Astron. Astrophys. 288, 665–674 (1994)

ASTRONOMY AND ASTROPHYSICS

The optimum open fraction of coded apertures. With an application to the wide field X-ray cameras of SAX

J.J.M. in 't Zand[1,2], J. Heise[1], and R. Jager[1]

[1] SRON-Laboratory for Space Research, Sorbonnelaan 2, NL-3584 CA Utrecht, The Netherlands
[2] Los Alamos National Laboratory, Astrophysics and Radiation Measurements, MS D436, Los Alamos, NM 87545, USA

Received 8 December 1993 / Accepted 31 January 1994

Abstract. We consider issues that concern the mathematical description of coded aperture patterns. Primarily this involves the relation between the open fraction of such patterns and the signal-to-noise ratio of imaged point sources. A refinement of the corresponding theory is introduced, taking into account the spatial response of the coded aperture camera. From this we predict that patterns with an open fraction of less than 0.5 can enhance the performance of coded aperture cameras to *bright* point sources, as opposed to what was previously thought.

As an application of the refined theory, we tested candidate open fractions in the instrumental configuration of two identical, wide field, coded aperture X-ray cameras (1.8–30 keV), that will be part of the X-ray satellite SAX (to be launched in late 1995). These tests consisted of numerical simulations of several types of observations, and show that open fractions between 0.25 and 0.33 are to be favored for the SAX cameras. The improvement in signal-to-noise ratio with respect to the commonly used open fraction of 0.5 is up to 30%. Whenever telemetry limits the data coverage, this profit may well be larger.

We also address additional aperture constraints as applied to the SAX cameras, such as the aperture geometry and pattern. From this analysis we propose a new type of coded aperture pattern for the SAX cameras with an open fraction equal to 0.33, which possesses near-ideal intrinsic mathematical properties.

Key words: instrumentation: detectors – techniques: image processing – telescopes – X-rays: general

1. Introduction

Coded aperture cameras (for reviews, see e.g. Skinner 1984; Caroli et al. 1987) are based on the principle of coding the spatial information contained in the object with a coded aperture. The coded image, as measured by a position-sensitive detector, can be decoded back to the object domain through a computational cross correlation with a mathematical representation of the coded aperture (usually this is done *after* the actual measurement). This two-stepped imaging technique is particularly suitable at photon energies where conventional imaging with a lens or mirror is very difficult or impossible. Therefore, the importance of this imaging technique to high-energy astronomy (X- and γ-ray) is evident.

The coded aperture pattern obviously plays a crucial role in the imaging performance of the camera. Independently, Gunson & Polychronopulos (1976) and Fenimore & Cannon (1978) proposed aperture patterns which have better characteristics than the random patterns proposed by e.g. Dicke (1968). These patterns are commonly known as Uniformly Redundant Arrays (URAs). They are based on cyclic difference sets (Baumert 1971) and are characterized by a two-valued cyclic autocorrelation function (i.e. a repeated delta-function). This means that the employment of such patterns in principle results in a complete suppression of 'false' peaks.

An important aspect of the aperture patterns is the fraction t of the number of all pattern elements that are partially or fully transparent. We will call t the 'open fraction' of the aperture (the actual fraction of open *area* may be somewhat smaller than t because of technical considerations). Gunson & Polychronopulos (1976) showed that the value of t which is optimum with respect to the signal-to-noise ratio of imaged point sources, $t_{\rm opt}$, depends on the ratio of the point source to background intensity, and concluded that $t_{\rm opt} = 0.5$ if the background level dominates over the source intensity. However, Fenimore (1978) demonstrated that $t_{\rm opt}$ decreases for faint sources if the total intensity of all sources increases, though the associated gain in signal-to-noise ratio is limited to a factor $\sqrt{2}$ at most.

In this paper we refine the theory of coded aperture camera performance as a function of open fraction. We will show analytically that coded apertures with an open fraction of less than 0.5 can enhance the performance of coded aperture cameras even for *bright* point sources, contrary to what was previously thought.

Send offprint requests to: J.J.M. in 't Zand, Los Alamos

Relatively little is known about URA patterns with open fractions less than 0.5. Recently, Skinner & Grindlay (1993) investigated URA-like patterns with $t = 0.25$ for application in instruments with a very broad passband. These patterns, however, do not sufficiently meet our needs. We discuss this and probe the range of low open fractions for patterns which have mathematical properties that are nearer to the ideal URA ones.

The study described here was performed as part of a program to find optimum aperture patterns for two wide field cameras (WFCs) that will form a Dutch contribution to the Italian X-ray astronomy satellite *'Satellite per Astronomia a raggi X'* (SAX, Scarsi 1993), to be launched in late 1995. Consequently, we tested our analytical analysis with numerical simulations of observations with these WFCs and confined the investigation of low open fractions for URA-like patterns within the limits imposed by the SAX-WFCs.

Customized coded apertures need to meet a number of other conditions. These pertain to the geometry and to the performance of the instrument when used as a collimator only (i.e. disregarding the coding by the aperture; in certain situations this may be worthwhile). We will address such conditions in the context of the SAX-WFCs.

2. Theoretical analysis of open fraction

Here we extend the theory behind the optimum open fraction $t_{\rm opt}$, introduced by Gunson & Polychronopulos (1976) and Fenimore (1978), taking additionally into account the spatial response of the complete coded aperture imaging system as described by the point spread function (PSF). We derive a revised analytical expression for $t_{\rm opt}$. In order to do so, we follow the process of the reconstruction of a point source and evaluate the expected value for its signal-to-noise ratio S/N, given the expected sources of radiation. For this analysis, we disregard any dependencies on the responses of collimator or support structures, since we are at this point only interested quantitatively in relative results. Furthermore, we assume that every subsection of the pattern has an open fraction equal to t. In Sect. 2.1, we describe the theory for the ideal case. The well-known results are extended in Sect. 2.2 to include the effect of the spatial response of the instrument.

2.1. Ideal PSF

The position-sensitive detector is assumed to be exposed to basically three components of radiation:

1. N point sources of intensity S_i cts per total detector area ($i = 1...N$) for the duration of the observation.
2. An isotropic diffuse sky background of B cts per detector area per field of view (FOV).
3. An internal detector background of b cts per detector area (this includes all counts that have *not* reached the detector via the aperture).

