The Mutual Information Criterion for SPECT Aperture Evaluation and Design

LINGXIONG SHAO, ALFRED O. HERO, MEMBER, IEEE, W. LESLIE ROGERS, MEMBER, IEEE, AND NEAL H. CLINTHORNE, MEMBER, IEEE

Abstract—**An aperture performance criterion for single photon emission computed tomography (SPECT) is proposed based on the mutual information (MI) between the source and detector processes. The MI is a measure of the reduction in uncertainty of the emitter location given the detector data, and it takes account of the inherent tradeoffs between the effects of sensitivity and resolution on source estimation accuracy.**

Specific expressions for the MI are derived for one-dimensional linear geometries and two-dimensional, parallel slice, ring geometries under the following assumptions: Poisson emission times; uniform emission angles; no scattering; and a known lost-count correction factor. For one-dimensional geometries a necessary and sufficient condition for an aperture to maximize the mutual information is given. Then MI-optimal apertures are derived for various source distributions using an iterative maximization procedure. The MI is then numerically calculated for various ring apertures associated with the parallel slice SPRINT II system.

NOMENCLATURE

T	Acquisition time.
τ_i	Detection time—time the gamma-ray detection occurs.
n	The number of emissions.
t_i	The time the ith emission occurs.
N	Emission process, consisting of n and $\{t_i\}_{i=1}^n$.
X	$X = \{X_i\}_{i=1}^n$ source emitter spatial positions.
Y	$Y = \{Y_i\}_{i=1}^n$ incident gamma-ray positions on the detector.
F	Failure to detect an emission.
W	$W = \{W_i\}_{i=1}^n$, observation symbols. If a gamma-ray detected at Y_i, $W_i = Y_i$, $W_i = F$ otherwise.
$a(x, y)$	Aperture indicator function as a function of x and y.
$a(z)$	Aperture indicator function as a function of position on aperture axis z.
$f_x(x)$	Mean emitter distribution, a probability density.
$f_y(y)$	Fluence distribution at the detector surface, an unnormalized probability density.
$P(F)$	Probability of failure to detect an emitter.
$f_o(y \mid x)$	Conditional probability of y given x without an aperture; a probability density.
$f(y \mid x)$	Conditional probability y given x with an aperture; an unnormalized probability density.
$f_o(z)$	Fluence distribution along the aperture axis.
$f(x \mid w)$	Conditional probability of x given observation w; a probability density.
$H(X)$	Entropy of the source distribution $f_x(x)$.
$H(n)$	Entropy of the Poisson random variable n with rate Λ.
$I(X; W)$	Mutual information between X and W.
$H(X \mid W)$	Conditional entropy of X given observation W.
$q_1(z)$	Information delivered to the detector by a gamma-ray which passes through the aperture.
$q_2(z)$	Information delivered by the gamma-ray failure events F.

Manuscript received November 15, 1988; revised March 31, 1989. This work was supported in part by the National Cancer Institute, DHHS, under PHS Grant CA32856.

L. Shao is with the Bioengineering Program, University of Michigan, Ann Arbor, MI 48109.

A. O. Hero is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109.

W. L. Rogers and N. H. Clinthorne are with the Division of Nuclear Medicine, University Hospital, Ann Arbor, MI 48109.

IEEE Log Number 8928435.

I. INTRODUCTION

THE principal components of gamma-ray imaging systems consist of an aperture constituted of lead or other gamma-ray absorbing material and a position-sensitive gamma-ray detector. Gamma-rays are detected as discrete events which are characterized by the coordinates where the gamma-ray interacts with the detector surface. The average number of detected events per time interval as a function of position will be referred to as the fluence distribution. This fluence distribution corresponds to a projection of the source distribution through the aperture. The purpose of the aperture is to restrict the set of possible emitter locations which can correspond to a gamma-ray's incident position on the detector. This effectively reduces the uncertainty in the possible emitter locations for a given detected event, however, at the expense of decreased fluence.

0278-0062/89/1200-0322$01.00 © 1989 IEEE

For an ideal uniform parallel hole collimator, the FWHM (full width at half maximum) of a point spread function (PSF) provides a measure of intrinsic spatial resolution of the detection system. Likewise, the area under the PSF, which is the average fluence, provides a measure of intrinsic sensitivity of the detection system. The sensitivity can be related to the detection signal-to-noise ratio (SNR) where the PSF plays the role of a signal and the conditional Poisson statistics in the measurement constitute the noise. Sensitivity is an important parameter for nuclear medicine imaging instruments since the administered radioactive dose to the patient has to be minimized and imaging must be restricted to a time interval over which the emission statistics are stationary.

For parallel-hole collimators, resolution and sensitivity are invariant to the spatial position of the point source. It is well known that improved sensitivity can be achieved by using nonuniform "coded" apertures at the expense of "multiplexing" the image data onto the detector [1]–[4]. For these nonuniform apertures the spatial resolution must be calculated from the PSF of the reconstructed image or its modulation transfer function (MTF), which is generally not invariant to source position and depends on the reconstruction algorithm [5], [6]. This system resolution is sometimes called the recoverable resolution. The conventional design strategy has been to find apertures which strike a reasonable compromise between resolution and detection sensitivity.

Although sensitivity and spatial resolution (FWHM) are convenient performance parameters for conventional aperture systems, they do not adequately account for the shape and spatial dependence of the point source response function. In addition, the proper tradeoff between sensitivity and resolution for a given object or class of objects is not simply determined. One way around this problem is to quantify the performance of the SPECT system by way of its accomplishment of specific tasks [7], [8]. This has the disadvantage of depending on the specification of a particular estimation, classification, or detection algorithm. In this paper we propose a mutual information (MI) measure as a yardstick for gauging the effect of the tradeoff between resolution and sensitivity on the transfer of information from the source to the detector. Our approach is similar in concept to the approaches described in [9] and [10]. However, our approach differs from these previous approaches in the following significant ways. First, in [9] and [10] the imaging system was modeled as a linear, spatially invariant channel with additive Gaussian noise statistically independent of the source signal. While this model may be appropriate to the CT and photographic imaging systems considered in the above references, it is inappropriate for the spatially varying channel with signal-dependent non-Gaussian noise which characterizes the SPECT detection process. Second, in [9] and [10] the channel capacity rather than the MI is considered. While the channel capacity provides a source independent measure of information transfer, it provides no indication of the degree to which a particular detection system is matched to a particular source distribution. The MI presented in this paper accounts for the non-Gaussian spatially varying channel characteristics of SPECT. Furthermore, the source independent channel capacity can be obtained from the MI presented here by means of a functional maximization over possible source distributions.

The MI can be put into context by the following comments. First, the MI can be interpreted as a measure of concentration of the *a posteriori* distribution of the emitter locations about the true locations. Second, the MI can be directly related to conventional measures of resolution and fluence associated with the projections. Third, the MI specifies a fundamental limit on the achievable mean square error (MSE) of any estimator of emitter locations through the rate distortion lower bound of information theory.

The main results of the paper are the following.

- An exact analytic expression for the MI is derived under the assumption of Poisson processes, no scattering, and a known count correction factor. The count correction factor is the ratio of the total number of emitted counts to the total number of detected counts. In the case where a count correction factor is not available, the data processing theorem [11] implies that the MI expression which is derived here is an upper bound on the actual MI associated with the reconstruction system. The degree to which the expression derived overestimates the MI is treated in a separate study [12].
- The MI can be put into a form which brings out the dependence of the information transfer from the source to the detector on resolution and fluence. The resolution is manifested in the MI as the conditional entropy of the emission distribution given the detector measurements.
- We derive a necessary and sufficient condition for a particular aperture to maximize the MI for a given source. This condition indicates the MI-optimal aperture should have openings at locations where the information delivered to the detector by passing gamma-rays exceeds the information delivered to the detector by blocking gamma-rays.
- For one-dimensional linear geometries, we numerically maximize the MI and find the MI-optimal apertures for uniform, Gaussian, two-level, and bimodal sources. The results show that a substantial (factor of three) gain in MI can be achieved in some cases by using coded apertures as opposed to parallel hole collimators.
- The robustness of an optimal aperture is studied by using an optimal aperture to image a different source than the one for which the aperture was optimized. The results suggest that the MI-optimal aperture is relatively robust (within 0.3 percent of the optimal MI) to small deviations of the source from the assumed source distribution.
- We derive an explicit form of the optimal MI for parallel slice ring geometries, such as SPRINT II. It is found that for a uniform ten-slit ring aperture and a uniform cylindrical source, the MI is maximized when the aperture

openings have a duty cycle of 43 percent. This duty cycle is similar to that of the optimal coded apertures found in [4] via maximization of signal-to-noise ratio.

