Dead Time Compensation for High-Flux Ranging

Joshua Rapp, *Student Member, IEEE,* Yanting Ma, *Member, IEEE,* Robin M. A. Dawson, *Member, IEEE,*
and Vivek K Goyal, *Fellow, IEEE*

Abstract—Dead time effects have been considered a major limitation for fast data acquisition in various time-correlated single photon counting applications, since a commonly adopted approach for dead time mitigation is to operate in the low-flux regime where dead time effects can be ignored. Through the application of lidar ranging, this work explores the empirical distribution of detection times in the presence of dead time and demonstrates that an accurate statistical model can result in reduced ranging error with shorter data acquisition time when operating in the high-flux regime. Specifically, we show that the empirical distribution of detection times converges to the stationary distribution of a Markov chain. Depth estimation can then be performed by passing the empirical distribution through a filter matched to the stationary distribution. Moreover, based on the Markov chain model, we formulate the recovery of arrival distribution from detection distribution as a nonlinear inverse problem and solve it via provably convergent mathematical optimization. By comparing per-detection Fisher information for depth estimation from high- and low-flux detection time distributions, we provide an analytical basis for possible improvement of ranging performance resulting from the presence of dead time. Finally, we demonstrate the effectiveness of our formulation and algorithm via simulations of lidar ranging.

Index Terms—Dead time, high-flux ranging, lidar, Markov chain, nonlinear inverse, single-photon detection.

I. INTRODUCTION

Time-correlated single photon counting (TCSPC) is a powerful technique for measuring the fast, time-dependent responses of actively illuminated systems. Commonly used for fluorescence lifetime imaging (FLIM) [1], TCSPC has also been applied to optical quantum information applications [2], light detection and ranging (lidar) [3], and non-line-of-sight (NLOS) imaging [4], [5], among others [6]. TCSPC is particularly useful for lidar because the single-photon sensitivity allows for lower-intensity signal returns, either from weaker illuminations or from distant, oblique-angled, or otherwise non-cooperative targets [7]. As a result, TCSPC lidar is one of the competing technologies currently being developed for commercial autonomous vehicles.

One of the main hardware limitations of photon counting is that the instrumentation has a dead time, a period of insensitivity after a photon detection during which arriving photons cannot be registered. Typical TCSPC applications

use a laser repeatedly pulsed with period t_r and build up a histogram of detection times relative to the most recent illumination. Without compensation for the effects of dead time, detection time histograms appear as distorted versions of the incident light intensity waveform, leading to erroneous depth estimates. The usual approach to dealing with dead time is to limit the acquisition flux so that photons are detected in at most 5% of illumination periods; this reduces the probability of photons arriving during the dead time, thus limiting the number of arrivals that are "lost." However, attenuating the flux incident on a detector is an inefficient use of the light reflected from a scene and slows down the acquisition of sufficient photons for accurate ranging. Allowing for higher incident flux and compensating for the resulting distortions due to dead time would enable faster acquisition without loss of accuracy and has been the subject of several recent works [8], [9], although their models of dead time assume a single detection can be recorded per illumination period, an assumption which does not necessarily hold for modern timing electronics. Our aim is to accurately model the effects of dead time on the photon detection process so that photons can be detected at a much higher rate and distortions introduced due to dead time can be predicted and corrected. Eventually, this approach should lead to the possibility of higher laser powers, shorter acquisition time, and more accurate depth estimation.

A. Dead Time Characterization

The source of the dead time and resulting behavior of the system may vary greatly depending on the implementation. Our work studies dead time correction for modern TCSPC systems with asynchronous electronics (such as the HydraHarp 400 [10] or TimeHarp 260 [11] from PicoQuant) and a nonparalyzable detector (e.g., a PDM-series [12] or Fast-gated [13] Single Photon Avalanche Diode (SPAD) from Micro Photon Devices, operated in free-running mode). In the following, we formally define paralyzability, the sources of the dead time, and synchronization so as to clarify the dead time model we consider.

Paralyzability. The dead times of event-counting detectors have been studied since at least the 1940s [14]–[17], with Feller first classifying detectors in terms of their paralyzability [18]. *Nonparalyzable (Type I)* detectors are dead for a fixed time t_d after a detection, regardless of whether additional photons arrive during the dead time. On the other hand, when photons arrive during the dead time of a *paralyzable (Type II)* detector, the dead time restarts and extends for at least another t_d. We consider only nonparalyzable detectors in our work.

Source. TCSPC systems suffer dead times from various components, each with different behavior. The timing electronics in *classical* TCSPC systems (in both nonreversed and

reversed start-stop modes) are only able to record a single detection time per illumination period [1]. *Modern* TCSPC electronics allow for multiple detections per period, but the duration of the signal recording process still forces the timing electronics to be insensitive to additional photon arrivals [19]. In addition, the detectors themselves suffer from dead times. For instance, the commonly used SPAD detectors are reverse-biased photodiodes that are single-photon sensitive because they are operated above the breakdown voltage. Incident photons cause an avalanche of carriers that is directly detectable as a precise digital signal, but which must be quenched in order to reset the detector to a photon-sensitive state [1], so the SPAD is dead during the quenching. The time during which the SPAD is "held-off" is further extended to prevent afterpulses, which are avalanches caused by the release of trapped carriers from previous avalanches [20]. The total hold-off time for active quenching circuits is thus a trade-off between shorter dead times versus lower afterpulsing probabilities; this parameter is often fixed in the detection circuitry design but may be left variable in some devices for tuning at the user's discretion (e.g., [13]). In this work, we assume that the hold-off time is sufficiently long such that any afterpulses can be considered indistinguishable from ambient detections and dark counts.[1]

Synchronization. The different sources of dead time further suggest two modes of TCSPC system operation. We call systems *synchronous* if they ensure that the end of a dead time is synchronized with the start of an illumination period. Synchronous operation is often built into the hardware, such as in classical TCSPC systems or in the gated mode of fast-gated SPADs [9], [13]. In reversed start-stop mode, classical timing electronics may become active in the middle of an illumination period, but that recovery time is consistent, as the dead time is synchronized to the stop signal (either a delayed version of the current pulse or the next pulse). Modern TCSPC electronics enable *asynchronous* operation, in which there is no enforced synchronization between the dead time and the timing electronics. If a photon is detected towards the end of a cycle and the dead time continues after the next laser excitation, there is no mechanism preventing the detector from becoming active in the middle of that cycle. In other words, the end of the dead time is no longer dependent on the cycle synchronization, but on the arrival time of the most recently detected photon. The synchronous and asynchronous architectures correspond to the "clock-driven" and "event-driven" SPAD recharge mechanisms, which were explored in [22], [23]. While most existing work on the effect of dead time assumes synchronous operation, we consider only asynchronous systems in this work.

B. Dead Time Compensation Approaches

Yu and Fessler outline a number of general strategies for handling the effects of dead time [24], with the simplest approach being to simply ignore the dead time. Most commonly, the total photon flux at the detector is changed such that the

dead time effects are actually negligible and can be ignored. Since the effect of dead time is that photon arrivals within t_d of a detection are missed, a straightforward approach is to reduce the total photon flux, either by lowering the laser power and ambient light if possible, or by attenuating with a filter at the detector. The suggestion of O'Connor and Phillips is to keep the fraction of excitations causing a detection to be at most 5% to avoid dead time effects [25], a recommendation that electronics manufacturers have adopted. Reducing the flux inevitably leads to longer time needed to acquire the same number of photons. As a result, several recent works have focused on reducing the number of photons per pixel needed for accurate depth imaging by incorporating parameterized probability models of detection times and priors on the spatial structure of natural scenes [26]–[29]. Other approaches have tried to ignore dead time by changing the hardware setup, such as using multiple detectors so that there is more likely to be a detector not in the reset state when a photon arrives [1].

Rather than attenuate the flux at the detector to avoid dead time effects, another strategy is to correct the distortions in the high-flux data after acquisition. Most algorithm-based attempts at dead time compensation consider synchronous systems due not only to the systems that have historically been available, but also for the convenient property that detection times are statistically independent of each other in different cycles [27]. One of the first methods for dealing with dead time in synchronous systems is that of Coates [30]. Coates's basic algorithm was designed for lifetime measurement, with later work adapting the algorithm to include background subtraction [31]. The basic principle of Coates's algorithm is that for any bin i in a histogram, the detections in the preceding bins spanning t_d represent excitation cycles when no photon could have been detected in bin i because the detector was dead. The number of cycles in which the detector thus must have been dead is used to adjust computation of the photon arrival probability in each bin. Recent work has rederived Coates's expression, which is the ML estimator for the number of photon arrivals in each bin of a histogram in a synchronous system, in order to include priors for maximum a posteriori (MAP) estimation [9]. A few models [32], [33] consider histogram corrections for a hybrid of synchronous and asynchronous systems, which do allow for multiple statistically dependent detections per illumination cycle but without the dependency carrying over into different cycles. A handful of papers address special cases of dead time effects in asynchronous systems: Antolovic et al. consider detection rate estimation for homogeneous arrival processes in [22], [23], whereas Cominelli et al. explore the special case when t_d equals an integer multiple of t_r and no correction is needed [34]. However, these approaches are insufficient to address the typical lidar acquisition mode with inhomogenous arrivals and in which t_d and t_r cannot necessarily be arbitrarily adjusted. The only work the authors are aware of that addresses asynchronous systems generally is that of Isbaner et al. [35], which effectively treats the detection process as a time-dependent attenuation of the arrival process intensity. Although they model both the electronics and detector dead times t_e and t_d, respectively, they note that such a system simplifies to having only one source of dead

---

[1] If the quenching time is too short, afterpulses can no longer be considered an independent Poisson process, as their occurrence is correlated with the previous detection time [21].

time when $t_e < t_d$. They use an iterative procedure to estimate the attenuation, which is then used to correct the detection histogram.

The last strategy for dealing with dead time is to use the data as acquired but to incorporate dead time into the detection model. In this vein, Heide et al. adjust their parameter estimation procedure to include dead time effects [8]. However, the synchronous system assumption they use is technically only valid for their asynchronous timing electronics (PicoQuant PicoHarp 300, [36]) if zero ambient light is present, which guarantees that the detector will be reset for the next signal pulse.

C. Other Related Work

In addition to missing photons that arrive during a dead time, events can fail to be registered in the recording of point processes through other forms of "counting loss," which depends on the measurement system design [1]. A well-documented form of counting loss is "pile-up," referring to the problem of the rising edge of a pulse overlapping with the tail of a previous pulse, such that the later pulse is not registered by a discriminator. Each piled-up pulse prolongs the duration in which new events cannot be detected, making the discriminator a Type II detector. Pile-up is present in some TCSPC system designs, such as those using passively-recharged SPADs [23] or hybrid photodetectors with negligible dead time [37]. Several approaches correcting for pile-up have recently been proposed for nuclear spectrometry, in which both the pulse times and energies are of interest [38]–[40]. Confusingly, the term "pile-up" is also sometimes used to describe the effect of dead time in synchronous TCSPC systems (e.g., [8], [9]) because the effect of earlier detections preventing later detections is similar. Another form of counting loss, named Type III in [24], may also occur in some systems (e.g., [41]) when two pulses occur close together and neither one gets recorded.

In lidar applications, the dead time-affected acquisition results in closer apparent distances, which has a similar effect to the intensity-dependent change in perceived depth known as "range walk error" [42]–[44]. Range walk is the result of using a discriminator to trigger in the leading edge of a signal pulse; a stronger signal with a steeper rising edge will be detected earlier than a weaker signal with a smaller slope. Due to the similarity with high-flux ranging, approaches correcting for range walk error could be adapted to compensate for dead time. Several optics-based methods aimed at range walk error correction attempt to experimentally measure and then correct for the bias in depth estimation. He et al. first calibrate the amount of range walk incurred as a function of the detection rate [45]. Then the conventional depth estimation procedure is performed with the dead time-distorted data, and the range walk bias is subtracted off to correct the depth estimate. Ye et al. use a similar method, except they first split the incident light with a 90:10 beamsplitter to two SPADs, using the lower-flux channel for simultaneous bias estimation to subtract off from the lower-variance high-flux estimate [46].

D. Main Contributions

1) Markov Chain Detection Time Model: We rigorously construct a Markov chain model to characterize the empirical distribution of detection times in asynchronous TCSPC systems. Analyzing the stationary distribution of the Markov chain directly leads to a simple log-matched filter estimator for depth estimation.

2) Arrival Intensity Reconstruction: We derive a nonlinear inverse formulation for arrival intensity estimation from the detection distribution, where the formulation is based on the stationary condition of the Markov chain and the nonlinear inverse problem is solved by a provably convergent optimization algorithm; the estimated arrival intensity can then be used for depth estimation and other tasks.

3) Accurate High-Flux Ranging: Using our Markov chain-based methods, we show that depth estimation from high-flux detection data can achieve lower error than using low-flux data for the same acquisition time or can alternatively achieve the same error from much faster acquisitions. Furthermore, our methods outperform the method of [35] applied to high-flux detection data.

4) Demonstration of Dead Time Benefits: By comparing the Fisher information per detected photon for depth estimation from the high- and low-flux detection time distributions, we show that when the background rate is low and the signal rate is sufficiently high, the presence of dead time may lead to improvement in ranging accuracy for a fixed number of detections. Moreover, when the dead time t_d is only slightly smaller than the illumination period t_r, such improvement can extend to higher background rate scenarios, since the dead time acts as a signal-triggered gate in this case.

II. EMPIRICAL DISTRIBUTION OF DETECTION TIMES

The challenge of studying the detection time distribution for the asynchronous dead time model is that the detection times are statistically dependent. In this section, we show that the dependency is Markovian and provide the explicit transition probability density function (PDF). From the transition PDF, we can analyze the stationary condition and obtain the stationary distribution, from which our high-flux ranging algorithms are derived.

