# Time-Delay Estimation for Filtered Poisson Processes Using an EM-Type Algorithm

Nikolaos Antoniadis and Alfred O. Hero, *Member, IEEE*

*Abstract*—**In this paper, we develop a modified EM algorithm to estimate a nonrandom time shift parameter of an intensity associated with an inhomogeneous Poisson process $N_t$, whose points are only partially observed as a noise-contaminated output $X$ of a linear time-invariant filter excited by a train of delta functions—A filtered Poisson process. The exact EM algorithm for computing the maximum likelihood time shift estimate generates a sequence of estimates each of which attempt to maximize a measure of similarity between the assumed shifted intensity and the conditional mean estimate of the Poisson increment $dN_t$. We modify the EM algorithm by using a linear approximation to this conditional mean esitmate. The asymptotic performance of the modified EM algorithm is investigated by an asymptotic estimator consistency analysis. We present simulation results that show that the linearized EM algorithm converges rapidly and achieves an improvement over conventional time-delay estimation methods, such as linear matched filtering and leading edge thresholding. In these simulations our algorithm gives estimates of time delay whose mean square error virtually achieves the CR lower bound for high count rates.**

## I. Introduction

FILTERED point processes are widely used to model a broad range of physical phenomena, such as radioactive emission measurements [1], opto-electronic signals [2], electro-physiological measurements [3], and acoustic reverberation [4]. The filtered point process model assumed here specifies the measured waveform as the superposition of single point responses (SPR's), each excited by a point of an inhomogeneous Poisson point process and each multiplied by a random SPR gain, and additive Gaussian (thermal) noise. The unknown parameters of interest are imbedded in the Poisson intensity function composed of a time-varying signal intensity and a constant dark current intensity. The filtered point process model arises in several important applications: nuclear particle detection systems, such as conventional and time-of-flight positron emission tomography (PET); optical communications receivers; impulsive "shot noise" waveform channels; neuron discharge; and seismic geophysics. In each of the aforementioned applications, detection of the point process of interest, and subsequent estimation of related intensity parameters, are of central importance.

In PET, $\gamma$-ray interactions with a scintillator crystal produce secondary photon emissions detected at the output of a high-gain photo-multiplier tube (PMT). Here the objective is to estimate the $\gamma$-ray interaction time $\tau_0$ that shifts the intensity of the secondary photon emission process. Traditional methods, such as leading-edge and constant fraction timing, are conceptually simple and have been used extensively. However, they fail in several important cases [5]. For example, high dark-current intensity and/or thermal noise may severely bias the leading edge and the constant fraction estimators. More recently, maximum likelihood (ML) methods have gained increasing popularity because of their associated optimality properties [6], [7]. However, analytical expressions for the likelihood function have been derived only for the ideal direct-observation case where the arrival times are observed exactly. Due to finite detector bandwidth and the various "noise" factors (random filter gains, thermal noise, dark current), this ideal model may be unrealistic. Various estimator structures have been proposed to approximate the exact likelihood function in the Poisson limited (low intensity) and Gaussian limited (high intensity) regimes [8]–[10]. However, the function maximization required for the MLE renders these techniques difficult to implement numerically.

In this paper we present a modified expectation-maximization (EM) algorithm that recursively updates the time-delay estimate to iteratively approximate the ML estimate. This algorithm explicitly accounts for the effects of finite bandwidth and noise inherent to the measurements. Like the exact form of the EM ML algorithm [11], our algorithm alternates between estimation of the complete data log-likelihood function and maximization of this estimate over the range of $\tau_0$. The difference between our algorithm and the exact EM algorithm is that we implement a linear approximation to the "$E$-step" of the algorithm. Specifically the algorithm first calculates the optimal linear estimate of the point process as the output of a linear time-varying filter. The output of the filter is then used in place of the ideal direct observation to estimate the time shift of the intensity function via the implementation of the standard MLE. While the exact EM algorithm converges to the MLE and is therefore a consistent estimator, the linearized EM algorithm is not guaranteed to be consistent and may stagnate or diverge. We introduce a general methodology for studying estimator consistency that involves studying convergence of a nonrandom estimator sequence obtained from the mean square

Manuscript received October 21, 1992; revised September 21, 1993. This work was supported by National Cancer Institute, DHHS, under grant CA-46622-01, and by the Al. Onassis Scholarship 1987–88, 1991–92. The associate editor coordinating the review of this paper and approving it for publication was Prof. John Goutsias.

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA.

IEEE Log Number 9401907.

limit of the linearized EM objective function. For the special case of triexponential intensity, monoexponential single point response, i.i.d. Gaussian single electron gain sequence, and negligible thermal noise, we establish that unless the ratio of the squared mean and the variance of the gains is above a certain threshold the linearized EM algorithm stagnates and the resultant time-delay estimate is inconsistent. Finally we present simulations that indicate that significant performance gains relative to previous methods are possible by using our technique. Furthermore, the simulations show that the MSE of the delay estimates virtually achieves the CR lower bound for high count rates.

## II. Statement of the Problem

Let $t_1, t_2, \cdots, t_n$ be the arrival times of events of an inhomogeneous Poisson process $\{dN_t : t \in [0,T]\}$ where $n = \int_0^T dN_t$ is the number of Poisson events or points during the observation interval $[0,T]$. Let the intensity function be given as the superposition

$$\lambda(t) = \lambda_s(t;\tau_0) + \lambda_0, \qquad t \in [0,T]$$

where $\lambda_s(t;\tau_0) = \lambda_s(t-\tau_0)$ is the intensity of an inhomogeneous signal point process, $\tau_0 \in \mathcal{T}$ is an unknown time shift parameter lying in a known interval $\mathcal{T}$, and $\lambda_0$ is a constant intensity governing a homogeneous "dark current" noise process. It is assumed that for all $\tau \in \mathcal{T}$ the mean signal count rate

$$\Lambda_s = \int_0^T \lambda_s(t-\tau)dt$$

is functionally independent of time shift $\tau \in \mathcal{T}$. The above assumption is valid when the $t$-support of $\lambda_s(t-\tau)$ is contained in $[0,T]$ for all $\tau \in \mathcal{T}$, i.e., the time-width $T_\lambda$ of $\lambda_s(t)$ is much shorter than the total observation time $T$.

Available for observation over $[0,T]$ is the continuous waveform $X$

$$X(t) = \sum_{i=1}^{n} a_i p(t-t_i) + w(t) \tag{1}$$

where the $\{a_i\}$ are i.i.d. Gaussian gains associated with each event and $w(t)$ represents the observation noise, which is taken to be white, zero-mean Gaussian with variance $\sigma_w^2$. The impulse response wavelet $p(t)$, normalized to have unit energy, is called the single point response (SPR) and models the inherent finite bandwidth of the measurement system. The first additive term in (1) is known as a *Poisson shot noise process* [2].

We adopt the following notation:

- $\Lambda = \Lambda_s + \Lambda_0$ : mean event count rate over $[0,T]$;
- $\rho = \Lambda_s/\Lambda_0$: signal-to-noise ratio;
- $\mu_a, \sigma_a^2$: mean and variance of the Gaussian gains;
- $\sigma_w^2 = \frac{N_0}{2}$: thermal noise spectral level;
- $\gamma = \mu_a^2/\sigma_w^2$: average SPR pulse-to-noise ratio.

There are two related objectives. First, a detector has to decide whether the "signal" $\lambda_s$ is present or not. If the detector decides that the signal is present, an estimation algorithm should give an estimate $\hat{\tau}$ of the time delay $\tau_0$. In this paper, we focus on post-detection processing of the observations, i.e., time-delay estimation.

## III. Previous Approaches and Approximations

### A. Conventional Timing Methods

In nuclear particle detection, fast and simple timing estimation methods have been used for a long time [1]. The common feature of this group of methods is the detection of the time that a linearly filtered version $X(t)*h(t) = \int_0^T X(u)h(t-u)du$ of the observations first crosses a preset threshold. Among these methods, the leading edge estimator is the easiest and most direct timing method. In this method a percentage of the peak value of the mean of $X(t) * h(t)$ is selected as an amplitude threshold. The selection is based on factors such as signal and noise statistics and practical experience. The estimate of time delay is then chosen as the first threshold crossing by $X(t) * h(t)$ over $[0,T]$. This method is quite effective when the variance of $X(t) * h(t)$ is not large. However, when the variance is large, "amplitude walk" can occur and estimation performance degrades. Dark current also affects the performance, causing early threshold triggering that can introduce large negative bias.

Matched filter-type estimators have also been used, e.g., in optical communications [2]. The optical matched filter (OMF) detects the location over $[0,T]$ of the peak output of a linear filter $h^{\mathrm{OMF}}$ matched to the mean $E[X(t)]$ in the absence of dark current: $h^{\mathrm{OMF}}(-t) = \mu_a\lambda_s(t) * p(t)$. This estimator performs well when the count rate is high, since in this case the standard deviation $\mathrm{var}^{\frac{1}{2}}[X(t)]$ is small relative to the expected response $E[X(t)]$. However, for lower count rates or high dark current levels, occasional outliers may severely bias the estimate [15].

