ELSEVIER

Nuclear Instruments and Methods in Physics Research A 474 (2001) 273–284

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

www.elsevier.com/locate/nima

Optimal coded aperture patterns for improved SNR in nuclear medicine imaging

Roberto Accorsi[a], Francesca Gasparini[b], Richard C. Lanza[a,*]

[a] *Nuclear Engineering Department, Massachusetts Institute of Technology, Room NW13-221, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*

[b] *CeSNEF, Dipartimento di Ingengeria Nucleare, Politecnico di Milano, Milano, Italy*

Received 22 November 2000; received in revised form 15 May 2001; accepted 16 May 2001

Abstract

During a Nuclear Medicine project that called for the optimal design of a coded aperture we found that low-throughput masks do not always provide a Signal-to-Noise Ratio (SNR) advantage. In this paper, we present the simulations of the performance of some coded aperture patterns chosen from different families and compare the results with theoretical predictions. A general expression for the SNR and its particular form for different patterns are provided. The choice of the optimal pattern family is discussed with reference to the characteristics of the object to be imaged and in light of the effect of near-field artifacts. No-Two-Holes-Touching (NTHT) arrays based on Modified Uniformly Redundant Arrays (MURAs) proved to offer the best compromise between SNR performance and practical fabrication constraints.

PACS: 42.15.−i; 42.30.Va; 42.79.Ag

Keywords: Coded apertures; Signal-to-noise ratio; Near-field imaging

1. Introduction

The Signal-to-Noise Ratio (SNR) properties of the Uniformly Redundant Arrays (URAs, [1]) have long been known [2]. Such apertures do not always perform better than pinhole systems, but, for concentrated bright objects in a high background environment, the SNR advantage of a coded aperture can be considerable. Also, for low background, the SNR depends on the open fraction ρ of the coded aperture, larger objects requiring lower ρ. Even if the maximum SNR loss in using a suboptimal open fraction is small, it can be significant in applications, such as Nuclear Medicine, where dose savings are important. Unfortunately, we do not know of any URAs with an open fraction significantly different from 50% [3], but researchers have proposed a number of other coded aperture families [4–13] which extend the attainable range of open fractions while maintaining ideal imaging properties.

In an effort to design a coded aperture system to enhance the resolution of an existing state-of-the-art Anger camera, we considered several of these

*Corresponding author. Tel.: +1-617-253-2399; fax: +1-617-253-2343.

E-mail address: lanza@mit.edu (R.C. Lanza).

0168-9002/01/$ - see front matter
PII: S0168-9002(01)01326-2

families. In particular, we were attracted by some arrays suggested for use in Nuclear Medicine problems, such as ours, which, on the basis of the URA result and by virtue of their low-throughput, were expected to provide an enhanced SNR [8]. In addition to the perfect correlation properties and range of available sizes and open fractions, we were also attracted by the self-supporting design of these arrays. Our preliminary simulations had a surprising outcome: not only did these low-throughput arrays not outperform half-open URAs, but they were also outperformed by other designs.

In this paper we present this result and a theoretical justification of these findings. We close the paper describing the design that we eventually chose for fabrication obtaining considerable SNR and near-field artifact reduction advantages.

2. Preliminary considerations

The goal of our research was to explore the potential of increasing the resolution of a pre-existing Anger camera, a Siemens E-cam, for planar single-photon imaging problems. Potential applications are imaging of organs in children and functional studies in small animals. Current technology relies on high-resolution collimators or pinhole apertures, but these techniques are limited by their sensitivity. The resolution currently available is in the 4–6 mm range. Just like a pinhole, coded apertures can be used in a magnifying geometry to improve resolution. Focusing on small animal imaging problems, we designed the system for a Field of View (FoV) of $9 \times 9\,\mathrm{cm}^2$, and, choosing not to push the limits too early in the project, a resolution of about 1.5 mm.

Many mask properties were investigated: the choice of the number of pixels in the mask pattern, their size and the mask thickness are outside the scope of this paper. Our main concern here is the choice of the mask pattern. Due to the number of mask sizes available, we started focusing on twin-prime [1] and *m*-sequence [4] URAs, MURAs [5], PNP [6], M–P, M–M [7] and the "new system" arrays of Ref. [8], a generalization of those of Refs. [6,7]. All these families (and, within each family, all patterns) have ideal imaging properties.

Since the resolution limit is ultimately set by the size of the mask holes, independent of magnification, we chose to design ~ 1 mm pinholes. Small holes, in turn, set a limit on mask thickness if collimation effects are to be avoided. Under these conditions the use of tungsten in place of lead to achieve lower thickness at constant transparency seemed to be worth the extra cost. By simulation, we estimated that the trade-off between opaqueness and collimation effects was optimum for 1.5-mm-thick tungsten. Small holes in such a material suggested fabrication by etching. The drawback of etched tungsten is that non-self-supporting arrays are difficult to fabricate: we were now bound to use self-supporting patterns. This precluded the use of URAs, of which we know only half-open, non-self-supporting patterns. At first, this did not seem to be completely to our disadvantage. On the contrary, as we shall see in the next section, the objects of our interest are imaged with higher SNR if low-throughput masks are used. Since these masks also tend to be self-supporting, we were hoping to achieve two goals at the same time.

