Recovering Photon Intensity Information from Continuous Photo-Detector Measurements[1]

Alfred O. Hero

Dept. of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109

ABSTRACT

The problem of estimating parameters of a photon intensity image from continuous spatio-temporal photo-detector measurements is considered. The photo-detector output is modeled as a spatio-temporal random field consisting of two additive components: a superposition of single photon responses associated with photons incident on the photo-detector surface; and a broadband thermal noise. In this paper we derive bounds on estimation MSE based on rate distortion theory. The rate distortion bound simplifies considerably for the case of position estimation, spherically symmetric photon intensity, and large detector surface area. For this case we can identify the form of a SNR threshold which separates performance into two regimes: the *photon-noise-limited regime*, where estimator performance equal to that achievable for noiseless direct photon detection is possible; and the *thermal-noise-limited regime*, where such direct detection estimator performance is not achievable. An explicit limiting form for this threshold is given for special case of a Gaussian spatial beamshape.

I. Introduction

In most treatments of quantum limited optical signal processing it is assumed that error-free direct observations of the photon point process incident on the photo-detectors are available. For direct photon observations, optimal estimators of photon intensity parameters and lower bounds on estimator performance can easily be derived. However, practical photon measurement systems, e.g. photo-multiplier tubes or CCD arrays, consist of a photo-detector whose single photon-response (SPR) is temporally and spatially bandlimited and is contaminated by thermal noise. For this continuous observation regime methods developed for the photon observation regime are not directly applicable.

In this paper we study the problem of recovering photon intensity information from measurements of the bandlimited and noise contaminated photo-detector output. We derive general lower bounds on estimator MSE in order to study the impact of finite bandwidth and thermal noise on achievable error. These bounds simplify considerably for the case of beam position estimation, spherically symmetric beam shape, and large detector surface area. For this case we can identify a SNR threshold which separates performance into two regimes: the *photon-noise-limited regime*, where estimator performance equal to that achievable for noiseless direct photon detection is possible; and the *thermal-noise-limited regime*, where such direct detection estimator performance is not achievable.

Our photo-detector model is a spatio-temporal extension of the one-dimensional temporal model developed in [9, 7] for timing estimation in PET imaging. This one-dimensional model has been experimentally validated for a photo-detector system consisting of a Burle 8575 photo-multiplier tube (PMT) and BGO scintillator in [12]. Our measurement model is similar to models proposed in other applications, e.g. [2, 11, 5], for which the results of this paper may also be applicable.

II. Problem Statement

Let $\underline{\tau} = [\tau_1, \ldots, \tau_L]^T \in \mathcal{T} \subset \mathbb{R}^L$ be an unknown parameter vector with probability density function $p_{\underline{\tau}}$. Define the spatio-temporal coordinates $(t, \underline{u}) \in I$, $I \stackrel{\text{def}}{=} [-\frac{T}{2}, \frac{T}{2}] \times A$, of photon incidences on the surface $A = [-a, a] \times [-a, a]$ of a photodetector over a time interval $[-\frac{T}{2}, \frac{T}{2}]$. Conditioned on $\underline{\tau}$, let $\lambda = \{\lambda(t, \underline{u}|\underline{\tau}) : (t, \underline{u}) \in I\}$ be the intensity of the photon point process $d\mathrm{N} = \{d\mathrm{N}(t, \underline{u}) : (t, \underline{u}) \in I\}$. We assume that the point process $d\mathrm{N}$, equivalently the counting process N, is conditionally Poisson given $\underline{\tau}$. Define $\mathbf{n} = \mathrm{N}(I)$ the total number of incident photons over I. Let $\{(\mathbf{t}_i, \underline{\mathbf{u}}_i)\}_{i=1}^{\mathbf{n}}$ be the $\mathbf{n}$ coordinates of these photons. Let $\mathbf{g}_i$ denote the total induced charge on a photo-detector resulting from a photon interaction at spatial position $\underline{\mathbf{u}}_i$ and at time $\mathbf{t}_i$. Conditioned on $\mathbf{n}$, $\{\mathbf{g}_i\}_{i=1}^{\mathbf{n}}$ is a sequence of i.i.d. random variables with range $\mathcal{G} = \mathbb{R}$ and probability density p_g. The mean and variance of the density p_g specify the mean efficiency and energy spread, respectively, of the photon-collection process. In this paper we assume that the $\mathbf{g}_i$'s are statistically independent of $\underline{\tau}$ given $\mathbf{n}$. It is also assumed that the intensity $\lambda(t, \underline{u}|\underline{\tau})$ is constant in t but varying in $\underline{u}$ and that, conditioned on $\underline{\tau}$, the integrated intensity $\Lambda \stackrel{\text{def}}{=} \int_I \lambda = E[\mathbf{n}|\underline{\tau}]$ of N is functionally independent of $\underline{\tau}$. The former assumption corresponds to a point process conditionally homogeneous over time but inhomogeneous over space; modeling the problem of imaging a temporally stationary image source. The latter assumption, called conditional energy invariance in [8], is appropriate in cases where the average rate Λ is known *a priori* and the spatial support $\{\underline{u} \in \mathbb{R}^2 : \lambda(t, \underline{u}|\underline{\tau}) > 0\}$ is contained in A for all $\underline{\tau}$. For example this holds for beam spatial-position estimation, where $\lambda(t, \underline{u}|\underline{\tau}) = \lambda(t, \underline{u} - \underline{\tau}|\underline{0})$, to be considered in the sequel.

