UNIVERSITY OF SOUTHAMPTON

FACULTY OF SCIENCE

DEPARTMENT OF PHYSICS

CODED APERTURE IMAGING WITH A HURA CODED APERTURE AND A DISCRETE PIXEL DETECTOR

by

Kevin Byard

A thesis submitted for the degree of
Doctor of Philosophy

November 1989

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE
PHYSICS

Doctor of Philosophy

GAMMA RAY IMAGING WITH A HURA CODED APERTURE AND A DISCRETE PIXEL DETECTOR

by Kevin Byard

An investigation into the gamma ray imaging properties of a hexagonal uniformly redundant array (HURA) coded aperture and a detector consisting of discrete pixels constituted the major research effort associated with this thesis. Such a system offers distinct advantages for the development of advanced gamma ray astronomical telescopes in terms of the provision of high quality sky images in conjunction with an imager plane which has the capacity to reject background noise efficiently. Much of the research was performed as part of an ESA sponsored study into a prospective space astronomy mission (GRASP). The effort involved both computer simulations and a series of laboratory test images.

A detailed analysis of the system point spread function (SPSF) of imaging planes which incorporate discrete pixel arrays is presented and the imaging quality quantified in terms of the signal to noise ratio (SNR). Computer simulations of weak point sources in the presence of detector background noise were also investigated and revealed that an inherent source of noise in the form of a coding error is present in the vast majority of observations, diminishing the quality of the deconvolved images. The coding errors could be completely removed for systems which employ hexagonal pixels, provided the mask element shadow area was correctly oriented and dimensioned with respect to the imager pixels.

Theories developed during the study were evaluated by a series of experimental measurements with a ^{57}Co gamma ray point source, an Anger camera detector, and a rotating HURA mask. These tests were complemented by computer simulations designed to reproduce, as close as possible, the experimental conditions. The 60° antisymmetry property of HURAs has also been employed to remove noise due to detector systematic effects present in the experimental images, and rendered a more realistic comparison of the laboratory tests with the computer simulations.

Plateau removal and weighted deconvolution techniques were also investigated as methods for the reduction of the coding error noise associated with the gamma ray images. These techniques are shown to be effective, particularly when used together, but only on images in which coding errors are present. No improvement in image quality is offered by these techniques in the "perfect overlap" situation of the HPD.

Publications

Some of the work presented here is to be published elsewhere as follows:

Chapters 5 and 6

K. Byard, A.J. Dean, A. Goldwurm, C.J. Hall, J.S.J. Harding, F. Lei. "Imaging Using HURA Coded Apertures With Discrete Pixel Detector Arrays." *Astronomy and Astrophysics,* **227**, 634-639 (1990).

A. Goldwurm, K. Byard, A.J. Dean, C.J. Hall, J.S.J. Harding. "Laboratory Images With HURA Coded Apertures." *Astronomy and Astrophysics,* **227**, 640-648 (1990).

Other Publications

A.J. Dean, Lei Fan, K. Byard, A. Goldwurm, C.J. Hall. "Radioactivity Induced Background Noise in Space Borne Astronomical Gamma Ray Telescopes Employing Inorganic Scintillation Spectrometers." *Experimental Astronomy,* **1**, 35-45 (1989).

A.J. Dean, Lei Fan, K. Byard, A. Goldwurm, C.J. Hall, J.S.J. Harding. "The Gamma Ray Emissivity of the Earth's Atmosphere." *Astronomy and Astrophysics,* **219**, 358-361 (1989).

C.J. Hall, B. Agrinier, K. Byard, A.J. Dean, A. Goldwurm, J.S.J. Harding, C. Labanti. "Beam Tests of a 3-D Position Sensitive Scintillation Detector." *IEEE Transactions on Nuclear Science,* NS-36, 863-865 (1989).

For my Mum and Dad.

Contents

Chapter 1

Gamma Ray Astronomy.

1.1 Introduction

The extreme technical difficulties associated with the design of effective gamma ray telescopes coupled with the necessity to perform observations above the earth's atmosphere has, until recently, rendered gamma ray astronomy virtually impossible. Consequently the development of gamma ray astronomy has been slow compared to that at lower photon frequencies. However, since the advent of balloon and satellite borne experiments, gamma ray astronomy has become a feasible proposition. Furthermore the development of novel techniques such as coded aperture imaging which enable arcminute source position location, has firmly established gamma ray astronomy as an observational science. Early experiments such as the omnidirectional Apollo Gamma Ray Spectrometer have paved the way for a whole new generation of detectors and telescopes. Consequently the relatively unexplored gamma ray domain, which in the past has been limited mainly to speculation, promises to provide a wealth of new information with the help of a number of new and exciting projects, such as Gamma-1, SIGMA, and the ambitious Gamma Ray Observatory to name a few. The next few years will therefore undoubtedly yield many new discoveries and the above missions render gamma ray astronomy a very exciting branch of observational astrophysics.

Although gamma ray telescope design is very difficult, once the detector design problems have been overcome the bonuses associated with high energy photons are of immense value. Firstly the unique property of gamma ray photons of being unhampered by absorption over

cosmological distances, and in particular through the Galactic disk, enables the study of objects which are invisible to astronomers who make observations in other wavebands. The most striking example of this is the Galactic centre, large areas of which are invisible in the optical domain but can be observed in the gamma ray waveband. This property also makes it possible to see the processes occurring at the centre of dense regions, such as the Galactic centre and active galactic nuclei, which in many cases is impossible at lower frequencies. Furthermore, gamma rays are have the unique property that they are produced by non thermal mechanisms which relate to energetic particle-particle and particle-field interactions, or by processes related to the nuclear structure of matter (e.g. radioactive decay), unlike the lower energy photons such as in the optical and ultraviolet domains, which relate to atomic and thermal processes. Gamma ray observations imply that, for the first time, one can have direct access to study some of the most fundamental energetic processes taking place in astrophysical situations, such as the sun, a multitude of Galactic objects, the interstellar medium, and, over cosmological distances, many extragalactic objects.

1.2 Gamma Ray Emission Mechanisms

1.2.1 Synchrotron Radiation

In high energy astrophysics the emission which dominates non thermal radiation mechanisms is synchrotron radiation. Synchrotron radiation is the emission from a charged particle which gyrates at relativistic velocity in a magnetic field. In the rest frame of the particle of charge e, the rate of energy loss is given by

$$-\left(\frac{dE}{dt}\right)' = \frac{e^2 a'^2}{6\pi\epsilon_0 c^3} \tag{1.1}$$

where a' is the acceleration of the particle in the particle's rest frame. The force on the particle F' is perpendicular to both the magnetic field lines and the particle velocity and is given by

$$F' = ma' = vB\gamma \sin\theta \tag{1.2}$$

where v is the velocity of the particle as seen in the observer's frame, m is the particle rest mass, B is the magnetic field strength and $\gamma = (1 - v^2/c^2)^{-1/2}$ is the Lorentz factor. The angle θ is the pitch angle of the motion of the particle to the magnetic field lines as shown

in fig. 1.1. Substituting for a' from 1.2 into 1.1 gives

$$-\left(\frac{dE}{dt}\right)' = \frac{e^2}{6\pi\epsilon_0 c^3}\frac{(vB\gamma\sin\theta)^2}{m^2}. \tag{1.3}$$

Therefore, all other factors being equal, the rate of energy loss is proportional to m^{-2} and hence is dominated by electrons, where the emission is of the order of 10^6 higher than that due to protons. The value of the Thomson cross section for electrons is $\sigma_T = e^4/6\pi\epsilon_0^2 c^4 m_e^2$ (m_e being the electron mass). This combined with the expression for magnetic energy density, $U_{mag} = B^2/2\mu_0$, and $c^2 = (\mu_0\epsilon_0)^{-1}$ gives

$$-\left(\frac{dE}{dt}\right)' = 2\sigma_T c\gamma^2 U_{mag}\left(\frac{v}{c}\right)^2\sin^2\theta. \tag{1.4}$$

This is the rate of energy loss of an electron as seen in the electron's rest frame. However, it can be shown that the energy loss rate is Lorentz invariant ([89] chapter 3) and so 1.4 is also the correct expression for the energy loss as seen by the observer. However 1.4 is for a single electron moving at a specific pitch angle θ. If the angle θ is averaged for many electrons over all pitch angles the desired relationship is

$$\begin{aligned} -\left(\frac{dE}{dt}\right)_{\text{average}} &= -\int_0^\pi \left(\frac{dE}{dt}\right)'\frac{1}{2}\sin\theta\, d\theta \\ &= \frac{4}{3}\sigma_T c\gamma^2\left(\frac{v}{c}\right)^2 U_{mag}. \end{aligned} \tag{1.5}$$

For a continuous power law input of relativistic electrons of the form

$$N(E) = k_1 E^{-x} \tag{1.6}$$

the resultant photon spectrum is also a power law flux of the form

$$I(\nu) = k_2\nu^{-(x-1)/2}. \tag{1.7}$$

Therefore synchrotron radiation from a celestial source can be identified by this power law, which gives a gradient of $-(x-1)/2$ when plotted on a logarithmic scale.

1.2.2 Synchrotron Self Absorption

If the density of the emitting region is high then some longer wavelength photons will be unable to escape from the source medium and instead are re-absorbed by the particles. The extent of the absorption can be calculated if the low energy Rayleigh-Jeans approximation is used for the particle kinetic energy temperature T_e:

$$T_e \propto \nu^{1/2}. \tag{1.8}$$

Absorption becomes significant if the photon brightness temperature T_b approaches T_e:

$$T_b \simeq T_e. \tag{1.9}$$

Now, the photon flux S_ν for low energy of wavelength λ is given by

$$S_\nu \propto \frac{T_b}{\lambda^2} \tag{1.10}$$

which combined with 1.8 and 1.9 gives

$$S_\nu \propto \nu^{5/2}$$

at low frequency. Therefore the self absorption part of the emission curve has a slope of 5/2 on a logarithmic scale as shown in fig. 1.2.

1.2.3 Bremsstrahlung Radiation

Bremsstrahlung radiation is the radiation emitted by a charged particle which is accelerated in the Coulomb field of an ion. The electrostatic field E_{es} in the vicinity of an ion as seen by a non relativistic charged particle is given by

$$E_{es} = \frac{Ze}{4\pi\epsilon_0 b^2} \tag{1.11}$$

where Ze is the ion charge and b is the collision parameter of the particle with the ion. For a relativistically moving particle 1.11 becomes

$$E'_{es} = \gamma E_{es} = \frac{\gamma Ze}{4\pi\epsilon_0 b^2} \tag{1.12}$$

and for a particle of charge e, the force on the particle is therefore given by

$$eE'_{es} = ma' = \frac{\gamma Ze^2}{4\pi\epsilon_0 b^2} \tag{1.13}$$

Substituting for a' from 1.13 into the expression for the power output of an accelerated charged particle (equation 1.1), and noting that the collision duration time is $\sim b/c$ gives the following expression for the energy output of the particle:

$$E_{\text{particle}} = \frac{e^2}{6\pi\epsilon_0 c^3}\left(\frac{\gamma Ze^2}{4\pi\epsilon_0 b^2 m}\right)^2 \frac{b}{c} \tag{1.14}$$

Therefore, as is the case for synchrotron radiation, electron emission dominates since the emitted energy is proportional to m^{-2}.

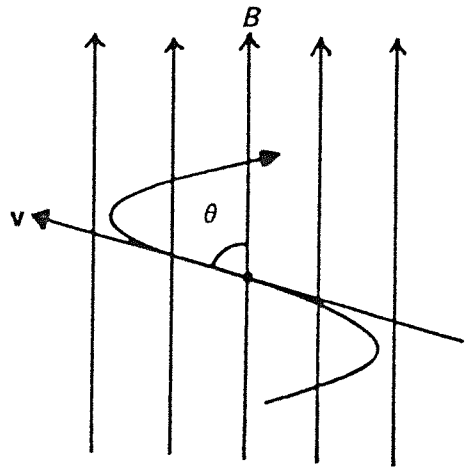


Fig. 1.1. *Definition of pitch angle θ in synchrotron radiation (from Longair [89]).*

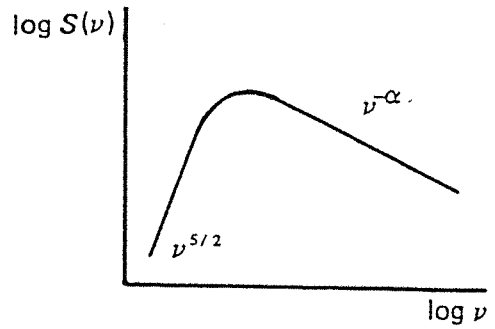


Fig. 1.2. *The spectrum of a source of synchrotron radiation exhibiting synchrotron self absorption (from Longair [89]).*

When there are many electrons and ions, the total energy output is given by

$$E_{\text{tot}} = N_e N_Z c \int_{b_{min}}^{\infty} E_{\text{particle}} 2\pi b \, db$$

giving

$$Etot = \frac{e^6 \gamma^2 N_e N_Z Z^2}{48\pi^2 \epsilon_0^3 c^3 m_e^2 b_{min}} \tag{1.15}$$

where N_e and N_Z are electron and ion concentrations respectively. The value of b_{min} can be calculated using Heisenberg's Uncertainty Principle:

$$p b_{min} \simeq \hbar$$

where p is the electron momentum $= \gamma m_e c$. Therefore

$$b_{min} = \frac{\hbar}{\gamma m_e c}$$

giving

$$E_{tot} = \frac{e^6 \gamma^3 N_e N_Z Z^2}{48\pi^2 \epsilon_0^3 c^2 m_e \hbar} \tag{1.16}$$

An exact quantum mechanical calculation of the total energy emitted by a cloud of electrons travelling at velocities given by a Maxwellian distribution gives an emissivity for bremsstrahlung radiation of

$$I(\nu) = \text{constant.} \frac{Z^2 e^6}{m_e^2} \left(\frac{m_e}{kT}\right)^{1/2} g(\nu, T) N_e N_Z \tag{1.17}$$

where $g(\nu, T)$ is the correction factor known as the Gaunt factor and T is the electron temperature. At frequencies $h\nu \ll kT$ the Gaunt factor has only a logarithmic dependence on frequency (and therefore is approximately equal to 1 since ν is low) whereas for $h\nu \gg kT$, $g(\nu, T)$ is approximately $(h\nu/kT)^{1/2}$.

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. However if the electron has non zero velocity it is possible for the photon to gain energy from the collision and be boosted to a higher energy, a process known as *inverse Compton scattering.* Inverse Compton scattering is most often encountered when, in the centre of momentum frame of the collision, the photon energy is much less than the electron rest mass energy, i.e. $\gamma h\nu \ll m_e c^2$.

If E and E' are the initial energies of the photon in the laboratory frame and the electron rest mass frame respectively, and θ is the laboratory frame collision angle of incidence, then

$$E' = E\gamma\left(1 + \frac{v}{c}\cos\theta\right). \tag{1.18}$$

Now, provided that $\gamma h\nu \ll m_e c^2$ the Compton interaction in the electron rest frame is simply ordinary Thomson scattering. Therefore integrating over all angles assuming isotropic photon distribution, and transforming into the laboratory frame gives an emission rate of

$$\frac{dE}{dt} = \frac{4}{3}\sigma_T c U_{rad}\gamma^2 \tag{1.19}$$

where U_{rad} is the radiation radiation energy density as seen by the laboratory frame ([89] page 282).

The spectrum of the scattered radiation $I(\nu)$ is given by

$$I(\nu) = \frac{3\sigma_T c}{16\gamma^4}\frac{N(\nu_0)}{\nu_0^2}\nu\left[2\nu\ln\left(\frac{\nu}{4\gamma^2\nu_0}\right) + \nu + 4\gamma^2\nu_0 - \frac{\nu^2}{2\gamma^2\nu_0}\right] \tag{1.20}$$

where ν_0 is the frequency of the unscattered isotropic radiation field, and $N(\nu_0)$ is the photon number density. At low frequencies $\nu \ll \gamma^2\nu_0$, the term in square brackets in 1.20 goes to a constant value and hence $I(\nu) \propto \nu$.

1.2.5 Pi-Meson Decay

Neutral pi-mesons are produced by the collision between cosmic rays (high energy protons) and the interstellar gas (hydrogen):

$$\mathrm{p{+}p} \rightarrow \mathrm{p{+}p} + n_1(\pi^+ + \pi^-) + n_2\pi^0$$

or from the collision of protons with antiprotons:

$$\mathrm{p{+}\bar{p}} \rightarrow n_3\pi^0$$

where n_1, n_2 and n_3 are small integers. The threshold for this reaction to take place is given by the total rest mass of the pi-mesons produced. For example if $n_1 = n_2 = 1$ the incoming cosmic ray protons in the first reaction must have an energy $\geq$290 MeV.

The neutral pi-meson which is produced subsequently decays to produce two gamma ray photons:

$$\pi^0 \rightarrow \gamma + \gamma$$

with a reaction half life of 10^{-6} s. In the rest frame of the pi-meson each photon has an energy of 69.8 MeV, which is half the neutral pi meson rest mass energy, but in the frame of a stationary observer the gamma rays have a spread of values dependent upon the kinetic energy and direction of the pi-mesons. This in turn depends on the energy and direction of motion of the incoming cosmic rays. The resultant spectrum is of the form shown in fig. 1.3 where the peak is centred on 69.8 MeV [128].

1.2.6 Electron Positron Annihilation

An electron and a positron at rest will annihilate to produce two gamma ray photons:

$$e^+ + e^- \rightarrow \gamma + \gamma.$$

Two photons must be produced which travel in opposite directions so that the momentum of the collision is conserved. Each photon has the energy of the rest mass of the electron, $m_e c^2$ =511 keV, and hence is relatively simple to locate as a spectral line feature at this energy. The line may be broadened by Doppler motion, as was noticed by Riegler when observing the Galactic centre region, fig. 1.4 [120]. Also in lower density regions it is occasionally possible for the metastable positronium to be formed. Here the electron and positron are in orbit around each other and may lead to three photon production with the photons having a range of values up to, but not exceeding, 511 keV.

1.2.7 Nuclear Gamma Ray Lines

Gamma rays can be produced by the decay of radioisotopes. The principal gamma ray production isotopes are shown in table 1.1, taken from Ramaty and Lingenfelter [118]. Examples of observations of such decay chains are from novae and supernovae explosions, where the ^{56}Ni - ^{56}Co - ^{56}Fe chain, and in particular the longer lived ^{44}Ti - ^{44}Sc - ^{44}Ca chain are to be expected. Another important decay observed in the Galactic centre region by the Max Planck Institut Compton telescope is the ^{26}Al - ^{26}Mg decay giving a line at 1.809 MeV [135,152].

It is also possible for gamma ray emission to occur from excited nuclei which undergo transition from one energy level to a state corresponding to a lower energy level. Since the energy levels of such nuclei are discrete the photons produced are emitted at well defined

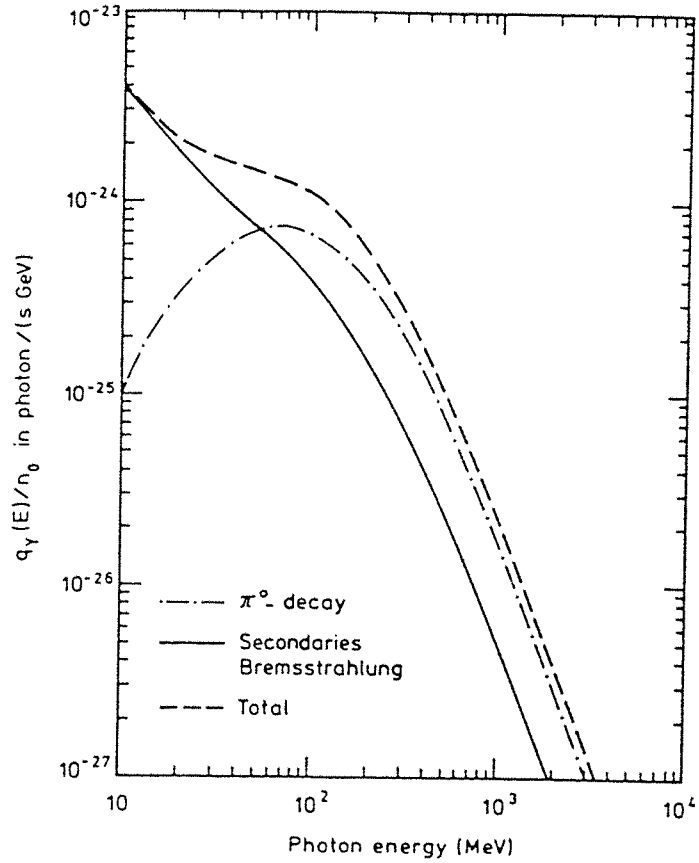


Fig. 1.3. *Spectrum of π^0 meson decay (from Schlickeiser [128]).*

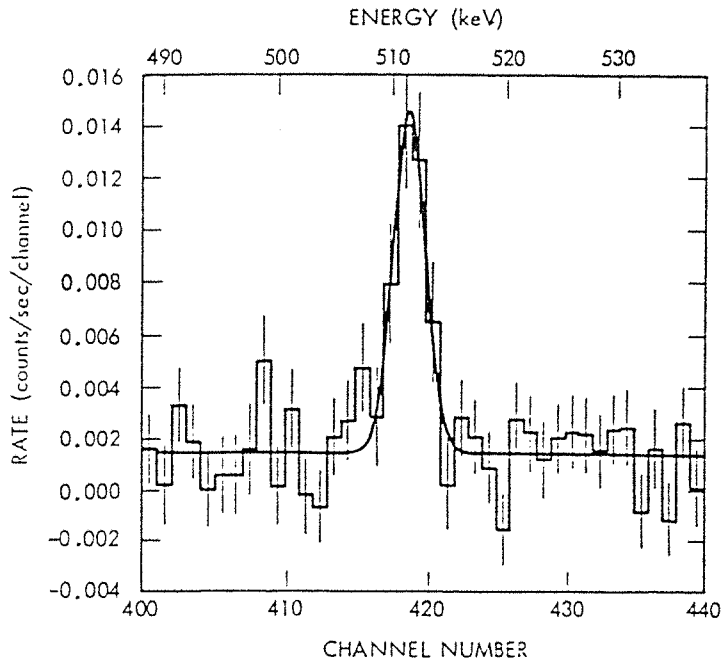


Fig. 1.4. *Net Galactic centre spectrum around 511 keV from HEAO-3 observations (from Riegler et al. [120]).*

energies and the resulting gamma rays are again observed as a spectral line feature. Nuclear excitation can occur in a number of ways:

(i) Radioactive decay, in which the gamma rays come from a daughter nucleus which has been formed in an excited state.

(ii) Collisions with fast moving nucleons which cause the target nuclei to fragment and produce gamma ray lines corresponding to the fragments as well as the target nuclei.

(iii) The capture of thermal neutrons, for example capture by hydrogen to form deuterium:

$$n + {}^1H \rightarrow {}^2D + \gamma$$

where the gamma ray has an energy of 2.23 MeV, the binding energy of the deuterium nucleus.

1.3 A Review of Gamma Ray Astronomical Observations

In recent years results have been obtained indicating that many diverse types of astronomical objects emit gamma radiation. On the Galactic side a variety of known astronomical objects are gamma ray emitters, for example compact objects in the form of pulsars and X ray binary systems, and diffuse emission such as that observed from cloud complexes and from large scale interstellar matter. On the extragalactic side examples of radio galaxies, Seyfert galaxies and quasars have been detected as gamma ray sources, all of which are so gamma ray luminous that their power of emission in this waveband surpasses that in any other. The remainder of this chapter briefly describes a number of past gamma ray observations and speculates upon the astrophysical advances which will be made possible following observations with a sophisticated spectral and imaging device such as the GRASP instrument described in chapter 7.

The two imaging experiments which pioneered gamma ray astronomy were the satellite borne SAS-2 and COS-B instruments. Both employed spark chambers and had a lower energy threshold of ~35 MeV. SAS-2 identified four definite gamma ray point sources: the Crab

and Vela pulsars, Cygnus X-3 and a fourth source with no obvious counterpart (Geminga). COS-B detected 25 discrete sources, among which were the above mentioned Crab and Vela pulsars and the quasar 3C273. However, the poor angular resolution of these instruments made identification of counterparts difficult. Later experiments such as the Birmingham University X Ray Telescope (XRT) and the California Institute of Technology Gamma Ray Imaging Payload (GRIP) have made finer angular resolution observations which have led to improved knowledge of the Galactic centre and the supernova SN1987A as well as a number of other objects.

1.4 Galactic Gamma Ray Astronomy

1.4.1 Galactic Centre

Positron annihilation radiation from the direction of the Galactic centre has been observed on a number of occasions [73,67,83,120,121]. Due to poor directionality of the observing instruments however, the origin of the emission cannot be located to much better than 10°, although the six month time variability observed by the HEAO-3 gamma ray spectrometer implies that the sources responsible for the emission are pointlike rather than diffuse [120] with size $\lesssim 10^{16}$ m. Moreover the 511 keV flux observed in the Galactic centre region by the Bell/Sandia group in 1977 and 1979 exceeds the expected diffuse flux at the 3.5σ and 2.7σ significance level adding more weight to the theory that the source is pointlike in nature.

Continuum emission from the Galactic centre has also been observed. In particular a hard spectrum source was reported by Levine [84] from the HEAO-1 A4 survey, located to within 2° of the centre of the Galaxy in the range 80-180 keV. However, recent observations by the Birmingham group from Spacelab-2 showed another source to be significantly brighter than this spectrum would suggest, instead being compatible with a much softer spectrum of index α =0.5 [138]. This illustrates the difficulty of assigning the variable MeV flux to any specific source and the current situation is far from clear.

1.4.2 The Interstellar Medium

The interstellar medium is responsible for three different components of gamma ray emission.

The first component is a continuum emission from interactions of high energy cosmic rays with interstellar matter. This emission around 100 MeV has been extensively studied by SAS-2 and COS-B. The COS-B map of the Galactic plane from 70 MeV to 5 GeV shows a few discrete sources, fig. 1.5 [97], but most of the continuum emission is presumed to be produced in interstellar space. This diffuse Galactic emission is thought to consist of three components; a neutral pi meson decay component from nuclear reactions of cosmic ray protons with interstellar matter, an electron bremsstrahlung component, and a weak inverse Compton component from collisions of cosmic ray electrons with low energy photons. At MeV energies the pi meson decay component can be neglected.

The second component is a line component arising from the interactions of low energy cosmic ray nuclei with interstellar matter. This has been discussed by Ramaty [116]. A variety of lines with narrow and broad profiles are expected, the most prominent ones being those at 4.4 MeV (^{12}C) and 6.15 MeV (^{16}O) as can be seen in fig. 1.6.

The third component is a line component from radioactive nuclei which were produced during various nucleosynthesis processes in celestial objects and which then diffused into interstellar space. The most important decay chains which occur during nucleosynthesis are shown in table 1.1, but so far only the ^{56}Ni - ^{56}Co - ^{56}Fe chain at 0.847 MeV and 1.238 MeV and the ^{26}Al - ^{26}Mg chain at 1.809 MeV have been detected, the former from the recent supernova SN1987A and the latter from the general direction of the Galactic centre. The 1.809 MeV ^{26}Al line was discovered in 1982 by the HEAO-3 spacecraft [93] with subsequent observations being reported by the SMM gamma ray spectrometer [135] and the MPI Compton telescope [152]. Originally, Ramaty and Lingenfelter [117] had considered supernovae to be the production sites of the line, but nowadays other objects such as novae, red giants and Wolf-Rayet stars are also thought to contribute to the emission.

Another important interstellar line emission is the 511 keV annihilation line. To date, measurements are consistent with the existence of a diffuse line component produced in interstellar space with an angular size of the order of $\pm 60°$. Positrons are thought to be supplied by ^{56}Ni, ^{44}Ti, and ^{26}Al decay chains shown in table 1.1.

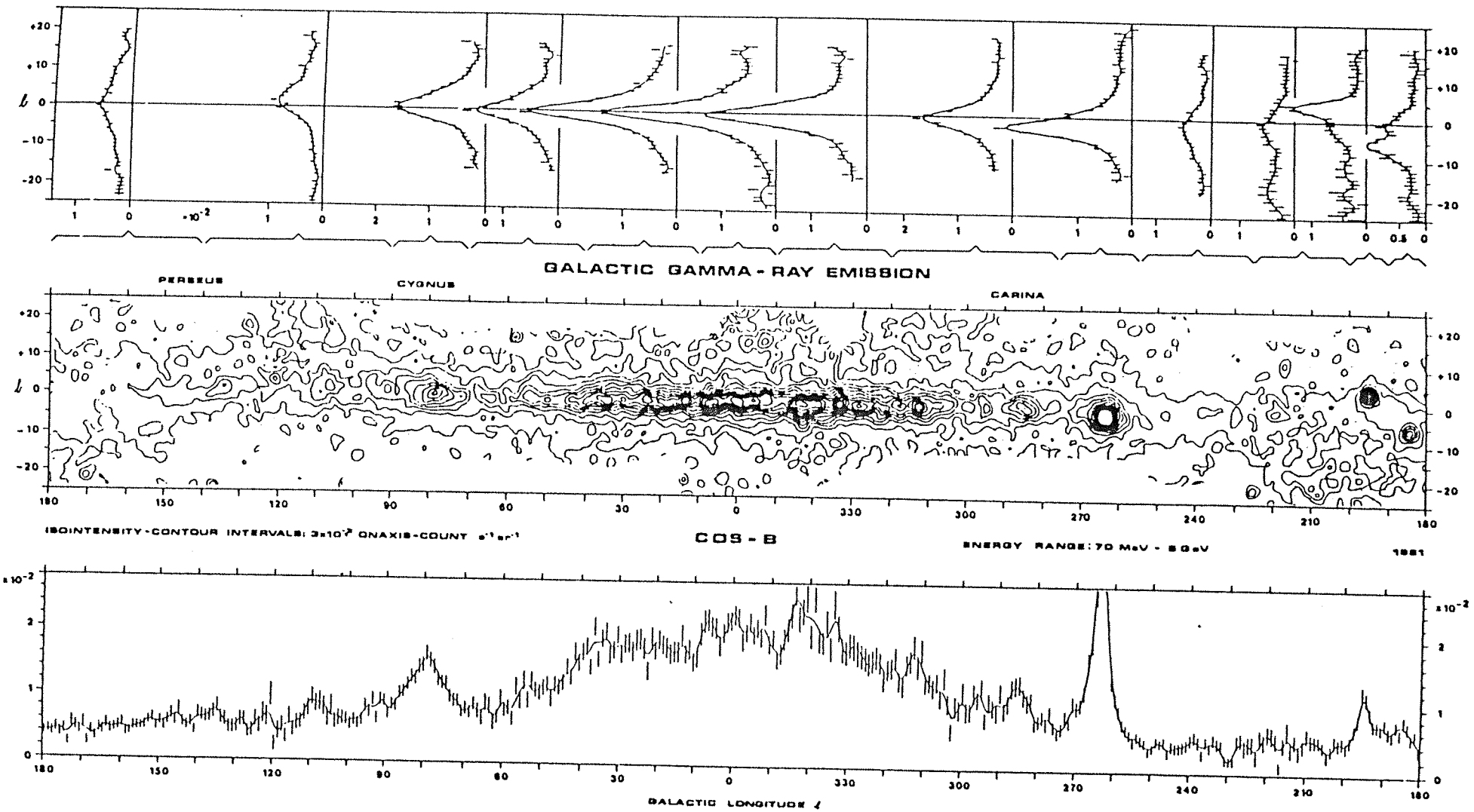


Fig. 1.5. *COS-B sky map of the Galactic plane for gamma rays in the energy range 70 MeV - 5 GeV (from Mayer-Hasselwander et al. [97]).*

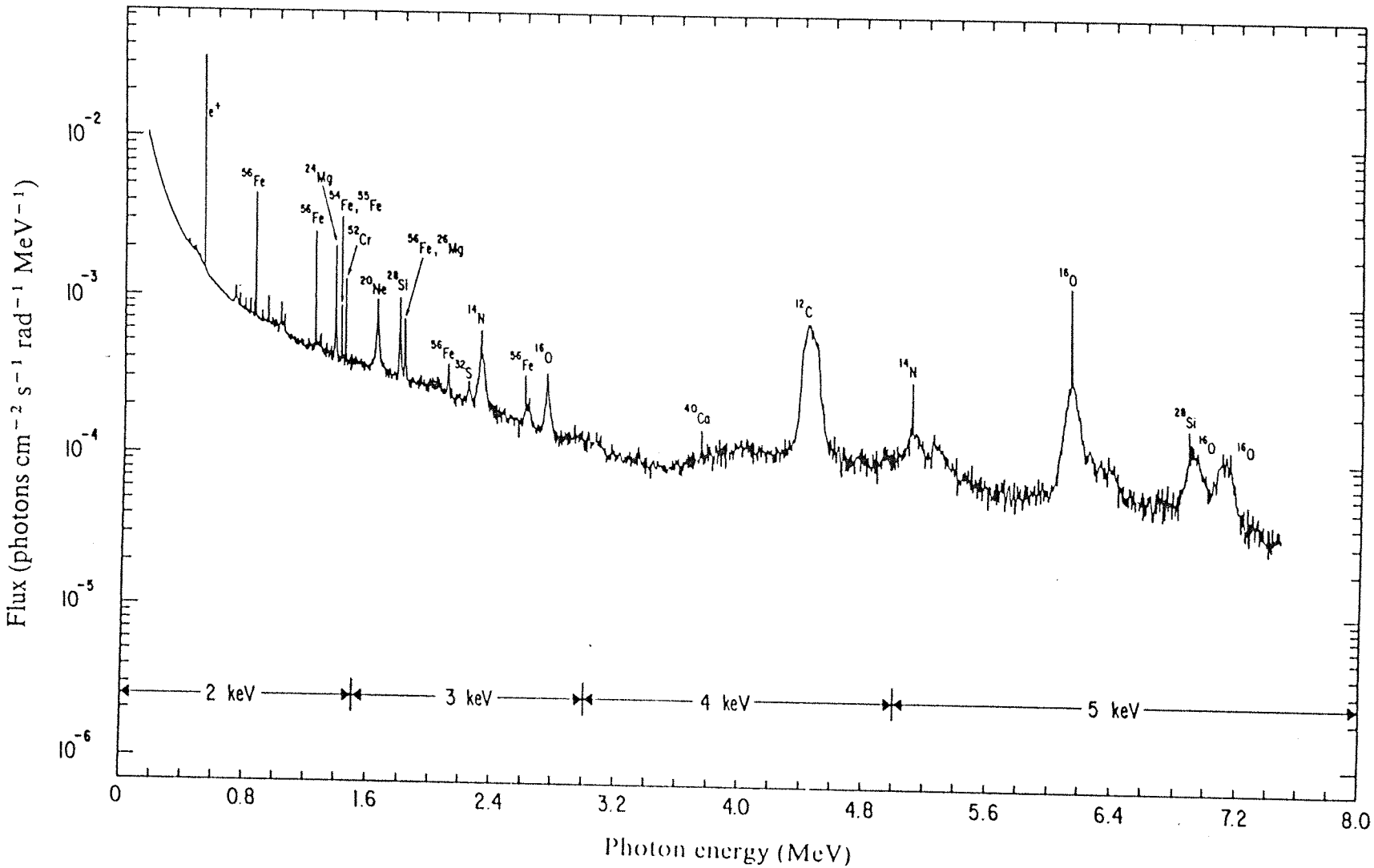


Fig. 1.6. *Predicted line emission from nuclear reactions of energetic particles with interstellar matter from the Galactic centre region (from Ramaty et al. [116]).*

Decay chain	Mean life (years)	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	0.122, 0.014
$^{22}Na \rightarrow {}^{22}Ne$	3.8	e^+, 1.275
$^{44}Ti \rightarrow {}^{44}Sc \rightarrow {}^{44}Ca$	68	e^+, 1.156, 0.078, 0.068
$^{60}Fe \rightarrow {}^{60}Co \rightarrow {}^{60}Ni$	2.2×10^6	1.332, 1.173, 0.059
$^{26}Al \rightarrow {}^{26}Mg$	1.1×10^6	e^+, 1.809

Table 1.1. *Principal gamma ray producing isotopes expected from astrophysical sources.*

1.4.3 Radio Pulsars

The idea that neutron stars are ideal gamma ray emitters arose after the discovery of the Crab and Vela pulsars by the SAS-2 team in the early 1970s. To date there is still no consistent explanation of the phenomena exhibited by pulsar magnetospheres, where the power output peaks in the gamma ray domain [146]. The gamma ray light curve of the Vela pulsar PSR0833-45 has been compared to the optical and radio curves by Kanbach et al. [74], fig. 1.7. The radio emission shows a single peak, similar to the emission of many other pulsars, whilst optical and gamma ray data each have two peaks delayed in phase with respect to the radio peak. This seems to indicate completely different emission mechanisms in each energy band. Also, new analysis of the COS-B Vela pulsar data has suggested the possibility that gamma radiation is almost totally linearly polarised [29] indicating that photon emission is from electromagnetic processes such as magnetobremsstrahlung or curvature radiation (i.e. radiation arising from the relativistic motion of charged particles along curved magnetic field lines), rather than of nuclear origin.

In the case of the Crab pulsar, PSR0531+21, a sharp line feature in the vicinity of 400 keV and width $\lesssim$3 keV was reported by Leventhal [82] with other observations by Yoshimori [165] and Ayre [7] apparently confirming it as a true feature. The fact that an observation by Ling [86] failed to detect this line may mean that the emission is variable. Leventhal interprets the line as possible gravitational redshift of the 511 keV positron annihilation line produced near the surface of the neutron star with a redshift of 0.28. Future observations of the Crab pulsar should confirm the existence of the line, whilst failure to observe lines with a high sensitivity instrument may indicate variability in line emission over a period of years.

1.4.4 X ray Binaries

At present there are about 100 X ray binaries known to emit in the energy range up to 25 keV in the Galaxy, and the Large Magellanic Cloud. The current study of X ray binaries is mainly concerned with binary pulsators and black hole candidates.

The binary pulsators are systems containing a rotating magnetised neutron star accreting from a binary companion. A striking feature of the emission from several of these systems is the appearance of pulsed cyclotron line features. For example such a feature was seen by

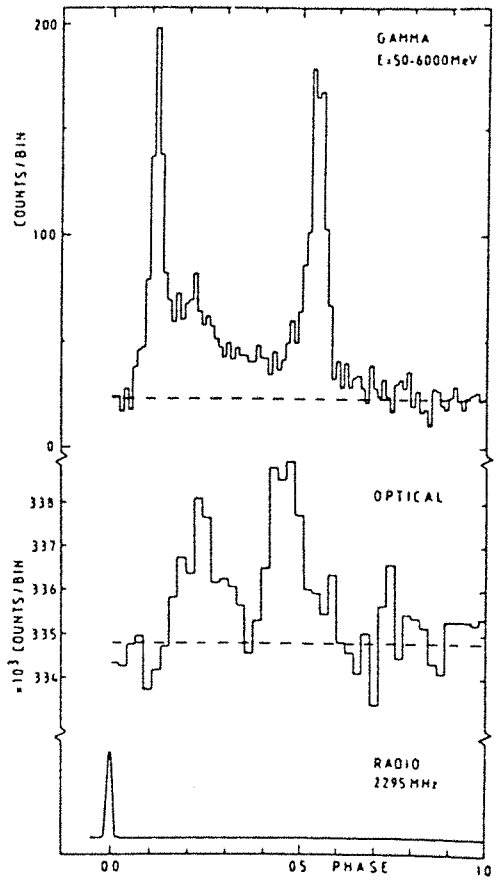


Fig. 1.7. *Light curve of the Vela pulsar in the gamma ray, optical and radio domains (from Kanbach et al. [74]).*

the MPI/AIT group [149] when carrying out a balloon observation of Hercules X-1 in 1976. A line was observed at 58±5 keV with an estimated width of $\lesssim$12 keV, though the spectral resolution of the experiment was poor. The MPI/AIT group interpreted this line feature as a cyclotron emission line from the high magnetic fields ($\sim 10^{12}$ gauss) near the pole of the neutron star. In 1980 a fine resolution measurement was made of Hercules X-1 [150] and the existence of a line at 39 keV was reported. The authors concluded that the two observations are directly comparable after data analysis procedures are considered.

Three binary X ray sources display jets at various wavelengths. These are Sco X-1, Cyg X-3 and SS433 [105]. SS433 has been shown to produce intense optical emission lines which display periodically varying redshifts and blueshifts with full amplitude of approximately 80,000 km s^{-1}. This varying optical emission has been explained as being due to two diametrically opposing relativistic jets travelling at $0.26c$ which precess on a cone of 20° half angle and 164 day period [1]. The X ray emission from SS433 has been studied by Watson [155] using data obtained by the Einstein Observatory imaging detectors. The results showed two bright diffuse lobes of X ray emission symmetrically displaced on either side of the centre of the object up to an angular distance of $\sim$40 arcminutes, as shown in fig. 1.8. SS433 has been associated with the supernova remnant W50 [14], and it seems likely that the X ray emission from the lobes arises from the interaction of the particles of the jets of SS433 with the particle cloud of W50. This is also apparent from radio maps of W50 from Downes et al. [41], where the radio "shell" is seen to be elongated at the exact positions of the X ray lobes. The nature of the emission has been discussed by Watson, and although the observed spectra were consistent with both thermal and power law emission, the thermal emission model is preferred since the pressure within the lobes in the synchrotron interpretation seems too small to produce the observed penetration in the SNR shell. Gamma ray lines have been reported from SS433 [79] at about 1.5 and 1.2 MeV, but their existence has not been confirmed [90]. Lamb [79] interprets the lines in terms of the above kinematic model as blue- and red-shifted components of the 1.369 MeV line from a nuclear transition of ^{24}Mg.

The three X ray binaries which are currently considered to be excellent black hole candidates are Cyg X-1, LMC X-3 and A0620-00. Of the three only Cyg X-1 has been detected in hard X rays ($\geq$300 keV) [103], a region in which the emission shows a significant excess. This is shown in fig. 1.9 where theoretical single temperature and two temperature models appear to fit the data well below 200 keV, but fall significantly short above 200 keV. Observations

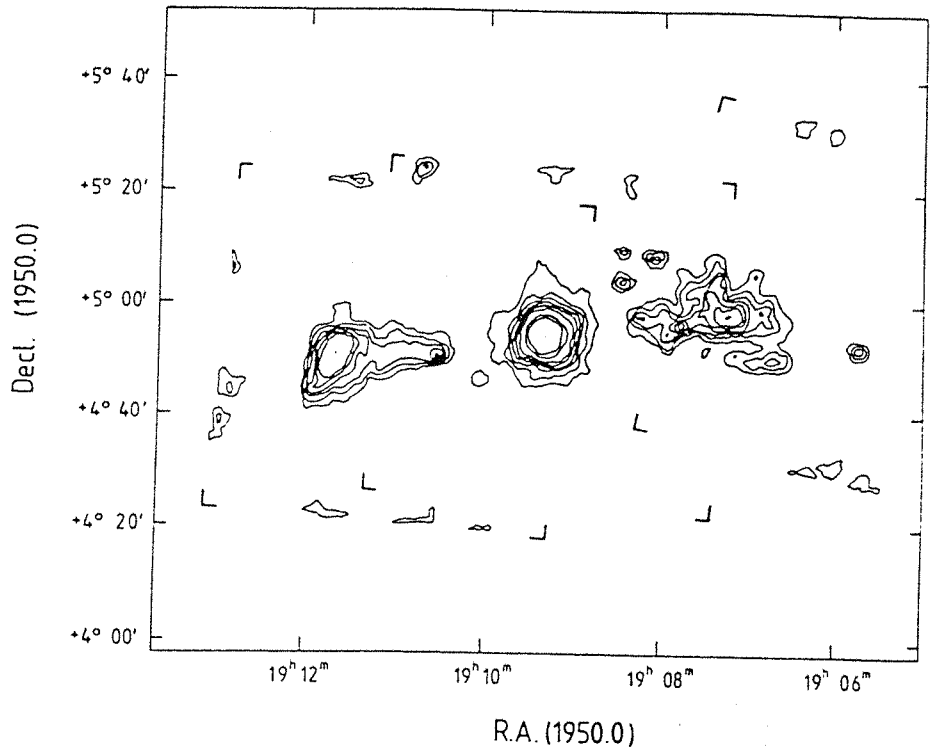


Fig. 1.8. *X ray contour map of SS433 clearly showing the presence of jets in the form of lobes (from Watson et al. [155]).*

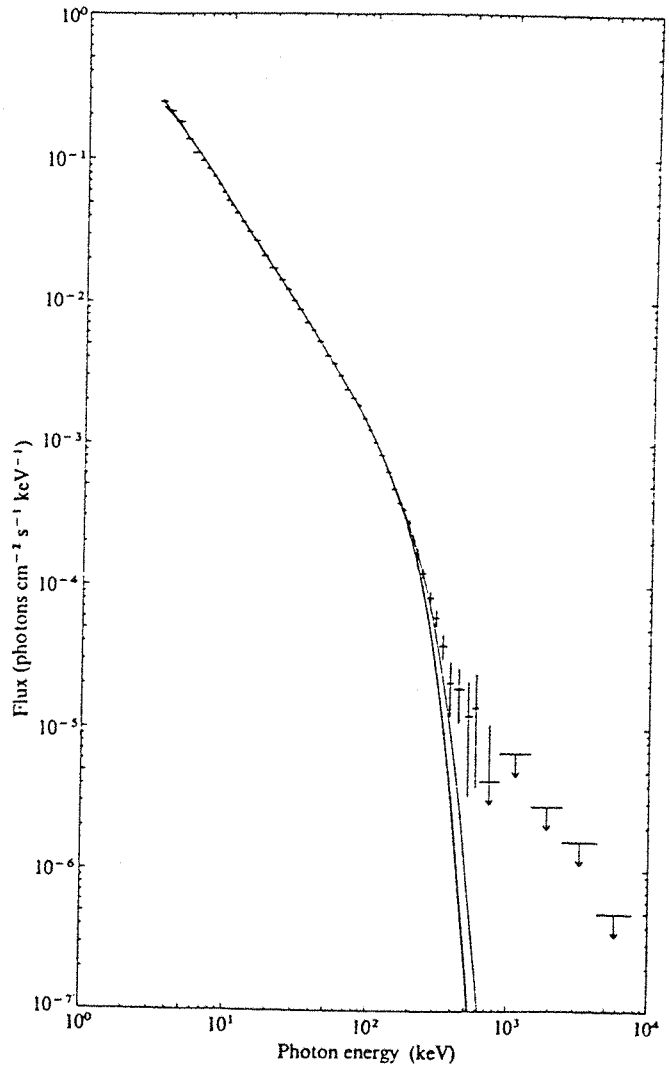


Fig. 1.9. *The average photon spectrum of Cygnus X-1, showing a large excess at high energies (from Nolan et al. [103]).*

by the HEAO-3 experiment in 1979 and 1980 in the range 50 keV - 10 MeV showed a time varying spectrum [87]. Over a period from September 1979 to May 1980 the 100 keV flux increased and the MeV flux disappeared while the overall spectrum softened, pivoting at around 400 keV. This is shown in fig. 1.10 where the solid line is the best fit for the earliest observation, the dashed line is the best fit line for the second observation and the dashed-dot line is for the third observation. Two possible reasons for the MeV component in the first observation may be inverse Compton scattering of copious soft photons by relativistic electrons, or from blueshifted positron annihilation radiation produced in a relativistic electron-positron plasma.

1.4.5 The "Missing Link"

The discovery of the eclipsing millisecond pulsar PSR1957+20 by the Colombia group using the Arecibo radio telescope [53] may identify the "missing link" in pulsar evolution, and connect radio pulsars with low mass X ray binaries. The size of the eclipsing region of the companion star combined with the size of its orbit around the central pulsar led to an apparent paradox, since these dimensions should mean that the companion star is disintegrated by the tidal field of the pulsar. The resolution of this problem was postulated before the discovery of PSR1957+20, by Ruderman [123,124]. The suggestion is that energy released by the gradual spin down of a millisecond pulsar is transferred to the companion in the form of MeV photons, which gives rise to a strong wind from the companion. This wind blocks the radio emission from the pulsar up to a distance many times the radius of the companion, enabling the companion in this case to be compact enough not to be tidally disrupted. The blocking of the pulsar radio emission by the wind causes the observed eclipses.

It is thought that within 10^8 years the companion will be completely evaporated, the pulsar spin down will not be significant, and what will remain is an isolated millisecond pulsar.

1.4.6 Geminga-Like Objects

The gamma ray source Geminga discovered by the SAS-2 satellite [48] and confirmed and studied in greater detail by COS-B [145] represents a special case amongst gamma ray objects. It is the second brightest source in the sky after the Vela pulsar, and the one which has the smallest COS-B error box, with a radius of approximately 0.4°. As such it has

attracted many searches for counterparts at other wavelengths, and an identification has recently been proposed which links it with a peculiar X ray source 1E0630+178 following work done by Bignami et al. [18] using Einstein data. A series of searches for an optical counterpart includes a blue optical identification by Halpern and Tytler [64] and observations using the European Southern Observatory by Bignami et al. [21], who's results agree well with those of Halpern and Tytler but it is apparent that their three datapoints (for B, V and R filters) cannot be fitted to either a blackbody or a nonthermal spectrum, fig. 1.11. It is speculated in [21] that the optical emission could be a proton cyclotron line feature. The value of L_X/L_{opt} is very large, $\sim 3,000$, which combined with the distance to the object points to a very underluminous optical counterpart, typically a neutron star. This would also explain the 59 second periodicity in the soft X ray emission which was reported by Bignami [19].

Radio observations by Sieber and Schlickeiser [137] using the MPI für Radioastronomie 100 m telescope failed to detect a counterpart at frequencies ranging from 1,666 to 10,700 MHz, and so the implication is that Geminga is a neutron star in the non pulsar phase.

The fact that there are of the order of 10^9 undetected neutron stars in the Galactic interstellar medium means that it is likely that many more "Geminga-like" objects exist. Future high sensitivity gamma ray astronomy experiments should provide answers to the fundamental questions of pulsar evolution.

1.4.7 Supernovae

The study of supernova remnants and supernova explosions is of fundamental importance in astrophysics as a test of the theories of explosive nucleosynthesis of the elements. Of particular importance in this context are gamma ray observations, where gamma ray line emissions provide a direct means of determining the quantity of a specific isotope generated in a supernova explosion. Furthermore the shapes of the gamma ray line profiles provide information about the expansion velocity of the inner ejecta, whilst relative line intensities relate to the physical environment, for example density distributions, at the time of production.

It was proposed by Clayton [33] that the exponential light curves of type I supernovae are

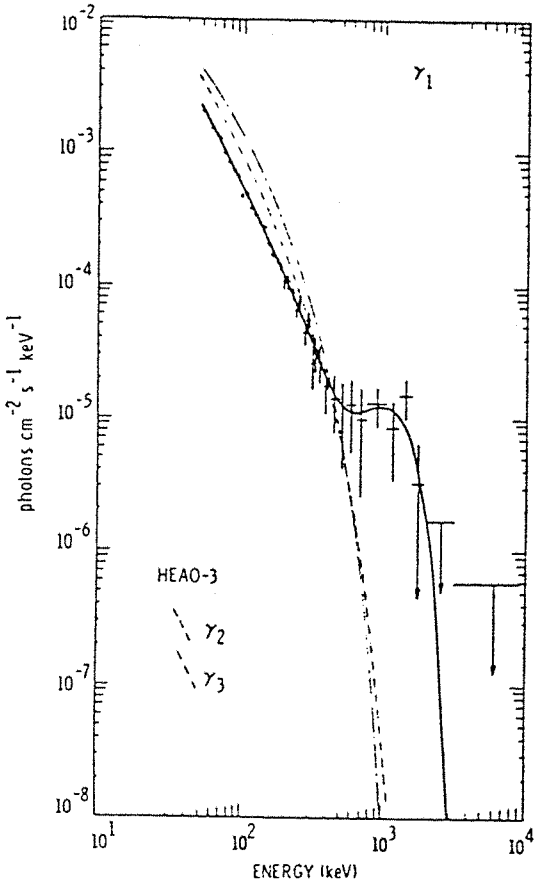


Fig. 1.10. *Variations in the gamma ray spectrum of Cygnus X-1 from September 1979 to May 1980 (from Ling et al. [87]).*

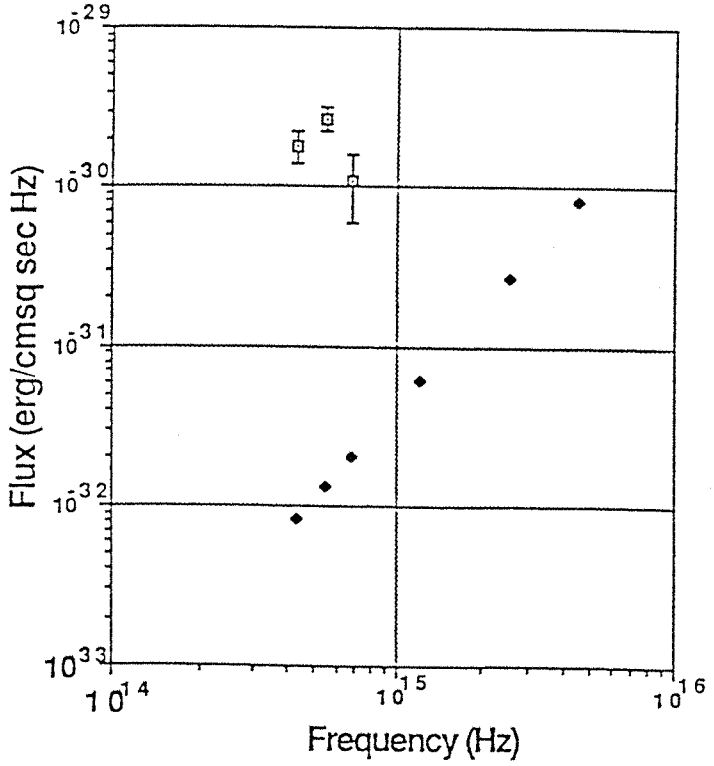


Fig. 1.11. *The observed datapoints of Geminga by Bignami et al. [21]. Also shown is the appropriate Rayleigh-Jeans extrapolation of the 9×10^5 K blackbody curve which best fits the Geminga X ray flux.*

powered by the delayed energy injection resulting from the ^{56}Ni - ^{56}Co - ^{56}Fe decay chain:

$$^{56}\mathrm{Ni} + \mathrm{e}^- \rightarrow {}^{56}\mathrm{Co}$$
$$^{56}\mathrm{Co} \rightarrow {}^{56}\mathrm{Fe} + \mathrm{e}^+ + \gamma.$$

The ^{56}Ni decay has a half life of 6.1 days whilst the ^{56}Co decay, with its principal lines at 0.847 MeV (100%) and 1.238 MeV (68%), has a 77 day half life. However, the gamma rays, which are generated near to the centre of the star, will not penetrate the envelope until the ejecta is sufficiently transparent. This, coupled with the relatively short half lives of the lines, means that the observed line flux will be low. Woosley et al. [162] have estimated that the peak 0.847 MeV line flux from a type I supernova at a distance d Mpc will be

$$F_{0.847} = 10^{-4}(d/10)^{-2} \text{ photons cm}^{-2} \text{ s}^{-1} \tag{1.21}$$

which means that the best sensitivity currently possible restricts 0.847 MeV line detection to galaxies at $d \lesssim 20$ Mpc. Nevertheless, this distance still encompasses the Virgo cluster where type I supernovae are estimated to occur at least three times per year [161].

A longer lived decay chain which in the future gives better prospects of nucleosynthesis line observation within the Galaxy is the ^{44}Ti - ^{44}Sc - ^{44}Ca chain with a half life of 48 years. The excited state of ^{44}Sc, produced from the decaying ^{44}Ti, cascades to its ground state emitting lines at 78 and 68 keV. The ^{44}Sc subsequently decays into the 1.156 MeV state of ^{44}Ca (with 3.9 hour half life). Ramaty and Lingenfelter [119] have calculated the 1.156 MeV flux from a 50 year old supernova remnant at a distance of 10 kpc as

$$F_{1.156} = 10^{-4} \text{ photons cm}^{-2} \text{ s}^{-1} \tag{1.22}$$

and they predict that this line should be clearly seen by the Oriented Scintillation Spectrometer Experiment (OSSE) on board the Gamma Ray Observatory (see section 2.3.1).

SN1987A

An exciting astrophysical event which occurred in 1987 was the observation of a supernova explosion in the Large Magellanic Cloud leaving the now famous object SN1987A. The occurrence of SN1987A provided the first opportunity to study the hard X ray and gamma ray emission from a type II supernova. The first concrete proof of the synthesis of the elements in supernovae came with the detection of gamma ray line emission by Matz et al. who used data from the Gamma Ray Spectrometer on NASA's Solar Maximum Mission [96].

