

Sensitivity of coded mask telescopes

Gerald K. Skinner

University of Maryland, College Park, Maryland 20742, USA, and Center for Research and Exploration in Space Science and Technology, and NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

(skinner@milkyway.gsfc.nasa.gov)

Received 9 January 2008; accepted 25 February 2008;
posted 2 April 2008 (Doc. ID 91482); published 12 May 2008

Simple formulas are often used to estimate the sensitivity of coded mask x-ray or gamma-ray telescopes, but these are strictly applicable only if a number of basic assumptions are met. Complications arise, for example, if a grid structure is used to support the mask elements, if the detector spatial resolution is not good enough to completely resolve all the detail in the shadow of the mask, or if any of a number of other simplifying conditions are not fulfilled. We derive more general expressions for the Poisson-noise-limited sensitivity of astronomical telescopes using the coded mask technique, noting explicitly in what circumstances they are applicable. The emphasis is on using nomenclature and techniques that result in simple and revealing results. Where no convenient expression is available a procedure is given that allows the calculation of the sensitivity. We consider certain aspects of the optimization of the design of a coded mask telescope and show that when the detector spatial resolution and the mask to detector separation are fixed, the best source location accuracy is obtained when the mask elements are equal in size to the detector pixels. © 2008 Optical Society of America

OCIS codes: 340.7430, 100.1830, 110.4280.

1. Introduction

Coded mask telescopes have been widely used in x-ray and gamma-ray astronomy, particularly at those energies where other imaging techniques are not available or where the wide field of view possible with the technique is important. Recent examples of astronomical applications of the technique include the Burst Alert Telescope (BAT) instrument on the Swift spacecraft [1] and three of the instruments on the International Gamma-Ray Astrophysics Laboratory (INTEGRAL) [2]; other examples are described in [3–6]. The technique is based on recording the shadow of a mask containing both transparent and opaque regions in a pattern that allows an image of the source of the radiation to be reconstructed. Coded mask imaging has been reviewed by Caroli *et al.* [7], with more recent discussion of some aspects of the technique by Skinner [8].

Many possible mask patterns have been discussed. The mask may simply contain randomly placed holes [9,10] or it may be based on geometric patterns

[11,12]; indeed, almost any design can be used without losing the imaging capability [8]. Most work has been based on patterns comprising holes placed on a regular rectangular or hexagonal grid according to some algorithm. Discussion of the choice of algorithm for placing the holes has concentrated on designs in which the (cyclic) autocorrelation function of the pattern, sampled at shifts corresponding to a whole number of cells of the grid, is bivalued with a central peak and flat wings. An extensive literature [13–28] exists on arrays which have this property, which are usually termed uniformly redundant arrays (URAs). For URA-based masks, in certain well-defined circumstances, cross correlation of the recorded data with an array which corresponds to the mask pattern (with a scaling and offset applied) leads to images with a point source response function (PSF) having a central peak and perfectly flat sidelobes. As image reconstruction by cross correlation can be shown (again in specific circumstances, to be discussed below) also to yield the best possible signal-to-noise ratio, such solutions have attracted widespread attention.

Variants of URAs have been proposed (e.g., modified uniformly redundant arrays, MURAs [29]; see

also [30]) in which the same ideal PSF is obtained when the reconstructing array differs marginally from the coding pattern. Provided the number of elements is large, the signal-to-noise ratio is essentially the same as for URAs.

URAs (and MURAs) provide a mathematically satisfying solution to the problem of mask design, but their advantages in practice are less evident. The circumstances in which the ideal response is obtained relate to the cyclic nature of the patterns. The PSF is free from spurious responses (ghosts or sidelobes) only if the shadow recorded is always of a whole number of cycles of a repeating pattern. It can be arranged that this condition is met over a limited field [the so-called fully coded field of view (FCFOV)] but for sources outside this region [in the partially coded field of view (PCFOV)] the shadow of the edge of the mask will appear in the recorded data. Sources in the PCFOV can produce spurious responses within the FCFOV and vice versa. It is possible to block with a collimator consisting of slats or tubes the flux from sources in the PCFOV but, as pointed out by [31], this not only narrows the observable field but results in attenuation of the recorded signal even within the FCFOV. Thus, it is often the case that random mask patterns are as good as any other and indeed a random pattern was selected for the mask of the very successful BAT instrument on the Swift satellite [32].

The ghost images and other imaging artifacts that arise from partial coding, from nonuniform background, or from other effects present in real instruments can be alleviated provided one has a good understanding of the instrument and adequate computing resources. Although for sufficiently long observations or combinations of observations these systematic errors will inevitably become important, many methods are available to reduce them [33–43]. Typically these involve fitting and subtracting bright sources, or otherwise taking them into account, and carefully modeling background nonuniformities.

Even with advanced image reconstruction techniques Poisson (or photon) noise due to limited counting statistics leads to random errors, placing a limit on the sensitivity. This limit is particularly important for very short observations of relatively bright sources, as is the case in the detection of gamma-ray bursts, for example. However, even for long-term survey observations it places an intrinsic limit on the sensitivity that can be achieved, however sophisticated the analysis technique.

We here consider the calculation of the statistical limit to the significance with which a point source can be observed in the presence of Poisson noise on both the flux from the source and the detector background. We pay particular attention to the assumptions that are made in the derivation of the formulas and the circumstances in which they are valid. In Section 2 we list the assumptions that have been made, explicitly or implicitly, in many previous approaches to this problem. In successive sections we attempt to provide useful expressions for the

signal-to-noise ratio where subsets of these assumptions do not hold.

As we deal only with the Poisson-noise-limited sensitivity of the instrument, the formulas given here only place an upper bound on the signal-to-noise ratio that can be obtained. The flux is assumed to be assessed by a procedure which maximizes the signal-to-noise ratio for the source in the presence of an unknown uniform detector background. In effect, data from all the detector pixels are combined, with weights that achieve this objective, but that consequently do not necessarily minimize imaging artifacts. Artifacts due to imperfect modeling of the background, to imaging in regions of the PCFOV where the instrument imaging response is intrinsically poor, or to using in the analysis an insufficiently precise description of the instrument response, may add systematic noise to the Poisson noise. If analysis methods are adopted that are designed to reduce such artifacts (for example, by reducing or eliminating ghost responses in images of fields containing extended sources or multiple point sources), they may increase the effect of Poisson noise. In the limit inverse matrix techniques (or inverse filtering) may completely remove systematic errors but are well known to lead to noise amplification.

By not considering errors other than statistical ones we effectively suppose that observations are short enough that systematic errors are well below the limit imposed by statistics and ignore the fact that for sufficiently long integration times they will eventually become important. In a well-designed instrument and with appropriate treatment of the data, performance close to the Poisson limit can nevertheless be achieved even for comparatively long observations, particularly if the telescope orientation is dithered or scanned during the observation to reduce the systematic noise, as forms part of the INTEGRAL observing strategy [2], is important for Swift/BAT survey work [44], and is planned for the Energetic X-ray Imaging Survey Telescope (EXIST) [45,46].

