Information Analysis of Single Photon Emission Computed Tomography With Count Losses

ALFRED O. HERO AND LING SHAO

Abstract—This paper presents an analysis of the information transfer from emitter space to detector space in Single Photon Computed Tomography (SPECT) systems which, unlike the study in [1], takes into account the fact that count loss "side information" is generally not available at the detector. This side information corresponds to the number of lost or deleted γ-rays due to lack of interaction with the detector. We show that the information transfer depends on the structure of the likelihood function of the emitter locations associated with the detector data. This likelihood function is the average of a set of ideal-detection likelihood functions each matched to a particular set of possible deleted γ-ray paths. We derive a lower bound on the information gain due to the incorporation of the count loss side information at the detector which is shown to be significant under either of the following conditions: 1) when the mean emission rate is small; 2) when the γ-ray deletion probability is strongly dependent on emitter location. Numerical evaluations of the mutual information, with and without side information, associated with information optimal apertures and uniform parallel-hole collimators are then presented.

I. Introduction

SINGLE Photon Emission Computed Tomography (SPECT) is a diagnostic imaging system based on reconstruction from projections of a γ-ray emitting source. SPECT systems are composed of a radioactive γ-ray emitting spatial source, a set of position sensitive detector surfaces which detect incident γ-rays, and a perforated lead aperture placed between the source and the detector in order to reduce the residual uncertainty in emitter position associated with the detected γ-rays. The set of projections are defined as the sequence of incident γ-ray positions along the set of detector surfaces. The emission process is modeled as a marked Poisson process over the spatial field of view of the detectors, and the detection process is obtained from the emission process from the randomly directed line paths of the γ-rays. Not all of the line paths pass through the perforations in the aperture and are detected, however. The loss of detected counts can be due to absorption into aperture septa, noninteraction with the detector scintillation crystal, or nonintersection of the γ-ray path with the detector surfaces. Furthermore, for high count rate studies, detected counts may not be registered due to detector, or system, dead time [2]. In this manner, the projections form a thinned marked Poisson process which is obtained by random deletions of certain γ-ray paths.

In Emission Computed Tomography (ECT), the overall performance of image reconstruction and classification algorithms is limited by the quality of the projections provided by the aperture/detector geometry. These projections contain the information on emitter locations which is essential for accurate reconstruction of the mean emitter distribution constituting the object of interest. In the context of SPECT [1], and other inverse problems [3]–[7], the mutual information has been used as an objective measure of system performance where other more conventional measures have been difficult to apply. This approach has been justified based on the role of the mutual information as a measure of intrinsic invertibility of a system transfer function, i.e., a measure of the degree to which the observations determine the variables to be recovered. In addition, the mutual information is related to reconstruction error through the rate distortion lower bound of information theory: uniformly low mutual information necessarily implies poor reconstruction performance under any convex penalty criterion [8].

In [1] a mutual information analysis was applied to an ideal-detection SPECT system having the capability of detecting the occurrence of an emission even if the emitted γ-ray is not detected on the detector surface. Presently, practical SPECT detection systems do not record the "side information" concerning the occurrence of undetected γ-rays. In this paper, we generalize the information analysis of [1] to the case where some counts are lost due to undetected γ-rays, which we call random deletions. We present results concerning the effect of the random deletions on the mutual information via comparisons of the results obtained here and the results obtained in [1].

II. Background and Main Assumptions

With few exceptions, throughout the paper we use the conventions that capital and bold letters denote random variables, e.g., X_i, $\boldsymbol{n}$, while lower case letters denote their realizations, e.g., x_i and n.

Let $\mathfrak{X}$ and $\mathfrak{Y}$ denote the "source space" and the "detector space," respectively. As in [1] over a fixed time interval $[0, T]$ a source generates γ-ray emissions at $\boldsymbol{n}$ random locations $\underline{\boldsymbol{X}} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n]^T$, $\boldsymbol{X}_i \in \mathfrak{X}$, where,

Manuscript received August 18, 1989; revised November 20, 1989. This work was supported in part by the National Cancer Institute, DHHS, under PHS Grant CA32856, PI W. L. Rogers.

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

L. Shao was with the Bioengineering Program, The University of Michigan, Ann Arbor, MI 48109. He is now with the Radiology Dept., Hospital of the University of Pennsylvania, Philadelphia, PA 19104.

IEEE Log Number 9034415.

0278-0062/90/0600-0117$01.00 © 1990 IEEE

conditioned on $n = n$, each emitter location X_i, $i = 1, \cdots, n$, is independent and identically distributed with density f_x. These emissions occur at times $\underline{t}^x = [t_1^x, \cdots, t_n^x]^T$ which are points of a temporally homogeneous Poisson process $N_x = \{N_x(t): t \in [0, T]\}$ with rate Λ equal to the average number of points over $[0, T]$: $\Lambda = E[N_x(T)] = E[n]$. Define the joint distributions of $(\underline{X}, n)$ at $(\underline{x}, n)$ as the differential probability $dP(x_1, \cdots, x_n, n) = P(X_1 \in (x_1, x_1 + dx_1], \cdots, X_N \in (x_n, x_n + dx_n], N_x(T) = n)$. Then:

$$dP(\underline{X}, n) = P_{N_x(T)}(n) \prod_{i=1}^{n} f_x(X_i)\, d\underline{x} \qquad (1)$$

where $d\underline{x} = dx_1, \cdots, dx_n$, and $P_{N_x(T)}(n)$ is the "Poisson-Λ" distribution:

$$P_{N_x(T)}(n) = \frac{\Lambda^n}{n!} e^{-\Lambda} \qquad n = 0, 1, \cdots. \qquad (2)$$

A γ-ray emitted at location X_i can either be detected at a location $Y_i \in \mathcal{Y}$ on some detector surface, or it can go undetected, an event denoted "F" for deletion or failure. The conditional probability of F given $X_i = x$ is denoted by $p_F(x)$, and the unconditional probability of F is the expectation:

$$\bar{p}_F = \int_{\mathcal{X}} p_F(x) f_x(x)\, dx. \qquad (3)$$

The conditional density of Y_i given $X_i = x$ and "*not* F" (no deletion) is denoted by $f(y|x)$ for $y \in \mathcal{Y}$. The conditional density of Y_i given "*not* F" is as follows:

$$f_y(y) = \frac{1}{1 - \bar{p}_F} \int_{\mathcal{X}} [1 - p_F(x)] f(y|x) f_x(x)\, dx. \qquad (4)$$

Both $f(y|x)$ and $f_y(y)$ are normalized densities, i.e., they have unit mass when integrated over $y \in \mathcal{Y}$. Observe that $p_F(x)$ and $f(y|x)$ are functions specified by the system geometry, while $\bar{p}_F$ and f_y are also functions of f_x.

Let $\underline{W} = [W_1, \cdots, W_n]^T$ be the sequence of detected positions and deletions, $W_i = Y_i$ or $W_i = F$, respectively. Let $N_y(T) = m$ be the number of detected γ-rays, and denote by $N_y = \{N_y(t): t \in [0, T]\}$ the sequence of detection times $\underline{t}^y = [t_1^y, \cdots, t_m^y]^T$. We refer to $(N_y, \underline{Y})$ and $(N_x, \underline{W})$ as detector processes. $(N_x, \underline{W})$ corresponds to an ideal-detector process for which the sequence of deletions $\{i: W_i = F\}$ are known to the observer. It can be shown [8], since all processes are stationary, that the process distributions depend on the sequence of deletions only through the number of deletions $n - m = \#\{i: W_i = F\}$. Hence $(N_x, \underline{W})$ corresponds to a detector process for which the number $n - m$ is known, or, equivalently, for which the total number of emissions n is known. In the sequel we refer to this knowledge as the "count loss side infomation."

The process $(N_x, \underline{W})$ is a marked homogeneous Poisson process [9] with mark space $\mathcal{W} = \mathcal{Y} \cup F$ and rate Λ:

$$dP(N_x, \underline{W}) = \frac{1}{T^n} P_{N_x(T)}(n) \left[\prod_{i=1}^{m} f_y(Y_i)\right] \cdot [1 - \bar{p}_F]^m [\bar{p}_F]^{n-m}. \qquad (5)$$

In (5) the Y_i's have been reindexed over $i \in \{1, \cdots, m\}$. The point process $(N_y, \underline{Y})$ can be obtained from $(N_x, \underline{W})$ by deletion of the events $\{W_i: W_i = F\}$, called "thinning" [9], and it can be shown that $(N_y, \underline{Y})$ is a marked homogeneous Poisson process with mark space $\mathcal{Y}$ and rate $\Lambda(1 - \bar{p}_F)$:

$$dP(N_y, \underline{Y}) = \frac{1}{T^m} P_{N_y(T)}(m) \prod_{i=1}^{m} f_y(Y_i). \qquad (6)$$

In (6) $\{P_{N_y(T)}(m)\}_{m \geq 0}$ is given by (2) with Λ replaced by $(1 - \bar{p}_F)\Lambda$, i.e., m is a Poisson-$(1 - \bar{p}_F)\Lambda$ distributed random variable.

