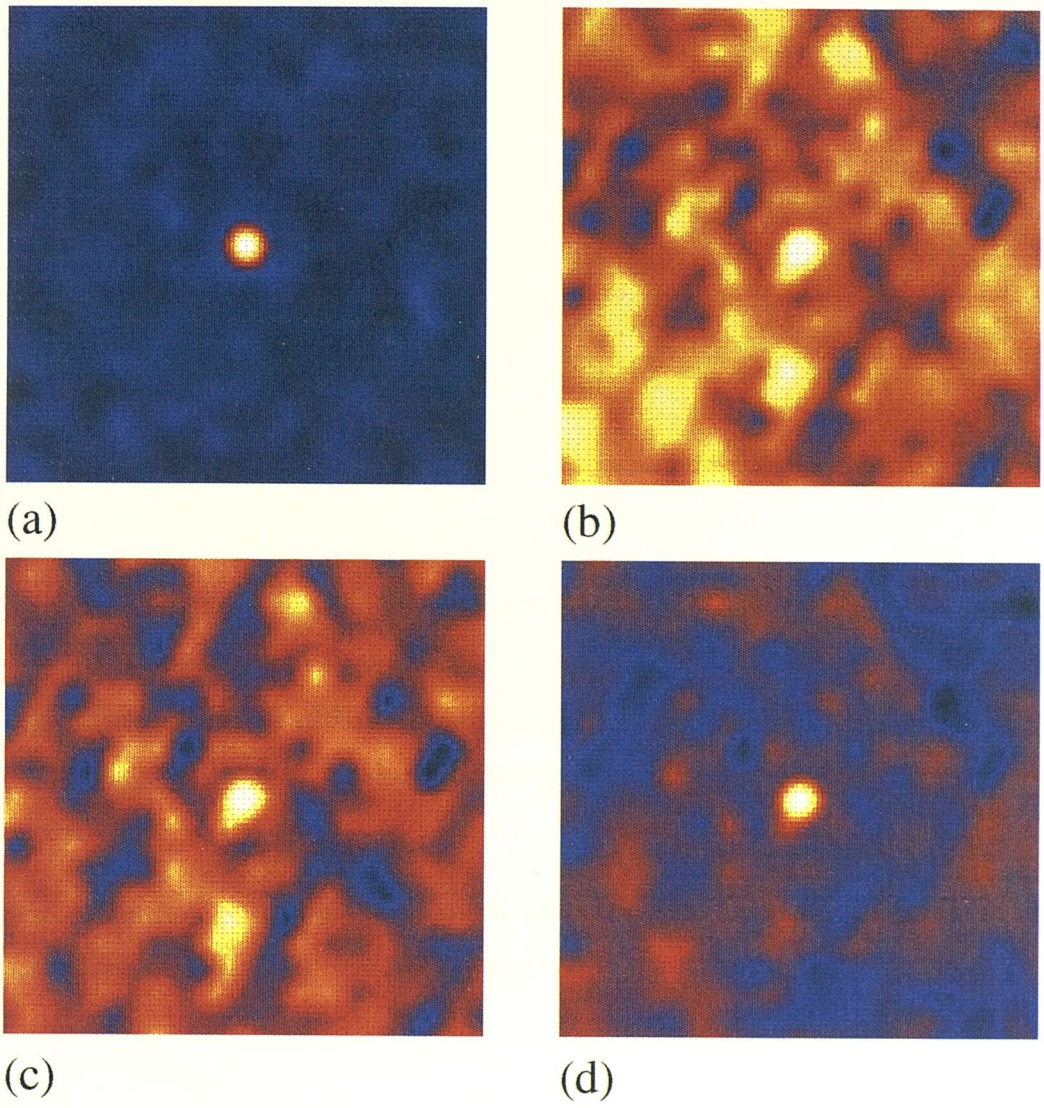


Figure 7.15: (a) The mask-antimask image of ^{133}Ba over the 0 – 450 keV band (high level systematic effects). (b) Unprocessed image. (c) Plateau removal image. (d) Flat-field image.

