

## Task performance based on the posterior probability of maximum-entropy reconstructions obtained with MEMSYS 3

K.J. Myers<sup>1</sup> and K.M. Hanson<sup>2</sup>

<sup>1</sup>Center for Devices and Radiological Health, FDA  
HFZ-142, 12720 Twinbrook Parkway, Rockville, MD 20857

<sup>2</sup>Los Alamos National Laboratory, MS P940, Los Alamos, NM 87545

### ABSTRACT

We have described previously<sup>1,2</sup> how image reconstruction algorithms can be evaluated on the basis of how well binary-discrimination tasks can be performed using the reconstructions. The test statistic in the detection task was the estimated activity within the object, also known as the non-prewhitening matched-filter output. This approximation to the likelihood function was used because a full characterization of the posterior probability function had not yet been performed.

A more complete approach is possible when the reconstruction procedure is founded on the Bayesian method. In this case the reconstruction is chosen to maximize the posterior probability and task performance involves using the posterior probability of the various alternatives as the decision variable. This full Bayesian approach should lead to optimal results because the posterior probability incorporates the full dependence on the measurements and constraints, yet is based on the relatively simple likelihood and prior probability distributions.

The commercially available code MEMSYS 3 provides Bayesian image reconstructions based on an entropy prior.<sup>3</sup> This paper details a method of image reconstruction evaluation based on the full posterior probability ratio, and describes results obtained using images derived from the MEMSYS 3 algorithm for a simple binary detection task. Our results demonstrate the improvement in detection performance that can be achieved when the full posterior probability function is incorporated into the decision variable.

### 1. INTRODUCTION

The usual purpose of a medical imaging system is to provide visual information to a radiologist for interpretation. Evaluation of the quality of a medical image should therefore quantitate how well the image conveys to the radiologist the information needed for the prescribed task. Such a quantitation can not be done simply on the basis of visual inspection. One method for deriving a figure of merit that describes how well the image enables the observer to perform some task is to perform human-observer studies using well-known psychophysical methods. In such studies, the observer is asked to perform some visual task, and through training develops some unknown decision strategy for doing so. Human observer performance can be nearly optimal for signals that are exactly known and embedded in white noise.<sup>4</sup> However, when the signal is not known exactly or the noise is correlated, the efficiency of the human observer relative to a machine observer can degrade markedly.<sup>5</sup>

A second method for evaluating image quality is to model a machine observer performing the visual task in place of the human observer. In this manner we have complete control over the decision strategy employed by the machine algorithm. Many trials can be performed for images simulated under one particular set of imaging system conditions to obtain a statistically significant measure of image quality. In this way any aspect of the imaging system can be evaluated and optimized, be it the data-taking hardware or the reconstruction algorithm.

In this paper we present a method for evaluating image reconstructions based on binary decisions made by a machine observer that evaluates the posterior probability ratio of the underlying hypotheses. This is the optimal decision strategy according to Bayesian theory. The images being evaluated were reconstructed from simulated CT projection data using a commercially-available code called MEMSYS 3. This code yields Bayesian reconstructions based on an entropy prior. A Monte Carlo simulation of the complete imaging process, including the generation of a set of object scenes, followed by data-taking, reconstruction, and performance of the specified task, has been used as the backbone of this image-evaluation research.<sup>1,2</sup>

## 2. IMAGE RECONSTRUCTION

The imaging problem we shall consider is the reconstruction of a 2D image from a data set of 1D projections. We assume that  $M$  individual projection measurements are acquired of the unknown scene  $\mathbf{f}_o$ , with the boldface character denoting a vector. We consider the measurements to be linear in the data and containing additive noise, so that the data vector for the  $i^{\text{th}}$  projection can be written:

$$\mathbf{g}_i = \mathbf{H}_i \mathbf{f}_o + \mathbf{n}_i \quad , \quad i=1, \dots, M, \quad (1)$$

where  $\mathbf{H}_i$  is the corresponding row of the measurement matrix and  $\mathbf{n}_i$  is the noise vector for that projection.

