Resolution Limit of Single-Photon LiDAR

Stanley H. Chan[†1] Hashan K. Weerasooriya[1] Weijian Zhang[1] Pamela Abshire[2]
Istvan Gyongy[3] Robert K. Henderson[3]
[1] Purdue University, [2] Univ. Maryland College Park, [3] University of Edinburgh

Abstract

Single-photon Light Detection and Ranging (LiDAR) systems are often equipped with an array of detectors for improved spatial resolution and sensing speed. However, given a fixed amount of flux produced by the laser transmitter across the scene, the per-pixel Signal-to-Noise Ratio (SNR) will decrease when more pixels are packed in a unit space. This presents a fundamental trade-off between the spatial resolution of the sensor array and the SNR received at each pixel. Theoretical characterization of this fundamental limit is explored. By deriving the photon arrival statistics and introducing a series of new approximation techniques, the Mean Squared Error (MSE) of the maximum-likelihood estimator of the time delay is derived. The theoretical predictions align well with simulations and real data.

1. Introduction

Single-photon LiDAR has a wide range of applications in navigation and object identification [21, 24–26, 30, 32]. By actively illuminating the scene with a laser pulse of a known shape, we measure the time delays of single photons upon their return, which correspond to the distance of the object [4, 19, 36]. The advancement of photo detectors has significantly improved the resolution of today's LiDAR [8, 15, 17, 33, 39–41]. Moreover, algorithms have shown how to reconstruct both the scene reflectivity and 3D structure [2, 6, 16, 20, 22, 23, 29, 36, 38, 42, 43].

As an imaging device, a photodetector used in LiDAR faces similar problems as any other CCD or CMOS pixels. Packing more pixels into a unit space decreases the SNR because the amount of photon flux seen by each pixel diminishes [12]. This fundamental limit is linked to the stochastic nature of the underlying Poisson arrival process of the photons [11, 37]. Unless noise mitigation schemes are employed [2, 14, 22, 31], there is a trade-off between the number of pixels one can pack in a unit space and the SNR we will observe at each pixel. The situation can be visualized in Fig. 1, where we highlight the phenomenon that if we

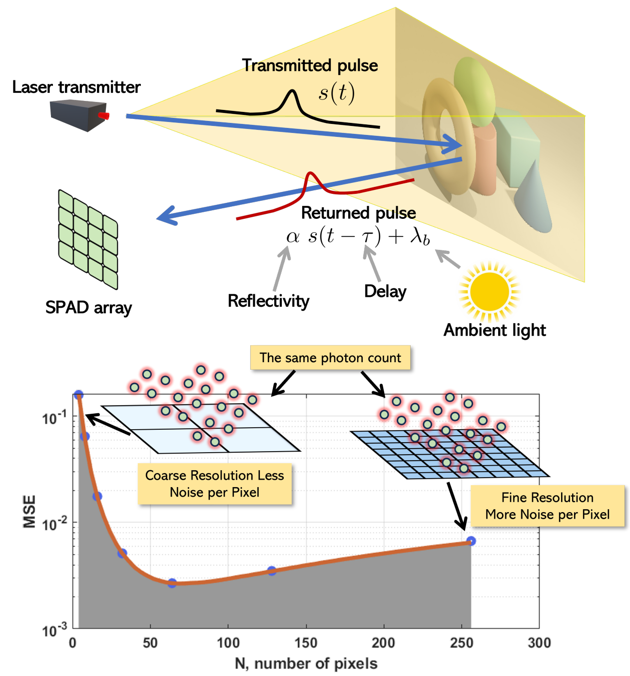

Figure 1. As we pack more pixels in a unit space, we gain the spatial resolution with a reduction in the SNR. The goal of this paper is to understand the trade-off between the two factors.

use many small pixels, the spatial resolution is good but the per pixel noise caused by the random fluctuation of photons will be high. The bias and variance trade-off will then lead to a performance curve that tells us how the accuracy of the depth estimate will behave as we vary the spatial resolution.

The goal of this paper is to rigorously derive the above phenomenon. In particular, we want to answer the following question:

> Can we theoretically derive, ideally in closed-form, the mean squared error of the LiDAR depth estimate as a function of the number of pixels per unit space?

The theoretical analysis presented in this paper is unique from several perspectives:

- **Beyond Single Pixel**. The majority of the computer vision papers in single-photon LiDAR are algorithmic. Few papers have theoretical derivations, but they all focus on a single pixel [2, 14, 22, 31], of which the foundation can be traced back to the original work of Bar-David (1969) [3]. Our paper departs from these results by generalizing the mean square estimation to an array of pixels.

- **New Proof Techniques**. A brute force derivation of the mean squared error is notoriously difficult. We overcome the hurdles by introducing a series of new theoretical approximation techniques in terms of modeling depth, approximating pixels, and utilizing convolutions.
- **Closed-form Results**. Under appropriate assumptions about the scene and sensors, our result has a simple interpretable *closed-form* expression that provides an excellent match with the practical scenarios in both real-world and simulated experiments.

2. Background: Photon Arrival Statistics

In this section, we discuss the mathematical preliminaries. Our result is based on Bar-David [3] which precedes many of the more recently published work [13, 31, 37]. For notation simplicity, our models are derived in 1D. Moreover, to make the main text concise, proofs of theorems are presented in the supplementary material.

2.1. Pulse Model

Let $c = 3 \times 10^8$ [m/s] be the speed of light, and let $d(x)$ be the distance [m] of the object at coordinate $x \in \mathbb{R}$. Hence, the total time [s] for the pulse to travel forward and then back is $\tau(x) = \frac{2d(x)}{c}$. We assume that $\tau(x)$ is a continuous-space function with a continuous amplitude.

The laser pulse is defined as a symmetric time-invarying function $s(t)$. Given a delay τ, the shifted pulse is $s(t-\tau)$.

Example 1. *If the pulse is Gaussian, then*

$$s(t-\tau) = \underbrace{\frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{(t-\tau)^2}{2\sigma_t^2}\right\}}_{=\mathcal{N}(t\,|\,\tau,\sigma_t^2)}, \tag{1}$$

where σ_t denotes the standard deviation.

For simplicity, we ignore the boundary conditions by assuming that the observation interval $(-T, T)$ is significantly larger than the width of the pulse, i.e., $\sigma_t \ll T$. Moreover, We assume that the delay τ lies well inside the observation interval, and the pulse is normalized so that

$$\int_{-T}^{T} s(t-\tau)\, dt = 1. \tag{2}$$

As the pulse reaches an object and is reflected back to the receiver, the received pulse takes the form of

$$\lambda(t) = \alpha \cdot s(t-\tau) + \lambda_b, \tag{3}$$

In this equation, α denotes the reflectivity of the object. For simplicity, we assume that α is a constant. The constant $\lambda_b \in \mathbb{R}$ denotes the background flux due to ambient light. The energy Q carried by $\lambda(t)$ is measured by

$$Q \stackrel{\text{def}}{=} \int_{-T}^{T} \lambda(t)\, dt = \alpha + 2T\lambda_b, \tag{4}$$

Which can be obtained by inserting Eq. (3) into the integrand shown in Eq. (4), and then using Eq. (2) to evaluate the integral.