Chapter 3 (Section 3.6.2) that rather than fitting the image SPSF directly to sources in the correlation map (which have already been subject to several stages of pre-processing and possibly post-processing such as CLEAN to remove ghost sources), it would be more efficient to fit directly to the raw shadowgram data using only the source positions identified in the correlation map. Consequently, Chapter 3 suggested the reintroduction of generalised inverse techniques such as SVD to perform the maximum likelihood fitting. Ultimately, iterative fitting will be required to retrieve the 'true' spectra from the 'raw' spectra obtained by fitting source fluxes in many energy bins. As a result, iteration in the early deconvolution stages must and can be minimised by this approach. Thus, for

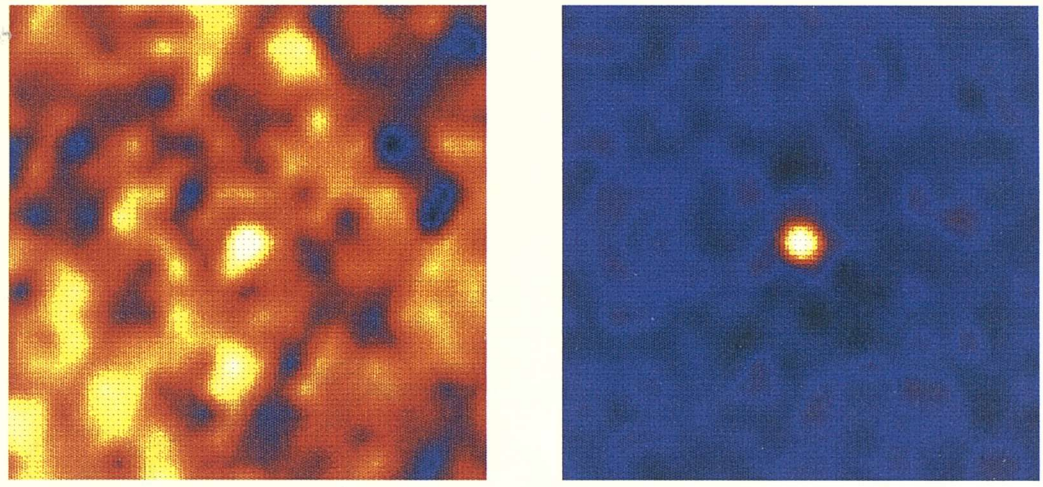


Figure 7.16: As in Fig. 7.15 but for dithered image reconstruction. The FOM for the unprocessed correlation (a) image in the 0 – 450 keV band is $\approx 6\sigma$. The dithered imaging strategy, because of its success at suppressing SPSF coding noise and pixel to pixel background variations, produces a correlation image (b) in which the same source is detected at the $\approx 41\sigma$ level.

the next generation of discrete pixel coded aperture telescopes, the image and spectral reconstruction sequence may well follow the example shown in Fig. 7.17.