In this paper only the cases of infinite detector resolution and ideal 100 percent detector photon collection efficiency are considered. For a treatment of nonideal detector resolution and sensitivity the reader is referred to [13]. The MI analysis in this paper holds for the ideal case of known number of emissions over the observation interval. Hence, the MI introduced here is actually an upper bound on the MI of nonideal detection systems. For a study of the impact of this additional knowledge on the MI, see [12].

The paper is organized as follows. In Section II, a mathematical model for the gamma-ray detection system is described. In Section III, the general form of MI is defined, and its relation to estimation error, resolution, and sensitivity is discussed. In Section IV, the problem of finding MI optimal apertures for a fixed source distribution is considered, and several numerical examples are presented. Finally, Section V concludes with suggestions for future work.

II. Background and Mathematical Model

Two examples of three-dimensional SPECT systems are illustrated in Figs. 1 and 2. In both figures the object is a radioactive source composed of a set of gamma-ray emitters which are randomly located at positions $\{x \in \mathfrak{X}\}$ according to the probability density function $f_x(x)$. Each emitter generates a gamma-ray at a random angle $(\theta, \psi) \in [0, 2\pi] \times [0, \pi]$ which is uniformly distributed, and may be detected at some position $y \in \mathfrak{Y}$ on one of the detector surfaces. In order to reduce the effect of angular uncertainty on the determination of emitter positions from detected events, a perforated lead aperture may be introduced between the object and the detector. The shape, depth, and placement of these perforations determine a set of projections of the source onto the detector. In the parallel-slice ring geometry of Fig. 2, the parallel collimator collimates the gamma-rays in the axial direction, while the aperture may be arbitrarily specified along the ring circumference. This reduces the imaging task to reconstruction of a set of two-dimensional slices from a set of one-dimensional projections taken off the circumference of the detector ring $\{y \in \mathfrak{Y}\}$. The sample distributions of the locations of incident gamma-rays along each detector surface, which are the planes in Fig. 1 and the thin ring in Fig. 2, comprise the projections of the sample distribution of emitters. The mean distributions along each detector surface, called the fluence distributions $f_y(y)$, comprise the projections of the (mean) source distribution $f_x(x)$. Since both the emission and detection processes are random, the mean performance of the detection system is determined by the nature of the joint probability distribution of the emission and detection random processes. In the sequel we focus on the derivation of this distribution and we compute the MI associated with a general detection system.

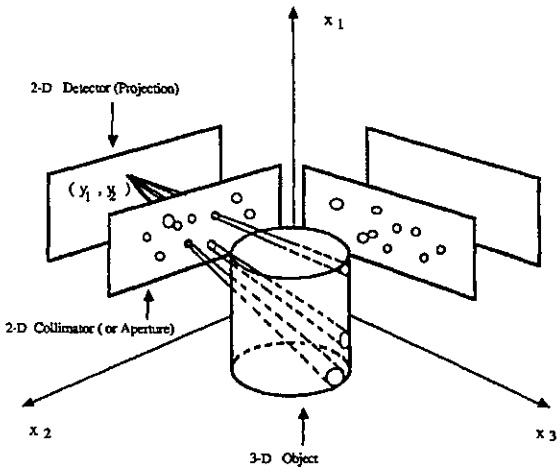

Fig. 1. A generalized rectangular parallel-detector three-dimensional geometry for SPECT.

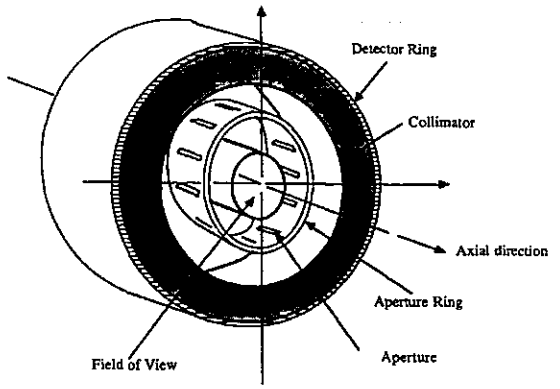

Fig. 2. Ring geometry SPECT system.

The relevant statistical quantities which characterize the process of acquiring data during a finite interval of time $[0, T]$ along the detector surface are: 1) the emission counting process N consisting of the number n of emissions and the emission times $\{t_i\}_{i=1}^n$, $t_i \in [0, T]$; 2) the source emitter spatial position X_i, at time t_i, taking values $x \in \mathfrak{X}$; 3) the position Y_i of an incident gamma-ray on the detector at detection time τ_i, taking values $y \in \mathfrak{Y}$; 4) the failure F of the gamma-ray at time t_i to make it to the detector; 5) the observation symbol W_i, at time τ_i, taking values $w \in \{F\} \cup \mathfrak{Y}$. It can be shown that τ_i can be taken as t_i, $i = 1, \cdots, n$ for the SPECT problem without loss of generality.

We assume the following model for the processes N, $\{X_i\}_{i=1}^n$, and $\{W_i\}_{i=1}^n$: 1) the point process N is inhomogeneous Poisson with intensity $\lambda(t)$ over the data acquisition time interval $[0, T]$; 2) given N, the emission locations $\{X_i\}_{i=1}^n$ are independent identically distributed (i.i.d.) random variables with probability density function (p.d.f.) $f_x(x)$; 3) the gamma-ray emitted from position $X_i = x$ travels along the straight line path $\vec{xy}$ (see Fig. 3)

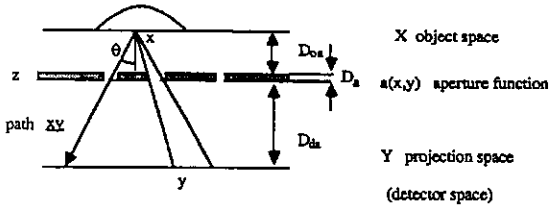

Fig. 3. One-dimensional detector system. The emission process can be described at an emission time $\{t_i\}$ by a line intersecting the source emitter location x and oriented at gamma-ray emission angle θ. The gamma-ray is detected at y if the aperture is open at every point along the segment $\vec{xy}$, i.e., $a(x, y) = 1$, while a failure occurs if the aperture is closed at any point along the segment $\vec{xy}$, i.e., $a(x, y) = 0$.

whose direction is drawn independently from a uniform distribution of angles; 4) if the path $\vec{xy}$ passes through an opening in the aperture, then the observation $W_i = y$, otherwise $W_i = F$ (failure).

Note that the straight line path assumption 3) is valid only for the case where there is no scattering. Note also that assumption 4) requires measurement of the "failures" $W_i = F$ over $[0, T]$. Presently, systems do not incorporate such measurements. On the other hand, it is conceivable that methods for generating lost-count corrections may be incorporated into future instruments. In any case, the MI to be derived under the above model is an upper bound on the MI of any system which does not incorporate lost-count corrections. For a rigorous analysis of the effect of lost-counts on MI, refer to [12].

For convenience, we define the aperture function $a(x, y)$ as follows:

$$a(x, y) = \begin{cases} 1 & \text{the gamma-ray path } \vec{xy} \text{ passes through an opening in the aperture} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Fig. 3 provides an illustration for the one-dimensional case.

In the next two subsections we derive the joint distributions for X_i and W_i for one- and two-dimensional geometries.

A. Joint Distributions: One-Dimensional Geometries

We first consider the case of one-dimensional linear geometries and infinite detector resolution. The type of one-dimensional geometries considered is a mathematical idealization of a single projection in a parallel slice tomographic system. Fig. 3 shows a line source which is parallel to a linear aperture and a linear detector, all lying in the same plane. In the figure, D_{oa} denotes the source–aperture distance, D_{da} denotes the detector–aperture distance, and D_{od} denotes the object–detector distance. For this case x and y become the scalars x and y, respectively, where, conditioned on an emission at time t_i, X_i is distributed according to the mean emitter distribution $f_x(x)$, and, conditioned on emitter location, the emission angle, denoted θ in Fig. 3, is uniformly distributed on $[-(\pi/2), (\pi/2)]$.

Under the above statistical model, several distributions must be derived for the MI calculation.