### B. Maximum Likelihood Approach

For given $\tau_0 \in \mathcal{T}$ we denote the probability density function of $X = \{X(t) : t \in [0,T]\}$ by $f(X;\tau_0)$. In the ML approach the log-likelihood function $l(\tau) \stackrel{\mathrm{def}}{=} \ln f(X;\tau)$ is derived and the estimator $\hat{\tau}$ is obtained by maximizing $l(\tau)$ over $\mathcal{T}$. When the observation waveform is specified by (1) the conditional density of the observations $X$ given $n$, the random gains $\{a_i\}_{i=1}^n$, and the arrival times $\{t_i\}_{i=1}^n$, is Gaussian and is specified (up to a constant scale factor) by the Cameron-Martin formula [13]

$$f(X \mid n, \{a_i\}_{i=1}^n, \{t_i\}_{i=1}^n) = \exp\left(\frac{2}{N_0}\sum_{i=1}^{n}\int_0^T X(t)\cdot a_i p(t-t_i)dt - \frac{1}{N_0}\sum_{i,j=1}^{n}\int_0^T a_i p(t-t_i)\cdot a_j p(t-t_j)dt\right). \tag{2}$$

Note that the above conditional density is functionally independent of the time delay $\tau_0$. When the random gains are themselves i.i.d. Gaussian, the unnormalized conditional density given $\{t_i\}_{i=1}^n$ and $n$ can be written as [15]

$$f(X \mid \{t_i\}_{i=1}^n, n) = \exp\left[\frac{\mu_a^2}{2\sigma_a^2}([\alpha \underline{X}_f + \underline{1}]^T [I + R_p\alpha]^{-1} \times [\alpha \underline{X}_f + \underline{1}] - n)\right] \quad (3)$$

where $\alpha \stackrel{\text{def}}{=} \gamma\sigma_a^2/\mu_a^2 = \sigma_a^2/\sigma_w$, $\underline{X}_f = [X_f(t_1), \cdots, X_f(t_n)]^T$ is a vector of $n$ time samples of the normalized and filtered measurements $X$ sampled at the event times $\{t_i\}_{i=1}^n$

$$X_f(t_i) = \frac{1}{\mu_a}\int_0^T X(t)p(t-t_i)dt, \qquad i = 1, \cdots, n$$

and $R_p = [R_p(t_i, t_j)]_{i,j=1,\cdots,n}$ is an $n \times n$ matrix of temporal correlations of $p$

$$R_p(t_i, t_j) = \int_0^T p(t-t_i)p(t-t_j)dt.$$

For given $\tau_0 \in \mathcal{T}$ the joint density of $\{t_i\}_{i=1}^n$ and $n$ is [12]

$$f(\{t_i\}_{i=1}^n, n; \tau_0) = \left(\prod_{i=1}^n \lambda(t_i; \tau_0)\right)\exp(-\Lambda). \quad (4)$$

In (4), $\tau_0$ is included in the arguments of $\lambda$ in order to emphasize the time shift dependence of the intensity function.

To obtain the unconditional density of the observations we must take the expectation of (3) over $\{t_i\}_{i=1}^n$ and $n$

$$f(X; \tau_0) = \sum_{n=0}^{\infty}\int_{t_1,\cdots,t_n} f(X \mid \{t_i\}_{i=1}^n, n) \cdot f(t_1, \cdots, t_n, n; \tau_0)dt_1 \cdots dt_n. \quad (5)$$

The function $f(X;\tau)$ of $\tau$ in (5) has no known analytic form and therefore cannot be implemented as a likelihood function. To deal with this difficulty, approximations to $f(X;\tau)$ have been made for various regimes of interest. It is useful to categorize the approximations into two general classes: 1) small pulse overlap approximations and 2) large pulse overlap approximations.

In the absence of temporal overlap of the pulses $\{p(t-t_i)\}_{i=1}^n$, and for sufficiently low thermal noise level $\alpha \gg 1$, accurate estimation of the event times $\{t_i\}_{i=1}^n$ can be performed, e.g., by using a simple level-crossing detector. If the event time estimates, $\{\hat{t}_i\}_{i=1}^n$, are used to approximate $\{t_i\}_{i=1}^n$ then ideal "direct detection" techniques such as ML [12] or least squares [6] can be used to estimate $\tau_0$ from the $\{\hat{t}_i\}_{i=1}^n$. The obvious shortcoming of this approximation is that it becomes increasingly inaccurate when there is significant pulse overlap since level-crossing thresholds cannot reliably be used to separate the overlapping pulses $\{p(t-t_i)\}_{i=1}^n$.

In [9], an asymptotic approximation to the likelihood function (5) in the small overlap regime was derived. The approximation is increasingly accurate as the correlation matrix $R_p$ in (3) approaches the identity $I$, an asymptotic regime corresponding to small $\Lambda$ and small $T_p/T$. However, several difficulties arise for its implementation. First, the evaluation of the objective function necessitates a double integration over $[0, T]$. Second, the integrand of this objective function involves evaluating an undamped exponential function of $X$ that can cause numerical instability.

When pulse overlap and $\Lambda$ are large the shot noise process (1) approaches a superposition of a large number of i.i.d. components and its distribution is closely approximated by the Gaussian, at least in regions of $[0, T]$ where the signal intensity $\lambda_s(t-\tau_0)$ is large [12]. Using characteristic functions, the first and second order statistics can be readily derived and the log-likelihood function can thus be specified. The Gaussian approximation is the basis for a number of methods for random and nonrandom shift estimation [8], [14].

In [15] the above and several other approximations were compared and simulation results were presented. It was found that the performance of these approximate estimators degraded rapidly, e.g., even for moderate pulse overlap no significant improvement over the simple leading edge timing techniques was achieved.

## IV. Formulation of the EM Algorithm

The EM algorithm is an iterative method to maximize the likelihood function. The method was formally introduced by Dempster *et al.* [11].

Suppose we are interested in estimating a vector of parameters $\underline{\theta}_0 \in \Theta$ based on measurement data $X \in \mathcal{X}$, having probability density $f(X; \underline{\theta}_0), \underline{\theta}_0 \in \Theta$, where $\mathcal{X}$ is a measurement space, called the incomplete data space. The computation of the maximum likelihood estimate (MLE) $\hat{\theta}$ of the parameter vector $\underline{\theta}_0$ may be difficult, e.g., due to an intractable form of the log-likelihood $l(\underline{\theta}) \stackrel{\triangle}{=} \ln f(X; \underline{\theta})$ as a function of $\underline{\theta}$. To overcome this difficulty, we postulate a different data space $\mathcal{Y}$, called the complete data space, for the measurements such that the log-likelihood $\ln f(Y; \underline{\theta})$ is better behaved as a function of $\underline{\theta}$. Once $Y$ and $f(Y; \underline{\theta})$ are specified, we implement a recursive algorithm that alternates between *estimating* the log-likelihood function $\ln f(Y; \underline{\theta})$ given $(X, \underline{\theta}^{(k)})$ and *maximizing* the estimate of $\ln f(Y; \underline{\theta})$ over $\underline{\theta} \in \Theta$ to obtain $\underline{\theta}^{(k+1)}$.

More specifically, if the complete data $Y$ were available, we would maximize the complete data log-likelihood function: $\hat{\underline{\theta}} = \arg\max_{\underline{\theta}\in\Theta} \ln f(Y; \underline{\theta})$. However, since $Y$ is not available we maximize the "best guess"of $\ln f(Y; \underline{\theta})$ given the actual observations $X$. If an estimate $\underline{\theta}^{(k)}$ is available, a good guess for $\ln f(Y; \underline{\theta})$ is one that minimizes the mean-square error given $X$ and $\underline{\theta}^{(k)}$. The minimum mean-square error (MMSE) estimator of $\ln f(Y; \underline{\theta})$ is the conditional mean

$$(E) \quad Q\left(\underline{\theta} \mid \underline{\theta}^{(k)}\right) = E\left[\ln f(Y; \underline{\theta}) \mid X; \underline{\theta}^{(k)}\right] \quad (6)$$

where $Q\left(\underline{\theta} \mid \underline{\theta}^{(k)}\right)$ is called the *objective function* of the EM algorithm. Next we maximize the objective function over $\underline{\theta}$ to obtain an updated estimate $\underline{\theta}^{(k+1)}$

$$(M) \quad \underline{\theta}^{(k+1)} = \arg\max_{\underline{\theta}\in\Theta} Q\left(\underline{\theta} \mid \underline{\theta}^{(k)}\right). \quad (7)$$

Alternation between $(E)$ and $(M)$ yields a sequence of values $\underline{\theta}^{(0)}, \underline{\theta}^{(1)}, \cdots$ and is called the EM algorithm. This algorithm has the monotonic property that the sequence $l(\underline{\theta}^{(k)}) = \ln f(X; \underline{\theta}^{(k)})$, $k = 1, 2, \cdots$, is nondecreasing at each stage. As a consequence the algorithm converges if the likelihood function is bounded. However, if there are multiple local maxima, it may not converge to the global maximum. The standard requirements for the EM algorithm to converge are: 1) The p.d.f.'s on both spaces $\mathcal{X}$ and $\mathcal{Y}$ are governed by the same parameters $\underline{\theta}$ and 2) $\mathcal{Y}$ must be related to $\mathcal{X}$ through a relation $g(Y) = X$, where $g$ is a deterministic function [11].

In our problem $\underline{\theta}$ is a scalar time shift parameter $\tau$, and the incomplete data are identified as the observation waveform $X = \{X(t) : t \in [0, T]\}$ . We define our complete-data $Y$ as the incomplete data supplemented by the set of arrival times: $[t_1, t_2, \cdots, t_n]^T, n$.

