Lower variance bounds for time estimation in TOF-PET.

Michel Defrise[1], Johan Nuyts[2], Emilie Roncali[3], Carlotta Trigila[3], Stefan Gundacker[4], Paul Lecoq[5].

Abstract. Data quality in TOF-PET depends on the accuracy of the detector time resolution (DTR). Given a model of the time distribution of the optical photons emission and of the photon transport and detection, the best achievable DTR can be obtained from a lower bound on the variance of any unbiased estimator of the interaction time of the incident annihilation photon. With current detectors hundreds of optical photons are typically detected and the Cramer-Rao bound (CRB) can be used because of its asymptotic tightness property. With some novel fast detectors involving e.g. Cerenkov photons the emission rate is low and the CRB yields an over-optimistic value of the DTR. We compare three methods to estimate the DTR given a model of the probability density function of the optical photon emission. The first method uses an approximation of the Barankin bound [12], [14] (BB). The two other methods use Monte-Carlo simulation to calculate the variance of the maximum likelihood (ML) estimator and the minimum mean square error (MMSE) estimator. Results for two detector models with a mixture of prompt and of slow emissions show that the CRB is not reliable when the number of detected optical photons is small. The DTR estimated using the BB, ML and MMSE methods are in all cases close to each other.

I. Introduction

One key parameter in TOF-PET is the accuracy with which the interaction time of an annihilation photon in the detector can be estimated by measuring the emitted optical photons. Currently, detector time resolutions (DTR) around 160 ps FWHM are available in clinical scanners, and results from several groups open the perspective of breakthroughs, aiming ideally at the goal of 10 ps coincidence time resolution (CTR) [1]–[3]. Promising approaches are based on the measurement of Cerenkov light or on the use of metamaterials, which interleave thin layers of dense scintillators and of fast scintillators. These systems lead to different patterns for the time distribution of emitted photons, as a function of the number of prompt photons detected on top of the standard scintillation emission. Depending on their origin (Cerenkov, cross-luminescence, quantum-confined scintillation in nanocrystals), the number of these prompt photons can vary from a handful to several hundreds.

[1]Vrije Universiteit Brussel (Belgium) [2]KULeuven (Belgium), [3]UC Davis, California (USA), [4]RWTH Aachen University (Germany), [5]Polytechnic University of Valencia (Spain).

The DTR is primarily determined by the physical characteristics of the detector, including light transport, photosensors, and electronics [4]. The digital or analogue algorithm used to produce an estimate $\hat{\theta}$ of the interaction time θ of the γ photon is a final component, which also influences the DTR but is easier to modify than the detector. Hence the question: given the physical characteristics of a detector, what is the best achievable intrinsic - i.e. independent of the algorithm - DTR? We consider this problem for an ideal situation where the detection times $t_1, t_2, \ldots, t_N$ of the N detected optical photons for each γ event can be measured. It is also assumed that the t_n are independent identically distributed (i.i.d.) random variables with a known probability density function $p(t|\theta)$, which models the detector timing characteristics. We characterize the DTR by the square root of the variance of $\hat{\theta}$. One way to answer this standard question in estimation problems is to derive a lower bound on the variance of any unbiased estimator of the parameter.

The Cramer-Rao lower bound (CRB) on the variance of estimators (see e.g. [5]) has been applied to the interaction time estimation for TOF-PET detectors [6]–[9]. An attractive asymptotic property of the CRB is that when N is large it is tight and is reached by the maximum-likelihood (ML) estimator, therefore yielding reliable values for the intrinsic DTR. However, when N is lower than $\simeq 100$ there is in general no guarantee that an unbiased estimator with variance equal to the CRB exists; one says that the CRB is not tight and it would in that case lead to over-optimistic values of the DTR, as noted in 1990 by Clinthorne et al [10], [11], see also [2]. As will be seen below, even the often assumed $1/\sqrt{N}$ dependence of the DTR on the number of detected photons does not hold for low N. Another limitation of the CRB is that its derivation requires assumptions on the differentiability of $p(t|\theta)$, which are not always satisfied at $t = \theta$ (see e.g. [5]).

This work revisits an alternative lower variance bound, the Barankin bound (BB) [12]–[14]. Contrary to the CRB, the BB is tight. Unfortunately the BB is much more complex to compute than the CRB and in practice one uses lower approximations of the BB. Section III presents the Barankin bound for the estimation of a single parameter and describes the approximation used here, which is similar to that proposed in [14]. A tight, or almost tight, lower variance bound is useful not only by yielding a reliable value of the DTR, but also because it allows verifying the efficiency of an algorithm proposed to estimate the gamma interaction time from the detected scintillation photons: if the variance obtained with this estimator is higher than the bound, we know that there is

room for improvement, i.e. the proposed algorithm does not optimally exploits the detector.

We will compare the CRB and Barankin bounds with DTR estimates obtained using the Monte-Carlo method and a specific algorithm to estimate for each simulated data set the arrival time. Besides being time consuming, the Monte-Carlo method is clearly not independent of the algorithm. It yields a valid estimate of the intrinsic DTR only if the chosen estimation algorithm is known to achieve the lowest possible estimator variance. Fortunately the estimator with minimum mean square error (MMSE) can be calculated explicitely for the time estimation problem (section IV), and it will be used below to verify the coherence with the Barankin bound. The Monte-Carlo method will also be applied with the maximum likelihood (ML) estimator, which is easier to calculate than the MMSE and is by far the most widely used algorithm for time estimation. The ML estimator is only asymptotically efficient, and one goal of this paper is to verify if it optimally exploits the data for detectors with low photon yields.

Two simple models of the probability density function (pdf) of the photon emission have been considered in the simulation. The modeled detectors combine two different emission mechanisms, each one described by the smoothed bi-exponential pdf used e.g. in [4], [6]–[9], [11], [18], [19]. We use the same gaussian optical transport function as [6], [8] and do not model dark current. The first pdf roughly simulates a metamaterial detector interleaving layers of LYSO and of the plastic scintillator EJ232 [3]. The second detector has a bi-modal pdf chosen to further challenge the ML and MMSE estimators. The last part of the results section VII considers the same models with two types of emission mechanisms, but now assuming that the detector is able to identify the emission mechanism of each detected photon. A preliminary version of this work was presented in [20]. This paper includes in addition theory and results for the minimum mean-square error estimator, and results for detectors which allow identifying the type of each detected optical photon.