2.2. Time of Arrival

Given $\lambda(t)$, we assume that M number of time stamps are generated over $[-T, T]$. Denote these time stamps as $\mathbf{t}_M = \{t_j\}_{j=1}^M$, where $-T \le t_1 < t_2 < \ldots < t_M \le T$. The joint distribution of $\mathbf{t}_M$ and M is as follows.

Theorem 1 ([3] Joint distribution of M time stamps)**.** *Let $\mathbf{t}_M = \{t_j\}_{j=1}^M$ such that $-T \le t_1 < t_2 < \ldots < t_M \le T$. For $M \ge 1$,*

$$p(\mathbf{t}_M, M) = e^{-Q} \prod_{j=1}^{M} \lambda(t_j). \tag{5}$$

The number M is a random variable. The probability mass function of M can be computed by marginalizing the joint distribution.

Corollary 1 (Probability of M occurrence)**.** *For any $M \ge 1$, the probability that there are M occurrences is*

$$p(M) = \frac{e^{-Q} Q^M}{M!}. \tag{6}$$

If $M = 0$, then there is no occurrence in $[-T, T]$. In this case, the probability is defined as

$$p(\mathbf{t}_0, 0) = e^{-Q}. \tag{7}$$

Example 2. *Suppose $s(t)$ is a Gaussian pulse and assume that $\lambda_b = 0$ and $\alpha = 1$. Then,*

$$p(\mathbf{t}_M, M) = \frac{e^{-Q}}{(\sqrt{2\pi\sigma_t^2})^M} \exp\left\{-\sum_{j=1}^{M} \frac{(t_j-\tau)^2}{2\sigma_t^2}\right\}.$$

The conditional probability of seeing $\mathbf{t}_M$ given M can be obtained by taking the ratio of the joint distribution $p(\mathbf{t}_M, M)$ and $p(M)$, yielding the following result.

$$p(\mathbf{t}_M \,|\, M) = \frac{p(\mathbf{t}_M, M)}{p(M)} = Q^{-M} M! \prod_{j=1}^{M} \lambda(t_j). \tag{8}$$

The other conditional probability of seeing M given $\mathbf{t}_M$ is 1. Putting these together, we can show that

$$p(\mathbf{t}_M, M) = p(M \,|\, \mathbf{t}_M) p(\mathbf{t}_M) = p(\mathbf{t}_M). \tag{9}$$

We can show that the integration of $p(\mathbf{t}_M, M)$ over the entire sample space is 1:

Corollary 2 (Probability over the sample space)**.**

$$\sum_{M=0}^{\infty} \int_{\Omega_M} p(\mathbf{t}_M, M)\, d\mathbf{t}_M = 1, \tag{10}$$

where $\Omega_M = \{\mathbf{t}_M \,|\, -T \leq t_1 < t_2 < \ldots t_M \leq T\}$.

2.3. Sampling from $p(t_M)$

When the pulse is Gaussian, Monte Carlo simulations of the time stamps can be performed in a two-step process:

- Step 1: Determine the number of samples M. This can be done by recognizing that the total energy of the pulse is $Q = \alpha + 2T\lambda_b$. The total number of samples M is a Poisson random variable such that $M \sim \text{Poisson}(Q)$. However, since the two summands of Q are independent, Raikov theorem states that M can be decomposed into a sum of two independent Poisson random variables. Thus, the number of samples is determined based on

$$M_s \sim \text{Poisson}(\alpha), \qquad M_b \sim \text{Poisson}(2T\lambda_b). \tag{11}$$

  We let $M = M_s + M_b$.
- Step 2: Draw M_s samples from $\mathcal{N}(t \,|\, \tau, \sigma^2)$ and M_b samples from a uniform distribution of a PDF:

$$\begin{aligned} t_j \,|\, M_s &\sim \mathcal{N}(t \,|\, \tau, \sigma^2), \qquad j = 1, \ldots, M_s, \\ t_i \,|\, M_b &\sim \text{Uniform}(-T, T), \qquad i = 1, \ldots, M_b. \end{aligned}$$

  The overall set of samples is $\mathbf{t}_M = \{t_j\}_{j=1}^{M_s} \cup \{t_i\}_{i=1}^{M_b}$.

As we can see, the distribution of the samples is nothing but the shape of the pulse. This is consistent with the literature where we draw a random number representing the height of each histogram bin. In our sampling procedure, we draw the time stamps *without* quantizing them into bins. For pulses of an arbitrary shape, we can perform an inverse CDF technique outlined in the supplementary material.

2.4. Assumptions For Theoretical Analysis

The goal of this paper is to derive *closed-form* results. As such, a series of assumptions are required to minimize the notational burden. Our assumptions are summarized below:

- We do not assume any dark current. In the supplementary material, we have a discussion about the dark current effects.
- We assume that α is a constant. To relax this assumption, we can replace α with $\alpha(x)$ in the proof. However, the final equation will involve an integration over $\alpha(x)$.
- Dead-time and Pile-up [7, 13, 14, 27, 28]. We assume there is no dead-time and hence no pile-up. The empirical analysis in the supplementary material, however, includes a case study that involves pile-up effect.
- Self-excitation process. Prior work such as [31] and [13] use self-excitation process (a variant of the Markov chain) to model the photon arrivals [37]. While this is accurate, deriving closed-form expressions is infeasible. Since we do not assume any dead-time, we follow Bar-David's inhomogeneous Poisson process [3] instead.
- Single-bounce and no multiple path. This is a standard assumption in LiDAR theory.

3. MSE Analysis

3.1. Single-Pixel MSE

To quantify the performance of a LiDAR pixel, we recall that the decision process involves estimating the delay τ given the measurements $\mathbf{t}_M$. Therefore, we need to specify the estimation procedure. Based on the estimates, we can then discuss the performance by evaluating the variance of the estimate.

Maximum-Likelihood Estimation (MLE). When no knowledge about τ is known a priori, we use MLE [18, 36]. MLE has been thoroughly exploited in single-depth estimation problems [2]. Given the measured time stamps $\mathbf{t}_M = [t_1, \ldots, t_M]$, we consider the log-likelihood

$$\widehat{\tau} = \underset{\tau}{\text{argmax}} \quad \mathcal{L}(\tau) \overset{\text{def}}{=} \sum_{j=1}^{M} \log\left[\alpha s(t_j - \tau) + \lambda_b\right],$$

Since the variable τ in the ML estimation is the time shift, the optimization can be solved by running a matched filter. Given the shape $s(t)$, we shift the pulse left and right until we see the best match with the data. Fig. 2 shows a pictorial illustration.