The complete-data likelihood has the simple form

$$f(Y; \tau) = f(X \mid \{t_i\}_{i=1}^n, n) \cdot f(\{t_i\}_{i=1}^n, n; \tau). \tag{8}$$

Using (3) and (4) in (8), we have

$$\begin{aligned} \ln f(Y \mid \tau) = \ln f(X \mid \{t_i\}_{n=1}^n, n) + \ln f(\{t_i\}_{i=1}^n, n; \tau) \\ \frac{\mu_a^2}{2\sigma_a^2}\left([\alpha \underline{X}_f + \underline{1}]^T [I + R_p \alpha]^{-1} [\alpha \underline{X}_f + \underline{1}] - n\right) \\ - \Lambda + \sum_{i=1}^n \ln \lambda(t_i; \tau) \end{aligned} \tag{9}$$

Taking the conditional expectation of (9) we obtain

$$\begin{aligned} E\left[\ln f(Y; \tau) \middle| X; \tau^{(k)}\right] \\ = E\left[\frac{\mu_a^2}{2\sigma_a^2}\left([\alpha \underline{X}_f + \underline{1}]^T [I + R_p \alpha]^{-1}\right.\right. \\ \left.\left.[\alpha \underline{X}_f + \underline{1}] - n\right) \middle| X; \tau^{(k)}\right] \\ - \Lambda + E\left[\sum_{i=1}^n \ln \lambda(t_i; \tau) \middle| X; \tau^{(k)}\right]. \end{aligned} \tag{10}$$

In (10) there is a sum of three terms: the first is a function of the observations and the event arrival times, but not a function of $\tau$. Hence, the conditional mean of this first term is a function of $\tau^{(k)}$ *only*. Consequently, this term does not affect the maximization with respect to $\tau$, performed in the $(M)$ step given by (7). The second term $\Lambda$ is also independent of $\tau$ due to the assumption that the time delay does not introduce truncation of the signal intensity in $[0, T]$. Therefore, we may ignore these two terms. For the third term, using the stochastic integral representation of the Poisson summation (see Appendix B for a brief overview), we obtain an equivalent $Q$ function for the EM algorithm

$$\begin{aligned} Q(\tau \mid \tau^{(k)}) \stackrel{\text{def}}{=} E\left[\sum_{i=1}^n \ln \lambda(t_i; \tau) \mid X; \tau^{(k)}\right] \\ = E\left[\int_0^T dN_t \ln \lambda(t; \tau) \mid X; \tau^{(k)}\right]. \end{aligned} \tag{11}$$

Assuming the exchange of expectation and integration is valid this is equivalent to

$$\begin{aligned} Q(\tau \mid \tau^{(k)}) &= \int_0^T E\left[dN_t \mid X, \tau^{(k)}\right] \ln \lambda(t; \tau) \\ &= \int_0^T \widehat{dN}_t\left(X, \tau^{(k)}\right) \ln \lambda(t; \tau) \end{aligned} \tag{12}$$

where $\widehat{dN}_t\left(X, \tau^{(k)}\right)$ denotes the minimum mean-square-error estimator (MMSEE) of the point process $dN_t$, given the observations $X$ and the current estimate $\tau^{(k)}$ of the time delay.

Using (12) in (7), the EM algorithm takes the following simple recursive form

$$\begin{aligned} \tau^{(k+1)} &= \arg\max_{\tau \in \mathcal{T}}\left[Q(\tau \mid \tau^{(k)})\right] \\ &= \arg\max_{\tau \in \mathcal{T}}\left[\int \widehat{dN}_t\left(X, \tau^{(k)}\right) \ln \lambda(t_i; \tau)\right]. \end{aligned} \tag{13}$$

In view of (13), we can make useful interpretations that elucidate the asymptotic behavior of the EM estimation algorithm. When there is no observation noise and $p(t)$ approaches a Dirac delta function, corresponding to an infinite bandwidth measurement system, then $dN_t$ is perfectly observable and $\widehat{dN}_t\left(X, \tau^{(k)}\right) = dN_t$. In this case

$$\begin{aligned} \int \ln \lambda(t; \tau) \widehat{dN}_t\left(X, \tau^{(k)}\right) &= \int \ln \lambda(t; \tau) dN_t \\ &= \sum_{i=1}^n \ln \lambda(t_i; \tau) \end{aligned}$$

which, when substituted back into (13), yields the well-known form of the maximum-likelihood estimator of $\tau_0$ when the point process is perfectly observable [6]. For this case the EM algorithm converges in a single step to the MLE.

On the other hand, when the noise level is high, it can be expected that the conditional mean is approximately equal to the unconditional mean: $\widehat{dN}_t\left(X, \tau^{(k)}\right) \approx E[dN_t \mid \tau^{(k)}] = \lambda\left(t; \tau^{(k)}\right) dt$. Therefore, from (12)

$$Q\left(\tau \mid \tau^{(k)}\right) \approx \int_0^T dt \lambda\left(t; \tau^{(k)}\right) \ln \lambda(t; \tau) \tag{14}$$

which has a maximum at $\tau = \tau^{(k)}$. To see this, we make use of the inequality $\ln z \leq z - 1$ as follows

$$\begin{aligned} \int_0^T dt \lambda(t; \tau^{(k)}) \ln \lambda(t; \tau) - \int_0^T dt \lambda(t; \tau^{(k)}) \ln \lambda(t; \tau^{(k)}) \\ = \int_0^T dt \lambda(t; \tau^{(k)}) \ln \frac{\lambda(t; \tau)}{\lambda(t; \tau^{(k)})} \\ \leq \int_0^T dt \lambda(t; \tau) - \int_0^T dt \lambda(t; \tau^{(k)}) \\ = \Lambda - \Lambda = 0. \end{aligned} \tag{15}$$

Therefore, we conclude that when the thermal noise level is high, the maximization operation is (13) is degenerate and the algorithm will stagnate on the initial estimate: $\tau^{(k)} = \tau^{(0)}$, $k = 1, 2, \cdots$.

## V. Linearization of Conditional Mean

The EM algorithm presented in the last section has a deceptively simple looking iterative form (13). The main difficulty for implementation is the calculation of the conditional mean estimate (11) of $S(\tau) \stackrel{\text{def}}{=} \int_0^T dN_t \ln \lambda(t;\tau)$, or, equivalently, the calculation of the conditional mean estimate (12) of $dN_t$. In general, the estimate of $S$ will be the output of an infinite-order nonlinear filter operating on the observations $X$. An important exception is the case of jointly Gaussian $X$ and $S$, where the conditional mean estimate is the *linear* minimum mean square error estimate (LMMSE). The jointly Gaussian approximation is reasonable for $X$ and $S$ when $\lambda(t;\tau)$ is uniformly large over $t \in [0,T]$ for all $\tau \in \mathcal{T}$. However, in general the conditional mean estimate of $S$ given $X$ is not linear. In this section we derive a linearized EM algorithm by finding the best linear approximation to the objective function $Q(\tau \mid \tau^{(k)})$ in (13). In the next section we investigate some of the limitations of this linearized algorithm.

Specifically, we have the approximate EM algorithm

$$\tau^{(k+1)} = \arg\max_{\tau \in \mathcal{T}} \tilde{Q}\left(\tau \mid \tau^{(k)}\right) \tag{16}$$

where

$$\tilde{Q}\left(\tau \mid \tau^{(k)}\right) = \int_0^T \widetilde{dN}_t\left(X, \tau^{(k)}\right) \ln \lambda(t;\tau) \tag{17}$$

and the optimal linear estimate $\widetilde{dN}_t(X, \tau^{(k)})$ of the event sequence is given by [16]

$$\begin{aligned}\widetilde{dN}_t(X,\tau^{(k)}) = &\int_0^T du\, R_{dN_t,X}\left(t,u;\tau^{(k)}\right) \\ &\times \int_0^T dv R_X^{-1}\left(u,v;\tau^{(k)}\right) \\ &\times \left[X(v) - \mu_X\left(v;\tau^{(k)}\right)\right] + E\left[dN_t \middle| \tau^{(k)}\right]\end{aligned} \tag{18}$$

where:

- $R_X(u,v;\tau)$ is the autocovariance function of the observations $X$ sampled at times $u$ and $v$ when the true time shift is $\tau$;
- $R_{dNt,X}(t,u;\tau)$ is the cross-covariance function between $dN_t$ and $X(u)$ when the true time shift is $\tau$;
- $\mu_X(u;\tau)$ is the mean of $X(u)$ when the true shift is $\tau$;
- $E[dN_t \mid \tau] = \lambda(t;\tau)dt$ is the mean of the point process when the true time shift is $\tau$;
- $R_X^{-1}$ denotes the inverse kernel function associated with $R_X$.