It is emphasized that the sensitivity considered is that for a point source at a known position or when measuring the flux in a particular pixel of an image; any flux from other point sources, or from extended emission away from the pixel under consideration, is handled by treating it as additional background.

2. Assumptions Frequently Made

Simple analyses of the signal-to-noise ratio obtainable with a coded mask telescope often assume, explicitly or implicitly, that the following conditions are met:

1. Half of the mask elements are open and half are closed (mask element open fraction $f_e = \frac{1}{2}$).
2. The holes are identical and equal in size to the pitch of the grid on which they are placed. For example, there is no supporting structure of the sort illustrated in Fig. 1 and the overall open fraction f is then equal to f_e .
3. The measurement uncertainty is the same for every detector element. We here characterize the

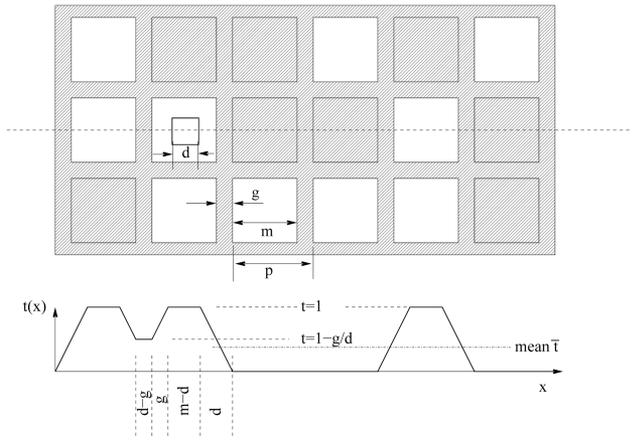


Fig. 1. Part of a mask in which the opaque elements are supported by a grid with bar width g and pitch p , leaving holes of width m . The plot beneath shows the response of a square detector pixel of side d as it is moved across the mask shadow along the line shown. The widths of the transition regions are indicated in the case where $g < d < m$.

source strength by the number of counts per unit area of detector where the mask is open, S , and the background by $B_{f,t}$, the number of background events per unit area of detector. This implies assuming that $B_{f,t} \gg S$. The subscript f,t is a reminder that as the background generally contains a significant contribution from diffuse sky emission, $B_{f,t}$ will generally be a function of f , of the solid angle of the field of view, and of the transmission (discussed below) of the mask elements. It is often convenient to assume that other sources in the field of view give rise to flux that is uncorrelated with the shadow pattern corresponding to the source under consideration. Their contribution to the detector counts can then be considered to be smeared out and included in $B_{f,t}$. Any such component of $B_{f,t}$, too, will depend on the f and the t values.

4. Each mask element is either totally opaque ($t_0 = 0$) or totally transparent ($t_1 = 1$). If this condition is not met, and if there is a background component due to the sky or other sources, then $B_{f,t}$ will also depend on the actual values of t_0 and t_1 .

5. The number of events is such that Poisson statistics may be treated as Gaussian. It will be shown in Section 5 that this supposition is a good one except in the most extreme circumstances, so that even where it is formally still being made below, it may often be ignored.

6. The detector has perfect spatial resolution so that the exact position of arrival of each photon is known, as opposed to having finite size pixels or a realistic continuous position readout with some measurement uncertainty.

The signal-to-noise ratio most often evaluated is the estimate of the intensity of the flux from a source relative to the uncertainty in its measurement (S/σ_s in the terminology used below). This can be different from the value relative to the noise in the absence of the source (S/σ_I below). We will generally consider

the former parameter because the latter can readily be obtained from the same formulas, but where relevant we will note as an assumption that it is indeed S/σ_s that is required:

7. The relevant signal-to-noise ratio is the estimate of the intensity of the flux from a source relative to the uncertainty in its measurement. This assumption is never needed if assumption 3 is made, as the two signal-to-noise ratio estimates are then the same.

There are two more simplifying assumptions that we will generally continue to make:

8. The sensitivity to be discussed is a typical value over part or all of the region imaged and/or the mask elements through which radiation is received are sufficiently numerous that numbers based on the average open fraction may be used. Results are then the same for (M)URA-based masks and random ones.

9. The measurement uncertainty due to background counting statistics is uniform across the detector plane. Sometimes a highly nonuniform background may be modeled and subtracted out. Even if the expectation level of the residual background is then everywhere zero, the random fluctuations can be more important in some regions than others, in which case this assumption is not valid.

An example of when assumption 8 is important arises when the detector plane consists of pixels that are on the same pitch as the mask elements or one that is a submultiple of it. If one considers only source directions such that the detector pixels are either fully shadowed or fully illuminated, then the sensitivity is the same as if the detector had perfect spatial resolution. However, in other directions the sensitivity is up to a factor of 2 poorer. It is generally most useful to average out such effects. An exception arises if an observation, or sequence of observations, is planned such that, for a particular source selected for study, the shadow boundaries always fall between detector pixels (e.g., the seven point hexagonal pointings of the INTEGRAL/Spectrometre pour Integral (SPI) instrument [47]).

Some aspects of a real system may invalidate several of these assumptions. For example, at high energies masks are likely to be partially transparent, contrary to assumption 4, and the large thicknesses which are employed to minimize the consequent loss in sensitivity mean that the apparent hole size and shape become functions of an off-axis angle and one no longer has the simple situation assumed in 2.

3. Relaxing Conditions 1, 2, and 3: Allowing Masks with Arbitrary Pattern and Detector Background not Necessarily Dominant

We will consider first a coded mask telescope in which the mask pattern is not necessarily 50% open and 50% closed (breaking assumption 1). Furthermore, we

suppose that the elements are not necessarily simple squares or hexagons and the mask pattern may contain structures other than the elements themselves (breaking assumption 2), an important example being the presence of a supporting grid as shown in Fig. 1. At the same time we will consider the general case where it is not necessarily true that $S \ll B_{f,t}$, breaking assumption 3. This case, and the associated issue of the optimum open fraction of the mask when $S \not\ll B_{f,t}$, have been discussed in the literature [10,13,28,48–51] but we here try to provide a unified approach and to correct some errors that have arisen.