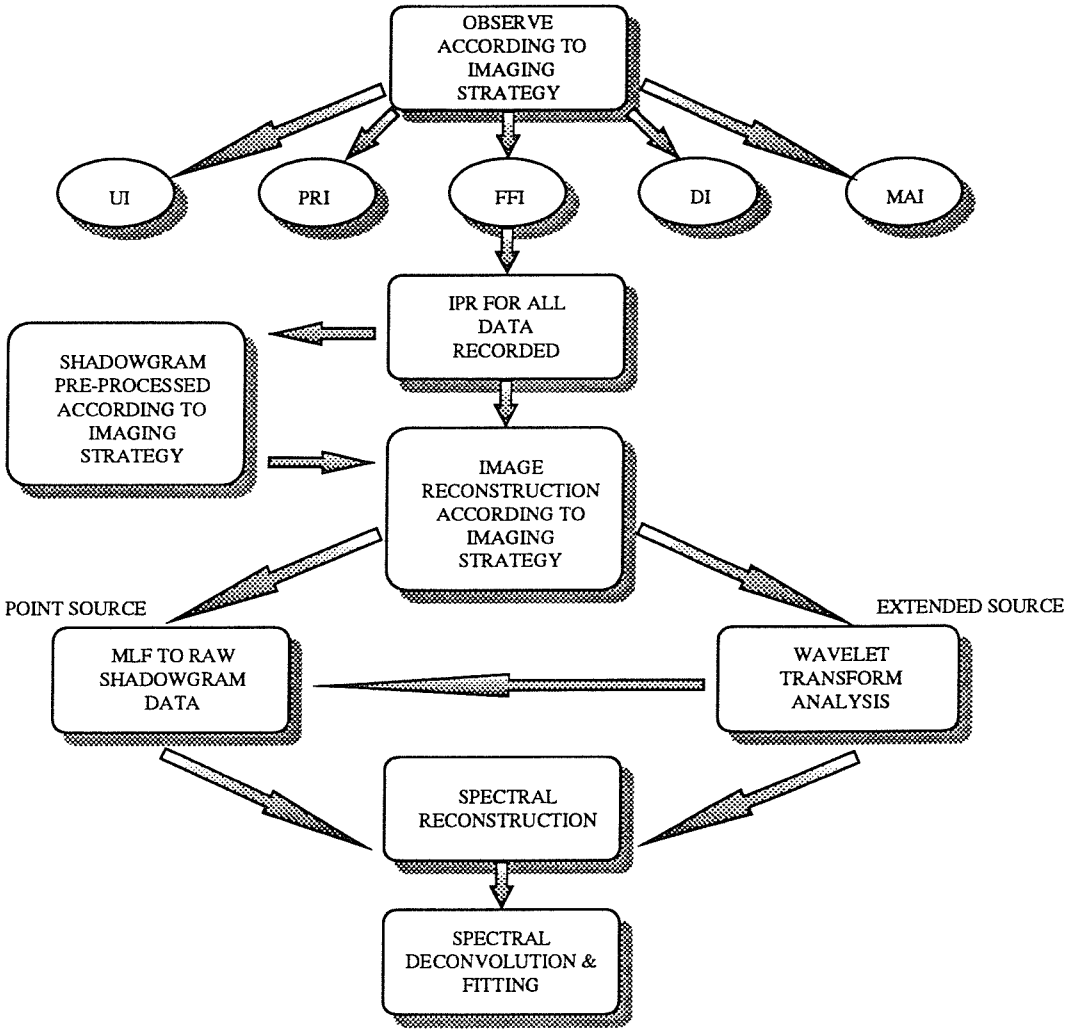


Figure 7.17: A possible solution to the reconstruction and analysis of data from discrete pixel coded aperture telescopes.

Bibliography

[1] G.B. Rybicki & A.P. Lightman, *Radiative Processes in Astrophysics* **Wiley.**

[2] G.R. Blumenthal & R.J. Gould, *Rev. Mod. Phys.* **42** 237 (1970).

[3] R. Ramaty *et al.*, *Ap. J. Suppl.* **40** 487 (1979).

[4] M. Cropper, *Space Sci. Rev.* **54** 195 (1990).

[5] M.S. Briggs, *et al.*, *IAU Circular* **5229** (1991).

[6] M.S. Briggs, *private communication* (1993).

[7] D. Barret *et al.*, *Ap. J.* **394** 615 (1992).

[8] W.R. Cook *et al.*, *Ap. J. Lett.* **372** L75 (1991).

[9] P. Mandrou *et al.*, *Astr. Ap. Suppl.* **97** 1 (1993).

[10] J. Trümper *et al.*, *Ap. J. Lett.* **219** L105 (1978).

[11] F. Nagase, *Frontiers of X-ray Astronomy* **U. Acad. Press** 79 (1992).

[12] I. Grenier *et al.*, *Astron. Astrophys* **204** 117 (1988).

[13] K.S. Cheng *et al.*, *Ap. J.* **300** 500 (1986).

[14] R. Sunyaev *et al.*, *IAU Circular* **5481** (1992).

[15] E. Massaro *et al.*, *Ap. J. Lett.* **376** L11 (1991).

[16] A. Goldwurm *et al.*, *Ap. J. Lett.* **389** L79 (1992).

[17] J.C. Ling & W.A. Wheaton, *Ap. J. Lett.* **343** L57 (1989).

[18] I. Kovner, *Astr. Ap.* **141** 341 (1984).

[19] E.P. Liang & C.D. Dermer, *Ap. J. Lett.* **325** L39 (1988).

[20] R.E. Ligenfelter & C. Hua, *Ap. J.* **381** 426 (1991).

[21] J.E. McClintock & R. Remillard, *Ap. J.* **308** 110 (1986).

[22] J.C. Ling *et al.*, *Ap. J. Lett.* **321** L117 (1987).

