LA-UR-77-2369

CONF-771023--4

TITLE: A CLASS OF NEAR-PERFECT CODED APERTURES

MASTER

AUTHOR(S): T. M. Cannon, E. E. Fenimore

SUBMITTED TO: For presentation at the IEEE-1977 Nuclear Science Symposium, October 19-21, 1977, San Francisco, CA.

los alamos scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87545

An Affirmative Action/Equal Opportunity Employer

Form No. 836
St. No. 2629
1/75

UNITED STATES
ENERGY RESEARCH AND
DEVELOPMENT ADMINISTRATION
CONTRACT W-7405-ENG. 36

A CLASS OF NEAR-PERFECT CODED APERTURES*

T. M. Cannon
E. E. Fenimore
University of California
Los Alamos Scientific Laboratory
Los Alamos, New Mexico 87545

Summary

Coded aperture imaging of gamma ray sources has long promised an improvement in the sensitivity of various detector systems. The promise has remained largely unfulfilled, however, for either one of two reasons. First, the encoding/decoding method produces artifacts, which even in the absence of quantum noise, restrict the quality of the reconstructed image. This is true of most correlation-type methods. Second, if the decoding procedure is of the deconvolution variety, small terms in the transfer function of the aperture can lead to excessive noise in the reconstructed image.

We propose to circumvent both of these problems by use of a uniformly redundant array (URA) as the coded aperture in conjunction with a special correlation decoding method. The correlation of the decoding array with the aperture results in a delta function with deterministically zero sidelobes. The properties of the encoding/decoding method are similar to those of the nonredundant pinhole array (NRA), however, the URA can be composed of thousands of holes whereas the NRA contains less than 40. In short, the URA offers the transmission advantage of the random array or Fresnel zone plate without introducing the artifacts typically seen when those apertures and others are used.

It is shown that the reconstructed image in the URA system contains virtually uniform noise regardless of the structure in the original source. Therefore, the *improvement* over a single pinhole camera will be relatively larger for the brighter points in the source than for the low intensity points. In the case of a large detector background noise the URA will always do much better than the single pinhole regardless of the structure of the object. In the case of a low detector background noise, the improvement of the URA over the single pinhole will have a lower limit of approximately $(1/2f)^{1/2}$ where f is the fraction of the field of view which is uniformly filled by the object.

Introduction

Coded aperture imaging was introduced by Dicke[1] and Ables[2] for application in x-ray astronomy. The basic concept is to replace the single opening of a simple pinhole camera with many openings called collectively the aperture. The recorded picture will thus consist of many overlapping images of the emitting object and in general bears no resemblance to the object. Computer or optical processing[1-3] of the picture is required in order to produce the "reconstructed object" which hopefully will resemble the original source.

The analysis methods developed so far can be categorized as either a deconvolution or a correlation. The following is a heuristic view of these two methods.

*Work performed under the auspices of the U.S. Department of Energy, Contract No. W7405-ENG-36, and NASA Grant S-57094A.

If the recorded picture is represented by the function P, the aperture by A and the object by O, then

$$P = (O * A) + n \tag{1}$$

where * is the correlation operator and n is some noise function. These terms are more fully defined in Ref. 4, hereafter referred to as Paper I. In the deconvolution methods, the object is solved for by

$$\begin{aligned}\hat{O} &= R\mathcal{F}^{-1}\left\{\mathcal{F}(P)/\mathcal{F}(A)\right\} \\ &= O + R\mathcal{F}^{-1}\left\{\mathcal{F}(n)/\mathcal{F}(A)\right\}\end{aligned} \tag{2}$$

where $\mathcal{F}, \mathcal{F}^{-1}$, and R are respectively the Fourier transform, the inverse Fourier transform and the reflection operator.

The main problem with deconvolution methods is that $\mathcal{F}(A)$ might have small terms. For example, we have empirically determined that roughly 15% of the Fourier transforms of 32 by 32 binary random arrays have at least one term which is zero. Although it is possible to avoid these particular arrays, it appears that it is a general property of large binary random arrays to have some small terms in their Fourier transform. These small terms can cause the noise to dominate the reconstructed object. The situation with deconvolution methods has been improved, however, by using Wiener filtering (Wood et al.[5]). Because of the poor noise-handling characteristics of the deconvolution methods, this paper will emphasize the correlation method.