2.1. The signal-to-noise ratio of Coded Aperture Cameras

The most widely recognized technique to calculate the SNR of a Coded Aperture Camera is that of Fenimore [2]. We were interested in imaging ^{99m}Tc. At 140 keV, 1.5 mm of tungsten still allow 1% penetration. To see if this was enough to influence the design, we modified Fenimore's formula to include the effect of mask transparency. With his methods, we obtained the expression

$$\mathrm{SNR}_{ij} = \frac{\sqrt{N_\mathrm{T} I_\mathrm{T}}\sqrt{\rho(1-\rho)}(1-t)\psi_{ij}}{\sqrt{(1-t)[\rho+(1-2\rho)\psi_{ij}]+t+\xi}} \qquad (1)$$

where N_T is the total number of pixels in the mask, I_T is the total number of counts due to the source, ρ the open fraction of the pattern (number of total open pixels of the mask N divided by the total number of pixels N_T), t the transparency of the mask (in our case 1%), ψ_{ij} the fraction of total

source counts due to the source present at the ijth reconstruction position and ξ the ratio of the counts at one reconstruction position due to uncoded background to I_T (background is assumed constant over the detector, so there is no need for subscripts for ξ). This formula differs from Fenimore's by the factor 2 at the denominator. Since all definitions are consistent, we believe this difference to come from the inadvertent substitution, in the original derivation by Fenimore, of a sum over all sources except the ijth with I_T, which includes all sources, causing an overestimate of the variance not relevant for most cases, but with visible effects for large ψ_{ij}. For verification, note that for a point source with neither transparency nor uncoded background the SNR at the source location is $\sqrt{\rho N_T I_T} = \sqrt{N I_T}$. In fact, all the coded aperture does is to count the same source N times and there is no need to justify a $\sqrt{2}$ reduction in SNR.

Taking the derivative of Eq. (1) with respect to ρ yields the optimal open fraction ρ_{opt}

$$\rho_{\mathrm{opt},ij} = \frac{t + \xi + \psi_{ij}(1-2t) - \sqrt{[t+\xi+\psi_{ij}(1-2t)]^2 - (1-t)(2\psi_{ij}-1)[\psi_{ij}(1-t)+t+\xi]}}{(1-t)(2\psi_{ij}-1)}. \quad (2)$$

Fig. 1 shows the result: for low background and objects with a small ψ_{ij}, the optimal fraction is significantly lower than 0.5. These are the expected conditions of Nuclear Medicine measurements, where images of extended objects are taken in low-background environments. Experimentally, we have measured values of ξ in the 10^{-4}–10^{-5} range. To have some sense of the values for the parameter ψ_{ij}, consider a 64×64 image with only 1% of the pixels having some signal. The average ψ_{ij} would be only about 0.02. It is true that if a pixel were brighter than average, its ψ_{ij}—and the optimal open fraction—would be higher, but it is also true that, since the sum of all ψ_{ij}'s is normalized to one, all other points would have an even lower ψ_{ij}. Second, it has been pointed out that even in the case of a single point source, the non-ideal response of the detector must be taken into account: a point source can still be spread over several detector elements and its ψ_{ij} would still be lower than 1 [3]. In conclusion, ψ_{ij} is typically as low as 10^{-2}, possibly 10^{-5} for extended objects or even less for weak regions of the source.

From Fig. 2 we see that the SNR loss for using a half-open pattern instead of an optimal one is higher the lower the background is. In our cases the loss can be as much as 20%. While it is true that this is not much, it still amounts to a factor of 1.5 in exposure time or dose, which can be significant in clinical practice. This seemed to justify some additional effort to find patterns with the best open fraction.

2.2. Pattern choice

For purposes of comparison with literature and previous results obtained by our group, we chose as test object a noisy image of a thyroid phantom (Fig. 3). This case is probably not the best benchmark for a coded aperture, because the object takes about 45% of the FoV so that the average value of ψ_{ij} is 2.6×10^{-4}. ξ was estimated from direct background measurements at 5.5×10^{-4}. With these values we produced, from Eq. (1), Fig. 4, where we have plotted the SNR as a function of the open fraction. For $t = 0.01$ the optimum aperture is 9.42%: we set out to find a pattern about 10% open.

From the size of the detector and its intrinsic resolution, we realized that we needed candidate patterns to have no more than 80×80 pixels. For the arrays of Ref. [8] we found a 77×77 pattern 9.7% open. Mainly for this characteristic, this array family is expressly indicated for Nuclear Medicine applications [8]. For comparison we chose a 79×79 50% open MURA. These patterns are shown in Fig. 5 along with the pattern that was eventually chosen, a 62×62 No-Two-Holes-Touching (NTHT, [9]) pattern based on a 31×31 MURA [5]. Further details on this last pattern are in Section 3.1.

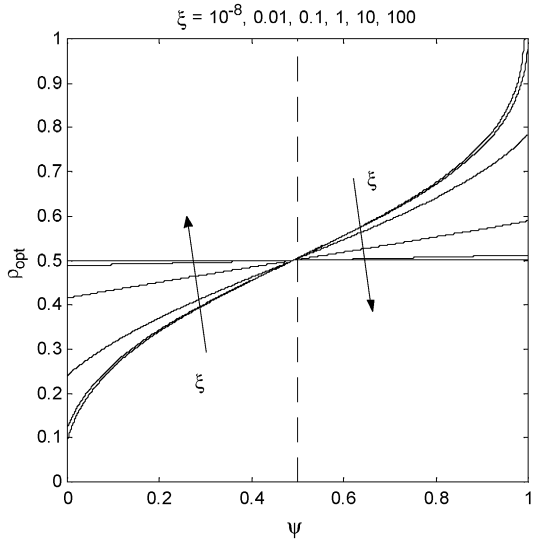

Fig. 1. Optimal open fraction as a function of ψ for different backgrounds for (M)URAs. 1% transparency assumed. Note that, especially at low background, the optimal open fraction for low ψ (extended objects) can be significantly less than 0.5. Due to transparency, for $\xi = 10^{-8}$, $\rho_{\text{opt}} = 1$ before $\rho = 1$.

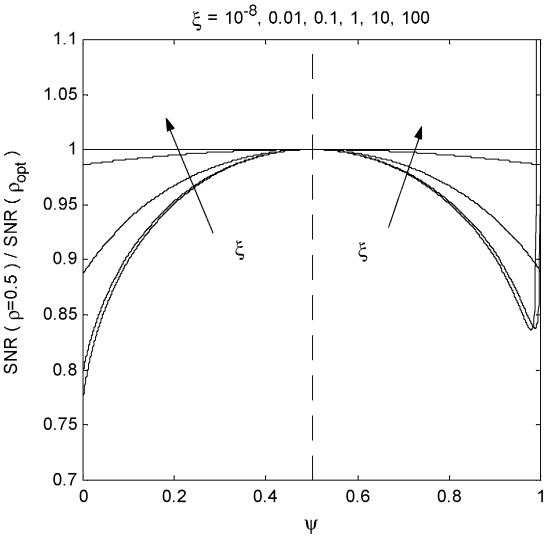

Fig. 2. SNR loss in using the half-open pattern in place of the optimal, for different backgrounds. 1% transparency assumed. As first reported, the loss is never more than 25%, [1]. Due to transparency, at low background $\rho_{\text{opt}} = 1$ before $\psi = 1$, both SNRs go to zero and an asymptote deflects the curves.