II. Notations

We denote $p(t|\theta)$ the pdf of the detection time t of a scintillation photon emitted after the interaction of a gamma in the detector at time θ. We assume that $p(t|\theta) = 0$ when $t < \theta$; if needed this assumption can always be satisfied by shifting the definition of θ by the largest possible error of the acquisition electronics. Invariance for time translation will be assumed so that $p(t|\theta) = p(t-\theta)$, using the same symbol p with some abuse of notations. The cumulative distribution function (cdf) is denoted

$$P(t|\theta) = \int_\theta^t p(t'|\theta)\, dt'. \tag{1}$$

The goal is the estimation of the interaction time θ, given the detection times $\vec{t} = (t_1, .., t_N) \in \mathbb{R}^N$ of N optical photons. All emissions are supposed independent, hence the joint pdf is the product

$$\mathcal{L}(\vec{t}, \theta) = \Pi_{j=1}^N p(t_j|\theta). \tag{2}$$

Variance bounds will be computed for a fixed N. As noted in [6] this has a limited impact when N is large, otherwise one should take additional expectations on the distribution of N.

Because early scintillation emissions carry most timing information on the gamma interaction, good estimates of the parameter θ can often be obtained using only the few first photons, as demonstrated in [6], [18]. A practical benefit of this approach is that only a small number of early photon detection times must be recorded, the following photons must only be counted to determine N and estimate the energy of the γ photon. Following [21] we note the time of the n-th detected photon by $t_{(n)}$, so that the sorted data is $t_{(1)} \le t_{(2)} \le t_{(3)} \cdots \le t_{(N)}$. The joint pdf of the $n \le N$ first photons is, for $\theta = 0$ [21],

$$\begin{aligned} & p_{n,N}(t_{(1)}, t_{(2)}, \ldots, t_{(n)}) \\ = \; & \frac{N!}{(N-n)!}(1 - P(t_{(n)}))^{N-n}\, \Pi_{j=1}^n p(t_{(j)}). \end{aligned} \tag{3}$$

An estimator $\hat{\theta} = F(\vec{t})$, defined by some function $F : \mathbb{R}^N \to \mathbb{R}$, is useful only if it shares the invariance of the problem for time translations, i.e. if for any $s \in \mathbb{R}$, $F(t_1 - s, \ldots, t_N - s) = F(t_1, \ldots, t_N) - s$. The bias

$$E(F(\vec{t})|\theta) - \theta = \int_\theta^\infty \ldots \int_\theta^\infty F(\vec{t})\mathcal{L}(\vec{t}, \theta) dt_1 \ldots dt_N - \theta \tag{4}$$

of such an equivariant estimator is independent of θ and can be corrected without impact on the variance of $\hat{\theta}$. Only unbiased estimators are considered in this paper.

III. Barankin bound

Contrary to the Cramer-Rao bound, the Barankin bound [12], [14], [22] requires no differentiability condition on the pdf, and it applies without restriction when -as in our case- the region where $p(t|\theta) > 0$ depends on the unknown parameter θ. Even when the Cramer-Rao bound is applicable, the Barankin bound provides tighter bounds on the variance of an estimator.

Consider first a single measurement t, with any pdf $p(t|\theta)$ which depends on a single parameter θ and is such that $p(t|\theta) = 0$ when $t < \theta$ and $p(t|\theta) > 0$ for $t > \theta$. A discrete approximation of the Barankin bound is defined by choosing some positive integer J and by selecting J positive parameters $\vec{\delta} = (\delta_1, \cdots, \delta_J)$, the choice of which will be discussed later. Define the functions

$$\eta_j(t, \theta) = \begin{cases} \frac{p(t|\theta+\delta_j)}{p(t|\theta)} - 1 & t \ge \theta \\ 0 & t < \theta \end{cases} \qquad j = 1, \ldots, J \tag{5}$$

Because $p(t|\theta + \delta_j) = 0$ when $t < \theta + \delta_j$, one has $\eta_j(t, \theta) = -1$ when $\theta < t < \theta + \delta_j$ and the functions defined by (5) are bounded. Define also the $J \times J$ symmetric non-negative definite matrix U with elements

$$\begin{aligned} U_{j,j'}(\theta) &= E\left(\eta_j(t, \theta)\eta_{j'}(t, \theta)\,|\,\theta\right) \qquad j, j' = 1, \ldots, J \\ &= \int_\theta^\infty \eta_j(t, \theta)\eta_{j'}(t, \theta) p(t|\theta) dt \end{aligned} \tag{6}$$

For the timing problem $\eta_j(t, \theta) = \eta_j(t - \theta, 0)$ and therefore the matrix U is independent of θ.

One shows that the variance of any unbiased estimator $\hat{\theta} = F(t)$ is bounded below by ([12], see also eqn (7) in [14]):

$$\text{Var}(\hat{\theta}) \geq \vec{\delta}^T U^\dagger(\theta)\vec{\delta} = \sum_{j=1}^{J}\sum_{j'=1}^{J} \delta_j U^\dagger_{j,j'}(\theta)\delta_{j'} \qquad (7)$$

where $U^\dagger(\theta)$ is the pseudo-inverse of the matrix $U(\theta)$. A link with the CRB (when it is applicable) is seen by noting that for $J = 1$, $\delta^{-2}U_{1,1}(\theta)$ tends to the Fisher information when $\delta \to 0$. By using multiple finite δ_j the Barankin bound can not only look at the local curvature, but also at the shape of $p(t|\theta)$ far away from the solution.

