Coded aperture imaging: predicted performance of uniformly redundant arrays

E. E. Fenimore

Uniformly redundant arrays (URA) have autocorrelation functions with perfectly flat sidelobes. The URA combines the high-transmission characteristics of the random array with the flat sidelobe advantage of the nonredundant pinhole arrays. A general expression for the signal-to-noise ratio (SNR) has been developed for the URA as a function of the type of object being imaged and the design parameters of the aperture. The SNR expression is used to obtain an expression for the optimum aperture transmission. Currently, the only 2-D URAs known have a transmission of ½. This, however, is not a severe limitation because the use of the nonoptimum transmission of ½ never causes a reduction in the SNR of more than 30%. The predicted performance of the URA system is compared to the image obtainable from a single pinhole camera. Because the reconstructed image of the URA contains virtually uniform noise regardless of the original object's structure, the improvement over the single pinhole camera is much larger for the bright points than it is for the low intensity points. For a detector with high background noise, the URA will always give a much better image than the single pinhole camera regardless of the structure of the object. In the case of a detector with low background noise, the improvement of the URA relative to the single pinhole camera will have a lower limit of $\sim(2f)^{-1/2}$, where f is the fraction of the field of view that is uniformly filled by the object.

I. Introduction

Coded aperture imaging with Fresnel zone plates was first conceived by Mertz and Young.[1] Dicke[2] subsequently suggested random pinhole arrays to achieve the same basic concept. This concept is to replace the single opening of a simple pinhole camera with many openings called, collectively, the aperture. The recorded picture will consist of many overlapping images of the emitting object and, in general, bears no resemblance to the object. Computer or optical processing of the picture is required to produce the reconstructed object which should resemble the original object.

The motivation for using many openings is to improve the signal-to-noise ratio (SNR) in the reconstructed object by increasing the collecting area while maintaining the angular resolution of the small opening. Adding more openings does not guarantee an improved SNR; the possible improvement depends on the object, the pattern of openings, and the processing techniques used. In this paper we present a SNR relationship for the uniformly redundant array (URA) patterns discussed by Fenimore and Cannon[3] (hereafter referred to as Paper 1).

URAs are nonrandom patterns of openings that have the distinctive feature that the number of times a particular separation occurs between pairs of openings is the same for all separations. The autocorrelation of the aperture function (defined as one where there is an opening and as zero otherwise) has perfectly flat sidelobes. It was shown in Paper 1 that this feature of the autocorrelation function can be exploited to produce a coded aperture imaging system that is superior to previously used systems. Details of implementation are to be found in Paper 1. For the purposes here, the following heuristic review will suffice.

Figure 1 shows two different methods for implementating the URA. Figure 1(A) shows a standard method of implementation for coded aperture imaging that could be used with any aperture pattern. A point on the source projects a shadow of the aperture onto a detector. The direction to the point is determined by the direction in which the shadow pattern is offset from the center of the detector, and the intensity of the point is determined by the integrated signal produced in the detector. In this implementation, the detector should be large enough to observe the entire shadow produced by any source within the field of view.

The special properties of the URA give rise to an alternate implementation geometry [Fig. 1(B)]. Here,

The author is with University of California, Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545.

Received 20 May 1978.
0003-6935/78/1115-3562$0.50/0.

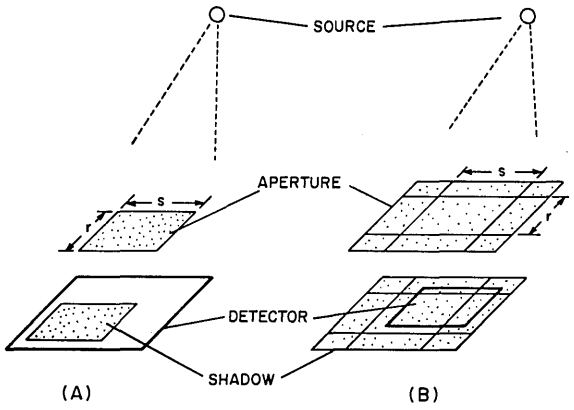

Fig. 1. (A) This coded aperture arrangement employs only the basic r by s pattern for the aperture and has the disadvantage that the detector must be large enough to contain the image from the full field of view. (B) This arrangement employs a $2r$ by $2s$ aperture composed of a mosaic of basic r by s patterns. Emitting points in the source produce shadows of cyclic permutations of the basic aperture pattern, and thus the detector needs to be only r by s in size.

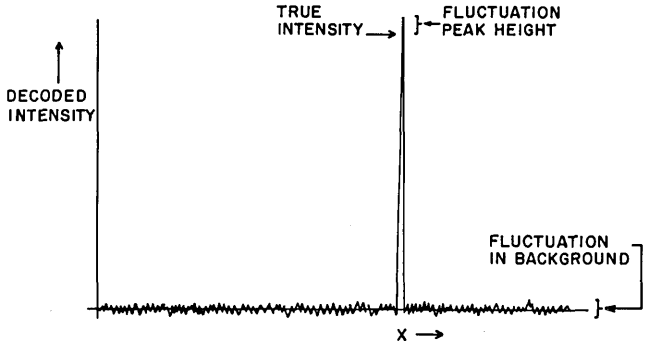

Fig. 2. A 1-D slice through a reconstructed object for a point source. The SNR is defined to be the peak height divided by the expected size of the fluctuations of the peak and the background.

the aperture is a section of an infinite mosaic of a basic URA pattern. The dimensions of the basic URA are r by s elements, and the aperture is a $2r$ by $2s$ section centered on a basic pattern. Paper 1 gives detailed expressions for calculating the aperture function, and they will not be repeated here. We will treat the aperture function as a $2r$ by $2s$ binary array, where a one represents an opening in the aperture, and a zero denotes opaque material. Thus, the ones in the aperture array have the same pattern as the openings in the actual aperture.