[23] M. Bouchet *et al.*, *Ap. J. Lett.* **383** L45 (1991).

[24] R. Sunyaev *et al.*, *Ap. J. Lett.* **383** L49 (1991).

[25] M.A. Alpar *et al.*, *Astr. Ap.* **257** 627 (1992).

[26] W.H.G. Lewin *et al.*, *Space Sci. Rev.* **46** 273 (1988).

[27] A. Vikhlinin *et al.*, *Ap. J.* **424** 395 (1994).

[28] S.M. Matz *et al.*, *Nature* **331** 416 (1988).

[29] B.J. Teegarden *et al.*, *Nature* **339** 122 (1989).

[30] J. Tueller *et al.*, *Ap. J. Lett.* **351** L41 (1990).

[31] P.A. Pinto & S.E. Woosley, *Ap. J.* **329** 820 (1988).

[32] K.W. Chan & R.E. Lingenfelter, *Ap. J.* **368** 515 (1991).

[33] S.E. Woosley *et al.*, *Ap. J.* **301** 601 (1986).

[34] K. Nomoto *et al.*, *Ap. J.* **286** 644 (1984).

[35] M.D. Leising & G.H. Share, *Ap. J. 328* 755 (1988).

[36] D.D. Clayton & F. Hoyle, *Ap. J. Lett.* **187** L101 (1974).

[37] R. Ramaty & R.E. Lingenfelter, *Ap. J. Lett.* **213** L5 (1977).

[38] W.A. Mahoney *et al.*, *Ap. J.* **286** 578 (1984).

[39] R. Diehl *et al.*, *Ap. J. Suppl.* **92** 429 (1994).

[40] W. Chen *et al.*, *Ap. J. Lett.* **440** L57 (1995).

[41] R. Diehl *et al.*, *HEA Division Conference, AAS* San Diego, California (1996).

[42] T.A. Weaver & S.E. Woosley, *Phys. Rep.* **227** 65 (1993).

[43] H. Bloemen *et al.*, *Astr. Ap. Lett.* **281** L5 (1994).

[44] S.E. Woosley *et al.*, *Ap. J.* **411** 823 (1993).

[45] A.M. Bykov & G.D. Fleishman, *Mon. Not. R. Astr. Soc.* **255** 269 (1992).

[46] A.M. Bykov & H. Bloemen, *Astr. Ap. Lett.* **283** L1 (1994).

[47] M. Leventhal *et al.*, *Ap. J. Lett.* **225** L11 (1978).

[48] G.K. Skinner *et al.*, *Nature* **330** 544 (1987).

[49] A. Goldwurm *et al.*, *Nature* **371** 589 (1994).

[50] G.R. Riegler *et al.*, *Ap. J. Lett.* **294** L13 (1985).

[51] I.F. Mirabel *et al.*, *Nature* **358** 215 (1992).

[52] W.R. Purcell *et al.*, *The 2nd Compton Symposium* AIP Conference (1994).

[53] N. Guessom *et al.*, *Ap. J.* **378** 170 (1991).

[54] N. Gehrels & J. Tueller, *Ap. J.* **407** 597 (1993).

[55] C.A. Meegan *et al.*, *Nature* **355** 143 (1992).

[56] C.A. Meegan *et al.*, in *Gamma-Ray Bursts : 2nd Workshop, Huntsville, AL* eds. G.J. Fishman, J.J. Brainerd & K. Hurely **New York : AIP** 3 (1994).

[57] M.S. Briggs, *Ap. J.* **407** 126 (1993).

[58] M.S. Briggs *et al.*, *Ap. J.* **459** 40 (1996).

[59] J.E. Hakkila *et al.*, *Ap. J.* **422** 659 (1994).

[60] J.P. Norris *et al.*, *Ap. J.* **424** 540 (1994).

[61] F. Ma & B Xie, *Ap. J. Lett.* **462** L63 (1996).

[62] R. Antonucci, *Ann. Rev. Astr. Ap.* **31** 473 (1993).

[63] B.L. Fanaroff & J.M. Riley, *Mon. Not. R. Astr. Soc.* **167** 31 (1974).

[64] P.D. Barthel, *Nature* **362** 326 (1993).

[65] A.J. Lawson, *Ph.D. Thesis* 122 (1995).