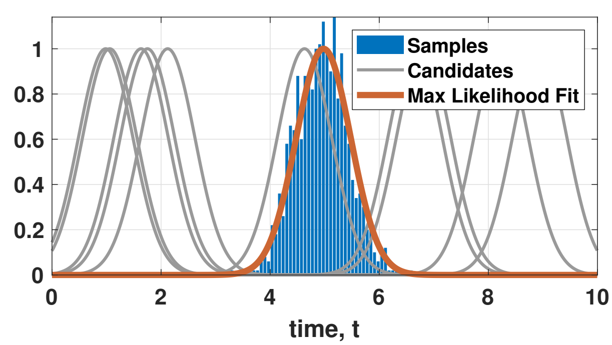

Figure 2. Matched filter: Given a known pulse shape, we shift the pulse until it matches with the measured samples.

If $\widehat{\tau}$ is the ML estimate, it is necessary that

$$\left.\frac{d\mathcal{L}}{d\tau}\right|_{\tau=\widehat{\tau}} = \left.\sum_{j=1}^{M} \frac{\alpha \dot{s}(t_j-\tau)}{\alpha s(t_j-\tau)+\lambda_b}\right|_{\tau=\widehat{\tau}} = 0. \quad (12)$$

This result is used in the implementation. Details can be found in the supplementary material.

MSE calculation.When τ_0 denotes the true time of arrival, the Taylor expansion of $\dot{\mathcal{L}}(\tau) = d\mathcal{L}/d\tau$ will give us

$$\dot{\mathcal{L}}(\tau) = \dot{\mathcal{L}}(\tau_0) + (\tau-\tau_0)\ddot{\mathcal{L}}(\tau_0) + \dots$$

Substituting $\tau = \widehat{\tau}$, and using the fact that $\dot{\mathcal{L}}(\widehat{\tau}) = 0$ because $\widehat{\tau}$ is the maximizer, we can show that

$$0 = \dot{\mathcal{L}}(\widehat{\tau}) = \dot{\mathcal{L}}(\tau_0) + (\widehat{\tau}-\tau_0)\ddot{\mathcal{L}}(\tau_0) + \dots$$

Therefore, the error is $\widehat{\tau}-\tau_0 \approx -\frac{\dot{\mathcal{L}}(\tau_0)}{\ddot{\mathcal{L}}(\tau_0)}$. By using this result, the variance of the estimate $\widehat{\tau}$ can be shown as follows.

Theorem 2. *[1, 3] Let $\lambda(t) = \alpha s(t-\tau_0) + \lambda_b$. Then*

$$\mathbb{E}[(\widehat{\tau}-\tau_0)^2] = \left[\int_{-T}^{T} \frac{(\alpha \dot{s}(t))^2}{\alpha s(t)+\lambda_b}\, dt\right]^{-1}, \quad (13)$$

where $\dot{s}(t)$ is the derivative of s with respect to t.

Example 3. *In the special case where $s(t) = \mathcal{N}(t \,|\, \tau_0, \sigma_t^2)$, and assume that $\lambda_b = 0$, we have*

$$\mathbb{E}[(\widehat{\tau}-\tau_0)^2] = \left[\int_{-T}^{T} \frac{(\alpha \dot{s}(t))^2}{\alpha s(t)+\lambda_b}\, dt\right]^{-1}$$

$$= \left[\int_{-T}^{T} \frac{\left(-\frac{t}{\sigma_t^2}\cdot\frac{\alpha}{\sqrt{2\pi\sigma_t^2}}e^{-\frac{t^2}{2\sigma_t^2}}\right)^2}{\frac{\alpha}{\sqrt{2\pi\sigma_t^2}}e^{-\frac{t^2}{2\sigma_t^2}}}\, dt\right]^{-1}$$

$$= \left[\int_{-T}^{T} \frac{t^2}{\sigma_t^4}\frac{\alpha}{\sqrt{2\pi\sigma_t^2}}e^{-\frac{t^2}{2\sigma_t^2}}\, dt\right]^{-1} \approx \left(\frac{\alpha}{\sigma_t^2}\right)^{-1} = \frac{\sigma_t^2}{\alpha}.$$

The last integration is the second moment of a zero-mean Gaussian, which will give us σ_t^2.

We remark that the per-pixel error calculated in Theorem 2 reaches the equality of the Cramer-Rao lower bound previously reported in [10, 34, 35]. Thus, no other unbiased estimator is better than what is reported here.

3.2. Space-Time Model

Continuous $\lambda(x,t)$. Our resolution-noise trade-off analysis requires a model of an *array* of pixels. To this end, we need to generalize from a single time delay τ to a function of time of arrivals $\tau(x)$ where x is the spatial coordinate. Thus, at every location x, and given the pulse shape $s(t)$, the ideal return pulse is

$$\lambda(x,t) = \alpha \cdot s(t-\tau(x)) + \lambda_b. \quad (14)$$

Fig. 3 shows a typical $\lambda(x,t)$ where the time delay $\tau(x)$ is translated to a space-time signal with a Gaussian pulse at every x. The discretization of $\lambda(x,t)$ will play a key role in our analysis.

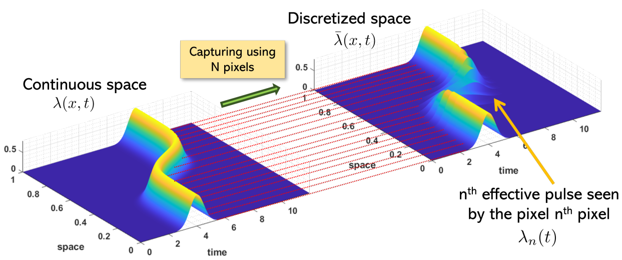

Figure 3. The space-time signal $\lambda(x,t)$ in the unit length $0 \le x \le 1$ and time span $[0,T]$, and its corresponding "effective" returned pulse $\overline{\lambda}(x,t)$ where each individual returned pulse is $\lambda_n(t)$.