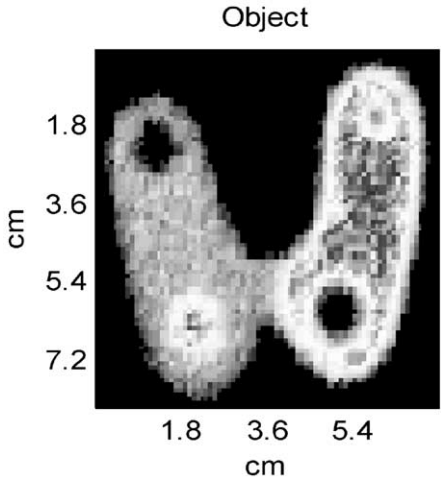

Fig. 3. Test object: average $\psi_{ij} = 2.6 \times 10^{-4}$. Maximum $\psi_{ij} = 3.7 \times 10^{-4}$. This image was taken with a high-resolution collimator and then used as simulation object. Pixel size: 0.9 mm.

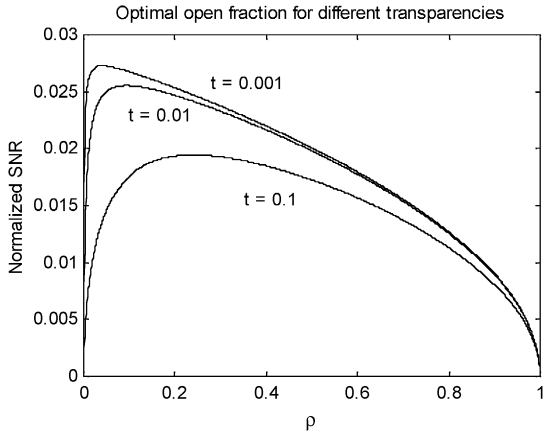

Fig. 4. SNR as a function of the open fraction for different mask transparencies (0.1, 0.01 and 0.001) according to Eq. (1). For the case of our interest, $\psi = 2.6 \times 10^{-4}$, $\xi = 5.5 \times 10^{-4}$ and $t = 0.01$. The result is normalized to the SNR of a pinhole with $t = 0$ (see Table 1) and $\sqrt{N_{\text{T}}}$. Therefore, to have a rough estimate of the SNR advantage for the coded apertures of this paper, one should multiply the ordinates by about 60.

2.3. Simulation results

Our computer simulations take into account several factors: detector intrinsic resolution, discrete detector pixels, near-field geometry, mask thickness and transparency, Poisson statistics,

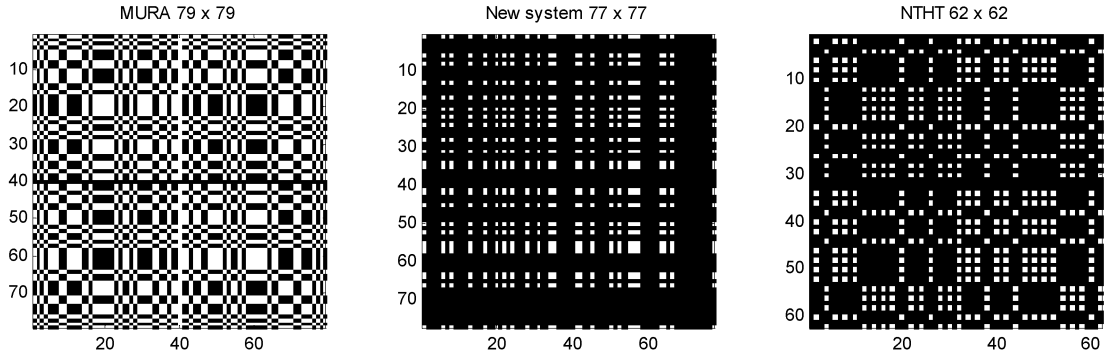

Fig. 5. Coded apertures used for the simulations of Section 2.3. Left: 79×79 MURA. Open fraction: 50%. Mask pixel size: 0.9 mm. Design resolution: 1.1 mm. Center: 77×77 array derived from a 7 and an 11 1-D URA, following the method of Ref. [8]. Open fraction: 9.7%. Mask pixel size: 0.9 mm. Design resolution: 1.2 mm. Right: 62×62 No-Two-Holes-Touching pattern based on a 31×31 MURA, with $e = 2$. Open fraction: 12.5%. Mask pixel size: 1.1 mm. Design resolution: 1.5 mm. For all masks the field of view is 9×9 cm and the base period of a mosaicked pattern is shown. As the thickness was 1.5 mm for all masks, more collimation effects were expected for the masks with smaller pixels and can potentially affect a visual comparison of performance. However, this effect is also compensated for by near-field artifact correction and all images can be fairly compared directly.

quantization noise and ambient background. They have been validated with experimental data from the detector that will eventually be used. We simulated a set of 17.5-min exposures of the test object with 40 cm object-to-detector distance. The activity was 200 μCi of ^{99m}Tc (140 keV).

On the basis of Section 2.1, given the low average ψ_{ij} of the test object, and that the open fraction of the "new system" array was chosen as close as possible to the optimum value, we expected this coded aperture to perform better than a half-open URA. To our surprise, the simulations predicted a much poorer performance for the "new system" array, to the point that other patterns had to be considered. The results of the simulations are shown in Fig. 6.