III. The Mutual Information Measure

The general formula for the mutual information between two random quantities U and V is [10]:

$$I(U; V) \overset{\text{def}}{=} E\left[\ln \frac{dP(U, V)}{dP(U)\, dP(V)}\right]. \qquad (7)$$

The mutual information measure can be applied to any estimation, detection, or other statistical problem which can be imbedded in a communication paradigm involving a set of "source symbols," $\{U\}$, which are the input to a "communication channel" yielding a set of output "destination symbols," $\{V\}$, which can be observed. In the context of tomography, the mutual information should be defined relative to a particular task by suitable definition of the source symbols. In [8, Ch. 6] image classification and detection tasks are considered, here we concentrate on tomographic reconstruction.

The overall tomographic system relating the "input" f_x to the "output" $\hat{f}_x$ can be represented as the cascade of three channels (see Fig. 1): the first generates the emitter locations $(\underline{X}, n)$ from the emitter distribution f_x; the second generates the detection process $(N_y, \underline{Y})$ from its input $(\underline{X}, n)$; and the third generates the reconstruction $\hat{f}_x$ from its input $(N_y, \underline{Y})$. The first channel is essentially a randomization mechanism following the statistics of a marked Poisson process. The second channel is characterized by the physical mechanisms underlying γ-ray production and detection, which is partially determined by the system geometry. The third channel is determined by the particular reconstruction algorithm which is applied to the measured data. For tomographic reconstruction, a worthwhile objective would be to maximize the mutual information between the mean emitter distribution, $U = f_x$, and the γ-ray detection process $V = (N_y, \underline{Y})$. If a set of possible mean emitter distributions, $\{f_x\}$ can be specified, along with prior probabilities on the members of the set, the mutual information $I(f_x; (N_y, \underline{Y}))$ can in principle be de-

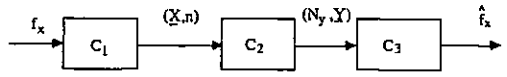

Fig. 1. Channel model for SPECT. C_1 is capacity of channel which maps mean emitter distribution f_x to realized emitter locations $(\underline{X}, n)$, C_2 is capacity of channel which maps $(\underline{X}, n)$ to detection process $(N_y, \underline{Y})$, and C_3 is capacity of channel which maps $(N_y, \underline{Y})$ to reconstruction, $\hat{f}_x$, of mean emitter distribution, f_x.

rived. As in [1], we focus on the simpler task of characterizing the mutual information, related to C_2, between the emitter locations $U = (\underline{X}, n)$ and $V = (N_y, \underline{Y})$. The data processing theorem [10] gives the result as follows:

$$C \leq \min\{C_1, C_2, C_3\}$$

where C is the information capacity of the cascaded channel constituting the transfer of source information from f_x to $\hat{f}_x$, and C_1, C_2, C_3 are information capacities of each of the three channels indicated in Fig. 1. Hence, while a large information capacity C_2 is not sufficient for the overall capacity C to be large, a large C_2 is necessary for C to be large. In this sense a characterization of the mutual information $I((\underline{X}, n); (N_y, \underline{Y}))$ is relevant to the transfer of information across the system.

In [1] the ideal-detection mutual information $I((\underline{X}, n); (\underline{W}, N_x))$ was studied. The ideal-detection case corresponds to a system which has access to the side information consisting of the sequence of deletions $W_i = F$ in the process $(N_x, \underline{W})$. It was shown in [1] that conventional measures of SPECT system performance, such as sensitivity and spatial resolution, system invertibility, and MSE of the reconstruction, are related to this ideal-detection mutual information. In this paper, we are interested in quantifying the loss in mutual-information when the detection system does not have access to the deletion side information. For this case the relevant detection process is $V = (N_y, \underline{Y})$ and the mutual information is defined as follows:

$$I((\underline{X}, n); (N_y, \underline{Y})) = E\left[\ln \frac{dP(\underline{X}, n, N_y, \underline{Y})}{dP(\underline{X}, n)\, dP(N_y, \underline{Y})}\right] = E\left[\ln \frac{dP(\underline{X}, n | N_y, \underline{Y})}{dP(\underline{X}, n)}\right]. \quad (8)$$

From the above expression it is clear that to evaluate the mutual information the *posterior* likelihood function $dP(\underline{X}, n | N_y, \underline{Y})$ is needed.

IV. The Posterior Distribution

Define the binary "deletion indicator function" d_i as follows:

$$d_i = \begin{cases} 1, & \text{if } W_i \neq F \\ 0, & \text{if } W_i = F. \end{cases} \quad (9)$$

With this notation define the column vector $\underline{d} = [d_1, \cdots, d_n]^T$. $\underline{d}$ is an n bit binary word whose one and zero entries specify the sequence in which the emissions are detected, $W_i = Y_i$, and deleted, $W_i = F$, respectively. Also define the position $\gamma(i)$ of the ith 1 in the vector $\underline{d}$: $\gamma(i) \stackrel{\text{def}}{=} \min\{k: \Sigma_{j=1}^{k} d_j \geq i\}$, $i = 1, \cdots, m$. $\gamma(i)$ is the time index at which the ith detection occurred, in particular: $t_i^y = t_{\gamma(i)}^x$.

Using the Poisson property that, conditioned on $n = n$, the pairs $\{(W_i, X_i)\}_{i=1}^n$ are independent, the joint distribution of $(\underline{X}, n)$ and the ideal observations $(N_x, \underline{W})$ is as follows [8]:

$$dP((\underline{X}, n), (N_x, \underline{W})) = \frac{1}{T^n} P_{N_x(T)}(n) \prod_{i=1}^{m} f(Y_i | X_{\gamma(i)}) f_x(X_{\gamma(i)}) \prod_{i=1}^{n} \cdot [1 - p_F(X_i)]^{d_i} [p_F(X_i) f_x(X_i)]^{1-d_i}\, d\underline{x}\, d\underline{y}\, dt. \quad (10)$$

By summing the joint distribution $dP((\underline{X}, n), (N_x, \underline{W}))$ over the possible ways that $n - m$ deletions $W_i = F$ can occur in the sequence $\underline{W}$, and dividing by $dP(N_y, \underline{Y})$ (6), the conditional distribution of $(\underline{X}, n)$ given the detector process $(N_y, \underline{Y})$ is, for $n = m, m+1, \cdots$ [8]:

$$dP(\underline{X}, n | \underline{Y}, N_y) = l_m(\underline{X}, n) \frac{P_{N_x(T)}(n) \prod_{i=1}^{n} f_x(X_i)}{P_{N_x(T)}(m) \prod_{i=1}^{m} f_y(Y_i)}\, d\underline{x} \quad (11)$$

where

$$l_m(\underline{x}, n) \stackrel{\text{def}}{=} \sum_{\underline{d}: \underline{d}^T \underline{1} = m} \prod_{i=1}^{m} f(Y_i | x_{\gamma(i)}) \cdot \prod_{i=1}^{n} p_F^{1-d_i}(x_i) (1 - p_F(x_i))^{d_i}. \quad (12)$$

By convention, $\Pi_{i=1}^{n}$ and $\Sigma_{\underline{d}: \underline{d}^T \underline{1} = 0}$ are equal to one when $n = 0$ so that: $l_m(\underline{x}, n) = 1$ when $m, n = 0$.