The sequence $\{(\mathbf{t}_i, \underline{\mathbf{u}}_i), \mathbf{g}_i\}_{i=1}^{\mathbf{n}}$ of photon arrival coordinates $(\mathbf{t}_i, \underline{\mathbf{u}}_i) \in [-\frac{T}{2}, \frac{T}{2}] \times A$ and photo-detector gains $\mathbf{g}_i \in \mathcal{G}$ defines a marked point process $d\mathrm{M}$ with index set $I = [-\frac{T}{2}, \frac{T}{2}] \times A$ and mark space $\mathcal{G}$. In the sequel we refer to the process M as the *direct photon detection data* or the *photon-noise-limited regime*. By contrast the actual measurement corresponds to the following finite bandwidth and thermal noise contaminated observations over $(t, \underline{u}) \in I$:

[1] This research was supported in part by the National Cancer Institute under grant R01-CA-46622-01.

$$\begin{aligned}\mathrm{X}(t,\underline{u}) &= \sum_{i=1}^{\mathrm{n}} \mathrm{g}_i p(t-\mathrm{t}_i, \underline{u}-\underline{\mathrm{u}}_i) + \mathrm{w}(t,\underline{u}) \qquad (1)\\ &= \iiint_I p(t-v, \underline{u}-\underline{\xi}) d\mathrm{M}(v,\underline{\xi}) + \mathrm{w}(t,\underline{u}),\end{aligned}$$

where w is a spatio-temporal white Gaussian noise. In (1) $p(t,\underline{u})$ is a unit energy SPR pulse with standard temporal width $T_p \ll T$ and standard spatial width $\sigma_p^2 \ll |A|$.

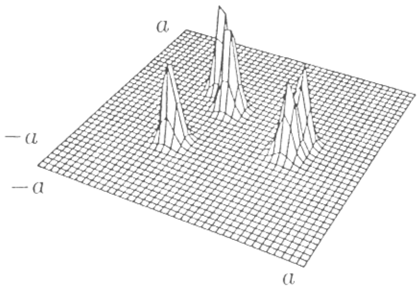

Figure 1: Snapshot of X at time $t \in [0,T]$.

The estimation objective is to determine the value of the vector parameter $\underline{\tau}$ based on the photo-detector waveform output X given in (1). For example, the MAP estimator $\hat{\underline{\tau}}^{MAP} = \hat{\underline{\tau}}^{MAP}(\mathrm{X})$ is obtained by maximizing the *a posteriori* log-likelihood function $\ln p_{\mathrm{X}|\underline{\tau}} p_{\underline{\tau}}$ over $\underline{\tau}$, where $p_{\mathrm{X}|\underline{\tau}}$ is the probability density of X given $\underline{\tau}$. Unfortunately, while a number of approximations have been proposed [9, 4], no analytical form for $p_{\mathrm{X}|\underline{\tau}}$ has been found to date. This also makes it difficult to derive analytical lower bounds on estimator MSE useful for establishing performance losses associated with a practical but perhaps sub-optimal estimator. This is to be contrasted with the case of direct photon detection for which the log-likelihood function $\ln p_{\mathrm{M}|\underline{\tau}}$ has an analytical form. Specifically, the direct detection log- likelihood function is specified by the joint densities of $\{(\mathrm{t}_i, \underline{\mathrm{u}}_i), \mathrm{g}_i\}_{i=1}^{\mathrm{n}}$:

$$p_{\mathrm{M}|\underline{\tau}}(\mathrm{M}|\underline{\tau}) = \begin{cases} \prod_{i=1}^{n} \lambda(t_i, \underline{u}_i|\underline{\tau}) f_g(g_i)\, e^{-\Lambda}, & n>0 \\ e^{-\Lambda}, & n=0 \end{cases}. \qquad (2)$$

III. Lower Bounds On MSE

In this section we develop lower bounds on the mean of component MSE's of any estimator $\hat{\underline{\tau}} = \hat{\underline{\tau}}(\mathrm{X})$ of $\underline{\tau}$:

$$\overline{\mathrm{MSE}}(\mathrm{X}) \stackrel{\text{def}}{=} \frac{1}{L}\sum_{i=1}^{L} E[(\tau_i - \hat{\tau}_i(\mathrm{X}))^2], \qquad (3)$$

III.1 Cramer-Rao Lower Bounds

Under some smoothness conditions on $p_{\mathrm{X}|\underline{\tau}}$ the CR bound on the MSE of any estimator $\hat{\tau}_1 = \hat{\tau}_1(\mathrm{X})$ for τ_1 is [1]:

$$\begin{aligned} E[(\tau_1 - \hat{\tau}_1(\mathrm{X}))^2] &\geq B_{cr}^1(\mathrm{X}) \qquad (4)\\ B_{cr}^1(\mathrm{X}) &\stackrel{\text{def}}{=} \left(E\left[\left(\frac{\partial}{\partial \tau_1}\ln p_{\mathrm{X},\tau_1}\right)^2\right]\right)^{-1}\\ &= -\left(E\left[\frac{\partial^2}{\partial \tau_1^2}\ln p_{\mathrm{X},\tau_1}\right]\right)^{-1},\end{aligned}$$

where $p_{\mathrm{X},\tau_1} = \int d\tau_2 \cdots \int d\tau_L\, p_{\mathrm{X}|\underline{\tau}} p_{\underline{\tau}}$. Since this function is not of analytical form the CR bound (4) cannot be computed. On the other hand, for direct photon observations the CR bound on the MSE of any estimator $\hat{\tau}_1 = \hat{\tau}_1(\mathrm{M})$ is:

$$\begin{aligned} E[(\tau_1 - \hat{\tau}_1(\mathrm{M}))^2] &\geq B_{cr}^1(\mathrm{M}) \qquad (5)\\ B_{cr}^1(\mathrm{M}) &\stackrel{\text{def}}{=} \left(E\left[\left(\frac{\partial}{\partial \tau_1}\ln p_{\mathrm{M},\tau_1}\right)^2\right]\right)^{-1}\\ &= -\left(E\left[\frac{\partial^2}{\partial \tau_1^2}\ln p_{\mathrm{M},\tau_1}\right]\right)^{-1},\end{aligned}$$

where $p_{\mathrm{M},\tau_1} = \int d\tau_2 \cdots \int d\tau_L\, p_{\mathrm{M}|\underline{\tau}} p_{\underline{\tau}}$. When $p_{\underline{\tau}}$ is uniform and the intensity λ is separable in τ_1 and $\tau_2, \ldots, \tau_L$, i.e when $\lambda(t,\underline{u}|\underline{\tau}) = \lambda(t,\underline{u}|\tau_1)\lambda(t,\underline{u}|\tau_2,\ldots,\tau_L)$, it is easily shown using (2) that the CR bound (5) reduces to:

$$\begin{aligned} B_{cr}^1(\mathrm{M}) &= \left(\iiint_I dt d\underline{u}\, E\left[\left(\frac{\partial \ln \lambda(t,\underline{u}|\underline{\tau})}{\partial \tau_1}\right)^2 \lambda(t,\underline{u}|\underline{\tau})\right]\right)^{-1} \qquad (6)\\ &= -\left(\iiint_I dt d\underline{u}\, E\left[\frac{\partial^2 \ln \lambda(t,\underline{u}|\underline{\tau})}{\partial \tau_1^2}\lambda(t,\underline{u}|\underline{\tau})\right]\right)^{-1}.\end{aligned}$$