A source will produce a shadow of the aperture that has $2r$ by $2s$ picture elements and is offset from the center of the imaging system. However, because the aperture function is periodic, the central r by s region of the detector plane will contain a particular cyclic permutation of the basic r by s pattern. (This is true so long as certain conditions are met concerning the relative sizes of the basic array, the aperture array, and the source; see Paper 1 for details.) Thus, one needs to use a detector only as large as the projected basic aperture pattern in order to determine the object completely. The direction to any point on the source is determined by which particular cyclic permutation is projected onto the detector, and the intensity of the point is determined by the integrated signal produced in the detector. The smaller required detector will be especially beneficial to applications in x-ray astronomy where the weight and size of instrumentation are normally limited. Due to this important advantage, this paper will emphasize the geometry of Fig. 1(b).

The emitting source is modeled by an array S. Each resolution element of S (i.e., S_{ij}) represents the integrated number of photons seen by a single opening in the aperture from one resolution element on the object. A resolution element is defined as the area of a source that can be seen through one opening by one point on the detector. Given the definition of the aperture A and S, the recorded picture array is (see Paper 1)

$$P = S * A + D, \tag{1}$$

where $*$ is the correlation operator, and D is the detector background noise. The reconstructed object is defined as

$$\hat{S} \equiv P * G = S * (A * G) + D * G, \tag{2}$$

where G is an array chosen such that $A * G$ has desired properties. In particular, for the URAs used in Paper 1, if G is defined as

$$G_{ij} = 1 \text{ if } A_{ij} = 1; \quad G_{ij} = -1 \text{ if } A_{ij} = 0, \tag{3}$$

$A * G$ is a delta function, and $S * \delta$ is exactly the original object S. Thus, although G was defined for a correlation, it is also the convolutional inverse of A, which means the source will be reconstructed perfectly (i.e., no artifacts). Since G contains terms that are all of the same magnitude, it will treat the noise well (i.e., $D * G$ will be well behaved).

In Paper 1 a particular set of URAs was used in which approximately one-half of the elements of A were equal to one, that is, the aperture had a transmission of $\frac{1}{2}$. In anticipation of a generalization of the URA functions, we have defined the following terms. Let A be the periodic aperture function equal to an infinite mosaic of basic (r by s) URA patterns. Let A' be the basic aperture pattern (i.e., an r by s array). Then $A * A'$ will have a central peak of height N, where N is equal to the number of ones in A'. For the arrays of Paper 1, N is equal to $(r \cdot s + 1)/2$. The correlation $A * A'$ will have sidelobes that are perfectly flat and which we define to have a value of M. For the arrays in Paper 1, M is equal to $(r \cdot s + 1)/4$. If G is defined as

$$G_{ij} = 1 \text{ if } A_{ij} = 1; \quad G_{ij} = M/(M - N) \text{ if } A_{ij} = 0, \tag{4}$$

$A * G$ is a function with a central spike equal to N, and the sidelobes are equal to zero (i.e., a delta function). For the arrays of Paper 1, $M/(M - N)$ is -1.

II. Signal-to-Noise Ratio for URAs

Figure 2 shows a 1-D slice through a reconstructed object for a hypothetical point source. The system point-spread function (SPSF, see Paper 1) predicts that the expected reconstructed object will be one nonzero term equal to the intensity of the point source with zeros

for the rest of the decoded object. (For the purpose of demonstration, it has been assumed that there is no detector background noise, i.e., no signal-independent spurious counts.) The actual reconstructed object will have noise due to the Poisson statistics involved with the photons going through each pinhole. These statistics give rise to both the background fluctuations and the variation in the peak height (Fig. 2).

Basically the SNR will be the ratio of the expected height of the peak to some noise term. There are several ways to define a noise term for Fig. 2. If one is solely interested in the intensity of the point source, an appropriate noise term might be the variance of the fluctuations of the peak height.[4] The problem with such a definition is that the resulting SNR expression predicts the best aperture is one that is completely open, that is, one big hole. A single large hole would have little angular resolution and would not give a location for the source. A noise term which is appropriate for locating sources is the variance of the background in which the peak must be found.[2] However, such a definition leads to an expression that implies that the best aperture for imaging a point source is one with only a single pinhole.

The problems of determining an intensity and of finding a location for a point source are in practice usually coupled together. Thus our noise term due to a point source will include[5,6] both the variance of the peak height and the variance of the fluctuations in which the peak must be located. (Fortunately, with the URA, the variance of the fluctuations in the vicinity of the peak is independent of the size of the area for which it is evaluated.) As we shall see in Sec. III, such a definition gives reasonable results for the optimum density.

An alternative definition would be the ratio of power in the true image to the power of the noise in the reconstructed image. However, as shown in Fig. 2, the integrated power in the noise might be comparable to the power in the point source and thus give a deceivingly low value. Instead the SNR will be defined as a function of the position in the reconstructed object as

$$\mathrm{SNR}_u(i,j) = \frac{E(S_{ij})}{\{\mathrm{VAR}[C(S_{ij},\hat{S}_{ij})] + \sum_k \sum_l \mathrm{VAR}[C(S_{kl},\hat{S}_{ij})] + \mathrm{VAR}[C(D,\hat{S}_{ij})]\}^{1/2}}, \quad (5)$$

where $E(S_{ij})$ is the expected (true) value for the ij-point in the reconstructed object, $\mathrm{VAR}[C(S_{ij},\hat{S}_{ij})]$ is the variance of the contribution of the ijth object point to the ijth reconstructed point,

$$\sum_k \sum_l \mathrm{VAR}[C(S_{kl},\hat{S}_{ij})]$$

is the variance in the vicinity of $\hat{S}_{ij}$ due to all the sources within the field of view, and $\mathrm{VAR}[C(D,\hat{S}_{ij})]$ is the variance due to the background noise D. We define $[\mathrm{VAR}(\hat{S}_{ij})]^{1/2}$ to represent the denominator of Eq. (5). Note that Eq. (5) is similar to the square root of the ratio of the power in the object to the power of the noise, except it is taken on a point-by-point basis.

