

Markov-Renewal Single-Photon LiDAR Simulator

Weijian Zhang[✉], Prateek Chennuri[✉], Hashan K Weerasooriya[✉], Bole Ma[✉],
and Stanley H. Chan[✉]

Purdue University, West Lafayette IN 47906, USA
{zhan5056, pchennur, hweeraso, ma929, stanchan}@purdue.edu

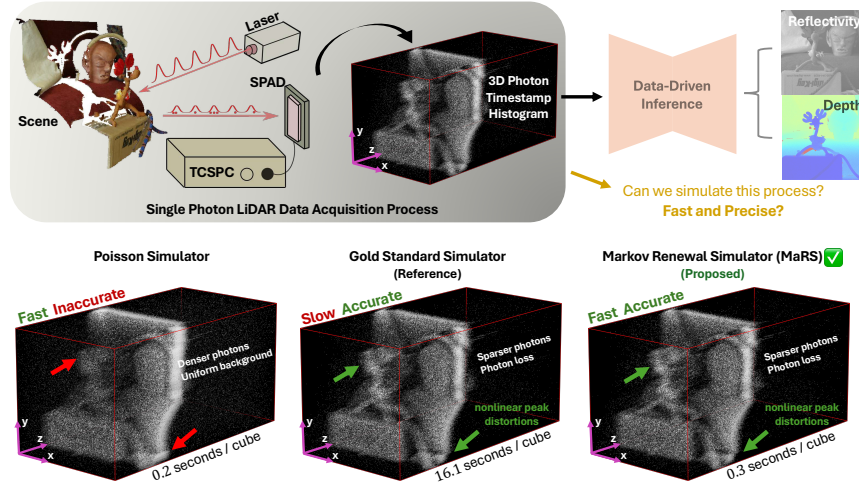


Fig. 1: The SP-LiDAR Simulation Dilemma: Speed *vs.* Accuracy. Hardware dead time in the physical acquisition process (*top*) forces a trade-off in simulation. (*Left*) The Poisson model is fast (0.2 s) but physically wrong, as it ignores detector dead time. Background photons are uniform. (*Center*) The sequential “gold-standard” models dead time and produces physically correct nonlinear distortions, but it is computationally slow (16.1 s). (*Right*) Our Markov-renewal simulator analytically models dead time, achieving the same high fidelity as the gold-standard at a practical compute cost (0.3 s). Our framework breaks the compromise, enabling fast and physically correct simulation to develop data-driven downstream solutions. *Best viewed in zoom.*

Abstract. Single-photon LiDAR (SP-LiDAR) simulators face a dilemma: fast but inaccurate Poisson models or accurate but prohibitively slow sequential models. This paper breaks that compromise. We present a simulator that achieves both fidelity and speed by focusing on the critical, yet overlooked, component of simulation: the photon count statistics. Our key contribution is a Markov-renewal process (MRP) formulation that, for the first time, analytically predicts the distribution of registered photon counts under dead time. To make this MRP model computationally tractable, we introduce a spectral truncation rule that efficiently computes the complex covariance statistics. By proving the shift-invariance of the process, we extend this per-pixel model to full histogram cube generation via a precomputed lookup table. Our method generates 3D

cubes indistinguishable from the sequential gold standard, yet is orders of magnitude faster. This finally enables large-scale, physically faithful data generation for learning-based SP-LiDAR reconstruction. Project page: <https://mars-splidar.github.io/>.

Keywords: Single-Photon LiDAR · Dead Time · Markov-Renewal Process · Physics-Based Simulator

1 Introduction

Single-photon LiDAR (SP-LiDAR) is a key technology for high-resolution 3D imaging [7, 11, 14, 15, 18, 23, 25, 35, 36]. Its single-photon sensitivity, however, introduces complex measurement statistics. Photon detections follow *stochastic point-process dynamics* [26] that are fundamentally coupled to detector state. Specifically, single-photon avalanche diodes (SPADs) enter *dead time* after each registration, ignoring subsequent photons. This effect is non-linear: it destroys the statistical independence of detections and creates complex temporal dependencies. The result is a fundamental distortion of both the photon count and the temporal probability density function (PDF), invalidating simple Poisson models [39]. Any simulator that ignores this is physically wrong.

Existing simulators face a practical speed–accuracy trade-off. On one hand, Poisson-multinomial models are fast but physically incorrect, as they ignore dead time. On the other hand, sequential “gold-standard” simulators are physically correct, as they brute-force the rejection sampling of photons during dead time. This sequential approach is computationally intractable, requiring per-photon loops that do not scale in flux, spatial resolutions, or realizations. It is a bottleneck that has prevented the generation of large-scale, physically accurate datasets required to train modern deep learning pipelines for SP-LiDAR.

This paper resolves the speed-vs-accuracy conflict. We present the **Markov Renewal SP-LiDAR Simulator (MaRS)**. The core idea is to *decouple* the simulation problem: the temporal PDF is well-studied, but the primary bottleneck is the *photon count*, which prior work has failed to model effectively. We theoretically show that by modeling the count statistics as a *Markov-renewal process* (MRP) [4, 42], we can preserve tractability while rigorously accounting for the state-dependent correlations caused by dead time. This MRP formulation is the key, providing an analytic path to the count statistics that bypasses the intractable sequential simulation. Our contributions are twofold:

- A new theoretical model of the photon count statistics under dead time using a Markov-renewal process (MRP).
- A new, high-throughput simulator (MaRS) that implements this model using spectral truncation and pixel parallelization.

As shown in Fig. 1, MaRS produces physically correct histogram cubes at a computation cost on par with the physically incorrect Poisson model, and orders of magnitude faster than the gold standard. It is the first SP-LiDAR simulator that provides the accuracy of a sequential model at the speed of an analytic one.

2 Related Work

Existing Simulators. Prior work has modeled SPAD device physics and temporal response [5, 9, 21, 26, 38]. For dead-time-aware SP-LiDAR simulation, the *gold standard* or *conventional simulator* first generates photon arrivals using multinomial Poisson sampling, and then sequentially rejects detections falling within the dead-time window [28, 40]. While physically accurate, this sequential rejection requires per-photon, per-pixel iteration and is computationally prohibitive for megapixel arrays or large-scale data generation.

Zhang *et al.* recently attempted to address this bottleneck by directly modeling dead-time-distorted statistics, incorporating both the temporal PDF and photon count [39]. However, their autoencoder-based estimation of the temporal PDF must be retrained for each dead-time setting, their count prediction exhibits limited accuracy in capturing variance, and the simulation pipeline remains pixel-wise rather than fully parallel. In contrast, our simulator provides a unified and flexible framework that enables fully parallel synthesis of entire histogram cubes with improved statistical fidelity.

Statistical Modeling in Presence of Dead Time. Faithful histogram simulation demands two things: the temporal PDF and the total photon count.