It can also be shown that $B_{cr}^1(\mathrm{M}) = B_{cr}^1(\mathrm{N})$, i.e. the observations M are no more informative than N. This is a consequence of our assumption that the g_i's are independent of $\underline{\tau}$ given n. Note that if $\ln\lambda$ is not sufficiently smooth its derivative may not square integrable in which case (6) will yield the trivial bound $B_{cr}^1 = 0$.

The following Lemma states the fact that the bound $B_{cr}^1(\mathrm{M})$ for the photon-noise-limited regime can be used as a weaker bound than $B_{cr}^1(\mathrm{X})$ for the thermal-noise-limited regime.

Lemma 1 *Let τ_1 be a random parameter. Let X and M be two measurement processes such that M is a refinement of X in the sense that the conditional density $p_{\mathrm{X}|\mathrm{M},\tau_1}$ is functionally independent of τ_1. Then, assuming the CR bounds $B_{cr}^1(\mathrm{X})$ (4) and $B_{cr}^1(\mathrm{M})$ (5) exist: $B_{cr}^1(\mathrm{X}) \geq B_{cr}^1(\mathrm{M})$. Thus:*

$$E[(\tau_1 - \hat{\tau}_1(\mathrm{X}))^2] \geq B_{cr}^1(\mathrm{M}). \qquad (7)$$

Lemma 1 can be established along the lines of the following argument:

$$\begin{aligned} 0 &\leq E\left[\left(\frac{\partial}{\partial \tau_1}\ln p_{\mathrm{M}|\mathrm{X},\tau_1}\right)^2\right]\\ &= E\left[-\frac{\partial^2}{\partial \tau_1^2}\ln p_{\mathrm{M}|\mathrm{X},\tau_1}\right]\\ &= E\left[-\frac{\partial^2}{\partial \tau_1^2}\ln \frac{p_{\mathrm{X}|\mathrm{M},\tau_1} p_{\mathrm{M},\tau_1}}{p_{\mathrm{X},\tau_1}}\right]\\ &= E\left[-\frac{\partial^2}{\partial \tau_1^2}\ln p_{\mathrm{M},\tau_1}\right] - E\left[-\frac{\partial^2}{\partial \tau_1^2}\ln p_{\mathrm{X},\tau_1}\right]\\ &\quad + E\left[-\frac{\partial^2}{\partial \tau_1^2}\ln p_{\mathrm{X}|\mathrm{M},\tau_1}\right]\\ &= [B_{cr}^1(\mathrm{M})]^{-1} - [B_{cr}^1(\mathrm{X})]^{-1},\end{aligned}$$

where in the second to last line we used the fact that $\frac{\partial}{\partial \tau_1} \ln p_{\mathbf{X}|\mathbf{M},\tau_1}$ is zero under the assumptions of the lemma.

With $B^i_{cr}(\mathbf{M})$ the corresponding CR bound (5) on $E[(\tau_i - \hat{\tau}_i(\mathbf{M}))^2]$ we have the following bound on $\overline{\mathrm{MSE}}$ (3):

$$\overline{\mathrm{MSE}}(\mathbf{X}) \geq B_{cr}(\mathbf{M}) \stackrel{\text{def}}{=} \frac{1}{L} \sum_{i=1}^{L} B^i_{cr}(\mathbf{M}). \tag{8}$$

III.2 Rate Distortion Lower Bounds

The rate distortion bound is based on Shannon's fundamental result:

$$\inf_{P_{\mathbf{Z}|\mathbf{V}}: \overline{\rho}(\mathbf{V},\mathbf{Z}) \leq d} I(\mathbf{V};\mathbf{Z}) = R_{\overline{\rho}}(d) \leq C = \sup_{P_{\mathbf{V}}} I(\mathbf{V};\mathbf{Z}), \tag{9}$$

where $I(\mathbf{V};\mathbf{Z}) \stackrel{\text{def}}{=} E[\ln P_{\mathbf{Z}|\mathbf{V}}(\mathbf{Z}|\mathbf{V})/P_{\mathbf{Z}}(\mathbf{Z})]$ is the mutual information between source process $\mathbf{V}$ and observed process $\mathbf{Z}$. In (9) C is the channel capacity and $R_{\overline{\rho}}$ is the rate distortion function associated with the average distortion $\overline{\rho}$. Here $\mathbf{Z}$ is identified with $\mathbf{X}$, $\mathbf{V}$ is identified with $\underline{\tau}$, and $\overline{\rho}$ is identified with $\overline{\mathrm{MSE}}$ defined in (3). Defining the entropy function $H(\underline{\tau}) = -E[\ln p_{\underline{\tau}}(\underline{\tau})]$, for this case Shannon's lower bound [6] gives: $R_{\overline{\rho}}(d) \geq H(\underline{\tau}) - \frac{L}{2}\ln(2\pi e d)$. Due to monotonicity in d, the combination of this lower bound on $R_{\overline{\rho}}(d)$ and the bound (9) gives the following rate distortion lower bound on $\overline{\mathrm{MSE}}$:

$$\overline{\mathrm{MSE}}(\mathbf{X}) \geq \frac{1}{2\pi e} e^{\frac{2}{L}H(\underline{\tau})} e^{-\frac{2}{L}C}. \tag{10}$$