First, the observation W_i is a mixed random variable whose conditional distribution has a continuous component $f(y \mid x)$ and a discrete component $P(F \mid x)$

$$dP(w \mid x) = \begin{cases} f(y \mid x)\, dy, & \text{if } w \in [y, y + dy]; \\ P(F \mid x) & \text{if } w = F. \end{cases} \tag{2}$$

Referring to Fig. 3, the conditional distribution of Y_i given $X_i = x$ for the case of no aperture, $f_o(y \mid x)$, which will be called the "nominal conditional distribution," is [13]

$$f_o(y \mid x) = \frac{1}{\pi D_{od}\left(1 + \left(\frac{x - y}{D_{od}}\right)^2\right)}. \tag{3}$$

Now with an aperture (1), the unnormalized conditional density $f(y \mid x)$ of y given x is given by

$$f(y \mid x) = a(x, y) f_o(y \mid x). \tag{4}$$

$f(y \mid x)$ is unnormalized since the probability that $W_i = F$ is nonzero. For an aperture of zero thickness (Fig. 4), the aperture function $a(x, y)$ is a function of (x, y) only through the point of intersection z of the path $\vec{xy}$ and the line supporting the aperture

$$z = \alpha x + (1 - \alpha) y$$

or

$$a(x, y) = a(\alpha x + (1 - \alpha) y)$$

where

$$\alpha = \frac{D_{da}}{(D_{oa} + D_{da})}. \tag{5}$$

Second, the detector fluence distribution $f_y(y)$ is obtained by averaging $f(y \mid x)$ over x

$$f_y(y) = \int_{-\infty}^{+\infty} f(y \mid x) f_x(x)\, dx. \tag{6}$$

$f_y(y)$ is unnormalized with mass $1 - P(F)$ since $f(y \mid x)$ is unnormalized.

Third, conditioned on x, the failure probability is the sum of the probabilities that the directed path $\vec{xy}$ intersects the aperture in regions such that $a(x, y) = 0$. This is simply the integral of $[1 - a(x, y)] f_o(y \mid x)$ over y

$$P(F \mid x) = \int_{-\infty}^{+\infty} dy [1 - a(x, y)] f_o(y \mid x). \tag{7}$$

The total probability of getting a failure over the ensemble of possible emitter locations is therefore

$$P(F) = \int_{-\infty}^{+\infty} P(F \mid x) f_x(x)\, dx. \tag{8}$$

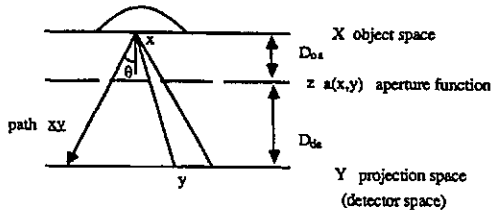

Fig. 4. Same system as in Fig. 3 but with zero thickness aperture. Point of intersection of $\overrightarrow{xy}$ with aperture is at the aperture coordinate $z = \alpha x + (1 - \alpha) y$ where $\alpha = D_{oa}/(D_{da})$.

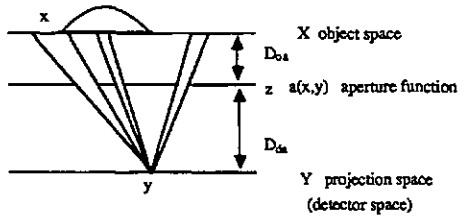

Fig. 5. The *a posteriori* distribution of emitter location given the observation $W = y$ is supported by the set of possible source locations which may pass through the aperture to yield a detector event at y.

Finally, the *a posteriori* distribution of the emitter location $X_i = x$ given a single observation $W_i = w$ is given by Bayes' rule [14]

$$f(x \mid w) = \begin{cases} \dfrac{f(y \mid x) f_x(x)}{f_y(y)}, & w = y; \\ \dfrac{P(F \mid x) f_x(x)}{P(F)}, & w = F. \end{cases} \tag{9}$$

The probability distribution $f(x \mid w)$ is an assignment of probabilities to the possible emitter locations x given w. The concentration or width of $f(x \mid w)$ is a measure of the resolution of the emitter location provided by the observation. Fig. 5 shows the support of $f(x \mid w)$ for $w = y$ and a zero thickness aperture.

B. Joint Distribution: Ring Geometries

In tomographic imaging system design, there are several constraints which may be imposed to reduce the range of aperture configurations. First, the field of view may be limited to a cylindrical region corresponding to the diameter of a human brain or torso. Second, the detector configuration may also take on cylindrical geometry in order that the detector solid angle may be maximized with the minimum detector area. A third constraint, which may be imposed to simplify the reconstruction, is to limit the axial incidence angles so that collimation along the axis of the cylindrical field of view is limited to nearly parallel slices. SPRINT II is a three-dimensional imaging system which incorporates these constraints. The detection system as shown in Fig. 2 consists of a slit-aperture ring, a detector ring, and a field of view inside the ring. In the axial direction the portion of the object which can be imaged is constrained to lie in a single slice by the parallel collimator.

The ring geometry emission/detection model is illustrated in Figs. 6 and 7. Conditioned on an emission at time t_i, the emitter location X_i is distributed over a planar region within the field of view according to the emitter distribution $f_x(x)$, and, conditioned on emitter location $X_i = x$, the emission angle (θ, ψ) is uniformly distributed.

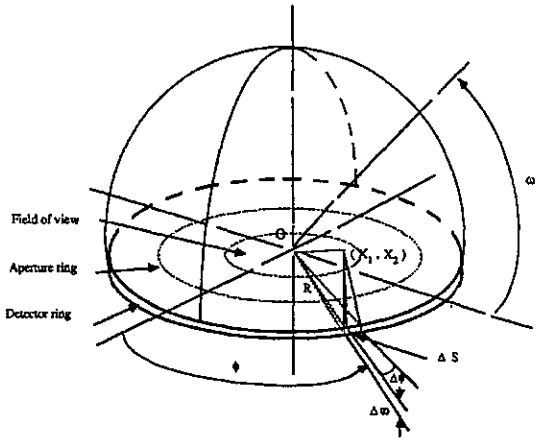

Fig. 6. A planar source is detected via incident gamma-rays along a finite width detector ring of radius R in a ring geometry tomographic system. The chisel shape region indicates the possible gamma-ray paths for which a source at position $x = (x_1, x_2)$ may be detected in the region ΔS along the detector ring.

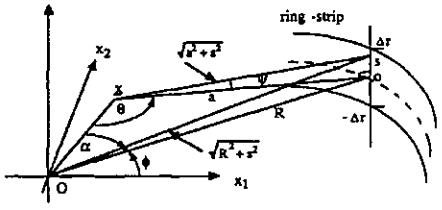

Fig. 7. A blow-up of the chisel shaped region of Fig. 6. The possible positions on the detector ring are parameterized by the coordinates (ϕ, s) relative to the origins O and o, respectively. The emission angles of the emitter at position $x = (x_1, x_2)$ are the uniformly distributed angles (θ, ψ) in the figure.

We consider the ideal case of infinite detector resolution. Since the aperture collimates along the vertical axis of Fig. 6, the possible detection positions are constrained to a narrow ring-strip of width $2\Delta r$, shown in Fig. 7. We parameterize the positions along this ring-strip by the angular variable ϕ, and the lateral position along the inside of the ring by s, relative to the origins O and o in Fig. 7,

respectively. Under the practical assumption that the radius R is much greater than the ring width $2\Delta r$ (the thickness of a slice), we have from (47) in Appendix B

$$f_o(\phi, s \mid x) = \frac{1}{2\pi^2} \frac{R^2 - R(x_1 \cos\phi + x_2 \sin\phi)}{\left((x_1 - R\cos\phi)^2 + (x_2 - R\sin\phi)^2\right)^{3/2}} \tag{10}$$

for $\phi \in [0, 2\pi]$ and $-\infty < s < \infty$. The conditional distribution with an aperture is found analogously to the one-dimensional case

$$f(\phi, s \mid x) = A(x, \phi, s) f_o(\phi, s \mid x) \tag{11}$$

where $A(x, \phi, s)$ is an aperture function which has the value one if there is a path $\vec{xy}$ from x to $y \triangleq (\phi, s)$ and is equal to zero otherwise.

III. Theory of Mutual Information

A. Mutual Information Measures

We briefly introduce the MI concept in this section. More details concerning the mathematical attributes of MI [15], and its role in estimation and detection theory [11], [16], can be found in the references.