Coded apertures have ideal imaging properties in the far-field case only, i.e. when the modulation of the projection of the mask onto the detector due to varying incidence angles is negligible. In near-field applications, as is ours, this is not true, and, as one moves the object closer to the detector to improve geometric efficiency, incidence angles and, thus, artifacts increase. We have developed a means of suppressing such artifacts. Since these are different for different masks, we present the result before and after artifact suppression. This technique is outside the scope of this paper, and is worth mentioning here only as long as it is necessary to a fair comparison of the images. Artifact suppression is achieved by adding two images, one taken with a mask and one with its anti-mask (the mask obtained by replacing open with closed holes and vice versa). Since suppression seems equally good in the three cases, the difference must be attributed to the SNR properties of the arrays. Best results are achieved for the half-open MURA pattern, closely followed by the NTHT pattern, while the performance of the "new system" array is clearly worst, despite the lower open fraction. It must be said that the anti-mask picture taken to correct the artifacts present in the "new array" image was very noisy, making the SNR of the corrected image even worse. However, it is clear that our considerations also apply for the uncorrected images, so that the poor performance cannot be blamed exclusively on the correction procedure.

A theoretical investigation of the SNR of the "new system" arrays was needed to confirm these results. Before moving to its exposition, we close this section noting that the symmetry of the pattern is of special interest in artifact reduction and influenced the choice of the pattern as much as SNR considerations. In fact, if an anti-symmetric mask is used, one can rotate the mask, instead of using a different one, to take the anti-mask picture, with cost and space savings. A related technique

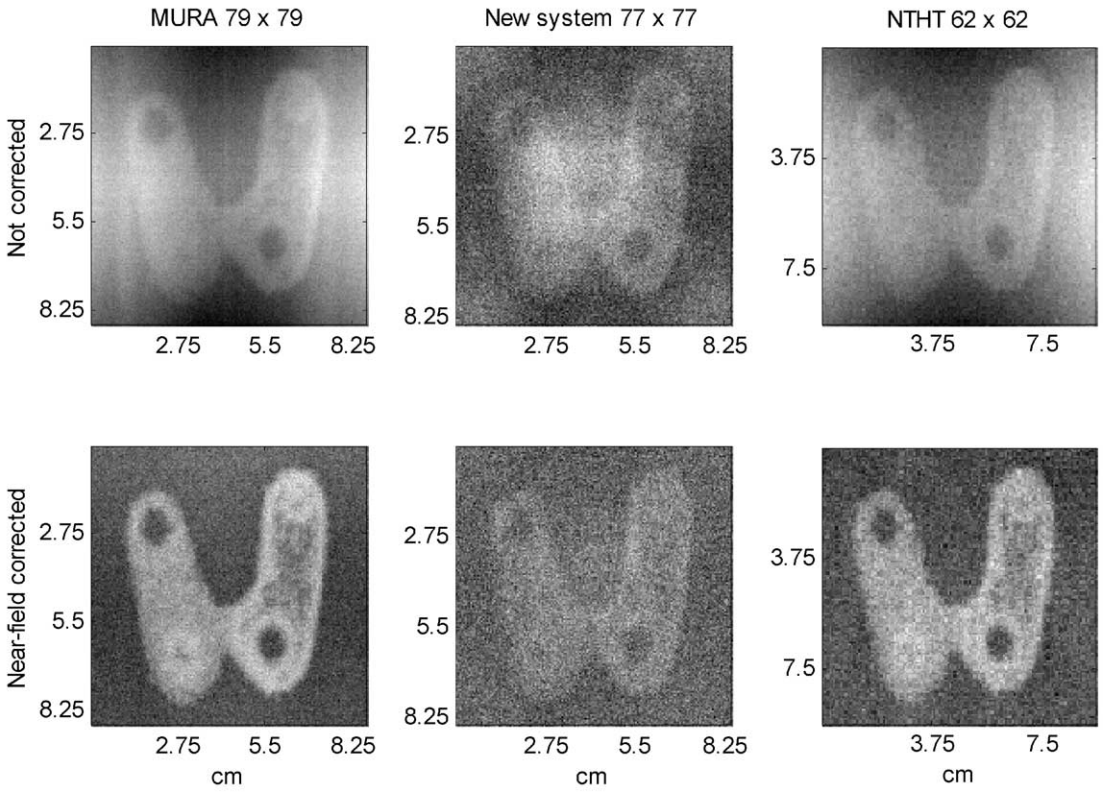

Fig. 6. Simulation results with and without near-field artifacts suppression. All corrected images enjoy a double exposure time, but the comparison is consistent across the three mask patterns. The "new system" corrected image is particularly noisy. This is mainly due to the second picture taken (not shown). However, the SNR advantage of the other two images is also evident from single exposures.

has been used in the past for the reduction of non-uniform background [14–16]), but we believe this to be the first time that it was ever applied to near-field artifact correction. We did not choose the largest NTHT pattern available (74×74, design resolution 1.2 mm) because it is not anti-symmetric and we would have had to fabricate two masks.

3. General formulation of the SNR for coded aperture cameras

The results above do not mean that the SNR is an unreliable figure of merit nor that Eq. (1) is wrong. Rather, it should be recalled that Eq. (1) was derived for URAs only and, thus, it should have not been applied to other patterns. Indeed the SNR of PNP arrays, a family strictly related to the "new system" arrays, was found, under particular conditions (no mask transparency and zero background), to compare unfavorably with URAs [6]. We needed to generalize Eq. (1) so that it would be applicable to other array families. At the same time, we wanted to investigate the effect of different decoding strategies and of mask transparency. As for the former, several decoding arrays have been proposed in the past. In particular the discussion has focused on matched and balanced decoding [1]. These two strategies can be shown to lead to a linear rescaling of the image. From a practical point of view, this does not affect image quality, because it is equivalent to a contrast and brightness adjustment usually overrun by display equipment. From a theoretical point of view, this is important because any linear transformation of the decoding array is a valid decoding array, which makes two constants available for optimization. Skinner and Ponman

[17] have compiled a list of desirable image properties (vanishing background, vanishing sidelobe level, total-count conservation, peak value equal to peak height above its own pedestal, statistical independence of pixels) and derived the conditions for the constants to satisfy one or more of these constraints. We point out that in SNR calculations it is convenient to set two such constants so that

$$\mathbf{A}\otimes\mathbf{G} = N\delta \tag{3}$$

where $\mathbf{A}$ is the mask, $\mathbf{G}$ the decoding array and $\otimes$ the periodic correlation operator. With this definition one can define the SNR as

$$\mathrm{SNR}_{ij} = \frac{\sqrt{I_\mathrm{T}}(1-t)N\psi_{ij}}{\sqrt{(1-t)\sum_{k,l}\psi_{kl}\sum_{u,v}\mathbf{A}_{kl}\mathbf{G}^2_{ij} + (t+\xi)\sum_{u,v}\mathbf{G}^2_{ij}}} \tag{4}$$

where u and v are the variables of all functions and the pairs (k, l) and (i, j) indicate, respectively, the shift of the mask pattern associated with the source reconstructed at (k, l) and the decoding position.