Using the independence between the thermal noise $w(t)$ and the point process $dN_t$, we can calculate the quantities needed in (18). The mean $\mu_X(t;\tau^{(k)})$ is calculated as follows

$$\begin{aligned}\mu_X\left(t;\tau^{(k)}\right) &= E\left[\sum_{i=1}^n a_i p(t-t_i) \middle| \tau^{(k)}\right] + E\left[w(t) \middle| \tau^{(k)}\right] \\ &= \mu_a \cdot \int_0^T \lambda\left(\xi;\tau^{(k)}\right) p(t-\xi) d\xi.\end{aligned} \tag{19}$$

The autocovariance function of the observations can be expressed as [12]

$$\begin{aligned}R_X\left(u,v;\tau^{(k)}\right) &= \sigma_w^2 \delta(u-v) + (\sigma_a^2 + \mu_a^2) \\ &\quad \times \int p(u-\zeta)\lambda\left(\zeta;\tau^{(k)}\right) p(v-\zeta) d\zeta.\end{aligned} \tag{20}$$

On the other hand, we have by definition of the cross-covariance function

$$\begin{aligned}&R_{dN_t,X}\left(t,u;\tau^{(k)}\right) \\ &\quad = E\left[dN_t \sum_{i=1}^n a_i p(u-t_i) \middle| \tau^{(k)}\right] - E\left[dN_t \middle| \tau^{(k)}\right] \\ &\qquad \times E\left[\sum_{i=1}^n a_i p(u-t_i) \middle| \tau^{(k)}\right].\end{aligned} \tag{21}$$

Using the stochastic integral representation, (44) in Appendix B, and making use of the independence between the point process and the random gains, we can write the two terms on the right-hand side of (21) as follows

$$\begin{aligned}&E\left[dN_t \cdot \sum_{i=1}^n a_i p(u-t_i) \middle| \tau^{(k)}\right] \\ &\quad = E\left[a_i \mid \tau^{(k)}\right] \cdot E\left[dN_t \cdot \sum_{t=1}^n p(u-t_i) \middle| \tau^{(k)}\right] \\ &\quad = \mu_a \cdot E\left[dN_t \cdot \int_{\{\zeta \in [0,T]\}} p(u-\zeta) dN_\zeta \middle| \tau^{(k)}\right] \\ &\quad = \mu_a \cdot \int_{\{\zeta \in [0,T]\}} p(u-\zeta) E\left[dN_t dN_\zeta \middle| \tau^{(k)}\right].\end{aligned} \tag{22}$$

$$\begin{aligned}&E\left[dN_t \mid \tau^{(k)}\right] E\left[\sum_{i=1}^n a_i p(u-t_i) \middle| \tau^{(k)}\right] \\ &\quad = \mu_a E\left[dN_t \mid \tau^{(k)}\right] E\left[\int_{\{\zeta \in [0,T]\}} p(u-\zeta) dN_\zeta \middle| \tau^{(k)}\right] \\ &\quad = \mu_a E\left[dN_t \mid \tau^{(k)}\right] \int_{\{\zeta \in [0,T]\}} p(u-\zeta) E\left[dN_\zeta \middle| \tau^{(k)}\right] \\ &\quad = \mu_a \lambda(t;\tau^{(k)}) dt \int_{\{\zeta \in [0,T]\}} p(u-\zeta)\lambda(\zeta;\tau^{(k)}) d\zeta.\end{aligned} \tag{23}$$

When $t \neq \zeta$, the two differentials $dN_t$ and $dN_\zeta$ in (22) are statistically independent, while when $t = \zeta$ we have $dN_t\, dN_\zeta = dN_t$, since $dN_t \in \{0,1\}$. Therefore

$$\begin{aligned}&E\left[dN_t \cdot \sum_{i=1}^n a_i p(u-t_i) \middle| \tau^{(k)}\right] \\ &\quad = \mu_a \int_{\{\zeta \in [0.T]\} - \{t\}} p(u-\zeta) E\left[dN_t \middle| \tau^{(k)}\right] \\ &\qquad \times E\left[dN_\zeta \mid \tau^{(k)}\right] + \mu_a p(u-t) E\left[dN_t \middle| \tau^{(k)}\right] \\ &\quad = \lambda(t;\tau^{(k)}) dt \mu_a \int_{\zeta \neq t} p(u-\zeta)\lambda(\zeta;\tau^{(k)}) d\zeta \\ &\qquad + \mu_a p(u-t)\lambda\left(t;\tau^{(k)}\right) dt\end{aligned} \tag{24}$$

where the integral on the RHS of the last equalities in (23) and (24) is a standard (Riemann) integral. Now, since the integrand $p(u-\zeta)\lambda(\zeta;\tau^{(k)})$ of this integral contains no delta functions, we have

$$\int_{\{\zeta\in[0,T]\}-\{t\}} p(u-\zeta)\lambda(\zeta;\tau^{(k)})\zeta = \int_{\{\zeta\in[0,T]\}} p(u-\zeta)\lambda(\zeta;\tau^{(k)})d\zeta.$$

Plugging (23) and (24) into (21), we obtain

$$R_{dN_t,X}\left(t,u;\tau^{(k)}\right) = \mu_a p(u-t)\lambda\left(t;\tau^{(k)}\right)dt. \tag{25}$$

Using (25) and (19) in (18) we obtain the final linear filtering formula

$$\widetilde{dN}_t(X,\tau^{(k)}) = \lambda(t;\tau^{(k)})dt\cdot\left\{1+\mu_a\int_0^T dup(u-t)\cdot\int_0^T dvR_X^{-1}\left(u,v;\tau^{(k)}\right)\cdot\left[X(v)-\mu_a\int_0^T\lambda\left(\xi;\tau^{(k)}\right)p(v-\xi)d\xi\right]\right\}. \tag{26}$$

Finally plugging (26) into (17) we obtain

$$\tilde{Q}\left(\tau\mid\tau^{(k)}\right) = \int_0^T dt\lambda(t;\tau^{(k)})\ln\lambda(t;\tau)\cdot\left\{1+\mu_a\int_0^T du\,p(u-t)\cdot\int_0^T dvR_X^{-1}\left(u,v;\tau^{(k)}\right)\times\left[X(v)-\mu_a\int\lambda\left(\xi;\tau^{(k)}\right)p(v-\xi)d\xi\right]\right\}. \tag{27}$$

*Discussion*

By examining the form of the optimal linear approximation (27) to the E step of the EM algorithm, we can make some useful interpretations. First, observe that when the observation waveform $X$ closely resembles its expected value under the current estimate $\tau^{(k)}$, which may occur if $\tau^{(k)}$ is close to the true time shift $\tau_0$ and the variance of $X$ is low, the argument of the quantity (27) under the double integral will have a value close to zero. Then $\widetilde{dN}_t \approx \lambda\left(t;\tau^{(k)}\right)dt$ and (27) becomes: $\tilde{Q}\left(\tau\mid\tau^{(k)}\right)\approx\int_0^T dt\,\lambda(t;\tau^{(k)})\ln\lambda(t;\tau)$, which, as was seen in (15), has a maximum at $\tau=\tau^{(k)}$. This suggests that, if the variance of $X$ is low, once the algorithm gives an estimate close to the true value of the parameter, successive iterates $\tau^{(k+1)},\tau^{(k+2)},\cdots$ will remain in the neighborhood of the parameter.

Second, observe that the difference $X(t)-\mu_a\int\lambda\left(\xi;\tau^{(k)}\right)p(t-\xi)d\xi$ should take on positive values in those time intervals where the value of $\lambda(t;\tau^{(k)})$ is small and the value of $\lambda(t;\tau_0)$ is large, whereas it should take on negative values where the opposite is true. Since this difference is passed through a filter involving the inverse of the assumed intensity function, the positive values are amplified, while the negative values are suppressed. As a result, the matching process will tend to shift the current estimate towards areas where $\lambda(t;\tau_0)$ is high, as expected.

However, the linearized EM algorithm can be seen to degenerate as the mean $\mu_a$ of the gain sequence goes to zero. Observe from (27) that the output of the linear filter, which is supposed to act as a correction factor for the estimator update from $\tau^{(k)}$ to $\tau^{(k+1)}$, has a multiplicative factor $\mu_a$. When $\mu_a$ is close to zero the observations $X(t)$ lose their influence on the update, $\tilde{Q}\left(\tau\mid\tau^{(k)}\right)\approx\int_0^T dt\lambda(t;\tau^{(k)})\ln\lambda(t;\tau)$ and, as explained above, the linearized EM algorithm will therefore stagnate.

## VI. Consistency of Linearized EM Algorithm

A natural way of trying to better understand the attributes of the linearized algorithm is to look at the iterative algorithm behavior under "best case" noiseless conditions. In particular, we will investigate the EM sequence $\tau^{(1)},\tau^{(2)},\cdots$ as $\Lambda$, the total average number of Poisson events, goes to infinity and under the simplifying assumption that $\sigma_w^2=0$. If the sequence converges to the true delay $\tau_0$ in the asymptotic limit of large $\Lambda$, then the limit point of the EM sequence is said to be a consistent estimator and otherwise it is inconsistent. Note that if the estimate is established to be inconsistent under the assumption $\sigma_w^2=0$, it will certainly be inconsistent for the more difficult case $\sigma_w^2>0$, although the converse may not hold.

Consider the equivalent EM algorithm

$$\tau^{(k+1)} = \arg\max_{\tau\in\mathcal{T}}\frac{1}{\Lambda}\tilde{Q}\left(\tau\mid\tau^{(k)}\right). \tag{28}$$

To investigate this algorithm as $\Lambda\to\infty$ we normalize both sides of (17) by the total count $\Lambda$

$$\frac{1}{\Lambda}\tilde{Q}\left(\tau\mid\tau^{(k)}\right) = \int_0^T\frac{1}{\Lambda}\widetilde{dN}_t(X,\tau^{(k)})\ln\lambda(t;\tau) = \int_0^T\frac{1}{\Lambda}\widetilde{dN}_t(X,\tau^{(k)})\ln\left(\frac{1}{\Lambda}\lambda(t;\tau)\right) + \ln\Lambda\int_0^T\frac{1}{\Lambda}\widetilde{dN}_t(X,\tau^{(k)}). \tag{29}$$

Since the second term on the right-hand side of (29) does not depend on the variable $\tau$, the linearized EM algorithm (16) is equivalent to

$$\tau^{(k+1)} = \arg\max_{\tau\in\mathcal{T}}\tilde{\tilde{Q}}\left(\tau\mid\tau^{(k)}\right) \tag{30}$$

where

$$\tilde{\tilde{Q}}\left(\tau\mid\tau^{(k)}\right) = \int_0^T\frac{1}{\Lambda}\widetilde{dN}_t\left(X,\tau^{(k)}\right)\ln\frac{1}{\Lambda}\lambda(t;\tau). \tag{31}$$