The objective of any reconstruction algorithm is to find an estimate of the scene that produced the measurement values. The Bayesian approach to image reconstruction maximizes the posterior probability of the reconstruction given the data. Maximizing the posterior probability is equivalent to minimizing the negative logarithm of the posterior probability  $\phi(\mathbf{f})$ , which can be written as

$$\phi(\mathbf{f}) = \Lambda(\mathbf{f}) + \Pi(\mathbf{f}) \quad , \quad (2)$$

where  $\mathbf{f}$  is an estimate of the scene. The first term in Eq. 2 represents the likelihood of a particular reconstruction given the data. When the noise is modelled as additive Gaussian, we can write this likelihood as

$$\Lambda(\mathbf{f}) = (1/2) (\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{R}_n^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) \quad , \quad (3)$$

where  $\mathbf{R}_n$  signifies the noise covariance matrix. Eq. 3 represents the misfit between the reconstruction and the detected data values as measured in the data domain. The likelihood is proportional to  $\chi^2$  in the usual way:

$$\Lambda(\mathbf{f}) = (1/2) \chi^2 \quad . \quad (4)$$

The second term in  $\phi(\mathbf{f})$ , which we have denoted by  $\Pi(\mathbf{f})$ , comes from the prior probability distribution. This term represents the known characteristics of the scene, including any constraints. While the likelihood term tells how well some  $\mathbf{f}$  matches the data, the prior term tells how well a reconstructed scene is consistent with prior knowledge about the object class.

Under certain assumptions about the object (positivity, additivity), several authors<sup>6-8</sup> have argued that the appropriate prior probability distribution is the entropy, defined to be

$$S(\mathbf{f}) = \sum_j f_j - m_j - f_j \log((f_j)/(m_j)) \quad , \quad (5)$$

where  $m_j$  is called the default value for  $f_j$ . The entropy prior requires the reconstruction to be positive through the logarithmic term in Eq. 5. The global maximum of  $S(\mathbf{f})$  occurs when  $f_j=m_j$  for all  $j$ , so that  $S(\mathbf{f})=0$ .

As the relative weight of the prior is not known beforehand, we introduce a scaling parameter  $\alpha$  so that the prior and likelihood terms combine to form a minimization problem on  $\phi(\mathbf{f})$ :

$$\text{Minimize } \left\{ \phi(\mathbf{f}) = -\alpha S(\mathbf{f}) + \frac{1}{2} \chi^2 \right\} \text{ with respect to } \mathbf{f} . \quad (6)$$

The resulting solution is the maximum a posteriori, or MAP, reconstruction  $\hat{\mathbf{f}}$ . The parameter  $\alpha$  is critical in this and other reconstruction methods employing regularization. Larger values of  $\alpha$  imply solutions with greater emphasis on the entropy term, while smaller values of  $\alpha$  indicate reconstructions more closely fitted to the data.

The maximum entropy algorithm employed here has been described by Skilling.<sup>9</sup> This algorithm can be run in either of two modes. In the "classic" mode,  $\alpha$  is adjusted to allow the solution to move away from the model  $\mathbf{m}$  only to the extent justified by the data, given their assumed accuracy. Stated another way, the classic method varies  $\alpha$  until it reaches its most probable value, where the number of good measurements, as determined by the noise in the data, is equal to a dimensionless measure of structure in the image, given by  $-2\alpha S(\mathbf{f})$ .<sup>3</sup> This is in contrast to "historic" maximum entropy, where the user specifies an aimed-for  $\chi^2$  value, indicating the desired agreement between the reconstruction and the data, and the algorithm maximizes the entropy under that constraint.

The results we presented at this meeting last year were derived from images reconstructed in the historic mode.<sup>2</sup> We showed how  $\alpha$  could be optimized to obtain the best discrimination results based on the estimated activity within a region of interest, also known as the non-prewhitening matched filter. Similarly, maximum a posteriori or MAP reconstructions for a Gaussian prior for the object class have been presented by Hanson.<sup>10</sup> This study also showed the effect of varying the regularization parameter on the final reconstruction, and the evaluation of the images was again based on a non-Bayesian decision variable.

### 3. IMAGE RECONSTRUCTION EVALUATION

In the previous section we described the maximum a posteriori method for reconstructing an object from its projection data. In Section 3.1 we describe the Monte Carlo method used in the simulation of the entire imaging chain. Section 3.2 presents the method for formulating the Bayesian decision variable based on the full posterior probability function. Section 3.3 describes the figure of merit used to summarize the Bayesian observer's task performance.

#### 3.1 Monte Carlo Simulation

We have chosen a Monte Carlo method for image evaluation because through it we can derive a statistically meaningful measure of image quality based on task performance for images that contain either random noise due to quantum fluctuations, or artifacts due to insufficient sampling, or both. The Monte Carlo method for image evaluation has been described in detail by Hanson,<sup>1</sup> and will only be summarized here.