A. Signal-to-Noise Ratio

The important parameter in this case is the open fraction of the mask, f . This takes into account the fraction of mask elements that are open, f_e , but may also be affected by other aspects of the design. Thus, $f = f_e(m/p)^2$ in the case of the example structure in Fig. 1. As for the moment we still ignore any effects of finite detector resolution, the total detector area A may be considered to be divided into an area fA that sees the source plus detector background (cosmic and particle) and an area $(1-f)A$ that measures only background. If the source is in the PCFOV then A should be taken as the area of that part of the detector that would, but for the mask, see the source. The expectation values for the counts measured in the two regions are

$$C_S = fA(S + B_{f,t}), \quad (1)$$

$$C_B = (1-f)AB_{f,t}. \quad (2)$$

Our estimate of the source strength is then

$$\hat{S} = \frac{C_S}{fA} - \frac{C_B}{(1-f)A}, \quad (3)$$

with variance

$$\sigma_S^2 = \frac{C_S}{(fA)^2} + \frac{C_B}{(1-f)^2A^2}, \quad (4)$$

$$= \frac{S + B_{f,t}}{fA} + \frac{B_{f,t}}{(1-f)A}. \quad (5)$$

The signal-to-noise ratio of the source flux measurement is thus

$$\frac{\hat{S}}{\sigma_S} = S \sqrt{\frac{f(1-f)A}{(1-f)S + B_{f,t}}} \quad (6)$$

(assumptions 4–9).

Note that in this case we obtain the same result whether f deviates from $\frac{1}{2}$ because of a supporting grid

as in Fig. 1, or whether it simply reflects the fraction of mask elements that are open (f_e), or a combination of the two. Indeed, it applies to an arbitrary mask design.

Some particular cases are (a) the limiting case $B_{f,t} \gg S$ for which

$$\frac{\hat{S}}{\sigma_S} = S \sqrt{\frac{f(1-f)A}{B_{f,t}}} \quad (7)$$

(assumptions 3–9),

and (b) the special case $f = \frac{1}{2}$ for which the signal-to-noise ratio can be written

$$\frac{\hat{S}}{\sigma_S} = \frac{(S/2)A}{\sqrt{[(S/2) + B_{1/2}]A}} = \left(\frac{S}{\sigma_S}\right)_{\text{ref}} \quad (8)$$

(assumptions 1, 4–9),

which is simply the number of counts due to the source divided by the square root of all the counts (source plus background, the latter including any contribution from other sources in the field of view). Below we will use this value as a reference against which to compare the sensitivity in other cases. Finally, in (c) when both of these conditions apply we have the widely quoted expression for the signal-to-noise ratio for an ideal 50% open coded mask instrument in the background dominated case

$$\frac{\hat{S}}{\sigma_S} = \frac{S}{2} \sqrt{\frac{A}{B_{1/2}}} \quad (9)$$

(assumptions 1, 3–9).

The signal-to-noise ratio as defined above is the ratio of the source strength to the uncertainty in its measure. For knowing whether a source is significantly detected or not, a more appropriate measure is the ratio of the measured flux to the noise in the surrounding region of an image. Consider a test position away from the true source position. The expected distribution of events for a hypothetical source at this position should ideally be uncorrelated with the actual distribution due to the real source. If there is some residual correlation, then systematic effects (ghosts or sidelobes) will result. However, we are here concerned with random noise so we may suppose that all the recorded events will be divided between the region measuring the flux from the hypothetical source and the region measuring the background, in proportion with the areas of the two regions

$$C'_S = fA(fS + B_{f,t}), \quad (10)$$

$$C'_B = (1-f)A(fS + B_{f,t}). \quad (11)$$

With these values Eq. (3) gives an expectation value of zero [52,53] with variance

$$\sigma_I^2 = \frac{1}{A(1-f)} \left(S + \frac{B_{f,t}}{f} \right), \quad (12)$$

leading to

$$\frac{\hat{S}}{\sigma_I} = S \sqrt{\frac{f(1-f)A}{fS + B_{f,t}}} \quad (13)$$

(assumptions 4–6, 8, 9),

which differs from Eq. (6) only in the factor multiplying S and can be obtained from it by omitting that term and including the source flux in the background. The two are identical if $B_{f,t} \gg S$ or if $f = \frac{1}{2}$.

B. Optimum Choice of f

If $B_{f,t} \gg S$ and if $B_{f,t}$ is independent of S , then consideration of Eq. (7) shows that the optimum open fraction is 50%. However, in general the background may not dominate. $B_{f,t}$ is the combination of a component intrinsic to the detector plus one due to a combination of diffuse sky emission and the smeared effect of sources, other than the one of interest, in the field of view. Thus, we may write $B_{f,t} = B_{\text{det}} + fB_{\text{sky}}$. Putting $b = (B_{\text{sky}}/B_{\text{det}})$ and $s = (S/B_{\text{det}})$ and solving for the optimum value of f one finds

$$f_{\text{opt}}(b, s) = \frac{1}{1 + \sqrt{\frac{1+b}{1+s}}} \quad (14)$$

(assumptions 4–9).

Equation (14) is equivalent to the expression given by in 't Zand *et al.* [51]. Although their result is correct, those authors state that it is the same as that of Fenimore [49], which is in fact different. The latter contains an additional factor of 2, noted by Accorsi *et al.* [28] as an error. It is in fact attributable to an attempt to combine the two different signal-to-noise ratios in Eqs. (6) and (13) above in a single expression, which in retrospect is probably not useful.

We can measure the advantage g of using the optimum open fraction as the signal-to-noise ratio with $f = f_{\text{opt}}$ relative to that for $f = \frac{1}{2}$ given by Eq. (8). After much manipulation it turns out that g depends only on the value of f_{opt} and is independent of the particular combination of s and b which led to that value. It is simply given by

$$g(b, s)^2 = 1 + 4 \left(f_{\text{opt}}(b, s) - \frac{1}{2} \right)^2. \quad (15)$$

Figure 2 provides a convenient nomogram for f_{opt} and g , which also illustrates some conclusions that can be drawn. One sees, for example, how large b (strong background from the sky or from sources other than that of interest) favors low f , moving toward the single pinhole camera extreme. On the other hand, for studying a bright source (large s) high f , more like an open light bucket, are preferable.

As has been noted by other authors, the advantage in signal-to-noise ratio to be obtained by using a value of f other than $\frac{1}{2}$ is small except in the most extreme circumstances. We note, however, that the low values of f marginally favored from this point of view when $S \ll B_{\text{det}} \ll B_{\text{sky}}$ can lead to important data handling and telemetry reductions, particularly when information about each event is recorded.

If the source to be studied dominates over the effects of intrinsic background by more than does the combination of all other sources and the diffuse sky emission ($s > b$), the optimum fraction can be larger than $\frac{1}{2}$. The circumstances in which this is most likely to be relevant is when the objective is to obtain information very quickly, for example, when studying short bursts of emission. We note, however, that this conclusion depends on the definition of signal-to-noise ratio.