Both background components are assumed to have a flat detector response. The detector data is sampled in bins equal in size to the aperture elements.

Assume that one is interested in the reconstruction of the intensity of one point source, say $i = 1$. The reconstruction involves a cross correlation of the detector data with the digitized aperture pattern, and may be split in two: one with the 'open' pattern (allocating a 0 to each closed aperture element and a 1 to each open one), and one with the 'closed' pattern (reversing the allocation). Thus, by performing the 'open' cross correlation all counts behind open elements are accumulated, while in the 'closed' cross correlation all counts behind closed elements are accumulated, as seen from the sky position to be reconstructed.

Given the sources of radiation, the 'open' cross correlation at the position k of the considered point source is expected to have the value

$$C_{k,\rm open} = S_1\, t + t\left\{B\, t + b + \sum_{i\neq 1} S_i t\right\} \text{ cts} \quad (1)$$

and the 'closed' correlation

$$C_{k,\rm closed} = (1-t)\left\{B\, t + b + \sum_{i\neq 1} S_i t\right\} \text{ cts.} \quad (2)$$

To obtain an estimate C_k of the intensity S_1 one needs to normalize the 'open' correlation with the corresponding fraction t of the aperture and the 'closed' correlation with $1 - t$, and subtract the latter from the former:

$$C_k = \frac{C_{k,\rm open}}{t} - \frac{C_{k,\rm closed}}{1-t} \text{ cts per detector area,} \quad (3)$$

so that $C_k = S_1$ is expected.

The variance $\sigma^2_{C_k}$ of the Poisson noise at the position of the considered point source can be extracted from Eqs. (1) and (2) which specify statistically independent measurements of counts:

$$\sigma^2_{C_k} = \frac{C_{k,\rm open}}{t^2} + \frac{C_{k,\rm closed}}{(1-t)^2} = \frac{S_1 + b}{t} + \frac{B + \sum_{i\neq 1} S_i + b}{1-t}. \quad (4)$$

Thus, the signal-to-noise ratio S/N of the point source is expected to be

$$S/N = \frac{S_1}{\sigma_{C_k}} = \frac{S_1}{\sqrt{\frac{S_1+b}{t} + \frac{B+\sum_{i\neq 1} S_i + b}{1-t}}}. \quad (5)$$

The optimum value of t, $t_{\rm opt}$ (when S/N is maximum), can be derived from the conditions:

$$\left[\frac{\partial(S/N)}{\partial t}\right]_{t=t_{\rm opt}} = 0 \quad \text{and} \quad \left[\frac{\partial^2(S/N)}{\partial^2 t}\right]_{t=t_{\rm opt}} < 0. \quad (6)$$

Working out these conditions yields

$$t_{\rm opt} = \frac{1}{1+\sqrt{f}} \quad (7)$$

with

$$f = \frac{B + \sum_{i\neq 1} S_i + b}{S_1 + b} \quad (8)$$

(f may be interpreted as a measure of the ratio of background to source intensity). This result is the same as Fenimore (1978) obtained, except that now the diffuse sky background term has been explicitly incorporated.

2.2. Non-ideal PSF

The spatial response of the system is described by the PSF p_i, the fraction of the source intensity which is found in the $i^{\rm th}$ reconstruction pixel (see Appendix). It takes into account all effects which influence the imaging of a point source such as for instance the detector resolution. These effects may introduce dependencies on the off-axis angle and spectrum of the point source. For a detailed description of these issues in the case of a typical coded aperture X-ray device, see In 't Zand (1992).

Due to the PSF, the 'open' and 'closed' cross correlations will change in value. A fraction $1 - p_k$ of the $S_1\ t$ counts is lost to neighboring detector pixels of which a fraction t again will correlate with the open pattern. The remainder of the $S_1\ t$ counts will contribute to the 'closed' correlation. The PSF has no impact on the B and b terms since the detector is assumed to have a flat response to these terms. Regarding the PSFs that apply to sources $i \neq 1$, these are irrelevant in the cross correlation considered because the probability of contributing to the closed or open component is the same for every element of those PSFs.

Eqs. (1) and (2) become:

$$C_{k,\rm open} = S_1\ t\ p_k + S_1\ t\ (1 - p_k)\ t + \\ + t \left\{ B\ t + b + \sum_{i\neq 1} S_i t \right\} \text{ cts} \quad (9)$$

and

$$C_{k,\rm closed} = S_1\ t\ (1 - p_k)\ (1 - t) + \\ + (1 - t) \left\{ B\ t + b + \sum_{i\neq 1} S_i t \right\} \text{ cts}. \quad (10)$$

The cross correlation value at k now is $C_k = S_1 p_k$ and the Poisson noise variance in C_k is

$$\sigma^2_{C_k} = \frac{S_1 p_k + b}{t} + \frac{S_1(1 - p_k) + B + \sum_{i\neq 1} S_i + b}{1 - t}. \quad (11)$$

Due to the presence of a PSF, intensities of point sources cannot simply be determined by evaluating pixel values, as was done in the above ideal case. Instead, one needs to test the measured PSF against a model PSF for a set of different intensities and positions via a least-squares method. The mathematical framework of this test is elaborated in the Appendix. The resulting 1σ error in the intensity is expected to be

$$\sigma^2_{\rm PSF} = \frac{1}{\sum_k \frac{p_k^2}{\sigma^2_{C_k}}} \quad (12)$$

where $\sigma^2_{C_k}$ is the Poisson variance for reconstruction pixel k and the summation is taken over the extent of the PSF. The expected signal-to-noise ratio of the considered point source over the PSF is

$$S/N = \frac{S_1}{\sigma_{\rm PSF}}$$

where $\sigma_{\rm PSF}$ is specified by Eqs. (12) and (11) so that

$$S/N = S_1 \sqrt{\sum_k \frac{p_k^2}{\frac{S_1 p_k + b}{t} + \frac{S_1(1-p_k) + B + \sum_{i\neq 1} S_i + b}{1-t}}} \quad (13)$$

The optimum open fraction is again given by Eq. (7) but now

$$f = \frac{S_1(1 - \sum p_k^3 / \sum p_k^2) + B + \sum_{i\neq 1} S_i + b}{S_1(\sum p_k^3 / \sum p_k^2) + b}. \quad (14)$$

The results for an ideal PSF (Eq. 8) naturally follow from those for a non-ideal PSF: the ideal PSF is a delta-function which is represented by $p_k = 1$ and $p_i = 0$ for $i \neq k$.

