The All Sky Monitor for the X-ray Timing Explorer

John P. Doty
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

The great complex behavior of compact galactic X-ray sources complicates their study. Since these objects are highly variable, and exhibit a multiplicity of "states", it is often impossible to anticipate when a desired observation can be performed. "Transient" objects appear without warning and may be detectable for only a short period of time. These difficulties can be reduced by an instrument which can monitor a large number of objects simultaneously, or nearly so.

The function of an "All Sky Monitor" is to record a photometric history of known objects, and to detect and locate transient objects. The classic techniques of this field are the pinhole camera and the scanned slat collimator. For both of these approaches, there is an unfortunate tradeoff between spatial resolution and "photon starvation": a high resolution instrument will capture very few photons. The multiple pinhole approach avoids this problem, but designs based on the Uniformly Redundant Array (URA) principle tend to be difficult to build and require a large telemetry bandwidth.

The All Sky Monitor (ASM) for the X-ray Timing Explorer (XTE) avoids these problems. It uses the multiple pinhole approach to avoid photon starvation for good photometric capability together with good resolution and an excellent capability for locating transients. While the XTE ASM doesn't have the ideal imaging qualities of URA based designs, it is simpler to build and requires a modest telemetry bandwidth; since the ASM is intended for detection, location, and photometry of discrete, point sources, imaging is of secondary importance.

History and Operating Principals

The general history of coded aperture cameras is covered well in a recent review paper. This section will describe the process by which the XTE ASM was invented and configured.

In 1976, George Ricker circulated a preprint, *A Diffraction-Limited X-ray Shadow Camera.* This paper was unique in its emphasis on the photometric and astrometric properties of the coded aperture

Consider an array of detector elements behind an aperture which contains a mask. Holes cover half of the area of the mask, while the other half blocks x-rays. A point source in the field of view will illuminate half of the detector elements. The other half will see only backgound. The intensity of the source may thus be measured by subtracting the intensity observed in the "background" elements from those in the "source" elements.

Of course, which elements are "source" and "background" depends on where the source is in the field of view. If an interfering source is present, some of its flux will appear in the "source" elements for the desired source, and some will appear in the "background " elements for the desired source. If the mask holes are randomly located, then half of the flux received from the interfering source is expected to go into "source" elements for the desired source while the other half is expected to go to "background" elements. Thus, the interfering source is equivalent to an increased background, as far as the measurement of the desired source is concerned.

The two major subtleties ignored in the above analysis are:

(1) Since the detector cannot perfectly determine where a photon strikes, detector elements can't be cleanly classified as "source" and "background". Each element exhibits a different effective area to the source.

(2) The fact that for a random mask an interfering source is *expected* to split its flux between "source" and "background" elements doesn't mean that an interfering source at a specific location will do so exactly.

These effects are handled by the least squares technique described in the "Data Analysis" section.

approach, rather than on its imaging properties. He advocated the use of one-dimensional systems, and even anticipated Fenimore and Cannon's URA concept by advocating ''pseudorandom shift register sequences'' as mask patterns (although he did not pursue this idea in detail). This paper was never published, but it had a seminal effect at MIT.

The next two years saw important developments in coded aperture imaging from the Birmingham and Los Alamos groups. Although the the design of the aperture mask was the stated subject of both these papers, the wider issues of overall imaging camera architecture were discussed. These papers made it clear to me that unambiguous imaging with these systems required not only a good mask design, but an architecture which limits the field of view, so that the number of unknowns in the image is smaller than or equal to the number of measurements (see box).

But why shouldn't the deconvolution be straightforward without special tricks? After all, deconvolution using Fourier transforms is straightforward and well behaved as long as the convolving function has no zeros in its power spectrum. The reason that the coded aperture system's deconvolution is

tricky comes from the fact that while Fourier deconvolution works on cyclic functions, the image captured by the coded aperture camera is not cyclic.

But it could be! I had the idea that one might scan a one-dimensional shadow camera in its imaging direction. The shadows on the detector would move, of course, until you had scanned 360°, at which point the original shadow pattern would be restored. Thus, the sequence of shadowgrams would form a cycle, whose pattern would be the intensity of the sky along the scan *cyclically* convolved with the mask pattern. This is the "scanning shadow camera" (SSC) architecture.

For those interested in two dimensional imaging, I note that this idea does not work in two dimensions. The reason is topological: the 2D Fourier transform has toroidal topology, unlike the spherical topology of the sky.