If it is the detectability of a source that is important, rather than the precision with which its intensity can

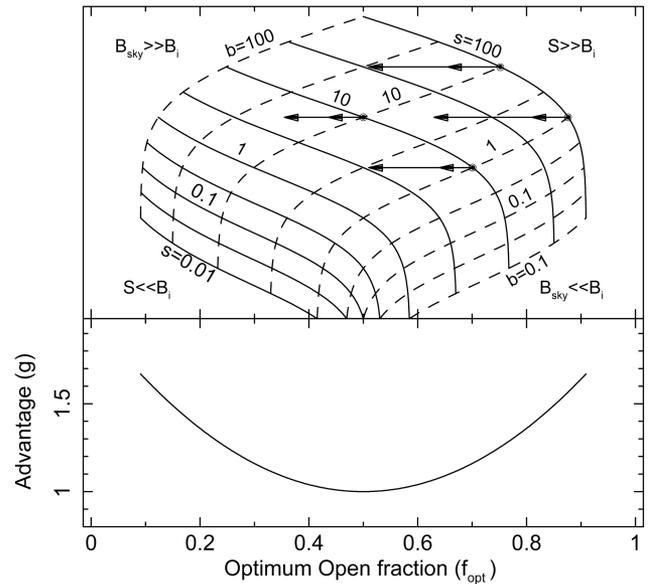


Fig. 2. Top: nomogram for determining the optimum mask open fraction f_{opt} given the parameters b and s , which are, respectively, the sky background and the strength of the observed source, relative to the intrinsic detector background. f_{opt} can be read from the horizontal scale at the intersection of lines of constant s (continuous lines, logarithmically spaced) and of constant b (dashed lines, also logarithmically spaced). The arrows illustrate the effect of a finite detector resolution (see Subsection 6.C). Bottom: signal-to-noise ratio when using $f = f_{\text{opt}}$ relative to that with $f = \frac{1}{2}$.

be measured, then it is \hat{S}/σ_I [Eq. (13)] that should be optimized rather than \hat{S}/σ_S [Eq. (6)]. The flux from the source should then be included in B_{sky} , not S , and Eqs. (14) and (15) and Fig. 2 used with s set to zero (or a small value). Figure 2 shows that f_{opt} is always less than $\frac{1}{2}$ in this case.

4. Imperfect Mask Opacity/Transparency: Relaxing Assumption 4

If the mask elements are not perfectly opaque and transparent but have transmissions t_0 and t_1 , respectively, Eqs. (1) and (2) take the form

$$C_S = fA(t_1S + B_{f,t}), \quad (16)$$

$$C_B = (1-f)A(t_0S + B_{f,t}). \quad (17)$$

Note that $B_{f,t}$ will in this case be a function of t_0 and t_1 , as well as of f . Following the same logic as above one finds

$$\frac{\hat{S}}{\sigma_S} = S(t_1 - t_0) \sqrt{\frac{f(1-f)A}{[(1-f)t_1 + ft_0]S + B_{f,t}}} \quad (18)$$

(assumptions 5–9),

and for source detection

$$\frac{\hat{S}}{\sigma_I} = S(t_1 - t_0) \sqrt{\frac{f(1-f)A}{[(1-f)t_0 + ft_1]S + B_{f,t}}} \quad (19)$$

(assumptions 5, 6, 8, 9).

Thus, the only changes necessary to allow for a uniform absorption in the nominally open areas ($t_1 < 1$) and/or for uniform leakage through the closed ones ($t_0 > 0$) are to multiply the signal-to-noise ratio by a factor of $t_1 - t_0$ and to correct the noise contribution due to source counts if this is not negligible. The reason for the simple form of the multiplying factor will become evident in Section 6.

Accorsi *et al.* [28] have treated the question of the optimum f when the mask is leaky ($t_0 > 0$), but unfortunately the expression they give contains some typographical errors, as well as being rather complex (although their equation for the signal-to-noise ratio is correct, that for f_{opt} twice has $2t$ in place of t). The general case, $t_0 > 0$ and $t_1 < 1$, can, however, be handled by using the equations and nomogram of Subsection 3.B with adjusted parameters

$$s' = t_1s + t_0b, \quad b' = t_0s + t_1b \quad (20)$$

in place of s and b .

5. Gaussian or Poissonian Statistics: Assumption 5

In the above, the only respect in which it has been assumed that Gaussian statistics are applicable is in characterizing the signal-to-noise ratio and the significance of detection in terms of a standard deviation calculated as the root of the sum of the variances [Eq. (4) or its equivalents in other cases]. If both C_S and C_B are small this is a slight simplification as the distribution of \hat{S} will not be strictly Gaussian (or Poissonian). The resulting effects are tiny except where only a very few events in total are involved. A potentially relevant case arises in the detection of a very brief burst—one occurring in so short a time that the background is very small. In this case the significance of detection should ideally be expressed in terms of likelihood. Often in these circumstances the background rate (and even its distribution over the detector) will be well determined by considering data before, and perhaps after, the burst. However, to place the problem in the same context as the above discussion we consider the case where there is no information available outside the time of the event itself. For the same reason we will consider the significance of detection at a given position without considering the degrees of freedom associated with finding the location of the event.

If we observe counts C_S and C_B in the exposed and shadowed parts of the detector [54], the difference in the Cash likelihood statistic [55] between the null (background only) hypothesis and the hypothesis in which a source is present at the supposed position is found to be

$$\begin{aligned} \frac{\delta C}{2} &= C_S \ln\left(\frac{C_S}{f}\right) + C_B \ln\left(\frac{C_B}{1-f}\right) \\ &\quad - (C_S + C_B) \ln(C_S + C_B) \end{aligned} \quad (21)$$

(assumptions 4, 6, 8, 9).

Figure 3 shows, for two examples of f , the combinations of numbers of events that give particular levels of confidence in the detection of a source, calculated according to Eq. (21). For comparison the corresponding contours can be calculated on the Gaussian assumption by evaluating the χ^2 parameter

$$\begin{aligned} \delta\chi^2 &= \frac{(C_S - f(C_S + C_B))^2}{f(C_S + C_B)} + \frac{(C_B - (1-f)(C_S + C_B))^2}{(1-f)(C_S + C_B)} \\ &= \frac{(C_S - f(C_S + C_B))^2}{f(1-f)(C_S + C_B)} \end{aligned} \quad (22)$$

(assumptions 4–6, 8, 9).

In fact, $\delta\chi^2$ in Eq. (22) is just the square of \hat{S}/σ_i from Eq. (13).