Temporal PDF. Synchronous systems are well modeled due to their independent increments [6, 8, 20, 24]. The real challenge is *asynchronous* operation [12, 17], where stochastic arrivals and dead time create complex, dependent increments. Prior work addressed this by using Markov chains to model relative registration states [28, 29], later accelerated in [41]. This line of work is useful, but it *only* predicts the temporal PDF. It offers no insight into *how many* photons are registered, making it an incomplete solution for a full simulator.

Photon count. This remains an open research problem. Modeling the count requires tracking detections in *absolute* time. The aforementioned Markov chain framework is the wrong tool for this job; it captures relative transitions, discarding the absolute timing needed to count events. Other frameworks are no better: self-exciting point processes are computationally intractable [32], and standard renewal theory [6, 37] breaks down because our inter-registration times are dependent, not independent and identically distributed (i.i.d.). We bridge this gap with a *Markov-renewal process* (MRP). The MRP formulation is the first to preserve renewal-like tractability while correctly handling the state-dependent correlations. This is the machinery required to derive the statistics of the absolute registration times and, by inversion, solve for the photon count.

Utility for Downstream Tasks. Learning-based SP-LiDAR methods have shown strong performance for depth and reflectivity recovery [1, 16, 22, 33, 34]. However, existing training pipelines rely on low-flux or simplified simulators that do not capture dead-time statistics. MaRS fills this gap by enabling efficient, training-scale histogram simulation in the high-flux dead-time regime.

High-fidelity simulation matters because reconstruction inherits the statistics of its training data. While prior models provide accurate temporal distributions, inaccurate photon-count statistics can still change the radiometric scale of the simulated histograms. This directly biases reflectivity estimation and can also

degrade depth recovery by altering histogram reliability. We therefore use depth and reflectivity estimation to isolate the downstream effect of count modeling. Networks trained with MaRS transfer more faithfully to gold-standard simulations than those trained with inaccurate count statistics, showing that dead-time-correct photon counts are not only a distributional improvement but also important for learning-based SP-LiDAR reconstruction.

3 MaRS: Design Philosophy

The goal of MaRS is to provide a unified simulation framework that remains both physically accurate and computationally efficient. Rather than simulating every photon registration individually, MaRS formulates the SP-LiDAR measurement process through analytical and statistical models that explicitly account for detector dead time and temporal correlations.

3.1 Dead-Time-Distorted Forward Model

The measurement process in SP-LiDAR can be viewed as two sequential stages: photons first *arrive* at the detector and are then *registered* by the sensor and electronics, forming a histogram over one laser repetition period.

Photon arrivals. Following prior works [2,3,13,27,31], we model photon arrivals as an inhomogeneous Poisson process with a per-cycle rate

$$\lambda(t) = S s(t - \tau) + \lambda_b, \quad 0 \leq t < t_r, \quad (1)$$

where $s(t)$ is the emitted laser pulse, S is the signal level that includes object reflectivity and laser energy, τ is the depth-induced delay, and λ_b is the background rate [19,30]. The transmitted pulse is modeled as a Gaussian $s(t) = \mathcal{N}(t; 0, \sigma_t^2)$. We assume that N_r pulses are emitted, with a repetition period t_r , resulting in a total exposure time of $N_r t_r$. The background level is defined as $B = t_r \lambda_b$. The total photon flux per period is $A = S + B$, and the signal-to-background ratio (SBR) is $\text{SBR} = S/B$. See the Supplement for a more general forward model.

Registration under dead time. A SPAD introduces a dead time t_d after each detection, during which incoming photons are lost. In asynchronous operation, the detector becomes active again immediately after t_d . The TCSPC module records timestamps $X_k = T_k \bmod t_r$ to one of its n_b bins, where T_k are the absolute detection times. When operating under dead time, the recorded timestamps deviate from the true arrival distribution: temporal correlations arise, the photon count rate decreases, and the histogram becomes distorted.

Input parameters. We define the complete system configuration as $\boldsymbol{\theta} = (t_r, N_r, \sigma_t, t_d, n_b, \tau, S, B, N_{\text{iter}})$, where N_{iter} represents the number of histogram realizations to be generated. In the presence of dead time, the histogram becomes a nonlinear function of $\boldsymbol{\theta}$. Since closed-form models for the joint distribution of all histogram bins are typically unavailable under free-running, inhomogeneous conditions, a decoupled modeling strategy is adopted in the following section.

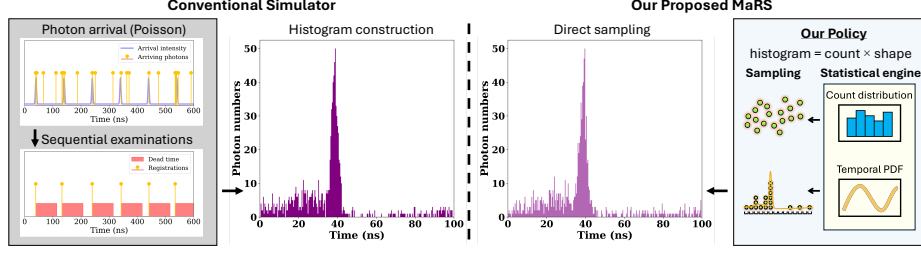


Fig. 2: MaRS: Design Philosophy. (left) Conventional high-flux SP-LiDAR simulation requires Poisson sampling, sequential dead-time enforcement, and timestamp binning, which is physically accurate but slow and hard to scale. (right) MaRS models a histogram as a product of *count* and *shape*. A statistical engine accurately predicts both the photon-count distribution and temporal PDF, enabling *one-shot sampling*.

3.2 Histogram Synthesis Decoupling

Unlike the conventional two-step simulation, we decouple the histogram synthesis process into two components: photon count and temporal shape, as illustrated in Fig. 2. Specifically, we first estimate the distorted photon count and the corresponding temporal probability density function (PDF), and then directly sample histograms based on these statistics.

The conventional simulator iterates through all photons and realizations sequentially to produce different histograms for the same scene. In contrast, the proposed simulation framework relies on statistical modeling. Once the model is established, it can be reused to generate multiple histograms in parallel.

Stationary Temporal PDF. The Markov chain model proposed by Rapp *et al.* has been shown to accurately describe the stationary temporal PDF of photon registrations [28, 29]. It is a discrete approximation that divides the continuous repetition period $[0, t_r)$ into a set of states $\mathcal{S} = \{s_1, \dots, s_{n_b}\}$, where n_b is the number of histogram bins. At its core is a transition matrix $\tilde{\mathbf{P}}$ that models the probability of transitions between the current and next states. Recently, Zhang *et al.* reformulated the large transition matrix by separating the role of dead time as row shifting [41]. As a result, the matrix $\tilde{\mathbf{P}}$ can be efficiently built, and the temporal PDF can be obtained by solving the leading left eigenvector $\boldsymbol{\pi} \tilde{\mathbf{P}} = \boldsymbol{\pi}$.

Photon count: an open problem. Although the solution to the temporal PDF is available, the photon count modeling is still an open problem. Without accurately estimating it, the scale of the generated histogram fail to reflect the true statistics, introducing bias in sensor characterization or downstream neural network training. We solve this problem in Sec. 4.

