System Tradeoffs in Gamma-Ray Detection Utilizing SPAD Arrays and Scintillators

Matthew W. Fishburn and Edoardo Charbon

Abstract—We present a statistical analysis of the tradeoffs between detector jitter and light detection efficiency for TOF PET gamma-ray detectors based on SPAD arrays and crystal scintillators. Results show that increasing the light detection probability is more important to improving the coincidence timing resolution than decreasing the detector jitter for modern scintillators. For a SPAD TDC array with a fill factor around 15% and a detector jitter of 120 ps, it is shown that a SiPM with a 30% fill factor and a jitter of 240 ps will produce a better timing resolution when using a LYSO crystal. Results also imply that SPAD TDC arrays might be competitive in TOF PET with SiPMs if faster scintillators are developed in the future and SPAD TDC arrays maintain a much lower jitter than SiPMs.

Index Terms—Silicon photomultiplier, SPAD TDC Array, TOF PET.

I. INTRODUCTION

NEARLY all positron emission tomography (PET) systems use scintillators in their gamma-ray detection components. To detect a scintillator's light output, most PET systems use photomultiplier tubes (PMTs), which rarely work in strong magnetic fields. Single-photon avalanche diode (SPAD) arrays, however, are known to be unaffected by strong magnetic fields [1]. These detectors create the possibility of a dual MRI/PET system, an advantage that PMTs do not currently share.

With the recent advent in scintillator-based systems of silicon photo-multipliers (SiPMs), large arrays of SPADs and quenching circuitry in parallel, it seems likely that prototype PET systems using SPAD time-to-digital converter (TDC) arrays will soon appear. Rather than place the SPADs in parallel like in an SiPM, a SPAD TDC array contains a TDC connected to every SPAD or shared by some number of SPADs. SPAD TDC arrays are generally products of standard CMOS processes [2]–[4]. SPAD TDC arrays can have better single-photon spatial resolution, but at the expense of a complex digital interface and a reduced fill factor due to the placement of the TDC electronics on-chip. Additionally, as reducing the parasitic capacitance coupled to a SPAD can improve timing resolution by reducing the required ionization, SPAD TDC arrays may have better timing performance than SiPMs [5].

M. W. Fishburn is with the Delft University of Technology, Delft 2628 CD, Netherlands (e-mail: m.w.fishburn@tudelft.nl; e.charbon@tudelft.nl).

E. Charbon is with the Delft University of Technology, Delft 2628 CD, Netherlands.

Recently, microlens-recovery has been proposed to address the fill factor problem [6]. These arrays could have fill factors that are comparable to sparser SiPMs, and this paper examines whether current SPAD TDC arrays with microlens-recovered fill factors can compete with SiPMs in time-of-flight (TOF) PET applications, though this approach should also apply to the more general fill factor versus detector jitter tradeoff. The approach is based on knowledge gained from use of the Megaframe system [7].

II. OVERVIEW OF COMPONENTS

This section reviews the properties of SPAD arrays, and then discusses the basics of gamma-ray detection when these arrays are coupled to a scintillator.

A. Review of SPAD Arrays

The data received from a SPAD TDC array is a series of time stamps, usually streamed so the location of the photon arrival can be discerned. The time stamps correspond to avalanches in the SPADs, either originating from noise or photoelectrons. Photoelectron diffusion and digitization errors distort the observations of the initiation times of the avalanches. In comparison, the signal from an SiPM is generally the current into parallel SPAD and quenching circuitry pairs. Aside from the same types of SPAD jitter, analog sources of jitter, such as *RC* delays or amplifier non-idealities, can distort SiPM signals.

To prevent a SPAD from heating and destroying itself after an avalanche, the electric field in the diode must be lowered to quench the avalanche. While the electric field builds up again, the diode is not sensitive to photoelectrons, and so the time following an avalanche is termed dead time. Dead time will deteriorate the number of observed photons, as photons could impinge on a detector with a low electric field and not cause an avalanche. There is a tradeoff between dead time and after-pulsing probability, as some time must pass for traps to empty in the diode, otherwise after-pulsing will increase the noise. Dead time in passively quenched SPADs ranges from roughly 50 ns to several microseconds, though active quenching can decrease the dead time [8].

Another factor degrading SPAD performance is noise. Due to the inherent noise of SPADs, not all observed avalanches are true photon arrivals. Some avalanches correspond to noise from manufacturing defects, band-to-band quantum tunneling, crosstalk, or even stray light. These false events will have an adverse effect on estimating the properties of incident gamma-rays. Many SPAD arrays have noise rates in the low MHz range, but many factors can increase or decrease this noise rate.

B. Scintillator to SPAD Array Coupling

Much like a PMT, SPAD arrays are not very sensitive to gamma-rays, and thus are usually coupled to a scintillator. When absorbing a gamma-ray, the scintillator is assumed to give off n_s scintillation photons that are observed by the SPAD array with probability $P(\lambda) \cdot G$, termed the light detection efficiency. $P(\lambda)$ is the probability of an avalanche given that a photon with wavelength λ impinges on the SPAD's active region, and G is the probability that a scintillation photon will impinge on a SPAD's active region. $P(\lambda)$ is known as the photon detection probability (PDP), and G is commonly called the geometry factor. G is a function of the optical coupling efficiency and the fill factor of the SPAD array. It should be noted that fill factors and PDPs quoted in the literature for SPAD arrays generally assume light to be orthogonally incident to the array; various effects deteriorate these numbers for non-orthogonal light.

Since similar terms describe errors in estimates of both the gamma-ray arrival time and the arrival time of the scintillation photons, the terms "gamma-ray timing resolution" or "gamma-ray jitter" will be used to refer to errors in the gamma-ray arrival time estimate, whereas the terms "SPAD timing resolution" or "SPAD jitter" will refer to the error in the arrival time estimates of scintillation photons.