Contours of constant χ^2 , calculated according to Eq. (22), are also shown in Fig. 3. As the two statis-

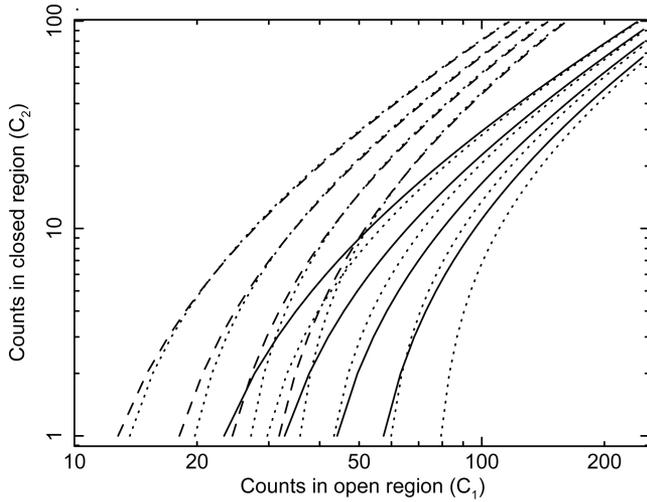


Fig. 3. Numbers of events needed in the region of the detector plane that are exposed to the source (C_S) and that shadowed by the mask (C_B) needed to achieve a given level of detection significance. The dashed and continuous curves are calculated using the Cash statistic [Eq. (21)] which correctly handles Poisson statistics and are for $f = 0.4$ and $f = 0.6$, respectively. Dotted curves are based on $\delta\chi^2$ [Eq. (22)] and so use the approximation of Gaussian statistics. The curves for each family are at levels of 17.3, 26.3, 37.4, and 50.4 (left to right), which for both the Cash statistic and χ^2 correspond to 4σ , 5σ , 6σ , and 7σ , respectively, for 1 degree of freedom (appropriate if the source position is known).

tics both follow the $\chi^2(1)$ distribution, a direct comparison can be made. It can be seen that there is relatively little difference between the two except in extreme cases where the total number of events in the background region number of events is quite low. The Gaussian assumption can still be good even if the number of events per detector element is small (even $\ll 1$), provided the total number of background events, C_B , is more than a dozen or so. For this reason assumption 5 is almost always valid.

6. Finite Detector Resolution: Assumption 6

There remains the assumption that the detector has perfect spatial resolution. Unfortunately, relaxing this assumption has a major impact. The situation is simplest if we again assume background dominated conditions (assumption 3). We consider this case first.

A. Background Dominated Case

The case where the mask shadow is recorded by a detector with limited spatial resolution and where $B_{f,t} \gg S$ is considered by Skinner [8]. It is shown there that in this case the sensitivity relative to that of a reference system with $f = \frac{1}{2}$ [Eq. (8)] is given by the coding power Δ so that

$$\left(\frac{S}{\sigma_S}\right) = \Delta \left(\frac{S}{\sigma_S}\right)_{\text{ref}} = \Delta \frac{S}{2} \sqrt{\frac{A}{B_{f,t}}}, \quad (23)$$

where

$$\frac{\Delta^2}{4} = \frac{1}{n} \sum_i P_i^2 - \left(\frac{1}{n} \sum_i P_i\right)^2 \quad (24)$$

(assumptions 3, 5, 8, 9),

and where P_i is the response in detector element i to a source at the position under consideration, relative to that for a fully exposed element to the same source (so that $0 < P_i < 1$). Assumption 4 is not needed as t_0 and t_1 can be taken into account in calculating the P_i . Equation (24) shows that Δ is simply twice the rms value of P_i . It can be quite different even for two close directions as the shadows of the edges of mask elements may fall differently with respect to the detector elements. If there is a large number of detector elements and their pitch is not commensurate with that of the mask elements then all relative phases of the two arrays will occur with about the same frequency and any such variations will be small. If there is a simple ratio between the pitches, then an average value of Δ may be used as a measure of the mean sensitivity, averaged over different sky directions (different shifts of the mask shadow), i.e., we invoke assumption 8.

The concept of coding power is a very useful one. It can be used to derive Eq. (7), which is the limiting form of Eq. (6) when $B_{f,t} \gg S$, or the corresponding form of Eq. (18) or Eq. (19) in the same limit. But it also allows quantitative treatment of the loss in sensitivity due to finite detector resolution in any background limited case. A detector with limited spatial resolution records only a blurred version of the shadow of the mask, as illustrated in Fig. 1 and the rms deviation of P is consequently reduced. This approach was used in [8] to obtain an expression for the sensitivity of a telescope having square elements of side m and a 50% open fraction when the detector has square pixels of finite size d . Generalizing the result obtained there to allow for masks with any open fraction and with imperfect transmission and opacity, one finds that the sensitivity must be multiplied by a coding power factor

$$\begin{aligned} \Delta &= \left(1 - \frac{d}{3m}\right) (t_1 - t_0) \sqrt{4f(1-f)} \quad \text{if } m \geq d \\ &= \frac{m}{d} \left(1 - \frac{m}{3d}\right) (t_1 - t_0) \sqrt{4f(1-f)} \quad \text{if } m \leq d \end{aligned} \quad (25)$$

(assumptions 3, 5, 8, 9).

Analytic solutions in even more general cases are messy and not very revealing, but numerical calculation of Δ allows the sensitivity of a proposed or actual system to be estimated. As an example, we take the case shown in Fig. 1, which is relevant to the EXIST project. The shadow of a mask with square elements of side m supported by a grid structure with bar width g is imagined to be recorded using a detector having

square pixels of side d . The variation of P along one particular line is illustrated in the case $g < d < m$.

The calculation can be simplified by noting that the mask pattern can be described as the convolution of a single mask hole, side m , with a sparse bed-of-nails function (or two-dimensional Shah function [56]), pitch $p = m + g$, in which only a fraction f_e of the spikes are present. The response function is then obtained by a further convolution with the form of a detector element. Use of the convolution theorem allows the Fourier transform of the response to be obtained and Parseval's theorem then gives its mean square value.

Example results are shown in Fig. 4. When the supporting grid is present there is an important loss in sensitivity unless the detector pixels are very small. In effect, the loss is due to an increased fraction of intermediate gray levels because the shadow of the fine grid is poorly resolved.

In real conditions the shadow of the mask cast by an off-axis source may not simply be a translation of that for an on-axis source. The finite thickness of the mask elements and/or that of a supporting grid, or the partial transparency of the structure, may modify the off-axis response. Grindlay and Hong [45] have discussed approaches to some of the problems associated with such complications.