In the correlation method the reconstructed object is defined to be

$$\hat{O} = P * G = RO * (A * G) + n * G \tag{3}$$

where G is called the postprocessing array[6,7] and is chosen such that A * G approximates a delta function. In general, we do not mean G to be the convolutional inverse function (A^{-1}), rather G will be selected in an *ad hoc* manner such that A * G has desirable properties. Normally G will be a binary array (as is A). If A * G *is* a delta function, then $\hat{O} = O + n * G$, and the object has been perfectly reconstructed except for the presence of the noise term. Note that the noise term in Eq. (3) will not have singularities as in the deconvolution method.

The original expectation of obtaining a roughly $\sqrt{N}$ improvement in the signal-to-noise ratio (due to the N openings in the aperture) has not been realized because A * G in general will not be a delta function. A point on the object will contribute to the reconstructed object the distribution A * G instead of a delta function.

Thus, even if there is no background noise and the source is intense enough such that shot noise is not a problem, the SNR for a point source becomes a fixed number regardless of the exposure time. The SNR becomes the ratio of the central peak in A * G to the "noise" in A * G, that is, the square root of the variance of the sidelobes. The resulting artifacts place a limit on the possible SNR improvement. The situation is much worse when the object is not a point source but is extended. In the extended case, the artifacts from *all* points in the object contribute noise to *each* point in the reconstructed object. The result is a low SNR which cannot be improved because the noise is set by the structure in A * G rather than counting statistics or background levels. In fact, the SNR for the coded aperture technique can be smaller than the SNR for a comparative single pinhole camera if the object is extended.

Choice of Encoding and Decoding Functions

There are a few arrays such that A * G is *effectively* a delta function (assuming A * G is sampled on the same scale as the size of the pinholes). Nonredundant arrays (NRA) have the the property that their autocorrelations (i.e., A * A) consist of a central spike with the sidelobes equal to unity out to some particular lag, L, and either zero or unity beyond that.[8] A true delta function would have all sidelobes cut to infinite lags equal to zero. If all the sidelobes are equal to a constant value (such as unity), then the only effect on the reconstructed object is the addition of a removable dc level. However, the sidelobes of the NRA are not all equal to the same value and thus the reconstructed object will contain artifacts unless the object is extremely small.

There is a class of arrays called pseudo-noise arrays[9,10] from which an A and G can be generated such that A * G *is* a delta function. In the pseudo-noise arrays the number of times that a particular separation (between a pair of "holes" or one in the aperture array) occurs is a constant regardless of which separation distance is under consideration, that is, the separations are uniformly redundant. We have labeled all arrays (Paper I) for which all separations (less than some maximum L) between pairs of holes occur a constant number of times as "uniformly redundant arrays (URA)." Thus, both the NRA and the arrays derived from pseudo-noise arrays are uniformly redundant arrays.

The URA aperture will be a section of an infinite uniformly redundant array consisting of a mosaic of identical basic arrays. The benefits and details of mosaicing are outlined in Paper I. These arrays follow from the pseudo-noise arrays described by Calabro and Wolf.[9] The basic array will have dimensions r by s where r and s are prime numbers and r-s equals 2. Thus, $A(i,j) = A(I,J)$, where $I = \mathrm{mod}_r i$ and $J = \mathrm{mod}_s j$. Furthermore,

$$A(I,J) = \begin{cases} 0 & \text{if } I = 0 \\ 1 & \text{if } J = 0,\ I \neq 0 \\ 1 & \text{if } C_r(I)C_s(J) = 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$C_r(I) = \begin{cases} 1 & \text{if there exists an integer } x, 1 \leq x < r \\ & \text{such that } I = \mathrm{mod}_r x^2 \\ -1 & \text{otherwise.} \end{cases} \qquad (4)$$

The postprocessing array, G, will be a section of the function

$$G(i,j) = \begin{cases} 1 & \text{if } A(i,j) = 1 \\ -1 & \text{if } A(i,j) = 0 \end{cases} \qquad (5)$$

which is used because it can be shown that the correlation of A with G is a mosaic of delta functions with zero sidelobes. Figure 1 shows a URA array with r = 43 and s = 41.