Neglecting dead time and noise, the SPAD array is expected to observe $\bar{n} = n_s \cdot P(\lambda) \cdot G$ photons. Each of these photons generates a photoelectron, and the term primary photoelectron will be used to describe a scintillation photon-generated photoelectron that would cause an avalanche if dead time is ignored. The actual number of primary photoelectrons during detection of a gamma-ray will be denoted by n.

Based on the array output of both noisy and true events, a system can estimate a gamma-ray's arrival time. This paper will model the distribution of the gamma-ray's arrival time for various SPAD jitters and light detection efficiencies. Standard SPAD jitters range from roughly 100 ps FWHM in fully digital SPAD TDC arrays to several hundred picoseconds in silicon photo-multipliers, though jitters as small as 70 ps sigma have been measured in SiPMs [9]–[12]. Light detection efficiencies range from over 25% for optimized silicon photo-multipliers with high fill factors and PDPs to under 5% in some modern SPAD TDC arrays with low fill factors [11], [13] . When coupled to a LYSO crystal that gives off roughly 14 500 scintillation photons per 511 keV gamma-ray, the number of primary photoelectrons would range from roughly 3600 for a detector with a light detection efficiency of 25% to under 150 for a detector with a light detection efficiency of 1% [14].

C. Energy Resolution

Assuming that each scintillation photon has an equal probability of being observed and the total number of observed photons is used, then the energy resolution following from statistics is roughly $2.35/\sqrt{\bar{n}}$, with $\bar{n}$ being the expected number of observed photons. A 20% energy resolution, used in this paper as an example resolution, corresponds to observing approximately 135 photons. Dead time, crosstalk and after-pulsing degrade this estimate, and systems usually correct for these factors.

Some systems use the photopeak out of the detector rather than the total number of photons. If the detector's SPADs' dead times are an order of magnitude larger than the decay time of the scintillator, then the current into the array will closely track the number of observed photons, and the photopeak from the SiPM will be a good approximation of the total number of observed photons. For the slowest scintillator examined in this study, LYSO, the dead time should be around 500 ns to meet this approximation. If the dead time is the same order of magnitude as the scintillator decay, then the mean number of primary photoelectrons will be higher than the number derived from the energy resolution, with the relation varying based on the dead time.

III. Method for Modeling Arrival Time Estimation

Using a model based on non-parametric statistics, this section will derive an estimate of the FWHM coincidence timing error of the two anti-parallel gamma-rays from an annihilation event. First, an overview of the scintillator light output and order statistics is given. Next, a simplified model is used to display the general use of order statistics in estimating this error. Finally, a more complete model with simulated results is presented, along with a discussion of factors which degrade the estimate.

A. Order Statistics Model

The probability of n arrivals between times 0 and t is often modeled as

$$P_n(t) = \frac{R(t)^n \cdot \exp(-R(t))}{n!}$$

with $R(t)$ being the expected number of primary photoelectrons observed between 0 and t. If the normalized number of primary photoelectrons is $F(t) = R(t)/R(\infty)$ with the final, expected number of observed primary photoelectrons being $R(\infty)$, then usually $F(t)$ is modeled as a single exponential

$$F(t) = 1 - \exp\left(-\frac{t}{\tau_d}\right) \tag{1}$$

with τ_d being a decay constant [15]. As electronics have improved and faster scintillators have become available, newer work has modeled $F(t)$ as a double-exponential with an additional rise component, τ_r [16]

$$F(t) = 1 - \frac{\tau_d \cdot \exp\left(-\frac{t}{\tau_d}\right) - \tau_r \cdot \exp\left(-\frac{t}{\tau_r}\right)}{\tau_d - \tau_r}. \tag{2}$$

The differences between modeling the scintillator emission as mono- or bi-exponential have been studied in detail elsewhere [17].

A scintillator works by transferring the gamma-ray's energy over a number of electrons, which usually emit photons during recombination [18]. Assuming the electron's recombination timings are independently and identically distributed and the number of primary photoelectrons, n, is available, then the kth primary photoelectron's generation time's probability density

function (PDF) $f_{k:n}(t)$ can be computed using the standard result [19]

$$f_{k:n}(t) = n \binom{n-1}{k-1} f(t) F(t)^{k-1} (1 - F(t))^{n-k} \tag{3}$$

where $f(t)$ and $F(t)$ are the PDF and cumulative density function (CDF), respectively, of the photoelectron generation process. $F(t)$ is presented in (1) for the single exponential case, while the CDF for the double-exponential case is presented in (2). The PDF for the double-exponential case is

$$\frac{dF(t)}{dt} = f(t) = \frac{\exp\left(-\frac{t}{\tau_d}\right) - \exp\left(-\frac{t}{\tau_r}\right)}{\tau_d - \tau_r}. \tag{4}$$

B. Simplified Case

This section will derive the minimum-variance, gamma-ray arrival time estimator based on a specific order statistic given by data from an ideal SPAD TDC array coupled to a scintillator with a single exponential decay. An ideal SiPM would have the same derivation, though the method for obtaining time stamps would differ.

Ignoring electronic- and geometry-induced inaccuracies in the arrival time, assuming no prior information about the gamma-ray arrival time, and assuming that the emitted scintillator photons are independently and identically distributed from the exponential process with PDF $f(t) = 1/\tau_d \exp(-(t - t_0)/\tau_d)$ when $t > t_0$, otherwise $f(t) = 0$, and unknown arrival time t_0, then the problem of estimating the arrival time becomes tractable using only elementary results from the theory of order statistics with an elegant, closed-form solution.