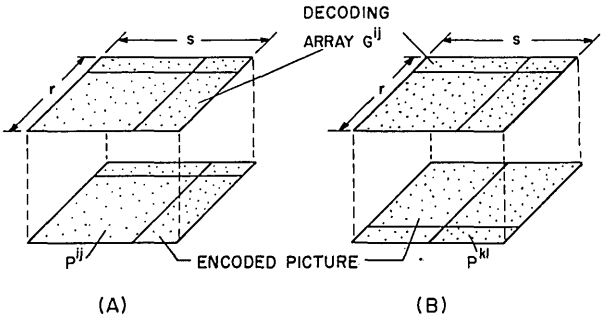

Fig. 3. (A) Orientation of the G array and the picture from the S_{ij} source when calculating the $\hat{S}_{ij}$ point in the reconstructed object. (B) Same as A except the picture is due to the S_{kl} source.

The expected value in the reconstructed object due to a particular source S_{ij} is simply NS_{ij}. The N arises from the fact that we defined the SPSF in Paper 1 as $A * G$ with a central height of N. Of course, we could have divided the SPSF by N to obtain a delta function with a height of unity so that the reconstructed image would be normalized to that which would be seen through a single opening. However, we wanted to normalize the reconstructed object to an image with an intensity commensurate with the actual open area of the aperture. Such a normalization also simplifies the expressions that lead to the SNR equation.

In order to calculate $\mathrm{VAR}(\hat{S}_{ij})$ we must first calculate $E(\hat{S}_{ij})$, the expected value of $\hat{S}_{ij}$. There are three possible contributors to a point in the reconstructed object. Certainly S_{ij} will contribute; in addition, there might be contributions from other sources in the field of view (for example S_{kl}) and from the background noise [D in Eq. (2)]. Each contribution will depend explicitly on the details of the decoding process. Because the geometry of Fig. 1(B) is assumed, the mosaic aperture will allow a smaller picture than might be assumed from Eq. (1). Since the Fig. 1(B) geometry is a nonstandard implementation, details of the correlation in Eq. (2) and its effect upon the SNR will be specified.

To calculate a term in a correlation of two functions, one performs a point-by-point multiplication and then adds up the resulting products. The correlation is a function of the lag, which is how much one function is shifted relative to the other function before the point-by-point multiplication. In the case of decoding a picture from the Fig. 1(B) geometry, one of these functions (the picture P) is truncated such that it consists of only an r by s array. Figure 3(A) shows how the $\hat{S}_{ij}$ term is calculated. The encoded picture in Fig. 3(A) is what one would expect from the source S_{ij}, that is, a shadow cast by an r by s section of the aperture offset from the center of the aperture [see Fig. 1(B)]. As mentioned above, an offset r by s section will be a cyclic permutation of the basic aperture pattern. We will denote that particular cyclic version as A^{ij}. (For the

purposes of this paper, subscripts will be used to denote an element of an array, whereas superscripts will be used to denote an array function.) The picture from the S_{ij} source (i.e., P^{ij}) will have an expected value of $S_{ij} \cdot A^{ij}$.

Figure 3(A) also shows the orientation of the G array for calculating the $\hat{S}_{ij}$ term. The G array has been lagged with respect to the aperture to correspond to the $\hat{S}_{ij}$ term in the correlation in Eq. (2). Because P is zero outside its r by s region, only the r by s section of the G array shown in Fig. 3(A) will contribute to $\hat{S}_{ij}$. That section, called G^{ij}, is related to A^{ij} by an equation similar to Eq. (4). In Fig. 3(A), $\hat{S}_{ij}$ is obtained by adding up the products of all the G^{ij} and P^{ij} elements that are vertically aligned (i.e., the point-by-point multiplication). The contribution of the S_{ij} element of the source to the $\hat{S}_{ij}$ element of the reconstructed object is, therefore,

$$C(S_{ij},\hat{S}_{ij}) = \sum_u \sum_v P^{ij}_{uv} G^{ij}_{uv}. \tag{6}$$

$C(S_{ij},\hat{S}_{ij})$ is a weighted sum of the elements of P^{ij}. The nonzero terms of P^{ij} are random variables with a mean of S_{ij} and a variance of S_{ij} (assuming Poisson statistics). Thus, the expected value of $C(S_{ij},\hat{S}_{ij})$ is just the sum of the weighting coefficients (G^{ij}_{uv}) times the expected value of P^{ij}_{uv}, that is,

$$E[C(S_{ij},\hat{S}_{ij})] = S_{\bar{y}} \sum_u \sum_v A^{ij}_{uv} G^{ij}_{uv}. \tag{7}$$

Referring to Fig. 3(A) to evaluate Eq. (7), we find that of the $r \cdot s$ terms in A^{ij}, N of those terms are one, and each is aligned with a one in the G^{ij} array. Because the remaining $r \cdot s - N$ terms in A^{ij} are zero [each aligned with a $M/(M-N)$ term in G^{ij}], they do not contribute to the summation in Eq. (7). Therefore, Eq. (7) becomes

$$E[C(S_{ij},\hat{S}_{ij})] = NS_{ij}. \tag{8}$$

Figure 3(B) will be used to calculate the expected contribution of the S_{kl} source to the $\hat{S}_{ij}$ term. The G function is the same as in Fig. 3(A), but the picture function now corresponds to an off-axis source S_{kl}, that is, its expected value is $S_{kl} \cdot A^{kl}$. Following arguments similar to above, the expected contribution of S_{kl} to $\hat{S}_{ij}$ is

$$E[C(S_{kl},\hat{S}_{ij})] = S_{kl} \sum_u \sum_v A^{kl}_{uv} G^{ij}_{uv}. \tag{9}$$