The expression (11) is the general posterior likelihood function when only m out of n of the points of the marked Poisson process $(N_x, \underline{X})$ are detected. The summation index in (12) iterates over all $\binom{n}{m}$ ways that the detector locations $Y_1, \cdots, Y_m$ can be associated with subsets of m of the emission locations $X_1, \cdots, X_n$. The function $l_m(\underline{x}, n)$ (12) is the likelihood function of the n γ-ray emission locations $x_1, \cdots, x_n$ given the m detected points $\underline{Y}$. If $p_F(x) = 0$ for all x, no deletions occur, $n = m$, and the likelihood function associated with the m emitter locations $x_1, \cdots, x_m$ is given by the following "uncompensated likelihood function":

$$l_m^{\text{un-comp}}(\underline{x}, m) = \prod_{j=1}^{m} f(Y_j | x_j). \quad (13)$$

Using this notation, from (12) and (13) we have:

$$l_m(\underline{x}, n) = \sum_{\underline{d}:\underline{d}^T\underline{1}=m} \alpha_{\underline{d}} l_m^{\text{un-comp}}(x_{\gamma(1)}, \cdots, x_{\gamma(m)}, m) \tag{14}$$

where $\alpha_{\underline{d}}$ is a weight associated with the *prior* probability that the particular length m subsequence of emitter positions $x_{\gamma(1)}, \cdots, x_{\gamma(m)}$ generated the m detected γ-rays:

$$\alpha_{\underline{d}} \stackrel{\text{def}}{=} \prod_{i=1}^{m} \left[1 - p_F(x_i)\right]^{d_i} \left[p_F(x_i)\right]^{1-d_i}. \tag{15}$$

It is important to note that the likelihood functions (12) and (11) differ from the likelihood function derived in [11]. This is because the approach in [11] sets up the likelihood functions for the mean emitter distribution f_x, while in the present context the likelihood functions apply to the particular realization of the emitter distribution, i.e., the actual locations of the emitters.

V. Effect of Random Deletions on the Mutual Information

The main results of this paper are given in the form of Propositions which are proven in the Appendix. First we give an equivalent expression for the mutual information associated with the ideal detection process whose form differs slightly from the form derived in [1].

Proposition 1: Let $(\underline{X}, n)$ be the spatial locations and number of γ-ray emissions and let $(N_x, \underline{W})$ be the ideal detection process defined at the end of Section II where the distributions $dP(\underline{X}, n)$ and $dP(\underline{X}, n | N_x, \underline{W})$ are given in (1) and (11). Then the mutual information, I_o, between the emission locations and the ideal detector process is the function:

$$I_o \stackrel{\text{def}}{=} I\big((\underline{X}, n); (N_x, \underline{W})\big) = H(n) + \Lambda I(X_i; W_i). \tag{16}$$

In (16) $H(n) = h_{\text{Poisson}}(\Lambda)$ is the entropy of the Poisson-Λ number of emissions; $I(X_i; W_i)$ is the mutual information between an emission location X_i and the detector/deletion symbol W_i:

$$I(X_i; W_i) = (1 - \bar{p}_F)\, I(X_i; Y_i) + D\big(p_F(x) \| \bar{p}_F\big) \tag{17}$$

where $I(X_i; Y_i)$ is the mutual information for a single emission-detection pair:

$$I(X_i; Y_i) = \int_{\mathcal{Y}} dy\, f_y(y) \int_{\mathcal{X}} dx\, f(x|y) \ln \frac{f(x|y)}{f_x(x)}; \tag{18}$$

and $D(p_F(x) \| \bar{p}_F)$ is the information divergence between the conditional probability of deletion, $p_F(x)$, and the average probability of deletion, $\bar{p}_F$:

$$D\big(p_F(x) \| \bar{p}_F\big) = \int_{\mathcal{X}} f_x(x) \left[p_F(x) \ln \frac{p_F(x)}{\bar{p}_F} \right] dx. \tag{19}$$

The entropy $H(n)$ is present in (16) since perfect information is provided by the ideal detection process $(N_x, \underline{W})$ concerning the number of emissions.

The following are the main results of this paper.

Proposition 2: Let $(\underline{X}, n)$ and $(N_x, \underline{W})$ be as in Proposition 1. Let $(N_y, \underline{Y})$ be the thinned process obtained from $(N_x, \underline{W})$ by deleting the points associated with the symbols "$W_i = F$." Then the mutual information I_{ed}, between the emission locations and the detection process is the function:

$$\begin{aligned} I_{ed} &\stackrel{\text{def}}{=} I\big((\underline{X}, n); (N_y, \underline{Y})\big) \\ &= H(m) + \Lambda(1 - \bar{p}_F)\, I(X_i; Y_i) \\ &\quad - E\left[\ln \frac{l_m^{\text{un-comp}}(\underline{X}, m)}{l_m(\underline{X}, n)} \right] \end{aligned} \tag{20}$$

where $H(m) = h_{\text{Poisson}}((1 - \bar{p}_F)\Lambda)$ is the entropy of the Poisson-$((1 - \bar{p}_F)\Lambda)$ number of detections; and $l_m(\underline{x}, n)$ and $l_m^{\text{un-comp}}(\underline{x}, m)$ are the likelihood function (12) and the uncompensated likelihood function (13).

I_{ed} is the mutual information between the emitter locations and the detection processes when the count loss side information is not available at the detector. An approximation to I_{ed} is obtained by neglecting the third term on the right hand side of the equality (20):

$$I_{ed} \approx I_1 \stackrel{\text{def}}{=} H(m) + \Lambda(1 - \bar{p}_F)\, I(X_i; Y_i). \tag{21}$$

The right hand side of (21) corresponds to the information between any m emitter locations $X_{\gamma(1)}, \cdots, X_{\gamma(m)}$ and the m observables $\underline{Y}$. The next proposition shows that the approximation (21) is actually an upper bound on I_{ed}. Furthermore this upper bound can be significantly less than the ideal detection information I_o in Proposition 1. These two facts will be used in the sequel to bound the information loss incurred by omitting count loss side information.

Proposition 3: The mutual information, $I_{ed} = I((\underline{X}, n); (N_y, \underline{Y}))$ satisfies the following chain of upper bounds:

$$I_{ed} \le I_1 - H(m|\underline{X}) \tag{22}$$

$$\le I_1 \tag{23}$$

$$= I_o - \left[H(n) - H(m) + \Lambda D\big(p_F(x) \| \bar{p}_F\big)\right] \tag{24}$$

$$\le I_o \tag{25}$$

where I_1 is the mutual information quantity defined in (21): I_o is the mutual information (16) between the emitter locations and the ideal detection process, and $D(p_F(x) \| \bar{p}_F)$ is the information divergence (19) between the conditional probability of deletion, $p_F(x)$, and the average probability of deletion, $\bar{p}_F$.

The proof of the above proposition uses an application of the *log-sum lemma* [12]. The loosest inequality, $I_{ed} \le I_o$, of Proposition 3 expresses the obvious fact that omis-

sion of the count loss side information can only decrease the mutual-information. The amount of decrease per-emission can be measured by the loss:

$$\text{loss} \stackrel{\text{def}}{=} \frac{I_o - I_{ed}}{\Lambda}. \tag{26}$$

The inequalities (22), (24), and (25), of Proposition 2 give the chain of lower bounds on the loss:

$$\text{loss} \geq \frac{I_o - I_1}{\Lambda} \tag{27}$$

$$\geq \frac{H(n) - H(m)}{\Lambda} + D(p_F(x) \| \bar{p}_F)$$

$$\geq 0. \tag{28}$$

The lower bound (28) is composed of the sum of two terms: 1) the scaled difference $[H(n) - H(m)]/\Lambda$ between the entropies of the number of emissions and number of detections; and 2) the information divergence $D(p_F(x) \| \bar{p}_F)$, given in (19). While each of these terms are nonnegative, it will be seen below that for large Λ, the information divergence term dominates the loss lower bound (28). The information divergence term can be interpreted as an (asymmetric) measure of the distance between the conditional deletion probability $p_F(x)$ and the average deletion probability $\bar{p}_F$ in the sense that (19) is zero if the two probabilities are equal while (19) increases as $\int p_F(x) \mid p_F(x) - \bar{p}_F \mid dx$ increases [12]. Furthermore, $D(p_F(x) \| \bar{p}_F)$ has the interpretation of information, $I(X_i; F)$, between the deletion event "F" and the emitter location X_i. Therefore, if "F" conveys no information about emitter position, in the sense that $D(p_F(x) \| \bar{p}_F) = 0$, the resultant information loss bound reduces to the difference $[H(n) - H(m)]/\Lambda$.