Observing $\lambda(x,t)$ through N pixels. Suppose that we allocate N pixels in the unit space to measure the returned time of arrivals. These times of arrivals are generated according to the joint distribution specified in Eq. (3). However, since at the n^{th} pixel the function $\lambda(x,t)$ occupies the interval $\frac{n}{N} \le x \le \frac{n+1}{N}$, we can define the *effective* return pulse $\lambda_n(t)$ by absorbing the coordinate x through integration. Specifically, we define $\lambda_n(t)$ as

$$\lambda_n(t) = \int_{\frac{n}{N}}^{\frac{n+1}{N}} \lambda(x,t)dx. \quad n = 0,\dots,N-1 \quad (15)$$

The resulting space-time approximation of $\lambda(x,t)$ is thus a piecewise function

$$\overline{\lambda}(x,t) = N\sum_{n=0}^{N-1} \lambda_n(t)\varphi(Nx-n), \quad (16)$$

where $\varphi(x)$ is a boxcar function defined as

$$\varphi(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

The definition here is consistent with how the spatial-oversampled quanta image sensor was defined [9, 44].

Remark 1: Can we approximate $\tau(x)$ instead? By looking at Eq. (15), it is tempted to think that we can approximate $\tau(x)$ via a piecewise *constant* function

$$\tau(x) \approx \overline{\tau}(x) = \sum_{n=0}^{N-1} \overline{\tau}_n\varphi(Nx-n), \quad (18)$$

where

$$\overline{\tau}_n = N \int_{\frac{n}{N}}^{\frac{n+1}{N}} \tau(x) dx. \tag{19}$$

This will give us a plausible candidate for $\lambda_n(t)$:

$$\lambda_n(t) = \alpha s(t - \overline{\tau}_n) + \lambda_b, \quad n = 0, \ldots, N-1. \tag{20}$$

However, the problem with this approximation is that physically it is invalid. As light propagates, the energy carried by the wave follows the "scattering" process via the superposition of the electromagnetic field [5]. When energy is distributed from the source, we need to integrate $\lambda(x, t)$ and not $\tau(x)$.

3.3. New Approximation Techniques

While Eq. (15) is a physically valid way to perform spatial discretization, it does not have a simple analytic expression. For example, when $\alpha = 1$ and $\lambda_b = 0$, if we plug a Gaussian pulse $s(t) = \mathcal{N}(t \,|\, 0, \sigma_t^2)$ into Eq. (15), we will need to evaluate the integral

$$\lambda_n(t) = \int_{\frac{n}{N}}^{\frac{n+1}{N}} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{(t-\tau(x))^2}{2\sigma_t^2}\right\} dx.$$

Since $\tau(x)$ is a function of x, it is impossible to arrive at a closed-form expression.

Our plan of deriving the theoretical bound involves several steps. At the core of our proof technique is the approximation of the boxcar function using a Gaussian kernel, as illustrated in Fig. 4. If we assume that the pulse is also a Gaussian, then a convolution of two Gaussians will remain a Gaussian. This will substantially improve the tractability of our equations.

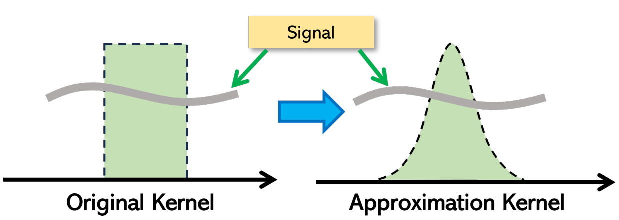

Figure 4. Our core proof involves an approximation of the boxcar kernel by a Gaussian. Doing so will allow us to replace the integration with a convolution.

Approximation 1: Linearize $\boldsymbol{\tau(x)}$. We approximate the time of arrival function $\tau(x)$ by a piecewise *linear* function. Suppose that there are N pixels in $[0, 1]$. We define the mid point x_n of each interval $[\frac{n}{N}, \frac{n+1}{N}]$ as

$$x_n \overset{\text{def}}{=} \frac{\frac{n}{N} + \frac{n+1}{N}}{2} = \frac{2n+1}{2N}.$$

Expanding $\tau(x)$ around x_n will give us

$$\tau(x) \approx \underbrace{\tau(x_n)}_{\overset{\text{def}}{=}\tau_n} + \underbrace{\tau'(x_n)}_{\overset{\text{def}}{=}c_n}(x - x_n), \qquad \tfrac{n}{N} \le x \le \tfrac{n+1}{N}. \tag{21}$$

Thus, for the entire $0 \le x \le 1$, $\tau(x)$ is approximated by

$$\tau(x) \approx \sum_{n=0}^{N-1} [\tau_n + c_n(x - x_n)] \varphi(Nx - n), \tag{22}$$

where φ is the boxcar function defined in Eq. (17).

Approximation 2: Replace boxcar by Gaussian. The second approximation is to give up the boxcar function $\varphi(x)$ because it does not allow us to derive a closed-form expression of $\lambda_n(t)$. We replace it with a Gaussian $\phi(x)$:

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{-\frac{x^2}{2\sigma_x^2}\right\} = \mathcal{N}(x \,|\, 0, \sigma_x^2). \tag{23}$$

However, if we want to approximate a boxcar function (with width W) by a Gaussian (with a standard deviation σ_x), what should be the relationship between W and σ_x so that the approximation is optimized? The answer is $\sigma_x = W/\sqrt{12}$.

Lemma 1. *Let $\varphi(x)$ be a boxcar function over the interval $[-\frac{W}{2}, \frac{W}{2}]$ and $\phi(x) = \mathcal{N}(x \,|\, 0, \sigma_x^2)$ be a Gaussian function. The optimal σ_x that offers the best match between $\varphi(x)$ and $\phi(x)$ is*

$$\sigma_x = \frac{W}{\sqrt{12}}. \tag{24}$$

If there are N pixels in $[0, 1]$, then the width of each pixel is $1/N$. This means that $\varphi(Nx - n)$ has a width of $1/N$. Therefore, the standard deviation of the shifted Gaussian $\phi(Nx - n)$ is $\sigma_x = 1/(\sqrt{12}N)$.

Approximation 3: Replace projection by convolution. One of the difficulties in Eq. (15) is the integration over the spatial interval. With the introduction of the Gaussian kernel, we replace the projection step by a spatially invariant convolution:

$$\begin{aligned}
\widetilde{\lambda}(x,t) &= \phi(x) \circledast \lambda(x,t) \qquad \text{[previously it was } \varphi(x)\text{]} \\
&= \mathcal{N}(x \,|\, 0, \sigma_x^2) \circledast [\alpha \mathcal{N}(t \,|\, \tau(x), \sigma_t^2) + \lambda_b] \\
&= \alpha\Big(\mathcal{N}(x \,|\, 0, \sigma_x^2) \circledast \mathcal{N}(t \,|\, \tau(x), \sigma_t^2)\Big) + \lambda_b.
\end{aligned}$$

The resulting $\lambda_n(t)$ can then be determined as the value of $\widetilde{\lambda}(x, t)$ at the mid point x_n of each pixel interval, i.e.,

$$\lambda_n(t) = \widetilde{\lambda}(x_n, t). \tag{25}$$

The following theorem summarizes the result of this series of approximations:

Theorem 3. *Under Approximations 1-3, the effective return pulse received by the n^{th} pixel is*

$$\lambda_n(t) = \alpha \cdot \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{-\frac{(t-\tau_n)^2}{2\sigma_n^2}\right\} + \lambda_b, \quad (26)$$

where $\tau_n = \tau(x_n)$*, and* $\sigma_n^2 = c_n^2\sigma_x^2 + \sigma_t^2$.