At this point the special properties of the URA play an important role. Regardless of the values of k, l and i, j (as long as k, l are not equal to i, j), there will be N nonzero terms in A^{kl} (all equal to unity), and exactly M of these terms will be aligned with a one in G^{ij}. The remaining N-M nonzero terms will always be aligned with the $M/(M-N)$ terms in G^{ij}. Thus, Eq. (9) becomes

$$E[C(S_{kl},\hat{S}_{ij})] = S_{kl}[M \cdot 1 + (N-M) \cdot M/(M-N)] = 0, \tag{10}$$

and there will be no expected contribution to $\hat{S}_{ij}$ from any of the other resolution elements of the source. The crucial attribute of the URA is that Eq. (10) is true for all k, l. It is now clear why we chose the value $M/(M-N)$ in Eq. (4); it was chosen after knowing the form of Eq. (10) in order to eliminate the contributions from the other sources.

The final possible contributor to $\hat{S}_{ij}$ is the detector noise term D in Eq. (2). The array D contains the background noise, that is, the signal that would be present even if the aperture were completely opaque. Dark currents, electronic noise, cosmic rays in proportional counters, and/or fog on film are examples of the types of background noise being considered. Assuming such noise is uniform across the detector, we define B as the average number of false counts per element of the picture array. Therefore, $D_{uv} = B$, and the expected contribution of the D to $\hat{S}_{ij}$ is

$$\begin{aligned} E[C(D,\hat{S}_{ij})] &= \sum_u \sum_v D_{uv} G^{ij}_{uv} = B \sum_u \sum_v G_{uv} \\ &= B \cdot [N \cdot 1 + (rs - N)M/(M-N)], \end{aligned} \tag{11}$$

which, for the arrays of Paper 1, is just B.

We can now total all the contributions to $\hat{S}_{ij}$. The term $E[C(D,\hat{S}_{ij})]$, which is constant for all ij, adds only a removable dc level, which will not effect the SNR. Since none of the other sources except S_{ij} contribute, we obtain

$$E(\hat{S}_{ij}) = NS_{ij} + \text{removable dc term}. \tag{12}$$

Once $E(\hat{S}_{ij})$ is known, the denominator of Eq. (5) can be evaluated. Because $E(\hat{S}_{ij})$ is the same (to within a removable constant) as $E(S_{ij})$, the VAR terms in Eq. (5) will depend only on the fluctuations of $\hat{S}_{ij}$ and the background caused by the quantum noise.

Since $\hat{S}_{ij}$ is equal to a sum of weighting factors (G^{ij}_{uv}) times a random variable (P^{ij}_{uv}), the variance of that sum will be the sum of the square of the weighting factors times the expected variance of the random variable ($S_{ij}A^{ij}_{uv}$, assuming Poisson statistics). Following similar arguments that led to Eq. (8),

$$\text{VAR}[C(S_{ij},\hat{S}_{ij})] = NS_{ij}. \tag{13}$$

Similarly, the variance caused by all the sources and the background are, respectively [cf. Eqs. (10) and (11)],

$$\begin{aligned} &\text{VAR}[C(S_{kl},\hat{S}_{ij})] = S_{kl}[M \cdot 1^2 + (N-M)(M/M-N)^2], \\ &\text{VAR}[C(D,\hat{S}_{ij})] \equiv \lambda B = [N \cdot 1^2 + (rs-N)(M/M-N)^2]B. \end{aligned} \tag{14}$$

Using Eqs. (13) and (14) in Eq. (5) gives

$$\text{SNR}_u(i,j) = \frac{NS_{ij}}{\left(NS_{ij} + \dfrac{MN}{N-M} I_T + \lambda B\right)^{1/2}}, \tag{15}$$

where I_T is the integrated intensity of the source

$$\sum_k \sum_l S_{kl}.$$

If the geometry of Fig. 1(A) is used instead of Fig. 1(B), by following similar arguments, one can show that the only difference in Eq. (15) is that λ is increased by the ratio of the detector area to $r \cdot s$.

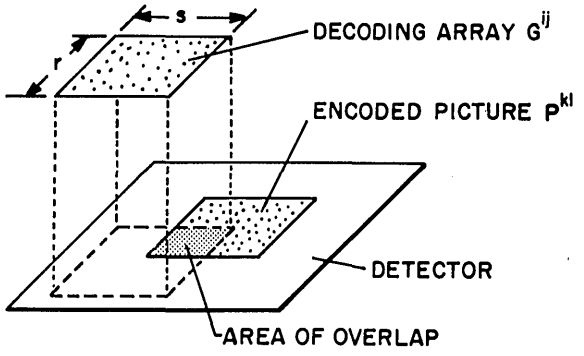

Fig. 4. Orientation of the G array and the picture from the S_{kl} source when calculating the $\hat{S}_{ij}$ point in the reconstructed object. The random array geometry of Fig. 1(A) is assumed.