Assume there are two random processes $X \in \mathfrak{X}$ and $W \in \mathfrak{W}$ with marginal probability distribution $P_x(x)$, $P_w(w)$ and a joint probability distribution $P(x, w)$. We want to find a quantitative measure of how much the occurrence of a particular event, $W = w$ say, tells us about the probability of some alternative, $X = x$ say. MI is such a measure. The general MI formula is

$$I(X; W) = \int dP(x, w) \log \frac{dP(x, w)}{dP(x)\, dP(w)}. \tag{12}$$

The random processes can be continuous or discrete. For this paper, since we consider infinite resolution detection processes, the case of interest is when one random process is continuous, say X with p.d.f. $f_x(x)$, and the other is a mixture of discrete and continuous, say W, with distribution $f_w(w)$ and $P_w(w)$ over the continuous and discrete components of W. In this case the MI has the form

$$\begin{aligned} I(X; W) = {} & \int_{\text{continuous } \mathfrak{W}} f_w(w) \int_X f(x \mid w) \\ & \cdot \log \frac{f(x \mid w)}{f_x(x)}\, dx\, dw \\ & + \sum_{\text{discrete } \mathfrak{W}} P_w(w) \int_X f(x \mid w) \\ & \cdot \log \frac{f(x \mid w)}{f_x(x)}\, dx. \end{aligned} \tag{13}$$

B. Qualitative Interpretations of Mutual Information

MI can be viewed in terms of entropy (discrete case) or differential entropy (continuous case). Both entropy and differential entropy are measures of the inherent uncertainty associated with the values taken on by a random process. For concreteness we consider the discrete case here. The entropy of a discrete process is defined as [15]

$$H(X) = -\sum_x p_x(x) \log p_x(x) \tag{14}$$

and the conditional entropy is

$$H(X \mid W) = -\sum_{x, w} p(x, w) \log p(x \mid w). \tag{15}$$

Substituting (14) and (15) in (12), we obtain the following classical relation between the MI and the entropies

$$I(X; W) = H(X) - H(X \mid W). \tag{16}$$

Several simple conclusions can be made from (16). $H(X)$ represents the (*a priori*) uncertainty in X and $H(X \mid W)$ represents the (*a posteriori*) conditional uncertainty of X after W is observed. If X and W are independent, then $H(X \mid W) = H(X)$, i.e., there is no reduction in uncertainty of X, which leads to $I(X; W) = 0$. On the other hand, if there is a one-to-one correspondence between X and W, then $H(X \mid W) = 0$, and the MI will be the same as the entropy of X. If W is a noisy observation of X over which one has some control, one can attempt to maximize MI by reducing *a posteriori* uncertainty, i.e., making $H(X \mid W)$ small.

For the case where X is continuous, similar interpretations can be made using a relation of the MI to differential entropy functions.

C. Relation to Mean Square Error

The MI can be related to a lower bound on the mean-square error of the reconstructed emitter locations via rate-distortion theory. Specifically, it can be shown [17] that if X is a random variable and $\hat{X}$ is any estimator of X based on W

$$E(X - \hat{X})^2 \geq \frac{1}{2\pi e} e^{2H(X)} e^{-2\max_{f_x} I(X; W)}.$$

Hence, the limiting case of low MI necessarily implies poor mean square error performance. Similar bounds can be established for a wide class of error measures.

D. Relations to Resolution and Fluence

The application of resolution and sensitivity to system design is rendered difficult due to the fact that resolution and sensitivity are spatially dependent and coupled. In many cases higher resolution can only be obtained at the expense of lower fluence. These factors make system op-

timization difficult to perform based solely on the criteria of resolution and fluence. A systematic approach to imaging evaluation and design is provided by the MI measure, which forms a composite resolution/sensitivity criterion relevant to the inherent estimatibility of emitter position for a given geometry.

Specifically, in view of (16), the system geometry affects the MI only through the conditional entropy $H(X \mid W)$. For a gamma-ray imaging system, the conditional entropy (15) can be expressed as

$$\begin{aligned} H(X \mid W) &= -\int dP(w) \int dx\, f(x \mid w) \log f(x \mid w) \\ &= -\int dy\, f_y(y) \int dx\, f(x \mid y) \log f(x \mid y) \\ &\quad - P(F) \int dx\, f(x \mid F) \log f(x \mid F) \\ &= \int dy\, f_y(y)\, H(X \mid W = y) \\ &\quad + P(F)\, H(X \mid W = F). \end{aligned} \tag{17}$$

In (17), $f_y(y)$ is the fluence distribution along the detector, i.e., the integral of $f_y(y)$ is the total detector fluence, and $H(X \mid W = y)$ and $H(X \mid W = F)$ are the conditional entropy densities given the observation $W = y$ and $W = F$, respectively. The conditional entropy densities are measures of the concentration of the *a posteriori* distributions $f(x \mid y)$ and $f(x \mid F)$ over emitter locations x given the observations y and F, respectively. Specifically, a highly concentrated *a posteriori* distribution has low conditional entropy, while a diffuse distribution has high entropy. Hence, the conditional entropy plays the role of resolution in as far as resolution affects the MI measure. Since $I(X, W) = H(X) - H(X \mid W)$, (17) implies the intuitive result that in order to achieve high MI, the imaging system should be such that fluence density $f_y(y)$ be small at points y along the detector which are associated with inherently low resolution, i.e., points y where $H(X \mid W = y)$ is high.

As an illustration, consider the special case where x and y are scalars and $f(x \mid y)$ is closely approximated by a Gaussian function

$$f(x \mid y) = \frac{1}{\sqrt{2\pi\sigma^2_{x \mid y}}}\, e^{(x-\mu_{x \mid y})^2/2\sigma^2_{x \mid y}} \tag{18}$$

where $\mu_{x \mid y}$ and $\sigma^2_{x \mid y}$ are the conditional mean and variance of x given y. For this case the conditional entropy can be calculated [11]

$$H(X \mid W = y) = \tfrac{1}{2} \ln (2\pi e \sigma^2_{x \mid y}) = \ln \sigma_{x \mid y} + c. \tag{19}$$

In the context of SPECT, $\sigma_{x \mid y}\, 2\sqrt{\ln (4)}$ is the FWHM measure of resolution of X given $W = y$. In light of (19), modulo an additive constant, $H(X \mid W = y)$ is equal to the conventional FWHM resolution measure plotted on a log (dB) scale. The MI is a composite measure of resolution and fluence which is obtained by averaging a function of the FWHM resolution $H(X \mid W = w)$ over the possible observation outcomes w.

IV. Application to Aperture Evaluation

In this section we use the results of Sections II and III to compute the MI between the observation $(N, W) = \{(t_1, W_1), \cdots, (t_n, W_n)\}$ and the emitter locations $X = \{X_1, \cdots, X_n\}$, where $n = N([0, T])$ is the total number of observations. Since the number of observations n is a random variable, the relevant source random vector is the $(n + 1) \times 1$ vector (X, n). In [13] we derive the following:

$$I((X, n); (N, W)) = \Lambda I(X_i, W_i) + H(n) \tag{20}$$

where $\Lambda = \int_{[0,T]} \lambda$ is the average number of counts, and $H(n)$ is entropy of the Poisson random variable n with rate Λ.

An important implication of the expression (20) is that the MI only depends on the joint emitter and observation distribution $dP(X_i, W_i)$ through the λ-independent MI $I(X_i, W_i)$. This is due to the fact that Λ and $H(n)$ are independent of $f_x(x)$ and $f(y \mid x)$.

For infinite detector resolution W_i is a mixed random variable taking on the continuous set of values $\{y\}$ and the discrete value F so that from (13)

$$\begin{aligned} I(X_i; W_i) &= \int_{\mathfrak{X}} dx\, f_x(x) \int_{\mathfrak{Y}} dy\, f(y \mid x) \log \frac{f(y \mid x)}{f_y(y)} \\ &\quad + \int_{\mathfrak{Y}} dx\, f_x(x)\, P(F \mid x) \log \frac{P(F \mid x)}{P(F)}. \end{aligned} \tag{21}$$

The expression (21) is a general MI formula for the ith emission–observation pair which will be applied to SPECT geometries in the following section. Since in general the MI depends on the source distribution $f_x(x)$, the relative MI will be used to permit comparisons between the MI associated with different sources. The relative MI is defined as the ratio of the MI $I(X_i, W_i)$ to the entropy of the source distribution $H(X_i)$

$$I_r(X_i; W_i) = \frac{I(X_i; W_i)}{H(X_i)}. \tag{22}$$

The relative MI is a measure of the transfer of information per bit of source entropy.