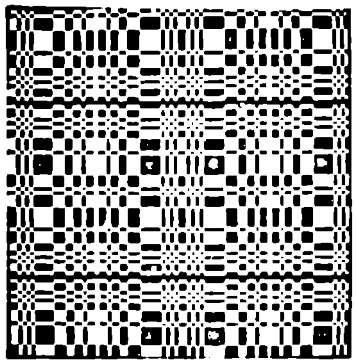


Fig. 1. A 43 by 41 uniformly redundant array (mosaicked)

The details of implementing A and G, including a mosaicking method that permits the aperture to be larger than previous coded aperture arrangements, are given in Paper I.

System Point-Spread Functions

Many of the characteristics of an imaging system can be seen in the system point-spread function (SPSF). The SPSF is defined to be the reconstructed object resulting from imaging a point source. From Eq. (3),

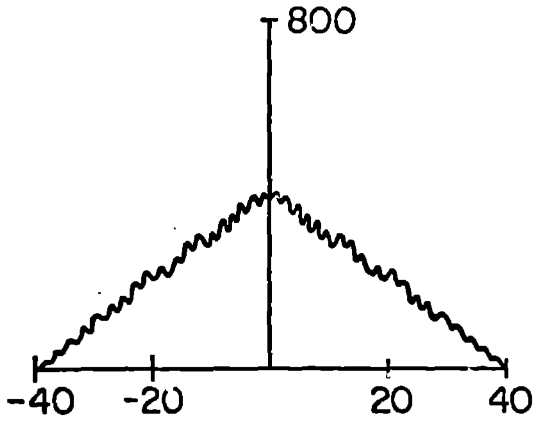
$$\mathrm{SPSF} = A * G \qquad (6)$$

The SPSF's for three different coded aperture systems are outlined below.

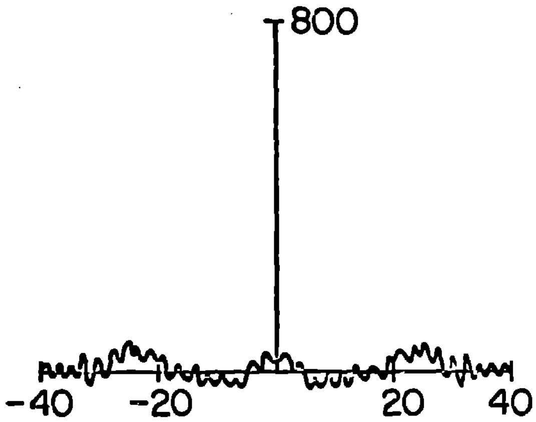
For the matched analysis process,[1,6] G is identical to A, and thus the SPSF is the autocorrelation of the aperture array. Figure 2a is a one-dimensional slice through a typical SPSF for the matched process. The two-dimensional SPSF is a spike on top of a pyramid. The ratio of the height of the spike to the height of the pyramid is the ratio of the number of pinholes to the number of possible pinhole positions (the "density," .5 in our example).

The reconstructed image will consist of the original image (source distribution) plus the image convolved with the pyramid. This will cause a severe degradation of the spatial resolution, especially if the source distribution is of low contrast. This degradation persists regardless of exposure time, and thus represents an upper limit on image quality.

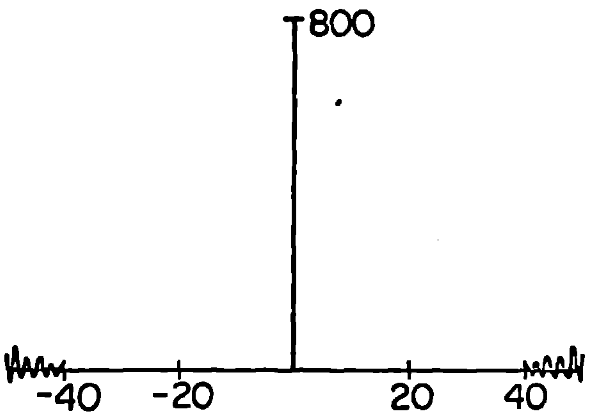
The SPSF for the matched process can be greatly improved by replacing the zeroes in G with -1's. Paper I refers to this as balanced correlation, which is similar to the mismatch method of Brown.[6,11] The SPSF for the balanced correlation method is a delta



(a) Random array (matched Process).