B. Finite Detector Resolution: Background Not Dominant

The approach used in [8] and Subsection 6.A for deriving the expression for the sensitivity considers the problem as equivalent to finding the gradient of the best-fit straight line in a data space relating the observed counts in a detector pixel, C_i , to the corresponding P_i . In the case treated there $B_{f,t} \gg S$ so

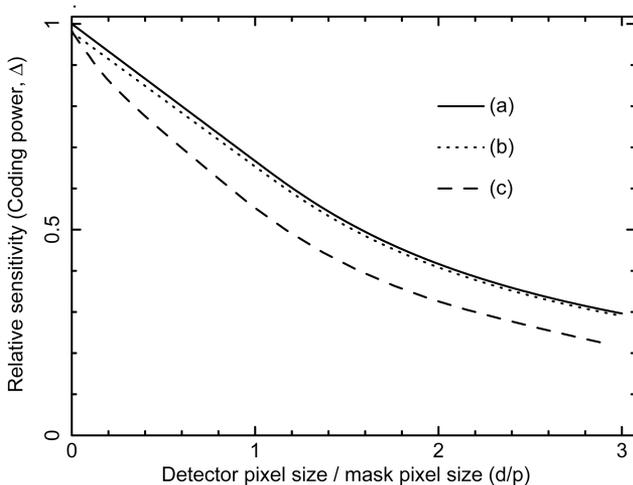


Fig. 4. Loss in sensitivity if the detector spatial resolution is not perfect. The solid curve (a) is for a simple mask with square elements having open fraction $f = 0.5$ and a detector with square pixels [Eq. (24)]. The dotted curve (b) shows the corresponding curve with $f = 0.4$. The dashed curve (c) is for a mask in which 50% of the elements are open but in which a supporting grid like that in Fig. 1 reduces the transparency to $f = 0.4$. If the detector resolution is good (low d/p) the grid provides coding and so the curve approaches (b). If the detector resolution is poor, the grid simply attenuates the flux and reduces the sensitivity (assumptions 3–5, 8, and 9).

the errors on each point are the same. In the general case the number of counts expected in detector pixel i of surface area a is [57]

$$C_i = a(B_{f,t} + P_i S) \pm \sqrt{C_i}, \quad (26)$$

and the best estimate of S can be found as the gradient of the straight line that is a weighted best fit to the points (C_i, P_i) . The uncertainty in the gradient is given by

$$\sigma_S^2 = \frac{2}{\Delta_w} \left(\sum \frac{1}{C_i} \right)^{-1}, \quad (27)$$

where Δ_w is simply the weighted equivalent of Δ :

$$\frac{\Delta_w^2}{4} = \left(\sum \frac{P_i^2}{C_i} \right) / \left(\sum \frac{1}{C_i} \right) - \left[\left(\sum \frac{P_i}{C_i} \right) / \left(\sum \frac{1}{C_i} \right) \right]^2$$

(assumptions 5, 7–9).

C. Optimum Open Fraction with Finite Detector Resolution

In 't Zand *et al.* [51] have noted how imperfect detector spatial resolution tends to decrease the optimum open fraction. For the case they discussed, that of the Beppo-SAX wide-field cameras, they concluded that f in the range of 0.25–0.33 was the best choice. However, this conclusion was more due to the fact that the fields simulated contain several sources whose flux led to a high b than to the effects of the detector resolution.

The shift of f_{opt} to lower values due to imperfect detector resolution can nevertheless be important when studying a single strong source. The arrows in Fig. 2 illustrate the effect with $m/d = 2$ (shorter arrows) and with $d = m$ (a case discussed in Section 7).

D. Poisson Statistics and Finite Detector Resolution

Finally, if the detector resolution is finite and the number of events is so small that Poisson statistics must be used, the source flux can be obtained by optimizing the Cash likelihood statistic

$$\frac{\mathcal{C}}{2} = - \sum C_i \ln(P_i S + B_{f,t}), \quad (28)$$

and the confidence limits obtained by finding the S for which \mathcal{C} changes by the required amount (with $B_{f,t}$ refitted). For calculating the confidence with which the null (background only) hypothesis can be rejected one can calculate

$$\frac{\delta \mathcal{C}}{2} = - \sum C_i \ln(P_i S + B_{f,t}) + 2n C_i \ln(C_i) \quad (29)$$

(only assumptions 7–9 necessary),

n being the number of detector pixels. Thus, although no useful explicit formulas are available in this case, use of this statistic provides a method for dealing with any particular example.

7. Optimizing Source Position Determination Accuracy

For a given mask to detector separation, dictated perhaps by spacecraft accommodation considerations and/or a minimum required field of view, the angular resolution of a coded mask telescope depends on the mask pixel size and also on that of the detector. Practical issues usually limit how small the pixels of a detector may be made, while the mask design is usually less subject to constraints. If we suppose the detector pixel size to be fixed, the angular resolution will formally continue to improve as the mask pixels are made smaller and smaller but, as was seen above, with low m/d the significance with which sources are detected will suffer.

Simulations confirm that a good approximation to the angular resolution is obtained by taking the Pythagorean sum of the angle subtended by a mask pixel at the detector and that subtended by a detector pixel at the mask. Near the center of the field of view

$$\delta\theta^2 = (m/l)^2 + (d/l)^2, \quad (30)$$

where l is the mask–detector separation. Note that in a case such as that illustrated in Fig. 1 it is the size of the holes (m) that is important, not their pitch.

The accuracy with which a source can be located is better than this by a factor approximately proportional to the signal-to-noise ratio S/σ_S of the source [58] so

$$\begin{aligned} \delta\alpha &= \left(\frac{\sigma_S}{S}\right) k [(m/l)^2 + (d/l)^2]^{\frac{1}{2}} \\ &= \Delta \left(\frac{\sigma_S}{S}\right)_{d=0} k [(m/l)^2 + (d/l)^2]^{\frac{1}{2}}, \end{aligned} \quad (31)$$

where k is a constant of the order of unity that depends on the exact definition of location accuracy. Substituting Δ from Eq. (25) one finds that, on the assumptions under which these equations apply (3, 5, 8, and 9), the lowest position uncertainty is obtained with $m = d$. This is illustrated in Fig. 5.

In the design of an instrument the number of objects detectable is likely to also be a consideration. As the minimum is relatively shallow, choosing a slightly higher m/d will allow additional faint sources to be detected at the expense of only a small loss in positioning accuracy for brighter ones. The BAT instrument on Swift uses 5/4; a value of 2/1 is baselined for EXIST.

8. Conclusions

The formulas presented above offer insight into the way in which the inevitable uncertainties due to Poisson statistics affect measurements with coded mask telescopes. In some cases the differences between a simplified treatment and the more precise

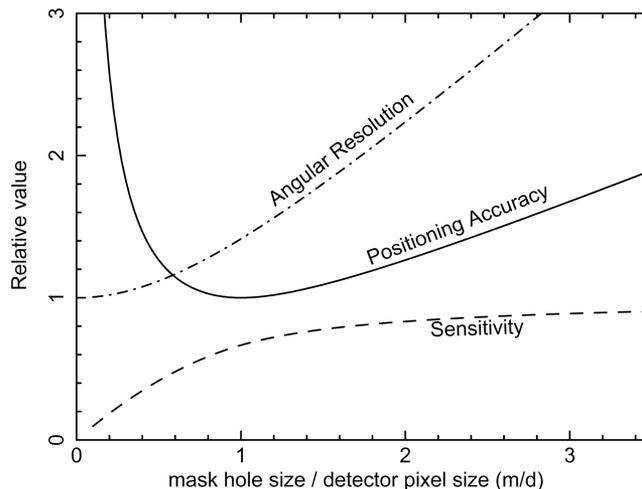


Fig. 5. Variation of relative values of signal-to-noise ratio, angular resolution, and source location accuracy with mask element size for a fixed detector element size. Assumptions 3, 5, 8, and 9 are made; in particular, noise is assumed to be background dominated ($B_{f,t} \gg S$).

one can be quite large. For example, with the choice of $m/d = 1$, shown in Section 7 to give the best source location accuracy, the sensitivity is worse by a factor of 2/3 than that which would be expected by blind application of a simplified formula. Often in astronomy the number of objects observed depends on the $-3/2$ power of the detection threshold, so use of the simplified approach would lead to an overestimate of that number by a factor of 1.8.