Matz reported observations of a significant ($> 5\sigma$) net line flux at 0.847 MeV, along with evidence for a line at 1.238 MeV, see fig. 1.12. These are the two main lines from the decay of ^{56}Co, thus confirming the predictions of Clayton [33], amongst others. However the lines appeared approximately six months earlier than expected, a phenomenon which was explained by Pinto and Woosley [111] as being due to ^{56}Co mixing. Previous models assumed that there was no mixing in the supernova, but the observed light curve with its early appearance of gamma rays implies that some ^{56}Co lies at a smaller column depth than earlier assumed. However the ^{56}Co must have been made from ^{56}Ni in the bottom most layers ejected by the supernova, and so mixing of ^{56}Co with the outer helium envelope must have occurred during expansion.

As SN1987A ages its spectrum is observed to change as shown in fig. 1.13. Fig. 1.13 shows the spectrum around the 0.847 MeV line at three different times. As the supernova ages the lines become narrower and redder as radiation from deeper material having lower velocity is able to escape due to the increasing transparency of the expanding envelope.

The first gamma ray imaging observations of SN1987A by Cook et al. confirmed that the bulk of the gamma ray emission comes from the supernova and not from any other object in the Large Magellanic Cloud [37] (see fig. 2.12).

1.4.8 Novae

Novae are accretion binary systems in which a dense white dwarf accretes hydrogen from a cool companion. The hydrogen starts to burn, but since the white dwarf is degenerate the increase in temperature does not change the pressure but instead yields a large increase in burning. Therefore the nuclear reactions run out of control and a nova explosion results. This thermonuclear "runaway" is expected to create significant amounts of ^{22}Na and ^{26}Al which emit gamma ray lines at 1.275 MeV and 1.809 MeV respectively. The half lives of these isotopes is 4 years for ^{22}Na and 10^6 years for ^{26}Al and so the mean time between successive Galactic novae of approximately 10 days means that many novae will contribute to the observable line fluxes. The ^{26}Al gamma rays appear as a diffuse Galactic emission (as discussed in section 1.4.2) whereas the ^{22}Na emission will be dominated by ~ 100 recent novae and hence will appear less diffuse.

Lack of knowledge about the physical conditions inside white dwarfs give uncertainties in

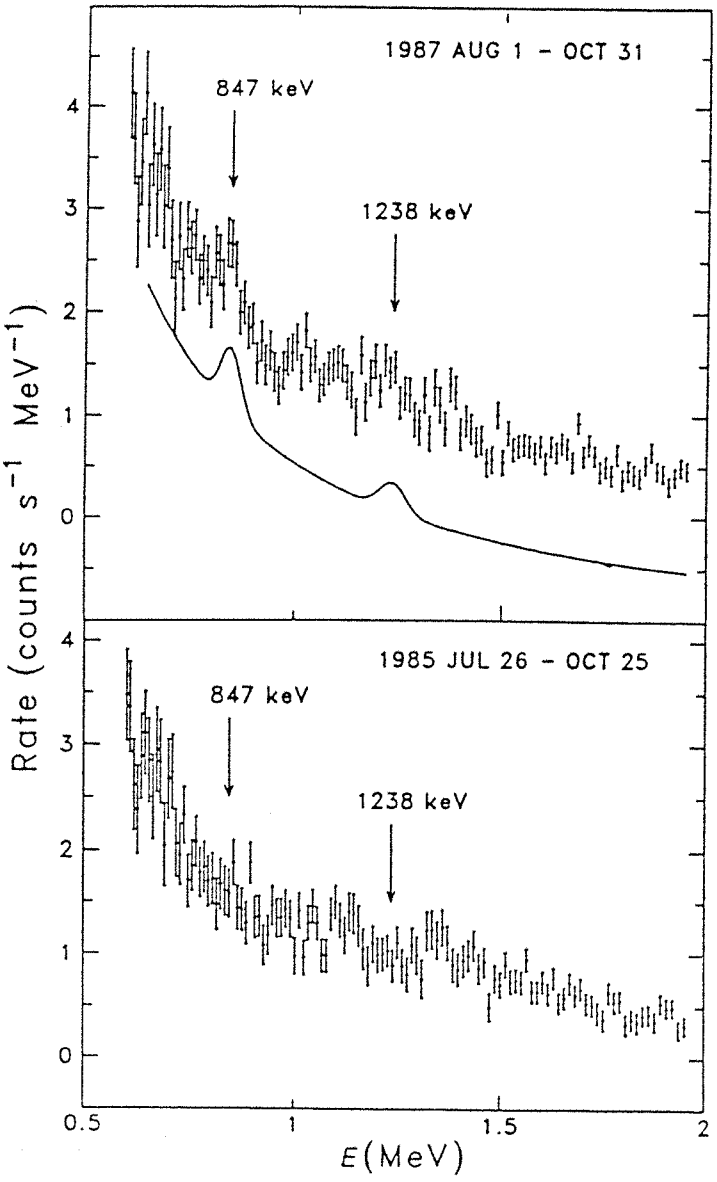


Fig. 1.12. *Spectrum of SN1987A accumulated from 1 August to 31 October 1987 (from Matz et al. [96]).*

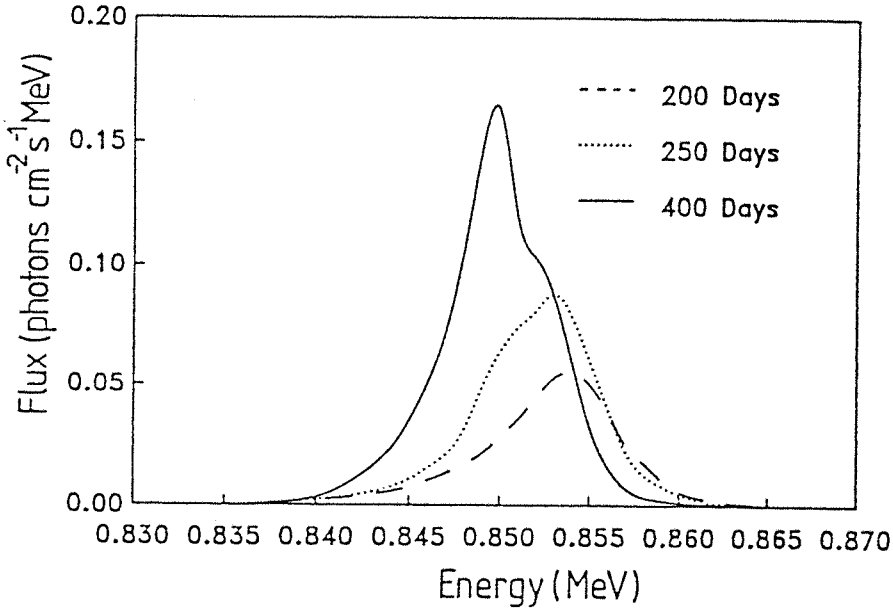


Fig. 1.13. *The gamma ray spectrum of SN1987A in the 0.847 MeV region including only those photons which escape without interaction (from Pinto and Woosley [111]).*

the expected novae ^{22}Na emission, but large variations are expected since production is dependant on the extent of mixing within the white dwarf. Future sensitive gamma ray instruments such as GRASP should be able to search for short lived isotopes since nova explosions become transparent to gamma rays within about a day. This should offer detailed insight into the mechanisms and environment of nova explosions and nova nucleosynthesis.

Also production of ^{7}Li is thought to occur in novae as a result of hydrogen burning sequences at about 1.5×10^{8} K. ^{7}Li is formed by electron capture of ^{7}Be emitting a line at 0.478 MeV, the ^{7}Be being formed via reactions of ^{3}He with ^{4}He.

1.4.9 Hydrostatic Nucleosynthesis Gamma Ray Lines

Apart from supernovae and novae nucleosynthesis can also occur in a non explosive manner. For example ^{26}Al production (1.809 MeV) by Wolf-Rayet stars has been proposed by Prantzos and Cassé [113]. In the most favourable case, they calculate a possible total mass of ^{26}Al due to WR stars of $2.2M_{\odot}$, and postulate that this could account for the total amount of ^{26}Al in the Galaxy. However according to current evaluations the proposed sites cannot account for the total of about $3M_{\odot}$ implied by the observations [58] and higher resolution observations are required to resolve this problem.

Another gamma ray emitting radioactive isotope which is expected to be formed by nucleosynthesis in stars is ^{60}Fe. ^{60}Fe is produced by slow neutron capture during central or shell He burning in massive stars and WR stars. Because ^{60}Fe production in WR stars is 4-5 orders of magnitude lower than that of ^{26}Al, then its 58.6 keV line is undetectable by present gamma ray instruments. This is borne out in the results of a search for the ^{60}Fe line by Mahoney et al. [93] using the HEAO-3 gamma ray spectroscopy experiment. The search failed to detect any such emission.

1.4.10 Gamma Ray Bursts

Since the initial discovery of gamma ray bursts [76] the body of evidence on the spectral behaviour of these events has steadily grown. In their early history, gamma ray burst measurements came mostly from instruments designed for other purposes and were detected serendipitously. The first measurements of gamma ray burst spectra were carried out by

Cline and Desai [34] and have since been measured on a number of occasions by different instruments. Mazets used the burst detector Konus on board the Soviet Venera 11 and Venera 12 space probes to measure the spectrum of a burst which occurred on 22 June 1975 [98].

Gamma ray burst spectra are known to contain line features, and this can be seen in the Konus results in fig. 1.14 where both an absorption feature and an emission feature are evident. The absorption line is located at ~50-70 keV and is probably a cyclotron line feature, whilst the emission line located at ~400 keV is possibly redshifted electron-positron annihilation. Both of these phenomena are associated with dense neutron stars.

The Burst and Transient Source Experiment (BATSE) on board the Gamma Ray Observatory is expected to continuously monitor the whole unocculted region of the sky for any burst phenomena and may help to solve the problem of the exact origin of gamma ray bursts [51] (see section 2.3.1).

1.5 Extragalactic Gamma Ray Astronomy

1.5.1 Active Galaxies

To date only a small number of active galactic nuclei (AGN) of different classes have been detected at X and gamma ray wavelengths. The data so far obtained on AGNs exhibit a number of similar characteristics, the most important of which for the gamma ray astronomer being that the power output for all these objects appears to peak in the gamma ray spectral region. This can be seen in fig. 1.15 where the spectra of some AGNs are shown [12]. This property combined with the extreme penetrating power of gamma ray photons makes them ideal for the study of the physical environment at the AGN nucleus. The measured gamma ray variability is of the order of approximately 6 months indicating that the emission is from a relatively compact region intimately related to the central power house.

Another feature of AGNs is that there appears to be a break in the spectrum at a few MeV [17]. This is possibly due to high energy gamma rays suffering an absorption *en route* by pair production, in photon-photon collisions with lower energy (X ray) photons present in intergalactic and interstellar space, or within the sources themselves [71]. This can occur

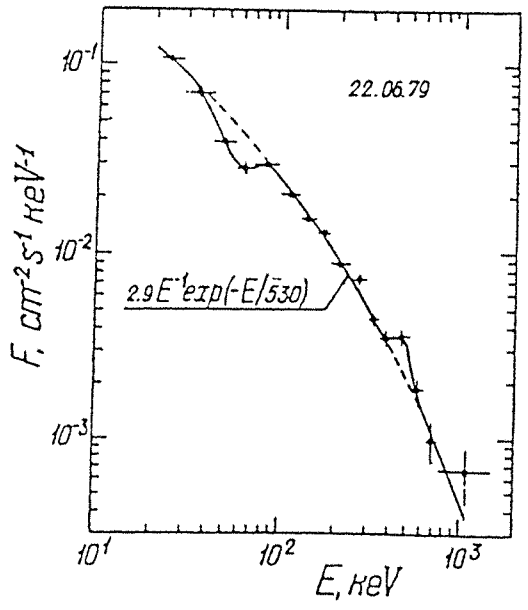


Fig. 1.14. *A gamma ray burst spectrum showing both an absorption line (~ 50 keV) and an emission line (~ 400 keV) (from Mazets et al. [98]).*

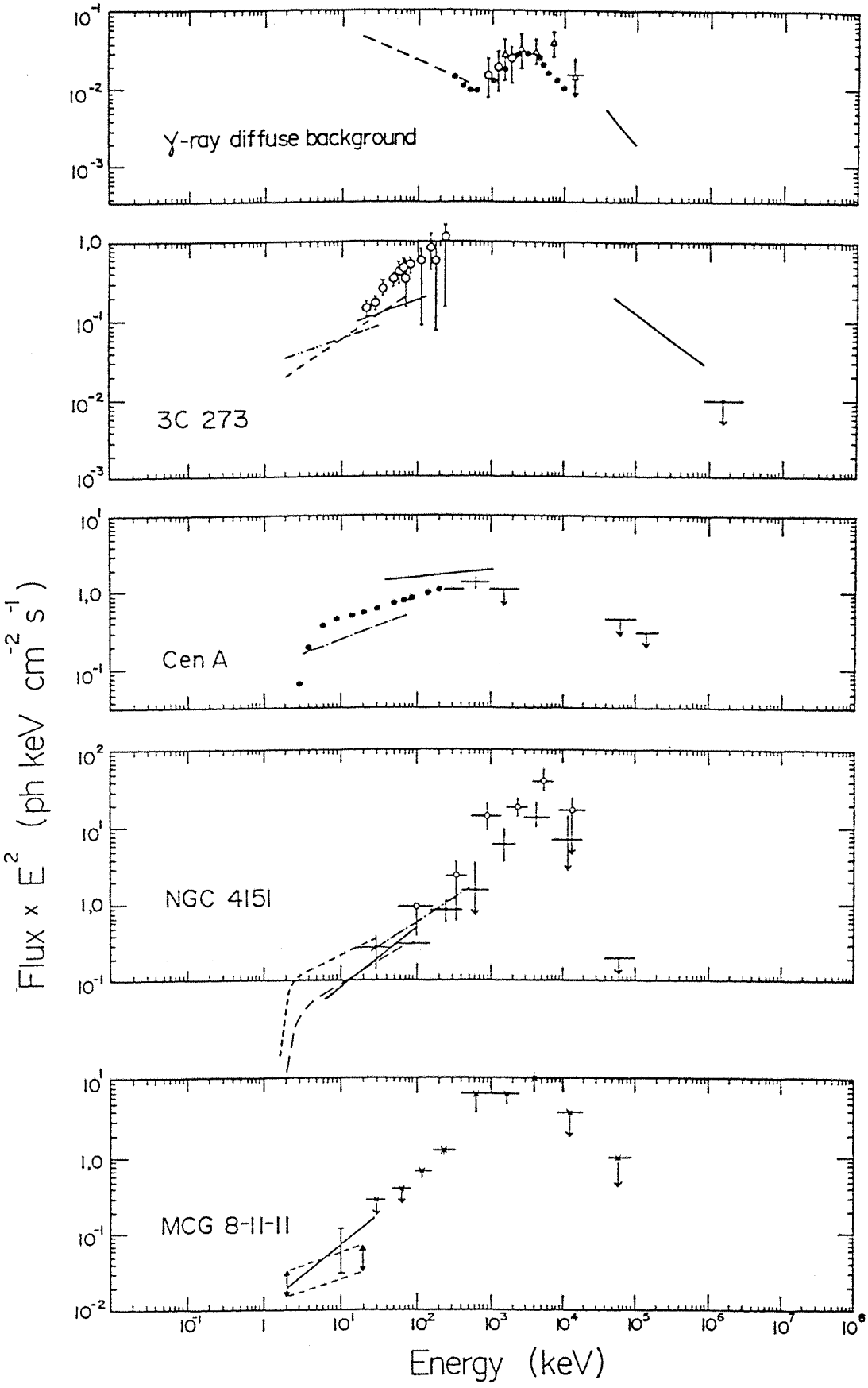


Fig. 1.15. *Energy spectrum of a sample of AGNs and of the cosmic diffuse background (from Bassani and Dean [12]).*

for two photons of energy E_X, E_γ provided that

$$E_X E_\gamma \geq \frac{2(m_e c^2)^2}{1 - \cos\alpha} \tag{1.23}$$

where α is the angle between the directions of motion of the two photons.

A few well known gamma ray emitting AGNs are now described.

The Seyfert Galaxies NGC4151 and MCG08-11-11

NGC4151 has been observed several times in the gamma ray region but at 1 MeV the results prior to 1981 largely may only be represented by upper limits (see Dean and Ramsden [39] for a review of observations). However, a gamma ray excess was measured by the MISO telescope in the energy range 0.02 - 19 MeV. Assuming a distance to NGC4151 of 19 Mpc (from redshift of 0.0032) gives a total luminosity in the range 0.5 - 5 MeV of 5×10^{44} erg s^{-1} [11]. The spectrum is shown in fig. 1.16 along with previous measurements. Assuming that the SAS-2 upper limits [17] and the MISO results are both valid then there must be a break in the spectrum above 1 MeV. In a previous MISO observation in 1977 the measured flux was higher, and Bassani and Dean conclude that the gamma ray flux may be variable with a timescale of less than 6 months [12]. Variability of NGC4151 has also been observed at other wavelengths, including repeated X ray flaring reported by Lawrence from Ariel V data, possibly due to accretion of matter onto a compact object [80].

More recent observations of NGC4151 were carried out by the HEAO-1 experiment [9]. The best fit model for the results is a power law spectrum with an index α =1.6±0.1 in the 2 keV - 2 MeV spectral range. The same authors also plot the 100 keV flux versus the power law spectral index α for a number of previous observations of NGC4151, see fig. 1.17. A weak correlation was found but the small number of datapoints warrants caution in drawing any definite conclusions about the hardening of the spectrum at increased luminosity.

The galaxy MCG08-11-11 was the first Seyfert galaxy to be discovered by means of its X ray emission [154]. The galaxy was found inside the 0.093 square degrees error box of the X ray source 2A0551+466 from the Ariel V catalogue of X ray sources [38] and its redshift of 0.0205 measured by Ward [154] places it at a distance of 123 Mpc. Ward also measured an X ray flux variability on a timescale of less than a month, further demonstrated by the presence of a compact radio source of 3 arcseconds angular diameter. A gamma ray measurement by Perotti et al. [109] using the MISO telescope exhibits spectral characteristics of those

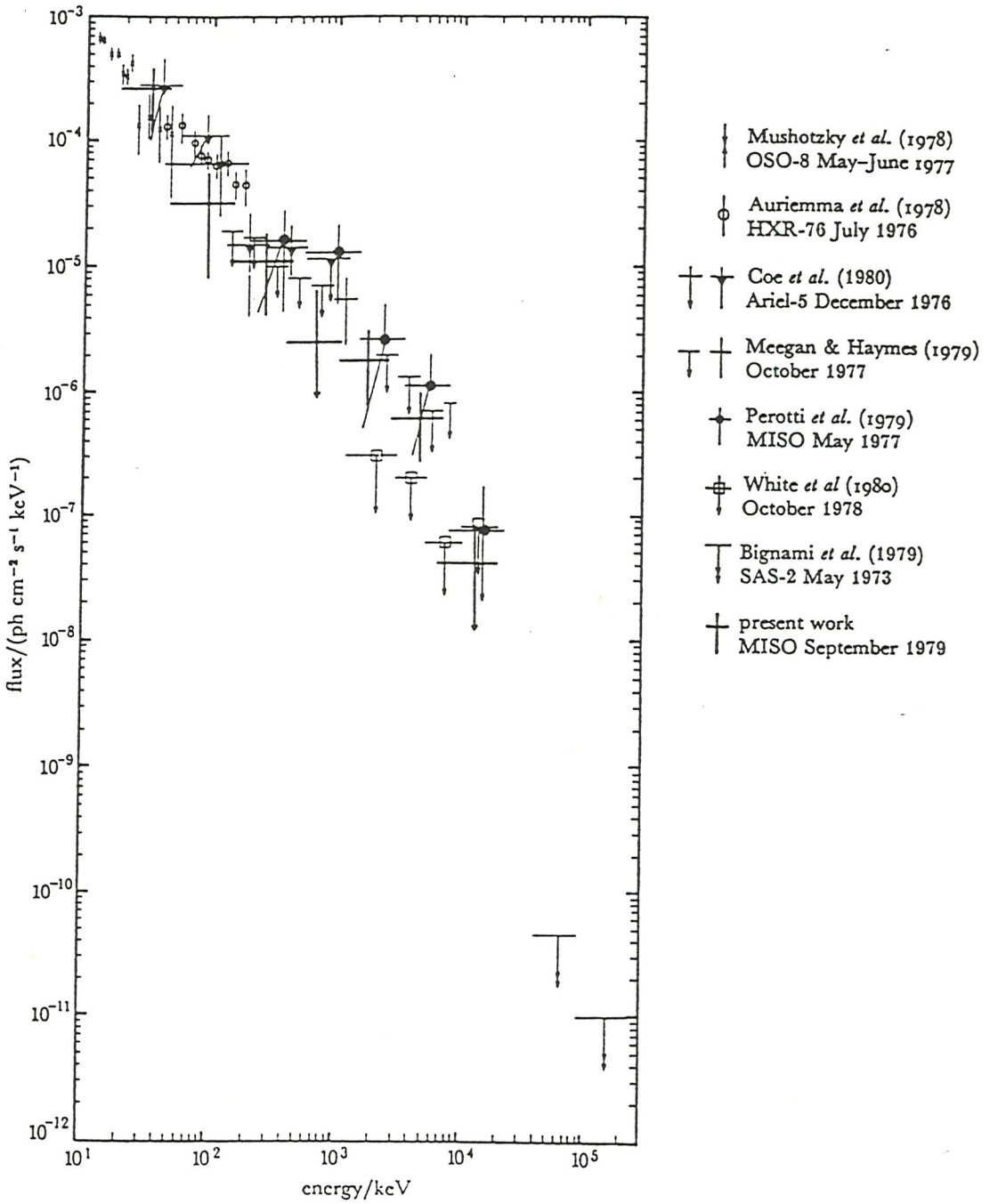


Fig. 1.16. *Photon spectrum of NGC4151 above 10 keV for some observations (from Baker et al. [11]).*

observed from NGC4151, with a break in the power law spectrum somewhere between 2 and 4 MeV, as shown in fig. 1.18.

Centaurus A (NGC5128)

Centaurus A is one of the brightest radio sources and is the nearest active galaxy at a distance of approximately 5 Mpc. It was one of the first extragalactic objects to be identified as an X ray source. An early spectral gamma ray measurement by the Rice University balloon experiment in 1974 yielded a power law spectrum in the range 0.033 - 12.25 MeV of index α =1.9±0.1 [63]. The same observation also gave the first detection of gamma ray lines from an extragalactic source, both 3.3σ above the continuum, and at 1.6 MeV and 4.5 MeV. Since the feature at 1.6 MeV is broad it was postulated by Hall [63] that this line is due to either Doppler broadened de-excitation of ^{20}Ne (1.63 MeV), or from line multiplicity of three de-excitations, namely ^{24}Mg* (1.37 MeV), ^{20}Ne* (1.63 MeV), and ^{28}Si* (1.78 MeV). Hall explains the 4.5 MeV line as probably a de-excitation of ^{12}C* (4.43 MeV), since this is the only abundant isotope which radiates near the observed line energy.

Further observations by the HEAO-1 instrument (2 keV - 2.3 MeV) and by the NASA Goddard Space Flight Centre Low Energy Gamma Ray Spectrometer (LEGS, 70 keV - 8 MeV) also yielded spectral data. The HEAO-1 measurements [8] gave data which fitted a two power law spectrum with $\alpha_1 = 1.60 \pm 0.03$ and $\alpha_2 = 2.0 \pm 0.2$ with a break at approximately 140 keV. Searches for line emission showed no significant excess above the continuum at 511 keV and 1.6 MeV. A six month interval between flux measurements showed a 50% intensity variation but no significant variability occurred on timescales of about 10 days. The NASA/GSFC LEGS data [54] fitted a power law with $\alpha = 1.59 \pm 0.30$, but there was no apparent spectral break nor any evidence of line emission, a significant decrease (a factor of 8) below the Rice University 1974 results. The lack of observed 511 keV annihilation radiation from Centaurus A limits the size of the emission region to be between 10^{11} and 5×10^{15} metres.

The most recent observation of Centaurus A was made by the balloon borne MPI Compton telescope [153]. Above 700 keV the statistical significance of the source detection was 4.1σ and the shape of the energy spectrum strongly suggests that the detected radiation is related to Centaurus A. The spectrum extends well into the MeV range at least up to 8 MeV and connects well to the previous power law spectral measurements as shown in fig. 1.19. If the

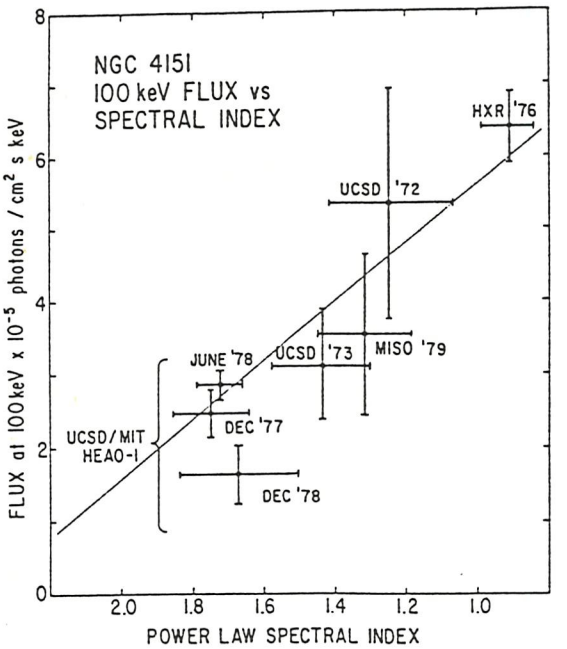


Fig. 1.17. *100 keV flux versus spectral index α for a range of observations of NGC4151 (from Baity et al. [9]).*

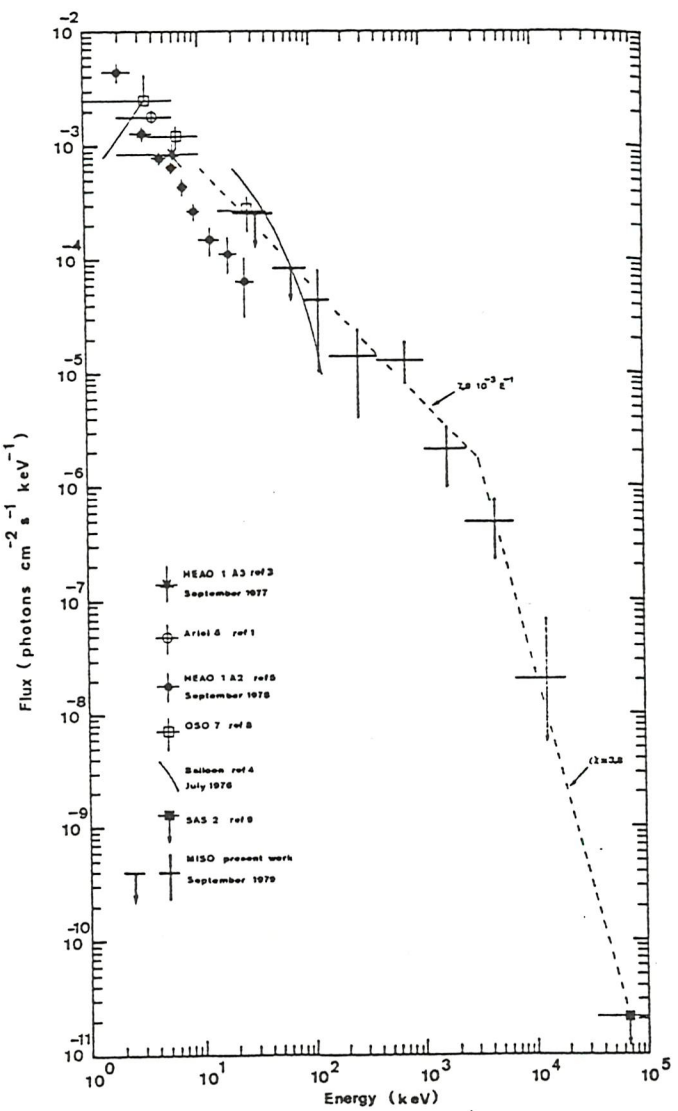


Fig. 1.18. *Photon spectrum of MCG08-11-11 from a number of observations, showing a break at $\sim$ 3 MeV (from Perotti et al. [109]).*

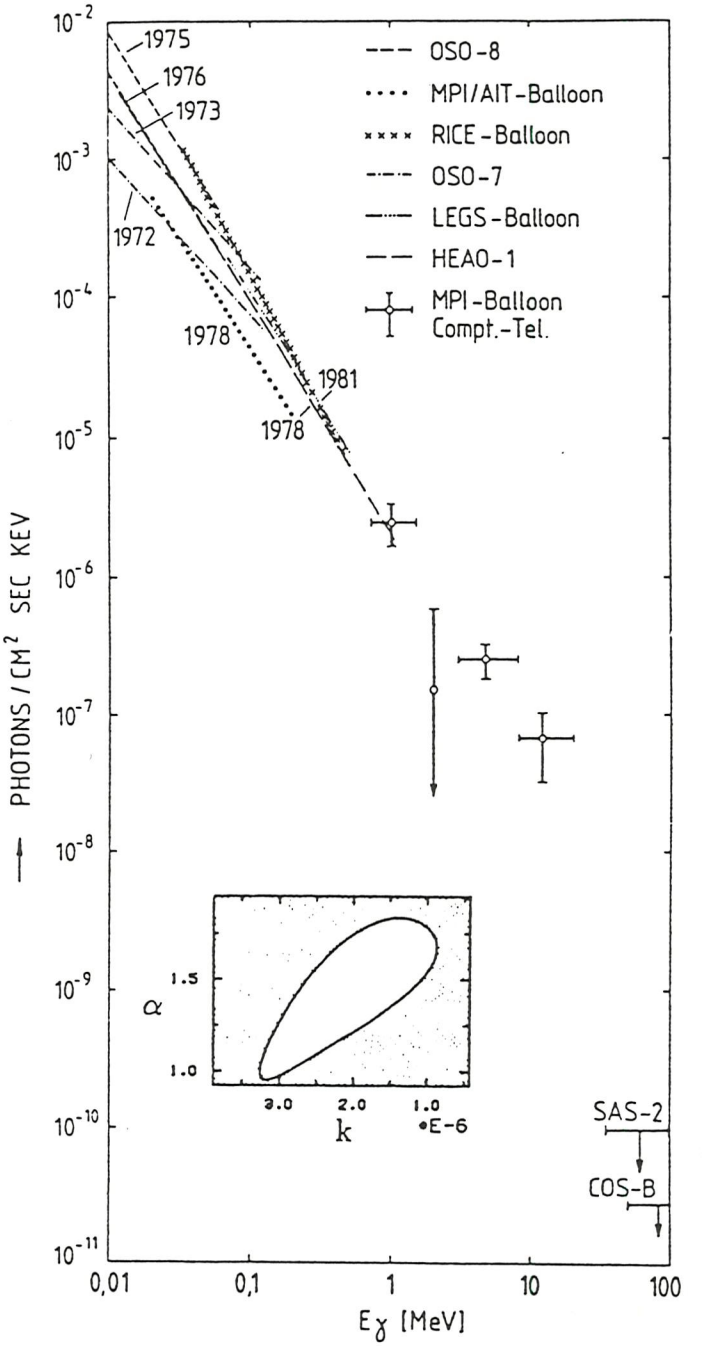


Fig. 1.19. *Photon spectrum of Centaurus A at X and gamma ray energies (from von Ballmoos et al. [153]).*

upper limits from SAS-2 and COS-B above 35 MeV and 50 MeV respectively were also valid during the Compton telescope flight then the spectrum must break at around 10 MeV. No line features were observed, thus giving a source radius lower limit of 2.5×10^{12} metres.

3C273

The quasar 3C273 has been studied and positively identified in all wavebands of the electromagnetic spectrum up to energies of 1 GeV. It is the brightest quasar as viewed from the earth and one of the closest with a redshift Z =0.158. Like several other active galaxies evidence for jets is apparent [52].

High energy gamma rays (50-500 MeV) from 3C273 were first reported in 1978 by Swanenburg [144] using COS-B data, and in the following year Primini published results of the first hard X ray detection in the range 13 - 180 keV using data from the A4 detectors on board HEAO-1 [114]. The X ray spectrum is well described, one observation example being given by Worrall et al. [163], see fig. 1.20. The soft X ray emission is well fitted by a power law photon index of 1.5 which shows a tendency to flatten to about 0.9 at photon energies of $\sim$9 keV. Generally the hard X ray emission fits well to a power law of slope $\sim$1.5, although the AIT/MPI balloon experiment showed an enhancement in the luminosity of about a factor of three greater than measurements made using the same instrument in 1978 [16]. In order for the hard X ray measurements to be consistent with the COS-B results which fit a power law of index 2.05, the maximum power output from the quasar must occur at photon energies close to 1 MeV.

3C273 has been shown to be variable at all wavelengths from radio to hard X rays on timescales ranging from less than one day to years. In particular the X ray emission has been observed to vary by a factor of 2 within a timescale of half a day [95] indicating that the X ray emission is derived from a compact zone ($\sim 10^{13}$ m), possibly an accreting massive black hole at the nucleus.

1.5.2 Normal Galaxies

In the near future gamma ray astronomy promises new means of examining normal galaxies compared to the numerous previous observations at longer wavelengths. For the more distant

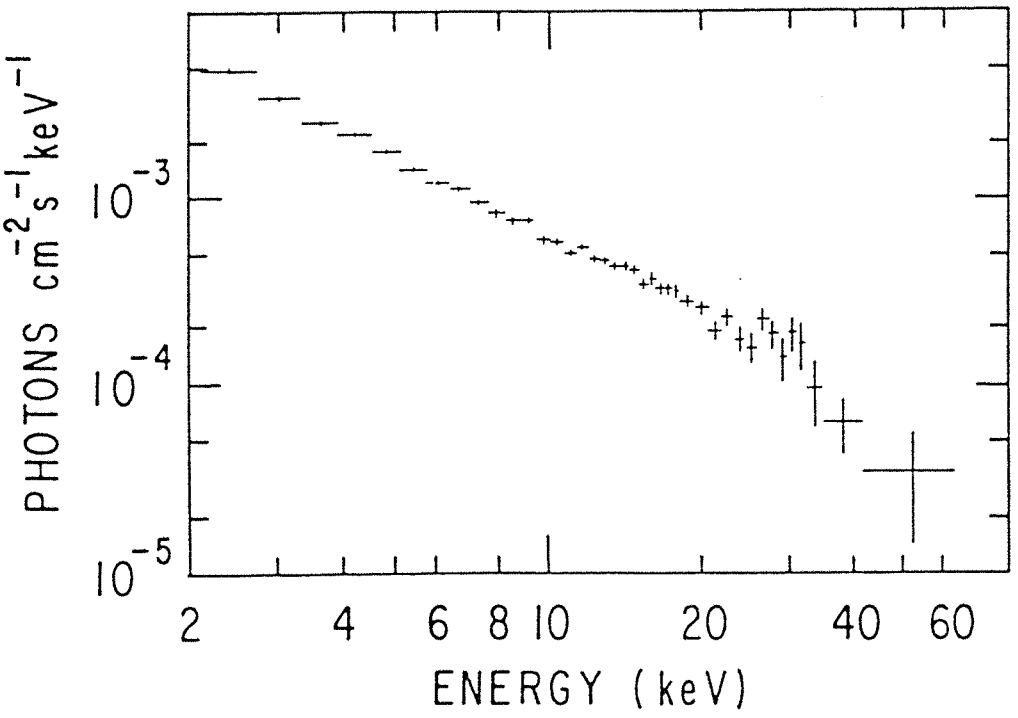


Fig. 1.20. *Spectrum of 3C273 as seen by the HEAO-1 X ray spectrometer (from Worrall et al. [163]).*

galaxies the best hope of useful study lies in searching for those hosting a supernova, since the rate of supernovae in our Galaxy is, on average, only two per century. The main priority however, is study of the local group (<1 Mpc), and in particular M31 and the Large and Small Magellanic Clouds (LMC, SMC). Although there are yet to be any definite results at high energies on the diffuse radiation from the closest galaxies it is hoped that the Gamma Ray Observatory [15], Gamma-1 [3] and GRASP will have the ability to detect the diffuse gamma radiation from the closest galaxies and indicate the emission distribution. Such measurements have the potential of revealing the presence of cosmic rays through their interactions as well as their density and distribution within the galaxy.

Attempts to observe the local group galaxies have been carried out by SAS-2 [48] and COS-B [112] but no emission from any galaxy was obtained and so only upper limits were available. These results are not surprising in the light of calculations by Ozel and Fichtel [107] who used theoretical cosmic ray distributions and magnetic field strengths to conclude that the expected fluxes (E >100MeV) were below the SAS-2 and COS-B upper limits, see fig. 1.21.

Stellar and interstellar processes should be occurring in M31 and in the Magellanic clouds producing diffuse gamma ray emission from both the 1.806 MeV ^{26}Al line as well as continuum processes resulting from the interactions of cosmic rays with the interstellar medium as in our own Galaxy. The ^{26}Al luminosity of our Galaxy is about 2×10^{37} erg s^{-1} from the centre [152], which implies a flux from the LMC of 2×10^{-5} photons cm^{-2} s^{-1}, which should be clearly visible by GRASP and therefore giving clues on stellar nucleosynthesis in this galaxy. Also of interest in the LMC is, of course, SN1987A discussed in section 1.4.7, as well as the black hole candidate LMC X-3 and a possible gamma ray burst supernova remnant N49.

1.5.3 Cosmic Diffuse Radiation

The existence of an intense cosmic diffuse background emission (CDB) extending from photon energies of a few keV to more than 100 MeV has been evident since the period of the earliest observations. A variety of theories has been postulated to explain the emission, ranging from those which attribute it to truly diffuse mechanisms of particle interactions [142], to those which consider the emission to be derived from a large number of unresolved point sources [17].

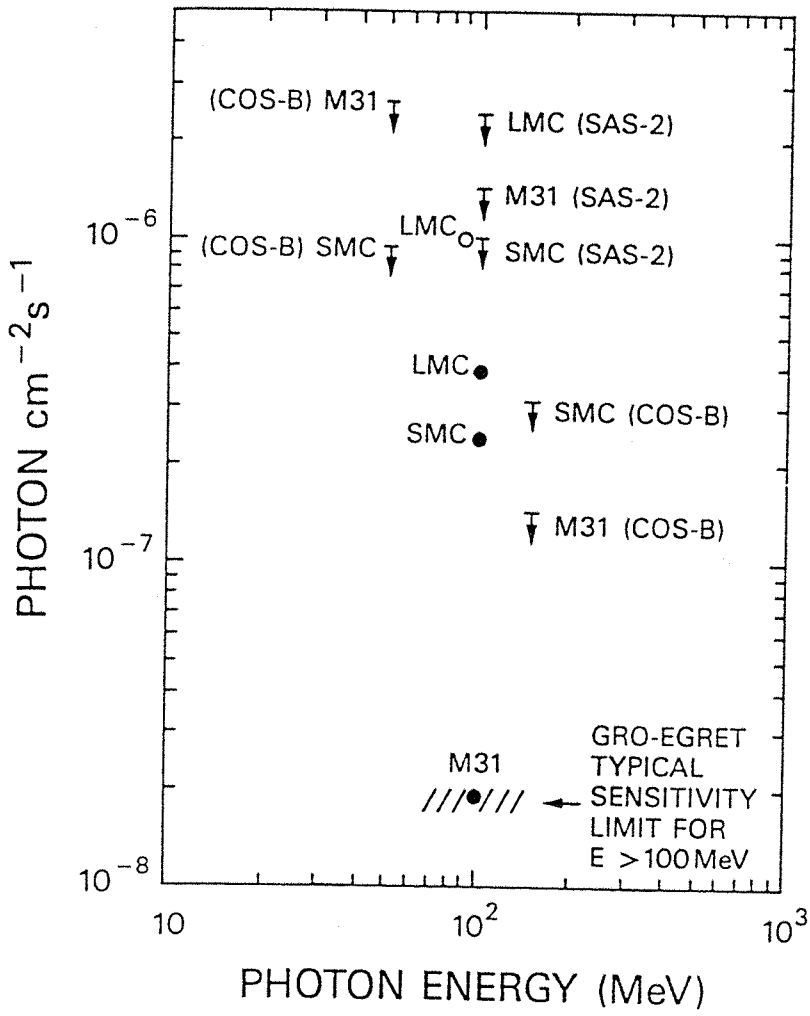


Fig. 1.21. *Upper limits from the SAS-2 and COS-B experiments for M31, the LMC and the SMC along with calculated fluxes (from Ozel and Fichtel [107]).*

The spectrum of the CDB is shown in fig. 1.22 and it is apparent that some structure is present. So far all experimental measurements show a "bump" in the spectrum at about 1 MeV [132,148]. This bump in the CDB spectrum is reminiscent of the MeV excesses shown by all AGNs, as is evident in fig. 1.15. Therefore the preferred model for the origin of the CDB is contribution from AGNs above 50 keV. This theory is given further weight by the discovery of a break in the CDB spectrum at MeV energies by a number of experiments. The scale of the AGN contribution is still unknown and more gamma ray observations of AGNs are required in order to answer this question.

One possibility of determining the absolute flux of the CDB by future gamma ray astronomical telescopes is to use the moon to occult parts of the field of view. The intrinsic background experienced by all gamma ray instruments makes absolute CDB measurements very difficult. However since the moon is opaque to gamma radiation, and is itself not a significant gamma ray source, then an imaging instrument such as GRASP will show a "negative source" or hole in the background when the moon is viewed, enabling the CDB to be accurately measured. The sun may also be used for the same purpose as it too is not a significant gamma ray source during flare minimum.

1.5.4 Clusters of Galaxies

A number of clusters of galaxies have been observed by hard X ray telescopes and a pattern has emerged in which the spectral emission up to about 100 keV may be described by a two component model, i.e. a thermal emission with kT ~few keV, and a power law extending to higher energies. A good example of this is the Virgo cluster observed by the HEAO-1 gamma ray detectors [81], the spectrum of which is shown in fig. 1.23 from 0.2-100 keV. The excess over the thermal emission above photon energies of about 8 keV is clear. Below 8 keV the best fit thermal bremsstrahlung model is for kT =2.05 keV (dashed line), and above 8 keV the spectrum is of a power law type, of index α =2.4 (dashed-dot line). It is thought that the hard X ray excess is due to one or more AGNs associated with the cluster. Strong and Bignami reported a similar gamma ray excess in the COS-B data of the Perseus cluster, and interpreted this component as possibly synchrotron self Compton emission from the nucleus of an AGN in the cluster [143].

Observations of radio haloes in some clusters suggest the presence of intergalactic high energy electrons and so some of the hard X ray flux is expected to arise from the inverse

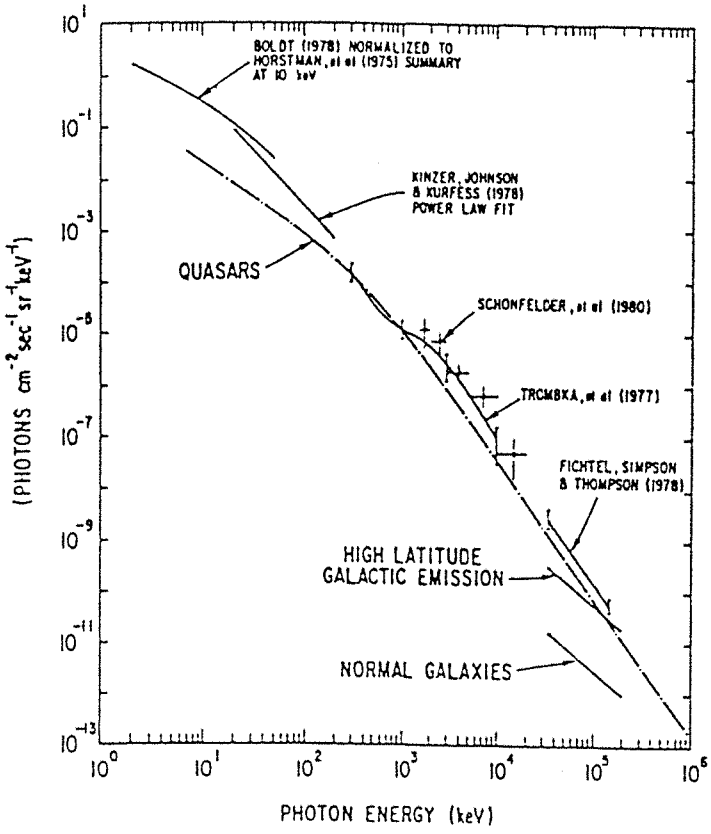


Fig. 1.22. *The diffuse gamma ray background spectra estimated for quasars and normal galaxies compared to some recent experimental data (from Fichtel and Trombka [49]).*

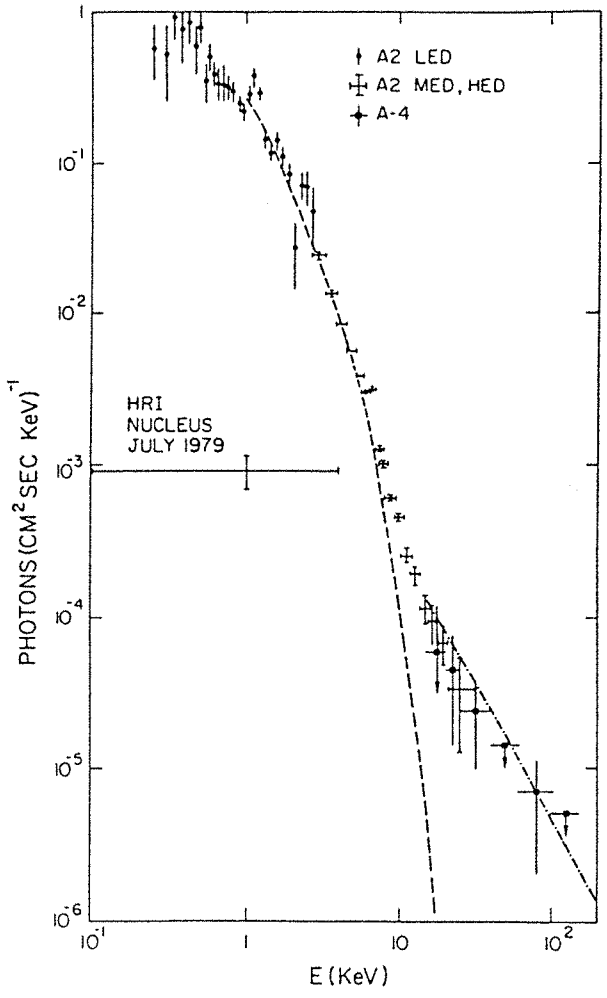


Fig. 1.23. *The X ray spectrum of the Virgo cluster of galaxies (from Lea et al. [81]).*

Compton process. Therefore, future measurements of this flux will set a lower limit to the intergalactic magnetic field strength inside a cluster.

Chapter 2

Gamma Ray Astronomical Telescopes.

2.1 Introduction

The earth's atmosphere is a strong attenuator of high energy photons, and so any photons from celestial gamma ray sources will not reach the earth's surface. Gamma ray telescopes, therefore, must be operated above the earth's atmosphere. The very low photon fluxes detectable from celestial gamma ray sources compared to the high background dictates that gamma ray astronomy must be performed in an environment of extremely low signal to noise. Therefore shielding from unwanted background and charged particle events is essential. To complicate matters further, the penetrating power of gamma rays rules out image formation by conventional imaging techniques such as with mirrors and lenses, necessitating that new techniques have to be developed. These problems have proven to be formidable and consequently astronomy using gamma rays has lagged behind other wavelengths.

The first high energy photon telescope was an X ray detector carried on a sounding rocket. However the use of sounding rockets for astronomy is limited due to the severe weight restrictions imposed by this vehicle and the short observation time available (of the order of a few minutes). Balloons and satellites offer much better prospects for gamma ray astronomy. Stratospheric balloons are capable of lifting payloads of 2000 kg up to 40 km altitude (where

X ray transmission is approximately 78% at 50 keV) and so payload weight is not a limiting problem. Observation times however, are still short, typically 24 hours. The best chance of performing high sensitivity X and gamma ray astronomy is undoubtedly by the use of satellites, as they do not suffer from atmospheric absorption and serious observation time restrictions. With the advent more powerful vehicles such as the space shuttle, larger systems can be placed into orbit thus providing the possibility to perform astronomy with more sophisticated and sensitive instrumentation.

2.2 Gamma Ray Telescopes, Past and Present

2.2.1 Non Position Sensitive Instruments

Most of the non position sensitive gamma ray telescopes that have been built are based on one basic design. The gamma ray detector itself is a single crystal scintillator which is surrounded by an anticoincidence shield in order to enable the rejection of events produced by charged particles. A typical example is given by Peterson [110], who describes an instrument which was used to measure the gamma ray emission from the earth's upper atmosphere in the photon energy range 200 keV - 10 MeV. The gamma ray detector was a cylinder of sodium iodide doped with thallium (NaI(Tl)), of 7.6 cm diameter by 7.6 cm length, mounted vertically with a 3 inch photomultiplier tube (PMT) on the bottom face (fig. 2.1). Apart from this face the NaI crystal was surrounded by a 2 cm thick plastic anticoincidence scintillator. The plastic was transparent to photons in the energy range of interest, and so the NaI detected photons from all directions. The instrument made five balloon flights between 1964 and 1966 with observation times ranging from approximately 6 to 14 hours, collecting gamma ray flux data from the earth's atmosphere through a range of altitudes.

A similar instrument, the Apollo Gamma Ray Spectrometer (AGRS), was flown on the Apollo 15 and 16 spacecrafts in 1971 and 1972 respectively [65]. The AGRS was successfully used to observe both the lunar surface composition from the gamma ray emission of the upper 30 cm of material, and cosmic gamma radiation. The detector was again a cylinder of NaI(Tl) of 7 cm diameter by 7 cm length optically coupled to a PMT, and the remaining sides surrounded by a passive stainless steel casing (fig. 2.2). The steel casing was surrounded by a 0.8 cm thick plastic anticoincidence shield which, like Peterson's atmospheric experiment, was transparent to gamma rays, thus enabling the NaI to collect photons from all directions.

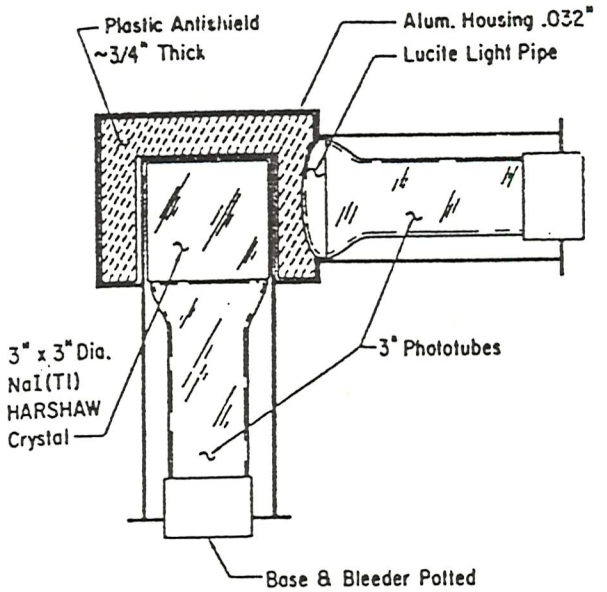


Fig. 2.1. *Peterson's atmospheric gamma ray experiment. A 7.6 × 7.6 cm NaI(Tl) shielded scintillation counter [110].*

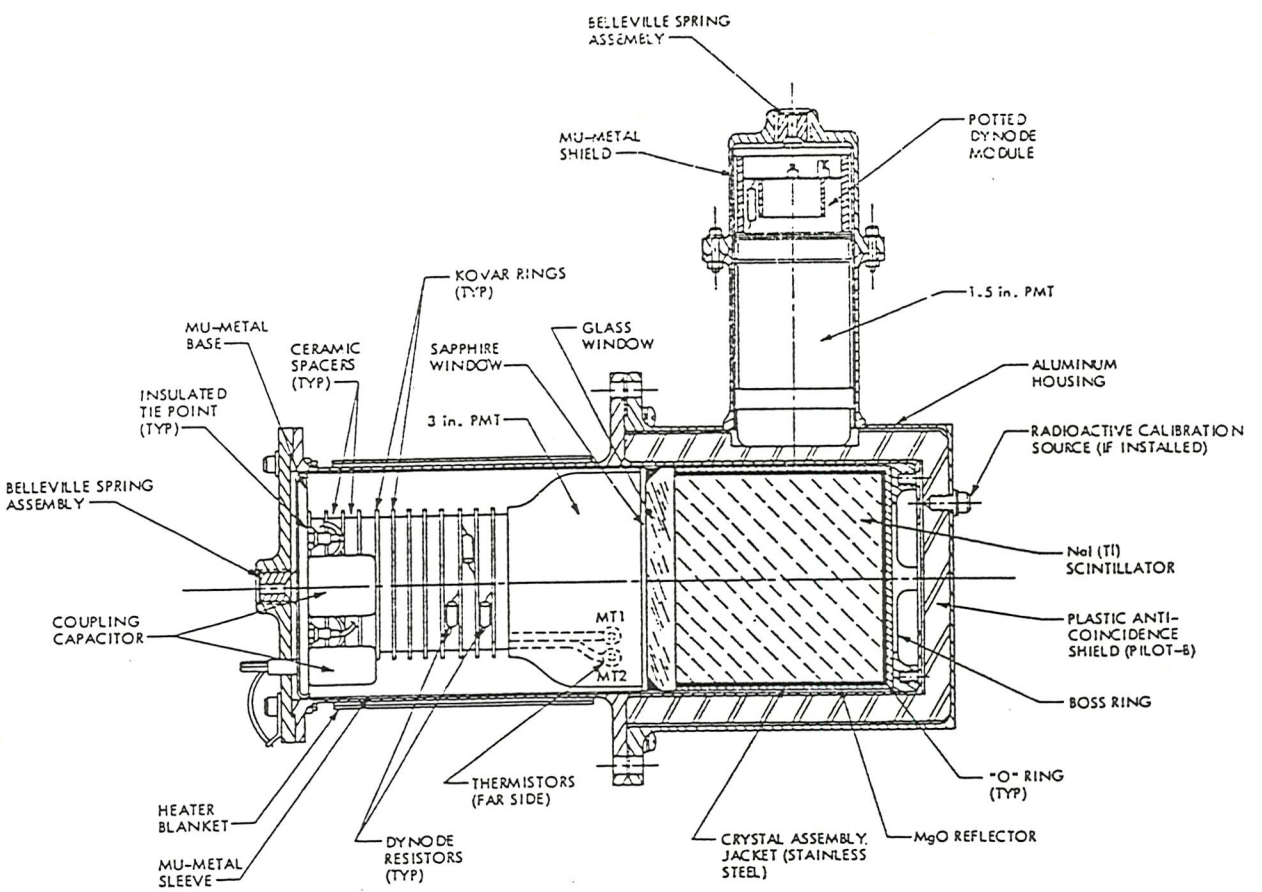


Fig. 2.2. *Schematic view of the Apollo Gamma Ray Spectrometer (AGRS) detector (from Harrington et al. [65]).*

The AGRS mapped 22% of the lunar surface and obtained cosmic gamma ray flux data over an energy range of 65 keV to 27.5 MeV. The results from the AGRS experiment [147] have been used to evaluate the background counting rate on the Southampton University gamma ray telescope Zebra [164] due to the cosmic diffuse background.

2.2.2 Compton Telescopes

In order to conduct any really useful gamma ray astronomy, telescopes capable of resolving sources to within at least half a degree are essential. One method of gamma ray astronomical imaging is by the use of Compton telescopes, which all use the same principle as explained by Schönfelder et al. [130]. A schematic drawing of the Compton telescope is shown in fig. 2.3. An incoming photon undergoes a Compton collision inside the upper plastic scintillator block, depositing energy in the form of visible light, the pulse height of which is measured by a PMT. The scattered photon undergoes a second collision in the lower scintillator, which is at a known fixed distance below the upper scintillator, and the energy deposited is again measured. The interaction sequence is measured, and only events which satisfy the time of flight requirements from upper to lower detectors are accepted. This enables rejection of a large proportion of the background events produced by neutrons since their velocity is much lower than the velocity of light, whilst charged particle events are rejected by an anticoincidence scintillator shield placed between the detectors.