4 Markov-Renewal Photon Count Prediction

The stationary temporal PDF describes *where* timestamps are across histogram bins, but it does *not* specify *how many* photons are registered. The total photon

count $N(t)$ is a random variable driven by inhomogeneous arrivals and detector dead time. Classical Poisson or renewal models break down because inter-registration times are state-dependent and temporally correlated.

In this section, we develop a probabilistic model for the number of photon registrations. We first identify the limitations of the Markov chain formulation and bridge the gap with Markov-renewal theory. Then, we show that the photon count asymptotically converges to a Gaussian distribution. Finally, we derive closed-form expressions for its mean and variance.

4.1 From Markov Chain to Markov Renewal

The photon registration process is a non-decreasing counting process $\{N(t) : t \geq 0\}$ where $N(t)$ is the number of photon registrations up to time t . Let T_k denote the *absolute* timestamp of the k th registration, and W_k the inter-registration time between the $(k-1)$ th and the k th photons: $W_k = T_k - T_{k-1}$ for $k \geq 2$, and $W_1 = T_1$. The *relative* timestamp of the k th registration within one laser period t_r is $X_k = T_k \bmod t_r$, ignoring quantization alignment to bin centers.

The Markov chain model in Sec. 3.2 captures transitions among the *relative* states X_k and accurately characterizes the stationary temporal PDF. However, it discards the *absolute* timing information contained in T_k and thus cannot predict the total number of events during a given exposure. While it can reproduce histogram *shapes*, it cannot recover the correct histogram *scale* determined by the underlying photon flux and system parameters.

To overcome this limitation, we formulate the registration process as a *Markov-renewal process* (MRP): a stochastic process $\{(X_k, W_k)\}_{k \in \mathbb{N}}$ that jointly models the Markovian state transitions and the random holding times between them on the absolute scale, as shown in Fig. 3 (1).

The *transition kernel* of the MRP is a joint probability that the process jumps to state j within time t , conditioned on state i :

$$Q_{ij}(t) = \Pr(X_{k+1} = j, W_{k+1} \leq t \mid X_k = i) \quad (2)$$

$$\stackrel{\text{def}}{=} \underbrace{\Pr(j \mid i)}_{\tilde{P}_{ij}} \underbrace{\Pr(W_{k+1} \leq t \mid i \rightarrow j)}_{:=F_{ij}(t)}, \quad (3)$$

where $\tilde{P}_{ij}$ is the transition probability mentioned in Sec. 3.2, and $F_{ij}(t)$ is the conditional cumulative density function (CDF) of the holding time for transition $i \rightarrow j$.

Directly modeling the distribution of the photon count $N(t)$ is difficult, because dead time couples successive registrations. The key observation is that $N(t)$ is the inverse process of the cumulative holding time: more frequent registrations correspond to smaller increments W_k . Crucially, the MRP formulation makes this connection explicit by jointly capturing state transitions and holding times. As illustrated in Fig. 3 (2), although the holding times are dependent, the underlying Markov states are strongly mixing, so the cumulative time $T_n = \sum_{k=1}^n W_k$ satisfies a central limit theorem (CTL). Inverting this asymptotic relation then yields a Gaussian approximation for the photon count $N(t)$.

Statistical engine: temporal PDF and count prediction

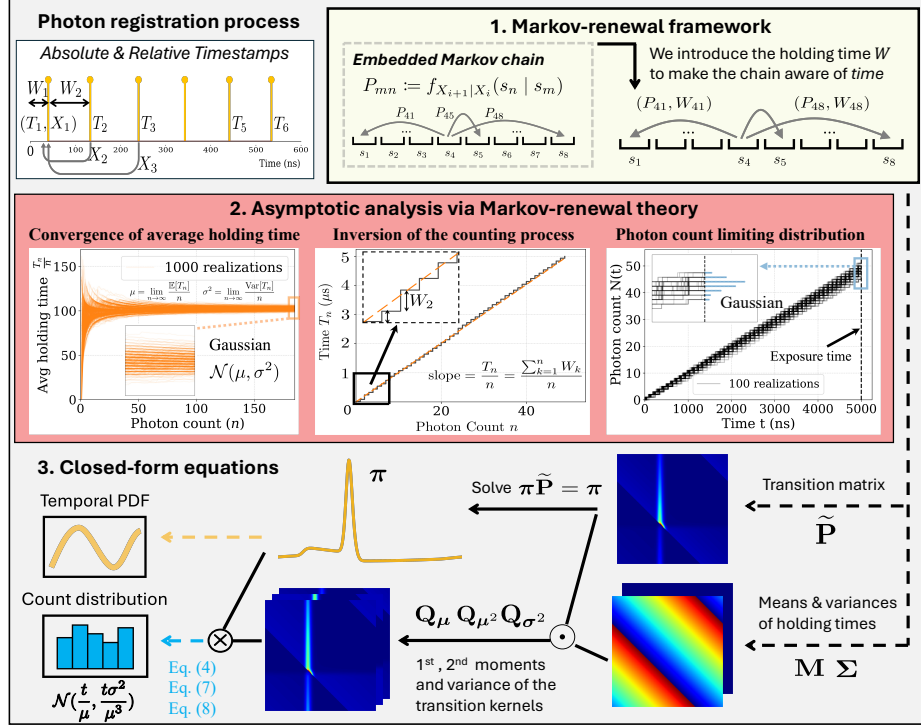


Fig. 3: Statistical Engine Overview. The core engine is built on a *Markov-renewal process* (MRP). (1) The MRP augments the embedded Markov chain with renewal timing on the absolute scale, enabling photon-count prediction besides the temporal PDF. (2) Photon counting is viewed as the inverse of cumulative holding time. The mixing Markov states yield a CLT for $T_n = \sum_{k=1}^n W_k$, whose inversion gives a Gaussian approximation for $N(t)$. (3) Together, these provide analytic prediction of both histogram shape and scale. *Best viewed in zoom.*

Theorem 1 (Convergence of photon count $N(t)$). *The photon count $N(t)$ asymptotically converges in distribution to a Gaussian:*

$$N(t) \xrightarrow{d} \mathcal{N}\left(\frac{t}{\mu}, \frac{t\sigma^2}{\mu^3}\right), \quad t \rightarrow \infty, \quad (4)$$

where the effective mean and variance of the holding times W_k are defined as

$$\mu \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \frac{\mathbb{E}[T_n]}{n} = \lim_{n \rightarrow \infty} \frac{\mathbb{E}[\sum_{k=1}^n W_k]}{n}, \quad (5)$$

$$\sigma^2 \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \frac{\text{Var}[T_n]}{n} = \lim_{n \rightarrow \infty} \frac{\text{Var}[\sum_{k=1}^n W_k]}{n}. \quad (6)$$

(Proof in Supplement; obtained from the CLT for T_n and inversion.)

4.2 Closed-Form Equations for Mean and Variance

Although Theorem 1 establishes the asymptotic law for the photon count $N(t)$, we still require explicit expressions for t , μ , and σ^2 .