Consider the behavior of the loss lower bound (28) as $\bar{p}_F$ goes to either zero or one. In either case, the divergence $D(p_F(x) \| \bar{p}_F)$ (19) goes to zero since $\bar{p}_F = 0$ ($\bar{p}_F = 1$) implies $p_F(x) = 0$ ($p_F(x) = 1$). The Poisson entropy function is monotonically nondecreasing from zero to infinity as the Poisson rate increases [8, Appendix E]. Therefore, $H(m) = h_{\text{Poisson}}(\Lambda[1 - \bar{p}_F])$ approaches $H(n) = h_{\text{Poisson}}(\Lambda)$ monotonically from below as $\bar{p}_F$ approaches zero. On the other hand $h_{\text{Poisson}}([1 - \bar{p}_F]\Lambda)$ approaches zero as $\bar{p}_F$ approaches one. Hence, the bound on information loss (28) approaches zero as $\bar{p}_F$ decreases to zero, while the bound approaches $H(n)/\Lambda$, the normalized entropy of the number of emissions, as $\bar{p}_F$ increases to one. Since an information value of $H(n)$ corresponds to maximum uncertainty in the number of emitted counts, this behavior is consistent with intuition, and suggests that, at least for large or small $\bar{p}_F$, the bounds (24) and (28) are fairly tight.

It is interesting to consider the relative magnitude of the loss as compared to the ideal-detection mutual information as a function of emission rate Λ. It can be shown [8, Ch. 5] that the following asymptotic forms of (28) hold:

$$\text{loss} \geq D(p_F(x) \| \bar{p}_F), \qquad (\Lambda \gg 1)$$

$$\text{loss} \geq \frac{H(n) - H(m)}{\Lambda} = \bar{p}_F \ln \frac{1}{\Lambda}, \qquad (\Lambda \ll 1). \tag{29}$$

Based on the asymptotic forms (29) of the per-emission information loss, the following implications can be deduced. For low mean emission rates, Λ, the per-emission information loss is a logarithmic function of $1/\Lambda$, which can be significant if the average probability of deletion, $\bar{p}_F$, is high. In particular, for very low mean emission rates the loss becomes unbounded. On the other hand, for high mean emission rates the per-emission information loss is a constant independent of Λ: the information divergence. If the divergence is large the absolute loss $I_o - I_{ed}$ can be significant. In terms of per-emission information loss it can be concluded that there is a diminishing return on incorporating count loss side information into the observations as Λ increases: for large Λ the return cannot be greater than $D(p_F(x) \| \bar{p}_F)$.

VI. Numerical Examples

To illustrate the sensitivity of the mutual information criterion to the effect of random deletions we numerically computed the mutual information for ideal detection, I_o (16), investigated in [1], the upper bound on the mutual information in the absence of count loss side information, I_1 (21), and the lower bound, *loss*, (28), on the per-emission information loss. To evaluate the loss, the entropy $h_{\text{Poisson}}(\Lambda)$ of a Poisson random variable must be computed. The Poisson entropy function is not of analytic form [13], [14], and in general the loss must be computed numerically. For small Λ this can be accomplished via truncation of the series $-\sum_{k=1}^{\infty} p(k) \ln p(k)$, while for large Λ the standard Gaussian approximation [15, Theorem 20] to the Poisson distribution can be used, for which the difference $H(n) - H(m) = h_{\text{Poisson}}(\Lambda) - h_{\text{Poisson}}([1 - \bar{p}_F]\Lambda)$ is approximated by $-(1/2) \ln (1 - \bar{p}_F)$. In these studies we considered two simple cases of one-dimensional line sources, referred to as objects in the sequel: 1) a single uniform (constant spatial intensity) source occupying 25 percent of the field of view, 2) two spatially separated uniform sources, called bimodal sources in the figures, each occupying 10 percent of the field of view.

We found optimal apertures by numerically maximizing the ideal-detection mutual information, I_o, and the nonideal detection mutual information bound I_1. For the cases studied, it was determined that the optimal apertures resulting from maximization of each of the criteria, I_o and I_1, were virtually identical despite the different functional forms of these two criteria. This is significant since efficient methods for maximization of I_o, which have been studied in [1], can be used to maximize I_1. Furthermore,

the optimal strategy displayed by I_o-maximizing apertures, discussed in [1], carries over to I_1-maximizing apertures.

In Figs. 2–12 the following quantities were held constant: detector and aperture lengths are 200 mm, source–detector distance is 60 mm, and minimal aperture opening width is 1 mm. These values were chosen to represent typical parameters for small linear one-dimensional SPECT geometries.

In Fig. 2 optimal apertures for a uniform source over a 40 mm center region are displayed for various object-to-aperture distances D_{oa}. The apertures have thickness $\Delta = 2$ mm. In this and other figures the openings of the aperture are represented by white spaces while the septa are represented by dark spaces. Also shown are the associated mutual information per-emission, I_1/Λ and, the probability of detection, $1 - \bar{p}_F$, which can be interpreted as the detector fluence-per-emission, for $\Lambda = 1000$. Note that a decrease in mutual information is accompanied by a decrease in fluence as the object-to-aperture distance is increased, however the rates of decrease are different. The apertures reveal the following optimal strategy, outlined in [1], for maximizing the information transfer from source to detector. Over regions of the aperture where the fluence is high enough so that fluence can be sacrificed for better resolution, e.g., near the center region, the aperture performs collimation. Over lower fluence regions the aperture opens up to allow a higher fluence to attain the detector. The regions of the aperture near the boundaries, denoted by solid dark lines in Fig. 2, are regions where opening or closing of the aperture has little or no effect on the projections since the possible γ-ray paths which intersect these regions do not intersect the detector. Over these regions we have arbitrarily set the aperture to a closed state.

In Fig. 3 optimal apertures for a bimodal source are displayed as a function of object-to-aperture distance D_{oa}. The bimodal source has intersource separation $d = 80$ mm and each source occupies 20 mm. The aperture thickness is $\Delta = 2$ mm. Note that over regions of the aperture which would otherwise cause severe multiplexing between the two sources on the detector, e.g., the center region, the aperture is closed.

In Fig. 4 the middle aperture of Fig. 3 is displayed along with the unnormalized detector fluence distribution $(1 - \bar{p}_F) f_y$. For this aperture I_1/Λ is 0.163 bits/emission and the total fluence-per-emission is 0.151. Observe that from the point of view of detector resolution, e.g., as measured by the concentration bandwidth of $(1 - \bar{p}_F) f_y$, the optimal aperture is quite poor. In Fig. 5 a uniform parallel-hole collimator, placed at an identical distance 36 mm from the line source, is shown along with its unnormalized detector fluence distribution. The detector fluence distributions are plotted on the same scale for Figs. 4 and 5. The thickness of the collimator is 2 mm, the width of each opening is 1 mm, and the duty cycle is 50 percent. Note that, despite the fact that the detector resolution appears superior for the collimator as compared to the optimal aperture of Fig. 4, the fluence is inferior, as is the mutual information. In particular there is more than twice as much emitter related information, I_1, at the detector for the optimal aperture than for the collimator aperture. An important unsolved problem is the specification of a decoding algorithm which can extract this additional information.

In Fig. 6 the results of maximization of the mutual information via specification of optimal apertures are dis-

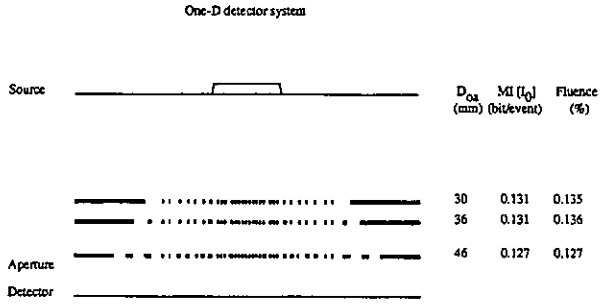

Fig. 2. MI-optimal apertures as a function of object-to-aperture distance D_{oa} for a uniform source of width 40 mm. Aperture and detector lengths are 200 mm, aperture thickness is $\Delta = 2$ mm, object-to-detector distance is 60 mm. For MI and fluence calculations $\Lambda = 1000$.

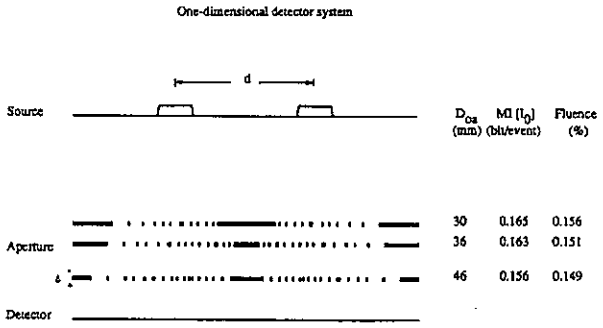

Fig. 3. MI-optimal apertures as a function of object-to-aperture distance D_{oa} for a pair of line sources (bimodal source) each of width 20 mm, and intersource spacing $d = 80$ mm. Aperture and detector length are 200 mm, aperture thickness is 2 mm, object-to-detector distance is 60 mm. For MI and fluence calculations $\Lambda = 1000$.

