2022 IEEE International Conference on Computational Photography (ICCP) | 978-1-6654-5851-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCP54855.2022.9887683

First Arrival Differential LiDAR

Tianyi Zhang, Mel J. White*, Akshat Dave*, Shahaboddin Ghajari, Ankit Raghuram, Alyosha C. Molnar,
Ashok Veeraraghavan

Abstract—Single-photon avalanche diode (SPAD) based LiDAR is becoming the de-facto choice for 3D imaging in many emerging applications. However, they suffer from three significant limitations: (a) the additional time-of-arrival dimension results in a data throughput bottleneck, (b) limited spatial resolution due to either low fill-factor (flash LiDAR) or scanning time (scanning-based LiDAR), and (c) coarse depth resolution due to quantization of photon timing by existing SPAD timing circuitries. In this paper, we present a novel, in-pixel computing architecture that we term *first arrival differential (FAD) LiDAR*, where instead of recording quantized time-of-arrival information at individual pixels, we record a temporal differential measurement between pairs of pixels. FAD captures relative order of photon arrivals at the two pixels (within a cycle or laser period) and creates a one-to-one mapping between this differential measurement and depth differences between the two pixels. We perform detailed system analysis and characterization using Monte Carlo simulation, and experimental emulation using a scanning-based single-photon avalanche diode. FAD pixels can result in a 10-100x reduction in per-pixel data throughput compared to TDC-based pixels. Under the same bandwidth constraints, FAD-LiDAR achieves better depth resolution and/or range than several state-of-the-art TDC-based LiDAR baselines.

Index Terms—In-pixel computing, single photon avalanche diode (SPAD), exotic sensors, computational photography, LiDAR, 3D imaging

◆

1 INTRODUCTION

3D IMAGING is crucial for a wide range of applications such as autonomous vehicles, augmented/virtual reality, robotics, entertainment and mobile imaging [1], [2], [3]. In particular, LiDAR has become a popular choice for many of these applications due to its speed and sensitivity. LiDAR is a class of active 3D imaging methods that relies on actively sending out laser waveforms and inferring the time of flight based on the phase shift or delay in the received waveform. While a variety of techniques exist for LiDAR (e.g. CW-TOF [4] and pulsed TOF based on APDs [5]), SPAD-based LiDAR has become the default for applications requiring single-photon sensitivity, high timing resolution, and CMOS compatibility [6], [7]. Promising results on low-light, long distance ranging have been shown in recent works [8], [9], [10], [11].

Types of LiDAR. LiDAR is typically categorized as either scanning LiDAR or flash (array-based) LiDAR. In scanning LiDAR, a single line is illuminated at a time and a line sensor records the 3D data. Sequential scanning of the line by rotating the illumination beam allows for the complete capture of a 2D range image. In these applications scan time becomes the primary limitation, limiting acquisition to about 3 million samples per second [12] [13]. Even assuming a nominal 30 Hz operation, this yields images with a resolution of less than 0.1 megapixels. The mechanical scanning parts also makes the system less durable.

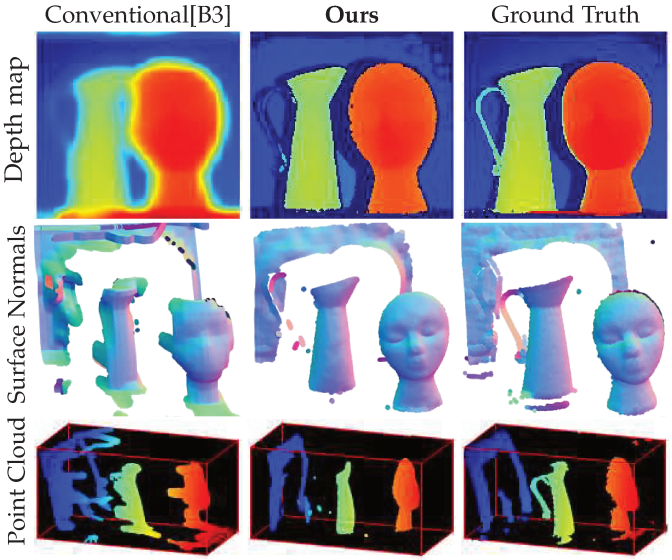


Fig. 1. **FAD LiDAR design emulated from experimental capture of real-world scenes**. Conventional (per-pixel TDC based) flash LiDAR sensors suffer from low spatial resolution due to poor pixel fill factor (the leftmost column corresponds to a typical liDAR baseline, with details in Table 1). Our approach exploits lightweight differential operations to perform high resolution 3D imaging.

Flash LiDAR utilizes a two-dimensional array detector coupled with flash illumination, requiring no mechanical scanning [7]. It is considered one of the most promising alternatives to achieve higher resolution operation. However, the current-generation flash LiDAR has low spatial resolution due to the large chip area occupied by the in-pixel circuitry required to generate and store photon timestamps, thereby reducing the overall fill factor. Furthermore, the addition of a temporal dimension increases the required data throughput rate per pixel by two orders of magnitude, making high spatial resolution intractable [14].

Key idea. We propose a novel flash LiDAR approach that

- *T. Zhang, A. Dave, A. Raghuram, A. Veeraraghavan are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77005.*
  E-mail: tz12@rice.edu
- *M. J. White, S. Ghajari, A. C. Molnar are with the Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY, 14850.*
- ** equal contributions*

978-1-6654-5851-1/22/$31.00 ©2022 IEEE

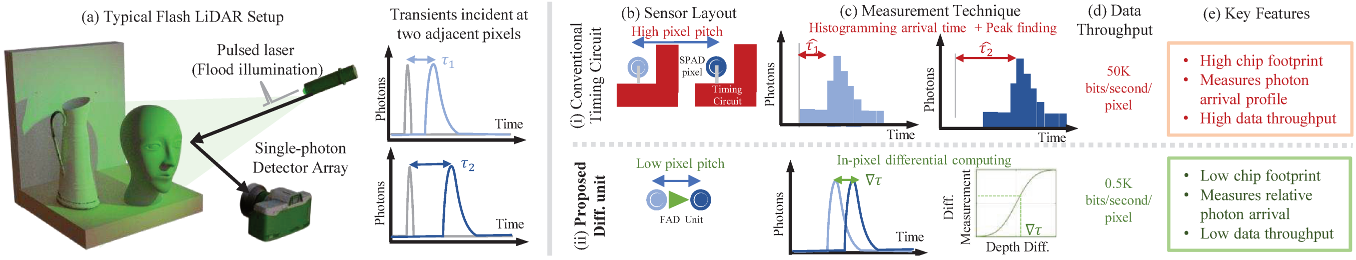


Fig. 2. **Key Idea**: Flash LiDAR setup typically involves flood illuminating the scene with a pulsed source and acquiring photon arrival information with a single photon detector array. (a) Conventional flash LiDAR sensors comprise of a single-photon detector (SPAD detector) pixel and timing circuit to measure the photon arrival profile. (b) The timing circuit has a high spatial and memory footprint resulting in high pixel pitch and low latency (i). In this work, we leverage the novel in-pixel construct of FAD units (ii) that can measure the relative depth difference (c) between two pixels while requiring low spatial and memory footprint. (d,e) Our FAD LiDAR performs high resolution 3D imaging at two orders of magnitude lower throughput than conventional flash LiDAR.

exploits the sparsity in depth gradients commonly observed in natural scenes by efficiently computing the *relative* arrival time differences, instead of the data-intensive *absolute* arrival time stamps or timing histograms. We introduce a lightweight in-pixel computing element called the first arrival differential (FAD) unit. A FAD unit compares and accumulates the stochastic photon arrival orders between adjacent pixel pairs over a number of cycles or laser periods. The output of a FAD is a 2D-intensity like measurement, each forming a monotonic response to depth gradient, as shown in Fig. 2.

These FAD counts allow us to extract depth gradients, detect depth edges, and estimate surface normals at high resolutions (Fig. 14). We can also generate high resolution depth maps by coupling the high resolution relative depth differences from these FAD units with the sparse absolute depth references from two orders of magnitude fewer timing circuits (Fig. 15). The highlights and benefits of our sensor, which we term as FAD LiDAR, include:

- **Data throughput and scalability.** Each FAD unit requires only a single count to be stored between a pair of pixels. This results in a 100x reduction in data throughput rate requirement per pixel (compared to computing absolute time stamps and histograms). Apart from lower power and readout circuit complexity, this data throughput reduction is also critical for scalability, making higher spatio-temporal sampling rates feasible.
- **Depth resolution.** FAD units have a unique, well-behaved, one-to-one mapping between the measured counter value and the depth difference between the two pixels. As the depth difference is encoded in an intensity-like unit (count of photons), this allows us to achieve depth resolution comparable (or sometimes better) than conventional LiDAR that need to quantize time-of-arrival into a discrete number of time bins.
- **Simple in-pixel computing architecture.** A FAD consists of a single Set/Reset Latch (SR latch) to identify the pixel with first photon arrival and a single up-down counter. Our preliminary design layouts suggest a 10x reduction in area compared to traditional histogram generation circuitry.