The system run time is given by the total exposure time $t = N_r t_r$. Denote by π (row vector) the stationary distribution and by $\tilde{\mathbf{P}}$ the embedded transition matrix (both from Sec. 3.2). To compute μ and σ^2 , we construct from the transition kernels in Eq. (3) the first- and second-moment kernels $\mathbf{Q}_\mu = \tilde{\mathbf{P}} \odot \mathbf{M}$, $\mathbf{Q}_{\mu^2} = \tilde{\mathbf{P}} \odot \mathbf{M} \odot \mathbf{M}$, and $\mathbf{Q}_{\sigma^2} = \tilde{\mathbf{P}} \odot \mathbf{\Sigma}$, where $\odot$ denotes element-wise multiplication and $\mathbf{M} = (\mu_{ij})$, $\mathbf{\Sigma} = (\sigma_{ij}^2)$ contain the transition-dependent mean and variance of the holding times:

$$\mu_{ij} = \mathbb{E}[W \mid i \rightarrow j], \quad \sigma_{ij}^2 = \text{Var}[W \mid i \rightarrow j].$$

Lemma 1. *Conditioned on transition $i \rightarrow j$, the mean and variance of the holding time W are*

$$\mu_{ij} = t_d + t_{i'j} + \frac{e^{-\Lambda}}{1 - e^{-\Lambda}} t_r, \quad \sigma_{ij}^2 = \frac{e^{-\Lambda}}{(1 - e^{-\Lambda})^2} t_r^2,$$

where Λ is the photon arrival flux in one period and $t_{i'j} = (x_j - x_{i'}) \bmod t_r$ is the minimum awake interval between the reactivation time $x_{i'}$, and the next registration x_j .

(Proof in Supplement; holding times combine a deterministic offset with a stochastic number of full repetition periods.)

Lemma 1 provides all quantities needed to compute $\mathbf{M}$ and $\mathbf{\Sigma}$, and thus $\mathbf{Q}_\mu$, $\mathbf{Q}_{\mu^2}$, and $\mathbf{Q}_{\sigma^2}$, enabling direct evaluation of μ and σ^2 .

Theorem 2 (Effective mean). *The effective mean holding time defined in Eq. (5) is*

$$\mu = \pi \mathbf{Q}_\mu \mathbf{1}. \quad (7)$$

(Proof in Supplement; by the law of total expectation.)

Theorem 3 (Effective variance). *The effective variance of holding times defined in Eq. (6) is*

$$\sigma^2 = \sigma_{\text{ss}}^2 + 2\gamma_{\text{ss}}, \quad (8)$$

which includes steady-state variance and covariance terms given by

$$\sigma_{\text{ss}}^2 = \pi(\mathbf{Q}_{\mu^2} + \mathbf{Q}_{\sigma^2})\mathbf{1} - \mu^2, \quad (9)$$

$$\gamma_{\text{ss}} = \sum_{l=1}^{\infty} \gamma_l \approx \sum_{l=1}^c (\pi \mathbf{Q}_\mu \tilde{\mathbf{P}}^{l-1} \mathbf{Q}_\mu \mathbf{1} - \mu^2), \quad (10)$$

respectively, where c denotes a cutoff such that $\gamma_l \approx 0$ for $l > c$. Intuition: As the order l increases, the covariance term $\gamma_l \rightarrow 0$ geometrically with a rate set by the spectral gap of $\tilde{\mathbf{P}}$ (see Lemma 2 in Supplement).

(Proof in Supplement; via variance expansion + mixing.)

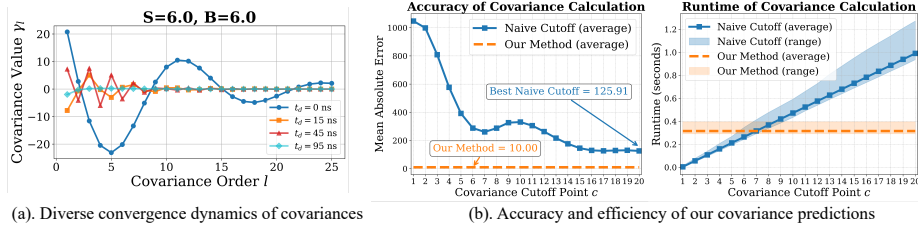


Fig. 4: Spectral Truncation for Long-Lag Covariances. (a) A common cutoff c cannot easily capture all cases. (b) Our spectral truncation rule matches the ground-truth covariance while significantly reducing computation. *Best viewed in zoom.*

5 Accurate and Efficient Histogram Cube Synthesis

To generate histogram cubes accurately and at scale, two challenges must be addressed: modeling the long-lag statistical effects introduced by dead time, and avoiding pixel-by-pixel sequential generation for an image. This section introduces the mechanisms that resolve these issues. We propose spectral modeling of the covariance structure for accuracy and a lookup-table-based algorithm for efficient reuse of per-pixel statistics, enabling fast and faithful cube synthesis.

5.1 Spectral Analysis of Long-Lag Covariances

While closed-form expressions for the variance in Eqs. (8) and (10) require summing the covariance series $\gamma_{ss} = \sum_{l \geq 1} \gamma_l$, the terms $\{\gamma_l\}$ exhibit widely varying decay patterns across different (S, B, t_d) , as shown in Fig. 4 (a) and more in the supplement. This makes a universal cutoff c unreliable: large c is computationally expensive, while small c introduces bias. A data-adaptive rule is thus required.

Motivation. Zhang *et al.* [41] showed that the long-term dynamics of the transition matrix $\tilde{\mathbf{P}}$ are governed by the magnitude and phase of a few dominant eigenvalues. Thus, although the early covariance terms depend on *all* eigenmodes, the long-range correlations are effectively controlled by only the leading ones. We exploit this insight to derive a compact spectral approximation for γ_{ss} .

Low-rank spectral model. Let the leading p eigenpairs of $\tilde{\mathbf{P}}$ be $\{(\lambda_i, \mathbf{v}_i, \mathbf{u}_i)\}_{i=1}^p$ with biorthogonality $\mathbf{U}_p^\top \mathbf{V}_p = \mathbf{I}_p$. We approximate $\tilde{\mathbf{P}}^{l-1} \approx \mathbf{V}_p \mathbf{\Lambda}_p^{l-1} \mathbf{U}_p^\top$, which is accurate for large l because the remaining eigenvalues mix rapidly toward 0. Define $\alpha_i := (\pi \mathbf{Q}_\mu \mathbf{V}_p)_i$ and $\beta_i := (\mathbf{U}_p^\top \mathbf{Q}_\mu \mathbf{1})_i$.

Proposition 1 (Spectral truncation rule). *For a chosen exact cutoff L and spectral rank p , the covariance series γ_{ss} can be approximated as*

$$\hat{\gamma}_{ss}(L, p) = \sum_{l=1}^L \gamma_l + \sum_{i=2}^p \frac{\alpha_i \beta_i \lambda_i^L}{1 - \lambda_i}. \quad (11)$$

The first term accounts for short-range correlations (computed exactly), while the second is a closed-form estimate of the tail using dominant eigenmodes.