Following absorption of a gamma-ray by the scintillator, the stream of n time stamps $T_1 \dots T_n$ from the SPAD TDC array can be sorted to give the order statistics of an exponential distribution, $T_{1:n} \dots T_{n:n}$. These statistics can be written using the Rènyi representation [19], a special case of (3)

$$T_{i:n} = t_0 + \tau_d \sum_{k=1}^{i} \frac{Z_k}{n - k + 1} \tag{5}$$

where the Z_k variables are independently and identically distributed variables from a exponential process with a decay constant $\tau_Z = 1 = \lambda_Z$. As the order statistics are the sum of independent variables with $E[Z_k] = 1$ and $\text{var}(Z_k) = 1$, the mean, variances and covariances are relatively straight-forward to derive from the characteristics of the independent variables

$$E[T_{i:n}] = t_0 + \tau_d \sum_{k=1}^{i} \frac{1}{n - k + 1}$$

$$\text{var}(T_{i:n}) = \tau_d^2 \sum_{k=1}^{i} \frac{1}{(n - k + 1)^2}$$

$$\text{covar}(T_{i:n}, T_{j:n}) = \text{var}(T_{\min(i,j):n}).$$

Of particular note is that the covariance between two samples is the same as the variance of the smaller sample.

The unbiased estimator of t_0 for one particular order statistic will be $T_{i:n} - \tau_d \sum_{k=1}^{i} 1/n - k + 1$, with a variance of $\text{var}(T_{i:n})$. The general form for the unbiased, linear estimator of t_0 will be

$$\tilde{t}_0 = \sum_{k=1}^{n} \alpha_k \left(T_{k:n} - \tau_d \sum_{k=1}^{i} \frac{1}{n - k + 1} \right)$$

with the constraint that the α_k variables sum to one. First it will be shown that the estimator variance for any two variable combinations is minimized when only the first order statistic is used, and then this will be generalized to all linear estimators. Thus, to minimize the variance in the estimate of t_0, all of the α variables should be 0 except for α_1, which should be one. This minimum-variance, linear estimator uses only the first order statistic in the estimate of t_0, which corresponds to using only the first photon arrival in estimating the gamma-ray's arrival time.

Let $T = \alpha \cdot T_{i:n} + (1 - \alpha) \cdot T_{j:n}$ with $i < j$. We wish to minimize $\text{var}(T)$, or

$$\alpha^2 \cdot \text{var}(T_{i:n}) + (1 - \alpha)^2 \cdot \text{var}(T_{j:n})$$
$$+ 2 \cdot \alpha \cdot (1 - \alpha) \cdot \text{covar}(T_{i:n}, T_{j:n}).$$

Replacing the covariance using the result above, this can be simplified further to

$$\text{var}(T) = \text{var}(T_{i:n}) + (1 - \alpha)^2 \cdot [\text{var}(T_{j:n}) - \text{var}(T_{i:n})].$$

Since $i < j$ and hence $\text{var}(T_{i:n}) < \text{var}(T_{j:n})$, this expression is minimized when $\alpha = 1$, corresponding to using only the order statistic with the smaller variance. Hence, for all pairs, using only the first order statistic will give the minimum-variance, linear, unbiased estimate of the arrival time. This property generalizes to an estimator utilizing m order statistics having a higher variance than an estimator utilizing $m - 1$ order statistics if $m > 1$ and the weight associated with the order statistic of highest rank is instead transferred to the order statistic with lowest rank. By induction, the minimum-variance, linear estimator of t_0 will use only the 1st order statistic.

The expression for the unbiased, linear, minimum-variance estimator of t_0 is $T_{1:n} - \tau_d/n$, with variance $\text{var}(T_{i:n}) = (\tau_d/n)^2$.

Subtracting the arrival time estimates of the two gamma-rays from an annihilation event gives the coincidence timing estimate. If both detectors observe the same number of primary photoelectrons, this estimate is the difference of two independently and identically distributed variables, and the resulting PDF is the auto-correlation of the original distribution. The PDF is thus $n/2\tau_d \exp(-|tn/\tau_d|)$, with the FWHM being $2\tau_d/n \ln(2)$. The asymptotic behavior of this PDF is different than that of a normal distribution, and hence the FWHM of this distribution is somewhat misleading when other non-idealities distort the distribution.

In reality the detectors will rarely observe the same number of events, as there is shot noise in n. With shot noise, the PDF of the initial primary photoelectron's arrival time can be approximated

as a weighted sum of exponentials with similar decay time constants. Instead of the FWHM having exactly $1/n$ behavior, it will be approximately $1/n$. However, a more advanced model will show that the $1/n$ behavior may not hold when other considerations, such as SPAD jitter, are taken into account.

As the generating PDF deviates from a single exponential and as other effects, such as SPAD jitter, are included, the PDFs of the order statistics also deviate further and further from the Rènyi representation. Because of the absence of SPAD jitter, noise and the complex light output of scintillators in this model, the timing errors obtained are too low. Furthermore, there is no guarantee that the first order statistic will still be the optimal estimator of the gamma-ray's arrival time when including distortions from other effects. However, this model is still useful for validating new models' behaviors when a single exponential is a good approximation for the scintillator light output, as is the case when detectors have a poor light detection efficiency.

C. Advanced Model Overview

This section presents results from a model which includes digitization errors, timing jitter, and a double exponential PDF for the scintillator photon generation. All modeling was performed with version 0.6 of the Scipy library [20] and version 2.5 of the Python programming language [21]. Light emission was modeled for LYSO, LaBr$_3$ and an imaginary, extremely fast scintillator. Scintillator light emission was assumed to be a double-exponential as in (4). LYSO was assumed to have rise and decay times of 0.5 ns and 40 ns, respectively, whereas the rise and decay time of LaBr$_3$ was assumed to be 0.2 ns and 17 ns [16]. The imaginary scintillator was assumed to have 0.05 ns and 10 ns rise and decay times.

