

Resolution Limit of Single-Photon LiDAR

Stanley H. Chan^{†1}

Hashan K. Weerasooriya¹

Weijian Zhang¹

Pamela Abshire²

Istvan Gyongy³

Robert K. Henderson³

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 [22, 25–27, 31, 33]. 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, 20, 37]. The advancement of photo detectors has significantly improved the resolution of today’s LiDAR [9, 16, 18, 34, 41–43]. Moreover, algorithms have shown how to reconstruct both the scene reflectivity and 3D structure [2, 7, 17, 21, 23, 24, 30, 37, 39, 44, 45].

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 [13]. This fundamental limit is linked to the stochastic nature of the underlying Poisson arrival process of the photons [12, 38]. Unless noise mitigation schemes are employed [2, 15, 23, 32], 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

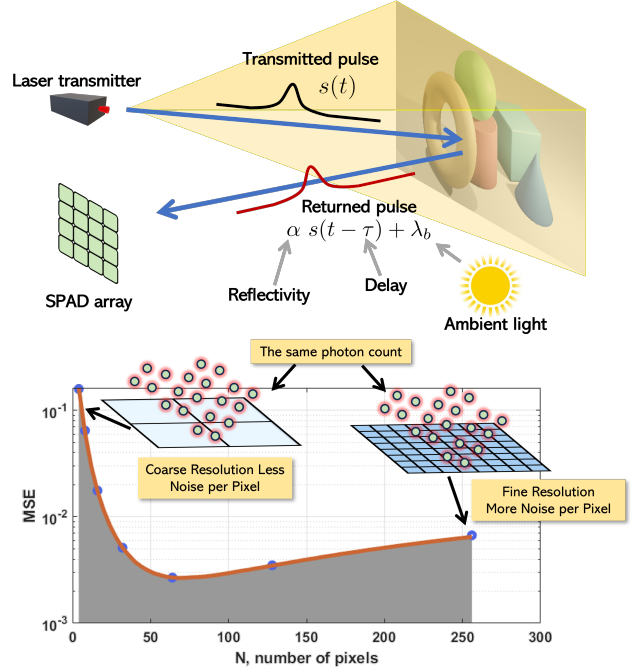


Figure 1. As we pack more pixels in a unit space, we gain the spatial resolution with a reduction in the SNR. The goal of this paper is to understand the trade-off between the two factors.

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 vi-

sion papers in single-photon LiDAR are algorithmic. Few papers have theoretical derivations, but they all focus on a single pixel [2, 15, 23, 32], 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 [14, 32, 38]. 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-invariant function $s(t)$. Given a delay τ , the shifted pulse is $s(t - \tau)$.

Example 1. If the pulse is Gaussian, then

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

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

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

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

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

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

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

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 $-T \leq t_1 < t_2 < \dots < t_M \leq T$. 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 $-T \leq t_1 < t_2 < \dots < t_M \leq T$. For $M \geq 1$,

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

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 $M \geq 1$, the probability that there are M occurrences is

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

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

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

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.

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

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

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

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).

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

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

2.3. Sampling from $p(\mathbf{t}_M)$

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

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

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:

$$t_j | M_s \sim \mathcal{N}(t | \tau, \sigma^2), \quad j = 1, \dots, M_s, \\ t_i | M_b \sim \text{Uniform}(-T, T), \quad i = 1, \dots, M_b.$$

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 [8, 14, 15, 28, 29]. 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 [32] and [14] use self-excitation process (a variant of the Markov chain) to model the photon arrivals [38]. 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 [19, 37]. MLE has been thoroughly exploited in single-depth estimation problems [2]. Given the measured time stamps $\mathbf{t}_M = [t_1, \dots, t_M]$, we consider the log-likelihood

$$\hat{\tau} = \underset{\tau}{\operatorname{argmax}} \mathcal{L}(\tau) \stackrel{\text{def}}{=} \sum_{j=1}^M \log [\alpha s(t_j - \tau) + \lambda_b],$$

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.

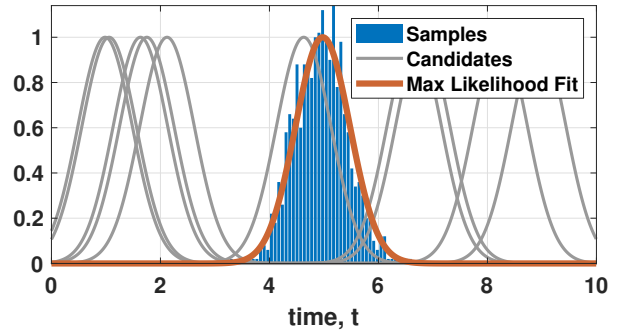


Figure 2. Matched filter: Given a known pulse shape, we shift the pulse until it matches with the measured samples.

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

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

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

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

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

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

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

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

$$\begin{aligned} \mathbb{E}[(\hat{\tau} - \tau_0)^2] &= \left[\int_{-T}^T \frac{(\alpha \dot{s}(t))^2}{\alpha s(t) + \lambda_b} dt \right]^{-1} \\ &= \left[\int_{-T}^T \frac{\left(-\frac{t}{\sigma_t^2} \cdot \frac{\alpha}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{t^2}{2\sigma_t^2}} \right)^2}{\frac{\alpha}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{t^2}{2\sigma_t^2}}} dt \right]^{-1} \\ &= \left[\int_{-T}^T \frac{t^2}{\sigma_t^4} \frac{\alpha}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{t^2}{2\sigma_t^2}} dt \right]^{-1} \approx \left(\frac{\alpha}{\sigma_t^2} \right)^{-1} = \frac{\sigma_t^2}{\alpha}. \end{aligned}$$

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 [11, 35, 36]. 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.

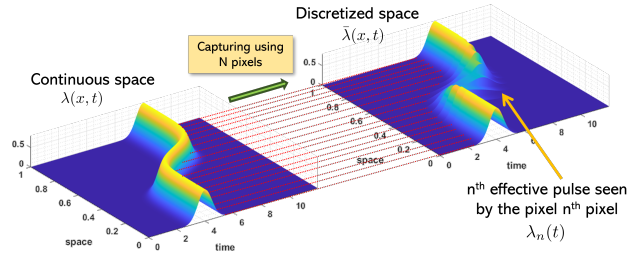


Figure 3. The space-time signal $\lambda(x, t)$ in the unit length $0 \leq x \leq 1$ and time span $[0, T]$, and its corresponding “effective” returned pulse $\bar{\lambda}(x, t)$ where each individual returned pulse is $\lambda_n(t)$.

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 $\frac{n}{N} \leq x \leq \frac{n+1}{N}$, 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

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

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

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

The definition here is consistent with how the spatial-oversampled quanta image sensor was defined [10, 46].

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

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

where

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

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

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

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 [6]. 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.

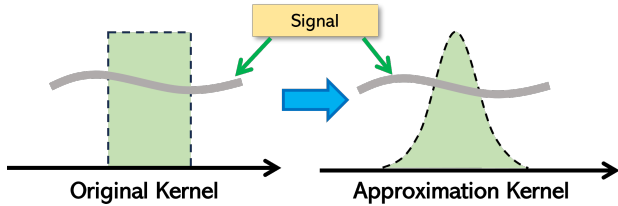


Figure 4. Our core proof involves an approximation of the boxcar kernel by a Gaussian. Doing so will allow us to replace the integration with a convolution.

Approximation 1: Linearize $\tau(x)$. 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

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

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

$$\tau(x) \approx \underbrace{\tau(x_n)}_{\stackrel{\text{def}}{=} \tau_n} + \underbrace{\tau'(x_n)}_{\stackrel{\text{def}}{=} c_n} (x - x_n), \quad \frac{n}{N} \leq x \leq \frac{n+1}{N}. \quad (21)$$

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

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

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 $\phi(x)$:

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

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

$$\sigma_x = \frac{W}{\sqrt{12}}. \quad (24)$$

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:

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

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

$$\lambda_n(t) = \tilde{\lambda}(x_n, t). \quad (25)$$

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 $\hat{\tau}(x)$:

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

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

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

where $\hat{\tau}_n$ 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:

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

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

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

Theorem 4 (Overall MSE). The MSE is

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

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 $\hat{\tau}_n$, which is then used to construct the reconstructed delay profile $\hat{\tau}(x) = \sum_{n=0}^{N-1} \hat{\tau}_n \varphi(Nx - n)$. We numerically compute the MSE for this $\hat{\tau}(x)$.

Theory. The theoretical prediction follows the equation $\text{MSE}(\hat{\tau}, \tau)$ 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 $N = 64$.

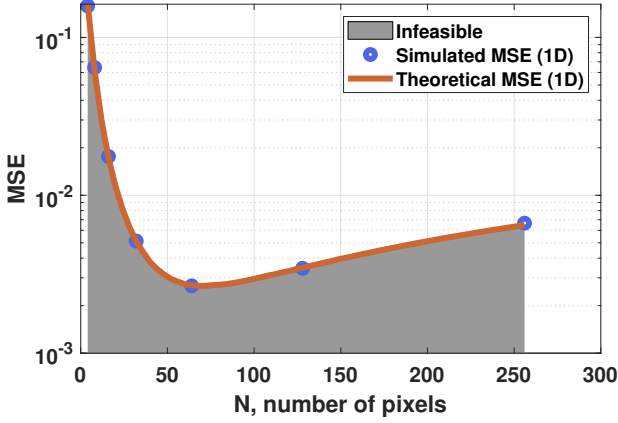


Figure 5. 1D simulation. Comparing simulation and the theoretically predicted MSE. Note the excellent match between the theory and the simulation.

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)

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

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

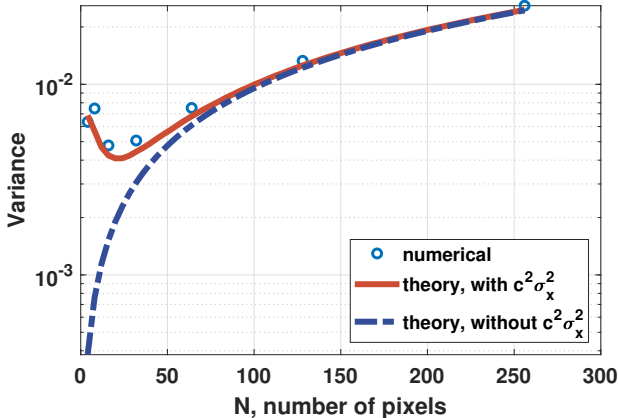


Figure 6. We compare two theoretical bounds: One with $c^2\sigma_x^2$ included (which is our full model), and one with $c^2\sigma_x^2$ missing (which is the simplified model). Note the excellent match between the theoretical prediction and the simulation result.

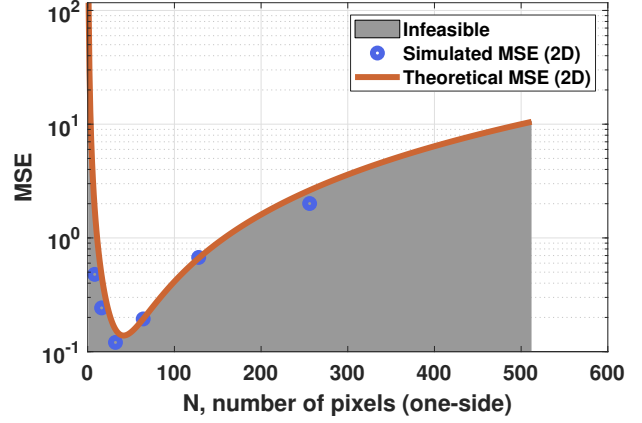


Figure 7. 2D simulation. Comparing simulation and the theoretically predicted MSE. Note the excellent match between the theory and the simulation.

with $c^2\sigma_x^2$ included can match with 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)

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

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 [18]. 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

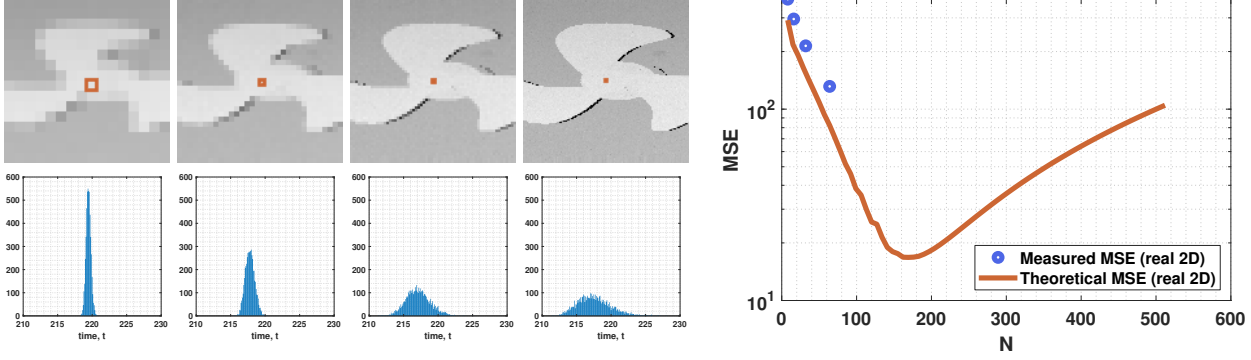


Figure 8. Real 2D experiment using a 192×128 SPAD reported in [18]. [Left-Top] ML estimate of the time of arrivals at different resolutions. As we reduce the spatial resolution of the SPAD, the noise per pixel reduces whereas the resolution becomes worse. [Left-Bottom] The distribution of ML estimate 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. [18].

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 $N = \left(\frac{\sqrt{\alpha_0} \|c\|}{\sqrt{12}\sigma_t} \right)^{1/2}$. 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 Shabali 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/StamleyChanGroup/>.