Limitations. A physical prototype implementing the technique we present here is yet to be fabricated. Instead, to emulate the operation of FAD LiDAR on real scenes, we acquire photon time stamps with a single-pixel SPAD-based scanning hardware and then compute the emulated FAD response post-capture. The current FAD design performs best under moderate to high illumination levels, and our current model is designed to handle single reflections only. We will analyze these limitations–as well as potential mitigation–in detail in latter sections.

2 RELATED WORK

Conventional flash LiDAR consists of SPAD pixels and timing circuitry. A measurement involves generating quantized time stamps representing photon arrival times using time-to-digital converters (TDCs). A histogram is computed from these time stamps either on-chip or off-chip to generate a photon arrival profile (also known as a transient). Depth can then be computed by locating peaks in the transient as illustrated in Fig. 2. An *ideal* SPAD-based LiDAR array would contain many closely packed SPADs and provide precise time stamps at every pixel. But such an array is not physically realizable due to the sheer amount of data produced and chip area required by the support circuitry as shown in Fig. 3. Here we explain the throughput and spatial footprint challenges associated with conventional SPAD-based LiDAR techniques. We also describe some recent efforts to alleviate these challenges by trading off other design metrics.

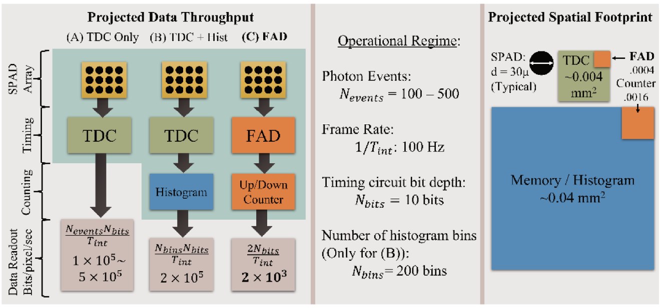


Fig. 3. **Comparison of flash LiDAR with proposed FAD device**. (A) Timestamps are reported by a TDC and offloaded from the chip. (B) Timestamps are histogrammed on-chip before being read. (C) Proposed FAD architecture. Under typical LiDAR operating conditions (Middle), FAD requires two orders of magnitude lower data throughput, while simultaneously achieving a significantly lower chip footprint compared to conventional LIDAR architectures (Right). Area estimate for 8-bit TDC (green) is based on transistor count from [15] and a 180 nm process. Memory size is estimated from reported area in [16].

On-chip TDC and off-chip histogramming. If the TDC timestamps are quantized with a bin width of B_{TDC} and are defined for a range R_{TDC}, then the TDC requires $\log_2\left(\frac{R_{\text{TDC}}}{B_{\text{TDC}}}\right)$ bits. If N_{events} photon arrival events are accumulated and the chip is operated for integration time of T_{int}, the data throughput in bits per pixel per second is given as:

$$\text{TP}_{\text{TDC}} = \frac{N_{\text{events}}}{T_{\text{int}}}\log_2\left(\frac{R_{\text{TDC}}}{B_{\text{TDC}}}\right) \tag{1}$$

The data throughput of TDCs under practical LiDAR regimes (Fig. 3) can easily go up to 1 Mbps per pixel – thousands of Gbps for a mega-pixel array [14]. Moreover, TDCs typically are large and must be situated close to the pixel for accuracy. This results in large pixel pitch, as well as poor fill factor – usually less than 20%, unless techniques such as microlenses or 3D stacking are incorporated [17]. Fig. 3 shows typical sizes for support circuitry.

On-chip TDC with Histogramming. Reading out the TDC time-stamps at every cycle and then histogramming is highly data-intensive. One way to reduce throughput is to directly accumulate histograms of photon arrivals on the chip. If the number of bins in the histogram is N_{bins}, the throughput of on-chip TDC with histogramming can be defined as

$$\text{TP}_{\text{Hist}} = \frac{N_{\text{bins}}}{T_{int}}\log_2\left(N_{\text{events}}\right) \tag{2}$$

When few bins suffice, on-chip histogramming provides a throughput advantage over on-chip TDC-only techniques. Adaptive peak finding from coarse to fine bin resolutions can further limit the on-chip memory and data readout [14]. However, as shown in Fig. 3, the spatial footprint of on-chip histogramming circuit is 10× higher than the TDC. See supplementary material for details.

Performance Trade-Offs. The throughput calculation (Eq. 1, 2) shows that for a fixed bandwidth, improvement in any of the performance factors: range, bin width, or spatial resolution, results in degradation of the others. This performance trade-off is visualized in Fig. 4 for conventional flash LiDAR chip implementations, Photonforce PF32 [18], Perenzoni 2017 [19], and Sony 2021 [20]. Efforts for further improving throughput or spatial footprint typically require more sophisticated on-chip processing, as described next.

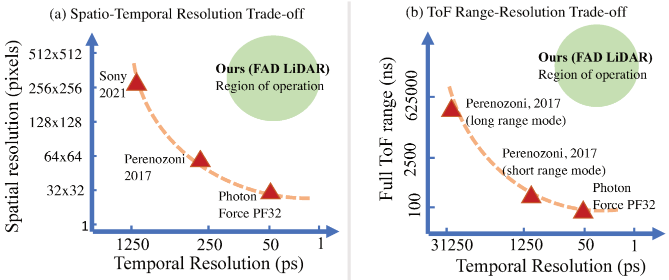


Fig. 4. **Trade-offs in conventional LiDAR implementations.** Given high bandwidth and chip footprint of timing circuits, for a fixed throughput, improving on spatial resolution, temporal resolution or the TOF range results in degradation of the other factors. This trend is also evident in existing LiDAR implementations.

To mitigate the problem, one solution is TDC sharing – connecting multiple SPAD diodes to a shared TDC. Another common architecture is to use coarser time bins and large pulse width. However, even with these improvements, the best SPAD LiDAR array resolution is still limited to ∼ 100k pixels, since densely placed TDCs are still needed. Next we introduce on-chip processing efforts to solve this challenge.

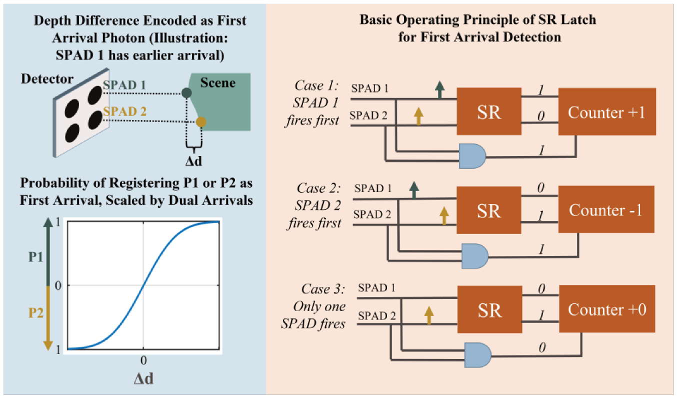


Fig. 5. **Imaging principle of FAD**. **Left:** Mapping of depth to the probability of an up/down count. **Right:** Various cases of photon arrival ordering.

On-chip processing for smarter SPAD arrays. As CMOS processes have matured, there has been an emergence of “exotic” 2D CMOS sensors, which record transforms or features computed within the sensing hardware rather than standard intensity measurements [21], [22]. These sensors usually have significant advantages in speed, imaging quality, and data throughput. Such trends have also extended to SPAD designs. For example, motion-triggered SPADs [23] are an event-based implementation for SPADs, such that TOF frames are recorded only if inter-frame intensity changes are present. Walker *et al.* directly output depth values, albeit requiring more on-chip logic [24]. To improve spatial resolution, Martin *et al.* present a hybrid array in which fine intensity values are used to upsample depth by 16x in 2D via a neural network [25].

Another approach is to find compact statistics or features that can be extracted in-pixel. For example, sketching techniques can be used to compress photon events on-chip [26]. Bhandari et al. [27] proposes a one-bit time-resolved imaging scheme. In addition to TOF, in-pixel computing schemes are also proposed for passive HDR imaging [28].

We present a novel on-chip processing construct for LiDAR based on the observation that the spatial gradients of depth and thus arrival times are sparse, analogous to how 2D differential sensors exploit sparsity in intensity gradients [29]. Our approach results in a significant performance boost for a given bandwidth as visualized in Fig. 4. To our knowledge, this technique is the first of its kind.

3 DESIGN PRINCIPLE AND FORWARD MODELING FOR FAD

3.1 FAD unit encodes relative depth difference

We introduce a novel FAD unit, which computes a *stochastic* measurement that encodes relative depth gradient between adjacent SPADs. Fig. 5 (top left) illustrates how a FAD unit works with two scene points separated by Δd.

Within a time window, photons reaching pixel 1 (corresponding to a nearer point) are more likely to arrive earlier than photons from pixel 2. Over a large number of cycles, the relative frequency of first arrival photons between the pixels can capture information about depth difference Δd.

With a FAD unit, there are three types of events that can occur in a cycle: (1) *dual arrivals*: both SPADs fire, (2) *single arrivals*: only one SPAD fires, and (3) *null*: neither SPAD fires. In the case of dual arrivals, comparing the arrival order of the two SPADs gives us information about their relative depth. We can achieve this via a simple Set/Reset latch (SR latch), which holds the value of the first input to go high (e.g., record a photon event).