Although the discussion here has been in terms of measuring the flux from a particular direction, for example, that from a point source, in many cases the results can be applied to extended sources by considering the flux per angular resolution element from the source.

It should be noted that systematic errors due to (uncorrected) variations in the background level across the detector plane have not been considered nor have been those due to ghosts (sidelobes) of other sources. Thus, the results are most directly applicable in the case of short observations of relatively bright sources (e.g., gamma-ray bursts) or where the design or observation strategy is such as to minimize such errors (e.g., through use of scanning, combined with a very large number of detector pixels, as in the proposed EXIST black hole finder mission).

The author thanks Roberto Accorsi, David Band, Jean in 't Zand, Ed Fenimore, and Craig Markwardt for helpful discussions.

References

1. S. D. Barthelmy, L. M. Barbier, J. R. Cummings, E. E. Fenimore, N. Gehrels, D. Hullinger, H. A. Krimm, C. B. Markwardt, D. M. Palmer, A. Parsons, G. Sato, M. Suzuki, T. Takahashi, M. Tashiro, and J. Tueller, "The Burst Alert Telescope (BAT) on the SWIFT Midex mission," *Space Sci. Rev.* **120**, 143–164 (2005).
2. C. Winkler, T. J.-L. Courvoisier, G. Di Cocco, N. Gehrels, A. Giménez, S. Grebenev, W. Hermsen, J. M. Mas-Hesse,

- F. Lebrun, N. Lund, G. G. C. Palumbo, J. Paul, J.-P. Roques, H. Schnopper, V. Schönfelder, R. Sunyaev, B. Teegarden, P. Ubertini, G. Vedrenne, and A. J. Dean, "The INTEGRAL mission," *Astron. Astrophys.* **411**, L1–L6 (2003).
3. A. M. Levine, H. Bradt, W. Cui, J. G. Jernigan, E. H. Morgan, R. Remillard, R. E. Shirey, and D. A. Smith, "First results from the all-sky monitor on the Rossi x-ray timing explorer," *Astrophys. J.* **469**, L33–L36 (1996).
 4. F. Tokanai, M. Matsuoka, N. Kawai, A. Yoshida, M. Yamauchi, K. Takagishi, I. Hatsukade, E. E. Fenimore, M. Galassi, and J. J. in 't Zand, I. A. Bond, K. Morimoto, and H. Nakamura, "Performance of wide-field x-ray monitor on board HETE (High-Energy Transient Experiment)," *Proc. SPIE* **2808**, 563–570 (1996).
 5. R. Jager, W. A. Mels, A. C. Brinkman, M. Y. Galama, H. Goulooze, J. Heise, P. Lowes, J. M. Muller, A. Naber, A. Rook, R. Schuurhof, J. J. Schuurmans, and G. Wiersma, "The wide field cameras on-board the BeppoSAX x-ray astronomy satellite," *Astron. Astrophys. Suppl. Ser.* **125**, 557–572 (1997).
 6. R. Vanderspek, J. Villaseñor, J. Doty, J. G. Jernigan, A. Levine, G. Monnelly, and G. R. Ricker, "GRB observations with the HETE soft x-ray cameras," *Astron. Astrophys.* **138**, 565–566 (1999).
 7. E. Caroli, J. B. Stephen, G. di Cocco, L. Natalucci, and A. Spizzichino, "Coded aperture imaging in x- and gamma-ray astronomy," *Space Sci. Rev.* **45**, 349–403 (1987).
 8. G. K. Skinner, "Coding (and decoding) coded mask telescopes," *Exp. Astron.* **6**, 1–7 (1995).
 9. J. G. Ables, "Fourier transform photography: a new method for x-ray astronomy," *Proc. Astron. Soc. Australia* **1**, 172–173 (1968).
 10. R. H. Dicke, "Scatter-hole cameras for x-rays and gamma rays," *Astrophys. J.* **153**, L101 (1968).
 11. L. Mertz, *Transformations in Optics* (Wiley, 1965).
 12. L. Mertz, "Ancestry of indirect techniques for x-ray imaging," *Proc. SPIE* **1159**, 14–17 (1989).
 13. J. Gunson and B. Polychronopoulos, "Optimum design of a coded mask x-ray telescope for rocket applications," *Mon. Not. R. Astron. Soc.* **177**, 485–497 (1976).
 14. E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Appl. Opt.* **17**, 337–347 (1978).
 15. R. J. Proctor, G. K. Skinner, and A. P. Willmore, "The design of optimum coded mask x-ray telescopes," *Mon. Not. R. Astron. Soc.* **187**, 633–643 (1979).
 16. A. B. Giles, "Self-supporting perfect masks for 2-D infrared and x-ray imaging," *Appl. Opt.* **20**, 3068–3072 (1981).
 17. M. Oda, "X-ray imaging techniques—modulation collimator and coded mask," *Adv. Space Res.* **2**, 207–216 (1982).
 18. W. J. Wild, "Dilute uniformly redundant sequences for use in coded-aperture imaging," *Opt. Lett.* **8**, 247–249 (1983).
 19. A. R. Gourlay and N. G. S. Young, "Coded aperture imaging: a class of flexible mask designs," *Appl. Opt.* **23**, 4111–4117 (1984).
 20. M. H. Finger and T. A. Prince, "Hexagonal uniformly redundant arrays for coded-aperture imaging," in *Proceedings of the International Cosmic Ray Conference*, F. C. Jones, ed. (NASA, 1985), pp. 295–298.
 21. S. R. Gottesman and E. J. Schneid, "PNP—A new class of coded aperture arrays," *IEEE Trans. Nucl. Sci.* **33**, 745–749 (1986).
 22. K. Byard, "Square element antisymmetric coded apertures," *Exp. Astron.* **2**, 227–232 (1992).
 23. K. Byard, "On self-supporting coded aperture arrays," *Nucl. Instrum. Methods Phys. Res. A* **322**, 97–100 (1992).
 24. L. E. Kopilovich and L. G. Sodin, "Synthesis of coded masks for gamma-ray and x-ray telescopes," *Mon. Not. R. Astron. Soc.* **266**, 357–359 (1994).
 25. L. G. Sodin, "Synthesis of coded masks for x-ray and gamma-ray telescopes," *Astron. Lett.* **21**, 423–427 (1995).
 26. H. D. Lüke and A. Busboom, "Binary arrays with perfect odd-periodic autocorrelation," *Appl. Opt.* **36**, 6612–6619 (1997).
 27. A. Busboom, H. D. Schotten, and H. Elders-Boll, "Coded aperture imaging with multiple measurements," *J. Opt. Soc. Am. A* **14**, 1058–1065 (1997).
 28. R. Accorsi, F. Gasparini, and R. C. Lanza, "Optimal coded aperture patterns for improved SNR in nuclear medicine imaging," *Nucl. Instrum. Methods Phys. Res. A* **474**, 273–284 (2001).
 29. S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Appl. Opt.* **28**, 4344–4352 (1989).
 30. H. D. Lüke and A. Busboom, "Mismatched filtering of periodic and odd-periodic binary arrays," *Appl. Opt.* **37**, 856–864 (1998).
 31. M. R. Sims, M. J. L. Turner, and R. Willingale, "The influence of disturbing effects on the performance of a wide field coded mass x-ray camera," *Nucl. Instrum. Methods Phys. Res. A* **228**, 512–531 (1985).
 32. D. Vigneau and D. W. Robinson, "Large coded aperture mask for spaceflight hard x-ray images," *Proc. SPIE* **4851**, 1326–1335 (2003).
 33. A. P. Hammersley and G. K. Skinner, "Data processing of imperfectly coded images," *Nucl. Instrum. Methods Phys. Res. A* **221**, 45–48 (1984).
 34. R. Willingale, "Use of the maximum entropy method in x-ray astronomy," *Mon. Not. R. Astron. Soc.* **194**, 359–364 (1981).
 35. R. Willingale, M. R. Sims, and M. J. L. Turner, "Advanced deconvolution techniques for coded aperture imaging," *Nucl. Instrum. Methods Phys. Res. A* **221**, 60–66 (1984).
 36. M. L. McConnell, E. L. Chupp, P. P. Dunphy, D. J. Forrest, and A. Owens, "The problem of nonuniform background rates in a coded aperture gamma-ray telescope," in *Proceedings of the International Cosmic Ray Conference* (Nauka, 1987), p. 309–311.
 37. A. P. Willmore, D. Bertram, M. P. Watt, G. K. Skinner, T. J. Ponman, M. J. Church, J. R. H. Herring, and C. J. Eyles, "Image correction in a coded mask x-ray telescope," *Mon. Not. R. Astron. Soc.* **258**, 621–628 (1992).
 38. J. E. Grindlay, D. Barret, K. S. K. Lum, R. P. Manandhar, B. Robbason, and S. Vance, "New wavelet methods for flatfielding coded aperture images," in *Imaging in High Energy Astronomy*, L. Bassani and G. DiCocco, eds. (Kluwer Academic, 1995), p. 213–220.
 39. R. M. Rideout and G. K. Skinner, "Minimum error image reconstruction for coded mask telescopes," *Astron. Astrophys.* **120**, 579–585 (1996).
 40. L. Bouchet, J. P. Roques, J. Ballet, A. Goldwurm, and J. Paul, "The SIGMA/Granat telescope: calibration and data reduction," *Astrophys. J.* **548**, 990–1009 (2001).
 41. G. Skinner and P. Connell, "The Spiros imaging software for the INTEGRAL SPI spectrometer," *Astron. Astrophys.* **411**, L123–L126 (2003).
 42. A. W. Strong, "Maximum entropy imaging with INTEGRAL/SPI data," *Astron. Astrophys.* **411**, L127–L129 (2003).
 43. B. M. Schäfer and N. Kawai, "A Fourier-based algorithm for modelling aberrations in HETE-2's imaging system," *Nucl. Instrum. Methods Phys. Res. A* **500**, 263–271 (2003).
 44. C. B. Markwardt, J. Tueller, G. K. Skinner, N. Gehrels, S. D. Barthelmy, and R. F. Mushotzky, "The SWIFT/BAT high-latitude survey: first results," *Astrophys. J.* **633**, L77–L80 (2005).
 45. J. E. Grindlay and J. Hong, "Optimizing wide-field coded aperture imaging: radial mask holes and scanning," *Proc. SPIE* **5168**, 402–410 (2004).
 46. D. L. Band, J. E. Grindlay, J. Hong, G. Fishman, D. H. Hartmann, A. Garson III, H. Krawczynski, S. Barthelmy, N. Gehrels, and