A. Photon Arrival Process

It is well known that photon arrival times at a detector are described by a Poisson process [47]. For TCSPC, the repeated illumination with period t_r makes the arrival process an inhomogeneous Poisson process with periodic intensity function $\lambda(t)$. In general, $\lambda(t)$ is composed of two parts:

$$\lambda(t) = \lambda_s(t) + \lambda_b(t), \quad (1)$$

where $\lambda_s(t)$ is the time-varying intensity of a *signal* process and $\lambda_b(t)$ is the intensity due to *background* (ambient light), which is assumed to be constant λ_b in this work. For the application of ranging, $\lambda_s(t)$ is often described parametrically in one period as the scaled and time-shifted illumination pulse $\lambda_s(t) = \alpha\beta s(t - 2z/c)$, where α is the target reflectivity, z is

the target depth, β is a gain factor corresponding to the illumination power, $s(t) = \exp(-t^2/2\sigma^2)/\int_0^{t_r} \exp(-\tau^2/2\sigma^2)d\tau$ is the Gaussian pulse shape approximation with half pulse width σ, and c is the speed of light. Within one period, the signal photon arrival rate S and background photon arrival rate B are defined as $S := \int_0^{t_r} \lambda_s(\tau)\,d\tau$ and $B := \lambda_b t_r$, respectively. The total flux is given by $\Lambda := S + B$, and the signal to background ratio is defined as $\mathrm{SBR} := S/B$.

B. Markov Chain Model for Detection Times

If there were no dead time effects, the detection process would be equivalent to the arrival process, which is Poisson with intensity $\lambda(t)$. Conditioned on the total number of detections, the absolute detection times[2] would be order statistics of i.i.d. random variables with common probability density function $\propto \lambda(t)$ [47, Section 2.3.3]. However, in the presence of dead time effects, the detection process is no longer Poisson, since the detection intensity, denoted by $\mu(t)$, is now a random process depending on the history of the detection process; such a detection process is referred to as a self-exciting process [47]. Specifically, let $\{N(t) : t \geq 0\}$ denote the detection process with (random) intensity $\mu(t)$, where $\{N(t) : t \geq 0\}$ is characterized by the number of detections $N(t)$ at time t and a sequence of absolute detection times $T_1, \ldots, T_{N(t)}$. The conditional PDF of T_{i+1} given $T_1, \ldots, T_i$ is [47, (6.15) (6.16)]

$$f_{T_{i+1}|T_1,\ldots,T_i}(t|t_1,\ldots,t_i) = \mu(t)\exp\left(-\int_{t_i}^t \mu(\tau)\,d\tau\right). \quad (2)$$

For a general self-exciting process, $\mu(t)$ can depend on the entire history of the process $\{N(\tau) : 0 \leq \tau < t\}$. For the specific detection process considered in this work, we have

$$\mu(t) = \begin{cases} \lambda(t), & \text{if } t > T_{N(t)} + t_d; \\ 0, & \text{if } T_{N(t)} < t \leq T_{N(t)} + t_d, \end{cases} \quad (3)$$

where we introduce the notation $T_0 := -\infty$. We can see that $\mu(t)$ only depends on the latest detection time. Therefore, for the $\mu(t)$ defined in (3), the right-hand side (RHS) of (2) depends on t_i but not on $t_1, \ldots, t_{i-1}$. That is, the absolute detection times form a Markov chain with transition PDF

$$f_{T_{i+1}|T_i}(t|t_i) = \lambda(t)\exp\left(-\int_{t_i+t_d}^t \lambda(\tau)\,d\tau\right)\mathbb{I}\{t > t_i+t_d\}, \quad (4)$$

where $\mathbb{I}$ is the indicator function. An illustration of a realization of the detection process is shown in Fig. 1.

Define two sequences of random variables, $\{K_i\}_{i\in\mathbb{N}}$ and $\{X_i\}_{i\in\mathbb{N}}$, such that $K_i := \lfloor T_i/t_r \rfloor$, where $\lfloor a \rfloor$ is the integer part of $a \in \mathbb{R}$, and $X_i := T_i \bmod t_r$, hence $T_i = K_i t_r + X_i$. That is, K_i is the number of illumination periods before T_i and X_i is the location of absolute detection time T_i within illumination period $K_i + 1$, which is referred to as detection time in this paper. Note that if there were no dead time effects,

[2]Throughout the paper, *absolute detection time* refers to the time when periodicity is not taken into consideration (i.e., with n_r illumination cycles, we have $0 < t_1 \leq t_2, \ldots, \leq t_{N(t)} \leq n_r t_r$), whereas *detection time* refers to the time of detection relative to the most recent illumination pulse, which is the absolute detection time modulo t_r. Therefore, the limiting empirical distribution of detection times would be $\propto \lambda(t)$.
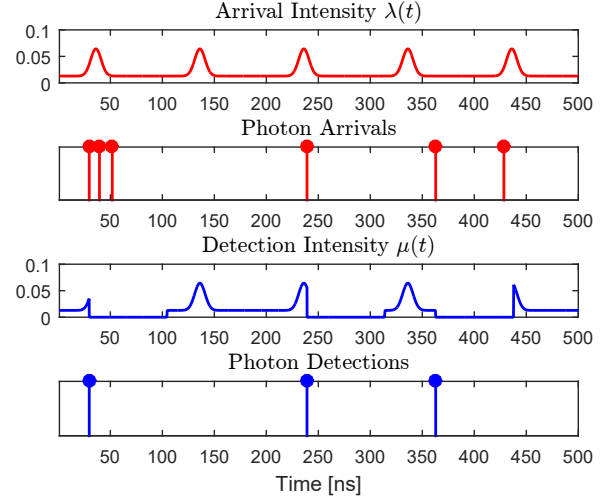

Fig. 1: Illustration of the effect of dead times on the detection process for $t_r = 100$ ns, $t_d = 75$ ns, $\sigma = 5$ ns, $S = 0.5$, and $B = 1$. Photon arrival times are generated according to the arrival intensity $\lambda(t)$. The detection intensity $\mu(t)$ is equal to the arrival intensity $\lambda(t)$ except immediately following a photon detection when $\mu(t) = 0$, so the detection times are a subset of the arrival times and detection is not a Poisson process.

the empirical distribution of X_i's would be identical to the arrival time PDF, given by

$$f_{X_A}(x) = \lambda(x)/\Lambda, \quad \text{for } x \in [0, t_r). \quad (5)$$

The following proposition provides statistical characterization of $\{X_i\}_{i\in\mathbb{N}}$ in the presence of dead time.

Proposition 1. *Suppose that the photon arrival process is an inhomogeneous Poisson process with periodic intensity function $\lambda(t)$, whose period is t_r, and the detector has dead time t_d. Define $x_d := t_d \bmod t_r$. Let the random sequence $\{T_i\}_{i\in\mathbb{N}}$ denote absolute detection times and define detection times as $X_i := T_i \bmod t_r$, for all $i \in \mathbb{N}$. Then the random sequence $\{X_i\}_{i\in\mathbb{N}}$ forms a Markov chain with state space $[0, t_r)$ and transition PDF*

$$f_{X_{i+1}|X_i}(x_{i+1}|x_i) = \frac{\lambda(x_{i+1})}{1 - \exp(-\Lambda)}$$

$$\exp\left(-\int_{x_i+x_d}^{\left\lceil \frac{x_i+x_d-x_{i+1}}{t_r} \right\rceil t_r + x_{i+1}} \lambda(\tau)\,d\tau\right), \quad (6)$$

where $\lceil a \rceil := \lfloor a \rfloor + 1$ and $\Lambda := \int_0^{t_r} \lambda(\tau)\,d\tau$.

Proof. See Appendix A. □

We can check that $\{X_i\}_{i\in\mathbb{N}}$ is ψ-irreducible, recurrent, and aperiodic, and hence it has a unique stationary PDF [48, Proposition 10.4.2]. Denoting the stationary PDF by f_{X_D}, then for all $x \in [0, t_r)$, f_{X_D} satisfies

$$f_{X_D}(x) = \int_0^{t_r} f_{X_D}(y) f_{X_{i+1}|X_i}(x|y)\,dy. \quad (7)$$

That is, f_{X_D} is the eigenfunction corresponding to eigenvalue 1 of the linear operator $\mathcal{P}$ defined as

$$\mathcal{P}f(x) := \int_0^{t_r} f(y) f_{X_{i+1}|X_i}(x|y)\,\mathrm{d}y. \qquad (8)$$

For the special case where $x_d = 0$, we show in the following that the arrival PDF f_{X_A} defined in (5) satisfies the stationary condition, meaning that dead time does not cause any distortion in detection time distribution; this result has also been noted in [34] with a different derivation. With $f_{X_\mathrm{D}}(x) = \lambda(x)/\Lambda$ and $x_d = 0$, the RHS of (7) is

$$\int_0^{t_r} \frac{\lambda(y)}{\Lambda} f_{X_{i+1}|X_i}(x|y)\mathrm{d}y$$
$$= \int_0^x \frac{\lambda(y)}{\Lambda} \frac{\lambda(x)}{1-\exp(-\Lambda)} \exp\left(-\int_y^x \lambda(\tau)\mathrm{d}\tau\right)$$
$$+ \int_x^{t_r} \frac{\lambda(y)}{\Lambda} \frac{\lambda(x)}{1-\exp(-\Lambda)} \exp\left(-\int_y^{t_r+x} \lambda(\tau)\mathrm{d}\tau\right)$$
$$= \frac{\lambda(x)}{\Lambda}\left[\frac{\int_0^x \lambda(y)\exp\left(-\int_y^x \lambda(\tau)\mathrm{d}\tau\right)\mathrm{d}y}{1-\exp(-\Lambda)} \right.$$
$$\left. + \frac{\int_x^{t_r} \lambda(y)\exp\left(-\int_y^{t_r+x} \lambda(\tau)\mathrm{d}\tau\right)\mathrm{d}y}{1-\exp(-\Lambda)} \right].$$

Label the two terms in the square brackets as A_1 and A_2. Using the chain rule and the Leibniz rule for differentiation, for any constant a that does not depend on y, we have that

$$\frac{\mathrm{d}}{\mathrm{d}y}\exp\left(-\int_y^a \lambda(\tau)\mathrm{d}\tau\right) = \lambda(y)\exp\left(-\int_y^a \lambda(\tau)\mathrm{d}\tau\right).$$

Letting $a = x$, we have

$$A_1 = \frac{\exp\left(-\int_y^x \lambda(\tau)\mathrm{d}\tau\right)\Big|_0^x}{1-\exp(-\Lambda)} = \frac{1-\exp\left(-\int_0^x \lambda(\tau)\mathrm{d}\tau\right)}{1-\exp(-\Lambda)}.$$

Similarly, let $a = t_r + x$, then

$$A_2 = \frac{\exp\left(-\int_0^x \lambda(\tau)\mathrm{d}\tau\right) - \exp(-\Lambda)}{1-\exp(-\Lambda)}.$$

It follows that $A_1 + A_2 = 1$, and so $\lambda(x)/\Lambda$ is the stationary distribution of the Markov chain when $x_d = 0$.

To numerically demonstrate the correctness of (6) for general x_d, we partition the state space $[0, t_r)$ into n_b equally spaced time bins with bin centers $\{b_n\}_{n=1}^{n_b}$ and approximate the linear operator $\mathcal{P}$ defined in (8) with an $n_b \times n_b$ matrix $\mathbf{P}$, where $P_{m,n} := f_{X_{i+1}|X_i}(b_n|b_m)$ with $f_{X_{i+1}|X_i}$ defined in (6). The matrix $\mathbf{P}$ is then normalized to have row sum equal to 1 so that it becomes a probability transition matrix $\widetilde{\mathbf{P}}$. A discrete approximation of f_{X_D}, denoted by a length-n_b row vector $\mathbf{f}_{X_\mathrm{D}}$, is then obtained as the leading left eigenvector of $\widetilde{\mathbf{P}}$, since $\mathbf{f}_{X_\mathrm{D}}$ should satisfy the Markov chain stationary condition $\mathbf{f}_{X_\mathrm{D}} = \mathbf{f}_{X_\mathrm{D}}\widetilde{\mathbf{P}}$. Moreover, if the second largest (in terms of magnitude) eigenvalue of $\widetilde{\mathbf{P}}$ is strictly less than one, in other words, $\widetilde{\mathbf{P}}$ admits a spectral gap, then the corresponding Markov chain converges to its stationary distribution geometrically fast. We have verified that in all parameter settings considered in this paper, $\widetilde{\mathbf{P}}$ admits a spectral gap, thus confirming the convergence of the chain. Finally, $\mathbf{f}_{X_\mathrm{D}}$ is compared with the histogram of a set of simulated detection times, where we expect a close match between the simulated histogram and $\mathbf{f}_{X_\mathrm{D}}$. Detection times are simulated by first generating arrival times according to (1). Then starting with detection of the first arrival time generated, subsequent arrivals are culled from the sequence if they are within t_d of the previous absolute detection time, as in [49]. Note that unlike in [49], both background photons and dark counts are considered to trigger dead times in the same manner as signal detections.

Comparisons between a histogram of detection times collected from simulation and the corresponding $\mathbf{f}_{X_\mathrm{D}}$ are shown in Fig. 2. In each simulation, the number of illuminations is $n_r = 50000$ and the half pulse width is $\sigma = 2$ ns. The close matches between predicted detection PDFs and the simulated histogram results validate the effectiveness of the Markov chain model in deriving the limiting distribution. The figure further illustrates the effect that dead time has on TCSPC. The first column of Fig. 2 shows results with $S = B = 0.1$, so the total flux Λ is low enough that few photons arrive during the detector dead time, and the arrival and detection densities are almost identical. If just the signal flux is increased, e.g., by increasing the illumination laser power (second column), the photon detection density narrows and shifts slightly toward earlier detection times (similar to the phenomenon of range walk error), due to early arrivals from the pulse blocking later photons from being detected. When the background flux increases, the distortions in the density due to dead time become more apparent. However, these distortions also depend on the particular values of t_r and t_d. When t_r is slightly larger than t_d (such as for $t_r = 80$ in the top row of Fig. 2), the dead time triggered after a signal detection will reset just before signal photons from the next pulse arrive at the detector. The dead time thus behaves as a signal-triggered gate, blocking detection of many background photons while allowing detection of additional signal. On the other hand, increasing t_r by just 20 ns (bottom row) causes a significant ripple in the detection PDF a duration t_d after the main signal peak (modulo t_r). The dead time is again often triggered by signal photons when S is large, but the reset of the detector in the next cycle allows incident background photons to be detected, amplifying the apparent background intensity at that part of the cycle. Note that this pre-pulse ripple could easily be mistaken for optical system inter-reflections or poor electronics thresholding if detector dead time were not taken into account.