[66] A.A. Zdziarski *et al.*, *Mon. Not. R. Astr. Soc. Lett.* **269** L55 (1994).

[67] K. Nandra & K. Pounds, *Mon. Not. R. Astr. Soc.* **268** 405 (1994).

[68] E. Jourdain *et al.*, *Astr. Ap. Lett.* **256** L38 (1992).

[69] A.A. Zdziarski *et al.*, *Ap. J. Lett.* **363** L1 (1990).

[70] F. Perotti *et al.*, *Ap. J. Lett.* **247** L63 (1981).

[71] F. Perotti *et al.*, *Ap. J.* **373** 75 (1991).

[72] M. Maisack *et al.*, *Ap. J. Lett.* **407** L61 (1993).

[73] G.M. Madejski *et al.*, *Ap. J.* **438** 672 (1995).

[74] M. Maisack *et al.*, *Astr. Ap.* **284** 28 (1994).

[75] R.D. Blandford, in *Theory of Accretion Disks* eds. P. Meyer, W. Duschl, J. Frank & E. Meyer-Hofmeister (1989).

[76] K.S. Thorne *et al.*, *Black Holes : The Membrane Paradigm* **New Haven : Yale University Press** (1986).

[77] P. Padovani & C.M. Urry, *Ap. J.* **356** 75 (1990).

[78] P.D. Barthel, *Ap. J.* **336** 606 (1989).

[79] E. Jourdain *et al.*, *Ap. J.* **412** 586 (1993).

[80] G. Ghisellini, *Mon. Not. R. Astr. Soc.* **236** 341 (1989).

[81] M.C. Begelman & M. Sikora, *Ap. J.* **322** 650 (1987).

[82] F. Melia & A. Königl, *Ap. J.* **340** 162 (1989).

[83] S.D. Bloom & A.P. Marscher, in *Compton Gamma-Ray Observatory* eds. N. Gehrels & M. Friedlander **New York : AIP** 578 (1993).

[84] R.D. Blandford & A. Levinson, *Ap. J.* **441** 79 (1995).

[85] R.A. Cameron *et al.*, in *Compton Gamma-Ray Observatory* eds. N. Gehrels & M. Friedlander **New York : AIP** 478 (1993).

[86] Y. Rephaeli & O. Goldshmidt, *Ap. J.* **397** 438 (1992).

[87] A. Bazzano *et al.*, *23rd ESLAB Symposium* **ESA SP-296** 901 (1989).

[88] C. Hanson *et al.*, *Mon. Not. R. Astr. Soc.* **242** 262 (1990).

[89] D.E. Gruber, in *The X-Ray Background* eds. A. Barcons & A. Fabian **CUP** (1992).

[90] A. Comastri *et al.*, *Astr. Ap.* **296** 1 (1995).

[91] A.A. Zdziarski it et al., *Ap. J. Lett.* **438** L63 (1995).

[92] A.A. Zdziarski, *Mon. Not. R. Astr. Soc. Lett.* **281** L9 (1996).

[93] F.W. Stecker *et al.*, *Phys. Rev. Lett.* **27** 1469 (1971).

[94] V. Schonfelder *et al.*, *Ap. J.* **240** 350 (1980)

[95] W.N. Johnson *et al.*, in *Proceedings of the 2nd Compton Symposium* eds. C.E. Fichtel, N. Gehrels & J.P. Norris **AIP Conf. Proc. 304** 515 (1994).

[96] A. Comastri *et al.*, in *Proceedings of the 3rd Compton Symposium, Munich* in press (1995).

[97] S.C. Kappadath *et al.*, in *Proceedings of the 3rd Compton Symposium, Munich* in press (1995).

[98] P. von Ballmoos, *Exp. Astr.* **6** 85 (1995).

[99] B. Aschenbach, *Rep. Prog. Phys.* **48** 579 (1985).

[100] L.P. Van Speybroeck, *Ap. Lett. Comm.* **26** 127 (1987).

[101] P. Gorenstein & K. Joensen, *Exp. Astr.* **6** 109 (1995).

[102] C.E. Fichtel *et al.*, *Ap. J.* **198** 163 (1975).

[103] K. Bennett *et al.*, *NASA* **CP-002** 27 (1977).