For N i.i.d. measurements $t_j, j = 1, \ldots, N$ the variance of any unbiased estimator $\hat{\theta} = F(\vec{t})$ satisfies the same inequality (7), with U replaced by the matrix U_N with elements (see Appendix 1):

$$U_{N,j,j'}(\theta) = \left(U_{j,j'}(\theta) + 1\right)^N - 1 \qquad j, j' = 1, \ldots, J. \qquad (8)$$

Equation (7) yields a valid variance bound for any choice of the J shift values δ_j. To obtain the tightest possible bound for a fixed J, one needs to maximize the bound with respect to the shift values δ_j but this is difficult because no guidance is available on the optimal choice. In addition, as will be illustrated below, the bound improves with increasing number of shifts J and one is facing the numerical difficulty that U tends to be ill-conditionned for large J.

IV. The MMSE estimator

To define the minimum mean square error estimator, one introduces a prior probability density function $\pi(\theta)$ of the interaction time θ. Given measured data $\vec{t} = (t_1, \ldots, t_N)$ the posterior distribution of θ is

$$\mathcal{P}(\theta|\vec{t}) = \frac{\pi(\theta)\mathcal{L}(\vec{t},\theta)}{\int_{-\infty}^{\infty}\pi(\theta')\mathcal{L}(\vec{t},\theta')d\theta'} \qquad (9)$$

where $\mathcal{L}(\vec{t},\theta)$ is the data likelihood. The MMSE estimator is the posterior mean of the parameter [5], [15], [16],

$$\begin{aligned}\hat{\theta}_{MMSE}(\vec{t}) &= E(\theta|\vec{t}) = \int_{-\infty}^{\infty}\mathcal{P}(\vec{t}|\theta)\,\theta\,d\theta \\ &= \frac{\int_{-\infty}^{\infty}\pi(\theta)\,\mathcal{L}(\vec{t},\theta)\,\theta\,d\theta}{\int_{-\infty}^{\infty}\pi(\theta')\,\mathcal{L}(\vec{t},\theta')\,d\theta'}, \qquad (10)\end{aligned}$$

and it minimizes the mean square error

$$e^2(\hat{\theta}) = E((\hat{\theta}-\theta)^2) = \int \pi(\theta)\int \mathcal{L}(\vec{t},\theta)(\hat{\theta}(\vec{t})-\theta)^2\,d\vec{t}\,d\theta \qquad (11)$$

i.e. $e^2(\hat{\theta}_{MMSE}) \leq e^2(\hat{\theta})$ for any estimator $\hat{\theta}$.

For the timing problem with detection of N optical photons measured at times t_j i.i.d. with pdf $p(t-\theta)$, the likelihood is given by eq. (2) and, assuming that no additional information is available, one takes the limit where the function $\pi(\theta)$ is constant within a sufficiently large time interval. The prior pdf $\pi(\theta)$ then cancels in (10) and the MMSE estimator becomes (eq. (6) in [11]):

$$\hat{\theta}_{MMSE} = \frac{\int_{-\infty}^{t_{(1)}}\theta\,\Pi_{j=1}^{N}p(t_j-\theta)\,d\theta}{\int_{-\infty}^{t_{(1)}}\Pi_{j=1}^{N}p(t_j-\theta')\,d\theta'} \qquad (12)$$

TABLE I: Parameters of the prompt and scintillation emissions (ns). Decay time τ_d, rise time τ_r, and mean t_{tr} and standard deviation σ_{tr} of the gaussian model $\mathcal{N}(t_{tr}, \sigma_{tr})$ of the optical transport. See equations (23, 24) in the Appendix.

	τ_d	τ_r	t_{tr}	σ_{tr}
LYSO/EJ232				
prompts	1.6	0.35	0.179	0.081
scintillation	40	0.01	0.179	0.081
BIMOD				
prompts	1	0.01	0.179	0.081
scintillation	20	5	0.179	0.081

where the upper limit $t_{(1)} = \min\{t_1, \ldots, t_N\}$ of the integrals reflects the fact that, for the timing problem, $p(t) = 0$ for $t < 0$. Equation (12) can be rewritten as

$$\hat{\theta}_{MMSE} = t_{(1)} - \frac{\int_0^{\infty} u\,\Pi_{j=1}^{N}p(t_j - t_{(1)} + u)\,du}{\int_0^{\infty}\Pi_{j=1}^{N}p(t_j - t_{(1)} + u)\,du} \qquad (13)$$

where $u = t_{(1)} - \theta$.

When the interaction time is estimated using only the $n < N$ first photons $t_{(1)}, \ldots, t_{(n)}$ the MMSE estimator is obtained by inserting the pdf (3) in (13):

$$\begin{aligned}\hat{\theta}_{MMSE,n} &= t_{(1)} \\ &- \frac{\int_0^{\infty} u\,p_{n,N}(u, t_{(2)} - t_{(1)} + u, \ldots, t_{(n)} - t_{(1)} + u)\,du}{\int_0^{\infty} p_{n,N}(u, t_{(2)} - t_{(1)} + u, \ldots, t_{(n)} - t_{(1)} + u)\,du}. \qquad (14)\end{aligned}$$

The MMSE estimator (13) is unbiased and is therefore also the minimum variance unbiased estimator.

V. The simulated emission models

We consider two hypothetical detectors which emit both prompt photons and scintillation photons. The two emission pdf $p^c(t|\theta)$ and $p^s(t|\theta)$ are modeled by the smoothed bi-exponential model used in [6] (eqn. (5)) and in [7], see Appendix 2. The optical transport in the crystal is modeled by a gaussian $\mathcal{N}(0.179$ ns, 0.081 ns). The first detector roughly approximates interleaved layers of LYSO and of the plastic scintillator EJ232 [1], [3]. The parameters of the second detector, named BIMOD, are chosen to obtain a bi-modal pdf, which more severely tests the estimators. The parameters of these two models are given in Table 1, and the corresponding pdf $p(t|\theta) = \alpha p^c(t|\theta) + (1-\alpha)p^s(t|\theta)$ are shown in Figure 1 for a prompt fraction rate $\alpha = 0.05$.