The biggest difference between Eq. (26) and Eq. (20) is the standard deviation of the Gaussian. In Eq. (20), the pulse width is always σ_t. Thus, the shape of the Gaussian is never changed no matter which pixel we consider. This problem is fixed in Eq. (26) where the standard deviation now depends on three things: (i) the temporal pulse width σ_t, (ii) the width of the pixel σ_x, (iii) the first derivative c_n of the time of arrival function $\tau(x)$.

3.4. Derivation of the MSE

Bias-Variance Decomposition. We are now in the position to derive the overall MSE. The MSE is measured between the true function $\tau(x)$ and the reconstructed function $\widehat{\tau}(x)$:

$$\text{MSE}(\widehat{\tau}, \tau) \overset{\text{def}}{=} \mathbb{E}\left[\int_0^1 (\widehat{\tau}(x) - \tau(x))^2 \, dx\right]. \quad (27)$$

In this equation, the reconstructed function $\widehat{\tau}(x)$ is a piecewise constant function defined by

$$\widehat{\tau}(x) \overset{\text{def}}{=} \sum_{n=0}^{N-1} \widehat{\tau}_n \varphi(Nx - n),$$

where $\widehat{\tau}_n$ is the ML estimate of the time of arrival at the n^{th} pixel, and $\varphi(x)$ is the boxcar function.

As will be shown in the supplementary material, the MSE defined Eq. (27) can be decomposed into bias and variance:

$$\text{MSE}(\widehat{\tau}, \tau) = \underbrace{\|\tau - \overline{\tau}\|_{L_2}^2}_{\text{bias}} + \underbrace{\mathbb{E}\left[\|\widehat{\tau} - \overline{\tau}\|_{L_2}^2\right]}_{\text{variance}}.$$

The bias measures how much resolution will drop when we use piecewise constant function $\overline{\tau}$ to approximate the continuous τ. The variance measures the noise fluctuation caused by the random ML estimate $\widehat{\tau}$.

Main Theoretical Result. The main result is stated in the theorem below.

Theorem 4 (Overall MSE)**.** *The MSE is*

$$MSE(\widehat{\tau}, \tau) = \underbrace{\frac{c^2}{12N^2}}_{bias} + \underbrace{\frac{N}{\alpha_0}\left(c^2\sigma_x^2 + \sigma_t^2\right)}_{variance}. \quad (28)$$

where $c^2 = (1/N)\sum_{j=1}^N c_n^2$*, and* α_0 *is the total flux of the scene.*

When deriving this main result, we assume that the pulse is Gaussian and the floor noise λ_b is zero. We will relax these assumptions in the supplementary material to consider more realistic situations.

Significance of Theorem 4. The main result is the first *closed-form expression* about the noise-resolution trade-off that we are aware of. As we will demonstrate in the experiment section, this simple formula matches well with the Monte Carlo simulation, albeit with minor numerical precision errors.

The closed-form expression in Theorem 4 offers many important insights about the behavior of the problem.

- α_0: Since α_0 is the total flux of the scene, a large α_0 will generate more time stamps which will in turn improve the variance. α_0 has no impact on the bias.
- σ_t: The pulse width determines the uncertainty of the time of arrivals, which affects the variance. σ_t does not affect the bias because the bias is independent of t.
- c: The slope of $\tau(x)$ specifies "how difficult" the scene is. In the easiest case where the scene is flat so that $c_n = 0$, the bias term drops to zero. If the slope is large, both bias and variance will suffer.
- σ_x: The parameter σ_x is a modeling constant. σ_x can be considered as a proxy to any diffraction limit caused by the optical system. A large point spread function of the optics will result in a large σ_x.

4. Experiments

4.1. Simulated 1D Experiment

We consider multiple 1D ground truth time of arrival functions $\tau(x)$ outlined in the supplementary material Sec. 9. The configurations can be found in Tab 1, also in the supplementary material.

Simulation. During simulation, we construct a space-time function $\lambda(x, t)$ with a very fine-grained spatial grid. At each x in the grid, there is a pulse function $s(t - \tau(x))$. We integrate $\lambda(x, t)$ for $\frac{n}{N} \leq x \leq \frac{n+1}{N}$ for each interval n to obtain the effective pulse $\lambda_n(t)$. M random time stamps are drawn from the inverse CDF of $\lambda_n(t)$, where M is a Poisson random variable with a rate α_0/N. The M time stamps (per each n) will give us an estimate $\widehat{\tau}_n$, which is then used to construct the reconstructed delay profile $\widehat{\tau}(x) = \sum_{n=0}^{N-1} \widehat{\tau}_n \varphi(Nx - n)$. We numerically compute the MSE for this $\widehat{\tau}(x)$.

Theory. The theoretical prediction follows the equation $\text{MSE}(\widehat{\tau}, \tau)$ described in Theorem 4. This is a one-line formula.

Result. The result of our experiment is reported in Fig. 5. As evident from the figure, the theoretical prediction matches very well with the simulation. The optimal number of pixels for this particular problem is around $N = 64$.

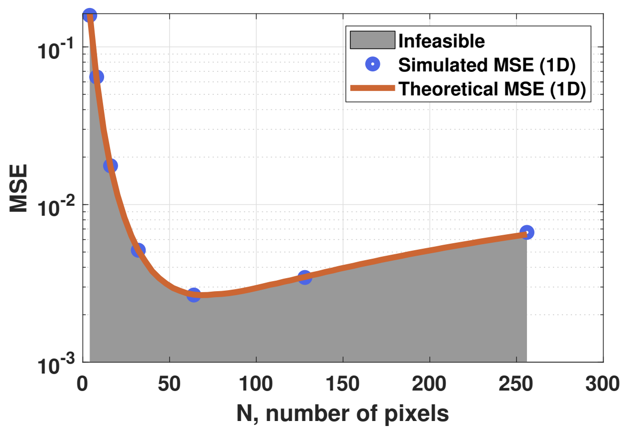

Figure 5. 1D simulation. Comparing simulation and the theoretically predicted MSE. Note the excellent match between the theory and the simulation.