It is important that $\mathbf{G}$ meets condition (3) so as not to introduce the variance of perfectly flat pedestals (which do not involve any noise) at the denominator of Eq. (4). With this caveat, the SNR expression is independent, as it should, of linear rescalings of the decoding array. From Eq. (4), we see that, once a pattern $\mathbf{A}$ and its decoding array $\mathbf{G}$ are given, all we need to calculate its SNR are $\sum_{u,v}\mathbf{A}_{kl}\mathbf{G}^2_{ij}$ and $\sum_{u,v}\mathbf{G}^2_{ij}$. Except for the considerations on linear transformations and mask transparency, the same formula was reached by Gottesman and Schneid with a heuristic derivation [6].

In Table 1 are summarized the data relative to the patterns discussed in this paper. We will not report the details of the calculations for all arrays. As an example, we will present the simple but important case of the pinhole and that of the NTHT because it is non-trivial, it was the pattern we eventually chose to fabricate and we are not aware of a similar result in the literature.

3.1. The SNR of the pinhole

A pinhole aperture can be considered as the limiting case of a coded aperture. In fact, since $\mathbf{A}$ is a δ-function, the choice $\mathbf{G} = \mathbf{A}$ is readily seen to satisfy Eq. (3) where, obviously, $N = 1$. In this case, then

$$\sum_{u,v}\mathbf{G}^2_{ij} = 1 \quad \text{and} \quad \sum_{u,v}\mathbf{A}_{kl}\mathbf{G}^2_{ij} = \delta(i-k, j-l) \tag{5}$$

which leads to

$$\mathrm{SNR}_{ij} = \frac{\sqrt{I_\mathrm{T}}(1-t)\psi_{ij}}{\sqrt{(1-t)\psi_{ij} + t + \xi}} \tag{6}$$

in complete agreement with a direct calculation and Refs. [2,6] providing an additional validation of Eqs. (3) and (4).

3.2. NTHT (M)URAs

Following the rules of Refs. [1,5] only about half-open (M)URAs can be generated. This limitation can be overcome by inserting a number $e-1$ of opaque columns (rows) between all columns (rows) of a (M)URA, to form NTHT (M)URAs [9]. The mask pattern now looks like a square grid of zeros in which we placed the elements of the original array (see also Fig. 5). This is the same as substituting each hole with a smaller one surrounded by mask material. ρ can only be smaller than that of the originating array. Since what matters to self-correlation properties is mainly the spacing of the holes of a URA and not their shape [18], one can think of using a non-integer e as well, but for simplicity we assumed e integer, greater than 1. The maximum non-trivial ρ is then obtained for $e = 2$ and is about 0.125. Lines of zeros have to be inserted accordingly in the decoding array. In decoding, when $\mathbf{G}$ is shifted so that its blank lines cover the non-blank lines of $\mathbf{A}$, the result of the multiplication is zero. For all other shifts, the result is the same as a normal (M)URA case, and one will have a peak and a sidelobe value which depend on the type of decoding adopted. So, if care is not taken that the sidelobe value be zero as well, $\mathbf{A}\otimes\mathbf{G}$ may well not be a δ function. The coefficients used in balanced decoding [1] are then forced in the

Table 1
Expressions needed for the calculation of the SNR via Eq. (4) for the arrays discussed in this paper. For NTHT (M)URAs N_T^0 is the total number of positions of the original array and $N_T = e^2\ N_T^0$ is still the total number of positions in the pattern. See text for e

Array family	$\sum \mathbf{A}_{kl}\mathbf{G}_{ij}^2$	$\sum \mathbf{G}_{ij}^2$	SNR
(M)URA	$\frac{\rho N_T}{1-\rho}\left[\rho + \delta_{i-k,j-l}(1-2\rho)\right]$	$\frac{\rho N_T}{1-\rho}$	$\frac{\sqrt{N_T I_T}\sqrt{\rho(1-\rho)}(1-t)\boldsymbol{\psi}_{ij}}{\sqrt{(1-t)\left[\rho + (1-2\rho)\boldsymbol{\psi}_{ij}\right] + t + \xi}}$
Product arrays (PNP, MP, MM, new system)	N	N_T	$\frac{\sqrt{N_T I_T}\rho(1-t)\boldsymbol{\psi}_{ij}}{\sqrt{(1-t)\rho + t + \xi}}$
Negative product arrays	$N_T\frac{\rho}{(1-\rho)^2}\left(\rho^2 - \frac{1}{N_T^2} + \frac{2}{N_T^2}N\delta\right)$	$\frac{N_T}{(1-\rho)^2}\left(\rho^2 + \frac{1-2\rho}{N_T^2}\right)$	$\frac{\rho(1-\rho)\sqrt{N_T I_T}(1-t)\boldsymbol{\psi}_{ij}}{\sqrt{(1-t)\left(\rho^3 - \rho/N_T^2 + 2\rho(1-\rho)/N_T\boldsymbol{\psi}_{ij}\right) + (t+\xi)\left(\rho^2 + (1-2\rho)/N_T^2\right)}}$
NTHT (M)URA	$\begin{cases} 0 & \text{for } N_T^0(e^2-1) \text{ shifts} \\ N & \text{for } N_T^0 \text{ shifts} \end{cases}$	N_T^0	$\frac{\sqrt{N_T I_T}\sqrt{\rho/2}(1-t)\boldsymbol{\psi}_{ij}}{\sqrt{(1-t)\rho + t + \xi}}$
Pinhole	$\delta_{i-k,j-l}$	1	$\frac{\sqrt{I_T}(1-t)\boldsymbol{\psi}_{ij}}{\sqrt{(1-t)\boldsymbol{\psi}_{ij} + t + \xi}}$

non-zero positions of **G**. For a MURA these are -1 and $+1$ [5].