C. Comparison of Fisher Information

High-flux acquisition enables detection of more photons than low-flux acquisition for a fixed number of illuminations. Although the detection time distribution is distorted in the sense that it is different from the arrival time distribution, our Markov chain model allows us to accurately predict the distortion. Therefore, it is expected that for a fixed number of illuminations, high-flux acquisition with our probabilistic model for detection times can improve ranging performance over the 5% low-flux acquisition rule. Another interesting
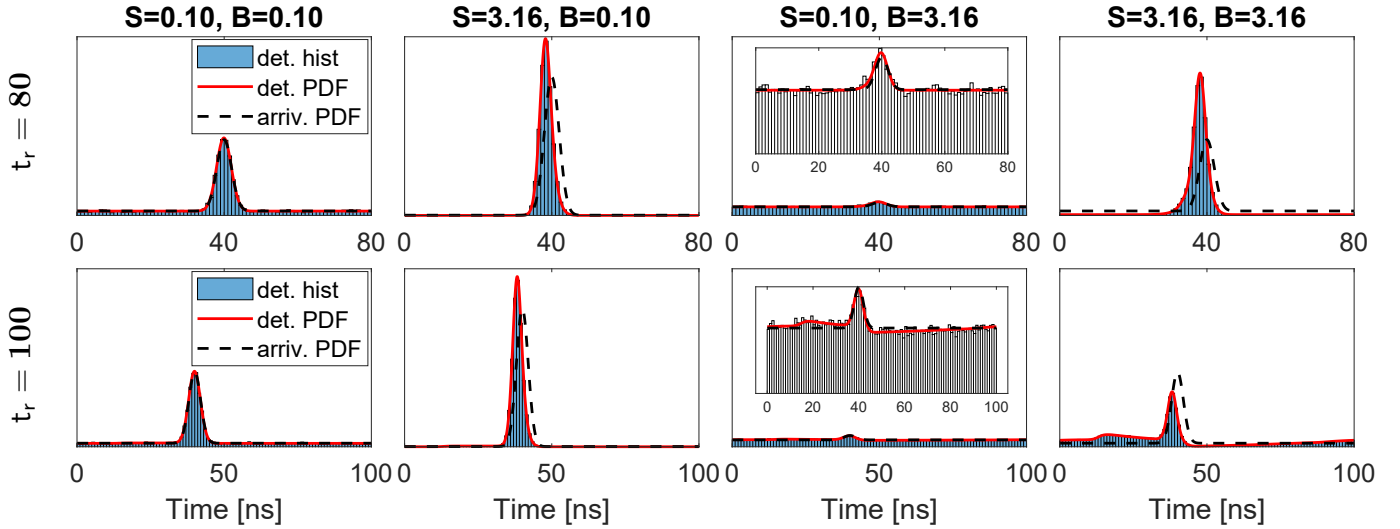
Fig. 2: Comparisons between histograms of simulated detection times, predicted detection time PDFs, and arrival time PDFs illustrate how dead time affects the detection process. In addition to a shift in the mode toward earlier detection times, the dead time may also cause a ripple in the detection PDF relative to the arrival PDF. Plots are shown for $\sigma = 2$ ns, $t_{bin} = 50$ ps, $n_r = 50000$ illuminations, and $t_d = 75$ ns. The vertical axis scale is constant for each row. An inset with a different vertical scale is included for each plot in the third column to emphasize the ripple that is not easily seen in the original scale.

aspect is to compare estimates from low-flux and high-flux acquisitions for a fixed number of detections. By comparing the arrival PDF, which is equivalent to the low-flux detection PDF, and the high-flux detection PDF in Fig. 2, we notice that dead time results in a "narrowing" of the pulse, especially for large S. Thus, we speculate that the distortion may in fact be favorable for depth estimation in some cases.

To verify this somewhat counter-intuitive speculation, we compare the Fisher information per detection for estimating the depth z from the low-flux PDF f_{X_A} (5) and high-flux PDF f_{X_D} (the stationary distribution of the Markov chain defined in Proposition 1),[3] which are denoted by FI_A and FI_D, respectively, and are computed as (derivation is provided in Appendix B)

$$FI_A = \int_0^{t_r} \left(\frac{\partial}{\partial z} f_{X_A}(x; z) \right)^2 \frac{1}{f_{X_A}(x; z)} \, dx,$$

$$FI_D = \int_0^{t_r} \left(\frac{\partial}{\partial z} f_{X_D}(x; z) \right)^2 \frac{1}{f_{X_D}(x; z)} \, dx,$$

where the derivative of f_{X_D} is computed numerically. Note that while realizations of detection times are not i.i.d. samples of f_{X_A} or f_{X_D}, most ranging algorithms only use the empirical distribution of detection times for depth estimation. Therefore, it is reasonable to consider Fisher information of the limiting empirical distributions f_{X_A} and f_{X_D} rather than that of the joint distributions. Fig. 3 presents the Fisher information ratio FI_D/FI_A for $t_d = 75$ ns and with t_r varying from 50 to 500 ns. By (6), we notice that only $x_d := t_d \mod t_r$ affects the detection time distribution. Hence, for the case where $t_r = 50$ ns, the effective dead time in terms of detection time distribution is 25 ns. In the regions where the ratio is greater than one, f_{X_D} is more informative about the depth z than f_{X_A}

(i.e., the dead time effect is beneficial) in the sense that the per-detection Fisher information is higher. We notice that such a region usually appears when B is not too large and the SBR is sufficiently high. A potential reduction in depth error variance was likewise noted by Heide et al. [8], but that analysis assumed zero background, which is a naïve assumption for most applications and which our analysis shows is not a necessary condition for dead time to be beneficial. When t_r is slightly larger than t_d (as for $t_r = 80$ in Fig. 3), the signal-triggered gating extends the region in which dead time is beneficial to larger B compared to the cases where t_r is much larger than t_d. Together with the plots in Fig. 2, this suggests that the most photon-efficient benefit from dead time is achieved when t_r is slightly larger than t_d.[4] This condition may be difficult to achieve in practice as the dead time is not tunable in many devices and adjustment of the illumination period is limited by the required maximum unambiguous range.

III. ARRIVAL INTENSITY ESTIMATION ALGORITHM

In this section, we first derive a maximum likelihood (ML) estimator for estimating the total flux Λ from absolute detection times. Then we derive an algorithm for estimating the arrival intensity $\lambda(x)$ for $x \in [0, t_r)$ from a histogram of detection times assuming that Λ is known; one may implement our algorithm with a calibrated Λ when available or with a Λ estimated by, for example, our ML estimator.

A. Maximum Likelihood Estimator for Λ

In [35], Isbaner et al. note the necessity of estimating Λ in order to correctly reconstruct the histogram of photon arrival times. Define the interdetection period R_i as the number of

---

[3]The reciprocal of the Fisher information is a lower bound for the mean squared error (MSE) of any unbiased estimator; this is also known as the Cramér-Rao inequality [50, Theorem 3.1].

[4]Note that more photons would be detected with shorter t_d, but each detection would likely be less informative of the depth.

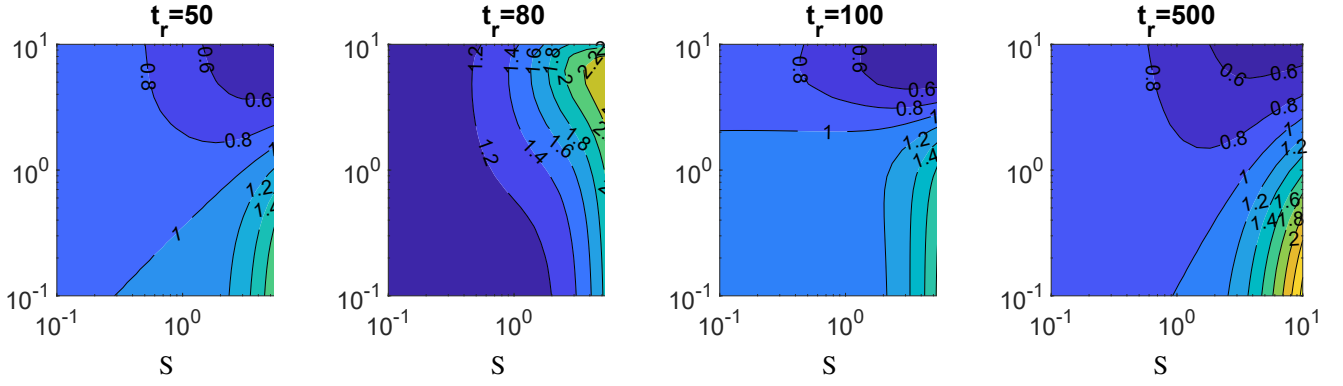Fig. 3: The Fisher information ratio $\text{FI}_\text{D}/\text{FI}_\text{A}$ indicates the performance improvement that may be gained for the same number of detections when high-flux data is used instead of low-flux data. The plots show for various signal rate S and background rate B and for $\sigma = 0.2$ ns, $\text{t}_\text{bin} = 10$ ps, and $\text{t}_\text{d} = 75$ ns that when SBR is sufficiently high and B is not too large, the effect of dead time is beneficial for range estimation.

completed periods after the detector reset at $T_i + \text{t}_\text{d}$ before another photon is detected at time T_{i+1}:

$$R_i := \left\lfloor \frac{T_{i+1} - (T_i + \text{t}_\text{d})}{\text{t}_\text{r}} \right\rfloor, \tag{9}$$

where T_i's are absolute detection times. Isbaner et al. claim that $P(R_i = r) \propto \exp(-r\Lambda)$ and use weighted least squares to fit an exponential function. In the following proposition, we verify the claim using properties of Poisson processes and the Markov nature of detections with dead time. Moreover, we show that R_i's are independent, and so an ML estimator for Λ can be easily computed from a realization of R_i's.

Proposition 2. *The random variables R_i's defined in* (9) *are i.i.d. with the same probability distribution as R, where*

$$P(R = r) = (1 - \exp(-\Lambda)) \exp(-r\Lambda), \quad r \in \{0\} \cup \mathbb{N}. \tag{10}$$

Proof. See Appendix C. □

By Proposition 2, given a realization of interdetection periods $\{r_i\}_{1 \le i \le n}$, the log-likelihood function is

$$\mathcal{L}(\{r_i\}_{i=1}^n; \Lambda) = -\Lambda \sum_{i=1}^n r_i + n \ln\left(1 - \exp(-\Lambda)\right).$$

Setting the derivative of $\mathcal{L}(\{r_i\}_{i=1}^n; \Lambda)$ with respect to Λ to zero, we obtain the ML estimator for Λ as

$$\widehat{\Lambda}^{\text{ML}} = -\ln\left(\frac{\sum_{i=1}^n r_i}{n + \sum_{i=1}^n r_i}\right). \tag{11}$$

Note that the distribution of R can be understood as follows. The number of photon arrivals per period is Poisson with parameter Λ, so $p = 1 - \exp(-\Lambda)$ is the probability of at least one photon arriving in a period. Then R has a geometric distribution $P(R = r) = (1 - p)^r p$, which matches (10).

In what follows, we consider estimating the arrival intensity λ from a detection time histogram assuming that Λ is known.

B. Relationship between Arrival and Detection Distributions

Plugging (6) into (7), we have

$$f_{X_\text{D}}(x) = \lambda(x) \left[\int_0^{x-\text{x}_\text{d}} f_{X_\text{D}}(y) \frac{\exp\left(-\int_{y+\text{x}_\text{d}}^x \lambda(\tau)\,\text{d}\tau\right)}{1 - \exp(-\Lambda)}\,\text{d}y \right.$$
$$\left. + \int_{x-\text{x}_\text{d}}^{\text{t}_\text{r}} f_{X_\text{D}}(y) \frac{\exp\left(-\int_{y+\text{x}_\text{d}}^{\text{t}_\text{r}+x} \lambda(\tau)\,\text{d}\tau\right)}{1 - \exp(-\Lambda)}\,\text{d}y \right] \tag{12}$$

for $x > \text{x}_\text{d}$, and

$$f_{X_\text{D}}(x) = \lambda(x) \left[\int_0^{\text{t}_\text{r}+x-\text{x}_\text{d}} f_{X_\text{D}}(y) \frac{\exp\left(-\int_{y+\text{x}_\text{d}}^{\text{t}_\text{r}+x} \lambda(\tau)\,\text{d}\tau\right)}{1 - \exp(-\Lambda)}\,\text{d}y \right.$$
$$\left. + \int_{\text{t}_\text{r}+x-\text{x}_\text{d}}^{\text{t}_\text{r}} f_{X_\text{D}}(y) \frac{\exp\left(-\int_{y+\text{x}_\text{d}}^{2\text{t}_\text{r}+x} \lambda(\tau)\,\text{d}\tau\right)}{1 - \exp(-\Lambda)}\,\text{d}y \right] \tag{13}$$

for $x \le \text{x}_\text{d}$. In (12) and (13), denote the factors in the brackets as $a(x)$ and we can then write $f_{X_\text{D}}(x) = \lambda(x)a(x)$, where $a(x)$ can be interpreted as the attenuation effect on the arrival intensity due to dead time. It is worth mentioning that similar factorization of f_{X_D} was also used in Isbaner et al. [35] for the derivation of their dead time correction algorithm. However, such a factorization is assumed at the beginning of their derivation, whereas we arrive at this factorization naturally from the stationary condition of a Markov chain.