2.3. Discussion

In Eq. (14), f depends on S_i ($i = 1, ..., N$). Therefore, no unique value for $t_{\rm opt}$ can be defined. Particularly for a *wide field* camera circumstances may differ widely from observation to observation, provided B and b are not generally dominating over S_i. For example, the SAX-WFCs may observe fields with only a couple of faint point sources (e.g. the Large Magellanic Cloud field) or fields with tens of bright point sources (e.g. the galactic center field), see Sect. 3. However, one can distinguish the following limiting cases for an ideal PSF (see Fig. 1):

1. If S_1 dominates over $\sum_{i\neq 1} S_i$, B and b, then f approaches 0 and $t_{\rm opt}$ approaches 1. This case is trivial: the detector effectively only measures counts from S_1, the use of a coded aperture then has no benefit.
2. If b dominates over S_i ($i = 1, ..., N$) and B, then f approaches 1 and $t_{\rm opt}$ approaches 0.5.
3. If $\sum_{i\neq 1} S_i + B$ dominates over S_1 and b, then f approaches ∞ and $t_{\rm opt}$ approaches $1/\sqrt{f}$.

For a non-ideal PSF, the first conclusion is cancelled, that is when only one point source illuminates the detector, $t_{\rm opt}$ is not 1 but depends on the PSF and will be less than 1. In this limit, Eq. (14) reduces to

$$f = \frac{1 - (\sum p_k^3 / \sum p_k^2)}{\sum p_k^3 / \sum p_k^2} \quad (15)$$

and $t_{\rm opt}$ thus becomes smaller, for instance $t_{\rm opt} = 0.5$ if $\sum p_k^3 / \sum p_k^2 = 0.5$. In other words, even for very bright point sources it is better to use a coded pattern which does *not* have the maximum open fraction.

How can this result be understood? One may envisage the image of a PSF-smeared point source as a group of fainter ideal point sources (say 'sub sources'). If the PSF-smeared point source is dominating in intensity, the sub sources may well not

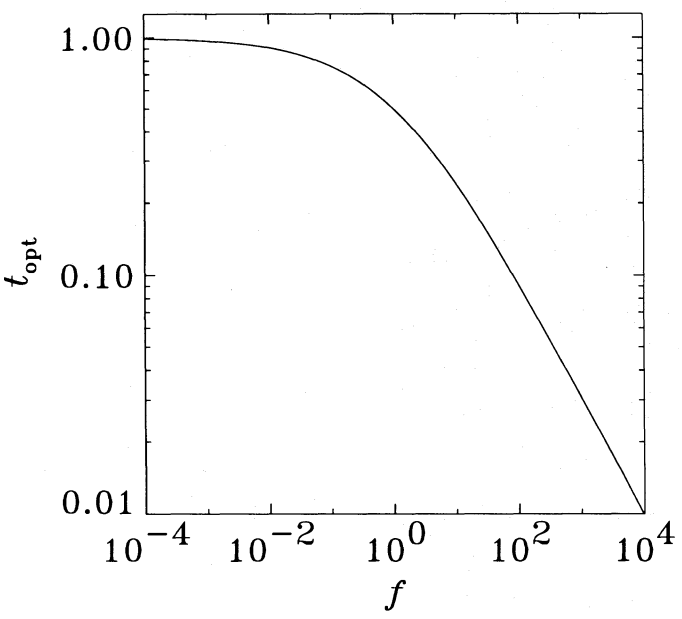

Fig. 1. The relation between t_{opt} and f

be. This is also clear when comparing Eq. (8) with Eq. (14): a fraction $1 - \sum p_k^3 / \sum p_k^2$ of S_1 is lost to $\sum_{i \neq 1} S_i$ in the numerator. Now, if every sub source complies to the above case 3, so will the combination of all these sub sources and as a result case 1 will break down to case 3 (the extend of this break down is of course dependent on the PSF and may be intermediate).

As is clear from Eq. (14), PSF dependencies only apply to bright sources. Hence, a change of the PSF will only affect the signal-to-noise ratio or t_{opt} of bright sources. One could consider altering the PSF to obtain another t_{opt} for bright sources. This may be accomplished by selection of another spacing of the pattern grid because the values of p_k depend on this. For example, if the spacing is chosen to be larger (thus making the pattern coarser), the PSF will tend more to an ideal delta function and $\sum p_k^3 / \sum p_k^2$ will become larger. Thus, for bright sources, t_{opt} will be larger. However, the fact that this will deteriorate the angular resolution might be experienced as a strong disadvantage.

In Sect. 1, t is defined as the fraction of elements that are partially or fully transparent. In order to improve the mechanical stability of coded apertures they are often manufactured with a support grid which has the same spacing as the pattern grid. As a result, the *area* of each transparent element is somewhat smaller, say by a factor of α ($\alpha < 1$). It can easily be shown that a decrease of α in most cases acts counterproductive in the signal-to-noise ratio: taking account of α converts Eq. (13) to

$$S/N = \sqrt{\alpha} S_1 \sqrt{\sum_k \frac{p_k^2}{\frac{S_1 p_k + b/\alpha}{t} + \frac{S_1(1-p_k) + B + \sum_{i \neq 1} S_i + b/\alpha}{1-t}}} \quad (16)$$

Table 1. Characteristics of each SAX-WFC

Detector area (=aperture area)	255×255 mm^2
Active detector area	530 cm^2
Effective area	$430t$ cm^2 at 6 keV
Distance Aperture-Detector	703 mm
Field of View (FWZR)	$40° \times 40°$ (0.47 sr)
Aperture element size	$\sim 1 \times 1$ mm^2
Angular Resolution (FWHM), on-axis	$5'$
Active photon energy range	1.8 – 30 keV
Photon energy resolution (FWHM)	18% at 6 keV
Photon detector depth	32 mm

(all radiation coming through the aperture is reduced by a factor of α). This equation shows that, for all combinations of S_i, B and b, S/N decreases with α[1].