Furthermore

$$\frac{1}{\Lambda}d\widetilde{N}_t(X,\tau^{(k)}) = \frac{1}{\Lambda}\lambda(t;\tau^{(k)})dt \cdot \left\{1+\mu_a\int_0^T du\, p(u-t)\int_0^T dv\left[\frac{1}{\Lambda}R_X\left(u,v;\tau^{(k)}\right)\right]^{-1} \times\left[\frac{1}{\Lambda}X(v)-\mu_a\int_0^T\frac{1}{\Lambda}\lambda\left(\xi;\tau^{(k)}\right)p(v-\xi)d\xi\right]\right\}. \quad (32)$$

Since $\Lambda = \int_0^T \lambda(t;\tau^{(k)})dt$, the quantity

$$\frac{1}{\Lambda}\lambda\left(t;\tau^{(k)}\right) = \lambda^*\left(t;\tau^{(k)}\right) \quad (33)$$

in (30) and (32) is merely a normalized version of the point process intensity, which integrates to 1 over $t \in [0,T]$ and that is functionally independent of $\Lambda$. On the other hand, for $\sigma_w^2 = 0$ from (20) we have

$$\frac{1}{\Lambda}R_X\left(u,v;\tau^{(k)}\right) = (\sigma_a^2+\mu_a^2)R_X^*(u,v;\tau^{(k)}) \quad (34)$$

where

$$R_X^*\left(u,v;\tau^{(k)}\right)\stackrel{\text{def}}{=}\int p(u-\zeta)\frac{1}{\Lambda}\lambda\left(\zeta;\tau^{(k)}\right)p(v-\zeta)d\zeta.$$

Next we show that, for true delay $\tau_0$, the normalized observation $\frac{1}{\Lambda}X(t)$ converges in the mean-square sense to its mean value $\mu_a\cdot\mu_X^*(t;\tau_0)$, where

$$\mu_X^*(t;\tau)\stackrel{\text{def}}{=}\int_0^T\lambda^*(\xi;\tau)p(t-\xi)d\xi. \quad (35)$$

It will suffice to show that the variance of $\frac{1}{\Lambda}X(t)$ goes to zero. Using (20) with $u=v=t$, $\tau^{(k)}=\tau_0$, and $\sigma_w^2=0$ we obtain

$$\mathrm{Var}\left(\frac{1}{\Lambda}X(t)\right) = \frac{1}{\Lambda^2}R_X(t,t;\tau_0) \quad (36)$$

$$= \frac{\mu_a^2+\sigma_a^2}{\Lambda}\int_0^T p^2(t-\xi)\lambda^*(\xi;\tau_0)d\xi. \quad (37)$$

Since the integral is finite, we obtain $\mathrm{Var}\left(\frac{1}{\Lambda}X(t)\right)\to 0$ and mean-square convergence is proved.

Using (32), (33), and (35) in (31) we obtain as $\Lambda\to\infty$

$$\tilde{\tilde{Q}}\left(\tau\mid\tau^{(k)}\right)\to\overline{Q}\left(\tau\mid\tau^{(k)}\right)$$

where

$$\overline{Q}\left(\tau\mid\tau^{(k)}\right)\stackrel{\text{def}}{=}\int_0^T dt\lambda^*\left(t;\tau^{(k)}\right)\ln\lambda^*(t;\tau)\cdot \times\left[1+\frac{\mu_a^2}{\mu_a^2+\sigma_a^2}\int du\, p(u-t) \times\int dv R_X^{*-1}\left(u,v,\tau^{(k)}\right) \times\left[\mu_X^*(v;\tau_0)-\mu_X^*\left(v;\tau^{(k)}\right)\right]\right]$$

and convergence is in the mean square (m.s.) sense for fixed true time-delay parameter $\tau_0$.

Define the sequence $\{\overline{\tau}^{(k)}\}$ by the following recursion

$$\overline{\tau}^{(k+1)} = \arg\max_{u\in\mathcal{T}}\overline{Q}(u\mid\overline{\tau}^{(k)}) \quad (38)$$

where $u$ is the search variable. The asymptotic consistency of the linearized EM sequence $\tau^{(1)},\tau^{(2)},\cdots$ can be studied by investigating the limit points of the sequence, $\overline{\tau}^{(1)},\overline{\tau}^{(2)},\cdots$

*Consistency of* $\{\tau^{(k)}\}$: Using Proposition 1 of Appendix C, if $\overline{Q}(u\mid\tau^{(k)})$ is differentiable and unimodal then the mean square convergence $\tilde{\tilde{Q}}(u\mid\tau^{(k)})\to\overline{Q}(u\mid\tau^{(k)})$ implies convergence in probability (i.p.) of the sequence of estimates $\tau^{(1)},\tau^{(2)},\cdots$, obtained from successive $u$-maximization of $\tilde{\tilde{Q}}(u\mid\tau^{(k)})$, to the sequence of estimates $\overline{\tau}^{(1)},\overline{\tau}^{(2)},\cdots$, obtained from successive maximization of $\overline{Q}(u|\tau^{(k)})$. Recall that for an ergodic stationary process a parameter estimator is said to be consistent if it converges i.p. to the true parameter as the total observation time goes to infinity. For the nonstationary Poisson process considered here, a natural extension of the definition of consistency for the linearized EM estimator is: The estimator is said to be consistent if $\lim_k\tau^{(k)}$ converges i.p. to the true parameter $\tau_0$; as $\Lambda\to\infty$. Under this definition, since $\lim_k\tau^{(k)}$ converges i.p. to $\lim_k\overline{\tau}^{(k)}$ defined above, the sequence $\tau^{(1)},\tau^{(2)},\cdots$, of linearized EM estimators is consistent if and only if the limit $\lim_k\overline{\tau}^{(k)}$ of the deterministic sequence $\overline{\tau}^{(1)},\overline{\tau}^{(2)},\cdots$, is equal to $\tau_0$.

First we note the following general consistency property. From the right-hand side of (38) stagnation occurs in the sequence $\{\overline{\tau}^{(k)}\}$ when the ratio $\mu_a^2/\sigma_a^2$ approaches zero since in this case the objective function $\overline{Q}(u\mid\tau^{(k)})$ reduces to the form (15). Therefore, unless the ratio $\mu_a/\sigma_a$, is sufficiently large to avoid stagnation of the sequence $\{\overline{\tau}^{(k)}\}$, the limit $\lim_{k\to\infty}\tau^{(k)}$ will be an inconsistent estimator.

By inspecting the surface $\{\overline{Q}(u\mid v): u,v\in\mathcal{T}\}$ we can establish consistency or inconsistency of the linearized EM algorithm for a specific intensity $\lambda$, pulse shape $p(t)$, and operating conditions, $\mu_a^2/\sigma_a^2,\rho$. Given a previous iterate $\overline{\tau}^{(k)}$, the next iterate $\overline{\tau}^{(k+1)}$ is given by the location along the line $\{(u,v): v=\overline{\tau}^{(k)}\}$ of the global maximum of $\overline{Q}(u\mid v)$. The surface $\{\overline{Q}(u\mid v): u,v\in\mathcal{T}\}$ for $\mathcal{T}=[2250\text{ ns}, 2750\text{ ns}]\times[2250\text{ ns}, 2750\text{ ns}]$ and for the true time delay $\tau_0=2500$ ns was generated for the case of triexponential signal intensity, monoexponential SPR, and the same parameters used for the simulations presented in the next section. Note that, since $\overline{Q}(u\mid v)$ is simply shifted by a change in $\tau_0$, the choice of $\tau_0$ within $\mathcal{T}$ is immaterial as long as the support of $\lambda_s(t-\tau_0)$ is contained in $[0,T]$. The signal intensity function is the $t$-function $\lambda_s(t-\tau)$ where

$$\lambda_s(t) = e^{-t/\kappa_1}\left[e^{-t/\kappa_2}-e^{-t/\kappa_3}\right] \quad (39)$$

and $0<\kappa_3<\kappa_2<\kappa_1$. By choosing $\kappa_1,\kappa_1,\kappa_3$ the slow component ($k_1$) and the rise time (approximately $\kappa_2-\kappa_3$) can be fitted to a large number of smooth unimodal intensity functions. To facilitate graphical interpretation, the objective function $\overline{Q}(u\mid v)$ displayed in Figs. 1 and 2 are normalized with the maximum value of $\overline{Q}(u\mid v)$ over $u$ for each $v$. For $\mu_a/\sigma_a=5$, we see in Fig. 1 that, no matter where $\overline{\tau}^{(k)}$ is located on the $v$ axis, the maximum along the line $\{(u,v): v=\overline{\tau}_{(k)}\}$ occurs at the true time delay $\tau_0$, since the

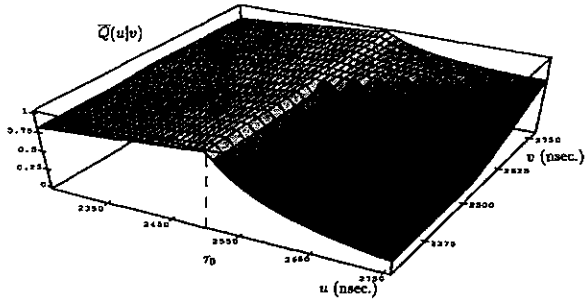



Fig. 1. Objective function for the triexponential intensity, $\mu_a/\sigma_a = 5$.