Plugging $f_{X_\text{D}}(x) = \lambda(x)a(x)$ into (12), we have

$$a(x) = \int_0^{x-\text{x}_\text{d}} \lambda(y)a(y) \frac{\exp\left(-\int_{y+\text{x}_\text{d}}^x \lambda(\tau)\,\text{d}\tau\right)}{1 - \exp(-\Lambda)}\,\text{d}y$$
$$+ \int_{x-\text{x}_\text{d}}^{\text{t}_\text{r}} \lambda(y)a(y) \frac{\exp\left(-\int_{y+\text{x}_\text{d}}^{\text{t}_\text{r}+x} \lambda(\tau)\,\text{d}\tau\right)}{1 - \exp(-\Lambda)}\,\text{d}y. \tag{14}$$

Differentiating both sides of the above equation with respect to x:

$$a'(x) \overset{(a)}{=} \frac{1}{1 - \exp(-\Lambda)} \left[\lambda(x - \mathsf{x_d}) a(x - \mathsf{x_d}) \left(1 - \exp(-\Lambda)\right) \right.$$
$$- \lambda(x) \left(\int_0^{x - \mathsf{x_d}} \lambda(y) a(y) \exp\left(-\int_{y + \mathsf{x_d}}^x \lambda(\tau) \, \mathsf{d}\tau \right) \mathsf{d}y \right.$$
$$\left. \left. + \int_{x - \mathsf{x_d}}^{\mathsf{t_r}} \lambda(y) a(y) \exp\left(-\int_{y + \mathsf{x_d}}^{\mathsf{t_r} + x} \lambda(\tau) \, \mathsf{d}\tau \right) \mathsf{d}y \right) \right]$$
$$\overset{(b)}{=} \lambda(x - \mathsf{x_d}) a(x - \mathsf{x_d}) - \lambda(x) a(x), \quad (15)$$

where step (a) uses the Leibniz rule and the fact that $\lambda(\mathsf{t_r} + x) = \lambda(x)$ and step (b) follows by noticing from (14) that the sum of the two integrals equals $(1 - \exp(-\Lambda)) a(x)$. Similarly, we can obtain from (13) that

$$a'(x) = \lambda(\mathsf{t_r} + x - \mathsf{x_d}) a(\mathsf{t_r} + x - \mathsf{x_d}) - \lambda(x) a(x). \quad (16)$$

Note that if we consider periodic extensions of $a(x)$ and $f_{X_{\mathrm{D}}}(x)$, then (15) and (16) are identical. In the following, $a(x)$ and $f_{X_{\mathrm{D}}}(x)$ are considered as their periodic extensions.

Integrating both sides of (15), we have that

$$a(x) = -\int_{x - \mathsf{x_d}}^x \lambda(\tau) a(\tau) \, \mathsf{d}\tau + C, \quad (17)$$

where C is a constant. Multiplying both sides of (17) by $\lambda(x)$, we have that

$$f_{X_{\mathrm{D}}}(x) = -\lambda(x) \int_{x - \mathsf{x_d}}^x f_{X_{\mathrm{D}}}(\tau) \, \mathsf{d}\tau + C\lambda(x). \quad (18)$$

Define

$$g(x) := \int_{x - \mathsf{x_d}}^x f_{X_{\mathrm{D}}}(\tau) \, \mathsf{d}\tau. \quad (19)$$

Since $f_{X_{\mathrm{D}}}(x)$ is a proper probability density function on the interval $[0, \mathsf{t_r})$, it satisfies

$$1 = \int_0^{\mathsf{t_r}} f_{X_{\mathrm{D}}}(x) \, \mathsf{d}x = -\int_0^{\mathsf{t_r}} \lambda(x) g(x) \, \mathsf{d}x + C \int_0^{\mathsf{t_r}} \lambda(x) \mathsf{d}x.$$

It follows that

$$C = \frac{1 + \int_0^{\mathsf{t_r}} \lambda(x) g(x) \, \mathsf{d}x}{\int_0^{\mathsf{t_r}} \lambda(x) \, \mathsf{d}x} = \frac{1 + \int_0^{\mathsf{t_r}} \lambda(x) g(x) \, \mathsf{d}x}{\Lambda}. \quad (20)$$

Plugging (19) and (20) into (18), we have the following relationship between the arrival intensity function $\lambda(x)$ and the limiting distribution of the detection times $f_{X_{\mathrm{D}}}(x)$:

$$f_{X_{\mathrm{D}}}(x) = -\lambda(x) g(x) + \frac{1 + \int_0^{\mathsf{t_r}} \lambda(x) g(x) \, \mathsf{d}x}{\Lambda} \lambda(x). \quad (21)$$

C. Nonlinear Inverse Formulation and Algorithm

Suppose that the time interval $[0, \mathsf{t_r})$ is partitioned into $\mathsf{n_b}$ equally spaced time bins with bin size $\mathsf{t_{bin}}$, which is the case in TCSPC.[5] Define $\mathsf{n_d} := \mathsf{x_d}/\mathsf{t_{bin}}$. Let the normalized histogram

---

[5]TCSPC systems digitize photon detection times using time-to-amplitude converters (TACs) or time-to-digital converters (TDCs) [1]. The number of bits allocated for an event record is usually fixed, so there is an inherent tradeoff between the acquisition resolution and repetition period. For instance, the HydraHarp 400 dedicates 15 bits per record (32 768 bins) and has a base bin resolution of 1 ps, which can be multiplied by powers of 2 [10].

of detection times be denoted by $\mathbf{h} = (h_1, \ldots, h_{\mathsf{n_b}})$, where $\sum_{i=1}^{\mathsf{n_b}} h_i = 1$. A discrete model for (21) is then

$$\mathbf{h} = -\mathrm{diag}(\mathbf{g})\boldsymbol{\lambda} + \Lambda^{-1}\boldsymbol{\lambda} + \Lambda^{-1}(\mathbf{g}^T \boldsymbol{\lambda})\boldsymbol{\lambda} + \Delta,$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{\mathsf{n_b}})$ is a discretization of $\lambda(t)$; Λ is the total flux and is assumed to be known; $g_i = \sum_{k=i-\mathsf{n_d}}^{i-1} h_k$ for $i > \mathsf{n_d}$ and $g_i = \sum_{k=i-\mathsf{n_d}+\mathsf{n_b}}^{\mathsf{n_b}} h_k + \sum_{k=1}^{i-1} h_k$ for $i \leq \mathsf{n_d}$, which follows from (19); $\mathrm{diag}(\mathbf{g})$ is a diagonal matrix with $\mathbf{g}$ on its diagonal; and Δ represents the error due to discretization and the difference between the finite-sample empirical distribution and the limiting distribution. For any fixed $\mathbf{h}$ (hence fixed $\mathbf{g}$), define an operator $\mathcal{T}(\cdot; \mathbf{h}) : \mathbb{R}^{\mathsf{n_b}} \to \mathbb{R}^{\mathsf{n_b}}$ as

$$\boldsymbol{\lambda} \mapsto \mathcal{T}(\boldsymbol{\lambda}; \mathbf{h}) := -\mathrm{diag}(\mathbf{g})\boldsymbol{\lambda} + \Lambda^{-1}\boldsymbol{\lambda} + \Lambda^{-1}(\mathbf{g}^T \boldsymbol{\lambda})\boldsymbol{\lambda}. \quad (22)$$

The inverse problem that we need to solve is then to estimate $\boldsymbol{\lambda}$ from the nonlinear system $\mathbf{h} = \mathcal{T}(\boldsymbol{\lambda}; \mathbf{h}) + \Delta$ given a measurement vector $\mathbf{h}$. Define the optimization problem:

$$\min_{\boldsymbol{\lambda}} \left\{ F(\boldsymbol{\lambda}) := D(\boldsymbol{\lambda}) + \delta_{[0,M]^{\mathsf{n_b}}}(\boldsymbol{\lambda}) \right\}, \quad (23)$$

where $D(\boldsymbol{\lambda}) := \frac{1}{2} \|\mathbf{h} - \mathcal{T}(\boldsymbol{\lambda}; \mathbf{h})\|^2$ with $\|\cdot\|$ being the Euclidean norm and $\delta_{[0,M]^{\mathsf{n_b}}}$ the indicator function of the bounded hypercube $[0, M]^{\mathsf{n_b}}$ for some constant M. Note that while one may include stronger regularizers to reflect prior knowledge about $\boldsymbol{\lambda}$, our goal here is to demonstrate that the proposed method can reconstruct the arrival intensity without any prior knowledge other than the intensity being non-negative; the method is thus applicable to a broader class of applications where the arrival intensity is less predictable such as in NLOS imaging. We use a monotone accelerated proximal gradient (APG) algorithm [51] to solve (23). Note that the proximal operator for $\delta_{[0,M]^{\mathsf{n_b}}}$ is the orthogonal projector onto $[0, M]^{\mathsf{n_b}}$, denoted by $\Pi_{[0,M]^{\mathsf{n_b}}}(\cdot)$, and the gradient of $D(\boldsymbol{\lambda})$ is computed as follows:

$$\nabla D(\boldsymbol{\lambda}) = \mathbf{J}_{\mathcal{T}}^T \left(\mathcal{T}(\boldsymbol{\lambda}; \mathbf{h}) - \mathbf{h} \right)$$
$$= \left(\frac{\mathbf{g}\boldsymbol{\lambda}^T}{\Lambda} + \frac{1 + \mathbf{g}^T \boldsymbol{\lambda}}{\Lambda} \mathbf{I} - \mathrm{diag}(\mathbf{g}) \right) \left(\mathcal{T}(\boldsymbol{\lambda}; \mathbf{h}) - \mathbf{h} \right), \quad (24)$$

where $\mathbf{J}_{\mathcal{T}}$ is the Jacobian matrix of $\mathcal{T}$ and $\mathbf{I}$ is the identity matrix. We emphasize that $\mathbf{h}$, $\mathbf{g}$, and Λ are fixed throughout the algorithm, thus they are treated as constant instead of functions of $\boldsymbol{\lambda}$ when computing the gradient $\nabla D(\boldsymbol{\lambda})$ in (24). The convergence of the monotone APG algorithm relies on an appropriate choice of the step size γ, which should satisfy $\gamma < 1/L$, where L is the Lipschitz constant of $\nabla D(\cdot)$ [51]. The following proposition provides an upper-bound L_u for L.

Proposition 3. *The Lipschitz constant L of the function $\nabla D(\cdot)$ defined in (24) is upper-bounded by L_u on $[0, M]^{\mathsf{n_b}}$, where L_u is defined as*

$$L_u := 2\Lambda^{-2}\mathsf{n_b}M^2 + \left(2\Lambda^{-2} + 2 + 6\Lambda^{-1}\right)\sqrt{\mathsf{n_b}}M + 4\Lambda^{-1} + 2.$$

Proof. See Appendix D. □

Setting the step size $\gamma = 1/L_u$, starting with some initialization $\boldsymbol{\lambda}^0 = \boldsymbol{\lambda}^1 = \mathbf{z}^1 \in [0, M]^{\mathsf{n_b}}$ and $q_0 = 0, q_1 = 1$, for
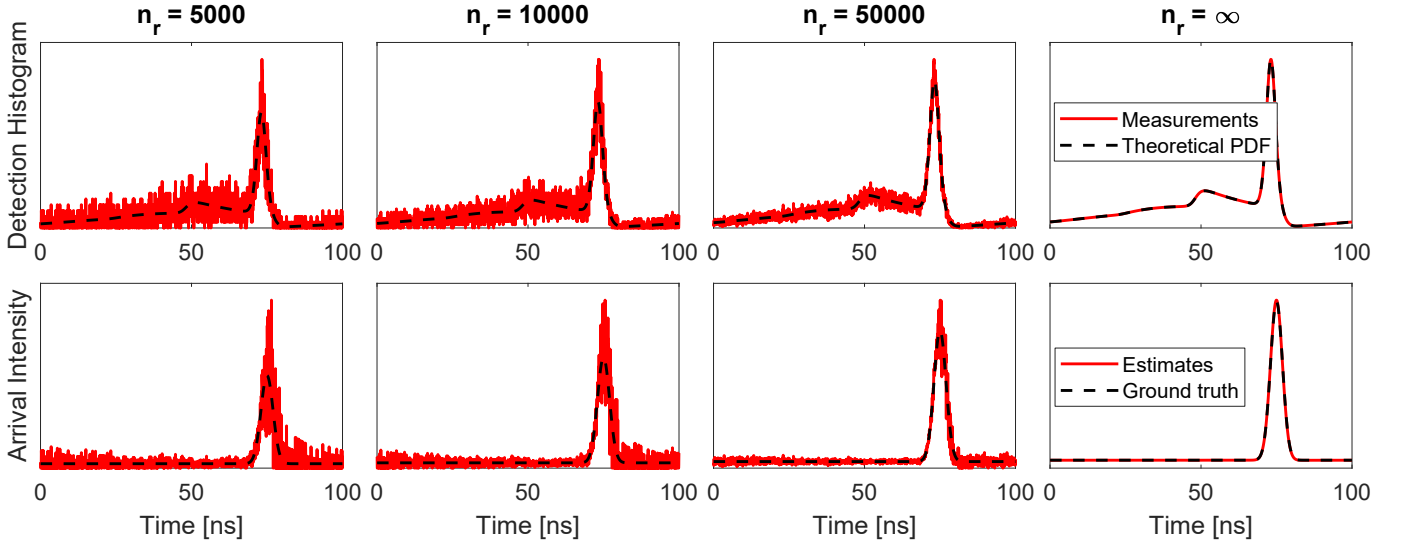
Fig. 4: Estimation of arrival intensity (bottom) from detection histogram (top) when $S = B = 3.16$, $\sigma = 2$ ns, $t_{bin} = 50$ ps, $t_r = 100$ ns, and $t_d = 75$ ns. From left to right: increased number of illuminations (n_r), where in the last column, theoretical detection histogram is used as measurement.