- G. Skinner, "EXIST's gamma-ray burst sensitivity," *Astrophys. J.* **673**, 1225–1232 (2008).
47. P. L. Jensen, K. Clausen, C. Cassi, F. Ravera, G. Janin, C. Winkler, and R. Much, "The INTEGRAL spacecraft—in-orbit performance," *Astron. Astrophys.* **411**, L7–L17 (2003).
 48. T. M. Palmieri, "An x-ray telescope sensitive at high energies," *Ap. Space Sci.* **26**, 431–445 (1974).
 49. E. E. Fenimore, "Coded aperture imaging—predicted performance of uniformly redundant arrays," *Appl. Opt.* **17**, 3562–3570 (1978).
 50. E. Caroli, R. C. Butler, G. di Cocco, P. P. Maggioli, L. Natalucci, and A. Spizzichino, "Coded masks in x- and gamma-ray astronomy—The problem of the signal-to-noise ratio evaluation," *Nuovo Cimento C* **7**, 786–804 (1984).
 51. J. J. M. in 't Zand, J. Heise, and R. Jager, "The optimum open fraction of coded apertures: with an application to the wide field x-ray cameras of SAX," *Astron. Astrophys.* **288**, 665–674 (1994).
 52. This illustrates an approximation in the approach used here. The shadow of a source at trial positions away from the peak cannot be totally independent of that expected for a source at the peak position [53]. The process described here is equivalent to correlation with a mean-subtracted form of the expected count distribution, which must produce a function with a mean of zero. Thus, a positive peak implies a negative mean level elsewhere. The effect goes inversely with the number of resolution elements in the field of view and becomes negligible for sufficiently large N .
 53. G. K. Skinner and T. J. Ponman, "On the properties of images from coded mask telescopes," *Mon. Not. R. Astron. Soc.* **267**, 518 (1994).
 54. The same symbols are used indiscriminately here for the observed and expected numbers of events as the one is the best estimate of the other.
 55. W. Cash, "Parameter estimation in astronomy through application of the likelihood ratio," *Astrophys. J.* **228**, 939–947 (1979).
 56. R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed. (McGraw-Hill, 1986)
 57. Note that the C_i are counts per pixel where C_S and C_B were totals for all the pixels of a particular category.
 58. Sometimes $S/(\sigma_S - 1)$ is used for a better approximation but the differences are small in practical cases.