Let the incoming photon of energy E_γ arrive at an angle θ to the zenith, and deposit energy E'_e in the upper scintillator. If the scattered photon has an energy E'_γ, a scatter angle of ϕ, and deposits energy E''_e in the lower scintillator, then the following equations hold:

$$E_\gamma = E'_e + E'_\gamma \tag{2.1}$$

$$\cos\phi = 1 - \frac{m_e c^2}{E'_\gamma} + \frac{m_e c^2}{E_\gamma} \tag{2.2}$$

where m_e is the electron rest mass. The energy E'_e is determined by the pulse height of the upper detector. The scattered photon energy E'_γ can be only approximately determined by the energy deposit E''_e in the lower detector, which in turn implies an uncertainty in the scatter angle ϕ. As a first approximation assume that all of the scattered photon energy is deposited in the lower detector, and so $E'_\gamma = E''_e$. Using this and 2.1 gives a modification of 2.2 where the scatter angle becomes $\hat{\phi}$, an estimate of ϕ:

$$\cos\hat{\phi} = 1 - \frac{m_e c^2}{E''_e} + \frac{m_e c^2}{E''_e + E'_e} \tag{2.3}$$

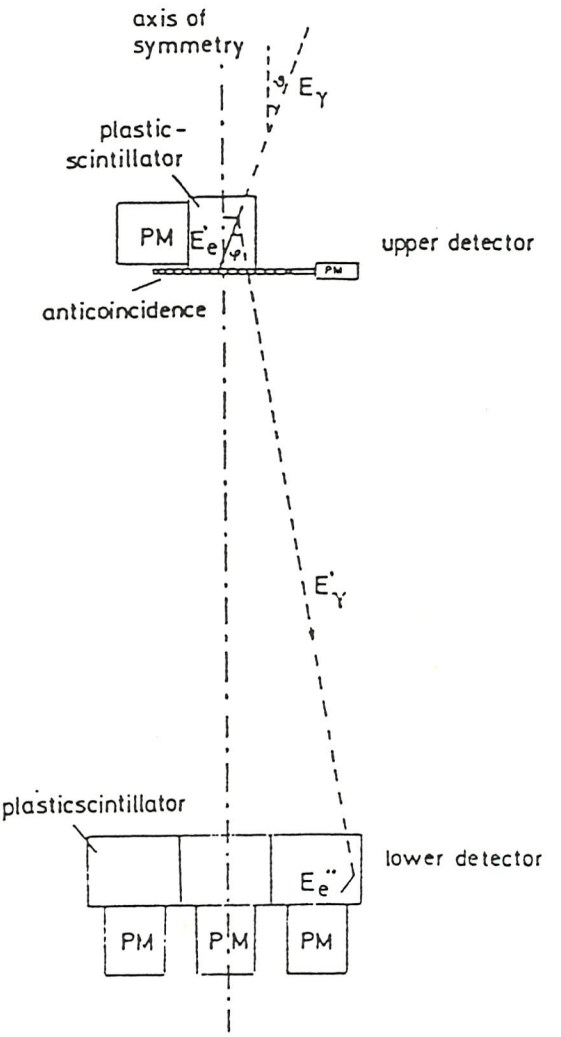


Fig. 2.3. *Illustration of the Compton telescope (from Schönfelder et al. [130]).*

If both of the detectors are subdivided into arrays of discrete pixels, each pixel with its own PMT, then the track of the scattered photon can be easily traced from upper to lower detector and so 2.3 can be immediately translated into an expression for $\hat{\theta}$ to give an approximation to the source position angle.

The method of data interpretation and analysis is given in [133] for the Max Planck Institut (MPI) balloon borne Compton telescope. In this instrument the upper detector was composed of 16 cells of low Z liquid scintillator so that Compton scattering would readily take place. 1.2 metres below this there was an array of 32 blocks of high Z (NaI(Tl)) scintillator, chosen so that the scattered gamma ray would be totally absorbed. Both detectors were completely surrounded by 8 mm thick plastic anticoincidence shields. A balloon flight has yielded an image of the Galactic anticentre from 1-10 MeV energy range (fig. 2.4) [133]. The MPI telescope forms the basis of the COMPTEL Compton telescope on board the Gamma Ray Observatory mission, see section 2.3.1.

Lockwood [88] describes another Compton telescope which used scintillators as the upper detector and a NaI crystal as the lower detector. It was launched on a balloon in 1978 to measure the atmospheric and diffuse cosmic gamma ray flux in the energy range 400 keV - 7 MeV. Other Compton telescopes are described elsewhere in the literature [69,168].

2.2.3 Coded Aperture Instruments

The idea of using coded apertures to carry out imaging of celestial X and gamma ray sources has led to great technical advances in telescope design and thus rapidly increasing interest in high energy astrophysics. The first astronomical high energy coded aperture imaging experiment of any real note was a solar observation in 1972 by Blake [22]. The aperture used was a square 61 $\times$ 61 element random array with the transparent element positions iteratively improved to minimise the system point spread function side lobes, and it was placed in front of a photographic film detector. The instrument was launched on a rocket and it made one observation of the sun, which after image reconstruction, showed structure to within one arcminute resolution, which agreed well with the general shape and intensity of the calcium K-line region from Hale Observatories.

More recent coded aperture telescopes have been more sophisticated, using uniformly redundant arrays (URA) rather than random arrays, so that no inherent noise is present in the

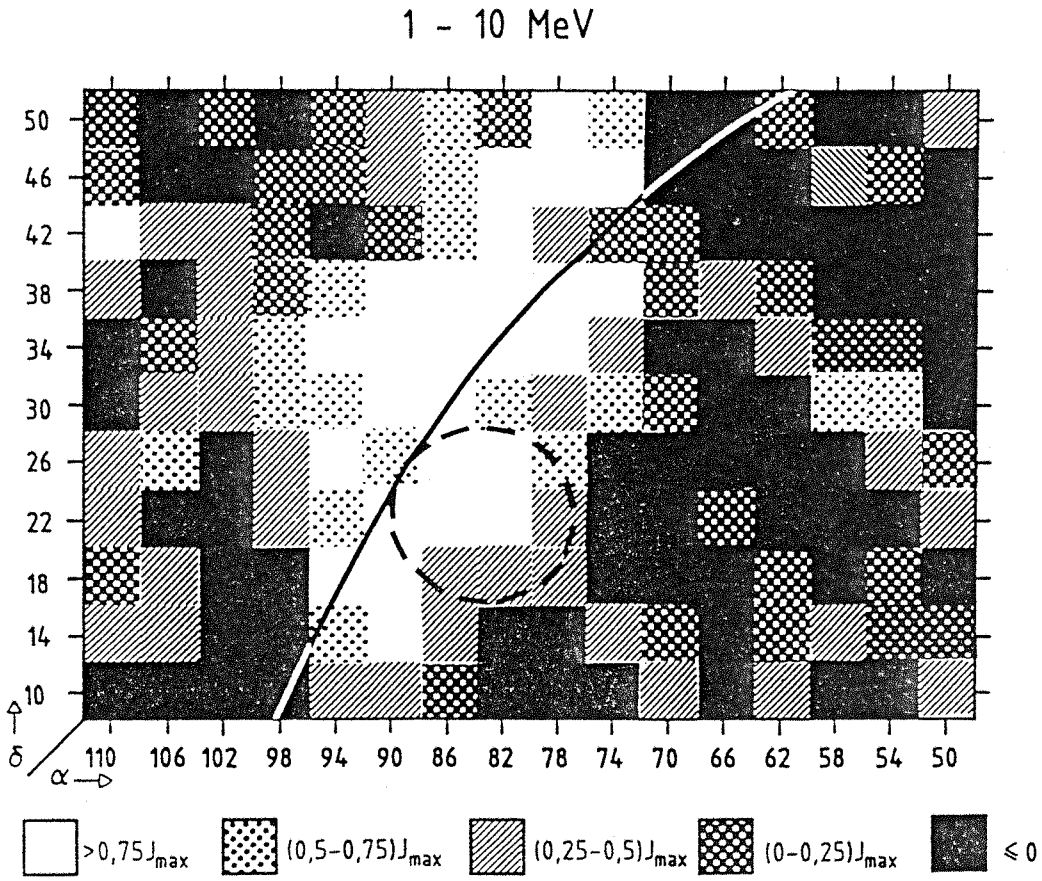


Fig. 2.4. *Image of the Galactic anticentre region using the MPI Compton Telescope from 1-10 MeV (from Schönfelder et al. [133]).*

deconvolution. The Directional Gamma Ray Telescope (DGT) designed at the University of New Hampshire used such an array. The DGT is shown schematically in fig. 2.5 [42]. The aperture used was a conventional 2 × 2 mosaic of a 5 × 7 twin prime URA, each element being 5.6 × 5.6 cm and thickness 1.9 cm, positioned 84 cm above the detector plane, giving an angular resolution of 3.8° and a fully coded field of view of 15° × 33°. The detector plane consisted of 35 BGO detectors, each being cylinders of 5.1 cm diameter and 2.0 cm long, with the whole plane being actively shielded from the sides and from below by 10 cm of NaI(Tl). The main results obtained from the DGT are of the Crab nebula, as shown in fig. 2.6, and the Cygnus region [99,106], and the Seyfert Galaxy NGC4151 [106].

The Birmingham University coded aperture X ray telescope (XRT) was flown on the Spacelab-2 Shuttle mission in 1985 lasting 8 days. The instrument itself [160] consisted of a pair of telescopes (fig. 2.7) both based on Hadamard URA patterns. The two apertures were placed approximately 3 metres from the detectors, one based on a 31 × 33 pseudo noise (PN) sequence giving a 12 arcminute angular resolution, and the other based on a 127 × 129 twin prime array giving a 3 arcminute angular resolution. The detectors used were gas proportional counters, able to detect an X ray photon event to within 0.8 mm position resolution. The fully coded field of view was 6.4° × 6.4° and the operational energy range was 2.5 - 30 keV. During the mission, observations were successfully made of some clusters of galaxies, the Vela supernova remnant and the Galactic centre region. To date most of the results are preliminary with observations of, for example, the Perseus cluster of galaxies (fig. 2.8), and the Galactic centre [139]. However, a more detailed analysis of the Galactic centre data has been presented by Skinner et al. [138], identifying at least three previously unreported sources as well as some known ones, see fig. 2.9.

The first gamma ray telescope to make use of a hexagonal uniformly redundant array (HURA) coded aperture was the Gamma Ray Imaging Payload (GRIP) designed at the California Institute of Technology. The detector [36] was a 41 cm diameter NaI Anger camera viewed by 19 PMTs shielded by a plastic scintillator approximately 16 cm thick and having an energy range of 30 keV to 5 MeV. It was situated 2.5 metres below a HURA coded aperture of basic pattern 127 elements [4]. The aperture elements were made of lead, and were hexagonal of dimension 2.5 cm flat to flat and 2 cm thick giving a geometric angular resolution of 0.6°. The aperture was able to rotate with respect to the detector, enabling source ambiguities to be removed, and thus allowing extension of the field of view up to 20° by repeating the aperture pattern to give that shown in fig. 2.10. The true source is

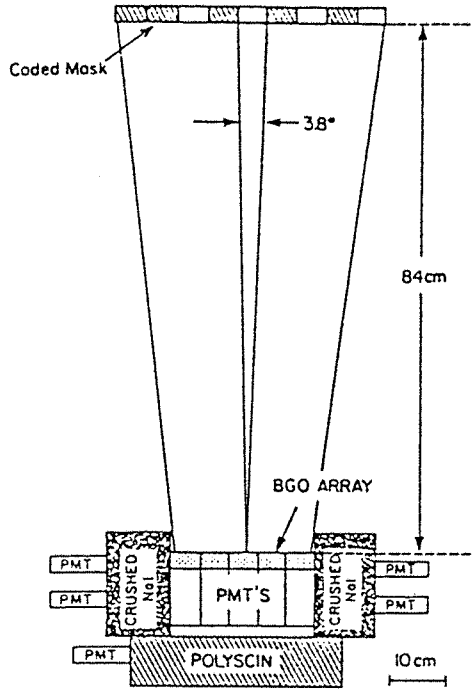


Fig. 2.5. *The New Hampshire Directional Gamma Ray Telescope (DGT, from Dunphy et al. [42]).*

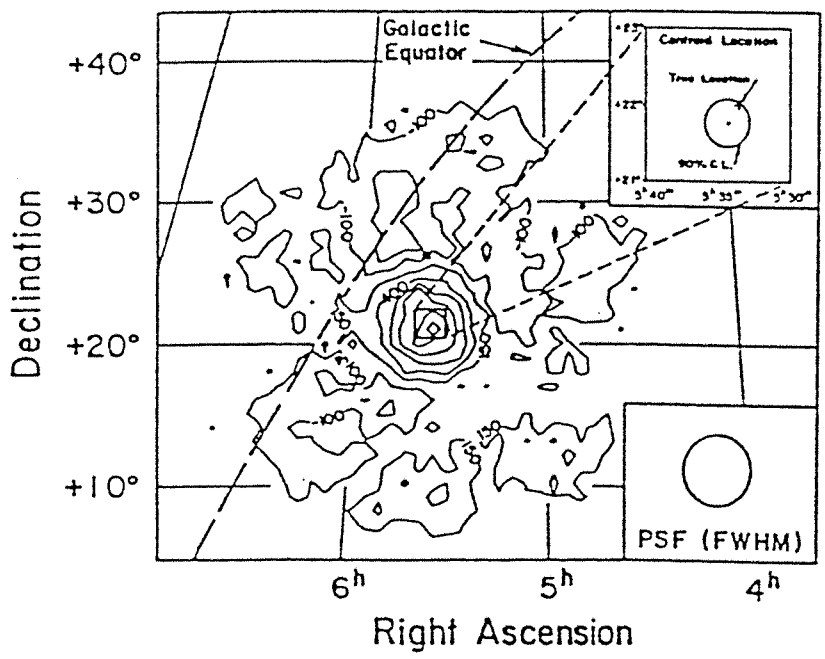


Fig. 2.6. *Composite image of the Crab region integrated over the energy range 200-600 keV (from McConnell et al. [99]).*

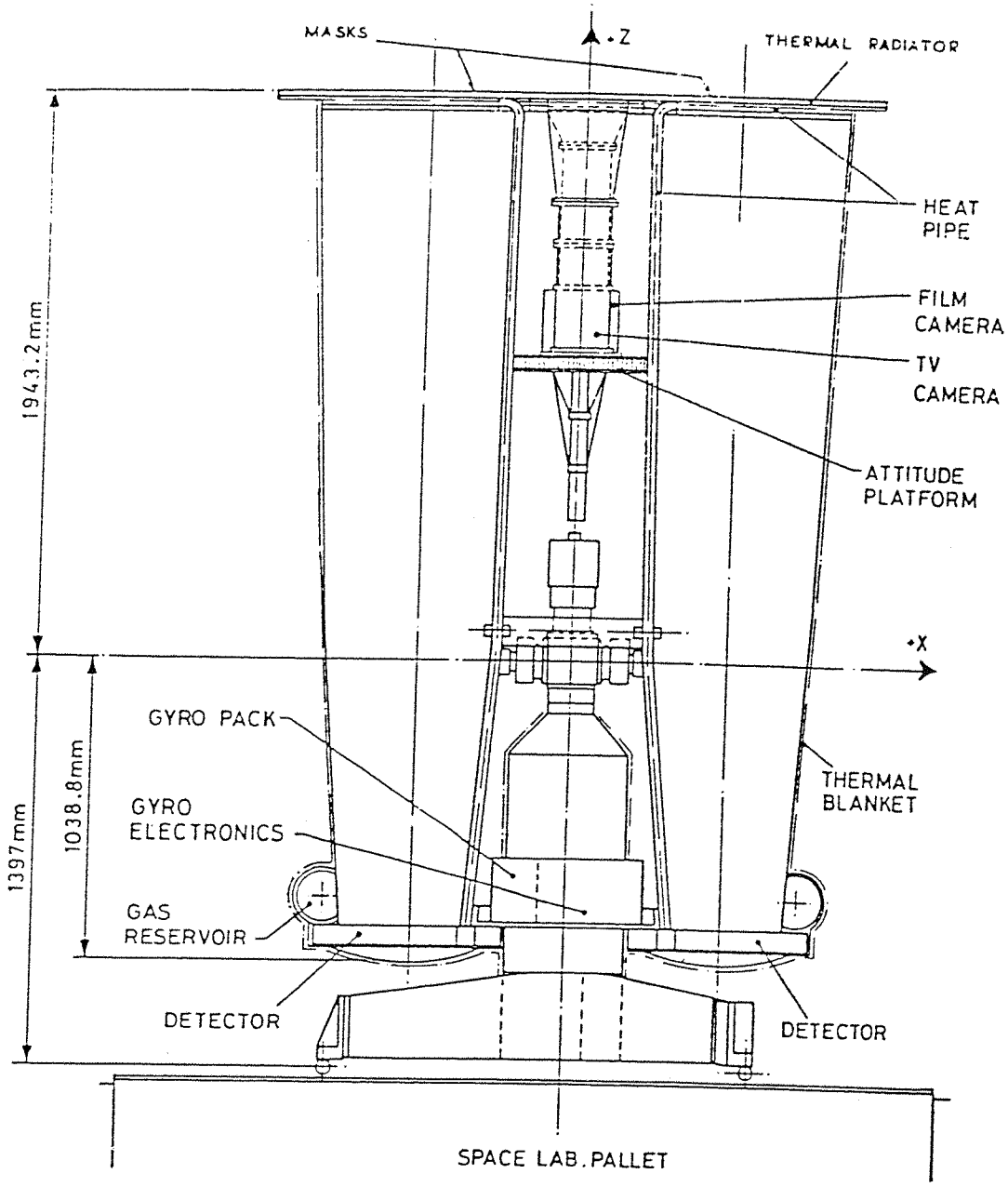


Fig. 2.7. *Schematic drawing of the Birmingham University X Ray Telescope (XRT, from Eyles et al. [43]).*

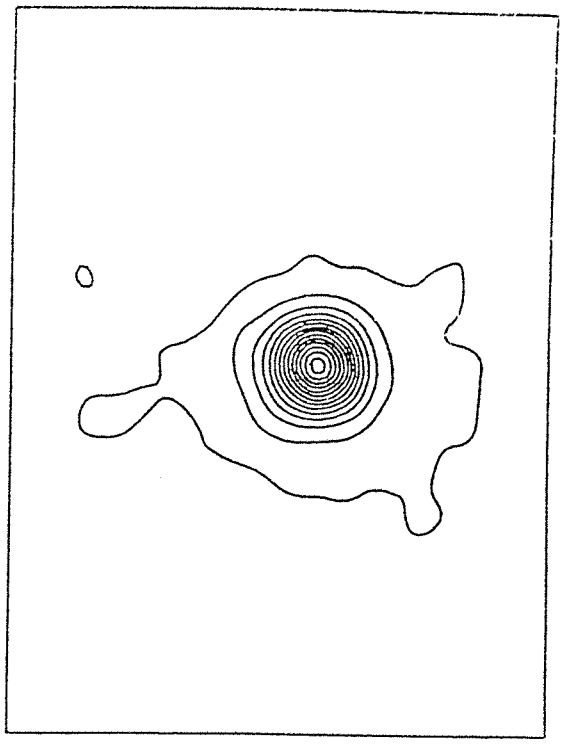


Fig. 2.8. *The Perseus cluster of galaxies in the range 2.5-25 keV (from Skinner et al. [139]).*

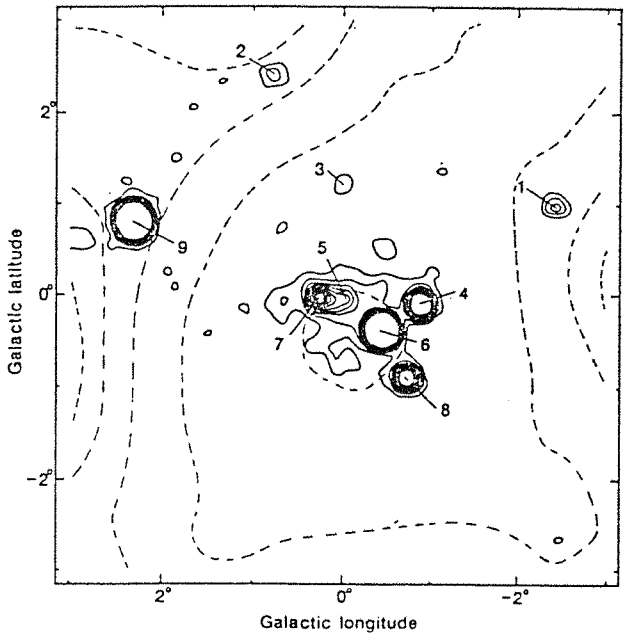


Fig. 2.9. *Contour plot of the Galactic centre region in the range 2.5-25 keV (from Skinner et al. [138]).*

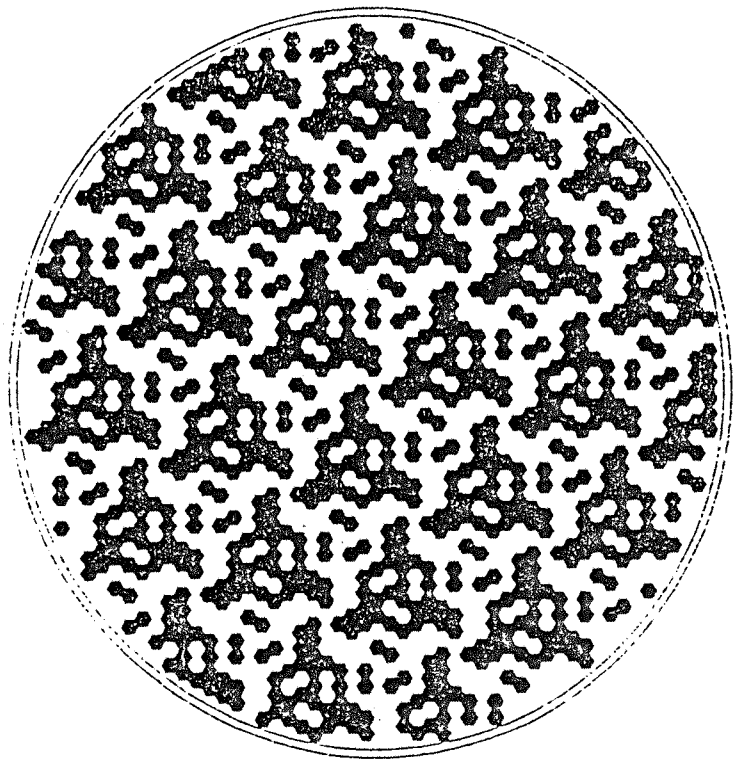


Fig. 2.10. *Coded aperture used on the GRIP telescope (from Althouse et al. [4]).*

easily seen by aperture rotation, as shown in fig. 2.11. The original image gives a series of periodic peaks (a), but aperture rotation blurs the ambiguous peaks into a series of low rings (b), which can be removed to give a clean deconvolved image (c). The first balloon flight of GRIP was in 1986 for 28 hours [5], in which time both the Crab and Cygnus regions were observed. A particularly interesting flight of GRIP took place in 1987 to observe the supernova SN1987A. Spectral and imaging observations of the supernova were made and results were reported by Cook et al. [37]. Fig. 2.12 shows an image of the SN1987A region along with a large cross indicating its expected position. Spectral data gives evidence for possible line emission due to ^{56}Co decay at 847 keV and 1.238 MeV (fig. 2.13).

The Zebra telescope is a Southampton University gamma ray astronomy project in collaboration with the Milano Istituto di Fisica Cosmica, the Bologna Istituto TESRE, and the Frascati Istituto Astrofisica Spaziale. It is a balloon borne instrument working in the energy range 0.2 - 10 MeV, having a fully coded field of view of 7° × 9° and angular resolution of approximately 1° [151]. In fig. 2.14 a schematic view of the Zebra arrangement for the first flight is shown. The aperture is a standard 2 × 2 repetition of a 9 × 7 URA generated from a PN sequence, with one row and one column removed to avoid ambiguities. The opaque elements are made from tungsten and have dimension 6.7 × 7.0 cm and are 2.0 cm thick. The basic 9 × 7 pattern has dimension 60.3 × 49.0 cm, the same as that of the useful area of the detector. 3.5 metres from the aperture lies the detection plane consisting of an array of nine position sensitive NaI(Tl) scintillation bars, each of dimension 55 × 5.8 × 5 cm thick, and each optically isolated from its neighbours [23]. The surface of each bar has been treated so that scintillation light produced by a photon interaction is attenuated exponentially along the length of the bar. This allows the reconstruction of both the photon interaction positions and their energy by measuring attenuated light with PMTs at each end of the bar [30]. The position sensitive bars are actively shielded below and at the side by 13 similar, but non position sensitive, bars and from above by a 6 mm thick plastic scintillator, as shown in fig. 2.15. The position resolution along the bars has typically a 1σ value of 10 mm at 1 MeV.

2.3 Future Gamma Ray Astronomy Projects

To date, gamma ray astronomical observations have been mainly carried out by balloon and limited satellite experiments. Nevertheless the results are very interesting, showing that

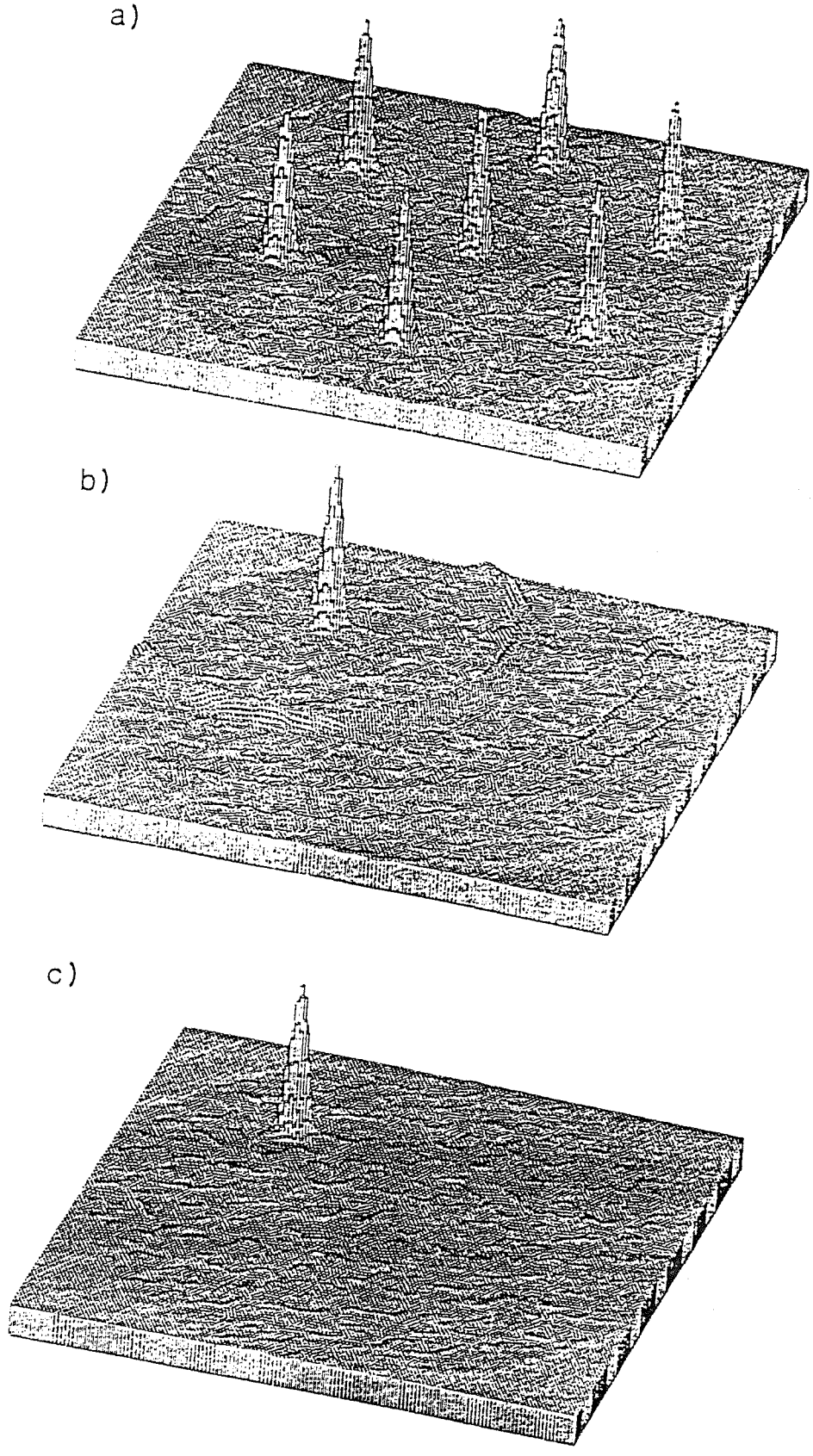


Fig. 2.11. *Images of a point source using a HURA: (a) stationary aperture, (b) rotating aperture, (c) rotating aperture image after ring removal (from Cook et al. [35]).*

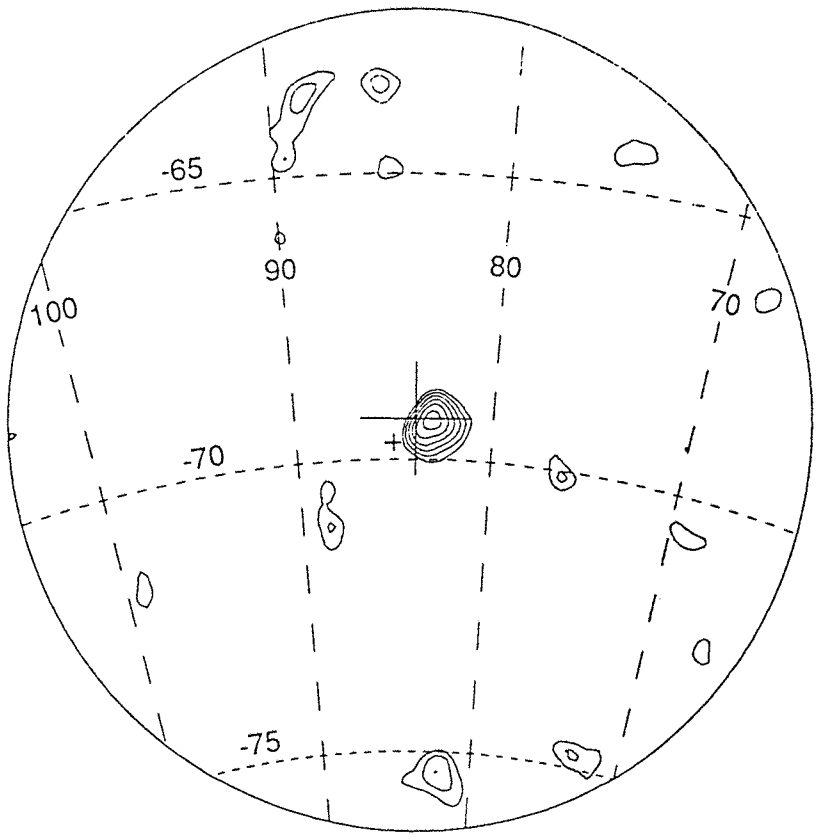


Fig. 2.12. *Images of the SN1987A region using GRIP. The large cross indicates the expected position of SN1987A (from Cook et al. [37]).*

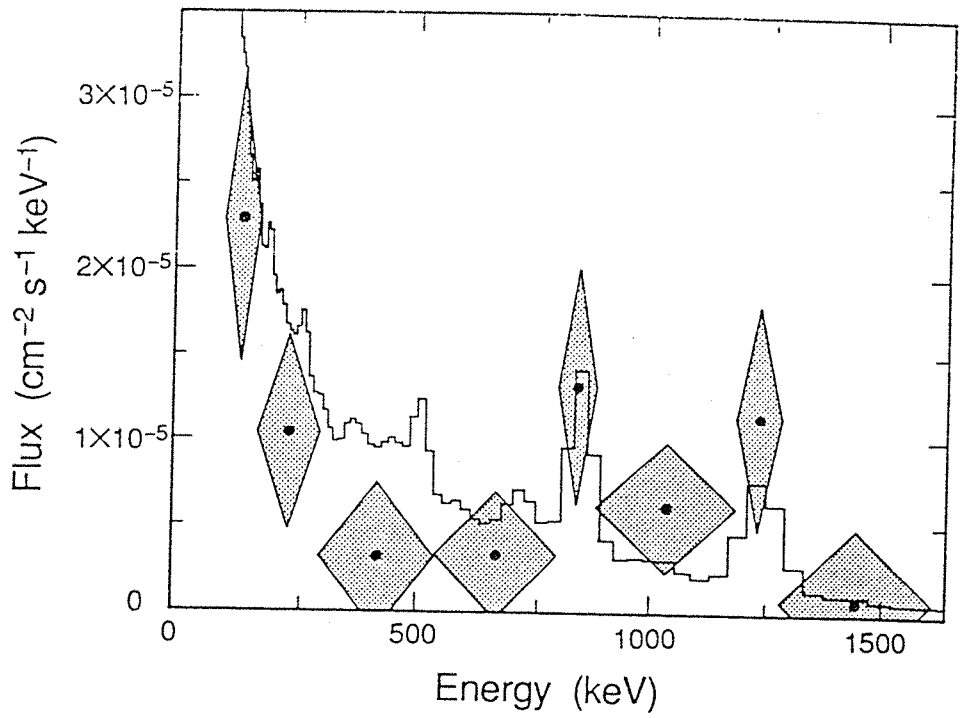


Fig. 2.13. *Energy spectrum of SN1987A measured by GRIP (from Cook et al. [37]).*

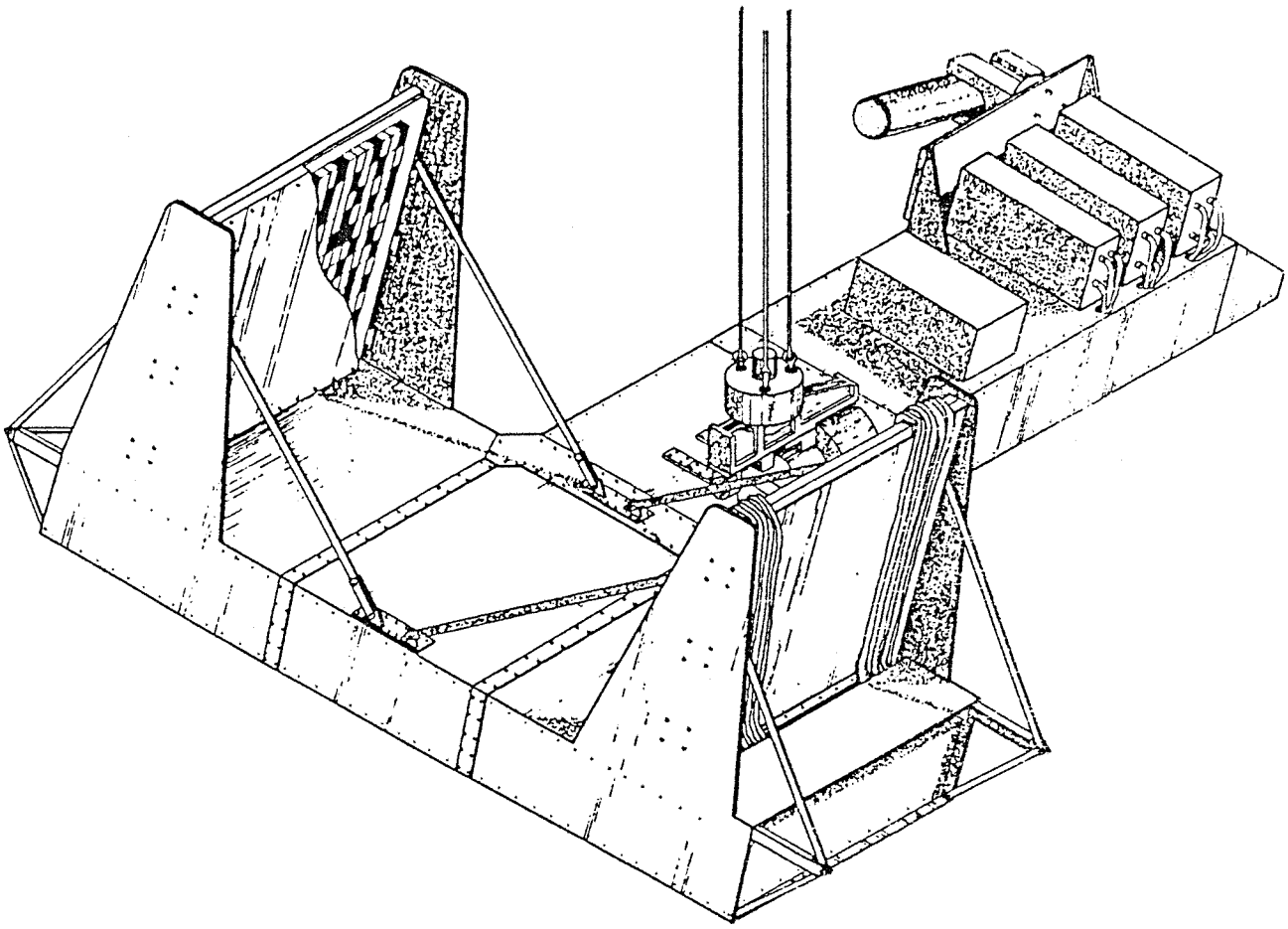


Fig. 2.14. *Schematic drawing of the Zebra telescope (from Villa et al. [151]).*

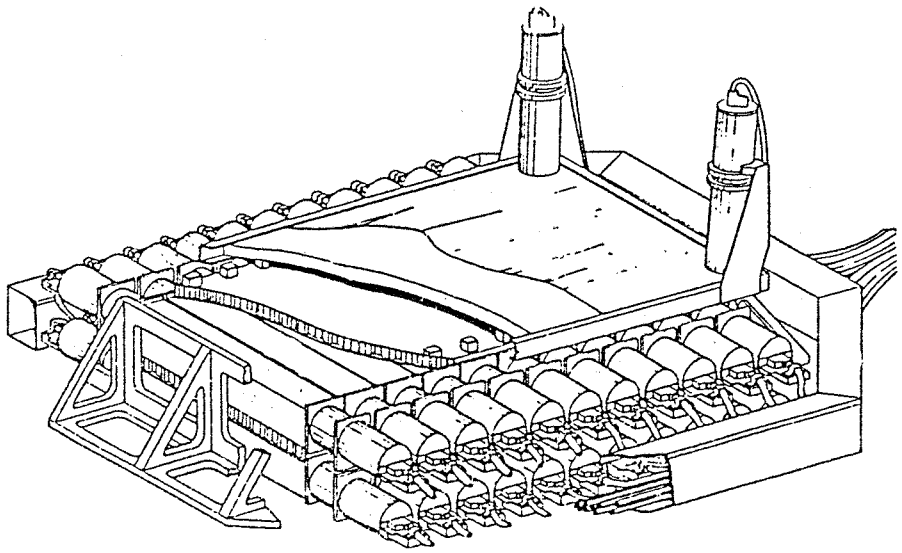


Fig. 2.15. *The Zebra detection plane (from Villa et al. [151]).*

a wide variety of galactic and extragalactic objects are gamma ray emitters. Much more will be known in the near future when new and more advanced gamma ray telescopes are launched. The main ones are now described.

2.3.1 The Gamma Ray Observatory

The Gamma Ray Observatory (GRO) is a NASA mission which is to be launched on the space shuttle. It consists of a support platform and four instruments covering the energy range 50 keV - 3 x 10^4 MeV [15]. The two year GRO mission is expected to increase the number of detected gamma ray sources by one or two orders of magnitude. The objectives of the mission include:

(i) A study of compact gamma ray sources such as neutron stars and black holes.

(ii) A search for evidence of the build up of heavy elements by nucleosynthesis.

(iii) To study the structural and dynamical features of the Galaxy.

(iv) To study other galaxies, such as radio galaxies, Seyferts and quasars.

(v) To measure the diffuse cosmic gamma ray background.

(vi) To observe gamma ray bursts.

The four experiments on board the GRO are now described.

The Oriented Scintillation Spectrometer Experiment (OSSE)

The OSSE consists of four identical phoswich detector systems, each one designed to operate in the energy range 50 keV - 10 MeV [78] (for a description and an explanation of phoswich detectors see Heath et al. [68]). Each detector has a 5° x 11° FWHM rectangular field of view provided by a conventional passive tungsten collimator (fig. 2.16). The primary element of each detector is a 13 inch diameter by 4 inch thick cylinder of NaI(Tl) scintillator, behind which is a CsI(Na) crystal of 3 inch thickness in the form of a phoswich assembly. The CsI(Na) scintillator is viewed by 7 PMTs, with the position of an event being measured by the relative photon fluxes from each tube. The whole detector system is actively shielded by a 2.75 inch thick CsI anticoincidence shield. The OSSE is designed to make comprehensive

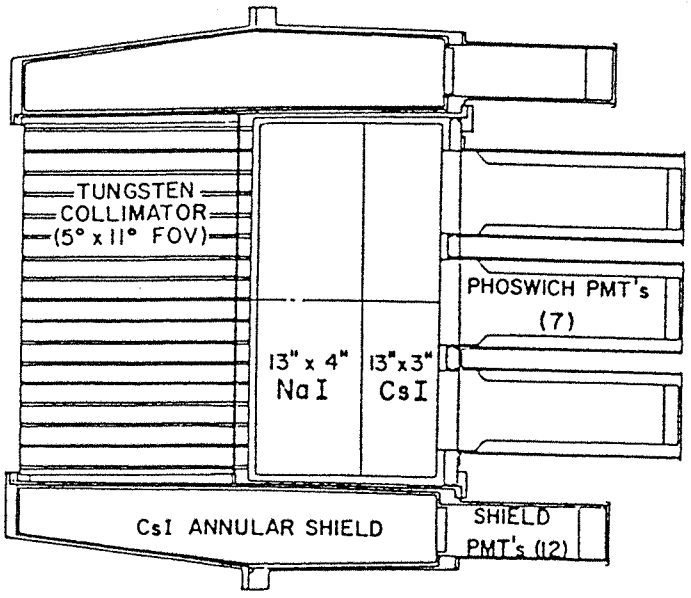


Fig. 2.16. *One OSSE detector (from Kurfess et al. [78]).*

observations of astrophysical sources from 50 keV - 10 MeV as well as gamma ray and neutron measurements from 10-150 MeV from solar flares.

The Imaging Compton Telescope (COMPTEL)

A sketch of COMPTEL is shown in fig. 2.17 with the principal dimensions indicated [134]. The instrument consists of two detector planes separated by a distance of 1.5 metres and surrounded by a system of anticoincidence shields. The observation energy range will be in the region of 1 - 30 MeV. The method of photon event selection is based on Compton kinematics and is explained in section 2.2.2, as well as by Schönfelder et al. [130]. The expected FWHM angular resolution is about 2° and the field of view will be approximately 1 steradian. The main objectives of COMPTEL are to study point sources, the diffuse Galactic emission and broadened line emission.

The Energetic Gamma Ray Experiment Telescope (EGRET)

EGRET is designed to study sources at the top end of the GRO energy range from 20 - 3×10^4 MeV. The position resolution is of the order of 5 - 10 arcminutes depending on the strength of the source [70]. The operating principle of the instrument is shown in fig. 2.18. An incoming gamma ray enters the top of the telescope and interacts in one of the 27 thin tantalum sheets interleaved between the closely spaced spark chamber modules producing an electron - positron pair. The tracks of the electrons are followed by a 3 dimensional reconstruction of the spark positions. Re-tracing of the trajectories gives the original incident direction and position of interaction of the photon. Further information on the progress of the electrons is measured by 6 more widely spaced spark chambers lower down. Charged particle rejection is provided by active shielding using an anticoincidence dome surrounding the spark chamber system. EGRET will study faint point sources, as well as some stronger ones, and the Galactic diffuse emission will also be observed.

Burst and Transient Source Experiment (BATSE)

BATSE is designed to monitor continuously the entire non occulted sky for a wide variety of gamma ray burst and other transient events. Burst event location is important in order

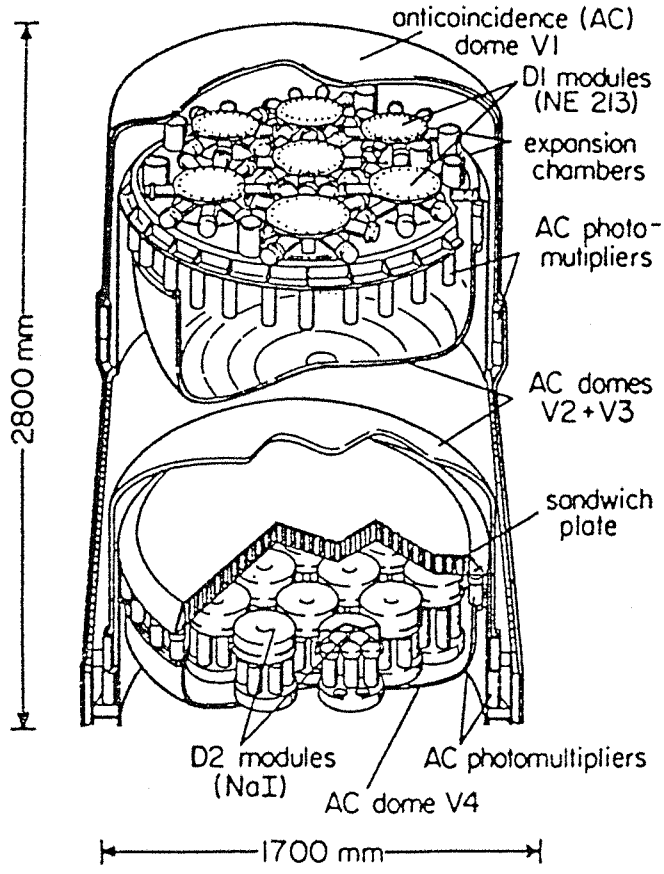


Fig. 2.17. *Schematic drawing of the COMPTEL instrument (from Schönfelder et al. [134]).*

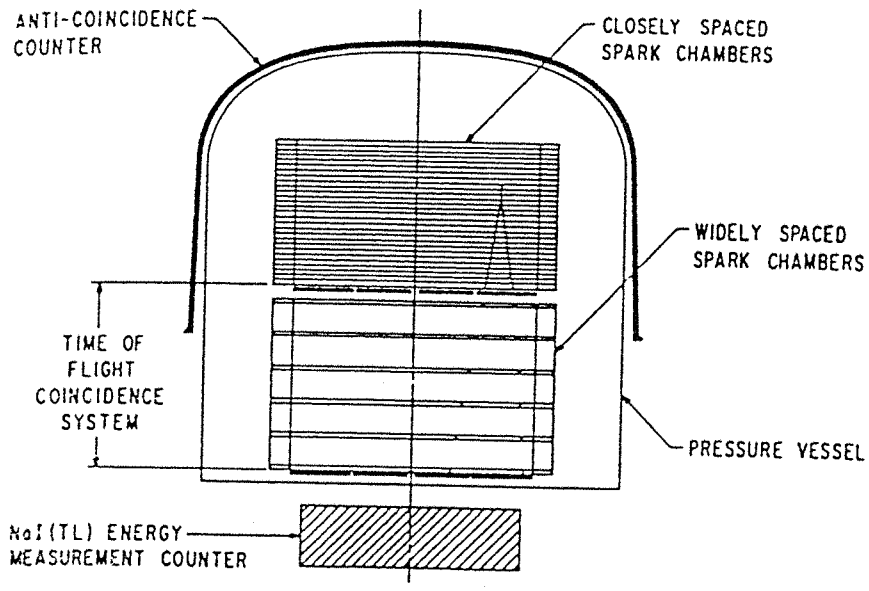


Fig. 2.18. *Schematic drawing of the EGRET telescope (from Hughes et al. [70]).*

that optical and radio counterparts can be found and studied. BATSE is expected to be a key element of a future interplanetary network of detectors each detecting gamma ray bursts and calculating the burst locations by triangulation methods. The telescope [51] is an all sky monitoring experiment operating in the energy range 20 keV - 10 MeV, and it consists of 8 uncollimated detector modules positioned around the GRO spacecraft. One such module is shown in fig. 2.19. A high detector sensitivity is provided for weak bursts by a 50.8 cm diameter × 1.3 cm thick flat scintillator crystal of NaI(Tl). A spectroscopy scintillation detector is also included in each detector module, optimised to give better spectral resolution of bursts over a wider energy range than the flat detector.

2.3.2 Gamma-1

Gamma-1 is a co-operative gamma ray project between numerous Soviet and French organisations, and the instrument is expected to be flown on a SOYUZ spacecraft in the late 1980s observing sources between 50 and 5,000 MeV. Fig. 2.20 schematically illustrates the design of the Gamma-1 telescope [3]. The basic detector is a 50 × 50 cm twelve gap spark chamber (MLSC) having gaps of 3 cm. Gamma ray events are selected on the basis that the photon converts into an electron-positron pair in the spark chamber and the electron is subsequently registered in the plastic upper and lower counters (UC and LC) and the gas Čerenkov counter (GČC) which define the event selection telescope. Time of flight analysis, accurate to approximately 2 ms, is used to reject slow protons or upward moving particles, whilst charged particle events are rejected by a set of anticoincidence counters surrounding the system (OD, ED and LD_i). Below the GČC lies a two-gap spark chamber (ASC) to track electron-positron pairs not visible in the upper chambers. Below the ASC is an ionisation calorimeter (Cal) consisting of four identical sections, each section consisting of six 60×60 cm lead plates 0.2 cm thick with 0.5 cm thick scintillation plates between. This calorimeter is designed to measure the energies in the range mentioned above from the electron-photon cascades produced inside it. An optional tungsten coded aperture (designated SKAT; a Russian acronym) is also included in the system which, on command, can be incorporated with the detector. Gamma-1 is expected to study and identify COS-B sources, observing them with a FWHM angular resolution of 1.2° without the aperture, and approximately 20 arcminutes with the aperture.

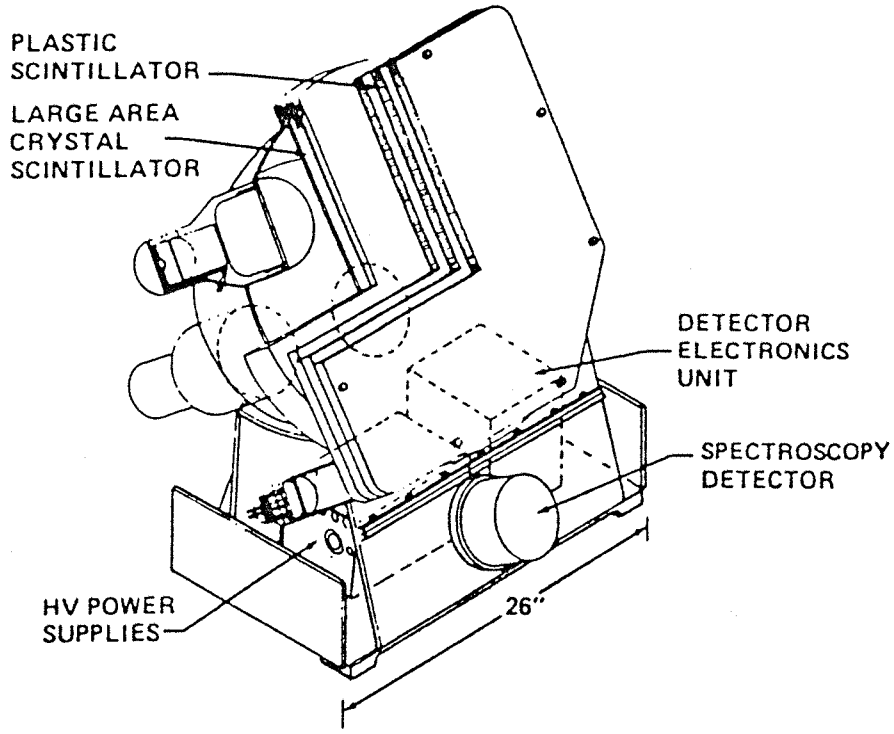


Fig. 2.19. *One BATSE detector module (from Fishman et al. [51]).*

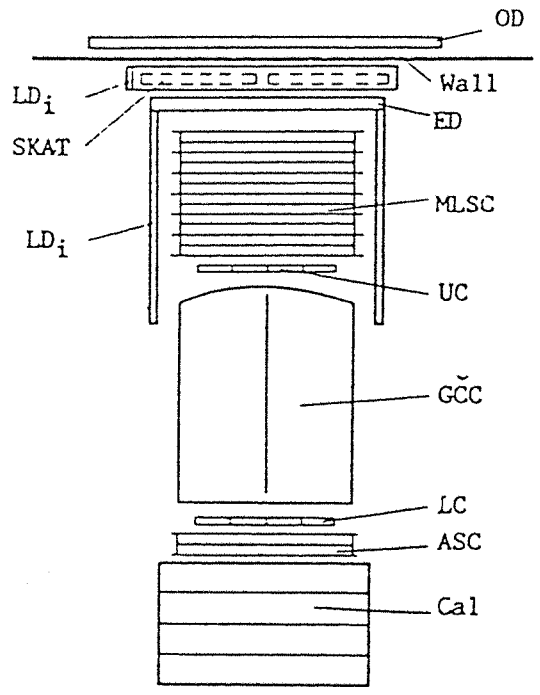


Fig. 2.20. *The Gamma-1 telescope (from Agrinier et al. [3]).*

2.3.3 SIGMA

The SIGMA experiment (Système d'Imagerie Gamma à Masque Aléatòire, i.e. random aperture gamma ray imaging device) is another French-Soviet mission which, it is expected, will be launched in 1989 on board the Soviet spacecraft GRANAT. The operational energy range will be 30 keV - 2 MeV. The telescope design [94] is based on a standard coded aperture system (fig. 2.21). The aperture to be used is rectangular of 49×53 elements generated from a basic 29×31 twin prime URA, with opaque elements being made of tungsten and having dimensions 9.4×9.4 mm ×1.5 cm thick. 2.5 metres below is a position sensitive high energy photon detector based on the Anger camera principle. The Anger camera is a block of NaI(Tl) of total area 825 cm^2 and 1.25 cm thick, and is viewed by 61 hexagonal PMTs glued to the crystal. Intensities measured by each PMT are weighted to determine the event position location to within 5 mm FWHM at 122 keV. The detector is surrounded by an active CsI anticoincidence shield limiting the field of view to about 1 steradian, whilst a 5 mm thick plastic scintillator, covering the field of view, rejects incoming charged particles. The angular resolution of SIGMA is approximately 7 arcminutes FWHM. The main objectives of SIGMA are to map accurately the distribution of the 511 keV electron-positron annihilation line first discovered in the Galactic centre by the HEAO-3 telescope [120], and to study AGNs and quasars by determining their overall luminosities and spectral features.

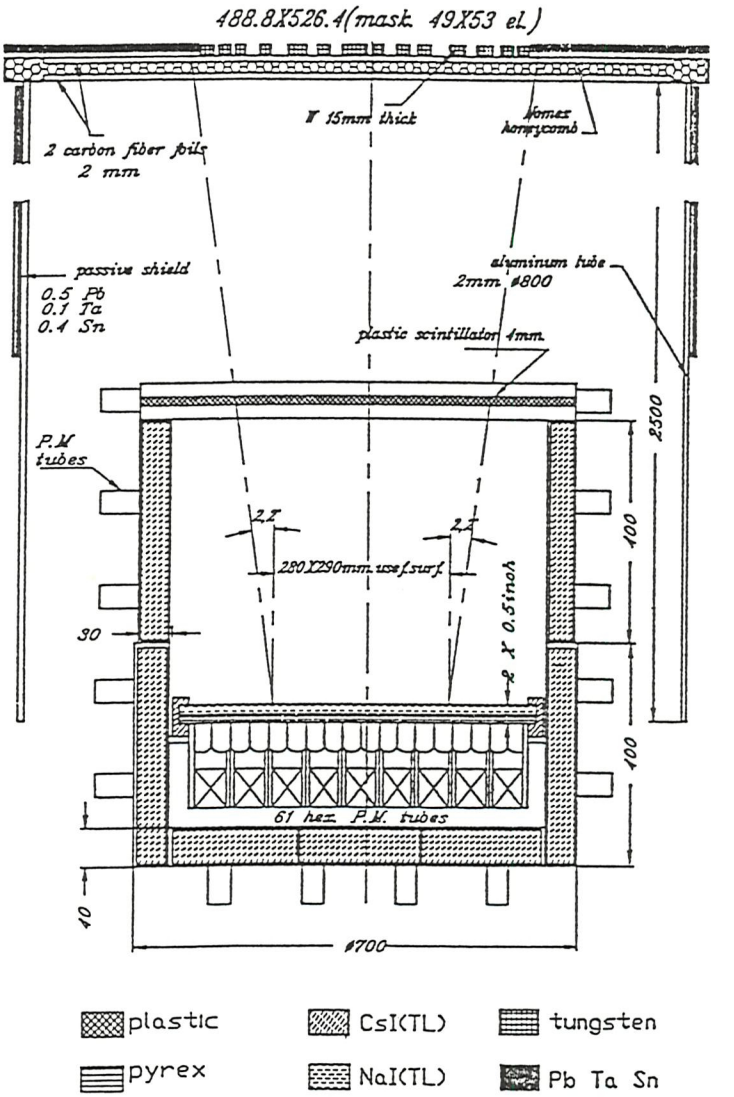


Fig. 2.21. *The SIGMA telescope (from Mandrou et al. [94]).*

Chapter 3

Coded Aperture Imaging.