4.2. Analysis of Variance

Our second experiment concerns about the validity of the approximations. Suppose we take a "lazy" route by using the "cheap" approximation outlined in Eq. (20). Then, under the condition that $s(t)$ is Gaussian and $\lambda_b = 0$, Theorem 2 will give us (via Example 3)

$$\text{MSE}(\tau, \widehat{\tau}) = \underbrace{\frac{c^2}{12N^2}}_{\text{bias}} + \underbrace{\frac{N}{\alpha_0}\sigma_t^2}_{\text{variance}}. \tag{29}$$

Compared with Theorem 4, the term $c^2\sigma_x^2$ is omitted. For the particular example shown in Fig. 5, we show in Fig. 6 a side-by-side comparison when the term $c^2\sigma_x^2$ is included and not included. It is clear from the figure that only the one with $c^2\sigma_x^2$ included can match with the simulation.

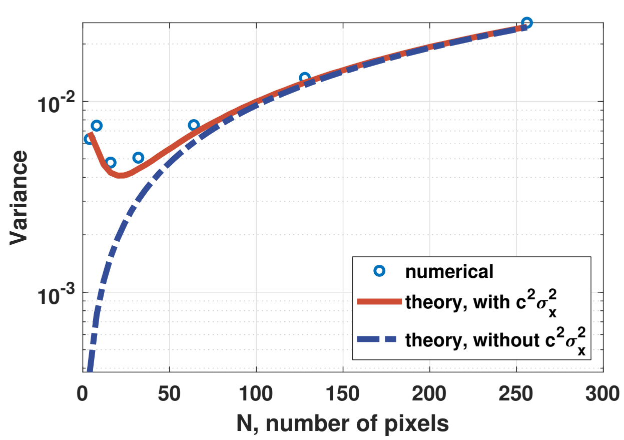

Figure 6. We compare two theoretical bounds: One with $c^2\sigma_x^2$ included (which is our full model), and one with $c^2\sigma_x^2$ missing (which is the simplified model). Note the excellent match between the theoretical prediction and the simulation result.

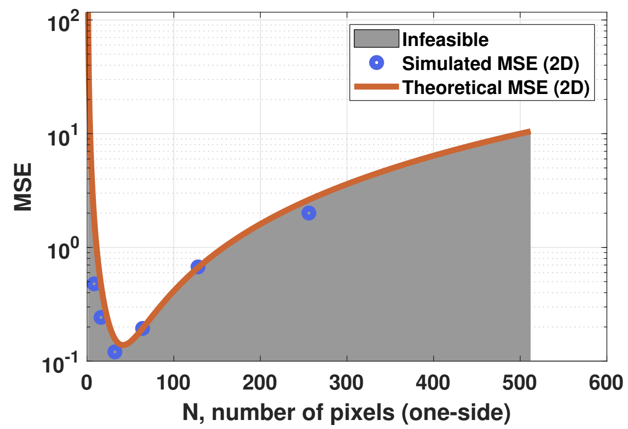

Figure 7. 2D simulation. Comparing simulation and the theoretically predicted MSE. Note the excellent match between the theory and the simulation.

4.3. Simulated 2D Experiment

Simulation. For 2D experiments, we use a ground truth depth map to generate the true time of arrival signal $\tau(\mathbf{x})$. Then, following a similar procedure outlined for the 1D case, we generate time stamps according to the required spatial resolution. For simplicity, we assume that the pulses are Gaussian, and that there is no noise floor. A piecewise constant 2D signal is reconstructed and the MSE is calculated.

Theory. The derivation of the theoretical MSE is outlined in the supplementary material Sec 14. Summarizing it here, the MSE is (with N being the number of pixels in one direction)

$$\text{MSE}(\tau, \widehat{\tau}) = \frac{\|\mathbf{c}\|^2}{12N^2} + \frac{N^2}{\alpha_0}\left(\|\mathbf{c}\|^2\sigma_x^2 + \sigma_t^2\right), \tag{30}$$

where $\|\mathbf{c}\|^2 = \int_{[0,1]^2} \|\nabla\tau(\mathbf{x})\|^2 d\mathbf{x}$ is the average gradient of the 2D time of arrival function.

Result. The result is outlined in Fig. 7. As we can see, the theoretical MSE again provides an excellent match with the simulated MSE.

4.4. Real 2D Experiment

In this experiment we analyze the real SPAD data collected by a sensor reported in [17]. The indoor scene consists of a static fan with a flat background, which is flood-illuminated using a picosecond pulsed laser source (Picoquant LDH series 670 nm laser diode with 1nJ pulse energy, operated with 25 MHz repetition rate). An f/4 objective was used in front of the SPAD, and binary time stamp frames (with a maximum of a single time stamp per pixel) were captured

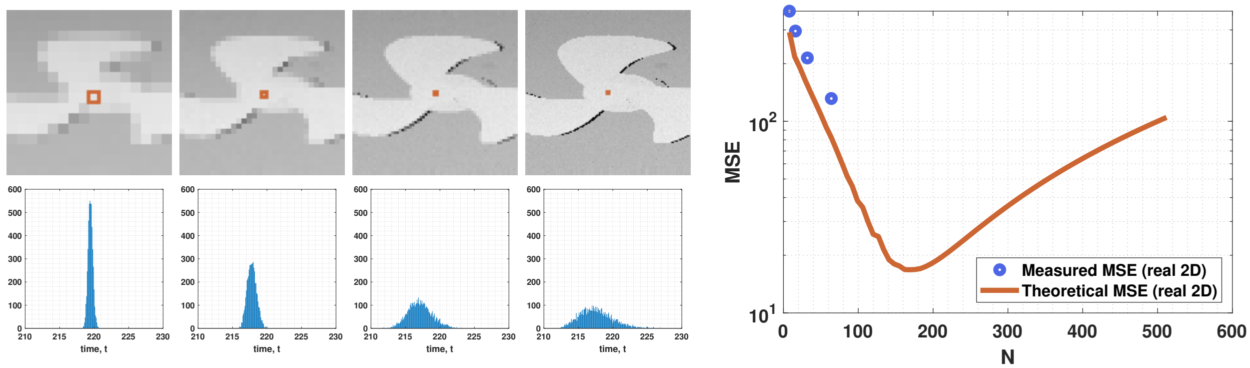

Figure 8. Real 2D experiment using a 192 × 128 SPAD reported in [17]. [Left-Top] ML estimate of the time of arrivals at different resolutions. As we reduce the spatial resolution of the SPAD, the noise per pixel reduces whereas the resolution becomes worse. [Left-Bottom] The distribution of ML estimate at the orange location. As we use a larger pixel, the variance of the estimated time of arrival reduces. [Right] The MSE curve compares the estimate and the pseudo ground truth, and the corresponding theoretical predictions.