The FAD unit, illustrated in Fig. 5, distinguishes dual arrival from single arrival events via an AND gate connected between the SPADs and counter. This mechanism discounts any single-arrival events by not allowing the counter to increment. In a FAD, the SR latch replaces the TDC and an up-down counter takes the place of histogramming and on-chip memory. The entire circuitry for the FAD (excluding the counter) is projected to only occupy 0.00042 mm^2 [30]. Please refer to the supplementary material for a preliminary device layout design. In Fig. 3 we perform a back-of-the-envelope comparison of circuit foot- prints with standard methods.

3.2 Mathematical modeling for the FAD structure

Upon integrating for a sufficient number of laser cycles, the mean of the FAD counts is a smooth monotonic function of the photon arrival time difference (Fig. 5). We term the mean of FAD counts as the FAD response. Here we show how the FAD response enables us to measure depth differences.

Notation. We start by modeling the temporal response of the laser and SPAD jitter combined as a Gaussian pulse $s[n]$ with standard deviation σ, where n denotes a discrete time bin [10]. Let α_1, α_2 and τ_1, τ_2 denote the scene intensity and the time bin of the peak mapped onto these two pixels, respectively. We also absorb common scaling factors such as the photon detection efficiency into α_1, α_2. Let the time-of-flight difference be defined as $\Delta\tau = \tau_1 - \tau_2$.

Assumptions on SPAD operation In our emulation, we assume all pixels are reset synchronously at each cycle, which can be practically implemented on-chip. Thus, all pixels become active at the beginning of each laser cycle, and the deadtime at one particular cycle doesn't affect later cycles. This assumption holds in our modeling, simulation and emulation experiments (operated under low photon rates to avoid pileup).

Photon arrival statistics The photon arrival process at each pixel can be modeled as an inhomogenous Poisson process, which is well studied [10]. At each time bin n, the number of photon arrivals approximately follows a Poisson distribution parameterized by rate function λ. Let the Poisson rate functions for the photon arrival at SPAD 1, 2 be $\lambda_1[n], \lambda_2[n]$. Here $\lambda_i[n] = \alpha_i s[n-\tau_i] + B$, where $s[n]$ represents the laser waveform, α_i represents the reflectivity (includes quantum efficiency term), and B the background term (dark count and ambient light). α, B values are all in photons/cycle. Let the active duration per cycle be T bins. The probability of detecting a photon at time bin n depends on the rate function for the time bin n as well as the prior time bins, and can be written as

$$P_{\text{i,detection}}[n] = \exp\left(-\sum_{1}^{n-1}\lambda_i[l]\right)\left(1-\exp\left(-\lambda_i[n]\right)\right) \quad (3)$$

Although this derivation is for Gaussian signals, the results still apply to non-Gaussian waveforms, e.g., including exponential tails (see Section 5).

Multinomial modeling of up/down events In each cycle, one of three possible events occur: up, down or null (no detection). Thus the comparison and count update process can be represented as a multinomial distribution sampled over N_{cycles}. With the set of counts as $\mathbf{C} = [C_{\text{up}}, C_{\text{down}}, C_{null}]$, where the entries represent up, down counts and empty cycles over N_{cycles} respectively, we have

$$\mathbf{C} \sim \textbf{mult}(P_{\text{up}}, P_{\text{down}}, 1 - P_{\text{up}} - P_{\text{down}}, N_{\text{cycles}}) \quad (4)$$

$P_{\text{up}}, P_{\text{down}}$ can be modeled based on photon arrival statistics at each pixel. Specifically, let $P_{i,\text{detection}}[l]$ represent the probability of receiving a detection at pixel i, bin l at any cycle. We derive P_{up} and P_{down} as

$$P_{\text{up}} = \sum_{0<l<T}\sum_{0<k<T} P_{1,\text{detection}}[l]P_{2,\text{detection}}[k]\mathbb{1}(l \leq k) \quad (5)$$

$$P_{\text{down}} = \sum_{0<l<T}\sum_{0<k<T} P_{1,\text{detection}}[l]P_{2,\text{detection}}[k]\mathbb{1}(l \geq k) \quad (6)$$

Directly computing the summations in Eq. 5, 6 is intractable due to high data dimensions. To simplify the analytical model, we separate photon arrivals from the background and the laser pulse. Specifically, we categorize the arrival events as four types:

- **Type I.** Photons at both SPADs arrive from pulses. This event gives us differential TOF information.
- **Type II.** Photons at both SPADs arrive from background. Since background photons follow uniform distributions, on average, they will not contribute to up and down count probability.
- **Type III.** The photon at SPAD 1 is from a pulse, and the photon at SPAD 2 is from background.
- **Type IV.** The photon at SPAD 1 is from background, and the photon at SPAD 2 is from a pulse.

Let us consider the case with no pile-up, i.e., total flux in photons per cycle is < 0.05. The detection probability at each bin can be simplified to be $P_{i,\text{detection}}[n] \approx \lambda_i[l]$ [31].

Type I is most useful in providing differential information, as we directly compare the TOF order between pulse photons at two pixels. Note that:

$$P_{\text{up}} = Pr\{\text{dual arrival}\} \cdot Pr\{l - k \leq 0 | \text{dual arrival}\} \quad (7)$$

Let $d = l - k$, the photon arrival time difference. Given that both pixels have photon arrivals within the period, $Pr\{l - k \leq 0|\text{dualarrival}\}$ models TOF difference as

$$P_{d|\text{dual arrival,Type I}} \sim N(\Delta\tau, 2\sigma^2) \quad (8)$$

$$P_{\text{dual arrival,Type I}} \approx \alpha_1\alpha_2 \quad (9)$$

We can use the erf notation for integrating the *up* probability $P_{d|\text{dualarrival}}$.

$$P_{\text{up,TypeI}} = \frac{\alpha_1\alpha_2}{2}\left[1 - \operatorname{erf}\left(\frac{\Delta\tau}{2\sigma}\right)\right] \tag{10}$$

$$P_{\text{down,TypeI}} = \frac{\alpha_1\alpha_2}{2}\left[1 + \operatorname{erf}\left(\frac{\Delta\tau}{2\sigma}\right)\right] \tag{11}$$

The contribution to the expected FAD value this case is hence:

$$E[\text{FAD, Type I}] = N_{\text{cycles}} \cdot (P_{\text{up,TypeI}} - P_{\text{down,Type I}}) \tag{12}$$

$$= -N_{\text{cycles}}\alpha_1\alpha_2 \operatorname{erf}\left(\frac{\Delta\tau}{2\sigma}\right) \tag{13}$$

The expected value is therefore a monotonic function of depth difference when background is absent. This property holds even if the pulse shape is not perfectly Gaussian.

Type II events will not contribute to the up and down counts. Since the background is uniformly distributed over the active windows, the up and down counts should have the same occurence probability and $E[\text{FAD, TypeIII}] = 0$. Please refer to supplementary document for more details.

Type III, IV events could potentially contribute to up and down counts. We have:

$$P_{\text{up,Type III}} = \alpha_1\left(1 - \frac{\tau_1}{T}\right)B; \quad P_{\text{down,Type III}} = \alpha_1\frac{\tau_1}{T}B \tag{14}$$

$$P_{\text{up,Type IV}} = \alpha_2\frac{\tau_2}{T}B; \quad P_{\text{down,Type IV}} = \alpha_2\left(1 - \frac{\tau_2}{T}\right)B \tag{15}$$

Therefore,

$$E[\text{FAD, Type III, IV}]$$

$$= N_{\text{cycles}} \cdot (P_{\text{up,Type III,IV}} - P_{\text{down,Type III,IV}}) \tag{16}$$

$$= N_{\text{cycles}}B\left[\alpha_1\left(1 - \frac{2\tau_1}{T}\right) - \alpha_2\left(1 - \frac{2\tau_2}{T}\right)\right] \tag{17}$$

$$= N_{\text{cycles}}B\left[(\alpha_1 - \alpha_2)\left(1 - \frac{2\bar{\tau}}{T}\right) - (\alpha_1 + \alpha_2)\frac{\Delta\tau}{2T}\right] \tag{18}$$

where $\bar{\tau} = \frac{\tau_1+\tau_2}{2}$. One observation is that when the two points are on a continuous surface, the TOF difference is small compared to the full range, i.e., $\Delta\tau << T$. Eq. 18 becomes:

$$E[\text{FAD, Type III, IV}] = N_{\text{cycles}}B\left[(\alpha_1 - \alpha_2)\left(1 - \frac{2\bar{\tau}}{T}\right)\right] \tag{19}$$

For initial analysis, we assume B is the same across all pixels as in [10]. For detailed derivations on up and down probabilities, please refer to the supplementary document.

From the Eq. 19, we can observe that the bias caused by Type III and IV events are ignorable when (1) reflectivity or albedo variation is small OR (2) background is low OR (3) pulse TOF values are close to the center of the laser period. With both high reflectivity variation and background present, there could be a large bias in FAD, as shown in Section 5. We will discuss ways to compensate for the bias in Section 4.