These ideas were still in my mind when, in 1980, we began work on the MIT XTE proposal. It was clear to us that some sort of ‘‘all sky monitor’’ was very important to the XTE mission. We had used slat collimators as sky monitors on SAS-3, and so I initially looked at the possibility of using them on XTE. The difficulty with slats is that it is hard to make an efficient collimator with a width of <1°, yet 1° is not enough resolution to dissect the dense concentration of sources in the galactic bulge. I therefore thought of using the high resolution of the one-dimensional shadow camera to help with this region. My original idea was to combine a single SSC with a pair of collimated detectors.

I discussed this idea with Garrett Jernigan, and, after he had thought about it, he came back to me with a a different idea: eliminate the slats and use two slightly tilted one-dimensional shadow cameras. He had calculated that the shadow camera was not only better at resolving sources, but better at measuring their intensities as well. The reason is that a narrow scanning slat collimated detector is ‘‘photon starved’’: its signal to noise ratio is, for many objects of interest, dominated by the statistics of the source photons themselves. The shadow camera captures more photons, and therefore exhibits a better signal to noise ratio.

Data Reduction

In this section, the method of least squares will be used to find data reduction algorithms with maximal signal to noise ratio. Least squares methods have been extensively used in X-ray astronomy.

The above figure illustrates the convolutional model of the coded aperture system. While it is, in some sense, an accurate representation of what's happening geometrically, it is a poor model in several respects:

(1) As the figure shows, the detector cannot capture the entire convolution of the field of view. Enlarging the detector doesn't help, since the field of view is also enlarged. If the functions above are replaced by discrete samples (which is what one does in practice), it may be seen that the number of elements in the field of view, N_f, is given by:

$$N_f = N_m + N_d - 1$$

where N_m and N_d are the number of mask and detector elements, respectively. Thus, the number of measurements, N_d, is always less than the number of unknowns, N_f, and the unknowns can not be computed unambiguously.

Some coded aperture designs have attempted to avoid this difficulty by using collimators to limit the field of view. This has disadvantages. Collimation reduces the effective area, especially off axis. It also introduces additional shadows which are not modeled by the convolutional model, but which must be taken into account if photometric precision is required.

(2) The detector is assumed to be an ideal flat plane. In reality, the detector exhibits many non-convolutional effects. Sensitivity will vary across its surface due to window support structure. X-rays pentetrate a random distance before interacting: this produces an angle and energy dependent smearing of the shadow pattern. The XTE ASM will use a position sensitive proportional counter similar to that used by the HEAO-B3 experiment: it will use charge division by a resistive anode to locate x-ray events. The location is computed by taking the ratio of the pulse height at one end of the wire to the sum of the pulse heights from both ends: this procedure produces an error distribution which is skewed toward the center of the detector, another non-convolutional effect.

Although the convolutional model may be adequate for coded aperture imaging, a more sophisticated detector model is required for coded aperture photometry.

Justification for the use of these methods, and their limitations when applied to Poisson distributed data have been discussed in the literature (refs).

The coded aperture camera is a linear system whose response may be described by the matrix equation:

$$\langle \mathbf{h} \rangle = \mathbf{A}\mathbf{x}$$

The raw measurements are a histogram of events with expectation values given by the column vector $\langle \mathbf{h} \rangle$. The column vector $\mathbf{x}$ represents the (unknown) source fluxes and background flux, while the matrix $\mathbf{A}$ represents the behavior of the camera. Each column of A represents the response of the camera to a unit flux from the corresponding source. Given an actual measured histogram, $\mathbf{h}$, we may estimate the unknowns by solving the equation:

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{h}$$

The vector $\hat{\mathbf{x}}$ is an estimate of $\mathbf{x}$. If σ^2 represents the variance of each element of $\mathbf{h}$, then the matrix $\sigma^2(\mathbf{A}^T\mathbf{A})^{-1}$ (the "covariance" or "error" matrix) represents the covariances of the elements of $\hat{\mathbf{x}}$. In particular, the diagonal elements of the error matrix are the variances of the elements of $\hat{\mathbf{x}}$ (Refs.). In practice, since we must normally estimate the variance of actual measured data, the error matrix must be considered an estimate of the true covariances.

The error matrix technique is especially useful when comparing the performance of different camera designs, since it allows the designer to avoid extensive Monte-Carlo testing. In this case, variances may be accurately calculated rather than estimated.