The SPAD jitter PDF can be convolved with the scintillator emission PDF to obtain a PDF for the observed arrival time of any photon, or the expected shape of the detected light curve. Assuming each scintillation photon is assumed to be independently and identically distributed from the same process, and given the number of primary photoelectrons, the PDF for the arrival time of the kth primary photoelectron (the kth order statistic) can be created using (3) and a sampled version of the detection PDF, in this case given by the convolution of (4) with a Gaussian curve representing the SPAD jitter. Fig. 1 displays the scintillator light output and the system light output along with the SPAD jitter. The SPAD jitter originates in several sources, as described in Section II-A. For this paper the FWHM of the SPAD jitter was assumed to be between 25 and 450 ps, with the distribution being Gaussian. In reality a SPAD's jitter will not be exactly Gaussian, with a wide range of factors distorting the jitter distribution, though a Gaussian fit has been shown to be a good approximation and allows comparison of a wide range of systems [5], [8].

By examining the FWHMs of the order statistics' PDFs, the minimum-FWHM, single-order-statistic-based estimator for the gamma-ray's arrival time can be found. In these simulations, the variances of the first five order statistics were examined. If a local minimum was found in first four variances, then the order statistics with this minimum variance was assumed to be the optimal one for the TOF measurement. If no local minimum of the
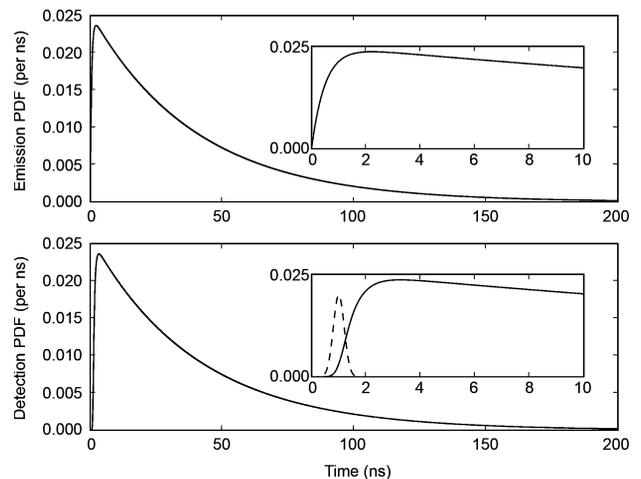
Fig. 1. Modeled emission and detection processes—Plotted against time are the PDF of a scintillation photon's emission from LYSO after absorbing a gamma-ray at time zero (solid curves in top plot and top inset) and the PDF of the observation time given that a SPAD array observes the photon (solid curves in bottom plot and bottom inset) with the detector having a Gaussian jitter (dashed curve with a .01 scaling factor in bottom inset). The emission PDF is (4) with a rise time of 0.5 ns and a decay time of 40 ns. The modeled jitter is a Gaussian curve with a FWHM of 0.45 ns in this example, though this value is variable. The detection PDF is the convolution of the emission PDF and detector jitter.
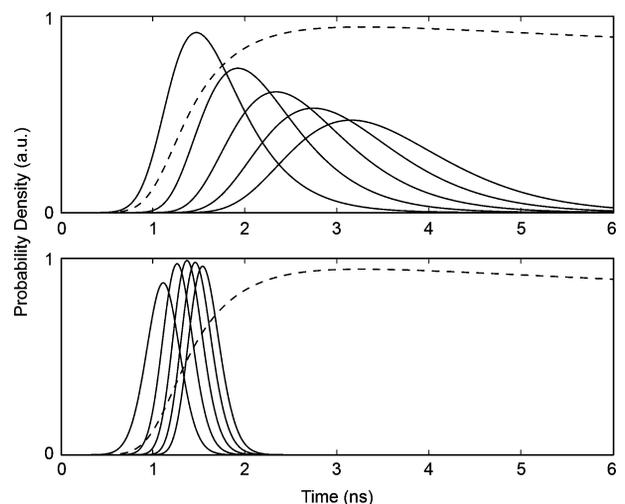
Fig. 2. Photon observation timing curves—The initial primary photoelectron observation time PDFs (solid) are overlaid on the detection PDF (dashed) for 100 (top) or 800 (bottom) primary photoelectrons in the example system from Fig. 1. The PDFs were scaled with the same factor within a subfigure to preserve shape information relative to one another, but the scaling is different between the top and bottom subfigures.

variance was found, then subsequent order statistics were examined until a local minimum was found. This estimator models thresholding techniques discussed in Section III-D. Fig. 2 superimposes the PDFs of the first- to fifth-order statistics onto the expected system light output for two systems with light detection efficiencies. For a specific system jitter, Fig. 3 displays the gamma-ray arrival time estimation error for various numbers of primary photoelectrons.

After the rank, k, of the order statistic which minimized the FWHM of the gamma-ray's arrival time for n primary
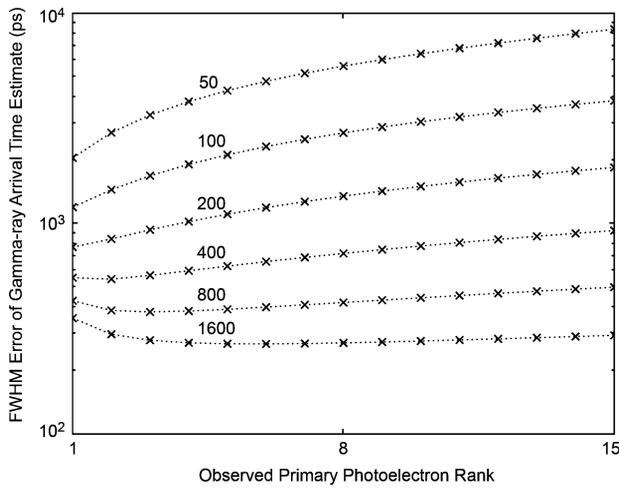
Fig. 3. Arrival time variances—The observation time variances of the initial primary photoelectrons are plotted versus the rank of the primary photoelectron when observed with a noise-free detector. As the light detection efficiency increases, the variances of the initial primary photoelectrons are dominated by the Gaussian jitter rather than the exponential emission, and hence the initial primary photoelectron with the minimum-variance arrival time is not always the first primary photoelectron. The curves are labeled with the mean number of primary photoelectrons, and were created with the same system conditions as in Figs. 1 and 2.