$k \geq 1$, the monotone APG algorithm for solving (23) proceeds as follows:

$$
\begin{aligned}
\mathbf{y}^k &= \boldsymbol{\lambda}^k + \frac{q_{k-1}}{q_k}\big(\mathbf{z}^k - \boldsymbol{\lambda}^k\big) + \frac{q_{k-1}-1}{q_k}\big(\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\big), \\
\mathbf{z}^{k+1} &= \Pi_{[0,M]^{n_b}}\big(\mathbf{y}^k - \gamma \nabla D(\mathbf{y}^k)\big), \\
\mathbf{x}^{k+1} &= \Pi_{[0,M]^{n_b}}\big(\boldsymbol{\lambda}^k - \gamma \nabla D(\boldsymbol{\lambda}^k)\big), \\
q_{k+1} &= \frac{\sqrt{4q_k^2 + 1} + 1}{2}, \\
\boldsymbol{\lambda}^{k+1} &= \begin{cases} \mathbf{z}^{k+1}, & \text{if } F(\mathbf{z}^{k+1}) \leq F(\mathbf{x}^{k+1}); \\ \mathbf{x}^{k+1}, & \text{otherwise.} \end{cases}
\end{aligned}
\tag{25}
$$

Since (23) is a nonconvex optimization problem, a good initialization is important to avoid converging to local minima that are not global minima. We now introduce an initialization scheme. Let C_λ be a scalar that depends on λ through $C_\lambda = \int_0^{t_r} \lambda(x) g(x)\, dx$. Then (21) can be written as $f_{X_D}(x) = -\lambda(x) g(x) + (1 + C_\lambda)\lambda(x)/\Lambda$, which implies

$$
\lambda(x) = \frac{f_{X_D}(x)}{(1 + C_\lambda)/\Lambda - g(x)}.
\tag{26}
$$

Plugging (26) back into the definition of C_λ, we obtain a fixed point equation for C_λ:

$$
C_\lambda = \int_0^{t_r} \frac{f_{X_D}(x) g(x)}{(1 + C_\lambda)/\Lambda - g(x)}\, dx.
\tag{27}
$$

Notice that by (26) the feasible set for C_λ is $\mathcal{C} = \{C \in \mathbb{R} : (1+C)/\Lambda - g(x) > 0, \forall x \in [0, t_r]\}$. It follows that the RHS of (27) is positive on $\mathcal{C}$ and monotone decreasing to zero as C_λ goes to infinity. Since the LHS of (27) is linearly increasing, a graph will easily show that (27) has a unique fixed point on $\mathcal{C}$. Therefore, if f_{X_D} is known perfectly, then one can solve (27) for C_λ and plug C_λ into (26) to have a perfect reconstruction of λ. However, in practice, we only have a histogram formed by a limited number of measured

detection times. Nevertheless, it is plausible to estimate C_λ as the fixed point of

$$
\widehat{C}_\lambda = \sum_{i=1}^{n_b} \frac{h_i\, g_i}{(1 + \widehat{C}_\lambda)/\Lambda - g_i}.
\tag{28}
$$

We can then use $\boldsymbol{\lambda}^0$ with the ith entry being defined as

$$
\lambda_i^0 = \frac{h_i}{(1 + \widehat{C}_\lambda)/\Lambda - g_i}
\tag{29}
$$

as the initialization of the nonconvex optimization problem. While we do not have a theoretical guarantee for convergence to the global minimum, in our simulation, both $\mathbf{h}$ plus random perturbation and the more principled initialization (29) lead to good estimates. Because solving (28) is easy and the initial estimate (29) is usually close to the solution, using the principled initialization allows the algorithm to converge faster.

Fig. 4 presents simulated detection histograms and the corresponding arrival intensity estimates using (25), where $S = B = 3.16$ and $\sigma = 2$ ns. We notice that as n_r increases, the detection histogram approaches f_{X_D}. Our estimated arrival intensity likewise approaches the true arrival intensity as n_r increases. It is interesting to note that while the error in the detection histogram resembles Poisson noise in that the variance increases as the mean increases, the error in the arrival intensity estimate is signal-dependent in a different way. We observe that the error variance is roughly proportional to the pointwise ratio of f_{X_D} and λ. Although we have no theoretical results supporting this hypothesis, the observation suggests that the portions of the arrival intensity easiest to reconstruct are those least attenuated by the dead time effects, and vice versa.

IV. APPLICATION TO RANGING

We now explore how the theory and algorithm developed in Sections II and III can be used for depth estimation. In Section IV-A, we assume that the acquisition parameters S, B, Λ

are known from accurate calibration and compare different methods using the true parameter values. In Section IV-B, we provide estimators for B and S, and together with the estimator for Λ introduced in Section III-A, we test our proposed methods using the estimated parameters.

A. Ranging with True Acquisition Parameters

The ML depth estimator for the Poisson arrival process passes the set of arrivals through a log-matched filter that is matched to the arrival intensity $\lambda(t)$, where the log-matched filter is defined as $v(t) := \log(\lambda(t)) = \log\left(f_{X_A}(t)\right) + \log(\Lambda)$ [52, (33)]. Given a set of low-flux detection times $\{t_i\}_{i=1}^n$, where photon loss due to dead time is negligible, the log-matched filter for estimating the depth z is defined as

$$\widehat{z}\left(\{t_i\}_{i=1}^n; f_{X_A}\right) := \arg\max_z \sum_{i=1}^n \int_0^{t_r} \delta(t - t_i) v\left(t + 2z/c\right) \mathrm{d}t$$

$$= \arg\max_z \sum_{i=1}^n \log\left(f_{X_A}\left(t_i + 2z/c\right)\right).$$

For practical implementation, $\{t_i\}_{i=1}^n$ may be quantized into n_b equally spaced time bins over $[0, t_r)$ with bin centers $\{b_k\}_{k=1}^{n_b}$. A histogram $\mathbf{h} = (h_1, \ldots, h_{n_b})$ for the low-flux detection times can then be obtained from the quantized data $\{\bar{t}_k\}_{k=1}^n$. Moreover, instead of estimating the depth z, we can estimate the time delay $\tau := 2z/c$, since the mapping from z to τ is one-to-one. The estimator for τ is then

$$\widehat{\tau}\left(\mathbf{h}; f_{X_A}\right) := \arg\max_{\tau \in \Gamma} \left\{ \sum_{k=1}^{n_b} h_k \log\left(f_{X_A}\left(b_k + \tau\right)\right) \right\} + b_{n_b/2},$$

where $\Gamma := \{-b_{n_b/2}, \ldots, -b_1, 0, b_1, \ldots, b_{n_b/2}\}$ is a set of on-grid relative time delays, and f_{X_A} is the arrival PDF assuming $b_{n_b/2}$ is the true delay.

In Section II, we have derived the limit of the empirical distribution of high-flux detection times f_{X_D}. Hence, we can similarly define a log-matched filter matched to f_{X_D} and define an estimator as

$$\widehat{\tau}\left(\mathbf{h}; f_{X_D}\right) := \arg\max_{\tau \in \Gamma} \left\{ \sum_{k=1}^{n_b} h_k \log\left(f_{X_D}\left(b_k + \tau\right)\right) \right\} + b_{n_b/2},$$

which is preferable if $\mathbf{h}$ is obtained via high-flux acquisition with non-negligible photon loss due to dead time. Note however that $\widehat{\tau}\left(\mathbf{h}; f_{X_D}\right)$ is not the ML estimator with dead time effects (even without quantization error), because in this case the joint PDF does not factorize as product of the marginals. While one can obtain the exact joint PDF from the transition PDF (6) and the marginal PDF f_{X_D}, the true ML estimator is inconvenient to implement. Therefore, $\widehat{\tau}\left(\mathbf{h}; f_{X_D}\right)$ is used in our simulations.

In Section III, we have derived an algorithm for estimating the arrival time distribution from the detection time distribution. Hence, given a detection histogram, our algorithm can compute an estimate for the arrival histogram $\widehat{\mathbf{h}}^A$, and then $\tau(\widehat{\mathbf{h}}^A; f_{X_A})$ can be used for depth estimation.

Based on the above discussion, letting $\mathbf{h}^{LF}$ and $\mathbf{h}^{HF}$ denote the detection time histograms obtained via low-flux and high-flux acquisitions, respectively, we compare six depth estimation methods applicable to asynchronous TCSPC systems. The methods are as follows:

1) **LF**: The low-flux approach first attenuates the incident flux (in practice by applying a neutral density filter) to limit the total flux arriving at the detector to 5% so that dead time effects can be ignored. Since the low-flux detection histogram $\mathbf{h}^{LF}$ can then be considered to be the same as the arrival histogram, it then uses $\widehat{\tau}(\mathbf{h}^{LF}; f_{X_A})$ as the estimator.

2) **HF**: The high-flux method naïvely assumes that dead time has no effect on the acquisition and uses the estimator $\widehat{\tau}(\mathbf{h}^{HF}; f_{X_A})$, even when $\mathbf{h}^{HF}$ is not a good approximation to the arrival histogram.

3) **SC**: Shift correction assumes that the dead time only adds a bias to the estimate and that the bias can be computed and subsequently subtracted away. In practice, this is equivalent to the optical calibration procedure in [45]; for our simulations, we compute the shift in the mode of f_{X_D} compared to that of f_{X_A} and subtract the shift correction from the HF estimate.

4) **Isbaner**: This method, based on the work of Isbaner et al., first estimates the arrival histogram $\widehat{\mathbf{h}}^A$ from $\mathbf{h}^{HF}$ using the algorithm in [35], which has publicly available code,[6] and then applies the estimator $\widehat{\tau}(\widehat{\mathbf{h}}^A; f_{X_A})$. While [35] can estimate Λ from data, we provide the algorithm with the true Λ for fair comparison.

5) **Proposed method 1 – MCPDF**: Our first method computes the Markov chain-based PDF f_{X_D} to directly apply $\widehat{\tau}(\mathbf{h}^{HF}; f_{X_D})$.

6) **Proposed method 2 – MCHC**: Our second method is similar to that of Isbaner et al., except it first estimates the arrival histogram $\widehat{\mathbf{h}}^A$ from $\mathbf{h}^{HF}$ using the Markov chain-based histogram correction algorithm introduced in Section III and then uses $\widehat{\tau}(\widehat{\mathbf{h}}^A; f_{X_A})$ as the estimator.

We perform Monte Carlo simulations with $t_r = 100$ ns, $t_d = 75$ ns, $\sigma = 0.2$ ns, and bin duration $t_{bin} = 5$ ps, which are reasonable experimental parameters for some laboratory settings. For each combination of S and B, we generate 600 realizations of the arrival process with $n_r = 10^4$ illuminations. Starting with the first arrival, the high-flux detection sequence is generated by removing subsequent arrivals if they occur within t_d of the previous detection. Generation of the corresponding low-flux detection sequence proceeds in the same manner, but the arrival process is first attenuated via Bernoulli thinning, so photons arrive in only 5% of illumination periods on average. For each method, the log-matched filtering is performed via circular cross-correlation (circular convolution of the histogram with the time-reversed PDF). This is due to the asynchronous dead time preserving the shift invariance of the arrival process.

Fig. 5 compares the MSE for time delay estimation achieved by the six compared methods as a function of the number of illuminations. We observe that MCPDF usually achieves
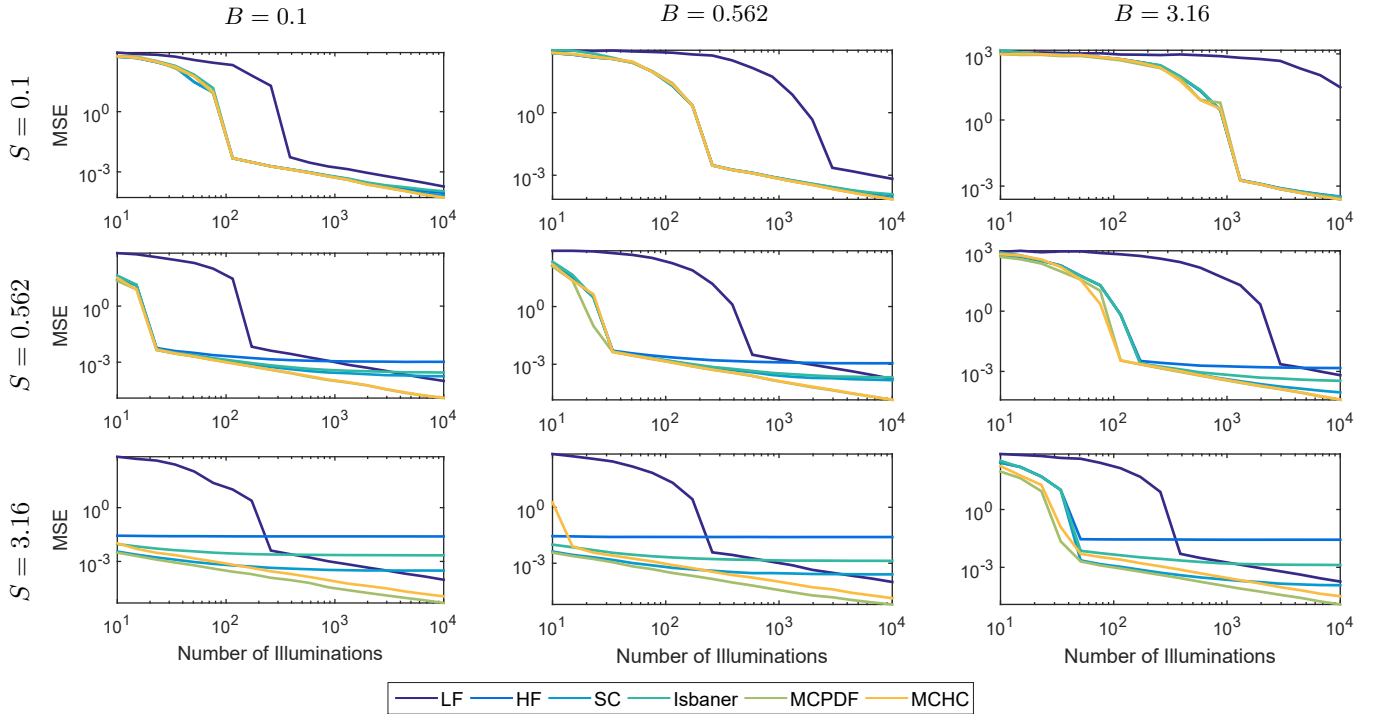
---

[6]http://projects.gwdg.de/projects/deadtimecorrectiontcspc

Fig. 5: Plots of the MSE for ranging as a function of n_r for $t_r = 100$ ns, $t_d = 75$ ns, $\sigma = 0.2$ ns, $t_{bin} = 5$ ps and various S and B values. Our proposed methods (MCPDF and MCHC) take advantage of the increased detection rate to perform more accurate ranging than with the low-flux acquisition for all values of S, B, and n_r.

the lowest MSE, since it directly performs parameter estimation with the updated detection model. Equivalently, MCPDF needs the fewest illuminations to achieve a given MSE, hence enabling the fastest acquisition. The MSE of MCHC is comparable to that of MCPDF, limited only in that it must first invert the histogram before estimating the depth. Compared to the LF approach, both MCPDF and MCHC require fewer illuminations to achieve the same MSE, and that time efficiency increases as S and B increase and dead time has a more significant impact. Regarding the other approaches, HF is more effective only for low numbers of illuminations but the estimate quickly becomes biased and is therefore not suited to precise depth measurement. Correcting for this bias with SC is quite effective for extending to somewhat higher n_r, although eventually more accurate modeling is necessary for more precise estimates. The state-of-the-art method for dealing with asynchronous dead-time models by Isbaner et al. [35] achieves low MSE when the total flux is low or moderate, while the accuracy degrades in high-flux scenarios. The performance degradation is due to their approximation of the detection time distribution being less accurate in high-flux settings (S. Isbaner, personal communication, May 14, 2018).