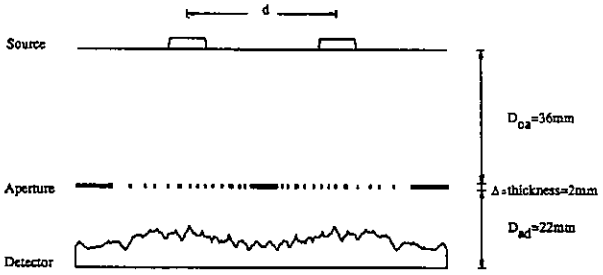

Fig. 4. MI-optimal aperture of Fig. 3 for $D_{oa} = 36$ mm. Also plotted is unnormalized detector fluence distribution (scaled for visual presentation), $(1 - \bar{p}_F) f_y(y)$. Total peremission fluence is 0.151 and peremission MI (I_o/Λ) is 0.63. MI is higher for this aperture than for collimator in Fig. 5 despite comparatively poorer detector resolution.

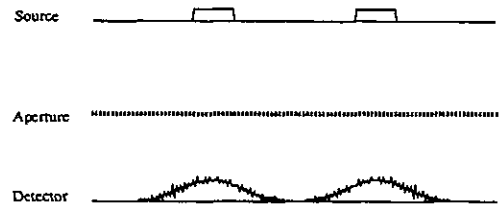

Fig. 5. Parallel-hole collimator aperture for identical system configurations as in Fig. 4. Width of each opening is 1 mm, duty cycle is 50 percent, and collimator thickness is $\Delta = 2$ mm. Unnormalized detector fluence distribution $(1 - \bar{p}_F) f_y(y)$ is plotted on same scale as in Fig. 4. Total peremission fluence is 0.0789 and peremission MI (I_n/Λ) is 0.0382.

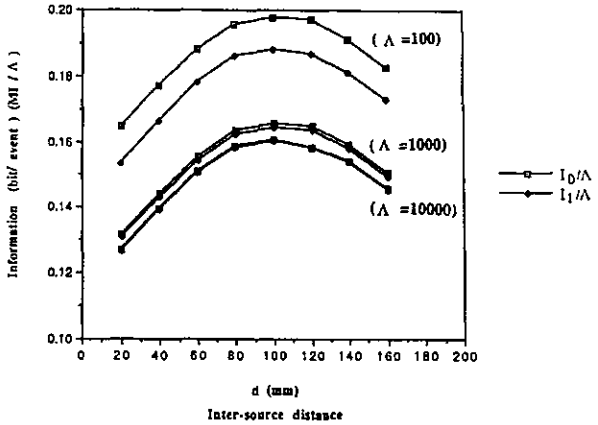

Fig. 6. Plots of MI peremission, I_n/Λ and I_1/Λ, for optimizing apertures, represented in Fig. 4 for $d = 80$ mm, as a function of source separation d for mean emission rates $\Lambda = 100$, $\Lambda = 1000$, and $\Lambda = 10{,}000$. For each d, MI-optimal apertures were found and associated MI calculated. Plots show maximum achievable MI peremission under each of criteria I_n and I_1. Comparison between each pair of curves indicates decreasing importance of count loss as Λ increases.

played for the bimodal source in Fig. 3 and for $\Lambda = 100$, $\Lambda = 1000$, and $\Lambda = 10{,}000$, respectively. The horizontal axis in the figure is the intersource separation d and the vertical axis is the information per emission under criteria I_o and I_1. The figures show I_o and I_1 based on their strongest possible showings: the maximum I_1 is compared to the maximum I_o for each value of d. Note that the curves, Fig. 6, are not achievable with any fixed aperture since for each d the apertures giving each of the points on these curves are different. For each value of Λ studied, the ideal mutual information, I_o/Λ, is greater than the mutual information, I_1/Λ, as was predicted by Proposition 3. The distance between I_1/Λ and I_1/Λ curves represents the information loss (peremission), due to deletions of the failure events "F" from the observations, as a function of Λ. It is significant that the difference between I_o/Λ and I_1/Λ increases as Λ decreases. This is consistent with the theoretical prediction, (29), that the count loss side information is more significant at low mean count rates than at high mean count rates. In particular for $\Lambda = 100$, and optimal apertures, the lack of count loss side information incurs a loss of approximately 5 percent of the mutual

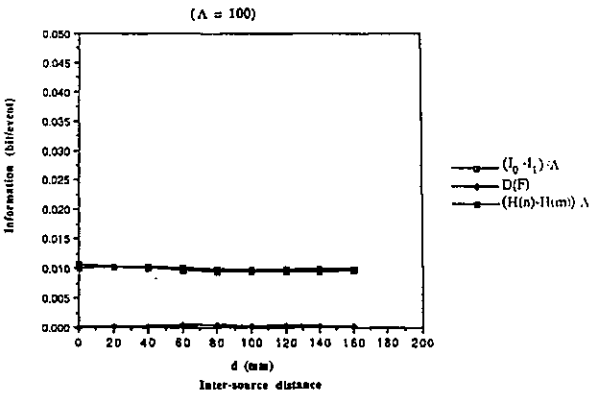

Fig. 7. Plot of components of MI loss per-emission, $[I_n - I_1]/\Lambda = [H(n) - H(m)]/\Lambda + D(p_F(x) \| \bar{p}_F)$, corresponding to difference between upper and lower curves in Fig. 6, as a function of intersource separation d, for $\Lambda = 100$. For this case MI loss is almost entirely due to entropy difference $[H(n) - H(m)]/\Lambda$.

information, I_o/Λ, available for ideal detection. Conversely, for higher Λ and optimal apertures, the count loss side information can be expected to have minimal impact on the total information transfer from the emitter locations to the detector.

In Figs. 7 and 8, the relative contributions of the entropy difference term $[H(n) - H(m)]\Lambda$ and the information divergence term $D(p_F(x) \| \bar{p}_F)$ to the loss $[I_n - I_1]/\Lambda$ (28), are studied for $\Lambda = 100$ and $\Lambda = 10{,}000$, respectively, and the same parameters as in Fig. 6. The top curves in Figs. 7 and 8 correspond to the difference between the top and bottom curves in Fig. 6. Comparison between Figs. 7 and 8 quantitatively illustrates (29): for low mean emission rate the information loss is dominated by the entropy difference $[H(n) - H(m)]\Lambda$, while for high mean emission rate the information loss is dominated by the information divergence $D(p_F(x) \| \bar{p}_F)$.

In Fig. 9 the mutual information per emission is plotted for $\Lambda = 100$ and $\Lambda = 1000$, respectively, and for the case where the aperture is optimized for an intersource distance $d = 80$ mm, yielding the aperture of Fig. 4, while the true intersource distance varies over the range 20 mm to 160 mm. The increased magnitude curvature of these curves relative to the lower two pairs of curves in Fig. 6 reflects the degree to which the fixed aperture is suboptimal, due to mismatch, when d is different from the value for which it was optimized.

In Fig. 10 the mutual information per emission is plotted as a function of intersource distance d for $\Lambda = 100$ and $\Lambda = 1000$, respectively, and for the parallel-hole collimator aperture of Fig. 5. Comparison between these curves and the curves in Fig. 9 indicates that the collimator is more robust to changes in d, in the sense that the mutual information curves in Fig. 10 are virtually constant over a large range of d around $d = 80$ mm. It is to be recalled, however, that the collimator has only half as much mutual information as the optimal aperature when $d = 80$ mm.

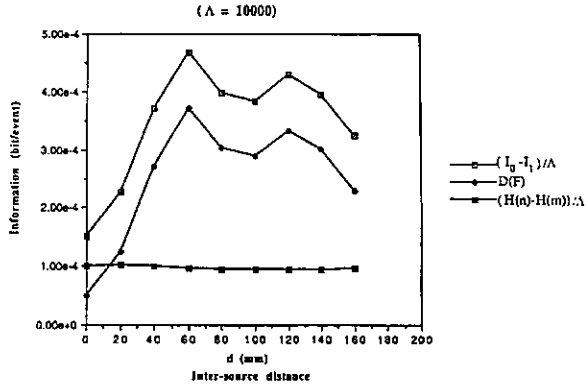

Fig. 8. Plot of components of MI loss per-emission, $[I_o - I_1]/\Lambda = [H(n) - H(m)]/\Lambda + D(p_F(x) \| \bar{p}_F)$, corresponding to difference between upper and lower curves in Fig. 6, as a function of intersource separation d, for $\Lambda = 10{,}000$. Here MI loss is almost entirely due to information divergence $D(p_F(x) \| \bar{p}_F)$.