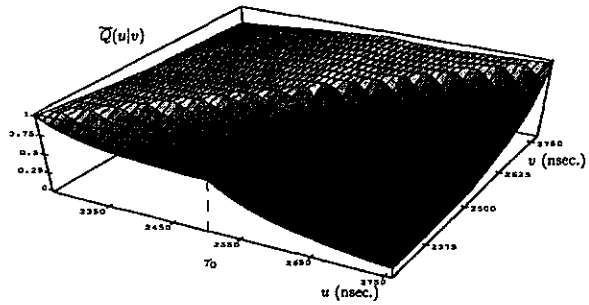



Fig. 2. Objective function for the triexponential intensity $\mu_a/\sigma_a = 1$.

maximum of $\overline{Q}(u \mid v)$ occurs along the horizontal ridge $u = \tau_0$ for all $v$. Thus for this high value of $\mu_a^2/\sigma_a^2$ the EM algorithm converges in a single step to $\tau = \tau_0$, no matter what the value of the initial point $\overline{\tau}^{(0)}$ and for this case the linearized EM algorithm is consistent. On the other hand, for $\mu_a/\sigma_a = 1$, we see in Fig. 2 that the maximum over $u$ of $\overline{Q}(u \mid v)$ occurs at $u = \overline{\tau}^{(k)}$, since $\overline{Q}(u \mid v)$ takes its maximum along the diagonal ridge $u = v$ for all $v$. Thus for this low value of $\mu_a^2/\sigma_a^2$ the EM algorithm stagnates at the initial estimate $\overline{\tau}^{(0)}$ and for this case the linearized EM algorithm is inconsistent.

Now, to further illustrate the asymptotic performance of the linearized EM algorithm we will consider the limiting case where $p(t)$ approaches a delta function. Then it can easily be shown that the optimal linear estimate (27) of the Poisson point process takes the asymptotic form

$$\begin{aligned}\lim_{\Lambda\to\infty} \frac{1}{\Lambda}\widetilde{dN_t} &= \lambda^*\left(t;\tau^{(k)}\right)dt\left[1 + \frac{\mu_a^2}{\mu_a^2+\sigma_a^2}\right. \\ &\quad \times \int du\, \delta(t-u) \int dv\, \delta(u-v). \\ &\quad \left. \times \frac{1}{\lambda^*(u;\tau^{(k)})}\left[\lambda^*(v;\tau_0) - \lambda^*(v;\tau^{(k)})\right]\right] \\ &= \frac{\sigma_a^2}{\mu_a^2+\sigma_a^2}\lambda^*(t;\tau^{(k)})dt + \frac{\mu_a^2}{\mu_a^2+\sigma_a^2}\lambda^*(t;\tau_0)dt. \end{aligned} \tag{40}$$

In the limit, the estimate of the point process reduces to a weighted average of the *true* intensity $\lambda^*(t;\tau_0)$ and the *assumed* intensity $\lambda^*(t;\tau^{(k)})$. When $\sigma_a^2 \gg \mu_a^2$ we see that the observations have less influence than the previous estimate on the estimation process. This is readily seen for the case $\mu_a = 0$, when the algorithm stagnates on the initial estimate. On the other hand, when $\mu_a^2 \gg \sigma_a^2$, the observations dominate and the estimation process will drive the estimate towards the true parameter value $\tau_0$. This fact suggests that the linear approximation is insufficient for regimes where $\sigma_a$ dominates $\mu_a$. In this case, a quadratic approximation [17] will perform better than the linear approximation introduced here.

## VII. Simulations

We applied our linearized EM algorithm to simulated data and compared its performance with that of two leading edge estimators and the linear OMF estimator of the time delay $\tau_0$.

In these simulations, the sampling interval was set at $T_s = 0.5$ ns. We used the triexponential intensity (39) with a rise time of approximately 2.5 ns and a quick decay of 60 ns along with a longer 300 ns decay component. More than 90% of the events are generated by this longer decay component. The overall decay component is approximately 260 ns. This intensity function is typical of a BGO scintillator used in PET imaging [18]. Our SPR $p(t)$ is an energy-normalized one-sided exponential pulse, $p(t) = \sqrt{\frac{1}{2T_p}}\exp(-t/T_p)$, with $T_p = 5$ ns. The random gains were generated as positive i.i.d. random variables with a truncated Gaussian distribution and $\mu_a/\sigma_a = 4.5$. It was verified that for this value the linearized EM algorithm gives a consistent estimator. The gain distribution and the pulse shape are analytic approximations to real measurement data for a Burle/RCA 8850 Photomultiplier Tube used in PET imaging [18]. The estimator's prior search region was centered at 2500 ns and was 500 ns wide. Each of the two leading edge estimators was implemented to trigger on the first crossing of thresholds $\beta_1$ and $\beta_2$, respectively. The thresholds were set at the 20% and 80% points between the peak of the mean output waveform $\mu_a\lambda(t) * p(t)$ and the mean background noise level $\mu_a\lambda_0 * p(t)$. The simulations were performed for $\gamma = 20$ db, $\rho = 20$ db and $\Lambda$ varying from 15 to 25 db (approximately 30 to 300 total event count).

The Poisson events were generated by a simple log transformation of pseudo random uniform variates to produce a Poisson process. The integrals in (13) and (18) were discretized and a high SNR approximation to the inverse kernel $R_X^{-1}$ was used (see Appendix A). 200 independent experiments were run for each value of $\Lambda = 15, 16, \cdots, 25$ db. The leading edge and optical matched filter, and linearized EM estimator were implemented for each experiment. For each experiment the linearized EM estimate was iterated until convergence was achieved. Since estimator performance is shift-invariant, we kept the value of the parameter $\tau_0$ constant throughout the simulations. The sample mean of the individual squared errors was then calculated and normalized with respect to the variance of a uniform random variable over the prior interval.

In Fig. 3(a) a simulated observation waveform is plotted as a function of $t$ for $\Lambda = 23$ dB (200 events). The corresponding optimal linear estimate $\widetilde{dN_t}$ discussed in Section III is plotted in Fig. 3(b) as a function of $t$. Ideally, if the estimate is accurate, $\widetilde{dN_t}$ should resemble a train of impulses located at the arrival times $\{t_i\}_{i=1}^n$. Observe from Fig. 3 that $\widetilde{dN_t}$ appears to operate as a high-pass filter that sharpens the peaks

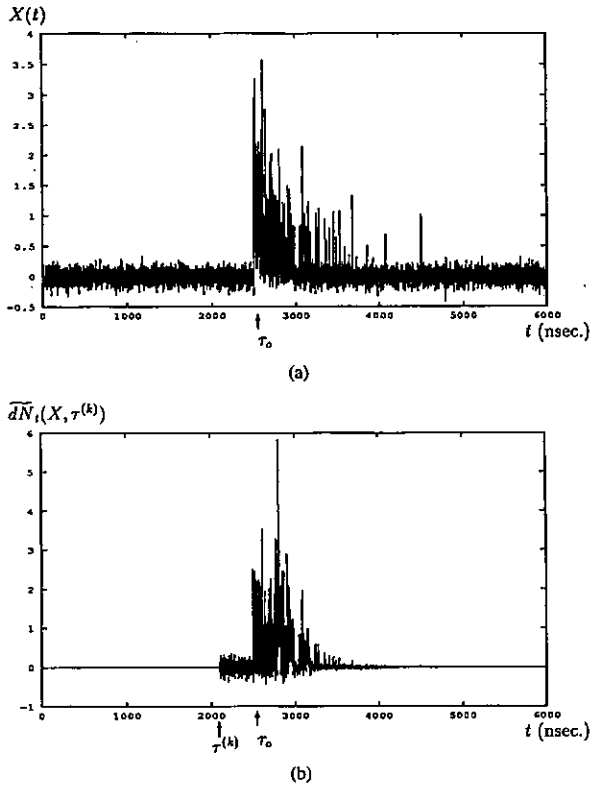



Fig. 3. (a) Simulated observation data for $\Lambda = 23$ dB,$\tau_0 = 2.5$ $\mu$s. (b) Output of the linear estimator od $dN_t$ for the data of Fig. 3(a) and $\tau^{(k)} = 2100$ ns.

of the signal, thereby enhancing visibility of the event times of the Poisson process.

In [8] it was shown that the Cramer-Rao (CR) lower bound on timing performance is trivial in several situations (e.g., when the scintillation pulse is rapidly varying) and a tighter lower bound on the post-detection timing error was derived. The RMS errors of all four estimators are plotted in Fig. 4 along with the unbiased form of the CR lower bound as a function of $\Lambda$. In Figs. 5 and 6, the bias of all four estimators is plotted. We see that the bias of the linearized EM estimator is very close to zero and that the linearized EM estimator virtually achieves this CR bound for high values of the mean rate and comes very close for moderate values of $\Lambda$ as well.

Based on these simulations we make the following conclusions for the above example:

- Our EM-type algorithm significantly outperforms both the leading edge and the matched filter estimators for moderate values of $\Lambda$. For very high values of $\Lambda$ (not shown in our graphs), all estimators have almost identical performace.
- The EM estimator comes very close to the lower bound derived in [18]. Hence, it is a virtually optimal estimation scheme.
- Unlike the other estimators studied, the bias of $\hat{\tau}^{\mathrm{EM}}$ is very low, even for moderate values of $\Lambda$.
- The EM algorithm converges very quickly, usually requiring only two or three iterations.