Fig. 4. Shot noise degradation—This graph depicts the PDFs of the third order statistic for 770, 800, or 830 primary photoelectrons (solid). The third order statistic best estimated the gamma-ray's arrival time for 800 primary photoelectrons. Assuming the shot noise is Gaussian with mean 800 and sigma $\sqrt{800}$, the degradation can be estimated by another Gaussian (dashed), though this estimation was not used in the model, only in this figure for visual comparison. Section III-C contains an explanation for how shot noise was included in the simulations.

photoelectrons and a specific SPAD jitter was found, shot noise was added as follows. The distribution of the observed number of primary photoelectrons was now assumed to be a Gaussian distribution with a mean of n and a sigma of $\sqrt{n}$. This normal distribution was sampled to discretize it, and the kth order statistic's PDF (the same order statistic rank was used) was found for each sample location within three standard deviations of n. These PDFs were then weighted by the probability of observing that specific number of primary photoelectrons, and then the weighted PDFs were summed to achieve a final PDF. The FWHM of the weighted-sum PDF was used for the error in the gamma-ray's arrival time for that particular SPAD jitter and expected number of primary photoelectrons. Shot noise's inclusion changed the FWHM by less than 5% in all of the modeled cases. Fig. 4 shows the effect of shot noise for one simulated case.

Repeating these calculations for a range of SPAD jitters and light detection efficiencies allows the creation of a contour plot which depicts the tradeoff between the SPAD jitter and light detection efficiency. As both SiPMs and SPAD TDC arrays are large arrays of SPADs, the same model should be valid for both types of arrays, and the location in the contour plot should estimate the error in the estimate of the gamma-ray's arrival time. The number of detected, primary photoelectrons was assumed to range between 50 and 6,500, corresponding to the number of primary photoelectrons for a low fill factor SPAD TDC array and a high fill factor SiPM. Fig. 5 shows such a contour plot for LYSO, displaying the tradeoff between different SPAD jitters and light detection efficiency. The same plot for LaBr_3 scintillators is shown in Fig. 6, and for an imaginary, fast scintillator in Fig. 8. The rank of the optimal order statistic for LaBr_3 is shown in Fig. 7

The final concerns for the model are the dead time, crosstalk and after-pulsing of SPADs. Dead time refers to the dead time

Fig. 5. Gamma-ray FWHM timing resolution (ps) for LYSO—Contours of the FWHM timing resolution for one gamma-ray detector are shown in picoseconds. The x axis depicts the mean number of primary photoelectrons, corresponding to the light detection efficiency, with the y axis depicting the SPAD jitter. As discussed in Section III-D, dashed curves use only order statistics with a rank of 10 or lower to provide a confidence level that dead time or small amounts of crosstalk would not change the results.

of an individual SPAD (sometimes called a micro-cell), not the dead time of the entire array. Crosstalk and after-pulsing are covered later, in Section III-D.

Modeling micro-cell dead time is complex, as another photon incident onto a SPAD's active region during an avalanche can change the timing characteristics of the avalanche's properties [22]. Instead of a full model of dead time, results are presented for when the dead time distorts arrival estimates of fewer than 5% of gamma-rays. As dead times are much larger than the time differences between the initial scintillation-photon arrivals, quantifying the probability of missing a primary photoelectron will relate to the probability that a
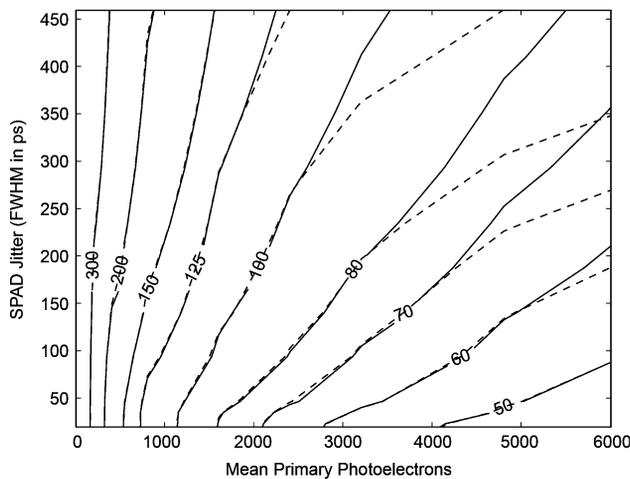
Fig. 6. Gamma-ray FWHM timing resolution (ps) for $LaBr_3$—Contours of the FWHM timing resolution for a single detector are shown in picoseconds. See the caption of Fig. 5 for a longer discussion of the figure. The optimal primary photoelectron's rank is shown in Fig. 7.

Fig. 8. Gamma-ray FWHM timing resolution (ps) for an imaginary scintillator—Contours of the FWHM timing resolution are shown in ps units. The imaginary scintillator was assumed to have a 10 ns decay time and a 0.050 ns rise time. See the caption of Fig. 5 for a longer discussion of the figure.

Fig. 7. Rank of optimal primary photoelectron for $LaBr_3$—Contours of the rank of the primary photoelectron with lowest variance are shown for $LaBr_3$. Dead time and crosstalk effects are ignored in this plot; see Fig. 6 for a discussion of those effects.

primary photoelectron is incident on a SPAD that has already fired. The probability of having two photons incident on the same SPAD is analogous to the probability of two people having the same birthday in a group of people, which is the general birthday problem [23]. For a 1000 SPAD array and 95% probability that a photon does not impinge on a fired SPAD, there must be 10 or fewer avalanches before dead time is expected to distort the results of more than 5% of the gamma-ray's arrival estimates. Figs. 6 and 8 contain dashed lines showing the effects from estimating the gamma-ray arrival time using only the first 10 scintillation photon arrival times. The contour plot for systems with LYSO in Fig. 5 also contains these dashed lines.