For similar reasons as before, an exact expression for the bound (10) is intractable, requiring evaluation of $p_{\mathbf{X}|\underline{\tau}}$ to compute the mutual information. To arrive at a tractable bound, as in [7] we use the "data processing theorem" applied to the cascaded channel representation of C shown in Fig. 2. Referring to the figure, the channel C is a cascade of two separate channels C_1 and C_2, where C_1 maps the source symbols $\underline{\tau}$ into the direct photon-detection data $\mathbf{M}$, and C_2 maps $\mathbf{M}$ into the observations $\mathbf{X}$. C_1 is thus a point process channel with associated intensity λ and mark distribution f_g, while C_2 is an additive Gaussian noise channel. The data processing theorem gives an upper bound on the overall capacity C of this cascaded channel [6]:

$$C \leq \min\{C_1, C_2\}. \tag{11}$$

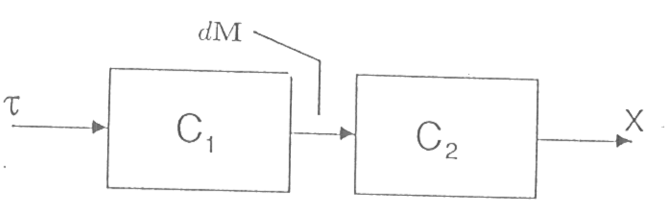

Figure 2: Cascaded Channel Representation for $\underline{\tau} \rightarrow \mathbf{X}$

Since the $\mathbf{g}_i$'s are independent of $\underline{\tau}$ given $\mathbf{n}$, $I(\mathbf{M},\underline{\tau}) = I(\mathbf{N},\underline{\tau})$ and we can apply an upper bound, C_1^*, on C_1 analogous to that obtained in [8, Lemma 4]:

$$C_1^* = \int d\underline{\tau}\, p^o(\underline{\tau}) \iiint_I dt d\underline{u}\, \lambda(t,\underline{u}|\underline{\tau}) \ln \frac{\lambda(t,\underline{u}|\underline{\tau})}{\overline{\lambda}^o(t,\underline{u}|\underline{\tau})} \tag{12}$$

$$\overline{\lambda}^o(t,\underline{u}|\underline{\tau}) \stackrel{\text{def}}{=} \int_T d\underline{\tau}\, p^o(\underline{\tau}) \lambda(t,\underline{u}|\underline{\tau}),$$

where p^o is a (density) function of $\underline{\tau}$ arising from a set of Kuhn-Tucker conditions.

An explicit upper bound C_2^* on C_2 is obtained by using a simple fact: a continuous state channel with covariance constrained output has capacity which is upper bounded by a similarly constrained Gaussian channel. This fact follows from the entropy decomposition formula for $I(\mathbf{M};\mathbf{X})$ where the channel output $\mathbf{X}$ is the sum of a "signal" $\mathbf{S}(t,\underline{u}) \stackrel{\text{def}}{=} \sum_{i=1}^{\mathbf{n}} \mathbf{g}_i p(t - \mathbf{t}_i, \underline{u} - \underline{\mathbf{u}}_i)$ plus an independent additive Gaussian noise $\mathbf{w}$:

$$\begin{aligned} C_2 &= \sup_{p_{\mathbf{M}}} \{H(\mathbf{X}) - H(\mathbf{X}|\mathbf{M})\} \\ &= \sup_{p_{\mathbf{M}}} \{H(\mathbf{X}) - H(\mathbf{w}|\mathbf{M})\} \\ &= \sup_{p_{\mathbf{M}}} \{H(\mathbf{X})\} - H(\mathbf{w}). \end{aligned} \tag{13}$$

In (13) we have used the facts: $\mathbf{S}$ is non-random given $\mathbf{M}$; $\mathbf{M}$ and $\mathbf{w}$ are independent; and $H(\mathbf{w})$ is functionally independent of $p_{\mathbf{M}}$. On the other hand, it is well known that, for a fixed output autocovariance function $K_X = K_s + K_w$, the entropy $H(\mathbf{X})$ is maximized for Gaussian $\mathbf{X}$. Hence the channel capacity C_2 is upper bounded by the capacity C_2^* of a Gaussian channel.

The autocovariance, $K_s(\underline{z}_1,\underline{z}_2) \stackrel{\text{def}}{=} \mathrm{cov}[\mathbf{S}(\underline{z}_1)\mathbf{S}(\underline{z}_2)]$, evaluated at two arbitrary spatio-temporal coordinates $\underline{z}_1 = (t_1,\underline{u}_1)$ and $\underline{z}_2 = (t_2,\underline{u}_2)$ in $I = [-\frac{T}{2},\frac{T}{2}] \times A$ is straightforwardly computed:

$$\begin{aligned} K_s(\underline{z}_1,\underline{z}_2) &= \mu_g^2 \iiint_I p(\underline{z}_1 - \underline{\zeta}) p(\underline{z}_2 - \underline{\zeta}) \overline{\lambda}(\underline{\zeta}) d\underline{\zeta} \\ &+ (\mu_g^2 + \sigma_g^2) \iiint_I d\underline{\zeta}_1 \iiint_I d\underline{\zeta}_2 p(\underline{z}_1 - \underline{\zeta}_1) p(\underline{z}_2 - \underline{\zeta}_2) K_\lambda(\underline{\zeta}_1,\underline{\zeta}_2), \end{aligned} \tag{14}$$

where:

$$K_\lambda(\underline{z}_1,\underline{z}_2) \stackrel{\text{def}}{=} \mathrm{cov}[\lambda(\underline{z}_1|\underline{\tau}),\lambda(\underline{z}_2|\underline{\tau})], \tag{15}$$

and $\overline{\lambda}(\underline{z}) \stackrel{\text{def}}{=} E[\lambda(\underline{z}|\underline{\tau})]$ is the intensity averaged over $\underline{\tau}$.