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, 11, 12, 14
- [4] Alberto Boretti. A perspective on single-photon LiDAR systems. *Microwave and Optical Technology Letters*, 66(1):e33918, 2024. 1
- [5] Stanley H. Chan. *Introduction to Probability for Data Science*. Michigan Publishing, 2021. 20
- [6] 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
- [7] 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
- [8] 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
- [9] Neale A. W. Dutton, Istvan Gyongy, Luca Parmesan, Salvatore Gnechi, 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
- [10] 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
- [11] Baris I. Erkmén 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
- [12] Eric R. Fossum. *The Invention and Development of CMOS Image Sensors*, chapter 23, pages 281–291. John Wiley & Sons, Ltd, 2023. 1
- [13] Eric R. Fossum, Jiaju Ma, Saleh Masoodian, Leo Anzagira, and Rachel Zizza. The quanta image sensor: Every photon counts. *Sensors*, 16(8), 2016. 1
- [14] 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
- [15] 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
- [16] 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
- [17] 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
- [18] 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, 28
- [19] Steven Kay. *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*. Prentice-Hall PTR, 2013. 3
- [20] 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
- [21] 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
- [22] 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
- [23] 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
- [24] Sizhuo Ma, Shantanu Gupta, Arin C. Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. Quanta burst photography. *ACM Trans. Graph.*, 39(4), 2020. 1
- [25] 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
- [26] Germán Mora-Martín, Alex Turpin, Alice Ruget, Abderahim 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.
- [27] 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
- [28] Adithya K. Pediredla, Aswin C. Sankaranarayanan, Mauro Buttavava, Alberto Tosi, and Ashok Veeraraghavan. Sig-

- nal processing based pile-up compensation for gated single-photon avalanche diodes. 2018. Available at <https://arxiv.org/abs/1806.07437>. 3
- [29] 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
- [30] 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
- [31] 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
- [32] 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
- [33] 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
- [34] 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
- [35] 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
- [36] Donggeek 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, 15
- [37] Donggeek 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
- [38] Donald L. Snyder and Michael I. Miller. *Random Point Processes in Time and Space*. Springer, 1991. 1, 2, 3, 11
- [39] 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
- [40] Alessandro Tontini, Sonia Mazzucchi, Roberto Passerone, Nicolò Broseghini, and Leonardo Gasparini. Histogram-less LiDAR through SPAD response linearization. *IEEE Sensors Journal*, 24(4):4656–4669, 2024. 24
- [41] 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
- [42] 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.
- [43] 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
- [44] 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
- [45] 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
- [46] 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

Resolution Limit of Single-Photon LiDAR

Supplementary Material

Contents

6 . Proof of Theorems	11
6.1 . Axioms	11
6.2 . Proof of Theorem 1	11
6.3 . Proof of Theorem 2	12
6.4 . Proof of Theorem 3	12
6.5 . Proof of Theorem 4	12
7 . Proof of Additional Results	14
7.1 . Proof of Corollary 1	14
7.2 . Proof of Corollary 2	14
7.3 . Proof of Lemma 1	14
7.4 . Cramer Rao Lower Bound	15
8 . Auxiliary Results and Proofs	15
8.1 . Lemma about Product of Functions	15
8.2 . Lemma about Characteristic Function	16
8.3 . Lemma for μ_1 and F_1	16
8.4 . Lemma for μ_2 and F_2	17
9 . Detailed Setups of Experiments	17
9.1 . Ground Truth Time-of-Arrival Function	17
9.2 . What if we use a different $\tau(x)$?	18
10. Unit Conversion	19
11. Sampling Procedure	20
11.1. Inverse CDF Method	20
11.2. Implementation and Demonstration	20
12. Non-zero Noise Floor	20
12.1. Maximum Likelihood Estimation	20
12.2. MSE Calculation	21
12.3. Solving the ML Estimation	22
12.4. Experimental Results	23
13. Pile-up Effects	24
13.1. Distribution	24
13.2. MLE	24
13.3. Theoretical MSE	24
14. Dark Count	25
15. Extension to 2D	25
15.1. Bias	25
15.2. Variance	26
16. Real 2D Experiment	27
16.1. Snapshot of Real Data	28
16.2. Pseudo Ground Truth	28

16.3. Estimating σ_t and α_0	28
16.4. Variance Estimation via Bootstrap	29
16.5. Bias Estimation via Numerical Integration	29

17. Q&A 29

6. Proof of Theorems

In this section, we present the proofs of the theorems. To clarify the contributions of this paper, we add comments to each proof to highlight whether this is based on existing theory or it is a new proof.

6.1. Axioms

We first state the three axioms for inhomogeneous Poisson processes [38].

Definition 1. *Axioms for inhomogeneous Poisson process*

- The probability of one occurrence in an infinitesimal interval Δt is given by

$$\mathbb{P}[1, \Delta t] = \lambda(t)\Delta t, \quad \Delta t \rightarrow 0. \quad (31)$$

- The probability of more than one occurrence in Δt is zero for $\Delta t \rightarrow 0$. Therefore, $\mathbb{P}[0, \Delta t]$ is the complement of $\mathbb{P}[1, \Delta t]$, which will give us

$$\mathbb{P}[0, \Delta t] = 1 - \lambda(t)\Delta t, \quad \Delta t \rightarrow 0. \quad (32)$$

- The number of occurrences in any interval is independent of those in all other disjoint intervals.

6.2. Proof of Theorem 1

✂ *Remark: This proof is adopted from Bar-David [3] with new elaborations provided for each step.*

Consider a set of time stamps $-T \leq t_1 < t_2 < \dots < t_M \leq T$. For each timestamp t_j , we consider an infinitesimal width $\pm \Delta t_j/2$. The probability that one and only one

photon falls in each of the intervals and none outside is

$$\begin{aligned}
p(\mathbf{t}_M, M) \Delta t_1 \Delta t_2 \dots \Delta t_M \\
&= \mathbb{P}[\text{observing one and only one in each interval}] \\
&= \mathbb{P}[0, (-T, t_1 - \frac{1}{2}\Delta t_1)] \\
&\quad \times \mathbb{P}[1, (t_1 - \frac{1}{2}\Delta t_1, t_1 + \frac{1}{2}\Delta t_1)] \\
&\quad \times \mathbb{P}[0, (t_1 + \frac{1}{2}\Delta t_1, t_2 - \frac{1}{2}\Delta t_2)] \\
&\quad \dots \\
&\quad \times \mathbb{P}[1, (t_M - \frac{1}{2}\Delta t_M, t_M + \frac{1}{2}\Delta t_M)] \\
&\quad \times \mathbb{P}[0, (t_M + \frac{1}{2}\Delta t_M, T)] \\
&= \mathbb{P}[0, (-T, t_1 - \frac{1}{2}\Delta t_1)] \\
&\quad \times \lambda(t_1) \Delta t_1 \times \mathbb{P}[0, (t_1 + \frac{1}{2}\Delta t_1, t_2 - \frac{1}{2}\Delta t_2)] \\
&\quad \dots \\
&\quad \times \lambda(t_M) \Delta t_M \times \mathbb{P}[0, (t_M + \frac{1}{2}\Delta t_M, T)]
\end{aligned}$$

As $\Delta t_j \rightarrow 0$, the intervals merge to become $(t_1 + \frac{1}{2}\Delta t_1, t_2 - \frac{1}{2}\Delta t_2) \rightarrow (t_1, t_2)$. Therefore,

$$\begin{aligned}
&\mathbb{P}[0, (-T, t_1 - \frac{1}{2}\Delta t_1)] \times \dots \times \mathbb{P}[0, (t_M + \frac{1}{2}\Delta t_M, T)] \\
&= \mathbb{P}[0, (-T, t_1)] \times \mathbb{P}[0, (t_1, t_2)] \dots \times \mathbb{P}[0, (t_M, T)] \\
&= \mathbb{P}[0, (-T, T)] \\
&= \exp \left[- \int_{-T}^T \lambda(t) dt \right] = e^{-Q}.
\end{aligned}$$

Regrouping the terms, we can show that

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

6.3. Proof of Theorem 2

✂ *Remark: This proof is adopted from Bar-David [3].*

Let's assume that $F(\tau)$ is a twice differentiable function. Then, let $F_1 = \dot{F}(\tau_0)$, and decompose F_1 as $F_1 = \mu_1 + F'_1$, where F'_1 is a zero-mean random variable. Similarly, decompose $F_2 = \ddot{F}(\tau_0)$ as $F_2 = \mu_2 + F'_2$. Then,

$$\epsilon = -\frac{\dot{F}(\tau_0)}{\ddot{F}(\tau_0)} = -\frac{F_1}{F_2} = -\frac{\mu_1 + F'_1}{\mu_2 + F'_2}.$$

Using Lemma 7, we know that $\mu_1 = 0$. Moreover, let's assume that $F'_2 \ll \mu_2$. We can, hence, simplify the above as

$$\begin{aligned}
\epsilon^2 &= \left(\frac{F'_1}{\mu_2 + F'_2} \right)^2 = \left(\frac{F'_1}{\mu_2} \left[\frac{1}{1 + \frac{F'_2}{\mu_2}} \right] \right)^2 \\
&\approx \left(\frac{F'_1}{\mu_2} \left[1 - \frac{F'_2}{\mu_2} \right] \right)^2.
\end{aligned}$$

Taking the expectation will give

$$\mathbb{E}[\epsilon^2] \approx \frac{\mathbb{E}[F_1'^2]}{\mu_2^2} - 2 \underbrace{\frac{\mathbb{E}[(F_1'^2 F_2')]}{\mu_2^3}}_{\rightarrow 0} + \underbrace{\frac{\mathbb{E}[F_1'^2 F_2'^2]}{\mu_2^4}}_{\rightarrow 0}.$$

Using Lemma 7 and Lemma 8, and assuming that $\tau_0 = 0$ without loss of generality, it follows that

$$\begin{aligned}
\mathbb{E}[\epsilon^2] &= \frac{\mathbb{E}[F_1'^2]}{\mu_2^2} = \frac{\int_{-T}^T \frac{(\alpha \dot{s}(t))^2}{\alpha s(t) + \lambda_b} dt}{\left(\int_{-T}^T \frac{(\alpha \dot{s}(t))^2}{\alpha s(t) + \lambda_b} dt \right)^2} \\
&= \left[\int_{-T}^T \frac{(\alpha \dot{s}(t))^2}{\alpha s(t) + \lambda_b} dt \right]^{-1}.
\end{aligned}$$

6.4. Proof of Theorem 3

✂ *Remark: This is a new proof we make for this paper.*

We consider $s(x, t) = s(t - \tau(x))$. Using Approximation 1, we write

$$\begin{aligned}
s(x, t) &= s(t - \tau(x)) = \mathcal{N}(t | \tau(x), \sigma_t^2) \\
&= \mathcal{N}(t | \tau_n + c_n(x - x_n), \sigma_t^2).
\end{aligned}$$

By Approximation 3, we can show that

$$\begin{aligned}
\tilde{s}(x, t) &\stackrel{\text{def}}{=} \phi(x) \otimes s(x, t) \\
&= \mathcal{N}(x | 0, \sigma_x^2) \otimes \mathcal{N}(t | \tau_n + c_n(x - x_n), \sigma_t^2) \\
&\stackrel{(a)}{=} \mathcal{N}(x | 0, \sigma_x^2) \otimes \frac{1}{c_n} \mathcal{N}\left(x \mid \frac{t}{c_n} - \frac{\tau_n}{c_n} + x_n, \frac{\sigma_t^2}{c_n^2}\right) \\
&\stackrel{(b)}{=} \frac{1}{c_n} \mathcal{N}\left(x \mid \frac{t}{c_n} - \frac{\tau_n}{c_n} + x_n, \sigma_n^2\right) \\
&= \mathcal{N}(t | \tau_n + c_n(x - x_n), \sigma_n^2),
\end{aligned}$$

where (a) is based on a simple switch between the roles of x and t in a Gaussian distribution, and (b) is based on the fact that the convolution of two Gaussian probability density functions is equivalent to adding two Gaussian random variables.

If we restrict ourselves to $x = x_n$, then we will obtain $s_n(t) \stackrel{\text{def}}{=} \tilde{s}(x_n, t) = \mathcal{N}(t | \tau_n, \sigma_n^2)$. Therefore, using the fact that $\phi(x) \otimes \lambda_b = \int_{-\infty}^{\infty} \phi(x) \lambda_b dx = \lambda_b$ since $\phi(x)$ integrates to 1, we can show that

$$\begin{aligned}
\lambda_n(t) &= \alpha \{ \phi(x) \otimes s(x, t) \} |_{x=x_n} + \lambda_b \\
&= \alpha \mathcal{N}(t | \tau_n, \sigma_n^2) + \lambda_b.
\end{aligned}$$

6.5. Proof of Theorem 4

In proving Theorem 4, we need a few lemmas. The first lemma is the classical bias-variance trade-off.

Lemma 2 (MSE decomposition). *The MSE can be decomposed as*

$$MSE(\hat{\tau}, \tau) \stackrel{\text{def}}{=} \text{bias} + \text{var}, \quad (33)$$

where

$$\begin{aligned} \text{bias} &= \int_0^1 (\bar{\tau}(x) - \tau(x))^2 dx \\ \text{var} &= \mathbb{E} \left[\int_0^1 (\hat{\tau}(x) - \bar{\tau}(x))^2 dx \right]. \end{aligned}$$

✂ *Remark: The decomposition of the MSE into bias and variance can be found in most of the machine learning textbooks. For completeness and in the context of our paper, we provide the derivations here.*

Proof of Lemma 2. We start with the definition of the MSE:

$$\begin{aligned} MSE(\hat{\tau}, \tau) &\stackrel{\text{def}}{=} \mathbb{E} \left[\int_0^1 (\hat{\tau}(x) - \tau(x))^2 dx \right] \\ &= \mathbb{E} \left[\int_0^1 (\hat{\tau}(x) - \bar{\tau}(x) + \bar{\tau}(x) - \tau(x))^2 dx \right] \end{aligned}$$

Expanding the terms, we can show that

$$\begin{aligned} MSE(\hat{\tau}, \tau) &= \mathbb{E} \left[\underbrace{\int_0^1 (\hat{\tau}(x) - \bar{\tau}(x))^2 dx}_{\text{variance}} \right] \\ &\quad + \underbrace{\mathbb{E} \left[\int_0^1 (\hat{\tau}(x) - \bar{\tau}(x)) (\bar{\tau}(x) - \tau(x)) dx \right]}_{=0} \\ &\quad + \underbrace{\int_0^1 (\bar{\tau}(x) - \tau(x))^2 dx}_{\text{bias}}. \end{aligned}$$

The second expectation in the above is zero because $\mathbb{E}[\hat{\tau}(x)] = \bar{\tau}(x)$ for every x according to Lemma 7.

Lemma 3 (Bias Term). *Let $\bar{\tau}(x) = \sum_{n=0}^{N-1} \bar{\tau}_n \varphi(Nx - n)$, and let N be the number of pixels in a unit space $0 \leq x \leq 1$. The bias is approximately*

$$\text{bias} = \frac{c^2}{12N^2}, \quad (34)$$

where $c_n = \tau'(x_n)$ is the slope of τ at x_n , and $c^2 = \frac{1}{N} \sum_{n=0}^{N-1} c_n^2$.