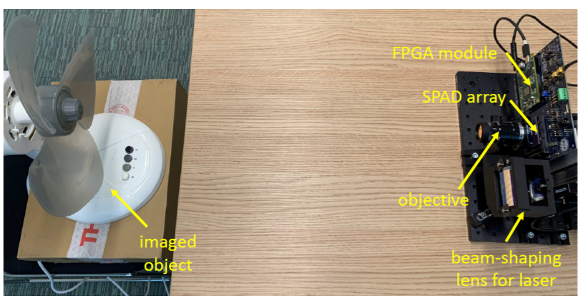

Figure 9. Experimental setup to capture the real SPAD data. The sensor we used here is 192 × 128 SPAD developed by Henderson et al. [17].

with an exposure time of 1 ms per frame. A total of 10,000 time stamps with a timing resolution of 35ps were thereby collected for each pixel. Pre-processing is performed to remove outliers. More descriptions of how this is done can be found in the supplementary material. The outcomes of the real 2D experiment are depicted in Fig. 8, whereas the schematic diagram of the experimental setup is shown in Fig. 9.

The top row of Fig. 8 shows the estimated depth map at four different resolutions. The estimation is done using the ML estimation. The bottom row of Fig. 8 shows the distribution of the ML estimates. This distribution is obtained through a bootstrap procedure where we sample with replacement $M = 3$ time stamps to estimate the time, and we bootstrap for $5,000$ times. The shrinking variance confirms that as we use fewer pixels, the estimation quality improves. The right hand side of Fig. 8 shows the theoretically predicted MSE and the measured MSE. The measured MSE is obtained by first constructing a pseudo ground truth from all the 10,000 frames (with pre-processing). We draw $M = 3$ samples from each pixel, sampled with replacement repeatedly 100 times, to compute the MSE.

The result in Fig. 8 does not show a valley. This is because the optimal N, by taking derivative of Eq. (30), is $N = \left(\frac{\sqrt{\alpha_0}\|\mathbf{c}\|}{\sqrt{12}\sigma_t}\right)^{1/2}$. Therefore, if the pulse width is short so σ_t is small, it is possible that the optimal N is larger than the physical resolution of the SPAD. In this case, maximizing the resolution is the best option.

5. Conclusion

A closed-form expression of the resolution limit for a SPAD sensor array is presented. It is found that the MSE decreases when the total amount of flux is high, the scene is smooth, and the pulse width is small. The MSE demonstrates a U-shape as a function of the number of pixels N in a unit space. When the optimal N is beyond the physical resolution of the sensor, no binning would be required. Extension of the theory to pile-up effects and non-Gaussian pulse is achievable with numerical integration. Advanced post-processing can possibly outperform the predicted bound which is based on ML estimation.

Acknowledgement

We thank the anonymous reviewers and the area chair for the meticulous review of this paper and, for offering insightful and constructive feedback. We are particularly thankful to Reviewer `SxN3` for trusting us on this revision. We thank Abdullah Al Shabili for the initial investigation of the LiDAR statistics. We thank Jonathan Leach's lab at Heriot-Watt University, and German Mora-Martin of the University of Edinburgh for collecting and sharing the data with us.

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

All plots presented in this paper can be reproduced using the code provided at https://github.itap.purdue.edu/StanleyChanGroup/.

References

[1] James B. Abshire and Jan F. McGarry. Estimating the arrival times of photon-limited laser pulses in the presence of shot and speckle noise. *J. Opt. Soc. Am. A*, 4(6):1080–1088, 1987. 4

[2] Yoann Altmann, Ximing Ren, Aongus McCarthy, Gerald S. Buller, and Steve McLaughlin. LiDAR waveform-based analysis of depth images constructed using sparse single-photon data. *IEEE Transactions on Image Processing*, 25 (5):1935–1946, 2016. 1, 2, 3

[3] Israel Bar-David. Communication under the poisson regime. *IEEE Transactions on Information Theory*, 15(1):31–37, 1969. 2, 3, 4

[4] Alberto Boretti. A perspective on single-photon LiDAR systems. *Microwave and Optical Technology Letters*, 66(1): e33918, 2024. 1

[5] Nicholas Chimitt, Xingguang Zhang, Yiheng Chi, and Stanley H. Chan. Scattering and gathering for spatially varying blurs. *IEEE Transactions on Signal Processing*, 2024. To appear. Available online at https://arxiv.org/abs/2303.05687. 5

[6] Joon Hee Choi, Omar A. Elgendy, and Stanley H. Chan. Image reconstruction for quanta image sensors using deep neural networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6543–6547, 2018. 1

[7] Peter Coates. The correction for photon 'pile-up' in the measurement of radiative lifetimes. *Journal of Physics E: Scientific Instruments*, 1(8):878, 1968. 3

[8] Neale A. W. Dutton, Istvan Gyongy, Luca Parmesan, Salvatore Gnecchi, Neil Calder, Bruce R. Rae, Sara Pellegrini, Lindsay A. Grant, and Robert K. Henderson. A SPAD-based QVGA image sensor for single-photon counting and quanta imaging. *IEEE Transactions on Electron Devices*, 63(1): 189–196, 2016. 1

[9] Omar A. Elgendy and Stanley H. Chan. Optimal threshold design for quanta image sensor. *IEEE Transactions on Computational Imaging*, 4(1):99–111, 2018. 4

[10] Baris I. Erkmen and Bruce Moision. Maximum likelihood time-of-arrival estimation of optical pulses via photon-counting photodetectors. In *2009 IEEE International Symposium on Information Theory*, pages 1909–1913, 2009. 4

[11] Eric R. Fossum. *The Invention and Development of CMOS Image Sensors*, chapter 23, pages 281–291. John Wiley & Sons, Ltd, 2023. 1

[12] Eric R. Fossum, Jiaju Ma, Saleh Masoodian, Leo Anzagira, and Rachel Zizza. The quanta image sensor: Every photon counts. *Sensors*, 16(8), 2016. 1

[13] Anant Gupta, Atul Ingle, and Mohit Gupta. Asynchronous single-photon 3D imaging. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7908–7917, 2019. 2, 3

[14] Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. Photon-flooded single-photon 3D cameras. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6763–6772, 2019. 1, 2, 3

[15] Robert H. Hadfield, Jonathan Leach, Fiona Fleming, Douglas J. Paul, Chee Hing Tan, Jo Shien Ng, Robert K. Henderson, and Gerald S. Buller. Single-photon detection for long-range imaging and sensing. *Optica*, 10(9):1124–1141, 2023. 1

[16] Felix Heide, Steven Diamond, David B. Lindell, and Gordon Wetzstein. Sub-picosecond photon-efficient 3D imaging using single-photon sensors. *Scientific Reports*, 8(1):17726, 2018. 1