3.3 2D array constructs using FAD units

The depth difference computation abilities of FAD units can be leveraged to perform 3D imaging by constructing 2D sensor arrays composed of FAD units. First, every pair of neighboring pixels, both horizontally and vertically, are connected by FAD units to acquire independent orthogonal depth gradient information. Second, we also need **additional intensity information** for capturing per-pixel reflectivity and background information. This can be achieved with either an additional counter per pixel or by integrating linear mode pixels (which tend to be 100x smaller in area) next to every SPAD pixel. Third, to provide robustness of operation and the ability to handle large depth discontinuities, we also add sparsely distributed pixels with individual TDCs. Typically, we only need one such pixel for every 10x10 array of FAD pixels. In an actual implementation, the dense FAD connections can be achieved by using the box clique structure [29] or tiling. We also discuss a more spatially efficient multi-scale chip design in the supplementary document.

3.4 Throughput analysis

Now that we have the forward model, we perform a back-of-the-envelope estimate for data throughput. FAD units only require one up-down counter each for accumulating horizontal and vertical differential measurements. The total number of arrival events is $N_{\text{events}} = \alpha_1\alpha_2 N_{\text{cycles}}$. Each up-down counter requires $\log_2(N_{\text{events}})$ bits. Thus the throughput for FAD array is

$$\text{TP}_{\text{FAD}} = \frac{2}{T_{\text{int}}}\log_2(N_{\text{events}}) \tag{20}$$

The intensity pixels are projected to have very small impact on our FAD bandwidth estimation. Even including the intensity bandwidth, the total bandwidth would still be within the same order of magnitude – accumulating a 2D intensity count over an integration time is still substantially smaller than time-stamping every photon with a per-pixel TDC. With background, neither would the throughput change much as we only have an additional bias term. In addition to compressing spatio-temporal information to 2D intensity-like information, the sparsity of natural albedo and depth gradients can also further reduce the bits needed by the differential counts. For realistic operating regimes such as (Fig. 3 (middle)), FAD units offer two orders of magnitude lower throughput than conventional flash LiDAR.

4 RAW DEPTH GRADIENT ESTIMATION AND EDGE DETECTION

In this section, we study how FAD counts could give us raw per-pixel estimates of edges, depth gradients and normals.

The chip arrangement in section 3.3 allow us to obtain background ($\hat{B}$) and reflectivity estimates ($\hat{\alpha_1}, \hat{\alpha_2}$). When background correction is needed, background estimates can be obtained by capturing intensity measurements with the lasers off. The pulse intensities can be estimated by turning the laser on and subtracting the background from each count - let them be $\hat{\alpha_1}, \hat{\alpha_2}$. For 2D measurements, a denoising step could be further applied to refine the intensity estimations. We start by considering edge detection and gradient estimation with low background, then discuss how to modify these steps when significant background is present.

Edge detection We can generate raw edge detection by thresholding the FAD values. When background is low, we only need to look at Type I events. From Eq. 13,

$$E[FAD] = -N_{\text{cycles}}\alpha_1\alpha_2 \operatorname{erf}\left(\frac{\Delta\tau}{2\sigma}\right) \tag{21}$$

To decouple FAD counts with intensity and exposure, we introduce a normalized FAD count, nFAD:

$$nFAD = \frac{FAD}{N_{\text{cycles}}\widehat{\alpha}_1\widehat{\alpha}_2} \tag{22}$$

We can get a set of potential edge points by thresholding nFAD. Fig. 5 (left) shows that the nFAD counts saturate to 1 or -1 whenever the depth differences are too high. In common scenes, high depth differences occur at depth edges. Thus, nFAD saturation can be used as an indicator for pixels at depth edges. We can simply threshold the absolute value nFAD counts (Fig. 10(b)). Depth edges close to hard intensity shadows can sometimes lead to spurious depth edges being detected. These can be alleviated when we have 2D nFAD measurements and apply common edge detection schemes, e.g. Canny edge detector.

Depth gradient estimation. For two adjacent points on a continuous surface, using Eq. 21, we perform moment-matching to get a coarse estimation of the depth gradients: $\hat{\Delta\tau} = -2\sigma \operatorname{inverf}(nFAD)$, where inverf is the inverse error response function.

Given depth gradients in both the x and y directions, we can get an initial estimate of per-pixel surface normal maps, as shown in Fig. 10(d) [32]. The normal vector $\mathbf{n}$ at each pixel can be computed as follows:

$$\mathbf{n}_u = [\widehat{\partial}_x z, \widehat{\partial}_y z, -1]^t, \mathbf{n} = \frac{\mathbf{n}_u}{||\mathbf{n}_u||} \tag{23}$$

$\hat{\partial_x} z, \hat{\partial_y} z$ corresponds to depth partials along x and y directions and can be estimated by normalizing $\hat{\Delta\tau}$ by dividing $(\text{pixel pitch} \times \text{lens magnification factor})$. $\hat{\Delta\tau}$ corresponds to the path difference and not exactly depth gradient, especially as the FOV becomes larger. In that case, a simple perspective projection step can be applied to $\hat{\Delta\tau}$ [32]. For the notation used in this paper, we assume $\hat{\Delta\tau}$ already contains this correction and corresponds to depth gradient values. So far, we only have normal estimates centered at one pixel. When we have normal values across a 2D grid, we can integrate and further refine the object structures. This will be discussed in latter sections.

Correcting for background bias. When background is high and the scene contains large albedo variation, we need to correct for the bias caused by background (as depicted in Fig. 9) before computing nFAD. Note that now the bias depends not only on $\Delta\tau$, but also on the absolute τ values. The absolute values in practice can be obtained with the help of sparse TDC measurements, which we will discuss more in detail in Section 6 and 7. Assume we have some estimate $\hat{\bar{\tau}}$, the edge detection criteria becomes

$$FAD > B(\hat{\alpha_1} - \hat{\alpha_2})\left(1 - \frac{2\bar{\tau}}{T}\right) + \omega_1 B(\hat{\alpha_1} + \hat{\alpha_2}) + \omega_2\hat{\alpha_1}\hat{\alpha_2} \tag{24}$$

Where ω_1, ω_2 are weights for the two terms involving $\Delta\tau$.

The corrected nFAD value on a smooth surface becomes

$$nFAD_{\text{corrected}} = \frac{FAD/N_{\text{cycles}} - bias}{\widehat{\alpha}_1\widehat{\alpha}_2} \tag{25}$$

where $bias = B(\hat{\alpha_1} - \hat{\alpha_2})(1 - \frac{2\hat{\tau}}{T})$. In the pairwise characterization (Section 5), we assume the rough position of the peaks are known. In the inference and normal imaging sections (6 and 7), we use depth interpolated from sparsely attached TDCs as $\hat{\bar{\tau}}$.

Note that the moment matching approach would result in noisy depth gradient estimates, since we are using the mean at a single pixel, without any spatial priors. However, the normal integration techniques we introduce in Section 7 can help alleviate noisy gradients or outliers.

5 CHARACTERIZING THE FAD RESPONSE

In this section, we characterize the raw FAD and model performance under different conditions (signal power, albedo variation, background, and depth gradient). In this section, we use only the simple inversion schemes in the previous section for a single pair of pixels, without introducing any 2D priors.

5.1 Simulation framework

We perform a simulation for a pair of SPAD pixels. For each pixel, the basic transient shape is based on a realistic transient from our MPD SPAD, which contains a single Gaussian peak exponential tail (see supp.). This template is scaled, shifted and added with a temporally uniform background offset to match with the reflectivity, depth difference and background values desired. In a single trial, we simulate measurements for an integration time of 30 ms, with a repetition rate of 40MHz (fixed cycle duration of 25 ns, with active period being 15 ns and a fixed reset time of 10 ns during which no photons are detected). We repeat this trial 100 times and plot the nFAD mean and noise, as well as the mean absolute error of our proposed estimation method.

For a given set of reflectivity and background, we simulate the arrival and differential logic for pairs of pixels. We use a Monte-Carlo simulation to generate photon arrivals (pulse and background) for a number of cycles at each pixel [33]. In each cycle, we use the differential logic (see supp.) to process photon arrivals into FAD counts. The photon counts corresponding to total pulse intensity and background are also collected for intensity estimation.

In all the inversion steps below, the σ value used is 104 ps (same calibrated value as emulation, see supp. for details). The background offset correction is also applied, with the absolute depth estimate $\hat{\tau}$ set to the location of SPAD 1. $\tau_1 = 5000$ ps in this analysis – we choose a location close to the center of the laser period, but the particular choice does not change our conclusions in the analysis.

We also introduce two quantities here:

- **effective signal to background ratio (SBR)**: commonly in SPAD literature, SBR is defined as the ratio between received pulse photons and signal photons for a single pixel [10] [34]. Here, because a pair of pixels is involved, we define the overall SBR as the geometric mean of the individual SBRs at each pixel.

- **reflectivity or albedo ratio**: corresponds to α_2/α_1. It represents the amount of reflectivity variation between a pair of pixel.

5.2 Effect of pulse power

We assume the dark count rate to be 1000cps [35] and no additional ambient light is present. We assume both pixels have the same power (photon rate), and vary them from 3e-3 up to 1e-2 photon/cycle ($SBR > 100$). We also vary the peak difference to range from -200ps to +200ps with a step size of 20ps. The results are shown in Fig. 6. The mean values of the nFAD responses match for all three cases (left). From the error analysis (right), we can see that the depth estimation errors increase as power decreases. The errors also increase as $\Delta\tau$ approaches the saturation points, since the nFAD response starts to flatten and the depth estimation uncertainty increases.