✂ *Remark: This is a new proof we make for this paper.*

Proof of Lemma 3

Integrating over the unit space $[0, 1]$, we can show that

$$\begin{aligned} &\int_0^1 (\bar{\tau}(x) - \tau(x))^2 dx \\ &= \int_0^1 \left(\sum_{n=0}^{N-1} \tau_n \varphi(Nx - n) - \tau(x) \right)^2 dx \\ &= \sum_{n=0}^{N-1} \underbrace{\int_{n/N}^{(n+1)/N} (\tau_n - \tau(x))^2 dx}_{=e_n^2}. \end{aligned}$$

Let's calculate the individual error e_n^2 . Let x_n be the midpoint of n/N and $(n+1)/N$. That is,

$$x_n = \frac{1}{2} \left(\frac{n}{N} + \frac{n+1}{N} \right) = \frac{2n+1}{2N}.$$

Then, we can take the first order approximation of $\tau(x)$ around x_n :

$$\tau(x) = \tau(x_n) + \tau'(x_n)(x - x_n).$$

Since $\tau(x_n) = \tau_n$ by construction, it follows that

$$\begin{aligned} e_n^2 &= \int_{n/N}^{(n+1)/N} (\tau_n - \tau(x))^2 dx \\ &= \int_{n/N}^{(n+1)/N} \left(\tau_n - (\tau(x_n) + \tau'(x_n)(x - x_n)) \right)^2 dx \\ &= \int_{n/N}^{(n+1)/N} \left(\tau_n - (\tau_n + \tau'(x_n)(x - x_n)) \right)^2 dx \\ &= \int_{n/N}^{(n+1)/N} \left(\tau'(x_n)(x - x_n) \right)^2 dx \\ &= [\tau'(x_n)]^2 \frac{\left(\frac{n+1}{N} - x_n\right)^3 - \left(\frac{n}{N} - x_n\right)^3}{3}, \quad x_n = \frac{2n+1}{2N} \\ &= \frac{[\tau'(x_n)]^2}{12N^3}. \end{aligned}$$

Summing over all n 's will give us

$$\int_0^1 (\bar{\tau}(x) - \tau(x))^2 dx = \sum_{n=0}^{N-1} e_n^2 = \frac{c^2}{12N^2},$$

where $c^2 = \frac{1}{N} \sum_{n=0}^{N-1} c_n^2$.

Lemma 4 (Variance Term). *Assume $s(t)$ is Gaussian and $\lambda_b = 0$. The variance is approximately*

$$\text{var} = \frac{N}{\alpha_0} (c^2 \sigma_x^2 + \sigma_t^2), \quad (35)$$

where $c_n = \tau'(x_n)$ is the slope of τ at x_n , and $c^2 = \frac{1}{N} \sum_{n=0}^{N-1} c_n^2$.

✂ *Remark: This is a new proof we make for this paper.*

Proof of Lemma 4.

Firstly, we use Theorem 3 and apply it to the single-pixel case in Theorem 2. Then, under the assumption that $\lambda_b = 0$, we can follow Example 3 to show that

$$\begin{aligned}\mathbb{E}[(\hat{\tau}_n - \tau_n)^2] &= \frac{\sigma_n^2}{\alpha} = \frac{c_n^2 \sigma_x^2 + \sigma_t^2}{\alpha} \\ &= \frac{N}{\alpha_0} (c_n^2 \sigma_x^2 + \sigma_t^2).\end{aligned}$$

Therefore, for a unit length, it follows that

$$\begin{aligned}\text{var} &= \mathbb{E} \left[\int_0^1 (\hat{\tau}(x) - \bar{\tau}(x))^2 dx \right] \\ &= \mathbb{E} \left[\sum_{n=0}^{N-1} \int_{n/N}^{(n+1)/N} (\hat{\tau}(x) - \bar{\tau}(x))^2 dx \right] \\ &= \sum_{n=0}^{N-1} \int_{n/N}^{(n+1)/N} \mathbb{E} \left[(\hat{\tau}(x) - \bar{\tau}(x))^2 \right] dx \\ &= \sum_{n=0}^{N-1} \int_{n/N}^{(n+1)/N} \mathbb{E} \left[(\hat{\tau}_n - \bar{\tau}_n)^2 \right] dx \\ &= \sum_{n=0}^{N-1} \int_{n/N}^{(n+1)/N} \frac{N}{\alpha_0} (c_n^2 \sigma_x^2 + \sigma_t^2) dx \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \frac{N}{\alpha_0} (c_n^2 \sigma_x^2 + \sigma_t^2) \\ &= \frac{N}{\alpha_0} (c^2 \sigma_x^2 + \sigma_t^2).\end{aligned}$$

7. Proof of Additional Results

7.1. Proof of Corollary 1

✂ *Remark: This proof is adopted from [3].*

The probability of observing M occurrence is

$$p(M) = \int_{-T}^T dt_1 \int_{t_1}^T dt_2 \dots \int_{t_{M-1}}^T dt_M p(\mathbf{t}_M, M).$$

The integration limit comes from the fact that the time stamps follow the order $-T \leq t_1 < t_2 < \dots < t_M \leq T$. The integral is equivalent to $1/M!$ of the hypercube $(-T, T)$. Thus,

$$\begin{aligned}p(M) &= \frac{1}{M!} \int_{-T}^T dt_1 \int_{-T}^T dt_2 \dots \int_{-T}^T dt_M e^{-Q} \prod_{j=1}^M \lambda(t_j) \\ &= \frac{e^{-Q}}{M!} \prod_{j=1}^M \int_{-T}^T \lambda(t_j) dt_j = \frac{e^{-Q} Q^M}{M!}.\end{aligned}$$

7.2. Proof of Corollary 2

✂ *Remark: Bar-David hinted the result in [3], but the proof was missing. We provide the proof here.*

The sum of the integrals can be written as follows.

$$\begin{aligned}&\sum_{m=0}^M \int_{\Omega_M} p(\mathbf{t}_M, M) d\mathbf{t}_M \\ &= \sum_{m=0}^M \int_{-T}^T dt_1 \int_{t_1}^T dt_2 \dots \int_{t_{M-1}}^T dt_M p(\mathbf{t}_M, M).\end{aligned}$$

Since \mathbf{t}_M is an ordered sequence of time stamps, we can rewrite the integration limits by taking into consideration the $M!$ combinations of the orders. This will give us

$$\begin{aligned}&\sum_{m=0}^M \int_{\Omega_M} p(\mathbf{t}_M, M) d\mathbf{t}_M \\ &= \sum_{m=0}^M \int_{-T}^T dt_1 \int_{-T}^T dt_2 \dots \int_{-T}^T dt_M \frac{1}{M!} p(\mathbf{t}_M, M) \\ &= \sum_{m=0}^M \frac{e^{-Q}}{M!} \prod_{j=1}^M \left[\int_{-T}^T \lambda(t_j) dt_j \right] = \sum_{m=0}^M \frac{e^Q}{M!} Q^M = 1.\end{aligned}$$

7.3. Proof of Lemma 1

✂ *Remark: This is a new proof we make for this paper*

The KL-divergence of the two distributions is

$$\begin{aligned}\text{KL}(\varphi \parallel \phi) &= \int_{-\infty}^{\infty} \varphi(x) \log \frac{\varphi(x)}{\phi(x)} dx \\ &= \int_{-\frac{W}{2}}^{\frac{W}{2}} -\frac{1}{W} \left[\log W + \log \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \right\} \right] dx \\ &= \int_{-\frac{W}{2}}^{\frac{W}{2}} -\frac{\log W}{W} + \frac{1}{W} \log(\sqrt{2\pi\sigma^2}) + \frac{x^2}{2\sigma^2 W} dx \\ &= -\log W + \log(\sqrt{2\pi\sigma^2}) + \frac{W^2}{24\sigma^2}.\end{aligned}$$

Taking the derivative with respect to σ^2 will yield

$$\frac{d}{d\sigma^2} \text{KL}(\varphi \parallel \phi) = \frac{1}{2\sigma^2} - \frac{W^2}{24\sigma^4}.$$

Equating this to zero and rearranging terms will give us

$$\sigma^2 = \frac{W^2}{12}. \quad (36)$$

7.4. Cramer Rao Lower Bound

Corollary 3 (Cramer Rao Lower Bound). *The Cramer-Rao lower bound for the ML estimate $\hat{\tau}$ given $\mathcal{L}(\tau)$ is*

$$\begin{aligned}\mathbb{E}[(\hat{\tau} - \tau_0)^2] &\geq \left(-\mathbb{E} \left[\frac{\partial^2 \mathcal{L}}{\partial \tau^2} \right] \right)^{-1} \\ &= \left[\int_{-T}^T \frac{(\alpha \dot{s}(t))^2}{\alpha s(t) + \lambda_b} dt \right]^{-1}.\end{aligned}$$

✂ *Remark: The Cramer-Rao lower bound for single-photon LiDAR has been previously mentioned in [36]. We provide the proof here for completion.*

Proof of Corollary 3.

Let $\mathcal{L}(\tau) = \log[\alpha s(t - \tau) + \lambda_b]$. Then

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \tau} &= -\frac{\alpha \dot{s}(t - \tau)}{\alpha s(t - \tau) + \lambda_b} \\ \frac{\partial^2 \mathcal{L}}{\partial \tau^2} &= -\frac{-(\alpha s(t - \tau) + \lambda_b) \ddot{s}(t - \tau) + (\alpha \dot{s}(t - \tau))^2}{(\alpha s(t - \tau) + \lambda_b)^2} \\ &= \frac{\ddot{s}(t - \tau)}{\alpha s(t - \tau) + \lambda_b} - \frac{(\alpha \dot{s}(t - \tau))^2}{(\alpha s(t - \tau) + \lambda_b)^2}\end{aligned}$$

Taking the expectation over t , we have

$$\begin{aligned}\mathbb{E} \left[-\frac{\partial^2 \mathcal{L}}{\partial \tau^2} \right] &= -\int_{-T}^T \frac{\partial^2 \mathcal{L}}{\partial \tau^2} \cdot [\alpha s(t - \tau) + \lambda_b] dt \\ &= -\underbrace{\int_{-T}^T \ddot{s}(t - \tau) dt}_{=0} + \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau))^2}{\alpha s(t - \tau) + \lambda_b} dt \\ &= \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau))^2}{\alpha s(t - \tau) + \lambda_b} dt.\end{aligned}$$

Without loss of generality, we set $\tau = 0$. Taking the reciprocal completes the proof.

8. Auxiliary Results and Proofs

In this section, we present additional results and proofs that are essential to the development of our theories.

8.1. Lemma about Product of Functions

Lemma 5. *Consider a function $f(t)$ and define its product:*

$$f_\pi(\mathbf{t}_M) = \prod_{j=1}^M f(t_j). \quad (37)$$

Assume $\mathbf{t}_M \sim p(\mathbf{t}_M)$. The expectation of f_π is

$$\mathbb{E}[f_\pi(\mathbf{t}_M)] = e^{-Q+G(T)}, \quad (38)$$

where $G(T) = \int_{-T}^T f(t)\lambda(t)dt$.

✂ *Remark: This proof is adopted from Bar-David.*

Proof of Lemma 5.

The expectation can be shown as follows.

$$\begin{aligned}\mathbb{E}[f_\pi(\mathbf{t}_M)] &= \int_{\Omega} f_\pi(\mathbf{t}_M) p(\mathbf{t}_M) d\mathbf{t}_M \\ &= \sum_{M=0}^{\infty} \underbrace{\int_{\Omega_M} f_\pi(\mathbf{t}_M) p(\mathbf{t}_M) d\mathbf{t}_M}_{\stackrel{\text{def}}{=} \mathbb{E}_M(f_\pi)}.\end{aligned}$$

Substituting the definition of $f_\pi(\mathbf{t}_M)$ into the expression above, we can show that

$$\begin{aligned}\mathbb{E}_M(f_\pi) &= \int_{\Omega_M} f_\pi(\mathbf{t}_M) p(\mathbf{t}_M, M) d\mathbf{t}_M \\ &= e^{-Q} \int_{-T}^T dt_1 \int_{t_1}^T dt_2 \dots \int_{t_{M-1}}^T dt_M \prod_{j=1}^M g(t_j),\end{aligned}$$

where $g(t)$ is defined as $g(t) = f(t)\lambda(t)$.

Now, consider a slightly different integration

$$\begin{aligned}&e^{-Q} \int_{-T}^T dt_1 \int_{-T}^T dt_2 \dots \int_{-T}^T dt_M \prod_{j=1}^M g(t_j) \\ &= e^{-Q} \left[\int_{-T}^T g(t) dt \right]^M \\ &= e^{-Q} G^M(T),\end{aligned}$$

where $G(T) = \int_{-T}^T g(t)dt$. The difference between this integral and the previous integral is that the previous integral is based $-T \leq t_1 \leq \dots \leq t_M \leq T$. To account for the shape from this triangle to the full hypercube, we add the $M!$ factors. This will give us

$$\mathbb{E}_M(f_\pi) = \frac{e^Q G^M(T)}{M!}.$$

Therefore,

$$\mathbb{E}[f_\pi(\mathbf{t}_M)] = e^{-Q} \sum_{j=1}^M \frac{G^M(T)}{M!} = e^{-Q+G(T)}.$$

8.2. Lemma about Characteristic Function

Lemma 6. For the function $f(t; iu)$ defined as, where $i = \sqrt{-1}$,

$$f(t; iu) = \exp \left\{ iu \cdot \frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right\}. \quad (39)$$

The function $G(T) = \int_{-T}^T f(t; iu) \lambda(t) dt$ has derivatives

$$\begin{aligned} G(T) \Big|_{iu=0} &= \int_{-T}^T \lambda(t) dt = Q \\ \frac{d}{d(iu)} G(T) \Big|_{iu=0} &= \int_{-T}^T \alpha \dot{s}(t - \tau_n) dt \\ \frac{d^2}{d(iu)^2} G(T) \Big|_{iu=0} &= \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\alpha s(t - \tau_n) + \lambda_b} dt \end{aligned}$$

✂ Remark: Bar-David mentioned the idea but we could not find the proof. Therefore, the proof here is new.

Proof of Lemma 6.

For the $f(t; iu)$ defined, we can show that

$$\begin{aligned} \frac{d}{d(iu)} f(t; iu) &= \frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \cdot \exp \left\{ iu \cdot \frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right\} \\ &= \frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \cdot f(t; iu) \end{aligned}$$

If we restrict $iu = 0$, then $f(t; iu)|_{iu=0} = 1$. Therefore,

$$\begin{aligned} \frac{dG(T)}{d(iu)} \Big|_{iu=0} &= \int_{-T}^T \frac{d}{d(iu)} f(t; iu) \cdot \lambda(t) dt \Big|_{iu=0} \\ &= \int_{-T}^T f(t; iu) \cdot \frac{\alpha \dot{s}(t - \tau_n)}{\lambda(t)} \cdot \lambda(t) dt \Big|_{iu=0} \\ &= \int_{-T}^T \alpha \dot{s}(t - \tau_n) dt = 0, \end{aligned}$$

where the last equality holds whenever $s(t)$ is a symmetric function.

$$\begin{aligned} \frac{d^2 G(T)}{d(iu)^2} \Big|_{iu=0} &= \int_{-T}^T \frac{d}{d(iu)} f(t; iu) \cdot \frac{\alpha \dot{s}(t - \tau_n)}{\lambda(t)} \cdot \lambda(t) dt \Big|_{iu=0} \\ &= \int_{-T}^T f(t; iu) \cdot \left(\frac{\alpha \dot{s}(t - \tau_n)}{\lambda(t)} \right)^2 \cdot \lambda(t) dt \Big|_{iu=0} \\ &= \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\alpha s(t - \tau_n) + \lambda_b} dt. \end{aligned}$$

8.3. Lemma for μ_1 and F_1

Lemma 7. Let $F_1 = \dot{F}(\tau_n)$, where the derivative is taken with respect to τ . Write $F_1 = \mu_1 + F'_1$. It holds that

$$\mu_1 = 0, \quad \mathbb{E}[F_1'^2] = \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\alpha s(t - \tau_n) + \lambda_b} dt.$$

✂ Remark: Similar to the previous Lemma, Bar-David mentioned the idea but we could not find the proof. Therefore, the proof here is new.