In this study the object class consists of a set of 10 scenes, each containing many randomly placed, non-overlapping disks on a zero background. Specifically, each scene contains 10 low-contrast (amplitude=0.1) disks and 10 high-contrast (amplitude=1.0) disks. The high- and low-contrast disks are all

8 pixels in diameter in an overall field of 128 pixels in diameter. One sample scene from the ensemble is shown in Figure 1a. The task is the detection of the low-contrast disks. The high-contrast disks are placed in the scene to probe the effect of object-dependent artifacts.

The measurement data consist of a number of parallel projections equally spaced over  $180^\circ$ , each containing 128 samples. The noise in the data is assumed to be additive, zero-mean Gaussian. These simulated noisy projection data become the input to the MEMSYS 3 reconstruction algorithm. We have investigated two measurement geometries. In the first case, the system consisted of 100 parallel projections, with noise standard deviation in the projections equal to 8. The second study used only 8 views, with a noise standard deviation equal to 2.

The reconstruction algorithm requires specification of two additional input parameters along with the projection data. The first is the model value in each pixel,  $m_j$  in Eq. 5. Since the randomly placed disks in the scene are equally likely to occur in any location, the underlying model is a uniform grey level set equal to the average pixel brightness in the scene. The second input parameter required is the assumed noise in the input data. The reconstruction algorithm employs a pre-smoothing filter that is a 3-pixel FWHM triangular window, reducing the rms noise in the data by a factor of 0.484. Hence, the assumed noise in the data we input to the algorithm was half the rms noise before smoothing. The purpose of the pre-smoothing filter is to avoid over-fitting the high-frequency components in the input data that can result in ringing. For this imaging geometry and noise amplitude we have found the number of iterations used by the MEMSYS 3 for the reconstruction of each scene to be less than 30. A sample reconstruction of the scene shown in Figure 1a is presented in Figure 1b for the 100-view geometry.

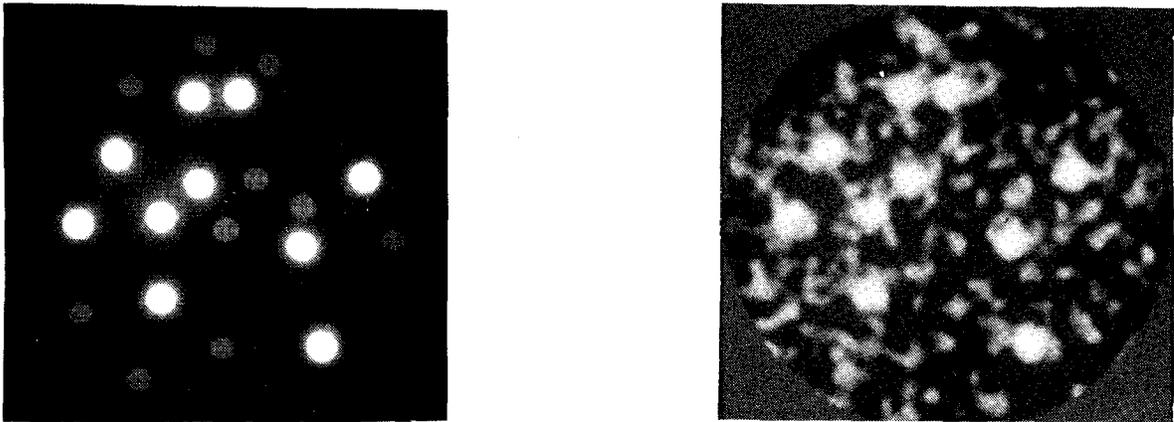


Figure 1. On the left is a sample input scene, with 10 high-contrast (amplitude=1.0) and 10 low-contrast disks (amplitude=0.1) on a zero background. On the right is a reconstruction of the scene from data consisting of 100 parallel projections with  $\sigma_n=8$ .

The first step in the Monte Carlo process is the generation of a random scene containing the disk objects. Simulated projection data for that scene are then input to the reconstruction algorithm, and an estimate of the scene is obtained. Finally, the task is performed by the machine observer, as described in Section 3.2. This process is repeated for each of 10 scenes so that an accurate estimate of the statistics of the task performance can be computed. From those statistics, a figure of merit is determined that summarizes the usefulness of the imaging system, including the reconstruction algorithm, for the defined task. The figure of merit we use as a summary measure is the detectability index  $d_A$ , described more fully in Section 3.3.