(b) Random array (balanced correlation).



(c) Uniformly redundant array (balanced correlation).

Fig. 2.

function with sidelobes having an *expected* value of zero. Figure 2b is a one-dimensional slice through a typical SPSF for the balanced correlation method.

The URA coded aperture system can be implemented with either the matched decoding me nod or balanced correlation. In either case the SPSF will be effectively a delta function with perfectly flat sidelobes. In the former case the sidelobe value is $(r \cdot s + 1)/4$, in the latter case it is zero. (See Fig. 2c.)

Simulations

We have performed computer simulations in order to demonstrate the differences between these various methods of coded aperture imaging. The matched and balanced correlation methods will be simulated with no source or detector noise in order to show that those procedures have artifacts which dominate the reconstructed object. An absence of noise is equivalent to exposing for a very long time with a perfect detector. The simulation of the URA system will include the noise and signal characteristics of an Anger camera viewing a one-millicurie source. Even under these conditions, the URA approach will be superior to the random array techniques when they are applied under perfectly noiseless conditions.

The test object, Fig. 3a, is a high-contrast object in the shape of a man. The man consists of 164 equally intense points in a 40 by 40 array. H's integrated intensity is approximately 1 millicurie. If the aperture to object separation is 3 cm and each pinhole is about .62 cm square, then each resolution element on the man emits about 210 photons/sec through each pinhole.

A simulation using a 40 by 40 random array and the matched process resulted in a reconstructed object with artifacts that were approximately 100 times larger than the true signal. One effect of these artifacts is a high dc level which can be removed. Fig. 3b shows the result of having done this. There still remain sufficient artifacts to render the man indiscernible. Since these artifacts are related to the convolution of the source with the pyramid in Fig. 2a, the reconstruction cannot be improved without *a-priori* knowledge of the source.

Fig. 3c shows the improvement attainable by using the balanced correlation method. Some improvements can be made to this result, but it still represents an upper limit on noiseless reconstruction quality.

The result of using a URA system with balanced correlation to image the man is shown in Fig. 3d. Had the simulation of the URA been performed under the same noiseless conditions as those leading to Figs. 3b and 3c, the reconstruction would have been a perfect reconstruction of Fig. 3a. The faint noise in the background of Fig. 3d is due to the quantum uncertainty included in this simulation.

Signa -to-Noise Ratio

The SNR will be defined as a function of the position in the reconstructed object. If O is the original source distribution and $\hat{O}$ the reconstruction, then

$$SNR(i,j) = \frac{E(O_{ij})}{[VAR(\hat{O}_{ij})]^{1/2}} \tag{7}$$

where $E(O_{ij})$ is the expected true value for the ij-th point in the reconstructed object and $VAR(\hat{O}_{ij})$ is the variance at that point. Note that Eq. (7) 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.

A more thorough definition of the terms used here as well as the derivation of the following signal-to-ratio expression can be found in Ref. 12 (hereafter referred to as Paper II).

$$SNR_{ij} = \frac{N^{1/2} O_{ij}}{[O_{ij} + I_t + 2B]^{1/2}} \tag{8}$$

where $I_t = \sum_i \sum_j O_{ij}$ and B is the number of detector background "counts" in each element of the encoded picture.

The major improvement resulting from the URA is the elimination of artifacts. In other coded aperture systems a fourth term is present in Eq. (8). This artifact term is proportional to the square of the integrated signal and will almost always dominate the other three terms. Its presence alone is the determining factor in the SNR expression (for correlation decoding methods). Since this term is proportional to the square of the signal and the noise is proportional to the square root of the denominator, the signal strength term in the numerator is canceled out by the artifact term. The result is that non-URA systems have a basic SNR limit which cannot be improved through longer exposure times.

The URA has no artifact term in the denominator, leaving just the three terms shown in Eq. (8). The limiting factor now becomes the I_t term, which may produce sufficient noise to offset the advantage of the increased signal through the many pinholes. This new limiting factor is orders of magnitude less severe than that of the artifact term, and thus many more objects will be amenable for viewing by a URA system than by other coded aperture systems.