Remark. See proof of Eq. (11) in the Supplement. Early covariance terms encode complex local behavior and are inexpensive to compute directly. Long-range correlations, however, are dictated by the slowest eigenmodes of $\tilde{\mathbf{P}}$. The spectral tail in Eq. (11) captures this effect analytically.

Choice of (L, p) . Empirically, across a range of (S, B, t_d) settings, eigenvalues beyond the fifth have magnitude $\lesssim 0.7$, making their contribution negligible after $L \approx 6$ steps. We therefore set $L = 6$ and retain $p = 5$ modes, which balances accuracy and runtime for all tested regimes. See more details in the Supplement.

Validation. Across diverse (S, B, t_d) settings, we compare our spectral truncation rule with deterministic summation of covariance terms. As shown in Fig. 4, in several challenging cases, even including 20 covariance terms fails to capture the long-range tail needed for accurate variance prediction, whereas our spectral rule recovers it precisely with negligible computational overhead. This demonstrates that a few dominant eigenmodes are sufficient to model the full correlation structure. See detailed experiment settings in the supplement.

5.2 Parallel Generation via Lookup Table and Delay Invariance

Even with the spectral rule in Proposition 1, two obstacles remain for a practical *high-throughput* simulator: (i) accurate on-the-fly count statistics are expensive at native temporal resolution (*e.g.*, $n_b = 2^{14}$), and (ii) per-pixel heterogeneity in (S, B, τ) makes any pixel-by-pixel recomputation non-scalable.

Why a naive LUT is infeasible. Since our MRP model fully specifies the per-pixel statistics, a natural idea is to precompute and reuse them via a lookup table (LUT), which is fundamentally faster than the gold-standard simulator (Monte Carlo). However, tabulating all three variables (S, B, τ) is prohibitive. While S and B can be discretized over a bounded range (*e.g.*, indoor flux levels), τ must span the repetition period with fine resolution. For a modest grid with $n_S=256$, $n_B=100$, $n_\tau=1024$, and $n_b=1024$, each entry stores (μ, σ^2) and a length- n_b PDF $\boldsymbol{\pi}$ (*i.e.*, $n_b+2=1026$ floats). The total storage is $256 \cdot 100 \cdot 1024 \cdot 1026 \cdot 4 \approx 108$ GB (32 bits), which is impractical and bandwidth-unfriendly.

Key observation: delay only shifts time. Let the per-pixel arrival rate be $\lambda_\tau(t) = S s(t - \tau) + B/t_r$, where S is the *effective* signal amplitude (absorbing range fall-off and albedo) and B is background. Shifting τ changes only the time origin of the flux. It does not change the renewal structure, hence not the count statistics. We formalize this property below.

Theorem 4 (Delay Invariance). Let μ_τ and σ_τ^2 denote the mean and variance of the registered photon count, and let $\boldsymbol{\pi}_\tau$ denote the temporal PDF, under arrival rate $\lambda_\tau(t) = S s(t - \tau) + B/t_r$. Then for any delays τ, τ' ,

$$\mu_\tau = \mu_{\tau'}, \quad \sigma_\tau^2 = \sigma_{\tau'}^2, \quad \boldsymbol{\pi}_{\tau'} = \text{shift}_{(\tau' - \tau)}(\boldsymbol{\pi}_\tau),$$

(Depth-dependent effects such as distance fall-off act by modifying S (and/or B) and can be accounted for as a preprocessing step before applying Theorem 4. See the Supplement for details and the proof of Theorem 4.)

Algorithm 1: Histogram Cube Generation

Input: System parameters $(t_r, N_r, \sigma_t, t_d, n_b)$, scene parameters $(\mathbf{Z}, \mathbf{R}, B)$, and realization N_{iter} , where $\mathbf{Z}, \mathbf{R} \in \mathbb{R}^{H \times W}$ are depth and reflectivity images.

Output: N_{iter} histogram cubes $\mathbf{H}_i \in \mathbb{R}^{H \times W \times n_b}$.

Step 1: Batched lookup. Retrieve stationary distributions and MRP parameters: $(\boldsymbol{\Pi}_0, \boldsymbol{\mu}_{\text{MRP}}, \boldsymbol{\sigma}_{\text{MRP}}^2) = \text{Lookup}(\mathbf{R}, B)$.

Step 2: Temporal alignment. Shift PDFs to account for scene-dependent delays: $\boldsymbol{\Pi} = \text{Shift}(\boldsymbol{\Pi}_0, \mathbf{Z} - t_r/2)$.

Step 3: Histogram cube sampling. Draw N_{iter} i.i.d. photon count maps: $\mathbf{n}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\text{MRP}}, \boldsymbol{\sigma}_{\text{MRP}}^2)$. Sample histogram cubes via $\mathbf{H}_i \sim \text{Multinomial}(\mathbf{n}_i, \boldsymbol{\Pi})$.

Consequence: remove τ from the LUT. By Theorem 4, we tabulate only over (S, B) at a canonical delay (pulse centered at $t_r/2$):

$$\mathcal{T}(S, B) = (\mu(S, B), \sigma^2(S, B), \boldsymbol{\pi}_0(S, B)).$$

The memory drops to $256 \cdot 100 \cdot 1026 \cdot 4 \approx 105$ MB (32 bits), making batched access feasible and cache-friendly.

Lookup-based parallel cube synthesis. Each pixel retrieves its (S, B) entry from the lookup table, applies the appropriate temporal shift determined by τ , and samples its histogram directly in a parallel fashion, as shown in Algorithm 1. This yields full-cube generation via vectorized memory reads plus cheap shifts and sampling, which is orders of magnitude faster than sequential rejection sampling, while preserving dead-time fidelity through the MRP model.

Remark. The LUT is *not* universal: it is generated for a given system configuration $(t_r, n_b, \sigma_t, t_d)$ and pulse shape $s(\cdot)$, and can be re-generated for different operating regimes, *i.e.*, different feasible ranges and discretizations of (S, B) .

6 Experiments

We evaluate the proposed simulator along three key dimensions: accuracy, efficiency, and practical utility. We first verify its ability to reproduce photon count statistics under dead time, then measure runtime scaling with flux, realizations, and pixels, and finally assess its impact by training depth and reflectivity reconstruction networks using data generated by our model.

6.1 Accuracy of Photon Count Prediction

We first assess the accuracy of our photon count model by comparing its predicted distribution to the empirical distribution obtained from the conventional sequential simulator, treated as ground truth. Baselines include: (i) a renewal approximation that replaces the inhomogeneous flux with a constant-rate model of equal per-cycle energy, and (ii) Zhang et al. [39].