3. Simulations of SAX-WFC observations

The theoretical analysis above shows the effect of t on the signal-to-noise ratio of point sources. Some simplifications were made to provide a clear-cut analysis, such as the omission of the collimator response and of a locally varying t over the aperture. However, to be able to perform a more rigorous quantitative analysis based on realistic conditions of astronomical observations with a coded aperture camera, simulations are needed. This was done by employing the wide field X-ray cameras of the X-ray satellite SAX, with the additional purpose to find an optimum pattern for these instruments.

3.1. SAX-WFC instrument description

For a recent and general overview of the SAX mission and its instruments, see Scarsi (1993). Details about the WFCs are reviewed by Jager et al. (1989, 1992). Here we summarize the key characteristics of the WFCs (see Table 1). The scientific objective of the WFCs is twofold:

- spatially-resolved simultaneous monitoring of compact X-ray sources in crowded fields with high sensitivity (i.e. monitoring of spectral variability over a large range of time scales);
- monitoring of large regions of the sky to detect the occurrence of X-ray transient sources and signal the necessity for follow-up studies with higher-sensitivity narrow-field instruments on board of SAX.

Both WFCs are identical and of the noncyclic box type (Ponman et al. 1987) which has been shown to be more sensitive

[1] We disregard here the point that p_k also changes with α and actually counteracts a diminishing S/N because the PSF becomes sharper. However, in most practical circumstances the PSF is dominated by several other factors and the sharper PSF does not quite neutralize the S/N reduction introduced by α alone

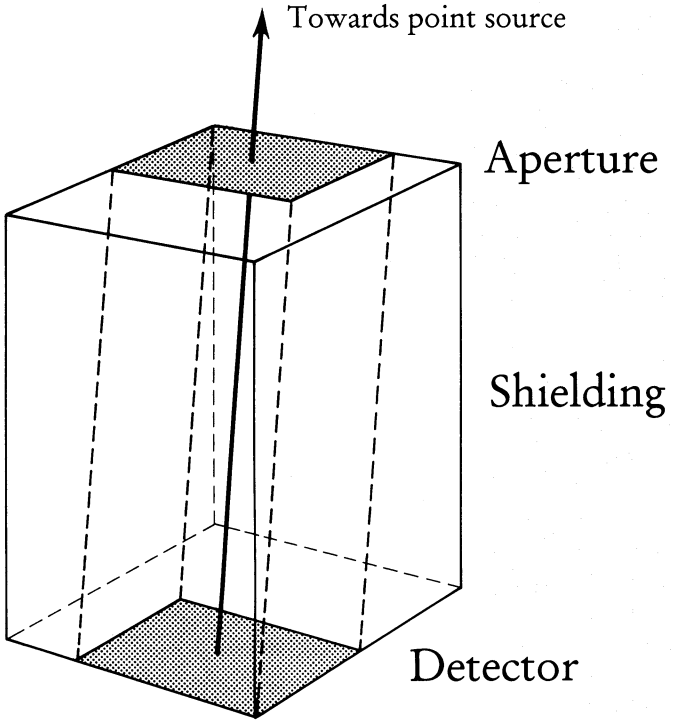

Fig. 2. Schematic drawing of a box-type coded aperture camera. The aperture and detector are equally sized and the aperture pattern is non-cyclic. An off-axis point source projects part of the aperture onto the detector

over cyclic types at wide FOVs (Sims et al. 1980), although its imaging capability is less because of coding errors. However, in most circumstances coding errors can satisfactorily be dealt with (see e.g. Hammersley et al. 1992). Figure 2 presents a schematic drawing of a WFC.

The detectors are multi-wire proportional counters. Since calibration measurements of the detectors were not yet available at the time this work was carried out, a model was used which is based on that of COMIS/TTM (In 't Zand 1992), a similar device currently flying on the *Mir* space station (having a FOV about one sixth times that of each SAX-WFC). This model is thought to be a fair representation of the SAX-WFC detectors since most of the technology is the same as that of COMIS/TTM (being built at the same laboratory) and the goal specifications (e.g. spatial and spectral resolution) are equal.

The PSF in the sky image as recorded with a SAX-WFC is determined by the coded aperture, the detector spatial resolution, photon penetration in the detector gas chamber, the geometry of the internal detector wires, the readout resolution of photon positions and the spatial response to the cross correlation. Due to the detector resolution and photon penetration, the PSF depends on the sky source spectrum and off-axis angle. The maximum extent of the PSF (at high photon energies and large off-axis angles) is about 1°. For an on-axis source with a Crab-like spectrum, the factor $\sum p_k^3 / \sum p_k^2 \approx 0.4$ (Eq. 14) so that, following Eqs. (7) and (15), $t_{\rm opt} \approx 0.45$ for the limiting case that S_1 dominates over all other terms in Eq. (14).

Table 2. Characteristics of sky fields used for the simulations

		Sky field		
Target		galactic center	Large Magellanic Cloud (LMC)	Crab nebula
Pointing	α	265°.6108	83°.8938	82°.8496
(1950.0)	δ	-28°.9168	-69°.2791	21°.5760
No. of sources		27	6	2
Exposure (s)		10^4	10^4	10^4

The models of both the internal detector background and the diffuse sky background were also extrapolated from those measured in flight with COMIS/TTM. The internal detector background level (1.8–30 keV) is 19.3 c s^{-1} per detector area and the level of the isotropic diffuse sky background is 276 t c s^{-1} per detector area. The internal detector background was assumed to have a flat detector response; the sky background was modulated by the detector entrance window support structure (i.e. by the field of view as seen from any detector position). From these values for b and B, $t_{\rm opt}$ for faint on-axis sources, as determined by Eq. (7) in a field with only faint sources, is 0.20.