3.1 Introduction

The penetrating nature of hard X rays and gamma rays ($E \geq 10$ keV) renders the imaging of astronomical objects in these wavebands impossible by conventional means. Novel techniques must therefore be employed if high resolution images of gamma ray sources are to be obtained. Attempts to image hard X rays have been made using grazing incidence mirrors [55] which, unfortunately, tend to have unfavourable aperture to mass ratio if used on a satellite. In the past, collimators, which have low resolving power or poor imaging properties, have provided the basis for instrument designs. However, in recent years, there has been a new approach to the imaging of high energy photon sources by a process based on the use of coded apertures. This technique is now generally accepted as the best method to create images at high photon energies. Systems based on this principle have already been used successfully in X ray astronomy [22,115,37,138], as well as medical diagnosis [122].

A single pinhole in an otherwise opaque material will cast a perfect, albeit inverted, image of the object onto a position sensitive detector with good angular resolution. However, this small effective aperture leads to a very low sensitivity. The coded aperture imaging technique increases the sensitivity by replacing this simple pinhole by many pinholes. Thus the main advantage of coded aperture imaging is the ability to maintain high angular resolution, such as would be given by a single pinhole camera, whilst keeping a wide field of view and a reasonably short observation time.

3.2 Basic Theory

Image formation using a coded aperture is a two stage process which relies on the rectilinear optics of photons. The first stage is to use an aperture consisting of elements having varying degrees of photon opacity to cast a shadow of a self luminous source onto a position sensitive photon detector. This is referred to as the source *convolution.* The second stage involves the decoding of the shadowgram detector image, by computational means, to give an estimated reconstruction of the original source distribution. This process is called the *deconvolution.* As already mentioned, a single pinhole in an otherwise opaque material will simply project an inverted image of the source distribution onto the detector and no decoding is necessary. However, the low signal to noise ratios experienced in gamma ray astronomy makes such a system impractical due to the very long observation times which would be required to create an acceptable image.

The coded aperture imaging process is shown schematically in fig. 3.1 and is explained below.

3.2.1 Convolution

If $\underline{w}$, $\underline{u_k}$, and $\underline{u}$ are vectors in the object plane, aperture plane and detector plane respectively, then the convolution equation is given by

$$P(\underline{u}) = O(\underline{w}) * A(\underline{u_k}) \tag{3.1}$$

where $P(\underline{u})$ is the detector image intensity distribution, $O(\underline{w})$ is the source distribution, $A(\underline{u_k})$ is the aperture transmission function and $*$ is the correlation operator.

If the system is quantized so that the detector bin (or pixel) size is equal to the aperture element size, then the following mathematical expression for the convolution is obtained

$$P(i,j) = \sum_x \sum_y O(x,y)A(i-x, j-y). \tag{3.2}$$

However, as pointed out by Fenimore and Cannon [46], all imaging systems suffer from an inherent noise which is not modulated by the aperture. Therefore, equation 3.1 needs to be modified by including a detector noise term $N(\underline{u})$, giving

$$P(\underline{u}) = O(\underline{w}) * A(\underline{u_k}) + N(\underline{u}). \tag{3.3}$$

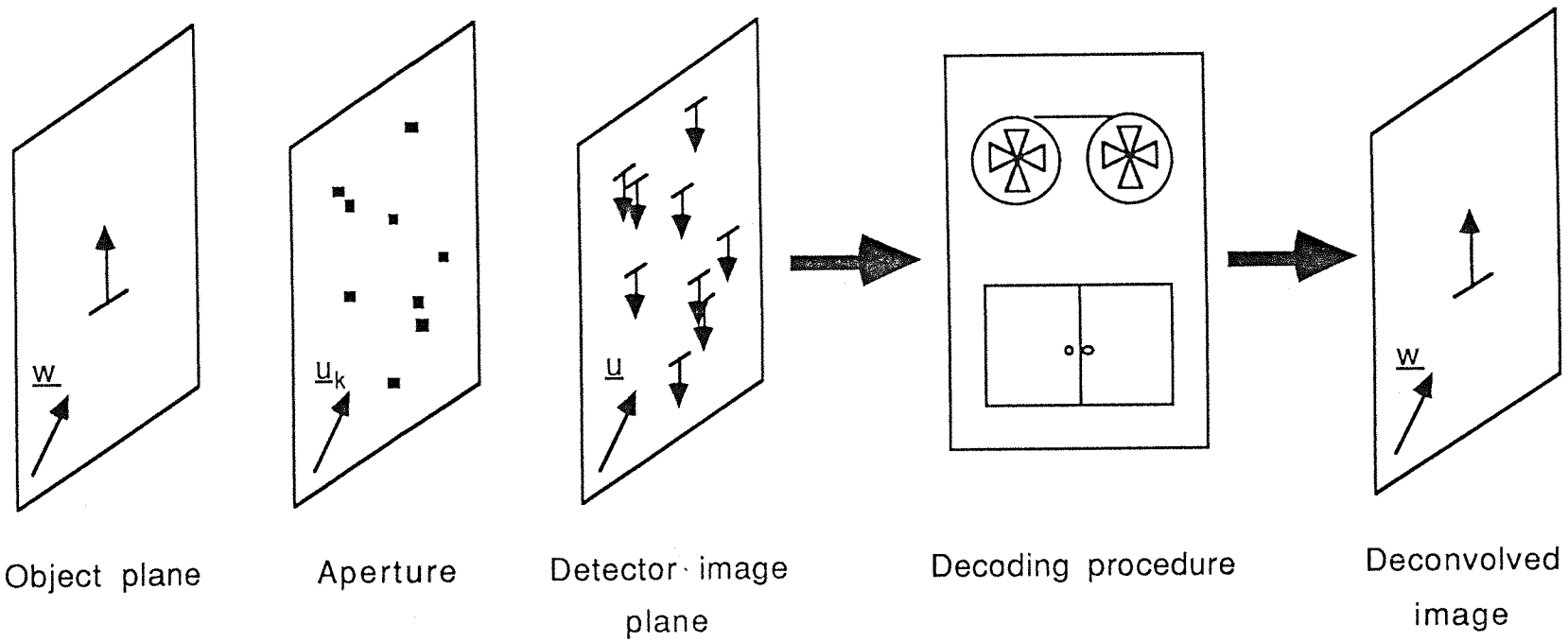


Fig. 3.1. *The basic steps of coded aperture imaging.*

3.2.2 Deconvolution

The recorded detector shadowgram normally bears no resemblance to the original source distribution, and a decoding procedure is therefore necessary. If $G(\underline{u}_k)$ is the decoding function, called the *deconvolution function*, then the reconstructed source distribution $\hat{O}(\underline{w})$ is given by

$$\hat{O}(\underline{w}) = P(\underline{u}) * G(\underline{u}_k). \tag{3.4}$$

Combining 3.3 and 3.4 gives

$$\hat{O}(\underline{w}) = O(\underline{w}) * A(\underline{u}_k) * G(\underline{u}_k) + N(\underline{u}) * G(\underline{u}_k). \tag{3.5}$$

Therefore, if the aperture $A(\underline{u}_k)$ and the deconvolution function $G(\underline{u}_k)$ are chosen such that $A(\underline{u}_k) * G(\underline{u}_k)$ is a delta function, then the deconvolved image $\hat{O}(\underline{w})$ will be a perfect reconstruction of the original source distribution $O(\underline{w})$, except for the presence of the noise term.

$A(\underline{u}_k) * G(\underline{u}_k)$ is referred to as the *system point spread function*, (SPSF) and for a quantized two dimensional system is given by

$$\text{SPSF}(i,j) = \sum_x \sum_y A(x,y)G(i-x, j-y) \tag{3.6}$$

and for a continuous system by

$$\text{SPSF}(i,j) = \int\int A(x,y)G(i-x, j-y)dx\,dy. \tag{3.7}$$

3.3 Historical Development of Coded Aperture Imaging

The development of coded aperture imaging has been concerned with choosing suitable functions for the aperture A, and deconvolution function G, to give the best approximation of the SPSF to a delta function. Sidelobes produced by imperfect SPSFs can cause problems by obliterating weak sources or even being mistaken for true sources. These problems are particularly serious in the gamma ray astronomical environment, where source fluxes are generally low compared to ever present background noise.

3.3.1 Fresnel Zone Plate

The idea of imaging by the use of a coded aperture was first conceived by Mertz and Young [102] who used a Fresnel zone plate (FZP) as the aperture. A FZP is characterised by the aperture function

$$A(r) = \cos(r^2) \tag{3.8}$$

where $r = (x^2 + y^2)^{1/2}$.

The motivation behind the use of the FZP as the aperture lies firstly in the fact that its relative aperture (i.e. the ratio of transparent area to total area) is 0.5 giving optimum source modulation, and secondly in its autocorrelation function $R_a(j)$, which is given by

$$\begin{aligned} R_a(j) &= \lim_{L\to\infty} \frac{1}{L} \int_{-L}^{L} \cos(r^2)\cos(j-r)^2 dr \\ &= \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 \end{cases} \end{aligned} \tag{3.9}$$

(c.f. equation 3.7). i.e. $R_a(j)$ is a perfect delta function. However, the continuous nature of such an aperture function, $A(r)$, makes construction of the FZP very difficult. Therefore, in practice, the FZP is approximated by using the binary aperture transmission function

$$A(r) = \begin{cases} 1 & \text{if } \cos(r^2) > 0 \\ 0 & \text{if } \cos(r^2) < 0 \end{cases} .$$

This gives a system of concentric rings of alternating transparent and opaque material, where the radius of the nth ring is equal to $r_i\sqrt{n}$, r_i being the radius of the innermost ring.

Two problems arise with this aperture. The first is that in practice the FZP and detector can only be of a finite size (i.e. $L \neq \infty$). Correlating the recorded image with an "infinite" deconvolution array will give a SPSF which has a spike corresponding to the source position, surrounded by a system of "ripples" which are intrinsic to the system. The second is due to the binary approximation which further degrades the deconvolved images. These problems can be partly overcome by using large apertures having many rings (apertures with $\lesssim 20$ opaque rings tend to give an undesirable image [28]) and FZPs have been used successfully in medical imaging [122].

3.3.2 Modulation Collimators

The idea of the modulation collimator was introduced by Oda [104]. The modulation collimator consists of several grids of alternating transparent and opaque material positioned in front of a detector which need not be position sensitive. Scanning the sky gives a detector "light curve" which enables source position to be uniquely determined within a periodic redundancy. The basic idea is shown in fig. 3.2. Fig. 3.2 (a) shows a two grid collimator and the transmission as a function of source angle. The source could have originated from every other sky pixel and so the redundancy is high. However, redundancy can be halved by placing another similar grid between the first two (3.2 (b)) so that the distance between adjacent windows is doubled. Redundancy can be further reduced by using more grids. (3.2 (c)).

An alternative method of achieving modulation is by the use of the rotation modulation collimator (RMC), first proposed by Mertz [101] and developed by Schnopper [129]. The RMC consists of a normal two grid modulation collimator, which rather than scanning the sky, the grid system is rotated. This means that any off axis source in the object plane will give a time varying response that can be analysed to obtain the source position accurately to within as close as ten seconds of arc.

3.3.3 Random Pinhole Arrays

The random pinhole array was first proposed by Dicke [40]. Here the single pinhole is replaced by an array of pinholes randomly placed in an otherwise opaque material, giving an aperture transmission over the entire aperture of approximately 50%. The motivation for using many pinholes is to improve the signal to noise ratio (SNR) in the reconstructed object by increasing the open aperture area, whilst maintaining the angular resolution of the single pinhole. The discrete nature of such an aperture has the advantage that it is easier to construct than the continuous FZP. Defining transparent aperture elements to be equal to one and opaque aperture elements to be zero, the autocorrelation function of an infinite random array consists of a spike with flat sidelobes. However, in practice only finite sized arrays are possible, and so statistical fluctuations in the random numbers used cause the sidelobes to be imperfect. Therefore, although overcoming the problems of the FZP "ripple effect", spurious peaks in the sidelobes are present. Such fluctuations can be reduced by

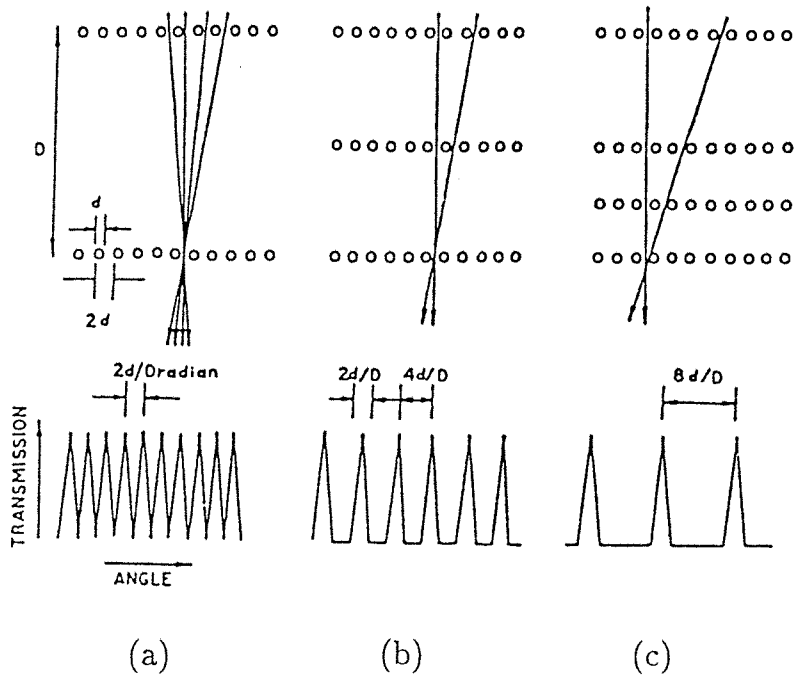


Fig. 3.2. *Increasing the redundancy when using a modulation collimator to image a gamma ray source (from Oda [104]).*

using smaller pinholes, but the detector positional resolution ultimately puts a lower limit to their size.

In Dicke's original system, the aperture is the same size as the detector and so the autocorrelation function is a spike on top of a pyramid, as shown in fig. 3.3 (a) for a one dimensional system. The sidelobes can be made to have approximately a constant value by making the aperture larger than the detector and restricting the field of view so as not to include any deconvolved sky pixel for which the detector is not wholly correlated with the aperture. The mean side lobe value in this case is approximately half of the height of the central spike, but spurious peaks are present. An example for a 300 element one dimensional aperture is given in fig. 3.3 (b).

Each source in the field of view produces sidelobes of intensity proportional to the source flux, and therefore may completely obliterate weaker sources. Another method of image reconstruction which considerably reduces the sidelobes was first proposed by Brown [25]. This method uses a cross correlation technique rather than the autocorrelation. Here, the deconvolution array, $G(i,j)$ used is defined by

$$\begin{aligned} G(i,j) &= 1 \quad \text{if} \quad A(i,j) = 1 \\ G(i,j) &= -1 \quad \text{if} \quad A(i,j) = 0 \end{aligned} \quad . \tag{3.10}$$

This cross correlation is known as *balanced correlation* [46]. The advantage of using balanced correlation is that the build up of of the side lobes is prevented, whilst the central spike retains the height of that produced by autocorrelation. Balanced correlation profiles for one dimensional 300 element apertures similar to those used to generate the profiles in fig. 3.3 are shown in fig. 3.4. The mean side lobe value is now approximately zero, but there are still statistical fluctuations which are inherent to the random aperture used.

Nevertheless, the random pinhole array has found application in the field of X ray astronomy [22].

3.3.4 Uniformly Redundant Arrays

The height and position of spurious peaks in a random pinhole camera image depend on the random number sequence used for the construction of the aperture. It was therefore natural to wonder if there was a set of "random" numbers giving an array with a perfect

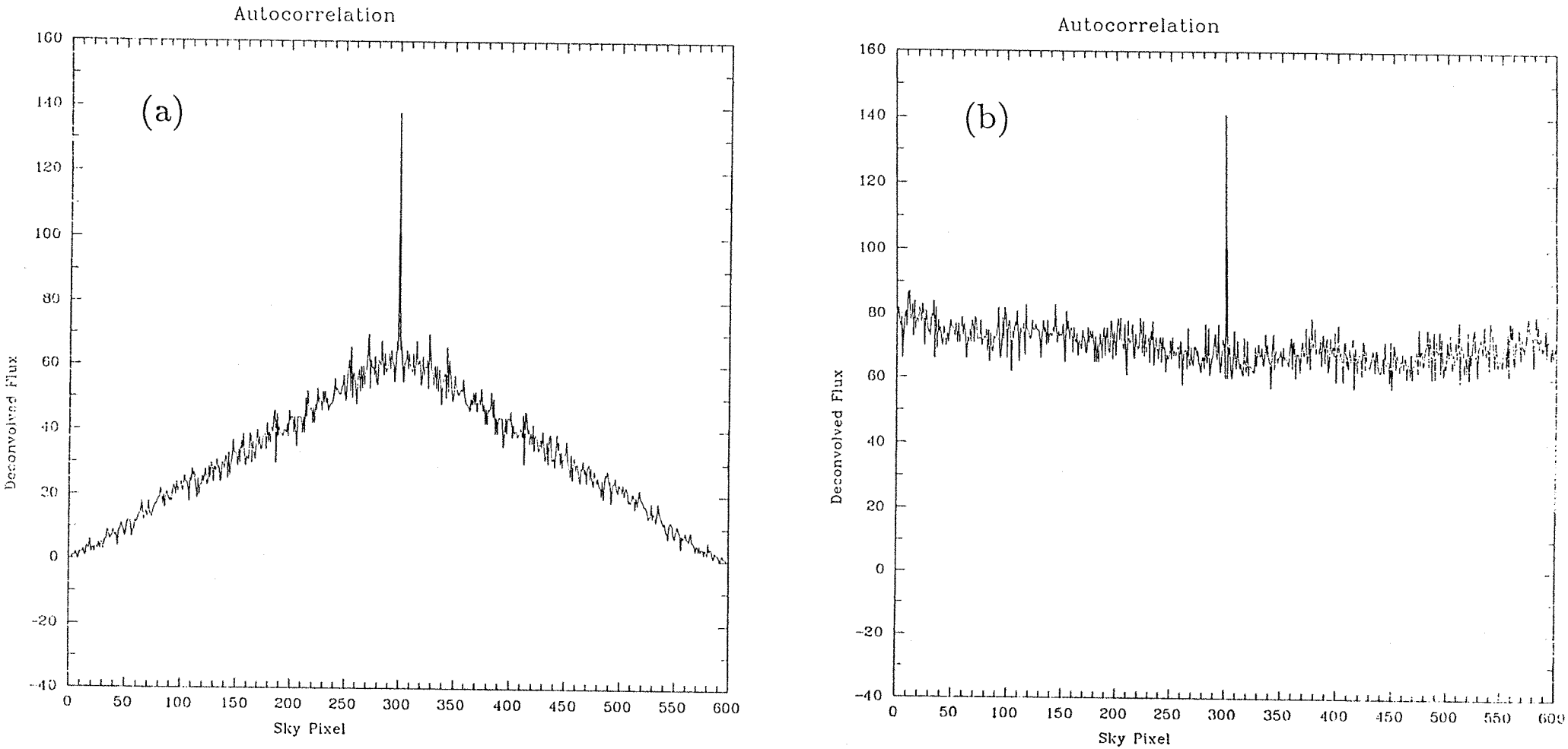


Fig. 3.3. *The SPSF of a 300 element random pinhole camera when using autocorrelation: (a) Dicke's original method with the aperture and detector the same size, (b) with the aperture larger than the detector.*

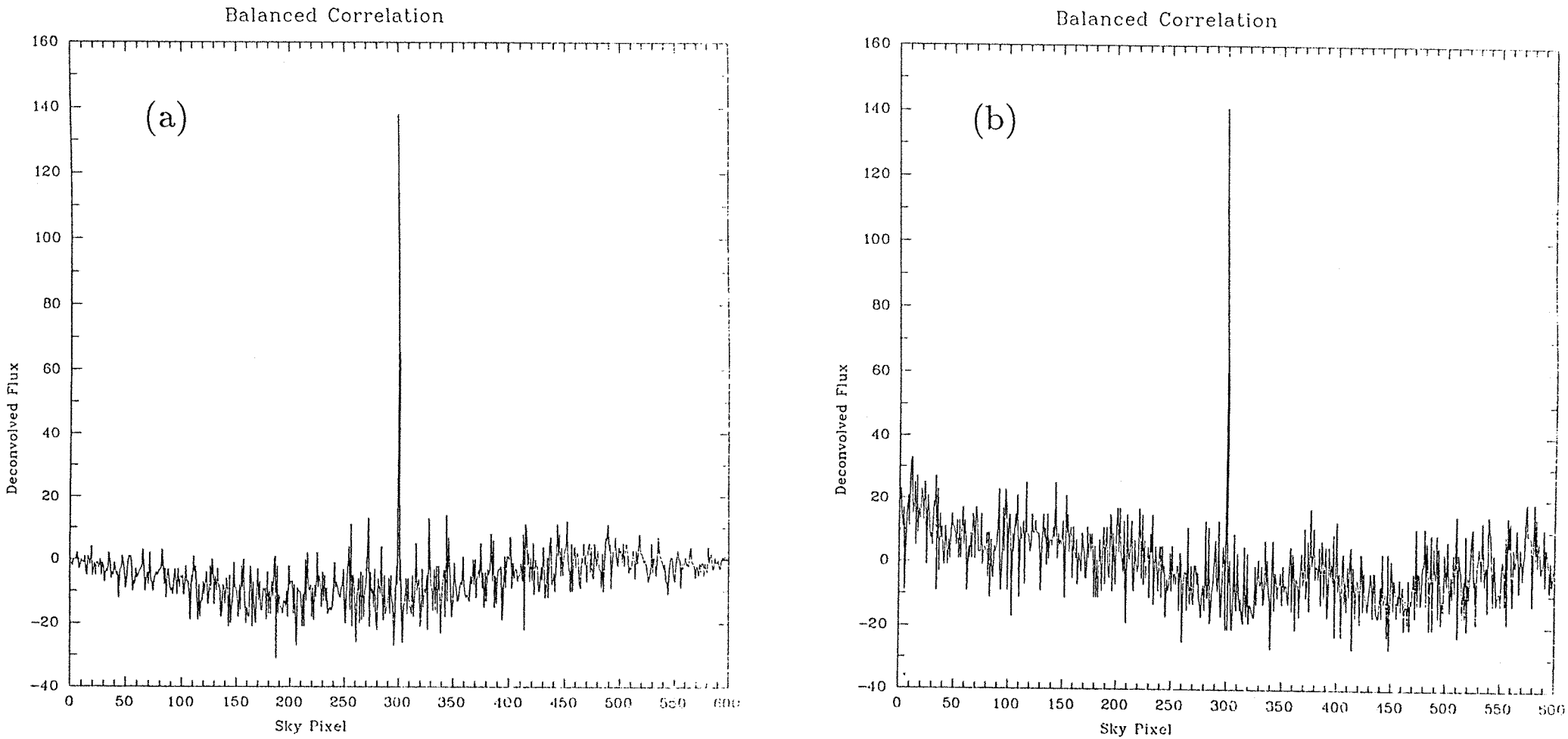


Fig. 3.4. *The SPSF of a 300 element random pinhole camera when using Brown's balanced correlation method: (a) with the aperture and detector the same size, (b) with the aperture larger than the detector.*

delta function SPSF.

If the condition of having an infinite size aperture is relaxed, then there is a class of array having perfect SPSFs with flat sidelobes. Such arrays are called *uniformly redundant arrays* (URA), and were proposed for use in 1976 by Gunson and Polychronopulos [60], and have since been extensively studied by Fenimore and Cannon [28,45] amongst others.

The reason that the sidelobes of a URA are flat can be seen if all the separations between all possible pairs of holes is measured (out to some maximum). Each separation occurs a uniform number of times and thus the separations are uniformly redundant.

A URA is a two dimensional representation of the one dimensional class of array known as the (v, k, λ) difference set. A thorough discussion of difference sets is given in chapter 4, though their definition is as follows. A (v, k, λ) difference set is a collection of k residues, modulo v, such that for any non zero residue d, the congruence $a_i - a_j \equiv d(\mathrm{mod} v)$ has exactly λ solution pairs (a_i, a_j) with a_i, a_j belonging to the difference set.

The important property of difference sets in connection with coded aperture imaging is the autocorrelation function $R_a(j)$ given in equation 4.16. It can be seen that $R_a(j)$ consists of a spike of height k at $j \equiv 0(\mathrm{mod}\, v)$ with flat sidelobes of height λ at $j \not\equiv 0(\mathrm{mod}\, v)$, as shown in fig. 3.5. Therefore the autocorrelation function of the incidence vector (see section 4.2.2) of a cyclically repeated (v, k, λ) difference set is periodic. Thus, cyclically repeating a URA aperture pattern also gives a periodic autocorrelation function for the imaging system.

Balanced correlation of the incidence vector gives sidelobes of value $2\lambda - k$ which, as will be shown in chapter 4, are equal to zero if the complement of this difference set is of the Hadamard type, i.e. $v = 4n - 1, k = 2n - 1, \lambda = n - 1$. Such difference sets are also referred to as *pseudo noise* (PN) sequences. Thus, Hadamard difference sets give rise to perfect periodic delta function SPSFs when using balanced correlation. Therefore, for the small price of restricting the field of view to include just one period, perfect imaging can be achieved with no source ambiguities.

Although Gunson and Polychronopulos first postulated the use of difference sets (and hence URAs) for coded aperture imaging in X ray astronomy, a special type of URA was previously considered for use in 1974 by Klemperer [77] in the field of baseline interferometry. Antenna arrays, such as the Very Large Array in New Mexico, employ the earth's rotation to acquire information from a large number of spatial frequencies, or baselines, so that over a period

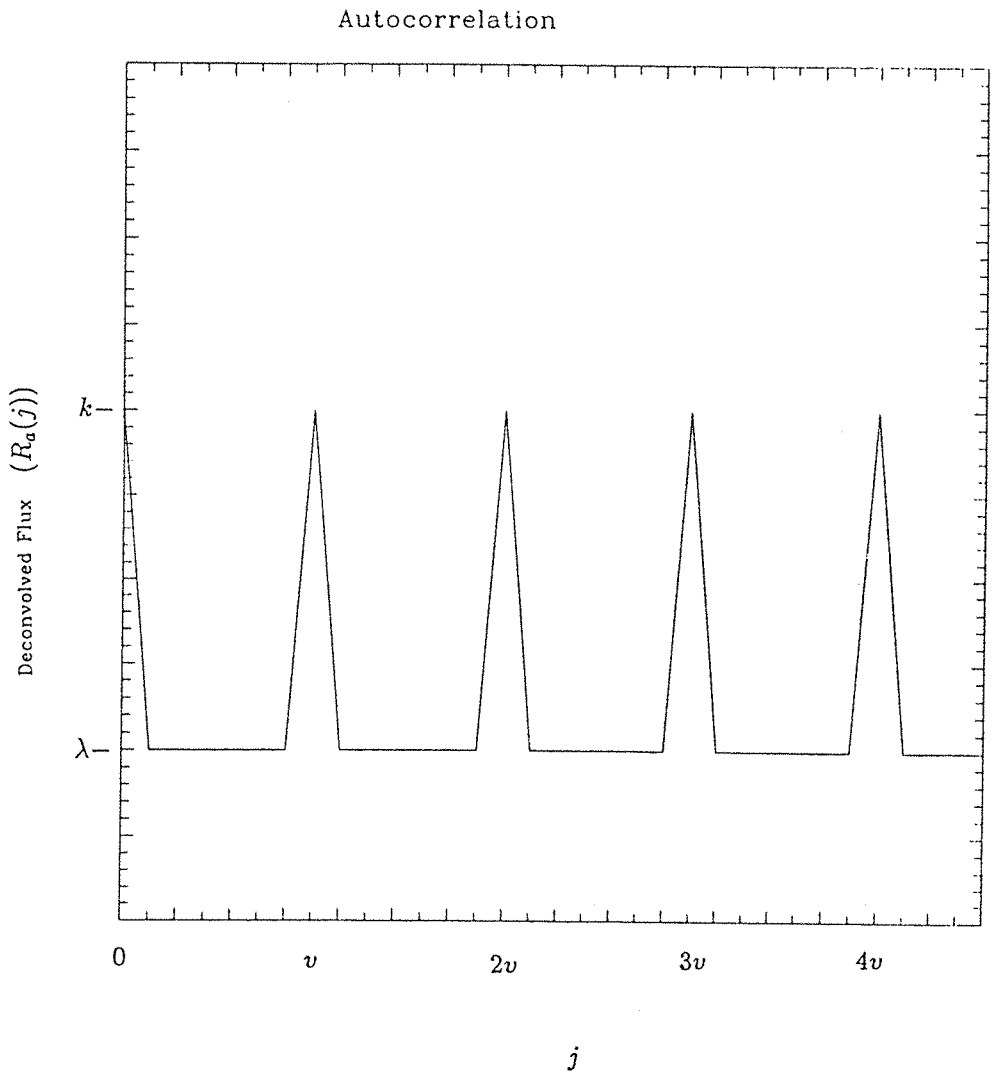


Fig. 3.5. *The autocorrelation function of a difference set.*

of about twelve hours, a complete map of a source can be built. However, rapidly varying sources cannot be mapped this way, and so Klemperer's idea was to dispense with the earth's rotation synthesis and sample all spatial frequencies simultaneously in a non redundant array (NRA). A NRA is simply a URA where each spatial frequency is sampled only once, out to some maximum. In the context of coded aperture imaging therefore, if k is the number of transparent elements, then for the difference set corresponding to the NRA, $\lambda = 1$ (i.e. a planar difference set).

Although suitable for baseline interferometry the NRA has such a low transparency (the largest of Klemperer's has twenty four holes with a relative aperture transparency of only 4%) that the observation time required for a reasonable gamma ray image would be too long.

Hadamard difference sets, and hence PN sequences, can be constructed by various methods. In the case of an aperture consisting of square elements, the method of shift registers is normally used [91,66], the construction details being given in section 4.3.1. Here, it is usually convenient to choose a PN sequence of order $v = 2^r - 1$ which is not a prime number, so that the basic pattern can be mosaiced onto an $n_1 \times n_2$ rectangular detector where n_1 and n_2 are relatively prime (i.e. they have no common factors) and $v = n_1 n_2$. When mosaicing the aperture from the PN sequence, all v elements of the sequence must appear inside each subset of the aperture which has the same size and shape of the detector. This is most conveniently achieved by placing the first PN sequence element into the $(0,0)$ position of the detector array, the second element into the $(1,1)$ position, and so on until the detector array is filled [164]. Thus, the ith element is placed in position $(i-1, i-1) \bmod (n_1, n_2)$ of the detector. The aperture is now simply a cyclic repetition of this basic pattern, constructed as large as possible so as to not contain the same basic $n_1 \times n_2$ subset more than once; i.e. of size $(2n_1 - 1) \times (2n_2 - 1)$. This will give a fully coded field of view of $n_1 \times n_2$ sky pixels. The Zebra telescope designed at Southampton University is mosaiced in this fashion from a PN sequence of $v = 63$ elements, onto a basic pattern of 9×7 elements, and full aperture of 17×13 elements [151].

3.3.5 Hexagonal Uniformly Redundant Arrays

To date, most of the studies of coded aperture imaging with URAs has been focussed on apertures based on square and rectangular lattices. Recently however, much attention has

been given to a new class of URA pattern based on a hexagonal lattice, which have particularly desirable properties for gamma ray astronomical imaging. These apertures, along with their imaging properties, have been studied by Cook et al. [35], and by Finger and Prince [50], and are constructed from a special type of Hadamard difference set; the *skew Hadamard* difference set.

Skew Hadamard difference sets are difference sets of order v that have the property that one of the residues a or $v-a$, but not both, appears in the set for $1 \leq a \leq v-1$. This property makes possible the construction of URAs which are antisymmetric upon a 180° rotation by the following procedure given by Finger and Prince [50].

1. Choose a prime number v of the form $v = 4n-1$. This is the order of the difference set, and hence is determined by the number of pixels on the detector.

2. Construct the skew Hadamard difference set, D, of order v by taking the quadratic residues modulo v as follows:

$$D = \left\{ 1^2, 2^2, \ldots, \left(\frac{v-1}{2}\right)^2 \right\} (\mathrm{mod}\, v).$$

   The proof that this is a skew Hadamard difference set is given in section 4.3.2.

3. Define the hexagonal lattice vectors $\underline{e}_0$ and $\underline{e}_1$ at 60° to each other, and choose an integer r such that the aperture element centred at $[i\underline{e}_0, j\underline{e}_1]$ (square brackets indicating a hexagonal co-ordinate system) is given the label l, where

$$l \equiv i + rj (\mathrm{mod}\, v).$$

4. Give all the aperture elements having $l \in D$ (l belonging to the set D) a value of zero (i.e. opaque elements), and the rest a value of one (transparent elements).

The URA constructed is periodic with the shape of each period being the same as the unit pattern of order v, which in turn is determined by the shape of the detector.

The fact that the URA is antisymmetric on 180° can be seen if one considers labels l_0 at 0° and l_{180} at 180°. Since $l_0 \equiv i_0 + rj_0 (\mathrm{mod} v)$, then $l_{180} \equiv i_{180} + j_{180} \equiv -(i_0 + j_0)(\mathrm{mod} v)$. Thus, $l_{180} \equiv -l_0 (\mathrm{mod} v)$. However, by the definition of a skew Hadamard difference set, if $l_0 \in D$, then $-l_0 \equiv l_{180} \notin D$, and vice-versa. Thus, skew Hadamard URAs are antisymmetric upon a 180° rotation, except for the central cell and its repetitions. This

property is particularly useful in gamma ray astronomical imaging. Systematic variations in the detector background, which may be larger than the detected source fluxes, can be effectively removed by taking two images at 180° aperture rotation intervals, and subtracting one detector image from the other prior to deconvolution.

Although this is a desirable property, the very large density of material required to absorb high energy photons means that the aperture will be very massive and so there may be mechanical problems in trying to rotate such an aperture through 180° (for example, the relatively small aperture used on the Zebra telescope has a mass of approximately 200 kg). There is, however, a subset of the skew Hadamard URAs which partly overcome this problem by requiring only a 60° rotation in order to become antisymmetric, depending on the value of the multiplier r chosen. These are referred to as *hexagonal uniformly redundant arrays* (HURA).

Consider the following. An aperture element has a label l at 0°. Assume that there is an integer m such that on a 60° rotation the element has a label $ml(\mathrm{mod}v)$, on 120° a label $m^2l(\mathrm{\ mod\ }v)$, and so on, as shown in fig. 3.6. As has already been pointed out, there is a 180° antisymmetry, and so using the Legendre symbol (a/v) for quadratic residues (i.e. $(a/v) = 1$ if a is a quadratic residue $(\mathrm{mod}v)$, $(a/v) = -1$ if a is a quadratic nonresidue $(\mathrm{mod}v)$, and $(a/v) = 0$ if v divides a, see section 4.3.2), then if l is a quadratic residue, m^3l is a quadratic nonresidue and vice-versa. Therefore

$$(m^3l/v) = -(l/v). \tag{3.11}$$

However, in number theory, Fermat's theorem (corollary 4.8) states that $m^{v-1} \equiv 1(\mathrm{mod}\ v)$ for prime v. Therefore $(m^2)^{(v-1)/2} \equiv 1(\mathrm{mod}\ v)$ which, multiplying by $l^{(v-1)/2}$ gives

$$(m^2l)^{(v-1)/2} \equiv l^{(v-1)/2}(\mathrm{mod}\,v)$$

which, by Euler's criterion for quadratic residues (theorem 4.9) gives

$$(m^2l/v) = (l/v). \tag{3.12}$$

Now, combining 3.11 and 3.12 gives

$$(m^3l/v) = -(m^2l/v). \tag{3.13}$$

Theorem 4.11 shows that the Legendre symbol is multiplicative, and so equation 3.13 becomes

$$(m^2/v)(ml/v) = -(m^2/v)(l/v)$$

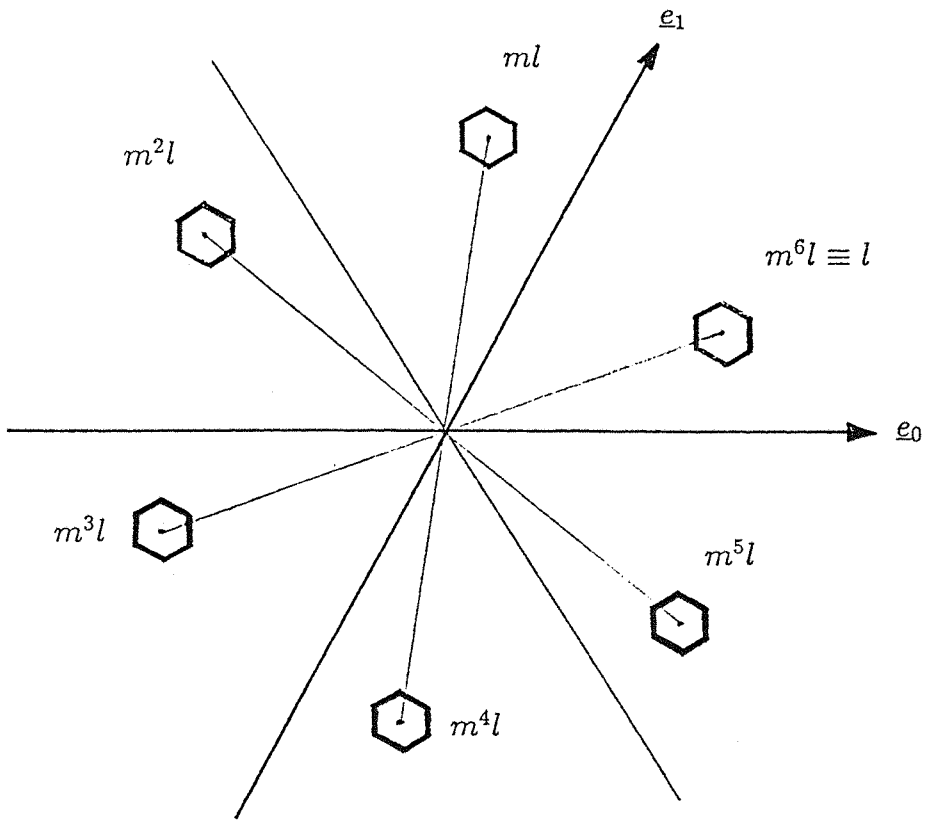


Fig. 3.6. *Hexagons in a hexagonal array which, when rotated through* 60°, *have their labels changed by a factor of* $m \pmod{v}$.

giving

$$(ml/v) = -(l/v). \tag{3.14}$$

Therefore if an element labelled l has the label $ml(\mathrm{mod}v)$ on 60° rotation, then the URA will be antisymmetric on 60° rotation, since the Legendre symbols of l and ml will be of opposite sign as given by 3.14. How can a value of m be calculated ?

The hexagonal co-ordinates $[i_1, j_1]$ of an element at 0° become $[i_2, j_2]$ on a 60° rotation, where

$$\left.\begin{array}{rcl} i_2 & = & -j_1 \\ j_2 & = & i_1 + j_1 \end{array}\right\}. \tag{3.15}$$

Therefore the label at 0° is

$$l \equiv i_1 + rj_1(\mathrm{mod}v) \tag{3.16}$$

and at 60° becomes

$$ml \equiv i_2 + rj_2 \equiv ri_1 + (r-1)j_1(\mathrm{mod}v). \tag{3.17}$$

Combining 3.16 and 3.17 gives

$$ml \equiv mi_1 + mrj_1 \equiv ri_1 + (r-1)j_1(\mathrm{mod}v). \tag{3.18}$$

Comparing coefficients in 3.18 for i_1 and j_1 gives

$$\begin{array}{rcll} m & \equiv & r & (\mathrm{mod}v) \\ mr & \equiv & r-1 & (\mathrm{mod}v) \end{array}$$

giving

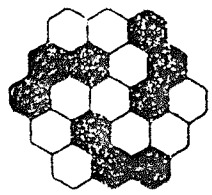
$$r^2 \equiv r - 1(\mathrm{mod}v). \tag{3.19}$$

Therefore, if the multiplier r is chosen such that 3.19 holds, then the URA will be anti-symmetric upon 60° rotation as well as 180°, except for the central cell ($l \equiv 0$) and its repetitions, and will thus be a HURA.

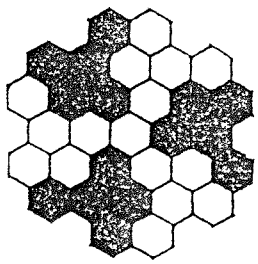
HURAs are a subset of the skew Hadamard URAs, since not all values of v give a solution for r in 3.19. In fact, since the basic unit pattern is of a hexagonal symmetry, it must have a multiple of six elements in total plus the central element, i.e. $v = 6t+1$ (integer t). Also, since $v = 4n-1$, then HURAs exist for any prime of the form $v = 12s+7$. This leaves , as possible orders for HURAs, the set of primes $v = 7, 19, 31, 43, 67, 79, 103, 127, 139, 163, 199, \ldots$. The first few are shown in fig. 3.7.

HURAs have already found application in real gamma ray astronomy on the Gamma Ray Imaging Payload (GRIP) [4] to observe, for example, the supernova SN1987A [37].

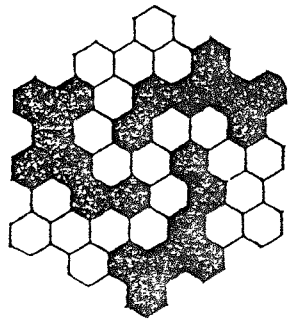
$v = 7$



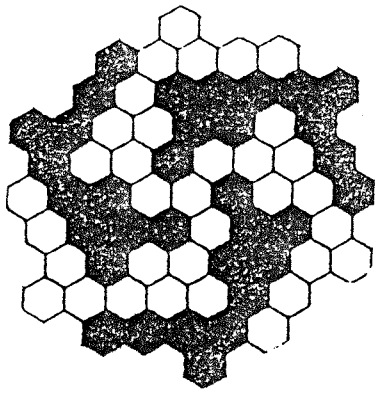
$v = 19$



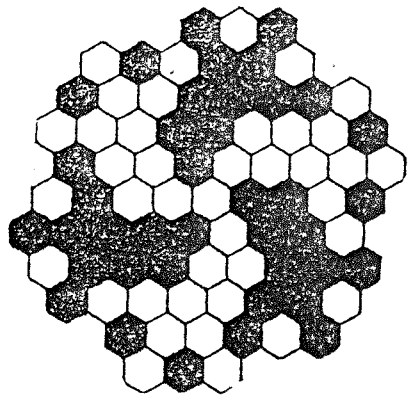
$v = 31$



$v = 43$



$v = 67$



$v = 79$

Fig. 3.7. *Basic patterns of the first six HURAs.*

3.4 Signal to Noise Ratio Considerations

After an image of an object has been reconstructed, it is necessary to quantify the image quality by measuring the significance of the increase of the source above the background level. Considering the whole object as an array of point sources, Fenimore [44] has presented an analysis of the signal to noise ratio (SNR) for a general case of a coded aperture having a flat side lobe SPSF.

Assuming that there is no systematic variation in the number of detector counts, there are two main contributions to the background noise. The first is the detector background which is not modulated by the aperture, and the second is due to the source photons which do undergo aperture modulation. As defined by Fenimore [44], the SNR at a point (i, j) in the reconstructed object is

$$\mathrm{SNR}(i,j) = \frac{\hat{O}(i,j)}{(\sigma^2)^{1/2}} = \frac{\hat{O}(i,j)}{({\sigma_1}^2 + {\sigma_2}^2 + {\sigma_3}^2)^{1/2}} \tag{3.20}$$

where $\hat{O}(i,j)$ is the reconstructed object (above the mean background level), and σ^2 is the total variance in $\hat{O}(i,j)$. The total variance is a combination of three separate variance factors: ${\sigma_1}^2$ is the variance of the contribution of $O(i,j)$ (the object flux at (i,j)) to $\hat{O}(i,j)$, ${\sigma_2}^2$ is the variance of the contribution of all other sources in the field of view to $\hat{O}(i,j)$, and ${\sigma_3}^2$ is the variance of the contribution of the background to $\hat{O}(i,j)$.

Defining the source flux and the detector background in terms of a single detector pixel, and assuming a perfectly efficient detector and perfectly opaque closed aperture elements gives

$$\mathrm{SNR}(i,j) = \frac{(v-k)S_{ij}}{[(v-k)S_{ij} + (v-k)\sum_l \sum_m S_{lm} + vD]^{1/2}}. \tag{3.21}$$

Here, v is the total number of aperture elements in the unit pattern, k is the number of opaque aperture elements per unit pattern, S_{ij} is the source flux per detector pixel from the source positioned at the (i,j) point of the object, and D is the average background counts per detector pixel.

In the case of a single point source positioned at (i,j), the value of $\sum_l \sum_m S_{lm}$ is zero, since there is no contribution from any other source in the field of view. Moreover the total counts, S, over the detector from the source is equal to $(v-k)S_{ij}$, and the total counts, B, from the background is vD. Therefore 3.21 becomes

$$\mathrm{SNR}(i,j) = \frac{S}{(S+B)^{1/2}}. \tag{3.22}$$

This is the expected SNR of a point source in the presence of non varying statistical background noise, and in the absence of systematic effects.

Chapter 4

Difference Sets. Their Properties and Construction Methods.

4.1 Introduction

Following work done by Mertz [100], the early days of coded aperture imaging saw X and gamma ray observations being carried out by using a Fresnel zone plate as the aperture. However, the difficulty in building such a system, which requires an aperture of continuously varying transmission function in order to obtain a perfect image, meant that this "ideal" aperture had to be replaced by one having binary transmission values, constructed in such a way as to give the best possible approximation. This, of course, gave images containing inherent noise which takes the form of a "ripple" effect of concentric rings centred around the peak. The random pinhole camera first suggested by Dicke [40] overcomes this ripple problem but suffers from spurious peaks in the autocorrelation function arising from statistical fluctuations.

The low photon fluxes which are experienced in gamma ray astronomical observations require an imaging system which minimises such inherent noise so that weak sources can be seen. In 1976 Gunson and Polychronopulos [60] introduced the use of arrays having no inherent

noise and hence perfect delta function autocorrelations which are now extensively used in gamma ray imaging. These one dimensional arrays are called *difference sets*, and the optimum difference sets to use are the *Hadamard* difference sets (which give rise to uniformly redundant arrays) which maximise the height of the delta function peak with respect to the height of the side lobes.

Difference sets arise naturally from combinatorial configurations known as block designs, and in order to fully understand difference sets, a knowledge of the properties of block designs is essential. Therefore section 4.2.1 is devoted to a discussion of block designs. Section 4.2.2 discusses the difference sets themselves, along with their properties, with a view to choosing the configuration most suitable for gamma ray imaging (i.e. the Hadamard configuration). For more information on the subject matter in section 4.2 see for example [127] chapters 8 and 9, and in particular [62] chapters 10 and 11. For the definitive work on difference sets see [13].

Section 4.3 deals with the two main Hadamard difference set construction methods. In section 4.3.1 the properties of shift register sequences are discussed, the proof that they lead to Hadamard difference sets, and the difference set construction method. For a more thorough discussion see also [66] and [57].

Finally in section 4.3.2 the quadratic residue construction method is given. Firstly the essential number theory is briefly described and then the properties of quadratic residues followed by the actual Hadamard difference set construction method. Quadratic residues are discussed in many books on number theory, but particularly good are [2] and [75].

A third commonly used construction method is the so called "twin prime" construction which gives Hadamard difference sets of order $p(p+2)$ where both p and $p+2$ are prime. Construction methods can be found in [59,31,24,27] and in particular [141].

4.2 Combinatorial Configurations

4.2.1 Block Designs

General Discussion

A problem which often appears in combinatorial mathematics is that of arranging objects into a specified number of sets so that the ith object appears r_i times in all the sets taken together, the jth set contains k_j objects, and so that groups of objects such as pairs and triplets occur a specified number of times. Such an arrangement may be called an *incidence system.* An example of such a system is called a *balanced incomplete block design*, also sometimes referred to more briefly as a *block design.*

A balanced incomplete block design is an arrangement of v objects (a_i), into b sets, called *blocks*, (B_j) such that:

(i) each block contains exactly k distinct objects

(ii) each object occurs in exactly r different blocks,

(iii) any pair of distinct objects occurs together in exactly λ different blocks.

(The term "incomplete" refers to the fact that not every combination of the v objects taken k at a time is considered as a block of the design. However, a portion of these combinations is "balanced" so far as comparisons between pairs of objects are concerned). Therefore a balanced incomplete block design has five parameters, and as a result is sometimes also referred to as a (b, v, r, k, λ) configuration.

It is possible to count the total number of incidences (i.e. the number of occurrences of one object in a block) in two different ways. One way is to see that there are v objects each occurring in r blocks giving vr incidences, and the other is to see that there are b blocks, each containing k objects giving bk incidences. It therefore follows that

$$bk = vr \tag{4.1}$$

Another relationship between the parameters can be derived by counting the occurrences of pairs containing a particular object, a_1. The object a_1 occurs in r blocks, and in each of

these is a pair with the $(k-1)$ remaining objects, whilst on the other hand a_1 is paired λ times with each of the remaining $(v-1)$ objects. Therefore

$$r(k-1) = \lambda(v-1) \tag{4.2}$$

Equations 4.1 and 4.2 are the two elementary relations on the five parameters of a balanced incomplete block design.

A block design is described by its *incidence matrix*, $A = (a_{ij})$, where $i = 1, \ldots, v$ and $j = 1, \ldots, b$. If $a_1, \ldots, a_v$ are the objects and $B_1, \ldots, B_b$ are the blocks, then

$$\left.\begin{aligned} a_{ij} &= 1 \quad \text{if } a_i \in B_j \\ a_{ij} &= 0 \quad \text{if } a_i \notin B_j \end{aligned}\right\} \tag{4.3}$$

Therefore, the columns of A are effectively the blocks of the design, B_j . The basic properties of the block design imply

$$w_v A = k w_b \tag{4.4}$$

$$AA^T = B = (r-\lambda)I_v + \lambda J_v \tag{4.5}$$

where w_v and w_b are the vectors of v and b 1's respectively, I_v is the $v \times v$ unit matrix and J_v is the $v \times v$ matrix of all 1's. Equation 4.4 expresses the fact that each column of A contains k 1's and therefore that each block contains k objects. The element b_{ij} of the matrix B in equation 4.5 is the inner product of the ith row of A with the jth row of A. Therefore b_{ii} is the number of 1's in the ith row of A, equal to r. But if $j \neq i$ then both the ith and jth rows have a 1 in column t if a_i and a_j both belong to B_t. Therefore, an off diagonal element b_{ij} of B counts the occurrences of the pair a_i, a_j, and is therefore λ.

The determinant of the matrix B can be easily calculated by using the property that adding or subtracting any row to any other row or any column to any other column of a matrix does not alter its determinant. Therefore, to calculate det B, subtract the first column of B from all the others and then add rows $2, \ldots, v$ to the first row. This gives a matrix with $r + (v-1)\lambda$ as the first entry on the main diagonal with the rest being $r - \lambda$, whilst all entries above the main diagonal are zero. The determinant is now simple to evaluate:

$$\det B = \det (AA^T) = (r-\lambda)^{v-1}(v\lambda - \lambda + r) \tag{4.6}$$

There are two very important inequalities relating to the parameters of a block design. The first can be easily derived from the trivial inequality that, for an incomplete design, $k < v$.

Using this and equation 4.2 gives us the result

$$\lambda < r \tag{4.7}$$

The second, also known as Fisher's inequality, makes use of equation 4.6. We know that A is if size $v \times b$. Now assume that $b < v$, and adjoin $v - b$ columns of zeros to A to obtain a square matrix A^* of order v such that

$$A^* A^{*T} = AA^T = (r - \lambda)I_v + \lambda J_v.$$

This is identical to equation 4.5 since, by matrix algebra, MM^{T} is invariant under addition of columns of zeros to the matrix M. Therefore, by equation 4.6

$$\det(A^* A^{*T}) = \det(A^*)\det(A^{*T}) = 0 = (r - \lambda)^{v-1}(v\lambda - \lambda + r)$$

which, by equation 4.7, means that $\lambda = r/(1 - v)$ giving a negative value for λ which is impossible. Therefore, the assumption "$b < v$" is false, which gives us Fisher's inequality:

$$b \geq v, \text{ and hence } r \geq k \tag{4.8}$$

Symmetric Block Designs

Fisher's inequality gives a lower bound for the number of blocks needed to construct a block design for a fixed number of objects. If this bound is achieved and $b = v$ (and $k = r$) then the block design is said to be symmetric, and is referred to as a *symmetric balanced incomplete block design*, or as it is sometimes called, a (v, k, λ) configuration. Therefore, by 4.2, symmetric block designs satisfy

$$k(k - 1) = \lambda(v - 1) \tag{4.9}$$

and the incidence matrix is a $v \times v$ square matrix satisfying

$$AA^T = B = (k - \lambda)I_v + \lambda J_v \tag{4.10}$$

obtained by substituting for $r = k$ into 4.5. Two other relations for A are given by

$$J_v A = kJ_v \tag{4.11}$$

$$AJ_v = kJ_v \tag{4.12}$$

The first says that every column of A contains k 1's, this being the number k of objects in each block. The second says that every row of A contains k 1's, this being the number $r = k$

of blocks containing each object. Combining 4.6, 4.7 and 4.9 for $r = k$ gives the result that $\det(AA^T) = \det^2 A \neq 0$. Thus, A is non singular and therefore possesses an inverse, A^{-1}. This enables us to prove the following very important result due to Ryser [126]. If 4.10 is multiplied by A^{-1} this gives

$$\begin{aligned} A^T &= (k-\lambda)A^{-1} + \lambda A^{-1} J_v \\ &= (k-\lambda)A^{-1} + \lambda k^{-1} J_v \end{aligned}$$

using 4.12. Therefore

$$A^T A = (k-\lambda)I_v + \lambda k^{-1} J_v A$$

which, by 4.11 gives

$$A^T A = AA^T = (k-\lambda)I_v + \lambda J_v \tag{4.13}$$

Therefore the incidence matrix of a symmetric block design is normal (i.e. it commutes with its transpose under multiplication). This property asserts the very important consequences that:

(i) any two distinct blocks of the design have exactly λ elements in common.

(ii) A^T is also the incidence matrix of a symmetric block design.

An example of a symmetric block design and its incidence matrix is given in appendix A. In the example in appendix A, $\lambda = 1$ and such designs are known in the terminology as finite projective planes, where the objects are usually referred to as "points" and the blocks as "lines", satisfying the three following axioms:

(i) two distinct points are contained in a unique line.

(ii) two distinct lines contain a unique common point.

(iii) each line contains at least three points.

For a more complete explanation of finite projective geometry see [62], chapter 12.

Existence Criteria for Symmetric Block Designs

The existence question for symmetric block designs is, at present, unsolved. That is, for a given parameter set v, k, λ, it is, in general, impossible to decide (short of an exhaus-

tive search) whether a symmetric block design with these parameters exists. Nevertheless significant progress has been made.

The obvious existence test is to see if these parameters satisfy equation 4.9. However there are many parameter sets which do satisfy 4.9 but for which it is known that no symmetric block design exists. Therefore more stringent conditions are required. To date the strongest known conditions on the existence of symmetric block designs is the following theorem, called the Bruck-Ryser-Chowla Theorem.

Theorem 4.1 (Bruck-Ryser-Chowla Theorem) *If a symmetric block design exists with parameters v, k, λ then, writing $n = k - \lambda$:*

1. *If v is even, n is a square*

2. *If v is odd, then the Diophantine equation*

$$z^2 = nx^2 + (-1)^{(v-1)/2}\lambda y^2$$

has a solution in integers x, y, z not all zero.

This Theorem was first proven for finite projective planes (i.e. $\lambda = 1$) in 1949 by Bruck and Ryser [26]. The following year, the theorem in its full form for general λ was proven by Chowla and Ryser [32].