An additional effect can distort the signal from an SiPMs when the dead time is too small. If a SPAD in an SiPM has a small dead time, the SPAD will begin to recharge very quickly and the SPAD's current will drop rapidly after an avalanche. If

the timing of the scintillation photons' arrivals are based on the current into the array, such as in an SiPM, the drop in current can cause incorrect labeling of the rank of primary photoelectrons. As described in the next section, the timing estimates that are critical to estimating the gamma-ray's arrival time occur in the first few nanoseconds, meaning that dead times that are tens of nanoseconds should be sufficient to avoid this effect.

D. Effects of False Events

This section describes the effects of events that are not caused by scintillation photons, but rather are caused by other sources, such as noise, after-pulsing and crosstalk.

To insulate from the effects of noise, a trigger criteria needs to be examined for SPAD TDC arrays and SiPMs. We propose the following for a SPAD TDC array: all events in a small time window are stored in a buffer, and whenever more than a set number of events, say 5, occurs within short time window, say 10 ns, then a gamma-ray is assumed to have arrived. The device then compares the time difference between previous events, and the event that is at least 4 ns from the previous event is assumed to be the first primary photoelectron. Similar techniques have been used with SiPMs to insulate the detectors from the effects of noise [24].

To check this trigger criteria is appropriate, the event rate from scintillation photons must be high enough to trigger the threshold and the noise rate must be low enough that a false event is unlikely to occur starting a few nanoseconds before the first primary photoelectron's arrival and ending with the arrival of the optimal primary photoelectron for the gamma-ray arrival time estimation. The justification that the event rate is high enough is clear in Fig. 2, as the slowest, dimmest scintillator with the worst optical coupling has a very high probability of having at least five events in the 10 ns following detection of the first primary photoelectron. The justification that the noise rate is low enough is more complex. Most modern SiPM and SPAD TDC arrays have noise numbers in the low megahertz range. For a 99% probability that no noise events occur for 5 ns before and 5 ns after a gamma-ray's arrival, the noise rate λ must

satisfy $\exp(-\lambda \cdot 10 \text{ ns}) > 0.99$. A noise rate of roughly 1 MHz satisfies this relation.

When the energy resolution was above 20%, the time between the primary photoelectron observation time used in the gamma-ray arrival time estimation and the optimal primary photoelectron always had a 95th percentile below 4 ns, justifying the 5 ns threshold. Fig. 2 depicts examples of arrival times for LYSO, though this criteria was checked for all scintillators. When the noise rate is increased to 5 MHz, the probability drops from 99% to 95% that noise does not effect the gamma-ray arrival time estimate. The effected percentage can be lowered with smaller time windows around the first primary photoelectron, which is possible with better light detection efficiencies.

In Section II, after-pulsing and crosstalk were ignored without justification. Since the initial photons arrive within a few nanoseconds of one another, and after-pulsing occurs after the dead time, after-pulsing will have an insignificant effect on the timing model. After-pulsing will effect the energy resolution if the total number of arrivals is used instead of the photopeak, and this distortion would require compensation.

The last noise factor, crosstalk, is not a problem for SPAD TDC arrays. This is ostensibly due to the large distance between the SPADs themselves in such arrays. In SiPMs, however, the close spacing of the diodes can produce crosstalk. The effects of crosstalk on initial event timing can be modeled in a fashion similar to the dead time, with a probability that the gamma-ray arrival time estimate is not effected when only order statistics of a certain rank or below are considered in estimates of the gamma-ray's arrival time. For low numbers of fired SPADs, the probability of crosstalk will increase in a roughly linear fashion with the number of fired SPADs. Assuming a 0.5% chance of crosstalk per avalanche, a ten-event threshold will produce crosstalk in less than 5% of cases. Thus, the same ten event threshold used for the effects of the dead time can also be used for the effects of small amounts of crosstalk; the rank of the primary photoelectron used in the estimation must be between 1 and 10, with the degradation as shown with the dashed lines in Figs. 5, 6, and 8.

IV. RESULTS

As the contour plots in Figs. 5, 6, and 8 show, SPAD jitter becomes increasingly important for extremely fast scintillators coupled to efficient detectors. In these figures, vertical contours correspond to the y axis variable having no effect on the timing resolution, while horizontal contours display that the x axis variable having no effect.

Two predictions merit explanation: the initial primary photoelectron does not always have the lowest arrival time jitter, and the gamma-ray's arrival time jitter can be lower than the SPAD jitter.

It has been experimentally shown that later primary photoelectrons may have a lower variance than the initial primary photoelectron [24]. Imagine, as a thought experiment, that the emitted light was a normal distribution, rather than an exponential one. In this thought experiment, the lowest variance primary photoelectrons will not be the initial observed events nor the last observed events, but rather the events observed in the middle [19]. Also, as the number of samples increases, the mean of the distribution can be more accurately estimated than the variance, as the distribution is oversampled. This explains the cause of having a lower gamma-ray arrival time jitter than SPAD jitter.

In reality, as the number of primary photoelectrons increases, the Gaussian distribution of the jitter begins to have a larger effect on the variances of the initial primary photoelectrons, explaining the increase in rank of the optimal primary photoelectron for estimating the gamma-ray's arrival time.

Fig. 7 shows the rank of the optimal order statistic for $LaBr_3$. The optimal rank is seen to stay low when either the light detection efficiency is very poor or the SPAD jitter is very good. The optimal rank increases for increasing SPAD jitter, and does so more quickly for systems with good light detection efficiency. High optimal ranks coupled with the effects of crosstalk could deteriorate the estimation of the gamma-ray's arrival time, as a higher rank implies more avalanches, increasing the likelihood of crosstalk. As described in Section III-D, even small crosstalk probabilities of 0.5% begin to introduce problems when the optimal rank increases above 10. Decreasing crosstalk can be accomplished with additional optical isolation between SPADs in an SiPM, and the model suggests more isolation should increase TOF PET performance of SiPMs with high SPAD jitters and high crosstalk probabilities. This effect merits further study.