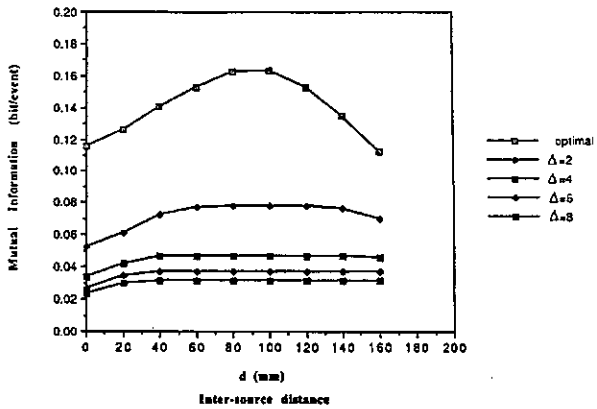

Fig. 11. Plots of MI peremission, I_1/Λ, as a function of d for fixed MI-optimal aperture of Fig. 4 and uniform parallel-hole collimators with fixed 2 mm opening width, 50 percent duty cycle, and varying collimator thicknesses Δ.

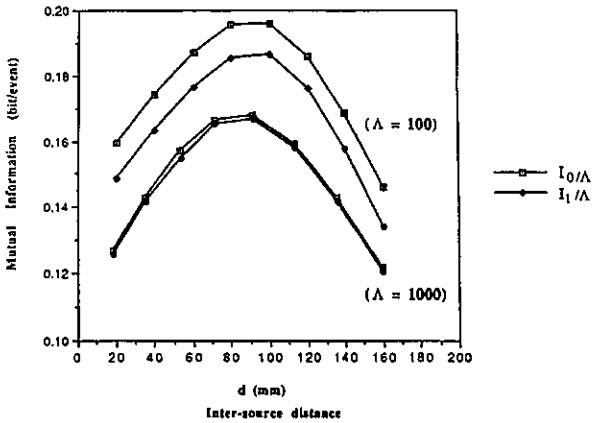

Fig. 9. Plots of MI per-emission, I_o/Λ and I_1/Λ, as a function of source separation d for optimal aperture of Fig. 4 which is optimized only for $d = 80$ mm. $\Lambda = 100$ and $\Lambda = 1000$, respectively.

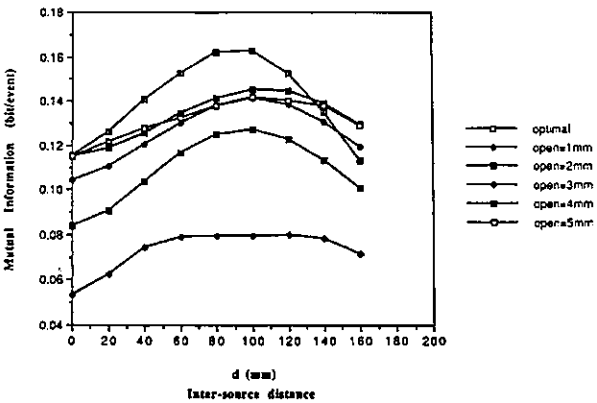

Fig. 12. Plots of MI per-emission, I_1/Λ, as a function of d for fixed MI-optimal aperture of Fig. 4 and uniform parallel-hole collimators with fixed 2 mm collimator thickness, 50 percent duty cycle, and varying widths of collimator openings.

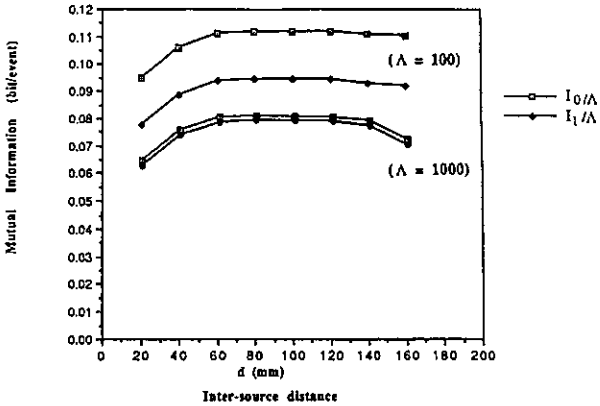

Fig. 10. Plots of I_o/Λ and I_1/Λ as a function of source separation d for collimator aperture of Fig. 5 for $\Lambda = 100$ and $\Lambda = 1000$, respectively. Comparison of the $\Lambda = 100$ curves to the analogous curves in Fig. 9 indicates a 100 percent gain in MI is possible by using an MI-optimal aperture. This gain is only achievable if true separation $d = 80$ mm is known.

The robustness issue is further studied in Figs. 11 and 12. In Fig. 11 the mutual information (I_1) of the fixed optimal aperture for $d = 80$ mm, shown in Fig. 4, is compared to the mutual information of uniform parallel-hole collimators with 1 mm aperture opening width, 50 percent duty cycle, and varying thicknesses Δ. For all collimator thicknesses studied, the collimators had significantly lower mutual information than the mismatched optimal aperture, labeled "optimal" in the figure, the deficiency increasing as the thickness increases. For the collimators shown, it is better, in an information sense, to use a mismatched optimal aperture. In Fig. 12, a similar comparison of mutual information (I_1) is shown for a set of collimators with fixed thickness $\Delta = 2$ mm, 50 percent duty cycle, and various widths of the aperture openings. As the width was increased, the mutual information increased to a maximum at $\Delta = 4$ mm and then began to decrease for larger values of Δ. Comparing the collimator mutual information curve associated with this optimal value $\Delta = 4$

mm, and the mismatched aperture mutual information curve, labeled "optimal" in the figure, it is evident that the robustness of this collimator begins to pay off as the true value of d increases beyond 140 mm. This implies that if the uncertainty in the actual intersource separation is significant, e.g., greater than 50 percent of the nominal value of 80 mm, the collimator with $\Delta = 4$ mm may have better information transfer properties than the mismatched aperture. On the other hand, if the uncertainty in d is less than 20 percent the mismatched aperture can still guarantee at least a 10 percent gain in mutual information over the collimators studied.

VII. Conclusion

We have investigated the performance of SPECT from the point of view of information transfer from the object to the projections on the detector which takes account of count loss side information. This side information corresponds to the number of lost γ-rays due to lack of interaction with the detector, e.g., through SPECT aperture absorption, out-of-field γ-ray paths, or incomplete energy deposition into the scintillator. The mutual information between the n γ-ray emissions and the $m \leq n$ γ-ray detections involves the likelihood function of the n emitter locations. This likelihood function can be represented as the weighted average of $\binom{n}{m}$ uncompensated likelihood functions, each of which is the likelihood of a possible subset of m of the n emission locations. An upper bound on the mutual information was then given, and a lower bound on the information loss due to the deletions of γ-ray paths was given. This lower bound indicated that the impact of count loss side information on the information transfer can be significant if either the mean emission rate, Λ, is small or if the deletion probability is strongly dependent on spatial emitter location. The small Λ regime may be relevant for dynamical studies where successive imaging times are limited to intervals over which the emitter distribution is nearly time independent. The deletion probability can be strongly dependent on emitter location for SPECT systems with spatially variant mean response, such as those incorporating coded apertures.

Results of a numerical study of two one-dimensional line sources were presented which indicated that the SPECT aperture which maximizes the mutual information is virtually identical to the ideal-detection information-optimal SPECT aperture, studied in [1]. For a mean emission rate of 100 events, and a source made up of two spatially separated line sources, called a bimodal source, the count loss accounts for a 5 percent drop in mutual information per-emission. As the mean emission rate increases to 10,000 the mutual information per-emission is not significantly affected by the count loss. Furthermore, for the bimodal source studied, the information-optimal aperture provides more than a 10 percent gain in information relative to an information-optimized uniform parallel-hole collimator when the source separation is known to within 20 percent of its true value. This implies that the information-optimal aperture is relatively robust to deviations of the source away from the assumed aperture design point. This suggests that the mutual information criterion might be relevant to aperture design when one has a good initial guess of the true mean distribution.