Condition 1 is proven by using equation 4.6 for $r = k$ and a square matrix A. This gives

$$\det(AA^T) = \det^2 A = (k - \lambda)^{v-1}(\lambda(v - 1) + k)$$

which, using 4.9 and writing $n = k - \lambda$ gives

$$\det^2 A = k^2 n^{v-1}.$$

Both sides of this equation are squares of an integer. Therefore, if v is even then $n = k - \lambda$ must be a square proving condition 1. The proof of condition 2 is very complicated but can be found in [32] or in [13], chapter 2 amongst others.

Although the Bruck-Ryser-Chowla theorem is only a necessary condition for the existence of symmetric design, it should be noted that there is no parameter set v, k, λ which satisfies this

theorem and equation 4.9 for which it is known that no symmetric design exists. Therefore it is conceivable that the Bruck-Ryser-Chowla conditions are even sufficient for the existence of the design but this has not been proved.

Cyclic Block Designs

A special type of symmetric block design which often occurs is a *cyclic* symmetric block design. This is a block design where if block B_i contains the objects $\{a_i\}$, then block B_{i+s} contains the objects $\{a_{i+s}\}$, where s is an integer representing a phase shift, and $i+s$ is taken modulo v. This kind of phase shifting can be thought of as a mapping of objects and blocks and such a mapping is called an *automorphism* of the block design. That is, if all subscripts i of objects and blocks are replaced by $i+s$ modulo v, then the objects and blocks of the design are simply permuted in a cycle of length v, and the incidence matrix A remains the same. The block design in appendix A is an example of a cyclic symmetric block design.

4.2.2 Difference sets

Definition and Relationship with Block Designs

A set of k residues $D : \{a_1, \ldots, a_k\}$ modulo v is called a *difference set* if for every residue $d \not\equiv 0 (\mathrm{mod} v)$ there are exactly λ solution pairs (a_i, a_j) with $a_i, a_j \in D$ such that

$$a_i - a_j \equiv d (\mathrm{mod} v).$$

An example of a difference set is the following:

$$D = \{1, 2, 4\} (\mathrm{mod} 7)$$

where $v = 7$, and $k = 3$. If every difference $a_i - a_j (\mathrm{mod} 7)$ is taken from this set, then it will be found that each non zero difference modulo 7 occurs once only. Therefore $\lambda = 1$.

Difference sets and block designs are very closely interlinked as is shown by the following theorem:

Theorem 4.2 *A set of k residues $D : \{a_1, \ldots, a_k\}$ modulo v is a difference set if and only*

if the sets $B_i : \{a_1+i, \ldots, a_k+i\}$ *modulo* v, $i = 0, \ldots, v-1$, *are a cyclic symmetric balanced incomplete block design.*

Proof. Suppose that $D : \{a_1, \ldots, a_k\}$ is a difference set and that $B_i : \{a_1 + i, \ldots, a_k + i\}$ are sets of residues modulo v with $i = 0, \ldots, v-1$. Now, let r and s be residues such that $r - s \equiv d \not\equiv 0 (\bmod v)$. Then, by the definition of a difference set, there are exactly λ ordered pairs $a_i, a_j \in D$ such that

$$r - s \equiv a_i - a_j \equiv d (\bmod v).$$

Define t such that $r \equiv a_i + t$ and $s \equiv a_j + t$ and so r and s both belong to B_t, where t is determined by $t \equiv r - a_i \equiv s - a_j (\bmod v)$. Since there are λ ordered pairs (a_i, a_j) for each non zero difference, d, then, to keep r and s constant, there are λ values of t. Therefore there are λ sets B_t which contain the pair r and s. Therefore The sets B_i form a cyclic symmetric balanced incomplete block design or a cyclic (v, k, λ) configuration, since each pair (r, s) occur in λ sets.

Because difference sets appear in this way from cyclic (v, k, λ) configurations, they are sometimes called *cyclic difference sets* or (v, k, λ) *difference sets*, and the example of a difference set given above is, in fact, block B_1 of the cyclic symmetric block design in appendix A, with $(v, k, \lambda) = (7, 3, 1)$.

An immediate consequence of theorem 4.2 is that the parameters of a (v, k, λ) difference set satisfy equation 4.9 for symmetric block designs.

General Discussion and Properties of Difference Sets

Consider the two following difference sets:

$$\begin{aligned} D &: \{1, 2, 4\} \quad \text{modulo } 7 \\ D^* &: \{0, 3, 5, 6\} \quad \text{modulo } 7. \end{aligned}$$

The set $D + D^*$ is, therefore, a complete set of residues modulo 7. The sets D and D^* are said to be *complementary* difference sets. In general if D is any difference set with parameters v, k, λ, then its *complement*, D^*, is also a difference set with parameters v^*, k^*, λ^* satisfying

$$\left.\begin{aligned} v^* &= v \\ k^* &= v - k \\ \lambda^* &= v - 2k + \lambda \end{aligned}\right\} \tag{4.14}$$

(the third equation being derived from the first two and equation 4.9). Recalling also that $n = k - \lambda$ it can be seen that both v and n are invariant under complementation. For most purposes it is sufficient to consider only one of a pair of complementary difference sets. This is frequently done by insisting that k be less than $v/2$ (equation 4.9 shows that $k = v/2$ cannot occur since λ could then never be an integer). Thus of the four difference set parameters (v, k, λ, n), v and n are the most fundamental. For this reason it is sometimes useful to express k and λ in terms of them. Using 4.9:

$$kk^* = k(v-k) = n(v-1)$$

but, since $k = v/2$ cannot occur, kk^* is maximised when $k = (v-1)/2$ and therefore $k^* = (v+1)/2$. Therefore in general:

$$kk^* = n(v-1) \leq \frac{(v-1)}{2}\frac{(v+1)}{2}$$

giving

$$v \geq 4n - 1.$$

An upper limit for v can be found from 4.9 by observing that, for the lowest possible λ for non trivial difference sets $\lambda = 1$, we get $v = n^2 + n + 1$ and $k = n - 1$. Therefore:

$$n^2 + n + 1 \geq v \geq 4n - 1. \tag{4.15}$$

Therefore, for a given n, equation 4.15 gives the lower and upper limits for v. The two extreme cases, $v = 4n - 1$ and $v = n^2 + n + 1$, correspond to difference sets of the *Hadamard* type and to the difference sets associated with finite projective planes (where $\lambda = 1$) respectively. When $v = 4n - 1$ the difference set is said to be a Hadamard difference set, and when $v = n^2 + n + 1$ the term *planar* difference set is sometimes used.

The existence question for difference sets remains open, but by theorem 4.2, a difference set may be regarded as a special type of symmetric block design, and consequently the Bruck-Ryser-Chowla theorem (theorem 4.1) holds for difference sets. However, theorem 4.1 together with the independent condition of equation 4.9 are not sufficient to ensure the existence of at least one (v, k, λ) difference set. For example no $(49, 16, 5)$ difference set exists although these parameters satisfy both theorem 4.1 and equation 4.9.

Incidence Vector and Autocorrelation Functions

Each difference set D is associated with a binary $(1, 0)$ sequence $\{a_i\}$ $(i = 0, \ldots, v-1)$ given by $a_i = 1$ if $i \in D$ and $a_i = 0$ otherwise. This binary sequence is called the *characteristic*

function or the *incidence vector* of the difference set D. It is exactly equivalent to one column (or block) of the cyclic block design from which D originates.

One of the most important properties of difference sets and their incidence vector is the autocorrelation function $R_a(j)$:

$$R_a(j) = \sum_{i=0}^{v-1} a_i a_{i+j} \ (i+j \text{ taken modulo } v). \tag{4.16}$$

$R_a(j)$ can be easily evaluated by using equation 4.10 for symmetric block designs. Therefore

$$R_a(j) = \begin{cases} k & \text{if } j \equiv 0 (\text{mod} v) \\ \lambda & \text{if } j \not\equiv 0 (\text{mod} v) \end{cases} . \tag{4.17}$$

Often D is represented by a binary $(-1, 1)$ sequence $\{b_i\}$, where $b_i = 2a_i - 1$ replacing zeros by minus ones. The autocorrelation function $R_b(j)$ of this sequence is

$$R_b(j) = \sum_{i=0}^{v-1} b_i b_{i+j} = \begin{cases} v & \text{if } j \equiv 0 (\text{mod} v) \\ v - 4n & \text{if } j \not\equiv 0 (\text{mod} v) \end{cases} . \tag{4.18}$$

$(i + j$ taken modulo $v)$. As can be seen these difference set autocorrelation functions are two valued and hence have extensive applications in digital space communications [56], and in coded aperture imaging.

A third possible correlation function, a cross correlation, can be achieved by correlating the vectors $\{a_i\}$ and $\{b_i\}$ to give a correlation function $R_c(j)$ given by:

$$R_c(j) = \sum_{i=0}^{v-1} a_i b_{i+j} \ (i+j) \text{ taken modulo } v) \tag{4.19}$$

which has two values:

$$R_c(j) = \begin{cases} k & \text{if } j \equiv 0 (\text{mod} v) \\ 2\lambda - k & \text{if } j \not\equiv 0 (\text{mod} v) \end{cases} . \tag{4.20}$$

Since the detector flux incident on a photon detector in a coded aperture system cannot be negative then the two correlations applicable to coded aperture imaging are equations 4.16 and 4.19. The important quantity to consider is the "peak minus the mean background" (or the peak minus the plateau), P_g, that is the quantity:

$$P_g = R_g(0) - R_g(j), \ j \not\equiv 0 (\text{mod} v)$$

where R_g is the correlation function of type g. Therefore by 4.17 and 4.20:

$$P_a = R_a(0) - R_a(j) = k - \lambda = n \tag{4.21}$$

$$P_c = R_c(0) - R_c(j) = k - (2\lambda - k) = 2n. \tag{4.22}$$

Therefore, since n is always positive the cross correlation R_c gives a peak minus mean background which is twice that of the straight forward autocorrelation R_a, and a better coded aperture deconvolved image is obtained. Also, since $P_c = 2n$, then the best image is achieved when n is a maximum for a given v (i.e. for a given detector size). However by the inequality given in 4.15, $n \leq (v+1)/4$ and so $n = (v+1)/4$ represents the maximum value of n, and hence the best image. Therefore $v = 4n - 1$ and so the best kind of difference set to use in coded aperture imaging is a Hadamard type. Substituting for $n = (v + 1)/4$ into 4.9 and noting that $n = k - \lambda$ gives, for a Hadamard difference set, $k = (v \pm 1)/2$, which for $k < v/2$ gives $k = (v-1)/2$. Applying this and maximising the coded aperture transparency (to give the maximum flux incident on the detector) means that the residues which belong to the difference set correspond to *opaque* aperture elements. Note that if the value of 1 is taken for open aperture elements and 0 for closed aperture elements (i.e. the complement of the difference set is taken) then the off-axis cross correlation $R_c(j) = 2\lambda^* - k^*$, is in fact equal to zero in the case of a Hadamard difference set (as can be shown by 4.14 and 4.23 below). Therefore the cross correlation $R_c(j)$ is a perfect delta function with zero sidelobes for Hadamard difference sets.

Hadamard Difference Sets

As already mentioned, difference sets of the form $v = 4n - 1$, where $n = k - \lambda$ is an integer, are called Hadamard difference sets. Using this and equation 4.9 gives the following difference set parameters:

$$\left.\begin{aligned} v &= 4n - 1 \\ k &= 2n - 1 \quad = (v-1)/2 \\ \lambda &= n - 1 \quad = (v-3)/4 \end{aligned}\right\} \tag{4.23}$$

and any difference set with these parameters for any $n \geq 2$ is a Hadamard difference set.

These difference sets have been extensively studied for several reasons, the main ones being (i) their relative abundance, (ii) the fact that they represent the lower limit for v in equation 4.15 (and hence extreme value of λ) and (iii) the minimal cross correlation function of their incidence vectors, see above.

Any symmetric block design (not necessarily cyclic) with parameters $(v, k, \lambda) = (4n-1, 2n-1, n-1)$ is called a Hadamard design which, if the design is cyclic, gives rise to the Hadamard

difference set, which is a block of the design. If the 0's of the incidence matrix of this cyclic design are replaced by -1's then the resultant matrix is such that the inner product of any two rows is -1, as can be seen by equation 4.18. Therefore by adding a row and column of $+1$'s to this matrix, a new $(v+1)$ by $(v+1)$ matrix H of $+1$'s and -1's is constructed such that

$$HH^T = (v+1)I_{v+1}. \tag{4.24}$$

Any matrix which satisfies 4.24 is called a *Hadamard matrix* (since it achieves the upper bound for the determinant as given by Hadamard's determinant theorem, i.e. abs. val. det $H = (4n)^{2n}$ [157]) and the inner product of any two distinct rows is zero.

The known Hadamard difference sets can be classified according to the form of v. The three most important ones are:

(a) $v = 2^j - 1$, $j \geq 2$. Such sets correspond to *pseudo noise sequences* which are discussed in section 4.3.1.

(b) $v = 4n - 1$ is prime. These difference sets can be constructed from quadratic residues (section 4.3.2).

(c) $v = p(p+2)$ where p and $p+2$ are twin primes.

Other types of Hadamard difference sets have been studied and can be found in the literature. See for example [157,108,158,141] and [156].

A special type of Hadamard difference set is the skew Hadamard difference set. This is a Hadamard difference set which contains one and only one of the residues $a_i, v - a_i$ for $1 \leq a_i \leq v - 1$. Therefore the incidence matrix A of the corresponding cyclic block design obeys the following:

$$A + A^T = J_v - I_v. \tag{4.25}$$

This "skew symmetry" of these sets is very important in coded aperture imaging. It enables removal of systematic detector background by using a simple 180° rotation of the mask pattern with respect to the detection plane [50].

Skew Hadamard difference sets can be constructed by the quadratic residue method, method (b) above (see section 4.3.2 and [108]). Johnsen [72] has shown that this construction gives the complete set of skew Hadamard difference sets, as well as proving many other of their properties.

In what follows the above difference set construction methods (a) and (b) are described.

4.3 Constructions of Hadamard Difference Sets

4.3.1 Shift Register Sequences

General Discussion

A *shift register* is an arrangement of r little boxes or "tubes" each containing a number, with the contents of each tube moving to the next tube at regular intervals, or pulses. A *shift register sequence* is the sequence of numbers appearing in any one of the r tubes. After the first pulse the first tube will be empty, as eventually will all the other tubes after r pulses, at which time the shift register sequence will stop. The shift register can be kept going, however, by feeding the contents of some of the tubes back into the first tube. Such an arrangement is called a *feedback shift register*. An example of a feedback shift register is given in fig. 4.1, where the contents of each tube is either a 0 ("off") or a 1 ("on") and the contents of some of the tubes are fed through a mod2 adder and inserted back into the first tube (addition mod2 is defined by $1+1=0+0=0$, $1+0=0+1=1$). It can easily be reasoned that such a binary shift register sequence is periodic, since each state of the shift register is determined by its previous state. Therefore, any two states which are the same will have identical following states and, since there are only a finite number of states for an r tube shift register, then periodicity will be established. For an r tube shift register of 0's and 1's there are 2^r possible states. However if the state "all 0's" ever occurs then all subsequent states will remain this way and the period, p will be equal to 1. Therefore, for a "long" period shift register this "all 0's" state cannot occur, which means that the period p is given by

$$p \leq 2^r - 1 \tag{4.26}$$

Recurrence Relation

If we focus our attention on the first tube of the shift register and assume that its succession of states is given by the binary $(0,1)$ sequence $\{a_n\} = \{a_0, a_1, a_2, \ldots\}$. The nth state of the first tube, a_n, is determined by the contents of some of the other tubes at the $(n-1)$th

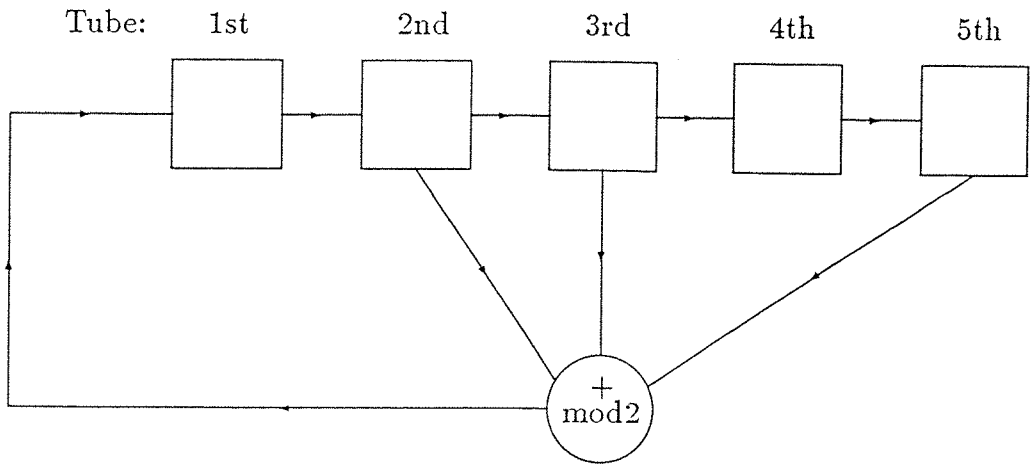


Fig. 4.1. *An example of a 5 tube shift register. Each term is shifted one place to the right in time with a clock pulse, with the new value of the first tube being the* mod2 *sum of the contents of tubes 2, 3, and 5 at the preceeding state.*

state, or alternatively by the contents of the first tube itself at some of the previous r states. That is, a_n is given by

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_r a_{n-r}$$

or

$$a_n = \sum_{i=1}^{r} c_i a_{n-i} \tag{4.27}$$

where the coefficients c_i are all 1's and 0's and are independent of n. Equation 4.27 is called the *linear recurrence relation* of the shift register sequence. An example of a shift register sequence and its recurrence relation is given later in fig. 4.2.

Generating Function and Characteristic Polynomial

Consider the shift register sequence $\{a_n\} = \{a_0, a_1, a_2, \ldots\}$ which describes the succession of states of the first tube. Then we may associate this sequence with its *generating function* $G(x)$, given by

$$G(x) = \sum_{n=0}^{\infty} a_n x^n. \tag{4.28}$$

Let the initial conditions of the shift register be

$$a_{-1}, a_{-2}, \ldots, a_{-r}. \tag{4.29}$$

Therefore, by 4.27 and 4.28:

$$\begin{aligned} G(x) &= \sum_{n=0}^{\infty} \sum_{i=1}^{r} c_i a_{n-i} x^n \\ &= \sum_{i=1}^{r} c_i x^i [a_i x^i + \cdots + a_{-1} x^{-1} + \sum_{n=0}^{\infty} a_n x^n] \end{aligned}$$

which, by 4.28, gives

$$G(x) = \frac{\sum_{i=1}^{r} c_i x^i (a_{-i} x^{-i} + \cdots + a_{-1} x^{-1})}{1 - \sum_{i=1}^{r} c_i x^i} \tag{4.30}$$

expressing $G(x)$ in terms of the initial conditions 4.29 and the feedback coefficients c_i only. Expressing 4.30 in another way:

$$G(x) = \frac{g(x)}{f(x)} \tag{4.31}$$

where

$$f(x) = 1 - \sum_{i=1}^{r} c_i x^i. \tag{4.32}$$

$f(x)$ is a very important quantity in generating shift register sequences, and is called the *characteristic polynomial* of the shift register sequence $\{a_n\}$, and is independent of the initial conditions. Note that $f(x) = 1 - \sum x^j$ where the sum is taken over the values of j for which the jth tube feeds back into the mod2 adder.

Period

The numerator in 4.30, which is $g(x)$, does depend on the initial conditions. By 4.30 and 4.31:

$$g(x) = \sum_{i=1}^{r} c_i x^i (a_{-i}x^{-i} + \cdots + a_{-1}x^{-1}). \tag{4.33}$$

This equation, along with 4.31 and 4.32 is very important in determining exactly the period of the shift register sequence, an upper bound to which was given in 4.26. This is shown in the following theorem.

Theorem 4.3 *If the characteristic polynomial $f(x)$ of an r tube shift register sequence $\{a_n\}$ has no factors in common with $g(x)$, then the period is the smallest positive integer p for which $f(x)$ divides $1 - x^p$.*

Proof. The fact that shift register sequences are periodic has already been established with 4.26 giving an upper limit to p. Now, by 4.28 and 4.31:

$$G(x) = \frac{g(x)}{f(x)} = \sum_{n=0}^{\infty} a_n x^n$$

$$\begin{aligned} = \quad & (a_0 + a_1x + \cdots + a_{p-1}x^{p-1}) \\ + \quad & x^p(a_0 + a_1x + \cdots + a_{p-1}x^{p-1}) \\ + \quad & x^{2p}(a_0 + a_1x + \cdots + a_{p-1}x^{p-1}) \\ + \quad & \cdots \end{aligned}$$

$$= (a_0 + a_1x + \cdots + a_{p-1}x^{p-1})(1 + x^p + x^{2p} + \cdots)$$

$$= \frac{a_0 + a_1x + \cdots + a_{p-1}x^{p-1}}{1 - x^p}$$

which gives

$$f(x)(a_0 + a_1x + \cdots + a_{p-1}x^{p-1}) = g(x)(1 - x^p).$$

Therefore if $f(x)$ has no factors in common with $g(x)$ then $f(x)$ divides $1 - x^p$.

One method of assuring that $f(x)$ and $g(x)$ have no common factors is to assume the initial conditions $a_{-1} = a_{-2} = \ldots = a_{1-r} = 0$, $a_{-r} = 1$, whereby $g(x) = 1$ (since c_r must equal 1 if the shift register is making use of the fact that it has r tubes) and no common factors exist. Another very important case is when $f(x)$ is *irreducible* (i.e. it can not be factored) in which case it can have no factors in common with $g(x)$ (unless $g(x) = 0$ which corresponds to the initial condition "all 0's"). In this case the period p does not depend on the initial conditions (excepting only the initial condition "all 0's").

Maximal Length Shift Registers

Equation 4.26 gives an upper bound for the period of an r tube shift register sequence. If this upper bound actually occurs and $p = 2^r - 1$ then the sequence $\{a_n\}$ is said to be of *maximum length*, and the shift register which produced it is called a *maximal length shift register*. Such sequences are also known as *pseudo noise sequences* (PN sequences) due to the similarity of the properties of these sequences with binary $(0, 1)$ random sequences.

A necessary condition for a shift register sequence to be of maximum length is that its characteristic polynomial $f(x)$ must be *irreducible* as is shown by the following theorem.

Theorem 4.4 *If a shift register sequence has maximum length then its characteristic polynomial $f(x)$ is irreducible.*

Proof. Let $A = \{a_n\}$ be the maximal length shift register sequence. Then A runs through all $2^r - 1$ non zero terms before repeating, and so any of these terms can be taken as the initial conditions. Therefore taking the initial conditions to be $a_{-1} = a_{-2} = \ldots = a_{1-r} = 0$, $a_{-r} = 1$ gives $g(x) = 1$ and $G(x) = 1/f(x)$ which has periodic coefficients a_n with period p. Now, assume that $f(x)$ is factorable: $f(x) = s(x)t(x)$ where $s(x)$ and $t(x)$ have degrees in x of $r_s > 0$ and $r_t > 0$ respectively, where $r_s + r_t = r$. By partial fractions, $1/f(x) = \alpha(x)/s(x) + \beta(x)/t(x)$. $\alpha(x)/s(x)$ is now a power series having periodic coefficients with period $p_s \leq 2^{r_s} - 1$, and $\beta(x)/t(x)$ is also a power series having periodic coefficients of period $p_t \leq 2^{r_t} - 1$. Therefore $1/f(x) = \alpha(x)/s(x) + \beta(x)/t(x)$ represents a power series with coefficients having period p at most the least common multiple of p_s and p_t which in

turn cannot exceed $p_s p_t$. Therefore

$$\begin{aligned} p \leq p_s p_t &\leq (2^{r_s}-1)(2^{r_t}-1) \\ &= 2^{r_s+r_t} - 2^{r_s} - 2^{r_t} + 1 \\ &= 2^r - 2^{r_s} - 2^{r_t} + 1 \\ &\leq 2^r - 3 \end{aligned}$$

giving

$$p = 2^r - 1 \leq 2^r - 3.$$

This is, of course, impossible and the resulting contradiction shows that the original assumption that $f(x)$ is factorable was a false one. Therefore $f(x)$ must be irreducible if maximum length is to be achieved.

However, irreducible characteristic polynomials do not necessarily lead to maximum length sequences. For example $f(x) = 1 + x + x^2 + x^3 + x^4$ is irreducible, but since it divides $1 - x^5$ then it has a period $p = 5$ due to theorem 4.3, rather than $2^4 - 1 = 15$. Therefore the irreducible characteristic polynomial is shown to be a necessary condition for maximum length by theorem 4.4, but is an insufficient one.

An r tube shift register sequence can be thought of as a binary $r-$digit number ranging from 1 to $2^r - 1$ (if the "all 0's" case is not included). In the "units digit" of this binary number, 1 corresponds to an odd number and 0 to an even number. From 1 to $2^r - 1$ there are 2^{r-1} odd numbers and $2^{r-1} - 1$ even numbers. Therefore in any maximum length sequence there are 2^{r-1} 1's and $2^{r-1} - 1$ 0's.

Autocorrelation of Maximum Length Sequences

The most important quantity which is associated with maximal length shift register or pseudo noise sequences is their autocorrelation function. In order to evaluate the autocorrelation function however, it must first be observed that two pseudo noise sequences, with the same period and recurrence relation, when added together term by term modulo 2, give the same sequence but with a simple phase shift. This is seen as follows. If $A = \{a_0, a_1, a_2, \ldots\}$ and $B = \{b_0, b_1, b_2, \ldots\}$ are two pseudo noise sequences of the same period and recurrence relation then by 4.27:

$$a_n = \sum_{i=1}^{r} c_i a_{n-i}$$

$$b_n = \sum_{i=1}^{r} c_i b_{n-i}$$

which are the two recurrence relations, but both having the same coefficients c_i. Therefore, adding the two terms:

$$\begin{aligned} a_n + b_n &= \sum_{i=1}^{r} c_i a_{n-i} + \sum_{i=1}^{r} c_i b_{n-i} \\ &= \sum_{i=1}^{r} c_i (a_{n-i} + b_{n-i}) (\mathrm{mod} 2) \end{aligned}$$

and so the recurrence relation still holds for $A + B$ as well as just for A and B. If the initial conditions of A and B are $\{a_{-1}, a_{-2}, \ldots, a_{-r}\}$ and $\{b_{-1}, b_{-2}, \ldots, b_{-r}\}$ respectively then the initial conditions of $A+B$ are $\{a_{-1}+b_{-1}, a_{-2}+b_{-2}, \ldots, a_{-r}+b_{-r}\}\,(\mathrm{mod}\,2)$. Now, whatever these r terms are, they are the same as the r terms of exactly one of the $2^r - 1$ non zero sequences of length r, which is equivalent to simply using different initial conditions for the shift register than for those used to generate A and B. This, combined with the same recurrence relation, gives the same pseudo noise sequence but with a phase shift. We can now prove the following very important theorem about pseudo noise sequences, which is done via the autocorrelation function.

Theorem 4.5 *Every pseudo noise sequence is the complement of a Hadamard difference set.*

Proof. Let $\{a_j\}$ be a pseudo noise sequence with period p. Then $\{a_j + a_{j+\tau}\}(\mathrm{mod}2)$ is a pseudo noise sequence containing $2^{r-1} = (p+1)/2$ 1's and $2^{r-1} - 1 = (p-1)/2$ 0's in every period, where τ is a phase shift and $j + \tau$ is taken modulo p. Now, assume that $a_j + a_{j+\tau}$ occurs x times in the form $1 + 1 = 0$. Then the other $2^{r-1} - 1 - x$ 0's from $\{a_j + a_{j+\tau}\}$ must occur as $0 + 0 = 0$. There are now $2^{r-1} - x$ 1's in the first term which must occur as $1 + 0 = 1$ and likewise $2^{r-1} - x$ 1's in the second term occurring as $0 + 1 = 1$. Altogether there are $p = 2^r - 1$ terms in the sequence $\{a_j + a_{j+\tau}\}$. Therefore

$$p = 2^r - 1 = (x) + (2^{r-1} - 1 - x) + (2^{r-1} - x) + (2^{r-1} - x)$$

which gives $x = 2^{r-2}$. Therefore the number of 1-1 pairs in $\{a_j\}$ and $\{a_{j+\tau}\}$ is $x = 2^{r-2}$, which is a constant, independent of τ for $\tau \neq 0$. Also, if $\tau = 0$ then the number of 1-1 pairs is the same as the number of 1's in the sequence, i.e. 2^{r-1}. Therefore the autocorrelation function $R_a(\tau)$ is two valued, as in 4.17 for difference sets:

$$R_a(\tau) = \sum_{j=1}^{p} a_j a_{j+\tau} = \begin{cases} 2^{r-1} = k^* & \text{if } \tau \equiv 0 (\mathrm{mod} p) \\ 2^{r-2} = \lambda^* & \text{if } \tau \not\equiv 0 (\mathrm{mod} p) \end{cases} \tag{4.34}$$

($j + \tau$ taken modulo p). Therefore, comparing with 4.17, pseudo noise sequences are in fact (v^*, k^*, λ^*) difference sets, where $v^* = p = 2^r - 1$, $k^* = 2^{r-1} = (v^* + 1)/2$, $\lambda^* = 2^{r-2} = (v^* + 1)/4$. Now, if the complement of this difference set is taken, a (v, k, λ) difference set is obtained with parameters $v = v^*$, $k = (v - 1)/2$, $\lambda = (v - 3)/4$, which are the parameters of a Hadamard difference set (equation 4.23). Thus a pseudo noise sequence, or a maximal length shift register sequence, is the complement of a Hadamard difference set.

Construction of Hadamard Difference Sets of the Form $v = 2^r - 1$

The construction of Hadamard difference sets of the form $v = 2^r - 1$ from shift register sequences follows immediately from theorem 4.4, which is a theorem on maximum length sequences stating that maximum length is achieved by the use of an irreducible polynomial. Examples of irreducible polynomials can be found in [66,91] and [140], and much higher order polynomials can be found in [166] and [167].

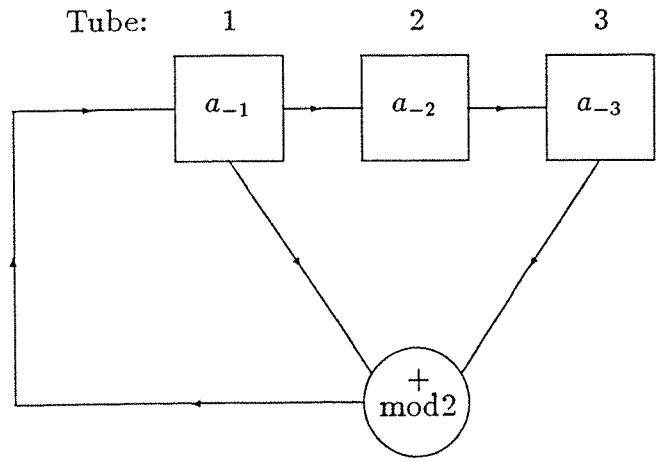
The basic method can be seen by following the example below which generates the maximum length sequence with $r = 3$ and period $p = 2^3 - 1 = 7$. The corresponding irreducible polynomial is $f(x) = 1+x+x^3$ (note that by 4.32 $f(x)$ should be written as $f(x) = 1-x-x^3$, but since all mathematical operations on the sequences are modulo 2, it is possible to change all the minus signs to plus signs since $1 - h(x) \equiv 1 + h(x) \,(\mathrm{mod}\, 2)$). The recurrence relation coefficients c_i for this $f(x)$ are, by 4.32, $c_1 = c_3 = 1$, $c_2 = 0$, giving the recurrence relation (by 4.27 modulo 2) of

$$a_n = a_{n-1} + a_{n-3}.$$

Therefore the shift register required is the one in fig. 4.2 (a), with the first and third tubes feeding into the mod2 adder before being inserted back into the first tube. Because $f(x)$ is irreducible the initial conditions chosen do not effect the period (so long as the condition $\{0, 0, 0\}$ is not chosen). For the initial condition $\{1, 1, 1\}$ the succession of states for the shift register is that shown in fig. 4.2 (b). The eighth state is the same as the first state giving a period $p = 7$. Every non zero binary 3-digit number from 1 to 7 is given in the succession of states. The succession of states of the first tube $\{a_0, a_1, \ldots, a_6\}$ is the pseudo noise sequence $\{0, 1, 0, 0, 1, 1, 1\}$ which, taking its complement, gives the Hadamard difference set $\{1, 0, 1, 1, 0, 0, 0\}$ having parameters $v = p = 7, k = (v - 1)/2 = 3, \lambda = (v - 3)/4 = 1$. Note that this is the same as one block of the symmetric $(7, 3, 1)$ block design in appendix A (with a phase shift). Constructions of maximal length shift register sequences are also

(a)

Recurrence relation: $a_n \equiv a_{n-1} + a_{n-3} \pmod 2$



(b)

State			
−1	1	1	1
0	0	1	1
1	1	0	1
2	0	1	0
3	0	0	1
4	1	0	0
5	1	1	0
6	1	1	1

Fig. 4.2. *(a) Shift register corresponding to the characteristic polynomial* $f(x) = 1+x+x^3$. *(b) Succession of states of the shift register.*

shown in [66] and [91].

4.3.2 Quadratic Residue Construction of Hadamard Difference Sets

Congruence

Let a and b be integers, and m an integer greater than zero. If m divides $a - b$ (written $m \mid a - b$) then a and b are said to be *congruent modulo* m. This is written $a \equiv b(\mathrm{mod} m)$. The integer m is called the *modulus*.

The arithmetic of congruence is as follows:

1. If $a \equiv b(\mathrm{mod} m)$ and $c \equiv d(\mathrm{mod} m)$ then $a + c \equiv b + d(\mathrm{mod} m)$.
2. If $a \equiv b(\mathrm{mod} m)$ and k is an integer then $ka \equiv kb(\mathrm{mod} m)$.
3. If $a \equiv b(\mathrm{mod} m)$ and $c \equiv d(\mathrm{mod} m)$ then $ac \equiv bd(\mathrm{mod} m)$.
4. If $ka \equiv kb(\mathrm{mod}\, m)$ and the greatest common divisor of k and m is 1 (written $(k, m) = 1$) then $a \equiv b(\mathrm{mod} m)$.

If M_m is the set of the multiples of m (e.g. $M_7 = \{0, 7, 14, 21, \ldots\}$), then the sets

$$M_m + \{r\} = \{qm + r : q = 0, \pm 1, \pm 2, \ldots\}$$

are called the *residue classes* mod m. The elements of these sets are called the *residues* mod$m$. If an integer a belongs to the set $M_m + \{r\}$, where $0 \leq r \leq m - 1$, then r is the least non negative residue congruent to $a(\mathrm{mod} m)$.

Complete and Reduced Residue Systems

A set of m integers that consists of exactly one integer from each of the residue classes mod m is called a *complete residue system* mod m. The set {0,1,2,...,m-1} is the least non negative residue system mod$m$. A *reduced residue system* mod$m$ is a set of integers consisting of exactly one integer from each residue class mod$m$ whose elements are relatively prime to m, i.e. they have no factors in common except for 1, i.e. $(a, m) = 1$. For example if

$m = 12$, then there are four residue classes mod 12 which consist of integers relatively prime to 12, e.g. $\{1, 5, 7, 11\}$ and $\{5, 25, 35, 55\}$ are reduced residue systems mod 12. The number of elements in a reduced residue system mod m is designated by $\phi(m)$, a quantity known as *Euler's ϕ-function.* $\phi(m)$ is therefore also equal to the number of integers in the set $\{1, 2, 3, \ldots, m-1, m\}$ that are relatively prime to m. In the above example therefore, $\phi(12) = 4$. If m is a prime number $m = p$, then $(a, p) = 1$ for every a which has $1 \leq a \leq p-1$, and so

$$\phi(p) = p - 1. \tag{4.35}$$

The Euler Fermat Theorem

A very important theorem in number theory is the Euler Fermat Theorem which is proved below. However, the following theorem on reduced residue systems is first required.

Theorem 4.6 *If $\{r_1, r_2, \ldots, r_{\phi(m)}\}$ is a reduced residue system* mod$m$ *and a is an integer such that $(a, m) = 1$, then the set of integers $\{ar_1, ar_2, \ldots, ar_{\phi(m)}\}$ is also a reduced residue system* mod$m$.

Proof. If a and r_i are relatively prime to m, then so must be ar_i. Also, if $ar_i \equiv ar_j (\bmod m)$ then, since $(a, m) = 1$, $r_i \equiv r_j (\bmod m)$ by the arithmetic of congruence. Therefore, if r_i and r_j are distinct then so must be ar_i and ar_j. Thus the set $\{ar_1, ar_2, \ldots, ar_{\phi(m)}\}$ consists of $\phi(m)$ integers from distinct residue classes, each of which are relatively prime to m. Thus the set $\{ar_1, ar_2, \ldots, ar_{\phi(m)}\}$ constitutes a reduced residue system mod$m$.

The Euler Fermat theorem can now be proved.

Theorem 4.7 (The Euler Fermat Theorem) *Let a be an integer such that $(a, m) = 1$. Then:*

$$a^{\phi(m)} \equiv 1 (\bmod m)$$

Proof. Let $S = \{r_1, r_2, \ldots, r_{\phi(m)}\}$ be a reduced residue system mod$m$. Then the set $T = \{ar_1, ar_2, \ldots, ar_{\phi(m)}\}$ is also a reduced residue system mod$m$ by theorem 4.6 since $(a, m) = 1$. This means that each element of S is congruent to an element of $T (\bmod m)$,

and each element of T is is congruent to an element of $S(\mathrm{mod}m)$. Therefore

$$ar_1.ar_2.\ldots.ar_{\phi(m)} \equiv r_1 r_2 \ldots r_{\phi(m)}(\mathrm{mod}m)$$

i.e.

$$a^{\phi(m)}(r_1 r_2 \ldots r_{\phi(m)}) \equiv (r_1 r_2 \ldots r_{\phi(m)})(\mathrm{mod}m)$$

which, by the division algorithm (algorithm 4) of the arithmetic of congruence for $(r_i, m) = 1$ gives the Euler Fermat Theorem:

$$a^{\phi(m)} \equiv 1(\mathrm{mod}m). \tag{4.36}$$

A corollary to this theorem is Fermat's theorem:

Corollary 4.8 (Fermat's Theorem) *If $m = p$ is a prime and a any integer then:*

$$a^p \equiv a(\mathrm{mod}p)$$

or:

$$a^{p-1} \equiv 1(\mathrm{mod}p). \tag{4.37}$$

This is easily proven by theorem 4.7 and equation 4.35.

The necessary mathematics is now complete for the study of quadratic residues with a view to the construction of Hadamard difference sets and Hadamard matrices.

Quadratic Residues. Definition and Legendre Symbol

If the congruence

$$x^2 \equiv a(\mathrm{mod}m)$$

has a solution for integer x then a is said to be a *quadratic residue modulo m.* If there is no solution for x to this congruence then a is said to be a *quadratic non residue modulo m.* A very useful notational device used in studying quadratic residues is the *Legendre symbol*, (a/p), where p is an odd prime, defined by:

$(a/p) = 1$ if a is a quadratic residue mod$p$.

$(a/p) = -1$ if a is a quadratic non residue mod$p$.

$(a/p) = 0$ if a is a multiple of p.

It is possible to evaluate (a/p) for any integer a and prime modulus p as follows. By Fermat's theorem (4.8) for $p \nmid a$ (i.e. p not dividing a):

$$a^{p-1} - 1 \equiv 0(\mathrm{mod}p)$$

which factorising gives

$$(a^{(p-1)/2} - 1)(a^{(p-1)/2} + 1) \equiv 0(\mathrm{mod}p)$$

giving two solutions for $a^{(p-1)/2}$:

$$a^{(p-1)/2} \equiv 1(\mathrm{mod}p) \tag{4.38}$$

or:

$$a^{(p-1)/2} \equiv -1(\mathrm{mod}p). \tag{4.39}$$

Now, assume that a is a quadratic residue mod$p$, then $(a/p) = 1$ and

$$x^2 \equiv a(\mathrm{mod}p)$$

has a solution. Raising both sides to the power $(p-1)/2$ (which is possible since $2 \mid p-1$ for any odd prime p) gives

$$a^{(p-1)/2} \equiv x^{p-1}(\mathrm{mod}p)$$

which by Fermat's theorem gives

$$a^{(p-1)/2} \equiv 1(\mathrm{mod}p)$$

which is 4.38. Therefore 4.38 holds for $(a/p) = 1$. It can also be shown that 4.39 holds if a is a quadratic non residue mod$p$, i.e. $(a/p) = -1$ ([75] chapter 25). Therefore we have the following theorem known as *Euler's criterion for the Legendre symbol.*

Theorem 4.9 (Euler's Criterion)

$$(a/p) \equiv a^{(p-1)/2}(\mathrm{mod}p).$$

Note that Euler's criterion holds also if $p \mid a$ and $(a/p) = 0$.

Considering the Legendre symbol for $a = -1$ gives an important corollary to Euler's criterion:

$$(-1/p) \equiv (-1)^{(p-1)/2} (\mathrm{mod} p)$$

and so:

Corollary 4.10 *-1 is a quadratic residue of primes of the form $p = 4n - 3$ (n being an integer) and -1 is a quadratic non residue of primes of the form $p = 4n - 1$.*

An immediately obvious property of the Legendre symbol is that if $a \equiv b (\mathrm{mod} p)$ then $(a/p) = (b/p)$. This is obvious since a and b are from the same residue class mod$p$. The Legendre symbol is multiplicative, as is shown by using Euler's criterion:

Theorem 4.11

$$(a/p)(b/p) = (ab/p)$$

Proof. By Euler's criterion

$$\begin{aligned} (a/p)(b/p) &\equiv a^{(p-1)/2} b^{(p-1)/2} (\mathrm{mod} p) \\ &\equiv (ab)^{(p-1)/2} (\mathrm{mod} p) \\ &\equiv (ab/p). \end{aligned}$$

In other words theorem 4.11 states that the product of two quadratic residues or of two quadratic non residues is a quadratic residue, and the product of a quadratic residue and a quadratic non residue is a quadratic nonresidue.

Autocorrelation of Quadratic Residues

To find the quadratic residues mod$p$ it is enough to consider the squares of the numbers from 1 to $p - 1$. In fact since $(p - a)^2 \equiv (-a)^2 \equiv a^2 (\mathrm{mod} p)$ it is enough to consider the squares

$$1^2, 2^2, \ldots, \left(\frac{p-1}{2}\right)^2.$$

These are all distinct, for if $i^2 \equiv j^2 (\bmod p)$ with $1 \leq i \leq (p-1)/2$ and $1 \leq j \leq (p-1)/2$, then p divides $(i+j)(i-j)$ which is only possible if $i = j$. Therefore there are $(p-1)/2$ quadratic residues mod$p$. The $(p-1)/2$ remaining non zero residues mod$p$ are the quadratic non residues. Zero is neither a quadratic residue nor a quadratic non residue, and $(0/p) = 0$. Using these facts and Euler's criterion gives

$$\sum_{a=0}^{p-1}(a/p) = 0. \tag{4.40}$$

The autocorrelation function $R_q(b)$ for quadratic residues is as follows:

$$R_q(b) = \sum_{a=0}^{p-1}(a/p)(a+b/p)$$

where $b \not\equiv 0(\bmod p)$, which by Euler's criterion gives

$$\begin{aligned} R_q(b) &\equiv \sum_{a=0}^{p-1} a^{(p-1)/2}(a+b)^{(p-1)/2}(\bmod p) \\ &\equiv \sum_{a=0}^{p-1} a^{(p-1)}(\bmod p) + \sum_{a=0}^{p-1}(ab)^{(p-1)/2}(\bmod p). \end{aligned}$$

Now, using Fermat's theorem (corollary 4.8), $a^{p-1} \equiv 1(\bmod p)$ for all $a \not\equiv 0(\bmod p)$, means that the first term on the right hand side is equal to $p-1$. The second term is simply a summation of the Legendre symbols of all of the elements of a reduced residue system mod p, as shown by theorem 4.6. Therefore $\sum_{a=0}^{p-1}(ab)^{(p-1)/2} \equiv \sum_{a=0}^{p-1} a^{(p-1)/2} \equiv \sum_{a=0}^{p-1}(a/p)(\bmod p)$. Thus:

$$R_q(b) \equiv \left[p - 1 + \sum_{a=0}^{p-1}(a/p)\right](\bmod p)$$

which by 4.40 gives

$$R_q(b) \equiv -1(\bmod p). \tag{4.41}$$

Now, $R_q(b)$ is a summation over p terms, two of which are zero (corresponding to when $(a/p) = 0$ and when $(a + b/p) = 0$), whilst the rest have a magnitude of 1. Therefore the magnitude of $R_q(b)$ cannot exceed $p-2$. This, along with 4.41 gives $R_q(b) = -1$ when $b \not\equiv 0(\bmod p)$. Note also that when $b \equiv 0(\bmod p)$, $R_q(0) = \sum_{a=0}^{p-1}(a/p)^2 = p-1$. Therefore

$$R_q(b) = \sum_{a=0}^{p-1}(a/p)(a+b/p) = \begin{cases} p-1 & \text{if } b \equiv 0(\bmod p) \\ -1 & \text{if } b \not\equiv 0(\bmod p) \end{cases}. \tag{4.42}$$

This two valued autocorrelation function is similar to that of difference sets (see section 4.2.2, equation 4.18), and is therefore of value in coded aperture imaging via the construction of Hadamard matrices, which follows.

Quadratic Residue Construction of Hadamard Matrices

The following construction, due to Paley [108], gives Hadamard matrices of any order $N = p+1 = 4n$ (where p is an odd prime and n an integer) and therefore of Hadamard difference sets of the form $v = p = 4n - 1$. It is known as the *Paley construction*, and can be found in [108] and also in [92] chapter 2.

First choose a prime p of the form $p = 4n - 1$ and construct the $p \times p$ matrix $Q = (q_{ij})$ with rows and columns labelled $0, 1, \ldots, p-1$. Denote (a/p) by $\chi(a)$ and then let $q_{ij} = \chi(j-i)$. For example, see fig. 4.3 for $p = 7$. Since p is of the form $4n-1$ then -1 is a quadratic non residue modulo p (corollary 4.10), and hence by theorem 4.11, $\chi(j-i) = \chi(-1)\chi(i-j) = -\chi(i-j)$, which means that Q is *skew symmetric*, i.e. $Q^T = -Q$. Now, consider the following:

$$P = (p_{ij}) = QQ^T$$

where:

$$p_{ii} = \sum_{k=0}^{p-1} q_{ik}^2 = p - 1 \tag{4.43}$$

and for $i \neq j$:

$$\begin{aligned} p_{ij} &= \sum_{k=0}^{p-1} q_{ik}q_{jk} \\ &= \sum_{k=0}^{p-1} \chi(k-i)\chi(k-j). \end{aligned}$$

Now, writing $a = k - i$ and $b = i - j$ we have

$$p_{ij} = \sum_{a=0}^{p-1} \chi(a)\chi(a+b)$$

which, by the quadratic residue autocorrelation function (equation 4.42) gives

$$p_{ij} = -1 \text{ for } i \neq j. \tag{4.44}$$

Therefore, by 4.43 and 4.44:

$$P = QQ^T = pI_p - J_p. \tag{4.45}$$

Another (obvious) relationship for Q is

$$QJ_p = J_pQ = 0 \tag{4.46}$$

since each row and column of Q contains $(p-1)/2$ $+1$'s and $(p-1)/2$ -1's, (see equation 4.40). Now, let

$$H = \begin{bmatrix} 1 & w_p \\ w_p^T & Q - I_p \end{bmatrix} \tag{4.47}$$

$$
Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\left[\begin{array}{ccccccc}
0 & 1 & 1 & -1 & 1 & -1 & -1 \\
-1 & 0 & 1 & 1 & -1 & 1 & -1 \\
-1 & -1 & 0 & 1 & 1 & -1 & 1 \\
1 & -1 & -1 & 0 & 1 & 1 & -1 \\
-1 & 1 & -1 & -1 & 0 & 1 & 1 \\
1 & -1 & 1 & -1 & -1 & 0 & 1 \\
1 & 1 & -1 & 1 & -1 & -1 & 0
\end{array}\right]
\end{array}
$$

Fig. 4.3. *The $p \times p$ matrix $Q = (q_{ij})$. $q_{ij} = \chi(j - i)$.*

where w_p is the vector of p 1's. Thus

$$\begin{aligned} HH^T &= \begin{bmatrix} 1 & w_p \\ w_p^T & Q - I_p \end{bmatrix} \begin{bmatrix} 1 & w_p \\ w_p^T & Q^T - I_p \end{bmatrix} \\ &= \begin{bmatrix} 1 & z_p \\ z_p^T & J_p + (Q - I_p)(Q^T - I_p) \end{bmatrix} \end{aligned}$$

where z_p is the vector of p 0's. Thus by 4.45

$$HH^T = (p+1)I_{p+1}. \tag{4.48}$$

Thus, the matrix H given in equation 4.47 is a Hadamard matrix of order $p + 1 = 4n$.

The corresponding difference set is called a *skew Hadamard difference set* and is simply either the set of quadratic residues or quadratic non residues modulo a prime p of the form $p = 4n - 1$. For more information about this kind of difference set see Johnsen [72]. Such skew symmetry is useful for detector background subtraction in coded aperture imaging.

Chapter 5

Imaging Using a 19 Element HURA Coded Aperture with Discrete Pixel Detector Arrays.

5.1 Gamma Ray Telescope Design. Problems to Overcome

In order to carry out gamma ray astronomical observations, great difficulties must be overcome. Firstly the interaction process of gamma ray photons with detector material is mostly by the Compton process. Not all of the photon's original energy is measured by the detector system and so many corrections are required in order to obtain a deconvolved photon spectrum from the photon energy loss spectrum. Secondly there is great difficulty in designing detectors which are well shielded from the high background noise due to particle fluxes experienced in the space environment in which a gamma ray telescope must function. Thirdly the position of the photon interaction has to be accurately determined in order to provide a high source angular resolution, a particularly important requirement when observing crowded source regions such as the Galactic centre. Finally the expected photon fluxes from celestial sources are very low compared to the background. This is an important consideration in view of an effective limit to the detector area imposed by weight constraints

of satellite and balloon vehicles.

The minimum source flux $\phi(E)$ that can be detected by a gamma ray telescope to a statistical significance of n standard deviations is given by

$$\phi(E) = \frac{n}{\epsilon(E)} \left(\frac{2B(E)}{AT\Delta E} \right)^{1/2} \text{ photons cm}^{-2} \text{ s}^{-1} \text{ keV}^{-1} \tag{5.1}$$

where $\epsilon(E)$ is the detector efficiency, $B(E)$ is the observed detector background level (counts cm^{-2} s^{-1} keV^{-1}), A is the useful detector area (cm^2), T is the observation time (s), and ΔE is the bandwidth of the detector (keV). Therefore, from a hardware point of view the minimum source flux that can be detected depends on three main parameters: a) the detector efficiency, b) the sensitive detector area, and c) the level of the background noise in the detector.

The main considerations when designing a telescope, therefore, are to optimise these three parameters to give the best sensitivity to weak sources; i.e. to maximise a) and b) and to keep c) as low as possible. In particular the elimination of any spatially varying systematic detector background noise is of great importance, one method being to make use of HURAs which are antisymmetric upon 60° rotation. This possibility has been studied by Cook et al. [35]. There are three main sources of background in gamma ray telescopes. These are photon induced background, neutron induced background and background from cosmic rays. The photon induced background has two main components. The first is the diffuse cosmic background which has been measured by the Apollo 16 and 17 missions [148] and by Schönfelder et al. [131], who calculates an energy spectrum of

$$\frac{dN}{dE} = 11\, E^{-2.3} \text{ photons cm}^{-2} \text{ s}^{-1} \text{ steradian}^{-1} \text{ MeV}^{-1} \tag{5.2}$$

derived from emission in the earth's atmosphere. The second component, which is dominant in high altitude balloon flights, originates from the decay of neutral pi mesons (above 50 MeV) or from the bremsstrahlung of relativistic electrons. The energy spectrum calculated is

$$\frac{dN}{dE} = 21\, E^{-1.65} \text{ photons cm}^{-2} \text{ s}^{-1} \text{ steradian}^{-1} \text{ keV}^{-1} \tag{5.3}$$

[132]. The neutron induced background arises from neutrons which, after being produced by the interaction of cosmic rays with atmospheric nuclei, go on to interact with nuclei in the detector material producing gamma rays or charged particles. These particles then interact with the detector to register an event count contributing to the background. Possible intensity values have been calculated by Archer [6].

Finally the cosmic ray induced background is produced by high energy cosmic ray protons interacting with the nuclei of the detector material which subsequently undergo radioactive decay. The decaying nuclides emit high energy photons which, if detected in another region of the detector, are registered as background events. Such reactions have been described by Rudstam [125] and possible background have been estimated by Baker et al. [10].

The design of a gamma ray telescope must take into account the above sources of background noise which, in the past, have proven to be formidable obstacles for the gamma ray astronomer to overcome. Although all of these problems have been considered in the case of the GRASP telescope design (see [58]), the introduction of further sources of noise which may occur in the form of an "intrinsic noise" as a result of the complex GRASP detector plane will be most unwelcome. Therefore a crucial aspect of the development of GRASP is the need to investigate the imaging performance of HURA coded apertures and discrete pixel detector arrays, and this is the subject matter which comprises this and the next chapter.

5.2 Imaging Using a 19 Element HURA Coded Aperture with Discrete Pixel Detector Arrays

The GRASP gamma ray astronomy telescope is intended to be an instrument capable of performing both imaging and spectroscopy measurements. Consequently the detector plane is to be constructed using two main types of detector:

(i) An array of germanium spectrometers situated at the centre of the detector having low event position resolution.

(ii) Caesium iodide (CsI) imager bars, having fine event position resolution, placed around the germanium spectrometers.

The fact that the two detector types are different and need to be coupled together for image deconvolution, plus the fact that the CsI detectors will be arranged as an array of bars, means that it is necessary to study the imaging properties of a discrete pixel detector. Also, for satellite gamma ray imaging experiments, HURA coded apertures are particularly suitable, one reason for this being that their basic patterns, and hence the detector plane, are almost circular. This arrangement is compatible with the cylindrical geometries of rocket systems.

Another reason is their antisymmetry on 60° rotation enabling subtraction of systematic detector background noise and improving image quality [35].

The design of the GRASP telescope, therefore, required a study of the imaging performance of HURA coded apertures with discrete pixel detector arrays.

5.3 Imaging Requirements

5.3.1 Imaging System Configurations and Parameters

Monte Carlo simulation programs have been developed and used to study the basic imaging performance of an ideal coded aperture system consisting of a HURA aperture, or *mask*, and a perfect discrete pixel detector plane. The mask used is shown in fig. 5.1. It has a basic pattern of $v = 19$ elements and it has 217 elements in total. The mask elements have been placed on a hexagonal lattice with unit vectors $\underline{e}_0$ and $\underline{e}_1$ at 60° to each other, constructed as given in section 3.3.5, and is antisymmetric upon a 60° rotation. This mask is able to rotate with respect to the detector and has been used in conjunction with both a square pixel detector array (SPD) and a hexagonal pixel detector array (HPD).

The detectors are circular with the useful region having the same size and orientation as the shadow of the basic mask pattern cast on it from the source radiation. The detector efficiency and the closed mask element opacity have both been assumed to be 100%.

Mask rotation has been simulated in steps with the *mask orientation angle* α being the angle of orientation of the mask with respect to the detector plane. In the SPD case, α is the angle of the $\underline{e}_0$ axis of the hexagonal mask grid with respect to the x axis of the SPD grid as shown in fig. 5.2, so that when $\alpha = 0°$ these two axes are coincident. In the HPD case, special detector configurations have been chosen such that two mask orientation angles give a near perfect binning of the mask element shadows by the detector pixels (fig. 5.3). One of these angles is chosen as $\alpha = 0°$ and all other mask orientations are measured from this starting point. Because of the difference in the geometry of the two detectors, there is no relationship between the value of α of one detector and that of the other.