It would be instructive to perform the analogous calculation using a random array for the aperture. The geometry of Fig. 1(A) is assumed as well as a balanced correlation decoding technique (see Paper 1). The expected contribution of the S_{ij} to $\hat{S}_{ij}$ will remain NS_{ij}, but now there will be a net contribution from the other sources in the field of view. For example, Fig. 4 shows the picture array from some S_{kl} source. Because of the geometry, the contribution to the total picture (i.e., the detector) is an r by s region which is smaller than the total picture. In calculating the $\hat{S}_{ij}$ point, the G array is lagged with respect to the picture and will be aligned with only a portion of the P^{kl} pattern. The area in common is approximately α, where α is equal to $(r - |i - k|)$ times $(s - |j - l|)$. If we define the density of the array to be $\rho = N/rs$, the area of overlap will have approximately $\rho\alpha$ nonzero (unity) terms. We now define γ (roughly equal to $\rho^2\alpha$) to be the number of the $\rho\alpha$ terms that are aligned with ones in the G array and η (roughly equal to $\rho\alpha - \rho^2\alpha$) to be the number aligned with the $\rho/\rho - 1$ terms in G [see Paper 1, Eq. (13)]. The expected contribution of S_{kl} to $\hat{S}_{ij}$ in the random array case is therefore

$$E[C_R(S_{kl},\hat{S}_{ij})] = [\gamma \cdot 1 + \eta\rho/(\rho - 1)]S_{kl} \equiv \beta S_{kl}. \tag{16}$$

From Paper 1, β is recognized as just one of the sidelobes in the SPSF, $A * G$. β will vary from point to point causing artifacts that mimic noise. (In Paper 1 we referred to these artifacts as inherent noise, which, in retrospect, was probably a misleading choice of terminology.) The artifacts cause $\hat{S}_{ij}$ to deviate systematically from $E(S_{ij})$ (in addition to randomly), and the variance in Eq. (5) becomes very large. In fact, the artifacts are the dominant source of error in the reconstructed object and constitute the basic limiting factor in coded aperture imaging without URAs. The artifacts are proportional to the signal; thus, their contributions to the variance [relative to $E(S_{ij})$] are not changed by a longer exposure or a reduction of the background. The SNR for the non-URA case will have a limiting value (usually very low), which is set by the artifacts.

The similarity between Eqs. (10) and (16) indicates that the origin of the URA's advantage lies in the fact that Eq. (10) gives a zero all the time, whereas in Eq. (16), β is roughly zero (but not exactly zero). The nonzero β in Eq. (16) means that there will be cross talk between the signals in different parts of the reconstructed object resulting in artifacts. The URA produces no cross talk between the signals in the object, but there is noise cross talk between different points. The noise cross talk gives rise to the I_T term in the denominator of Eq. (15). In non-URA systems, the artifacts dominate the SNR and form the basic limiting feature of coded aperture imaging. The URA removes this limitation, and the new basic limitation is orders of magnitude smaller. In the URA systems the new limitation depends on whether the stronger signal gained by the larger open area (relative to a single pinhole) is offset by the $[MN/(N - M)]I_T$ term in the noise [see Eq. (15)].

III. Optimum Aperture Transmission

In this section, we investigate the effect of varying the transmission or density of the array. We make the following approximations and definitions:

$$\rho \equiv N/(r \cdot s), \quad \psi_{ij} \equiv S_{ij}/I_T, \tag{17}$$
$$M \approx \rho^2(r \cdot s), \quad \xi \equiv B/I_T.$$

Substituting Eqs. (17) and (14) into Eq. (15) gives

$$\mathrm{SNR}_u(i,j) = \frac{(1 - \rho)^{1/2}\rho\psi_{ij}(r \cdot s \cdot I_T)^{1/2}}{[\rho^2(1 - \psi_{ij}) + \rho(\psi_{ij} + \xi)]^{1/2}}. \tag{18}$$

Equation (18) gives the SNR for the ijth point primarily as a function of the density and two dimensionless variables. ξ is the ratio of the background to I_T, the integrated intensity of the source, and ψ is the ratio of the source under consideration (S_{ij}) to I_T. These two dimensionless variables can in general parameterize any source. We wish to find the density (as a function of ξ and ψ) for which the SNR is a maximum. From Eq. (18), one can show that the optimum density is

$$\rho_T = \frac{[(\psi_{ij} + \xi)^2 + (1 - \psi_{ij})(\psi_{ij} + \xi)]^{1/2} - (\psi_{ij} + \xi)}{1 - \psi_{ij}}. \tag{19}$$

Figure 5(A) shows the optimum density as a function of ψ for select values of ξ. Because ξ is the ratio of the background noise in one picture element to the integrated intensity (through a single opening) of the entire field of view, ξ will usually be very small. Values of 10^{-5} are not unusual. However, certain situations in x-ray astronomy might have ξ as large as 1. Figure 5(A) shows that for small values of ψ and normal values of ξ (<0.1), the optimum SNR is often obtained by an aperture with a density much less than $\frac{1}{2}$. (Unless the source consists of a few point sources, ψ will also normally be quite small, ~0.02.) Fortunately, the effects of using a density of $\frac{1}{2}$ are not severe. Figure 5(B) shows R as a function of ψ and ξ, where R is the ratio of the SNR with $\rho = \frac{1}{2}$ to the SNR with $\rho = \rho_T$. In no case will the SNR resulting from using a density of $\frac{1}{2}$ be less than 70% of the SNR with an optimum density. It is fortunate that a nonoptimum density of $\frac{1}{2}$ does not produce severe effects because at this time the only 2-D URAs are those discussed in Paper 1, and they always have a density of $\frac{1}{2}$.