VI. Implementation

To approximate the BB we defined a fixed set of 80 values of the shifts δ, with a geometric progression between two empirically chosen values δ_{min} and δ_{max}. For each value of J we selected randomly V_δ subsets of J different shifts from this fixed set. Table II gives the parameters used for the two detector models. For each such subset $\delta_1, \cdots, \delta_J$ the matrix U_N was calculated by numerical integration of eqn. (6) with uniform step $dt = 0.1$ ps. This yielded V_δ BB values; the

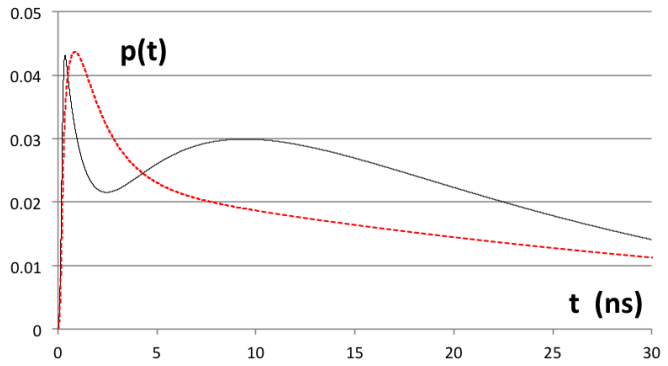

Fig. 1: The probability density function of the emission of detected optical photons for two detectors. LYSO/EJ232: red dotted, BIMOD: black.

data below report the highest of these values. Even though this crude sub-optimal optimization is not guaranteed to yield the tightest bound, each computed value is a valid bound because the inequality (7) holds for any choice of the J shifts. The pseudo-inverse $U_N^\dagger$ was calculated by singular value decomposition. To limit the risk of numerical errors singular values smaller than 10^{-8} were not included when calculating (7); this omission lowers the computed BB value because U_N is definite positive; therefore the obtained value remains a valid lower bound for the variance of $\hat{\theta}$. Test runs showed that the obtained BB, but also the ill-conditioning of U_N, increase with increasing number of shifts J, before reaching a plateau. An example is shown by Figure 2. The value $J = 16$ was selected as a compromise between convergence, reasonable condition number, and computation time.

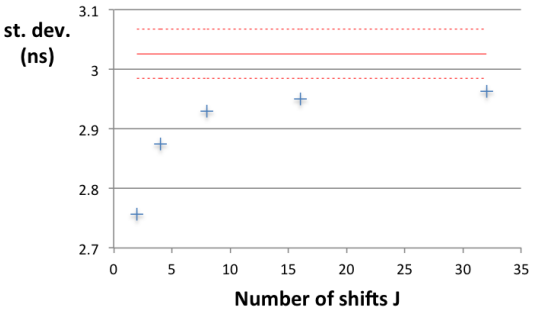

Fig. 2: The Barankin bound (+) as function of J for the BIMOD detector with $\alpha = 5$ % prompt fraction and $N = 10$ detected photons. The red lines indicate the value of the MMSE estimators $\pm$ one standard error.

The Cramer-Rao bound

$$\mathrm{CRB} = \frac{1}{N}\left\{\int_\theta^\infty \frac{(\partial p(t-\theta)/\partial\theta)^2}{p(t-\theta)}dt\right\}^{-1} \qquad (15)$$

was calculated by integration with uniform steps $dt = 0.1$ ps and truncation at $t_{max} = 200$ ns, and the partial derivative was estimated with a two point finite difference.

TABLE II: Range of the shift values (ns) used to compute the approximation of the Barankin bound, and number V_δ of tested sets of shifts, for different number N of detected photons.

N	4	10	40	100	900
LYSO/EJ232					
δ_{min}	0.001	0.001	0.001	0.0005	0.0005
δ_{max}	100	100	10	10	10
V_δ	8000	8000	8000	8000	30000
BIMOD					
δ_{min}	0.005	0.005	0.005	0.005	0.005
δ_{max}	25	25	25	25	25
V_δ	10000	10000	10000	10000	10000

A Monte-Carlo simulation was used to estimate the variance of the ML and MMSE estimators of the arrival time, with respectively 5000 and 10000 Monte-Carlo trials for the LYSO and BIMOD detectors, each with a fixed number of detected photons N. For each trial the detection times $t_1, \ldots, t_N$ of the optical photons were generated using a lookup table of size 20000 of the inverse function of the cumulative distribution function $\alpha P^c(t|\theta) + (1-\alpha)P^s(t|\theta)$, built with a time sampling 0.1 ps. The statistical errors were estimated by dividing the 5000 trials into 10 independent subsets.

The ML estimate was obtained by exhaustive search with step 1 ps, in an interval $[t_{(1)} - 15 \text{ ns}, t_{(1)}]$, where $t_{(1)} = \min(t_1, \ldots, t_N)$. The obtained estimate $\hat{\theta}$ was refined by repeating the exhaustive search in a smaller interval $\hat{\theta} \pm 0.75$ ns and with a step 0.1 ps. The ML estimator is biased but, as noted in section II, the bias is independent of θ and can be corrected without impacting the variance.

The MMSE estimator was calculated by integrating (13) with a uniform sampling step 0.1 ps; the integrals were truncated at $t_{max} = 200$ ns. To avoid numerical underflow when computing the products $\Pi_j p(t_j - t_{(1)} + t)$ for large N, all the probabilities in the numerator and denominator in (13) where rescaled by their value at $t = t_{tr}$. This scaling was not needed when computing the MMSE using only the first n detected photons, as long as $n <\simeq 50$ (see (14) and (3)).

The computation times for to estimate the DTR for the BIMOD model ($N = 100$) were 1988s for the Barankin bound (with 8000 tested sets), 45600s for ML estimation (5000 Monte-Carlo trials) and 228200s for MMSE (5000 Monte-Carlo trials). These figures for a non-optimized implementation on a MacPro processor (2,3 GHz Intel Core i5) only give a rough estimate of the computational complexity but indicate that the Barankin bound is the fastest method to estimate the DTR, given a model of the emission pdf.