### 3.2 Bayesian Strategy for Task Performance

The task we have chosen for the evaluation of these images is the detection of the low-contrast disks. At last year's SPIE meeting<sup>2</sup> we presented results where the observer was a non-prewhitening matched filter. The non-prewhitening matched filter is the optimal observer (in the Bayesian sense) when the disk signal and background are known exactly and the noise is uncorrelated and Gaussian. This observer bases its decision on the likelihood ratio of the two alternative hypotheses, effectively forming the decision variable  $\lambda$  given by

$$\lambda = \sum_j (f_j \cdot \hat{f}_j) \quad , \quad (7)$$

where  $f_j$  is the expected disk amplitude at pixel  $j$ . The test statistic is therefore the estimated activity in the region of the disk.

More recently, we presented discrimination results for an observer that had uncertainty about the objects being discriminated. The observer first estimated parameters describing a model of the object under the two hypotheses.<sup>11</sup> The parameters were fit to minimize the mean-square difference between the reconstructed region and the model for the object under each hypothesis. The decision variable was the difference between the mean-square error for the two hypotheses:

$$\begin{aligned} \lambda &= |f_1 - \hat{f}|^2 - |f_2 - \hat{f}|^2 \\ &= \sum_j [(f_1)_j - \hat{f}_j]^2 - \sum_j [(f_2)_j - \hat{f}_j]^2 \quad , \quad (8) \end{aligned}$$

where  $f_1$  is the object specified by the best-fit parameters to the model under hypothesis 1, and similarly for  $f_2$ .

The difference in mean-square errors is the proper Bayesian decision variable when the fluctuations in the image values are independent, stationary, and Gaussian distributed. However, this is often not the case in our reconstructed images. Each test region in the reconstruction consists of a set of  $N$  pixels and can therefore be represented as one point in an  $N$ -dimensional space. The pixels have random values due to the quantum fluctuations in the incoming radiation stream as well as any random, scene-dependent artifacts. Therefore, the set of reconstructions gives a "cloud" (the posterior probability distribution) in the  $N$ -dimensional space. It is correlations in that  $N$ -dimensional space that the Bayesian observer accounts for. The locations where the disk is present in the scene give one cloud in the  $N$ -dimensional space, and the locations where the disk is absent yield a second cloud. The Bayesian observer uses all information about the distribution of these clouds to choose the optimal hypersurface to separate them. It is these correlations in the image fluctuations that the Bayesian observer accounts for, in addition to the usual spatial correlation in the original scene that defines a disk.

The Bayesian observer bases its decision on the posterior probability ratio of the image estimate under the two hypotheses. The Bayesian observer's decision variable requires knowledge of the full probability density function on the image when the disk is present and absent. We are now able to calculate the full posterior probability ratio for some hypothesized object given the reconstruction based on the entropy prior assumed for the scene, using Eq. 6.

The final reconstruction minimizes the log posterior probability ratio given in Eq. 6. Near the solution  $\hat{\mathbf{f}}$ , the log posterior probability of some arbitrary object  $\mathbf{f}$  can be calculated using Eq. 6 and the calculational simplicity found by expanding in a Taylor series about the minimum, giving:

$$\begin{aligned} \phi(\mathbf{f}) &\simeq \phi(\hat{\mathbf{f}}) + \frac{1}{2} [\mathbf{f} - \hat{\mathbf{f}}]^T [\alpha(\text{Diag}(\hat{\mathbf{f}}))^{-1} + \mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H}] (\mathbf{f} - \hat{\mathbf{f}}) \\ &\simeq \phi(\hat{\mathbf{f}}) + \frac{\alpha}{2} \sum_j \frac{(\mathbf{f}_j - \hat{\mathbf{f}}_j)^2}{\hat{\mathbf{f}}_j} + \frac{1}{2} [\mathbf{H}(\mathbf{f} - \hat{\mathbf{f}})]^T \mathbf{R}_n^{-1} [\mathbf{H}(\mathbf{f} - \hat{\mathbf{f}})] \end{aligned} \quad (9)$$

For a reconstruction  $\hat{\mathbf{f}}$ , Eq. 9 gives the log posterior probability of some test object  $\mathbf{f}$ . No linear term is present in the expression because it is an expansion about a minimum. Eq. 9 is in essence a Gaussian approximation to the log posterior probability ratio with two covariance terms. The second term in the expression represents the variability in the object class represented in the entropy prior. The third term looks like  $\chi^2$ , but here it is relative to the projection of  $\hat{\mathbf{f}}$  rather than the data since this is the basis we have chosen for this quadratic approximation.