Using a Karhunen-Loeve expansion of $\mathbf{X}$ in terms of the eigenfunctions $\Psi(\underline{z})$ of $K_s(\underline{z}_1,\underline{z}_2)$, the channel capacity C_2^* can be shown to have the form [6]:

$$C_2^* = \frac{1}{2}\sum_{i=1}^{\infty} \ln\left(1 + \frac{2\eta_i}{N_o}\right) \tag{16}$$

where $N_o/2$ is the white noise spectral level, $|I| = T|A|$ is the volume of I, and η_i, $i = 1,2,\ldots$, are the eigenvalues of K_s:

$$\eta_i \Psi(\underline{z}) = \iiint_I K_s(\underline{z},\underline{\xi}) \Psi(\underline{\xi}) d\underline{\xi}.$$

Combination of (11) and the upper bounds (12) and (16) on C_1 and C_2 gives the following rate distortion lower bound on $\overline{\mathrm{MSE}}$ defined in (3):

$$\overline{\mathrm{MSE}}(\mathbf{X}) \geq B_{rdlb}(\mathbf{X}) \stackrel{\text{def}}{=} \frac{1}{2\pi e} e^{\frac{2}{L}H(\underline{\tau})} e^{-\frac{2}{L}\min\{C_1^*, C_2^*\}}. \quad (17)$$

III.3 Homogeneous and Isotropic Fields

To obtain a tractable analysis of C_2^* we specialize to the case of homogeneous isotropic K_s over $[0,T] \times A$, large observation time T, and large photo-detector surface area $|A|$. If $K(\underline{z}_1, \underline{z}_2)$ depends only on the difference $\underline{z}_1 - \underline{z}_2$ then K_s is said to be homogeneous. In this case we abuse notation and write $K_s(\underline{\zeta} + \underline{z}, \underline{\zeta}) = K_s(\underline{z})$. If furthermore $K_s(\underline{z})$ depends only on the magnitude $\|\underline{z}\|$ then K_s is said to be isotropic. The homogeneity and isotropy of K_s will depend on the form of $\lambda(t, \underline{u}|\underline{\tau})$ and $p(t, \underline{u})$, the size of the index set I relative to the standard widths of $\lambda(t, \underline{u}|\underline{\tau})$ and $p(t, \underline{u})$, and the prior $p_{\underline{\tau}}$. Note that this does not require the unreasonable properties that the field $\mathbf{X}$ be either conditionally homogeneous or conditionally isotropic given $\underline{\tau}$.

For homogeneous K_s the theory of Toeplitz forms provides a spectral domain description of the capacity C_2^* (16) in terms of the limiting capacity-per-unit-volume $\overline{C}_2^* \stackrel{\text{def}}{=} \lim_{a,T\to\infty} C_2^*/|I|$ (recall that $I = [-\frac{T}{2}, \frac{T}{2}] \times A$, $A = [-a, a] \times [-a, a]$). In particular [13]:

$$\overline{C}_2^* = \frac{1}{2} \iiint_{-\infty}^{\infty} df\, d\nu_1 d\nu_2 \ln\left(1 + \frac{2\, G_s(f, \nu_1, \nu_2)}{N_o}\right), \quad (18)$$

where G_s is the power spectral density associated with $K_s(\underline{z})$:

$$G_s(f, \nu_1, \nu_2) \quad (19)$$
$$= \iiint_{-\infty}^{\infty} dt\, du_1 du_2 K_s(t, u_1, u_2) e^{-j2\pi(ft + \nu_1 u_1 + \nu_2 u_2)}.$$

In the Fourier transform (19) f is temporal frequency and ν_1, ν_2 are spatial frequencies. The limiting form (18) for the normalized capacity $C_2^*/|I|$ gives the large T and large a approximation:

$$C_2^* \approx |I| \overline{C}_2^* \quad (20)$$

to be used in the sequel.

Under the assumption that $p(t, \underline{u})$ is a decaying mono-exponential over t:

$$p(t, \underline{u}) = \begin{cases} \frac{1}{T_p} e^{-t/T_p} \cdot p(\underline{u}), & t \geq 0 \\ 0, & t < 0 \end{cases}, \quad (21)$$

further simplification of $\overline{C}_2^*$ can be made by using the integral identity:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \ln\left(1 + \frac{\alpha}{1 + (w/\rho)^2}\right) dw = \frac{\rho}{2}\left[\sqrt{1+\alpha} - 1\right]. \quad (22)$$

Specifically, applying this identity to (18) we obtain:

$$\overline{C}_2^* = \frac{1}{2T_p} \iint_{-\infty}^{\infty} \left[\sqrt{1 + Q(\nu_1, \nu_2)} - 1\right] d\nu_1 d\nu_2, \quad (23)$$

where

$$Q(\nu_1, \nu_2) \stackrel{\text{def}}{=} \frac{2}{N_o} \int_{-\infty}^{\infty} G_s(f, \nu_1, \nu_2) df \quad (24)$$

is the noise-power normalized spatial spectrum of $\mathbf{X}$.

Finally, if Q is spherically symmetric, $Q(\nu_1, \nu_2) = \tilde{Q}(r)$ where $r = \sqrt{\nu_1^2 + \nu_2^2}$, as occurs for a spatially isotropic random field, the double integral in (23) can be reduced to a single integral via a rectangular-to-polar transformation of variables $(\nu_1, \nu_2) \to (r, \theta)$. This results in:

$$\overline{C}_2^* = \frac{\pi}{T_p} \int_{-\infty}^{\infty} r \left[\sqrt{1 + \tilde{Q}(r)} - 1\right] dr. \quad (25)$$

In (25) the spherically symmetric spatial spectrum $\tilde{Q}$ can either be calculated directly from the 3-D Fourier transform G_s of K_s using (24), or from the 1-D Hankel transform of $\tilde{K}_s(v) \stackrel{\text{def}}{=} K_s(0, \underline{u})$ where $v \stackrel{\text{def}}{=} \|\underline{u}\|$ [13]. Using (20) we thus have the following approximation to the unormalized capacity C_2^*:

$$C_2^* \approx \frac{\pi |I|}{T_p} \int_{-\infty}^{\infty} r \left[\sqrt{1 + \tilde{Q}(r)} - 1\right] dr. \quad (26)$$

which is accurate for large T and large a.