Of course, this structure influences the quantities of Table 1, which now depend on the shift of **G** relative to **A**, in particular on whether or not the blank lines of **G** overlap with the non-blank lines of **A**. The second term of the variance in Eq. (4) is easily calculated from the decoding coefficients

$$\sum_{u,v} \mathbf{G}_{ij}^2 = N_{\mathrm{T}}^0 \tag{7}$$

where N_{T}^0 is the total number of positions of the original (M)URA array. The first term is more complicated

$$\sum_{k,l} \boldsymbol{\psi}_{kl} \sum_{u,v} \mathbf{A}_{kl}\mathbf{G}_{ij}^2 = N \sum_{k,l}^{N_{\mathrm{T}}^0} \boldsymbol{\psi}_{kl} \tag{8}$$

where the summation on the right-hand side indicates a sum over pixels of the image reconstructed for shifts of **G** which superimpose its blank lines to those of **A**. If the source has no particular structure, $\sum_{k,l}^{N_{\mathrm{T}}^0} \boldsymbol{\psi}_{kl}$ is simply the normalized activity of a $1/e^2$ fraction of it. One can substitute this sum over a partial number of elements with a sum over all elements

$$\sum_{k,l} \boldsymbol{\psi}_{kl} \sum_{u,v} \mathbf{A}_{kl}\mathbf{G}_{ij}^2 = \frac{N}{e^2} \sum_{k,l}^{N_{\mathrm{T}}} \boldsymbol{\psi}_{kl}. \tag{9}$$

To avoid confusion between the density of the (M)URA and of the NTHT pattern we carried out the calculation only for the case of 50% open original arrays, which, anyway, are the only ones we know. The open fraction of the NTHT array now depends on e only: $\rho = (2e^2)^{-1}$. With a little work, the result of Table 1 is reached.

One potential disadvantage of NTHT patterns is that of all patterns for which the square of the decoding coefficients is not constant over the positions. Gottesman and Schneid [6] pointed out that this leads to a dependence of the SNR on the different points of the reconstruction. To quantify it, we performed the test they proposed: given an object made of two point sources, we keep one fixed and measure the variation of the SNR at this source as we move the other continuously from one pixel to a first neighbor. Let the shift be denoted by R. Measured in pixels, it ranges from 0, for no shift, to 1. In Fig. 7 the dependence of the SNR is evident, but is not as dramatic as for geometric arrays [13,6]. It should be stressed that this test is performed with point sources and no background, conditions that maximize the effect. Moreover, we have already pointed out that in real cases even point sources spread over a few pixels, depending on how the detector samples the mask shadow. For NTHT arrays, if a source covers an $e \times e$ square of pixels, no dependence is found at all. Our simulations and experimental results showed no structures due to this SNR dependence.

3.3. Comparing the performance of different arrays

From the formulae of Table 1 we produced the graphs of Fig. 8. Here, as well as analytically, it can be recognized that there is no optimum ρ for "new system" and NTHT arrays for any combination of ξ and $\boldsymbol{\psi}_{ij}$. For both families, higher the ρ, the better. Also, for $\rho < 0.5$, a half-open (and even more an optimally open) URA will always perform best. However, this is not so for $\rho > 0.5$, where "new system" arrays have an SNR higher than that of all other arrays. Starting from the encouraging observation that the negative (i.e. an array where open and closed elements are swapped) of a good array—decoding array pair (**A**, **G**) is again a good pair, we tried using the negative of a "new system" array. This is indeed an array of very high open fraction (90.3%) and its

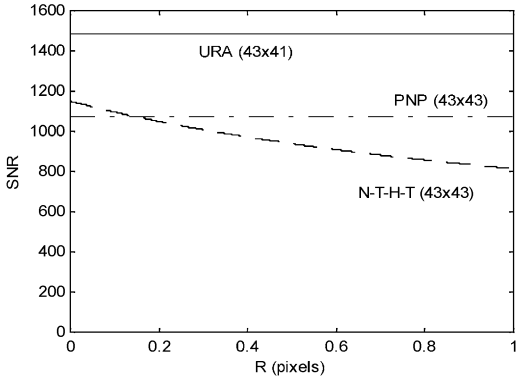

Fig. 7. Dependence of the SNR on the position of sources. $I_{\mathrm{T}} = 10^4$ and $\xi = 0$ as in Ref. [6]. For high backgrounds and non-point-like sources the dependence is not as strong.

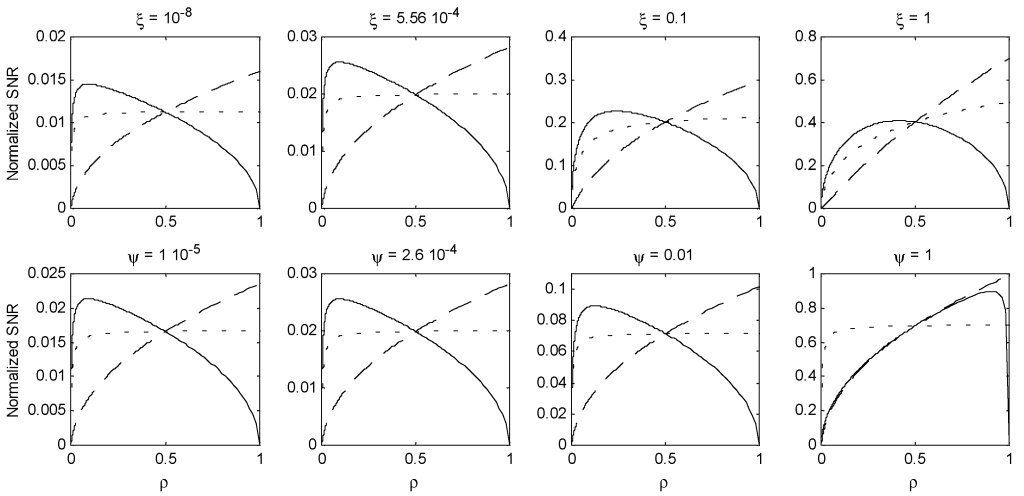

Fig. 8. Comparison of the SNR of different masks for $t = 0.01$. The top row is drawn for $\psi = 2.6 \times 10^{-4}$. The bottom row is drawn for $\xi = 5.56 \times 10^{-4}$. The continuous line refers to (M)URAs, the dotted to NTHT arrays and the dashed to the "new system" arrays of Ref. [8]. The SNR is normalized as in Fig. 4: the advantage over the pinhole increases with ξ and ψ.

study is also relevant to our near-field artifact correction method. Unfortunately, the SNR formula applicable to negative patterns is not the same as that of the original family (see Table 1). It turns out that in this case, the SNR for the negative array is about 2 times worse than that of the original.