Both stationary mask and rotating mask observations have been made. A stationary mask image is simply an image of a source taken with the mask at a constant orientation. A

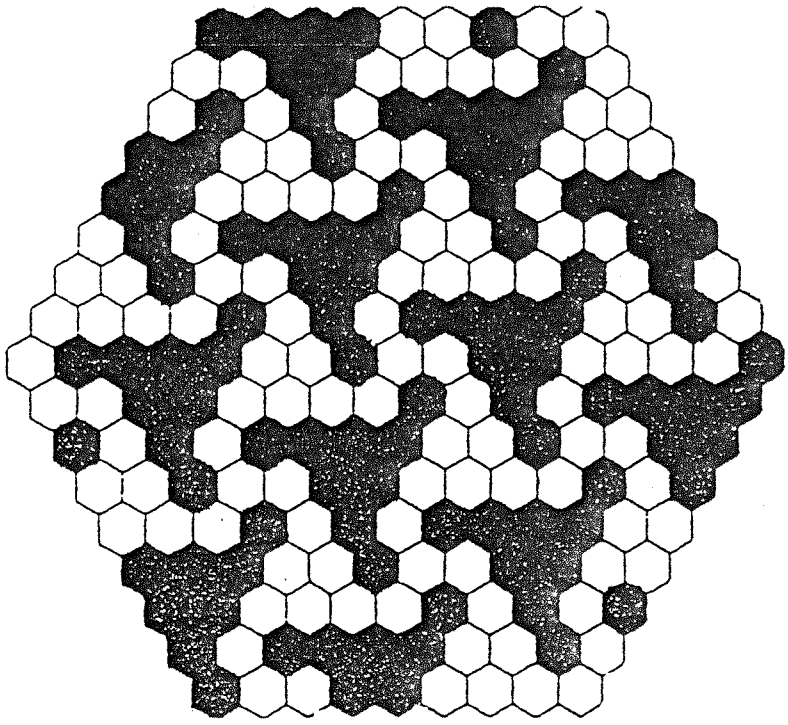


Fig. 5.1. *Mask pattern used in the simulations. The basic pattern has 19 elements, and there are 217 elements in total.*

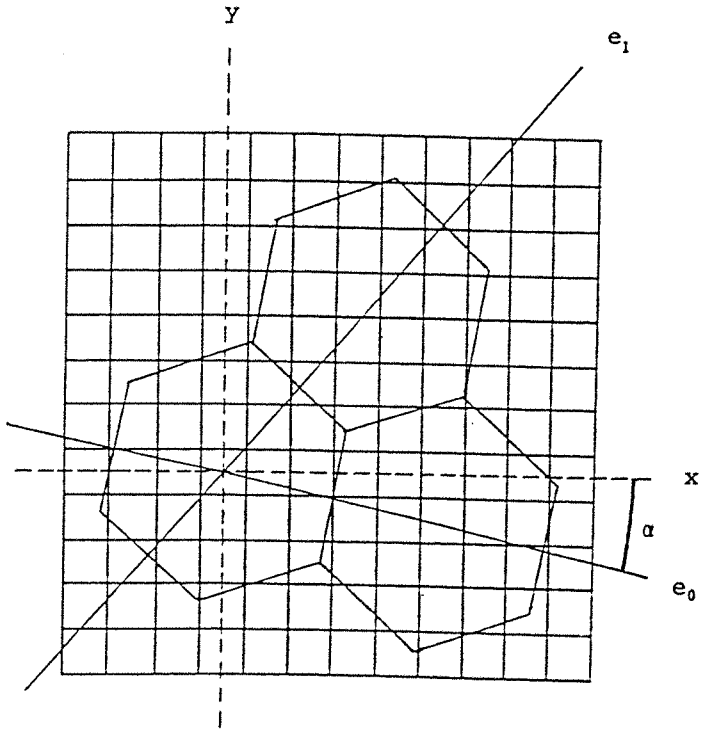


Fig. 5.2. *Mask element shadow and detector pixel configuration for the SPD for* $A_r = 21.7$.

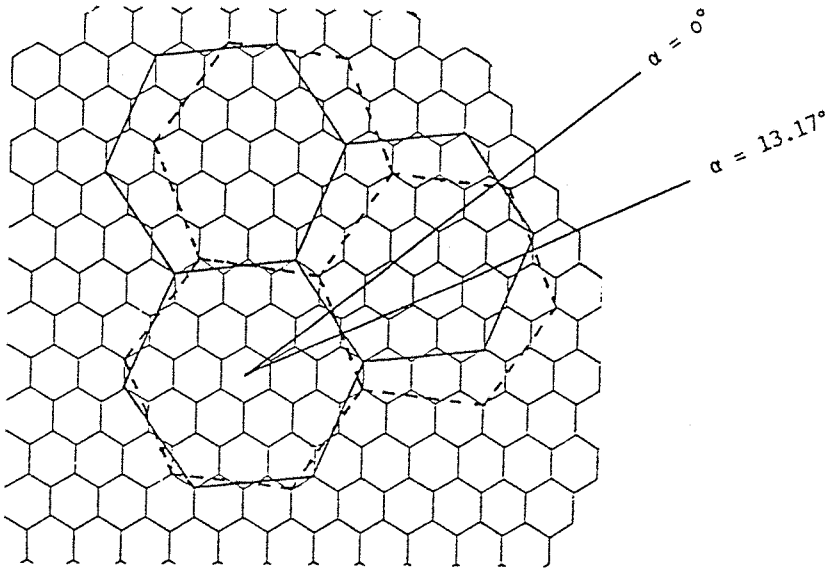


Fig. 5.3. *Mask element shadow and detector pixel configuration for the HPD for* $A_r = 19$.

rotating mask image consists of the sum of six separate stationary mask images at 10° to each other.

The detector resolution is defined by the *relative area* A_r. A_r is the area of the projection of one mask element onto the detector by the source, divided by the area of one detector pixel. The simulations in this section have been carried out having the mask element projection size (flat to flat) equal to N times the detector pixel size for the SPD (where N is an integer), and equal to $\sqrt{\frac{3}{2}(M+1)(\frac{3}{2}[M+1]-1)}-1$ times the detector pixel size (flat to flat) for the HPD (where M is an odd integer). Typical values used in the simulations were $N = M = 5$, giving slightly different relative areas in the two detector plane cases ($A_r = 21.7$ for the SPD as in fig. 5.2, and $A_r = 19$ for the HPD as in fig. 5.3).

5.3.2 Convolution

The convolution of the radiation by the mask has been simulated by calculating the area of each detector pixel which was obscured by the mask in a given system configuration. This was done by splitting the detector pixels into smaller "virtual pixels" and considering those virtual pixels whose centre lay in the shadow region of the mask projection to be totally obscured. The square detector pixels were simply split into smaller squares whilst the hexagonal detector pixels were split into smaller hexagons by using the same configuration as that of the mask element and detector pixels for the perfect overlap situation as mentioned above (fig. 5.3). This convolution method gave an estimate of the mask shadow pattern on the detector to a sufficient level of approximation in an acceptable computing time. Each square detector pixel was split into 7 x 7 square virtual pixels, and each hexagonal detector pixel was split into 91 hexagonal virtual pixels.

5.3.3 Deconvolution

Deconvolution has been performed by using the finely sampled balanced correlation technique [46,47], where each detector pixel flux is projected back to each source bin and either added or subtracted depending on whether the mask element is open or closed, see section 3.2.2. Each detector pixel is correlated with the mask element in which its centre is projected back. Because of the non perfect overlap of the mask element shadows with the detector pixels, this standard deconvolution technique is expected to produce some system-

atic noise in the form of a *coding error* noise, which is inherent in the imaging system. This is not the case for an optimum system using, for example, square mask elements and a square pixel detector, where the overlap is perfect and no coding noise is present [60].

The stationary mask deconvolved images exhibit a peak with a broad base (similar to a cone) corresponding to the source position, and repetitions of this source periodically spaced throughout the image with hexagonal periodicity. One such image is shown in fig. 5.4 (a) for the SPD with $A_r = 21.7$. The rotated mask images enable the true source peak to be easily identified because the repetitions are blurred out into a low ring centred on the source peak [35]. A rotated mask image for the SPD with $A_r = 21.7$ is shown in fig. 5.4 (b).

5.3.4 Signal to Noise Ratio

The image quality has been quantified by using the standard expression for the *signal to noise ratio* (SNR) defined by

$$\mathrm{SNR} = \frac{P - B_D}{\sigma_B} \tag{5.4}$$

where P is the deconvolved source peak pixel flux value, B_D is the mean background noise per deconvolved pixel and σ_B is the standard deviation in the deconvolved background. For a statistical point source in the presence of background noise, the expected SNR, SNR_{exp}, of a deconvolved image taken using an optimum coded system is, as given in equation 3.22 in section 3.4:

$$\mathrm{SNR}_{exp} = \frac{S}{(S+B)^{1/2}} \tag{5.5}$$

where S is the total detector source counts and B is the total detector background counts. When a systematic effect, such as coding error, produces a noise comparable to or higher than statistical noise, the SNR becomes lower than SNR_{exp}.

Because of the nature of the deconvolved images the SNR has been calculated using the area of the image lying just outside the base of the source peak and inside a circle stretching out as far as possible without including any part of any of the source repetitions.

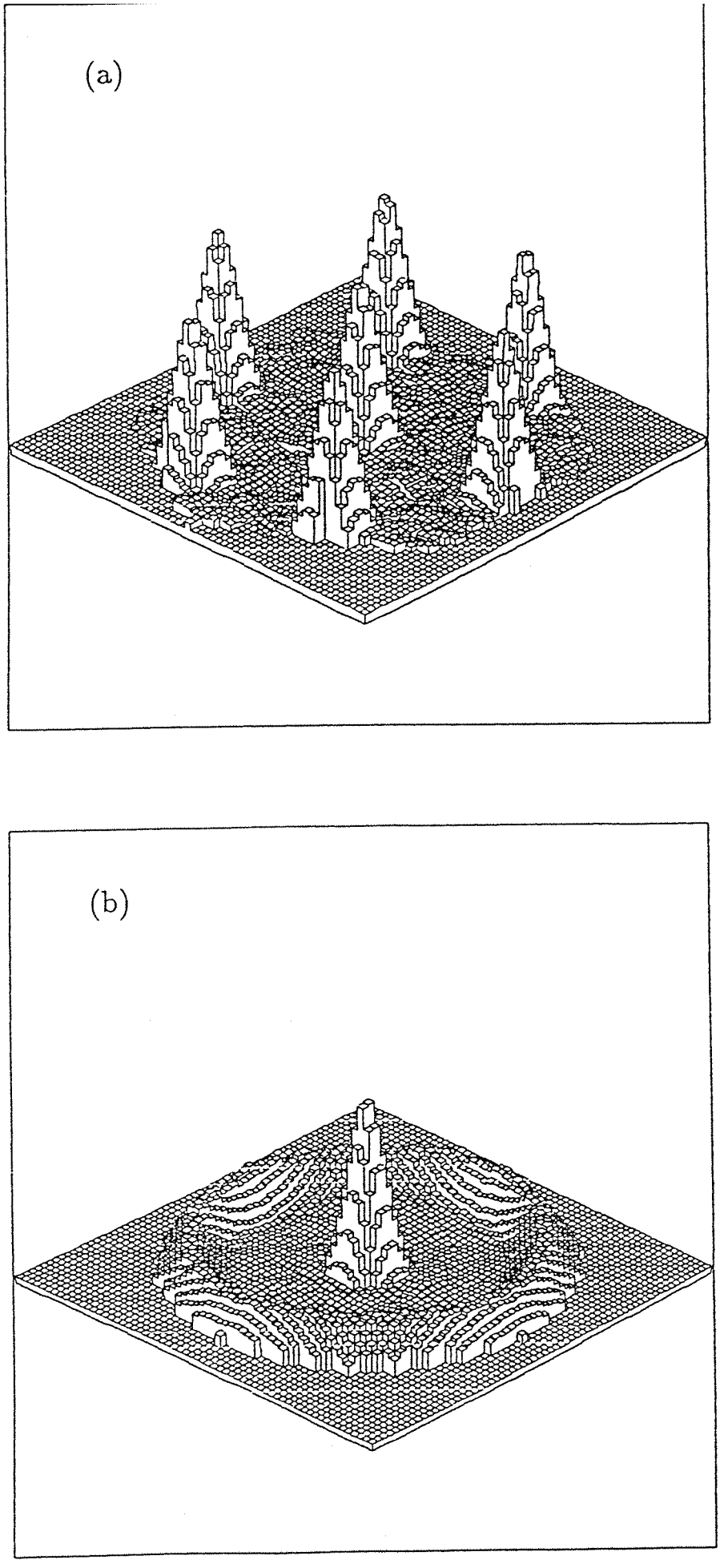


Fig. 5.4. *Deconvolved images for the SPD for $A_r = 21.7$ for (a) stationary mask, and (b) rotated mask.*

5.4 Computer Simulations and Comparison of the SPD with the HPD

The computer simulations have been divided into three sections:

1. System point spread function.
2. Statistical source and background.
3. Case of nearly perfect overlap for the HPD.

The system point spread function (SPSF) of an imaging system is the response of the system when observing a non statistical point source in the fully coded field of view in the absence of background noise. This is an alternative, but equivalent, definition to that given in section 3.2.2. Therefore, for a perfect imaging system the SPSF is a delta function and the SNR if infinite. For an imperfect system the coding error noise varies with different observational situations and this noise has been studied and quantified neglecting statistical variations and background noise. For the SPSF the imaging quality has been studied under conditions of varying mask orientation angle, different source position, and varying detector position resolution. Simulations have also been carried out of observations of a statistical point source in the presence of background noise for both stationary and rotating mask images. Finally the case of nearly perfect mask element binning by the HPD has been studied to see the possible reduction in coding error noise using this special geometry. In all cases the simulations assumed that a source lay at the very centre of a sky pixel.

5.4.1 System Point Spread Function

Mask Orientation

The SNR of the SPSF has been calculated for a randomly positioned point source in the fully coded field of view at different mask orientations α between 0° and 60° using both detector geometries. The relative areas used were $A_r = 21.7$ for the SPD (fig. 5.2) and $A_r = 19$ for the HPD (fig. 5.3). The results are shown in figs. 5.5 and 5.6 where fig. 5.5 is the graph of SNR vs. α produced by the SPD and fig. 5.6 is the corresponding graph for the HPD. In both cases the SNR varies irregularly and rapidly with no apparent periodicity within the

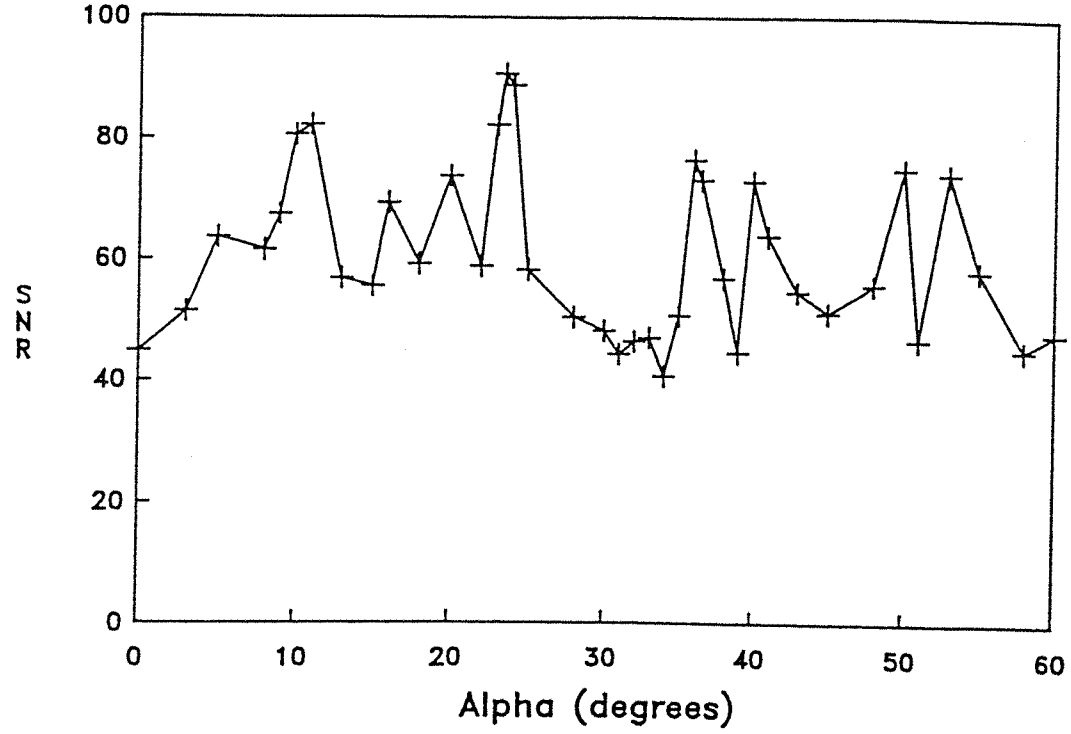


Fig. 5.5. *SNR of SPSF versus mask orientation α for the SPD for $A_r = 21.7$.*

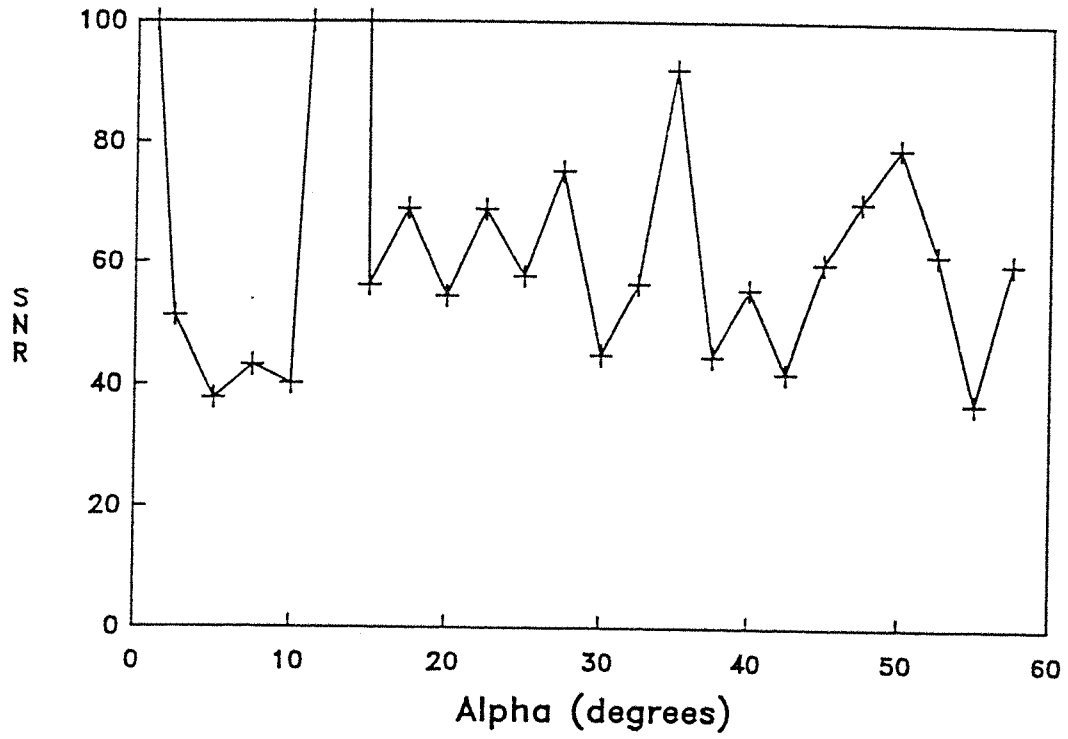


Fig. 5.6. *SNR of SPSF versus mask orientation α for the HPD for $A_r = 19$.*

60° range studied. The SNR fluctuates up to 40° from the mid SNR (i.e. the value mid way between the highest and lowest SNR values). For a perfect system the SNR is infinite. However the SNR here is not infinite due to the non perfect overlap of detector pixels with the mask element shadows, producing coding error noise. The source of the coding error noise is illustrated qualitatively in fig. 5.7. Fig. 5.7 (a) and (b) show the mask oriented at 0° and 10° respectively whilst (c) and (d) show their corresponding deconvolution arrays for the SPD when $A_r = 21.7$. Because of the need to have the deconvolution array in the same geometry as the detector array, the hexagonal mask patterns shown in (a) and (b) have been approximated to those shown in (c) and (d) respectively. It is clear that this approximation will produce coding noise in the deconvolution process and keep the SNR finite. The two deconvolution arrays for $\alpha = 0°$ and 10° can be seen to have markedly different shapes. Thus the coding error is different for the two different situations and hence the SNR fluctuates with α, giving the profile in fig. 5.5. For a fixed source position significant changes in geometrical correlation between detector pixels and mask element shadows is expected to occur for changes in the mask orientation of $\Delta\alpha >$ half pixel width/mask radius, which is about 1° in this case, which correlates well with the results where high variations in the SNR are observed on angular scales of $> 1°$.

In the HPD case, however, the SNR approaches infinity at the two mask orientation angles $\alpha = 0°$ and $\alpha = 13.2°$. These are the two angles for which the detector pixels are well matched by the mask element shadows, as shown in fig. 5.3, indicating that the coding error noise is effectively absent in these cases.

Source Position

The SNR has been studied for different source positions around the centre of the field of view, and for larger displacements in a straight line across the whole field of view. Source movements inside an area of approximately one mask element gave only a small SNR range in each case, of value < 7%. Higher variations have been observed for larger displacements. For example fig. 5.8 shows the variations in SNR when moving the source across the field of view in the case of the HPD for three different mask orientations ($\alpha = 35°, 45°, 55°$). Changes of up to 20% from the mid SNR are observed, but the fact that the three different curves don't cross each other at any point shows that the mean noise is clearly determined by the mask orientation than by the source position. The SPD gives very similar results.

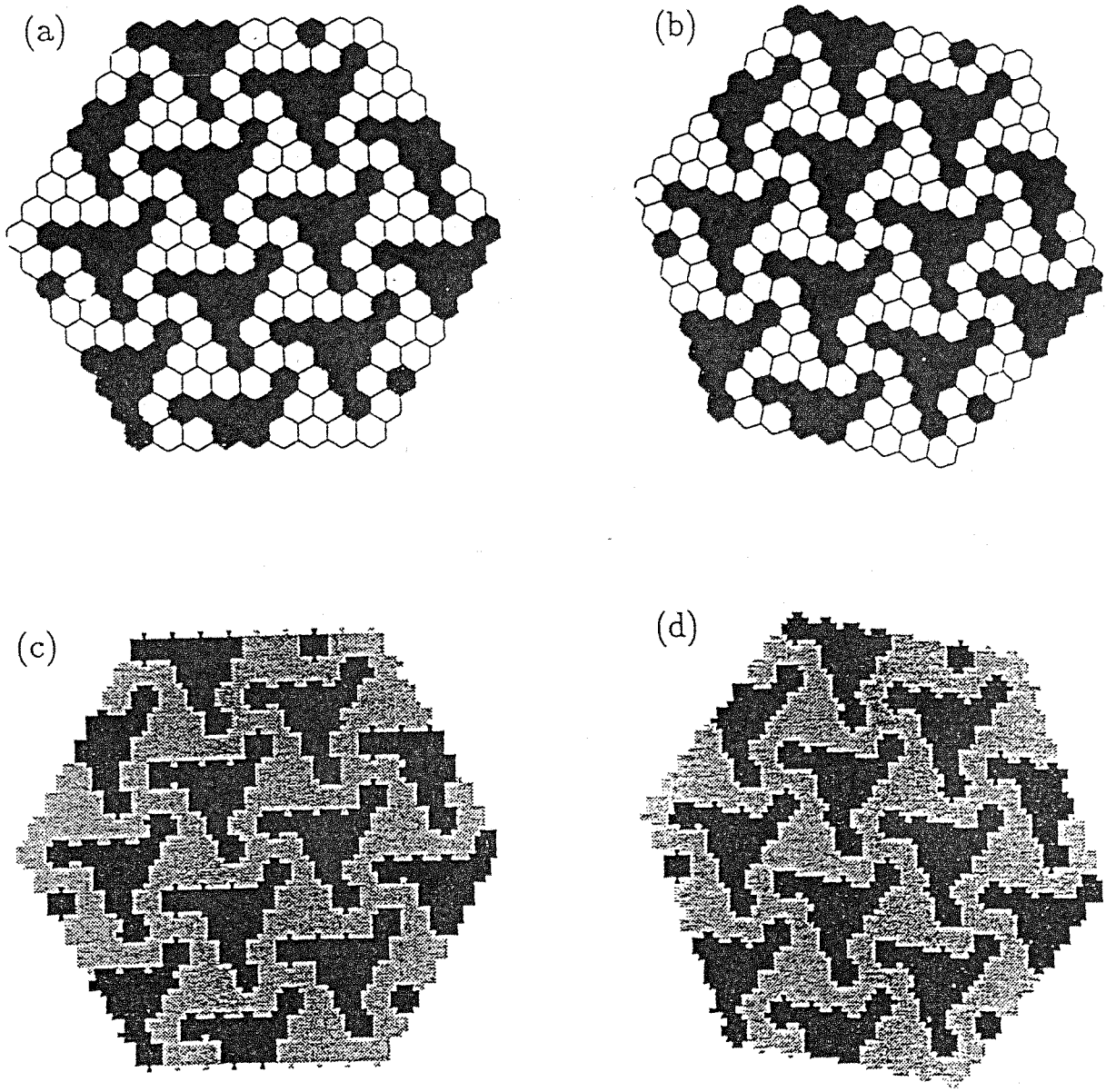


Fig. 5.7. *Masks for (a)* $\alpha = 0°$ *and (b)* $\alpha = 10°$*, and deconvolution arrays for (c)* $\alpha = 0°$ *and (d)* $\alpha = 10°$ *for the SPD at* $A_r = 21.7$ *(light areas = open elements, dark areas = closed elements).*

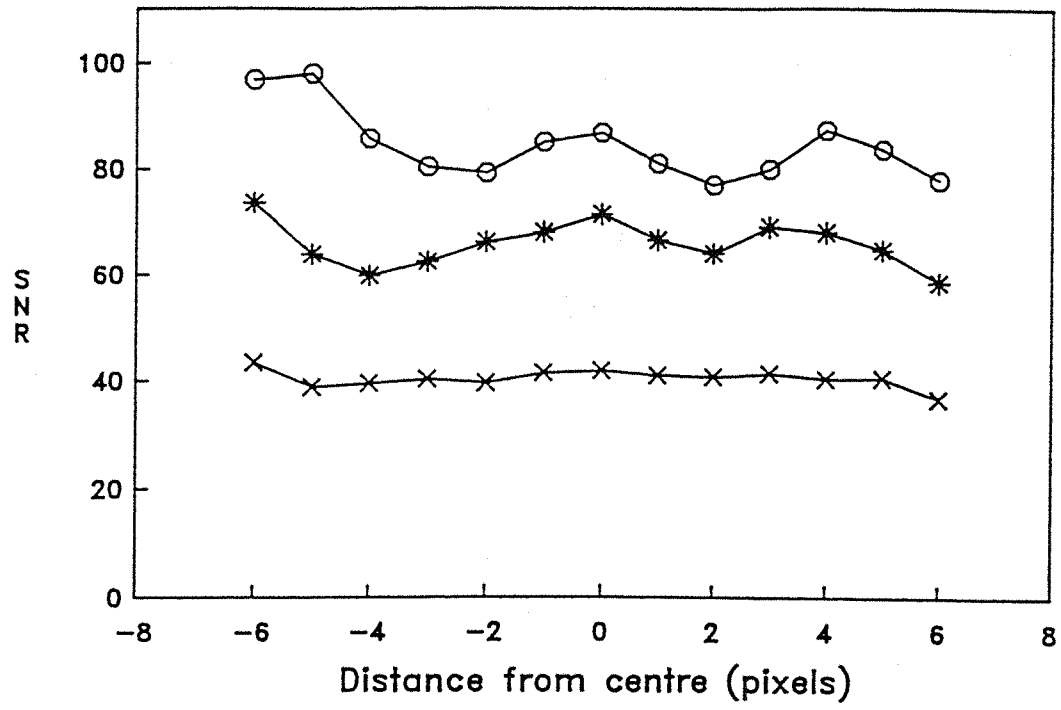


Fig. 5.8. *SNR of SPSF versus source position for the HPD (A_r = 19). Each curve corresponds to a different value of α (35° = circles, 45° = asterisks, 55° = crosses).*

Relative Area

A point source in the centre of the field of view has been observed using different position resolutions for both a stationary and a rotated mask for both detector geometries. The graphs of SNR versus relative area A_r using stationary masks at a fixed orientation are shown in figs. 5.9 and 5.10 for the SPD and HPD respectively. As would be expected, the image quality generally increases with increasing detector positional resolution due to a better approximation to the "ideal" coding system being achieved at higher resolution. However, at any given mask orientation geometrical artifacts due to the coding error may occur leading to unexpected trends. For example there is a decrease in SNR for the SPD from $A_r = 7.8$ to $A_r = 13.9$ when $\alpha = 43°$. Neglecting such artifacts, the general trend seems to be approximately linear. This is evidently the case also for rotating mask images where, in addition, the geometrical artifacts are averaged by the rotation, as can be seen from the results shown graphically in fig. 5.11 for both detector geometries. A linear fit to each dataset is also shown, the fit being particularly good for the HPD where a regression analysis has a very high correlation coefficient (= 0.98). The gradients of the lines are 5.0 for the SPD and 6.2 for the HPD.

The values of SNR for rotated mask images at $A_r = 21.7$ for the SPD and $A_r = 19$ for the HPD are 125.2 and 118.0 respectively. Therefore for rotating mask images with low background level the SNR of the SPD is expected to be about 10% higher than for the HPD at these A_r values. This is an important consideration when comparing the two detector geometries, and is referred to in the following section.

5.4.2 Statistical Source and Background

The same imaging systems have been used to simulate observations of a non central point source in the presence of varying levels of background noise for both a stationary and a rotating mask. The relative areas were $A_r = 21.7$ for the SPD and $A_r = 19$ for the HPD. In all simulations the mean source flux was 100 photons per fully illuminated square pixel of the SPD, and the background levels were 0.0, 0.1, 0.5, 1.0, 3.0, 5.0, and 10.0 times the mean source flux. The corresponding fluxes used for the HPD were normalised to those of the the SPD so that the total detector counts were kept constant for each geometry. Mask orientations were $\alpha = 3°, 13°, 23°, 33°, 43°$ and $53°$ for the SPD, and $\alpha =$

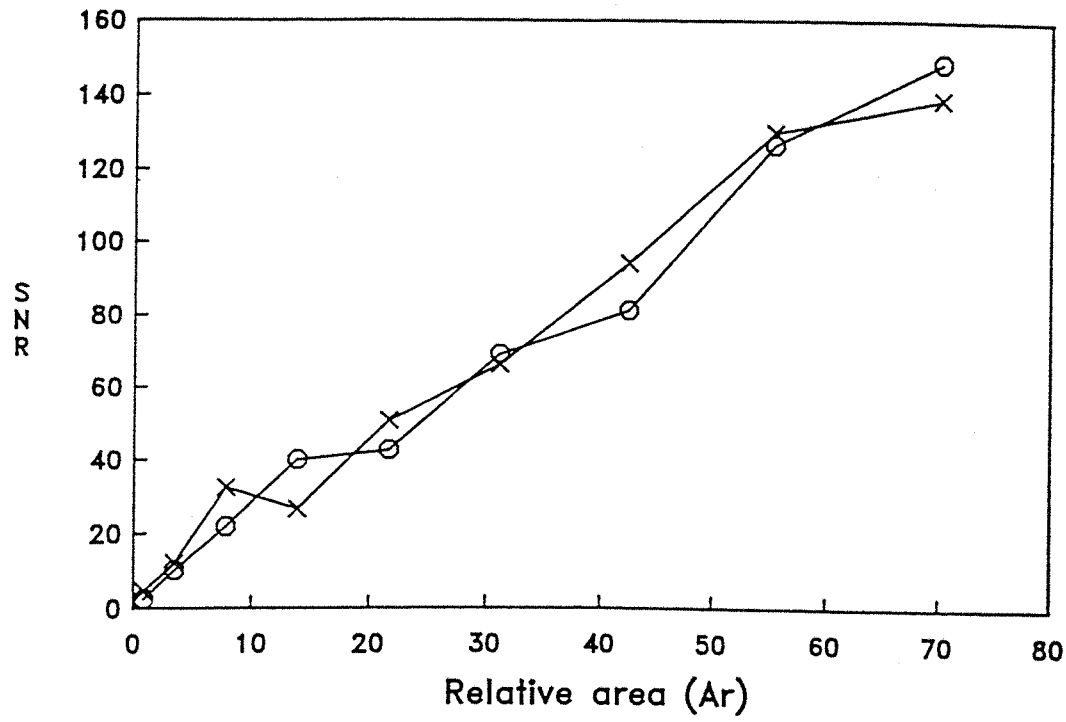


Fig. 5.9. *SNR of SPSF versus A_r for the SPD using a stationary mask for $\alpha = 33°$ (circles) and 43° (crosses).*

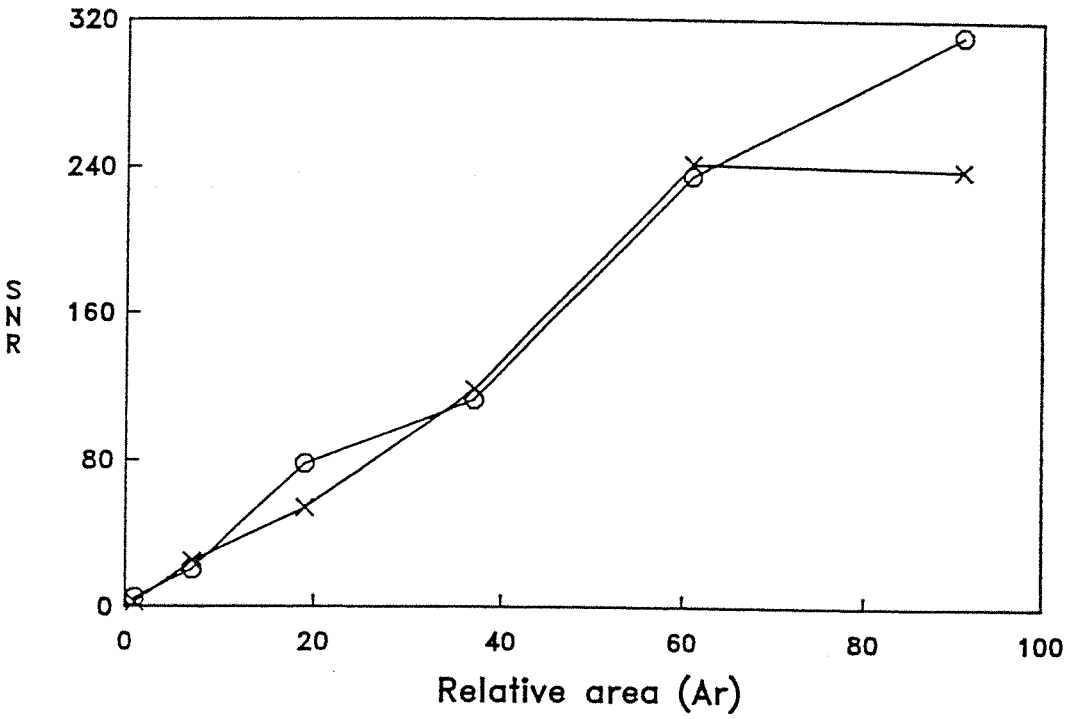


Fig. 5.10. *SNR of SPSF versus A_r for the HPD using a stationary mask for $\alpha = 25°$ (crosses) and $\alpha = 15°$ (circles).*

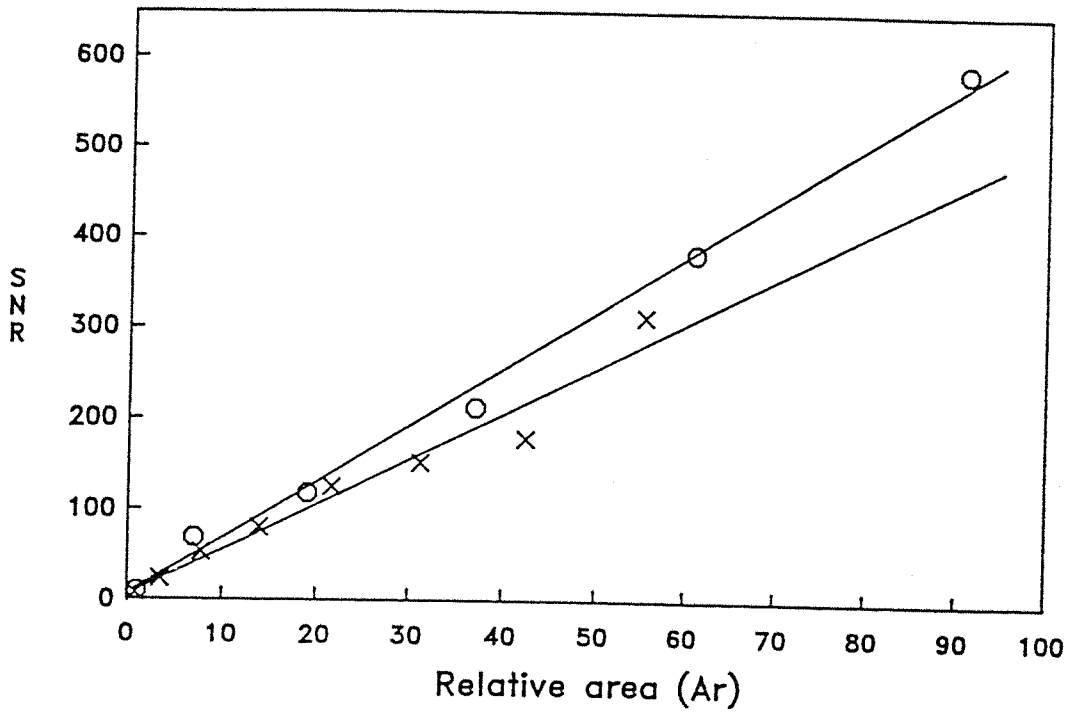


Fig. 5.11. *SNR versus A_r for the SPD and the HPD for rotating mask. The datapoints and lines of best fit are shown for the HPD (circles) and SPD (crosses).*

$0°, 5°, 15°, 25°, 35°, 45°$ and $55°$ for the HPD.

The graphs of SNR vs α for six of the seven background levels are shown in fig. 5.12 for the SPD and fig. 5.13 for the HPD. Also indicated are the expected SNR values, SNR_{exp}, for a perfect system as given by equation 5.5. Because of the presence of coding error noise the observed SNR is always less than SNR_{exp} for both detectors. The observed decrease ranges from ~50% at low background up to ~90% at higher background. The exceptions to this are at the two angles of near perfect overlap for the HPD, the results of which are discussed in the following section.

Figs. 5.12 and 5.13 show that the SNR decreases with increasing background noise, as would be expected, but the profiles vs. α have the same general trend as those of the SPSFs (for example peaks at $\alpha = 23°$ for the SPD and at $\alpha = 35°$ for the HPD). This shows that the noise due to the coding error is always overwhelmingly dominant over statistical noise, even in the case of high background noise. Moreover the background noise level seems to heavily magnify the coding error noise.

For the rotated mask five trials of observations of a point source away from the centre of the field of view have been simulated for each background level, and the results, which are shown along with errors in table 5.1, have been plotted on a logarithmic scale graph, shown in fig. 5.14. The results for both detectors are plotted, along with the expected SNR values for the rotated mask images. As expected the SNR decreases with increasing background noise for both detectors, and in all cases is less than SNR_{exp}. This shows that the coding error noise is also present in rotated mask images and therefore reduces the image quality. Table 5.1 also shows the ratio of SNR to SNR_{exp} for each simulation and it can be seen that this ratio decreases with increasing background noise. This implies that the background level magnifies the coding error noise in the deconvolved images.

The SNR is further reduced by an effective loss of source photons. These are photons which originate from the source but are incident on detector pixels which are correlated with opaque mask elements when deconvolving the sky pixel in which the source lies. The proportion of these "lost" photons is 5% for the SPD at $A_r = 21.7$, with a similar value for the HPD. Thus, by equation 5.5 for SNR_{exp}, the SNR is expected to fall by 5%. This is a negligible decrease, however, compared to that produced by the coding error noise, with values of 50% to 90% as mentioned above.

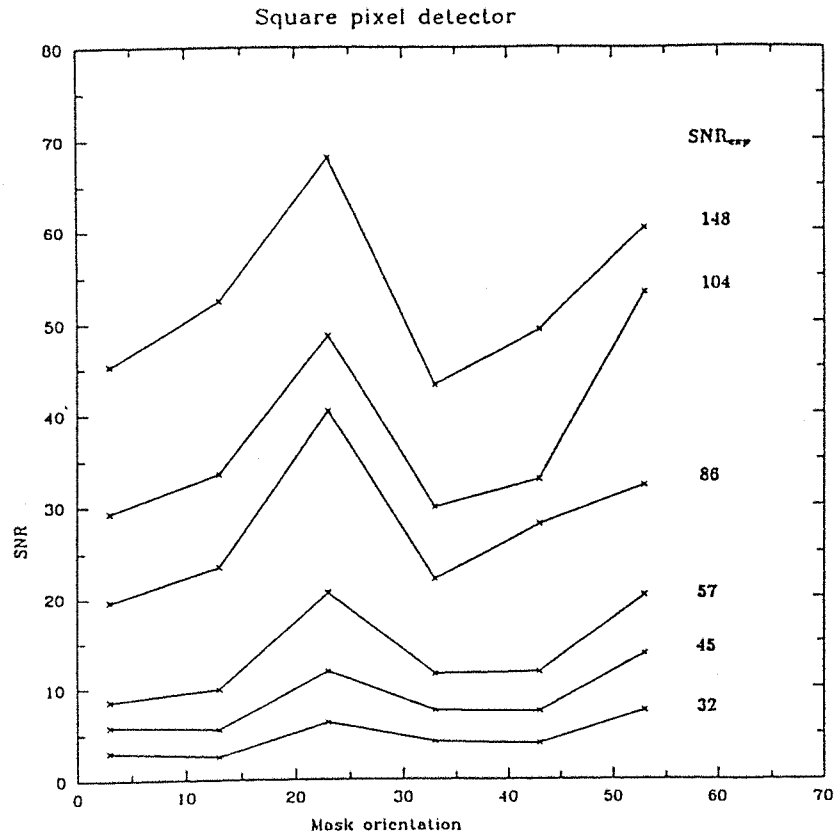


Fig. 5.12. *SNR versus α for six background levels using the SPD. Also indicated are the corresponding values of SNR_{exp}.*

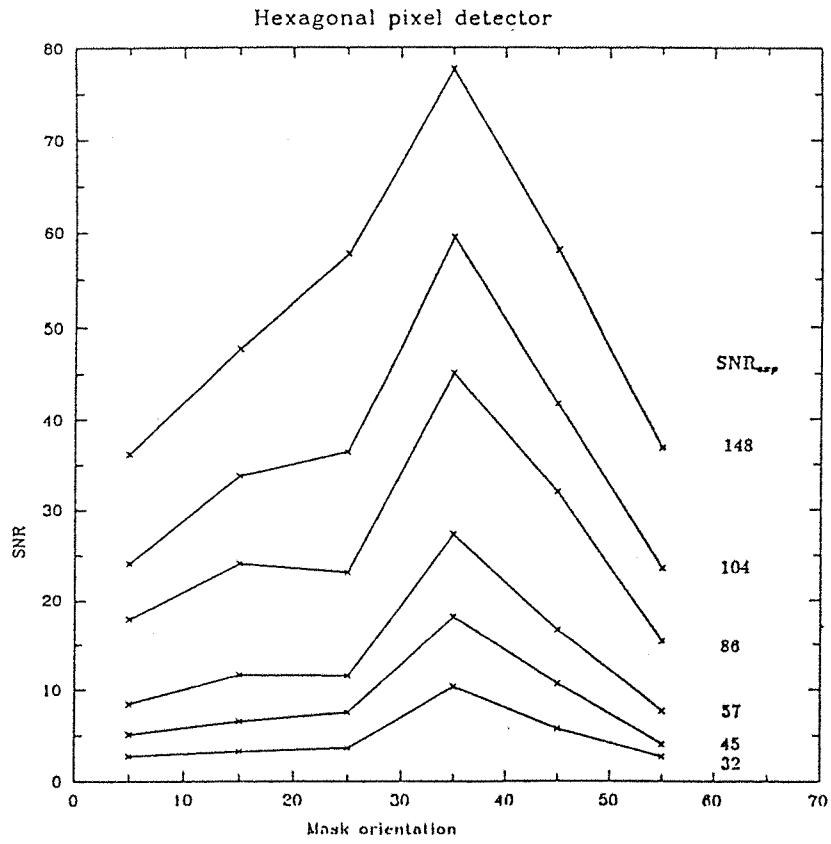


Fig. 5.13. *SNR versus α for six background levels using the HPD with SNR_{exp} values indicated.*

Bkg flux (source flux units)	SNR_{exp}	SPD		HPD	
		SNR	SNR/SNR_{exp}	SNR	SNR/SNR_{exp}
0.0	362	132.3 ± 6.2	0.37	118.6 ± 3.7	0.33
0.1	330	127.7 ± 3.7	0.39	108.2 ± 5.8	0.36
0.5	256	91.8 ± 8.0	0.36	74.6 ± 3.6	0.29
1.0	210	63.8 ± 2.6	0.30	53.9 ± 1.3	0.26
3.0	139	31.7 ± 0.6	0.23	24.8 ± 1.2	0.18
5.0	111	20.0 ± 0.4	0.18	15.5 ± 0.5	0.14
10.0	80	10.2 ± 0.3	0.13	8.1 ± 0.2	0.10

Table 5.1. *SNR of rotated mask images for the SPD at $A_r = 21.7$ and the HPD at $A_r = 19$ at 7 different background levels. Also given are the SNR_{exp} values at each level and the quantity SNR/SNR_{exp}.*

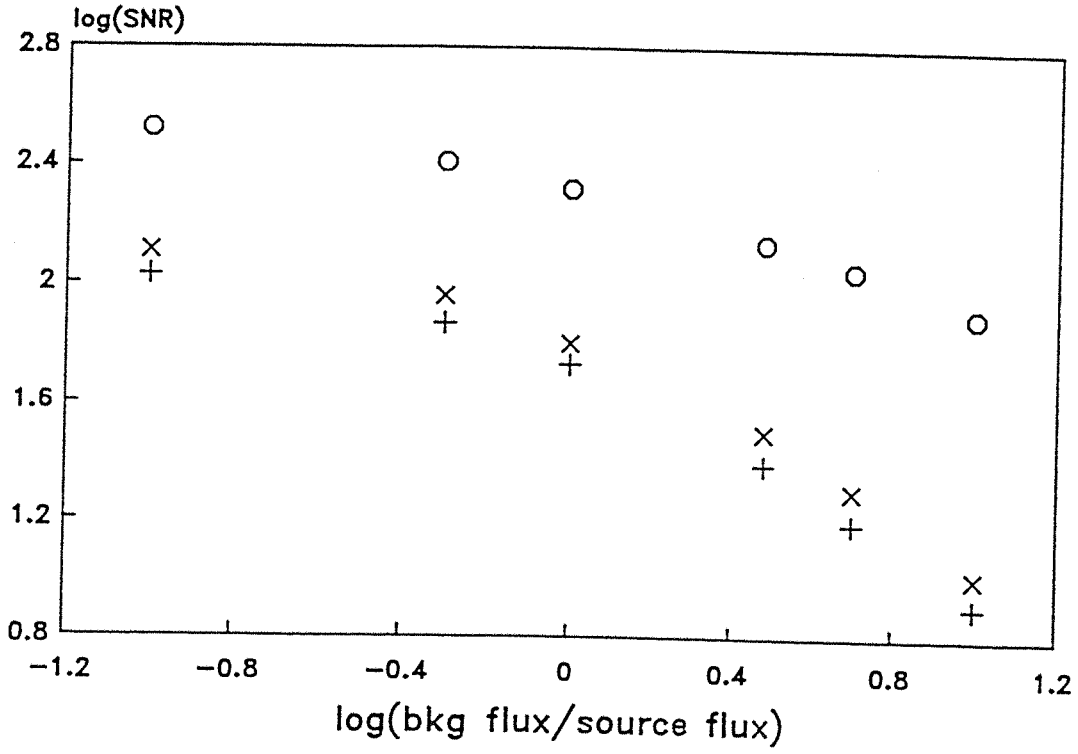


Fig. 5.14. *Logarithmic scale plot of SNR versus Background flux/Source flux for a non central point source using the SPD ($A_r = 21.\overline{7}$, crosses) and the HPD ($A_r = 19$, pluses). Also shown are the SNR_{exp} values in the absence of coding error noise (circles).*

The observed SNR values for the SPD are between 10% and 20% higher than those of the HPD for rotated mask images. This is probably due to A_r being higher for the SPD giving a finer positional resolution. As was pointed out in the previous section such differences in resolution can lead to the changes observed.

5.4.3 Case of Near Perfect Overlap for the HPD

Five observation trials of the same non central point source were simulated for the HPD with $A_r = 19$, using the mask orientations of near perfect overlap $\alpha = 0°$ and $\alpha = 13.17°$ (fig. 5.3). The same background levels as before were used. The results, with errors, for $\alpha = 13.17°$ are given in table 5.2. The expected SNR for each background level lies inside the estimated errors calculated from the five trials at each level. This indicates that with the HPD and a stationary mask oriented at either of the two angles of near perfect overlap the coding error noise is not in evidence. Although there is no coding error noise, there is still an effective loss of source photons at these angles because the overlap is still not perfect which should give a SNR less than SNR_{exp}. This reduction is again very small, less than 5%, and therefore not apparent in the results given in table 5.2, since statistical fluctuations are of the order of 10%. Other simulations of the SPSF of this configuration have shown that the SNR tends to infinity, and does not change with source position nor detector resolution.

A rotating mask incorporating these optimum angles was also used to observe a central point source, and the results are shown in table 5.3. A clear improvement (up to $\sim 40\%$) is seen only when using both of the optimum angles to build a rotating mask image. However when incorporating only one of these angles inconsistent results were obtained, depending on which other orientation angles were used.

5.5 Discussion and Conclusions

When a point source is observed using a HURA mask and a discrete pixel detector, a coding error noise is observed to be present in the deconvolved image. This is due to the non perfect overlap of detector pixels with the mask element shadows. Such a mismatch causes the recorded image of the mask shadow to be *asymmetrically* distorted by the finite resolution and discrete nature of the detector. It must be emphasised that this blurring is

Bkg flux (source flux units)	SNR_{exp}	SNR
0.0	148	167 ± 19
0.1	135	142 ± 17
0.5	104	112 ± 17
1.0	86	83 ± 11
3.0	57	57 ± 9
5.0	45	46 ± 5
10.0	32	34 ± 4

Table 5.2 *SNR of stationary mask images for the HPD at* $\alpha = 13.17°$ *and* $A_r = 19$ *for different background levels, along with* SNR_{exp} *values.*

Bkg levels	SNR					SNR_{exp}
	A	B	C	D	E	
0.0	126.4	121.8	145.3	170.0	118.6 ± 3.7	362
0.1	123.8	105.7	133.4	150.1	108.2 ± 5.8	330
0.5	89.2	69.4	103.7	102.0	74.6 ± 3.6	256
1.0	63.9	53.9	79.9	78.4	53.9 ± 1.3	210
3.0	31.2	24.1	37.4	37.5	24.8 ± 1.2	139
5.0	18.4	15.1	23.9	23.2	15.5 ± 0.5	111
10.0	10.5	8.1	12.9	12.4	8.1 ± 0.2	80

Table 5.3. *SNR of rotated mask images for the HPD at $A_r = 19$ incorporating the special angles of perfect overlap, along with SNR_{exp}, at different background levels:*

A: 0°, 10°, 20°, 30°, 40°, 50°.

B: 5°, 13.2°, 25°, 35°, 45°, 55°.

C: 0°, 5°, 15°, 25°, 35°, 45°.

D: 0°, 13.2°, 25°, 35°, 45°, 55°.

E: 5°, 15°, 25°, 35°, 45°, 55°.

completely different from that which occurs when a source lies on the border of two sky pixels in the field of view of an optimum coded system (also sometimes referred to as a phasing error, as in [47]). In this case the blurring is *symmetrical* in the sense that it occurs in the same way for any mask element shadow, so that the only effect is a blurring of the source peak and no coding error noise is generated in the deconvolved images. Moreover, in a simple cross correlation deconvolution, the deconvolution array has the same geometry as the detector array and therefore it is only an approximation of the real hexagonal mask pattern (fig. 5.7 (c)). The sky pixels are reconstructed by correlating the detector array with different parts of this deconvolution array. Because of the discrete subdivision of the mask, some photons which should be correlated with an open mask element instead become correlated with a closed one and vice versa, and an error is introduced in the reconstruction of the source counts for off source sky positions. Because different sky pixels correspond to different subsets of the deconvolution array, the number of closed cells varies from subset to subset. Thus, different proportions of the background will be subtracted for different sky pixels giving a coding error for the background, the magnitude of which depends on the background level, and being dominant for higher background counting rates.

As expected the higher the detector resolution the better is the geometrical correlation between detector pixels and mask element shadows, and therefore the lower the coding error noise. The noise was found to decrease linearly with the number of pixels per mask element, with gradients of SNR vs. A_r of 5.0 for the SPD and 6.2 for the HPD in the case of the rotated mask, both curves passing close to the origin (fig. 5.11). For both detector geometries the coding error noise also depends on the mask orientation and, to a lesser extent, on the source position, and is highly magnified by the background level. In the case of a detector with about 20 pixels per mask element, the SNR of the SPSF fluctuates irregularly with the mask orientation (up to 50% from the mid SNR, figs. 5.5 an 5.6), and with the source position (up to 20%, fig. 5.8). For both a stationary and a rotating mask the reduction of the SNR with respect to the expected SNR for statistical noise only (SNR_{exp}) ranges from 60% at low background levels up to 90% for input signal to noise ratios of 0.1. Therefore the sensitivity of a telescope using this coded system can be drastically reduced, particularly if it works in the energy range of gamma ray astronomy where the input signal to noise ratios are typically of the order of $10^{-3} - 10^{-2}$.

In the case of a rotating mask, the problem cannot be avoided by using a particular detector geometry, and more sophisticated deconvolution techniques must be applied. Though

other deconvolution techniques (such as the use of a Wiener filter, or the maximum entropy method [159]) may give better results, one can improve the correlation by reducing the error due to the decoding array both for the source and for the background counts, as will be shown in the following chapter. For a stationary mask however, it has been found that for a detector array of hexagonal pixels with a special ratio of detector pixel size to mask element size, the mask can be oriented in such a way that the overlap between mask elements and detector pixels is nearly perfect and hence the coding error noise is completely absent.

Chapter 6

Experimental Imaging Tests Using a 19 Element HURA Coded Aperture.

6.1 Introduction

In order to investigate further the theories developed in the previous chapter a series of experimental tests was carried out using a HURA of basic pattern $v = 19$ and a gamma ray photon detector to view a radioactive gamma ray point source. These tests, along with correlated computer simulations are now described.

6.2 Experimental Set Up

The experimental tests were carried out using a HURA mask, an Anger camera detector and a radioactive source. The mask used was the same pattern as that shown in fig. 5.1, but slightly smaller, having 127 elements in total. The opaque mask elements consisted of 2 mm thick hexagons of lead which were 30 mm flat to flat. These elements were glued to a 53 cm diameter circular perspex disk support, which in turn was placed on a metallic structure.

The structure was constructed in such a way that the mask could be rotated about the very centre of the central mask element, and with a graded angular scale thus allowing the mask to be oriented at the desired angles. The mask was placed at half distance between the source and the detector so that the dimension of the projected shadows of the elements was approximately 60 mm flat to flat, and the basic pattern was occupying most of the sensitive detector area.

The detector was a circular Anger type gamma camera [136] at the Southampton General Hospital. It has a circular sensitive region of 37 cm diameter, an effective energy range of 80-500 keV, and a positional resolution of 6 mm FWHM. Data of any observation were recorded and digitised in maps of 128 × 128 square pixels (2.9 mm each) and stored on magnetic tapes to be subsequently analysed.

6.3 Laboratory Tests

To simulate a celestial source a 5.14 μCi ^{57}Co source was used. The 122 keV emission line is 99.9% absorbed by the 2 mm of lead, whilst the perspex support is effectively transparent. This source was placed approximately at the centre of the field of view at a distance of 1 metre from the detector, and was imaged 12 times for 12 different mask orientation angles from 0° to 110° separated by 10° intervals. At any mask orientation the source was viewed for 2 min 45 sec with a total of approximately 6.3×10^4 photons collected over the useful detector area (i.e. the area of the detector corresponding to the basic 19 element mask pattern). It was calculated that about 18% of the detected events were not modulated by the mask, and hence represent a residual background. This is probably due to several different effects, such as scattered radiation, mask transparency and environmental background. Nevertheless with these values of source and background the value of SNR_{exp} for a stationary mask image due to statistical noise only would be approximately 200, and because any coding error effect limits the SNR to well under 100-150 (see figs. 5.5 and 5.6) this ensures that the effect being studied will not be hidden by the statistical noise.