A. Linear Geometries

We first consider the application of the MI to a one-dimensional zero thickness aperture and a one-dimensional infinite resolution linear detector for imaging a simple line source. The case of finite length will be discussed in the sequel. In view of the distributions $f_x(x)$, $f_y(y)$, $f(y \mid x)$, and $P(f \mid y)$ found in the previous section, we substitute (4), (6), and (7) into (21) to obtain $I(X_i, W_i)$

explicitly in terms of the aperture function $a(z) = a(x, y)$

$$I(X_i; W_i) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx\,dy \Bigg\{ a(\alpha x + (1-\alpha)y) f_o(y \mid x) f_x(x) \times \log \frac{a(\alpha x + (1-\alpha)y) f_o(y \mid x)}{\int_{-\infty}^{+\infty} a(\alpha v + (1-\alpha)y) f_o(y \mid v) f_x(v)\, dv} \Bigg\} + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx\,dy \Bigg\{ [1 - a(\alpha x + (1-\alpha)y)] f_o(x \mid y) f_x(x) \times \log \frac{\int_{-\infty}^{+\infty} du [1 - a(\alpha x + (1-\alpha)u)] f_o(u \mid x)}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dv\,du [1 - a(\alpha v + (1-\alpha)u)] f_o(u \mid v) f_x(v)} \Bigg\}. \tag{23}$$

Consider the objective of finding an aperture function $a(z)$ that maximizes the MI in (23). In order to make the dependency structure of $I(X_i, W_i)$ on $a(z)$ more explicit, we make a change of variable and rearrange formula (23)

$$I(X_i, W_i) = \int_{-\infty}^{+\infty} dz \big[a(z)\, q_1(z) + (1 - a(z))\, q_2(z) \big] \tag{24}$$

where $q_1(z)$ and $q_2(z)$ are functions, given in Appendix A, of $a(z)$, α, $f_x(x)$, and $f(y \mid x)$.

Specifically, $q_1(z)$ corresponds to the information delivered to the detector by the gamma-rays which pass through the aperture at position $z(a(z) = 1)$, while $q_2(z)$ corresponds to the information delivered by the gamma-rays which are blocked by the aperture at position $z(a(z) = 0)$. In other words, $q_1(z)$ and $q_2(z)$ are measures of the information delivered by the detection events and failure events, respectively.

B. A Global Condition for the Optimal Aperture

The expression (24) gives a necessary and sufficient condition which must be satisfied by an optimal aperture $a(z) = a(x, y)$ which maximizes the MI. Because $a(z)$ is a piecewise binary function, only one of the nonnegative quantities $a(z)\, q_1(z)$ or $(1 - a(z))\, q_2(z)$ can be nonzero for a fixed value of z. Define a_o as an optimum aperture and let this aperture have associated MI: $I_{a_o}(X_i; W_i) = \max_a I_a(X_i; W_i)$. Also define $q_1(z; a_o)$ and $q_2(z; a_o)$ to be the evaluation of q_1 (32) and q_2 (33) of Appendix A, respectively, for the optimum aperture a_o. Then, for any aperture a it can be shown [13]

$$I_a(X_i; W_i) = \int_Z dz \big[a(z)\, q_1(z; a_o) + (1 - a(z))\, q_2(z; a_o) \big] + O(\|a - a_o\|) \tag{25}$$

where $\|a - a_o\| \triangleq \int_Z |a(z) - a_o(z)|\, dz$ is a measure of distance between the two apertures a and a_o. If a is close to a_o, then the term $O(\|a - a_o\|)$ can be neglected, and therefore [13] the optimal aperture $a = a_o$ must satisfy the threshold rule

$$a(z) = \begin{cases} 1, & q_1(z) \geq q_2(z); \\ 0, & q_1(z) < q_2(z). \end{cases} \tag{26}$$

The rule (26) specifies that the optimal aperture is such that $a(z)$ has an opening at z if and only if the information contributed by passing gamma-rays at z exceeds the information contributed by absorbing gamma-rays at z. The threshold rule is valid for apertures which can be completely described on a line, i.e., zero thickness apertures. For apertures, $a(x, y)$, of nonzero thickness, an analogous threshold-type optimality rule holds

$$a(x, y) = \begin{cases} 1, & q_1(x, y) \geq q_2(x, y); \\ 0, & q_1(x, y) < q_2(x, y) \end{cases} \tag{27}$$

where $q_1(x, y)$ and $q_2(x, y)$ are functions of $a(x, y)$, α, and $f(y \mid x)$ given in [13].

C. Aperture Optimization

Although the conditions (26) and (27) characterize the optimal aperture, from our simulation results, it was found that (26) and (27) were difficult to mechanize to search for the optimal aperture. However, the optimality of any designated aperture can be verified *post facto* by using properties (26) and (27).

When we numerically optimized the aperture, practical constraints were imposed on (24) and (23). In the case of a finite one-dimensional detector system, we assumed the detector had length $2l_d$, the object field had length $2l_s$, and the aperture had thickness D_a, length $2l_a$, and a finite number of openings. For an aperture with minimum opening width K, the maximum possible number of openings in the aperture is $n = 2l_a/K$.

We tried several ways to search the space of aperture functions $\mathcal{A} = \{a(z)\}$ for the optimal aperture including hill climbing algorithms. In the numerical study described below we used a local hill climbing algorithm which searched over the MI function for a sequence of monotonically improving apertures. After each iteration of search, only the aperture which yielded the highest MI was retained, and the search was continued until the MI reached a maximum value.

The algorithm is given as follows.

Specify an initial aperture $a_0(z)$. Compute its MI, $I_0(X_i; W_i)$ by using formula (24).

2) For $j = 1, 2, \cdots$, find an updated aperture a_j by toggling $a_{j-1}(z)$ for some z. In the numerical study we used a simple sequential toggling of adjacent aperture elements z_1 to z_n.

3) Calculate the MI, $I_j(X_i; W_i)$, associated with the updated aperture a_j using the formula (24). If $I_j > I_{j-1}$, set $a_{j+1} = a_j$, and go to step 4. If $I_j \leq I_{j-1}$, set $a_{j+1} = a_{j-1}$ and go back to step 2.

4) Pick a stopping rule:

a) Check the rate of increase of MI; if there is not much increase in MI, i.e., the difference between $I_{j-1}(X_i; W_i)$ and $I_j(X_i; W_i)$ is within some tolerance, stop the search. Otherwise go back to step 2.

b) Check the optimality of a_{j+1} with the condition (26). If optimal, stop the search. Otherwise go back to step 2.

For the above algorithm the initial starting point for the aperture was found to be important. If the initial aperture is far away from the optimal aperture, it is possible that the search will converge to a local maximum instead of a global maximum of the MI. In the numerical results section to follow, the global optimality of the aperture obtained from the above algorithm was determined by repeatedly running the algorithm for several different initial apertures. With some hindsight, provided by the results of the numerical results section, the observed structure of the optimal apertures suggests the following strategy for choosing the initial aperture. Let the fluence distribution along the aperture be denoted $f_o(z)$. Select the initial aperture as a parallel hole collimator for those values z where $f_o(z)$ is large (sacrificing detector fluence for better resolution), and as an open aperture for those values z where $f_o(z)$ is small (sacrificing resolution for detector fluence).

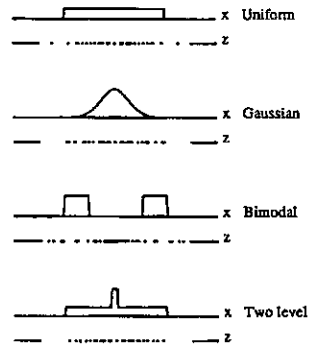

Fig. 8. Source distributions, and their associated mutual information optimal apertures, used in the simulation study of one-dimensional linear geometries. Here, D_{oa} = 30 mm, D_{od} = 29 mm, and the apertures have 160 mm in length, 1 mm in thickness, and 1.6 mm for the minimal opening element.

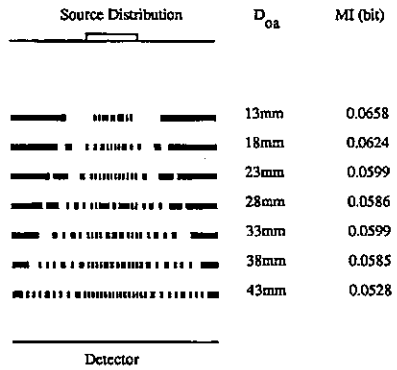

Fig. 9. A series of optimal apertures which show how the optimal aperture pattern changes when the location of aperture changes between the object and detector. The object-detector distance is fixed (50 mm), the aperture has 200 mm in length, 2 mm in thickness, and 1 mm for the minimal opening element. The source distribution is uniform over 50 mm width.

In the numerical study, to be discussed in the next section, we will apply the above hill climbing algorithm to aperture optimization for several one-dimensional object distributions, including uniform, Gaussian, bimodal uniform, and two-level uniform, which are shown in Fig. 8 along with the associated optimal apertures. The effect of varying the aperture location between the object and detector will be studied for the uniform source, Fig. 9. Then, the sensitivity of the optimal aperture to variations in the source distribution will be studied.