In Fig. 5, which shows the results for LYSO, the SPAD jitter has a smaller effect on the timing resolution than the light detection efficiency. In the left-hand side of the figure, corresponding to relatively few photons being detected, the contours are nearly vertical, implying that the timing resolution is limited by the number of detected photons. When detecting more than 1000 photons, SPAD jitter begins to have a small effect, but marginal improvements in photon detection can compensate for sizable increases in SPAD jitter. For LYSO-based detectors in TOF PET, it appears that the main constraint on timing resolution is the detector's ability to capture light efficiently.

For an imaginary, fast scintillator with a 50 ps rise time and 10 ns decay time, Fig. 8 shows that the error in the gamma-ray's arrival time can depend on both the SPAD jitter and the light detection efficiency. While the light detection efficiency can still be a bottleneck, for systems with a high SPAD jitter it is less of an issue as the number of samples increases. For configurations that capture more light, halving the number of samples and halving the SPAD jitter would have little effect on the estimation error. In this case, marginal improvements in SPAD jitter and light detection efficiency would have the same effect. The dashed lines in Fig. 8 show the effect of dead time or crosstalk in this model, and this effect could be significant. Development of faster, bright scintillators would merit further study of this effect.

The tradeoff for $LaBr_3$ is shown in Fig. 6, and the results lie between the other two contour plots. As with the other two plots, there is a sample-constrained region in the left-hand side of the plot. In the right hand-side of the plot, doubling the number of detected photons and increasing the SPAD jitter by a factor of four would keep the same arrival time estimation error, so while the SPAD jitter plays an increasingly important

role, marginal improvements in the light detection efficiency would produce better timing resolutions than improvements in the SPAD jitter.

The contour plots can be compared to values reported in the literature [10], [24]–[28], though the number of detected photons and SPAD jitters often need to be estimated based on the fill factors, PDPs and optical coupling efficiencies.

In general, all plots have a sample-constrained region in which the SPAD jitter has little to no effect on the timing resolution. In these cases the single exponential model is a good one, and the variance from sampling the single exponential dominates all other factors. As the decay time and rise time of the scintillator decrease, the SPAD jitter plays an increasingly important role. Thus, as faster, brighter scintillators are developed, the detector jitter versus light detection efficiency tradeoff will play an increasingly important role in estimating the gamma-ray's arrival time.

A linear estimation based on multiple scintillator photon arrival times is an improvement which could reduce the gamma-ray arrival time estimation error. From the simplified model, the gamma-ray arrival time estimation error FWHM should follow a $1/n$ pattern, however the data deviates from this pattern for efficient detectors. Depending on the uncorrelated component in the jitter of different scintillation photon observations, a linear estimator might reduce the estimation error of the gamma-ray's arrival time, but at the cost of increased system complexity.

For a microlens-recovered SPAD TDC array, the fill factor varies from over 30% for light at orthogonal angles to nearly 0% for light incident at high angles [6]. Assuming a 120 ps jitter, a 15% average fill factor, a 40% PDP and the same optical coupling efficiency as SiPMs, the model implies a SPAD TDC array coupled to a LYSO crystal would have a worse timing error than a LYSO crystal and SiPM detector with a 60% fill factor, 40% PDP and 340 ps detector jitter. For $LaBr_3$ and perfect optical coupling efficiency, the SiPM would still yield better results. The model predicts that microlenses would need to recover about 30% more fill factor for a $LaBr_3$-SPAD TDC array combination to be competitive with a $LaBr_3$-SiPM combination. For a possible scintillator as bright as $LaBr_3$ but with a faster decay, the SPAD TDC might produce better estimation errors; however, a range of factors such as linear estimators, crosstalk and geometry could change the expected results.

V. CONCLUSION

While SPAD TDC arrays utilizing microlens-recovery can have high fill factors, models suggest that the current generation of SPAD TDC arrays detect too few photons to compete with SiPMs in TOF PET systems utilizing LYSO. Adding microlens recovery to SiPMs should further widen the performance gap. For faster scintillators, such as $LaBr_3$, the recovery in fill factor from using microlenses appears to be too low to warrant using SPAD TDC arrays in place of SiPMs with a low crosstalk probability, though a SPAD TDC array may be competitive with an SiPM that has a high probability of crosstalk. However, as the constraints of SPAD TDC arrays change, including microlens

technology advancing, TDC jitter improving, and faster scintillators developing, SPAD TDC arrays may become competitive with SiPMs in TOF PET applications.

REFERENCES

[1] R. Hawkes, A. Lucas, J. Stevick, G. Llosa, S. Marcatili, C. Piemonte, A. Del Guerra, and T. Carpenter, "Silicon photomultiplier performance tests in magnetic resonance pulsed fields," in *IEEE Nuclear Science Symp. Conf. Rec.*, Nov. 3, 2007, vol. 5, pp. 3400–3403.

[2] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. Henderson, "A 32 × 32 50 ps resolution 10 bit time to digital converter array in 130 nm CMOS for time correlated imaging," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 2009, pp. 77–80.

[3] M. Gersbach, Y. Maruyama, E. Labonne, J. Richardson, R. Walker, L. Grant, R. Henderson, F. Borghetti, D. Stoppa, and E. Charbon, "A parallel 32 × 32 time-to-digital converter array fabricated in a 130 nm imaging CMOS technology," in *Proc. Eur. Solid-State Circuits Conf. (ESSCIRC)*, 2009, pp. 196–199.