6.4 Data Analysis

6.4.1 Rebinning in Square Detector Pixels

The recorded gamma camera data consists of maps of the counts detected in each observation in each of the 128 × 128 square gamma camera pixels. Before deconvolution the pixels were rebinned to simulate the response of a discrete pixel detector of the desired positional resolution. The rebinning was firstly carried out with square detector pixels, grouping squares of 7 × 7, 4 × 4 and 3 × 3 camera pixels (higher resolutions would be comparable to, or greater than, the actual camera position resolution of 6 mm). These gave detector resolutions having relative areas A_r =7.8, 23.9 and 42.4. These rebinned detector images are deconvolved by cross correlating them with the corresponding deconvolution array to get a final deconvolved image. Fig. 6.1 shows the steps of the deconvolution process. Fig. 6.1 (a) shows the mask pattern ($v = 19$), (b) shows the deconvolution array ($\alpha = 0°$) (c) shows the recorded gamma camera image (128 × 128 pixels), (d) shows the rebinned detector image ready for deconvolution (e.g. for 4 × 4 camera pixels per detector pixel, $A_r = 23.9$), and the final deconvolved image is shown in (e). Rotating mask images are obtained by summing deconvolved images for six different mask orientation angles between 0° and 50°, or between 60° and 110°. A rotated mask image is shown in fig. 6.1 (f).

6.4.2 Rebinning in Hexagonal Detector Pixels

As pointed out in section 5.4.1, there are special configurations for which no coding error is present when using a HURA mask in conjunction with a HPD. In order to verify this result the data has been deconvolved after rebinning the gamma camera pixels into hexagonal detector pixels according to this optimum geometrical configuration. Therefore one angle of perfect overlap was taken as $\alpha = 0°$ with other angles being measured relative to this starting point. Theoretically there is another angle ($< 60°$) at which this situation again occurs (section 5.4.3) but the 10° rotation step of the mask meant that perfect overlap occurred again only after 60°. At all other angles the mismatch problem arises and there is coding error noise.

Because the square gamma camera pixels cannot be rebinned perfectly into hexagons, the required HPD configuration is well approximated only if the hexagonal pixels contain a

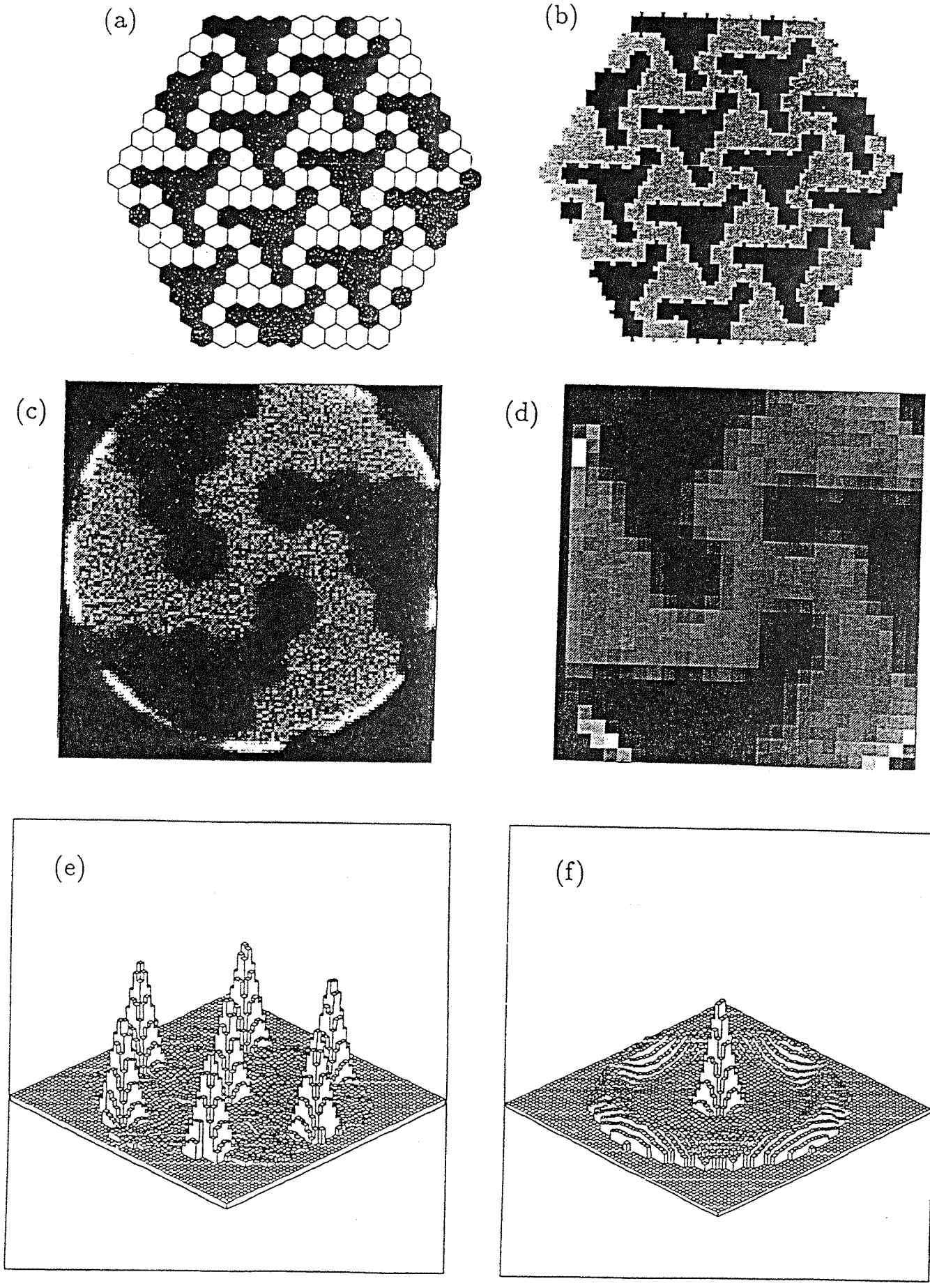


Fig. 6.1. *Steps of the deconvolution process for the SPD: (a) mask pattern ($v = 19$), (b) deconvolution array (c) recorded gamma camera image (128×128 pixels), (d) rebinned detector image ready for deconvolution (4×4 camera pixels per detector pixel; $A_r = 23.9$), (e) final deconvolved image (stationary mask), (f) rotating mask image (summation of six stationary mask images at $\alpha = 0°, 10°, 20°, 30°, 40°, 50°$).*

large number of camera pixels. This, and the geometrical constraints of the configuration which exists only for $A_r = 1, 7, 19, \ldots, 3n(n-1)+1$ $(n = 1, 2, \ldots)$ limit the rebinning to be satisfactory only for small values of A_r ($A_r = 7$ and $A_r = 19$).

6.5 Correlated Simulations

The laboratory images have also been compared to Monte Carlo simulation results. In each case five simulation trials have been carried out and the mean value of the SNR and the errors have been calculated. To obtain the correct input parameters for the simulations the detector images have been analysed and the exact source position and flux, and the mean background level was determined. To determine source flux and background values the numerical matrix corresponding to fig 6.1 (d) was obtained, where each pixel flux is given in number of counts. This matrix is shown in fig. 6.2 for the SPD with $A_r = 23.9$ and $\alpha = 0°$. From this matrix, the pixel fluxes for the 9 closed areas and the 10 open areas of the 19 element basic pattern were found and the mean source flux and background values were obtained. Although this is quite an accurate method of determining the required fluxes for the simulations, systematic effects intrinsic to the gamma camera and environmental background may well be present in the collected data, changing with time and hence with the mask orientation. This sometimes makes it difficult to make a direct comparison with the simulations and to estimate the effect of coding error under different observing conditions. In order to clean the images of this noise, the antimask procedure of Cook et al. [35] has also been used. Here the systematic detector noise due to the background can be almost completely removed by taking a detector image at mask orientation α, and subtracting from it the detector image at $\alpha + 60°$ before deconvolution. This is possible because of the property that HURA masks are antisymmetric upon 60° rotation; i.e. open elements substitute closed elements and vice versa. Here the deconvolution array assumes the value zero for the central mask element and its repetitions. Antimask images are very effective in reducing systematic effects due to the background, and this procedure has enabled the coding error noise to be studied under more preferable conditions.

Two points must be kept in mind when comparing antimask images to normal images. Firstly, although the observation time is double when the antimask image is used, the noise due to the coding error is generally overwhelmingly dominant over the statistical noise and so the higher statistics can be neglected for most of the analysis. Secondly however, the

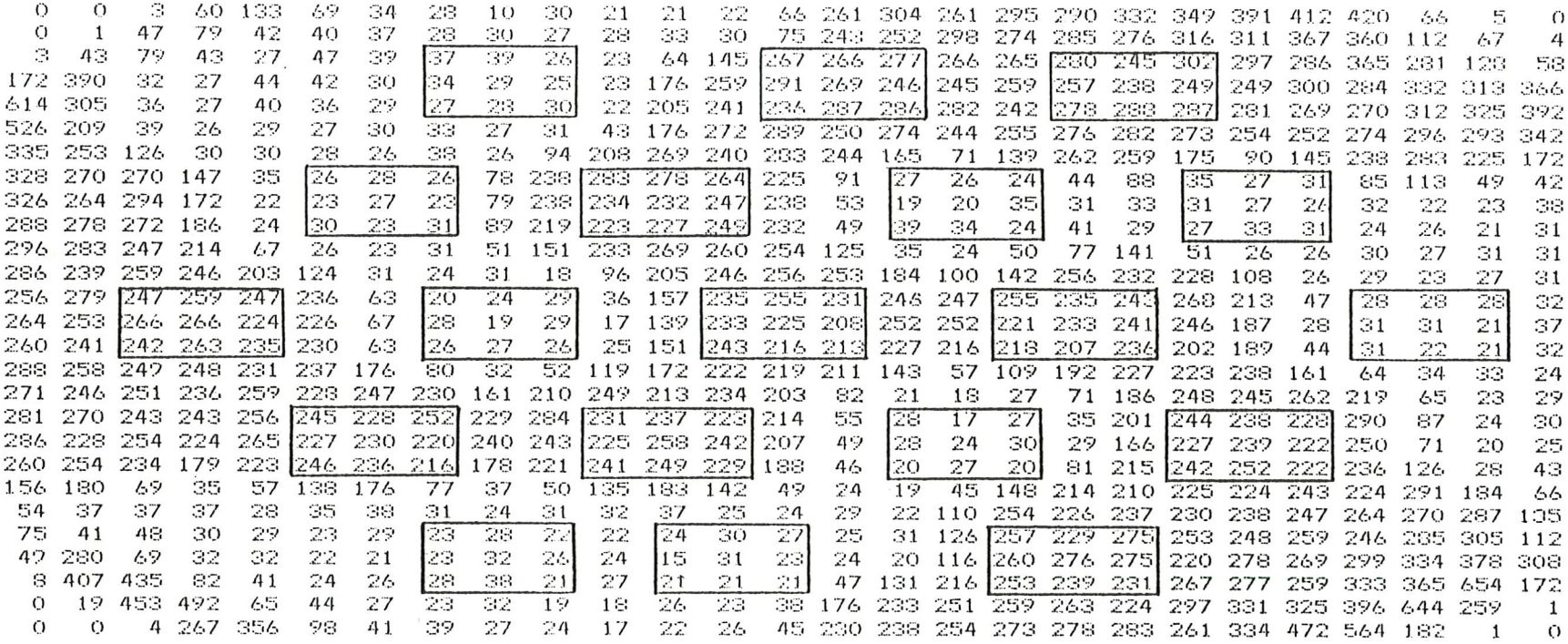


Fig. 6.2. *Numerical matrix of detector image corresponding to* $\alpha = 0°$ *and* $A_r = 23.9$ *for the SPD (fig. 6.1 (d)).*

use of antimask also reduces the coding error in precisely the same way as does the plateau removal technique described in section 6.7.1 (step (a)).

All simulations assumed, as before, a detector efficiency and opaque mask element opacity of 100%, and no systematically varying background. Also the source was always considered to lie at the centre of a sky pixel.

6.6 Results

6.6.1 Square Pixel Detector

For the SPD the gamma camera pixels were rebinned into three different position resolutions. From the analysis of the recorded data a source contribution of 221 photons/pixel and a residual background of 27 photons/pixel were calculated for $A_r = 23.9$. These values, properly scaled for different values of A_r, have been used for the correlated simulations and to calculate the expected SNR values.

Table 6.1 shows the complete results for a source in the centre of the field of view imaged with a stationary mask at 12 different mask orientation angles and with a rotating mask. Here the SNR values are shown vs. α for three different values of A_r ($A_r = 7.8, 23.9$ and 42.4), along with the SNR for the rotating mask. Also shown are SNR values, with standard deviations, for the simulations. The main features of the results are more evident in the graphs of figs. 6.3, 6.4 and 6.5 where the SNR vs. α is plotted for the three different position resolutions. The trend of SNR vs. α obtained from the experimental data is consistent with that of the simulations for all three cases. The SNR varies with α by as much as 50% from the average value, and is well below the expected value of SNR_{exp}=206. Because the source is at the centre of the field of view an approximate periodicity of 30° is observed. Different behaviour is observed if the data is analysed assuming a different source position (i.e. using data from a different part of the detector). The reduction of the SNR with respect to SNR_{exp} is mainly due, for low resolution, to the coding error noise. However, for higher resolutions, a strong contribution of some systematic effects inherent in the gamma camera become evident, as the SNR does not increase as much as for the simulations (fig. 6.5). This is particularly well illustrated in fig. 6.6 showing the SNR of rotated mask images versus A_r. The graph shows clearly that no increase in SNR from $A_r = 23.9$ to $A_r = 42.4$

Mask orien- -tation	SNR stationary mask (SNR_{exp}=206)					
	$A_r = 7.8$		$A_r = 23.9$		$A_r = 42.4$	
	Experim.	Simul.	Experim.	Simul.	Experim.	Simul.
0°	13.5	16.0 ± 0.3	34.3	40.8 ± 1.2	27.0	45.0 ± 1.7
10°	25.5	30.5 ± 0.3	25.7	34.7 ± 0.6	36.7	87.1 ± 14.6
20°	22.1	29.3 ± 0.8	23.3	35.2 ± 0.6	34.4	75.6 ± 4.1
30°	15.0	17.8 ± 0.3	30.6	40.3 ± 1.1	28.4	48.1 ± 2.8
40°	25.3	30.8 ± 0.2	26.6	34.5 ± 0.8	42.2	84.1 ± 7.8
50°	24.0	29.0 ± 0.5	25.0	36.0 ± 0.9	34.9	70.1 ± 8.7
60°	14.9	18.0 ± 0.2	28.1	43.6 ± 1.7	27.6	50.3 ± 1.9
70°	22.8	31.0 ± 0.8	23.9	33.8 ± 0.7	30.4	83.4 ± 4.6
80°	26.5	29.5 ± 0.8	24.6	34.7 ± 1.0	36.2	73.6 ± 1.5
90°	14.7	17.7 ± 0.3	35.3	41.6 ± 1.7	29.6	51.9 ± 3.1
100°	26.2	30.9 ± 1.1	26.5	34.2 ± 0.8	37.9	86.9 ± 4.3
110°	23.8	29.3 ± 1.0	25.4	35.4 ± 1.1	38.2	74.6 ± 3.0
α	Rotating mask (SNR_{exp}=505)					
0°-50°	46.4	61.4 ± 1.9	57.3	96.9 ± 1.6	58.9	145.0 ± 7.7
60°-110°	44.0	62.0 ± 0.7	53.7	99.0 ± 1.0	56.0	150.2 ± 8.4

Table 6.1 *SNR of the SPD for 3 different values of A_r at 12 different values of α for both simulated and experimental data using normal deconvolution. Results for both stationary and rotating mask are shown.*

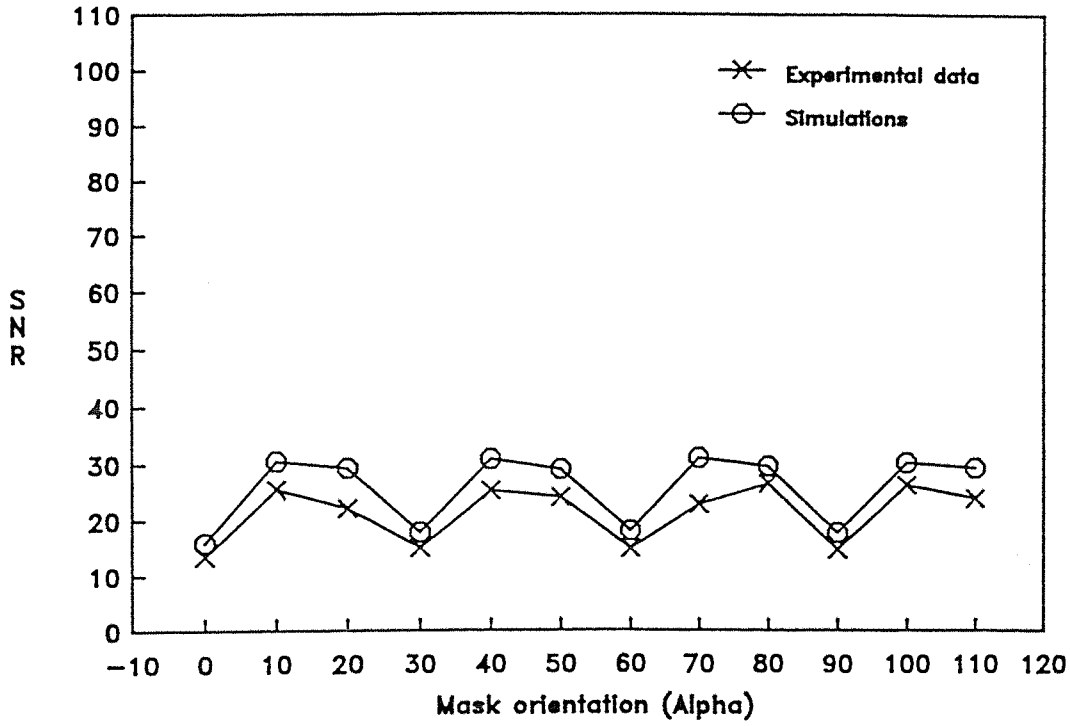


Fig. 6.3. *SNR versus α for the SPD for $A_r = 7.8$. Experimental and simulated results are shown.*

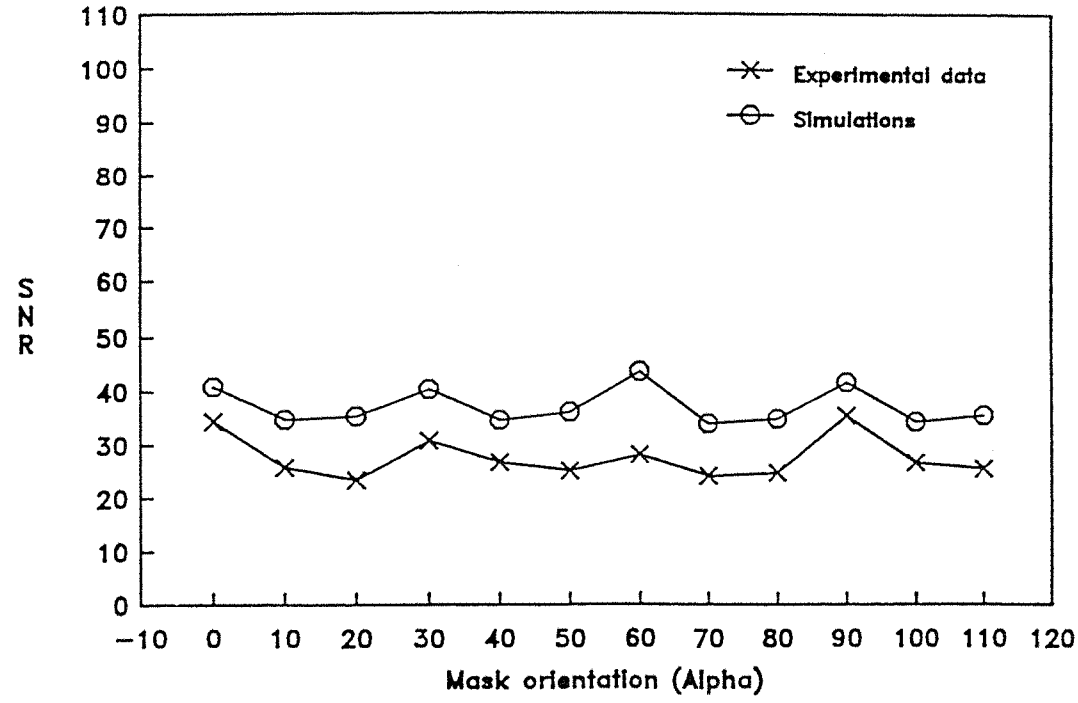


Fig. 6.4. *SNR versus α for the SPD for $A_r = 23.9$. Experimental and simulated results are shown.*

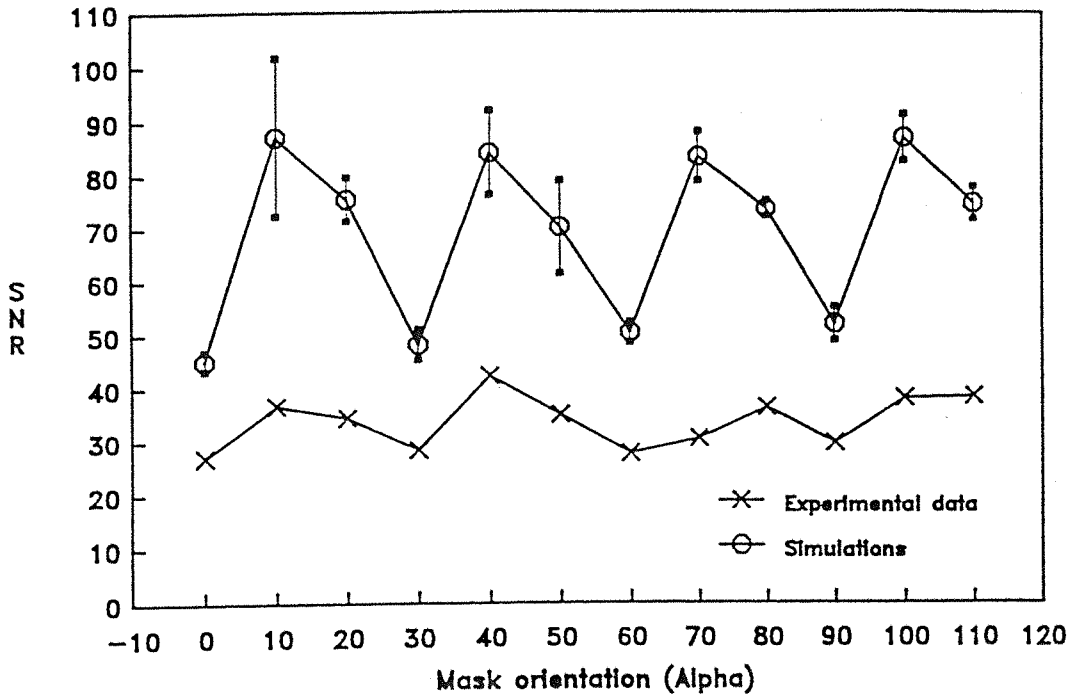


Fig. 6.5. *SNR versus α for the SPD for $A_r = 42.4$. Experimental and simulated results are shown.*

is observed, and the difference between SNR for experimental images and for simulations increases with A_r. However, using antimask images (fig. 6.7) reduces the systematic effect and the experimental data results approximate better to the simulation results, and the SNR increases linearly with A_r. The use of antimask does not correct the systematic effects in the source photons and this may explain the residual difference observed between experimental and simulated images.

6.6.2 Hexagonal Pixel Detector

When the configuration of a HPD with $A_r = 19$ was used to rebin the gamma camera pixels before deconvolution as described in section 6.4.2, approximately 268 photons/pixel were found to be the source contribution and 32 photons/pixel were found to be the background contribution. The resulting SNR values obtained both for stationary mask observations and for mask and antimask observations are shown in table 6.2, along with simulations and corresponding errors. Also shown are values using a rotating mask.

Large differences between the experimental and the simulation values are observed, particularly for $\alpha = 0°$ and 60°. At these angles, because of the absence of coding error noise the SNR of the simulated images approaches the value expected in the absence of any systematic noise (SNR_{exp}=206). Although at $\alpha = 0°$ the observed SNR reaches a maximum there is no clear evidence in the experimental data that at this angle coding error is absent. Moreover at 60°, where another peak is expected, the SNR is not as high as that at 0°. The reason for these discrepancies is the systematic detector noise which lowers the value of SNR below that of SNR_{exp}. However when the antimask procedure is applied (last two columns of table 6.2) the large increase in the SNR at $\alpha = 0°$ becomes visible due to the reduction of systematic noise intrinsic in the gamma camera when there is no coding error. Fig. 6.8 shows the trend of the experimental SNR vs. α compared to the simulation values when antimask is used. All the points at $\alpha \neq 0°$ are almost consistent with the simulations. At these angles the coding error is the main source of noise and it affects both the experimental and simulated images in the same way. At $\alpha = 0°$ the coding error is not present and also the systematic detector background noise is reduced by antimask. The difference observed here between the experimental and the simulated images can be attributed to the residual systematic noise in the detector source photons which the use of antimask cannot reduce.

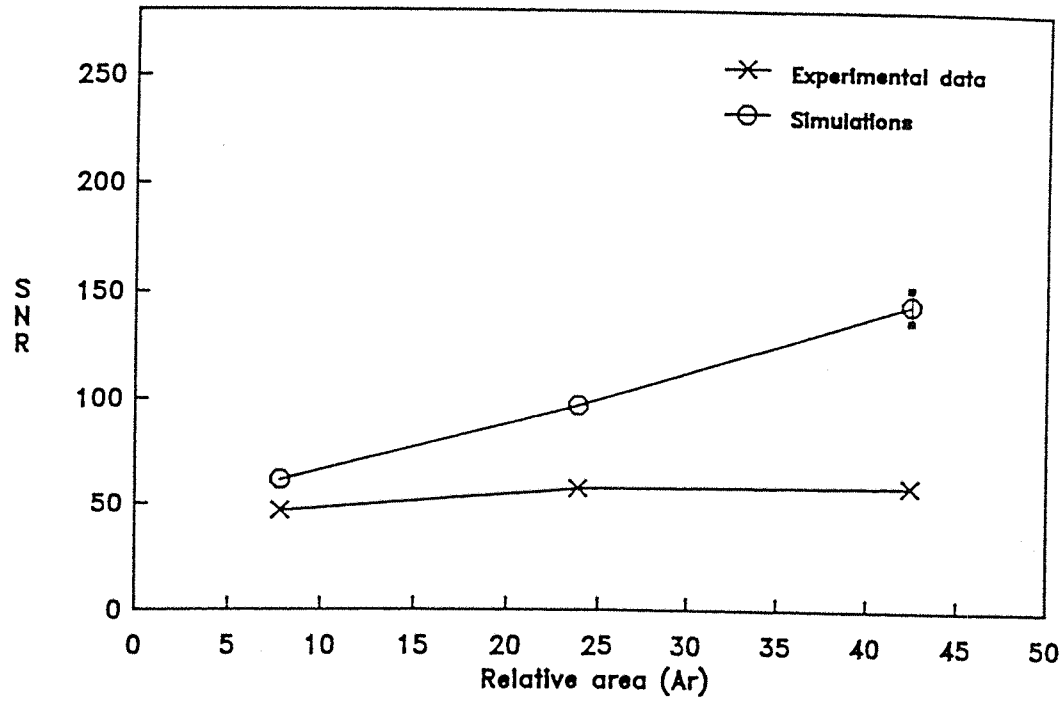


Fig. 6.6. *SNR versus A_r for the SPD for rotating mask (experimental and simulations).*

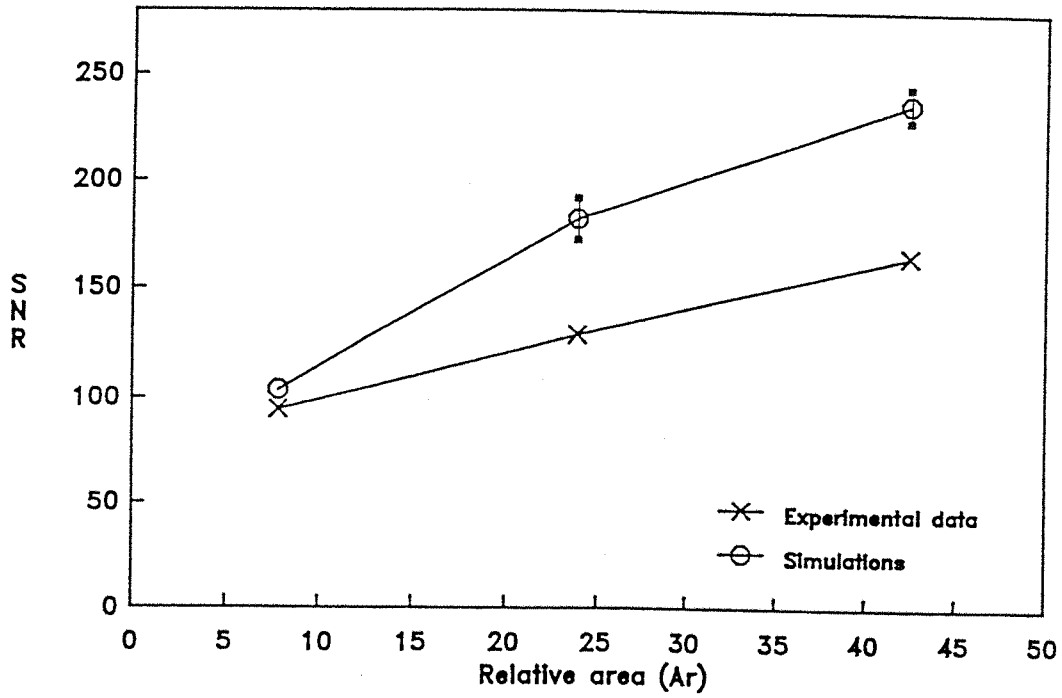


Fig. 6.7. *SNR versus A_r for the SPD for rotating mask and antimask (experimental and simulations).*

Mask orientation α	SNR			
	Mask only (SNR_{exp}=205)		Mask and antimask (SNR_{exp}=290)	
	Experimental	Simulations	Experimental	Simulations
0°	49.1	186.5 ± 31.7	116.7	287.7 ± 18.5
10°	30.9	42.9 ± 0.6	47.4	52.0 ± 1.8
20°	25.5	44.0 ± 1.2	45.0	54.8 ± 1.0
30°	43.2	49.5 ± 1.3	74.3	86.4 ± 3.5
40°	38.8	36.8 ± 2.0	77.8	84.5 ± 1.1
50°	40.1	54.1 ± 3.0	64.4	71.4 ± 2.1
60°	38.4	190.3 ± 13.2		
70°	28.0	43.0 ± 1.6		
80°	26.2	43.4 ± 2.2		
90°	46.7	51.6 ± 2.0		
100°	45.5	39.0 ± 1.7		
110°	42.9	54.4 ± 1.6		
Rotating mask				
0°-50°	72.2	126.8 ± 5.2	131.1	200.9 ± 6.6
60°-110°	76.0	128.9 ± 3.7		

Table 6.2 *SNR of the HPD at $A_r = 19$ both with and without the use of antimask. Experimental and simulated data are given both for stationary and rotating mask.*

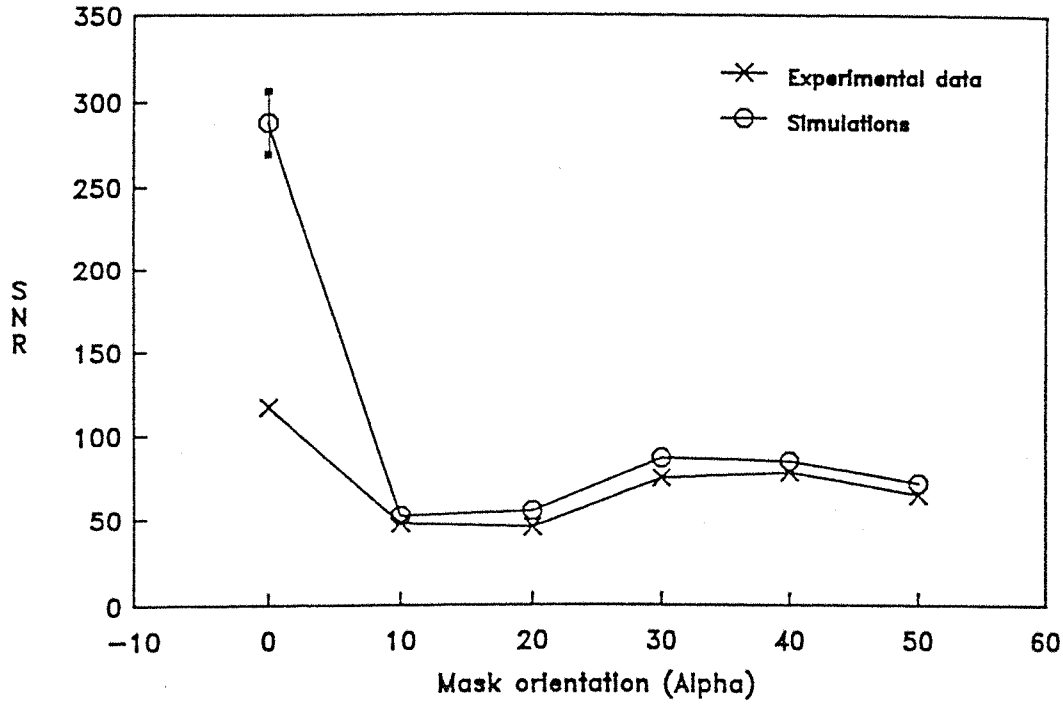


Fig. 6.8. *SNR versus α of the experimental images for the HPD ($A_r = 19$) using antimask. Simulation values are also shown.*

6.7 Improving Image Quality

The noise due to coding error presents a serious problem when the mask element shadows are not perfectly binned by the detector pixels, particularly so when low source fluxes are experienced. What follows is an attempt to improve the image quality by a two step process: (a) plateau removal, and (b) weighted deconvolution. The ideas behind the use of these two steps are now explained.

6.7.1 Step (a). Plateau Removal

One of the main conclusions drawn from the analysis of imaging a point source at different background levels (see section 5.4.2) was that the coding error noise is highly magnified by the background level. Therefore, by this reasoning, it should be possible to remove a constant background level, i.e. perform a *plateau removal*, from every detector pixel prior to deconvolution in order to get an improved SNR. This has in fact been shown to be the case. Note that for an optimum coded system (i.e. perfect binning of mask element shadows by the detector pixels), a uniform noise level does not require such a plateau removal because the only effect is the addition of a corresponding dc level in the deconvolved image. However in the case where there is coding error, any background dc level increases this error and thus decreases the SNR.

The optimum value of removal of a plateau can be generalised, and it depends on the values of the source and background fluxes per detector pixel. Figs. 6.9 and 6.10 show the SNR vs. plateau removal values for two non statistical point sources in different non statistical background levels using the SPD. The absence of statistics in this analysis is used to show the optimum plateau removal values. In both cases the source flux was 100 photons/pixel and $A_r = 21.7$. Fig. 6.9 (a) is for $\alpha = 0°$ and a background flux of 50 photons/pixel, whilst fig. 6.10 (a) is for $\alpha = 10°$ and a background flux of 100 photons/pixel. Graphs in figs. 6.9 (b) and 6.10 (b) show corresponding fine structure about the approximate peak SNRs. Although there is a slight fluctuation depending on the configuration it is clear that the optimum value to subtract from each pixel is approximately R_p where

$$R_p = D_B + \frac{D_S}{2} \tag{6.1}$$

and D_B is the mean background photons/pixel on the detector, and D_S is the mean source photons/illuminated pixel on the detector. In a practical situation however, the values of

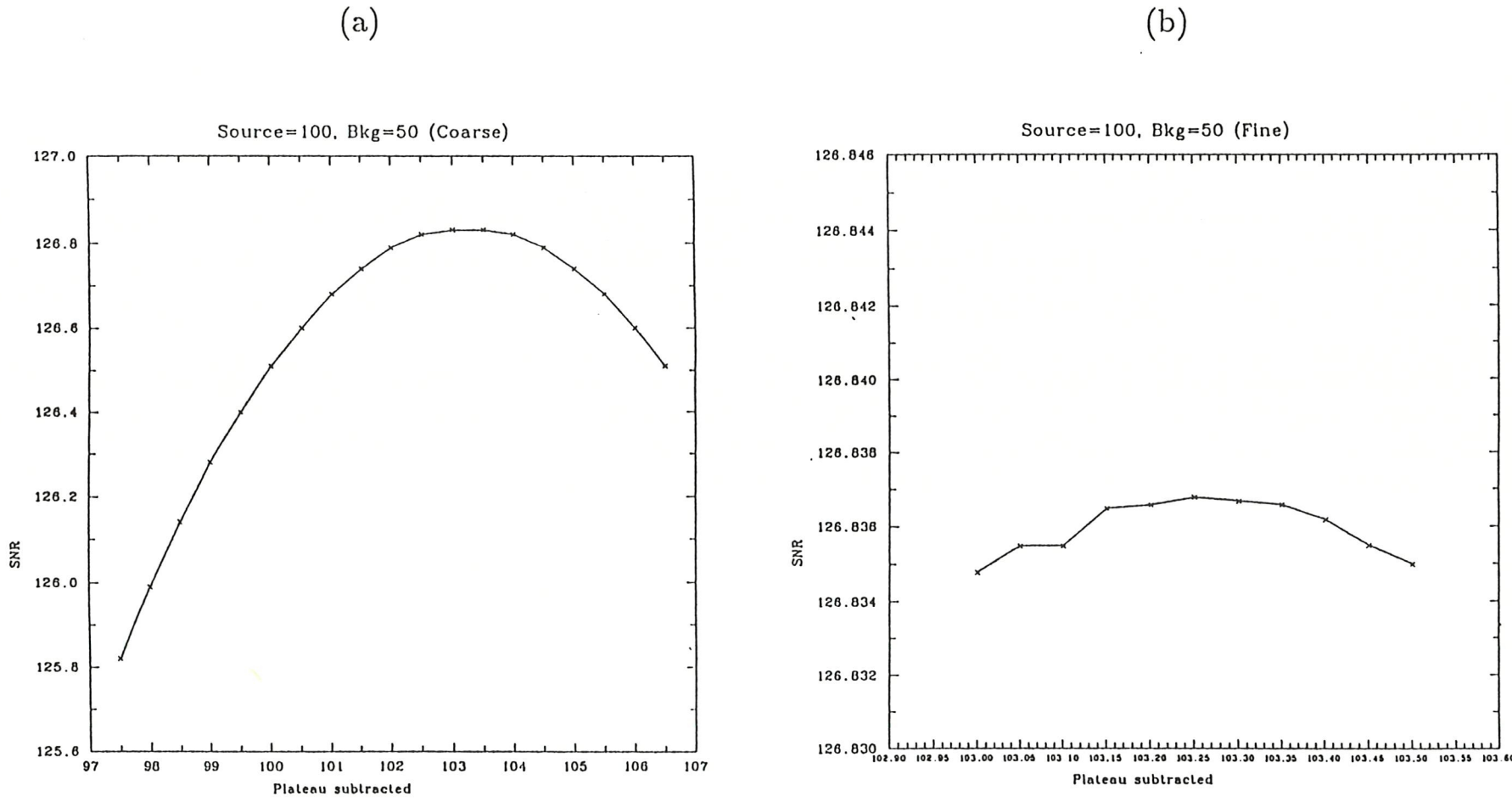


Fig. 6.9. *SNR versus plateau removal for the SPD at $A_r = 21.7$ and $\alpha = 0°$. Source flux = 100 photons/pixel, background flux = 50 photons/pixel: (a) coarse structure, (b) fine structure about SNR peak.*

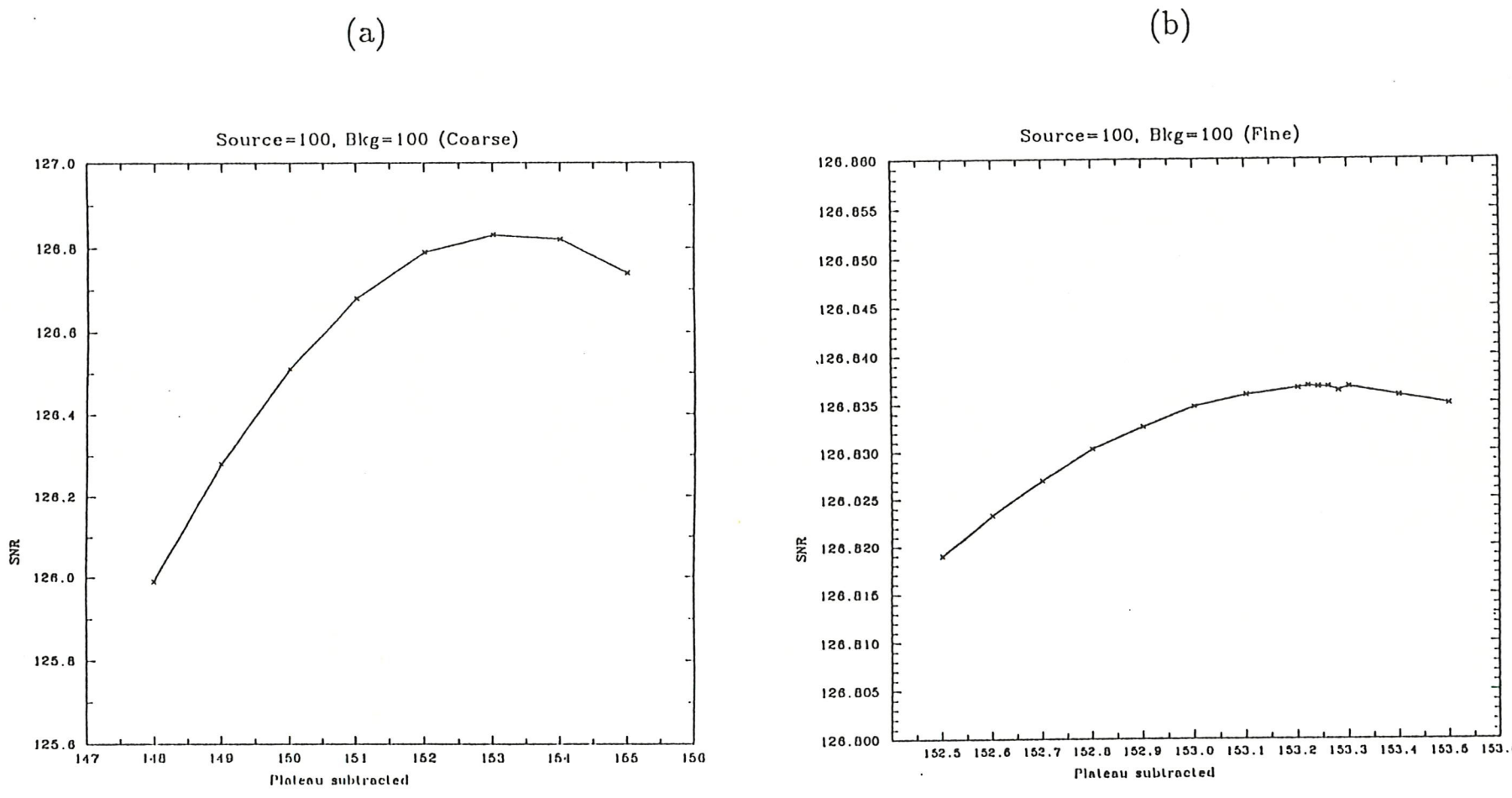


Fig. 6.10. *SNR versus plateau removal for the SPD at $A_r = 21.7$ and $\alpha = 10°$. Source flux=100 photons/pixel, background flux = 100 photons/pixel: (a) coarse structure, (b) fine structure about SNR peak.*

D_B and D_S are not known. Therefore it is a sufficiently good approximation for most purposes if R_p takes on the value of the mean detector pixel count level. The detector plane image before deconvolution therefore takes the form of fig. 6.11, where the detector plane is represented in one dimension. Note that this procedure reduces the coding error in the same way as the use of antimask when there is no systematic noise (see section 6.5).

6.7.2 Step (b). Weighted Deconvolution

The coding error noise depends both on the recorded detector image and on the deconvolution array. The deconvolution array G is determined by the system configuration, i.e. by the relative area A_r and the mask orientation angle α. Fig. 5.7, shows clearly that by using the previous definition for the deconvolution array, the resulting array is only an approximation to the real hexagonal mask pattern. In particular different opaque or transparent areas do not have the same shape and this introduces errors in the deconvolution. However it is possible to build an array G which takes into account that some pixels, when projected back in the cross correlation, might fall between two or even three mask elements and that they should be weighted before being added or subtracted. The balanced correlation deconvolution array G of this *weighted deconvolution* algorithm is therefore given by:

$$G(i,j) = +(F_o(i,j) - F_c(i,j)) \tag{6.2}$$

where $F_o(i,j)$ is the fraction of the (i,j) pixel area which is projected onto an open mask element, and $F_c(i,j)$ is the fraction projected onto a closed one. This can be thought of as an extension of the standard cross correlation because when the hexagonal mask elements are completely binned by the pixels $G(i,j)$ assumes only the values of +1 or -1. When the antimask images are used the fraction of detector pixel area projected on the central element, or one of its repetitions, does not contribute to the value of G (i.e. this fraction is taken as zero).

6.7.3 Results

Both of the above steps have been applied to the data in order to reduce the effect of the coding error. The results were very similar for both the SPD and the HPD, with the main principles being illustrated well by the HPD results. The results obtained both for the experimental and the simulated images in the case of the hexagonal pixel rebinning

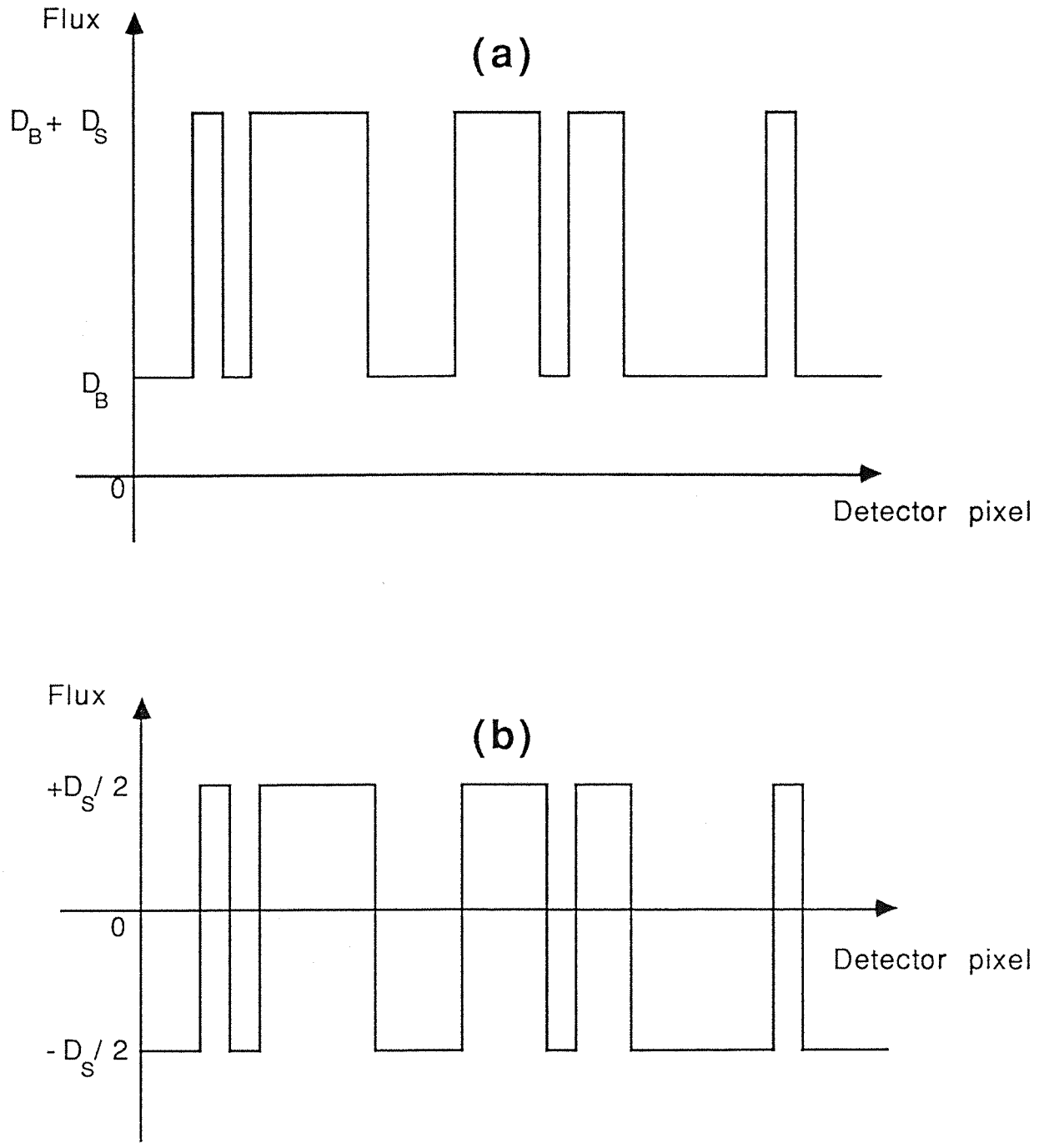


Fig. 6.11. *Schematic one dimensional illustration of the optimum plateau removal prior to deconvolution: (a) before plateau removal, (b) after plateau removal.*

with $A_r = 19$ are shown in table 6.3. The values in the 2nd and 3rd columns (plateau removal, step (a)) of table 6.3 are the SNRs of the images obtained using plateau removal (step(a)) only, whilst the next two columns show the results using both plateau removal and weighted deconvolution (steps (a) and (b)). The final two columns show results obtained using antimask and weighted deconvolution.

Comparing the values of simulated antimask images of table 6.2 and the values of the simulations using plateau removal (3rd column of table 6.3) shows that the effect of the use of antimask on the coding error is equivalent to the use of plateau removal before deconvolution. The graph in fig. 6.12 shows the HPD experimental results for three cases: normal deconvolution, step (a) only, and steps (a) and (b). The graph shows how the SNR increases when applying the steps of the deconvolution for different mask orientations. From fig. 6.12 it can be seen that at $\alpha = 0°$ and $\alpha = 60°$ no increase in SNR is achieved using step (a) or steps (a) and (b), and this confirms that no coding error is present at these angles and the SNR is limited only by the systematic gamma camera background noise. The systematic noise is also responsible for the fact that the increase in SNR is much lower than the corresponding increase observed in the simulated images. The graph in fig. 6.13 is the SNR vs. α using antimask images for the HPD. The curves shown are for experimental images both with and without the use of steps (a) and (b), and for the simulations using steps (a) and (b). The graph shows that by using antimask, and therefore reducing external systematic effects, the coding error is highly reduced in the laboratory images, with the exception of when $\alpha = 0°$ where there is no coding error. The difference in experimental and simulated values is explained by the fact that here the main source of noise for the experimental images is no longer the coding error, but the residual systematic effects of the gamma camera from the source photons which antimask cannot reduce. At $\alpha = 0°$ no coding error is present and the simulated image SNR is close to the expected value for statistical noise only (SNR_{exp}=290), while the experimental value is not improved by the use of steps (a) and (b) and is limited to $\sim$ 120 by systematic effects.

Use of steps (a) and (b) for rotated mask images is even more effective. Again the experimental results are limited by external systematic effects but the simulation values (see table 6.3) show that the SPSF reaches a very high SNR ($\sim$ 300) because the coding error is smoothed by the mask rotation.

Plateau removal and weighted deconvolution have been applied to the experimental data images and simulated images assuming a SPD. When the difference in positional resolution

Mask orient--ation α	SNR					
	Mask only				Mask and antimask	
	Step (a)		Steps (a) and (b)		Weighted deconv.	
	Experim.	Simul.	Experim.	Simul.	Experim.	Simul.
0°	49.1	197.9 ± 24.4	48.5	198.9 ± 15.3	117.5	264.6 ± 15.8
10°	38.7	53.5 ± 1.5	42.1	102.8 ± 6.1	66.0	110.3 ± 3.9
20°	31.0	55.5 ± 1.6	35.6	127.5 ± 10.7	76.4	156.5 ± 19.1
30°	45.0	78.8 ± 5.9	50.4	130.3 ± 14.9	117.3	169.1 ± 16.0
40°	43.6	80.4 ± 3.1	47.3	129.0 ± 7.8	105.6	143.4 ± 16.7
50°	39.8	68.9 ± 2.6	42.9	137.1 ± 8.7	94.4	143.2 ± 12.5
60°	38.4	194.2 ± 37.3	38.0	188.4 ± 41.7		
70°	33.7	51.4 ± 1.0	35.5	100.2 ± 2.0		
80°	31.7	55.1 ± 1.0	38.9	132.0 ± 18.9		
90°	49.0	87.0 ± 5.4	59.3	143.0 ± 8.8		
100°	47.5	82.0 ± 3.7	53.1	119.9 ± 12.7		
110°	43.1	64.6 ± 2.9	48.1	135.0 ± 9.6		
Rotating mask						
0°-50°	70.0	194.7 ± 13.7	72.0	306.4 ± 20.3	180.0	364.6 ± 48.9
60°-110°	74.6	192.8 ± 6.6	76.5	319.2 ± 20.3		

Table 6.3. *SNR of the HPD for $A_r = 19$ using plateau removal only (step (a)), and both plateau removal and weighted deconvolution (steps (a) and (b)) on the simulated and experimental data. Stationary and rotated mask results are given.*

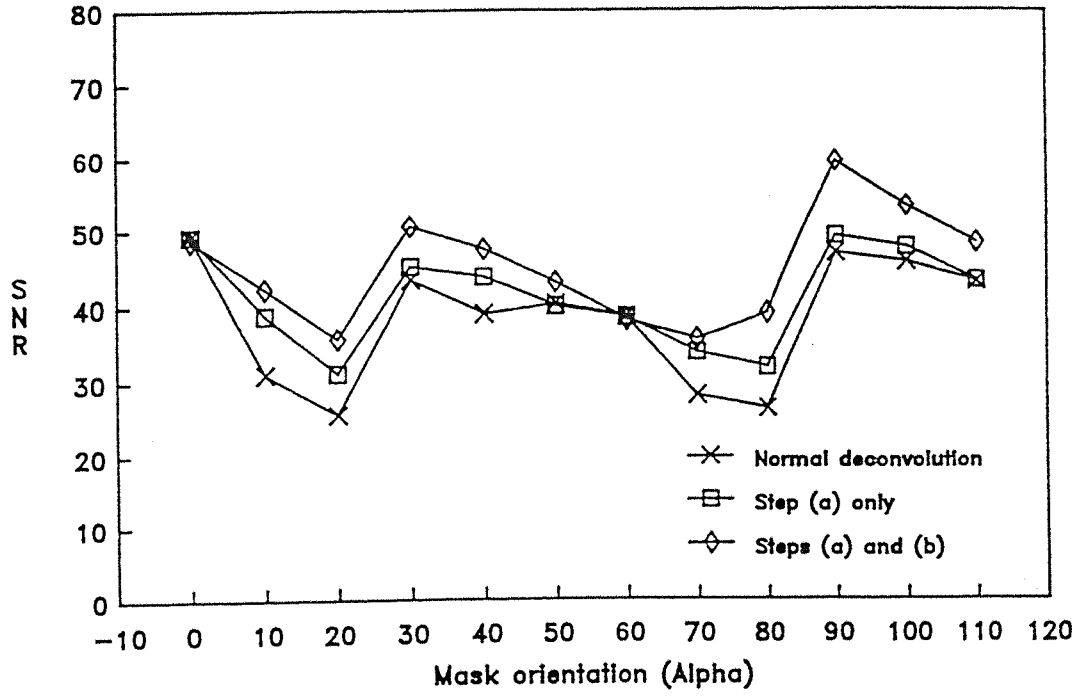


Fig. 6.12. *SNR versus α of experimental images for the HPD with $A_r = 19$ when using standard deconvolution (crosses), plateau removal (step (a)) only (squares), and plateau removal and weighted deconvolution (steps (a) and (b), diamonds).*

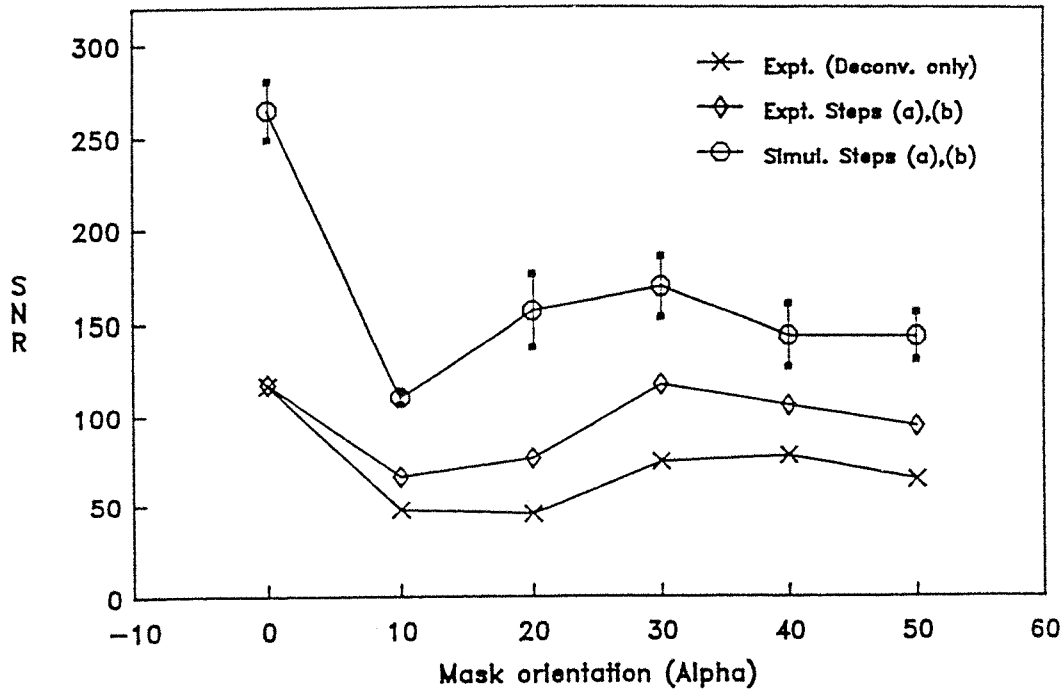


Fig. 6.13. *SNR versus α of experimental images for the HPD with $A_r = 19$ using antimask for both standard deconvolution (crosses), and when using steps (a) and (b) (diamonds). For comparison the SNRs of the simulated images for steps (a) and (b) are also shown (circles).*

is taken into account the results agree well with those of the HPD.