While the information based approach taken in this paper can indicate potential information gains, it does not tell us how to use the measurements to achieve the information gain. This situation is inherent to the Shannon theory of communication: the theory only establishes the existence of optimal source coding which can achieve maximum information transfer. For practical realization of the benefits of the information gains reported here more work needs to be done. As a first step, we are currently considering the implementation of the maximum likelihood estimator for the emitter locations discussed in Section IV of this paper.

Appendix

Here we prove the three propositions given in this paper.

Proof of Proposition 1: Identification of $(\underline{X}, n)$ and $(N_x, \underline{W})$ with U and V in (7) gives:

$$I\big((\underline{X}, n); (N_x, \underline{W})\big) = E\left[\ln \frac{dP(\underline{X}, n \,|\, N_x, \underline{W})}{dP(\underline{X}, n)}\right]. \quad (30)$$

The posterior distribution $dP(\underline{X}, n \,|\, N_x, \underline{W})$ can be obtained from dividing $dP((\underline{X}, n), (N_x, \underline{W}))$ (10) by $dP(N_x, \underline{W})$ (5):

$$dP(\underline{X}, n \,|\, N_x, \underline{W}) = \frac{1}{[1 - \bar{p}_F]^m [\bar{p}_F]^{n-m}} \prod_{i=1}^{m} f(X_{\gamma(i)} \,|\, Y_i) \prod_{i=1}^{n} \cdot \big[1 - p_F(X_i)\big]^{d_i} \big[p_F(X_i) f_x(X_i)\big]^{1-d_i} d\underline{x}\, d\underline{y}\, dt, \quad (31)$$

Using (1) and (31) the mutual information formula (30) becomes:

$$\begin{aligned} I\big((\underline{X}, n); (N_x, \underline{W})\big) &= E\left[\ln \prod_{i=1}^{n} \left[\frac{f(X_i \,|\, Y_i)}{f_x(X_i)}\right]^{d_i} \left[\frac{p_F(X_i)}{\bar{p}_F}\right]^{1-d_i}\right] \\ &\quad - E\big[\ln P_{N_\Lambda(T)}(n)\big] \\ &= E\left[\sum_{i=1}^{n} d_i \ln \frac{f(X_i \,|\, Y_i)}{f_x(X_i)} + (1 - d_i) \ln \frac{p_F(X_i)}{\bar{p}_F}\right] \\ &\quad + H(n). \end{aligned} \quad (32)$$

Now since, given n, $\{(W_i, X_i)\}_{i=1}^{n}$ are i.i.d., $\Sigma_{i=1}^{n} d_i = m$, $E[n] = \Lambda$, and $E[m] = (1 - \bar{p}_F)\Lambda$, the above re-

duces to:

$$
\begin{aligned}
&I((\underline{X}, n); (N_x, \underline{W})) \\
&= E[m]\, E\left[\ln \frac{f(X_i \mid Y_i)}{f_x(X_i)} \middle| d_i = 1\right] \\
&\quad + E[n - m]\, E\left[\ln \frac{p_F(X_i)}{\bar{p}_F} \middle| d_i = 0\right] + H(n) \\
&= \Lambda(1 - \bar{p}_F) \int_{\mathcal{Y}} dy\, f_y(x) \int_{\mathcal{X}} dx\, f(x \mid y) \ln \frac{f(x \mid y)}{f_x(x)} \\
&\quad + \Lambda \bar{p}_F \int_{\mathcal{X}} dx\, f_x(x) \frac{p_F(x)}{\bar{p}_F} \ln \frac{p_F(x)}{\bar{p}_F} + H(n) \\
&= \Lambda\left[(1 - \bar{p}_F)\, I(X_i; Y_i) + D(p_F(x) \| \bar{p}_F)\right] + H(n).
\end{aligned} \tag{33}
$$

This establishes Proposition 1.

Proof of Proposition 2: Using (1) and (11), the information (8) becomes:

$$
\begin{aligned}
&I((\underline{X}, n); (N_y, \underline{Y})) \\
&= E\left[\ln \frac{l_m(\underline{X}, n) \dfrac{P_{N_x(T)}(n) \prod_{i=1}^{n} f_x(X_i)}{m!}}{\dfrac{P_{N_y(T)}(m) \prod_{i=1}^{m} f_y(Y_i)}{n!}} \Big/ P_{N_x(T)}(n) \prod_{i=1}^{n} f_x(X_i)\right] \\
&= E[\ln l_m(\underline{X}, n)] - E\left[\ln P_{N_y(T)}(m) \prod_{i=1}^{m} f_y(Y_i)\right]
\end{aligned} \tag{34}
$$

From the identity [10] $I(X_i, Y_i) = H(Y_i) - H(Y_i \mid X_i)$, the second quantity on the right hand side of the equality above is:

$$
\begin{aligned}
&E\left[\ln P_{N_y(T)}(m) \prod_{i=1}^{m} f_y(Y_i)\right] \\
&\quad = E[\ln P_{N_y(T)}(m)] + E\left[\sum_{i=1}^{m} \ln f_y(Y_i)\right] \\
&\quad = -H(m) + E[m] \int f_y(y) \ln f_y(y)\, dy \\
&\quad = -H(m) - \Lambda(1 - \bar{p}_F)\, H(Y_i) \\
&\quad = -H(m) - \Lambda(1 - \bar{p}_F)\, I(X_i; Y_i) \\
&\qquad - \Lambda(1 - \bar{p}_F)\, H(Y_i \mid X_i)
\end{aligned} \tag{35}
$$

where $H(m)$ is the entropy of a Poisson random variable with rate $\Lambda(1 - \bar{p}_F)$, $H(Y_i) = E[-\ln f_y(Y_i)]$ is the (differential) entropy of Y_i, and $H(Y_i \mid X_i) = E[-\ln f(Y_i \mid X_i)]$ is the conditional entropy of Y_i given X_i. Using the definition of $l_m^{\text{un-comp}}$ (13) we also have $E[\ln l_m^{\text{un-comp}}(\underline{X}, m)] = E[\ln \Pi_{i=1}^{m} f(Y_i \mid X_i)] = -\Lambda(1 - \bar{p}_F) H(Y_i \mid X_i)$. Combining this with (34) and (35) establishes identity (20) of Proposition 1.

Proof of Proposition 3: In view of (20) we first show:

$$
E\left[\ln \frac{l_m^{\text{un-comp}}(\underline{X}, m)}{l_m(\underline{X}, n)}\right] \geq H(m \mid \underline{X}), \tag{36}
$$

to establish (22). Since the entropy of a discrete random variable is nonnegative [10], $H(m \mid \underline{X}) \geq 0$ and the inequality (24) will immediately follow.

In the proof of Proposition 2 it was shown that $E[\ln l_m^{\text{un-comp}}(\underline{X}, m)] = -\Lambda(1 - \bar{p}_F) H(Y_i \mid X_i)$. Hence:

$$
\begin{aligned}
&E\left[\ln \frac{l_m^{\text{un-comp}}(\underline{X}, m)}{l_m(\underline{X}, n)}\right] \\
&\quad = -\Lambda(1 - \bar{p}_F) H(Y_i \mid X_i) - E[\ln l_m(\underline{X}, n)].
\end{aligned} \tag{37}
$$

Now explicitly express the expectation:

$$
E[\ln l_m(\underline{X}, n)] = \int dP(\underline{y}, m, \underline{x}, n) \ln l_m(\underline{x}, n). \tag{38}
$$

To obtain the differential probability $dP(\underline{y}, m, \underline{x}, n)$ integrate the product of $dP(N_y, \underline{Y})$ (6) and $dP(\underline{X}, n \mid N_y, \underline{Y})$ (11) over $\underline{t}^y \in \times_{i=1}^{m}[0, T]$. The substitution of this result into (38) yields:

$$
\begin{aligned}
&dP(\underline{y}, m, \underline{x}, n) \\
&\quad = P_{N_x(T)}(n) \prod_{i=1}^{n} f_x(x_i) \sum_{\underline{d}: \underline{d}^T \underline{1} = m} \prod_{i=1}^{m} f(y_i \mid x_{\gamma(i)}) \prod_{i=1}^{n} \\
&\qquad \cdot p_F^{1-d_i}(x_i) \left(1 - p_F(x_i)\right)^{d_i} d\underline{y}\, d\underline{x}.
\end{aligned} \tag{39}
$$