IV. Beam Localization

Here we specialize to the case where $L = 2$ and $\underline{\tau} = [\tau_1, \tau_2]^T$ corresponds to a spatial translation of λ:

$$\lambda(t, \underline{u}|\underline{\tau}) = \frac{1}{T} \lambda(\underline{u} - \underline{\tau}), \quad (27)$$

where $\lambda(\underline{u}) = \lambda(\|\underline{u}\|)$ is a known spherically symmetric spatial intensity with $\iint_A \lambda(\underline{u}) d\underline{u} = \Lambda$. We assume that $\underline{\tau}$ is uniformly distributed over $\mathcal{T} = A$ and that the detector surface $A = [-a, a] \times [-a, a]$ is large. The estimation of beam location $\underline{\tau}$ over the surface of a photo-detector is important in fields such as laser radar, star tracking, and sub-atomic particle detection [14, 10, 3]

For this beam localization problem several simplifications occur due to our assumptions on λ and $\underline{\tau}$.

First the CR bound (6) applies which reduces to:

$$B_{cr}^1(\mathbf{M}) = \left[\iint_A \left(\frac{d \ln \lambda(u_1)}{du_1}\right)^2 \lambda(u_1) du_1\right]^{-1}, \quad (28)$$

where

$$\lambda(u_1) = \int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{-a}^{a} \lambda(t, u_1, u_2|\underline{\tau}) dt du_2 = \int_{-a}^{a} \lambda(u_1, u_2) du_2.$$

Second, the rate distortion bound (17) simplifies due to the following: 1) the capacity C_1^* (12) reduces to [8, Sec. VI.A]:

$$C_1^* = \iint_A \lambda(\underline{u}) \ln \frac{\lambda(\underline{u})}{\overline{\lambda}(\underline{u})} d\underline{u}, \quad (29)$$

where the mean spatial intensity $\overline{\lambda}(\underline{u})$ in (14) is constant over A:

$$\overline{\lambda}(\underline{u}) = \begin{cases} \frac{\Lambda}{|A|}, & \underline{u} \in A \\ 0, & o.w. \end{cases}; \quad (30)$$

and 2) the covariance K_λ in (15) is homogeneous over I: $K_\lambda(\underline{z}_1, \underline{z}_2) = K_\lambda(\underline{z}_1 - \underline{z}_2)$, $\underline{z}_1, \underline{z}_2 \in I$. Consequently, under the large A assumption, the spectrum $G_s(f, \nu_1, \nu_2)$ can be derived using (14) for K_s and (21) for $p(t, u_1, u_2)$:

$$G_s(f, \nu_1, \nu_2) = \frac{1}{1 + (2\pi f T_p)^2} \cdot \gamma \hat{G}_s(\nu_1, \nu_2), \quad (31)$$

where $\sqrt{\gamma}$ is the rms single-photon-response-to-noise power (SNR) ratio:

$$\gamma \stackrel{\text{def}}{=} \frac{2(\mu_g^2 + \sigma_g^2)}{N_o},$$

μ_g and σ_g^2 are the mean and variance of the SPR gains $\mathbf{g}_i$. In (31) $\hat{G}_s$ is the SNR normalized spatial spectrum:

$$\hat{G}_s(\nu_1, \nu_2) \stackrel{\text{def}}{=} \quad (32)$$
$$|P(\nu_1, \nu_2)|^2 \left[\Lambda + \frac{\mu_g^2}{\mu_g^2 + \sigma_g^2} |\Lambda(\nu_1, \nu_2) - \overline{\Lambda}(\nu_1, \nu_2)|^2\right],$$

where $P(\nu_1, \nu_2)$ is the 2-D Fourier transform of the spatial component $p(\underline{u}) = p(u_1, u_2)$ of the SPR in (21), $\Lambda(\nu_1, \nu_2)$ is the 2-D Fourier transform of the unshifted spatial intensity $\lambda(\underline{u})$, and $\overline{\Lambda}(\nu_1, \nu_2)$ is the 2-D Fourier transform of the piecewise constant spatial intensity (30). As before $\Lambda \stackrel{\text{def}}{=} \Lambda(0, 0)$ is the integrated intensity. Applying the integral identity (22) with (31), we obtain (23) with $Q(\nu_1, \nu_2) = \gamma \hat{G}_s(\nu_1, \nu_2)$. Combining all of these results the rate distortion bound (17) takes the form:

$$B_{rdlb}(\mathbf{X}) \quad (33)$$
$$= \begin{cases} \frac{|A|}{2\pi e} e^{-\iint_A \lambda(\underline{u}) \ln \frac{\lambda(\underline{u})}{\overline{\lambda}(\underline{u})} d\underline{u}}, & \gamma > \gamma_o \\ \frac{|A|}{2\pi e} e^{-\frac{|I|}{2T_p} \iint_{-\infty}^{\infty} \left[\sqrt{1+\gamma \hat{G}_s(\underline{\nu})} - 1\right] d\underline{\nu}}, & \gamma \leq \gamma_o \end{cases},$$

where $\underline{\nu} = (\nu_1, \nu_2)$, and γ_o is a SNR threshold determined by the condition $C_1^* = C_2^*$ in (17). Specifically, γ_o is the solution $\gamma = \gamma_o$ of the equation:

$$\iint_A \lambda(\underline{u}) \ln \frac{\lambda(\underline{u})}{\overline{\lambda}(\underline{u})} d\underline{u} = \frac{|I|}{2T_p} \iint_{-\infty}^{\infty} \left[\sqrt{1 + \gamma \hat{G}_s(\underline{\nu})} - 1\right] d\underline{\nu}, \quad (34)$$

when the solution exists. An important case where the solution does not exist is when the spatial filter response approaches a delta function, i.e. $|P(\underline{\nu})| \rightarrow$ *constant*. In this case, since the right hand side of the above equation becomes unbounded, $C_1^* < C_2^*$, and hence $\min\{C_1^*, C_2^*\} = C_1^*$, for all $\gamma > 0$. Therefore $\gamma_o = 0$ and the lowerbound (33) is identical to the direct detection bound $B_{rdlb}(\mathbf{M}) = |A|/(2\pi e) \exp(-C_1^*)$.