[104] C.E. Fichtel *et al.*, *Astr. Ap. Suppl.* **97** 13 (1993).

[105] V. Schonfelder *et al.*, *IEEE Trans. Nucl. Sci.* **NS-31** 766 (1984).

[106] L. Mertz, in *Proceeds of the Symposium on Modern Physics* ed. J. Fox **Brooklyn : Polytechnic Press** 787 (1968).

[107] H.W. Schnopper *et al.*, *Space Sci. Rev.* **8** 534 (1968).

[108] N. Lund, *Ap. Space Sci.* **75** 145 (1981).

[109] S. Brandt & N. Lund, *Adv. Space Res.* **10** 239 (1990).

[110] A.P. Willmore, *Mon. Not. R. Astr. Soc.* **147** 387 (1970).

[111] A. Cruise, *Mon. Not. R. Astr. Soc.* **171** 305 (1975).

[112] J.A. Hogböm, *Astr. Ap. Suppl.* **15** 417 (1974).

[113] D. Palmer & T.A. Prince, *IEEE Trans. Nucl. Sci.* **NS-34** 71 (1987).

[114] A. Adams *et al.*, *Astr. Ap.* **20** 121 (1972).

[115] L. Mertz & N.O. Young, in *Proceedings of the Conference on Optical Instruments & Techniques* ed. K.J. Habell **London : Chapman & Hill** 305 (1961).

[116] W.L. Rogers *et al.*, *J. Nucl. Med.* **21** 371 (1980).

[117] R.H. Dicke, *Ap. J. Lett.* **153** L101 (1968).

[118] J.G. Ables, *Proc. Astr. Soc. Aus.* **1** 172 (1968).

[119] R.L. Blake *et al.*, *Rev. Sci. Instr.* **45** 513 (1974).

[120] J. Gunson & B. Polychronopulos, *Mon. Not. R. Astr. Soc.* **177** 485 (1976).

[121] L.D. Baumert, *J. Appl. Math.* **17** 826 (1969).

[122] L.D. Baumert, *Lecture Notes in Mathematics No. 182 : Cyclic Difference Sets* **Berlin : Springer** (1971).

[123] R.J. Procter *et al.*, *Mon. Not. R. Astr. Soc.* **187** 633 (1979).

[124] M. Hall, *Combinational Theory* **Waltham : Blaisdell** (1967).

[125] J. Singer, *Trans. Am. Math. Soc.* **43** 377 (1938).

[126] E.C. Johnsen, *J. Algebra* 4 388 (1966).

[127] M.H. Finger & T.A. Prince, in *Proceedings of the 19th International Cosmic Ray Conference, La Jolla, USA* **9** 295 (1985).

[128] E.E. Fenimore & T.M. Cannon, *Appl. Opt.* **17** 337 (1978).

[129] F.J. MacWilliams & N.J.A. Sloane, *Proc. IEEE* **64** 1715 (1976).

[130] S. Miyamoto, *Space Sci. Instr.* **3** 473 (1977).

[131] A.B. Giles, *Appl. Opt.* **20** 3068 (1981).

[132] S.R. Gottesman & E.E. Fenimore, *Appl. Opt.* **28** 4344 (1989).

[133] S.R. Gottesman & E.J. Schneid, *IEEE Trans. Nucl. Sci.* **NS-33** 745 (1986).

[134] K. Byard, *Nucl. Instr. Meth.* **A322** 97 (1992).

[135] K. Byard, *Nucl. Instr. Meth.* **A336** 262 (1993).

[136] G.K. Skinner & R.M. Rideout, in *Imaging in High Energy Astronomy* eds. L. Bassani & G. di Cocco **Kluwer Academic** 177 (1995).

[137] H.F. Van Beek *et al.*, *IEEE Trans. Nucl. Sci.* **NS-31** 791 (1984).

[138] P. Ubertini *et al.*, *Space Sci. Instr.* **5** 237 (1981).

[139] P. Ubertini *et al.*, *Nucl. Instr. Meth.* **217** 97 (1983).

[140] P. Ubertini *et al.*, *Space Sci. Rev.* **46** 1 (1987).

[141] J.E. Bateman *et al.*, *Nucl. Instr. Meth.* **217** 77 (1983).