Substitution of (39) and (12) into (38) gives:

$$
\begin{aligned}
&E[\ln l_m(\underline{X}, n)] \\
&\quad = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \int d\underline{x} \int d\underline{y}\, P_{N_x(T)}(n) \prod_{i=1}^{n} f_x(x_i) \\
&\qquad \cdot \sum_{\underline{d}: \underline{d}^T \underline{1} = m} \prod_{i=1}^{m} f(y_i \mid x_{\gamma(i)}) \prod_{i=1}^{n} \\
&\qquad \cdot p_F^{1-d_i}(x_i) \left(1 - p_F(x_i)\right)^{d_i} \\
&\qquad \times \ln \sum_{\underline{d}: \underline{d}^T \underline{1} = m} \prod_{i=1}^{m} f(y_i \mid x_{\gamma(i)}) \prod_{i=1}^{n} \\
&\qquad \cdot p_F^{1-d_i}(x_i) \left(1 - p_F(x_i)\right)^{d_i} \\
&\quad = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \int d\underline{x} \int d\underline{y}\, P_{N_x(T)}(n) \\
&\qquad \prod_{i=1}^{n} f_x(x_i) \left(\sum_{\underline{d}: \underline{d}^T \underline{1} = m} q_{\underline{d}} \ln \sum_{\underline{d}: \underline{d}^T \underline{1} = m} q_{\underline{d}}\right)
\end{aligned} \tag{40}
$$

where $q_{\underline{d}}$ is defined as the quantity:

$$
q_{\underline{d}} \stackrel{\text{def}}{=} \prod_{i=1}^{m} f(y_i \mid x_{\gamma(i)}) \prod_{i=1}^{n} p_F^{1-d_i}(x_i) \left(1 - p_F(x_i)\right)^{d_i}. \tag{41}
$$

We next invoke the *log-sum inequality* [12, Lemma 3.1]:

$$\sum_k q_k \ln \frac{\sum_k q_k}{\sum_k p_k} \le \sum_k q_k \ln \frac{q_k}{p_k} \tag{42}$$

where $\{q_k\}$ and $\{p_k\}$ are arbitrary, nonnegative, indexed sets of numbers. Letting the index $k = k(\underline{d})$ in (42) range over the $\binom{n}{m}$ indexes of $\Sigma_{\underline{d}:\underline{d}^T\underline{1}=m}$ in the integrand (40) and setting:

$$p_{k(\underline{d})} \stackrel{\text{def}}{=} \prod_{i=1}^{n} p_F^{1-d_i}(x_i)\left(1 - p_F(x_i)\right)^{d_i}, \tag{43}$$

one obtains the following bound on the right hand side of the equality (40):

$$\begin{aligned} &E\left[\ln l_m(\underline{X}, n)\right] \\ &\le \sum_{n=0}^{\infty}\sum_{m=0}^{n}\int d\underline{x}\int d\underline{y}\, P_{N_x(T)}(n)\prod_{i=1}^{n} f_x(x_i)\sum_{\underline{d}:\underline{d}^T\underline{1}=m} \\ &\quad \cdot q_{\underline{d}}\left[\ln\frac{q_{\underline{d}}}{p_{\underline{d}}} + \sum_{\underline{d}:\underline{d}^T\underline{1}=m}\ln p_{\underline{d}}\right] \\ &= \int dP(\underline{y}, m, \underline{x}, n)\left[\ln\frac{q_{\underline{d}}}{p_{\underline{d}}} + \ln p_{\underline{d}}\right] \\ &= E\left[\ln\frac{q_{\underline{d}}}{p_{\underline{d}}}\right] + E\left[\ln\sum_{\underline{d}:\underline{d}^T\underline{1}=m} p_{\underline{d}}\right] \end{aligned} \tag{44}$$

Using the expressions (41) and (43) for $q_{\underline{d}}$ and $p_{\underline{d}}$ with n and m replaced by $\boldsymbol{n}$ and $\boldsymbol{m}$:

$$\begin{aligned} E\left[\ln\frac{q_{\underline{d}}}{p_{\underline{d}}}\right] &= E\left[\ln\prod_{i=1}^{m} f\left(Y_i \middle| X_{d(i)}\right)\right] \\ &= \Lambda(1-\bar{p}_F)\, H\left(Y_i \middle| X_i\right), \end{aligned} \tag{45}$$

and, observing that the sum in the argument of the "ln" on the right side of equality (45) is the conditional probability of $\boldsymbol{m}$ given $\underline{X}$ [8]:

$$\begin{aligned} E\left[\ln\sum_{\underline{d}:\underline{d}^T\underline{1}=m} p_{\underline{d}}\right] &= E\left[\ln\sum_{\underline{d}:\underline{d}^T\underline{1}=m}\prod_{i=1}^{n} p_F^{1-d_i}(X_i)\right. \\ &\qquad \left.\cdot\left(1-p_F(x_i)\right)^{d_i}\right] \\ &= E\left[\ln P\left(\boldsymbol{m}\middle|\underline{X}\right)\right] \\ &= -H\left(\boldsymbol{m}\middle|\underline{X}\right). \end{aligned} \tag{46}$$

Substitution of (45) and (46) into (44), yields the upper bound:

$$E\left[\ln l_m(\underline{X}, n)\right] \le \Lambda(1-\bar{p}_F)\, H\left(Y_i\middle|X_i\right) - H\left(\boldsymbol{m}\middle|\underline{X}\right),$$

which, when substituted into (37) establishes (36).

The identity (24) follows immediately from the identities (16), (17), and (21):

$$\begin{aligned} I_o - I_1 &= H(\boldsymbol{n}) - H(\boldsymbol{m}) + \Lambda\left[I(X_i; W_i)\right. \\ &\qquad \left. - (1-\bar{p}_F)\, I(X_i; Y_i)\right] \\ &= H(\boldsymbol{n}) - H(\boldsymbol{m}) + \Lambda D\left(p_F(x)\middle\|\bar{p}_F\right). \end{aligned}$$

Since the entropy of a Poisson variable is monotone nondecreasing in the rate parameter [8, Appendix E]: $H(\boldsymbol{n}) - H(\boldsymbol{m}) = h_{\text{Poisson}}(\Lambda) - h_{\text{Poisson}}([1-\bar{p}_F]\Lambda)$ is nonnegative. Furthermore the divergence $D\left(p_F(x)\middle\|\bar{p}_F\right)$ is nonnegative [12]. Hence the difference $I_o - I_1$ cannot be less than zero and inequality (28) is established.

References

[1] L. Shao, A. O. Hero, W. L. Rogers, and N. H. Clinthorne, "The mutual information criterion for SPECT aperture evaluation and design," *IEEE Trans. Med. Imaging*, vol. 8, pp. 322–336, Dec. 1989.

[2] D. A. Mankoff, G. Muehllehner, and J. S. Karp, "The effect of detector performance on high count rate PET imaging with a tomograph based on position sensitive detectors," *IEEE Trans. Nucl. Sci.*, vol. 35, no. 1, pp. 592–596, Feb. 1988.

[3] R. R. Mohler and C. S. Hwang, "Nonlinear data observability and information," *J. Franklin Instit.*, vol. 325, no. 4, pp. 443–464, 1988.

[4] A. Tarantola and B. Valette, "Inverse problems = a quest for information," *J. Geophys.*, vol. 50, pp. 159–170, 1982.

[5] R. F. Wagner, D. G. Brown, and M. S. Pastel, "Application of information theory to the assessment of computed tomography," *Med. Phys.*, vol. 6, no. 2, pp. 83–94, Mar./Apr. 1979.

[6] N. J. Mars and G. W. Van-Arragon, "Time delay estimation in nonlinear systems using average amount of mutual information analysis," *Sig. Proc.*, vol. 4, no. 2, pp. 139–153, Apr. 1982.

[7] M. Fuderer, "The information content of MR images," *IEEE Trans. Med. Imaging*, vol. 7, no. 4, pp. 368–380, Dec. 1988.

[8] L. Shao, "Mutual information optimization and evaluation of single photon emission computed tomography," Ph.D. dissertation, Univ. of Michigan, Ann Arbor, MI 48109, Oct. 1989.

[9] D. L. Snyder, *Random Point Processes*. New York: Wiley, 1975.

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

[11] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-1, pp. 113–122, Oct. 1982.

[12] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL: Academic Press, 1981.

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

[14] I. Rubin, "Information rates for Poisson sequences," *IEEE Trans. Info. Theory*, pp. 283–294, May 1973.

[15] A. M. Mood, F. A. Greybill, and D. C. Boes, *Introduction to the Theory of Statistics*. New York: McGraw Hill, 1976.