We evaluate a grid of (S, B, t_d) values and summarize qualitative results in Fig. 5 and quantitative metrics in Tab. 1. Figure 5 shows that the renewal

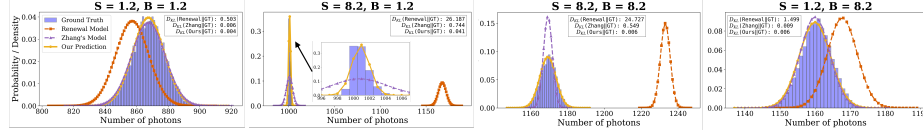


Fig. 5: Qualitative Count Prediction Accuracy. We show comparison of photon count distributions under representative (S, B) combinations with $t_d = 75$ ns. MaRS closely matches the empirical ground truth in both shape and spread, whereas the renewal model and Zhang *et al.* [39] exhibit noticeable deviations. *Best viewed in zoom.*

Table 1: Quantitative Accuracy of Photon Count Prediction. Metrics are averaged over a grid of (S, B, t_d) values. See details in the supplement.

Methods	Distribution Metrics		Moment Metrics	
	Wass Dist ($\downarrow$)	KL Div ($\downarrow$)	Mean Diff ($\downarrow$)	Var Diff ($\downarrow$)
Poisson	7966.130	24.208	7966.130	9257.464
Renewal	207.565	18.821	207.564	45.927
Zhang [39]	2.337	0.409	0.282	64.558
MaRS	0.309	0.013	0.269	1.588

model struggles to capture both the mean and variance, Zhang’s model provides accurate mean estimates but mismatches variance under high signal levels, while MaRS achieves close matches robustly in various SBRs. Table 1 reflects the same results under broader (S, B, t_d) cases and evaluation metrics.

Real-data validation. We further validate the proposed count model on real SPAD measurements. We use a 7.5 MHz pulsed laser and a SPAD with 85 ns dead time, record the number of detections over 9975 laser cycles as one realization, and repeat this process for 10000 realizations. Following the calibration procedure of Rapp *et al.* [29], we estimate the signal and background levels and predict the photon-count distribution. As shown in Fig. 6, MaRS closely matches the empirical count distributions for both bright and dark targets under fixed distance and ambient illumination. This provides direct hardware evidence that the proposed MRP count model captures real dead-time count statistics, beyond agreement with the sequential gold-standard simulator.

6.2 Efficiency of MaRS

We benchmark ¹ the efficiency of MaRS by measuring runtime scaling with flux level, number of realizations, and number of pixels, as shown in Fig. 7.

For flux and realization scaling, we use the single-pixel simulator and compare three variants: the conventional sequential simulator, MaRS without the lookup table (requiring large matrix computations on the fly), and MaRS with the lookup table. For pixel scaling, we compare MaRS with lookup against the conventional simulator. The lookup table specification is in Tab. 2.

¹ All experiments are conducted in NumPy on a CPU machine equipped with an AMD Ryzen™ 7 9700X (8 cores / 16 threads) and 32GB DDR5 RAM.

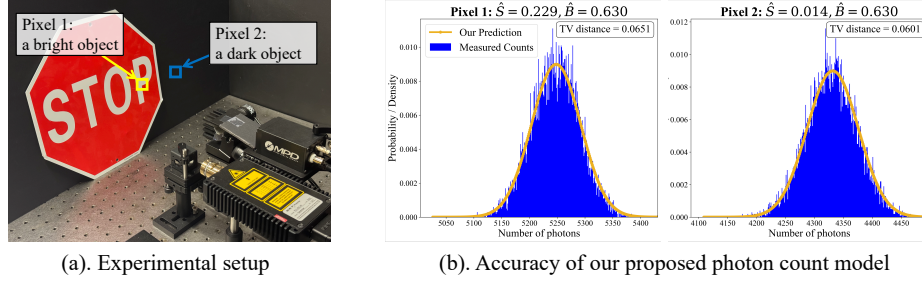


Fig. 6: Real-data validation of photon-count statistics. Photon-count distributions measured from real hardware closely match our predictions. *Best viewed in zoom.*

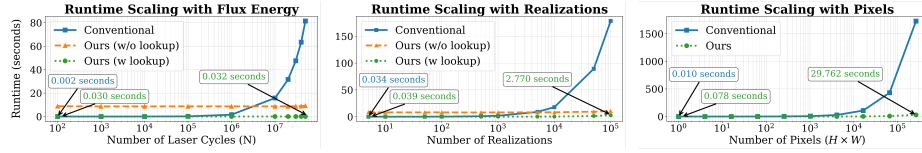


Fig. 7: Simulator Efficiency Comparison Analysis. We compute the scaling of simulation runtime with flux, realizations, and pixels. MaRS with the lookup table maintains near-constant runtime across all settings, while MaRS without lookup and the conventional simulator slow down significantly, demonstrating the benefit of the lookup-table design. *Best viewed in zoom.*

Across all experiments, MaRS with the lookup table exhibits near-constant runtime, as all statistics are retrieved from precomputed entries regardless of flux, realizations, or pixel count. MaRS without the lookup table degrades due to repeated high-resolution matrix operations, and the conventional sequential simulator scales poorly in all cases. These results confirm the effectiveness of the lookup-table design in enabling fast, scalable histogram cube synthesis.

6.3 Downstream Task: Depth and Reflectivity Estimation

A key goal of this work is to validate the utility of MaRS to train neural networks for downstream tasks. We investigate the role of an accurate photon count prediction on depth and reflectivity recovery. Besides, we benchmark the runtime of MaRS in training tasks.

Implementation. We use a DDPM U-Net [10], modified to ingest the full spatio-temporal histogram cube (details in the supplementary). The same architec-

Table 2: Lookup table specification for experiments in Secs. 6.2 and 6.3.

Parameter	S range (# points)	B range (# points)	Temporal bins n_b	Memory
Value	[0.01, 10] (256)	[0.01, 10] (100)	1024	100.4 MB

Table 3: Comparison of simulator properties. Our contributions are highlighted.

Simulator	Dead-time-correct statistics			Use of LUT
	μ	σ^2	π	
Poisson	✗	✗	✗	✗
Renewal	✗	✗	✓	✓
Zhang et al. [39]	✓	✗	✓	✓
MaRS	✓	✓	✓	✓

Table 4: Runtime breakdown: generating a $128 \times 128 \times 1024$ cube.

Stage	Time (s)
Batched lookup	0.020 ± 0.005
Temporal alignment	0.144 ± 0.008
Histogram sampling	0.849 ± 0.070
Total (MaRS)	1.013 ± 0.071
Gold standard (GS)	40.831 ± 8.045

Table 5: Depth and Reflectivity Estimation Results Across Simulators. The same network architecture is trained on data generated by the baseline and the proposed MaRS simulators for depth and reflectivity estimation separately until convergence. Pink rows correspond to the *same-simulator* evaluation, while blue rows denote testing on the GS simulator (*cross-simulator* setup). MaRS shows the strongest generalization to the GS domain, demonstrating the benefit of our accurate MRP model for unbiased recovery and the LUT-based simulator that enables efficient training.