[142] J.E. Bateman *et al.*, *Nucl. Instr. Meth.* **225** 209 (1984).

[143] I. Dorion *et al.*, *IEEE Trans. Nucl. Sci.* **NS-34** 442 (1987).

[144] H. Nguyen Ngoc *et al.*, *Nucl. Instr. Meth.* **172** 603 (1980).

[145] J. Davelaar *et al.*, *Space Sci. Rev.* **30** 591 (1981).

[146] J. Davelaar *et al.*, *IEEE Trans. Nucl. Sci.* **NS-29** 142 (1982).

[147] A. Smith *et al.*, *IEEE Trans. Nucl. Sci.* **NS-34** 57 (1987).

[148] EIDOS, A Proposal Submitted to ESA for the INTEGRAL M2 Mission **Part 1 : Scientific & Technical Plan** (1995).

[149] H.O. Anger, *Rev. Sci. Instr.* **29** 27 (1958).

[150] T. Carter, *Ph.D Thesis* **University of Southampton** (1995).

[151] I.B. Pleasants, *Ph.D Thesis* **University of Southampton** (1994).

[152] R.M. Rideout & G.K. Skinner, *Astr. Ap.* **in press** (1996).

[153] C.W. Helstrom, *J. Opt. Soc. Am.* **57** 297 (1967).

[154] R. Willingale *et al.*, *Nucl. Instr. Meth.* **221** 60 (1984).

[155] W.K. Pratt, *Digital Image Processing* **New York : John Wiley & Sons** (1978).

[156] J. Maeda & K. Murata, *Appl. Opt.* **83** 857 (1984).

[157] E. Caroli *et al.*, *Space Sci. Rev.* **45** 349 (1987).

[158] G.K. Skinner & M.R. Nottingham, *Nucl. Instr. Meth.* **A333** 540 (1993).

[159] P.R. Bevington & D.K. Robinson, *Data Reduction & Error Analysis for the Physical Sciences* **New York : McGraw-Hill** (1994).

[160] W.H. Richardson, *J. Opt. Soc. Am.* **62** 55 (1972).

[161] L.B. Lucy, *Astr. J.* **79** 745 (1974).

[162] F. Murtagh *et al.*, *Astr. Ap. Suppl.* **112** 179 (1995).

[163] R. Willingale, *Mon. Not. R. Astr. Soc.* **194** 359 (1981).

[164] B.R. Frieden, *J. Opt. Soc. Am.* **62** 511 (1972).

[165] S.F. Gull & G.J. Daniell, *Nature* **272** 686 (1978).

[166] A.C. Fabian *et al.*, *Mon. Not. R. Astr. Soc.* **193** 175 (1980).

[167] J. Skilling *et al.*, *Mon. Not. R. Astr. Soc.* **187** 145 (1979).

[168] R. Narayan & R. Nityananda, *Ann. Rev. Astr. Ap.* **24** 127 (1986).

[169] N. Weir & S. Djorgovski, in *The Restoration of HST Images & Spectra, Proceedings of the STScI Workshop* eds. R.L. White & R.J. Allen **Baltimore : STScI** (1990).

[170] Tj.R. Bontekoe *et al.*, *Astr. Ap.* **248** 328 (1991).

[171] A.W. Strong, *Exp. Astr.* **6** 97 (1995).

[172] T. Bayes, *Phil. Trans. R. Soc. London* **53** 330 (1763).

[173] J. Skilling, in *Maximum Entropy & Bayesian Methods* ed. J. Skilling **Dordrecht : Kluwer Academics** (1989).

[174] S.F. Gull, in *Maximum Entropy & Bayesian Methods* ed. J. Skilling **Dordrecht : Kluwer Academics** (1989).

[175] N. Weir, *Proceedings of the ESO/ST-ECF Data Analysis Workshop, ST-ECF Newsletter* **16** 8 (1991).

[176] R.K. Piña & R.C. Puetter, *Pub. Astr. Soc. Pac.* **104** 1096 (1992).

[177] R.K. Piña & R.C. Puetter, *Pub. Astr. Soc. Pac.* **105** 630 (1993).

[178] R. Diehl, *Exp. Astr.* **6** 103 (1995).

[179] J-L. Starck *et al.*, *Astr. Ap. Suppl.* **118** 575 (1996).