After  $\hat{\mathbf{f}}$  has been determined by the reconstruction algorithm, taking the difference of the log posterior probability for two test objects is equivalent to computing the posterior probability ratio of the two hypothesized objects. The Bayesian decision variable is thus formed:

$$\lambda = \phi(\mathbf{f}_1) - \phi(\mathbf{f}_2) \quad , \quad (10)$$

where  $\mathbf{f}_i$  is the test object for hypothesis  $i$ . This Bayesian decision variable is fundamentally different from the decision variable represented by Eq. 8 because of the covariance terms contained in Eq. 9.

While we cannot visualize the full N-dimensional cloud given by the posterior probability distribution, we can get some idea of its behavior by considering the joint probability of pixels taken two at a time. Figure 2 is a contour plot of the posterior probability for two pixels in the region of a disk. This plot was obtained by varying each pixel's value about its reconstructed value and using Eq. 9. Figure 2 graphically shows the posterior probability cloud centered on the reconstruction values, and in particular, shows the correlation in the posterior probability distribution. (The pixels are 3.6 pixels apart.) It is this correlation that is accounted for by the Bayesian observer that uses the full posterior probability distribution.

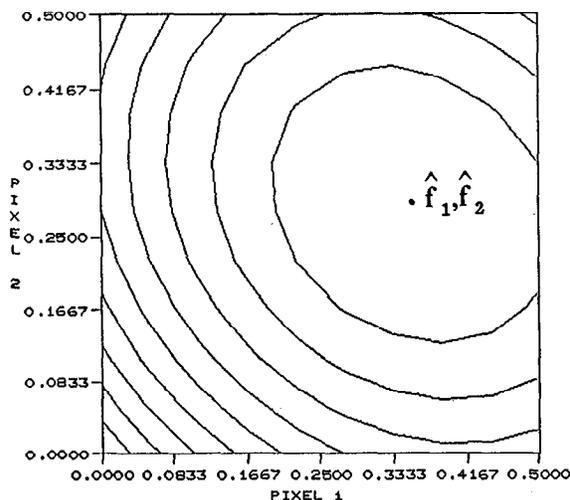


Figure 2. Plot of the joint posterior probability of two pixels in the region of a disk in the reconstruction. The pixels are 3.6 pixels apart.

In the results section we shall present detectability indices for four decision strategies:

- 1) An observer that fits the background and disk amplitude (both constrained to be non-negative) by minimizing the mean-squared error, and uses the estimated disk contrast (activity above the background) as a decision variable. This strategy is similar to the non-prewhitening matched filter, except that the background is assumed to be unknown to the observer.
- 2) An observer that fits the background and disk region (both constrained to be non-negative) using the posterior probability (minimizing Eq. 9) and uses the estimated disk contrast (activity above the background) as a decision variable.
- 3) An observer that forms the posterior probability ratio for two test objects, disk present and absent, with the background assumed to be zero in each case. (Eq. 9 is evaluated for two test objects, disk amplitude=.1 and disk amplitude=0., both on a zero background, and the difference in the results is the test statistic.)
- 4) An observer that fits the background (constrained to be non-negative) to form two test objects, disk present (amplitude=.1) and disk absent (amplitude=0.) by minimizing Eq. 9 under each hypothesis. The difference in log posterior probabilities determined from Eq. 10 is the decision variable.

### 3.3 Figure of Merit for Task Performance

The machine observer in our Monte Carlo simulation is presented with 100 locations where the disk is actually present (10 disks in 10 scenes) and 100 locations where the disk is absent. The observer calculates its decision variable and declares the signal to be present, a positive response, if the decision variable is above its decision threshold. By applying this strategy to each location where the signal is actually absent, and again where the signal is actually present, we can derive histograms of the decision variable under both the signal-present and signal-absent conditions. The traditional receiver operating characteristic (ROC) curve is obtained by plotting the fraction of true-positive responses versus the fraction of false-positive responses from the histograms as the observer's decision threshold is varied. The area under the ROC curve  $A$  can then be used to compute the detectability index  $d_A$ , according to the formula:

$$d_A = 2 \operatorname{erf}^{-1}\{2(1-A)\} \quad , \quad (11)$$

where  $\operatorname{erf}^{-1}$  is the inverse error function. The area under the ROC curve has the value 0.5 when the decision variable histograms completely overlap, yielding a detectability index of 0. When there is complete separation of the two histograms, the area under the ROC curve is 1.0, and the corresponding value of the detectability index in this limit is  $\infty$ .