A. Labeling

Consider an hypothetical detector, which has two emission mechanisms but is able to identify the type (prompt or scintillation) of each detected photon. The ML estimate of θ is obtained by maximizing the likelihood

$$\mathcal{L}(\vec{t}, N_c, N_s, \theta) = \Pi_{j=1}^{N_c} p^c(t_j|\theta)\, \Pi_{j=1}^{N_s} p^s(t_{j+N_c}|\theta) \qquad (16)$$

in the same way as without labeling. In equation (16) the indices of the detected photons were reordered in such a way

that $t_1, t_2, \cdots, t_{N_c}$ are the times of the N_c prompt photons and $t_{N_c+1}, \cdots, t_N$ are the times of the $N_s = N - N_c$ scintillation photons. One difference with the case without labeling is that N_c and N_s are known and therefore the ML estimator can be corrected using for each event the specific bias $\Delta\theta(N_c, N_s)$ that corresponds to N_c and N_s:

$$\Delta\theta(N_c, N_s) = E\left(\arg\max_{\theta'} \mathcal{L}(\vec{t}, N_c, N_s, \theta') \middle| N_c, N_s, \theta\right) - \theta. \tag{17}$$

This bias was estimated by Monte-Carlo simulation (with given prompt fraction α). This bias correction need not be done for the MMSE because it is unbiased.

When photons are labeled, the Barankin bound is computed separately for each (N_c, N_s), and the obtained bounds are averaged with a binomial weight $\mathcal{B}(N, \alpha)$.

Numerical results with labeled detection are presented in section VII.C for the BIMOD detector.

VII. Results

A. LYSO/EJ232

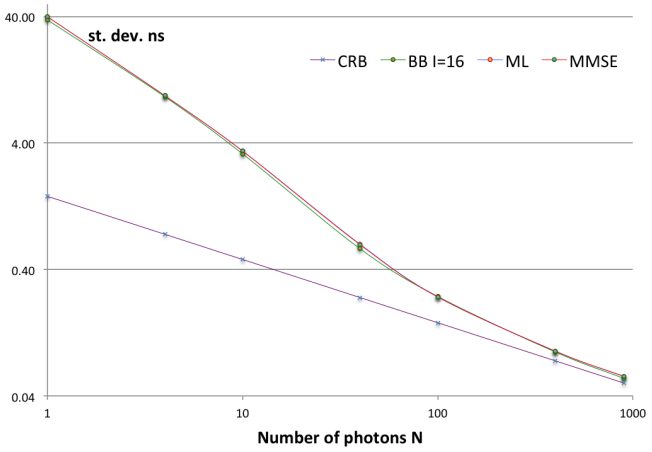

Fig. 3: The standard deviation (ns) of the maximum-likelihood and minimum mean square error estimators of the arrival time θ for the LYSO/EJ232 detector with $\alpha = 5$ % prompt fraction, and the corresponding Cramer-Rao and Barankin lower bounds. Log-log scale.

Figure 3 shows the standard deviation (ns) of the estimators $\hat{\theta}_{MMSE}$ and $\hat{\theta}_{ML}$ and the Cramer-Rao and Barankin (J=16) lower bounds, as a function of the number of detected optical photons in the LYSO/EJ232 detector model. The data points are linked with smooth lines, which - except for the CRB - almost overlap on this logarithmic vertical scale. For better visualization, Figure 4 plots for the same data the ratios with the BB with displayed error bars ($\pm$ 1 standard error). A power fit of the standard deviation of $\hat{\theta}_{ML}$ for $1 \leq N \leq 40$ in Figure 3 yields $N^{-1.12}$ compared to the usual $N^{-0.5}$ for the CRB. The power fit for for $100 \leq N \leq 900$ yields $N^{-0.67}$. No local maximum was observed during the exhaustive likelihood optimization. The data were also computed for the LYSO/EJ232 detector without prompts. Figure 5 shows the DTR gain due to the introduction of prompts photons, calculated here as the ratio between the standard deviation of the MMSE estimator with $\alpha = 0$ and with $\alpha = 0.05$.

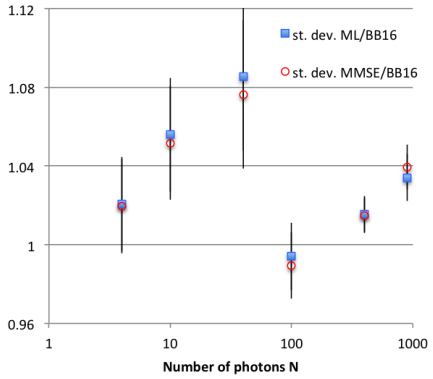

Fig. 4: The ratio between the standard deviation (ns) of the maximum-likelihood and minimum mean square error estimators and the Barankin lower bounds for the LYSO/EJ232 detector with $\alpha = 5$ % prompt fraction.

Figure 6 confirms the observation made by previous authors [6], [18] that the arrival time can be estimated using only the first detected optical photons, without significantly degrading the DTR. This simplifies data acquisition and significantly accelerates data processing. The data are shown for the maximum-likelihood estimate but the same result is observed for the MMSE estimate, and also for the BIMOD detector.

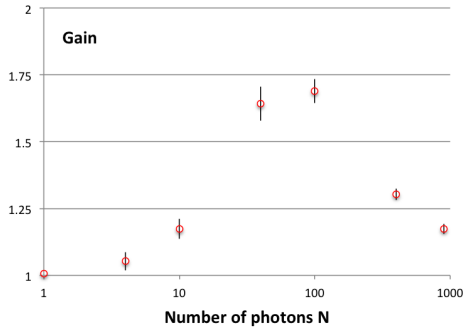

Fig. 5: Ratio between the standard deviation of the MMSE estimator with $\alpha = 0.0$ and with $\alpha = 0.05$, for the LYSO/EJ232 detector.

B. BIMOD

Figures 7 and 8 show qualitatively similar results for the BIMOD detector. As expected with this bi-modal pdf (see Figure 1), at least one local maximum of the likelihood was observed in respectively $5281, 475, 2$ of the Monte-Carlo trials (out of 10000) for $N = 4, 10, 40$.