If different observations have different variances, the a weighted least squares method is appropriate:

$$\mathbf{A}^T\mathbf{W}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{W}\mathbf{h}$$

$\mathbf{W}$ is a diagonal matrix of "weights"; each diagonal element of $\mathbf{W}$ should be the inverse of the variance of the corresponding observation. Since for Poisson statistics, the variance and expectation value are the same:

$$W_{ii} = \langle h_i \rangle^{-1}$$

Unfortunately, this equation cannot be applied directly to actual data since $\langle\mathbf{h}\rangle$ isn't something that can be measured.

Thus, a way to estimate the appropriate weights is needed. Possible methods include:

(1) Set $W_{ii} = h_i^{-1}$. This works well if each h_i is large. However, it produces biased results if any of the h_i is too small, and doesn't work at all if any of the h_i is zero.

(2) Use the unweighted method to estimate $\hat{\mathbf{x}}$. Then set $\tilde{\mathbf{h}} = \mathbf{A}\hat{\mathbf{x}}$; $\tilde{\mathbf{h}}$ is an estimate for $\langle\mathbf{h}\rangle$, and thus an estimate for $\mathbf{W}$. Use the weighted method to obtain a new estimate for $\hat{\mathbf{x}}$. Iterate until convergence is achived. This is the standard method in X-ray astronomy. It behaves well and converges very rapidly as long as the total number of counts in $\langle\mathbf{h}\rangle$ isn't too small.

(3) Use an independent data set to estimate $\mathbf{W}$, or don't weight at all. This method is unbiased, but may not yield the best signal to noise ratio. It may be useful, however, when an estimate is

required that is "in the noise": an estimate which, while insignificant by itself, will be combined later with estimates based on other measurements.

The success of the above methods depends on the observed fact that in most practical cases, the results are only weakly dependent on the weights. Methods (1) and (2) misbehave for small numbers of counts, when **W** becomes extremely dependent on the data. Note that if the number of counts is large, methods (1) and (2) are nearly linear in the data, and method (3) is exactly linear.

Criteria for determining confidence intervals for estimates based on method (2) have been described in the literature. These criteria are applicable in both nonlinear and linear cases. In the case of a linear or nearly linear method, it is often more convenient to use the error matrix to estimate the signal to noise ratio. If the weight matrix **W** accurately represents the inverses of the variances of **h**, then the matrix $(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}$ is the error matrix for the weighted method.

The solution to the weighted lest squares method may be written:

$$\hat{\mathbf{x}} = \mathbf{U}\mathbf{h}$$

where:

$$\mathbf{U} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}$$

This is closely related to Fenimore's "URA-tagging" method for coded aperture photometry. Individual elements of **x** may be computed by computing the dot product of the corresponding row of **U** with the measurements **h**. In Fenimore's terminology, the rows of *U* are "tag arrays". Thus, an individual source intensity may be efficiently computed if its tag array is known beforehand. This is especially true if **h** is sparse, which in this case means that it contains many fewer photons than elements. Fenimore suggests forming the dot product photon by photon in such cases: this is essentially the well known technique for multiplying a "randomly packed" sparse matrix by another matrix.

The least squares method described above provides a straightforward way to calculate tag arrays which yield optimal signal to noise. The method also demonstates that "URA tagging" is not restricted to URA based cameras: the special properties of the URA were not invoked at any point in the derivation.

An important special case is the case where there is just one source and no weighting is needed. In this case, the matrix **A** is reduced to a vector, **a**, and the vector **x** is reduced to a scalar, x. The least squares estimate of x is just the correlation of **a** with the data, **h**, normalized by dividing by the correlation of **a** with itself:

$$\hat{x} = \frac{\mathbf{a}\cdot\mathbf{h}}{\mathbf{a}\cdot\mathbf{a}}$$

This result demonstrates the "matched filter" principle: *the optimum way to detect a signal in noisy data is to correlate the data with the expected signal.* This principle is applicable whenever the noise is of constant amplitude ($\sigma^2(h_i) = \sigma^2(h_j)$), and the noise in separate measurements is uncorrelated.

The error "matrix" is just the scalar $\frac{\sigma^2}{\mathbf{a}\cdot\mathbf{a}}$. Since this is the variance of $\hat{x}$, its square root represents the noise in the measurement. The signal to noise ratio is thus:

$$\frac{x}{\sigma}\sqrt{\mathbf{a}\cdot\mathbf{a}}$$

The correlation of **a** with itself, **a**·**a**, represents the signal power which results from a source of unit flux. Keeping track of the signal power will prove to be an effective way to understand the sensitivity of coded aperture cameras.