Training Simulator	SBR = 0.6						SBR = 1.0					
	Reflectivity			Depth (m)			Reflectivity			Depth (m)		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	RMSE $\downarrow$	MAE $\downarrow$		PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	RMSE $\downarrow$	MAE $\downarrow$	
Poisson	36.4946	0.9668	0.0225	0.0722	0.0516		33.6827	0.9616	0.0231	0.0606	0.0468	
	7.0237	-0.0753	0.5160	1.6537	1.4517		7.2490	-0.0145	0.5056	0.7853	0.6448	
Renewal	29.7230	0.9192	0.0544	0.0543	0.0397		30.9704	0.9329	0.0413	0.0514	0.0388	
	12.1210	0.6033	0.3030	0.0959	0.0637		11.6988	0.6279	0.2903	0.0967	0.0646	
Zhang et al. [39]	22.7444	0.8908	0.1220	0.0526	0.0385		21.9506	0.8820	0.1128	0.0506	0.0382	
	22.5421	0.8931	0.1010	0.0820	0.0526		22.2137	0.9036	0.0894	0.1193	0.0620	
MaRS	24.0321	0.8893	0.1315	0.0532	0.0400		24.2821	0.8946	0.1063	0.0553	0.0426	
	24.2488	0.9084	0.1159	0.0739	0.0471		23.5776	0.9113	0.0899	0.0949	0.0521	

ture is trained to recover depth and reflectivity maps separately. We consider the same lookup table specifications in Tab. 2.

Training and baselines. We train four versions of the same U-Net under different simulators: (1) Poisson, (2) Renewal, (3) Zhang et al. [39], and (4) the proposed **MaRS**. Note that we accompany Renewal and Zhang et al. [39] with the accurate temporal PDF model from Rapp *et al.* [28], but they cannot work efficiently without our LUT approach, as shown in Tab. 3.

Evaluation. We test each model on data from its own simulator to confirm stable learning, then evaluate all models on the **Gold-Standard (GS)** simulator, which best reflects real sensor behavior.

Results. Table 4 quantifies MaRS’s efficiency against the prohibitively slow GS simulator. Table 5 and Figure 8 show that **MaRS** yields the most accurate depth and reflectivity estimates, and its minimal validation-to-test difference indicates that it matches detector statistics most faithfully. Although scenes may not exactly match the LUT’s discrete (S, B, τ) grid, the resulting quantization error is minor and is readily absorbed by the reconstruction network.

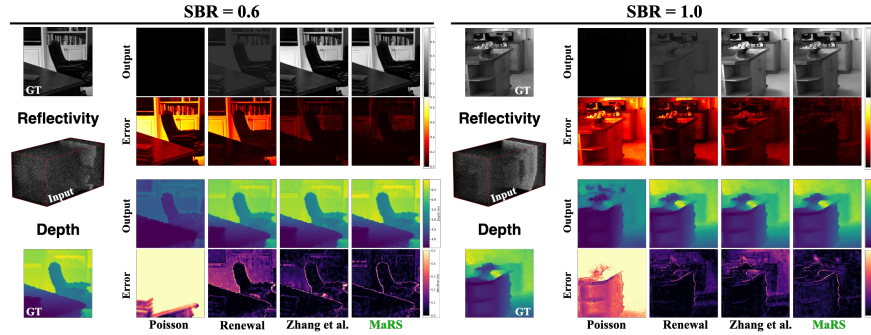


Fig. 8: Depth and Reflectivity Estimation Visual Results Across Simulators. We compare the visual quality of models trained using baseline simulators and the proposed MaRS, evaluated on data generated by the GS simulator. *Best viewed in zoom.*

7 Conclusion

SP-LiDAR simulation has been stuck between slow-but-accurate sequential models and fast-but-wrong Poisson approximations under dead time. We remove this trade-off with MaRS: an MRP-based simulator that (i) analytically models dead-time photon statistics, (ii) accurately predicts challenging covariances via spectral truncation, and (iii) synthesizes full histogram cubes without sequential sampling. MaRS matches gold-standard statistics while running $> 40\times$ faster for a $128\times128\times1024$ cube, enabling fast and physically correct data generation for learning-based SP-LiDAR.

Acknowledgements

The work is supported, in part, by the DARPA / SRC CogniSense JUMP 2.0 Center, NSF IIS-2133032, and NSF ECCS-2030570.

References

1. Altmann, Y., McLaughlin, S., Davies, M.E.: Fast Online 3D Reconstruction of Dynamic Scenes From Individual Single-Photon Detection Events. *IEEE Transactions on Image Processing (TIP)* **29**, 2666–2675 (2020)
2. Bar-David, I.: Communication under the Poisson Regime. *IEEE Transactions on Information Theory* **15**(1), 31–37 (1969)
3. Chan, S.H., Weerasooriya, H.K., Zhang, W., Abshire, P., Gyongy, I., Henderson, R.K.: Resolution Limit of Single-Photon LiDAR. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 25307–25316 (2024)
4. Fox, B.: *Semi-Markov Processes: A Primer* (1968)