3.2. Outline of simulations and used aperture patterns

The simulations of SAX-WFC observations were performed within the full active energy band (1.8–30 keV). The point sources were taken from a database which was constructed out of measurements with COMIS/TTM (In 't Zand 1992, chapter 7). This database gives positions, intensities and rough spectral characteristics of 60 X-ray sources in the same energy band. The spectrum of each source was modeled by either a power-law spectrum or a thermal bremsstrahlung spectrum, depending on the best description of the measured COMIS/TTM spectrum.

Three sky fields were used for the simulations, representing extremes of typical observational conditions. The characteristics of these fields are listed in Table 2. The exposure time was taken to be 10^4 s (the expected typical continuous exposure time spent on one sky field with SAX is about 3×10^4 s). The galactic center field consists of 27 point sources with a wide variety of intensities, the LMC field contains 6 faint sources and the Crab field contains 2 bright sources. These particular examples of 'extreme' fields were chosen because COMIS/TTM data is available on them.

In the case of the galactic center field we also simulated the detector response to two components of diffuse galactic emission: according to the finite disk model by Warwick et al. (1985) and to the thin hot plasma model of Yamauchi et al. (1990). The former has an extend of $\sim 60°$ in galactic longitude by ~0°.6 in latitude, while the latter is sized 1°.0×1°.8, centered on the galactic center with an inclination to the galactic plane of 21°.

Five values of t were employed for simulations of observations of these three sky fields: t =0.05, 0.125, 0.25, 0.33 and 0.5. In order to avoid difficulties with respect to bin phases of the cross correlation and to be independent of the limited

possibilities of constructing URA-based patterns for arbitrary open fractions, we used patterns in which the open elements are *randomly* distributed. Although the imaging capability of such patterns is worse than that for patterns based on URAs, this is of no consequence to our analysis because we assume full knowledge of the modeled sky source distribution and thus our analysis is not disturbed by the crosstalk from unforeseen X-ray sources as is common to these types of patterns. The apertures consist of 250×250 elements each 1.02×1.02 mm^2 in size. We note that without full knowledge of the sky distribution, URAs will provide superior performance to that of random patterns and eventually do need to be employed (see Sect. 5).

3.3. Results

The results for any combination of sky field and open fraction are given in Table 3. To compare the results for the different open fractions, we listed the measured signal-to-noise ratios for all significantly detected sources as well as the averaged values. Furthermore, for the galactic center field, we averaged separately two groups of sources: those that have a signal-to-noise ratio less or more than 50 at $t = 0.050$.

For the galactic center field, the best average signal-to-noise ratio, within a 3% margin, is achieved for an open fraction between 0.125 and 0.333. The two other tested open fractions are at least 10% worse in average signal-to-noise ratio. The result differs somewhat between faint and bright sources, the faint sources having better signal-to-noise ratios at lower open fractions than the bright sources, but this is not very significant within the above mentioned margin. The same result applies to the other two fields, preference exists for open fractions only slightly larger than the galactic center field. Rounding up the results for all three fields, the conclusion is that $t_{\rm opt}$ lies between 0.250 and 0.333.

The simulations show that improvements in signal-to-noise ratios, relative to the conventional choice of $t = 0.5$, are on average 20%. This means a considerable benefit in exposure time, up to 40% on average, to obtain similar signal-to-noise ratios. Other criteria, originating in data handling issues, may boost the argument for smaller open fractions even further. For example, it is evident that lower t values result in less data to be telemetred down to Earth, at *no expense* of sensitivity. Furthermore, if it is common that telemetry limits the data coverage for $t = 0.5$, resulting in loss of data, the profit in signal-to-noise ratio of $t = t_{\rm opt}$ data with respect to $t = 0.5$ data may even be larger than already concluded from the above analysis.

4. Additional constraints in the aperture choice

There are a few additional considerations which need to be addressed before the final choice for an aperture. First, there is the geometry of the aperture. With respect to geometry, a coded aperture pattern is defined by the shapes and sizes of the complete aperture plate and those of the pattern elements. The shape and size of the SAX aperture plates is matched with that of the square detectors (in order to create a 'box' type coded aperture camera, see Sect. 3.1). The choice for the size of the pattern elements is motivated by the goal to achieve the best angular resolution possible within the limits dictated by the spatial resolution of the detectors (0.5 mm FWHM) and the possibilities of the etching technique used to fabricate the holes in an aluminium plate. This results in a choice of square aperture elements of about 1×1 mm^2. This size, combined with the total aperture size of 255×255 mm^2, dictates the number of pattern elements.

The simulations show a favorable t range of 0.25–0.33. This result is based on the assumption that the sky reconstruction is performed with a cross correlation algorithm. However, one might consider reconstruction of the intensity without the use of the coded aperture pattern, from the count rate versus time diagram (the 'lightcurve'). This is feasible if an X-ray source produces a unique feature in this diagram. X-ray burst sources are strong candidates for such a reconstruction. Hence, this option clearly matters for SAX, since the study of X-ray bursts is part of the scientific objective of the WFCs.

The sensitivity to an event in the lightcurve is set by the noise level of the background, where the background is the count rate due to all celestial X-ray sources and the internal detector background. The standard deviation of this background level is given by:

$$\sigma_{\rm BG} = \sqrt{\left(B + \sum_i S_i\right) t + b} \quad \text{cts per time unit.} \qquad (17)$$

The ratio V of the signal-to-noise ratio of the event as seen in the light curve (say at the peak intensity) to that as seen in a cross correlation reconstruction is, therefore, given by

$$V = \sqrt{\frac{t}{(1-t)\sum_k p_k^2}}. \qquad (18)$$

Calculations show that, for the SAX-WFCs, $V > 1$ if $t > 0.20$. Therefore, for $0.25 < t < 0.33$, the total-detector lightcurve is more sensitive to clearly recognizable events than the cross correlation. Furthermore, this advantage increases with t and it is best to have an open fraction as high as possible from this viewpoint. Therefore, within the range for t defined above, we consider $t = 0.33$ to be the optimum choice.

5. The aperture pattern

Having narrowed down the open fraction and geometry of the coded aperture, we need to address the arrangement of open and closed pattern elements in the aperture. As discussed above (see Sect. 2) it is very desirable to use uniformly redundant arrays (URAs). The question now is: given the selected open fraction of 0.33 for the SAX-WFCs, what choice is there in URA patterns complying to this value?