Figure 3 shows another technique for visualizing the decision-variable space for this problem. The axes are the log posterior probability calculated for the two hypotheses, signal present and signal absent. This particular plot was constructed with the Bayesian observer having full knowledge of the signal and background (strategy 3 in the previous section). Each symbol corresponds to a pair of log posterior probability values calculated by the Bayesian observer at each test location, one for each hypothesis. We show two classes of points on the graph - the horizontal dashes correspond to locations where the disk is actually absent, and the vertical dashes are locations where the disk is actually present. The Bayesian observer takes the difference between the log probability values to decide whether to call the disk present or absent. Therefore, one particular decision threshold is represented as a line on this graph of slope 1, with intercept given by the threshold value. The ROC curve is obtained by calculating the true-positive and false-positive fractions as the decision threshold is varied, corresponding to varying the intercept of the discriminant function, which is the line of slope 1. Once the area under the resulting curve is determined, the detectability index is given by Eq. 11.

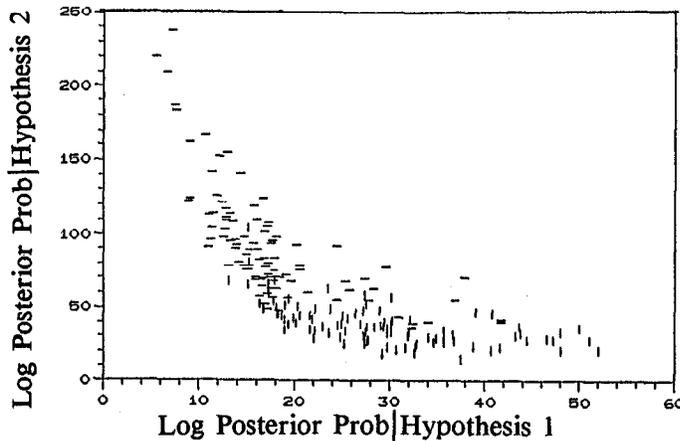


Figure 3. Plot of pairs of log posterior probability values, one for each hypothesis, for each test region in the reconstructions. Horizontal dashes indicate locations where the disk is actually absent and vertical dashes indicate locations where the disk is actually present.

We have deliberately chosen not to use the index  $d'$  as the figure of merit for task performance. As discussed by Wagner *et al.*,<sup>12</sup>  $d'$  is a useful figure of merit only when the histograms of the decision variable under each hypothesis are Gaussian. The decision-variable distributions seen in Figure 3 are very much non-Gaussian, so  $d'$  is not an appropriate figure of merit for this case.

#### 4. RESULTS

Table 1 summarizes the results for the 100-view case with rms noise of 8. The table contains two columns for two different values for  $\alpha$  (corresponding to different strengths for the regularization term). The column with the smaller value of  $\alpha$  ( $\alpha=0.25$ ) corresponds to reconstructed images obtained from the classic mode of MEMSYS 3, in which  $\alpha$  is chosen by the algorithm. The average rms residual for the 10 scenes in this case was equal to 3.79. The second column corresponds to images reconstructed using MEMSYS 3 in the historic mode for  $\alpha=4.28$ , giving an average rms residual in this case of 4.20. The higher values for  $\alpha$  and rms residual in this latter case indicate that these reconstructions are not as close to the maximum likelihood solution as the classic reconstructions.

The rows of Table 1 give detectability indices for four different decision methods, as described in Section 3.2 above. The first row gives the detectability index for an observer that uses the disk contrast as a decision variable, after first fitting the background by minimizing the mean-square error. The last three rows of the table are all results based on the full posterior probability handled in different ways. The second row gives the detectability that results when the estimated disk contrast, based on the full posterior probability distribution, is used as a decision variable. The fitted disk amplitude and background values were constrained to be non-negative. The third row gives the detectability index for a Bayesian observer that made full use of the posterior probability distribution associated with the reconstruction and assumed that the background was zero. The last row gives the detectability index based on the full posterior probability ratio, just as in row 3, but this observer had to estimate the background. The fitted background was based on the full posterior probability and was assumed to be non-negative.