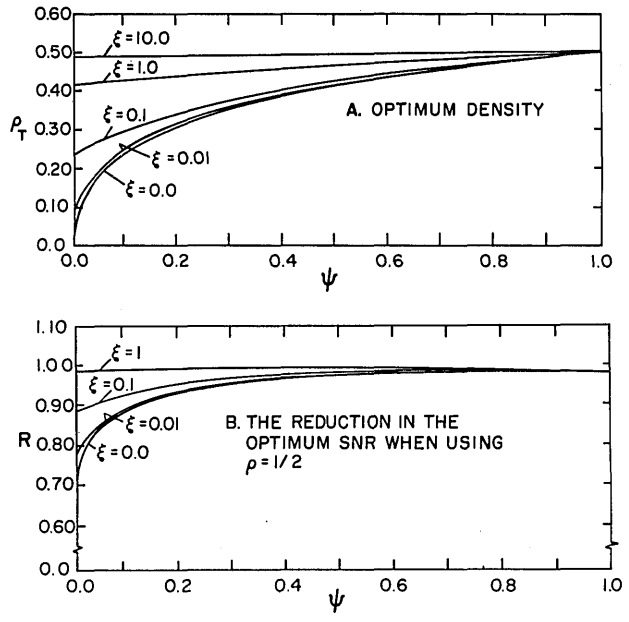

Fig. 5. (A) Optimum density of the aperture as a function of the relative intensity of a point in the source ψ for select values of the detector background noise ξ. (B) The ratio of the SNR with a density of ½ to the SNR with the optimum density as a function of ψ and ξ. Note that using a density of ½ never causes a large effect.

Gunson and Polychronopulos[7] have suggested a 1-D x-ray system involving arrays similar to the URA. These 1-D arrays can have several densities, and they can be developed into 2-D URAs (Gunson; private communication). However, as Fig. 5(B) shows, this probably will not be necessary because the advantage gained will always be small. On the other hand, a lower density would reduce the count rate, which is an important limiting factor in Anger cameras and proportional counters.

Gunson and Polychronopulos[7] also proposed an optimum density equation. Their equation, however, seems to agree with ours only if ψ is equal to zero. In particular, for point sources and low backgrounds ($\xi \sim 0$), their equation predicts an optimum density near zero, whereas it is clear that ½ gives a much better image. This problem stems from the fact that they include the fluctuations of the background but not the peak height [see discussion of Eq. (5)].

Even though the arrays with densities of ½ give SNRs near the optimum value, this should not be taken to indicate that the resulting image will have higher quality than that of a single pinhole. Figure 5(A) suggests that as ξ and ψ go to zero, the aperture should become a single opening (i.e., ρ approaches zero). This implies that for ξ and ψ near zero, the URA system will not do better than a single pinhole.

IV. Predicted Performance for the URAs of Paper 1

In this section, we will investigate the relative merit of the URA system. The URAs of Paper 1 always have ρ approximately equal to ½ and M equal to $N/2$. Under these conditions, Eq. (15) becomes

$$\mathrm{SNR}_u(i,j) = \frac{N^{1/2}S_{ij}}{(S_{ij}+I_T+2B)^{1/2}} = \frac{N^{1/2}\psi_{ij}I_T{}^{1/2}}{(1+2\xi+\psi_{ij})^{1/2}}. \quad (20)$$

Equation (20) should be used to predict the performance of a URA system.

To obtain a measure of the relative merit of the URA system, the URA image will be compared with the image obtainable by a single pinhole camera. Although we will be specifically considering a single pinhole, it should be realized that the SNR from a single pinhole is comparable with that from a rastering collimator or an imaging (i.e., diverging pencil beams) collimator. This approach has been the standard method of comparison for coded aperture imaging.[2,5–7]

A comparison between the URA and a single pinhole can at times be misleading. A larger SNR for the URA does not necessarily mean that it is the best system to use in a particular situation. Often a more sophisticated imaging system than a single pinhole might be better than the URA system. For example, the grazing incidence telescope concentrates the x rays such that a small detector (therefore less background) can be used. However, the grazing incidence telescope is ineffective above 3 keV, whereas the URA will be effective as long as closed portions of the aperture remain opaque.

Some systems operate in two stages. They have a set of coarse collimators that locate the source and then direct a set of fine collimators to image it (foveal system). In this manner, the instrument does not have to image the entire region to locate the source, rather it just needs to image (at the fine resolution) an area typical of the sources under consideration. The foveal system necessitates a splitting of the available space for the instrument between the coarse and the fine collimators. Therefore, relatively less collecting area can be dedicated to the imaging of the source at the fine resolution. An additional disadvantage of the foveal system is that it is ineffective with sources which vary on a time scale comparable with the time required to direct the set of fine collimators.

Modulation collimators (e.g., Oda[8]) are similar to coded aperture imaging systems, except they modulate the signal in time rather than spatially. However, they become ineffective with extended objects and transients that vary on a time scale comparable with the time scale of motion for the modulation collimator.

The most meaningful comparison with the pinhole image will be for sources that cannot take advantage of the more sophisticated imaging systems. In particular, the most appropriate comparison is for high-energy (>3 keV) extended or transient sources. For other sources, such a comparison will give us a basis for further (more detailed) comparisons.

The SNR for a single pinhole is (assuming Poisson statistics)

$$\mathrm{SNR}_p(i,j) = \frac{S_{ij}}{(S_{ij}+B)^{1/2}} = \frac{\psi_{ij}I_T{}^{1/2}}{(\psi_{ij}+\xi)^{1/2}}. \quad (21)$$

The ratio of the SNR for the URA to that of the single pinhole is referred to as the figure of merit or the multiplexing advantage.[5]

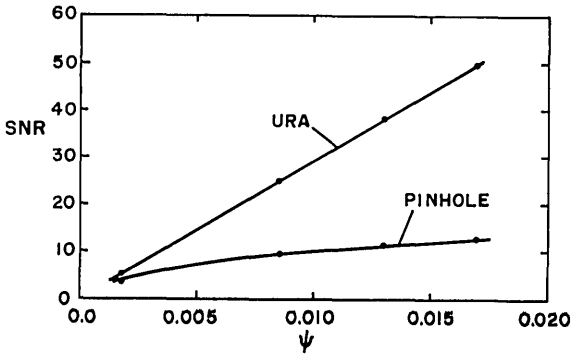

Fig. 6. Predicted SNR of a URA system and a pinhole camera for a source described in the text. Note that the improvement for the brighter points in the source (ψ large) is much better than for the low-intensity points.