The lower bound (33) separates the MSE performance into two SNR regimes: the photon-noise-limited regime ($\gamma > \gamma_o$); and the thermal-noise-limited regime ($\gamma \leq \gamma_o$). In the photon-noise-limited regime $B_{rdlb}(\mathbf{X})$ decreases exponentially as a function of the information divergence:

$$d(\lambda, \overline{\lambda}) \stackrel{\text{def}}{=} \iint_A \lambda(\underline{u}) \ln \frac{\lambda(\underline{u})}{\overline{\lambda}(\underline{u})} d\underline{u}$$

which measures the difference between the spatial intensities $\lambda(\underline{u})$ (27) and $\overline{\lambda}(\underline{u})$ (30). The closer $\lambda(\underline{u})$ is to the uninformative uniform intensity $\overline{\lambda}(\underline{u})$, the poorer becomes the estimator MSE. On the other hand, in the thermal-noise-limited-regime regime $B_{rdlb}(\mathbf{X})$ decreases exponentially in the rms deviation between the 2-D Fourier transforms of $\lambda(\underline{u})$ and $\overline{\lambda}(\underline{u})$. Observe also that the maximum value of $B_{rdlb}(\mathbf{X})$ is the "entropy power" $\frac{1}{2\pi e} e^{H(\underline{\tau})} = |A|/(2\pi e)$ which is slightly larger than the *a priori* variance $\sigma_{\tau_i}^2 = |A|/12$ of τ_1 and τ_2.

Gaussian Beam Model

Here we specialize to the following spatially symmetric Gaussian beam and SPR models:

$$\lambda(u_1, u_2) = \frac{\Lambda}{2\pi\sigma_\lambda^2} e^{-\frac{u_1^2 + u_2^2}{2\sigma_\lambda^2}}$$
$$p(u_1, u_2) = \frac{1}{2\pi\sigma_p^2} e^{-\frac{u_1^2 + u_2^2}{2\sigma_p^2}}.$$

From (29)

$$C_1^* = \Lambda \ln \left[\frac{|A|}{2\pi e \sigma_\lambda^2}\right]. \quad (35)$$

Next we derive an upper bound on the SNR threshold γ_o defined as the solution $\gamma = \gamma_o$ of the equation $C_1^* = C_2^*$ (34). Specifically, since $\mu_g^2/(\mu_g^2 + \sigma_g^2)|\Lambda(\underline{\nu}) - \overline{\Lambda}(\underline{\nu})|^2$ is non-negative, the solution of $C_1^* = C_2^*$ with $\hat{G}_s(\nu_1, \nu_2)$ (32) in C_2^* replaced by $|P(\nu_1, \nu_2)|^2 \Lambda$ gives an upper bound on γ_o. This upper bound becomes tight when $\mu_g^2/(\mu_g^2 + \sigma_g^2)|\Lambda(\underline{\nu}) - \overline{\Lambda}(\underline{\nu})|^2 \ll \Lambda$ which occurs for either: $\sigma_g^2 \gg \mu_g^2$, i.e wide variation in the amplitudes of the photo-responses, or σ_λ^2 is large, i.e. broad spatial intensity width. Making this substitution, $\hat{G}_s(\nu_1, \nu_2)$ becomes spherically symmetric and, identifying: $Q(\nu_1, \nu_2) = \gamma \hat{G}_s(\nu_1, \nu_2)$ and $|P(\nu_1, \nu_2)|^2 = \exp(-\sigma_p^2(2\pi r)^2)$, (26) applies:

$$C_2^* = \frac{\pi|I|}{T_p} \int_{-\infty}^{\infty} r \left[\sqrt{1 + \beta e^{-\sigma_p^2(2\pi r)^2}} - 1\right] dr, \quad (36)$$

where $\beta \stackrel{\text{def}}{=} \gamma\Lambda$. The integral in (36) can be expressed in analytical form:

$$C_2^* = \frac{|I|}{4\pi T_p \sigma_p^2} \left[\sqrt{1+\beta} - 1 + \frac{1}{2} \ln \left(\frac{4(\sqrt{1+\beta} - 1)}{\beta(\sqrt{1+\beta} + 1)}\right)\right]. \quad (37)$$

Asymptotic analysis of the right side of (37) gives the following large and small β representations:

$$C_2^* = \begin{cases} \sqrt{\beta} \cdot \frac{|I|}{4\pi T_p \sigma_p^2}[1 + o(1)], & \beta \gg 1 \\ \frac{\beta}{2} \cdot \frac{|I|}{4\pi T_p \sigma_p^2} + o(\beta), & \beta \ll 1 \end{cases} \quad (38)$$

where $o(1) \rightarrow 0$ as $\beta \rightarrow \infty$ and $\frac{1}{\beta} o(\beta) \rightarrow 0$ as $\beta \rightarrow 0$. Using $\beta = \gamma\Lambda$, equating (35) and (38), and solving for $\gamma = \gamma_o$ yields the following upper bound on the SNR threshold γ_o:

$$\gamma_o \le \begin{cases} (4\pi\frac{T_p\sigma_p^2}{|I|})^2 \cdot \left(\ln\left[\frac{|A|}{2\pi e\sigma_\lambda^2}\right]\right)^2 \cdot \Lambda, & \gamma_o\Lambda \gg 1 \\ 8\pi\frac{T_p\sigma_p^2}{|I|} \cdot \ln\left[\frac{|A|}{2\pi e\sigma_\lambda^2}\right], & \gamma_o\Lambda \ll 1 \end{cases}. \quad (39)$$