D. Ring Geometry

An MI formula can be derived for the ring geometry, which differs somewhat from (23). A distinctive feature of the SPRINT ring geometry is that during acquisition the aperture ring rotates to attain sufficient angular sam-

pling within the field of view. For a rotating ring the data acquisition interval $[0, T]$ is divided into m equal intervals $\{[T_k, T_{k+1}]\}$ over which the average source activity is Λ_k. At the end of each subinterval, the aperture function $a_i(x, \phi, s)$ changes incrementally to the kth rotated version of the original aperture. Since, conditioned on the emission times $\{t_i\}$, the emission/detection process (X_i, W_i) is independent over the time index i, we have the expression

$$I(X; W) = \sum_{k=1}^{m} \Lambda_k \int dx \int d\phi \int ds \left[a_k(x, \phi, s) f_0(\phi, s \mid x) f_x(x) \cdot \log \frac{a_k(x, \phi, s) f_0(\phi, s \mid x)}{\int a_k(x', \phi, s) f_0(\phi, s \mid x') f_x(x')\, dx} \right] + \int dx \int d\phi \int ds \left[(1 - a_k(x, \phi, s)) f_0(\phi, s \mid x) f_x(x) \cdot \log \frac{\int d\phi' \int ds' (1 - a_k(x, \phi', s')) f_0(\phi', s' \mid x)}{\int dx \int d\phi' \int ds' (1 - a_k(x', \phi', s')) f_0(\phi', s' \mid x') f_x(x')} \right] \tag{28}$$

where $f_x(x)$ is the two-dimensional object distribution and $f_0(\phi, s \mid x)$ is the nominal conditional distribution (10).

In the next section, two aperture rings of different diameters containing nine-slits and ten-slits, respectively, will be compared based on the MI measure (28). In both rings, the slits are so designed that the systems have the same object resolution at the center of the field of view. It is important to note that the off-axis resolution of the two rings are different. In the numerical evaluations of (28), we first make comparisons between the two rings for different sources. Then, the change in MI as a function of slit-width for the ten-slit aperture ring will be studied.

V. Numerical Results

A. One-Dimensional Linear Geometry

As previously mentioned, for the linear geometry, optimization of the one-dimensional detector system was investigated for the following line sources: uniform, Gaussian, biomodal uniform, and two-level uniform.

A series of optimal apertures derived for a uniform source distribution is shown in Fig. 9 as a function of object–aperture distance. General observations relating to Fig. 9 are as follows. As the distance between the source and the aperture decreases from 43 mm to 13 mm, the opening and closing aperture transitions become increasingly clustered near the center or mean of the source distribution. This is due to the finite length of the detector which makes the state of the aperture irrelevant for gamma-rays on paths which do not intersect the detector surface. It is also noteworthy that, generally speaking, the width of the openings in the aperture is relatively small and constant over the center region of the source distribution, while the width is increasingly large outside of the range of support of the source distribution. This structure can be related to the greater need to reduce the uncertainty in the center region than is needed for photons detected farther from the center. The structure of these optimal apertures illustrates the following optimal strategy. In the center region where fluence at the aperture is relatively high, the detector views the source through a parallel hole collimator aperture which sacrifices some fluence for higher detector resolution. On the other hand, outside of the center region where nominal fluence is comparatively low, the detector views the source through an aperture with a large opening, which sacrifices some resolution while maintaining fluence. It was observed that when the aperture is at the midpoint between the object and detector, the optimal aperture has three times as much MI as the uniform parallel collimator with equivalent smallest opening width. This indicates the importance of maintaining an optimal tradeoff between resolution and fluence at the detector.

Dependency of the optimal aperture on source distributions is illustrated in Fig. 8. Four one-dimensional optimal apertures derived for the uniform, the equivalent variance Gaussian, the bimodal uniform, and the two-level source distributions are shown. The optimal apertures for the uniform source and the Gaussian source are very similar. Both optimal apertures have rapid transitions near the center of the distributions, and similar relative MI. This suggests that the variance is the principal attribute of uni-

modal source distributions which affects the structure of optimal apertures. On the other hand, for the bimodal source distribution, the optimal aperture differs significantly from the optimal apertures for the unimodal sources. Also, the two-level uniform distribution has an optimal aperture which concentrates its structure about the center of the source distribution to a somewhat greater degree than for the one-level uniform source previously discussed.

Examples illustrating the loss of relative MI when an optimal aperture is used to image a different source than the one for which it was optimized are shown in Table I. In the examples the performance of optimal apertures for the two-level source, the bimodal source, and the uniform source of Fig. 8 are compared when the sources are switched. It can be seen that, between two-level and uniform sources, the relative MI percentage loss is less than 0.3 percent for both cases. On the other hand, between uniform and bimodal sources or two-level and bimodal sources, the relative loss is significantly higher. This suggests that, while the MI is certainly dependent on the underlying distribution, the MI optimal aperture matched to the uniform source is relatively robust to deviations of the source from the assumed distribution used to derive the optimal aperture.

TABLE I
RELATIVE MUTUAL INFORMATION FOR TRUE AND MISMATCHED SOURCES RELATIVE TO AN OPTIMAL APERTURE

Source distribution \ Optimal aperture	Uniform optimal aperture	Two-level optimal aperture	Bimodal optimal aperture
Uniform source	MI_r =0.0333 Loss=0.0 %	MI_r =0.0332 Loss=0.3 %	MI_r =0.0248 Loss=25.5 %
Two-level source	MI_r =0.0335 Loss=0.3 %	MI_r =0.0336 Loss=0.0 %	MI_r =0.0245 Loss=27.1 %
Bimodal source	MI_r =0.0356 Loss=27.3 %	MI_r =0.0339 Loss=30.8 %	MI_r =0.0490 Loss=0.0 %

TABLE II
COMPARISON OF THE MUTUAL INFORMATION AND THE FLUENCE BETWEEN NINE- AND TEN-SLIT RING APERTURES FOR VARIOUS SOURCES

		Mutual information $\times 10^{-4}$ (bit) I(X,W)	Fluence $\times 10^{-4}$	$\frac{MI_{9\text{-slit}}}{MI_{10\text{-slit}}}$
22cm Diameter Uniform	10-slit	2.070	1.487	1.386
	9-slit	2.870	2.044	
1.5cm diameter on-center	10-slit	0.412	1.373	1.399
	9-slit	0.576	1.889	
1.5cm diameter off-center	10-slit	0.567	1.474	1.394
	9-slit	0.797	2.020	
1.5cm diameter multipl source	10-slit	1.642	1.415	1.390
	9-slit	2.282	1.946	

TABLE III
RELATIVE MUTUAL INFORMATION PER BIT OF SOURCE ENTROPY FOR THE NINE- AND TEN-SLIT APERTURES AND THE IDENTICAL SOURCES AS IN TABLE II

Source distributions	$MI_r = I(X;W)/H(X)$ $\times 10^{-4}$ (1/BIT)	
	10 slit aperture	9 slit aperture
22cm diameter uniform	0.803	1.112
1.5cm diameter on-center	0.875	1.224
1.5cm diameter off-center	1.091	1.521
1.5cm diameter multiple source	0.952	1.323

B. Two-Dimensional Extended Sources

The evaluation of two aperture rings using MI is presented in Table II. The first aperture ring has a diameter of 306 mm with nine equally spaced slits over the aperture with slit width of 3.4 mm (see Fig. 2). The second aperture ring has a diameter of 340 mm with ten equally spaced slits of slit width 2.4 mm. Both imaging systems have the same fields of view, 220 mm in diameter, the same detector ring diameters, 500 mm in diameter, and the same geometric resolutions of 9.2 mm on-axis. Four different sources were simulated, including a uniform cylindrical source which filled the field of view, a small central source, multiple small sources, and a small off-center source. The MI and fluence for the two rings with these sources is shown in Table II. These results indicate that the nine-slit ring has higher MI than the ten-slit ring for all sources. The ratio of the MI of the nine-slit ring to the ten-slit ring (denoted by $R = I_9/I_{10}$) is very close to the ratio of the respective fluences. This is significant in that it suggests that for fixed geometric resolution the MI is essentially a measure of fluence. Table III shows the relative MI for the same sources as in Table II. For both aperture rings, the small off-center source gives the highest relative MI. This can be attributed to the higher geometric resolution of the ring geometry for off-center sources and essentially equal fluences [18]. This implies that the small off-center source is more easily estimated from detected data than other sources. These results are in agreement with conventional arguments [18].