Proof of Lemma 7.

Let the characteristic function of the random variable F_1 be

$$\Phi(iu) = \mathbb{E} \left[\exp \left\{ iu \cdot \sum_{j=1}^M \frac{\alpha \dot{s}(t_j - \tau_n)}{\alpha s(t_j - \tau_n) + \lambda_b} \right\} \right]. \quad (40)$$

Thus, by Lemma 5, we have $\Phi(iu) = \mathbb{E}[f_\pi(\mathbf{t}_M)] = e^{-Q+G(T)}$ with f_π defined by Lemma 6.

The first moment of F_1 is therefore

$$\begin{aligned} \mu_1 &= \frac{d}{d(iu)} \Phi(iu) \Big|_{iu=0} = \frac{d}{d(iu)} e^{-Q+G(T)} \Big|_{iu=0} \\ &= e^{-Q+G(T)} \frac{d}{d(iu)} G(T) \Big|_{iu=0} = 0, \end{aligned}$$

where the last equality is due to Lemma 6.

For $\mathbb{E}[F_1'^2]$, we can show that

$$\begin{aligned} \mathbb{E}[F_1'^2] &= \frac{d^2}{d(iu)^2} \Phi(iu) \Big|_{iu=0} \\ &= e^{-Q} \frac{d}{d(iu)} \left[e^{G(T)} \frac{d}{d(iu)} G(T) \right] \Big|_{iu=0} \\ &= e^{-Q} \left[e^{G(T)} \left(\frac{d}{d(iu)} G(T) \right)^2 + e^{G(T)} \frac{d^2}{d(iu)^2} G(T) \right] \Big|_{iu=0} \\ &= e^{-Q} \left[e^Q (0)^2 + e^Q \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\lambda(t)} dt \right] \\ &= \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\lambda(t)} dt = \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\alpha s(t - \tau_n) + \lambda_b} dt. \end{aligned}$$

Since the width of the pulse is much smaller than the interval $(-T, T)$, we can make $\tau_n = 0$ without loss of generality.

8.4. Lemma for μ_2 and F_2

Lemma 8. Let $F_2 = \ddot{F}(\tau_n)$. Write $F_2 = \mu_2 + F_2'$. We can show that

$$\mu_2 = \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\alpha s(t - \tau_n) + \lambda_b} dt,$$

$$\mathbb{E}[F_2'^2] = \int_{-T}^T \left(\frac{d}{d\tau_n} \left(\frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right) \right)^2 \lambda(t) dt.$$

✠ *Remark:* Similar to the previous Lemma, Bar-David mentioned the idea but we could not find the proof. Therefore, the proof here is new.

Proof of Lemma 8.

The characteristic function of F_2 is

$$\Phi(iu) = \mathbb{E} \left[\exp \left\{ iu \cdot \sum_{j=1}^M \frac{d}{d\tau_n} \frac{\alpha \dot{s}(t_j - \tau_n)}{\alpha s(t_j - \tau_n) + \lambda_b} \right\} \right]. \quad (41)$$

Similar to Lemma 7, we define

$$h(t) = \exp \left\{ iu \cdot \frac{d}{d\tau_n} \frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right\}. \quad (42)$$

Then, the function $H(T) = \int_{-T}^T h(t) \lambda(t) dt$, where $\lambda(t) = \alpha s(t - \tau_n) + \lambda_b$, and its derivatives are

$$H(T) \Big|_{iu=0} = \int_{-T}^T [\alpha s(t - \tau_n) + \lambda_b] dt$$

$$\frac{d}{d(iu)} H(T) \Big|_{iu=0} = \int_{-T}^T \frac{d}{d\tau_n} \left(\frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right) \lambda(t) dt$$

$$\frac{d^2}{d(iu)^2} H(T) \Big|_{iu=0} = \int_{-T}^T \left(\frac{d}{d\tau_n} \left(\frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right) \right)^2 \lambda(t) dt$$

With some simplifications, we can show that

$$\begin{aligned} \mu_2 &= \frac{d}{d(iu)} H(T) \Big|_{iu=0} \\ &= \int_{-T}^T \frac{d}{d\tau_n} \left(\frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right) \lambda(t) dt \\ &= \int_{-T}^T \left(\frac{-\lambda(t) \alpha \dot{s}(t - \tau_n) - \alpha \dot{s}(t - \tau_n) \dot{\lambda}(t)}{\lambda(t)^2} \right) \lambda(t) dt \\ &= \underbrace{- \int_{-T}^T \alpha \dot{s}(t) dt}_{=0} + \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\alpha s(t - \tau_n) + \lambda_b} dt. \end{aligned}$$

Therefore,

$$\mu_2 = \int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\alpha s(t - \tau_n) + \lambda_b} dt.$$

Similarly, we can show that

$$\begin{aligned} \mathbb{E}[F_2'^2] &= \frac{d^2}{d(iu)^2} \Phi(iu) \Big|_{iu=0} \\ &= e^{-Q} \left[e^{H(T)} \left(\frac{d}{d(iu)} H(T) \right)^2 + e^{H(T)} \frac{d^2}{d(iu)^2} H(T) \right] \Big|_{iu=0} \\ &= \left(\int_{-T}^T \frac{(\alpha \dot{s}(t - \tau_n))^2}{\alpha s(t - \tau_n) + \lambda_b} dt \right)^2 \\ &\quad + \int_{-T}^T \left(\frac{d}{d\tau_n} \left(\frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right) \right)^2 (\alpha s(t - \tau_n) + \lambda_b) dt. \end{aligned}$$

Subtracting the mean square will give us

$$\begin{aligned} \mathbb{E}[F_2'^2] &= \mathbb{E}[F_2^2] - \mu_2^2 \\ &= \int_{-T}^T \left(\frac{d}{d\tau_n} \left(\frac{\alpha \dot{s}(t - \tau_n)}{\alpha s(t - \tau_n) + \lambda_b} \right) \right)^2 (\alpha s(t - \tau_n) + \lambda_b) dt. \end{aligned}$$

9. Detailed Setups of Experiments

For the experiments we presented in the main text, the parameters are configured as Table 1. The parameters here are unit-free in the sense they are picked to elaborate the theory.

Table 1. Parameters used in 1D experiments

N	number of pixels	
x	spatial coordinate	$0 \leq x \leq 1$
t	temporal duration	$0 \leq t \leq 10$
ΔN	width of each pixel	$1/N$
Δx	spatial grid spacing	$1/2048$
Δt	temporal grid spacing	$1/256$
σ_t	pulse width	0.5
α_0	overall scene flux	10000
λ_b	noise floor	0
σ_x	spatial Gaussian radius	$1/(\sqrt{12}N)$
c_n	n th pixel of $\tau'(x)$	use gradient

9.1. Ground Truth Time-of-Arrival Function

The true time of arrival function $\tau(x)$ can be any function defined on $0 \leq x \leq 1$. For example, in the main text, we consider a two-step scene with a smooth transition. Therefore, we can pick a sigmoid function of the following form.

$$\tau(x) = \frac{4}{1 + e^{-20(x-0.5)}} + 4, \quad (43)$$

The shape for this particular model is shown in Fig. 11 below. We stress that this is just one of the many models we can pick.

To compute the gradient $\tau'(x)$, the computationally efficient way is to perform finite difference instead of the analytic form. In MATLAB, this can be done using the command `gradient`.

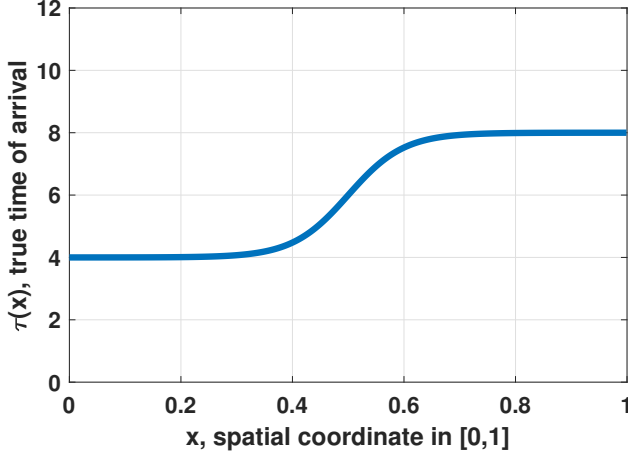


Figure 11. The ground truth time of arrival function $\tau(x)$ as a function of the coordinate $0 \leq x \leq 1$.

```
tau = 4./(1+exp(-20*(x-0.5)))+4;
tau_slope = gradient(tau)/dx;
```

The division by dx at the end is a reflection of the finite difference operation. For example, at x_k , the true gradient is

$$\tau'(x_k) = \lim_{\Delta x \rightarrow 0} \frac{\tau(x_k + \Delta x) - \tau(x_k)}{\Delta x} \approx \frac{\tau(x_{k+1}) - \tau(x_k)}{\Delta x}.$$

Since finite difference `gradient` only computes $\tau(x_{k+1}) - \tau(x_k)$, we need to divide the result by Δx to compensate for the spatial interval.

9.2. What if we use a different $\tau(x)$?

A critical observation of our theory is that the MSE can be decomposed into bias and variance. The variance has some subtle but important dependency on the shape of τ , but the bias term will experience more impact.

According to Lemma 3, the bias can be approximated by

$$\text{bias} = \int_0^1 [\tau(x) - \bar{\tau}(x)]^2 dx \approx \frac{c^2}{12N^2}, \quad (44)$$

where $c^2 = \frac{1}{N} \sum_{n=0}^{N-1} c_n^2$, with $c_n = \tau'(x_n)$ being the gradient of τ . Therefore, as τ changes, c^2 will change accordingly. Figure 12 illustrates two examples. In the same figure, we also plot the theoretically predicted bias and the simulated result. When $\tau(x)$ is relatively smooth, the theory has an excellent match with the simulation.

Limitations. With no surprise, the piecewise constant approximation has limitations. (1) When the function $\tau(x)$ is intrinsically discontinuous, the theory will not be able to

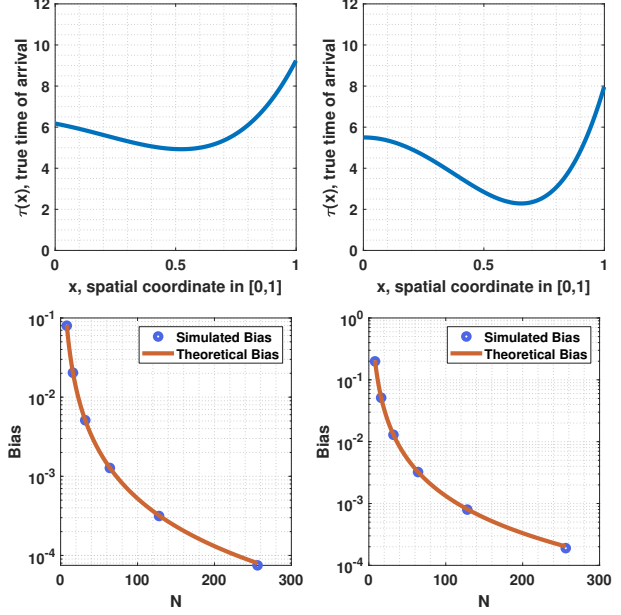


Figure 12. $\tau(x)$ and the bias.

keep track of the trend. (2) When the function $\tau(x)$ is noisy, then the gradient will be overestimated. Both will lead to degradation of the bias estimation, as illustrated in Fig. 13.

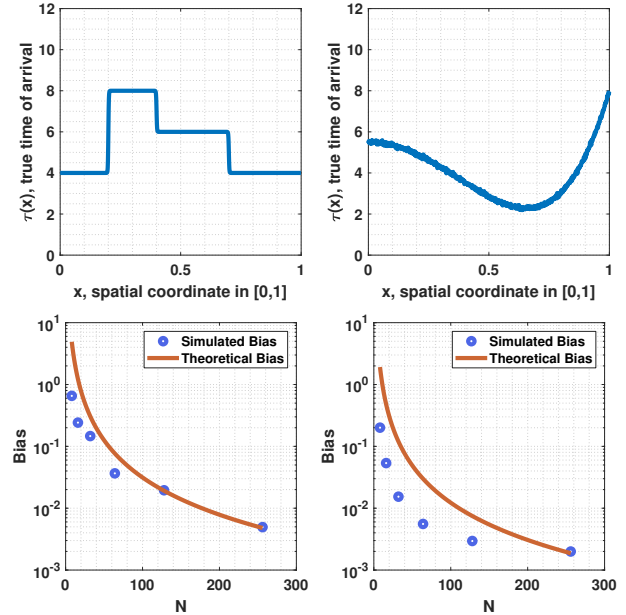


Figure 13. Limitations of the approximations we presented in this paper. In the presence of discontinuous edges, and/or in the presence of noise, the performance of the theoretical prediction degrades.

Not all limitations can be analytically mitigated. For ex-

ample, mitigating the discontinuity requires us to know the location of the discontinuous points. In practice, the likely solution is to numerically integrate the error. However, we will not be able to retain a clean, interpretable, and elegant formula.

For the noise problem, we can model the true time of arrival function as

$$\tau(x) = \tau_0(x) + \eta(x),$$

where $\eta(x)$ is a random process denoting the noise, e.g., $\eta(x) \sim \mathcal{N}(0, \sigma_e^2)$. In this case, the bias will become

$$\text{bias} = \sum_{n=0}^{N-1} \int_{\frac{n}{N}}^{\frac{n+1}{N}} [\tau'(x_n)(x - x_n) + e(x)]^2 dx.$$

With some calculations, we can show that

$$\text{bias} = \frac{c^2}{12N^2} + \sigma_e^2,$$

where $c^2 = \frac{1}{N} \sum_{n=0}^{N-1} c_n^2$, with $c_n = \tau'_0(x_n)$ being the slope of the *clean* signal $\tau_0(x)$. Therefore, if we are given the noisy time of arrival $\tau(x)$, we first need to recover the clean $\tau_0(x)$ so that we can estimate the correct c^2 . Then, the noise variance σ_e^2 is added to compensate for the noise term. The correction can be visualized in Fig. 14.

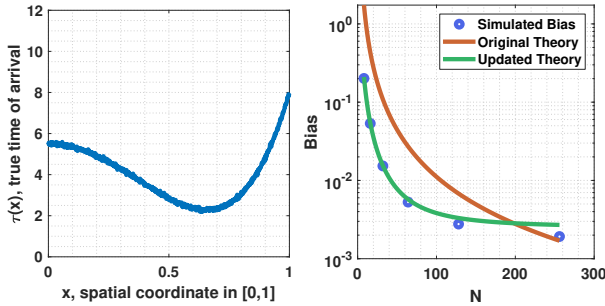


Figure 14. When the true time of arrival is noise, it is possible to obtain a better bias estimate by compensating for the noise. However, this is rather an *oracle* scenario because in practice, we never know the true noise-free $\tau(x)$ and we never know the probability distribution of noise.