[180] INTEGRAL, International Gamma-Ray Astrophysics Laboratory **Report on the Phase A Study, ESA/NASA, SCI(93)1** (1993).

[181] J.P. Roques *et al.*, *Adv. Space Res.* **10** 223 (1990).

[182] G. Villa *et al.*, *IEEE Trans. Nucl. Sci.* **NS-34** 26 (1987).

[183] C.J. Eyles *et al.*, *J. Brit. Interplan. Soc.* **40** 159 (1985).

[184] G.F. Bignami *et al.*, *IEEE Trans. Nucl. Sci.* **NS-34** 31 (1987).

[185] IBIS, Experiment Interface Document – B **ISSUE 2.2** (1996).

[186] E.E. Fenimore, *Appl. Opt.* **17** 3562 (1978).

[187] E.E. Fenimore & G.S. Weston, *Appl. Opt.* **20** 3058 (1981).

[188] M.H. Finger, *Ph.D. Thesis* **California Institute of Technology** (1988).

[189] J.B. Stephen, *Ph.D. Thesis* **University of Southampton** (1988).

[190] F. Perotti *et al.*, *Ap. J.* **356** 467 (1990).

[191] T. Carter *et al.*, *Nucl. Instr. Meth.* **A348** 567 (1994).

[192] K. Byard, *Ph.D. Thesis* **University of Southampton** (1989).

[193] W.E. Althouse *et al.*, *Proceedings of the 20th International Cosmic Ray Conference, Moscow, USSR* **1** 84 (1987).

[194] A. Goldwurm, *Exp. Astr.* **6** 9 (1995).

[195] C.E. Covault *et al.*, *IEEE Trans. Nucl. Sci.* **NS-38** 591 (1991).

[196] G.E. Forsyth *et al.*, in *Computer Methods for Mathematical Computations* ed. N.J. Cliffs **Englewood : Prentice-Hill** Chapter 9 (1977).

[197] G.K. Skinner *et al.*, *Ap. Lett. Comm.* **27** 199 (1988).

[198] GRASP, Gamma-Ray Astronomy with Spectroscopy & Positioning **Report on the Phase A Study, ESA, SCI(88)2** (1988).

[199] P. Goupillaud *et al.*, *Geoexploration* **23** 85 (1984).

[200] F. Argoul *et al.*, *Phys. Lett.* **A135** 327 (1989).

[201] S. Grebenev *et al.*, *Ap. J.* **445** 607 (1995).

[202] J. Grindlay *et al.*, in *Imaging in High Energy Astronomy* eds. L. Bassani & G. di Cocco **Kluwer Academic** 213 (1995).

[203] I. Daubechies, *IEEE Trans. Inf. Theory* **36** 961 (1990).

[204] E. Escalera & A. Mazure, *Ap. J.* **388** 23 (1992).

[205] E. Slezak *et al.*, *Astr. Ap.* **227** 301 (1990).

[206] M. Lampton *et al.*, *Ap. J.* **208** 177 (1976).

[207] R. Brun, *GEANT - Detector Description and Simulation Tool* **W1503** (1994).

[208] K.S. Krane, *Introductory Nuclear Physics* **New York : John Wiley & Sons** (1988).

[209] F. Lebrun *et al.*, in *Imaging in High Energy Astronomy* eds. L. Bassani & G. di Cocco **Kluwer Academic** 339 (1995).

[210] W.R. Cook *et al.*, *IEEE Trans. Nucl. Sci.* **NS-31** 771 (1984).

[211] K. Byard & D. Ramsden, *Nucl. Instr. Meth.* **A342** 600 (1994).

[212] D. Ustundag *et al.*, in *Maximum Entropy & Bayesian Methods* eds. W.T. Grandy & L.H. Schick **Kluwer Academic** 295 (1991).

[213] Ph. Laudet & J.P. Roques, *Nucl. Instr. Meth.* **A267** 212 (1988).

[214] F.O. Seward, *J. Brit. Interplan. Soc.* **31** 83 (1978).

[215] P.H. Connell & G.K. Skinner, *INTEGRAL Document* **SPI/Bir/TN005 - 2.1** (1995).

[216] B. Swinyard, *Ph.D. Thesis* **University of Southampton** (1993).

[217] I.B. Pleasants, *priv. comm.* (1993).