In Paper II a figure of merit is obtained for the URA by comparing it to a single pinhole camera. We define F_{ij} as the advantage of the URA over the single pinhole as a function of position in the reconstructed object:

$$F_{ij} = \frac{SNR_{URA}}{SNR_{pinhole}} = \left[\frac{N(O_{ij} + B)}{I_t + O_{ij} + 2B}\right]^{1/2} \tag{9}$$

From this it can be shown that the URA will have a net advantage whenever B (the background noise) is large. In the case of low detector background noise, the improvement of the URA over the pinhole will have a lower limit of approximately $(1/2f)^{1/2}$, where f is the fraction of the field of view which is uniformly filled by the object.

Conclusions

Coded aperture techniques were originally introduced to obtain an improved signal-to-noise ratio (SNR) for low-intensity sources (particularly x-ray sources) while maintaining high angular resolution. An improved SNR can be obtained with the matched process if the emitting object consists of a few bright point sources. However, as the object becomes complex the random array methods no longer give an improved SNR (for example see Fig. 3b).

We have pointed out that the matched process can be improved by just a slight change in the analysis procedure. The balanced correlation method is used with the same recorded picture as the matched process. The balanced correlation method subtracts out the high-contrast inherent background as the reconstruction object is being calculated and thus does not have the object-dependent, high-contrast background characteristic of the matched process. Figure 3c demonstrates the improvement possible by using the balanced correlation method.

The uniformly redundant arrays (URA) offer still further improvements. The URA combines the high-transmission characteristics of the random array with the flat sidelobe advantage of the nonredundant arrays. The high transmission provides a capability to image very low-intensity sources, and the flat sidelobes mean that there will be no artifacts to obscure low-contrast sources.

The simulations show that the URA with shot and background noise produces a much better reconstructed object than the random arrays with no shot or background noise (see Fig. 3d). Furthermore, since there is no limiting SNR set by the artifacts (see Eqs. (8) and (9)) with a longer exposure time one can see smaller and smaller contrast changes in the reconstructed object.

Acknowledgments

We wish to thank H. Andrews and J. Grindlay for independently pointing out the nice properties of pseudo-noise arrays. We also acknowledge helpful discussions and encouragement from R. Blake.

References

1. R. H. Dicke, Astrophys. J. 153, L101 (1968).
2. J. G. Ables, Proc. Astron. Soc. Aust. 4, 172 (1968).
3. H. H. Barrett and G. D. DeMeester, Applied Opt. 13, 1100 (1974).
4. E. E. Fenimore and T. M. Cannon, Applied Opt. 16, (1978) in press.
5. J. W. Woods, M. P. Ekstrom, T. M. Palmieri and R. E. Twogood, IEEE Trans. Nuc. Sci. NS-22, 379 (1975).
6. C. M. Brown, Ph.D. Thesis, "Multiplex Imaging and Random Arrays," U. of Chicago (1972).
7. R. L. Blake, A. J. Burek, E. E. Fenimore and R. Puetter, Rev. Sci. Instr. 45, 513 (1973).
8. M.J.E. Golay, JOSA 61, 272 (1970).
9. D. Calabro and J. K. Wolf, Inform. and Control 11, 537 (1968).
10. F. J. MacWilliams and N.J.A. Sloane, Proc. of IEEE 64, 1715 (1976).
11. C. M. Brown, J. Applied Phys. 45, 1806 (1973).
12. E. E. Fenimore, "Coded Aperture Imaging: Predicted Performance of Uniformly Redundant Arrays," submitted to Applied Optics.