C. BIMOD with labeling

Consider an hypothetical detector, which is able to identify the emission mechanism (prompt or scintillation) of each detected photon. To which extent would this additional information improve the timing resolution? To verify this we ran

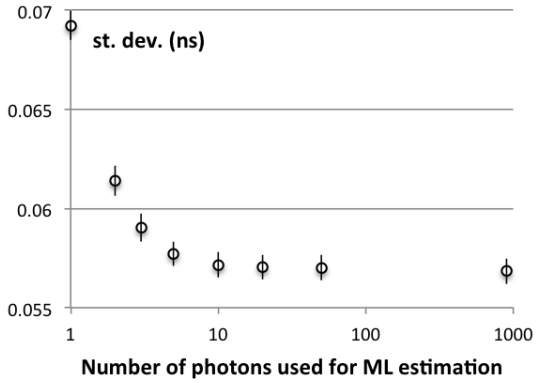

Fig. 6: Standard deviation of the ML estimator versus the number of first photons used to calculate the likelihood, for the LYSO/EJ232 detector with 5% prompts and a total of $N = 900$ photons detected.

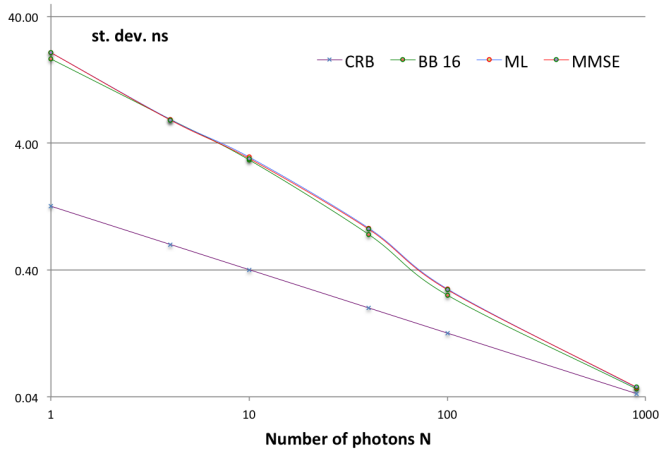

Fig. 7: The standard deviation (ns) of the maximum-likelihood and minimum mean square error estimators of the arrival time θ for the BIMOD detector with $\alpha = 5$ % prompt fraction, and the corresponding Cramer-Rao and Barankin lower bounds. Log-log scale.

the same Monte-Carlo simulation as above for the BIMOD parameters, using as in the previous subsections a fixed total number of photons N. The interaction time θ was estimated by maximizing the likelihood, with four different modes:

- i/ using all detected photons but without labeling (same data as in Figure 7),
- ii/ using all detected photons with labeling,
- iii/ using only the prompts photons,
- iv/ using only the scintillation photons.

Note that mode iii/ provides no estimate when no prompt photon is detected; this situation frequently occurs for small N because of the small prompt fraction $\alpha = 0.05$. Therefore, the estimate with mode iii/ used the scintillation photons for all events with no prompt. For all modes with labeling the ML estimate was bias corrected, with the bias corresponding to the known number of prompt photons in each event, thus according to (17).

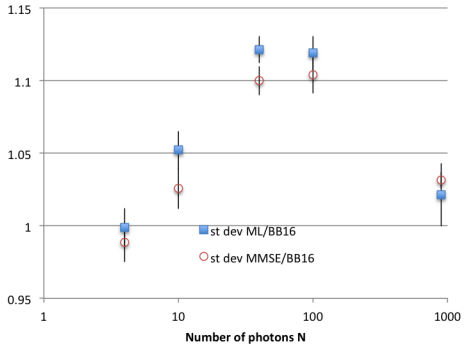

Fig. 8: Ratio between the standard deviation (ns) of the maximum-likelihood and minimum mean square error estimators and the Barankin lower bounds for the BIMOD detector with $\alpha = 5$ % prompt fraction.

Results are shown in Figure 9. To reduce the computation burden, the Barankin bound was calculated for $J = 8$ instead of $J = 16$ in previous subsections. This choice is justified a posteriori by the small gap observed in Figure 9 between the Barankin bound and the standard deviation of the ML estimate using all photons with labels. For these data, the difference between the standard deviation of the ML and MMSE estimators does not exceed 1.5%. These observations indicate that maximum likelihood estimation is close to optimal also when each event can be labeled.

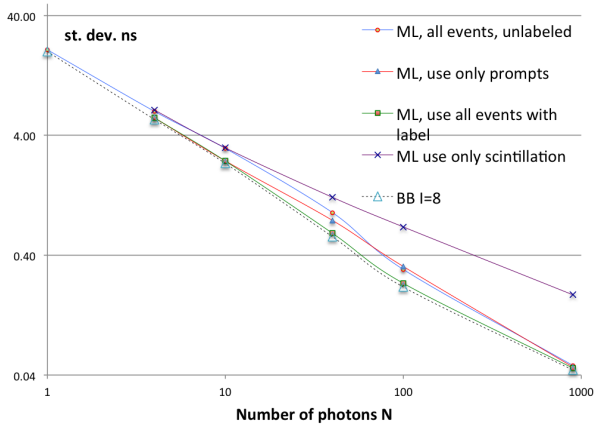

Fig. 9: The standard deviation (ns) of the maximum-likelihood of the arrival time θ for the BIMOD detector with $\alpha = 5$ % prompt fraction, and the corresponding Barankin lower bounds. The detector is able to identify the type of each detected photon as either prompt or scintillation. Log-log scale.

VIII. Conclusion

This paper compares methods to calculate the time resolution of PET detectors, given an analytical or numerical model of the probability density function of the optical photons emitted after the interaction of an annihilation photon. Todate clinical PET scanners are based on detectors with high photon

yields, typically emitting several hundred scintillation photons. Asymptotic statistical properties hold to a good approximation in this case and the Cramer-Rao lower variance bound yields a reliable estimate of the DTR. Alternatively, a Monte-Carlo estimate of the DTR can be obtained by applying the maximum likelihood estimation to simulated data. For this reason the applicability of the CRB and ML estimation is often taken for granted.