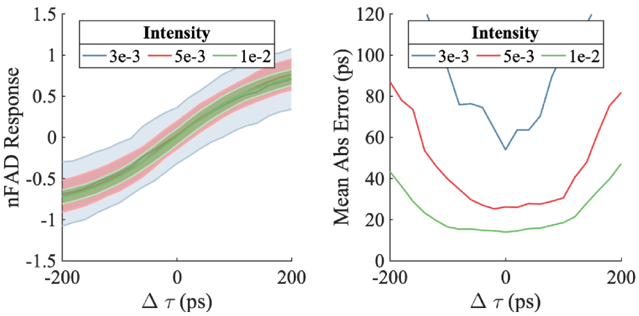


Fig. 6. **Effect of pulse power.** Left: nFAD counts (normalized with intensities) over different depth gradient values (x-axis using round trip time as unit), with solid lines indicating the mean and shaded regions indicating standard deviation. As signal power lowers, the noise in nFAD increases significantly, since there are fewer dual arrivals. Right: we visualize the depth estimation error, by directly applying the per-pixel gradient inversion formula. The overall error increases from 20ps, to up to $\approx$80ps, as we decrease intensity from 1e-2 to 3e-3 photons/cycle . At the point of 3e-3, it becomes difficult to resolve gradient accurately between a single pair.

5.3 Effect of albedo variation

In this section we study the effect of albedo variation. We fix pulse intensity of the SPAD 1 to be 0.01. We vary the albedo ratio, using 0.3, 0.5, 0.8. Here only dark counts are present. The results are shown in Fig. 7.

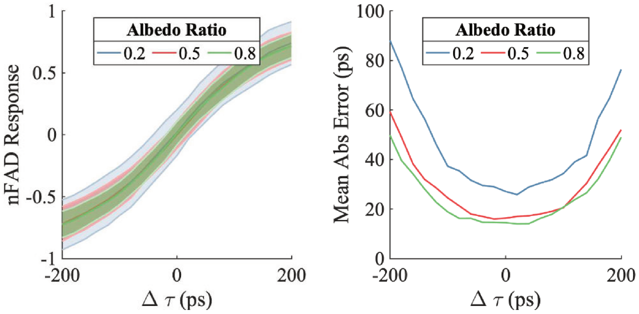


Fig. 7. **Effect of albedo variation under low ambient light (1000cps dark counts only)**. Here α_1 is fixed and vary the albedo ratio α_2/α_1. We can observe that as albedo ratio increases, the variance in nAD and depth inversion error both increases, due to fewer dual arrivals. The albedo variation causes no bias due to low background.

When high background is present, larger albedo variation introduces biases, but are correctable. Please see supp. for this experiment.

5.4 Effect of background

Here we consider the background effect under two scenarios: (1) no albedo variation (both SPADs are 0.01 photon/cycle), and (2) an albedo ratio of 0.5 (0.01 and 0.05 photon/cycle). For each scenario, we consider background cases from dark counts only to a low SBR of 0.1. The results are shown in Fig. 8 and Fig. 9. From Fig. 8, we can observe that under no albedo variation, background arrivals introduce variance without shifting the mean nFAD curves. The estimation errors increase as we decrease SBR. From Fig. 9, we can clearly see a bias or shift introduced to the nFAD response when the correction term is not applied. However, applying corrections will still yield reasonable gradient estimations (See Fig. 9, right).

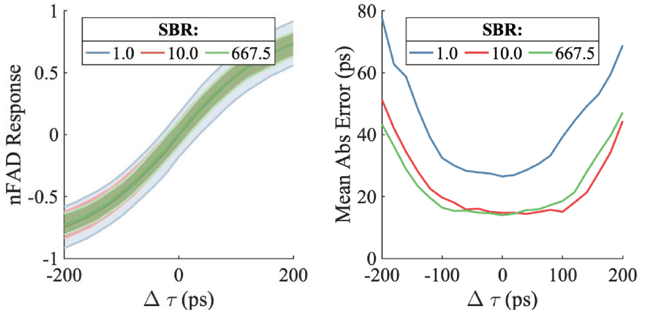


Fig. 8. **Effect of background under no albedo variation**. In this case, there is little distortion of the expected nFAD curve, except that as SBR increases (background lowers), there is less variance. Per-pixel estimation works up to an SBR of around 1, and if SBR decreases further the errors become significant.

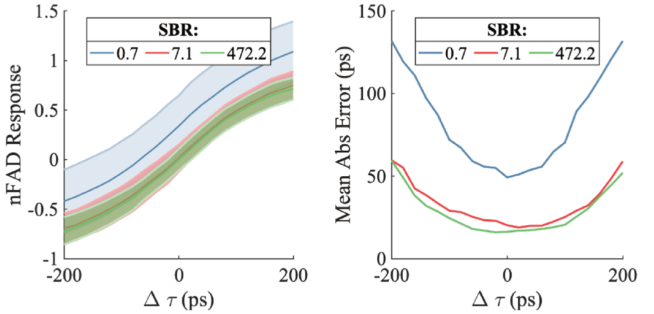


Fig. 9. **Effect of background under high albedo variation**. In this case as background increases, the uncorrected nFAD is clearly biased as background increases (left). On the right side, we show the depth errors after applying the correction. We can still observe an increase in gradient estimation errors as SBR becomes low. With albedo ratio of 0.5, per-pixel estimation works up to an SBR of around 1.

6 3D SCENE INFERENCE

In Section 4, we see how nFAD units provide a convenient way to measure high-resolution depth variation at each pixel (Eq. 21). Here we show how an array of nFAD counts can be exploited for 3D inference tasks such as depth edge detection, object segmentation, and measuring surface normals. Fig. 10 illustrates our reconstruction pipeline and 3D applications on a real-world scene.

Depth Gradients and Surface Normals. The normalization of FAD counts to produce the nFAD can introduce significant noise at low-intensity regions. Thus we start our processing pipeline by denoising nFAD and intensity counts with an edge-preserving filter such as non-local means, before preforming the division. Using Equation 4, we can

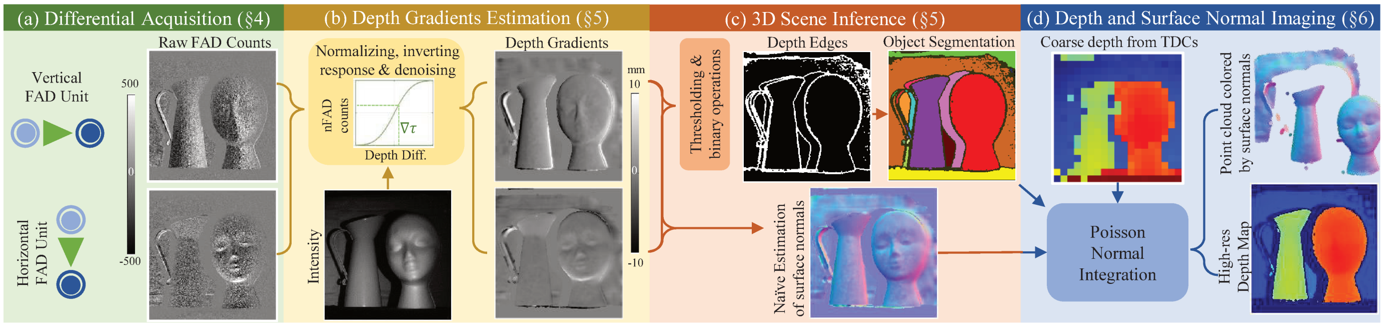


Fig. 10. **Our reconstruction pipeline.** Depth gradients (a) are obtained by normalizing the raw FAD counts (b) and then inverting the nFAD response. By thresholding the depth differences and processing, we detect depth edges in the scene, segment objects, and estimate surface normals (c). With high-res depth gradients, object segmentation, and low-res absolute depth measurements, we obtain high-res depth reconstruction (d).

then obtain a per-pixel raw estimate of the depth gradients. An example of the resulting depth gradients are shown in Fig. 10.

Depth edge detection. After the denoising step, we can simply threshold the absolute value of nFAD counts (Fig. 10(b)), similar to the per-pixel case. Depth edges close to hard intensity shadows can sometimes lead to spurious depth edges being detected, due to the sparseness of dual arrivals.

Segmentation. The depth edges correspond to object contours. We fill the gaps in the object contours by performing standard morphological operations on the binary contour map. We then find the connected components in this binary map using MATLAB's `bwconncomp` function. Each connected component should approximately correspond to a different object in the scene (Fig. 3(d)).

7 DEPTH AND SURFACE NORMAL IMAGING

While the local nFAD measurements enable us to infer local surface normals and depth edges at a high resolution, these counts are agnostic to the absolute depth of the objects. For estimating the absolute depth, we need to also rely on sparse pixels with TDCs that capture absolute depth at sparse locations. Using this, we can recover a high-resolution depth map of the scene (Fig. 10(d)). An overview of our reconstruction algorithm follows. Please refer to the supplementary material for implementation details.