In addition to enabling faster acquisition, we explored whether dead time could lead to more accurate ranging for an equal number of detected photons. The Fisher information analysis in Section II-C has provided a theoretical prediction that for sufficiently high SBR, estimating depth from the dead time-distorted detection time distribution can yield lower MSE than that from the arrival time distribution. Although the estimators in our Monte Carlo simulation are not guaranteed

to achieve the Cramér-Rao lower bound (i.e., the reciprocal of Fisher information), we would like to see whether the reduction of ranging error due to dead time also exists with simple and commonly used estimators. Fig. 6 compares the MSE for time delay estimation by the six methods as a function of the number of detections. We notice that for the high SBR cases where $S = 3.16, B = 0.1$ and $S = 3.16, B = 0.562$, MCPDF outperforms LF, which provides numerical evidence that dead time can be beneficial when properly modeled.

B. Ranging with Estimated Acquisition Parameters

The results in Section IV-A use methods that compute f_{X_D}, f_{X_A}, and $\widehat{h}^A$ assuming the true values of B, S, and Λ are known. However, in most practical scenarios, this information will not be available *a priori*. We describe here some strategies that can be used for determining those parameters.

1) Maximum Likelihood Estimator for B: We begin by assuming that background calibration measurements can occasionally be made within the ranging process, for which photons are detected while the laser is turned off. For sequences of ranging measurements such as in 3D imaging, such background-only acquisitions could be made for each laser position or for sets of laser points (e.g., once per row or once per image). Since the background process is homogeneous with $\lambda(t) = \lambda_b$, we can rewrite (4) as

$$f_{T_{i+1}|T_i}(t|t_i) = \lambda_b \exp\big(- \lambda_b\big(t - (t_i + t_d)\big)\big)\mathbb{I}\{t > t_i + t_d\}.$$
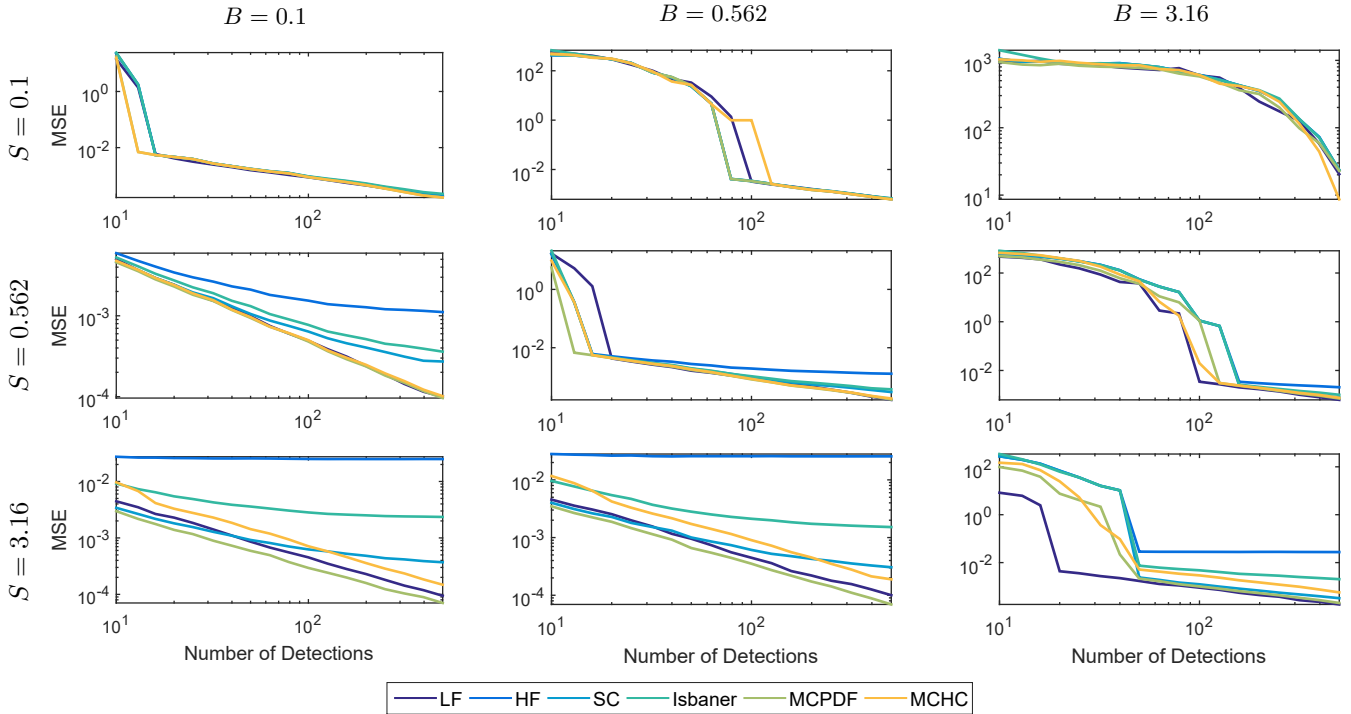
Fig. 6: Plots of the MSE for ranging as a function of the number of detections for $t_r = 100$ ns, $t_d = 75$ ns, $\sigma = 0.2$ ns, $t_{bin} = 5$ ps and various S and B values. For high S and B not too large, the presence of dead time actually improves ranging accuracy with our MCPDF method relative to the low flux measurements due to the narrowing of the signal pulse.

Then the conditional distribution of T_i's given $T_1 = t_1$ is

$$f_{T_2,\ldots,T_n|T_1}(t_2,\ldots,t_n|t_1) = \prod_{i=1}^{n-1} f_{T_{i+1}|T_i}(t_{i+1}|t_i)$$

$$= \lambda_b^{n-1} \exp(-\lambda_b(t_n - t_1) + (n-1)\lambda_b t_d).$$

Given a set of absolute detection times $\{t_i\}_{i=1}^n$, the (conditional) log-likelihood function $\ln\left(f_{T_2,\ldots,T_n|T_1}\right)$ is

$$\mathcal{L}(\{t_i\}_{i=1}^n; \lambda_b) = (n-1)\ln(\lambda_b) - \lambda_b(t_n - t_1) + (n-1)\lambda_b t_d.$$

Setting the derivative of $\mathcal{L}(\{t_i\}_{i=1}^n; \lambda_b)$ with respect to λ_b to zero, we obtain the (conditional) ML estimator for λ_b as

$$\widehat{\lambda}_b^{ML} = \frac{n-1}{(t_n - t_1) - (n-1)t_d}.$$

It follows that $\widehat{B}^{ML} = \widehat{\lambda}_b^{ML} t_r$.

2) Estimating S: From $\widehat{\Lambda}^{ML}$ and $\widehat{B}^{ML}$, we could also compute $\widehat{S} = \max(\widehat{\Lambda}^{ML} - \widehat{B}^{ML}, 0)$ to ensure non-negativity. However, setting $\widehat{S} = 0$ whenever $\widehat{B}^{ML} > \widehat{\Lambda}^{ML}$ is not informative for depth estimation with the log-matched filter, since f_{X_D} and f_{X_A} would be uniform PDFs. Instead, one can assume that there is always at least some small amount of signal and background in the ranging process, so we choose to set minimum values of $S_{min} = B_{min} = 0.01$. Then $\widehat{B} = \max(\widehat{B}^{ML}, B_{min})$, $\widehat{\Lambda} = \max(\widehat{\Lambda}^{ML}, \widehat{B} + S_{min})$, and $\widehat{S} = \widehat{\Lambda} - \widehat{B}$.

Fig. 7 shows one example of estimates using this strategy for 500 Monte Carlo trials with $S = B = 0.562$, $t_r = 100$ ns, $t_d = 75$ ns, $\sigma = 0.2$ ns, and $t_{bin} = 10$ ps. In Fig. 7a, the $\widehat{B}$, $\widehat{\Lambda}$, and $\widehat{S}$ estimates consistently improve as n_r increases beyond a very small number of detections. The resulting depth
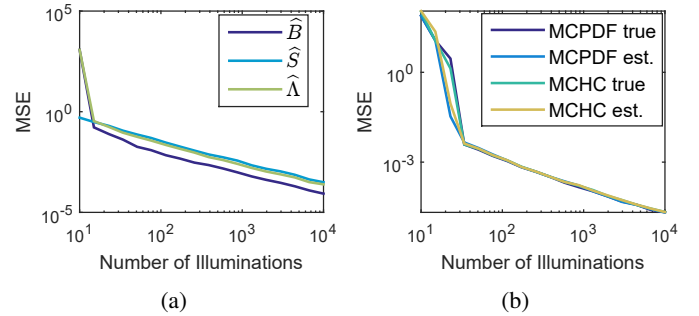

(a)   (b)

Fig. 7: Plots of the parameter estimates as a function of n_r for $S = B = 0.562$, $t_r = 100$ ns, $t_d = 75$ ns, $\sigma = 0.2$ ns, and $t_{bin} = 10$ ps. In (a), the estimates $\widehat{B}$, $\widehat{\Lambda}$, and $\widehat{S}$ improve as n_r increases. The ranging results in (b) using estimated parameters show no degradation in performance compared to the methods with parameters known *a priori*.

estimates shown in Fig. 7b are virtually indistinguishable from the methods using the known parameter values. It is worth noting that, while $\widehat{B}^{ML}$ applies for any value of B, $\widehat{\Lambda}^{ML}$ becomes less reliable for large Λ since $P(R = 0)$ in (10) approaches unity. If the number of illuminations was not fixed in advance, one could pursue an adaptive acquisition strategy as in [53]. Alternatively, for 3D imaging, one could take advantage of spatial correlations to estimate Λ, for example, to borrow measurements from neighboring pixels.

V. CONCLUSION

This work studied dead time compensation for a modern, asynchronous, nonparalyzable detector. By using a Markov

chain model for detection times, we obtained the limit of the empirical distribution of detection times as the stationary distribution of the Markov chain. We found that the Fisher information per detection can be higher for this limiting distribution than for the arrival distribution, which suggests that the distortion due to dead time can be beneficial for depth estimation if used properly. Indeed, simulation results showed that our first proposed method MCPDF, which is a log-matched filter matched to the limiting distribution, achieved lower error than the low-flux method for a fixed number of detections when the SBR is sufficiently high. By exploiting the stationary condition for the Markov chain, we derived our second proposed method MCHC, which estimates the arrival distribution from the detection distribution by solving a nonlinear inverse problem with a provably convergent optimization algorithm, and then the corrected histogram is used in a log-matched filter estimator. Although we only tested MCHC in the context of ranging, it makes no assumptions about the arrival intensity and should thus be applicable to other dead-time-limited TCSPC applications, including FLIM and NLOS imaging; we leave such extensions as future work.

APPENDIX A
PROOF OF PROPOSITION 1

First, we show that $\{X_i\}_{i\in\mathbb{N}}$ is a Markov chain. Define $\mathcal{B}_i := \{X_i \leq x_i\}, \forall i \in \mathbb{N}$. We need to establish that

$$P(\mathcal{B}_{i+1}|X_k = x_k, \forall k \leq i) = P(\mathcal{B}_{i+1}|X_i = x_i). \quad (30)$$

The following equivalence of events will be useful in the proof, as it relates the sets defined by elements of $\{X_i\}_{i\in\mathbb{N}}$ to those of $\{T_i\}_{i\in\mathbb{N}}$, which has a known transition density (4):

$$\begin{aligned}
\{X_{i+1} \leq x_{i+1}\} &= \cup_{k=0}^{\infty}\{k\mathsf{t_r} < T_{i+1} \leq k\mathsf{t_r} + x_{i+1}\}, \\
\{T_i = k_i\mathsf{t_r} + x_i\} &= \{K_i = k_i\} \cap \{X_i = x_i\}.
\end{aligned} \quad (31)$$

Let $\mathsf{t_d} = \mathsf{k_d}\mathsf{t_r} + \mathsf{x_d}$, where $\mathsf{k_d} = \lfloor \mathsf{t_d}/\mathsf{t_r} \rfloor$ and $\mathsf{x_d} = \mathsf{t_d} \mod \mathsf{t_r}$. Moreover, define $\mathcal{A}_k := \{k\mathsf{t_r} < T_{i+1} \leq k\mathsf{t_r} + x_{i+1}\}, \forall k \in \mathbb{N} \cup \{0\}$. Then we have

$$\begin{aligned}
&P(\mathcal{B}_{i+1}|X_j = x_j, \forall j \leq i) \\
&= \sum_{k_1,\ldots,k_i} \Bigg(P(\mathcal{B}_{i+1}|X_j = x_j, K_j = k_j, \forall j \leq i) \\
&\qquad\qquad \cdot P(K_j = k_j, \forall j \leq i|X_j = x_j, \forall j \leq i)\Bigg) \\
&= \sum_{k_1,\ldots,k_i} \Bigg(P(\cup_{k=0}^{\infty}\mathcal{A}_k|T_j = k_j\mathsf{t_r} + x_j, \forall j \leq i) \\
&\qquad\qquad \cdot P(K_j = k_j, \forall j \leq i|X_j = x_j, \forall j \leq i)\Bigg) \quad (32)
\end{aligned}$$

where the summation is over all $0 \leq k_1 \leq \cdots \leq k_i < \infty$ and the last equality follows by (31). In the following, we show

that the first probability in (32) only depends on x_i and $\mathsf{x_d}$.