6.8 Discussion and Conclusions

When antimask images are used to reduce the systematic effects due to the experimental conditions the laboratory tests of a coded aperture system consisting of a HURA mask and a discrete pixel detector confirm the results obtained by correlated computer simulations. Moreover, these results also confirm the analysis of the SPSF in chapter 5. The coding error due to the mismatch between mask element shadows and detector pixels introduces a coding error noise which highly reduces the quality of the SPSF. The noise has been found to vary rapidly and irregularly with the mask orientation, and, to a lesser extent, with the source position, and is found to decrease when the positional resolution is increased. The observed variations are as predicted by the computer simulations and no large differences have been observed between square and hexagonal pixel detectors for general mask orientations or for a rotating mask. However with a HPD it is possible to find a configuration for which at a certain mask orientation the mask elements are nearly perfectly binned by the hexagonal detector pixels, and when the standard cross correlation deconvolution is applied, no coding error is present. It has been found that rebinning the camera pixels into hexagonal pixels for this configuration raises the SNR of the experimental images to large values, but does not reach the expected SNR value for just statistical noise only because of external systematic effects which cannot be completely reduced by the antimask procedure. The experimental data seems to confirm that this configuration gives an optimum coded system. For rotating masks however, or when a proper design of the coded system is not feasible, the simple cross correlation deconvolution must be replaced by a more suitable algorithm in order to give a higher image quality. Several advanced techniques for the deconvolution and restoration of images obtained with non optimum coded systems have already been developed [159] but they involve the use of time consuming algorithms and usually require large quantities of computing power. The algorithm that has been developed here is to perform a subtraction of the mean count level per detector pixel, and then a cross correlation of the detector image with a weighted deconvolution array which depends on the mask orientation. The use of this technique in deconvolving the experimental images has been shown to be very effective in reducing the coding error and therefore in improving the quality of the deconvolved image.

For stationary masks, the optimum choice for a coded aperture system consisting of a HURA

mask and a discrete pixel detector is a system with a configuration of hexagonal pixels and a properly oriented mask as explained in section 5.4.3, for which no coding error is present and the SPSF is a perfect triangular function. This configuration has been adopted for the coded aperture system of the GRASP telescope, in which a stationary HURA mask (of basic pattern $v = 79$) is coupled to a circular composite detector plane consisting of caesium iodide hexagonal bars and circular germanium spectrometers (see chapter 7). For rotating masks the use of plateau removal and weighted deconvolution, and the effect of the rotation combine to reduce almost completely the effect due to the coding error, and the pixel geometry of the detector plane (for equivalent positional resolutions) is no longer of importance for the performance of the system.

Chapter 7

Gamma Ray Astronomy with Spectroscopy and Positioning. The GRASP Telescope.

7.1 Introduction

Gamma ray astronomy is just beginning to emerge from the exploratory phase and is now being established as an observational science. Telescopes such as the balloon borne Gamma Ray Imaging Payload (GRIP) designed at the California Institute of Technology [4] and the New Hampshire Directional Gamma Ray Telescope (DGT) [42] have led to advances in the knowledge of the high energy photon emission from celestial objects. However there is still a need for a gamma ray astronomy mission devoted both to imaging and spectroscopy.

Imaging at the sub-MeV energy level is to be performed with moderate spectral resolution using the SIGMA telescope [94]. At higher energies the Gamma-1 mission [3], although advancing the study of COS-B sources, will probably suffer from orbit and lifetime limitations. The Gamma Ray Observatory (GRO) [15] will have excellent sensitivity in the gamma ray region both in the low energy (COMPTEL) and high energy (EGRET) ranges, but will carry telescopes having a limited angular resolution in each waveband.

The GRASP gamma ray astronomy project [20,58] hopes to provide both imaging and spectral data from a wide range of objects in order to advance further the understanding of the astrophysics of those objects which have already been detected, and to discover new objects in the high energy domain. The energy range of GRASP will be from 4 keV to $\gtrsim$100 MeV, with an expected energy resolution of $E/\Delta E \simeq 500$ at 1 MeV. It is expected that the instrument will locate celestial point sources to within 1 arcminute resolution inside a field of view of 50 square degrees.

7.2 Scientific Objectives

The scientific objectives of the GRASP mission can be split into three main sections: solar observations, Galactic observations and extragalactic observations.

7.2.1 Solar Observations

The GRASP solar observations will mainly be of phenomena related to solar flares. To date high resolution gamma ray line spectroscopy has never been applied to radiation from solar flares. It is hoped that the GRASP germanium spectrometer will provide for spectral observations of flares, hence giving detailed information about the energy released by accelerated particles such as protons in ^{12}C by the isospin zero (4.438 MeV line) and one (15.1 MeV line) levels. Also, GRASP will study the earth's albedo from the interaction of solar particles with the earth's atmosphere.

7.2.2 Galactic Observations

GRASP is expected to make a large number of observations of objects within the Galaxy. Of particular importance is the Galactic centre where there are very strong radio and infrared sources known to exist, but in the gamma ray region the situation is unclear. The origin of the continuum emission from the Galactic centre is only known to within 10° but GRASP should give more accurate position locations of the sources causing the emission. GRASP also hopes to map the Galactic centre line emission from the 511 keV electron-positron annihilation, and the 1.809 MeV ^{26}Al line observed by the Max Planck Institut Compton

telescope [152].

Gamma ray emission from the interstellar medium will be observed by GRASP. Such emission has three components. The first is a continuum component from interactions of high energy cosmic rays with interstellar matter, the second is a line emission from interactions of low energy cosmic ray nuclei with interstellar matter, and the third is also a line emission, from radioactive nuclei produced during various nucleosynthesis processes in certain celestial objects which then diffuse into interstellar space.

The study of compact objects emitting gamma rays will be carried out using GRASP. The main underlying theme is the study of the physics of neutron stars, since they appear in a range of compact objects such as radio pulsars, X ray binaries, possibly gamma ray bursters, and "Geminga-like" objects. Fine spectroscopy of nuclear lines (in particular the 511 keV annihilation line) from neutron stars, which will be possible using GRASP, enables the measurement of the surface gravity by means of redshift, putting significant constraints on the allowable types of neutron star equation of state [85].

The search for radioactive nuclei represents a test for the theories of heavy element production by explosive nucleosynthesis in novae and supernovae. GRASP hopes to carry out observations of known supernova remnants and also to discover recent (<100 years) supernova remnants which are, as yet, undetected. Such observations will provide answers to the question of explosive nucleosynthesis by studying the line emission from the ^{56}Ni - ^{56}Co - ^{56}Fe decay chain at 0.847 MeV (probably from other galaxies, as ^{56}Co has the relatively short half life of 77 days), and from the ^{44}Ti - ^{44}Sc - ^{44}Ca decay chain at 0.068 MeV, 0.078 MeV and 1.156 MeV. Of particular interest will be repeated observations of the recent nearby supernova SN1987A observed in the Large Magellanic Cloud. So far a number of balloon borne gamma ray observations have been made of SN1987A, detecting ^{56}Co at 0.847 MeV, 1.238 MeV and 2.599 MeV. The emission of such gamma ray lines is encouraging for the observation of future supernovae by GRASP. Hydrogen burning at high temperatures in nova explosions is expected to produce significant amounts of both ^{22}Na and ^{26}Al whose subsequent decay in the interstellar medium provides line emission at 1.275 MeV and 1.809 MeV respectively. Observations of these lines by GRASP will provide important information guiding theoretical studies of nova events.

7.2.3 Extragalactic Observations

The observations of extragalactic objects by GRASP will include active galaxies and, in particular, active galactic nuclei (AGN). So far only a small number of AGNs have been detected because extragalactic gamma ray astronomy is still in its infancy. GRASP will observe known active galaxies and quasars, and also may detect new AGNs in the large field of view whilst observing other objects of interest. The penetrating nature of the high energy photons ($\sim$ 1 MeV) make them very useful for studying the region close to the centre of the galaxy and therefore essential if the physical processes inside AGNs are to be understood.

As well as AGNs, normal galaxies, like our own, will be a GRASP target. Supernovae in nearby galaxies will be searched for to study explosive nucleosynthesis processes. Individual bright sources inside galaxies in the local group (<1Mpc) are hoped to be observed, the three main candidates being M31 and the Large and Small Magellanic Clouds. It should be possible to detect inside all three of these galaxies both the diffuse 1.809 MeV ^{26}Al emission line, and the continuum emission resulting from cosmic ray interactions with the interstellar medium. Einstein data has shown the presence of many bright X ray sources near the central region of M31, whilst the Magellanic clouds have their own targets of interest, such as the bright X ray binary LMC X-1 and the black hole candidate LMC X-3, as well as SN1987A.

A number of clusters of galaxies (e.g. Virgo, Perseus, Coma, 3C129, A2142) have been observed by hard X ray telescopes and the preferred two component model consistent with the data suggests the presence of one or more AGNs in the cluster [81] (see also section 1.5.4). The imaging capability of GRASP will enable location of any such AGNs thus explaining the hard X ray flux. Also, measurement of the power law flux by GRASP may be used to set a lower limit to any magnetic fields associated with a cluster.

With GRASP it should be possible to study cosmology by using the 511 keV electron-positron annihilation line. The presence of a massive black hole and its associated accretion disk in AGNs could be a source of intense positron production. Therefore the annihilation line may well be detectable by GRASP, whose energy resolution allows measurement of any redshift of this line for AGNs with $Z \geq 0.01$. This covers the majority of X ray emitting Seyferts, BL-Lacs and quasars. For example 3C273 has a redshift $Z = 0.158$ which will produce an annihilation line at $\sim$ 430 keV if the redshift is entirely cosmological. Comparison of GRASP high energy data with existing optical data may prove whether the redshift is purely cosmological, or whether there is a non-cosmological component in the redshift.

The measurement of the absolute flux of the cosmic diffuse background has always proven difficult because of contamination from intrinsic background associated with the detector and the environment. However, it is intended to use GRASP to measure the cosmic diffuse background by lunar occultation of parts of the field of view, as explained in section 1.5.3. It is expected that the germanium detectors on board GRASP will measure the background in the energy range 4-200 keV. It is possible that the sun may also be used for the same purpose during flare minimum.

7.3 The GRASP Telescope Configuration

The way in which the GRASP telescope will differ from any other gamma ray astronomical instrument to date will be its ability to perform both fine spectral analysis and high angular resolution observations whilst maintaining a wide field of view. In order for such a device to function in this manner with current technology available, it is necessary to use two detector types, one for spectroscopy and one for imaging. The GRASP spectrometer will be an array of germanium detectors, which are surrounded by an annulus of imaging bars made of caesium iodide. These two components constitute the useful detector plane which is shielded from unwanted events, such as charged particle events, by an active anticoincidence scintillator, or *veto system.* Four metres above the detector plane is the tungsten coded aperture for source modulation [58].

A schematic view of the GRASP telescope is shown in fig. 7.1, with performance parameters of each detector system given in table 7.1, and the main components described below.

7.3.1 Germanium Spectrometer

A significant problem to overcome in designing the germanium spectrometer is to minimise as much as possible the background noise. For the case of a well shielded germanium system, and over the spectral range 200 keV to a few MeV, the major contribution to the background noise is due to beta decays derived from particle activation of the germanium crystals. The path length of beta decay electrons within the germanium is very short ($\sim$ few mm) and may be considered as *single site events* (fig. 7.2). In contrast, above 200 keV, the gamma ray photons are absorbed in germanium by multiple spatially separated interactions, consisting

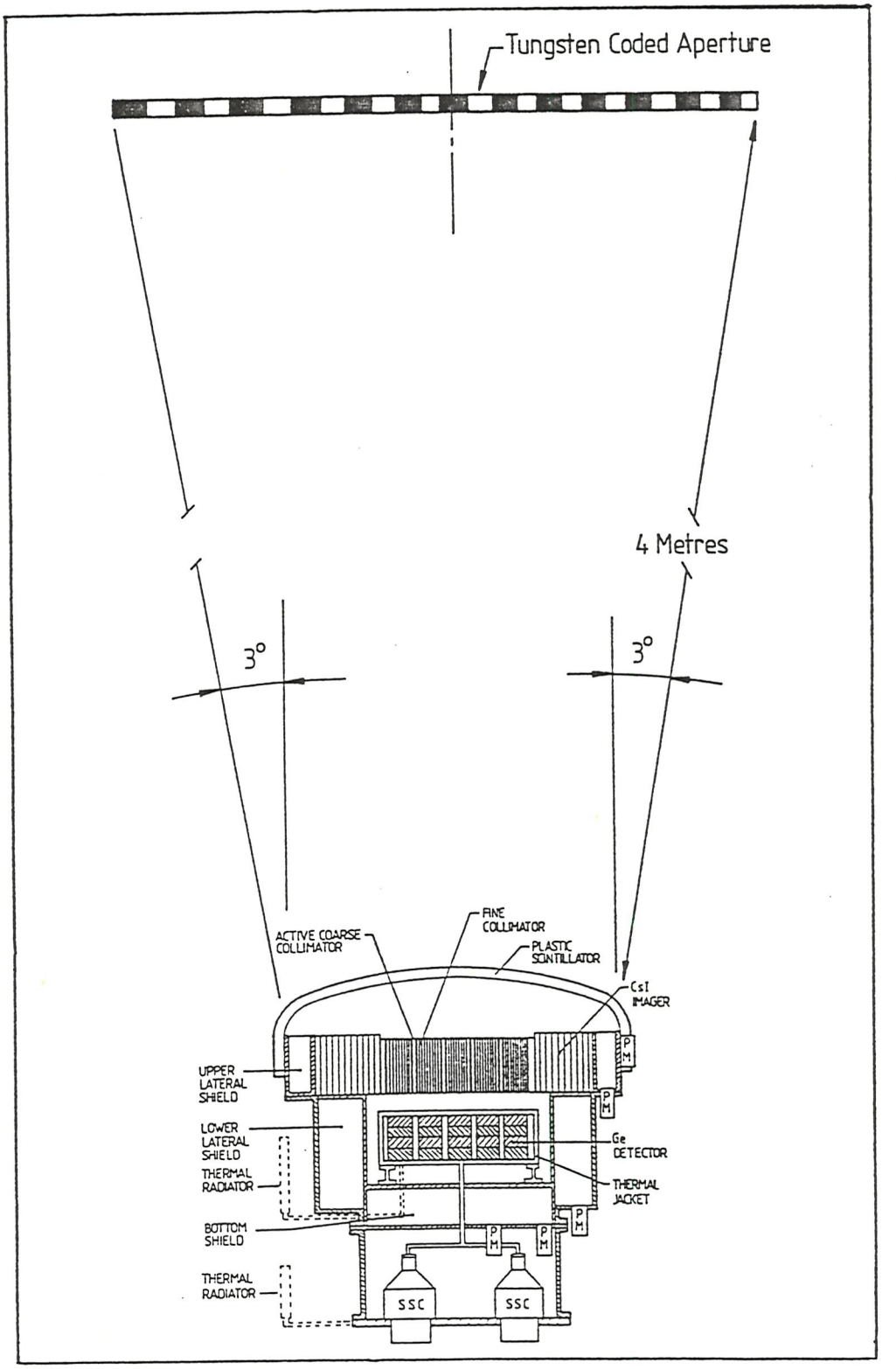


Fig. 7.1. *Schematic drawing of the GRASP telescope (from [58]).*

Performance Parameter	Ge Array	CsI Array
Energy Range	20 keV → 10 MeV	300 keV →> 100 MeV
Detection Area	400 cm^2	2400 cm^2
Spectral Resolution ((E/ΔE) at 1 MeV)	~ 500	~ 10
Field of View fully coded partially (50%) coded	 6° FWHM 14° FWHM	 6° 14°
Typical Point Source Location Capability (5σ source)	15'	5'
Temporal Resolution	100 μs	100 μs
Line Detection Sensitivity	(at 1 MeV, narrow line) (line width ~ 2 keV)	(at 1 MeV, broad line) (line width ~ 20 keV)
3σ in 10^5 s 3σ in 10^6 s	2 10^{-5} ph cm^{-2} s^{-1} 6 10^{-6} ph cm^{-2} s^{-1}	3 10^{-5} ph cm^{-2} s^{-1} 10^{-5} ph cm^{-2} s^{-1}
Continuum Sensitivity 3σ in 10^5 s 3σ in 10^6 s	(at 100 keV, ΔE = 100 keV) 5.6 10^{-7} ph cm^{-2} s^{-1} keV^{-1} 1.7 10^{-7} ph cm^{-2} s^{-1} keV^{-1}	(at 10 MeV, ΔE = 10 MeV) 10^{-9} ph cm^{-2} s^{-1} keV^{-1} 3 10^{-10} ph cm^{-2} s^{-1} keV^{-1}
Performance Parameter	**Optical Transient Camera**	**X-Ray Monitor**
Wavelength/Energy Range	450 nm → 750 nm	4 keV → 25 keV
Detection area	CCD: 385 x 288 pixels (8.5 mm x 6.5 mm)	250 cm^2
Field of View	16° x 12°	6° FWHM
Typical Point Source Location Capability	2.'5	5'
Temporal Resolution	1 s	0.5 s
Sensitivity	12^m for 1 s	< 1 mCrab (10^5 s)

Table 7.1. *The expected performance parameters of the GRASP telescope.*

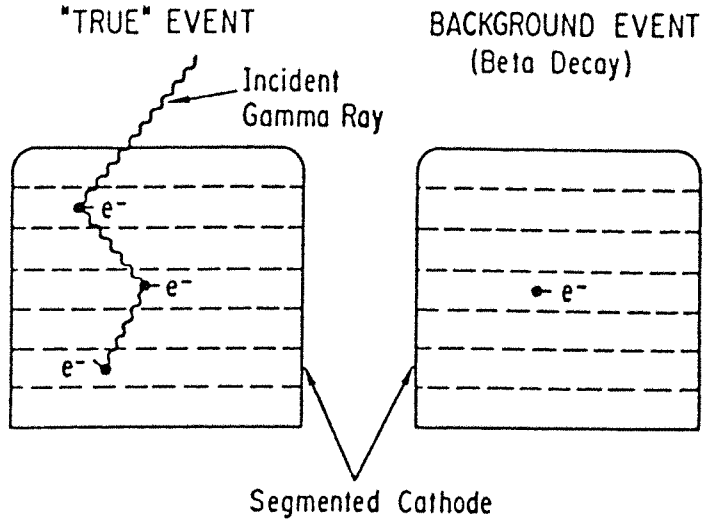


Fig. 7.2. *Principal discrimination between multiple site gamma camera events and single site background events in germanium (from [58]).*

of one or more Compton scatters followed by a photoelectric absorption. These are called *multiple site interactions*. If the single site and multiple site events can be distinguished then the major background contribution will be reduced. There are three techniques which can be used to achieve this: coaxial detector segmentation, pulse shape discrimination, and using a stack of planar germanium detectors.

A single coaxial germanium detector can be electrostatically separated into horizontal segments, and events can be resolved spatially by measuring the charge output from the detector. Single site events will only be measured in one segment and can easily be rejected. The second method, pulse shape discrimination, is possible because of the detector's finite charge drift velocity which results in a sharp charge pulse for single events and a broad pulse for multiple events.

The third technique of using a stack of planar germanium detectors has been chosen for use as the GRASP germanium spectrometer. The principle is the same as the use of a segmented coaxial detector, except that the separation between layers is physical (by, for example, some substance or vacuum) rather than electrical. Each germanium stack, or module, for GRASP consists of four hyperplane germanium detectors with an ultrathin boron implanted outer contact on the front and sides, and a lithium diffused spot on the near face. Each planar detector is 54 mm total diameter with a useful diameter of 52 mm and 15 mm thick, and the separation between the planars varies between 2 mm and 4 mm. The whole module is enclosed in a sealed vacuum container and kept cool at 85 K by a cryogenic cooler. The spectral resolution of the germanium $E/\Delta E$ is expected to be $\sim$500 at 1 MeV.

The whole GRASP germanium spectrometer consists of 19 such modules arranged in a hexagonal pattern (see fig. 7.3) such that the module spacing is exactly equal to the mask element spacing of the coded aperture, thus giving the best possibility to view celestial sources at infinity.

7.3.2 Caesium Iodide Imager

The caesium iodide position sensitive detector plane is constructed from a three dimensional array of gamma ray detector elements in the form of bars of caesium iodide which has been doped with thallium. Each bar is capable of both locating and measuring the energy deposited by particle interactions throughout the entire detector volume. The three dimen-

sional position sensitivity is very useful. It enables finer imaging pixels to be constructed, the separation of multiple site events from single site beta decay background events, and event tracking from Compton kinematics, in turn enabling the initial interaction position to be accurately determined. The caesium iodide imager is 15 cm deep ensuring almost 100% detection efficiency in the range 300 keV - 100 MeV.

Each bar is 15 cm long, of hexagonal cross section 1.6 cm flat to flat. These dimensions have been chosen so that when the bars are stacked vertically into a hexagonal array, there is a match with both the germanium spectrometer array and the coded aperture, with 19 hexagonal bars to each hexagonal mask element ($A_r = 19$. See fig. 7.7).

The position of a photon interaction inside a bar is calculated from the relative pulse heights measured by two photodiodes, one placed at each end of the bar. The event position resolution, as measured experimentally for a bar of square cross section, has been estimated as typically 10 mm FWHM, see fig. 7.4 [58,61].

The caesium iodide detector plane construction is modular in form with six identical modules each containing approximately 200 position sensitive bars, each bar being optically isolated from its neighbour [20]. The modules are arranged around the outside of the germanium array as shown in fig. 7.3 and thus provide a 15 cm thick active shield for the germanium spectrometer. The total sensitive area of the caesium iodide imager plane is 2,400 cm^2.

7.3.3 Veto System

The GRASP veto system is an important element of the instrument, since it has to reduce the detector background by orders of magnitude. The GRASP detector configuration is shown schematically in fig. 7.5, and the six main parts of the veto system are indicated. Of these six parts only the front plastic shield rejects mainly charged particles. All the other parts of the veto system reject photons with high efficiency. The basic elements are as follows:

1. Front plastic shield. This is a large dome with a thickness of about 0.5-1.0 cm, viewed by 8 photomultiplier tubes (PMT).
2. Upper lateral shield. Made of CsI(Tl), thickness 3 cm, 8 segments with 2 PMTs per segment.

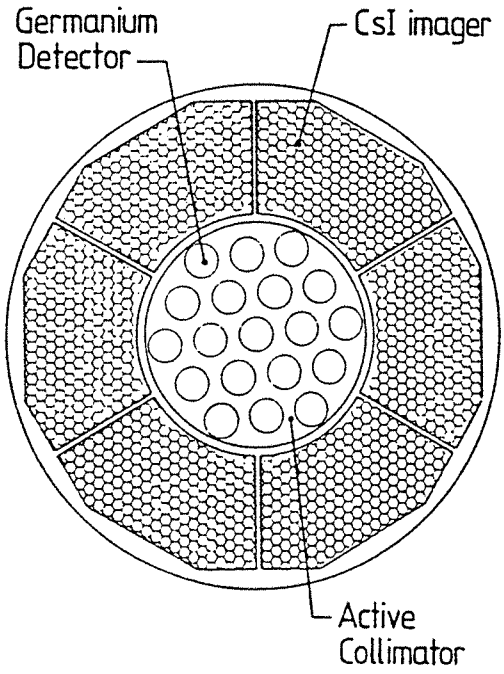


Fig. 7.3. *Schematic plan view of the GRASP detector plane (from [58]).*

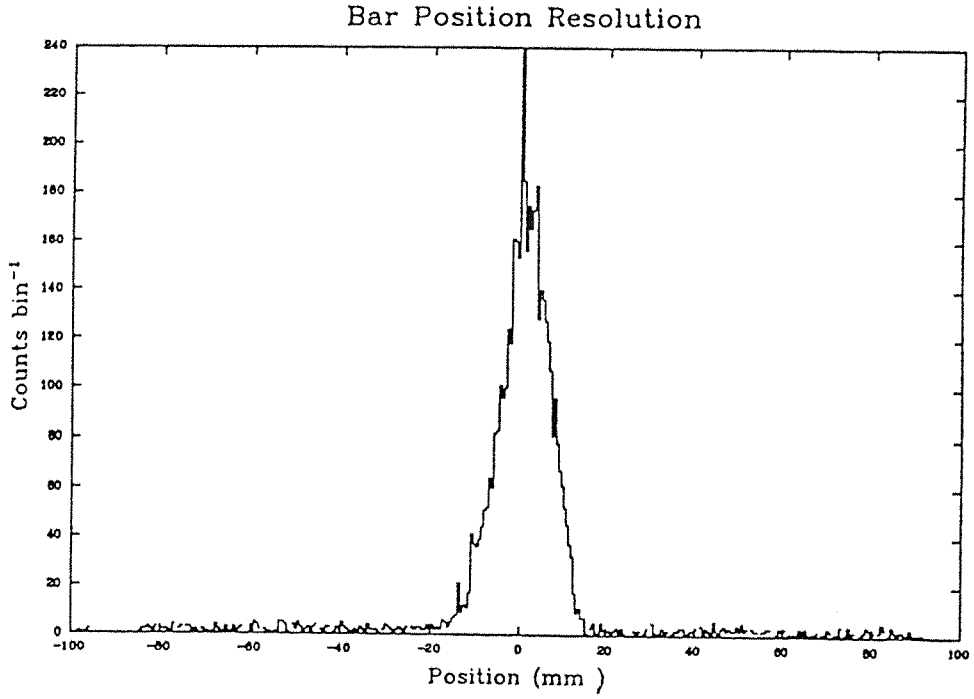


Fig. 7.4. *Positional resolution of a caesium iodide imager bar detector (from [58]).*

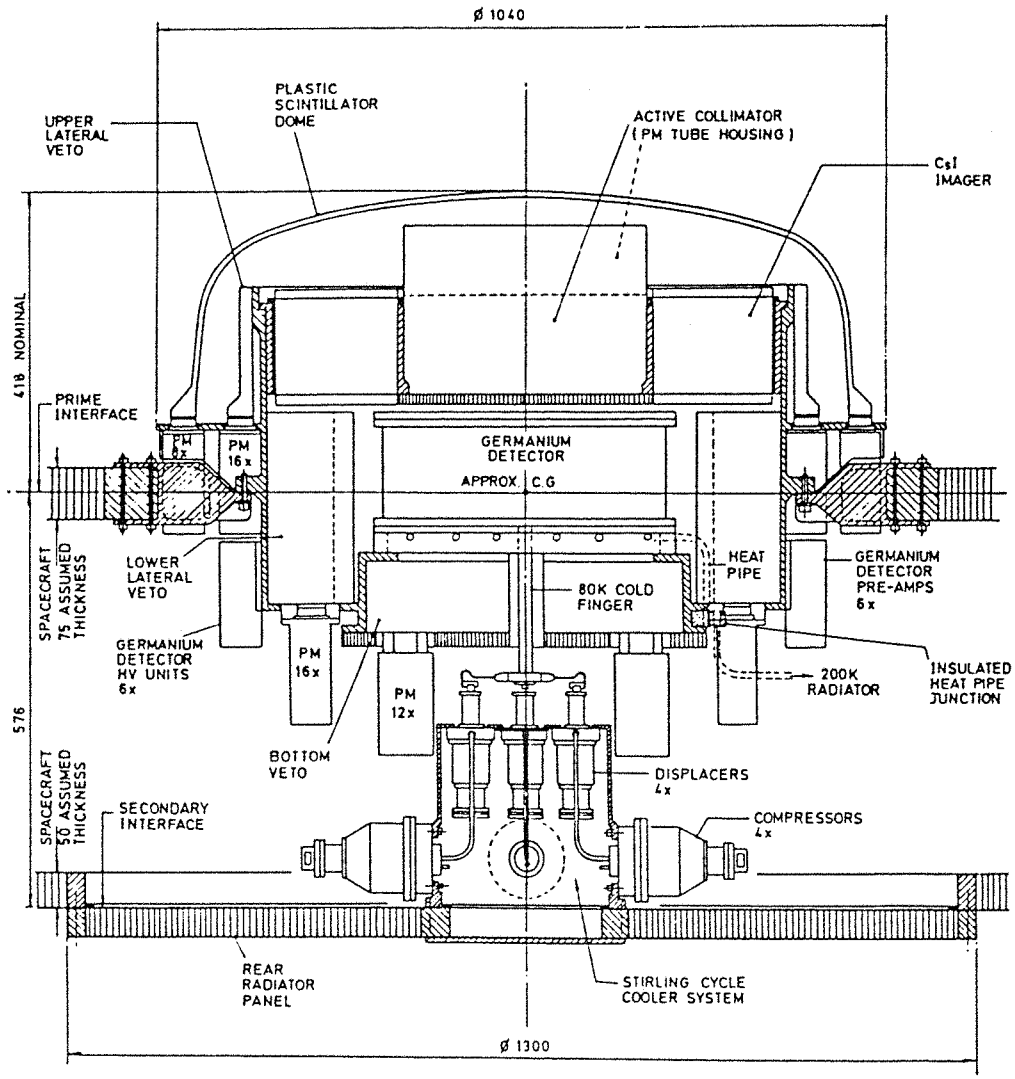


Fig. 7.5. *The GRASP detector (from [58]).*

3. Lower lateral shield. Made of CsI(Tl), thickness 12 cm, 8 segments with 2 PMTs per segment.

4. Bottom shield. Made of CsI(Tl), thickness 12 cm, 6 segments + 1 or 2 central parts, 2 PMTs per segment.

5. Active collimator. Made of drilled CsI(Tl) crystal or separate elements of CsI(Tl) or bismuth germenate (BGO), field of view about 25° FWHM, minimum wall thickness about 1 cm, height about 12 cm, viewed by 25 PMTs.

6. Passive collimator. Consists of 19 separate units mounted in the holes of the active collimator, hexagonal cells, from tungsten or tantalum with 0.01 cm thickness, cell size 1.2 cm, field of view about 6° FWHM, height about 12 cm.

The segmentation of the shields is necessary to reduce the dead time of the veto system after the passage of a charged particle with high energy loss.

7.3.4 Coded Aperture

Coded apertures have already been used to perform useful gamma ray astronomy (see section 2.2.3) and will be used in the forthcoming SIGMA mission [94]. Using a coded aperture enables instruments to be designed which have a large sensitive area, a high angular resolution whilst being able to maintain a wide field of view, and the ability to measure both the source and the detector background simultaneously. A wide variety of mask pattern types and sizes are available. In the case of the GRASP experiment, the mask pattern has been chosen to match the complex detector configuration shown in fig. 7.3, which uses detector elements of different sizes. The mask pattern chosen is a hexagonal uniformly redundant array (HURA), a type discovered by Finger and Prince [50] and discussed in section 3.3.5. The mask has 79 elements in its basic pattern and 211 elements in total, as shown in fig. 7.6. The hexagonal element size corresponds to the dimensions of the germanium detector array whilst the cell size for the caesium iodide imager is such that 19 caesium iodide elements fit into one mask element as shown in fig. 7.7. This configuration preserves the properties of the mask pattern, and was shown in chapter 5 that it results in an optimum coded system in the sense that it is free from coding errors.

The point source location error for a distance of 4 metres between the mask and the detector is typically 5 arcminutes for weak (5σ) sources, whilst the mask diameter of 110 cm gives a

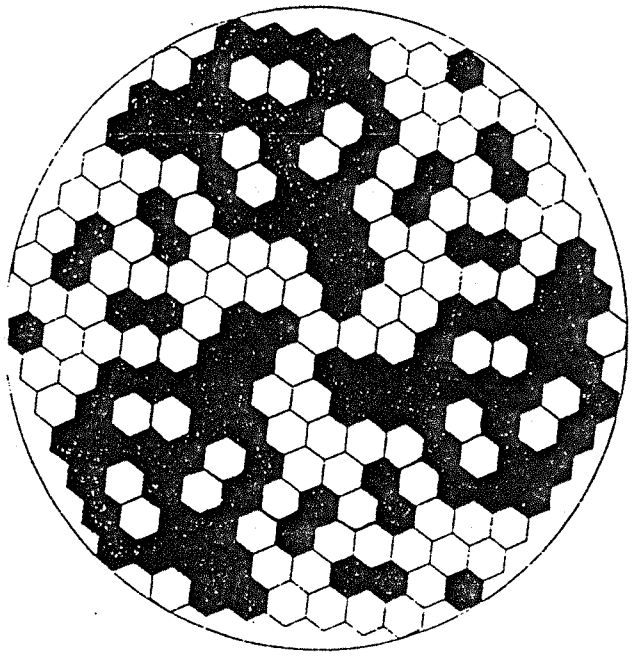


Fig. 7.6. *The GRASP coded aperture ($v = 79$, 211 elements in total, from [58]).*

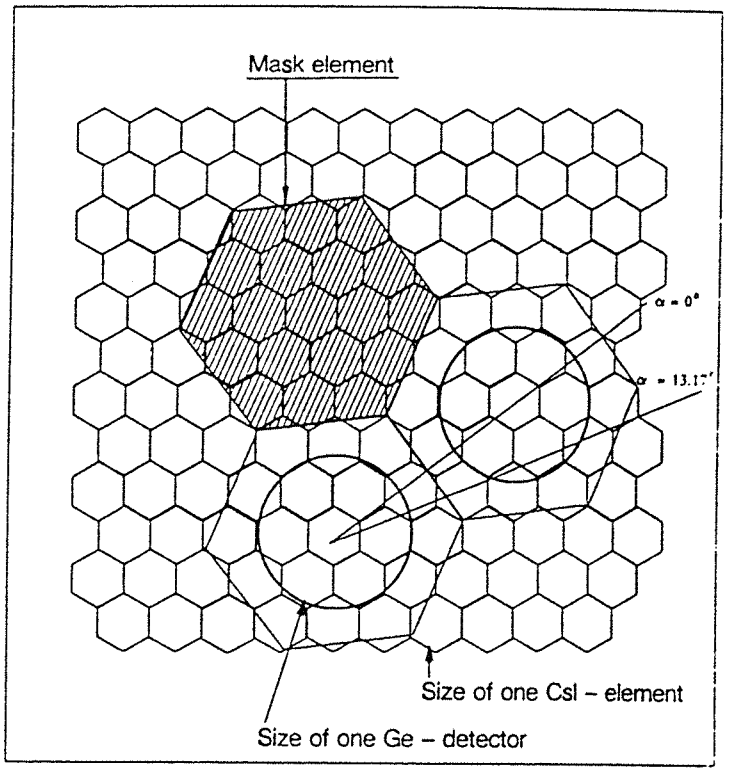


Fig. 7.7. *The GRASP mask element and detector plane configuration showing both the caesium iodide bars and the circular germanium spectrometers (from [58]).*

6° fully coded field of view. However, the point source sensitivity for the germanium array will vary significantly according to whether the source is well centred in the mask shadow elements or located near the hexagon borderlines. In this respect the caesium iodide imager will perform better due to the small pixel size, and therefore detection and localisation of serendipitous sources are most likely to result from data from the caesium iodide imager.

The opaque mask elements are made of 1.5 cm thick tungsten alloy which achieves 70% mask opacity even in the difficult band between 2 and 3 MeV. These tungsten elements are supported on both sides by carbon fibre reinforced plastic honeycomb panels on the top and bottom face.

Appendix A

Block Designs.

In this appendix is given an example of a cyclic symmetric balanced incomplete block design and the associated difference set, as explained in section 4.2.

The block design has parameters $b = v = 7$, $r = k = 3$, $\lambda = 1$. The objects $a_i = i$ are the integers from 1 to 7 ($\equiv 0$ since the design is cyclic of period 7) and the blocks B_i are defined as follows:

$$B_1 : \{1, 2, 4\}$$
$$B_2 : \{2, 3, 5\}$$
$$B_3 : \{3, 4, 6\}$$
$$B_4 : \{4, 5, 0\}$$
$$B_5 : \{5, 6, 1\}$$
$$B_6 : \{6, 0, 2\}$$
$$B_7 = B_0 : \{0, 1, 3\}$$

The incidence matrix $A = \{a_{ij}\}$ obeying the condition 4.3 is given by

$$A = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 = a_0 \end{array} \begin{array}{c} \begin{array}{ccccccc} B_1 & B_2 & B_3 & B_4 & B_5 & B_6 & B_7 = B_0 \end{array} \\ \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right] \end{array}$$

Note that equations 4.2 and 4.9 hold for this design, and that

$$B = AA^{\mathrm{T}} = 2I_7 + J_7$$

which, as can be seen, agrees both with 4.5 and with 4.10.

The associated difference set is one of the blocks of the design (section 4.2.2, theorem 4.2), say block B_1. The incidence vector of this difference set B_1 is

$$B_1 : \{1, 1, 0, 1, 0, 0, 0\}$$

which has a two valued autocorrelation function $R_a(j)$ given by:

$$R_a(j) = \begin{cases} 3 & \text{if } j \equiv 0 (\mathrm{mod} 7) \\ 1 & \text{if } j \not\equiv 0 (\mathrm{mod} 7) \end{cases}$$

obeying 4.17.

Appendix B

Glossary of Mathematical Symbols.

$\in$	belongs to (the set)
$\notin$	does not belong to (the set)
I_a	unit matrix of order a
J_a	square matrix of 1's of order a
$\equiv$	is congruent to
$\not\equiv$	is not congruent to
(a, b)	greatest common divisor of a and b
M_a	set of multiples of a
$\phi(n)$	Euler's ϕ-function for the integer n (section 4.3.2)
$a \mid b$	a divides b
$a \nmid b$	a does not divide b
(a/p)	Legendre symbol of a with respect to p (section 4.3.2)

Bibliography

[1] G.O. Abell and B. Margon. Nature, **279**, 701-703 (1979).

[2] J. Agnew. "Explorations in Number Theory.", Brooks/Cole Publishing Co. (1972).

[3] B. Agrinier et al. Soviet Astronomy, **30**, 508-514 (1986).

[4] W.E. Althouse et al. Proc. 19th Int. Cosmic Ray Conf., **3**, 299-302 (1985).

[5] W.E. Althouse et al. Proc. 20th Int. Cosmic Ray Conf., **1**, 84-87 (1987).

[6] N.P. Archer et al. Nucl. Phys., **83**, 241-258 (1966).

[7] C.A. Ayre et al. Mon. Not. R.A.S., **205**, 285-295 (1983).

[8] W.A. Baity et al. Ap. J., **244**, 429-435 (1981).

[9] W.A. Baity et al. Ap. J., **279**, 555-562 (1984).

[10] R.E. Baker et al. Nucl. Instr. Meth., **158**, 595-604 (1979).

[11] R.E. Baker et al. Phil. Trans. Royal Soc. London, **301A**, 603-606 (1981).

[12] L. Bassani and A.J. Dean. Spa. Sci. Rev., **35**, 367-398 (1983).

[13] L.D. Baumert. Lecture Notes in Mathematics, **182**, "Cyclic Difference Sets.", Springer-Verlag (1971).

[14] M.C. Begelman et al. Ap. J., **238**, 722-730 (1980).

[15] D.L. Bertsch. Adv. Spa. Research, **3**, 515-523 (1984).

[16] M. Bezler et al. Astron. Astrop., **136**, 351-354 (1984).

[17] G.F. Bignami et al. Ap. J., **232**, 649-658 (1979).

[18] G.F. Bignami et al. Ap. J. Lett., **272**, L.9-13 (1983).

[19] G.F. Bignami et al. Nature, **310**, 464-469 (1984).

[20] G.F. Bignami et al. IEEE Trans. Nucl. Sci., **NS-34**, 31-35 (1987).

[21] G.F. Bignami et al. Astron. Astrop. Lett., **202**, L.1-4 (1988).

[22] R.L. Blake et al. Rev. Sci. Instr., **45**, 513-516 (1974).

[23] G. Boella et al. IEEE Trans. Nucl. Sci., **NS-33**, 755-758 (1986).

[24] A. Brauer. Math. Zeitschrift, **58**, 219-225 (1953).

[25] C. M. Brown. J. Appl. Phys., **45**, 1806-1811 (1974).

[26] R.H. Bruck and H.J. Ryser. Can. J. Math., **1**, 88-93 (1949).

[27] D. Calabro and J.K. Wolf. Inform. Control, **11**, 537-560 (1968).

[28] T.M. Cannon and E.E. Fenimore. Opt. Eng., **19**, 283-289 (1980).

[29] P.A. Caraveo et al. Ap. J., **327**, 203-213 (1988).

[30] P. Charalambous et al. Nucl. Instr. Meth., **221**, 183-186 (1984).

[31] S. Chowla. Math. Student, **12**, 82-85 (1945).

[32] S. Chowla and H.J. Ryser. Can. J. Math., **2**, 93-99 (1950).

[33] D.A. Clayton et al. Ap. J., **155**, 75-82 (1969).

[34] T.L. Cline and U.D. Desai. Ap. J. Lett., **196**, L.43-46 (1975).

[35] W.R. Cook et al. IEEE Trans. Nucl. Sci., **NS-31**, 771-775 (1984).

[36] W.R. Cook et al. IEEE Trans. Nucl. Sci., **NS-32**, 129-133 (1985).

[37] W.R. Cook et al. Ap. J. Lett., **334**, L.87-90 (1988).

[38] B.A. Cooke et al. Mon. Not. R.A.S., **182**, 489-515 (1978).

[39] A.J. Dean and D. Ramsden. Phil. Trans. Royal Soc. London, **301A**, 577-602 (1981).

[40] R.H. Dicke. Ap. J. Lett., **153**, L. 101-106 (1968).

[41] A.J.B. Downes et al. Astron. Astrop., **103**, 277-287 (1981).

[42] P.P. Dunphy et al. Nucl. Instr. Meth., **274**, 362-379 (1989).

[43] C.J. Eyles et al. J. British Interplan. Soc., **40**, 159-162 (1987).

[44] E.E. Fenimore. Appl. Opt., **17**, 3562-3570 (1978).

[45] E.E. Fenimore. Appl. Opt., **19**, 2465-2471 (1980).

[46] E.E. Fenimore and T.M. Cannon. Appl. Opt., **17**, 337-347 (1978).

[47] E.E. Fenimore and G.S. Weston. Appl. Opt., **20**, 3058-3067 (1981).

[48] C.E. Fichtel et al. Ap. J., **198** 163-182 (1975).

[49] C.E. Fichtel and J.I. Trombka. "Gamma Ray Astrophysics.", NASA - SP - 453 (1981).

[50] M.H. Finger and T.A. Prince. Proc. 19th Int. Cosmic Ray Conf., **3**, 295-298 (1985).

[51] G.J. Fishman et al. Proc. 19th Int. Cosmic Ray Conf., **3**, 343-346 (1985).

[52] A.R. Foley et al. Mon. Not. R.A.S., **216**, 679-690 (1985).

[53] A.S. Fruchter et al. Nature, **333**, 237-239 (1988).

[54] N. Gehrels et al. Ap. J., **278**, 112-118 (1984).

[55] R. Giacconi et al. Ap. J., **230**, 540-550 (1979).

[56] "Digital Communications with Space Applications.", Prentice-Hall, Engelwood Cliffs, New Jersey (1964).

[57] S.W. Golomb. "Shift Register Sequences.", Aegean Park Press (1982).

[58] "GRASP. Gamma Ray Astronomy with Spectroscopy and Positioning.", Report on the Phase A Study. ESA Code:SCI(88)2 (1988).

[59] W. Gruner. Comment. Math. Helv., **12**, 149-152 (1939).

[60] J. Gunson and B. Polychronopulos. Mon. Not. R.A.S., **177**, 485-497 (1976).

[61] C.J.Hall et al. IEEE Trans. Nucl. Sci., **NS-36**, 863-865 (1989).

[62] Marshall Hall Jr. "Combinatorial Theory.", Blaisdell (1967).

[63] R.D. Hall et al. Ap. J., **210**, 631-641 (1976).

[64] J.H. Halpern and D. Tytler. Ap. J., **330**, 201-217 (1988).

[65] T.M. Harrington et al. Nucl. Instr. Meth., **118**, 401-411 (1974).

[66] M. Harwit and N.J.A. Sloane. "Hadamard Transform Optics.", Academic Press (1979).

[67] R.C. Haymes et al. Ap. J., **201**, 593-602 (1975).

[68] R.L. Heath et al. Nucl. Instr. Meth., **162**, 431-476 (1979).

[69] D. Herzo et al. Nucl. Instr. Meth., **123**, 583-597 (1975).

[70] E.B. Hughes et al. IEEE Trans. Nucl. Sci., **NS-27**, 364-369 (1980).

[71] T.V. Jelley. Nature, **211**, 472-475 (1966).

[72] E.C. Johnsen. J. Algebra, **4**, 388-402 (1966).

[73] W.N. Johnson and R.C. Haymes. Ap. J., **184**, 103-125 (1973).

[74] G. Kanbach et al. Astron. Astrop., **90**, 163-169 (1980).

[75] A.M. Kirch. "Elementary Number Theory.", Intext Educational Publishers, New York and London (1974).

[76] R.W. Klebesadel et al. Ap. J. Lett., **182**, L.85-88 (1973).

[77] W.K. Klemperer. Astron. Astrop. Suppl., **15**, 449-451 (1974).

[78] J.D. Kurfess et al. Adv. Spa. Res., **3**, 109-112 (1983).

[79] R.C. Lamb et al. Nature, **305**, 37-39 (1983).

[80] A. Lawrence. Mon. Not. R.A.S., **192**, 83-94 (1980).

[81] S.M. Lea et al. Ap. J., **246**, 369-375 (1981).

[82] M. Leventhal et al. Ap. J., **216**, 491-502 (1977).

[83] M. Leventhal et al. Ap. J. Lett., **225**, L.11-14 (1978).

[84] A. Levine et al. Bull. A.A.S., **11**, 429 (1979).

[85] E.P. Liang. Ap. J., **304**, 682-687 (1986).

[86] J.C. Ling et al. Nature, **270**, 36-37 (1977).

[87] J.C. Ling et al. Ap. J. Lett., **321**, L.117-122 (1987).

[88] J.A. Lockwood et al. Ap. J., **248**, 1194-1201 (1981).

[89] M.S. Longair. "High Energy Astrophysics.", Cambridge University Press (1981).

[90] C.J. MacCallum. Ap. J., **291**, 486-491 (1985).

[91] F.J. MacWilliams and N.J.A. Sloane. Proc. IEEE, **64**, 1715-1731 (1976).

[92] F.J. MacWilliams and N.J.A. Sloane. "The Theory of Error Correcting Codes.", Amsterdam, North Holland Publishing (1977).

[93] W.A. Mahoney et al. Ap. J., **262**, 742-748 (1982).

[94] P. Mandrou. Adv. Spa. Res., **3**, 525-531 (1984).

[95] N. Marshall et al. Mon. Not. R.A.S., **194**, 987-1002 (1981).

[96] S.M. Matz et al. Nature, **331**, 416-418 (1988).

[97] H.A. Mayer-Hasselwander et al. Astron. Astrop., **105**, 164-175 (1982).

[98] E.P. Mazets et al. Astrop. Spa. Sci., **82**, 261-282 (1982).

[99] M.L. McConnell et al. Proc. 20th Int. Cosmic Ray Conf., **1**, 66-69 (1987).

[100] L. Mertz. "Transformations in Optics.", J. Wiley and Sons Inc. (1965).

[101] L. Mertz. Proc. Symp. Mod. Opt., Polytechnic Institute, Brooklyn (1968).

[102] L. Mertz and N.Young. Proc. Int. Conf. Opt. Instr. Tech., page 305 (1961).

[103] P.L. Nolan et al. Nature, **293**, 275-277 (1981).

[104] M. Oda. Appl. Opt., **4**, 143 (1965).

[105] H. Ogelman. Proc. NATO ASI: "Astrophysical Jets and their Engines.", ed. W. Kunt, 67. (1987).

[106] A. Owens et al. Proc. 20th Int. Cosmic Ray Conf., **1**, 172-175 (1987).

[107] M.E. Ozel and C.E. Fichtel. Ap. J., **335**, 135-141 (1988).

[108] R.E.A.C. Paley. J. Math. Phys., **12**, 311-320 (1933).

[109] F. Perotti et al. Nature, **292**, 133-135 (1981).

[110] L.E. Peterson et al. J. Geophys. Res., **78**, 7942-7958 (1973).

[111] P.A. Pinto and S.E. Woosley. Nature, **333**, 534-537 (1988).

[112] A.M.T. Pollock et al. Astron. Astrop., **94**, 116-120 (1981).

[113] N. Prantzos and M. Cassé. Ap. J., **307**, 324-331 (1986).

[114] F.A. Primini et al. Nature, **278**, 234-235 (1979).

[115] R.J. Proctor et al. Mon. Not. R.A.S., **187**, 633-643 (1979).

[116] R. Ramaty et al. Ap. J. Suppl., **40**, 487-526 (1979).

[117] R. Ramaty and R.E. Lingenfelter. Ap. J. Lett., **213**, L.5-7 (1977).

[118] R. Ramaty and R.E. Lingenfelter. Nature, **278** 127-132 (1979).

[119] R. Ramaty and R.E. Lingenfelter. Phil. Trans. Royal Soc. London, **301A**, 671-686 (1981).

[120] G.R. Riegler et al. Ap. J. Lett., **248**, L.13-16 (1981).

[121] G.R. Riegler et al. Ap. J. Lett., **294**, L.13-15 (1985).

[122] W.L. Rogers et al. J. Nucl. Medicine, **21**, 371 (1980).

[123] M. Ruderman et al. Ap. J., **336**, 507-518 (1989).

[124] M. Ruderman et al. Ap. J. (In press) (1989).

[125] G. Rudstam et al. Zeitschrift für Naturf., **21a**, 1027 (1966).

[126] H.J. Ryser. Proc. Amer. Math. Soc., **1**, 422-424 (1950).

[127] H.J. Ryser. "Combinatorial Mathematics.", Carus Mathematical Monographs, No. 14. Wiley, New York (1963).

[128] R. Schlickeiser. Astron. Astrop. Lett., **106**, L.5-8 (1982).

[129] H.W. Schnopper et al. Space Sci. Rev., **8**, 534 (1968).

[130] V. Schönfelder et al. Nucl. Instr. Meth., **107**, 385-394 (1973).

[131] V. Schönfelder et al. Ap. J., **217**, 306-319 (1977).

[132] V. Schönfelder et al. Ap. J., **240**, 350-362 (1980).

[133] V. Schönfelder et al. Astron. Astrop., **110**, 138-151 (1982).

[134] V. Schönfelder et al. IEEE Trans. Nucl. Sci., **NS-31**, 766-770 (1984).

[135] G.H. Share et al. Ap. J. Lett., **292**, L.61-65 (1985).

[136] M.D. Short. Nucl. Instr. Meth., **221**, 142-149 (1984).

[137] W. Sieber and R. Schlickeiser. Astron. Astrop., **113**, 314-323 (1982).

[138] G.K. Skinner et al. Nature, **330**, 544-547 (1987).

[139] G.K. Skinner et al. Phys. Bull., **38**, 10-11 (1987).

[140] W. Stahnke. Math. Comput., **27**, 977-980 (1973).

[141] R.G. Stanton and D.A. Sprott. Can. J. Math., **10**, 73-77 (1958).

[142] F.W. Stecker et al. Phys. Rev. Lett., **27**, 1469-1472 (1971).

[143] A.W. Strong and G.F. Bignami. Ap. J., **274**, 549-557 (1983).

[144] B.N. Swanenburg et al. Nature, **275**, 298 (1978).

[145] B.N. Swanenburg et al. Ap. J. Lett., **243**, L.69-73 (1981).

[146] D.J. Thompson et al. Ap. J. Lett., **200**, L.79-82 (1975).

[147] J.I. Trombka et al. Ap. J., **181**, 737-746 (1973).

[148] J.I. Trombka et al. Ap. J., **212**, 925-935 (1977).

[149] J. Trümper et al. Ap. J. Lett., **219**, L.105-110 (1978).

[150] J. Tueller et al. Ap. J., **279**, 177-183 (1984).

[151] G. Villa et al. IEEE Trans. Nucl. Sci., **NS-34**, 26-30 (1987).

[152] P. von Ballmoos et al. Ap. J., **318**, 654-663 (1987).

[153] P. von Ballmoos et al. Ap. J., **312**, 134-142 (1987).

[154] J. Ward et al. Astron. Astrop. Lett., **59**, L.19-21 (1977).

[155] M.G. Watson et al. Ap. J., **273**, 688-696 (1983).

[156] A.L. Whiteman. Illinois J. Math., **6**, 107-121 (1962).

[157] J. Williamson. Duke Math. J., **11**, 65-81 (1944).

[158] J. Williamson. Bull. Amer. Math. Soc., **53**, 608-613 (1947).

[159] R. Willingale et al. Nucl. Instr. Meth., **221**, 60-66 (1984).

[160] A.P. Willmore et al. Nucl. Instr. Meth., **221**, 284-287 (1984).

[161] S.E. Woosley et al. Comments Nucl. Part. Phys., **9**, 185-197 (1981).

[162] S.E. Woosley et al. Ap. J., **301**, 601-623 (1986).

[163] D.M. Worrall et al. Ap. J., **232**, 683-688 (1979).

[164] M. Yearworth. Ph.D Thesis, University of Southampton (1987).

[165] M. Yoshimori et al. Australian J. Phys., **32**, 375-382 (1979).

[166] N. Zierler and J. Brillhart. Inform. Control, **13**, 541-544 (1968).

[167] N. Zierler and J. Brillhart. Inform. Control, **14**, 566-569 (1969).

[168] A.D. Zych et al. IEEE Trans. Nucl. Sci., **NS-26**, 506-512 (1979).

Acknowledgements

It is a pleasure to thank the many people who have contributed to this work in one way or another.

I am greatly indebted to my supervisor Dr. A. J. Dean for his guidance and enthusiasm throughout this work. Thanks also to all other members of the Astrophysics Group at Southampton University, past and present, who have made my time there an enjoyable one. In particular I would like to thank Dr. A. Goldwurm, my research partner for the main part of these three years, who produced some of the results presented here. Special thanks also to Dr. M. Yearworth for his great help and innumerable useful discussions, to S. Duncan for saving much of my time by performing deconvolutions on the transputer array, and to Dr. P. Maggioli for his guidance and tuition when I first arrived, making my introduction to research a lot easier.

Thanks are also due to Dr. J. Fleming and the staff of the Department of Nuclear Medicine at Southampton General Hospital for enabling me to use their gamma ray imaging facilities, and to Dr. Keith Lloyd of the Department of Mathematics at Southampton University for his guidance regarding the mathematical section of this thesis.

Finally, and above all, I would like to thank my Mum and Dad, without who's constant help and encouragement this work would certainly never have been possible. It is to them whom this thesis is dedicated.