It is also important to note that the curves of Fig. 8 are not concerned with the actual existence of patterns. If one sketched the same curves of Fig. 8 for existing arrays only, the situation would be that of Fig. 9, which gives an explanation of our choice of the pattern. Note that for NTHT patterns a valid negative pattern is obtained by inverting the elements of the original (M)URA pattern only. For existing 50% open patterns, this does not change the open fraction, so that the SNR is the same of the positive NTHT pattern.

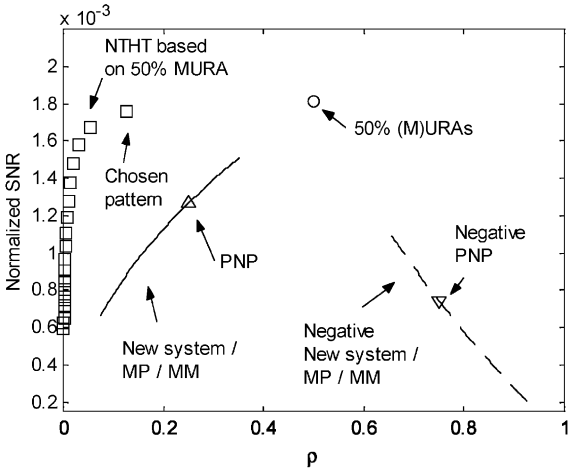

Fig. 9. SNR as a function of ρ for existing arrays. The SNR is normalized as in Fig. 4. For this graph $t = 0.01$, $\psi = 2.6 \times 10^{-4}$ and $\xi = 5.56 \times 10^{-4}$, as applies to the case study. The negative of a NTHT array has the same SNR as its positive. For the NTHT mask of Fig. 5 the advantage over the pinhole is about 1.2 in SNR terms (1.4 in time or activity) for this "average" image pixel. The PNP and negative PNP array singled out are those simulated in Section 2.

3.4. When does a coded aperture pay off?

Given the number of parameters involved, a number of comparisons can be made. In general the result is that, as already commonly accepted, coded apertures are favored for high background and concentrated sources. We were interested in a quantitative comparison between the pinhole and the NTHT mask we built. For simplicity, in the following we assume $t = 0$. From Table 1, the coded aperture has an SNR higher than the

pinhole if

$$(2 - N_T\rho)\xi \leqslant N_T\rho\psi - 2\rho. \quad (10)$$

Since the smallest URA is a 5×3 array with 8 open positions, the smallest open fraction is $\rho_{min} = 8/N_T$, so

$$\xi \geqslant \frac{\rho N_T}{2 - \rho N_T}\psi_{ij} - \frac{2\rho}{2 - \rho N_T}. \quad (11)$$

This inequality can be interpreted as the separation of the (ψ, ξ) plane in two regions. The right-hand side is the family of straight lines passing through the point $(2/N_T, 0)$. To each ρ corresponds a different line, identified by its intercept $2\rho/(\rho N_T - 2)$. Since ρ must be greater than ρ_{min} but less than $\rho_{MAX} = 0.5$ (URA case), not all the lines of the family have a physical meaning. In Fig. 10 the solid lines represent these limiting cases. The line for $\rho = \rho_{opt} = 0.125$ is the one for which an actual NTHT array gives best performance. An NTHT coded aperture will have an SNR higher than the pinhole depending on its parameters (N_T and ρ) and those of the image(ψ, and ξ). The coded aperture is favored for all points of the image lying above this line. First, the coded aperture offers an advantage for all points such that $\psi_i \geqslant 2/N_T$, for all ξ. Second, the higher the background, the better the performance of the coded aperture, for all ψ_{ij}. Third, for increasing ρ, the intercept drops expanding the region of coded aperture better performance, in agreement with the results of Fig. 8.

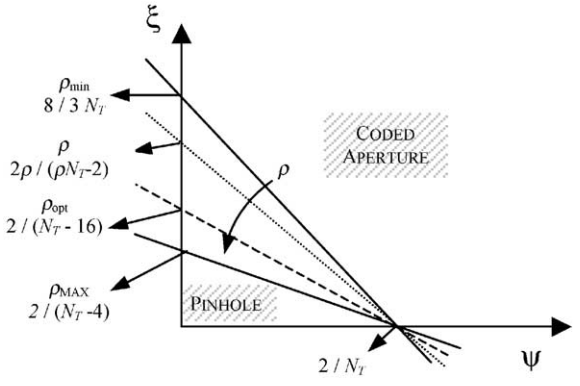

Fig. 10. $\psi - \xi$ plane indicating the cases for which a coded aperture offers a SNR higher than the pinhole.

The elementary sources ψ_{ij} of which an object is made can be located on the same graph. Since ξ is constant for all points, any object is represented by a horizontal line going from the minimum (usually 0) to the maximum ψ_{ij}. In Fig. 11 the case of the thyroid case study is shown. The background is high enough to suggest a small advantage for all points. These numbers are consistent with the small advantage seen in practice and in past applications [19,20].