Observe that for low photon intensity levels ($\gamma_o\Lambda \ll 1$) the upper bound on γ_o is constant as a function of Λ but increases proportionally to the relative spatio-temporal width $(T_p\sigma_p^2)/|I|$ of the SPR pulses $\{p(t-\mathbf{t}_i, \underline{u}-\underline{\mathbf{u}}_i)\}_{i=1}^{\mathbf{n}}$. In this low Λ regime there is little SPR overlap and performance degradation relative to direct detection can be attributed to loss of single photon resolution due to finite SPR pulse width. On the other hand for high photon intensity levels ($\gamma_o\Lambda \gg 1$) the upper bound on γ_o increases linearly in Λ and quadratically in the relative spatio-temporal SPR pulse width. Note also that the bound decreases quadratically in the relative width $\sigma_\lambda^2/|A|$ of the spatial intensity $\lambda(\underline{u}-\underline{\tau})$. The behavior of the γ_o-bound reflects the difficulty of estimating photon arrival coordinates $\{\mathbf{t}_i, \underline{\mathbf{u}}_i\}^{\mathbf{n}}$ in the presence of increasing amounts of spatio-temporal SPR pulse overlap, which can be due to broad SPR responses (large $(T_p\sigma_p^2)/|I|$), high photon count rate (large Λ), or densely packed photon arrivals (small $\sigma_\lambda^2/|A|$). Thus, for large Λ, the relation (39) quantifies the degree to which SPR pulse overlap compromises the attainment of photon-limited estimation performance. Interestingly, γ_o is much more sensitive to SPR width, in which the bound on γ_o increases quadratically, than it is to intensity width, in which the bound decreases only logarithmically.

Under the symmetric Gaussian beam model $B_{cr}^1(\mathbf{M}) = B_{cr}^2(\mathbf{M})$ and the CR bound (8) takes the simple form:

$$B_{cr}(\mathbf{M}) = \sigma_\lambda^2/\Lambda. \quad (40)$$

Assume $\mu_g^2/(\mu_g^2+\sigma_g^2)|\Lambda(\underline{\nu})-\overline{\Lambda}(\underline{\nu})|^2 \ll \Lambda$ and focus on the case $\gamma_o\Lambda \gg 1$ for concreteness. From (35) and (38) we have:

$$B_{rdlb}(\mathbf{X}) = \begin{cases} \frac{|A|}{2\pi e} \cdot \left(\frac{2\pi e\sigma_\lambda^2}{|A|}\right)^\Lambda, & \gamma > \gamma_o \\ \frac{|A|}{2\pi e} \cdot e^{-\sqrt{\gamma\Lambda}\frac{|I|}{4\pi T_p\sigma_p^2}}, & \gamma \le \gamma_o \end{cases}. \quad (41)$$

On the basis of (40) and (41) we conclude:

1. Unlike B_{rdlb} which converges to a value $|A|/(2\pi e)$ close to the prior variance $\sigma_{\tau_1}^2$, B_{cr} converges to the unreasonable limit of ∞ as $\Lambda \to 0$. Indeed, the CR bound is not valid for large estimation errors due to the non-differentiability of the uniform prior $p_{\underline{\tau}}(\underline{\tau})$ at the boundary of the prior region $\mathcal{T} = A$. While a smoother prior would eliminate the problem, this is a significant deficiency in the CR bound.

2. As $\sigma_\lambda^2 \to 0$: $B_{cr} \to 0$ while $B_{rdlb} \to |A|/(2\pi e)\cdot\exp(-\sqrt{\gamma\Lambda}\cdot |I|/[4\pi T_p\sigma_p^2])$ since $\gamma_o \to \infty$. Hence B_{rdlb} is generally a tighter bound for small σ_λ^2.

3. B_{cr} decreases in inverse proportion to Λ while B_{rdlb} decreases exponentially in $\gamma\Lambda$. Hence, B_{cr} can be expected to be a tighter bound for large Λ and large γ.

References

[1] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramer-Rao bounds," *Annals of Statistics*, vol. 15, no. 4, pp. 1421–1438, 1987.

[2] A. M. Bruckstein, T. J. Shan, and T. Kailath, "The resolution of overlapping echos," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol. ASSP-33, pp. 1357–1368, Dec. 1985.

[3] N. H. Clinthorne, W. L. Rogers, L. Shao, and K. F. Koral, "A hybrid maximum likelihood position computer for scintillation cameras," *IEEE Trans. Nuclear Science*, vol. NS-37, no. 2, pp. 658–663, 1990.

[4] M. T. El-Hadidi and B. Hirosaki, "The Bayes optimal receiver for digital fibre optic communications systems," *Optical and Quantum Electronics*, vol. 13, pp. 469–486, 1981.

[5] P. Faure, "Theoretical models of reverberation noise," *J. Acoust. Soc. Am.*, vol. 36, pp. 259–268, 1964.

[6] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.

[7] A. O. Hero, "Timing estimation for a filtered Poisson process in Gaussian noise," *IEEE Trans. on Inform. Theory*, vol. 37, no. 1, pp. 92–106, Jan. 1991.

[8] A. O. Hero, "Lower bounds on estimator performance for energy invariant parameters of multi-dimensional Poisson processes," *IEEE Trans. on Inform. Theory*, vol. 35, pp. 843–858, July 1989.

[9] A. O. Hero, N. Antoniadis, N. H. Clinthorne, W. L. Rogers, and G. D. Hutchins, "Optimal and sub-optimal post-detection timing estimators for PET," *IEEE Trans. Nuclear Science*, vol. NS-37, no. 2, pp. 725–729, 1990.

[10] L. Kazovsky, "Beam position estimation by means of detector arrays," *Opt.·Quantum Electron.*, vol. 13, pp. 201–208, 1981.

[11] J. M. Mendel, "White noise estimators for seismic data processing in oil exploration," *IEEE Trans. Automatic Control*, vol. AC-22, no. 5, pp. 694–706, Oct. 1977.

[12] N. Petrick, N. H. Clinthorne, W. L. Rogers, and A. O. Hero, "First photoelectron timing error evaluation of a new scintillation detector model," *IEEE Trans. Nuclear Science*, vol. 38, no. 2, , to appear April 1991.

[13] D. J. Sakrison and V. R. Algazi, "Comparison of line-by-line and two-dimensional encoding of random images," *IEEE Trans. on Inform. Theory*, vol. IT-17, no. 4, pp. 386–398, July 1971.

[14] P. Salomon and T. Glavich, "Image signal processing in sub-pixel accuracy star trackers," in *Proc. Soc. Phot-Opt. Instrum. Eng.*, pp. 64–74, 1980.