Fig. 10 shows the relative MI as a function of slit width. The slit width is specified by duty cycle which is the number of aperture elements which are open relative to the total number of elements. For both small on- and off-center sources, the relative MI does not change significantly except at the extremes of 4.5 and 95.5 percent, where the geometric resolution is inadequate relative to the diameters of the sources. The relative MI for uniform and multiple sources increases as duty cycle increases and reach their maximum at duty cycle of about 43 and 60 percent, respectively. This maximum provides the optimal tradeoff of fluence and resolution for maximum MI.

The results indicated in Fig. 10 are very similar to the results obtained in [4] where the relative standard deviation was compared for different size on-axis uniform sources as a function of mean code transparency (duty cycle of the aperture openings). This similarity is consistent with the theoretical remarks in Section III concerning the relation of the MI to fluence and resolution.

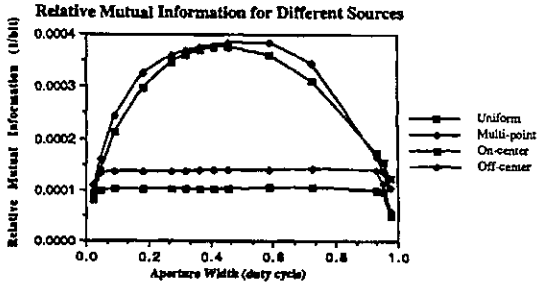

Fig. 10. The mutual information for the ten-slit SPRINT II ring geometry for varying width slits. The horizontal axis is expressed in duty cycle of openings in the ring aperture. The relative MI for point sources are fairly constant, while the MI for uniform and multiple-point sources have a maximum at duty cycle values of 43 and 60 percent, respectively.

VI. Conclusion

The MI criterion can be used to evaluate and optimize apertures based on the fundamental measure of transfer of source information across the aperture to the detector. While the conventional parallel-hole collimator weights the emitter positions within the object equally, the optimal MI aperture strives for an optimal tradeoff between the fluence and resolution, depending on the object distribution and the system geometry.

One major issue is information robustness of given apertures. For the cases studied, our results show that, while specifically designed to match a particular mean emitter distribution, the optimal aperture appears fairly robust relative to small deviations in the assumed source distribution. This implies that when a source is only partially known, an optimal aperture designed under this partial knowledge should be nearly optimal for the true source, as long as the true source does not differ too much from the design point.

As presented in this paper, the MI is primarily useful for the evaluation of a prespecified aperture for a prespecified mean emitter distribution. For practical application of the MI criterion to optimal aperture design, the model must be extended to unknown sources. There are several methods which are currently being investigated, including maximum-likelihood iteration and min–max design. Other areas of investigation worth pursuing are the extensions of the MI formulation to emission/detection processes with attenuation due to Compton scattering and absorption, and to dynamic processes where the mean emitter distribution may change over time. Finally, we are considering the formulation of MI for characterizing the inherent ability of tomographic systems to perform specific tasks. These tasks may not be limited to tomographic reconstruction. For example, classification, detection, or feature extraction may be of primary interest.

A final issue is the relation between the MI and "image quality." As a theoretical issue, it is known that high MI is necessary for good performance of any classifier, estimator, or detector, regardless of the specific form of the penalty criterion (see Section III-C). As a practical issue, however, high MI will not be sufficient for high quality images unless good reconstruction algorithms are specified for the MI-optimal apertures. While, due to the irregular aperture hole patterns, parallel beam algorithms based on reconstruction from line integrals are not directly applicable to these MI apertures, algorithms such as maximum likelihood (ML) [19] can be used. The issue of image quality improvement of high MI detection systems can only be resolved through experimental studies of joint implementation of MI-optimal apertures and associated reconstruction algorithms.

Appendix A

In this Appendix the form for the mutual information (24) is derived. Recalling the form for the mutual information (23) we have ($f \triangleq fx$)

$$
\begin{aligned}
I(X;W) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx\,dy \Bigg\{ & a(\alpha x + (1-\alpha)y)\,F_{1+2}(x-y)\,f(x) \\
& \times \log \frac{a(\alpha x + (1-\alpha)y)\,F_{1+2}(x-y)}{\int_{-\infty}^{+\infty} a(av + (1-\alpha)y)\,F_{1+2}(v-y)\,f(v)\,dv} \Bigg\} \\
+ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx\,dy \Bigg\{ & \left[1 - a(\alpha x + (1-\alpha)y)\right] F_{1+2}(x-y)\,f(x) \\
& \times \log \frac{\int_{-\infty}^{+\infty} du\left[1 - a(\alpha x + (1-\alpha)u)\right] F_{1+2}(x-u)}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dv\,du\left[1 - a(\alpha v + (1-\alpha)u)\right] F_{1+2}(v-u)\,f(v)} \Bigg\}
\end{aligned}
\tag{29}
$$

where, to simplify via change of variables, $F_{1+2}(y-x)$ has been defined

$$F_{1+2}(y-x) = \frac{1}{2\pi D_{od}\left(1+\left(\frac{x-y}{D_{od}}\right)^2\right)}.$$

Let

$$\alpha x + (1-\alpha)y \triangleq z, \qquad \alpha x + (1-\alpha)u \triangleq u'.$$

Then

$$dx = \frac{dz}{\alpha}, \qquad x = \frac{z-(1-\alpha)y}{\alpha}$$

$$du' = \frac{du}{(1-\alpha)}, \qquad u = \frac{u'-z+(1-\alpha)y}{(1-\alpha)}$$

$$x-y = \frac{z-y}{\alpha}, \qquad x-u = \frac{z-(1-\alpha)y-\alpha u'}{\alpha(1-\alpha)}.$$

Substituting these variables into (29)

$$I(X;W) = \int_{-\infty}^{+\infty} dz \int_{-\infty}^{+\infty} dy \left\{ a(z)\frac{1}{\alpha}F_{1+2}\left(\frac{z-y}{\alpha}\right) f\left(\frac{z-(1-\alpha)y}{\alpha}\right) \times \log \frac{a(z)F_{1+2}\left(\frac{z-y}{\alpha}\right)}{\int_{-\infty}^{+\infty} a(z')\frac{1}{\alpha}F_{1+2}\left(\frac{z'-y}{\alpha}\right) f\left(\frac{z'-(1-\alpha)y}{\alpha}\right) dz'} \right\}$$

$$+ \int_{-\infty}^{+\infty} dz \int_{-\infty}^{+\infty} dy \left\{ [1-a(z)]\frac{1}{\alpha}F_{1+2}\left(\frac{z-y}{\alpha}\right) f\left(\frac{z-(1-\alpha)y}{\alpha}\right) \times \log \frac{\int_{-\infty}^{+\infty} du'[1-a(u')]\frac{1}{1-\alpha}F_{1+2}\left(\frac{z-(1-\alpha)y-\alpha u'}{\alpha(1-\alpha)}\right)}{\int_{-\infty}^{+\infty} dz' \int_{-\infty}^{+\infty} du[1-a(z')]\frac{1}{\alpha}F_{1+2}\left(\frac{z'-u}{\alpha}\right) f\left(\frac{z'-(1-\alpha)u}{\alpha}\right)} \right\}. \tag{30}$$

Now, define

$$g(z,y) \triangleq \frac{1}{\alpha}F_{1+2}\left(\frac{z-y}{\alpha}\right) f\left(\frac{z-(1-\alpha)y}{\alpha}\right) \tag{31}$$

and

$$q_1(z) \triangleq \int_{-\infty}^{+\infty} dy\, g(z,y) \log \frac{a(z)F_{1+2}\left(\frac{z-y}{\alpha}\right)}{\int_{-\infty}^{+\infty} dz' a(z') g(z',y)} \tag{32}$$

$$q_2(z) \triangleq \int_{-\infty}^{+\infty} dy\, g(z, y) \log \frac{\int_{-\infty}^{+\infty} du'\left[1 - a(u')\right] \frac{1}{1-\alpha} F_{1+2}\left(\frac{z - (1-\alpha)\, y - \alpha u'}{\alpha(1-\alpha)}\right)}{\int_{-\infty}^{+\infty} dz' \int_{-\infty}^{+\infty} du\left[1 - \alpha(z')\right] g(z', u)}. \tag{33}$$

Finally, substitute (31), (32), and (33) into (30) to obtain (24)

$$I(X; W) = \int_{-\infty}^{+\infty} dz\left[a(z)\, q_1(z) + \left(1 - a(z)\right) q_2(z)\right].$$

Appendix B

In this Appendix the conditional probability distribution of the gamma-ray path intersection on a ring detector is derived. Fig. 6 shows the coordinate parameterization for the ring while Fig. 7 is a blow-up of the chisel shaped region in Fig. 6. Referring to Fig. 7 we make the following definitions. An emitter located at a point lying in the plane, $x = (x_1, x_2)$, emits a gamma-ray at azimuthal angle ψ and longitudinal angle θ whose path intersects a flat annular detector ring at ring coordinates (ϕ, s). The emission angles are referenced to an arbitrary zero angle, e.g., in Fig. 7 ϕ is measured relative to the indicated (horizontal) baseline, and θ is measured relative to the line connecting the emitter position x to the origin O. Positions on the detector ring are parameterized by the angle ϕ measured relative to the origin O in Fig. 7 and the horizontal position on the interior of the ring s is relative to the center line o in the figure. The emission angle (θ, ψ) is assumed to be uniform over $[0, 2\pi] \times [-\pi/2, \pi/2]$. The path length from the emitter location x to the detector intersection (ϕ, s) is denoted by a.