For the classic run, the full Bayesian decision rule based on the posterior probability ratio gave a detectability index of 1.68 with and without background fitting. This value is considerably higher than the detectability index of 1.14 resulting from a decision rule based on the estimated disk contrast using Eq. 9 in the fitting process. However, we found that these decision variables that utilized Eq. 9 in some fashion were outdone by a simple estimate of the disk contrast based on the mean-square error, which gave a  $d_A$  equal to 1.74.

The ranking of decision strategies in the classic run did not hold when  $\alpha$  was increased, as seen from the second column of results for the 100-view case. Now the three results based on decision variables that

use the posterior probability outscore the simple contrast estimate. In addition, every detectability index calculated here is higher than the highest value obtained for the classic maximum entropy reconstructions. The ideal-observer SNR can be calculated for the detection of the low-contrast disks in the data domain, under the usual SKE assumptions and using the formalism presented by Wagner and Brown.<sup>13</sup> The resulting SNR, found to be equal to 2.36, is remarkably close to the best  $d_A$  value determined for the Bayesian observer using the reconstructed images and the posterior probability ratio.

Table 2 summarizes the results for the 8-view,  $\sigma_n=2.0$  case. Results are presented for the classic mode ( $\alpha=.208$ , average rms residual=.542) and one historic-mode run ( $\alpha=.532$ , average rms residual=.499). The trends in the table are similar to those seen in Table 1, that is, the full posterior probability ratio can give better performance than decision variables based on contrast estimates, but for both geometries it is clear that the choice of  $\alpha$  is critical.

Table 1. Results for the 100-view,  $\sigma_n=8$  case.

Test Statistic	$d_A$	
	Classic $\alpha=0.25$	Historic $\alpha=4.28$
Estimated disk contrast (no prior)	1.74	1.85
Estimated disk contrast (entropy prior)	1.14	2.11
Posterior probability ratio (Background known)	1.68	2.13
Posterior probability ratio (Fitted background)	1.68	2.34

Table 2. Results for the 8-view,  $\sigma_n=2$  case.

Test Statistic	$d_A$	
	Classic $\alpha=0.21$	Historic $\alpha=0.53$
Estimated disk contrast (no prior)	1.38	1.46
Estimated disk contrast (entropy prior)	1.36	0.86
Posterior probability ratio (Background known)	0.95	1.87
Posterior probability ratio (Fitted background)	0.95	1.87

## 5. DISCUSSION

We have described how the same posterior probability used in a reconstruction step can be used to form the full Bayesian decision variable based on the posterior probability ratio of two alternative hypotheses. Several interesting questions remain to be answered. While the best discriminability for each imaging system (geometry) was obtained using a decision variable based on the full posterior probability ratio, the choice of  $\alpha$  was crucial to that result. Other values of  $\alpha$  can give images where contrast estimates give a better decision variable. Further investigation into the choice of  $\alpha$  is needed.

Further study into the proper choice for the model and prior are also necessary. The entropy prior assumes the scene is strictly positive, while the background in the original scene was actually zero. This discrepancy leads to a bias in the reconstruction away from the actual background value. We can perhaps see a related effect in the  $d_A$  values given in Table 1 for the historic run. We see that the posterior probability ratio gives best performance when the background has to be fitted. Generally an observer that

has uncertainty, and must therefore perform a fitting operation, does not perform as well as one with no uncertainty. These results show the opposite to be true here. We believe this is because the observer that "knows" the background is zero is handicapped, because in the reconstruction the background values are always greater than zero due to the bias inherent in the entropic prior. The prior is inappropriate for the scene, giving "misinformation" to the observer. The implications of this and the previous paragraph are that we still have much to learn about determining the best decision variable for a given imaging system and task.

The model observer in these studies was presented with only a 14-pixel diameter subregion of the scene centered on the test area. However, the structures found in images reconstructed from projections can have long-range correlation. One simple example is the streak artifacts found in an image reconstructed from a limited number of views. If more of the image had been made available to the observer, some gain in performance might have been manifested, although it is unknown at present how much difference in performance would be found.

It is not currently known how close a human observer's detection performance would come to the detectability indices presented here, especially for the geometry presented with a very limited number of views. Whenever task performance can be shown to differ greatly between human and model observers, it is an indication that the available information is far from optimally presented for the human observer. When that is the case, the images should be either processed in some way to make the information more available to the human observer, or a machine observer should be used instead. We plan to test human performance in evaluating these images in the near future. Performance as a function of number of views and subregion size will be investigated.