Sensitivity and the Single Source

For a coded aperture camera illuminated by a single source and background, the expected number of counts in each detector element is given by:

$$\langle h_i \rangle = \frac{AT}{N}(Is_i + r\,\Omega B_d + B_p)$$

where:

I is the source intensity.
r is the rato of pinhole area to detector area.
N is the number of elements in the position senstive detector.
A is the detector area.
T is the integration time.
B_d is the diffuse background rate.
B_p is the intrinsic detector background rate.
Ω is the solid angle of the detector, as seen from the aperture.

The vector **s** could be called the "shadow" vector: in the ideal case s_i is 1 where the source illuminates

the detector through the aperture, and 0 elsewhere.

In the case of a dominant, known background, the matched filter method can be used to estimate I from the data. The total background per detector element is:

$$B = \frac{AT}{N}(r\Omega B_d + B_p)$$

To implement the matched filter, the background must be subtracted from **h**, yielding **h′**:

$$h_i' = h_i - B$$

The matched filter vector is then given by:

$$\mathbf{a} = \frac{AT}{N}\mathbf{s}$$

and the estimated intensity is:

$$\hat{I} = \frac{N}{AT}\frac{\mathbf{s}\cdot\mathbf{h}'}{\mathbf{s}\cdot\mathbf{s}} = \frac{1}{ATS}\mathbf{s}\cdot\mathbf{h}'$$

where:

$$S = \frac{\mathbf{s}\cdot\mathbf{s}}{N}$$

S is the "power" in the shadow pattern (in a mathematical rather than a physical sense).

Assuming that the background dominates the noise, the signal to noise ratio may be approximated by:

$$\Psi = I\sqrt{\frac{\overline{\mathbf{a}\cdot\mathbf{a}}}{B}} = I\sqrt{\frac{\overline{ATS}}{(r\Omega B_d + B_p)}}$$

If the projected aperture lies entirely within the boundary of the detector and the observed shadow pattern is perfectly sharp, $S = r$. If in addition the particle background is a negligible contributer to the noise, r drops out of the signal to noise equation:

$$\Psi_0 = I\sqrt{\frac{\overline{AT}}{B_d\Omega}}$$

Ψ_0 represents an upper limit on the sensitivity of the coded aperture system.

If the background is uncertain, the expected detector response may be written as:

$$\langle \mathbf{h} \rangle = \frac{AT}{N} \begin{bmatrix} 1 & s_1 \\ 1 & s_2 \\ 1 & s_3 \\ \cdots & \cdots \end{bmatrix} \begin{bmatrix} \beta \\ I \end{bmatrix}$$

where:

$$\beta = r\,\Omega B_d + B_p$$

I'll call the first column of the detector response matrix the "background vector", or **b**; b_i=1 for all i. The least squares estimates of I and β may be obtained by solving the equation:

$$\frac{AT}{N} \begin{bmatrix} \mathbf{b}\cdot\mathbf{b} & \mathbf{b}\cdot\mathbf{s} \\ \mathbf{b}\cdot\mathbf{s} & \mathbf{s}\cdot\mathbf{s} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{I} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{s} \end{bmatrix} \mathbf{h}$$

At this point, it is convenient to introduce two more definitions:

$$R = \frac{\mathbf{b}\cdot\mathbf{s}}{N}$$

$$\tilde{S} = S - R^2$$

R is the average value in the shadow pattern **s**; thus R^2 is the portion of the power in the shadow pattern due to its nonzero average (the "DC" power), and $\tilde{S}$ is the rest of the power in the shadow pattern (the "AC" power). Rewriting the least squares equation in terms of R and S (also noting that $\mathbf{b}\cdot\mathbf{b}=N$):

$$AT \begin{bmatrix} 1 & R \\ R & S \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{I} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{s} \end{bmatrix} \mathbf{h}$$

The solution is:

$$\begin{bmatrix} \hat{\beta} \\ \hat{I} \end{bmatrix} = \frac{1}{AT\tilde{S}} \begin{bmatrix} S & -R \\ -R & 1 \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{s} \end{bmatrix} \mathbf{h} = \frac{1}{AT\tilde{S}} \begin{bmatrix} S\,\mathbf{b} - R\,\mathbf{s} \\ \mathbf{s} - R\,\mathbf{b} \end{bmatrix} \mathbf{h}$$

The estimated source intensity is:

$$\hat{I} = \frac{1}{AT\tilde{S}} \tilde{\mathbf{s}}\cdot h$$

where:

$$\tilde{\mathbf{s}} = \mathbf{s} - R\,\mathbf{b}$$