Segmentation-aided normal integration We can obtain surface normal maps solely from the nFAD measurements (Section 6). If the scene does not contain depth discontinuities, the surface can be reconstructed, up to an unknown depth offset, by integrating the normal map [32]. But for realistic scenes involving depth discontinuities such as in Fig. 10, directly integrating the normal map would fail at the sharp edge discontinuities. The FAD units also provide an object segmentation map (Fig. 14) by detecting these discontinuities. We use this segmentation map to create masked normal maps for each object and integrate each of these normal maps. Here we choose use the FFT-based Poisson normal integration by Frankot and Chellappa [36], as it achieves a decent balance between speed and robustness, and requires no parameter tuning.

Absolute depth offsets from sparse TDCs The above step results in depth maps of the objects up to a constant offset. We then use least squares to fit this depth offset per object to match the depth values obtained from the sparse TDCs. The depth map of the scene is computed as the composition of depth maps of each segmented object (Fig. 10(d)).

8 RESULTS FROM SIMULATED DATA

To quantify errors for objects with a larger scale, we simulate FAD measurements for a scene containing a bunny object spanning around 1mx1m area (300x300 resolution). We reconstruct the depth and compute the angular errors of the resulting surface normals.

We use the MitsubaTOF renderer [37] to obtain transients for a particular scene, with collocated source and detector. We sample ∼25k time bins with 1 ps resolution. To speed up the Monte-Carlo simulation, we take the dominant peak per transient and convolve with a Gaussian IRF to include jitter effects. In this experiment, instead of a full jitter with long trailing tails, we adopt a Gaussian IRF function with 150 ps FWHM to reduce time bins. We simulated measurements for 30 ms with a 40 MHz repetition rate. In addition, we assume a spatially uniform background corrupting the FAD measurements.

Fig. 11 shows surface reconstruction errors for a base case as well as additional cases where (1) lower power is used, and (2) higher background is present. When computing the average errors, we exclude the saturated depth edges and background. We obtain quite accurate surface normal estimates along continuous surfaces.

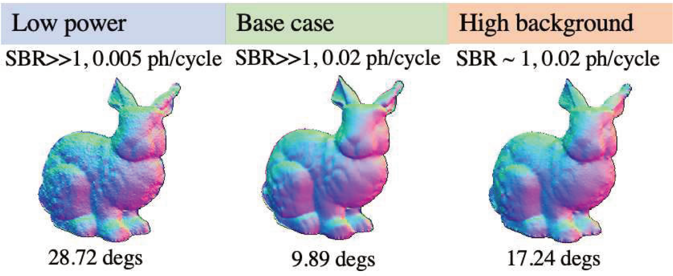


Fig. 11. **Surface reconstruction results for rabbit object**. From left to right are: low power case, base case and high background case. When average photon rate falls below 5e-3 the reconstruction starts failing. Right: reconstruction with non-trivial background. Here the background is set to be uniform, and equals to the max photon rate from the pulse returns. Our bias correction is successful even under such non-trivial background value. We also compare the mean angle error in surface normals compared to the ground truth.

9 RESULTS FROM REAL DATA

9.1 Hardware Emulation of flash LiDAR

To experimentally emulate the chip, we built a scanning hardware prototype shown in Fig. 12. We collect photon arrival time tags per cycle, using a Hydroharp and MPD FastGatedSPAD. The ToF range is 12500 ps and the bin resolution is 1 ps. The detector is scanned using a galvo mirror at a resolution of 384 × 512 with a pitch of 0.8 mm. The scene is lit using a fiber coupled laser passed through a diffuser (inset of Fig. 12). The laser source is a Chameleon Discovery operated at 480 nm wavelength and 80 MHz rep. rate (12.5 ns period). The average photon rate per pulse is 0.01 photons per cycle. We then compute the FAD counts from time tags based on the model introduced in Section 3.2. See supplementary for details .

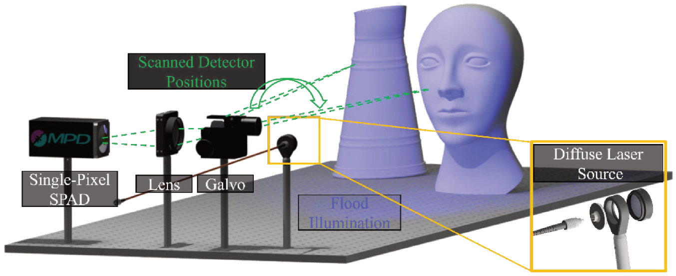


Fig. 12. **Experimental setup used for chip emulation.** A transient detector is focused onto the scene with a lens, and scanned using a galvo mirror. A diffuse light source is created by passing the output of a fiber optic through a ground glass diffuser.

9.2 Characterization with simple structures

First, we quantify the resolution of depth edge detection and surface normal reconstruction. As shown in the top row of Fig. 13, we scan a single line across (1) a staircase object with increasing depth gaps, (2) a plane tilted 45° away from the optical axis, and (3) a sphere (center line with maximum diameter). A long integration time of ~300 ms is used per pixel location. nFAD counts are computed along the lines. The third and fourth rows in Fig. 13 show the depth gradients and depth value estimations. The gradients here are directly estimated using the per-pixel inversion equation, and depth is directly integrated from gradient values (without Poisson integration). In this example, we can reconstruct depth differences around a few mm.

9.3 3D Scene Inference

In Fig. 14 we show 3D scene inference results. Columns (b)-(c) show depth gradients reconstructed from nFAD measurements. The objects shown in the scene are around 10–20 cm in scale and span a depth of 0.7–1.2 m. The reconstructed gradients capture fine structures such as the ribs on the pumpkin, nose of the pig and hands of the figurine. With gradients, we can directly take the cross product and produce high quality surface normals as shown in column (d). Notice that the FAD approach can even capture fine wrinkles on the background cloth. (e) shows the depth edge masks, which allows us to produce clean segmentation results for both scenes in (f). We have only used relative depth information extracted from nFAD, assuming no TDCs.

Comparison Baselines. Table 1 lists the specifications of the full single-pixel scanning and conventional LiDAR baselines, where the net throughput is kept fixed. The specifications are chosen based on state-of-the-art LiDAR works [18], [19], [20]. We emulate coarser spatial resolution by spatially downsampling the high resolution depth map from our scanning-based setup. The TDC ranges and bin width is achieved by cropping the range and quantizing the ground truth depths respectively. We also perform spatio-temporal smoothing and filtering to improve the quality and visibility of baseline data. Fig. 15 lists emulated results.

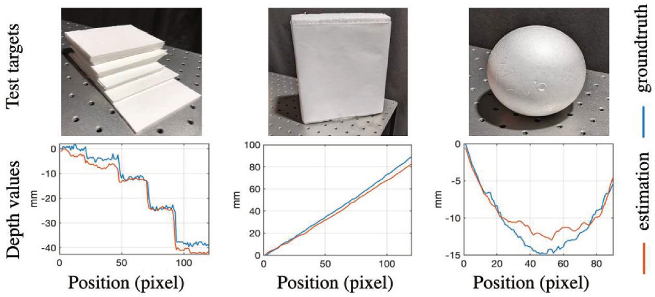


Fig. 13. **Quantative evaluation on simple structures.** The top row shows the targets and bottom row shows the ground truth and estimated depth values for a 1D scan.

Table 1. **Specifications of emulated designs**. Ground truth (GT) denotes the high spatio-temporal resolution obtained by scanning LiDAR. The conventional LiDAR baselines (B1, B2, B3) and our FAD LiDAR are designed with the same net throughput (net TP). B1, B2, B3 sacrifice depth resolution, depth range, and spatial resolution respectively to match this throughput. Our design has 256x higher resolution for the same net TP without sacrificing on any other factor.

	TDC Resolution	Bin Width (ps)	TDC Range (ns)	Net TP (Gb/s)
GT	384 × 512	1	12	10657
B1	96 × 128	1200	12	163.3
B2	96 × 128	50	0.5	163.3
B3	48 × 64	1	12	163.3
Ours	384 × 512	1	12	163.3

Imaging results Fig.15 demonstrates that the depth imaging quality of our system clearly improves upon conventional baselines. FAD LiDAR preserves fine structures and provides a large depth range. This improved depth resolution and range is achieved along with a much smaller data throughput rate (50-100x reduction). Baseline B1 (first column) sacrifices timing resolution for range and spatial resolution. However, time is sampled so coarsely that finer structures cannot be inferred. Baseline B2 maintains the same spatial resolution but chooses smaller bins over range, and the depth range stops at the middle of the object. Here locations out of the TDC measurement range are set to have 0 depth. The background cloth is completely missed. In B3, finer bins and long-range are chosen over spatial resolution. Small features such as the arms of the ghost are completely lost. In contrast, our FAD-LiDAR is able to achieve high spatial resolution, temporal resolution, and range, thanks to the differential nature and compactness of the FAD circuit.