10. Unit Conversion

For our theoretical analysis to match a realistic sensor, a conversion between the size of a physical pixel and the size of a point in the numerical grid needs to be established. In what follows, we present a specific example with some specific numbers. These can be easily translated to other configurations.

Suppose that the overall size of the SPAD array is 10mm as shown in Fig. 15 below. This 10mm would correspond

to 1 unit space. Suppose that we use 1024 points in the numerical grid to measure the array. Then, the width of each pixel is

$$\Delta x \stackrel{\text{def}}{=} 1 \text{ pixel} = \frac{10\text{mm}}{1024} = 9.76\mu\text{m} = \frac{1}{1024} \text{ unit space}.$$

For this paper's analysis, let's assume that we group 32 points (for example) as one *super-pixel*. The width of each super-pixel is therefore

$$1 \text{ super-pixel} = 32 \text{ pixels} = 312.5\mu\text{m} = \frac{32}{1024} \text{ unit space}.$$

Now, if we want to use a Gaussian to approximate the boxcar kernel, then the standard deviation of the Gaussian should be

$$\begin{aligned} \sigma_x &= \frac{1 \text{ super-pixel}}{\sqrt{12}} \\ &= 9.2372 \text{ pixels} = 90.21\mu\text{m} = \frac{32}{1024\sqrt{12}} \text{ unit space}. \end{aligned}$$

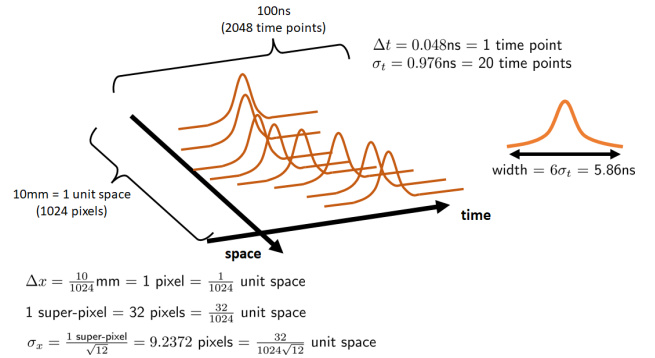


Figure 15. Conversion between the physical units to the number of grid points used on a computer.

For the temporal axis, we take a similar approach. Suppose that the total duration of the measurement is 100ns, and we use 2048 time points to measure the total duration. Then, each time point would correspond to

$$\Delta t = 1 \text{ time point} = \frac{100\text{ns}}{2048} = 0.048\text{ns}.$$

A typical pulse has a width of around 2ns-6ns. So, suppose that we define the pulse standard deviation as $\sigma_t = 20$ time points, then

$$\sigma_t = 0.976\text{ns} = 20 \text{ time points}.$$

Since $6\sigma_t$ of a Gaussian can capture 99% of the energy, we can safely say that the width of the pulse is around

$$\text{pulse width} = 6\sigma_t = 120 \text{ time points} = 5.86\text{ns},$$

which is reasonably inside the 2ns-6ns range.

11. Sampling Procedure

In the main text, we outlined a procedure to generate time stamps according to a Gaussian pulse. This section discusses how to extend the idea to arbitrary pulse shapes.

11.1. Inverse CDF Method

When the pulse $\lambda(t)$ has a completely arbitrary shape and when the noise floor is not a constant, we will not be able to draw samples from distributions with known formulae. In this case, the sampling can be done using the inverse cumulative distribution function (inverse CDF) technique [5, Ch.4.9].

The concept of sampling from a known PDF is to leverage the cumulative distribution function (CDF) and a uniform random variable. A classical result shows the following. Suppose that there is a random variable X generated from a distribution with a CDF F_X . If $U \sim \text{Uniform}[0, 1]$, and if we send U to the inverse CDF F_X^{-1} , then the transformed random variable $F_X^{-1}(U)$ will follow the distribution F_X .

The specific steps to perform the inverse CDF are as follows.

- Step 1: Compute the CDF $\Lambda(t) = \int_{-\infty}^t \lambda(r)dr$. Assuming that $\Lambda(t)$ is invertible, compute the inverse mapping Λ^{-1} . To obtain $\Lambda^{-1}(p)$ for a given p , we numerically build a lookup table.
- Step 2: Compute the number of samples $M \sim \text{Poisson}(Q)$. Then, generate random samples by sending uniform random variables: $t_j = \Lambda^{-1}(U_j)$, where $U_j \sim \text{Uniform}[0, 1]$, and $j = 1, \dots, M$. Then $\{t_j\}_{j=1}^M$ will follow the probability distribution $\lambda(t)$.

11.2. Implementation and Demonstration

In terms of implementation, we can use the following MATLAB code. Suppose that we are given the received pulse $\lambda(t) = \alpha s(t - \tau) + \lambda_b$. Then, we can integrate $\lambda(t)$ to obtain the CDF $\Lambda(t) = \int_{-\infty}^t \lambda(r)dr$. On a computer, the command `cumsum` will serve the purpose of generating the CDF. Once the CDF is numerically determined, we send M uniform random variables to Λ^{-1} . The inversion is performed numerically by finding the closest t_j such that $t_j = \Lambda^{-1}(p)$ can give $\Lambda(t_j) = p$, where p is the uniform random variable.

```
function time_stamps = ...
    generate_time_stamps(lambda,t,M)

    c = cumsum(lambda)/sum(lambda);
    p = rand(1,M);
    c = repmat(c(:), [1,M]);
    [~,pos] = min(abs(c-p));
    time_stamps = t(pos);
end
```

Fig. 16 shows a demonstration where we generate M time stamps. The underlying pulse shape is Gaussian, but we also assume a nonlinear noise floor. Specifically, the noise floor mimics a scattering medium which is modeled by a Gamma distribution.

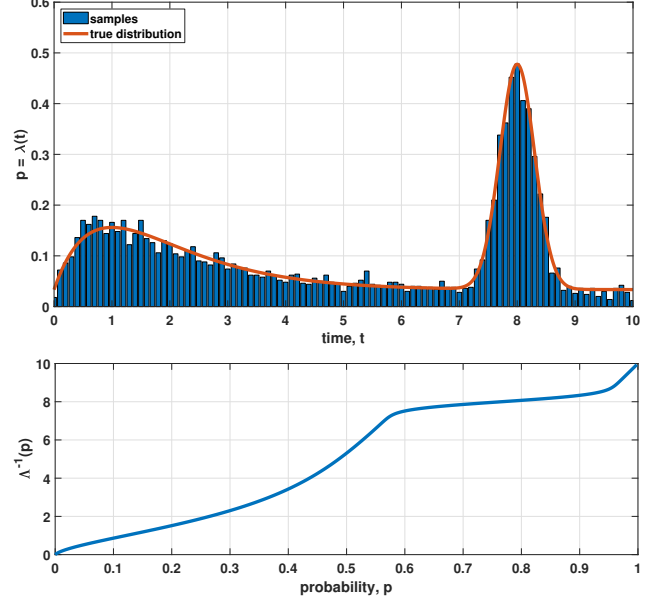


Figure 16. [Top] Random samples t_M drawn from a pulse $\lambda(t)$ which consists of a non-uniform noise floor and a Gaussian pulse. [Bottom] The inverse CDF $\Lambda^{-1}(p)$ is used to generate the samples from a uniform distribution. $\Lambda^{-1}(p)$ is generated numerically from the known equation $\lambda(t)$.

```
s1 = 0.2*pdf('normal', t, tau, sigma_t);
s2 = 0.2*pdf('gamma', t, 2, 1);
lambda_b = 0.02*ones(1,length(t));
lambda = s1+s2+lambda_b;

lambda_pdf = lambda/sum(lambda*dt);
lambda_cdf = cumsum(lambda_pdf);
dp = 1/500;
p = 0:dp:1-dp;
c = repmat(lambda_cdf(:), [1,length(p)]);
[~,pos] = min(abs(c-p));
lambda_icdf = t(pos);
```

12. Non-zero Noise Floor

The goal of the main text is to derive a simple and informative MSE equation. This is achieved by assuming that the pulse is Gaussian and there is no noise. In this section, we discuss how the theoretical results can be derived for arbitrary pulses, by means of a numerical scheme.

12.1. Maximum Likelihood Estimation

Assuming that the transmitted pulse is $s(t)$ and the received pulse is $\lambda(x, t) = \alpha s(t - \tau(x)) + \lambda_b$. Here, we assume that

$s(t)$ can take an arbitrary shape and $\lambda_b > 0$. We also assume that there is a spatial kernel $\phi(x)$ which will be applied to $\lambda(x, t)$ to yield the resulting space-time function $\tilde{\lambda}(x, t)$:

$$\tilde{\lambda}(x, t) = \phi(x) \otimes \lambda(x, t).$$

The convolution \otimes can be implemented numerically. In MATLAB, we can use the command `imfilter`:

```
lambda = alpha*s + lambda_b
h = ones(dN,1)/dN;
lambda_tilde = imfilter(lambda, h, ...
    replicate');
```

Following the same derivations as the main text, we let $x_n = (2n + 1)/(2N)$ be the mid point of each spatial interval and define the effective pulse $\tilde{\lambda}_n(t)$ as

$$\tilde{\lambda}_n(t; \tau_n) = \tilde{\lambda}(x_n, t; \tau_n),$$

where we emphasize that $\tilde{\lambda}_n$ has an underlying parameter τ_n which needs to be estimated. Then, the time of arrivals $\mathbf{t}_M = [t_1, \dots, t_M]$ will follow the distribution specified by the effective return pulse $\tilde{\lambda}_n(t)$.

The questions we need to ask now are (1). how to estimate the time of arrival $\hat{\tau}$, (2). what is the variance $\mathbb{E}[(\hat{\tau}_n - \tau_n)^2]$?

For the purpose of deriving theoretical bounds, the estimator we use is the maximum likelihood estimator. The ML estimator is the one that maximizes the likelihood function. In our case, it is

$$\hat{\tau}_n = \underset{\tau_n}{\operatorname{argmax}} \underbrace{\sum_{j=1}^M \log \tilde{\lambda}_n(t_j; \tau_n)}_{\mathcal{L}(\tau_n)}, \quad (45)$$

The shape of the likelihood function $\mathcal{L}(\tau_n)$ as a function of the parameter τ_n is shown in Fig. 17, for a typical Gaussian pulse with a non-zero noise floor. The likelihood function has a clear maximum (minimum if we take the negative log likelihood) around the true parameter. For the specific example shown in Fig. 17, we use the following configurations: $\alpha = 100$, $\lambda_b = 30$, $s(t, \tau_0) = \mathcal{N}(t | \tau_0, \sigma_t^2)$ where $\tau_0 = 5$ and $\sigma_t = 0.5$. We randomly generate $K = 100$ random vectors $\mathbf{t}_M^{(k)}$ for $k = 1, \dots, K$. We plot the negative log-likelihood $\mathcal{L}(\tau_n | \mathbf{t}_M^{(k)})$ as a function of the parameter τ_n . The negative log-likelihood functions are random because $\mathbf{t}_M^{(k)}$ are random. We take the expectation of these $K = 100$ random negative log-likelihood functions to visualize the mean function.

An equivalent alternative approach is to perform the estimation by solving a zero-finding problem. Taking the derivative of \mathcal{L} with respect to τ_n , we can show that the

solution $\hat{\tau}_n$ must satisfy the equation

$$\left. \frac{d\mathcal{L}}{d\tau_n} \right|_{\tau_n=\hat{\tau}_n} = \sum_{j=1}^M \frac{\dot{\tilde{\lambda}}_n(t_j; \tau_n)}{\tilde{\lambda}_n(t_j; \tau_n)} \bigg|_{\tau_n=\hat{\tau}_n} = 0. \quad (46)$$

Then, the maximum likelihood estimate $\hat{\tau}_n$ is determined by finding the zero-crossing of the derivative such that $\dot{\mathcal{L}}(\hat{\tau}_n) = 0$.

Example 4. Suppose that $\tilde{\lambda}_n(t; \tau_n) = \alpha s(t - \tau_n) + \lambda_b$, then the likelihood function is

$$\mathcal{L}(\tau_n) = \sum_{j=1}^M \log \{ \alpha s(t_j - \tau_n) + \lambda_b \}.$$

The derivative of the likelihood is

$$\frac{d\mathcal{L}}{d\tau_n} = - \sum_{j=1}^M \frac{\alpha s(t_j - \tau_n)}{\alpha s(t_j - \tau_n) + \lambda_b}.$$

The shape of the likelihood derivative can be seen in Fig. 18, for a typical example outlined in the description of Fig. 17. The maximum likelihood estimate is the zero-crossing of the derivative. We emphasize that for arbitrary pulse shape and a non-zero floor noise, neither the zero-finding approach nor the matched filter approach would have an analytic solution. We will discuss the numerical implementation shortly.

12.2. MSE Calculation

Given the effective return pulse $\tilde{\lambda}_n(t)$, the MSE follows from Theorem 2. To make our notations consistent, we as-

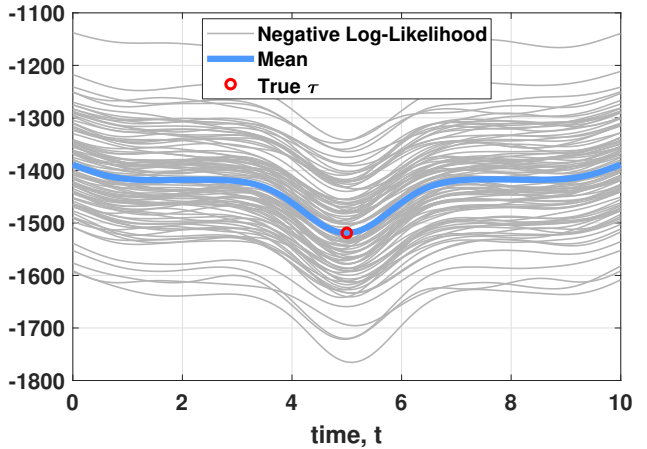


Figure 17. The shape of the negative log-likelihood function $\mathcal{L}(\tau)$ as a function of the unknown parameter τ .