This study was motivated by recent developments of detectors with low photon yields. When the number of detected photons is low the asymptotic properties do not hold and the Cramer-Rao bound is not necessarily tight. This was observed in [10] for the detector timing problem, and the results in Figs. 3 and 7 show that the CRB can underestimate the DTR by more than one order of magnitude. Even with $N = 100$ photons, a relatively large photon yield, the underestimation is significant. This observation can be tentatively explained by the fact (illustrated by Fig. 6) that the DTR is mostly determined by the relatively rare fast events, so the number of "important" photons is actually smaller than the total number.

With low photon yields, there is also no guarantee that ML estimation is the method of choice. Nevertheless, our results show that the gap between the standard deviation of the ML estimate, the standard deviation of the MMSE estimate, and the Barankin bound is small even for low numbers of detected photons. One can conclude that the ML estimate almost optimally uses the data, even with low light yields. Interestingly this was also observed for the BIMOD detector despite the existence of multiple local maxima of the likelihood. With low photon yields the DTR for both detector models improves much faster with increasing number of detected photons than the familiar $1/\sqrt{N}$ dependence.

We used two methods to obtain the intrinsic -i.e. the best possible - DTR: the Barankin bound and the minimum-mean-square-error estimator. Both methods have their drawback. The Barankin bound is complex to calculate and only an approximation was obtained here; this approximation is a valid lower bound for the DTR but is not guaranteed to be the tightest possible bound. The MMSE has a convenient closed form expression for this problem but a time consuming Monte-Carlo simulation is required to obtain the DTR with sufficient statistical precision. Using these two independent methods allowed us to verify that the differences between the DTR values obtained from the ML estimate, MMSE and the Barankin approximation are probably not significant in practice in view of the unavoidable approximations in the emission model itself. This should of course be verified for each specific model.

As usual we used the variance of the estimated arrival time of the annihilation photon to caracterize the DTR. The CTR, measured by the variance of the TOF profile, can be computed from the DTR of the two detectors in coincidence. The relation is straightforward if the probability distribution function of the estimated arrival times is gaussian for the two detectors. Otherwise, knowing the two DTRs is not sufficient, and a Monte Carlo simulation is needed to estimate the two probability distribution functions, which must then be convolved.

The two probability density functions in the simulations use an unrealistic gaussian model of the optical transport and neglect the effect of dark current. Therefore the specific values of the DTR in section VII are only illustrations, but we expect that the methodological conclusions summarized above would be valid for more realistic detector models.

Appendix 1: Barankin bound for N i.i.d. measurements.

We derive here equation (8), which relates the matrix U_N for N detected photons to its expression for $N = 1$. See problem 5.23 p. 139 in [5].

In section III the Barankin bound was described for a single measurement t. Equation (7) also applies to the case of a N dimensional data $\vec{t} = (t_1, \ldots, t_N)$ by replacing everywhere t by $\vec{t}$ and the pdf $p(t|\theta)$ by the joint pdf $\mathcal{L}(\vec{t}|\theta)$. The matrix U_N is thus defined by

$$U_{N,j,j'}(\theta) = \int \ldots \int \eta_{N,j}(\vec{t},\theta)\eta_{N,j'}(\vec{t},\theta)\mathcal{L}(\vec{t}|\theta)dt_1 \ldots dt_N \quad (18)$$

In our case, the times $t_i, i = 1, \ldots, N$ of the N optical photons are i.i.d. with joint pdf given by (2), and the functions η in (5) become

$$\eta_{N,j}(\vec{t},\theta) = \begin{cases} \Pi_{i=1}^{N} \frac{p(t_i|\theta+\delta_j)}{p(t_i|\theta)} - 1 & t \geq \theta \\ 0 & t < \theta \end{cases} \qquad j = 1, \ldots, J \quad (19)$$

Using the pdf (2) and noting that all integrals are from $t = \theta$ to $t = \infty$ and that, with $\delta_j > 0$, $p(t|\theta + \delta_j) = 0$ when $t < \theta$, one has

$$\begin{aligned} & \int \ldots \int \Pi_{i=1}^{N} \frac{p(t_i|\theta+\delta_{j'})}{p(t_i|\theta)} \times \\ & \Pi_{i=1}^{N} \frac{p(t_i|\theta+\delta_j)}{p(t_i|\theta)} \mathcal{L}(\vec{t}|\theta)dt_1 \ldots dt_N \\ = & \left(\int \frac{p(t|\theta+\delta_j)}{p(t|\theta)} \frac{p(t|\theta+\delta_{j'})}{p(t|\theta)} p(t|\theta)dt \right)^N \\ = & \left(U_{j,j'}(\theta) + 1\right)^N \end{aligned} \quad (20)$$

and

$$\begin{aligned} & \int \ldots \int \Pi_{i=1}^{N} \frac{p(t_i|\theta+\delta_j)}{p(t_i|\theta)} \mathcal{L}(\vec{t}|\theta)dt_1 \ldots dt_N \\ = & \left(\int p(t|\theta+\delta_j)dt \right)^N = 1. \end{aligned} \quad (21)$$

Inserting this into (18) one obtains equation (8).

Appendix 2: the bi-exponential model

The probability density function of the photon emission after interaction of a γ at $\theta = 0$, is modeled following [4], [6]–[9], [11], [18], [19]. The bi-exponential model

$$\begin{aligned} p_{biexp}(t|0) &= \int_0^t dt' \frac{1}{\tau_r} e^{-t'/\tau_r} \frac{1}{\tau_d} e^{-(t-t')/\tau_d} \\ &= \begin{cases} \frac{1}{\tau_d - \tau_r} \left(e^{-t/\tau_d} - e^{-t/\tau_r} \right) & t \geq 0 \\ 0 & t < 0 \end{cases} \end{aligned} \quad (22)$$

where $\tau_r < \tau_d$, is convolved with a gaussian model $\mathcal{N}(t_{tr}, \sigma_{tr})$ of the optical transport time to obtain the probability density function used in [6] (eqs. (5,7,10)) and in section VII:

$$p(t|0) = \begin{cases} C\frac{1}{\tau_d - \tau_r}\left(a(t,\tau_d) - a(t,\tau_r)\right) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (23)$$