5. Gulinatti, A., Rech, I., Assanelli, M., Ghioni, M., Cova, S.: A physically based model for evaluating the photon detection efficiency and the temporal response of SPAD detectors. *Journal of Modern Optics* **58**(3-4), 210–224 (2011)
6. Gupta, A., Ingle, A., Gupta, M.: Asynchronous Single-Photon 3D Imaging. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7908–7917 (2019)
7. Gyongy, I., Dutton, N.A.W., Henderson, R.K.: Direct Time-of-Flight Single-Photon Imaging. *IEEE Transactions on Electron Devices* **69**(6), 2794–2805 (2022)
8. Heide, F., Diamond, S., Lindell, D.B., Wetzstein, G.: Sub-picosecond Photon-Efficient 3D Imaging using Single-Photon Sensors. *Scientific reports* **8**(1), 17726 (2018)
9. Hernandez, Q., Gutierrez, D., Jarabo, A.: A computational model of a single-photon avalanche diode sensor for transient imaging. *arXiv preprint arXiv:1703.02635* (2017)
10. Ho, J., Jain, A., Abbeel, P.: Denoising Diffusion Probabilistic Models. In: *Advances in Neural Information Processing Systems*. vol. 33, pp. 6840–6851 (2020)
11. Jiang, P.Y., Li, Z.P., Ye, W.L., Hong, Y., Dai, C., Huang, X., Xi, S.Q., Lu, J., Cui, D.J., Cao, Y., Xu, F., Pan, J.W.: Long Range 3D Imaging through Atmospheric Obscurants using Array-based Single-Photon LiDAR. *Opt. Express* **31**(10), 16054–16066 (2023)
12. Kitichotkul, R., Bharadwaj, S., Rapp, J., Ma, Y., Mehta, A., Goyal, V.K.: Free-running vs. Synchronous: Single-Photon Lidar for High-flux 3D Imaging. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2025)
13. Kitichotkul, R., Rapp, J., Goyal, V.K.: The Role of Detection Times in Reflectivity Estimation with Single-Photon LiDAR. *IEEE Journal of Selected Topics in Quantum Electronics* **30**(1), 1–14 (2024)
14. Kitichotkul, R., Rapp, J., Ma, Y., Mansour, H.: Simultaneous range and velocity measurement with Doppler single-photon lidar. *Optica* **12**(5), 604–613 (May 2025)
15. Li, Z.P., Ye, J.T., Huang, X., Jiang, P.Y., Cao, Y., Hong, Y., Yu, C., Zhang, J., Zhang, Q., Peng, C.Z., Xu, F., Pan, J.W.: Single-Photon Imaging over 200km. *Optica* **8**(3), 344–349 (2021)
16. Lindell, D.B., O’Toole, M., Wetzstein, G.: Single-Photon 3D Imaging with Deep Sensor Fusion. *ACM Trans. Graph.* **37**(4), 113.1–113.12 (2018)
17. Lyu, L., Li, D., Wu, T., Mi, Q., Yang, Y., Zhou, W., Xu, L.: High-Counting-Rate Single-Photon LiDAR With Free-Running InGaAs/InP SPAD. *IEEE Transactions on Instrumentation and Measurement* **74**, 1–12 (2025)
18. McCarthy, A., Taylor, G.G., Garcia-Armenta, J., Korzh, B., Morozov, D.V., Beyer, A.D., Briggs, R.M., Allmaras, J.P., Bumble, B., Colangelo, M., Zhu, D., Berggren, K.K., Shaw, M.D., Hadfield, R.H., Buller, G.S.: High-Resolution Long-Distance Depth Imaging LiDAR with Ultra-Low Timing Jitter Superconducting Nanowire Single-photon Detectors. *Optica* **12**(2), 168–177 (2025)
19. Padmanabhan, P., Zhang, C., Charbon, E.: Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications. *MDPI, Sensors* **19**(24), 5464 (2019)
20. Pediredla, A.K., Sankaranarayanan, A.C., Buttafava, M., Tosi, A., Veeraraghavan, A.: Signal Processing Based Pile-up Compensation for Gated Single-Photon Avalanche Diodes. *arXiv preprint arXiv: 1806.07437* (2018)
21. Peña, M., Gutierrez, D., Marco, J.: Looking around flatland: End-to-end 2D real-time NLOS imaging. *IEEE Transactions on Computational Imaging* **11**, 189–200 (2025)

22. Peng, J., Xiong, Z., Huang, X., Li, Z.P., Liu, D., Xu, F.: Photon-Efficient 3D Imaging with A Non-local Neural Network. In: European Conference on Computer Vision (ECCV). pp. 225–241 (2020)
23. Piron, F., Morrison, D., Yuce, M.R., Redouté, J.M.: A Review of Single-Photon Avalanche Diode Time-of-Flight Imaging Sensor Arrays. *IEEE Sensors Journal* **21**(11), 12654–12666 (2021)
24. Po, R., Pediredla, A., Gkioulekas, I.: Adaptive Gating for Single-Photon 3D Imaging. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16333–16342 (2022)
25. Qian, X., Jiang, W., Deen, M.J.: Single Photon Detectors for Automotive LiDAR Applications: State-of-the-Art and Research Challenges. *IEEE Journal of Selected Topics in Quantum Electronics* **30**, 1–20 (2024)
26. Qian, X., Jiang, W., Elsharabasy, A., Deen, M.J.: Modeling for Single-Photon Avalanche Diodes: State-of-the-Art and Research Challenges. *Sensors* **23**(7) (2023)
27. Rapp, J., Goyal, V.K.: A Few Photons Among Many: Unmixing Signal and Noise for Photon-Efficient Active Imaging. *IEEE Transactions on Computational Imaging (TCI)* **3**(3), 445–459 (2017)
28. Rapp, J., Ma, Y., Dawson, R.M.A., Goyal, V.K.: Dead Time Compensation for High-Flux Ranging. *IEEE Transactions on Signal Processing* **67**(13), 3471–3486 (2019)
29. Rapp, J., Ma, Y., Dawson, R.M., Goyal, V.K.: High-flux Single-Photon LiDAR. *Optica* **8**(1), 30–39 (2021)
30. Scholes, S., Mora-Martin, G., Zhu, F., Gyongy, I., Soan, P., Leach, J.: Fundamental Limits to Depth Imaging with Single-photon Detector Array Sensors. *Nature Scientific Reports* **13**(1), 176 (2023)
31. Shin, D., Kirmani, A., Goyal, V.K., Shapiro, J.H.: Photon-Efficient Computational 3-D and Reflectivity Imaging With Single-Photon Detectors. *IEEE Transactions on Computational Imaging (TCI)* **1**(2), 112–125 (2015)
32. Snyder, D.L., Miller, M.I.: Random Point Processes in Time and Space. Springer (1991)
33. Sun, Z., Lindell, D.B., Solgaard, O., Wetzstein, G.: SPADnet: Deep RGB-SPAD Sensor Fusion Assisted by Monocular Depth Estimation. *Opt. Express* **28**(10), 14948–14962 (2020)
34. Tachella, J., Altmann, Y., Mellado, N., McCarthy, A., Tobin, R., Buller, G.S., Tourneret, J.Y., McLaughlin, S.: Real-time 3D Reconstruction from Single-photon LiDAR Data using Plug-and-Play Point Cloud Denoisers. *Nature Communications* **10**(1) (2019)
35. Weerasooriya, H.K., Chennuri, P., Zhang, W., Gyongy, I., Chan, S.H.: Joint Depth and Reflectivity Estimation using Single-Photon LiDAR. *arXiv preprint* (2025)
36. Young, A., Batagoda, N.M., Zhang, H., Dave, A., Pediredla, A., Negrut, D., Raskar, R.: Enhancing Autonomous Navigation by Imaging Hidden Objects using Single-Photon LiDAR (2025)
37. Yu, D.F., Fessler, J.A.: Mean and Variance of Single Photon Counting with Dead-time. *Physics in Medicine & Biology* **45**(7), 2043 (2000)
38. Zappa, F., Tosi, A., Dalla Mora, A., Tisa, S.: SPICE modeling of single photon avalanche diodes. *Sensors and Actuators A: Physical* **153**(2), 197–204 (2009)
39. Zhang, W., Weerasooriya, H.K., Chan, S.: Ultrafast High-Flux Single-Photon Lidar Simulator via Neural Mapping. In: IEEE International Conference on Image Processing (ICIP). pp. 2772–2777 (2025)

- 40. Zhang, W., Weerasooriya, H.K., Chennuri, P., Chan, S.H.: Parametric Modeling and Estimation of Photon Registrations for 3D Imaging. In: IEEE 26th International Workshop on Multimedia Signal Processing (MMSP). pp. 1–6 (2024)
- 41. Zhang, W., Weerasooriya, H.K., Chennuri, P., Chan, S.H.: Real-Time Markov Modeling for Single-Photon LiDAR: $1000\times$ Acceleration and Convergence Analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2026)
- 42. Çinlar, E.: Markov Renewal Theory. *Advances in Applied Probability* **1**(2), 123–187 (1969)