## 6.SUMMARY

The commercially available code MEMSYS 3 has been used to form reconstructions from noisy CT projection data. Using a Monte Carlo simulation of the entire imaging chain, we have evaluated the imaging system based on the detectability of low-contrast disks in the reconstructions. Detection indices were reported for four observers: 1) an observer that used the estimated disk contrast as a decision variable after estimating the disk amplitude and background using a mean-square error criterion, 2) an observer that used the estimated disk contrast as a decision variable after estimating the disk amplitude and background by minimizing the log posterior probability, 3) a Bayesian observer who made full use of the posterior probability distribution associated with the reconstruction and assumed a known background, and 4) a Bayesian observer who used the posterior probability ratio as the decision variable and had to estimate the background.

We have found that the MAP reconstructions obtained with the "classic" maximum entropy algorithm are very close to maximum-likelihood images, in that the algorithm relies very heavily on the data to form a solution associated with a very small value for  $\alpha$ . These images are reconstructed in a Bayesian estimation sense, but they do not yield optimum images for this detection task. In fact, simple estimates of disk contrast, perhaps surprisingly, give better detection performance than a decision variable based on the posterior probability ratio. Images with a larger value for the parameter  $\alpha$  than that chosen by the classic algorithm have been found to yield higher detectability indices for the full Bayesian observer.

The Bayesian decision variable based on the posterior probability distribution of a reconstruction can indeed be calculated, as long as the prior and likelihood terms that contribute to the reconstruction solution are known. Improved task performance can then be found, since the posterior probability ratio incorporates the full complexity of the effect of the measurement matrix and any constraints.

## 7. ACKNOWLEDGMENTS

The authors wish to thank Robert F. Wagner for numerous illuminating discussions. We would also like to acknowledge Stephen F. Gull and John Skilling. Many of the ideas presented here were provoked by the manual they wrote to accompany MEMSYS 3. This work was supported in part by the United States Department of Energy under contract number W-7405-ENG-36.

## 8. REFERENCES

1. K.M. Hanson, "Method of evaluating image-recovery algorithms based on task performance," *J. Opt. Soc. Am. A* **7**, 1294-1304 (1990).
2. K.J. Myers and K.M. Hanson, "Comparison of the algebraic reconstruction technique with the maximum entropy reconstruction technique for a variety of detection tasks," *Proc. SPIE* **1231**, 176-187 (1990).
3. S.F. Gull and J. Skilling, *Quantified Maximum Entropy - MEMSYS 3 Users' Manual*, Maximum Entropy Data Consultants Ltd., Royston, England (1989).
4. A.E. Burgess, R.F. Wagner, R.J. Jennings, and H.B. Barlow, "Efficiency of human visual discrimination," *Science* **214**, 93-94 (1981).
5. K.J. Myers, H.H. Barrett, M.C. Borgstrom, D.D. Patton, and G.W. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," *J. Opt. Soc. Am. A* **2**, 1752-1759 (1985).
6. J.E. Shore and R.W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Info. Theory* **IT-26**, 26-39 and **IT-29**, 942-943 (1980).
7. Y. Tikochinsky, N.Z. Tishby, and R.D. Levine, "Consistent inference of probabilities for reproducible experiments," *Phys. Rev. Lett.* **52**, 1357-1360 (1984).
8. S.F. Gull and J. Skilling, "Maximum entropy method in image processing," *IEE Proc.* **131(F)**, 646-659 (1984).
9. J. Skilling, "Quantified maximum entropy," in *Maximum Entropy and Bayesian Methods*, P.F. Fougère, ed., Kluwer (1990).
10. K.M. Hanson, "Object detection and amplitude estimation based on maximum a posteriori reconstructions," *Proc. SPIE* **1231**, 164-175 (1990).
11. K.M. Hanson and K.J. Myers, "Rayleigh task performance as a method to evaluate image reconstruction algorithms," to be published in the proceedings of the Tenth Annual Maxent Workshop, Univ. of Wyoming (July 30-August 3, 1990).
12. R.F. Wagner, K.J. Myers, M.J. Tapiovaara, D.G. Brown, and A.E. Burgess, "Maximum a posteriori detection and figures of merit for detection under uncertainty," *Proc. SPIE* **1231**, 195-204 (1990).
13. R.F. Wagner and D.G. Brown, "Unified SNR analysis of medical imaging systems," *Phys. Med. Biol.* **30**, 489-518 (1985).