with

$$\begin{aligned} a(t,\tau) &= \frac{1}{\sqrt{2\pi}\sigma_{tr}} \int_0^t dt'\, e^{-t'/\tau} e^{-(t-t'-t_{tr})^2/2\sigma_{tr}^2} \\ &= \frac{1}{2} e^{(-\frac{t-t_{tr}}{\tau} + \frac{\sigma_{tr}^2}{2\tau^2})} \times \\ &\times \left(\text{erf}(\frac{t - t_{tr} - \frac{\sigma_{tr}^2}{\tau}}{\sqrt{2}\sigma_{tr}}) + \text{erf}(\frac{t_{tr} + \frac{\sigma_{tr}^2}{\tau}}{\sqrt{2}\sigma_{tr}}) \right) \\ C &= \frac{2}{1 + \text{erf}(\frac{t_{tr}}{\sqrt{2}\sigma_{tr}})} \\ \text{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2)\, dy \end{aligned} \quad (24)$$

An alternative form to improve numerical stability is

$$\begin{aligned} a(t,\tau) &= \frac{1}{2} e^{(-\frac{t-t_{tr}}{\tau} + \frac{\sigma_{tr}^2}{2\tau^2})} \times \\ &\times \left(\text{erfc}(\frac{t_{tr} + \frac{\sigma_{tr}^2}{\tau} - t}{\sqrt{2}\sigma_{tr}}) - \text{erfc}(\frac{t_{tr} + \frac{\sigma_{tr}^2}{\tau}}{\sqrt{2}\sigma_{tr}}) \right) \end{aligned} \quad (25)$$

For $t = \theta$

$$p(t|\theta) = \frac{\partial p(t|\theta)}{\partial t}|_{t=\theta} = 0 \quad (26)$$

Therefore the pdf is C^1 in $t = \theta$, hence the integrand for the calculation of the Fisher information has no singularity and the CRB is applicable.

REFERENCES

[1] P. Lecoq, Pushing the Limits in Time-of-Flight PET Imaging, IEEE Trans. Nucl. Sc. (2017), 1 (6), 473-485.

[2] D. R. Schaart, G. Schramm, J. Nuyts, S. Surti, "Time of Flight in Perspective: Instrumental and Computational Aspects of Time Resolution in Positron Emission Tomography", IEEE Trans. Rad. Plasma Med. Sc., 5 (5), 598-618, 2021.

[3] P. Lecoq et al, "Metascintillators: New results for TOFPET applications", IEEE Transactions on Radiation and Plasma Medical Sciences, 2022, 6 (5), 510-516.

[4] S. Derenzo, WS Choong and W.W. Mooses, "Fundamental limits of scintillation detector timing precision", Phys. Med. Biol. 59 (2014) 32613286.

[5] E. L. Lehmann and G. Casella, Theory of point estimation, Springer 1998.

[6] S. Seifert, H. T van Dam and D. R Schaart, Lower bound on the timing resolution of scintillation detectors, Phys Med Biol 57 (2012), 1797-1814.

[7] S. Gundacker, Time resolution in scintillator based detectors for positron emission tomography, PhD TU Wien, 2014

[8] S. Gundacker, R.M. Turtos, E. Auffray, P. Lecoq, Precise rise and decay time measurements of inorganic scintillators by means of X-ray and 511 keV excitation, Nucl. Instr. Meth. in Phys. A 891 (2018) 4252.

[9] F. Loignon-Houle, E. Bertrand, M. Toussaint, F. Camirand Lemyre and R. Lecomte, Lower bound on timing resolution of TOF-PET detectors producing prompt photons, IEEE NSS-MIC 2021.

[10] N. H. Clinthorne, A. O. Hero, N. A. Petrick, W. L. Rogers, Nuclear Instruments and Methods in Physics Research A299 (1990) 157- 161

[11] N. H. Clinthorne, N. A. Petrick, W. L. Rogers, and A. 0. Hero, A Fundamental Limit on Timing Performance with Scintillation Detectors, IEEE Trans. Nucl. Sc. (1990), 658-663.

[12] E. W. Barankin, Locally best unbiased estimate, Ann. Math. Statist. 20(4): 477-501, 1949.

[13] J. Albuquerque, The Barankin bound: A geometric interpretation (Corresp.), IEEE Transactions on Information Theory, vol. 19, no. 4, pp. 559-561, July 1973, doi: 10.1109/TIT.1973.1055038.

[14] J. D. Gorman, A. O. Hero, Lower Bounds For Parametric Estimation with Constraints, IEEE Trans. Inf. Theory, 26 (6), 1990, 1285-1301.

[15] P. H. Garthwaite, I. T. Jolliffe and B. Jones, Statistical Inference, Oxford Science Publications 2002

[16] H. H. Barrett, K. J. Myers, Foundations of Image Science, Wiley 2004.

[17] M. Ruis-Gonzalez, V. Bora and L. Furenlid, "Maximum-Likelihood Estimation of Scintillation Pulse Timing", IEEE Trans Radiat Plasma Med Sci. 2018 January ; 2(1): 16. doi:10.1109/TRPMS.2017.2765316

[18] R. Vinke, P. D. Olcott, J. W. Cates and C. S. Levin, The lower timing resolution bound for scintillators with non-negligible optical photon transport in time-of-flight PET, Phys. Med. Biol. 59 (2014), 6215-6229.

[19] J. W. Cates, R. Vinke and Craig S Levin, Analytical calculation of the lower bound on timing resolution for PET scintillation detectors comprising high-aspect-ratio crystal elements, Phys. Med. Biol. 60 (2015), 5141-5161.

[20] M. Defrise, J. Nuyts, E. Roncali, C. Trigila, S. Gundacker and P. Lecoq, The Barankin Bound for Time Estimation in TOF-PET, 2022 IEEE Nuclear Science Symposium, Medical Imaging and Room-Temperatyure Semi-conductor Detector Workshop.

[21] H. A. David , Order Statistics, Wiley 1970

[22] Douglas G. Chapman, Herbert Robbins, Minimum Variance Estimation Without Regularity Assumptions, Ann. Math. Statist. 22(4): 581-586 (1951)