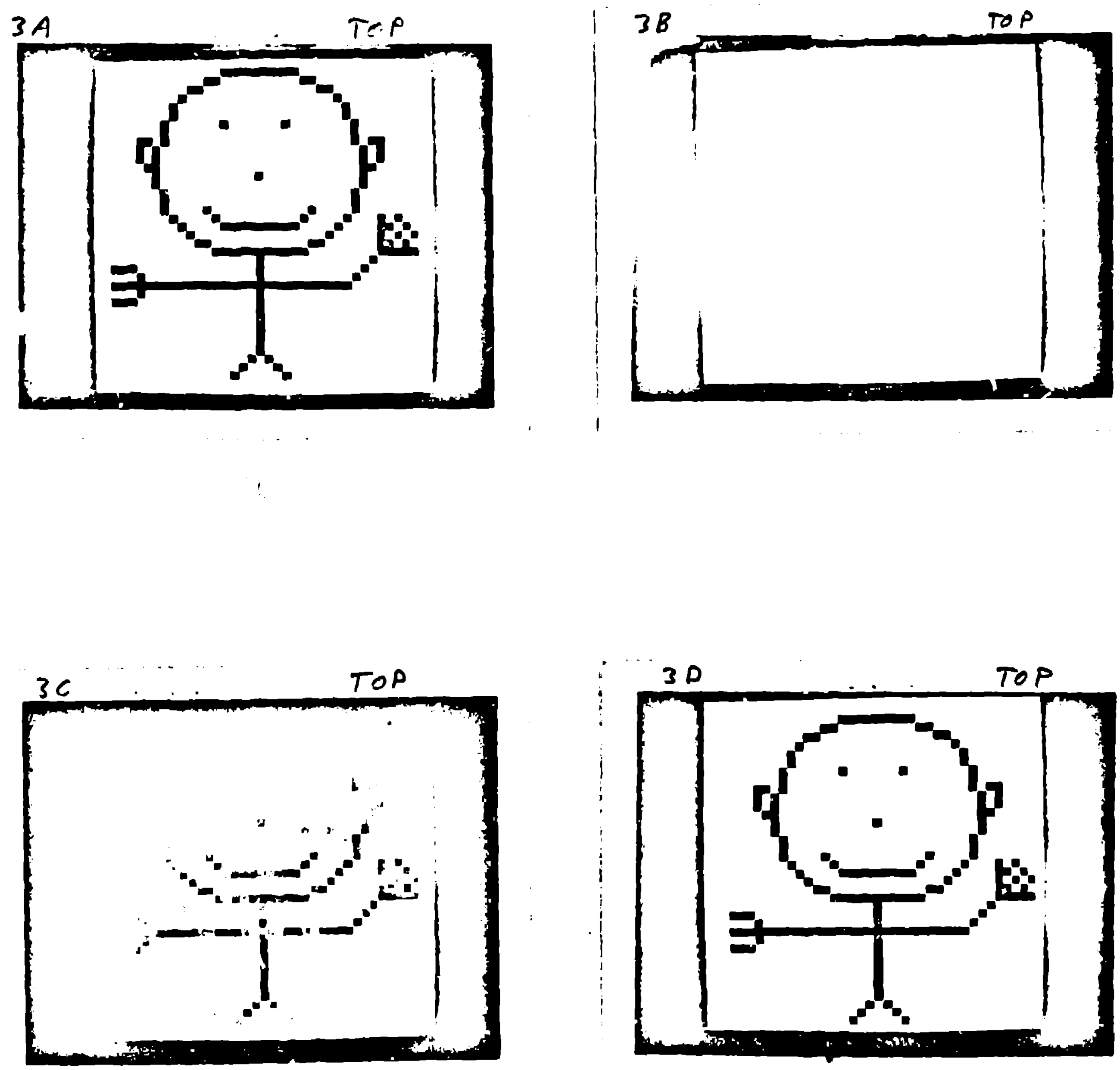


Fig. 3a, b, c, d, will require special handling. Captions on next page.

Fig. 3a. Shown above is the test object used in the computer simulations of this paper. Each point in the man emits 210 photons per second per pinhole.

Fig. 3b. This figure represents the result of having imaged the man (Fig. 3a) through a random pinhole aperture and then having decoded using the matched decoding process. The high background bias, which is signal-dependent, nearly obliterates the man. The simulation was noise-free, hence the bias stems entirely from the nature of the SPSF of Fig. 2a. In some cases in which the distribution of the object is partially known, an attempt could be made to mitigate the bias effects.

Fig. 3c. This figure is the result of having imaged the man through a random aperture and then having decoded using the balanced correlation method. This was a noise-free simulation and hence represents an upper limit on the obtainable image quality.

Fig. 3d. This picture demonstrates the result of having used a uniformly redundant array and the balanced correlation decoding method. Quantum statistics on the source as well as background noise were included in the simulation. Even higher quality is obtainable through longer exposure time.