[4] D. Stoppa, F. Borghetti, J. Richardson, R. Walker, L. Grant, R. Henderson, M. Gersbach, and E. Charbon, "A 32 × 32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain," in *Proc. Eur. Solid-State Circuits Conference (ESSCIRC)*, 2009, pp. 204–207.

[5] A. Spinelli and A. Lacaita, "Physics and numerical simulation of single photon avalanche diodes," *IEEE Trans. Electron Devices*, vol. 44, pp. 1931–1943, Nov. 1997.

[6] S. Donati, G. Martini, and M. Norgia, "Microconcentrators to recover fill-factor in image photodetectors with pixel on-boardprocessing circuits," *Opt. Exp.*, vol. 15, no. 26, pp. 18066–18075, 2007.

[7] Million Frame Per Second, Time-Correlated Single Photon Camera [IST FP6 FET Open] [Online]. Available: http://www.megaframe.eu

[8] A. Rochas, ""Single-photon avalanche diodes in CMOS technology," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2003.

[9] C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, "A 128 × 128 single-photon image sensor with column-level 10-bit time-to-digital converter array," *IEEE J. Solid-State Circuits*, vol. 43, pp. 2977–2989, Dec. 2008.

[10] C. L. Kim, G.-C. Wang, and S. Dolinsky, "Multi-pixel photon counters for TOF PET detector and its challenges," in *IEEE Nuclear Science Symp. Conf. Rec.*, Oct. 2008, pp. 3586–3590.

[11] MPPC Catalog Hamamatsu Photonics, 2008.

[12] G. Llosa, R. Battiston, N. Belcari, M. Boscardin, G. Collazuol, F. Corsi, G.-F. D. Betta, A. Del Guerra, N. Dinu, G. Levi, S. Marcatili, S. Moehrs, C. Marzocca, C. Piemonte, and A. Pozza, "Novel silicon photomultipliers for PET applications," *IEEE Trans. Nucl. Sci.*, vol. 55, pp. 877–881, Jun. 2008.

[13] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, "Toward a 3-D camera based on single photon avalanche diodes," *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, pp. 796–802, Jul./Aug. 2004.

[14] M. Kapusta, M. Moszynski, M. Balcerzyk, J. Braziewicz, D. Wolski, J. Pawelke, and W. Klamra, "Comparison of the scintillation properties of LSO:Ce manufactured by different laboratories and of LGSO:Ce," *IEEE Trans. Nucl. Sci.*, vol. 47, pp. 1341–1345, Aug. 2000.

[15] R. F. Post and L. I. Schiff, "Statistical limitations on the resolving time of a scintillation counter," *Phys. Rev.*, vol. 80, no. 6, p. 1113, Dec. 1950.

[16] J. Glodo, W. Moses, W. Higgins, E. van Loef, P. Wong, S. Derenzo, M. Weber, and K. Shah, "Effects of Ce concentration on scintillation properties of $LaBr_3 : Ce$," *IEEE Trans. Nucl. Sci.*, vol. 52, pp. 1805–1808, Oct. 2005.

[17] Y. Shao, "A new timing model for calculating the intrinsic timing resolution of a scintillator detector," *Phys. Med. Biol.*, vol. 52, no. 4, p. 1103, 2007.

[18] W. R. Leo, *Techniques for Nuclear and Particle Physics Experiments*. New York: Springer, 1994.

[19] B. Arnold, *A First Course in Order Statistics*. New York: Wiley-Interscience, 1992.

[20] SciPy, [Online]. Available: http://www.scipy.org

[21] Python Programming Language Official Website [Online]. Available: http://www.python.org

[22] A. Lacaita, S. Cova, A. Spinelli, and F. Zappa, "Photon-assisted avalanche spreading in reach-through photodiodes," *Appl. Phys. Lett.*, vol. 62, no. 6, pp. 606–608, 1993.

[23] E. H. Mckinney, "Generalized birthday problem," *Amer. Math. Monthly*, vol. 73, no. 4, pp. 385–387, 1966.

[24] D. R. Schaart, S. Seifert, R. Vinke, H. T. van Dam, P. Dendooven, H. Löhner, and F. J. Beekman, "$LaBr_3 : Ce$ and SiPMs for time-of-flight PET: Achieving 100 ps coincidence resolving time," *Phys. Med. Biol.*, vol. 55, no. 7, p. N179, 2010.

[25] D. R. Schaart, H. T. van Dam, S. Seifert, R. Vinke, P. Dendooven, H. Lohner, and F. J. Beekman, "A novel, SiPM-array-based, monolithic scintillator detector for PET," *Phys. Med. Biol.*, vol. 54, no. 11, pp. 3501–3512, 2009.

[26] P. Buzhan, B. Dolgoshein, L. Filatov, A. Ilyin, V. Kaplin, A. Karakash, S. Klemin, R. Mirzoyan, A. Otte, E. Popova, V. Sosnovtsev, and M. Teshima, "Large area silicon photomultipliers: Performance and applications," *Nucl. Instrum. Methods Phys. Res. A*, vol. 567, no. 1, pp. 78–82, 2006.

[27] G. Llosa, N. Belcari, M. Bisogni, G. Collazuol, S. Marcatili, S. Moehrs, F. Morsani, C. Piemonte, and A. Del Guerra, "Energy and timing resolution studies with silicon photomultipliers (SiPMs) and 4-pixel SiPM matrices for PET," *IEEE Trans. Nucl. Sci.*, vol. 56, pp. 543–548, Jun. 2009.

[28] A. Nassalski, M. Moszynski, A. Syntfeld-Kazuch, T. Szczesniak, L. Swiderski, D. Wolski, T. Batsch, and J. Baszak, "Silicon photomultiplier as an alternative for APD in PET/MRI applications," in *IEEE 2008 Nuclear Science Symp. Conf. Rec.*, Oct. 2008, pp. 1620–1625.