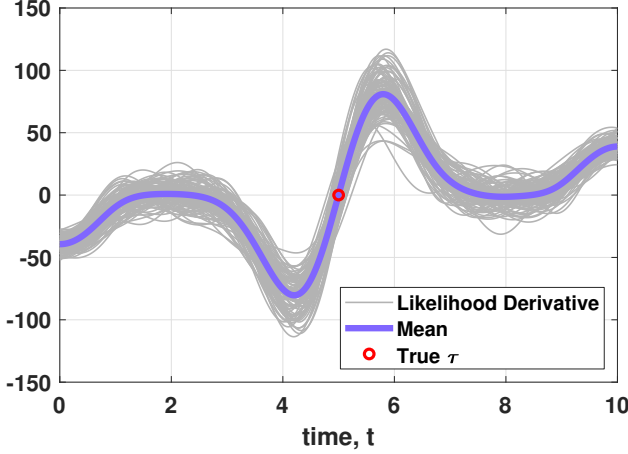


Figure 18. The shape of the negative log-likelihood function's derivative $\frac{d}{d\tau}\mathcal{L}(\tau)$ as a function of the unknown parameter τ .

sume that $\tilde{\lambda}_n(t)$ takes the form

$$\tilde{\lambda}_n(t) = \alpha s_n(t - \tau_n) + \lambda_b, \quad (47)$$

for some function $s_n(t)$. This assumption is valid whenever α and λ_b are independent of x so that any spatial convolution applied to construct $\tilde{\lambda}(t)$ will not affect α . Under this notation, we follow Theorem 2 and write the variance of the estimator

$$\mathbb{E}[(\hat{\tau}_n - \tau_n)^2] = \left[\int_{-T}^T \frac{(\alpha \dot{s}_n(t))^2}{\alpha s_n(t) + \lambda_b} dt \right]^{-1}, \quad (48)$$

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

For arbitrary pulse shape, $s_n(t)$ does not have any analytic expression, so does $\dot{s}_n(t)$. Nevertheless, the integration in Eq. (49) can still be performed numerically. The MATLAB code below shows how we can do it numerically.

```
tau = 4./(1+exp(-20*(x-0.5)))+4;
tau_slope = gradient(tau)/dx;
for n=1:N
    c(n) = mean(tau_slope(dN*(n-1)+[1:dN]));
    s_dot(n,:) = gradient(s(n,:))/dt;
    var_n(n) = 1/sum((alpha*s_dot(n,:)).^2) ...
        ./ (alpha*s(n,:)+lambda_b)*dt);
end
var_theory = mean(var_n);
bias_theory = mean((c.^2)/(12*N^2));
mse_theory = bias_theory + var_theory;
```

With Eq. (48) defined, the overall MSE as a function of the number of pixels is as follows.

Theorem 5. For arbitrary pulse shape and noise floor, the MSE takes the form

$$MSE = \frac{c^2}{12N^2} + \frac{N}{\alpha_0} \mathbb{E}[(\hat{\tau}_n - \tau_n)^2]. \quad (49)$$

Here, we use the fact that $\mathbb{E}[(\hat{\tau}_n - \tau_n)^2]$ has the same value for all n so it does not matter which n we use to calculate the variance. Although Eq. (49) is no longer a closed-form expression, it still preserves the shape of the resolution limit where the MSE decays quadratically with respect to $1/N^2$ and increases linearly with respect to N .

12.3. Solving the ML Estimation

In this subsection, we explain how to implement the maximum likelihood estimation for an arbitrary pulse and noise floor. There are three ways to do this.

Approach 1: Gradient-based Likelihood Maximization.

For pulses with a known analytic expression, the most straightforward approach is to perform gradient descent with the gradient $\frac{d}{d\tau}\mathcal{L}$:

$$\tau^{(k+1)} = \tau^{(k)} - \gamma^{(k)} \frac{d}{d\tau}\mathcal{L}(\tau^{(k)}),$$

for some step size $\gamma^{(k)}$. Most numerical software such as MATLAB has built-in optimization packages to accomplish this step. For example, using `fminunc`, we will be able to find the maximum of the likelihood function.

The example below assumes a Gaussian pulse so that we have a closed-form expression for the likelihood function.

```
function L = myL(tau, tj, ...
    sigma_t, alpha, lambda_b)
L = -sum(log(alpha * ...
    1/sqrt(2*pi*sigma_t^2) * ...
    * exp(-(tj-tau).^2/(2*sigma_t^2)) + ...
    + lambda_b));
```

To solve the actual maximization we do

```
fun = @(tau) myL(tau, tj, ...
    sigma_t, alpha, lambda_b);
tau_hat = fminunc(myL, tau0, option);
```

where `tau0` is the initial guess. For the purpose of deriving the “oracle” maximum likelihood estimate (so that we are not penalized by having a bad algorithm), we use the ground truth τ_0 as the initial guess. Readers may worry that this would return us the ground truth τ_0 as the maximum likelihood estimate, i.e., $\hat{\tau} = \tau_0$. We note that this will never happen because the likelihood \mathcal{L} needs to fit the measured time stamps t_M and so \mathcal{L} is a random function. Since \mathcal{L} is a random function, the maximum location is a random variable too. Therefore, while the expected value of the estimate $\hat{\tau}$ is the true τ_0 , for a particular realization $\hat{\tau}$ will never be the same as τ_0 .

Approach 2: Search-based Likelihood Maximization.

For arbitrary pulse shape, the gradient-based approach is difficult to implement because the pulse $s(t - \tau)$ is numerically a different function when we use a different τ . The viable approach, as we mentioned in the previous subsection, is to run a matched filter. A matched filter requires

us to shift the known pulse to the left or to the right until we see the best fit to the data. On computers, we need to perform two steps:

- Given a current estimate τ and a perturbation $\Delta\tau$, translate it to the time index of the numerical array $\tilde{\lambda}_n(t; \tau)$.
- Shift the index based on the amount corresponding to $\Delta\tau$, and calculate $\tilde{\lambda}_n(t; \tau + \Delta\tau)$.

In MATLAB, the shifting operation can be implemented using the commands below.

```
function L = myL(tau, tj, lambda_bar_n, ...
    t, tau0)
lengthT = length(t);
[~, pos_new] = min(abs(t-tau));
[~, pos_ini] = min(abs(t-tau0));
lambda1_pad = [lambda_bar_n(1)...
    *ones(1,lengthT), ...
    lambda_bar_n, ...
    lambda_bar_n(end) ...
    *ones(1,lengthT)];
lambda2_pad = circshift(...
    lambda1_pad, ...
    pos_new-pos_ini);
lambda2 = lambda2_pad(...
    lengthT+1:2*lengthT);
L = -sum(log(interp1(t, ...
    lambda2, tj, 'spline')));
```

The padding is a band-aid for a MATLAB specific command `circshift` which circularly shifts the indices. If we do not pad the array appropriately, a pulse located near the end of the time axis will cause erroneous values to the likelihood function after they get circularly flipped to the beginning of time.

Because of the shifting operations defined above, the function call is significantly harder for automatic differentiation unless we customize the step size. The reason is that if the default step size is smaller than the temporal grid we used to define the pulse, two adjacent indices will remain the same. Thus the gradient will be zero. A workaround solution is to go with a non-gradient optimization. For MATLAB, we can use the package `fminsearch`.

```
fun = @(tau) myL(tau, tj, lambda_n, t, tau0);
tau_hat = fminsearch(fun, t0, option);
```

Approach 3: Zero-finding of the Likelihood Gradient.

The third approach we can use is to directly solve for the maximum likelihood solution. Recall from Eq. (46) that the maximum likelihood estimate must be the zero-crossing point of Eq. (46), we can directly implement the derivative as a function call such as the example below.

```
function L = myL_dt(tau, tj, sigma_t, ...
    alpha, lambda_b)
my_s = alpha ...
    * (1/sqrt(2*pi*sigma_t^2)) ...
    * exp(-(tj-tau).^2/(2*sigma_t^2)) ...
    + lambda_b;
my_s_dot = -alpha*(1/sqrt(2*pi*sigma_t^2)) ...
```

```
* exp(-(tj-tau).^2/(2*sigma_t^2)) ...
.* ((tj-tau)./(sigma_t^2));
L = sum(my_s_dot./my_s);
```

Then, we can use `fzero` to find the zero-crossing point.

```
fun = @(tau) myL_dt(tau, tj, sigma_t, ...
    alpha, lambda_b);
tau_hat = fzero(fun, tau0);
```

Which approach is faster? Since the goal of this paper is not to provide an algorithm, we skip a formal complexity analysis of the methods. To give readers a rough idea of the comparison between the three approaches, on a typical experiment using the same machine, the runtime is as follows.

	Method	Command	Runtime
1	Gradient descent	<code>fminunc</code>	0.064 sec
2	Matched filter	<code>fminsearch</code>	0.034 sec
3	Zero crossing	<code>fzero</code>	0.016 sec

We remark that all three methods give nearly identical solutions up to the precision of the numerical grid we set.

12.4. Experimental Results

To convince readers that Eq. (49) can be numerically implemented and matches with simulation, we show in Fig. 19 the comparison between the simulation and the theoretical prediction for a range of λ_b values. As is evident from the plot, the theoretical prediction matches very well with the simulation.

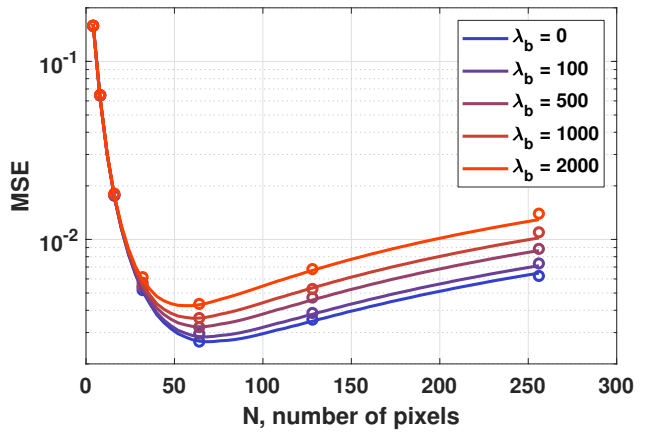


Figure 19. Excellent match between the theoretical bound and the simulation for various levels of ambient noise.

As we have demonstrated in this example, it is sometimes possible to numerically compute the bias and variance so that the theory will match with the simulation. However, by doing so, we will no longer be able to write down the resolution limit in a simple and interpretable closed-form

expression. This is not a deficiency of our theory, it is just a sacrifice of clarity and interpretability in exchange for better theoretical precision.

13. Pile-up Effects

Pile up effects refer to the situation where the photo detector responds earlier than the actual arrival of the pulse signal, typically due to the presence of a strong background.

13.1. Distribution

For simplicity, we model the pile up effect as an exponentially decaying distribution in the background, that is,

$$\lambda(x, t) = \alpha s(t - \tau(x)) + \beta \lambda_p(t) + \lambda_b, \quad (50)$$

where $\lambda_p(t) = \gamma e^{-\gamma t}$ is an exponential function parameterized by the reflected laser event rate γ , and the background event rate λ_b [40]. In this case, the sampling of the time stamps will follow from three steps:

- Draw $M_s = \text{Poisson}(\alpha)$ samples. The distribution of these M_s samples is the shape of the pulse $s(x, t)$. The parameter α specifies the average number of incident signal photons.
- Draw $M_p = \text{Poisson}(\beta\gamma)$ samples. These M_p samples follow the distribution $\text{Exponential}(\gamma)$ (whose mean is $1/\gamma$). The parameter $\beta\gamma$ specifies the average number of pile-up photons (coming from the background).
- Draw $M_b = \text{Poisson}(\lambda_b T)$ samples, assuming that $0 \leq t \leq T$. These M_b samples follow the distribution $\text{Uniform}(\lambda_b T)$.

An example of the time stamp histogram is shown in Fig. 20.

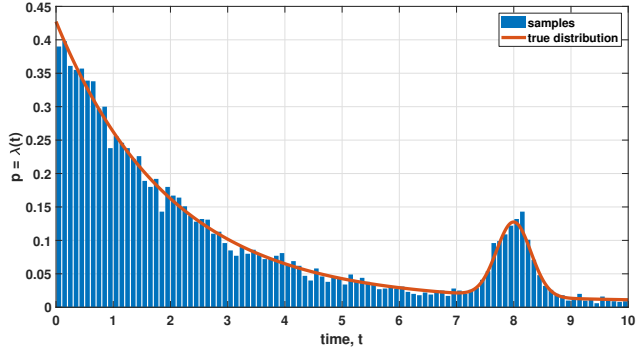


Figure 20. Pile up effect can be modeled as an exponential decaying distribution. Added to the signal pulse $s(t - \tau)$, pile up effect makes the ML estimation of the true delay significantly harder.

13.2. MLE

The maximum likelihood estimation in the presence of the pile up effect needs a separation of background and signal. Because of the simplified model we choose (for the purpose

of theoretical illustration), we can exploit the oracle knowledge we know about the pile up and background. Given $\lambda(x, t)$, we know that the binning is such that

$$\begin{aligned} \lambda_n(t) &= \int_{\frac{n}{N}}^{\frac{n+1}{N}} \lambda(x, t) dx \\ &= \underbrace{\alpha \int_{\frac{n}{N}}^{\frac{n+1}{N}} s(t - \tau(x)) dx}_{=s_n(t - \tau_n)} + \int_{\frac{n}{N}}^{\frac{n+1}{N}} [\beta \lambda_p(t) + \lambda_b] dx \\ &= \alpha s_n(t - \tau_n) + \frac{\beta \lambda_p(t) + \lambda_b}{N}, \end{aligned}$$

Where we approximate the integral by a simplified function $s_n(t)$ delayed by a time constant τ_n .

Given this decomposition of $\lambda_n(t)$ as the signal $s_n(t)$ and the noise part $\beta \lambda_p(t) + \lambda_b$, we use Theorem 2 to estimate the variance of the n th pixel:

$$\mathbb{E}[(\hat{\tau}_n - \tau_n)^2] = \int_{-T}^T \frac{[\alpha \dot{s}_n(t)]^2}{\lambda_n(t)} dt.$$

This integration needs to be evaluated numerically.

In terms of simulation, we need to run the ML estimation. The ML estimation requires us to search for an optimal τ_n such that the candidate distribution matches with the measurement. To this end, we follow the procedure outlined in Sec. 12.3 to numerically run a search scheme to pick up the ML estimate.

13.3. Theoretical MSE

A comparison between the simulated MSE and the theoretically predicted MSE is shown in Fig. 21. For this particular example, we set $\alpha = 1 \times 10^4$, $\beta = 1 \times 10^5$, and $\lambda_b = 100$. The pile up distribution has a constant $\gamma = 4$. Again, we emphasize that these units are unit-free, in the sense they are chosen to illustrate the validity of the theorem rather than matching any particular real sensor.

As we can see in Fig. 21, the theoretical prediction matches extremely well with the simulated MSE. The small deviation towards a larger N is due to the fact that the ML estimation faces difficulty when the total number of samples is few. As an example, when $N = 256$, the number of samples per pixel is around 400 whereas when $N = 8$, the number of pixels is around 28,000. When there is no pile up noise, the variance of ML estimate is basically limited by the number of samples. But in the presence of complicated noise (such as pile up), the ML estimation procedure has limitations where it cannot differentiate signal from noise. Therefore, when the number of samples is few, the simulation performance will be worse than what the theory predicts. Nevertheless, Fig. 21 still shows a high degree of match between the theory and the simulation.