$$P(\cup_{k=0}^{\infty}\mathcal{A}_k|T_j = k_j\mathsf{t_r} + x_j, \forall j \leq i)$$

$$\overset{(a)}{=} \sum_{k=0}^{\infty} P(\mathcal{A}_k|T_i = k_i\mathsf{t_r} + x_i)$$

$$\overset{(b)}{=} \int_{(k_i+\mathsf{k_d})\mathsf{t_r}+x_i+x_d}^{(k_i+\mathsf{k_d})\mathsf{t_r}+x_{i+1}} \lambda(t) \exp\left(-\int_{(k_i+\mathsf{k_d})\mathsf{t_r}+x_i+\mathsf{x_d}}^{t} \lambda(\tau)\mathsf{d}\tau\right)\mathsf{d}t$$

$$+ \sum_{k=k_i+\mathsf{k_d}+1}^{\infty} \int_{k\mathsf{t_r}}^{k\mathsf{t_r}+x_{i+1}} \lambda(t) \exp\left(-\int_{(k_i+\mathsf{k_d})\mathsf{t_r}+x_i+\mathsf{x_d}}^{t} \lambda(\tau)\mathsf{d}\tau\right)\mathsf{d}t, \quad (33)$$

where step (a) follows by the Markov property of $\{T_i\}_{i=1}^{\infty}$ and $\{\mathcal{A}_k\}$ being disjoint and in step (b), we have plugged in (4) and assumed that $x_i + \mathsf{x_d} \leq x_{i+1} \leq \mathsf{t_r}$. Note that other relationships between x_i, x_{i+1}, $\mathsf{x_d}$, $\mathsf{t_r}$ may lead to slightly different expression, but the derivation follows similarly. (We will see that the expression does not depend on $\mathsf{k_d}$.) Label the two terms in (33) as S_1 and S_2. First, consider S_1:

$$\begin{aligned}
S_1 &= \int_{x_i+\mathsf{x_d}}^{x_{i+1}} \lambda(t) \exp\left(-\int_{k_i\mathsf{t_r}+x_i+\mathsf{x_d}}^{k_i\mathsf{t_r}+t} \lambda(\tau)\mathsf{d}\tau\right)\mathsf{d}t \\
&= \int_{x_i+\mathsf{x_d}}^{x_{i+1}} \lambda(t) \exp\left(-\int_{x_i+\mathsf{x_d}}^{t} \lambda(\tau)\mathsf{d}\tau\right)\mathsf{d}t,
\end{aligned}$$

which follows by change of variable and $\lambda(t + k_i\mathsf{t_r}) = \lambda(t)$. Next consider S_2:

$$\begin{aligned}
S_2 &= \sum_{k=k_i+\mathsf{k_d}+1}^{\infty} \int_0^{x_{i+1}} \lambda(t) \exp\left(-\int_{(k_i+\mathsf{k_d})\mathsf{t_r}+x_i+\mathsf{x_d}}^{k\mathsf{t_r}+t} \lambda(\tau)\mathsf{d}\tau\right)\mathsf{d}t \\
&= \sum_{k=0}^{\infty} \int_0^{x_{i+1}} \lambda(t) \exp\left(-\int_{(k_i+\mathsf{k_d})\mathsf{t_r}+x_i+\mathsf{x_d}}^{(k+k_i+\mathsf{k_d}+1)\mathsf{t_r}+t} \lambda(\tau)\,\mathsf{d}\tau\right)\mathsf{d}t \\
&= \sum_{k=0}^{\infty} (\exp(-\Lambda))^k \int_0^{x_{i+1}} \lambda(t) \exp\left(-\int_{x_i+\mathsf{x_d}}^{\mathsf{t_r}+t} \lambda(\tau)\,\mathsf{d}\tau\right)\mathsf{d}t \\
&= \frac{\int_0^{x_{i+1}} \lambda(t) \exp\left(-\int_{x_i+\mathsf{x_d}}^{\mathsf{t_r}+t} \lambda(\tau)\,\mathsf{d}\tau\right)\mathsf{d}t}{1 - \exp(-\Lambda)}.
\end{aligned}$$

Notice that neither S_1 nor S_2 depends on $\mathsf{k_d}$, $\{K_j\}_{j\leq i}$, or $\{X_j\}_{j<i}$. Plugging S_1 and S_2 back into (32), we have that

$$\begin{aligned}
&P(\mathcal{B}_{i+1}|X_j = x_j, \forall j \leq i) \\
&= \sum_{k_1,\ldots,k_i} (S_1 + S_2)P(K_j = k_j, \forall j \leq i|X_j = x_j, \forall j \leq i) \\
&= S_1 + S_2,
\end{aligned}$$

where the last equality holds since $P(\cdot|X_j = x_j, \forall j \leq i)$ is a probability measure and that the summation $\sum_{k_1,\ldots,k_i}$ is over all $0 \leq k_1 \leq \ldots \leq k_i < \infty$. Hence, we have established (30), and therefore proved that $\{X_i\}_{i\in\mathbb{N}}$ is a Markov chain.

Next, we compute the transition PDF to justify (6):

$$f_{X_{i+1}|X_i}(x_{i+1}|x_i) = \frac{\mathsf{d}}{\mathsf{d}x_{i+1}} P(X_{i+1} \leq x_{i+1}|X_i = x_i)$$

$$= \frac{\mathsf{d}}{\mathsf{d}x_{i+1}}S_1 + \frac{\mathsf{d}}{\mathsf{d}x_{i+1}}S_2 = \frac{\lambda(x_{i+1}) \exp\left(-\int_{x_i+\mathsf{x_d}}^{x_{i+1}} \lambda(\tau)\,\mathsf{d}\tau\right)}{1 - \exp(-\Lambda)}.$$

Recall that we have assumed $x_i + \mathsf{x_d} \leq x_{i+1} \leq \mathsf{t_r}$ in the derivation above, and we can check that it matches (6) for this case. Other cases can be derived similarly.

APPENDIX B
DERIVATION OF FISHER INFORMATION

We present the derivation for $\mathrm{FI_D}$; the derivation for $\mathrm{FI_A}$ follows similarly. By definition of Fisher information:

$$\mathrm{FI_D} = \int_0^{\mathsf{t_r}} \left(-\frac{\partial^2}{\partial z^2} \log \left(f_{X_D}(x;z) \right) \right) f_{X_D}(x;z) \, \mathrm{d}x$$

$$= \int_0^{\mathsf{t_r}} \frac{\left(\frac{\partial}{\partial z} f_{X_D}(x;z) \right)^2}{f_{X_D}(x;z)} \, \mathrm{d}x - \int_0^{\mathsf{t_r}} \frac{\partial^2}{\partial z^2} f_{X_D}(x;z) \, \mathrm{d}x$$

$$\overset{(a)}{=} \int_0^{\mathsf{t_r}} \frac{\left(\frac{\partial}{\partial z} f_{X_D}(x;z) \right)^2}{f_{X_D}(x;z)} \, \mathrm{d}x - \frac{\partial^2}{\partial z^2} \left(\int_0^{\mathsf{t_r}} f_{X_D}(x;z) \, \mathrm{d}x \right)$$

$$= \int_0^{\mathsf{t_r}} \frac{\left(\frac{\partial}{\partial z} f_{X_D}(x;z) \right)^2}{f_{X_D}(x;z)} \, \mathrm{d}x,$$

where the interchange of derivative and integral in step (a) holds trivially, since the range of the integral is finite.

APPENDIX C
PROOF OF PROPOSITION 2

In the following, we will show that

$$P\left(R_i = r_i, R_{i-1} = r_{i-1} \right) = \left(1 - \exp(-\Lambda) \right)^2 \prod_{j=i-1}^{i} \exp(-r_j \Lambda),$$

which would imply that Proposition 2 is valid for R_i and R_{i-1}; the proof for more than two R_i's follows similarly.

Define event $\mathcal{E}_j$ for $j = i-1, i$ as

$$\mathcal{E}_j := \left\{ r_j \mathsf{t_r} + T_j + \mathsf{t_d} \leq T_{j+1} < (r_j + 1)\mathsf{t_r} + T_j + \mathsf{t_d} \right\}.$$

By definition of R_i in (9),

$$P\left(R_i = r_i, R_{i-1} = r_{i-1} \right) = P\left(\mathcal{E}_i \cap \mathcal{E}_{i-1} \right)$$
$$= \mathbb{E}\left[P\left(\mathcal{E}_i \cap \mathcal{E}_{i-1} \,\middle|\, T_{i-1} \right) \right].$$

Note that by the Markov property of absolute detection times as discussed in Section II-B, the joint PDF of T_{i+1}, T_i given $T_{i-1} = t_{i-1}$ is

$$f_{T_{i+1}, T_i | T_{i-1}}(t_{i+1}, t_i | t_{i-1}) = f_{T_{i+1} | T_i}(t_{i+1} | t_i) f_{T_i | T_{i-1}}(t_t | t_{i-1}).$$

Let $a_j := r_j \mathsf{t_r} + t_j + \mathsf{t_d}$ for $j = i-1, i$. Then

$$P\left(\mathcal{E}_1, \mathcal{E}_2 \,\middle|\, T_{i-1} = t_{i-1} \right)$$
$$= \int_{a_{i-1}}^{\mathsf{t_r} + a_{i-1}} \int_{a_i}^{\mathsf{t_r} + a_i} f_{T_{i+1} | T_i}(t_{i+1} | t_i) f_{T_i | T_{i-1}}(t_i | t_{i-1}) \, \mathrm{d}t_{i+1} \, \mathrm{d}t_i.$$

First consider the inner integral:

$$\int_{a_i}^{\mathsf{t_r} + a_i} f_{T_{i+1} | T_i}(t_{i+1} | t_i) \, \mathrm{d}t_{i+1}$$
$$= \int_{a_i}^{\mathsf{t_r} + a_i} \lambda(t_{i+1}) \exp\left(-\int_{t_i + \mathsf{t_d}}^{t_{i+1}} \lambda(\tau) \, \mathrm{d}\tau \right) \mathrm{d}t_{i+1}$$
$$= -\exp\left(-\int_{t_i + \mathsf{t_d}}^{t_{i+1}} \lambda(\tau) \, \mathrm{d}\tau \right) \Bigg|_{t_{i+1} = a_i}^{t_{i+1} = \mathsf{t_r} + a_i}$$
$$= (1 - \exp(-\Lambda)) \exp(-r_i \Lambda).$$

Note that the inner integral does not depend on t_i. Using similar calculation, we have that the outer integral does not depend on t_{i-1}. Hence,

$$P\left(\mathcal{E}_1 \cap \mathcal{E}_2 \,\middle|\, T_{i-1} = t_{i-1} \right) = P\left(\mathcal{E}_1 \cap \mathcal{E}_2 \right)$$
$$= \left(1 - \exp(-\Lambda) \right)^2 \prod_{j=i-1}^{i} \exp(-r_j \Lambda),$$

which is the desired result.

APPENDIX D
PROOF OF PROPOSITION 3

In the following, we will find an upper bound for the Lipschitz constant L of $\nabla D(\cdot)$ defined in (24). For brevity, we omit the dependence on $\mathbf{h}$ in the notation for $\mathcal{T}$. By (24), we have

$$\Lambda \nabla D(\boldsymbol{\lambda}) = \mathbf{g} \boldsymbol{\lambda}^T \mathcal{T}(\boldsymbol{\lambda}) + \mathcal{T}(\boldsymbol{\lambda}) + \mathbf{g}^T \boldsymbol{\lambda} \mathcal{T}(\boldsymbol{\lambda}) - \Lambda \operatorname{diag}(\mathbf{g}) \mathcal{T}(\boldsymbol{\lambda})$$
$$- \mathbf{g} \boldsymbol{\lambda}^T \mathbf{h} - \mathbf{h} - \mathbf{g}^T \boldsymbol{\lambda} \mathbf{h} + \Lambda \operatorname{diag}(\mathbf{g}) \mathbf{h}.$$

It follows that for any $\mathbf{u}, \mathbf{v} \in [0, M]^{\mathsf{n_b}}$, we have by triangle inequality that

$$\Lambda \| \nabla D(\mathbf{u}) - \nabla D(\mathbf{v}) \|$$
$$\leq \| \mathbf{g} \mathbf{u}^T \mathcal{T}(\mathbf{u}) - \mathbf{g} \mathbf{v}^T \mathcal{T}(\mathbf{v}) \| + \| \mathcal{T}(\mathbf{u}) - \mathcal{T}(\mathbf{v}) \|$$
$$+ \| \mathbf{g}^T \mathbf{u} \mathcal{T}(\mathbf{u}) - \mathbf{g}^T \mathbf{v} \mathcal{T}(\mathbf{v}) \|$$
$$+ \Lambda \| \operatorname{diag}(\mathbf{g}) \mathcal{T}(\mathbf{u}) - \operatorname{diag}(\mathbf{g}) \mathcal{T}(\mathbf{v}) \|$$
$$+ \| \mathbf{g} \mathbf{u}^T \mathbf{h} - \mathbf{g} \mathbf{v}^T \mathbf{h} \| + \| \mathbf{g}^T \mathbf{u} \mathbf{h} - \mathbf{g}^T \mathbf{v} \mathbf{h} \|.$$

Label the six terms on the right hand side as $T_1, \ldots, T_6$. We will show that there exist constants $L_1, \ldots, L_6 < \infty$ such that $T_i \leq L_i \| \mathbf{u} - \mathbf{v} \|, \forall i = 1, \ldots, 6$. Then the Lipschitz constant L of the gradient ∇D is upper bounded by $\Lambda^{-1} \sum_{i=1}^{6} L_i$.

First consider T_2. Let $\hat{g} := \max_{i \in [\mathsf{n_b}]} g_i$. Then

$$T_2 \overset{(a)}{\leq} \| \mathbf{g}^T \mathbf{u} \mathbf{u} - \mathbf{g}^T \mathbf{v} \mathbf{v} \| + \| \mathbf{u} - \mathbf{v} \| + \| \operatorname{diag}(\mathbf{g})(\mathbf{u} - \mathbf{v}) \|$$
$$\overset{(b)}{\leq} \| \mathbf{g}^T \mathbf{u} \mathbf{u} - \mathbf{g}^T \mathbf{u} \mathbf{v} \| + \| \mathbf{g}^T \mathbf{u} \mathbf{v} - \mathbf{g}^T \mathbf{v} \mathbf{v} \| + (1 + \hat{g}) \| \mathbf{u} - \mathbf{v} \|$$
$$\overset{(c)}{\leq} \| \mathbf{g} \| \| \mathbf{u} \| \| \mathbf{u} - \mathbf{v} \| + \| \mathbf{g} \| \| \mathbf{v} \| \| \mathbf{u} - \mathbf{v} \| + (1 + \hat{g}) \| \mathbf{u} - \mathbf{v} \|$$
$$\overset{(d)}{\leq} 2\sqrt{\mathsf{n_b}} M \| \mathbf{u} - \mathbf{v} \| + 2 \| \mathbf{u} - \mathbf{v} \| = 2(\sqrt{\mathsf{n_b}} M + 1) \| \mathbf{u} - \mathbf{v} \|,$$

where step (a) follows by triangle inequality, step (b) follows by triangle inequality and the fact that the largest eigenvalue of a diagonal matrix equals to the largest entry on its diagonal, and step (c) follows by Cauchy–Schwarz. To see step (d), notice that $\| \mathbf{u} \|, \| \mathbf{v} \| \leq \sqrt{\mathsf{n_b}} M$, since $\mathbf{u}, \mathbf{v} \in [0, M]^{\mathsf{n_b}}$ and $\max_{i \in [\mathsf{n_b}]} g_i \leq \| \mathbf{g} \| \leq \| \mathbf{g} \|_1 \leq \| \mathbf{h} \|_1 = 1$ (the second inequality follows by the fact that $\mathbf{g}$ is non-negative and so $\| \mathbf{g} \| = \sqrt{\sum_{i=1}^{\mathsf{n_d}} g_i^2} \leq \sqrt{(\sum_{i=1}^{\mathsf{n_d}} g_i)^2} = \| \mathbf{g} \|_1$, the third inequality assumed $\mathsf{t_d} \leq \mathsf{t_r}$ and the last equality follows by $\mathbf{h}$ being a proper probability density function). Similarly, we can show that $T_4 \leq 2\Lambda(\sqrt{\mathsf{n_b}} M + 1) \| \mathbf{u} - \mathbf{v} \|$, $T_5 \leq \| \mathbf{u} - \mathbf{v} \|$, and $T_6 \leq \| \mathbf{u} - \mathbf{v} \|$.