10 LIMITATIONS AND DISCUSSIONS

Sparse dual arrivals due to low reflectivity. Since a dual arrival detection requires both pixels to respond in a cycle, low reflectivity regions can result in fewer dual arrivals. One way to mitigate this is by performing differentials across groups of pixels. Another option is to use lasers with higher peak powers. Unlike typical TCSPC, FAD LiDAR

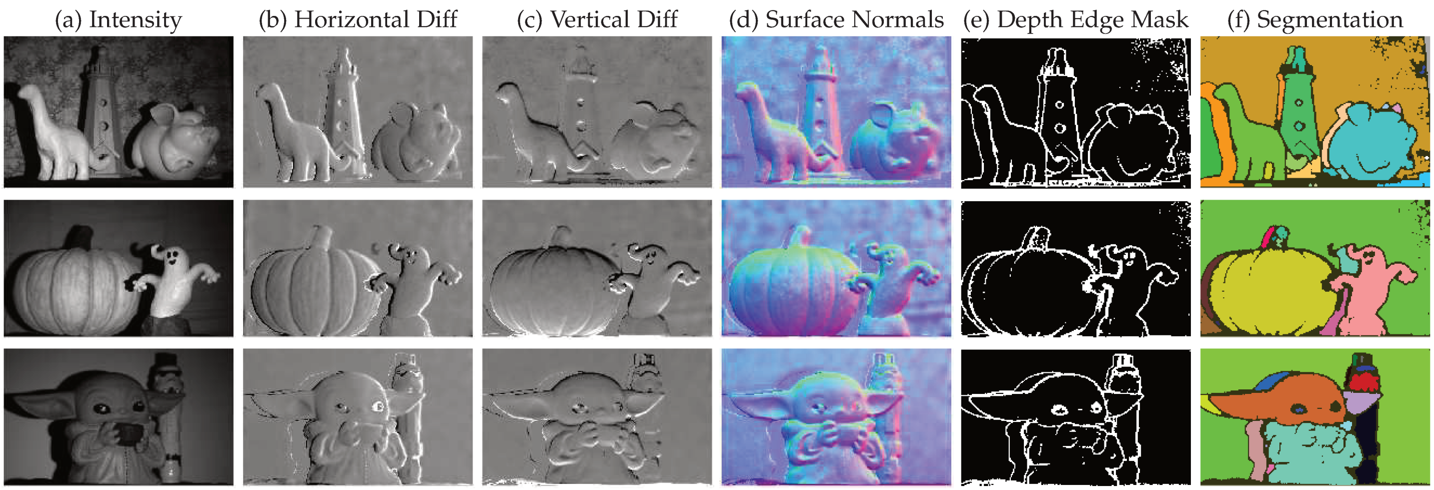


Fig. 14. **3D Scene Inference** from FAD-SPAD array on real-world scenes. From the FAD unit measurements, we estimate the depth differences (b) and (c). These differences enable inferring the surface normal orientation of the objects (d). The saturation of FAD response corresponds to the depth edges (e) which can be used to segment objects in the scene (f).

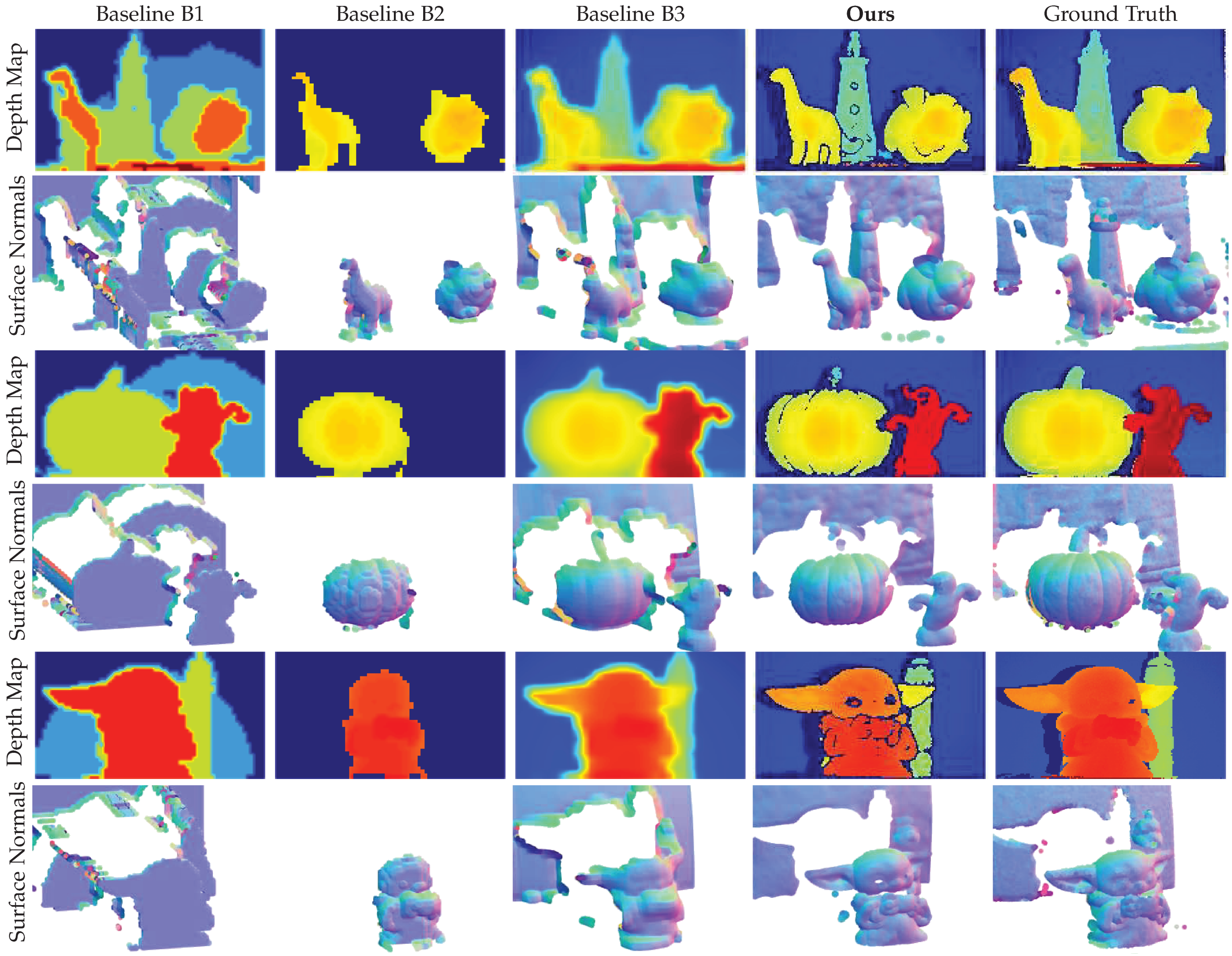


Fig. 15. **High resolution 3D imaging and surface normals by differential LiDAR** emulated using our hardware setup. Reconstruction from a scanning-based LiDAR design is denoted as Ground Truth. Conventional flash LiDAR designs B1, B2, and B3 suffer from performace tradeoffs resulting in either poor depth resolution (B1), low depth range (B2) or poor spatial resolution (B3). Our differential flash design offers significantly better reconstruction quality for the same data throughput as conventional baselines.

has a dependence on laser pulse energy in addition to average laser power. For the same average laser power, increasing pulse energy by M and simultaneously reducing the pulse repetition rate by M, results in M times more dual arrivals, and consequently better performance. This is an additional degree of freedom that needs to be considered and optimized while setting up a FAD-LIDAR system.

Algorithm Improvements. Currently we adopt a sequential reconstruction pipeline that involves mostly traditional processing methods. We expect the performance to improve with the integration of end-to-end or learning-based methods.

Non-linear distortions and multiple reflections. The current model and reconstruction do not account for multiple reflections. In our simulated scenes and experimental captures, however, the first order returns seem to dominate and the presence of higher orders does not degrade visual quality. In cases where there is significant multi-bounce effects, mitigation techniques such as gating [38] or an improved model that incorporates multiple returns will be needed, which we leave to future work.

11 CONCLUSIONS

We have demonstrated that the local difference of first-arrival photon counts carries enough information to reconstruct a scene without requiring full transient information, and that this can be computed in-pixel with minimal supporting circuitry and extremely small silicon footprint. Our theoretical models, simulations, and emulations indicate that the FAD unit approach could provide significantly improved spatial resolution and reduced data throughput in SPAD-based LiDAR arrays, while maintaining temporal resolution and range. Moreover, this work opens up a new space of design around the differential connections between pixels. Here we have detailed one such possibility–connecting each SPAD to its nearest neighbors–but many other schemes can be built with the same principle, and may offer different advantages depending on the application.

ACKNOWLEDGMENTS

This work is supported by NSF Expeditions (CCF-1730574), NSF CAREER Award (IIS-1652633), MTEC grant (MTEC-21-06-MPAI-077) and DARPA grant (N66001-19-C-4020). Additionally, we want to thank Prof. Richard Baranuik and Adithya Pediredla for very helpful discussions on the project.

REFERENCES

[1] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 1907–1915.

[2] C. Steger, M. Ulrich, and C. Wiedemann, *Machine vision algorithms and applications*. John Wiley & Sons, 2018.

[3] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "Bethere: 3d mobile collaboration with spatial input," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 179–188.

[4] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of quantum electronics*, vol. 37, no. 3, pp. 390–397, 2001.

[5] G. Adamo and A. Busacca, "Time of flight measurements via two lidar systems with sipm and apd," in *2016 AEIT International Annual Conference (AEIT)*. IEEE, 2016, pp. 1–5.

[6] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Applied optics*, vol. 35, no. 12, pp. 1956–1976, 1996.