[17] Robert K. Henderson, Nick Johnston, Francescopaolo Mattioli Della Rocca, Haochang Chen, David Day-Uei Li, Graham Hungerford, Richard Hirsch, David Mcloskey, Philip Yip, and David J. S. Birch. A 192×128 time correlated SPAD image sensor in 40-nm CMOS technology. *IEEE Journal of Solid-State Circuits*, 54(7):1907–1916, 2019. 1, 7, 8

[18] Steven Kay. *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*. Prentice-Hall PTR, 2013. 3

[19] Ahmed Kirmani, Dheera Venkatraman, Dongeek Shin, Andrea Colaço, Franco N. C. Wong, Jeffrey H. Shapiro, and Vivek K Goyal. First-photon imaging. *Science*, 343(6166): 58–61, 2014. 1

[20] Jongho Lee, Atul Ingle, Jenu V. Chacko, Kevin W. Eliceiri, and Mohit Gupta. CASPI: collaborative photon processing for active single-photon imaging. *Nature Communications*, 14(1):3158, 2023. 1

[21] Zhaohui Li, Haifeng Pan, Guangyue Shen, Didi Zhai, Weihua Zhang, Lei Yang, and Guang Wu. Single-photon LiDAR for canopy detection with a multi-channel Si SPAD at 1064 nm. *Optics & Laser Technology*, 157:108749, 2023. 1

[22] David B. Lindell, Matthew O'Toole, and Gordon Wetzstein. Single-photon 3D imaging with deep sensor fusion. *ACM Trans. Graph.*, 37(4):113.1–113.12, 2018. 1, 2

[23] Sizhuo Ma, Shantanu Gupta, Arin C. Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. Quanta burst photography. *ACM Trans. Graph.*, 39(4), 2020. 1

[24] Aongus McCarthy, Nils J. Krichel, Nathan R. Gemmell, Ximing Ren, Michael G. Tanner, Sander N. Dorenbos, Val Zwiller, Robert H. Hadfield, and Gerald S. Buller. Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. *Opt. Express*, 21(7): 8904–8915, 2013. 1

[25] Germán Mora-Martín, Alex Turpin, Alice Ruget, Abderrahim Halimi, Robert Henderson, Jonathan Leach, and Istvan Gyongy. High-speed object detection with a single-photon time-of-flight image sensor. *Opt. Express*, 29(21): 33184–33196, 2021.

[26] Kazuhiro Morimoto, Andrei Ardelean, Ming-Lo Wu, Arin Can Ulku, Ivan Michel Antolovic, Claudio Bruschini, and Edoardo Charbon. Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications. *Optica*, 7(4): 346–354, 2020. 1

[27] Adithya K. Pediredla, Aswin C. Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. Signal processing based pile-up compensation for gated single-photon avalanche diodes. 2018. Available at https://arxiv.org/abs/1806.07437. 3

[28] Ryan Po, Adithya Pediredla, and Ioannis Gkioulekas. Adaptive gating for single-photon 3D imaging. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16333–16342, 2022. 3

[29] Valentin Poisson, Van Thien Nguyen, William Guicquero, and Gilles Sicard. Luminance-depth reconstruction from compressed time-of-flight histograms. *IEEE Transactions on Computational Imaging*, 8:148–161, 2022. 1

[30] Xuanyu Qian, Wei Jiang, and M. Jamal Deen. Single photon detectors for automotive LiDAR applications: State-of-the-art and research challenges. *IEEE Journal of Selected Topics in Quantum Electronics*, 30(1: Single-Photon Technologies and Applications):1–20, 2024. 1

[31] Joshua Rapp and Vivek K Goyal. A few photons among many: Unmixing signal and noise for photon-efficient active imaging. *IEEE Transactions on Computational Imaging*, 3 (3):445–459, 2017. 1, 2, 3

[32] Joshua Rapp, Julian Tachella, Yoann Altmann, Stephen McLaughlin, and Vivek K Goyal. Advances in single-photon Lidar for autonomous vehicles: Working principles, challenges, and recent advances. *IEEE Signal Processing Magazine*, 37(4):62–71, 2020. 1

[33] Justin Richardson, Richard Walker, Lindsay Grant, David Stoppa, Fausto Borghetti, Edoardo Charbon, Marek Gersbach, and Robert K. Henderson. A 32×32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging. In *2009 IEEE Custom Integrated Circuits Conference*, pages 77–80, 2009. 1

[34] Stirling Scholes, German Mora-Martin, Feng Zhu, Istvan Gyongy, Phil Soan, and Jonathan Leach. Fundamental limits to depth imaging with single-photon detector array sensors. *Nature Scientific Reports*, 13(1):176, 2023. 4

[35] Dongeek Shin, Ahmed Kirmani, Vivek K Goyal, and Jeffrey H. Shapiro. Information in a photon: Relating entropy and maximum-likelihood range estimation using single-photon counting detectors. In *2013 IEEE International Conference on Image Processing*, pages 83–87, 2013. 4

[36] Dongeek Shin, Ahmed Kirmani, Vivek K Goyal, and Jeffrey H. Shapiro. Photon-efficient computational 3-D and reflectivity imaging with single-photon detectors. *IEEE Transactions on Computational Imaging*, 1(2):112–125, 2015. 1, 3

[37] Donald L. Snyder and Michael I. Miller. *Random Point Processes in Time and Space*. Springer, 1991. 1, 2, 3

[38] Julián Tachella, Yoann Altmann, Nicolas Mellado, Aongus McCarthy, Rachael Tobin, Gerald S. Buller, Jean-Yves Tourneret, and Stephen McLaughlin. Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers. *Nature Communications*, 10(1):4984, 2019. 1

[39] Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolović, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. A 512 × 512 SPAD image sensor with integrated gating for widefield FLIM. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(1):1–12, 2019. 1

[40] Federica Villa, Fabio Severini, Francesca Madonini, and Franco Zappa. SPADs and SiPMs arrays for long-range high-speed light detection and ranging LiDAR. *Sensors*, 21(11), 2021.

[41] Michael Wayne, Arin Ulku, Andrei Ardelean, Paul Mos, Claudio Bruschini, and Edoardo Charbon. A 500 × 500 dual-gate SPAD imager with 100 *IEEE Transactions on Electron Devices*, 69(6):2865–2872, 2022. 1

[42] Mian Wei, Sotiris Nousias, Rahul Gulve, David B. Lindell, and Kiriakos N. Kutulakos. Passive ultra-wideband single-photon imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8135–8146, 2023. 1

[43] Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. A theory of fermat paths for non-line-of-sight shape reconstruction. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6793–6802, 2019. 1

[44] Feng Yang, Yue M. Lu, Luciano Sbaiz, and Martin Vetterli. Bits from photons: Oversampled image acquisition using binary poisson statistics. *IEEE Transactions on Image Processing*, 21(4):1421–1436, 2012. 4