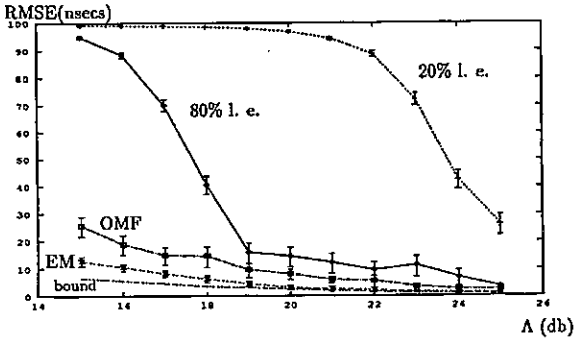



Fig. 4. Root mean square error (RMSE) of two leading edge (l.e.), the optical matched filter (OMF), and the linearized EM (EM) estimators (95% confidence intervals are indicated). The unbiased CR bound (bound) is plotted for comparison.

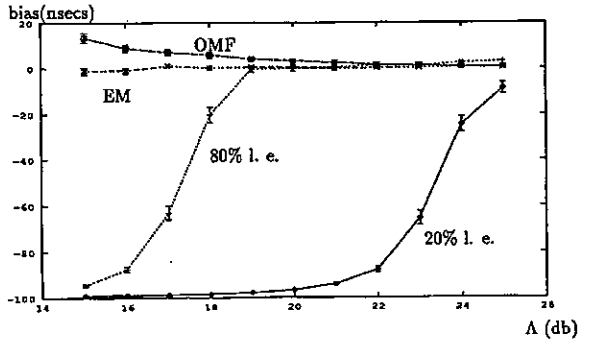



Fig. 5. Bias of two leading edge (l.e.), the optical matched filter (OMF), and the linearized EM (EM) estimators (95% confidence intervals are indicated).

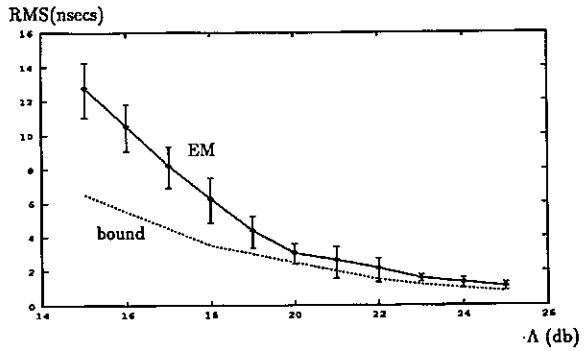



Fig. 6. Blow up of Fig. 5 to show error of linearized EM estimates (EM) and the unbiased CR lower bound (bound).

## VIII. Conclusion

We have presented an approach that makes use of the EM algorithm to estimate the time delay of the inhomogeneous component of a filtered Poisson process. For large square mean-to-variance ratio $\mu_a^2/\sigma_a^2$ of the gain sequence, the algorithm exhibits quick convergence and holds a clear advantage in terms of MSE over previous timing techniques. Our model may be applied to any time-varying intensity function and the approach easily generalizes to arbitrary parameters $\tau$ of the

intensity $\lambda$. Robustness issues remain to be addressed, since the algorithm assumes that the shapes of the intensity and the SPR pulse are known.

It is useful to contrast the observation model considered here with other similar models considered in geophysical exploration (detection of underground explosions, seismic deconvolution for geologic strata analysis, etc.). In these latter applications, observations are taken in discrete time and the intensity of the process $dN_t$ is constant. The objective is to estimate the arrival times and the associated gains. Several optimal and suboptimal deconvolution algorithms have been developed, such as single most likely replacement (SMLR) and its variations [19], [20]. They give satisfactory results, but do not extend to the time-varying intensity case. However, in the time-delay estimation context, the event sequence is not explicitly needed. The algorithm makes use of an *estimate* of the arrival times, $\widetilde{dN}_t$ only insofar as it is needed to produce an estimate of time shift. Indeed, as may be seen from Fig. 3(b), the linear estimate $\widetilde{dN}_t$ is not restricted to the range [0, 1] and it may even take on negative values. It is remarkable that such an estimate is sufficient to give a highly accurate estimate of an intensity shift parameter.

## APPENDIX A

An analytical expression for the inverse $R_X^{-1}(t,u)$ of the autocovariance function $R_X(t,u)$ is difficult to obtain in the general case, since it is required that an integral equation of the form

$$\int R_X(t,s)\phi(s)ds = \lambda\phi(t) \tag{41}$$

be solved, where $\phi(t)$ are orthonormal eigenfunctions associated with $R_X$ and $\lambda$ are corresponding eigenvalues [21]. (41) is not easy to solve and an exact solution can be derived only when the function is stationary and has a rational Fourier transform. Obviously, this is a restricted class of functions and it does not contain the covariance function of a filtered Poisson process with inhomogeneous intensity. However, under the assumption of a causal, exponentially decaying $p(t)$, we may compute an approximate inverse. First we discretize the temporal arguments $t$ and $u$ : $t, u \in \{j\Delta\}_{j=0}^{M}$, where $\Delta > 0$ and $M\Delta = T$. In order to account for the effect of sampling, we will integrate $p(t)$ and $\lambda(t)$ between successive samples. Using (20), the discrete-time equivalent covariance matrix $\Sigma_x = ((R_X(i\Delta, j\Delta)))_{i,j=0,\cdots,M-1}$ is

$$\Sigma_X = (\sigma_a^2 + \mu_a^2)\cdot P \cdot \operatorname{diag}\{\tilde{\lambda}(i)\}\cdot P^T + \sigma_w^2 I$$

where $I$ is the $M \times M$ identity matrix and $P$ is a matrix whose columns are shifted versions of the integrated samples $\tilde{p}(k) = \int_{k\Delta}^{(k+1)\Delta} p(t)dt, k = 0, \cdots, M-1$. For causal $P(t)$, $P$ has the form

$$P = \begin{bmatrix} \tilde{p}(0) & 0 & & \cdots & 0 \\ \tilde{p}(1) & \tilde{p}(0) & 0 & \cdots & \\ \tilde{p}(2) & \tilde{p}(1) & \tilde{p}(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \tilde{p}(M-1) & \tilde{p}(M-2) & \cdots & & \tilde{p}(0) \end{bmatrix}.$$

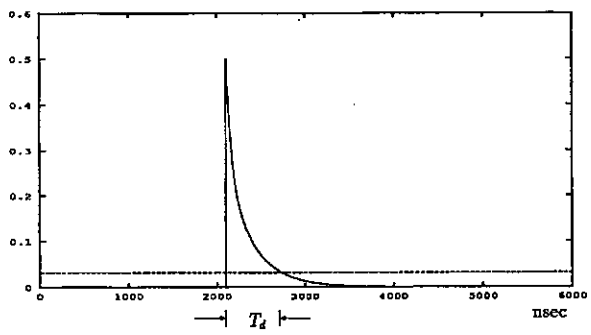



Fig. 7. The time interval of length $T_d$ defining the blocks in (42) for a monoexponential intensity function.

The matrix $\Sigma_X$ can be approximated by a block diagonal matrix under the assumptions of high SNR. In particular, when the covariance of the signal process is much larger than the variance of the thermal noise and the dark current is low, we may be able to find a time interval where $(\sigma_a^2 + \mu_a^2)P \operatorname{diag}\{\tilde{\lambda}(i)\}P^T$ dominates, while the white noise dominates the dark current component in the rest of the observation interval. Then we may write

$$\Sigma_X \approx \begin{bmatrix} \sigma_w^2 I & 0 & 0 \\ 0 & (\mu_a^2 + \sigma_a^2)P_d\tilde{\Lambda}_d P_d^T & 0 \\ 0 & 0 & \sigma_w^2 I \end{bmatrix} \tag{42}$$

where $P_d$ and $\tilde{\Lambda}_d$ are $T_d \times T_d$ submatrices of $P$ and diag $\left(\tilde{\lambda}(i)\right)$, respectively, and $T_d$ is defined as the length of the time interval over which the covariance of the signal process dominates the noise. For any given value of the relevant parameters it is straightforward to find $T_d$ and the resulting time intervals, as is evident from Fig. 7.

Under the above assumptions, the inverse matrix is written as follows

$$\Sigma_X^{-1} \approx \begin{bmatrix} \frac{1}{\sigma_w^2} I & 0 & 0 \\ 0 & \frac{1}{\mu_a^2+\sigma_a^2} P_d^{-T}\tilde{\Lambda}_d^{-1}P_d^{-1} & 0 \\ 0 & 0 & \frac{1}{\sigma_w^2} I \end{bmatrix}. \tag{43}$$

Let $p(t)$ be of exponential type: $p(t) = \frac{1}{T_p}e^{-\frac{t}{T_p}}, t \geq 0$. Then the discrete-time energy equivalent pulse is

$$p(k) = \begin{cases} cd^k, & k \geq 0 \\ 0, & k < 0 \end{cases}$$

where $d = e^{-\frac{\Delta}{T_p}}$ and $c = \sqrt{2\Delta/\left[(1-d^2)\ln\left(\frac{1}{d}\right)\right]}$. It can then be easily verified that

$$P^{-1} = \frac{1}{c}\cdot \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ -d & 1 & 0 & & \\ \vdots & \ddots & \ddots & \ddots & \\ & & -d & 1 & 0 \\ 0 & \cdots & & -d & 1 \end{bmatrix}.$$

Now the matrix $P_d^{-T}\Lambda_d^{-1}P_d^{-1}$ in (43) is written as follows (see equation at the bottom of the next page), where $\tau_d$ is the time point where the interval of length $T_d$ starts.

## APPENDIX B

Let $\{N_t; 0 \le t < T\}$ be a Poisson process and $\{t_i\}_{i=1}^n$ the arrival times of its events. We define the differential increment of the process as

$$dN_t = \lim_{\Delta t \to 0+} [N_{t+\Delta t} - N_t].$$

The increment may take only the values 0 or 1.