[7] F. Villa, F. Severini, F. Madonini, and F. Zappa, "SPADs and SiPMs arrays for long-range high-speed light detection and ranging (LiDAR)," *Sensors*, vol. 21, no. 11, p. 3839, 2021.

[8] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-photon three-dimensional imaging at up to 10 kilometers range," *Optics express*, vol. 25, no. 10, pp. 11 919–11 931, 2017.

[9] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, "Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," *Nature Comm.*, vol. 10, no. 1, pp. 1–6, 2019.

[10] D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. Wong, and J. H. Shapiro, "Photon-efficient imaging with a single-photon camera," *Nature communications*, vol. 7, no. 1, pp. 1–8, 2016.

[11] Z.-P. Li, J.-T. Ye, X. Huang, P.-Y. Jiang, Y. Cao, Y. Hong, C. Yu, J. Zhang, Q. Zhang, C.-Z. Peng *et al.*, "Single-photon imaging over 200 km," *Optica*, vol. 8, no. 3, pp. 344–349, 2021.

[12] Velodyne, "Hdl-32e." [Online]. Available: https://velodynelidar.com/products/hdl-32e/

[13] Neuviton, "Titan m1-pro." [Online]. Available: https://www.neuvition.com/products/titan-m1-pro.html

[14] C. Zhang, S. Lindner, I. M. Antolović, J. M. Pavia, M. Wolf, and E. Charbon, "A 30-frames/s, 252×144 SPAD flash LiDAR with 1728 dual-clock 48.8-ps tdcs, and pixel-wise integrated histogramming," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1137–1151, 2018.

[15] J. SzyduczyÅski, D. KoÅcielnik, and M. Miskowicz, "A successive approximation time-to-digital converter with single set of delay lines for time interval measurements," *Sensors*, vol. 19, p. 1109, 03 2019.

[16] A. T. Erdogan, R. Walker, N. Finlayson, N. KrstajiÄ, G. Williams, J. Girkin, and R. Henderson, "A CMOS SPAD line sensor with per-pixel histogramming tdc for time-resolved multispectral imaging," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 6, pp. 1705–1719, 2019.

[17] C. Bruschini, H. Homulle, I. M. Antolovic, S. Burri, and E. Charbon, "Single-photon avalanche diode imagers in biophotonics: review and outlook," *Light: Science & Applications*, vol. 8, no. 1, pp. 1–28, 2019.

[18] L. Pancheri, N. Massari, F. Borghetti, and D. Stoppa, "A 32x32 spad pixel array with nanosecond gating and analog readout," in *International Image Sensor Workshop (IISW)*, 2011.

[19] M. Perenzoni, D. Perenzoni, and D. Stoppa, "A 64x64-pixels digital silicon photomultiplier direct tof sensor with 100-mphotons/s/pixel background rejection and imaging/altimeter mode with 0.14% precision up to 6 km for spacecraft navigation and landing," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 1, pp. 151–160, 2016.

[20] O. Kumagai, J. Ohmachi, M. Matsumura, S. Yagi, K. Tayu, K. Amagawa, T. Matsukawa, O. Ozawa, D. Hirono, Y. Shinozuka *et al.*, "A 189x600 back-illuminated stacked spad direct time-of-flight depth sensor for automotive lidar systems," in *2021 IEEE International Solid-State Circuits Conference*, vol. 64, 2021, pp. 110–112.

[21] A. Moini, "Vision chips or seeing silicon," *Center for High Performance Integrated Technologies and Systems, University of Adelaide, Adelaide, Australia, Technical Report*, 1997.

[22] J. N. Martel, L. K. Mueller, S. J. Carey, P. Dudek, and G. Wetzstein, "Neural sensors: Learning pixel exposures for HDR imaging and video compressive sensing with programmable sensors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 7, pp. 1642–1653, 2020.

[23] F. M. Della Rocca, H. Mai, S. W. Hutchings, T. Al Abbas, K. Buckbee, A. Tsiamis, P. Lomax, I. Gyongy, N. A. Dutton, and R. K. Henderson, "A 128×128 SPAD motion-triggered time-of-flight image sensor with in-pixel histogram and column-parallel vision processor," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 7, pp. 1762–1775, 2020.

[24] R. J. Walker, J. A. Richardson, and R. K. Henderson, "A 128×96 pixel event-driven phase-domain $\delta\sigma$-based fully digital 3d camera in 0.13 μm cmos imaging technology," in *2011 IEEE International Solid-State Circuits Conference*. IEEE, 2011, pp. 410–412.

[25] G. M. Martín, A. Turpin, A. Ruget, A. Halimi, R. Henderson, J. Leach, and I. Gyongy, "High-speed object detection using SPAD

sensors," in *Photonic Instrumentation Engineering VIII*, vol. 11693. International Society for Optics and Photonics, 2021, p. 116930L.

[26] M. P. Sheehan, J. Tachella, and M. E. Davies, "A sketching framework for reduced data transfer in photon counting lidar," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 989–1004, 2021.

[27] A. Bhandari, M. H. Conde, and O. Loffeld, "One-bit time-resolved imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1630–1641, 2020.

[28] A. Ingle, T. Seets, M. Buttafava, S. Gupta, A. Tosi, M. Gupta, and A. Velten, "Passive inter-photon imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8585–8595.

[29] J. Tumblin, A. Agrawal, and R. Raskar, "Why i want a gradient camera," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 103–110.

[30] M. White, S. Ghajari, T. Zhang, A. Dave, A. Veeraraghavan, and A. Molnar, "A differential spad array in 0.18μm cmos for hdr imaging," *International Symposium on Circuits and Systems*, June 2022.

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

[32] Y. Quéau, J.-D. Durou, and J.-F. Aujol, "Normal integration: a survey," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 4, pp. 576–593, 2018.

[33] J. Rapp, Y. Ma, R. M. Dawson, and V. K. Goyal, "High-flux single-photon lidar," *Optica*, vol. 8, no. 1, pp. 30–39, 2021.

[34] A. Gupta, A. Ingle, A. Velten, and M. Gupta, "Photon-flooded single-photon 3d cameras," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6770–6779.

[35] C. Zhang, "CMOS SPAD sensors for 3d time-of-flight imaging, LiDAR and ultra-high speed cameras," 2019.

[36] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 10, no. 4, pp. 439–451, 1988.

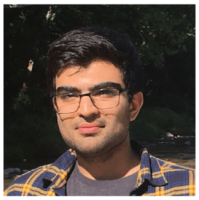
[37] A. Pediredla, A. Veeraraghavan, and I. Gkioulekas, "Ellipsoidal path connections for time-gated rendering," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–12, 2019.

[38] K. Morimoto, A. Ardelean, M.-L. Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini, and E. Charbon, "Megapixel time-gated SPAD image sensor for 2d and 3d imaging applications," *Optica*, vol. 7, no. 4, pp. 346–354, 2020.

Tianyi Zhang is currently a Ph.D student at Rice University, where she also received her M.S. degree in 2022, B.S. degree (Hons.) in 2019 in electrical engineering. Her research interests including exotic image sensors, SPADs, quanta image sensors, computational imaging and computer vision.

Mel White is a Postdoctoral Associate at Cornell University, where she also completed her PhD. Her doctoral work was on integrated circuit design for image sensors, and her current research extends to compressive and lensless techniques for computational imaging systems. She holds a BS in Electrical and Computer Engineering from RPI and a BA in International Relations from William & Mary.

Akshat Dave is currently a Ph.D. student at Rice University. He received his Bachelors and Masters degree in Electrical Engineering from Indian Institute of Technology Madras in 2017. His research interests include time-of-flight, polarization, computational imaging, inverse rendering and neural rendering.



Shahaboddin Ghajari received the B.S. degree from the Ferdowsi University of Mashhad, and the M.S. degree in electrical engineering from the Sharif University of Technology in 2015 and 2017, respectively. He is currently a Ph.D. candidate in the ECE department with the Molnar Group at Cornell University, and his research interests include circuit design for biomedical and imaging applications.

Ankit Raghuram received his B.S. degree in Biomedical Engineering from the Georgia Institute of Technology in 2015 and his Ph.D degree in ECE from Rice University in 2022.He is currently a Research Engineer at Rice University, and his interests include biomedical imaging, microscopy, diffuse imaging, time-of-flight, and computational imaging.

Alyosha Christopher Molnar received the B.S. degree (Hons.) in engineering from the Swarthmore College in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley in 2003 and 2007, respectively. From 1998 to 2002, he was with the RF IC Group, Conexant Systems, Inc. (now Skyworks Solutions, Inc.), Newport Beach, CA. He joined the Faculty at Cornell University in 2007, where he is currently an Associate Professor and focused on software-defined radios, neural interface circuits, integrated imaging techniques, and ultralow-power sensing.

Ashok Veeraraghavan received the B.S. degree in electrical engineering from the Indian Institute of Technology, in 2002, and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Maryland, in 2004 and 2008, respectively. He is currently an Associate Professor of electrical and computer engineering, Rice University. Before joining Rice University, he spent three years as a Research Scientist with Mitsubishi Electric Research Labs, Cambridge, MA, USA. His research interests include computational imaging, computer vision, machine learning, and robotics.