The objective is to find the nominal conditional distribution $f_o(\phi, s \mid x)$ of the detector hitting point without an aperture. Since the transformation from (θ, ψ) to (ϕ, s) is one-to-one we have the following:

$$f_o(\phi, s \mid x) = f(\theta, \psi \mid x) \frac{1}{|J|} \tag{34}$$

where J is the Jacobian matrix associated with the transformation. Referring to Fig. 7 it is evident that J is a triangular matrix since variations in ψ produce no variations in ϕ. Hence,

$$|J| = \left|\frac{\partial \phi}{\partial \theta}\right| \left|\frac{\partial s}{\partial \psi}\right|. \tag{35}$$

Furthermore, due to the assumed uniform distribution of (θ, ψ), (34) becomes

$$f_o(\phi, s \mid x) = \frac{1}{2\pi^2} \left|\frac{\partial \theta}{\partial \phi}\right| \left|\frac{\partial \psi}{\partial s}\right|,$$

$$\phi \in [0, 2\pi],\ s \in (-\infty, \infty). \tag{36}$$

In reference to Fig. 7, we have

$$a \sin(\theta) = R \sin(\alpha), \qquad \text{(Law of sines)}; \tag{37}$$

$$s = a \tan \psi, \tag{38}$$

$$a = \sqrt{(x_1 - R\cos\phi)^2 + (x_2 - R \sin \phi)^2}, \tag{39}$$

$$\alpha = \tan^{-1}\left(\frac{x_2}{x_1}\right) - \phi. \tag{40}$$

Now from (37) we have

$$\theta = \sin^{-1}\left(\frac{R}{a} \sin \alpha\right). \tag{41}$$

Substitution of (40) into (41) gives the following [13]:

$$\frac{d\theta}{d\phi} = \frac{R^2\left(\frac{a^2}{R^2}(x_1 \cos\phi + x_2 \sin\phi) - \frac{1}{R}(x_1 \cos\phi + x_2 \sin\phi)^2 + \frac{\|x\|^2}{R}\right)}{a^2\left(x_1 \cos\phi + x_2 \sin\phi - \frac{\|x\|^2}{R}\right)} \tag{42}$$

and an expression for $ds/d\psi$ is obtained from (38)

$$\frac{ds}{d\psi} = a\frac{1}{\cos^2\psi} = a\frac{1}{\cos\left(\sin^{-1}\frac{s}{a}\right)} = a\left(1+\left(\frac{s}{a}\right)^2\right). \tag{43}$$

Substitute (42) and (43) into (36) to obtain the final form for the joint distribution of ϕ and s given x:

$$f_o(\phi, s\,|\,x) = \frac{1}{2\pi^2}\left|\frac{R^2\left(-\frac{a^2}{R^2}(x_1\cos\phi + x_2\sin\phi) - \frac{1}{R}(x_1\cos\phi + x_2\sin\phi)^2 + \frac{\|x\|^2}{R}\right)}{a^3\left(1+\left(\frac{s}{a}\right)^2\right)\left(x_1\cos\phi + x_2\sin\phi - \frac{\|x\|^2}{R}\right)}\right| \tag{44}$$

for $\phi \in [0, 2\pi]$ and $s \in (-\infty, \infty)$.

Substitution of a^2 into the numerator of (44), use of (39)

$$f_o(\phi, s\,|\,x) = \frac{1}{2\pi^2}\left|\frac{R^2\left(\frac{1}{R}(x_1\cos\phi + x_2\sin\phi)\left(x_1\cos\phi + x_2\sin\phi - \frac{\|x\|^2}{R}\right) - \left(x_1\cos\phi + x_2\sin\phi - \frac{\|x\|^2}{R}\right)\right)}{a^3\left(1+\left(\frac{s}{a}\right)^2\right)\left(x_1\cos\phi + x_2\sin\phi - \frac{\|x\|^2}{R}\right)}\right|. \tag{45}$$

and after some algebraic manipulations

$$f_o(\phi, s\,|\,x) = \frac{1}{2\pi^2}\frac{R^2 - R(x_1\cos\phi + x_2\sin\phi)}{a^3\left(1+\left(\frac{s}{a}\right)^2\right)}. \tag{46}$$

In the case of a far-field source and a thin axial slice, $s^2/a^2 \ll 1$, and hence, using (39)

$$f_o(\phi, s\,|\,x) = \frac{1}{2\pi^2}\frac{R^2 - R(x_1\cos\phi + x_2\sin\phi)}{\left((x_1 - R\cos\phi)^2 + (x_2 - R\sin\phi)^2\right)^{3/2}}. \tag{47}$$

References

[1] W. E. Smith, R. G. Paxman, and H. H. Barrett, "Image reconstruction from coded data: I. Reconstruction algorithms and experimental results," *J. Opt. Soc. Amer.*, vol. A.2, pp. 758–761, 1985.

[2] G. F. Knoll, W. L. Rogers, K. F. Koral, J. S. Stamos, and N. H. Clinthorne, "Application of coded aperture in tomographic head scanning," *Nucl. Instr. Meth. Phys. Res.*, vol. 221, pp. 226–232, 1984.

[3] K. F. Koral, W. L. Rogers, and G. F. Knoll, "Digital tomographic imaging with time modulated pseudorandom coded aperture and Anger camera," *J. Nucl. Med.*, vol. 16, pp. 402–413, May 1975.

[4] W. L. Rogers and R. S. Adler, "Time-coded aperture design for nuclear medicine imaging: A study of signal-to-noise ratio," *Appl. Opt.*, vol. 21, pp. 324–333, 1982.

[5] G. F. Knoll, "Single photon emission computed tomography," *IEEE Proc.*, vol. 71, pp. 320–329, Mar. 1983.

[6] B. M. Tsui and R. J. Jaszczak, "Interactions of collimation, sampling and filtering on SPECT spatial resolution," *IEEE Trans. Nucl. Sci.*, vol. NS-31, pp. 527–531, Feb. 1984.

[7] H. H. Barrett, W. E. Smith, K. J. Meyers, T. D. Milster, and R. D. Fiete, "Quantifying the performance of imaging systems," in *Proc. SPIE*, pp. 65–69, 1986.

[8] J. A. Swets, "ROC analysis applied to the evaluation of medical imaging techniques," *Invest. Radiol.*, vol. 14, pp. 109–121, 1979.

[9] N. M. Moody, W. Paul, and M. L. Joy, "A survey of medical gamma-ray cameras," *IEEE Proc.*, vol. 58, pp. 217–242, Feb. 1970.

[10] R. F. Wagner, D. G. Brown, and M. S. Pastel, "Application of information theory to the assessment of computed tomography," *Med. Phys.*, vol. 6, pp. 83–94, Mar./Apr. 1979.

[11] R. Blahut, *Applications of Information Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[12] A. O. Hero, "Unobserved counts in ECT: Information loss and likelihood corrections," Commun. Signal Processing Lab. (CSPL), Dep. EECS, Univ. Michigan, Tech. Memo., 1989.

[13] L. Shao and A. O. Hero, "Information criteria for SPECT image reconstruction and aperture optimization," Commun. Signal Processing Lab., Dep. EECS, Univ. Michigan, Tech. Rep. 261, Oct. 1988.

[14] H. Stark and J. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[15] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[16] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1978.

[17] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[18] W. L. Rogers, N. H. Clinthorne, L. Shao, P. Chiao, J. Stamos, and K. F. Koral, "SPRINT II, a second generation single photon ring tomograph," *IEEE Trans. Med. Imaging*, vol. 7, pp. 291–297, Dec. 1988.

[19] M. I. Miller, D. L. Snyder, and T. R. Miller, "Maximum-likelihood reconstruction for single photon emission computed tomography," *IEEE Trans. Nucl. Sci.*, vol. NS-32, pp. 769–778, Feb. 1985.