Next consider T_1:

$$T_1 \overset{(a)}{\le} \left| \mathbf{u}^T \mathcal{T}(\mathbf{u}) - \mathbf{v}^T \mathcal{T}(\mathbf{v}) \right|$$

$$\overset{(b)}{\le} \left| \mathbf{u}^T \mathcal{T}(\mathbf{u}) - \mathbf{v}^T \mathcal{T}(\mathbf{u}) \right| + \left| \mathbf{v}^T \mathcal{T}(\mathbf{u}) - \mathbf{v}^T \mathcal{T}(\mathbf{v}) \right|$$

$$\overset{(c)}{\le} \|\mathcal{T}(\mathbf{u})\| \|\mathbf{u} - \mathbf{v}\| + \|\mathbf{v}\| \|\mathcal{T}(\mathbf{u}) - \mathcal{T}(\mathbf{v})\|$$

$$\overset{(d)}{\le} \left(\Lambda^{-1} n B^2 + \left(\Lambda^{-1} + 2 \right) \sqrt{\mathsf{n_d}} M \right) \|\mathbf{u} - \mathbf{v}\|,$$

where step (a) follows by $\|\mathbf{g}\| \le 1$ as established before, step (b) follows by triangle inequality, step (c) follows by Cauchy–Schwarz, and step (d) follows by

$$\|\mathcal{T}(\mathbf{u})\| = \|\Lambda^{-1} \mathbf{g}^T \mathbf{u}\mathbf{u} + \Lambda^{-1} \mathbf{u} - \mathrm{diag}(\mathbf{g}) \mathbf{u}\|$$

$$\le \Lambda^{-1} \|\mathbf{g}\| \|\mathbf{u}\|^2 + \Lambda^{-1} \|\mathbf{u}\| + \hat{g} \|\mathbf{u}\|$$

$$\le \Lambda^{-1} \mathsf{n_b} M^2 + \Lambda^{-1} \sqrt{\mathsf{n_b}} M + \sqrt{\mathsf{n_b}} M.$$

Similarly, we can show that

$$T_3 \le \left(\Lambda^{-1} \mathsf{n_b} M^2 + \left(\Lambda^{-1} + 2 \right) \sqrt{\mathsf{n_b}} M \right) \|\mathbf{u} - \mathbf{v}\|.$$

Proposition 3 is then obtained by combining the upper bounds for T_1 through T_6.

ACKNOWLEDGMENT

REFERENCES

[1] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*. Berlin: Springer, 2005.

[2] R. H. Hadfield, "Single-photon detectors for optical quantum information applications," *Nature Photonics*, vol. 3, no. 12, pp. 696–705, Dec. 2009.

[3] J. S. Massa, A. M. Wallace, G. S. Buller, S. J. Fancey, and A. C. Walker, "Laser depth measurement based on time-correlated single-photon counting," *Optics Letters*, vol. 22, no. 8, pp. 543–545, 1997.

[4] M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, "Non-line-of-sight imaging using a time-gated single photon avalanche diode," *Optics Express*, vol. 23, no. 16, 2015.

[5] M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," *Nature*, pp. 338–341, Mar. 2018.

[6] Y. Altmann, S. McLaughlin, M. J. Padgett, V. K. Goyal, A. O. Hero, and D. Faccio, "Quantum-inspired computational imaging," *Science*, vol. 361, no. 6403, p. 660, 2018.

[7] S. Pellegrini, G. S. Buller, J. M. Smith, A. M. Wallace, and S. Cova, "Laser-based distance measurement using picosecond resolution time-correlated single-photon counting," *Measurement Science & Technology*, vol. 11, no. 6, pp. 712–716, 2000.

[8] F. Heide, S. Diamond, D. B. Lindell, and G. Wetzstein, "Sub-picosecond photon-efficient 3D imaging using single-photon sensors," *Scientific Reports*, vol. 8, no. 1, p. 17726, Dec. 2018.

[9] A. K. Pediredla, A. C. Sankaranarayanan, M. Buttafava, A. Tosi, and A. Veeraraghavan, "Signal processing based pile-up compensation for gated single-photon avalanche diodes," arXiv:1806.07437 [physics.ins-det], Jun. 2018.

[10] M. Wahl, H. J. Rahn, T. Röhlicke, G. Kell, D. Nettels, F. Hillger, B. Schuler, and R. Erdmann, "Scalable time-correlated photon counting system with multiple independent input channels," *Review of Scientific Instruments*, vol. 79, no. 12, p. 123113, Dec. 2008.

[11] M. Wahl, T. Röhlicke, H.-J. Rahn, R. Erdmann, G. Kell, A. Ahlrichs, M. Kernbach, A. W. Schell, and O. Benson, "Integrated multichannel photon timing instrument with very short dead time and high throughput," *Review of Scientific Instruments*, vol. 84, no. 4, p. 043102, Apr. 2013.

[12] A. Giudice, M. Ghioni, R. Biasi, F. Zappa, S. Cova, P. Maccagnani, and A. Gulinatti, "High-rate photon counting and picosecond timing with silicon-SPAD based compact detector modules," *Journal of Modern Optics*, vol. 54, no. 2-3, pp. 225–237, 2007.

[13] M. Buttafava, G. Boso, A. Ruggeri, A. Dalla Mora, and A. Tosi, "Time-gated single-photon detection module with 110 ps transition time and up to 80 MHz repetition rate," *Review of Scientific Instruments*, vol. 85, no. 8, 2014.

[14] B. V. Gnedenko, "On the theory of Geiger-Müller counters," *Zh. exp. teor. Fiz*, vol. 11, pp. 101–106, 1941.

[15] L. Kosten, "On the frequency distribution of the number of discharges counted by a Geiger-Müller counter in a constant interval," *Physica*, vol. 10, no. 9, pp. 749–756, Nov. 1943.

[16] C. Levert and W. Scheen, "Probability fluctuations of discharges in a Geiger-Müller counter produced by cosmic radiation," *Physica*, vol. 10, no. 4, pp. 225–238, Apr. 1943.

[17] J. D. Kurbatov and H. B. Mann, "Correction of G-M counter data," *Physical Review*, vol. 68, no. 1-2, pp. 40–43, 1945.

[18] W. Feller, "On probability problems in the theory of counters," in *Selected Papers I*, R. L. Schilling, Z. Vondraček, and W. A. Woyczyński, Eds. Cham, Switzerland: Springer International Publishing, 2015, pp. 751–759.

[19] J. Arlt, D. Tyndall, B. R. Rae, D. D.-U. Li, J. A. Richardson, and R. K. Henderson, "A study of pile-up in integrated time-correlated single photon counting systems," *Review of Scientific Instruments*, vol. 84, no. 10, p. 103105, Oct. 2013.

[20] S. Cova, M. Ghioni, M. A. Itzler, J. C. Bienfang, and A. Restelli, "Semiconductor-based detectors," in *Single-Photon Generation and Detection*, A. Migdall, S. V. Polyakov, J. Fan, and J. C. Bienfang, Eds. Academic Press, 2013, ch. 4, pp. 83–146.

[21] G. Humer, M. Peev, C. Schaeff, S. Ramelow, M. Stipčević, and R. Ursin, "A simple and robust method for estimating afterpulsing in single photon detectors," *Journal of Lightwave Technology*, vol. 33, no. 14, pp. 3098–3107, Jul. 2015.

[22] I. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe, and E. Charbon, "Nonuniformity analysis of a 65-kpixel CMOS SPAD imager," *IEEE Transactions on Electron Devices*, vol. 63, no. 1, pp. 57–64, Jan. 2016.

[23] I. M. Antolovic, C. Bruschini, and E. Charbon, "Dynamic range extension for photon counting arrays," *Optics Express*, vol. 26, no. 17, p. 22234, Aug. 2018.

[24] D. F. Yu and J. A. Fessler, "Mean and variance of single photon counting with deadtime," *Physics in Medicine and Biology*, vol. 45, no. 7, pp. 2043–2056, Jul. 2000.

[25] D. V. O'Connor and D. Phillips, *Time-Correlated Single Photon Counting*. Orlando, FL: Academic Press, 1984.

[26] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, "First-photon imaging," *Science*, vol. 343, no. 6166, pp. 58–61, 2014.

[27] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-efficient computational 3-D and reflectivity imaging with single-photon detectors," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 112–125, Jun. 2015.

[28] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform-based analysis of depth images constructed using sparse single-photon data," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 1935–1946, May 2016.

[29] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 445–459, 2017.

[30] P. B. Coates, "The correction for photon 'pile-up' in the measurement of radiative lifetimes," *Journal of Physics E: Scientific Instruments*, vol. 1, no. 8, pp. 878–879, Aug. 1968.

[31] C. C. Davis and T. A. King, "Photon pile-up corrections in the study of time-varying light sources," *Journal of Physics E: Scientific Instruments*, vol. 5, no. 11, pp. 1072–1074, 1972.

[32] L. Xu, Y. Zhang, Y. Zhang, C. Yang, X. Yang, and Y. Zhao, "Restraint of range walk error in a Geiger-mode avalanche photodiode lidar to acquire high-precision depth and intensity information," *Applied Optics*, vol. 55, no. 7, p. 1683, Mar. 2016.

[33] G. Verma and R. J. Drost, "Inhomogeneous Poisson process rate function inference from dead-time limited observations," *Journal of the Optical Society of America A*, vol. 34, no. 5, p. 770, May 2017.

[34] A. Cominelli, G. Acconcia, P. Peronio, M. Ghioni, and I. Rech, "High-speed and low-distortion solution for time-correlated single photon counting measurements: A theoretical analysis," *Rev. Scientific Instruments*, vol. 88, no. 12, p. 123701, Dec. 2017.

[35] S. Isbaner, N. Karedla, D. Ruhlandt, S. C. Stein, A. Chizhik, I. Gregor, and J. Enderlein, "Dead-time correction of fluorescence lifetime measurements and fluorescence lifetime imaging," *Optics Express*, vol. 24, no. 9, p. 9429, 2016.

[36] M. Wahl, H. J. Rahn, I. Gregor, R. Erdmann, and J. Enderlein, "Dead-time optimized time-correlated photon counting instrument with synchronized, independent timing channels," *Review of Scientific Instruments*, vol. 78, no. 3, 2007.

[37] M. Patting, P. Reisch, M. Sackrow, R. Dowler, M. Koenig, and M. Wahl, "Fluorescence decay data analysis correcting for detector pulse pile-up at very high count rates," *Optical Engineering*, vol. 57, no. 03, p. 1, Jan. 2018.

[38] T. Trigano, T. Dautremer, E. Barat, and A. Souloumiac, "Pile-up correction algorithms for nuclear spectrometry," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. IV, 2005, pp. 441–444.

[39] Y. Sepulcre, T. Trigano, and Y. Ritov, "Sparse regression algorithm for activity estimation in γ spectrometry," *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4347–4359, Sep. 2013.

[40] T. Trigano, Y. Sepulcre, and Y. Ritov, "Sparse reconstruction algorithm for nonhomogeneous counting rate estimation," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 372–385, Jan. 2017.

[41] S. Pommé, R. Fitzgerald, and J. Keightley, "Uncertainty of nuclear counting," *Metrologia*, vol. 52, no. 3, pp. S3–S17, Jun. 2015.

[42] P. Palojärvi, T. Ruotsalainen, and J. Kostamovaara, "A 250-MHz BiCMOS receiver channel with leading edge timing discriminator for a pulsed time-of-flight laser rangefinder," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 6, pp. 1341–1349, Jun. 2005.

[43] S. Kurtti and J. Kostamovaara, "Pulse width time walk compensation method for a pulsed time-of-flight laser rangefinder," in *2009 IEEE Instrumentation and Measurement Technology Conference*, May 2009, pp. 1059–1062.

[44] H. S. Cho, C. H. Kim, and S. G. Lee, "A high-sensitivity and low-walk error LADAR receiver for military application," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 10, pp. 3007–3015, Oct. 2014.

[45] W. He, B. Sima, H. Chen, Yunfei a Dai, Q. Chen, and G. Gu, "A correction method for range walk error in photon counting 3D imaging LIDAR," *Optics Communications*, vol. 308, pp. 211–217, Nov. 2013.

[46] L. Ye, G. Gu, W. He, H. Dai, and Q. Chen, "A real-time restraint method for range walk error in 3-D imaging lidar via dual detection," *IEEE Photonics Journal*, vol. 10, no. 2, 2018.

[47] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. Springer Science & Business Media, 2012.

[48] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 2012.

[49] Q. Hernandez, D. Gutierrez, and A. Jarabo, "A computational model of a single-photon avalanche diode sensor for transient imaging," arXiv:1703.02635 [physics.ins-det], Feb. 2017.

[50] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.

[51] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," in *Proc. Advances in Neural Information Processing Systems 28*, Montreal, Canada, December 7–12 2015.

[52] I. Bar-David, "Communication under the Poisson regime," *IEEE Trans. Inf. Theory*, vol. 15, no. 1, pp. 31–37, 1969.

[53] S. C. Medin, J. Murray-Bruce, D. Castañón, and V. K. Goyal, "Beyond binomial and negative binomial: Adaptation in Bernoulli parameter estimation," arXiv:1809.08801 [stat.AP], Sep. 2018.