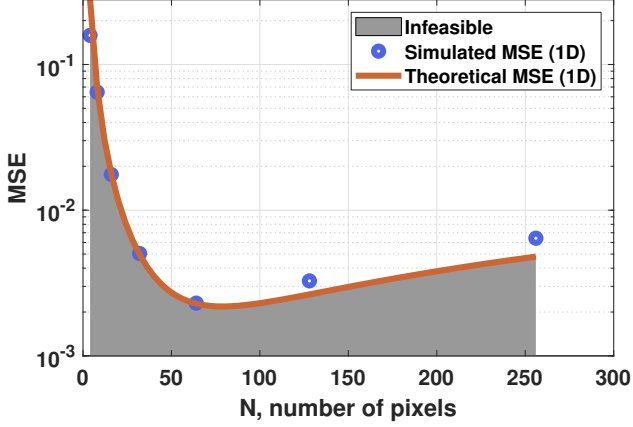


Figure 21. MSE comparison in the presence of pile up. The theoretically predicted MSE has a very good match with the simulated MSE.

14. Dark Count

The model we presented in the main text does not consider dark count. In this section, we briefly discuss how it can be included.

Dark count is a type of sensor noise caused by the random generation of electrons in the depletion region. These random electrons are present even when there is no photoelectric event. Therefore, one way to model dark count is to treat it as a uniform noise floor that does not change over time and does not depend on the scene flux. However, dark counts scale with the sensing area. So, if there are N pixels and if we assume that the sensing area scales linearly with N , then the dark counts will be reduced by N times on average. Therefore, a simple model in the context of this paper is to write

$$\lambda(t) = \frac{\alpha}{N} s(t - \tau) + \frac{\lambda_b}{N} + \frac{\lambda_{\text{dark}}}{N}. \quad (51)$$

Because of this simple addition to the background, we can treat λ_{dark} as part of the background noise. In a typical scenario when the dark count rate is in the order 10 per second while a SPAD can count up to 1 million photons per second (e.g., ThorLab's single-photon detectors), the impact of dark current is not visible unless we operate it in extremely low-light.

15. Extension to 2D

In this section, we briefly summarize how the results can be extended to 2D. We will derive the bias and the variance.

15.1. Bias

For the bias term, in 2D, we need to consider a 2D coordinate $\mathbf{x} \in [0, 1] \times [0, 1]$. The time of arrival function τ is

denoted as $\tau(\mathbf{x})$. The average function $\bar{\tau}(\mathbf{x})$ is

$$\bar{\tau}(\mathbf{x}) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \tau_{m,n} \varphi(Mx - m) \varphi(Ny - n),$$

Where $\tau_{m,n}$ is the (m, n) th value defined as

$$\tau_{m,n} = \int_{\frac{m}{M}}^{\frac{m+1}{M}} \int_{\frac{n}{N}}^{\frac{n+1}{N}} \tau(x, y) dx dy.$$

The derivative of the time of arrival function, in the 2D case, will become the gradient. This means that

$$\nabla_{\mathbf{x}} \tau(\mathbf{x}) = \left[\frac{\partial \tau}{\partial x}, \frac{\partial \tau}{\partial y} \right]^T.$$

For notational simplicity, we write $\nabla_{\mathbf{x}} \tau(\mathbf{x}) = \mathbf{c} = [c^x, c^y]^T$. If we are interested in the gradient of the (m, n) th pixel, we write

$$\mathbf{c}_{m,n} = [c_{m,n}^x, c_{m,n}^y]^T = \nabla_{\mathbf{x}} \tau(\mathbf{x}_{m,n}).$$

The magnitude square of the gradient is denoted as $\|\mathbf{c}_{m,n}\|^2 = (c_{m,n}^x)^2 + (c_{m,n}^y)^2$, and the overall gradient summed over all the pixels is

$$\|\mathbf{c}\|^2 = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \|\mathbf{c}_{m,n}\|^2 \approx \int_0^1 \int_0^1 \|\nabla_{\mathbf{x}} \tau(\mathbf{x})\|^2 d\mathbf{x},$$

where the last approximation holds by assuming that (m, n) covers the unit squares.

The main result for the bias term in the 2D case is summarized in the lemma below.

Lemma 9. *Let $\tau(\mathbf{x})$ be a 2D time of arrival function over the unit square $\mathbf{x} \in [0, 1] \times [0, 1]$. Suppose that we use $M \times N$ pixels to approximate the unit square, and for simplicity assume that $M = N$. If we approximate $\tau(\mathbf{x})$ using a piecewise constant function $\bar{\tau}(\mathbf{x})$, the bias is given by*

$$\text{Bias} = \frac{\|\mathbf{c}\|^2}{12N^2}. \quad (52)$$

Proof of Lemma 9.

We start with the definition of the bias:

$$\begin{aligned} \text{Bias} &= \int_0^1 \int_0^1 (\bar{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2 d\mathbf{x} \\ &= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \underbrace{\int_{\frac{m}{M}}^{\frac{m+1}{M}} \int_{\frac{n}{N}}^{\frac{n+1}{N}} (\tau_{m,n} - \tau(x, y))^2 dx dy}_{=e_{m,n}^2} \end{aligned}$$

Let $\mathbf{x}_{m,n}$ be the mid-point such that

$$\mathbf{x}_{m,n} = \left[\frac{2m+1}{2M}, \frac{2n+1}{2N} \right].$$

We can approximate the first order approximation to $\tau(\mathbf{x})$:

$$\tau(\mathbf{x}) = \tau(\mathbf{x}_{m,n}) + \mathbf{c}_{m,n}^T (\mathbf{x} - \mathbf{x}_{m,n}),$$

where $\mathbf{c}_{m,n} = \nabla \tau(\mathbf{x}_{m,n})$ is the gradient. Then, the error is

$$\begin{aligned} e_{m,n}^2 &= \int_{\frac{m}{M}}^{\frac{m+1}{M}} \int_{\frac{n}{N}}^{\frac{n+1}{N}} (\tau_{m,n} - \tau(\mathbf{x}))^2 d\mathbf{x} \\ &= \int_{\frac{m}{M}}^{\frac{m+1}{M}} \int_{\frac{n}{N}}^{\frac{n+1}{N}} \left(\nabla \tau(\mathbf{x}_{m,n})^T (\mathbf{x} - \mathbf{x}_{m,n}) \right)^2 d\mathbf{x} \\ &= [c_{m,n}^x]^2 \frac{1}{12M^3N} + [c_{m,n}^y]^2 \frac{1}{12N^3M}. \end{aligned}$$

Assuming $M = N$, the above is simplified to

$$e_{m,n}^2 = \frac{[c_{m,n}^x]^2 + [c_{m,n}^y]^2}{12N^4} = \frac{\|\mathbf{c}_{m,n}\|^2}{12N^4}.$$

Summing over all (m, n) 's will give us

$$\begin{aligned} \text{Bias} &= \int_0^1 \int_0^1 (\bar{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2 d\mathbf{x} \\ &= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \frac{\|\mathbf{c}_{m,n}\|^2}{12N^4} = \frac{\|\mathbf{c}\|^2}{12N^2}. \end{aligned}$$

Visualization of the Bias. A visualization of the bias can be seen in Fig. 22. In this example, we use a real depth map as the ground truth $\tau(\mathbf{x})$. We scale the depth map so that the maximum value is 20, and the minimum value is 10. The full resolution is 512×512 . To reduce the discontinuity of the depth map at object boundaries, we apply a spatial low-pass filter to smooth out the edges. The estimated gradient magnitude is $\|\mathbf{c}\|^2 = \int_0^1 \nabla_{\mathbf{x}} \tau(\mathbf{x}) d\mathbf{x} = 1.42 \times 10^3$. We consider four different resolutions 8×8 , 16×16 , 32×32 , and 128×128 . For each resolution, we use the equation outlined in Lemma 9 to calculate the theoretically predicted bias. We compare this predicted bias with the simulation. As we can see from Fig. 22, the theoretical prediction offers an excellent match with the simulation.

15.2. Variance

For the variance term, we note that the derivation follows similarly to the 1D case:

$$\begin{aligned} \text{Var} &= \mathbb{E} \left[\int_0^1 \int_0^1 (\hat{\tau}(\mathbf{x}) - \bar{\tau}(\mathbf{x}))^2 d\mathbf{x} \right] \\ &= \mathbb{E} \left[\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \int_{\frac{m}{M}}^{\frac{m+1}{M}} \int_{\frac{n}{N}}^{\frac{n+1}{N}} (\hat{\tau}(\mathbf{x}) - \bar{\tau}(\mathbf{x}))^2 d\mathbf{x} \right] \\ &= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \int_{\frac{m}{M}}^{\frac{m+1}{M}} \int_{\frac{n}{N}}^{\frac{n+1}{N}} \mathbb{E} \left[(\hat{\tau}_{m,n} - \bar{\tau}_{m,n})^2 \right] d\mathbf{x}. \end{aligned}$$

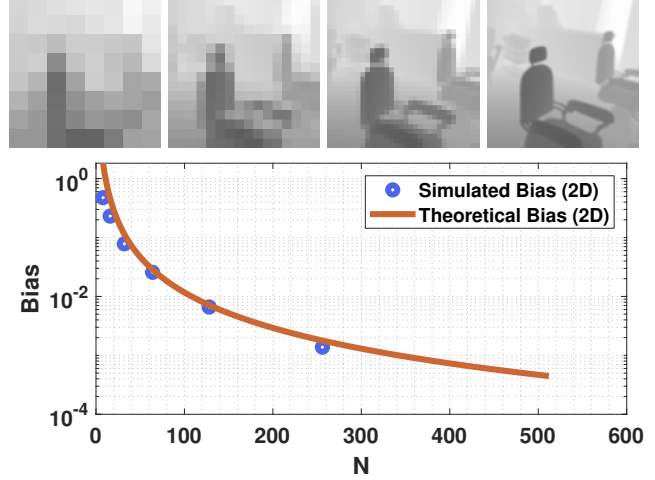


Figure 22. Visualization of the bias in 2D. The full resolution of the depth map is 512×512 . We consider multiple resolutions to study the effect of the bias. This plot highlights the excellent match between the theoretical prediction and the simulation results.

Therefore, as long as we can calculate $\mathbb{E} \left[(\hat{\tau}_{m,n} - \bar{\tau}_{m,n})^2 \right]$, we will be able to determine the variance.

Based on the arguments we described before stating the lemma, we shall focus on deriving the variance for each individual pixel. To this end, we need to derive the effective return pulse $\lambda_{m,n}(\mathbf{x}, t)$. Following the steps of the proof of Theorem 3, the critical step is the convolution of a 2D spatial Gaussian and a 1D temporal Gaussian

$$\text{2D spatial Gaussian} = \mathcal{N}(\mathbf{x} | 0, \sigma_s^2 \mathbf{I})$$

$$\text{1D temporal Gaussian} = \mathcal{N}(t | \tau_0 + \mathbf{c}_0^T (\mathbf{x} - \mathbf{x}_0), \sigma_t^2).$$

Here, $\sigma_s = 1/(\sqrt{12}N)$ is the spatial radius of the 2D Gaussian, which serves the same role as σ_x in the 1D case. The linear approximation $\tau_0 + \mathbf{c}_0^T (\mathbf{x} - \mathbf{x}_0)$ is the 2D version of the 1D linear approximation $\tau_0 + \tau'(x_0)(x - x_0)$, where \mathbf{x}_0 is the center pixel in each interval of the grid, and \mathbf{c}_0 is the gradient at \mathbf{x}_0 .

Before we proceed to the proof, we state the main result.

Lemma 10. *Let $\tau(\mathbf{x})$ be a 2D time of arrival function over the unit square $\mathbf{x} \in [0, 1] \times [0, 1]$. Suppose that we use $M \times N$ pixels to approximate the unit square, and for simplicity assume that $M = N$. Let α_0 be the total flux over the unit square, then The variance is given by*

$$\text{Var} = \frac{N^2}{\alpha_0} (\|\mathbf{c}\|^2 \sigma_s^2 + \sigma_t^2), \quad (53)$$

where $\sigma_s = 1/(\sqrt{12}N)$ is the spatial radius of the Gaussian approximation, and σ_t is the temporal spread of the pulse.

Proof of Lemma 10. The core derivation is the convolution between the 2D Gaussian in space and the 1D Gaussian in time. By separability of a 2D Gaussian, we can write the 2D Gaussian as the convolution of two orthogonal 1D Gaussian functions:

$$\mathcal{N}(\mathbf{x} | 0, \sigma_s^2 \mathbf{I}) = \mathcal{N}(x | 0, \sigma_s^2) \otimes \mathcal{N}(y | 0, \sigma_s^2).$$

The order of x versus y does not matter. Then, substituting it into the main convolution, we can perform two consecutive convolutions:

$$\begin{aligned} \mathcal{N}(t | \tau_0 + \mathbf{c}_0^T(\mathbf{x} - \mathbf{x}_0), \sigma_t^2) \otimes \mathcal{N}(\mathbf{x} | 0, \sigma_s^2 \mathbf{I}) \\ = \mathcal{N}(t | \tau_0 + \mathbf{c}_0^T(\mathbf{x} - \mathbf{x}_0), \sigma_t^2) \otimes \mathcal{N}(x | 0, \sigma_s^2) \\ \otimes \mathcal{N}(y | 0, \sigma_s^2) \end{aligned}$$

The first convolution, following the derivation of Theorem 3, will give us

$$\begin{aligned} \mathcal{N}(t | \tau_0 + \mathbf{c}_0^T(\mathbf{x} - \mathbf{x}_0), \sigma_t^2) \otimes \mathcal{N}(x | 0, \sigma_s^2) \\ = \mathcal{N}(t | \tau_0 + \mathbf{c}_0^T(\mathbf{x} - \mathbf{x}_0), \sigma_t^2 + [c_0^x]^2 \sigma_s^2), \end{aligned}$$

where c_0^x is the horizontal component of the gradient $\mathbf{c}_0 = [c_0^x, c_0^y]^T$.