$$F_{ij} \equiv \frac{\mathrm{SNR}_u}{\mathrm{SNR}_p} = \left[N \frac{(\psi_{ij} + \xi)}{2\xi + \psi_{ij} + 1} \right]^{1/2}. \tag{22}$$

From Eq. (22), the maximum improvement is $(N/2)^{1/2}$. This is lower than the often quoted improvement of $\sqrt{N}$ because the high dc background associated with autocorrelation analyses (see Paper 1) has been subtracted off in order to give the actual intensity of the object. The improvement is $\sqrt{N}$ if the dc background is not subtracted off. Since N will probably be as large as 10^4 or 10^5, there will be a net advantage even if ψ and ξ are as small as 0.05. The URA will be superior for a collection of point sources, or extended sources (especially those with a patchy appearance), and when the detector background is high.

We now briefly describe several possible applications and compare for each the predicted performance of the URA to a single pinhole. To be fair, the competing systems should use the same resources. Thus, it is assumed that they will operate with the same size detector for the same period of time. The single pinhole camera has an optimum configuration when the image from the entire field of view just fills the detector area, because this results in a maximum collecting area per resolution element. Since distances from the detector to aperture and aperture to object are assumed to be the same in both systems, the image from a single opening in the URA system will also just fill the area of the detector. Considering the rules for implementation of the URA (see Paper 1), the field of view should consist of $r \cdot s$ resolution elements.

X-ray images of the sun (e.g., Underwood *et al.*[9]) and images of laser implosion pellets (e.g., Godwin[10]) are morphologically similar. In both cases there is a region of possible emission (for example, the disk of the sun) that defines the field of view with the $r \cdot s$ resolution elements. However, at any one time, only a fraction of the field of view is actually emitting, giving the sources a patchy appearance. When a solar flare occurs, only a fraction f will have appreciable emission. Although ψ will probably have a wide range of values, ψ will be $\sim 1/(f \cdot r \cdot s)$, and the advantage [assuming ξ is small in Eq. (22)] of using the URA will be about $(2f)^{-1/2}$. For the sun, F might be as large as 10, meaning a 100 times reduction in the time it takes to obtain an image with a particular SNR.

The improvement of the URA over the single pinhole is relatively greater for the bright points than it is for the low-intensity points. The denominator of Eq. (15) (i.e., the noise) is virtually independent of the position in the reconstructed object; thus, the bright points have the same noise level as do the low-intensity points. In contrast, the single pinhole noise is directly related (by Poisson statistics) to the intensity of the source within the resolution element under consideration. For this reason we explicitly defined Eq. (22) as a function of the position in the reconstructed object. No one number will characterize the improvement over the single pinhole; rather, the bright points will always have a better improvement F than the weak points.

As an example, consider a system with N equal to 882 ($r = 41$, $s = 43$) and a source that consists of 10 points with $\psi = 0.017$, 20 points with $\psi = 0.013$, 40 points with $\psi = 0.0085$, 130 points with $\psi = 0.0017$, and the remaining points (1563) have ψ equal to zero. Figure 6 shows the predicted performance of the URA and the single pinhole. We have assumed that ξ is zero (this will favor the single pinhole) and that I_T is 10^4. Although the URA and single pinhole give comparable SNRs when ψ is small, the URA has imaged the bright points comparably better than the low-intensity points.

Figure 6 shows that one number cannot characterize the improvement obtainable by the URA system. In fact, the estimate that F is $(2f)^{-1/2}$ is usually a lower limit. For the above example, f is approximately 200/1763, which gives 2.1 as an estimate of the improvement. Figure 6 shows an advantage of 1.21 for the weakest points ($\psi = 0.0017$), and that advantage grows to 3.85 for the brightest points ($\psi = 0.017$). In most cases, $(2f)^{-1/2}$ will be close to a lower limit for the improvement. To actually determine the relative merit of a URA system, plots similar to Fig. 6 must be generated for the source being considered.

In both solar and laser-fusion imaging, the diagnostics are most sensitive to the location of the bright points. A URA system will be able to provide a larger improvement at the locations in the image where that improvement will do the most good. Thus, the URA system should be well suited to both solar and laser-fusion imaging.

There are several possible applications of coded aperture imaging in the medical field.[11] Typically, a radioisotope is injected into a patient, and its eventual presence or absence in various parts of the body is the basis for a diagnosis. The applicability of the URA to medical sources depends on how the radioisotope distributes itself in the patient. We will consider two situations, one of which would be an unlikely candidate for the URA (injection of thallium-201 in order to study cardiac infarcts) and one for which the URA would be a worthwhile technique (bone scans).

When a patient is suspected of having a cardiac infarct, one standard diagnostic method is to inject the patient with the radioisotope thallium-201, which

preferentially collects in the blood and muscles of the heart. Because the site of the infarct contains damaged tissue, relatively less of the radioisotope will collect there. The x-ray imaging system should provide the physician with the location of the infarct by detecting its lower emission. There are two factors which make this an unlikely candidate for a URA system. First, a large fraction of the radioisotope is in other nearby organs and tissue. The total source (heart plus surrounding tissue) will be a source which approximately fills the field of view uniformly, and thus ψ will be $\sim 1/(r \cdot s)$. Equation (22) indicates that the resultant URA image will be inferior to that of the single pinhole by a factor of about $(\frac{1}{2})^{-1/2}$. Second, the diagnosis depends on the details of the low-intensity region of the source for which the improvement due to the URA is relatively small (see discussion of Fig. 6). However, there are radioisotopes other than thallium-201, which will preferentially collect in the infarct, thus mitigating these problems.