Recently, Skinner & Grindlay (1993) proposed patterns to be utilized in instruments covering a very wide photon energy passband (2 to 3 decades). The best patterns are build out of two URAs with different spatial scales. As a natural consequence

Table 3. Signal-to-noise ratios for point sources in simulated SAX observations of three sky fields, the exposure time is 10^4 s.

a. 24 sources in the galactic center field

	Open fraction t				
Source	0.050	0.125	0.250	0.333	0.500
X1700-377	14.8	23.7	26.7	26.9	24.2
X1702-429	10.9	13.3	12.5	13.3	11.9
X1702-363	273.0	324.7	336.5	328.5	296.7
X1705-440	66.1	73.3	69.3	68.6	60.5
X1708-407	16.4	17.1	15.9	15.5	16.5
X1724-307	12.7	14.7	12.7	13.9	9.1
X1728-337	53.6	60.1	59.1	57.3	48.5
X1728-247	6.9	10.0	10.2	10.3	7.7
X1728-169	75.6	86.9	86.9	83.4	74.6
X1731-260	64.4	65.6	62.6	58.7	52.7
X1735-269	14.6	15.6	13.7	13.8	11.0
X1735-444	37.1	41.1	39.8	40.0	32.8
X1742-294	21.8	24.5	21.9	22.1	16.0
X1744-300	23.4	30.0	27.9	26.1	22.1
X1744-361	34.4	35.4	33.3	31.0	26.9
X1744-265	234.8	263.2	256.1	246.5	216.6
X1746-370	9.5	8.6	8.2	9.1	9.6
X1755-338	34.9	38.0	31.7	31.0	28.8
X1758-250	417.2	515.3	536.4	529.7	477.7
X1758-205	250.0	302.9	310.7	301.6	267.5
X1811-171	82.2	87.2	80.1	79.5	65.4
X1813-140	108.9	122.0	120.2	115.0	101.9
X1820-303	152.9	177.0	177.1	170.7	150.5
X1822-371	8.5	10.8	12.4	12.6	9.7
Average	84.4	98.4	98.4	96.0	85.0
Average $< 50.0_{t=0.05}$	18.9	21.8	20.5	20.4	17.4
Average $> 50.0_{t=0.05}$	161.7	188.9	190.5	185.4	164.8
Total counts	1.6×10^6	3.7×10^6	7.3×10^6	9.7×10^6	1.5×10^7

b. The LMC field

	Open fraction t				
Source	0.050	0.125	0.250	0.333	0.500
X0521-720	25.2	29.7	32.0	30.8	29.0
X0532-664	13.7	23.0	26.4	26.5	25.5
X0538-641	18.1	21.3	22.0	20.0	20.0
X0540-697	27.7	30.3	30.8	29.0	26.0
Average	21.2	26.1	27.8	26.6	25.1
Total counts	3.1×10^5	5.5×10^5	9.7×10^5	1.2×10^6	1.8×10^6

c. The Crab nebula field

	Open fraction t				
Source	0.050	0.125	0.250	0.333	0.500
X0531+219	497.5	733.9	883.8	912.8	885.1
X0535+262	320.3	493.1	580.4	589.6	554.9
Average	408.9	613.5	732.1	751.2	720.0
Total counts	9.1×10^5	2.3×10^6	4.8×10^6	6.5×10^6	9.8×10^6

these patterns have open fractions less than 0.5. In particular, Skinner & Grindlay investigated performances of $t = 0.25$ patterns (combining two $t \approx 0.5$ patterns), with emphasis on the minimization of false peaks. The performance of these patterns was found to be better than those of random patterns, although worse than that of a single URA. Thus, a compromise was established between performance and applicability in a wide photon energy passband. For the relatively narrow passband of the SAX-WFCs, there is no need for such a compromise. Also, we anticipate that the large scale structures in those patterns will, due to interference with large structures of the detector (e.g. support structure bars), introduce extra coding problems. Therefore, we pursue possible solutions of *single* URAs with low open fractions and employ the following reasoning.

A URA a_i (a_i is either 0 or 1) is related to a cyclic difference set through its indices: all i ($i = 0, ..., n-1$) for which $a_i = 1$ need to be a member of a cyclic difference set. a_i is then characterized by inter-element spacings of $a_i = 1$ that are uniformly covered for all spacings. The cyclic autocorrelation c_l of the URA is two valued (Baumert 1971):

$$c_l = \sum_{i=0}^{n-1} a_i a_{\mathrm{mod}(i+l,n)} = \begin{cases} k & \text{if } \mathrm{mod}(l,n) = 0 \\ \frac{k(k-1)}{n-1} & \text{if } \mathrm{mod}(l,n) \neq 0 \end{cases} \quad (19)$$

(i.e. a periodic delta function), where k is the number of members of the set (thus, the membership fraction within n is given by $t = k/n$, this parameter defines the open fraction of patterns to be constructed from this array). URAs can be folded in two dimensions. The folding procedure should follow some rules to preserve the ideal autocorrelation in two dimensions (see e.g. Miyamoto 1977 and Proctor et al. 1979). The resulting patterns are proven to be advantageous against purely random patterns (see e.g. Gunson & Polychronopulos 1976 and Fenimore & Cannon 1978), whose cyclic autocorrelations are k-valued. This advantage is not relevant in the simulations discussed in Sect. 3 because we proceeded from full knowledge of the observed sky.

Fenimore & Cannon (1978) introduced URA patterns from so-called *quadratic residue* cyclic difference sets, which are defined as follows: if there exists an integer u such that $n = 4u - 1$ is prime, the i for which $a_i = 1$ are given by all values $\mathrm{mod}(j^2, n)$ ($j = 1, ..., n-1$). For such a pattern, $k = \frac{n}{2} - \frac{1}{2}$ so that for $n \gg 1$, t approaches $\frac{1}{2}$.

There is also a class of URA patterns with $t \approx 0.25$. These are based on *biquadratic residue* cyclic difference sets (see e.g. Baumert 1971) and their definition is analogous to that of quadratic residue sets: if there exists an odd integer u such that $n = 4u^2 + 1$ is prime, the i for which $a_i = 1$ are given by all values $\mathrm{mod}(j^4, n)$ ($j = 1, ..., n-1$). For such patterns $k = \frac{n}{4} - \frac{1}{4}$ so that for $n \gg 1$, t approaches $\frac{1}{4}$.