The second convolution, which is applied after the first convolution, will give us

$$\begin{aligned} \mathcal{N}(t | \tau_0 + \mathbf{c}_0^T(\mathbf{x} - \mathbf{x}_0), \sigma_t^2 + [c_0^x]^2 \sigma_s^2) \otimes \mathcal{N}(y | 0, \sigma_s^2) \\ = \mathcal{N}(t | \tau_0 + \mathbf{c}_0^T(\mathbf{x} - \mathbf{x}_0), \sigma_t^2 + [c_0^x]^2 \sigma_s^2 + [c_0^y]^2 \sigma_s^2) \\ = \mathcal{N}(t | \tau_0 + \mathbf{c}_0^T(\mathbf{x} - \mathbf{x}_0), \sigma_t^2 + \|\mathbf{c}_0\|^2 \sigma_s^2). \end{aligned}$$

Restricting ourselves to $\mathbf{x} = \mathbf{x}_0$, we can show that the effective return pulse is

$$\lambda_{m,n}(t) = \alpha \mathcal{N}(t | \tau_{m,n}, \|\mathbf{c}_{m,n}\|^2 \sigma_s^2 + \sigma_t^2) + \lambda_b.$$

Finally, we recognize that since there are MN pixels in the unit space, (and assuming that $M = N$ for simplicity), then each pixel will see a total flux of α_0/N^2 . Therefore,

$$\begin{aligned} \mathbb{E} \left[(\hat{\tau}_{m,n} - \bar{\tau}_{m,n})^2 \right] &= \frac{\text{variance of } \lambda_{m,n}(t)}{\alpha} \\ &= \frac{N^2}{\alpha_0} (\|\mathbf{c}_{m,n}\|^2 \sigma_s^2 + \sigma_t^2). \end{aligned}$$

Summing over m and n will give us

$$\begin{aligned} \text{Var} &= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \int_{\frac{m}{M}}^{\frac{m+1}{M}} \int_{\frac{n}{N}}^{\frac{n+1}{N}} \mathbb{E} \left[(\hat{\tau}_{m,n} - \bar{\tau}_{m,n})^2 \right] d\mathbf{x} \\ &= \frac{N^2}{\alpha_0} (\|\mathbf{c}\|^2 \sigma_s^2 + \sigma_t^2). \end{aligned}$$

Visualizing the Variance. Fig. 23 shows the visualization of the variance as a function of N . The experimental configuration is identical to the one we used in the 2D bias analysis. For the purpose of analyzing the variance, we need to

further set up a Monte Carlo simulation. To this end, we first set the spatial spacing of the grid as $\Delta x = 1/512$ along each direction. The total amount of flux is set as $\alpha_0 = 1e^6$. Suppose that we have N pixels on each side, the unit square will have N^2 pixels. Each pixel will therefore observe α_0/N^2 amount of flux. When $N = 8$, the per pixel flux is 15625 photons per pixel. When $N = 256$, the per pixel flux is 15.26 photons per pixel.

For simplicity, we assume that the pulse $s(t)$ is Gaussian. The width of the pulse is 2 arbitrary units, which is roughly 10 percent of the true depth value. We also assume that the noise floor is $\lambda_b = 0$. These two assumptions allow us to use a simple ensemble average as the maximum likelihood estimator. Otherwise, we need to numerically implement the matched filter.

In this particular experiment, we use $\sigma_s = 1/(\sqrt{12}N)$ as the spatial radius, and $\|\mathbf{c}\|^2 = 1.42 \times 10^3$. The variance is calculated for each pixel and then summed over the entire unit square. A total number of 1000 random trials are conducted to obtain the variance per pixel (and for the whole image).

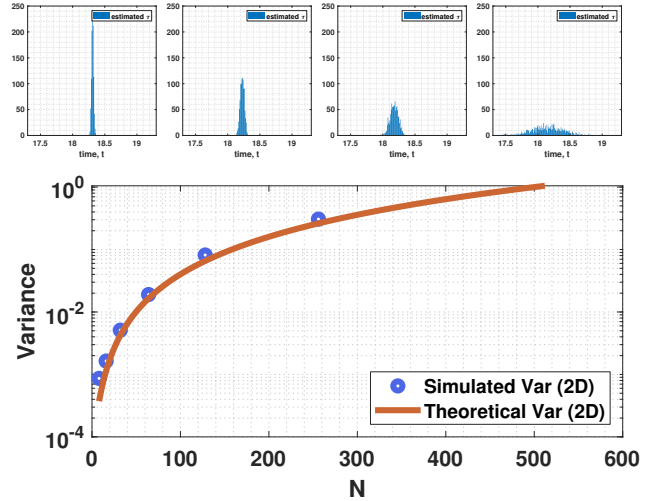


Figure 23. Visualization of the variance for a 2D example. For every pixel in the scene, we run a Monte Carlo simulation to compute the variance of the estimated depth values. The histograms shown in this figure correspond to the four resolutions we showed in the bias case. As we increase the spatial resolution (so that each pixel becomes smaller), the histograms become flat and so the variance increases. The plot at the bottom highlights the excellent match between the theoretically predicted variance and the simulated variance.

16. Real 2D Experiment

In this section, we provide additional detail on how real 2D experiments are conducted.

16.1. Snapshot of Real Data

As depicted in the main text, the real data is collected by a 192×128 SPAD array previously reported by Henderson et al. [18]. The system has a time-to-digital converter (TDC) timing resolution of 35 ps. For convenience, we crop the center 128×128 region for analysis. Fig. 24 shows a one-frame snapshot of the real data, and the 10,000-frame average. We observe that the 10,000-frame average is noisy, although the shape of the object is visible.

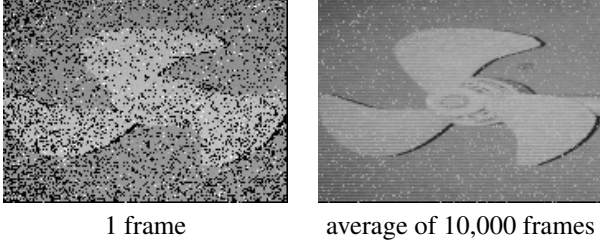


Figure 24. Raw data produced by the SPAD array.

Inspecting the source of the noise, we plot the histogram of one of the pixels of the data volume in Fig. 25. We recognize two issues in the data: (i) There is a secondary pulse, likely caused by secondary bounces from the background; (ii) there is a strong peak happening at the beginning (and sometimes at the end) of the sensing period, likely caused by failed detections of the SPAD. We did not show these samples in the histogram because it is just a strong spike in the histogram. The presence of these two issues makes the average of the time stamps problematic, hence generating noise.

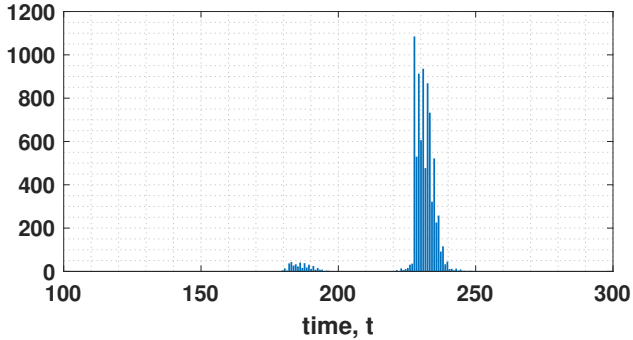


Figure 25. Histogram of one of the pixels in the SPAD data array.

16.2. Pseudo Ground Truth

For the purpose of MSE analysis, we need to construct a pseudo ground truth — a ground truth signal generated from all the available samples, hoping as noise free as possible. To achieve this goal, we need to reject the outliers so that we can retain only the primary pulse.

Our pre-processing step is done by identifying the center of the primary pulse. Once the center is identified, we crop around the center histogram with $\pm 3\sigma_t$ where σ_t is the standard deviation of the pulse.

Two remarks are worth mentioning: (i) The identification of the center of the primary pulse involves a few steps. The first step is to find a coarse estimate of the center. For this purpose, we compute the mean of the entire 10,000 data points and crop a large temporal window around the mean. We smooth out the histogram so that we can pick a more reliable peak. The smoothing step is done by adding Gaussian random variables. (ii) Once the peak is identified, we move to the second stage by putting a small temporal window around the peak. We reject all samples that fall outside this small temporal window, thus effectively removing the secondary pulse and the spike in the beginning and at the end.

After we have cleaned up the data, we can plot the resulting histogram as shown in Fig. 26. The pulse is significantly more concentrated around the true peak. The resulting depth map is slightly over-smoothed. However, since all our subsequent analysis is done based on the processed histogram (as we will be re-sampling with replacement from the processed histogram), computing the MSE with respect to this smoothed depth map will not cause issues to the analysis.

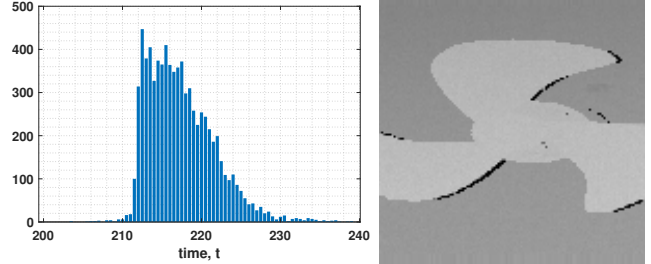


Figure 26. [Left] The processed histogram of one of the pixels and [Right] the resulting depth map.

16.3. Estimating σ_t and α_0

In the theoretical analysis of the variance, we need to know the pulse width σ_t and the total amount of flux α_0 .

For σ_t , we use the histogram shown in Fig. 26 to compute the variance at every pixel location. Although the pulse is asymmetric, we treat it as a symmetric pulse and compute the standard deviation regardless. We repeat the process for every pixel, and we generate a map of σ_t as shown in Fig. 27. For simplicity, we use the average of the σ_t in this map as the true value for our theoretical analysis. Our estimated σ_t is $\sigma_t = 3.93$.

For α_0 , we again use the histogram shown in Fig. 26 to count the number of elements in the histogram. Recall that

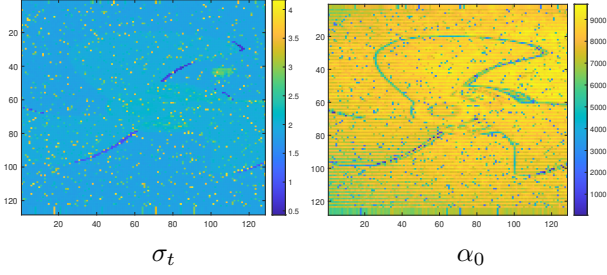


Figure 27. The values of σ_t and α_0 over the unit square.

during the pre-processing stage, we have already eliminated all the outliers from the raw data. This will leave us with a smaller set of samples compared to the original data array. The number of active samples will inform us about the number of photons arriving at the sensor. During our analysis, we will sample with replacement from the histogram. For each trial, we draw K samples from the available data points per pixel. Therefore, the total flux we have in the data is the sum of the values in Fig. 27, divided by 10,000 because we have 10,000 frames, and multiplied by K . Approximately our α_0 is 3.86×10^4 .

16.4. Variance Estimation via Bootstrap

The main tool we use to evaluate the variance is bootstrap. The idea is that given a dataset of M data points, we sample with replacement K samples and calculate the estimate of interest. The variance of the estimate is called the bootstrapped variance. When we repeat the random trials for long enough, the bootstrapped variance will converge to the true variance in probability.

Following this idea, for every pixel, we sample with replacement $K = 3$ times to generate the samples. From these samples, we construct the ML estimate (for every pixel) where the ML estimation is done by taking the simple average of the samples. Here, we assume that the pulse is symmetric even though it is not. We find that this approximation does not cause too many issues. Once the ML estimate is obtained, we repeat the process 20 times to evaluate the variance.

For lower resolutions, we bin the samples to form a bigger pool. If we use a 2×2 bin, then the total number of samples to be generated for the bootstrap purpose is $K \times 2^2 = 12$ samples; If we use a 3×3 bin, then the total number of samples is $K \times 3^2 = 27$ samples. These samples will give us an estimate of the real variance of the data.

The theoretical formula of the variance is based on a simpler form $\sigma_t^2 N^2 / \alpha_0$. We omitted the $\|\mathbf{c}\|^2 \sigma_s^2$ because $\|\mathbf{c}\|^2 \sigma_s^2$ is small compared to σ_t^2 . (Recall that $\sigma_s = 1/(\sqrt{12N})$).

16.5. Bias Estimation via Numerical Integration

The bias is computed numerically via integration. The reason is that the pseudo ground truth depth map is overall a piecewise constant function. As we explained in the limitation subsection of the Bias section, there is no simple analytic formula for calculating the bias for piecewise constant functions unless we perform the numerical integration. The numerical integration is done by downsampling and upsampling the cleaned raw data volume. For example, to calculate the bias caused by a 2×2 bin, we take the average over all the available samples in the $2 \times 2 \times T$ (where T is the number of available samples). After binning, we scale it back to the full resolution and compute the sum squared error with respect to the pseudo ground truth.

17. Q&A

In this section, we list a few questions and answers which might be of interest to readers.

Q1. Can you just use as many pixels as possible during acquisition, and then perform binning (of the time stamps) as post-processing?

Answer: Yes, this is completely possible. In fact, in our real data analysis, we see that when the pulse is short enough, the MSE decays monotonically with the N . Theoretically, the optimal N still exists but this N could be larger than what the physical resolution of the sensor can support. In this case, we should just maximize the resolution.

From a practical point of view, we also agree that post-processing of noisy time stamps can be cost-effective. Analog processing on the sensor front could also be a solution.

Q2. What if you denoise the estimated time of arrival map? Will it beat your MSE bound?

Answer: Yes, it will. The theory we presented here uses the maximum likelihood (ML) estimation. ML estimation allows us to say things concretely so that we can offer a simple and interpretable MSE estimate. If we do denoising, then we will be doing maximum a posteriori or minimum mean square estimation. In those cases, we need to specify the prior distribution for which no one has a formula. Even if we do pick a prior (e.g., total variation), the derivation of the MSE will be substantially harder if not intractable. So, we lose the capability of writing a simple and interpretable formula. In short, while we are almost certain that a well-defined post-processing will “beat” our MSE bound, this “victory” offers little to no theoretical benefit to the scope of this paper.

Q3. What is the utility of this paper?

Answer: The MSE we show in this paper is, in our honest opinion, simple, interpretable, and elegant. For the first time in the LiDAR literature, we provide the exact characterization of the resolution limit.

Q4. You need to show more comparisons by sending the raw time stamps through a state-of-the-art depth reconstruction neural network.

Answer: Beating a SOTA depth reconstruction neural network is not the purpose of this paper.

Q5. Is MSE the right metric?

Answer: Yes, if you want to derive equations, especially a simple equation to give you intuitions about the problem. No, if you are more interested in practical scenarios. There is no better or worse.

Q6. Your model is inaccurate. It made too many assumptions such as Gaussian pulse, single-bounce, no dark current, constant reflectivity, etc.

Answer: While we also want to be as accurate as possible, accuracy and simplicity are mutually exclusive in this paper. As we have explained in the supplementary material, all these situations can be handled *numerically* by integrations. But this will defeat the purpose of deriving a closed-form expression.

Q7. Some papers in the literature use the Markov chain/self-excitation process to model the photon arrivals. Why are you skipping all these?

Answer: We are just assuming that there is no dead time. If there is dead time, then you are correct that self-excitation processes are needed to provide a better model. However, this would be substantially more complicated than what we present here. By assuming no dead time, we can go back to the standard inhomogeneous Poisson process by exploiting independence. Even in this significantly simplified case, we see that the derivation of the MSE is non-trivial.

Q8. Can your model handle fog?

Answer: Yes, but it will be complicated. Scattering medium such as fog affects the reflectivity α_0 and causes additional background as shown in Figure 16. You will need numerical integration to evaluate the bound.