In dealing with Poisson processes, the following stochastic function is of particular interest

$$f(t) = \sum_{i=1}^{n} g(t - t_i)$$

where $n$ is the number of the Poisson events in the time interval $[0, T]$ and $g$ is a deterministic function of $t$. Using the binary nature of $dN_t$ we may rewrite (44) in a more convenient way as follows

$$f(t) = \int_0^T dN_u \; g(t-u).$$

The integral can be interpreted as a Riemann-Stieltjes integral having the evaluation

$$\int_0^T dN_u g(t-u) = \begin{cases} 0, & N_T = 0 \\ \sum_{i=1}^{n} g(t-t_i), & N_T \ge 1 \end{cases} \quad (44)$$

which is called the *stochastic integral representation* of the Poisson process.

## APPENDIX C

*Proposition 1:* Fix the point $v$ and let $\tilde{Q}(u \mid v)$ take its maximum value at the point $u = \hat{\tau}$. Assume that $\frac{1}{\Lambda}\tilde{Q}(u \mid v)$ is differentiable in $u$ and that as $\Lambda \to \infty$ it converges mean square to a nonrandom function $\bar{Q}(u \mid v)$ which is differentiable and unimodal taking its unique maximum value at $u = \overline{\tau}$. Then $\hat{\tau}$ converges in probability to $\overline{\tau}$.

*Proof:* By differentiability we can define $q(u) = \partial/\partial_u \frac{1}{\Lambda}\tilde{Q}(u \mid v)$, $\overline{q}(u) = \partial/\partial_u \bar{Q}(u \mid v)$, and $\Delta q(u) = q(u) - \overline{q}(u)$. By unimodality $\overline{q}(u)$ is a nonincreasing function of $u$ with unique zero at $u = \hat{\tau}$, i.e., for all $u > 0$

$$\overline{q}(\overline{\tau} + u) < 0 < \overline{q}(\overline{\tau} - u). \quad (45)$$

For arbitrary $u > 0$ consider the probability

$$\begin{aligned} &P(q(\overline{\tau}+u) < 0 < q(\overline{\tau}-u)) \\ &\quad = P(\Delta q(\overline{\tau}+u) + \overline{q}(\overline{\tau}+u) \\ &\qquad < 0 < \Delta q(\overline{\tau}-u) + \overline{q}(\overline{\tau}-u)) \\ &\quad = P(\Delta q(\overline{\tau}+u) < -\overline{q}(\overline{\tau}+u), -\Delta q(\overline{\tau}-u) < \overline{q}(\overline{\tau}-u)) \\ &\quad \ge P\Big(\sqrt{|\Delta q(\overline{\tau}+u)|^2 + |\Delta q(\overline{\tau}-u)|^2} \\ &\quad < \min\{-q(\overline{\tau}+u), \overline{q}(\overline{\tau}-u)\}\Big). \end{aligned}$$

Where the above inequality follows from (45). Using the Tchebychev inequality we obtain for any $u > 0$

$$\begin{aligned} &P(q(\overline{\tau}+u) < 0 < q(\overline{\tau}-u)) \\ &\qquad \ge 1 - \frac{E\left[|\Delta q(\overline{\tau}+u)|^2\right] + E\left[|\Delta q(\overline{\tau}-u)|^2\right]}{\min\{|\overline{q}(\overline{\tau}+u)|^2, |\overline{q}(\overline{\tau}-u)|^2\}}. \end{aligned}$$

By the m.s. convergence assumption the expectations on the right side of the Tchebychev inequality converge to zero. Therefore we have established that for all $u > 0$

$$\lim_{\Lambda \to \infty} P(q(\overline{\tau}+u) < 0 < q(\overline{\tau}-u)) = 1. \quad (46)$$

and hence $\hat{\tau}$ converges to $\overline{\tau}$ in probability [22]. □

## REFERENCES


[1] F. Gatti and V. Svelto, "Review of theories and experiments of resolving time with scintillation counters," *Nucl. Instruments, Methods*, vol. 43, pp. 248–268, 1966.

[2] R. M. Gagliardi and S. Karp, *Optical Communications.* New York: Wiley, 1976.

[3] R. Schoonhoven, D. F. Stegeman, and A. van Oosterom, "The forward problem in electroneurography-II: Comparison of models," *IEEE Trans. Biomed. Eng.* vol. BME-33, pp. 335–341, 1986.

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

[5] G. F. Knoll, *Radiation, Detection and Measurement.* New York: Wiley, 1979.

[6] I. Bar-David, "Minimum-mean-square-error estimation of photon pulse delay," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 3, pp. 326–330, 1975.

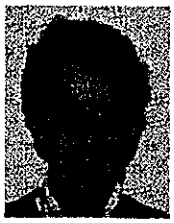

$$P_d^{-T}\Lambda_d^{-1}P_d^{-1} \approx \frac{1}{c^2}$$

$$\begin{bmatrix} \frac{1}{\lambda(\tau_d)} + \frac{d^2}{\lambda(\tau_d+\Delta)} & -\frac{d}{\lambda(\tau_d)+\Delta} & 0 & \cdots & 0 \\ -\frac{1}{\lambda(\tau_d+\Delta)} & \frac{1}{\lambda(\tau_d+\Delta)} + \frac{d^2}{\lambda(\tau_d+2\Delta)} & -\frac{1}{\lambda(\tau_d+2\Delta)} & \ddots & \vdots \\ 0 & -\frac{1}{\lambda(\tau_d+2\Delta)} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \frac{1}{\lambda(\tau_d+(T_d-1)\Delta)} + \frac{d^2}{\lambda(\tau_d+T_d\Delta)} & -\frac{d}{\lambda(\tau_d+T_d\Delta)} \\ 0 & \cdots & 0 & -\frac{d}{\lambda(\tau_d+T_d\Delta)} & \frac{1}{\lambda(\tau_d+T_d\Delta)} \end{bmatrix}$$

[7] T. Tomitani, "A maximum likelihood approach to timing in scintillation counters," in *Proc. IEEE Workshop Time-of-Flight Tomography*, 1982, pp. 89–93.
[8] N. A. Petrick, A. O. Hero, N. H. Clinthorne, and W. L. Rogers, "A method for improved time-of-arrival estimation for weak optical pulses with applications to scintillation detectors," *IEEE Trans. Nucl. Sci.*, vol. 38, no. 2, pp. 174–177, 1991.
[9] A. O. Hero, "Timing estimation for a filtered Poisson process in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 37, no. 1, pp. 92–106, 1991.
[10] T. T. Kadota, "Approximately optimum detection of deterministic signal in Gaussian and compound Poisson noise," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1517–1527, 1988.
[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc. B*, vol. 39, pp. 1–37, 1977.
[12] D. L. Snyder, *Random Point Processes*. New York: Wiley, 1975.
[13] V. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
[14] H. T. El-Hadidi and B. Hirosaki, "The Bayes optimal receiver for digital fiber optic communications systems," *Opt. Quantum Electron.*, vol. 13, pp. 469–486, 1981.
[15] A. O. Hero, N. Antoniadis, N. Clinthorne, W. L. Rogers, and G. D. Hutchins, "Optimal and sub-optimal post-detection timing estimators for PET," *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 725–729, 1990.
[16] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
[17] N. Antoniadis, "Time delay estimation for inhomogeneous Poisson processes in the presence of Gaussian noise," EECS Dept., The University of Michigan, Ann Arbor, MI, Sept. 1992.
[18] N. Clinthorne, N. Petrick, W. L. Rogers, and A. O. Hero, "A fundamental limit on timing performance with scintillation detectors," *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 658–663, 1990.
[19] J. J. Kormylo and J. M. Mendel, "Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 3, pp. 482–488, 1982.
[20] C.-Y. Chi and J. M. Mendel, "Improved maximum-likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 429–435, Mar. 1984.
[21] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill, 1958.
[22] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.

**Nikolaos Antoniadis** received the B.S. degree from Ptychion, Aristotelian University, Thessaloniki, Greece, in 1986. He received the M.S.E. in electrical engineering: systems, the M.S. in bioengineering, and the Ph.D. in electrical engineering: systems from the University of Michigan, Ann Arbor, in 1989, 1990, and 1992, respectively.

While at the University of Michigan, he was supported by a Research Assistantship (1988–1991 and 1991–1992). Since 1992, he has been serving in the Greek Army. His current interests are in statistical signal processing, biosystems, estimation and detection theory, and tomographic imaging.



**Alfred O. Hero III** (S'80–M'84) was born in Boston, MA, on December 5, 1955. He received the B.S. degree (summa cum laude) from Boston University, Boston, MA, and the Ph.D. degree from Princeton University, Princeton, NJ, in 1980 and 1984, respectively, both in electrical engineering.

He is presently Associate Professor of Electrical Engineering and Computer Science (EECS) at the University of Michigan, Ann Arbor. He has held positions of Visiting Scientist at M.I.T. Lincoln Laboratory, Lexington, MA (1987–1989), Visiting Professor at the Ecole Nationale de Techniques Avancees (ENSTA), Paris, France (1991), and William Clay Ford Fellow at Ford Motor Company (1993). His present research interests are in the areas of detection and estimation theory, applications of statistical signal processing in manufacturing, tomographic imaging, and antenna array processing.

Dr. Hero is a member of Tau Beta Pi, the New York Academy of Sciences, and Commision C of the International Union of Radio Science (URSI). He was Chairman for Publicity for the 1986 IEEE International Symposium on Information Theory. He is General Chairman for the 1995 International Conference on Acoustics, Speech, and Signal Processing.