We are interested in URA patterns with an open fraction of 0.33. Furthermore, we here choose not to be too strict in the conditions for the pattern since in a practical situation the geometry of a proportional counter prevents the recording of the complete pattern at all times anyway. This is even more true in box-type coded aperture cameras such as the SAX-WFCs, where off-axis sky sources principally project just part of the pattern on the detector (see Fig. 2).

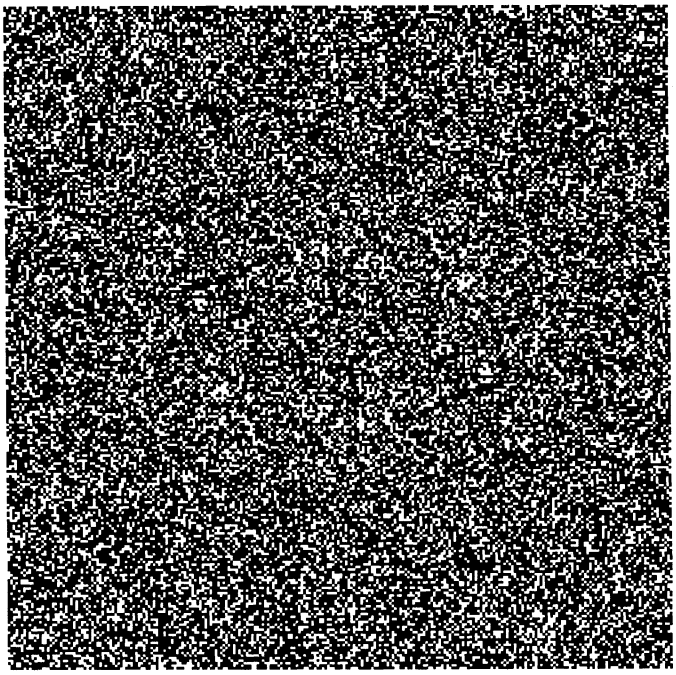

Fig. 3. A picture of the coded aperture pattern chosen for both SAX-WFCs. White squares represent holes in the pattern

To our knowledge, there are no genuine URA patterns with $t = 0.33$. However, following a routine extrapolated from the two mentioned above, we have found a pattern type with $t \approx 0.33$ that is very close to a URA-type. Instead of two valued, the cyclic autocorrelation of the resulting array is four valued. This pattern is defined as follows: if there is an integer u for which $n = 3u + 1$ is prime, the indices i of all $a_i = 1$ are given by all values $\mathrm{mod}(j^3, n)$ ($j = 1, ..., n-1$). We call this type of array a '*triadic residue*' array. The zero-shift autocorrelation value (number of open elements) is u. The three non-zero-shift autocorrelation values, λ_1, λ_2 and λ_3, occur close (within 1% for $u \gg 1$) to $\frac{1}{3}u$ and $\sum_{i=1}^{3} \lambda_i = u - 1$. In order to size such a pattern to about 250×250, there is ample choice in u: for instance $u = 21840$. In this case the array may be folded row wise in a 256×256 pattern. The array is just 15 elements too short to fill the complete pattern. Another, even more favorable choice is $u = 21846$. We find this array to have the property $\lambda_1 = \lambda_2$, so that the cyclic autocorrelation for this particular case is 3 instead of 4 valued. The array can be folded row wise also in a 256×256 pattern, having only 3 elements short to complete the pattern. The latter pattern is a very favorable one from our point of view. Therefore, we consider this pattern to be the best choice for the apertures of both SAX wide field cameras[2]. Figure 3 shows a binary representation of this pattern.

[2] No distinction is made here between each camera since no separate observing programs are foreseen for each camera

6. Summary

We have shown, both theoretically and through numerical simulations of observations with the SAX-WFCs, that there exists a dependency of the optimum open fraction of a coded aperture (with respect to the signal-to-noise ratio of point sources) on the spatial response of the instrument to point sources. As a result, bright point sources follow faint sources in the tendency to have higher signal-to-noise ratios at lower aperture open fractions.

A unique optimum value for the open fraction, applicable to diverse sky fields as observed with one instrument, can in general not be given. However, for the SAX-WFCs the optimum is limited to within a certain range where the signal-to-noise ratio, averaged over the point sources in the field of view, does not deviate by more than a few percent from the optimum value. Although profit in signal-to-noise ratios is limited to a factor $\sqrt{2}$ at most, we stress that this results in a considerable benefit in the exposure time needed to arrive at similar signal-to-noise ratios and that telemetry coverage considerations may be at least as important as signal-to-noise considerations.

In the case of the wide field cameras to be employed in SAX, simulations show a preferred open fraction between roughly 0.25 and 0.33. Combining this preference with the consideration of good sensitivity to X-ray bursts in the total count rate of the detector, the choice has been made for a pattern with $t = 0.33$ and nearly ideal cyclic cross correlation characteristics. The chosen pattern is based on a 'triadic' residue set with $u = 21846$ which is folded 2-dimensionally row wise and completed to a 256×256 pattern with 3 closed pattern elements.

Acknowledgements. We would sincerely like to thank Jaap Schuurmans (SRON) for the help in the setup of the software used for this study. Also, we thank him and Gerard Stevens (SRON) for many useful discussions.