In a more general case, objects that take the whole FoV have an average ψ_{ij} of $\langle\psi\rangle = 1/N_T$: since this is less than $2/N_T$, whether or not there is an advantage over the pinhole is a matter of background. However, if the object fills only a fraction f^2 of the area of the FoV (so that f is the 1D reduction in the FoV), $\langle\psi\rangle = 1/f^2N_T$. For $f < 1/\sqrt{2}$ the advantage is present, independent of background. The smaller f, the higher the advantage: if the object takes about 5% of the FoV (e.g. a $2 \times 2\,cm^2$ area in a $9 \times 9\,cm^2$ FoV), then the SNR advantage would be a factor of almost 3.2, which is a 10-fold reduction in time or dose at constant image quality.

4. Conclusions and future work

Our study was concerned with finding an optimal coded aperture pattern for Nuclear Medicine applications. In this field, typical conditions, low background and sparse objects, are

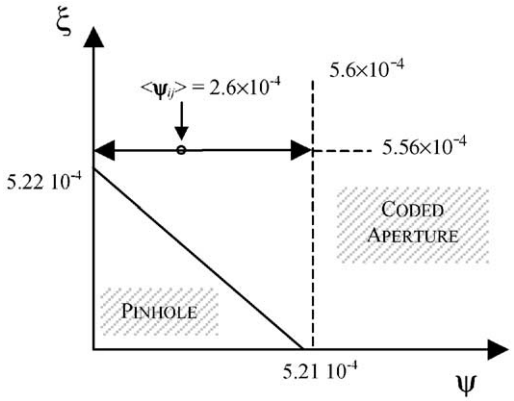

Fig. 11. Application to the thyroid case study. Even if the advantage is small, all image points are imaged with better SNR than they would have been with a pinhole.

those for which the benefits of using an optimal open fraction are the greatest, up to a factor of 1.5 in exposure time or activity. In theory the best SNR is provided by 10–30% open URAs. We know a few 1D URA sequences with low open fraction: these can be folded in 2D arrays, but they are either too sparse, too short or do not provide acceptable pattern sizes [21]. Other low-throughput families have been suggested in literature for specific use in Nuclear Medicine, but both simulation and theory have shown that the theoretical result derived for URAs only does not apply to these new patterns, whose SNR actually decreases with the open fraction. This is also true for NTHT patterns based on (M)URAs, but the reduction in SNR is not as dramatic. We do not know of any pattern that can achieve the SNR predicted for less-than-half-open URAs. The best artifact-free array we know, as far as SNR is concerned, is still a half-open URA. These patterns, however, are not self-supporting and, in cases like ours, cannot be used. Self-supporting NTHT patterns present the opportunity of having almost the same SNR offered by half-open URAs.

From the normalized SNR one can calculate that the SNR advantage over the pinhole for the thyroid case study is never large for patterns of size <60. However, the advantage would increase for objects with a higher ψ. This opens two opportunities. First, one can target the application of coded apertures to pathologies that are typically detected by small hot spots on a dark background. A good example of this would be whole body FDG uptake imaging of tumors. Second, if only a small part of the FoV were used, $\langle\psi\rangle$ would increase. This can be achieved by shielding part of the FoV. For instance, in small animal imaging a case of great interest is that of the activity distribution, e.g. in a mouse's head. The animal could be shielded with a lead foil in which is cut an aperture of the size of the head. In this way we are trading a part of the FoV (and, thus, of the detector) for increased SNR, an opportunity not offered by pinholes. For this case, preliminary experiments have showed that it is indeed possible to image with a resolution <1.5 mm (FWHM) and preliminary theoretical investigation suggests a potential advantage of 10 in exposure time. The resolution limit has not been pushed yet: at present, more aggressive designs expected to reach sub-millimeter resolution seem to be achievable.

In the near future we intend to carry out experimental trials to confirm the opportunities predicted by theory.

Acknowledgements

The authors would like to thank Professor Berthold Horn of the Artificial Intelligence Laboratory and the Department of Electrical Engineering and Computer Science of the Massachusetts Institute of Technology for his careful review of the manuscript. This material is based on work supported by the Office of National Drug Control Policy under Contract No. DAAD07-98-C-0117 and by the Federal Aviation Administration under Grant No. 93-G-053.

References

[1] E.E. Fenimore, T.M. Cannon, Appl. Opt. 17 (1978) 337.
[2] E.E. Fenimore, Appl. Opt. 17 (1978) 3562.
[3] J.J.M. in't Zand, J. Heise, R. Jager, Astronom. Astrophys. 288 (1994) 665.
[4] T.M. Cannon, E.E. Fenimore, Appl. Opt. 18 (1978) 1052.
[5] S.R. Gottesman, E.E. Fenimore, Appl. Opt. 28 (1989) 4344.
[6] S.R. Gottesman, E.J. Schneid, IEEE Trans. Nucl. Sci. NS-33 (1986) 745.
[7] K. Byard, Nucl. Instrum. and Meth. A 322 (1992) 97.
[8] K. Byard, Nucl. Instrum. and Meth. A 336 (1993) 262.
[9] E.E. Fenimore, T.M. Cannon, Appl. Opt. 20 (1981) 1858.
[10] L.E. Kopilovich, Opt. Commun. 68 (1988) 7.
[11] L.E. Kopilovich, G.S. Sodin, Monthly Notices Roy. Astronom. Soc. 266 (1993) 357.
[12] A.B. Giles, Appl. Opt. 20 (1981) 3068.
[13] A.R. Gourlay, N.G. Young, Appl. Opt. 23 (1984) 4111.
[14] U.B. Jayanthi, J. Braga, Nucl. Instrum. and Meth. A 310 (1991) 685.
[15] P.P. Dunphy, et al., Nucl. Instrum. and Meth. A 274 (1989) 362.
[16] M.L. McConnell, et al., IEEE Trans. Nucl. Sci. NS-29 (1982) 155.
[17] G.K. Skinner, T.J. Ponman, Monthly Notices Roy. Astronom. Soc. 267 (1994) 518.
[18] E.E. Fenimore, Appl. Opt. 19 (1980) 2465.
[19] K.F. Koral, W.L. Rogers, G.F. Knoll, J. Nucl. Med. 16 (1975) 402.
[20] K.F. Koral, et al. J. Nucl. Med. 20 (1979) 345.
[21] W.J. Wild, Opt. Lett. 8 (1983) 247.