In contrast to the above case, bone scans should be a good candidate for a URA system. Because the skeleton takes up a small fraction of the body and there are radioisotopes which strongly prefer bone structure, a bone scan would involve a source with ψ significantly larger than $1/(r \cdot s)$. Equation (22) predicts that the URA will provide a better image than a single pinhole camera. In addition, many bone diseases are characterized by excessive absorption of the radioisotope; thus, the diagnosis will depend on points in the object that have the largest improvement over the single pinhole.

The above comparisons might be unfair to the single pinhole. In many medical applications a converging collimator combined with an Anger camera could be placed close to the patient and thus improve the SNR because there would be a larger number of received counts due to a smaller inverse square of the distance. The URA has to be farther away to avoid a large magnification constant which reduces the number of resolution elements when the size of the detector is fixed (see Paper 1).

So far we have only discussed cases where the detector background noise was assumed to be zero. Equation (20) shows that if ξ is large ($\gtrsim 0.01$), the URA advantage will be appreciable regardless of ψ, that is, regardless of how extended or complex the source is. This will be true even if the object fills the field of view; therefore, for the information received, instrumentation more sophisticated than the single pinhole probably would not reduce the advantage of the URA. In many situations and energy bandpasses, detectors can be built with small ξ; therefore, an experiment with high ξ might be uncommon. However, in x-ray astronomy it is not unusual to have sources that are much less intense than the detector background; thus, the URA will be particularly effective when used on extended x-ray or gamma-ray celestial sources.

Another typical situation that occurs in x-ray astronomy is a field of view containing a collection of point sources. ψ is expected to be on the order of at least 0.01, and this, combined with the high ξ expected, means that a URA system would locate such sources easily. A particularly important application would be the study of transient x-ray sources. These sources emit for such a short time that it is difficult to point a more sophisticated system at them. The combination of a large field of view yet good angular resolution with the ability to operate in extreme detector noise makes the URA a very promising instrument for x-ray astronomy.

Finally several experimental effects which would tend to decrease the SNR for all types of coded apertures have not been included. When film is used errors will probably be generated in the densitometer; also, there could be an alignment problem between the picture and the densitometer. Fabrication errors in the aperture could also produce errors in the reconstructed image. The most serious source of error might be due to nonuniformity of the detector's spatial efficiency. Equation (10) represents a balancing of signals over the entire detector area. If one section of the detector is more efficient than the rest, artifacts will result.

V. Summary

A SNR equation has been developed [Eq. (18)] to assist in predicting the performance of the URA. The SNR expression differs from those developed for other coded apertures[2,4] in that the noise term includes both the variance of the peak height and the variance of the fluctuations in which the peak must be located [see discussion of Eq. (5)]. The SNR has been formulated as a function of the transmission or density of the aperture, the ratio of the intensity of a resolution element to the integrated intensity of the source (ψ), and the ratio of the detector's background noise to the integrated intensity (ξ). The optimum density was found as a function of ψ and ξ [see Eq. (19) and Fig. 5(A)]. Figure 5(B) shows that the use of an aperture with a density of $\frac{1}{2}$ is not a severe limitation because the nonoptimum arrays with densities of $\frac{1}{2}$ never give an SNR that is more than 30% smaller than the optimum density. The one exception is when the source produces an unacceptable count rate in the detector. In that case, it might be worthwhile to use an array of lower density. We conclude that the URAs presented by Fenimore and Cannon[3] (Paper 1) can be used in most situations.

We presented an SNR equation [Eq. (20)] for the URAs discussed in Paper 1 and compared the URA system to the single pinhole camera in Eq. (22). Our conclusions concerning the usefulness of the URA system follow.

(1) Because the encoding and decoding process produces virtually uniform noise in the reconstructed object, the improvement over a single pinhole is larger for the bright points than the low-intensity points (see discussion of Fig. 6).

(2) Although one number generally will not characterize the improvement possible with the URA, a lower limit is roughly $(2f)^{-1/2}$, where f is the fraction of the field of view that is uniformly filled by the object. Because one number will not truly represent the advantage of the URA, plots similar to Fig. 6 must be generated for the sources being considered.

(3) The URA will always give a better image than that from a single pinhole in the case of a few (10 to 20) point sources ($\psi \gtrsim 0.01$).

(4) The URA will always give a better image if the detector background noise is significant ($\xi > 0.01$).

In general, the URA will be best suited for locating and imaging sources that emit from less than one-third of the field of view, especially if there is significant detector background noise.

The author thanks Harry Barrett and Paul Harper for helpful discussions and information on the medical applications. Richard Blake and Michael Cannon were also very instrumental in the development of this work. This work was done under the auspices of the U.S. Department of Energy and the National Aeronautics and Space Administration grant S57094A.

References

1. L. Mertz and N. Young, in *Proceedings of the International Conference on Optical Instruments and Techniques* (Chapman and Hall, London, 1961), p. 305.
2. R. H. Dicke, Astrophys. J. **153**, L101 (1968).
3. E. E. Fenimore and T. M. Cannon, Appl. Opt. **17**, 337 (1978).
4. H. H. Barrett and G. D. DeMeester, Appl. Opt. **13**, 1100 (1974).
5. C. M. Brown, Ph.D. thesis, "Multiplex Imaging and Random Arrays," U. Chicago (1972).
6. C. M. Brown, J. Appl. Phys. **45**, 1806 (1973).
7. F. Gunson and B. Polychronopulos, Mon. Not. R. Astron. Soc. **177**, 485 (1976).
8. M. Oda, Appl. Opt. **4**, 143 (1965).
9. J. G. Underwood, J. E. Milligan, A. C. Deloach, and R. B. Hoover, Appl. Opt. **16**, 858 (1977).
10. R. P. Godwin, Adv. X-Ray Anal. **19**, 533 (1975).
11. H. H. Barrett, J. Nucl. Med. **13**, 382 (1972).