Appendix A: the point spread function (PSF)

In its continuous form, the PSF is characterized by the normalized function $f(x-x_0, y-y_0)$ which specifies per unit area the fraction of the point-source intensity contained in the infinitesimal area $([x-x_0]...[x-x_0]+\mathrm{d}x, [y-y_0]...[y-y_0]+\mathrm{d}y)$, when the point source is located at position (x_0, y_0). In this continuous form the point-source image is given by:

$$m(x,y) = s\, f(x-x_0, y-y_0) + l \qquad \text{intensity/area} \qquad \text{(A1)}$$

where s is the total point-source intensity and accounting for a constant background level l. Actually, the PSF will be integrated on a regular grid with bin sizes $\Delta x \times \Delta y$ and sampled at (x_i, y_i), yielding a reconstructed sky of:

$$m_i = s\, p_i + L \qquad \text{intensity} \qquad \text{(A2)}$$

where $L = l\ \Delta x\ \Delta y$ and

$$p_i = \int_{x_i-\frac{1}{2}\Delta x}^{x_i+\frac{1}{2}\Delta x} \int_{y_i-\frac{1}{2}\Delta y}^{y_i+\frac{1}{2}\Delta y} f(x-x_0, y-y_0)\mathrm{d}x\ \mathrm{d}y. \qquad \text{(A3)}$$

The point-source intensity and position may be obtained by applying a least-squares method on the appropriate part of the reconstructed sky. This involves minimizing the S-statistic (Lampton et al. 1976)

$$S = \sum_{i=1}^{N} \frac{(C_i - m_i)^2}{\sigma_{C_i}^2}, \qquad \text{(A4)}$$

leaving free simultaneously s, x_0, y_0 and L (N is the number of pixels in the appropriate part of the sky, C_i the observed value of the reconstructed sky at element i and $\sigma_{C_i}^2$ its variance). If the model m_i is consistent with the data C_i and the deviations $\frac{C_i - m_i}{\sigma_{C_i}}$ are independent and approximately Gaussian distributed with a zero mean (see e.g. Bevington 1969), the minimum value $S_{\min}$ is a sample from the probability distribution of Pearson's χ^2-statistic for $N - N_p$ degrees of freedom (N_p is the number of parameters for which a solution is searched and is 4 in this case). The model is regarded to be discrepant with the data if the integrated probability

$$\alpha \equiv \int_{S_{\min}}^{\infty} f(\chi^2_{N-N_p})\mathrm{d}\chi^2 \qquad \text{(A5)}$$

is smaller than a certain threshold value, where $f(\chi^2_{N-p})$ is the probability distribution of χ^2 for $N - N_p$ degrees of freedom. Lampton et al. argue that a threshold value of 10% is a reasonable compromise between security of conclusions and sensitivity toward detecting false models.

Given a correct model, errors on all parameters with a certain confidence level P may be determined. The P confidence-level region in parameter space is conditioned by the contour (Lampton et al. 1976):

$$S = S_{\min} + \chi^2_{N_p}(P) \qquad \text{(A6)}$$

with χ^2 distributed for a number of freedoms equal to the number of parameters N_p. For a simultaneous fit of all four parameters, a confidence level of 68% is equivalent with $\chi^2_4(68\%) = 4.7$. Often, so-called '1-sigma' errors are quoted, because they are easier to infer. For a parameter with a fitted value of p the 1-sigma error σ_p is defined by:

$$S(p+\sigma_p) = S(p) + 1.00 \qquad \text{(A7)}$$

For the four parameters considered here, working out Eq. (A7) (to the second order Taylor expansion) and including Eqs. (A2) and (A4) results in:

$$\sigma_s^2 = \frac{1}{\sum \frac{(p_i)^2}{\sigma_{C_i}^2}}, \qquad \text{(A8)}$$

$$\sigma_{x_0}^2 = \frac{1}{s^2 \sum \frac{\left(\frac{\partial p}{\partial x}\right)_i^2}{\sigma_{C_i}^2}}, \qquad \text{(A9)}$$

$$\sigma_{y_0}^2 = \frac{1}{s^2 \sum \frac{\left(\frac{\partial p}{\partial y}\right)_i^2}{\sigma_{C_i}^2}} \quad \text{and} \qquad \text{(A10)}$$

$$\sigma_L^2 = \frac{1}{\sum \frac{1}{\sigma_{C_i}^2}}. \qquad \text{(A11)}$$

These 1-sigma errors indicate the accuracy of estimating all parameters separately at a 68% confidence level, or simultaneously at a confidence level of 21%.

References

Baumert L.D., 1971, "Lecture Notes in Mathematics No. 182: Cyclic Difference Sets", Springer, Berlin

Bevington P.R., 1969, Data reduction and error analysis for the physical sciences, McGraw-Hill, New York

Caroli E., Stephen J.B., Di Cocco G., Natalucci L., Spizzichino A., 1987, Space Sci. Rev. 45, 349

Dicke R.H., 1968, ApJ, 153, L101

Fenimore E.E., 1978, Appl. Opt. 17, 3562

Fenimore E.E., Cannon T.M., 1978, Appl. Opt. 17, 337

Gunson J., Polychronopulos B, 1976, MNRAS 177, 485

Hammersley A.P., Ponman T.J., Skinner G.K., 1992, Nucl. Instrum. Meth. A311, 585

In 't Zand J.J.M., 1992, A coded-mask imager as monitor of galactic X-ray sources, Ph.D. Thesis, University of Utrecht

Jager R., In 't Zand J.J.M., Schuurmans J.J., Heise J., Mels W.A., Brinkman A.C., 1989, in "EUV, X-ray, and gamma-ray instrumentation for astronomy and astrophysics", SPIE Proc. Ser. 1159, 2

Jager R., Heise J., In 't Zand J.J.M., Brinkman A.C., 1993, Adv. Space Res. 13, (12)315

Lampton M., Margon B., Bowyer S., 1976, ApJ 208, 177

Miyamoto S., 1977, Space Sci. Instrum. 3, 473

Ponman T., Hammersley A., Skinner G., 1987, Nucl. Instrum. Meth. A262, 419

Proctor R.J., Skinner G.K., Willmore A.P., 1979, MNRAS 187, 633

Scarsi L., 1993, A&AS 97, 371

Sims M.R., Turner M.J.L., Willingale R., 1980, Space Sci. Rev. 5, 109

Skinner G.K., 1984, Nucl. Instrum. Meth. 221, 33

Skinner G.K., Grindlay J.E., 1993, A&A 276, 673

Yamauchi S., Kawada M., Koyama K., Kunieda H., Tawara Y., 1990, ApJ 365, 532

Warwick R.S., Turner M.J.L., Watson M.G., Willingale R., 1985, Nat 317, 218

This article was processed by the author using Springer-Verlag LaTeX A&A style file version 3.